Letter - Discoveries

The stepwise increase in the number of transcription factor families in the Precambrian predated the diversification of plants on land

Bruno Catarino[1*], Alexander J. Hetherington[1*], David M. Emms[1], Steven Kelly[1#] and Liam Dolan[1#]

[1]Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK

[*]Joint first authors.

[#]Corresponding authors. Email: steven.kelly@plants.ox.ac.uk; liam.dolan@plants.ox.ac.uk

Abstract

The colonisation of the land by streptophytes and their subsequent radiation is a major event in Earth history. We report a stepwise increase in the number of transcription factor (TF) families and subfamilies in Archaeplastida before the colonisation of the land. The subsequent increase in TF number on land was through duplication within existing TF families and subfamilies. Almost all subfamilies of the Homeodomain (HD) and basic Helix-Loop-Helix (bHLH) had evolved before the radiation of extant land plant lineages from a common ancestor. We demonstrate that the evolution of these TF families independently followed similar trends in both plants and metazoans; almost all extant HD and bHLH subfamilies were present in the first land plants and in the last common ancestor of bilaterians. These findings reveal that the majority of innovation in plant and metazoan TF families occurred in the Precambrian before the Phanerozoic radiation of land plants and metazoans.

Introduction

Complex multicellular plants and metazoans diversified at the taxonomic, morphological and genetic levels during the Phanerozoic eon (541 million years ago to present) (Carroll, 2001; Droser and Gehling, 2015; Erwin et al., 2011; Kenrick and Crane, 1997a; Powell and Kowalewski, 2002). Transcription factors (TFs) control a plethora of developmental mechanisms in eukaryotic organisms that appear during the course of evolutionary radiations and transitions (Holland, 2013; Menand et al., 2007; de Mendoza et al., 2013). The role of TF diversification around one of these major transitions– the colonisation of the land by streptophyte plants – is not understood (Mukherjee et al., 2009; Pires and Dolan, 2010). The move to land by plants was paralleled by a major body plan transition; the evolution of complex 3D tissues and key morphological characters crucial for life on land such as the epidermis, rooting and shooting systems, stomata, and water conducting tissues that are associated with the function of characteristic TFs (Graham et al., 2000; Kenrick and Crane, 1997b; MacAlister and Bergmann, 2011; Menand et al., 2007; Pillitteri et al., 2007; Pires et al., 2013; De Rybel et al., 2013; Sakakibara, 2016; Tam et al., 2015; Xu et al., 2014). The recent availability of the genome sequence of an aquatic streptophyte alga from a clade that is sister to the land plants (Wickett et al., 2014), *Klebsormidium flaccidum* (Hori et al., 2014), and the liverwort, one of the earliest divergent lineages of land plants (Kenrick and Crane, 1997b; Wickett et al., 2014), *Marchantia polymorpha* (NCBI GenBank accession LVLJ00000000.1) allowed us to track the change in the number and size of TF families and subfamilies during streptophyte evolution. Here we demonstrate that the origin of the majority of TF families and subfamilies predate the radiation of extant land plant lineages.

Results and Discussion

47 of the 48 land plant TF families evolved before the colonisation of the land

We identified 48 (see Supplementary Material) TF families in the genomes of 15 species of Archaeplastida (the lineage that includes red algae, chlorophytes and streptophytes) (Adl et al., 2005; Archibald, 2009). These taxa comprise the red algae *Cyanidioschyzon merolae*; *Porphyridium purpureum*, *Chondrus crispus* the chlorophytes *Ostreococcus tauri*, *Micromonas pusilla*, *Chlorella variabilis*, *Coccomyxa subellipsoidea*, *Volvox carteri* and *Chlamydomonas reinhardtii*; the streptophyte alga *Klebsormidium flaccidum* and the streptophyte land plants *Marchantia polymorpha*, *Physcomitrella patens*, *Selaginella moellendorffii*, *Oryza sativa* and *Arabidopsis thaliana* (see Supplementary Material, Supplementary table 1). TF genes were defined using established rules for the identification and categorisation of TFs on the basis of their domain architecture (Finn et al. 2014; Jin et al. 2014, Supplementary Material). The number of TF families that are present at each internal node in the phylogeny was inferred using parsimony. This revealed that the number of TF families

increased progressively from node 1 to node 6 (fig 1). There were 17 TF families in the common ancestor of all species in this analysis (node 1 fig 1) and 28 in the common ancestor of chlorophytes and streptophytes (node 2 fig 1). 39 TF families were present in the aquatic common ancestor of *K. flaccidum* and land plants (node 3 fig 1) and 47 TF families had evolved in the first land plants (node 4 fig. 1). The stepwise origin of TF families in the aquatic ancestors of land plants – from 17 TFs in the last common ancestor of all species at node 1 to 47 in the first land plants – contrasts strikingly with the subsequent evolution of TF families on land. During the radiation of plants in the terrestrial realm TF number increased from 47 to 48; only one TF family, GeBP (node 5 fig. 1), evolved after the divergence of bryophytes and vascular plants (fig. 1). This indicates that there was a stepwise increase in the number of TF families in Archaeplastida before the radiation of land plants and since that time there has been relatively little change in TF family number (fig. 1).

Land plants radiated both taxonomically and morphologically after the colonisation of the land (Bateman et al., 1998; Finet et al., 2010; Kenrick and Crane, 1997b; Nickrent et al., 2000; Qiu et al., 2006; Wickett et al., 2014). However, this analysis revealed that this radiation was not accompanied by an increase in the number of novel TF families (fig. 1). Instead, the number of TFs within each family increased as land plants evolved after the colonisation of the land (de Mendoza et al., 2013; Mukherjee et al., 2009; Pires and Dolan, 2010). It is possible that TF evolution was accompanied by the origin of new subfamilies during land plant evolution. To test the hypothesis that the origin of novel TF subfamilies paralleled the radiation of plants on land a gene tree analysis was conducted on two of the major pan-eukaryotic TF families (included in the 48 TF families examined above), Homeodomain (HD) and basic Helix-Loop-Helix (bHLH) (Degnan et al., 2009; Gyoja, 2014; Holland, 2013; Mukherjee et al., 2009; Nam and Nei, 2005; Pires and Dolan, 2010; Sakakibara, 2016; Simionato et al., 2007).

13 of the 14 subfamilies of Homeodomain transcription factors are present in streptophyte algae

Phylogenetic analysis of Archaeplastida HD proteins (fig. 2A, supplementary fig. S1) resolved the 14 monophyletic subfamilies previously reported by Mukherjee and colleagues (2009). 11 of the 14 subfamilies are supported by Shimodaira-Hasegawa-like approximate likelihood ratio test (SH-like aLRT) values over 0.85 (supplementary fig. S1). There are 26 HD TFs in 13 subfamilies in the genome of the streptophyte algae *K. flaccidum* (fig. 2A, Supplementary table 2); NDX is the only land plant subfamily that is not present the *K. flaccidum* genome. This indicates that at least 13 HD subfamilies had originated in an aquatic streptophyte ancestor before colonisation of the land. Phylogenetic analysis of the 19 *M. polymorpha* HD TFs indicates that all 14 HD subfamilies previously described in land plants are present in this early diverging land plant. Furthermore, there is a single gene in the majority of *K. flaccidum* and *M. polymorpha* subfamilies, whereas there are multiple genes in the same subfamilies of angiosperms (Supplementary table 2). The majority (13/14) of TF subfamilies was also present before the colonization of the land, as described above in the analysis of all TF families. Thus the colonization and subsequent diversification of plants on land was not accompanied by the substantial evolution of new HD subfamilies.

26 of the 30 land plant basic helix-loop-helix subfamilies were present in the first land plants

The analysis of the gene tree of Archaeplastida bHLH genes indicated that the major families and subfamilies had evolved before the colonisation of the land as observed with HD subfamilies. This analysis revealed that there is a single bHLH protein in the red alga, *C. merolae* and 3 bHLH proteins encoded in the genomes of the chlorophytes *C. reinhardtii and V. carteri* (Supplementary table 3). However, none of the 30 streptophyte bHLH subfamilies were found in either red algae or chlorophytes (fig.2B). There are 10 *K. flaccidum* bHLH TFs in 6 subfamilies all of which were

previously described in land plants – IVb, IVc,Vb, VII(a+b), XI and XIII. The genome of the liverwort *M. polymorpha* encodes 49 bHLH proteins which comprise 25 of the 30 plant bHLH subfamilies (fig. 2B, Supplementary table 3). The *M. polymorpha* sequences, together with the sequences of *P. patens*, *S. moelendorffii*, *O. sativa* and *A. thaliana*, defined 4 new monophyletic bHLH subfamilies: IVd(2), VIIIc(3), XVI and XVII. 23 of the 30 bHLH subfamilies are monophyletic clades supported by SH-like aLRT values greater than 0.85 (supplementary fig. S2). The numbers of bHLH subfamilies in the earliest divergent land plants (*M. polymorpha* and *P. patens*) allow us to infer that 26 bHLH subfamilies were present in the first land plants.

These data indicate that some new bHLH subfamilies evolved while other subfamilies were lost during the course of land plant evolution. Subfamilies VIIIc(3) and XVI are present only in bryophytes (liverworts and moss) suggesting that subfamilies VIIIc(3) and XVI were either lost before the origin of vascular plants or that they evolved independently in the case that bryophyte constitute a monophyletic clade (Cox et al., 2014; Wickett et al., 2014). There are only non-seed plant proteins in subfamily XVII. The most parsimonious interpretation of this result is that subfamily XVII was lost in the lineage giving rise to the seed plants after the divergence of the lycophytes from the seed plants (fig. 2B, Supplementary table 3). Moreover, subfamilies Ib(2), IVd(1), IVd(2) and XV are present in angiosperms only (fig. 2B, Supplementary table 3), suggesting that these 4 subfamilies evolved in the lineage leading to the angiosperms after the divergence of lycophytes and seed plants. These data suggest that in contrast to HD subfamilies evolution, there were losses and origins of new bHLH subfamilies during the evolution of plants on land. The presence of 6 bHLH subfamilies in *K. flaccidum* compared to 26 bHLH subfamilies in bryophytes (*M. polymorpha* and *P. patens*) suggests that the majority of bHLH subfamilies originated around the time that plants colonised the land, or at very least after the time when *K. flaccidum* and early diverging land plants last shared a common ancestor.

Precambrian evolution of the majority of the HD and bHLH subfamilies in plants and metazoans

To independently assess if morphological diversification was predated by the evolution of the majority of TF subfamilies in another eukaryotic radiation we contrasted the evolution of both HD and bHLH families in metazoans (Degnan et al., 2009; Larroux et al., 2008; Ryan et al., 2010). Strikingly, in metazoans, like plants, there was an early origin followed by a stepwise increase of the majority of HD and bHLH subfamilies before the radiation of bilaterians (fig. 3). This observation indicates that the evolution of the majority of HD and bHLH subfamilies in the Precambrian predated the major morphological diversification in both plants (radiation of land plants) and metazoans (radiation of bilaterians) in the Phanerozoic (Clarke et al., 2011; Erwin et al., 2011; Kenrick and Crane, 1997a; Parfrey et al., 2011).

We demonstrate that there was a stepwise increase in the number of TF families during the course of Archaeplastida evolution before the colonisation of the land by plants. The majority of plant TF families had already evolved by the time the first plants colonised the land. This stepwise increase in the number of families before the radiation of land plants is paralleled by a similar stepwise increase in the number of subfamilies in two of the major pan-eukaryotic TF families (HD and bHLH). Taken together, these data demonstrate that there was a gradual increase in the number of TF families and subfamilies before the radiation of plants on land. This was followed by a relatively small increase in the number of TF families and subfamilies during and after the land plant radiation. Therefore the morphological radiation of land plants was accompanied by an increase in number of TF proteins within these highly conserved TF subfamilies. Similarly metazoan TFs originated in the Precambrian before the bilaterian radiation. This suggests that TF diversification in the Precambrian preceded two major radiations in organismal diversity in distant branches of the tree of life.

Materials and Methods are described in Supplementary Material.

Acknowledgements

This work was supported by the PLANTORIGINS Marie Curie Network and the EVO500 European Research Council-Advanced Grant to L.D (EVO500 Project No: 25028) which funded B.C.; A.J.H. was funded by a Doctoral Training Partnership Scholarship from the Biotechnology and Biological Research Council (BB/J014427/1). This project has also received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 637765 (DME). SK is a Royal Society University Research Fellow.

References

Adl, S.M., Simpson, A.G.B., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., Bowser, S.S., Brugerolle, G., Fensome, R.A., Fredericq, S., et al. (2005). The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J. Eukaryot. Microbiol. *52*, 399–451.

Archibald, J.M. (2009). The Puzzle of Plastid Evolution. Curr. Biol. *19*, 81–88.

Bateman, R.M., Crane, P.R., DiMichele, W.A., Kenrick, P.R., Rowe, N.P., Speck, T., and Stein, W.E. (1998). EARLY EVOLUTION OF LAND PLANTS: Phylogeny, Physiology, and Ecology of the Primary Terrestrial Radiation. Annu. Rev. Ecol. Syst. *29*, 263–292.

Carroll, S.B. (2001). Chance and necessity: the evolution of morphological complexity and diversity. Nature *409*, 1102–1109.

Clarke, J.T., Warnock, R.C.M., and Donoghue, P.C.J. (2011). Establishing a time-scale for plant evolution. New Phytol. *192*, 266–301.

Cox, C.J., Li, B., Foster, P.G., Embley, T.M., and Civáň, P. (2014). Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. Syst. Biol. *63*, 272–279.

Degnan, B.M., Vervoort, M., Larroux, C., and Richards, G.S. (2009). Early evolution of metazoan transcription factors. Curr. Opin. Genet. Dev. *19*, 591–599.

Droser, M.L., and Gehling, J.G. (2015). The advent of animals: The view from the Ediacaran. Proc. Natl. Acad. Sci. *112*, 4865–4870.

Erwin, D.H., Laflamme, M., Tweedt, S.M., Sperling, E.A., Pisani, D., and Peterson, K.J. (2011). The Cambrian conundrum: early divergence and later ecological success in the early history of animals. Science *334*, 1091–1097.

Finet, C., Timme, R.E., Delwiche, C.F., and Marlétaz, F. (2010). Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. Curr. Biol. *20*, 2217–2222.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: The protein families database. Nucleic Acids Res. *42*, 222–230.

Graham, L.E., Cook, M.E., and Busse, J.S. (2000). The origin of plants: body plan changes contributing to a major evolutionary radiation. Proc. Natl. Acad. Sci. U. S. A. *97*, 4535–4540.

Gyoja, F. (2014). A genome-wide survey of bHLH transcription factors in the Placozoan Trichoplax adhaerens reveals the ancient repertoire of this gene family in metazoan. Gene *542*, 29–37.

Holland, P.W.H. (2013). Evolution of homeobox genes. Wiley Interdiscip. Rev. Dev. Biol. *2*, 31–45.

Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., Sato, S., Yamada, T.,

Mori, H., Tajima, N., et al. (2014). Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. Nat. Commun. *5*, 3978.

Jin, J., Zhang, H., Kong, L., Gao, G., and Luo, J. (2014). PlantTFDB 3.0: A portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res. *42*, 1–6.

Kenrick, P., and Crane, P.R. (1997a). The Origin and Early Diversification of Land Plants – A Cladistic Study (Washington DC: Smithsonian Institute Press).

Kenrick, P., and Crane, P.R. (1997b). The origin and early evolution of plants on land. Nature *389*, 33–39.

Larroux, C., Luke, G.N., Koopman, P., Rokhsar, D.S., Shimeld, S.M., and Degnan, B.M. (2008). Genesis and expansion of metazoan transcription factor gene classes. Mol. Biol. Evol. *25*, 980–996.

MacAlister, C.A., and Bergmann, D.C. (2011). Sequence and function of basic helix-loop-helix proteins required for stomatal development in Arabidopsis are deeply conserved in land plants. Evol. Dev. *13*, 182–192.

Menand, B., Yi, K., Jouannic, S., Hoffmann, L., Ryan, E., Linstead, P., Schaefer, D.G., and Dolan, L. (2007). An ancient mechanism controls the development of cells with a rooting function in land plants. Science *316*, 1477–1480.

de Mendoza, A., Sebé-Pedrós, A., Šestak, M.S., Matejcic, M., Torruella, G., Domazet-Loso, T., and Ruiz-Trillo, I. (2013). Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proc. Natl. Acad. Sci. U. S. A. *110*, 1–9.

Mukherjee, K., Brocchieri, L., and Bürglin, T.R. (2009). A comprehensive classification and evolutionary analysis of plant homeobox genes. Mol. Biol. Evol. *26*, 2775–2794.

Nam, J., and Nei, M. (2005). Evolutionary change of the numbers of homeobox genes in bilateral animals. Mol. Biol. Evol. *22*, 2386–2394.

Nickrent, D.L., Parkinson, C.L., Palmer, J.D., and Duff, R.J. (2000). Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. Mol. Biol. Evol. *17*, 1885–1895.

Parfrey, L.W., Lahr, D.J.G., Knoll, A.H., and Katz, L. a (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc. Natl. Acad. Sci. U. S. A. *108*, 13624–13629.

Pillitteri, L.J., Sloan, D.B., Bogenschutz, N.L., and Torii, K.U. (2007). Termination of asymmetric cell division and differentiation of stomata. Nature *445*, 501–505.

Pires, N., and Dolan, L. (2010). Origin and diversification of basic-helix-loop-helix proteins in plants. Mol. Biol. Evol. *27*, 862–874.

Pires, N.D., Yi, K., Breuninger, H., Catarino, B., Menand, B., and Dolan, L. (2013). Recruitment and remodeling of an ancient gene regulatory network during land plant evolution. Proc. Natl. Acad. Sci. *110*, 9571–9576.

Powell, M.G., and Kowalewski, M. (2002). Increase in evenness and sampled alpha diversity through the Phanerozoic: Comparison of early Paleozoic and Cenozoic marine fossil assemblages. Geology *30*, 331–334.

Qiu, Y.-L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J., et al. (2006). The deepest divergences in land plants inferred from phylogenomic evidence. Proc. Natl. Acad. Sci. U. S. A. *103*, 15511–15516.

Ryan, J.F., Pang, K., Mullikin, J.C., Martindale, M.Q., and Baxevanis, A.D. (2010). The homeodomain complement of the ctenophore Mnemiopsis leidyi suggests that Ctenophora and

Porifera diverged prior to the ParaHoxozoa. Evodevo *1*, 9.

De Rybel, B., Möller, B., Yoshida, S., Grabowicz, I., Barbier de Reuille, P., Boeren, S., Smith, R.S., Borst, J.W., and Weijers, D. (2013). A bHLH complex controls embryonic vascular tissue establishment and indeterminate growth in Arabidopsis. Dev. Cell *24*, 426–437.

Sakakibara, K. (2016). Chapter One - Technological Innovations Give Rise to a New Era of Plant Evolutionary Developmental Biology (Elsevier Ltd).

Simionato, E., Ledent, V., Richards, G., Thomas-Chollier, M., Kerner, P., Coornaert, D., Degnan, B.M., and Vervoort, M. (2007). Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. BMC Evol. Biol. *7*, 33.

Tam, T.H.Y., Catarino, B., and Dolan, L. (2015). Conserved regulatory mechanism controls the development of cells with rooting functions in land plants. Proc. Natl. Acad. Sci. 201416324.

Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. *111*, E4859–E4868.

Xu, B., Ohtani, M., Yamaguchi, M., Toyooka, K., Wakazaki, M., Sato, M., Kubo, M., Nakano, Y., Sano, R., Hiwatashi, Y., et al. (2014). Contribution of NAC transcription factors to plant adaptation to land. Science *343*, 1505–1508.

Figure legends

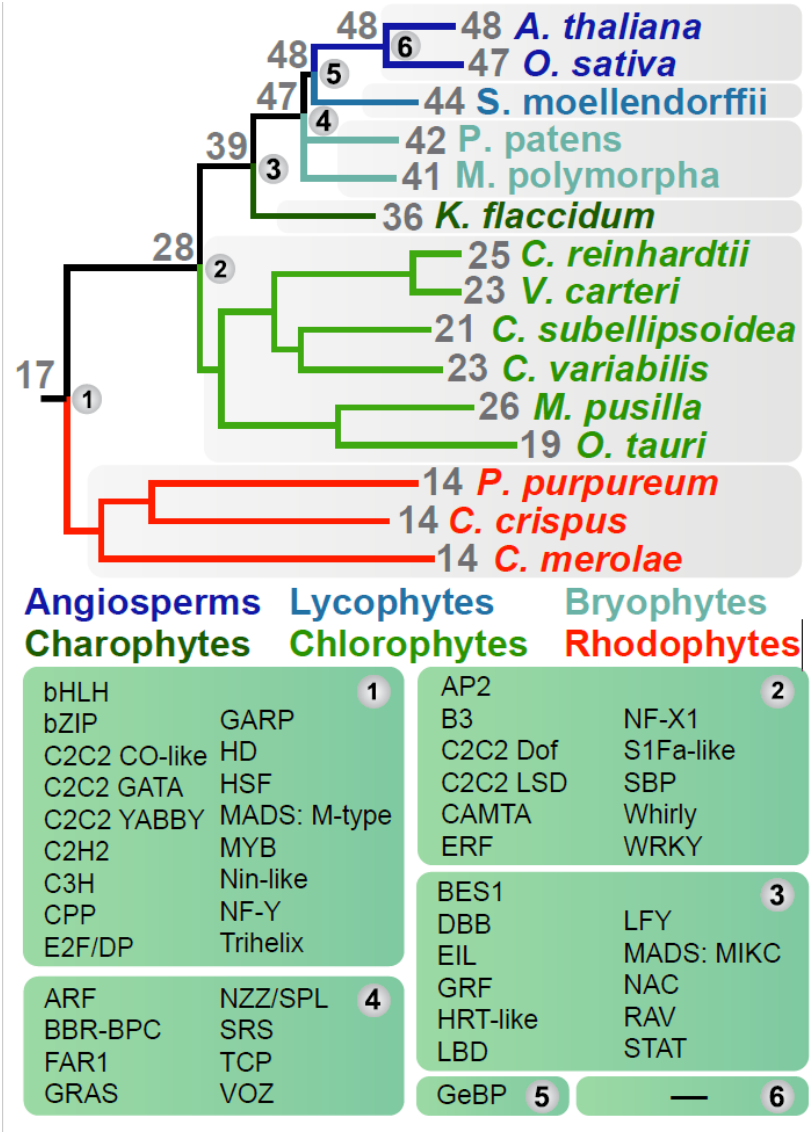

Figure 1. Distribution of transcription factor families in plants. Cladogram of Archaeplastida phylogeny (based on Finet et al. 2010, Cox et al. 2014, Wickett et al. 2014) with the origin of the transcription factor families (within green boxes) and total number of transcription factor families (grey) present at each ancestral node (circled numbers) and in extant species (coloured according to clade).

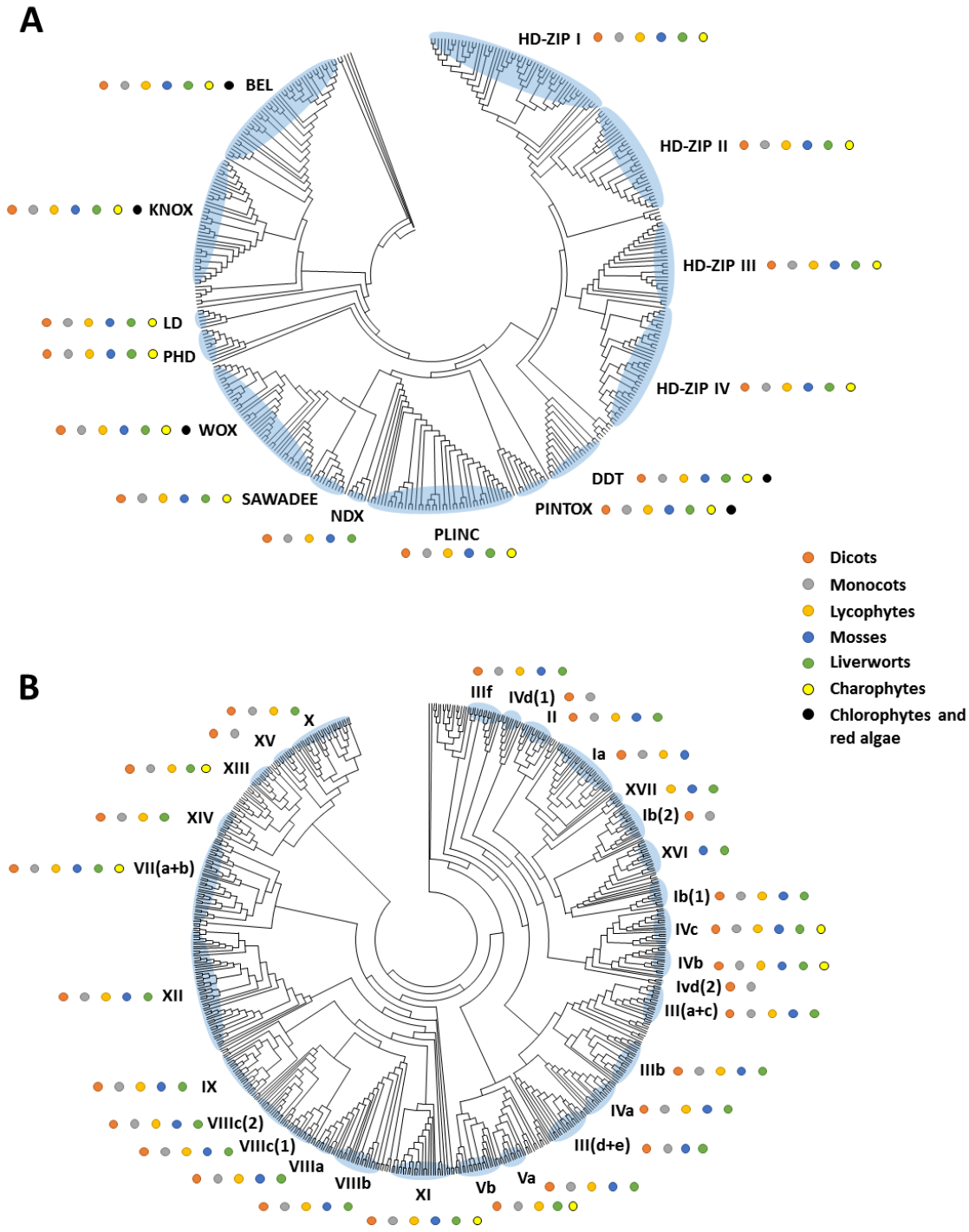

Figure 2. Phylogenetic analysis of Homeodomain and bHLH Archaeplastida transcription factors. Circular cladograms representing the Maximum Likelyhood (ML) analysis of Archaeplastida homeodomain (A) and basic Helix-Loop-Helix (B) transcription factor proteins. The trees are unrooted ML generated with the software PhyML 3.0, (Guindon *et al.*, 2010) using the JTT model of amino acid substitution, an estimated gamma distribution parameter and a SH-like aLRT test. Subfamiles are grouped with blue ellipses. Coloured dots represent protein presence in each subfamily. *Arabidopsis thaliana* (orange); *Oryza sativa* (grey); *Selaginella moellendorffii* (gold); *Physcomitrella patens* (blue)*; Marchantia polymorpha* (green); *K. flaccidum* (yellow) Chlorophytes-

Chlamydomonas reinhardtii; *Ostreococcus tauri*; *Volvox carteri* (black). See supplementary fig. S1 and S2 for fully annotated trees.

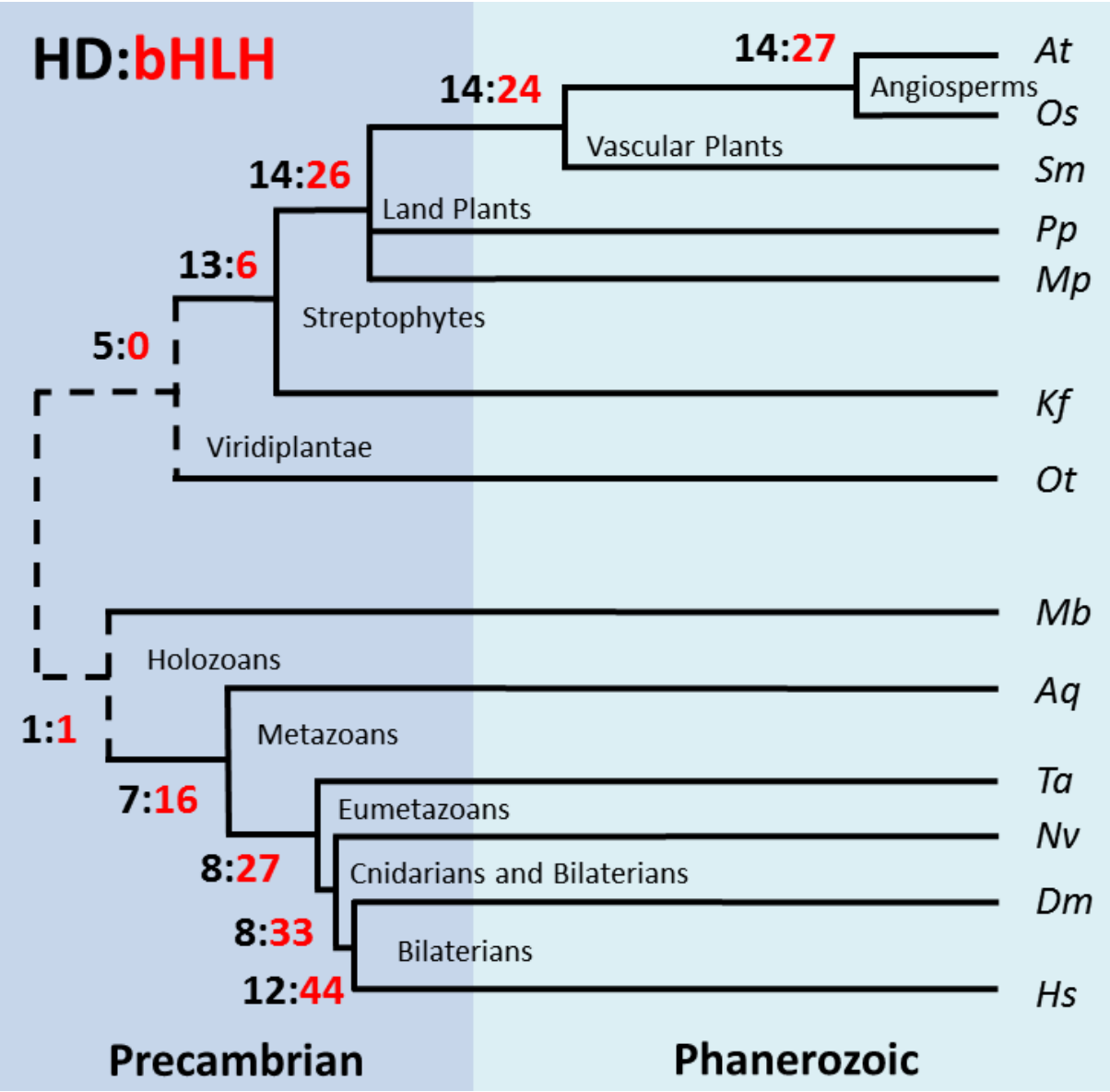

Figure 3. A Precambrian origin the majority of HD and bHLH Transcription Factor subfamilies. The number of HD (black) and bHLH (red) subfamilies present at each node of a simplified time calibrated phylogeny of animals and Archaeplastida. Subfamily number for Archaeplastida based on the current study and for animals based on (Degnan et al., 2009; Larroux et al., 2008; Ryan et al., 2010). Phylogenetic tree for Archaeplastida based on (Cox et al., 2014; Wickett et al., 2014), time calibrations for land plants based on (Clarke et al., 2011), time calibration for the origin of streptophytes based on (Parfrey et al., 2011). Phylogeny and time calibration of animals based on (Erwin et al., 2011). Dashed lines lack time calibrations. Abbreviated taxa names are *At, Arabidopsis thaliana*; *Os, Oryza sativa*; *Sm, Selaginella moellendorffii*; *Pp, Physcomitrella patens*; *Mp, Marchantia polymorpha; Kf ,Klebsormidium flaccidum; Ot, Ostreococcus tauri*; *Aq, Amphimedon queenslandica; Nv, Nematostella vectensis; Sp, Strongylocentrotus purpuratus; Hs, Homo sapiens; Dm, Drosophila melanogaster; Lg, Lottia gigantea*.