Molecular Plant
Research Article

CellPress
Partner Journal

Major episodes of horizontal gene transfer drove the evolution of land plants

Jianchao Ma[1,7], Shuanghua Wang[2,7], Xiaojing Zhu[1], Guiling Sun[1], Guanxiao Chang[1], Linhong Li[3], Xiangyang Hu[4], Shouzhou Zhang[5], Yun Zhou[1], Chun-Peng Song[1,*] and Jinling Huang[1,2,6,*]

[1]State Key Laboratory of Crop Stress Adaptation and Improvement, School of Life Sciences, Henan University, Kaifeng 475004, China

[2]Key Laboratory for Plant Diversity and Biogeography of East Asia, Yunnan Key Laboratory for Fungal Diversity and Green Development, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

[3]Department of Entomology, China Agricultural University, Beijing 100193, China

[4]Shanghai Key Laboratory of Bio-Energy Crops, School of Life Sciences, Shanghai University, Shanghai 200444, China

[5]Fairy Lake Botanical Garden, Shenzhen & Chinese Academy of Sciences, Shenzhen 518004, China

[6]Department of Biology, East Carolina University, Greenville, NC 27858, USA

[7]These authors contributed equally to this article.

***Correspondence: Jinling Huang (huangj@ecu.edu), Chun-Peng Song (songcp@henu.edu.cn)**

https://doi.org/10.1016/j.molp.2022.02.001

ABSTRACT

How horizontal gene transfer (HGT) has contributed to the evolution of animals and plants remains a major puzzle. Despite recent progress, defining the overall scale and pattern of HGT events in land plants has been largely elusive. In this study, we performed systematic analyses for acquired genes in different plant groups and throughout land plant evolution. We found that relatively recent HGT events occurred in charophytes and all major land plant groups, but their frequency declined rapidly in seed plants. Two major episodes of HGT events occurred in land plant evolution, corresponding to the early evolution of streptophytes and the origin of land plants, respectively. Importantly, a vast majority of the genes acquired in the two episodes have been retained in descendant groups, affecting numerous activities and processes of land plants. We analyzed some of the acquired genes involved in stress responses, ion and metabolite transport, growth and development, and specialized metabolism, and further assessed the cumulative effects of HGT in land plants.

Key words: plant evolution, adaptation, streptophytes, cumulative effect, stress response, growth and development

Ma J., Wang S., Zhu X., Sun G., Chang G., Li L., Hu X., Zhang S., Zhou Y., Song C.-P., and Huang J. (2022). Major episodes of horizontal gene transfer drove the evolution of land plants. Mol. Plant. **15**, 857–871.

INTRODUCTION

The biosphere represents a knitted and regulated network of living organisms, where individuals interact and co-evolve in dynamic environments. Organismal interactions sometimes lead to genetic integration, occasionally in the form of intracellular organelles (e.g., mitochondria and plastids) (Dyall et al., 2004), but more frequently through acquisition of individual genes (Soucy et al., 2015; Husnik and McCutcheon, 2018). Horizontal gene transfer (HGT), the exchange of genetic material across species boundaries, is often considered a driving force in prokaryotic evolution, but has also gained increasing appreciation in eukaryotes (Huang, 2013; Soucy et al., 2015), particularly thanks to the rapid accumulation of genome data in recent years. By spreading evolutionary success across lineages, HGT may provide recipient organisms new functions or phenotypes to better adapt to their environments. Nonetheless, although HGT is rampant in prokaryotes and relatively common in unicellular eukaryotes (Dagan et al., 2008; Husnik and McCutcheon, 2018), the role of HGT in complex multicellular eukaryotes, including plants and animals, is far less clear and sometimes considered anecdotal.

Land plants evolved within the charophyte green algae (Delwiche and Cooper, 2015). The evolution of land plants witnessed an

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

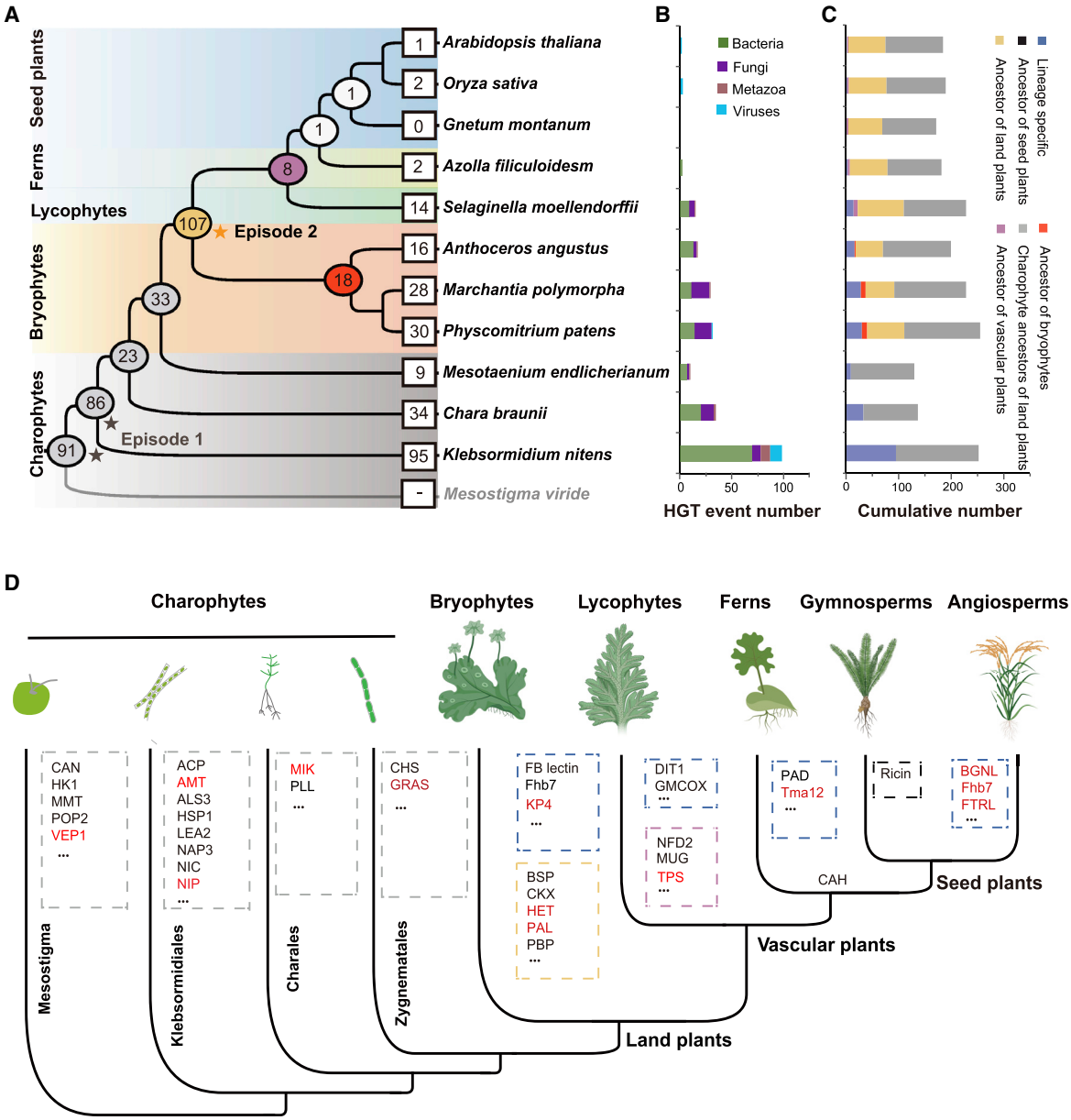

Figure 1. Summary of HGT event numbers in 12 representative species of charophytes and land plants.

(**A**) HGTs in 12 sampled species. Numbers in ovals represent gene families transferred to the ancestors of individual groups, and those in squares represent gene families transferred to sampled species. Gene families transferred specifically to *M. viride* (or Mesostigomatophyceae) are not included because of the low assembly quality. Ovals in gray and red represent gene families transferred to charophyte ancestors of land plants and the ancestral bryophyte, whereas those in yellow and purple show gene families transferred to the last common ancestor of land plants and vascular plants, respectively. The two major episodes of historical HGT events are marked by stars. The phylogenetic tree was constructed based on 215 single-copy genes from the sampled species.

(**B**) Lineage-specific HGT events. The x axis represents the number of acquired gene families.

(**C**) Cumulative HGT events in each sampled species. Colors represent the numbers of gene families acquired by different ancestral nodes over time.

(**D**) Summary of acquired genes in charophytes and major land plant groups. Images of individual organisms are adopted from BioRender (https://biorender.com/). The different colors of the frames represent HGT events that occurred during different evolutionary stages of plants, as shown in (**C**). Ellipses in boxes indicate additional acquired genes that are not listed due to space limit. Genes in red show HGT events that were also reported previously. Abbreviations are as follows: ACP, acid phosphatase; ALS3, aluminum-sensitive 3; AMT, ammonium transporter; BGNL, β-1,6-glucanase like; BSP, basic secretory protein; CAH, cyanamide hydratase; CAN, Ca^{2+}-dependent nuclease; CHS, chalcone synthase; CKX, cytokinin oxidase; DIT1, dityrosine biosynthesis protein 1; FB lectin, fruit body lectin; Fhb7, fusarium head blight 7; FTRL, fungal transcriptional regulatory-like; GDA, guanine deaminase; GMCOX, glucose methanol choline oxidoreductase; GRAS, GRAS family transcription factor; HET, HET domain-containing protein; HK1,

(legend continued on next page)

increasing level of structural and physiological complexity, from charophyte algae with only unicellular zygotes in the diploid stage and bryophytes with dominant gametophytes and parasitic, multicellular sporophytes eventually to seed plants with dominant sporophytes and reduced, often structurally internalized gametophytes. Over their course of evolution, land plants interacted with environmental microbes and other organisms, particularly rhizobacteria and mycorrhizal (or mycorrhiza-like) fungi (Martin et al., 2017). Because physical association often facilitates HGT, it is expected that such intricate interactions would occasionally lead to acquisition of foreign genes in land plants.

Theoretically, acquired genes in plants (loosely referring to land plants and their charophyte relatives, hereafter) could be derived from any foreign source, including other plant species, but they are not equally important in their functional roles. Prokaryotes, fungi, and other distantly related organisms often have very different gene contents and, as such, are often considered to be a major source of novel functions (Huang and Yue, 2013; Soucy et al., 2015). The acquired novel genes (or functions) may influence plant physiology and development, ultimately leading to better adaptation of plants to their environments. Indeed, it has been shown that HGT occurred at various stages of land plant evolution (Soucy et al., 2015; Chen et al., 2021). For instance, the acquired genes have played an important role in phytohormone biosynthesis and signaling (Yue et al., 2014; Cheng et al., 2019; Wang et al., 2020c), terpene synthesis (Jia et al., 2016), transcription regulation (Cheng et al., 2019), biotic and abiotic stress responses (Hoang et al., 2009; Li et al., 2018a; Guan et al., 2018; Wang et al., 2020a), and other activities of land plants (Yang et al., 2015; Bowman et al., 2017; Wang et al., 2020b). Nonetheless, a complete picture of HGT in the evolution of land plants, particularly seed plants, is still lacking.

In this study, we performed comprehensive analyses of HGT in major land plant groups and charophytes. In particular, we conducted a considerable amount of comparative genomic analyses and PCR verification to improve the quality of HGT detection. We discuss the overall scale and pattern of HGT events, their cumulative effects, and the activities affected by HGT in land plants. Remarkably, we found that two major episodes of historical HGT events have significantly shaped the long-term evolution of land plants.

RESULTS

Contrasting scales of relatively recent HGT events between seedless and seed plants

To identify acquired genes over the course of land plant evolution, we first selected 12 representative species from charophytes and major land plant groups, including four charophytes (*Mesostigma viride, Klebsormidium nitens*, *Chara braunii*, and *Mesotaenium endlicherianum*) and eight land plants (bryophytes *Physcomitrium patens*, *Marchantia polymorpha*, and *Anthoceros angustus*, the lycophyte *Selaginella moellendorffii*, the fern *Azolla filiculoides*, the gymnosperm *Gnetum montanum*, and angiosperms *Oryza sativa* and *Arabidopsis thaliana*) (Supplemental Figure 1). Annotated protein sequences from each of the 12 representative species were subjected to detailed and rigorous analyses for genes of prokaryotic, viral, fungal, or animal origin (Supplemental Figure 2). To further understand the overall scope of HGT in seed plants, we also selected 19 additional seed plants, including 2 gymnosperms, 13 grasses, and 4 species from the family Solanaceae, to screen for genes of foreign origin. In total, 31 charophyte and land plant genomes were analyzed to understand the role of HGT in land plant evolution. These analyses identified a total of 593 gene families that were transferred to streptophytes (charophytes and land plants) (Figure 1A and Supplemental Table 1). Genes acquired by their earlier ancestors (e.g., the last common ancestors of Plantae or green plants) are not included in the current study.

Our data indicate that HGT is active in charophytes and all major groups of land plants, but there is a contrasting scale between seedless and seed plants (Figure 1B). Among all sampled genomes, many more recently acquired genes were identified in charophytes and bryophytes. In particular, 95 gene families of foreign ancestry were identified in the filamentous charophyte *K. nitens*, a number far greater than reported thus far in any green plant to our knowledge, and 30 gene families in *P. patens* and other mosses (Figure 1A). All identified genes in *K. nitens* are located on long contigs, usually flanked by sequences of vertical descent, and a vast majority of them are identical to annotations reported in the published draft genome sequence (Hori et al., 2014). Likewise, all identified genes in *P. patens* are located on the assembled chromosomes according to Phytozome (Goodstein et al., 2012). Furthermore, most of the recently acquired genes identified in the two genomes also share the highest percentage identity with their homologs from *Interfilum paradoxum* (a close relative to *K. nitens*) and *Physcomitrium sphaericum* (a congeneric species of *P. patens*), respectively (Supplemental Table 2).

Compared with charophytes and bryophytes, a lower number of recently acquired genes was found in the sampled lycophytes and ferns (Figure 1A). We identified 14 acquired gene families in the lycophyte *S. moellendorffii*, most of which were also found in other sequenced genomes or transcriptomes of *Selaginella* (Figure 1A and Supplemental Table 2). Notably, the dityrosine biosynthesis gene *DIT1*, which is involved in fungal spore wall maturation, was transferred from fungi to lycophytes (Figure 2A). This *DIT1* gene is highly duplicated in *Selaginella* and upregulated during root development (Figure 2B). In the fern *A. filiculoides*, two gene families of foreign origin were identified.

Genes acquired relatively recently were also identified in seed plants, although the number is significantly lower (Supplemental Table 3). For example, the makes caterpillars

histidine kinase 1; HSP1, heat-stable protein 1; KP4, fungal killer protein 4; LEA2, late embryogenesis abundant protein, group 2; MIK, MYO-inositol kinase; MMT, methionine S-methyltransferase; MUG, meiotically upregulated gene; NAP3, non-intrinsic ABC protein 3; NFD2, nuclear fusion-defective 2; NIC, nicotinamidase; NIP, nodulin 26-like intrinsic protein; PAD, phenolic acid decarboxylase; PAL, phenylalanine ammonia lyase; PBP, penicillin binding protein; PLL, pectin lyase-like; POP2, pollen-pistil interaction 2; TPS, terpene synthase family; VEP1, vein patterning protein 1.

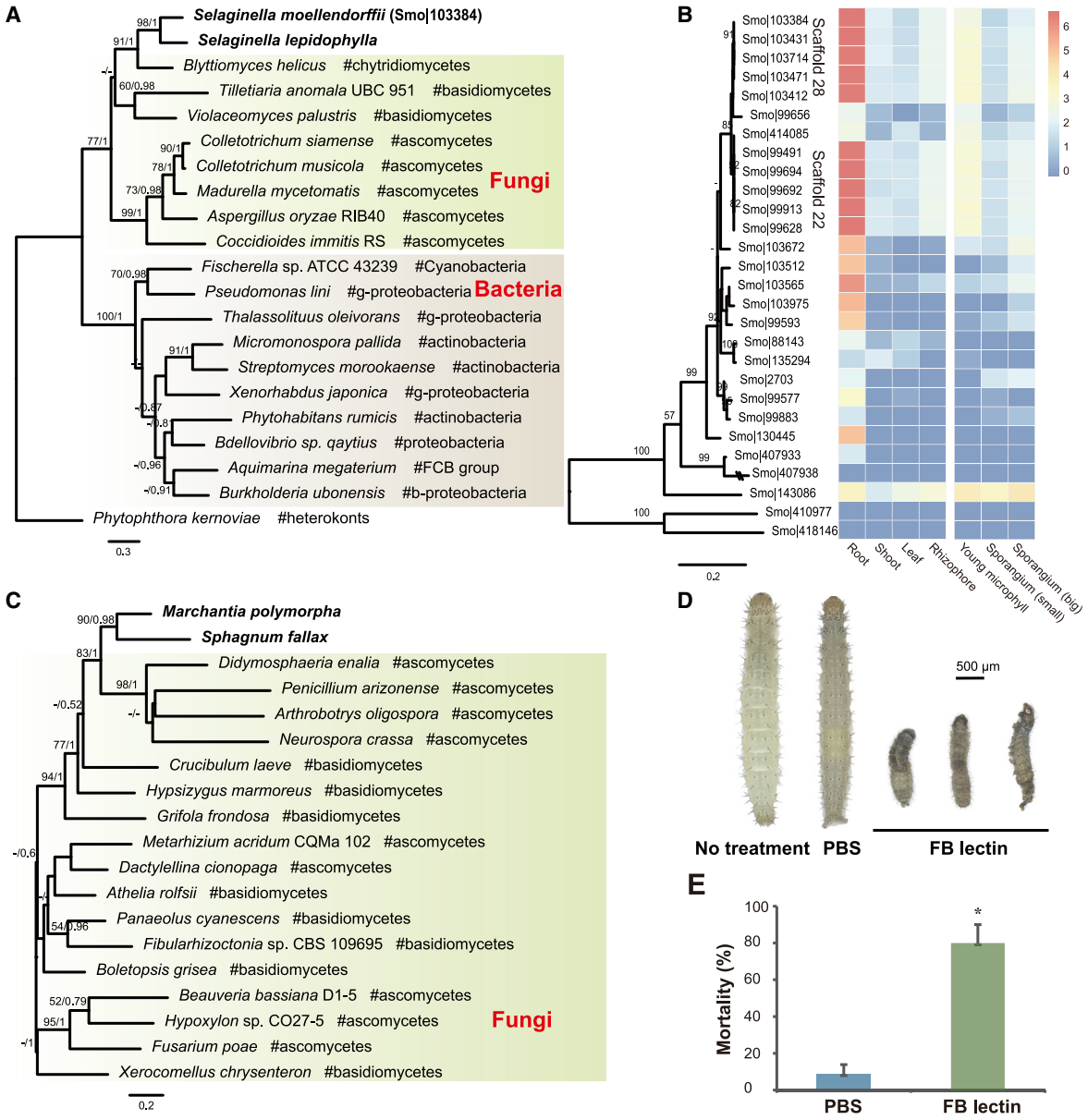

Figure 2. Analyses of dityrosine biosynthesis protein 1 (DIT1) from lycophytes and fungal FB lectin from the liverwort *M. polymorpha*.
(**A**) Phylogenetic analyses of DIT1 protein sequences.
(**B**) Molecular phylogeny and expression profile of *DIT1* in *Selaginella*. Log2 fold-change FPKM values were used for constructing the heatmap.
(**C**) Molecular phylogeny of FB lectin. Numbers beside branches on each tree represent bootstrap values from maximum likelihood and posterior probabilities from Bayesian analyses, respectively. Dashes indicate values lower than 50%.
(**D**) Larvae of *P. xylostella* after treatment with FB lectin. The larvae injected with PBS were used as controls.
(**E**) Mortality of *P. xylostella* after being treated with FB lectin and PBS. Statistically significant difference is indicated (Student's *t*-test, *p < 0.05).

floppy (*mcf*) gene was transferred from either bacteria or fungi to *Cycas*, an arborescent and evergreen gymnosperm genus (unpublished data). In grasses, other than multiple genes of viral origin, four genes acquired from fungal endophytes were also identified, confirming the findings of earlier studies (Shinozuka et al., 2017, 2020; Wang et al., 2020a). Some additional genes of possibly bacterial origin were identified, but most of them could not be confirmed because of the lack of homologous genes in related species or even in related strains of the same species. For instance, the secondary half of a receptor-like kinase gene in maize (MaizeGDB: GRMZM5G870291; https://www.maizegdb.org/gene_center/gene/GRMZM5G870291) was found to be of bacterial origin, but no homologous sequences could be identified in other maize strains. We listed these potential cases of HGT in Supplemental Table 3.

Major episodes of historical HGT events

Considering the evidence of relatively recent HGT events occurring in different land plant groups and charophytes, we performed additional analyses on historical HGT events during the earlier

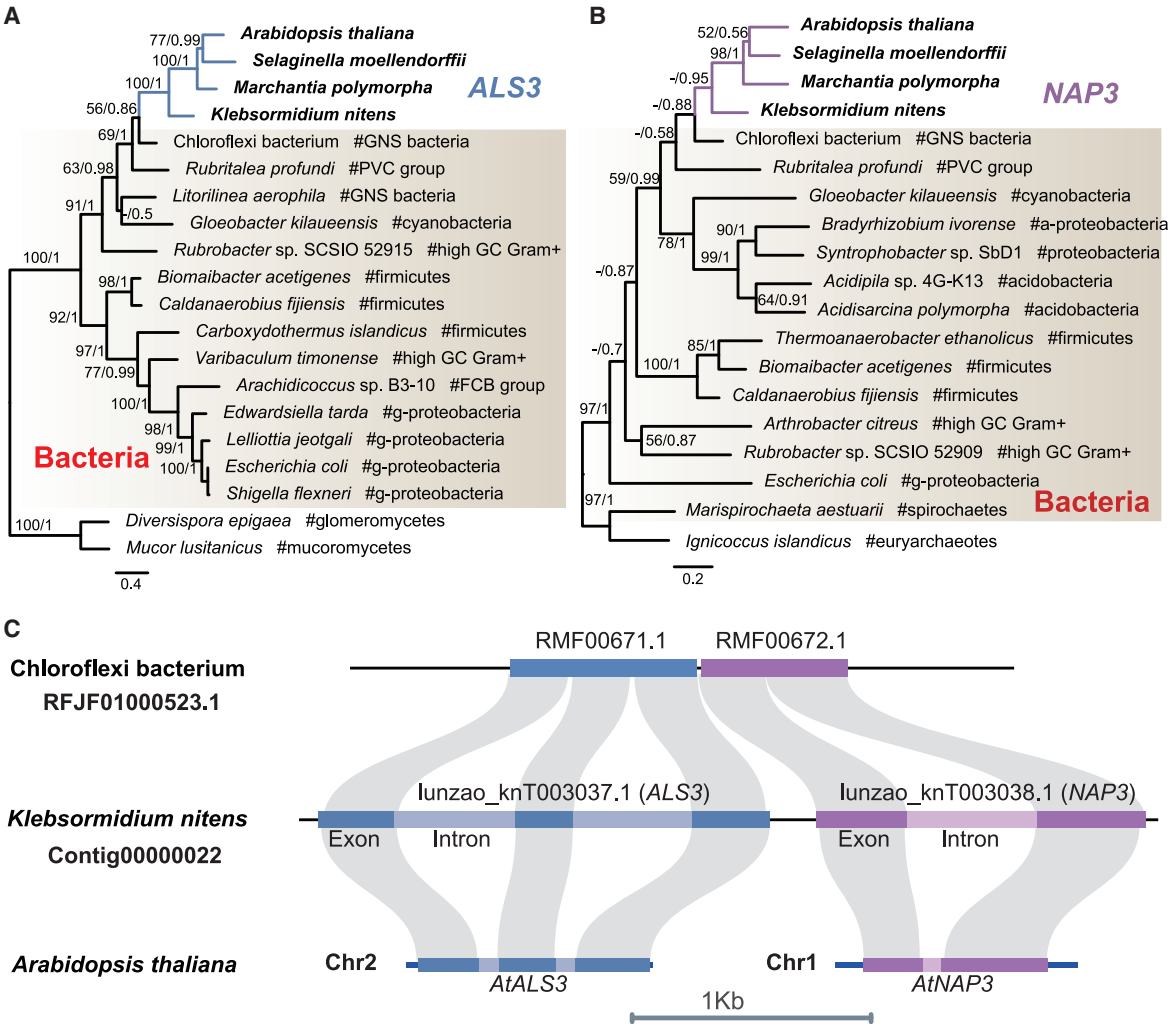

Figure 3. Horizontal transfer of the *ALS3*/*NAP3* ABC transporter complex genes from bacteria during the early evolution of streptophytes.

(**A** and **B**) Molecular phylogenies of ALS3 (**A**) and NAP3 (**B**) protein sequences. Numbers beside branches represent bootstrap values from maximum likelihood and posterior probabilities from Bayesian analyses, respectively. Dashes indicate values lower than 50%.

(**C**) Collinearity of *ALS3* and *NAP3* among a Chloroflexi bacterium, *K. nitens*, and *A. thaliana*. Blue and purple represent the structures of *ALS3* and *NAP3* genes, respectively. Scale bar, 1000 bp.

stages of land plant evolution. Our analyses indicate that historical HGT events not only occurred, but also can be confidently identified in many cases. For instance, an ABC transporter complex is encoded by two genes, Aluminum Sensitive 3 (*ALS3*) and Non-intrinsic ABC Protein 3 (*NAP3*) (Belal et al., 2015). Phylogenetic analyses show that, for both *ALS3* and *NAP3*, streptophyte sequences form a monophyletic group that in turn is derived from bacteria (Figure 3A and 3B). In many bacteria (e.g., Chloroflexi and firmicutes) and *K. nitens*, *ALS3* and *NAP3* are positioned in tandem on the chromosome (Figure 3C), suggesting that the two genes were most likely co-transferred from bacteria during the early evolution of streptophytes. Another example is a guanine deaminase (*GDA*) gene in gymnosperms, which is identifiable in *Ginkgo biloba* and conifers. The gymnosperm guanine deaminase protein sequences are considerably more similar to homologs from bacteria than to those from other eukaryotes (77% versus less than 60% identity). Phylogenetic analyses strongly suggest a bacterial origin of gymnosperm guanine deaminase (Supplemental Figure 3).

Despite their continuous occurrence, historical HGT events did not occur uniformly in the evolution of land plants. We found that two major episodes of historical HGT events, consisting of 177 and 107 gene families, respectively, are responsible for a majority of genes acquired during the early stages of land plant evolution (Figure 1A). These two episodes correspond to the early evolution of streptophytes (episode 1) and the origin of land plants (episode 2), again pointing to the role of HGT in facilitating some major transitions of plant evolution (Huang and Gogarten, 2008; Price et al., 2012). These findings are largely consistent with the two bursts of genome novelty in the origin of land plants and confirm the speculation on the potential involvement of HGT (Bowles et al., 2020), but suggest a much lower scale of historical HGT.

Sources of acquired genes

Overall, genes of bacterial origin account for a majority of acquired genes identified in all sampled genomes. However, a large fraction of acquired genes are also derived from fungi. Especially in bryophytes and lycophytes, about 48% and 36%, respectively, of the identified genes were acquired from fungi (Figure 1B). Several acquired genes, such as those encoding killer protein 4 (KP4), heterokaryon incompatibility (HET) domain-containing protein (Sun et al., 2020), fungal fruit body (FB) lectin, and spore wall maturation protein DIT1 (Figure 2), are otherwise related to fungus-specific activities. In particular, some of these fungus-derived genes (e.g., *DIT1*, FB lectin, and HET domain-containing protein) are highly duplicated in plants (Figure 2B and Supplemental Figure 4A). It is also noteworthy that at least 19 gene families are exclusively detected in fungi, bacteria (usually multiple phyla), and streptophytes (Supplemental Table 4), suggesting that bacteria, given their ubiquitous presence and early origin, might be the ultimate source, and serial HGT events might have been involved. Such a scenario of serial HGT events has also been observed for other genes or organisms (Emiliani et al., 2009; Sun et al., 2010). Sequences of viral affiliation were detected in certain sampled genomes, including at least 11 genes of viral origin in *K. nitens*, most of which are associated with plankton viruses. Several other genes in seed plants are related to viruses, such as mitoviruses or rice tungro bacilliform viruses (Supplemental Tables 2 and 3).

HGT from animals to streptophytes has been previously reported (Hoang et al., 2009). We therefore performed analyses for sequences that were likely acquired from animals. In *K. nitens*, homologs for at least nine genes are restricted to animals and sometimes bacteria or other charophytes (Supplemental Table 2). Intriguingly, although these homologs are usually widely distributed in animals, those from aquatic animals often are among the top hits to *K. nitens* sequences (e.g., poly(ADP-ribose) polymerase 14-like protein, NCBI accession no. GAQ84501). A similar distribution pattern was also observed for sequences of bryophytes and other land plants (e.g., NCBI accession no. PNR44297) (Supplemental Figure 5), including a bryophyte actinoporin gene reported earlier (Hoang et al., 2009). Whether such sequence similarity between streptophytes and aquatic animals reflects gene exchanges between organisms inhabiting similar habitats remains to be further investigated.

The cumulative effects of HGT events in plant evolution

Given the dynamic occurrence of HGT over time, we further assessed the cumulative effect of HGT in plant evolution. Such a cumulative effect of HGT is particularly pronounced in seed plants. Although recently acquired genes specific to seed plants are relatively rare, genes derived from earlier HGT events, especially the two major episodes, are common. Specifically, about 58% (103/177) of the genes acquired during episode 1 and 71% (76/107) of those during episode 2 have been retained in seed plants (Figure 1C). Furthermore, genes acquired prior to the origin of streptophytes might have also been accumulated in the genomes of seed plants (Huang and Gogarten, 2008; Qiu et al., 2013), even though they are not included in the current study.

Many genes acquired early in land plant evolution are known to participate in a wide range of activities and processes in seed plants (Figure 1D). For instance, the patatin-like phospholipase gene, likely acquired from bacteria during episode 1 (Supplemental Figure 6), has been retained in all land plant lineages. In plants, patatins are involved in multiple processes, including fat metabolism, stress response, and auxin and elicitor signal transduction (Scherer et al., 2010; Yang et al., 2012). The bark storage protein gene, also acquired during episode 1 (Supplemental Figure 7), is involved in short-term nitrogen storage in barks and parenchyma cells between seasons in poplars and other temperate trees, thus playing a key role in nitrogen cycling (Wetzel et al., 1989). Another example is the plant flotillin gene, likely acquired from fungi during episode 1 (Supplemental Figure 8). In angiosperms, flotilins participate in endocytosis, nitrogen-fixing bacterial infection, and seedling development (Haney and Long, 2010; Li et al., 2012).

Some of the genes derived from historical HGT events became exapted for newer functions, whereas others were highly duplicated in land plants after the transfer event, leading to significant functional differentiation. For example, the subtilase gene family, which was transferred from bacteria to the last common ancestor of Coleochaetophyceae, Zygnematophyceae, and land plants (Yue et al., 2012; Xu et al., 2019) (Supplemental Table 2), was subjected to multiple rounds of duplication, ultimately resulting in 56 gene copies in *A. thaliana* (Schaller et al., 2018; Xu et al., 2019) and 80 in grape (Cao et al., 2014). In angiosperms, this subtilase gene family is involved in numerous activities, including biotic interactions, regulation of cell death, and control of growth and development (Schaller et al., 2018). To a certain extent, these acquired genes provide the genetic basis for the remarkable diversity in form and ecology that land plants ultimately evolved (see following sections).

Activities and processes in plants affected by HGT

A majority of the acquired genes identified in our analyses have not been functionally investigated in land plants. Based on available data, we here highlight some of the notable genes involved in different plant physiological and developmental processes.

Stress responses

It has long been known that HGT facilitates the adaptation of recipient organisms to a shifting environment (Gogarten et al., 2002; Husnik and McCutcheon, 2018; Soucy et al., 2015). Because of the major habitat transition from water to land, adaptive traits (e.g., water conductance, stomata-like structures, and embryos) evolved in plants to overcome terrestrial stresses (e.g., UV irradiance, drought, and drastic temperature fluctuation) (Furst-Jansen et al., 2020). It is not surprising that new genes, especially those acquired from environmental organisms, play roles in stress responses of land plants. In mosses, an actinoporin gene of animal origin and a hemerythrin gene of fungal origin have been experimentally demonstrated to confer dehydration tolerance to plants (Guan et al., 2018; Hoang et al., 2009). Our current analyses identified additional genes important in dehydration resistance and abiotic stress responses. For instance, the LEA proteins are known for their role in desiccation resistance of land plants (Candat et al., 2014). The *LEA2* gene family was likely acquired from bacteria

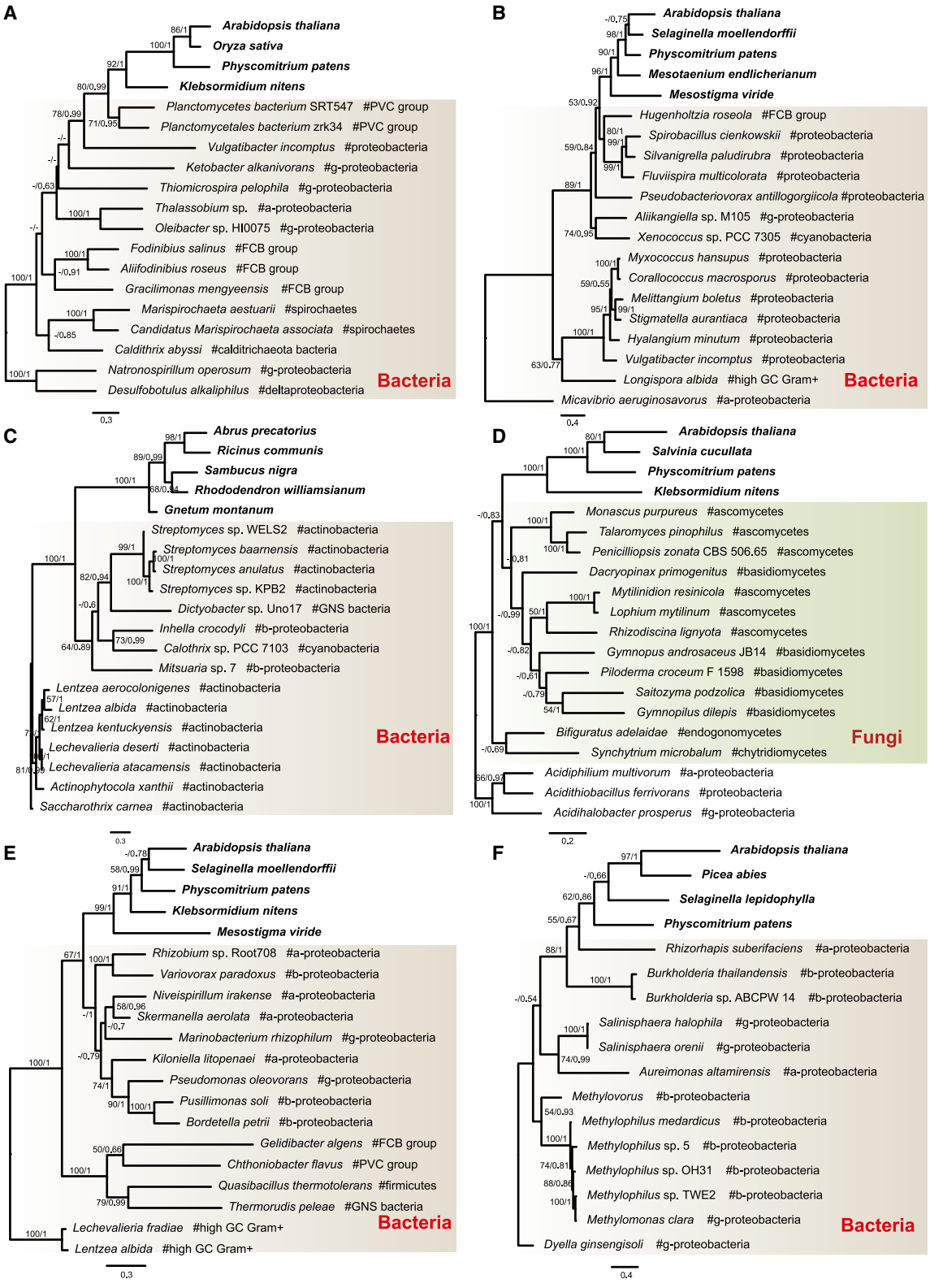

Figure 4. Additional examples of horizontally acquired genes in plants.
(**A–F**) Molecular phylogenies of LEA2 (**A**), methionine S-methyltransferase-like protein (**B**), ricin/abrin (**C**), ammonia transporter (**D**), pollen–pistil interaction 2 (**E**), and lachrymatory factor synthase (**F**). Numbers beside branches represent bootstrap values from maximum likelihood and posterior probabilities from Bayesian analyses, respectively. Dashes indicate values lower than 50%.

during episode 1 and is now widely distributed in land plants (Figure 4A). Similarly, the *ALS3/NAP3* ABC transporter complex (Figure 3) has also been retained in all major land plant groups. In angiosperms, this ABC transporter complex is required for aluminum resistance and phosphate-deficiency response (Larsen et al., 2005; Belal et al., 2015). In addition, the methionine S-methyltransferase-like gene (*MMT*) and an osmotically inducible protein C gene (*OsmC*), which are related to plant salinity tolerance and bacterial oxidative stress resistance, respectively (Atichartpongkul et al., 2001; Ogawa and Mitsuya, 2012), were also acquired during episode 1 (Figure 4B and Supplemental Figure 9).

Some acquired genes identified in our analyses are involved in responses to biotic stresses in land plants. Other than the patatin gene family (Supplemental Figure 6), which is involved in oxylipin biosynthesis and pathogen resistance (La Camera et al., 2009), the plant basic secretory protein (*BSP*) gene family, acquired from bacteria during episode 2 (Supplemental Figure 10), is often thought to be related to pathogen resistance (Marquez et al., 2018). The HET domain presumably is a major component of the self-recognition system in filamentous fungi (Hall et al., 2010). It has been shown that these fungus-related HET domain-containing genes are upregulated in mosses in response to fungal infection (Reboledo et al., 2021; Sun et al., 2020). In liverworts and mosses, the acquired fungal FB lectin (Figure 2C) contains an FB domain (pfam 07367) that is likely involved in insecticidal activities (Wang et al., 2002). Our experimental data show that injection of liverwort FB lectin led to an average of 80% mortality in the larvae of diamondback moths (*Plutella xylostella*), significantly higher than in the control group (8.8% mortality) (Figure 2D and 2E).

Although genes acquired by seed plants are relatively rare, they are disproportionally involved in biotic stress responses. A glutathione S-transferase gene, acquired by wheatgrasses (*Thinopyrum* spp.) from fungal endophytes, has recently been shown to confer resistance to *Fusarium* head blight (Wang et al., 2020a). Ricin and abrin are ribosome-inactivating enzymes involved in plant defense against herbivores and microbial pathogens (Peumans and Van Damme, 1995; Lannoo and Van Damme, 2014). Intriguingly, our data show that ricin and abrin toxins are encoded by a gene family acquired from bacteria during the evolution of seed plants (Figure 4C).

Metabolite and ion transport

As often found in other groups (Whitaker et al., 2009; Yue et al., 2013), many genes identified in our analyses encode enzymes catalyzing various metabolic processes, particularly carbohydrate and amino acid metabolism (Supplemental Table 5).

We identified at least 10 transporter gene families of foreign ancestry in our analyses. In addition to the *ALS3/NAP3* ABC transporter complex (Figure 3), other acquired gene families include zinc transporter, magnesium transporter, inorganic phosphate transporter, cation-transporting P-type ATPase, ammonium transporter, oligopeptide transporter, nuclear transport factor 2 (*NTF2*), major facilitator proteins, and nodulin 26-like intrinsic protein (NIP) (Supplemental Table 5). Among these transporters, the NIP family, widely distributed in land plants, transport glycerol, water, boric acid, urea, and other small molecules across cell membranes (Wallace et al., 2006; Wang et al., 2016). Our phylogenetic analyses show that the NIP family in land plants is derived from an earlier HGT event from bacteria during episode 1 (Supplemental Figure 11), consistent with findings from previous studies (Abascal et al., 2014; Zardoya et al., 2002). In addition, the ammonium transporter gene family, which is important in plant uptake of nitrogen from soil, was found to be acquired from fungi during episode 1 (Figure 4D), again confirming a previous report (McDonald et al., 2012). We note here that, except for a zinc transporter that appears to be recently transferred to *K. nitens* and close relatives, all other identified transporter gene families were likely acquired, either from fungi or from bacteria, during the two major episodes and are now retained in descendant plant groups (Supplemental Table 2). These acquired transporter genes influence related pathways and processes by facilitating the exchange of metabolic products between cellular compartments, setting the stage for the evolution of greater developmental complexity of land plants.

Growth and development

The growth and development of land plants is largely controlled by phytohormone signaling and transcriptional regulation. It has already been reported that HGT played a role in the evolution of auxin biosynthesis (Yue et al., 2014), cytokinin degradation (Wang et al., 2020c), and ABA signal perception (Cheng et al., 2019). Our current analyses also show that the karrikin and strigolactone receptor family *KAI2/D14* (Wang et al., 2021) and plant histidine kinase genes, commonly involved in the signaling of cytokinin and ethylene, were also likely acquired from bacteria (Supplemental Figure 12).

Some acquired genes identified in our analyses directly participate in the structural development of land plants. The bacterial ribonuclease III family was likely acquired by the last common ancestor of vascular plants (Supplemental Figure 13). In *A. thaliana*, this family is essential for the fusion of polar nuclei, a key process of megagametophyte development in angiosperms (Portereiko et al., 2006). The pollen–pistil interaction 2 gene (*POP2*), involved in fertilization regulation in angiosperms (Renault et al., 2011), was acquired from bacteria during episode 1 (Figure 4E). EVE domain-containing proteins (or thymocyte nuclear-like proteins), widely distributed in streptophytes and bacteria, are thought to be functionally related to programmed cell death (Bell et al., 2020). Consistent with an earlier finding (Bell et al., 2020), our data show that plant EVE domain-containing protein genes are derived from bacteria (Supplemental Figure 14). At least three acquired genes have been shown experimentally to modulate vascular development (Jun et al., 2002; Yang et al., 2015; Ribeiro et al., 2020) (Supplemental Table 5). In addition, a gene of fungal origin also regulates protonemal and gametophore development, cell reprogramming, and other related processes in mosses (Wang et al., 2020b).

Land plant cell walls not only protect plants from environmental stresses, but also play a pivotal role in plant cell differentiation and upright growth (Sarkar et al., 2009). In particular, land plant cell walls contain unique components such as pectin, rhamnogalacturonans, and xyloglucan. Our analyses

identified acquired genes that function in cell wall formation and metabolism, including galacturonic acid phosphate kinase, xyloglucan endotransglycosylase, alpha-L-fucosidase, rhamnogalacturonate lyase, and pectin esterase (Supplemental Table 5). The pectin esterase gene family was likely acquired from bacteria and vastly amplified in land plants (Supplemental Figure 15). Other than their roles in cell wall metabolism, pectin esterases are also involved in other developmental and physiological processes of plants, such as stem elongation, fruit ripening, and heat stress response (Kameshwar and Qin, 2018; Wu et al., 2018).

Specialized metabolism

Plants produce a rich diversity of specialized metabolites, many of which are known for their roles in defense against pathogens and herbivores or as protective pigmentation against UV irradiation (Hartmann, 2007). Several earlier studies already reported horizontally acquired genes in the biosynthesis of terpenoids, vitamin K, and phenylpropanoids (Emiliani et al., 2009; Widhalm et al., 2012; Jia et al., 2016). Our analyses confirmed the above earlier findings and suggest a far greater impact of HGT on plant specialized metabolism. Some of our identified genes related to specialized metabolism include lachrymatory factor synthase, which is responsible for the formation of the irritating sulfine released from the chopped onion (Imai et al., 2002) (Figure 4F), fungal KP4 in mosses, and pore-forming hemolysin in lycophytes (Supplemental Figure 16). In fact, a large fraction (about 27.8%) of acquired genes identified in our analyses are likely involved in specialized metabolism, mirroring their restricted or patchy distributions, frequently at the genus or family level.

The carbohydrate-binding protein lectins are ubiquitous in plants and contain some of the deadliest toxins, such as ricin (Chrispeels and Raikhel, 1991). Other than an FB lectin gene acquired by bryophytes and the ricin/abrin gene family acquired by seed plants (Figures 2C and 4C), a concanavalin A lectin gene was also acquired from bacteria by the last common ancestor of land plants (Supplemental Figure 17).

Phenylpropanoid metabolism is responsible for the production of a large number of phenolic compounds such as coumarins, lignans, suberin, and flavonoids (Dixon and Paiva, 1995). The gene encoding phenylalanine ammonia lyase (PAL), which catalyzes the first step of the phenylpropanoid pathway, was reportedly acquired from fungi by the ancestral land plant (Emiliani et al., 2009), but a recent study paints a more complex picture of PAL evolution (de Vries et al., 2021). Our analyses show that at least two additional key gene families in phenylpropanoid metabolism, chalcone synthase (CHS) and stilbene synthase (STS), are also derived from bacteria (Supplemental Figure 18). The two families share a common origin and, after acquisition from bacteria, were further duplicated and functionally differentiated. CHS catalyzes the first step in the biosynthesis of flavonoids, the largest group of phenylpropanoids, whereas STS catalyzes the formation of stilbenes, compounds commonly known for their potent antimicrobial activities (Chong et al., 2009).

DISCUSSION

In this study, we performed comprehensive and rigorous analyses to understand the role of HGT in land plant evolution. Several issues, particularly insufficient sequence data from some groups and uncertain phylogenetic framework, could complicate data interpretation (Huang and Yue, 2013). Our identification of acquired genes is based on available sequence data and current understanding of land plant phylogeny (Puttick et al., 2018; One Thousand Plant Transcriptomes Initiative, 2019; Li et al., 2020). It is possible that the interpretation of identified genes will differ when additional sequencing data become available or land plant phylogeny is revised. Therefore, our results provide only an overall pattern and functional significance of HGT in land plant evolution, instead of a detailed list of acquired genes in individual taxa.

Contamination from environmental microbes during genome sequencing is always a significant concern in HGT detection. Especially because of the lack of genomic sequence data from closely related species, it is extremely challenging to distinguish species-specific HGT events from contamination. In our analyses, we relied on multiple lines of evidence to exclude contamination, such as gene location on the scaffold, distribution in closely related taxa, neighboring genes of vertical descent, and PCR or RT-PCR amplification. In particular, we performed careful comparative genomic analyses and identification of recently acquired genes in the charophyte *K. nitens*, the moss *P. patens*, the lycophyte *S. moellendorffii*, and grasses, based on our internally generated genome sequences or publicly available data from closely related species. These comparative genome analyses, to a large extent, alleviated the complication arising from sequencing contamination, allowing us to confidently identify recently acquired genes in these taxa. However, it is very likely that certain recently acquired genes in other sampled species might have escaped our detection.

Our comprehensive analyses provide much-needed information to understand the importance of HGT in the evolution of land plants and of eukaryotes in general. Our data, together with those reported earlier (Shinozuka et al., 2017, 2020; Wang et al., 2020a), clearly show that relatively recent HGT has occurred in charophytes and all major groups of land plants. In particular, about 1.5% of genes in the genome of the charophyte *K. nitens* were likely acquired from foreign sources. Even in seed plants, relatively recent HGT events from environmental microbes (e.g., fungi and bacteria) to families or lower-rank taxonomic groups can be confidently identified. In addition, the foreign nature of some genes acquired early in land plant evolution, such as *ALS3/NAP3* ABC transporter complex (Figure 3), has also been vividly documented in their conserved molecular features between the donor and the recipient group. Although the scale of HGT differs among groups and at different evolutionary stages, there appears to be no insurmountable barriers to HGT in any major group of land plants. Notably, a vast majority of genes acquired during the two major episodes of historical HGT are retained in numerous extant land plants (Figure 1C). Given the common belief that horizontally acquired genes are usually transient and secondarily lost (Keeling and Palmer, 2008; Ku et al., 2015), our data point to a far greater scope of historical HGT than we currently realize.

The involvement of acquired genes in numerous physiological and developmental processes demonstrates a significant role

for HGT in land plant evolution. On the other hand, it is also possible that other factors, such as historical contingency, might have led to the retention of certain acquired genes in plants. It should be noted again that most foreign genes in metabolite and ion transport, as well as in growth and development, identified in our analyses were acquired during the two episodes. These genes have often been accumulated, duplicated, or functionally differentiated in descendant groups, thus contributing to the diversification and long-term evolution of land plants. Historically, gene or genome duplication, mutation, and new gene origination are thought to be the major genetic mechanisms in plant evolution. Our data demand that HGT also be considered as an important force, along with other mechanisms of plant evolution.

The relatively large number of acquired genes in charophytes and bryophytes is consistent with the weak-link model (Huang, 2013), which postulates that organisms with exposed reproductive structures or early developmental stages tend to acquire more genes from organisms sharing their environments. Alternatively, apical cell damage and regeneration in bryophytes could also facilitate the introduction of foreign genes. The dramatically smaller number of acquired genes specific to seed plants, however, is somewhat perplexing. Presumably, this might be attributable to multiple issues, such as structurally internalized gametes and lack of sufficient time for gene recruitment, especially considering the younger age of many seed plants compared with charophytes and bryophytes. However, these scenarios might not sufficiently explain the paucity of recently acquired genes overall in seed plants, as multiple genes of fungal, viral, or bacterial origin can be confidently identified in the grass family. Coincidently, some of the genes acquired early in land plant evolution, notably those related to stress responses, were secondarily lost in seed plants (Supplemental Table 6) (Hoang et al., 2009; Guan et al., 2018; Sun et al., 2020). Therefore, we speculate that the rapid decline in the number of acquired genes in seed plants might have resulted from relatively stabilized physical environments or a developmental or physiological shift, such as improved adaptative mechanisms for terrestrial habitats, instead of barriers to HGT *per se*. We notice again that the acquired genes specific to seed plants, although relatively rare, are disproportionally related to biotic stress responses. This observation likely reflects the increasing diversity of pathogens or herbivores for plants, again indicating that, even in seed plants, acquisition of genes from foreign sources can be a mechanism of rapid adaptation under environmental stresses.

METHODS

Data sources

Annotated protein sequences and related data for sampled genomes were downloaded from Phytozome (Goodstein et al., 2012) and their respective sequencing centers (Supplemental Table 7). An internal customized database was created, including NCBI non-redundant protein sequences (downloaded on February 23, 2020, and updated on November 25, 2020) and annotated protein sequences from 14 recently published charophyte and land plant genomes (Supplemental Table 7). OneKP was also used in BLAST searches to determine the taxonomic distributions of acquired genes.

Phylogenetic tree construction using single-copy genes

Gene family clustering was performed for all protein sequences from the 12 representative species (Figure 1A and Supplemental Figure 1) using OrthoFinder (v.2.4.0), with default settings (Emms and Kelly, 2019). A total of 215 single-copy genes were obtained. Protein sequences for each single-copy gene were aligned using MAFFT (v.7.467) with default settings. These protein sequence alignments were concatenated using Geneious 10.0.2 (Biomatters, New Zealand). Poorly aligned regions and gaps were removed by trimAl (v.1.4) with parameters "-automated1" (Capella-Gutierrez et al., 2009). A phylogenetic tree was constructed from concatenated sequences using IQ-TREE, with an automatically selected best-fit amino acid substitution model (Figure 1A). Ultrafast bootstrap analyses were performed using 1000 replicates (-bb 1000 -m TEST).

Identification of horizontally acquired genes

Twelve species representing major groups of charophytes and land plants were first selected for HGT identification. Annotated protein sequences for these 12 species were subjected to genome screening for candidates of horizontally acquired genes using our HGT identification pipeline. Each of the candidate genes was then subjected to careful taxonomic sampling, manual sequence inspection and annotation, and rigorous phylogenetic analyses (Supplemental Figure 2). To better understand the scope of HGT in seed plants, 19 additional seed plant species (2 gymnosperms, 13 grasses, and 4 species from the family Solanaceae) were also included for genome screening and phylogenetic analyses (Supplemental Figure 1). In total, 31 charophyte and land-plant species were analyzed for acquired genes in this study.

HGT identification in our analyses included three main steps: whole-genome screening to identify candidate acquired genes, elimination of contamination, and phylogenetic analyses of candidate acquired genes (see following sections for details). A detailed flowchart of the analyses can be found in Supplemental Figure 2. For each candidate acquired gene identified in our analyses, we used multiple lines of evidence to assess its existence as a true gene in the sampled species. Particularly for candidate acquired genes that were found in a single sampled species, if these genes were located in short contigs and their homologs could not be confirmed in close relatives, they would be excluded from our dataset. Based on our internally generated and other publicly available genome sequence data, we were able to perform detailed and comprehensive comparative genome analyses of species- or taxon-specific HGT events in the charophyte *K. nitens*, the moss *P. patens*, the lycophyte *S. moellendorffii*, and grasses (also see following section "elimination of contamination").

Cumulative HGT in each sampled species was calculated by adding all gene families acquired specifically by the subject species and those acquired earlier by its ancestors. The timing of each transfer event was carefully inferred based on the taxonomic distribution of the acquired gene in the recipient plant group and the topology of the subject gene tree. All HGT events occurring prior to the origin of streptophytes were excluded from our data. To estimate the transfer direction, we manually inspected the distribution of the subject gene in the donor group,

and then assessed whether the gene was possibly present in the ancestor of the donor group (or in the ancestor of major subgroups of the donor group). Ages of the donor and the recipient were estimated based on the TimeTree of Life (timetree.org) or other available information. The donor taxon was required to be at least as old as the recipient (Huang and Gogarten, 2006). For instance, many acquired genes in land plants are exclusively detected in multiple major groups of bacteria (e.g., proteobacteria and actinobacteria) or fungi (e.g., both ascomycetes and basidiomycetes), suggesting that they were present in the last common ancestor of these groups. Because the ancestors of these groups emerged much earlier than land plants, the subject gene would most likely be transferred from bacteria or fungi to land plants.

A final list of acquired genes for each sampled species was determined after rigorous phylogenetic analyses for candidate acquired genes, in combination of rare molecular characters (e.g., unique gene structures, shared insertions/deletions between sequences from donor and recipient groups), and restricted taxonomic distribution (Supplemental Figure 2). Functions of the acquired genes identified in the analyses were classified into categories based on EggNOG annotation results (Huerta-Cepas et al., 2019).

Genome screening for candidate acquired genes

To identify candidates of horizontally acquired genes in each sampled genome, we performed BLAST searches against our internal customized sequence database (see "data sources"), using the annotated protein sequences of the sampled species as query (Supplemental Table 7). The following parameters were used in BLAST searches: evalue=1e-5 and max_target_seqs=50000. Screening for candidates of acquired genes was then performed using an earlier described HGT detection method (https://github.com/waterml/blast2hgt) (Li et al., 2018b). In brief, all hit species from BLAST searches were roughly divided into seven categories according to NCBI taxonomy: bacteria, fungi, archaea, viruses, metazoans, green plants, and other eukaryote species. The putative recipient was set to be groups of various taxonomic ranks represented by each sampled species (e.g., Streptophyta, Embryophyta, Tracheophyta, Spermatophyta, etc.) (Supplemental Table 7). The putative donor group was set to be bacteria (excluding cyanobacteria), fungi, archaea, viruses, or metazoans. If the top hit of BLAST results was from a putative donor group and the number of top hits (excluding hits from the recipient group) was more than 5, the query gene would be retained. Genes transferred between plants (e.g., those between grasses) were excluded from our analyses.

The following criteria were used to filter the retained genes from each sampled species: (1) protein sequence identity between the putative donor group and the recipient group must be above 20%; (2) except for viruses, the gene is found in at least 30 species of the donor group; (3) except the recipient taxon, the top 50 species in the BLAST output should overwhelmingly belong to putative donor groups (prokaryotes, viruses, fungi, or animals). An index, calculated in the following equation, was used to evaluate the distribution of the top 50 species in different groups (Yanai et al., 2005):

$$\text{index} = \frac{\sum_{i=1}^{n}\left(1 - \frac{\text{Exp } i}{\text{Exp } max}\right)}{n-1},$$

where n is the number of donor groups, Exp$i$ is the number of species in each group, and Expmax is the maximal number of species among all groups. An index value greater than 0.8 would indicate that most of the top 50 hit species belong to a given donor group.

Elimination of contamination

To exclude sequencing contamination, we first inspected the flanking regions for each candidate acquired gene on the assembled scaffold. Unless independent evidence for its existence as a true endogenous gene was available, if the candidate acquired gene was located at the end of a scaffold or in a scaffold that contains only genes of potential foreign ancestry (e.g., significantly more similar to fungal sequences than to plant sequences), it would be eliminated from further analyses. Several genomes (e.g., *M. viride* and *Picea abies*) had poor assembly quality and, as such, only the candidate genes that are also distributed in closely related species were retained (e.g., those distributed in *M. viride* and other charophytes); the remaining genes could not be confirmed without independent lines of evidence and are not discussed in this work. For candidate acquired genes having over 80% identities with homologs from the putative donor, we manually examined whether orthologs were present in closely related species; if no orthologs were found in closely related species, the candidate gene was excluded from further analyses. Except for the charophyte *K. nitens*, the moss *P. patens*, the lycophyte *S. moellendorffii*, and grasses, a vast majority of the sampled species lack quality genome sequence data from closely related taxa thus far. For *S. moellendorffii* and grasses, complete genome sequence data for multiple closely related species are publicly available, which allows comparative genome analyses to distinguish species- or taxon-specific transfer events from sequencing contamination. The identity of acquired genes in *K. nitens* was confirmed by comparing with our resequencing data for the species and our internally generated genome sequences for *I. paradoxum*, a close relative of *K. nitens*. Similarly, the identity of acquired genes in *P. patens* was confirmed by comparing with our internally generated genome sequences for *P. sphaericum*, a congeneric species of *P. patens* (see Supplemental Notes).

RT-PCR or genomic PCR verification was performed for a subset of candidate acquired genes identified in six sampled species of which we were able to obtain genomic DNA or living materials. These six species include the charophyte *K. nitens*, the bryophytes *P. patens* and *M. polymorpha*, the lycophyte *S. moellendorffii*, the fern *A. filiculoides*, and the gymnosperm *G. montanum* (Supplemental Table 8). We performed RT-PCR or genomic PCR verification for the following candidate acquired genes: (1) lacking homologs in close relatives, (2) located on a short scaffold with fewer than five genes, (3) located at ends of scaffolds/chromosomes, (4) having over 80% identity with protein sequences from the putative donor. A total of 57 candidate acquired genes were selected from the above six species for RT-PCR or genomic PCR verification.

Phylogenetic analyses of candidate acquired genes

For each candidate acquired gene identified from genome screening, all hit species generated from BLAST searches against our customized database were divided into seven categories according to the NCBI taxonomy database: bacteria, fungi, archaea, viruses, metazoans, green plants, and other eukaryotic species. The seven categories were then divided into subgroups, some of which were further divided into groups of lower taxonomic ranks based on their sizes (Supplemental Table 9). At least one sequence of each subgroup was selected to ensure sufficient and balanced samplings. Subgroups with fewer than three species were manually inspected to exclude sequencing contamination, such as mislabeled sequences in the non-redundant database.

Multiple sequence alignments were performed using MAFFT with default setting (v.7.467) (Katoh and Standley, 2013). Poorly aligned regions and gaps were removed manually or using trimAl (v.1.4) with parameters "-automated1" (Capella-Gutierrez et al., 2009). Maximum likelihood trees were constructed using IQ-TREE (v.2.0.3) (Nguyen et al., 2015), with automatically selected best-fit amino acid substitution model. Bootstrap analyses were performed using standard non-parametric bootstrap (-b 1000). Bayesian analyses were conducted using MrBayes (v.3.2.7), with the generally best-fit model inferred from IQ-TREE (Ronquist et al., 2012). Two independent runs with four chains each were calculated simultaneously for 10 million generations, sampling every 100 generations. The average standard deviation of split frequencies below 0.01 was used to ensure convergence of the runs. Posterior probabilities were generated after discarding the first 25% of the sampled trees. Phylogenetic trees were visualized using FigTree (v.1.4.4) (http://tree.bio.ed.ac.uk/software/figtree).

For convenience of display, trees were rooted based on the following principles: (1) whenever possible, ancient paralogs were used as the outgroup; (2) the root should not be placed between the most similar (or potentially the most closely related) sequences; and (3) the resulting topology should largely reflect the overall species relationships. However, all trees should be strictly interpreted as unrooted.

Expression analysis of *DIT1* gene in *Selaginella*

Transcriptome data of the lycophyte *S. moellendorffii* were downloaded from the GenBank SRA database (Supplemental Table 10), and reads were mapped to the reference genome using Histat2 (Kim et al., 2019). The output results were sorted and fragments per kilobase of exon per million mapped fragments (FPKM) values were extracted using StringTie (v.2.1.2) (Pertea et al., 2015). The heatmap was built using the pheatmap (v.1.0.8) package in R.

FB lectin purification and treatment of *P. xylostella*

The full-length coding sequence of FB lectin from the liverwort *M. polymorpha* (Mapoly0118s0047) was amplified and ligated to the pET-28a vector and then transformed in *Escherichia coli* BL 21(DE3) pLysS cells. After incubation at 37°C overnight and 0.01 mM isopropyl β-D-1-thiogalactopyranoside induction at 28°C for 6 h, the transformed strain was harvested and disrupted by sonication at 4°C. The cell lysate was centrifuged at 12 000 *g* for 40 min at 4°C in an RC5 Plus centrifuge. The supernatant from the previous step was loaded on Ni-NTA agarose column preequilibrated with Tris–NaCl buffer at 4°C. The column was washed extensively with Tris–NaCl buffer containing 20 mM imidazole, and the 6×His-tagged protein was eluted with Tris–NaCl buffer containing 250 mM imidazole. The elution product containing pure protein was washed three times with Tris–NaCl buffer and concentrated using a Centricon (Millipore PM10) (Supplemental Figure 4B).

The diamondback moth *P. xylostella* was purchased from the PilotScale Base of Bio-Pesticides, Institute of Zoology, Chinese Academy of Sciences, in 2021. *P. xylostella* was kept in a growth chamber at 27°C ± 1°C, with a relative 40%–60% humidity and 16 h light:8 h dark photoperiod, and was fed on radish seedlings (*Raphanus sativus*). The purified FB lectin from *M. polymorpha* was dissolved in phosphate-buffered saline (PBS) and injected into 2-day-old third instars of *P. xylostella* (5.41 μg/individual). The larvae injected with PBS were used as controls. A total of 30 larvae were injected for each replicate, with three replicates being used, and the mortality was recorded 24 h post injection.

RT-PCR and genomic PCR amplification

Total RNA or DNA extraction, RT-PCR, and genomic PCR were performed as previously described (Wang et al., 2020b) to eliminate the possibility of contamination from fungal endophytes or bacterial endosymbionts. Information on primers used in the amplification is provided in Supplemental Table 8.

DATA AVAILABILITY

The phylogenetic tree constructed based on 215 single-copy genes, gene IDs, protein sequences for all acquired genes identified in this work, and multiple sequence alignments used in phylogenetic analyses were uploaded into FigShare (https://doi.org/10.6084/m9.figshare.16831777). Genome and annotation files of *K. nitens*, *I. paradoxum*, and *P. sphaericum* have been submitted to the NGDC database under project ID PRJCA004838.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

FUNDING

This work is funded in part by the National Natural Science Foundation of China (31970248 and 32000176).

AUTHOR CONTRIBUTIONS

J.H. and C.P.S. conceived the study. J.M., S.W., and J.H. performed the analyses and PCR verification. X.Z., G.S., X.H., S.Z., and Y.Z. participated in the analyses. L.L. and G.C. performed experiments on fungal FB lectin. J.H., J.M., and S.W. wrote the manuscript.

ACKNOWLEDGMENTS

We thank John Stiller, Yang Liu, and Qia Wang for comments and suggestions to improve the manuscript. We are also grateful to Bojian Zhong for providing living material of *Chara braunii*, and Pei Liang and Yanlong Guan for help with related experiments. No conflict of interest is declared.

Received: August 17, 2021
Revised: December 10, 2021
Accepted: January 26, 2022
Published: March 1, 2022

REFERENCES

Abascal, F., Irisarri, I., and Zardoya, R. (2014). Diversity and evolution of membrane intrinsic proteins. Biochim. Biophys. Acta **1840**:1468–1481.

Atichartpongkul, S., Loprasert, S., Vattanaviboon, P., Whangsuk, W., Helmann, J.D., and Mongkolsuk, S. (2001). Bacterial Ohr and OsmC paralogues define two protein families with distinct functions and patterns of expression. Microbiology (Reading) **147**:1775–1782.

Belal, R., Tang, R., Li, Y., Mabrouk, Y., Badr, E., and Luan, S. (2015). An ABC transporter complex encoded by aluminum sensitive 3 and NAP3 is required for phosphate deficiency responses in arabidopsis. Biochem. Biophys. Res. Commun. **463**:18–23.

Bell, R.T., Wolf, Y.I., and Koonin, E.V. (2020). Modified base-binding EVE and DCD domains: striking diversity of genomic contexts in prokaryotes and predicted involvement in a variety of cellular processes. BMC Biol. **18**:159.

Bowles, A.M.C., Bechtold, U., and Paps, J. (2020). The origin of land plants is rooted in two bursts of genomic novelty. Curr. Biol. **30**:530–536.e2.

Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., et al. (2017). Insights into land plant evolution garnered from the Marchantia polymorpha genome. Cell **171**:287–304.e5.

La Camera, S., Balagué, C., Göbel, C., Geoffroy, P., Legrand, M., Feussner, I., Roby, D., and Heitz, T. (2009). The Arabidopsis patatin-like protein 2 (PLP2) plays an essential role in cell death execution and differentially affects biosynthesis of oxylipins and resistance to pathogens. Mol. Plant Microbe Interact. **22**:469–481.

Candat, A., Paszkiewicz, G., Neveu, M., Gautier, R., Logan, D.C., Avelange-Macherel, M.H., and Macherel, D. (2014). The ubiquitous distribution of late embryogenesis abundant proteins across cell compartments in Arabidopsis offers tailored protection against abiotic stress. Plant Cell **26**:3148–3166.

Cao, J., Han, X., Zhang, T., Yang, Y., Huang, J., and Hu, X. (2014). Genome-wide and molecular evolution analysis of the subtilase gene family in Vitis vinifera. BMC Genomics **15**:1116.

Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics **25**:1972–1973.

Chen, R., Huangfu, L., Lu, Y., Fang, H., Xu, Y., Li, P., Zhou, Y., Xu, C., Huang, J., and Yang, Z. (2021). Adaptive innovation of green plants by horizontal gene transfer. Biotechnol. Adv. **46**:107671.

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

Chong, J., Poutaraud, A., and Hugueney, P. (2009). Metabolism and roles of stilbenes in plants. Plant Sci. **177**:143–155.

Chrispeels, M.J., and Raikhel, N.V. (1991). Lectins, lectin genes, and their role in plant defense. Plant Cell **3**:1–9.

Dagan, T., Artzy-Randrup, Y., and Martin, W. (2008). Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. Proc. Natl. Acad. Sci. U S A **105**:10039–10044.

Delwiche, C.F., and Cooper, E.D. (2015). The evolutionary origin of a terrestrial flora. Curr. Biol. **25**:R899–R910.

Dixon, R.A., and Paiva, N.L. (1995). Stress-induced phenylpropanoid metabolism. Plant Cell **7**:1085–1097.

Dyall, S.D., Brown, M.T., and Johnson, P.J. (2004). Ancient invasions: from endosymbionts to organelles. Science **304**:253–257.

Emiliani, G., Fondi, M., Fani, R., and Gribaldo, S. (2009). A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. Biol. Direct. **4**:7.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. **20**:238.

Furst-Jansen, J.M.R., de Vries, S., and de Vries, J. (2020). Evo-physio: on stress responses and the earliest land plants. J. Exp. Bot. **71**:3254–3269.

Gogarten, J.P., Doolittle, W.F., and Lawrence, J.G. (2002). Prokaryotic evolution in light of gene transfer. Mol. Biol. Evol. **19**:2226–2238.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., et al. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. **40**:D1178–D1186.

Guan, Y., Liu, L., Wang, Q., Zhao, J., Li, P., Hu, J., Yang, Z., Running, M.P., Sun, H., and Huang, J. (2018). Gene refashioning through innovative shifting of reading frames in mosses. Nat. Commun. **9**:1555.

Hall, C., Welch, J., Kowbel, D.J., and Glass, N.L. (2010). Evolution and diversity of a fungal self/nonself recognition locus. PLoS One **5**:e14055.

Haney, C.H., and Long, S.R. (2010). Plant flotillins are required for infection by nitrogen-fixing bacteria. Proc. Natl. Acad. Sci. U S A **107**:478–483.

Hartmann, T. (2007). From waste products to ecochemicals: fifty years research of plant secondary metabolism. Phytochemistry **68**:2831–2846.

Hoang, Q.T., Cho, S.H., McDaniel, S.F., Ok, S.H., Quatrano, R.S., and Shin, J.S. (2009). An actinoporin plays a key role in water stress in the moss Physcomitrella patens. New Phytol. **184**:502–510.

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.

Huang, J. (2013). Horizontal gene transfer in eukaryotes: the weak-link model. Bioessays **35**:868–875.

Huang, J., and Gogarten, J.P. (2006). Ancient horizontal gene transfer can benefit phylogenetic reconstruction. Trends Genet. **22**:361–366.

Huang, J., and Gogarten, J.P. (2008). Concerted gene recruitment in early plant evolution. Genome Biol. **9**:R109.

Huang, J., and Yue, J. (2013). Horizontal gene transfer in the evolution of photosynthetic eukaryotes. J. Syst. Evol. **51**:13–29.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. **47**:D309–D314.

Husnik, F., and McCutcheon, J.P. (2018). Functional horizontal gene transfer from bacteria to eukaryotes. Nat. Rev. Microbiol. **16**:67–79.

Imai, S., Tsuge, N., Tomotake, M., Nagatome, Y., Sawada, H., Nagata, T., and Kumagai, H. (2002). Plant biochemistry: an onion enzyme that makes the eyes water. Nature **419**:685.

Jia, Q., Li, G., Kollner, T.G., Fu, J., Chen, X., Xiong, W., Crandall-Stotler, B.J., Bowman, J.L., Weston, D.J., Zhang, Y., et al. (2016). Microbial-type terpene synthase genes occur widely in nonseed land plants, but not in seed plants. Proc. Natl. Acad. Sci. U S A **113**:12328–12333.

Jun, J.H., Ha, C.M., and Nam, H.G. (2002). Involvement of the VEP1 gene in vascular strand development in Arabidopsis thaliana. Plant Cell Physiol. **43**:323–330.

Kameshwar, A.K.S., and Qin, W. (2018). Structural and functional properties of pectin and lignin–carbohydrate complexes de-esterases: a review. Bioresour. Bioprocess. **5**:43.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. **30**:772–780.

Keeling, P.J., and Palmer, J.D. (2008). Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Genet. **9**:605–618.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. **37**:907–915.

Ku, C., Nelson-Sathi, S., Roettger, M., Sousa, F.L., Lockhart, P.J., Bryant, D., Hazkani-Covo, E., McInerney, J.O., Landan, G., and Martin, W.F. (2015). Endosymbiotic origin and differential loss of eukaryotic genes. Nature **524**:427–432.

Lannoo, N., and Van Damme, E.J. (2014). Lectin domains at the frontiers of plant defense. Front. Plant Sci. **5**:397.

Larsen, P.B., Geisler, M.J., Jones, C.A., Williams, K.M., and Cancel, J.D. (2005). ALS3 encodes a phloem-localized ABC transporter-like protein that is required for aluminum tolerance in Arabidopsis. Plant J. **41**:353–363.

Li, R., Liu, P., Wan, Y., Chen, T., Wang, Q., Mettbach, U., Baluska, F., Samaj, J., Fang, X., Lucas, W.J., et al. (2012). A membrane microdomain-associated protein, Arabidopsis Flot1, is involved in a clathrin-independent endocytic pathway and is required for seedling development. Plant Cell **24**:2105–2122.

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. (2018a). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. Nat. Plants **4**:460–472.

Li, M., Zhao, J., Tang, N., Sun, H., and Huang, J. (2018b). Horizontal gene transfer from bacteria and plants to the arbuscular mycorrhizal fungus rhizophagus irregularis. Front. Plant Sci. **9**:701.

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

Marquez, N., Giachero, M.L., Gallou, A., Debat, H.J., Cranenbrouck, S., Di Rienzo, J.A., Pozo, M.J., Ducasse, D.A., and Declerck, S. (2018). Transcriptional changes in mycorrhizal and nonmycorrhizal soybean plants upon infection with the fungal pathogen macrophomina phaseolina. Mol. Plant Microbe Interact. **31**:842–855.

Martin, F.M., Uroz, S., and Barker, D.G. (2017). Ancestral alliances: plant mutualistic symbioses with fungi and bacteria. Science **356**, eaad4501.

McDonald, T.R., Dietrich, F.S., and Lutzoni, F. (2012). Multiple horizontal gene transfers of ammonium transporters/ammonia permeases from prokaryotes to eukaryotes: toward a new functional and evolutionary classification. Mol. Biol. Evol. **29**:51–60.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. **32**:268–274.

Ogawa, S., and Mitsuya, S. (2012). S-methylmethionine is involved in the salinity tolerance of Arabidopsis thaliana plants at germination and early growth stages. Physiol. Plant **144**:13–19.

One Thousand Plant Transcriptomes Initiative. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. Nature **574**:679–685.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. **33**:290–295.

Peumans, W.J., and Van Damme, E.J. (1995). Lectins as plant defense proteins. Plant Physiol. **109**:347–352.

Portereiko, M.F., Sandaklie-Nikolova, L., Lloyd, A., Dever, C.A., Otsuga, D., and Drews, G.N. (2006). Nuclear fusion defective1 encodes the Arabidopsis RPL21M protein and is required for karyogamy during female gametophyte development and fertilization. Plant Physiol. **141**:957–965.

Price, D.C., Chan, C.X., Yoon, H.S., Yang, E.C., Qiu, H., Weber, A.P., Schwacke, R., Gross, J., Blouin, N.A., Lane, C., et al. (2012). Cyanophora paradoxa genome elucidates origin of photosynthesis in algae and plants. Science **335**:843–847.

Puttick, M.N., Morris, J.L., Williams, T.A., Cox, C.J., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Schneider, H., Pisani, D., et al. (2018). The interrelationships of land plants and the nature of the ancestral embryophyte. Curr. Biol. **28**:733–745.e2.

Qiu, H., Price, D.C., Weber, A.P., Facchinelli, F., Yoon, H.S., and Bhattacharya, D. (2013). Assessing the bacterial contribution to the plastid proteome. Trends Plant Sci. **18**:680–687.

Reboledo, G., Agorio, A.D., Vignale, L., Batista-García, R.A., and Ponce De León, I. (2021). Transcriptional profiling reveals conserved and species-specific plant defense responses during the interaction of Physcomitrium patens with Botrytis cinerea. Plant Mol. Biol. **107**:365–385.

Renault, H., El Amrani, A., Palanivelu, R., Updegraff, E.P., Yu, A., Renou, J.P., Preuss, D., Bouchereau, A., and Deleu, C. (2011). GABA accumulation causes cell elongation defects and a decrease in expression of genes encoding secreted and cell wall-related proteins in Arabidopsis thaliana. Plant Cell Physiol. **52**:894–908.

Ribeiro, C.L., Conde, D., Balmant, K.M., Dervinis, C., Johnson, M.G., McGrath, A.P., Szewczyk, P., Unda, F., Finegan, C.A., Schmidt, H.W., et al. (2020). The uncharacterized gene EVE contributes to vessel element dimensions in Populus. Proc. Natl. Acad. Sci. U S A **117**:5059–5066.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. **61**:539–542.

Sarkar, P., Bosneaga, E., and Auer, M. (2009). Plant cell walls throughout evolution: towards a molecular understanding of their design principles. J. Exp. Bot. **60**:3615–3635.

Schaller, A., Stintzi, A., Rivas, S., Serrano, I., Chichkova, N.V., Vartapetian, A.B., Martínez, D., Guiamét, J.J., Sueldo, D.J., van der Hoorn, R.A.L., et al. (2018). From structure to function - a family portrait of plant subtilases. New Phytol. **218**:901–915.

Scherer, G.F., Ryu, S.B., Wang, X., Matos, A.R., and Heitz, T. (2010). Patatin-related phospholipase A: nomenclature, subfamilies and functions in plants. Trends Plant Sci. **15**:693–700.

Shinozuka, H., Hettiarachchige, I.K., Shinozuka, M., Cogan, N.O.I., Spangenberg, G.C., Cocks, B.G., Forster, J.W., and Sawbridge, T.I. (2017). Horizontal transfer of a ß-1,6-glucanase gene from an ancestral species of fungal endophyte to a cool-season grass host. Sci. Rep. **7**:9024.

Shinozuka, H., Shinozuka, M., de Vries, E.M., Sawbridge, T.I., Spangenberg, G.C., and Cocks, B.G. (2020). Fungus-originated genes in the genomes of cereal and pasture grasses acquired through ancient lateral transfer. Sci. Rep. **10**:19883.

Soucy, S.M., Huang, J., and Gogarten, J.P. (2015). Horizontal gene transfer: building the web of life. Nat. Rev. Genet. **16**:472–482.

Sun, G., Yang, Z., Ishwar, A., and Huang, J. (2010). Algal genes in the closest relatives of animals. Mol. Biol. Evol. **27**:2879–2889.

Sun, G., Bai, S., Guan, Y., Wang, S., Wang, Q., Liu, Y., Liu, H., Goffinet, B., Zhou, Y., Paoletti, M., et al. (2020). Are fungi-derived genomic regions related to antagonism toward fungi in mosses? New Phytol. **228**:1169–1175.

de Vries, S., Furst-Jansen, J.M.R., Irisarri, I., Dhabalia Ashok, A., Ischebeck, T., Feussner, K., Abreu, I.N., Petersen, M., Feussner, I., and de Vries, J. (2021). The evolution of the phenylpropanoid pathway entailed pronounced radiations and divergences of enzyme families. Plant J. **107**:975–1002.

Wallace, I.S., Choi, W.G., and Roberts, D.M. (2006). The structure, function and regulation of the nodulin 26-like intrinsic protein family of plant aquaglyceroporins. Biochim. Biophys. Acta **1758**:1165–1175.

Wang, M., Triguéros, V., Paquereau, L., Chavant, L., and Fournier, D. (2002). Proteins as active compounds involved in insecticidal activity of mushroom fruitbodies. J. Econ. Entomol. **95**:603–607.

Wang, M., Ding, L., Gao, L., Li, Y., Shen, Q., and Guo, S. (2016). The interactions of aquaporins and mineral nutrients in higher plants. Int. J. Mol. Sci. **17**:1229.

Wang, H., Sun, S., Ge, W., Zhao, L., Hou, B., Wang, K., Lyu, Z., Chen, L., Xu, S., Guo, J., et al. (2020a). Horizontal gene transfer of Fhb7 from fungus underlies Fusarium head blight resistance in wheat. Science **368**, eaba5435.

Wang, S., Guan, Y., Wang, Q., Zhao, J., Sun, G., Hu, X., Running, M.P., Sun, H., and Huang, J. (2020b). A mycorrhizae-like gene regulates stem cell and gametophore development in mosses. Nat. Commun. **11**:2030.

Wang, X., Ding, J., Lin, S., Liu, D., Gu, T., Wu, H., Trigiano, R.N., McAvoy, R., Huang, J., and Li, Y. (2020c). Evolution and roles of cytokinin genes in angiosperms 2: do ancient CKXs play housekeeping roles while non-ancient CKXs play regulatory roles? Hortic. Res. **7**:29.

Wang, Q., Smith, S.M., and Huang, J. (2021). Origins of strigolactone and karrikin signaling in plants. Trends Plant Sci. https://doi.org/10.1016/j.tplants.2021.11.009.

Wetzel, S., Demmers, C., and Greenwood, J.S. (1989). Seasonally fluctuating bark proteins are a potential form of nitrogen storage in three temperate hardwoods. Planta **178**:275–281.

Whitaker, J.W., McConkey, G.A., and Westhead, D.R. (2009). The transferome of metabolic genes explored: analysis of the horizontal transfer of enzyme encoding genes in unicellular eukaryotes. Genome Biol. **10**:R36.

Widhalm, J.R., Ducluzeau, A.L., Buller, N.E., Elowsky, C.G., Olsen, L.J., and Basset, G.J. (2012). Phylloquinone (vitamin K(1)) biosynthesis in plants: two peroxisomal thioesterases of Lactobacillales origin hydrolyze 1,4-dihydroxy-2-naphthoyl-CoA. Plant J. **71**:205–215.

Wu, H.C., Bulgakov, V.P., and Jinn, T.L. (2018). Pectin methylesterases: cell wall remodeling proteins are required for plant response to heat stress. Front. Plant Sci. **9**:1612.

Xu, Y., Wang, S., Li, L., Sahu, S.K., Petersen, M., Liu, X., Melkonian, M., Zhang, G., and Liu, H. (2019). Molecular evidence for origin, diversification and ancient gene duplication of plant subtilases (SBTs). Sci. Rep. **9**:12485.

Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., et al. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics **21**:650–659.

Yang, W.Y., Zheng, Y., Bahn, S.C., Pan, X.Q., Li, M.Y., Vu, H.S., Roth, M.R., Scheu, B., Welti, R., Hong, Y.Y., et al. (2012). The patatin-containing phospholipase A pPLAIIα modulates oxylipin formation and water loss in Arabidopsis thaliana. Mol. Plant **5**:452–460.

Yang, Z., Zhou, Y., Huang, J., Hu, Y., Zhang, E., Xie, Z., Ma, S., Gao, Y., Song, S., Xu, C., et al. (2015). Ancient horizontal transfer of transaldolase-like protein gene and its role in plant vascular development. New Phytol. **206**:807–816.

Yue, J., Hu, X., Sun, H., Yang, Y., and Huang, J. (2012). Widespread impact of horizontal gene transfer on plant colonization of land. Nat. Commun. **3**:1152.

Yue, J., Sun, G., Hu, X., and Huang, J. (2013). The scale and evolutionary significance of horizontal gene transfer in the choanoflagellate Monosiga brevicollis. BMC Genomics **14**:729.

Yue, J., Hu, X., and Huang, J. (2014). Origin of plant auxin biosynthesis. Trends Plant Sci. **19**:764–770.

Zardoya, R., Ding, X., Kitagawa, Y., and Chrispeels, M.J. (2002). Origin of plant glycerol transporters by horizontal gene transfer and functional recruitment. Proc. Natl. Acad. Sci. U S A. **99**:14893–14896.