

Regulatory features determine the evolutionary fate of laterally acquired genes in plants

Catherine F. Collins ¹, Benjamin T. Alston ¹, Samuel G.S. Hibdige ^{1,2}, Pauline Raimondeau ^{1,3}, Emily R. Baker ¹, Graciela Sotelo ^{1,4}, Alexander S.T. Papadopoulos ⁵, Pascal-Antoine Christin ¹, Lara Pereira ^{1*}, Luke T. Dunning ^{1*}

¹Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

²Environment and Agrifood, Faculty of Engineering and Applied Sciences, Cranfield University, Cranfield MK43 0AL, UK

³Department of Ecology and Evolutionary Biology, Yale University, P.O. Box 208106, New Haven, CT 06520, USA

⁴Departamento de Ecología e Biología Animal, Universidade de Vigo, Vigo 36310, Spain

⁵Molecular Ecology and Evolution Group, School of Environmental and Natural Sciences, Bangor University, Bangor LL57 2UW, UK

*Corresponding authors: E-mails: l.dunning@sheffield.ac.uk; lara.pereira.garcia@sheffield.ac.uk.

Associate editor: Flavia Mascagni

Abstract

Lateral gene transfer (LGT) is widespread in eukaryotes, including in animals and plants where it can fuel adaptive evolution and innovation. However, the factors that influence the integration and long-term retention of transferred genes remain poorly understood. The pangenome of the grass *Alloteropsis* has a high turnover of laterally acquired genes, and here we combine expression, methylation, and genomic data to identify factors promoting their long-term persistence. Most transferred genes appear to be degenerating, showing lower expression levels and/or greater sequence truncation compared to their vertically inherited homologs. These degenerating genes also show significantly higher levels of DNA methylation, potentially indicating transcriptional silencing. The likelihood of a transferred gene being retained will be influenced by how easily it can be expressed in the recipient genome. In *Alloteropsis*, putatively functional laterally acquired genes had expression levels significantly more similar to their donor ortholog than to their vertically inherited homolog. Transferred genes carry *cis*-regulatory elements encoded on the fragment of DNA that moves between species, likely facilitating their expression in the new genomic context. Evolutionary novelty may also increase the likelihood that selection retains a transferred gene. However, only a significant difference in expression level, not sequence divergence, between donor orthologs and vertically inherited homologs is associated with successful lateral gene transfer. Overall, our results show that most transferred genes degrade over time. However, those capable of regulating their own expression are more likely to persist and contribute to long-term evolutionary innovation.

Keywords horizontal gene transfer, HGT, gene expression, selection, adaptation

Introduction

Horizontal or lateral gene transfer (LGT) is the acquisition of genetic material outside of sexual reproduction (Doolittle 1999). The importance of this process for prokaryote evolution has long been established (Freschi et al. 2019; Arnold et al. 2022; Gophna et al. 2022), and over the last 2 decades it has become increasingly apparent that LGT can also play an important role in eukaryote evolution (Van Etten and Bhattacharya 2020; Keeling 2024). LGT effectively expands a species gene pool, enabling it to use genetic variation from other species (Brockhurst et al. 2019). This increased standing variation can accelerate adaptation (Barrett and Schluter 2008), and acquired genetic information has underpinned several evolutionary innovations in eukaryotes, including the colonization of land by plants

(Ma et al. 2022), the silencing of host defenses by parasites (Xi et al. 2012; Gilbert and Maumus 2023), and the optimization of core metabolism (Christin et al. 2012). However, the specific genes transferred are random, and therefore the proportion of evolutionarily advantageous events is likely to be small. Indeed, in prokaryotes most transfers are neutral or deleterious, and rapidly lost in the recipient species through drift or selection (Sorek et al. 2007; Acar Kirit et al. 2020; Arnold et al. 2022). While LGT is known to occur in a wide range of plants and animals, we are yet to test the factors that influence the successful integration and long-term retention of transferred genes.

The probability of successful LGT depends critically on whether the recipient can express the transferred gene and produce a functional protein (Keeling 2024). LGT is particularly prevalent in parasitic plants (Xi et al. 2012; Yang et al. 2016,

Received: December 1, 2025. Accepted: January 21, 2026

© The Author(s) 2026. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1 The number of laterally acquired genes identified in *Alloteropsis*.

Species	Accession	Location	Raimondeau et al. (2023)		
			Read mapping	Phylogenetics	Used in this study
<i>A. semialata</i>	AUS1	Australia	50	50 [53]	50 [52]
	RSA5-03	South Africa	33	29 [31]	26 [28]
	TAN1-04B	Tanzania	45	43 [51]	43 [51]
	ZAM1505-10	Zambia	100	95 [125]	94 [124]
<i>A. angusta</i>	UGA4	Uganda	32	32 [33]	26 [27]
Total			168	168 [293]	167 [282]

The laterally acquired genes were identified by Raimondeau et al. (2023) using both read mapping and phylogenetic approaches, and in this study we restricted our analysis to only include those identified by both methods, present in the original genome annotation, and not 100% identical to another sequence. For each accession, we show the number of unique laterally acquired genes (i.e. not counting recent duplicates that could have arisen prior or post transfer separately), and the total number of LGT in each genome annotation (counting recent duplicates separately) given in parenthesis. The total row (in bold) indicates the number of unique laterally acquired genes across the *Alloteropsis* lineage, meaning that those shared by multiple reference genomes are only counted once. The number in the brackets on this row counts the laterally acquired genes in all 5 reference genomes, and their duplicates, separately.

2019; Vogel et al. 2018; Yoshida et al. 2019), where intimate physical contact with the host via the haustorium serves as a conduit for gene transfer. Despite the haustoria acting as a highway for mRNA movement between the species (Kim et al. 2014), most transfers are surprisingly driven by the movement of genomic DNA (Yang et al. 2016). This is thought to be because DNA fragments can also transfer intact promoters, increasing the likelihood that the gene remains functional in the recipient species (Yang et al. 2016). In grasses (Hibdige et al. 2021) and insects (Peccoud et al. 2017), the number of successful transfers increases as phylogenetic distance decreases, possibly also due to ease-of-use, with more closely related species more likely to share regulatory mechanisms (Hibdige et al. 2021). If the regulation of a gene is encoded on the transferred DNA fragment, then any advantageous regulatory feature from the donor may also be transferred. Therefore, genes which are predominantly influenced by *cis*-regulatory elements will be more easily expressed by the recipient and result in successful LGT.

The likelihood of retaining a laterally transferred gene will also increase if it adds functional novelty (Dunning and Christin 2020; Keeling 2024), such as introducing a gene that the recipient lacks. For example, ferns acquired a chimeric neochrome photoreceptor that was unique to hornworts, enabling them to successfully grow in low-light conditions (Li et al. 2014). However, not all transfers involve novel acquisitions. In grasses, nearly 80% of transferred genes co-exist with a vertically inherited homolog in the recipient's genome (Hibdige et al. 2021). In these cases, functional novelty may arise from genetic variation that accumulated since the donor and recipient lineages diverged. These differences can accumulate neutrally over time through genetic drift, and when the gene is transferred into a new genomic environment, it can enhance standing genetic variation that can later fuel adaptation (Olofsson et al. 2019). Alternatively, the donor gene may have undergone episodic positive selection prior to transfer. For example, phosphoenolpyruvate carboxylase (PEPC) is a key enzyme in C_4 photosynthesis, an adaptive trait that increases carbon fixation in hot and high light environments (Sage 2004). Three different genes encoding this enzyme were laterally acquired by separate populations of the grass *Alloteropsis semialata* (Christin et al. 2012; Dunning et al. 2017). These genes had been optimized for C_4 metabolism in older C_4

grass lineages through positive selection (Dunning et al. 2017), and their recurrent acquisition accelerated metabolic adaptation in the recipient (Phansopa et al. 2020). While there is some anecdotal evidence of its importance in eukaryotes, it has yet to be formally tested whether the likelihood of retaining a laterally acquired gene is correlated with the functional novelty it provides to the recipient.

LGT appears to be particularly widespread within the grass family (Vallenback et al. 2008; Christin et al. 2012; Prentice et al. 2015; Mahelka et al. 2017, 2021; Dunning et al. 2019; Hibdige et al. 2021; Park et al. 2021; Wu et al. 2022; Raimondeau et al. 2023), making it a useful system to determine the factors that influence the integration and retention of laterally acquired genes. Importantly, these studies focus on plant-plant transfers and are distinct from recent work suggesting that interdomain LGT into plants may be overestimated (Aguirre-Carvajal and Armijos-Jaramillo 2025). In grasses, *Alloteropsis* is an emerging model system for studying LGT (Christin et al. 2012; Dunning et al. 2017, 2019; Olofsson et al. 2019; Phansopa et al. 2020; Hibdige et al. 2021), and a recent in-depth phylogenetic analysis of grass-to-grass transfers in the *Alloteropsis* pangenome identified 168 unique laterally acquired genes in 5 reference genomes (Table 1), revealing a high turnover with continual gene gain and loss (Raimondeau et al. 2023). The distribution of the laterally acquired genes was then determined across the lineage using additional whole genome re-sequencing datasets, indicating that only a small proportion are old acquisitions retained over time, and most are relatively recent and geographically restricted (Raimondeau et al. 2023). Their apparent transience is further supported by the rate at which they are lost from the recipient species, which is more than 500 times higher than that of vertically inherited genes (Raimondeau et al. 2023).

Here, we expand on this previous study in *Alloteropsis* by comparing the patterns of expression and sequence evolution of the laterally acquired genes identified in the *Alloteropsis* pangenome with the orthologs in the donor species and vertically inherited homologs in the recipient. Using these data, we specifically test the following hypotheses: (1) laterally acquired genes are generally expressed at lower levels, more truncated and more heavily methylated than the corresponding homolog in the recipient species, consistent with the expectation that most are

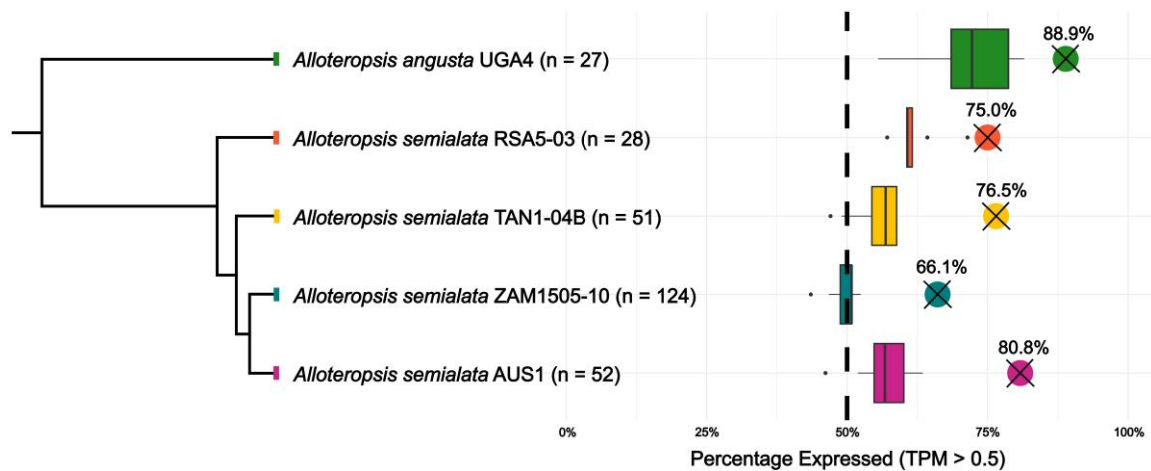


Figure 1 Expression of laterally acquired genes in *Alloteropsis*. A phylogenetic tree adapted from Bianconi et al. (2020) shows the evolutionary relationships of the *Alloteropsis* accessions, with the number of laterally acquired genes in each genome indicated (n). The dots with crosses show the percentage of genes that have detectable expression (TPM > 0.5) in at least 1 RNA-seq replicate for that accession. Boxplots show the median and interquartile range of the number of laterally acquired genes expressed in each individual RNA-seq dataset.

not adaptive and are undergoing functional degradation. (2) For the minority of laterally acquired genes that remain functional, their expression patterns will more closely resemble those of the donor ortholog than the recipient homolog, suggesting that *cis*-regulatory elements are important for their retention. Finally, (3) the donor ortholog and recipient homolog associated with functional laterally acquired genes are expected to have greater sequence and expression divergence than those associated with laterally acquired genes that degenerate postacquisition, suggesting evolutionary novelty increases their chances of long-term retention.

Results

Laterally acquired genes are expressed at relatively low levels

We generated 55 Illumina RNA-Seq data sets for the same 5 *Alloteropsis* accessions that have reference genomes and were used by Raimondeau et al. (2023) to identify laterally acquired genes (Table 1). For each accession, we sequenced up to 3 replicated samples from 4 tissues (leaf base, leaf tip, leaf sheath, and root) (Table S1). To infer gene expression levels (transcripts per million; TPM) we used a reference-based approach, mapping each accession to its respective reference genomes. Out of the 282 individual laterally acquired genes identified by both approaches in Raimondeau et al. (2023), and present in the original genome annotations (Table 1), 208 (73.8%) were expressed (≥ 0.5 TPM) in at least one of the 55 RNA-Seq datasets, with a similar proportion expressed in each of the 5 reference genomes (mean = 77.5%; SD = 8.3%; Fig. 1; Table S2). Most laterally acquired genes co-exist in the genome with their vertically inherited homolog (per accession mean = 72.1%, SD = 10.7%), with a significantly higher proportion ($P = 0.008$; Wilcoxon rank sum test) of recent duplication in the latter (laterally acquired genes: per accession mean = 10.3%; SD = 9.1%; vertical homolog: per accession mean = 31.9%; SD = 9.1%). The presence of a vertical

homolog is not associated with the expression of laterally acquired genes ($P = 0.830$; Pearson χ^2 test).

To compare the overall expression of laterally acquired genes and their vertically inherited homologs, we first calculated the log-TPM difference between them independently for all RNA-seq datasets from the different accessions and tissues (Fig. 2a). The log-TPM differences were then combined into a single linear mixed-effects model (LMM) that included accession and orthology to account for nonindependence (a similar approach was taken for all linear models). Overall, laterally acquired genes are expressed at significantly lower levels compared to vertically inherited homologs ($P = 1.87 \times 10^{-14}$; Fig. 2a). Using the same approach with an additional 19 RNA-seq datasets from the donor species (Table S1), we also showed that the laterally acquired genes are generally expressed at significantly lower levels (LMM; $P < 2 \times 10^{-16}$; Fig. 2b) than their ortholog in the donor. For the donor analysis, we used *Themeda triandra* as a proxy for transfers from Andropogoneae species ($n = 42$ genes in the 5 genomes), and *Setaria italica* for transfers from Cenchrinae species ($n = 69$ in the 5 genomes). *Themeda triandra* is a known donor, and *S. italica* belongs to the same genus as inferred Cenchrinae donors (Dunning et al. 2019). These species were therefore used as representatives of their respective clades, although not all inferred transfers necessarily originated from these specific species or genera.

There were 107 laterally acquired genes (range of 10 to 63 per accession) for which both the vertically inherited homolog and donor ortholog were present and expressed (TPM > 0.5; Table S2). When the laterally acquired genes were compared to both of these combining all tissues, over half (per accession mean = 53.8%; SD = 6.1%) were expressed at a significantly lower level (adjusted $P < 0.05$; Wilcoxon rank sum test; Table S3). Those with significantly reduced expression were classified as “degenerating,” with either the transfer itself inhibiting activity, or the gene being actively downregulated postacquisition. The remaining genes were classified as “putatively stable”. In total, 69.6% of the laterally acquired genes shared between accessions were placed in the same category (Table S3).

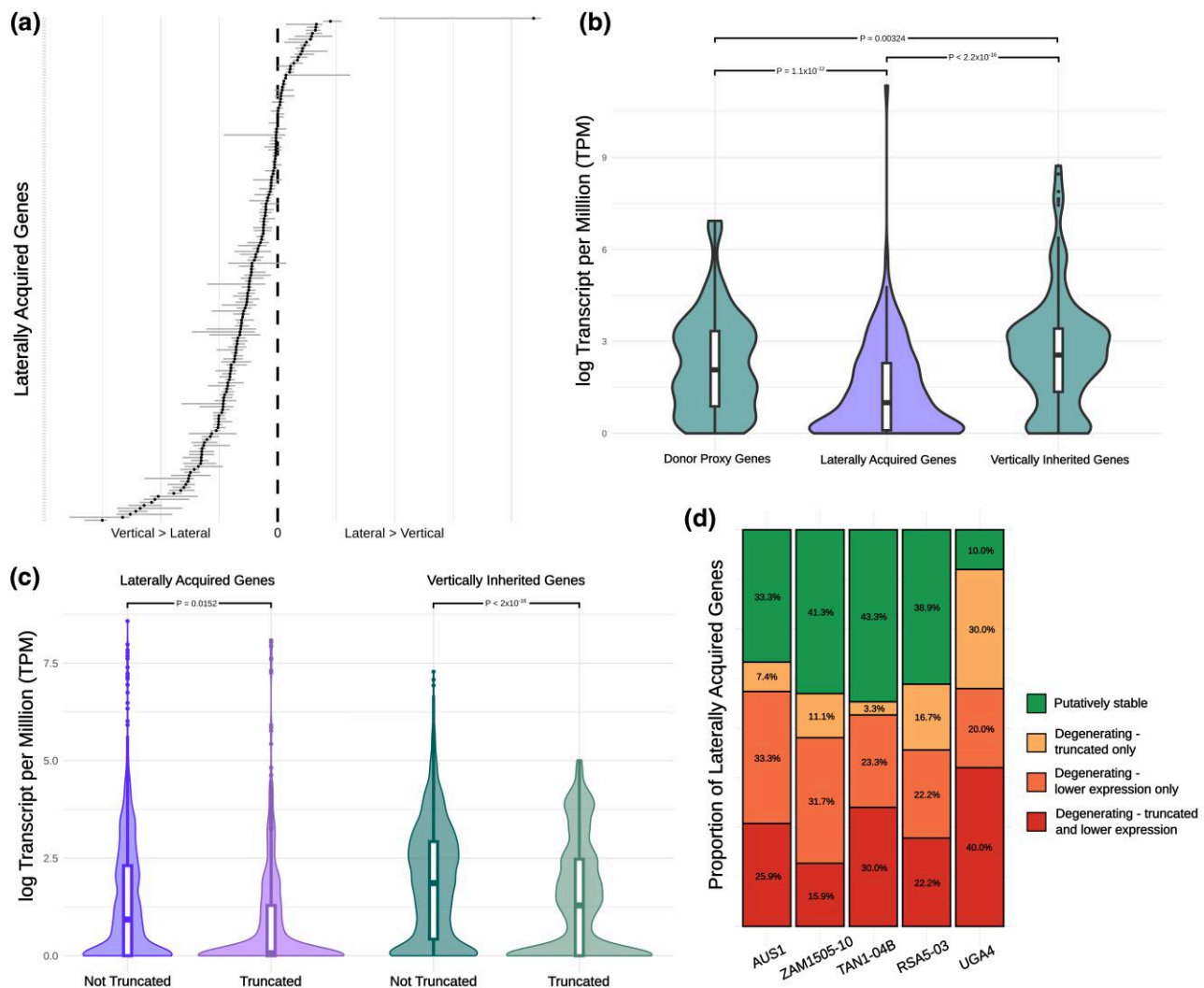


Figure 2 Expression patterns of laterally acquired genes. a) Comparison of expression levels between laterally acquired and vertically inherited homologs, all tissues combined. For each gene pair ($n = 169$) in each replicate, the log-transformed expression of the vertically inherited gene was subtracted from that of its corresponding laterally acquired homolog. Each dot indicates the median for the laterally acquired gene pair's replicates (paralog expression summed), the horizontal indicates interquartile range. The vertical dotted line corresponds to a difference of zero between lateral and vertical expression. b) Comparison of log-transformed expression levels between laterally acquired, vertically inherited and donor orthologs, with significant differences indicated. Laterally acquired genes show a mean 59.7% decrease in expression compared to the vertically inherited homologs, and 40.6% decrease compared to the donor orthologs. c) Comparison in log-transformed expression levels between truncated and nontruncated genes, for laterally acquired and vertically inherited homologs. d) Laterally acquired genes were divided into 4 categories based on whether they (i) had significantly lower expression levels than the vertically inherited and donor proxy gene and (ii) were classified as truncated ($< 70\%$ sequence length in comparison to nearest orthologs).

Laterally acquired genes have higher gene body methylation

Epigenetic silencing is one mechanism by which transgenes are rapidly silenced in genetically modified plants (Rajeevkumar et al. 2015), and laterally acquired genes may be subjected to similar processes. To investigate this, we generated bisulfite sequencing data from leaf tissue for the 4 *A. semialata* accessions, and calculated the proportion of methylated cytosines (CpG) for the gene body, exons, and 1 kb up- and down-stream. To test for a difference in the proportion of CpG methylation between laterally acquired genes and their vertically inherited

homologs, we again used a LMM. Laterally acquired genes had significantly higher CpG methylation across the gene body ($P = 0.001$), exons ($P = 0.004$), and 1 kb down-stream ($P = 0.026$; Fig. 3a). This difference was most pronounced across the gene body, with a mean of 2.7% increase in CpG methylated sites (Fig. 3a).

We repeated the LMM analysis to compare CpG methylation between the degenerating and putatively stable laterally acquired genes. Those that are classed as degenerating had a significantly higher proportion of CpG methylated sites in the gene body ($P = 0.003$; Fig. 3b). This pattern was also observed on the exons ($P = 0.034$).

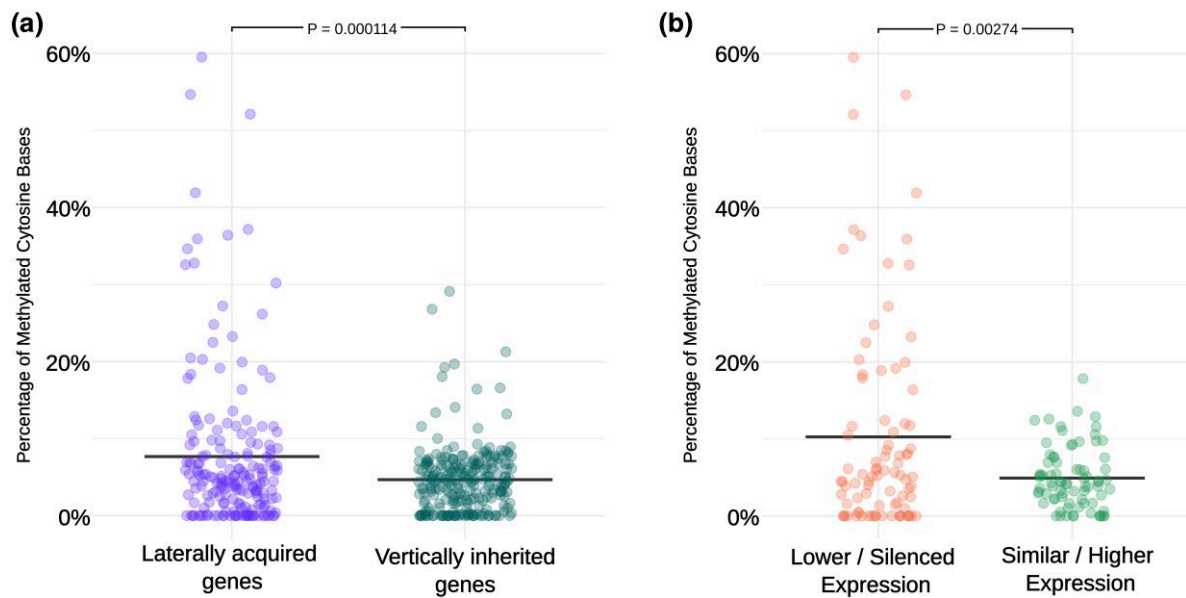


Figure 3 Methylation of laterally acquired genes. Comparison of proportion of methylated cytosine sites across the full gene region between a) laterally acquired and vertically inherited genes and b) laterally acquired genes with silenced or lower expression than their corresponding vertically inherited and donor proxy genes, and those with similar or higher expression. Horizontal bars indicate the mean.

Laterally acquired genes are frequently truncated

The coding sequences of laterally acquired genes and their vertically inherited homologs in the *Alloteropsis* reference genomes were compared with those of other model grasses to assess whether either were truncated. They were classed as truncated if: (i) the gene was less than 70% of the length of all other orthologs, (ii) the annotated sequence did not begin with a start codon, or (iii) the annotated sequence did not end with a stop codon. Across all 5 genomes, 33.3% of the laterally acquired genes were classified as truncated compared to 16.2% of their vertically inherited homologs (note only 2 genes had no start/stop codon and were not >30% truncated). Even though there were 10 truncated vertically inherited homologs with full length laterally acquired genes, the increase in the overall proportion of full length laterally acquired genes with truncated vertically inherited homologs was not significant ($P = 0.607$). A generalized linear mixed model (GLMM) indicated that vertically inherited homologs have significantly lower odds of being truncated than laterally acquired genes ($P = 4.44 \times 10^{-4}$). A separate LMM also showed that truncated genes had significantly lower expression levels than nontruncated genes ($P = 3.43 \times 10^{-16}$; Fig. 2c). The TPM expression measure accounts for gene length. Finally, degenerating genes were significantly more likely to be truncated than those that are putatively stable (GLMM $P = 0.012$). Given structural decay suggests functional loss, we reclassified truncated genes previously considered putatively stable as degenerating (if paralogs were present all copies had to be truncated). This adjustment further demonstrates that a substantial majority of laterally acquired genes are degenerating (mean = 66.6%, SD = 13.6%; Fig. 2d).

Functional laterally acquired genes have donor-like expression levels

We conducted a more detailed analysis on the expression pattern of the 56 laterally acquired genes that appear to be putatively stable in *Alloteropsis* to test if their expression pattern was more similar to the vertically inherited homolog or to the ortholog in the donor (Fig. 4a). The 56 laterally acquired genes represent 45 unique genes when accounting for orthologs present in more than 1 *Alloteropsis* accession. To test for a significant difference between the laterally acquired genes and their vertically inherited homolog/donor ortholog we used Kruskal-Wallis tests (FDR-corrected), followed by Dunn's post hoc test for pairwise differences using the FSA package (Ogle et al. 2025) in R. Based on these results, the laterally acquired genes were then placed into 1 of 5 different categories: (1) "No difference"—no significant difference among the 3 groups; (2) "Higher"—expression significantly higher than both donor orthologs and vertical homologs; (3) "Intermediate"—expression significantly different from both, but intermediate; and either (4) "recipient-like" or (5) "donor-like" if significantly different from only one but not the other. We then used binomial tests to look for significant differences in the number of genes assigned to each category.

A majority of putatively stable laterally acquired genes showed no difference in expression to the vertically inherited homolog or to the donor ortholog ($P = 7.58 \times 10^{-6}$, exact binomial test). The "no difference" category is not informative as to whether the laterally acquired gene expression level is dictated by *cis*-acting regulatory elements linked to the laterally acquired gene DNA fragment, or if its expression level is now governed by the recipient's own regulatory machinery. We therefore repeated the analysis without this classification, and out of the remaining

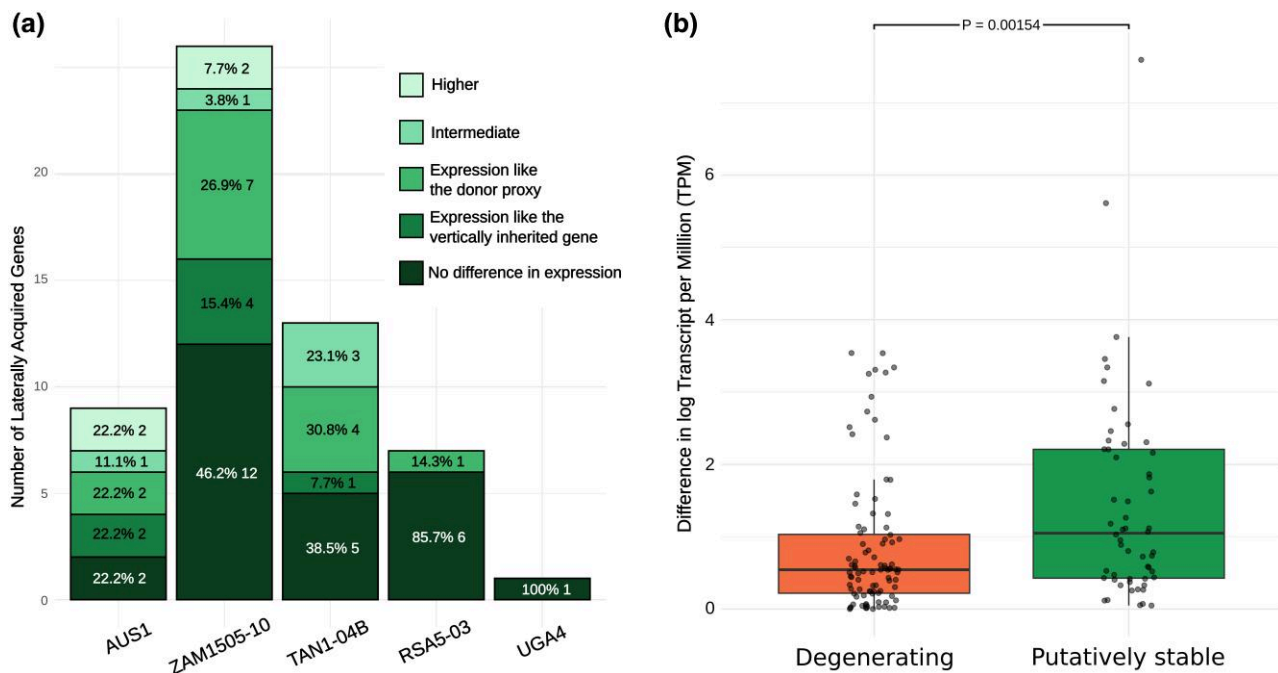


Figure 4 Expression patterns using vertically inherited and donor proxy genes as a comparison. a) Expression of laterally acquired genes classified as “putatively stable,” i.e. gene expression of the gene is not significantly lower than both the vertically inherited homolog and donor ortholog, and it has not been classified as truncated. Categories refer to the relationship between the laterally acquired gene and its corresponding vertically inherited homolog and donor orthologs. Higher: laterally acquired gene significantly higher in expression than the corresponding homologous genes; intermediate: laterally acquired gene intermediate in expression between the homologous genes. b) Difference in log transformed expression levels between corresponding vertically inherited and donor proxy genes for laterally acquired genes classified as “Degenerating” or “Putatively stable.” *P* value from Wilcoxon rank-sum test is shown above boxplots.

4 categories, putatively stable genes where expression levels were more similar to the donor proxy were significantly over-represented ($P = 0.008$, exact binomial test). This pattern is largely driven by 3 out of the 5 accessions (TAN1-04B, UGA4, and ZAM1505-10) (Fig. 4a).

To test if *cis*-acting regulatory elements received from the donor were present in the promoters of laterally acquired genes, we mapped short-read data from the donor to the AUS1 reference genome and showed that high-quality mapping overlaps with predicted transcription factor binding sites motifs (Figure S1).

Evidence of post-transfer expression modification

Three *Alloteropsis semialata* accessions possess a laterally acquired phosphoenolpyruvate carboxykinase (PCK) gene, which forms part of the C_4 photosynthetic pathway (Christin et al. 2012). In the exclusively C_4 accessions ZAM1505-10 and AUS1, expression of this laterally acquired PCK in the leaf and sheath tissues is comparable to the donor PCK ($P = 0.232$; $P = 0.885$; Figure S2). However, it is expressed significantly lower than the donor and vertical PCK in the roots ($P = 4.76 \times 10^{-6}$; $P = 0.004$). In these accessions, the vertically inherited PCK has lower expression in leaf and sheath ($P = 5.34 \times 10^{-6}$; $P = 0.002$), but is expressed at similar levels to the donor PCK in the root ($P = 0.262$; $P = 0.138$). This suggests that the transfer of PCK has driven sub-functionalization, with the laterally acquired PCK now only

responsible for C_4 photosynthesis, and the vertically inherited PCK restricted to a ubiquitous housekeeping role. Contrasting results were observed in TAN1-04B, a $C_3 + C_4$ intermediate, in which only part of its carbon fixation is through the C_4 pathway that requires high PCK expression. In this case, the laterally acquired PCK gene shows similar expression to the vertically inherited copy ($P = 0.841$), both being expressed significantly lower than the donor PCK ($P = 9.71 \times 10^{-5}$).

Stable transfers have donor orthologs and recipient homologs with increased expression divergence

We calculated the absolute difference in mean expression (logTPM) between vertically inherited homologs and donor orthologs for each laterally acquired gene (Table S4). The differences were significantly larger for orthologs associated with “putatively stable” (mean = 1.46; SD = 1.43) rather than “degenerating” (mean = 0.844; SD = 0.944) laterally acquired genes (Wilcoxon rank-sum test: $P = 0.002$; Fig. 4b). We also fitted a LMM to test the interaction between gene status (degenerating vs putatively stable) and gene type (vertically inherited vs donor) on expression levels, including accession and orthology to account for nonindependence. The interaction was highly significant ($P = 9.52 \times 10^{-12}$), confirming that difference in gene expression between the vertically inherited homolog and donor ortholog is significantly greater in putatively stable laterally

acquired genes than in degenerating ones. For putatively stable genes, vertical gene expression is generally higher than donor ortholog expression (Table S4).

Stable transfers do not have increased nucleotide divergence between donor orthologs and recipient homologs

We calculated the pairwise neutral genetic divergence (the rate of synonymous substitutions [dS]) between vertically inherited homologs and donor orthologs for each laterally acquired gene (Table S5). There was no significant difference associated with “putatively stable” (mean = 0.447; SD = 0.379) or “degenerating” (mean = 0.474; SD = 0.568) lateral acquired genes (Wilcoxon rank-sum test: $P = 0.837$; Figure S3). There was also no difference between the rate of nonsynonymous substitutions (dN, Wilcoxon rank-sum test: $P = 0.658$) or the overall selection pressure (dN/dS, Wilcoxon rank-sum test: $P = 0.645$). LMMs also showed no significant interaction between gene status (degenerating vs putatively stable) and gene type (vertically inherited vs donor) for dN ($P = 0.641$), dS ($P = 0.575$), or dN/dS ($P = 0.89$).

Laterally acquired genes show increased relaxed selection compared to the vertically inherited homolog

Using branch and branch-site models in CodeML (Yang 2007), we calculated the dN/dS ratio (ω) for laterally acquired and vertically inherited homologs since they coalesced in the gene tree (Table S5). For the laterally acquired gene, the ω value reflects its history of selection in the donor lineage and the time since it was transferred into the recipient genome. The branch model identified significant shifts in ω for 39/116 (14 decreased ω , 25 increased ω) laterally acquired genes ($P < 0.05$; Fig. 5a), with similar proportions for those classified as degenerating (10.8% decreased ω , 24.1% increased ω) and functional (11.5% decreased ω , 19.2% increased ω). In comparison, only 14/116 vertically inherited homologs showed significant shifts in ω compared to the background tree ($P < 0.05$; 10 decreased ω , 4 increased ω ; Fig. 5a). Subsequent branch-site analysis of the genes showing relaxed selection (increased ω) identified significant positive selection in 6/25 (3 degenerating, 2 stable, 1 with no identified donor proxy gene) laterally acquired genes, and 2/4 of the vertically inherited genes ($P < 0.05$; Fig. 5b).

Overall, laterally acquired genes were significantly more likely to show relaxed constraint than vertically inherited genes (Fisher’s exact test, $P = 1.12 \times 10^{-4}$), although there was no difference between those classified as degenerating or putatively stable (Fisher’s exact test, $P = 0.503$). It should be noted that these results do not include the 3 laterally acquired PEPC genes in *Alloteropsis* that were acquired from Andropogoneae, Cenchrinae and Melinidinae species, and have been shown to contain positively selected amino acid residues (Christin et al. 2012). This exclusion was because in Raimondeau et al. (2023) the genes either did not pass the stringent phylogenetic filters used to identify LGT (Cenchrinae and Melinidinae PEPC in ZAM1505-10), or the inferred gene tree lacked a vertically

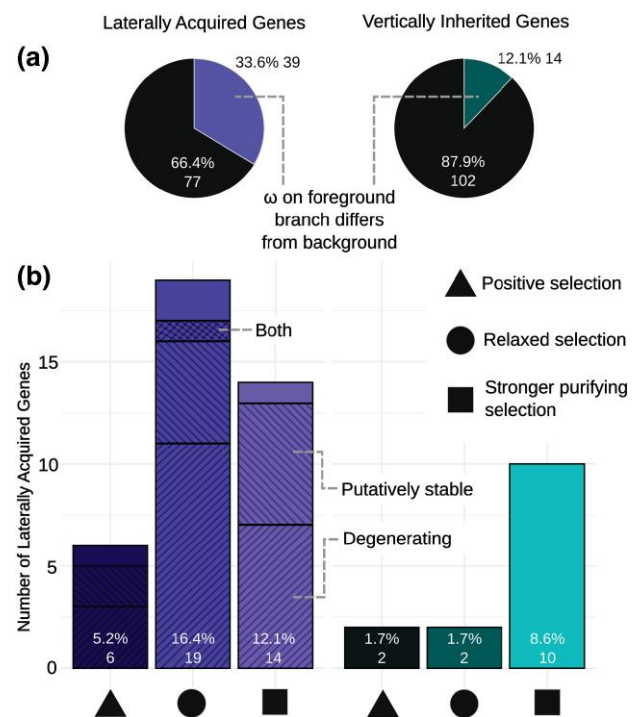


Figure 5 Selection patterns in laterally acquired and vertically inherited genes. a) Colored segments indicate genes where the foreground branch experienced significantly different selection pressures compared to background branches, as determined by the CodeML branch model ($P < 0.05$). b) Genes significantly different were further categorized as: stronger purifying selection, where the dN/dS ratio (ω) for the gene is lower than the background ω ; positive selection, where ω for the gene is higher than the background ω and evidence for sites under positive selection has been found using the PAML CodeML branch site model ($P < 0.05$); and relaxed selection, where ω for the gene is greater than the background ω but not significant in the branch site model. Percentages represent the proportion of overall genes ($n = 116$) in each category. For laterally acquired genes, hashing indicates “degenerating” and “putatively stable” categorization. “Both” indicates where categorization differed across accessions. No hash indicates genes where a donor copy was not present.

inherited copy in the same accession for comparison (Andropogoneae PEPC in AUS1).

Time since acquisition is not correlated with expression level

Molecular dates for the timing of each acquisition were obtained from Raimondeau et al. (2023) (Table S6). Overall, laterally acquired genes tend to be young, with 86 genes (54.4%) less than 2 million years old (Ma), 36 genes (22.8%) in the 2 to 4 Ma range, 32 genes (20.3%) in the 4 to 6 Ma range, and only 14 genes (8.9%) in the “more than 6 Ma” category. Although overall there is a correlation between the time since acquisition and expression level (LM; $P = 3.56 \times 10^{-5}$; $R^2 = 0.005$), this relationship is not significant when accounting for nonindependence in the data (orthology and accession) using a LMM. At the accession level, significant time-expression level relationships were detected in AUS1 (LM $P < 2 \times 10^{-16}$), TAN1-04B (LM $P = 0.000753$),

ZAM1505-10 (LM $P = 3.64 \times 10^{-6}$), and UGA4 (LM $P = 0.000259$), but only AUS1 had a significant relationship when accounting for non-independence within the LMM ($P = 0.00463$; Figure S4a). RSA5-03 showed no significant associations in either model. When age was compared between “Degenerating” and “Putatively stable,” no significant difference was found ($P = 0.1964$; Wilcoxon rank-sum Test). At the accession level, only ZAM1505-10 found a significant difference in age between these categories ($P = 0.0172$); all other accessions showed no significant difference (Figure S4b).

Discussion

Most lateral gene transfers are unsuccessful

Most laterally acquired genes are likely to be rapidly lost. Those that remain are predominately expressed, although a majority show significantly reduced expression compared to their vertically inherited homolog and donor ortholog. This pattern suggests that either the transfer of the gene itself has inhibited activity, or the gene has been downregulated following acquisition. The transferred genes also have a significantly higher sequence truncation rate (>2 -fold) in the genome compared to their vertically inherited counterparts. We collectively termed the relatively lowly expressed and/or truncated genes as “degenerating,” as their reduced expression and structural decay suggest functional loss. Degenerating genes represented a substantial proportion of the lateral gene content across the five genomes (mean = 68.3%, SD = 13.0%; Fig. 2d). The abolition of gene expression and nonsense mutations that generate truncated proteins are both processes that lead to pseudogenisation and ultimately the loss of a gene from a genome (Albalat et al. 2016). Previous findings in *Alloteropsis* have shown that laterally acquired genes are lost at a rate 500 times higher than vertically inherited loci (Raimondeau et al. 2023). Most laterally acquired genes in *Alloteropsis* appear to be nonadaptive and are only transient components of the recipient’s genome, mirroring patterns previously seen in prokaryotes (Sorek et al. 2007; Acar Kirit et al. 2020; Arnold et al. 2022). This means the genes detected likely represent only the briefly visible fraction of a much more extensive and ongoing process of LGT (Keeling 2024), with only a small subset occasionally retained through positive selection (Olofsson et al. 2019).

Epigenetic silencing pathways can target and rapidly suppress transgenic DNA (Neupert et al. 2020), and a similar mechanism may act to reduce the expression of laterally acquired genes, which are functionally analogous to transgenes. In *Alloteropsis*, the laterally acquired genes have significantly higher gene body methylation than their vertically inherited homologs. Furthermore, degenerating transferred genes have significantly more gene body methylation than those that remain transcriptionally active (Fig. 3). However, there is no significant difference in the number of methylated sites in the 1 kb upstream promoter region, a region typically associated with transcriptional suppression (Zhang et al. 2018). Given that these transfers happened many thousand or even millions of years ago, the promoter methylation may have been rendered obsolete in this time. Whilst the function of gene body methylation is debated (Muyle et al. 2022), methylation levels

are generally associated with chromatin accessibility (Zhong et al. 2021), and elevated gene body methylation may reflect chromatin remodeling inhibiting the activity of the transferred genes. This may reduce the expression and any potential deleterious effect the gene has, rendering its persistence in the genome subject to neutral evolutionary forces.

Here, we only assessed expression levels in 3 tissues (leaf blade, leaf sheath and root) of mature plants, under a single growth condition. It is possible that the laterally acquired genes with relatively low expression levels have been subfunctionalised and may remain functional in different tissues or at different developmental stages. However, the significant correlation between low expression and increased sequence truncation (Fig. 2c) further supports the conclusion that these genes are predominately degenerating and it is not mainly driven by undetected subfunctionalisation.

Successful transfers encode their own regulation

The putatively functional laterally acquired genes exhibit expression patterns that are more often similar to the donor ortholog than to the vertically inherited homolog ($P = 0.008$, exact binomial test; Fig. 4a). The genes themselves are transferred as part of large DNA fragments (Dunning et al. 2019), and this result implies that successful transfers are more likely to be transcriptional units that carry linked *cis*-regulatory elements, such as promoters. Indeed, for 3 of the genes that belong to clearly delineated fragments of foreign DNA (Dunning et al. 2019), all predicted transcription factor binding site motifs located up to 1-kb upstream from the first intron are from the donor (Figure S1). This process mirrors successful transgenesis, in which promoter sequences are critical for regulating the spatial and temporal expression of the transgene (Dutt et al. 2014). Genes that heavily depend on *trans*-regulatory mechanisms are less likely to be functionally retained after transfer, as they will become decoupled from the necessary regulatory landscape during the process. *Cis*-regulatory differences are also more commonly responsible for adaptive evolution (Signor and Nuzhdin 2018).

Whilst initial “plug in and play” may have an important role in successful LGT, we do see evidence of post-acquisition modification of gene expression patterns. PCK is present in plants where it is ubiquitously expressed in all tissues and plays a role in gluconeogenesis. In some species, PCK is also co-opted for the C_4 photosynthetic cycle, meaning that it becomes very highly expressed in leaf tissue. In the C_4 *Alloteropsis*, the lateral acquisition of an additional gene encoding PCK has enabled subfunctionalisation. The vertically inherited copy retains its ancestral gluconeogenesis function (moderately expressed in leaf and root tissue), whereas the laterally acquired copy is specialized C_4 photosynthesis (highly expressed in leaves, lowly expressed in roots) (SI Appendix, Figure S2). This post-acquisition specialization demonstrates that both the genomic and regulatory contexts into which a gene is inserted play a critical role in shaping its evolutionary fate. Furthermore, the likelihood of retaining a transferred gene may increase if the vertically inherited homolog was under balancing selection for multiple functions prior to the initial transfer event.

Functional laterally acquired genes are enriched for expression rather than coding divergence

The likelihood of successful LGT increases when it introduces functional novelty that is maintained by selection (Keeling 2024). Although the underlying mechanisms differ, this also applies to adaptive introgression, such as the spread of insecticide resistant genes in mosquitoes (Weill et al., 2000). Functional novelty can arise through divergence in the coding sequence or how the gene is expressed. Here, we find that functional lateral gene transfers are more common for genes exhibiting greater expression divergence between donor orthologs and recipient homologs (Fig. 4b). The acquisition of genes with novel expression patterns can be adaptive, for example the introgression of a gene with a *cis*-regulatory mutation has been associated with adaptive shifts in flower color in the monkeyflower *Mimulus aurantiacus* (Stankowski et al. 2015). In contrast, functional lateral gene transfers are not enriched for genes with elevated coding sequence divergence between donor and recipient copies (Figure S3). The relative contribution of protein-coding versus regulatory change to adaptive evolution has been debated, but data from Murid rodents suggests while adaptive noncoding regulatory changes may be more frequent, substitutions in proteins may have the largest effect on phenotypic evolution (Halligan et al. 2013). This may explain the overall enrichment of functional genes with expression divergence between donor and recipient copies, but not coding divergence, as the increased frequency of the former means the recipient is more likely to acquire genes with adaptive regulatory variation that can be selectively maintained.

Selection for transferred genes with adaptive protein coding divergence is still likely to be important in their retention. For example, the repeated acquisition of positively selected PEPC genes by *Alloteropsis* (Christin et al. 2012; Dunning et al. 2017) supports this, assuming there is no bias in the genes transferred. Our results did show that laterally acquired genes were more likely to be under relaxed selection than vertically inherited genes, possibly reflecting a shift to relaxed selection in the recipient as they are likely either deleterious or neutral. However, we did detect positive selection in 3-fold more laterally acquired genes compared to vertical inherited homologs (Fig. 5), but it should be noted that the absolute numbers are extremely low and there was no significant difference between those that appeared to be functional versus degenerating.

No clear relationship between gene age and expression decay

Based on our finding that most laterally acquired genes are degenerating, we would expect their expression levels to decay over time, with only those selectively retained remaining expressed in the long-term. However, overall we found no significant correlation between time since transfer and either expression levels or the classification of genes as either degenerating or putatively stable over the time scales considered here (Figure S4). Although it should be noted that there was a

significant association at the individual level for the Australian accession with higher expression in older LGT (Figure S4), which could be driven by differences in demographic history, local selective pressures, and genomic context, culminating in population-specific trajectories in the expression and retention of laterally acquired genes. Several factors could explain the lack of a relationship between gene age and expression decay. Laterally acquired genes are typically transferred as part of large genomic fragments (Dunning et al. 2019), potentially allowing neutral or slightly deleterious genes to persist through hitchhiking with beneficial ones (Olofsson et al. 2019). Similarly, if inserted into regions of the genome with a lower recombination rate it may take longer for them to be silenced and purged. In both instances, selection will take even longer to act if the laterally acquired gene is weakly deleterious or even selectively neutral. Finally, fluctuating selection in changing environments can maintain genetic variation, with genes initially retained for their adaptive potential but later becoming nonfunctional (Holstad et al. 2024). Similarly, it has been shown that 1 hitchhiking gene in *A. semialata* had a delayed selective impact, contributing to standing variation that became advantageous in a different geographic context (Olofsson et al. 2019).

Whilst this study focuses on *Alloteropsis*, the results are likely to have broad implications for other plants. Indeed, it has been previously shown that results in this species (Dunning et al. 2017, 2019) have been later mirrored in others (Hibdige et al. 2021). The apparently high levels of LGT reported in grasses, relative to other nonparasitic plant groups, is likely in part due to the extensive genomic resources available for this family. Because the most likely mechanism facilitating LGT in grasses is reproductive contamination (Pereira et al. 2023), similar levels of LGT may also occur in other wind-pollinated plant groups. The result that transferred genes are more likely to persist if they encode their own regulation should apply broadly, as gene regulation is universally critical for integration into new genomic contexts regardless of how the gene actually moved. This is relevant not only for natural LGT across plants, but also for transgenic plants and introgression loci between species. Ultimately, the persistence of a gene is dependent on its regulatory compatibility and its functional novelty.

Conclusion

LGT in *Alloteropsis* is frequent but largely transient, with most acquired genes showing signs of degradation through reduced expression and structural decay. However, a subset of acquired genes remain functional, particularly those with donor-like expression. This indicates that the presence of *cis*-regulatory elements on the fragments of DNA moving between species may increase the chances of their retention. The functional laterally acquired genes also have donor orthologs and recipient homologs with greater expression divergence, potentially indicating selection for regulatory novelty. The lack of a clear relationship between gene age and degeneration suggests that factors such as genomic context, fluctuating selection, and hitchhiking may mediate their long-term retention. Overall, while most laterally acquired genes are short-lived in the recipient species, those that persist can contribute to regulatory innovation and adaptive evolution.

Methods

RNA extraction, library generation and sequencing

Plants were grown in climatically controlled greenhouse conditions (12 h daylight, 25/20 °C day/night temperature) at the Arthur Willis Environmental Centre (AWEC) at The University of Sheffield. *Alloteropsis semialata* (AUS1 [Australian C₄], RSA5-03 [South African C₃], TAN1-04B [Tanzanian C₃ + C₄] and ZAM1505 [Zambian C₄]) and *Alloteropsis angusta* (UGA4) accessions were sampled from a living collection maintained at AWEC. *Themeda triandra* samples were taken from a Filipino accession (Dunning et al. 2022) (TtPh16-4), and *S. italica* was grown from seed (Herbiseed, UK). These species were chosen as *T. triandra* and *Setaria* sp. are known donors for some of the laterally acquired genes (Dunning et al. 2019). Samples from leaf tip, leaf base, leaf sheath, and root were taken from mature plants on the same day, flash frozen in liquid nitrogen and kept at −80 °C. Total RNA was extracted using a Qiagen RNeasy plant kit (Qiagen, Hilden, Germany) with an on-column DNA digestion step (RNase-Free DNase Set; Qiagen). RNA sequencing was performed in 2 batches. Libraries for the first 28 samples were generated using the TruSeq RNA Library Preparation Kit v2 (Illumina, San Diego, CA, USA), and 100 bp paired-end sequenced on 2 Illumina HiSeq 2500 flow cells (pooled with 20 samples from an unrelated project) at the Sheffield Diagnostic Genetics Service. For the remaining 46 samples, libraries were prepared and sequenced by Novogene, generating ~5 Gb of 150 bp paired-end reads per sample using the Illumina NovaSeq 6000 s4 platform.

Quantifying gene expression

Read quality was assessed before and after trimming with Fastqc v.0.11.8 (Andrews 2010) and Multiqc v.1.14 (Ewels et al. 2016). RNA-seq data was cleaned using Trimmomatic v.0.38 (Bolger et al. 2014) to remove adaptor contamination, low quality or ambiguous bases from both ends of each read (Phred quality score < 3), trim any remaining low quality bases (4 bp sliding window with mean Phred quality score < 30), and finally remove short reads (< 50 bp). The *Alloteropsis* RNA-seq data was mapped to reference genomes previously assembled for the same individual: AUS1 (GCA_004135705.1), RSA5-03 (GCA_036972165.1), TAN1-04B (GCA_036785585.1), ZAM1505 (GCA_036785565.1), UGA4 (GCA_037127165.1). *S. italica* was mapped to the *Setaria italica* v2.0 reference (GCA_000263155.2). *T. triandra* was mapped to *Sorghum bicolor* (*Sorghum bicolor*_NCBIv3 GCA_000003195.3), a closely related grass from the same subfamily (Andropogoneae), as the only reference for *T. triandra* is fragmented (N50 = 13.4 kb), incomplete (81.6% BUSCO complete) and unannotated (Dunning et al. 2022). Paired-end reads were quasi-mapped onto their respective reference genome and transcript expression counts quantified using Salmon v.1.4.0 (Patro et al. 2017). Transcript counts were converted to gene level TPM using the R package tximport v.1.3.9 (Soneson et al. 2015).

All sequencing datasets were also mapped to the *A. semialata* AUS1 reference genome for batch effect analysis. Principal

component analysis (PCA) was undertaken using the package DESeq2 v.1.40.0 (Love et al. 2014) to compare variability in expression profiles between samples. The resulting PCA did not highlight an effect of sequencing batch, but showed tissue-specific clustering, particularly separating leaf (sheath, base and tip) samples from root samples (SI Appendix, Figure S5).

Identifying vertical homologs and donor orthologs

This study uses grass-to-grass laterally acquired genes identified in the five *Alloteropsis* genomes by Raimondeau et al. (2023). This relied on a phylogenetic pipeline approach that requires well-resolved gene trees. Therefore, most inferred transfers are of relatively old genes present in multiple grass species/clades, meaning the laterally acquired genes often co-occur with ancestral copies in the recipient genome. We restricted our analysis to only those present in the genome annotation (i.e. not those additionally identified using Blast (Camacho et al. 2009) by Raimondeau et al. (2023), or those that were 100% identical that could cause problems quantifying their expression (1 identical copy was removed). We revisited the trees generated by Raimondeau et al. (2023) to classify each laterally acquired gene into 3 groups: (i) vertically inherited homolog and donor ortholog present; (ii) vertically inherited homolog present, donor ortholog absent; and (iii) both vertically inherited homolog and donor ortholog absent. For donors, we only considered transfers from Cenchrinae and Andropogoneae, with either *S. italica* or *S. bicolor* sequences present, respectively. Corresponding vertically inherited homologs had to belong to the same accession.

Laterally acquired gene expression analysis

For the gene expression analyses presented in the main text, we summed the TPM values of recent paralogs (i.e. sequences that formed a monophyletic *Alloteropsis* clade in the gene tree). We then used this to calculate log-TPM differences between laterally acquired genes, vertically inherited homologs and donor orthologs (Table 1). The log-TPM differences were calculated independently for each RNA-seq library, and we combined all tissues into each analysis so all genes regardless of tissue specificity could be included at once. To determine if summing paralog expression values influenced our results, we repeated all analyses using: (i) only the most highly expressed paralog, and (ii) only considering transfers without paralogs, i.e. neither the laterally acquired gene, vertical homolog or donor ortholog were duplicated in the gene tree (Table S2). Given the separation of root versus leaf tissues on principal component 1 (SI Appendix, Figure S4), we also repeated all analyses excluding roots (Table S2). In the main text we only refer to the results using summed TPM values across all tissue samples, but all the additional analyses are presented in Table S7, with only the coding sequence methylation analysis having predominantly conflicting results.

All statistical analyses were conducted in R (version 4.4.1) using the dplyr (Wickham et al. 2023) and broom (Robinson et al. 2022) packages for data manipulation. Linear mixed-effects

models (LMM) were fitted using the lme4 package in R (Bates et al. 2015). To test for a significant difference in expression between laterally acquired genes and vertically inherited homologs we fitted a LMM with the following structure:

$$\log \text{TPM difference} \sim 1 + (\text{LGT/Accession})$$

In this model, accession and LGT orthology were included as random effects to account for nonindependence of our data. Model fitting was performed using restricted maximum likelihood, and statistical significance was assessed using the lmerTest package (Kuznetsova et al. 2017). To test for a difference in expression between laterally acquired genes and their orthologs in the donor, another LMM was fitted with the following structure:

$$\log \text{TPM} \sim \text{Type} + (1 | \text{LGT/Accession})$$

In this model, *Type* (laterally acquired or donor ortholog) is included as a fixed effect. The nonindependence of the data is again accounted for as above.

To classify each laterally acquired gene as either “putatively functional” or “degenerating” we conducted 2 pairwise comparisons: (i) paired Wilcoxon signed-rank test to compare logCount between laterally acquired gene and vertically inherited homolog and (ii) unpaired Wilcoxon rank-sum test between laterally acquired gene and donor ortholog. Multiple testing corrections were performed using the BH method (Benjamini and Hochberg 1995). Laterally acquired genes with significantly lower expression compared to both vertical homologs and donor ortholog were categorized as “degenerating.”

Cis-regulatory elements prediction

To determine if *cis*-regulatory elements are also acquired from the donor species we looked at the noncoding flanking regions of 3 laterally acquired genes (LGT-082, LGT-083, and LGT-084 in Table S2) that belong to clearly delineated fragments of foreign DNA (Dunning et al. 2019) (note this is only possible for a small subset of the more recent transfers). First we predicted transcription factor binding sites in the 1 kb upstream promoter region for the 3 laterally acquired genes with PlantPAN 4.0 (Chow et al. 2024), using the PlantPAN 3.0 database constructed with *Zea mays*, *Oryza sativa*, and *Sorghum bicolor*. We then mapped whole genome re-sequencing data with Bowtie2 v2.4.5 (Langmead and Salzberg 2012), from the donor (*Iseilema vaginiflorum* NCBI SRA: SRR23559206), and an *A. semialata* accession without this laterally-acquired fragment (TPE1-10, SRA SRX4397332), and compared the coverage along the promoter. The results were visualized in the genome viewer IGV 2.5.2 (Thorvaldsdóttir et al. 2013).

Truncation analysis

To determine if the laterally acquired or vertically inherited orthologs were potentially nonfunctional, we compared their coding sequences in *Alloteropsis* to that of five other model grasses: *Brachypodium distachyon* (GCA_000005505.4), *O. sativa* (GCA_001433935.1), *S. italica* (GCA_000263155.2), *S. bicolor* (GCA_000003195.3), *Z. mays* (GCA_902167145.1). Blastn (Camacho et al. 2009) was used to identify the top-hit match in each of the model grasses, before the full length sequences were aligned using the Geneious Prime 2024.0 (<https://www.geneious.com>) multiple

sequence aligner. Each alignment was then manually inspected, and the *Alloteropsis* sequences classified as truncated if: (i) the *Alloteropsis* sequence was <70% of the length of the other sequences; (ii) no start codon at the beginning of the *Alloteropsis* sequence (present in model grasses); (iii) no stop codon at the end of the *Alloteropsis* sequence (present in model grasses).

A GLMM was used to examine the relationship between truncation and whether the gene is laterally acquired or vertically inherited. The model was specified as follows:

$$\text{Truncation} \sim \text{Type} + (1 | \text{LGT/Type})$$

Where *Truncation* represents a binary option, and to account for the hierarchical structure of the data, *LGT* (paralogy) and *Type* (laterally acquired or vertically inherited) were included as random intercepts. The model was fitted using the glmer() function from the lme4 package in R (Bates et al. 2015), with a binomial family and logit link function. A similar GLMM was additionally used to evaluate the effect of truncation status on laterally acquired gene expression category (degenerating vs putatively stable), specified as:

$$\text{Category} \sim \text{Truncation} + (1 | \text{LGT})$$

Methylation

DNA was extracted from the 4 *A. semialata* accessions and sent to Novogene for whole genome bisulfite library preparation and sequencing. For each library, ~35 Gb of data was generated (150 bp paired-end reads) using the Illumina NovaSeq 6000 S4 platform (coverage 33.2x–50.0x; Table S8). The bisulfite reads were trimmed using TrimGalore v0.6.10, a wrapper script that automates trimming (adapter and low quality bases), as well as the removal of biased methylation positions using Cutadapt v2.6 (Martin 2011) and FastQC. The methylation analysis was performed using Bismark v0.24.2 (Krueger and Andrews 2011). First, the reference genomes were indexed using the bismark_genome_preparation script, transforming the assembly into a bisulfite-converted version (C-to-T and G-to-A converted). Then, the Bismark aligner performed read mapping for each individual to its respective bisulfite-treated reference genome, using Bowtie2 v2.4.1 (Langmead and Salzberg 2012). Methylation data extraction was carried out using Bismark’s methylation extractor tool, with the following options: –no_overlap –comprehensive –bedGraph. Methylated cytosine site information was extracted from the bedGraph file and filtered to remove cytosines with fewer than 10 or more than 50 reads. These data were further filtered to identify methylated cytosine sites within the whole gene body, the coding sequence, and 1 kb up- and down-stream from the gene. The proportion of methylated cytosines was calculated based on the total number of cytosines with aligned reads in the region and the number of cytosines where at least 50% of reads were methylated.

To compare the proportion of methylated sites between laterally acquired and vertically inherited homologs, and between degenerating and putatively functional laterally acquired genes (based on gene expression only), we fitted LMMs using the lme4 package in R (Bates et al. 2015). Accession and LGT orthology were included as random effects to account for nonindependence of our data.

Selective constraint analysis

For the 116 laterally acquired genes that have a vertically inherited homolog present, we inferred the selection pressure they have been evolving under. A single representative from all *Alloteropsis* accessions was used as a reference, selected based on overall gene length and with a preference for selecting the same accession for both homologs. Pairwise dN, dS, and dN/dS (ω) were calculated using the Yang and Nielsen (2000) method implemented in yn00, part of the PAML v.4.10.7 package (Yang 2007). To detect signatures of adaptive or purifying selection, we further applied both branch and branch-site models in CodeML, following Yang et al. (2016) protocol. We used the sequence alignments and phylogenetic gene trees from Raimondeau et al. (2023), filtered for full length sequences. For direct comparison between the lateral and vertical homolog, ingroup taxa were removed to make the 2 homologs sister to each other in the tree. We then independently set the branches leading to each of these as foreground branches. The inferred ω reflects the cumulative selection pressure acting on each gene since the point of divergence. For the laterally acquired genes, this includes both the evolutionary history within the donor lineage, and the time since integration into the recipient genome. To test for significant shifts in ω , a likelihood ratio test was performed against the null model (single ω across all branches). For branches under significant relaxed selection (i.e. significantly higher ω) we subsequently ran branch-site models (parameters: model=2, NSsites=2, fix_omega=0) to identify codons under positive selection compared to the null model (fix_omega=1).

Molecular dating

The time since transfer for each laterally acquired gene was extracted from Raimondeau et al. (2023) to test if this is correlated with expression level or whether the gene is functional. To do this, we fitted a LMM using the lmer function from the lme4 package in R (Bates et al. 2015). Random intercepts were included for laterally acquired genes and nested accession to account for hierarchical structure and repeated measures within laterally acquired genes and accessions. Differences in time since acquisition between our expression categories were assessed using a Wilcoxon rank-sum test.

Author contributions

C.F.C., P.-A.C., L.P., and L.T.D. designed the study. S.G.S.H. and P.R. sampled plants for RNA work. E.B., L.P., G.S., and P.R. performed RNA lab work. L.T.D. generated bisulfite data. C.F.C. analysed the data with the help of B.T.A., L.P., and L.T.D. All authors helped interpret the results. C.F.C., L.P., and L.T.D. wrote the manuscript with the help of all authors.

Supplementary material

Supplementary material is available at *Molecular Biology and Evolution* online.

Funding

This work was funded by a Natural Environment Research Council (NERC) grant (NE/V000012/1). C.F.C. is supported by a University of Sheffield PhD scholarship, G.S. is supported by the MAPAS project (ERC grant agreement no 947921), P.-A.C. is supported by a Royal Society University Research Fellowship (URF\R\180022), and L.T.D. is supported by a NERC Independent Research Fellowship (NE/T011025/1).

Data availability

All raw RNA and bisulfite sequencing data have been deposited with NCBI under Bioproject PRJNA1299365. All scripts used in this study are available on GitHub: <https://github.com/Sheffield-Plant-Evolutionary-Genomics/>.

References

- Acar Kirit H, Lagator M, Bollback JP. Experimental determination of evolutionary barriers to horizontal gene transfer. *BMC Microbiol.* 2020;20:326. <https://doi.org/10.1186/s12866-020-01983-5>.
- Aguirre-Carvajal K, Armijos-Jaramillo V. Reassessing interkingdom horizontal gene transfer suggests limited influence on plant genomes. *Ecol Evol.* 2025;15:e72653. <https://doi.org/10.1002/ece3.72653>.
- Albalat R, Cañestro C. Evolution by gene loss. *Nat Rev Genet.* 2016;17:379–391. <https://doi.org/10.1038/nrg.2016.39>.
- Andrews S. FastQC: A Quality Control tool for High Throughput Sequence Data. 2010. Available at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Arnold BJ, Huang I-T, Hanage WP. Horizontal gene transfer and adaptive evolution in bacteria. *Nat Rev Microbiol.* 2022;20:206–218. <https://doi.org/10.1038/s41579-021-00650-4>.
- Barrett RD, Schluter D. Adaptation from standing genetic variation. *Trends Ecol Evol.* 2008;23:38–44. <https://doi.org/10.1016/j.tree.2007.09.008>.
- Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw.* 2015;67:1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Statistical Society. Series B (Methodological).* 1995;57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Bianconi ME et al. Contrasted histories of organelle and nuclear genomes underlying physiological diversification in a grass species. *Proc Biol Sci.* 2020;287:20201960. <https://doi.org/10.1098/rspb.2020.1960>.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Brockhurst MA et al. The ecology and evolution of pangenomes. *Curr Biol.* 2019;29:R1094–R1103. <https://doi.org/10.1016/j.cub.2019.08.012>.
- Camacho C et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421. <https://doi.org/10.1186/1471-2105-10-421>.

- Chow C-N *et al.* PlantPAN 4.0: updated database for identifying conserved non-coding sequences and exploring dynamic transcriptional regulation in plant promoters. *Nucleic Acids Res.* 2024;52:D1569–D1578. <https://doi.org/10.1093/nar/gkad945>.
- Christin P-A *et al.* Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. *Curr Biol.* 2012;22:445–449. <https://doi.org/10.1016/j.cub.2012.01.054>.
- Doolittle WF. Lateral genomics. *Trends Cell Biol.* 1999;9:M5–M8. [https://doi.org/10.1016/S0962-8924\(99\)01664-5](https://doi.org/10.1016/S0962-8924(99)01664-5).
- Dunning LT, Christin P-A. Reticulate evolution, lateral gene transfer, and innovation in plants. *Am J Bot.* 2020;107:541–544. <https://doi.org/10.1002/ajb2.1452>.
- Dunning LT *et al.* Introgression and repeated co-option facilitated the recurrent emergence of C_4 photosynthesis among close relatives. *Evolution.* 2017;71:1541–1555. <https://doi.org/10.1111/evo.13250>.
- Dunning LT *et al.* Lateral transfers of large DNA fragments spread functional genes among grasses. *Proc Natl Acad Sci U S A.* 2019;116:4416–4425. <https://doi.org/10.1073/pnas.1810031116>.
- Dunning LT *et al.* Hybridisation and chloroplast capture between distinct *Themeda triandra* lineages in Australia. *Mol Ecol.* 2022;31:5846–5860. <https://doi.org/10.1111/mec.16691>.
- Dutt M, Dhekney SA, Soriano L, Kandel R, Grosser JW. Temporal and spatial control of gene expression in horticultural crops. *Hortic Res.* 2014;1:14047. <https://doi.org/10.1038/hortres.2014.47>.
- Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* 2016;32:3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>.
- Freschi L *et al.* The *Pseudomonas aeruginosa* pan-genome provides new insights on its population structure, horizontal gene transfer, and pathogenicity. *Genome Biol Evol.* 2019;11:109–120. <https://doi.org/10.1093/gbe/evy259>.
- Gilbert C, Maumus F. Sidestepping Darwin: horizontal gene transfer from plants to insects. *Curr Opin Insect Sci.* 2023;57:101035. <https://doi.org/10.1016/j.cois.2023.101035>.
- Gophna U, Altman-Price N. Horizontal gene transfer in archaea—from mechanisms to genome evolution. *Annu Rev Microbiol.* 2022;76:481–502. <https://doi.org/10.1146/annurev-micro-040820-124627>.
- Halligan DL *et al.* Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. *PLoS Genet.* 2013;9:e1003995. <https://doi.org/10.1371/journal.pgen.1003995>.
- Hibdigé SGS, Raimondeau P, Christin P-A, Dunning LT. Widespread lateral gene transfer among grasses. *New Phytol.* 2021;230:2474–2486. <https://doi.org/10.1111/nph.17328>.
- Holstad A *et al.* Evolvability predicts macroevolution under fluctuating selection. *Science.* 2024;384:688–693. <https://doi.org/10.1126/science.adi8722>.
- Keeling PJ. Horizontal gene transfer in eukaryotes: aligning theory with data. *Nat Rev Genet.* 2024;25:416–430. <https://doi.org/10.1038/s41576-023-00688-5>.
- Kim G, LeBlanc ML, Wafula EK, DePamphilis CW, Westwood JH. Genomic-scale exchange of mRNA between a parasitic plant and its hosts. *Science.* 2014;345:808–811. <https://doi.org/10.1126/science.1253122>.
- Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics.* 2011;27:1571–1572. <https://doi.org/10.1093/bioinformatics/btr167>.
- Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. *J Stat Softw.* 2017;82:1–26. <https://doi.org/10.18637/jss.v082.i13>.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357–359. <https://doi.org/10.1038/nmeth.1923>.
- Li FW *et al.* Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. *Proc Natl Acad Sci U S A.* 2014;111:6672–6677. <https://doi.org/10.1073/pnas.1319929111>.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550. <https://doi.org/10.1186/s13059-014-0550-8>.
- Ma J *et al.* Major episodes of horizontal gene transfer drove the evolution of land plants. *Mol Plant.* 2022;15:857–871. <https://doi.org/10.1016/j.molp.2022.02.001>.
- Mahelka V *et al.* Multiple horizontal transfers of nuclear ribosomal genes between phylogenetically distinct grass lineages. *Proc Natl Acad Sci U S A.* 2017;114:1726–1731. <https://doi.org/10.1073/pnas.1613375114>.
- Mahelka V *et al.* A *Panicum*-derived chromosomal segment captured by *Hordeum* a few million years ago preserves a set of stress-related genes. *Plant J.* 2021;105:1141–1164. <https://doi.org/10.1111/tpj.15167>.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17:10. <https://doi.org/10.14806/ej.17.1.200>.
- Muyle AM, Seymour DK, Lv Y, Huettel B, Gaut BS. Gene body methylation in plants: mechanisms, functions, and important implications for understanding evolutionary processes. *Genome Biol Evol.* 2022;14:evac038. <https://doi.org/10.1093/gbe/evac038>.
- Neupert J *et al.* An epigenetic gene silencing pathway selectively acting on transgenic DNA in the green alga *Chlamydomonas*. *Nat Commun.* 2020;11:6269. <https://doi.org/10.1038/s41467-020-19983-4>.
- Ogle DH, Doll JC, Wheeler AP, Dinno A. FSA: Simple Fisheries Stock Assessment Methods. R package version 0.9.6. 2025.
- Olofsson JK *et al.* Population-specific selection on standing variation generated by lateral gene transfers in a grass. *Curr Biol.* 2019;29:3921–3927.e5. <https://doi.org/10.1016/j.cub.2019.09.023>.
- Park M, Christin P-A, Bennetzen JL. Sample sequence analysis uncovers recurrent horizontal transfers of transposable elements among grasses. *Mol Biol Evol.* 2021;38:3664–3675. <https://doi.org/10.1093/molbev/msab133>.
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods.* 2017;14:417–419. <https://doi.org/10.1038/nmeth.4197>.
- Peccoud J, Loiseau V, Cordaux R, Gilbert C. Massive horizontal transfer of transposable elements in insects. *Proc Natl Acad Sci U S A.* 2017;114:4721–4726. <https://doi.org/10.1073/pnas.1621178114>.
- Pereira L, Christin P-A, Dunning LT. The mechanisms underpinning lateral gene transfer between grasses. *Plants People Planet.* 2023;5:672–682. <https://doi.org/10.1002/ppp3.10347>.
- Phansopa C, Dunning LT, Reid JD, Christin P-A. Lateral gene transfer acts as an evolutionary shortcut to efficient C_4

- biochemistry. *Mol Biol Evol.* 2020;37:3094–3104. <https://doi.org/10.1093/molbev/msaa143>.
- 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. *Proc R Soc Lond B Biol Sci.* 2015;282:20152453. <https://doi.org/10.1098/rspb.2015.2453>.
- Raimondeau P *et al.* Lateral gene transfer generates accessory genes that accumulate at different rates within a grass lineage. *New Phytol.* 2023;240:2072–2084. <https://doi.org/10.1111/nph.19272>.
- Rajeevkumar S, Anunanthini P, Sathishkumar R. Epigenetic silencing in transgenic plants. *Front Plant Sci.* 2015;6:693. <https://doi.org/10.3389/fpls.2015.00693>.
- Robinson D, Hayes A, Couch S. broom: Convert Statistical Objects into Tidy Tibbles. 2022. <https://broom.tidymodels.org>.
- Sage RF. The evolution of C₄ photosynthesis. *New Phytol.* 2004;161:341–370. <https://doi.org/10.1111/j.1469-8137.2004.00974.x>.
- Signor SA, Nuzhdin SV. The evolution of gene expression in cis and trans. *Trends Genet.* 2018;34:532–544. <https://doi.org/10.1016/j.tig.2018.03.007>.
- Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* 2015;4:1521. <https://doi.org/10.12688/f1000research.7563.1>.
- Sorek R *et al.* Genome-wide experimental determination of barriers to horizontal gene transfer. *Science.* 2007;318:1449–1452. <https://doi.org/10.1126/science.1147112>.
- Stankowski S, Streisfeld MA. Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkey-flowers. *Proc R Soc Lond B Biol Sci.* 2015;282:20151666. <https://doi.org/10.1098/rspb.2015.1666>.
- Thorvaldsdóttir E, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 2013;14:178–192. <https://doi.org/10.1093/bib/bbs017>.
- Vallenback P, Jaarola M, Ghatnekar L, Bengtsson BO. Origin and timing of the horizontal transfer of a *PgiC* gene from *Poa* to *Festuca ovina*. *Mol Phylogenet Evol.* 2008;46:890–896. <https://doi.org/10.1016/j.ympev.2007.11.031>.
- Van Etten J, Bhattacharya D. Horizontal gene transfer in eukaryotes: not if, but how much? *Trends Genet.* 2020;36:915–925. <https://doi.org/10.1016/j.tig.2020.08.006>.
- Vogel A *et al.* Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris*. *Nat Commun.* 2018;9:2515. <https://doi.org/10.1038/s41467-018-04344-z>.
- Weill M *et al.* The kdr mutation occurs in the Mopti form of *Anopheles gambiae* s. s. through introgression. *Insect Mol Biol.* 2000;9:451–455. <https://doi.org/10.1046/j.1365-2583.2000.00206.x>.
- Wickham H, François R, Henry L, Müller K, Vaughan D. dplyr: A Grammar of Data Manipulation. R package version 1.1.4. 2023. <https://dplyr.tidyverse.org>.
- Wu D *et al.* Lateral transfers lead to the birth of momilactone biosynthetic gene clusters in grass. *Plant J.* 2022;111:1354–1367. <https://doi.org/10.1111/tjp.15893>.
- Xi Z *et al.* Horizontal transfer of expressed genes in a parasitic flowering plant. *BMC Genomics.* 2012;13:227. <https://doi.org/10.1186/1471-2164-13-227>.
- Yang Z. PAML 4: phylogenetic analysis by Maximum likelihood. *Mol Biol Evol.* 2007;24:1586–1591. <https://doi.org/10.1093/molbev/msm088>.
- Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol.* 2000;17:32–43. <https://doi.org/10.1093/oxfordjournals.molbev.a026236>.
- Yang Z *et al.* Horizontal gene transfer is more frequent with increased heterotrophy and contributes to parasite adaptation. *Proc Natl Acad Sci U S A.* 2016;113:E7010–E7019. <https://doi.org/10.1073/pnas.1608765113>.
- Yang Z *et al.* Convergent horizontal gene transfer and cross-talk of mobile nucleic acids in parasitic plants. *Nat Plants.* 2019;5:991–1001. <https://doi.org/10.1038/s41477-019-0458-0>.
- Yoshida S *et al.* Genome sequence of *Striga asiatica* provides insight into the evolution of plant parasitism. *Curr Biol.* 2019;29:3041–3052.e4. <https://doi.org/10.1016/j.cub.2019.07.086>.
- Zhang H, Lang Z, Zhu J-K. Dynamics and function of DNA methylation in plants. *Nat Rev Mol Cell Biol.* 2018;19:489–506. <https://doi.org/10.1038/s41580-018-0016-z>.
- Zhong Z *et al.* DNA methylation-linked chromatin accessibility affects genomic architecture in Arabidopsis. *Proc Natl Acad Sci U S A.* 2021;118:e2023347118. <https://doi.org/10.1073/pnas.2023347118>.



DSPACE

<https://dspace.org/>

Regulatory features determine the evolutionary fate of laterally acquired genes in plants

Collins, Catherine F.; Alston, Benjamin T.; Hibdige, Samuel; Raimondeau, Pauline; Baker, Emily R. ...

Mascagni, Flavia

2026-02

Oxford University Press (OUP)

<https://doi.org/10.1093/molbev/msag042>

Collins CF, Alston BT, Hibdige SGS, et al., (2026) Regulatory features determine the evolutionary fate of laterally acquired genes in plants. *Molecular Biology and Evolution*, Volume 43, Issue 2, February 2026, Article number msag042