In complex multicellular eukaryotes such as animals and plants, horizontal gene transfer is commonly considered rare with very limited evolutionary significance. Here we show that horizontal gene transfer is a dynamic process occurring frequently in the early evolution of land plants. Our genome analyses of the moss *Physcomitrella patens* identified 57 families of nuclear genes that were acquired from prokaryotes, fungi or viruses. Many of these gene families were transferred to the ancestors of green or land plants. Available experimental evidence shows that these anciently acquired genes are involved in some essential or plant-specific activities such as xylem formation, plant defence, nitrogen recycling as well as the biosynthesis of starch, polyamines, hormones and glutathione. These findings suggest that horizontal gene transfer had a critical role in the transition of plants from aquatic to terrestrial environments. On the basis of these findings, we propose a model of horizontal gene transfer mechanism in nonvascular and seedless vascular plants.
Horizontal gene transfer (HGT) is the process of genetic movement between species. Traditionally considered to be predominant in prokaryotes\(^1\), HGT now appears to be widespread in microbial eukaryotes\(^2\). As an efficient mechanism to spread evolutionary success, HGT may introduce genetic novelties to recipient organisms, thus facilitating phenotypic variation and adaptation to shifting environments or allowing access to new resources. The novelties introduced by HGT range from virulence factors in pathogens\(^3,4\), food digestive enzymes in nematodes and rumen ciliates\(^5,6\), to anaerobic metabolism in intracellular parasites\(^7,8\).

Although HGT in prokaryotes and unicellular eukaryotes has been under some extensive studies and well documented\(^9-10\), how HGT has contributed to the evolution of complex multicellular eukaryotes, such as animals and plants, remains elusive. Presumably because of the barrier of germline in animals and apical meristem in plants\(^11\), HGT is generally believed to be rare and insignificant in complex multicellular eukaryotes, except for organisms in a symbiotic relationship\(^12,13\) and for plant mitochondrial genes\(^14,15\). This belief, however, has been cast in doubt by reports of acquired genes in invertebrates and plants from free-living organisms\(^16-18\). Importantly, because all multicellular eukaryotes are derived from unicellular ancestors, this belief largely discounts the dynamic nature of HGT and the contribution of ancient HGT to the evolution of multicellular lineages\(^19\). Therefore, to better understand the role of HGT in eukaryotic evolution, it is critical to reassess the occurrence and biological functions of horizontally acquired genes in multicellular eukaryotes.

Land plants emerged from charophycean green algae about 480–490 million years ago\(^20\). During their colonization of land, plants gradually evolved complex regulatory systems, body plan and phenotypic novelties that facilitated their adaptation to and radiation in terrestrial environments\(^21\). Because of the importance of HGT in the adaptation of organisms to new niches, we decided to investigate whether such habitat and developmental transition was aided by acquisition of novel genes, especially those during early evolution of land plants. Thus far, although the role of HGT in the evolution of land plants, especially flowering plants, has long been speculated\(^22\), there are very few reported cases of HGT in land plants that are related to nuclear gene\(^23-25\). We here present evidence for the widespread and significant impact of HGT of nuclear genes on plant colonization of land based on analyses of the moss Physcomitrella patens, an extant representative of early land plants.

We further propose a model for gene acquisition in nonvascular and seedless vascular plants and discuss the cumulative impact of HGT on multicellular eukaryotes.

### Results

**HGT-derived genes in land plants.** Eukaryotic genomes contain many genes of prokaryotic origin, most of which are derived from mitochondria and plastids\(^26\). Gene transfer from these organelles to the nucleus, often called endosymbiotic gene transfer (EGT), has been studied in many eukaryotic groups\(^27-29\) and will not be included here. In this study, we identified genes in land plants that were acquired independently from other sources, primarily based on phylogenomic analyses of the moss *Physcomitrella patens*. Whenever possible, independent evidence such as restricted taxonomic distribution and uniquely shared genomic characters (for example, indels or domain structures) were also considered. To reduce the complication arising from differential gene losses, we focused on identifying genes acquired from prokaryotes and viruses. Genes acquired from fungi were also identified because of the role of mycorrhizae in land plant evolution\(^30\) and available evidence for HGT between mycorrhizal partners\(^25\). Furthermore, because genes acquired by the common ancestor of Plantae (green plants, red algae and glaucophytes) have been under some detailed analyses\(^31-33\), this study only identified genes in *P. patens* that were acquired after the separation of green plants from red algae and glaucophytes.

With the annotated protein sequences of *P. patens* as input, 910 genes were identified using AlienG\(^34\) as potentially of prokaryotic, fungal or viral origin. Among these 910 genes, 394 were removed from further analyses because of their locations on short scaffolds or their high percent-identities with cyanobacterial sequences, which is often suggestive of plastid origin. Of the remaining 516 genes, 32 genes of four families had identifiable homologues only in prokaryotes or fungi; 96 genes of 53 families showed a monophyletic relationship between sequences of green plants and those from prokaryotes, fungi or viruses in phylogenetic analyses, with bootstrap support of 80% or higher from either maximum likelihood or distance analyses or both. In total, 128 genes of 57 families were identified as derived from prokaryotes, fungi or viruses (Table 1; Figs 1 and 2; Supplementary Information). Twenty-four of these gene families in green plants also share unique indels and amino acid residues with their putative donors. The online Supplementary Data show the taxonomic distributions, multiple sequence alignments, molecular phylogenies and other relevant information for the 57 gene families we have identified in this study.

Of the 57 gene families, 18 are present in both green algae and land plants, suggesting that they were likely acquired before the origin of land plants. The remaining 39 gene families are not found in green algae and might have been acquired during or after the origin of land plants. Notably, 19 of the identified gene families are only found in *P. patens* and their putative donors (prokaryotes, viruses or fungi) (Table 1; Supplementary Table S1). All of these 19 families are located on large genomic scaffolds, indicating that they are unlikely to be bacterial contaminants. As *P. patens* is the only moss whose complete genome sequence is available, it is unclear whether these families also exist in other mosses or nonvascular land plants. However, the lack of homologues for these gene families in vascular plants suggests that they were likely transferred more recently to either *P. patens* or its close relatives.

The vast majority of acquired gene families identified in our analyses are derived from miscellaneous bacterial lineages. Ten families are derived from fungi, and only one family is from archaea and viruses, respectively. As expected for land plants, which have often undergone frequent duplication events, 25 of identified gene families contain multiple copies in *P. patens*. In some cases, both acquired genes and endogenous homologues co-exist in *P. patens*. For instance, the gene family encoding FAD-linked oxidase comprises three identifiable copies in *P. patens*, two of which are closely related to CFB bacterial homologues and one may have been vertically inherited in eukaryotes (Supplementary Fig. S1). A similar evolutionary scenario is also observed for the gene encoding phosphoenolpyruvate carboxylase (PEPCase). In this case, two PEPCase gene copies exist in *P. patens*, one of which is clearly related to proteobacterial sequences, whereas the other to those from photosynthetic eukaryotes, the chytrid fungus *Spizellomyces*, and other bacteria (Supplementary Fig. S2).

As HGT identification can be prone to errors owing to poor data quality and methodological limitations\(^19\), we have taken very cautious measures to alleviate these issues. These measures include construction of a comprehensive database, broad and balanced taxonomic sampling, careful inspection of alignments, determination of optimal protein substitution matrix for each data set and detection of other molecular characters consistent with the identified relationships. Such measures may have reduced most of the artifacts commonly encountered in HGT detection. It is critical to note that, although differential gene loss, sometimes associated with hidden paralogy, can always be invoked as an alternative explanation, HGT is the most parsimonious interpretation for the genes identified in Table 1. This interpretation is consistent with independent evidence such as shared indels and amino-acid residues for many identified
Table 1 | Horizontally acquired genes identified in Physcomitrella patens.

| Putative gene product | Putative donor | Functional category | Figure | Homologous locus in Arabidopsis |
|-----------------------|----------------|---------------------|--------|-------------------------------|
| Subtilase family (10) | Bacteria       | Proteolysis         |        | AT4G30020                     |
| Arginase              | Bacteria       | Polyamine biosynthesis |      | AT4G08900                    |
| Acyl-activating enzyme 18 (AAEI8) (3) | Bacteria | Auxin biosynthesis | Figure S3 | AT1G55320 |
| YUCCA family monoxygenase (YUC3) (5) | Bacteria | Auxin biosynthesis | Figure S4 | AT1G04610 |
| Glutamate-cysteine ligase (GCL) (3) | Proteobacteria | Glutathione synthesis | Figure S5 | AT4G23100 |
| Wound-responsive family protein (6) | Bacteria | Defense response | Figure S39 | AT1G19660 |
| HAD superfamily, subfamily IIIb acid phosphatase (4) | Bacteria | Herbivorous insect resistance | Figure S42 | AT4G29260 |
| NRPS-like enzyme      | Fungi          | Oxidative stress resistance | Figure S44 | AT4G18540 |
| N-acetyl-gamma-glutamyl-phosphate reductase (argC) (2) | Alpha-proteobacteria | Cadmium stress response | Figure S43 | AT2G19940 |
| HAD-superfamily hydrolyase | Bacteria | Cold stress response | Figure S18 | AT5G48960 |
| Killer toxin Protein (KP4) (2) | Ascomycetes | Pathogen resistance | Figure S49 | No |
| Flotilloin-like protein | Ascomycetes | Endocytosis | Figure S45 | AT1G25260 |
| Allantoate amidohydrolase (AAH) (2) | Bacteria | Purine degradation | Figure S7 | AT4G20070 |
| Ureidoglycolate amidohydrolase (UAH) | Bacteria | Purine degradation | Figure S7 | AT5G43600 |
| Guanine deaminase (GDA) | Alpha-proteobacteria | Purine degradation | Figure S6 | No |
| PfKb family kinase (3) | Delta-proteobacteria | Vitamin B6 salvaging | Figure S36 | AT5G58730 |
| Methionine gamma-lyase (MGL) (2) | CFB bacteria | l-methionine degradation | Figure S20 | AT1G64660 |
| Glutamine synthetase (GS) | CFB bacteria | Glutamine biosynthesis | Figure S8 | No |
| 3,4-Dihydroxy-2-butanoate 4-phosphate synthase (ribB) | Euryarchaeotes | Riboflavin biosynthesis | Figure S13 | No |
| Hemerythrin HHE domain protein | Ascomycetes | Iron homeostasis | Figure S46 | No |
| Hydroxyprolylpeptidase reductase 2 (HPR2) (2) | Alpha-proteobacteria | Pollen germination and tube growth | Figure S27 | AT4G17370 |
| Inositol 2-dehydrogenase like protein | Ascomycetes | Peptidoglycan binding | Figure S50 | No |
| Peptidoglycan-binding domain containing protein | Alpha-proteobacteria | Sugar binding | Figure S24 | AT5G52190 |
| Sugar isomerase (SIS) family (2) | Bacteria | Starch biosynthesis | Figure S14 | AT5G04360 |
| Limit dextrinase (LDA) | Bacteria | Cellulose degradation | Figure S15 | AT5G04885 |
| Beta-glucosidase (2) | Ascomycetes | Carbohydrate metabolism | Figure S47 | AT3G26140 |
| Gycosyl hydrolase family (2) | Delta-proteobacteria | Carbohydrate metabolism | Figure S23 | No |
| Glycoside hydrolase | Gamma-proteobacteria | Carbohydrate metabolism | Figure S25 | AT3G54440 |
| Alpha-l-rhamnosidase | CFB bacteria | Carbohydrate metabolism | Figure S33 | No |
| FAD-linked oxidase (2) | CFB bacteria | Oxygen-dependent oxidoreductases | Figure S1 | No |
| Short-chain dehydrogenase/reductase SDR | Proteobacteria | Oxidation-reduction | Figure S28 | No |
| Fatty acyl-ACP thioesterases B (FATB)(5) | Bacteria | Fatty acid biosynthesis | Figure S41 | AT1G08510 |
| 1,4-dihydroxy-2-naphthoate octaprenyltransferase | Delta-proteobacteria | Menaquinone biosynthesis | Figure S37 | No |
| Phosphoenolpyruvate carboxylase (PEPCase) | Gamma-proteobacteria | Carbon fixation | Figure S2 | No |
| GroES-like zinc-binding alcohol dehydrogenase family | High GC gram + | Glycolysis | Figure S16 | AT5G63620 |
| Pyruvate kinase (2) | Bacteria | Glycolysis | Figure S21 | AT3G49160 |
| Phosphoglycerate kinase (PGK) (2) | Delta-proteobacteria | Glycolysis | Figure S17 | No |
| ATP-binding cassette II (ABCII) transporter | Bacteria | Molecular transport | Figure S29 | AT1G63270 |
| Uracil permease (2) | Bacteria | Nucleobase transport | Figure S30 | AT5G03555 |
| l-fucose permease* | Ascomycetes | Sugar transport | Figure S51 | No |
| Beta-1,4-mannosyl-glycoprotein (2) | Basidiozymes | Glycosyl transferring | Figure S48 | AT5G14480 |
| DNA repair family protein | Ascomycetes | DNA replication | Figure S12 | No |
| Toprim domain-containing protein | Bacteria | DNA replication | Figure S12 | No |
| DNA topoisomerase I | Protoproteobacteria | DNA replication | Figure S10 | AT4G31210 |
| Phage/plasmid primase, P4 family (5) | Viruses | DNA replication | Figure S11 | No |
| Ribosomal protein S6 | Beta-proteobacteria | RNA binding | Figure S40 | No |
| M6 family peptidase (3) | Bacteria | Peptidase activity | Figure S35 | No |
| Amidohydrolase family | Bacteria | Hydrolase activity | Figure S31 | No |
| Amidase family protein (2) | Bacteria | Acrylonitrile metabolism | Figure S32 | AT5G07360 |
| d-alanine-d-alanine ligase family | Chloramiidae/CFB | Peptidoglycan biosynthesis | Figure S34 | AT3G08840 |
| Diene lactone hydrolase family | Bacteria | Hydrolase activity | Figure S38 | No |
| Vein Patterning 1 (VEP1) | Bacteria | Vascular development | Figure 1B | AT4G24220 |
| Heterokaryon incompatibility (HET) superfamily (20) | Fungi | Heterokaryon formation | No | No |
| ybU protein | High GC gram + | Unknown | Figure S22 | No |
| Acyl-CoA N-acetyltransferase | Alpha-proteobacteria | Unknown | Figure S19 | AT2G23390 |
| Hypothetical protein* | Ascomycetes | Unknown | Figure S52 | No |

*Note: numbers in the brackets indicate the numbers of genes within each family. The heterokaryon incompatibility (HET) superfamily was identified based on its restricted taxonomic distribution.

*Genes that were also reported by earlier studies.
gene families (Fig. 2; Supplementary Information). On the other hand, the number of acquired genes in *P. patens* may have been underestimated in this study for several reasons. First, our study is primarily based on phylogenetic analysis, which, despite being considered the most reliable approach for HGT detection, tends to have more false negatives owing to the lack of sufficient phylogenetic signal in many data sets. Second, only genes transferred from prokaryotes, viruses and fungi to plants were included in our results, those from other eukaryotes were not detected. Third, our results only include genes derived from a single HGT event (that is, genes transferred directly from their ultimate donors to mosses or to recent common ancestors of green plants). This might overlook genes involved in secondary or recurrent transfer events, which often lead to complex and patchy distributions. Finally, our results are based solely on the analyses of the *P. patens* genome. Acquired genes in other plant species or secondarily lost in *P. patens* are not included. Therefore, our current results may only be viewed as a glimpse of acquired genes in land plants.

**HGT in plant development and adaptation.** Many of the genes identified in our analyses are related to essential or plant-specific metabolic and developmental processes (Table 1). Multiple gene families related to carbohydrate metabolism were acquired from bacteria, and they are involved in starch biosynthesis, cellulose degradation, pollen and seed germination as well as other activities in *Arabidopsis*. Another notable example is the large and versatile subtilase gene family. With subtilases of *P. patens* as queries, we were able to identify homologues only in bacteria and other land plants. Such sequence similarity is consistent with earlier reports that plant subtilases differ significantly from those of fungi and animals. Further phylogenetic analyses indicate that land plant subtilases are derived from a single HGT event from bacteria, followed by rapid gene duplication (Fig. 1a).

Our analyses also identified genes related to biosynthesis of plant polyamines and hormones. The gene encoding arginine is responsible for degrading arginine into ornithine, a major precursor for the biosynthesis of polyamines. Sequences of land plant arginase share 32–48% identities with those of bacterial agmatinase, but only 25–28% identities with arginine of other organisms. Consistent with the results of sequence comparisons, phylogenetic analyses indicate that land plant arginase evolved from bacterial agmatinase (Supplementary Fig. S3). At least two acquired gene families, including those encoding acyl-activating enzyme 18 (AAE18) and YUCCA flavin monooxygenase (YUC3), are involved in the biosynthesis of auxin, a hormone that regulates abscission suppression, apical dominance, cell elongation and xylem differentiation. Both AAE18 and YUC3 families were likely acquired from bacteria (Fig. 2; Supplementary Fig. S4). In particular, plant AAE18 sequences share multiple conserved amino-acid residues and indels with homologues from plant communities, verrucomicrobia and CFB bacteria.
Figure 2 | Multiple sequence alignment (a) and molecular phylogeny (b) of acyl-activating enzymes 18 (AAE18). Boxed columns indicate the amino-acid residues and indels shared by bacterial and green plant AAE18 sequences. Numbers above branches show bootstrap values from maximum likelihood and distance analyses, respectively. Asterisks indicate values < 50%.
(Fig. 2a). Intriguingly, both the production and inhibition of auxin may be affected by the expression of acquired genes. In Arabidopsis, the bacteria-derived arginase (see above) may negatively regulate the production of auxin by reducing the level of nitric oxide, which in turn mediates the induction of auxin in roots.41

Several other acquired gene families identified in our analyses are related to plant defence and stress tolerance (Table 1). Notably, glutathione is essential for plant disease resistance, photo-oxidative stress defence and heavy metal detoxification. Glutamate-cysteine ligase (GCL) is the first of the two enzymes catalysing the formation of glutathione. Identifiable homologues of P. patens GCL are only present in green plants and bacteria. Our phylogenetic analyses also show that the GCL gene was acquired from bacteria (Supplementary Fig. S5), which is consistent with an earlier report.42 In addition, at least three gene families acquired from bacteria, including guanine deaminase, allantoiate amidohydrolase and ureidoglycolate amidohydrolase,43 are involved in purine degradation and nitrogen recycling (Table 1; Supplementary Figs S6 and S7). Furthermore, another acquired gene, glutamine synthetase, is directly responsible for assimilating ammonia into amino acids in plants (Supplementary Fig. S8).

Discussion

Conventional belief is that HGT is frequent in unicellular eukaryotes but rare in multicellular eukaryotes because of the barriers of germline and apical meristem. Although evidence of HGT in multicellular eukaryotes is still limited, there have been numerous reports of acquired genes (including those of viral and bacterial origins) in mitochondrial genomes of seed plants.44-46 These viral and bacterial genes were integrated into mitochondria and passed onto descendants ultimately through the apical meristem. Such observations, combined with other relatively recent HGT events reported in plants,13,17 and animals,16,18,46 suggest that neither germline nor apical meristem constitutes an insurmountable barrier to HGT.

The finding of 18 recently acquired gene families in mosses also raises questions why more foreign genes exist in this lineage and whether recent HGT of nuclear genes also occurs in other land plants. We reason that the acquisition of genes by mosses might largely be attributed to the unique evolutionary position and biological features of this lineage. As mosses were among the first dwellers on land, they might have encountered hostile environments with intense ultraviolet radiation,47 which could break large DNA molecules into small fragments and release them into the environment. It is also known that mosses are effective in DNA transformation.48 This ability to uptake foreign DNA, including beneficial genes from co-inhabitants, likely facilitated the establishment of these early land plants in a hostile and shifting environment. In addition, these early land plants formed mycorrhizal association with diverse fungal species,30,49 and this symbiotic relationship provided further opportunities for gene transfer between fungi and early land plants.25

Mosses also have distinct and dominant gametophytes in their lifecycle. As one of the earliest plant groups on land, mosses lack true vascular systems and complex protective structures for gametes and zygotes. We hypothesize that at least two entry points exist for foreign genes into mosses and zygotes. We infer that the exterior environment to exposed zygotes and young embryos will likely be fixed and passed onto adult sporophytes. The second likely entry point for acquired genes in mosses is the stage of fertilization and early embryo development. Unlike seed plants where eggs are protected within ovules and fertilization entails a precise mechanism for pollen tube elongation and sperm delivery, mosses conceal eggs in single-layered and hollow archegonia, which are open during fertilization. Any foreign genes transferred from the exterior environment to exposed zygotes and young embryos will likely be fixed and passed onto adult sporophytes.

The above model presumes that organisms with unprotected or weakly protected zygotes in their lifecycles are prone to HGT. This model predicts the existence of recently acquired genes in plants with independent, though sometimes reduced, gametophytes such as nonvascular and seedless vascular plants. The gradual transition of these early-branching land plants toward seed plants, this model also predicts the existence of anciently acquired genes in gymnosperms and angiosperms, where fertilization and embryogenesis are structurally internalized. It should be noted here that even such structural internalization might not entirely exclude recent HGT in gymnosperms and angiosperms. It is conceivable that pollen grains from distantly related plants may be deposited on the stigma of another plant, allowing foreign pollen DNA the chance to be transformed into the zygote and the young embryo.17

The increasing structural complexity of land plants has been accompanied by diversified metabolic pathways and their chemical output. Like other complex multicellular eukaryotes, plants are able to form distinctive structures and coordinate development throughout their lifecycle. Our data clearly show that HGT contributed greatly to the metabolism, development and regulation of land plants. For example, members of the subtilase family participate in many biological processes, including protein degradation in seeds and fruits, lateral root formation, xylem differentiation, cuticle and epidermal development, and stomata pattern formation.30-50 Likewise, polyamines are involved in numerous important biological activities in plants such as translation, cell proliferation and signaling, ion channel regulation, and stress response.53,54 Furthermore, plant hormones have a vital role in regulating cell differentiation and structural development.

Land plants are also diverse in morphology, life history and habitat, and they have evolved many adaptive traits essential for their survival and development. Particularly during their transition from aquatic to terrestrial environments, plants evolved features to not only tolerate abiotic stress such as desiccation, fluctuating temperature and nutrient limitation, but also defend themselves against herbivory and microbial infection. Many of the acquired genes identified in our analyses are either directly or indirectly related to plant defence and stress tolerance. For instance, polyamines not only regulate calcium homeostasis and stomatal closure, but also are involved in plant tolerance to abiotic stress such as drought, salt and cold.55 Given the role of arginase in polyamine biosynthesis, the acquisition of the arginase gene might benefit plants greatly as they adapt to water shortages, salinity and fluctuating temperatures on land. Similarly, the involvement of subtilases in the development of lateral roots, cuticle and stomatal cells also points to an important role of this gene family in water conduction as well as protection from desiccation and microbial infection in land plants. Additionally, several gene families identified in our analyses are functionally related to DNA replication and repair (Table 1; Supplementary Figs S9, S10, S11 and S12). Given the fact that early land plants faced ubiquitous and intense ultraviolet radiation on earth surface (which might cause DNA damage and consequently interrupt the normal cell cycle of plants), the acquisition of these genes may have conferred early land plants additional abilities to fix DNA damage and facilitate their survival. Such DNA repair-related genes have also been demonstrated to be of preferential uptake in some bacteria.55
Eukaryotic evolution has been significantly shaped by the origin of mitochondria and plastids, which routed numerous bacterial genes to the nucleus. Although such EGT events are often considered to be a dominant force in eukaryotic genome evolution, the sources of transferred genes are intrinsically constrained by the gene pool of mitochondria and plastids. With organellar genomes becoming increasingly reduced, the process of EGT will eventually approach a dead end. The lack of such constraint for HGT, on the other hand, may potentially introduce genes of numerous distantly related organisms than to those from close relatives, it will be considered a candidate of acquired genes. Genes that are only detected in the query and potentially homologues was also investigated. Because of the bacterial nature of mitochondria and plastids, we also investigated if other eukaryotic homologues were mitochondrial or plastid precursors, which often suggest a bacterial origin (see Supplementary Information). Additionally, each alignment was carefully inspected for potential contamination in the process of genome sequencing, any candidate of acquired genes located on a short scaffold was removed from further consideration. Detailed phylogenetic analyses, including sequence re-sampling from our internal customized sequence database, were performed for each of the remaining candidates. Taxonomic distribution of sequence homologues was also investigated. Because of the bacterial nature of mitochondria and plastids, we also investigated if other eukaryotic homologues were mitochondrial or plastid precursors, which often suggest a bacterial origin (see Supplementary Information). Additionally, each alignment was carefully inspected for rare genomic characters that might indicate a close affinity between the candidate gene and homologues from the putative donor. A candidate gene was determined to be horizontally acquired based on (1) gene tree topology that shows a green plant/donor clade with bootstrap support of over 80% from maximum likelihood.

Determining the origin for candidates of acquired genes. For each candidate of acquired genes identified by AlienG, we first checked the scaffold on which the gene was located. Because of the potential contamination in the process of genome sequencing, any candidate of acquired genes located on a short scaffold was removed from further consideration. Detailed phylogenetic analyses, including sequence re-sampling from our internal customized sequence database, were performed for each of the remaining candidates. Taxonomic distribution of sequence homologues was also investigated. Because of the bacterial nature of mitochondria and plastids, we also investigated if other eukaryotic homologues were mitochondrial or plastid precursors, which often suggest a bacterial origin (see Supplementary Information). Additionally, each alignment was carefully inspected for rare genomic characters that might indicate a close affinity between the candidate gene and homologues from the putative donor. A candidate gene was determined to be horizontally acquired based on (1) gene tree topology that shows a green plant/donor clade with bootstrap support of over 80% from maximum likelihood.
or distance analyses or both, (2) taxonomic distribution of homologues only in the putative donor group (bacteria, archaea, viruses or fungi) and (3) unique domain structures, indels or amino-acid residues shared with homologues from the putative donor group.

**Phylogenetic analyses.** Multiple protein sequence alignments were performed using MUSCLE and clustalX, followed by manual refinement. Gaps and ambiguously aligned sites were removed manually (alignments are available from the authors on request). Sequences that caused aberrant alignments and whose real identity could not be confirmed were also removed from alignments. Phylogenetic analyses were performed with a maximum likelihood method using PhyML 3.0 (ref. 56) and a distance method using neighbour of PHYLIPEW v.3.68 (ref. 57) in EMBOSS package. ModelGenerator was used to select the available model of protein substitution and rate heterogeneity that best fit each data set. Bootstrap support values were estimated using 100 pseudo-replicates. Maximum likelihood distances for distance analyses were calculated using TREE-PUZZLE v.5.2 (ref. 59) and PUZZLEBOOT v.1.03 (A. Roger and M. Holder, www.tree-puzzle.de). The models used in maximum likelihood and distance analyses are the same in most cases. If the best model selected by ModelGenerator was not implemented in TREE-PUZZLE, the second best model was used. All other parameters in the analyses used default settings.

**Functional annotation.** Whenever possible, functional annotation of the acquired genes followed the information provided by The Arabidopsis Information Resources (TAIR) (www.arabidopsis.org) and published experimental data. Homologous gene loci in Arabidopsis were also obtained from TAIR.

**References**

1. Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* **19**, 2226–2238 (2002).
2. Andersson, J. O. Gene transfer and diversification of microbial eukaryotes. *Annu. Rev. Microbiol.* **63**, 177–199 (2009).
3. Belbahri, L., Calmin, G., Mauch, F. & Andersson, J. O. Evolution of the cutinase gene family: evidence for lateral gene transfer of a candidate Phytophthora virulence factor. *Gene* **408**, 1–8 (2008).
4. Sun, G., Yang, Z., Kosch, T., Summers, K. & Huang, J. Evidence for acquisition of virulence effectors in pathogenic chytrids. *J. Biol. Sci.** **41**, 195 (2011).
5. Danchin, E. G. et al. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc. Natl Acad. Sci. USA* **107**, 17651–17656 (2010).
6. Ricard, G. et al. Horizontal gene transfer from Bacteria to rumen Ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. *BMC Genomics* **7**, 22 (2006).
7. Andersson, J. O., Sjogren, A. M., Davis, L. A., Emlery, T. M. & Roger, A. J. Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. *Curr. Biol.* **13**, 94–104 (2003).
8. Ochman, H., Lawrence, J. G. & Grossman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).
9. Andersson, J. O. Lateral gene transfer in eukaryotes. *Cell Mol. Life Sci.* **62**, 1182–1197 (2005).
10. Keeling, P. J. & Palmer, J. D. Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* **9**, 605–618 (2008).
11. Bock, R. The give-and-take of DNA: horizontal gene transfer in plants. *Trends Plant Sci.* **15**, 11–22 (2010).
12. Klasson, L., Kambris, Z., Cook, P. E., Walker, T. & Sinkins, S. P. Horizontal gene transfer between Wolbachia and the mosquito *Aedes aegypti*. *BMC Genomics* **10**, 33 (2009).
13. Yoshida, S., Maruyama, S., Norzaki, H. & Shirasu, K. Horizontal gene transfer by the parasitic plant *Striga hermonthica*. *Science* **328**, 1128 (2010).
14. Berghorothsson, U., Adams, K. L., Thomason, B. & Palmer, J. D. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* **424**, 197–201 (2003).
15. Richardson, A. O. & Palmer, J. D. Horizontal gene transfer in plants. *J. Exp. Bot.* **58**, 1–9 (2007).
16. Moran, N. A. & Jarvik, T. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* **326**, 624–627 (2010).
17. Christin, P. A. et al. Adaptive evolution of C4 photosynthesis through recurrent lateral gene transfer. *Curr. Biol.* **22**, 445–449 (2012).
18. Nl, T. et al. Ancient gene transfer from algae to animalic mechanisms and evolutionary significance. *BMC Evol. Biol.* **12**, 83 (2012).
19. Huang, J. & Gogarten, J. P. Ancient horizontal gene transfer can benefit phylogenetic reconstruction. *Trends Genet.* **22**, 361–366 (2006).
20. Sanderson, M. J., Thorne, J. L., Wikstrom, N. & Bremer, K. Molecular evidence on plant divergence times. *Am. J. Bot.* **91**, 1656–1665 (2004).
21. Graham, L. E., Cook, M. E. & Busse, J. S. The origin of plants: body plan changes contributing to a major evolutionary radiation. *Proc. Natl Acad. Sci. USA* **97**, 4535–4540 (2000).
22. Syvanen, M. Horizontal gene transfer: evidence and possible consequences. *Annu. Rev. Genet.* **28**, 237–261 (1994).
23. Copley, S. D. & Dhillion, J. K. Lateral gene transfer and parallel evolution in the history of glutathione biosynthesis genes. *Gene* **197**, 237–261 (1994).
24. Emiliani, G., Fumii, M., Fani, R. & Gribaldo, S. A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. *Biol. Direct* **4**, 7 (2009).
25. Richards, T. A. et al. Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *Plant Cell* **21**, 1087–1091 (2009).
26. Pisani, D., Cotton, J. A. & McNerney, J. O. Superreeves disentangle the chimerical origin of eukaryotic genomes. *Mol. Biol. Evol.* **24**, 1752–1760 (2007).
27. Martin, W. et al. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl Acad. Sci. USA* **99**, 12246–12251 (2002).
28. Hackett, J. D. et al. Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. *Corr. Biol.* 14, 213–218 (2004).

29. Eiser, C. et al. A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* 21, 1643–1660 (2004).

30. Wang, B. et al. Presence of three mycorrhizal genes in the common ancestor of land plants suggests a key role of mycorrhizas in the colonization of land by plants. *New Phytol.* 186, 514–525 (2010).

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

32. Price, D. C. et al. Cyanophora paradoxa genome elucidates origin of photosynthesis in algae and plants. *Science* 335, 843–847 (2012).

33. Huang, J. & Gogarten, J. P. Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biol.* 8, R99 (2007).

34. Tian, J. et al. AlienG: an effective computational tool for phylogenetic identification of horizontally transferred genes The third International Conference on Bioinformatics and Computational Biology (BICoB): New Orleans, Louisiana (2011).

35. Philippe, H. & Douady, C. J. Horizontal gene transfer and phylogenetics. *Curr. Opin. Microbiol.* 6, 498–505 (2003).

36. Andersson, J. O. Evolution of patchily distributed proteins shared between eukaryotes and prokaryotes: Dictyostelium as a case study. *J. Mol. Microbiol. Biotechnol.* 20, 83–95 (2011).

37. Sun, G., Yang, Z., Ishwar, A. & Huang, J. Algal genes in the closest relatives of animals. *Mol. Biol. Evol.* 27, 2879–2890 (2010).

38. Schaller, A., Stintzi, A. & Graff, L. Subtilases - versatile tools for protein turnover, plant development, and interactions with the environment. *Physiol. Plant.* 145, 52–66 (2012).

39. Cheng, Y., Dai, X. & Zhao, Y. Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in Arabidopsis. *Genes Dev.* 20, 1790–1799 (2006).

40. Wiszniewski, A. A., Zhou, W. & Smith, S. M. & Bussell, J. D. Identification of two arginase-negative mutants of Arabidopsis exhibit increased nitric oxide signaling in root development. *Plant Physiol.* 147, 1936–1946 (2008).

41. Noctor, G. et al. Glutathione in plants: an integrated overview. *Plant Cell Environ.* 35, 454–484 (2012).

42. Werner, A. K., Romeis, T. & Witte, C. P. Ureide catabolism in Arabidopsis thaliana and Escherichia coli. *Nat. Chem. Biol.* 6, 19–21 (2010).

43. Goremykin, V. V., Salamini, F., Velasco, R. & Viola, R. Mitochondrial DNA of Arabidopsis thaliana is expressed at sites of lateral root emergence. *DNA Res.* 6, 13–19 (1999).

44. Koulintchenko, M., Konstantinov, Y. & Dietrich, A. Plant mitochondria actively import DNA via the permeability transition pore complex. *EMBO J.* 22, 1245–1254 (2003).

45. Zhu, B. et al. Horizontal gene transfer in silkworm, Bombyx mori. *BMC Genomics* 12, 248 (2011).

46. Lowry, B., Lee, D. & Hébant, C. The Origin of Land Plants: A New Look at an Old Problem 29 (1980).

47. Cove, D. The moss Physcomitrella patens. *Annu Rev Genet* 39, 339–358 (2005).

48. Brundrett, M. C. Coevolution of roots and mycorrhizas of land plants. *New Phytol.* 154, 275–304 (2002).

49. Neuteboom, L. W., Veth-Tello, L. M., Clijdesdale, O. R., Hooykaas, P. J. & van der Zaal, B. J. A novel subtilisin-like protease gene from Arabidopsis thaliana is expressed at sites of lateral root emergence. *DNA Res.* 6, 13–19 (1999).

50. Tanaka, H. et al. A subtilisin-like serine protease is required for epidermal surface formation in Arabidopsis embryos and juvenile plants. *Development* 128, 4681–4689 (2001).

51. Zhao, C., Johnson, B. J., Kostitsup, B. & Beers, E. P. Exploiting secondary growth in Arabidopsis. Construction of xylem and bark cDNA libraries and cloning of three xylem endopeptidases. *Plant Physiol.* 123, 1185–1196 (2000).

52. Alcazar, R. et al. Polyamines: molecules with regulatory functions in plant abiotic stress tolerance. *Planta* 231, 1237–1249 (2010).

53. Philippe, H. & Douady, C. J. Horizontal gene transfer and phylogenetics. *Curr. Opin Microbiol.* 14, 583–595 (2011).

54. Price, D. C. et al. Cyanophora paradoxa genome elucidates origin of photosynthesis in algae and plants. *Science* 335, 843–847 (2012).

55. Wiszniewski, A. A., Zhou, W. & Smith, S. M. & Bussell, J. D. Identification of two Arabidopsis genes encoding a peroxisomal oxidoreductase-like protein and an acyl-CoA synthetase-like protein that are required for responses to pro-auxins. *Plant Mol. Biol.* 69, 503–515 (2009).

56. Flores, T. et al. Arginase-negative mutants of Arabidopsis exhibit increased nitric oxide signaling in root development. *Plant Physiol.* 147, 1936–1946 (2008).

57. Noctor, G. et al. Glutathione in plants: an integrated overview. *Plant Cell Environ.* 35, 454–484 (2012).

58. Werner, A. K., Romeis, T. & Witte, C. P. Ureide catabolism in Arabidopsis thaliana and Escherichia coli. *Nat. Chem. Biol.* 6, 19–21 (2010).

59. Goremykin, V. V., Salamini, F., Velasco, R. & Viola, R. Mitochondrial DNA of Vitis vinifera and the issue of rampant horizontal gene transfer. *Mol. Biol. Evol.* 26, 99–110 (2009).

60. Koulintchenko, M., Konstantinov, Y. & Dietrich, A. Plant mitochondria actively import DNA via the permeability transition pore complex. *EMBO J.* 22, 1245–1254 (2003).

61. Zhu, B. et al. Horizontal gene transfer in silkworm, Bombyx mori. *BMC Genomics* 12, 248 (2011).

62. Lowry, B., Lee, D. & Hébant, C. The Origin of Land Plants: A New Look at an Old Problem 29 (1980).

63. Cove, D. The moss Physcomitrella patens. *Annu Rev Genet* 39, 339–358 (2005).

64. Brundrett, M. C. Coevolution of roots and mycorrhizas of land plants. *New Phytol.* 154, 275–304 (2002).