Application of next-generation sequencing to analysis of the genetic basis of complex traits in plants, with particular focus on nickel hyperaccumulation in the *Alyssum serpyllifolium* species complex

Maria Kinga Sobczyk

Linacre College
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Trinity Term 2015

Abstract

Reliable, high-throughput and low-cost next-generation sequencing technologies have invigorated genetic research into non-model organisms over the last decade. In this work, RNA-Seq was employed to obtain the first-ever transcriptomes of two groups of closely related plant taxa possessing distinctive complex physiological traits, namely metal hyperaccumulation and C_4 photosynthesis.

Metal hyperaccumulator plants possess an extraordinary ability to take up trace elements from the soil and accumulate them to high concentrations in their shoots, probably to serve as a type of elemental defence against natural enemies. Taxonomically, the most common form of metal hyperaccumulation, nickel hyperaccumulation, is encountered on nickel-rich ultramafic (serpentine) soils, and is found with the highest frequency (ca. 51 species) in the genus *Alyssum* (family Brassicaceae). Here, the genetic basis and evolutionary history of nickel tolerance and hyperaccumulation was investigated in the *Alyssum serpyllifolium* Desf. species complex, which contains both serpentine and non-serpentine populations of unresolved phylogenetic relationships on the Iberian Peninsula. Genome scans for outlier loci and differential expression analyses identified a number of candidate hyperaccumulator genes common to two serpentine populations found in Portugal and Spain, but the majority of adaptive variation was of local origin. Phylogenetic and population-genetic inferences based on neutral and putatively adaptive loci suggested that the key genes for the nickel hyperaccumulation trait evolved once and spread across serpentine populations early in the history of this species, with no genetic isolation but continued recent gene flow between serpentine and non-serpentine populations.

To test the power of next-generation sequencing for analysing the genetic basis of a separate complex trait, a cross-species comparison was performed using RNA-Seq of two congeneric tropical species, the C_4 plant *Alternanthera pungens* Kunth and the C_3 plant *Alternanthera philoxeroides* (Mart.) Griseb. f. *angustifolia* Suess. (family Amaranthaceae). These species were cultivated at two different temperatures and showed significant differences in levels of overall gene expression plasticity and isoform switching in certain photosynthesis genes, which it is proposed may explain the observed difference in the ability of these two species to acclimate to low and high growth temperatures.

Application of next-generation sequencing to analysis of the genetic basis of complex traits in plants, with particular focus on nickel hyperaccumulation in the *Alyssum serpyllifolium* species complex

Maria Kinga Sobczyk

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2015

Acknowledgements

I would like to thank my main thesis supervisor Professor JAC Smith for his constant support, incredible incisiveness and enthusiasm throughout the project. I have learned a lot and tremendously enjoyed being part of the Smith lab. My thanks also go to Dr Dmitry Filatov, my second supervisor, for offering insightful comments and computing facilities along with funding for an additional RNA-Seq run. I am especially thankful to other current and former members of the Department of Plant Sciences (in alphabetical order): Pedro Bota for introducing me to GS-MS, Dr Mark Chapman for introducing me to microsatellite genotyping, Dr José Escarré (CNRS, Montpellier) for help in locating French populations of *Alyssum serpyllifolium*, Dr Tom Flynn for briefing me on his *Alyssum* thesis and phylogenetic analysis, Dr Maxim Kapralov for sharing his raw Caryophyllales RNA-Seq data with me, Dr Steve Kelly for informal bioinformatics help during the initial part of the project, and Lynne Richardson (Zoology) for technical advice on sequencing. My time as a graduate student in Oxford would also not be the same without the company of fellow DPhil students at my home department, members of the Linacre College Common Room and friends from other colleges.

The work undertaken in the thesis was financially supported by the Biotechnology and Biological Sciences Research Council studentship award to MK Sobczyk and Natural Environment Research Council NBAF small project grant award to JAC Smith as well as The Genetics' Society Heredity Fieldwork Grant award to MK Sobczyk. This contribution is gratefully acknowledged.

Abstract

Reliable, high-throughput and low-cost next-generation sequencing technologies have invigorated genetic research into non-model organisms over the last decade. In this work, RNA-Seq was employed to obtain the first-ever transcriptomes of two groups of closely related plant taxa possessing distinctive complex physiological traits, namely metal hyperaccumulation and C_4 photosynthesis.

Metal hyperaccumulator plants possess an extraordinary ability to take up trace elements from the soil and accumulate them to high concentrations in their shoots, probably to serve as a type of elemental defence against natural enemies. Taxonomically, the most common form of metal hyperaccumulation, nickel hyperaccumulation, is encountered on nickel-rich ultramafic (serpentine) soils, and is found with the highest frequency (ca. 51 species) in the genus *Alyssum* (family Brassicaceae). Here, the genetic basis and evolutionary history of nickel tolerance and hyperaccumulation was investigated in the *Alyssum serpyllifolium* Desf. species complex, which contains both serpentine and non-serpentine populations of unresolved phylogenetic relationships on the Iberian Peninsula. Genome scans for outlier loci and differential expression analyses identified a number of candidate hyperaccumulator genes common to two serpentine populations found in Portugal and Spain, but the majority of adaptive variation was of local origin. Phylogenetic and population-genetic inferences based on neutral and putatively adaptive loci suggested that the key genes for the nickel hyperaccumulation trait evolved once and spread across serpentine populations early in the history of this species, with no genetic isolation but continued recent gene flow between serpentine and non-serpentine populations.

To test the power of next-generation sequencing for analysing the genetic basis of a separate complex trait, a cross-species comparison was performed using RNA-Seq of two congeneric tropical species, the C_4 plant *Alternanthera pungens* Kunth and the C_3 plant *Alternanthera philoxeroides* (Mart.) Griseb. f. *angustifolia* Suess. (family Amaranthaceae). These species were cultivated at two different temperatures and showed significant differences in levels of overall gene expression plasticity and isoform switching in certain photosynthesis genes, which it is proposed may explain the observed difference in the ability of these two species to acclimate to low and high growth temperatures.

Contents

Chapter 1

Introduction

1.1 Overview of next-generation sequencing methods

1.1.1 Advent of high-throughput, low-cost DNA sequencing technologies

High-throughput sequencing (HTS), also known as next-generation sequencing (NGS) technologies, such as Illumina sequencing or 454 pyrosequencing, have enabled researchers to generate short sequence reads (typically 100-300 bp but up to 14,000 bp) in a still growing range of contexts, such as DNA-Seq, RNA-Seq, ChiP-Seq, Bisulphite-Seq (Fonseca *et al.*, 2012) at an unprecedented scale and low cost over the past 5 years (Wang *et al.*, 2009**a**; Metzker, 2010; Oshlack *et al.*, 2010). As such, these methods have often displaced older techniques, such as 'first generation' Sanger sequencing in certain contexts (Metzker, 2010). For instance, RNA-Seq has largely replaced DNA microarrays for gene expression profiling (Finotello & DiCamillo, 2014) and has become the dominant technology for the study of gene expression (Wang *et al.*, 2009**c**). It has been shown that well-designed RNA-Seq and microarray experiments show good concordance (Nookaew *et al.*, 2012), and RNA-Seq results from technical and biological replicates are highly reproducible and correlate well with qPCR results (Cloonan *et al.*, 2008; Nagalakshmi *et al.*, 2008).

RNA-Seq protocols start with reverse transcription of total RNA to cDNA, product amplification, shearing, fragment selection, followed by steps specific to each of the RNA-Seq technologies (Illumina, Roche 454, Pacific Biosciences etc.). In the case of the most popular method, Illumina, also adopted here, adaptors are then ligated at the ends of

fragments, which allows attachment to the flow cell surface and bridge-amplification in solid phase (Marguerat & Bähler, 2010; Metzker, 2010). Clusters containing the same fragment cycle through four-colour cyclic reversible termination, whereby one of the four nucleotides with attached fluorophore is added to the flow cell, followed by a washing step and imagining of clusters where the given nucleotide finds a complementary nucleotide, cleavage of the terminating group and the dye, and finally re-start of the cycle.
Reads produced in this cyclic procedure from the application of a base-calling algorithm to the signal generated by excited fluorophores are then mapped either to reference sequences available or generated *de novo* from the reads themselves. The resulting mapped read counts provide a digital measure of gene expression in the sample. Read counts can then be used as input in differential gene expression studies aiming to characterize organismal, tissue, or even cellular response to different extrinsic or intrinsic factors. Crucially, RNA-Seq provides a window into both gene expression phenotype and underlying gene sequences at the same time, enabling the latter to be utilized concomitantly for population genetics studies.

1.1.2 Application of RNA-seq to research in non-model organisms: gene expression profiling and population genetics

The advent of next-generation sequencing methods has also allowed researchers to tackle reconstruction of the transcriptomes of non-model plant (and animal) species with little-to-no prior sequence information (Gowik *et al.*, 2011; Ness *et al.*, 2011; Su *et al.*, 2011; Huang *et al.*, 2012; Sloan *et al.*, 2012). This is evident in the publication record of projects, such as the *1000 Plants* (1 KP, www.onekp.com) initiative, which have sequenced hundreds of non-model plant species. Lack of a requirement for prior knowledge of the transcript sequence is a huge advantage compared to previously widespread microarrays, where expression quantification was performed by hybridization with pre-defined gene probes and so the transcript sequence in a given species or a close relative had to be known from previous research (Garber *et al.*, 2011). Furthermore, gene probes used in microarrays span only a small fragment of the entire transcript, whereas in RNA-Seq the sequence and expression of the entire transcript are potentially obtainable from scratch (Bullard

et al., 2010; Finotello & DiCamillo, 2014). The short length of the microarray probes can frequently result in cross-hybridisation of homologous sequences. Nevertheless, there is an analogous problem with RNA-Seq, where challenge remains in assigning reads to closely related paralog and transcript isoforms. The presence of background noise and saturation, which are unique consequences of the microarray technology, lead to much lower dynamic range than in RNA-Seq, whose dynamic range can span five orders of magnitude (Mortazavi *et al.*, 2008). While RNA-Seq allows single base pair resolution and control over which reads contribute to quantification of expression of any given gene, microarrays return only signal intensities, which are less quantitative in nature (Marguerat & Bähler, 2010). Lastly, results from different microarrays are not directly comparable and complex normalization is involved prior to re-analysis, which is not usually the case in RNA-Seq experiments.

RNA-Seq certainly possesses potential pitfalls of its own, such as biases introduced at different stages of sample preparation and inherent to the sequencing platform. As to the first point, depending on where primers used for reverse transcription of mRNA bind, we expect changes to the representation of different transcript ends in the sequencing output. Random hexamer primers will result in the over-abundance of 5′- ends, while poly-dT result oligomers in over-abundance of 3′- ends (Nagalakshmi *et al.*, 2008). Sequencing depth can influence estimates of gene expression and the power to detect differential expression (Robinson *et al.*, 2010). Highly expressed genes will be sampled and sequenced disproportionately more often than the genes showing lower levels of expression given limited number of sequence reads generated and the effect is even exacerbated for genes producing long transcripts. As to the second point, different sequencing platform show specific error profiles, including basepair miscalling (all platforms, typically at 0.1% to 0.01% rates in Illumina; Yang *et al.*, 2010**b**), artificial introduction of tracts of homopolymers (454 pyrosequencing), bias against regions with extreme base composition, such as GC-rich regions (Illumina) (Aird *et al.*, 2011) etc. Fortunately, effective *post hoc* methods have been developed to correct for most of the RNA-Seq errors and biases listed above (Oshlack *et al.*, 2010; Finotello & DiCamillo, 2014) and their use will be described later in analyzing the

experimental results obtained.

RNA-Seq is a low-cost technique resulting in huge reduction in the sequence complexity as compared to genome sequencing but at the same time allowing genome-wide sampling of genetic variation. Importantly, this genetic variation is limited to functional regions and as such, it benefits from low indel frequency and predictable codon structure. For that reason, RNA-Seq can be used to survey the genetic basis of adaptation, using such approaches as genome-wide scans for selection. Genome-wide scans, which involve comparison of between-population and within-population variation based on SNPs at each locus to look for signatures of selection and identification of possibly adaptive loci, have only become possible in the last few years due to the advent of NGS. Previously used markers in non-model organisms, such as AFLP (Amplified Fragment Length Polymorphism), RFLP (Restriction Fragment Length Polymorphism) and microsatellites, suffered from many drawbacks. They were time-consuming to generate, with anonymous positions in the genome, preventing any functional inference about potentially adaptive loci (Davey *et al.*, 2011). Conversely, marker production using NGS can result in parallel scoring at millions of markers in an unlimited number of individuals following a single sequencing experiment.

RNA-Seq technologies nevertheless possess certain disadvantages. High quality sequence data is biased only towards genes expressed at moderate-to-high levels in all samples. Further complications arise from allele-specific expression, alternative splicing, paralogous loci and stochastic drop-out of alleles when sequencing at low-coverage, which can make it difficult to establish genotypes (Nielsen *et al.*, 2011; De Wit *et al.*, 2015). However, at least one potential drawback of using transcriptomes for calculating allele frequency differences, such as required for genome scans - namely the bias introduced by allele-specific expression in different ecotypes - could in fact point to real underlying sequence differences in the gene regulatory region (Westram *et al.*, 2014).

1.2 Introducing metal hyperaccumulation in plants

1.2.1 Metal hyperaccumulation – definition, distribution in space and on the tree of life

One of the most extraordinary adaptations known in the plant kingdom is the ability of certain plants to hyperaccumulate trace elements in their above-ground biomass. This trait is present in only about 500 species, representing approximately 0.2% of all angiosperm species (Plaza *et al.*, 2015). In contrast to metal excluders, whose strategy is to control the uptake of metals into the root and prevent metal translocation to aerial organs, hyperaccumulators accumulate metals in the shoot to levels toxic to other plants (Baker & Brooks, 1989; Rascio & Navari-Izzo, 2011). This is remarkable since the photosynthetic apparatus is one of the major targets of transition metal phytotoxicity typically resulting in severe symptoms such as chlorosis and necrosis, wilting, reduced development and growth (Pandey & Sharma, 2002; Rahman *et al.*, 2005). These general signs of poor plant health are a product of numerous harmful interactions at the cellular level (Haydon & Cobbett, 2007), including: non-specific binding of metals to enzyme functional groups and displacement of other metals from their binding sites, generation of reactive oxygen species by redox-active metals, which can lead to disruption of the electron-transport chain (Qadir *et al.*, 2004), lipid peroxidation and subsequent disruption of membrane integrity (Pandolfini *et al.*, 1992; Ros *et al.*, 1992; Gonnelli *et al.*, 2001).

The majority (~80%) of hyperaccumulator plants are known to hyperaccumulate nickel (Reeves & Baker, 2000; Verbruggen *et al.*, 2009; Krämer, 2010). Nickel's sole function in plants is as a cofactor in the enzyme urease, and as a result it is the least abundant (along with molybdenum) essential trace element, reaching concentrations of typically 0.1–10 mg/kg dry biomass in non-accumulator plants (Reeves, 1992; Marschner, 1995; Palmer & Guerinot, 2009). Other metals for which hyperaccumulating plants have been discovered include: zinc, cadmium, copper, cobalt, manganese, and the metalloids arsenic, thallium and selenium, with nickel being the most commonly hyperaccumulated metal due to its prevalence in ultramafic rocks over all the continents (Van der Ent *et al.*, 2013). When ultramafic rocks have weathered and undergone metamorphic hydration of the minerals

olivine and pyroxene (Mengoni *et al.*, 2009), they form serpentine soils that have have distinctive physical and chemical characteristics.

About 90% of nickel hyperaccumulators are species endemic to serpentine soils (Reeves & Baker, 2000). The reverse is not true: nickel hyperaccumulation is quite rare on serpentine soils – only approximately 2% of serpentine species are nickel hyperaccumulators (Kazakou *et al.*, 2008). Serpentine soils display particularly high contents of nickel, cobalt and chromium, low Ca/Mg ratio, and low levels of essential minerals: nitrogen, phosphorus and potassium (Brooks, 1987). Other types of soils have nickel concentration of typically between 7 and 50 mg/kg, while in serpentine soils the nickel content usually ranges from 700 to 8000 mg/kg (Reeves, 1992; Reeves & Baker, 2000).

This characteristic geochemistry, combined with a typically thin soil cover and granular texture, poor water-holding capacity, easy erosion and exposure to high light intensity, makes serpentine soils a notoriously difficult environment for plant growth (Proctor, 1975; Freitas *et al.*, 2004). As a result, serpentine outcrops are treated as ecological islands inhabited by specialized floras. Indeed, the discontinuity between serpentine outcrops, with their sparse vegetation, and normal neighboring soils can be clearly delimited from afar (Brady *et al.*, 2005; Kruckeberg, 2004).

Despite the considerable interest in nickel hyperaccumulation in nature, this phenomenon was only recorded in the scientific literature for the first time in 1948 (Minguzzi & Vergnano, 1948). The species in question was *Alyssum bertolonii* Desv. growing on serpentine soils in Tuscany, Italy. However, the term 'hyperaccumulator' was not coined until 1976 (Jaffré *et al.*, 1976). Brooks *et al.* (1977) first set the threshold for nickel hyperaccumulation at 1,000 mg/kg shoot dry biomass (all Ni concentration values are given per shoot dry mass unless stated otherwise) as this is 10-100 times higher than nickel concentration found in non-accumulator plants growing on serpentine (typically 10-100 mg/kg, Reeves, 1992), and 100-1,000 times higher than in plants growing on other soils (typically 0.1-5 mg/kg, Reeves, 1992; Palmer & Guerinot, 2009). This threshold is not entirely arbitrary, as evidenced by

the non-overlapping bimodal distribution of shoot nickel content in hyperaccumulators and non-accumulators with a boundary between the two groups at around 0.1% Ni in the genus *Alyssum* L. (Brooks, 1987; Reeves, 1992; Pollard *et al.*, 2002).

Taxonomically, nickel hyperaccumulation ability is most prevalent in eudicots – only four monocot species, as contrasted with 363 eudicots, have so far been found to exhibit this trait (Burge & Barker, 2010). In fact, distribution of the nickel hyperaccumulation trait is heavily skewed towards three families containing a large number of hyperaccumulators: Brassicaceae (87 species in 10 genera, including the genus *Alyssum* section *Odontarrhena*, Euphorbiaceae (83 species, including many Cuban endemics, such as species of *Phyllanthus*) and Asteraceae (46 species, including *Senecio coronatus* and *Berkheya coddii*). There are a further 18 families containing 2-20 species each, and 23 families each with a single hyperaccumulator species (Flynn, 2013).

1.2.2 Adaptive value of metal hyperaccumulation

Of four original hypotheses proposed to explain the possible adaptive advantage conferred by hyperaccumulation of metals (Boyd & Martens, 1992), only the pathogen/herbivore "elemental defense" hypothesis has gathered plentiful supporting experimental evidence (Boyd, 2004). There have been many reports of decreased pathogen infection rates and herbivory on plants hyperaccumulating metals (Poschenrieder *et al.*, 2006; Boyd, 2007; Fones *et al.*, 2010), but some negative evidence of no protective effect against herbivory exists in the literature, too (Noret *et al.*, 2007). More field and laboratory studies are underway in this active field of research to further test the different hypotheses concerning the selective advantage offered by metal hyperaccumulation. For instance, protection against environmental stresses has also been proposed as the role for high Ni concentration in above-ground tissue in plants. According to the hypothesis, perennial shrubs living in xeric habitats (such as *Alyssum* Ni hyperaccumulators) could be protected from UV-induced oxidative stress by nickel absorbance (de la Fuente *et al.*, 2007), but this hypothesis has not yet been tested in plants.

1.2.3 Metal hyperaccumulation as a complex trait

A complex trait "refers to any phenotype that does not exhibit classical Mendelian recessive or dominant inheritance attributable to a single gene locus" (Lander & Schork, 1994). The ability to hyperaccumulate metals requires considerable modification of basic physiological and biochemical pathways, as outlined below, which necessitates the involvement of multiple genes and pathways, e.g. as shown by previous QTL (Quantitative Trait Locus) analyses in zinc hyperaccumulators (Deniau *et al.*, 2006; Willems *et al.*, 2007; Frérot *et al.*, 2010) which led to identification of multiple major QTLs for zinc accumulation and tolerance. Thus, metal hyperaccumulation warrants a place on a list of important complex traits found in plants (Hanikenne & Nouet, 2011), such as CAM (Crassulacean Acid Metabolism) and C_4 photosynthesis (Christin *et al.*, 2011; Flood *et al.*, 2011; Langdale, 2011), and salt tolerance (Flowers *et al.*, 1997; Yamaguchi & Blumwald, 2005). No QTL analyses have been conducted on any nickel hyperaccumulating species so far but only on nickel-tolerant serpentine ecotypes of *Silene vulgaris* (Bratteler *et al.*, 2006**b**,**a**) and *Caulanthus amplexicaulis* (Burrell *et al.*, 2012), which invites new studies into the genetic architecture of nickel hyperaccumulation, including this one.

1.2.4 Physiology of metal hyperaccumulation

1.2.4.1 Overview

Hyperaccumulator plants accumulate metals to high concentrations in the shoot, so their shoot:root metal concentration ratios are typically above 1, unlike other plants when exposed to metals. This preferential accumulation of metals in the shoot entails modification to basic physiological mechanisms governing metal transport: these include increased rates of root metal uptake from the environment, decreased vacuolar sequestration in the root (Lasat *et al.*, 1998; Richau *et al.*, 2009) and efficient xylem loading (Krämer *et al.*, 1996; Hanikenne *et al.*, 2008). As metals need to be preferentially moved to the shoot in hyperaccumulators, vacuolar sequestration in the root is two to three times lower, and zinc efflux out of the vascular parenchyma cells into the xylem twice as fast in *Noccaea caerulescens* compared to non-accumulating plants, for instance (Lasat *et al.*, 2000; Richau *et al.*, 2009). This is followed by increased metal uptake into shoot cells (Küpper *et al.*,

2001) and sequestration in the cell wall and shoot vacuoles with high storage capacity (Küpper *et al.*, 2000; Ma *et al.*, 2005; Gustin *et al.*, 2009). Hyperaccumulator plants are obviously hypertolerant to the metal they accumulate: they are able to detoxify the metal when being transported through the symplast, and can store it in large quantities in their aerial organs (Pollard *et al.*, 2002; Krämer, 2010).

1.2.4.2 Mechanisms of countering metal toxicity

Metals pose a serious risk to cell metabolism because of their tendency to bind to free functional groups on proteins (Krämer *et al.*, 2007). As described by the Irving–Williams series, a given metal can displace another with lower complex stability constant: for example, nickel can displace iron, manganese and magnesium (da Silva & Williams, 1991); this phenomenon is generated by a decrease in the radii of high spin octahedral ions across the periodic table (Irving & Williams, 1953). In addition, transition metals have especially high binding affinities for N- and S-donor ligands (amino and thiol functional groups, respectively), so the cell must exercise tight control over the availability of free metal ions (Waldron *et al.*, 2009; Foster *et al.*, 2014). Consequently, transition metals in the cytoplasm are almost entirely bound by proteins with chaperone functions or by other specific chelators, resulting in e.g. sub-nanomolar free Ni^{2+} concentrations (Williams, 2007; Smart *et al.*, 2010). We know little about metallochaperones presence in plants apart from the copper ATX chaperones, which are highly conserved in eukaryotes (Burkhead *et al.*, 2009; Puig & Peñarrubia, 2009). On the other hand, low-molecular-weight chelators such as nicotianamine, histidine, citrate and malate have been shown to be necessary for sequestration of metals in specific cell compartments and/or for long-distance transport. Depending on the cell compartment or tissue, a given metal species will be chelated by a different ligand, forming a complex of different stability, facilitating further transport or immobilisation.

Furthermore, oxidative stress resulting from metal toxicity seems to require increased production of antioxidants, such as glutathione (Schützendübel & Polle, 2002; Jozefczak *et al.*, 2012). In the shoots of nickel-hyperaccumulating *Noccaea* species, salicylic acid

levels are constitutively elevated and promote activation of genes involved in glutathione synthesis and turnover resulting in increased levels of reduced glutathione, which appears to contribute to nickel tolerance (Freeman *et al.*, 2004, 2005; Freeman & Salt, 2007). This was shown in *Arabidopsis thaliana* plants expressing *Noccaea goesingense* serine acetyltransferase that overproduce salicylic acid and become highly resistant towards nickel (Freeman & Salt, 2007).

1.2.4.3 Metal chelation in transit

One such small molecule is the amino acid histidine, which is the only key constitutive component of nickel hyperaccumulation identified to date in *Alyssum*. Histidine acts as a very effective chelator for nickel cations through its carboxylate, imidazole and amine functions: of all the common carboxylic acids and proteinogenic amino acids, histidine has the highest association constant for binding to nickel at typical cytoplasmic pH values around 7.5 (Martell & Smith, 1989). Krämer *et al.* (1996) showed that nickel induced a dose-dependent increase in histidine in the xylem sap of three hyperaccumulator species of *Alyssum*, a response absent in the congeneric non-accumulator *Alyssum montanum*; furthermore, supplying *A. montanum* with exogenous histidine phenocopied the nickel hyperaccumulation trait. Thus, histidine is likely important for detoxifying nickel not only by chelation, but also by facilitating loading of nickel into the xylem (possibly as a nickel-histidine complex) and its transport to the shoot. Histidine-dependent transfer of nickel to the shoot might be specific to only some genera in the Brassicaceae, since *Arabidopsis thaliana* lines overproducing histidine show only increased nickel tolerance but not hyperaccumulation of nickel in the shoots, despite the presence of elevated concentrations of histidine in the xylem (Ingle *et al.*, 2005**a**). Nickel speciation analysis conducted by Alves *et al.* (2011) and Centofanti *et al.* (2013) showed that nickel in the xylem travels mostly as a free hydrated cation (70%) in *Alyssum pintodasilave*. In contrast, in the congeneric nickel hyperaccumulator *A. lesbiacum*, a considerable proportion (up to 19%) of the nickel being translocated in the xylem can be complexed to histidine, depending on the nickel concentration to which the plant is exposed (Krämer *et al.*, 1996). However, there are doubts about the feasibility of methods employed by Alves *et al.* (2011) to monitor

the native *in vivo* Ni complexation state, rising the possibility that the free, dissociated Ni state measured is a technical artefact.

Studies on nickel hyperaccumulators from the genus *Noccaea* provide more clues on the potential role of histidine in nickel hyperaccumulation, indicating that nickel chelation by histidine can prevent its sequestration in the root vacuole. Richau *et al.* (2009) showed that nickel-histidine complexes are poorly taken up by root- and shoot- derived tonoplast vesicles from nickel-hyperaccumulating *Noccaea caerulescens*, unlike the non-accumulator *Thlaspi arvense*. Much better sequestration was obtained for nickel-citrate complexes in the hyperaccumulator, especially in shoot vacuoles. Thus, it appears that due to prevention of sequestration of the nickel-histidine complex in root vacuoles of the hyperaccumulator species, detoxified nickel is available for export to the shoot, where it is eventually bound with organic acids such as citrate.

Another ligand important for early detoxification of nickel on entering the plant can be nicotianamine (NA), and a considerable proportion of Ni pool in the root and xylem in *N. caerulescens* is found in a complex with NA (Vacchina *et al.*, 2003; Mari *et al.*, 2006; Callahan *et al.*, 2007). Supporting the hypothesis that the nickel-nicotianamine complex is one form in which nickel travels from root to shoot in *N. caerulescens*, *A. thaliana* plants expressing the *N. caerulescens* nicotianamine synthase (the enzyme catalyzing nicotianamine formation from three S-adenosylmethionine molecules) became significantly more nickel tolerant and accumulated more nickel in the shoot (Pianelli *et al.*, 2005). Detoxification of nickel by complexation with nicotianamine and enhancement of xylem loading of nickel by nicotianamine has also been reported in *Arabidopsis halleri*, even though this plant is normally only a hyperaccumulator of zinc and cadmium (Cornu *et al.*, 2015).

1.2.4.4 Metal chelation *in situ*

Inside the shoot vacuoles, organic acids are quantitatively the major chelating agents for metals in hyperaccumulator plants, as shown using X-ray absorption techniques (Callahan *et al.*, 2006; Haydon & Cobbett, 2007). For instance, in *Arabidopsis halleri*, most of shoot

zinc pool is associated with malate (Sarret *et al.*, 2002), whereas in *Noccaea caerulescens* most of the zinc is bound with citrate (Küpper *et al.*, 2004). Citrate has also been shown to be the dominant ligand for chelation of nickel in shoots of several species of nickel hyperaccumulators (Lee *et al.*, 1978; Homer *et al.*, 1991). Summarizing a number of earlier studies, Brooks & Robinson (1998) concluded that the major chelators of nickel in shoots of hyperaccumulator species of *Alyssum* are malic, malonic and citric acids.

In the shoot, Ni has been shown to accumulate preferentially inside the epidermal cells on both adaxial and abaxial surfaces and at the base of trichomes in hyperaccumulating species of *Alyssum* (Küpper *et al.*, 2001; de la Fuente *et al.*, 2007; Smart *et al.*, 2010). In trichomes, nickel is distributed under a layer of calcareous deposits (Krämer *et al.*, 1997; Smart *et al.*, 2007). Nickel is also further deposited inside epidermal cell walls, mesophyll and vascular bundle, in that order of preference. Interestingly, in the hyperaccumulator species *Alyssum murale*, in contrast to the non-accumulator *Alyssum montanum*, nickel is specifically directed into mitochondria, where it interacts with Krebs cycle to increase synthesis of its side-products citric and malic acids, which are used to chelate nickel (Agrawal *et al.*, 2013).

1.2.5 Use of metal hyperaccumulating plants by humans

In addition to naturally occurring metal-rich soils, hyperaccumulator plants often flourish on sites contaminated as a result of human activity such as mining and smelting. This has suggested the potential application of metal hyperaccumulator plants in cleaning up metal-contaminated sites (phytoremediation) and of extracting metals from low-grade ores or mineralized soil (phytomining, agromining) (Brooks & Robinson, 1998). While field trials have shown economic viability of nickel phytomining using e.g. *Alyssum corsicum* under certain conditions (Li *et al.*, 2003**b**), novel plant genotypes with superior ability to accumulate multiple metals as well as production of high biomass combined with efficient metal extraction and storage are still sought after. In order to develop such plants, we need better understanding of the exact genetic pathways implicated in metal hyperaccumulation and general marginal land tolerance (Chaney *et al.*, 2007, 2010; Sheoran *et al.*, 2010; Escande *et al.*, 2014; Van der Ent *et al.*, 2015). Although not the primary focus of the present work, it is to be hoped that further insight into the genomic basis of metal hyperaccumulation

might provide novel targets for breeding strategies aimed at producing plants more suited to a variety of applications in phytoremediation and phytomining.

1.3 *A. serpyllifolium* as a model species for investigating nickel hyperaccumulation

1.3.1 Limited historical research into nickel hyperaccumulation

Despite the greater abundance of nickel hyperaccumulator species (>400) in nature, much more research has so far been conducted on the molecular basis and evolution of metal hyperaccumulation in zinc and cadmium hyperaccumulators. However, only 15 described species display zinc hyperaccumulation (Meerts & Van Isacker, 1997; Bert *et al.*, 2000; Krämer, 2010) and just 4 cadmium hyperaccumulation (Brown *et al.*, 1995; Küpper *et al.*, 2000; Verbruggen *et al.*, 2009). Two species have become *de facto* model organisms for studying molecular mechanisms of hyperaccumulation: *Arabidopsis halleri* (L.) O'Kane & Al-Shehbaz (referred to in the most of the earlier literature prior to 1997 as *Cardaminopsis halleri* (L.) Hayek) and *Noccaea caerulescens* (J.Presl & C.Presl) F.K.Mey. (referred to in the earlier literature as *Thlaspi caerulescens* J.Presl & C.Presl), which are constitutive zinc hyperaccumulators in both metallicolous and non-metallicolous populations. Some populations of the two species hyperaccumulate cadmium (Robinson *et al.*, 1998; Reeves *et al.*, 2001; Bert *et al.*, 2003; Roosens *et al.*, 2003), and a number of *Noccaea caerulescens* accessions can hyperaccumulate nickel as well, albeit to a slightly lower degree (Reeves & Brooks, 1983; Reeves *et al.*, 2001). However, as nickel hyperaccumulation is the more common form in terms of number of species across the plant kingdom (368 species according to Burge & Barker, 2010 and an estimate of 450 species by Van der Ent *et al.*, 2015) and the most efficient form of metal hyperaccumulation, exclusively nickel hyperaccumulating taxa could benefit from more focused studies (Kazakou *et al.*, 2008). So far, the genus *Alyssum* L. containing nickel hyperaccumulators along with non-accumulator species (Krämer, 2010; Pollard *et al.*, 2014), has been the favoured taxon in nickel hyperaccumulation research using modern genetics and molecular biology approaches.

1.3.2 Occurrence of nickel hyperaccumulation in *Alyssum* L.

The genus *Alyssum* L. was first described by Linnaeus in 1753 based on *Alyssum montanum* (Dudley, 1966). As currently circumscribed, the genus contains about 190 species, and is particularly notable for the fact that it contains the highest number of hyperaccumulator species of all genera in both absolute and relative terms – 51 hyperaccumulators, contrasted with 98 non-accumulators and 39 non-characterised taxa (Burge & Barker, 2010). Flynn (2013) (Figure 1.1) showed that all 51 *Alyssum* hyperaccumulator species (34 exclusively hyperaccumulators, 17 facultative hyperaccumulators) belong to a single monophyletic clade that recovers almost exclusively (except for 5 species) the traditional section *Odontarrhena*, containing all the previously identified nickel hyperaccumulators in *Alyssum* (Brooks *et al.*, 1979), with the sole exception of the hyperaccumulator *A. lenense* Adams. However, this clade shows poor internal resolution at the species level due to insufficient phylogenetically informative variation for the four sequenced chloroplast DNA regions (*trnL–F*, *rps16–trnK*, *matK*, *trnD–T*), and so it has not been possible to establish the phylogenetic relationships between the hyperaccumulators and non-accumulators within this clade with any degree of confidence (Flynn, 2013).

1.3.3 *Alyssum serpyllifolium* – facultative hyperaccumulator

Several taxa within the genus *Alyssum* are classed as facultative hyperaccumulators or pseudometallophytes (Pollard *et al.*, 2014), i.e. containing populations occurring on metalliferous and non-metalliferous substrates with distinct differences in ability to hyperaccumulate metals. The *Alyssum serpyllifolium* species complex is the best studied case of facultative nickel hyperaccumulation in *Alyssum*, with clear morphological and physiological differences between serpentine and non-serpentine populations being evident under 'common garden' conditions in the laboratory (Brooks *et al.*, 1981; Pollard *et al.*, 2002; Ingle *et al.*, 2005**a**). *Alyssum serpyllifolium sensu lato* is a widespread species in south-western Europe, with a distribution centred on the Iberian Peninsula, that is found growing on both serpentine and non-serpentine substrates (Dudley, 1964; Tutin, 1993). At various times, it has been proposed that the serpentine populations should be accorded the status of distinct subspecies, or even raised to the rank of species (Dudley,

1986**a**,**b**), but so far phylogenetic analyses based on chloroplast DNA sequences have not been able to establish clearly resolved relationships between the different populations (Cecchi *et al.*, 2013; Flynn, 2013). The *Alyssum serpyllifolium* species complex was chosen for investigation in the present study, because as a taxon with distinct serpentine and non-serpentine ecotypes, with close but as yet unresolved relationships, it lends itself well to the study of the evolution of the hyperaccumulation trait from a number of perspectives.

1.3.4 A singular innovation in *Alyssum* L. - nickel hyperaccumulation?

In theory, two simplified scenarios could be invoked to explain the origin and wide geographic distribution of the nickel hyperaccumulation trait within the genus *Alyssum*, given that hyperaccumulator species are found on serpentine sites scattered across the entire Mediterranean basin, from the Iberian peninsula in the west to the Irano-Turanian region in the east (Brooks, 1987). Hyperaccumulation ability might either have arisen at each serpentine site independently from a local (adjacent) ancestral non-accumulating population, or alternatively the hyperaccumulation trait might have arisen just once in an ancestral population and then spread by dispersal and range expansion to serpentine sites scattered over a wide geographical area. Serpentine soil biochemistry and the specialised plant adaptations required to thrive on such substrates suggest that serpentine-adapted plants may not be successful competitors on non-serpentine substrates (Brooks, 1987; Elmendorf & Moore, 2007; Anacker & Harrison, 2012; Anacker, 2014). However, given appropriate dispersal abilities, they should be able to colonise other serpentine islands on encountering suitable habitats. With time, the populations occupying such serpentine islands would become genetically distinct and could, as edaphic endemics, eventually give rise to new taxa through the process of allopatric speciation.

According to the phylogenetic work of Flynn (2013) based on chloroplast DNA sequences (Figure 1.1), nickel hyperaccumulation appears to have evolved in the genus *Alyssum* on the branch leading to a single monophyletic clade that includes the section *Odontarrhena* (subclade A1a, with a crown node radiation estimated from ultrametric trees to date to 3.3 MYA). This single origin of nickel hyperaccumulation could require up to 15 apparent losses of the

nickel hyperaccumulation trait within the clade (depending on how the non-accumulator species are coded), but multiple independent origins of nickel hyperaccumulation could not be ruled out. Sub-clade A1a radiated in the Pliocene Epoch, and Flynn (2013) inferred a 2.5-fold increase in the species net diversification rate on the branch leading to this clade compared with the background rate in the rest of the tribe Alysseae using a binary state speciation and extinction (BiSSE) analysis (Maddison *et al.*, 2007). Flynn (2013) therefore concluded that nickel hyperaccumulation qualified as a 'key innovation' in *Alyssum*, fulfilling the four criteria required to meet such a designation: synapomorphy (single origin of the trait and its persistence); functional advantage (tolerance of serpentine soils, and possibly also enhanced defence against herbivores and pathogens); range expansion (widespread distribution across the Mediterranean basin); and higher net diversification rate of lineages possessing the trait. In practice, as the nickel hyperaccumulation trait always co-occurs with serpentine tolerance within *Alyssum*, it is difficult to distinguish between the two in terms of what constitutes the actual key innovation in the lineage, and this issue will be re-examined in the research presented here. Specifically, *Alyssum* hyperaccumulators are expected to show genetic adaptations not only related to nickel hyperaccumulation, but also to the other challenges posed by the serpentine soils, such as high chromium and cobalt, low calcium and high magnesium, low levels of potassium, nitrogen and phosphorus, and low water availability due to the poor water-retention ability of the soils (Proctor, 1975; Brady *et al.*, 2005; Kazakou *et al.*, 2008).

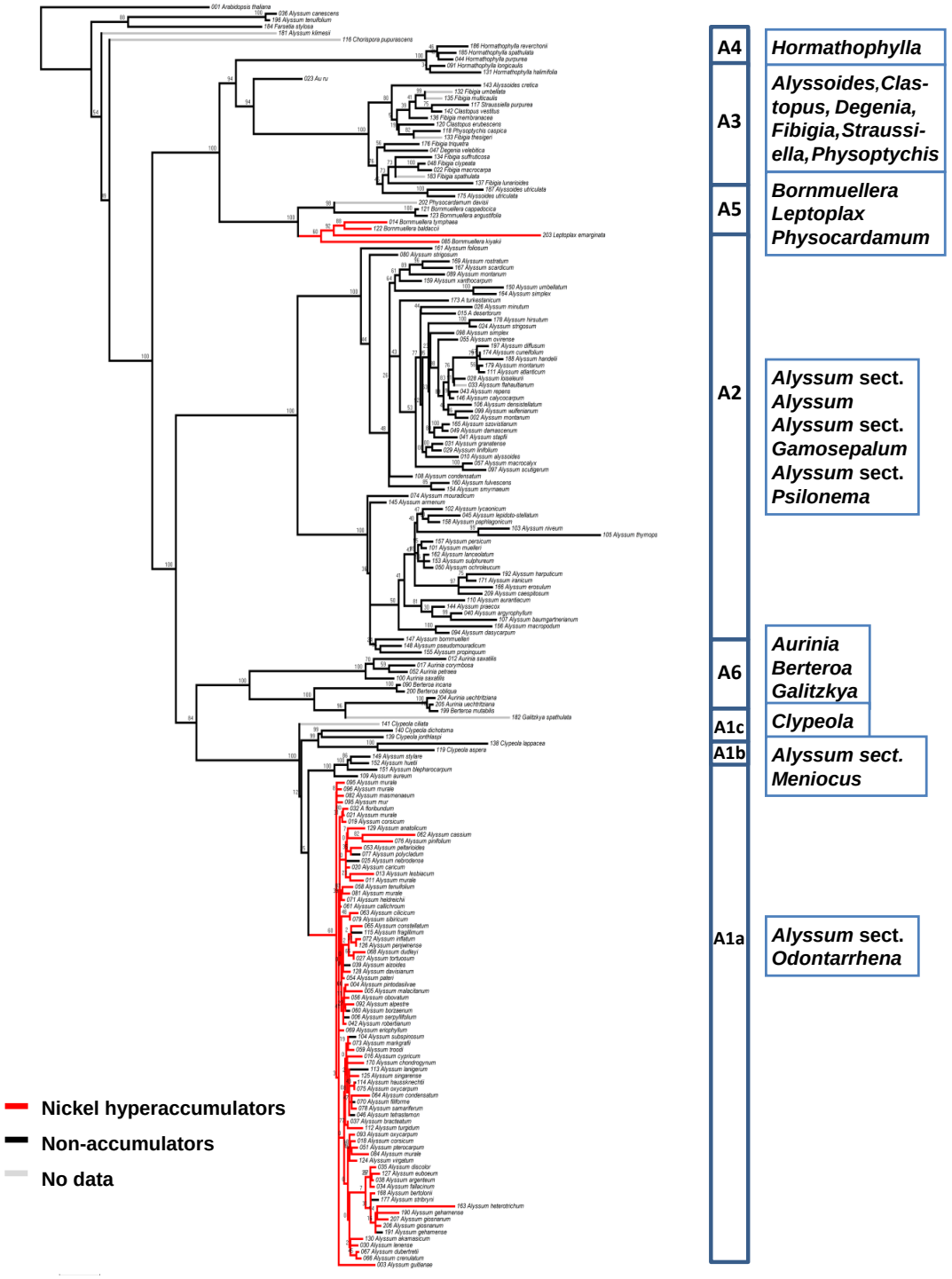


Figure 1.1: A maximum likelihood tree of the tribe *Alysseae*, modified with permission from Flynn (2013). To produce the tree, 170 out of 255 species in the tribe *Alysseae* were sequenced for at least four loci (*matK*, *rps16–trnK*, *trnD–T* and *trnL–F*), amounting to 5.8 kb of cpDNA sequence in total. Note that *Alyssum*, as traditionally circumscribed, is not monophyletic, and that *Alyssum* section *Odontarrhena* (A1a), which contains the *Alyssum serpyllifolium* species complex, is sister to *Clypeola* (A1c).

Table 1.1: Names used at various times at species and subspecies ranks to describe distinct populations within the *Alyssum serpyllifolium* species complex. Suffix -S denotes serpentine populations and -NS non-serpentine populations.

Species name	Subspecies name	Type locality	Serpentine?	Populations investigated
Alyssum serpyllifolium Desf. (Desfontaines, 1798)	ssp. *serpyllifolium*	Atlas Mountains, northern Algeria	No	Alhaurín-NS, León-NS, Morata-NS, Rubia-NS, Anduze-NS, Bédarieux-NS, Malaucène-NS
Alyssum pintodasilvae T.R. Dudley (Dudley, 1986**a**)	ssp. *lusitanicum* T.R. Dudley & P. Silva (Dudley, 1966)	Bragança, Trás-os-Montes e Alto Douro, northeastern Portugal	Yes	Samil-S
Alyssum malacitanum (Rivas Goday) (Dudley, 1986**b**)	ssp. *malacitanum* Rivas Goday (Rivas Goday & Esteve, 1972)	western Betic Cordillera of Málaga, Andalusia, southern Spain	Yes	Carratraca-S, Sierra Bermeja-S
Alyssum malacitanum (Rivas Goday) (Dudley, 1986**b**) *'Alyssum guitianae' nomen nudum*	N/A	Melide, Capelada-Serro do Careón, A Coruña, Galicia, northwestern Spain	Yes	Barazón-S

1.3.5 Morphological differences within *Alyssum serpyllifolium* species complex

Alyssum serpyllifolium Desf. is a highly variable, exclusively sexual perennial and iteroparous herb species, first described based on non-serpentine populations (Desfontaines, 1798), but variously split at different times into distinct subspecies or even species (Dudley, 1986**a**,**b**; Rodriguez-Oubina & Ortiz, 1991; Tutin, 1993). It is characterised by oblanceolate or obovate-spathulate leaves (Cecchi *et al.*, 2013), which along with stems are covered by unicellular stellate (stalk-less, multi-branched) trichomes with a variety of forms (Dudley, 1964) that give a tomentose-whitish appearance to the plant (de la Fuente *et al.*, 2007). Field-collected *Alyssum serpyllifolium* specimens from different populations (Figure 1.2 A, B, C and Figure 1.3 *top*) display differences in morphology in terms of trichome cover, plant stature, leaf shape and coloration, which led E.J. Nyárády (1931, 1932, 1949) to delimit as many as 19 intraspecific taxa within the species (Dudley, 1966). This is now thought to represent a grossly inflated estimate of the number of infraspecific taxa, but the taxonomic status of different populations within the *Alyssum serpyllifolium* species complex is still unresolved and is summarised in Table 1.1. As all species/subspecies descriptions were made from field material, some of the differences in morphology seen could be due to phenotypic plasticity rather than genetic differences (Sultan, 2000). On the other hand, differential nickel hyperaccumulation and tolerance, along with certain differences in morphology in standardised pot (Brooks *et al.*, 1981; Dudley, 1986**a**) and hydroponic trials (A.J. Pollard & J.A.C. Smith, unpublished results) do support the existence of key genetic differences between serpentine and non-serpentine populations. In the present work, different accessions are typically referred to by their population names derived from nearby geographic features, and the name is always followed by a suffix denoting a serpentine (-S) or non-serpentine (-NS) origin of the population. Species names (*Alyssum pintodasilvae*, *A. malacitanum*, *A. guitianae*) are also sometimes used to refer to the various serpentine populations of *A. serpyllifolium* Desf. from the Iberian peninsula.

Differences in plant stature, as well as leaf shape and arrangement, were used as diagnostic characters in the original descriptions of the serpentine taxa *Alyssum malacitanum*

(Carratraca-S) and *A. pintodasilvae* (Samil-S). *Alyssum malacitanum* was described as having flat leaves (Dudley, 1986**b**) and *A. pintodasilvae* as having conduplicate leaves (folded together lengthwise with the upper surface within; Dudley, 1986**a**). Results from a recent common garden experiment confirmed these differences in leaf morphology between the Samil-S/Barazón-S and Carratraca-S populations, and also revealed that Samil-S and Barazón-S populations shared an obovate-spathulate leaf shape, whereas the Carratraca-S population possessed an oblanceolate leaf shape (unpublished, Graham, 2013). Amongst three non-serpentine populations examined, all three showed conduplicate leaves, but while one population displayed oblanceolate leaf shape (Morata-NS), the other two showed obovate-spathulate leaves (Rubia-NS and Alhaurín-NS), so these morphological characters did not differentiate consistently between serpentine and non-serpentine populations. Similarly, trichome morphology, which has been previously employed to delineate species boundaries in Brassicaceae (Ančev & Goranova, 2006; Beilstein *et al.*, 2006, 2008), was found to show only marginal differences between the various populations of *Alyssum serpyllifolium*, with the three serpentine populations having a significantly higher abaxial trichome area density (and correspondingly smaller trichomes) than the three non-serpentine populations in the study (unpublished, Graham, 2013). Finally, morphometric analysis of seeds and silicles using nine descriptors did not reveal any consistent differences between *A. malacitanum*, *A. pintodasilvae* and *A. serpyllifolium sensu stricto*. These results suggest that morphological differences between populations within the *A. serpyllifolium* complex are marginal at best and highlight the need for phylogenetic investigations of the taxonomic status of these different entities.

1.3.6 Physiological differences within *Alyssum serpyllifolium* species complex

Measurements of shoot nickel concentration in *Alyssum serpyllifolium* material collected in the field have established that non-serpentine populations of *Alyssum serpyllifolium* growing on limestone soils are non-accumulators containing only low amounts of Ni in the shoot (2 to 4 mg/kg; Brooks & Radford, 1978; Brooks *et al.*, 1979; de la Fuente *et al.*, 2007; Becerra-Castro, 2012; Cabello-Conejo *et al.*, 2014. Conversely, specimens collected

from serpentine soils showed nickel concentrations in the shoot always exceeding the conventional hyperaccumulation threshold of 0.1 % (w/w): *Alyssum malacitanum* - mean (± SD) of 10,391 ± 774 mg/kg (Becerra-Castro, 2012); *Alyssum guitianae* - mean of 15,464 ± 719 mg/kg (Becerra-Castro, 2012); *Alyssum pintodasilvae* - range from 1,940 to 9,000 mg/kg (Brooks and Radford, 1978), or mean of 11,022 ± 1,169 mg/kg (Becerra-Castro, 2012), with a maximum of 38,105 mg/kg (Freitas *et al.*, 2004). In the field, plants restricted to non-serpentine soils will not have the opportunity to hyperaccumulate nickel, so the only way to determine the physiological differences between these populations is by means of common-garden experiments under controlled conditions. Nickel tolerance and accumulation tests carried out by Brooks *et al.* (1981) in the laboratory setting first questioned the simple division between hyperaccumulating serpentine *Alyssum* populations/species and non-serpentine non-accumulators. A nickel tolerance test based on the maximum tolerated Ni concentration in pot trials with amended soils showed substantial range of values from serpentine *Alyssum pintodasilvae* (maximum tolerance of 3,160 mg/kg soil Ni) to *A. malacitanum* (max. 1,000 mg/kg) and the non-serpentine *Alyssum serpyllifolium* Desf. (max. 600 mg/kg). Ni tolerance assays measuring germination rate and root growth in excised shoots resulted in the same rank order of taxa (Brooks et al., 1981b). Importantly, when grown in artificially Ni-enriched soil in pots, the maximum leaf Ni concentration in *Alyssum serpyllifolium* Desf. plants could reach the hyperaccumulation threshold of 1,000 mg/kg (Brooks *et al.*, 1981; Reeves & Brooks, 1983).

As a prelude to the present investigation of the genomic basis of nickel hyperaccumulation in *Alyssum*, and to characterise the physiological differences between populations of *A. serpyllifolium* in greater detail, an investigation of nickel tolerance and hyperaccumulation in six accessions of this taxon was recently conducted for hydroponically cultivated plants by A.J. Pollard and J.A.C. Smith, the results of which are reported briefly here. In solution-culture experiments, the exposure to precise nickel concentrations can be controlled more readily than in amended soils, and a range of concentrations can be tested covering incipient to severe toxicity. Six different Iberian populations of *A. serpyllifolium sensu lato* were used in this experiment (the sources of which are shown in Figure 1.4): Carratraca-S (*A.*

malacitanum) and Samil-S (*A. pintodasilvae*), Barazon-S (*'A. guitianae'*), Morata-NS, Rubía-NS and Alhaurín-NS (all *A. serpyllifolium* Desf.), together with two other species from the tribe Alysseae as outgroups for comparison: *A. montanum* L. and *Clypeola jonthlaspi* L. *Clypeola jonthlaspi* is a non-hyperaccumulating annual herb (Figure 1.2 and Figure 1.3 *bottom*), widely distributed across Europe, North Africa and south-west Asia (Tutin, 1993; Keshavarzi *et al.*, 2012). The phylogenetic work of Flynn (2013) using chloroplast DNA sequences (Figure 1.1) indicates that the genus *Clypeola* is nested within the genus *Alyssum* as traditionally circumscribed, in a sister-group position to the subclade A1a containing the entire section *Odontarrhena* (including *Alyssum serpyllifolium*); that close relationship was confirmed in other Alysseae phylogenies (Warwick *et al.*, 2008; Rešetnik *et al.*, 2013; Li *et al.*, 2015**b**) with even merger of the two subclades proposed (Warwick *et al.*, 2008; Rešetnik *et al.*, 2013). *Alyssum montanum* is more distantly related to *A. serpyllifolium* and *Clypeola jonthlaspi*, as it is part of section *Alyssum* within the traditionally circumscribed genus *Alyssum*, and it is another species complex showing a wide geographic distribution and displaying considerable intraspecific variability (Tutin, 1993; Španiel *et al.*, 2012).

As assessed by their relative dry biomass production in nickel-containing solutions (Figure 1.5), all six accessions of *Alyssum serpyllifolium sensu lato* tested showed a considerably higher degree of nickel tolerance than *A. montanum*, for which growth was already approximately 50% inhibited at 10 μM Ni. *Clypeola jonthlaspi* was similar in its nickel tolerance to the two least-tolerant accessions of *A. serpyllifolium* (Morata-NS and Rubía-NS), but was significantly more nickel-tolerant than *A. montanum* (data not shown). On average, the three serpentine populations showed a higher degree of nickel tolerance than the three non-serpentine populations, but a wide range of nickel tolerance was evident across the six accessions. Growth of the two most tolerant serpentine accessions (Carratraca-S and Samil-S) was not significantly reduced even at the highest nickel concentration tested (300 μM Ni), whereas a slight growth inhibition was detectable for the two least-tolerant accessions of *A. serpyllifolium* (Morata-NS and Rubía-NS) at 30 μM Ni ($p < 0.05$). Interestingly, two accessions – the most nickel tolerant of the non-serpentine populations (Alhaurín-NS) and the least tolerant of the serpentine populations (Barazón-S) – showed an intermediate, and

rather similar, degree of nickel tolerance. This suggests that a wide range of intermediate phenotypes exist amongst the natural populations of *A. serpyllifolium*, rather than a strict bimodal distribution of tolerance values in serpentine versus non-serpentine populations.

In terms of hyperaccumulation, the highest levels of nickel hyperaccumulation were shown by the three serpentine accessions, which accumulated significantly ($p < 0.05$) greater nickel concentrations in their leaves (≥ 10,000 mg/kg) at 300 μM Ni than the three non-serpentine accessions (Figure 1.5). However, at solution Ni concentrations of 100 μM or less, all six accessions showed very similar levels of nickel accumulation in the leaf tissues, and indeed all six exceeded the conventional nickel-hyperaccumulation threshold of 1,000 mg/kg shoot dry biomass at a solution Ni concentration of 30 μM, at which the non-serpentine accessions showed only mild toxicity symptoms. All six accessions of *A. serpyllifolium* accumulated considerably higher concentrations of foliar nickel than *A. montanum*, which remained below the nickel-hyperaccumulation threshold at all but the most toxic concentration of nickel.

These findings provide some further insight into the nickel tolerance and hyperaccumulation of natural populations of *Alyssum serpyllifolium* relative to the original studies on this species complex published by Brooks *et al.* (1978, 1979, 1981a, 1981b) and Morrison *et al.* (1980). First, although the serpentine populations on average are more physiologically nickel-tolerant than the non-serpentine populations, there appears to be a continuous range of intermediate phenotypes present within the species complex, with the least tolerant of the serpentine populations (Barazón-S) being only marginally more nickel-tolerant than the most tolerant of the non-serpentine populations (Alhaurín-NS).

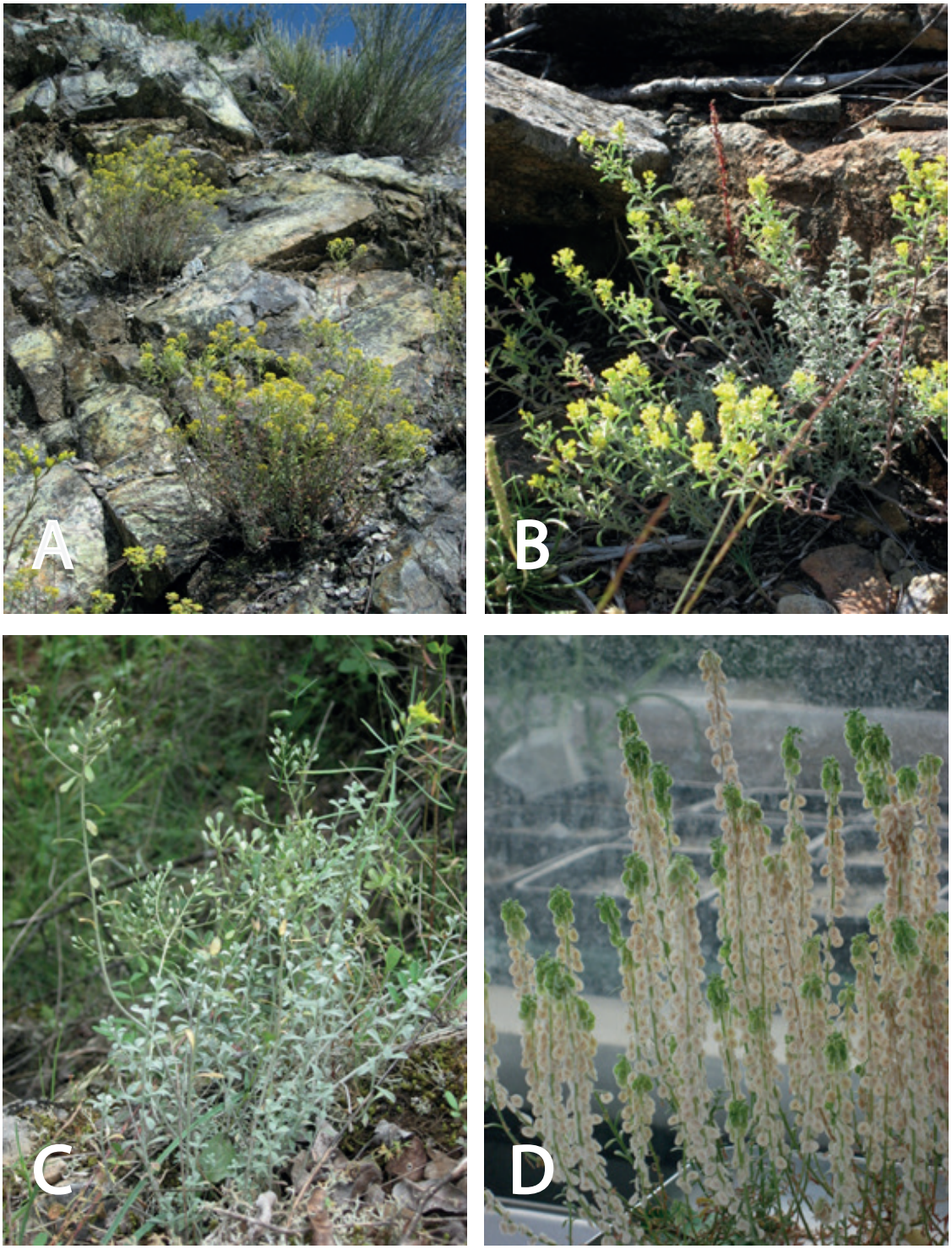


Figure 1.2: Serpentine taxa: **A)** *Alyssum pintodasilvae* T.R.Dudley (= *A. serpyllifolium* Desf. ssp. *lusitanicum* T.R.Dudley & P.Silva) growing at Samil, near Bragança, Trás-os-Montes, Portugal; **B)** *Alyssum malacitanum* (Rivas Goday) T.R.Dudley (= *A. serpyllifolium* Desf. ssp. *malacitanum* Rivas Goday) growing at Carratraca, Málaga, Andalucía, Spain.
Non-serpentine taxa: **C)** *Alyssum serpyllifolium* Desf. ssp. *serpyllifolium*) growing at Covas, Rubiá, Orense, Galicia, Spain; **D)** *Clypeola jonthlaspi* L. cultivated in the glasshouse, grown from seed originating from Jarash, Jordan, kindly provided by the Millennium Seed Bank, Royal Botanic Gardens, Kew (serial no. 413532). Photographs courtesy of J.A.C. Smith, used with permission.

Figure 1.3: **Top**: Inflorescence of glass-house cultivated *Alyssum malacitanum* (Carratraca-S). Photograph courtesy of John Baker, used with permission.
Bottom: Inflorescence-infructescence of *Clypeola jonthlaspi*. Photograph courtesy of Sarah Gregg, used under Creative Commons license with permission.

Figure 1.4: Distribution of the sampled *Alyssum serpyllifolium* populations in the Mediterranean region. Serpentine populations are marked in blue and non-serpentine populations in red.

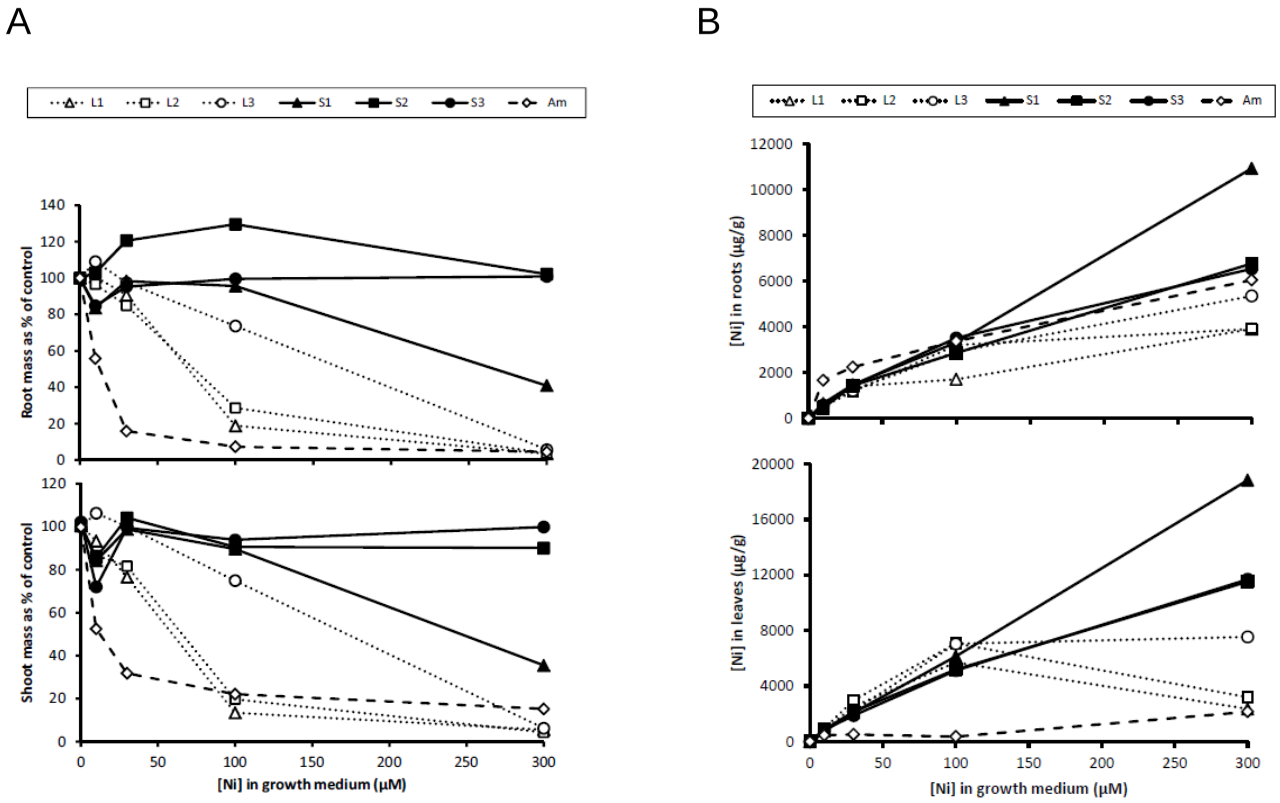


Figure 1.5: Nickel tolerance (A) and hyperaccumulation (B) assayed in a solution-culture experiment with six accessions of *Alyssum serpyllifolium sensu lato* (L1 – 3: three limestone populations; S1 – 3: three serpentine populations) and one accession of *Alyssum montanum* (Am). Plants were germinated from seed for 2 weeks on moist sand before transfer to a hydroponic solution based on 0.1-strength Hoagland solution macronutrients plus modified 0.2-strength Long Ashton micronutrients as described by Roosens et al. (2003). After 3 weeks, the hydroponic solution was supplemented with $NiSO_4$ to a final concentration of 0.1 (control), 10, 30, 100, or 300 *µM*. Solutions were replaced weekly and after 8 further weeks plants were harvested and oven-dried at 80 °C for 72 h for biomass determination. Nickel content was determined on sub-samples of root and shoot dry biomass following extraction with concentrated HNO_3 and appropriate dilution by atomic absorption spectrophotometry. Populations of *A. serpyllifolium* were as follows: L1: Rubía, Galicia, Spain; L2: Morata, Madrid, Spain; L3 Alhaurín, Andalucía, Spain; S1: Barazón, Galicia, Spain; S2: Samil, Trás-os-Montes, Portugal; S3: Carratraca, Andalucía, Spain. Seeds of *A. montanum* were obtained from a commercial source (Dobies of Devon, Paignton, UK). Unpublished results of A.J. Pollard and J.A.C. Smith, used with permission.

Second, even the least tolerant of the *A. serpyllifolium* populations sampled from non-serpentine sites were found to be considerably more nickel-tolerant than *A. montanum*; this is contrary to the view of Brooks et al. (1978, 1979, 1981a, 1981b), who considered the taxa to be almost equally nickel-sensitive, although their protocols did not in fact permit a very precise estimate of the effect of moderate concentrations of nickel. Third, both the serpentine and non-serpentine accessions have the physiological potential to hyperaccumulate nickel in their shoots to concentrations above 0.1 % (w/w) before the onset of pronounced toxicity symptoms, even though the latter do not have the opportunity to express this capacity when growing on non-ultramafic substrates. In line with the clear Ni hyperaccumulator status of both serpentine and non-serpentine populations of *A. serpyllifolium* revealed under laboratory conditions, the population will be grouped based on their edaphic origin. These intraspecific differences in nickel hyperaccumulation and tolerance found within the *Alyssum serpyllifolium* species complex thus provide a very good starting point for dissection of the genetic mechanisms and genomic architecture underpinning these traits, which will form the principal topic of this thesis. While a certain degree of intraspecific variation in Ni hyperaccumulation and tolerance has previously been demonstrated in the serpentine endemic hyperaccumulator *Alyssum bertolonii* native to the Apennine mountains in Italy (Galardi *et al.*, 2007), and indeed the non-accumulator *A. montanum* is also very variable and occurs both on and off serpentine substrates (Tutin, 1993; Španiel *et al.*, 2012), neither of these taxa shows the wide range of accumulator phenotypes exhibited by the *Alyssum serpyllifolium* species complex across the variety of habitats in which it is found in the field.

1.3.7 *Alyssum serpyllifolium* species complex – distribution of different cytotypes

The geographical distribution of *Alyssum serpyllifolium* range spans the Iberian Peninsula, southern France, and most likely North Africa (Tutin, 1993; Jalas *et al.*, 1996). Original reports exist of the species' presence in north-west Africa (Algeria, Morocco; Cecchi *et al.*, 2013), but the African populations do not appear to have been investigated since the early 1990s (Vogt & Oberprieler, 1994). The traditional hyperaccumulator populations

are endemic to a few ultramafic outcrops in Spain and Portugal (Figure 1.4, Brooks, 1987; Cecchi *et al.*, 2013. *Alyssum malacitanum* (= *A. serpyllifolium* ssp. *malacitanum*) is part of secondary scrub communities growing on soils of dolomite and peridotite ultramafic rocks of the Baetic Province (Dudley, 1986**b**; Rivas Martínez, 1991), specifically the endemic alliance Staehelino-Ulicion baetici located in the western Baetic Cordillera of Málaga (Sierra Bermeja-S and Carratraca-S populations). This is an area of 30,000 hectares of ultramafic rocks (Asensi *et al.*, 2004), typically displaying soil Ni concentrations of between 1,532 and 4,254 mg/kg (de la Fuente *et al.*, 2007). *Alyssum pintodasilvae* (Samil-S) is restricted to the 8,000 hectare serpentine area in north-eastern Portugal in the region of Trás-os-Montes e Alto Douro near Bragança, where soil Ni concentrations vary from 851 to 2,962 mg/kg (Peterson *et al.*, 2003). Serpentinophytes such as *Alyssum pintodasilvae* dominate the landscape there due to extensive tillage, preventing establishment of climax dwarf scrub vegetation (Brooks et al., 1981b; Dudley, 1986a). The Galician serpentine populations in northwestern Spain near Melide (Barazon-S) have sometimes been described as *'Alyssum guitianae'* (Rodriguez-Oubina & Ortiz, 1991), but this name has never been validly published, and more recently the tendency has been to regard these populations as part of *A. pintodasilvae*.

Serpentine areas inhabited by *Alyssum serpyllifolium* populations take the form of relatively small 'islands' of specialized substrate up to tens of kilometres in diameter (Brooks, 1987; Flynn, 2013). As they are surrounded by vast stretches of non-serpentine 'sea' with different geochemistry, they have previously been described as edaphic islands (Stebbins, 1942; Mayer *et al.*, 1994), by analogy to classical archipelagos of islands harbouring distinct floras. Here, instead of large distances between islands limiting migration and promoting speciation, there are not easily surmountable ecological boundaries presented by substrates of different geochemistry in the intervening areas. In the *Alyssum serpyllifolium* species complex, this is coupled with the fact that the typical serpentine outcrops may also be constrained by altitude, as they often occur in mountain ranges such as Sierra de Aguas (Carratraca-S: *A. malacitanum*), Sierra Bermeja de Estapona (Sierra Bermeja-S: *A. malacitanum*), and Serra do Careón (Barazón-S: *'A. guitianae'*).

The *Alyssum serpyllifolium* species complex also shows two ploidy levels, as is common in the Brassicaceae family and also within the genus *Alyssum* (Španiel *et al.*, 2015 and cf. the *A. montanum* species complex: Španiel *et al.*, 2011; Zozomová-Lihová *et al.*, 2014). The populations on the Iberian peninsula (*A. serpyllifolium*: Cecchi *et al.*, 2013; *A. pintodasilvae*: Fernandes & Queirós, 1973; Küpfer, 1974; Cecchi *et al.*, 2013; *A. malacitanum* and *'A. guitianae'*: Cecchi *et al.*, 2013) and Africa (Galland & Love, 1984; Galland, 1990; Vogt & Oberprieler, 1994) show ploidy $n = 8$, but both diploid $n = 8$ (Küpfer, 1974) and tetraploid $n = 16$ (Puech, 1963; Bonnet, 1963) populations in France have been reported.

1.4 Caryophyllales as a model order for investigating C_4 photosynthesis

1.4.1 Multiple origins of C_4 photosynthesis in Caryophyllales

The Caryophyllales is a big plant order with 11,510 species distributed among 34 families (Bremer *et al.*, 2009). Found across all the continents, members of this order show adaptations to a variety of environments: dry deserts (e.g. succulent cacti), acidic substrates (e.g. sundews), high salinity environments (e.g. members of the Chenopodiaceae family - *Salicornia*, *Salsola*; Yang *et al.*, 2015). Widespread presence in high light and dry environments made Caryophyllales a hotspot for the evolution of CO_2-concentrating complex traits on the tree of life (Christin *et al.*, 2014): C_4 photosynthesis has arisen at least 23 times out of minimum 62 times across the flowering plants in total (Sage *et al.*, 2011) and multiple CAM origins have also been noted. Disproportionately frequent C_4 photosynthesis emergence in the Caryophyllales has been hypothesised to be a result of a combination of ecological, life history, anatomical and genomic factors (Sage, 2001; Christin *et al.*, 2013; Griffiths *et al.*, 2013); therefore this group lends itself well to the study of the trait in order to define the minimum number of changes required to the plant's genome to gain it. In grasses, the transition from C_3 to C_4 photosynthesis was statistically linked to the migration to more arid, open-space environments such as savannas from forest understory (Edwards & Smith, 2010) but no direct link to abiotic stress was uncovered. In contrast, in the Amaranth family, faster evolution of C_4 photosynthesis

was directly correlated with drought resistance (Kadereit *et al.*, 2012). Also, in grasses, certain anatomical changes present in one grass subclade's common ancestor were shown to facilitate multiple origins of C_4 photosynthesis within it (Griffiths *et al.*, 2013).

1.4.2 Physiology of C_4 photosynthesis

C_4 and CAM plants fix inorganic carbon by the coupled action of carbonic anhydrase and PEPC into a four-carbon organic acid in the mesophyll cells (Osmond, 1978; Hatch, 1987). In C_4 plants, the acid is then exported for subsequent CO_2 release for re-fixation in the C_3 cycle by ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) inside bundle sheath cells with morphology allowing elevated CO_2 concentration. The CO_2-concentrating mechanism of C_4 plants results in 10-100-fold increase of $[CO_2]$ around Rubisco in bundle sheath cells, compared to ambient air (Hatch, 1987). As such, the C_4 cycle can be viewed as an ancillary cycle to the Calvin cycle, which through elevated CO_2 concentrations, eliminates photorespiratory losses due to wasteful oxygenation reaction of Rubisco at low $[CO_2]$, however, at the expense of additional ATP used in the cycle (Sage *et al.*, 2012).

The evolution of CO_2-concentrating mechanisms provides an advantage in environments leading to low intracellular concentrations of CO_2 through one, or a combination of factors (Sage *et al.*, 2012). This is because the enzyme Rubisco responsible for initial fixation of CO_2 in the first step of the Calvin cycle also shows affinity for oxygen, which out-competes CO_2 for Rubisco active sites at lower $[CO_2]$. The oxygenation activity of Rubisco results in decreased photosynthesis efficiency by up to a third, because C_3 cycle product 3-phosphoglycerate is produced at a reduced rate and a toxic by-product 2-phosphoglycolate is generated which needs to be recycled using an additional C_2 cycle – this is the process of photorespiration, which incurs the net loss cost of at least 25% of the initially fixed carbon, nitrogen in the form of ammonia, and additional expenditure of ATP and NAD(P)H (Ogren, 1984). It has to be noted that at the same time photorespiration's inefficiency can have a positive role in dissipating extra light energy and thus limiting photooxidative damage and photoinhibition, particularly under drought and high salinity conditions when the stomata are closed leading to reduced rates of CO_2 assimilation (Wingler

et al., 1999, 2000). Other factors that promote photorespiration are: low atmospheric [CO_2], high temperatures decreasing CO_2 solubility at a faster rate than that of oxygen (Sage *et al.*, 2012; Lundgren *et al.*, 2015). However, as a result, C_4 plants have higher photosynthetic efficiency than C_3 plants at higher temperatures: at 25°C and the current atmospheric [CO_2] of 400 ppm, C_4 plants reach 30% higher maximum photosynthetic energy conversion efficiencies than C_3 plants (Zhu *et al.*, 2008**b**). C_4 photosynthesis allows almost total elimination of photorespiration (Skillman, 2008), but the extra energy expenditure in C_4 photosynthesis results only in increased net photosynthetic efficiency under conditions of high photorespiration. Therefore, the trait is found among species inhabiting areas promoting high photorespiration, such as tropical savannas and grasslands, which contribute as much as one quarter to terrestrial primary production (Still *et al.*, 2003).

The key anatomical change seen in C_4 plants is termed Kranz (German: wreath) anatomy, and is an arrangement of two layers made up of different cell types in a concentrical manner around the leaf vascular bundle, which is necessary to isolate Rubisco from the ambient air (Hatch, 1987; Nelson, 2011). The outermost layer in contact with the intercellular air space consists of mesophyll cells, and the innermost of enlarged bundle sheath cells forming a ring around the vasculature, which are connected to the mesophyll via numerous plasmdesmata facilitating metabolite flux. C_4 photosynthesis is partitioned between the two cell types with the initial carbon fixation into the four-carbon acid oxaloacetate (OAA) taking place in the mesophyll cell cytoplasm through the action of PEPC. PEPC substrate's phospho*enol*pyruvate (PEP) is regenerated from pyruvate on the return leg of the C_4 cycle in the mesophyll chloroplasts by pyruvate inorganic phosphate dikinase (PPDK).

Subsequently, OAA is reduced in the mesophyll cells to malate or aminated to aspartate for export to the bundle sheath cells by diffusion through plasmodesmata. The C_4 cycle can then be completed in the bundle sheath cells by decarboxylation through one of the three pathways (Furbank, 2011), one of which always dominates (but not necessarily exclusively) in a given species: NAD-malic enzyme (NAD-ME), NADP-malic enzyme (NADP-ME) and phosphoenolpyruvate carboxykinase (PEP-CK). NAD-ME species reduce OAA to malate

in the mesophyll chloroplasts, whereas NADP-ME and PEP-CK species aminate OAA to aspartate in the mesophyll cytoplasm. In NAD-ME species, malate in the bundle sheath is decarboxylated in the mitochondria to release CO_2, which is then fixed by Rubisco in the Calvin cycle, and pyruvate, which is recycled to the mesophyll. More steps are involved in NADP-ME and PEP-CK-type C_4 photosynthesis. In the NADP-ME-type, aspartate is first trans-aminated to OAA, which is then converted to malate by NADP malate dehydrogenase (NADP-MDH), and finally malate is decarboxylated in the chloroplasts by NADP-ME to pyruvate. To complete the cycle, pyruvate is trans-aminated to alanine (with incoming aspartate used as substrate) and alanine is transported back into the mesophyll, where the reaction is reversed and pyruvate released. In PEP-CK-type plants, it is OAA rather than malate that is directly decarboxylated – this reaction takes place in the cytosol and is catalysed by PEP-CK. The product of the reaction, PEP, is then converted to pyruvate, which then follows the same route as in NADP-ME type C_4 photosynthesis.

1.4.3 Human interest in dissecting the C_4 photosynthesis complex trait

Higher water-, nitrogen-use efficiencies and yield exhibited by C_4 plants at higher ambient temperatures make them ideal crops for the predicted changes in many ecosystems due to future climate change, chief of which is the predicted 1.1–6.4°C increase in the world average temperature in the next 100 years (Xu *et al.*, 2013). Consequently, considerable efforts have been expended to engineer C_4 photosynthesis into C_3 crops such as rice, to complement the limited roster of C_4 crops, such as maize, sugar cane and sorghum. C_4 photosynthesis is a classic case of plant complex trait, as the effect on the plant's phenotype is underlaid by changes to numerous loci. The function of dozens of photosynthetic enzymes is changed by regulation of their expression patterns and amino-acid substitutions altering their catalytic properties in the new metabolic context, as exemplified by phosphoenolpyruvate carboxylase (PEPC) evolution in C_4 plants (Christin *et al.*, 2014), and development of the leaf is changed resulting in the novel Kranz anatomy (Dengler *et al.*, 1994). Despite the immense importance of C_4 photosynthesis to life on Earth, only its enzymatic machinery has been well characterised. Because C_4 photosynthesis takes place over two cell types and in three subcellular compartments (cytosol, mitochondria and chloroplasts), the

involvement of many metabolite transporters is needed in shuttling intermediate products between different compartments; however, the molecular identity of relatively few of these transporters is known (Bräutigam *et al.*, 2011). Moreover, regulation of C_4 photosynthesis enzymes and the genetic specification of Kranz anatomy is still incompletely understood (Külahoglu *et al.*, 2014). Recent years have delivered a spate of cross-species transcriptomic studies trying to identify novel genes involved in C_4 photosynthesis, with varying success (Bräutigam *et al.*, 2011; Gowik *et al.*, 2011; Aubry *et al.*, 2014; Bräutigam *et al.*, 2014; Külahoglu *et al.*, 2014; Wang *et al.*, 2014).

1.4.4 Differential temperature acclimation ability in C_3 and C_4 photosynthesis

Overall, C_4 plants can assimilate effectively at much higher temperatures than C_3 plants (largely because of very low rates of photorespiration) and also exhibit better photosynthetic rate at high growth than low growth temperatures. In contrast, C_3 plants demonstrate better stability in photosynthetic rate across a wider range of temperatures (Yamori *et al.*, 2014). In general, C_3 plants have a much greater capacity for acclimation, with the potential optimum temperature of photosynthesis ranging from approximately 10°C to 35°C, as opposed to C_4 photosynthesis with possible optima ranging from 26°C to 43°C. While the acclimation ranges of photosynthetically optimal temperature can differ widely in C_3 species, they remain consistent in C_4 plants and show little interspecific differentiation (Yamori *et al.*, 2014). At temperatures above the optimum, photosynthesis is hypothesised to proceed more slowly due to a reduced activation state of Rubisco because of decreased activity of its activating enzyme Rubisco activase and/or limitations in electron transport (Yamori *et al.*, 2014). Regarding the minimum thermal barrier of acclimation response, C_4 photosynthesis typically becomes inoperative at temperatures < 15°C with no photosynthetic activity (Wang *et al.*, 2008**a**; Sage *et al.*, 2011). The progressive loss of photosynthetic capacity observed in most C_4 species (but not in all, see e.g. the C_4 grass *Miscanthus* × *giganteus*) at temperatures below 18°C is due to slow photoinhibition of the photosystems (Long *et al.*, 1994). As the capacity for assimilation of carbon at low temperatures is decreased, excitation pressure on the thylakoid membrane is increased which promotes photoinhibition. Another factor

influencing poorer temperature plasticity of C_4 photosynthesis is the lability of C_4 cycle enzymes, in particular PPDK (Naidu *et al.*, 2003; Wang *et al.*, 2008**a**).

1.4.5 Selection of C_3 and C_4 Caryophyllales species to study temperature acclimation

Analysis of gene expression differences important for physiological adaptation in C_3 and C_4 species will be undertaken for two widespread species in the Amaranthaceae: *Alternanthera pungens* Kunth (common name: khakiweed; C_4 species) and *Alternanthera philoxeroides* (Mart.) Griseb. f. *angustifolia* Suess. (common name: alligator weed; C_3 species, referred to as *Alternanthera angustifolia* here). Secondly, another goal is to perform the first-ever comparison of transcriptomes of pairs of closely related Caryophyllales species separated across ecological boundaries: warm and cold/salinity-adapted C_3 species - *Schiedea globosa* H. Mann and *Honckenya peploides* (L.) Ehrh. (common name: sea sandwort). The transcriptomes of the four species will be studied at different temperatures: 18°C and 30°C, and also 25°C in the case of *H. peploides* and *S. globosa*. This is because temperature has been shown to exert a differential effect on the photosynthetic activity of plants and the abiotic stress response depending on the adaptations to native conditions found in the species (Berry & Björkman, 1980).

Alternanthera Forssk. is the second-largest genus in subfamily Gomphrenoideae of Amaranthaceae, with the number of species estimated at 80-200, depending on the circumscription. The high morphological variation found within the genus has also led to a number of subspecies, varieties and forms described in individual species over time. The native range of *Alternanthera* is mainly constrained to the Neotropics (Mears, 1977; Sánchez-Del Pino *et al.*, 2012), with a number of species introduced elsewhere. The genus contains a single independent derived lineage of C_4 species (NADP-ME subgroup: Sage *et al.*, 2007, 2011), to which *A. pungens* belongs, and interestingly C_3–C_4 intermediate species are also found in a clade containing C_3 species (Gowik *et al.*, 2006; Sage *et al.*, 2007). According to Sánchez-Del Pino *et al.* (2012), the phylogeny of *Alternanthera* shows that *A. pungens* is situated in a subclade B3, which is sister to a 100% supported subclade B2 containing the C_3

species *A. angustifolia*. Consequently, the species pair represents an appropriate contrast for studying transcriptomic differences between C_3 and C_4 species from the phylogenetic perspective.

Both *Honckenya peploides* (Figure 1.7 *top*) and *Schiedea globosa* (Figure 1.7 *bottom*) are perennial C_3 herbs (Gagné & Houle, 2002; Wallace *et al.*, 2009), with *Honckenya peploides* and *Wilhelmsia physodes* forming the extant sister group to *S. globosa* according to a chloroplast phylogeny (Sakai *et al.*, 2006). Because of this relatively close phylogenetic relationship, *S. globosa* and *H. peploides* form an appropriate pair for studying the transcriptomic differences important for adaptation to their native climates: tropical and temperate-subarctic, respectively.

The two of the *Alternanthera* species in this study are procumbent herbs that have spread widely from their native ranges. *A. angustifolia* (Figure 1.6 *top*) is an invasive species originating in the Parana River region of South America (Gunasekera & Bonila, 2001), which has now spread to almost all tropical and subtropical parts of Australia, Asia and North America (Julien *et al.*, 1995) and has become established as a problem weed of 10 crop species in 32 countries (Geng *et al.*, 2006; Wang *et al.*, 2009**b**). One of the characteristics of *A. angustifolia* that has allowed it to colonise diverse habitats is its adaptability to different water regimes: it is mainly an aquatic species found in streams, swamps and irrigation channels, but it can also propagate in drier habitats, such as croplands and roadsides (Pan *et al.*, 2005; Wang *et al.*, 2005). In aquatic ecosystems, this alien invader can overgrow the entire surface of the water body, making access to boat routes more difficult and having a negative impact on the native ecosystem's food web due to reduced oxygen exchange and water flow (Wang *et al.*, 2008**b**; Dong *et al.*, 2012); in terrestrial ecosystems, its growth can lead to loss of crop yield and reduction in the diversity of the native flora (Sainty *et al.*, 1997). This species has also been shown to be quite resistant to heavy metal toxicity, salinity and herbicides (Balagtas-Burow *et al.*, 1993; Naqvi *et al.*, 1993; Eberbach & Bowmer, 1995; Naqvi & Rizvi, 2000). As with most weeds, it can reproduce vegetatively with roots and stolons, in addition to sexual reproduction (Buckingham, 1996) and asexual mode of

reproduction facilitates its spread across novel habitats (Julien *et al.*, 1992).
Alternanthera pungens (Figure 1.6 *bottom*) is another, although much less studied tropical invasive weedy species (Robertson, 1981; Sánchez-Del Pino *et al.*, 2012) that carries out C_4 photosynthesis (Rajagopalan *et al.*, 1993). This plant is widely used in folk medicine and as such it has mostly been researched because of its putative medicinal qualities (Calderón *et al.*, 1997; Djohan *et al.*, 2009; Guédé *et al.*, 2010).

Schiedea globosa is one of the 32 species in the genus *Schiedea*, which is a group endemic to Hawaii and a result of the single adaptive radiation event that has resulted in a cluster of species showing high diversity of mating systems, morphologies and habitats (Weller *et al.*, 1995; Filatov & Burke, 2004). The genus *Schiedea* shows the highest diversity in breeding systems of all Hawaiian angiosperm genera (Sakai *et al.*, 2006) and *Schiedea globosa* has so far attracted scientific interest because of its unusual mating system - it is subdioecious, meaning it has 'female' or 'male' individuals with the occasional occurrence of hermaphrodite flowers in staminate inflorescences (Weller *et al.*, 1990). *S. globosa* is the most widespread species in the genus and it can be found on steep, north-facing coastal rocky slopes or cliffs (Wallace *et al.*, 2009).
Honckenya peploides is also found in coastal environments but in much colder climates, from arctic to temperate regions (Sánchez-Vilas *et al.*, 2010). *H. peploides* is an important pioneer species on upper beaches, where its clonal clumps form so-called embryo dunes, the youngest type of dune formation (Sánchez-Vilas *et al.*, 2012). There, the plants face a number of stresses, such as low nutrient availability, drought, sand movement and erosion, and high salinity both from the saltwater spray and substrate (Gagné & Houle, 2002). *H. peploides* is a halophyte well-adapted to those harsh environmental conditions and has been studied as an example of stress-tolerant species facilitating dune establishment. *H. peploides* is subdioecious, just as *S. globosa*, and it can also reproduce clonally through stoloniferous rhizomes (Sánchez-Vilas *et al.*, 2010).

Figure 1.6: **Top:** *Alternanthera philoxeroides* (Mart.) Griseb. f. *angustifolia* Suess. growing in New South Wales, Australia. Photograph courtesy of Doug Beckers, used under Creative Commons license with permission.
Bottom: *Alternanthera pungens* Kunth growing in Maui, Hawaii. Photograph courtesy of Forest & Kim Starr, used under Creative Commons license with permission.

Figure 1.7: Top: *Honckenya peploides* (L.) Ehrh. growing in Örö, Finland. Photograph courtesy of Henna Ke, used under Creative Commons license with permission.
Bottom: *Schiedea globosa* H. Mann growing in O'ahu, Hawaii. Photograph courtesy of David Eickhoff, used under Creative Commons license with permission.

1.5 Research aims statement

The rapid spread of next-generation sequencing into almost every part of biological research has in particular left its mark on study of non-model organisms by providing ready access to previously unavailable genetic blueprint underlying the diversity of ecological adaptation observed in nature. From butterfly mimicry to dessication tolerance in resurrection plants, complex traits are building blocks of the most compelling natural phenomena, and their dissection in non-model organisms has been made considerably easier in the age of a $1,000 human genome. Drawing on this opportunity, two highly ecologically relevant complex traits will be investigated for the first time in non-model species in this work.

First of all, the *Alyssum serpyllifolium* species complex is a prime target for studying the evolution of nickel hyperaccumulation complex trait, as it contains both serpentine and non-serpentine populations with well documented differences in Ni hyperaccumulation and tolerance ability. Here, RNA-Seq will be used for the first time to evaluate genetic divergence between serpentine and non-serpentine *Alyssum serpyllifolium* populations in a genome-wide manner, as reflected in the degree of gene expression differences and sequence polymorphism. Species-wide and population-specific candidate genes important for the Ni hyperaccumulation phenotype and serpentine adaptation will be proposed based on statistical interpretation of the data obtained. Detected patterns of neutral and adaptive variation between serpentine and non-serpentine populations will then be employed to evaluate simple evolutionary scenarios of hyperaccumulation trait origin and spread that could have resulted in the modern species complex.

Second of all, the power of next-generation sequencing for illuminating the genetic basis of another complex trait – C_4 photosynthesis – will be tested through application of the RNA-Seq bioinformatics approaches developed earlier to two non-model *Alternanthera* species in the plant order Caryophyllales. Using differential gene expression analysis, the study will investigate the genetic basis for differential temperature acclimation ability in C_3 and C_4 species, including mechanisms such as alternative splicing.

Chapter 2

Identification of Candidate Ni Hyperaccumulation and Serpentine Adaptation Genes Using Next-Generation Sequencing

2.1 Introduction

It is expected that differences between ecotypes/species due to adaptation to different environments should, at least partially, be reflected in gene expression differences and previous research has confirmed that (López-Maury *et al.*, 2008). The majority of genes shown to be important for the zinc hyperaccumulation trait were revealed to be top differentially expressed genes when comparing *Arabidopsis halleri*/*Noccae caerulescens* and *Arabidopsis thaliana* transcriptomes using cDNA oligonucleotide microarrays (Becher *et al.*, 2004; Weber *et al.*, 2004; Verbruggen *et al.*, 2013). Knowing that relative high gene expression level seems a good indication of functional importance in hyperaccumulator plants, as is also the case with the only gene with well established function in nickel hyperaccumulation in *Alyssum* - *ATP-PHOSPHORIBOSYLTRANSFERASE* (*ATP-PRT 1*, Ingle *et al.*, 2005**a**), novel candidate genes responsible for the nickel-hyperaccumulating phenotype and serpentine adaptation in *Alyssum serpyllifolium* will be identified in a RNA-Seq comparison of two replicate limestone and serpentine populations.

Comparison of gene expression variation in closely related, conspecific individuals with differences in a given trait - here, chiefly nickel accumulation and tolerance levels, should

support identification of genes adaptive for it. The conclusion from previous studies on the two zinc/cadmium hyperaccumulator species is that a complex physiological trait such as metal hyperaccumulation can arise due to concerted changes in the expression of key genes (Becher *et al.*, 2004). Based on data available so far, *Arabidopsis halleri* and *Noccaea caerulescens* utilize the genes involved in metal homeostasis in non-accumulator species – ZIP, MTP and HMA family members – but increasing their expression through gene duplication and promoter enhancement (Verbruggen *et al.*, 2009; Krämer, 2010). In the experiment, plants were cultivated under common garden conditions with Ni exposure in serpentine populations, which should highlight candidate genes important for nickel homeostasis rather than other edaphic adaptations, unless their expression is non-plastic and constitutive.

Secondly, genome-wide comparisons of allele frequencies between recently diverged populations can also help us identify the key candidate genes important for adaptation with no *a priori* information. Identification of peaks of high differentiation relative to average genomic background is the first step in establishing genes under divergent selection in populations in unbiased manner (Beaumont, 2005). Genome-wide scans for signatures of selection, in particular fixed sequence differences between serpentine and non-serpentine populations will be conducted here for the first time in the context of nickel hyperaccumulation in *Alyssum*. One way to identify them is by finding genomic regions which fall outside of the expected normal distribution of divergence - so called outliers. Highly diverged loci are likely to have been under the influence of disruptive selection (Beaumont & Nichols, 1996) and the aim was to find them by comparing individual serpentine populations to the two combined non-serpentine populations. It has to be kept in mind that some of the differences seen will be just strictly due to drift and combination of demographic and other selective forces (e.g. background selection in subdivided populations, Lotterhos & Whitlock, 2014), and previous knowledge of functional importance of some of the targets in e.g. metal tolerance, helps to restrict the analysis to more true positive outliers. Furthermore, focusing only on loci showing non-synonymous replacements and belonging to overrepresented categories should further

reduce the number of false positives (De Wit *et al.*, 2015).

Thirdly, variation along various environmental axes is omnipresent in nature and because of that populations of many species display adaptations to selective pressures stemming from locally encountered abiotic and biotic factors (Savolainen *et al.*, 2013; Tiffin & Ross-Ibarra, 2014). Local adaptation to particular environments can be a first step towards ecological speciation driven by divergence in selection pressures between different environments or ecological niches (Rundle & Nosil, 2005), which can occur in the face of continuing gene flow between incipient species. Here, NGS will be also employed to survey the amount of local adaptation seen in the two serpentine populations of *Alyssum serpyllifolium*, using both gene expression profiles and results of genome scans for candidate genes as proxies. Both sampled serpentine populations have previously been classified as populations, subspecies or even distinct species (see Chapter 1, section 1.4) and so it will be interesting to see if the estimated degree of local adaptation seen in each population suggests that they have been indeed on very distinct evolutionary trajectories for some time now.

Aims and objectives

- Development, testing and execution of bioinformatics pipelines to generate and quantify the first *Alyssum* transcriptome based on RNA-Seq data.
- Evaluation of a degree of gene expression differences observed between/within serpentine and non-serpentine populations of *Alyssum serpyllifolium* grown under common laboratory conditions.
- Analysis of the footprints of selection left in the transcriptomes of the *A. serpyllifolium* populations.
- Identification of candidate genes important for adaptation across the serpentine populations, as revealed by the joint evidence of transcriptomics and tests for selection with corresponding isolation of neutral variation present in the genome and useful for phylogenetic and demography inference.
- Identification of key overrepresented pathways and genes among serpentine populations, both as shared between them and population-specific (local).
- Providing first estimates of the degree of isolation, mating system and gene flow between serpentine and non-serpentine *A. serpyllifolium* populations using identified genome-wide neutral genetic markers (SNPs).

2.2 Materials and methods

2.2.1 Plant material

In the pilot experiment, *Alyssum pintodasilvae* T.R.Dudley, henceforth referred to by its population name as Samil-S, and *Clypeola jonthlaspi* seeds (Tables 2.1 and 2.2) were germinated on moist grit sand mixed with coarse vermiculite. Plants were grown in a glasshouse with natural solar irradiation supplemented by sodium-vapour lamps to maintain a 16 hour light period (minimum 23°C daytime, 19°C nighttime temperature). After 2 weeks, seedlings were transferred to 1.2-litre polystyrene boxes (10 plants/box). Plants were grown for 3 weeks on 0.1-strength modified Hoagland solution according to Roosens *et al.* (2003), which was replaced every week. After that time, solution strength was doubled and $NiSO_4$ was added to a final concentration of 30 μM in boxes with Samil-S only to elicit expression of any nickel-inducible genes. Following 3 weeks of nickel exposure, plants were harvested and snap-frozen in liquid nitrogen.

The main experiment included also two non-serpentine populations of *Alyssum serpyllifolium* Desf. (Morata-NS and Rubía-NS), as well as one serpentine population of *Alyssum malacitanum* Rivas Goday (Carratraca-S) (Tables 2.1 and 2.2), in addition to Samil-S and *Clypeola jonthlaspi*. In this experiment, after 3 weeks since germination, seedlings were transferred to hydroponic culture with 0.1-strength Hoagland solution (Roosens *et al.*, 2003) modified to include 0.2 μM $ZnSO_4$. After further 2 weeks, $NiSO_4$ was added to a final concentration of 30 μM in boxes with Samil-S and Carratraca-S, and plants were exposed to nickel for 3 weeks before harvest (Figure 2.1) and cryopreservation in liquid nitrogen.

2.2.2 Population sites and their selection

The economical set-up employed allowed assessment of the differences in gene expression and underlying sequences in the two serpentine *Alyssum serpyllifolium* populations with very similar, high nickel hyperaccumulation/tolerance ability but at the same time of clearly allopatric origin in the species' Iberian range ends in NE Portugal (Samil-S) and SW Spain (Carratraca-S). The comparison was made against two non-serpentine populations

showing similar, relatively low Ni hyperaccumulation/tolerance and situated in allopatry relative to the two serpentine populations and each other.

Previous analyses revealed that the two serpentine areas are generally not dissimilar in terms of soil properties, as compared to non-serpentine areas inhabited by *Alyssum serpyllifolium*. Here, results for a population from the calcareous-dolomitic area of Sierra Nevada (parapatric to Carratraca-S, situated ca. 125 km to the east) taken from Becerra-Castro (2012) are reported, as the only ones available. It has to be also noted that Becerra-Castro (2012) tested a different population of *Alyssum malacitanum* - Sierra Bermeja-S, situated ca. 55 km to the south of Carratraca-S.

Both serpentine soils were slightly acidic (Sierra Bermeja-S: pH 5.6, Samil-S: pH 6.3), while the Sierra Nevada-NS soil was slightly alkaline at pH 7.2. The serpentine soils were characterised by low Ca/Mg ratios < 1 (Sierra Bermeja-S: 0.5, Samil-S: 0.6) with Mg predominant in the exchange complex, as expected (Brooks, 1987), and as compared to Sierra Nevada-'s soil Ca/Mg ratio of 2.9. Carbon content of the serpentine soils was equally low in Samil-S (mean 3.05%) and Sierra Bermeja-S (mean 2.42%) likely reflecting their low organic content and/or increased erosion relative to the non-serpentine reference soil (mean 9.73%). Reduced fertility of serpentine soils was also evident in lower nitrogen content of the serpentine soils (Samil-S: mean 0.28%, Sierra Bermeja-S: mean: 0.21%) relative to the Sierra Nevada-NS soil (mean 0.46%).

Nickel concentration was significantly higher in the Samil-S habitat (mean 3,028 mg/kg) than in Sierra Bermeja-S (mean 2,098 mg/kg), but both soils contained two orders of magnitude more nickel than the non-serpentine Sierra Nevada soil (mean 32 mg/kg). Serpentine soils were also found to contain high total chromium (Samil-S: mean 847 mg/kg, Sierra Bermeja-S: mean 1,407 mg/kg) and cobalt concentrations (Samil-S: mean 141 mg/kg, Sierra Bermeja: mean 104 mg/kg), however, neither of the two metals is readily available for plants in soluble form (Kidd & Monterroso, 2005; Cabello-Conejo *et al.*, 2014).

Mineral content analysis of laboratory- and field-grown specimen of *Alyssum serpyllifolium* has only been conducted in *Alyssum pintodasilvae* (Samil-S) (Freitas *et al.*, 2004; Lazaro

et al., 2006; Cabello-Conejo *et al.*, 2014). Studies in native soils (Lazaro *et al.*, 2006) and metal amended soils (De Varennes *et al.*, 1996) revealed that Samil-S did not display specific root-to-shoot transfer and accumulation of any metal (Cd, Cr, Cu, Mn, Pb, Zn) other than Ni, although like in other serpentinophytes in the community, root:soil metal concentration ratios exceeded 1 and the concentration of Zn and Cr in its tissue was above the toxicity threshold for non-metallicolous plants.

2.2.3 RNA extraction and sequencing

In the pilot experiment, total RNA was extracted with TRIzol reagent (Invitrogen, Carlsbad, USA) from pooled shoots and roots from 20 plants for Samil-S and 10 plants for *Clypeola*. Total RNA integrity was checked on a polyacrylamide gel. One 10 µg pooled total RNA sample separately from root and shoot from each species was used in preparing multiplexed cDNA library and the libraries were sequenced with a HiSeq 2000 instrument (Illumina, San Francisco, USA) in 100 bp paired-end mode at Wellcome Trust Centre for Human Genetics (Oxford, UK) with each sample sequenced over $\frac{1}{12}$ of a lane. Base calls were encoded with Illumina version 1.5+ format (Phred64) but converted to Phred33 before analysis.

In the main experiment, total RNA was extracted from 50 mg fresh mass of tissue removed from the tip of the primary shoot and 100 mg fresh mass of mature root, all collected in the morning (Figure 2.1). RNA was extracted using TRIzol reagent and the aqueous phase containing RNA further purified with RNeasy Mini Kit (Qiagen, Hilden, Germany). For Morata-NS and Carratraca-S, 3 root and 3 shoot samples taken from independent plants were prepared; for Samil-S and Clypeola 2 root and 2 shoot samples taken from independent plants were prepared. Finally, for Rubía-NS, 3 root samples from independent plants were harvested. Total RNA integrity was checked with a Bioanalyzer 2100 (Agilent, Santa Clara, USA). Between 8 and 42 µg of total RNA per sample was used in preparation of each multiplexed cDNA library. All samples except for Rubía-NS were sequenced together twice over two lanes on a HiSeq 2000 in 100 paired-end mode at the NERC Biomolecular Analysis Facility (Edinburgh, UK). Rubía-NS root samples were then sequenced with a

Table 2.1: Taxa investigated with RNA-Seq

Species	Synonym	Population name	Soil type
Alyssum serpyllifolium Desf.	*Alyssum pintodasilvae* T.R.Dudley	Samil-S	serpentine
Alyssum serpyllifolium Desf.	*Alyssum malacitanum* Rivas Goday	Carratraca-S	serpentine
Alyssum serpyllifolium Desf.	-	Morata-NS	limestone
Alyssum serpyllifolium Desf.	-	Rubía-NS	limestone
Clypeola jonthlaspi L.	-	-	non-serpentine

Table 2.2: Sources of seeds used in the study

Sample name	Latitude / Longitude	Seed collection
Samil-S	41.46° N, 6.45° W	A.J. Pollard (Furman University, USA)
Carratraca-S	36.70° N, 4.60° W	A.J. Pollard, J.A.C. Smith
Morata-NS	40.29° N, 3.49° W	R. D. Reeves (Sheffield University, UK)
Rubía-NS	42.54° N, 6.82° W	T.A. Flynn, J.A.C. Smith
Clypeola	N/A (near Jarash, Jordan)	Millennium Seed Bank (Royal Botanic Gardens, Kew, UK)

HiSeq 2000 in 100 bp paired-end mode at Wellcome Trust Centre for Human Genetics (Oxford, UK) with each sample sequenced over $\frac{1}{16}$ of a lane. In all cases, base calls were encoded with Illumina version 1.8+ format (Phred33).

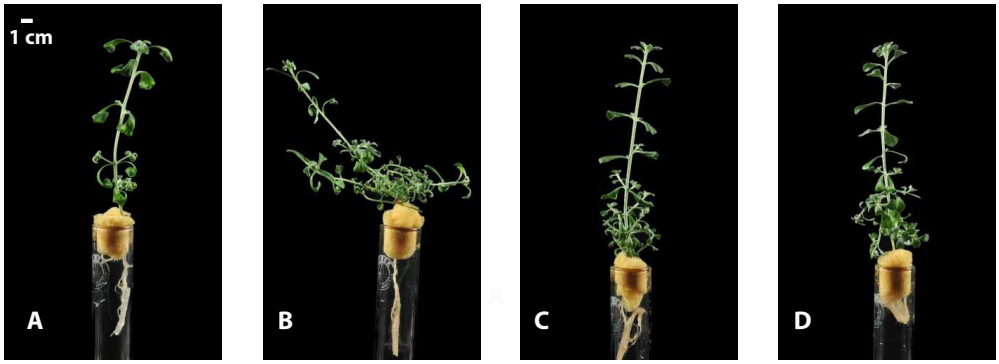


Figure 2.1: Serpentine taxa: **A)** *Alyssum malacitanum* (Carratraca-S), **B)** *Alyssum pintodasilvae* (Samil-S) and non-serpentine taxa: **C)** *Alyssum serpyllifolium* (Morata-NS), **D)** *Alyssum serpyllifolium* (Rubía-NS) grown in hydroponic culture at 8 weeks since germination.

2.2.4 Read quality control

The goal of read pre-processing was to minimise the influence of sequencing errors on the assembly and prevent misassembly (El-Metwally *et al.*, 2013). All FASTQ files with sequencing reads were first examined using Prinseq 0.20.3 (Schmieder & Edwards,

2011) and FastQC 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to confirm the quality of sequencing - ie. consistently high quality scores across the read length, low sequence duplication and sequence over representation levels. Since FastQC detected Illumina adapters and PCR primers amongst over-represented sequences, all such contaminating sequences were then trimmed away with Trimmomatic ver. 0.32 using the authors' recommended settings (Bolger *et al.*, 2014), along with stretches of bases with low Phred quality score (Q < 15 averaged over a 5-bp sliding window) at the ends of reads. Fqtrim 0.94 (https://ccb.jhu.edu/software/fqtrim/) was then used to remove A/T homopolymer stretches (n > 5 bp) indicative of polyA tails, to select reads based on size (n > 20 bp) and to filter out the reads with high N-character (undefined base) content (n >= 5%). Any possible eukaryotic or prokaryotic rRNA contamination was filtered out with SortMeRNA 1.8 (Kopylova *et al.*, 2012) and finally the remaining reads were corrected for sequencing errors with Racer 1.0 (Ilie & Molnar, 2013), which shows very good performance in comparison with other software employing k-mer count but also k-mer spectrum and multiple alignment error correction strategies. A study has shown that adding on an error correction pre-processing step improved assembly using different assemblers, including Velvet (Salzberg *et al.*, 2012) and reduced the number of errors should diminish the number of false positive SNPs identified (De Wit *et al.*, 2015).

2.2.5 Assembly and mapping

In the study, two different types of read mappers were used: based on Burrows-Wheeler transform (BWA) and seed method (Stampy). Stampy was employed in mapping reads to reference sequences from different species: *Alyssum* and *Clypeola* reads in the Arabidopsis transcriptome pre-mapping step prior to the assembly and later when mapping *Clypeola* reads to the *Alyssum* reference to define the ancestral status of each allele. Seed methods utilise only short initial match (ie. seed), then further extended in both directions, which allows high sensitivity required for interspecific read mapping. On the other hand, for standard read mapping of reads from each species to their own reference Burrows-Wheeler Aligner (BWA) was employed, as it uses hash tables search for first-pass identification of candidate alignment locations, which results in huge speed-up at vastly

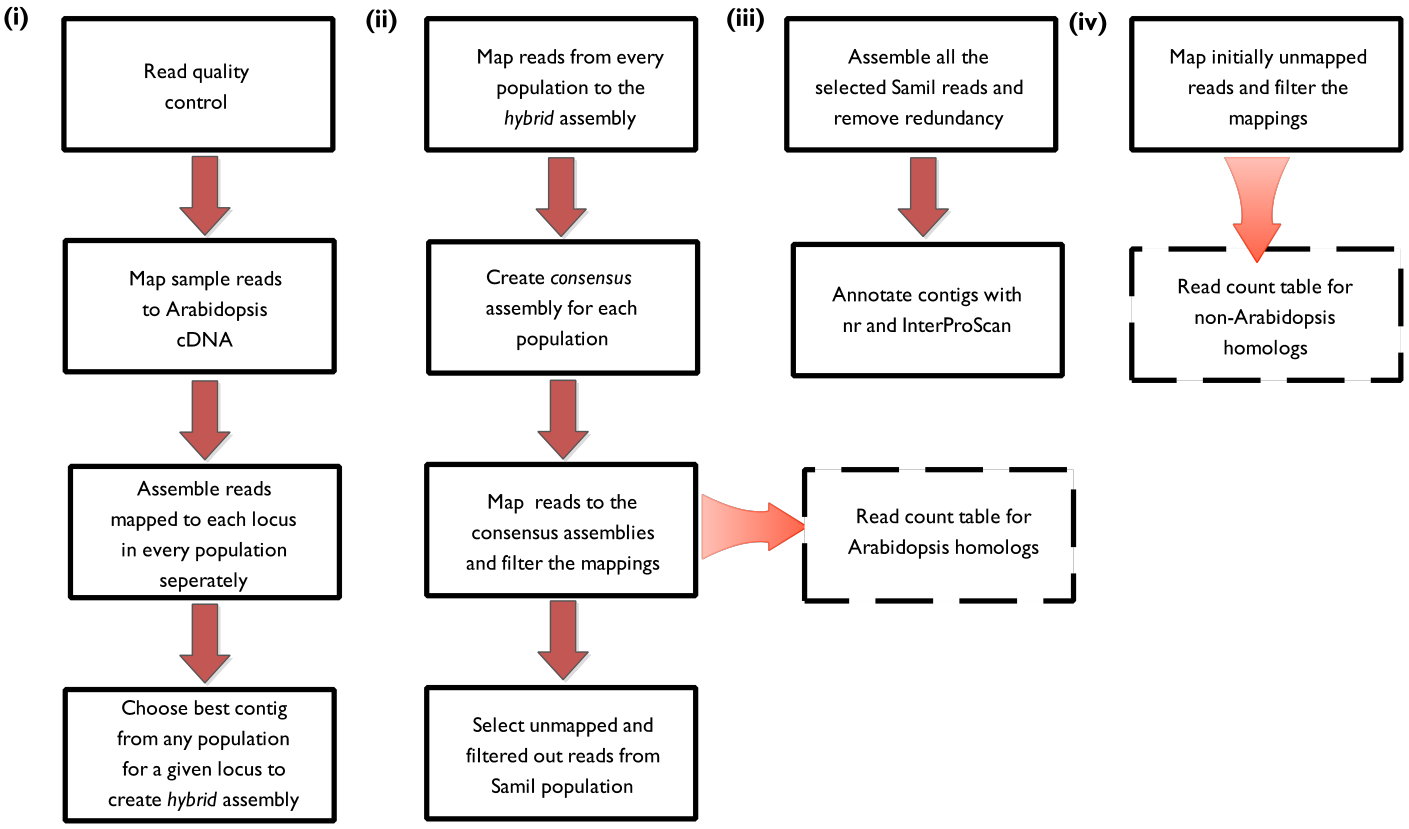


Figure 2.2: Outline of the workflow used in generating the *Alyssum serpyllifolium* transcript sequences and expression estimates

higher computational cost when more diverged reads are considered (Oshlack *et al.*, 2010; Garber *et al.*, 2011).

Reads were first mapped to *Arabidopsis thaliana* TAIR10 (Lamesch *et al.*, 2011) representative BLAST set of transcripts using Stampy 1.0.20 (Lunter & Goodson, 2011) independently for each individual (default settings, except for: `-substitutionrate 0.15`) as it employs hash-based indexing and can accurately map sequences with at least up to 15% divergence to the reference (Sedlazeck *et al.*, 2013; Vijay *et al.*, 2013). Mapping output was examined using ea-utils.1.1.2 (https://code.google.com/p/ea-utils/) and custom Perl scripts to confirm good mapping rate. The read pairs where both reads mapped uniquely to the same gene in all the single individual samples were extracted. All such extracted reads mapping to a given gene from across all the individuals in a population were pooled and reads assembled

using Oases 0.2.8 (Schulz *et al.*, 2012) with the following settings:

```
$ velveth 31 -fastq -shortPaired
$ velvetg -read_trkg yes -cov_cutoff 4 -scaffolding yes -min_contig_lgth
200 -ins_length 180 -ins_length_sd 50 -min_pair_count 8
$ oases -ins_length 180 -ins_length_sd 50 -scaffolding yes -cov_cutoff
4 -min_trans_lgth 300 -edgeFractionCutoff 0.15 -min_pair_count 8
```

Contigs output by Oases were subsequently annotated using the BLASTx algorithm (Altschul *et al.*, 1997) implemented in BLAST 2.2.26+ suite with *Arabidopsis thaliana* TAIR10 proteome used as a reference with maximum 10^{-4} hit Expect value (e-value) cut-off. All contigs with a top hit to a given gene were then evaluated with custom Perl scripts to select the representative sequence for each *Arabidopsis* locus present in the assemblies. Since coverage and nucleotide diversity for each gene can differ across populations, some transcripts were missing in some populations or assembled to transcripts of lower quality (low contiguity etc.) than in others, so it was not possible to select one population assembly as a reference. Instead, a 'hybrid' assembly comprising transcripts from all the four populations was constructed as follows. The top bit-score contigs for each gene in each of the four *Alyssum* populations were considered to select one with:

- no N characters
- no premature stop codons within the high-scoring segment pairs (HSPs)
- no HSPs in different frames
- highest bit score (and shortest length if more than one contig of the same score was found)

Only one population of *Clypeola* was sequenced, so reads were assembled using two additional *k*-mer sizes (21 and 41), instead, because using multiple *k*-mer sizes can improve assembly of genes showing different levels of expression (Surget-Groba & Montoya-Burgos, 2010; Zhao *et al.*, 2011). The best *Clypeola* contigs were then selected in an analogous

manner using criteria listed above.

The selected contigs were then annotated with putative 5′ UTR, CDS and 3′ UTR regions. If a start codon aligning with the 1 position of the *Arabidopsis* reference protein was not found in the alignment, any first methionine present upstream of the aligned portion of the protein was treated as the start codon, or no 5′ UTR region was defined in the transcript unless a start codon was found. Similarly, if no stop codon was present downstream from the aligned portion of the gene, no 3′ UTR region was defined and all the nucleotides up to the 3′ end were treated as protein-coding.

To try to improve mapping efficiency within individual *Alyssum* populations, individual 'consensus' assemblies were created for each population based on the 'hybrid' reference assembly described above (Figure 4.2 (ii)). Reads from all the samples in a given population were mapped to the 'hybrid' reference using STAR 2.3.0 (Dobin *et al.*, 2013) with the following settings:

```
$ star –outFilterMismatchNmax 100 –seedSearchStartLmax 25
–outFilterMultimapNmax 1
```

SAMtools 0.1.17 (Li *et al.*, 2009**a**) was then used to convert the SAM output files to BAM format, and to generate population consensus sequence based on mpileup-called variants, using the following commands:

```
$ samtools mpileup –q 20 –d 10000 –L 10000 –Q 20 –m 5 –F 0.2 –o 20
–uf | bcftools view –cg – | vcfutils.pl vcf2fq | seqtk fq2fa
```

The resulting contigs were parsed with a custom Perl script to create the final 'consensus' transcriptome for every *Alyssum* population. If the new 'consensus' contig did not contain any ambiguity IUPAC characters, it superseded the 'hybrid' assembly contig; in cases in which ambiguous characters were present in the population 'consensus' assembly was

preferentially used whenever possible.

Mapping of reads to the individual population 'consensus' assemblies was then carried out with BWA mem 0.7.7 (options `-U 15 -T 0`; Li, 2013). Converted BAM files were then filtered with SAMtools to remove chimeric mappings and discordant read pairs as follows:

```
$ samtools view -b -f 2 -F 2048
```

Each BAM file was then inspected in Tablet 1.13 (Milne *et al.*, 2013) and read count table exported as input for differential expression analysis. Remaining reads, ie. those initially unmapped and removed after the filtering step, were then pooled from all the single individual Samil-S samples and assembled to resolve any transcripts with weak or no homology to *Arabidopsis* or fragments from *Arabidopsis* homologs already present in the dataset (Figure 4.2 (iii)). Samil-S reads were chosen because of the primary interest in highly expressed genes in hyperaccumulators; and introducing additional polymorphism from other populations would result in more fragmented and redundant assembly (Haas *et al.*, 2013; Singhal, 2013). Reads were assembled with Oases as before, but this time using *k*-mers 21 to 51, step size 10. The four assemblies were then merged with `merged` command in Oases to remove redundant transcripts in the following way:

```
$ velveth merged 31 -long
$ velvetg merged -read_trkg yes -scaffolding yes -conserveLong yes
-min_contig_lgth 300
$ oases merged -merge yes -scaffolding yes -min_trans_lgth 300
```

The merged transcripts were then clustered additionally with USEARCH 7.0 (Edgar, 2010) using the settings below and finally contigs extended with CAP3 (Huang & Madan, 1999) using default settings:

```
$ usearch -cluster_smallmem -id 0.9 -centroids
```

The resultant assembly, henceforth referred to as 'supplementary', was then searched against the *Arabidopsis* proteome, using BLASTx algorithm as previously, to find contigs with significant hits to *Arabidopsis*. Contigs not annotated after that procedure were then searched with BLASTx against nr and InterPro reference databases (Jones *et al.*, 2014) with Blast2GO 2.7.1 (Götz *et al.*, 2008), default settings but maximum e-value of 10^{-5}. All the remaining *Alyssum* reads were mapped to the output contigs separately for each individual using BWA mem (Figure 4.2 (iv)), with identical settings to those used previously. In the case of *Clypeola*, reads were mapped with Stampy as previously described, making use of its higher sensitivity when mapping to more diverged sequences. Sedlazeck *et al.* (2013) have shown that Stampy shows no decline in read mapping rate up to at least 10% divergence and mean coding region sequence divergence between *Alyssum* and *Clypeola* stands at 5.5% (Figure 2.6). All the BAM files were again filtered to remove discordant read pairs and chimeric reads with SAMtools. Reads mapping to the *Arabidopsis* loci already present in the read count matrix were simply summed and new entries created for contigs annotated using nr or InterProScan hits.

2.2.6 Differential expression analysis

Read count matrix including all sequenced samples was imported into edgeR 3.6 (Robinson *et al.*, 2010) and DESeq2 1.4 (Love *et al.*, 2014) R packages. These two packages were selected, because of their consistent top performance in various comparative studies (Guo *et al.*, 2013; Rapaport *et al.*, 2013; Robles *et al.*, 2012; Seyednasrollah *et al.*, 2013). Genes in the data set were filtered to include only those showing expression higher than 1 count per million in at least 3 samples, resulting in 20 483 out of 21 149 genes being retained for differential expression analysis. Extremely lowly expressed genes are of little biological interest and are unlikely to be found differentially expressed but eliminating them and thus reducing the number of tests made will increase power of the remaining statistical comparisons (Chen *et al.*, 2014**b**). `PlotMDS` and `plotPCA` functions in edgeR and DESeq2, respectively, as well as a core R function (R Core Team, 2014) `prcomp` were then used to visualise the general gene expression similarity of all the samples.

Generalized linear model (GLM) design feature from the two packages was used to fit a model including all the samples with two specified factors (*tissue* and *population*) in order to establish better gene dispersion estimates and facilitate making multiple contrasts (Chen *et al.*, 2014**b**). In DESeq2, function `DESeq` was called to carry out the default statistical analysis on the filtered data set and contrasts of interest were investigated with `results` function. In edgeR, library sizes were first normalised with `calcNormFactors` function followed by dispersion estimation with `estimateGLMCommonDisp`, `estimateGLMTrendedDisp` and `estimateGLMTagwiseDisp` before fitting gene-wise negative binomial GLMs with `glmFit` function using the calculated dispersion values. Individual pairwise comparisons were then accessed using `makeContrasts` and `glmLRT` functions. DESeq and edgeR utilise similar strategies to detect differentially expressed (DE) genes, but with no clear superiority of either method (Soneson & Delorenzi, 2013; Robles *et al.*, 2012), and edgeR especially can suffer from a high false positive rate. To mitigate that, genes output as significantly differentially expressed, were only retained if they remained so after applying Benjamini-Hochberg procedure to control for false discovery rate (FDR) < 0.01 in both packages.

Finally, the results from all the individual contrasts were grouped into various general patterns of gene expresssion: upregulated/downregulated/not differentially expressed in the shoot/root based on 'all the serpentine versus all the non-serpentine' (S vs NS), ie. Carratraca-S, Samil-S versus Morata-NS, Rubía-NS, *Clypeola*; or 'all *Alyssum* versus *Clypeola*' comparisons, ie. all *Alyssum serpyllifolium* populations versus *Clypeola*. For 'all the serpentine versus all the non-serpentine' comparison, mean-centred $\log_2$ reads per kilobase of exon model per million mapped reads (RPKM) (Mortazavi *et al.*, 2008) values were plotted as a line graph with R `plot` function. For 'all *Alyssum* versus *Clypeola*' comparison individual gene expression patterns were first hierarchically clustered using Euclidean distance measure and average-linkage clustering criterion with R `hclust` function and plotted as a heatmap with `heatmap.3` function. Enrichment of gene ontology (GO) terms (Rhee *et al.*, 2008), gene families, PlantCyc pathways (Caspi *et al.*, 2013),

and literature datasets in the different groupings was investigated with PlantGSEA web-server (Yi *et al.*, 2013) (Fisher's exact test with FDR < 0.05 significance cut-off), while enrichment of KEGG pathways (Kanehisa *et al.*, 2014) and InterPro domains was analysed with GeneCodis3 web-server (Tabas-Madrid *et al.*, 2012) (hypergeometric test with FDR < 0.05 significance cut-off). Fisher's exact test and its equivalent, hypergeomteric test, were chosen since they can be appplied to samples of any size (Primmer *et al.*, 2013).

2.2.7 SNP and genotype calling

All the *Alyssum* samples were first mapped to the Carratraca-S 'consensus' transcriptome and the 'supplementary' transcriptome used previously in DE analysis with BWA mem (settings as before). *Clypeola* samples were first mapped to the Carratraca-S 'consensus' transcriptome with Stampy (settings as before) and separately to the *Clypeola* transcriptome with BWA mem; the former also included mapping initially unmapped *Clypeola* reads to the 'supplementary' transcriptome, done with Stampy.
In addition to applying the previous filtering strategy to remove chimeric mappings and discordant read pairs from the resultant BAM files, any multimapping reads with MAPQ = 0 were filtered using the following command in SAMtools:

```
$ samtools view -b -q 1
```

PCR and optical duplicates were marked in all the samples with MarkDuplicates module from Picard Tools ver. 1.114 (http://picard.sourceforge.net/) to exclude them from further analysis (`MAX_FILE_HANDLES_FOR_READ_ENDS_MAP=1000 OPTICAL_DUPLICATE_PIXEL_DISTANCE=10` options). Increase in coverage of one allele due to read duplication can otherwise result in miscalling heterozygotes as homozygotes and errors introduced in PCR duplicates can be treated as true variants by genotype callers (Andrews & Luikart, 2014). To allow multi-sample SNP calling, all the individual BAM files were tagged with sample and group names with bamaddrg (https://github.com/ekg/bamaddrg, default settings). SNP calling was carried out on all the *Alyssum* samples simultaneously using 3 popular and accurate variant callers (Cheng *et al.*, 2014): freebayes ver. 0.9.14-2 (Garrison & Marth,

2012) using default settings, GATK Unified Genotyper ver. 3.1-1 (DePristo *et al.*, 2011) (options `-stand_call_conf 5 -stand_emit_conf 5` to relax stringency of filtering of genotypes to be output), and SAMtools:

```
$ samtools mpileup -d 1000 -L 1000 -o 20 -C 50 -u -D
$ bcftools view -c -g -v
```

Variant calls from the three programs were then integrated with BAYSIC (Cantarel *et al.*, 2014), leading to retention of only calls made by all of the three variant callers. This step was undertaken due to low concordance of different variant-calling pipelines reported in the literature (O'Rawe *et al.*, 2013) to reduce the number of false positive variants. These were then further parsed to exclude variant calls with the quality score below 20. Next, individual genotypes produced by SAMtools were chosen for further analysis (minimum read depth: 8; minimum genotype quality: 20) and, when unavailable, freebayes genotypes were used. In all the analyses listed below, sole focus was placed on simple biallelic SNPs and excluded loci for which high observed heterozygosity deviating from Hardy-Weinberg equilibrium ($H_e > 0.5$) in any of the populations suggested the possibility of unresolved paralogy, which would create artefactual genetic variation (Stapley *et al.*, 2010).

SNPs in a given contig were phased using HapSeq ver. 2 (Zhang & Zhi, 2013). The following options were used: set Hidden Markov Model (HMM) rounds to 50, with 10 discarded as burn-in (`rounds` and `burn-in` options). In addition, `readForwOpt` was used to perform HMM search with haplotype data, and `mhPhasing` and `mhRounds` to perform 10 Metropolis-Hastings sampling rounds using the read data, with 5 rounds used as burn-in (`mhBurnin` option). Sequencing error rate (`seqError` option) was set to 0.01, and 36 possible reference haplotype states were considered in HMM computation (`states` option).

2.2.8 Population structure

Principal Component Analysis (PCA) was conducted on the matrix containing all simple biallelic SNPs genotyped in all the individuals using the `smartpca` module in the EIGENSOFT package ver 5.0.2 (Patterson *et al.*, 2006). Structure ver. 2.3 (Pritchard *et al.*, 2000; Falush *et al.*, 2003) was used to confirm clear delimitation of the four populations under study. As recommended in the manual, λ parameter was first inferred once for $K = 1$, which converged stably around 1, and that value was used for analyses with other K. In total, 13,050 independent SNPs resident in different putatively neutral transcripts were sampled. F_{ST} outliers were excluded, and five replicates for each value of K from 1 to 5 were run with 50,000 burn-in and 500,000 iterations. Structure output was analysed and summarised with Structure Harvester Web ver. 0.6.94 (Earl & vonHoldt, 2012). To find the optimal alignment of clusters across all replicate runs and visualise it, the Greedy search algorithm was applied in CLUMPP ver. 1.1.2 (Jakobsson & Rosenberg, 2007) using the G' pairwise matrix similarity statistic with 1000 random input orders tested, weighted by the number of individuals in each population.

2.2.9 Tests for selection

A combination of PopGenome ver. 2.0.7 (Pfeifer *et al.*, 2014) and custom Perl scripts was used to calculate basic SNP frequency statistics, including number of segregating sites and minor allele frequency spectra, for each *Alyssum* population. PopGenome was also used in calculating various population genetics summary statistics, such as Tajima's D (Tajima, 1989), Fu and Li's D (Fu & Li, 1993), Fay and Wu's H (Fay & Wu, 2000), Zeng's E (Zeng *et al.*, 2006) and Watterson estimator (Watterson, 1975), as well as carry out a McDonald-Kreitman (MK) test (McDonald & Kreitman, 1991) with *Clypeola* as an outgroup. Ancestral status for each SNP was established by mapping *Clypeola* reads to the *Alyssum* reference transcriptome using Samtools, as described previously. P-values in the McDonald-Kreitman test were calculated using Fisher's Exact Test (`fisher.test`) function in R. In some analyses, Rubía-NS population was excluded due to low sample size (three individuals), which would introduce additional bias. All simple biallelic SNPs genotyped in at least five individuals in Morata-NS and five individuals in Carratraca-S or

Samil-S populations were scanned for F_{ST} outliers using Bayescan ver. 2.1 (Foll & Gaggiotti, 2008). Bayescan analysis allows comparison of posterior probabilities of two models for each SNP, one with selection and one without, using posterior odds ratio. The method results in low number of false positives and robust power under different demography scenarios (Mita *et al.*, 2013; Lotterhos & Whitlock, 2014; Villemereuil *et al.*, 2014). For each test, Bayescan was run for 100,000 iterations with 50,000 burn-in. FDR cut-off of 5% was used. The significance of the extent of overlap between outlier SNPs and loci in Carratraca-S and Samil-S populations was checked by bootstrapping 1,000 times in R.

Pairwise d*N*/d*S* ratios were calculated for each population against *Clypeola* orthologs, in turn, to detect genes with high rates of nonsynonymous substitutions. d*N*/d*S* comparisons within the two species pairs were as follows: a reciprocal Blast approach (BLASTn, E-value threshold = 10^{-10}) was used to identify orthologue contig pairs amongst the two assemblies, and coding regions were subsequently aligned using transAlign ver 1.2 script (Bininda-Emonds, 2005). TransAlign determines the optimal codon-based alignment of protein-coding sequences based on multiple frame alignments of translated sequences with ClustalW (Larkin *et al.*, 2007). Nucleotide alignments as determined by transAlign were then fed into the `codeml` program from PAML package ver. 4.8 (Yang, 2007) and d*N* and d*S* rates calculated using mode -2 and the F3x4 codon table. `Codeml` was run twice: once with constrained ω value ($\omega = 1$) and once with calculation of ω enabled, and lnL was compared between the two runs using a log-likelihood ratio test.

Enrichment for different gene ontologies was tested using Plant GSEA (Gene Set Enrichment Analysis, (Yi *et al.*, 2013)) webserver (*A. thaliana* genome as the background reference), as before for differential expression using 0.05 as the FDR cut-off.

2.3 Results and discussion

2.3.1 Sequencing and read quality control

In total, 27 separate libraries were sequenced in the study: 3 shoot and 3 root samples from each of the serpentine populations of *Alyssum serpyllifolium* - Carratraca-S and Samil-S, as

well as a non-serpentine population Morata-NS and a non-accumulator outgroup species *Clypeola jonthlaspi*, which belongs to the same tribe Alysseae in the Brassicaceae family as *Alyssum* (Warwick *et al.*, 2008). In addition, 3 root samples from the non-serpentine *Alyssum serpyllifolium* Rubía-NS population were sampled (Table 2.3).

Four samples from the pilot experiment containing mixed RNA from a number of individuals were sequenced to obtain between 9.35 to 10.19 million read pairs (Table 2.3). For these samples, only 56.8% to 67.1% read pairs were retained after quality control, compared with typically more than 93% for samples in the main experiment. This higher level of the discarded reads in the pilot experiment can be attributed to rRNA contamination in the sequenced library (data not shown). Nevertheless, high percentage of discarded reads is a common malady of RNA-Seq studies, with typically 21% of reads removed during QC step in a survey of 21 studies conducted by Dewoody *et al.* (2013).

The sequencing depth was higher in the main experiment, with the exception of Rubía-NS samples, and the range spanned 15.28 to 22.84 million sequenced paired-end reads. Quality control trimming resulted in similar reduction of read length in all samples to between 95.9 and 97.2 bp (data not shown). Phred quality score of reads after trimming reached an average of 36.22 (4.75 SD, data not shown). In general, sequencing depth achieved in the study was in the region of 10 million paired-end reads and more, and thus should provide us with reliable expression quantification of transcripts across a range of abundance levels (Vijay *et al.*, 2013). Nevertheless, to create a comprehensive catalog of genes expressed across the full range of expression levels, a depth of 30-60 million reads would be required, as demonstrated in RNA-Seq studies in various animals (Schulz *et al.*, 2012; Francis *et al.*, 2013). Shoot and root were sequenced separately, as samples from different tissues and developmental stages provide more fine-scale picture of gene and isoform expression and recovery of more transcripts at a lower sequencing depth becomes more feasible. Given specialised physiological adaptations to hyperaccumulation and serpentine, in addition to the obvious ones due to distinct morphology and functions, the expectation was to see pronounced differences in gene expression profiles in the two tissues.

Table 2.3: Summary of RNA-Seq reads obtained in the study

Population	Tissue	Number of samples	Raw read pairs (in millions)	% retained reads	% retained bases
Pilot experiment					
Clypeola	root	1 (mixed)[1]	9.95	57.7	55.5
Clypeola	shoot	1 (mixed)[1]	9.35	64.1	61.9
Samil-S	root	1 (mixed)[2]	10.11	67.1	64.4
Samil-S	shoot	1 (mixed)[2]	10.19	56.8	54.7
Main experiment					
Clypeola	root	2	15.49 - 18.19	96.1 - 96.7	93.4 - 94.0
Clypeola	shoot	2	15.28 - 18.16	96.6 - 97.2	93.9 - 93.2
Samil-S	root	2	20.16 - 20.54	97.1 - 97.5	94.1 - 94.6
Samil-S	shoot	2	14.98 - 21.56	97.6 - 97.7	94.7 - 94.8
Carratraca-S	root	3	16.96 - 19.97	95.8 - 96.2	92.9 - 93.5
Carratraca-S	shoot	3	16.98 - 19.21	96.3 - 97.4	93.2 - 93.5
Morata-NS	root	3	20.48 - 22.84	94.7 - 95.8	93.0 - 94.0
Morata-NS	shoot	3	20.09 - 21.60	90.4 - 99.3	88.2 - 96.8
Rubia-NS	root	3	6.75 - 8.14	93.1 - 94.5	89.5 - 91.3

[1] contains pooled tissue from 10 individuals; [2] contains pooled tissue from 20 individuals

Table 2.4: Overview of RNA-Seq 'consensus' assembly and its structural annotation based on *A. thaliana* transcriptome.

	Alyssum serpyllifolium	*Clypeola jonthlaspi*	*Arabidopsis thaliana*
Number of transcripts:	18 566	17 222	27 416
Mean transcript length:	1 486	1 477	1 610
N50:	1 767	1 750	1 930
Total base content:	27 598 528	25 452 040	44 153 096
Mean GC content :			
5′ UTR	39.34%	39.14%	38.17%
CDS	44.77%	44.77%	44.11%
3′ UTR	34.31%	34.45%	33.32%
Number of contigs containing:			
5′ UTR + CDS + 3′ UTR	13 690 (6 987[1] / 3 282[2] / 2 224[3])	12 683 (6 304[1] / 3 098[2] / 2 047[3])	
5′ UTR + CDS	1 074 (941[2])	943 (850[2])	
CDS + 3′ UTR	2 589 (1 586[3])	2 533 (1 576[3])	
(partial) CDS	1 212	1 063	

Overlap of start and/or stop codon of the query protein with *Arabidopsis thaliana* homologue when aligned: [1] both start and stop stop codon, [2] only start codon, [3] only stop codon.

2.3.2 Assembly, mapping and transcriptome annotation

Vijay *et al.* (2013) and Nadeau *et al.* (2014) showed that when including a pre-mapping step to a reference transcriptome (up to 15% divergence), the resulting transcriptome and gene expression estimates are improved compared to when using a *de novo* assembly. In the current data set, mapping the reads to *Arabidopsis thaliana* transcriptome resulted in mapping to between 24 086 and 26 799 *Arabidopsis* loci and an average of 80% mapped reads across all the samples (data not shown), despite even higher sequence divergence from *Arabidopsis* ranging from 18% to 20%. This compares favourably with previous studies - Dewoody *et al.* (2013) found that studies using a reference genome had typically 74% reads mapped (n = 30) and those using a *de novo* assembly had 65% reads mapped (n = 20). Following the assembly of the selected reads for each mapped-to locus in individual populations and selection of the best available contig per each locus, 'hybrid' assemblies containing 18 565/17 222 sequences longer than 300 bp and mean contig length of 1 486 bp/1 477 bp for *Alyssum* and *Clypeola*, respectively, were created (Table 2.4).

APVO single copy genes shared by *A. thaliana, O. sativa, P. trichocarpa, V. vinifera* (Duarte *et al.*, 2010) and especially Core Eukaryotic Genes (CEG; Parra *et al.*, 2007) are highly conserved genes and thus can be used as a proxy for asssembly completeness (Bradnam *et al.*, 2013). *Alyssum* assembly contained 904 out of 959 APVO genes (94%), whereas *Clypeola* assembly contained 875 genes (91%). In the case of CEG, 394 genes out of 458 (86%) were detected in the *Alyssum* assembly and 390 (85%) in the *Clypeola* assembly. While roughly only two-thirds of *Arabidopsis* protein-coding genes were sampled in the assemblies, the genes that did get assembled were reconstituted to a nearly full size judging by mean contig length (Table 2.4).

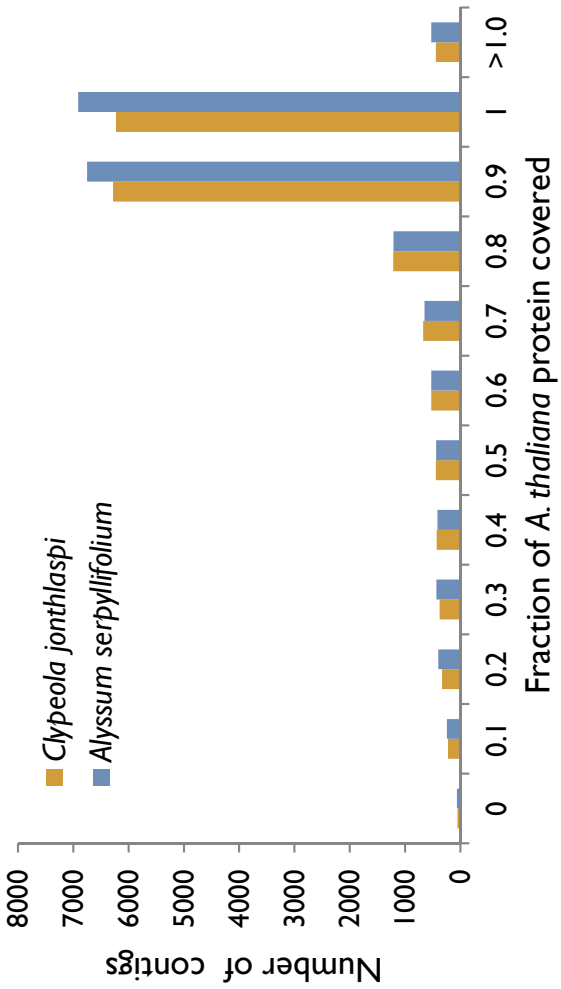


Figure 2.3: Coverage of *C. jonthlaspi* and *A. serpyllifolium* 'consensus' contigs relative to *A. thaliana* reference.

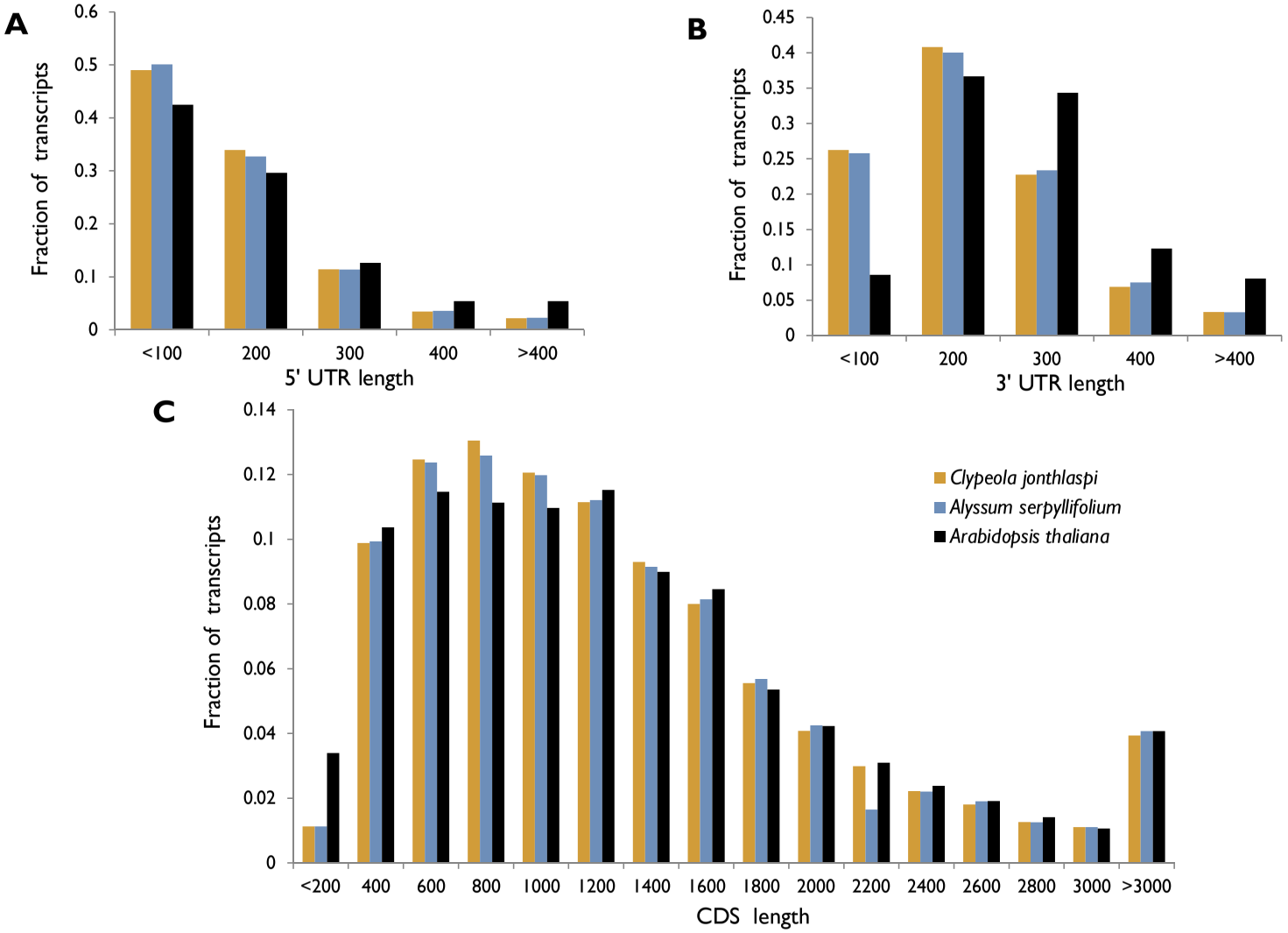


Figure 2.4: Comparison of distribution of **A)** 5′ UTR; **B)** 3′ UTR, **C)** CDS region length among *C. jonthlaspi*, *A. serpyllifolium* 'consensus' assemblies and the reference *A. thaliana* transcriptome.

This conclusion is further supported by coverage of the *A. thaliana* transcripts by the assembly contigs which centred at 90% both in *Alyssum* and *Clypeola* (Figure 2.3). Coverage for the selection of conserved orthologs was even higher - mean 96% for genes in the APVO and CEG datasets (data not shown).

As well as this, the majority of contigs were found to contain the coding sequence (CDS) together with 5′ and 3′ untranslated regions (UTR): 13 690 in *Alyssum* and 12 683 in *Clypeola* (Table 2.4). In the majority of cases, at least the start or stop of the protein was deemed high confidence by the virtue of aligning with *A. thaliana* homolog's start or stop position. Furthermore, 1 074/2 589 *Alyssum* contigs and 943/2 533 *Clypeola* contigs contained either of the 5′ and 3′ regions in addition to CDS, respectively. Veracity of the annotation procedure is supported by differences in GC nucleotide content across the three regions in the assemblies closely tracking that of *A. thaliana*'s (Table 2.4) When looking at the size distribution of 5′/3′ UTR and CDS regions in the three species, 5′ UTR and CDS appear to have assembled better than 3′ UTR (Figure 2.4). Relative reduction in the number of longer 3′ UTRs could be explained by certain degradation of input mRNA used in preparation of the sequenced cDNA libraries. Lastly, functional annotation was copied over from

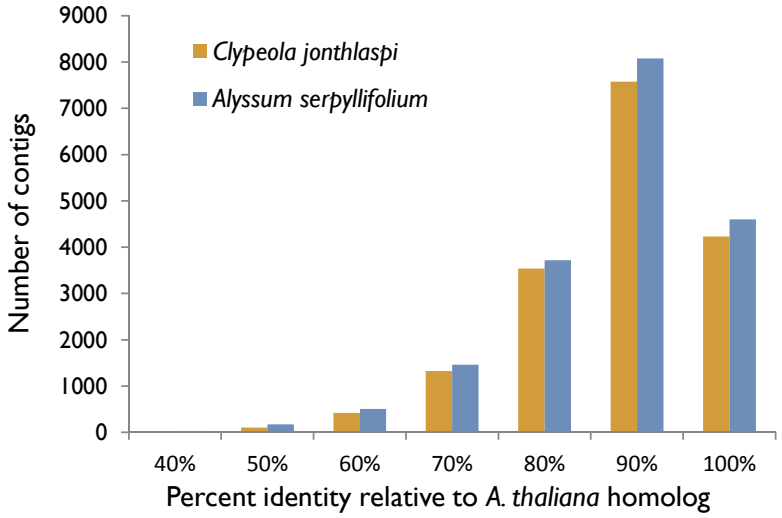


Figure 2.5: Distribution of amino acid identity of *C. jonthlaspi* and *A. serpyllifolium* 'consensus' contigs relative to *A. thaliana* reference proteins.

Arabidopsis thaliana reference loci. Mean amino acid percent identity for all the contigs for both species is 83% (Figure 2.5), while nucleotide percent identity in the coding region

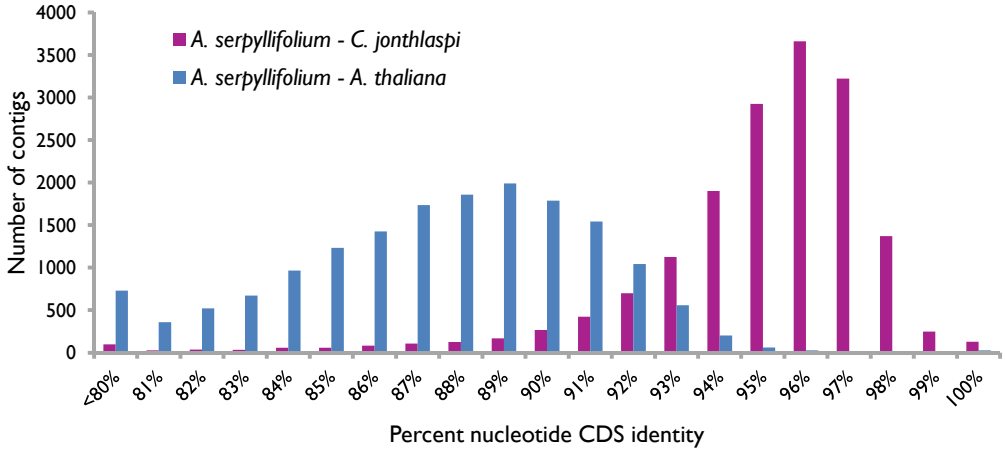


Figure 2.6: Distribution of nucleotide identity of *A. serpyllifolium* 'consensus' contigs relative to *A. thaliana* and *C. jonthlaspi* in the coding region.

was at 87% (Figure 2.6). It has been shown that protein sequence similarity in the range of 40 - 60% allows for reliable functional annotation of the homologs (Addou *et al.*, 2009) so the functional annotation of the dataset should be broadly correct.

The 'hybrid' assembly could then be used as a reference for mapping of reads from each population and creating a population-specific 'consensus' transcriptome based on detected variants. This strategy was possible given moderate levels of polymorphism across populations: nucleotide diversity (π) in a select subset of 26 stably-expressed, single-copy APVO loci centred at 0.0053 with 0.0033 SD. BWA mem used for mapping the reads from each population to the 'hybrid' assembly and thus key to creating 'consensus' transcriptome has been shown to handle much higher levels of polymorphism (at least up to 1.5%) with no decline in performance (Li, 2013). The resulting four 'consensus' transcriptomes featured sequence changes in between 7 875 (Rubía-NS), 9 778 (Samil-S), 10 207 (Carratraca-S) and 10 244 (Samil-S) contigs but using them as a reference resulted only in a small increase in the fraction of mapped reads (0.8-1.6 percentage point, data not shown). This was expected given moderate nucleotide diversity levels across the populations.

Reads unmapped to the 'consensus' transcriptome or filtered due to multi-mapping or

chimeric alignments were then rescued by independent 'supplementary' assembly which they were then mapped to. The majority of new contigs showed similarity to *Arabidopsis thaliana* genes already present in the 'consensus' assembly; only 680 out of 4 570 contigs with top blast hit to *Arabidopsis* mapped to new *Arabidopsis* loci. The former can represent fragments from the original transcripts that were not assembled as a contiguous region in the 'consensus' assembly and paralogs. Out of the remaining 1 048 contigs, 987 had at least a single InterPro domain match and further 199 showed significant similarity to a sequence from nr database. The highest number of hits belonged to two other Brassicaceae species: *Eutrema salsugineum* (30 hits) and *Capsella rubella* (13 hits) as expected given phylogenetic proximity, however, dozens of hits to invertebrate animals and even plant viruses were also found (data not shown).

Re-using the initially unmapped reads in that way resulted in the increase of mapped reads by on average 5.9 percentage points (pp; SD = 0.4 pp) in all the root samples and 4.9 pp in the shoot samples (SD = 0.2 pp; data not shown). Ultimately, 40× average read coverage was achieved in the study but variation in the expression intensity was wide (Figure 2.7) and the distribution was skewed towards genes expressed at a low level. A considerable number of genes was expressed at a high level of RPKM > 100: on average 1 901 in *Alyssum* and 2 067 in *Clypeola*. However, the biggest fraction of genes showed barely detectable expression (RPKM < 5): 6 735 on average in *Alyssum* samples and 7 214 in *Clypeola*. Higher number of *Clypeola* contigs with no mapped reads - mean 3 098 compared to 1 661 in *Alyssum* can be attributed to 1 344 fewer *Arabidopsis* loci assembled in the *Clypeola* transcriptome. The majority of *Clypeola* genes found not to be expressed are expected to be of little biological interest since they appear to be expressed at a low level in both species, as evidenced by significantly higher number of *Alyssum* genes showing expression below 5 RPKM (t-test, $p < 0.01$) - on average 958 more than in *Clypeola*.
One reason for fewer transcripts having been reconstituted in the *Clypeola* transcriptome is the underlying lower mapping rate to the *Arabidopsis* transcriptome in the first step of the assembly process (Figure 4.2). While the shoot *Clypeola* samples mapped well at 82% (data not shown), root samples showed only 70% mapping rate, for unclear reasons. This

effect carried on downstream to final mapping of reads to the 'consensus' transcriptome and additional contig set based on unmapped reads. Whereas an average of 82% and 85% of reads were mapped in the *Alyssum* samples in the root and shoot, respectively, the percentage in *Clypeola* was only 71% and 79%, respectively. This artifact did not lead to an elevated number of DE genes in the root, though - if anything, more genes were found to be differentially expressed in the shoot in all comparisons made (see below).

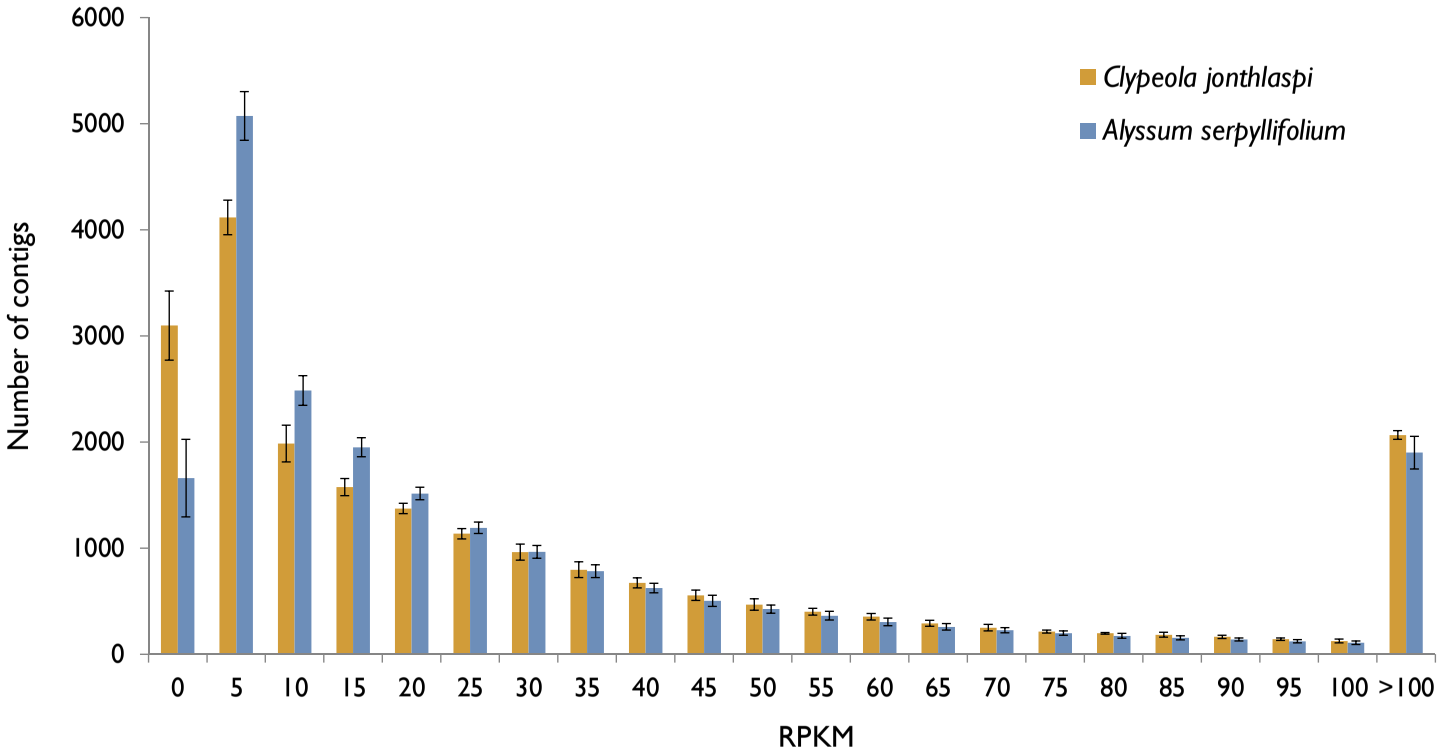


Figure 2.7: Distribution of RPKM values, representing gene expression strength, across all 'consensus' contigs of *Alyssum serpyllifolium* and *Clypeola jonthlaspi*. Error bars represent SD based on the number of contigs in each bin among all the sequenced samples.

2.3.3 Differential expression analysis

2.3.3.1 MDS and PCA analysis

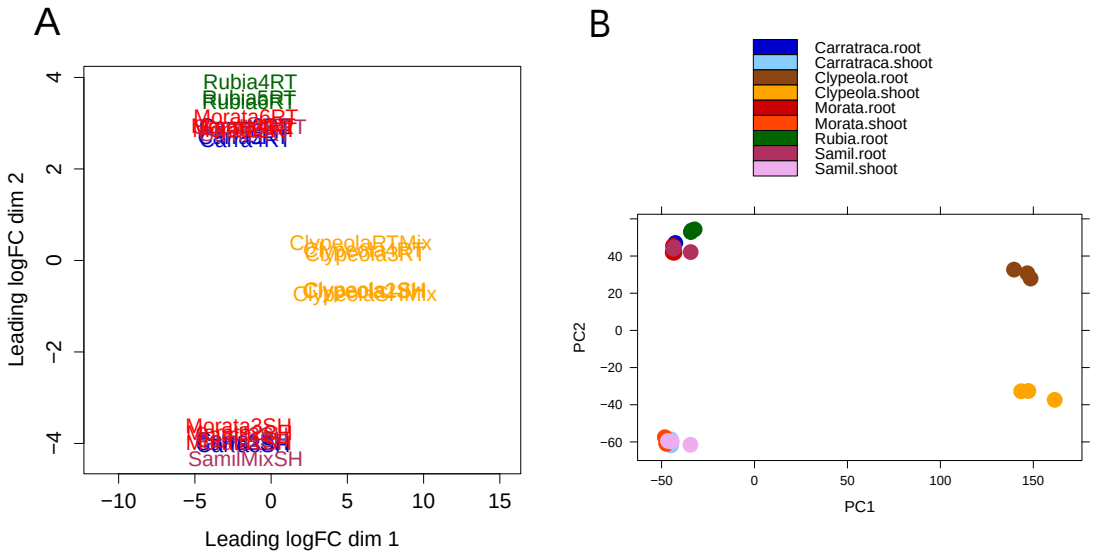


Figure 2.8: Results of **A)** MDS analysis in edgeR and **B)** PCA in DESeq2 based on expression of 500 genes with the highest variance in the dataset. *SH* denotes shoot samples and *RT* denotes root samples.

To investigate overall similarity in patterns of gene expression across the samples, distances between the samples for the top 500 genes with the highest variance were visualised using two dimensionality reduction methods: multidimensional scaling (MDS, Figure 2.8 A) and principle component analysis (PCA, Figure 2.8 B). The results were similar using both approaches in clearly separating the samples along two components - tissue and species, with bigger difference between tissues in *Alyssum* compared to *Clypeola* visible in the MDS plot. This difference disappeared when looking at the plot based on all the genes in the dataset (data not shown), suggesting the existence of greater tissue-specificity in the highly variable genes between *Alyssum* and *Clypeola*.

Crucially, all the *Alyssum* samples in a given tissue clustered together, regardless of the edaphic status of the source population - vast interspecific expression differences between *Alyssum* and *Clypeola* drown out intraspecific differences within *Alyssum*.

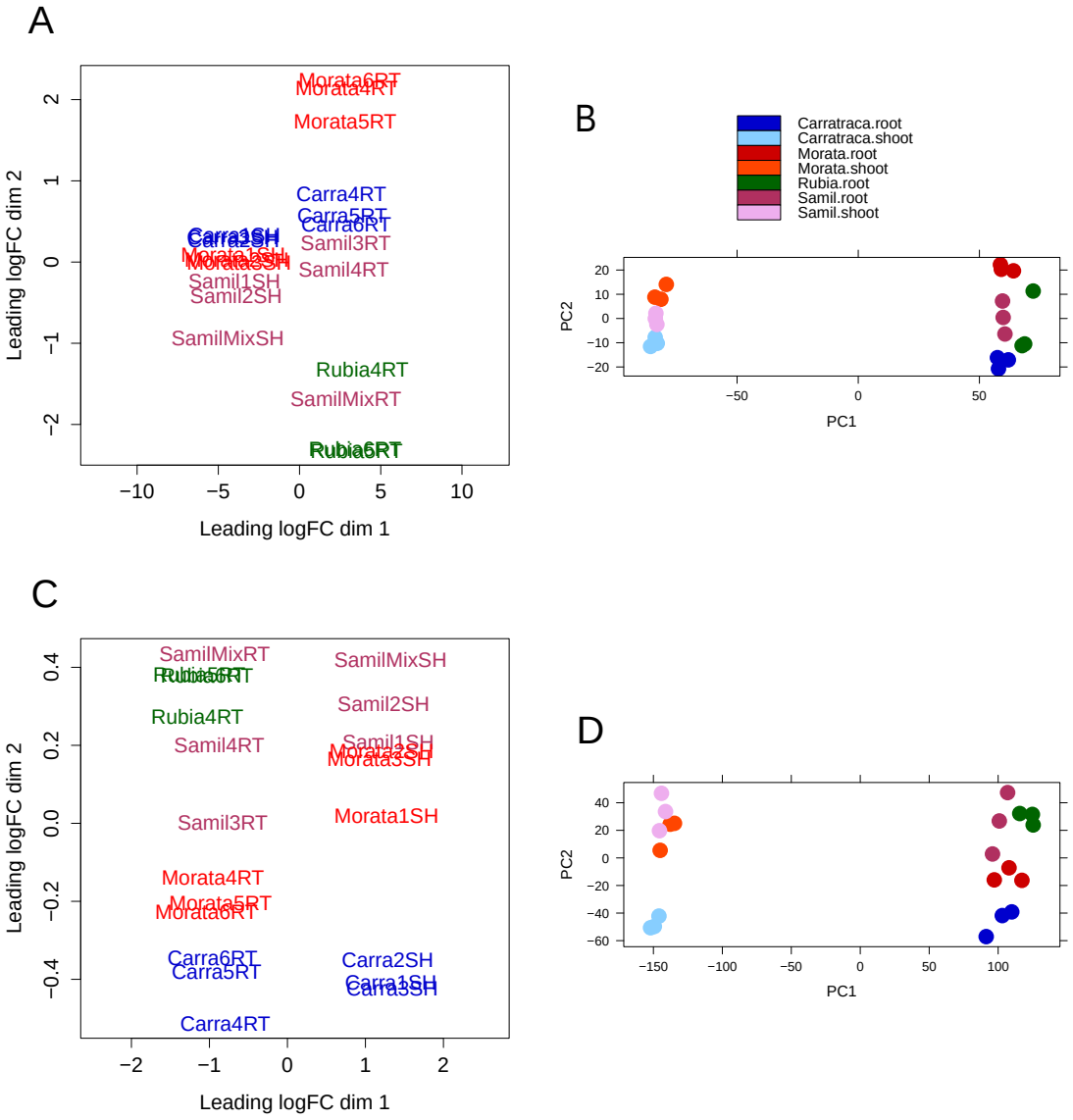


Figure 2.9: Results of **A), C)** MDS analysis in edgeR and **B), D)** PCA in DESeq2 based on expression of **A), B)** 500 genes with the highest variance and **C), D)** all the genes in the *Alyssum* dataset. *SH* denotes shoot samples and *RT* denotes root samples.

In order to focus on the relationships between *Alyssum* samples more closely, *Clypeola* samples were removed and PCA/MDS analysis repeated. Looking at the results based on the top 500 genes, there is greater scatter among root samples compared to the shoot samples (Figure 2.9 A and B) indicative of greater biological variability in the root gene expression in all of the populations. Root samples from serpentine populations Carratraca-S and Samil-S tended to group together away from the non-serpentine populations Rubía-NS and Morata-NS. However, when including all the genes in the analysis, this effect disappeared

and the distances between population samples in each tissue correlated with geographical distances between the source populations (Figure 2.9 C and D).

2.3.3.2 DE genes shared between the serpentine populations

The analysis then proceeded to identify specific genes contributing to the hyperaccumulator serpentine phenotype, and associated serpentine adaptation, also known as *serpentine syndrome* (Brady *et al.*, 2005) in *Alyssum*. First, every possible one-on-one contrast between taxa was made for root and shoot tissues separately. Genes significantly differentially expressed (FDR < 0.01) in all of the serpentine versus non-serpentine comparisons were then grouped into four classes: upregulated in the root, downregulated in the root, upregulated in the shoot, downregulated in the shoot in the two *Alyssum* serpentine populations relative to the non-serpentine *Alyssum* and *Clypeola* (Figure 2.10). In total, only 52 genes met that criteria which, along with the results of the PCA and MDS analysis suggests that few gene expression differences are shared by the serpentine *Alyssum* populations relative to the non-serpentine ones.

Root Ten genes were found to be upregulated in the serpentine roots (Figure 2.10 A) and they were enriched for the *membrane* category. Out of the 6 proteins annotated with the *membrane* term in the group, 3 show transporter activity: *IRON REGULATED 1* (*IREG1*) and *IRON REGULATED 2* (*IREG2*), and an aminoacid/auxin permease (AAAP) family member. The other three genes in the *membrane* category belong to the yet uncharacterised family whose 10 members contain cystathionine-β-synthase (CBS) domains. Importantly, among the ten upregulated genes, *IREG1*, the AAAP family member and a galactose oxidase/kelch repeat superfamily member had been shown to belong to the gene set showing strong upregulation in response to iron deprivation (Dinneny *et al.*, 2008; Buckhout *et al.*, 2009; Long *et al.*, 2010).

IREG1 belongs to the small family of 3 *Arabidopsis* homologs of mammal ferroportin IREGs, which functions as iron exporters. *Arabidopsis IREG1* is localised to the plasma membrane, mainly expressed in the stele (Morrissey *et al.*, 2009) and is able to export

nickel and cobalt. It has been shown to be important for nickel and cobalt detoxification (Kirchner, 2009), consistent with increased nickel tolerance in *IREG1* overexpressor lines and increased nickel sensitivity in *ireg1* knockout lines.

IREG2 in *Arabidopsis* is targeted to the tonoplast in the two outermost root layers and functions in nickel and cobalt sequestration. It is expressed under iron deficiency conditions and its proposed role in *Arabidopsis* is to protect the plant against toxic effects of increased unspecific influx of nickel and cobalt due to *IRON-REGULATED TRANSPORTER 1* (*IRT1*, transporter responsible for iron uptake from the soil, described later) overexpression under low iron (Schaaf *et al.*, 2006; Morrissey *et al.*, 2009). There is evidence for higher expression of *IREG2* in the roots of *A. halleri* relative to *A. thaliana* (Talke *et al.*, 2006) as well as higher expression in nickel hyperaccumulating accession of *N. caerulescens* (Halimaa *et al.*, 2014). Additionally, a newly characterised *IREG* family member in non-Brassicaceae nickel hyperaccumulator *Psychotria gabriellae* appears to be functionally redundant with *At*IREG2 (Merlot *et al.*, 2014) – it is targeted to the tonoplast and confers increased nickel resistance in WT and *ireg2 Arabidopsis* plants. Therefore, IREG family members are likely to have convergently evolved functional importance in separate lineages of nickel hyperaccumulators. Lastly, the AAAP family member (AT5G38820) is a putative amino acid transporter regulated by iron deficiency (Buckhout *et al.*, 2009; García *et al.*, 2010).

Surprisingly, only one gene, a cytochrome 450 gene was found to be downregulated in all the serpentine relative to all non-serpentine in the root (Figure 2.10 B).

Shoot Comparison of the shoot transcriptomes yielded more shared DE genes between the serpentine and non-serpentine than in the root: 28 upregulated and 13 downregulated. This could be partially down to the fact that samples for only one (Morata-NS) of the two non-serpentine populations were available, and Rubía-NS expression data was missing. Two genes showed upregulation in the shoot as well as in the root (Figure 2.10 C): *IREG2*, the AAAP family member, which together with *ATP-PRT1* and *IRT1*, contributed to the over-representation of genes showing strong upregulation in response to iron deprivation, seen also in the root. Five out of 28 upregulated genes were associated with *metal ion transport*

gene ontology term: *IREG2*, *IRT1*, *DICARBOXYLATE TRANSPORT 2.2* (*DIT2.2*), *L-TYPE LECTIN RECEPTOR KINASE S.5* (*LECRK-S.5*) and a lectin pectin lyase-like superfamily member. Furthermore, 3 genes with carboxylic acid transmembrane transporter activity were detected: *CATIONIC AMINO ACID TRANSPORTER 4* (*CAT4*), *DIT2.2*, and the AAAP family member. Two more genes belonging to the galactose oxidase/kelch family with 152 members were found among upregulated genes in the shoot in addition to the one gene in the root. CAT4, a member of the cationic amino acid transporter family (Su *et al.*, 2004) was recently shown to transport histidine into the vacuole and increase nickel hyperaccumulation and tolerance when expressed in yeast (C.J. Snowden, S.T. Mugford & J.A.C Smith, unpublished).

Among 13 downregulated genes in all serpentine (Figure 2.10 D), two overrepresented gene ontology terms were found with two associated genes, each: *response to karrikin* - *SMALL AND BASIC INTRINSIC PROTEIN 2* (*SIP2*) and *CELL WALL INVERTASE 1* (*CWINV1*) and *respiratory burst involved in defense response* - *CWINV1* and an α-1,4-glycosyltransferase family gene.

2.3.3.3 Population-specific serpentine DE genes

It is also quite possible that population-specific (ie. present in only one of the serpentine populations relative to all the non-serpentine taxa) differentially expressed genes also contribute to the hyperaccumulation trait as well as adaptation to the serpentine in a given population, and so they were examined more closely.

Root upregulated genes Comparison of Carratraca-S- and Samil-S- specific upregulated genes in the root revealed similarly small number of genes - 15 and 10, respectively. In Carratraca-S, 2 genes were annotated with *amino acid transport* term (Figure 2.11 A): *CAT4* (upregulated also in the shoot of the two serpentine populations) and a mitochondrial substrate carrier family protein. In Samil-S, *HEAVY METAL ATPASE 5* (*HMA5*) was found amongst upregulated root genes (Figure 2.12 A). *HMA5* has been shown to play a key role in Cu detoxification in the roots and interact with copper chaperones *COPPER CHAPERONE*

(*CCH*) and *HOMOLOG OF ANTI-OXIDANT 1* (*ATX1*) in *Arabidopsis* (Andrés-Colás *et al.*, 2006). In rice, *HMA5* homolog is a plasma membrane transporter responsible for xylem loading of copper in both vegetative and reproductive structures (Deng *et al.*, 2013).

Root downregulated genes Twelve and eight population-specific downregulated genes in the root in Carratraca-S and Samil-S did not appear to share common functions, either. In Carratraca-S, these genes were enriched for *glutathione degradation pathway* (Figure 2.11 B) - including *GAMMA-GLUTAMYL TRANSPEPTIDASE 3* (*GGT3*), and two were oxoglutarate/iron-dependent oxygenase.

Shoot upregulated genes Just as in 'all serpentine versus all non-serpentine' comparison, a higher number of population-specific differentially expressed genes was found in the shoot. Seventy-seven and 68 genes were upregulated in Carratraca-S and Samil-S, respectively. In the case of Carratraca-S, two of the CBS DUF21 domain genes upregulated in the root in the two serpentine populations were also found to be so in the shoot (Figure 2.11 C). Four genes involved in ubiquitin ligase complex formation also stood out. Amongst Samil-S-specific genes, two glutathione transferases were found: *GLUTATHIONE S-TRANSFERASE TAU 3* (*GSTU3*) and *GLUTATHIONE S-TRANSFERASE F4* (*GSTF4*) (Figure 2.12 C). Both in Carratraca-S and in Samil-S, *membrane transporters* were an overrepresented category: 8 in Carratraca-S and 5 in Samil-S. Different ZRT, IRT-like Protein (ZIP) family members were upregulated in each population: *ZIP12* in Carratraca-S, and *ZIP9* and *ZIP11* in Samil-S.

ZIP9 is a marker for zinc deficiency response in the root in *Arabidopsis* (Gustin *et al.*, 2009; Yang *et al.*, 2010**a**), *N. caerulescens* (van de Mortel *et al.*, 2006) and *A. halleri* (Talke *et al.*, 2006; Chiang *et al.*, 2006), but is also constitutively upregulated in *A. halleri* roots relative to *A. thaliana* (Grotz *et al.*, 1998; Weber *et al.*, 2004; Talke *et al.*, 2006). In addition, *AtZIP9* was able to complement Mn uptake-deficient mutant in yeast (Milner *et al.*, 2013). *ZIP11* shows constitutively higher expression in *A. halleri* in all tissues relative to *A. thaliana*, and shows especially big difference in leaves (Chiang *et al.*, 2006). There is conflicting evidence regarding *ZIP11* upregulation under zinc deficiency conditions in *Arabidopsis*

(van de Mortel *et al.*, 2006; Jain *et al.*, 2013) but it is likely not to be responsive to zinc deficiency because of the lack of zinc deficiency responsive element in its promoter which is present in the promoters of other zinc deficiency-regulated ZIP family members (Jain *et al.*, 2013). Lastly, both *ZIP11* and *ZIP12* have the ability to rescue a zinc-uptake mutant in yeast (Milner *et al.*, 2013).

Additionally, population-specific upregulated transport genes included *METALLOTHIONEIN 1A* (*MT1A*), *GLUTAMATE RECEPTOR 2.4* (*GLR2.4*, uncharacterised glutamate transporter), *PLEIOTROPIC DRUG RESISTANCE 5* (*PDR5*), *VHP2;2* and two nucleobase permeases in Carratraca-S, while *NON-INTRINSIC ABC PROTEIN 3* (*NAP3*) and a proton-dependent oligopeptide transporter (POT) family representative were found in Samil-S. In *Arabidopsis*, (*MT1A*) is strongly expressed in the phloem and senescent leaves, and is induced by copper (Guo *et al.*, 2008). It has been shown to take part in copper accumulation in roots, and together with phytochelatins mediates copper tolerance (Guo *et al.*, 2008) as well as copper mobilisation during seed maturation (Benatti *et al.*, 2014). *NAP3* is an ortholog of rice SENSITIVE TO ALUMINIUM RHIZOTOXICITY *STAR1* protein necessary for aluminium tolerance (Huang *et al.*, 2009) and despite different expression pattern, the *Arabidopsis* ortholog also participates in aluminium tolerance and shows redundancy with *OsSTAR1* function (Huang *et al.*, 2010).

Shoot downregulated genes Over twice as many genes appeared to be downregulated in the shoot specifically in Carratraca-S: 142, than in Samil-S: 68. Interestingly, despite different individual genes in each enriched category, some processes seemed to be downregulated in both populations, namely: *response to stress*, *defence response*, *tryptophan/indokylamine catabolism* and *oxidation-reduction*. Forty-three and 20 of the downregulated genes in Carratraca-S and Samil-S, respectively, were annotated with the category *response to stress* and since most of the genes in it were also found to be over-represented under more specific categories, such as *defense response* described below, they will not be listed here.

Twenty-three downregulated defense genes in Carratraca-S included (Figure 2.11 D): *METHYL ESTERASE 9* (*MES9*), two cytochrome P450 genes and two peroxidases, *UDP-DEPENDENT GLYCOSYLTRANSFERASE 76B1* (*UGT76B1*), *KUNITZ TRYPSIN INHIBITOR*

1 (*KTI1*), *BETA GLUCOSIDASE 26* (*BGLU26*) and *DOWNY MILDEW RESISTANT 6* (*DMR6*). Samil-S-specific downregulated defense response genes were, amongst others (Figure 2.12 D), *METHYL ESTERASE 1* (*MES1*), *RESISTANCE TO PSEUDOMONAS SYRINGAE 3* (*RPS3*), *GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE C2* (*GAPC2*), *MYB DOMAIN PROTEIN 108* (*MYB108*) and *PROLINE DEHYDROGENASE 1* (*PDH1*).
KT1 is a serine protease, which is induced late following exposure to plant pathogens and salicylic acid. It negatively regulates programmed cell death when exposed to bacterial and fungal pathogens (Li *et al.*, 2008**a**).
A single different MES (Methyl Esterase) family member in each of the serpentine population transcriptome's was downregulated. Methyl esterases hydrolyse salicylic acid derivative methyl salicylate into salicylic acid (Manosalva *et al.*, 2010). In plants, methyl salicylate is a hypothesised long-range systemic acquired resistance signal transported in the phloem away from the site of infection and converted back into biologically active salicylic acid (Liu *et al.*, 2011; Shah & Zeier, 2013).
RPS3 is one of the R genes encoding cytoplasmic receptor-like protein recognising specific avirulence genes produced by *Pseudomonas* bacteria, inducing hypersensitive response at the sites of infection and thus conferring resistance to strains expressing these effector genes (Grant *et al.*, 1995; Boyes *et al.*, 1998; Grant *et al.*, 2000; Rose *et al.*, 2012). Interestingly, *RPS3* also showed significant sub-population-specific expression variation amongst *N. caerulescens* from different sites showing marked variation in morphology and Ni hyperaccumulation ability (Visioli *et al.*, 2012).

Five out of 79 *Arabidopsis* genes involved in *tryptophan/indokylamine catabolism* were found in Carratraca-S: *SULFURTRANSFERASE PROTEIN 16* (*STR16*), *CYTOCHROME P450, FAMILY 79, SUBFAMILY B, POLYPEPTIDE 2* (*CYP79B2*), *GLUTAMATE RECEPTOR 1.1* (*GLR1.1*), *PEARLI4* and *BGLU26*. In Samil-S, three genes were found: avirulence induced gene 2-like (AIG2-like) family member, *IQ-MOTIF PROTEIN 1* (*IQM1*) and *MAP KINASE KINASE 9* (*MKK9*). *CYP79B2* is one of the two genes catalysing the conversion of tryptophan into indole-3-acetaldoxime, from which point either plant hormone indole acetic acid or secondary metabolites with defence properties: indole glucosinolates or

camalexins (the main phytoalexin) can be synthesised (Hull *et al.*, 2000; Glawischnig *et al.*, 2004).

BGLU26 (also known as *PENETRATION 2*) participates in the pre-invasion phase of nonhost resistance as well as resistance to adapted fungal pathogens (Hiruma *et al.*, 2013). *BGLU26* encodes a peroxisome-bound myrosinase which activates a certain type of indole glucosinolates with antifungal properties as well as increases callose deposition to restrict pathogen growth (Lipka *et al.*, 2005; Bednarek *et al.*, 2009; Clay *et al.*, 2009).

The last shared GO category between Carratraca-S and Samil-S was *oxidoreductase activity*. Ten genes in that category in Carratraca-S featured six cytochrome 450 genes, three 2-oxoglutarate-dependent oxygenases and *DOWNY MILDEW RESISTANT 6* (*DMR6*). Amongst eight genes in Samil-S, only two belonged to cytochrome 450 family and the rest featured *FERRITIN2* (*FER2*), *COPPER CHAPERONE FOR SOD1* (*CCS*), *GAPC2*, *PDH1*. *FER2* is one of the plant genes with homology to ferritin proteins, which function as iron stores in animals. Evidence from plants indicates that rather than providing iron storage for development, ferritins protect plants against iron-induced ROS (Ravet *et al.*, 2009**a**,**b**). In *N. caerulescens*, *FER2*'s upregulation is associated with high zinc accumulating ecotypes (Halimaa *et al.*, 2014) and the gene is constitutively upregulated in *A. halleri* relative to *A. thaliana* (Talke *et al.*, 2006).

CCS is a copper chaperone which delivers the copper co-factor to superoxide dismutase in the chloroplast stroma, cytoplasm and peroxisome. In turn, copper superoxide dismutases take part in detoxifying superoxide ions by catalysing dismutation to hydrogen peroxide (Abdel-Ghany *et al.*, 2005; Chu *et al.*, 2005). In *Arabidopsis*, copper superoxide dismutases and *CCS* are downregulated by miR398 under thermal stress, which appears critical especially for the protection of reproductive structures (Guan *et al.*, 2013).

In Samil-S, only one more gene ontology term was found to be overrepresented among downregulated genes - response to ethylene stimulus (4 genes). Seven of the downregulated genes in Samil-S had also appeared as downregulated in a *WRKY DNA-BINDING PROTEIN 15* (*WRKY15*, ROS-inducible transcription factor modulating salt and oxidative

stress responses) overexpressor line (Vanderauwera *et al.*, 2012).

Four specific enriched categories among genes downregulated in the shoot in Carratraca-S concerned broadly amino acid metabolism: *transport of nitrate* (9 genes, including *PHOSPHATE 1* - transporter mediating inorganic phosphate xylem loading for transport to the shoot, Hamburger *et al.*, 2002), *cellular modified amino acid metabolic process* (15 genes, including *NICOTIANAMINE SYNTHASE 1 - NAS1*), *amino acid transport* (8 genes), *phenylpropanoid biosynthetic process* (8 genes). In stark contast to Samil-S, where one τ- and one ϕ- class glutathione S-transferase were found to be upregulated in the shoot, two τ (*GLUTATHIONE S-TRANSFERASE TAU 11 and 12 - GSTU11, GSTU12*) and one ϕ (*GLUTATHIONE S-TRANSFERASE F3 - GSTF3*) glutathione S-transferases were downregulated in Carratraca-S. Three more downregulated transferases in Carratraca-S were *ISOPENTYLTRANSFERASES* (*IPT3* and *IPT5*), which function as cytokinin synthetases, and *NAS1*. The three downregulated GSTs were also overrepresented under the *toxin metabolism* category along with five more genes: *DIHYDROSPHINGOSINE PHOSPHATE LYASE 1* (*DPL1*), *NAC DOMAIN CONTAINING PROTEIN 53* (*NAC053*), an O-acyltransferase and *WRKY DNA-BINDING PROTEIN 28* (*WRKY28*).

NAS1 is an enzyme synthesising a non-proteogenic amino acid chelator nicotianamine from three molecules of S-adenosyl methionine. Nicotianamine (NA) can form stable complexes with a variety of metal cations: Fe^{2+}, Fe^{3+}, Cu^{2+}, Ni^{2+}, Co^{2+}, Zn^{2+}, and Mn^{2+} (ordered by decreasing affinity). Due to its ability to bind those metals with high affinity and low molecular weight, nicotianamine is one of the critical components of metal homoeostasis through roles in metal remobilisation, long range transport and chelation for detoxification (Verbruggen *et al.*, 2009; Maestri *et al.*, 2010). In particular, nicotianamine is involved in remobilising iron from vasculature into iron sinks in young leaves and floral tissues as well as seeds, and transport of zinc into the vacuole for zinc tolerance in *Arabidopsis* (Klatte *et al.*, 2009; Haydon *et al.*, 2012; Schuler *et al.*, 2012). The importance of nicotianamine in the two model hyperaccumulator species *N. caerulesces* and *A. halleri* is well established (Verbruggen *et al.*, 2009; Maestri *et al.*, 2010; Hanikenne & Nouet, 2011). It is produced and accumulates in *Thlaspi caerulescens* (Vacchina *et al.*, 2003) and *Arabidopsis halleri*

(Weber *et al.*, 2004) to much higher concentrations than in congeneric non-accumulators. Thus, finding *NAS1* amongst downregulated genes in Carratraca-S is unexpected, given the upregulation of *NAS1* in *Arabidopsis* on exposure to excess zinc (van de Mortel *et al.*, 2006), in the shoot transcriptome of *N. caerulescens* relative to the non-accumulator *T. arvense* (Hammond *et al.*, 2006), and constitutive high expression of different NAS isoforms in *A. halleri* relative to *A. thaliana* in the root and the shoot (Becher *et al.*, 2004; Weber *et al.*, 2004; Talke *et al.*, 2006). In particular, research suggests that *NAS2* transcript levels in *N. caerulescens* correlate with zinc and cadmium tolerance (Halimaa *et al.*, 2014) and Zn/Cd hyperaccumulation in *A. halleri* - when *NAS2* expression in *A. halleri* is knocked down by RNAi, zinc and cadmium foliar concentrations are vastly diminished, since nicotianamine is proposed to complex Zn/Cd and facilitate passage towards the xylem in the root (Deinlein *et al.*, 2012). Lately, Cornu *et al.* (2015) further confirmed the importance of nicotinamine (and consequently, *NAS*) in establishing the zinc hyperaccumulator phenotype through facilitating xylem loading step in *A. halleri*. However, they also surprisingly showed that nickel translocation to the shoot was actually increased by the lack of *Ah*NAS2, as opposed to zinc.

Conversely, overexpression of *Nc*NAS1 in *Arabidopsis* increased nickel tolerance and accumulation (Pianelli *et al.*, 2005), as expected. In *N. caerulescens*, *Nc*NAS1 expression and so nicotianamine production was found to be induced after Ni exposure and nicotianamine likely mediates nickel root to shoot translocation in the xylem, where stable Ni^{2+}-NA complexes are formed only in the hyperaccumulator *N. caerulescenes* and not in the non-accumulator *T. arvense* (Mari *et al.*, 2006); positive correlation exists between foliar nicotianamine and shoot nickel content in different *Noccaea* species (Callahan *et al.*, 2007). Interestingly, overexpression of *NAS* from non-accumulator plants can also lead to induction of nicotianamine synthesis after metal treatment, higher transition metal tolerance, and even elevated shoot metal accumulation (Douchkov *et al.*, 2005; Kim *et al.*, 2005).

Glutathione-S-transferases (GST) form a family of enzymes with six subclasses in *Arabidopsis* that conjugate reduced glutathione via a a sulfhydryl group to xenobiotics and

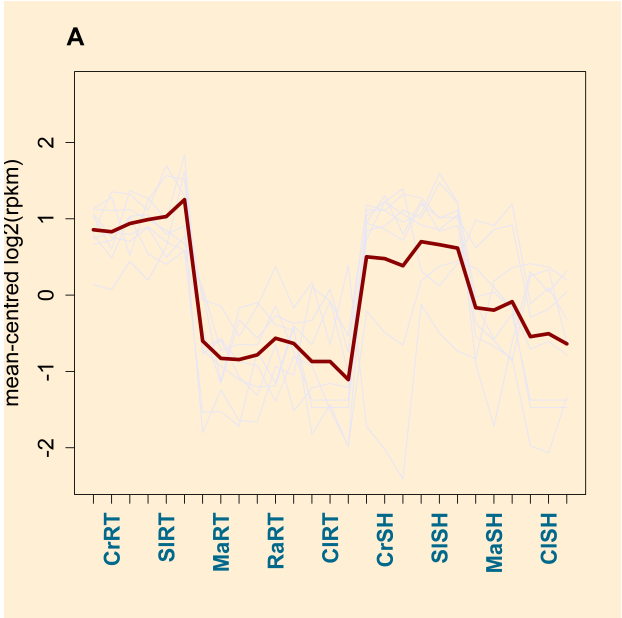
peroxidised lipids for detoxification as well act as peroxidases or reductases to ROS (Dixon *et al.*, 2009). The production of this antioxidant is upregulated in response to salicylic acid, hydrogen peroxide and metal stress in plants (Dixon *et al.*, 2002; van de Mortel *et al.*, 2006; Tuomainen *et al.*, 2010; Jozefczak *et al.*, 2012). As to the last point, expression of a number of GSTs has been shown to increase on exposure to zinc in the leaf epidermis in *N. caerulescens* (Schneider *et al.*, 2012). Also, overexpression of a tobacco GST gene in *Arabidopsis* led to increased aluminium tolerance supporting hypothesised role of GSTs in metal induced stress (Ezaki *et al.*, 2000). Consistent with the importance of glutathione in metal hyperaccumulators, previous proteomics research in the Ni hyperaccumulator *Alyssum lesbiacum* identified glutathione S-transferase to be upregulated on short exposure to nickel (Ingle *et al.*, 2005**b**).

The case for GST's importance for adaptation to nickel hyperaccumulation is particularly strong in *N. caerulescens*. In the shoots of nickel-hyperaccumulating *Noccaea* species, salicylic acid levels are constitutively elevated and signal activation of genes in the glutathione synthesis and turnover pathway; this results in increased production of reduced glutathione, which appears to contribute to nickel tolerance (Freeman *et al.*, 2004, 2005). This was shown in *Arabidopsis thaliana* plants expressing *Noccaea caerulescens* serine acetyltransferase that overproduce glutathione and become highly resistant towards nickel (Freeman & Salt, 2007). Interestingly, GSTF3 protein levels have been shown to be increased in the root of low zinc accumulating *N. caerulescens* ecotype relative to a high zinc accumulating ecotype (Tuomainen *et al.*, 2010), yet the gene shows higher expression in *N. caerulescens* shoot than in the non-accumulator *T. arvense* (Hammond *et al.*, 2006).

DPL1 catalyses the degradation of sphingoid long chain bases phosphates (LCB-Ps) to C_{16} fatty aldehydes and phosphoethanolamine in plants (Tsegaye *et al.*, 2007; Nishikawa *et al.*, 2008). LCB-Ps are hypothesised to function as signalling molecules just like in animals (Worrall *et al.*, 2008) and have been shown to play a negative role in reducing transpiration under thermal stress (Nishikawa *et al.*, 2008). *NAC053* is a transcriptional activator stimulating expression of ROS biosynthetic genes when activated under drough-induced senescent conditions (Lee *et al.*, 2012). In particular, *NAC053* participates in a

positive feedback loop mechanism promoting programmed cell death on exposure to heat stress, whereby *NAC053* activation induces hydrogen peroxide production and *vice versa* (Shih *et al.*, 2014).

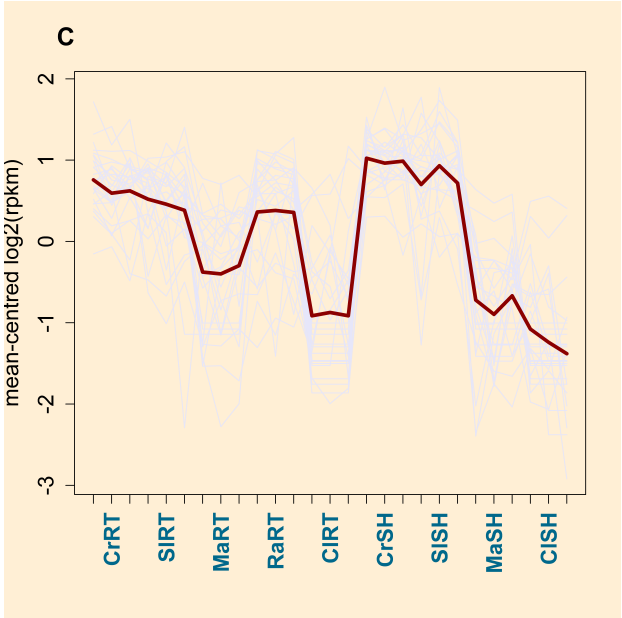
Finally, 15 genes related to salicylic acid (SA) signalling were downregulated in Carratraca-S. They had been discovered among 473 genes showing induction of expression upon salicylic acid application (Krinke *et al.*, 2007) and include transcription factors *WRKY75* and *WRKY28*. Both were shown to be upregulated in response to pathogenicity-determinant oxalic acid secreted by necrotrophic fungi (Chen *et al.*, 2013). The two genes regulate downstream defence response genes in the salicylic and jasmonic acid pathways. *WRKY28* acts a transcriptional activator of salicylic acid precursor biosynthesis gene *ISOCHORISMATE SYNTHASE 1* (*ICS1*, van Verk *et al.*, 2011) and both genes serve to induce an oxidative burst which slows down fungal growth and increases resistance to infection (Chen *et al.*, 2013). Overexpression of *WKRY28* with a transcription factor *bHLH17* in *Arabidopsis* confers increased tolerance to abiotic stresses: drought, salinity and oxidative stress (Babitha *et al.*, 2013). *WRKY75* possess also roles in root cell morphogenesis - suppresses root hair development in non-root hair files (Rishmawi *et al.*, 2014) and regulates lateral root architecture in general (Devaiah *et al.*, 2007). The last function has been linked to *WRKY75*'s role as a regulator of Pi deficiency response - it is induced under low Pi, acts as a positive regulator of many phosphate deficiency genes and is required for low Pi tolerance (Devaiah *et al.*, 2007).



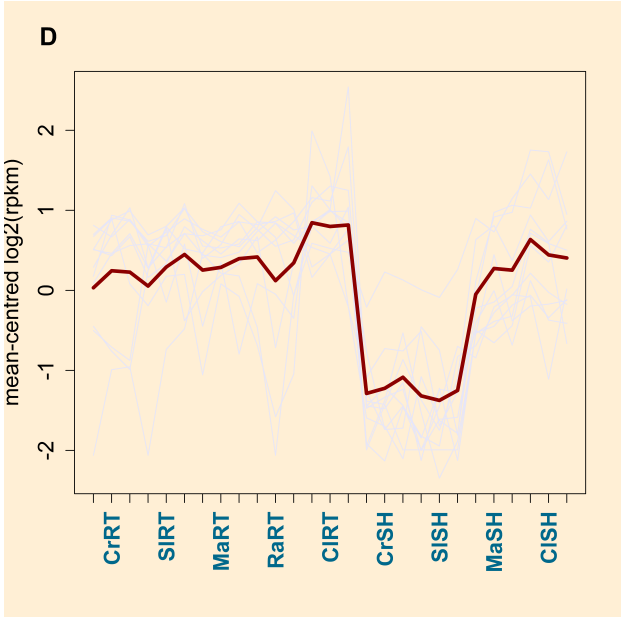
Total number of genes in the cluster: 10	
Gene ontology	**Genes:**
transmembrane transporter activity (3/1015)	AT2G38460 (*IREG1*), AT5G03570 (*IREG2*), AT5G38820 (AAAP family)
membrane (6/6079)	AT2G38460 (*IREG1*), AT5G03570 (*IREG2*), AT5G38820 (AAAP family), AT4G14230 (CBS domain-containing), AT4G33700 (CBS domain-containing), AT1G47330 (CBS domain-containing)
InterPro	
galactose oxidase/kelch, beta-propeller (1/152)	AT3G07720 (galactose oxidase/kelch repeat superfamily)
cystathionine beta-synthase, core (3/35)	AT4G14230 (CBS domain-containing), AT4G33700 (CBS domain-containing), AT1G47330 (CBS domain-containing)
domain of unknown function DUF21 (3/9)	AT4G14230 (CBS domain-containing), AT4G33700 (CBS domain-containing), AT1G47330 (CBS domain-containing)
Literature	
strong upregulation in response to iron deprivation:	
Dinneny et al., 2008 (3/51) and Long et al., 2010 (3/359)	AT3G07720 (galactose oxidase/kelch repeat superfamily), AT5G38820 (AAAP family), AT5G03570 (*IREG2*)
Buckhout *et al.,* 2009 (2/23)	AT3G07720 (galactose oxidase/kelch repeat superfamily), AT5G38820 (AAAP family)

B
mean-centred log2(rpkm)
-1.5
-1.0
-0.5
0.0
0.5
1.0
CrRT
SlRT
MaRT
RaRT
ClRT
CrSH
SlSH
MaSH
ClSH

Total number of genes in the cluster: 1
InterPro
cytochrome P450 (1/251)
AT3G20120 (CYP705A21)



Total number of genes in the cluster: 28	
Gene Ontology	**Genes:**
metal ion transport (5/573)	AT4G19690 (*IRT1*), AT5G64280 (*DIT2.2*), AT5G06740 (*LECRK-S.5*), AT3G16850 (pectin lyase-like superfamily), AT5G03570 (*IREG2*)
carboxylic acid transmembrane transporter activity (3/92)	AT5G64280 (*DIT2.2*), AT3G03720 (*CAT4*), AT5G38820 (AAAP family)
KEGG	
histidine biosynthesis (1/18)	AT1G58080 (*ATP-PRT1*)
InterPro	
galactose oxidase/kelch, beta-propeller (2/152)	AT3G04250 (F-box associated ubiquitination effector family), AT4G39600 (galactose oxidase/kelch repeat superfamily)
Literature	
strong upregulation in response to iron deprivation:	
Dinneny *et al.*, 2008 (4/51) and Long *et al*, 2010 (4/359)	AT4G19690 (*IRT1*), AT1G58080 (*ATP-PRT1*), AT5G38820 (AAAP family), AT5G03570 (*IREG2*)
Buckhout *et al.*, 2009 (2/23)	AT4G19690 (*IRT1*), AT5G38820 (AAAP family)



Total number of genes in the cluster: 13	
Gene Ontology	**Genes:**
response to karrikin (2/128)	AT3G57520 (*SIP2*), AT3G13790 (*CWINV1*)
respiratory burst involved in defense response (2/121)	AT3G09020 (alpha 1,4-glycosyltransferase family), AT3G13790 (*CWINV1*)

Figure 2.10: Overview of significantly (FDR < 0.01) differentially expressed genes when comparing *Alyssum* serpentine samples to *Alyssum* non-serpentine and *Clypeola* samples from a given tissue. RPKM values were first $\log_2$ transformed and then median-centered to allow plotting of genes with wide dynamic range of expression and highlight common patterns of expression across the samples, respectively. Individual genes are show in lavender and trendline (mean value for each sample) in dark red. Tables show select significantly enriched (FDR < 0.05) GO, KEGG and InterPro categories and literature datasets in each gene expression pattern grouping with associated DE gene names; the figures in parenthesis represent the number of genes in the dataset matching a given category compared to the *Arabidopsis* genome background. Legend: **A)** upregulated in all serpentine roots, **B)** downregulated in all serpentine roots, **C)** upregulated in all serpentine shoots, **D)** downregulated in all serpentine shoots. *Cr* - Carratraca-S, *Sl* - Samil-S, *Ma* - Morata-NS, *Cl* - Clypeola, *RT* - root samples, *SH* - shoot samples.

Total number of genes in the cluster: 15

Gene Ontology **Genes:**

amino acid transport (2/266) AT3G03720 (*CAT4*), AT1G07030 (mitochondrial substrate carrier family)

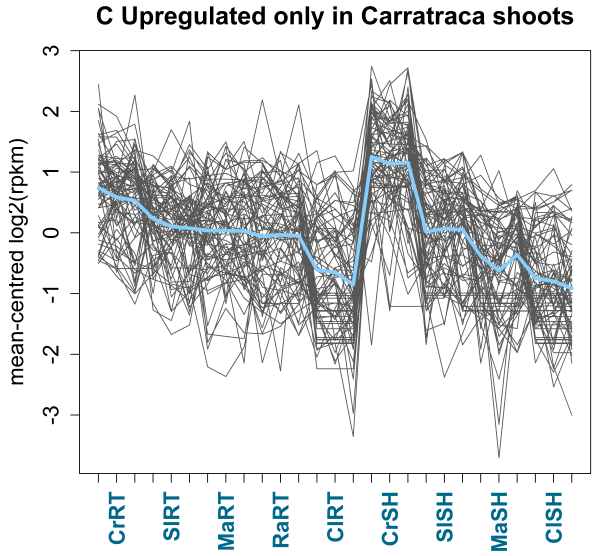
Total number of genes in the cluster: 12

KEGG

glutathione degredation (1/3)	AT4G29210 (*GGT3*)

InterPro

oxoglutarate/iron-dependent oxygenase (2/123)	AT1G06620 (2-oxoglutarate-dependent dioxygenase), AT4G16330 (2-oxoglutarate-Fe(II)-dependent oxygenase)



Total number of genes in the cluster: 77	
Gene Ontology	**Genes:**
transmembrane transporter activity (8/77)	AT2G37280 (*PDR5*), AT3G14410 (nucleotide/sugar transporter family), AT4G31710 (*GLR2.4*), AT1G20840 (*MT1*), AT1G09860 (*PUP16*), AT5G62160 (*ZIP12*), AT1G16780 (*VHP2;2*), AT2G34190 (xanthine/uracil permease family)
ubiquitin ligase complex (4/234)	AT1G56040 (HEAT/U-box domain-containing protein), AT1G11160 (transducin/WD40 repeat-like superfamily), AT5G51270 (U-box domain-containing protein kinase family), AT5G42190 (*ASK2*)
KEGG	
protein processing in endoplasmic reticulum (2/110)	AT5G41340 (*UBC4*), AT5G42190 (*ASK2*)
InterPro	
domain of unknown function DUF21 (2/9)	AT4G14230 (CBS domain-containing), AT4G33700 (CBS domain-containing)

Total number of genes in the cluster: 142	
Gene Ontology	**Genes:**
response to stress (43/4037)	AT1G06620, AT1G27980, AT2G34830, AT3G04570, AT4G37150, AT5G64905, AT1G71030, AT5G04080, AT2G34940, AT3G19030, AT4G39950, AT3G11340, AT2G23030, AT5G05340, AT1G70850, AT1G64160, AT1G74440, AT5G07100, AT5G26340, AT3G23430, AT4G14465, AT4G04955, AT2G02930, AT1G73260, AT4G19810, AT5G64120, AT5G47630, AT4G03400, AT2G46340, AT1G69930, AT2G37130, AT2G44490, AT3G26210, AT5G19875, AT3G54420, AT3G10500, AT3G51960, AT5G22500, AT5G24530, AT1G03220, AT4G23310, AT1G51790
nitrate transport (9/197)	AT2G34830 (*WRKY35*), AT2G23030 (*SNRK2.9*), AT1G07610 (*MTIC*), AT2G44490 (*BGLU26*), AT3G54420 (*EP3*), AT3G49210 (O-acyltransferase (WSD1-like) family), AT1G73260 (*KTI1*), AT5G64120 (*PRX71*), AT3G23430 (*PHO1*)
cellular modified amino acid metabolic process (15/708)	AT1G71030 (*MYBL2*), AT5G66040 (*STR16*), AT5G04080, AT3G19030, AT4G39950 (*CYP79B2*), AT3G04110 (*GLR1.1*), AT2G20960 (*PEARLI4*), AT5G22500 (*FAR1*), AT5G24530 (*DMR6*), AT2G44490 (*BGLU26*), AT1G69920 (*GSTU12*), AT1G09490 (alcohol dehydrogenase), AT1G64160 (*DIR5*), AT5G04950 (*NAS1*), AT3G23430 (*PHO1*)
defense response (23/1644)	AT3G04570 (*AHL19*), AT4G37150 (*MES9*), AT5G64905 (*PROPEP3*), AT5G04080, AT2G34940

	(*VSR5*), AT3G19030, AT4G39950 (*CYP79B2*), AT3G11340 (*UGT76B1*), AT5G24530 (*DMR6*), AT4G23310 (*CRK23*), AT1G51790 (leucine-rich repeat protein kinase family), AT2G37130 (peroxidase superfamily), AT2G02930 (*GSTF3*), AT1G73260 (*KTI1*), AT5G64120 (*PRX71*), AT2G44490 (*BGLU26*), AT3G26210 (*CYP71B23*), AT3G54420 (*EP3*), AT5G47630 (*MTACP3)* , AT1G70850 (*MLP34*), AT1G64160 (*DIR5*), AT1G74440 (DUF962 protein), AT4G14465 (*AHL20*)
toxin metabolic process (8/211)	AT1G27980 (*DPL1*), AT4G18170 (*WRKY28*), AT3G10500 (*NAC053*), AT1G69930 (*GSTU11*), AT1G69920 (*GSTU12*), AT5G59530 (2-oxoglutarate and Fe(II)-dependent oxygenase superfamily), AT3G49210 (O-acyltransferase (WSD1-like) family), AT2G02930 (*GSTF3*)
tryptophan / indokylamine catabolic process (5/79)	AT5G66040 (*STR16*), AT4G39950 (*CYP79B2*), AT3G04110 (*GLR1.1*), AT2G20960 (*PEARLI4*), AT2G44490 (*BGLU26*)
amino acid transport (8/266)	AT4G18170 (*WRKY28*), AT1G08230 (*GAT1*), AT1G51790 (leucine-rich repeat protein kinase family), AT3G04070 (*NAC047*), AT3G54420 (*EP3*), AT3G21690 (MATE efflux family), AT3G49210 (O-acyltransferase (WSD1-like) family), AT5G26340 (*STP13*)
phenylpropanoid biosynthetic process (8/326)	AT1G71030 (*MYBL2*), AT5G04080, AT3G19030, AT5G22500 (*FAR1*), AT5G24530 (*DMR6*), AT1G09490 (alcohol dehydrogenase), AT1G64160 (*DIR5*), AT3G23430 (*PHO1*)
transferase activity, transferring alkyl or aryl groups (6/120)	AT5G19040 (*IPT5*), AT1G69930 (*GSTU11*), AT1G69920 (*GSTU12*), AT2G02930 (*GSTF3*), AT3G63110 (*IPT3*), AT5G04950 (*NAS1*)
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen (10/461)	AT1G06620 (2-oxoglutarate-dependent dioxygenase-like), AT4G39950 (*CYP79B2*), AT5G24530 (*DMR6*), AT3G26210 (*CYP71B23*), AT3G20120 (*CYP705A21*), AT2G30830 (2-oxoglutarate-dependent dioxygenase-like), AT5G59530 (2-oxoglutarate and Fe(II)-dependent oxygenase superfamily), AT2G45510 (*CYP704A2*), AT5G25140 (*CYP71B13*), AT4G22710 (*CYP706A2*)
Literature	
induced expressed genes upon SA treatment: Krinke *et. al.*, 2007 (15/473)	AT1G65490, AT3G13062 (polyketide cyclase/dehydrase and lipid transport superfamily), AT3G11340 (*UGT76B1*), AT1G15380 (*GLYI4*), AT5G24530 (*DMR6*), AT4G39780 (ERF/AP2 transcription factor family), AT3G04070 (*SHG*), AT3G54420 (*EP3*), AT1G66920 (protein kinase superfamily), AT3G29034, AT4G36040 (*DJC23*), AT1G32460, AT1G26800 (RING/U-box superfamily), AT5G26340 (*STP13*), AT5G13080 (*WRKY75*)

Figure 2.11: Overview of significantly (FDR < 0.01) differentially expressed genes when comparing *Alyssum* Carratraca-S samples to *Alyssum* serpentine and *Clypeola* samples from a given tissue. Only genes that are specific to Carratraca-S and not DE in *Alyssum* Samil-S are shown here. RPKM values were first $\log_2$ transformed and then median-centered to allow plotting of genes with wide dynamic range of expression and highlight common patterns of expression across the samples, respectively. Individual genes are show in gray and trendline (mean value for each sample) in blue. Tables show select significantly enriched (FDR < 0.05) GO, KEGG and InterPro categories and literature datasets in each gene expression pattern grouping in each gene expression pattern grouping with associated DE gene names; the figures in parenthesis represent the number of genes in the dataset matching a given category compared to the *Arabidopsis* genome background. Legend: **A)** upregulated only in Carratraca-S roots, **B)** downregulated only in Carratraca-S roots, **C)** upregulated only in Carratraca-S shoots, **D)** downregulated only in Carratraca-S shoots. *Cr* - Carratraca-S, *Sl* - Samil-S, *Ma* - Morata-NS, *Cl* - Clypeola, *RT* - root samples, *SH* - shoot samples.

Total number of genes in the cluster: 10

Gene Ontology	Genes:
copper-transporting ATPase activity (1)	AT1G63440 (*HMA5*)
ribonuclease H-like (2/112)	AT4G29090 (ribonuclease H-like superfamily), AT2G27880 (*AGO5*)

Total number of genes in the cluster: 8

No enriched gene category.

Total number of genes in the cluster: 68	
Gene Ontology	**Genes:**
glutathione transferase activity (2/48)	AT2G29470 (*GSTU3*), AT1G02950 (*GSTF4*)
substrate-specific transmembrane transporter activity (5/822)	AT4G33020 (*ZIP9*), AT1G67940 (*NAP3*), AT2G38100 (POT family), AT1G55910 (*ZIP11*), AT5G25400 (nucleotide-sugar transporter family)

Total number of genes in the cluster: 68	
Gene Ontology	**Genes:**
response to stress (20/4037)	AT3G28940, AT4G03560, AT4G31120, AT3G07040,0 AT1G13440, AT3G54820, AT1G60710, AT4G12000, AT1G80130, AT3G11050, AT5G49570, AT1G12520, AT1G42700, AT3G06490, AT2G23620, AT4G33050, AT1G73500, AT1G08940, AT2G44180, AT3G30775,
response to ethylene stimulus (4/353)	AT3G53280 (*CYP71B5*), AT3G28580 (P-loop containing nucleoside triphosphate hydrolases superfamily), AT3G06490 (*MYB108*), AT1G73500 (*MKK9*)
defense response (9/1644)	AT3G28940 (AIG2-like family), AT3G07040 (*RPS3*), AT1G13440 (*GAPC2*), AT3G06490 (*MYB108*), AT3G30775 (*PDH1*), AT2G23620 (*MES1*), AT4G33050 (*IQM1*), AT1G08940 (phosphoglycerate mutase family), AT2G44180 (*MAP2A*)
oxidation-reduction process (8/1368)	AT3G53280 (*CYP71B5*), AT1G13440 (*GAPC2*), AT1G60710 (*ATB2*), AT3G11050 (*FER2*), AT1G12520 (*CCS*), AT3G30775 (*PDH1*), AT1G55700 (cysteine/histidine-rich C1 domain family), AT1G33730 (*CYP76C5*)
tryptophan / indokylamine catabolic process (3/79)	AT3G28940 (AIG2-like family), AT4G33050 (*IQM1*), AT1G73500 (*MKK9*)
Literature	
downregulated in *WRKY15* overexpressor line: Vanderauwera *et al*., 2012 (7/750)	AT4G16630 (DEA(D/H)-box RNA helicase family protein), AT3G53280 (*CYP71B5*), AT3G07040 (*RPS3*), AT1G13440 (*GAPC2*), AT1G29940 (*NRPA2*), AT1G63780 (*IMP4*), AT1G73500 (*MKK9*)

Figure 2.12: Overview of significantly (FDR < 0.01) differentially expressed genes when comparing *Alyssum* Samil-S samples to *Alyssum* non-serpentine and *Clypeola* samples from a given tissue. Only genes that are specific to Samil-S and not DE in *Alyssum* Carratraca-S are shown here. RPKM values were first $\log_2$ transformed and then median-centered to allow plotting of genes with wide dynamic range of expression and highlight common patterns of expression across the samples, respectively. Individual genes are show in gray and trendline (mean value for each sample) in purple. Tables show select significantly enriched (FDR < 0.05) GO, KEGG and InterPro categories and literature datasets in each gene expression pattern grouping with associated DE gene names in each gene expression pattern grouping with associated DE gene names; the figures in parenthesis represent the number of genes in the dataset matching a given category compared to the *Arabidopsis* genome background. Legend:: **A)** upregulated only in Samil-S roots, **B)** downregulated only in Samil-S roots, **C)** upregulated only in Samil-S shoots, **D)** downregulated only in Samil-S shoots. *Cr* - Carratraca-S, *Sl* - Samil-S, *Ma* - Morata-NS, *Cl* - Clypeola, *RT* - root samples, *SH* - shoot samples.

2.3.4 SNP calling

In total, 1,042,563 biallelic SNPs were detected in the four *Alyssum serpyllifolium* populations under study (Table 2.5): 894,219 were found in 18,344 contigs in the consensus assembly, followed by 148,344 located in 7,228 contigs in the supplementary assembly. Looking at individual populations: Carratraca-S (Table 2.6), Morata-NS (Table 2.8), Samil-S (Table 2.7) and Rubía-NS (Table 2.9) contained 480,059 (193,947 private), 502,900 (196,046 private), 500,328 (207,139 private) and 198,826 (57,471 private) SNPs, respectively. The lower number of SNPs in Rubía-NS is a direct consequence of the lower number of individuals sequenced in that population - 3, as opposed to 6 in all the other populations.

Table 2.5: SNP statistics on all *Alyssum* transcriptomes combined

	main	***contigs no.***	***suppl.***	***contigs no.***
all, including low frequency (< 5%) SNPs				
biallelic	894,219	18,344	148,344	7,228
multiallelic	31,224	11,428	6,808	2,967
min. 5% frequency variants				
biallelic	472,311	17,964	74,765	6,819
multiallelic	22,153	9,550	4,424	2,261
min. 5% frequency variants and min. 5 genotypes/population				
biallelic	292,755	12,934	17,710	2,258
multiallelic	12,200	6,164	858	551

When low-frequency variants (< 5%, effectively present just once in the whole dataset) were excluded, the number of biallelic SNPs almost halved, down to 547,076 SNPs: 472,311 across 17,964 contigs in the consensus assembly, and 74,765 across 6,819 contigs in the supplementary assembly. Consequently, a greater proportion of the remaining SNPs was shared across populations: Carratraca-S contained 327,645 SNPs (54,170 private), Morata-NS 343,532 (50,845 private), Samil-S 335,122 (55,359 private) and Rubía-NS 156,556 (19,305 private).

Table 2.6: SNP statistics on the Carratraca-S transcriptome

	main	***(private)***	***suppl.***	***(private)***
all, including low frequency (< 5%) SNPs				
biallelic	412,836	164,236	67,223	29,711
multiallelic	18,029	4,159	3,412	1,009
min. 5% frequency variants				
biallelic	284,234	46,792	43,411	7,378
multiallelic	14,982	1,585	2,684	335
min. 5% frequency variants and min. 5 genotypes/population				
biallelic	175,671	27,450	11272	1780
multiallelic	8,350	786	609	63

Table 2.7: SNP statistics on the Samil-S transcriptome

	main	***(private)***	***suppl.***	***(private)***
all, including low frequency (< 5%) SNPs				
biallelic	435,935	176,394	64,393	30,745
multiallelic	17,877	3,769	3,449	1,163
min. 5% frequency variants				
biallelic	292,609	45,230	42,513	10,129
multiallelic	14,947	1,315	2,805	565
min. 5% frequency variants and min. 5 genotypes/population				
biallelic	184,058	28,500	10,625	1,552
multiallelic	8,391	668	592	55

Table 2.8: SNP statistics on the Morata-NS transcriptome

	main	***(private)***	***suppl.***	***(private)***
all, including low frequency (< 5%) SNPs				
biallelic	432,117	165,489	70,783	30,557
multiallelic	18,742	4,212	3,802	1,226
min. 5% frequency variants				
biallelic	297,574	42,961	45,958	7,884
multiallelic	15,516	1,473	2,860	396
min. 5% frequency variants and min. 5 genotypes/population				
biallelic	183,331	25,094	11,546	1,522
multiallelic	8,615	752	622	64

Table 2.9: SNP statistics on the Rubia-NS transcriptome

	main	***(private)***	***suppl.***	***(private)***
all, including low frequency (< 5%) SNPs				
biallelic	174,603	49,379	24,223	8,092
multiallelic	8,078	1,019	1,279	266
min. 5% frequency variants				
biallelic	138,485	17,000	18,071	2,305
multiallelic	7,381	483	1,095	98
min. 5% frequency variants and min. 5 genotypes/population				
biallelic	102,657	12,568	6,432	770
multiallelic	4,995	292	372	28

Many of the analyses undertaken required that only a low amount of data is missing, and in these instances, a further filtered dataset of 310,465 biallelic SNPs from across 15,192 contigs was used. All of these SNPs were genotyped in at least 5 individuals in Carratraca-S, Morata-NS and Samil-S populations and in all 3 individuals in Rubía-NS. In the individual populations in this filtered data set, the number of biallelic SNPs was: 186,943 in Carratraca-S (29,230 private), 194,877 in Morata-NS (26,616 private), 194,683 in Samil-S (30,052 private) and 109,089 in Rubía-NS (13,338 private).

Also, over 70,000 indels spread among approximately half of the contigs (data not shown) were detected but they were ignored in the current analysis.

2.3.5 Population assignment

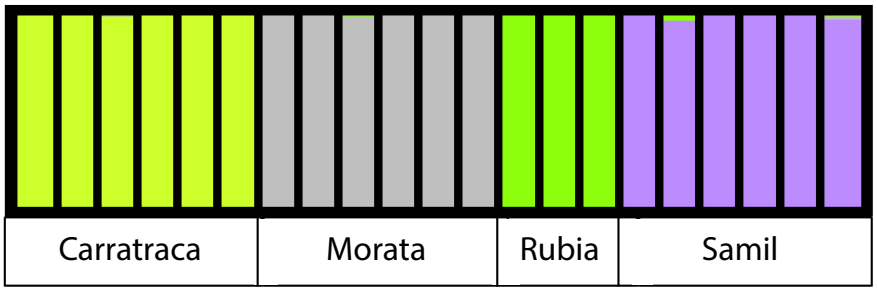


Figure 2.13: Cluster assignment to individuals in Carratraca-S, Morata-NS, Rubia-NS, Samil-S according to $K = 4$

To test that four clearly delimited Hardy-Weinberg populations individually with limited admixture were being dealt with, Structure software was first used to partition the *Alyssum serpyllifolium* individuals into a different number of clusters ($K = 1$ to $K = 5$) based on a matrix containing 13,050 putatively neutral, unlinked markers. Structure analysis results based on Evanno's ΔK method (Evanno *et al.*, 2005) as well as mean LnP(D) change (data not shown) point to a four-population structure in the current study, confirming the Carratraca-S, Morata-NS, Rubía-NS and Samil-S labels (Figure 2.13). Interestingly, alternative K values ($K = 2$) and ($K = 3$) grouped individuals according to their geographic proximity (Figure 2.14). Similar conclusions were drawn from PCA analysis

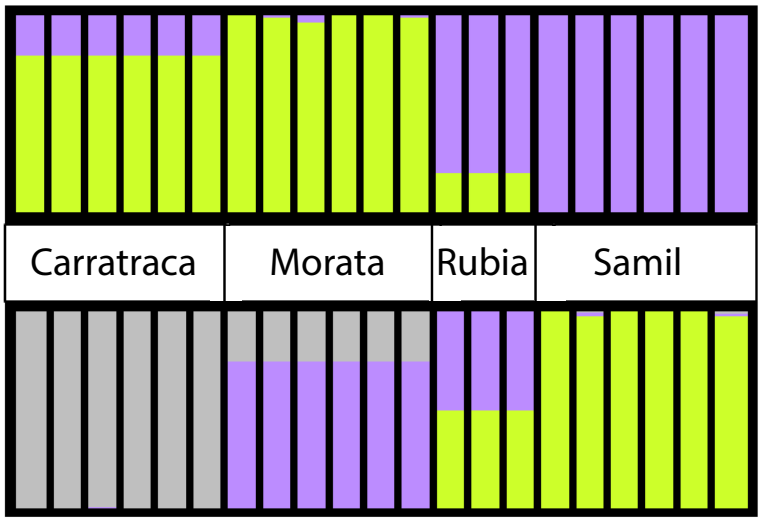


Figure 2.14: Cluster assignment to individuals in Carratraca-S, Morata-NS, Rubia-NS, Samil-S according to $K = 2$ (top) $K = 3$ (bottom)

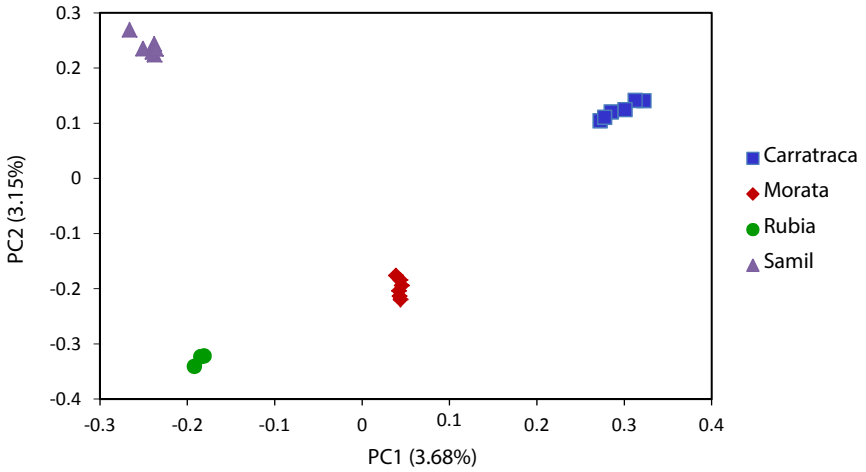


Figure 2.15: Results of PCA analysis on SNPs genotyped in all individuals

using extended sampling of 265,775 SNPs – all biallelic SNPs containing no missing data in any individual (Figure 2.15).

2.3.6 Comparison of population genetics parameters

Little is known about the mating system, population structure and degree of gene flow between populations of *Alyssum serpyllifolium*. In general, species in the genus *Alyssum* are mostly allogamous with the exception of some perennials (Cecchi *et al.*, 2013). The mating system of *Alyssum serpyllifolium* has not been targeted by any research previously, to the

best of my knowledge. Other Brassicaceae metal hyperaccumulators have either totally self-incompatible, fully outcrossing sexual mating system (*Arabidopsis halleri*, Llaurens *et al.*, 2008) or a mixed mating system (*Noccea caerulescens*, Mousset *et al.*, 2016, selfing rates ranging from 0.2 to 0.5).

Here, this perennial plant (Küpfer & Nieto Feliner, 1993) is shown to have a high inbreeding coefficient F_{IS} (Nei *et al.*, 1983) amongst individuals in all the populations, with mean inbreeding coefficient for individuals ranging from 0.42 to 0.53 and no significant difference observed between populations (data not shown). High inbreeding coefficient indicates more inbreeding than expected assuming random mating and can be a product of (partial) selfing, small population size, population subdivision (Wahlund effect) and habitat fragmentation (Charlesworth & Charlesworth, 1987; Young *et al.*, 1996). Slightly smaller F_{IS} values: mean 0.38 ± 0.01 and ranging from 0.19 to 0.37, were discovered in comprehensive microsatellite surveys of *N. caerulescens* populations from metalliferous and non-metalliferous localities (Besnard *et al.*, 2009; Mousset *et al.*, 2016). This species has been shown to be self-compatible with strong variation in outcrossing rate amongst populations (Riley, 1956; Dubois *et al.*, 2003; Mousset *et al.*, 2016). Previous studies of pseudometallophytes showed elevated inbreeding in metallicolous populations, e.g. in *Anthoxanthum odoratum*, *Agrostis tenuis* (Antonovics & Bradshaw, 1968; Antonovics, 1972), *Armeria maritima* (Lefèbvre, 1976). Self-fertility in metallophytes has been hypothesised to be selected for to reduce gene flow from non-adapted populations (Antonovics & Bradshaw, 1968) and/or for reproductive assurance amongst the founders of new metalliferous sites (Lefèbvre, 1976) but the hypothesis has not gained support in the literature (Dubois *et al.*, 2003; Leimu & Fischer, 2008; Hereford, 2010; Mousset *et al.*, 2016).

Elements of *A. serpyllifolium* morphology (flower size etc.) and F_{IS} values suggest it to be (at least partially) out-crossing. The results presented below indicate that different *Alyssum serpyllifolium* populations are capable of interbreeding, or were until recently as evidenced by limited levels of divergence observed. Furthermore, moderate nucleotide diversity observed does not support total shift to autogamy (Leffler *et al.*, 2012) in *Alyssum*

serpyllifolium: Nei's π (Nei & Li, 1979) ranged from median 0.00490 and mean 0.00604 (± 0.00511 SD) in Samil-S, through median 0.00504 and mean 0.00618 (± 0.00508 SD) in Morata-NS to 0.00506 and 0.00625 (± 0.00529 SD) in Carratraca-S in the coding region, respectively (Figure 2.16).

Taking into account all the SNPs regardless of frequency, average density of synonymous SNPs was higher than that of non-synonymous SNPs in all the *Alyssum serpyllifolium* populations (Figure 2.17), later shown to be likely a consequence of pervasive genome-wide purifying selection encountered in many species (Mitchell-Olds *et al.*, 2007). Between 35.6% and 37.5% contigs contained 0 synonymous SNPs on average per 100 bp, while for non-synonymous SNPs the fraction increased to between 58.6% and 59.9%. The majority of contigs contained 1 synonymous SNP per 100 bp – between 44.7% and 46.9%, while for non-synonymous SNPs the fraction decreased to between 33.7% and 34.5%. Again, much higher proportion of contigs contained 2 synonymous SNPs per 100 bp than non-synonymous SNPs – between 13.7% and 14.3% versus 4.8% and 5.3%, respectively. Lastly, between 3.4% and 4.1% contigs contained > 2 synonymous SNPs, and between 1.6% and 1.7% contigs contained > 2 non-synonymous SNPs.

Absolute divergence D_{xy} (Nei, 1987) between populations (Figure 2.18) varied significantly between population comparisons when making each of the possible three comparisons (paired *t*-test, two-tailed *p*-value < 10^{-15}). Highest divergence was found between the two serpentine populations, Carratraca-S and Samil-S: mean 0.01037 (± 0.00693 SD), followed by Morata-NS and Samil-S: mean 0.01013 (± 0.00701 SD) with the lowest divergence found between Carratraca-S and Morata-NS: mean 0.00985 (±0.00693 SD). Top 1% of loci with the highest D_{xy} in Morata-NS-Carratraca-S comparison contained a slight overrepresentation of serpentine differentially expressed genes (Fisher's exact test, *p*-value < 0.009).

Figure 2.16: Nucleotide diversity (π) distribution across all the loci in Carratraca-S, Samil-S and Morata-NS

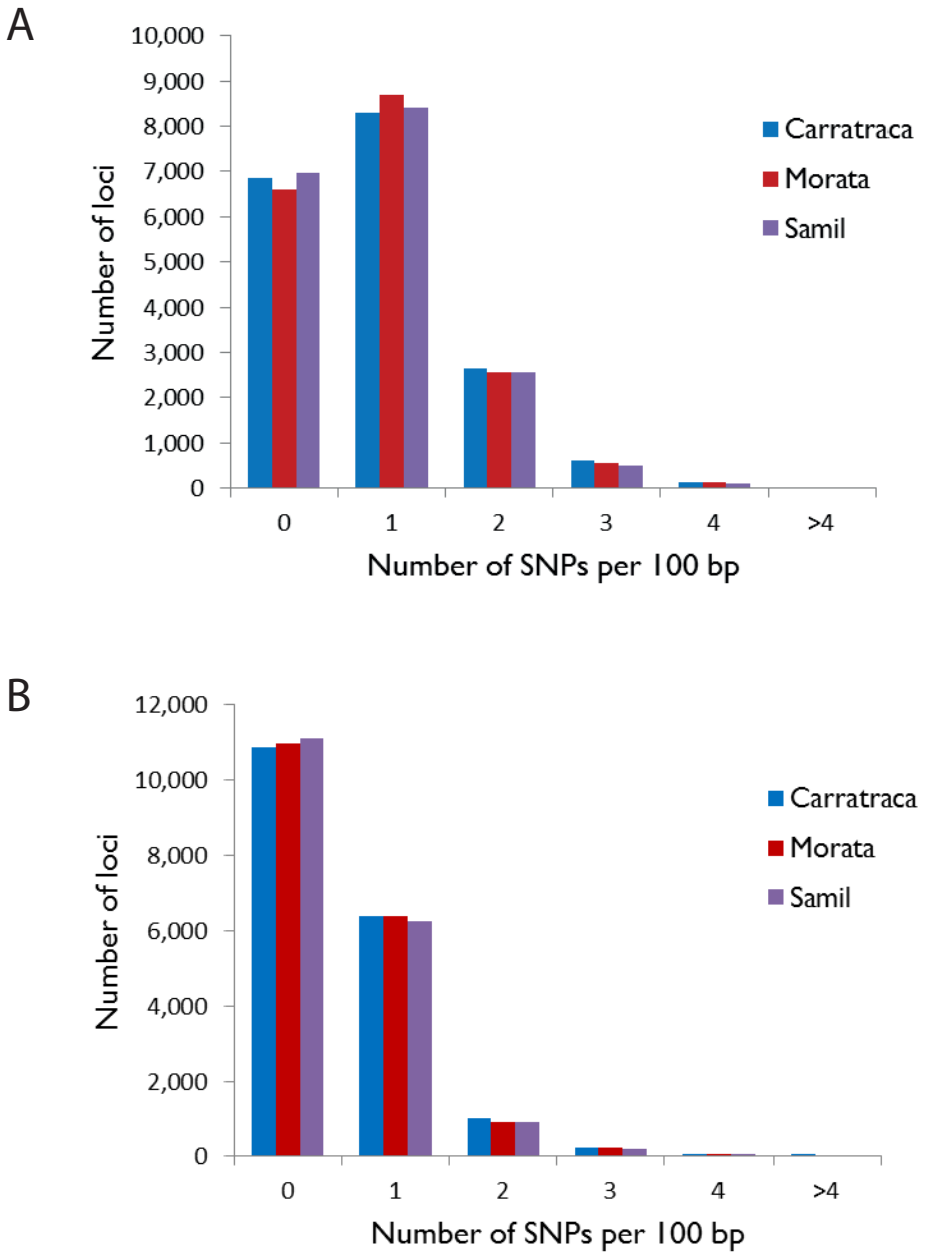


Figure 2.17: Average frequency of occurrence of **(A)** synonymous and **(B)** non-synonymous SNPs per 100 bp of coding sequence across loci in the *Alyssum* assembly

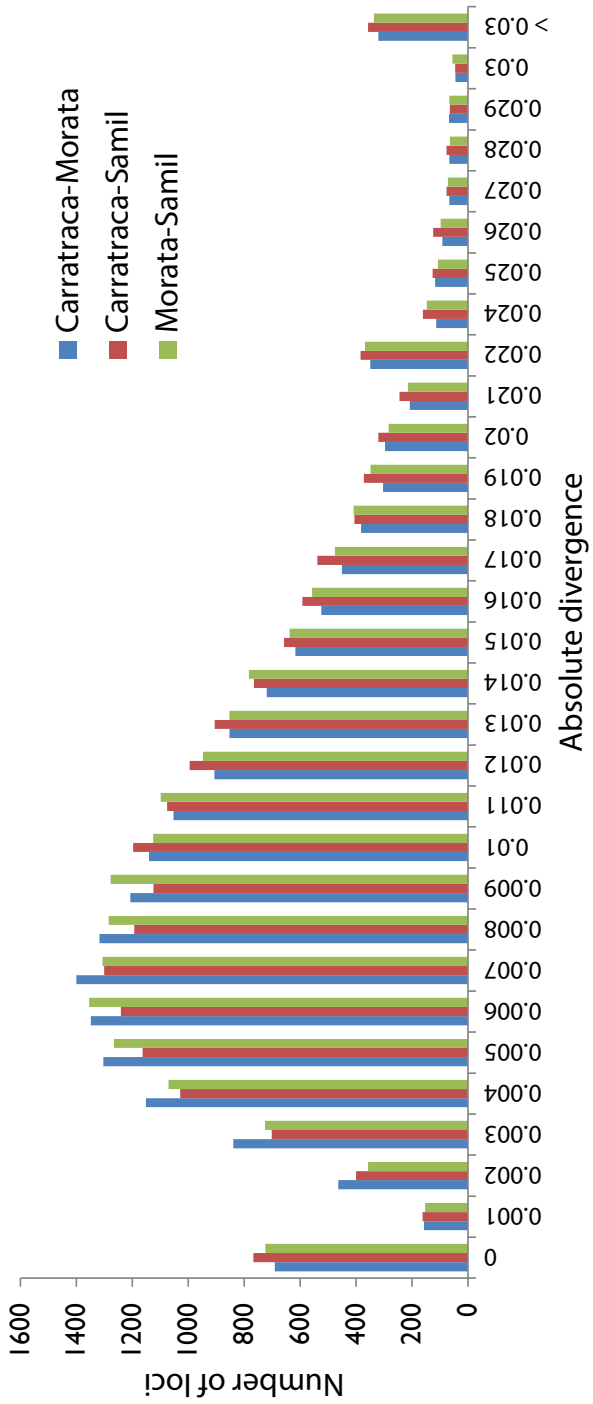


Figure 2.18: Comparison of distribution of pairwise D_{xy} distances between the *Alyssum serpyllifolium* populations

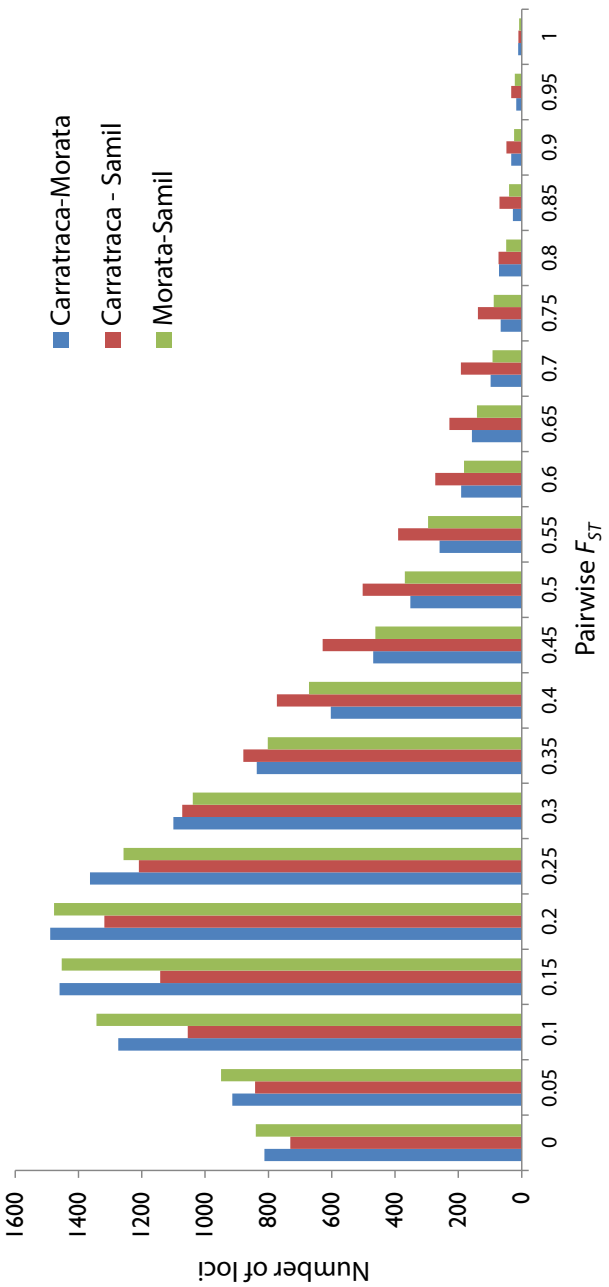


Figure 2.19: Comparison of distribution of pairwise F_{ST} distances between the *Alyssum serpyllifolium* populations

Similarly, when using Weir and Cockerham's (1983) F_{ST} as a distance metric (Figure 2.19), Carratraca-S-Samil-S had the highest mean locus F_{ST} at 0.223 (± 0.179 SD), which was significantly higher than in both non-serpentine-serpentine comparisons: Morata-NS-Carratraca-S (mean: 0.191 ± 0.159, paired t-test, two-tailed p-value < 10^{-15}) and Morata-NS-Samil-S (mean 0.190 ± 0.159, paired t-test, two-tailed p-value < 10^{-15}); however no significant difference was detected between the former and the latter's F_{ST} distributions. F_{ST} values for individual populations (Figure 2.20) pointed to the highest differentiation in Carratraca-S: 0.257 (± 0.159), followed by Samil-S: 0.240 (± 0.153) and Morata-NS: 0.230 (± 0.141 SD), and the differences between populations were significant when compared in pairwise manner at p-value < 10^{-15}. Much lower differentiation was found in a study of five Alpine populations of *Arabidopsis halleri* (Fischer *et al.*, 2013), where mean F_{ST} lay at 0.038, and ranged 0.026–0.048 over 2 million SNPs.

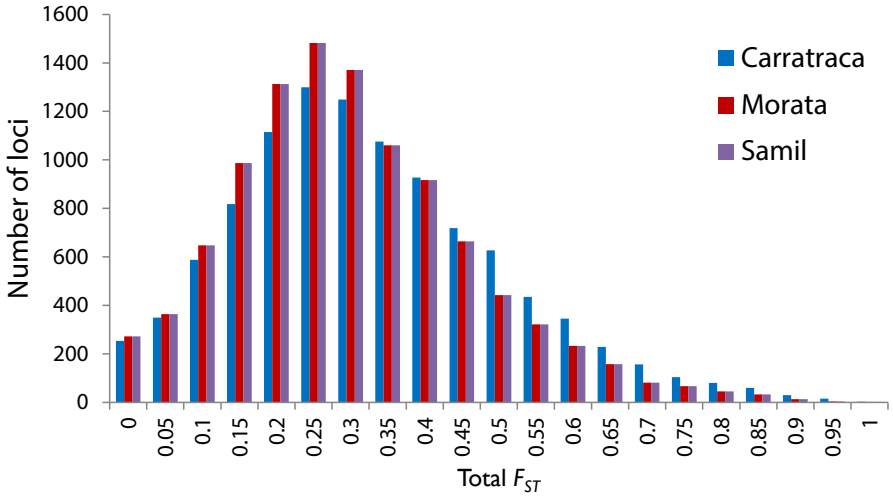


Figure 2.20: Total F_{ST} distribution across all the loci in Carratraca-S, Samil-S and Morata-NS

2.3.7 Evidence for recent selection in the *Alyssum* transcriptomes

Different histories of selective pressure and drift between demes can result in increased genetic differentiation, commonly measured by F_{ST}. A number of Bayesian methods have been developed which help pinpoint areas of the genome that show elevated F_{ST} and therefore were likely targeted by positive directional selection (Stephan, 2015). To detect genes under recent positive selection in serpentine relative to non-serpentine

populations, Bayescan was first run in two-population mode, taking Morata-NS and each of the serpentine populations in turn. The SNP loci in the far right tail of the distribution were then identified. All of the outlier SNPs showed fixation for divergent alleles in the serpentine and non-serpentine populations. SNPs containing the serpentine allele in the non-serpentine population Rubía-NS which had not been included in the Bayescan analysis were excluded. Two datasets containing alleles specific to Samil-S or Carratraca-S relative to the two non-serpentine populations were finally arrived at, some of which are likely to be important for hyperaccumulation and serpentine adaptation (referred to as "Bayescan outliers"). A complementary dataset was then created containing all SNPs showing the same pattern of allele frequencies between each serpentine and the two non-serpentine populations (referred to as "Fixed outliers") – fixation for divergent alleles in the dataset, regardless of the fact that some of them were omitted by Bayescan, which does not output all the loci simply containing the highest allele frequency differences. Obviously, the label of outlier and non-outlier is dependent on the detection methods employed, and could be misleading in the expected false positive and negative cases, which can be explained by a multitude of factors, such as low heterozygosity rates and hence inflated F_{ST} estimates in a given genomic regions without significant contribution from divergent selection (Cruickshank & Hahn, 2014) or demographic history factors, such as hidden hierarchical population structure (Excoffier *et al.*, 2009).

In total, 1357 fixed SNPs (1291 Bayescan outliers) were detected in Carratraca-S population, while 1324 fixed SNPs (340 Bayescan outliers) were detected in Samil-S population. When it came to actual transcripts, the fixed SNPs were distributed amongst 480 genes in Carratraca-S (448 Bayescan outliers) and amongst 481 genes in Samil-S (193 Bayescan outliers). Genes where all the alleles present in the serpentine population were ancestral relative to *Clypeola* and/or missing were then eliminated, but such genes formed a small fraction of all outlier loci (92 genes/154 SNPs in Carratraca-S and 112 genes/197 SNPs in Samil-S) and are henceforth referred to ancestral outlier loci, as opposed to serpentine outlier loci. Unless explicitly stated otherwise, from now on only serpentine outlier loci will be discussed.

Further support for functional role in adaptation to hyperaccumulation and serpentine comes from comparing the ancestral status of the serpentine outlier alleles with that of the genome background. High number of Bayescan/fixed outlier loci with derived alleles (serpentine outlier loci) mentioned in the previous paragraph contrasts with low incidence (15%) of derived alleles amongst major alleles in the rest of the genome (Fisher's exact test, p-value < 10^{-15}). Also, the number of non-synonymous substitutions was significantly higher (Fisher's exact test, p-value < 10^{-5}) amongst the serpentine outliers (43-45%) than that expected given the incidence of non-synonymous substitutions in the rest of the genome (36%).

The magnitude of overlap at the SNP and locus level was highly significant between the two serpentine populations. Less than 1 in 5 of the fixed SNPs in each population were shared between Carratraca-S and Samil-S: 142 fixed (maximum 15 expected due to chance), distributed over 64 transcripts and 42 Bayescan (maximum 7 expected due to chance) outliers. At the level of the entire transcript, 77 genes common to both fixed datasets (maximum 28 expected due to chance) were found and 36 (maximum 20 expected due to chance) when looking at Bayescan outliers. Also, in both populations 45% of outlier alleles were only detected in one of the serpentine populations and so are absent/segregating at low frequencies below detection limit in the other serpentine population. The most overrepresented functional category amongst fixed and Bayescan outliers (data not shown) in the two populations was *metal ion transmembrane activity* (7 genes out of 157 available in *A. thaliana* genome), and other categories related to cation/sugar substrate-specific transport and membranes were highly overrepresented. In general, categories significantly enriched among the shared and population-specific outlier genes were strikingly similar. They included those related to leaf and root development – e.g. t*richome and hair cell elongation and differentiation, meristem initiation*; defence – e.g. *response to wounding, response to bacterium/nematode/virus/fungus/chitin, systemic acquired resistance*; *response to abiotic stresses*, including *salt stress*, *dehydration, heat, cold and redox stress*; regulation of protein activity – e.g. *kinase activity*, epigenetic modifications (*methylation, acetylation)*

and last but not least metal and nutrient homeostasis – e.g. *metal transport, iron deficiency response*.

Lastly, there was significantly increased representation of genes upregulated in serpentine populations amongst Bayescan and fixed outliers in both Samil-S and Carratraca-S populations as well as amongst outliers shared by the two serpentinophytes (Fisher's exact test, p-value < 0.05). Such genes are amongst the strongest candidates for functional importance in adaptation to serpentine environment and, include *CAT4, IREG1, CBS-DUF21* family members, *ATP-PRT1*, and a SGNH hydrolase-type esterase superfamily protein in the two serpentine populations, in addition to *HMA5* in Samil-S. While the majority of SNPs defining the expression levels occur in the regulatory regions which are not well-sampled in RNA-Seq studies, this significant overrepresentation of upregulated genes amongst the outliers could be indicative of the presence of an indirect link, due to tight linkage disequilibrium influenced by e.g. a selective sweep. Nucleotide diversity was significantly lower amongst Bayescan/fixed outlier genes compared to the remainder of the genome, which supports the hypothesis that outliers arose throught the action of selection and not other evolutionary processes. In Carratraca-S, mean nucleotide diversity in the non-outliers was 0.0063, whereas Bayescan and fixed outliers's π centered at 0.00342 and 0.00353 (paired t-test, two-tailed p-value < 10^{-15}), accordingly. Similarly, in Samil-S, mean π was 0.00621 and in Bayescan and fixed outliers, mean π centered at 0.00430 (paired t-test, two-tailed p-value < 10^{-15}) and 0.00503 (paired t-test, two-tailed p-value < 10^{-10}), respectively. Following the selective sweep, new rare mutations start accumulating in the previously homogenous region of low genetic diversity. This distortion in the allele frequency spectrum is detectable for thousands of generations and a range of population genetics tests were devised to detect it, the first and most of popular of which is Tajima's D (Tajima, 1989). Tajima's D compares the average nucleotide diversity (π) with a total number of segregating sites in the sample (S) by subtracting S from π. Under the assumptions of neutrality, these two values should be equal. An excess of rare alleles will result in disproportionate increase in S in relation to π and in more negative D values, which could either indicate previous action of positive selection or population expansion

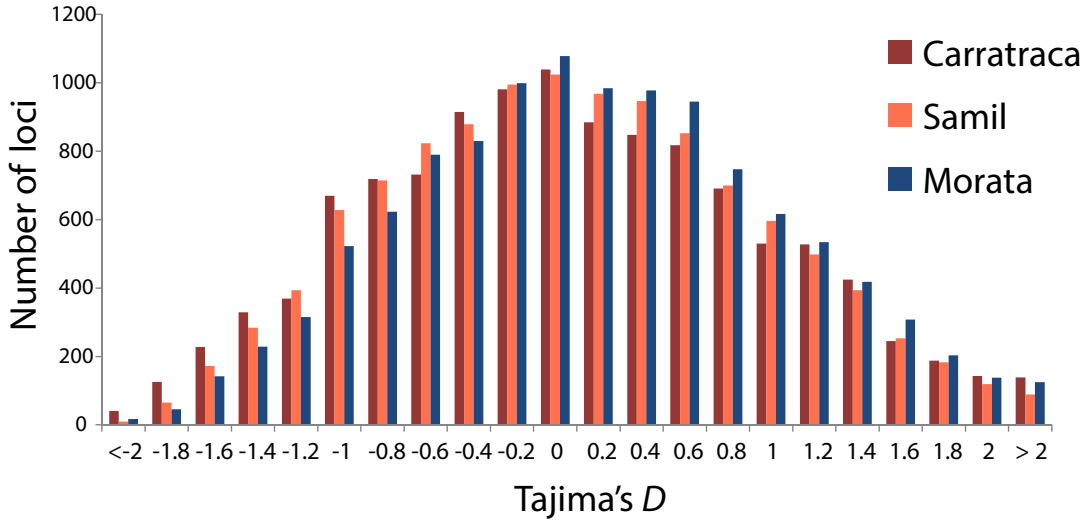


Figure 2.21: Tajima's *D* distribution across all the loci in Carratraca-S, Samil-S and Morata-NS

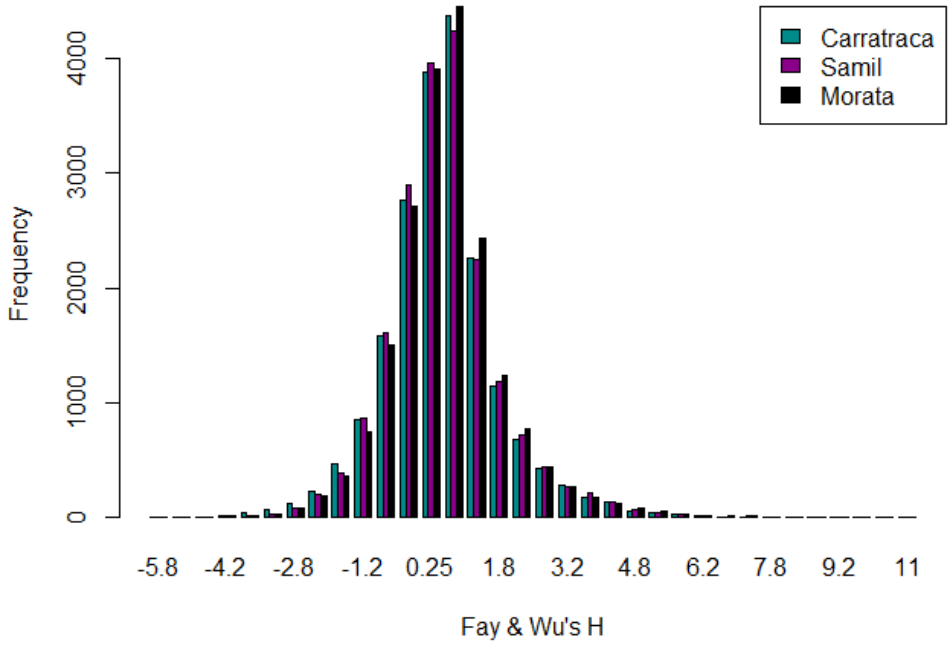


Figure 2.22: Fay & Wu's *H* distribution across all the loci in Carratraca-S, Samil-S and Morata-NS

(Hartl & Clark, 2007; Vitti *et al.*, 2013). One of the tests which are extensions to Tajima's *D* approach, Fay and Wu's *H* (Fay & Wu, 2000) is often used to distinguish the two alternative reasons for significantly negative Tajima's *D* values. Fay and Wu's *H* takes advantage of an outgroup which specifies the ancestral and derived states of alleles. Under the scenario of a

selective sweep, we expect to find derived alleles in linkage disequilibrium with the causal allele to hitch-hike to high frequency. The resulting excess of high-frequency derived alleles is reflected in small values of *H* and confirms the action of positive selection in the gene.

One more statistic, *DH*, which combines the strengths of Tajima's *D* and Fay and Wu's *H* was introduced by Zeng *et al.* (2006). Furthermore, Zeng *et al.* (2006) introduced a new statistic, Zeng's *E*, contrasting the high and low frequency portions of the frequency spectrum, which is unlike Fay and Wu's *H* and Tajima's *D* using intermediate frequency alleles as the baseline for comparison. Zeng's *E* is enhanced by the co-option of Ewens-Watterson test (Ewens, 1972; Watterson, 1975), which contrasts the observed levels of Hardy-Weinberg homozygosity with that expected under a neutral model and thus makes the test mostly insensitive to recombination (Vitti *et al.*, 2013).

Tajima's *D* distribution of Morata-NS and the two serpentine populations was significantly different (Figure 2.21) with Morata-NS's *D* centred at 0.068 (± 0.852 SD), and significantly lower (paired *t*-test, two-tailed *p*-value < 10^{-15}) at mean −0.012 (± 0.900 SD) in Samil-S and −0.029 (± 0.852 SD) in Carratraca-S. Tajima's *D* in Samil-S was also slightly but significantly higher than that of Carratraca-S (paired *t*-test, two-tailed *p*-value < 0.005). An analogous trend in significant differences between populations was seen in the distribution of Tajima's *D* extension test Fay & Wu's *H* (Figure 2.22). This implies a higher number of rare alleles segregating in the serpentine populations relative to Morata-NS, but the degree of difference observed is not so big as to suggest a stronger influence of population expansion and selective sweeps in the recent history of those populations.

In two other tests related to Tajima's *D*: Fu & Li's *D* (Figure 2.23) and Zeng's *E* (Figure 2.24), significantly lower values (paired *t*-test, two-tailed *p*-value < 10^{-15}) were obtained in Samil-S (mean Fu & Li's *D*: 0.021 ± 0.886 SD, mean Zeng's *E*: -0.541 ± 1.254 SD) relative to Carratraca-S (mean Fu & Li's *D*: 0.438 ± 1.173 SD, mean Zeng's *E*: -0.407 ± 1.235 SD) and Morata-NS (mean Fu & Li's *D*: 0.398 ± 0.884 SD, mean Zeng's *E*: -0.324 ± 1.223 SD).

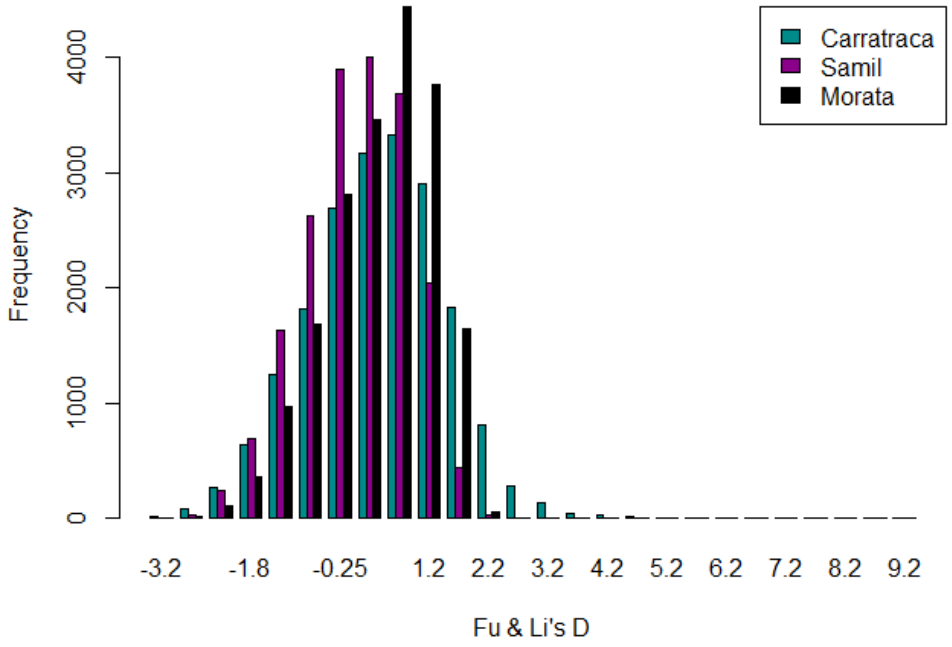


Figure 2.23: Fu & Li's *D* distribution across all the loci in Carratraca-S, Samil-S and Morata-NS

However, an opposite trend exists between Carratraca-S and Morata-NS populations when comparing the two statistics. Carratraca-S's Fu & Li's *D* values are slightly higher than that of Morata-NS (paired *t*-test, two-tailed *p*-value < 0.05), however for Zeng's *E*, Carratraca-S values are highly significantly smaller than Morata-NS's (paired *t*-test, two-tailed *p*-value $< 10^{-15}$).

It has been suggested that combining Tajima's *D* and Fay and Wu's *H* can be a means towards distinguishing population size expansion and purifying selection (Yang, 2006; Zeng *et al.*, 2006), since *H* is not very sensitive to the former but to the latter (i.e. by the virtue of focusing on high frequency derived alleles). Taking genes from the left 5% tails of Tajima's *D* and Fay & Wu's *H* in each population's statistic distribution, 259, 232 and 179 loci were found to be in bottom 5% values for both statistics in Carratraca-S, Morata-NS and Samil-S, respectively. For overlap between 5% tails of Tajima's *D* and Fu & Li's *D*, 384, 407 and 370 genes were found in Carratraca-S, Morata-NS and Samil-S, respectively. As can be seen, Samil-S had an appreciably lower number of transcripts meeting bottom 5%

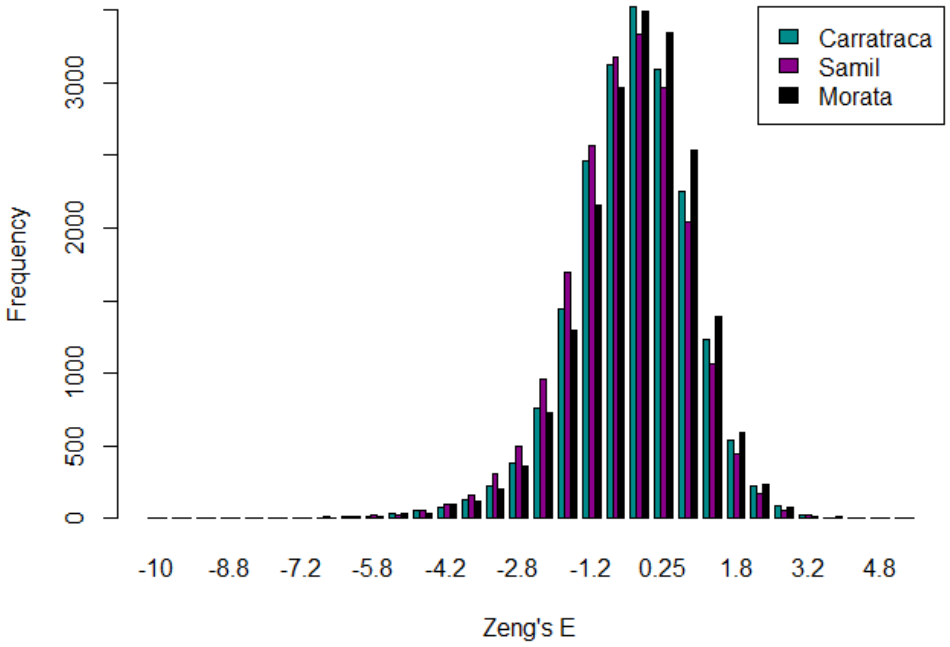


Figure 2.24: Zeng's *E* distribution across all the loci in Carratraca-S, Samil-S and Morata-NS

requirement for Tajima's *D* and Fay & Wu's *H*. The lower number of loci in the bottom 5% tail of Tajima's *D* and Fay & Wu's *H* distribution, combined with significantly lower values of Fu & Li's *D* and Zeng's *E*, statistics focusing on low frequency variants, suggests higher role of population size expansion than positive selection in Samil-S relative to Carratraca-S. No overlap was found between Tajima's *D*-Fay & Wu's *H* or Tajima's *D*-Fu & Li's *D* and Bayescan/fixed outlier transcripts, which is not surprising given that the two approaches look for relatively earlier (F_{ST} outlier) and more recent (Tajima's *D*, Fay & Wu's *H*, Fu & Li's *D*) marks left by directional selection (Quintana-Murci & Clark, 2013).

Table 2.10: Shared outlier genes amongst the two serpentine populations. SNPs present in Carratraca-S

			Polarity status in relation to *Clypeola*					Substitution effect			
TAIR identifier	Gene name	SNP no.	Derived	Ancestral	Missing	Seg.[¶]	Overlap[§]	NS*	S**	UTR	DE?[‡]
Metal transport and homeostasis											
AT5G67330	*NATURAL RESISTANCE ASSOCIATED MACROPHAGE PROTEIN 4 (NRAMP4)*	10	5	3	2	0	10	5	5	0	No
AT5G13740	*ZINC-INDUCED FACILITATOR 1 (ZIF1)*	9	8	0	1	0	1	4	5	0	No
AT2G38460	*IRON-REGULATED PROTEIN 1 (IREG1)*	7	5	0	2	0	3	3	2	2	RT: up
AT4G16370	*OLIGOPEPTIDE TRANSPORTER 3 (OPT3)*	4	2	1	1	0	1	1	3	0	No
AT4G19960	*K+ UPTAKE PERMEASE 9 (KUP9)*	2	0	0	2	0	2	0	2	0	No
AT2G26650	*K+ TRANSPORTER 1 (KT1)*	1	1	0	0	0	0	1	0	0	No
AT3G08040	*FERRIC REDUCTASE DEFECTIVE 3 (FRD3)*	1	0	0	0	1	1	1	0	0	No
AT1G66240	*ANTI-OXIDANT PROTEIN 1 (ATX1)*	1	1	0	0	0	1	1	0	0	No
AT2G38170	*CATION EXCHANGER 1 (CAX1)*	2	1	0	0	0	1	0	1	0	No
AT5G13750	*ZINC-INDUCED FACILITATOR -LIKE 1 (ZIFL1)*	3	2	1	0	0	1	1	2	0	No
Amino acid metabolism, phosphate and nitrate nutrition											
AT1G58080	*ATP PHOSPHORIBOSYL TRANSFERASE 1 (ATP-PRT1)*	2	0	0	1	1	0	2	0	0	SH: up
AT3G03720	*CATIONIC AMINO ACID TRANSPORTER 4 (CAT4)*	10	7	2	1	0	2	7	3	0	SH: up, RT: up
AT5G43360	*PHOSPHATE TRANSPORTER 1;3 (PHT1;3)*	7	0	0	7	0	5	6	1	0	No
Lipid metabolism											
AT3G08510	*PHOSPHOLIPASE C 2 (PLC2)*	3	1	1	1	0	0	2	1	0	No

			Polarity status in relation to *Clypeola*					Substitution effect			
TAIR identifier	Gene name	SNP no.	Derived	Ancestral	Missing	Seg.¶	Overlap§	NS*	S**	UTR	DE?‡
Redox control											
AT2G38180	SGNH hydrolase-type esterase superfamily protein	2	2	0	0	0	0	0	2	0	SH: up
AT1G11530	*C-TERMINAL CYSTEINE RESIDUE IS CHANGED TO A SERINE 1 (CXXS1)*	5	2	2	1	0	3	4	1	0	SH: down
AT3G13080	*ATP-BINDING CASSETTE C3 (ABCC3)*	1	1	0	0	0	0	0	1	0	No
AT2G39480	*ATP-BINDING CASSETTE B6 (ABCB6)*	3	3	0	0	0	1	0	3	0	No
Response to abiotic stressors											
AT2G20990	*SYNAPTOTAGMIN A (SYTA)*	1	1	0	0	0	1	1	0	0	No
AT1G35720	*ANNEXIN 1 (ANN1)*	1	0	1	0	0	1	1	0	0	No
Defense response to pathogens											
AT2G39200	*MILDEW RESISTANCE LOCUS O 12 (MLO12)*	8	6	1	1	0	3	3	5	0	No
Regulation of flowering, leaf and root morphogenesis											
AT5G67100	*INCURVATA 2 (ICU2)*	1	1	0	0	0	1	0	1	0	No
AT2G38440	*SCAR HOMOLOG 2 (SCAR2)*	8	4	1	3	0	6	4	4	0	No
AT3G44750	*HISTONE DEACETYLASE 2A (HDA2A)*	9	5	3	1	0	1	4	5	0	No
Other											
AT1G47330	CBS-DUF21 family	5	2	0	3	0	3	4	1	0	RT: up
AT4G36670	*POLYOL/MONOSACCHARIDE TRANSPORTER 6 (PMT6)*	2	1	1	0	0	2	1	1	0	No

¶: number of polymorphisms segregating in *Clypeola*

§: number of outlier SNPs segregating in the other serpentine population - Samil-S

* : non-synonymous; **: synonymous;

‡: differentially expressed gene in the population relative to non-serpentine populations; SH - shoot; RT- root

Table 2.11: Shared outlier genes amongst the two serpentine populations. SNPs present in Samil-S

			Polarity status in relation to *Clypeola*					Substitution effect			
TAIR identifier	Gene name	SNP no.	Derived	Ancestral	Missing	Seg.[¶]	Overlap[§]	NS*	S**	UTR	DE?[‡]
Metal transport and homeostasis											
AT5G67330	*NATURAL RESISTANCE ASSOCIATED MACROPHAGE PROTEIN 4 (NRAMP4)*	10	5	3	2	0	10	5	5	0	No
AT5G13740	*ZINC-INDUCED FACILITATOR 1 (ZIF1)*	1	1	0	0	0	1	0	1	0	No
AT2G38460	*IRON-REGULATED PROTEIN 1 (IREG1)*	3	1	0	2	0	3	1	0	2	RT: up
AT4G16370	*OLIGOPEPTIDE TRANSPORTER 3 (OPT3)*	1	0	1	0	0	1	0	1	0	No
AT4G19960	*K+ UPTAKE PERMEASE 9 (KUP9)*	3	0	0	3	0	2	1	2	0	No
AT2G26650	*K+ TRANSPORTER 1 (KT1)*	2	0	1	1	0	0	1	1	0	No
AT3G08040	*FERRIC REDUCTASE DEFECTIVE 3 (FRD3)*	1	0	0	0	1	1	1	0	0	No
AT1G66240	*ANTI-OXIDANT PROTEIN 1 (ATX1)*	1	0	0	1	0	1	1	0	0	No
AT2G38170	*CATION EXCHANGER 1 (CAX1)*	2	1	1	0	0	1	1	1	0	No
AT5G13750	*ZINC-INDUCED FACILITATOR -LIKE 1 (ZIFL1)*	1	0	1	0	0	1	1	0	0	No
Amino acid metabolism, phosphate and nitrate nutrition											
AT1G58080	*ATP PHOSPHORIBOSYL TRANSFERASE 1 (ATP-PRT1)*	3	2	0	0	1	3	0	3	0	SH: up
AT3G03720	*CATIONIC AMINO ACID TRANSPORTER 4 (CAT4)*	2	0	0	2	0	2	2	0	0	SH: up
AT5G43360	*PHOSPHATE TRANSPORTER 1;3 (PHT1;3)*	5	0	0	5	0	5	4	1	0	No
Lipid metabolism											
AT3G08510	*PHOSPHOLIPASE C 2 (PLC2)*	1	1	0	0	0	0	1	0	0	No

			Polarity status in relation to *Clypeola*					Substitution effect			
TAIR identifier	Gene name	SNP no.	Derived	Ancestral	Missing	Seg.¶	Overlap§	NS*	S**	UTR	DE?‡
Redox control											
AT2G38180	SGNH hydrolase-type esterase superfamily protein	2	1	1	0	0	2	0	2	0	SH: up
AT1G11530	*C-TERMINAL CYSTEINE RESIDUE IS CHANGED TO A SERINE 1 (CXXS1)*	3	1	1	1	0	3	3	0	0	No
AT3G13080	*ATP-BINDING CASSETTE C3 (ABCC3)*	1	0	0	1	0	1	0	1	0	No
AT2G39480	*ATP-BINDING CASSETTE B6 (ABCB6)*	1	1	0	0	0	1	0	1	0	No
Response to abiotic stressors											
AT2G20990	*SYNAPTOTAGMIN A (SYTA)*	2	2	0	0	0	1	2	0	0	No
AT1G35720	*ANNEXIN 1 (ANN1)*	1	0	1	0	0	1	1	0	0	No
Defense response to pathogens											
AT2G39200	*MILDEW RESISTANCE LOCUS O 12 (MLO12)*	4	2	1	1	0	4	3	1	0	No
Regulation of flowering, leaf and root morphogenesis											
AT5G67100	*INCURVATA 2 (ICU2)*	4	0	2	2	0	1	1	3	0	No
AT2G38440	*SCAR HOMOLOG 2 (SCAR2)*	6	2	1	3	0	6	3	3	0	No
AT3G44750	*HISTONE DEACETYLASE 2A (HDA2A)*	1	1	0	0	0	1	0	1	0	No
Other											
AT1G47330	CBS-DUF21 family	4	2	0	2	0	4	3	1	0	RT: up
AT4G36670	*POLYOL/MONOSACCHARIDE TRANSPORTER 6 (PMT6)*	2	1	1	0	0	2	1	1	0	No

¶: number of polymorphisms segregating in *Clypeola*

§: number of outlier SNPs present in the other serpentine populations - Carratraca-S

* : non-synonymous; **: synonymous;

‡: - differentially expressed gene in the population relative to non-serpentine populations; SH - shoot; RT- root

Table 2.12: Population-specific outlier genes: Carratraca-S

			Polarity status in relation to *Clypeola*				Substitution effect				
TAIR identifier	Gene name	SNP no.	Derived	Ancestral	Missing	Seg.[¶]	NS*	S**	UTR	Present in S?[§]	DE?[‡]
Metal transport and homeostasis											
AT2G19110	*HEAVY METAL ATPASE 4 (HMA4)*	1	1	0	0	0	0	1	0	0	No
AT1G16010	*MAGNESIUM TRANSPORTER 2 (MGT2)*	2	1	0	1	0	1	1	0	0	No
AT4G19690	*IRON-REGULATED TRANSPORTER 1 (IRT1)*	10	0	0	10	0	2	1	7	10	No
AT5G24380	*YELLOW STRIPE LIKE 2 (YSL2)*	1	1	0	0	0	0	1	0	0	No
AT1G07610	*METALLOTHIONEIN 1C (MT1C)*	1	0	0	0	1	1	0	0	1	SH: down
Amino acid metabolism, phosphate and nitrate nutrition											
AT2G33770	*PHOSPHATE 2 (PHO2)*	1	1	0	0	0	1	0	0	0	No
AT2G01830	*WOODEN LEG (WOL)*	4	1	2	1	0	2	2	0	0	No
AT5G20410	*MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE 2 (MGD2)*	1	1	0	0	0	0	1	0	1	No
Cell wall biosynthesis											
AT3G16920	*CHITINASE-LIKE PROTEIN 2 (CTL2)*	1	1	0	0	0	1	0	0	1	No
AT5G13000	*GLUCAN SYNTHASE-LIKE 12 (GSL12)*	26	13	11	2	0	6	20	0	0	No
AT2G22125	*CELLULOSE SYNTHASE-INTERACTIVE PROTEIN 1 (CSI1)*	3	1	1	1	0	0	3	0	1	No
AT1G05850	*CHITINASE-LIKE PROTEIN 1 (CTL1)*	1	1	0	0	0	0	0	1	0	No
AT1G52760	*CAFFEOYL SHIKIMATE ESTERASE (CSE)*	1	1	0	0	0	1	0	0	1	No
AT3G58100	*PLASMODESMATA CALLOSE -BINDING PROTEIN 5 (PDCB5)*	2	2	0	0	0	1	1	0	1	No

			Polarity status in relation to *Clypeola*				Substitution effect				
TAIR identifier	Gene name	SNP no.	Derived	Ancestral	Missing	Seg.¶	NS*	S**	UTR	Present in S?§	DE?‡
Lipid metabolism											
AT3G07020	*STEROL GLUCOSYLTRANSFERASE (SGT)*	2	1	0	1	0	1	0	1	1	No
AT2G38040	*ACETYL COA CARBOXYLASE CARBO-ZYLTRANSFERASE ALPHA (CAC3)*	1	1	0	0	0	0	1	0	1	No
AT4G16760	*ACYL-COA OXIDASE 1 (ACX1)*	1	1	0	0	0	0	1	0	0	No
Redox control											
AT4G03560	*TWO-PORE CHANNEL 1 (TPC1)*	4	1	1	2	0	3	1	0	4	No
AT4G35310	*CALMODULIN-DOMAIN PROTEIN KINASE 5 (CPK5)*	1	1	0	0	0	1	0	0	0	No
AT5G06690	*WCRKC THIOREDOXIN 1 (WCRKC1)*	1	1	0	0	0	0	1	0	1	No
AT5G23310	*FE SUPEROXIDE DISMUTASE 3 (FSD3)*	1	1	0	0	0	1	0	0	0	No
AT4G33000	*CALCINEURIN B-LIKE PROTEIN 10 (CBL10)*	1	1	0	0	0	0	1	0	0	No
AT4G29210	*GAMMA-GLUTAMYL TRANSPEPTIDASE 4 (GGT4)*	2	2	0	0	0	1	1	0	2	No
Response to abiotic stressors											
AT5G22080	*Chaperone DnaJ-domain superfamily protein*	2	1	0	1	0	0	0	2	2	No
AT2G01980	*SALT OVERLY SENSITIVE 1 (SOS1)*	2	2	0	0	0	1	1	0	0	No
AT1G62970	*Chaperone DnaJ-domain superfamily protein*	1	0	1	0	0	0	0	1	0	No
AT2G31470	*DROUGHT TOLERANCE REPRESSOR (DOR)*	2	0	0	2	0	1	0	0	0	No
AT4G16660	*heat shock protein 70 (Hsp 70) family protein*	1	1	0	0	0	0	1	0	0	No

			Polarity status in relation to *Clypeola*				Substitution effect				
TAIR identifier	Gene name	SNP no.	Derived	Ancestral	Missing	Seg.¶	NS*	S**	UTR	Present in S?§	DE?‡
Defense response to pathogens, herbivores											
AT5G23020	*2-ISOPROPYLMALATE SYNTHASE 2 (IMS2)*	2	1	1	0	0	0	2	0	0	No
AT2G46400	*WRKY DNA-BINDING PROTEIN 46 (WRKY46)*	16	11	2	2	1	8	8	0	3	No
AT2G37710	*RECEPTOR LECTIN KINASE (RLK)*	5	3	2	0	0	3	2	0	4	No
AT1G12280	*SUPPRESSOR OF MKK1 MKK2 2 (SUMM2)*	1	0	0	0	1	1	0	0	1	No
AT3G10985	*SENESCENCE ASSOCIATED GENE 20 (SAG20)*	1	1	0	0	0	0	0	1	1	No
AT3G04720	*PATHOGENESIS-RELATED 4 (PR4)*	1	1	0	0	0	0	1	0	1	No
Regulation of flowering, leaf and root morphogenesis											
AT1G30330	*AUXIN RESPONSE FACTOR 6 (ARF6)*	5	3	0	2	0	0	1	4	1	No
AT3G12980	*HISTONE ACETYLTRANSFERASE OF THE CBP FAMILY 5 (HAC5)*	5	3	0	2	0	3	1	1	0	No
AT2G22540	*SHORT VEGETATIVE PHASE (SVP)*	4	1	2	1	0	3	1	0	0	No
AT4G01710	*CROOKED (CRK)*	4	1	0	2	1	1	2	1	3	No
AT4G02020	*SWINGER (SWN)*	4	4	0	0	0	3	1	0	0	No
AT1G79000	*HISTONE ACETYLTRANSFERASE OF THE CBP FAMILY 1 (HAC1)*	2	2	0	0	0	2	0	0	0	No
AT1G05830	*TRITHORAX-LIKE PROTEIN 2 (ATX2)*	1	1	0	0	0	0	1	0	1	No
Other											
AT1G51740	*SYNTAXIN OF PLANTS 81 (SYP81)*	1	1	0	0	0	1	0	0	0	No

¶: number of polymorphisms segregating in *Clypeola*

* : non-synonymous; **: synonymous;

§: number of outlier SNPs segregating in the other serpentine population- Samil-S

‡: differentially expressed gene in the population relative to non-serpentine populations; SH - shoot; RT- root

Table 2.13: Population-specific outlier genes: Samil-S

			Polarity status in relation to *Clypeola*				Substitution effect				
TAIR identifier	Gene name	SNP no.	Derived	Ancestral	Missing	Seg.[¶]	NS*	S**	UTR	Present in C?[§]	DE?[‡]
Metal transport and homeostasis											
AT1G05300	*ZINC TRANSPORTER 5 PRECURSOR (ZIP5)*	2	2	0	0	0	1	1	0	2	No
AT5G09400	*K+ UPTAKE PERMEASE 7 (KUP7)*	2	1	0	1	0	1	0	1	1	No
AT4G33520	*HEAVY METAL ATPASE 6 (HMA6)*	1	0	0	0	1	0	1	0	1	No
AT5G09690	*MAGNESIUM TRANSPORTER 7 (MGT7)*	1	0	1	0	0	0	1	0	1	No
AT1G10130	*ENDOPLASMIC RETICULUM-TYPE CALCIUM-TRANSPORTING ATPASE 3 (ECA3)*	1	1	0	0	0	0	0	1	0	No
AT2G47160	*REQUIRES HIGH BORON 1 (BOR1)*	1	1	0	0	0	1	0	0	0	No
AT4G10380	*NOD26-LIKE INTRINSIC PROTEIN 5;1 (NIP5;1)*	1	1	0	0	0	0	0	1	1	No
AT3G18440	*ALUMINUM-ACTIVATED MALATE TRANSPORTER 9 (ALMT9)*	2	1	0	1	0	2	0	0	0	No
Amino acid metabolism, phosphate and nitrate nutrition											
AT1G31860	*HISTIDINE BIOSYNTHESIS 2 (HISN2)*	5	1	2	1	1	2	3	0	0	No
AT1G09795	*ATP PHOSPHORIBOSYL TRANSFERASE 2 (ATP-PRT2)*	1	1	0	0	0	0	1	0	1	No
AT4G33680	*ABERRANT GROWTH AND DEATH 2 (AGD2)*	1	1	0	0	0	0	1	0	1	No
AT1G76430	*PHOSPHATE TRANSPORTER 1;9 (PHT1;9)*	2	2	0	0	0	2	0	0	0	No
Cell wall biosynthesis											
AT4G04970	*GLUCAN SYNTHASE-LIKE 1 (GSL1)*	13	11	0	2	0	3	10	0	8	No
AT4G36220	*FERULIC ACID 5-HYDROXYLASE 1 (FAH1)*	1	1	0	0	0	0	0	1	0	No
AT1G05250	*PEROXIDASE 2 (PRX2)*	16	0	0	16	0	10	5	1	16	No

			Polarity status in relation to *Clypeola*				Substitution effect				
TAIR identifier	Gene name	SNP no.	Derived	Ancestral	Missing	Seg.[¶]	NS*	S**	UTR	Present in C?[§]	DE?[‡]
Lipid metabolism											
AT3G15730	*PHOSPHOLIPASE D ALPHA 1 (PLDALPHA1)*	4	3	1	0	0	3	1	0	0	No
AT5G57800	*ECERIFERUM 3 (CER3)*	3	3	0	0	0	1	2	0	0	No
AT2G31360	*16:0DELTA9 DESATURASE 2 (ADS2)*	1	1	0	0	0	1	0	0	1	No
Redox control											
AT5G61420	*HIGH ALIPHATIC GLUCOSINOLATE 1 (HAG1)*	1	1	0	0	0	1	0	0	0	No
AT3G10572	*ABERRANT PEROXISOME MORPHOLOGY 9 (APEM9)*	7	1	3	2	1	3	4	0	7	No
AT4G25100	*FE SUPEROXIDE DISMUTASE 1 (FSD1)*	1	1	0	0	0	0	1	0	0	No
Response to abiotic stressors											
AT5G67320	*HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 15 (HOS15)*	6	3	3	0	0	1	4	1	6	No
AT1G14000	*VH1-INTERACTING KINASE (VIK)*	3	1	1	1	0	2	1	0	3	No
AT3G07770	*HEAT SHOCK PROTEIN 89.1*	1	1	0	0	0	1	0	0	0	No
AT1G10350	*DNAJ heat shock family*	21	10	9	2	0	6	12	3	5	No
AT4G39150	*DNAJ heat shock family*	4	3	0	1	0	3	1	0	1	No
AT1G06040	*SALT TOLERANCE (STO)*	4	3	1	0	0	1	3	0	0	No
AT5G35410	*SALT OVERLY SENSITIVE 2 (SOS2)*	1	1	0	0	0	1	0	0	0	No
AT3G06370	*SODIUM HYDROGEN EXCHANGER 4 (NHX4)*	6	6	0	0	0	3	3	0	1	No
AT3G12360	*INCREASED TOLERANCE TO NACL (ITN1)*	3	1	1	1	0	0	1	2	2	No
AT3G08760	*ATSIK*	2	1	1	0	0	1	1	0	2	No
Defense response to pathogens, herbivores											
AT1G11310	*MILDEW RESISTANCE LOCUS O 2 (MLO2)*	11	5	6	0	0	6	5	0	11	No
AT1G75040	*PATHOGENESIS-RELATED GENE 5*	3	1	0	1	1	2	1	0	0	No

			Polarity status in relation to *Clypeola*				Substitution effect				
TAIR identifier	Gene name	SNP no.	Derived	Ancestral	Missing	Seg.¶	NS*	S**	UTR	Present in C?§	DE?‡
AT1G28380	*NECROTIC SPOTTED LESIONS 1 (NSL1)*	3	2	1	0	0	0	3	0	0	No
AT1G10170	*NF-X-LIKE 1 (NFXL1)*	3	2	1	0	0	1	2	0	1	No
AT4G11280	*1-AMINOCYCLOPROPANE-1-CARBOXYLIC ACID SYNTHASE 6 (ACS6)*	3	1	1	1	0	0	2	1	0	No
AT5G05680	*MODIFIER OF SNC1,7 (MOS7)*	1	1	0	0	0	1	0	0	0	No
AT4G23190	*CYSTEINE-RICH RECEPTOR-LIKE PROTEIN KINASE 11 (CRK11)*	1	1	0	0	0	0	1	0	0	No
AT1G31280	*ARGONAUTE 2 (AGO2)*	1	1	0	0	0	1	0	0	0	No
AT3G11840	*PLANT U-BOX 24 (PUB24)*	1	1	0	0	0	0	1	0	0	No
Regulation of flowering, leaf and root morphogenesis											
AT3G07650	*CONSTANS-LIKE 9 (COL9)*	12	5	6	1	0	1	10	1	0	No
AT3G01460	*METHYL-CPG-BINDING DOMAIN 9 (MBD9)*	5	5	0	0	0	3	2	0	0	No
AT5G23150	*ENHANCER OF AG-4 2 (HUA2)*	3	1	1	1	0	1	1	1	3	No
AT4G32551	*LEUNIG (LUG)*	3	2	0	1	0	0	0	3	0	No
AT5G20490	*MYOSIN XIK*	3	3	0	0	0	2	1	0	1	No
AT3G12400	*ELC*	1	1	0	0	0	1	0	0	0	No
Other											
AT1G28490	*SYNTAXIN OF PLANTS 61 (SYP61)*	6	4	1	1	0	1	5	0	6	No
AT1G16240	*SYNTAXIN OF PLANTS 51 (SYP51)*	1	1	0	0	0	0	1	0	0	No
AT4G02195	*SYNTAXIN OF PLANTS 42 (SYP42)*	1	1	0	0	0	1	0	0	0	No
AT4G33700	*CBS-DUF21 family*	6	4	0	1	1	3	3	0	0	No
AT4G14240	*CBS-DUF21 family*	1	1	0	0	0	1	0	0	0	No
AT4G14230	*CBS-DUF21 family*	1	1	0	0	0	0	0	1	1	SH & RT: up

¶: number of polymorphisms segregating in *Clypeola*

* : non-synonymous; **: synonymous;

§: number of outlier SNPs segregating in the other serpentine population - Carratraca-S

‡: differentially expressed gene in the population relative to non-serpentine populations; SH - shoot; RT- root

2.3.8 Identification of *Alyssum* outlier genes of further research interest

Highly diverged loci are likely to have been under the influence of disruptive selection (Beaumont & Nichols, 1996) and the aim here was to find them by comparing individual serpentine populations to the two combined non-serpentine populations. It has to be kept in mind that some of the differences seen will be just strictly due to drift and a combination of demographic and alternative selective forces (e.g. background selection in subdivided populations; Lotterhos & Whitlock, 2014), and accrued knowledge of functional importance of some of the putative selection targets in, e.g. metal tolerance, helps to restrict the analysis to the outliers more likely to be true positives. Furthermore, focusing only on loci showing non-synonymous replacements and belonging to overrepresented categories should provide additional evidence to put the candidate gene forward as a true positive (De Wit *et al.*, 2015). This section provides a selection of Bayescan outlier genes, containing derived alleles (relative to *Clypeola*) in serpentine populations and thus likely under directional selection. The genes are grouped under labels representing highly significantly overrepresented gene ontology categories amongst the outlier genes and further grouped under subheadings: Shared, i.e. candidate outlier genes detected both in Carratraca-S (Table 2.10) and Samil-S (Table 2.11), and Population-specific - candidate outlier genes detected only in a single serpentine population - Carratraca-S (Table 2.12) or Samil-S (Table 2.13).

2.3.8.1 Metal transport and homeostasis

Shared Among the genes in this category, a number of transporter genes previously implicated in metal tolerance and hyperaccumulation were found. Shared Carratraca-S and Samil-S targets included: *NATURAL RESISTANCE ASSOCIATED MACROPHAGE PROTEIN 4 (NRAMP4), ZINC-INDUCED FACILITATOR 1 (ZIF1), IRON-REGULATED PROTEIN 1 (IREG1), OLIGOPEPTIDE TRANSPORTER 3 (OPT3), K+ UPTAKE PERMEASE 9 (KUP9), K+ TRANSPORTER 1 (KT1), FERRIC REDUCTASE DEFECTIVE 3 (FRD3), ANTI-OXIDANT PROTEIN 1 (ATX1)* and *CATION EXCHANGER 1 (CAX1)*.

Members of *NRAMP* family of metal transporters have been shown to be strongly expressed in various hyperaccumulating taxa. In particular, *NRAMP4* has been shown to be both constitutively more highly expressed and exhibit inducible expression on cadmium exposure in the zinc hyperaccumulator ecotype when compared to non-accumulating ecotype in *Sedum alfredii* in the Crassulaceae family (Gao *et al.*, 2013), and its expression to correlate with nickel tolerance in *Noccaea caerulescens* (Halimaa *et al.*, 2014; Visioli *et al.*, 2014) as well as show significantly elevated expression in *N. caerulescens* versus *Arabidopsis thaliana* both in the shoots and in the roots (Oomen *et al.*, 2009). The last study also revealed that the *N. caerulescens* homolog displays the same general functional properties as that of *A. thaliana* – namely is capable of Zn, Fe, Mn and Cd transport and is localised to the tonoplast (Thomine *et al.*, 2000, 2003; Lanquar *et al.*, 2005, 2010; Pottier *et al.*, 2015). Furthermore, *NRAMP4* overexpression in *Arabidopsis thaliana* promotes Zn and Cd remobilisation from root vacuoles and decreased metal accumulation in the root (Pottier *et al.*, 2015). What is particularly interesting given the number of outlier SNPs identified in the gene coding region in serpentine populations of *Alyssum serpyllifolium*, is that NRAMP4's ability to transport different metals is strongly affected by point mutations in *Arabidopsis thaliana*, as is typical of other members of this gene family (Pottier *et al.*, 2015). This invites the hypothesis of potential adaptive relevance of replacement substitutions in the serpentine populations in altering the transporter metal specificity in the context of Ni hyperaccumulation.

Another vacuolar transporter, *ZIF1* has been shown to be important for iron and zinc homeostasis in *Arabidopsis* (Haydon *et al.*, 2012). It is capable of changing the root and shoot partitioning of metal chelator nicotianamine and in this way influencing zinc accumulation and tolerance as well as iron deficiency response. *ZIF1* overexpressor line contains increased levels of nicotianamine in the root vacuoles which correlates with the increase of zinc in the root vacuoles and at the same time produces Fe deficiency response in the leaves. Another gene in the family, *ZIF2*, has been recently shown to be targeted to the root tonoplast and contribute to zinc tolerance (Remy *et al.*, 2014). *ZIF1* along with *ZINC-INDUCED FACILITATOR-LIKE 1 (ZIFL1)*, another shared target of selection, belong to a large family of poorly characterised Major Facilitator Superfamily (MFS) transporters with

128 members in the *Arabidopsis* genome (Remy *et al.*, 2013). Two different isoforms are produced from the At*ZIFL1* locus: full length *ZIFL1.1* and truncated *ZIFL1.3*, and they both show differential expression patterns and functions, however, both can transport potassium when expressed in yeast (Remy *et al.*, 2013). *ZIFL1.1* is targeted to the tonoplast of root cells and influences auxin efflux from the root tip towards the shoot, whereas *ZIFL1.2* sits at the plasma membrane of leaf stomatal guard cells and is involved in regulation of drought stress tolerance by modulating stomatal closure (Remy *et al.*, 2013). Another member of the family and paralog of *ZIFL1*, *ZIFL2* has been linked to cadmium/zinc tolerance in *N. caerulescens* due to its expression pattern across different *N. caerulescens* ecotypes (Halimaa *et al.*, 2014) as well as was revealed to be a target of local adaptation to serpentine in *A. lyrata* (Turner *et al.*, 2010).

OPT3 is an oligopeptide transporter upregulated in *Arabidopsis thaliana* in response to excess zinc (van de Mortel *et al.*, 2006). Conversely, an *OPT3* homolog from *N. caerulescens* (Hu *et al.*, 2012**b**) has been shown to be upregulated in response to zinc and iron deficiency. It is a plasma-membrane localised transporter and is expressed primarily in the shoot, with strong expression in the pericycle in particular, and can function as a heavy metal transporter with broad substrate specificity (Fe/Zn/Cd/Cu) in yeast. The role of *OPT3* in systemic iron signalling in *Arabidopsis* between leaves and roots has been firmly established now (Mendoza-Cózatl *et al.*, 2014; Zhai *et al.*, 2014) - it mediates iron phloem loading as well as iron redistribution from mature to developing tissues.
Another gene involved in iron homeostasis, *FRD3* has been previously found to be constitutively expressed in *Noccaea caerulescens* (van de Mortel *et al.*, 2006) and in *Arabidopsis halleri* (Talke *et al.*, 2006) at a much higher level than in *A. thaliana*. *FRD3* is responsible for the efflux of metal chelator citrate into root vasculature which contributes to iron root-to-shoot translocation (Durrett *et al.*, 2007) as well as to zinc transport and tolerance (Pineau *et al.*, 2012).

Two potassium transporters were identified as candidate loci in both serpentine populations: *KUP9* and *KT1*. *KUP9* locus was previously identified amongst targets of local

adaptation to serpentine in *A. lyrata* (Turner *et al.*, 2010). *KT1* encodes an inward rectifying channel predominantly expressed in root hairs and endodermis in *Arabidopsis* which functions as one of the chief routes of potassium entry at intermediate to high (>10 μM) concentration in the medium (Rubio *et al.*, 2010; Alemán *et al.*, 2011; Nieves-Cordones *et al.*, 2014). In addition, another KUP family member, *KUP7* is an outlier candidate with polymorphism specific to Samil-S.

Lastly, one copper chaperone - *ATX1* and tonoplast calcium antiporter - *CAX1* were identified in both serpentine populations. *ATX1* is crucial for copper homeostasis under both copper deficiency and excess (Shin *et al.*, 2012) based on the knockout and overexpressor line phenotypes and interacts with copper transporters *HMA5* (Andrés-Colás *et al.*, 2006) and *HMA7* (Puig *et al.*, 2007). *CAX1* has been initially linked to serpentine syndrome in an *Arabidopsis* mutant screen conducted on media with low Ca:Mg ratio which is characteristic of that type of soils (Bradshaw, 2005). A mutant identified in the screen required low Ca:Mg ratio for normal growth. In the field, Turner *et al.* (2010) identified a polymorphism situated between *CAX7* and *CAX8* paralogs specific to serpentine *A. lyrata*, suggestive of importance of CAX family members to serpentine plants. Recently, *CAX1* has been unmasked as one of the genes behind a major QTL for cadmium tolerance in *A. halleri* conditional on low Ca environment (Baliardini *et al.*, 2015).

Lastly, one gene predicted to function in metal transport but with no established role was identified: AT5G27690 (heavy metal transport/detoxification superfamily).

Population-specific Two different members of the Heavy Metal ATPase family are candidate targets of selection in the two serpentine populations: *HEAVY METAL ATPASE 4* (*HMA4*) in Carratraca-S and *HEAVY METAL ATPASE 6* (*HMA6*) in Samil-S. *HMA4* colocalizes with a major QTL for cadmium tolerance in the zinc and cadmium hyperaccumulator *Arabidopsis halleri* and shows elevated expression in comparison to *A. thaliana* (Becher *et al.*, 2004; Weber *et al.*, 2004) and non-serpentine *A. lyrata* ssp. *petraea* in the root and in the shoot (Courbot *et al.*, 2007). It is necessary for zinc and cadmium hyperaccumulation, as it mediates root to shoot flux of zinc in Brassicaceae family members *A. thaliana* (Hussain *et al.*, 2004; Mills *et al.*, 2005), *A. halleri* (Talke *et al.*, 2006; Hanikenne

et al., 2008) and *N. caerulescens* (Ó Lochlainn *et al.*, 2011; Craciun *et al.*, 2012), likely by facilitating xylem loading of the metal, as it is a plasma membrane transporter located in root pericycle and xylem parenchyma (Hanikenne *et al.*, 2008). It is also constitutively upregulated in cadmium hyperaccumulating ecotype of *Sedum alfredii* relative to the non-accumulating ecotype (Gao *et al.*, 2013).
HMA6's possible role is hyperaccumulation is less clear – so far it has only been shown to be 4.6-fold upregulated in *A. halleri* shoots in comparison to *A. thaliana* (Talke *et al.*, 2006). In *Arabidopsis*, HMA6 functions as a copper transporter localised to the chloroplast envelope and responsible for copper co-factor delivery to the chloroplast (Shikanai *et al.*, 2003; Boutigny *et al.*, 2014).

Similarly, two different magnesium transporters belonging to the same family were found in each of the two *Alyssum* serpentine populations: high affinity MAGNESIUM TRANSPORTER 2 (*MGT2*, Gebert *et al.*, 2009) in Carratraca-S and low affinity MAGNESIUM TRANSPORTER 7 (*MGT7*, Mao *et al.*, 2008) in Samil-S – likely false positive, as it contains the ancestral allele. *MGT2* homolog shows constitutively high and also Cd-inducible expression in cadmium accumulator *Solanum nigrum* in comparison with the non-accumulator *Solanum torvum* (Xu *et al.*, 2012), while polymorphisms in *MGT7* homolog (as well as *MGT5* and *MGT9*) have been associated with adaptation to serpentine in *A. lyrata* (Turner *et al.*, 2010). Interestingly, *MGT2* has been implicated in tolerance of serpentine low Ca:Mg ratios in *A. thaliana*. At*MGT2* was found to be targeted to tonoplast (Conn *et al.*, 2011), and is chiefly expressed in the mesophyll vacuoles, where magnesium is preferentially deposited. Knockout *mgt2* lines have lower Mg content under serpentine conditions and *MGT* gene expression levels in various ecotypes correlate with the degree of Mg accumulation under serpentine-like Ca:Mg ratios (Conn *et al.*, 2011).

Two more transporters with possible links to hyperaccumulator and serpentine adaptation phenotype were found in each population – *IRON-REGULATED TRANSPORTER 1 (IRT1)* and *YELLOW STRIPE LIKE 2 (YSL2)* in Carratraca-S, *ZINC TRANSPORTER 5 PRECURSOR (ZIP5)* and *K+ UPTAKE PERMEASE 7 (KUP7)* in Samil-S (mentioned

earlier).

IRT1 transporter serves as the chief route of entry of iron into the root cells in eudicots (Eide *et al.*, 1996; Varotto *et al.*, 2002; Vert *et al.*, 2002), however, it is also capable of transporting a variety of other metal ions - zinc, manganese, cobalt, cadmium and crucially nickel (Korshunova *et al.*, 1999; Rogers *et al.*, 2000; Nishida *et al.*, 2011). Differential expression of two *IRT1* homologs in different *Noccaea caerulescens* ecotypes likely contributes to their difference in cadmium (Plaza *et al.*, 2007) and nickel accumulating ability (Halimaa *et al.*, 2014).

YSL2 is a plasma membrane transporter of nicotianamine-Fe^{2+} complexes that is primarily expressed in xylem parenchyma in leaves (DiDonato *et al.*, 2004; Chu *et al.*, 2010). In *N. caerulescens*, a zinc hyperaccumulation QTL has been detected in *YSL2*'s region on the chromosome (Assunção *et al.*, 2006). More is known about metal homeostasis roles played by the other closely related YSL family members. Plasma membrane transporter *YSL1* shows functional redundancy with *YSL3* and both are Fe, Cu, Zn transporters with roles in mobilising iron and other metals from the senescent leaves into developing inflorescence and seeds; thus they are required for proper development of those structures (Waters *et al.*, 2006; Chu *et al.*, 2010). Recently, *YSL3* has been implicated in defence response: it is activated by *NONEXPRESSOR OF PR1 (NPR1)* on exposure to pathogens in salicylic acid-dependent pathway, and *ysl3* mutant is more susceptible to infection than wild type (Chen *et al.*, 2014**a**).

ZIP5 belongs to the same family of metal transporters as *IRT1*. Many ZIP genes have been implicated in the hyperaccumulator phenotype in *N. caerulescens* and *A. halleri* (Halimaa *et al.*, 2014) but *ZIP5*'s function is yet to be established. *ZIP5* has been found to be upregulated in response to zinc deficiency in *Brassica rapa* (Li *et al.*, 2014**a**) and wild-type *Arabidopsis* (van de Mortel *et al.*, 2006) as well as under the scenario of zinc deficiency driven by increased deposition of zinc to the shoot vacuoles in transgenic plants (Gustin *et al.*, 2009). *OsZIP5* has been shown to be a plasma membrane-localised zinc transporter capable of influencing overall zinc distribution in the plant (Lee *et al.*, 2010).

Other metal transport-related genes specific to Samil-S include: *ENDOPLASMIC RETICULUM-TYPE CALCIUM-TRANSPORTING ATPASE 3 (ECA3)*, *REQUIRES HIGH BORON 1 (BOR1)* and *NOD26-LIKE INTRINSIC PROTEIN 5;1 (NIP5;1)*. *ECA3* is a Ca^{2+}/Mn^{2+} pump localized to post-Golgi compartments which is important for manganese detoxification under Mn excess (Li *et al.*, 2008**b**) but also required under Mn deficiency conditions (Mills *et al.*, 2008). *BOR1* and *NIP5;1* are two transporters necessary for borate uptake. *NIP5;1* is the main plasma membrane borate channel which belongs to the aquaporin family, is expressed in the root cap and epidermal cells and possesses outside facing polarity for borate import (Takano *et al.*, 2006; Tanaka *et al.*, 2011). It is upregulated under boron deficiency conditions and *nip5;1* mutant/*NIP5;1* overexpressor lines show decreased/increased tolerance to low boron, respectively (Takano *et al.*, 2006; Kato *et al.*, 2009). Interestingly, *NIP5;1* also showed very high genetic differentiation in serpentine and granitic accessions of *A. lyrata*, indicating a possible role in the adaptation to serpentine soil (Turner *et al.*, 2008).
BOR1 is a root plasma membrane-localised transporter with inwards-facing polarity responsible for export of borate towards the stele for xylem loading under boron deficiency conditions (Takano *et al.*, 2002, 2010).

Lastly, two genes important for malate transport and response to aluminium ion were detected in Samil-S: *ALUMINUM-ACTIVATED MALATE TRANSPORTER 9 (ALMT9)* and an aluminium activated malate transporter family protein (AT1G08440). *At*ALMT9 is targeted to the vacuole in the mesophyll cells where it mediates malate (also tartrate in *V. vinifera*; De Angeli *et al.*, 2013) fluxes into the vacuole (Kovermann *et al.*, 2007). In guard cells, it has a role as a malate-activated chloride channel needed for stomatal opening (De Angeli *et al.*, 2013). Gene expression of another member of the family, *ALMT12* has been shown to correlate with nickel tolerance in *N. caerulescens* accessions (Halimaa *et al.*, 2014). Malate is one of the chief Ni chelators in serpentine species of *Alyssum* (Brooks *et al.*, 1981) and induction of another ALTM transporter (*ALMT1*) at a low Ni concentration correlates with higher tolerance and accumulation of nickel in the shoot in serpentine *Arabidopsis thaliana* ecotypes (Agrawal *et al.*, 2012).

In Carratraca-S, *METALLOTHIONEIN 1C (MT1C)* was also found. *MTIC* encodes one of the three metallothioneins (small peptides binding metals and important for metal tolerance) of class I; the *Arabidopsis* genome contains 5 more metallothionein genes belonging to 3 other classes (Zimeri *et al.*, 2005). All members of class I in *Arabidopsis* are involved in Cd tolerance as well as As and Zn accumulation based on the phenotype of their knockdown lines (Zimeri *et al.*, 2005). Metallothioneins of different classes, in particular II and III have also been shown to be implicated in metal tolerance, particularly of Cd, Cu and Zn (Hassinen *et al.*, 2009; Roosens *et al.*, 2004; Visioli *et al.*, 2014). In *A. halleri*, *MT2A, MT2B, MT3* are all upregulated in comparison with *A. thaliana* (Chiang *et al.*, 2006), while the three metallothionein genes along with MT1A are constitutively upregulated in metallicolous *N. caerulescens* relative to non-metallicolous *N. caerulescens* accessions and *A. thaliana* (Roosens *et al.*, 2004; van de Mortel *et al.*, 2006; Hassinen *et al.*, 2009; Halimaa *et al.*, 2014). In addition, *MT2* from *Sedum alfredii* concomitantly increases accumulation and tolerance to Cd and Zn when heterologously expressed in yeast and tobacco (Zhang *et al.*, 2014**b**). *MT1A* and *MT1C*'s role in metal hyperaccumulating plants further suggests that the high F_{ST} window in the region harbouring these genes in the comparison between the serpentine and non-serpentine population of *A. lyrata* could be a result of directional selection (Turner *et al.*, 2010).

2.3.8.2 Amino acid metabolism, phosphate and nitrate nutrition

Shared The key enzyme involved in the first and second step in the histidine biosynthesis pathway and constitutively highly expressed in the *Alyssum* hyperaccumulators, including *Alyssum pintodasilvae* (Samil-S) - *ATP-PRT 1* (Ingle *et al.*, 2005**a**; Rees *et al.*, 2009), was found to be a shared candidate gene in the two serpentine populations. Also, a gene encoding the enzyme's another isoform, *ATP-PRT2*, was found to be an outlier only in Samil-S. Histidine has been demonstrated to play an important role both in nickel chelation in the root as well as in xylem loading for transfer towards the shoot in *Alyssum* (Krämer *et al.*, 1996; Kerkeb & Krämer, 2003). Furthermore, in *N. caerulescens*, histidine was shown to inhibit vacuolar Ni sequestration in the root, as evidenced by blocked uptake of Ni:histidine complex into isolated tonoplast vesicles (Richau *et al.*, 2009).

The other shared gene in the category was *PHOSPHATE TRANSPORTER 1;3 (PHT1;3)*, a phosphate transporter belonging to a family of nine *Arabidopsis* Pi transporters with homology to yeast high-affinity phosphate transporter PHO84 (Mudge *et al.*, 2002). It has been shown to be constitutively upregulated in *N. caerulescens* roots in comparison with *Arabidopsis thaliana* (van de Mortel *et al.*, 2006). In Samil-S, another phosphate transporter in the family was additionally found – *PHT1;9* which functions as a high affinity Pi/H^+ symporter important for root-to-shoot inorganic phosphate translocation and response to Pi supply, as well as mediates arsenate uptake from the soil (Remy *et al.*, 2012; Lapis-Gaza *et al.*, 2014). The gene is upregulated in metallicolous *N. caerulescens* and its expression is associated with cadmium hyperaccumulation and tolerance (Halimaa *et al.*, 2014).

Samil-S *HISTIDINE BIOSYNTHESIS 2 (HISN2)* is one of the enzymes in the histidine biosynthesis pathway and possesses dual phosphoribosyl-ATP pyrophosphohydrolase and phosphoribosyl-AMP cyclohydrolase activity responsible for catalysing the second and third step following the ATP-PRT-catalysed first step (Fujimori & Ohta, 1998).
ABERRANT GROWTH AND DEATH 2 (AGD2) encodes a chloroplast-localised L,L-diaminopimelate aminotransferase which converts tetrahydrodipicolinate to LL-DAP in the lysine biosynthesis pathway in plants (Hudson *et al.*, 2006). Its dwarf mutant shows constitutive resistance to *Pseudomonas syringae* and high salicylic acid content; knockout of the gene results in inviable embryo (Rate & Greenberg, 2001; Song *et al.*, 2004). Based on that and other evidence, it has been speculated that AGD2 is involved in the synthesis of amino-acid derived compound which can suppress defence response and promote development (Song *et al.*, 2004).

Carratraca-S In Carratraca-S, one out of three enzyme isoforms important for adaptation to low Pi conditions was found - *MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE 2 (MGD2)*. Along with *MGD3*, it is expressed under low phosphate nutrition to drive the galactolipid digalactosyl diacyl glycerol (DGDG) biosynthesis, which substitutes

phospholipids in plant membranes under Pi-deprived conditions (Dörmann & Benning, 2002; Kobayashi *et al.*, 2006, 2009).

PHOSPHATE 2 (PHO2) is a ubiquitin E2-conjugase whose activity is essential for maintaining Pi nutrition homeostasis by tightly regulating degradation of *PHO1*, a proposed transporter required for Pi root to shoot translocation (Liu *et al.*, 2012**a**).

WOODEN LEG (WOL) is an important histidine kinase receptor transmitting cytokinin signalling with a wide variety of functions: negative regulation of phosphate starvation response (Franco-Zorrilla *et al.*, 2002), establishment of root vasculature in the embryo (Mähönen *et al.*, 2000) and negative regulation of sulfate assimilation through downregulation of high-affinity sulfate transporters (Maruyama-Nakashita *et al.*, 2004).

2.3.8.3 Cell wall biosynthesis

Cell walls serve as an important sink of metals in both metal tolerant and intolerant plants (Li *et al.*, 2015**a**). In particular, cell wall polysaccharides, such as low methylestrified pectins provide functional groups to which divalent and trivalent metal cations can bind. Below, outlier genes with functions related to cell wall biosynthesis and maintenance and thus potentially important for nickel accumulation in hyperaccumulator cell walls are pin-pointed.

First of all, different *GLUCAN SYNTHASE-LIKE* genes were found to be putative targets of selection in Carratraca-S and Samil-S: *GSL12* in Carratraca-S and *GSL1* in Samil-S. *GSL* is a family of 12 callose synthase genes. Callose, a β-1,3-glucan is deposited at the plasmodesmata, cell plate and in phloem cells after wounding and pathogen attack, and is also important for proper pollen development (Enns *et al.*, 2005). *GSL1* has a partially redundant role in forming the callose wall separating the microspores of the tetrad and in pollen grain maturation (Enns *et al.*, 2005). *GSL12* is involved in plasmodesmata aperture formation by callose deposition and thus essential for cell signalling (Vatén *et al.*, 2011). Also, a GPI-anchored *PLASMODESMATA CALLOSE-BINDING PROTEIN 5 (PDCB5)*, identified as an outlier gene in Carratraca-S, is localised to the plasmodesmata and predicted to bind callose (Simpson *et al.*, 2009).

Genes involved in lignin and chitin metabolism in Carratraca-S included: *CHITINASE-LIKE PROTEIN 1 (CTL1), CHITINASE-LIKE PROTEIN 2 (CTL2), CELLULOSE SYNTHASE-INTERACTIVE PROTEIN 1 (CSI1)*, and *CAFFEOYL SHIKIMATE ESTERASE (CSE)*; while in Samil-S: *PEROXIDASE 2 (PRX2), FERULIC ACID 5-HYDROXYLASE 1 (FAH1)*. Two homologs - CTL1 and CTL2 are chitinase-like proteins secreted to the apoplast whose proposed, mutually redundant, role is in glucan chain assembly, facilitating interactions between hemicelluloses and cellulose microfibrils (Sánchez-Rodríguez *et al.*, 2012). *CTL2* was also found to be downregulated in the shoot only in the Carratraca-S population. *CS1* interacts both with cortical microtubules and with cellulose synthase complexes involved in primary cell wall biosynthesis, providing a link between the two and together forming a scaffold required for guiding cellulose biosynthesis orientation (Gu *et al.*, 2010; Bringmann *et al.*, 2012; Li *et al.*, 2012**b**; Landrein *et al.*, 2013). *CSE* has a novel role in lignin biosynthesis and it is responsible for hydrolysis of caffeoyl shikimate into caffeate and shikimate (Vanholme *et al.*, 2013).

PRX2 is hypothesised to function in the last step of lignin biosynthesis in oxidative coupling of monolignols to the growing lignin polymer in the stem, based on analysis of mutant phenotypes and characterisation of chemical activity of a recombinant protein produced in *E. coli* (Shigeto *et al.*, 2013, 2014). *FAH1* functions in lignin biosynthesis pathway as well and its role lies in shuttling guaiacyl lignin pathway intermediates into the pathway leading to production of lignin composed of syringyl units (Meyer *et al.*, 1998; Humphreys *et al.*, 1999; Marita *et al.*, 1999; Weng *et al.*, 2010). The gene is downregulated in the shoot in the Samil-S population relative to others.

2.3.8.4 Lipid metabolism

Shared One lipase was identified amongst shared candidate genes in the two serpentine populations: *PHOSPHOLIPASE C 2 (PLC2)*. It belongs to a large group of phosphoinositide-specific phospholipase genes and hydrolyses phosphatidylinositol 4,5-bisphosphate into inositol 1,4,5-trisphosphate and diacylglycerol, two important second messengers (Hirayama *et al.*, 1997; Otterhag *et al.*, 2001). In addition, Samil-S's candidate list contains *PHOSPHOLIPASE D ALPHA 1 (PLDALPHA1)*, which hydrolyses glycerophosholipids to

phosphatidic acid, then quickly converted to diacylglycerol (Fan *et al.*, 1997; Pappan *et al.*, 1997) as well as an uncharacterised mono/di-acylglycerol lipase.

Carratraca-S Two additional genes directly involved in lipid metabolism discovered in Carratraca-S were: *ACETYL CO-ENZYME A CARBOXYLASE CARBOXYLTRANSFERASE ALPHA SUBUNIT (CAC3)* and *ACYL-COA OXIDASE 1 (ACX1)*. *CAC3* encodes the alpha subunit of acetyl-CoA carboxylase catalysing a critical step in *de novo* fatty acid biosynthesis – biosynthesis of malonyl-CoA from acetyl-CoA and bicarbonate (Baud *et al.*, 2003; Mu *et al.*, 2008). On the other hand, *ACX1* is one of the six enzymes responsible for catalysing the first rate-limiting step of beta oxidation of long chain fatty acids – conversion from acyl-CoA to trans-2-enoyl-CoA (Adham *et al.*, 2005; Pedersen & Henriksen, 2005). *STEROL GLUCOSYTRANSFERASE (SGT)* is an enzyme whose function is glycosylating sterols (Stucky *et al.*, 2015). The end product of that reaction – sterol glucosides play a structural role in plant membranes but they are also hypothesised to function in pollen coat maturation (Choi *et al.*, 2014), sugar transport, storage and signalling (Grille *et al.*, 2010) as well as tolerance to abiotic stress: salt, heat, cold (Mishra *et al.*, 2013).

Samil-S *ECERIFERUM3 (CER3)* forms a complex catalysing the synthesis of very long chain alkanes from very long chain acyl-CoAs and thus is required for production of the main cuticular wax component (Bernard *et al.*, 2012; Lam *et al.*, 2012). *16:0 DELTA9 DESATURASE 2 (ADS2)* is a 16:0 desaturase of monogalactosyl diacylglycerol and phosphatiadylglycerol, and is essential for plasma membrane acclimation to cold and freezing through the shift to higher unsaturated lipid content (Chen & Thelen, 2013).

2.3.8.5 Response to abiotic stress

Shared *SYNAPTOTAGMIN A (SYTA)* is a member of a plant family homologous to animal membrane trafficking proteins, important for, amongst other things, plasma membrane resealing. *SYTA* is an integral membrane protein which has been shown to be a crucial for plasma membrane integrity, and rapidly upregulated in response to freezing and osmotic stress (Schapire *et al.*, 2008; Yamazaki *et al.*, 2008). *SYTA* also regulates endocytic recycling pathway and as a consequence trafficking of viral proteins (Lewis & Lazarowitz, 2010).

A number of heat shock proteins and chaperones were found amongst the outlier genes. Carratraca-S and Samil-S contained two DnaJ-domain (HSP40) chaperones, each, as well as a HSP70 and HSP90 protein, respectively. This underlies the importance of maintaining such aspects of cellular homeostasis like correct protein folding, membrane and protein stability and protein degradation (Wang *et al.*, 2004) in *Alyssum* hyperaccumulators' cells which can face many abiotic stressors, such as Ni, redox and thermal stress. Previous proteomic research showed increase in expression of one HSP70 and one HSP90 gene on nickel exposure in the hyperaccumulator *Alyssum lesbiacum* (Ingle *et al.*, 2005**b**)

Carratraca-S *TWO-PORE CHANNEL 1 (TPC1)* forms the Slow Vacuolar (SV) channel in *Arabidopsis* whose function is still debated (Gutla *et al.*, 2012). *TPC1* is regulated by magnesium, sodium, reducing agents, and importantly, vacuolar calcium (Dadacz-Narloch *et al.*, 2011). Recently, *TPC1* has been demonstrated to be a key component in rapid systemic stress signaling dependent on Ca^{2+} wave propagation (Choi *et al.*, 2014). *TPC1* is both a SNP-based candidate in Carratraca-S and shoot downregulated gene in Samil-S. Interestingly, it is yet another *Arabidopsis lyrata*'s candidate gene for serpentine adaptation overlapping with candidate genes from the current dataset (Turner *et al.*, 2010). *SALT OVERLY SENSITIVE 1 (SOS1)* is a key element of salt stress protection system in plants. It is a plasma membrane Na^+/H^+ antiporter responsible for Na^+ efflux and thus keeping Na^+ levels in the cytoplasm at a non-toxic level as well as long-distance Na^+ translocation from root to the shoot (Zhang & Shi, 2013). *SOS1* is positively regulated by a kinase complex comprised of Ca^{2+} activated protein *SOS3* and the kinase *SOS2* (Quintero *et al.*, 2002, 2011; Zhang & Shi, 2013), which tellingly, is a candidate gene in the Samil-S population. Under conditions of NaCl stress, *SOS2* phosphorylates *SOS1* which increases its transport activity.

Samil-S *SODIUM HYDROGEN EXCHANGER 4 (NHX4)* is one of the six *Arabidopsis* NHX intracellular Na^+/H^+ antiporters (Li *et al.*, 2009**b**; Bassil *et al.*, 2011). It belongs to the vacuolar group and its expression is responsive to salt stress. Evidence for a role in salt

stress response comes from salt hypersensitivity of *NHX4* overexpressor lines and *E. coli* expressing *NHX4* as well as increased salt tolerance of *nhx4* plants (Li *et al.*, 2009**b**).
Two other outlier genes of note in Samil-S were: *SALT INDUCIBLE KINASE (SIK)*, a kinase involved in the negative regulation of osmotic stress genes (Hwang *et al.*, 2002) and *HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 15 (HOS15)*, a WD40-repeat protein, which acts as a repressor of genes associated with cold tolerance through histone H4 deacetylation (Zhu *et al.*, 2008**a**).

2.3.8.6 Defence response to pathogens and herbivores

Shared *MILDEW RESISTANCE LOCUS 2* (*MLO2*, shared candidate) and *MLO12* (candidate specific to Samil-S) are amongst 15 genes with homology to barley mildew resistance locus o. They form part of an ancient system modulating resistance against nonadapted powdery mildew pathogens (Humphry *et al.*, 2010). *MLO2* and *MLO12* together with *MLO6* are required for fungal penetration of the cell wall (Consonni *et al.*, 2006, 2010), and while *MLO2* knockout confers partial immunity, *mlo2 mlo6 mlo12* mutant is totally immune against powdery mildews. This resistance is mediated through indolic metabolites: glucosinolates and camalexins (Consonni *et al.*, 2010).
Moreover, Carratraca-S and Samil-S had different putative disease resistance genes amongst the outliers: 5 and 15 genes belonging to TIR-NBS-LRR class, respectively.

Samil-S In Samil-S, *HIGH ALIPHATIC GLUCOSINOLATE* (*HAG1*, also known as *MYB28*), was discovered amongst the outlier loci. It is one of the three MYB factors co-ordinately trans-activating production of methionine-derived (aliphatic) glucosinolates and down-regulating expression of indole glucosinolate biosynthesis genes (Gigolashvili *et al.*, 2007, 2008). Glucosinolates are Brassicales-specific group of nitrogen- and sulphur-rich secondary metabolites with effective anti-pest and anti-pathogen properties (Halkier & Gershenzon, 2006).

Carratraca-S Interestingly, another gene involved in the production of aliphatic glucosinolates was found in Carratraca-S. *2-ISOPROPYLMALATE SYNTHASE (IMS2)* is

a chloroplast-localised methylthioalkylmalate synthase, which is responsible for the first step in the aliphatic glucosinolates biosynthesis pathway – elongation of the chain by insertion of methylene groups (Textor *et al.*, 2007).

WRKY46, a transcription factor which plays a variety of roles in protection against biotic and abiotic stressors, was also found in Carratraca-S. It is rapidly induced by salinity, drought and hydrogen peroxide, and regulates a number of downstream targets involved in cell redox homeostasis (Ding *et al.*, 2014). *WRKY46* expression is upregulated by salicylic acid and pathogen infection and it positively regulates resistance to pathogenic bacteria, likely by activating *ICS1*, an enzyme in the SA biosynthesis pathway, amongst its downstream targets (van Verk *et al.*, 2011; Hu *et al.*, 2012**a**). In addition, *WRKY46* is expressed in the guard cells where it regulates light-dependent starch metabolism and consequently affects opening of the stomatal aperture (Ding *et al.*, 2014).

Furthermore, Ding *et al.* (2013) have unequivocally shown that *WRKY46* is a negative regulator of root malate exporter *ALMT1* which is one of the key Al tolerance genes.

2.3.8.7 Regulation of flowering, leaf and root morphogenesis

Shared *HISTONE DEACETYLASE 2A (HDA2A)* belongs to a group of four plant-specific histone deacetylases in the *Arabidopsis* genome. They are part of a pathway helping to establish adaxial-abaxial leaf polarity (Wu *et al.*, 2000; Ueno *et al.*, 2007). *SCAR2* is one of four SCAR/WAVE genes involved in positive activation of Arp2/3 complex whose role is inducing nucleation and rearrangement of actin cytoskeleton (Zhang *et al.*, 2005**a**,**b**). It is also known as *IRREGULAR TRICHOME BRANCH 1* due to the mutant phenotype's resemblance to the *distorted* mutant phenotype (misshapen trichomes due to aberrant expansion of the trichome stalk and stunted branches).

Population-specific In Carratraca-S, *CROOKED (CRK)*, one of the members of the Arp2/3 complex, turned out to be a possible target of directional selection. As such, CRK is involved in functions relating to the actin cytoskeleton in development and physiology, such as the recently uncovered feedback loop mechanism between H_2O_2 production and cytoskeleton remodelling in guard cells leading towards stomatal closure (Li *et al.*, 2003**a**,

2014**b**). Similarly to SCAR2, the *crk* mutant belongs to the class of mutants exhibiting *distorted* mutant trichome phenotype because of disruption in fine actin polymerisation (Mathur *et al.*, 1999, 2003).

Another gene related to the cytoskeleton – *MYOSIN XI K (XIK)* was also discovered amongst outliers in Samil-S. It is one of 13 class XI myosin genes responsible for actin filament shape change and turnover necessary for e.g. organelle and vesicle trafficking important for cell expansion (Peremyslov *et al.*, 2012, 2013; Cai *et al.*, 2014), polar tip growth in root hairs (Peremyslov *et al.*, 2008; Prokhnevsky *et al.*, 2008; Park & Nebenführ, 2013), development of different leaf epidermis cell types, such as trichomes (Ojangu *et al.*, 2012) and resistance against fungal penetration (Yang *et al.*, 2014).

2.3.8.8 Other

Shared *POLYOL/MONOSACCHARIDE TRANSPORTER 6 (PMT6)* belongs to a subfamily of six *Arabidopsis* genes in a family of 60 monosaccharide transporter-like (MST-like) genes. While the gene itself has not been characterised yet, co-family members are symporters transporting a wide range of sugar substrates: linear polyols, e.g. xylitol, sorbitol but also *myo*-inositol and many hexoses and pentoses present in roots and leaf vasculature (Klepek *et al.*, 2005, 2010).

Different *SYNTAXIN* genes were found among the outliers in the two serpentine populations: *SYP81* in Samil-S and *SYP42*, *SYP51*, *SYP61* in Carratraca-S. Also knowns as Qa-SNAREs, they mark target plant membranes to safeguard precise vesicle-associated membrane fusion events (Bassham *et al.*, 2008). SYP42, along with two other SYP proteins - SYP41 and SYP43, belongs to the SYP4 group of syntaxins in *Arabidopsis*, which localize to the trans-Golgi network (TGN) (Uemura *et al.*, 2012). TGN is an organelle responsible for sorting of the endocytosed and secretory cargos and forming the early endosome (Dettmer *et al.*, 2006; Viotti *et al.*, 2010) and SYP4 proteins regulate the transport pathways downstream from TGN. Likely because of their involvement with secretion, SYP4 proteins are required for *PEN1/2*-dependent resistance to fungal pathogens (Uemura *et al.*, 2012). *SYP61* has been hypothesised to participate in vesicle transport required in response to

(a)biotic stressors (Drakakaki *et al.*, 2011). SYP61 interacts with another syntaxin SYP121 to facilitate post-Golgi trafficking of PIP2;7 aquaporin to the plasma membrane and *syp1* mutants show high sensitivity to osmotic stress (Zhu *et al.*, 2002; Hachez *et al.*, 2014). Furthermore, *SYP61* is hypothesised to facilitate exocytosis of the cell wall components (Drakakaki *et al.*, 2011). SYP51 together with three other proteins determines membrane fusion events between the prevacuolar compartment and the vacuole which is required for e.g. proper seed development (Ebine *et al.*, 2008).

2.3.9 d*N*/d*S* and McDonald-Kreitman test

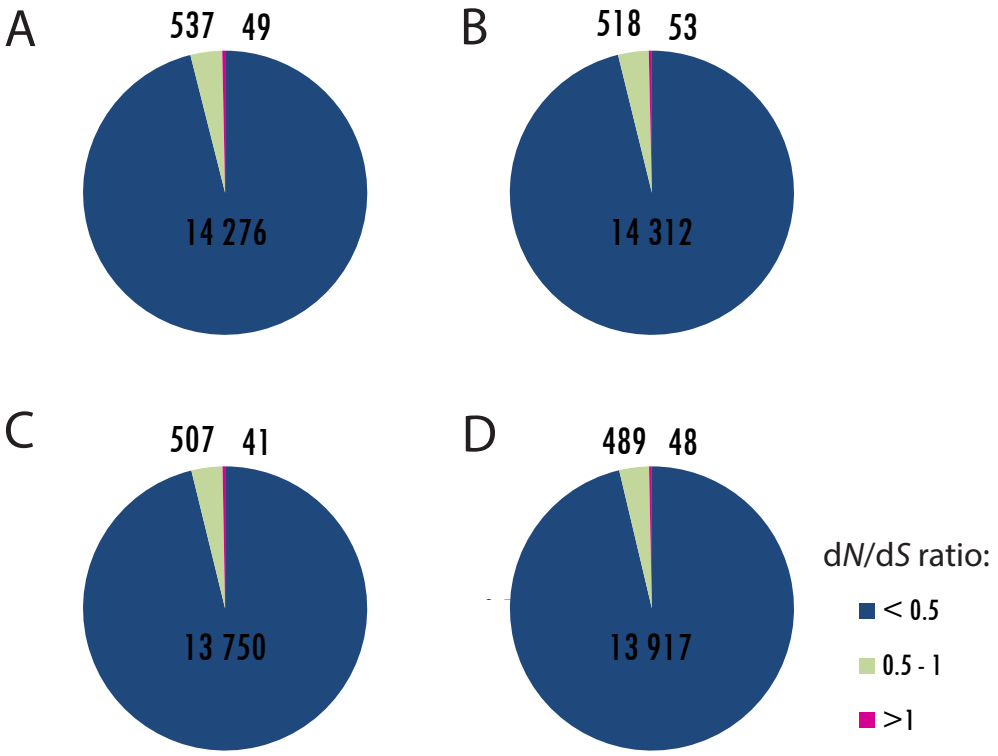


Figure 2.25: Breakdown of d*N*/d*S* ratios as calculated against *Clypeola* for: **A** Carratracca-S, **B** Morata-NS **C** Rubia-NS **D** Samil-S

Rates of substitution for nonsynonymous and synonymous sites between homologous sequences from non-accumulator outgroup *Clypeola jonthlaspi* and every *Alyssum serpyllifolium* population were compared in turn. As expected, the vast majority of loci showed d*N*/d*S* < 0.5 (Figure 2.25) and were thus experiencing long-term purifying selection.

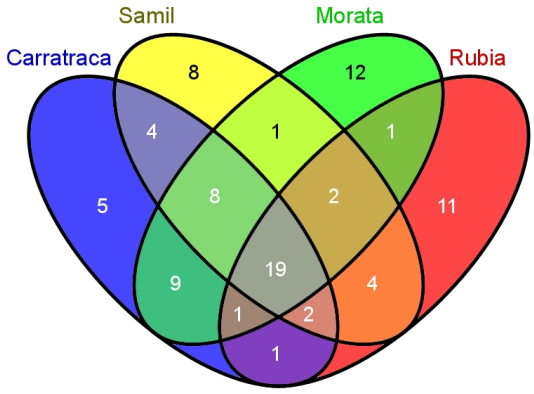


Figure 2.26: The extent of overlap of genes with elevated (>1.1) d*N*/d*S* ratios across the *Alyssum serpyllifolium* populations: Carratraca-S, Samil-S, Morata-NS, Rubia-NS

Between 41 and 53 genes showed elevated d*N*/d*S* ratios of > 1.1 in the individual *Alyssum* populations indicating possible historical action of positive selection. The majority of these genes were shared between the populations (Figure 2.26) and only a few of them showed statistically significant log-likelihood ratio test result (χ_2, $p < 0.05$), when comparing ω fixed at 1 (neutrality) and the empirical ω. In Carratraca-S and Samil-S populations, one gene belonging to a family of unknown function (DUF577), which was earlier shown to be upregulated both in the shoot and in the root in all the *Alyssum* populations relative to *Clypeola*, was found. In Morata-NS, *PATHOGENESIS-RELATED 1* (*PR1*), a defence response marker for systemic acquired resistance induced by plant pathogens, belonged to that category. However, both genes showed elevated d*N*/d*S* ratio of > 1 in all four *Alyssum* populations, so they are not candidate genes for serpentine adaptation.

The McDonald-Kreitman test was then carried out to jointly contrast the amount of nonsynonymous and synonymous divergence and polymorphism between each *Alyssum* population and *Clypeola jonthlaspi*. The test is based on the null assumption that the levels of interspecific divergence and intraspecific polymorphism observed in a gene are expected to be correlated because of constant mutation and substitution rate unless some other force, such as selection, is acting. A significantly higher rate of nonsynonumous to synonomous substitutions in the between-species comparison than within-species comparison suggests the action of positive selection, whereas the opposite suggest the action of balancing selection (Vitti *et al.*, 2013). Contigs with neutrality index significantly ($p < 0.05$) higher or

lower than 1 were extracted for further analysis as putative targets of negative and positive selection, accordingly. In total, 399, 387, 339 and 303 targets of negative selection were detected in Carratraca-S, Samil-S, Morata-NS and Rubía-NS populations, respectively, whereas 68, 68, 82 and 64 targets of positive selection were found in those populations.

2.3.9.1 Negative selection

The McDonald-Kreitman test allows detection of negative selection, i.e. purification of a gene from new mutations, more often than not deleterious, which maintains low gene diversity in the region. Genes under negative selection are understood to be of vital functional importance as their sequence needs to remain conserved. In the context of current research, the aim was to find out the identity of genes under particularly strong negative selection in the *Alyssum* lineage, but especially specific to the serpentine lineages.

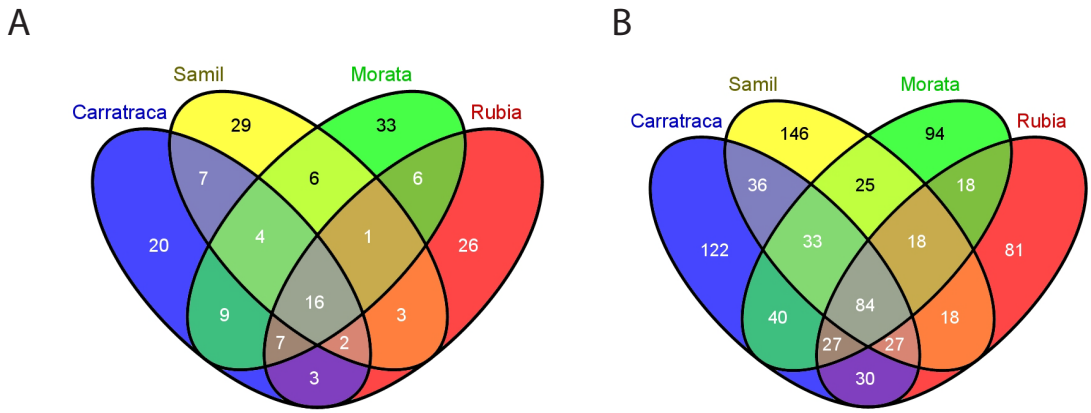


Figure 2.27: The extent of overlap of candidate genes experiencing **A** positive **B** negative selection according to the McDonald-Kreitman test across the *Alyssum serpyllifolium* populations: Carratraca-S, Samil-S, Morata-NS, Rubia-NS

All *Alyssum* populations A limited number of targets of both positive (Figure 2.27 A) and negative selection (Figure 2.27 B) was found to be overlapping between serpentine and non-serpentine populations despite long shared evolutionary history relative to the *Clypeola* outgroup. Eighty-four targets of negative selection were found to be shared among all *Alyssum* populations. The functions of these genes were enriched for *protein desumoylation* (6 genes) and *hydrogen peroxide biosynthetic pathway* (4 genes), *amine catabolic process* (5 genes, including *THA1* and 2 out of 3 genes involved in proline degradation: *P5CDH* and *PDH1*), and *trichome morphogenesis* (*DEK1*, *GNOM*, *HEN2*, *VIL2*).

Proline, in addition to its essential protein building block function, protects against osmotic and metal stress (Anjum *et al.*, 2014) and serves as a transient nitrogen sink (Deuschle *et al.*, 2001; Cecchini *et al.*, 2011). Mitochondrial enzymes *PROLINE DEHYDROGENASE 1* (*PDH1*) and *DELTA1-PYRROLINE-5-CARBOXYLATE DEHYDROGENASE (P5CDH*) are the first and the second enzymes in the proline catabolism pathway. Proline is first converted into toxic intermediate pyrroline-5-carboxylate by the action of *PDH1* which can then be oxidised by *P5CDH* into glutamate or recycled back into proline by pyrroline-5-carboxylate reductase (Deuschle *et al.*, 2001, 2004). The latter route results in an increased production of ROS in cells destined for programmed cell death through hypersensitive response on exposure to plant pathogens (Cecchini *et al.*, 2011; Monteoliva *et al.*, 2014), while the former route is intensified in response to water deprivation in growing tissues (Sharma *et al.*, 2011).

THREONINE ALDOLASE 1 (*THA1*) takes part in degradation of threonine into glycine. It is the threonine aldolase isoform expressed mainly in seeds and seedlings (Joshi *et al.*, 2006).

Shared between serpentine populations Thirty-six genes under negative selection (e.g. *LONG-CHAIN ACYL-COA SYNTHETASE 4 - LACS4, LOW PHOSPHATE ROOT1 - LPR1, DEFECTIVE IN MERISTEM SILENCING 5 - DMS5* and *METHIONINE AMINOPEPTIDASE 2A - MAP2A*) shared between Carratraca-S and Samil-S contained a higher proportion of *transferase* genes than expected by chance (11 genes – including *LYSM-CONTAINING RECEPTOR-LIKE KINASE 3 - LYK3, RNA-DEPENDENT RNA POLYMERASE 1- RDR1*, a glutathione S-transferase AT5G44000, *CBL-INTERACTING PROTEIN KINASE 3 - CIPK3/9*) as well as genes in the *vesicle docking involved in exocytosis* category (6 genes, including *EXOCYST SUBUNIT EXO70 FAMILY PROTEIN* genes: *EXO70A1, EXO70D3, EXO70H6*, *VACUOLAR PROTEIN SORTING 33 - VPS33*) and *glutamate metabolic process* (4 genes, including *DIT2.1, DELTA1-PYRROLINE-5-CARBOXYLATE DEHYDROGENASE - P5CDH*). More general categories connected to *development* and *reproduction* were also overrepresented in both populations.

LPR1 is an ER-localised multicopper oxidase, which together with its paralog and a P5-type

ATPase *PDR2*, relays low Pi status signal to *SCARECROW*, a key regulator of root meristem patterning, to optimize root architecture (shorter primary root, more lateral branching) under Pi deficiency conditions (Svistoonoff *et al.*, 2007; Ticconi *et al.*, 2009).
DIT2.1 and *DIT2.2* genes are a result of tandem genome duplication in *Arabidopsis*. *DIT2* is a chloroplastic glutamate/malate translocator crucial for primary and secondary ammonia assimilation in the photorespiratory pathway (Taniguchi *et al.*, 2002; Renné *et al.*, 2003).

The two *CBL-INTERACTING PROTEIN* kinases found to be under negative selection in the two serpentine populations (as well as non-serpentine Rubía-NS): *CIPK3* and *CIPK9* were recently shown to be key in regulating magnesium homoeostasis in the cell under condition of excess magnesium in relation to calcium, as experienced by plants under serpentine conditions (Tang *et al.*, 2015). CIPK3 and CIPK9 both interact with tonoplast-localised CALCINEURIN-B LIKE Ca^{2+} sensors CBL2 and CBL3 and along with two other CIPK genes form a signalling pathway essential for extra magnesium sequestration in the vacuole. On its own, CIPK3 forms a complex with CBL3 and participates in regulatory pathway maintaining potassium homoeostasis under low K^+ conditions in *Arabidopsis* (Pandey *et al.*, 2007; Liu *et al.*, 2013). In addition, one CBL gene - *CBL10*, was detected as a target of negative selection, only in the Samil-S population. CBL10 interacts with CIPK24 (SOS2) forming a vacuolar complex regulating the activity of Na^+ channels in response to shoot salt stress which promotes vacuolar sodium sequestration and thus salt tolerance (Kim *et al.*, 2007; Quan *et al.*, 2007; Lin *et al.*, 2009). CBL10 has been also implicated in K^+ homoeostasis by negatively regulating the activity of potassium channel KT1 which mediates K^+ uptake from the soil (Lin *et al.*, 2009; Ren *et al.*, 2013).

Carratraca-S Genes under negative selection in Carratraca-S belonged to the following overrepresented categories: *protein/DNA methylation* (10 genes, including *HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENE 1 - HOS1*), *protein ubiquitination* (11 genes, including *AXR1, THO6* and *NUCLEOSOME/CHROMATIN ASSEMBLY FACTOR GROUP C 3 - NFC3*) *regulation of flower development* (14 genes, including *CANDIDATE G-PROTEIN COUPLED RECEPTOR 1 - CAND1, HUA ENHANCER 1 - HEN1, HEN1*

SUPPRESSOR 1 - HESO1, PHOTOPERIOD-INDEPENDENT EARLY FLOWERING 1 - PIE1, RECEPTOR-LIKE PROTEIN KINASE 2 - RPK2 and *SPLAYED*), *trichome and root hair cell morphogenesis* (8 genes, including *SPIKE1*), *production of siRNA involved in RNA interference* (8 genes), *response to DNA damage and DNA repair* (13 genes, including *BREAST CANCER SUSCEPTIBILITY 1 - BRCA1, RAD54, REPLICATION PROTEIN A 1B - RPA70B,* AT4G32700), *response to red or far red light* (14 genes, including *SIGMA FACTOR 5 - SIG5, TIMING OF CAB EXPRESSION 1 - TOC1*), *response to freezing* (6 genes), *lipid storage* (6 genes, including *FATTY ACID DESATURASE 6 - FAD6*) and in general *post-translational protein modification* (49 genes – mainly *kinases* and *phosphatases*).

HOS1 is a RING finger E3 ubiquitin ligase, functioning as a negative regulator of cold responses by promoting degradation of *INDUCER OF CBF EXPRESSION 1 (ICE1)*, a cold-induced MYB transcription factor (Ishitani *et al.*, 1998; Lee *et al.*, 2001; Dong *et al.*, 2006). *HOS1* also participates in maintaining repression of flowering following short-term cold exposure by inducing expression of a flowering repressor *FLOWERING LOCUS C (FLC)* through attenuating the action of *HISTONE DEACETYLASE 6 (HDA6)*, which silences *FLC* chromatin (Jung *et al.*, 2013). Furthermore, *ICE1* also regulates flowering in response to photoperiod by negatively regulating the abundance of *CONSTANS (CO)*, one of the flowering promoters (Lazaro *et al.*, 2012). Recently, *HOS1* has been implicated in the maintenance of circadian rhythm by promoting mRNA export from the nucleus (MacGregor *et al.*, 2013).

SPIKE1 is a DOCK family guanine nucleotide exchange factor localised to the surface of the ER which is important for cytoskeletal organisation in the epidermal cells and thus essential for proper trichome and pavement cell morphogenesis (Qiu *et al.*, 2002; Zhang *et al.*, 2010). *SPIKE1* is part of a signalling complex which forms specialised domains in the ER called exit sites controlling WAVE/SCAR and ARP2/3 downstream signalling (Basu *et al.*, 2008; Zhang *et al.*, 2010).

FAD6 converts oleic acid (18:1) to linoleic acid (18:2) in the biosynthesis pathway of polyunsaturated fatty acid–containing lipids (Falcone *et al.*, 1994; Zhang *et al.*, 2009). *FAD6* is required for recovery from photoinhibition at low temperatures (Vijayan & Browse, 2002), temperature acclimation in membrane lipid composition (Falcone *et al.*, 2004),

high salt stress tolerance (Zhang *et al.*, 2009) and it also participates in the regulation of *SUPPRESSOR OF SA INSENSITIVE 2*-mediated defence signalling in chloroplasts (Kachroo *et al.*, 2003; Nandi *et al.*, 2003).

Samil-S In Samil-S, the following overrepresented functional categories were encountered among genes putatively under purifying selection in that lineage: *ATP binding* (46 genes), *RNA processing* (26 genes, such as *BSM, PTAC2*, *RHON, OTP84* and *SVR7* – in total, 3 out of 13 chloroplast RNA processing genes), *regulation of potassium transport* (2 out of 3 annotated genes in the category in the genome, including *CBL10*) and 2 *chaperones*: *MGE1* and *PFD4*.

2.3.9.2 Positive selection

Shared between serpentine populations Amongst the 7 genes likely to have been influenced by positive selection in Carratraca-S and Samil-S, 4 genes with *transmembrane transporter* activity were found (MATE efflux family protein AT1G61890, *AMINO ACID PERMEASE 5 - AAP5, NAP1, YELLOW STRIPE LIKE 1 - YSL1*), two of which are also involved in response to iron (*NAP1* and *YSL1*). *NAP1* is an iron-stimulated ATPase which is part of a complex participating in the assembly of plastidial sulphur-iron clusters (Xu *et al.*, 2005) and also plays a role in chlorophyll catabolism (Nagane *et al.*, 2010). Other genes included: *CELLULOSE SYNTHASE 5 (CESA5)*, which is necessary for the production of seed mucilage cellulose required for adhesion of mucilage pectin components to the seed (Harpaz-Saad *et al.*, 2011; Sullivan *et al.*, 2011). *AAP5* is a plasma membrane amino acid transporter mediating root uptake of basic amino acids arginine and lysine (Svennerstam *et al.*, 2008, 2011) as well as histidine in heterologous expression systems (Boorer & Fischer, 1997; Fischer *et al.*, 2002).

Carratraca-S The top category overrepresented amongst Carratraca-S's targets of positive selection was again *transmembrane transporter activity* - 11 genes, 5 of which are metal transporters (such as *NA+/H+ EXCHANGER 1 - NHX1* and two potassium transporters *KT1, KUP9* and a predicted potassium transporter AT1G60160), while two are AAP (amino

acid permease) family members – *LYSINE HISTIDINE TRANSPORTER 1 - LHT1* and *AAP5*. Lastly, two MAPKK kinases (MEKK subfamily) were found: *MITOGEN-ACTIVATED PROTEIN KINASE KINASE KINASE 7 - MAPKKK7, MAPK/ERK KINASE KINASE 1 - MEKK1*. *LHT1* was the first transporter functioning in root amino acid uptake identified in plants (Chen & Bush, 1997; Hirner *et al.*, 2006). *LHT1* has high affinity for histidine, glutamine and several other amino acids at low concentrations typically encountered in soils (Svennerstam *et al.*, 2008, 2011). *NHX1* is a tonoplast Na^+ and K^+/H^+ antiporter responsible for vacuolar compartmentalisation of Na^+ and important for salt tolerance (Apse *et al.*, 1999). Crucially, *NHX1* plays a vital role in K^+ accumulation in the guard cell vacuole, essential for the regulation of stomatal aperture opening (Barragán *et al.*, 2012; Andrés *et al.*, 2014).

Samil-S In Samil-S, significant overrepresentation of genes involved in *kinase* activity was revealed (10 genes, including *MEKK1*), in addition to individual targets, such as: *MILDEW RESISTANCE LOCUS O 15 - MLO15, NONEXPRESSER OF PR GENES 3 - NPR3* and *TRICHOME BIREFRINGENCE-LIKE 37 - TBL37*. *MEKK1* is part of MAPK signalling cascade known to be involved in stress signalling networks in animals and plants (Bjornson *et al.*, 2014). Depending on the identity of its downstream and upstream partners, *MEKK1* can participate in establishing plant response to different stressors, e.g. wounding (Hadiarto *et al.*, 2006), pathogens (Asai *et al.*, 2002), cold (Furuya *et al.*, 2014) and salt (Teige *et al.*, 2004) and is also part of a glutamate sensing pathway postulated to help adapt root architecture to local nitrogen supply in the soil (Forde *et al.*, 2013). *NONEXPRESSER OF PR GENES 3 (NPR3)* participates in the negative regulation of defence response to pathogens. It functions as a salicylic acid receptor and then promotes *NPR1*'s (key promoter of defence response) proteosomal degradation in a salicylic-acid concentration-dependent manner through its function as an adaptor of the Cullin 3 ubiquitin E3 ligase (Fu *et al.*, 2012; Shi *et al.*, 2013).

As can be seen, similar categories of genes were detected by McDonald-Kreitman test to those detected by differential gene expression analysis and search for F_{ST} outliers, which

is why the extent of precise overlap between the results of those tests was compared. In Carratraca-S. two putative targets of positive selection were also simultaneously F_{ST} outliers – potassium transporters *KT1* and *KUP9*. In Samil-S, two such genes were found: a protein phosphatase *HOMOLOGY TO ABI2* (*HAB2*) and *ATP-PRT1*, upregulated in all serpentine populations. Twelve genes deemed to be under negative selection by MK test were detected among Carratraca-S F_{ST} outliers and 13 amongst Samil-S outliers, and some of them were MK test outliers in both serpentine populations (*ATSIK, SULFOQUINOVOSYLDIACYLGLYCEROL 1 - SQD1, PROBABLE CDP-DIACYLGLYCEROL–INOSITOL 3-PHOSPHATIDYLTRANSFERASE 2 - PIS2*). Lastly, a possible link between differentially expressed genes in the *Alyssum* populations and candidate targets of selection based on McDonald-Kreitman test was tested. No significant overrepresentation of either upregulated or downregulated genes was discovered amongst genes with neutrality index significantly deviating from 1 in either direction (p-value < 0.01).

2.4 Key findings

- Thanks to the power of next-generation sequencing, the biggest library of *Alyssum* sequences yet has been generated, totalling 27.6 Mbp across 18,566 transcript sequences. High quality transcriptome assembly was achieved, with over 75% of transcripts reconstructed within over 90% of their predicted length.

- Sequencing 21 individual samples taken from four two serpentine (Carratraca-S, Samil-S) and two non-serpentine (Morata-NS, Rubia-NS) *Alyssum serpyllifolium* populations allowed investigation of intraspecific genetic variation related to edaphic adaptation, such as increased nickel tolerance and hyperaccumulation. Separate sequencing of two main plant parts: shoot and root was helpful in determining targeted gene expression changes as predicted due to their roles in plant physiology, particularly in metal homeostasis of interest here.

- Serpentine and non-serpentine *A. serpyllifolium* root and shoot transcriptomes were virtually indistinguishable against the non-accumulator *Clypeola jonthlaspi* outgroup. When analysed on their own, overall gene expression similarity across *A. serpyllifolium* populations correlated with geographic distance rather than ecotype. However, the most variable genes across the populations showed clustering by adaptation in the root, where gene expression changes of larger magnitude than in the shoot were observed across the populations. Larger dispersion in gene expression values in the root could reflect the challenges posed by direct contact with serpentine soils and necessary adaptation in the root proteome to maintain homeostasis.

- A relatively small number of tissue-wide transcriptome alterations was needed to acquire the full hyperaccumulator and serpentine-adapted state in *Alyssum*. As a first of its kind, the study identified both constitutive and Ni-inducible components of common gene expression differences between serpentine and non-serpentine ecotypes of *Alyssum serpyllifolium*, which proved to be limited to 52 genes. Future experiments could distinguish the two components by including samples under a variety of nickel

concentrations, from excess to no nickel. Putative locally adaptive gene expression differences were also identified (246 genes in Carratraca-S, 154 genes in Samil-S).

- Genes commonly upregulated both in the shoot (7 out of 28 genes) and in the root (3 out of 10 genes) in the two serpentine populations were significantly enriched for transporters, underlining their prominence in establishing effective metal hyperaccumulation. Transporters with firmly established roles in metal homeostasis were identified: *IREG1*, *IREG2*, *IRT1*, along with novel putative aminoacid transporters: *CAT4* and an AAAP-family member. A novel gene family of membrane proteins: CBS-DUF21 was heavily overrepresented, with three out of seven members upregulated in the roots of serpentine *A. serpyllifolium* populations.

- A number of high quality, non-singleton, bi-allelic SNP makers were identified across the four populations: 292,755 SNPs distributed across 12,934 genes. Each population appeared genetically distinct in PCA and STRUCTURE analyses carried out using a sample of SNP markers.

- A simple scan for directional selection revealed many outlier SNPs with unusually large differences in allele frequency (F_{ST} outliers) across serpentine and non-serpentine populations. In total, 388 and 369 genes fixed for at least one derived variant relative to *Clypeola* and non-serpentine populations were found in Carratraca-S and Samil-S, respectively. Similar proportion of genes and individual outlier SNPs - approx. one in ten - overlapped between the two serpentine populations, a result highly significantly deviating from chance. General support for adaptive significance of many outlier variants was provided by higher frequency of non-synonymous changes among outliers (43%) than genomewide (36%), and significantly reduced nucleotide diversity of the outlier genes.

- In both serpentine populations, F_{ST} outlier loci, McDonald-Kreitman test outliers and differentially expressed genes were enriched for similar functional categories, lending further support for their adaptive significance. They included: metal ion transport, oligopeptide transport, defence response (including salicylic acid signalling

pathway), response to various abiotic stressors (including iron deprivation response) as well as histidine biosynthesis and transport. In addition, putative targets of selection responsible for morphological differentiation between populations are likely contained within overrepresented categories related to leaf and root development – e.g. trichome and hair cell elongation and differentiation, meristem initiation, flowering.

- At the gene level, a significant number of genes was found to be simultaneously upregulated in the serpentine populations as well as contain outlier SNPs. These included *CAT4* and *IREG1* transporters, *ATP-PRT1* histidine biosynthesis enzyme and one member of the novel CBS-DUF21 family. In addition, some of these genes were previously found to be directly implicated in metal homeostasis and Ni hyperaccumulation, so they present a strong case for candidate genes involved in serpentine adaptation in *Alyssum*.

- The local adaptation process likely generated a considerable proportion of putative adaptive outlier polymorphisms and gene expression differences. There is evidence for substantial local adaptation within each serpentine population, with 45% of outlier alleles specific to a given serpentine population and absent in any other S or NS population.

- Genome-wide divergence, as measured by pairwise F_{ST} and D_{XY} distance metrics, is significantly higher between serpentine populations than between serpentine and non-serpentine populations. Significantly positive F_{IS} values across serpentine and non-serpentine populations indicate partial inbreeding and together with moderate nucleotide diversity are suggestive of a mixed mating system in *Alyssum serpyllifolium*.

Chapter 3

Tracing the Evolutionary History of the *Alyssum serpyllifolium* Species Complex

3.1 Introduction

The *Alyssum serpyllifolium* species complex belongs to a distinct group of metal-tolerant plants often described as pseudometallophytes and facultative metallophytes because the complex encompasses both metallicolous and non-metallicolous populations (Pollard *et al.*, 2014). The two main model species for metal hyperaccumulation, *Arabidopsis halleri* and *Noccaea caerulescens* also belong to the group of facultative metallophytes. Two broad hypotheses have been considered in literature so far regarding the origin and spread of metallicolous populations (Pauwels *et al.*, 2005). One hypothesis is based on the observation of often huge geographic distances separating populations inhabiting metalliferous outcrops, which would limit dispersal of metallicolous populations and instead favour evolution of metallicolous populations from nearby non-metalliferous sites (Schat *et al.*, 1996) driven by mechanisms of ecological speciation (Rundle & Nosil, 2005). The alternative hypothesis proposes a single origin of a metallicolous population, its spread across outlying metalliferous sites and subsequent differentiation between newly established metallicolous populations due to genetic drift. In this case, non-metallicolous populations could be either ancestral as in the previous scenario but also locally descendant from metallicolous populations.

Both metallicolous and non-metallicolous populations of *A. halleri* and *N. caerulescens*

display a basic constitutive level of zinc tolerance and hyperaccumulation (Meerts & Van Isacker, 1997; Bert *et al.*, 2000; Escarré *et al.*, 2000; Frérot *et al.*, 2003; Pauwels *et al.*, 2006; Meyer *et al.*, 2010). However, the degree of hyperaccumulation and tolerance correlates with substrate metal concentration, and tends to be lowest for non-metallicolous populations (Pauwels *et al.*, 2005). A similar pattern of a constitutive level of nickel tolerance and accumulation has been recently revealed in a common garden experiment with different populations of *Alyssum serpyllifolium* (A.J. Pollard & J.A.C. Smith, submitted). While it is generally believed that the emergence of the hyperaccumulation trait in the two model species was driven by and co-incident with appearance of anthropogenic metal polluted sites in the mining regions of Europe (Pauwels *et al.*, 2006; Jiménez-Ambriz *et al.*, 2007; Besnard *et al.*, 2009), recent research in *A. halleri* supports a much more ancient appearance of the trait or at least its *HMA4* component during actual speciation hundreds of thousands of years ago, and pre-dating any human industrial activity (Roux *et al.*, 2011). It has been proposed that the major loci of tolerance and accumulation, such as *HMA4* were selected early on in the history of the species (Pauwels *et al.*, 2005; Roux *et al.*, 2011) leading to constitutive zinc tolerance and accumulation in *A. halleri*, with local selection acting on additional genes in each established metallophyte population later to boost metal hyperaccumulation and tolerance ability. This has been supported by results of genome scan analyses, which showed the presence of outliers specific to different metallicolous *A. halleri* populations (Meyer *et al.*, 2009). High variation in zinc tolerance within non-metallicolous populations and individuals can then be explained by either local gene flow from metallicolous to non-metallicolous populations, or as a result of ancestral standing genetic variation present within the non-metallicolous populations which later was seized upon as a basis for adaptation in metallicolous populations.

The serpentine species of *Alyssum* occupy isolated ultramafic outcrops that have been exposed for many millions of years, at least since the Miocene, and so probably have a different evolutionary history from *Arabidopsis halleri* and *Noccaea caerulescens*, which are typically found in metal-contaminated sites created in the last two or three millennia as a result of anthropogenic disturbance (such as mine sites). Additionally, the majority

of range of *Alyssum serpyllifolium* was beyond the maximum advance of the main polar ice cap during the Pleistocene glaciations (Reeves, 1992; Hewitt, 1999), unlike the ranges of *A. halleri* and *N. caeruelscens*, possibly offering a more stable habitat and leading to a different evolutionary trajectory. However, glacial ice covered parts of the current range of *A. serpyllifolium*, including southern Spain, certain regions in central Spain, as as well as the Pyrenees and the south of France (Levin, 2006) so the Last Glacial Maximum cannot be entirely ruled out as a factor in the phylogenetic history of the *Alyssum serpyllifolium* species complex. Influence of climatic fluctuations can be important for population dynamics of serpentine species (Kolář *et al.*, 2012). One scenario describes formerly widespread metallophyte species being pushed out of non-serpentine sites during post-glacial reforestation due to its low competitiveness on non-metalliferous soils. Meanwhile, serpentine populations become separated and due to lack of or low gene flow progressively differentiate due to drift and selection. This depleted (or paleoendemic) species scenario (proposed by Stebbins, 1942; Stebbins & Major, 1965 and considered in the serpentine context by Kruckeberg, 1954 has been shown to have taken place in serpentine subspecies *Minuartia laricifolia* subsp. *ophiolitica* inhabiting the Alps (Moore *et al.*, 2013) and in *Strepthantus glandulus* complex (Mayer *et al.*, 1994).

This chapter continues the investigation of genetic data in the *Alyssum serpyllifolium* species complex, moving from comparing the patterns of gene expression and identifying footprints of selection to trying to establish the phylogenetic and demographic history which resulted in divergence of different metallicolous and non-metallicolous populations, and ultimately arriving at a hypothesis regarding the origin and spread of nickel hyperaccumulation trait, such as described above for zinc hyperaccumulators. To do that all, available main methods of phylogenomic reconstruction: mixture trees based on concatenated SNPs, population trees, species trees and phylogenetic networks will be first employed on transcriptome sequence dataset to infer the deep neutral phylogenetic history between serpentine and non-serpentine populations.

Then, three different genetic species delimitation approaches will be utilised to ascertain if

the neutral level of genetic divergence observed merits elevation of any of the populations within this complex to the species level. The *Alyssum serpyllifolium* species complex has so far been the focus of a limited number of studies trying to define the taxonomic status of populations/sub-species within it (Dudley, 1986**a**,**b**) as a part of a mission to attain a stable taxonomy of the genus. However, the sole focus on chiefly morphological and physiological character traits - such as the degree of nickel hyperaccumulation and tolerance observed under common laboratory conditions - as well as poor experimental design render the species delimitations of the hyperaccumulators *Alyssum malacitanum* and *Alyssum pintodasilvae* highly controversial. However, the framework of integrative taxonomy asks us to consider other types of genetic evidence when delimiting species (Leaché *et al.*, 2014), chief amongst them genetic data, which has not yet been applied in the context of the *A. serpyllifolium* species complex. As much as the species is a fundamental concept in biology, it lacks a single strict and objective definition which would allow testing the suitability of different species delimitations scenarios (De Queiroz, 2007). This has changed with the introduction of genetic data-based species delimitation methods (Fujita *et al.*, 2012). In contrast to species concepts linked to maintenance of genetic isolation between species, as in the biological species concept (Mayr, 1942) - defined as: "a species is a group of interbreeding populations that is reproductively isolated from other such groups", molecular species delimitation focuses on the endpoint of speciation as reflected directly in DNA sequences - involving a very high degree of genetic isolation between independently evolving metapopulations (De Queiroz, 2007) and falling under phylogenetic species concept, defined as: "a species an irreducible (basal) cluster of organisms, diagnosably distinct from other such clusters, and within which there is a parental pattern of ancestry and descent" (Cracraft, 1989).

While whole-genome data has had a deep impact on advances in describing the phyla on the tree of life, smaller- scale evolutionary changes resulting in speciation events had been recalcitrant to capture by quantitative sequence-based methods. However, adoption of the multispecies coalescent model has provided objectivity and statistical rigour, as well as a general impetus to speciation research (Carstens *et al.*, 2013**a**), despite

persistent computational problems in dealing with genome-wide datasets, with typical studies limited to 20, and no more than 100 loci (Fujita *et al.*, 2012). In traditional phylogenetic analysis, the same species tree is supposed to underlie all individual gene trees, while in the multispecies coalescent, coalescent processes in the ancestral and modern species and conflicts between species trees and gene trees can be accounted for. Strong phylogenetic signal can thereby be derived despite the low power of individual loci due to uncertainties in the branch lengths and topologies, which is especially important for studying closely related populations/species. Here, both Bayesian guided approaches requiring a prior consensus species tree (BP&P, Yang & Rannala, 2010) and unguided approaches (SNAPP-based Bayes' Factor Determination, Leaché *et al.*, 2014; STACEY, Jones, 2015) were compared. Species delimitation in the simplest terms is performed by establishing a threshold on the posterior node heights in the species tree, with small heights indicating evidence for collapsing a node and lumping the populations into a single species, or retaining the node as a speciation node for nodes with large heights showing reciprocal monophyly across many loci. Within this framework, models lumping and splitting populations into species in different ways can then be directly ranked.

Microsatellite genotyping of an extended sampling of a total of eight serpentine and non-serpentine Alyssum serpyllifolium populations in this chapter will serve to shed light on the shallower events in their demographic history. Microsatellite repeats are reliable and economical genetic markers with widespread use in population and conservation genetics/genomics (Selkoe & Toonen, 2006). They have typically high mutation rates (between 10^{-2} and 10^{-6} mutations per locus per generation), which makes them suitable to study recent species history (Schlötterer, 2000). Nowadays, high-throughput sequencing allows *in silico* detection of genetic markers with ease (Guichoux *et al.*, 2011), as is the case here with RNA-Seq-derived EST-SSRs. EST-SSRs are specifically believed to offer an advantage over intergenic-SSRs due to the high success rate of amplification and inter-species transferability (Varshney *et al.*, 2005; Barbará *et al.*, 2007; Ellis & Burke, 2007). This is a desirable feature in design of studies encompassing populations of higher levels of divergence, as in the case of *Alyssum serpyllifolium*. Furthermore, *de novo* assemblies have

proven to be a robust source of microsatellite markers, and authors routinely report PCR validation success rates of between 86% and 100% using EST-SSRs developed in this way (Garg *et al.*, 2011; Kaur *et al.*, 2011; Wei *et al.*, 2011; Zhang *et al.*, 2012). Therefore, EST-SSRs were chosen as markers to support analysis of a wider sampling of *Alyssum serpyllifolium* populations to investigate the more recent phylogenetic relationships and history of gene flow between serpentine and non-serpentine populations.

Lastly, genetic inference based on neutral variation will be compared with that of putatively adaptive variation. This will be achieved by an initial phylogenetic analysis involving genes putatively under positive selection in the serpentine populations and resequencing of two selected candidate genes and one neutral gene across a larger sample from the eight *Alyssum serpyllifolium* populations to contrast the observed sequence variation patterns. This work aims to establish the cause of the apparent genetic convergence in significant overlap of candidate serpentine adaptation genes identified in Chapter 2. In general, three mechanisms can be responsible for convergent molecular evolution: 1) independent *de novo* mutations in the same gene ('hotspots of evolution': Martin *et al.*, 2013) in different populations or species; 2) utilisation of pre-existing neutral standing genetic variation inherited from a recent common ancestor by selection as a consequence of new environmental pressures that render the variation adaptive; and 3) introduction of adaptive variation present in one population to the other by introgression. The last two mechanisms are obviously only possible in closely related lineages, and as such have been termed collateral evolution by Stern (2013), as opposed to parallel evolution describing the first mechanism, which in a sense is true homoplasy at a genotypic level, since it requires generation of independent mutations and not just the recycling/propagation of existing mutations Martin *et al.*, 2013. Current research suggests a far more prominent role for collateral evolution from standing genetic variation and introgression in sexually reproducing eukaryotes, in contrast to a relatively higher contribution of independent *de novo* mutations in unicellular organisms (Burke *et al.*, 2010). Simple prokaryotes have short generation times and are fast evolvers, which means that supply of *de novo* mutations is not a limiting factor in such organisms, but more complex organisms with longer life cycles

tend to depend on pre-exisiting genetic variation (Barrett & Schluter, 2008). Other factors influencing the emergence of parallel evolution include those related to selection, like the probability and magnitude of fitness increase by mutation at the locus (Chevin *et al.*, 2010; Conte *et al.*, 2012), but also those related to genetic architecture, such as the mutation and recombination rates at the locus (Chan *et al.*, 2010; Renaut *et al.*, 2014). In collateral evolution from standing variation, the amount of shared standing variation is key, which is in turn influenced by facets of demographic history, such as effective population sizes, history of gene flow, and selection (Westram *et al.*, 2014). For collateral evolution from introgression, geographical distances and species dispersal range would be of particular importance.

Neutral and divergent loci face different constraints in their histories and can inform us about distinct evolutionary processes (Colosimo *et al.*, 2005; Mitchell-Olds *et al.*, 2007; Nosil *et al.*, 2009; Johannesson *et al.*, 2010; Jones *et al.*, 2012). Neutral loci are expected to reflect the average history of diversification and gene flow of populations under study, while relationships estimated based on outlier loci will not necessarily reflect that history but also selective pressures - when these are shared between populations, they can group such populations together in phylogenies. In theory, it is possible to distinguish between parallel evolution from *de novo* mutation and collateral evolution by examining the gene region surrounding the focal allele(s) of convergent evolution (Stern, 2013). In the case of parallel evolution, phylogenies constructed from the candidate gene will be identical with the phylogenies from other genomic regions. On the other hand, the prediction for collateral evolution by hybridisation and from standing variation would be that the phylogenies from the gene under convergent evolution and closely linked regions would show inconsistency with the other gene trees.

In the case of *Alyssum serpyllifolium*, it should be possible to distinguish between evolution of serpentine adaptation/hyperaccumulation ability in parallel at each site from *de novo* mutations as contrasted with a single origin of the putatively adaptive genes (either from standing genetic variation, or by historical introgression) by studying the phylogenies of

outlier and neutral genes. For a single origin of adaptive alleles, phylogenetic trees based on candidate outlier alleles along with linked sites should show clustering by adaptation and not by geography (distance), as would be expected in the gene trees of neutral loci, assuming there is gene flow between adjacent populations. On the other hand, under the parallel evolution scenario, trees based on the sites linked to the outliers and on neutral genes should both result in similar clustering of the populations, likely by virtue of geography (Roda *et al.*, 2013).

Aims and objectives

- Genome-wide phylogenetic analyses to gauge the timing and patterns of ancestry and descent between a small sampling of serpentine and non-serpentine populations in the *A. serpyllifolium* species complex.
- First attempt at clarification of the taxonomic status of different serpentine and non-serpentine *A. serpyllifolium* populations using the multispecies coalescent species delimitation framework.
- Development of new microsatellite markers based on the *A. serpyllifolium* transcriptome. Application of the EST-SSRs in a denser sampling of serpentine and non-serpentine populations to estimate their genetic diversity and differentiation between populations, and assessing correlation of these with ecotype and geography.
- Resequencing of firm candidate genes proposed to be adaptive for hyperaccumulation ability, along with a neutral gene for comparison, across additional replicate serpentine and non-serpentine populations to determine the origin of adaptive variation.

3.2 Materials and methods

3.2.1 Demography and phylogenetics analyses

Taking advantage of the availability of thousands of neutral SNP markers for which 21 individuals in the four *Alyssum serpyllifolium* populations were genotyped, a number of

methods were used to infer the genetic relationships between them. As defined in Chapter 2, the neutral dataset consists of 13,050 putatively neutral contigs, which did not appear to deviate significantly from neutrality according to any of the tests for selection carried out previously. Only sub-samples of neutral unlinked markers located in different contigs are included in all the analyses described below, unless stated otherwise.

3.2.1.1 SNP-based analyses

The D_A genetic distance measure (Nei *et al.*, 1983) was used to evaluate the relationships among the populations and the outgroup using a down-sampled dataset of 1,305 neutral SNPs in POPTREE2 (Takezaki *et al.*, 2010) with 1000 bootstrap replicates. The rest of the analyses described in this subsection were carried out using sub-samples of 13,050 unlinked neutral SNPs (one SNP/locus) genotyped with no missing data in any individual. A neighbour-joining tree was constructed based on a pairwise genetic distance matrix (Saitou & Nei, 1987) in MEGA 6 (Tamura *et al.*, 2013) with 1000 bootstrap replicates. A NeighborNet phylogenetic network was generated based on the UncorrectedP method (proportion of sites that are different in each pair of sequences) in SplitsTree ver. 4.13.1 (Huson & Bryant, 2006) with 1000 bootstrap replicates using default settings. Generation of a maximum likelihood (ML) tree based on concatenated SNPs was also attempted to obtain an overview of relationships among all the individuals, despite caveats involved in creating artificial concatenated sequence from SNPs located in loci with various evolutionary histories (Lemmon & Lemmon, 2013). This analysis was run on the RAxML BlackBox webserver (Stamatakis *et al.*, 2008) using the gamma model of rate heterogeneity and a search for best ML tree followed by 100 bootstrap replicates. The SNAPP ver. 1.2 (Bryant *et al.*, 2012) coalescent Markov Chain Monte Carlo (MCMC) sampler, which is, conversely, designed to infer species trees from independent biallelic markers while side-stepping gene tree construction, was then employed. Sampling was done every 1000 generations and 2 million generations were run in total. Convergence of parameters was assessed with Tracer ver. 1.6 (http://beast.bio.ed.ac.uk/Tracer).

Population trees were also surveyed in a different way using Treemix ver. 1.12 (Pickrell & Pritchard, 2012). Treemix implements a maximum likelihood method using allele

frequency data under a Gaussian approximation to genetic drift to infer population ancestry graphs and historical admixture events; populations are connected based on more incongruent allele-sharing proportions than expected by chance. In the analysis, *Clypeola* provided the root, and the presence of between 0 to 6 admixture events was tested. Finally, the results were bootstrapped 1000 times and the explanatory power of the seven models with different number of admixture events was compared using the Aikake Information Criterion (AIC).

3.2.1.2 Gene-based phylogenies in *BEAST

The gene trees derived from shared outlier loci and neutral loci were investigated with multi-locus multi-individual analysis in *BEAST ver. 2.3 (Bouckaert *et al.*, 2014). The analyses were run using two different sets of neutral and outlier loci: shared outliers with functions related to 1) metal transport and tolerance (10 genes), and 2) other functions (13 genes). Secondly, 30 putatively neutral genes that were of comparable genetic diversity to the putatively selected loci were chosen at random. In order to prevent bias introduced by artefactual genetic variation stemming from undetected gene duplication events, the analysis was also repeated using a subset of fifteen putatively neutral APVO genes (Duarte *et al.*, 2010), which belong to a set of confirmed single-copy genes from across the angiosperms. For *BEAST runs including the larger neutral gene dataset, 300 millions MCMC iterations were run due to larger size of the dataset, and for the three other datasets, 200 million MCMC iterations were run, sampled every 10 000 chains. The molecular clock was set to strict due to infra-specific sampling and consequent expectation of lower inter-branch rate variation (Brown & Yang, 2011), and population size model to follow linear growth with constant root. Yule prior was placed on species tree and best-fit sequence evolution model for each gene (HKY with empirical base frequencies) was determined with PartitionFinder ver. 1.1.1 (Lanfear *et al.*, 2014), using BIC (Bayesian Information Criterion). Divergence time calibration in the neutral gene phylogeny was based on A1 crown node age (i.e. the split between *Clypeola* and *Alyssum* section *Odontarrhena*, which contains all the known hyperaccumulator species of *Alyssum*) of 11.65 MYA and a 95% HPD confidence limit of 7.49–16.5 MYA taken from Flynn (2013). The divergence time parameters were sampled

from a log normal distribution to prevent incorrect sampling of negative values, and the distribution was defined by coefficients M = 1.066 and S = 0.08. The first 10% of results was discarded as burn-in, and run convergence and stationarity were inspected in Tracer ver 1.6 to confirm that ESS scores for all estimated parameters reached at least stable 200, followed by generation of maximum clade credibility tree with median heights in TreeAnnotator and visualisation of the species trees in DensiTree ver. 2.2.1 (Bouckaert, 2010) and FigTree ver. 1.4 (http://tree.bio.ed.ac.uk/software/figtree/). Run convergence across multiple runs was established by performing three independent runs and checking the variability of results across runs; a single representative run is reported here.

3.2.2 Species delimitation

More objective and quantitative molecular species delimitation methods with explicit model assumptions have grown in popularity in recent times (Carstens *et al.*, 2013**a**). Generation of multiple gene sequences from RNA-Seq allowed application of these methods to the problem of species delimitation within the *Alyssum serpyllifolium* species complex.

3.2.2.1 BP&P

Firstly, joint Bayesian species tree estimation and delimitation was carried out in the program BP&P ver. 3.1 (Yang & Rannala, 2014). Its multispecies coalescent model (Rannala & Yang, 2003) accounts for incomplete lineage sorting, incongruity between gene trees and species trees, but as with other methods, it cannot handle migration post-speciation. Both fixed guide tree (here, *BEAST tree based on neutral loci was used) and newly implemented unguided species delimitation methods were tested in BP&P ver. 3.1 (Yang & Rannala, 2014). Diffuse priors differing by two orders of magnitude, as recommended by the authors, were chosen to explore the effect of priors on the results. The authors' simulations indicate that prior values do not influence the ranking of models, but only their posterior probabilities. For the gamma function describing Θ, the prior was set to G(1, 10) - large ancestral population sizes, or G(2, 2000) - small ancestral population sizes. For the gamma function describing τ at the root of the tree, the prior was set to G(1, 10) - deep divergence, or G(2, 2000) - shallow divergence. Other divergence times were

assigned the Dirichlet prior as indicated by the authors. All possible combination of Θ and τ priors were used, as their values can affect the outcome of the species delimitation procedure, with large Θ and small τ favouring conservative solutions with fewer species (Yang & Rannala, 2014).

Guide trees with one, two, three or four delimited species were considered and algorithm 0 or 1 with auto fine-tune selected to define reversal jump MCMC (rjMCMC) moves responsible for node collapse or split between species delimitation models in the fixed guide tree analyses, while the nearest-neighbour interchange algorithm was used for branch swapping in the unguided analysis. For both guided and unguided methods, each possible combination of four parameters (Θ prior, τ prior, number of delimited species, algorithm 0 or 1) was used, resulting in 32 runs in total. The sequence dataset analysed consisted of 116 putatively neutral loci with no missing data in each of the 21 individuals, with one haplotype sampled per individual. Genes with pairwise D_{xy} varying uniformly across one order of magnitude (between 0.002 and 0.021 on average) were used to ensure sufficient polymorphism information content between populations. Each analysis of 200,000 MCMC generations (sampling every two generations) was run with a burn-in period of 10,000. Care was taken to ensure that different starting trees were considered in the unguided analysis.

3.2.2.2 STACEY

The second approach utilising a multispecies coalescent model was STACEY ver. 1.0.4 (Jones, 2015; Jones *et al.*, 2015). It is an unbiased unguided approach with no *a priori* assignment of individuals into populations. It is a modification to the BEAST 2 software (Bouckaert *et al.*, 2014) which also includes a new built-in birth-death-collapse prior modification of the standard birth-death prior in BEAST (Jones, 2015). The addition of the prior results in a model in which the tip nodes of the species tree are effectively minimal possible clusters of individuals. As opposed to BP&P, STACEY does not make use of rjMCMC moves, but a new NodesNudge MCMC move, which can lead to better convergence. A sub-sample of 20 genes used in the BP&P analysis were selected based on

the top average pairwise D_{xy}, as an analysis of more loci would be too computationally intensive. Each of the 10 replicates was run for 200 million MCMC generations, stored every 5,000 generations. Otherwise, default settings recommended in the STACEY paper (Jones, 2015) were applied. Species Delimitation Analyser ver. 1.8.0 was then used to define the clusterings of individuals into putative species from MCMC tree samples (Jones, 2015) with default settings and 10% burn-in in the ten replicates.

3.2.2.3 Bayes' Factor Delimitation (BFD) in SNAPP

SNAPP (Bryant *et al.*, 2012) provides an alternative unguided method for estimation of species trees without MCMC integrating over or sampling gene trees, but instead using thousands of genome-wide markers and estimating the likelihood of change in allele frequencies across phylogenetic nodes. The advantage of the method lies in genome-wide sampling, instead of use of a handful of loci, and also in the lack of requirement for guide tree specification, which can introduce bias when the correctness of the tree is uncertain (Leaché *et al.*, 2014). Finally, the Bayesian approach allows straightforward comparison of alternative models of species delimitation using Marginal Likelihood Estimation (MLE) values, which first need to be adjusted with proportionality constants. Here, path sampling MLE approach was used, according to the framework introduced by Leaché *et al.* (2014). The competing species delimitation models were ranked according to MLE values, and Bayes' Factor (BF) was used to compare each model to the *4 species* null model.

Six realistic models for species delimitation, informed by previous studies, were tested. The null model was a *4 species* model, in which every population as identified by Structure analysis constitutes a separate species, and non-serpentine populations are equally distinct as serpentine populations. This model is not supported by any authority in the literature, as only differences between serpentine, and not non-serpentine populations, were ever considered meriting elevation to a subspecies or species rank. In the *3 species A* model, both non-serpentine populations Morata-NS and Rubía-NS constitute a single species, while Carratraca-S (*Alyssum malacitanum*) and Samil- S (*Alyssum pintodasilvae*) are separate species. This model is consistent with the taxonomic treatment of Dudley (1986a,b),

based chiefly on Ni tolerance and accumulation differences, with a small component of morphological variation. Conversely, in the *3 species B* model, the serpentine populations Carratraca-S and Samil-S belong to one species, while Morata-NS and Rubía-NS are separate entitites. In the *2 species A* model, the two serpentine and two non-serpentine populations each form a species, and in the *2 species B* model, the populations are grouped according to the consensus phylogeny described in this chapter: species 1 - Carratraca-S and Morata-NS, species 2 - Samil-S and Rubía-NS. In the case of each model, 200,000 MCMC replicates were run per step, and 48 steps in total on a sample of 1,149 unlinked neutral SNPs, as suggested by the authors due to computational limits. Default settings and priors were used, with the exception of alpha (=2) and beta (=1000), which places the prior for the population sizes at the lower limit in line with fairly low values measured for the Alyssum populations. Leaché et al. (2014) have shown that the two priors do not appear to have a significant effect on species delimitation outcome. The number of individuals sampled should be sufficient, as established by the robustness of results in simulation studies involving five individuals per species (Leaché et al., 2014). Path sampling results were summarised with Path Sample Analyser package from BEAST ver. 2.3 with alpha set at 0.3 and 10% burn-in.

3.2.3 Microsatellite development, genotyping and analysis

3.2.3.1 DNA extraction, EST-SSR detection and PCR amplification of EST-SSR fragments

Genomic DNA was extracted using a modification of the CTAB method (Rogers & Bendich, 1985; Porebski *et al.*, 1997) from up to 40 individuals in each of the eight sampled *Alyssum serpyllifolium* populations (Table 3.1). The concentration of DNA in each sample was measured using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, USA) following extraction.

Perfect microsatellites were detected in the consensus *Alyssum* RNA-Seq assembly with SciRoKo ver. 3.4 (Kofler *et al.*, 2007) using the following default settings: di-nucleotide pattern – minimum 7 repeats; tri-nucleotide – minimum 5 repeats; tetra-, penta-, hexa-

nucleotide – minimum 4 repeats. Each individual was then genotyped for 8 developed microsatellites. The PCR mixture for each individual amplification reaction was assembled with the following reagents: 1.5 µl of 10x buffer (New England Biolabs, Hitchin, UK), 0.3 µl of 10 mM total dNTPs (Thermo Scientific, Hemel Hempstead, UK), 1.2 µl of 25 mM $MgCl_2$ (New England Biolabs), 0.06 µl of 10 μM forward primer, 0.3 µl of 10 μM reverse primer, 0.3 µl of FAM/HEX/NED labelled universal primer, 0.1 µl of *Taq* Polymerase (New England Biolabs), 1-2 µl DNA solution (depending on the concentration, 10-100 ng contained in the reaction) and double-distilled water to 15 µl total volume. Oligonucleotide primers were supplied by Eurofins MWG Operon (Ebersberg, Germany), and in addition to the gene-specific sequences listed in Table 3.4, forward primers contained a 5′ overhang fragment of 5′-CACGACGTTGTAAAACGAC-3′ to facilitate incorporation of the dye-labelled universal primer in subsequent PCR cycles. PCR cycling was carried out using an MJ Research PTC-225 Peltier Thermal Cycler (MJ Research, Waltham, USA) or Eppendorf Mastercycler gradient (Eppendorf, Stevenage, UK) using the following conditions: (1) initial denaturation (3 minutes at 95°C), (2) 10 cycles of touch-down PCR (30 s at 95°C, 30 s at 65°C [reduced by one degree per cycle], 1 min at 68°C), (3) 30 cycles of standard amplification (30 s at 95°C, 30 s at 55°C, 1 min at 68°C), and (4) final extension (5 min at 68°C).

3.2.3.2 Genotype scoring

PCR-amplified fragments containing target microsatellite repeats were run on an Applied Biosystems Capillary Genetic Analyzer machine at the Department of Zoology, University of Oxford, UK, with GeneScan 500 LIZ Size Standard added as a reference and results collected using a DS-30 filter. Microsatellite genotypes were scored manually using GeneMarker ver. 2.6.3 (SoftGenetics, Pennsylvania, USA) following the default settings for plant microsatellite scoring. In cases where a genotype could not be established with confidence, PCR products were re-run a second time on the ABI machine. MICRO-CHECKER ver. 2.2.3. (Van Oosterhout *et al.*, 2004) was then employed to test for genotyping artefacts – stuttering, large allele drop-out and presence of null alleles. MICRO-CHECKER's combined use of deviation from Hardy-Weinberg equilibrium (HWE) across

multiple loci with homozygote allele size and genotype allele size difference distributions allows distinguishing homozygote excess due to inbreeding or Wahlund effect from that generated by genotyping errors. All of the individual genotypes were exported from MS Excel into CREATE ver. 1.37 (Coombs *et al.*, 2008), which was then used to convert the data matrix into formats required for various programs.

3.2.3.3 Data analysis

Firstly, departures from HWE and linkage disequilibrium (LD) were tested in Arlequin ver. 3.5 (Excoffier *et al.*, 2005) and FSTAT ver. 2.9.3.2 (Goudet, 1995). Subsequently, two programs were used in calculating basic population genetics parameters: allele frequency, *F*-statistics, heterozygosity, as well as AMOVA (Arlequin). The presence of an isolation-by-distance pattern (Wright, 1943) was investigated using the Mantel test with 1000 permutations (Mantel, 1967) in Genepop ver. 4.2 (Raymond & Rousset, 1995) by examining the correlation between log-transformed Euclidean distances between pairs of populations and linearized pairwise genetic distances (F_{ST}/(1- F_{ST})). The PHYLIP package ver. 3.69 (Felsenstein, 2005) was employed to calculate various inter-population genetic distance metrics with bootstrapping (1000 replicates), while Factorial Correspondence Analysis was carried out in Genetix ver. 4.05 using the default settings (http://www.genetix.univ-montp2.fr/).

Bayesian analysis in Structure ver. 2.3.3 (Pritchard *et al.*, 2000; Falush *et al.*, 2003) was used to assign individuals to genetic clusters (K) and to estimate admixture proportions (Q) for each individual. A model was chosen in which individuals had admixed ancestries and correlated allele frequencies, which should allow detection of more ancient admixture events (Falush *et al.*, 2003). Choosing independent allele frequencies made no significant difference to the structure detected, and so results based on correlated allele frequencies are reported here. The LocPrior clustering method implemented in the latest version of Structure was used, which is not only based on the individual multilocus genotypes but also takes into account the sampling locations. The model is recommended by the authors when the genetic data is not highly informative in order to detect population structure,

and was chosen because of a moderate number of microsatellite markers employed in the present study.

The value of K was set from a minimum of 2 to a maximum of 10, and five simulations were run for each K value with a burn-in of 100,000 and a main run of 1,000,000 MCMC iterations. To define the most probable number of genetic clusters (K) present in the data, the method proposed by Evanno *et al.* (2005) was initially used, which is based on an *ad hoc* measure ΔK that depends on the rate of change in the log probability of the data between successive values of K. Secondly, the number of true clusters was also inferred following L(K) over a number of clusters, and then looking for the signature of L(K) starting to plateau with diminished variance in L(K) (Pritchard *et al.*, 2000). These calculations were carried out by Structure Harvester Web ver. 0.6.94 (Earl & vonHoldt, 2012), which also generated CLUMPP input files. Subsequently, cluster assignments from across the replicate runs were aligned and averaged using CLUMPP ver. 1.1 (Jakobsson & Rosenberg, 2007) run with the Greedy algorithm and 1000 permutations of randomized input order. Resulting final cluster assignments were visualized using the program DISTRUCT ver. 1.1 (Rosenberg, 2004).

3.2.4 Sanger-based resequencing of candidate loci

3.2.4.1 Locus selection and primer development

Four candidate genes with outlier SNPs were tested: *ATP-PRT1*, *CAT4*, *NRAMP4* and *IREG1* because of their formerly investigated role in nickel hyperaccumulation in *Alyssum*: *ATP-PRT1* (Ingle *et al.*, 2005**a**), *CAT4* (S.T. Mugford, C.J. Snowden and J.A.C. Smith, unpublished results), or in other hyperaccumulators: *NRAMP4* (Oomen *et al.*, 2009; Halimaa *et al.*, 2014). *IREG1* was also included because of its proposed role in Ni transport and tolerance in *Arabidopsis thaliana* (Kirchner, 2009), as well as its differential expression between serpentine and non-serpentine populations in the *Alyssum* RNA-Seq results.

Primers in conserved regions surrounding target SNPs were developed using multiple sequence alignments from sequences of all individuals in the four sequenced *Alyssum serpyllifolium* populations using the PrimerDesign-M web server (Yoon & Leitner, 2015)

with default settings. For *ATP-PRT1*, *CAT4*, *IREG1* and *NRAMP4* amplification, 2, 3, 4 and 2 external and 1, 2, 1 and 2 internal primer pairs, respectively, were designed and tested in all possible combinations of forward and reverse primers. Genes with neutral genetic pattern of variation were selected for Sanger sequencing from the neutral gene set described before - the subset of stably expressed genes with no missing coverage for any sequenced individual, to increase the likelihood of future successful amplification in the extended sampling of populations, with a particular focus on genes with similar π and D_{xy} values and gene lengths to those of the candidate hyperaccumulator genes. Two primer pairs were designed and tested for each of the following genes: *6B-INTERACTING PROTEIN 1-LIKE 1* (*ASIL1*), *NON-ATPASE SUBUNIT 9* (*ATS9*). Neither of these genes is expected to play a specific role in ecological adaptation in *Alyssum*: ASIL1 is a transcriptional repressor of seed maturation genes in germinating seeds and seedlings (Gao *et al.*, 2009), whereas ATS9 is a subunit of the 19S regulatory complex from the 26S proteasome (Peng *et al.*, 2001).

Table 3.1: Extended sampling of *Alyssum serpyllifolium* populations used in Sanger resequencing of candidate genes and surveyed with EST-SSR

Population name	Serpentine?	Latitude, Longitude	Material collection
Alhaurín	N	36.91° N, 4.80° W	A.J. Pollard, J.A.C. Smith
Barazón	Y	42.92° N, 8.00° W	T.A. Flynn, J.A.C. Smith
Carratraca	Y	36.70° N, 4.60° W	A.J. Pollard, J.A.C. Smith
León	N	42.38° N, 5.43° W	A.J. Pollard (Furman University, USA)
Morata	N	40.30° N, 3.48° W	T.A. Flynn, J.A.C. Smith
Rubía	N	42.54° N, 6.82° W	T.A. Flynn, J.A.C. Smith
Samil	Y	41.46° N, 6.45° W	A.J. Pollard, J.A.C. Smith
Sierra Bermeja	Y	36.48 ° N, 5.20° W	C. Quintela-Sabarís (University of Santiago de Compostela, Spain)
Anduze	N	44.04° N, 3.96° E	M.K. Sobczyk
Bédarieux	N	43.62° N, 3.21° E	M.K. Sobczyk
Malaucène	N	44.04° N, 5.18° E	M.K. Sobczyk

3.2.4.2 DNA extraction, gene fragment amplification and Sanger sequencing

Genomic DNA was extracted and quantified as described in section 3.2.3.1. PCR mixture (30 µl) for each sample was assembled with the following reagents: 3 µl of 10x buffer (New England Biolabs), 0.6 µl of 10 mM total dNTPs (Thermo Scientific), 2.4 µl of 25 mM $MgCl_2$ (New England Biolabs), 0.6 µl of 10 μM forward primer, 0.6 µl of 10 μM reverse primer, 0.2 µl of *Taq* Polymerase (New England Biolabs), 1 µl DNA solution (25-250 ng), and 21.6 µl double-distilled water. In the case of *ASIL1* PCR, 4.8 µl of 5M betaine was added to the final concentration of 0.8 M with concomitant reduction of the water volume, to increase primer annealing specificity and fragment amplification rate (Ralser *et al.*, 2006). Custom oligos were supplied by Sigma Aldrich (Gillingham, UK). PCR cycling was carried out using an MJ Research PTC-225 Peltier Thermal Cycler or Eppendorf Mastercycler gradient using the following conditions: (1) initial denaturation (3 min at 95°C); (2) 10 cycles of touch-down PCR (30 s at 95°C, 30 s at 65°C [reduced by one degree per cycle], 1 min at 68°C); (3) 33 cycles (IREG1) or 40 cycles (*NRAMP4*) of amplification (30 s at 95°C, 1 min at 45–55°C, 4 min at 68°C), or in the case of *ASIL1* the first 10 cycles carried out at 42°C and the final 30 cycles with 50°C annealing temperature to facilitate permissive primer binding and target region amplification in the initial steps of the PCR reaction; and (4) final extension step (5 min at 68°C). To confirm the success of PCR reactions, PCR samples were separated with agarose gel electrophoresis, using 1.5% (w/v) agarose gel stained with ethidium bromide (1 μg/ml). Gels were typically run for 30 min to 1 h at 100–150 V and the DNA separation results visualised using Kodak Gel Logic 2200 Imaging System connected to a UV illuminator. PCR products of correct size, high purity and concentration were then sent to the DNA Sequencing Unit at the Department of Zoology, University of Oxford for clean-up, sequencing using BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Carlsbad, USA), purification of the products of sequencing reactions and subsequent run on an ABI 3730xl DNA Analyzer sequencing machine.

3.2.4.3 Data analysis

Raw ABI output chromatograms were examined and exported to the FASTA format using FinchTV software ver. 1.4 (Perkin-Elmer, Beaconsfield, UK). Consensus DNA contigs from

combined forward primer and reverse primer sequencing results were created manually in MEGA6 (Tamura *et al.*, 2013) from sequences trimmed on base call quality. DnaSP ver. 5 (Librado & Rozas, 2009) was employed to calculate the relevant population genetics parameters, carry out tests for selection, and phase the diploid sequences with PHASE algorithm (Stephens *et al.*, 2001). Lastly, AMOVA was performed in Arlequin ver. 3.5 (Excoffier *et al.*, 2005) and gene trees based on Nei's D_A (Nei *et al.*, 1983) were prepared in POPTREE2 (Takezaki *et al.*, 2010) with 1000 bootstrap replicates.

3.3 Results and discussion

3.3.1 Phylogenies based on neutral markers

Methods sampling both the entire neutral gene coding sequences (BEAST) and independent SNP markers distributed throughout thousands of neutral genes (Treemix, SplitsTree, RaxML, NJ, Nei's D_A) were employed to establish the phylogenetic relationships between the four *Alyssum* populations in the RNA-Seq study. Such dense sampling is expected to allow resolution of phylogenies regardless of short branch lengths or population sizes (Lemmon & Lemmon, 2013).

Both stochastic and non-stochastic processes inherent in evolution, such as drift, lineage sorting and introgression, often result in incongruence between different gene trees, and make individual gene trees unreliable for determination of relationships between populations/species that have recently diverged or are involved in rapid radiations (Cutter, 2013), and so this phenomenon was expected to be of note in closely related populations of *Alyssum serpyllifolium*. In such cases, coalescent, multi-individual, multi-locus methods such as *BEAST, which can simultaneously estimate individual gene trees and the overarching species tree while accounting for gene tree uncertainty and lineage sorting, as opposed to concatenation-based methods, are particularly well-suited (Kutschera *et al.*, 2014).

*BEAST results based on a random subsample of putatively neutral loci and a subsample of APVO (Duarte et al., 2010) neutral loci – that is, conserved single copy genes across eudicot and monocot genera (*Arabidopsis*, *Populus*, *Vitis*, *Oryza*), produced perfect agreement.

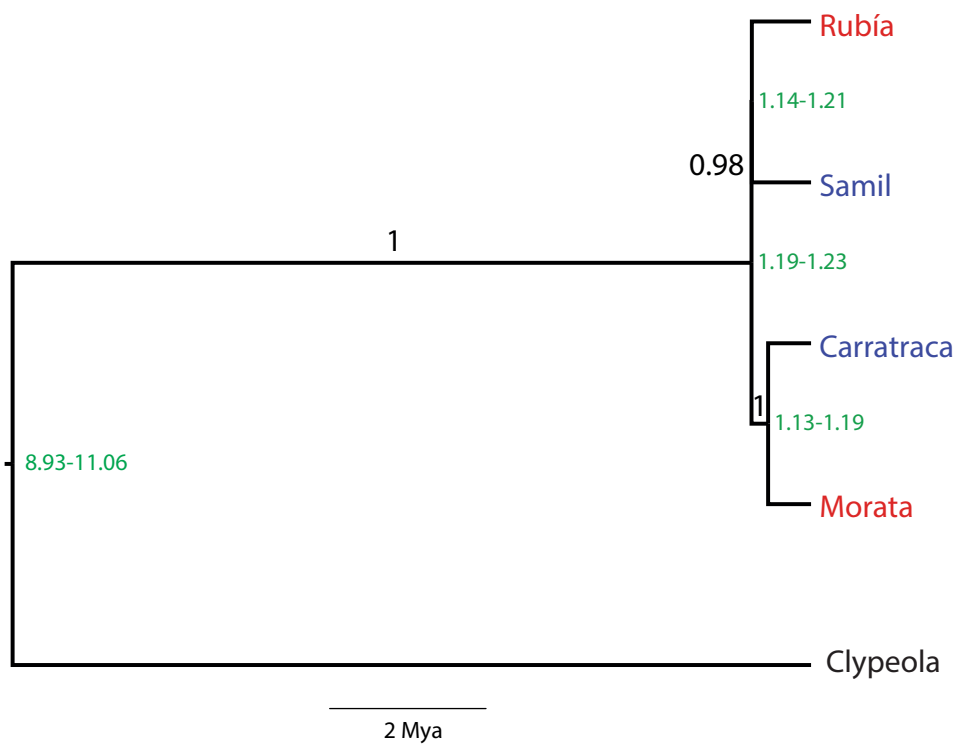


Figure 3.1: *Clypeola*-rooted *BEAST maximum clade credibility *Alyssum serpyllifolium* population tree based on 30 neutral genes. Posterior probabilities are given at each branch and 95% highest posterior density (HPD) intervals for node ages in Mya are marked in green. Serpentine populations are marked in blue and non-serpentine populations in red. Note that the apparent trichotomy is an illusion introduced by low node heights.

Furthermore, results from concatenated analysis of genome-wide sampling of SNPs in a ML framework in RAxML arrived at the same tree topology (data not shown). The trees recover two serpentine-non-serpentine subclades, (Rubía-NS, Samil-S) and (Carratraca-S, Morata-NS), with very high confidence (0.98 and 1.00 posterior probability, respectively: Figure 3.1). Secondary calibration from the *Alyssum* phylogeny of Flynn, (2013) was used to calibrate the split from the outgroup *Clypeola* based on the *Clypeola–Alyssum* section *Odontarrhena* crown node. The results indicate that the divergence between all four *Alyssum serpyllifolium* populations took place over a relatively short time in the overlapping interval of 1.23–1.13 MYA. The split between (Carratraca-S, Morata-NS) and (Rubía-NS, Samil-S) subclades is estimated to have taken place 1.23–1.19 MYA (95% HPD values), with subsequent divergence between 1.21–1.14 MYA (Rubía-NS, Samil-S) and 1.19–1.13 MYA (Carratraca-S, Morata-NS). The dating indicates an ancient origin of all *Alyssum*

serpyllifolium populations, which likely precedes emergence of metal hyperaccumulation in the model genera *Noccaea* and *Arabidopsis*. However, the timing in absolute terms at this stage does not have a high degree of certainty, given that the estimates are based on a secondary calibration from a single phylogeny of the genus *Alyssum* genus inferred from a handful of chloroplast loci originally calibrated using three verified Brassicales fossils and a molecular-phylogenetic tree of this order (Beilstein *et al.*, 2010). Therefore, *Alyssum serpyllifolium* divergence times presented here will reflect and amplify the distortions from true divergence times introduced at an earlier stage by secondary calibration. Population

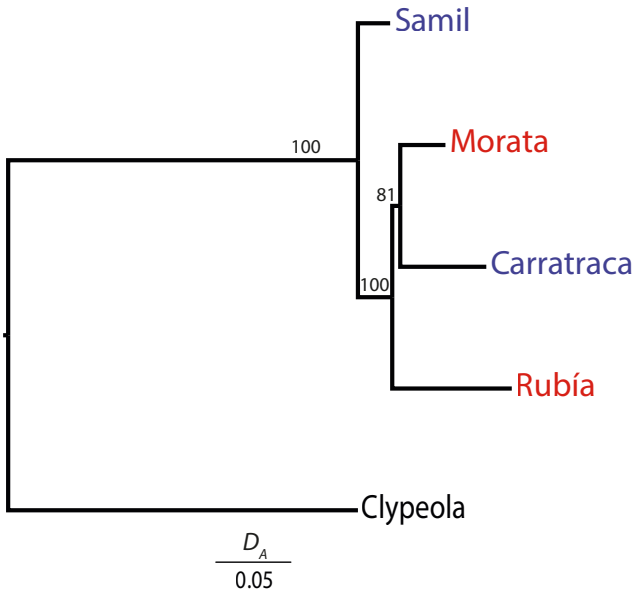


Figure 3.2: *A. serpyllifolium* population tree based on Nei's (1983) D_A calculated over 1015 unlinked SNPs and rooted on *Clypeola*. Branch values represent percentage support over 1000 bootstrap replicates. Serpentine populations are marked in blue and non-serpentine populations in red.

tree topologies based on D_A (Nei *et al.*, 1983) net sequence divergence per site (Figure 3.2) and the neighbour-joining algorithm (data not shown, Saitou & Nei, 1987) were not unlike the *BEAST analysis results, with the exception of the position of Rubía-NS, which now shares direct common descent with Carratraca-S and Morata-NS, and not Samil-S. To test for any deviation from strictly bifurcating phylogenies as indicated by slight incongruence between species tree and populations tree, two methods were applied: Treemix to detect the presence of gene flow, and NeighborNet network to reveal incomplete lineage sorting or gene flow. According to the AIC, the model with two unexpected moderate strength

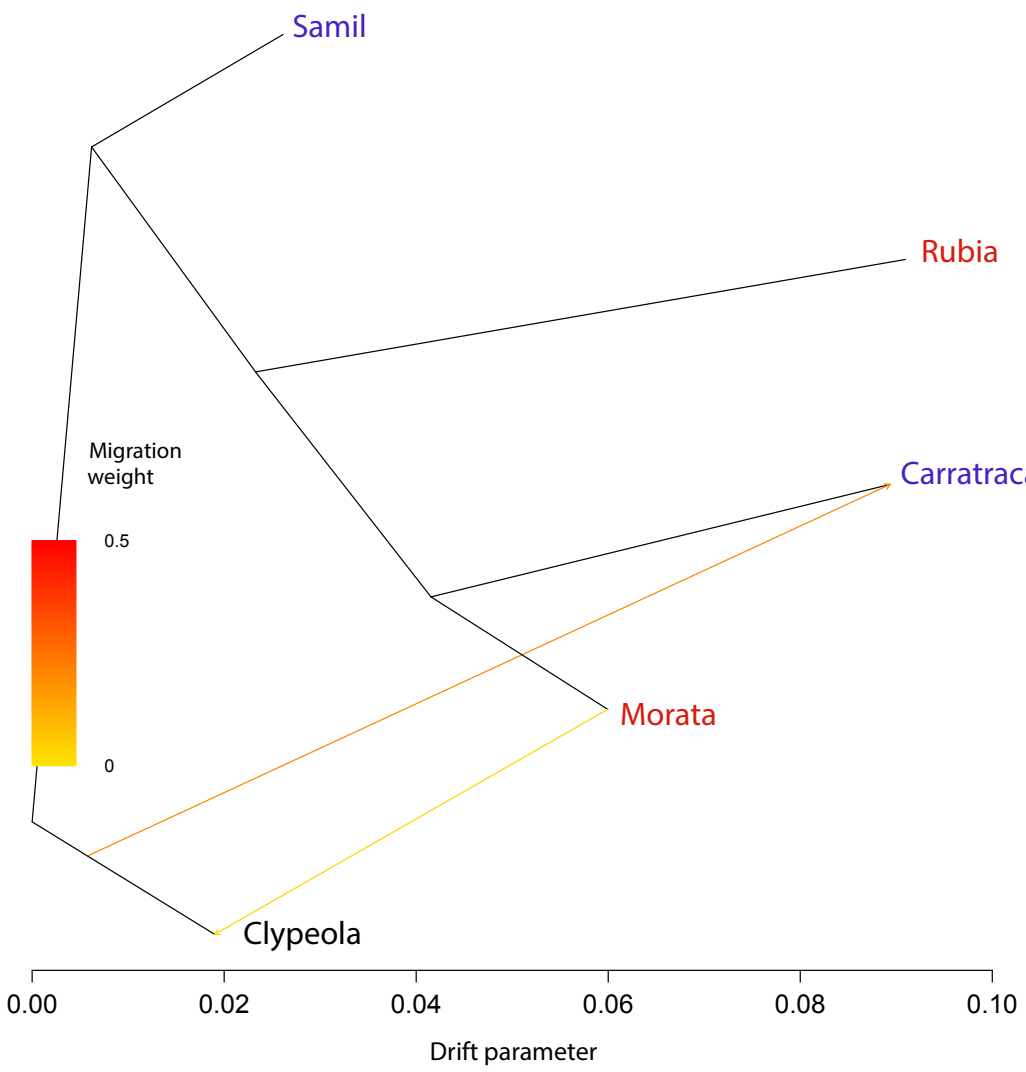


Figure 3.3: Treemix ancestry graph representing the relationships between *Alyssum serpyllifolium* populations and the most likely number of past admixture events (two), which are coloured according to their weight. The phylogeny is based on 13,050 unlinked SNPs. Serpentine populations are marked in blue and non-serpentine populations in red.

migration bands from *Clypeola* to Carratraca-S, and low intensity from Morata-NS to *Clypeola* was the best of the ones tested in Treemix (Figure 3.3), which is quite unexpected given the long evolutionary distance separating the species and lack in overlap of their native ranges, and so could be an overestimate due to false explanation of a pattern of allele frequency differences resulting from processes other than hybridization (Yu *et al.*, 2011). Furthermore, results for alternative models with a different number of migration events were close in terms of their AIC values: -191.1 for two migration events, through

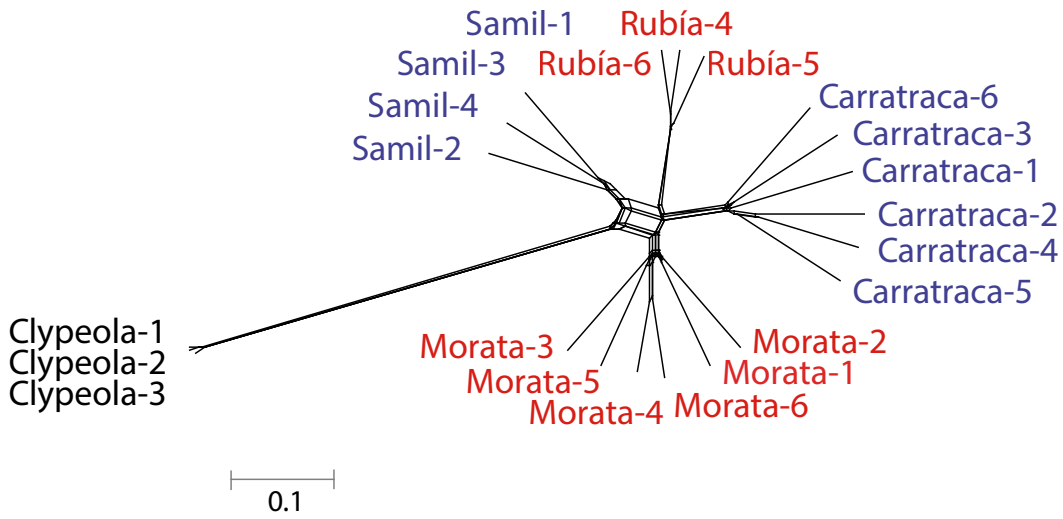


Figure 3.4: Consensus SplitsTree NeighborNet network of *Alyssum* individuals based on 13,050 unlinked SNPs, rooted on *Clypeola*. Serpentine populations are marked in blue and non-serpentine populations in red. All inter-population splits had 100% bootstrap support (not shown).

-189.2 to -182.5 for between 1 and 6 migration events, and -167.9 for 0 migration events. The graph topology is in line with results from the two previously described analyses, with Samil-S splitting first followed by Rubía-NS, and finally Carrataca-S as sister to Morata-NS. Morata-NS and Samil-S branches were quite short compared to Rubía-NS and Carratraca-S, but no such differences in the rate of evolution between populations are supported by other phylogenetic analyses.

Lastly, the NeighborNet algorithm in the SplitsTree software package was employed to uncover and visualise any possible reticulation events (Figure 3.4). The overall topology of the network emphasises the complex pattern of shared polymorphisms between the serpentine and non-serpentine populations as well as relatively high genetic diversity contained within *Alyssum serpyllifolium* individuals, judged by long branching leading to individuals in population sub-clades. Again, the serpentine and non-serpentine populations do not form two separate sub-clades, which is suggestive of equally long shared evolutionary histories between serpentine and non-serpentine populations.

Table 3.2: Species delimitation results from SNAPP Bayes' Factor Delimitation analysis. Bayes' Factor compares each model to the null model with 4 delimited species.

Model	MLE	Rank	BF
4 species	-13172.5	1	N/A
3 species A	-13672.2	2	-999.4
3 species B	-13710.1	3	-1075.3
2 species A	-14110.9	4	-1876.8
2 species B	-14215.6	5	-2086.3

MLE - Maximum Likelihood Estimation; BF - Bayes' Factor

3.3.2 Species delimitation

To test if the levels of genetic diversity observed between populations merited their designation as independently evolving lineages, and perhaps even raising them to the species level, three different molecular methods of species delimitation were used. In view of the conflicting results, species boundaries within the *Alyssum serpyllifolium* species complex continue to remain fuzzy. Overall, two extreme outcomes were obtained from species delimitation analyses: lumping all populations into a single species (STACEY) or retaining all 4 populations as putative species (BP&P, SNAPP). Different simplifications assumed by the models of species delimitation in the methods tested are likely responsible for the disparate outcomes (Carstens *et al.*, 2013**b**). Inclusion of more populations and outgroup species in the analyses to gain more information about relative levels of polymorphism and divergence between different lineages could be helpful in obtaining more conclusive species delimitations in the complex. It has to be emphasised that the results reported below are based simply on nucleotide sequence data, whereas it has been suggested that molecular species delimitation methods could benefit from explicit inclusion of classical taxonomic information about the species as a prior in order to arrive at comprehensive species delimitation outcomes (Jones, 2015). Nevertheless, it is not uncommon for molecular sequence information to show incongruence with species classification based on morphological traits, as in the case of three subspecies of *Cistus ladanifer*, which occurs in similar habitats (Quintela-Sabarís *et al.*, 2010).

3.3.2.1 BP&P

In total, 24 different replicate runs were carried out in BP&P, which were monitored for rjMCMC acceptance rates to remain within the range of 0.15–0.70 to avoid the chain becoming stuck in a non-optimal state. Consistent results were obtained across runs with high ESS (>500), always pointing to 4 species delimitation with perfect support (posterior probability = 1) and always recovering the consensus topology of ((Rubía-NS, Samil-S) (Carratraca-S, Morata-NS)) as the most probable species tree. The results are rather surprising given the fact that BP&P tends to be a species lumper rather than splitter, as its model does not take account of the homogenising effect of gene flow post-speciation, and the retained species splits are generally conservative (Yang & Rannala, 2014). Misspecification of the guide tree and reassignment of individuals to populations can cause errors (Zhang *et al.*, 2014**a**), but neither bias is likely to be present here, as the same results have been obtained in unguided runs, and assignment of individuals to populations was validated with Structure and PCA analysis.

3.3.2.2 Bayes' Factor Delimitation (BFD) with SNAPP

Apparent stationarity and large effective sample sizes (ESS) (>200) were achieved in the BFD analysis. According to Kass & Raftery (1995), Bayes Factor values above 10 offer decisive support in favour of a given hypothetical species delimitation model, and so in all cases the *4 species* null model was very strongly preferred (Table 3.2).

3.3.2.3 STACEY

Across 10 replicate STACEY runs, uniform results were obtained with >99.9% posterior probabilities, supporting a single cluster encompassing all *Alyssum* populations. Simulations conducted by the authors indicate that the method is conservative (Jones, 2015; Jones et al., 2015), with all the errors amounting to false species lumping, as in the case of BP&P.

3.3.3 EST-SSR detection

Overall, 3,942 di- to hexa-nucleotide EST-SSRs were discovered in the consensus *Alyssum* assembly, with tri-nucleotide SSRs (2,906, or 73.7% of the total number; top motif: AAG)

as the most abundant group, followed by di-nucleotide SSRs (752, or 19.1% of total number; top motif: AG). Average EST-SSR frequency was 1 in 10.5 kbp and average SSR length was approx. 17.9 bp. The prevalence of tri-nucleotide repeats has been found in EST-derived SSRs in other plant species, e.g. in various legumes (Garg *et al.*, 2011; Kaur *et al.*, 2011, 2012; Jhanwar *et al.*, 2012; Zhang *et al.*, 2012) and cereals (Varshney *et al.*, 2002; La Rota *et al.*, 2005), *Scabiosa columbaria* (Angeloni *et al.*, 2011), *Ammopiptanthus mongolicus* (Zhou *et al.*, 2012), *Artemisia tridentata* (Bajgain *et al.*, 2011) and *Ipomoea batatas* (Tao *et al.*, 2012). The prevalence of trimeric EST-SSRs has been explained as stemming from the pressure to maintain an unchanged reading frame in protein coding sequences (Metzgar *et al.*, 2000). On the other hand, estimates of total SSR frequency vary considerably according to detection parameters, e.g. from 1 in 0.74 kbp to 1 in 17.42 kbp for wheat EST-SSRs (Varshney *et al.*, 2002; Kaur *et al.*, 2011), and so are not very informative. The top *Alyssum* EST-SSR motifs AG and AAG have frequently been reported as the most abundant motifs in plants in general (Lawson & Zhang, 2006; Wei *et al.*, 2011; Zhou *et al.*, 2012); they occur preferentially in UTRs and were shown to be involved in transcriptional regulation (Zhang *et al.*, 2006).

Out of 3,942 microsatellites detected in the *Alyssum* assembly, 32 were selected based on repeat size (3 or 6 nucleotides, i.e. no introduction of frame-shift mutation), as well as the function of the target gene, and lack of differential expression pattern, as well as outlier SNPs associated with it (i.e. putatively neutral gene, not likely to be important for adaptation to serpentine). Primers were then designed around each locus, and amplification success along with level of polymorphism was tested in 3-4 individuals in each population. Four serpentine populations (Barazón-S, Carratraca-S, Samil-S, Sierra Bermeja-S) and four non-serpentine populations (Alhaurín-NS, León-NS, Morata-NS, Rubía-NS: Table 3.1) were eventually genotyped across eight selected microsatellite loci with mode 5 alleles found across 4 microsatellite loci and a range of 4-12 alleles (Tables 3.3 and 3.4). In total, 272 individuals were genotyped in the eight populations, and low amounts of missing data were recorded; between 250 and 266 individuals were scored for each of the eight microsatellite loci, all unbiased towards any population. As no evidence of genotyping artefacts was

found, and all loci are in linkage equilibrium, results from all eight microsatellite loci could be used in the subsequent analyses.

Table 3.3: Description of genes containing EST-SSRs used in the study of *Alyssum serpyllifolium* populations

Fragment ID	TAIR locus	Gene name and function
AP31733	AT2G46830	*CIRCADIAN CLOCK ASSOCIATED 1*, transcriptional repressor of *TOC1* involved in circadian rhythm regulation
AP32282	AT5G22090	Protein of unknown function (DUF3049)
AP31679	AT2G33730	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AP801	AT3G56590	Hydroxyproline-rich glycoprotein family protein
AP31640	AT1G29350	Kinase-related protein of unknown function (DUF1296)
AP34461	AT1G49975	Protein containing photosystem I reaction centre subunit N domain
AP10368	AT3G01780	*TPLATE*, cytokinesis protein participating in vesicle-trafficking important for pollen germination
AP5386	AT2G25180	*RESPONSE REGULATOR 12*, an *Arabidopsis* response regulator acting in cytokinin signalling pathway

Table 3.4: Summary of *Alyssum serpyllifolium* microsatellite loci used in this study

Fragment ID	TAIR locus	Repeat size	Fragment size range	Allele number	Forward primer 5′→3′	Reverse primer 5′→3′
AP31733	AT2G46830	3	187-202	5	CCGATTCCCAAAGATCCCGTG	GGTATTGCCGCAAAGTTTTTCA
AP32282	AT5G22090	3	294-315	5	TCCTCCTCACTTTCGCTGAA	GACGGTTGATGGCGGTTTTG
AP31679	AT2G33730	6	363-399	12	TCCTCACCAAAGCTCAGCG	CTTAGCCTCCTCCTCCCTCT
AP801	AT3G56590	6	208-229	5	TGGAGGTGGGATATGAGCAAA	CGAGCAGAGGAGACCAAAGA
AP31640	AT1G29350	3	253-292	10	CGAAAACTCTGCGATGTGGC	CGTCTCTTGGGTTTTGCTGC
AP34461	AT1G49975	3	147-168	7	AAGGGAAAGTCAGAAGCAGAGC	AGTTCTTCAAAGTTTCATAGACAACA
AP10368	AT3G01780	3	343-355	4	TCTAACTGACGGAGGGGTTG	TGCCATTCTTGAAGCACTGC
AP5386	AT2G25180	3	373-385	5	GGGATTCTTGGTCGGCTCAA	CTGATGGAGATGCTTGTGGGT

Table 3.5: Average F_{IS} (Weir & Cockerham, 1984), observed (H_o) and expected (H_e) heterozygosities with standard deviation (SD) in the microsatellite-genotyped individuals (n) in the eight *Alyssum serpyllifolium* populations. F_{IS} values deviating significantly from 0 are labelled with an asterisk (*).

Population	n	F_{IS}	H_o	SD	H_e	SD
Alhaurín-NS	36	0.35*	0.33	0.21	0.51	0.22
Barazón-S	35	0.07	0.40	0.27	0.43	0.25
Carratraca-S	33	0.27*	0.48	0.24	0.65	0.15
León-NS	22	0.22*	0.51	0.26	0.65	0.23
Morata-NS	30	0.22*	0.55	0.14	0.70	0.09
Rubía-NS	41	0.24*	0.44	0.21	0.58	0.13
Samil-S	36	0.16*	0.62	0.14	0.73	0.11
Sierra Bermeja-S	36	0.08	0.51	0.20	0.55	0.18

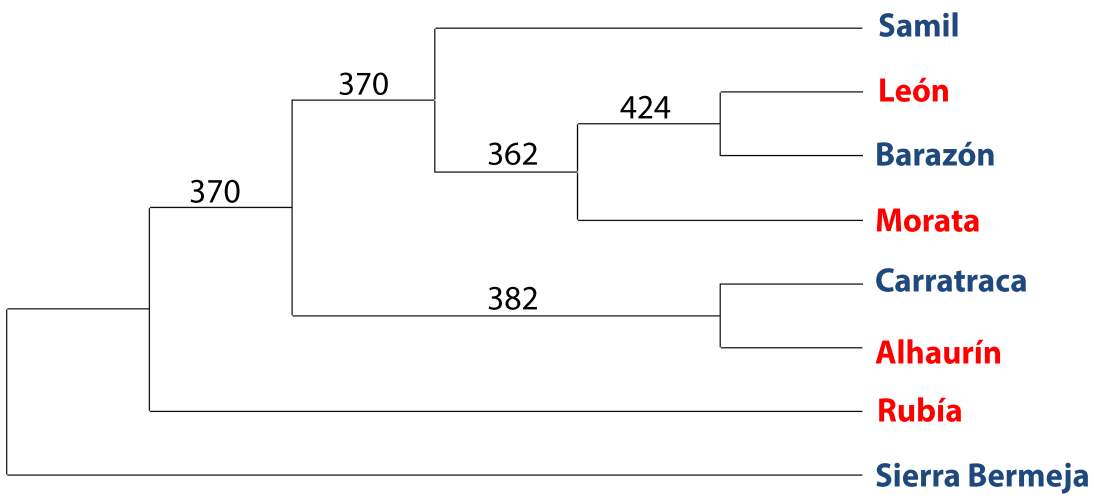


Figure 3.5: Extended majority rule consensus neighbour-joining tree based on Nei's (1972) D_{ST}. Values on the branches represent bootstrap support over 1000 replicates. Serpentine populations are marked in blue and non-serpentine in red.

Table 3.6: Averaged pairwise F_{ST} (Weir & Cockerham, 1984) between the eight *Alyssum serpyllifolium* populations according to the results for eight microsatellite loci

	Barazón-S	**Rubía-NS**	**Samil-S**	**Sierra Bermeja-S**	**Morata-NS**	**Carratraca-S**	**Alhaurín-NS**	**León-NS**
Barazón-S	0							
Rubía-NS	0.402	0						
Samil-S	0.231	0.246	0					
Sierra Bermeja-S	0.324	0.288	0.237	0				
Morata-NS	0.201	0.226	0.082	0.175	0			
Carratraca-S	0.288	0.248	0.133	0.159	0.095	0		
Alhaurín-NS	0.410	0.317	0.176	0.274	0.121	0.131	0	
León-NS	0.220	0.226	0.146	0.258	0.063	0.141	0.225	0

3.3.4 Levels of genetic diversity in S and NS populations of *A. serpyllifolium*

The results show high levels of population differentiation, with overall F_{ST} = 0.23 (95% CI 0.17 – 0.29), and echo the genome-wide estimates of high differentiation obtained for four select populations through the RNA-Seq analyses. Moderate-to-high values of F_{ST} in *A. serpyllifolium* are typical of values from other taxa endemic to specific substrates, such as *Primulina tabacum* (0.3936: Ni *et al.*, 2006), *Jurinea pinnata* (0.374: Salmerón-Sánchez *et al.*, 2014**a**) or in *Convolvulus boisseri* (0.395: Salmerón-Sánchez *et al.*, 2014**b**). An even higher F_{ST} range was encountered between Western Swiss populations of the model hyperaccumulator *N. caerulescens*, where F_{ST} reached an average of 0.591 (Besnard *et al.*, 2009). Values in that range provide evidence for moderate gene flow between the populations of *Alyssum serpyllfolium* probably determined by ecological (serpentine "islands" of endemism) and topographical barriers, such as mountains, resulting in fragmentation of the species range.

The average within-population gene diversity for microsatellite markers (H_O = 0.48) and heterozygote deficiency are typical of endemic species with small population sizes and disjunct distribution, as in the example of *Convolvulus boissieri* (Salmerón-Sánchez *et al.*, 2014**b**). F_{IS} values in the range of 0.07 – 0.35, and significantly different from 0 over 1000 permutations in all populations except for Barazón-S and Sierra Bermeja-S (Table 3.5), point to a degree of inbreeding. Significantly positive F_{IS} values for the four populations obtained through SNP genotyping from RNA-Seq further support the conclusion of partial inbreeding, which likely stems from population isolation and small effective sizes, but endogenous barriers to gene flow cannot be excluded until research is performed into the mating system of this taxon. Similar F_{IS} values were obtained in microsatellite surveys of metallicolous and non-metallicolous populations of *N. caerulescens*, a species with a mixed mating system (Mousset *et al.*, 2016). Nevertheless, these results are in line with expectations for serpentine endemics, reflecting small effective population sizes and small distribution ranges, and as a consequence low genetic diversity and high levels of inbreeding (Anacker *et al.*, 2011).

Previous studies of the population genetics of species harbouring metallicolous and non-metallicolous populations have often speculated that, given the scenario of local evolution of metallicolous populations from more tolerant, low-frequency genotypes in non-metallicolous populations, a founder effect may have been present in metallicolous populations when they originally diverged from non-metallicolous populations (Pauwels *et al.*, 2005), resulting in lower genetic diversity in the metallicolous populations. However, no differences in genetic diversity between metallicolous and non-metallicolous populations have been consistently found in the species investigated to date (Vekemans & Lefèbvre, 1997; Quintela-Sabarís *et al.*, 2010), such as *Arabidopsis halleri* (Pauwels *et al.*, 2005), *Alyssum bertolonii* (Mengoni *et al.*, 2003) and *Minuartia laricifolia* ssp. *ophiolitica* (Moore *et al.*, 2013); neither have they been found here (Table 3.5). However, the bottlenecks in question would have to have been quite recent to stand out in this way, and gene flow and accumulation of new mutations would later erode the signal of any putative bottlenecks (Vekemans & Lefèbvre, 1997).

AMOVA results indicated that 23.7% (Φ_{ST} = 0.24) of all genetic variation was partitioned between populations, while 16.7% (Φ_{IS} = 0.21) was partitioned within individuals within populations. Similar between-population partitioning of genetic variation (22%), was found in an ISSR (Inter-Simple Sequence Repeat) study of four populations of the nickel hyperaccumulator *Alyssum lesbiacum* (Adamidis *et al.*, 2014). In contrast, another *Alyssum* hyperaccumulator, *A. bertolonii*, had over a half (51%) of its genetic variation partitioned between populations in a study of 9 populations genotyped with chloroplast SSRs (Mengoni *et al.*, 2003). Between-population variation found in a chloroplast RFLP (Restriction Fragment Length Polymorphism)-based study of 28 *Arabidopsis halleri* populations (Pauwels *et al.*, 2005) was even higher and amounted to 68%. These examples support the conclusion that levels of between-population genetic variation in the *Alyssum serpyllifolium* species complex are far from unusually high or low among those found in metal hyperaccumulator species, indicating that the populations of *Alyssum serpyllifolium* have not been subject to extremely strong vicariant processes, unless the results are severely biased by homoplasy of microsatellite alleles (Pannell & Fields, 2014).

3.3.5 Population structure determination with EST-SSR

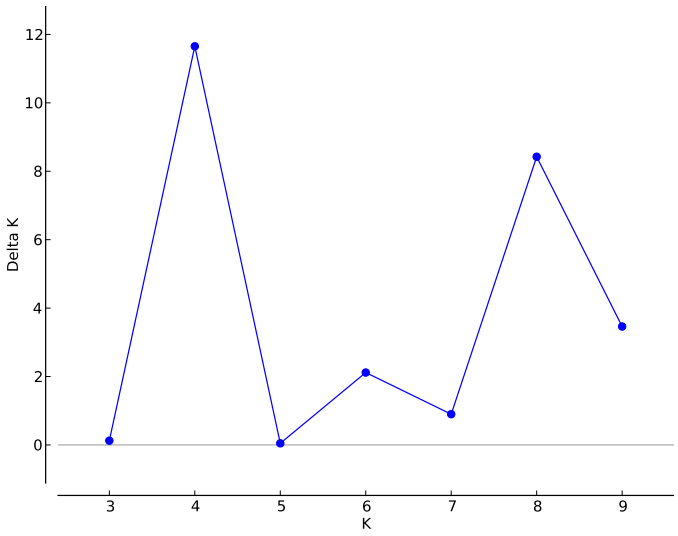


Figure 3.6: Probability of different numbers of true populations (K) identified using EST-SSR markers in the Structure analysis, as calculated by ΔK (Evanno *et al.*, 2005)

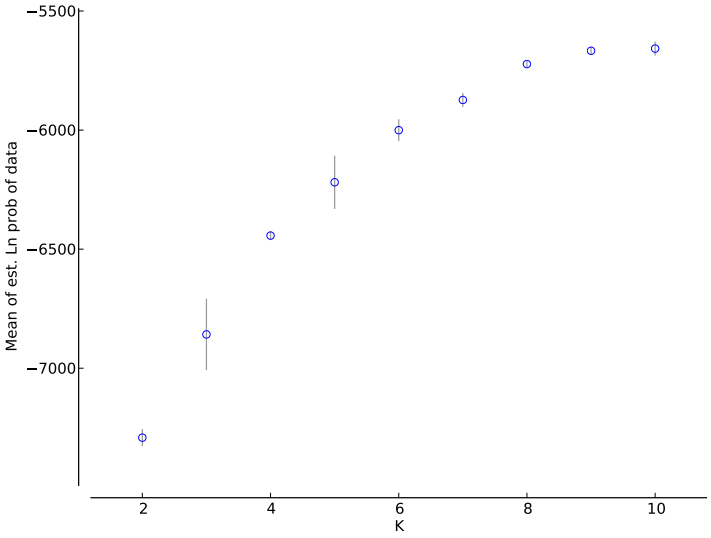


Figure 3.7: EST-SSR marker data fit in the Structure analysis (LnP(D)) to different numbers of hypothetical populations (K)

Population-genetic structure analysis with Structure which detects fairly recent genetic isolation (Yang *et al.*, 2010**b**), reveals two optimal numbers of genetic clusters (K): $K = 4$

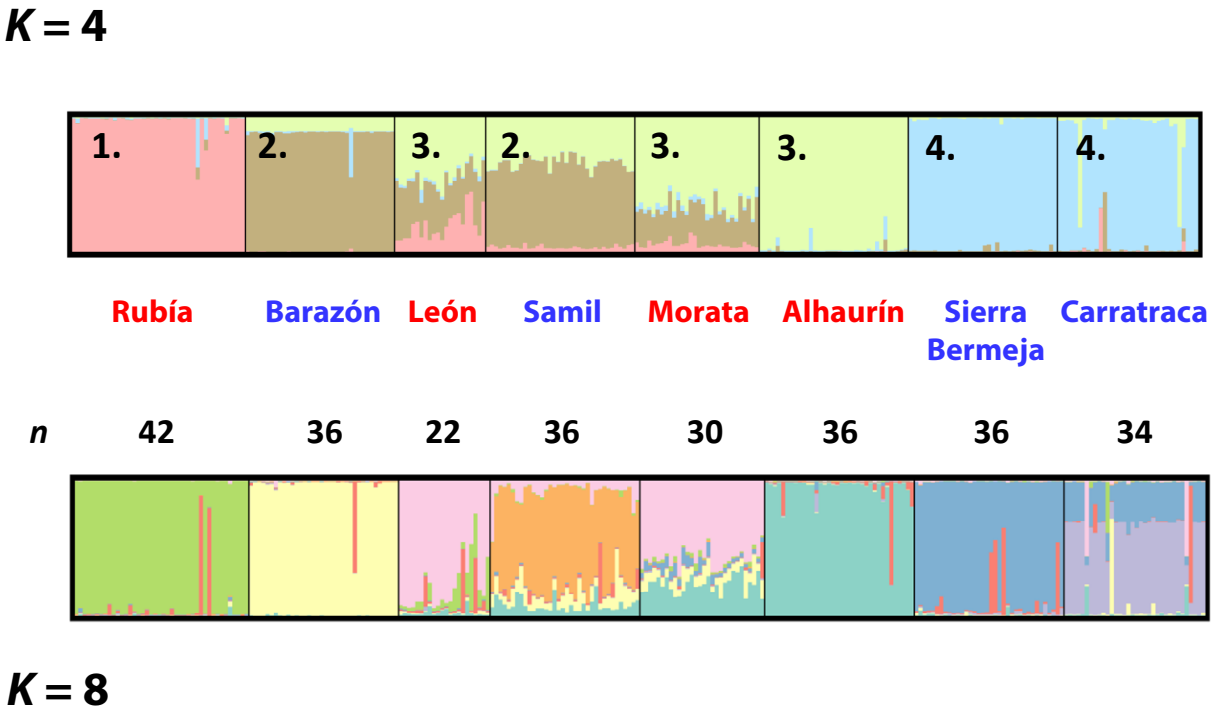


Figure 3.8: Assignment of genotypes of individuals in the eight *Alyssum serpyllifolium* populations to either 4 (*top*) or 8 (*bottom*) genetic clusters as defined to be most likely by Structure. Serpentine populations are marked in blue and non-serpentine in red. *n* - number of sampled individuals in each population

(as per ΔK, Figure 3.6) and $K = 8$ (as per LnP(D), Figure 3.7). $K = 4$ results show a trend towards grouping together geographically close populations of the same ecotype, while $K = 8$ simply place each population in its own separate cluster (Figure 3.8). Admixture proportions of individuals in different populations varied in the $K = 4$ solution. Individuals in non-serpentine populations (northern Rubía-NS and southern Alhaurín-NS) belonged to a single cluster each (pink and green clusters, respectively), and so did both *Alyssum malacitanum* populations in the south (Sierra Bermeja-S and Carratraca-S: blue cluster). The other four populations showed varying degrees of admixture. The northern '*Alyssum guitanii*' population (Barazón-S) and *Alyssum pintodasilvae* population (Samil-S) mostly belonged to the fourth cluster (brown), with more (Samil-S) or less (Barazón-S) admixture from Alhaurín-NS's green cluster. The northern and central non-serpentine populations Morata-NS and León-NS contained the highest admixture from Alhaurín-NS's green cluster, in addition to smaller admixture from Samil-S's and Barazón-S's brown cluster and

Rubía-NS's pink cluster (only in León-NS). The two most optimal *K* solutions are likely to reflect the complex levels of organisation in *Alyssum* population structure (Meirmans, 2015), with $K = 4$ identifying non-hierarchical relationships resulting from the interplay of geographic distance and edaphic factors, as evidenced by the four serpentine populations being evenly split into a northern (Samil-S, Barazón-S) and a southern (Carratraca-S, Sierra Bermeja-S) groups corresponding to species circumscriptions of *Alyssum pintodasilvae* and *Alyssum malacitanum*, accordingly, and isolation of nearby non-serpentine populations (Rubía-NS in the north and Alhaurín-NS in the south).

Consequently, to investigate if the pattern of isolation-by-distance (IBD) could be detected over all populations, a Mantel test was carried out on Weir and Cockerham's F_{ST} pairwise distances matrix (Table 3.6). This did not reveal any direct correlation, and the number of populations within each ecotype was too small to allow testing for IBD within them. No consensus neighbour-joining tree could be obtained with reliable branch support values based on Nei's (1972) D_{ST} distance metric (Figure 3.5) and no population structure could be detected using factor analysis in GENETIX (data not shown). Taken together, it was not possible to resolve the relationships between *Alyssum serpyllifolium* populations with only eight microsatellite loci. Further research is required with a higher number of genetic markers for increased power, as SNP-based results from RNA-Seq were sufficiently densely sampled to deliver high confidence population phylogenies (see section 3.3.1).

In general, previous research on pseudometallophytes has revealed isolation-by-distance irrespective of the edaphic type of population sites (Mayer *et al.*, 1994; Mengoni *et al.*, 2001; Berglund *et al.*, 2003; Pauwels *et al.*, 2005; Yokoo *et al.*, 2009). A study of different populations of *Arabidopsis halleri* using chloroplast RFLP markers (Pauwels *et al.*, 2005) found a strong pattern of isolation-by-distance: chlorotypes from metallicolous populations were more closely related to chlorotypes from parapatric non-metallicolous populations than allopatric metallicolous populations which suggested that they had been founded independently from nearby non-metallicolous populations. In *Jurinea pinnata*, whose populations occur on either gypsum or dolomite substrates, genetic differentiation

between population correlates with geographic distances and not the edaphic factor (Salmerón-Sánchez *et al.*, 2014**a**).

3.3.5.1 Polyploidy

None of the *Alyssum serpyllifolium* populations from the Iberian Peninsula showed any evidence of polyploidy when analysing microsatellite genotyping data, e.g. detection of more than two fragment sizes at a given microsatellite locus that was clearly not a technical stutter artefact. This was expected, given previous reports that had consistently pointed to their diploidy (see Chapter 1). Unfortunately, the karyotypes of the three French populations, for which material was collected in the field for the first time (Anduze, Bédarieux, Malaucène: Table 3.1), appeared to be polyploid based on the genotyping results of all the highly polymorphic microsatellite loci (data not shown). Population genetic analyses of polyploids is notoriously difficult, not only because it is almost never possible to establish the true genotypes, but also because few methods exist that allow co-analysis of populations of different ploidy levels (Dufresne *et al.*, 2014). For that reason, these three putatively polyploid populations were dropped from the sample of *Alyssum serpyllifolium* populations analysed.

While plant polyploidy in certain cases does seem to confer advantage in environments with extreme ion imbalance (Madlung, 2013) and can facilitate metal ion accumulation, as in the case of root autopolyploidy in natural populations of *Arabidopsis thaliana* (Chao *et al.*, 2013), polyploidy has so far not been associated with adaptation to serpentine environment in species with multiple ploidy levels (Kolář *et al.*, 2013), and this factor does not seem to play a role in *Alyssum serpyllifolium* either, given that French polyploid populations inhabit non-serpentine soils (Mengoni *et al.*, 2003).

3.3.6 Source of adaptive variation

In order to test if the shared outlier candidate genes for adaptation to nickel hyper-accumulation and serpentine environment tolerance were descended from the same common ancestor, or rather arose independently by convergence as *de novo* mutations,

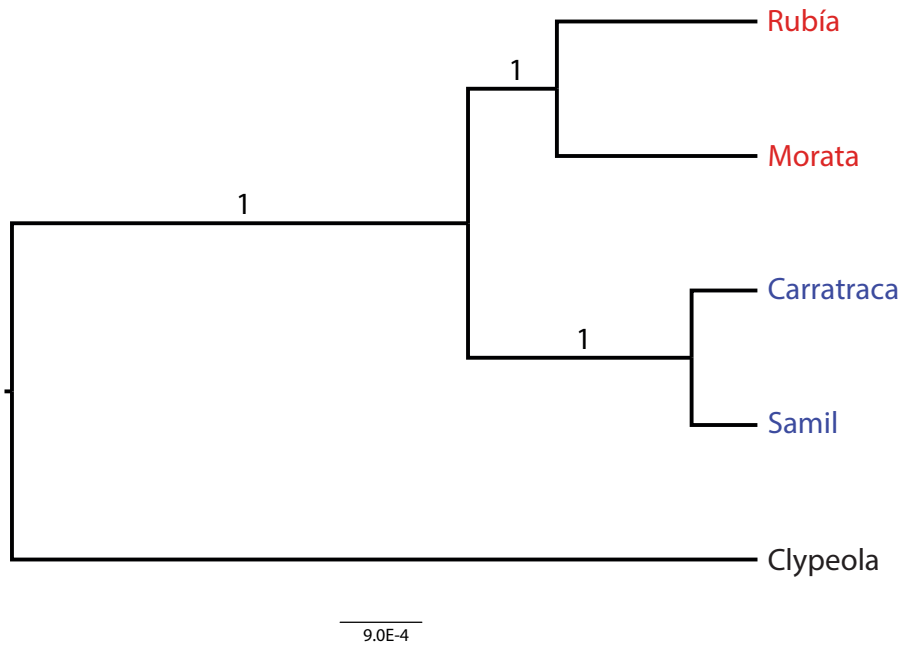


Figure 3.9: *Clypeola*-rooted *BEAST maximum clade credibility tree based on 11 *Alyssum serpyllifolium* outlier genes involved in metal homeostasis. Posterior probabilities are given at each branch. Serpentine populations are marked in blue and non-serpentine populations in red.

two additional multi-gene phylogenies were generated in *BEAST based on the RNA-Seq data. Firstly, a phylogeny based on 11 candidate outlier genes involved in metal homeostasis (*ATX1*, *AKT1*, *CAX1*, *IREG1*, *FRD3*, *OPT3*, *KUP9*, *ZIF1*, *ZIFL1*, *NRAMP4*, *CAT4*) was reconstructed to test if this differed from the species tree based on neutral markers and genes. A directly opposite outcome was found to the two subclades defined by neutral genes: the two serpentine populations Carratraca-S and Samil-S were clustered in one subclade and the two non-serpentine populations Morata-NS and Rubía-NS in the other, with branch lengths significantly shorter in the serpentine subclade (Figure 3.9). This striking difference demonstrates that common genes putatively important for hyperaccumulation or serpentine adaptation largely share closely related haplotypes, which could have arisen either through intogression or through parallel selection from standing genetic variation. An identical tree topology with the same branch support values (data not shown) was obtained when analysing the sequences of 13 non-metal homeostasis-related candidate genes for adaptation to serpentine/hyperaccumulation (*CXXS1*, AT1G47330 -

a CBS-DUF21 gene, *ATP-PRT1*, *SYTA*, a SGNH hydrolase-type esterase, *SCAR2*, *MLO12*, *PGP6*, *PLC2*, *MRP3*, *HDA3*, *PHT3*, *ICU2*).

3.3.7 Sanger-based resequencing of candidate loci

Identifying a set of candidate outlier loci under selection using a limited sampling of individuals and populations is only the first step towards confirming their importance. The next step is to extend the sampling in a larger survey of polymorphism in the loci to establish patterns of haplotype variation encountered in populations from across replicate environments (De Wit *et al.*, 2015). In the current study, four candidate genes - *ATP-PRT1*, *CAT4*, *IREG1* and *NRAMP4* were tested, out of which two - *IREG1* and *NRAMP4* - were finally selected due to reproducible sequencing results obtained across different *Alyssum serpyllifolium* populations.

Putatively neutral loci for comparison were selected based on three factors: their absence as putative targets of selection in the results of all tests applied in Chapter 2, their function, and levels of nucleotide diversity (π) and divergence (D_{xy}) in line with that of candidate loci. Amplification of three loci was tested, namely *ASIL1*, *ATS9* and *ERH3*. Out of the three tested neutral genes, *ASIL1* was selected because of a higher rate of successful amplification and sequencing than *ERH3* and higher sequence diversity than found in *ATS9*.

Primers were designed to amplify *IREG1* and *NRAMP4* gene fragments containing the highest possible number of outlier SNPs. All of 10 *NRAMP4* outlier SNPs were genotyped in this way, and in *IREG1* just one Carratraca-S outlier SNP was missing (which was over 1 kbp further downstream from the preceding outlier SNP). For both genes, amplification primers were also used as sequencing primers (Table 3.7), but in the neutral gene *ASIL1* a different reverse primer was used for amplification (ASIL1_R1) and sequencing (ASIL1_INTF1) due to noisy results from sequencing using the default amplification primer.

Sequenced regions in both *IREG1* and *NRAMP4* were found to contain putative introns: 82 bp long in *IREG1* and 145 bp long in *NRAMP4* (Table 3.8), which were eliminated from

the analyses described below, unless otherwise stated. The *NRAMP4* gene fragment was sequenced from both ends, resulting in coverage of positions 218-640 and 896-1507 in the original contig (Table 3.7), as a microsatellite present in the non-covered region resulted in sequence phase changes. However, cloning was not pursued to resolve the problem, as the missing region does not contain any outlier SNPs.

IREG1 sequencing did not proceed according to expectations, despite designing the primers specifically according to the conserved contig sequence, which was based on abundant sequencing reads and in full structural homology to the reference *IREG1* gene from *Arabidopsis thaliana*. Despite these precautions and PCR amplification resulting in a clean single product in each individual, the sequenced fragment was missing a region approximating to the full first exon, which is conserved across the Brassicaceae, moreover a premature stop codon was found in the 5′ region. This suggests that a non-functional pseudogene copy of the original *IREG1* was unexpectedly Sanger-sequenced. Nevertheless, the results of *IREG1* sequencing are reported here anyway, as the gene copy sequenced does not appear degraded beyond what was mentioned, and it can offer insight into the evolution of the original *IREG1*.

Table 3.7: Overview of the primers used in the amplification and sequencing of two putative hyperaccumulator genes (*IREG1*, *NRAMP4*) and the neutral *ASIL1* gene in eight *A. serpyllifolium* populations

Gene	Primer ID	Sequenced contig region	Annealing temperature	Sequence (5′→3′)
ASIL1	ASIL1_SEQF1	220-954	50°C	CAAAAAAAAAATCCCCWAATCTCAG
	ASIL1_R1			TTTTCGGATTCGAGTCTCAAATTG
	ASIL1_INTF1		55°C	GGTAATGGTGATGGTCCTAG
IREG1	IREG1(4)F	92-841	45°C	ATAGTAAATAAAAACATATTTATAG
	IREG1(4)R			CACTACGATACACATATATG
NRAMP4	NRAMP4_F3	218-640	55°C	ATAGTTGGAGTAGACGAAGAAGAAG
	NRAMP4_R3	896-1507		TCTTCTTGAAAGGTCTCTTTCC

Table 3.8: Comparison of sequence diversity statistics across the three loci (*ASIL1*, *IREG1*, *NRAMP4*) resequenced in the eight *Alyssum serpyllifolium* populations, calculated over nonsynonymous (NS) or synonymous (S) sites.

	ASIL1	***IREG1***	***NRAMP4***
coding sequence length	735	586	888
intron sequence length	N/A	82	145
number of segregating sites	30	24	33
number of segregating sites (NS)	17	11	11
number of segregating sites (S)	13	10	14
number of singletons	2	6	3
haplotype (gene diversity) ± SD	0.965 ± 0.009	0.834 ± 0.020	0.809 ± 0.026
π_{NS} **(%)**	0.318	0.333	0.348
π_S (%)	1.232	1.174	1.341
π_{NS} / π_S	0.258	0.284	0.260
θ_W **from S (per site) ± SD**	0.00739 ± 0.00212	0.00629 ± 0.00186	0.00532 ± 0.00023

Table 3.9: Detailed population-by-population comparison of basic sequence diversity statistics in resequenced *ASIL1*, *IREG1*, *NRAMP4* gene fragments.

	Alhaurín-NS	Barazón-S	Carratraca-S	León-NS	Morata-NS	Rubía-NS	Samil-S	Sierra Bermeja-S
ASIL1								
n	10	10	13	9	10	5	8	6
haplotype (gene diversity)	0.695	0.189	0.963	0.993	0.968	0.800	0.958	0.894
π_{NS} (%)	0.301	0.033	0.047	0.253	0.139	0.253	0.175	0.282
π_S (%)	0.287	0	1.562	1.449	1.315	0.349	0.651	0.541
θ_W from S (per site)	0.0023	0.0004	0.0029	0.0036	0.0031	0.0024	0.0029	0.0027
IREG1								
n	12	10	12	10	11	11	10	10
haplotype (gene diversity)	0.772	0	0.236	0.363	0.005	0.173	0.958	0.911
π_{NS} (%)	0.055	0	0.019	0.112	0.079	0	0.244	0.142
π_S (%)	0.601	0	0.069	0.044	0.525	0.076	0.846	1.050
θ_W from S (per site)	0.0028	0	0.0008	0.0021	0.0025	0.0004	0.0034	0.0030
NRAMP4								
n	10	6	11	2	10	7	8	11
haplotype (gene diversity)	0.484	0.848	0.574	-	0.233	0.615	0.400	0.363
π_{NS} (%)	0.024	0.122	0.024	-	0.030	0.056	0	0
π_S (%)	0.084	0.314	0.269	-	0	0.067	0.105	0.102
θ_W from S (per site)	0.0005	0.0017	0.0015	-	0.0003	0.0005	0.0003	0.0007

Table 3.10: F_{ST} (Hudson *et al.*, 1992) pairwise comparison between *Alyssum serpyllifolium* populations calculated based on the candidate gene *NRAMP4* (left corner, in bold) and the neutral gene *ASIL1* (right corner, in regular font)

	Alhaurín-NS	Barazón-S	Carratraca-S	León-NS	Morata-NS	Rubía-NS	Samil-S	Sierra Bermeja-S
Alhaurín-NS	-	0.513	0.400	0.296	0.419	0.597	0.246	0.562
Barazón-S	**0.926**	-	0.239	0.325	0.542	0.757	0.305	0.639
Carratraca-S	**0.942**	**0.703**	-	0.209	0.358	0.588	0.160	0.507
León-NS	**0.006**	**0.883**	**0.928**	-	0.284	0.416	0.178	0.429
Morata-NS	**0.106**	**0.933**	**0.951**	**0.333**	-	0.580	0.327	0.559
Rubía-NS	**0.748**	**0.917**	**0.936**	**0.583**	**0.545**	-	0.605	0.650
Samil-S	**0.967**	**0.764**	**0.102**	**0.965**	**0.977**	**0.961**	-	0.502
Sierra Bermeja-S	**0.968**	**0.763**	**0.056**	**0.965**	**0.978**	**0.963**	**0.123**	-

Overall, similar sequence diversity levels were obtained for the two candidate genes and the neutral gene (Table 3.8). Between 24, 30 and 33 segregating sites were found in *IREG1*, *ASIL1* and *NRAMP4*, respectively, which were distributed among a high number of haplotypes, resulting in very high haplotype diversity: 0.809 ± 0.026 SD in *NRAMP4*, 0.834 ± 0.020 SD in *IREG1*, and 0.965 ± 0.009 SD in *ASIL1*. Highly similar synonymous (*ASIL1* π_S 1.232%, *IREG1* π_S 1.174%, *NRAMP4* π_S 1.341%) and non-synonymous (*ASIL1* π_{NS} 0.318%, *IREG1* π_{NS} 0.333%, *NRAMP4* π_{NS} 0.348%) nucleotide diversity levels were recorded in the three genes, resulting in π_{NS}/π_S ratios of 0.258-0.284. Thus, the neutral gene *ASIL1* offers a good comparison to the putatively selected genes. Interestingly, *IREG1* does not show the relaxed selective constraint expected in a pseudogene which should be evident in relatively higher π_{NS} (Lynch & Force, 2000), so the sequenced gene may have undergone pseudogenisation only recently.

Focusing on results from individual populations (Table 3.9): between 5 and 13 individuals were sequenced for *ASIL1*, 10 to 12 individuals for *IREG1*, and 6 to 11 individuals for *NRAMP4*, with the exception of León-NS, for which only 2 individuals were successfully sequenced. For this reason, the León-NS population was removed from most of analyses centering on *NRAMP4*. Gene diversity varied widely for each gene in the populations sequenced, but with no clear pattern, and so did nucleotide diversity and θ_W (Table 3.9). However, genetic differentiation between populations as measured by F_{ST} (Table 3.10) and D_A (Figure 3.10) revealed striking disparities between serpentine and non-serpentine populations in the two candidate genes, in contrast to the neutral gene. In *NRAMP4*, two firmly supported sub-clades grouping serpentine and non-serpentine populations, respectively, were reconstructed based on Nei's (1983) D_A (Figure 3.10), and were also recovered in an equivalent analysis of *IREG1* (data not shown). An absence of genetic parallelism through independent selection from ancestral standing variation in individual populations in the case of *NRAMP4* is supported by the presence of two distinct serpentine and non-serpentine intron haplotypes, which group into two sub-clades on the population tree (data not shown). The tests for recent selection (Tajima's *D*, Fu & Li's *D* and *F*, Fu's *F*, Fay & Wu's *H*) did not detect any departures from neutrality in any of the populations,

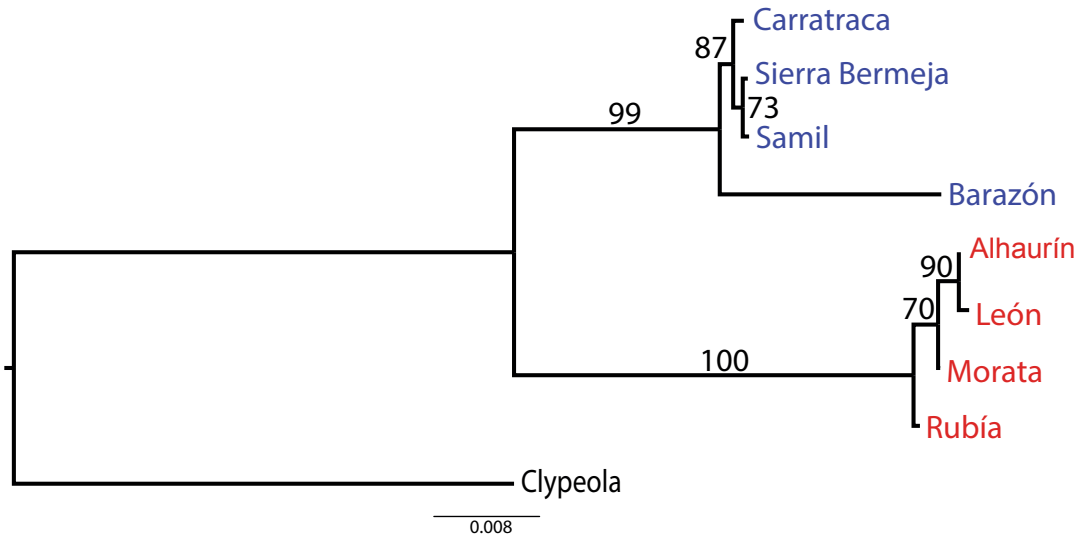


Figure 3.10: Gene tree based on Nei's (1983) D_A calculated for SNPs in the sequenced intronless *NRAMP4* gene fragment. Branch values represent percentage support over 1000 bootstrap replicates. Serpentine populations are marked in blue and non-serpentine populations in red.

which taken together implies that the single origin and fixation of a putatively adaptive haplotype through the action of divergent selection occurred relatively early in the history of the species complex.

Conversely, no well supported population tree (all bootstrap support values < 0.55) could be recovered based on *ASIL1*, with serpentine and non-serpentine populations intermixed on the tree (data not shown), emphasizing the far from simple evolutionary history of the species complex, which was also observed in the low power microsatellite analysis. This is also reflected in pairwise F_{ST} values for *ASIL1*, which did not vary significantly

Table 3.11: AMOVA results for the three resequenced genes (*ASIL1*, *IREG1*, *NRAMP4*) across eight populations of *Alyssum serpyllifolium*. Populations were split into two groups: serpentine and non-serpentine and partitioning of genetic variation compared among different levels.

	Percentage variation		
	Among groups	Among populations within groups	Within populations
ASIL1	4.5	41.1	54.4
IREG1	64.5	16.8	18.7
NRAMP4	89.7	4.6	5.6

when comparing populations from the two ecotypes (Table 3.10). Conversely, in *NRAMP4* pairwise comparisons of the serpentine and non-serpentine populations result in F_{ST} > 0.9, whereas for pairwise comparisons within serpentine populations, much lower differentiation was found, with F_{ST} typically < 0.125, with the exception of Barazón-S, where *NRAMP4* appears to have undergone additional longer sequence evolution than in other serpentine populations (Figure 3.10). AMOVA results (Table 3.11) further underscore the high degree of shared genetic variation in the candidate genes in populations of the same ecotype. When populations are grouped into serpentine and non-serpentine clusters, the overwhelming percentage (89.7% in *NRAMP4* and 64.5% in *IREG1*) of genetic variation is split between the two groups, while the rest is evenly distributed between *among populations within groups* and *within populations*. Almost the exactly opposite pattern is found for the neutral gene *ASIL1*, where no significant percentage of genetic variation (only 4.5%) is explained by the serpentine versus non-serpentine grouping, and the majority of variation is partitioned *among populations within groups* (41.1%) and *within populations* (54.4%).

Detailed breakdown of the outlier SNPs confirmed the pattern observed through AMOVA and population distance metrics (Table 3.12). Out of 10 outlier SNPs in *NRAMP4*, 8 are fixed for separate alleles in serpentine and non-serpentine populations as expected. Two SNPs show slight deviation: at one SNP position, the serpentine allele is segregating in Rubía-NS, and at the other SNP position, the serpentine allele is not fixed in Barazón-S, but both of these SNPs are at silent positions and so are not the focal variants which were initially under selection. Four out of the 8 SNPs showing serpentine versus non-serpentine polarity encoded non-synonymous changes, and in all cases the serpentine alleles are derived relative to non-accumulators *Arabidopsis thaliana* and *Clypeola jonthlaspi*, and so could likely be of adaptive importance. In particular, Asp→Asn amino acid replacement at position 1190 leads to a change from an amino acid with acidic to basic side-chain properties in the serpentine populations, which could influence protein function. A more complex picture emerges in the hypothesised *IREG1* pseudogene, where 1 out of 6 SNP positions is non-synonymous (position 509, Val→Ala) and a derived allele is fixed in

Barazón-S and Carratraca-S, but is segregating in Samil-S and Sierra Bermeja-S, as in the case of the other outlier SNPs in the (pseudo-)gene.

3.4 Key findings

- *BEAST multi-locus phylogenetic analysis of two serpentine and two non-serpentine *Alyssum serpyllifolium* populations revealed a rapid pattern of descent among all populations, tentatively dated at between 1.23-1.13 MYA (95% HPD), followed by genetic drift in the two mixed serpentine-non-serpentine population subclades.

- Application of various methods of phylogenetic inference (Neighbour-joining, Maximum Likelihood, NeighborNet, Treemix) on a selection of unlinked neutral SNPs supported a confident conclusion of a lack of clear separation of populations into serpentine versus non-serpentine subclades, indicating complex history of gene flow between serpentine and non-serpentine populations, with northern population's (Samil-S, Rubía-NS) lineages splitting off first.

- Species delimitation analysis of two serpentine and two non-serpentine populations with three different coalescent-based methods (BP&P, BFD, STACEY) proved inconclusive, with two extreme models favoured: single species model (STACEY) and four species model (BP&P, BFD). As the latter is an implausible model, supported neither by morphology, physiology nor ecology, the taxonomic recommendation is to retain all *Alyssum serpyllifolium* populations as a single species for now.

- 272 individuals distributed among eight Iberian *Alyssum serpyllifolium* populations were genotyped for eight putatively neutral EST-SSR markers developed from 3,942 di- to hexa- nucleotide microsatellites discovered through RNA-Seq.

- Results showed high levels of population differentiation, with overall high F_{ST} = 0.23 (0.17 – 0.29 95% CI) indicating restricted modern gene flow. F_{IS} values in the range of 0.07 – 0.35 indicate a moderate degree of inbreeding. Average within-population gene diversity (H_O = 0.48) and heterozygote deficiency are typical of endemic species with small population sizes.

- The power afforded by selected microsatellite loci was too low to establish the relationships between the serpentine and non-serpentine populations with any degree of confidence but STRUCTURE analysis pointed to a complex interplay of both geography and edaphic factor as likely drivers of genetic divergence.

- *BEAST multi-gene phylogenetic analyses of two subsets of shared candidate genes in the two serpentine populations - involved in metal homeostasis and other processes - arrived at a different outcome than that based on neutral genes, with serpentine and non-serpentine populations separated into two distinct subclades. This shared selected genetic variation arose once in *Alyssum serpyllifolium* and forms the basis for the superior Ni hyperaccumulation phenotype and serpentine adaptation seen in metallicolous populations.

- That pattern was further confirmed in an extended sampling of four serpentine and four non-serpentine populations sequenced for two select outlier genes involved in metal transport: *NRAMP4* and *IREG1* (with the caveat that a pseudogenised copy of this gene was likely sequenced). Phylogenetic analyses recovered one serpentine and one non-serpentine subclade, and no well-supported topology was established in the neutral gene tree. While the vast majority of genetic variation was accounted for by the serpentine-non-serpentine division in the AMOVA analyses: NRAMP4 (90%) and IREG1 (65%), in the neutral gene it only accounted for 4.5%. A number of non-synonymous derived alleles was found in the candidate serpentine genes, some of which are proposed to have a functional effect on the protein due to altered amino-acid physiochemical properties (Table 3.12).

Table 3.12: Detailed summary of the distribution of *IREG1* and *NRAMP4* outlier SNPs among S (serpentine) and NS (non-serpentine) *Alyssum serpyllifolium* populations. Positions of SNPs are given with reference to the original contig from the RNA-Seq assembly. The substitution effect is given for non-synonymous positions. Ancestral status of the alleles was determined based on *Clypeola* and *Arabidopsis* outgroups.

Position of candidate SNP	Fixed in all S populations?	Present in NS populations?	Substitution (NS→S)	Substitution effect	S allele derived?
IREG1					
171		region missing in sequenced amplicon			
203		region missing in sequenced amplicon			
493*	Segregating in SB and Samil	N	C → A	Ser → Arg	N
504	Absent in SB, segregating in Samil	N	A → G	-	Y
509	Segregating in SB and Samil	N	T → C	Val → Ala	Y
516	Segregating in SB and Samil	N	G → T	Arg → Ser	N
556	Segregating in SB and Samil	N	C → A	Leu → Ile	N
681*	Segregating in SB and Samil	N	G → A	-	Y
NRAMP4					
343	Y	Segregating in Rubía	C → G	-	Y
511	Y	N	T → G	-	Y
519	Y	N	C → G	Ala → Gly	Y
527	Y	N	G → A	Val → Ile	Y
571	Y	N	C → T	-	N
976	Y	N	T → C	-	N
1019	Y	N	A → C	Ile → Leu	Y
1060	Y	N	C → T	-	Y
1190	Y	N	G → A	Asp → Asn	Y
1222	Segregating in Barazón	N	T → C	-	N

SB - Sierra Bermeja

* - outlier SNP not detected in the original RNA-Seq dataset

Chapter 4

Application of the Devised RNA-Seq Pipeline in Caryophyllales

4.1 Introduction

Despite the diversity of ecological forms and importance of members of the Caryophyllales for humans as crops (e.g. buckwheat, beet, spinach, quinoa, jojoba, rhubab, *Amaranthus* spp.), the Caryophyllales has not substantially benefited from the on-going revolution in genome and transcriptome sequencing brought about by NGS until very recently. Up until now, only one high quality Caryophyllales genome - the sugar beet cultivar of *Beta vulgaris* (Dohm *et al.*, 2013) has been generated, and an early draft version of the *Amaranthus hypochondriacus* genome was released last year (Sunil *et al.*, 2014). Only a handful of Caryophyllales transcriptomes were also available until 2015 - buckwheat (Logacheva *et al.*, 2011), carnation (Tanase *et al.*, 2012), spinach (Dohm *et al.*, 2013) and venus flytrap (Jensen *et al.*, 2015). This all changed with a recent release of 27 genomes and 69 Caryophyllales transcriptomes (Yang *et al.*, 2015). However, the studies mentioned earlier concentrated either on generating sequences useful for breeding efforts in agriculture (e.g. Dohm *et al.*, 2013) and horticulture (e.g. Tanase *et al.*, 2012) or focused on the problems of phylogenomic nature (Brockington *et al.*, 2015; Christin *et al.*, 2015; Yang *et al.*, 2015), and never patterns of gene expression differences of ecological-evolutionary importance.

The study's first goal is to generate the first transcriptomes of four non-model plant species in Caryophyllales. To achieve this, the custom-designed bioinformatic pipeline

for transcriptome reconstruction and quantification developed in Chapter 2 for the plant taxon *Alysseae* will be applied and modified for this new objective. Subsequently, the first-ever comparison of closely related C_3 and C_4 Caryophyllales species will be attempted, in order to identify gene-expression differences that could be associated with their key traits and specific ecological adaptations. In the case of *A. angustifolia* and *A. pungens*, the focus lies in a comparison of the C_3 (*A. angustifolia*) and C_4 (*A. pungens*) modes of photosynthesis, whereas in the case of *H. peploides* and *S. globosa* the focus is adaptation to high salinity and low temperature (*H. peploides*) compared with more mesophytic traits (*S. globosa*).

Plant transcriptomes will be sampled at warm (25°C, 30°C) and colder (18°C) growth temperatures, which will allow identification of species-constitutive gene expression differences found across all the differing temperatures, but also those specific to either cold or warm temperature. Temperature-specific differences could illuminate our understanding of greater flexibility in acclimation to temperature and high abiotic stress tolerance found in *A. angustifolia*, which has resulted in its spread to colder, temperate regions and diverse habitats as a highly invasive weed, unlike the less invasive *A. pungens*, which is restricted to warmer regions. Specifically, temperature-induced gene expression changes in the C_3 and C_4 core photosynthetic enzymes will be of interest.

Aims and objectives

- Evaluation and application of a previously developed bioinformatics pipeline in reconstructing the transcriptomes of unrelated non-model species harbouring complex traits of ecological importance in an order where few genetic resources were previously available.

- Genome-wide *in silico* discovery of SNP and SSR genetic markers to facilitate future population genetics studies in *Caryophyllales* species of interest in weed science and photosynthesis research.

- Transcriptome-based comparison of gene expression plasticity in response to low and high temperature in C_3 and C_4 species, with particular focus on the photosynthesis machinery.

- Variant-focused discovery of evidence for possible isoform switching in photosynthesis genes in C_3 and C_4 species.

4.2 Materials and methods

4.2.1 Sample preparation

4.2.1.1 Plant material

Seeds of *Alternanthera pungens* and *Alternanthera angustifolia* were obtained from the Royal Botanic Gardens, Kew (UK). *Schiedea globosa* was kindly provided as cuttings by Dr Stephen G. Weller from greenhouse collections at the University of California, Irvine (USA). Accessions of *Honckenya peploides* were collected by Dr Maxim Kapralov north of the Arctic Circle near Tromsø (Norway). The *Alternanthera* species were germinated from seed at the beginning of the experiment, while *S. globosa* cuttings were rooted and field-collected *H. peploides* plants were kept in a greenhouse before the start of the experiment when they were transferred into three growth chambers. Growth chambers conditions were the day temperature of 18°C, 25°C (just *S. globosa* and *H. peploides*) and 30°C, and night temperature of 15°C, 20°C and 25°C, respectively. The day length was set to 16 hours. Plants were kept in the growth chambers for 4 months, after which leaf samples were harvested from plants in full vigour.

4.2.1.2 RNA extraction and sequencing

Three biological replicates of *Alternanthera philoxeroides* and *Alternanthera pungens* leaves were collected from separate plants grown at 18°C and 30°C. Two replicates were collected from *Honckenya peploides* and *Schiedea globosa* leaves, each, from separate plants grown at 18°C, 25°C and 30°C, each. In total, six RNA-Seq libraries were prepared for each species and sequenced. RNA was extracted from mature leaves at the end of 4-month acclimation period. Extraction was performed using the RNeasy Plant Mini Kit (QIAGEN)

following the manufacturer's protocol, and RNA was immediately snap frozen in liquid nitrogen and shipped to the Wellcome Trust Centre for Human Genetics (Oxford, UK) sequencing facility on dry ice.

4.2.2 Read quality control

All FASTQ files with sequencing reads output by the Illumina system were first examined using Prinseq 0.20.3 (Schmieder & Edwards, 2011) and FastQC 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to confirm the quality of sequencing – i.e. consistently high-quality scores across the length of reads, low sequence duplication and sequence over-representation levels. Since FastQC detected Illumina adapters and PCR primers amongst over-represented sequences, all such contaminating sequences were then trimmed with Trimmomatic ver. 0.32 using the authors' recommended settings (Bolger *et al.*, 2014), along with stretches of bases with low Phred quality score ($Q > 15$ averaged over a 5-bp sliding window) at the ends of reads. Fqtrim 0.94 (https://ccb.jhu.edu/software/fqtrim/) was then used to remove A/T homopolymer stretches ($n > 5$ bp) at the ends of reads, indicative of polyA tails; to select reads based on size ($n > 20$ bp), and to filter out the reads with high N (undefined base) content ($n >= 5\%$). Any possible eukaryotic or prokaryotic rRNA contamination was filtered with SortMeRNA 1.8 (Kopylova *et al.*, 2012) and finally the remaining reads were corrected for sequencing errors with Musket 1.0.6 (Liu *et al.*, 2012**b**).

4.2.3 Assembly, annotation and mapping

Corrected reads were then assembled using Oases 0.2.8 (Schulz *et al.*, 2012) with *k*-mer sizes ranging from 21 to 81 (step size: 10) on two data sets for each species: a) reads from all individuals in the species; b) reads from one individual grown at 18°C and one grown at 30°C. The latter option was considered in order to obtain representative sequences from genes expressed at different temperatures, but also to avoid assembly problems associated with increased polymorphism introduced by adding more individuals (increased redundancy and contig fragmentation, Ruttink *et al.*, 2013). Low *k*-mer sizes will result in quick termination of de Bruijn graph assembly, as closely related genes will share some

nodes, so higher *k*-mer values are generally preferred. However, the use of higher *k*-mer values for genes expressed at a low level will result in too low *k*-mer coverage for proper assembly, so a range of *k*-mer sizes for assembly construction is recommended (Gruenheit *et al.*, 2012).

Scaffolding was allowed with minimum support of eight bridging read pairs, and a transcript had to be at least 300 bp long to be retained. Contigs within each assembly were then merged with USEARCH 7.0 (Edgar, 2010) using the Greedy *smallmem* algorithm and clustering percentage identity set at 95%, 98% or finally 99% with centroid output. Subsequently, assemblies from different *k*-mer sizes in the range of 31 to 61 were merged and re-clustered with USEARCH to arrive at the final assembly with reduced redundancy.

Assemblies were annotated against the 1.1 release of the *Beta vulgaris* ssp. *vulgaris* (sugar beet) genome (Dohm *et al.*, 2013), as it was the only high quality genome available in the order Caryophyllales. Contigs were searched with the BLASTx algorithm (Altschul *et al.*, 1997) implemented in the BLAST 2.2.26+ suite against sugar beet protein-coding sequences, and hits with minimum e-value of 10^{-5} were accepted. Contigs not annotated after this procedure were then searched with BLASTx, this time against the *Arabidopsis thaliana* proteome (TAIR10 release, Lamesch *et al.*, 2011) and lastly against the nr database using Blast2GO 2.7.1 (Götz *et al.*, 2008) with default settings but minimum e-value of 10^{-5}. In addition, prior to clustering, transcripts used for population genetics analysis were filtered for those containing N (unknown) basepairs, premature stop codons within the high-scoring segment pairs (HSPs) and frameshift mutations as evidenced by HSPs in different frames. The selected contigs were reverse complemented to obtain plus-strand sequence, if needed, and then annotated with putative 5′ UTR, CDS and 3′ UTR regions with a custom Perl script. If a start codon aligning with the 1 position of the reference protein was not found in the alignment, any first methionine present upstream of the aligned portion of the protein was treated as the start codon, or no 5′ UTR was defined in the transcript unless a start codon was found. Similarly, if no stop codon was present downstream from the aligned portion of the gene, no 3′ UTR was defined and all the nucleotides up to the 3′ end were treated as protein-coding.

Final mapping was carried out with STAR 2.3.0 (Dobin *et al.*, 2013) with maximum alignment mismatch set at 10 and maximum number of alignments for a read set to 10 for differential expression analyses and set to 1 to eliminate multi-mapping reads for population genetics analyses. Mappings were subsequently filtered in SAMtools 0.1.17 (Li *et al.*, 2009**a**) to remove chimeric and discordant reads.

4.2.4 Differential expression analysis

Read abundance for each contig was estimated using eXpress ver. 4.0 (Roberts & Pachter, 2013), as its use prevents the loss of information resulting from discarding multi-mapping reads. Using the Expectation-Maximisation (EM) algorithm, eXpress re-assigns multi-mapping reads to alternative contig locations with appropriate weight, given each contig's expression based on uniquely mapping reads and other parameters. Rounded effective counts obtained from eXpress were used as input for differential expression analysis conducted using two R packages: edgeR ver. 3.6 (Robinson *et al.*, 2010) and DESeq2 ver. 1.4 (Love *et al.*, 2014), the intersection of whose results was treated as the final list of differentially expressed genes in the study. When conducting interspecific comparisons, read counts were adjusted in linear proportion to the contig coverage of its sugar beet or *Arabidopsis* homolog to correct for bias inherent when using two separate assemblies. Genes in the data set were filtered to include only those showing expression higher than 1 count per million in at least three samples. Extremely lowly expressed genes are of little biological interest and are unlikely to be found differentially expressed. but eliminating them and thus reducing the number of tests made will increase the power of the remaining statistical comparisons (Chen *et al.*, 2014**b**).

The Generalized Linear Model (GLM) design feature from the two packages was used to fit a model including all the samples with specified factors (*temperature*, and *species* - only in the interspecific comparisons) in order to establish better gene dispersion estimates and facilitate multiple contrasts (Chen *et al.*, 2014**b**). In DESeq2, the function `DESeq` was used to carry out the default statistical analysis on the filtered data set and contrasts

of interest were investigated with the `results` function. In edgeR, library sizes were first normalised with `calcNormFactors` function followed by dispersion estimation with `estimateGLMCommonDisp`, `estimateGLMTrendedDisp` and `estimateGLMTagwiseDisp` before fitting gene-wise negative binomial GLMs with `glmFit` function using the calculated dispersion values. Individual pairwise comparisons were then accessed using `makeContrasts` and `glmLRT` functions. DESeq and edgeR utilise similar strategies to detect differentially expressed (DE) genes, with no established clear superiority of either method (Robles *et al.*, 2012; Soneson & Delorenzi, 2013); however, edgeR can suffer from a high false positive rate. To mitigate against this possibility, genes output as significantly differentially expressed, were only retained if they remained so after applying the Benjamini-Hochberg procedure to control for a false discovery rate (FDR) < 0.01 in both packages. `PlotMDS` and `plotPCA` functions in edgeR and DESeq2, respectively, as well as a core R function (R Core Team, 2014) `prcomp` (input: $\log_2$ trimmed-mean normalised read counts), were finally used to obtain an overview of expression pattern similarity of all the samples.

Enrichment of gene ontology (Ashburner *et al.*, 2000) terms and metabolic pathways amongst upregulated and downregulated genes in a given contrast was tested using the GOSeq package ver. 1.18 (Young *et al.*, 2010), which aims to correct for bias in enrichment results related to over-representation of long genes amongst differentially expressed genes. Here, the results were standardised by the use of the CDS length of the model species homolog to be consistent and facilitate interspecific comparisons; otherwise default settings were followed. Lastly, all the enrichment test results were adjusted for multiple comparisons using Benjamini-Hochberg correction with an FDR cut-off of 0.05.

4.2.5 SNP calling and SSR identification

PCR and optical duplicates were marked in all the samples with the MarkDuplicates module from Picard Tools ver. 1.114 (http://picard.sourceforge.net/) to exclude them from further analysis. This step was required because increase in coverage of one allele due to read duplication can result in miscalling heterozygotes as homozygotes and errors

introduced in PCR duplicates can be treated as true variants by genotype callers (Andrews & Luikart, 2014).

To allow multi-sample SNP calling, all the individual BAM files were tagged with sample and group names with bamaddrg (http://github.com/ekg/bamaddrg, default settings). Then, SNP calling was carried out on all the samples in a given species simultaneously using three different variant callers: freebayes ver. 0.9.14-2 (Garrison & Marth, 2012) using default settings, GATK Unified Genotyper ver. 3.1-1 (DePristo *et al.*, 2011, options `-stand_call_conf 5 -stand_emit_conf 5` to relax stringency of filtering of genotypes to be output) and SAMtools. Variant calls from the three programs were then integrated with BAYSIC (Cantarel *et al.*, 2014) leading to retention of only calls made by at least two of the variant callers.

Microsatellites in all the assemblies were identified using SciRoKo ver. 3.4 (Kofler *et al.*, 2007) using the perfect search algorithm and specifying minimum total allele length to be 10 basepairs, combined with minimum repeat number of 10-5-5-4-4-3 (for mono- to hexa- nucleotide motifs, respectively).

4.2.6 Differential allele expression in Rubisco small subunit and Rubisco activase genes in response to temperature in *Alternanthera*

All the contigs with a best hit to either Rubisco small subunit (*rbcS*) or Rubisco activase (*rbca*) genes in the *B. vulgaris* proteome were extracted from each species' assembly used for differential expression analysis. The contigs were blastxed against the *A. thaliana* and *B. vulgaris* proteomes and the contigs with evidence of chimerism were eliminated. Both Rubisco small subunit and Rubisco activase are highly conserved across the plant kingdom, so any deviation from the transcript structure present in the model species was indicative of errors in the assembly. Subsequently, one representative contig for each gene was selected based on three criteria: highest similarity and coverage of the two model species' homologs as well as high mapping rate. Then, reads were mapped to modified assemblies, where all the contigs with homology to *rbcS* and Rubisco activase were substituted for by the chosen representative contig for each gene. Since nucleotide diversity was found to be 0.0735 for

the 5 *rbcS* copies found in *B. vulgaris* genome, the number of maximum mismatches was set at 20 for each read pair to allow for alignment from reads derived from all the copies of the *rbcS* gene in the genome. Otherwise, the same procedure was followed as given above for differential expression mapping. Finally, SNPs were called using GATK Unified Genotyper and allele depths for the SNPs within rbcS or Rubisco activase coding regions were found for all SNPs fixed for heterozygotes in all the samples from a given species. The read counts for each allele were adjusted in linear proportion to the total mapped reads in each sample using RPKM standardisation. This allowed a test of differential expression of gene isoforms at 18°C and 30°C in the two species without establishing the exact number and sequence of each gene isoform, which proved too challenging using *de novo* assembly approach. The difference in expression between reference and alternative alleles at a given position in response to temperature was tested with replicated G-test of independence in R with the package `Deducer` (Fellows, 2012).

4.3 Results and discussion

4.3.1 Sequencing and read quality control

Table 4.1: Summary of RNA-Seq reads obtained in the Caryophyllales study

	A. angustifolia	*A. pungens*	*H. peploides*	*S. globosa*
Number of samples grown at:				
18°C	3	3	2	2
25°C	0	0	2	2
30°C	3	3	2	2
Raw read pairs (in millions) / sample	4.52-5.48	4.22-5.93	4.5-5.51	4.33-5.55
Mean read length after QC / sample	95.7-96.1	96.0-96.1	96.4-96.6	96.5-96.8
% retained reads / sample	78.9-89.5	78.1-86.6	80.3-86.3	75.2-78.2

RNA-Seq has become the default method for measuring gene expression differences in the biological samples of interest. Here, the transcriptomes of four species in the order Caryophyllales were sequenced: 3 independent leaf samples of the two *Alternanthera* species - *A. angustifolia, A. pungens* grown at 18°C or 30°C; for *H. peploides* and *S. globosa* two leaf samples of plants grown at those temperatures were sequenced, each, and in

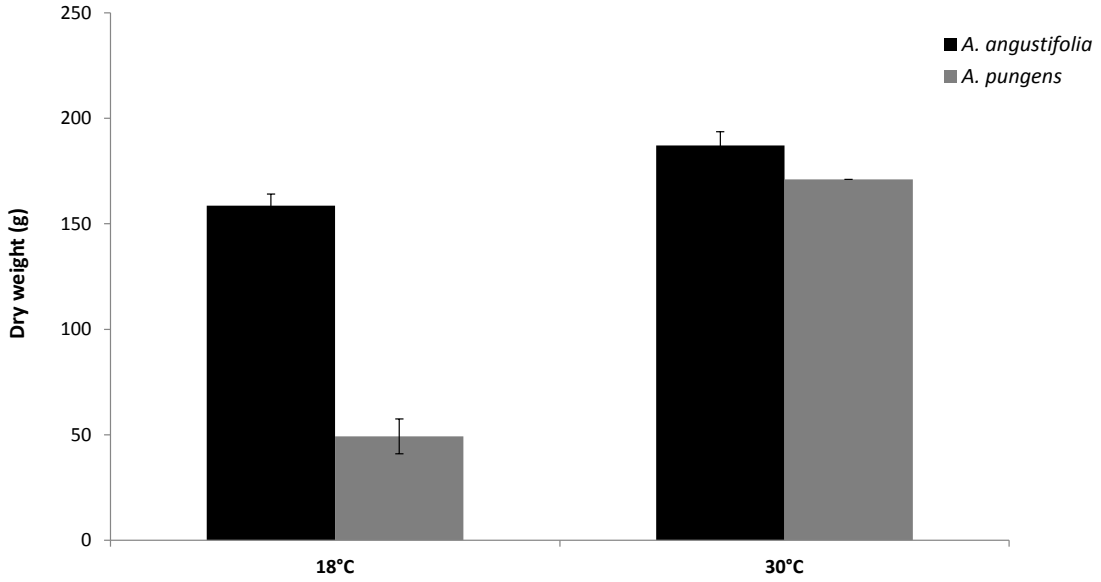


Figure 4.1: Mean dry weight of *A. angustifolia* (C_3) and *A. pungens* (C_4) plants grown for 4 months at 18°C and 30°C. Error bars represent standard error based on 3 replicates.

addition two more leaf samples from plants grown at 25°C were sequenced. In the case of *Alternanthera*, the principal focus here lies in differences exhibited in acclimation to temperature in the two tropical species differing in their mode of photosynthesis: C_3 (*A. angustifolia*) and C_4 (*A. pungens*). The C_4 species *A. pungens* shows significantly poorer ability to acclimate to the lower temperature, and experiences a 71% reduction in its mass when grown at 18°C, as compared to when grown at 30°C, while in the C_3 species *A. angustifolia* the relative reduction amounts to only 15% in mass (Figure 4.1). The comparison of *H. peploides* and *S. globosa* transcriptomes at different growth temperatures promises to offer insight into expression changes required for acclimation to temperature in plants naturally adapted to different climates and environmental conditions, namely: coastal, tropical (*S. globosa*) and dunelike, subarctic (*H. peploides*). In the investigation conducted, each sample consisting of leaves from a single individual was sequenced at the depth of between 4.22 million and 5.93 million reads (Table 4.1). After quality control, including removal of low-quality terminal basepairs, putative polyA-tails and extremely short reads as well as reads with high N content, average read length centred on 95.7 – 96.8 bp (Table 4.1). Between 75.2% and 89.5% of reads were kept for final analysis, with virtually all of the removed reads classified as rRNA contamination (data not shown).

Nevertheless, a high percentage of discarded reads is a common problem of RNA-Seq studies, with typically 21% of reads found to be removed during QC step in a survey of 21 studies conducted by Dewoody *et al.* (2013).

4.3.2 Assembly, mapping and transcriptome annotation

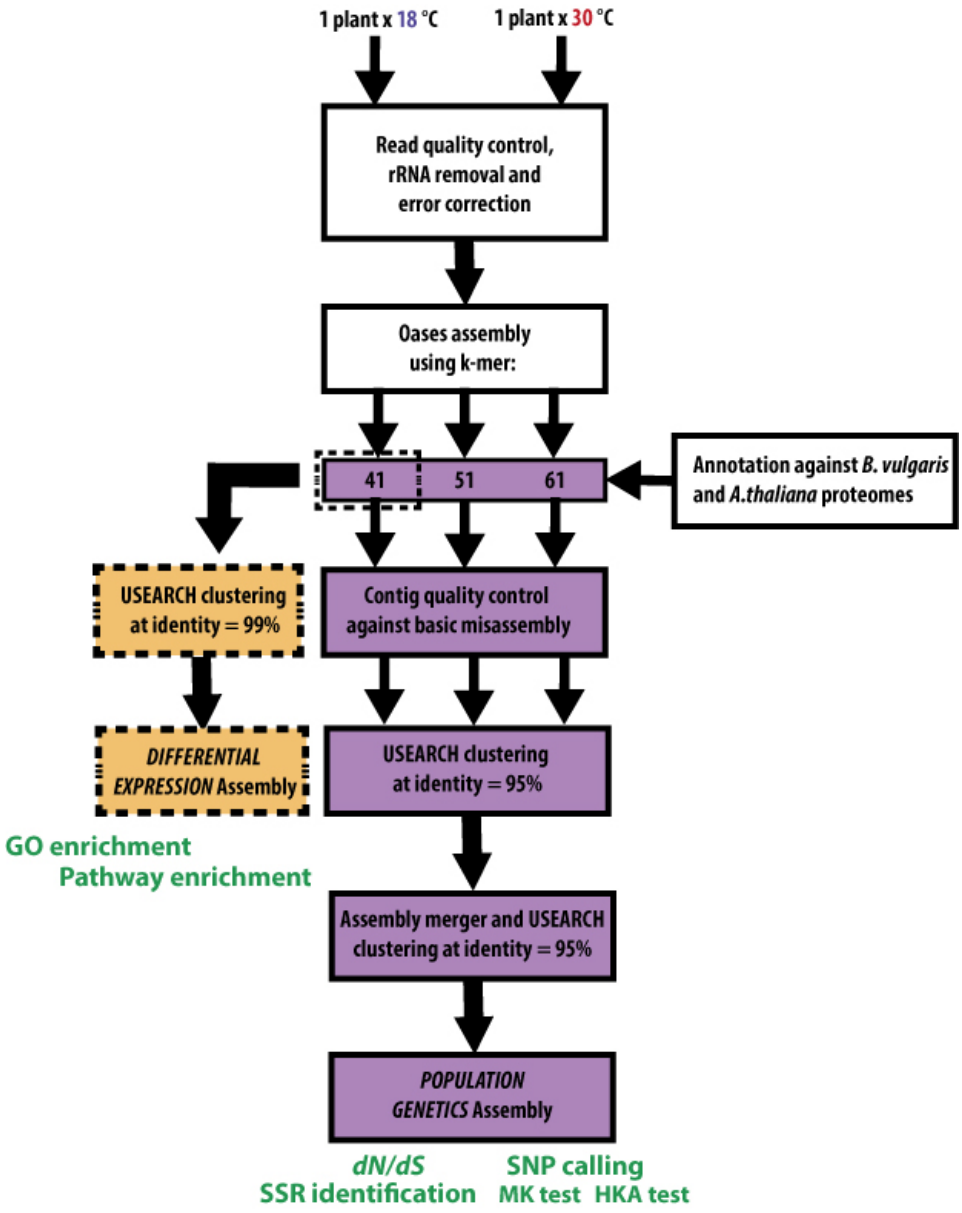


Figure 4.2: Outline of the workflow used in the analysis of Caryophyllales RNA-Seq data

Two assemblies were prepared: one for differential gene expression analysis (abbreviated as *expresssion* assembly) and a separate one for population-genetic analysis

(abbreviated as *popgen* assembly) (Figure 4.2). This separation was made because it was impossible to achieve a good balance between mapping rate and redundancy in *S. globosa*, *H. peploides* and *A. pungens* due to high nucleotide diversity (data not shown). When using a single *k*-mer assembly of 41 with and without a 99% clustering step carried out by USEARCH, total mapping rate was higher at around 80% on average, but the multi-mapping rate increased to around 40% on average (data not shown). When contigs were clustered to remove redundancy at the identity of 95%, the multi-mapping rate dropped to 10% but overall mapping rate decreased to 60% in the three species (data not shown). Building on the strengths of the two redundancy removal approaches, two assemblies were devised in the study. For differential gene expression analysis, highly redundant assembly based on reads from 2 individuals grown at 18°C and 30°C, respectively, assembled using *k*-mer = 41 and clustered at 99% identity was chosen (Figure 4.2). For population genetic analysis, more stringent criteria for contig selection were used to decrease mis-assembly and redundancy rate at the cost of lower mapping rate. Selected contigs from individual *k*-mer assemblies based on more conservative *k*-mer values of 41, 51 and 61 were first clustered at 95% identity, combined and re-clustered again at 95% (Figure 4.2). In the

Table 4.2: Assembly and mapping statistics for the *Alternanthera* transcriptomes in the study

	A. angustifolia		*A. pungens*	
	Expression	***Popgen***	***Expression***	***Popgen***
Number of contigs	19,306	15,921	26,726	17,460
Mean contig length	1,488	1,381	1,282	1,131
Median contig length	1,265	1,183	1,107	945
N50	1,877	1,729	1,612	1,430
Unique mapping rate	61.2-63.4%	61.2-64.0%	37.1-39.7%	48.3-50.9%
Multi-mapping rate	23.8-24.6%	11.5-12.0%	39.0-42.6%	16.4-17.8%

expression assemblies, the number of contigs varied from 19,306 in *A. angustifolia* (15,921 *popgen* assembly) to 25,695 in *S. globosa* (18,144 *popgen* assembly), 26,726 in *A. pungens* (17,460 *popgen* assembly) and 27, 962 in *H. peploides* (17,987 *popgen* assembly). *Expression* assemblies had longer contig length than the corresponding *popgen* assemblies in all cases (Tables 4.2, 4.3). Mean contig length ranged from 1,263 bp in *H. peploides* to 1,488 bp in

Table 4.3: Assembly and mapping statistics for *H. peploides* and *S. globosa* transcriptomes in the study

	H. peploides		*S. globosa*	
	Expression	***Popgen***	***Expression***	***Popgen***
Number of contigs	27,962	17,987	25,695	18,144
Mean contig length	1,263	1,093	1,353	1,196
Median contig length	1,066	896	1,151	1,001
N50	1,612	1,397	1,716	1,527
Unique mapping rate	36.5-38.4%	45.0-46.7%	41.3-46.6%	45.0-49.0%
Multi-mapping rate	36.2-39.0%	14.0-15.6%	29.5-33.6%	14.3-15.1%

A. angustifolia in the *expression* assembly, while it ranged from 1,093 bp in *H. peploides* to 1,381 bp in *A. angustifolia* in the *popgen* assembly. Analogous results were obtained for median contig length and N50 statistic (Tables 4.2, 4.3).

In terms of mapping rate, *A. angustifolia* stood out from the three other species in terms of its much higher unique mapping rate: 61.2%-64.0% reads mapped in both assemblies, and lowest multi-mapping rate. This coincided with the smallest number of contigs and the greatest length of all assemblies considered, showing that the lowest redundancy was achieved in this species' transcriptome. *S. globosa* was intermediate with 41.3-49% unique mapping rate, followed by *A. pungens* (37.1-50.9% unique mapping rate) and *H. peploides* (36.5-46.7% unique mapping rate) (Tables 4.2, 4.3). While multi-mapping rate decreased significantly in the *popgen* assembly in all the species: e.g. from 23.8-24.6% to 11.5-12.0% in *A. angustifolia* and from 39.0-42.6% down to 16.4-17.8% in *A. pungens*, this came with a trade-off of reduced overall mapping rate, i.e. precision over sensitivity. In total, *A. angustifolia* and *A. pungens* had the highest overall mapping rates in the *expression* assembly, at 85.0-88.0% and 76.1-82.3%, respectively, followed by *H. peploides* and *S. globosa* at 72.7-77.4% and 70.8-80.2%, respectively. The decline in mapping rate for *popgen* assemblies relative to *expression* assemblies ranged from 12-12.3 pp in *A. angustifolia*, 11.4-14.6 pp in *A. angustifolia*, 13.7-15.1 pp in *H. peploides* through to 11.1-15.9 pp in *S. globosa*. This is suggestive of not only reduced redundancy, but also sensitivity in the *popgen* assemblies, as non-redundant sequences were apparently also removed, resulting in reduced mapping

rate.

Contigs in both types of assemblies were first annotated using the only currently available high quality genome in the order Caryophyllales (sugar beet), to which all the species in the study belong. The vast majority of contigs could be annotated in this way: in the *A. angustifolia expression* assembly, 17,898 out of 19,306 (92.7%) of contigs had a sugar beet homolog, followed by 123 annotated with an *Arabidopsis thaliana* homolog among the remainder of unannotated contigs. In the *A. pungens expression* assembly, 25,193 out of 26,726 (94.3%) contigs were annotated with a sugar beet homolog and 155 with an *A. thaliana* homolog. In *H. peploides*, 25,690 out of 27,962 (91.9%) / 212 contigs, while in *S. globosa* 23,677 out of 25,695 (92.1%) / 203 contigs were annotated with sugar beet / *A. thaliana* homologs, respectively. An even higher fraction of contigs was annotated with a sugar beet homolog in the *popgen* assemblies (data not shown). In total, a moderate number of genes was annotated, most likely due to low sequencing depth (Tarazona *et al.*, 2011): 10,578 in *A. angustifolia*, 10,116 in *A. pungens*, 9,479 in *H. peploides*, 9,571 in *S. globosa*. *B. vulgaris* is predicted to contain 27,421 protein-coding genes (Dohm *et al.*, 2013), so slightly more than one-third of *B. vulgaris* homologs were detected in all the *expression* assemblies.

When comparing amino acid percentage identity relative to sugar beet proteins, it varied from mean 74.5-74.% (± 13.5% SD) in *S. globosa* and *H. peploides*, respectively, to 79.0- 79.5% (± 12.4% SD) in *A. pungens* and *A. angustifolia*, respectively. It has been shown that protein sequence similarity in the range of 40-60% allows for reliable functional annotation of the homologs (Addou *et al.*, 2009), so the annotation utilised in the study should be broadly correct. The completeness of the reconstructed transcripts was measured in two ways: by looking at the extent of coverage of reference protein homologs from *B. vulgaris* (Figure 4.4) and at distribution of lengths of 5′ UTRs, CDS and 3′ UTRs in the four species in comparison to *B. vulgaris* (Figure 4.3). Just as in terms of transcript length, *A. angustifolia* assemblies proved to be the best with median coverage of *B. vulgaris* proteins at 90.4% and 92.0% in *popgen* and *expression* assemblies, respectively (Figure 4.4). *A. angustifolia*

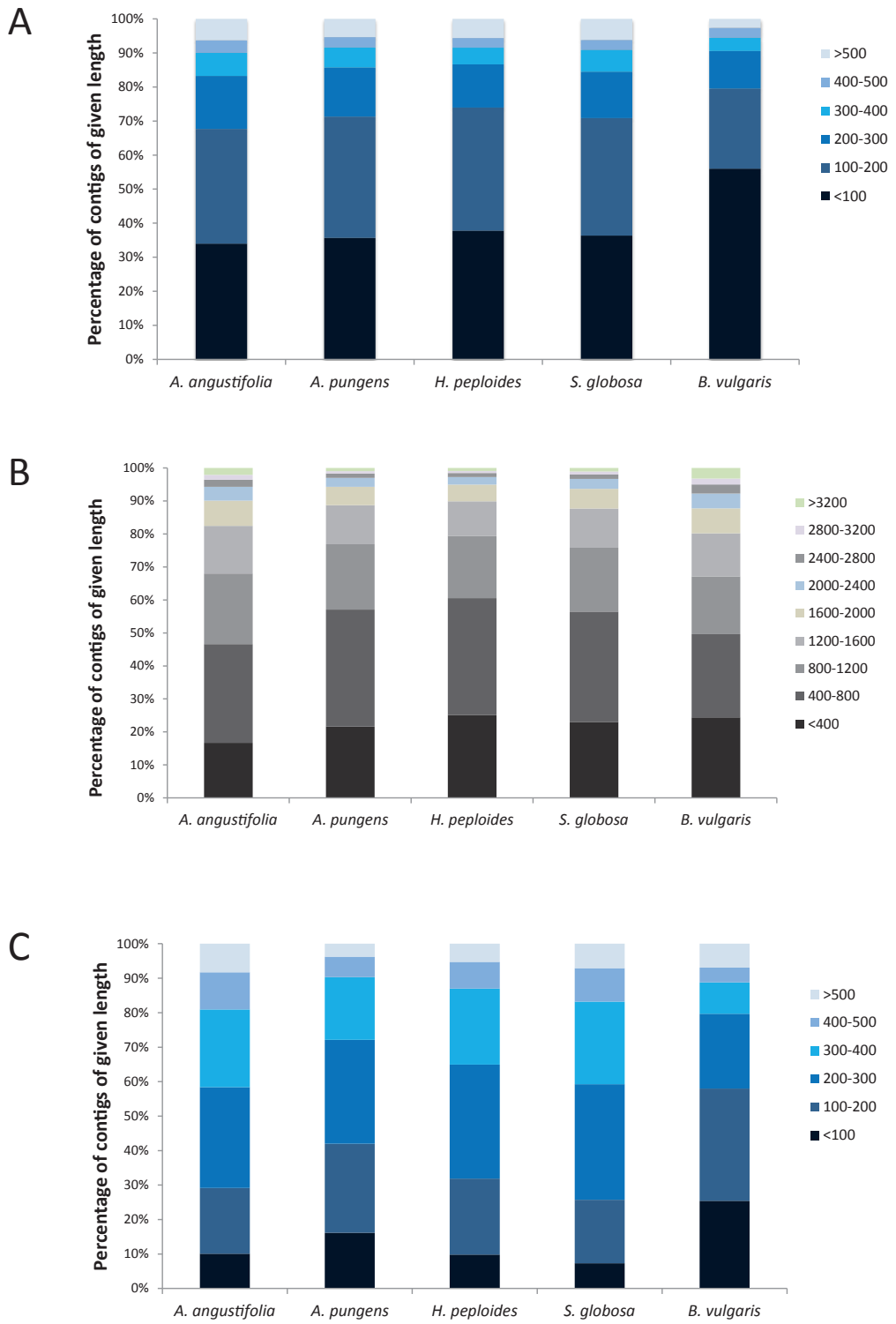


Figure 4.3: Comparison of distribution of **A)** 5′ UTR; **B)** 3′ UTR; **C)** CDS region length among transcripts in *A. angustifolia*, *A. pungens*, *H. peploides* and *S. globosa popgen* assemblies and the reference *B. vulgaris* transcriptome.

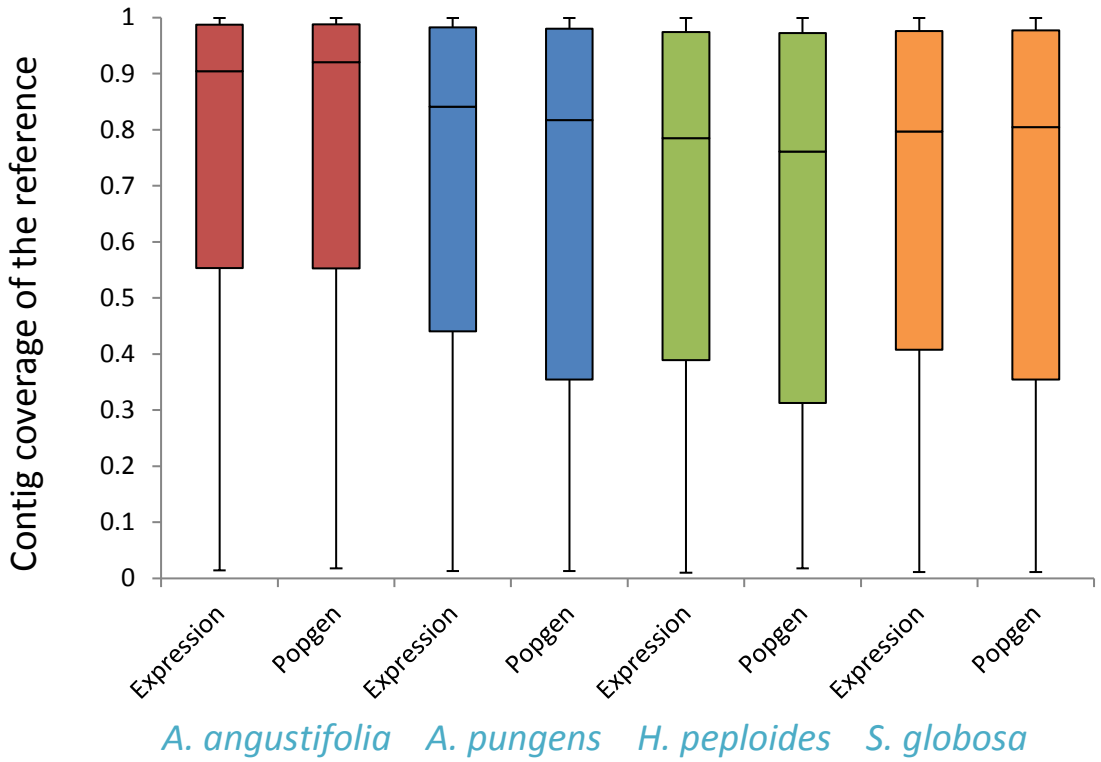


Figure 4.4: Coverage of the contigs in different assemblies relative to the reference sugar beet root proteins

was then followed by the sister species *A. pungens* at median 84.0% and 81.7% coverage, *S. globosa* (79.7% and 80.5%) and *H. peploides* (78.5% and 76.1%). While the range of coverage was similar across the four species' assemblies, *A. angustifolia* coverage distribution was skewed towards contigs with better coverage, as can be seen by the shift of the first quartile by between 11.3-24.0 pp relative to the rest of the transcriptomes. Reassuringly, the increase in the contig sizes in the *expression* assemblies relative to the *popgen* assemblies in *A. pungens*, *H. peploides* and *S. globosa* corresponded to the increase in the coverage of the sugar beet reference proteins, which supports the correctness of the assembly process. Next, the overall completeness of the transcripts was assessed by not only looking at the coding region but also at the 5′ and 3′ UTRs by comparing the length of each transcript region in the generated assemblies relative to the *B. vulgaris* reference transcriptome, which looked very similar to *B. vulgaris* for 5′ UTR and CDS regions, and less so for 3′ UTR. The over-representation of longer sequences among the 3′ UTR in the sequenced species could

be explained by a combination of factors: 3' decay of sequenced transcripts and incorrect annotation of contigs with premature stop codons (Figure 4.3).

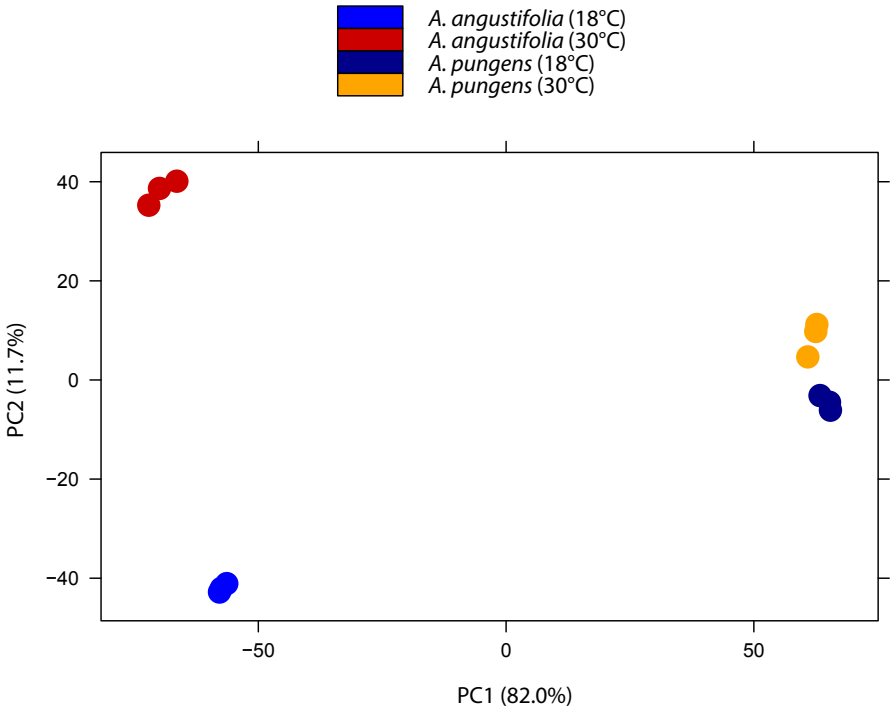
4.3.3 Differential gene expression analysis

MDS (multi-dimensional scaling) function in the edgeR package as well as the PCA (principal component analysis) function in DESeq2 and R-base were used to visualise the relationships between the samples to check the reproducibility of biological replicates (Figure 4.5). While the three replicates from the two temperatures, 18°C and 30°C, clustered together according to the temperature along the first two principal components in *A. angustifolia* and *A. pungens*, this was not the case for *H. peploides* and *S. globosa*. In the *A. angustifolia-A. pungens* PCA plot (Figure 4.5 a), the first principal component explaining the largest percentage of variance (82%) separates the samples according to the species divide, while the second component (11.7% variance) results in separation of the samples according to the temperature; they both combine to form four sample clusters on the plot representing the four experimental groups, as expected. Intriguingly, the variance attributed to the temperature can be seen to be much smaller in the C_4 species *A. pungens* than in the C_3 species *A. angustifolia*, suggesting more plasticity in the transcriptomic response to temperature in the latter. This is also evident in PCA plots when each species is analysed independently: the second principal component separating the samples according to the temperature explains 6.1% gene expression variance in *A. angustifolia*, while only 2.6% in *A. pungens* (data not shown). On the other hand, there is a clear lack of consistency of gene expression profiles from plants grown at different temperatures in *H. peploides* and *S. globosa* (Figure 4.5 4.5a). In *Honckenya*, only the two samples from plants grown at 25°C appear to show consistent gene expression pattern and cluster together, while those grown at 18°C and 30°C are intermixed. In *Schiedea*, while the two replicates from the 25°C and 30°C treatments are clearly delimited, the two 18°C replicates do not show a consistent response. Due to limited replication (two replicates) per treatment in the *H. peploides-S. globosa* comparison, it is impossible to establish which sample is more representative of the population of plants grown at a given temperature. Therefore, differential expression analysis including the affected samples will not be reliable and, unfortunately, had to be

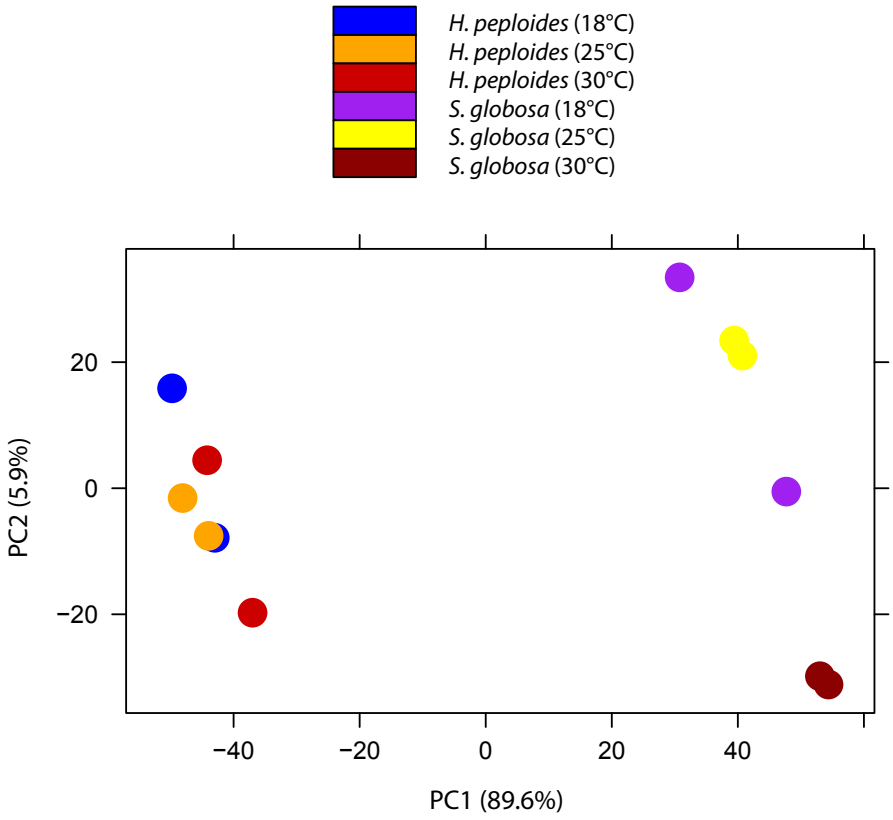
abandoned for this species pair.

In total, 10,578 and 10,116 genes were analysed in intraspecific comparisons within *A. angustifolia* and *A. pungens*, respectively. For interspecific *A. angustifolia*-*A. pungens* comparison, 8,576 genes with the same reference annotation were compared. Most contrasts yielded thousands of differentially expressed genes and so in order to limit the number of genes to those whose changes in gene expression are more likely to have a meaningful biological effect, it was thought best to retain only those with minimum 2-fold change and false discovery significance threshold of 0.01. In agreement with smaller differences in temperature response in *A. pungens* than in *A. angustifolia* seen in the PCA plot, intraspecific comparisons of samples in each species between 18°C and 30°C revealed only 9 downregulated genes at 30°C in *A. pungens* and 110 in *A. angustifolia*; none of them matching (Figure 4.6 b). Similarly, there were about ten times more upregulated genes at 30°C in *A. angustifolia* than in *A. pungens*: 226 and 21, respectively; five genes were shared, including two trypsin inhibitor genes (Figure 4.6 a). The apparent lack of plasticity in acclimation of gene expression to temperature in *A. pungens* could manifest itself finally in the 71% reduction in its biomass when grown at 18°C, as opposed to just 15% reduction in *A. angustifolia* (Figure 4.1).

Comparison of the two species' transcriptomes at each temperature, 18°C and 30°C, revealed a number of genes showing the same trend in expression at both temperatures: 163 upregulated and 207 downregulated in *A. pungens* relative to *A. angustifolia* (Figure 4.7). These genes are likely responsible for constitutive differences between the two species, including that of different modes of photosynthesis, as can be seen in Tables 4.4 and 4.5, which show the alterations in the expression of C_4 cycle genes. However, the same order of magnitude of gene number shows upregulation or downregulation only at a single experimental temperature, indicating their greater importance for plant acclimation. Amongst genes upregulated in *A. pungens* in comparison to *A. angustifolia*, 144 were only DE at 18°C and 219 at 30°C (Figure 4.7 4.7b); for downregulated genes the numbers were: 169 and 233, accordingly (Figure 4.7 4.7a).

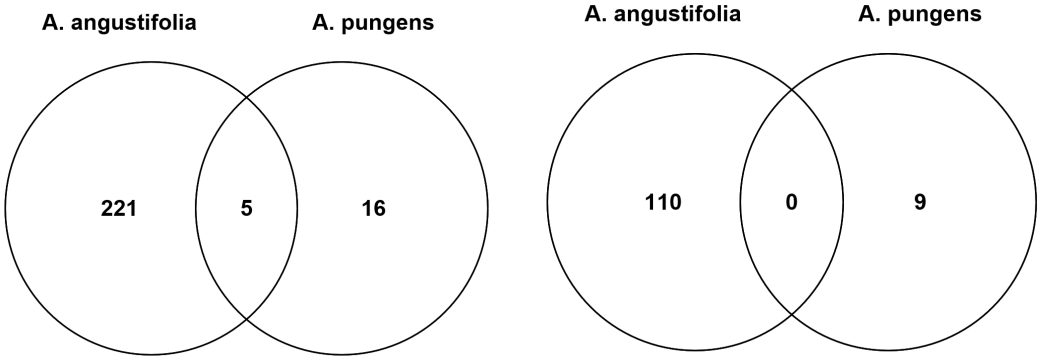


(a) *A. angustifolia* - *A. pungens*



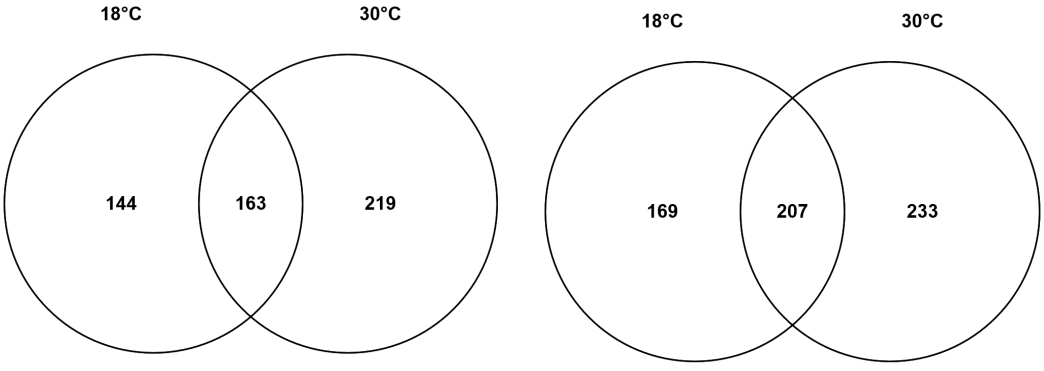
(b) *H. peploides* - *S. globosa*

Figure 4.5: PCA plot of variance of normalised expression of top 5,000 DEG in (a) *A. angustifolia* and *A. pungens*, (b) *H. peploides* and *S. globosa* comparison.



(a) Upregulated genes at 30°C **(b)** Downregulated genes at 30°C

Figure 4.6: Overlap of (a) upregulated and (**b**) downregulated genes in response to temperature in the two *Alternanthera* species in this study.



(a) Upregulated genes in *A. pungens* **(b)** Downregulated genes in *A. pungens*

Figure 4.7: Overlap of (a) upregulated and (**b**) downregulated genes in *A. pungens* relative to *A. angustifolia* at 18°C and 30°C

Table 4.4: Expression of genes with well established functions in C_4 photosynthesis in *A. pungens* and *A. angustifolia* grown at 18°C. Gene expression values are given in RPKM. Entries for genes significantly ($p < 0.01$) upregulated in *A. pungens* are printed in bold.

B. vulgaris ID	Gene name and description	AP	AA	Ratio AP/AA	Log fold-change
C_4 cycle					
Bv7_172270_kwer.t1	**Alanine aminotransferase 2 (AlaAT2)**	**289**	**61**	**4.7**	**0.68**
Bv9_214310_aeyz.t1	Aspartate aminotransferase 1 (ASP1), mitochondrial	77	86	0.9	-0.05
Bv8_201100_ezcs.t1	Aspartate aminotransferase 2 (ASP2), cytoplasmic	184	52	3.5	0.55
Bv2_031830_twkf.t1	**Aspartate aminotransferase 3 (AST3), chloroplastic**	**631**	**83**	**7.6**	**0.88**
Bv1_006100_wmrj.t1	NAD-dependent malic enzyme 59 kDa isoform (NAD-ME), mitochondrial	45	183	0.2	-0.61
Bv1u_019880_ofkj.t1	NAD-dependent malic enzyme 65 kDa isoform (NAD-ME), mitochondrial	70	54	1.3	0.11
Bv_33060_gkqg.t1	**NADP-dependent malate dehydrogenase (NADP-MDH)**	**2250**	**249**	**9.0**	**0.96**
Bv9u_231540_zmjw.t1	**NADP-dependent malic enzyme (NADP-ME)**	**1635**	**63**	**26.0**	**1.41**
Bv8_201720_sfck.t1	NADP-dependent malic enzyme (NADP-ME)	112	114	1.0	-0.01
Bv9_215080_xeaz.t1	**Phosphoenolpyruvate carboxylase 1 (PEPC1)**	**3024**	**360**	**8.4**	**0.92**
Bv4_084890_ihek.t1	Phosphoenolpyruvate carboxylase 2 (PEPC2)	61	114	0.5	-0.27
Bv1_015500_fjqs.t1	**Pyruvate, phosphate dikinase 2 (PPDK2), chloroplastic**	**5689**	**90**	**63.2**	**1.80**
Bv8_190370_jqid.t1	**Pyruvate, phosphate dikinase regulatory protein (PPDK RP), chloroplastic**	**115**	**33**	**3.5**	**0.54**
Transport					
Bv4_072820_xjai.t1	**2-oxoglutarate/malate translocator 1 (DIT1), chloroplastic**	**199**	**39**	**5.1**	**0.71**
Bv5_115710_yhsk.t1	**Triose phosphate/phosphate translocator 1 (PPT1), chloroplastic**	**983**	**121**	**8.1**	**0.91**

AP: *Alternanthera pungens*
AA: *Alternanthera angustifolia*

Table 4.5: Expression of genes with well established functions in C_4 photosynthesis in *A. pungens* and *A. angustifolia* grown at 30°C. Gene expression values are given in RPKM. Entries for genes significantly ($p < 0.01$) upregulated in *A. pungens* are printed in bold.

B. vulgaris ID	Gene name and description	AP	AA	Ratio AP/AA	Log fold-change
C_4 cycle					
Bv7_172270_kwer.t1	Alanine aminotransferase 2 (AlaAT2)	239	52	4.6	0.66
Bv9_214310_aeyz.t1	Aspartate aminotransferase 1 (ASP1), mitochondrial	49	58	0.8	-0.07
Bv8_201100_ezcs.t1	Aspartate aminotransferase 2 (ASP2), cytoplasmic	184	57	3.2	0.51
Bv2_031830_twkf.t1	**Aspartate aminotransferase 3 (AST3), chloroplastic**	**444**	**66**	**6.7**	**0.83**
Bv1_006100_wmrj.t1	NAD-dependent malic enzyme 59 kDa isoform (NAD-ME), mitochondrial	46	111	0.4	-0.38
Bv1u_019880_ofkj.t1	NAD-dependent malic enzyme 65 kDa isoform (NAD-ME), mitochondrial	65	53	1.2	0.09
Bv_33060_gkqg.t1	**NADP-dependent malate dehydrogenase (NADP-MDH)**	**1700**	**186**	**9.1**	**0.96**
Bv9u_231540_zmjw.t1	**NADP-dependent malic enzyme (NADP-ME)**	**1123**	**98**	**11.5**	**1.06**
Bv8_201720_sfck.t1	NADP-dependent malic enzyme (NADP-ME)	71	27	2.6	0.42
Bv9_215080_xeaz.t1	**Phosphoenolpyruvate carboxylase 1 (PEPC1)**	**2526**	**337**	**7.5**	**0.87**
Bv4_084890_ihek.t1	Phosphoenolpyruvate carboxylase 2 (PEPC2)	59	62	1.0	-0.02
Bv1_015500_fjqs.t1	**Pyruvate, phosphate dikinase 2 (PPDK2), chloroplastic**	**4472**	**71**	**63.0**	**1.80**
Bv8_190370_jqid.t1	**Pyruvate, phosphate dikinase regulatory protein (PPDK RP), chloroplastic**	**166**	**32**	**5.2**	**0.71**
Transport					
Bv4_072820_xjai.t1	2-oxoglutarate/malate translocator 1 (DIT1), chloroplastic	145	173	0.8	-0.08
Bv5_115710_yhsk.t1	**Triose phosphate/phosphate translocator 1 (PPT1), chloroplastic**	**790**	**88**	**9.0**	**0.95**

AP: *Alternanthera pungens*
AA: *Alternanthera angustifolia*

4.3.4 Exploring functional categories of genes important for temperature response in C_4 and C_3 photosynthesis

4.3.4.1 C_4 photosynthesis genes

C_4 plants express an additional suite of enzymes specific to the C_4 cycle, and a significantly higher abundance of core enzymes of the NADP-ME subtype of C_4 photosynthesis was expected to be found in *A. pungens* than in *A. angustifolia* regardless of the temperature. Indeed, as Tables 4.4 and 4.5 demonstrate, key C_4 enzymes: alanine aminotransferase, aspartate aminotransferase, NADP-MDH, NADP-ME and PEPC displayed constitutively higher expression (4.7-26.0 fold) in *A. pungens* at both temperatures, with no significant change across the temperature treatments. Amongst other genes related to C_4 photosynthesis, very high expression of *PYRUVATE, PHOSPHATE DIKINASE 2* (*PPDK2*) - 63 times higher in *A. pungens* than in *A. angustifolia* and the transporter *TRIOSE PHOSPHATE/PHOSPHATE TRANSLOCATOR* (*PPT1*) was found. It seems that *A. pungens* uses only (or overwhelmingly) NADP-ME subtype C_4 photosynthesis, as NAD-ME expression was low, and in fact was higher in *A. angustifolia*. Interestingly, *DIT1* was upregulated in *A. pungens* relative to *A. angustifolia* at 18°C but not at 30°C due to a sudden spike in gene expression in *A. angustifolia* at 30°C (possibly due to DIT1's role in the 'malate valve'), which is still nevertheless consistent with its proposed role in C_4 photosynthesis transport as a chloroplast OAA/malate shuttle in the mesophyll (Gowik *et al.*, 2011; Kinoshita *et al.*, 2011), since the expression of the gene is moderately high and stable in the C_4 species. Overall, expression of C_4 photosynthesis genes is in agreement with expectation in *A. pungens* and *A. angustifolia*, which increases confidence in the expression results for other genes, for which there is no such clear expectation.

4.3.4.2 Temperature acclimation in a C_3 plant

Intraspecific comparison in *A. angustifolia* and *A. pungens* across the two growth temperatures revealed very limited remodelling of the transcriptome in the C_4 species *A. pungens*, as opposed to *A. angustifolia* (Figure 4.6). A number of gene categories were more highly expressed in *A. angustifolia* at 30°C than at 18°C: genes involved in photosynthesis, heat and oxidative stress and cell wall biosynthesis. In the first category, 2 ribulose-bisphosphate

carboxylase small subunit genes and 3 Rubisco activase genes were significantly upregulated at 30°C, which can probably be explained by increased Rubisco catalytic rate and thus turnover, as well as decreased Rubisco activase activity at higher temperatures.

Similarly, a number of genes typically upregulated under the condition of heat stress were found to be more highly expressed at 30°C than at 18°C in *A. angustifolia* but not in *A. pungens*: 5 heat shock protein binding genes (including 4 heat shock 40 proteins), 5 genes involved in response to heat - *MULTIPROTEIN BRIDGING FACTOR 1C* (*MBF1C*), *ANNEXIN 4*, *HEAT SHOCK FACTOR 4* (*HSF 4*). In addition to 4 HSP40 chaperones guiding protein folding and refolding important for mitigating the effects of stress, such as high temperatures which promote protein damage and denaturation (Sun *et al.*, 2002), *HSF40*, a transcriptional co-activator which belongs to the family of 20 genes binding to heat stress elements in the promoters of heat shock proteins, was found. *HSF40* itself contributes to the development of acquired thermotolerance in the face of prolonged heat stress, but is also a positive regulator in the systemic acquired resistance defence response pathway (Ikeda *et al.*, 2011; Pick *et al.*, 2012). *MBF1C* acts as a transcriptional co-activator in the ethylene-response signal transduction pathway, which promotes tolerance to heat and osmotic stress (Suzuki *et al.*, 2005). Lastly, *ANNEXIN4* belongs to a class of calcium- and lipid- binding proteins, and along with *ANNEXIN1*, it has been shown to be a part of osmotic stress signalling pathway and act in a photoperiod-specific manner (Huh *et al.*, 2010).

More broadly, oxidation-reduction process-related genes were overrepresented (18 genes) among genes upregulated at 30°C compared to 18°C in *A. angustifolia*, including 6 cytochrome P450 genes, *ALKENAL/ONE OXIDOREDUCTASE* (*AOR*), *MYO-INOSITOL OXYGENASE 4* (*MIOX4*), and a gene with L-ascorbate oxidase function. Cytochrome P450 proteins catalyse a diverse set of oxidative reactions in metabolism of plant compounds and xenobiotics (Morant *et al.*, 2003), and their expression is highly responsive to abiotic stresses (Narusaka *et al.*, 2004), which may be one reason for their high expression at 30°C. AOR is a chloroplast-localised enzyme, which catalyzes the reduction of α,β-unsaturated bond in

reactive carbonyls in the chloroplast stroma, such as formed during lipid peroxidation. It is essential for countering oxidative stress in the chloroplast and preventing photosystem I inhibition (Yamauchi *et al.*, 2012). MIOX4 is an enzyme involved in an alternative L-ascorbic acid biosynthesis pathway in plants (Lorence *et al.*, 2004; Valpuesta & Botella, 2004). L-ascorbate oxidases oxidize ascorbic acid to monodehydroascorbate, which in turn is unstable and quickly disproportionates into dehydroascorbate, and again ascorbic acid (Pignocchi *et al.*, 2003). Ascorbate is one of the most important antioxidants, vital for defence against oxidative stress due to its role in the glutathione-ascorbate cycle, and it is therefore intriguing to see both catabolic and anabolic ascorbate pathway genes upregulated at elevated temperature.

Two additional functional groups stood out: 9 genes involved in carbohydrate metabolism and 12 transmembrane transporters. One of the genes in the first category is *XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE 23* (*XTH-23*), which belongs to a group of cell-wall modifying enzymes whose activity is needed for cell expansion (Miedes *et al.*, 2013). Another gene related to cell wall homeostasis was a caffeic acid 3-O-methyltransferase, which catalyzes methylation reactions of lignin precursors (Ma & Xu, 2008). Focusing on the transporters upregulated at 30°C in *A. angustifolia*, 2 sulfate transporters, 3 oligopeptide transporters and the aquaporin *PLASMA MEMBRANE INTRINSIC PROTEIN 2;1* (*PIP2;1*) were the subgroups discovered. Aquaporins form integral membrane water channels and *PIP2;1* expression is essential for increased guard cell permeability to water required by ABA-mediated stomatal closure triggered by heat and drought (García-Mata & Lamattina, 2001; Grondin *et al.*, 2015).

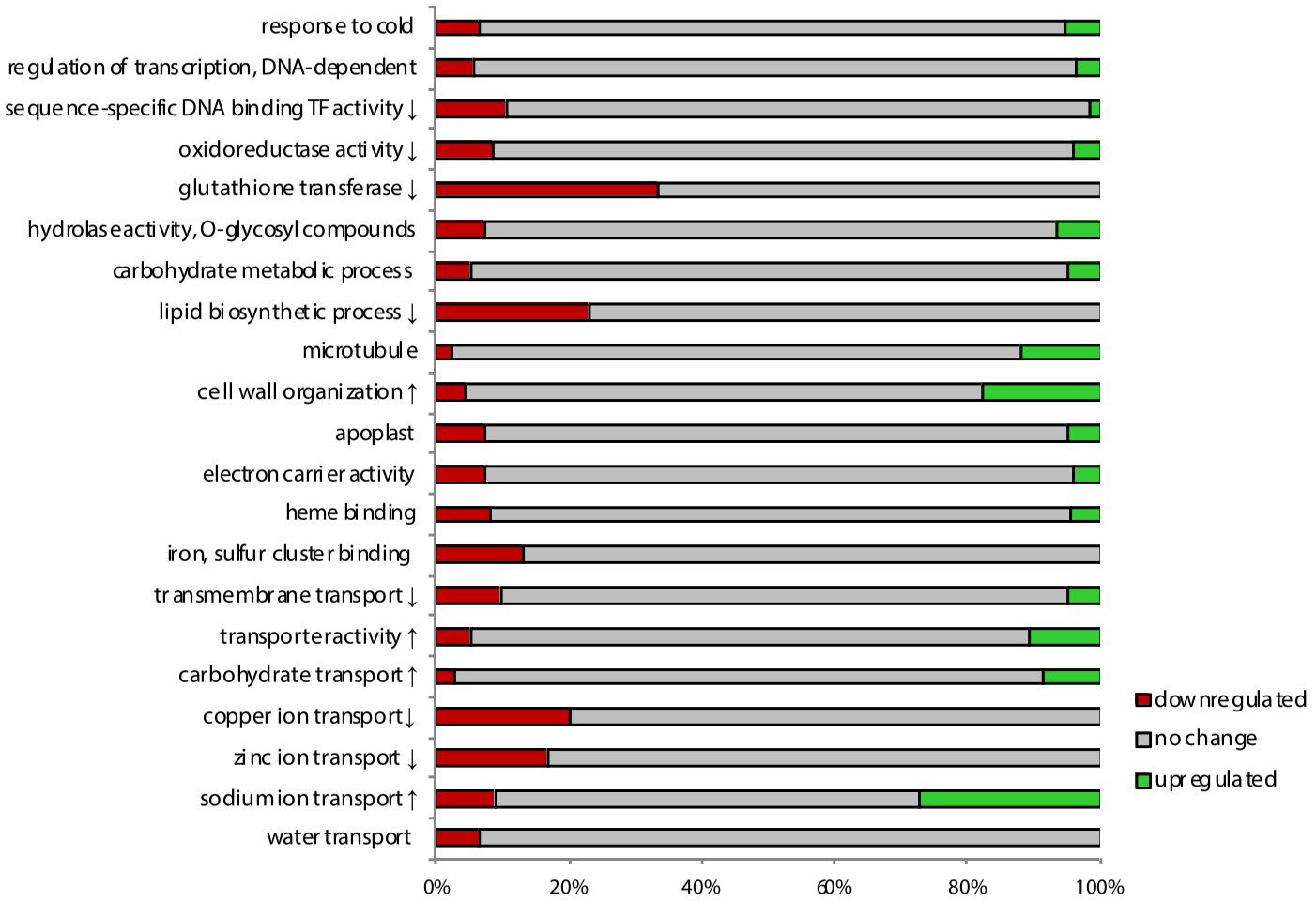


Figure 4.8: Select gene ontology terms showing significant over-representation amongst genes upregulated (indicated by ↑) or downregulated (indicated by ↓) in *A. pungens* relative to *A. angustifolia* at 18°C.

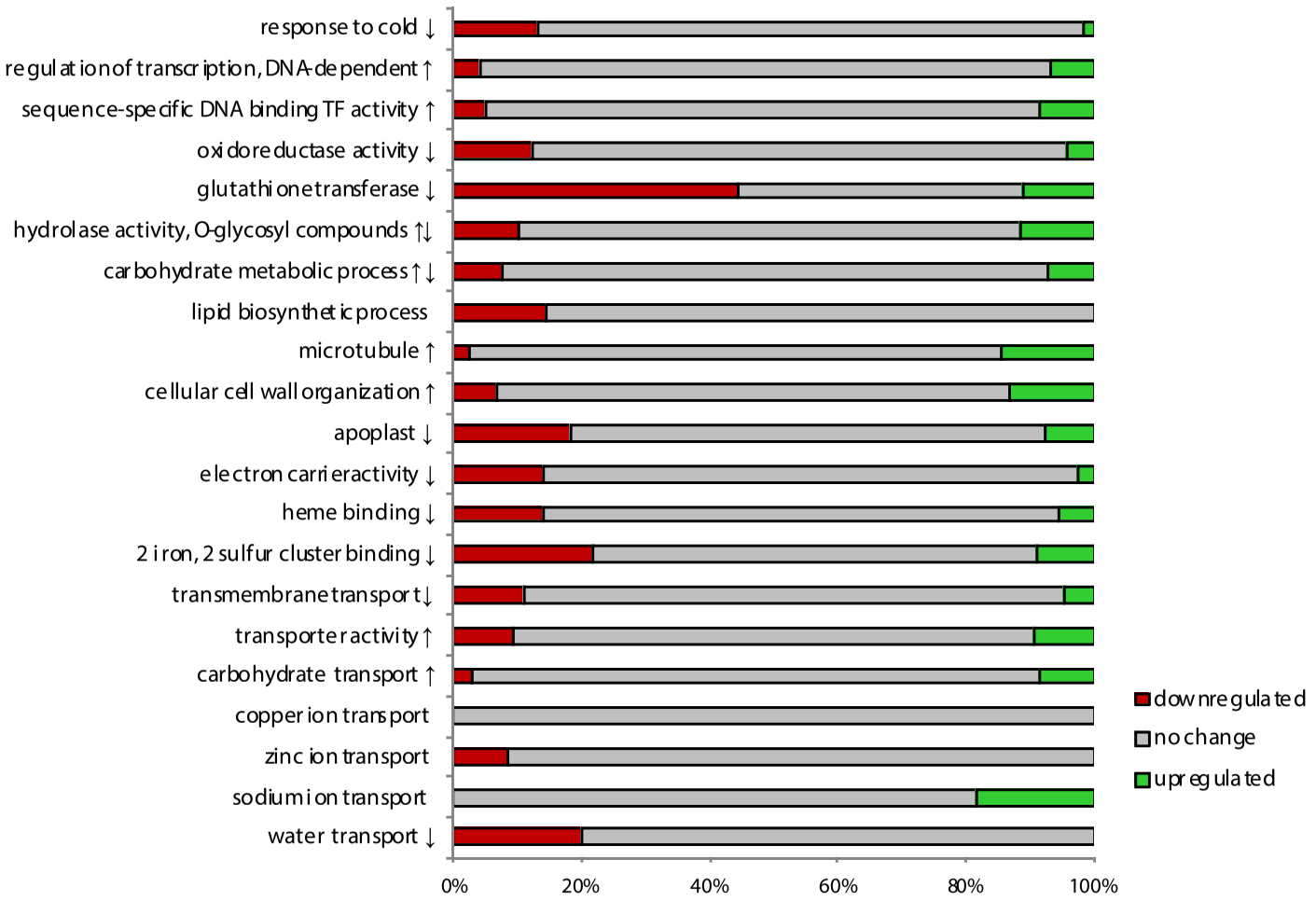


Figure 4.9: Select gene ontology terms showing significant over-representation amongst genes upregulated (indicated by ↑) or downregulated (indicated by ↓) in *A. pungens* relative to *A. angustifolia* at 30°C.

4.3.4.3 Interspecific differences between C_4 and C_3 plants in gene expression: constitutive and plastic

A number of gene ontology terms were significantly enriched amongst differentially expressed genes when conducting cross-species comparison of transcriptomes of *A. angustifolia* and *A. pungens* at either 18°C (Figure 4.8) or 30°C (Figure 4.9). In the following, selected genes in the categories hypothesised to play a role in both constitutive (i.e. found at both temperatures) interspecific differences and the plastic differences present just in acclimation to either 18°C or 30°C between the C_3 and the C_4 species are described in functional groupings. In all comparisons here after, gene expression level in *A. pungens* is compared with that of *A. angustifolia*. In general, heat and oxidative stress response genes as well as many genes involved in chloroplast biogenesis, function and turn-over were downregulated in *A. pungens* relative to *A. angustifolia*. On the other hand, genes involved in cell wall formation and modification were constitutively upregulated in the species.

Photosynthesis and chloroplast function Two genes known to participate in the photosystem repair were downregulated: *DEGRADATION OF PERIPLASMIC PROTEINS 8 (DEG8*, both at 18°C and 30°C) and *DEGRADATION OF PERIPLASMIC PROTEINS 5 (DEG5*, only at 18°C). These two proteases form a hexamer in the thylakoid lumen and cleave damaged protein D1 in the Photosystem II and thus are needed for the repair of Photosystem II following the exposure to high light and photoinhibition (Sun *et al.*, 2007). Another two genes important for protection against photoinhibition were downregulated at 18°C: chloroplastic *CHLOROPHYLL A OXYGENASE (CAO)* and *PROTON GRADIENT REGULATION 5 (PGR5)*. *CHLOROPHYLL A OXYGENASE* converts chlorophyllide a to chlorophyllide b (Tanaka *et al.*, 1998) which is required for formation of chlorophyll-protein antenna complexes around Photosystem II and adaptation to different levels of light intensity and resistance to photoinhibition. *PGR5* is needed for electron transfer in one of cyclic electron flow systems around Photosystem I and PSI protection from fluctuating light (Suorsa *et al.*, 2012). Finally, two photorespiration-related genes found to be downregulated both at 18°C and 30°C were *L-SERINE:GLYOXYLATE AMINOTRANSFERASE (SGAT)*,

which encodes an enzyme functioning in photorespiratory nitrogen cycling through performing transamination reactions (Kendziorek & Paszkowski, 2008) and *AOR*, mentioned in the previous section. Two genes were downregulated just at 18°C: *FERREDOXIN-DEPENDENT GLUTAMATE SYNTHASE 1 (GLU1)* and *GLYCOLATE OXIDASE 1 (GOX1)*. *GLU1* has been found to be important for detoxification of excess ammonium produced during photorespiration (Potel *et al.*, 2009; Jamai *et al.*, 2009) and *GOX1* is one of the GOX enzymes participating in photorespiration that oxidizes 2-phosphoglycolate to glyoxylate in the peroxisome, with concomitant production of hydrogen peroxide (Fahnenstich *et al.*, 2008; Rojas & Mysore, 2012). In contrast, two enzymes in the Calvin cycle were downregulated at 30°C as expected due to more constant photosynthesis rate in C_4 species: *GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE B SUBUNIT (GAPB)* and *PHOSPHORIBULOKINASE (PRK)* (Marri *et al.*, 2009). Different subunits of chloroplastic ATP synthase changed expression according to temperature: *ATP SYNTHASE DELTA CHAIN (ATPD)* and *ATP SYNTHASE CF1 BETA SUBUNIT* were constitutively upregulated, while *ATP SYNTHASE SUBUNIT BETA (ATPB)* was downregulated at 18°C.

Two genes involved in chlorophyll catabolism and more generally, leaf senescence, were downregulated at 18°C: *CHLOROPHYLLASE 2 (CLH2)* and *PHEOPHORBIDE A OXYGENASE (PAO)*. CLH2 is thought to catalyze the initial step in chloroplast-localised chlorophyll degradation pathway: hydrolysis of the ester bond to yield chlorophyllide and phytol (Lin *et al.*, 2014), while PAO catalyzes one of the later steps in the degradation of chlorophyll to an intermediate fluorescent chlorophyll catabolite (Sakuraba *et al.*, 2012). Lastly, two genes participating in chloroplast protein production were downregulated at 18°C and 30°C: *CHLOROPLAST RNA BINDING (CRB)*, and a chloroplastic elongation factor G gene. *CRB* along with its homolog hold roles in chloroplast rRNA metabolism and transcription, including stabilisation of essential chloroplast rRNAs (Qi *et al.*, 2012). Another downregulated gene important for expression of chloroplast genes was a chloroplastic elongation factor G GTPase, which catalyzes translocation of mRNA and tRNA during translation of chloroplast genes (Albrecht *et al.*, 2006).

Response to oxidative stress A number of genes important for cellular redox homeostasis were differentially expressed: 3 peroxidases were downregulated at 18°C, while 3 other peroxidase genes were upregulated at 30°C. Peroxidases utilize either hydrogen peroxide or molecular oxygen to oxidise a number of substrates and play a variety of roles during plant development, including roles in cell wall modification (Yoshida *et al.*, 2003). Two classes of genes with a variety of roles in catabolism and detoxification were downregulated at 30°C: 5 cytochrome P450 genes and 4 glutathione transferase genes, 3 of which were downregulated at 18°C. Genes with glutathione transferase function were significantly more often downregulated than expected by chance at both 18°C and 30°C (Figure 4.8 and 4.9). Glutathione transferases are ubiquitous proteins whose glutathione-dependent catalytic reactions are primary conjugation and detoxification of xenobiotics; glutathione transferases are also associated with responses to a number of abiotic stresses that generate redox imbalanace (Dixon *et al.*, 2010).

Proteins containing an iron-sulfur cluster were significantly overrepresented among genes downregulated at 30°C: chloroplastic choline monooxygenase, *GLUTAREDOXIN-C9 (GRXC9)*, *PGR5* and *CAO*. Three genes were also downregulated at 18°C (constitutively downregulated): amongst them a ferredoxin protein and *GLUTAREDOXIN 4 (GRX4)*. Choline monooxygenases are a type of oxidoreductases which convert choline to betaine aldehyde hydrate using ferredoxin as a reducing agent. Betaine aldehyde is an intermediate in the biosynthesis of glycine betaine, an important plant osmoprotectant (Yamada *et al.*, 2015). On the other hand, chloroplastic *BETAINE ALDEHYDE DEHYDROGENASE* which oxidises glycine betaine aldehyde into the osmoprotectant glycine betaine in the second step of glycine betaine biosynthesis was upregulated at both temperatures. This gene was shown to be upregulated during drought and salt stress in other plants. It is not known if glycine betaine is produced by *Alternanthera*; however, the compound was in fact discovered in sugar beet in the same order, so this is likely. In *Arabidopsis*, which does not accumulate glycine betaine, *BETAINE ALDEHYDE DEHYDROGENASE* has been implicated in stress tolerance as a detoxification enzyme controlling the level of aminoaldehydes (Missihoun *et al.*, 2011). Glutaredoxins are a type of disulfide oxidoreductases important for

cellular protection against oxidative stress. *GRXC9* is a chloroplast-localised glutaredoxin guarding against protein carbonylation within chloroplasts (Cheng *et al.*, 2006), while *GRX4* participates in salicylic acid/jasmonic acid cross-talk (Ndamukong *et al.*, 2007).

Oxido-reductases A number of oxido-reductase genes were downregulated at both temperatures - 23. However, one category of oxido-reductases was upregulated at 18°C - those acting on the CH-OH group of donors and using NAD or NADP as acceptor, such as NADP-dependent *ISOCITRATE DEHYDROGENASE (ICDH)*, which catalyzes the production of the respiratory intermediate 2-oxoglutarate (Boex-Fontvieille *et al.*, 2013). Downregulated oxidoreductases at 30°C included: *1-AMINOCYCLOPROPANE -1-CARBOXYLATE OXIDASE 3 (ACO4)*, two omega-3 fatty acid desaturases (chloroplastic and endoplasmic reticulum-located), *FERREDOXIN-DEPENDENT GLUTAMATE SYNTHASE 1 (GLU1)*, *GLYCOLATE OXIDASE 1 (GOX1)*. *ACO4* belongs to a group of oxidases catalysing the last committed step in the ethylene biosynthesis pathway (Gómez-Lim *et al.*, 1993). Omega-3 fatty acid desaturases' role is introduction of the third double bond in the synthesis of 16:3 and 18:3 fatty acids which could be of importance in adaptation to heat stress in *A. angustifolia*. *FERRITIN 3 (FER3)* was found among upregulated oxido-reductases at 30°C - unlike mammal ferritins, plant ferritins do not appear to constitute a major Fe sink but are needed for protection against oxidative stress (Ravet *et al.*, 2009**a**). *SENESCENCE-RELATED GENE 1 (SRG1)*, a member of the family and marker for senescence (Callard *et al.*, 1996), as well as a protein with a putative monodehydroascorbate reductase function in the glutathione-ascorbate cycle, were constitutively upregulated at both temperatures.

Cell wall formation Regarding cell wall biosynthesis genes, their upregulated expression was predicted in the C_4 species, given they may be needed to produce the Krantz anatomy through the structural reinforcement of bundle sheath cells. Indeed, genes labelled with GO term *cell wall organisation* were enriched for among upregulated genes constitutively at both temperatures (Figure 4.8 and 4.9). Among upregulated genes at both temperatures was *CELLULOSE SYNTHASE 8 (CESA8)*, whereas only at 18°C: *CELLULOSE SYNTHASE LIKE D3 (CSLD3)*, *GLUCAN SYNTHASE LIKE 1 (GSL1)* and *GLUCAN SYNTHASE-LIKE*

5 (GSL5), and just at 30°C: an endo-1,4-β glucanase. CESA8 is one of the three isozymes together responsible for cellulose biosynthesis in cellulose synthase complex and thus required for secondary cell wall formation (Hill *et al.*, 2014). On the other hand, *CELLULOSE SYNTHASE 3 (CESA3)*, which together with two other isozymes forms a complex for the synthesis of cellulose in the primary cell wall (Gonneau *et al.*, 2014), was downregulated at 30°C. Interestingly, in cases of both primary and secondary cell wall cellulose synthase complexes, the different enzyme subunits are found in 1:1:1 ratio, but no coordinated change in their expression was observed here. Cellulose and xyloglucan components of the cell wall contain β-1,4-glycosidic bonds that can be cleaved by endo-1,4-β glucanases, which are believed to play numerous roles in cell wall biosynthesis and modification (Glass *et al.*, 2015). *CSLD3* encodes a glycosyl transferase important for cellulose and xyloglucan organisation in the root hair cell walls (Galway *et al.*, 2011). The two glucan synthase genes *(GSL1, GSL5)* found participate in the synthesis of the (1,3)-β-glucan callose which is deposited as cell wall thickenings to prevent fungal infection (Ellinger *et al.*, 2013, 2014). Among constitutively downregulated genes associated with cell wall organisation were also two xyloglucan-endotransglucosylase, the first of which is responsible for cell wall loosening for leaf cell growth: *XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE 8 (XTH8)* and *XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE 23 (XTH23)* (Miura *et al.*, 2010). *XTH-23* was also upregulated at 30°C relative to 18°C in *A. angustifolia*.

Wax provides a mechanical barrier important for drought, UV light and pathogen resistance. Furthermore, cell walls of bundle sheath cells are impregnated with waxy suberin or reinforced with lignin, which helps to establish a barrier against CO_2 diffusion out of bundle sheath cells (Evert *et al.*, 1977; Wang *et al.*, 2014). Contrary to the expectation of high expression of wax biosynthesis genes, three genes involved in wax biosynthesis were downregulated at 18°C: *ECERIFERUM 3 (CER3)*, *WAX INDUCER 1 (WIN1* (also downregulated at 30°C), *3-KETOACYL-COA SYNTHASE (KCS1). CER3* is a transmembrane protein which promotes biosynthesis of very-long-chain alkane components of the cuticular wax (Mao *et al.*, 2012), while *KCS1* is a key enzyme in elongating fatty acids in wax biosynthesis (Ghanevati & Jaworski, 2001). *WIN1* is an ERF/AP2 transcription factor which promotes

expression of cutin biosynthesis genes and wax accumulation (Oshima *et al.*, 2013).

Transcription factors Whereas at 18°C there was an over-representation (Figures 4.8 and 4.9) of downregulated sequence-specific DNA binding transcription factors (15 genes), at 30°C this category was among the top represented amongst upregulated genes (12 genes). The more general category *regulation of transcription, DNA-dependent*, was also enriched for among upregulated genes at 30°C with 29 genes, and although not significant, the number of downregulated genes in this category was bigger than that of upregulated by a third at 18°C. Constitutively upregulated or downregulated transcription factors could be important for generating the differences in phenotypes between the two species, including the mode of photosynthesis.

Five constitutively downregulated transcription factors were found. Among them, homeobox-leucine zipper protein HAT2 is an auxin responsive factor which belongs to class II homeodomain-leucine zipper (HD-ZIPII) transcription factor family, which is known to participate in regulation of leaf morphogenesis, in particular shade avoidance response (Ciarbelli *et al.*, 2008). Moreover, previously uncharacterised, constitutively downregulated genes included *GATA9*, which along with the other members of the GATA family of obscure transcriptional activators, may be involved in regulation of light-responsive genes (Manfield *et al.*, 2007): *GATA6* and *GATA24* were also found amongst transcription factors downregulated only at 18°C, while *GATA7* was upregulated at 30°C. Another family of transcription factors was also heavily represented among differentially expressed genes: *WRKY40* was constitutively downregulated, *WRKY70* downregulated only at 18°C, while *WRKY7* and *WRKY20* constitutively upregulated. Both WRKY40 and WRKY7 are negative regulators of plant innate immune response (Kim *et al.*, 2006); WRKY40 acts also as a repressor of abiotic stress response, including expression of stress-responsive chloroplast and mitochondrial genes, in particular in high-light-induced signalling (Xu *et al.*, 2006; Van Aken *et al.*, 2013). In contrast, WRKY70 is a positive regulator of the plant defence response, which promotes ROS-induced plant cell death (Brosché *et al.*, 2014), but is also a negative regulator of osmotic stress tolerance by repressing stomatal closure and also leaf

senescence (Besseau *et al.*, 2012; Hu *et al.*, 2012**a**; Li *et al.*, 2013). Soybean WRKY20, on the other hand, has been shown to be a positive regulator of stomatal closure and drought tolerance, in addition to wax biosynthesis genes (Luo *et al.*, 2013).

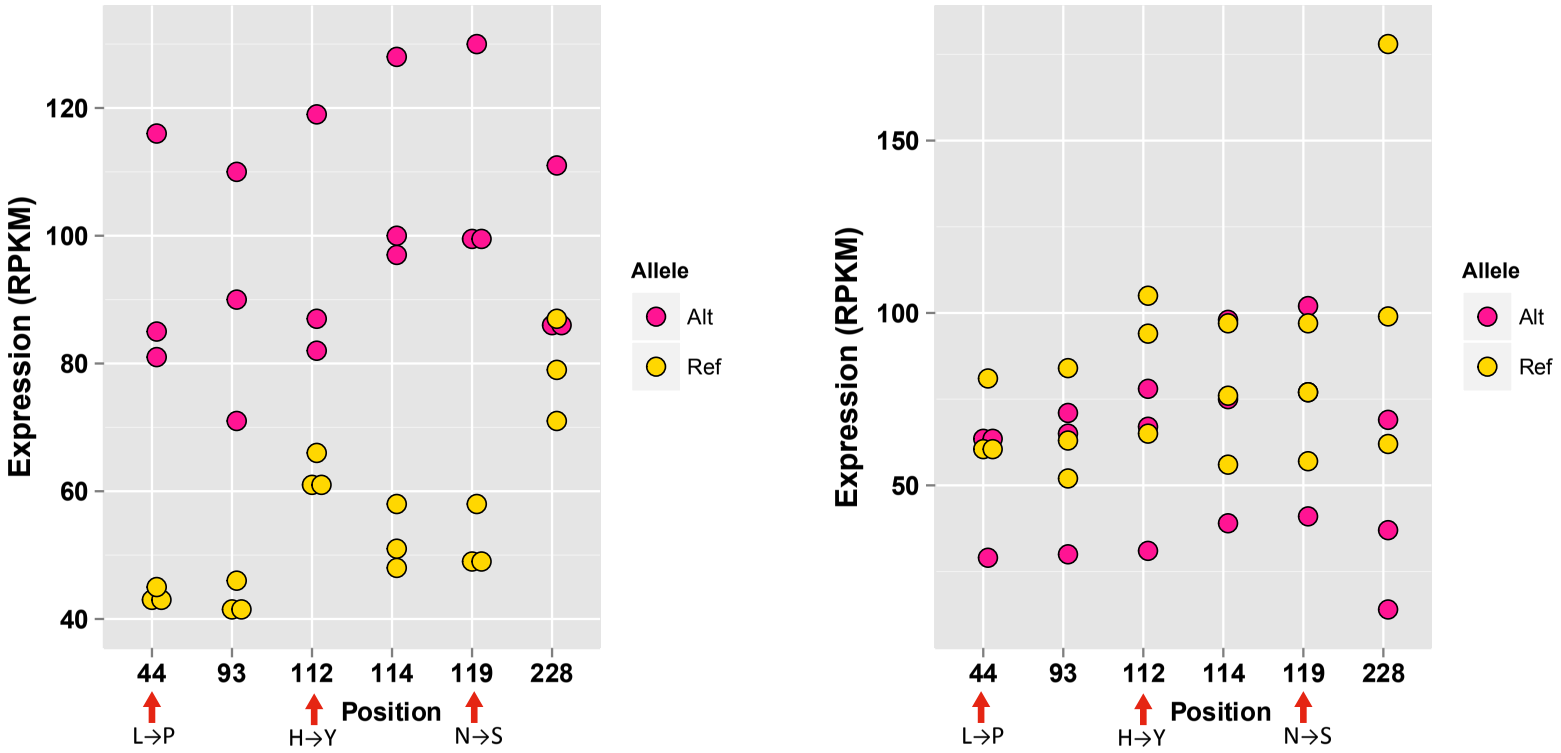
Two members of the ethylene response factor subfamily in the ERF/AP2 transcription factor family were downregulated: previously uncharacterised *ERF-1* as well as well-known *ETHYLENE RESPONSE FACTOR 72 (ERF72)*, a transcriptional activator of oxidative, osmotic and heat stress resistance in the ABA and ethylene signalling pathway (Papdi *et al.*, 2015). Among constitutively upregulated transcription factors, only 2 genes were found, one of which is *PATHOGENESIS RELATED HOMEODOMAIN PROTEIN A (PRHA)*. *PRHA* is strongly auxin-induced homeobox gene associated with regions of developing vasculature but unknown function (Plesch *et al.*, 1997).

Most transcription factors were only upregulated at 30°C and included: *HOMEOBOX PROTEIN 22 (HB22)*, *ARABIDOPSIS THALIANA HOMEOBOX 12 (ATHB-12)*. HB22 is a zinc finger homeobox protein, which can bind to the phosphoenolpyruvate carboxylase gene of *Flaveria trinervia* and has been hypothesised to regulate its mesophyll-specific expression (Windhövel *et al.*, 2001). ATHB-12 belongs to a homeodomain leucine zipper class I of proteins, and is a positive regulator of leaf cell expansion and endoreduplication during development as well as mediates negative feedback effect in drought-induced ABA signalling (Valdés *et al.*, 2012; Hur *et al.*, 2015). Interestingly, *SCARECROW (SCR)* turned up among upregulated transcription factors at 18°C, while *SHORT-ROOT (SHR)* was upregulated at 30°C. Both transcription factors are members of the GRAS family, and required for maintaining the quiescent centre and radial patterning, with *SCR* acting downstream of *SHR* (Helariutta *et al.*, 2000; Sabatini *et al.*, 2003; Koizumi & Gallagher, 2013). Crucially here, these two transcription factors seem to be conserved regulators for bundle sheath cell specification (Cui *et al.*, 2014) and further than that, *SCR* has been shown to be necessary for establishing correct Krantz anatomy in maize leaves (Slewinski *et al.*, 2012). *SCR* has also turned up as a shared candidate regulator of C_4 leaf development in *Zea mays* and *Cleome gynandra* (Aubry *et al.*, 2014). However,

none of the other transcription factors identified in the current study matched any on the shortlist of 119 candidate transcription factors from Aubry *et al.* (2014). While two transcription factor families repeatedly appeared in Aubry *et al.* (2014) and here (WRKY and ARABIDOPSIS THALIANA HOMEOBOX), the lack of convergence in the two studies might suggest recruitment of different transcription facto family members in C_4 evolution in Caryophyllales, which would be surprising given that the *Z. mays* - *C. gynandra* comparison is evolutionary very distant (monocot – eudicot).

Transporters Overall, genes involved in transmembrane transport were constitutively downregulated at both temperatures but different subcategories were also upregulated, such as carbohydrate transporters (Figures 4.8 and 4.9), e.g. constitutively upregulated *EARLY RESPONSE TO DEHYDRATION SIX-LIKE 1 (ESL1)*, *TRIOSE-PHOSPHATE/PHOSPHATE TRANSLOCATOR (TPT)*, and uncharacterised transporters in the nucleobase-ascorbate transporter group - *NUCLEOBASE-ASCORBATE TRANSPORTER 7 (NAT7)* and *NUCLEOBASE-ASCORBATE TRANSPORTER 11 (NAT11)*. ESL1 is a monosaccharide transporter induced by osmotic stress and hypothesised to regulate osmotic pressure (Yamada *et al.*, 2010). TPT is a chloroplast envelope transporter that exports triose phosphates from the Calvin cycle in the stroma to the cytosol for sucrose synthesis (Schneider *et al.*, 2002) and is also essential for acclimation to high light (Walters *et al.*, 2003; Schmitz *et al.*, 2012, 2014). This gene was expressed 8-9 fold more highly at both temperatures in *A. pungens* (Tables 4.4 and 4.5).

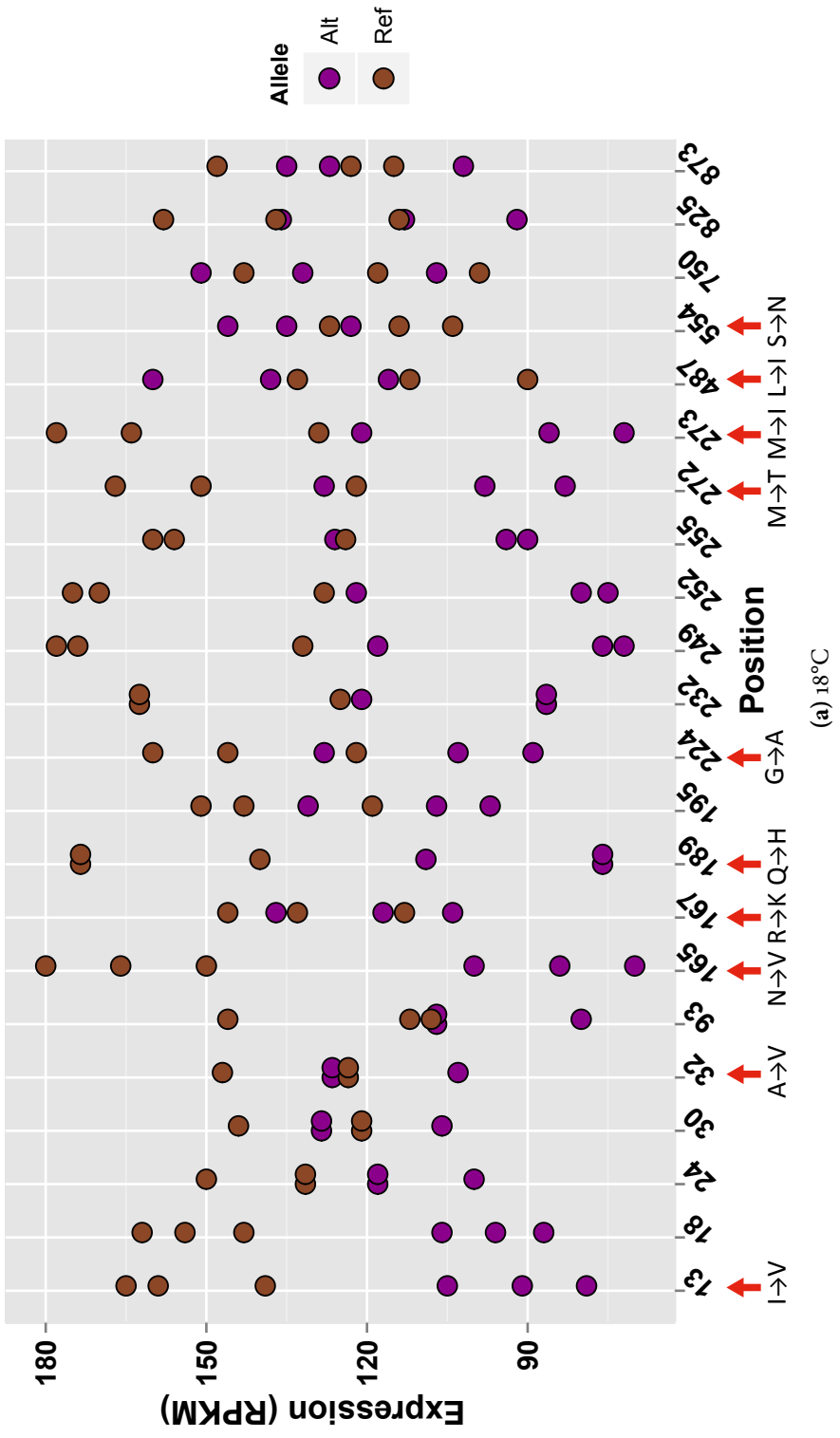
In contrast, aquaporin genes were downregulated at 30°C: a borate channel *NOD26-LIKE INTRINSIC PROTEIN 6;1 (NIP6;1*, Tanaka *et al.*, 2008), water channel *PLASMA MEMBRANE INTRINSIC PROTEIN 2;1 (PIP2;1)* - already found among genes upregulated at 30°C in *A. angustifolia*, and finally *TONOPLAST INTEGRAL PROTEIN 2;1 (TIP2;1)*, a vacuolar ammonium transporter (Loqué *et al.*, 2005), also downregulated at 18°C. Metal transport genes were downregulated at 18°C: three ZIP transporters and the copper-transporting ATPase *RAS-RELATED NUCLEAR PROTEIN 1 (RAN1)*, which is essential for ethylene signalling via the delivery of copper cofactors for ethylene receptors (Woeste & Kieber, 2000; Binder *et al.*, 2010).



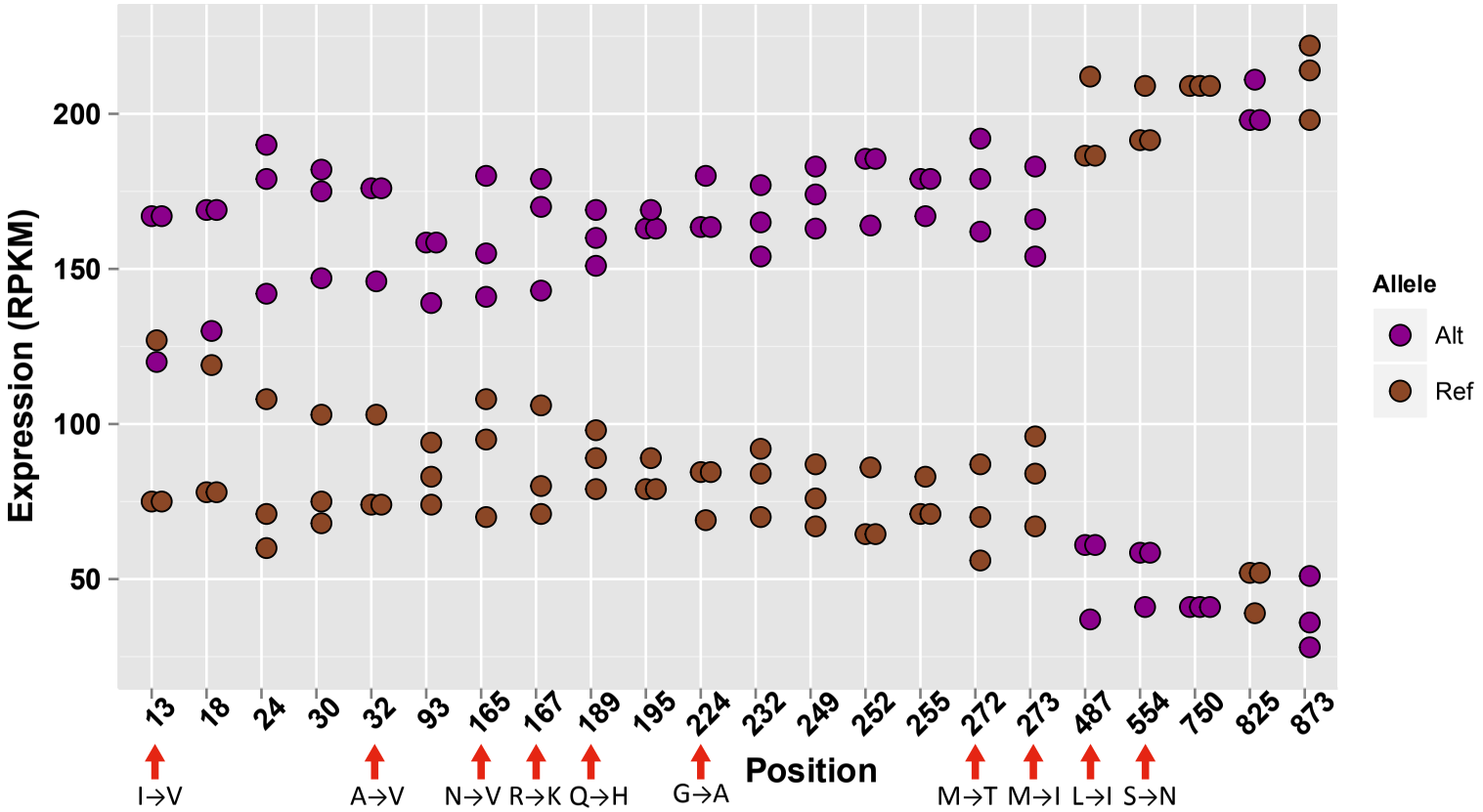
(a) 18°C

(b) 30°C

Figure 4.10: Allele-specific expression at SNPs in Rubisco activase showing opposite trends of differential expression at: (a) 18°C and (b) 30°C in *A. pungens*. Results from three replicate plants are shown per site and sites are numbered in reference to *A. thaliana* Rubisco activase coding region (AT2G39730). Amino acid changes (Ref → Alt) in non-synonymous substitutions are marked below the chart. Reference (Ref) and alternative (Alt) allele states were determined in reference to the assembly consensus sequence.


Allele
Alt
Ref
Expression (RPKM)
180
150
120
90
Position
I→V
A→V
N→V
R→K
Q→H
G→A
M→T
M→I
L→I
S→N
13
18
24
30
32
93
165
167
189
195
224
232
249
252
255
272
273
487
554
750
825
873

(a) 18°C



(a) 30°C

Figure 4.9: Allele-specific expression at SNPs in the Rubisco activase showing opposite trends of differential expression at: (a) 18°C and (b) 30°C in *A. angustifolia*. Results from three replicate plants are shown per site and sites are numbered in reference to *A. thaliana* Rubisco activase coding region (AT2G39730). Amino acid changes (Ref → Alt) in non-synonymous substitutions are marked below the chart. Reference (Ref) and alternative (Alt) allele states were determined in reference to the assembly consensus sequence.

4.3.5 Differential allele expression in Rubisco small subunit and Rubisco activase genes across temperature in *Alternanthera*

Rubisco, the most abundant protein on Earth, which constitutes 25% and upwards of leaf protein content, has a hexadecameric structure in all green plants. Rubisco is made up of eight large subunits (RbcL) arranged as a tetramer of antiparallel dimers, capped by four small subunits (RbcS) at the upper and four at the lower face (Andersson & Backlund, 2008). While not contributing to the active site, small subunits are required for structural support. Type I Rubisco found in green plants shows significant catalytic differences across plant species, especially between C_3 and C_4 species, however catalytic properties of only a handful of Rubiscos have been characterized. Therefore, the link between expressed sequence and catalytic diversity, and influence of environmental factors, such as temperature on the two has not been extensively studied (Whitney *et al.*, 2011).

Before it can carry out any catalytic reactions, Rubisco needs to be activated by carbamylation of the active site pocket which allows RuBP to be positioned in place for electrophilic attack by CO_2 (Andersson, 2008). In certain cases, because of formation of pentulose bisphosphate by-products by Rubisco misfiring or erroneous docking of RuBP before carbamylation reaction, Rubisco becomes inactive and needs to be remodelled by a chaperone Rubisco activase to release inhibitory sugars in order to regain catalytic activity (Mueller-Cajar *et al.*, 2014; Hauser *et al.*, 2015). Rubisco activase belongs to a group of AAA+ ATPases, which form a hexameric helical arrangement with a central pore, through which target protein's loop segments are pulled through and unfolded (Stotz *et al.*, 2011). According to the current model, Rubisco activase recognises Rubisco (residues 89-94 in spinach rbcL) using at least two regions of the protein: N-terminal domain and alpha-helical subdomain (helix 9) in the C-terminal domain (Portis *et al.*, 2008; Stotz *et al.*, 2011). However, it has been proposed that due to the diameter size of the Rubisco activase pore, it is highly unlikely that only a short six amino-acid peptide motif is engaged by Rubisco (Mueller-Cajar *et al.*, 2014).

Rubisco activase is a thermolabile enzyme, and as such, it has been linked to decreased activation states of Rubisco at elevated temperatures in C_3 plants but there has been no

confirmation of direct cause and effect (Salvucci & Crafts-Brandner, 2004**b**,**a**; Cen & Sage, 2005; Yamori *et al.*, 2006; Kurek *et al.*, 2007; Yamori *et al.*, 2012).

The *Arabidopsis* and *B. vulgaris* genomes contain one Rubisco activase gene, which gives rise to two different isoforms due to alternative splicing. Also, four Rubisco small subunit genes have been identified in the *Arabidopsis* genome, and five in the sugar beet genome (Dohm *et al.*, 2013). Here, the hypothesis is that C_3 and C_4 plants will show differences in response to temperature in terms of produced isoforms of Rubisco activase and Rubisco small subunit. This hypothesis has been informed by two observations. Firstly, Rubisco enzyme kinetics measurements showed that efficiency of carboxylation and specificity of carboxylation over oxygenation vary significantly across the two tested temperatures in the C_4 species (*A. pungens*), with more efficient and specific Rubisco activity at 30°C, and no significant difference across the two temperatures in the C_3 species *A. angustifolia*. These differences in Rubisco activity also correspond to significant differences in growth rates across the two temperatures between the two *Alternanthera* species, with *A. pungens* showing significantly better growth at 30°C and no significant effect of temperature on growth in *A. angustifolia*. Secondly, some C_3 species, such as spinach (also Caryophyllales), wheat and cotton can express an alternative Rubisco activase isoform at high temperatures, which apparently confers thermostability to the Rubisco activation process (Crafts-Brandner *et al.*, 1997; Law & Crafts-Brandner, 2001; Law *et al.*, 2001). The hypothesis was tested by identifying SNPs with alleles showing opposite pattern of expression at low and high temperature, e.g. higher expression of a given allele at 18°C and lower at 30°C than that of the alternative allele, as well as significantly varying across the two temperature conditions in absolute terms. Such pattern suggests existence of separate enzyme isoforms underlying the differences in expression of different alleles, which could not be resolved using our current methodology. Then, SNPs showing replacement instead of just synonymous substitutions were of interest as they can result in altered activity and interaction between the two proteins investigated.

For Rubisco activase, 81 and 75 SNPs were found which showed significant differential expression between the two temperatures in the gene coding region in *A. angustifolia* and

A. pungens, respectively. However, SNPs which showed opposite trends of expression at the two experimental temperatures were a fraction of those: 21 SNPs in *A. angustifolia* with 9 of them classed as replacement substitutions (Figure 4.9), and 6 SNPs in *A. pungens* with 3 of them replacement substitutions (Figure 4.10). While the numbers of SNPs showing differential expression of the two alleles in the coding region of Rubisco activase in *A. angustifolia* and *A. pungens* are comparable, four times more SNPs matching the expected pattern were identified in the C_3 species. For Rubisco small subunit itself, out of 45 SNPs showing differential expression in the coding region, 17 SNPs showed the expected pattern in *A. angustifolia* (9 of them replacement substitutions), while none did in *A. pungens*, out of 19 SNPs showing differences in expression at the two temperatures (data not shown).

Only one SNP encoding amino acid replacement in *rbcS* was found in the separate gene copies in *A. thaliana* and *B. vulgaris*, while the rest of SNPs were specific to each species in the analyses undertaken, supporting their species-specific importance in the *Alternanthera* lineage. On the other hand, most substitutions encoded amino acids with similar chemical properties. In *A. angustifolia* only three (2 x Ala ↔ Ser, Gly ↔ Ser) and four (Asn ↔ Val, His ↔ Gln, Thr ↔ Met, Asn ↔ Ser) changes for amino acids with different side chain types in rbcS and Rubisco activase were found, respectively. Nevertheless, in *A. pungens*, all of the three replacement substitutions encoded amino acids with more pronounced differences in chemical properties (Lys ↔ Pro, His ↔ Tyr, Asn ↔ Ser). The crystal structure of Rubisco and Rubisco activase interacting in a complex has not been solved yet, so only very simplistic predictions can be made regarding the functional importance of these substitutions based on their exposure on protein surface, which in this case support, or rather do not rule out the importance of the replacements listed above. Overall, the much higher frequency of SNPs with expected pattern of expression in *A. angustifolia* Rubisco activase, and their lack in *A. pungens* Rubisco small subunit agrees with the hypothesis that this C_3 species expresses different isoforms of one or both of those genes which function better at either low or high temperatures as part of the species' temperature acclimation response, guaranteeing stable photosynthesis rates. In contrast, in *A. pungens*, there is no

such plasticity in isoform switching, and consequently the rate of photosynthesis declines at the sub-optimal, low temperature.

4.3.6 Population genetics analyses

To facilitate future population genetics studies utilising molecular markers, SNPs and microsatellites were identified in the *popgen* assembly in each species in the study. A number of di- to hexa- nucleotide microsatellite loci were identified in the assemblies, from 1,797 SSRs in *H. peploides*, through 1,974 SSRs in *S. globosa*, 2,746 in *A. pungens* and 3,622 in *A. angustifolia* (Table 4.7). As expected, the top two microsatellite categories in each species were tri-nucleotide and hexa-nucleotide repeats, as they are not purged quickly by natural selection like the microsatellites, which can introduce frame-shift mutations (Metzgar *et al.*, 2000).

When it comes to SNP calling, again, there were significant differences between results obtained from the *A. angustifolia* assembly and the remainder. While 77,906 SNPs were found in *A. angustifolia*, 291,499 were found in *A. pungens*, 230,583 in *S. globosa* and 423,556 in *H. peploides* (Table 4.6). One basic method of filtering 'false positive' SNPs, which are a result of collapsing two homologous sequences into one contig, is to look for SNPs heterozygous in each individual sequenced (termed *fixed heterozygotes* here), which is a result strongly deviating from Hardy-Weinberg equilibrium. Such SNPs constitute less than 10% in *A. angustifolia*, but as many as 49.5% in *A. pungens*, 25.4% in *H. peploides*, 45.6% in *S. globosa* (Table 4.6).

Taken together, the higher number of SNPs in *A. pungens*, *H. peploides* and *S. globosa* could be the outcome of a number of evolutionary processes, including polyploidization events. In fact, *H. peploides* and *S. globosa* have a high ploidy of $n = 34$ chromosomes, and $n = 22–23$, respectively (Malling, 1957; Kapralov *et al.*, 2009), and divergence distribution of sequenced paralogs in *Schiedea* along with the *Schiedea-Honckenya* divergence support a single polyploidisation event prior to the *Honckenya - Schiedea* split (Kapralov *et al.*, 2009). Therefore, one needs to proceed with caution in any population genetic studies on

these three species. Also, because of the inability to establish correct genotypes, no tests for selection contrasting levels of intraspecific polymorphism with interspecific divergence, such as the HKA and MK test, could be applied to the species pairs sequenced in the study. Consequently, despite identification of 6,563 ortholog pairs using a reciprocal Blast approach (Hirsh & Fraser, 2001) between *A. angustifolia* and *A. pungens* (average nucleotide identity: 88.5%, SD ± 3.9), and of 5,349 ortholog pairs between *H. pungens* and *S. globosa* (average nucleotide identity: 92%, SD ± 3.9), no population genetics tests were carried out.

4.4 Key findings

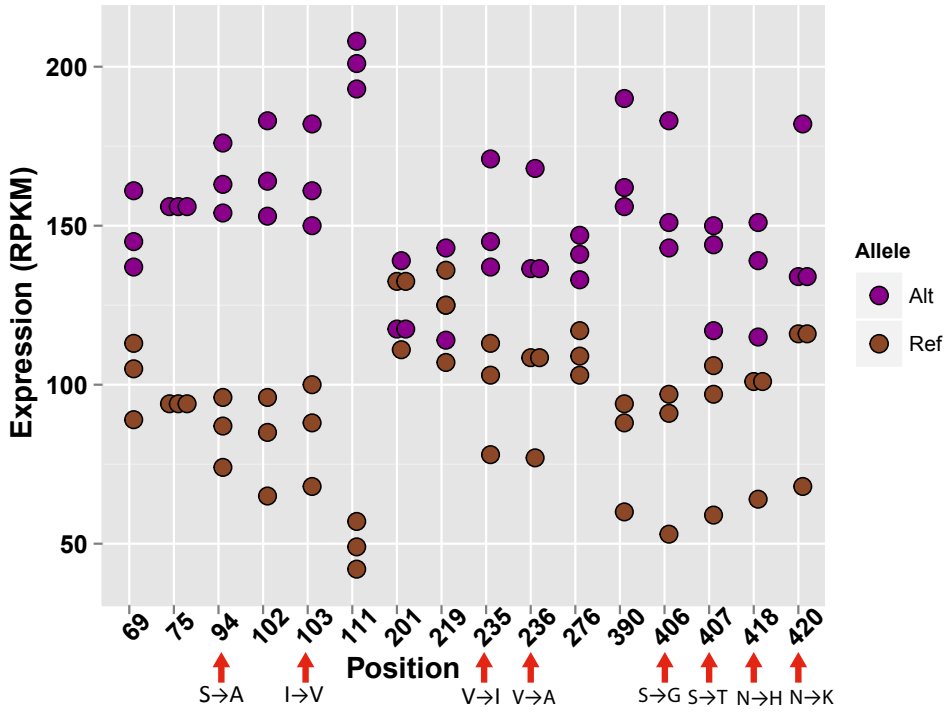
- Re-use - with modification - of a custom bioinformatics pipeline allowed assembly, annotation and quantification of four new transcriptomes of non-model species with complex traits of ecological interest in the order Caryophyllales. On average, ~10,000 reference genes were annotated in each assembly, representing the most highly expressed fraction of each species' transcriptome. Median coverage of the reference proteins by reconstructed transcripts was >75% in each species, allowing reliable gene identification and expression quantification.

- Generation of an additional assembly for each species with low redundancy and misassembly allowed identification of thousands of microsatellite and SNP markers but also uncovered indirect evidence for polyploidy in all but one (*A. angustifolia*) species in the study.

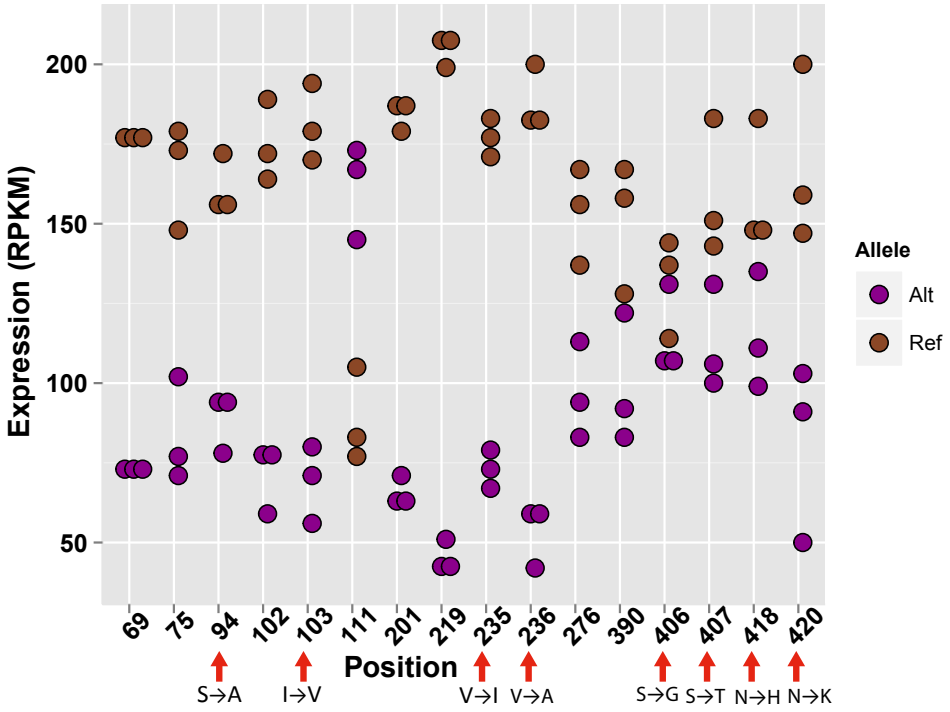
- Comparison of shoot transcriptomes of two *Alternanthera* species at two temperatures (18°C and 30°C) revealed substantial changes in the C_3 species (*A. angustifolia* - 336 genes) but not in the C_4 species (*A. pungens* - 30 genes), which correlated with stable plant performance and photosynthesis rate in *A. angustifolia* but poor acclimation to the low temperature in *A. pungens*. Relative to the C_3 species, *A. pungens* showed constitutive upregulation of C_4 photosynthesis genes as well as genes related to establishment of the Kranz anatomy, and downregulation of genes related to photorespiration, as expected. The temperature acclimation response in the C_3 species centred on changes

related to heat tolerance and redox stress, and crucially: photosynthetic enzymes, photosystem function and regulation of photoinhibition.

→ Indirect evidence for the role of isoform switching in temperature acclimation was found in the C_3 species. Alternative isoforms of thermostable Rubisco activase and Rubisco small subunit are likely expressed at the high temperature promoting Rubisco activation and efficient photosynthesis.



(a) 18°C



(b) 30°C

Figure 4.10: Allele-specific expression at SNPs in the Rubisco small subunit showing opposite trends of differential expression at: (a) 18°C and (b) 30°C in *A. angustifolia*. Results from three replicate plants are shown per site and sites are numbered in reference to *A. thaliana* Rubisco small subunit coding region (AT5G38410). Amino acid changes (Ref → Alt) in non-synonymous substitutions are marked below the chart. Reference (Ref) and alternative (Alt) allele states were determined in reference to the assembly consensus sequence.

Table 4.6: Summary of SNP calling results for the Caryophyllales species

	A. angustifolia	contig no.	*A. pungens*	contig no.	*H. peploides*	contig no.	*S. globosa*	contig no.
all SNPs								
biallelic	77,906	12,313	291,499	14,301	423,556	17,271	230,582	16,726
multiallelic	448	529	1,567	824	4,579	3,659	1,715	1,317
fixed heterozygotes eliminated								
biallelic	70,043	12,170	148,137	13,963	316,130	17,171	125,534	13,276
multiallelic	425	508	982	741	4,005	3,398	1,504	1,125
min. 5 genotypes								
biallelic	55,147	10,682	120,941	12,675	260,884	15,537	92,774	10,871
multiallelic	487	364	701	566	2,714	2,699	1,128	809

Table 4.7: Summary of microsatellites detected in *popgen* assembly in each species: AA - *A. angustifolia*, AP - *A. pungens*, HP - *H. peploides*, SG - *S. globosa*

	Count				Average length				Counts/Mbp			
	AA	*AP*	*HP*	*SG*	*AA*	*AP*	*HP*	*SG*	*AA*	*AP*	*HP*	*SG*
dinucleotide	580	299	294	292	13.47	13.33	13.36	13.31	26.36	15.13	14.95	13.45
trinucleotide	1835	1189	660	713	19.3	19.03	18.9	19.67	83.4	60.17	33.56	32.84
tetranucleotide	140	152	90	95	19.24	19.87	19.59	19.79	6.36	7.69	4.58	4.38
pentanucleotide	101	86	112	78	24.18	23.9	25.57	24.13	4.59	4.35	5.7	3.59
hexanucleotide	966	1020	641	796	22.79	23.53	24.58	24.78	43.91	51.62	32.6	36.66

Chapter 5

Conclusions and Future Recommendations

5.1 Features of the *A. serpyllifolium* transcriptome assembly

Faced with opportunity for colonisation of soils containing phytotoxic levels of heavy metals, certain plant lineages have evolved not only tolerance but also an active ability to take up and transport metals for accumulation in the shoot in the process of metal hyperaccumulation. This intriguing plant trait invites questions regarding its genetic basis and evolutionary routes taken to arrive at it. This study tackled aspects of both issues within the Iberian *Alyssum serpyllifolium* species complex, which contains populations endemic to serpentine soils with superior Ni hyperaccumulation ability, as well as Ni tolerant non-serpentine populations, with so far unexplored phylogenetic relationships linking the populations within the complex. The aim of this chiefly RNA-Seq-based work was to evaluate genetic differences between serpentine and non-serpentine *Alyssum serpyllifolium* populations as reflected in the degree of gene expression differences and sequence polymorphism. Also of interest was the nature of evolutionary forces acting on the candidate loci for hyperaccumulation and serpentine adaptation combined with demographic history of the serpentine and non-serpentine populations in the species complex.

Twenty-one *Alyssum serpyllifolium* individuals distributed among 2 serpentine (Carratraca-S, Samil-S) and 2 non-serpentine (Morata-Ns, Rubía-NS) populations were sequenced using RNA-Seq at 40× average coverage, along with 6 additional individuals from the outgroup *Clypeola jonthlaspi*. A hybrid assembly procedure combining a pre-mapping step to the *Arabidopsis thaliana* transcriptome prior to *de novo* assembly of each locus was employed. This resulted in 18,566 *Alyssum* contigs showing high coverage of the coding regions of the corresponding *Arabidopsis* orthologs - mean 90%. SNP calling yielded 292,755 biallelic SNPs genotyped with high confidence in at least 85% of individuals in each population.

5.2 Differential gene expression analysis in *Alyssum*

Differential gene expression analysis (FDR cut-off < 0.01) using RNA samples from both the root and the shoot revealed small gene expression differences between serpentine and non-serpentine populations of *Alyssum serpyllifolium*: 52 differentially expressed genes (DEG), as opposed to 5,087 DEG in all *Alyssum serpyllifolium* versus *Clypeola jonthlaspi* comparison. Greater variance in gene expression amongst the serpentine and non-serpentine *Alyssum* populations was observed in the root than in the shoot, as evidenced by the PCA analysis which suggests more flexibility in gene expression is needed in *Alyssum* in the root when faced with difficult soil chemistries.

Out of the 10 genes upregulated in serpentine populations in the root, 6 were membrane-targeted, chief amongst them 3 members of the yet uncharacterised Cystathionine-β-synthase domain-containing family (CBS-DUF21) and 2 members of the Iron Regulated (IREG) family: *IREG1* and *IREG2*, which show nickel and cobalt transport activity and appear important for Ni hyperaccumulation across different plant families. Twenty-eight genes upregulated in the shoot were enriched for metal transporters with another appearance from *IREG2*, but also the main soil iron importer *IRT1*. Population-specific upregulated genes in the shoot included members of the ZRT, IRT-like Protein (ZIP) family previously implicated in metal hyperaccumulation: *ZIP9*, *ZIP11* and *ZIP12*. Genes associated with histidine production (*ATP-PRT1*) and transport (*CAT4*) were also found

amongst upregulated genes in the shoots of both serpentine populations, as histidine has previously been shown to be an important nickel chelator in *Alyssum*, which facilitates nickel loading into the xylem for transport to the shoot. Interestingly, despite different individual genes in each enriched category, the same processes seemed to be downregulated in both serpentine populations in the shoot, namely: response to stress, defence response to bacteria and fungi, tryptophan/indokylamine catabolism, metal detoxification and oxidation-reduction.

5.3 Genome-wide scans for selection in *Alyssum*

The search for outlier loci, containing SNPs showing the highest F_{ST} values, was conducted between each serpentine population and the two non-serpentine populations. Both serpentine populations contained a similar number of candidate outliers displaying total polarisation in frequency between serpentine and non-serpentine alleles and derived state relative to the outgroup: Carratraca-S 1170 SNPs (in 388 genes), while Samil-S 1127 SNPs (in 369 genes). The high proportion of derived serpentine alleles as compared with just 15% amongst major alleles in the rest of the genome likely indicates strong action of directional selection in the serpentine lineage. This conclusion is further supported by a significantly higher number of non-synonymous substitutions amongst the serpentine outlier loci (43-45%), than in the rest of the genome (36%) as well as significantly lower nucleotide diversity of the serpentine outlier genes (mean 0.00342- 0.00430) compared with non-outliers (mean 0.00621 - 0.0063).

Only close to 1 in 10 of the outlier SNPs (142 SNPs in 77 genes) in each population was shared between the two serpentine populations. In both of them, 45% of outlier alleles were only detected in one of the serpentine populations and so are absent/segregating at very low frequencies below detection limit in the other serpentine population. This suggests the existence of local selective pressures resulting in divergence (Kaeuffer *et al.*, 2012) and possibly functional redundancy or flexibility in generating the same phenotype. QTL analyses conducted in other model hyperaccumulator and metallophyte species support the existence of few loci of major effect (e.g. Bratteler *et al.*, 2006**b**) so the non-shared

outlier loci are possibly involved in adaptation to other differences between sites, such as latitude, or have a tangential, non-essential role in adaptation to serpentine. Genetic crosses allow explicit testing of associations between genotype and phenotype under controlled environmental conditions (Stern, 2013). Such approaches were not undertaken before for *Alyssum* because of the long time commitment (reproduction only in the second year) and high cost required to maintain a lab population of different crosses. Given the right resources, however, QTL analysis could be used in identification of broad genomic regions harbouring the selected loci followed by time-consuming fine-mapping of specific genes (Stern, 2013). This method has the benefit of genome-wide discovery of candidate loci allowing estimation of their effect sizes (Conte *et al.*, 2012), unlike genome scans, which do not provide quantitative information about the magnitude of the effect of a given candidate locus.

Focusing on caveats of the analysis, a non-insignificant proportion of identified outliers are expected to be false positives arising out of both the methodology and experimental design used - here the very small number of replicates is crucial, as it can lead to distortion away from the true allele frequency spectrum. When it comes to methodology biases, the empirical approach for identifying outliers in genome scans can lead to both a high rate of false negatives, particularly in the instance of genome-wide purifying selection, standing genetic variation, previous population bottlenecks on one hand (Li *et al.*, 2012**a**) and from false positives hitch-hiking to a causal allele on the other hand (Teshima *et al.*, 2006). Moreover, protein activity does not necessarily always correlate with steady-state transcript levels due to e.g. differential mRNA stability as well as turn-over and translation rates (Marguerat & Bähler, 2010; Wolf, 2013), so in effect there is some potential for false positives and negatives among differentially expressed genes, as well.

5.4 Key pathways and genes for serpentine adaptation in *Alyssum* and beyond

In both serpentine populations, the outlier loci were enriched for similar functional categories as the differentially expressed genes: metal ion transport, iron deprivation response,

defence responses, histidine biosynthesis and oligopeptide transport. For instance, the second isoform of the ATP-PRT enzyme (*ATP-PRT2*) catalysing the first step of histidine biosynthesis in addition to the gene catalysing the second and third steps (*HISN2*) were found amongst outliers in Samil-S, with *ATP-PRT1* expression also upregulated in both serpentine populations. In addition, two histidine soil importers (*AAP5* and *LHT1*) showed traces of positive selection in the serpentine lineage according to the McDonald-Kreitman test. Next, shared candidate outlier loci important for metal homeostasis included two members of the Major Facilitator Superfamily (MFS) - *ZIF1* and *ZIFL1*, while two different members of the Heavy Metal ATPase family (*HMA4* and *HMA6*) were targeted in each population specifically. In the abiotic stress category, different components of the SOS salt tolerance pathway and in the defense category, different genes involved in the synthesis of aliphatic glucosinolates were found amongst outliers in the two serpentine populations.

In fact, genes in the same family (e.g. ZIP, MT, NAP) or even the same genes were pinpointed as candidate genes based on the two complementary approaches. There was a statistically significant increased representation of genes upregulated in serpentine populations amongst outlier loci (both population-specific and shared). Such loci are amongst the strongest candidates for functional importance in adaptation to a serpentine environment or hyperaccumulation and include *CAT4*, *IREG1*, CBS-DUF21 family members, *ATP-PRT1*, *HMA5*.

On a broader scale, there was significant convergence in the functional classes and families (e.g. IREG, NRAMP, HMA, ZIP, YSL) of identified genes potentially important for the hyperaccumulator phenotype in *Alyssum* when compared with results from previous studies on hyperaccumulators of other metals. Identification of many genes involved in zinc and cadmium hyperaccumulation is not surprising given the promiscuity of certain metal transporters and chelators towards metals with similar and even not-so-similar chemical properties but could also indicate altered functionalities and changed metal ion preference (Baxter & Dilkes, 2012). Furthermore, a number of outlier candidate genes appeared to show convergent evolution to the serpentine environment, as they had

been previously identified in an analogous way in serpentine accessions of *Arabidopsis lyrata* (Turner *et al.*, 2010). They include *KUP9* potassium transporter, *CAX1* antiporter important for growth under low Ca/Mg regimes encountered in the serpentine soils, *MGT* magnesium transporter as well as *MT1C* metallothionein protein and *TPC1* vacuolar channel in Carratraca-S, and *NIP5* transporter responsible for borate uptake from the soil in Samil-S. Significant convergence in the genetic mechanism of metal hyperaccumulation and serpentine endemism revealed is likely due to pleiotropic constrains surrounding the mechanism for metal tolerance and high conservation of metal transporters. Simply put, "specialized genes drive the evolution of specialized traits" (Martin *et al.*, 2013).

However, some gene families with proven important roles in other hyperaccumulators were missing in the study – e.g. the MTP family of transporters, nicotianamine synthase genes and genes involved in sulphate assimilation. Novel candidate genes not found before included *CAT4* - a member of the CAT family important for histidine transport and 4 out of the 7 CBS-DUF21 genes in the *A. thaliana* genome. CBS-domain containing proteins have been shown to play a role in protection against heavy metal and osmotic stress in plants (Singh *et al.*, 2012).

5.5 Proposed genetic model of serpentine adaptation in *A. serpyllifolium*

Alyssum serpyllifolium is an Ni hyperaccumulator exclusively (De Varennes *et al.*, 1996): it cannot hyperaccumulate Cr, Cu, Pb, Zn, which suggests specific molecular adaptation towards accumulation of Ni. Moreover, low bioavailability of chromium and cobalt, despite their abundance in the serpentine habitat of *A. serpyllifolium* suggests that specific adaptations targeted at countering their toxicity are not very likely. On the other hand, tolerance to these metals is elevated in serpentine *Alyssum serpyllifolium*. Furthermore, *Alyssum serpyllifolium*'s uptake of other metals (cobalt, manganese) present in its native environment is an order of magnitude higher than of the other serpentinophytes in the wild (Freitas *et al.*, 2004), which suggests partial presence of common mechanisms facilitating uptake of these metals as well as nickel. Plant metal homeostasis network is a complex

system with many in-built redundancies as studied by the field of metallomics (Mounicou *et al.*, 2009; Krämer, 2010), and revealed e.g. by promiscuity of many metal transporters such as IRT1, NRAMP4, which are capable of transporting various metal cations. With that in mind, genes potentially adaptive for nutrient challenges posed by serpentine soils will be discussed focusing only on their chief known functions, with implicit possible interactions in other related pathways and processes.

Integrating various strands of evidence from this study (differential gene expression analysis, scans for selection) and the literature, one is able to roughly define a partial common core of putative genes important for serpentine adaptation in *A. serpyllifolium* (Figure 5.1). It is centred on two permanent fixtures of serpentine soils: large mobile Ni fraction in the soil and low Ca:Mg ratio (Freitas *et al.*, 2004; Kidd & Monterroso, 2005; Cabello-Conejo *et al.*, 2014).

Low Ca:Mg ratios seen in the serpentine soil are reflected in plants: in serpentine *Alyssum serpyllifolium* populations (Sierra Bermeja-S, Samil-S, Barazón-S), Ca:Mg ratio ranges from 0.3 to 0.6 but increases to 2.6 in non-serpentine *Alyssum* (Becerra-Castro, 2012). To deal with magnesium excess and low calcium, *A. serpyllifolium* is hypothesised here to rely on three genes with established roles in metal homeostasis under low Ca:Mg ratios - *CAX1* as well as *CIPK3* and *CIPK9* kinases. *CAX1* was found among shared F_{ST} outliers in serpentine populations and is a tonoplast calcium antiporter with roles in supporting tolerance of low Ca condition (Bradshaw, 2005; Baliardini *et al.*, 2015). The two *CIPK* genes were found to be under strong purifying selection in the serpentine populations and participate in signalling pathway leading to storage of excess Mg in the vacuole (Tang *et al.*, 2015).

Moreover, a tonoplast Mg transporter *MGT2* hypothesised to transport Mg into the vacuole in the mesophyll for sequestration (Conn *et al.*, 2011) was found among F_{ST} outliers in Carratraca-S. Finally, a novel role in Mg homeostasis is proposed for the root-upregulated CBS-DUF21 proteins found in serpentine populations as the CBS domain features in a

number of bacterial and eukaryotic Mg transporters, and can act as a Mg sensor and affect Mg transport rate (Hattori *et al.*, 2007; Gómez-García *et al.*, 2012; Hardy *et al.*, 2015).

Moving onto adaptations allowing high nickel transfer and accumulation in serpentine populations, the first proposed mechanism is the activation of iron deprivation response across the plant, as evidenced by upregulation of iron transporter *IRT1* in the shoot, *IREG* family transporters in the root and signature of directional selection found in serpentine populations in transporters involved in Fe homeostasis: *OPT3*, *FRD3* and *ZIF1*. In direct analogy with zinc deficiency response activated in zinc hyperaccumulators *A. halleri* and *N. caerulescens* and driving increased zinc uptake into the root and the shoot (Hanikenne *et al.*, 2008; Gustin *et al.*, 2009), iron deficiency response-activated genes are hypothesised to facilitate Ni uptake and transport to the shoot. The main cause of iron deprivation response seen in *Alyssum* plants is expected to stem from nickel out-competing iron for IRT1 binding sites due to its high bioavailability in serpentine soils and conversely, poor iron solubility at near neutral pH found in serpentine soils (Nishida *et al.*, 2011; Becerra-Castro, 2012). Then, iron deficiency would elicit upregulation of IREG transporters with proposed roles in stimulating Ni hyperaccumulation: export of nickel from the root into the xylem by IREG1 (Kirchner, 2009) and temporary buffering of excess Ni in the root vacuole by IREG2 (Schaaf *et al.*, 2006). Under this scenario, remobilisation of Ni from the root vacuole for xylem loading would be done by NRAMP4, as it is a tonoplast transporter with wide metal transport abilities with confirmed role in Zn and Cd remobilisation in *A. thaliana* (Thomine *et al.*, 2000, 2003; Lanquar *et al.*, 2005, 2010; Pottier *et al.*, 2015). Particularly exciting in the context of this study, is the recent finding that NRAMP4's ability to transport different metals is strongly affected by point mutations in *Arabidopsis thaliana* (Pottier *et al.*, 2015), advancing the hypothesis of functional role of replacement substitutions seen across all serpentine populations in *A. serpyllifolium* in tweaking the transporter properties in favour of Ni transport.

Finally, further confirmation for the importance of histidine chelation-related mechanisms in Ni transport, tolerance and hyperaccumulation was found in the study. Elevated

expression of key histidine biosynthesis enzyme - *ATP-PRT1* and signature of selection seen in the gene across both serpentine populations as well as in *HISN2* (another enzyme in the histidine biosynthesis pathway) in Samil-S confirms results from Ingle *et al.* (2005**a**) and is consistent with suggested histidine-dependent Ni xylem loading (Krämer *et al.*, 1996; Kerkeb & Krämer, 2003). The identification of the *CAT4* transporter as being highly expressed in the shoot tissues of serpentine populations suggests that histidine transport might be important in aspects of Ni hyperaccumulation other than xylem loading, such as vacuolar sequestration in the shoot. Preliminary studies have shown that the *Alyssum CAT4* encodes a vacuolar histidine transporter, and is implicated in the sequestration of nickel into the vacuole (C.J. Snowden, S.T. Mugford & J.A.C Smith, unpublished).

In addition, the study found a highly significant over-representation of genes related to defence response and oxidative stress, such as involved in respiratory burst, among genes downregulated in the shoot across the serpentine populations and discovered as putative targets of positive selection, with scant overlap among the actual genes targeted. This could be explained by the proposed trade-off mechanism in the elemental defence hypothesis of metal hyperaccumulation (Boyd & Martens, 1992; Kazakou *et al.*, 2008), where Ni accumulation adaptation affords protection from biotic stresses and ROS stress at the same time allowing reduction in the expenditure for regular innate mechanisms of plant defence. This would be one of the first times such a gene regulation effect has been shown for metal hyperaccumulators (Farinati *et al.*, 2009). Quite the contrary, in *N. caerulescens*, increased shoot activation of salicylic acid synthesis and downstream synthesis/turnover of the anti-oxidant glutathione contributes to Ni tolerance and accumulation (Freeman *et al.*, 2004, 2005; Freeman & Salt, 2007). However, this pathway is overall downregulated in the shoot in serpentine *A. serpyllifolium* populations, with e.g. different Methyl Esterase genes involved in long-range salicylic acid signalling downregulated in each serpentine population and 15 genes responsive to salicylic acid downregulated in the Carratraca-S shoot.

Apart from high nickel and low Ca:Mg ratios, serpentine soils exhibit typically other

unfavourable traits towards plant growth. Low phosphorus and nitrogen content have been measured in Samil-S (Kidd & Monterroso, 2005) and Barazon-S habitats (Álvarez-López *et al.*, 2016). In agreement with that factor exerting selective pressure, two potassium transporters (*KUP9* and *KT1*), one phosphate transporter (*PHT1;3*) and one gene in the low Pi signalling relay (*LPR1*) were identified as shared targets of selection in both studied serpentine populations. Possible influence of local adaptation was reflected in the presence of population-specific outlier genes: a regulator of phosphate transport (*PHO2*), a DGDG biosynthesis gene expressed under low Pi (*MGD2*), a negative regulator of phosphate deficiency response (*WOL*) in Carratraca-S along with a potassium transporter (*KUP7*) and a negative regulator of the *KT1* potassium transporter (*CBL10*) in Samil-S.

Interestingly, a number of outlier loci in Samil-S were associated with regulating the uptake of other metals: boron (*BOR1*, *NIP5;1*), manganese (*ECA3*) and sodium (*SOS2*, *NHX4*, *SIK*). In certain serpentinophytes, the concentrations of these metals was found to have an inverse relationship with the high plant magnesium concentration (Brooks & Yang, 1984), however this antagonism in nutrient uptake should be also present in Carratraca-S, where such hypothetical adaptation appears absent.

5.6 Next steps towards establishing functional importance of candidate genes

While comparative transcriptomics and scans for selection provide circumstantial evidence for potential adaptive value of a given gene, *in vivo* characterisation of gene function can lend more direct support by revealing relevant phenotypes. In the future, in order to establish the role of candidate genes identified in the study in metal homeostasis, a number of steps could be taken. Establishing gene expression responsiveness to metal concentration using qRT-PCR or Nothern blotting could be a first step. For instance, a number of zinc transporters identified in *Arabidopsis* are zinc deficiency- or excess- inducible, among them candidate genes identified in this study: *ZIP9*, *ZIP12* (Jain *et al.*, 2013) and *ZIF1* (Haydon & Cobbett, 2007). Furthermore, impaired release of nickel and histidine from the root into the xylem in *A. lesbiacum* on addition of a translation inhibitor (Kerkeb

& Krämer, 2003) suggests a possible involvement of nickel-induced protein(s) in xylem loading in *Alyssum*.

However, the majority of the known key hyperaccumulator genes were shown to be constitutively expressed (Becher *et al.*, 2004; Freeman *et al.*, 2004; Weber *et al.*, 2004; Ingle *et al.*, 2005**a**; Hammond *et al.*, 2006), which is typical of endemic plants that complete their entire life cycle on metalliferous soils or other habitats with a high level of a specific stressor (see e.g. halophyte *Thellungiella salsuginea* - Gong *et al.*, 2005; Amtmann, 2009; Sun *et al.*, 2010). Therefore, lack of evidence for metal inducibility of a given gene's expression would not discount its role in metal homeostasis in hyperaccumulators. Next, responses of *Arabidopsis* loss-of-function mutants, and equally overexpressor lines featuring the *Alyssum* gene, to media with a range of metal concentrations would be characterised, particularly those showing atypical concentrations in serpentine soils. Plant performance and metal tolerance would be measured using such standard metrics as germination rate, root growth, biomass and chlorophyll content, and metal accumulation in different organs determined. Transport ability, selectivity and kinetics of predicted transporters could be tested using yeast functional complementation assays. In addition, in cases such as *NRAMP4*, where there is firm evidence of altered metal selectivity conferred by point site mutations (Pottier *et al.*, 2015), non-synonymous outlier alleles as discovered in the serpentine populations could be introduced through site-directed mutagenesis. Lastly, subcellular and tissue localisation would be investigated with transient expression of fluorescent protein-tagged constructs in *Arabidopsis* protoplasts or tobacco mesophyll cells, and with *in situ* hybridisation, GUS staining or immunostaining, respectively.

5.7 Potential role of CNV in serpentine adaptation in *Alyssum*

One unexplored factor in adaptation to hyperaccumulation and serpentine environments in this work is the phenomenon of copy number variation (CNV), which has proven of some importance in extremophytes (Wu *et al.*, 2012; Oh *et al.*, 2014), including other hyperaccumulators, in particular through increased expression or sub-functionalisation

of transporters - likely because of their importance for adaptation to atypical ionic environments. In the current study, steps were taken to eliminate polymorphisms with frequencies deviating from the Hardy-Weinberg equilibrium to ensure that the population genetics results reported pertain to a single locus at a time and not an amalgamation of a number of copies. However, gene expression changes detected could certainly be due to differences in the total expression of a number of unknown paralogs. There is considerable evidence for the importance of gene duplication in establishing the hyperaccumulator and metal tolerant phenotype, so far through promoting increased expression – triplication of *HMA4* in *A. halleri* (Hanikenne *et al.*, 2008), duplication-to-quadruplication of *HMA4* and pentaplication of *HMA3* in *N. caerulescens* (Ó Lochlainn *et al.*, 2011; Ueno *et al.*, 2011; Craciun *et al.*, 2012) and copy number variation in *ZIP3, ZIP6*, *ZIP9* and *MTP1* in *A. halleri* (Talke *et al.*, 2006).

Unbiased detection of CNVs would certainly be facilitated by the generation of an entire genome sequence in the future, whereas now targeted effort utilising molecular biology methods is required. Generation of a genome sequence of *Alyssum serpyllifolium* would allow application of linkage disequilibrium-based methods for detection of recent soft and hard selective sweeps, adaptations in regulatory regions and studying the physical architecture of the complex trait, such as role of inversions, CNV and distribution of selected genes in different chromosomal regions. For example, the genome of *Noccaea caerulescens* has been shown to contain a number of inversions grouping Zn and Fe homeostasis genes in a single chromosomal region that is absent in closely related species (Mandáková *et al.*, 2015). Moreover, provision of a good genome assembly would facilitate studies of the processes surrounding divergence between serpentine and non-serpentine populations. It would allow identification of genomic regions resistant to gene flow between them, and quantifying their size, number and distribution (Roda *et al.*, 2013).

5.8 Phylogenetic and demographic history of *A. serpyllifolium*

In future studies, more balanced selection of populations should be pursued, and non-serpentine populations within consistent distances from each of the three serpentine clusters of populations collected. For now, neutral gene flow between *Alyssum serpyllifolium* populations was assessed using a small set of neutral EST-SSR and SNP markers. Microsatellite analysis using an extended sampling of 8 populations and ~ 30 individuals/population genotyped for 8 neutral EST-SSR loci showed no trace of clear isolation-by-distance, or isolation-by-adaptation pattern based on pairwise F_{ST} and Nei's D_A (Nei *et al.*, 1983) and multifactorial analysis. However, a top solution in STRUCTURE analysis recovered 4 clusters of populations which group together parapatric populations on the same substrate. STRUCTURE results from genome-wide sampling of SNPs sourced through RNA-Seq indicated clustering by distance when different K values are considered. Also, expression profiles of serpentine and non-serpentine populations did not group by the hyperaccumulation ability but rather by distance, with the exception of the most variable genes in the root and shoot. Genome-wide, the largest pairwise absolute divergence D_{xy} (Nei, 1987) and F_{ST} (Weir & Cockerham, 1983) on average was found between the two serpentine populations, and was significantly higher than when comparing serpentine with non-serpentine populations. Neighbour-joining trees, Treemix ancestry graphs and RaxML ML trees all show one of the *Alyssum serpyllifolium* serpentine populations to be basal in relation to one subclade containing serpentine and non-serpentine population and another subclade containing parapatric non-serpentine population. Taken together, there is a lack of clear separation into one serpentine and one non-serpentine clade, indicating a complex history of past neutral gene flow between serpentine and non-serpentine populations. More sophisticated scenarios than just bifurcating phylogenies could be explicitly tested using various Markov chain Monte Carlo (MCMC) or Approximate Bayesian Computation (ABC) coalescence-based methods and/or allele frequency spectrum-based demography methods (Sousa & Hey, 2013) to establish migration rates, divergence times and ancestral population sizes between populations within the complex, which was attempted here but abandoned due to problems with run convergence.

On the other hand, investigating the portion of the *A. serpyllifolium* genome putatively adaptive for serpentine adaptation and Ni hyperaccumulation revealed substantial genetic convergence between different metallicolous populations. The number of shared outliers both at the gene and SNP level was significantly higher than expected by chance in the serpentine populations. Phylogenetic analysis of shared outliers indicates that these outlier loci arose as a result of a single sweep from *de novo* mutations or ancestral standing variation as evidenced by clear separation between serpentine and non-serpentine haplotypes, confirmed in a larger sample of eight populations for two select genes - *IREG1* and *NRAMP4*. Given that based on genome-wide neutral variation, the two sequenced serpentine populations group by geography and not by ecotype, overall high differentiation between all the *Alyssum* populations and their estimated ancient divergence (ca. 1.2 MYA) in quick succession, it is quite likely that the common genetic source of hyperaccumulation adaptation arose early in the history of the species and became quickly fixed in serpentine populations which then carried on diverging due to local adaptation and drift, but continued gene flow and introgression at later stages cannot be discounted.

5.9 Proposed model of evolution of edaphic adaptation in *A. serpyllifolium*

Assembling all evidence gathered in the thesis together, a gradual model of complex trait acquisition within the *Alyssum serpyllifolium* emerges involving three broad phases. This study helped to pinpoint adaptive changes in the transcriptomes of serpentine *Alyssum* populations which have occurred during the second (serpentine-wide) and third (local) phases. The first phase involved species-wide evolutionary processes which gave rise to constitutive basic Ni hyperaccumulation ability and tolerance across the non-serpentine and serpentine populations. The genetic variation behind this could partially come from pre-adaptation present also in other closely related Ni hyperaccumulators in Odontarrhena, where at least 70% species are Ni hyperaccumulators (Brooks *et al.*, 1979). This sort of clustering suggests presence of cryptic precursor traits (Marazzi *et al.*, 2012) or enabling traits (Edwards & Donoghue, 2013), which could include better tolerance of challenges

posed by serpentine soils, such as mineral imbalance (Brady *et al.*, 2005), water deficits (Hughes *et al.*, 2001) and facilitated by genetic architecture, anatomy and physiology of the plant lineage.

This scenario is comparable with that hypothesised to have frequently taken place in the evolution of the other complex trait of interest in this work. In C_4 photosynthesis research, it has been now recognised that simple binary encoding of plants as carrying either C_3 or C_4 photosynthesis is misguided in many cases of so-called C_3-C_4 intermediates with partial C_4 properties (Christin *et al.*, 2010). For example, in the widely studied genus *Flaveria*, some C_3-C_4 intermediates give independently rise to fully C_4 species, whereas others persist in various stages of intermediate state (McKown & Dengler, 2007) but none seems to return to their ancestral C_3 state. Indeed, Christin *et al.* (2010) postulated that the common ancestors at the bases of C_4-rich clades contained partial C_4 characteristics, such as leaf anatomy with suitable vein spacing, cell size or enzyme biochemical properties (Sage, 2001), which were seized upon and developed in C_4 lineages. Similarly, gradual acquisition of metal hyperaccumulation trait from an ancestor with some pre-hyperaccumulation genetic adaptation seems likely in *Alyssum*.

During the short second phase of selection which led to a split between serpentine and non-serpentine populations, additional genetic components were selected for resulting in core hyperaccumulator genes shared across all serpentine populations, such as *IREG1* and *NRAMP4*, which allowed them to colonise the serpentine outcrops. Lastly, since then there was a long period (third phase) of local adaptation, drift and limited migration. Serpentine soils acting as island-like habitats can be expected to be a hotbed of local adaptation due to pronounced environmental heterogeneity along with patchy distribution and sharp boundaries which are expected to limit gene flow (Brady *et al.*, 2005). Similar to *N. caerulescens* (Dubois *et al.*, 2003; Jiménez-Ambriz *et al.*, 2007; Gonneau *et al.*, 2014), the likely mixed mating system of *A. serpyllifolium* would serve as another barrier to gene flow. Nevertheless, geographic structuring within the species complex is apparent when sampling genome-wide neutral variation. This lack or small effect of ecotype on the

structure of neutral genetic variation is relatively common in pseudometallophytes (Mengoni *et al.*, 2000; Deng *et al.*, 2007; Bizoux *et al.*, 2008; Słomka *et al.*, 2011); geographical grouping sometimes explains neutral genetic variation better than ecotype characteristics (Baumbach & Hellwig, 2007; Pauwels *et al.*, 2012).

Such observations are congruent with the ecological model of speciation (Feder *et al.*, 2012, 2013), and it appears that *A. serpyllifolium* populations are positioned on the ecological speciation continuum spectrum. Gene flow from serpentine to non-serpentine populations would then only get established for the neutral part of the genome, as the maladaptive serpentine adaptation genes would be purged, resulting in the maintenance of two ecotypes with distinct Ni hyperaccumulation abilities. In the case of parapatric metallicolous and non-metallicolous populations and continued flow of migrants towards non-metallicolous populations, this process might not operate efficiently enough resulting in non-serpentine populations with cryptic intermediate Ni hyperaccumulation and tolernce phenotype. This effect has been possibly documented in the Alhaurín-NS population which is parapatric to *Alyssum malacitanum* populations: Sierra Bermeja-S and Carratraca-S (Figure 1.4) and shows significantly higher Ni hyperaccumulation ability and tolerance than other non-serpentine populations of *A. serpyllifolium* when grown in the laboratory (Figure 1.5)

The overarching model for metal hyperaccumulation and edaphic adaptation evolution in *A. serpyllifolium* outlined above bears resemblance to the proposed scenario for evolution of metal hyperaccumulation in *A. halleri* (Pauwels *et al.*, 2005, 2008, 2012). Sampling of 65 metallicolous and non-metallicolous populations of *Arabidopsis halleri* across its European range revealed no isolation-by-adaptation but isolation-by-distance, regardless of the ecotype (Pauwels *et al.*, 2005, 2012). This supports the hypothesis that founder events for different metallicolous populations and evolution of enhanced metal tolerance and hyperaccumulation were independent. However, estimates of co-incidence of Ah*HMA4* origin and duplication with *A. halleri* speciation 300 000 year ago (Roux *et al.*, 2011) along with basal level of zinc tolerance and hyperaccumulation constitutive across the entire species (Bert *et al.*, 2000, 2002; Pauwels *et al.*, 2006; Kashem *et al.*, 2010) suggest the

presence of ancient common core of species-wide genetic adaptations for increased zinc tolerance and accumulation shared across all the populations, such as that involving *HMA4* - ie. the result of the first phase (species-wide) of metal hyperaccumulation evolution. There is not much evidence for a second phase of adaptation involving all metallicolous populations, but rather only local adaptation on colonisation of individual regions or even just substrates (corresponding to the third phase described above). In short, Pauwels *et al.* (2012) speculate that adaptive genes in metallicolous populations do not share common ancestry in distinct phylogeographic units in the species but suggest convergent realisation of the same phenotype employing homologous and non-homologous genetic pathways. Lastly, high variation in zinc tolerance within non-metallicolous populations and individuals in *A. halleri* can then be explained by either local gene flow from metallicolous to non-metallicolous populations, or ancestral standing genetic variation segregating within the non-metallicolous populations which was seized upon as a basis for adaptation in parapatric metallicolous populations, which could be the case given the relatively young age of this species.

In fact, the timeline of events in the evolution of *Alyssum serpyllifolium* is expected to be much longer as the serpentine outcrops in the Iberian Peninsula have represented a stable and permanent fixture of the landscape for millennia rather than centuries as is the case with man-made metallicolous sites inhabited by *A. halleri*. In agreement with that, the current - admittedly, very uncertain estimates due to tertiary calibration and low taxon sampling - put the origin of *A. serpyllifolium* at ca. 1.2 MYA and divergence within the complex shortly there after, making serpentine populations of *A. serpyllifolium* much older than that of *A. halleri*. *A. halleri* speciated 300,000 years ago with much later emergence of metallicolous populations, possibly co-inciding with appearance of human metal mining activity and accordingly its populations appear much less diverged as judged by F_{ST} values (Fischer *et al.*, 2013).

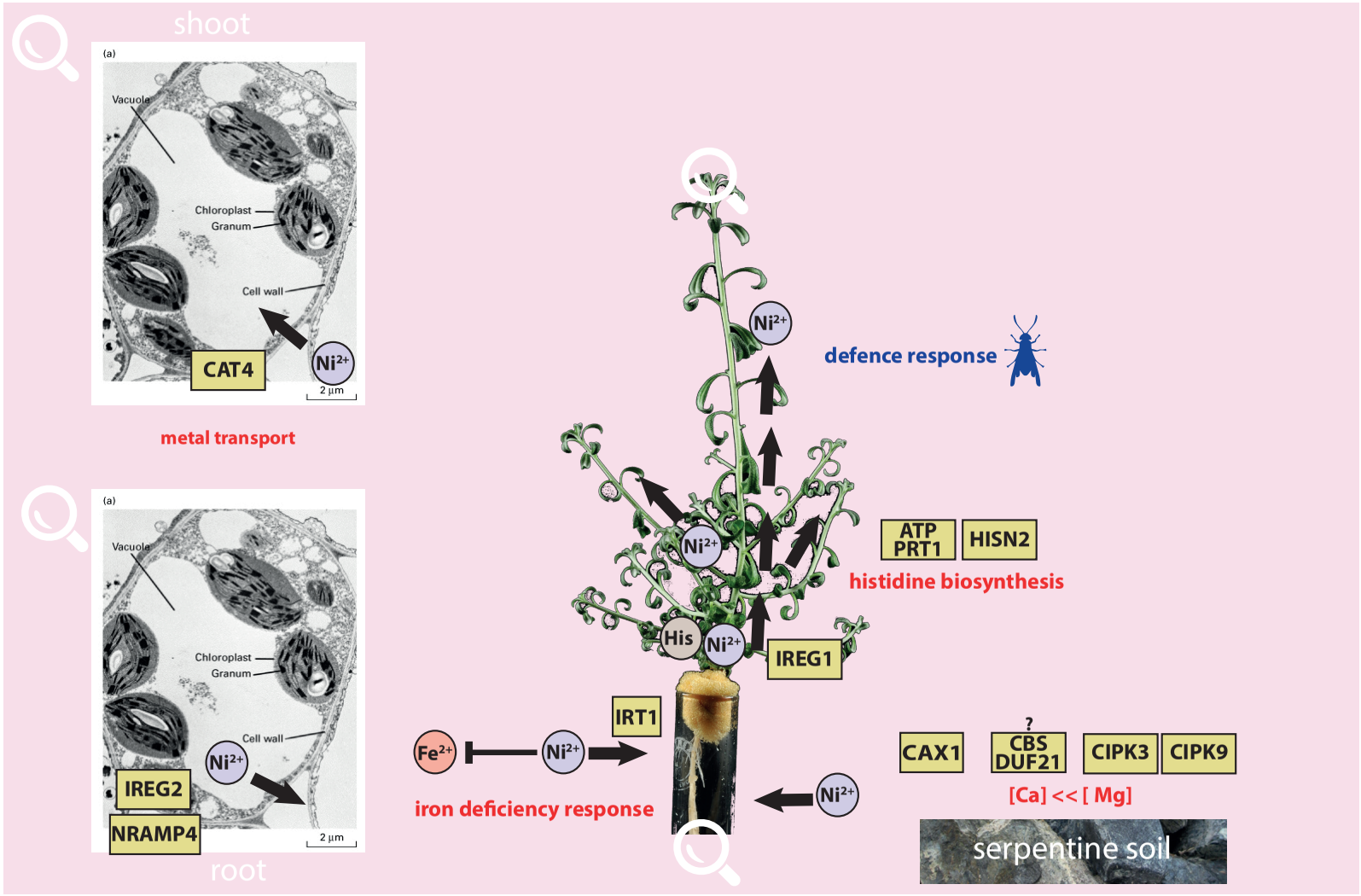


Figure 5.1: Summary of common key processes and candidate genes involved in adaptation to Ni hyperaccumulation and serpentine soil identified in both metallicolous populations of *Alyssum serpyllifolium* (Carratraca-S, Samil-S) in the study. Red font indicates GO terms overrepresented among upregulated genes, while blue font those overrepresented among downregulated genes.

5.10 Features of the new Caryophyllales transcriptome assemblies

The first RNA-Seq study of four species in the order Caryophyllales was carried out with the goal of measuring acclimation in gene expression in response to temperature in related C_4 and C_3 species. Two species pairs from a range of environments were sequenced. The first pair featured a C_4 species (*Alternanthera pungens*) and a weedy C_3 species (*Alternanthera angustifolia*), while the second pair two C_3 species: warm-adapted tropical (*Schiedea globosa*) and cold-adapted subarctic (*Honckenya peploides*). Plants in the experiment were grown under different temperatures to compare and contrast the acclimation response which is known to depend on native climate conditions and photosynthesis type (Berry & Björkman, 1980). Two temperatures were tested: 18°C and 30°C with three replicate leaf samples sequenced per treatment in the *Alternanthera* comparison. In contrast, three temperature conditions were used in the *Schiedea-Honckenya* comparison - 18°C, 25°C, 30°C with 2 replicate leaf samples sequenced per treatment. This proved to be a fatal flaw in the experimental design as reproducibility was shown to be lacking across the replicate samples in a PCA analysis, precluding differential gene expression analysis. While the new *S. globosa* and *H. peploides* transcriptomes remain as the outcome of this study, their quality metrics slightly below that of *Alternanthera* transcriptomes, these assemblies will not be discussed below.

Small RNA-Seq libraries were generated (~ 5 mln reads) in the experiment. Following a multi-step assembly procedure, 19,306 and 26,726 contigs were recovered in *A. angustifolia* and *A. pungens*, respectively. The transcripts were annotated based on the only available high quality genome in the Caryophyllales: sugar beet root, which had a significant hit for the majority of contigs in each assembly (92.7% in *A. angustifolia* and 94.3% in *A. pungens*). For the majority of transcripts in the assemblies, a considerable fraction of the reference protein was reconstructed: median 92.0% in *A. angustifolia* and 81.7% in *A. pungens*. Differential gene expression analysis was carried out comparing samples grown at different temperature inside each species and at the same temperatures across the two species (8,576 transcripts compared), and genes with a minimum 2-fold difference in

expression were only considered.

In addition, separate assemblies with lower contig redundancy and mis-assembly rate, achieved at a cost of significantly reducing mapping rate, were generated to facilitate SNP and EST-SSR discovery in a more unbiased manner. In *A. angustifolia*, 3,622 di- to hexa- nucleotide EST-SSRs were discovered, while 2,746 in *A. pungens*. SNP calling revealed 77,906 SNPs in *A. angustifolia* and almost four times as many in *A. pungens* - 291,499. However, almost half of *A. pungens* SNPs appear to be false positives derived from merging homologous sequences, which warrants caution in future analyses, and may indicate recent polyploidization in the species.

5.11 Constitutive differences between transcriptomes of C_3 and C_4 species

In the interspecific gene expression comparison, differentially expressed genes showing the same trend in expression at both temperatures (163 upregulated and 207 downregulated genes) likely represent genes mostly responsible for constitutive differences between the species related to e.g. morphology, life history and, last but not least, C_4 photosynthesis. Confirming correct quantification of gene expression in the cross-species comparison undertaken here, the relative level of C_4 photosynthesis genes agreed with expectations. It showed constitutively higher expression of C_4 photosynthesis (NADP-ME) subtype genes, such as alanine aminotransferase, aspartate aminotransferase, NADP-MDH, NADP-ME and PEPC in *A. pungens* relative to *A. angustifolia*. This result increased confidence in other gene expression estimates in the study. Moreover, in line with elimination of photorespiration in C_4 plants, genes involved in the process were downregulated in *A. pungens*: *SGAT*, *AOR*, *GLU1*, *GOX1*. Interestingly, genes with functions in cell wall formation and modification (such as *CESA8*, *GSL1*, *GSL5*) were constitutively upregulated in the C_4 species, which may serve e.g. towards structural reinforcement of bundle sheath cells in Kranz anatomy.

5.12 Genetic basis for temperature acclimation ability in C_3 and C_4 species

The majority of plants show some degree of alteration in their physiology in response to temperature during the course of their life cycle (so-called temperature acclimation), which is of particular importance for survival and reproduction in seasonal environments, as plants are sessile organisms (Yamori *et al.*, 2014). However, like other C_4 species, *A. pungens* has a much smaller temperature range of optimum photosynthesis function and shifted towards higher temperatures, than C_3 species which results in poorer acclimation to sub-optimal low temperatures. When grown at 18°C compared to 30°C, the C_4 species *A. pungens* showed significantly higher reduction in its biomass - 71%, than just 15% reduction seen in *A. angustifolia*. The gene expression phenotype behind this phenomenon was investigated in this study. Overall, the C_3 species displays much more plastic gene expression response to temperature (336 genes DEG), while the C_4 species gene expression profile remains fairly constant from 18°C to 30°C (30 DEG). In the interspecific comparison between *A. pungens* and *A. angustifolia* at each temperature, the same order of magnitude of genes were seen to be upregulated and downregulated exclusively at a single temperature (between 144 and 233 genes). Some of these genes are likely behind the differences in temperature acclimation response observed in the two species.

First and foremost, temperature acclimation is expected to affect the photosynthesis apparatus, resulting in a shift in the optimal temperature for photosynthetic activity and also more efficient photosynthesis (photon yield) at a given temperature. Processes involved in photosynthetic acclimation include increase in the quantity of photosynthetic enzymes at extreme temperatures to compensate for their decreased activity. In agreement with this hypothesis, *A. angustifolia* showed higher expression of certain photosynthesis genes at 30°C than at 18°C: 2 ribulose-bisphosphate carboxylase small subunit genes and 3 Rubisco activase genes. In addition, a number of genes directly involved in maintaining photosystem function and preventing photoinhibition were upregulated in *A. angustifolia* relative to *A. pungens* at 18°C: *DEG5*, *DEG8*, *CAO*, *PGR5*.

Secondly, a certain class of accessory proteins – heat-shock proteins (HSPs)/chaperones – is needed for heat tolerance, as these proteins increase cellular homeostasis through stabilization of membranes and proteins, and by helping in protein folding and assembly (Vierling, 1991; Yamori *et al.*, 2014). Supporting that notion, 5 heat shock protein binding genes (including 4 *HSP40*) and 5 genes involved in response to heat, including *HEAT SHOCK FACTOR 4*, were upregulated in *A. angustifolia* at 30°C relative to 18°C. Moreover, genes involved in combating redox stress were upregulated at 30°C in *A. angustifolia* relative to 18°C, including two genes in the ascorbate pathway and *AOR*, a chloroplast enzyme important for countering lipid peroxidation and preventing photosystem I inhibition (Yamauchi *et al.*, 2012). On the other hand, in *A. pungens*, significant over-representation of glutathione transferases was found among downregulated genes across the two temperatures.

Furthermore, alterations to membrane composition are highly important in acclimation to more extreme temperatures, as appropriate membrane fluidity needs to be maintained for the cell to carry out its normal functions, including photosynthesis. This is done through changes to the ratio of unsaturated and saturated fatty acids in membranes, which increases at low temperature and decreases at high temperature (Sung *et al.*, 2003; Falcone *et al.*, 2004), and increased expression of a omega-3 fatty acid desaturase at 30°C in *A. angustifolia* relative to *A. pungens* could be an example for this mechanism at play.

Interestingly, certain transcription factors tangentially involved in C_4 photosynthesis, such as *HB22*, *ATHB-12*, were upregulated in *A. pungens* relative to *A. angustifolia* at 30°C when plant growth is optimal and two transcription factors involved in, among other things, bundle sheath cell specification (*SCR*, *SHR*) were upregulated at either 18° or 30°C, each.

5.13 Temperature-dependent isoform switching in C_3 and C_4 photosynthesis

Evidence gathered in the study and previously prompted analysis of temperature-induced differential expression within genes - alternative enzyme isoforms with better performance

at low or high temperature. Detection of alternative enzyme isoforms was attempted indirectly through differential expression of SNP alleles across two temperatures within each species. Rubisco enzyme kinetics measurements indicate a drop in the efficiency and specificity of function at 18°C compared to 30°C in the C_4 species *A. pungens*, and no influence of temperature in the C_3 species *A. angustifolia* (M.V. Kapralov, unpublished), which corresponds to growth rate changes; different catalytic properties of Rubisco in C_3 and C_4 species are commonly found. Rubisco activation state is dependent on chaperone Rubisco activase, which remodels inhibited Rubisco to release it from inactive state (Mueller-Cajar *et al.*, 2014; Hauser *et al.*, 2015). Rubisco activase is generally a thermolabile enzyme but alternative, more thermostable Rubisco activase isoforms are expressed in some C_3 species, such as spinach (also Caryophyllales) at high temperatures (Crafts-Brandner *et al.*, 1997; Law & Crafts-Brandner, 2001; Law *et al.*, 2001). Consequently, it was hypothesised that the C_3 species *A. angustifolia* is able to maintain constant Rubisco activation and kinetics thanks to expression of alternative isoforms of Rubisco activase as well as its interacting parter Rubisco small subunit. In Rubisco activase, 9 SNPs with replacement substitutions in *A. angustifolia* and only 3 in *A. pungens* met the required pattern, while in Rubisco small subunit gene, 9 SNPs with replacement substitutions in *A. angustifolia* and 0 in *A. pungens* were found. The crystal structure of the Rubisco activase-rbS complex is not available but the exposure of the amino acids does not rule out their functional role in interaction between Rubisco small subunit and Rubisco activase. Therefore, the presence of a number of SNPs supports the hypothesis of temperature-induced isoform switching in *A. angustifolia* in two interacting partners, leading to stable Rubisco activation and photosynthesis rates across a wide range of temperatures seen *in vivo*.

Bibliography

Abdel-Ghany SE, **Burkhead JL**, **Gogolin KA**, **Andrés-Colás N**, **Bodecker JR**, **Puig S**, **Peñarrubia L**, **Pilon M**. **2005**. AtCCS is a functional homolog of the yeast copper chaperone Ccs1/Lys7. *FEBS Letters* **579**: 2307–2312.

Adamidis G, **Aloupi M**, **Kazakou E**, **Dimitrakopoulos P**. **2014**. Intra-specific variation in Ni tolerance, accumulation and translocation patterns in the Ni-hyperaccumulator *Alyssum lesbiacum*. *Chemosphere* **95**: 496–502.

Addou S, **Rentzsch R**, **Lee D**, **Orengo CA**. **2009**. Domain-based and family-specific sequence identity thresholds increase the levels of reliable protein function transfer. *Journal of Molecular Biology* **387**: 416–430.

Adham AR, **Zolman BK**, **Millius A**, **Bartel B**. **2005**. Mutations in *Arabidopsis* acyl-CoA oxidase genes reveal distinct and overlapping roles in beta-oxidation. *Plant Journal* **41**: 859–874.

Agrawal B, **Lakshmanan V**, **Kaushik S**, **Bais HP**. **2012**. Natural variation among *Arabidopsis* accessions reveals malic acid as a key mediator of nickel (Ni) tolerance. *Planta* **236**: 477–489.

Agrawal B, **Czymmek KJ**, **Sparks DL**, **Bais HP**. **2013**. Transient influx of nickel in root mitochondria modulates organic acid and reactive oxygen species production in nickel hyperaccumulator *Alyssum murale*. *Journal of Biological Chemistry* **288**: 7351–7362.

Aird D, **Ross MG**, **Chen WS**, **Danielsson M**, **Fennell T**, **Russ C**, **Jaffe DB**, **Nusbaum C**, **Gnirke A**. **2011**. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology* **12**: R18.

Albrecht V, **Ingenfeld A**, **Apel K**. **2006**. Characterization of the *Snowy cotyledon 1* mutant of *Arabidopsis thaliana*: the impact of chloroplast elongation factor G on chloroplast development and plant vitality. *Plant Molecular Biology* **60**: 507–518.

Alemán F, **Nieves-Cordones M**, **Martínez V**, **Rubio F**. **2011**. Root K^+ acquisition in plants: the *Arabidopsis thaliana* model. *Plant & Cell Physiology* **52**: 1603–1612.

Altschul SF, **Madden TL**, **Schäffer AA**, **Zhang J**, **Zhang Z**, **Miller W**, **Lipman DJ**. **1997**. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.

Álvarez-López V, **Prieto-Fernández Á**, **Cabello-Conejo M**, **Kidd P**. **2016**. Organic amendments for improving biomass production and metal yield of Ni-hyperaccumulating plants. *Science of the Total Environment* **548**: 370–379.

Alves S, **Nabais C**, **Simões Gonçalves MDL**, **Correia Dos Santos MM**. **2011**. Nickel speciation in the xylem sap of the hyperaccumulator *Alyssum serpyllifolium* ssp. *lusitanicum* growing on serpentine soils of northeast Portugal. *Journal of Plant Physiology* **168**: 1715–1722.

Amtmann A. **2009**. Learning from evolution: *Thellungiella* generates new knowledge on essential and critical components of abiotic stress tolerance in plants. *Molecular Plant* **2**: 3–12.

Anacker BL. **2014**. The nature of serpentine endemism. *American Journal of Botany* **101**: 219–224.

Anacker BL, **Harrison SP**. **2012**. Climate and the evolution of serpentine endemism in California. *Evolutionary Ecology* **26**: 1011–1023.

Anacker BL, **Whittall JB**, **Goldberg EE**, **Harrison SP**. **2011**. Origins and consequences of serpentine endemism in the California flora. *Evolution* **65**: 365–376.

Ančev M, **Goranova V**. **2006**. Trichome morphology of eleven genera of the tribe Alysseae (Brassicaceae) occurring in Bulgaria. *Willdenowia* **36**: 193–204.

Andersson I. **2008**. Catalysis and regulation in Rubisco. *Journal of Experimental Botany* **59**: 1555–1568.

Andersson I, **Backlund A**. **2008**. Structure and function of Rubisco. *Plant Physiology and Biochemistry* **46**: 275–291.

Andrews KR, **Luikart G**. **2014**. Recent novel approaches for population genomics data analysis. *Molecular Ecology* **23**: 1661–1667.

Andrés Z, **Pérez-Hormaeche J**, **Leidi EO**, **Schlücking K**, **Steinhorst L**, **McLachlan DH**, **Schumacher K**, **Hetherington AM**, **Kudla J**, **Cubero B**, **Pardo JM**. **2014**. Control of vacuolar dynamics and regulation of stomatal aperture by tonoplast potassium uptake. *Proceedings of the National Academy of Sciences* **111**: 1806–1814.

Andrés-Colás N, **Sancenón V**, **Rodríguez-Navarro S**, **Mayo S**, **Thiele DJ**, **Ecker JR**, **Puig S**, **Peñarrubia L**. **2006**. The *Arabidopsis* heavy metal P-type ATPase HMA5 interacts with metallochaperones and functions in copper detoxification of roots. *Plant Journal* **45**: 225–236.

Angeloni F, **Wagemaker CaM**, **Jetten MSM**, **Op den Camp HJM**, **Janssen-Megens EM**, **Francoijs KJ**, **Stunnenberg HG**, **Ouborg NJ**. **2011**. *De novo* transcriptome characterization and development of genomic tools for *Scabiosa columbaria* L. using next-generation sequencing techniques. *Molecular Ecology Resources* **11**: 662–674.

Anjum NA, **Aref IM**, **Duarte AC**, **Pereira E**, **Ahmad I**, **Iqbal M**. **2014**. Glutathione and proline can coordinately make plants withstand the joint attack of metal(loid) and salinity stresses. *Frontiers in Plant Science* **5**: 662.

Antonovics J. **1972**. Population dynamics of the grass *Anthoxanthum odoratum* on a zinc mine. *Journal of Ecology* **60**: 351–365.

Antonovics J, **Bradshaw A**. **1968**. Evolution in closely adjacent plant populations IV: Manifold effects of gene flow. *Heredity* **23**: 507–524.

Apse MP, **Aharon GS**, **Snedden WA**, **Blumwald E**. **1999**. Salt tolerance conferred by overexpression of a vacuolar Na^+/H^+ antiport in *Arabidopsis*. *Science* **285**: 1256–1258.

Asai T, **Tena G**, **Plotnikova J**, **Willmann MR**, **Chiu WL**, **Gomez-Gomez L**, **Boller T**, **Ausubel FM**, **Sheen J**. **2002**. MAP kinase signalling cascade in *Arabidopsis* innate immunity. *Nature* **415**: 977–983.

Asensi A, **Rodríguez N**, **Díez-Garretas B**, **Amils R**, **Boyd R**, **Baker A**, **Proctor J**, *et al.*. **2004**. Nickel hyperaccumulation of some subspecies of *Alyssum serpyllifolium* (Brassicaceae) from ultramafic soils on the Iberian Peninsula. In: *Ultramafic rocks: their soils, vegetation and fauna. Proceedings of the Fourth International Conference on Serpentine Ecology, Cuba, 21-26 April, 2003.*. Science Reviews. 263–265.

Ashburner M, **Ball CA**, **Blake JA**, **Botstein D**, **Butler H**, **Cherry JM**, **Davis AP**, **Dolinski K**, **Dwight SS**, **Eppig JT**, **Harris MA**, **Hill DP**, **Issel-Tarver L**, **Kasarskis A**, **Lewis S**, **Matese JC**, **Richardson JE**, **Ringwald M**, **Rubin GM**, **Sherlock G**. **2000**. Gene ontology: tool for the unification of biology. *Nature Genetics* **25**: 25–29.

Assunção AGL, **Pieper B**, **Vromans J**, **Lindhout P**, **Aarts MGM**, **Schat H**. **2006**. Construction of a genetic linkage map of *Thlaspi caerulescens* and quantitative trait loci analysis of zinc accumulation. *New Phytologist* **170**: 21–32.

Aubry S, **Kelly S**, **Kümpers BMC**, **Smith-Unna RD**, **Hibberd JM**. **2014**. Deep evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in two independent origins of C_4 photosynthesis. *PLoS Genetics* **10**: e1004365.

Babitha K, **Ramu S**, **Pruthvi V**, **Mahesh P**, **Nataraja K**, **Udayakumar M**. **2013**. Co-expression of *At*bHLH17 and *At*WRKY28 confers resistance to abiotic stress in *Arabidopsis*. *Transgenic Research* **22**: 327–341.

Bajgain P, **Richardson BA**, **Price JC**, **Cronn RC**, **Udall JA**. **2011**. Transcriptome characterization and polymorphism detection between subspecies of big sagebrush (*Artemisia tridentata*). *BMC Genomics* **12**: 370.

Baker A, **Brooks R**. **1989**. Terrestrial higher plants which hyperaccumulate metallic elements. A review of their distribution, ecology and phytochemistry. *Biorecovery* **1**: 81–126.

Balagtas-Burow GE, **Moroney JV**, **Longstreth DJ**. **1993**. Growth and osmotic adjustment of cultured suspension cells from *Alternanthera philoxeroides* (Mart.) Griseb after an abrupt increase in salinity. *Journal of Experimental Botany* **44**: 673–679.

Baliardini C, **Meyer CL**, **Salis P**, **Saumitou-Laprade P**, **Verbruggen N**. **2015**. CAX1 co-segregates with Cd tolerance in the metal hyperaccumulator *Arabidopsis halleri* and plays a role in limiting oxidative stress in *Arabidopsis*. *Plant Physiology* **169**: 549–559.

Barbará T, **Palma-Silva C**, **Paggi GM**, **Bered F**, **Fay MF**, **Lexer C**. **2007**. Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology* **16**: 3759–3767.

Barragán V, **Leidi EO**, **Andrés Z**, **Rubio L**, **De Luca A**, **Fernández JA**, **Cubero B**, **Pardo JM**. **2012**. Ion exchangers NHX1 and NHX2 mediate active potassium uptake into vacuoles to regulate cell turgor and stomatal function in *Arabidopsis*. *Plant Cell* **24**: 1127–1142.

Barrett RD, **Schluter D**. **2008**. Adaptation from standing genetic variation. *Trends in Ecology & Evolution* **23**: 38–44.

Bassham DC, **Brandizzi F**, **Otegui MS**, **Sanderfoot AA**. **2008**. The secretory system of *Arabidopsis*. *The Arabidopsis Book* **6**.

Bassil E, **Tajima H**, **Liang YC**, **Ohto Ma**, **Ushijima K**, **Nakano R**, **Esumi T**, **Coku A**, **Belmonte M**, **Blumwald E**. **2011**. The *Arabidopsis* Na^+/H^+ antiporters NHX1 and NHX2 control vacuolar pH and K^+ homeostasis to regulate growth, flower development, and reproduction. *Plant Cell* **23**: 3482–3497.

Basu D, **Le J**, **Zakharova T**, **Mallery EL**, **Szymanski DB**. **2008**. A SPIKE1 signaling complex controls actin-dependent cell morphogenesis through the heteromeric WAVE and ARP2/3 complexes. *Proceedings of the National Academy of Sciences* **105**: 4044–4049.

Baud S, **Guyon V**, **Kronenberger J**, **Wuillème S**, **Miquel M**, **Caboche M**, **Lepiniec L**, **Rochat C**. **2003**. Multifunctional acetyl-CoA carboxylase 1 is essential for very long chain fatty acid elongation and embryo development in *Arabidopsis*. *Plant Journal* **33**: 75–86.

Baumbach H, **Hellwig F**. **2007**. Genetic differentiation of metallicolous and non-metallicolous *Armeria maritima* (Mill.) Willd. taxa (Plumbaginaceae) in Central Europe. *Plant Systematics and Evolution* **269**: 245–258.

Baxter I, **Dilkes BP**. **2012**. Elemental profiles reflect plant adaptations to the environment. *Science* **336**: 1661–1663.

Beaumont MA. **2005**. Adaptation and speciation: what can F_{ST} tell us? *Trends in Ecology & Evolution* **20**: 435–440.

Beaumont MA, **Nichols RA**. **1996**. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **263**: 1619–1626.

Becerra-Castro C. **2012**. *Role of plant-associated bacteria in the remediation of contaminated soils*. Universidade de Santiago de Compostela PhD thesis.

Becher M, **Talke IN**, **Krall L**, **Krämer U**. **2004**. Cross-species microarray transcript profiling reveals high constitutive expression of metal homeostasis genes in shoots of the zinc hyperaccumulator *Arabidopsis halleri*. *Plant Journal* **37**: 251–268.

Bednarek P, **Piślewska-Bednarek M**, **Svatoš A**, **Schneider B**, **Doubský J**, **Mansurova M**,

Humphry M, **Consonni C**, **Panstruga R**, **Sanchez-Vallet A**, **Molina A**, **Schulze-Lefert P**. **2009**. A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science* **323**: 101–106.

Beilstein MA, **Al-Shehbaz IA**, **Kellogg EA**. **2006**. Brassicaceae phylogeny and trichome evolution. *American Journal of Botany* **93**: 607–619.

Beilstein MA, **Al-Shehbaz IA**, **Mathews S**, **Kellogg EA**. **2008**. Brassicaceae phylogeny inferred from *phytochrome A* and *ndhF* sequence data: tribes and trichomes revisited. *American Journal of Botany* **95**: 1307–1327.

Beilstein MA, **Nagalingum NS**, **Clements MD**, **Manchester SR**, **Mathews S**. **2010**. Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* **107**: 18724–18728.

Benatti M, **Yookongkaew N**, **Meetam M**, **Guo WJ**, **Punyasuk N**, **AbuQamar S**, **Goldsbrough P**. **2014**. Metallothionein deficiency impacts copper accumulation and redistribution in leaves and seeds of *Arabidopsis*. *New Phytologist* **202**: 940–951.

Berglund ABN, **Dahlgren S**, **Westerbergh A**. **2003**. Evidence for parallel evolution and site-specific selection of serpentine tolerance in *Cerastium alpinum* during the colonization of Scandinavia. *New Phytologist* **161**: 199–209.

Bernard A, **Domergue F**, **Pascal S**, **Jetter R**, **Renne C**, **Faure JD**, **Haslam RP**, **Napier JA**, **Lessire R**, **Joubès J**. **2012**. Reconstitution of plant alkane biosynthesis in yeast demonstrates that *Arabidopsis* ECERIFERUM1 and ECERIFERUM3 are core components of a very-long-chain alkane synthesis complex. *Plant Cell* **24**: 3106–3118.

Berry J, **Björkman O**. **1980**. Photosynthetic response and adaptation to temperature in higher plants. *Annual Review of Plant Physiology* **31**: 491–543.

Bert V, **Macnair MR**, **De Laguerie P**, **Saumitou-Laprade P**, **Petit D**. **2000**. Zinc tolerance and accumulation in metallicolous and nonmetallicolous populations of *Arabidopsis halleri* (Brassicaceae). *New Phytologist* **146**: 225–233.

Bert V, **Bonnin I**, **Saumitou-Laprade P**, **De Laguérie P**, **Petit D**. **2002**. Do *Arabidopsis halleri* from nonmetallicolous populations accumulate zinc and cadmium more effectively than those from metallicolous populations? *New Phytologist* **155**: 47–57.

Bert V, **Meerts P**, **Saumitou-Laprade P**, **Salis P**, **Gruber W**, **Verbruggen N**. **2003**. Genetic basis of Cd tolerance and hyperaccumulation in *Arabidopsis halleri*. *Plant and Soil* **249**: 9–18.

Besnard G, **Basic N**, **Christin PA**, **Savova-Bianchi D**, **Galland N**. **2009**. *Thlaspi caerulescens* (Brassicaceae) population genetics in western Switzerland: is the genetic structure affected by natural variation of soil heavy metal concentrations? *New Phytologist* **181**: 974–984.

Besseau S, **Li J**, **Palva ET**. **2012**. WRKY54 and WRKY70 co-operate as negative regulators of leaf senescence in *Arabidopsis thaliana*. *Journal of Experimental Botany* **63**: 2667–2679.

Binder BM, **Rodríguez FI**, **Bleecker AB**. **2010**. The copper transporter RAN1 is essential for biogenesis of ethylene receptors in *Arabidopsis*. *Journal of Biological Chemistry* **285**: 37263–37270.

Bininda-Emonds O. **2005**. transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. *BMC Bioinformatics* **6**: 156.

Bizoux JP, **Daïnou K**, **Raspé O**, **Lutts S**, **Mahy G**. **2008**. Fitness and genetic variation of *Viola calaminaria*, an endemic metallophyte: implications of population structure and history. *Plant Biology* **10**: 684–693.

Bjornson M, **Benn G**, **Song X**, **Comai L**, **Franz AK**, **Dandekar AM**, **Drakakaki G**, **Dehesh K**. **2014**. Distinct roles for mitogen-activated protein kinase signaling and CALMODULIN-BINDING TRANSCRIPTIONAL ACTIVATOR 3 in regulating the peak time and amplitude of the plant general stress response. *Plant Physiology* **166**: 988–996.

Boex-Fontvieille ERA, **Gauthier PPG**, **Gilard F**, **Hodges M**, **Tcherkez GGB**. **2013**. A new

anaplerotic respiratory pathway involving lysine biosynthesis in isocitrate dehydrogenase-deficient *Arabidopsis* mutants. *New Phytologist* **199**: 673–682.

Bolger AM, **Lohse M**, **Usadel B**. **2014**. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Bonnet ALM. **1963**. Contribution à l'étude caryologique du genre *Alyssum* L. (s. lat.). *Naturalia Monspeliensia. Serie Botanique* **15**: 41–52.

Boorer KJ, **Fischer WN**. **1997**. Specificity and stoichiometry of the *Arabidopsis* H^+/amino acid transporter AAP5. *Journal of Biological Chemistry* **272**: 13040–13046.

Bouckaert R, **Heled J**, **Kühnert D**, **Vaughan T**, **Wu CH**, **Xie D**, **Suchard MA**, **Rambaut A**, **Drummond AJ**. **2014**. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**: e1003537.

Bouckaert RR. **2010**. DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* **26**: 1372–1373.

Boutigny S, **Sautron E**, **Finazzi G**, **Rivasseau C**, **Frelet-Barrand A**, **Pilon M**, **Rolland N**, **Seigneurin-Berny D**. **2014**. HMA1 and PAA1, two chloroplast-envelope PIB-ATPases, play distinct roles in chloroplast copper homeostasis. *Journal of Experimental Botany* **65**: 1529–1540.

Boyd RS. **2004**. Ecology of metal hyperaccumulation. *New Phytologist* **162**: 563–567.

Boyd RS. **2007**. The defense hypothesis of elemental hyperaccumulation: status, challenges and new directions. *Plant and Soil* **293**: 153–176.

Boyd RS, **Martens S**. **1992**. The raison d'être for metal hyperaccumulation by plants. In: **Baker A**, **Proctor J**, **Reeves R**, (eds.) *The Vegetation of Ultramafic (Serpentine) Soils*. Andover, UK: Intercept Ltd. 279–289.

Boyes DC, **Nam J**, **Dangl JL**. **1998**. The *Arabidopsis thaliana* RPM1 disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response. *Proceedings of the National Academy of Sciences* **95**: 15849–15854.

Bradnam KR, **Fass JN**, **Alexandrov A**, **Baranay P**, **Bechner M**, **Birol I**, **Boisvert S**, **Chapman Ja**, **Chapuis G**, **Chikhi R**, **Chitsaz H**, **Chou WC**, **Corbeil J**, **Del Fabbro C**, **Docking TR**, **Durbin R**, **Earl D**, **Emrich S**, **Fedotov P**, **Fonseca Na**, **Ganapathy G**, **Gibbs Ra**, **Gnerre S**, **Godzaridis E**, **Goldstein S**, **Haimel M**, **Hall G**, **Haussler D**, **Hiatt JB**, **Ho IY**, **Howard J**, **Hunt M**, **Jackman SD**, **Jaffe DB**, **Jarvis E**, **Jiang H**, **Kazakov S**, **Kersey PJ**, **Kitzman JO**, **Knight JR**, **Koren S**, **Lam TW**, **Lavenier D**, **Laviolette F**, **Li Y**, **Li Z**, **Liu B**, **Liu Y**, **Luo R**, **Maccallum I**, **Macmanes MD**, **Maillet N**, **Melnikov S**, **Naquin D**, **Ning Z**, **Otto TD**, **Paten B**, **Paulo OS**, **Phillippy AM**, **Pina-Martins F**, **Place M**, **Przybylski D**, **Qin X**, **Qu C**, **Ribeiro FJ**, **Richards S**, **Rokhsar DS**, **Ruby JG**, **Scalabrin S**, **Schatz MC**, **Schwartz DC**, **Sergushichev A**, **Sharpe T**, **Shaw TI**, **Shendure J**, **Shi Y**, **Simpson JT**, **Song H**, **Tsarev F**, **Vezzi F**, **Vicedomini R**, **Vieira BM**, **Wang J**, **Worley KC**, **Yin S**, **Yiu SM**, **Yuan J**, **Zhang G**, **Zhang H**, **Zhou S**, **Korf IF**. **2013**. Assemblathon 2: evaluating *de novo* methods of genome assembly in three vertebrate species. *GigaScience* **2**: 10.

Bradshaw H. **2005**. Mutations in CAX1 produce phenotypes characteristic of plants tolerant to serpentine soils. *New Phytologist* **167**: 81–88.

Brady KU, **Kruckeberg AR**, **Bradshaw Jr H**. **2005**. Evolutionary ecology of plant adaptation to serpentine soils. *Annual Review of Ecology, Evolution, and Systematics* **36**: 243–266.

Bratteler M, **Baltsberger M**, **Widmer A**. **2006a**. QTL analysis of intraspecific differences between two *Silene vulgaris* ecotypes. *Annals of Botany* **98**: 411–419.

Bratteler M, **Lexer C**, **Widmer A**. **2006b**. Genetic architecture of traits associated with serpentine adaptation of *Silene vulgaris*. *Journal of Evolutionary Biology* **19**: 1149–1156.

Bräutigam A, **Kajala K**, **Wullenweber J**, **Sommer M**, **Gagneul D**, **Weber KL**, **Carr KM**, **Gowik U**, **Mass J**, **Lercher MJ**, **Westhoff P**, **Hibberd JM**, **Weber APM**. **2011**. An mRNA blueprint for

C_4 photosynthesis derived from comparative transcriptomics of closely related C_3 and C_4 species. *Plant Physiology* **155**: 142–156.

Bräutigam A, **Schliesky S**, **Külahoglu C**, **Osborne CP**, **Weber AP**. **2014**. Towards an integrative model of C_4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C_4 species. *Journal of Experimental Botany* **65**: 3579–3593.

Bremer B, **Bremer K**, **Chase M**, **Fay M**, **Reveal J**, **Soltis D**, **Soltis P**, **Stevens P**. **2009**. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* **161**: 105–121.

Bringmann M, **Li E**, **Sampathkumar A**, **Kocabek T**, **Hauser MT**, **Persson S**. **2012**. POM-POM2/CELLULOSE SYNTHASE INTERACTING 1 is essential for the functional association of cellulose synthase and microtubules in *Arabidopsis*. *Plant Cell* **24**: 163–177.

Brockington SF, **Yang Y**, **Gandia-Herrero F**, **Covshoff S**, **Hibberd JM**, **Sage RF**, **Wong GKS**, **Moore MJ**, **Smith SA**. **2015**. Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. *New Phytologist* **207**: 1170–1180.

Brooks RR. **1987**. *Serpentine and Its Vegetation: A Multidisciplinary Approach*. Portland, USA: Dioscorides Press.

Brooks RR, **Radford CC**. **1978**. Nickel accumulation by European species of the genus *Alyssum*. *Proceedings of the Royal Society of London B: Biological Sciences* **200**: 217–224.

Brooks RR, **Robinson B**. **1998**. *Plants That Hyperaccumulate Heavy Metals*. Wallingford, UK: CAB International.

Brooks RR, **Yang Xh**. **1984**. Elemental levels and relationships in the endemic serpentine flora of the Great Dyke, Zimbabwe and their significance as controlling factors for the flora. *Taxon* **33**: 392–399.

Brooks RR, **Lee J**, **Reeves RD**, **Jaffré T**. **1977**. Detection of nickeliferous rocks by analysis of herbarium specimens of indicator plants. *Journal of Geochemical Exploration* **7**: 49–57.

Brooks RR, **Morrison RS**, **Reeves RD**, **Dudley TR**, **Akman Y**. **1979**. Hyperaccumulation of nickel by *Alyssum* Linnaeus (Cruciferae). *Proceedings of the Royal Society of London. Series B, Biological Sciences* **203**: 387–403.

Brooks RR, **Shaw S**, **Asensi Marfil A**. **1981**. The chemical form and physiological function of nickel in some Iberian *Alyssum* species. *Physiologia Plantarum* **51**: 167–170.

Brosché M, **Blomster T**, **Salojärvi J**, **Cui F**, **Sipari N**, **Leppälä J**, **Lamminmäki A**, **Tomai G**, **Narayanasamy S**, **Reddy RA**, **Keinänen M**, **Overmyer K**, **Kangasjärvi J**. **2014**. Transcriptomics and functional genomics of ROS-induced cell death regulation by Radical-Induced Cell Death 1. *PLoS Genetics* **10**: e1004112.

Brown RP, **Yang Z**. **2011**. Rate variation and estimation of divergence times using strict and relaxed clocks. *BMC Evolutionary Biology* **11**: 271.

Brown S, **Angle J**, **Chaney R**, **Baker A**. **1995**. Zinc and cadmium uptake by hyperaccumulator *Thlaspi caerulescens* grown in nutrient solution. *Soil Science* **59**: 125–133.

Bryant D, **Bouckaert R**, **Felsenstein J**, **Rosenberg NA**, **RoyChoudhury A**. **2012**. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution* **29**: 1917–1932.

Buckhout T, **Yang T**, **Schmidt W**. **2009**. Early iron-deficiency-induced transcriptional changes in *Arabidopsis* roots as revealed by microarray analyses. *BMC Genomics* **10**: 147.

Buckingham GR. **1996**. Biological control of alligator weed, *Alternanthera philoxeroides*, the world's first aquatic weed success story. *Castanea* **61**: 232–243.

Bullard JH, **Purdom E**, **Hansen KD**, **Dudoit S**. **2010**. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**: 94.

Burge DO, **Barker WR**. **2010**. Evolution of nickel hyperaccumulation by *Stackhousia tryonii* (Celastraceae), a serpentinite-endemic plant from Queensland, Australia. *Australian Systematic Botany* **23**: 415–430.

Burke MK, **Dunham JP**, **Shahrestani P**, **Thornton KR**, **Rose MR**, **Long AD**. **2010**. Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature* **467**: 587–590.

Burkhead JL, **Gogolin Reynolds KA**, **Abdel-Ghany SE**, **Cohu CM**, **Pilon M**. **2009**. Copper homeostasis. *New Phytologist* **182**: 799–816.

Burrell AM, **Hawkins AK**, **Pepper AE**. **2012**. Genetic analyses of nickel tolerance in a North American serpentine endemic plant, *Caulanthus amplexicaulis* var. *barbarae* (Brassicaceae). *American Journal of Botany* **99**: 1875–1883.

Cabello-Conejo M, **Becerra-Castro C**, **Prieto-Fernández A**, **Monterroso C**, **Saavedra-Ferro A**, **Mench M**, **Kidd P**. **2014**. Rhizobacterial inoculants can improve nickel phytoextraction by the hyperaccumulator *Alyssum pintodasilvae*. *Plant and Soil* **379**: 35–50.

Cai C, **Henty-Ridilla JL**, **Szymanski DB**, **Staiger CJ**. **2014**. *Arabidopsis* myosin XI: a motor rules the tracks. *Plant Physiology* **166**: 1359–1370.

Calderón CP, **García Aseff SB**, **Fuentes LB**. **1997**. Evaluation of diuretic activity of *Alternanthera pungens* extract in rats. *Phytotherapy Research* **11**: 606–608.

Callahan DL, **Baker AJM**, **Kolev SD**, **Wedd AG**. **2006**. Metal ion ligands in hyperaccumulating plants. *Journal of Biological Inorganic Chemistry* **11**: 2–12.

Callahan DL, **Kolev SD**, **O'Hair RAJ**, **Salt DE**, **Baker AJM**. **2007**. Relationships of nicotianamine and other amino acids with nickel, zinc and iron in *Thlaspi* hyperaccumulators. *New Phytologist* **176**: 836–848.

Callard D, **Axelos M**, **Mazzolini L**. **1996**. Novel molecular markers for late phases of the growth cycle of *Arabidopsis thaliana* cell-suspension cultures are expressed during organ senescence. *Plant Physiology* **112**: 705–715.

Cantarel BL, **Weaver D**, **McNeill N**, **Zhang J**, **Mackey AJ**, **Reese J**. **2014**. BAYSIC: a Bayesian method for combining sets of genome variants with improved specificity and sensitivity. *BMC Bioinformatics* **15**: 104.

Carstens BC, **Brennan RS**, **Chua V**, **Duffie CV**, **Harvey MG**, **Koch Ra**, **McMahan CD**, **Nelson BJ**, **Newman CE**, **Satler JD**, **Seeholzer G**, **Posbic K**, **Tank DC**, **Sullivan J**. **2013a**. Model selection as a tool for phylogeographic inference: an example from the willow *Salix melanopsis*. *Molecular Ecology* **22**: 4014–4028.

Carstens BC, **Pelletier TA**, **Reid NM**, **Satler JD**. **2013b**. How to fail at species delimitation. *Molecular Ecology* **22**: 4369–4383.

Caspi R, **Altman T**, **Billington R**, **Dreher K**, **Foerster H**, **Fulcher CA**, **Holland TA**, **Keseler IM**, **Kothari A**, **Kubo A**, **Krummenacker M**, **Latendresse M**, **Mueller LA**, **Ong Q**, **Paley S**, **Subhraveti P**, **Weaver DS**, **Weerasinghe D**, **Zhang P**, **Karp PD**. **2013**. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research* **42**: 459–471.

Cecchi L, **Colzi I**, **Coppi A**, **Gonnelli C**, **Selvi F**. **2013**. Diversity and biogeography of Ni-hyperaccumulators of *Alyssum* section *Odontarrhena* (Brassicaceae) in the central western Mediterranean: evidence from karyology, morphology and DNA sequence data. *Botanical Journal of the Linnean Society* **173**: 269–289.

Cecchini NM, **Monteoliva MI**, **Alvarez ME**. **2011**. Proline dehydrogenase contributes to pathogen defense in *Arabidopsis*. *Plant Physiology* **155**: 1947–1959.

Cen YP, **Sage RF**. **2005**. The regulation of Rubisco activity in response to variation in temperature and atmospheric CO_2 partial pressure in sweet potato. *Plant Physiology* **139**: 979–990.

Centofanti T, **Sayers Z**, **Cabello-Conejo MI**, **Kidd P**, **Nishizawa NK**, **Kakei Y**, **Davis AP**, **Sicher RC**, **Chaney RL**. **2013**. Xylem exudate composition and root-to-shoot nickel translocation in *Alyssum* species. *Plant and Soil* **373**: 59–75.

Chan YF, **Marks ME**, **Jones FC**, **Villarreal G**, **Shapiro MD**, **Brady SD**, **Southwick AM**, **Absher DM**, **Grimwood J**, **Schmutz J**, *et al.*. **2010**. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**: 302–305.

Chaney RL, **Angle JS**, **Broadhurst CL**, **Peters CA**, **Tappero RV**, **Sparks DL**. **2007**. Improved understanding of hyperaccumulation yields commercial phytoextraction and phytomining technologies. *Journal of Environmental Quality* **36**: 1429–1443.

Chaney RL, **Broadhurst CL**, **Centofanti T**. **2010**. *Trace Elements in Soils*. Oxford, UK: Blackwell Publishing.

Chao DY, **Dilkes B**, **Luo H**, **Douglas A**, **Yakubova E**, **Lahner B**, **Salt DE**. **2013**. Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. *Science* **341**: 658–659.

Charlesworth D, **Charlesworth B**. **1987**. Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics* **18**: 237–268.

Chen Cc, **Chien WF**, **Lin NC**, **Yeh KC**. **2014a**. Alternative functions of *Arabidopsis* Yellow Stripe-Like 3: from metal translocation to pathogen defense. *PLoS One* **9**: e98008.

Chen L, **Bush DR**. **1997**. LHT1, a lysine- and histidine-specific amino acid transporter in *Arabidopsis*. *Plant Physiology* **115**: 1127–1134.

Chen M, **Thelen JJ**. **2013**. Acyl-Lipid Desaturase 2 is required for chilling and freezing tolerance in *Arabidopsis*. *Plant Cell* **25**: 1430–1444.

Chen X, **Liu J**, **Lin G**, **Wang A**, **Wang Z**, **Lu G**. **2013**. Overexpression of *At*WRKY28 and *At*WRKY75 in *Arabidopsis* enhances resistance to oxalic acid and *Sclerotinia sclerotiorum*. *Plant Cell Reports* **32**: 1589–1599.

Chen Y, **Lun A**, **Smyth G**. **2014b**. Differential expression analysis of complex RNA-seq experiments using edgeR. *Frontiers in Probability and the Statistical Sciences* : 51–74.

Cheng AY, **Teo YY**, **Ong RTH**. **2014**. Assessing single nucleotide variant detection and genotype calling on whole-genome sequenced individuals. *Bioinformatics* **30**: 1707–1713.

Cheng NH, **Liu JZ**, **Brock A**, **Nelson RS**, **Hirschi KD**. **2006**. *At*GRXcp, an *Arabidopsis* chloroplastic glutaredoxin, is critical for protection against protein oxidative damage. *Journal of Biological Chemistry* **281**: 26280–26288.

Chevin LM, **Martin G**, **Lenormand T**. **2010**. Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution* **64**: 3213–3231.

Chiang HC, **Lo JC**, **Yeh KC**. **2006**. Genes associated with heavy metal tolerance and accumulation in Zn/Cd hyperaccumulator *Arabidopsis halleri*: a genomic survey with cDNA microarray. *Environmental Science & Technology* **40**: 6792–6798.

Choi H, **Ohyama K**, **Kim YY**, **Jin JY**, **Lee SB**, **Yamaoka Y**, **Muranaka T**, **Suh MC**, **Fujioka S**, **Lee Y**. **2014**. The role of *Arabidopsis* ABCG9 and ABCG31 ATP binding cassette transporters in pollen fitness and the deposition of steryl glycosides on the pollen coat. *Plant Cell* **26**: 310–324.

Christin PA, **Freckleton RP**, **Osborne CP**. **2010**. Can phylogenetics identify C_4 origins and reversals? *Trends in Ecology & Evolution* **25**: 403–409.

Christin PA, **Sage TL**, **Edwards EJ**, **Ogburn RM**, **Khoshravesh R**, **Sage RF**. **2011**. Complex evolutionary transitions and the significance of C_3–C_4 intermediate forms of photosynthesis in Molluginaceae. *Evolution* **65**: 643–660.

Christin PA, **Osborne CP**, **Chatelet DS**, **Columbus JT**, **Besnard G**, **Hodkinson TR**, **Garrison LM**, **Vorontsova MS**, **Edwards EJ**. **2013**. Anatomical enablers and the evolution of C_4 photosynthesis in grasses. *Proceedings of the National Academy of Sciences* **110**: 1381–1386.

Christin PA, **Arakaki M**, **Osborne CP**, **Bräutigam A**, **Sage RF**, **Hibberd JM**, **Kelly S**, **Covshoff**

S, **Wong GKS**, **Hancock L**, **Edwards EJ**. **2014**. Shared origins of a key enzyme during the evolution of C_4 and CAM metabolism. *Journal of Experimental Botany* **65**: 3609–3621.

Christin PA, **Arakaki M**, **Osborne CP**, **Edwards EJ**. **2015**. Genetic enablers underlying the clustered evolutionary origins of C_4 photosynthesis in angiosperms. *Molecular Biology and Evolution* **32**: 846–858.

Chu CC, **Lee WC**, **Guo WY**, **Pan SM**, **Chen LJ**, **Li HM**, **Jinn TL**. **2005**. A copper chaperone for superoxide dismutase that confers three types of copper/zinc superoxide dismutase activity in *Arabidopsis*. *Plant Physiology* **139**: 425–436.

Chu HH, **Chiecko J**, **Punshon T**, **Lanzirotti A**, **Lahner B**, **Salt DE**, **Walker EL**. **2010**. Successful reproduction requires the function of *Arabidopsis* Yellow Stripe-Like 1 and Yellow Stripe-Like 3 metal-nicotianamine transporters in both vegetative and reproductive structures. *Plant Physiology* **154**: 197–210.

Ciarbelli AR, **Ciolfi A**, **Salvucci S**, **Ruzza V**, **Possenti M**, **Carabelli M**, **Fruscalzo A**, **Sessa G**, **Morelli G**, **Ruberti I**. **2008**. The *Arabidopsis* Homeodomain-leucine Zipper II gene family: diversity and redundancy. *Plant Molecular Biology* **68**: 465–478.

Clay NK, **Adio AM**, **Denoux C**, **Jander G**, **Ausubel FM**. **2009**. Glucosinolate metabolites required for an *Arabidopsis* innate immune response. *Science* **323**: 95–101.

Cloonan N, **Forrest ARR**, **Kolle G**, **Gardiner BBA**, **Faulkner GJ**, **Brown MK**, **Taylor DF**, **Steptoe AL**, **Wani S**, **Bethel G**, **Robertson AJ**, **Perkins AC**, **Bruce SJ**, **Lee CC**, **Ranade SS**, **Peckham HE**, **Manning JM**, **McKernan KJ**, **Grimmond SM**. **2008**. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods* **5**: 613–619.

Colosimo PF, **Hosemann KE**, **Balabhadra S**, **Villarreal G**, **Dickson M**, **Grimwood J**, **Schmutz J**, **Myers RM**, **Schluter D**, **Kingsley DM**. **2005**. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**: 1928–1933.

Conn SJ, **Conn V**, **Tyerman SD**, **Kaiser BN**, **Leigh RA**, **Gilliham M**. **2011**. Magnesium transporters, MGT2/MRS2-1 and MGT3/MRS2-5, are important for magnesium partitioning within *Arabidopsis thaliana* mesophyll vacuoles. *New Phytologist* **190**: 583–594.

Consonni C, **Humphry M**, **Hartmann H**, **Livaja M**, **Durner J**, **Westphal L**, **Vogel J**, **Lipka V**, **Kemmerling B**, **Schulze-Lefert P**, **Somerville S**, **Panstruga R**. **2006**. Conserved requirement for a plant host cell protein in powdery mildew pathogenesis. *Nature Genetics* **38**: 716–720.

Consonni C, **Bednarek P**, **Humphry M**, **Francocci F**, **Ferrari S**, **Harzen A**, **Ver Loren van Themaat E**, **Panstruga R**. **2010**. Tryptophan-derived metabolites are required for antifungal defense in the *Arabidopsis mlo2* mutant. *Plant Physiology* **152**: 1544–1561.

Conte GL, **Arnegard ME**, **Peichel CL**, **Schluter D**. **2012**. The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **279**: 5039–5047.

Coombs J, **Letcher B**, **Nislow K**. **2008**. CREATE: a software to create input files from diploid genotypic data for 52 genetic software programs. *Molecular Ecology Resources* **8**: 578–580.

Cornu JY, **Deinlein U**, **Höreth S**, **Braun M**, **Schmidt H**, **Weber M**, **Persson DP**, **Husted S**, **Schjoerring JK**, **Clemens S**. **2015**. Contrasting effects of nicotianamine synthase knockdown on zinc and nickel tolerance and accumulation in the zinc/cadmium hyperaccumulator *Arabidopsis halleri*. *New Phytologist* **206**: 738–750.

Courbot M, **Willems G**, **Motte P**, **Arvidsson S**, **Roosens N**, **Saumitou-Laprade P**, **Verbruggen N**. **2007**. A major quantitative trait locus for cadmium tolerance in *Arabidopsis halleri* colocalizes with *HMA4*, a gene encoding a heavy metal ATPase. *Plant Physiology* **144**: 1052–1065.

Craciun AR, **Meyer CL**, **Chen J**, **Roosens N**, **De Groodt R**, **Hilson P**, **Verbruggen N**. **2012**. Variation in *HMA4* gene copy number and expression among *Noccaea caerulescens* populations presenting different levels of Cd tolerance and accumulation. *Journal of Experimental Botany* **63**:

4179–4189.

Cracraft J. **1989**. Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In: **Otte D**, **Endler JA**, (eds.) *Speciation and Its Consequences*. Sunderland, USA: Sinauer Associates. 28–59.

Crafts-Brandner SJ, **van de Loo FJ**, **Salvucci ME**. **1997**. The two forms of ribulose-1, 5-bisphosphate carboxylase/oxygenase activase differ in sensitivity to elevated temperature. *Plant Physiology* **114**: 439–444.

Cruickshank TE, **Hahn MW**. **2014**. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* **23**: 3133–3157.

Cui H, **Kong D**, **Liu X**, **Hao Y**. **2014**. SCARECROW, SCR-LIKE 23 and SHORT-ROOT control bundle sheath cell fate and function in *Arabidopsis thaliana*. *Plant Journal* **78**: 319–327.

Cutter AD. **2013**. Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. *Molecular Phylogenetics and Evolution* **69**: 1172–1185.

da Silva J, **Williams R**. **1991**. *The Biological Chemistry of the Elements: The Inorganic Chemistry of Life*. Oxford, UK: Oxford University Press.

Dadacz-Narloch B, **Beyhl D**, **Larisch C**, **López-Sanjurjo EJ**, **Reski R**, **Kuchitsu K**, **Müller TD**, **Becker D**, **Schönknecht G**, **Hedrich R**. **2011**. A novel calcium binding site in the slow vacuolar cation channel TPC1 senses luminal calcium levels. *Plant Cell* **23**: 2696–2707.

Davey JW, **Hohenlohe PA**, **Etter PD**, **Boone JQ**, **Catchen JM**, **Blaxter ML**. **2011**. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**: 499–510.

De Angeli A, **Baetz U**, **Francisco R**, **Zhang J**, **Chaves M**, **Regalado A**. **2013**. The vacuolar channel *Vv*ALMT9 mediates malate and tartrate accumulation in berries of *Vitis vinifera*. *Planta* **238**: 283–291.

De Angeli A, **Zhang J**, **Meyer S**, **Martinoia E**. **2013**. *At*ALMT9 is a malate-activated vacuolar chloride channel required for stomatal opening in *Arabidopsis*. *Nature Communications* **4**: 1804.

de la Fuente V, **Rodríguez N**, **Díez-Garretas B**, **Rufo L**, **Asensi A**, **Amils R**. **2007**. Nickel distribution in the hyperaccumulator *Alyssum serpyllifolium* Desf. spp. from the Iberian Peninsula. *Plant Biosystems* **141**: 170–180.

De Queiroz K. **2007**. Species concepts and species delimitation. *Systematic Biology* **56**: 879–886.

De Varennes A, **Torres MO**, **Neto MMPM**, **Coutinho JF**, **Rocha MMGS**. **1996**. Effects of heavy metals on the growth and mineral composition of a nickel hyperaccumulator. *Journal of Plant Nutrition* **19**: 669–676.

De Wit P, **Pespeni MH**, **Palumbi SR**. **2015**. SNP genotyping and population genomics from expressed sequences –current advances and future possibilities. *Molecular Ecology* **24**: 2310–2323.

Deinlein U, **Weber M**, **Schmidt H**, **Rensch S**, **Trampczynska A**, **Hansen TH**, **Husted S**, **Schjoerring JK**, **Talke IN**, **Krämer U**, **Clemens S**. **2012**. Elevated nicotianamine levels in *Arabidopsis halleri* roots play a key role in zinc hyperaccumulation. *Plant Cell* **24**: 708–723.

Deng F, **Yamaji N**, **Xia J**, **Ma JF**. **2013**. A member of the heavy metal P-type ATPase *Os*HMA5 is involved in xylem loading of copper in rice. *Plant Physiology* **163**: 1353–1362.

Deng J, **Liao B**, **Ye M**, **Deng D**, **Lan C**, **Shu W**. **2007**. The effects of heavy metal pollution on genetic diversity in zinc/cad- mium hyperaccumulator *Sedum alfredii* populations. *Plant and Soil* **297**: 83–92.

Dengler NG, **Dengler RE**, **Donnelly PM**, **Hattersley PW**. **1994**. Quantitative leaf anatomy of C_3 and C_4 grasses (Poaceae): bundle sheath and mesophyll surface area relationships. *Annals of Botany* **73**: 241–255.

Deniau A, **Pieper B**, **Ten Bookum W**, **Lindhout P**, **Aarts M**, **Schat H**. **2006**. QTL analysis of cadmium and zinc accumulation in the heavy metal hyperaccumulator *Thlaspi caerulescens*.

Theoretical and Applied Genetics **113**: 907–920.

DePristo MA, **Banks E**, **Poplin R**, **Garimella KV**, **Maguire JR**, **Hartl C**, **Philippakis AA**, **del Angel G**, **Rivas MA**, **Hanna M**, **McKenna A**, **Fennell TJ**, **Kernytsky AM**, **Sivachenko AY**, **Cibulskis K**, **Gabriel SB**, **Altshuler D**, **Daly MJ**. **2011**. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**: 491–498.

Desfontaines R. **1798**. *Flora Atlantica, Sive Historia Plantarum, Quae in Atlante, Agro Tunetano et Algeriensi Crescunt*. volume 2.

Dettmer J, **Hong-Hermesdorf A**, **Stierhof YD**, **Schumacher K**. **2006**. Vacuolar H^+-ATPase activity is required for endocytic and secretory trafficking in *Arabidopsis*. *Plant Cell* **18**: 715–730.

Deuschle K, **Funck D**, **Hellmann H**, **Däschner K**, **Binder S**, **Frommer WB**. **2001**. A nuclear gene encoding mitochondrial 1-pyrroline-5-carboxylate dehydrogenase and its potential role in protection from proline toxicity. *Plant Journal* **27**: 345–356.

Deuschle K, **Funck D**, **Forlani G**, **Stransky H**, **Biehl A**, **Leister D**, **van der Graaff E**, **Kunze R**, **Frommer WB**. **2004**. The role of 1-pyrroline-5-carboxylate dehydrogenase in proline degradation. *Plant Cell* **16**: 3413–3425.

Devaiah BN, **Karthikeyan AS**, **Raghothama KG**. **2007**. WRKY75 transcription factor is a modulator of phosphate acquisition and root development in *Arabidopsis*. *Plant Physiology* **143**: 1789–1801.

Dewoody JA, **Abts KC**, **Fahey AL**, **Ji Y**, **Kimble SJ**, **Marra NJ**, **Wijayawardena BK**, **Willoughby JR**. **2013**. Of contigs and quagmires: next-generation sequencing pitfalls associated with transcriptomic studies. *Molecular Ecology Resources* **13**: 551–558.

DiDonato RJ, **Roberts LA**, **Sanderson T**, **Eisley RB**, **Walker EL**. **2004**. *Arabidopsis* Yellow Stripe-Like 2 (YSL2): a metal-regulated gene encoding a plasma membrane transporter of nicotianamine–metal complexes. *Plant Journal* **39**: 403–414.

Ding ZJ, **Yan JY**, **Xu XY**, **Li GX**, **Zheng SJ**. **2013**. WRKY46 functions as a transcriptional repressor of ALMT1, regulating aluminum-induced malate secretion in *Arabidopsis*. *Plant Journal* **76**: 825–835.

Ding ZJ, **Yan JY**, **Xu XY**, **Yu DQ**, **Li GX**, **Zhang SQ**, **Zheng SJ**. **2014**. Transcription factor WRKY46 regulates osmotic stress responses and stomatal movement independently in *Arabidopsis*. *Plant Journal* **79**: 13–27.

Dinneny JR, **Long TA**, **Wang JY**, **Jung JW**, **Mace D**, **Pointer S**, **Barron C**, **Brady SM**, **Schiefelbein J**, **Benfey PN**. **2008**. Cell identity mediates the response of *Arabidopsis* roots to abiotic stress. *Science* **320**: 942–945.

Dixon DP, **Lapthorn A**, **Edwards R**. **2002**. Plant glutathione transferases. *Genome Biology* **3**: 3004–3010.

Dixon DP, **Hawkins T**, **Hussey PJ**, **Edwards R**. **2009**. Enzyme activities and subcellular localization of members of the *Arabidopsis* glutathione transferase superfamily. *Journal of Experimental Botany* **60**: 1207–1218.

Dixon DP, **Skipsey M**, **Edwards R**. **2010**. Roles for glutathione transferases in plant secondary metabolism. *Phytochemistry* **71**: 338 – 350.

Djohan Y, **Camara C**, **Mondé A**, **Koffi G**, **Niamké G**, **Déré L**, **Tiahou G**, **Djessou P**, **Sess D**. **2009**. Interest of antioxidants in the care of the patients infected by the HIV: the experience of long term administration of *Alternanthera pungens* herb tea. *Annales de Biologie Clinique* **67**: 563–568.

Dobin A, **Davis Ca**, **Schlesinger F**, **Drenkow J**, **Zaleski C**, **Jha S**, **Batut P**, **Chaisson M**, **Gingeras TR**. **2013**. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

Dohm JC, **Minoche AE**, **Holtgräwe D**, **Capella-Gutiérrez S**, **Zakrzewski F**, **Tafer H**, **Rupp O**, **Sörensen TR**, **Stracke R**, **Reinhardt R**, **Goesmann A**, **Kraft T**, **Schulz B**, **Stadler PF**, **Schmidt**

T, **Gabaldón T**, **Lehrach H**, **Weisshaar B**, **Himmelbauer H**. **2013**. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature* **505**: 546–550.

Dong BC, **Alpert P**, **Guo W**, **Yu FH**. **2012**. Effects of fragmentation on the survival and growth of the invasive, clonal plant *Alternanthera philoxeroides*. *Biological Invasions* **14**: 1101–1110.

Dong CH, **Agarwal M**, **Zhang Y**, **Xie Q**, **Zhu JK**. **2006**. The negative regulator of plant cold responses, HOS1, is a RING E3 ligase that mediates the ubiquitination and degradation of ICE1. *Proceedings of the National Academy of Sciences* **103**: 8281–8286.

Douchkov D, **Gryczka C**, **Stephan UW**, **Hell R**, **Baumlein H**. **2005**. Ectopic expression of nicotianamine synthase genes results in improved iron accumulation and increased nickel tolerance in transgenic tobacco. *Plant, Cell & Environment* **28**: 365–374.

Drakakaki G, **van de Ven W**, **Pan S**, **Miao Y**, **Wang J**, **Keinath NF**, **Weatherly B**, **Jiang L**, **Schumacher K**, **Hicks G**, **Raikhel N**. **2011**. Isolation and proteomic analysis of the SYP61 compartment reveal its role in exocytic trafficking in *Arabidopsis*. *Cell Research* **22**: 413–424.

Dörmann P, **Benning C**. **2002**. Galactolipids rule in seed plants. *Trends in Plant Science* **7**: 112 – 118.

Duarte J, **Wall PK**, **Edger P**, **Landherr L**, **Ma H**, **Pires JC**, **Leebens-Mack J**, **dePamphilis C**. **2010**. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology* **10**: 61.

Dubois S, **Cheptou PO**, **Petit C**, **Meerts P**, **Poncelet M**, **Vekemans X**, **Lefèbvre C**, **Escarré J**. **2003**. Genetic structure and mating systems of metallicolous and nonmetallicolous populations of *Thlaspi caerulescens*. *New Phytologist* **157**: 633–641.

Dudley TR. **1964**. Synopsis of the genus *Alyssum*. *Journal of the Arnold Arboretum* **45**: 358–373.

Dudley TR. **1966**. A new Portuguese subspecies of *Alyssum serpyllifolium* Desfontaines. *Agronomia Lusitana* **28**: 71–78.

Dudley TR. **1986a**. A new nickelophilous species of *Alyssum* (Cruciferae) from Portugal, *Alyssum pintodasilvae* T.R. Dudley, sp. nov. *Feddes Repertorium* **97**: 135–138.

Dudley TR. **1986b**. A nickel hyperaccumulating species of *Alyssum* from Spain: *Alyssum malacitanum* (Rivas Goday) T.R. Dudley, comb. et stat. nov. *Feddes Repertorium* **97**: 139–142.

Dufresne F, **Stift M**, **Vergilino R**, **Mable BK**. **2014**. Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. *Molecular Ecology* **23**: 40–69.

Durrett TP, **Gassmann W**, **Rogers EE**. **2007**. The FRD3-mediated efflux of citrate into the root vasculature is necessary for efficient iron translocation. *Plant Physiology* **144**: 197–205.

Earl DA, **vonHoldt BM**. **2012**. Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conservation Genetics Resources* **4**: 359–361.

Eberbach PL, **Bowmer KH**. **1995**. Conversion of C-14 glyphosate to carbon dioxide by alligator weed. *Journal of Aquatic Plant Management* **33**: 27–29.

Ebine K, **Okatani Y**, **Uemura T**, **Goh T**, **Shoda K**, **Niihama M**, **Morita MT**, **Spitzer C**, **Otegui MS**, **Nakano A**, **Ueda T**. **2008**. A SNARE complex unique to seed plants is required for protein storage, vacuole biogenesis and seed development of *Arabidopsis thaliana*. *Plant Cell* **20**: 3006–3021.

Edgar RC. **2010**. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Edwards EJ, **Donoghue MJ**. **2013**. Is it easy to move and easy to evolve? Evolutionary accessibility and adaptation. *Journal of Experimental Botany* **64**: 4047–4052.

Edwards EJ, **Smith SA**. **2010**. Phylogenetic analyses reveal the shady history of C_4 grasses. *Proceedings of the National Academy of Sciences* **107**: 2532–2537.

Eide D, **Broderius M**, **Fett J**, **Guerinot ML**. **1996**. A novel iron-regulated metal transporter from plants identified by functional expression in yeast. *Proceedings of the National Academy of Sciences* **93**: 5624–5628.

El-Metwally S, **Hamza T**, **Zakaria M**, **Helmy M**. **2013**. Next-generation sequence assembly: four stages of data processing and computational challenges. *PLoS Computational Biology* **9**: e1003345.

Ellinger D, **Naumann M**, **Falter C**, **Zwikowics C**, **Jamrow T**, **Manisseri C**, **Somerville SC**, **Voigt CA**. **2013**. Elevated early callose deposition results in complete penetration resistance to powdery mildew in *Arabidopsis*. *Plant Physiology* **161**: 1433–1444.

Ellinger D, **Glöckner A**, **Koch J**, **Naumann M**, **Stürtz V**, **Schütt K**, **Manisseri C**, **Somerville SC**, **Voigt CA**. **2014**. Interaction of the *Arabidopsis* GTPase RabA4c with its effector PMR4 results in complete penetration resistance to powdery mildew. *Plant Cell* **26**: 3185–3200.

Ellis JR, **Burke JM**. **2007**. EST-SSRs as a resource for population genetic analyses. *Heredity* **99**: 125–132.

Elmendorf SC, **Moore KA**. **2007**. Plant competition varies with community composition in an edaphically complex landscape. *Ecology* **88**: 2640–2650.

Enns L, **Kanaoka M**, **Torii K**, **Comai L**, **Okada K**, **Cleland R**. **2005**. Two callose synthases, GSL1 and GSL5, play an essential and redundant role in plant and pollen development and in fertility. *Plant Molecular Biology* **58**: 333–349.

Escande V, **Garoux L**, **Grison C**, **Thillier Y**, **Debart F**, **Vasseur JJ**, **Boulanger C**, **Grison C**. **2014**. Ecological catalysis and phytoextraction: symbiosis for future. *Applied Catalysis B: Environmental* **146**: 279–288.

Escarré J, **Lefèbvre C**, **Gruber W**, **Leblanc M**, **Lepart J**, **Rivière Y**, **Delay B**. **2000**. Zinc and cadmium hyperaccumulation by *Thlaspi caerulescens* from metalliferous and nonmetalliferous sites in the Mediterranean area: implications for phytoremediation. *New Phytologist* **145**: 429–437.

Evanno G, **Regnaut S**, **Goudet J**. **2005**. Detecting the number of clusters of individuals using the software Structure: a simulation study. *Molecular Ecology* **14**: 2611–2620.

Evert R, **Eschrich W**, **Heyser W**. **1977**. Distribution and structure of the plasmodesmata in mesophyll and bundle-sheath cells of *Zea mays* L. *Planta* **136**: 77–89.

Ewens WJ. **1972**. The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**: 87–112.

Excoffier L, **Laval G**, **Schneider S**. **2005**. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* **1**: 47.

Excoffier L, **Hofer T**, **Foll M**. **2009**. Detecting loci under selection in a hierarchically structured population. *Heredity* **103**: 285–298.

Ezaki B, **Gardner RC**, **Ezaki Y**, **Matsumoto H**. **2000**. Expression of aluminum-induced genes in transgenic *Arabidopsis* plants can ameliorate aluminum stress and/or oxidative stress. *Plant Physiology* **122**: 657–666.

Fahnenstich H, **Scarpeci TE**, **Valle EM**, **Flügge UI**, **Maurino VG**. **2008**. Generation of hydrogen peroxide in chloroplasts of *Arabidopsis* overexpressing glycolate oxidase as an inducible system to study oxidative stress. *Plant Physiology* **148**: 719–729.

Falcone DL, **Gibson S**, **Lemieux B**, **Somerville C**. **1994**. Identification of a gene that complements an *Arabidopsis* mutant deficient in chloroplast omega-6 desaturase activity. *Plant Physiology* **106**: 1453–1459.

Falcone DL, **Ogas JP**, **Somerville CR**. **2004**. Regulation of membrane fatty acid composition by temperature in mutants of *Arabidopsis* with alterations in membrane lipid composition. *BMC Plant Biology* **4**: 17.

Falush D, **Stephens M**, **Pritchard JK**. **2003**. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.

Fan L, **Zheng S**, **Wang X**. **1997**. Antisense suppression of phospholipase D alpha retards abscisic acid- and ethylene-promoted senescence of postharvest *Arabidopsis* leaves. *Plant Cell* **9**: 2183–2196.

Farinati S, **DalCorso G**, **Bona E**, **Corbella M**, **Lampis S**, **Cecconi D**, **Polati R**, **Berta G**, **Vallini G**, **Furini A**. **2009**. Proteomic analysis of *Arabidopsis halleri* shoots in response to the heavy metals cadmium and zinc and rhizosphere microorganisms. *Proteomics* **9**: 4837–4850.

Fay JC, **Wu CI**. **2000**. Hitchhiking under positive Darwinian selection. *Genetics* **155**: 1405–1413.

Feder JL, **Egan SP**, **Nosil P**. **2012**. The genomics of speciation-with-gene-flow. *Trends in Genetics* **28**: 342–350.

Feder JL, **Flaxman SM**, **Egan SP**, **Comeault AA**, **Nosil P**. **2013**. Geographic mode of speciation and genomic divergence. *Annual Review of Ecology, Evolution, and Systematics* **44**: 73–97.

Fellows I. **2012**. Deducer: a data analysis GUI for R. *Journal of Statistical Software* **49**: 1–15.

Felsenstein J. **2005**. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author.

Fernandes AMFS, **Queirós M**. **1973**. Contribuiçao para o conhecimento citotaxonómico das spermatophyta de Portugal IX. *Boletim da Sociedade Broteriana ser. 2* **47**: 315–355.

Filatov D, **Burke S**. **2004**. DNA diversity in Hawaiian endemic plant *Schiedea globosa*. *Heredity* **92**: 452–458.

Finotello F, **DiCamillo B**. **2014**. Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. *Briefings in Functional Genomics* **14**: 130–142.

Fischer MC, **Rellstab C**, **Tedder A**, **Zoller S**, **Gugerli F**, **Shimizu KK**, **Holderegger R**, **Widmer A**. **2013**. Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Molecular Ecology* **22**: 5594–5607.

Fischer WN, **Loo DDF**, **Koch W**, **Ludewig U**, **Boorer KJ**, **Tegeder M**, **Rentsch D**, **Wright EM**, **Frommer WB**. **2002**. Low and high affinity amino acid H^+-cotransporters for cellular import of neutral and charged amino acids. *Plant Journal* **29**: 717–731.

Flood PJ, **Harbinson J**, **Aarts MG**. **2011**. Natural genetic variation in plant photosynthesis. *Trends in Plant Science* **16**: 327 – 335.

Flowers TJ, **Garcia A**, **Koyama M**, **Yeo AR**. **1997**. Breeding for salt tolerance in crop plants — the role of molecular biology. *Acta Physiologiae Plantarum* **19**: 427–433.

Flynn TA. **2013**. *Evolution of nickel hyperaccumulation in* Alyssum *L.* University of Oxford DPhil thesis.

Foll M, **Gaggiotti O**. **2008**. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**: 977–993.

Fones H, **Davis CAR**, **Rico A**, **Fang F**, **Smith JAC**, **Preston GM**. **2010**. Metal hyperaccumulation armors plants against disease. *PLoS Pathogens* **6**: e1001093.

Fonseca NA, **Rung J**, **Brazma A**, **Marioni JC**. **2012**. Tools for mapping high-throughput sequencing data. *Bioinformatics* **28**: 3169–3177.

Forde BG, **Cutler SR**, **Zaman N**, **Krysan PJ**. **2013**. Glutamate signalling via a MEKK1 kinase-dependent pathway induces changes in *Arabidopsis* root architecture. *Plant Journal* **75**: 1–10.

Foster AW, **Osman D**, **Robinson NJ**. **2014**. Metal preferences and metallation. *Journal of Biological Chemistry* **289**: 28095–28103.

Francis WR, **Christianson LM**, **Kiko R**, **Powers ML**, **Shaner NC**, **Haddock SH**. **2013**. A comparison across non-model animals suggests an optimal sequencing depth for *de novo* transcriptome assembly. *BMC Genomics* **14**: 167.

Franco-Zorrilla JM, **Martin AC**, **Solano R**, **Rubio V**, **Leyva A**, **Paz-Ares J**. **2002**. Mutations at

CRE1 impair cytokinin-induced repression of phosphate starvation responses in *Arabidopsis*. *Plant Journal* **32**: 353–360.

Freeman J, **Salt D**. **2007**. The metal tolerance profile of *Thlaspi goesingense* is mimicked in *Arabidopsis thaliana* heterologously expressing serine acetyl-transferase. *BMC Plant Biology* **7**: 63.

Freeman JL, **Persans MW**, **Nieman K**, **Albrecht C**, **Peer W**, **Pickering IJ**, **Salt DE**. **2004**. Increased glutathione biosynthesis plays a role in nickel tolerance in *Thlaspi* nickel hyperaccumulators. *Plant Cell* **16**: 2176–2191.

Freeman JL, **Garcia D**, **Kim D**, **Hopf A**, **Salt DE**. **2005**. Constitutively elevated salicylic acid signals glutathione-mediated nickel tolerance in *Thlaspi* nickel hyperaccumulators. *Plant Physiology* **137**: 1082–1091.

Freitas H, **Prasad M**, **Pratas J**. **2004**. Analysis of serpentinophytes from north–east of Portugal for trace metal accumulation – relevance to the management of mine environment. *Chemosphere* **54**: 1625–1642.

Frérot H, **Petit C**, **Lefèbvre C**, **Gruber W**, **Collin C**, **Escarré J**. **2003**. Zinc and cadmium accumulation in controlled crosses between metallicolous and nonmetallicolous populations of *Thlaspi caerulescens* (Brassicaceae). *New Phytologist* **157**: 643–648.

Frérot H, **Faucon MP**, **Willems G**, **Godé C**, **Courseaux A**, **Darracq A**, **Verbruggen N**, **Saumitou-Laprade P**. **2010**. Genetic architecture of zinc hyperaccumulation in *Arabidopsis halleri*: the essential role of QTL × environment interactions. *New Phytologist* **187**: 355–367.

Fu YX, **Li WH**. **1993**. Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.

Fu ZQ, **Yan S**, **Saleh A**, **Wang W**, **Ruble J**, **Oka N**, **Mohan R**, **Spoel SH**, **Tada Y**, **Zheng N**, **Dong X**. **2012**. NPR3 and NPR4 are receptors for the immune signal salicylic acid in plants. *Nature* **486**: 228–232.

Fujimori K, **Ohta D**. **1998**. Isolation and characterization of a histidine biosynthetic gene in *Arabidopsis* encoding a polypeptide with two separate domains for phosphoribosyl-ATP pyrophosphohydrolase and phosphoribosyl-AMP cyclohydrolase. *Plant Physiology* **118**: 275–283.

Fujita MK, **Leaché AD**, **Burbrink FT**, **McGuire JA**, **Moritz C**. **2012**. Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution* **27**: 480–488.

Furbank RT. **2011**. Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types? *Journal of Experimental Botany* **62**: 3103–3108.

Furuya T, **Matsuoka D**, **Nanmori T**. **2014**. Membrane rigidification functions upstream of the MEKK1-MKK2-MPK4 cascade during cold acclimation in *Arabidopsis thaliana*. *FEBS Letters* **588**: 2025–2030.

Gagné JM, **Houle G**. **2002**. Factors responsible for *Honckenya peploides* (Caryophyllaceae) and *Leymus mollis* (Poaceae) spatial segregation on subarctic coastal dunes. *American Journal of Botany* **89**: 479–485.

Galardi F, **Corrales I**, **Mengoni A**, **Pucci S**, **Barletti L**, **Barzanti R**, **Arnetoli M**, **Gabbrielli R**, **Gonnelli C**. **2007**. Intra-specific differences in nickel tolerance and accumulation in the Ni-hyperaccumulator *Alyssum bertolonii*. *Environmental and Experimental Botany* **60**: 377–384.

Galland N. **1990**. *Recherche sur l'origine de la flore orophile du Maroc : etude caryologique et cytogeographique*. Universite Mohammed V, Rabat PhD Thesis.

Galland N, **Love A**. **1984**. Chromosome number reports LXXXV. *Taxon* **33**: 756–760.

Galway M, **Eng R**, **Schiefelbein J**, **Wasteneys G**. **2011**. Root hair-specific disruption of cellulose and xyloglucan in *At*CSLD3 mutants, and factors affecting the post-rupture resumption of mutant root hair growth. *Planta* **233**: 985–999.

Gao J, **Sun L**, **Yang X**, **Liu JX**. **2013**. Transcriptomic analysis of cadmium stress response in the heavy metal hyperaccumulator *Sedum alfredii* Hance. *PLoS One* **8**: e64643.

Gao MJ, **Lydiate DJ**, **Li X**, **Lui H**, **Gjetvaj B**, **Hegedus DD**, **Rozwadowski K**. **2009**. Repression of seed maturation genes by a trihelix transcriptional repressor in *Arabidopsis* seedlings. *Plant Cell* **21**: 54–71.

Garber M, **Grabherr MG**, **Guttman M**, **Trapnell C**. **2011**. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature Methods* **8**: 469–477.

García MJ, **Lucena C**, **Romera FJ**, **Alcántara E**, **Pérez-Vicente R**. **2010**. Ethylene and nitric oxide involvement in the up-regulation of key genes related to iron acquisition and homeostasis in *Arabidopsis*. *Journal of Experimental Botany* **61**: 3885–3899.

García-Mata C, **Lamattina L**. **2001**. Nitric oxide induces stomatal closure and enhances the adaptive plant responses against drought stress. *Plant Physiology* **126**: 1196–1204.

Garg R, **Patel RK**, **Tyagi AK**, **Jain M**. **2011**. *De novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Research* **18**: 53–63.

Garrison E, **Marth G**. **2012**. Haplotype-based variant detection from short-read sequencing. *eprint arXiv:1207.3907* .

Gebert M, **Meschenmoser K**, **Svidová S**, **Weghuber J**, **Schweyen R**, **Eifler K**, **Lenz H**, **Weyand K**, **Knoop V**. **2009**. A root-expressed magnesium transporter of the MRS2/MGT gene family in *Arabidopsis thaliana* allows for growth in Low-Mg^{2+} environments. *Plant Cell* **21**: 4018–4030.

Geng YP, **Pan XY**, **Xu CY**, **Zhang WJ**, **Li B**, **Chen JK**. **2006**. Phenotypic plasticity of invasive *Alternanthera philoxeroides* in relation to different water availability, compared to its native congener. *Acta Oecologica* **30**: 380–385.

Ghanevati M, **Jaworski JG**. **2001**. Active-site residues of a plant membrane-bound fatty acid elongase beta-ketoacyl-CoA synthase, FAE1 KCS. *Biochimica et Biophysica Acta* **1530**: 77 – 85.

Gigolashvili T, **Yatusevich R**, **Berger B**, **Müller C**, **Flügge UI**. **2007**. The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in *Arabidopsis thaliana*. *Plant Journal* **51**: 247–261.

Gigolashvili T, **Engqvist M**, **Yatusevich R**, **Müller C**, **Flügge UI**. **2008**. HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in *Arabidopsis thaliana*. *New Phytologist* **177**: 627–642.

Glass M, **Barkwill S**, **Unda F**, **Mansfield SD**. **2015**. Endo-beta-1,4-glucanases impact plant cell wall development by influencing cellulose crystallization. *Journal of Integrative Plant Biology* **57**: 396–410.

Glawischnig E, **Hansen BG**, **Olsen CE**, **Halkier BA**. **2004**. Camalexin is synthesized from indole-3-acetaldoxime, a key branching point between primary and secondary metabolism in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 8245–8250.

Gómez-Lim MA, **Valdés-López V**, **Cruz-Hernandez A**, **Saucedo-Arias LJ**. **1993**. Isolation and characterization of a gene involved in ethylene biosynthesis from *Arabidopsis thaliana*. *Gene* **134**: 217–221.

Gómez-García I, **Stuiver M**, **Ereño J**, **Oyenarte I**, **Corral-Rodríguez MA**, **Müller D**, **Martínez-Cruz LA**. **2012**. Purification, crystallization and preliminary crystallographic analysis of the CBS-domain pair of cyclin M2 (CNNM2). *Acta Crystallographica Section F* **68**: 1198–1203.

Gong Q, **Li P**, **Ma S**, **Indu Rupassara S**, **Bohnert HJ**. **2005**. Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*. *Plant Journal* **44**: 826–839.

Gonneau M, **Desprez T**, **Guillot A**, **Vernhettes S**, **Höfte H**. **2014**. Catalytic subunit stoichiometry within the cellulose synthase complex. *Plant Physiology* **166**: 1709–1712.

Gonnelli C, **Galardi F**, **Gabbrielli R**. **2001**. Nickel and copper tolerance and toxicity in three Tuscan populations of *Silene paradoxa*. *Physiologia Plantarum* **113**: 507–514.

Goudet J. **1995**. FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity* **86**: 485–486.

Gowik U, **Engelmann S**, **Bläsing O**, **Raghavendra AS**, **Westhoff P**. **2006**. Evolution of C_4 phosphoenolpyruvate carboxylase in the genus *Alternanthera*: gene families and the enzymatic characteristics of the C_4 isozyme and its orthologues in C_3 and C_3/C_4 *Alternantheras*. *Planta* **223**: 359–368.

Gowik U, **Bräutigam A**, **Weber KL**, **Weber APM**, **Westhoff P**. **2011**. Evolution of C_4 photosynthesis in the genus *Flaveria*: how many and which genes does it take to make C_4? *Plant Cell* **23**: 2087–2105.

Grant M, **Godiard L**, **Straube E**, **Ashfield T**, **Lewald J**, **Sattler A**, **Innes R**, **Dangl J**. **1995**. Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. *Science* **269**: 843–846.

Grant M, **Brown I**, **Adams S**, **Knight M**, **Ainslie A**, **Mansfield J**. **2000**. The *RPM1* plant disease resistance gene facilitates a rapid and sustained increase in cytosolic calcium that is necessary for the oxidative burst and hypersensitive cell death. *Plant Journal* **23**: 441–450.

Griffiths H, **Weller G**, **Toy LF**, **Dennis RJ**. **2013**. You're so vein: bundle sheath physiology, phylogeny and evolution in C_3 and C_4 plants. *Plant, Cell & Environment* **36**: 249–261.

Grille S, **Zaslawski A**, **Thiele S**, **Plat J**, **Warnecke D**. **2010**. The functions of steryl glycosides come to those who wait: Recent advances in plants, fungi, bacteria and animals. *Progress in Lipid Research* **49**: 262–288.

Grondin A, **Rodrigues O**, **Verdoucq L**, **Merlot S**, **Leonhardt N**, **Maurel C**. **2015**. Aquaporins contribute to ABA-triggered stomatal closure through OST1-mediated phosphorylation. *Plant Cell* **27**: 1945–1954.

Grotz N, **Fox T**, **Connolly E**, **Park W**, **Guerinot ML**, **Eide D**. **1998**. Identification of a family of zinc transporter genes from *Arabidopsis* that respond to zinc deficiency. *Proceedings of the National Academy of Sciences* **95**: 7220–7224.

Gruenheit N, **Deusch O**, **Esser C**, **Becker M**, **Voelckel C**, **Lockhart PJ**. **2012**. Cutoffs and k-mers: implications from a transcriptome study in allopolyploid plants. *BMC Genomics* **13**: 92.

Götz S, **García-Gómez JM**, **Terol J**, **Williams TD**, **Nagaraj SH**, **Nueda MJ**, **Robles M**, **Talón M**, **Dopazo J**, **Conesa A**. **2008**. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* **36**: 3420–3435.

Gu Y, **Kaplinsky N**, **Bringmann M**, **Cobb A**, **Carroll A**, **Sampathkumar A**, **Baskin TI**, **Persson S**, **Somerville CR**. **2010**. Identification of a cellulose synthase-associated protein required for cellulose biosynthesis. *Proceedings of the National Academy of Sciences* **107**: 12866–12871.

Guan Q, **Lu X**, **Zeng H**, **Zhang Y**, **Zhu J**. **2013**. Heat stress induction of miR398 triggers a regulatory loop that is critical for thermotolerance in *Arabidopsis*. *Plant Journal* **74**: 840–851.

Guédé NZ, **N'guessan K**, **Dibié TE**, **Grellier P**. **2010**. Ethnopharmacological study of plants used to treat malaria, in traditional medicine, by Bete Populations of Issia (Côte d'Ivoire). *Journal of Pharmaceutical Sciences and Research* **2**: 216–227.

Guichoux E, **Lagache L**, **Wagner S**, **Chaumeil P**, **Léger P**, **Lepais O**, **Lepoittevin C**, **Malausa T**, **Revardel E**, **Salin F**, **Petit RJ**. **2011**. Current trends in microsatellite genotyping. *Molecular Ecology Resources* **11**: 591–611.

Gunasekera L, **Bonila J**. **2001**. Alligator weed: tasty vegetable in Australian backyards? *Journal of Aquatic Plant Management* **39**: 17–20.

Guo WJ, **Meetam M**, **Goldsbrough PB**. **2008**. Examining the specific contributions of individual *Arabidopsis* metallothioneins to copper distribution and metal tolerance. *Plant Physiology* **146**: 1697–1706.

Guo Y, **Li CI**, **Ye F**, **Shyr Y**. **2013**. Evaluation of read count based RNA-seq analysis methods.

BMC Genomics **14**: S2.

Gustin JL, **Loureiro ME**, **Kim D**, **Na G**, **Tikhonova M**, **Salt DE**. **2009**. MTP1-dependent Zn sequestration into shoot vacuoles suggests dual roles in Zn tolerance and accumulation in Zn-hyperaccumulating plants. *Plant Journal* **57**: 1116–1127.

Gutla PVK, **Boccaccio A**, **De Angeli A**, **Gambale F**, **Carpaneto A**. **2012**. Modulation of plant TPC channels by polyunsaturated fatty acids. *Journal of Experimental Botany* **63**: 6187–6197.

Haas BJ, **Papanicolaou A**, **Yassour M**, **Grabherr M**, **Blood PD**, **Bowden J**, **Couger MB**, **Eccles D**, **Li B**, **Lieber M**, **Macmanes MD**, **Ott M**, **Orvis J**, **Pochet N**, **Strozzi F**, **Weeks N**, **Westerman R**, **William T**, **Dewey CN**, **Henschel R**, **Leduc RD**, **Friedman N**, **Regev A**. **2013**. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**: 1494–1512.

Hachez C, **Laloux T**, **Reinhardt H**, **Cavez D**, **Degand H**, **Grefen C**, **De Rycke R**, **Inzé D**, **Blatt MR**, **Russinova E**, **Chaumont F**. **2014**. *Arabidopsis* SNAREs SYP61 and SYP121 coordinate the trafficking of plasma membrane aquaporin PIP2;7 to modulate the cell membrane water permeability. *Plant Cell* **26**: 3132–3147.

Hadiarto T, **Nanmori T**, **Matsuoka D**, **Iwasaki T**, **Sato Ki**, **Fukami Y**, **Azuma T**, **Yasuda T**. **2006**. Activation of *Arabidopsis* MAPK kinase kinase (*At*MEKK1) and induction of *At*MEKK1–*At*MEK1 pathway by wounding. *Planta* **223**: 708–713.

Halimaa P, **Lin YF**, **Ahonen VH**, **Blande D**, **Clemens S**, **Gyenesei A**, **Häikiö E**, **Kärenlampi SO**, **Laiho A**, **Aarts MGM**, **Pursiheimo JP**, **Schat H**, **Schmidt H**, **Tuomainen MH**, **Tervahauta AI**. **2014**. Gene expression differences between *Noccaea caerulescens* ecotypes help to identify candidate genes for metal phytoremediation. *Environmental Science & Technology* **48**: 3344–3353.

Halkier BA, **Gershenzon J**. **2006**. Biology and biochemistry of glucosinolates. *Annual Review of Plant Biology* **57**: 303–333.

Hamburger D, **Rezzonico E**, **Petétot JMC**, **Somerville C**, **Poirier Y**. **2002**. Identification and characterization of the *Arabidopsis* PHO1 gene involved in phosphate loading to the xylem. *Plant Cell* **14**: 889–902.

Hammond JP, **Bowen HC**, **White PJ**, **Mills V**, **Pyke Ka**, **Baker AJM**, **Whiting SN**, **May ST**, **Broadley MR**. **2006**. A comparison of the *Thlaspi caerulescens* and *Thlaspi arvense* shoot transcriptomes. *New Phytologist* **170**: 239–260.

Hanikenne M, **Nouet C**. **2011**. Metal hyperaccumulation and hypertolerance: a model for plant evolutionary genomics. *Current Opinion in Plant Biology* **14**: 252–259.

Hanikenne M, **Talke IN**, **Haydon MJ**, **Lanz C**, **Nolte A**, **Motte P**, **Kroymann J**, **Weigel D**, **Krämer U**. **2008**. Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of *HMA4*. *Nature* **453**: 391–395.

Hardy S, **Uetani N**, **Wong N**, **Kostantin E**, **Labbe D**, **Bégin L**, **Mes-Masson A**, **Miranda-Saavedra D**, **Tremblay M**. **2015**. The protein tyrosine phosphatase PRL-2 interacts with the magnesium transporter CNNM3 to promote oncogenesis. *Oncogene* **34**: 986–995.

Harpaz-Saad S, **McFarlane HE**, **Xu S**, **Divi UK**, **Forward B**, **Western TL**, **Kieber JJ**. **2011**. Cellulose synthesis via the FEI2 RLK/SOS5 pathway and Cellulose Synthase 5 is required for the structure of seed coat mucilage in *Arabidopsis*. *Plant Journal* **68**: 941–953.

Hartl D, **Clark A**. **2007**. *Principles of Population Genetics*. Sunderland, USA: Sinauer Associates.

Hassinen VH, **Tuomainen M**, **Peräniemi S**, **Schat H**, **Kärenlampi SO**, **Tervahauta AI**. **2009**. Metallothioneins 2 and 3 contribute to the metal-adapted phenotype but are not directly linked to Zn accumulation in the metal hyperaccumulator, *Thlaspi caerulescens*. *Journal of Experimental Botany* **60**: 187–196.

Hatch MD. **1987**. C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta* **895**: 81–106.

Hattori M, **Tanaka Y**, **Fukai S**, **Ishitani R**, **Nureki O**. **2007**. Crystal structure of the MgtE Mg^{2+} transporter. *Nature* **448**: 1072–1075.

Hauser T, **Popilka L**, **Hartl FU**, **Hayer-Hartl M**. **2015**. Role of auxiliary proteins in Rubisco biogenesis and function. *Nature Plants* **1**.

Haydon MJ, **Cobbett CS**. **2007**. Transporters of ligands for essential metal ions in plants. *New Phytologist* **174**: 499–506.

Haydon MJ, **Kawachi M**, **Wirtz M**, **Hillmer S**, **Hell R**, **Krämer U**. **2012**. Vacuolar nicotianamine has critical and distinct roles under iron deficiency and for zinc sequestration in *Arabidopsis*. *Plant Cell* **24**: 724–737.

Helariutta Y, **Fukaki H**, **Wysocka-Diller J**, **Nakajima K**, **Jung J**, **Sena G**, **Hauser MT**, **Benfey PN**. **2000**. The *SHORT-ROOT* gene controls radial patterning of the *Arabidopsis* root through radial signaling. *Cell* **101**: 555–567.

Hereford J. **2010**. Does selfing or outcrossing promote local adaptation? *American Journal of Botany* **97**: 298–302.

Hewitt GM. **1999**. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**: 87–112.

Hill JL, **Hammudi MB**, **Tien M**. **2014**. The *Arabidopsis* cellulose synthase complex: a proposed hexamer of CESA trimers in an equimolar stoichiometry. *Plant Cell* **26**: 4834–4842.

Hirayama T, **Mitsukawa N**, **Shibata D**, **Shinozaki K**. **1997**. *AtPLC2*, a gene encoding phosphoinositide-specific phospholipase C, is constitutively expressed in vegetative and floral tissues in *Arabidopsis thaliana*. *Plant Molecular Biology* **34**: 175–180.

Hirner A, **Ladwig F**, **Stransky H**, **Okumoto S**, **Keinath M**, **Harms A**, **Frommer WB**, **Koch W**. **2006**. *Arabidopsis* LHT1 is a high-affinity transporter for cellular amino acid uptake in both root epidermis and leaf mesophyll. *Plant Cell* **18**: 1931–1946.

Hirsh AE, **Fraser HB**. **2001**. Protein dispensability and rate of evolution. *Nature* **411**: 1046–1049.

Hiruma K, **Fukunaga S**, **Bednarek P**, **Piślewska-Bednarek M**, **Watanabe S**, **Narusaka Y**, **Shirasu K**, **Takano Y**. **2013**. Glutathione and tryptophan metabolism are required for *Arabidopsis* immunity during the hypersensitive response to hemibiotrophs. *Proceedings of the National Academy of Sciences* **110**: 9589–9594.

Homer FA, **Reeves RD**, **Brooks RR**, **Baker AJ**. **1991**. Characterization of the nickel-rich extract from the nickel hyperaccumulator *Dichapetalum gelonioides*. *Phytochemistry* **30**: 2141–2145.

Hu Y, **Dong Q**, **Yu D**. **2012a**. *Arabidopsis* WRKY46 coordinates with WRKY70 and WRKY53 in basal resistance against pathogen *Pseudomonas syringae*. *Plant Science* **185**: 288–297.

Hu YT, **Ming F**, **Chen WW**, **Yan JY**, **Xu ZY**, **Li GX**, **Xu CY**, **Yang JL**, **Zheng SJ**. **2012b**. *Tc*OPT3, a member of oligopeptide transporters from the hyperaccumulator *Thlaspi caerulescens*, is a novel Fe/Zn/Cd/Cu transporter. *PloS One* **7**: e38535.

Huang CF, **Yamaji N**, **Mitani N**, **Yano M**, **Nagamura Y**, **Ma JF**. **2009**. A bacterial-type ABC transporter is involved in aluminum tolerance in rice. *Plant Cell* **21**: 655–667.

Huang CF, **Yamaji N**, **Ma JF**. **2010**. Knockout of a bacterial-type ATP-binding cassette transporter gene, *AtSTAR1*, results in increased aluminum sensitivity in *Arabidopsis*. *Plant Physiology* **153**: 1669–1677.

Huang J, **Lu X**, **Yan H**, **Chen S**, **Zhang W**, **Huang R**, **Zheng Y**. **2012**. Transcriptome characterization and sequencing-based identification of salt-responsive genes in *Millettia pinnata*, a semi-mangrove plant. *DNA Research* **19**: 195–207.

Huang X, **Madan A**. **1999**. CAP3: a DNA sequence assembly program. *Genome Research* **9**: 868–877.

Hudson AO, **Singh BK**, **Leustek T**, **Gilvarg C**. **2006**. An ll-diaminopimelate aminotransferase defines a novel variant of the lysine biosynthesis pathway in plants. *Plant Physiology* **140**:

292–301.

Hudson RR, **Slatkin M**, **Maddison W**. **1992**. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583–589.

Hughes R, **Bachmann K**, **Smirnoff N**, **Macnair MR**. **2001**. The role of drought tolerance in serpentine tolerance in the *Mimulus guttatus* Fischer ex DC. complex. *South African Journal of Science* **97**: 581–586.

Huh SM, **Noh EK**, **Kim HG**, **Jeon BW**, **Bae K**, **Hu HC**, **Kwak JM**, **Park OK**. **2010**. *Arabidopsis* annexins AnnAt1 and AnnAt4 interact with each other and regulate drought and salt stress responses. *Plant and Cell Physiology* **51**: 1499–1514.

Hull AK, **Vij R**, **Celenza JL**. **2000**. *Arabidopsis* cytochrome P450s that catalyze the first step of tryptophan-dependent indole-3-acetic acid biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 2379–2384.

Humphreys JM, **Hemm MR**, **Chapple C**. **1999**. New routes for lignin biosynthesis defined by biochemical characterization of recombinant ferulate 5-hydroxylase, a multifunctional cytochrome P450-dependent monooxygenase. *Proceedings of the National Academy of Sciences* **96**: 10045–10050.

Humphry M, **Bednarek P**, **Kemmerling B**, **Koh S**, **Stein M**, **Göbel U**, **Stüber K**, **Piślewska-Bednarek M**, **Loraine A**, **Schulze-Lefert P**, **Somerville S**, **Panstruga R**. **2010**. A regulon conserved in monocot and dicot plants defines a functional module in antifungal plant immunity. *Proceedings of the National Academy of Sciences* **107**: 21896–21901.

Hur YS, **Um JH**, **Kim S**, **Kim K**, **Park HJ**, **Lim JS**, **Kim WY**, **Jun SE**, **Yoon EK**, **Lim J**, **Ohme-Takagi M**, **Kim D**, **Park J**, **Kim GT**, **Cheon CI**. **2015**. *Arabidopsis thaliana* homeobox 12 (*At*HB12), a homeodomain-leucine zipper protein, regulates leaf growth by promoting cell expansion and endoreduplication. *New Phytologist* **205**: 316–328.

Huson DH, **Bryant D**. **2006**. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**: 254–267.

Hussain D, **Haydon MJ**, **Wang Y**, **Wong E**, **Sherson SM**, **Young J**, **Camakaris J**, **Harper JF**, **Cobbett CS**. **2004**. P-Type ATPase heavy metal transporters with roles in essential zinc homeostasis in *Arabidopsis*. *Plant Cell* **16**: 1327–1339.

Hwang IH, **Lim JH**, **Pih KT**. **2002**. Osmotic stress-inducible protein functioning as a negative regulator in osmotic stress signaling pathway of plants. US Patent App. 10/239,919.

Ikeda M, **Mitsuda N**, **Ohme-Takagi M**. **2011**. *Arabidopsis* HsfB1 and HsfB2b act as repressors of the expression of heat-inducible Hsfs but positively regulate the acquired thermotolerance. *Plant Physiology* **157**: 1243–1254.

Ilie L, **Molnar M**. **2013**. RACER: Rapid and accurate correction of errors in reads. *Bioinformatics* **29**: 2490–2493.

Ingle RA, **Mugford ST**, **Rees JD**, **Campbell MM**, **Smith JAC**. **2005a**. Constitutively high expression of the histidine biosynthetic pathway contributes to nickel tolerance in hyperaccumulator plants. *Plant Cell* **17**: 2089–2106.

Ingle RA, **Smith JAC**, **Sweetlove LJ**. **2005b**. Responses to nickel in the proteome of the hyperaccumulator plant *Alyssum lesbiacum*. *Biometals* **18**: 627–641.

Irving HMNH, **Williams RJP**. **1953**. The stability of transition-metal complexes. *Journal of the Chemical Society* **637**: 3192–3210.

Ishitani M, **Xiong L**, **Lee H**, **Stevenson B**, **Zhu JK**. **1998**. *HOS1*, a genetic locus involved in cold-responsive gene expression in *Arabidopsis*. *Plant Cell* **10**: 1151–1161.

Jaffré T, **Brooks R**, **Lee J**, **Reeves R**. **1976**. *Sebertia acuminata*: a hyperaccumulator of nickel from New Caledonia. *Science* **193**: 579–580.

Jain A, **Sinilal B**, **Dhandapani G**, **Meagher RB**, **Sahi SV**. **2013**. Effects of deficiency and excess of

zinc on morphophysiological traits and spatiotemporal regulation of zinc-responsive genes reveal incidence of cross talk between micro- and macronutrients. *Environmental Science & Technology* **47**: 5327–5335.

Jakobsson M, **Rosenberg NA**. **2007**. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**: 1801–1806.

Jalas J, **Suominen J**, **Lampinen R**. **1996**. *Atlas Florae Europaeae. Distribution of vascular plants in Europe vol. 11 (Cruciferae: Ricotia to Raphanus)*. Helsinki, Finland: Helsinki University Printing House.

Jamai A, **Salomé PA**, **Schilling SH**, **Weber AP**, **McClung CR**. **2009**. *Arabidopsis* photorespiratory serine hydroxymethyltransferase activity requires the mitochondrial accumulation of ferredoxin-dependent glutamate synthase. *Plant Cell* **21**: 595–606.

Jensen MK, **Vogt JK**, **Bressendorff S**, **Seguin-Orlando A**, **Petersen M**, **Sicheritz-Pontén T**, **Mundy J**. **2015**. Transcriptome and genome size analysis of the venus flytrap. *PLoS One* **10**: e0123887.

Jhanwar S, **Priya P**, **Garg R**, **Parida SK**, **Tyagi AK**, **Jain M**. **2012**. Transcriptome sequencing of wild chickpea as a rich resource for marker development. *Plant Biotechnology Journal* **10**: 690–702.

Jiménez-Ambriz G, **Petit C**, **Bourrié I**, **Dubois S**, **Olivieri I**, **Ronce O**. **2007**. Life history variation in the heavy metal tolerant plant *Thlaspi caerulescens* growing in a network of contaminated and noncontaminated sites in southern France: role of gene flow, selection and phenotypic plasticity. *New Phytologist* **173**: 199–215.

Johannesson K, **Panova M**, **Kemppainen P**, **André C**, **Rolán-Alvarez E**, **Butlin RK**. **2010**. Repeated evolution of reproductive isolation in a marine snail: unveiling mechanisms of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 1735–1747.

Jones FC, **Grabherr MG**, **Chan YF**, **Russell P**, **Mauceli E**, **Johnson J**, **Swofford R**, **Pirun M**, **Zody MC**, **White S**, **Birney E**, **Searle S**, **Schmutz J**, **Grimwood J**, **Dickson MC**, **Myers RM**, **Miller CT**, **Summers BR**, **Knecht AK**, **Brady SD**, **Zhang H**, **Pollen AA**, **Howes T**, **Amemiya C**, **Lander ES**, **Di Palma F**, **Lindblad-Toh K**, **Kingsley DM**, **Platf BIGS**, **Team WGA**. **2012**. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**: 55–61.

Jones GR. **2015**. STACEY: species delimitation and phylogeny estimation under the multispecies coalescent. *bioRxiv* .

Jones GR, **Aydin Z**, **Oxelman B**. **2015**. DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics* **31**: 991–998.

Jones P, **Binns D**, **Chang HY**, **Fraser M**, **Li W**, **McAnulla C**, **McWilliam H**, **Maslen J**, **Mitchell A**, **Nuka G**, **Pesseat S**, **Quinn AF**, **Sangrador-Vegas A**, **Scheremetjew M**, **Yong SY**, **Lopez R**, **Hunter S**. **2014**. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240.

Joshi V, **Laubengayer KM**, **Schauer N**, **Fernie AR**, **Jander G**. **2006**. Two *Arabidopsis* threonine aldolases are nonredundant and compete with threonine deaminase for a common substrate pool. *Plant Cell* **18**: 3564–3575.

Jozefczak M, **Remans T**, **Vangronsveld J**, **Cuypers A**. **2012**. Glutathione is a key player in metal-induced oxidative stress defenses. *International Journal of Molecular Sciences* **13**: 3145–3175.

Julien M, **Bourne A**, **Low V**. **1992**. Growth of the weed *Alternanthera philoxeroides* (Martius) Grisebach (alligator weed) in aquatic and terrestrial habitats in Australia. *Plant Protection Quarterly (Australia)* **7**: 102–108.

Julien MH, **Skarratt B**, **Maywald G**. **1995**. Potential geographical distribution of alligator weed and its biological control by *Agasicles hygrophila*. *Journal of Aquatic Plant Management* **33**:

55–60.

Jung JH, **Park JH**, **Lee S**, **To TK**, **Kim JM**, **Seki M**, **Park CM**. **2013**. The cold signaling attenuator *HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENE 1* activates *FLOWERING LOCUS C* transcription via chromatin remodeling under short-term cold stress in *Arabidopsis*. *Plant Cell* **25**: 4378–4390.

Kachroo A, **Lapchyk L**, **Fukushige H**, **Hildebrand D**, **Klessig D**, **Kachroo P**. **2003**. Plastidial fatty acid signaling modulates salicylic acid– and jasmonic acid–mediated defense pathways in the *Arabidopsis ssi2* mutant. *Plant Cell* **15**: 2952–2965.

Kadereit G, **Ackerly D**, **Pirie MD**. **2012**. A broader model for C_4 photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae s.s.). *Proceedings of the Royal Society of London B: Biological Sciences* **279**: 3304–3311.

Kaeuffer R, **Peichel CL**, **Bolnick DI**, **Hendry AP**. **2012**. Parallel and nonparallel aspects of ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. *Evolution* **66**: 402–418.

Kanehisa M, **Goto S**, **Sato Y**, **Kawashima M**, **Furumichi M**, **Tanabe M**. **2014**. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research* **42**: D199–D205.

Kapralov MV, **Stift M**, **Filatov DA**. **2009**. Evolution of genome size in Hawaiian endemic genus *Schiedea* (Caryophyllaceae). *Tropical Plant Biology* **2**: 77–83.

Kashem MA, **Singh BR**, **Kubota H**, **Sugawara R**, **Kitajima N**, **Kondo T**, **Kawai S**. **2010**. Zinc tolerance and uptake by *Arabidopsis halleri ssp. gemmifera* grown in nutrient solution. *Environmental Science and Pollution Research* **17**: 1174–1176.

Kass RE, **Raftery AE**. **1995**. Bayes factors. *Journal of the American Statistical Association* **90**: 773–795.

Kato Y, **Miwa K**, **Takano J**, **Wada M**, **Fujiwara T**. **2009**. Highly boron deficiency-tolerant plants generated by enhanced expression of NIP5;1, a boric acid channel. *Plant and Cell Physiology* **50**: 58–66.

Kaur S, **Cogan NOI**, **Pembleton LW**, **Shinozuka M**, **Savin KW**, **Materne M**, **Forster JW**. **2011**. Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. *BMC Genomics* **12**: 265.

Kaur S, **Pembleton L**, **Cogan N**, **Savin K**, **Leonforte T**, **Paull J**, **Materne M**, **Forster J**. **2012**. Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. *BMC Genomics* **13**: 104.

Kazakou E, **Dimitrakopoulos PG**, **Baker aJM**, **Reeves RD**, **Troumbis aY**. **2008**. Hypotheses, mechanisms and trade-offs of tolerance and adaptation to serpentine soils: from species to ecosystem level. *Biological Reviews of the Cambridge Philosophical Society* **83**: 495–508.

Kendziorek M, **Paszkowski A**. **2008**. Properties of serine:glyoxylate aminotransferase purified from *Arabidopsis thaliana* leaves. *Acta Biochimica et Biophysica Sinica* **40**: 102–110.

Kerkeb L, **Krämer U**. **2003**. The role of free histidine in xylem loading of nickel in *Alyssum lesbiacum* and *Brassica juncea*. *Plant Physiology* **131**: 716–724.

Keshavarzi M, **Abassian S**, **Sheidai M**. **2012**. Pollen morphology of the genus *Clypeola* (Brassicaceae) in Iran. *Phytologia Balcanica* **18**: 17–24.

Kidd P, **Monterroso C**. **2005**. Metal extraction by *Alyssum serpyllifolium* ssp. *lusitanicum* on mine-spoil soils from Spain. *Science of the Total Environment* **336**: 1–11.

Kim BG, **Waadt R**, **Cheong YH**, **Pandey GK**, **Dominguez-Solis JR**, **Schültke S**, **Lee SC**, **Kudla J**, **Luan S**. **2007**. The calcium sensor CBL10 mediates salt tolerance by regulating ion homeostasis in *Arabidopsis*. *Plant Journal* **52**: 473–484.

Kim KC, **Fan B**, **Chen Z**. **2006**. Pathogen-induced *Arabidopsis* WRKY7 is a transcriptional repressor and enhances plant susceptibility to *Pseudomonas syringae*. *Plant Physiology* **142**:

1180–1192.

Kim S, **Takahashi M**, **Higuchi K**, **Tsunoda K**, **Nakanishi H**, **Yoshimura E**, **Mori S**, **Nishizawa NK**. **2005**. Increased nicotianamine biosynthesis confers enhanced tolerance of high levels of metals, in particular nickel, to plants. *Plant and Cell Physiology* **46**: 1809–1818.

Kinoshita H, **Nagasaki J**, **Yoshikawa N**, **Yamamoto A**, **Takito S**, **Kawasaki M**, **Sugiyama T**, **Miyake H**, **Weber AP**, **Taniguchi M**. **2011**. The chloroplastic 2-oxoglutarate/malate transporter has dual function as the malate valve and in carbon/nitrogen metabolism. *Plant Journal* **65**: 15–26.

Kirchner S. **2009**. *The* At*IREGs - Characterization of a new family of metal transporters in* Arabidopsis thaliana. University of Hohenheim PhD thesis.

Klatte M, **Schuler M**, **Wirtz M**, **Fink-Straube C**, **Hell R**, **Bauer P**. **2009**. The analysis of *Arabidopsis* nicotianamine synthase mutants reveals functions for nicotianamine in seed iron loading and iron deficiency responses. *Plant Physiology* **150**: 257–271.

Klepek YS, **Geiger D**, **Stadler R**, **Klebl F**, **Landouar-Arsivaud L**, **Lemoine R**, **Hedrich R**, **Sauer N**. **2005**. *Arabidopsis* POLYOL TRANSPORTER 5, a new member of the monosaccharide transporter-like superfamily, mediates H^+ -symport of numerous substrates, including myo-inositol, glycerol, and ribose. *Plant Cell* **17**: 204–218.

Klepek YS, **Volke M**, **Konrad KR**, **Wippel K**, **Hoth S**, **Hedrich R**, **Sauer N**. **2010**. *Arabidopsis thaliana* POLYOL/MONOSACCHARIDE TRANSPORTERS 1 and 2: fructose and xylitol/H^+ symporters in pollen and young xylem cells. *Journal of Experimental Botany* **61**: 537–550.

Kobayashi K, **Masuda T**, **Takamiya K**, **Ohta H**. **2006**. Membrane lipid alteration during phosphate starvation is regulated by phosphate signaling and auxin/cytokinin cross-talk. *Plant Journal* **47**: 238–248.

Kobayashi K, **Awai K**, **Nakamura M**, **Nagatani A**, **Masuda T**, **Ohta H**. **2009**. Type-B monogalactosyldiacylglycerol synthases are involved in phosphate starvation-induced lipid remodeling, and are crucial for low-phosphate adaptation. *Plant Journal* **57**: 322–331.

Kofler R, **Schlötterer C**, **Lelley T**. **2007**. SciRoKo: a new tool for whole genome microsatellite search and investigation. *Bioinformatics* **23**: 1683–1685.

Koizumi K, **Gallagher KL**. **2013**. Identification of SHRUBBY, a SHORT-ROOT and SCARECROW interacting protein that controls root growth and radial patterning. *Development* **140**: 1292–1300.

Kolář F, **Fér T**, **Štech M**, **Trávníček P**, **Dušková E**, **Schönswetter P**, **Suda J**. **2012**. Bringing together evolution on serpentine and polyploidy: spatiotemporal history of the diploid-tetraploid complex of *Knautia arvensis* (Dipsacaceae). *PLoS One* **7**: e39988.

Kolář F, **Dortová M**, **Lepš J**, **Pouzar M**, **Krejčová A**, **Štech M**. **2013**. Serpentine ecotypic differentiation in a polyploid plant complex: shared tolerance to Mg and Ni stress among di- and tetraploid serpentine populations of *Knautia arvensis* (Dipsacaceae). *Plant and Soil* **374**: 435–447.

Kopylova E, **Noé L**, **Touzet H**. **2012**. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**: 3211–3217.

Korshunova Y, **Eide D**, **Gregg Clark W**, **Lou Guerinot M**, **Pakrasi H**. **1999**. The IRT1 protein from *Arabidopsis thaliana* is a metal transporter with a broad substrate range. *Plant Molecular Biology* **40**: 37–44.

Kovermann P, **Meyer S**, **Hörtensteiner S**, **Picco C**, **Scholz-Starke J**, **Ravera S**, **Lee Y**, **Martinoia E**. **2007**. The *Arabidopsis* vacuolar malate channel is a member of the ALMT family. *Plant Journal* **52**: 1169–1180.

Krämer U. **2010**. Metal hyperaccumulation in plants. *Annual Review of Plant Biology* **61**: 517–534.

Krämer U, **Grime G**, **Smith J**, **Hawes C**, **Baker A**. **1997**. Micro-PIXE as a technique for studying nickel localization in leaves of the hyperaccumulator plant *Alyssum lesbiacum*. *Nuclear*

Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms **130**: 346–350.

Krämer U, **Talke IN**, **Hanikenne M**. **2007**. Transition metal transport. *FEBS Letters* **581**: 2263–2272.

Krinke O, **Ruelland E**, **Valentová O**, **Vergnolle C**, **Renou JP**, **Taconnat L**, **Flemr M**, **Burketová L**, **Zachowski A**. **2007**. Phosphatidylinositol 4-kinase activation is an early response to salicylic acid in *Arabidopsis* suspension cells. *Plant Physiology* **144**: 1347–1359.

Krämer U, **Cotter-Howells JD**, **Charnock JM**, **Baker AJM**, **Smith JAC**. **1996**. Free histidine as a metal chelator in plants that accumulate nickel. *Nature* **379**: 635–638.

Kruckeberg AR. **1954**. The ecology of serpentine soils. 3. Plant species in relation to serpentine soils. *Ecology* **35**: 267–274.

Kruckeberg AR. **2004**. *Geology and Plant Life: The Effects of Landforms and Rock Types on Plants*. University of Washington Press.

Külahoglu C, **Denton AK**, **Sommer M**, **Maß J**, **Schliesky S**, **Wrobel TJ**, **Berckmans B**, **Gongora-Castillo E**, **Buell CR**, **Simon R**, **De Veylder S**, **Bräutigam A**, **P WA**. **2014**. Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae C_3 and C_4 plant species. *Plant Cell* **26**: 3243–3260.

Küpfer P. **1974**. Recherches sur les liens de parenté entre la flore orophile des Alpes et celledes Pyrénées. *Boissiera* **23**: 3–322.

Küpfer P, **Nieto Feliner G**. **1993**. *Alyssum* L. *Flora Iberica* **4**: 167–184.

Küpper H, **Lombi E**, **Zhao FJ**, **McGrath SP**. **2000**. Cellular compartmentation of cadmium and zinc in relation to other elements in the hyperaccumulator *Arabidopsis halleri*. *Planta* **212**: 75–84.

Küpper H, **Lombi E**, **Zhao FJ**, **Wieshammer G**, **McGrath SP**. **2001**. Cellular compartmentation of nickel in the hyperaccumulators *Alyssum lesbiacum, Alyssum bertolonii* and *Thlaspi goesingense*. *Journal of Experimental Botany* **52**: 2291–2300.

Küpper H, **Mijovilovich A**, **Meyer-Klaucke W**, **Kroneck PM**. **2004**. Tissue-and age-dependent differences in the complexation of cadmium and zinc in the cadmium/zinc hyperaccumulator *Thlaspi caerulescens* (Ganges ecotype) revealed by X-ray absorption spectroscopy. *Plant Physiology* **134**: 748–757.

Kurek I, **Chang TK**, **Bertain SM**, **Madrigal A**, **Liu L**, **Lassner MW**, **Zhu G**. **2007**. Enhanced thermostability of *Arabidopsis* Rubisco activase improves photosynthesis and growth rates under moderate heat stress. *Plant Cell* **19**: 3230–3241.

Kutschera VE, **Bidon T**, **Hailer F**, **Rodi JL**, **Fain SR**, **Janke A**. **2014**. Bears in a forest of gene trees: phylogenetic inference is complicated by incomplete lineage sorting and gene flow. *Molecular Biology and Evolution* **31**: 2004–2017.

La Rota M, **Kantety RV**, **Yu JK**, **Sorrells ME**. **2005**. Non-random distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. *BMC Genomics* **6**: 23.

Lam P, **Zhao L**, **McFarlane HE**, **Aiga M**, **Lam V**, **Hooker TS**, **Kunst L**. **2012**. RDR1 and SGS3, components of RNA-mediated gene silencing, are required for the regulation of cuticular wax biosynthesis in developing inflorescence stems of *Arabidopsis*. *Plant Physiology* **159**: 1385–1395.

Lamesch P, **Berardini TZ**, **Li D**, **Swarbreck D**, **Wilks C**, **Sasidharan R**, **Muller R**, **Dreher K**, **Alexander DL**, **Garcia-Hernandez M**, **Karthikeyan AS**, **Lee CH**, **Nelson WD**, **Ploetz L**, **Singh S**, **Wensel A**, **Huala E**. **2011**. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* **40**: D1202–D1210.

Lander ES, **Schork NJ**. **1994**. Genetic dissection of complex traits. *Science* **265**: 2037–2048.

Landrein B, **Lathe R**, **Bringmann M**, **Vouillot C**, **Ivakov A**, **Boudaoud A**, **Persson S**, **Hamant O**. **2013**. Impaired cellulose synthase guidance leads to stem torsion and twists phyllotactic patterns in *Arabidopsis*. *Current Biology* **23**: 895 – 900.

Lanfear R, **Calcott B**, **Kainer D**, **Mayer C**, **Stamatakis A**. **2014**. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology* **14**: 82.

Langdale JA. **2011**. C_4 cycles: past, present, and future research on C_4 photosynthesis. *Plant Cell* **23**: 3879–3892.

Lanquar V, **Lelièvre F**, **Bolte S**, **Hamès C**, **Alcon C**, **Neumann D**, **Vansuyt G**, **Curie C**, **Schröder A**, **Krämer U**, *et al.*. **2005**. Mobilization of vacuolar iron by *At*NRAMP3 and *At*NRAMP4 is essential for seed germination on low iron. *The EMBO Journal* **24**: 4041–4051.

Lanquar V, **Ramos MS**, **Lelièvre F**, **Barbier-Brygoo H**, **Krieger-Liszkay A**, **Krämer U**, **Thomine S**. **2010**. Export of vacuolar manganese by *At*NRAMP3 and *At*NRAMP4 is required for optimal photosynthesis and growth under manganese deficiency. *Plant Physiology* **152**: 1986–1999.

Lapis-Gaza HR, **Jost R**, **Finnegan PM**. **2014**. *Arabidopsis PHOSPHATE TRANSPORTER1* genes *PHT1;8* and *PHT1;9* are involved in root-to-shoot translocation of orthophosphate. *BMC Plant Biology* **14**: 334.

Larkin M, **Blackshields G**, **Brown N**, **Chenna R**, **McGettigan P**, **McWilliam H**, **Valentin F**, **Wallace I**, **Wilm A**, **Lopez R**, **Thompson J**, **Gibson T**, **Higgins D**. **2007**. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.

Lasat MM, **Baker AJ**, **Kochian LSV**. **1998**. Altered Zn compartmentation in the root symplasm and stimulated Zn absorption into the leaf as mechanisms involved in Zn hyperaccumulation in *Thlaspi caerulescens*. *Plant Physiology* **118**: 875–883.

Lasat MM, **Pence NS**, **Garvin DF**, **Ebbs SD**, **Kochian LSV**. **2000**. Molecular physiology of zinc transport in the Zn hyperaccumulator *Thlaspi caerulescens*. *Journal of Experimental Botany* **51**: 71–79.

Law DR, **Crafts-Brandner SJ**, **Salvucci ME**. **2001**. Heat stress induces the synthesis of a new form of ribulose-1, 5-bisphosphate carboxylase/oxygenase activase in cotton leaves. *Planta* **214**: 117–125.

Law RD, **Crafts-Brandner SJ**. **2001**. High temperature stress increases the expression of wheat leaf ribulose-1, 5-bisphosphate carboxylase/oxygenase activase protein. *Archives of Biochemistry and Biophysics* **386**: 261–267.

Lawson MJ, **Zhang L**. **2006**. Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome Biology* **7**: R14.

Lazaro A, **Valverde F**, **Piñeiro M**, **Jarillo JA**. **2012**. The *Arabidopsis* E3 ubiquitin ligase HOS1 negatively regulates CONSTANS abundance in the photoperiodic control of flowering. *Plant Cell* **24**: 982–999.

Lazaro JD, **Kidd P**, **Martinez CM**. **2006**. A phytogeochemical study of the Trás-os-Montes region (NE Portugal): Possible species for plant-based soil remediation technologies. *Science of the Total Environment* **354**: 265–277.

Leaché AD, **Fujita MK**, **Minin VN**, **Bouckaert RR**. **2014**. Species delimitation using genome-wide SNP data. *Systematic Biology* **63**: 534–542.

Lee H, **Xiong L**, **Gong Z**, **Ishitani M**, **Stevenson B**, **Zhu JK**. **2001**. The *Arabidopsis HOS1* gene negatively regulates cold signal transduction and encodes a RING finger protein that displays cold-regulated nucleo–cytoplasmic partitioning. *Genes & Development* **15**: 912–924.

Lee J, **Reeves RD**, **Brooks RR**, **Jaffré T**. **1978**. The relation between nickel and citric acid in some nickel-accumulating plants. *Phytochemistry* **17**: 1033–1035.

Lee S, **Jeong H**, **Kim S**, **Lee J**, **Guerinot M**, **An G**. **2010**. *Os*ZIP5 is a plasma membrane zinc transporter in rice. *Plant Molecular Biology* **73**: 507–517.

Lee S, **Seo PJ**, **Lee HJ**, **Park CM**. **2012**. A NAC transcription factor NTL4 promotes reactive oxygen species production during drought-induced leaf senescence in *Arabidopsis*. *Plant Journal*

70: 831–844.

Lefèbvre C. 1976. Breeding system and population structure of *Armeria maritima* (Mill.) Willd. on a zinc-lead mine. *New Phytologist* **77**: 187–192.

Leffler EM, **Bullaughey K**, **Matute DR**, **Meyer WK**, **Ségurel L**, **Venkat A**, **Andolfatto P**, **Przeworski M. 2012**. Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biology* **10**: e1001388.

Leimu R, **Fischer M. 2008**. A meta-analysis of local adaptation in plants. *PLoS One* **3**: e4010.

Lemmon EM, **Lemmon AR. 2013**. High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* **44**: 99–121.

Levin HR. 2006. *The Earth Through Time*. New York, USA: Wiley.

Lewis JD, **Lazarowitz SG. 2010**. *Arabidopsis* synaptotagmin SYTA regulates endocytosis and virus movement protein cell-to-cell transport. *Proceedings of the National Academy of Sciences* **107**: 2491–2496.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv e-prints* .

Li H, **Handsaker B**, **Wysoker A**, **Fennell T**, **Ruan J**, **Homer N**, **Marth G**, **Abecasis G**, **Durbin R. 2009a**. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Li HT, **Liu H**, **Gao XS**, **Zhang H. 2009b**. Knock-out of *Arabidopsis AtNHX4* gene enhances tolerance to salt stress. *Biochemical and Biophysical Research Communications* **382**: 637–641.

Li J, **Brader G**, **Palva ET. 2008a**. Kunitz trypsin inhibitor: an antagonist of cell death triggered by phytopathogens and fumonisin B1 in *Arabidopsis*. *Molecular Plant* **1**: 482–495.

Li J, **Li H**, **Jakobsson M**, **Li S**, **Sjödin P**, **Lascoux M. 2012a**. Joint analysis of demography and selection in population genetics: where do we stand and where could we go? *Molecular Ecology* **21**: 28–44.

Li J, **Besseau S**, **Törönen P**, **Sipari N**, **Kollist H**, **Holm L**, **Tapio E. 2013**. Defense-related transcription factors WRKY70 and WRKY54 modulate osmotic stress tolerance by regulating stomatal aperture in *Arabidopsis*. *New Phytologist* **200**: 457–472.

Li J, **Liu B**, **Cheng F**, **Wang X**, **Aarts MGM**, **Wu J. 2014a**. Expression profiling reveals functionally redundant multiple-copy genes related to zinc, iron and cadmium responses in *Brassica rapa*. *New Phytologist* **203**: 182–194.

Li S, **Blanchoin L**, **Yang Z**, **Lord EM. 2003a**. The putative *Arabidopsis* Arp2/3 complex controls leaf cell morphogenesis. *Plant Physiology* **132**: 2034–2044.

Li T, **Tao Q**, **Shohag M**, **Yang X**, **Sparks D**, **Liang Y. 2015a**. Root cell wall polysaccharides are involved in cadmium hyperaccumulation in *Sedum alfredii*. *Plant and Soil* **389**: 387–399.

Li X, **Chanroj S**, **Wu Z**, **Romanowsky SM**, **Harper JF**, **Sze H. 2008b**. A distinct endosomal Ca^{2+}/Mn^{2+} pump affects root growth through the secretory process. *Plant Physiology* **147**: 1675–1689.

Li X, **Li JH**, **Wang W**, **Chen NZ**, **Ma TS**, **Xi YN**, **Zhang XL**, **Lin HF**, **Bai Y**, **Huang SJ**, **Chen YL. 2014b**. ARP2/3 complex-mediated actin dynamics is required for hydrogen peroxide-induced stomatal closure in *Arabidopsis*. *Plant, Cell & Environment* **37**: 1548–1560.

Li Y, **Feng Y**, **Lv G**, **Liu B**, **Qi A. 2015b**. The phylogeny of *Alyssum* (Brassicaceae) inferred from molecular data. *Nordic Journal of Botany* **33**: 715–721.

Li YM, **Chaney R**, **Brewer E**, **Roseberg R**, **Angle JS**, **Baker A**, **Reeves R**, **Nelkin J. 2003b**. Development of a technology for commercial phytoextraction of nickel: economic and technical considerations. *Plant and Soil* **249**: 107–115.

Li Z, **Chen Y**, **Mu D**, **Yuan J**, **Shi Y**, **Zhang H**, **Gan J**, **Li N**, **Hu X**, **Liu B**, **Yang B**, **Fan W. 2012b**. Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. *Briefings in Functional Genomics* **11**: 25–37.

Librado P, **Rozas J**. **2009**. DnaSP ver. 5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.

Lin H, **Yang Y**, **Quan R**, **Mendoza I**, **Wu Y**, **Du W**, **Zhao S**, **Schumaker KS**, **Pardo JM**, **Guo Y**. **2009**. Phosphorylation of SOS3-LIKE CALCIUM BINDING PROTEIN8 by SOS2 protein kinase stabilizes their protein complex and regulates salt tolerance in *Arabidopsis*. *Plant Cell* **21**: 1607–1619.

Lin YP, **Lee TY**, **Tanaka A**, **Charng YY**. **2014**. Analysis of an *Arabidopsis* heat-sensitive mutant reveals that chlorophyll synthase is involved in reutilization of chlorophyllide during chlorophyll turnover. *Plant Journal* **80**: 14–26.

Lipka V, **Dittgen J**, **Bednarek P**, **Bhat R**, **Wiermer M**, **Stein M**, **Landtag J**, **Brandt W**, **Rosahl S**, **Scheel D**, **Llorente F**, **Molina A**, **Parker J**, **Somerville S**, **Schulze-Lefert P**. **2005**. Pre- and postinvasion defenses both contribute to nonhost resistance in *Arabidopsis*. *Science* **310**: 1180–1183.

Liu LL, **Ren HM**, **Chen LQ**, **Wang Y**, **Wu WH**. **2013**. A protein kinase, CALCINEURIN B-LIKE PROTEIN-INTERACTING PROTEIN KINASE 9, interacts with CALCIUM SENSOR CALCINEURIN B-LIKE PROTEIN 3 and regulates potassium homeostasis under low-potassium stress in *Arabidopsis*. *Plant Physiology* **161**: 266–277.

Liu PP, **von Dahl CC**, **Park SW**, **Klessig DF**. **2011**. Interconnection between methyl salicylate and lipid-based long-distance signaling during the development of systemic acquired resistance in *Arabidopsis* and tobacco. *Plant Physiology* **155**: 1762–1768.

Liu TY, **Huang TK**, **Tseng CY**, **Lai YS**, **Lin SI**, **Lin WY**, **Chen JW**, **Chiou TJ**. **2012a**. PHO2-dependent degradation of PHO1 modulates phosphate homeostasis in *Arabidopsis*. *Plant Cell* **24**: 2168–2183.

Liu Y, **Schröder J**, **Schmidt B**. **2012b**. Musket: a multistage k-mer spectrum based error corrector for Illumina sequence data. *Bioinformatics* **29**: 308–315.

Llaurens V, **Castric V**, **Austerlitz F**, **Vekemans X**. **2008**. High paternal diversity in the self-incompatible herb *Arabidopsis halleri* despite clonal reproduction and spatially restricted pollen dispersal. *Molecular Ecology* **17**: 1577–1588.

Logacheva MD, **Kasianov AS**, **Vinogradov DV**, **Samigullin TH**, **Gelfand MS**, **Makeev VJ**, **Penin AA**. **2011**. *De novo* sequencing and characterization of floral transcriptome in two species of buckwheat *(Fagopyrum)*. *BMC Genomics* **12**: 30.

Long S, **Humphries S**, **Falkowski PG**. **1994**. Photoinhibition of photosynthesis in nature. *Annual Review of Plant Biology* **45**: 633–662.

Long TA, **Tsukagoshi H**, **Busch W**, **Lahner B**, **Salt DE**, **Benfey PN**. **2010**. The bHLH transcription factor POPEYE regulates response to iron deficiency in *Arabidopsis* roots. *Plant Cell* **22**: 2219–2236.

López-Maury L, **Marguerat S**, **Bähler J**. **2008**. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature Reviews Genetics* **9**: 583–593.

Loqué D, **Ludewig U**, **Yuan L**, **von Wirén N**. **2005**. Tonoplast intrinsic proteins *At*TIP2;1 and *At*TIP2;3 facilitate NH_3 transport into the vacuole. *Plant Physiology* **137**: 671–680.

Lorence A, **Chevone BI**, **Mendes P**, **Nessler CL**. **2004**. *myo*-Inositol oxygenase offers a possible entry point into plant ascorbate biosynthesis. *Plant Physiology* **134**: 1200–1205.

Lotterhos KE, **Whitlock MC**. **2014**. Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology* **23**: 2178–2192.

Love MI, **Huber W**, **Anders S**. **2014**. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550.

Lundgren MR, **Besnard G**, **Ripley BS**, **Lehmann CER**, **Chatelet DS**, **Kynast RG**, **Namaganda M**, **Vorontsova MS**, **Hall RC**, **Elia J**, **Osborne CP**, **Christin PA**. **2015**. Photosynthetic

innovation broadens the niche within a single species. *Ecology Letters* **18**: 1021–1029.
Lunter G, **Goodson M**. **2011**. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research* **21**: 936–939.
Luo X, **Bai X**, **Sun X**, **Zhu D**, **Liu B**, **Ji W**, **Cai H**, **Cao L**, **Wu J**, **Hu M**, **Liu X**, **Tang L**, **Zhu Y**. **2013**. Expression of wild soybean WRKY20 in *Arabidopsis* enhances drought tolerance and regulates ABA signalling. *Journal of Experimental Botany* **64**: 2155–2169.
Lynch M, **Force A**. **2000**. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**: 459–473.
Ma JF, **Ueno D**, **Zhao FJ**, **McGrath SP**. **2005**. Subcellular localisation of Cd and Zn in the leaves of a Cd-hyperaccumulating ecotype of *Thlaspi caerulescens*. *Planta* **220**: 731–736.
Ma QH, **Xu Y**. **2008**. Characterization of a caffeic acid 3-O-methyltransferase from wheat and its function in lignin biosynthesis. *Biochimie* **90**: 515–524.
MacGregor DR, **Gould P**, **Foreman J**, **Griffiths J**, **Bird S**, **Page R**, **Stewart K**, **Steel G**, **Young J**, **Paszkiewicz K**, **Millar AJ**, **Halliday KJ**, **Hall AJ**, **Penfield S**. **2013**. HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 1 is required for circadian periodicity through the promotion of nucleo-cytoplasmic mRNA export in *Arabidopsis*. *Plant Cell* **25**: 4391–4404.
Maddison WP, **Midford PE**, **Otto SP**. **2007**. Estimating a binary character's effect on speciation and extinction. *Systematic Biology* **56**: 701–710.
Madlung A. **2013**. Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity* **110**: 99–104.
Maestri E, **Marmiroli M**, **Visioli G**, **Marmiroli N**. **2010**. Metal tolerance and hyperaccumulation: costs and trade-offs between traits and environment. *Environmental and Experimental Botany* **68**: 1–13.
Malling H. **1957**. The chromosome number of *Honckenya peploides* (L.) Ehrh., with a note on its mode of sex determination. *Hereditas* **43**: 517–524.
Mandáková T, **Singh V**, **Krämer U**, **Lysak MA**. **2015**. Genome structure of the heavy metal hyperaccumulator *Noccaea caerulescens* and its stability on metalliferous and nonmetalliferous soils. *Plant Physiology* **169**: 674–689.
Manfield IW, **Devlin PF**, **Jen CH**, **Westhead DR**, **Gilmartin PM**. **2007**. Conservation, convergence, and divergence of light-responsive, circadian-regulated, and tissue-specific expression patterns during evolution of the *Arabidopsis* GATA gene family. *Plant Physiology* **143**: 941–958.
Manosalva PM, **Park SW**, **Forouhar F**, **Tong L**, **Fry WE**, **Klessig DF**. **2010**. Methyl esterase 1 (*St*MES1) is required for systemic acquired resistance in potato. *Molecular Plant-Microbe Interactions* **23**: 1151–1163.
Mantel N. **1967**. The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**: 209–220.
Mao B, **Cheng Z**, **Lei C**, **Xu F**, **Gao S**, **Ren Y**, **Wang J**, **Zhang X**, **Wang J**, **Wu F**, **Guo X**, **Liu X**, **Wu C**, **Wang H**, **Wan J**. **2012**. WAX CRYSTAL-SPARSE LEAF 2, a rice homologue of WAX2/GL1, is involved in synthesis of leaf cuticular wax. *Planta* **235**: 39–52.
Mao DD, **Tian LF**, **Li LG**, **Chen J**, **Deng PY**, **Li DP**, **Luan S**. **2008**. *At*MGT7: an *Arabidopsis* gene encoding a low-affinity magnesium transporter. *Journal of Integrative Plant Biology* **50**: 1530–1538.
Marazzi B, **Ané C**, **Simon MF**, **Delgado-Salinas A**, **Luckow M**, **Sanderson MJ**. **2012**. Locating evolutionary precursors on a phylogenetic tree. *Evolution* **66**: 3918–3930.
Marguerat S, **Bähler J**. **2010**. RNA-seq: from technology to biology. *Cellular and Molecular Life Sciences* **67**: 569–579.
Mari S, **Gendre D**, **Pianelli K**, **Ouerdane L**, **Lobinski R**, **Briat JF**, **Lebrun M**, **Czernic P**. **2006**.

Root-to-shoot long-distance circulation of nicotianamine and nicotianamine-nickel chelates in the metal hyperaccumulator *Thlaspi caerulescens*. *Journal of Experimental Botany* **57**: 4111–4122.

Marita JM, **Ralph J**, **Hatfield RD**, **Chapple C**. **1999**. NMR characterization of lignins in *Arabidopsis* altered in the activity of ferulate 5-hydroxylase. *Proceedings of the National Academy of Sciences* **96**: 12328–12332.

Marri L, **Zaffagnini M**, **Collin V**, **Issakidis-Bourguet E**, **Lemaire SD**, **Pupillo P**, **Sparla F**, **Miginiac-Maslow M**, **Trost P**. **2009**. Prompt and easy activation by specific thioredoxins of Calvin cycle enzymes of *Arabidopsis thaliana* associated in the GAPDH/CP12/PRK supramolecular complex. *Molecular Plant* **2**: 259–269.

Marschner H. **1995**. *Mineral Nutrition of Higher Plants*. Special Publications of the Society for General Microbiology. Cambridge, USA: Academic Press.

Martell AE, **Smith RM**. **1989**. *Critical Stability Constants*. volume 6. Plenum Press.

Martin SH, **Dasmahapatra KK**, **Nadeau NJ**, **Salazar C**, **Walters JR**, **Simpson F**, **Blaxter M**, **Manica A**, **Mallet J**, **Jiggins CD**. **2013**. Genome-wide evidence for speciation with gene flow in *[*Heliconius butterflies. *Genome Research* .

Maruyama-Nakashita A, **Nakamura Y**, **Yamaya T**, **Takahashi H**. **2004**. A novel regulatory pathway of sulfate uptake in *Arabidopsis* roots: implication of CRE1/WOL/AHK4-mediated cytokinin-dependent regulation. *Plant Journal* **38**: 779–789.

Mathur J, **Spielhofer P**, **Kost B**, **Chua N**. **1999**. The actin cytoskeleton is required to elaborate and maintain spatial patterning during trichome cell morphogenesis in *Arabidopsis thaliana*. *Development* **126**: 5559–5568.

Mathur J, **Mathur N**, **Kirik V**, **Kernebeck B**, **Srinivas BP**, **Hülskamp M**. **2003**. *Arabidopsis CROOKED* encodes for the smallest subunit of the ARP2/3 complex and controls cell shape by region specific fine F-actin formation. *Development* **130**: 3137–3146.

Mayer MS, **Soltis PS**, **Soltis DE**. **1994**. The evolution of the *Streptanthus glandulosus* complex (Cruciferae): genetic divergence and gene flow in serpentine endemics. *American Journal of Botany* **81**: 1288–1299.

Mayr E. **1942**. *Systematics and the origin of species, from the viewpoint of a zoologist*. Cambridge, USA: Harvard University Press.

McDonald JH, **Kreitman M**. **1991**. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.

McKown AD, **Dengler NG**. **2007**. Key innovations in the evolution of Kranz anatomy and C_4 vein pattern in *Flaveria* (Asteraceae). *American Journal of Botany* **94**: 382–399.

Mears JA. **1977**. The nomenclature and type collections of the widespread taxa of *Alternanthera* (Amaranthaceae). *Proceedings of the Academy of Natural Sciences of Philadelphia* **129**: 1–21.

Meerts P, **Van Isacker N**. **1997**. Heavy metal tolerance and accumulation in metallicolous and non-metallicolous populations of *Thlaspi caerulescens* from continental Europe. *Plant Ecology* **133**: 221–231.

Meirmans PG. **2015**. Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology* **24**: 3223–3231.

Mendoza-Cózatl DG, **Xie Q**, **Akmakjian GZ**, **Jobe TO**, **Patel A**, **Stacey MG**, **Song L**, **Demoin DW**, **Jurisson SS**, **Stacey G**, **Schroeder JI**. **2014**. OPT3 is a component of the iron-signaling network between leaves and roots and misregulation of OPT3 leads to an over-accumulation of cadmium in seeds. *Molecular Plant* **7**: 1455–1469.

Mengoni A, **Gonnelli C**, **Brocchini E**, **Galardi F**, **Pucci S**, **Bazzicalupo M**. **2000**. Genetic diversity and heavy metal tolerance in populations of *Silene paradoxa* L. (Caryophyllaceae): A random amplified polymorphic DNA analysis. *Molecular Ecology* **9**: 1319–1324.

Mengoni A, **Barabesi C**, **Gonnelli C**, **Galardi F**, **Gabbrielli R**, **Bazzicalupo M**. **2001**. Genetic

diversity of heavy metal-tolerant populations in *Silene paradoxa* L. (Caryophyllaceae): a chloroplast microsatellite analysis. *Molecular Ecology* **10**: 1909–1916.

Mengoni A, **Baker AJM**, **Bazzicalupo M**, **Reeves RD**, **Adigüzel N**, **Chianni E**, **Galardi F**, **Gabbrielli R**, **Gonnelli C**. **2003**. Evolutionary dynamics of nickel hyperaccumulation in *Alyssum* revealed by its nrDNA analysis. *New Phytologist* **159**: 691–699.

Mengoni A, **Schat H**, **Vangronsveld J**. **2009**. Plants as extreme environments? Ni-resistant bacteria and Ni-hyperaccumulators of serpentine flora. *Plant and Soil* **331**: 5–16.

Merlot S, **Hannibal L**, **Martins S**, **Martinelli L**, **Amir H**, **Lebrun M**, **Thomine S**. **2014**. The metal transporter *Pg*IREG1 from the hyperaccumulator *Psychotria gabriellae* is a candidate gene for nickel tolerance and accumulation. *Journal of Experimental Botany* **65**: 1551–1564.

Metzgar D, **Bytof J**, **Wills C**. **2000**. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Research* **10**: 72–80.

Metzker ML. **2010**. Sequencing technologies - the next generation. *Nature Reviews Genetics* **11**: 31–46.

Meyer CL, **Vitalis R**, **Saumitou-Laprade P**, **Castric V**. **2009**. Genomic pattern of adaptive divergence in *Arabidopsis halleri*, a model species for tolerance to heavy metal. *Molecular Ecology* **18**: 2050–2062.

Meyer CL, **Kostecka AA**, **Saumitou-Laprade P**, **Créach A**, **Castric V**, **Pauwels M**, **Frérot H**. **2010**. Variability of zinc tolerance among and within populations of the pseudometallophyte species *Arabidopsis halleri* and possible role of directional selection. *New Phytologist* **185**: 130–142.

Meyer K, **Shirley AM**, **Cusumano JC**, **Bell-Lelong DA**, **Chapple C**. **1998**. Lignin monomer composition is determined by the expression of a cytochrome P450-dependent monooxygenase in *Arabidopsis*. *Proceedings of the National Academy of Sciences* **95**: 6619–6623.

Mähönen AP, **Bonke M**, **Kauppinen L**, **Riikonen M**, **Benfey PN**, **Helariutta Y**. **2000**. A novel two-component hybrid molecule regulates vascular morphogenesis of the *Arabidopsis* root. *Genes & Development* **14**: 2938–2943.

Miedes E, **Suslov D**, **Vandenbussche F**, **Kenobi K**, **Ivakov A**, **Van Der Straeten D**, **Lorences EP**, **Mellerowicz EJ**, **Verbelen JP**, **Vissenberg K**. **2013**. Xyloglucan endotransglucosylase/hydrolase (XTH) overexpression affects growth and cell wall mechanics in etiolated *Arabidopsis* hypocotyls. *Journal of Experimental Botany* **64**: 2481–2497.

Mills RF, **Francini A**, **da Rocha PSF**, **Baccarini PJ**, **Aylett M**, **Krijger GC**, **Williams LE**. **2005**. The plant P1B-type ATPase *At*HMA4 transports Zn and Cd and plays a role in detoxification of transition metals supplied at elevated levels. *FEBS Letters* **579**: 783–791.

Mills RF, **Doherty ML**, **López-Marqués RL**, **Weimar T**, **Dupree P**, **Palmgren MG**, **Pittman JK**, **Williams LE**. **2008**. ECA3, a Golgi-localized P2A-type ATPase, plays a crucial role in manganese nutrition in *Arabidopsis*. *Plant Physiology* **146**: 116–128.

Milne I, **Stephen G**, **Bayer M**, **Cock PJ**, **Pritchard L**, **Cardle L**, **Shaw PD**, **Marshall D**. **2013**. Using Tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics* **14**: 193–202.

Milner MJ, **Seamon J**, **Craft E**, **Kochian LSV**. **2013**. Transport properties of members of the ZIP family in plants and their role in Zn and Mn homeostasis. *Journal of Experimental Botany* **64**: 369–381.

Minguzzi C, **Vergnano O**. **1948**. Il contenuto di nichel nelle ceneri di *Alyssum bertolonii*. *Atti della Societá Toscana Scienze Naturali* **55**: 49–74.

Mishra MK, **Chaturvedi P**, **Singh R**, **Singh G**, **Sharma LK**, **Pandey V**, **Kumari N**, **Misra P**. **2013**. Overexpression of *WsSGTL1* gene of *Withania somnifera* enhances salt tolerance, heat tolerance and cold acclimation ability in transgenic *Arabidopsis* plants. *PLoS One* **8**: e63064.

Missihoun TD, **Schmitz J**, **Klug R**, **Kirch HH**, **Bartels D**. **2011**. Betaine aldehyde dehydrogenase genes from *Arabidopsis* with different sub-cellular localization affect stress responses. *Planta* **233**: 369–382.

Mita S, **Thuillet AC**, **Gay L**, **Ahmadi N**, **Manel S**, **Ronfort J**, **Vigouroux Y**. **2013**. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology* **22**: 1383–1399.

Mitchell-Olds T, **Willis JH**, **Goldstein DB**. **2007**. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics* **8**: 845–856.

Miura K, **Lee J**, **Miura T**, **Hasegawa PM**. **2010**. SIZ1 controls cell growth and plant development in *Arabidopsis* through salicylic acid. *Plant & Cell Physiology* **51**: 103–113.

Monteoliva M, **Rizzi Y**, **Cecchini N**, **Hajirezaei MR**, **Alvarez M**. **2014**. Context of action of proline dehydrogenase (*PDH*) in the hypersensitive response of *Arabidopsis*. *BMC Plant Biology* **14**: 21.

Moore AJ, **Merges D**, **Kadereit JW**. **2013**. The origin of the serpentine endemic *Minuartia laricifolia* subsp. *ophiolitica* by vicariance and competitive exclusion. *Molecular Ecology* **22**: 2218–2231.

Morant M, **Bak S**, **Møller BL**, **Werck-Reichhart D**. **2003**. Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. *Current Opinion in Biotechnology* **14**: 151–162.

Morrison R, **Brooks R**, **Reeves R**. **1980**. Nickel uptake by *Alyssum* species. *Plant Science Letters* **17**: 451–457.

Morrissey J, **Baxter IR**, **Lee J**, **Li L**, **Lahner B**, **Grotz N**, **Kaplan J**, **Salt DE**, **Guerinot ML**. **2009**. The ferroportin metal efflux proteins function in iron and cobalt homeostasis in *Arabidopsis*. *Plant Cell* **21**: 3326–3338.

Mortazavi A, **Williams BA**, **McCue K**, **Schaeffer L**, **Wold B**. **2008**. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**: 621–628.

Mounicou S, **Szpunar J**, **Lobinski R**. **2009**. Metallomics: the concept and methodology. *Chemical Society Reviews* **38**: 1119–1138.

Mousset M, **David P**, **Petit C**, **Pouzadoux J**, **Hatt C**, **Flaven l**, **Ronce O**, **Mignot A**. **2016**. Lower selfing rates in metallicolous populations than in non-metallicolous populations of the pseudometallophyte *Noccaea caerulescens* (Brassicaceae) in Southern France. *Annals of Botany* .

Mu J, **Tan H**, **Zheng Q**, **Fu F**, **Liang Y**, **Zhang J**, **Yang X**, **Wang T**, **Chong K**, **Wang XJ**, **Zuo J**. **2008**. LEAFY COTYLEDON 1 is a key regulator of fatty acid biosynthesis in *Arabidopsis*. *Plant Physiology* **148**: 1042–1054.

Mudge SR, **Rae AL**, **Diatloff E**, **Smith FW**. **2002**. Expression analysis suggests novel roles for members of the PHT1 family of phosphate transporters in *Arabidopsis*. *Plant Journal* **31**: 341–353.

Mueller-Cajar O, **Stotz M**, **Bracher A**. **2014**. Maintaining photosynthetic CO_2 fixation via protein remodelling: the Rubisco activases. *Photosynthesis Research* **119**: 191–201.

Nadeau N, **Ruiz M**, **Salazar P**, **Counterman B**, **Medina JA**, **Ortiz-Zuazaga H**, **Morrison A**, **McMillan WO**, **Jiggins CD**, **Papa R**. **2014**. Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Research* **24**: 1316–1333.

Nagalakshmi U, **Wang Z**, **Waern K**, **Shou C**, **Raha D**, **Gerstein M**, **Snyder M**. **2008**. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**: 1344–1349.

Nagane T, **Tanaka A**, **Tanaka R**. **2010**. Involvement of *At*NAP1 in the regulation of chlorophyll degradation in *Arabidopsis thaliana*. *Planta* **231**: 939–949.

Naidu SL, **Moose SP**, **Al-Shoaibi AK**, **Raines CA**, **Long SP**. **2003**. Cold tolerance of C_4 photosynthesis in *Miscanthus* × *giganteus*: adaptation in amounts and sequence of C_4

photosynthetic enzymes. *Plant Physiology* **132**: 1688–1697.

Nandi A, **Krothapalli K**, **Buseman CM**, **Li M**, **Welti R**, **Enyedi A**, **Shah J**. **2003**. *Arabidopsis sfd* mutants affect plastidic lipid composition and suppress dwarfing, cell death, and the enhanced disease resistance phenotypes resulting from the deficiency of a fatty acid desaturase. *Plant Cell* **15**: 2383–2398.

Naqvi S, **Rizvi S**. **2000**. Accumulation of chromium and copper in three different soils and bioaccumulation in an aquatic plant, *Alternanthera philoxeroides*. *Bulletin of Environmental Contamination and Toxicology* **65**: 55–61.

Naqvi SM, **Howell RD**, **Sholas M**. **1993**. Cadmium and lead residues in field-collected red swamp crayfish (*Procambarus clarkii*) and uptake by alligator weed, *Alternanthera philoxiroides*. *Journal of Environmental Science & Health Part B* **28**: 473–485.

Narusaka Y, **Narusaka M**, **Seki M**, **Umezawa T**, **Ishida J**, **Nakajima M**, **Enju A**, **Shinozaki K**. **2004**. Crosstalk in the responses to abiotic and biotic stresses in *Arabidopsis*: analysis of gene expression in cytochrome P450 gene superfamily by cDNA microarray. *Plant Molecular Biology* **55**: 327–342.

Ndamukong I, **Abdallat AA**, **Thurow C**, **Fode B**, **Zander M**, **Weigel R**, **Gatz C**. **2007**. SA-inducible *Arabidopsis* glutaredoxin interacts with TGA factors and suppresses JA-responsive *PDF1.2* transcription. *Plant Journal* **50**: 128–139.

Nei M. **1987**. *Molecular Evolutionary Genetics*. New York, USA: Columbia University Press.

Nei M, **Li WH**. **1979**. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences* **76**: 5269–5273.

Nei M, **Tajima F**, **Tateno Y**. **1983**. Accuracy of estimated phylogenetic trees from molecular data. II Gene frequency data. *Journal of Molecular Evolution* **19**: 153–170.

Nelson T. **2011**. *Development of leaves in C_4 plants: anatomical features that support C_4 metabolism*. New York, USA: Springer.

Ness RW, **Siol M**, **Barrett SCH**. **2011**. *De novo* sequence assembly and characterization of the floral transcriptome in cross- and self-fertilizing plants. *BMC Genomics* **12**: 298.

Ni X, **Huang Y**, **Wu L**, **Zhou R**, **Deng S**, **Wu D**, **Wang B**, **Su G**, **Tang T**, **Shi S**. **2006**. Genetic diversity of the endangered Chinese endemic herb *Primulina tabacum* (Gesneriaceae) revealed by amplified fragment length polymorphism (AFLP). *Genetica* **127**: 177–183.

Nielsen R, **Paul JS**, **Albrechtsen A**, **Song YS**. **2011**. Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics* **12**: 443–451.

Nieves-Cordones M, **Alemán F**, **Martínez V**, **Rubio F**. **2014**. K^+ uptake in plant roots. The systems involved, their regulation and parallels in other organisms. *Journal of Plant Physiology* **171**: 688–695.

Nishida S, **Tsuzuki C**, **Kato A**, **Aisu A**, **Yoshida J**, **Mizuno T**. **2011**. *At*IRT1, the primary iron uptake transporter in the root, mediates excess nickel accumulation in *Arabidopsis thaliana*. *Plant & Cell Physiology* **52**: 1433–1442.

Nishikawa M, **Hosokawa K**, **Ishiguro M**, **Minamioka H**, **Tamura K**, **Hara-Nishimura I**, **Takahashi Y**, **Shimazaki Ki**, **Imai H**. **2008**. Degradation of sphingoid long-chain base 1-phosphates (LCB-1Ps): functional characterization and expression of *At*DPL1 encoding LCB-1P lyase involved in the dehydration stress response in *Arabidopsis*. *Plant & Cell Physiology* **49**: 1758–1763.

Nookaew I, **Papini M**, **Pornputtpong N**, **Scalcinati G**, **Fagerberg L**, **Uhlén M**, **Nielsen J**. **2012**. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic Acids Research* **40**: 10084–10097.

Noret N, **Meerts P**, **Vanhaelen M**, **Dos Santos A**, **Escarré J**. **2007**. Do metal-rich plants deter

herbivores? A field test of the defence hypothesis. *Oecologia* **152**: 92–100.
Nosil P, **Funk DJ**, **Ortiz-Barrientos D**. **2009**. Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* **18**: 375–402.
Ogren WL. **1984**. Photorespiration: pathways, regulation, and modification. *Annual Review of Plant Physiology* **35**: 415–442.
Oh DH, **Hong H**, **Lee SY**, **Yun DJ**, **Bohnert HJ**, **Dassanayake M**. **2014**. Genome structures and transcriptomes signify niche adaptation for the multiple-ion-tolerant extremophyte *Schrenkiella parvula*. *Plant Physiology* **164**: 2123–2138.
Ojangu EL, **Tanner K**, **Pata P**, **Jarve K**, **Holweg C**, **Truve E**, **Paves H**. **2012**. Myosins XI-K, XI-1, and XI-2 are required for development of pavement cells, trichomes, and stigmatic papillae in *Arabidopsis*. *BMC Plant Biology* **12**: 81.
Ó Lochlainn S, **Bowen HC**, **Fray RG**, **Hammond JP**, **King GJ**, **White PJ**, **Graham NS**, **Broadley MR**. **2011**. Tandem quadruplication of *HMA4* in the zinc (Zn) and cadmium (Cd) hyperaccumulator *Noccaea caerulescens*. *PloS One* **6**: e17814.
Oomen RJFJ, **Wu J**, **Lelièvre F**, **Blanchet S**, **Richaud P**, **Barbier-Brygoo H**, **Aarts MGM**, **Thomine S**. **2009**. Functional characterization of NRAMP3 and NRAMP4 from the metal hyperaccumulator *Thlaspi caerulescens*. *New Phytologist* **181**: 637–650.
O'Rawe J, **Jiang T**, **Sun G**, **Wu Y**, **Wang W**, **Hu J**, **Bodily P**, **Tian L**, **Hakonarson H**, **Johnson WE**, **Wei Z**, **Wang K**, **Lyon GJ**. **2013**. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Medicine* **5**: 28.
Oshima Y, **Shikata M**, **Koyama T**, **Ohtsubo N**, **Mitsuda N**, **Ohme-Takagi M**. **2013**. MIXTA-like transcription factors and WAX INDUCER1/SHINE1 coordinately regulate cuticle development in *Arabidopsis* and *Torenia fournieri*. *Plant Cell* **25**: 1609–1624.
Oshlack A, **Robinson MD**, **Young MD**. **2010**. From RNA-seq reads to differential expression results. *Genome Biology* **11**: 220.
Osmond C. **1978**. Crassulacean acid metabolism: a curiosity in context. *Annual Review of Plant Physiology* **29**: 379–414.
Otterhag L, **Sommarin M**, **Pical C**. **2001**. N-terminal EF-hand-like domain is required for phosphoinositide-specific phospholipase C activity in *Arabidopsis thaliana*. *FEBS Letters* **497**: 165–170.
Palmer CM, **Guerinot ML**. **2009**. Facing the challenges of Cu, Fe and Zn homeostasis in plants. *Nature Chemical Biology* **5**: 333–340.
Pan X, **Geng Y**, **Zhang W**, **Li B**, **Chen J**. **2005**. Cover shift and morphological plasticity of invasive *Alternanthera philoxeroides* along a riparian zone in south China. *Acta Phytoecological Sinica* **30**: 835–843.
Pandey GK, **Cheong YH**, **Kim BG**, **Grant JJ**, **Li L**, **Luan S**. **2007**. CIPK9: a calcium sensor-interacting protein kinase required for low-potassium tolerance in *Arabidopsis*. *Cell Research* **17**: 411–421.
Pandey N, **Sharma CP**. **2002**. Effect of heavy metals Co^{2+}, Ni^{2+} and Cd^{2+} on growth and metabolism of cabbage. *Plant Science* **163**: 753–758.
Pandolfini T, **Gabbrielli R**, **Comparini C**. **1992**. Nickel toxicity and peroxidase activity in seedlings of *Triticum aestivum* L. *Plant, Cell & Environment* **15**: 719–725.
Pannell JR, **Fields PD**. **2014**. Evolution in subdivided plant populations: concepts, recent advances and future directions. *New Phytologist* **201**: 417–432.
Papdi C, **Pérez-Salamó I**, **Joseph MP**, **Giuntoli B**, **Bögre L**, **Koncz C**, **Szabados L**. **2015**. The low oxygen, oxidative and osmotic stress responses synergistically act through the ethylene response factor VII genes *RAP2.12, RAP2.2* and *RAP2.3*. *Plant Journal* **82**: 772–784.
Pappan K, **Zheng S**, **Wang X**. **1997**. Identification and characterization of a novel plant

phospholipase D that requires polyphosphoinositides and submicromolar calcium for activity in *Arabidopsis*. *Journal of Biological Chemistry* **272**: 7048–7054.

Park E, **Nebenführ A**. **2013**. Myosin XIK of *Arabidopsis thaliana* accumulates at the root hair tip and is required for fast root hair growth. *PLoS One* **8**: e76745.

Parra G, **Bradnam K**, **Korf I**. **2007**. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**: 1061–1067.

Patterson N, **Price AL**, **Reich D**. **2006**. Population structure and eigenanalysis. *PLoS Genetics* **2**: e190.

Pauwels M, **Saumitou-Laprade P**, **Holl AC**, **Petit D**, **Bonnin I**. **2005**. Multiple origin of metallicolous populations of the pseudometallophyte *Arabidopsis halleri* (Brassicaceae) in central Europe: the cpDNA testimony. *Molecular Ecology* **14**: 4403–4414.

Pauwels M, **Frérot H**, **Bonnin I**, **Saumitou-Laprade P**. **2006**. A broad-scale analysis of population differentiation for Zn tolerance in an emerging model species for tolerance study: *Arabidopsis halleri* (Brassicaceae). *Journal of Evolutionary Biology* **19**: 1838–1850.

Pauwels M, **Willems G**, **Roosens N**, **Frerot H**, **Saumitou-Laprade P**. **2008**. Merging methods in molecular and ecological genetics to study the adaptation of plants to anthropogenic metal-polluted sites: implications for phytoremediation. *Molecular Ecology* **17**: 108–119.

Pauwels M, **Vekemans X**, **Godé C**, **Frérot H**, **Castric V**, **Saumitou-Laprade P**. **2012**. Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance, *Arabidopsis halleri* (Brassicaceae). *New Phytologist* **193**: 916–928.

Pedersen L, **Henriksen A**. **2005**. Acyl-CoA oxidase 1 from *Arabidopsis thaliana*. Structure of a key enzyme in plant lipid metabolism. *Journal of Molecular Biology* **345**: 487–500.

Peng Z, **Staub JM**, **Serino G**, **Kwok SF**, **Kurepa J**, **Bruce BD**, **Vierstra RD**, **Wei N**, **Deng XW**. **2001**. The cellular level of PR500, a protein complex related to the 19S regulatory particle of the proteasome, is regulated in response to stresses in plants. *Molecular Biology of the Cell* **12**: 383–392.

Peremyslov VV, **Prokhnevsky AI**, **Avisar D**, **Dolja VV**. **2008**. Two class XI myosins function in organelle trafficking and root hair development in *Arabidopsis*. *Plant Physiology* **146**: 1109–1116.

Peremyslov VV, **Klocko AL**, **Fowler JE**, **Dolja VV**. **2012**. *Arabidopsis* myosin XI-K localizes to the motile endomembrane vesicles associated with F-actin. *Frontiers in Plant Science* **3**.

Peremyslov VV, **Morgun EA**, **Kurth EG**, **Makarova KS**, **Koonin EV**, **Dolja VV**. **2013**. Identification of myosin XI receptors in *Arabidopsis* defines a distinct class of transport vesicles. *Plant Cell* **25**: 3022–3038.

Peterson LR, **Trivett V**, **Baker AJ**, **Aguiar C**, **Pollard AJ**. **2003**. Spread of metals through an invertebrate food chain as influenced by a plant that hyperaccumulates nickel. *Chemoecology* **13**: 103–108.

Pfeifer B, **Wittelsburger U**, **Ramos-Onsins SE**, **Lercher MJ**. **2014**. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution* **31**: 1929–1936.

Pianelli K, **Mari S**, **Marquès L**, **Lebrun M**, **Czernic P**. **2005**. Nicotianamine over-accumulation confers resistance to nickel in *Arabidopsis thaliana*. *Transgenic Research* **14**: 739–748.

Pick T, **Jaskiewicz M**, **Peterhänsel C**, **Conrath U**. **2012**. Heat shock factor HsfB1 primes gene transcription and systemic acquired resistance in *Arabidopsis*. *Plant Physiology* **159**: 52–55.

Pickrell JK, **Pritchard JK**. **2012**. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* **8**: e1002967.

Pignocchi C, **Fletcher JM**, **Wilkinson JE**, **Barnes JD**, **Foyer CH**. **2003**. The function of ascorbate oxidase in tobacco. *Plant Physiology* **132**: 1631–1641.

Pineau C, **Loubet S**, **Lefoulon C**, **Chalies C**, **Fizames C**, **Lacombe B**, **Ferrand M**, **Loudet O**,

Berthomieu P, **Richard O**. **2012**. Natural variation at the *FRD3* MATE transporter locus reveals cross-talk between Fe homeostasis and Zn tolerance in *Arabidopsis thaliana*. *PLoS Genetics* **8**: e1003120.

Plaza S, **Tearall KL**, **Zhao FJ**, **Buchner P**, **McGrath SP**, **Hawkesford MJ**. **2007**. Expression and functional analysis of metal transporter genes in two contrasting ecotypes of the hyperaccumulator *Thlaspi caerulescens*. *Journal of Experimental Botany* **58**: 1717–1728.

Plaza S, **Weber J**, **Pajonk S**, **Thomas J**, **Talke IN**, **Schellenberg M**, **Pradervand S**, **Burla B**, **Geisler M**, **Martinoia E**, **Kraemer U**. **2015**. Wounding of *Arabidopsis halleri* leaves enhances cadmium accumulation that acts as a defense against herbivory. *BioMetals* **28**: 521–528.

Plesch G, **Störmann K**, **Torres JT**, **Walden R**, **Somssich I**. **1997**. Developmental and auxin-induced expression of the *Arabidopsis PRHA* homeobox gene. *Plant Journal* **12**: 635–647.

Pollard AJ, **Powell KD**, **Harper FA**, **Smith JAC**. **2002**. The genetic basis of metal hyperaccumulation in plants. *Critical Reviews in Plant Sciences* **21**: 539–566.

Pollard AJ, **Reeves RD**, **Baker AJ**. **2014**. Facultative hyperaccumulation of heavy metals and metalloids. *Plant Science* **217**: 8–17.

Porebski S, **Bailey LG**, **Baum BR**. **1997**. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter* **15**: 8–15.

Portis AR, **Li C**, **Wang D**, **Salvucci ME**. **2008**. Regulation of Rubisco activase and its interaction with Rubisco. *Journal of Experimental Botany* **59**: 1597–1604.

Poschenrieder C, **Tolrà R**, **Barceló J**. **2006**. Can metals defend plants against biotic stress? *Trends in Plant Science* **11**: 288–295.

Potel F, **Valadier MH**, **Ferrario-Méry S**, **Grandjean O**, **Morin H**, **Gaufichon L**, **Boutet-Mercey S**, **Lothier J**, **Rothstein SJ**, **Hirose N**, **Suzuki A**. **2009**. Assimilation of excess ammonium into amino acids and nitrogen translocation in *Arabidopsis thaliana* – roles of glutamate synthases and carbamoylphosphate synthetase in leaves. *FEBS Journal* **276**: 4061–4076.

Pottier M, **Oomen R**, **Picco C**, **Giraudat J**, **Scholz-Starke J**, **Richaud P**, **Carpaneto A**, **Thomine S**. **2015**. Identification of mutations allowing Natural Resistance Associated Macrophage Proteins (NRAMP) to discriminate against cadmium. *Plant Journal* **83**: 625–637.

Primmer CR, **Papakostas S**, **Leder EH**, **Davis MJ**, **Ragan MA**. **2013**. Annotated genes and nonannotated genomes: cross-species use of Gene Ontology in ecology and evolution research. *Molecular Ecology* **22**: 3216–3241.

Pritchard JK, **Stephens M**, **Donnelly P**. **2000**. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Proctor J. **1975**. The ecology of serpentine soils. *Advances in Ecological Research* **9**: 255–366.

Prokhnevsky AI, **Peremyslov VV**, **Dolja VV**. **2008**. Overlapping functions of the four class XI myosins in *Arabidopsis* growth, root hair elongation, and organelle motility. *Proceedings of the National Academy of Sciences* **105**: 19744–19749.

Puech S. **1963**. Introduction à une monographie d'Anduze (Gard) : étude écologique et caryosystématique de quelques taxa cévenols. *Naturalia Monspeliensia. Serie botanique* **15**: 125–129.

Puig S, **Peñarrubia L**. **2009**. Placing metal micronutrients in context: transport and distribution in plants. *Current Opinion in Plant Biology* **12**: 299–306.

Puig S, **Mira H**, **Dorcey E**, **Sancenón V**, **Andrés-Colás N**, **Garcia-Molina A**, **Burkhead JL**, **Gogolin KA**, **Abdel-Ghany SE**, **Thiele DJ**, **Ecker JR**, **Pilon M**, **Peñarrubia L**. **2007**. Higher plants possess two different types of ATX1-like copper chaperones. *Biochemical and Biophysical Research Communications* **354**: 385–390.

Qadir S, **Qureshi M**, **Javed S**, **Abdin M**. **2004**. Genotypic variation in phytoremediation potential

of *Brassica juncea* cultivars exposed to Cd stress. *Plant Science* **167**: 1171–1181.

Qi Y, **Armbruster U**, **Schmitz-Linneweber C**, **Delannoy E**, **de Longevialle AF**, **Rühle T**, **Small I**, **Jahns P**, **Leister D**. **2012**. *Arabidopsis* CSP41 proteins form multimeric complexes that bind and stabilize distinct plastid transcripts. *Journal of Experimental Botany* **63**: 1251–1270.

Qiu JL, **Jilk R**, **Marks MD**, **Szymanski DB**. **2002**. The *Arabidopsis SPIKE1* gene is required for normal cell shape control and tissue development. *Plant Cell* **14**: 101–118.

Quan R, **Lin H**, **Mendoza I**, **Zhang Y**, **Cao W**, **Yang Y**, **Shang M**, **Chen S**, **Pardo JM**, **Guo Y**. **2007**. SCABP8/CBL10, a putative calcium sensor, interacts with the protein kinase SOS2 to protect *Arabidopsis* shoots from salt stress. *Plant Cell* **19**: 1415–1431.

Quintana-Murci L, **Clark AG**. **2013**. Population genetic tools for dissecting innate immunity in humans. *Nature Reviews Immunology* **13**: 280–293.

Quintela-Sabarís C, **Vendramin GG**, **Castro-Fernández D**, **Fraga MI**. **2010**. Chloroplast microsatellites reveal that metallicolous populations of the Mediterranean shrub *Cistus ladanifer* L. have multiple origins. *Plant and Soil* **334**: 161–174.

Quintero FJ, **Ohta M**, **Shi H**, **Zhu JK**, **Pardo JM**. **2002**. Reconstitution in yeast of the *Arabidopsis* SOS signaling pathway for Na^+ homeostasis. *Proceedings of the National Academy of Sciences* **99**: 9061–9066.

Quintero FJ, **Martinez-Atienza J**, **Villalta I**, **Jiang X**, **Kim WY**, **Ali Z**, **Fujii H**, **Mendoza I**, **Yun DJ**, **Zhu JK**, **Pardo JM**. **2011**. Activation of the plasma membrane Na/H antiporter Salt-Overly-Sensitive 1 (SOS1) by phosphorylation of an auto-inhibitory C-terminal domain. *Proceedings of the National Academy of Sciences* **108**: 2611–2616.

R Core Team. **2014**. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Rahman H, **Sabreen S**, **Alam S**, **Kawai S**. **2005**. Effects of nickel on growth and composition of metal micronutrients in barley plants grown in nutrient solution. *Journal of Plant Nutrition* **28**: 393–404.

Rajagopalan A, **Devi MT**, **Raghavendra A**. **1993**. Patterns of phosphoenolpyruvate carboxylase activity and cytosolic pH during light activation and dark deactivation in C_3 and C_4 plants. *Photosynthesis Research* **38**: 51–60.

Ralser M, **Querfurth R**, **Warnatz HJ**, **Lehrach H**, **Yaspo ML**, **Krobitsch S**. **2006**. An efficient and economic enhancer mix for PCR. *Biochemical and Biophysical Research Communications* **347**: 747–751.

Rannala B, **Yang Z**. **2003**. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**: 1645–1656.

Rapaport F, **Khanin R**, **Liang Y**, **Pirun M**, **Krek A**, **Zumbo P**, **Mason CE**, **Socci ND**, **Betel D**. **2013**. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology* **14**: R95.

Rascio N, **Navari-Izzo F**. **2011**. Heavy metal hyperaccumulating plants: how and why do they do it? And what makes them so interesting? *Plant Science* **180**: 169–181.

Rate DN, **Greenberg JT**. **2001**. The *Arabidopsis aberrant growth and death2* mutant shows resistance to *Pseudomonas syringae* and reveals a role for NPR1 in suppressing hypersensitive cell death. *Plant Journal* **27**: 203–211.

Ravet K, **Touraine B**, **Boucherez J**, **Briat JF**, **Gaymard F**, **Cellier F**. **2009a**. Ferritins control interaction between iron homeostasis and oxidative stress in *Arabidopsis*. *Plant Journal* **57**: 400–412.

Ravet K, **Touraine B**, **Kim SA**, **Cellier F**, **Thomine S**, **Guerinot ML**, **Briat JF**, **Gaymard F**. **2009b**. Post-translational regulation of *At*FER2 ferritin in response to intracellular iron trafficking during fruit development in *Arabidopsis*. *Molecular Plant* **2**: 1095–1106.

Raymond M, **Rousset F**. **1995**. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**: 248–249.

Rees JD, **Ingle RA**, **Smith JAC**. **2009**. Relative contributions of nine genes in the pathway of histidine biosynthesis to the control of free histidine concentrations in *Arabidopsis thaliana*. *Plant Biotechnology Journal* **7**: 499–511.

Reeves R. **1992**. The hyperaccumulation of nickel by serpentine plants. In: **Baker A**, **Proctor J**, **Reeves R**, (eds.) *The Vegetation of Ultramafic (Serpentine) Soils*. Andover, UK: Intercept Ltd. 253–277.

Reeves R, **Brooks R**. **1983**. European species of *Thlaspi* L. (Cruciferae) as indicators of nickel and zinc. *Journal of Geochemical Exploration* **18**: 275–283.

Reeves RD, **Baker AJ**. **2000**. *Phytoremediation of Toxic Metals: Using Plants to Clean Up the Environment*. New York, USA: Wiley.

Reeves RD, **Schwartz C**, **Morel JL**, **Edmondson J**. **2001**. Distribution and metal-accumulating behavior of *Thlaspi caerulescens* and associated metallophytes in France. *International Journal of Phytoremediation* **3**: 145–172.

Remy E, **Cabrito TR**, **Batista RA**, **Teixeira MC**, **Sá-Correia I**, **Duque P**. **2012**. The Pht1;9 and Pht1;8 transporters mediate inorganic phosphate acquisition by the *Arabidopsis thaliana* root during phosphorus starvation. *New Phytologist* **195**: 356–371.

Remy E, **Cabrito TR**, **Baster P**, **Batista RA**, **Teixeira MC**, **Friml J**, **Sá-Correia I**, **Duque P**. **2013**. A Major Facilitator Superfamily transporter plays a dual role in polar auxin transport and drought stress tolerance in *Arabidopsis*. *Plant Cell* **25**: 901–926.

Remy E, **Cabrito TR**, **Batista RA**, **Hussein MAM**, **Teixeira MC**, **Athanasiadis A**, **Sá-Correia I**, **Duque P**. **2014**. Intron retention in the 5'UTR of the novel *ZIF2* transporter enhances translation to promote zinc tolerance in *Arabidopsis*. *PLoS Genetics* **10**: e1004375.

Ren XL, **Qi GN**, **Feng HQ**, **Zhao S**, **Zhao SS**, **Wang Y**, **Wu WH**. **2013**. Calcineurin B-like protein CBL10 directly interacts with AKT1 and modulates K^+ homeostasis in *Arabidopsis*. *Plant Journal* **74**: 258–266.

Renaut S, **Owens GL**, **Rieseberg LH**. **2014**. Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Molecular Ecology* **23**: 311–324.

Renné P, **Dreßen U**, **Hebbeker U**, **Hille D**, **Flügge UI**, **Westhoff P**, **Weber APM**. **2003**. The *Arabidopsis* mutant *dct* is deficient in the plastidic glutamate/malate translocator DiT2. *Plant Journal* **35**: 316–331.

Rešetnik I, **Satovic Z**, **Schneeweiss GM**, **Liber Z**. **2013**. Phylogenetic relationships in Brassicaceae tribe Alysseae inferred from nuclear ribosomal and chloroplast DNA sequence data. *Molecular Phylogenetics and Evolution* **69**: 772–786.

Rhee SY, **Wood V**, **Dolinski K**, **Draghici S**. **2008**. Use and misuse of the gene ontology annotations. *Nature Reviews Genetics* **9**: 509–515.

Richau KH, **Kozhevnikova AD**, **Seregin IV**, **Vooijs R**, **Koevoets PLM**, **Smith JAC**, **Ivanov VB**, **Schat H**. **2009**. Chelation by histidine inhibits the vacuolar sequestration of nickel in roots of the hyperaccumulator *Thlaspi caerulescens*. *New Phytologist* **183**: 106–116.

Riley R. **1956**. The influence of breeding system on the genecology of *Thlaspi alpestre* L. *New Phytologist* **55**: 319–330.

Rishmawi L, **Pesch M**, **Juengst C**, **Schauss AC**, **Schrader A**, **Hülskamp M**. **2014**. Non-cell-autonomous regulation of root hair patterning genes by WRKY75 in *Arabidopsis*. *Plant Physiology* **165**: 186–195.

Rivas Goday S, **Esteve FC**. **1972**. Flora serpentinícola española, nota II. Nuevos edafismos y sus respectivas asociaciones del Reino de Granada. *Anales de la Real Academia de Farmacia* **38**: 409–462.

Rivas Martínez S. **1991**. Endemismos vasculares de Andalucia. *Rivasgodaya* **6**: 5–76.
Roberts A, **Pachter L**. **2013**. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature Methods* **10**: 71–73.
Robertson KR. **1981**. The genera of Amaranthaceae in the southeastern United States. *Journal of the Arnold Arboretum* **62**: 267–314.
Robinson BH, **Leblanc M**, **Petit D**, **Brooks RR**, **Kirkman JH**, **Gregg PE**. **1998**. The potential of *Thlaspi caerulescens* for phytoremediation of contaminated soils. *Plant and Soil* **203**: 47–56.
Robinson MD, **McCarthy DJ**, **Smyth GK**. **2010**. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
Robles J, **Qureshi S**, **Stephen S**, **Wilson S**, **Burden C**, **Taylor J**. **2012**. Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. *BMC Genomics* **13**: 484.
Roda F, **Ambrose L**, **Walter GM**, **Liu HL**, **Schaul A**, **Lowe A**, **Pelser PB**, **Prentis P**, **Rieseberg LH**, **Ortiz-Barrientos D**. **2013**. Genomic evidence for the parallel evolution of coastal forms in the *Senecio lautus* complex. *Molecular ecology* **22**: 2941–2952.
Rodriguez-Oubina J, **Ortiz S**. **1991**. Los pastizales pioneros vivaces de los suelos serpentínicos del NO ibérico. *Lazaroa* **12**: 333–344.
Rogers EE, **Eide DJ**, **Guerinot ML**. **2000**. Altered selectivity in an *Arabidopsis* metal transporter. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 12356–12360.
Rogers SO, **Bendich A**. **1985**. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Molecular Biology* **5**: 69–76.
Rojas CM, **Mysore KS**. **2012**. Glycolate oxidase is an alternative source for H_2O_2 production during plant defense responses and functions independently from NADPH oxidase. *Plant Signaling & Behavior* **7**: 752–755.
Roosens N, **Verbruggen N**, **Meerts P**, **Ximénez-Embún P**, **Smith JAC**. **2003**. Natural variation in cadmium tolerance and its relationship to metal hyperaccumulation for seven populations of *Thlaspi caerulescens* from western Europe. *Plant, Cell & Environment* **26**: 1657–1672.
Roosens NH, **Bernard C**, **Leplae R**, **Verbruggen N**. **2004**. Evidence for copper homeostasis function of metallothionein (MT3) in the hyperaccumulator *Thlaspi caerulescens*. *FEBS letters* **577**: 9–16.
Ros R, **Morales A**, **Segura J**, **Picazo I**. **1992**. *In vivo* and *in vitro* effects of nickel and cadmium on the plasmalemma ATPase from rice (*Oryza sativa* L.) shoots and roots. *Plant Science* **83**: 1–6.
Rose L, **Atwell S**, **Grant M**, **Holub EB**. **2012**. Parallel loss-of-function at the *RPM1* bacterial resistance locus in *Arabidopsis thaliana*. *Frontiers in plant science* **3**.
Rosenberg NA. **2004**. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**: 137–138.
Roux C, **Castric V**, **Pauwels M**, **Wright SI**, **Saumitou-Laprade P**, **Vekemans X**. **2011**. Does speciation between *Arabidopsis halleri* and *Arabidopsis lyrata* coincide with major changes in a molecular target of adaptation? *PloS one* **6**: e26872.
Rubio F, **Alemán F**, **Nieves-Cordones M**, **Martínez V**. **2010**. Studies on *Arabidopsis* A*thak5*, A*takt1* double mutants disclose the range of concentrations at which *At*HAK5, *At*AKT1 and unknown systems mediate K^+ uptake. *Physiologia Plantarum* **139**: 220–228.
Rundle HD, **Nosil P**. **2005**. Ecological speciation. *Ecology Letters* **8**: 336–352.
Ruttink T, **Sterck L**, **Rohde A**, **Bendixen C**, **Rouzé P**, **Asp T**, **Van de Peer Y**, **Roldan-Ruiz I**. **2013**. Orthology Guided Assembly in highly heterozygous crops: creating a reference transcriptome to uncover genetic diversity in *Lolium perenne*. *Plant Biotechnology Journal* **11**: 605–617.
Sabatini S, **Heidstra R**, **Wildwater M**, **Scheres B**. **2003**. SCARECROW is involved in positioning

the stem cell niche in the *Arabidopsis* root meristem. *Genes & Development* **17**: 354–358.
Sage R. 2001. Environmental and evolutionary preconditions for the origin and diversification of the C_4 photosynthetic syndrome. *Plant Biology* **3**: 202–213.
Sage RF, **Sage TL**, **Pearcy RW**, **Borsch T. 2007**. The taxonomic distribution of C_4 photosynthesis in Amaranthaceae sensu stricto. *American Journal of Botany* **94**: 1992–2003.
Sage RF, **Christin PA**, **Edwards EJ. 2011**. The C_4 plant lineages of planet Earth. *Journal of Experimental Botany* **62**: 3155–3169.
Sage RF, **Sage TL**, **Kocacinar F. 2012**. Photorespiration and the evolution of C_4 photosynthesis. *Annual Review of Plant Biology* **63**: 19–47.
Sainty G, **McCorkelle G**, **Julien M. 1997**. Control and spread of alligator weed *Alternanthera philoxeroides* (Mart.) Griseb., in Australia: lessons for other regions. *Wetlands Ecology and Management* **5**: 195–201.
Saitou N, **Nei M. 1987**. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406–425.
Sakai AK, **Weller SG**, **Wagner WL**, **Nepokroeff M**, **Culley TM. 2006**. Adaptive radiation and evolution of breeding systems in *Schiedea* (Caryophyllaceae), an endemic Hawaiian genus. *Annals of the Missouri Botanical Garden* **93**: 49–63.
Sakuraba Y, **Schelbert S**, **Park SY**, **Han SH**, **Lee BD**, **Andrès CB**, **Kessler F**, **Hörtensteiner S**, **Paek NC. 2012**. STAY-GREEN and chlorophyll catabolic enzymes interact at light-harvesting complex II for chlorophyll detoxification during leaf senescence in *Arabidopsis*. *Plant Cell* **24**: 507–518.
Salmerón-Sánchez E, **Martínez-Nieto M**, **Martínez-Hernández F**, **Garrido-Becerra J**, **Mendoza-Fernández A**, **de Carrasco C**, **Ramos-Miras J**, **Lozano R**, **Merlo M**, **Mota J. 2014a**. Ecology, genetic diversity and phylogeography of the Iberian endemic plant *Jurinea pinnata* (Lag.) DC. (Compositae) on two special edaphic substrates: dolomite and gypsum. *Plant and Soil* **374**: 233–250.
Salmerón-Sánchez E, **Merlo ME**, **Medina-Cazorla JM**, **Pérez-García FJ**, **Martínez-Hernández F**, **Garrido-Becerra JA**, **Mendoza-Fernández AJ**, **Valle F**, **Mota JF. 2014b**. Variability, genetic structure and phylogeography of the dolomitophilous species *Convolvulus boissieri* (Convolvulaceae) in the Baetic ranges, inferred from AFLPs, plastid DNA and ITS sequences. *Botanical Journal of the Linnean Society* **176**: 506–523.
Salvucci ME, **Crafts-Brandner SJ. 2004a**. Mechanism for deactivation of Rubisco under moderate heat stress. *Physiologia Plantarum* **122**: 513–519.
Salvucci ME, **Crafts-Brandner SJ. 2004b**. Relationship between the heat tolerance of photosynthesis and the thermal stability of Rubisco activase in plants from contrasting thermal environments. *Plant Physiology* **134**: 1460–1470.
Salzberg SL, **Phillippy AM**, **Zimin A**, **Puiu D**, **Magoc T**, **Koren S**, **Treangen TJ**, **Schatz MC**, **Delcher AL**, **Roberts M**, *et al.*. **2012**. GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Research* **22**: 557–567.
Sánchez-Del Pino I, **Motley TJ**, **Borsch T. 2012**. Molecular phylogenetics of *Alternanthera* (Gomphrenoideae, Amaranthaceae): resolving a complex taxonomic history caused by different interpretations of morphological characters in a lineage with C_4 and C_3– C_4 intermediate species. *Botanical Journal of the Linnean Society* **169**: 493–517.
Sánchez-Vilas J, **Philipp M**, **Retuerto R. 2010**. Unexpectedly high genetic variation in large unisexual clumps of the subdioecious plant *Honckenya peploides* (Caryophyllaceae). *Plant Biology* **12**: 518–525.
Sánchez-Vilas J, **Bermúdez R**, **Retuerto R. 2012**. Soil water content and patterns of allocation to below-and above-ground biomass in the sexes of the subdioecious plant *Honckenya peploides*.

Annals of Botany **110**: 839–848.

Sarret G, **Saumitou-Laprade P**, **Bert V**, **Proux O**, **Hazemann JL**, **Traverse A**, **Marcus MA**, **Manceau A**. **2002**. Forms of zinc accumulated in the hyperaccumulator *Arabidopsis halleri*. *Plant Physiology* **130**: 1815–1826.

Savolainen O, **Lascoux M**, **Merilä J**. **2013**. Ecological genomics of local adaptation. *Nature Reviews Genetics* **14**: 807–820.

Schaaf G, **Honsbein A**, **Meda AR**, **Kirchner S**, **Wipf D**, **von Wirén N**. **2006**. At*IREG2* encodes a tonoplast transport protein involved in iron-dependent nickel detoxification in *Arabidopsis thaliana* roots. *Journal of Biological Chemistry* **281**: 25532–25540.

Schapire AL, **Voigt B**, **Jasik J**, **Rosado A**, **Lopez-Cobollo R**, **Menzel D**, **Salinas J**, **Mancuso S**, **Valpuesta V**, **Baluska F**, **Botella MA**. **2008**. *Arabidopsis Synaptotagmin 1* is required for the maintenance of plasma membrane integrity and cell viability. *Plant Cell* **20**: 3374–3388.

Schat H, **Vooijs R**, **Kuiper E**. **1996**. Identical major gene loci for heavy metal tolerances that have independently evolved in different local populations and subspecies of *Silene vulgaris*. *Evolution* **50**: 1888–1895.

Schlötterer C. **2000**. Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**: 365–371.

Schmieder R, **Edwards R**. **2011**. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**: 863–864.

Schmitz J, **Schöttler MA**, **Krueger S**, **Geimer S**, **Schneider A**, **Kleine T**, **Leister D**, **Bell K**, **Flügge UI**, **Häusler RE**. **2012**. Defects in leaf carbohydrate metabolism compromise acclimation to high light and lead to a high chlorophyll fluorescence phenotype in *Arabidopsis thaliana*. *BMC Plant Biology* **12**: 8.

Schmitz J, **Heinrichs L**, **Scossa F**, **Fernie AR**, **Oelze ML**, **Dietz KJ**, **Rothbart M**, **Grimm B**, **Flügge UI**, **Häusler RE**. **2014**. The essential role of sugar metabolism in the acclimation response of *Arabidopsis thaliana* to high light intensities. *Journal of Experimental Botany* **65**: 1619–1636.

Schneider A, **Häusler RE**, **Kolukisaoglu Ü**, **Kunze R**, **Van Der Graaff E**, **Schwacke R**, **Catoni E**, **Desimone M**, **Flügge UI**. **2002**. An *Arabidopsis thaliana* knock-out mutant of the chloroplast triose phosphate/phosphate translocator is severely compromised only when starch synthesis, but not starch mobilisation is abolished. *Plant Journal* **32**: 685–699.

Schneider T, **Persson DP**, **Husted Sr**, **Schellenberg M**, **Gehrig P**, **Lee Y**, **Martinoia E**, **Schjoerring JK**, **Meyer S**. **2012**. A proteomics approach to investigate the process of Zn hyperaccumulation in *Noccaea caerulescens* (J & C. Presl) F.K. Meyer. *Plant Journal : for cell and molecular biology* : 131–142.

Schuler M, **Rellán-Álvarez R**, **Fink-Straube C**, **Abadía J**, **Bauer P**. **2012**. Nicotianamine functions in the phloem-based transport of iron to sink organs, in pollen development and pollen tube growth in *Arabidopsis*. *Plant Cell* **24**: 2380–400.

Schulz MH, **Zerbino DR**, **Vingron M**, **Birney E**. **2012**. Oases: Robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**: 1086–1092.

Schützendübel A, **Polle A**. **2002**. Plant responses to abiotic stresses: heavy metal-induced oxidative stress and protection by mycorrhization. *Journal of Experimental Botany* **53**: 1351–1365.

Sedlazeck FJ, **Rescheneder P**, **von Haeseler A**. **2013**. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics (Oxford, England)* **29**: 2790–2791.

Selkoe KA, **Toonen RJ**. **2006**. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**: 615–629.

Seyednasrollah F, **Laiho A**, **Elo LL**. **2013**. Comparison of software packages for detecting differential expression in RNA-seq studies. *Briefings in Bioinformatics* **16**: 59–70.

Shah J, **Zeier J**. **2013**. Long-distance communication and signal amplification in systemic acquired

resistance. *Frontiers in Plant Science* **4**.
Sharma S, **Villamor JG**, **Verslues PE**. **2011**. Essential role of tissue-specific proline synthesis and catabolism in growth and redox balance at low water potential. *Plant Physiology* **157**: 292–304.
Sheoran V, **Sheoran A**, **Poonia P**. **2010**. Role of hyperaccumulators in phytoextraction of metals from contaminated mining sites: a review. *Critical Reviews in Environmental Science and Technology* **41**: 168–214.
Shi Z, **Zhang Y**, **Maximova S**, **Guiltinan M**. **2013**. *Tc*NPR3 from *Theobroma cacao* functions as a repressor of the pathogen defense response. *BMC Plant Biology* **13**: 204.
Shigeto J, **Kiyonaga Y**, **Fujita K**, **Kondo R**, **Tsutsumi Y**. **2013**. Putative cationic cell-wall-bound peroxidase homologues in *Arabidopsis*, *At*Prx2, *At*Prx25, and *At*Prx71, are involved in lignification. *Journal of Agricultural and Food Chemistry* **61**: 3781–3788.
Shigeto J, **Nagano M**, **Fujita K**, **Tsutsumi Y**. **2014**. Catalytic profile of *Arabidopsis* peroxidases, *At*Prx-2, 25 and 71, contributing to stem lignification. *PLoS One* **9**: e105332.
Shih CF, **Hsu WH**, **Peng YJ**, **Yang CH**. **2014**. The NAC-like gene *ANTHER INDEHISCENCE FACTOR* acts as a repressor that controls anther dehiscence by regulating genes in the jasmonate biosynthesis pathway in *Arabidopsis*. *Journal of Experimental Botany* **65**: 621–639.
Shikanai T, **Müller-Moulé P**, **Munekage Y**, **Niyogi KK**, **Pilon M**. **2003**. PAA1, a P-type ATPase of *Arabidopsis*, functions in copper transport in chloroplasts. *Plant Cell* **15**: 1333–1346.
Shin LJ, **Lo JC**, **Yeh KC**. **2012**. Copper chaperone ANTIOXIDANT PROTEIN 1 is essential for copper homeostasis. *Plant Physiology* **159**: 1099–1110.
Simpson C, **Thomas C**, **Findlay K**, **Bayer E**, **Maule AJ**. **2009**. An *Arabidopsis* GPI-anchor plasmodesmal neck protein with callose binding activity and potential to regulate cell-to-cell trafficking. *Plant Cell* **21**: 581–594.
Singh AK, **Kumar R**, **Pareek A**, **Sopory SK**, **Singla-Pareek SL**. **2012**. Overexpression of rice CBS domain containing protein improves salinity, oxidative, and heavy metal tolerance in transgenic tobacco. *Molecular Biotechnology* **52**: 205–216.
Singhal S. **2013**. *De novo* transcriptomic analyses for non-model organisms: an evaluation of methods across a multi-species data set. *Molecular Ecology Resources* **13**: 403–416.
Skillman JB. **2008**. Quantum yield variation across the three pathways of photosynthesis: not yet out of the dark. *Journal of Experimental Botany* **59**: 1647–1661.
Slewinski TL, **Anderson AA**, **Zhang C**, **Turgeon R**. **2012**. *Scarecrow* plays a role in establishing Kranz anatomy in maize leaves. *Plant & Cell Physiology* **53**: 2030–2037.
Sloan DB, **Keller SR**, **Berardi AE**, **Sanderson BJ**, **Karpovich JF**, **Taylor DR**. **2012**. *De novo* transcriptome assembly and polymorphism detection in the flowering plant *Silene vulgaris* (Caryophyllaceae). *Molecular Ecology Resources* **12**: 333–343.
Słomka A, **Sutkowska A**, **Szczepaniak M**, **Malec P**, **Mitka J**, **Kuta E**. **2011**. Increased genetic diversity of *Viola tricolor* L. (Violaceae) in metal-polluted environments. *Chemosphere* **83**: 435–442.
Smart K, **Kilburn M**, **Salter C**, **Smith J**, **Grovenor C**. **2007**. NanoSIMS and EPMA analysis of nickel localisation in leaves of the hyperaccumulator plant *Alyssum lesbiacum*. *International Journal of Mass Spectrometry* **260**: 107–114.
Smart KE, **Smith JAC**, **Kilburn MR**, **Martin BG**, **Hawes C**, **Grovenor CR**. **2010**. High-resolution elemental localization in vacuolate plant cells by nanoscale secondary ion mass spectrometry. *Plant Journal* **63**: 870–879.
Sánchez-Rodríguez C, **Bauer S**, **Hématy K**, **Saxe F**, **Ibáñez AB**, **Vodermaier V**, **Konlechner C**, **Sampathkumar A**, **Rüggeberg M**, **Aichinger E**, **Neumetzler L**, **Burgert I**, **Somerville C**, **Hauser MT**, **Persson S**. **2012**. CHITINASE-LIKE1/POM-POM1 and its homolog CTL2 are glucan-interacting proteins important for cellulose biosynthesis in *Arabidopsis*. *Plant Cell* **24**:

589–607.

Soneson C, **Delorenzi M**. **2013**. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics* **14**: 91.

Song JT, **Lu H**, **Greenberg JT**. **2004**. Divergent roles in *Arabidopsis thaliana* development and defense of two homologous genes, *ABERRANT GROWTH AND DEATH2* and *AGD2-LIKE DEFENSE RESPONSE PROTEIN1*, encoding novel aminotransferases. *Plant Cell* **16**: 353–366.

Sousa V, **Hey J**. **2013**. Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics* **14**: 404–414.

Španiel S, **Kempa M**, **Salmerón-Sánchez E**, **Fuertes-Aguilar J**, **Mota JF**, **Al-Shehbaz IA**, **German DA**, **Olšavská K**, **Šingliarová B**, **Zozomová-Lihová J**, *et al.*. **2015**. AlyBase: database of names, chromosome numbers, and ploidy levels of Alysseae (Brassicaceae), with a new generic concept of the tribe. *Plant Systematics and Evolution* **301**: 2463–2491.

Stamatakis A, **Hoover P**, **Rougemont J**. **2008**. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* **57**: 758–771.

Stapley J, **Reger J**, **Feulner PGD**, **Smadja C**, **Galindo J**, **Ekblom R**, **Bennison C**, **Ball AD**, **Beckerman AP**, **Slate J**. **2010**. Adaptation genomics: the next generation. *Trends in Ecology & Evolution* **25**: 705–712.

Stebbins GL. **1942**. The genetic approach to problems of rare and endemic species. *Madrono* **60**: 302–319.

Stebbins GL, **Major J**. **1965**. Endemism and speciation in the California flora. *Ecological Monographs* : 2–35.

Stephan W. **2015**. Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Molecular Ecology* .

Stephens M, **Smith NJ**, **Donnelly P**. **2001**. A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics* **68**: 978–989.

Stern DL. **2013**. The genetic causes of convergent evolution. *Nature Reviews Genetics* **14**: 751–764.

Still CJ, **Berry JA**, **Collatz GJ**, **DeFries RS**. **2003**. Global distribution of C_3 and C_4 vegetation: carbon cycle implications. *Global Biogeochemical Cycles* **17**: 6–1–6–14.

Stotz M, **Mueller-Cajar O**, **Ciniawsky S**, **Wendler P**, **Hartl FU**, **Bracher A**, **Hayer-Hartl M**. **2011**. Structure of green-type Rubisco activase from tobacco. *Nature Structural & Molecular Biology* **18**: 1366–1370.

Stucky DF, **Arpin JC**, **Schrick K**. **2015**. Functional diversification of two UGT80 enzymes required for steryl glucoside synthesis in *Arabidopsis*. *Journal of Experimental Botany* **66**: 189–201.

Su Cl, **Chao YT**, **Alex Chang YC**, **Chen WC**, **Chen CY**, **Lee AY**, **Hwa KT**, **Shih MC**. **2011**. *De novo* assembly of expressed transcripts and global analysis of the *Phalaenopsis aphrodite* transcriptome. *Plant & Cell Physiology* **52**: 1501–1514.

Su YH, **Frommer WB**, **Ludewig U**. **2004**. Molecular and functional characterization of a family of amino acid transporters from *Arabidopsis*. *Plant Physiology* **136**: 3104–3113.

Sullivan S, **Ralet MC**, **Berger A**, **Diatloff E**, **Bischoff V**, **Gonneau M**, **Marion-Poll A**, **North HM**. **2011**. CESA5 is required for the synthesis of cellulose with a role in structuring the adherent mucilage of *Arabidopsis* seeds. *Plant Physiology* **156**: 1725–1739.

Sultan SE. **2000**. Phenotypic plasticity for plant development, function and life history. *Trends in Plant Science* **5**: 537–542.

Sun Q, **Gao F**, **Zhao L**, **Li K**, **Zhang J**. **2010**. Identification of a new 130 bp cis-acting element in the Ts*VP1* promoter involved in the salt stress response from *Thellungiella halophila*. *BMC Plant Biology* **10**: 90.

Sun W, **Van Montagu M**, **Verbruggen N**. **2002**. Small heat shock proteins and stress tolerance in

plants. *Biochimica et Biophysica Acta - Gene Structure and Expression* **1577**: 1–9.

Sun X, **Peng L**, **Guo J**, **Chi W**, **Ma J**, **Lu C**, **Zhang L**. **2007**. Formation of DEG5 and DEG8 complexes and their involvement in the degradation of photodamaged Photosystem II reaction center D1 protein in *Arabidopsis*. *Plant Cell* **19**: 1347–1361.

Sung DY, **Kaplan F**, **Lee KJ**, **Guy CL**. **2003**. Acquired tolerance to temperature extremes. *Trends in Plant Science* **8**: 179–187.

Sunil M, **Hariharan AK**, **Nayak S**, **Gupta S**, **Nambisan SR**, **Gupta RP**, **Panda B**, **Choudhary B**, **Srinivasan S**. **2014**. The draft genome and transcriptome of *Amaranthus hypochondriacus*: a C_4 dicot producing high-lysine edible pseudo-cereal. *DNA Research* **21**: 585–602.

Suorsa M, **Järvi S**, **Grieco M**, **Nurmi M**, **Pietrzykowska M**, **Rantala M**, **Kangasjärvi S**, **Paakkarinen V**, **Tikkanen M**, **Jansson S**, **Aro EM**. **2012**. PROTON GRADIENT REGULATION5 is essential for proper acclimation of *Arabidopsis* Photosystem I to naturally and artificially fluctuating light conditions. *Plant Cell* **24**: 2934–2948.

Surget-Groba Y, **Montoya-Burgos JI**. **2010**. Optimization of *de novo* transcriptome assembly from next-generation sequencing data. *Genome Research* **20**: 1432–1440.

Suzuki N, **Rizhsky L**, **Liang H**, **Shuman J**, **Shulaev V**, **Mittler R**. **2005**. Enhanced tolerance to environmental stress in transgenic plants expressing the transcriptional coactivator *Multiprotein Bridging Factor 1c*. *Plant Physiology* **139**: 1313–1322.

Svennerstam H, **Ganeteg U**, **Näsholm T**. **2008**. Root uptake of cationic amino acids by *Arabidopsis* depends on functional expression of *Amino Acid Permease 5*. *New Phytologist* **180**: 620–630.

Svennerstam H, **Jämtgård S**, **Ahmad I**, **Huss-Danell K**, **Näsholm T**, **Ganeteg U**. **2011**. Transporters in *Arabidopsis* roots mediating uptake of amino acids at naturally occurring concentrations. *New Phytologist* **191**: 459–467.

Svistoonoff S, **Creff A**, **Reymond M**, **Sigoillot-Claude C**, **Ricaud L**, **Blanchet A**, **Nussaume L**, **Desnos T**. **2007**. Root tip contact with low-phosphate media reprograms plant root architecture. *Nature Genetics* **39**: 792–796.

Tabas-Madrid D, **Nogales-Cadenas R**, **Pascual-Montano A**. **2012**. GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. *Nucleic Acids Research* **40**: 478–483.

Tajima F. **1989**. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.

Takano J, **Noguchi K**, **Yasumori M**, **Kobayashi M**, **Gajdos Z**, **Miwa K**, **Hayashi H**, **Yoneyama T**, **Fujiwara T**. **2002**. *Arabidopsis* boron transporter for xylem loading. *Nature* **420**: 337–340.

Takano J, **Wada M**, **Ludewig U**, **Schaaf G**, **von Wirén N**, **Fujiwara T**. **2006**. The *Arabidopsis* Major Intrinsic Protein NIP5;1 is essential for efficient boron uptake and plant development under boron limitation. *Plant Cell* **18**: 1498–1509.

Takano J, **Tanaka M**, **Toyoda A**, **Miwa K**, **Kasai K**, **Fuji K**, **Onouchi H**, **Naito S**, **Fujiwara T**. **2010**. Polar localization and degradation of *Arabidopsis* boron transporters through distinct trafficking pathways. *Proceedings of the National Academy of Sciences* **107**: 5220–5225.

Takezaki N, **Nei M**, **Tamura K**. **2010**. POPTREE2: software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Molecular Biology and Evolution* **27**: 747–752.

Talke IN, **Hanikenne M**, **Krämer U**. **2006**. Zinc-dependent global transcriptional control, transcriptional deregulation, and higher gene copy number for genes in metal homeostasis of the hyperaccumulator *Arabidopsis halleri*. *Plant Physiology* **142**: 148–167.

Tamura K, **Stecher G**, **Peterson D**, **Filipski A**, **Kumar S**. **2013**. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**: 2725–2729.

Tanaka A, **Ito H**, **Tanaka R**, **Tanaka NK**, **Yoshida K**, **Okada K**. **1998**. *Chlorophyll A Oxygenase (CAO)* is involved in chlorophyll b formation from chlorophyll a. *Proceedings of the National Academy of Sciences* **95**: 12719–12723.

Tanaka M, **Wallace IS**, **Takano J**, **Roberts DM**, **Fujiwara T**. **2008**. NIP6;1 is a boric acid channel for preferential transport of boron to growing shoot tissues in *Arabidopsis*. *Plant Cell* **20**: 2860–2875.

Tanaka M, **Takano J**, **Chiba Y**, **Lombardo F**, **Ogasawara Y**, **Onouchi H**, **Naito S**, **Fujiwara T**. **2011**. Boron-dependent degradation of *NIP5;1* mRNA for acclimation to excess boron conditions in *Arabidopsis*. *Plant Cell* **23**: 3547–3559.

Tanase K, **Nishitani C**, **Hirakawa H**, **Isobe S**, **Tabata S**, **Ohmiya A**, **Onozaki T**. **2012**. Transcriptome analysis of carnation (*Dianthus caryophyllus* L.) based on next-generation sequencing technology. *BMC Genomics* **13**.

Tang RJ, **Zhao FG**, **Garcia VJ**, **Kleist TJ**, **Yang L**, **Zhang HX**, **Luan S**. **2015**. Tonoplast CBL–CIPK calcium signaling network regulates magnesium homeostasis in *Arabidopsis*. *Proceedings of the National Academy of Sciences* **112**: 3134–3139.

Taniguchi M, **Taniguchi Y**, **Kawasaki M**, **Takeda S**, **Kato T**, **Sato S**, **Tabata S**, **Miyake H**, **Sugiyama T**. **2002**. Identifying and characterizing plastidic 2-oxoglutarate/malate and dicarboxylate transporters in *Arabidopsis thaliana*. *Plant & Cell Physiology* **43**: 706–717.

Tao X, **Gu YH**, **Wang HY**, **Zheng W**, **Li X**, **Zhao CW**, **Zhang YZ**. **2012**. Digital gene expression analysis based on integrated *de novo* transcriptome assembly of sweet potato *Ipomoea batatas* (L.). *PLoS One* **7**: e36234.

Tarazona S, **García-Alcalde F**, **Dopazo J**, **Ferrer A**, **Conesa A**. **2011**. Differential expression in RNA-seq: a matter of depth. *Genome Research* **21**: 2213–2223.

Teige M, **Scheikl E**, **Eulgem T**, **Dóczi R**, **Ichimura K**, **Shinozaki K**, **Dangl JL**, **Hirt H**. **2004**. The MKK2 pathway mediates cold and salt stress signaling in *Arabidopsis*. *Molecular Cell* **15**: 141–152.

Teshima KM, **Coop G**, **Przeworski M**. **2006**. How reliable are empirical genomic scans for selective sweeps? *Genome Research* **16**: 702–712.

Textor S, **de Kraker JW**, **Hause B**, **Gershenzon J**, **Tokuhisa JG**. **2007**. MAM3 catalyzes the formation of all aliphatic glucosinolate chain lengths in *Arabidopsis*. *Plant Physiology* **144**: 60–71.

Thomine S, **Wang R**, **Ward JM**, **Crawford NM**, **Schroeder JI**. **2000**. Cadmium and iron transport by members of a plant metal transporter family in *Arabidopsis* with homology to *Nramp* genes. *Proceedings of the National Academy of Sciences* **97**: 4991–4996.

Thomine S, **Lelièvre F**, **Debarbieux E**, **Schroeder JI**, **Barbier-Brygoo H**. **2003**. *At*NRAMP3, a multispecific vacuolar metal transporter involved in plant responses to iron deficiency. *Plant Journal* **34**: 685–695.

Ticconi CA, **Lucero RD**, **Sakhonwasee S**, **Adamson AW**, **Creff A**, **Nussaume L**, **Desnos T**, **Abel S**. **2009**. ER-resident proteins PDR2 and LPR1 mediate the developmental response of root meristems to phosphate availability. *Proceedings of the National Academy of Sciences* **106**: 14174–14179.

Tiffin P, **Ross-Ibarra J**. **2014**. Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution* **29**: 673–680.

Tsegaye Y, **Richardson CG**, **Bravo JE**, **Mulcahy BJ**, **Lynch DV**, **Markham JE**, **Jaworski JG**, **Chen M**, **Cahoon EB**, **Dunn TM**. **2007**. *Arabidopsis* mutants lacking long chain base phosphate lyase are fumonisin-sensitive and accumulate trihydroxy-18:1 long chain base phosphate. *Journal of Biological Chemistry* **282**: 28195–28206.

Tuomainen M, **Tervahauta A**, **Hassinen V**, **Schat H**, **Koistinen KM**, **Lehesranta S**, **Rantalainen K**, **Häyrinen J**, **Auriola S**, **Anttonen M**, **Kärenlampi S**. **2010**. Proteomics of *Thlaspi caerulescens* accessions and an inter-accession cross segregating for zinc accumulation. *Journal*

of Experimental Botany **61**: 1075–1087.

Turner TL, **von Wettberg EJ**, **Nuzhdin SV**. **2008**. Genomic analysis of differentiation between soil types reveals candidate genes for local adaptation in *Arabidopsis lyrata*. *PLoS One* **3**: e3183.

Turner TL, **Bourne EC**, **Von Wettberg EJ**, **Hu TT**, **Nuzhdin SV**. **2010**. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics* **42**: 260–263.

Tutin T. **1993**. *Flora Europaea: Psilotaceae to Platanaceae*. Flora Europaea 5 Volume Hardback Set. Cambridge, UK: Cambridge University Press.

Uemura T, **Kim H**, **Saito C**, **Ebine K**, **Ueda T**, **Schulze-Lefert P**, **Nakano A**. **2012**. Qa-SNAREs localized to the trans-Golgi network regulate multiple transport pathways and extracellular disease resistance in plants. *Proceedings of the National Academy of Sciences* **109**: 1784–1789.

Ueno D, **Milner MJ**, **Yamaji N**, **Yokosho K**, **Koyama E**, **Clemencia Zambrano M**, **Kaskie M**, **Ebbs S**, **Kochian LSV**, **Ma JF**. **2011**. Elevated expression of *Tc*HMA3 plays a key role in the extreme Cd tolerance in a Cd-hyperaccumulating ecotype of *Thlaspi caerulescens*. *Plant Journal* **66**: 852–862.

Ueno Y, **Ishikawa T**, **Watanabe K**, **Terakura S**, **Iwakawa H**, **Okada K**, **Machida C**, **Machida Y**. **2007**. Histone deacetylases and ASYMMETRIC LEAVES2 are involved in the establishment of polarity in leaves of *Arabidopsis*. *Plant Cell* **19**: 445–457.

Vacchina V, **Mari S**, **Czernic P**, **Marquès L**, **Pianelli K**, **Schaumlöffel D**, **Lebrun M**, **Lobinski R**. **2003**. Speciation of nickel in a hyperaccumulating plant by high-performance liquid chromatography-inductively coupled plasma mass spectrometry and electrospray MS/MS assisted by cloning using yeast complementation. *Analytical Chemistry* **75**: 2740–2745.

Valdés A, **Övernäs E**, **Johansson H**, **Rada-Iglesias A**, **Engström P**. **2012**. The homeodomain-leucine zipper (HD-Zip) class I transcription factors ATHB7 and ATHB12 modulate abscisic acid signalling by regulating protein PHOSPHATASE 2C and abscisic acid receptor gene activities. *Plant Molecular Biology* **80**: 405–418.

Valpuesta V, **Botella MA**. **2004**. Biosynthesis of L-ascorbic acid in plants: new pathways for an old antioxidant. *Trends in Plant Science* **9**: 573–577.

Van Aken O, **Zhang B**, **Law S**, **Narsai R**, **Whelan J**. **2013**. *At*WRKY40 and *At*WRKY63 modulate the expression of stress-responsive nuclear genes encoding mitochondrial and chloroplast proteins. *Plant Physiology* **162**: 254–271.

van de Mortel JE, **Almar Villanueva L**, **Schat H**, **Kwekkeboom J**, **Coughlan S**, **Moerland PD**, **Ver Loren van Themaat E**, **Koornneef M**, **Aarts MGM**. **2006**. Large expression differences in genes for iron and zinc homeostasis, stress response, and lignin biosynthesis distinguish roots of *Arabidopsis thaliana* and the related metal hyperaccumulator *Thlaspi caerulescens*. *Plant Physiology* **142**: 1127–1147.

Van der Ent A, **Baker AJM**, **Reeves RD**, **Pollard AJ**, **Schat H**. **2013**. Hyperaccumulators of metal and metalloid trace elements: facts and fiction. *Plant and Soil* **362**: 319–334.

Van der Ent A, **Baker AJ**, **Reeves RD**, **Chaney RL**, **Anderson CW**, **Meech JA**, **Erskine PD**, **Simonnot MO**, **Vaughan J**, **Morel JL**, *et al.*. **2015**. Agromining: farming for metals in the future? *Environmental Science & Technology* **49**: 4773–4780.

Van Oosterhout C, **Hutchinson WF**, **Wills DP**, **Shipley P**. **2004**. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**: 535–538.

van Verk M, **Bol J**, **Linthorst H**. **2011**. WRKY transcription factors involved in activation of SA biosynthesis genes. *BMC Plant Biology* **11**: 89.

Vanderauwera S, **Vandenbroucke K**, **Inzé A**, **van de Cotte B**, **Mühlenbock P**, **De Rycke R**, **Naouar N**, **Van Gaever T**, **Van Montagu MCE**, **Van Breusegem F**. **2012**. *At*WRKY15 perturbation abolishes the mitochondrial stress response that steers osmotic stress tolerance in

Arabidopsis. *Proceedings of the National Academy of Sciences* **109**: 20113–20118.

Vanholme R, **Cesarino I**, **Rataj K**, **Xiao Y**, **Sundin L**, **Goeminne G**, **Kim H**, **Cross J**, **Morreel K**, **Araujo P**, **Welsh L**, **Haustraete J**, **McClellan C**, **Vanholme B**, **Ralph J**, **Simpson GG**, **Halpin C**, **Boerjan W**. **2013**. Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. *Science* **341**: 1103–1106.

Varotto C, **Maiwald D**, **Pesaresi P**, **Jahns P**, **Salamini F**, **Leister D**. **2002**. The metal ion transporter IRT1 is necessary for iron homeostasis and efficient photosynthesis in *Arabidopsis thaliana*. *Plant Journal* **31**: 589–599.

Varshney RK, **Thiel T**, **Stein N**, **Langridge P**, **Graner A**. **2002**. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cellular and Molecular Biology Letters* **7**: 537–546.

Varshney RK, **Graner A**, **Sorrells ME**. **2005**. Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* **23**: 48–55.

Vatén A, **Dettmer J**, **Wu S**, **Stierhof YD**, **Miyashima S**, **Yadav SR**, **Roberts CJ**, **Campilho A**, **Bulone V**, **Lichtenberger R**, **Lehesranta S**, **Mähönen AP**, **Kim JY**, **Jokitalo E**, **Sauer N**, **Scheres B**, **Nakajima K**, **Carlsbecker A**, **Gallagher KL**, **Helariutta Y**. **2011**. Callose biosynthesis regulates symplastic trafficking during root development. *Developmental Cell* **21**: 1144–1155.

Vekemans X, **Lefèbvre C**. **1997**. On the evolution of heavy-metal tolerant populations in *Armeria maritima*: evidence from allozyme variation and reproductive barriers. *Journal of Evolutionary Biology* **10**: 175–191.

Verbruggen N, **Hermans C**, **Schat H**. **2009**. Molecular mechanisms of metal hyperaccumulation in plants. *New Phytologist* **181**: 759–776.

Verbruggen N, **Hanikenne M**, **Clemens S**. **2013**. A more complete picture of metal hyperaccumulation through next-generation sequencing technologies. *Frontiers in Plant Science* **4**: 388.

Vert G, **Grotz N**, **Dédaldéchamp F**, **Gaymard F**, **Guerinot ML**, **Briat JF**, **Curie C**. **2002**. IRT1, an *Arabidopsis* transporter essential for iron uptake from the soil and for plant growth. *Plant Cell* **14**: 1223–1233.

Vierling E. **1991**. The roles of heat shock proteins in plants. *Annual Review of Plant Biology* **42**: 579–620.

Vijay N, **Poelstra JW**, **Künstner A**, **Wolf JBW**. **2013**. Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive *in silico* assessment of RNA-seq experiments. *Molecular Ecology* **22**: 620–634.

Vijayan P, **Browse J**. **2002**. Photoinhibition in mutants of *Arabidopsis* deficient in thylakoid unsaturation. *Plant Physiology* **129**: 876–885.

Villemereuil P, **Frichot É**, **Bazin É**, **François O**, **Gaggiotti OE**. **2014**. Genome scan methods against more complex models: when and how much should we trust them? *Molecular Ecology* **23**: 2006–2019.

Viotti C, **Bubeck J**, **Stierhof YD**, **Krebs M**, **Langhans M**, **van den Berg W**, **van Dongen W**, **Richter S**, **Geldner N**, **Takano J**, **Jürgens G**, **de Vries SC**, **Robinson DG**, **Schumacher K**. **2010**. Endocytic and secretory traffic in *Arabidopsis* merge in the *trans*-Golgi network/early endosome, an independent and highly dynamic organelle. *Plant Cell* **22**: 1344–1357.

Visioli G, **Vincenzi S**, **Marmiroli M**, **Marmiroli N**. **2012**. Correlation between phenotype and proteome in the Ni hyperaccumulator *Noccaea caerulescens* subsp. *caerulescens*. *Environmental and Experimental Botany* **77**: 156–164.

Visioli G, **Gullì M**, **Marmiroli N**. **2014**. *Noccaea caerulescens* populations adapted to grow in metalliferous and non-metalliferous soils: Ni tolerance, accumulation and expression analysis of

genes involved in metal homeostasis. *Environmental and Experimental Botany* **105**: 10–17.
Vitti JJ, **Grossman SR**, **Sabeti PC**. **2013**. Detecting natural selection in genomic data. *Annual Review of Genetics* **47**: 97–120.
Vogt R, **Oberprieler C**. **1994**. Chromosome numbers of North African phanerogams 4. *Candollea* **49**: 549–570.
Španiel S, **Marhold K**, **Passalacqua NG**, **Zozomová-Lihová J**. **2011**. Intricate variation patterns in the diploid-polyploid complex of *Alyssum montanum-A. repens* (Brassicaceae) in the Apennine Peninsula: evidence for long-term persistence and diversification. *American Journal of Botany* **98**: 1887–1904.
Španiel S, **Marhold K**, **Thiv M**, **Zozomová-Lihová J**. **2012**. A new circumscription of *Alyssum montanum* ssp. *montanum* and *A. montanum* ssp. *gmelinii* (Brassicaceae) in Central Europe: molecular and morphological evidence. *Botanical Journal of the Linnean Society* **169**: 378–402.
Waldron KJ, **Rutherford JC**, **Ford D**, **Robinson NJ**. **2009**. Metalloproteins and metal sensing. *Nature* **460**: 823–830.
Wallace LE, **Weller SG**, **Wagner WL**, **Sakai AK**, **Nepokroeff M**. **2009**. Phylogeographic patterns and demographic history of *Schiedea globosa* (Caryophyllaceae) on the Hawaiian Islands. *American Journal of Botany* **96**: 958–967.
Walters RG, **Shephard F**, **Rogers JJ**, **Rolfe SA**, **Horton P**. **2003**. Identification of mutants of *Arabidopsis* defective in acclimation of photosynthesis to the light environment. *Plant Physiology* **131**: 472–481.
Wang B, **Li W**, **Wang J**. **2005**. Genetic diversity of *Alternanthera philoxeroides* in China. *Aquatic Botany* **81**: 277–283.
Wang D, **Portis AR**, **Moose SP**, **Long SP**. **2008a**. Cool C_4 photosynthesis: pyruvate pi dikinase expression and activity corresponds to the exceptional cold tolerance of carbon assimilation in *Miscanthus* × *giganteus*. *Plant Physiology* **148**: 557–567.
Wang H, **Feng T**, **Peng X**, **Yan M**, **Tang X**. **2009a**. Up-regulation of chloroplastic antioxidant capacity is involved in alleviation of nickel toxicity of *Zea mays* L. by exogenous salicylic acid. *Ecotoxicology and Environmental Safety* **72**: 1354–1362.
Wang L, **Czedik-Eysenberg A**, **Mertz RA**, **Si Y**, **Tohge T**, **Nunes-Nesi A**, **Arrivault S**, **Dedow LK**, **Bryant DW**, **Zhou W**, *et al.*. **2014**. Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of maize and rice. *Nature Biotechnology* **32**: 1158–1165.
Wang N, **Yu FH**, **Li PX**, **He WM**, **Liu FH**, **Liu JM**, **Dong M**. **2008b**. Clonal integration affects growth, photosynthetic efficiency and biomass allocation, but not the competitive ability, of the alien invasive *Alternanthera philoxeroides* under severe stress. *Annals of Botany* **101**: 671–678.
Wang N, **Yu FH**, **Li PX**, **He WM**, **Liu J**, **Yu GL**, **Song YB**, **Dong M**. **2009b**. Clonal integration supports the expansion from terrestrial to aquatic environments of the amphibious stoloniferous herb *Alternanthera philoxeroides*. *Plant Biology* **11**: 483–489.
Wang W, **Vinocur B**, **Shoseyov O**, **Altman A**. **2004**. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science* **9**: 244–252.
Wang Z, **Gerstein M**, **Snyder M**. **2009c**. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**: 57–63.
Warwick SI, **Sauder Ca**, **Al-Shehbaz Ia**. **2008**. Phylogenetic relationships in the tribe *Alysseae* (Brassicaceae) based on nuclear ribosomal ITS DNA sequences. *Botany* **86**: 315–336.
Waters BM, **Chu HH**, **DiDonato RJ**, **Roberts LA**, **Eisley RB**, **Lahner B**, **Salt DE**, **Walker EL**. **2006**. Mutations in *Arabidopsis Yellow Stripe-Like1* and *Yellow Stripe-Like3* reveal their roles in metal ion homeostasis and loading of metal ions in seeds. *Plant Physiology* **141**: 1446–1458.
Watterson GA. **1975**. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**: 256–276.

Weber M, **Harada E**, **Vess C**, **Roepenack-Lahaye EV**, **Clemens S**. **2004**. Comparative microarray analysis of *Arabidopsis thaliana* and *Arabidopsis halleri* roots identifies nicotianamine synthase, a ZIP transporter and other genes as potential metal hyperaccumulation factors. *Plant Journal* **37**: 269–281.

Wei W, **Qi X**, **Wang L**, **Zhang Y**, **Hua W**, **Li D**, **Lv H**, **Zhang X**. **2011**. Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genomics* **12**: 451.

Weir BS, **Cockerham CC**. **1984**. Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

Weller SG, **Sakai AK**, **Wagner WL**, **Herbst DR**. **1990**. Evolution of dioecy in *Schiedea* (Caryophyllaceae: Alsinoideae) in the Hawaiian Islands: biogeographical and ecological factors. *Systematic Botany* **15**: 266–276.

Weller SG, **Wagner WL**, **Sakai AK**. **1995**. A phylogenetic analysis of *Schiedea* and *Alsinidendron* (Caryophyllaceae: Alsinoideae): implications for the evolution of breeding systems. *Systematic Botany* **20**: 315–337.

Weng JK, **Mo H**, **Chapple C**. **2010**. Over-expression of F5H in COMT-deficient *Arabidopsis* leads to enrichment of an unusual lignin and disruption of pollen wall formation. *Plant Journal* **64**: 898–911.

Westram AM, **Galindo J**, **Alm Rosenblad M**, **Grahame JW**, **Panova M**, **Butlin RK**. **2014**. Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? *Molecular Ecology* **23**: 4603–4616.

Whitney SM, **Houtz RL**, **Alonso H**. **2011**. Advancing our understanding and capacity to engineer nature's CO_2-sequestering enzyme, Rubisco. *Plant Physiology* **155**: 27–35.

Willems G, **Dräger DB**, **Courbot M**, **Godé C**, **Verbruggen N**, **Saumitou-Laprade P**. **2007**. The genetic basis of zinc tolerance in the metallophyte *Arabidopsis halleri* ssp. *halleri* (Brassicaceae): an analysis of quantitative trait loci. *Genetics* **176**: 659–674.

Williams R. **2007**. A chemical systems approach to evolution. *Dalton Transactions* **10**: 991–1001.

Windhövel A, **Hein I**, **Dabrowa R**, **Stockhaus J**. **2001**. Characterization of a novel class of plant homeodomain proteins that bind to the C_4 phosphoenolpyruvate carboxylase gene of *Flaveria trinervia*. *Plant Molecular Biology* **45**: 201–214.

Wingler A, **Quick WP**, **Bungard RA**, **Bailey KJ**, **Lea PJ**, **Leegood RC**. **1999**. The role of photorespiration during drought stress: an analysis utilizing barley mutants with reduced activities of photorespiratory enzymes. *Plant, Cell & Environment* **22**: 361–373.

Wingler A, **Lea PJ**, **Quick WP**, **Leegood RC**. **2000**. Photorespiration: metabolic pathways and their role in stress protection. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **355**: 1517–1529.

Woeste KE, **Kieber JJ**. **2000**. A strong loss-of-function mutation in *RAN1* results in constitutive activation of the ethylene response pathway as well as a rosette-lethal phenotype. *Plant Cell* **12**: 443–455.

Wolf JBW. **2013**. Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Molecular Ecology Resources* **13**: 559–572.

Worrall D, **Liang YK**, **Alvarez S**, **Holroyd GH**, **Spiegel S**, **Panagopulos M**, **Gray JE**, **Hetherington AM**. **2008**. Involvement of sphingosine kinase in plant cell signalling. *Plant Journal* **56**: 64–72.

Wright S. **1943**. Isolation by distance. *Genetics* **28**: 114–138.

Wu HJ, **Zhang Z**, **Wang JY**, **Oh DH**, **Dassanayake M**, **Liu B**, **Huang Q**, **Sun HX**, **Xia R**, **Wu Y**, **Wang YN**, **Yang Z**, **Liu Y**, **Zhang W**, **Zhang H**, **Chu J**, **Yan C**, **Fang S**, **Zhang J**, **Wang Y**, **Zhang F**, **Wang G**, **Lee SY**, **Cheeseman JM**, **Yang B**, **Li B**, **Min J**, **Yang L**, **Wang J**, **Chu C**, **Chen SY**,

Bohnert HJ, **Zhu JK**, **Wang XJ**, **Xie Q**. **2012**. Insights into salt tolerance from the genome of *Thellungiella salsuginea*. *Proceedings of the National Academy of Sciences* **109**: 12219–12224.

Wu K, **Tian L**, **Malik K**, **Brown D**, **Miki B**. **2000**. Functional analysis of HD2 histone deacetylase homologues in *Arabidopsis thaliana*. *Plant Journal* **22**: 19–27.

Xu J, **Sun J**, **Du L**, **Liu X**. **2012**. Comparative transcriptome analysis of cadmium responses in *Solanum nigrum* and *Solanum torvum*. *New Phytologist* **196**: 110–124.

Xu J, **Li Y**, **Ma X**, **Ding J**, **Wang K**, **Wang S**, **Tian Y**, **Zhang H**, **Zhu XG**. **2013**. Whole transcriptome analysis using next-generation sequencing of model species *Setaria viridis* to support C_4 photosynthesis research. *Plant Molecular Biology* **83**: 77–87.

Xu X, **Chen C**, **Fan B**, **Chen Z**. **2006**. Physical and functional interactions between pathogen-induced *Arabidopsis* WRKY18, WRKY40, and WRKY60 transcription factors. *Plant Cell* **18**: 1310–1326.

Xu XM, **Adams S**, **Chua NH**, **Møller SG**. **2005**. *At*NAP1 represents an atypical SufB protein in *Arabidopsis* plastids. *Journal of Biological Chemistry* **280**: 6648–6654.

Yamada K, **Osakabe Y**, **Mizoi J**, **Nakashima K**, **Fujita Y**, **Shinozaki K**, **Yamaguchi-Shinozaki K**. **2010**. Functional analysis of an *Arabidopsis thaliana* abiotic stress-inducible facilitated diffusion transporter for monosaccharides. *Journal of Biological Chemistry* **285**: 1138–1146.

Yamada N, **Takahashi H**, **Kitou K**, **Sahashi K**, **Tamagake H**, **Tanaka Y**, **Takabe T**. **2015**. Suppressed expression of choline monooxygenase in sugar beet on the accumulation of glycine betaine. *Plant Physiology and Biochemistry* **96**: 217–221.

Yamaguchi T, **Blumwald E**. **2005**. Developing salt-tolerant crop plants: challenges and opportunities. *Trends in Plant Science* **10**: 615–620.

Yamauchi Y, **Hasegawa A**, **Mizutani M**, **Sugimoto Y**. **2012**. Chloroplastic NADPH-dependent alkenal/one oxidoreductase contributes to the detoxification of reactive carbonyls produced under oxidative stress. *FEBS Letters* **586**: 1208–1213.

Yamazaki T, **Kawamura Y**, **Minami A**, **Uemura M**. **2008**. Calcium-dependent freezing tolerance in *Arabidopsis* involves membrane resealing via synaptotagmin SYT1. *Plant Cell* **20**: 3389–3404.

Yamori W, **Suzuki K**, **Noguchi K**, **Nakai M**, **Terashima I**. **2006**. Effects of Rubisco kinetics and Rubisco activation state on the temperature dependence of the photosynthetic rate in spinach leaves from contrasting growth temperatures. *Plant, Cell & Environment* **29**: 1659–1670.

Yamori W, **Masumoto C**, **Fukayama H**, **Makino A**. **2012**. Rubisco activase is a key regulator of non-steady-state photosynthesis at any leaf temperature and, to a lesser extent, of steady-state photosynthesis at high temperature. *Plant Journal* **71**: 871–880.

Yamori W, **Hikosaka K**, **Way DA**. **2014**. Temperature response of photosynthesis in C_3, C_4, and CAM plants: temperature acclimation and temperature adaptation. *Photosynthesis Research* **119**: 101–117.

Yang L, **Qin L**, **Liu G**, **Peremyslov VV**, **Dolja VV**, **Wei Y**. **2014**. Myosins XI modulate host cellular responses and penetration resistance to fungal pathogens. *Proceedings of the National Academy of Sciences* **111**: 13996–14001.

Yang TJ, **Lin WD**, **Schmidt W**. **2010a**. Transcriptional profiling of the *Arabidopsis* iron deficiency response reveals conserved transition metal homeostasis networks. *Plant Physiology* **152**: 2130–2141.

Yang X, **Dorman KS**, **Aluru S**. **2010b**. Reptile: representative tiling for short read error correction. *Bioinformatics* **26**: 2526–2533.

Yang Y, **Moore MJ**, **Brockington SF**, **Soltis DE**, **Wong GKS**, **Carpenter EJ**, **Zhang Y**, **Chen L**, **Yan Z**, **Xie Y**, **Sage RF**, **Covshoff S**, **Hibberd JM**, **Nelson MN**, **Smith SA**. **2015**. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Molecular Biology and Evolution* **32**: 2001–2014.

Yang Z. **2006**. *Computational Molecular Evolution*. volume 21 of *Oxford Series in Ecology and Evolution*. Oxford, UK: Oxford University Press.

Yang Z. **2007**. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.

Yang Z, **Rannala B**. **2010**. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 9264–9.

Yang Z, **Rannala B**. **2014**. Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution* **31**: 3125–3135.

Yi X, **Du Z**, **Su Z**. **2013**. PlantGSEA: a gene set enrichment analysis toolkit for plant community. *Nucleic Acids Research* **41**: 98–103.

Yokoo T, **Kobayashi S**, **Oginuma K**, **Fujikawa K**, **Mitsui Y**, **Ikeda H**, **Setoguchi H**. **2009**. Genetic structure among and within populations of the serpentine endemic *Heteropappus hispidus* ssp. *leptocladus* (Compositae). *Biochemical Systematics and Ecology* **37**: 275–284.

Yoon H, **Leitner T**. **2015**. PrimerDesign-M: a multiple-alignment based multiple-primer design tool for walking across variable genomes. *Bioinformatics* **31**: 1472–1474.

Yoshida K, **Kaothien P**, **Matsui T**, **Kawaoka A**, **Shinmyo A**. **2003**. Molecular biology and application of plant peroxidase genes. *Applied Microbiology and Biotechnology* **60**: 665–670.

Young A, **Boyle T**, **Brown T**. **1996**. The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution* **11**: 413–418.

Young MD, **Wakefield MJ**, **Smyth GK**, **Oshlack A**. **2010**. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11**: R14.

Yu Y, **Than C**, **Degnan JH**, **Nakhleh L**. **2011**. Coalescent histories on phylogenetic networks and detection of hybridization despite incomplete lineage sorting. *Systematic Biology* **60**: 138–149.

Zeng K, **Fu YX**, **Shi S**, **Wu CI**. **2006**. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* **174**: 1431–1439.

Zhai Z, **Gayomba SR**, **Jung Hi**, **Vimalakumari NK**, **Piñeros M**, **Craft E**, **Rutzke MA**, **Danku J**, **Lahner B**, **Punshon T**, **Guerinot ML**, **Salt DE**, **Kochian LSV**, **Vatamaniuk OK**. **2014**. OPT3 is a phloem-specific iron transporter that is essential for systemic iron signaling and redistribution of iron and cadmium in *Arabidopsis*. *Plant Cell* **26**: 2249–2264.

Zhang C, **Kotchoni SO**, **Samuels AL**, **Szymanski DB**. **2010**. SPIKE1 signals originate from and assemble specialized domains of the endoplasmic reticulum. *Current Biology* **20**: 2144–2149.

Zhang C, **Rannala B**, **Yang Z**. **2014a**. Bayesian species delimitation can be robust to guide-tree inference errors. *Systematic Biology* **63**: 993–1004.

Zhang J, **Liang S**, **Duan J**, **Wang J**, **Chen S**, **Cheng Z**, **Zhang Q**, **Liang X**, **Li Y**. **2012**. *De novo* assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L.). *BMC Genomics* **13**: 90.

Zhang J, **Zhang M**, **Tian S**, **Lu L**, **Shohag MJI**, **Yang X**. **2014b**. Metallothionein 2 (*Sa*MT2) from *Sedum alfredii* Hance confers increased Cd tolerance and accumulation in yeast and tobacco. *PLoS One* **9**: e102750.

Zhang JL, **Shi H**. **2013**. Physiological and molecular mechanisms of plant salt tolerance. *Photosynthesis Research* **115**: 1–22.

Zhang JT, **Zhu JQ**, **Zhu Q**, **Liu H**, **Gao XS**, **Zhang HX**. **2009**. Fatty acid desaturase-6 (FAD6) is required for salt tolerance in *Arabidopsis thaliana*. *Biochemical and Biophysical Research Communications* **390**: 469–474.

Zhang K, **Zhi D**. **2013**. Joint haplotype phasing and genotype calling of multiple individuals using haplotype informative reads. *Bioinformatics* **29**: 2427–2434.

Zhang L, **Zuo K**, **Zhang F**, **Cao Y**, **Wang J**, **Zhang Y**, **Sun X**, **Tang K**. **2006**. Conservation of noncoding microsatellites in plants: implication for gene regulation. *BMC Genomics* **7**: 323.

Zhang X, **Dyachok J**, **Krishnakumar S**, **Smith LG**, **Oppenheimer DG**. **2005a**. *IRREGULAR TRICHOME BRANCH1* in *Arabidopsis* encodes a plant homolog of the Actin-Related Protein2/3 complex activator Scar/WAVE that regulates actin and microtubule organization. *Plant Cell* **17**: 2314–2326.

Zhang X, **Grey PH**, **Krishnakumar S**, **Oppenheimer DG**. **2005b**. The *IRREGULAR TRICHOME BRANCH* loci regulate trichome elongation in *Arabidopsis*. *Plant & Cell Physiology* **46**: 1549–1560.

Zhao QY, **Wang Y**, **Kong YM**, **Luo D**, **Li X**, **Hao P**. **2011**. Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinformatics* **12**: S2.

Zhou Y, **Gao F**, **Liu R**, **Feng J**, **Li H**. **2012**. *De novo* sequencing and analysis of root transcriptome using 454 pyrosequencing to discover putative genes associated with drought tolerance in *Ammopiptanthus mongolicus*. *BMC Genomics* **13**: 266.

Zhu J, **Gong Z**, **Zhang C**, **Song CP**, **Damsz B**, **Inan G**, **Koiwa H**, **Zhu JK**, **Hasegawa PM**, **Bressan RA**. **2002**. OSM1/SYP61: a syntaxin protein in *Arabidopsis* controls abscisic acid–mediated and non-abscisic acid–mediated responses to abiotic stress. *Plant Cell* **14**: 3009–3028.

Zhu J, **Jeong JC**, **Zhu Y**, **Sokolchik I**, **Miyazaki S**, **Zhu JK**, **Hasegawa PM**, **Bohnert HJ**, **Shi H**, **Yun DJ**, **Bressan RA**. **2008a**. Involvement of *Arabidopsis* HOS15 in histone deacetylation and cold tolerance. *Proceedings of the National Academy of Sciences* **105**: 4945–4950.

Zhu XG, **Long SP**, **Ort DR**. **2008b**. What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Current Opinion in Biotechnology* **19**: 153–159.

Zimeri A, **Dhankher O**, **McCaig B**, **Meagher R**. **2005**. The plant MT1 metallothioneins are stabilized by binding cadmiums and are required for cadmium tolerance and accumulation. *Plant Molecular Biology* **58**: 839–855.

Zozomová-Lihová J, **Marhold K**, **Španiel S**. **2014**. Taxonomy and evolutionary history of *Alyssum montanum* (Brassicaceae) and related taxa in southwestern Europe and Morocco: diversification driven by polyploidy, geographic and ecological isolation. *Taxon* **63**: 562–591.