# Water-related innovations in land plants evolved by different patterns of gene cooption and novelty

Alexander M. C. Bowles[1,2] (iD), Jordi Paps[1,3] (iD) and Ulrike Bechtold[1,4] (iD)

[1]School of Life Sciences, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK; [2]School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1RL, UK; [3]School of Biological Sciences, University of Bristol, 24 Tyndall Avenue, Bristol, BS8 1TQ, UK; [4]Present address: Department of Biosciences, Durham University, South Road, Durham, DH1 3LE, UK

Authors for correspondence:
*Jordi Paps*
*Email: jordi.paps@bristol.ac.uk*

*Ulrike Bechtold*
*Email: ulrike.bechtold@durham.ac.uk*

## Summary

- The origin of land plants and their descendants was marked by the evolution of key adaptations to life in terrestrial environments such as roots, vascular tissue and stomata. Though these innovations are well characterized, the evolution of the genetic toolkit underlying their development and function is poorly understood.
- We analysed molecular data from 532 species to investigate the evolutionary origin and diversification of genes involved in the development and regulation of these adaptations.
- We show that novel genes in the first land plants led to the single origin of stomata, but the stomatal closure of seed plants resulted from later gene expansions. By contrast, the major mechanism leading to the origin of vascular tissue was cooption of genes that emerged in the first land plants, enabling continuous water transport throughout the ancestral vascular plant. In turn, new key genes in the ancestors of plants with true leaves and seed plants led to the emergence of roots and lateral roots.
- The analysis highlights the different modes of evolution that enabled plants to conquer land, suggesting that gene expansion and cooption are the most common mechanisms of biological innovation in plant evolutionary history.

## Introduction

The first land plants and their descendants have adapted to a multitude of new environments (Morris *et al.*, 2018). Genome analysis has identified that gene homologues for major biological innovations often precede terrestrialization, suggesting that innovations once thought to be land plant-specific may have emerged in older ancestors (e.g. associations with substrate microbiota) (Hori *et al.*, 2014; Nishiyama *et al.*, 2018; Wang *et al.*, 2019; Bowles *et al.*, 2020). Though many genes evolved before the plant transition to land, genetic rewiring of developmental and stress response pathways occurred later in plant evolution, increasing the adaptive plasticity to water availability (Fürst-Jansen *et al.*, 2020). Innovations important for water regulation and transport have evolved in a stepwise manner (Fig. 1). These include the evolution of the cuticle in the ancestor of land plants which acts as an extracellular hydrophobic barrier providing desiccation protection, whilst convergent evolution of leaves within vascular plants (in lycophytes, ferns and seed plants) refined the control of water movement (Harrison, 2017). Three of the most important features for water regulation, and the focus of this study, are stomata, vascular tissue and roots.

Stomata, a key adaptation to dry environments, are pores in plant tissue that enable gas exchange and that regulate water loss (Sussmilch *et al.*, 2019). The origin of stomata is a key adaptation of the first land plants, enabling plants to proliferate in terrestrial environments (Chater *et al.*, 2017). The aperture of stomata is regulated by multiple factors including light, atmospheric $CO_2$ and abscisic acid (ABA) levels (Araújo *et al.*, 2011). Under drought stress, ABA accumulates, triggering a signalling pathway that leads to stomatal closure (Bharath *et al.*, 2021). This control enabled ancestral plants to optimize the balance between $CO_2$ uptake and water loss.

Although stomata emerged in the ancestor of land plants, stomatal function shows clear distinctions between different plant lineages. There is debate about the emergence of active stomatal closure, in which hormonal or environmental cues trigger a network of signalling proteins regulating stomatal aperture (Plackett *et al.*, 2021). Comparative analyses have suggested that ancestral land plants possessed actively controlled stomata (Harris *et al.*, 2020). Early evidence from physiological studies contradicts this, suggesting that stomata closure of vascular plants was hydropassive (Brodribb & McAdam, 2011). However, further physiological studies suggest that lycophytes and ferns do respond to experimentally introduced ABA (Ruszala *et al.*, 2011; Chater *et al.*, 2016; Cai *et al.*, 2017), although studies often used nonphysiological levels of ABA (Brodribb & McAdam, 2017). Regardless of where active stomatal closure emerges, it is undisputed that there are evolutionary distinctions between the stomatal control of distinct plant
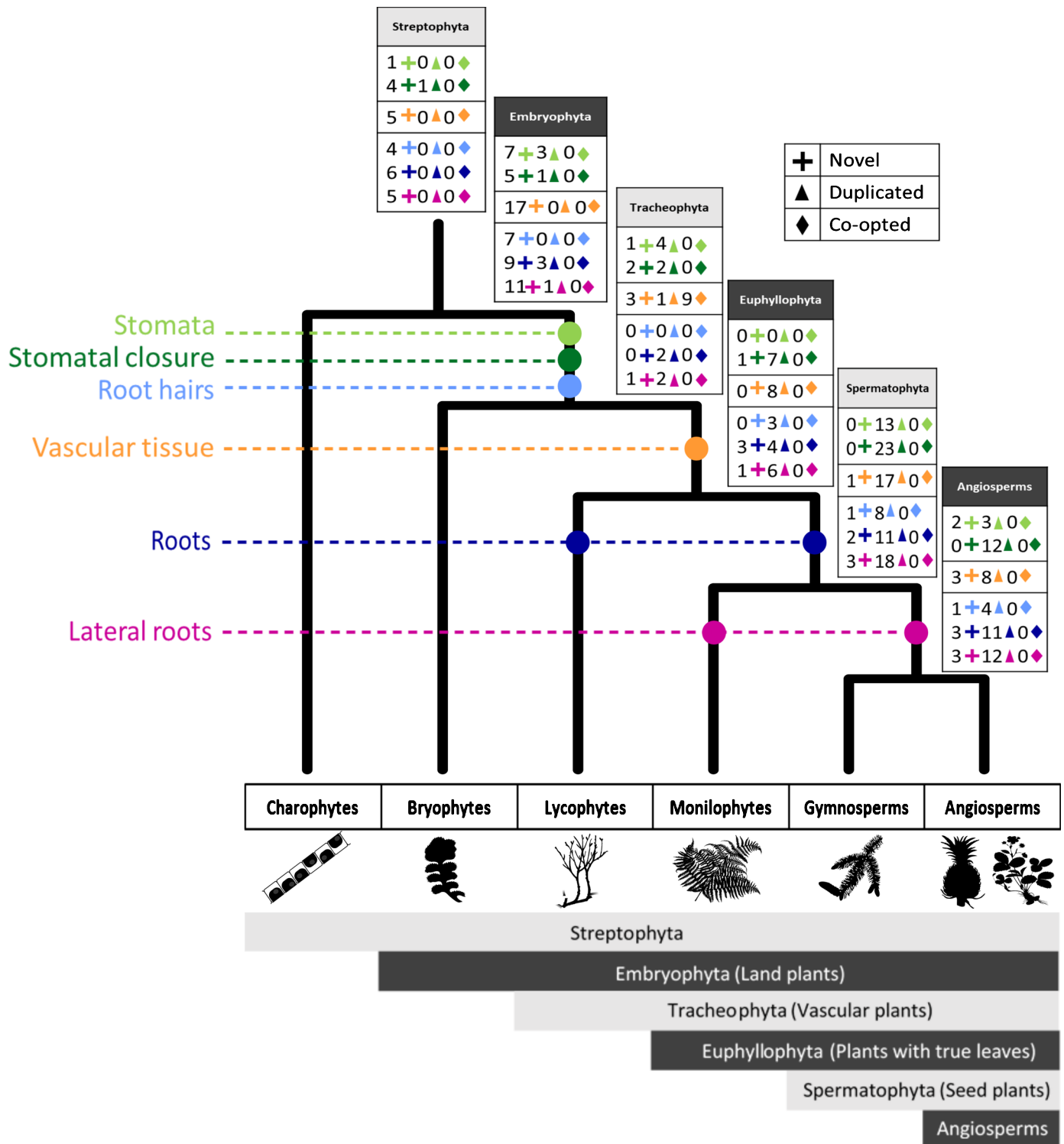
**Fig. 1** Plant–water relationships have evolved in a stepwise manner. The tree demonstrates the evolutionary relationships of plants with silhouettes below (sourced from phylopic.org) illustrating species in each group. The dashed lines leading to each different water regulatory innovation denote their origins. The colour is repeated within the boxes illustrating the genetic mechanisms associated with each innovation (light blue: stomatal development, dark blue: stomatal signalling, orange: vascular tissue development, light green: root hair, green: primary root, dark green: lateral root development). The key demonstrates the numbers of novel, expanded and coopted genes in the boxes. The major plant groups are classified at the bottom in the solid block colour (e.g. Embryophyta, Tracheophyta) and their common names are listed in the dashed boxes beside (e.g. Land plants, Vascular plants).

groups, with seed plant stomata responding rapidly to water availability (Brodribb & McAdam, 2017). This draws into question the origins of stomatal closure in the evolutionary history of plants.

In tracheophytes, also known as vascular plants, xylem and phloem tissues transport water through the plant, which is then evaporated through stomata (Lucas *et al.*, 2013). Lignified vascular tissue also provides mechanical support enabling plants to increase their size and stature and to dominate terrestrial habitats. Indeed, the first trees in the fossil record are early vascular plants (Stein *et al.*, 2007). The current understanding of vascular tissue development reveals a complex and highly coordinated set of genes controlling the specification, patterning and growth of distinct cell types (De Rybel *et al.*, 2016). However, due to this complexity, the evolutionary origin of these genes is poorly understood.

Rooting systems provide multiple functions including anchorage and water uptake (Raven & Edwards, 2001). Distinct components of the rooting system emerged at different points during plant evolution; root hairs were present in the first vascular plants (Kenrick & Strullu-Derrien, 2014), primary roots emerged convergently in lycophytes and plants with true leaves (Hetherington & Dolan, 2018), whilst lateral roots emerged convergently in ferns and seed plants (Hetherington *et al.*, 2020). The first land plants had a rhizoid-based rooting system that has been distinguished from the root hairs of vascular plants (Kenrick & Strullu-Derrien, 2014). Sequence and functional homology have been identified for root hair development genes in the rhizoid development of *Physcomitrium patens*, but the structural homology between rhizoids and root hairs is unclear (Menand *et al.*, 2007; Tam *et al.*, 2015; Honkanen *et al.*, 2016). The piecemeal and convergent evolution of the rooting system is underpinned by a toolkit of genes whose origin and molecular evolution is poorly understood.

As mentioned above, the evolution of these three innovations, namely roots, vascular tissue and stomata, is underpinned by plant genome evolution. As recently highlighted by the one thousand plant transcriptomes (1KP) project, which analysed patterns of gene diversification for the 23 largest plant gene families, ancestral streptophyte and embryophyte genomes were characterized by gene novelty whilst the genomes of younger plant ancestors were distinguished by gene and genome duplication (Leebens-Mack *et al.*, 2019). These patterns of novelty are also observed genome-wide by comparative genomic analyses using complete genomes (Bowles *et al.*, 2020). However, not much is known about the mode of evolution – gene novelty vs expansion vs cooption – of the genetic components involved in stomata, vascular tissue and roots. Understanding the evolution of key developmental and signalling genes is the first step in reconstructing the evolutionary history of these innovations. Here, to investigate the modes of genome evolution of important features for adapting to life on land, we use a comparative approach, incorporating genomic and transcriptomic data representing 532 species from across the plant tree of life. Our results show that different patterns of genome evolution drive the emergence of each of the genetic toolkits of stomata, vascular tissue and roots.

## Materials and Methods

### Homology assignment

The pipeline approach has previously been described and used to investigate genome evolution (Paps & Holland, 2018; Bowles *et al.*, 2020; Guijarro-Clarke *et al.*, 2020). Briefly, proteins were extracted for 208 genomes, including 178 land plant species and broad outgroup sampling, and similarity between proteins was identified with an all-vs-all BLAST search (Altschul *et al.*, 1990). Sequences were clustered into Homology Groups (HGs) using Markov Clustering (MCL) with a granularity score of 2 (Enright *et al.*, 2002). MCL uses graph theory and hidden Markov models to cluster proteins into groups based on the output of the BLAST comparisons. These HGs will contain large gene families, with orthologues and paralogues. This method does have limitations shared with similar BLAST-based approaches, including the lack of detection of lateral gene transfer or gene fusion/fissions. Further division of the HGs (e.g. split into orthologues and paralogues) was not pursued at this stage, instead being completed by gene tree analysis as described below.

### Expanding taxon sampling

To accommodate the ever-growing representation of plant omics data, the taxon sampling was expanded beyond our original sampling of 208 complete genomes (Bowles *et al.*, 2020). First, data from the 1KP project were sourced and binned into seven local BLAST databases (charophytes, hornworts, liverworts, mosses, lycophytes, ferns, gymnosperms) (Leebens-Mack *et al.*, 2019). Second, BLAST databases for newly published evolutionary significant genomes were constructed, grouped into streptophytes (Supporting Information Table S1; Dataset S1): *Chlorokybus atmophyticus*, *Mesostigma viride* (Wang *et al.*, 2019), *Chara braunii* (Nishiyama *et al.*, 2018), *Penium margaritaceum* (Jiao *et al.*, 2020), *Spirogloea muscicola* and *Mesotaenium endlicherianum* (Cheng *et al.*, 2019); bryophytes: *Anthoceros angustus* (Zhang *et al.*, 2020), *Anthoceros agrestis*, *Anthoceros punctactus* (Li *et al.*, 2020), *Fontinalis antipyretica* (Yu *et al.*, 2020a) and *Pleurozium schreberi* (Pederson *et al.*, 2019); ferns: *Azolla filiculoides* and *Salvinia cucullata* (Li *et al.*, 2018); and gymnosperms: *Abies alba* (Mosca *et al.*, 2019), *Picea glauca* (Warren *et al.*, 2015), *Picea lambertiana* (Stevens *et al.*, 2016), *Picea taeda* (Zimin *et al.*, 2014), *Pseudotsuga menziesii* (Neale *et al.*, 2017) and *Sequoiadendron giganteum* (Scott *et al.*, 2020). The *Arabidopsis thaliana* genes from each HG inferred from the 208 genomes dataset were extracted and searched against each of these databases using reciprocal best-hit BLAST, specifying a maximum of 20 species per BLAST database (Tatusov *et al.*, 1997). Therefore, in this analysis, in total, 227 genomes and 305 nonangiosperm plant transcriptomes were analysed, representing all major streptophyte lineages.

### Genetic toolkit of stomatal development and signalling, vascular tissue development and root development

The literature was searched to identify genes involved in developmental and signalling pathways of stomata, vascular tissue and

roots. For stomatal development genes, a composite list was made from Lau & Bergmann (2012); Le *et al.* (2014) and Chater *et al.* (2017). Stomatal signalling genes were identified from Cotelle & Leonhardt (2016); Albert *et al.* (2017) and Cai *et al.* (2017). For the development of vascular tissues, genes from Ruonala *et al.* (2017) were used. Root development genes were identified in Jung & McCouch (2013) whilst root hair development genes were identified in Vissenberg *et al.* (2020). For lateral root development, genes from Verstraeten *et al.* (2014); Oh *et al.* (2018) and Santos Teixeira *et al.* (2019) were used.

In the previous research, the 208 genomes dataset was queried based solely on taxonomic occupancy (Bowles *et al.*, 2020). New to this study, the dataset was queried for the list of candidate genes compiled above. Specifically, HGs were extracted based on the Uniprot IDs for each gene using a new computational script, MCL_search_by_gene_name_2.pl (Dataset S2) (UniProt Consortium, 2018). Note that gene counts for individual species may be overinflated by the merging of sequences with subspecies data and by multiple systems of nomenclature used during gene annotation. Information about the accession and sequence details for all HGs is available in Dataset S3, as discussed below.

## Detecting mechanisms of gene evolution

Three evolutionarily distinct mechanisms of gene evolution were subsequently identified. Novel HGs are here defined as a set of genes present in the last common ancestor (LCA) of a clade and absent in all outgroups; in the context of the investigated innovations, the novel genes and their related innovations emerge in the same node of the evolutionary tree. A coopted HG is defined as a set of genes traditionally associated with a biological innovation (e.g. vascular tissue) whose origin pre-dates the emergence of such innovation (e.g. HG originating in the LCA of land plants linked to vascular tissue development). To investigate the diversification of genes, gene phylogenies were inferred and examined (see next section; Dataset S3). From this, an expanded HG was defined as a group of genes that underwent gene duplications in the LCA of the clade of interest linked to a biological innovation.

## Gene tree inference

A curated list of species in the original genomic dataset was collated with representatives for each major plant group. These were *Cyanidioschyzon merolae* (Rhodophyta), *Cyanophora paradoxa* (Glaucophyta), *Bathycoccus prasinos* (Chlorophyta), *Klebsormidium flaccidum* (now called *K. nitens*, Charophyta), *Marchantia polymorpha*, *Physcomitrium patens* (bryophytes), *Selaginella moellendorffii* (Lycophyta), *Picea abies* (gymnosperms), *Amborella trichopoda* (ANA grade), *Oryza sativa indica*, *Brachypodium distachyon* (monocots) and *A. thaliana* (eudicots). These HGs were extracted using the MCL_search_by_gene_name_2.pl script (Dataset S2) (UniProt Consortium, 2018). A file containing all gene IDs for all species was then used to extract the protein sequences using the PERL one liner 'perl -ne 'if(/^>(\S+)/){$c=$i{$1}}$c?print:chomp;$i{$_}=1 if @ARGV' ids.file all_fasta.file'. To each HG fasta file, the sequences from the reciprocal BLAST

queries were added, incorporating data from the 1KP transcriptomes and additionally published genomes (Dataset S1).

To investigate expanded HGs, sequences from each HG were aligned using MAFFT using the –auto parameter, which automatically selects an appropriate alignment strategy as well as the leave gappy region parameter (Katoh *et al.*, 2002). Multiple sequence alignments were trimmed with TRIMAL using the gappyout method, to identify and remove poorly aligned positions. Gene trees (bootstrapped maximum-likelihood phylogenies) were inferred using IQ-TREE, using the inbuilt MODELFINDER (MFP) to select the best-fitting substitution model, specifying 1000 ultrafast bootstrap replicates (Nguyen *et al.*, 2015). Trees were rooted against the latest plant phylogeny to infer gene innovations, expansions and cooptions. Trees were visualized in iTOL (Letunic & Bork, 2019) and are available at https://itol.embl.de/shared/Bowles_et_al.

## Results

### Gene novelty suggests that stomata evolved once in the ancestor of land plants

Stomata in bryophytes (liverworts, mosses and hornworts) demonstrate a patchy distribution with absences in liverworts and some mosses, but are found in all vascular plants (Duckett & Pressel, 2018). This raises questions about the single origin of stomata in the first land plants or their convergent evolution in the ancestors of vascular plants, mosses and hornworts. Here, our analysis shows the stomatal development pathway originated in the LCA of land plants, based on inferences from a taxonomically broad representation of genomes and transcriptomes. Of the 23 stomatal development genes, 21 pre-dated or accompanied the origin of land plants (Figs 2a,b, S1; Dataset S2). In *A. thaliana*, the basic helix–loop–helix (bHLH) genes SPCH, MUTE and FAMA are required consecutively to determine stomatal development, aided by the bHLH transcription factor SCREAM (Fig. 2a,b). In our analysis, SCREAM was identified in all land plants, even the liverwort *M. polymorpha*, which does not have stomata. MUTE, SPCH and FAMA were present in the LCA of land plants but were not recovered in *M. polymorpha*, potentially accompanying stomatal loss in liverworts, as previously observed (Chater *et al.*, 2017). The evolutionary age of these 21 stomatal genes suggests that bryophyte stomata may develop in a similar manner to vascular plant stomata, and by extension the stomata of the first land plants. STOMAGEN, which regulates stomatal density (Sugano *et al.*, 2010), appears in the origin of vascular plants, the first time stomata appear on leaf-like structures (Fig. 2a,b). Of the three features investigated, stomata were the only innovation with a strong association with novel genes.

### Gene expansion enabled stomatal control in the ancestor of seed plants

Gene families involved in stomatal closure experienced multiple gene duplications, predominantly in the ancestor of seed plants, which could precede the neo- or subfunctionalization of the

**(a)**



**(b)**

**(c)**
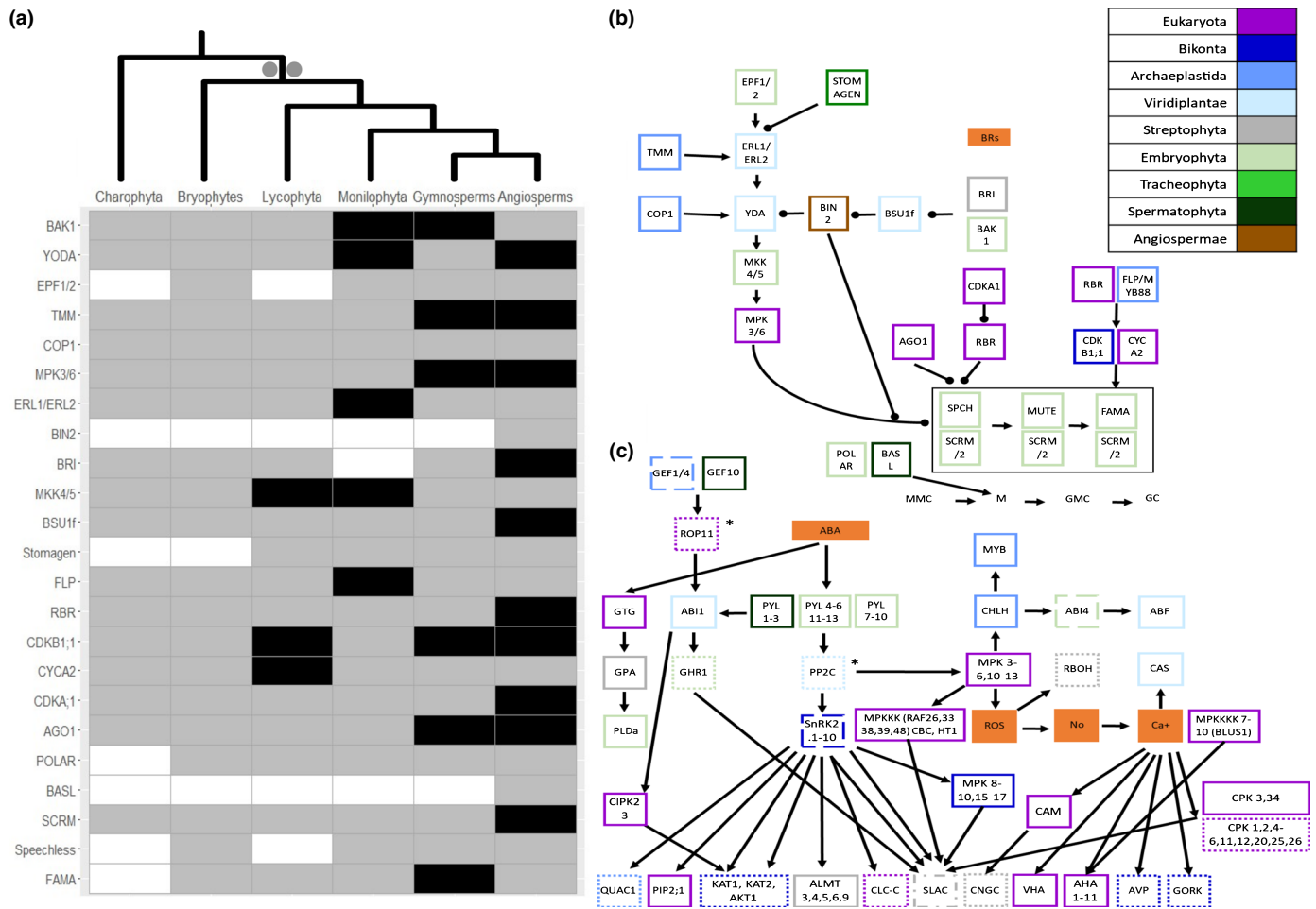
**Fig. 2** The genomic basis of the evolutionary development of stomata and stomata signalling. (a) Heatmap displaying absence (white), partial presence (grey) and presence (black) in all species for both transcriptomic and genomic data for genes involved in stomatal development (Supporting Information Figs S1, S2). The tree at the top illustrates plant evolutionary relationships and the two grey dots denote the origin of stomata and stomata signalling (grey). (b) Genetic network leading to the development of stomata. Each gene is coloured based on its phylogenetic appearance. Nongenetic components are coloured in solid orange. Lines ending in circles denote negative interactions whilst lines ending in arrows denote activation. The states of guard cell development are highlighted: MMC (meristemoid mother cell), M (meristemoid), GMC (guard mother cell) and GC (guard cell). (c) Genetic network involved in stomatal signalling. Expanded genes in the ancestor of Spermatophyta are highlighted by boxes with dotted edges. Expanded genes in the ancestor of Euphyllophyta are highlighted by boxes with dashed edges. Expanded genes in the ancestor of Tracheophyta are highlighted by boxes with dot–dash–dot edges. Asterisks indicate an HG that expands twice, once in the ancestor of Euphyllophyta and again in the ancestor of Spermatophyta. Nongenetic components are coloured in solid orange. Lines ending in circles denote negative interactions whilst lines ending in arrows denote activation.

duplicated genes (Figs 2c, S2; Dataset S3). Genes involved in the signalling of some potassium and anion channels (QUAC1, KAT2, AKT1, GORK) were present before the origin of land plants and expanded in the ancestor of seed plants (Fig. 2; Dataset S3). Additionally, important for stomatal closure is a group of PYLs (1–3) (PYR1-LIKE 1), an ABA receptor, which emerged in the ancestor of seed plants (Fig. 2). The remaining PYLs (4–13) emerged in the ancestor of land plants (Fig. 2c).

Although most gene expansions were identified in the LCA of seed plants, several notable expansions were found in the ancestor of plants with true leaves. One such example is SNRK2 (SNF1-RELATED PROTEIN KINASE 2), a family of genes including OST1 (SNRK2.6), which is a core element of the ABA=dependent signalling pathway (Fig. 2c). Furthermore, an HG containing PP2Cs (Protein Phosphatase 2Cs), which regulate ABA-dependent activation, expanded in the ancestor of plants with

true leaves (Fig. 2c). Our data specifically highlight the genetic rewiring of the PP2C–SnRK2 complex as the central regulator of ABA signalling in the ancestor of plants with true leaves. These findings suggest that gene expansions, especially in ABA signalling pathways, played an essential role in the evolution of stomatal closure, allowing plants to preserve water by restricting transpiration rates. These may be the product of whole genome duplications in the ancestor of seed plants (Leebens-Mack *et al.*, 2019) or frequent gene duplication (Ezoe *et al.*, 2020).

## Vascular tissue evolved through a complex of genetic mechanisms

Only two HGs (SACL1: SUPPRESSOR OF ACAULIS 51; and WOX1: WUSCHEL-related homeobox 1) involved in vascular tissue development originated in the ancestor of vascular plants

(Figs 3a, S3; Dataset S3). Several genes involved in vascular tissue development emerged in the ancestor of land plants, showing a patchy distribution in nonvascular plants but being present in all the vascular plant species sampled in our analyses (176 tracheophyte genomes). This retention in the LCA of vascular plants suggests a vital function in the biology of this group. Nine HGs fitted this criterion of cooption (Fig. 1). These HGs were found across all vascular development pathways, suggesting that vascular tissue emerged through a complex of evolutionary mechanisms. These findings demonstrate the importance of gene expansions and repurposing of old genes for novel functions in the evolutionary development of the vascular system. Together these genetic processes contributed to the origin and evolution of one of the most successful plant groups, ~450 million years ago (Morris *et al.*, 2018).

### Novel genes enabled the evolution of roots

Our analyses show that novel genes in land plants enabled the evolutionary development of root hairs (Figs 4, S4). CAPRICE (CPC), TRIPTYCHON (TRY) and ENHANCER OF TRY

AND CPC 1 (ETC1) are together in a single HG and emerge in the ancestor of seed plants, to promote root hair cell differentiation. This suggests that nonseed plants develop root hairs without these genes and that greater control of development evolved in the LCA of seed plants.

Fossil evidence supports at least two origins of roots in the evolutionary history of plants, once in the ancestor of lycophytes and again in the ancestor of plants with true leaves (Hetherington & Dolan, 2018). Analysis of genes involved in primary root development revealed that many were conserved across land plants (Figs 4a, S5). Two HGs, ARABIDOPSIS RESPONSE REGULATOR 12 (ARR12) and LATERAL ROOT ORGAN DEFECTIVE (LATD), have emerged in the LCA of euphyllophytes (Dataset S4). Both HGs have been shown to modulate primary and lateral root growth and development, with responses to ABA and water deprivation (Léran *et al.*, 2014; Nguyen *et al.*, 2016).

The data demonstrate that the majority of lateral root development genes pre-date the emergence of lateral roots, originating in the ancestors of Streptophyta (e.g. PIN1–3) and land plants (e.g. IAA 8, 14), which contribute to other functions in these rootless
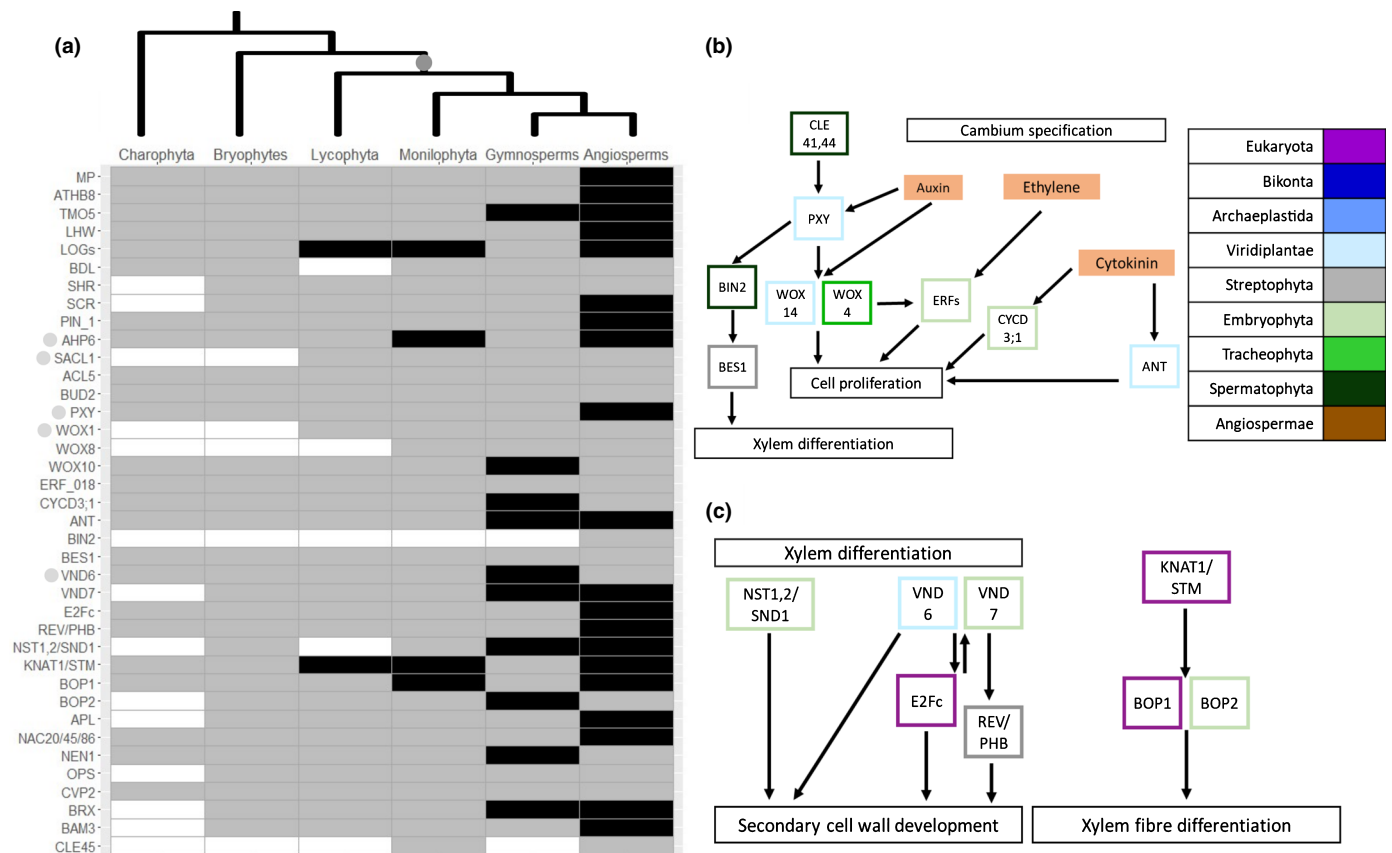


**Fig. 3** The genomic basis of the evolutionary development of vascular tissue. (a) Heatmap displaying absence (white), partial presence (grey) and presence (black) in all species for both transcriptomic and genomic data for genes involved in vascular tissue development (Supporting Information Fig. S3). The tree at the top illustrates plant evolutionary relationships and the origin of vascular tissue (grey). Genes discussed in the text are highlighted by a grey dot. (b) Genetic network involved in cambium specification. Each gene, in panels (b) and (c), is coloured based on its phylogenetic appearance. Expanded genes are highlighted by boxes with dotted edges and coopted genes are highlighted by boxes with dashed edges. Lines ending in circles denote negative interactions whilst lines ending in arrows denote activation. (c) Genetic network involved in xylem differentiation. Nongenetic components are coloured in solid orange. Lines ending in circles denote negative interactions whilst lines ending in arrows denote activation.
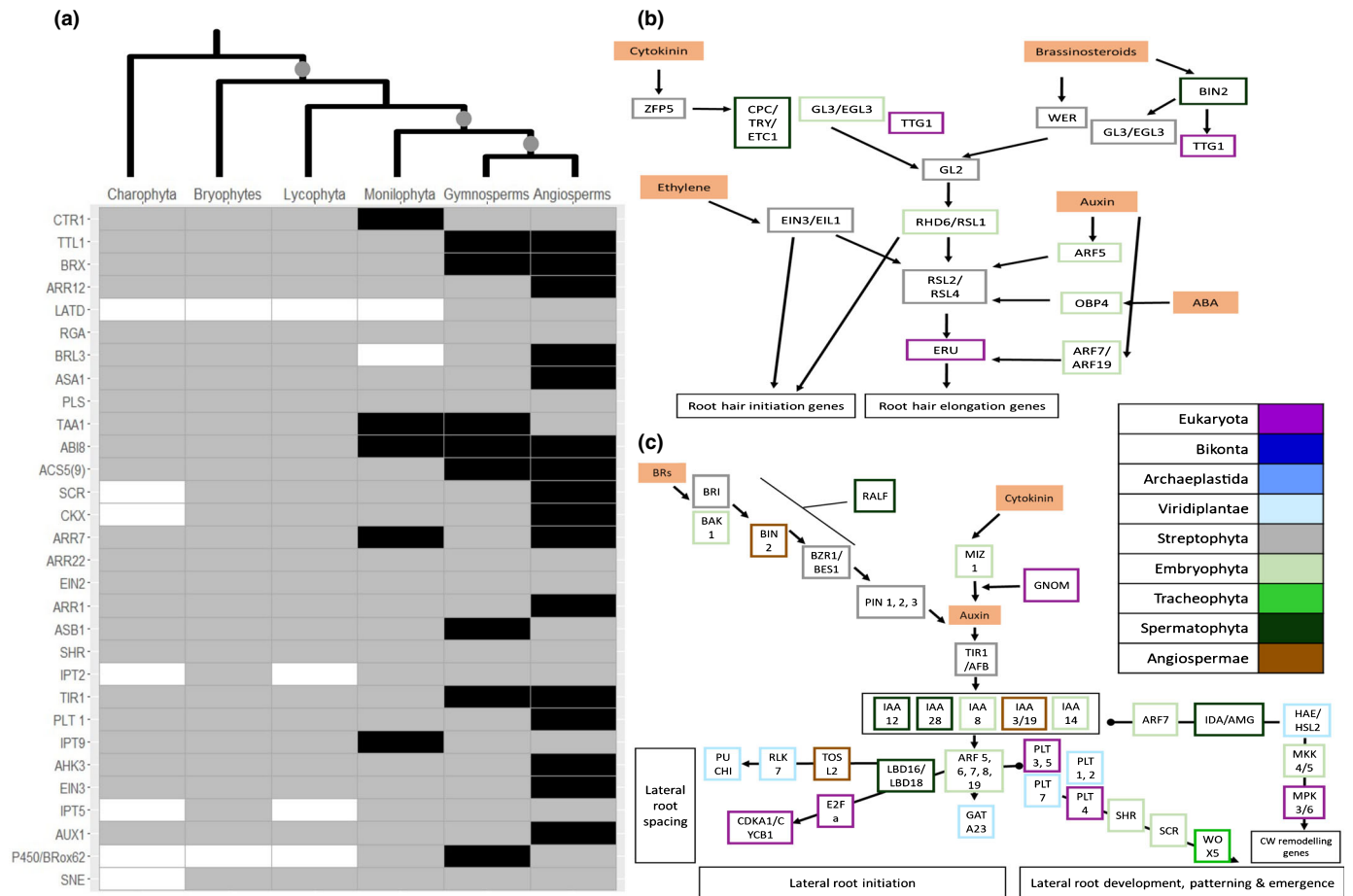
**Fig. 4** The genomic basis of the evolutionary development of roots, root hairs and lateral roots. (a) Heatmap displaying absence (white), partial presence (grey) and presence (black) in all species for both transcriptomic and genomic data for genes involved in primary root development (Supporting Information Figs S4–S6). The tree at the top illustrates plant evolutionary relationships and the grey dots denote the origin of root hairs, roots and lateral roots. (b) Genetic network leading to the development of root hairs. Each gene, in Fig. 2(b,c), is coloured based on its phylogenetic appearance. Nongenetic components are coloured in solid orange. Lines ending in circles denote negative interactions whilst lines ending in arrows denote activation. (c) Genetic network leading to the development of lateral roots. Lines ending in circles denote negative interactions whilst lines ending in arrows denote activation.

plants (Mutte *et al.*, 2018). Three key genes appeared with the origin of lateral roots in the ancestor of seed plants: INDOLEACETIC ACID-INDUCED PROTEIN 12 and 28 (IAA12, 28), INFLORESCENCE DEFICIENT IN ABSCISSION (IDA) and RAPID ALKALINIZATION FACTOR (RALF; Dataset S4). These three HGs are involved in multiple stages of lateral root development, suggesting their emergence contributed to the origin of lateral roots.

Hydrotropism is the directional growth of plants towards water, enabling acquisition of water under drought stress (Dietrich *et al.*, 2017), an important target trait for drought avoidance. The two key genes essential for hydrotropism are MIZU-KUSSEI 1 (MIZ1) and MIZU-KUSSEI 2 (MIZ2/GNOM). The data presented infer that MIZ1 emerged and diversified through expansion in the ancestor of land plants (Fig. 4c; Dataset S3). This suggests that rooting system hydrotropism may have been important during plant terrestrialization (Shkolnik *et al.*, 2018). The HG containing MIZ2/GNOM was present in the ancestor of eukaryotes, with genes expanding in the ancestor of seed

plants, potentially responsible for lateral root hydrotropism (Fig. 4c; Dataset S3). The development of increasingly morphologically complex rooting and response systems in the ancestors of land plants, plants with true leaves and seed plants has enabled access to previously unavailable water sources. The data presented above demonstrate that a combination of novel genes is responsible for the emergence of plants with true roots with lateral branches that are capable of responding to water gradients in the soil.

## Discussion

Overall, these analyses provide novel insights into the molecular evolution of three morphological innovations fundamental for plant life. By combining these inferences about patterns of gene evolution for stomata, vascular tissue and roots, insights into plant genome evolution more broadly can be garnered. The most striking revelation is that each of these morphological innovations is underpinned by different patterns of gene evolution. We

demonstrated that novel genes led to the origin of stomata (Fig. 2a,b) in the first land plants, but stomatal control developed over the course of plant evolutionary history, with gene expansions leading to rapid closure in the first seed plants (Fig. 2c). The development of vascular tissue, conversely, is mostly associated with cooption, the repurposing of old genes for new functions (Fig. 3). Another striking pattern is that, aside from stomata, very few novel genes emerged simultaneously with the morphological innovation with which they were associated (Fig. 1). This suggests that gene expansion and gene cooption are the more common mechanisms of biological innovation throughout plant evolutionary history.

Comparative analysis of the stomatal development pathway recovered most genes as originating in the ancestor of land plants, suggesting that the stomata of the first land plants and extant ones developed in a similar manner. This reinforces studies using transcriptome and other sequence data (Chater et al., 2017; Harris et al., 2020). Due to the single origin of stomata in the ancestor of land plants, this would suggest that reductive processes contributed to stomatal evolution in bryophytes as identified by Harris et al. (2020). As highlighted above, the evolution of stomatal closure has been widely debated. Perhaps most importantly, we find the vast majority of stomatal signalling genes in the LCA of land plants and older ancestors, suggesting an ancestral, conserved function in the land plant ancestor. These inferences are supported by physiological studies of early diverging land plants (Chater et al., 2011; Ruszala et al., 2011; Cai et al., 2017). These findings echo those of Harris et al. (2020) who conducted similar work for 18 stomatal function genes using a reduced sequence dataset compared to the 34 genes investigated in this study. A gene shared between these analyses is OST1 (SnRK2.6), which in Harris et al. (2020) is found in single copy emerging in the ancestor of Streptophyta. This finding is developed further in this study, which identifies all SnRK2s in a multigene family emerging in the ancestor of Viridiplantae. This is due to the homology approach compared to the orthology approach of Harris et al. (2020). Our results suggest that the evolutionary trajectory of stomata has been shaped by gene expansions leading to gene neo- or subfunctionalization in younger ancestors. Additional genes investigated include multiple PYLs and SnRK2s, key components in ABA signalling, the latter of which were found to duplicate in the ancestor of plants with true leaves (Fig. 2). This could explain why there is an evolutionary distinction between the stomatal control of plants with true leaves and older ancestors. The data here also support widescale gene expansion in the ancestor of seed plants, suggesting a difference between the stomatal control of seed plants and other plants.

Gene cooption was identified as the major mechanism leading to the evolution of vascular tissue (Fig. 3). This suggests that this new morphological structure was the product of the repurposing of a preexisting genetic toolkit already found in land plants. Importantly, there is recent experimental evidence for the redeployment of several individual genes essential for vascular cell development (e.g. TMO5/LHW; Lu et al., 2020), supporting our findings of gene cooption as a major mechanism in vascular tissue development. An alternative hypothesis to the origin of

vascular tissue in plants suggests that vascular-like tissue is present in bryophytes and therefore potentially in the LCA of land plants (Brodribb et al., 2020). This is based on the patchy distribution of water- and food-conducting cells found in the bryophytes (Ohtani et al., 2017), potentially having a similar evolutionary history to stomata, which have undergone reductive evolution in the bryophytes (Harris et al., 2020). Therefore, the coopted genes identified in this study could represent novel land plant genes if vascular-like tissue was present in the LCA of land plants.

All, bar one, root hair development genes were present in the ancestor of land plants. CPC, TRY and ETC1 are in a single HG and emerge in the ancestor of seed plants, to promote root hair cell differentiation by repressing GLABRA 2 and 3 (Tominaga et al., 2008). This suggests that nonspermatophyte land plants develop root hairs without these genes and that greater control of root hair development evolved in the LCA of seed plants. Only two genes, ARR12 and LATD, originated in the ancestor of plants with true leaves, accompanying the origin of primary roots. ARR12 regulates cell differentiation and meristem growth (Yokoyama et al., 2007; Moubayidin et al., 2010). Under drought, ARR12 is downregulated as an adaptive mechanism to control root growth to cope with water deficit (Nguyen et al., 2016). LATD is also required for lateral root and nodule meristem development (Léran et al., 2014) and, in concert with ABA, modulates primary root elongation (Zhang et al., 2014), thus suggesting a potential mechanism for controlling root growth under water stress conditions. These genes therefore play crucial roles in plant growth and development but also in response to water availability. Three HGs emerged at the same time as lateral roots, IAA12, IDA and RALF. IAA12 and IAA28 are auxin-responsive proteins important for the production of lateral root primordia and optimizing the distribution of new root organs (De Rybel et al., 2010; Stoeckle et al., 2018). IDA is required for cell wall dissolution, by facilitating the separation of epidermal tissues, enabling lateral root emergence (Zhu et al., 2019). RALF1 inhibits cell elongation for lateral root formation and density, particularly under drought stress (Li et al., 2019). Under drought stress, RALF1 modulates root hair growth and cell size, and together with other RALFs is implicated in stress responses of lateral roots (Murphy & De Smet, 2014; Zhu et al., 2020). These three HGs are involved sequentially in the initiation and formation of lateral roots, and therefore the emergence of these genes enabled the evolution of lateral roots (Fig. 4a,c).

In this study, we investigated the evolutionary development of stomata, vascular tissue and roots using genes characterized in flowering plants, specifically based on A. thaliana. As these genes are involved in a known function in A. thaliana, genes in other organisms within an HG were assumed to be functionally homologous. Indeed, several genes have been experimentally demonstrated to have conserved functions across the major plant lineages (i.e. ABA signalling genes in ferns (Cai et al., 2017), stomatal development genes in mosses (Caine et al., 2020), primary root development in ferns (Yu et al., 2020b)). However, there are limitations with using sequence homology as a predictor of gene function (Gabaldón & Koonin, 2013). Whilst comparative genomics can be used to predict functionally analogous

genes, experimental validation is needed to fully characterize gene function. In spite of these caveats, this study utilizes genome-scale data from the largest plant genome study to date (Bowles *et al.*, 2020), incorporating additional transcriptomic and genomic data providing broad taxonomic coverage for all major groups, to investigate the molecular evolution of the relationship of plants with water.

The ancestor of land plants had a very limited ability to regulate water content. The evolution of stomata, vascular tissue and roots increased the capacity of water transport and regulation of seed plants. The development of these features at every major step in the evolutionary history of plants highlights the role of water availability as a driver of plant evolution. Here the results demonstrate that gene novelty, expansion and cooption contribute differently to distinct steps in the evolution of water regulatory traits. Some of these genes evolved at the same time as the morphological innovation they are associated with (i.e. stomatal development genes), while others are older, indicating that cooption was concomitant with the evolution of these traits. Overall, our analyses shed new light on the genetic basis of the evolution of life on land, highlighting the role of genome dynamics in the diversification of the plant kingdom.

## Acknowledgements

## Author contributions

AMCB, JP and UB designed the study and analyses. AMCB performed the analyses. AMCB, JP and UB wrote the manuscript.

## ORCID

Ulrike Bechtold https://orcid.org/0000-0003-2320-3890
Alexander M. C. Bowles https://orcid.org/0000-0002-7487-3811
Jordi Paps https://orcid.org/0000-0003-2636-6636

## Data availability

The data that support the findings of this study are available in the Supporting Information of this article.

## References

**Albert R, Acharya BR, Jeon BW, Zañudo JGTT, Zhu M, Osman K, Assmann SM. 2017.** A new discrete dynamic model of ABA-induced stomatal closure predicts key feedback loops. *PLoS Biology* **15**: e2003451.

**Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990.** Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

**Araújo WL, Fernie AR, Nunes-Nesi A. 2011.** Control of stomatal aperture. *Plant Signaling & Behavior* **6**: 1305–1311.

**Bharath P, Gahir S, Raghavendra AS. 2021.** Abscisic acid-induced stomatal closure: an important component of plant defense against abiotic and biotic stress. *Frontiers in Plant Science* **12**: 615114.

**Bowles AMC, Bechtold U, Paps J. 2020.** The origin of land plants is rooted in two bursts of genomic novelty. *Current Biology* **30**: 530–536.

**Brodribb TJ, Carriquí M, Delzon S, McAdam SAM, Holbrook NM. 2020.** Advanced vascular function discovered in a widespread moss. *Nature Plants* **6**: 273–279.

**Brodribb TJ, McAdam SAM. 2011.** Passive origins of stomatal control in vascular plants. *Science* **331**: 582–585.

**Brodribb TJ, McAdam SAM. 2017.** Evolution of the stomatal regulation of plant water content. *Plant Physiology* **174**: 639–649.

**Cai S, Chen G, Wang Y, Huang Y, Marchant DB, Wang Y, Yang Q, Dai F, Hills A, Franks PJ et al. 2017.** Evolutionary conservation of ABA signaling for stomatal closure. *Plant Physiology* **174**: 732–747.

**Caine RS, Chater CCC, Fleming AJ, Gray JE. 2020.** Stomata and sporophytes of the model moss *Physcomitrium patens*. *Frontiers in Plant Science* **11**: 643.

**Chater CC, Caine RS, Fleming AJ, Gray JE. 2017.** Origins and evolution of stomatal development. *Plant Physiology* **174**: 624–638.

**Chater CC, Caine RS, Tomek M, Wallace S, Kamisugi Y, Cuming AC, Lang D, MacAlister CA, Casson S, Bergmann DC et al. 2016.** Origin and function of stomata in the moss *Physcomitrella patens*. *Nature Plants* **2**: 1–7.

**Chater C, Kamisugi Y, Movahedi M, Fleming A, Cuming AC, Gray JE, Beerling DJ. 2011.** Regulatory mechanism controlling stomatal behavior conserved across 400 million years of land plant evolution. *Current Biology* **21**: 1025–1029.

**Cheng S, Xian W, Fu Y, Marin B, Keller J, Wu T, Sun W, Li X, Xu Y, Zhang YU et al. 2019.** Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. *Cell* **179**: 1057–1067.

**Cotelle V, Leonhardt N. 2016.** 14-3-3 proteins in guard cell signaling. *Frontiers in Plant Science* **6**: 1210.

**De Rybel B, Mähönen AP, Helariutta Y, Weijers D. 2016.** Plant vascular development: from early specification to differentiation. *Nature Reviews Molecular Cell Biology* **17**: 30–40.

**De Rybel B, Vassileva V, Parizot B, Demeulenaere M, Grunewald W, Audenaert D, Van Campenhout J, Overvoorde P, Jansen L, Vanneste S et al. 2010.** A novel *Aux/IAA28* signaling cascade activates *GATA23*-dependent specification of lateral root founder cell identity. *Current Biology* **20**: 1697–1706.

**Dietrich D, Pang L, Kobayashi A, Fozard JA, Boudolf V, Bhosale R, Antoni R, Nguyen T, Hiratsuka S, Fujii N et al. 2017.** Root hydrotropism is controlled via a cortex-specific growth mechanism. *Nature Plants* **3**: 1–8.

**Duckett JG, Pressel S. 2018.** The evolution of the stomatal apparatus: intercellular spaces and sporophyte water relations in bryophytes—two ignored dimensions. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **373**: 20160498.

**Enright AJ, Van Dongen S, Ouzounis CA. 2002.** An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* **30**: 1575–1584.

**Ezoe A, Shirai K, Hanada K. 2020.** Degree of functional divergence in duplicates is associated with distinct roles in plant evolution. *Molecular Biology and Evolution* **38**: 1447–1459.

**Fürst-Jansen JMR, de Vries S, de Vries J. 2020.** Evo-physio: on stress responses and the earliest land plants. *Journal of Experimental Botany* **71**: 3254–3269.

**Gabaldón T, Koonin EV. 2013.** Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics* **14**: 360–366.

**Guijarro-Clarke C, Holland PWH, Paps J. 2020.** Widespread patterns of gene loss in the evolution of the animal kingdom. *Nature Ecology & Evolution* **4**: 519–523.

**Harris BJ, Harrison CJ, Hetherington AM, Williams TA. 2020.** Phylogenomic evidence for the monophyly of bryophytes and the reductive evolution of stomata. *Current Biology* **30**: 2001–2012.

Harrison CJ. 2017. Development and genetics in the evolution of land plant body plans. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 372: 20150490.

Hetherington AJ, Berry CM, Dolan L. 2020. Multiple origins of dichotomous and lateral branching during root evolution. *Nature Plants* 6: 454–459.

Hetherington AJ, Dolan L. 2018. Stepwise and independent origins of roots among land plants. *Nature* 561: 235–238.

Honkanen S, Jones VAS, Morieri G, Champion C, Hetherington AJ, Kelly S, Proust H, Saint-Marcoux D, Prescott H, Dolan L. 2016. The mechanism forming the cell surface of tip-growing rooting cells is conserved among land plants. *Current Biology* 26: 3238–3244.

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. *Nature Communications* 5: 3978.

Jiao C, Sørensen I, Sun X, Sun H, Behar H, Alseekh S, Philippe G, Palacio Lopez K, Sun LI, Reed R et al. 2020. The *Penium margaritaceum* genome: hallmarks of the origins of land plants. *Cell* 181: 1097–1111.

Jung JKHH, McCouch S. 2013. Getting to the roots of it: genetic and hormonal control of root architecture. *Frontiers in Plant Science* 4: 186.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30: 3059–3066.

Kenrick P, Strullu-Derrien C. 2014. The origin and early evolution of roots. *Plant Physiology* 166: 570–580.

Lau OS, Bergmann DC. 2012. Stomatal development: a plant's perspective on cell polarity, cell fate transitions and intercellular communication. *Development* 139: 3683–3692.

Le J, Zou J, Yang K, Wang M. 2014. Signaling to stomatal initiation and cell division. *Frontiers in Plant Science* 5: 297.

Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li Z, Melkonian M, Mirarab S et al. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574: 679–685.

Léran S, Varala K, Boyer J-C, Chiurazzi M, Crawford N, Daniel-Vedele F, David L, Dickstein R, Fernandez E, Forde B et al. 2014. A unified nomenclature of nitrate transporter 1/peptide transporter family members in plants. *Trends in Plant Science* 19: 5–9.

Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v.4: recent updates and new developments. *Nucleic Acids Research* 47: W256–W259.

Li F-W, Brouwer P, Carretero-Paulet L, Cheng S, de Vries J, Delaux P-M, Eily A, Koppers N, Kuo L-Y, Li Z et al. 2018. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants* 4: 460–472.

Li F-W, Nishiyama T, Waller M, Frangedakis E, Keller J, Li Z, Fernandez-Pozo N, Barker MS, Bennett T, Blázquez MA et al. 2020. *Anthoceros* genomes illuminate the origin of land plants and the unique biology of hornworts. *Nature Plants* 6: 259–272.

Li H-T, Yi T-S, Gao L-M, Ma P-F, Zhang T, Yang J-B, Gitzendanner MA, Fritsch PW, Cai J, Luo Y et al. 2019. Origin of angiosperms and the puzzle of the Jurassic gap. *Nature Plants* 5: 461–470.

Lu K-J, van 't Wout Hofland N, Mor E, Mutte S, Abrahams P, Kato H, Vandepoele K, Weijers D, De Rybel B. 2020. Evolution of vascular plants through redeployment of ancient developmental regulators. *Proceedings of the National Academy of Sciences, USA* 117: 733–740.

Lucas WJ, Groover A, Lichtenberger R, Furuta K, Yadav SR, Helariutta Y, He XQ, Fukuda H, Kang J, Brady SM et al. 2013. The plant vascular system: evolution, development and functions. *Journal of Integrative Plant Biology* 55: 294–388.

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

Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ. 2018. The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences, USA* 115: 2274–2283.

Mosca E, Cruz F, Gómez-Garrido J, Bianco L, Rellstab C, Brodbeck S, Csilléry K, Fady B, Fladung M, Fussi B et al. 2019. A reference genome sequence for the European silver fir (*Abies alba* Mill.): a community-generated genomic resource. *G3: Genes, Genomes, Genetics* 9: 2039–2049.

Moubayidin L, Perilli S, Dello Ioio R, Di Mambro R, Costantino P, Sabatini S. 2010. The rate of cell differentiation controls the Arabidopsis root meristem growth phase. *Current Biology* 20: 1138–1143.

Murphy E, De Smet I. 2014. Understanding the *RALF* family: a tale of many species. *Trends in Plant Science* 19: 664–671.

Mutte SK, Kato H, Rothfels C, Melkonian M, Wong GKS, Weijers D. 2018. Origin and evolution of the nuclear auxin response system. *eLife* 7: e33399.

Neale DB, McGuire PE, Wheeler NC, Stevens KA, Crepeau MW, Cardeno C, Zimin AV, Puiu D, Pertea GM, Sezen UU et al. 2017. The Douglas-Fir genome sequence reveals specialization of the photosynthetic apparatus in Pinaceae. *G3: Genes, Genomes, Genetics* 7: 3157–3167.

Nguyen KH, Van HC, Nishiyama R, Watanabe Y, Leyva-González MA, Fujita Y, Tran UT, Li W, Tanaka M, Seki M et al. 2016. Arabidopsis type B cytokinin response regulators *ARR1, ARR10,* and *ARR12* negatively regulate plant responses to drought. *Proceedings of the National Academy of Sciences, USA* 113: 3090–3095.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274.

Nishiyama T, Sakayama H, de Vries J, Buschmann H, Saint-Marcoux D, Ullrich KK, Haas FB, Vanderstraeten L, Becker D, Lang D et al. 2018. The *Chara* genome: secondary complexity and implications for plant terrestrialization. *Cell* 174: 448–464.

Oh E, Seo PJ, Kim J. 2018. Signaling peptides and receptors coordinating plant root development. *Trends in Plant Science* 23: 337–351.

Ohtani M, Akiyoshi N, Takenaka Y, Sano R, Demura T. 2017. Evolution of plant conducting cells: perspectives from key regulators of vascular cell differentiation. *Journal of Experimental Botany* 68: 17–26.

Paps J, Holland PWH. 2018. Reconstruction of the ancestral metazoan genome reveals an increase in genomic novelty. *Nature Communications* 9: 1730.

Pederson ERA, Warshan D, Rasmussen U. 2019. Genome sequencing of *Pleurozium schreberi*: the assembled and annotated draft genome of a pleurocarpous feather moss. *G3: Genes, Genomes, Genetics* 9: 2791–2797.

Plackett ARG, Emms DM, Kelly S, Hetherington AM, Langdale JA. 2021. Conditional stomatal closure in a fern shares molecular features with flowering plant active stomatal responses. *Current Biology* 31: 4560–4570.

Raven JA, Edwards D. 2001. Roots: evolutionary origins and biogeochemical significance. *Journal of Experimental Botany* 52: 381–401.

Ruonala R, Ko D, Helariutta Y. 2017. Genetic networks in plant vascular development. *Annual Review of Genetics* 51: 335–359.

Ruszala EM, Beerling DJ, Franks PJ, Chater C, Casson SA, Gray JE, Hetherington AM. 2011. Land plants acquired active stomatal control early in their evolutionary history. *Current Biology* 21: 1030–1035.

Santos Teixeira JA, ten Tusscher KH. 2019. The systems biology of lateral root formation: connecting the dots. *Molecular Plant* 12: 784–803.

Scott AD, Zimin AV, Puiu D, Workman R, Britton M, Zaman S, Caballero M, Read AC, Bogdanove AJ, Burns E et al. 2020. A reference genome sequence for giant sequoia. *G3: Genes, Genomes, Genetics* 10: 3907–3919.

Shkolnik D, Nuriel R, Bonza MC, Costa A, Fromm H. 2018. *MIZ1* regulates *ECA1* to generate a slow, long-distance phloem-transmitted $Ca^{2+}$ signal essential for root water tracking in Arabidopsis. *Proceedings of the National Academy of Sciences, USA* 115: 8031–8036.

Stein WE, Mannolini F, Hernick LV, Landing E, Berry CM. 2007. Giant cladoxylopsid trees resolve the enigma of the Earth's earliest forest stumps at Gilboa. *Nature* 446: 904–907.

Stevens KA, Wegrzyn JL, Zimin A, Puiu D, Crepeau M, Cardeno C, Paul R, Gonzalez-Ibeas D, Koriabine M, Holtz-Morris AE et al. 2016. Sequence of the sugar pine megagenome. *Genetics* 204: 1613–1626.

Stoeckle D, Thellmann M, Vermeer JE. 2018. Breakout – lateral root emergence in *Arabidopsis thaliana*. *Current Opinion in Plant Biology* 41: 67–72.

Sugano SS, Shimada T, Imai Y, Okawa K, Tamai A, Mori M, Hara-Nishimura I. 2010. Stomagen positively regulates stomatal density in *Arabidopsis*. *Nature* 463: 241–244.

Sussmilch FC, Schultz J, Hedrich R, Rob M, Roelfsema G. 2019. Acquiring control: the evolution of stomatal signalling pathways. *Trends in Plant Science* **24**: 342–351.

Tam THY, Catarino B, Dolan L. 2015. Conserved regulatory mechanism controls the development of cells with rooting functions in land plants. *Proceedings of the National Academy of Sciences, USA* **112**: E3959–E3968.

Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science* **278**: 631–637.

Tominaga R, Iwata M, Sano R, Inoue K, Okada K, Wada T. 2008. Arabidopsis *CAPRICE-LIKE MYB 3 (CPL3)* controls endoreduplication and flowering development in addition to trichome and root hair formation. *Development* **135**: 1335–1345.

UniProt Consortium T. 2018. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **46**: 2699.

Verstraeten I, Schotte S, Geelen D. 2014. Hypocotyl adventitious root organogenesis differs from lateral root development. *Frontiers in Plant Science* **5**: 495.

Vissenberg K, Claeijs N, Balcerowicz D, Schoenaers S. 2020. Hormonal regulation of root hair growth and responses to the environment in *Arabidopsis*. *Journal of Experimental Botany* **71**: 2412–2427.

Wang S, Li L, Li H, Sahu SK, Wang H, Xu Y, Xian W, Song BO, Liang H, Cheng S *et al.* 2019. Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nature Plants* **6**: 95–106.

Warren RL, Keeling CI, Saint YMM, Raymond A, Taylor GA, Vandervalk BP, Mohamadi H, Paulino D, Chiu R, Jackman SD *et al.* 2015. Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *The Plant Journal* **83**: 189–212.

Yokoyama A, Yamashino T, Amano Y-I, Tajima Y, Imamura A, Sakakibara H, Mizuno T. 2007. Type-B ARR transcription factors, *ARR10* and *ARR12*, are implicated in cytokinin-mediated regulation of protoxylem differentiation in roots of *Arabidopsis thaliana*. *Plant and Cell Physiology* **48**: 84–96.

Yu J, Li L, Wang S, Dong S, Chen Z, Patel N, Goffinet B, Chen H, Liu H, Liu Y. 2020a. Draft genome of the aquatic moss *Fontinalis antipyretica* (Fontinalaceae, Bryophyta). *Gigabyte* **2020**: 1–9.

Yu J, Zhang Y, Liu W, Wang H, Wen S, Zhang Y, Xu L. 2020b. Molecular evolution of auxin-mediated root initiation in plants. *Molecular Biology Evolution* **37**: 1387–1393.

Zhang C, Bousquet A, Harris JM. 2014. Abscisic acid and lateral root organ defective/numerous infections and polyphenolics modulate root elongation via reactive oxygen species in *Medicago truncatula*. *Plant Physiology* **166**: 644–658.

Zhang J, Fu X-X, Li R-Q, Zhao X, Liu Y, Li M-H, Zwaenepoel A, Ma H, Goffinet B, Guan Y-L *et al.* 2020. The hornwort genome and early land plant evolution. *Nature Plants* **6**: 107–118.

Zhu Q, Shao Y, Ge S, Zhang M, Zhang T, Hu X, Liu Y, Walker J, Zhang S, Xu J. 2019. A *MAPK* cascade downstream of *IDA–HAE/HSL2* ligand–receptor pair in lateral root emergence. *Nature Plants* **5**: 414–423.

Zhu S, Estévez JM, Liao H, Zhu Y, Yang T, Li C, Wang Y, Li L, Liu X, Pacheco JM *et al.* 2020. The *RALF1–FERONIA* complex phosphorylates *eIF4E1* to promote protein synthesis and polar root hair growth. *Molecular Plant* **13**: 698–716.

Zimin A, Stevens KA, Crepeau MW, Holtz-Morris A, Koriabine M, Marçais G, Puiu D, Roberts M, Wegrzyn JL, de Jong PJ *et al.* 2014. Sequencing and assembly of the 22-Gb Loblolly Pine Genome. *Genetics* **196**: 875–890.

## Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

**Dataset S1** Charophyte genome BLAST.

**Dataset S2** Gene occupancy.

**Dataset S3** FASTA alignments and phylogenetic trees.

**Dataset S4** List of novel, duplicated and coopted genes.

**Fig. S1** Heatmap displaying absence, partial presence and presence in all species for both transcriptomic and genomic data for genes involved in stomatal development.

**Fig. S2** Heatmap displaying absence, partial presence and presence in all species for both transcriptomic and genomic data for genes involved in stomatal signalling.

**Fig. S3** Heatmap displaying absence, partial presence and presence in all species for both transcriptomic and genomic data for genes involved in vascular tissue development.

**Fig. S4** Heatmap displaying absence, partial presence and presence in all species for both transcriptomic and genomic data for genes involved in root hair development.

**Fig. S5** Heatmap displaying absence, partial presence and presence in all species for both transcriptomic and genomic data for genes involved in primary root development.

**Fig. S6** Heatmap displaying absence, partial presence and presence in all species for both transcriptomic and genomic data for genes involved in lateral root development.

**Table S1** Additional genome data used in this study and sources of genome data.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.