Demographic history of speciation in a *Senecio* altitudinal hybrid zone on Mt. Etna

Dmitry A. Filatov*[1], Owen G. Osborne[1,2] and Alexander S. Papadopoulos[1,3]

[1] Department of Plant Sciences, University of Oxford, United Kingdom

[2] Current address: Department of Life Sciences, Imperial College London

[3] Kew Gardens, London UK

* Corresponding author:

E-mail: dmitry.filatov@plants.ox.ac.uk

Phone: (44)1865 275 051

Fax: (44)1865 275 074

Running title: Ecological speciation in *Senecio*

Key words: speciation, hybrid zone, Senecio, gene flow

Abstract

Hybrid zones typically form as a result of species coming into secondary contact, but can also be established *in situ* as an ecotonal hybrid zone, a situation which has been reported far less frequently. An altitudinal hybrid zone on Mount Etna between two ragwort species (the low elevation *Senecio chrysanthemifolius* and high elevation *S. aethnensis*) could potentially represent either of these possibilities. However, a scenario of secondary contact versus speciation-with-gene-flow has not been explicitly tested. Here we test these alternatives and demonstrate that the data do not support secondary contact. Furthermore, we report that the previous analyses of speciation history of these species were based on admixed populations, which has led to inflated estimates of ongoing, interspecific gene flow. Our new analyses, based on 'pure' *S. aethnensis* and *S. chrysanthemifolius* populations, reveal gene exchange of less than one effective migrant per generation, a level low enough to allow the species to accumulate neutral, genome wide differences. Overall, our results are consistent with a scenario of speciation-with-gene-flow and a divergence time which coincides with the rise of Mt. Etna to altitudes above 2000m (~150KY). Further work to quantify the role of adaptation to contrasting environments of high and low altitudes will be needed to support the scenario of recent ecological speciation in this system.

Introduction

Climatic or geological changes can drastically alter environments and threaten a species' existence, but they can also create new ecological opportunities. Adaptation to newly emerged niches can stimulate populations to diverge and become reproductively isolated (Schluter 2001). This model of "ecological" speciation is well established - drawing on Darwin's concept of new species formation "by means of natural selection". Species that evolve in this way may remain in contact and exchange genetic material, but few cases of clear-cut ecological speciation in the face of gene flow have been characterised (Arnegard *et al.* 2014; Ravinet *et al.* 2016; Schluter 2009; Soria-Carrasco *et al.* 2014).

In situations where the incipient species are not isolated, active sharing of genetic material between the populations opposes their gradual divergence. On the other hand, local adaptation can promote divergence, but only of traits and genes that are directly involved in local adaptation or are in tight physical linkage with such traits/genes. Thus, one may expect that genes under divergent/disruptive selection should be less prone to interspecific gene flow, compared to the rest of the genome, generating a mosaic of divergent and non-divergent regions (Feder *et al.* 2012; Wu 2001). Widespread heterogeneity in genomic differentiation has, indeed, been reported for several species (Poelstra *et al.* 2014; Renaut *et al.* 2012; Turner *et al.* 2005). It remains unclear, however, how such genomic mosaics proceed to the genome-wide divergence that is observed once speciation is complete (Feder & Nosil 2010; Flaxman *et al.* 2012; Via 2012). Situations in which reproductive isolation is not complete facilitate research into this process, as the interaction between selection, genetic divergence and reproductive barriers (both intrinsic and extrinsic) can be studied while divergence is taking place. Well studied examples of ecological speciation despite ongoing gene flow in animals include: marine/freshwater and limnetic/benthic ecotypes of the three-spine stickleback (Arnegard *et al.* 2014); host plant ecotypes in the phytophagous walking stick insect *Timema cristinae* (Soria-Carrasco *et al.* 2014) and divergent ecotypes in different tidal zones in the marine snail *Littorina saxatilis* (Johannesson *et al.* 1995; Ravinet *et al.* 2016). In plants, ecological speciation has been reported in dune/non-dune ecotypes of Helianthus (Andrew & Rieseberg 2013), coastal and inland forms of *Mimulus guttatus* (Lowry & Willis 2010) and in a number of species on Lord Howe Island (Papadopulos *et al.* 2013).

An emerging model of potential ecological speciation with gene flow in plants is the pair of recently diverged diploid (Comes & Abbott 2001) species of *Senecio* on Mount Etna in Sicily. *Senecio aethnensis* Jan ex DC. is found at high altitudes on Mt. Etna (>2000m), whereas *S. chrysanthemifolius* Poir. is largely confined to the volcano's lower slopes (<1000m; (Brennan *et al.* 2009)). Both species are pollinated by generalist insect pollinators and the seeds have parachute-like pappi, increasing the potential for dispersal and hybridisation (Brennan *et al.* 2009). At intermediate altitudes the two species form a stable hybrid zone (Chapman *et al.* 2005; James & Abbott 2005), where phenotypic traits that distinguish the two species show a range of intermediate phenotypes that tracks an altitudinal gradient (Brennan *et al.* 2009; James & Abbott 2005). The main habitats of the species are vastly different and vary in many environmental factors that are correlated with altitude (temperature, solar radiation, proportion of UVB radiation and partial pressures of atmospheric gases) (Korner 2007) and possibly many other variables (e.g. water

availability, soil composition, pollinator availability, herbivore presence, interspecific competition, frost etc (James & Abbott 2005; Ross 2010)). As expected from the physiological constraints imposed by their respective habitats, there is evidence for divergent selection between the species. Phenotypic and molecular cline centres are displaced, indicating a role of selection in cline maintenance and of the different ecological optima for each trait (Brennan *et al.* 2009; Brennan *et al.* 2013). In greenhouse conditions, germination temperature-limits are correlated with the altitude of the source population (Ross *et al.* 2012). The species are interfertile (Chapman *et al.* 2005), but there is evidence of modest levels of reproductive isolation in terms of intrinsic genomic incompatibilities (Brennan *et al.* 2014; Chapman *et al.* 2016), flowering time difference, and environmental selection against immigrants (Brennan *et al.* 2009; Ross 2010).

Previous population genetic reconstructions of the demographic history of *Senecio* speciation on Mt. Etna (Chapman *et al.* 2013; Muir *et al.* 2013; Osborne *et al.* 2013) have consistently reported high interspecific gene flow and relatively recent (<200 KYA) divergence times, regardless of the type of markers used. Strikingly, estimates of divergence time coincide with the timing of the growth of Mt. Etna to the elevations which delineate the species' ranges today (Osborne *et al.* 2013). However, it is possible that these studies underestimated the level of divergence and overestimated levels of gene flow due to the locations at which plant material representing the two species was sampled (Abbott & Brennan 2014). Phenotypically, the lower and higher altitude samples from previous studies (750- 870m and 2036-2471m, respectively) corresponded to 'pure' *S. chrysanthemifolius* and *S. aethnensis* (Figure 1). However, these populations may be admixed at the genetic level. This is most likely to be a problem for the high altitude *S. aethnensis* populations as these were sampled from half a kilometre below the boundary for 'pure' *S. aethnensis* established from previous admixture analysis of 26 RAPD/ISSR markers (Abbott & Brennan 2014; James & Abbott 2005). On the other hand, *S. chrysanthemifolius* populations used in the previous studies come from the proposed 'pure' range of that species. Still, there remains a possibility that sampling from the very top and the very bottom of the species range would lead to lower estimates of interspecific gene flow and higher estimates of the divergence time. Furthermore, it remains unknown whether the two species have always exchanged genes since their divergence, or if there was a period of complete geographic isolation during which speciation took place followed by more recent secondary contact. The previous studies (Chapman *et al.* 2013; Muir *et al.* 2013; Osborne *et al.* 2013) have not explicitly tested for the possibility of secondary contact, implicitly assuming that given the species divergence is very recent, any allopatric phases in their history were very brief and unimportant for their divergence. This is a critical issue as a significant period of allopatry would cast doubt on claims that the Mt. Etna *Senecio* have evolved as a result of ecological speciation. If intrinsic reproductive barriers (Brennan *et al.* 2014; Chapman *et al.* 2016) emerged during an allopatric phase, speciation could have been initiated without any direct influence of divergent selection. Conversely, if speciation did take place in the face of ongoing gene flow, it remains likely that divergent selection has directly lead to the onset of reproductive isolation (Seehausen *et al.* 2014; Smadja & Butlin 2011).

To clarify the demographic history of *S. chrysanthemifolius* and *S. aethnensis* speciation we

collected DNA polymorphism data for the populations of these species from the bottom (<600m) and the very top (>2500m) of the species ranges on Mt. Etna, using the so called “nextRAD” (Russello *et al.* 2015) – a primer based alternative to restriction site-associated DNA sequencing (RADseq) that is less prone to errors introduced by restriction enzyme based protocols, such as RADseq (Arnold *et al.* 2013). We demonstrate that samples from the extremes of the altitudinal cline are indeed less admixed than the *S. chrysanthemifolius* and *S. aethnensis* samples from the altitudes sampled previously (Chapman *et al.* 2013; Muir *et al.* 2013; Osborne *et al.* 2013). The estimation of species demography using these relatively ‘pure’ samples confirms that these species have split relatively recently and have actively exchanged genes ever since, providing additional support for the ecological speciation scenario.

Methods

For population genetic analyses, healthy leaf tissue was collected from two wild population of *S. aethnensis* and two populations of *S. chrysanthemifolius* (Table 1). The later were sampled in the ‘pure’ range of *S. chrysanthemifolius* at the lower end of Mt. Etna altitudinal cline (between 500 and 700 meters above sea mean level). One of the two *S. aethnensis* populations was sampled from the ‘pure’ range of that species at >2500m, while the second population was from lower altitude (~2000m) to match the *S. aethnensis* samples analysed in the previous study (Chapman *et al.* 2013). Upon collection the leaves were photographed and dried in the silica gel for DNA extraction. The photographs of 4 to 8 leaves per plant were taken to measure leaf perimeter and leaf area using ImageJ (Schneider *et al.* 2012). The leaf perimeter to area ratio calculated for each leaf was averaged across all leaves per plant.

Genomic DNA was extracted from dry leaf material using a modified CTAB protocol (Doyle & Doyle 1990), and cleaned using Qiagen DNAeasy mini spin columns. Preparation of nextRAD libraries (Russello *et al.* 2015) and single end Illumina sequencing (150b) was conducted by SNPsaurus (Oregon, USA). Raw reads were processed and aligned using Stacks v1.34 (Catchen *et al.* 2011). The process_radtags program was used to remove reads containing ambiguous base calls (Ns) and low quality reads. To create a reliable *de novo* catalogue of loci/haplotypes, all individuals were analysed with a minimum stack depth of 15 (-m option), distance between stacks of 2 (-M) and distance between loci of 2 (-n), removing highly repetitive stacks. Using this catalogue as a reference, all individuals were then genotyped using a minimum stack depth of 3 with haplotype verification.

SNPs with less than 20% missing genotypes were loaded into ProSeq3 (Filatov 2009) to obtain basic descriptive statistics and to generate input files for downstream analyses in *Structure* (Falush *et al.* 2003), *Arlequin* (Excoffier *et al.* 2005) and *dadi* (Gutenkunst *et al.* 2009). The program *Structure* was used to assess the clustering of individuals according to their population and species of origin. The distribution of genetic variation between different hierarchical levels of population structure was analysed with AMOVA, as implemented in *Arlequin*; statistical significance was evaluated by conducting 10,000 random permutations.

For demographic inference and visualisation of 2-dimensional site frequency spectra (2D-SFS)

we used the *dadi* package (Gutenkunst *et al.* 2009). In the absence of an outgroup to establish the ancestral state for each SNP, we used a 'folded' frequency spectrum for which all SNPs with frequency *x* and *n* - *x* (where x is SNP frequency and n is sample size) were pooled together. The analyses were conducted for all possible pairs of *S. chrysanthemifolius* and *S. aethnensis* populations (Table S1). Following the recommendation in the *dadi* manual, the sample sizes for each of the populations were 'projected down' to include 16 to 20 alleles per population to account for missing data.

Initially we conducted an exploratory analysis using a set of models of different complexity to find the simplest model that adequately describes the data. We started with a relatively parameter rich model, *IMpre*, that allows for population size changes before and after species split. We then progressively excluded free parameters as long as there was no significant reduction in the fit of the model to data. *IMpre* has eight free parameters (Figure 2A): time of population size change before the species split (T_b), population size before the species split (N_b), relative size of populations during species split (s), time of species split (T_s), modern population size of species 1 and 2 (N_1 and N_2) and effective migration rates in two directions (M_1 and M_2). Effective migration rates are equal to m - the fraction of individuals each generation in population 1 that are new migrants from population 2, times the effective population size of the ancestral population (i.e. $M = 2N_am$). Fixing the parameters at specific values allowed us to test demographic hypotheses using likelihood ratio tests. For example, by fixing migration rates at zero, we tested whether non-zero migration improved model fit to the data. Replicate runs with perturbed starting parameters (*perturb_params* function in *dadi*) were used to ensure the analysis reached the global maximum. To evaluate the robustness of the parameter estimates, we used 300 bootstrap replicates of the data, generated by ProSeq3 (Filatov 2009). The model was fitted to each of these replicates, parameters estimated and the confidence intervals calculated as $E \pm 1.96\sigma$, where E is the likelihood parameter estimate and σ is standard deviation for the parameter across the bootstrap replicates. Plotting of observed and modelled 2D-SFSs was done with *plot_2d_comp_multinom* command in *dadi* package. Loglikelihood profiles were generated by varying the parameter of interest while keeping all other parameters in the model at best parameter estimates for the given model. Loglikelihood values for each value of the model parameter of interest were recorded.

In addition to the standard set of *dadi* models we implemented two extensions for the standard *split_migr* model. The first extension, *split_migr2*, allowed for different migration rates in two directions (Figure 2B). This model was used in a likelihood ratio test with the standard *split_migr* model to test for significance of difference in migration rates in two directions. Another extension of the *split_migr* model allowed for varying migration rate over time since the species split. The standard models in *dadi* assume constant migration. To account for the possibility of secondary contact or varying rate of migration over time we implemented an additional model, *splitExpMigr*, that is an extension of the standard *split_migr* model, but allows the migration rate after split to change exponentially over time after species split (Figure 2C). Similar to *split_migr*, *splitExpMigr* model assumes a species (or population) split at time (T_s) into two populations of constant population sizes (N_1 and N_2 for the two species) and equal migration in both directions (parameter

M in *split_migr*). To allow for migration change over time *splitExpMigr* includes two migration parameters, migration immediately after the split (M_{split}) and migration at present ($M_{present}$). Thus, in total *splitExpMigr* includes five parameters: T_s, N_1 and N_2, M_{split} and $M_{present}$. Coalescent simulations for the secondary contact model were conducted in *ms* (Hudson 2002) with the following command line matching the sample sizes populations C1 and A2 (*ms* 38 1000 -t 300.0 -I 2 20 18 0.5 -em 0.1 1 2 0 -ej 1.0 1 2). The simulated samples were loaded in *dadi* and analysed with the *splitExpMigr* model.

The demographic history of the two species was further assessed using the composite-likelihood approach implemented in *fasctsimcoal2* (Excoffier *et al.* 2013). To find the global maximum, parameters values were estimated from the 2D-SFS 50 times for each of ten models and the best fitting model was determined from the Akaike Information Criterion (AIC) for the set of parameter values with the best fit for each model (Table S2). For the best-fitting model, 95% confidence intervals were estimated by non-parametric bootstrapping (100 simulated datasets, with 10 rounds of parameter estimation for each simulation). Models were calibrated using the *Asteraceae* per generation mutation rate of 1.0×10^{-8} mutations per site (Strasburg & Rieseberg, 2008). Three sets of models were assessed. The simplest model (*DivNoM*) has 4 parameters and describes an ancestral population of size N_a that splits into two descendent populations of size N_1 and N_2 at time T_0 with no migration. *DivNoM* was extended to an isolation-with-migration model (*IM*) by allowing asymmetrical migration between the descendent populations following divergence (6 parameters). Three further models were developed form the basic IM model by including two periods of differing levels of migration between the descendent populations: The *IM2R8S* model (Figure 2D) permitted different migration rates in two time periods (T_0 - T_1 and T_1 – Present; 9 parameters in total); The *IMSTOPS* model allowed migration between T_0 and T_1 but no migration between T_1 and the present (7 parameters in total); and the *IML8M* model did not allow migration between T_0 and T_1, but migration was permitted between T_1 and the present (i.e. a secondary contact model; 7 parameters in total). These five models were then extended by adding two extra parameters to each model to allow exponential population size change in each of the species between T_0 and the present. This brought the total number of models we implemented and analysed in *fasctsimcoal2* to ten (Table S2), with five of the models assuming constant population size (*DivNoM*, *IM*, *IM2R8S*, *IMSTOPS*, *IML8M*) and five models allowing exponential population size change following species split (*DivNoM_GR*, *IM_GR*, *IM2R8S_GR*, *IMSTOPS_GR*, *IML8M_GR*).

Results

To investigate *S. aethnensis* and *S. chrysanthemifolius* speciation and hybridisation, we analysed phenotypic and molecular diversity in two low altitude and two high altitude populations of *Senecio* from Mt. Etna (Table 1) and explored the evolutionary history of the species using demographic modelling.

Phenotypes

The analysis of phenotypic differences between the *Senecio* populations was focused on the most

conspicuous trait that varies along the altitudinal *Senecio* gradient on Mt Etna – leaf shape. The 'pure' *S. aethnensis* and *S. chrysanthemifolius* plants have non-dissected and highly dissected leaves, respectively (Figure 1). Plants from the hybrid zone at intermediate elevation display a range of intermediate phenotypes (Brennan *et al.* 2012; Brennan *et al.* 2009). We measured the leaf perimeter to leaf area ratio to ensure that the measure of leaf serration is normalised by leaf size. As expected, this ratio was much higher in the two low altitude populations that have more dissected leaves, compared with the high altitude samples (Table 1). All pairwise population comparisons demonstrate significant differences in leaf shape both within and between the species (*F*-tests, $P < 0.001$). Consistent with the previous reports of a phenotypic cline in leaf shape (Brennan *et al.* 2009), the lowest *S. chrysanthemifolius* population (C1, Table 1) had most dissected leaves of all, but contrary to the cline expectation, the *S. aethnensis* population growing at around 2000m (A1, Table 1) had lower perimeter to area ratio than the samples of that species from above 2500m (population A2, Table 1). However, the differences between the two sampled populations within each species are very small, compared with the interspecific differences.

Genetic polymorphism and differentiation

Overall we generated 72.9 million reads distributed among 200,657 RAD tags; per individual coverage ranged from 23.8 to 118.3. After single nucleotide polymorphism (SNP) calling and filtering (see methods) we included 6726 SNPs. Between 2012 and 2831 of the SNPs were polymorphic within each population (Table 1). The overall level of sequence diversity was higher in *S. aethnensis* populations (Table 1), consistent with previous reports (Chapman *et al.* 2013).

The majority of DNA diversity (41.9%) was distributed within individuals (heterozygosity), which is unsurprising given that a self-incompatibility system operates in *Senecio* (Hiscock 2000). Divergence between the two species is the second largest source of overall diversity (40.4%), while the contribution of inter-population differences within each species is very low (2.1%; Table 2). Accordingly, population differentiation analysis also revealed low levels of population differentiation within each species ($F_{st} < 0.05$; Table 3). Given the high level of hybridisation at intermediate altitudes, the finding that both population differentiation (F_{st}) and net sequence divergence (D_a) were considerably higher for interspecific pairwise population comparisons was unexpected (Table 3). All pairwise population comparisons within and between the species were significant regardless of the statistic used ($P < 0.001$). While the sampled individuals readily clustered according to their species assignment (Figure 3A), no clear clustering by population of origin was observed within the species (Figure 3B and C). Our *Structure* results were consistent with previous delimitation of 'pure' and admixed populations of the two species (Brennan *et al.* 2009; James & Abbott 2005): both *S. chrysanthemifolius* populations and the higher altitude *S. aethnensis* population (A2) were almost completely free of admixed individuals, while among the individuals in the lower *S. aethnensis* population (A1) 90% displayed minor admixture (Figure 3A).

Demographic modelling

To reconstruct the demographic changes that have occurred during the speciation of *S. aethnensis* and *S. chrysanthemifolius* and infer the timing of the divergence event we explored models of

population splitting with migration (Figure 2), as implemented in *dadi* (Gutenkunst *et al.* 2009) and *fastsimcoal2* (Excoffier *et al.* 2013). This method compares the observed site frequency spectrum (SFS) to that expected given a specific population split demographic model (Figure 4) in order to estimate model parameters and assess the likelihood of the observed data given the model. The analyses were focused on for the 'pure' populations from the extremes of the altitudinal cline (C1 and A2; Table 4). Due to computational restrictions only *dadi* was used for testing all other possible pairs of *S. aethnensis* and *S. chrysanthemifolius* populations (Table S1; Figure S1).

The most complex *dadi* model applied here (*IMpre*) included eight free parameters and allowed for population size change, both before and after divergence, as well as permitting different migration rates for each direction of migration (Figure 2A). This model fitted the data well, except a slight excess of low-frequency polymorphisms in the observed frequency spectrum of both species, compared to that predicted by the model (Figure 4). Using likelihood ratio tests on the nested models, the analyses revealed that a much simpler species split model with no population size changes fits the data equally well for the populations from the highest (A2) and lowest (C1) elevations (Table 4). Furthermore, separate migration rates do not differ to a great extent (Figure 5). An alternative model with a single migration parameter (*split_mig*) does not have noticeably lower likelihood than the models with two migration parameters (Table 4). The estimated time of species split (T = 1.3 bootstrap CI: 0.07 to 2.24) is expressed in units of $2N_a$ generations, where N_a is ancestral population size. To convert this to years one needs to assume a plausible value of N_a and the number of generations per year. Assuming N_a = 100,000 and 0.5 generations per year (see discussion), the two species diverged 128.3 (bootstrap CI: 7.6 to 249.1) thousand years, which roughly falls in the range of the estimates reported previously (Chapman *et al.* 2013; Muir *et al.* 2013; Osborne *et al.* 2013). *IM* model implemented in *fastsimcoal2* model is similar to the *dadi split_migr2* model, except that *IM* includes ancestral population size as a parameter and *fastsimcoal2* estimates N_a = 33,989 (Table S2). The best-fitting model, *IM2R8S*, estimates N_a = 413,191 (Table 5), but the confidence interval is fairly wide (95%CI: 122,827 to 465,594) and it is clear that estimates of this parameter are highly variable between the models (Table S2), making it problematic to use *fastsimcoal2* N_a estimates as a basis for rescaling parameters in *dadi* models.

Parameter estimates in *dadi* models did vary for other population pairs, but for all model types and populations used, fixing gene flow parameters to zero significantly reduced the fit of the models to data (Table S1). This confirms that gene flow between *S. aethnensis* and *S. chrysanthemifolius* has been a feature of their evolution. Using the populations from the highest and lowest elevations (A2 and C1) reveals modest level of interspecific introgression (M = 0.3 bootstrap CI: 0.19 to 0.41), while using lower altitude *S. aethnensis* population (A1) with either of the two *S. chrysanthemifolius* populations leads to higher gene flow estimates (Table S1). On the other hand, the estimated timing of speciation was robust to the choice of the populations for analysis and consistently points to a relatively recent origin of these species (~$2.5N_a$ generations ago Table S1). The estimate of species split time in the best fitting *IM2R8S fastsimcoal2* model is consistent with this conclusion (T_s = 126,407 generations ago, 95%CI: 102,642 to 306,413).

To test the alternative scenario of allopatric divergence followed by more recent hybrid zone

formation and gene flow, we implemented a *splitExpMigr* model in *dadi*, where the migration rate is allowed to vary exponentially over time (Figure 2C). The *splitExpMigr* model includes two migration parameters that describe the migration rate directly after species split (M_{split}) and the migration rate in the present ($M_{present}$). If hybridisation between two species started recently, then the estimate of M_{split} is expected to be much lower than that for $M_{present}$. However, the estimates for the two migration parameters are very similar (M_{split} = 0.25 bootstrap CI: 0 to 0.545; $M_{present}$ = 0.30 bootstrap CI: 0.221 to 0.446) and the exponential migration model does not have a better fit than the model with a single migration parameter for the populations from the highest and lowest elevations (A2 and C1; Table 4 and Figure 6A). This lack of support for a period of allopatry following the species split does not appear to be due to lack of power, as the log-likelihood drops rapidly for lower values of M_{split} (Figure 6A). Furthermore, this model was capable of detecting recent secondary contact based on simulated data (Figure 6B). In addition to *splitExpMigr* model, we implemented a set of *fastsimcoal2* models allowing for a step-wise change in migration rate. The best-fitting of these models, *IM2R8S*, includes four migration parameters to account for different migration rates in two directions before and after the step-wise migration change (Figure 2D). That model indicates relatively high migration following speciation ~126k generations ago, and a sharp drop in migration later on (Table 5), which is the opposite to the expectations of the secondary contact scenario and is consistent with a scenario of ecological speciation with gene flow.

Discussion

Our results are consistent with the divergence of *Senecio* on Mt. Etna as a recent instance of speciation with gene flow along an elevational gradient. Demographic analysis of populations of each species from the extremes of the distribution, suggest that gene flow across the hybrid zone has been less extensive than previously thought and is likely to have decreased substantially since divergence took place. Nevertheless, our results confirm that there is likely to have been continuous genetic exchange and migration throughout their evolutionary history. Current levels of migration are sufficiently low to permit neutral differentiation of the species, suggesting the evolution of reproductive isolation in the face of high gene flow between the species may have been a rapid process. Given the weight of evidence implicating divergent selection in the evolution of RI between the species it remains highly likely that this represents one of the few well-documented cases of ecological speciation with gene flow in plants.

Geographically, *S. aethnensis* is endemic to the upper slopes of Mount Etna, whereas *S. chrysanthemifolius* is found only on or around Mount Etna (Alexander 1979). As might be expected from the species' geographic context, they have been found to be sister species using multiple phylogenetic approaches based on genome-wide datasets (Osborne *et al*., 2016). Furthermore, there is both phenotypic (Brennan *et al.* 2009; Ross *et al.* 2012), and molecular (Brennan *et al.* 2009; Chapman *et al.* 2013; Muir *et al.* 2013) evidence for divergent selection between the species. Previous work had suggested that, despite the species' considerable phenotypic divergence, they diverged recently, in concert with the growth of Mount Etna, and that there had been gene flow since their split (Chapman *et al.* 2013; Muir *et al.* 2013; Osborne

et al. 2013). The demographic analyses conducted here are consistent with continuous gene flow throughout the evolutionary history of the species, eliminating the possibility of an allopatric phase. Importantly, the simulation analyses demonstrate that it is possible to discriminate between secondary contact and speciation with gene flow events using this method. Recent demographic research has called into question "classic" cases of sympatric/parapatric speciation in cichlid fish (Martin *et al.* 2015). The current results provide valuable support for speciation with gene flow in plants beyond those described from Lord Howe Island (Papadopulos *et al.* 2011).

There are several possible reasons for the differences in demographic analyses of *S. aethnensis* and *S. chrysanthemifolius* speciation. Our study used the same analytical approach as in (Chapman *et al.* 2013), so the opposing results from the two studies must stem from the differences in the data. The results presented here indicate that the use of admixed *S. aethnensis* populations from below the 'pure' range of that species is very likely to be the major source of inflated estimates of interspecific gene flow in previous work (Chapman *et al.* 2013; Muir *et al.* 2013; Osborne *et al.* 2013). However, our sampling of *S. aethnensis* population A1 from the same altitude as before (~2000m) revealed a very different frequency spectrum to that observed in (Chapman *et al.* 2013). In particular, the 2D-SFSs in our data (Figure 4 and S1) show a higher proportion of SNPs that are polymorphic in only one of the species and fixed in the other (irrespective of the populations included). Thus, sampling from a lower altitude does not fully account for the differences in the data. A second notable sampling difference is that (Chapman *et al.* 2013) sampled fewer individuals from multiple populations, while the current study focused on deeper sampling from fewer populations. This suggests that the 2D-SFS in the previous study reflects species-wide differences, while the spectra reported here incorporate the differences between the populations in addition to species differences. Given the very low levels of subdivision between the populations within each of the species (Tables 2 and 3), this is unlikely to significantly bias the analyses.

The two studies also used different kinds of high-throughput sequence data. Chapman et al (2013) used transcriptome sequencing (RNA-seq) from multiple individuals, while our study employed 'nextRAD', and sampled more individuals. Both collect polymorphism data from a fraction of the genome, but the two approaches differ with regard to what fraction of the genome they sample. RAD (or nextRAD) yields SNPs from random genomic regions, most of which are likely to be in non-coding regions, whereas, SNPs identified and genotyped with RNA-seq are necessarily in expressed regions in or next to protein coding genes. This may make transcriptome data more prone to effects of direct or indirect selection (e.g. hitchhiking with adaptive mutations in adjacent coding sequences) than RAD data. It is unclear to what extent selection obscured the results of demographic inference in the RNA- seq-based dataset, but this effect is likely to be relatively minor, as only 4-fold degenerate sites were used in the previous study to minimise the effects of selection (Chapman *et al.* 2013). Furthermore, non-random sampling of genomic regions by different types of markers may be another source of discrepancies. RNA-seq SNPs are more likely to be over-represented in gene-rich regions, while nextRAD-based SNPs are likely to reflect an overall genome-wide species differentiation. Chapman *et al* (2016) demonstrated that genes with high

interspecific differentiation cluster in the genetic map of *aethnensis* and *S. chrysanthemifolius*, though the actual physical size of these regions is unclear. If recombination in these differentiated regions is low (e.g. in pericentromeric regions or inversions), they would appear to be small in the genetic map, but actually represent a significant part of the genome. If this were the case, many of our nextRAD SNPs could come from gene-poor genomic regions with high interspecific differentiation, while RNA-seq-based SNPs used by Chapman *et al* (2013) would reflect lower species differentiation in the actively recombining gene-rich regions. Non- (or rarely) recombining regions tend to be gene poor and they are expected to show higher interspecific differentiation (Nachman & Payseur 2012). This could result in significant underestimation of genome-wide species divergence with SNPs genotyped by RNA-seq. Our current study, based on nextRAD, is free from this potential bias.

Consistent with this, our analyses of genetic polymorphism in *S. aethnensis* and *S. chrysanthemifolius* show that the species have accumulated substantial differences in allele frequencies. In particular, about 80% of SNPs were polymorphic only in one of the populations (the very left and very bottom rows of the observed frequency spectrum on figure 4). Such a degree of divergence is not expected if the two species are actively exchanging genes. For example, with more than three effective migrants per generation, the 2D-SFS converges to that expected under the standard neutral model with no isolation between the populations (Figure 7). Thus, if gene flow between *S. aethnensis* and *S. chrysanthemifolius* were as high as reported previously (*M* ranging from 3 to 15; (Chapman *et al.* 2013), diversifying selection throughout the genome would be necessary to maintain the allele frequency differences reported in our study. Although it is theoretically possible, diversifying selection at most SNPs analysed in our study seems highly improbable.

The demographic modelling framework employed in this project estimates the parameters scaled by the ancestral population size, N_a, which is unknown (e.g., the time of divergence, *T*, is given in units of $2N_a$ generations). Thus, the estimate of speciation time in years depends on an assumption of the ancestral population size and average generation time. As our models do not provide support for significant population size changes in *S. aethnensis* and *S. chrysanthemifolius* or their ancestor, we assumed that N_a must be of similar size to the effective population sizes at present. Given that *S. aethnensis* is endemic to high altitudes on Mt. Etna, present population size of this species is unlikely to be large. Although *S. chrysanthemifolius* is more widespread than *S. aethnensis*, its population size appears slightly smaller than that of *S. aethnensis* (Tables 4 and 5). We assumed N_a = 100,000, which is different to the assumption of N_a = 300,000 in the previous study (Chapman *et al.* 2013) but both fall within or close to the 95% CI estimated by *fastsimcoal2* with the best fitting model. However, the previous study had to adjust for significant population size reduction prior to species split and our estimates of the current population sizes are similar to that in Chapman et al (2013). Given these assumptions, we estimate that *S. aethnensis* and *S. chrysanthemifolius* have diverged around 128 thousand years ago, which is compatible with the estimates reported previously (Chapman *et al.* 2013; Muir *et al.* 2013; Osborne *et al.* 2013). Assuming higher or lower N_a will necessarily yield higher or lower time of species divergence. As

there is no way to obtain a more certain value for the ancestral population size, our estimate of speciation time has to be taken with caution. However, the estimate of species split time in the best fitting *IM2R8S fastsimcoal2* model is close to our estimates from *dadi* analyses and is consistent with the previously expressed hypothesis that *S. aethnensis* and *S. chrysanthemifolius* speciated at the time when the height of Mt. Etna has risen above 2km, reaching the altitudes inhabited by modern day *S. aethnensis* (Osborne *et al* 2013).

It now seems highly likely that the Mt. Etna *Senecio* species are an example of recent ecological speciation with gene flow. The system has many unique qualities making it an indispensable model for speciation research. Much of the recent scientific discourse regarding ecological speciation has concerned the existence and nature of genomic "islands of speciation". These are regions of the genome that are hypothesised to contain divergently selected loci and, as a consequence, are expected to have higher divergence and lower effective gene flow than the rest of the genome. However, some authors have argued that highly differentiated regions could result from selection following divergence rather than a local reduction in gene flow caused by divergent selection, but the demographic histories of the species are often poorly understood (Cruickshank & Hahn 2014; Noor & Bennett 2009; Turner & Hahn 2010). Lying across a steep ecological cline, the well characterised hybrid zone of *S. aethnensis* and *S. chrysanthemifolius* provides opportunities to understand the genomic changes that take place during speciation. The cline is likely to infer strong divergent selection as elevational clines are correlated with many biologically important environmental characteristics (Korner 2007). By sampling across the cline, geographic and genomic cline based approaches can be used which measure gene flow at individual loci far more directly than by statistics such as F_{ST}, which are also affected by factors other than gene flow (Barton & Hewitt 1985; Gompert & Buerkle 2011).

Linking ecological divergence to the emergence of reproductive isolation is a vital next step in determining the underlying causes of speciation in this system. Recent work has found evidence of a low level of intrinsic reproductive isolation between them (Brennan *et al.* 2014) and understanding whether this evolved as a direct consequence of ecological divergence or intrinsic reproductive (e.g. spontaneous mutation) barriers prior to ecological divergence will be crucial. If regions of the genome harbouring genomic incompatibilities also contain control ecologically important traits and/or genomic signatures of divergent selection a link between ecological divergence and reproductive isolation would be established. To this end, two significant obstacles have been overcome as a result of the research presented here. First, the sampling regime that had been called into question (Abbott & Brennan 2014) has been rectified. Second, the models applied have differentiated between a scenario of continuous gene flow since divergence as opposed to a period of isolation followed by secondary contact. In substantially bolstering the case for ecological speciation, we have laid the foundations for future research into the genomic basis of ecological speciation and the development of *S. aethnensis* and *S. chrysanthemifolius* as a major study system for evolutionary biology.

Acknowledgements

We thank Mark Chapman for advice and colleagues at SNPsaurus (Oregon, USA) for generating nextRAD data for this study. This work was supported by funding to DAF from NERC (NE/K004352/1) and a Sainsbury PhD studentship award to OGO from the Gatsby Charitable Foundation.

References

Abbott RJ, Brennan AC (2014) Altitudinal gradients, plant hybrid zones and evolutionary novelty. *Philos Trans R Soc Lond B Biol Sci* **369**. pii: 20130346.

Alexander JCM (1979) Mediterranean species of *Senecio* sections *Senecio* and *Delphinifolius*. *Notes from the Royal Botanic Garden Edinburgh* **37**, 387-428.

Andrew RL, Rieseberg LH (2013) Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. *Evolution* **67**, 2468-2482.

Arnegard ME, McGee MD, Matthews B*, et al.* (2014) Genetics of ecological divergence during speciation. *Nature* **511**, 307-311.

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Mol Ecol* **22**, 3179- 3190.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Ann Rev Ecol Syst* **16**, 113-148.

Brennan AC, Barker D, Hiscock SJ, Abbott RJ (2012) Molecular genetic and quantitative trait divergence associated with recent homoploid hybrid speciation: a study of *Senecio squalidus* (Asteraceae). *Heredity (Edinb)* **108**, 87-95.

Brennan AC, Bridle JR, Wang AL, Hiscock SJ, Abbott RJ (2009) Adaptation and selection in the *Senecio* (*Asteraceae*) hybrid zone on Mount Etna, Sicily. *New Phytol* **183**, 702-717.

Brennan AC, Harris SA, Hiscock SJ (2013) The population genetics of sporophytic self-incompatibility in three hybridizing *Senecio* (*Asteraceae*) species with contrasting population histories. *Evolution* **67**, 1347-1367.

Brennan AC, Hiscock SJ, Abbott RJ (2014) Interspecific crossing and genetic mapping reveal intrinsic genomic incompatibility between two *Senecio* species that form a hybrid zone on Mount Etna, Sicily. *Heredity (Edinb)* **113**, 195-204.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping Loci de novo from short-read sequences. *G3 (Bethesda)* **1**, 171-182.

Chapman MA, Forbes DG, Abbott RJ (2005) Pollen competition among two species of Senecio (Asteraceae) that form a hybrid zone on Mt. Etna, Sicily. *Am J Bot* **92**, 730-735.

Chapman MA, Hiscock SJ, Filatov DA (2013) Genomic divergence during speciation driven by adaptation to altitude. *Mol Biol Evol* **30**, 2553-2567.

Chapman MA, Hiscock SJ, Filatov DA (2015) The genomic bases of morphological divergence and reproductive isolation driven by ecological speciation in *Senecio* (*Asteraceae*). *J Evol Biol* **29**, 98-113.

Comes HP, Abbott RJ (2001) Molecular phylogeography, reticulation, and lineage sorting in Mediterranean *Senecio* sect. *Senecio* (*Asteraceae*). *Evolution* **55**, 1943-1962.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* **23**, 3133-3157.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* **12**, 13-15.

Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genet* **9**, e1003905.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online* **1**, 47-50.

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.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends Genet* **28**, 342- 350.

Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* **64**, 1729-1747.

Filatov DA (2009) Processing and population genetic analysis of multigenic datasets with ProSeq3 software. *Bioinformatics* **25**, 3189-3190.

Flaxman SM, Feder JL, Nosil P (2012) Spatially explicit models of divergence and genome hitchhiking. *J Evol Biol* **25**, 2633-2650.

Gompert Z, Buerkle CA (2011) Bayesian estimation of genomic clines. *Mol Ecol* **20**, 2111-2127.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* **5**, e1000695.

Hiscock SJ (2000) Self-incompatibility in *Senecio squalidus* (*Asteraceae*). *Ann Bot* **85**, 181-190.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337-338.

James JK, Abbott RJ (2005) Recent, allopatric, homoploid hybrid speciation: the origin of *Senecio squalidus* (*Asteraceae*) in the British Isles from a hybrid zone on Mount Etna, Sicily. *Evolution* **59**, 2533-2547.

Johannesson K, Johannesson B, Lundgren U (1995) Strong natural selection causes microscale allozyme variation in a marine snail. *Proc Natl Acad Sci U S A* **92**, 2602-2606.

Korner C (2007) The use of 'altitude' in ecological research. *Trends Ecol Evol* **22**, 569-574.

Lowry DB, Willis JH (2010) A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol* **8**, pii: e1000500.

Martin CH, Cutler JS, Friel JP*, et al.* (2015) Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. *Evolution* **69**, 1406-1422.

Muir G, Osborne OG, Sarasa-Marcuello J, Hiscock SJ, Filatov DA (2013) Recent ecological speciation in *Senecio*. *Evolution* **67**, 3032-3042.

Nachman MW, Payseur BA (2012) Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philos Trans R Soc Lond B Biol Sci* **367**, 409-421.

Noor MA, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity (Edinb)* **103**, 439-444.

Osborne OG, Batstone TE, Hiscock SJ, Filatov DA (2013) Rapid speciation with gene flow following the formation of Mt. Etna. *Genome Biol Evol* **5**, 1704-1715.

Osborne OG, Chapman MA, Nevado B, Filatov DA (2016) Maintenance of species boundaries despite ongoing gene flow in ragworts. *Genome Biol Evol*, accepted.

Papadopulos AS, Baker WJ, Crayn D*, et al.* (2011) Speciation with gene flow on Lord Howe Island. *Proc Natl Acad Sci U S A* **108**, 13188-13193.

Papadopulos AS, Price Z, Devaux C*, et al.* (2013) A comparative analysis of the mechanisms underlying speciation on Lord Howe Island. *J Evol Biol* **26**, 733-745.

Poelstra JW, Vijay N, Bossu CM*, et al.* (2014) The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344**, 1410-1414.

Ravinet M, Westram A, Johannesson K*, et al.* (2016) Shared and non-shared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Mol Ecol* **25**, 287-305.

Renaut S, Maillet N, Normandeau E*, et al.* (2012) Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philos Trans R Soc Lond B Biol Sci* **367**, 354-363.

Ross RI (2010) *Local adaptation and adaptive divergence in a hybrid species complex in Senecio*. DPhil thesis, University of Oxford.

Ross RI, Agren JA, Pannell JR (2012) Exogenous selection shapes germination behaviour and seedling traits of populations at different altitudes in a *Senecio* hybrid zone. *Ann Bot* **110**, 1439-1447.

Russello MA, Waterhouse MD, Etter PD, Johnson EA (2015) From promise to practice: pairing non-invasive sampling with genomics in conservation. *PeerJ* **3**, e1106.

Schluter D (2001) Ecology and the origin of species. *Trends Ecol Evol* **16**, 372-380.

Schluter D (2009) Evidence for ecological speciation and its alternative. *Science* **323**, 737-741.

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675.

Seehausen O, Butlin RK, Keller I*, et al.* (2014) Genomics and the origin of species. *Nat Rev Genet* **15**, 176-192.

Smadja CM, Butlin RK (2011) A framework for comparing processes of speciation in the presence of gene flow. *Mol Ecol* **20**, 5123-5140.

Soria-Carrasco V, Gompert Z, Comeault AA*, et al.* (2014) Stick insect genomes reveal natural selection's role in parallel speciation. *Science* **344**, 738-742.

Turner TL, Hahn MW (2010) Genomic islands of speciation or genomic islands and

speciation? *Mol Ecol* **19**, 848-850.

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol* **3**, e285.

Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos Trans R Soc Lond B Biol Sci* **367**, 451-460.

Watterson G (1975) On the number of segregating sites in genetic models without recombination. *Theor Popul Biol* **7**, 256-276.

Wu C-I (2001) The genic view of the process of speciation. *J Evol Biol* **14**, 851-865.

Data Accessibility:

All sequences were submitted to genbank under project accession numbers PRJNA313194 and PRJNA313255.

Table 1. Summary of population locations and diversity

Sample ID	Species	Elevation (meters)	Location		Sample size	Polym.[1)] SNPs	Watterson's θ ± SD [2)]		Leaf shape [3)]
			Latitude	Longitude			per sequence	per 1 kb	
C1	*S. chrys.*	587.5	N37.5951	E14.9793	14	2012	632.68 ± 14.105	1.1 ± 0.44	0.9991 ± 0.20652
C2	*S. chrys.*	706.3	N37.6161	E15.0320	12	2487	823.54 ± 16.514	1.5 ± 0.52	0.8528 ± 0.26723
A1	*S. aeth.*	2090.8	N37.7072	E15.0081	10	2221	785.09 ± 16.659	1.4 ± 0.52	0.2164 ± 0.03792
A2	*S. aeth.*	2636.5	N37.7257	E15.0038	16	2831	853.17 ± 16.035	1.5 ± 0.50	0.2906 ± 0.08632

[1)] Out of 560,850 alignment positions analysed 6,726 positions were polymorphic in the entire dataset of 52 individuals

[2)] Watterson's estimator of population-scaled mutation rate, θ (Watterson 1975) and its standard deviation (SD) calculated assuming free recombination

[3)] Leaf shape is expressed as ratio of perimeter to area (±SD)

Table 2. Distribution of molecular variation at the individual, population and species levels

Source of variation	df	Sum of squares	Variance components	Percentage of variation
Among species	1	9332.577	169.03091	40.41
Among populations within species	2	1058.66	8.86566	2.12
Among individuals within populations	48	14670.84	65.26356	15.6
Within individuals	52	9106	175.11538	41.87
Total	103	34168.077	418.27552	

Table 3. Pairwise distances between the populations. D_a is above and F_{ST} is below the diagonal. All pairwise population comparisons revealed significant ($P < 0.001$) population differentiation for both statistics.

	S. chrysanthemifolius		*S. aethnensis*	
Populations	**C1**	**C2**	**A1**	**A2**
C1	-	21.03	294.38	376.02
C2	0.047	-	325.627	413.41
A1	0.397	0.383	-	24.34
A2	0.461	0.454	0.046	-

Table 4. Estimates of species split parameters and likelihood ratio tests for *dadi* analyses in C1 and A2 population pair.

Model	log-likelihood	LRT[1] 2ΔLL	LRT[1] signif.?	free params	N_b [2]	T_b [3]	s [4]	N_{chr} [5]	N_{aet} [5]	T_s [3]	M_1 [6] (chr<=aet)	M_2 [6] (chr=>aet)
IMpre	-428.76			8	1.39	0.29	0.42	0.68	0.99	1.47	0.31	0.26
IM	-429.44	1.36	NS	6	1*	1*	0.40	0.57	0.84	1.03	0.38	0.30
*IM**	-429.71	0.55	NS	5	1*	1*	0.5*	0.51	0.80	0.80	0.35	0.33
IMnoMigr	-540.04	220.65	P<0.00001	3	1*	1*	0.5*	0.35	0.61	0.29	0*	0*
split_migr2	-428.59			5	no	no	no	0.61	0.87	1.26	0.33	0.28
split_migr	-429.04	0.90	NS	4	no	no	no	0.63	0.86	1.28	$M_1 = M_2 = 0.30$	
											M_{split} [7]	$M_{present}$ [7]
splitExpMigr	-428.94			5	no	no	no	0.65	0.88	1.30	0.25	0.30
*splitExpMigr**	-437.65	17.42	P<0.001	4	no	no	no	0.49	0.68	0.55	0.0001*	0.80

[1] In all cases LRTs are for the comparison with the model immediately above.

[2] N_b, population size before the species split, scaled in units of N_{anc}

[3] The time of of species split (T_s) and population size change before the species split (T_b), scaled in units of $2N_{anc}$ generations

[4] s relative size of populations during species split

[5] N_{chr} and N_{aet} are effective population sizes of the *S. chrysanthemifolius* and *S. aethnensis* at present, scaled in units of N_{anc}

[6] M_1 and M_2 are population scaled migration rates ($M = 2N_{anc}m$) from *S. aethnensis* to *S. chrysanthemifolius* and in opposite direction.

[7] For *splitExpMigr* models the migration parameters are M_{split} and $M_{present}$ (population scaled migration at the time of species split and at present, respectively) in the M_1 and M_2 columns, respectively.

Table 5. Parameter estimates for the best fitting model (*IM2R8S*) in *fastsimcoal2* analyses.

Parameters	Estimate	Lower 95% CI	Upper 95% CI	Notes
N_1	12770	11398	14039	Present effective population size of *S. chrysanthemifolius*
N_2	18531	16482	19925	Present effective population size of *S. aethnensis*
N_{anc}	413191	122827	465594	Ancestral effective population size
T_0	126407	102642	306413	Time of species split
T_1	60701	37432	83149	Time of step-wise migration rate change
M_1	0.24	0.18	0.29	*S.chrys*. => *S.aet*. population migration rate ($2Nm$) at time T_1
M_2	0.21	0.16	0.24	*S.chrys*. <= *S.aet*. population migration rate ($2Nm$) at time T_1
M_3	20.68	2.36	34.37	*S.chrys*. => *S.aet*. population migration rate ($2Nm$) before T_1
M_4	1.63	1.10	28.43	*S.chrys*. <= *S.aet*. population migration rate ($2Nm$) before T_1

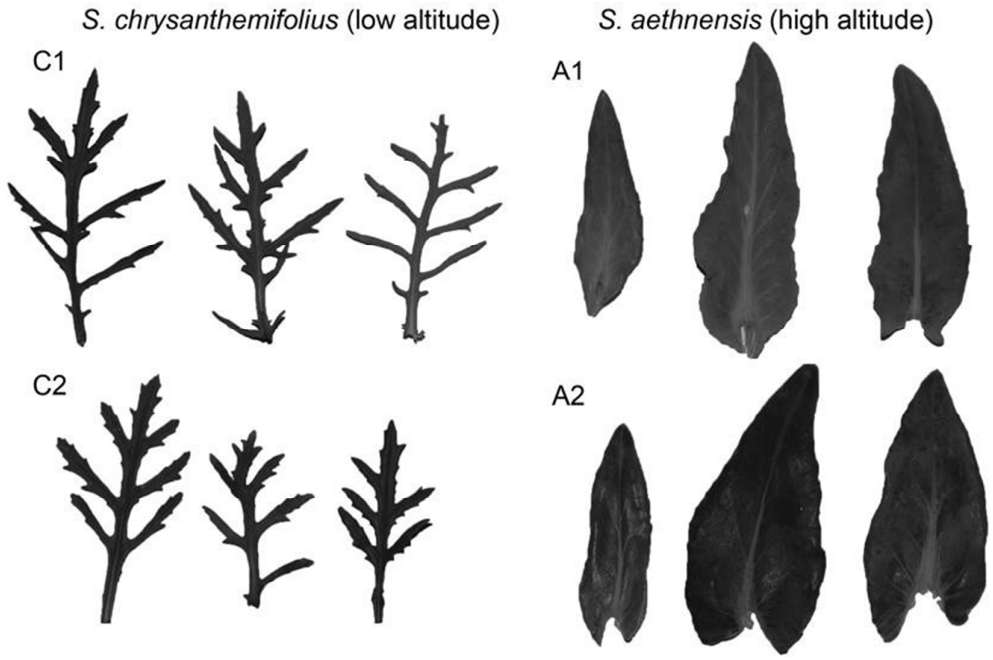


Figure 1. Leaf shapes in high and low altitude populations of *Senecio* on Mt. Etna. Three randomly chosen leaves are shown for each of the two *S. chrysanthemifolius* (C1 and C2) and two *S. aethnensis* (A1 and A2) populations.

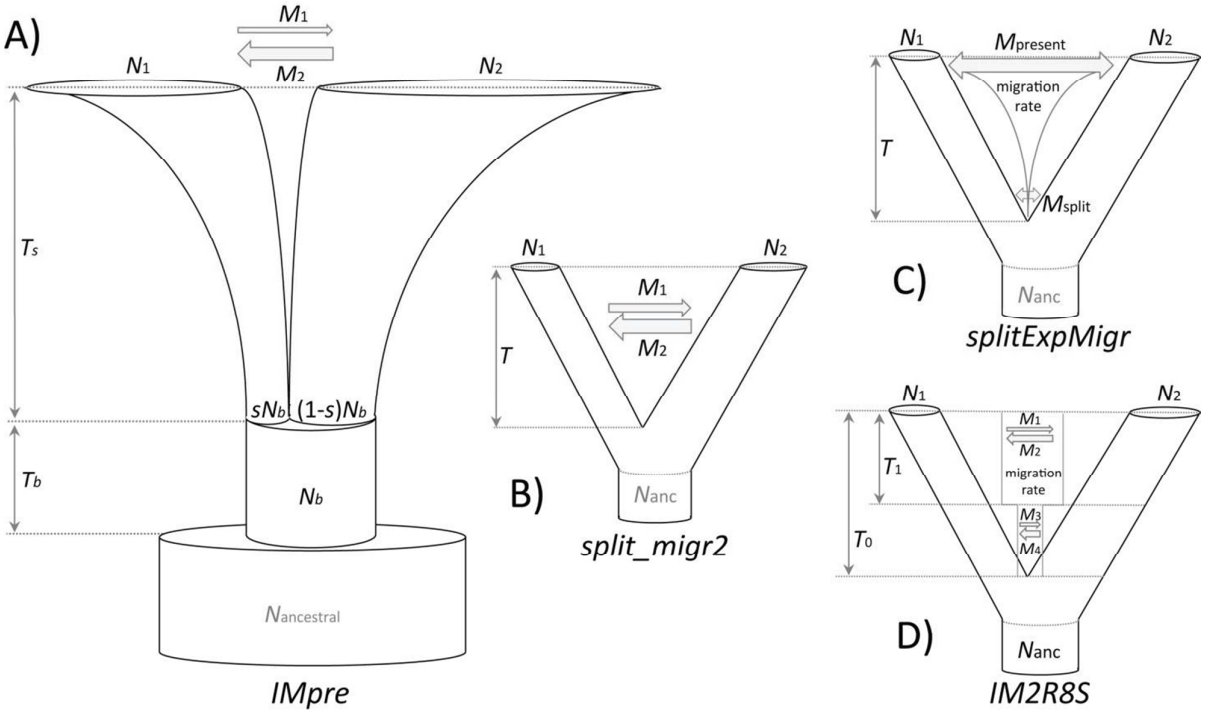


Figure 2. Schematic representation of demographic models used. A to C panels represent the models implemented in *dadi*, while panel D shows the best fitting model in *fastsimcoal2* analyses. The model names are given below each tree; horizontal arrows show the direction of gene flow and the width of the tree branch is proportional to the population size. Every model lists free parameters in black italic font. In *dadi* models the parameters are expressed in units of ancestral population size ($N_{ancestral}$, or N_{anc} on panels A-C) and N_{anc} is not estimated in these models (shown in grey). A) The most parameter rich *dadi* model, *IMpre*, includes eight parameters. Going forward in time, the ancestral population undergoes population size change (N_b) at time T_b and then splits into two species with sizes sN_b and $(1\text{-}s)N_b$, which exponentially grow (or decline) in size to present population sizes (N_1 and N_2) and exchange migrants at constant rate with migration rates M_1 and M_2. B) A simpler *dadi* model *split_migr2* is nested within *IMpre* and includes five parameters. C) *splitExpMigr* model is similar to split_*migr2*, but it allows for gene flow to exponentially change over time from rate M_{split} right after species split to the migration rate at present ($M_{present}$). Furthermore, splitExpMigr assumes that migration at any given time is equal in both directions. D) The *fastsimcoal2* model *IM2R8S* includes nine free parameters, including four separate migration parameters (M_1 - M_4) describing migration before and after a single stepwise migration rate change at time T_1. Similar to *split_migr2* model, the population sizes of the two species following the species split at time T_0 are constant and determined by modern population sizes (N_1 and N_2). Unlike the other models described above, ancestral population size (N_{anc}) is a free parameter estimated in *IM2R8S* model.

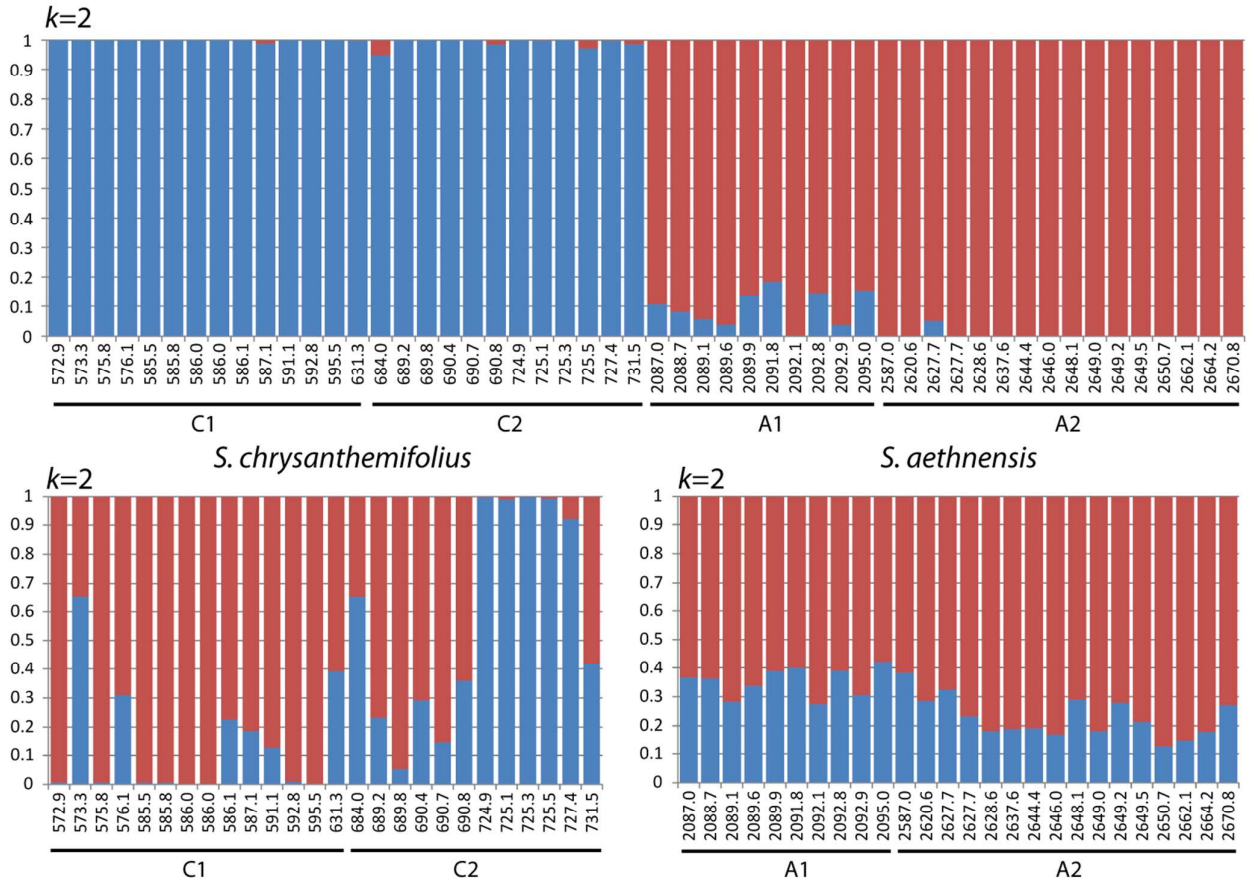


Figure 3. Clustering of samples by species (top plot) and populations within species (bottom plots), as revealed by *Structure* analysis.

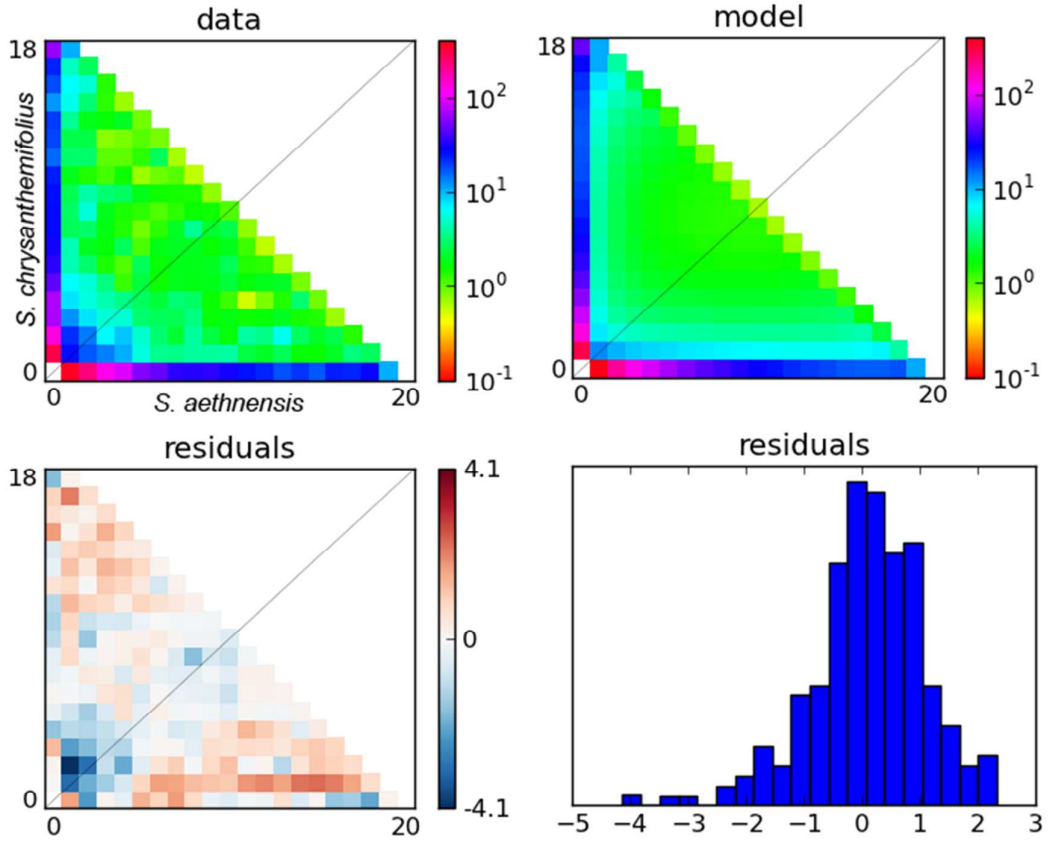


Figure 4. Observed and expected 2-dimentional site frequency spectra (2D-SFS) for *S. chrysanthemifolius* population C1 and *S. aethnensis* population A2. The expected 2D-SFS (top right) was generated for *IMpre* model. The residuals for fitting expected spectrum to data (observed spectrum) are shown in bottom plots.

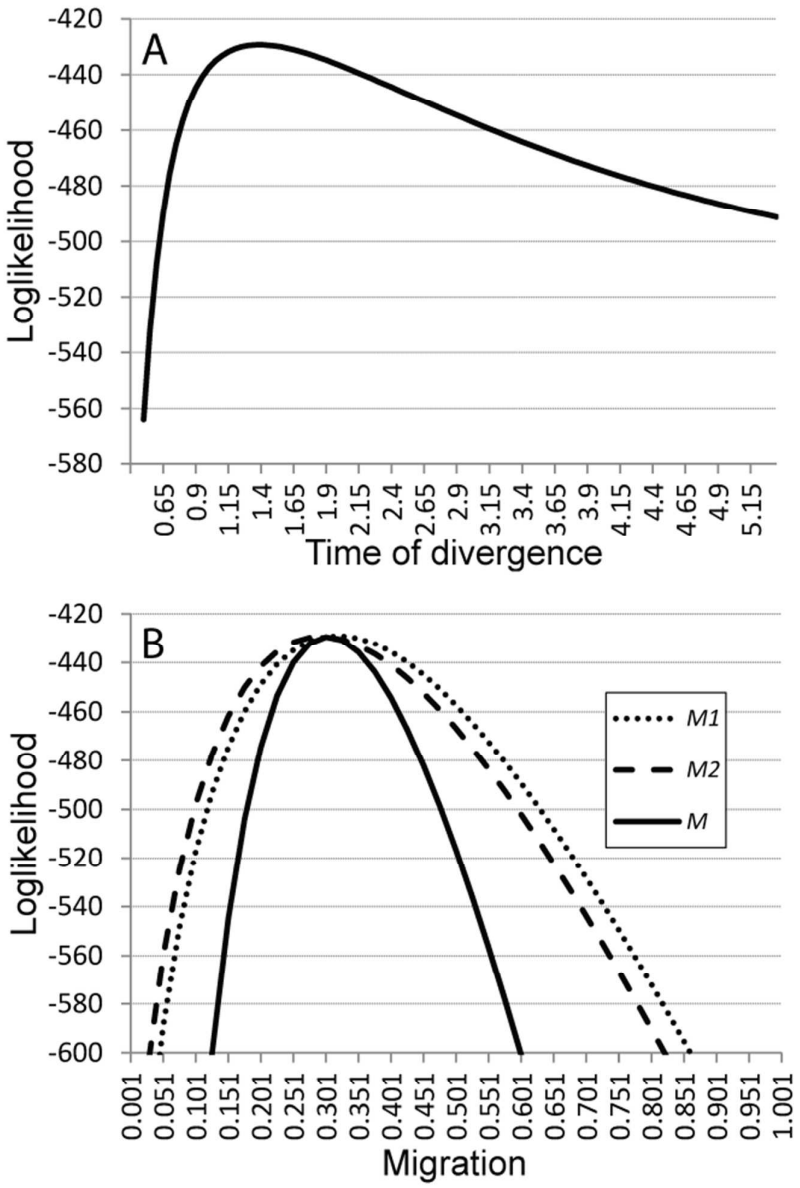


Figure 5. Likelihood profiles for the split with migration model for *S. chrysanthemifolius* population C1 and *S. aethnensis* population A2. Loglikelihoods are plotted for a range of values for the time of species divergence (A) and migration (B) parameters. Both panels are based on *split_migr2* model, except the solid curve on the panel B, which is based on model *split_migr*.

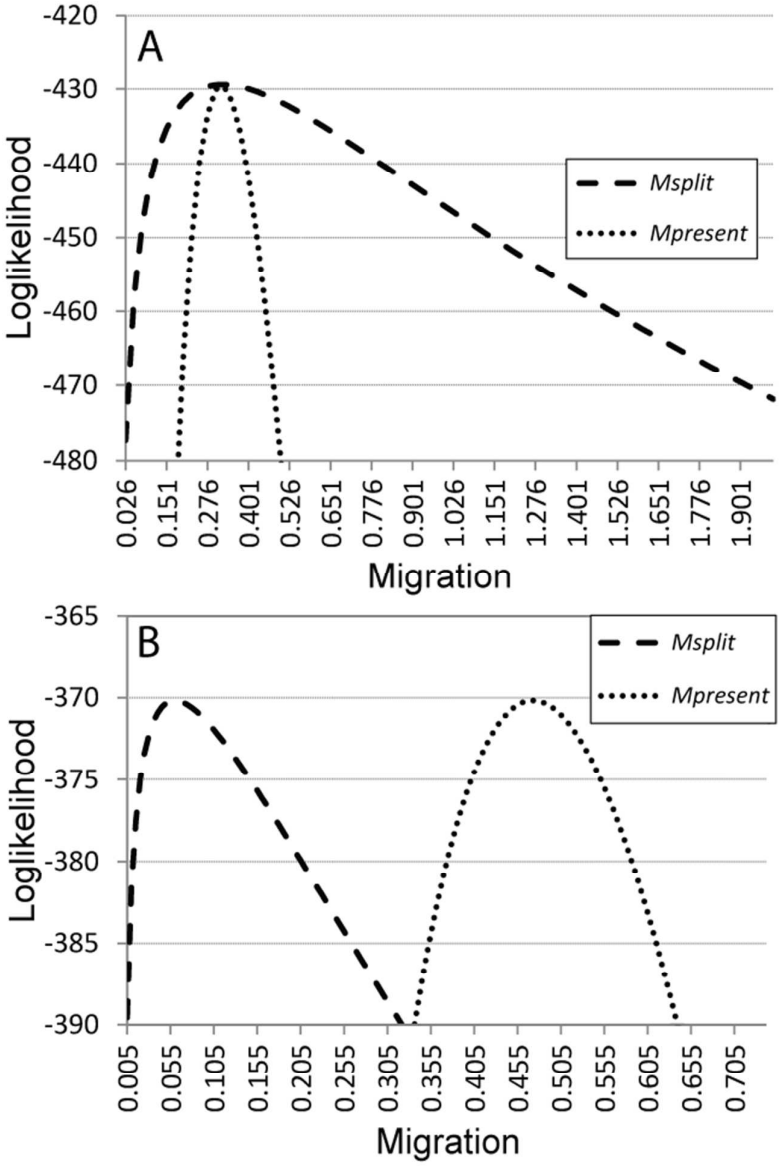


Figure 6. Discrimination of speciation with gene flow from secondary contact. Likelihood profiles for the split with migration model with exponentially growing migration after species split (*splitExpMigr*). In the model equal migration in both directions starts right after species split with rate M_{split} and it is allowed to exponentially change to rate $M_{present}$ – migration rate at present. (A) *S. chrysanthemifolius* population C1 and *S. aethnensis* population A2 (B) simulated species split with recent secondary contact.

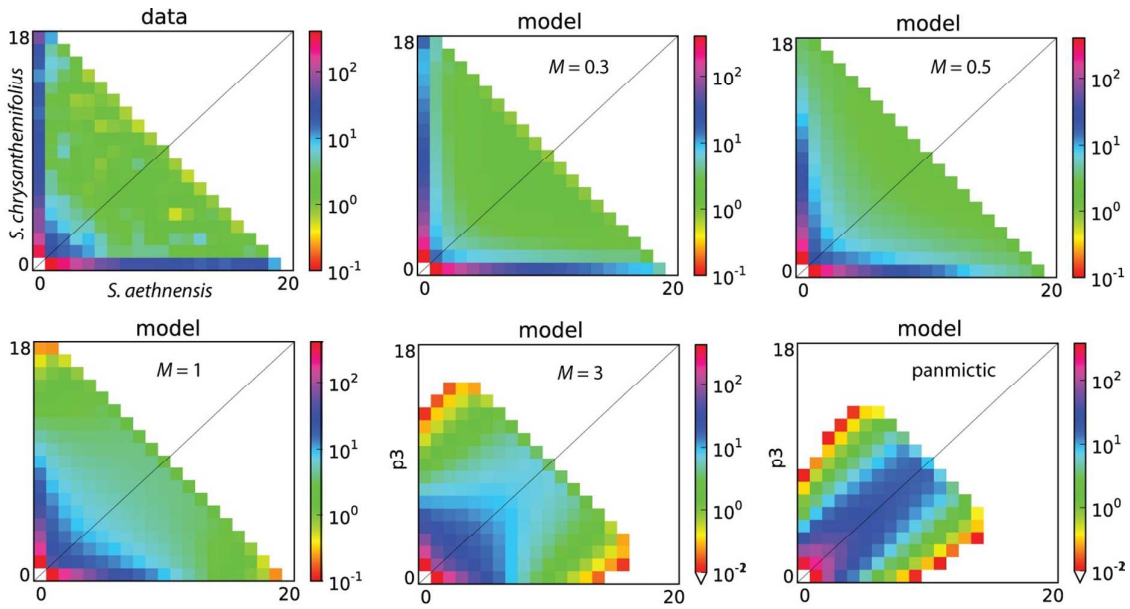


Figure 7. Observed and expected 2D-SFS with increasing amounts of gene flow. The top left panel shows the observed 2D-SFS for populations C1 and A2, while the other five panels show 2D-SFSs expected under population split model *split_migr* with migration rate increasing from M = 0.3 to a panmictic standard neutral model. All the simulations except the bottom right panel assumed that an ancestral population split into two populations of the same size $2N_e$ generations ago, and gene flow between the populations (M) remained constant until the present day. The bottom right panel assumed a standard neutral model with no population split.

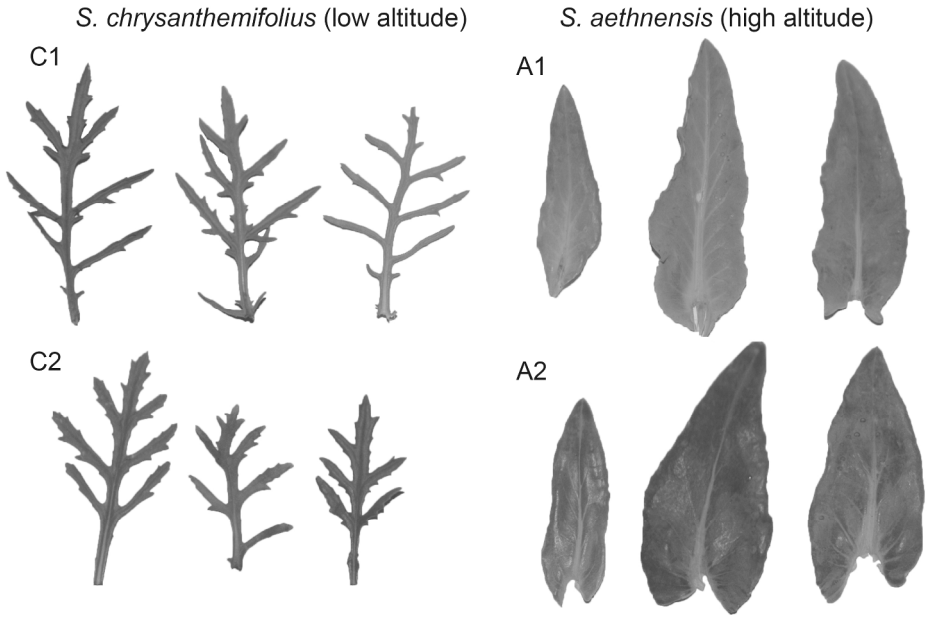


Figure 1. (high resolution)
250x157mm (300 x 300 DPI)

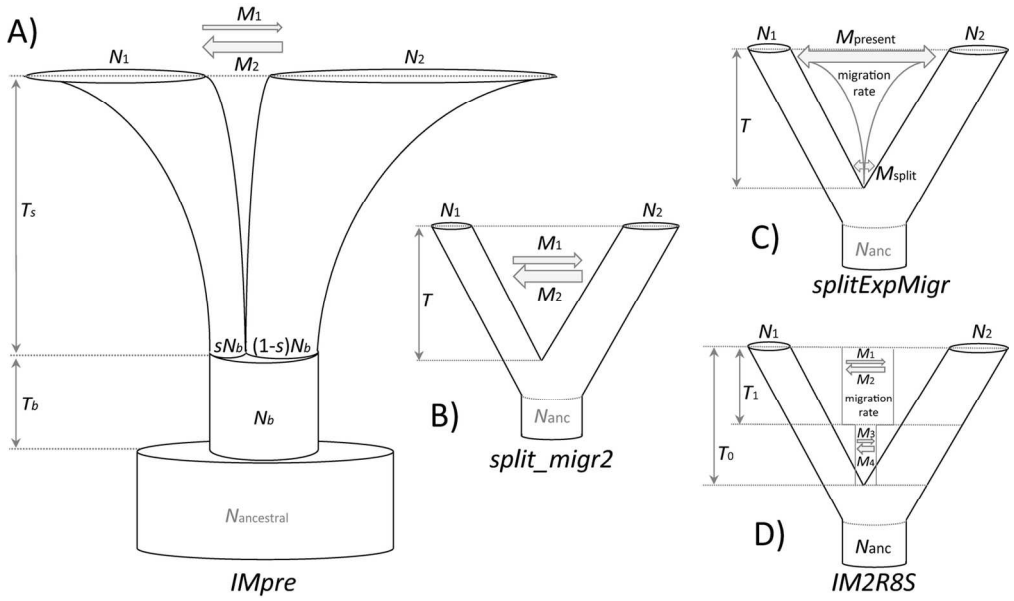


Figure 2. (high resolution)
139x82mm (300 x 300 DPI)

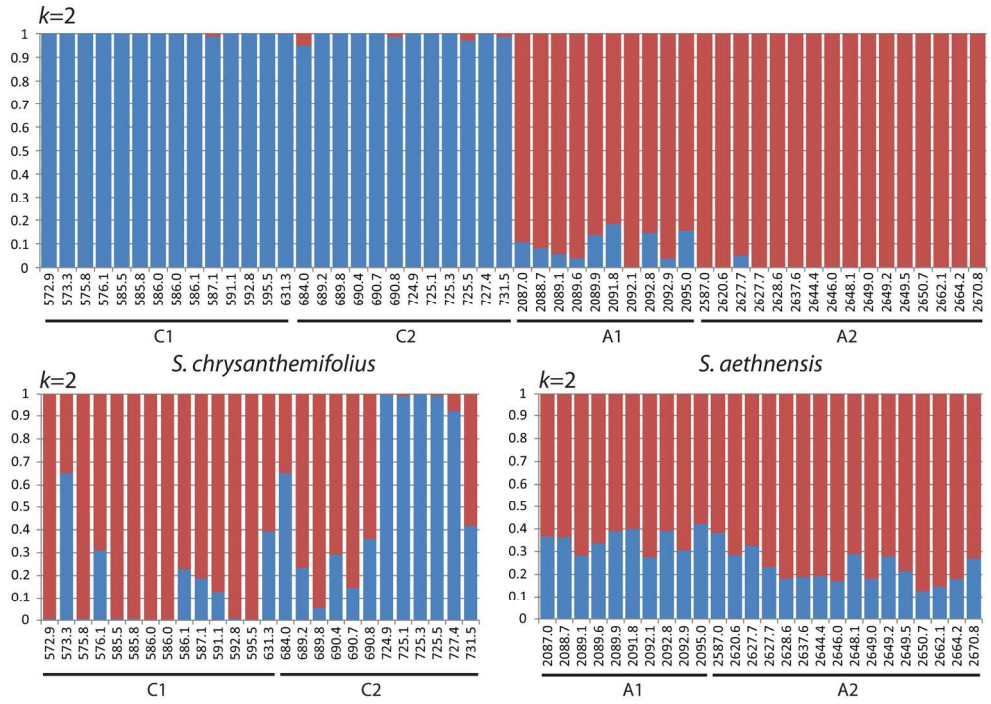


Figure 3. (high resolution)
184x128mm (300 x 300 DPI)

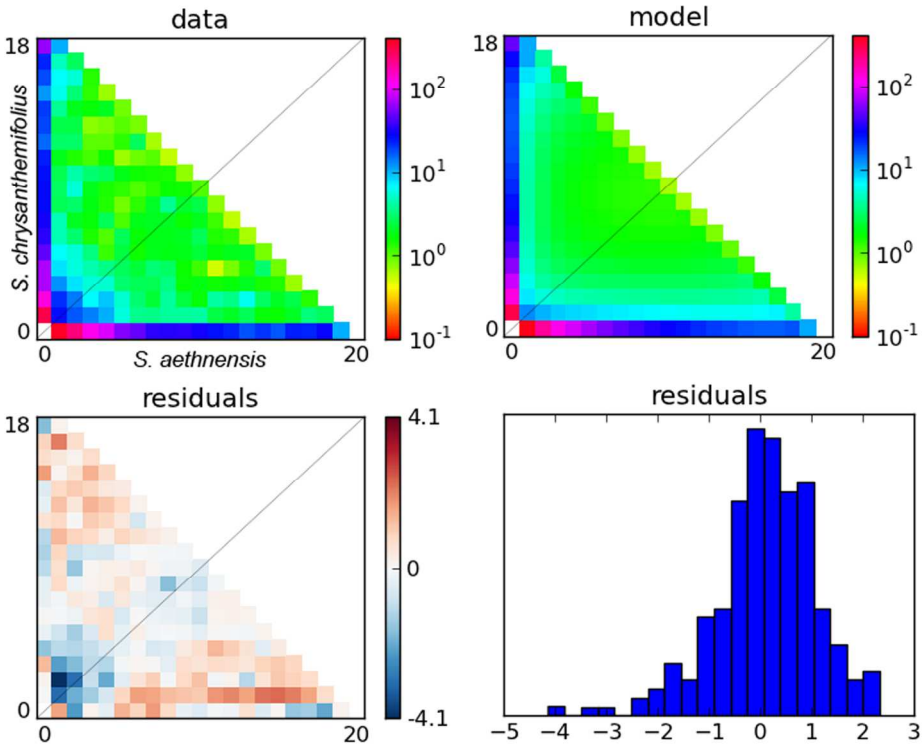


Figure 4. (high resolution)
193x151mm (300 x 300 DPI)

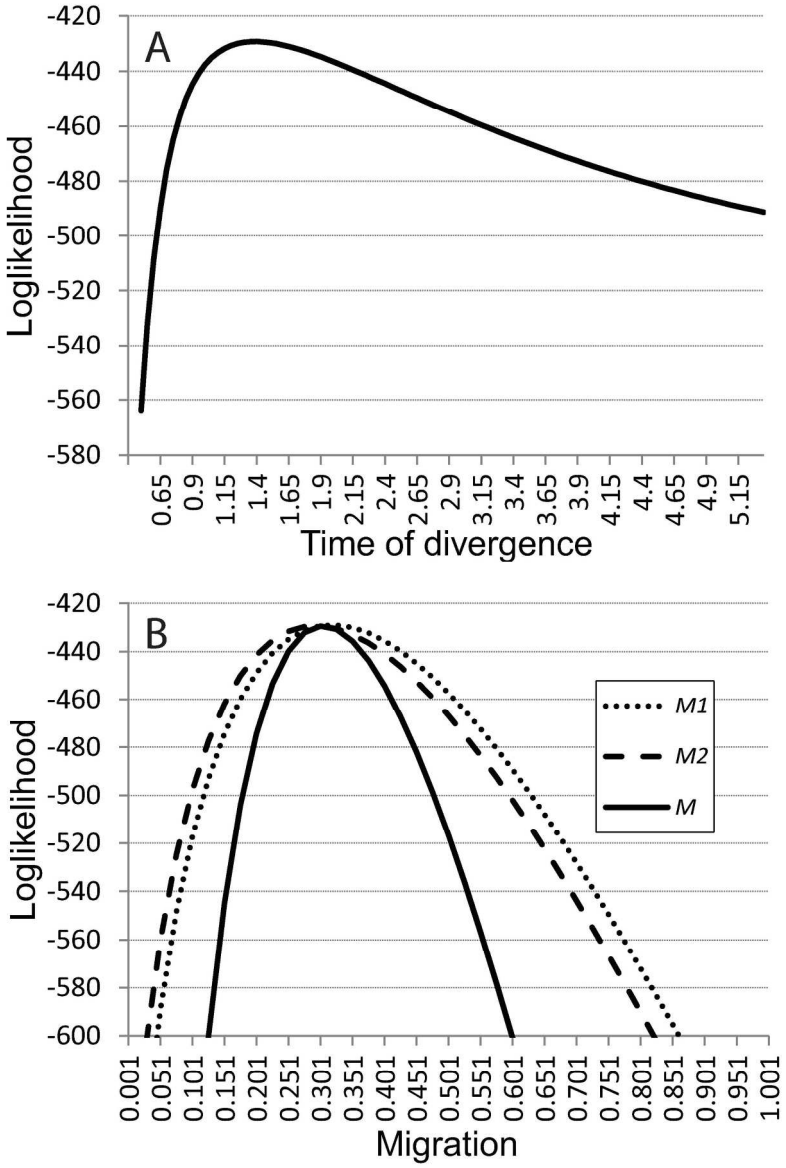


Figure 5. (high resolution)
171x254mm (300 x 300 DPI)

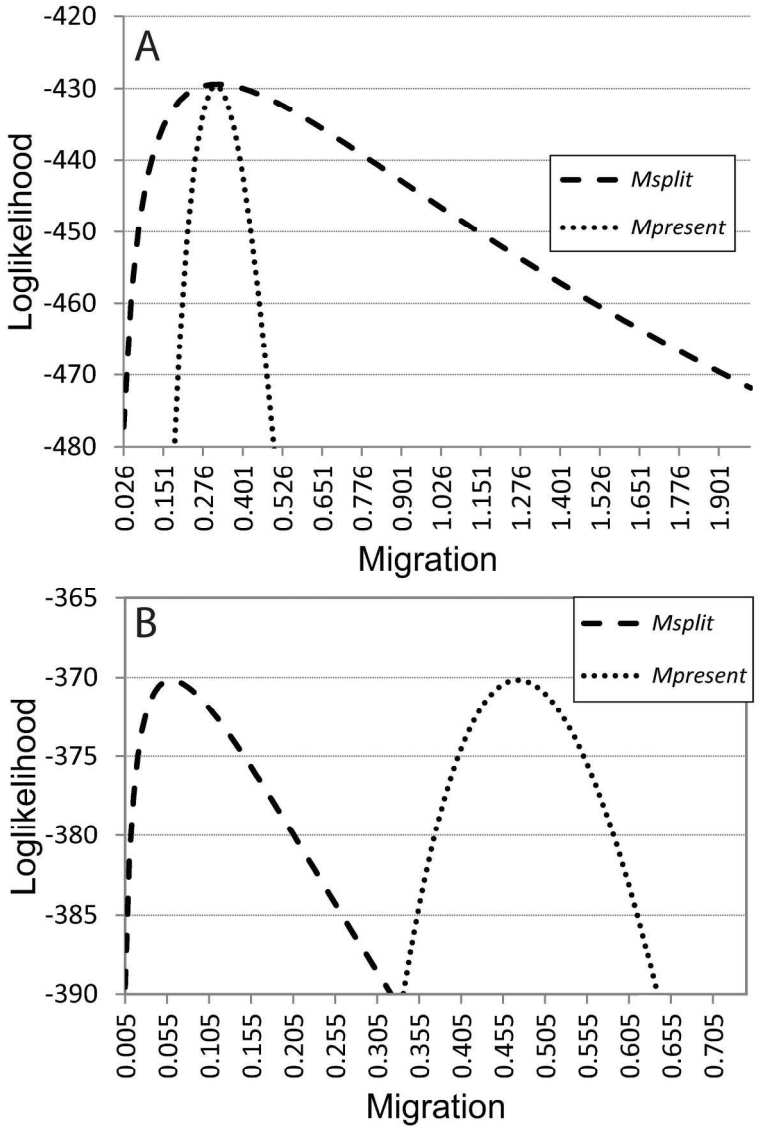


Figure 6. (high resolution)
174x254mm (300 x 300 DPI)

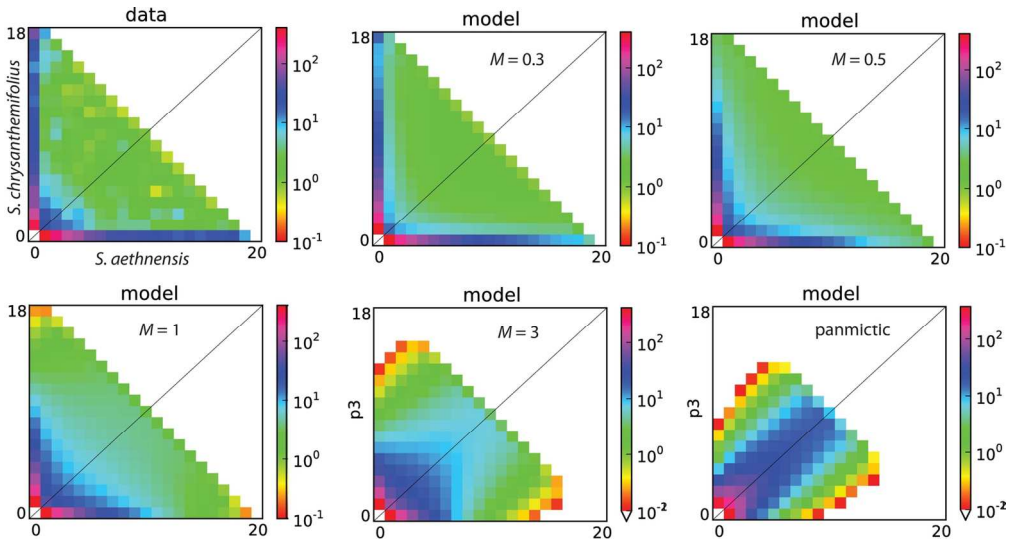


Figure 7. (high resolution)
144x76mm (300 x 300 DPI)

Supporting information for

Demographic history of speciation in a *Senecio* altitudinal hybrid zone on Mt. Etna

Dmitry A. Filatov, Owen G. Osborne and Alexander S. Papadopoulos

Supplementary Table S1. Demographic modelling of speciation for all combinations of *S. aethnensis* and *S. chrysanthemifolius* populations using *dadi*.

	log-likelihood	LRT[1)] 2ΔLL	LRT[1)] signif.?	theta	free params	N_b[2)]	T_b[2)]	s[2)]	N_{chr}[2)]	N_{aet}[2)]	T[2)]	M_1[2)] (chr<=aet)	M_2[2)] (chr=>aet)
population pair: C1 and A2													
IMpre	-428.76			420.30	8	1.39	0.29	0.42	0.68	0.99	1.47	0.31	0.26
IM	-429.44	1.36	NS	513.90	6	1*	1*	0.40	0.57	0.84	1.03	0.38	0.30
*IM**	-429.71	0.55	NS	547.40	5	1*	1*	0.5*	0.51	0.80	0.80	0.35	0.33
IMnoMigr	-540.04	220.65	*P*<0.00001	647.90	3	1*	1*	0.5*	0.35	0.61	0.29	0*	0*
split_migr2	-428.59			465.30	5	no	no	no	0.61	0.87	1.26	0.33	0.28
split_migr	-429.04	0.90	NS	455.52	4	no	no	no	0.63	0.86	1.28	$M_1 = M_2 = 0.30$	
population pair: C1 and A1													
IMpre	-358.51			464.76	8	0.89	1.05	0.90	0.54	0.71	1.10	0.29	1.15
IM	-358.54	0.06	NS	420.97	6	1*	1*	0.88	0.59	0.76	1.23	0.27	1.01
*IM**	-364.56	12.04	*P*<0.0001	431.11	5	1*	1*	0.5*	0.62	0.61	1.13	0.35	0.92
IMnoMigr	-471.06	212.98	*P*<0.00001	528.95	3	1*	1*	0.5*	0.31	0.51	0.19	0*	0*
split_migr2	-364.82			416.85	5	no	no	no	0.62	0.60	1.15	0.33	0.92
split_migr	-375.76	21.87	*P*<0.0001	415.69	4	no	no	no	0.52	0.70	1.13	$M_1 = M_2 = 0.61$	
population pair: C2 and A2													
IMpre	-466.30			1015.83	8	0.44	1.22	0.13	0.37	0.43	0.64	1.09	0.48
IM	-466.59	0.59	NS	513.41	6	1*	1*	0.15	0.71	0.81	1.13	0.55	0.25
*IM**	-476.88	20.57	*P*<0.0001	486.81	5	1*	1*	0.5*	0.63	0.95	1.29	0.44	0.31
IMnoMigr	-667.32	380.88	*P*<0.00001	673.10	3	1*	1*	0.5*	0.37	0.56	0.26	0*	0*
split_migr2	-478.16			496.55	5	no	no	no	0.56	0.82	1.19	0.49	0.29
split_migr	-482.68	9.03	*P*<0.01	484.40	4	no	no	no	0.63	0.78	1.27	$M_1 = M_2 = 0.37$	
population pair: C2 and A1													

IMpre	-396.72			450.38	8	0.83	0.62	0.18	0.76	0.49	0.70	0.76	0.49
IM	-397.11	0.78	NS	426.36	6	1*	1*	0.14	0.75	0.50	0.68	0.68	0.88
*IM**	-399.01	3.81	*P*<0.01	361.35	5	1*	1*	0.5*	0.72	0.64	1.14	0.41	0.96
IMnoMigr	-534.84	271.66	*P*<0.00001	452.97	3	1*	1*	0.5*	0.41	0.51	0.18	0*	0*
split_migr2	-401.74			338.48	5	no	no	no	0.70	0.64	1.39	0.39	0.94
split_migr	-411.63	19.79	*P*<0.0001	341.03	4	no	no	no	0.59	0.75	1.28	$M_1 = M_2 = 0.64$	

[1)] In all cases, LRTs are for the comparison with the model immediately above.

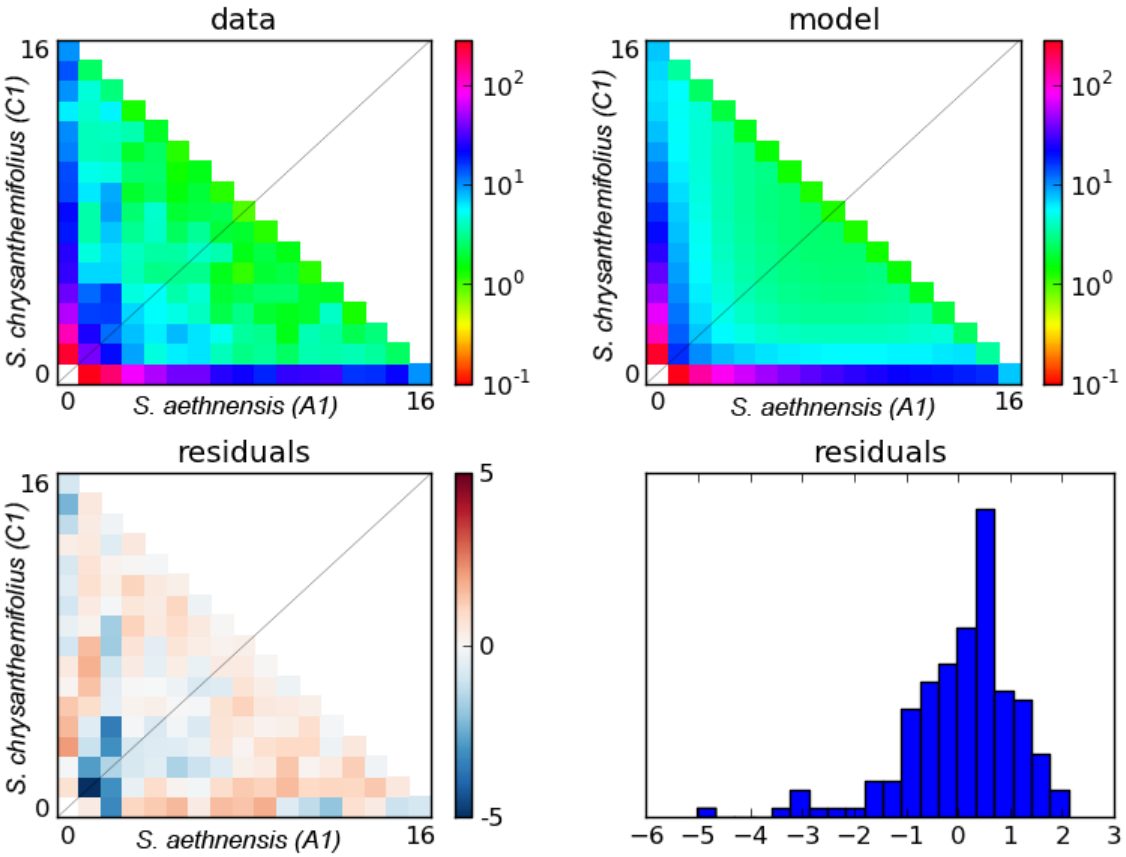
[2)] Parameters are as in table 4.

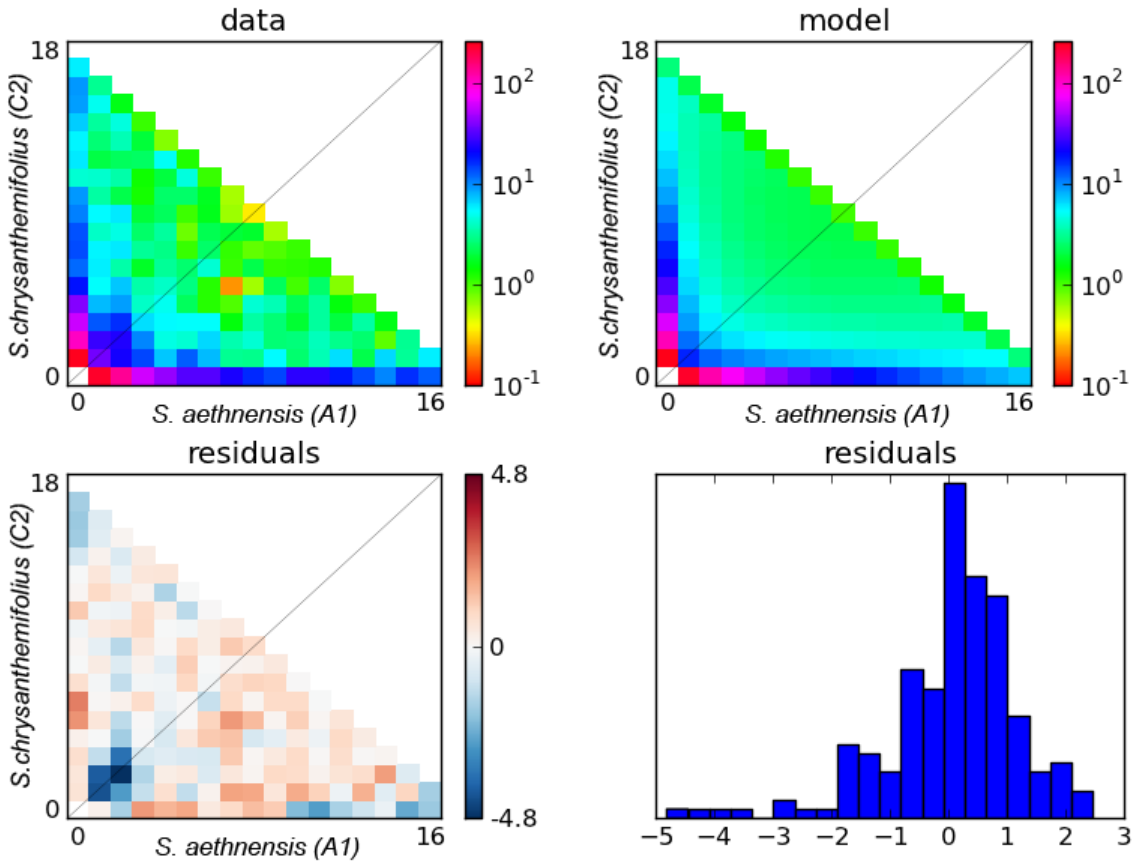
Supplementary Table S2. Demographic modelling of speciation for populations C1 and A2 using *fastsimcoal2*.

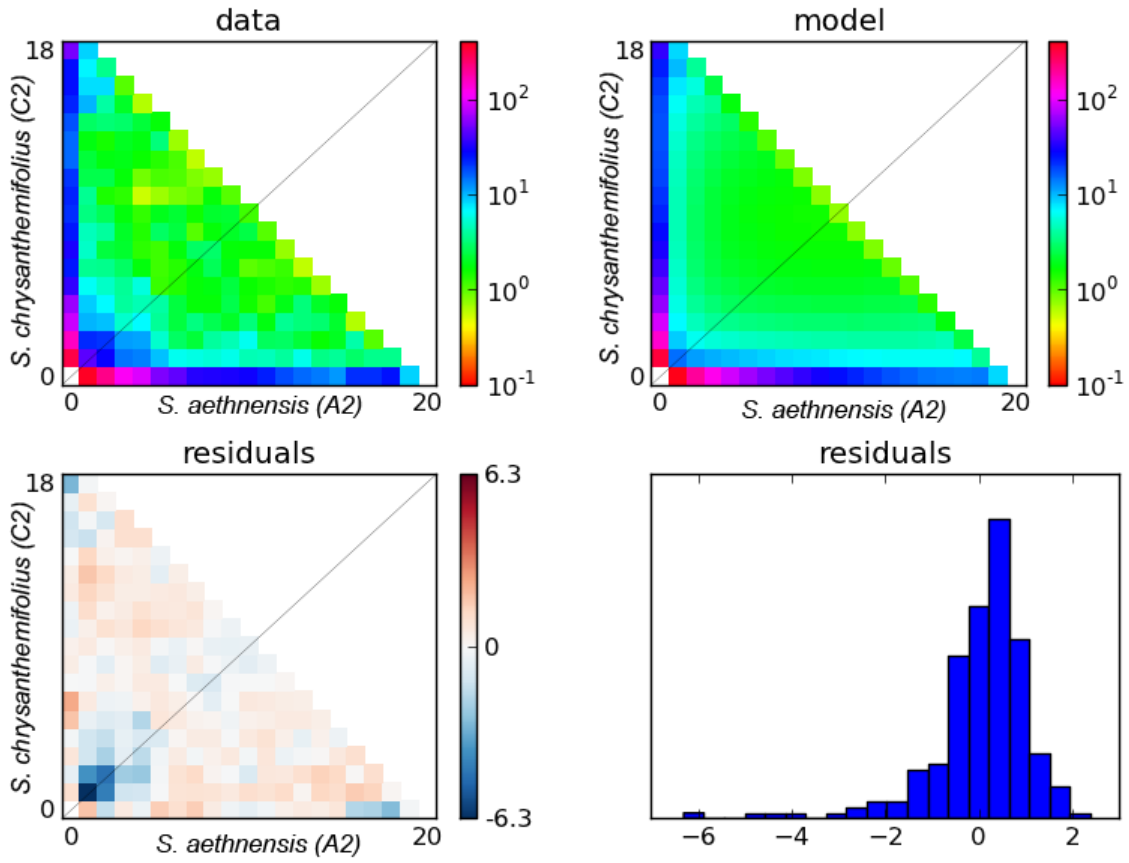
Models:	*DivNoM*	*IM*	***IM2R8S***[1)]	*IMSTOPS*	*IML8M*	*DivNoM_GR*	*IM_GR*	*IM2R8S_GR*	*IMSTOPS_GR*	*IML8M_GR*
ln(likelihood)	-28464	-28380	**-28345**	-28433	-28419	-28448	-28351	-28346	-28379	-28354
AIC	56936.3	56772.5	**56708.9**	56879.9	56852.6	56908.9	56718.5	56713.9	56775.6	56726.6
Δ AIC	227.4	63.6	**0.0**	170.9	143.7	200.0	9.6	5.0	66.6	17.7
Akaike weights of evidence	0.000	0.000	**0.916**	0.000	0.000	0.000	0.008	0.084	0.000	0.000
Change in migration rate	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Population size change	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Parameter number	4	6	9	7	7	6	8	11	9	9
Effective population sizes:										
Ancestral N_e	30638	33989	413191	108625	40001	670458	189537	299095	419407	429010
S. chrysanth. N_e present	11153	11638	12770	9397	11138	21583	13423	13253	16956	14298
S. aethnensis N_e present	15823	15905	18531	17285	14862	26321	21571	18362	20031	24973
S. chrysanth. N_e at time T_0	-[2)]	-	-	-	-	133271	12929	81418	287734	30638
S. aethnensis N_e at time T_0	-	-	-	-	-	118926	91535	18359	91380	97422
Time (generations):										
species split (T_0)	156060	334320	126407	205154	31558	12600	84813	287314	43547	29401
migration change (T_1)	-	-	60701	6457	25613	-	-	67431	3125	28901
Migration rates (m x10^{-5}):										
aet. to chr. (T_0 to present)	-	0.697	-	-	-	-	0.739	-	-	-
chr. to aet. (T_0 to present)	-	0.663	-	-	-	-	0.755	-	-	-
aet. to chr. (T_0 to T_1)	-	-	6.398	4.334	-	-	-	47.035	0.706	-
chrys. to aet. (T_0 to T_1)	-	-	55.809	1.012	-	-	-	100.44	1.69524	-
aet. to chr. (T_0 to present)	-	-	0.804	-	0.769	-	-	0.852	-	0.706
chr. to aet. (T_0 to present)	-	-	0.646	-	0.726	-	-	0.629	-	0.593

[1)] The best fitting model is shown in bold.

[2)] For models with no population size change N_e at present and at time T_0 is the same (single N_e parameter per species).


data
model
residuals
residuals
S. chrysanthemifolius (C1)
S. aethnensis (A1)
16
0
10^2
10^1
10^0
10^{-1}
5
0
-5
−6 −5 −4 −3 −2 −1 0 1 2 3


data
model
residuals
residuals
S.chrysanthemifolius (C2)
S. aethnensis (A1)
18
0
16
10^2
10^1
10^0
10^{-1}
4.8
-4.8
−5 −4 −3 −2 −1 0 1 2 3



Supplementary Figure S1. Observed and expected 2-dimentional site frequency spectra (2D-SFS) for population pairs C1/A1, C2/A1 and C2/A2. The expected 2D-SFS (top right) was generated for *IMpre* model. The residuals for fitting expected spectrum to data (observed spectrum) are shown in bottom plots.