Phylogenetic relatedness, co-occurrence, and rhizomes increase lateral gene transfer among grasses

Samuel G. S. Hibdige[1], Pauline Raimondeau[1,2], Pascal-Antoine Christin[1], Luke T. Dunning[1*]

Affiliations:

[1] Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom

[2] Present address: Laboratoire Évolution & Diversité Biologique (EDB UMR5174), CNRS, Institut de Recherche pour le Développement, F-31062 Toulouse, France

[*]**Corresponding author:** Luke T. Dunning; Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom; +44 (0)1142220027; l.dunning@sheffield.ac.uk funnings@hotmail.com

Abstract

Background:

Lateral gene transfer (LGT) has been documented in a broad range of eukaryotes, where it can promote adaptation. In plants, LGT of functional nuclear genes has been repeatedly reported in parasitic plants, ferns and grasses, but the exact extent of the phenomenon remains unknown. Systematic studies are now needed to identify the factors that govern the frequency of LGT among plants.

Results:

Here we scan the genomes of a diverse set of grass species that span more than 50 million years of divergence and include major crops. We identify protein-coding LGT in a majority of them (13 out of 17). There is variation among species in the amount of LGT received. Rhizomatous species received 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 amongst close relatives facilitating successful transfers. However, we also observe genetic exchanges among distantly related species that diverged shortly after the origin of the grass family when they co-occur in the wild, pointing to an additional role of biogeography. The dynamics of successful LGT in grasses therefore appear to be dependent on both opportunity (co-occurrence and rhizomes) and compatibility (phylogenetic distance).

Conclusion:

Overall, we show that LGT is a widespread phenomenon in grasses, which is boosted by repeated contact among related lineages. The process has moved functional genes across the entire grass family into domesticated and wild species alike.

Keywords: adaptation, evolution, genomics, horizontal gene transfer, phylogenomics, Poaceae.

1. Background

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 facilitating the exchange of 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 being 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 a variety of multicellular eukaryotes (reviewed in: Anderson 2005; Keeling and Palmer, 2008; Schönknecht et al. 2014; Husnik et al. 2018), including plants (reviewed in: Richardson and Palmer 2007; Gao et al. 2014; Wickell and Li 2019).

A majority of nuclear LGT among plants 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, LGT is not restricted to parasitic interactions, and it 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; Dunning et al. 2019). Grasses represent the best system to investigate factors promoting LGT in non-parasitic plants as they are the only group where multiple LGT recipients have been identified, and there is extensive genomic resources available due to 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). 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). While reports of LGT in other species in the group suggest it is a widespread phenomenon, its full distribution within the family remains to be assessed.

Grasses are very diverse (Soreng et al. 2015), and the more than 12,000 species exhibit extensive phenotypic variation that may influence LGT dynamics. In particular, the family contains both annuals and perennials. If LGT happens during vegetative growth (e.g. root-to-root inosculation), the number of LGT is predicted to be higher in perennial and rhizomatous species. Conversely, if LGT happens through illegitimate pollination, the number of LGT is not expected to vary with growth form. In addition, grasses are variable in both their genome size and ploidy levels (Caetano-Anollés 2005), and genome complexity coupled with high activity of transposable elements might facilitate the integration of foreign DNA. Finally, successful transfers might be more likely to occur between closely-related groups with similar genome features as observed in prokaryotes (Skippington and Ragan 2012; Soucy et al. 2015). Most of the grass diversity is clustered in the two BOP and PACMAD sister groups that diverged more than 50 million years ago (Christin et al. 2014). Each of the two groups has more than 5,000 taxa and includes model species with complete genomes (Soreng et al. 2015). The family therefore offers unparalleled opportunities to determine whether functional characteristics or phylogenetic distance determine the amount of LGT among non-parasitic 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 clade (Oryzoideae and Pooideae) and three from the PACMAD clade (Andropogoneae, Chloridoideae, and Paniceae). Together, these five groups capture more than 8,000 species or over

70% of the diversity within the whole family (Soreng et al. 2015). In our sampling, each of these five groups is represented by at least two species, allowing us to monitor the number of transfers among each group. In addition, the species represent a variety of domestication status, life-history strategies, genome sizes, and ploidy levels (**Table 1**). Using this sampling design, we (i) test whether LGT is more common in certain taxonomic groups, and (ii) test whether some plant characters are associated with a statistical increase of LGT. We then focus on the donors of the LGT received by the Paniceae tribe, a group for which seven genomes are available, to (iii) test whether the number of LGT increases with phylogenetic relatedness. Our work represents the first systematic quantification of LGT among members of a large group of plants and sheds new light on the conditions that promote genetic exchanges across species boundaries in plants.

2. Results

2.1 No unambiguous LGT from non-grass donors

We first screened 817,612 genes from the 17 grass genomes (**Table 1**) for potential LGT of non-grass origin based on sequence similarity. BLASTn was used to identify the top-hit match of each grass gene to a database containing grass and non-grass protein-coding sequences (including data from non-grass Poales; **Table S1**). In total, 481 grass genes had a top-hit match to a non-grass. To determine whether the evolutionary history of these genes was compatible with a LGT scenario, we inferred phylogenetic trees using homologous sequences from the species present in the initial blast database (17 grasses and 37 non-grasses). For 150 of these genes, the low number of species for which homologs were detected (less than six species per alignment) prevented meaningful phylogenetic analyses. Maximum-likelihood phylogenetic trees were generated for the remaining 331 candidates.

Out of 331 candidates, 135 produced a topology that was not suggestive of LGT. These were considered as false positives and discarded. For 163 of the remaining 196 candidates, the gene tree topology suggested a transfer from a grass to a non-grass species. All these phylogenetic patterns resulted from the presence of near complete grass chloroplast genomes in two eudicot genome assemblies: (i) *Aegilops tauschii* chloroplast sequence data in the *Brassica rapa IVFCAASv* assembly; and (ii) *Hordeum vulgare* chloroplast sequence database in the *Medicago truncatula* MedtrA17_4.0 assembly. In both cases, the grass plastomes are present as discrete scaffolds not incorporated into chromosomes, and it remains to be verified whether this foreign DNA indeed exists in the eudicot genomes or represents contamination.

The remaining 33 genes produced phylogenetic trees where the grouping of grass and non-grass genes was consistent with non-grass to grass LGT, although 11 of these lacked statistical support (<50% bootstrap support). The 22 supported candidates were from two sources: (i) chloroplast

genes from the peanut *Arachis hypogaea* present as discrete unassembled scaffolds in the *Cenchrus americanus* genome (n=2); and (ii) chloroplast (n=7) and nuclear (n=13) genes of Fagales origin on chromosome 5 of *Hordeum vulgare*. We checked the Illumina data used to assemble the *Hordeum vulgare* genome in combination with BAC-end sequencing and the existence of none of the genes of Fagales origin was supported by more than one sequencing batch (**Table S2**). We therefore conclude that, while these genes should be investigated further with more sequencing, there is currently no unambiguous evidence for LGT from non-grass species to grasses.

2.2 LGT occurs in all lineages of grasses, and especially in rhizomatous species

We considered LGT among five groups of grasses that were monophyletic and supported by a majority of gene trees in previous multigene coalescence analyses (Dunning et al. 2019). These clades diverged more than 20 million years ago (Christin et al. 2012a, 2014), and there are low risks of widespread incomplete lineage sorting or hybridization generating inconsistency between the gene and species trees. Each of these clades is represented by at least two genomes that capture early splits within them, allowing the identification of genes from other species nested within each group. All 17 species with complete genomes were successively considered as potential recipients of LGT from any of the four other clades (**Fig. 1**). To identify all LGT, we adapted a previously developed pipeline (Dunning et al. 2019), which involves two phylogenetic filtering steps using transcriptome and genome data from up to 85 grass species (**Table S1**). In brief, to be considered as an LGT each gene had: (i) to be nested within one of the other four groups of grasses (**Fig. 1**); (ii) their nesting had to be well supported (≥ 70% bootstrap support); (iii) potential parology problems had to be ruled out (i.e. discarding phylogenies with multiple apparent duplication events that can explain the phylogenetic incongruence); (iv) the nesting had to be supported by phylogenetic trees constructed solely from the 3rd codon positions, which are less subject to adaptive convergent evolution; and (v) where possible, the 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 LGT).

Out of the 817,612 grass genes screened, 55.89% had sufficient homologous grass sequences (≥ 10 taxa) for reliable phylogenetic reconstruction (**Table 2** & **Table S3**), and were tested for LGT (the numbers of genes tested are used when correcting for the number of genes in statistical analyses). 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 grass genomes in this initial analysis (**Table 2**; full results **Table S3**). For the majority of these LGT (74%; **Table S3**), the recipient genome also contained a native copy, which further argues against incomplete lineage sorting. We also conducted synteny analysis using seven diploid species, with at least one representative from each of the five clades, to rule out unrecognised parology problems in generating the observed phylogenetic patterns. For each LGT detected in these seven species we determined whether the genes from the other six species identified as orthologous in the phylogenetic tree were syntenic to the LGT or the native gene. In no instance was a gene from another model species syntenic with the LGT, with 74.6% being syntenic with the native copy, and 25.4% being syntenic to neither (**Table S4**). The synteny analyses therefore confirm that our phylogenetic trees identify true orthologs, and the patterns compatible with LGT cannot be explained by unrecognised paralogy.

Mapping of short-read data in multiple cases confirmed a high similarity between the putative donor and recipient on intron sequences in addition to exons (**Fig. 2**). However, because either the transfers are too ancient or because we lack whole genome data very close relatives of the donors, it is not possible to delimit with high precision the laterally acquired fragments detected here. Despite this, we still see intergenic regions with high similarity (**Fig. S1**), which together with intronic similarities rules out convergent evolution or other phylogenetic biases generating the gene/species tree discordance.

The number of LGT received varied among species (p-value < 0.01; Chi-square test; mean = 8.4; SD=9.0; range=0 − 34; **Table S3**), with the highest numbers observed in *Panicum virgatum* (n= 30), *Alloteropsis semialata* (n=20), and *Cenchrus americanus* (n=15). Only a subset of the 59 previously reported LGT in *Alloteropsis semialata* (Dunning et al. 2019) are recorded as the previous analysis examined additional groups of donors not considered here and used mapping of short reads to identify candidates flanking LGT detected as described here (see below). Despite the significant variation among species, the difference among the five groups was not significant when considering all LGT (p-value = 0.16, Kruskal-Wallis test). The test repeated solely on LGT for which AU tests could be performed (i.e. native copies of the recipient were detected – see Methods) indicated marginally significant variation in the amount of LGT received among groups (p-value = 0.03, Kruskal-Wallis test). Overall, our results show that LGT is widespread across the grass family and occurs in a majority of species (**Fig. 1; Table 2**). No LGT were detected in four of the 17 species considered here, but some LGT might remain undetected due to our stringent phylogenetic filtering, and because we are only considering transfers among the predefined five grass clades.

Among the 17 species screened, LGT is observed in all functional groups (**Fig. 3**). We detected LGT in wild species, but also in major crops. For instance, maize (*Zea mays*) has 11 LGT received from Chloridoideae and Paniceae, while wheat (*Triticum aestivum*) has ten LGT received from Andropogoneae, Chloridoideae and Paniceae (**Table 2**). The LGTs may be beneficial for the crops, with transferred loci including those with functions related to abiotic stress tolerance and disease resistance (**Table S3**). In the other comparisons, LGT seem to be more abundant in perennial, rhzomatous and C_4 species (**Fig. 3**). When accounting for the phylogenetic structure using phylogenetic generalized least squares (PGLS), three plant characteristics jointly increased the number of LGT: the number of genes tested ($p < 0.01$), the presence of rhizomes ($p = < 0.01$), and low ploidy levels ($p < 0.01$; combined adjusted $R^2 = 0.652$).

2.3 LGT are more commonly received from closely related species

Overall, some clades acted more frequently as donors (p-value < 0.01, Kruskal-Wallis test). Specifically, the Andropogoneae were the source of most transfers (**Table 2**). However, these were mainly received by members of Paniceae, which are the closest relatives of Andropogoneae in our dataset and are also represented by the most genomes (**Table 1**). While these patterns suggest that LGT occur more frequently among close relatives, directly comparing the rates is made difficult because the clades vary in their number of species, number of genomes available and age. However, for a given clade of recipients, it is possible to compare the frequency of different groups of donors. We therefore focused on the identity of donors of LGT to Paniceae, the group with the highest number of complete genomes.

Seven Paniceae genomes were used in this study, and this increased sample size further allows to detect intra-Paniceae LGT. We therefore reported the number of LGT transferred from the Panicinae and Cenchrinae subtribes (each represented by two genomes) to other Paniceae, in addition to those received from other groups. In total, we identify 129 LGT in the Paniceae, 35 of which were transferred from the Cenchrinae and Panicinae subtribes (**Table 3**; full results **Table S5**). When focusing on Paniceae recipients, some groups are 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 LGT given per species decreases with the phylogenetic distance to Paniceae, reaching lowest levels in the BOP clade (Pooideae and Oryzoideae; **Fig. 4**).

3. Discussion

Lateral gene transfer is a potent evolutionary force capable of having a profound impact on the evolutionary trajectory of a species. Here, we use grasses as a model system to investigate the factors that dictate the prevalence of LGT among plants. Using a combination of stringent phylogenetic and genomic analyses, we have identified a grand total of 170 genes (approximately 3.72 LGT per 10,000 genes; 135 in the first round of analyses and 35 among groups of Paniceae) that have been laterally transferred to 13 of the 17 complete grass genomes that were screened. Our analytical pipeline was developed to drastically reduce the amount of false positives, and is purposely very conservative. This enables us to minimise the effects of other evolutionary processes such as hybridisation and incomplete lineage sorting generating false positives. As a result the number of LGT identified is likely only a subset of those existing in the complete grass genomes. In addition, the phylogenetic filtering prevents us from detecting LGT from clades of grasses for which no genome is available. With the current sampling, approximately 30% of the grass diversity is not considered as potential LGT donors (Soreng et al. 2015).

Our phylogenetic approach prevents us from detecting LGT happening among members of the same group of grasses, such as the numerous exchanges among lineages of Paniceae previously detected (Dunning et al. 2019). This is perfectly exemplified by the case of *A. semialata*, in which 59 LGT were previously detected (Dunning et al. 2019). Here, only 20 were identified when considering solely LGT among the five higher groups, while a further 13 were detected when considering subgroups of Paniceae as potential donors. Improvements in the initial screening approach (i.e. removing an initial read-mapping step from Dunning et. al. 2019 method) resulted in the identification of four novel LGT. The remaining LGTs not detected in this study were a result of slight methodological differences (e.g. less donor clades considered and sampling more genomes means greater multiple testing correction), and screening the flanking regions of LGTs for additional genes with high sequence similarity was not preformed due to the larger scale of the

present analyses and varying genome assembly contiguity. The inability to detect LGT identified with more detailed analyses strongly suggest that the LGTs we report here concern only a small fraction of those existing in grass genomes. Finally, our pipeline precludes the detection of older LGT that are shared by multiple individuals within a clade. Despite these limitations we show that LGT is common in grasses, certain groups exchange more genes than others, the frequency of LGT increases in rhizomatous species, and there appears to be a role of phylogenetic distance underpinning the LGT dynamics.

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

LGT is common in grasses and is observed in each of the five groups investigated here. We detected LGT in domesticated and wild species alike, although it is currently unknown whether the LGTs occurred before or after domestication and whether these genes are associated with agronomic traits. The genetic exchanges are not restricted to any functional category of grasses (**Fig. 3**), and the ubiquity of the phenomenon provides some support for a breakdown in reproductive behaviour and illegitimate pollination as the mechanism responsible for the transfers as wind pollination is universal in this group. However, there is a statistical increase of the number of LGT in rhizomatous species and two of the three species with the highest numbers of LGT (*Alloteropsis semialata* and *Panicum virgatum*) are perennials that can propagate vegetatively via rhizomes. These patterns suggest that root-to-rhizome contacts (i.e. inosculation) provide increased opportunities for gene transfers directly into the germline. The genetic bottleneck and selection characterising rhizomes would increase the chance of LGT retention, especially if these provide an advantage (Yu et al. 2020). However, we did not detect LGT in the third rhizomatous species we sampled (*Zoysia japonica*; **Table 1**). Increased species sampling is now needed to precisely quantify the impact of the growth form on the amount of gene transfers and how it interacts with other factors.

3.2 It is easier to acquire genes from close relatives

Previously, LGT has been reported among very distant groups of plants (Li et al. 2014). In the case of grasses, our analyses did not detect any unambiguous case of a LGT from a non-grass donor (**Table 1**). This may indicate that the phylogenetic distance is too great and the genome features too different for successful gene transfers, or that the mechanism of transfer is grass specific. Within grasses, there is an effect of the phylogenetic distance on the number of transfers observed, as shown by the Paniceae receiving more LGT from closer relatives (**Fig. 4**). This pattern mirrors that observed in prokaryotes (Popa and Dagan 2011; Skippington and Ragan 2012; Soucy et al. 2015) and insects (Peccoud et al. 2017), where the frequency of transfers is higher between closely related species . In prokaryotes, this effect is thought to result from more similar DNA sequence promoting homologous replacement of the native copy (Skippington and Ragan 2012). This is unlikely to play a role in grasses as the LGTs are inserted in non-syntenic positions in the genome where they coexist with the native copy, often on different chromosomes (Dunning et al. 2019; **Table S4**). However, stretches of DNA similar between the donor and recipient (e.g. transposable elements) may still be involved in the incorporation of the LGT onto the chromosomes. Alternatively, the effect of the phylogenetic distance might stem from the regulation of the LGT post 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 the transfer rather than the rate of transfer. Overall, our analyses indicate that it is easier to either obtain LGT from close relatives or to use it after the transfers, thereby increasing the chance of selectively retaining it.

3.3 Grasses steal genes from their neighbours.

In addition to genomic compatibility, the probability of co-occurrence in the wild might decrease with divergence time, explaining the observed effect of phylogenetic distance (**Fig. 4**). Paniceae grasses generally grow in biodiverse savannas which are dominated by Andropogoneae species in

wetter areas, and Chloridoideae grasses in drier habitats (Lehmann et al. 2019). The effects of phylogenetic distance are therefore confounded with those of biogeography, but specific examples indicate that biogeography can take precedence.

We observe some transfers between Pooideae and Paniceae, two groups that diverged >50 Ma, representing one of the earliest splits within this family. This indicates that LGT are possible across the whole grass family. In our dataset, the only recipient of these transfers is *Dichanthelium oligosanthes*, a frost-tolerant grass from North America, which inhabits colder areas than other members of the Paniceae (Studer at al. 2016). In cold regions, *D. oligosanthes* co-occurs with members of the Pooideae, and this biogeographic pattern likely facilitated exchanges between the two groups of grasses.

Biogeography might also be responsible for differences in the identity of the LGT donors between the two closely related *Panicum* species. Indeed, a majority (75%) of LGT in *Panicum hallii* were received from Chloridoideae, while a majority (81%) of those in *Panicum virgatum* were received from Andropogoneae. 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 from which 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 a detailed description of the spatial distribution of each grass species, but also their abundance. Indeed, the likelihood of encounters will increase with the number of individuals of the donor species and not just its presence. Such ecological datasets coupled with genomic data sets for large numbers of grasses will be key to future studies of LGT dynamics in this group of plants.

4. Conclusions

Using stringent phylogenomic filtering, we have shown that lateral gene transfer (LGT) is a widespread evolutionary process in grasses, where it occurs in wild species as well as in widely cultivated crops (e.g. maize and wheat). LGT does not appear restricted to particular functional types, although it increases in rhizomatous species, where vegetative growth offers extra opportunities for gene transfers into the germline. In addition, we show that the amount of successful transfers decreases with phylogenetic distance. This effect of the phylogenetic distance might result from increased genomic compatibility among more related groups. Alternatively, groups that diverged more recently might be more likely to co-occur, offering more opportunities for genetic exchanges. Indeed, biogeography seems to have an overall effect on the frequency of LGT, as the only species from distant groups that exchanged genes are those that co-occur following recent niche shifts. Overall, our study shows that LGT occurs in a variety of grasses, and the frequent movements of functional genes can strongly impact the evolution of this group of plants.

5. Methods

5.1 Searching for non-grass LGT

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). We initially screened for non-grass LGT by identifying grass genes that had a top-hit BLASTn match (min. length 300 bp) to a non-grass species. The grasses in the BLAST database comprised the 17 species described above, and the non-grass species included coding sequences from 25 genomes extracted from the Ensembl Plants database (release 37), supplemented with data for 12 non-grass monocots, including three genomes [*Ananas comosus* - Poales (Ming et al. 2015), *Elaeis guineensis* – Arecales (Singh et al. 2013) and *Phalaenopsis equestris* – Asparagales (Cai et al. 2015)] and nine transcriptomes (all Poales; McKain et al. 2016; **Table S1**). To verify genes were of non-grass origin, we then generated phylogenetic trees for each gene with a top-hit BLASTn match to a non-grass species. For genes with matches in at least five other species (six species including the query), nucleotide alignments were generated by aligning the BLASTn matching regions to the query sequence using the 'add fragments' parameter in MAAFT v7.427 (Katoh and 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. Maximum likelihood phylogenetic trees with 100 bootstrap replicates were then inferred using PhyML v.21031022 (Guindon and Gascuel 2003), with the best substitution model identified by Smart Model Selection SMS v.1.8.1 (Lefort et al. 2017). Phylogenetic trees were visually inspected, and those genes of non-grass species nested within grasses with >50% bootstrap support were considered as LGT. To identify the potential donor of any candidate non-grass LGT, BLASTn searches were also preformed against the NCBI non-redundant nucleotide database to identify the most similar sequences in the database.

5.2 Detecting grass-to-grass LGT

We modified the approach previously used by Dunning et al. (2019) to identify grass-to-grass LGT.

Specifically, we did not use the initial mapping filtering step from Dunning et al. (2019), as it relied on the availability of high-coverage genome data for pairs of closely related species. Instead, our updated approach uses two stringent phylogenetic filtering steps to identify unambiguous LGT.

In the first step, 37-taxa trees are constructed using data from the 17 grass genomes described above, supplemented with transcriptome data for 20 additional species from across the grass family (Moreno-Villena et al. 2018). BLASTn was used to identify the top-match for each gene in the 36 other species with a minimum match length of 300bp (not necessarily a single continuous BLAST match). Each set of genes was aligned as described above using the 'add fragments' parameter in MAAFT v7.427. Alignments with less than ten species were considered non informative and consequently discarded (retained 55.9% of genes; total = 457,003 genes; mean per species 26,883 genes; SD = 13,042 genes; **Table S3**). For each alignment with ten species or more, a maximum-likelihood phylogenetic tree was inferred using PhyML v.20120412 with the GTR+G+I substitution model. Each topology was then mid-point rooted using the phytools package in R and perl scripts 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) that were supported by most gene trees in a previous mulitgene coalescent species tree analysis (**Fig. 1**; Dunning et al. 2019). The whole set of analyses were later repeated to detect LGT between well supported subclades within the Paniceae, considering LGT received from two clades represented by at least two genomes and supported by most gene trees in previous analyses (i.e. Cenchrinae and Panicinae, **Fig. 1**). To be considered as nested, the sister species of the query gene, and their combined sister group, had to belong to the same grass group. For genes that were nested, the analysis was repeated with 100 bootstrap replicates to verify that the nesting of the query sequence was supported by a minimum of 50% of the bootstrap replicates.

The second filtering step was applied to those candidates that passed the first phylogenetic filter. For those, 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 >300bp (not necessarily a single continuous blast match) in each of the 105 datasets. Alignments were generated as previously, before being re-aligned as codons using MAAFT 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 by Smart Model Selection SMS v.1.8.1. The trees were manually inspected and discarded if: i) there were less than three species within or outside the LGT donor clade; ii) the LGT candidate was not nested within another group of grasses with the increased taxon sampling; or iii) the tree had obvious paralogy problems due to gene duplication events. For retained candidates, we removed paralogs 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. The tree inference was then repeated with 100 bootstraps, and the trees were again manually inspected, discarding candidates where the nesting was supported by <70% bootstrap replicates. Finally, BLASTx was used to annotate the LGT candidates against the SwissProt database.

5.3 Secondary analyses of candidate LGT

After these two successive filters, retained candidates were subjected to extra validation. To verify the nesting of the candidate LGT was not due to convergent adaptive amino acid substitutions, we generated phylogenetic trees based solely on the 3rd 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 as described 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 and Hasegawa 2001) using the site-wise likelihood values generated by PhyML, and p-values were Bonferroni corrected to account for multiple testing. 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 the second validation, 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 sp.; **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.

5.4 Synteny analyses

Synteny analyses were performed for a subset of seven diploid grass genomes (*Brachypodium distachyon*, *Oryza sativa*, *Oropetium thomaeum*, *Panicum hallii*, *Setaria italica*, *Sorghum bicolor*, and *Zea mays*) using SynFind (Tang et al. 2015). For each LGT in these species, we determined whether genes from the other reference genomes identified as orthologs to the native copy in the phylogenetic trees were syntenic to the LGT or the native copy based on the highest syntelog score (**Table S4**).

5.5 Grass traits

Plant traits were obtained from a variety of 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 (Bennett and Leitch 2005), and climatic

information from Watcharamongkol et al. (2018). The climate data for *Oropetium thomaeum* was not included in Watcharamongkol et al. (2018), and was therefore retrieved from GBIF (GBIF.org (11th July 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.wyhtoo) and WorldClim (Harris et al. 2014; Fick and 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**). The Kruskal-Wallis tests performed used absolute LGT numbers, which were divided into donor groups when determining if some clades were more frequent donors than others. To determine if any trait or genome feature was associated with the numbers of LGT, we preformed phylogenetic generalized least squares (PGLS) to account for the relatedness between samples. The 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**). Iterative models were performed, removing the least significant variable until only significant variables remained (p-value <0.05).

6. Availability of data and materials

All analyses were preformed using publicly available data (**Table S1**). Phylogenetic trees and alignments are included as supplementary datasets.

7. Funding

This work was funded by an European Research Council Grant (ERC-2014-STG-638333) and a Royal Society Research Grant (RGF\EA\180247). P.-A.C. was supported by a Royal Society University Research Fellowship (URF\R\180022).

8. Authors' contributions

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

10. References

Andersson JO. Lateral gene transfer in eukaryotes. Cellular and Molecular Life Sciences CMLS. 2005;62(11):1182-1197.

Barrett RD, Schluter D. Adaptation from standing genetic variation. Trends in Ecology & Evolution. 2008;23(1):38-44.

Bennett MD, Leitch IJ. Plant DNA C-values database. 2005.

Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, ... Jenkins J. Reference genome sequence of the model plant *Setaria*. Nature Biotechnology. 2012;30(6):555.

Caetano-Anollés G. Evolution of genome size in the grasses. Crop Science. 2005;45(5):1809-1816.

Cai J, Liu X, Vanneste K, Proost S, Tsai WC, Liu KW, ... Zheng Z. The genome sequence of the orchid *Phalaenopsis equestris*. Nature Genetics. 2015;47(1):65.

Cannarozzi G, Plaza-Wüthrich S, Esfeld K, Larti S, Wilson YS, Girma D, ... Lyons E. Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (*Eragrostis tef*). BMC Genomics. 2014;15(1):581.

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

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

Christin PA, Edwards EJ, Besnard G, Boxall SF, Gregory R, Kellogg EA, ... Osborne CP. Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. Current Biology. 2012;22(5):445-449.

Christin PA, Spriggs E, Osborne CP, Strömberg CA, Salamin N, Edwards EJ. Molecular dating, evolutionary rates, and the age of the grasses. Systematic Biology. 2014;63(2):153-165.

Clayton WD, Vorontsova MS, Harman KT, Williamson H. GrassBase - The Online World Grass Flora. 2016.

Doolittle WF. Lateral genomics. Trends in Biochemical Sciences. 1999;24(12):M5-M8.

Dunning LT, Olofsson JK, Parisod C, Choudhury RR, Moreno-Villena JJ, Yang Y, ... Christin PA. Lateral transfers of large DNA fragments spread functional genes among grasses. Proceedings of the National Academy of Sciences. 2019;116(10):4416-4425.

Fick SE, Hijmans RJ. WorldClim 2: new 1□km spatial resolution climate surfaces for global land areas. International journal of climatology. 2017;*37*(12):4302-4315.

Freschi L, Vincent AT, Jeukens J, Emond-Rheault JG, Kukavica-Ibrulj I, Dupont MJ, ... Levesque RC. The *Pseudomonas aeruginosa* Pan-Genome Provides New Insights on Its Population Structure, Horizontal Gene Transfer, and Pathogenicity. Genome Biology and Evolution. 2018;11(1):109-120.

Gao C, Ren X, Mason AS, Liu H, Xiao M, Li J, Fu D. Horizontal gene transfer in plants. Functional & Integrative Genomics. 2014;14(1):23-29.

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, ... Hadley D. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science. 2002;296(5565);92-100.

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

Guo L, Qiu J, Ye C, Jin G, Mao L, Zhang H, ... Lin Z. *Echinochloa crus-galli* genome analysis provides insight into its adaptation and invasiveness as a weed. Nature Communications. 2017;8(1):1031.

Harris IPDJ, Jones PD, Osborn TJ, Lister DH. Updated high resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. *International journal of climatology*. 2014;*34*(3):623-642.

Husnik F, McCutcheon JP. Functional horizontal gene transfer from bacteria to eukaryotes. Nature Reviews Microbiology. 2018;16(2):67.

International Barley Genome Sequencing Consortium. A physical, genetic and functional sequence assembly of the barley genome. Nature. 2012;491(7426):711.

International Brachypodium Initiative. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature. 2010;463(7282):763.

International Wheat Genome Sequencing Consortium. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science. 2014;345(6194):1251788.

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

Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. Nature Reviews Genetics. 2008;9(8):605.

Kellogg EA, Buell CR. Splendor in the grasses. Plant physiology. 2009;149(1):1-3.

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

Lehmann CE, Griffith DM, Simpson KJ, Anderson TM, Archibald S, Beerling DJ, ... Fox DL. Functional diversification enabled grassy biomes to fill global climate space. BioRxiv. 2019;583625.

Li FW, Villarreal JC, Kelly S, Rothfels CJ, Melkonian M, Frangedakis E, ... Burge DO. Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. Proceedings of the National Academy of Sciences. 2014;111(18):6672-6677.

Lovell JT, Jenkins J, Lowry DB, Mamidi S, Sreedasyam A, Weng X, ... Gordon SP. The genomic landscape of molecular responses to natural drought stress in *Panicum hallii*. Nature Communications. 2018;9(1):5213.

Mahelka V, Krak K, Kopecký D, Fehrer J, Šafář J, Bartoš J, Blattner FR. Multiple horizontal transfers of nuclear ribosomal genes between phylogenetically distinct grass lineages. Proceedings of the National Academy of Sciences. 2017;114(7):1726-1731.

McKain MR, Tang H, McNeal JR, Ayyampalayam S, Davis JI, dePamphilis CW, ... Leebens-Mack JH. A phylogenomic assessment of ancient polyploidy and genome evolution across the Poales. Genome Biology and Evolution. 2016;8(4):1150-1164.

Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, ... Zhang J. The pineapple genome and the evolution of CAM photosynthesis. Nature Genetics. 2015;47(12):1435.

Moreno-Villena JJ, Dunning LT, Osborne CP, Christin PA. Highly expressed genes are preferentially co-opted for C_4 photosynthesis. Molecular Biology and Evolution. 2017;35(1):94-106.

Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature. 2000;405(6784):299.

Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S. The caper package: comparative analysis of phylogenetics and evolution in R. R package version. 2013;5(2):1-36.

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, ... Schmutz J. The *Sorghum bicolor* genome and the diversification of grasses. Nature. 2009;457(7229):551.

Peccoud J, Loiseau V, Cordaux R, & Gilbert C. Massive horizontal transfer of transposable elements in insects. Proceedings of the National Academy of Sciences, 2017;114(18):4721-4726.

Popa O & Dagan T. Trends and barriers to lateral gene transfer in prokaryotes. Current opinion in microbiology, 2011;14(5):615-623.

Richardson AO, Palmer JD. Horizontal gene transfer in plants. Journal of Experimental Botany, 2007;58(1):1-9.

Prentice HC, Li Y, Lönn M, Tunlid A, Ghatnekar L. A horizontally transferred nuclear gene is associated with microhabitat variation in a natural plant population. Proceedings of the Royal Society B: Biological Sciences. 2015;282(1821):20152453.

Schönknecht G, Weber AP, Lercher MJ. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. Bioessays. 2014;36(1):9-20.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, ... Minx P. The B73 maize genome: complexity, diversity, and dynamics. Science. 2009;326(5956):1112-1115.

Singh R, Ong-Abdullah M, Low ETL, Manaf MAA, Rosli R, Nookiah R, ... Azizi N. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. Nature. 2013;500(7462):335.

Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Zuloaga FO, Judziewicz EJ, ... Morrone O. A worldwide phylogenetic classification of the Poaceae (Gramineae). Journal of Systematics and Evolution. 2015;53(2):117-137.

Soucy SM, Huang J, Gogarten JP. Horizontal gene transfer: building the web of life. Nature Reviews Genetics. 2015;16(8):472.

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

Skippington E, Ragan MA. Phylogeny rather than ecology or lifestyle biases the construction of *Escherichia coli–Shigella* genetic exchange communities. Open Biology. 2012;2(9):120112.

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-1313.

Stein JC, Yu Y, Copetti D, Zwickl DJ, Zhang L, Zhang C, ... Wei S. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. Nature Genetics. 2018;50(2):285.

Studer AJ, Schnable JC, Weissmann S, Kolbe AR, McKain MR, Shao Y, ... Brutnell TP. The draft genome of the C_3 panicoid grass species *Dichanthelium oligosanthes*. Genome Biology. 2016;17(1):223.

Tanaka H, Hirakawa H, Kosugi S, Nakayama S, Ono A, Watanabe A, ... Shimizu K. Sequencing and comparative analyses of the genomes of zoysiagrasses. DNA Research. 2016;23(2):171-180.

Tang H, Bomhoff MD, Briones E, Zhang L, Schnable JC, Lyons E. SynFind: compiling syntenic regions across any set of genomes on demand. Genome Biology and Evolution. 2015;7(12):3286-3298.

Vallenback P, Jaarola M, Ghatnekar L, Bengtsson BO. Origin and timing of the horizontal transfer of a *PgiC* gene from *Poa* to *Festuca ovina*. Molecular Phylogenetics and Evolution. 2008;46(3):890-896.

VanBuren R, Bryant D, Edger PP, Tang H, Burgess D, Challabathula D, Freeling M. Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. Nature. 2015;527(7579):508.

Varshney RK, Shi C, Thudi M, Mariac C, Wallace J, Qi P, ... Srivastava RK. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. Nature Biotechnology. 2017;35(10):969.

Vogel A, Schwacke R, Denton AK, Usadel B, Hollmann J, Fischer K, ... Mayer KF. Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris*. Nature Communications. 2018;9(1):2515.

Watcharamongkol T, Christin PA, Osborne CP. C_4 photosynthesis evolved in warm climates but promoted migration to cooler ones. Ecology Letters. 2018;21(3):376-383.

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

Xi Z, Bradley RK, Wurdack KJ, Wong KM, Sugumaran M, Bomblies K, ... Davis CC. Horizontal transfer of expressed genes in a parasitic flowering plant. BMC Genomics. 2012;13(1):227.

Yang Z, Wafula EK, Kim G, Shahid S, McNeal JR, Ralph PE, ... Person TN. Convergent horizontal gene transfer and cross-talk of mobile nucleic acids in parasitic plants. Nature Plants. 2019;5:991-1001.

Yoshida S, Maruyama S, Nozaki H, Shirasu K. Horizontal gene transfer by the parasitic plant *Striga hermonthica*. Science. 2010;328(5982):1128-1128.

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

Tables

Table 1: Model species used in this study and associated traits.

Clade	Species	Ploid.	1C	#Genes tested	Cult.	LH	Clim.	Phot.	Cont.	Rhiz.
Pooideae	*Brachypodium distachyon*[1]	2n	0.31	17204	N	A	Temp	C_3	6	N
Pooideae	*Hordeum vulgare*[2]	2n	5.39	16192	Y	A	Temp	C_3	6	N
Pooideae	*Triticum aestivum*[3]	6n	16.95	56619	Y	A	Temp	C_3	6	N
Oryzoideae	*Oryza sativa*[4]	2n	0.49	19259	Y	A	Trop	C_3	6	N
Oryzoideae	*Leersia perrieri*[5]	2n	0.32	15777	N	A	Trop	C_3	1	N
Chloridoideae	*Eragrostis tef*[6]	4n	0.69	30605	Y	A	Temp	C_4	5	N
Chloridoideae	*Oropetium thomaeum*[7]	2n	0.29	15168	N	A	Temp	C_4	2	N
Chloridoideae	*Zoysia japonica*[8]	4n	0.42	20416	Y	P	Temp	C_4	1	Y
Andropogoneae	*Sorghum bicolor*[9]	2n	0.69	21962	Y	A	Temp	C_4	6	N
Andropogoneae	*Zea mays*[10]	2n	2.65	25866	Y	A	Temp	C_4	6	N
Paniceae	*Alloteropsis semialata*[11]	2n	1.10	23071	N	P	Trop	C_4	3	Y
Paniceae	*Cenchrus americanus*[12]	2n	2.65	20159	Y	A	Temp	C_4	4	N
Paniceae	*Dichanthelium oligosanthes*[13]	2n	0.96	17761	N	P	Cold	C_3	1	N
Paniceae	*Echinochloa crus-galli*[14]	6n	1.37	54181	N	A	Temp	C_4	6	N
Paniceae	*Panicum hallii*[15]	2n	0.55	30255	N	P	Temp	C_4	1	N
Paniceae	*Panicum virgatum*[16]	4n	1.89	45043	Y	P	Cold	C_4	3	Y
Paniceae	*Setaria italica*[17]	2n	0.49	27465	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; [3]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.

Table 2: Number of lateral gene transfers detected between the five clades.

			Donor clade					
Clade	**Species**	**# LGT**	**Non-grass**	**Pooid.**	**Ory.**	**Chlor.**	**Andro.**	**Pan.**
Pooideae	*Brachypodium distachyon*	4	0	-	0	0	4	0
Pooideae	*Hordeum vulgare*	0	0	-	0	0	0	0
Pooideae	*Triticum aestivum*	8(10)	0	-	0	5	0(2)	3
Oryzeae	*Oryza sativa*	0	0	0	-	0	0	0
Oryzeae	*Leersia perrieri*	1(4)	0	0	-	0	1	0(3)
Chloridoideae	*Eragrostis tef*	1(9)	0	0	0	-	0	1(9)
Chloridoideae	*Oropetium thomaeum*	0	0	0	0	-	0	0
Chloridoideae	*Zoysia japonica*	0	0	0	0	-	0	0
Andropogoneae	*Sorghum bicolor*	2(3)	0	0	0	0	-	2(3)
Andropogoneae	*Zea mays*	11	0	0	0	2	-	9
Paniceae	*Alloteropsis semialata*	20	0	0	0	4	16	-
Paniceae	*Cenchrus americanus*	15	0	0	0	5	10	-
Paniceae	*Dichanthelium oligosanthes*	4	0	4	0	0	0	-
Paniceae	*Echinochloa crus-galli*	10	0	0	0	3	7	-
Paniceae	*Panicum hallii*	8	0	0	0	6	2	-
Paniceae	*Panicum virgatum*	30	0	0	0	1	29	-
Paniceae	*Setaria italica*	7	0	0	0	0	7	-

The 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.

Table 3: Number of lateral gene transfers detected in Paniceae.

Subtribes	Species	# LGT	Pooid. (3,698 sp.)	Ory. (115 sp.)	Chlor. (1,602 sp.)	Andro. (1,202 sp.)	Cench. (287 sp.)	Pani. (157 sp.)
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

The 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.

Figures

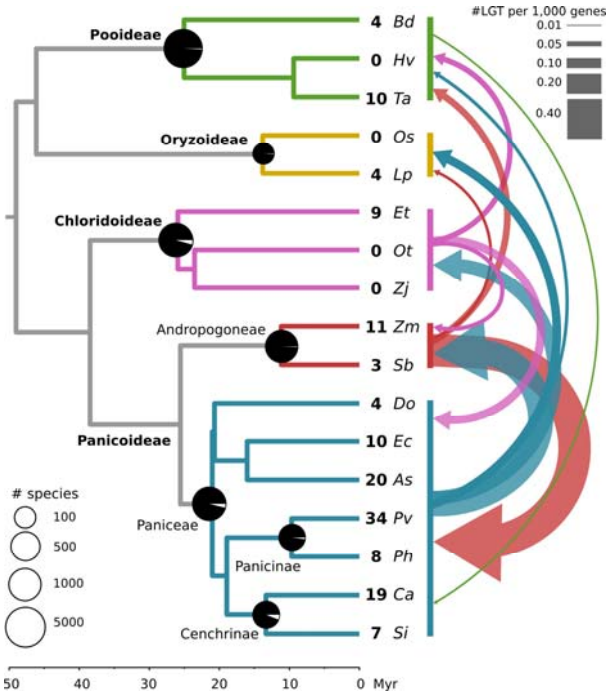

Fig 1: Distribution of lateral gene transfers among grasses.

Time-calibrated phylogenetic tree of 17 model grass species used in this study (extracted from Christin et al. 2014; scale in million years - Myr). The direction of LGT between grass clades is shown with arrows whose size is proportional to the number of LGT recived. 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 LGT detected in each genome.

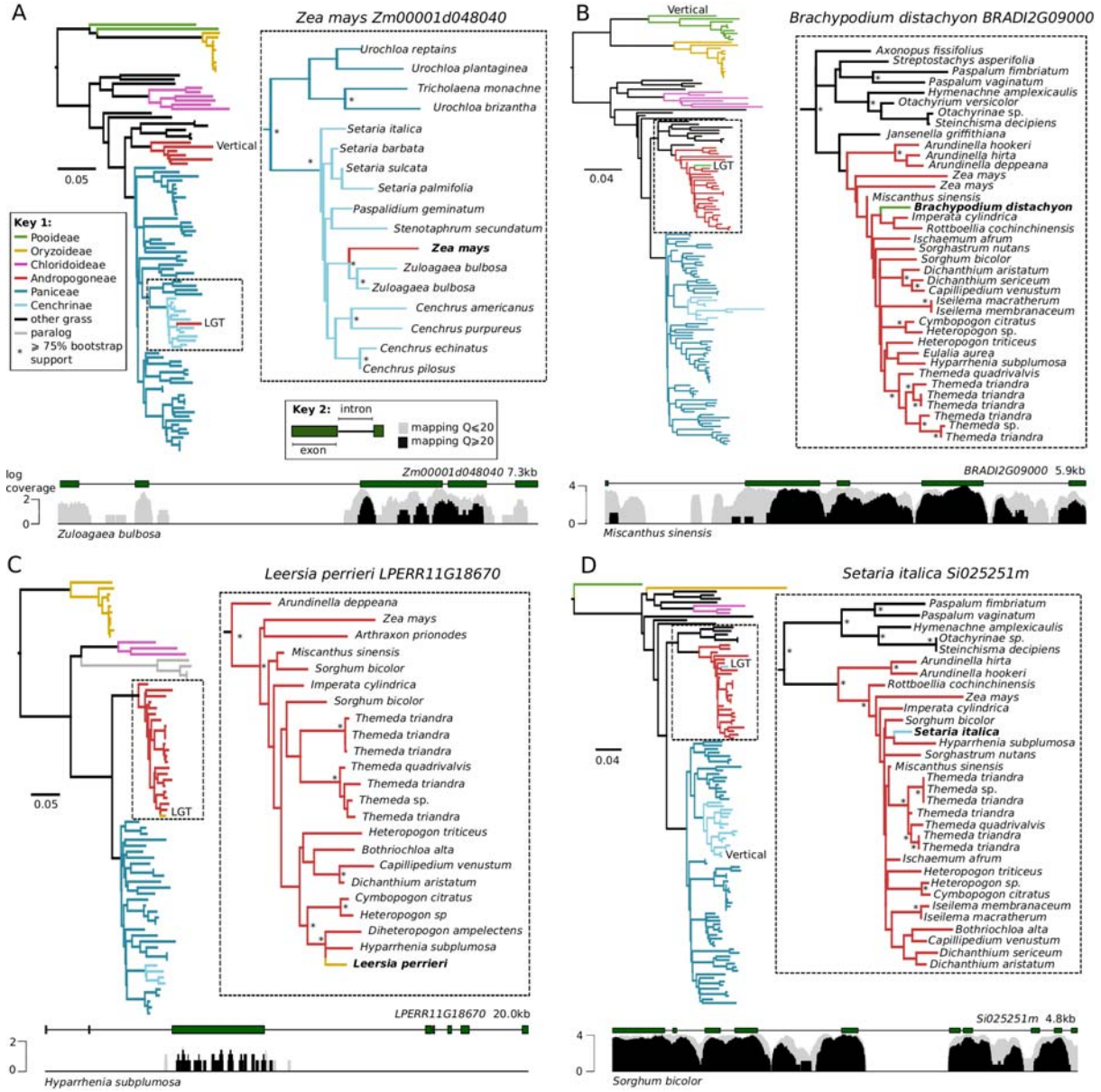

Fig 2: Four examples of grass-to-grass lateral gene transfer.

Each panel (A-D) shows an exemplar grass-to-grass LGT, with full and expanded regions of maximum likelihood phylogenies shown. 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.

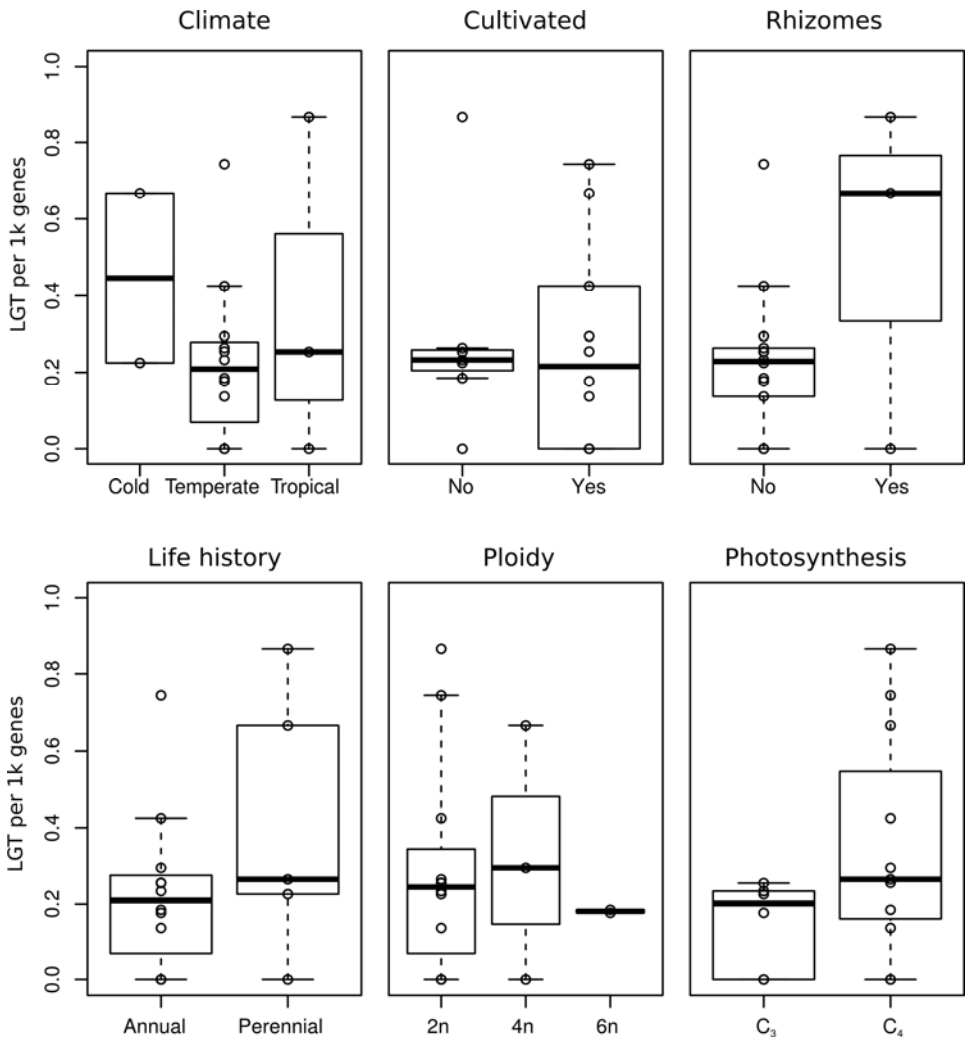

Fig. 3: Numbers of lateral gene transfer (LGT) received by different categories of grasses. For each group, the distribution of LGT numbers is shown with box plots connecting the median and the interquartile range, with whiskers showing 1.5 x the interquartile range.

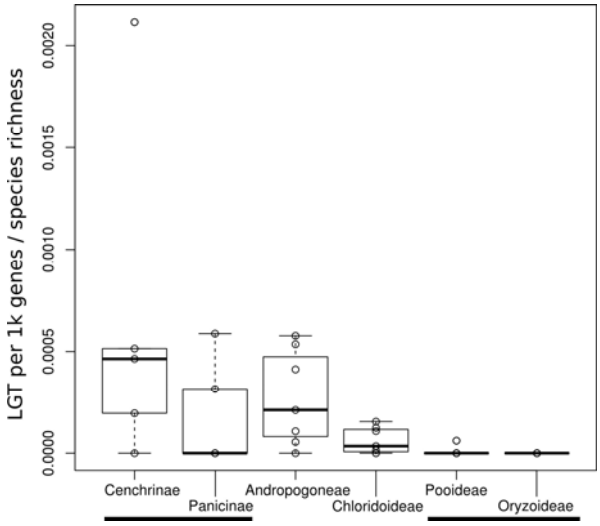

Fig. 4: Number of lateral gene transfer (LGT) received by Paniceae species from different clades.

The number of LGT in each Paniceae genome is corrected by the number of genes in the same genome 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 1.5 x interquartile range.