New Phytologist

Research

# Post-transfer adaptation of HGT-acquired genes and contribution to guanine metabolic diversification in land plants

**Jun-Jie Wu[1]*, Qian-Wen Deng[2,3]*, Yi-Yang Qiu[1]*, Chao Liu[2,4], Chen-Feng Lin[1], Ya-Lu Ru[1], Yue Sun[1], Jun Lai[1], Lu-Xian Liu[5], Xing-Xing Shen[2,4], Ronghui Pan[2,3] and Yun-Peng Zhao[1]**

[1]Systematic & Evolutionary Botany and Biodiversity Group, MOE Key Laboratory of Biosystem Homeostasis and Protection, College of Life Sciences, Zhejiang University, Hangzhou, 310058, China; [2]College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, 310058, China; [3]ZJU-Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou, 310027, China; [4]Center for Evolutionary & Organismal Biology, Zhejiang University, Hangzhou, 310058, China; [5]Laboratory of Plant Germplasm and Genetic Engineering, School of Life Sciences, Henan University, Kaifeng, 475001, China

Authors for correspondence:
*Ronghui Pan*
*Email: panr@zju.edu.cn*

*Yun-Peng Zhao*
*Email*: *ypzhao@zju.edu.cn*

Received: *8 May 2024*
Accepted: *24 July 2024*

*New Phytologist* (2024)
**doi**: 10.1111/nph.20040

**Key words:** guanine metabolism diversification, horizontal gene transfer, intron gains, land plant evolution, post-transfer adaptation.

## Summary

- Horizontal gene transfer (HGT) is a major driving force in the evolution of prokaryotic and eukaryotic genomes. Despite recent advances in distribution and ecological importance, the extensive pattern, especially in seed plants, and post-transfer adaptation of HGT-acquired genes in land plants remain elusive.
- We systematically identified 1150 foreign genes in 522 land plant genomes that were likely acquired via at least 322 distinct transfers from nonplant donors and confirmed that recent HGT events were unevenly distributed between seedless and seed plants.
- HGT-acquired genes evolved to be more similar to native genes in terms of average intron length due to intron gains, and HGT-acquired genes containing introns exhibited higher expression levels than those lacking introns, suggesting that intron gains may be involved in the post-transfer adaptation of HGT in land plants.
- Functional validation of bacteria-derived gene *GuaD* in mosses and gymnosperms revealed that the invasion of foreign genes introduced a novel bypass of guanine degradation and resulted in the loss of native pathway genes in some gymnosperms, eventually shaping three major types of guanine metabolism in land plants. We conclude that HGT has played a critical role in land plant evolution.

## Introduction

Land plants (embryophytes) originated in the Middle Paleozoic (*c.* 470 million years ago), or possibly even earlier, in the Middle Cambrian to Early Ordovician (Morris *et al.*, 2018; Bowman, 2022; Harris *et al.*, 2022). Major evolutionary transitions, including the colonization of land and the origin of vascular tissues, seeds, and flowers (Donoghue *et al.*, 2021), have resulted in the diversification of land plants into > 370 000 extant species, with profound and lasting impacts on Earth's ecosystems and climate change (Berry *et al.*, 2010; Lenton *et al.*, 2016). Each transition was accompanied by gene innovation and loss in land plant genomes (Bowles *et al.*, 2020). New genes may have been transferred from other species into land plant genomes via horizontal gene transfer (HGT). These transferred genes could theoretically be acquired from any foreign source, including other plants and nonplant species. Foreign genes acquired from distantly related species (e.g. bacteria and fungi) are often considered more important for novel functions in land plants (Huang & Yue, 2013; Soucy *et al.*, 2015). Previous studies have reported HGT from nonplant species to early diverged land plants (e.g. *Anthoceros angustus*, *Physcomitrella patens*, and *Marchantia polymorpha*; Yue *et al.*, 2012; Bowman *et al.*, 2017; Li *et al.*, 2020) and a few land plants of interest (e.g. *Ceratopteris richardii*, *Cycas panzhihuaensis*, *Torreya grandis*, and *Thinopyrum elongatum*; H. Wang *et al.*, 2020; Liu *et al.*, 2022; Marchant *et al.*, 2022; Lou *et al.*, 2023). A recent comprehensive analysis of 40 representative streptophytes (8 charophytes, 10 seedless plants, and 22 seed plants) revealed two major episodes of historical HGT events in land plant evolution, corresponding to the early evolution of streptophytes (Episode 1) and the origin of land plants (Episode 2), and contrasting scales of relatively recent HGT events between seedless and seed plants (Ma *et al.*, 2022). An accelerating number of available genomes provides an opportunity to uncover extensive patterns of HGT in land plants, especially in seed plants, a lineage containing > 90% species of land plants (Nic Lughadha *et al.*, 2016; Shi *et al.*, 2023). In addition, a limited number of studies have addressed post-transfer adaptation of HGT-acquired genes in land plants (Wang *et al.*, 2023). The foreign genes appear to gradually adapt to the recipient genomes

*These authors contributed equally to this work.

over time (Husnik & McCutcheon, 2018; Fan *et al.*, 2020). Nevertheless, post-transfer adaptation of HGT-acquired genes in land plants remains unclear.

In this study, we systematically identified and characterized horizontally acquired genes using a robust and conservative phylogeny-based approach in 522 high-quality land plant genomes (19 seedless plants and 503 seed plants) representing seven major clades of land plants, and conducted a comprehensive analysis of HGT-acquired genes to test the extensive pattern of HGT across major land plant clades, to explore post-transfer adaptation of HGT-acquired genes in land plants, and to examine the impact of foreign genes on native genes in land plants.

## Materials and Methods

### Data sources

To collect the set of land plant genomes as of December 2022, we used 'plant genome' as a search term in Web of Science (https://www.webofscience.com/) to screen for relevant papers and to obtain the basic information on species name, sequencing platform, accession number, assembly level, and download link. For species with multiple genomes sequenced, we only selected the genome with publicly available annotation, the highest assembly level, and the highest genome completeness (Table S1).

### Identification of HGTs in land plants

To detect horizontally acquired genes from nonplant species in land plants, we followed a robust and conservative phylogeny-based method (Keeling & Palmer, 2008) used in insects (Li *et al.*, 2022; Supporting Information Fig. S1). Briefly, to avoid potential spurious results arising from small genomic fragments of contaminant organisms in the genome assemblies, we only retained those genes that located in genomic contigs or scaffolds that were ≥ 100 kb. For protein sequence of each gene, we evaluated whether it was horizontally acquired by a two-step workflow. In Step 1, we first performed a BLASTP search in DIAMOND v.2.0.15.153 (Buchfink *et al.*, 2021) with an *e*-value cutoff of 1*e*–10 against a custom database (Refseq+) consisting of the reference protein sequences (Refseq) (last accessed on 10 June 2022) and 522 land plant protein sequences. HGTFINDER (v.1, https://github.com/xingxingshen/HGTfinder) (Shen *et al.*, 2018) was then used to: (1) assign taxonomic information to each BLASTP hit from NCBI Taxonomy database and (2) classify the BLASTP hits into three different lineages (RECIPIENT: land plants; GROUP: streptophytes excluding land plants; OUTGROUP: all species excluding streptophytes) based on their taxonomic information to obtain three values: bbhO (BLAST bitscore of the best hit in the OUTGROUP lineage), bbhG (bitscore of the best hit in GROUP lineage but not in the RECIPIENT lineage), and maxB (bitscore of the query to itself). Using these values, we calculated: (1) the Alien Index: AI = (*bbhO*/*maxB*) − (*bbhG*/*maxB*) and (2) the percentage of species from the OUTGROUP lineage (outg_pct) in the list of the top 1000 hits that have different taxonomic species names. Of the 22 268 484 genes analyzed, 84 615 genes passed the cutoffs of AI value > 0 and outg_pct > 80%. We also compared the number of candidate genes retained when outg_pct was set to 60%, 70%, and 80% to further justify the cutoff setting and found that there was only a slight difference (Fig. S2). In Step 2, we extracted the 1000 most similar homologs from the Refseq+ database by SEQKIT v.2.1.0 (Shen *et al.*, 2016). Multiple sequence alignments were performed by MAFFT v.7.310 (Katoh *et al.*, 2002) with '--auto' parameter. We used TRIMAL (v.1.4) (Capella-Gutiérrez *et al.*, 2009) to trim poor alignment with the '-automated1' parameter. IQ-TREE v.2.1.4 (Minh *et al.*, 2020) was used to infer maximum likelihood (ML) trees with '-B 1000' parameters. Each ML tree was rooted at the midpoint using the APE (Paradis *et al.*, 2004) and PHANGORN (Schliep, 2011) R packages and was visualized using the command version of ITOL v.4 (Letunic & Bork, 2019). To further identify reliable HGT-acquired genes, we conducted three additional analyses. First, for candidate-acquired genes with > 90% identity to the putative donors, we excluded those candidate-acquired genes for which no orthologs were found in closely related species. Second, we excluded candidate-acquired genes with homology to plastid sequences. Third, by manual inspection, we further excluded those ML trees for which nonplant species were not monophyletic due to donor sequence contamination. After inspection, we identified 1150 putative HGT-acquired genes in 522 land plant genomes.

Since HGT-acquired genes belonging to the same gene family may be acquired from diverse donors, we considered both phylogenetic relationships and donor sources and identified 322 gene families for 1150 HGT-acquired genes. These 322 gene families, which were assumed to be at least 322 distinct HGT events, were then divided into historical HGT events (with homologs in other land plant clades or Zygnematophyceae) and recent HGT events (with homologs only in the given clade) based on previous study and the land plant clades in the gene families. Foreign genes acquired by earlier ancestors of land plants (e.g. the last common ancestors of Plantae or green plants) were not included in this study. Considering that it is difficult to be absolutely certain about the real donors, the transferred genes in putative donor species were inferred as the best hits among the lineage of putative donors, using a monophyletic most similar homolog approach (Fan *et al.*, 2020; Li *et al.*, 2022).

### Validation of HGT-acquired genes

To ensure the reliability of these 1150 HGT-acquired genes, we first used CONTERMINATOR v.1.c74b5 (Steinegger & Salzberg, 2020) to detect potential contamination and found none for these HGT-acquired genes. Second, we examined the distribution of sequence lengths of the 479 genomic contigs containing HGT-acquired genes alongside the distribution of sequence lengths of the 49 696 genomic contigs that not containing HGT-acquired genes (Fig. S3a). We found that contigs containing the HGT-acquired genes were typically longer than contigs without them. Third, we examined the distribution of the proportions of HGT-acquired genes that reside in the 479

contigs and found that none of the 479 contigs contained HGT-acquired genes at a frequency > 9% (Fig. S3b). Fourth, we examined the protein sequence similarity between HGT-acquired genes in land plants and their closest homologs in nonplant donors for HGT-acquired genes and found that the similarity values ranged from 22% to 92%, with an average value of 51% (Fig. S3c). Fifth, we checked the genomic location of 1150 HGT-acquired genes and calculated the ratio of their gene start location to scaffold length (Fig. S3d). We found that HGT-acquired genes were roughly evenly distributed across the scaffolds. We then checked the donor source of HGT-acquired genes with intron and found that HGT-acquired genes with intron were mainly of bacterial/viral origin (64.0%; Fig. S3e). Finally, we examined 58 randomly selected HGT-acquired genes in the six land plant genomes representing five of the seven clades, using polymerase chain reaction and Sanger sequencing experiments. For each of the 58 HGT-acquired genes, two separate polymerase chain reaction reactions were used to amplify the upstream and downstream regions that flanked the foreign gene (Table S2). The size of each polymerase chain reaction target was *c.* 1500 bp (Fig. S4). We considered the HGT-acquired gene to be successfully validated only if its upstream and downstream regions were successfully amplified and their Sanger sequencing were nearly identical (identity ≥ 98.0%) to the DNA sequences. Our results showed that the percentage of successful validation ranged from 75.0% to 87.5% across six land plants tested, with an average of 84.5% (Fig. S3f). Given the impact of biological factors such as heterozygous sequences, repetitive sequences, and polyploidy on polymerase chain reaction and Sanger sequencing, this result remained acceptable and similar to a recent study in insects (Li *et al.*, 2022). Taken together, these results indicate that the set of 1150 HGT-acquired genes present in the scaffolds with a length of scaffolds ≥ 100 kb is reliable.

### Phylogenomic matrix construction

For concatenation-based ML tree inference, the protein sequences of 522 land plant genomes and seven outgroup genomes were evaluated respectively in Benchmarking Universal Single-Copy Orthologs (BUSCO v.5.4.7), and we used 1614 BUSCO genes from embryophyta_odb10 (2020-09-10) as a set of reliable markers of diverse lineages (Manni *et al.*, 2021). After screening, 1373 orthologous genes in at least 70% of 529 land plant genomes were concatenated for subsequent analysis.

### RNA extraction and real-time PCR

Total RNAs were extracted from different tissues of *Ginkgo biloba* using a plant RNA isolation kit (Vazyme, Nanjing, China). First-strand cDNA was synthesized from 1.0 μg of total RNA using RevertAid Reverse Transcriptase (Thermo Scientific, Waltham, MA, USA) with oligo (dT) primers. The genomic DNAs were extracted from leaf of *G. biloba* using the modified cetyltrimethylammonium bromide (CTAB). The primers used here are listed in Table S2.

### Origins and feature of introns

To identify the origin of the 1972 introns, we first extracted and masked the sequence of each intron using BEDTOOLS v.2.30.0 (Quinlan & Hall, 2010) and performed a BLASTN search against a custom database consisting of the Nucleotide (nt) database (last accessed on 10 June 2022) and 522 land plant genomes with an *e*-value cutoff of 1*e*–5 and the option '-task blastn-short'. Considering that intron sequences might be originated from different regions of the same genome, we counted the total query length of each subject sequence based on BLASTN hits and calculated the total query coverage (total query length/intron length) for each subject sequence. Only subject sequences with a total query coverage ≥ 50% were considered as the putative origin of the corresponding intron. The features of the introns were characterized using Extensive *de novo* TE Annotator (EDTA v.2.1.3; Ou *et al.*, 2019) with default settings.

### RNA sequencing and transcriptome analysis

Three tissues (leaf, stem, and root) of *G. biloba* were collected and frozen in liquid nitrogen with three biological replicates. Library construction and sequencing were performed on the Illumina HiSeq2000 platform to generate 150 bp pair-end reads. For transcriptome analysis, raw data were filtered by TRIMMOMATIC v.0.39 (Bolger *et al.*, 2014), and clean reads were mapped to cDNA using SALMON v.0.13.0 to directly quantify gene expression levels (TPM) (Patro *et al.*, 2017).

### Plant material and cultivation

All mutant/transgenic lines are in the *Arabidopsis thaliana* ecotype Columbia (Col-0) background. The *Arabidopsis* T-DNA mutants *gsda* (GK432D08) (Kleinboelting *et al.*, 2011) was obtained from the GABI-Kat collection. The coding regions of the *GbGuaD* were amplified using polymerase chain reaction and respectively cloned into the expression vector pCAMBIA1300 promoting by a 35S promoter. Complementation lines were obtained by transforming *gsda* with the *Agrobacterium tumefaciens* strain GV3101 carrying the recombinant plasmids via the floral dip method (Clough & Bent, 1998). *Arabidopsis* plants were grown on soil under long-day conditions (22°C, 16 h : 8 h light : dark) at 60% relative humidity. For dark stress treatment, seedlings of Col-0, *gsda* mutants, and complementation lines were covered for 6 d after 24 d of growth. Full rosettes of the respective genotypes were harvested after 6 d dark treatment for metabolite analysis (three pots per genotype as biological replicates).

### Protein purification and assay

The full-length coding sequence of *GbGuaD* was amplified by polymerase chain reaction and cloned into the EcoRI and XbaI sites of the pMAL-p2x expression vector containing the N-terminal MBP tag. For functional expression, *Escherichia coli* BL21 (DE3) cells were transformed with recombinant plasmids

and the corresponding pMAL-p2x vectors without insert as controls and grown in Luria-Bertani (LB) medium with 100 μg $ml^{-1}$ ampicillin at 37°C. A 1 : 100 dilution of the culture was chosen and expanded in 100 ml LB medium until the optical density at 600 nm reached *c.* 0.5. Recombinant protein expression was induced with 0.5 mM isopropyl 1-thio-β-d-galactopyranoside at 200 rpm, 15°C for 16 h. After incubation, the cells were harvested by centrifugation (at 20 000 ***g***, 4°C for 30 min). Lysis and protein purification were performed according to the manufacturer's protocol of Amylose Resin (New England Biolabs, NEB). Protein concentration was measured by the Bradford dye binding assay. The enzyme assay was performed in a 5 ml reaction mixture at 37°C for 40 min containing 10 mM $MgCl_2$, 0.1 mM guanine, and 40 μg GbGuaD protein. Controls were run without GbGuaD protein. After centrifugation, the supernatant was filtered through the 0.22 μm aqueous phase filtration membrane for metabolite analysis.

### Metabolite analysis

For seedlings, the extraction method was adapted from Heinemann *et al.* (2021). After freeze-drying and weighing, the plant material was ground with eight 2 mm steel beads at a frequency of 30 Hz for 5 min. One milliliter of extraction buffer (10 mM ammonium acetate, pH = 7.5) prewarmed to 60°C was used and samples were immediately incubated at 95°C for 10 min. After 5 min cooling on ice and 15 min centrifugation at 20 000 ***g*** at 4°C for the removal of beads and cell debris, 80% of the supernatant was filtered through the 0.22 μm aqueous phase filtration membrane. Following the filtration, 400 μl of the supernatant was used for the subsequent analysis. Guanine and guanosine were quantified by external standards (ESTDs, calibration curve) using an Agilent HPLC 1200 system with an Eclipse XDB-C18 150 × 4.6 mm column (Agilent Technologies, Santa Clara, CA, USA). For ESTDs, standard solutions of guanine (0, 0.1, 1, 5, 10, and 20 μg $ml^{-1}$) and guanosine (0, 0.2, 2, 10, and 20, 40 μg $ml^{-1}$) were added into the matrix. Ammonium acetate (10 mM, pH = 7.5) and 100% methanol served as mobile phase A and mobile phase B. The gradient on the column was as follows: 0 min, 5% B; 1.5 min, 5% B; 3.5 min, 15% B; 6 min, 100% B; 7 min, 100% B; 7.1 min, 5% B; and 13 min, 5% B. The injection volume was 20 μl and the flow rate was 0.8 ml $min^{-1}$. Signal-to-noise ratios < 10 were considered as not detected. For protein assay, mobile phase A was 0.2% acetic acid and mobile phase B was acetonitrile. The gradient was as follows: 0 min, 3% B; 8 min, 4% B; 18 min, 60% B; and 20 min, 3% B. Leaf Chl concentrations were measured by averaging the SPAD values of four identical leaves during dark stress for each sample in the SPAD-502 meter (Three samples per genotype as biological replicates).

### Subcellular localization

The subcellular localization of GbGuaD investigated by confocal microscopy in the leaves of *Nicotiana tabacum* transiently expressing the protein fused to N-terminal and C-terminal (Fig. S5a) mVenus-tag, was carried out as described previously (Deng *et al.*, 2022). A FluoView FV3000 confocal laser scanning microscope (Olympus) was used for image acquisition, with mVenus excited by 514 nm lasers and detected at 530–630 nm.

### Statistical analysis

Data are presented as mean ± SD in figures. The statistical analyses including testing the normality of data distribution were performed in R v.4.2.2 and OriginPro v.9.8.0.200 software. The Pearson correlation coefficient was used to summarize the strength of the linear relationship between two data sets. Statistical significance was determined by *t*-test with two-sided, and multiple comparisons was performed (ns, $P > 0.05$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). Further details of number of samples and replicates are provided above.

## Results

### Distribution of HGT in land plants

To systematically identify putative HGT-acquired genes in land plants, we selected 522 publicly available high-quality genomes representing all seven clades of land plants (Table S1), including hornworts (2), liverworts (7), mosses (2), clubmosses (3), ferns (5), gymnosperms (7), and angiosperms (496). Using a robust and conservative phylogeny-based method that we recently applied to insects (Fig. S1; Li *et al.*, 2022), *c.* 22 million sequences of protein-coding genes present in the scaffolds with a length of ≥ 100 kb from 522 land plant genomes were screened for putative HGT-acquired genes. We identified a total of 1150 genes (Table S3) that were transferred from nonplant sources into 64 land plants, likely via at least 322 distinct transfers (Fig. 1).

In terms of the average number of acquired gene families by the number of included genomes from each clade, hornworts acquired the highest average number of gene families (19.5 gene families per species), followed by mosses (19.3 gene families per species), clubmosses (17.3 gene families per species), liverworts (13 gene families per species), ferns (6.6 gene families per species), gymnosperms (3.3 gene families per species), and angiosperms (0.4 gene families per species) (Fig. S6a). When evaluating recent HGT events, the average number of recent HGT events was 15.7, 14.5, 11.0, 10.0, 5.2, 0.6, and 0.4 in mosses, hornworts, clubmosses, liverworts, ferns, gymnosperms, and angiosperms, respectively (Fig. S6a). Screening of putative donor organisms for the 322 acquired gene families using approaches of both BLAST and phylogeny (Table S4) suggested that bacteria (50.3%) and fungi (41.6%) were the major donor sources, with the bacterial phylum Proteobacteria (18.6%) and the fungal phylum Ascomycota (21.4%) being the most common donor organisms (Fig. 2c).

### Post-transfer adaptation of HGT in land plants

To explore the post-transfer adaptation of HGT-acquired genes in land plant genomes, we examined the gene structure of 1150

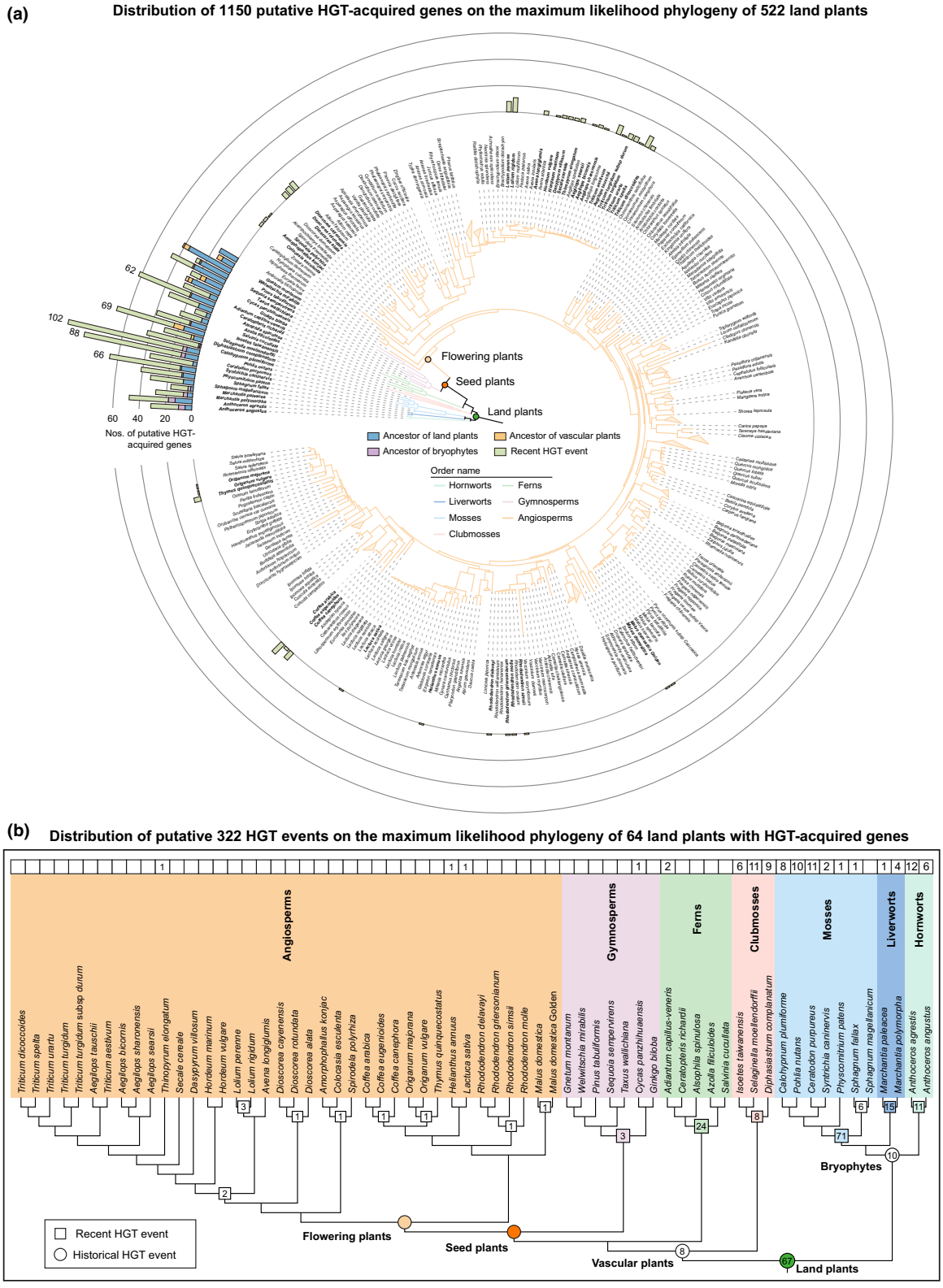


**Fig. 1** Distribution of the 1150 putative horizontal gene transfer (HGT)-acquired genes and 322 acquired gene families, which were assumed to be at least 322 distinct HGT events in land plants. (a) Distribution of the 1150 putative HGT-acquired genes on the maximum likelihood (ML) phylogeny of 522 land plants. (b) Distribution of the 322 acquired gene families on the ML phylogeny of 64 land plants with HGT-acquired genes.

New Phytologist

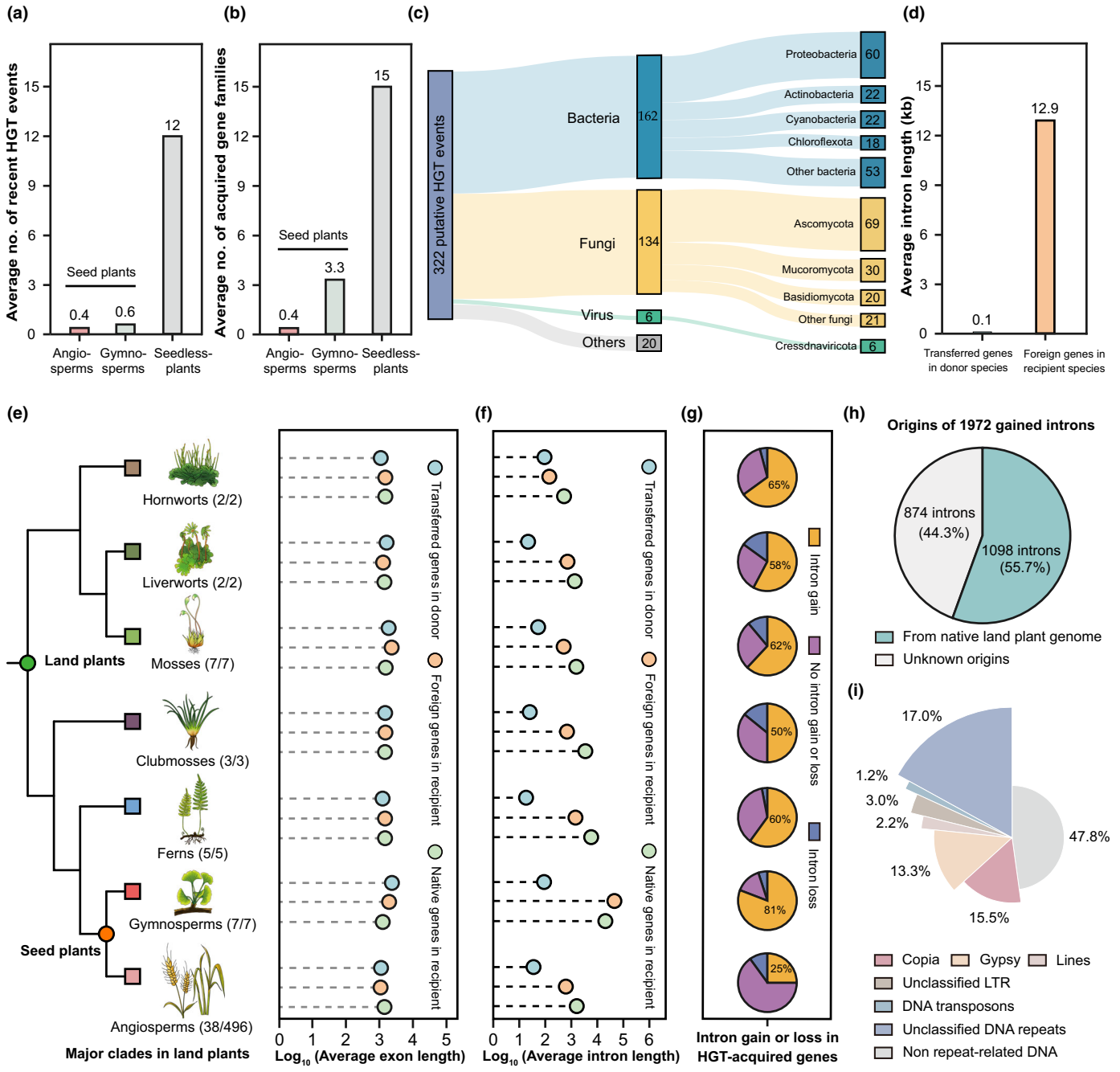


**Fig. 2** Horizontal gene transfer (HGT) was unevenly distributed in land plants and gene structure after transfers. (a) Comparisons of average number of recent HGT events between 19 seedless plants and 45 seed plants. (b) Comparisons of average number of acquired gene families between 19 seedless plants and 45 seed plants. (c) Distribution of 322 putative HGT events (blue, bacteria; yellow, fungi; green, viruses; gray, others). Numbers in the box represent the count of putative donor species; line width indicates the proportion. (d) Comparisons of average intron length of transferred genes in donor species and foreign genes in recipient species. (e) Comparisons of average exon length between transferred genes in donor species (light blue), foreign genes in recipient species (light orange), and native genes in recipient species (light green) across seven major clades of land plants. (f) Comparisons of average intron length between transferred genes in donor species (light blue), foreign genes in recipient species (light orange), and native genes in recipient species (light green) across seven major clades of land plants. (g) After integration into land plant genomes, the proportion of HGT-acquired genes with no intron gain or loss (neither foreign genes in land plants nor transferred genes in putative donors have introns), intron loss (transferred genes in putative donors with introns, whereas foreign genes in land plants without introns), and intron gain (foreign genes in land plants with introns). (h) The origins of 1972 gained introns. (i) The transposable element (TE) compositions of 1098 introns gained from native land plant genomes.

HGT-acquired genes in both putative donor and recipient genomes, as well as native genes in land plant genomes. We found that the average exon length of HGT-acquired genes in land plants was similar to that of the putative recipient and donor species (Fig. 2e), while their average intron length was considerably longer than that of the putative donor species (average: 12.9 kb

vs 0.1 kb) (Fig. 2d). In particular, the average intron length of HGT-acquired genes in gymnosperms increased > 500-fold after transfer from putative donor genomes (average: 82.9 vs 0.16 kb) (Fig. 2f). Comparison of the introns of HGT-acquired genes in putative HGT donor organisms and land plant recipients showed that all 1972 introns present in the 683/1150 (59.4%) HGT-acquired genes were gained after insertion into land plant genomes. Among land plants, gymnosperms had the highest net intron gain rate at 81.0%, while angiosperms had the lowest at 25.0% (Fig. 2g). Furthermore, we found that the 1098/1972 (55.7%) introns had the best hits from their native land plant genomes (Fig. 2h). TE annotation revealed that these 1098 gained introns from native land plant genomes are repeat-rich DNA sequences, including LTR transposons (33.0%), DNA transposons (1.2%), and unclassified repeats (17.0%) (Fig. 2i).

To investigate whether intron gain events were involved in the post-transfer adaptation of HGT-acquired genes in land plant genomes, we conducted three separate analyses. We first compared the identity between the HGT-acquired genes in land plants and the transferred genes in the donor species. We found that HGT-acquired genes with introns had lower identity than those without introns compared to the transferred genes in the donor species (Fig. 3a). This indicated that through intron gain events, HGT-acquired genes would become increasingly divergent from the transferred genes in the donor species. Second, we compared the expression levels between HGT-acquired genes with introns and those without introns, using 60 publicly available transcriptome datasets from 20 land plants (Table S5). Transcriptome analysis showed that 77.3% (573/741) of the HGT-acquired genes had expression levels, compared with 72.3% of the native genes. Notably, of the 20 land plants analyzed, 15 (75%) had on average threefold higher gene expression levels of HGT-acquired genes containing introns compared with those lacking introns (Fig. 3c). This suggested that intron gains can increase the expression levels of HGT-acquired genes in recipient genomes, consistent with our recent finding in insects (Li *et al.*, 2022).

Lastly, to further reveal whether the HGT-acquired genes containing introns shared a similar evolution with the native genes, we calculated the ratio of total intron/exon length of HGT-acquired genes containing introns and native genes and compared it with genome size. We found that the ratio of total intron/exon length in HGT-acquired genes ($r = 0.74$, $P = 3.6 \times 10^{-8}$) (Fig. 3b) had a higher correlation with the genome size than native genes ($r = 0.48$, $P = 0.0018$) (Fig. S6b), illustrating that, as with native genes in land plant genomes (Niu *et al.*, 2022), the genome expansion also occurred in the intergenic and genic regions of HGT-acquired genes.

## Foreign gene *GuaD* acquired from bacteria introduce a novel bypass for guanine degradation

KEGG pathway analysis of the 1150 HGT-acquired genes by eggnog-mapper (Cantalapiedra *et al.*, 2021) showed that most of them were associated with metabolic pathway (Fig. S7a). We found a wide range of metabolic functions brought by HGT-acquired genes in land plants, including but not limited to secondary metabolism, xenobiotics biodegradation, amino acid metabolism, carbohydrate metabolism, lipid metabolism, nucleotide metabolism, and energy metabolism (Fig. S7b). Several functional foreign genes involved in metabolic pathway in land plants have been reported (Yang *et al.*, 2015; Q. Wang *et al.*, 2022; Kfoury *et al.*, 2024), but little attention has been paid to the impact of HGT-acquired genes on native pathway gene. Thus, we explored the function of bacterially derived gene *GBM104943* enriched in nucleotide metabolism, which had a well-established research background of related pathway genes (Witte & Herde, 2019). Previous studies inferred its bacterial origin and potential function based on the incomplete protein sequence (93 aa/434 aa) of *GBM104943* (Ma *et al.*, 2022), but more functional evidence was needed for its role after HGT. Phylogenetic analysis revealed that *GBM104943* and other homologous genes in mosses and gymnosperms were embedded in the bacterial gene cluster (Fig. 4b). The homolog of *GBM104943* in the bacterial donors encodes the enzyme guanine deaminase (GuaD), which catalyzes the hydrolytic deamination of guanine to produce xanthine and ammonia. In contrast to that in bacteria, the current model of purine metabolism in plants suggests that guanine cannot be directly degraded. The degradation of guanine in plants is thought to be first catalyzed by a hypoxanthine-guanine phosphoribosyl transferase (HGPRT) and a hitherto unknown GMP phosphatase (GMPP) to form guanosine. Guanosine is then deaminated to xanthosine by a guanosine deaminase (GSDA) (Fig. 4a; Witte & Herde, 2019). The acquisition of *GuaD* in mosses and gymnosperms may confer a direct guanine-degrading function to these land plants.

To investigate the role of *GBM104943*, which we called *GuaD-Like* in plant physiological and developmental processes, we first examined the gene structure of *GbGuaD*, and compared the expression in different tissues using transcriptome data from *G. biloba*, one of the early diverged groups of gymnosperms. The foreign gene *GbGuaD* in *G. biloba* exhibited the same gene length as the putative bacterial donor, and neither contained introns (Fig. 4c). Quantification of the transcriptome showed that *GbGuaD* was expressed in multiple tissues (Fig. 4d), which was consistent with the reverse transcription polymerase chain reaction results (Fig. S5b). Our subcellular localization analysis showed that *GbGuaD* was present in the cytosol and nucleus (Fig. 4e), which is consistent with other enzymes involved in guanine metabolism, such as GSDA (Dahncke & Witte, 2013) and the nucleoside hydrolase heterocomplex NSH1/NSH2 (Riegler *et al.*, 2011).

To evaluate the guanine deaminase activity of *GbGuaD*, the protein was fused to an MBP tag at its N-terminus, affinity purified, and assayed with guanine as substrate. We found that *GbGuaD* could effectively catalyze the deamination of guanine to xanthine *in vitro* (Fig. 5a). To further explore the role of *GbGuaD* in guanine catabolism *in planta*, we tested whether guanine degradation could be restored by exogenously introducing *GbGuaD* into an *Arabidopsis gsda* mutant, which is sensitive to long-term darkness due to the toxic effects of excessive guanosine accumulation (Schroeder *et al.*, 2018). To this end, we generated two complementation lines

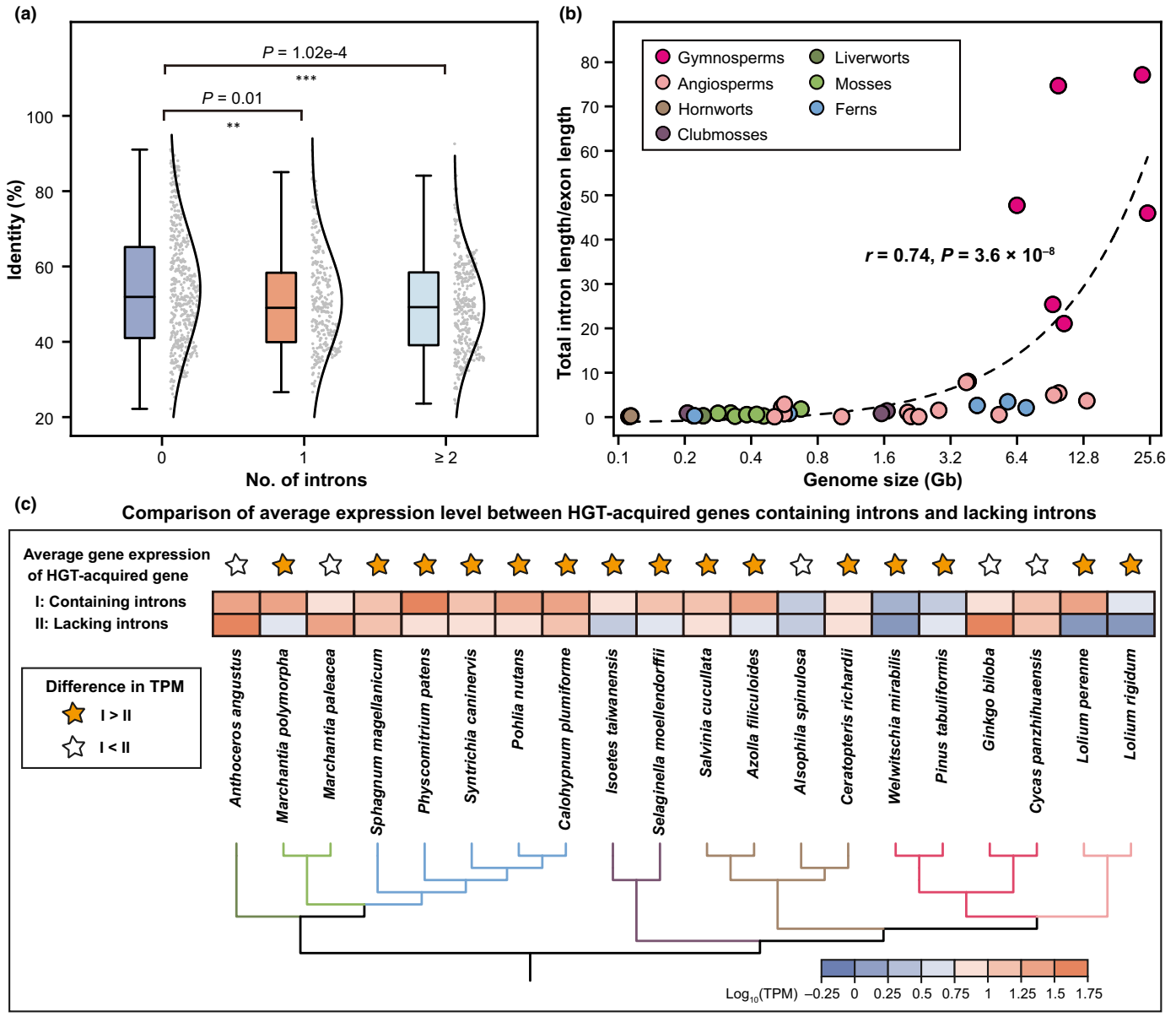


**Fig. 3** Repeat-rich intron gains from native land plant genomes were likely involved in horizontal gene transfer (HGT) adaptation. (a) Comparisons of identity between transferred genes in donor species and foreign genes in recipient species with no intron ($n$ = 467), with one intron ($n$ = 295), and with two or more introns ($n$ = 388). The horizontal lines represent the medians, and the whiskers extend from the hinges to the smallest or largest values within 1.5 times the interquartile range. (b) Correlation between the ratio of total intron/exon length of HGT-acquired genes containing introns and the size of land plant genomes. (c) Comparison of average expression level between HGT-acquired genes containing introns (I: first row) and HGT-acquired genes lacking introns (II: second row) for 20 land plants. Note that we compared HGT-acquired genes containing and lacking introns only within each transcriptome dataset (e.g. only using transcriptome data from the same tissue and the same stage for a given species). Student's $t$-test, asterisks indicate statistical significance, that is, *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

by transforming a *gsda* mutant with the full-length cDNA of *GbGuaD* driven by the CaMV-35s promoter. Compared to wild-type (WT) Col-0, the *gsda* mutant showed a chlorotic phenotype (Fig. 5b) and a significant reduction in Chl content after 6 d of darkness (Fig. S5c). Interestingly, expression of *GbGuaD* in the *gsda* mutant partially restored the WT-like green leaf phenotype (Fig. 5b). Furthermore, metabolite analysis of rosette leaves showed that the dark-induced accumulation of guanosine (Fig. 5c) and guanine (Fig. 5d) in *gsda* was significantly reduced by expression of *GbGuaD*, demonstrating that in the absence of guanosine degradation, *GbGuaD* can directly degrade guanine and reduce the amount of guanine salvaged to guanosine *in vivo*. Collectively, these results indicate that the foreign gene *GbGuaD*, acquired from bacteria, has introduced a novel bypass for guanine metabolism in *G. biloba* and most likely other gymnosperms and mosses by directly degrading guanine to xanthine.

Nucleosides can be converted to nucleotides, a process called salvage. Alternatively, they can be completely degraded and

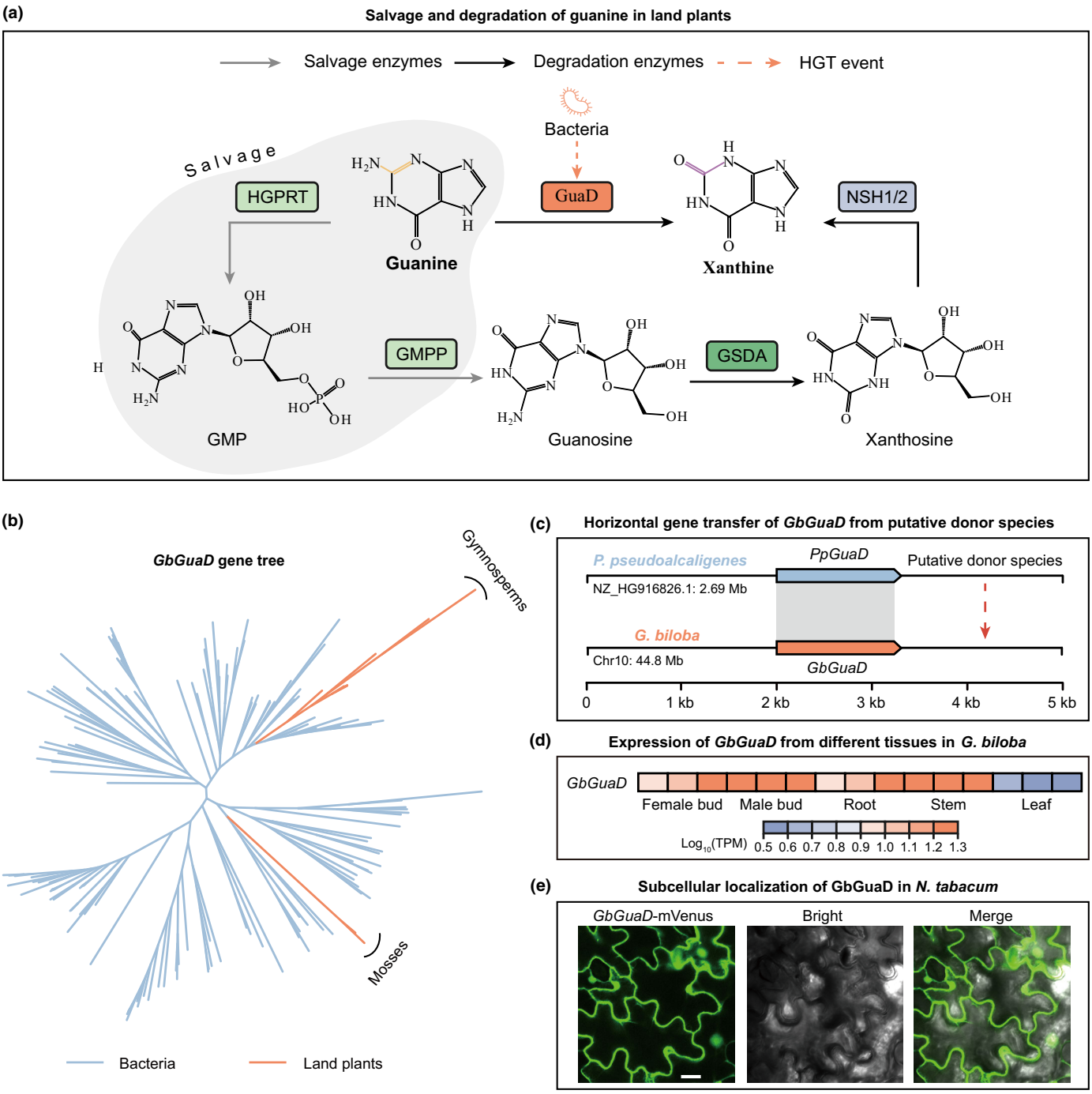


**Fig. 4** Origin and expression pattern of foreign gene *GuaD*. (a) Enzymes involved in the salvage and degradation of guanine in land plants. GMPP, unknown GMP phosphatase; GSDA, guanosine deaminase; GuaD, guanine deaminase enzyme; HGPRT, hypoxanthine-guanine phosphoribosyl transferase; NSH1/2, nucleoside hydrolase heterocomplex. (b) A simplified gene family phylogenetic tree of the gene *GbGuaD*. Blue branches indicate bacteria, while orange branches indicate mosses and gymnosperms. (c) Horizontal gene transfer of *GbGuaD*. The foreign gene *GbGuaD* in the *Ginkgo biloba* genome has the same gene length as the putative bacterial donor and neither contains introns. (d) Expression level of *GbGuaD* in different tissues using transcriptome data from *G. biloba*. (e) Confocal images of lower leaf epidermis cells of *Nicotiana tabacum* transiently expressing N-terminally GbGuaD-mVenus fusion protein. Bar, 20 μm.

utilized as nutrient sources in plants. For guanine, the reaction of salvage should precede degradation (Dahncke & Witte, 2013; Witte & Herde, 2019). An intriguing question is whether the novel bypass introduced by the HGT-acquired gene, which directly degrades guanine to xanthine, has an effect on the original salvage and degradation pathway genes in guanine metabolism. Therefore, we examined the distribution of the key guanine salvage enzyme HGPRT and the direct guanine degradation

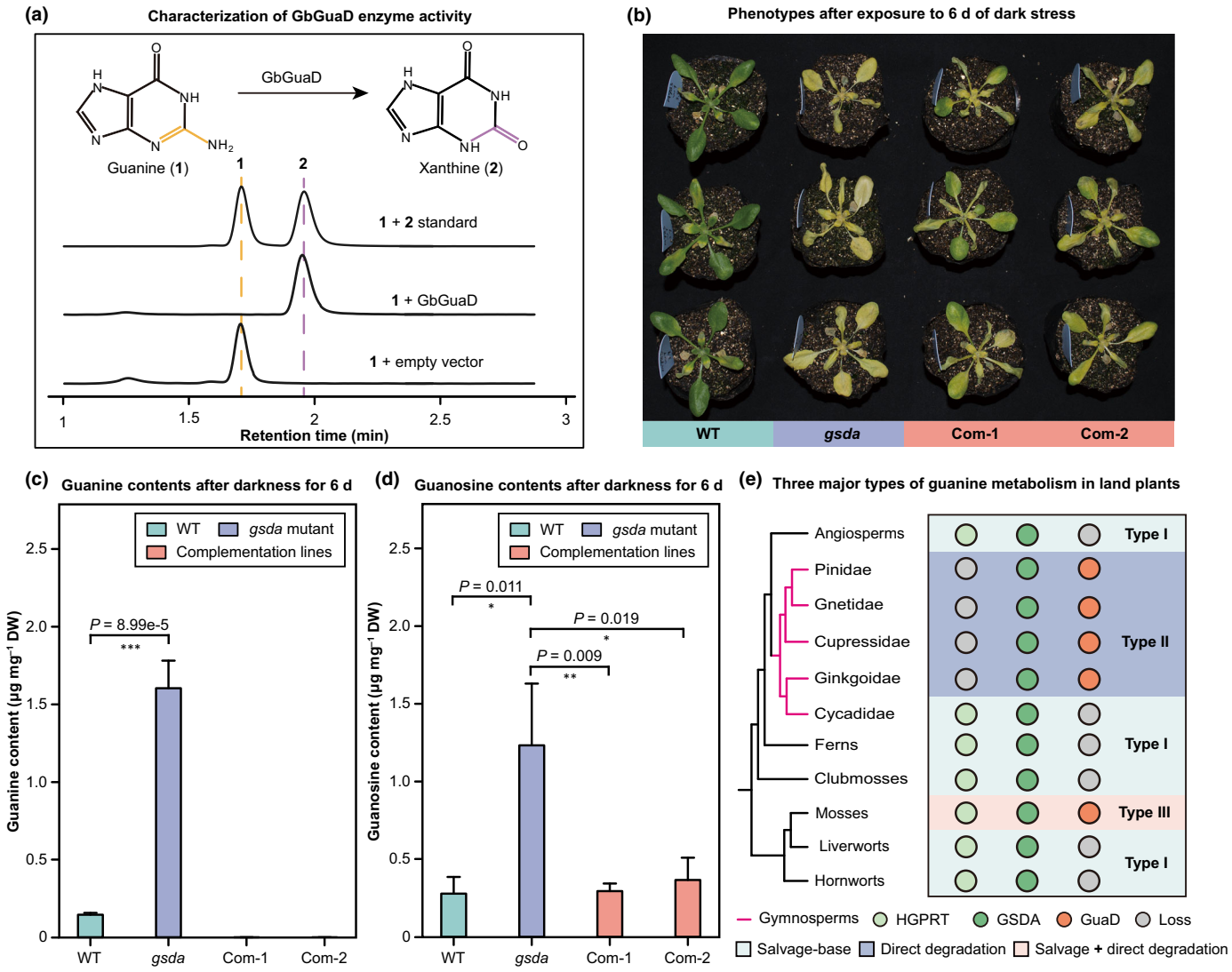


**Fig. 5** Foreign gene *GuaD* acquired from bacteria involved in guanine metabolism. (a) Chromatograms of GbGuaD-catalyzed deamination of guanine. Elution peaks for the standards guanine (**1**) and xanthine (**2**) were indicated. (b) Phenotypic response to darkness in *Arabidopsis thaliana* wild-type (WT), *gsda* mutant (*gsda*), and two complementation lines (Com-1 and Com-2). For the dark stress treatment, seedlings were covered for 6 d after 24 d of growth. (c) Mean ± SE, Guanine concentration in WT, *gsda* mutant (*gsda*), and two complementation lines (Com-1 and Com-2) after 6 d of darkness. (d) Mean ± SE, Guanosine concentration in WT, *gsda* mutant (*gsda*), and two complementation lines (Com-1 and Com-2) after 6 d of darkness. Three samples per genotype were used as biological replicates. (e) Three major types of guanine metabolism in land plants. There were three major types of guanine metabolism: primarily salvage-based (Type I), solely direct degradation (Type II), and a combination of salvage and direct degradation in land plants (Type III). Circles denote hypoxanthine-guanine phosphoribosyl transferase (HGPRT) (light green), GSDA (dark green), and GuaD (Orange) enzymes, respectively; gray circle indicates the loss of gene. Kruskal–Wallis, asterisks indicate statistical significance, that is, *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

enzyme GuaD in 522 land plant genomes. Our results showed that only mosses retained both HGPRT and GuaD in guanine metabolism. However, the native guanine salvage enzyme HGPRT was lost with the acquisition of GuaD via HGT in four of five clades of gymnosperms, including Ginkgoidae, Cupressidae, Gnetidae, and Pinidae. Conversely, in the Cycadidae clade of gymnosperms, GuaD was lost while HGPRT was retained. Overall, there are three major types of guanine metabolism in land plants: primarily salvage-based, solely direct degradation, and a combination of salvage and direct degradation (Fig. 5e).

## Discussion

### HGT is unevenly distributed in land plants

A recent analysis of HGTs in 40 green plants (8 charophytes, 10 seedless plants, and 22 seed plants) reported that there were contrasting scales of relatively recent HGT events between seedless and seed plants (Ma *et al.*, 2022). Using a broad and representative taxonomic sampling approach (especially among seed plants), our systematic identification of 1150 HGT-acquired

genes in 522 high-quality land plants also showed that the average number of recent HGT events occurring in seedless plants was also remarkably higher than in seed plants (Fig. 2a). Our results are congruent with previous study (Ma *et al.*, 2022), but offer a finer resolution on seed plants. We found that there were a considerable number of acquired gene families retained in gymnosperms, a group of early diverging seed plants (Yang *et al.*, 2022; Fig. 2b). Of these, 17 gene families may be the remnants of Episode 2, and the remaining six were likely acquired by the last common ancestor of vascular plants or gymnosperms (Table S4). These retained gene families often comprise multiple copies in gymnosperms, and are directly or indirectly involved in plant adaptation to challenging environments. For example, the cyanamide hydratase genes transferred from fungi to the last common ancestor of vascular plants may directly contribute to the biodegradation and utilization of nitriles, which are widely distributed in the environment and toxic to higher eukaryotes (Li *et al.*, 2015). Notably, the HGT-acquired genes encoding the glutathione-dependent formaldehyde-activating enzyme (GFA), acquired from bacteria by the last common ancestor of land plants, may indirectly enhance UV-B resistance by accelerating the degradation of formaldehyde induced by UV-B (Goenrich *et al.*, 2002; Y. Wang *et al.*, 2022). In addition, we also identified some previously unreported recent HGT events in certain clades of flowering plants (Table S4). For example, HGT of cold shock protein (CSP) from bacteria may improve the ability of triticeae plants to adapt to cold and dry habitats (Yu *et al.*, 2017; Kim *et al.*, 2021). This suggests that, although relatively rare in flowering plants, recent HGT events that facilitate adaptation to a changing environment are constantly occurring (Huang, 2013).

## Contributor to the adaptation of foreign genes in land plants

In general, many studies focus on the fundamental innovations that HGTs bring to land plants (Yue *et al.*, 2012; Bowman *et al.*, 2017; H. Wang *et al.*, 2020; S. Wang *et al.*, 2020; Chen *et al.*, 2022; Liu *et al.*, 2022; Marchant *et al.*, 2022; Lou *et al.*, 2023; Zhang *et al.*, 2024), but the factors that contribute to the post-transfer adaptation of these foreign genes in land plant genomes are still poorly understood. Some studies discussed that foreign genes that maintained in eukaryotic genomes for a long time could acquire eukaryotic features, and intron gains could increase the expression of foreign genes (Husnik & McCutcheon, 2018; Fan *et al.*, 2020). A recent study showed that intron gains enabled HGT-acquired genes to increase their length toward the average length of native genes in insect genomes, and HGT-acquired genes with introns had higher expression levels than HGT-acquired genes without introns (Li *et al.*, 2022). In our study, we found that 683/1150 HGT-acquired genes contain repeat-rich introns, which were likely gained from the native land plant genome after the initial gene transfers (Fig. 2). Furthermore, a comparison of gene structure shows that intron gain events brought HGT-acquired genes closer to native genes in terms of average intron length. More importantly, HGT-acquired genes containing introns also exhibited higher expression levels than those lacking intron in land plants (Fig. 3). This indicates that intron gains can increase expression levels, which is consistent with previous studies (Husnik *et al.*, 2013; Blondel *et al.*, 2020; Rossi *et al.*, 2020; Li *et al.*, 2022; Niu *et al.*, 2022). Overall, our results suggest that repeat-rich introns gained from the native land plant genome may have contributed to the adaptation of HGT-acquired genes in land plant genomes.

## Impact of foreign genes on conserved primary metabolism in land plants

Previous studies have evidenced the contribution of HGT-acquired genes to key traits in land plants (Chen *et al.*, 2021; Ma *et al.*, 2022; Prasad *et al.*, 2022; Wang *et al.*, 2023). Due to the challenges of genetic manipulation in nonmodel plants, only a few of these previously reported HGT-acquired genes have been well studied (H. Wang *et al.*, 2020; S. Wang *et al.*, 2020; Q. Wang *et al.*, 2022; Chen *et al.*, 2022; Liu *et al.*, 2022; Marchant *et al.*, 2022; Naumann *et al.*, 2022; Guan *et al.*, 2023; Lou *et al.*, 2023). Based on the available functional annotations and pathway analysis, we found that most of the HGT-acquired genes were enriched in the metabolism pathway, particularly primary metabolism, which is consistent with previous study (Yue *et al.*, 2012; Ma *et al.*, 2022). Primary metabolites generally play an important role in maintaining the basic life activities of plants, while a large number of secondary metabolites with complex structures are used to respond to changes in the shifting environment (Neilson *et al.*, 2013; Erb & Kliebenstein, 2020). Accordingly, primary metabolism was commonly assumed to be more evolutionarily conserved than secondary metabolism, which may make it more challenging for foreign genes involved in primary metabolism to maintain gene function or persist in the recipient genomes over time (Maeda & Fernie, 2021; S. Wang *et al.*, 2022). Interestingly, the results of our functional verification showed that the foreign gene *GuaD*, acquired from bacteria by mosses and gymnosperms, not only maintained its original function and plugged into the conserved primary metabolism, but also led to the loss of the native pathway gene *HGPRT* in four clades of gymnosperms (Fig. 5). The dynamic process of foreign gene transfer and loss of the original pathway gene or the HGT-acquired gene together shaped three major types of guanine metabolism (Fig. 5e). The diversification of guanine metabolism may reflect the adaptive evolution of land plants to heterogeneous habitats. Additionally, a recent study had also identified a guanine deaminase enzyme (OsGDA1) in rice (Gotarkar *et al.*, 2021). However, significant differences in homology (*c.* 21%) and structure of GbGuaD and OsGDA1 implied that they were isoenzymes with the same chemical reaction but different evolutionary origins (Fig. S5d). Accordingly, other potential types of guanine metabolism in land plants deserve further explorations.

To summarize, the uneven distribution pattern of recent HGT events was confirmed between seedless and seed plants (12 vs 0.5 per species) based on the most extensive sampling strategy. In seed plants, gymnosperms retained more foreign gene families than angiosperms (3.3 vs 0.4 per species). HGT-acquired genes evolved to be more structurally similar to native genes due to

intron gains which may also be involved in post-transfer adaptation. The bacteria-derived gene, *GbGuaD*, introduced a novel bypass for guanine degradation and resulted in the loss of the native pathway gene in some gymnosperms, shaping the diversified pathways of guanine metabolism in land plants.

## Acknowledgements

We thank Zheng-Zhang Zhu, Shi-Yi Lin, and Hao-Jie Jiang for assistance in the HGT identification and Xin-Yi Fan for assistance in the polymerase chain reaction validation. We also thank Tuo Yang, Tongzhou Tao, Ren-Yu Liao, and Guang-Jing Ma for kindly providing plant genomic DNA for polymerase chain reaction validations. We are grateful to Drs Jinling Huang and Qia Wang for their valuable comments and suggestions. This research was supported by the National Natural Science Foundation of China (Nos. 32071484, 32371691, 32200231), the Natural Science Foundation of Zhejiang Province (No. LZ21C030003, LZ23C020002, LR23C140001), the National Key R&D Program of China (2022YFD1401600), the Fundamental Research Funds for the Central Universities (226-2023-00021), and Beijing Life Science Academy (2023000CC0010).

## Competing interests

None declared.

## Author contributions

Y-PZ, X-XS and RP conceived and designed the study. J-JW and CL performed computational analyses. J-JW and C-FL collected genomes. Q-WD, Y-YQ and YS performed transgenic experiments. Y-YQ and JL performed metabolite analysis. J-JW wrote the original draft. Y-PZ, X-XS, RP, LL and YR edited the paper. J-JW, Q-WD and Y-YQ contributed equally to this work.

## ORCID

Ronghui Pan https://orcid.org/0000-0002-4264-5566
Xing-Xing Shen https://orcid.org/0000-0001-5765-1419
Yun-Peng Zhao https://orcid.org/0000-0003-4393-8472

## Data availability

All gene alignments and gene trees can be found in the figshare repository (https://figshare.com/s/6215cdfb1cb9df3c1118). RNA sequencing data of female buds and male buds of *Ginkgo biloba* are available in GinkgoDB under accession no.: T0001 (https://ginkgo.zju.edu.cn/project/T0001/). RNA sequencing data of leaf, stem, and root of *Ginkgo biloba* have been deposited in CNCB under Project ID: PRJCA018702 (https://www.cncb.ac.cn/).

## References

**Berry JA, Beerling DJ, Franks PJ. 2010.** Stomata: key players in the earth system, past and present. *Current Opinion in Plant Biology* **13**: 232–239.

**Blondel L, Jones TE, Extavour CG, Rokas A, Wittkopp PJ, Gazave E. 2020.** Bacterial contribution to genesis of the novel germ line determinant oskar. *eLife* **9**: e45539.

**Bolger AM, Lohse M, Usadel B. 2014.** TRIMMOMATIC: a flexible trimmer for illumina sequence data. *Bioinformatics* **30**: 2114–2120.

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

**Bowman JL. 2022.** The origin of a land flora. *Nature Plants* **8**: 1352–1369.

**Bowman JL, Kohchi T, Yamato KT, 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.

**Buchfink B, Reuter K, Drost H. 2021.** Sensitive protein alignments at tree-of-life scale using diamond. *Nature Methods* **18**: 366–368.

**Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021.** EGGNOG-MAPPER v.2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution* **38**: 5825–5829.

**Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón 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, Yang Z. 2021.** Adaptive innovation of green plants by horizontal gene transfer. *Biotechnology Advances* **46**: 107671.

**Chen R, Lu Y, Zhang E, Chen Z, Huangfu L, Zuo Z, Zhao Y, Zhu M, Zhang Z, Chuan M *et al.* 2022.** The plant streptolysin S (SLS)-associated gene B confers nitroaromatic tolerance and detoxification. *Journal of Hazardous Materials* **433**: 128779.

**Clough SJ, Bent AF. 1998.** Floral dip: a simplified method for agrobacterium-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal* **16**: 735–743.

**Dahncke K, Witte C. 2013.** Plant purine nucleoside catabolism employs a guanosine deaminase required for the generation of xanthosine in *Arabidopsis*. *Plant Cell* **25**: 4101–4109.

**Deng Q, Li H, Feng Y, Xu R, Li W, Zhu R, Akhter D, Shen X, Hu J, Jiang H *et al.* 2022.** Defining upstream enhancing and inhibiting sequence patterns for plant peroxisome targeting signal type 1 using large-scale *in silico* and *in vivo* analyses. *The Plant Journal* **111**: 567–582.

**Donoghue PCJ, Harrison CJ, Paps J, Schneider H. 2021.** The evolutionary emergence of land plants. *Current Biology* **31**: R1281–R1298.

**Erb M, Kliebenstein DJ. 2020.** Plant secondary metabolites as defenses, regulators, and primary metabolites: the blurred functional trichotomy. *Plant Physiology* **184**: 39–52.

**Fan X, Qiu H, Han W, Wang Y, Xu D, Zhang X, Bhattacharya D, Ye N. 2020.** Phytoplankton pangenome reveals extensive prokaryotic horizontal gene transfer of diverse functions. *Science Advances* **6**: eaba0111.

**Goenrich M, Bartoschek S, Hagemeier CH, Griesinger C, Vorholt JA. 2002.** A glutathione-dependent formaldehyde-activating enzyme (GFA) from *Paracoccus denitrificans* detected and purified via two-dimensional proton exchange NMR spectroscopy. *Journal of Biological Chemistry* **277**: 3069–3072.

**Gotarkar D, Longkumer T, Yamamoto N, Nanda AK, Iglesias T, Li L, Miro B, Blanco Gonzalez E, Montes Bayon M, Olsen KM *et al.* 2021.** A drought-responsive rice amidohydrolase is the elusive plant guanine deaminase with the potential to modulate the epigenome. *Physiologia Plantarum* **172**: 1853–1866.

**Guan Y, Ma L, Wang Q, Zhao J, Wang S, Wu J, Liu Y, Sun H, Huang J. 2023.** Horizontally acquired fungal killer protein genes affect cell development in mosses. *The Plant Journal* **113**: 665–676.

**Harris BJ, Clark JW, Schrempf D, Szöllősi GJ, Donoghue PCJ, Hetherington AM, Williams TA. 2022.** Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex common ancestor of land plants. *Nature Ecology & Evolution* **6**: 1634–1643.

**Heinemann KJ, Yang S, Straube H, Medina-Escobar N, Varbanova-Herde M, Herde M, Rhee S, Witte C. 2021.** Initiation of cytosolic plant purine nucleotide catabolism involves a monospecific xanthosine monophosphate phosphatase. *Nature Communications* **12**: 6846.

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

**Huang J, Yue J. 2013.** Horizontal gene transfer in the evolution of photosynthetic eukaryotes. *Journal of Systematics and Evolution* **51**: 13–29.

**Husnik F, McCutcheon JP. 2018.** Functional horizontal gene transfer from bacteria to eukaryotes. *Nature Reviews Microbiology* **16**: 67–79.

**Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, Tanaka M, Satoh N, Bachtrog D, Wilson ACC *et al.* 2013.** Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* **153**: 1567–1578.

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

**Keeling PJ, Palmer JD. 2008.** Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics* **9**: 605–618.

**Kfoury B, Rodrigues WFC, Kim S, Brandizzi F, Del-Bem L. 2024.** Multiple horizontal gene transfer events have shaped plant glycosyl hydrolase diversity and function. *New Phytologist* **242**: 809–824.

**Kim SY, Kim JS, Cho W, Jun KM, Du X, Kim KD, Kim YK, Lee GS. 2021.** A cold-shock protein from the south pole-dwelling soil bacterium *Arthrobacter* sp. confers cold tolerance to rice. *Genes* **12**: 1589.

**Kleinboelting N, Huep G, Kloetgen A, Viehoever P, Weisshaar B. 2011.** GABI-Kat simpleSearch: new features of the *Arabidopsis thaliana* T-DNA mutant database. *Nucleic Acids Research* **40**: D1211–D1215.

**Lenton TM, Dahl TW, Daines SJ, Mills BJW, Ozaki K, Saltzman MR, Porada P. 2016.** Earliest land plants created modern levels of atmospheric oxygen. *Proceedings of the National Academy of Sciences, USA* **113**: 9704–9709.

**Letunic I, Bork P. 2019.** Interactive tree of life (ITOL) v4: recent updates and new developments. *Nucleic Acids Research* **47**: W256–W259.

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

**Li J, Biss M, Fu Y, Xu X, Moore SA, Xiao W. 2015.** Two duplicated genes *ddi2* and *ddi3* in budding yeast encode a cyanamide hydratase and are induced by cyanamide. *Journal of Biological Chemistry* **290**: 12664–12675.

**Li Y, Liu Z, Liu C, Shi Z, Pang L, Chen C, Chen Y, Pan R, Zhou W, Chen X *et al.* 2022.** HGT is widespread in insects and contributes to male courtship in lepidopterans. *Cell* **185**: 2975–2987.

**Liu Y, Wang S, Li L, Yang T, Dong S, Wei T, Wu S, Liu Y, Gong Y, Feng X *et al.* 2022.** The *Cycas* genome and the early evolution of seed plants. *Nature Plants* **8**: 389–401.

**Lou H, Song L, Li X, Zi H, Chen W, Gao Y, Zheng S, Fei Z, Sun X, Wu J. 2023.** The *Torreya grandis* genome illuminates the origin and evolution of gymnosperm-specific sciadonic acid biosynthesis. *Nature Communications* **14**: 1315.

**Ma J, Wang S, Zhu X, Sun G, Chang G, Li L, Hu X, Zhang S, Zhou Y, Song C *et al.* 2022.** Major episodes of horizontal gene transfer drove the evolution of land plants. *Molecular Plant* **15**: 857–871.

**Maeda HA, Fernie AR. 2021.** Evolutionary history of plant metabolism. *Annual Review of Plant Biology* **72**: 185–216.

**Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021.** BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution* **38**: 4647–4654.

**Marchant DB, Chen G, Cai S, Chen F, Schafran P, Jenkins J, Shu S, Plott C, Webber J, Lovell JT *et al.* 2022.** Dynamic genome evolution in a model fern. *Nature Plants* **8**: 1038–1051.

**Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020.** IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37**: 1530–1534.

**Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ. 2018.** The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences, USA* **115**: E2274–E2283.

**Naumann C, Heisters M, Brandt W, Janitza P, Alfs C, Tang N, Toto Nienguesso A, Ziegler J, Imre R, Mechtler K *et al.* 2022.** Bacterial-type ferroxidase tunes iron-dependent phosphate sensing during *Arabidopsis* root development. *Current Biology* **32**: 2189–2205.

**Neilson EH, Goodger JQD, Woodrow IE, Møller BL. 2013.** Plant chemical defense: at what cost? *Trends in Plant Science* **18**: 250–258.

**Nic Lughadha E, Govaerts R, Belyaeva I, Black N, Lindon H, Allkin R, Magill R, Nicolson N. 2016.** Counting counts: revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of other recent estimates. *Phytotaxa* **272**: 82.

**Niu S, Li J, Bo W, Yang W, Zuccolo A, Giacomello S, Chen X, Han F, Yang J, Song Y *et al.* 2022.** The Chinese pine genome and methylome unveil key features of conifer evolution. *Cell* **185**: 204–217.

**Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T *et al.* 2019.** Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology* **20**: 275.

**Paradis E, Claude J, Strimmer K. 2004.** APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290.

**Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017.** Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**: 417–419.

**Prasad A, Chirom O, Prasad M. 2022.** Horizontal gene transfer and the evolution of land plants. *Trends in Plant Science* **27**: 1203–1205.

**Quinlan AR, Hall IM. 2010.** BEDTOOLS: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

**Riegler H, Geserick C, Zrenner R. 2011.** *Arabidopsis thaliana* nucleosidase mutants provide new insights into nucleoside degradation. *New Phytologist* **191**: 349–359.

**Rossi M, Hausmann AE, Thurman TJ, Montgomery SH, Papa R, Jiggins CD, McMillan WO, Merrill RM. 2020.** Visual mate preference evolution during butterfly speciation is linked to neural processing genes. *Nature Communications* **11**: 4763.

**Schliep KP. 2011.** PHANGORN: phylogenetic analysis in R. *Bioinformatics* **27**: 592–593.

**Schroeder RY, Zhu A, Eubel H, Dahncke K, Witte C. 2018.** The ribokinases of *Arabidopsis thaliana* and *Saccharomyces cerevisiae* are required for ribose recycling from nucleotide catabolism, which in plants is not essential to survive prolonged dark stress. *New Phytologist* **217**: 233–244.

**Shen W, Le S, Li Y, Hu F. 2016.** SEQKIT: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**: e0163962.

**Shen X, Opulente DA, Kominek J, Zhou X, Steenwyk JL, Buh KV, Haase MAB, Wisecaver JH, Wang M, Doering DT *et al.* 2018.** Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* **175**: 1533–1545.

**Shi J, Tian Z, Lai J, Huang X. 2023.** Plant pan-genomics and its applications. *Molecular Plant* **16**: 168–186.

**Soucy SM, Huang J, Gogarten JP. 2015.** Horizontal gene transfer: building the web of life. *Nature Reviews Genetics* **16**: 472–482.

**Steinegger M, Salzberg SL. 2020.** Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in genbank. *Genome Biology* **21**: 115.

**Wang H, Li Y, Zhang Z, Zhong B. 2023.** Horizontal gene transfer: driving the evolution and adaptation of plants. *Journal of Integrative Plant Biology* **65**: 613–616.

**Wang H, Sun S, Ge W, Zhao L, Hou B, Wang K, Lyu Z, Chen L, Xu S, Guo J *et al.* 2020.** Horizontal gene transfer of *fhb7* from fungus underlies *Fusarium* head blight resistance in wheat. *Science* **368**: eaba5435.

**Wang Q, Smith SM, Huang J. 2022.** Origins of strigolactone and karrikin signaling in plants. *Trends in Plant Science* **27**: 450–459.

**Wang S, Guan Y, Wang Q, Zhao J, Sun G, Hu X, Running MP, Sun H, Huang J. 2020.** A mycorrhizae-like gene regulates stem cell and gametophore development in mosses. *Nature Communications* **11**: 2030.

**Wang S, Li Y, He L, Yang J, Fernie AR, Luo J. 2022.** Natural variance at the interface of plant primary and specialized metabolism. *Current Opinion in Plant Biology* **67**: 102201.

**Wang Y, Wang J, Lv Q, He Y. 2022.** Adh2/GSNOR1 is a key player in limiting genotoxic damage mediated by formaldehyde and UV-B in *Arabidopsis*. *Plant, Cell & Environment* **45**: 378–391.

**Witte C, Herde M. 2019.** Nucleotide metabolism in plants. *Plant Physiology* **182**: 63–78.

**Yang Y, Ferguson DK, Liu B, Mao K, Gao L, Zhang S, Wan T, Rushforth K, Zhang Z. 2022.** Recent advances on phylogenomics of gymnosperms and a new classification. *Plant Diversity* **44**: 340–350.

**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 Phytologist* **206**: 807–816.

**Yu T, Xu Z, Guo J, Wang Y, Abernathy B, Fu J, Chen X, Zhou Y, Chen M, Ye X *et al.* 2017.** Improved drought tolerance in wheat plants overexpressing a synthetic bacterial cold shock protein gene *SeCspA*. *Scientific Reports* **7**: 44050.

**Yue J, Hu X, Sun H, Yang Y, Huang J. 2012.** Widespread impact of horizontal gene transfer on plant colonization of land. *Nature Communications* **3**: 1152.

**Zhang Z, Diao R, Sun J, Liu Y, Zhao M, Wang Q, Xu Z, Zhong B. 2024.** Diversified molecular adaptations of inorganic nitrogen assimilation and signaling machineries in plants. *New Phytologist* **241**: 2108–2123.

## Supporting Information

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

**Fig. S1** The workflow for inference and reliability verification of HGTs in land plants.

**Fig. S2** The sensitive analysis of the cutoff value of outg_pct.

**Fig. S3** Robustness of HGT inference.

**Fig. S4** Agarose gel electrophoresis of PCR products of 58 randomly selected HGT-acquired genes in six land plants representing five of seven clades.

**Fig. S5** Foreign gene *GuaD* acquired from bacteria involved in guanine metabolism.

**Fig. S6** Intron gains were likely involved in HGT adaptation.

**Fig. S7** KEGG analysis of 1150 HGT-acquired genes in our study.

**Table S1** The list of genome information for 522 land plants and seven outgroups used in our study.

**Table S2** The primers used in this study.

**Table S3** The list of 1150 HGT-acquired genes in land plants identified in our study.

**Table S4** The list of putative 322 HGT events in land plants identified in our study.

**Table S5** Gene expression of 573 HGT-acquired genes in 20 land plants for each transcriptome dataset.

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