



This is a repository copy of *Regulatory features determine the evolutionary fate of laterally acquired genes in plants.*

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/231722/>

Version: Preprint

---

**Preprint:**

Collins, C.F., Alston, B.T., Hibdige, S.G. et al. (7 more authors) (Submitted: 2025)  
Regulatory features determine the evolutionary fate of laterally acquired genes in plants.  
[Preprint - bioRxiv] (Submitted)

<https://doi.org/10.1101/2025.08.22.671697>

---

© 2025 The Author(s). This preprint is made available under a Creative Commons Attribution 4.0 International License. (<https://creativecommons.org/licenses/by/4.0/>)

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:  
<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

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

Catherine F Collins<sup>1</sup>, Benjamin T Alston<sup>1</sup>, Samuel GS Hibdige<sup>1,2</sup>, Pauline Raimondeau<sup>1,3</sup>, Emily Baker<sup>1</sup>, Graciela Sotelo<sup>1,4</sup>, Alexander S. T. Papadopoulos<sup>5</sup>, Pascal-Antoine Christin<sup>1</sup>, Lara Pereira<sup>1†</sup>, Luke T Dunning<sup>1†</sup>

## Affiliations:

<sup>1</sup> Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom

<sup>2</sup> Environment and Agrifood, Faculty of Engineering and Applied Sciences, Cranfield University, Cranfield MK43 0AL, United Kingdom

<sup>3</sup> Department of Ecology and Evolutionary Biology, Yale University, P.O. Box 208106, New Haven, Connecticut, 06520, USA

<sup>4</sup> Universidade de Vigo, Departamento de Ecoloxía e Bioloxía Animal, 36310 Vigo, Spain

<sup>5</sup> Molecular Ecology and Evolution Group, School of Environmental and Natural Sciences, Bangor University, Bangor LL57 2UW, United Kingdom

**†Co-corresponding author:** Luke T. Dunning; Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom; +44 (0) 1142220027; l.dunning@sheffield.ac.uk

**†Co-corresponding author:** Lara Pereira; Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom; lara.pereira.garcia@sheffield.ac.uk

## 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 xenolog than to their vertically inherited homolog. This pattern suggests that transferred genes may carry cis-regulatory elements encoded on the fragment of DNA that moves between species, 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 and recipient orthologs 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.

## Significance Statement

Lateral gene transfer (LGT) can introduce novel traits into plant genomes, yet most transferred genes are only transient residents and are degenerating, with reduced expression, truncation, and elevated DNA methylation. However, a minority persist and are more likely to resemble their donor counterparts in expression, suggesting co-transfer of cis-regulatory elements. These findings indicate that regulatory compatibility is key to their long-term survival.

## 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<sup>1</sup>. The importance of this process for prokaryote evolution has long been established<sup>2-4</sup>, and over the last two decades it has become increasingly apparent that LGT can also play an important role in eukaryote evolution<sup>5,6</sup>. LGT effectively expands a species gene pool, enabling it to use genetic variation from other species<sup>7</sup>. This increased standing variation can accelerate adaptation<sup>8</sup>, and acquired genetic information has underpinned several evolutionary innovations in eukaryotes, including the colonisation of land by plants<sup>9</sup>, the silencing of host defences by parasites<sup>10,11</sup>, and the optimisation of core metabolism<sup>12</sup>. 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<sup>3,13,14</sup>. 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<sup>5</sup>. LGT is particularly prevalent in parasitic plants<sup>10,15-18</sup>, 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<sup>19</sup>, most transfers are surprisingly driven by the movement of genomic DNA<sup>16</sup>. 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<sup>16</sup>. In grasses<sup>20</sup> and insects<sup>21</sup>, the number of successful transfers increases as phylogenetic distance decreases. This may also be due to ease-of-use, with more closely related species more likely to share regulatory mechanisms<sup>20</sup>. 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<sup>5,22</sup>, 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<sup>23</sup>. 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<sup>20</sup>. 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<sup>24</sup>. Alternatively, the donor ortholog may have undergone episodic positive selection. For example, phosphoenolpyruvate carboxylase (PEPC) is a key enzyme in C<sub>4</sub> photosynthesis, an adaptive trait that increases carbon fixation in hot and high light environments<sup>25</sup>. Three different genes encoding this enzyme were laterally acquired by separate populations of the grass *Alloteropsis semialata*<sup>12,26</sup>. These genes had been optimised for C<sub>4</sub> metabolism in older C<sub>4</sub> grass lineages through positive selection<sup>26</sup>, and their recurrent acquisition accelerated metabolic adaptation in the recipient<sup>27</sup>. 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 in the grass family<sup>12,20,28-35</sup>, making it a useful system to determine the factors that influence the integration and retention of laterally acquired genes. An in-depth phylogenetic analysis of grass-to-grass transfers in the *Alloteropsis* pangenome identified 168 unique laterally acquired genes in five reference genomes (Table 1), revealing a high turnover with continual gene gain and loss<sup>35</sup>. 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<sup>35</sup>. 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<sup>35</sup>.

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 xenologs 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 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 xenolog than the recipient homolog, suggesting that cis-regulatory elements are important for their retention. Finally, [3]

functional laterally acquired genes will correspond to orthologs with greater sequence and expression divergence between donor and recipient than those that degenerate post-acquisition, 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 five *Alloteropsis* accessions that have reference genomes and were used by Raimondeau *et al.*<sup>35</sup> to identify laterally acquired genes (Table 1). For each accession, we sequenced up to three replicated samples from four 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.*<sup>35</sup>, and present in the original genome annotations (Table 1), 208 (73.8%) were expressed ( $\geq 0.5$  TPM) in at least one of the 55 RNA-Seq datasets, with a similar proportion expressed in each of the five reference genomes (mean = 77.5%; SD = 8.3%; Figure 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 Chi-squared 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 (Figure 2a). The log-TPM differences were then combined into a single linear mixed-effects model (LMM) that included accession and orthology to account for non-independence (a similar approach was taken for all linear models). Overall, laterally acquired genes are expressed at significantly lower levels compared to vertically inherited homologs ( $p = 1.87 \times 10^{-14}$ ; Figure 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;  $p < 2 \times 10^{-16}$ ; Figure 2b) than their xenolog 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).

There were 107 laterally acquired genes (range of 10-63 per accession) for which both the vertically inherited homolog and donor xenolog 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 post acquisition. 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).

### **Laterally acquired genes have higher gene body methylation**

Epigenetic silencing is one mechanism by which transgenes are rapidly silenced in genetically modified plants<sup>36</sup>, and laterally acquired genes may be subjected to similar processes. To investigate this, we generated bisulfite sequencing data from leaf tissue for the four *A. semialata* accessions, and calculated the proportion of methylated cytosines (CpG) for the gene body, exons, and 1kb 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 ( $p = 0.026$ ; Figure 3a). This difference was most pronounced across the gene body, with a mean of 2.7% increase in CpG methylated sites (Figure 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 sites in the gene body ( $p = 0.003$ ; Figure 3b). This pattern was also observed on the exons ( $p = 0.034$ ).

### **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: (a) the gene was less than 70% of the length of all other orthologs, (b) the annotated sequence did not begin with a start codon, or (c) the annotated sequence did not end with a stop codon. Across all 5 genomes, 33.3% of the laterally acquired genes were classed as truncated compared to 16.2% of their vertically inherited homologs. 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 generalised 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 non-truncated genes ( $p = 3.43 \times 10^{-16}$ ; Figure 2c), n.b. 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%; Figure 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 xenolog in the donor (Figure 4a). The 56 laterally acquired genes represent 45 unique genes when accounting for orthologs present in more than one *Alloteropsis* accession. To test for a significant difference between the laterally acquired genes and their vertically inherited homolog/donor xenolog we used Kruskal-Wallis tests (FDR-corrected), followed by Dunn's post hoc test for pairwise differences using the FSA package<sup>37</sup> in R. Based on these results, the laterally acquired genes were then placed into one of five different categories: (1) 'No difference' - no significant difference among the three groups; (2) 'Higher' - expression significantly higher than both donor and vertical xenologs, (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 gene or to the donor xenologs ( $p = 7.58 \times 10^{-6}$ , 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 four 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 three out of the five accessions (TAN1-04B, UGA4 and ZAM1505-10) (Figure 4a).

### **Evidence of post-transfer expression modification**

Three *Allotroopsis semialata* accessions possess a laterally acquired phosphoenolpyruvate carboxykinase (PCK) gene, which forms part of the C<sub>4</sub> photosynthetic pathway<sup>12</sup>. In the exclusively C<sub>4</sub> accessions ZAM1505-10 and AUS1, expression of this laterally acquired PCK in the leaf and sheath tissues is comparable to the donor PCK ( $p = 0.232$ ;  $p = 0.885$ ). 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 subfunctionalization, with the laterally acquired PCK now only responsible for C<sub>4</sub> photosynthesis, and the vertically inherited PCK restricted to a ubiquitous housekeeping role. Contrasting results were observed in TAN1-04B, a C<sub>3</sub>+C<sub>4</sub> intermediate, in which only part of its carbon fixation is through the C<sub>4</sub> 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 and recipient orthologs with increased expression divergence.**

We calculated the absolute difference in mean expression (logTPM) between vertically inherited 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) lateral acquired genes (Wilcoxon rank-sum test:  $p = 0.002$ ; Figure 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 non-independence. The interaction was highly significant ( $p = 9.52 \times 10^{-12}$ ), confirming that difference in gene expression between the vertically inherited and donor homolog 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 gene expression (Table S4).

### **Stable transfers do not have increased nucleotide divergence between donor and recipient orthologs.**

We calculated the pairwise neutral genetic divergence (the rate of synonymous substitutions [dS]) between vertically inherited and donor orthologs for each laterally acquired gene (Table S5). There was no significant difference between orthologs 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 non-synonymous substitutions (dN, Wilcoxon rank-sum test:  $p = 0.658$ ) or the overall selection pressure (dN/dS, Wilcoxon rank-sum test:  $p = 0.645$ ). LMM's 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<sup>38</sup>, we calculated the dN/dS ratio ( $\omega$ ) for laterally acquired and vertically inherited homologs since they coalesced in the gene tree (Table S5). For the laterally acquired gene, the  $\omega$  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  $\omega$  for 39/116 (14 decreased  $\omega$ , 25 increased  $\omega$ ) laterally acquired genes ( $P < 0.05$ ; Figure 5a), with similar proportions for those classified as degenerating (10.8% decreased  $\omega$ , 24.1% increased  $\omega$ ) and functional (11.5% decreased  $\omega$ , 19.2% increased  $\omega$ ). In comparison, only 14/116 vertically inherited homologs showed significant shifts in  $\omega$  compared to the background tree ( $p < 0.05$ ; 10 decreased  $\omega$ , 4 increased  $\omega$ ; Figure 5a). Subsequent branch-site analysis of the genes showing relaxed selection (increased  $\omega$ ) 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 ( $p < 0.05$ ; Figure 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 functional (Fisher's exact test,  $p = 0.503$ ). It should be noted that these results do not include the three 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<sup>12</sup>. This exclusion was because in Raimondeau et al.<sup>35</sup> 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 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.<sup>35</sup> (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 2–4 Ma range, 32 genes (20.3%) in the 4–6 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 non-independence in the data (orthology and accession). At the accession level, significant time–expression level relationships were detected in AUS1 (LM  $p < 2 \times 10^{-16}$ ; LMER  $p = 0.00463$ ), TAN1-04B (LM  $p = 0.000753$ ), ZAM1505-10 (LM  $p = 3.64 \times 10^{-6}$ ), and USA4 (LM  $p = 0.000259$ ; Figure S3). KWT 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 S3).

# Discussion

## Most laterally gene transfers are unsuccessful

Most laterally acquired genes are expressed, however, a majority show significantly reduced expression compared to both their vertically inherited homolog and donor xenolog. This 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%; Figure 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<sup>39</sup>. Previous findings in *Alloteropsis* have shown that laterally acquired genes are lost at a rate 500 times higher than vertically inherited loci<sup>35</sup>. Most laterally acquired genes in *Alloteropsis* appear to be non-adaptive and are only transient components of the recipient's genome, mirroring patterns previously seen in prokaryotes<sup>3,13,14</sup>. This means the genes detected likely represent only the briefly visible fraction of a much more extensive and ongoing process of lateral gene transfer<sup>5</sup>, with only a small subset occasionally retained through positive selection<sup>24</sup>.

Epigenetic silencing pathways can target and rapidly suppress transgenic DNA<sup>40</sup>, 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 (Figure 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<sup>41</sup>. 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<sup>42</sup>, methylation levels are generally associated with chromatin accessibility<sup>43</sup>, and elevated gene body methylation may reflect chromatin remodelling 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.

## Successful transfers encode their own regulation

The putatively functional laterally acquired genes exhibit expression patterns that are more often similar to the donor xenolog than to the vertically inherited homolog ( $p = 0.008$ , exact binomial test; Figure 4a). The genes themselves are transferred as part of large DNA fragments<sup>32</sup>, and this result implies that successful transfers are more likely to be transcriptional units that carry linked cis-regulatory elements, such as promoters. This process mirrors successful transgenesis, in which the promoter sequences are critical for regulating the spatial and temporal expression of the transgene<sup>44</sup>. 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<sup>45</sup>.

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  $C_4$  *Allotetraploids*, 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 specialised  $C_4$  photosynthesis (highly expressed in leaves, lowly expressed in roots) (SI Appendix, Figure S1). This post-acquisition specialisation 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 novelty of gene expression profiles

The likelihood of successful lateral gene transfer increases when it introduces functional novelty that is maintained by selection<sup>5</sup>. Although the underlying mechanisms differ, this also applies to adaptive introgression, such as the spread of insecticide resistant genes in mosquitoes<sup>46</sup>. 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 and recipient orthologs (Figure 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 colour in the monkeyflower *Mimulus aurantiacus*<sup>47</sup>.

In contrast, functional lateral gene transfers are not enriched for genes with elevated coding sequence divergence between donor and recipient orthologs (Figure S2). 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<sup>48</sup>. This may explain the overall enrichment of functional genes with expression, 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 genes encoding PEPC by *Alloteropsis*, which were subject to strong positive selection in the donor lineages<sup>12,26</sup>. 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 (Figure 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.

### Time since transfer

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 (Figure S3), although it should be noted that there were significant associations at the individual level for some accessions (Figure S3). Several factors could explain this pattern. Laterally acquired genes are typically transferred as part of large genomic fragments<sup>32</sup>, potentially allowing neutral or slightly deleterious genes to persist through hitchhiking with beneficial ones<sup>24</sup>. 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 non-functional<sup>49</sup>. Similarly, it has been shown that one hitchhiking gene in *A. semialata* had a delayed selective impact, contributing to standing variation that became advantageous in a different geographic context<sup>24</sup>.



## Conclusion

Lateral gene transfer (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 and recipient orthologs 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 C<sub>4</sub>], RSA5-03 [South African C<sub>3</sub>], TAN1-04B [Tanzanian C<sub>3</sub>+C<sub>4</sub>] and ZAM1505 [Zambian C<sub>4</sub>]) and *Alloteropsis angusta* (UGA4) accessions were sampled from a living collection maintained at AWEC. *Themeda triandra* samples were taken from a Filipino accession<sup>50</sup> (TtPh16-4), and *Setaria 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<sup>32</sup>. Samples from leaf tip, leaf base, leaf sheath, and root were taken 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 two 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 two 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<sup>51</sup> and Multiqc v.1.14<sup>52</sup>. RNA-seq data was cleaned using Trimmomatic v.0.38<sup>53</sup> 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<sup>50</sup>. Paired-end reads were quasi-mapped onto their respective reference genome and transcript expression counts

quantified using Salmon v.1.4.0<sup>54</sup>. Transcript counts were converted to gene level transcripts per million (TPM) using the R package tximport v.1.3.9<sup>55</sup>.

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<sup>56</sup> 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 S4).

### Identifying vertical and donor orthologs

This study uses laterally acquired genes identified by Raimondeau et al<sup>35</sup>. We restricted our analysis to only those present in the genome annotation (i.e. not those additionally identified using Blast<sup>57</sup> by Raimondeau et al<sup>35</sup>), or those that were 100% identical that could cause problems quantifying their expression (one identical copy was removed). We revisited the trees generated by Raimondeau et al.<sup>35</sup> to classify each laterally acquired gene into three groups: [1] vertically inherited homolog and donor xenolog present; [2] vertically inherited homolog present, donor xenolog absent; and [3] both vertically inherited homolog and donor xenolog 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 xenologs (Table 1). The log-TPM differences were calculated independently for each RNA-seq library, and we combined all tissues into each analysis. To determine if summing paralog expression values influenced our results, we repeated all analyses using: [1] only the most highly expressed paralog, and [2] only considering transfers without paralogs, i.e. neither the laterally acquired gene, vertical homolog or donor xenolog 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<sup>58</sup> and broom<sup>59</sup> packages for data manipulation. Linear mixed-effects models (LMM) were fitted using the lme4 package in R<sup>60</sup>. To test for a significant difference in expression between laterally acquired genes and vertically inherited homologs we fitted a LMM with the following structure:

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

Where accession and LGT orthology were included as random effects to account for non-independence of our data. Model fitting was performed using restricted maximum likelihood (REML), and statistical significance was assessed using the lmerTest package<sup>61</sup>. To test for a difference in expression between laterally acquired genes and their xenologs in the donor, another LMM was fitted with the following structure:

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

Where *Type* (laterally acquired or donor xenolog) is included as a fixed effect. The non-independence of the data is again accounted for as above.

To classify each laterally acquired gene as either ‘putatively functional’ or ‘degenerating’ we conducted two pairwise comparisons: [1] paired Wilcoxon signed-rank test to compare logCount between laterally acquired gene and vertically inherited homolog and [2] unpaired Wilcoxon rank-sum test between laterally acquired gene and donor xenolog. Multiple testing corrections were performed using the BH method<sup>62</sup>. Laterally acquired genes with significantly lower expression compared to both vertical homologs and donor xenologs were categorised as ‘degenerating’.

### Truncation analysis

To determine if the laterally acquired or vertically inherited orthologs were potentially non-functional, we compared their coding sequences in *Alloteropsis* to that of five other model grasses: *Brachypodium distachyon* (GCA\_000005505.4), *Oryza sativa* (GCA\_001433935.1), *Setaria italica* (GCA\_000263155.2), *Sorghum bicolor* (GCA\_000003195.3), *Zea mays* (GCA\_902167145.1). Blastn<sup>57</sup> 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: [1] the *Alloteropsis* sequence was < 70% of the length of the other sequences;  
[2] no start codon at the beginning of the *Alloteropsis* sequence (present in model grasses);  
[3] 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:

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

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<sup>60</sup>, 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:

$$Category \sim Truncation + (1 | LGT)$$

## Methylation

DNA was extracted from the four *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<sup>63</sup> and FastQC. The methylation analysis was performed using Bismark v0.24.2<sup>64</sup>. 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<sup>65</sup>. 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<sup>60</sup>. Accession and LGT orthology were included as random effects to account for non-independence 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 ( $\omega$ ) were calculated using the Yang and Nielsen<sup>66</sup> method implemented in yn00, part of the PAML v.4.10.7 package<sup>38</sup>. To detect signatures of adaptive or purifying selection, we further applied both branch and branch-site models in CodeML, following Yang et al.<sup>16</sup> protocol. We used the sequence alignments and phylogenetic gene trees from Raimondeau et al.<sup>35</sup>, filtered for full length sequences. For direct comparison between the lateral and vertical homolog, ingroup taxa were removed to make the two homologs sister to each other in the tree. We then independently set the branches leading to each of these as foreground branches. The inferred  $\omega$  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  $\omega$ , a likelihood ratio test was performed against the null model (single  $\omega$  across all branches). For branches under significant relaxed selection (i.e. significantly higher  $\omega$ ) 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.<sup>35</sup> 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<sup>60</sup>. 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. To assess the relationship between time since acquisition and our expression categories, we fitted a GLMM using the glmer function from the lme4 package in R<sup>60</sup>. Again, a random intercept was included for laterally acquired genes to account for

non-independence among LGTs within the same laterally acquired gene. Both models were assessed for model significance using likelihood ratio tests. The molecular dating analyses were also performed at the individual level.

## Acknowledgements

This work was funded by a Natural Environment Research Council (NERC) grant (NE/V000012/1). CFC is supported by a University of Sheffield PhD scholarship, GS is supported by the MAPAS project (ERC grant agreement no 947921), P-AC is supported by a Royal Society University Research Fellowship (URF\R\180022), LP is supported by a NERC Standard Grant (NE/V000012/1), LTD is supported by a NERC Independent Research Fellowship (NE/T011025/1).

## Author contributions

CFC, P-AC, LP and LTD designed the study. SGSH and PR sampled plants for RNA work. EB, LP, GS and PR performed RNA lab work. LTD generated bisulfite data. CFC analysed the data with the help of BTA, LP and LTD. All authors helped interpret the results. CFC, LP and LTD wrote the manuscript with the help of all authors.

## 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

1. W.F. Doolittle, Lateral genomics. *Trends in Cell Biology*, **9**(12), M5–M8 (1999).
2. L. Freschi *et al.*, The *Pseudomonas aeruginosa* Pan-Genome Provides New Insights on Its Population Structure, Horizontal Gene Transfer, and Pathogenicity. *Genome Biology and Evolution*, **11**(1), 109–120 (2019).
3. B.J. Arnold, I.-T. Huang, W.P. Hanage, Horizontal gene transfer and adaptive evolution in bacteria. *Nature Reviews Microbiology*, **20**(4), 206–218 (2022).
4. U. Gophna, N. Altman-Price, Horizontal Gene Transfer in Archaea—From Mechanisms to Genome Evolution. *Annual Review of Microbiology*, **76**, 481–502 (2022).
5. P.J. Keeling, Horizontal gene transfer in eukaryotes: aligning theory with data. *Nature Reviews Genetics*, **25**(6), 416–430 (2024).
6. J. Van Etten, D. Bhattacharya. Horizontal Gene Transfer in Eukaryotes: Not if, but How Much?. *Trends in Genetics*, **36**(12), 915–925 (2020).
7. M.A. Brockhurst, *et al.* The Ecology and Evolution of Pangenomes. *Current Biology*, **29**(20), R1094–R1103 (2019).
8. R.D. Barrett, D. Schluter. Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, **23**(1), 38–44 (2008).
9. J. Ma *et al.* Major episodes of horizontal gene transfer drove the evolution of land plants. *Molecular Plant*, **15**(5), 857–871 (2022).
10. Z. Xi *et al.* Horizontal transfer of expressed genes in a parasitic flowering plant. *BMC Genomics*, **13**(1), 227 (2012).
11. C. Gilbert, F. Maumus. Sidestepping Darwin: horizontal gene transfer from plants to insects. *Current Opinion in Insect Science*, **57**, 101035 (2023).
12. P.-A. Christin *et al.* Adaptive Evolution of C<sub>4</sub> Photosynthesis through Recurrent Lateral Gene Transfer. *Current Biology*, **22**(5), 445–449 (2012).
13. R. Sorek *et al.* Genome-Wide Experimental Determination of Barriers to Horizontal Gene Transfer. *Science*, **318**(5855), 1449–1452 (2007).
14. H. Acar Kirit, M. Lagator, J.P. Bollback. Experimental determination of evolutionary barriers to horizontal gene transfer. *BMC Microbiology*, **20**(1), 326 (2020).
15. A. Vogel *et al.* Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris*. *Nature communications*, **9**(1), 2515 (2018).
16. Z. Yang *et al.* Horizontal gene transfer is more frequent with increased heterotrophy and contributes to parasite adaptation. *Proceedings of the National Academy of Sciences*, **113**(45), E7010–E7019 (2016).
17. Z. Yang *et al.* Convergent horizontal gene transfer and cross-talk of mobile nucleic acids in parasitic plants. *Nature Plants*, **5**(9), 991–1001 (2019).



18. S. Yoshida *et al.* Genome sequence of *Striga asiatica* provides insight into the evolution of plant parasitism. *Current Biology*, **29**(18), 3041-3052 (2019).
19. G. Kim, M.L. LeBlanc, E.K. Wafula, C.W. DePamphilis, J.H. Westwood. Genomic-scale exchange of mRNA between a parasitic plant and its hosts. *Science*, **345**(6198), 808-811 (2014).
20. S.G.S. Hibdige *et al.* Widespread lateral gene transfer among grasses. *New Phytologist*, **230**(6), 2474–2486 (2021).
21. J. Peccoud *et al.* Massive horizontal transfer of transposable elements in insects. *Proceedings of the National Academy of Sciences*, **114**(18), 4721–4726 (2017).
22. L.T. Dunning, P.-A. Christin. Reticulate evolution, lateral gene transfer, and innovation in plants. *American Journal of Botany*, **107**(4), 541-544 (2020).
23. F.W. Li *et al.* Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. *Proceedings of the National Academy of Sciences*, **111**(18), 6672-6677 (2014).
24. J.K. Olofsson *et al.* Population-Specific Selection on Standing Variation Generated by Lateral Gene Transfers in a Grass. *Current Biology*, **29**(22), 3921-3927.e5 (2019).
25. R.F. Sage. The evolution of C<sub>4</sub> photosynthesis. *New Phytologist*, **161**(2), 341-370 (2004).
26. L.T. Dunning *et al.* Introgression and repeated co-option facilitated the recurrent emergence of C<sub>4</sub> photosynthesis among close relatives. *Evolution*, **71**(6), 1541–1555 (2017).
27. C. Phansopa *et al.* Lateral Gene Transfer Acts As an Evolutionary Shortcut to Efficient C4 Biochemistry. *Molecular Biology and Evolution*, **37**(11), 3094–3104 (2020).
28. P. Vallenback, M. Jaarola, L. Ghatnekar, B.O. Bengtsson. Origin and timing of the horizontal transfer of a *PgiC* gene from *Poa* to *Festuca ovina*. *Molecular Phylogenetics and Evolution*, **46**(3), 890-896 (2008).
29. H.C. Prentice, Y. Li, M. Lönn, A. Tunlid, L. Ghatnekar. A horizontally transferred nuclear gene is associated with microhabitat variation in a natural plant population. *Proceedings of the Royal Society B: Biological Sciences*, **282**(1821), 20152453 (2015).
30. V. Mahelka *et al.* Multiple horizontal transfers of nuclear ribosomal genes between phylogenetically distinct grass lineages. *Proceedings of the National Academy of Sciences*, **114**(7), 1726-1731 (2017).
31. V. Mahelka *et al.* A *Panicum*-derived chromosomal segment captured by *Hordeum* a few million years ago preserves a set of stress-related genes. *The Plant Journal*, **105**(5), 1141–1164 (2021).
32. L.T. Dunning *et al.* Lateral transfers of large DNA fragments spread functional genes among grasses. *Proceedings of the National Academy of Sciences*, **116**(10), 4416–4425 (2019).
33. M. Park, P.-A. Christin, J.L. Bennetzen. Sample sequence analysis uncovers recurrent horizontal transfers of transposable elements among grasses. *Molecular Biology and Evolution*, **38**(9), 3664-3675 (2021).
34. D. Wu *et al.* Lateral transfers lead to the birth of momilactone biosynthetic gene clusters

in grass. *The Plant Journal*, **111**(5), 1354-1367 (2022).

35. P. Raimondeau *et al.* Lateral gene transfer generates accessory genes that accumulate at different rates within a grass lineage. *New Phytologist*, **240**(5), 2072–2084 (2023).

36. S. Rajeevkumar, P. Anunanthini, R. Sathishkumar. Epigenetic silencing in transgenic plants. *Frontiers in Plant Science*, **6**, 693 (2015).

37. D.H. Ogle, J.C. Doll, A.P. Wheeler, A. Dinno. FSA: Simple Fisheries Stock Assessment Methods. R package version 0.9.6, (2025).

38. Z. Yang. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, **24**(8), 1586–1591 (2007).

39. R. Albalat, C. Cañestro. Evolution by gene loss. *Nature Reviews Genetics*, **17**(7), 379–391 (2016).

40. J. Neupert *et al.* An epigenetic gene silencing pathway selectively acting on transgenic DNA in the green alga *Chlamydomonas*. *Nature Communications*, **11**(1), 6269 (2020).

41. H. Zhang, Z. Lang, J.-K. Zhu. Dynamics and function of DNA methylation in plants. *Nature Reviews Molecular Cell Biology*, **19**(8), pp. 489–506 (2018).

42. A. M. Muyle, D. K. Seymour, Y. Lv, B. Huettel, B.S. Gaut. Gene body methylation in plants: mechanisms, functions, and important implications for understanding evolutionary processes. *Genome Biology and Evolution*, **14**(4), evac038 (2022).

43. Z. Zhong *et al.* DNA methylation-linked chromatin accessibility affects genomic architecture in Arabidopsis. *Proceedings of the National Academy of Sciences*, **118**(5), e2023347118 (2021).

44. M. Dutt, S.A. Dhekney, L. Soriano, R. Kandel, J.W. Grosser. Temporal and spatial control of gene expression in horticultural crops. *Horticulture Research*, **1**, 14047 (2014).

45. S.A. Signor, S.V. Nuzhdin. The evolution of gene expression in cis and trans. *Trends in Genetics*, **34**(7), 532-544 (2018).

46. M. Weill *et al.* The kdr mutation occurs in the Mopti form of *Anopheles gambiae s. s.* through introgression. *Insect Molecular Biology*, **9**(5), 451-455 (2001).

47. S. Stankowski, M.A. Streisfeld. Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. *Proceedings of the Royal Society B: Biological Sciences*, **282**(1814), 20151666 (2015).

48. D.L. Halligan *et al.* Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. *PLOS Genetics*, **9**(12), e1003995 (2013).

49. Agnes Holstad *et al.* Evolvability predicts macroevolution under fluctuating selection. *Science* **384**, 688-693 (2024).

50. L.T. Dunning *et al.* Hybridisation and chloroplast capture between distinct *Themeda triandra* lineages in Australia. *Molecular Ecology*, **31**, 5846–5860 (2022).

51. S. Andrews. FastQC: A Quality Control tool for High Throughput Sequence Data. (2010). Available at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

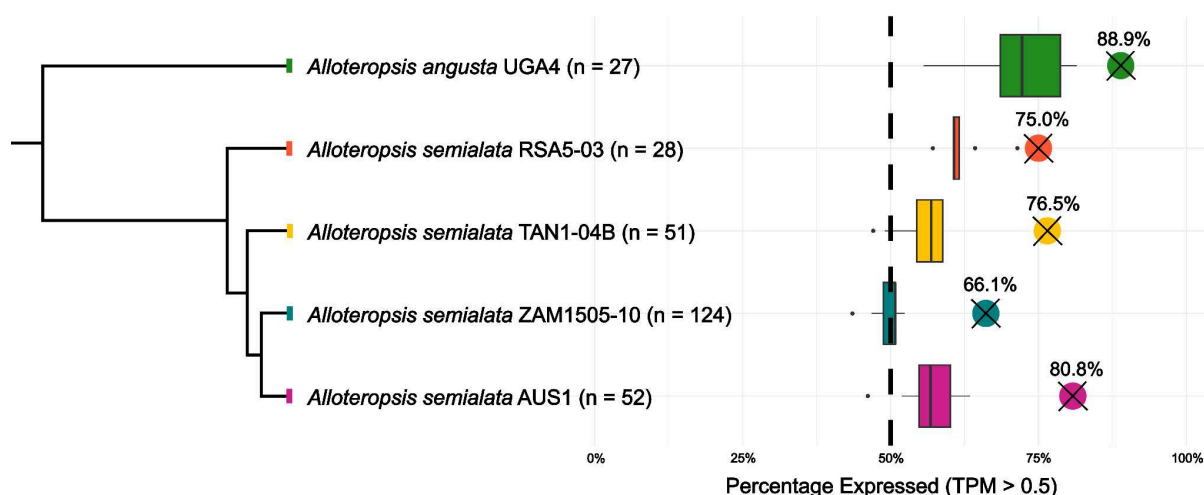
52. P. Ewels *et al.* MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, **32**(19), 3047–3048 (2016).
53. A.M. Bolger, M. Lohse, B. Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**(15), 2114–2120 (2014).
54. R. Patro *et al.* Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, **14**(4), 417–419 (2017).
55. C. Sonesson, M.I. Love, M.D. Robinson. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, **4**, 1521 (2015).
56. M.I. Love, W. Huber, S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**(12), 550 (2014).
57. C. Camacho *et al.* BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421 (2009).
58. H. Wickham, R. François, L. Henry, K. Müller, D. Vaughan. dplyr: A Grammar of Data Manipulation. R package version 1.1.4, <https://dplyr.tidyverse.org> (2023).
59. D. Robinson, A. Hayes, S. Couch. broom: Convert Statistical Objects into Tidy Tibbles. <https://broom.tidymodels.org> (2022).
60. D. Bates *et al.* Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, **67**, 1–48 (2015).
61. A. Kuznetsova, P.B. Brockhoff, R.H.B. Christensen. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, **82**, 1–26 (2017).
62. Y. Benjamini, Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**(1), 289–300 (1995).
63. M. Martin. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet.journal*, **17**(1), 10 (2011).
64. F. Krueger, S.R. Andrews. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, **27**(11), 1571–1572 (2011).
65. B. Langmead, S.L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**(4), 357–359 (2012).
66. Z. Yang, R. Nielsen. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution*, **17**, 32–43 (2000).
67. M.E. Bianconi *et al.* Contrasted histories of organelle and nuclear genomes underlying physiological diversification in a grass species. *Proceedings of the Royal Society B*, **287**, 20201960 (2020).

# Figures and Tables

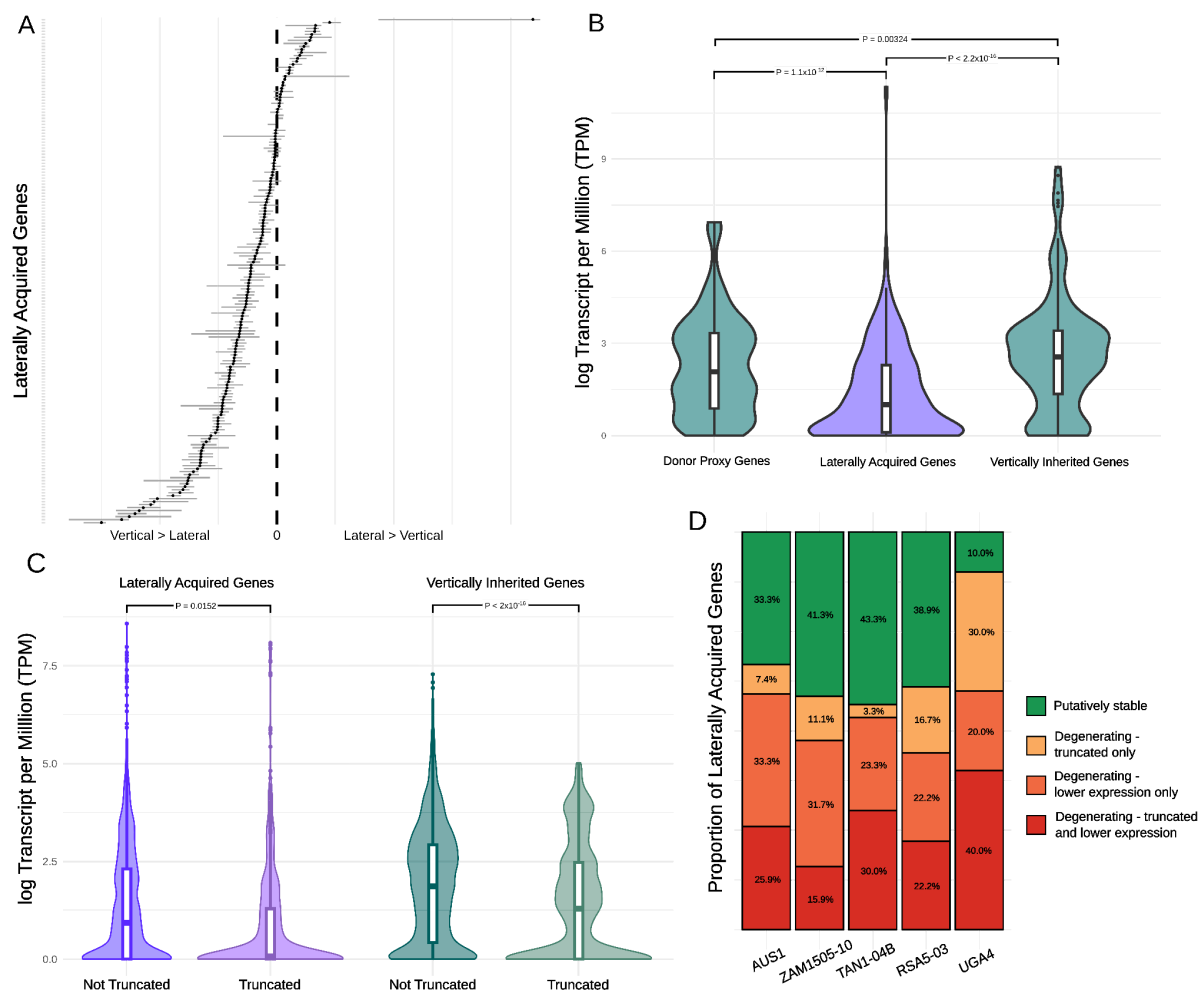
**Table 1: The number of laterally acquired genes identified in *Alloteropsis*.**

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, were present in the original genome annotation, and were not 100% identical to another sequence. For each accession the number of unique laterally acquired genes is shown, and the number of genes this corresponds to in the genome annotation is given in parenthesis (this counts duplicates that could have arisen prior or post lateral transfer separately). The total row 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 five reference genomes, and their duplicates, separately.

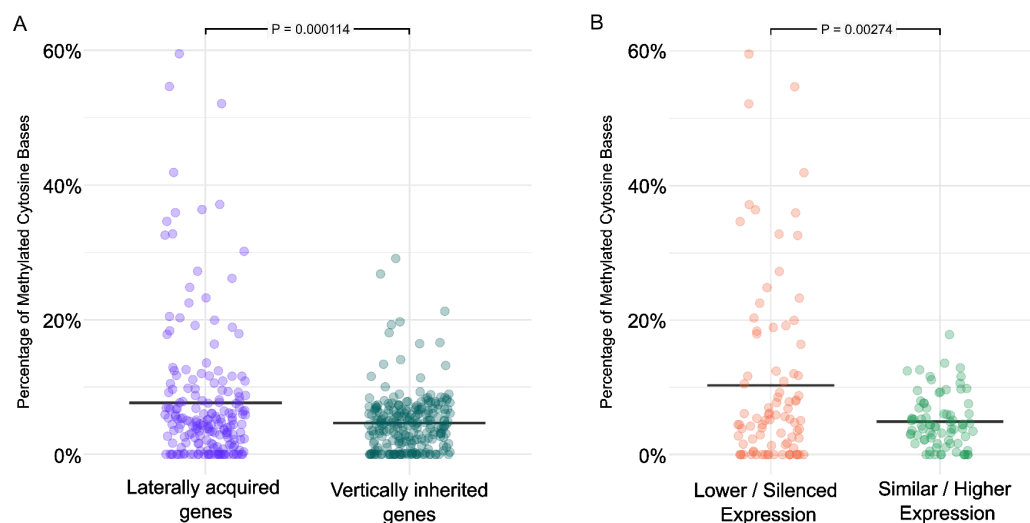
Species	Accession	Location	Raimondeau et at. (2023)		
			Read mapping	phylogenetics	Used in this study
<i>A. semialata</i>	AUS1	Australia	50	50 [53]	50 [52]
	RSA5-03	South Africa	33	29 [31]	26 [28]
	TAN1-04B	Tanzania	45	43 [51]	43 [51]
	ZAM1505-10	Zambia	100	95 [125]	94 [124]
<i>A. angusta</i>	UGA4	Uganda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



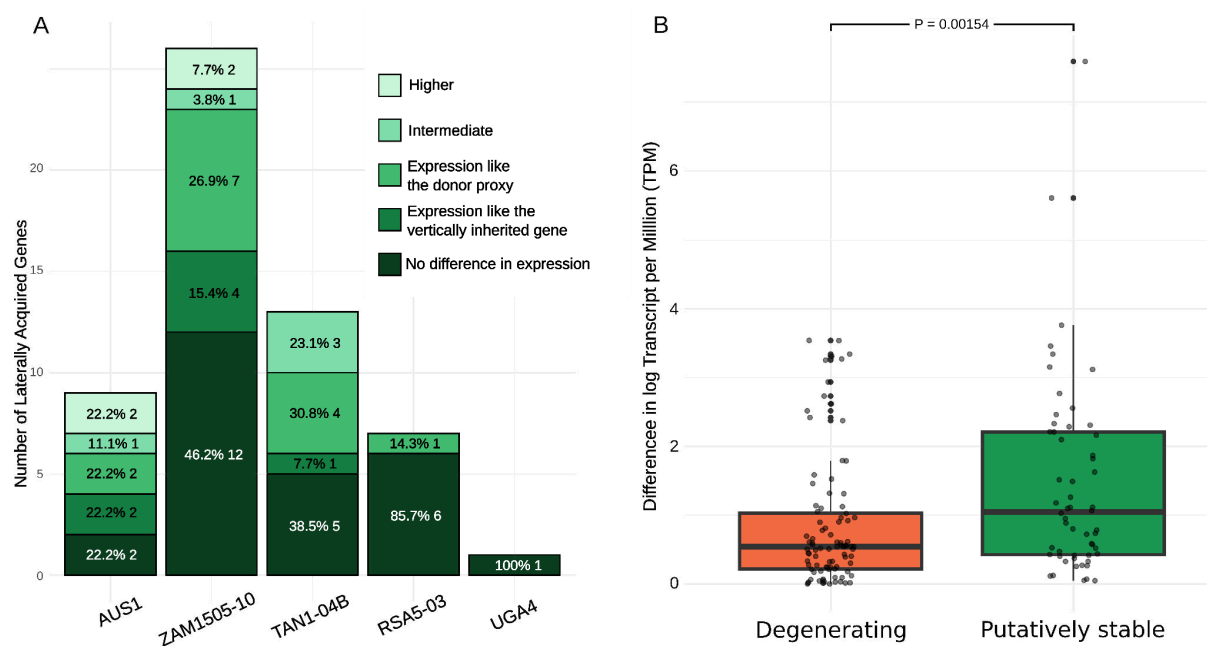
**Figure 1:** Expression of laterally acquired genes in *Alloteropsis*. A phylogenetic tree adapted from Bianconi et al.<sup>67</sup> shows the evolutionary relationships of the *Alloteropsis* accessions, with the number of laterally acquired genes in each genome indicated (n). The dots with crosses show the percentage of genes where at least one replicate passes the expression threshold (TPM > 0.5) in any tissue. Boxplots show the median and interquartile range of the number of laterally acquired genes expressed in each individual RNA-seq dataset.



**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 xenologs, 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 xenologs. C) Comparison in log-transformed expression levels between truncated and non-truncated genes, for laterally acquired and vertically inherited homologs. D) Laterally acquired genes were divided into four 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).

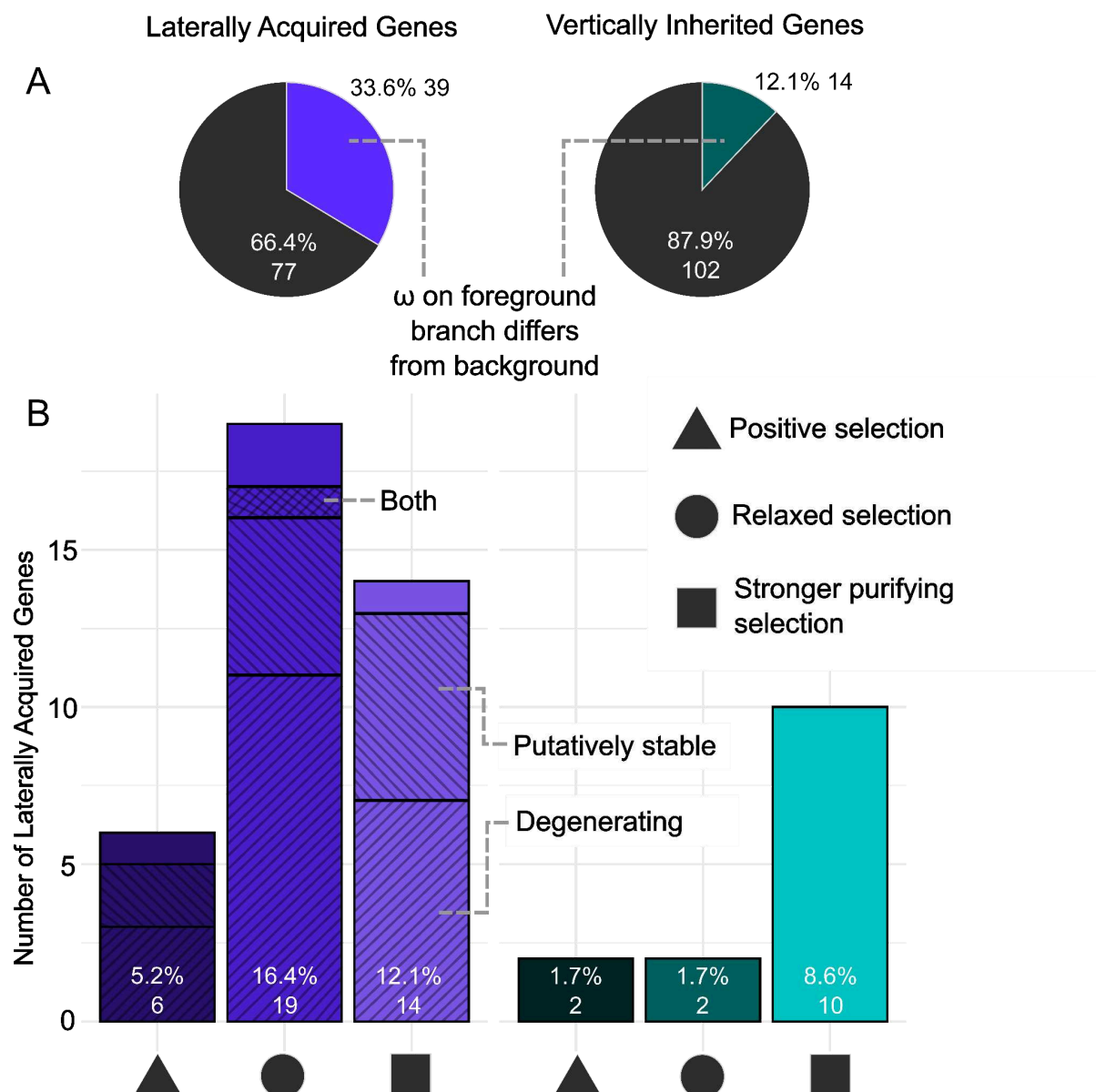


*Figure 3: Methylation of laterally acquired genes. Comparison of proportion of methylated cytosine sites across the full gene region between a) laterally acquired and vertically inherited genes and b) laterally acquired genes with silenced or lower expression than their corresponding vertically inherited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



**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 xenolog, 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 xenologs. 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’. Comparison of log transformed expression counts between vertically inherited and donor proxy genes corresponding to laterally acquired genes classified as ‘Degenerating’ or ‘Putatively stable’. P value from Wilcoxon Rank Sum test is shown above boxplots.





**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 categorised as: stronger purifying selection, where the dN/dS ratio ( $\omega$ ) for the gene is lower than the background  $\omega$ ; positive selection, where  $\omega$  for the gene is higher than the background  $\omega$  and evidence for sites under positive selection has been found using the PAML CodeML branch site model ( $p < 0.05$ ); and relaxed selection, where  $\omega$  for the gene is greater than the background  $\omega$  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' categorisation. 'Both' indicates where categorisation differed across accessions. No hash indicates genes where a donor copy was not present.

# Supplementary Information for

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

Catherine F Collins, Benjamin T Alston, Samuel GS Hibdige, Pauline Raimondeau, Emily Baker, Graciela Sotelo, Alexander S. T. Papadopoulos, Pascal-Antoine Christin, Lara Pereira,  
Luke T Dunning

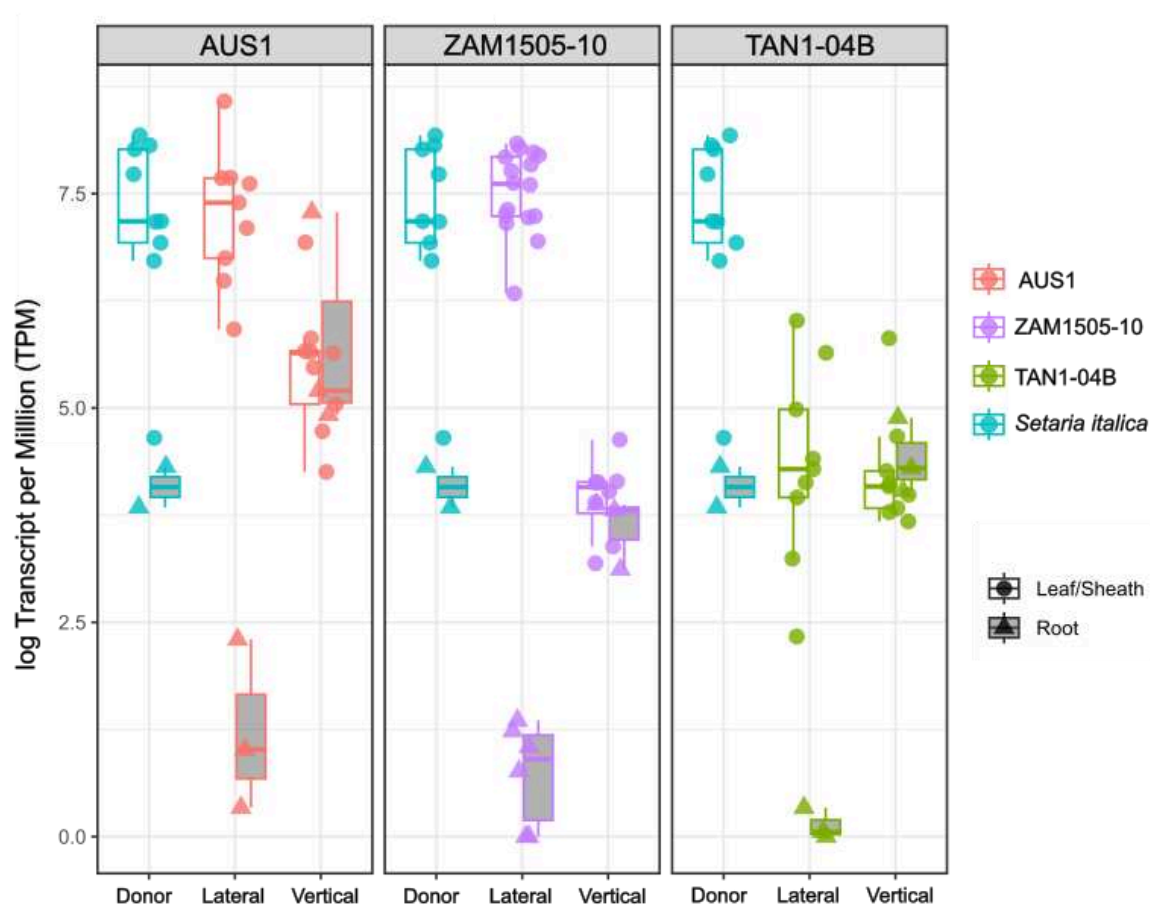
**This Supplementary Information includes:**

Supplementary Figures 1-4

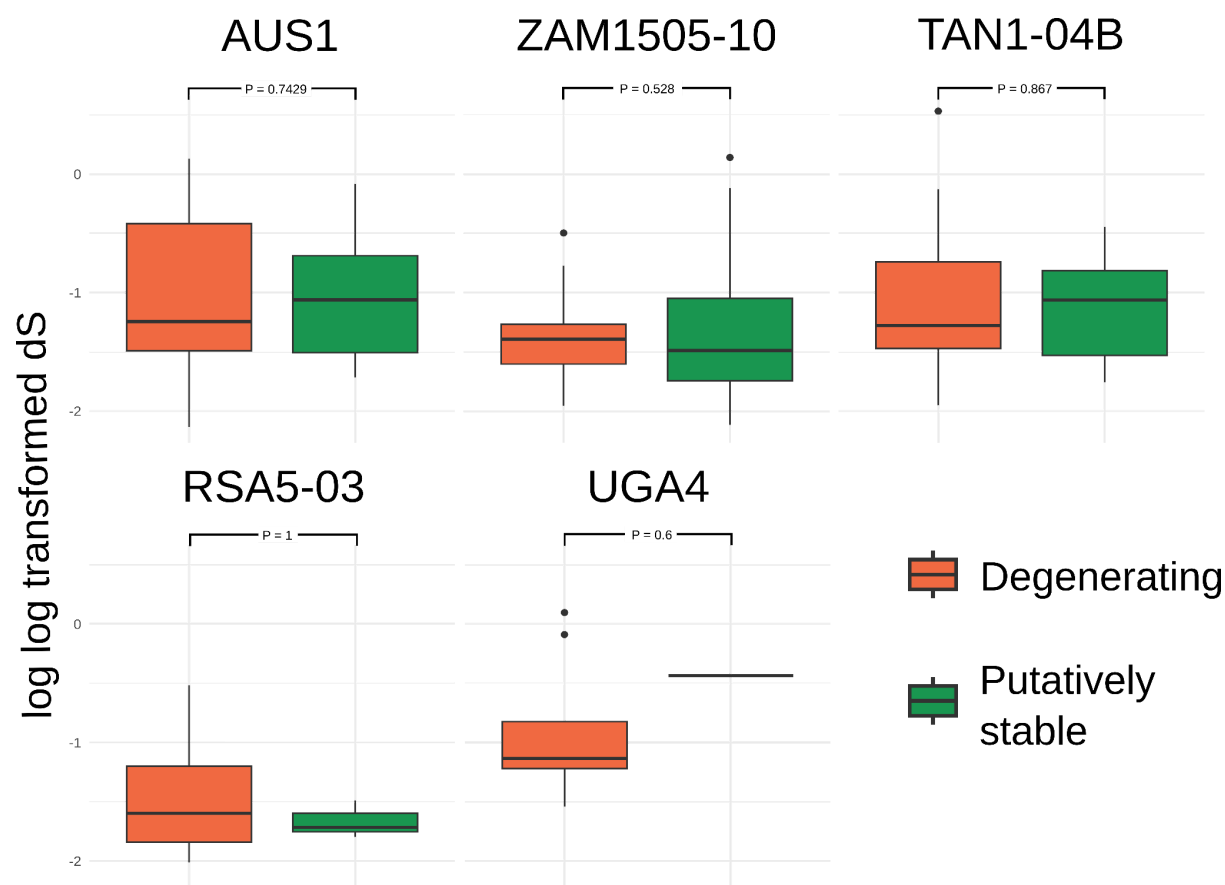
Supplementary Tables 1-8 are in a separated excel file.

## Supplementary Figures

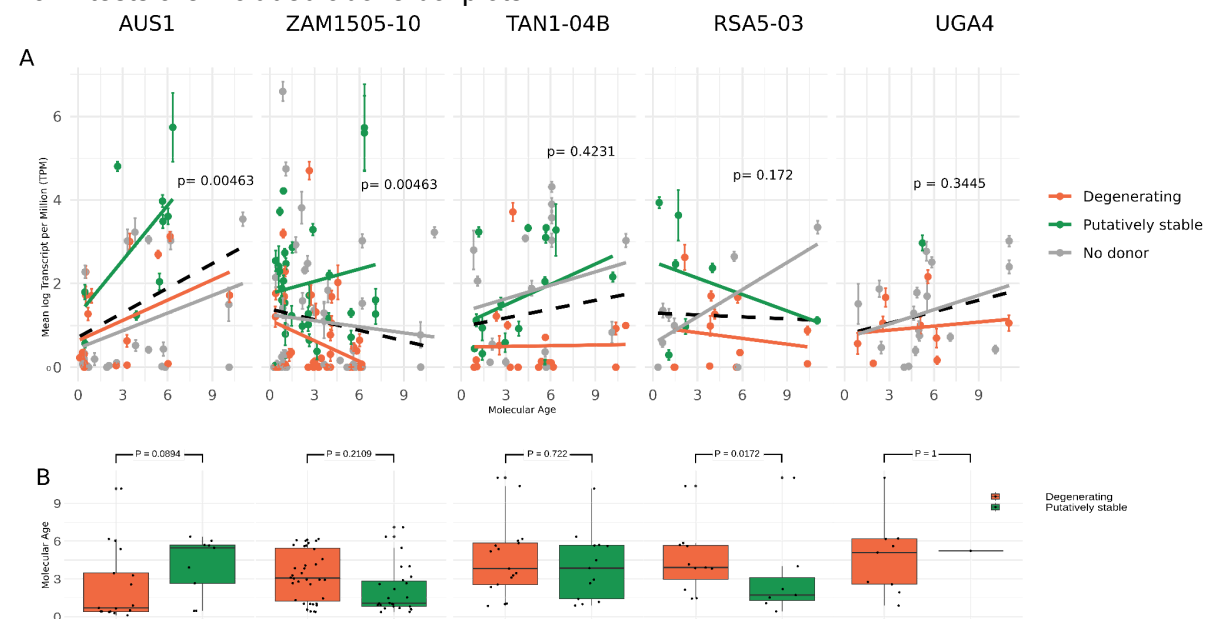
**Supplementary Figure 1** - Expression patterns of LGT-019, phosphoenolpyruvate carboxykinase (PCK). Boxplots show median and interquartile range



**Supplementary Figure 2:** Genetic divergence between corresponding vertically inherited and donor proxy genes for laterally acquired genes classified as ‘Degenerating’ and ‘Putatively stable’. dS indicates synonymous substitutions between the gene pairs; this value is twice log transformed to create a normal distribution of the data.



**Supplementary Figure 3** - Time since acquisition and expression patterns of laterally acquired genes. A) Relationship between age and expression. Error bars indicate variation across replicates per laterally acquired gene. Black dashed line indicates the linear relationship between age and TPM, with p values from linear mixed effects models included. Coloured lines indicate trends for laterally acquired genes classified as 'degenerating' or 'putatively stable'. No donor indicates laterally acquired genes where a donor proxy gene was not found in the gene phylogeny. B) Boxplots showing the range of molecular age across laterally acquired genes within expression categories. P values from Wilcoxon Signed Rank tests are included above boxplots.



**Supplementary Figure 4:** Principal component analysis (PCA) of RNA-seq data demonstrates tissue effect between the leaf and sheath tissue types (red and blue respectively) and the roots (green). Aang = *Alloteropsis angusta*, Asem = *Alloteropsis semialata*, SET = *Setaria italica*, THE = *Themeda triandra*.

