New Phytologist

Research

Widespread lateral gene transfer among grasses

Samuel G. S. Hibdige, Pauline Raimondeau, Pascal-Antoine Christin and Luke T. Dunning

Animal and Plant Sciences, University of Sheffield, Western Bank Sheffield S10 2TN, UK

Author for correspondence:
Luke T. Dunning
Email: l.dunning@sheffield.ac.uk

Received: *6 January 2021*
Accepted: *28 February 2021*

New Phytologist (2021)
doi: 10.1111/nph.17328

Key words: adaptation, evolution, genomics, horizontal gene transfer, phylogenomics, Poaceae.

Summary

- Lateral gene transfer (LGT) occurs in a broad range of prokaryotes and eukaryotes, occasionally promoting adaptation. LGT of functional nuclear genes has been reported among some plants, but systematic studies are needed to assess the frequency and facilitators of LGT.
- We scanned the genomes of a diverse set of 17 grass species that span more than 50 Ma of divergence and include major crops to identify grass-to-grass protein-coding LGT.
- We identified LGTs in 13 species, with significant variation in the amount each received. Rhizomatous species acquired statistically more genes, probably because this growth habit boosts opportunities for transfer into the germline. In addition, the amount of LGT increases with phylogenetic relatedness, which might reflect genomic compatibility among close relatives facilitating successful transfers. However, genetic exchanges among highly divergent species indicates that transfers can occur across almost the entire family.
- Overall, we showed that LGT is a widespread phenomenon in grasses that has moved functional genes across the grass family into domesticated and wild species alike. Successful LGTs appear to increase with both opportunity and compatibility.

Introduction

The adaptive potential of a species is limited by its evolutionary history, the amount of standing genetic variation and the rate of new mutations (Barrett & Schluter, 2008). Lateral gene transfer (LGT) enables organisms to overcome these limitations by exchanging genetic material between lineages that have evolved significant reproductive barriers (Doolittle, 1999). LGT is an important evolutionary force in prokaryotes, with up to 60% of genes within a species pan-genome acquired in this manner (Freschi *et al.*, 2018). The genes transferred can have a dramatic effect on adaptation, facilitating the colonisation of new niches and the development of novel phenotypes, as exemplified by the rapid spread of antibiotic resistance in bacteria (Ochman *et al.*, 2000). While LGT is more prevalent in prokaryotes, it has also been documented in various multicellular eukaryotes (reviewed in: Anderson, 2005; Keeling & Palmer, 2008; Schönknecht *et al.*, 2014; Husnik and McCutcheon, 2018; Van Etten & Bhattacharya, 2020), including plants (reviewed in: Richardson & Palmer, 2007; Gao *et al.*, 2014; Wickell & Li, 2019; Chen *et al.*, 2021).

DNA has been transferred into plants from prokaryotes, fungi and viruses, with recipients in particular in algae (Cheng *et al.*, 2019) and bryophytes (Yue *et al.*, 2012; Maumus *et al.*, 2014; Bowman *et al.*, 2017; Zhang *et al.*, 2020). As for plant-to-plant transfers, a majority of nuclear LGTs reported so far involve the transfer of genetic material between parasitic species and their hosts, with examples from the genera *Cuscuta* (Vogel *et al.*, 2018; Yang *et al.*, 2019), *Rafflesia* (Xi *et al.*, 2012), and *Striga* (Yoshida *et al.*, 2010). However, plant-to-plant LGT is not restricted to parasitic interactions and has been recorded in ferns (Li *et al.*, 2014) and eight different species of grasses (Vallenback *et al.*, 2008; Christin *et al.*, 2012a; Prentice *et al.*, 2015; Mahelka *et al.*, 2017, 2021; Dunning *et al.*, 2019). Grasses represent one of the best systems to investigate factors promoting LGT between nonparasitic plants as multiple transfers have been identified in the group and there are extensive genomic resources available as a result of their economic and ecological importance (Chen *et al.*, 2018). Early examples of grass-to-grass LGT were largely obtained incidentally, and only one grass genome (*Alloteropsis semialata*) has been comprehensively scanned, with 59 LGTs identified using stringent phylogenetic filters (Dunning *et al.*, 2019). These 59 protein-coding genes were transferred from at least nine different donors as part of 23 large fragments of foreign DNA (up to 170 kb per fragment). A majority of the acquired LGTs within *A. semialata* are expressed, with functions associated with photosynthesis, disease resistance and abiotic stress tolerance (Dunning *et al.*, 2019; Phansopa *et al.*, 2020). While reports of LGT in other species in the group suggest that it is a widespread phenomenon, its full distribution within the family remains to be assessed.

Grasses are very diverse (Soreng *et al.*, 2015), with almost 12 000 species exhibiting extensive phenotypic variation that may influence LGT dynamics. In particular, the family contains both annuals and perennials. If LGT takes place during vegetative growth (e.g. root-to-root inosculation (Dunning *et al.*, 2019), or other graft-like processes (Stegemann & Bock, 2009; Hertle *et al.*, 2021)), the number of LGTs is predicted to be higher in perennial and rhizomatous species. Conversely, if LGT occurs through illegitimate pollination (Christin *et al.*, 2012), the number of LGTs may not vary with growth form as the wind-pollinated syndrome is universal in this group, or it could be higher in annuals that produce seeds more frequently. The frequency of LGT between species is also likely to be influenced by their geographical distribution, as transfers require the physical movement of DNA. The mechanism of transfer will dictate whether the minimal distance lies within the zone of direct contact (e.g. for inosculation) or within the limits of pollen dispersal (e.g. for illegitimate pollination). Finally, successful transfers might be more likely to occur between closely related groups with similar genome features, as observed in prokaryotes (Skippington & Ragan, 2012; Soucy *et al.*, 2015). Most grass diversity is clustered in the two BOP and PACMAD sister groups that diverged more than 50 Ma (Christin *et al.*, 2014). Each of the two groups has more than 5000 taxa and includes model species with complete genomes (Soreng *et al.*, 2015). The family therefore offers unparalleled opportunities to assess whether functional characteristics or phylogenetic distance determines the amount of LGT among nonparasitic plants.

In this study, we use a phylogenomic approach to scan 17 different grass genomes and quantify LGT among them. The sampled species belong to five different clades of grasses: two from the BOP group (Oryzoideae and Pooideae) and three from the PACMAD group (Andropogoneae, Chloridoideae and Paniceae). Together, these five groups contain more than 8000 species or over 70% of the diversity within the whole family (Soreng *et al.*, 2015). In our sampling, each of these five groups was represented by at least two divergent species, allowing us to monitor the number of transfers between each pair of groups. In addition, the sampled species represented various domestication statuses, life-history strategies, genome sizes and ploidy levels (Table 1). Using this sampling design, we: (1) tested whether LGT is more common in certain phylogenetic lineages; and (2) tested whether some plant characteristics are associated with a statistical increase of LGT. We then focused on the donors of the LGTs received by the Paniceae tribe, a group for which seven genomes are available, to (3) test whether the number of LGTs increased with phylogenetic relatedness. Our work represents the first systematic quantification of LGT among members of a large group of nonparasitic plants and sheds new light on the conditions that promote genetic exchanges across species boundaries in plants.

Materials and Methods

Detecting grass-to-grass LGT

We modified the approach previously used by Dunning *et al.* (2019) to identify grass-to-grass LGT. Specifically, the initial mapping filtering step was discarded to avoid preferentially detecting LGTs in groups for which high-coverage genome data are available for multiple closely related species. In total, 17 genomes were scanned for LGT (Table 1), with all phylogenetic analyses based on coding sequences (total = 817 621 genes; mean per species = 48 095 genes; SD = 26 764 genes). Our analytical pipeline relied on BLAST searches followed by phylogenetic inference and filtering based on phylogenetic patterns, and is analogous to existing tools to identify putative orthologues (Emms and Kelly, 2015). Using our custom pipeline allowed us to tailor its details towards identifying putative LGT from any type of gene family. Furthermore, we performed additional synteny analysis to verify that our method recovers true orthologues.

Table 1 Species used in this study and associated traits.

Group	Species	Ploid.	1C	No. of genes tested	Cult.	LH	Clim.	Phot.	Cont.	Rhiz.
Pooideae	*Brachypodium distachyon*[1]	2n	0.31	17 204	N	A	Temp	C$_3$	6	N
Pooideae	*Hordeum vulgare*[2]	2n	5.39	16 192	Y	A	Temp	C$_3$	6	N
Pooideae	*Triticum aestivum*[3]	6n	16.95	56 619	Y	A	Temp	C$_3$	6	N
Oryzoideae	*Oryza sativa*[4]	2n	0.49	19 259	Y	A	Trop	C$_3$	6	N
Oryzoideae	*Leersia perrieri*[5]	2n	0.32	15 777	N	A	Trop	C$_3$	1	N
Chloridoideae	*Eragrostis tef*[6]	4n	0.69	30 605	Y	A	Temp	C$_4$	5	N
Chloridoideae	*Oropetium thomaeum*[7]	2n	0.29	15 168	N	A	Temp	C$_4$	2	N
Chloridoideae	*Zoysia japonica*[8]	4n	0.42	20 416	Y	P	Temp	C$_4$	1	Y
Andropogoneae	*Sorghum bicolor*[9]	2n	0.69	21 962	Y	A	Temp	C$_4$	6	N
Andropogoneae	*Zea mays*[10]	2n	2.65	25 866	Y	A	Temp	C$_4$	6	N
Paniceae	*Alloteropsis semialata*[11]	2n	1.10	23 071	N	P	Trop	C$_4$	3	Y
Paniceae	*Cenchrus americanus*[12]	2n	2.65	20 159	Y	A	Temp	C$_4$	4	N
Paniceae	*Dichanthelium oligosanthes*[13]	2n	0.96	17 761	N	P	Cold	C$_3$	1	N
Paniceae	*Echinochloa crus-galli*[14]	6n	1.37	54 181	N	A	Temp	C$_4$	6	N
Paniceae	*Panicum hallii*[15]	2n	0.55	30 255	N	P	Temp	C$_4$	1	N
Paniceae	*Panicum virgatum*[16]	4n	1.89	45 043	Y	P	Cold	C$_4$	3	Y
Paniceae	*Setaria italica*[17]	2n	0.49	27 465	Y	A	Temp	C$_4$	6	N

Ploid., Ploidy; 1C, 1C genome size in Gb; Cult., cultivated (Y, yes; N, no); LH, life history (A, annual; P, perennial); Clim., climate (Temp, temperate; Trop, tropical); Phot., photosynthetic type; Cont., number of continents; Rhiz., rhizomatous (Y, yes; N, no). [1]International Brachypodium Initiative (2010); [2]International Barley Genome Sequencing Consortium (2012); [3]International Wheat Genome Sequencing Consortium (2014); [4]Goff *et al*. (2002); [5]Stein *et al*. (2018); [6]Cannarozzi *et al*. (2014); [7]VanBuren *et al*. (2015); [8]Tanaka *et al*. (2016); [9]Patterson *et al*. (2009); [10]Schnable *et al*. (2009); [11]Dunning *et al*. (2019); [12]Varshney *et al*. (2017); [13]Studer *et al*. (2016); [14]Guo *et al*. (2017); [15]Lovell *et al*. (2018); [16]*Panicum virgatum* v4.1, DOE-JGI, http://phytozome.jgi.doe.gov/; [17]Bennetzen *et al*. (2012).

As a first step, we verified for each gene whether its relationships to the best-hit match from 36 other species were as expected based on the species tree, to rapidly discard genes that were clearly not LGT and focus subsequent analyses on plausible candidates. In this step, 37-taxa trees were constructed using data from the 17 grass genomes (Table 1), supplemented with transcriptome data for 20 additional species from across the grass family (Moreno-Villena *et al.*, 2018; Supporting information Table S1). For each gene, we used BLASTN to identify the best hit (highest bit-score) with a minimum match length of 300 bp (not necessarily a single continuous BLAST match) from each of the other 36 species. These sequences were then extracted and nucleotide alignments were generated by aligning the BLASTN matching regions to the query sequence using the 'add fragments' parameter in MAFFT v.7.427 (Katoh & Standley, 2013). If the BLASTN match for a species was fragmented, the different fragments were joined into a single sequence after they had been aligned. Alignments with less than 10 species were considered noninformative and consequently discarded (retained 55.9% of genes; total = 457 003 genes; mean per species = 26 883 genes; SD = 13 042 genes; Table S2). For each alignment with 10 species or more, a maximum-likelihood phylogenetic tree was inferred using PHYML v.20120412 (Guindon & Gascuel, 2003) with the GTR+G+I substitution model. Each topology was then midpoint rooted using the PHYTOOLS package in R and PERL scripts (available from GitHub: https://github.com/SamuelHibdige/) were used to identify genes from each focus species nested within a different group of grasses. We focused on five groups (Andropogoneae, Chloridoideae, Oryzoideae, Paniceae and Pooideae) represented by at least two complete genomes that were supported by most gene trees in a previous multigene coalescent species tree analysis (Fig. 1; Dunning *et al.*, 2019). The whole set of analyses was later repeated to detect LGT between well supported subclades within the Paniceae, the most densely sampled group with seven genomes spread across the group (Fig. 1). In these subsequent analyses, we considered LGTs received from two Paniceae clades represented by two genomes and supported by most gene trees in previous analyses (i.e. Cenchrinae and Panicinae, Fig. 1; Dunning *et al.*, 2019). To be considered as nested, the sister group of the query gene (joining at node 1), and their combined sister group (joining at node 2), had to belong to the same grass group to which the query gene does not belong. For genes that were nested, the analysis was repeated with 100 bootstrap replicates produced using PHYML to verify that nesting of the query sequence was supported by bootstrap node support values of at least 50% at either node 1 or node 2. A soft bootstrap node support threshold (50%) was used to retain all potential LGTs for the more stringent second filtering step (see Fig. S1 for the effect of varying this threshold).

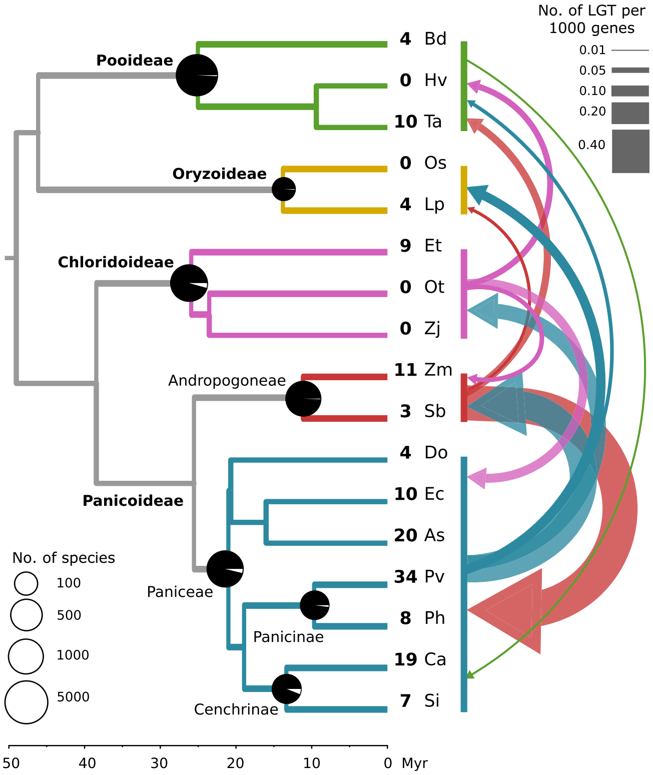

Fig. 1 Distribution of lateral gene transfers (LGTs) among grasses. A time-calibrated phylogenetic tree is shown for the 17 grass species used in this study (phylogenetic tree from Christin *et al*., 2014; scale in Myr). The direction of LGT between grass clades is shown with arrows whose size is proportional to the number of LGTs received. The black portion of pie charts on key nodes of the phylogeny indicates the quartet support for the observed topology based on a multigene coalescence analysis (Dunning *et al*., 2019). The size of each pie chart is proportional to the number of species within the clade (Soreng *et al*., 2015). Numbers at the tips are the number of LGTs detected in each genome.

For candidates that passed the first phylogenetic filter, we performed a second round of filtering using data from 105 genome/transcriptome datasets belonging to 85 species, including the datasets used for the 37-taxa trees (Table S1). For each LGT candidate, we used BLASTN to identify all matches (not just the best match) with a minimum alignment length of 300 bp (not necessarily a single continuous BLAST match) in each of the 105 datasets. Alignments were generated as previously described, before being re-aligned as codons using MAFFT and manually trimmed with a codon-preserving method to remove poorly aligned regions. Maximum-likelihood phylogenies were then inferred using PHYML v.21031022, with the best substitution model identified using Smart Model Selection SMS v.1.8.1 (Lefort *et al.*, 2017). Trees were inspected manually and discarded if: (1) there were too few taxa, with either less than three species within the LGT donor clade or less than three species outside the LGT donor clade; (2) the LGT candidate was not nested within another group of grasses with the increased taxon sampling; or (3) the tree had obvious paralogy problems as a result of gene duplication events. For retained candidates, we removed paralogues representing duplicates originating before the core grasses (BOP and PACMAD clades; Soreng *et al.*, 2015), and joined fragmented transcripts from a single data set if they were nested within the same phylogenetic group. To avoid merging recent paralogues, separate transcripts were retained if they overlapped significantly and had multiple nucleotide substitutions. Up to this point analyses were performed on each gene from each genome, and an individual phylogenetic tree was therefore computed for each gene belonging to a group of recent duplicates (e.g. generated by autopolyploidisation). For subsequent downstream analyses, we only retained one gene tree per group of homologous LGT candidates (e.g. taxon-specific duplicates). The tree inference was then repeated with 100 bootstraps, and the trees were again manually inspected, retaining candidates for which the placement of the LGT in a group was supported by at least one node with ≥ 70% bootstrap node support. Finally, Blastx was used to annotate the LGT candidates against the SwissProt database.

After these two successive filters, retained candidates were subjected to further validation. To verify that nesting of the candidate LGTs was not due to convergent adaptive amino acid substitutions, we generated phylogenetic trees based solely on third codon positions, which are less subject to positive selection (Christin *et al.*, 2012b). Phylogenetic trees were generated as above and were manually inspected to confirm the LGT scenario. To verify that the LGT scenario was statistically better than the species tree, we then conducted approximately unbiased (AU) topology tests that compared the maximum-likelihood topology with a topology representing the null hypothesis (forcing monophyly of the donor and recipient clades; recipients for the within-Paniceae analysis were constrained at the genus level if they did not belong to the Cenchrinae or Panicinae). The null topology was inferred by first constraining the clades and inferring a tree with the GTR + G model in RaxML v.8.2.12 (Stamatakis, 2014), before using this topology as a constraint for a maximum-likelihood phylogeny inferred with PhyML as described above. The AU tests were then performed in Consel v.1.20 (Shimodaira & Hasegawa, 2001) using the site-wise likelihood values generated using PhyML. *P*-values were Bonferroni corrected for multiple testing. LGT candidates with nonsignificant results ($P > 0.05$) were discarded. In some cases, no native copy was present in any species from the group containing the focus species, preventing AU tests. These genes were retained, although the numbers were recorded separately (Table 2; n.b. statistics reported and values quoted in the text include these genes).

For candidates retained after these extra validation steps, new phylogenetic trees were inferred with a denser species sampling to refine the identification of the potential donor. Illumina short-read data sets ($n = 71$; 65 spp.; Table S1) were added to the trees using the method described in Dunning *et al.* (2019). The dense trees were then manually inspected and any presenting strong discrepancies with the expected species relationships were discarded. All separate genes are counted in the final LGT tally for each species, so duplicates (e.g. via polyploidisation) arising after the transfer are counted separately (Table S2).

In summary, to be considered as an LGT each gene: (1) had to be nested within one of the other four groups of grasses (Fig. 1); (2) their nesting had to be well supported (≥ 70% bootstrap node support); (3) potential paralogy problems had to be ruled out (i.e. discarding phylogenies with multiple apparent duplication events that can explain the phylogenetic incongruence); (4) the nesting had to be supported by phylogenetic trees constructed solely from the third codon positions, which are less subject to adaptive convergent evolution; and (5) when possible, nesting had to be supported by approximately unbiased (AU) tests to confirm the LGT topology was a significantly better fit than a topology constrained to match the species tree (see Fig. 2 for exemplar LGTs). Alignments (Dataset S1), 85 taxa phylogenies (Dataset S2), third codon position phylogenies (Dataset S3) and phylogenies with short-read data added (Dataset S4) are included in the Supporting information. All analyses were preformed using publicly available data (Table S1).

Table 2 Number of lateral gene transfers (LGT) detected between the five groups.

			Donor clade				
Clade	Species	No. of LGT	Pooid.	Ory.	Chlor.	Andro.	Pan.
Pooideae	*Brachypodium distachyon*	4	–	0	0	4	0
Pooideae	*Hordeum vulgare*	0	–	0	0	0	0
Pooideae	*Triticum aestivum*	8(10)	–	0	5	0(2)	3
Oryzae	*Oryza sativa*	0	0	–	0	0	0
Oryzae	*Leersia perrieri*	1(4)	0	–	0	1	0(3)
Chloridoideae	*Eragrostis tef*	1(9)	0	0	–	0	1(9)
Chloridoideae	*Oropetium thomaeum*	0	0	0	–	0	0
Chloridoideae	*Zoysia japonica*	0	0	0	–	0	0
Andropogoneae	*Sorghum bicolor*	2(3)	0	0	0	–	2(3)
Andropogoneae	*Zea mays*	11	0	0	2	–	9
Paniceae	*Alloteropsis semialata*	20	0	0	4	16	–
Paniceae	*Cenchrus americanus*	15	0	0	5	10	–
Paniceae	*Dichanthelium oligosanthes*	4	4	0	0	0	–
Paniceae	*Echinochloa crus-galli*	10	0	0	3	7	–
Paniceae	*Panicum hallii*	8	0	0	6	2	–
Paniceae	*Panicum virgatum*	30	0	0	1	29	–
Paniceae	*Setaria italica*	7	0	0	0	7	–

Numbers in parentheses include genes for which approximate unbiased (AU) topology tests could not be performed as no native copy from the same clade was present to constrain the tree topology. Pooid., Pooideae; Ory., Oryzoideae; Chlor., Chloridoideae; Andro., Andropogoneae; Pan., Paniceae.

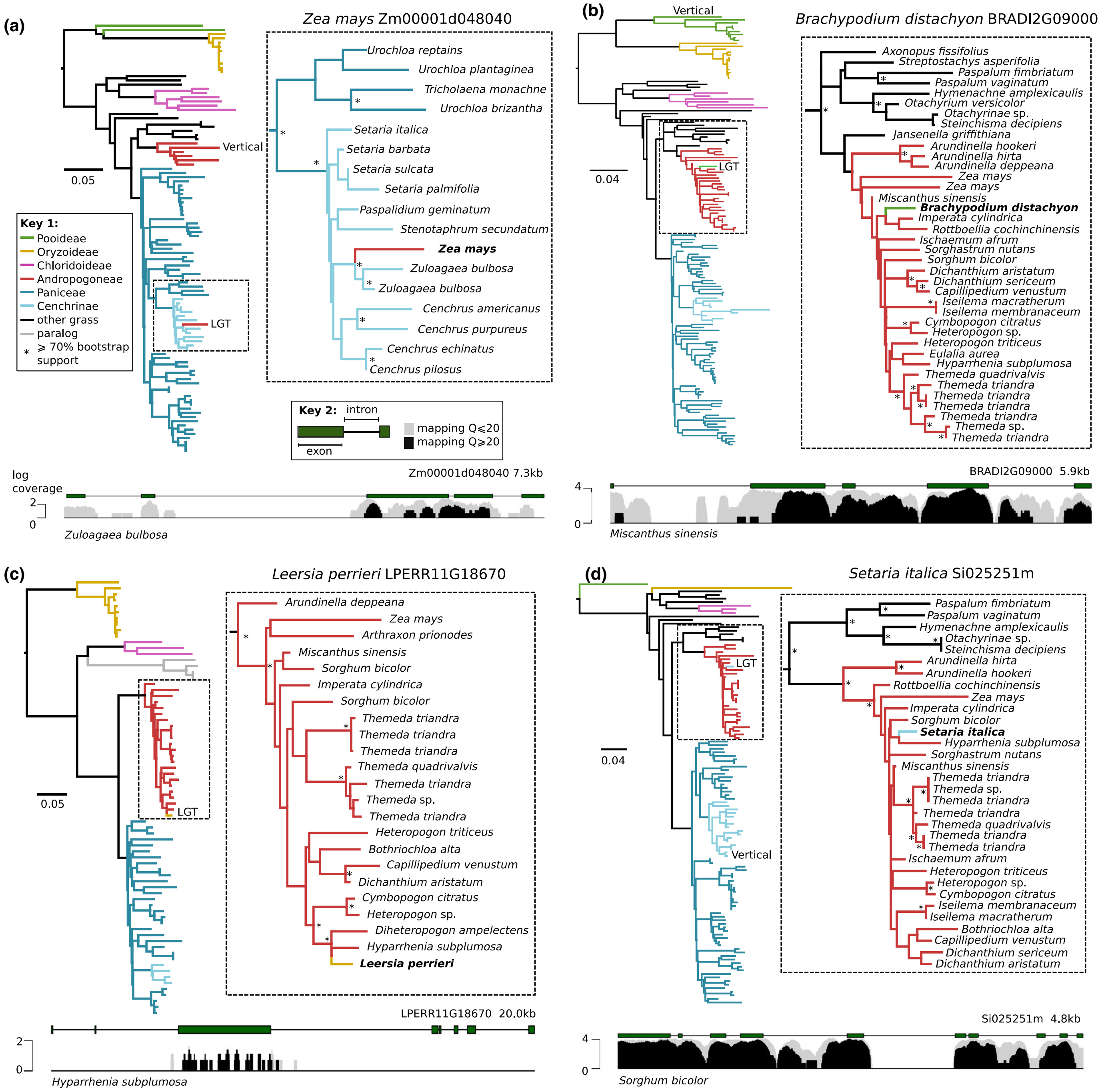

Fig. 2 Four examples of grass-to-grass lateral gene transfer. Each panel (a–d) shows an exemplar grass-to-grass LGT, with full and expanded regions of maximum-likelihood phylogenies shown. Asterisks denote nodes with bootstrap support values ≥ 70%, and branches are coloured per group. A coverage plot for each gene model is shown below, generated from short-read mapping data for a species closely related to the LGT donor.

Synteny analyses

Synteny analyses were performed with all genomes with reasonable contiguity (N50 ≥ 1 Mb; $n = 13$; Table S2) using SynFind (Tang *et al*., 2015) with default parameters. For each LGT in these species, we determined whether genes from the other reference genomes identified as orthologues to the native copy in the phylogenetic trees were syntenic to the LGT or the native copy based on the highest syntelog score (Table S3).

Analyses of replicate sequencing runs to check for potential contamination

Independently sequenced runs from the same accession or cultivar for each of the model species were screened for the presence of each LGT, as potential contaminations would not appear in multiple replicates derived from independent DNA samples. Paired-end Illumina whole-genome data were obtained from the NCBI Sequence Read Archive and mapped to the reference genome using Bowtie2 v.2.3.5.1 (Langmead & Salzberg, 2012) with default parameters. Mean coverage depths for the coding sequence of each gene in the genome were then calculated using Bedtools v.2.26.0 (Quinlan & Hall, 2010), with large BAM files downsampled with Picard Tools v.2.13.2-Snapshot (Broad Institute, 2019).

Confirming LGT scenario with similarity of noncoding regions

Due to rapid divergence, noncoding sequences can only be accurately compared among close relatives. In the case of LGT, similarity of noncoding DNA is therefore expected only when genome data are available for a close relative of the donor (see analyses of *A. semialata*; Dunning *et al.*, 2019; Olofsson *et al.*, 2019). We compared pairwise similarities of noncoding regions (intron and intergenic) of LGT regions versus the rest of the genome for a single multigene fragment from the *Setaria italica* genome. This fragment was selected as it had clear high-quality intergenic mapping ($Q \geq 20$) when using *Sorghum bicolor* data as a proxy for the donor, suggesting that sequence data in this case are available for a close relative of the donor. For this analysis, paired-end Illumina whole-genome data belonging to the putative donor group, as well as close relatives of the recipient, were mapped to the reference genome as described above. We then used Bedtools coverage to calculate the proportion of introns and intergenic regions with nonzero coverage with the different species, testing the hypothesis that coverage from the proxy donor is inflated around the putative LGTs. For introns, we restricted the analysis to those between 200 bp and 2 kb. For intergenic regions, we randomly generated windows using Bedtools shuffle, excluding gene regions from the analysis.

Grass traits and statistical analyses

Plant traits were obtained from various sources. Life history, distribution, growth form and the domestication status were retrieved from GrassBase (Clayton *et al.*, 2016). 1C Genome sizes were obtained from the Plant DNA *C*-values database (Pellicer & Leitch, 2020), and climatic information from Watcharamongkol *et al.* (2018). The climate data for *Oropetium thomaeum* were not included in Watcharamongkol *et al.* (2018), and were therefore retrieved from GBIF (GBIF.org; 11 July 2019 GBIF Occurrence Download https://doi.org/10.15468/dl.wyhtoo) and WorldClim (Harris *et al.*, 2014; Fick & Hijmans, 2017) using the same methods. All statistical tests were preformed in R v.3.0.2, with the expected frequencies for chi-squared tests based on the number of genes tested within each species (Table 1). Kruskal–Wallis tests were performed using absolute LGT numbers, which were divided into donor groups when testing whether some clades were more frequent donors than others. To determine if any trait or genome feature was associated with the number of LGTs, we preformed phylogenetic generalised least squares (PGLS) to account for the relatedness between samples. The PGLS analysis was preformed in R with the caper package (Orme *et al.*, 2013) using a time-calibrated phylogenetic tree retrieved from Christin *et al.* (2014), and various traits as explanatory variables (Table 1). Individual and iterative models were performed, removing the least significant variable until only significant variables remained ($P < 0.05$).

Results

LGT occurs in all lineages and functional types of grass

Out of the 817 612 genes from the 17 grass genomes (Table 1) screened, 55.89% had sufficient homologous grass sequences (≥ 10 taxa) for reliable phylogenetic reconstruction (Tables 2,S2), and were tested for LGT. A majority (99.73%) of the initial 37-taxa phylogenies did not support a scenario of LGT among the five grass groups, with successive filtering resulting in the identification of 135 LGT candidates across the 17 species (Table 2; full results Table S2). Expectedly, a higher bootstrap threshold would decrease the number of retained candidates, but even a very conservative threshold of 95% support would identify 99 LGTs (Fig. S1). The number of LGTs received varied among species ($P < 0.01$; chi-squared test; mean = 8.4; SD = 9.0; range = 0–30; Table S2), with the highest numbers observed in *Panicum virgatum* ($n = 30$), *Alloteropsis semialata* ($n = 20$), and *Cenchrus americanus* ($n = 15$). It should be noted that only a subset of the 59 previously reported LGTs in *Alloteropsis semialata* (Dunning *et al.*, 2019; Table S4) were retrieved, as the previous analysis examined more groups of donors not considered here, and secondary candidates based solely on read mapping patterns were not recorded in this study. Despite the significant variation between species, the difference among the five phylogenetic groups was not significant ($P = 0.16$, Kruskal–Wallis test). Overall, our results showed that LGT is widespread across the grass family and occurs in a majority of the species sampled here (Fig. 1; Table 2). No LGTs were detected in four of the 17 species analysed, but some LGT might remain undetected due to our stringent phylogenetic filtering and because only transfers among the predefined five grass clades were considered.

Among the 17 species screened, LGT was observed in all functional groups (Fig. 3). LGT was detected in wild species, but also in major crops. For instance, maize (*Zea mays*) had 11 LGTs received from Chloridoideae and Paniceae, while wheat (*Triticum aestivum*) had 10 LGTs received from Andropogoneae, Chloridoideae and Paniceae (Table 2). The LGTs may be beneficial for the crops, with transferred loci including some with functions related to abiotic stress tolerance and disease resistance (Table S2). Across all plant properties, some seemed to be associated with larger numbers of LGT (Fig. 3). A PGLS analysis was conducted to test whether any of the traits had a significant relationship with the amount of LGT, while accounting for phylogenetic effects. For this, a model was constructed to explain the absolute number of LGTs using nine traits as predictor variables (Table 1) and a time-calibrated phylogenetic tree retrieved from Christin *et al.* (2014). Initially, models were constructed for each predictor variable, with the amount of LGT shown to increase with the presence of rhizomes ($P = 0.026$, adjusted $R^2 = 0.243$) and the number of genes tested ($P = 0.038$, adjusted $R^2 = 0.207$). A

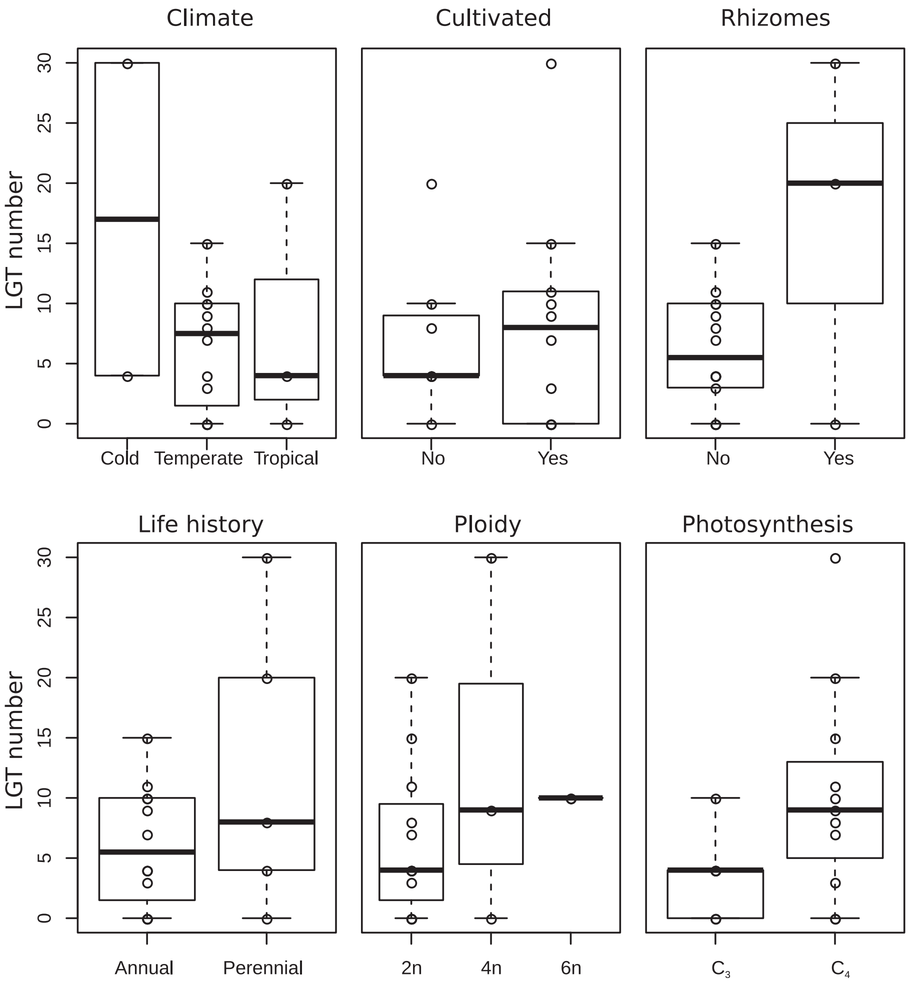

Fig. 3 Number of lateral gene transfers (LGTs) received by different categories of grasses. For each group, the distribution of LGT numbers is shown with box plots connecting the median and the interquartile range, with whiskers connecting the last points within 1.5× the interquartile range. Individual data points are shown with dots.

combined model with all explanatory variables was subsequently performed to test for their joint effects. Iterative models were performed, removing the least significant variable until only significant variables remained ($P < 0.05$). PGLS analysis (combined adjusted $R^2 = 0.652$) identified three characteristics that jointly explained the number of LGTs: the number of genes tested ($P < 0.001$), the presence of rhizomes ($P = 0.002$), and the ploidy level ($P = 0.006$). For LGT that occured before genome duplication, their number would be expected to double in tetraploids and triple in hexaploids, because each homoeologous chromosome would carry a copy of the LGT. While we noted that a majority of LGTs detected in the polyploids have been duplicated ($n = 43$), there were still multiple singletons ($n = 35$). These singletons were either acquired after genome duplication, or they were possibly orphaned as a result of the complexity of polyploid genome assembly. Future studies should use larger sample sizes to definitively show the effects, but our analyses suggested that some categories of grasses were more likely to be involved in LGT.

LGTs are more commonly received from closely related species

Overall, some clades acted more frequently as donors ($P < 0.01$, Kruskal–Wallis test). Specifically, the Andropogoneae were the source of most transfers (Table 2). However, these were mainly received by members of the Paniceae, which are the closest relatives of Andropogoneae in our dataset, and are also represented by the most genomes (Table 1). While these patterns suggested that LGT occurs more often among close relatives, directly comparing the rates was difficult because the clades varied in their number of species, number of genomes available and age. However, for a given clade of recipients, it was possible to compare the frequency of different groups of donors while controlling for the number of species. We therefore focused on the identity of donors of LGT to Paniceae, the group with the highest number of complete genomes from multiple genera.

Seven Paniceae genomes were used in this study, and this increased the sample size further and allowed the detection of intra-Paniceae LGT. We therefore reported the number of LGTs transferred from the Panicinae and Cenchrinae subgroups of Paniceae (each represented by two genomes; Fig. 1) to other Paniceae, in addition to those received from other groups. In total, we identified 129 LGTs across the seven Paniceae genomes, 35 of which were transferred from the Cenchrinae and Panicinae subgroups (Table 3; full results Table S5). When focusing on Paniceae recipients, some groups were more often LGT donors than others even after correcting for the number of species in each donor clade ($P < 0.01$, Kruskal–Wallis test). The number of

Table 3 Number of lateral gene transfers (LGT) detected in Paniceae.

Subgroup	Species	No. of LGT	Pooid. (3698 spp.)	Ory. (115 spp.)	Chlor. (1602 spp.)	Andro. (1202 spp.)	Cench. (287 spp.)	Pani. (157 spp.)
Cenchrinae	*Cenchrus americanus*	16	0	0	5	10	–	1
Cenchrinae	*Setaria italica*	7	0	0	0	7	–	0
Panicinae	*Panicum hallii*	8	0	0	6	2	0	–
Panicinae	*Panicum virgatum*	36	0	0	1	29	6	–
Other	*Alloteropsis semialata*	33(34)	0	0	4	16	13(14)	0
Other	*Dichanthelium oligosanthes*	5	4	0	0	0	1	0
Other	*Echinochloa crus-galli*	23	0	0	3	7	8	5

Number of species in each clade is indicated in parentheses, with values from Soreng *et al*. (2015); Pooid., Pooideae; Ory., Oryzoideae; Chlor., Chloridoideae; Andro., Andropogoneae; Cench., Cenchrinae; Pani., Panicinae.

LGTs given per species decreased with the phylogenetic distance to Paniceae, reaching lowest levels in the BOP clade (Pooideae and Oryzoideae; Fig. 4).

Ruling out alternative hypotheses

There are five main alternative hypotheses to LGT: (1) incomplete lineage sorting; (2) unrecognised paralogy; (3) hybridisation; (4) contamination; and (5) phylogenetic biases, such as convergent evolution. Evidence reducing the likelihood of these alternative explanations is discussed in the following paragraphs.

(**1**) Incomplete lineage sorting: for a majority of the LGTs we detected (79.4%), the recipient genome also contained a native copy, which argues against incomplete lineage sorting as an alternative hypothesis. However, as pseudogenisation of the native copy has been observed in cases in which the LGT acts as a functional replacement (Dunning *et al.*, 2019; Phansopa *et al.*, 2020), their continued coexistence should not always be expected. The coexistence of native and laterally acquired orthologues permitted the comparison of patterns of synteny in multiple species to rule out unrecognised paralogy problems.

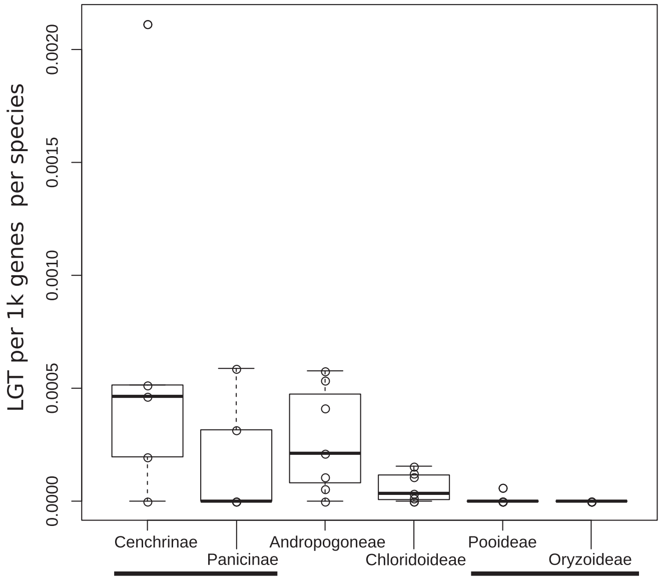

Fig. 4 Number of lateral gene transfers (LGT) received by Paniceae species from different groups. The number of LGTs in each Paniceae genome is corrected by the number of genes tested, as well as the number of species in the group of donors. The phylogenetic distance increases from left to right, with equidistant clades joined by solid bars. Box plots show median, interquartile range and whiskers connecting the last points within 1.5× the interquartile range, with individual data points shown with dots.

(**2**) Unrecognised paralogy: we used 13 species for this analysis, with at least two representatives from each of the five groups. For each LGT detected in these 13 species, we determined whether the genes from the other 12 species that were identified as orthologous in the phylogenetic tree were syntenic to the LGT or the native gene. In total, 76.2% of orthologues were syntenic with the native copy, 2.86% were syntenic with the LGT and 20.9% were syntenic to neither (Table S3). The 2.86% of orthologues syntenic to the LGT corresponded to three genes in *Echinochloa crus-galli* acquired from a Cenchrinae species and could result from technical (e.g. mis-assembly) or biological (e.g. homologous replacement) processes. Overall, the synteny analyses confirm that our phylogenetic trees identified true orthologues in most cases, and the phylogenetic patterns suggesting LGT cannot be explained by widespread unrecognised paralogy.

(**3**) Hybridisation: the patterns of synteny between the native and laterally acquired genes also argue against straightforward hybridisation through sexual reproduction and chromosomal recombination during the transfers, as already argued previously (Dunning *et al.*, 2019). With the exception of three genes in *Echinochloa crus-galli*, the LGTs appear to be inserted into the genome in random locations, often on different chromosomes as the native orthologues.

(**4**) Contamination: we ruled out contamination as the source of the foreign DNA in the genomes by confirming the presence of the laterally acquired DNA in multiple independent sequencing runs (Table S6; Fig. S2). For six of the reference genomes, 'gold-standard' datasets exist, that is whole-genome resequencing data sets for the same cultivar as the reference genome, but that were produced independently from the initial assembly project (Table S6). A further four genomes had multiple libraries from the original assembly project, and these were derived from independent DNA samples (Table S6). For the remaining three species with LGT, the available sequencing data cannot be used to rule out contamination as only the whole-genome data used to generate the reference assembly exist, and when there were multiple sequencing libraries/runs it was unclear whether they were derived from independent

DNA samples (Table S6). For each dataset, we compared the genome-wide mean per-base coverage for each gene to that of the identified LGT (Table S6; Fig. S2), with an expectation that a gene corresponding to sample contamination would have zero (or near zero) coverage in all but one independently produced sequencing runs used to assemble the reference genome, and in none of the sequencing runs produced independently of the reference genome. All LGTs had sequencing data in all independent datasets apart from one gene in *Z. mays*. For this species, we used seven datasets from the same cultivar that were produced independently in seven different labs. Only five out of these seven datasets supported the presence of the LGT *Zm00001d039537*, with the most parsimonious explanation being LGT variation between individuals, as previously documented in *Alloteropsis semialata* (Dunning *et al.*, 2019). A majority of LGTs had coverage depths greater than the 5th (97.0% of LGTs) and 2.5th (99.0% of LGTs) percentile of coverage depth for all genes in the genome (Table S6; Fig. S2). Overall, these results confirmed that, at least for species with independent replicates, contamination in the original reference genomes was not responsible for the presence of the LGTs in the sequence datasets.

(**5**) Phylogenetic bias: convergent evolution or other systematic biases in the data could lead to gene/species tree discordance (Chang & Campbell, 2000). In addition to confirming the patterns with phylogenetic trees built on third positions of codons, we assessed the similarity between the recipient and donor species in noncoding DNA. The mapping of short-read data to four genomes confirmed in some cases a high similarity between the putative donor and recipient on intron sequences of LGTs in addition to exons (Fig. 2). It was however not possible to delimit with high precision the laterally acquired fragments detected here (as carried out for *A. semialata* in Dunning *et al.*, 2019 and Olofsson *et al.*, 2019), either because the transfers were too ancient or because we lacked whole-genome data for very close relatives of the donors. However, we did detect a multigene fragment in *Setaria italica* that also appeared to have laterally acquired intergenic DNA when using *Sorghum bicolor* mapping data as a proxy for the unknown Andropogoneae donor (Fig. S3). For this fragment, we quantified the mapping rates between intron and intergenic LGT regions to the rest of the genome. Out of the 20 972 genes from *S. italica* with at least one intron between 200 and 2000 bp, only 164 had a higher proportion of bases covered by *S. bicolor* reads than the three LGTs in the *S. italica* fragment. Of these 164 genes, only 67 were covered by more reads of the species from the donor group (*S. bicolor*) than of the close relative of *S. italica* (*Cenchrus americanus*). The multigene fragment also included 4.5 kb of laterally acquired intergenic DNA with 91.3% nonzero coverage with the *S. bicolor* data (Fig. S3). This region was compared with 10 000 other randomly sampled 4.5 kb intergenic regions across the genome, all of which had a lower nonzero coverage than that of the LGT region (mean = 2.6%; SD = 7.1%). The observation of some intergenic regions with high similarity (Fig. S3), together with intronic similarities (Fig. 2), further ruled out convergent evolution or other phylogenetic biases (e.g. long branch attraction) as being responsible for all detected cases of gene/species tree discordance.

Discussion

Lateral gene transfer is a potent evolutionary force capable of having a profound effect on the evolutionary trajectory of a species and its descendants (Li *et al.*, 2014; Cheng *et al.*, 2019; Phansopa *et al.*, 2020; Chen *et al.*, 2021). Here, we use grasses as a model to investigate the factors that dictate the prevalence of LGT among plants. Using a combination of stringent phylogenetic and genomic analyses, a grand total of 170 genes were identified (*c.* 3.72 LGTs per 10 000 genes; 135 between the five large groups of grasses and 35 among groups of Paniceae) that had been laterally transferred to 13 of the 17 complete grass genomes that were screened (Tables 1, 3). Our approach was developed to drastically reduce the amount of false positives and was purposely very conservative. This enabled us to minimise the effects of other evolutionary processes such as hybridisation and incomplete lineage sorting. As a result, the number of LGTs identified is likely to be only a subset of those existing in the complete grass genomes. In addition, phylogenetic filtering prevents detection of LGT from clades of grasses for which no genome is available. With the current sampling, approximately 30% of the grass diversity was never considered as potential LGT donors (Soreng *et al.*, 2015). The number of detected LGTs therefore depends on the sampling of genomes, and future studies with additonal species representing more potential donors are likely to lead to further LGT discoveries. Our efforts have already indicated that the phenomenon is prevalent in the family.

Our phylogenetic pipeline prevented the detection of LGT among members of the same group of grasses, such as the numerous exchanges among lineages of Paniceae previously detected (Dunning *et al.*, 2019). This was perfectly exemplified in *A. semialata*, in which 26 LGTs had been detected previously based on phylogenetic analyses (referred to as ‘primary LGT’, with 33 ‘secondary candidates’ detected based on similarity in flanking regions in Dunning *et al.*, 2019; Table S4). Here, only 20 LGTs were identified when considering solely LGT among the five higher groups (Table 2), while a further 14 LGTs were detected when considering subgroups of Paniceae as potential donors (Table 3). Seven more LGT had previously been detected in *A. semialata* from a group of Paniceae (Melinidinae) but were not considered as putative donors here because of the absence of reference genomes (Table S4). These differences highlight the influence of the availability of genomes for putative donors on our ability to detect LGT. In addition, five of the 34 LGTs detected here in the genome of *A. semialata* were not identified in previous analyses of the same genome, showing that LGT detection depends on multiple factors. First, the removal of a first filter based on similarity analyses in this study allowed identification of additional LGTs (Table S4). Conversely, the increased number of genomes in this study influenced the correction of *P*-values for multiple testing, leading sometimes to nonsignificant topology tests (Table S4). In addition, the detection of LGT candidates based on secondary screening of flanking regions in the previous study (33 ‘secondary LGT candidates’ in Dunning *et al.*, 2019; Table S4) demonstrated that some LGTs could not be identified based solely on phylogenetic analyses, because they were too short or not present in enough species to infer robust phylogenetic trees. Finally, our approach precluded the detection of older LGTs that are shared by multiple species among the 17 reference genomes, as reported in other cases (e.g. Li *et al.*, 2014). We conclude that the LGTs reported here describe only a small fraction of those existing in grass genomes. Despite these limitations, we showed that LGT is common in grasses, certain groups exchange more genes compared with others, the frequency of LGT appears to increase in rhizomatous species and there may be a role of phylogenetic distance underpinning the LGT dynamics. Analyses based on more genomes will in the future refine our conclusions, and potentially provide greater statistical power to precisely quantify the effect of different factors on the rate of LGT.

LGT occurs in all functional groups, and is especially prevalent in rhizomatous species

LGT is common in grasses and was observed in each of the five groups investigated here (Fig. 1). We detected LGT in domesticated and wild species alike (Fig. 3), although it is unknown at this time whether the LGTs occurred before or after domestication and whether these genes were associated with agronomic traits. Genetic exchanges were not restricted to any functional category of grasses (Fig. 3), and the ubiquity of the phenomenon provides some support for the breakdown in reproductive behaviour and illegitimate pollination as the mechanism responsible for transfers, as wind pollination is universal in this group. Further work is required to determine how traits associated with the wind-pollinated syndrome (e.g. self compatibility, plant height and pollen longevity) could affect LGT among grasses. There was also a statistical increase of the number of LGTs in rhizomatous species, with two of the three species with the highest numbers of LGTs (*Alloteropsis semialata* and *Panicum virgatum*) are perennials that can propagate vegetatively via rhizomes (Tables 1,3). These patterns suggested that root-to-rhizome contact (i.e. inosculation) provides an increased opportunity for retaining gene transfers, as the integration of foreign DNA in rhizome tissue means that any subsequent plant material regrown from these cells, including reproductive tissue, will contain the LGT. This hypothesis is compatible with previous reports of genetic exchanges following grafts (Stegemann & Bock, 2009; Hertle *et al.*, 2021). In this instance, LGT is similar to somatic mutations occurring in clonal species, as documented in the seagrass *Zostera mariana* where they can ultimately enter the sexual cycle (Yu *et al.*, 2020). The genetic bottleneck and selection characterising rhizomes would further increase the chance of LGT retention, especially if these give a selective advantage (Yu *et al.*, 2020). However, we did not detect LGT in the third rhizomatous species that was sampled (*Zoysia japonica*; Table 1). Increased species sampling, particularly for rhizomatous species represented by only three genomes in this study, is now needed to confirm our conclusions and precisely quantify the effect of growth form on the amount of gene transfers and how this interacts with other factors.

It is easier to acquire genes from close relatives

Within grasses, there is an effect of the phylogenetic distance on the number of transfers observed, as shown by the Paniceae, which received more LGTs from closer relatives (Fig. 4). This pattern mirrors that observed in prokaryotes (Popa & Dagan, 2011; Skippington & Ragan, 2012; Soucy *et al.*, 2015) and insects (Peccoud *et al.*, 2017), in which the frequency of transfers is higher between closely related species. In prokaryotes, this effect is thought to result from more similar DNA sequences promoting homologous replacement of the native copy (Skippington & Ragan, 2012). This is unlikely to play a role in grasses as the LGTs are predominately inserted in nonsyntenic positions in the genome where they coexist with the native copy (Table S3). However, stretches of DNA similar between the donor and recipient (e.g. transposable elements) may still be involved in the incorporation of LGT into the chromosomes, a hypothesis that can be tested when genome assemblies for donor species become available. Alternatively, the effect of phylogenetic distance might stem from the regulation of LGT after acquisition, with genes transferred from closely related species more likely to share regulatory mechanisms. In such a scenario, the phylogenetic effect would reflect the utility of the LGT for the recipient species and therefore selection after transfer rather than the rate of transfer. Overall, our analyses indicates that it is easier to either obtain LGTs from close relatives or to use it after the transfers, thereby increasing the chance of selectively retaining it.

A potential role of overlapping distributions

We observed some transfers between Pooideae and Paniceae, two groups that diverged > 50 Ma, representing one of the earliest splits within this family (GWPGII, 2012). This indicated that LGT is possible across the whole grass family. In our dataset, the only recipient of these transfers was *Dichanthelium oligosanthes* (Table 2), a frost-tolerant grass from North America that inhabits colder areas than other members of the Paniceae (Studer *et al.*, 2016). In cold regions, *D. oligosanthes* can co-occur with members of the Pooideae, and this biogeographic pattern is likely to facilitate exchanges between the two groups of grasses. However, given the difficulties of identifying the donor at the species level (or even genus) with the current data, we cannot be sure that the specific donor and *D. oligosanthes* co-occur. As more whole-genome datasets become available for the diverse Pooideae, co-occurrence between donor and recipient species can be tested directly.

Biogeography might also cause differences in the identity of the LGT donors between the two closely related *Panicum* species. Indeed, a majority (75%) of LGTs in *Panicum hallii* were received from Chloridoideae, while a majority (81%) of those in *Panicum virgatum* were received from Andropogoneae (Table 3). This pattern mirrors the dominant grassland type (Chloridoideae vs Andropogoneae) for a majority of the range of each of the two species, and the area where the individual for the genome assembly was sampled (Lovell *et al.*, 2018; Lehmann *et al.*, 2019).

Quantifying the effects of biogeography as opposed to other factors requires both identifying the donor at the species level and a detailed description of the spatial distribution of each grass species, including their abundances. Indeed, the likelihood of encounters will increase with the number of individuals of the donor species and not just its presence. In addition, the scale of relevant interactions would depend on transfer mechanisms, with pollination-mediated or vector-mediated transfers potentially able to move genes across plants from a given region, while direct transfers between plants (e.g. via inosculation) would only happen among directly adjacent species. Detailed ecological datasets coupled with genomic data for many species are therefore needed to precisely assess the effect of biogeography on LGT dynamics in grasses.

Conclusion

Using stringent phylogenomic filtering, we have shown here that LGT is a widespread process in grasses, and it occurs in wild species as well as in widely cultivated crops (e.g. maize and wheat). LGT does not appear to be restricted to particular functional types, although it seems to increase in rhizomatous species, in which vegetative growth offers extra opportunities for gene transfers into the germline. In addition, we have shown that the amount of successful transfers decreases with phylogenetic distance. This effect of phylogenetic distance might result from increased genomic compatibility among more related groups. Thanks to the rapid accumulation of genome data for various groups of grasses, future studies of LGT will be able to sample densely the diversity of grasses and therefore refine our conclusions. However, with the current data we have shown that LGT occurs in a variety of grasses, highlighting the potential effect of the frequent movement of functional genes between species on the evolution of this critical group of plants.

Acknowledgements

This work was funded by a European Research Council grant (ERC-2014-STG-638333) and a Royal Society Research grant (RGF\EA\180247). P-AC is supported by a Royal Society University Research Fellowship (URF\R\180022). LTD is supported by a Natural Environment Research Council Independent Research Fellowship (NE/T011025/1).

Author contributions

All authors designed the project. SGSH, PR and LTD conducted the analyses. SGSH, PAC and LTD wrote the paper, with the help of PR.

ORCID

Pascal-Antoine Christin https://orcid.org/0000-0001-6292-8734
Luke T. Dunning https://orcid.org/0000-0002-4776-9568

References

Andersson JO. 2005. Lateral gene transfer in eukaryotes. *Cellular and Molecular Life Sciences CMLS* **62**: 1182–1197.

Barrett RD, Schluter D. 2008. Adaptation from standing genetic variation. *Trends in Ecology & Evolution* **23**: 38–44.

Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J *et al.* 2012. Reference genome sequence of the model plant *Setaria. Nature Biotechnology* **30**: 555.

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.

Broad Institute. 2019. Picard toolkit. *GitHub repository.* [WWW document] URL http://broadinstitute.github.io/picard/ [accessed 16 September 2020].

Cannarozzi G, Plaza-Wüthrich S, Esfeld K, Larti S, Wilson YS, Girma D, de Castro E, Chanyalew S, Blösch R, Farinelli L *et al.* 2014. Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (*Eragrostis tef*). *BMC Genomics* **15**: 1–21.

Chang BS, Campbell DL. 2000. Bias in phylogenetic reconstruction of vertebrate rhodopsin sequences. *Molecular Biology and Evolution* **17**: 1220–1231.

Chen F, Dong W, Zhang J, Guo X, Chen J, Wang Z, Lin Z, Tang H, Zhang L. 2018. The sequenced angiosperm genomes and genome databases. *Frontiers in Plant Science* **9**: 418.

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.

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

Christin PA, Besnard G, Edwards EJ, Salamin N. 2012. Effect of genetic convergence on phylogenetic inference. *Molecular Phylogenetics and Evolution* **62**: 921–927.

Christin PA, Edwards EJ, Besnard G, Boxall SF, Gregory R, Kellogg EA, Hartwell J, Osborne CP. 2012. Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. *Current Biology* **22**: 445–449.

Christin PA, Spriggs E, Osborne CP, Strömberg CA, Salamin N, Edwards EJ. 2014. Molecular dating, evolutionary rates, and the age of the grasses. *Systematic Biology* **63**: 153–165.

Clayton WD, Vorontsova MS, Harman KT, Williamson H. 2016. *GrassBase – The online world grass flora.* [WWW document] URL https://www.kew.org/data/grassbase/ [accessed 15 July 2019]

Doolittle WF. 1999. Lateral genomics. *Trends in Biochemical Sciences* **24**: M5–M8.

Dunning LT, Olofsson JK, Parisod C, Choudhury RR, Moreno-Villena JJ, Yang Y, Dinora J, Quick WP, Park M, Bennetzen JL *et al.* 2019. Lateral transfers of large DNA fragments spread functional genes among grasses. *Proceedings of the National Academy of Sciences, USA* **116**: 4416–4425.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves ortholog inference accuracy. *Genome Biology* **16**: 1–14.

Van Etten J, Bhattacharya D. 2020. Horizontal gene transfer in eukaryotes: not if, but how much? *Trends in Genetics* **36**: 915–925.

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1 km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**: 4302–4315.

Freschi L, Vincent AT, Jeukens J, Emond-Rheault JG, Kukavica-Ibrulj I, Dupont MJ, Charette SJ, Boyle B, Levesque RC. 2018. The *Pseudomonas aeruginosa* pan-genome provides new insights on its population structure, horizontal gene transfer, and pathogenicity. *Genome Biology and Evolution* **11**: 109–120.

Gao C, Ren X, Mason AS, Liu H, Xiao M, Li J, Fu D. 2014. Horizontal gene transfer in plants. *Functional & Integrative Genomics* **14**: 23–29.

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H *et al.* 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**: 92–100.

Grass Phylogeny Working Group II. 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New Phytologist* **193**: 304–312.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**: 696–704.

Guo L, Qiu J, Ye C, Jin G, Mao L, Zhang H, Yang X, Peng Q, Wang Y, Jia L *et al.* 2017. *Echinochloa crus-galli* genome analysis provides insight into its adaptation and invasiveness as a weed. *Nature Communications* **8**: 1–10.

Harris IPDJ, Jones PD, Osborn TJ, Lister DH. 2014. Updated high resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. *International Journal of Climatology* **34**: 623–642.

Hertle AP, Haberl B, Bock R. 2021. Horizontal genome transfer by cell-to-cell travel of whole organelles. *Science. Advances* **7**: eabd8215.

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

International Barley Genome Sequencing Consortium. 2012. A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**: 711.

International Brachypodium Initiative. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon. Nature* **463**: 763.

International Wheat Genome Sequencing Consortium. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**: 1251788.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

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

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

Lefort V, Longueville JE, Gascuel O. 2017. SMS: Smart model selection in PhyML. *Molecular Biology and Evolution* **34**: 2422–2424.

Lehmann CE, Griffith DM, Simpson KJ, Anderson TM, Archibald S, Beerling DJ, Bond WJ, Denton E, Edwards EJ, Forrestel EJ *et al.* 2019. Functional diversification enabled grassy biomes to fill global climate space. *BioRxiv.* 583625.

Li FW, Villarreal JC, Kelly S, Rothfels CJ, Melkonian M, Frangedakis E, Ruhsam M, Sigel EM, Der JP, Pittermann J, Burge DO. 2014. Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. *Proceedings of the National Academy of Sciences, USA* **111**: 6672–6677.

Lovell JT, Jenkins J, Lowry DB, Mamidi S, Sreedasyam A, Weng X, Barry K, Bonnette J, Campitelli B, Daum C *et al.* 2018. The genomic landscape of molecular responses to natural drought stress in *Panicum hallii. Nature Communications* **9**: 1–10.

Mahelka V, Krak K, Kopecký D, Fehrer J, Šafář J, Bartoš J, Blattner FR. 2017. Multiple horizontal transfers of nuclear ribosomal genes between phylogenetically distinct grass lineages. *Proceedings of the National Academy of Sciences, USA* **114**: 1726–1731.

Mahelka V, Krak K, Fehrer J, Caklová P, Nagy Nejedlá M, Čegan R, Kopecký D, Šafář J. 2021. A panicum-derived chromosomal segment captured by Hordeum a few million years ago preserves a set of stress-related genes. *The Plant Journal* **105**:1141–1164.

Maumus F, Epert A, Nogué F, Blanc G. 2014. Plant genomes enclose footprints of past infections by giant virus relatives. *Nature Communications* **5**: 1–10.

Moreno-Villena JJ, Dunning LT, Osborne CP, Christin PA. 2017. Highly expressed genes are preferentially co-opted for C_4 photosynthesis. *Molecular Biology and Evolution* **35**: 94–106.

Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299–304.

Olofsson JK, Dunning LT, Lundgren MR, Barton HJ, Thompson J, Cuff N, Ariyarathne M, Yakandawala D, Sotelo G, Zeng K *et al.* 2019. Population-specific selection on standing variation generated by lateral gene transfers in a grass. *Current Biology* **29**: 3921–3927.

Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S. 2013. The caper package: comparative analysis of phylogenetics and evolution in R. *R package version* **5**: 1–36.

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A *et al.* 2009. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**: 551–556.

Peccoud J, Loiseau V, Cordaux R, Gilbert C. 2017. Massive horizontal transfer of transposable elements in insects. *Proceedings of the National Academy of Sciences, USA* **114**: 4721–4726.

Pellicer J, Leitch IJ. 2020. The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. *New Phytologist* **226**: 301–305.

Phansopa C, Dunning LT, Reid JD, Christin PA. 2020. Lateral gene transfer acts as an evolutionary shortcut to efficient C_4 biochemistry. *Molecular Biology and Evolution* **37**: 3094–3104.

Popa O, Dagan T. 2011. Trends and barriers to lateral gene transfer in prokaryotes. *Current Opinion in Microbiology* **14**: 615–623.

Prentice HC, Li Y, Lönn M, Tunlid A, Ghatnekar L. 2015. A horizontally transferred nuclear gene is associated with microhabitat variation in a natural plant population. *Proceedings of the Royal Society B: Biological Sciences* **282**: 20152453.

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

Richardson AO, Palmer JD. 2007. Horizontal gene transfer in plants. *Journal of Experimental Botany* **58**: 1–9.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves T *et al.* 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**: 1112–1115.

Schönknecht G, Weber AP, Lercher MJ. 2014. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. *BioEssays* **36**: 9–20.

Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**: 1246–1247.

Skippington E, Ragan MA. 2012. Phylogeny rather than ecology or lifestyle biases the construction of *Escherichia coli–Shigella* genetic exchange communities. *Open Biology* **2**: 120112.

Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Zuloaga FO, Judziewicz EJ, Filgueiras TS, Davis JI, Morrone O. 2015. A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution* **53**: 117–137.

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

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Stegemann S, Bock R. 2009. Exchange of genetic material between cells in plant tissue grafts. *Science* **324**: 649–651.

Stein JC, Yu Y, Copetti D, Zwickl DJ, Zhang L, Zhang C, Chougule K, Gao D, Iwata A, Goicoechea JL *et al.* 2018. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza. Nature Genetics* **50**: 285–296.

Studer AJ, Schnable JC, Weissmann S, Kolbe AR, McKain MR, Shao Y, Cousins AB, Kellogg EA, Brutnell TP. 2016. The draft genome of the C_3 panicoid grass species *Dichanthelium oligosanthes. Genome Biology* **17**: 1–18.

Tanaka H, Hirakawa H, Kosugi S, Nakayama S, Ono A, Watanabe A, Hashiguchi M, Gondo T, Ishigaki G, Muguerza M *et al.* 2016. Sequencing and comparative analyses of the genomes of zoysiagrasses. *DNA Research* **23**: 171–180.

Tang H, Bomhoff MD, Briones E, Zhang L, Schnable JC, Lyons E. 2015. SynFind: compiling syntenic regions across any set of genomes on demand. *Genome Biology and Evolution* **7**: 3286–3298.

Vallenback P, Jaarola M, Ghatnekar L, Bengtsson BO. 2008. Origin and timing of the horizontal transfer of a *PgiC* gene from *Poa* to *Festuca ovina. Molecular Phylogenetics and Evolution* **46**: 890–896.

VanBuren R, Bryant D, Edger PP, Tang H, Burgess D, Challabathula D, Freeling M. 2015. Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum. Nature* **527**: 508–511.

Varshney RK, Shi C, Thudi M, Mariac C, Wallace J, Qi P, Zhang H, Zhao Y, Wang X, Rathore A *et al.* 2017. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nature Biotechnology* **35**: 969–976.

Vogel A, Schwacke R, Denton AK, Usadel B, Hollmann J, Fischer K, Bolger A, Schmidt MHW, Bolger ME, Gundlach H *et al.* 2018. Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris. Nature Communications* **9**: 1–11.

Watcharamongkol T, Christin PA, Osborne CP. 2018. C_4 photosynthesis evolved in warm climates but promoted migration to cooler ones. *Ecology Letters* **21**: 376–383.

Wickell DA, Li F-W. 2019. On the evolutionary significance of horizontal gene transfer in plants. *New Phytologist* **225**: 113–117.

Xi Z, Bradley RK, Wurdack KJ, Wong KM, Sugumaran M, Bomblies K, Rest JS, Davis CC. 2012. Horizontal transfer of expressed genes in a parasitic flowering plant. *BMC Genomics* **13**: 1–8.

Yang Z, Wafula EK, Kim G, Shahid S, McNeal JR, Ralph PE, Timilsena PR, Yu W, Kelly EA, Zhang H *et al.* 2019. Convergent horizontal gene transfer and cross-talk of mobile nucleic acids in parasitic plants. *Nature Plants* **5**: 991–1001.

Yoshida S, Maruyama S, Nozaki H, Shirasu K. 2010. Horizontal gene transfer by the parasitic plant *Striga hermonthica*. *Science* **328**: 1128.

Yu L, Boström C, Franzenburg S, Bayer T, Dagan T, Reusch TB. 2020. Somatic genetic drift and multilevel selection in a clonal seagrass. *Nature Ecology & Evolution* 1–11.

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**: 1–9.

Zhang J, Fu XX, Li RQ, Zhao X, Liu Y, Li MH, Zwaenepoel A, Ma H, Goffinet B, Guan YL *et al.* 2020. The hornwort genome and early land plant evolution. *Nature plants* **6**: 107–118.

Supporting Information

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

Dataset S1 Nucleotide alignments.

Dataset S2 LGT maximum-likelihood trees, 85 taxa.

Dataset S3 LGT maximum-likelihood trees, 85 taxa third codon position.

Dataset S4 LGT maximum-likelihood trees, including short-read data.

Fig. S1 Impact of varying filtering parameters on LGT detection.

Fig. S2 Coverage plots comparing independent sequencing runs.

Fig. S3 Details of a laterally acquired DNA fragment in the *Setaria italica* genome.

Table S1 List of data sets used in different steps.

Table S2 Results of the analysis of 17 reference genomes.

Table S3 Results of the synteny analyses.

Table S4 Comparisons of the LGT detected in *Alloteropsis semialata* with a previous study.

Table S5 Results of the analysis of LGT among Paniceae species.

Table S6 Coverage analyses of genes used to compare independently produced sequencing runs.

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