**Department of Animal and Plant Sciences**

**University of Sheffield**

Yellow stickers

**Registration Number**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 8 | 0 | 1 | 6 | 3 | 9 | 3 | 1 |

**Module number**

|  |  |
| --- | --- |
| APS | 61013 |

**Title**

|  |
| --- |
| Understanding the regulation of laterally transferred grass genes requires gene-specific context. |

**Word count**

|  |
| --- |
| 6294 |

*By submitting this document I certify that the work submitted is all my own, except where specifically indicated (e.g. by referencing, or quotation), and confirm that I have read and understood the University's rules relating to plagiarism. I understand that the Department reserves the right to run spot checks on all coursework using plagiarism software.*

**Abstract**

The regulation of genes occurs through interactions of cis-regulatory elements, which are physically linked to the gene and trans-regulatory elements which act on distant genes. Typically, these two elements evolve in parallel and share evolutionary history with the gene they act upon. However, in the event of lateral gene transfer (LGT) – a mechanism of gene transfer which can overcome species boundaries – the newly acquired gene, and possibly its associated cis-regulatory elements will have a mismatched phylogeny relative to the recipients’ trans-regulatory elements. This raises the possibility of unique patterns of gene expression. Among eukaryotes, LGT events are especially prevalent between grasses, such as the tropical grass *Alloteropsis semialata* where laterally transferred genes (LGTs) have been previously identified and linked to donors. We test whether *A. semialata* LGTs are uniquely regulated by quantifying and comparing their expression to orthologous native genes of the recipient *A. semialata* and one of two donor lineages. We first find that the expression of LGTs is not closer to either native orthologue, suggesting there is no cis-trans regulatory bias. Secondly, we find that the expression of LGTs is not uniquely diverged relative to their donor orthologues. However, various LGTs have expression values distinctly closer to either the recipient or donor, and the most expressed of all LGTs was closely related to a previously identified PEP carboxykinase gene important for C4 photosynthesis. While further studies of cis-trans regulatory effects may still prove useful in the context of LGTs, these results suggest gene-specific contexts are perhaps more important.

**Keywords:** Poaceae; horizontal gene transfer; gene expression; phylogenetics; regulatory divergence

**Introduction**

In evolution, the acquisition of novel traits can occur not only through changes in gene function (Kaessmann, 2010), but also by regulatory changes in gene expression (López-Maury *et al.,* 2008). Generally, the regulation of genes relies on the co-ordinated response of cis-regulatory elements, which are physically linked to neighbouring genes (Biłas *et al.* 2016), and trans-regulatory elements which can act on distant genes (Gilad, Rifkin and Pritchard, 2008). In vertical evolution, these two regulatory elements are known to evolve in parallel (Camino et al., 2015). However, in the case of interspecific gene transfer, the new gene and possibly cis-regulatory elements if they transferred as part of non-coding DNA (Kaplinsky *et al*., 2002; Dunning *et al.,* 2019)will have a distinct evolutionary history relative to the hosts regulatory elements, raising the potential of unique expression patterns.

Lateral gene transfer (LGT) is one mechanism in which interspecific gene transfer can occur, and it refers to the movement of genes between individuals without direct parent-offspring inheritance (Danchin, 2016). It is most widely studied in prokaryotes, where up to 60% of a populations collective genome (pangenome) can derive from LGT events (Freschi *et al*., 2019) and in bacteria provide major adaptive benefits such as antibiotic resistance, increased virulence and heightened resistance against heat and oxidative stressors (Ochman, Lawrence and Groisman, 2000; Deng *et al.,* 2019; Kamal *et al.,* 2021). Laterally transferred genes (LGTs) can also move from prokaryotes to unicellular eukaryotes and provide adaptive benefits; with prokaryote derived LGTs estimated to account for less than 0.5% of a yeast pangenome (Soanes and Richards, 2014), ~1% of red algae pangenomes (Rossoni *et al.,* 2019) and 0.16-1.44% of phytoplankton pangenomes (Fan *et al.,* 2020). Additionally, LGT within eukaryotes has been found to occur between fungi (Reynolds *et al.,* 2018), from plants to animals (Xia *et al.,* 2021) and between plants (Wickell and Li, 2019). Plant-to-plant transfers can have major evolutionary importance, such as the acquisition of a unique bryophyte photoreceptor by ferns (Li *et al.,* 2014) and genes which improve parasitism in host-parasite relationships (Yang *et al.,* 2016; 2019; Vogel *et al.,* 2018). However, LGT between grasses are perhaps the most well studied plant-to-plant transfer (Vallenback et al., 2008; Christin *et al.,* 2012a; 2012b; Prentice *et al.,* 2015; Olofsson *et al.,* 2016; 2019; Mahelka *et al.,* 2017; 2021; Dunning *et al.,* 2019; Bianconi *et al.,* 2020; Phansopa *et al.,* 2020; Hibdige *et al.,* 2021; Raimondeau *et al*., *unpublished*) and include genes with putative roles in photosynthesis, disease resistance and abiotic stress tolerance (Christin *et al.,* 2012a; Dunning *et al.,* 2019; Mahelka *et al.,* 2021; Hibdige *et al.,* 2021). However, the regulatory consequences of LGTs in grasses remains unknown.

To understand how LGTs are regulated in grasses, we quantify and compare the expression of LGTs to the expression of related native genes (orthologues) in both the donor and recipient genomes. We use the tropical grass species *Alloteropsis semialata* (Gibbs-Russell, 1983) to measure the expression of LGTs and recipient genes, as the highest number of grass LGTs have been recorded in this genus (Hibdige *et al.,* 2021; Raimondeau *et al., unpublished*). To quantify donor expression, we chose members of the Cenchrinae, a subtribe within the tribe Paniceae, which also contains *Alloteropsis* (Olofsson *et al.,* 2016) and Andropogoneae, a tribe which diverged from Paniceae 25~ mya (Christin *et al.,* 2012a) as they donate LGTs most frequently to *Alloteropsis* (Hibdige *et al*., 2021). Specifically, we used *Setaria italica* (Cenchrinae) which diverged from *Alloteropsis* 20~ mya and *Themeda triandra* (Andropogoneae) as these genera have been identified as direct donors of LGTs to *Alloteropsis* (Christin *et al.,* 2012a). While previous studies have analysed the expression of LGTs in *Alloteropsis* (Dunning *et al.,* 2019), it has not yet been compared to the donor species, preventing an evaluation of the regulatory effects of LGT.

In this study we aim to address this gap in two ways. First, by investigating whether LGTs are closer in expression to the native genes of the recipient or donor. If the expression of LGTs are conserved after LGT, and more influenced by cis-regulatory elements (transferred within the same genomic fragment), we would expect the expression of LGTs to be closer to donor genes. Conversely, if the expression is more dependent on trans-regulatory elements, we would expect the expression of LGTs to be closer to recipient genes. Secondly, we compare the divergence rate of LGTs and donor orthologues, if the divergence rate is unchanged, we expect the divergence of LGTs and recipient orthologues to be similar to that of donor and recipient orthologues. Conversely, if LGTs undergo rapid changes in expression, we expect LGTs to have changed more than expected based on divergence times.

**Materials and methods**

*Species selection*

Species were selected based on a study by Raimondeau *et al.* (*unpublished*) who previously identified and matched LGTs received by *Alloteropsis semialata* to various donor lineages. LGTs derived from the Andropogoneae tribe and Cenchrinae subtribe were chosen as the focus of our study. To measure the expression of LGTs and native *A. semialata* genes three geographically distinct accessions of *A. semialata* were used, originating from Australia (AUS1), South Africa (KWT3) and Zambia (ZAM15-05-10). Descriptions of the AUS1 and KWT3 accessions can be found in Lundgren *et al.* (2015), while descriptions of the ZAM15-05-10 accession can be found in Lundgren *et al.* (2019) and Bianconi *et al*. (2020). For measurement of donor orthologues, *Themeda triandra* was used as a representative for Andropogoneae, while *Setaria italica* was used as a representative for Cenchrinae.

*RNA extraction, library generation and sequence preparation.*

RNA was extracted from the roots and leaf tips of plants grown in controlled conditions at the Arthur Willis Environmental Centre, Sheffield as previously described by Moreno-Villena *et al.* (2018; also see Dunning *et al.* 2019). Three clones were produced for each *A. semialata* accession and for both donor species, although at the time of writing sequencing remains to be done for the roots of one *S. italica* triplicate, and the roots and leaf tip of one *T. triandra* triplicate. Individual cDNA libraries were prepared using the TruSeq RNA Library Preparation Kit v2 (Illumina, San Diego, CA, USA) and were sequenced as paired-end reads on a HiSeq 2500 platform (as described in Moreno-Villena *et al.* 2018). The quality of sequencing data was then checked using FastQC v 0.11.9 (Andrews, 2022), with all raw reads showing high quality scores. After this initial check, reads were then trimmed using Trimmomatic v 0.39 (Bolger, Lohse and Usadel, 2014) in order to remove sequencing adapters at the start of reads. Trimmed reads were then quality checked once more using FastQC.

*Reference genome alignment and extraction of expression data.*

Reference genomes were composed as coding sequences (Table 1) and paired end reads for the tip leaf and roots of each plant were mapped to their respective reference genomes using Bowtie2 v 2.4.4 (Langmead and Salzberg, 2012). The reference genome used for the AUS1 accession was generated in Dunning *et al*. (2019), while the reference genomes used for KWT3 and ZAM15-05-10 were generated in Raimondeau *et al.* (*unpublished*). For the donors, the *S. italica* reference genome was first published in (Bennetzen *et al*., 2012). *T. triandra* has no published reference genome, so the reference genome of *Sorghum bicolor*, another member of the Andropogoneae was used instead (Paterson *et al.,* 2009; McCormick *et al*., 2017). Alignments were run on two different *S. bicolor* coding sequences for *T. triandra* to match genes identified in Raimondeau *et al*. (*unpublished*). After samples were aligned to reference genomes, expression values (FPKM) were quantified using eXpress v 1.5.1 (Roberts and Pachter, 2013) and individual genes were assigned to orthogroups based on phylogenies outlined by Raimondeau *et al.* (*unpublished*). A list of scripts used for quantifying expression data and a table containing expression data for all genes used in our study can be found in the supplementary material.

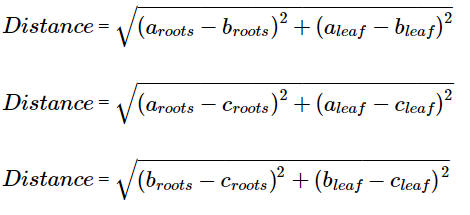
|  |  |  |
| --- | --- | --- |
| **Table 1.** Reference genomes used in this study | | |
| Species | Reference genome(s) | Availability |
| *A. semialata*  (AUS1) | ASEM\_C4\_v1.0.CDS.fasta | Dunning *et al.* (2019) |
| *A. semialata* (KWT3) | KWT\_v1.0.all.maker.CDS.fasta | Raimondeau *et al.* (*unpublished*) |
| *A. semialata* (ZAM15-05-10) | ASEM\_ZAM15-05-10.v1.0.all.maker.cds.fasta | Raimondeau *et al.* (*unpublished*) |
| *S. italica* | Setaria\_italica.JGIv2.0.cds.all.fa | http://ftp.ensemblgenomes.org/pub/plants/release-37/fasta/ |
| *T. triandra* | Sorghum\_bicolor.Sorghum\_bicolor\_v2.cds.all.fa  Sbicolor\_454\_v3.1.1.cds.fa | <http://ftp.ensemblgenomes.org/pub/plants/release-37/fasta/>  https://phytozome-next.jgi.doe.gov/info/Sbicolor\_v3\_1\_1 |

*Statistical analysis*

All statistical analysis and data visualisation were done using R v 4.1.0 (R Core Team, 2021) with the suite of packages found in ‘tidyverse’ v 1.3.1 (Wickham *et al.,* 2019). The R script and all required files can be found in the supplementary material. As data was not normalised only non-parametric statistical tests were used. Kruskal-Wallis tests were performed to check if the distribution of LGTs and native genes among *A. semialata* accessions was consistent. Spearman’s rank correlation tests were used to calculate the proportion of shared variance (*ρ2*) for root and tip leaf measurements. For LGTs three different correlations were calculated to match orthogroups shared by native *A. semialata*, *T. triandra* or *S. italica*. This was done to examine whether correlations of root and tip leaf measurements in LGTs changed with the donor or were more similar to native *A. semialata* genes. To assess whether expression was higher in the roots or tip leaf, Wilcoxon-signed rank tests were used and individual tests were performed for LGTs derived from Cenchrinae and Andropogoneae. Wilcoxon signed-rank tests were also used to compare whether LGT expression was significantly different from both donor and recipient orthologues, and to compare donors to recipients. As LGTs and native genes varied in frequency, data was filtered independently for each paired test, to ensure the tests used the largest sample size possible.

*Distance calculations*

To measure distances between orthologues the following calculation was used:



**Eq 1.** Calculating the expression distances between orthogroups.

Where the overall expression distance of two different orthogroups (‘a’ and ‘b’) can be calculated by the square root of the additive sum of ‘a’ minus ‘b’ squared for both root and tip leaf expression. This test was performed between each orthogroup type (donor, recipient, LGT) and independently for *S. italica* and *T. triandra* donors. The sum of distance values in this test is always positive, allowing for easier visualisation of data.

*Identification of highly expressed genes*

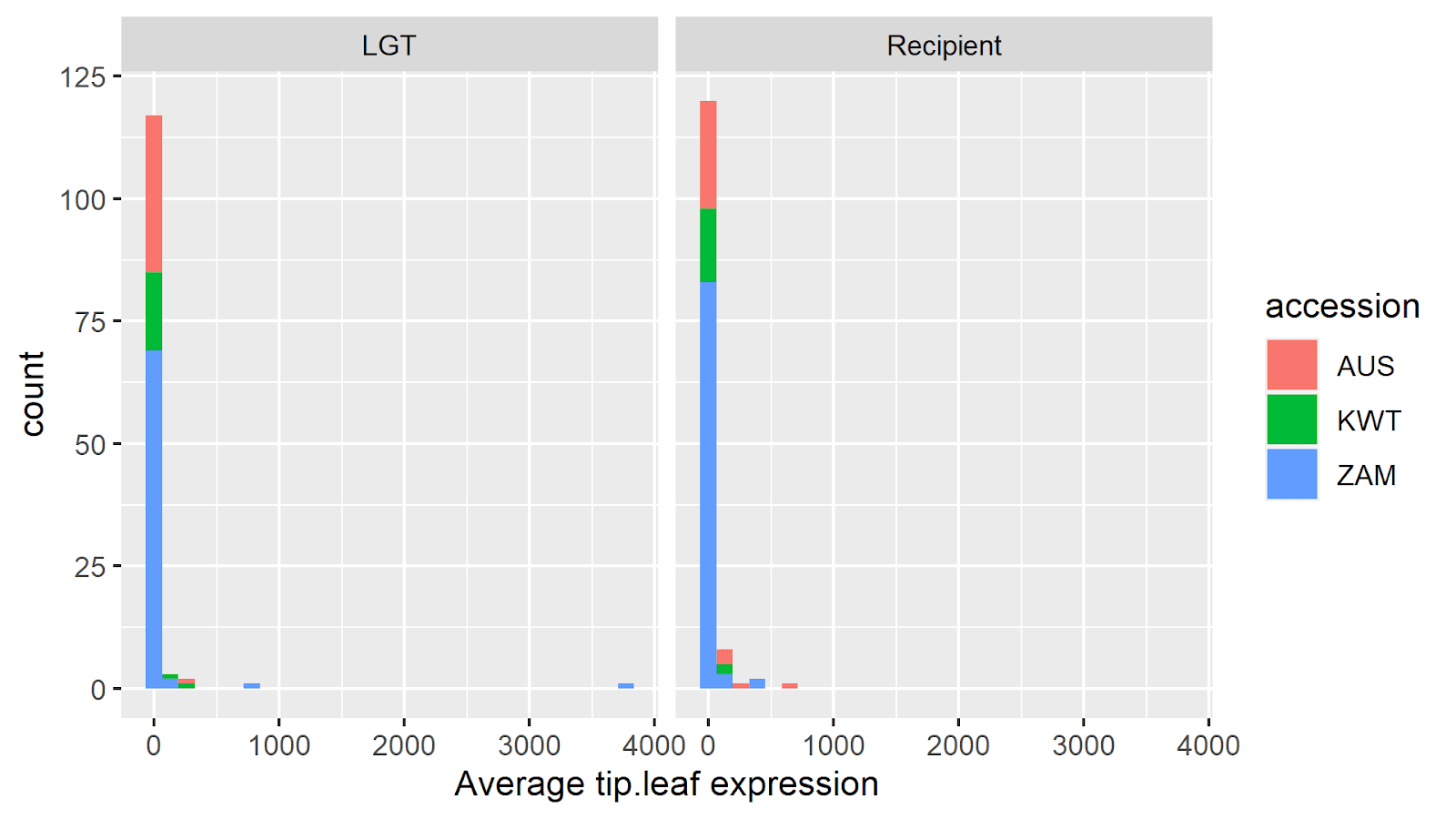
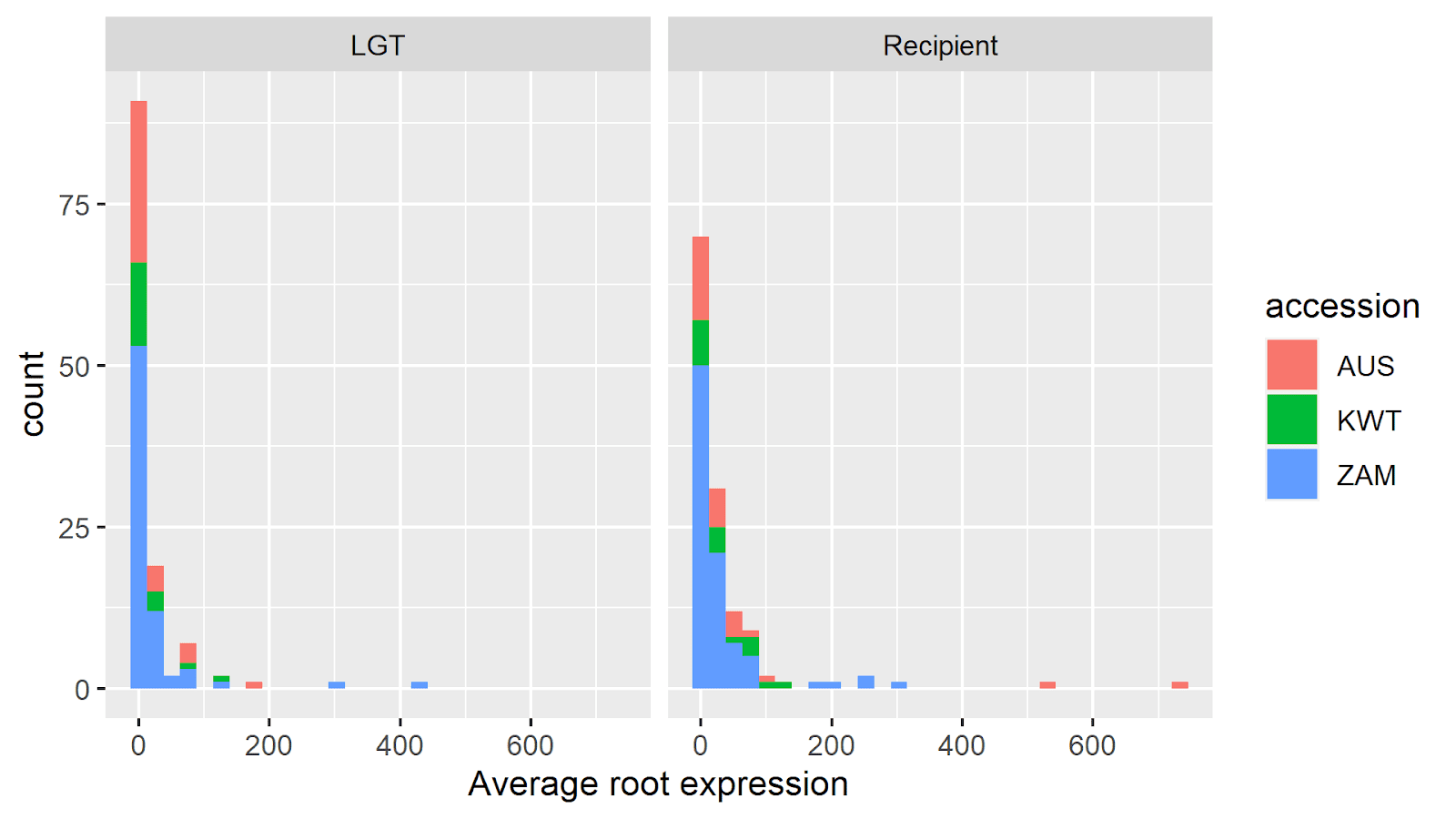
After quantification of expression levels, the most highly expressed genes from our results (> 250 FPKM) were assessed for functional significance using NCBI BLAST (Sayers *et al.,* 2021).

**Results**

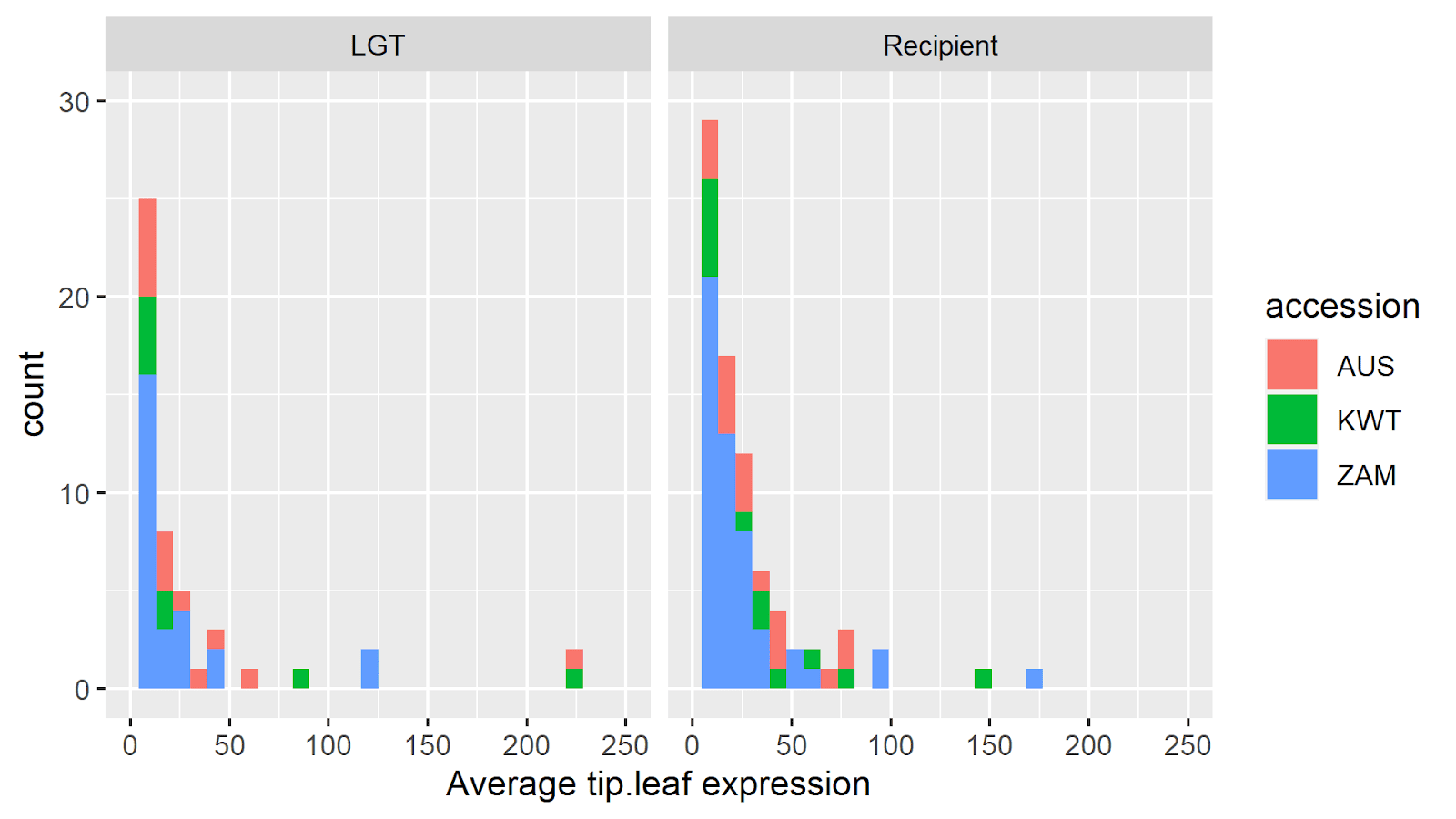
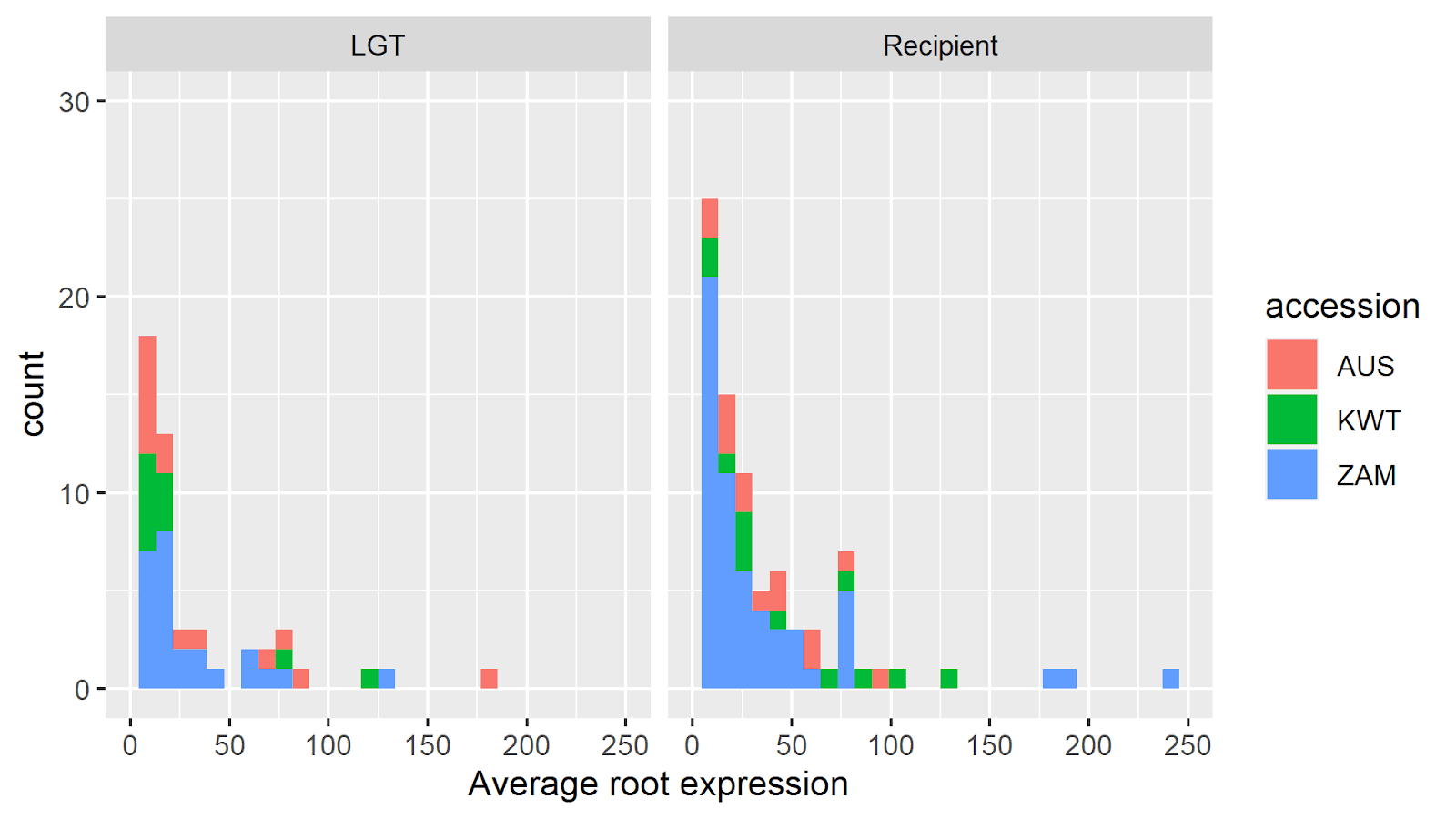
*Expression is consistent between A. semialata accessions*

In total, 124 LGTs and 132 recipient orthologues belonging to 101 orthogroups were found to be paired across three *A. semialata* accessions, with ZAM15-05 having the highest number of LGTs (n = 73), followed by AUS1 (n = 33) and KWT3 (n = 18). Among accessions, variation in the expression of LGTs was higher in the tip leaf than the roots and highest in ZAM15-05, where tip leaf expression reached 3772 FPKM in orthogroup 25 and root expression reached 430 FPKM in orthogroup 83 (Fig 1). Overall expression levels did not vary significantly across accessions for recipient genes in the roots (Kruskal-Wallis test: H(2) = 1.8117, p = 0.404) or leaves (Kruskal-Wallis test: H(2) = 4.4877, p = 0.106). Similarly, no significant differences in expression were found for LGTs among accessions in the roots (Kruskal-Wallis test: H(2) = 1.7932, p = 0.408) or leaves (Kruskal-Wallis test: H(2) = 0.5298, p = 0.767). These patterns suggest that expression patterns across the three accessions are similar, justifying their pooling for later tests. However, it should be noted that gene-specific patterns are apparent, as various orthogroups vary in expression across accessions (Fig 2).

**A B**

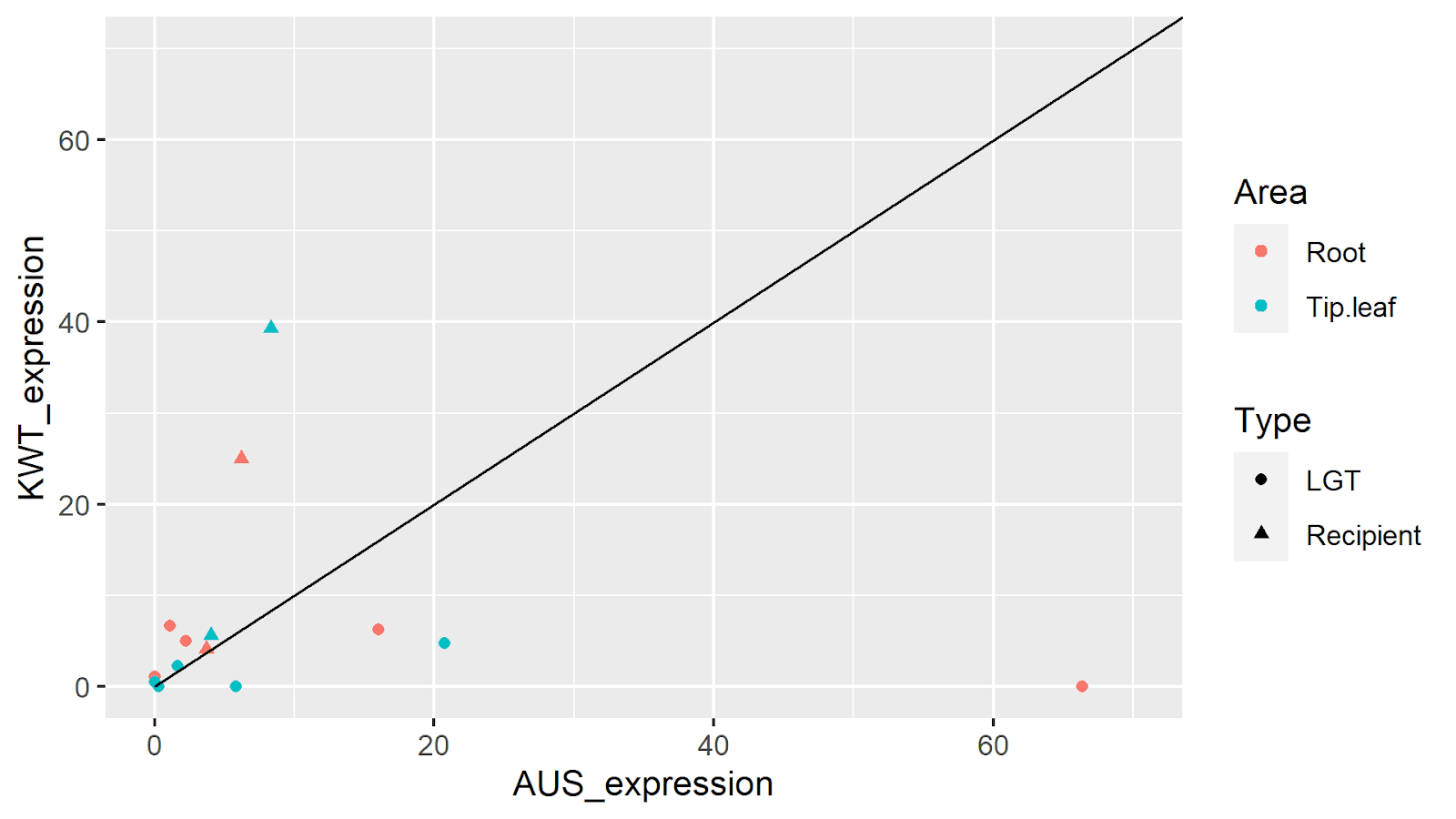


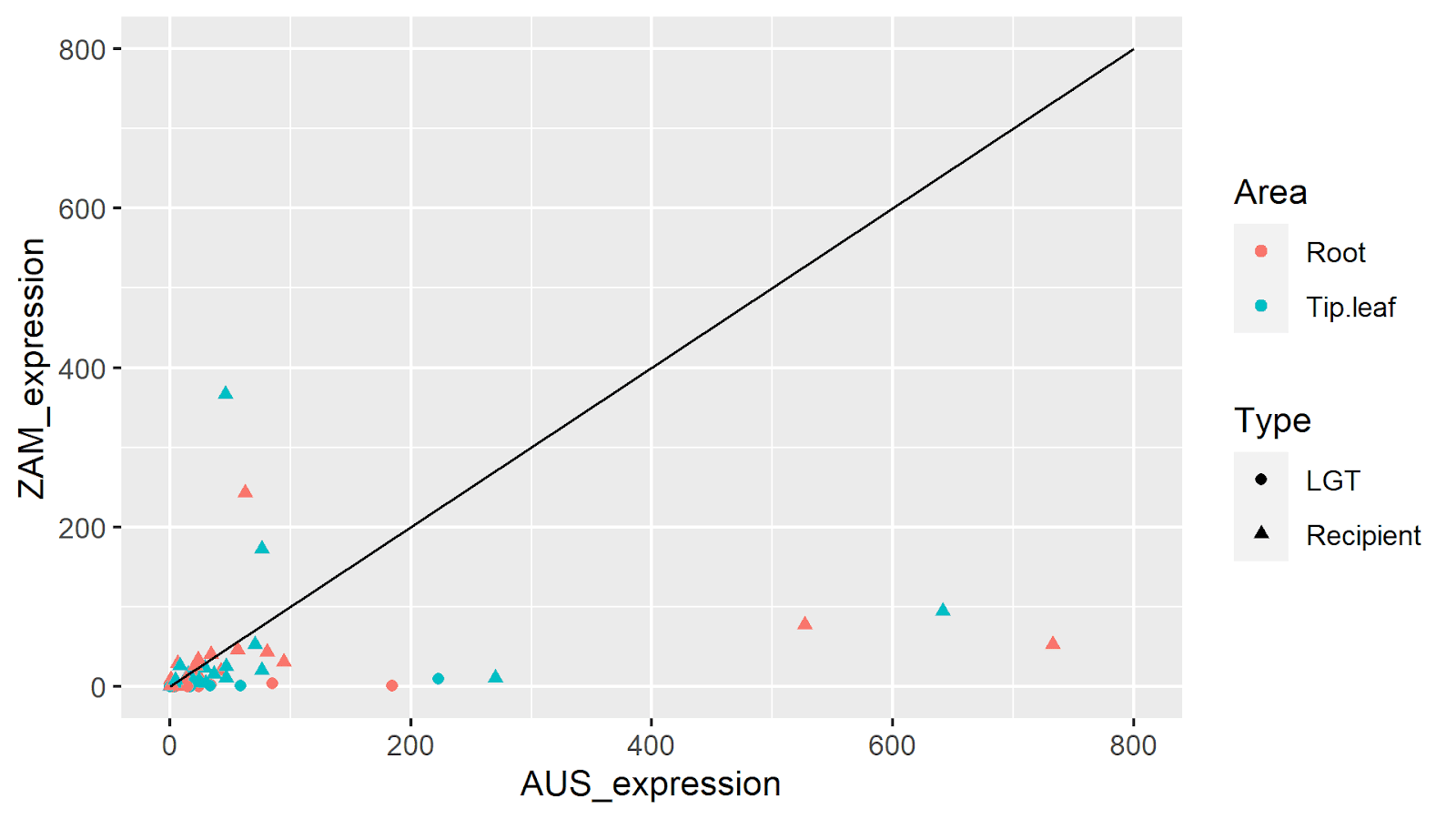
**C D**

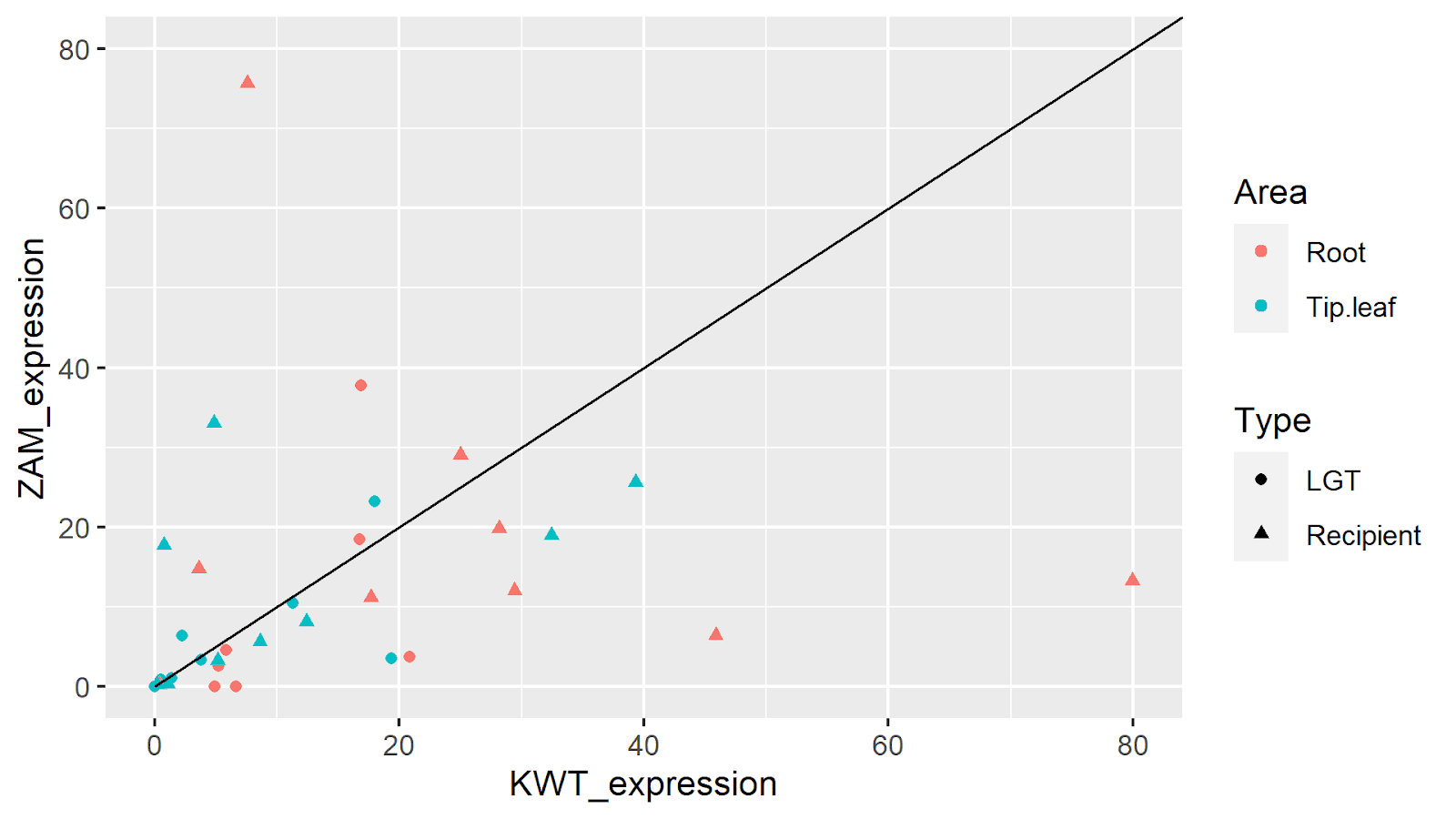


**Figure 1.** The distribution of root or tip leaf expression measured in FPKM (X axis) for recipient genes and LGTs across *A. semialata* accessions. Count (Y axis) refers to the total number of genes matching a particular interval of expression. Counts are coloured by accession, and count values are cumulative. Panels **A** and **B** show the full range of values, while for clarity purposes **C** and **D** are limited to genes with expression below 250 FPKM.

**A**

**B**

**C**



**Figure 2.** A comparison of expression (FPKM) between different accessions of *A. semialata*. Only orthogroups present in **(A)** AUS1 and KWT3 (n = 5); **(B)** AUS1 and ZAM15-05 (n = 12); or **(C)** KWT3 and ZAM15-05 (n = 9) are used in each respective plot. Each point represents an orthologous gene found in both accessions and is colour-coded by the area expression data was recorded. Shapes indicate whether genes are LGTs or native to the recipient.

*Root and leaf expression correlations are consistent across orthologues.*

Variation in root and tip leaf expression (FPKM) (Fig 3) were similar for LGTs when filtering for orthogroups which also had native expression data for *A. semialata* (Spearman’s rank correlation: LGTs - S = 71035, p = < 0.001, *ρ2* = 0.60; *A. semialata* - S = 103305, p = < 0.001, *ρ2* = 0.53), *S. italica* (Spearman’s rank correlation: LGTs - S = 25816, p = < 0.001, *ρ2* = 0.61; *S. italica* -S = 15809, p = < 0.001, *ρ2* = 0.51) and *T. triandra* (Spearman’s rank correlation: LGTs - S = 2607.7, p = < 0.001, *ρ2* = 0.57; *T. triandra* - = 1278.4, p = < 0.001, *ρ2* = 0.65). Additionally, root expression was found to be higher for both native genes (Wilcoxon signed-rank test:  V = 4969, p = 0.011) and LGTs (Wilcoxon signed-rank test: V = 3592, p = 0.003) in *A. semialata*. When filtered by donor, root expression was only significantly higher for LGTs derived from Cenchrinae (Wilcoxon signed-rank test: V = 1677, p = < 0.001) but not Andropogoneae (Wilcoxon signed-rank test: V = 365, p = 0.621). Neither *S. italica* (Wilcoxon signed-rank test: V = 1523, p = 0.0597) or *T. triandra* (Wilcoxon signed-rank test: V = 488, p = 0.460) were found to have significantly higher root expression. When comparing the interquartile range (IQR) of root and tip leaf expression data however, LGTs derived from Andropogoneae were the only group which had a higher IQR for tip leaf expression than root expression (Table 2).

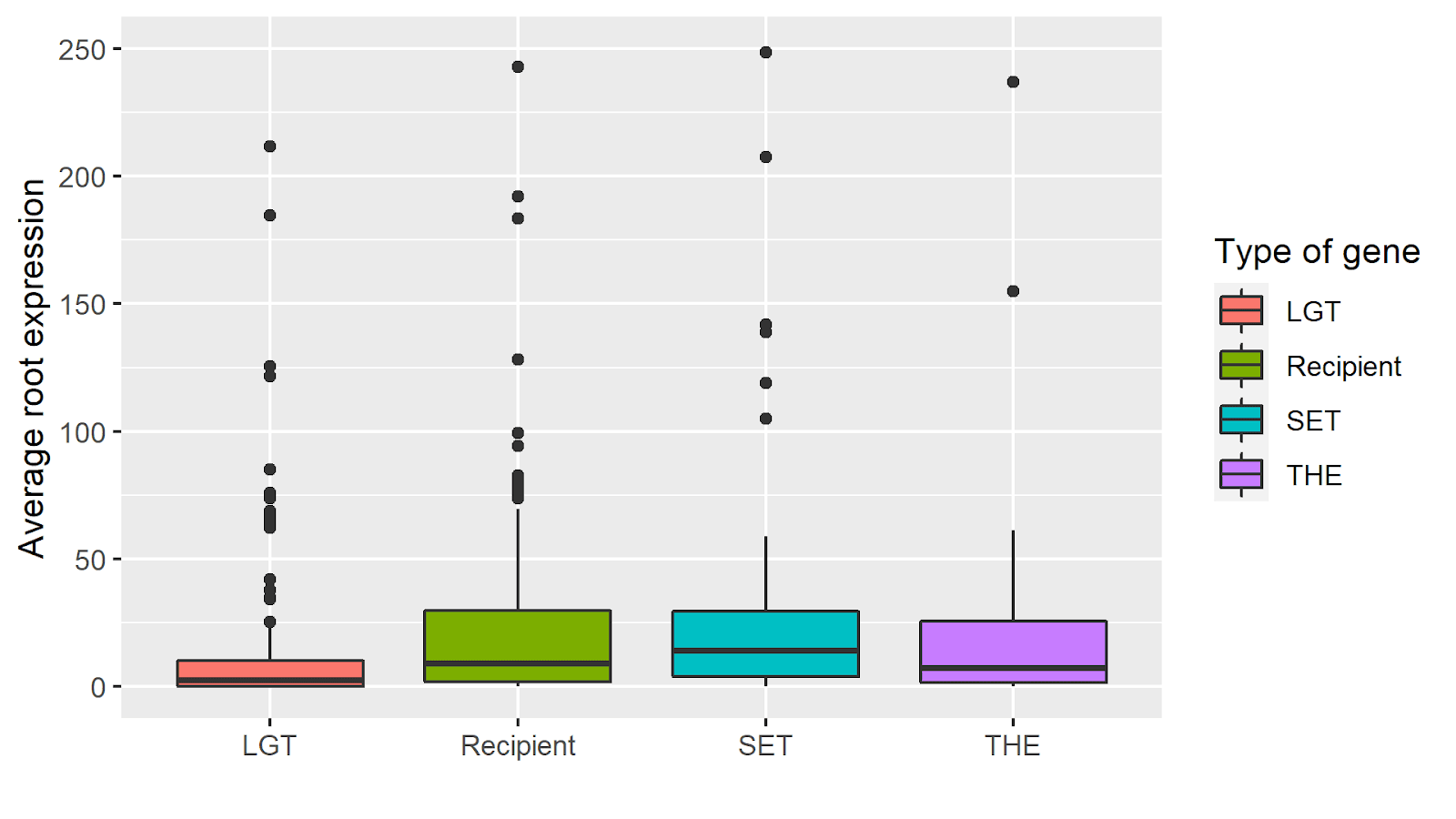


**Figure 3.** A comparison of root (Y axis) and tip leaf (X axis) expression (FPKM) for *A. semialata* (recipient/LGT), *Setaria italica* (SET) and *T. triandra* (THE). Each point represents an individual gene, colour coded to its source. For increased clarity, genes above 250 FPKM are excluded.

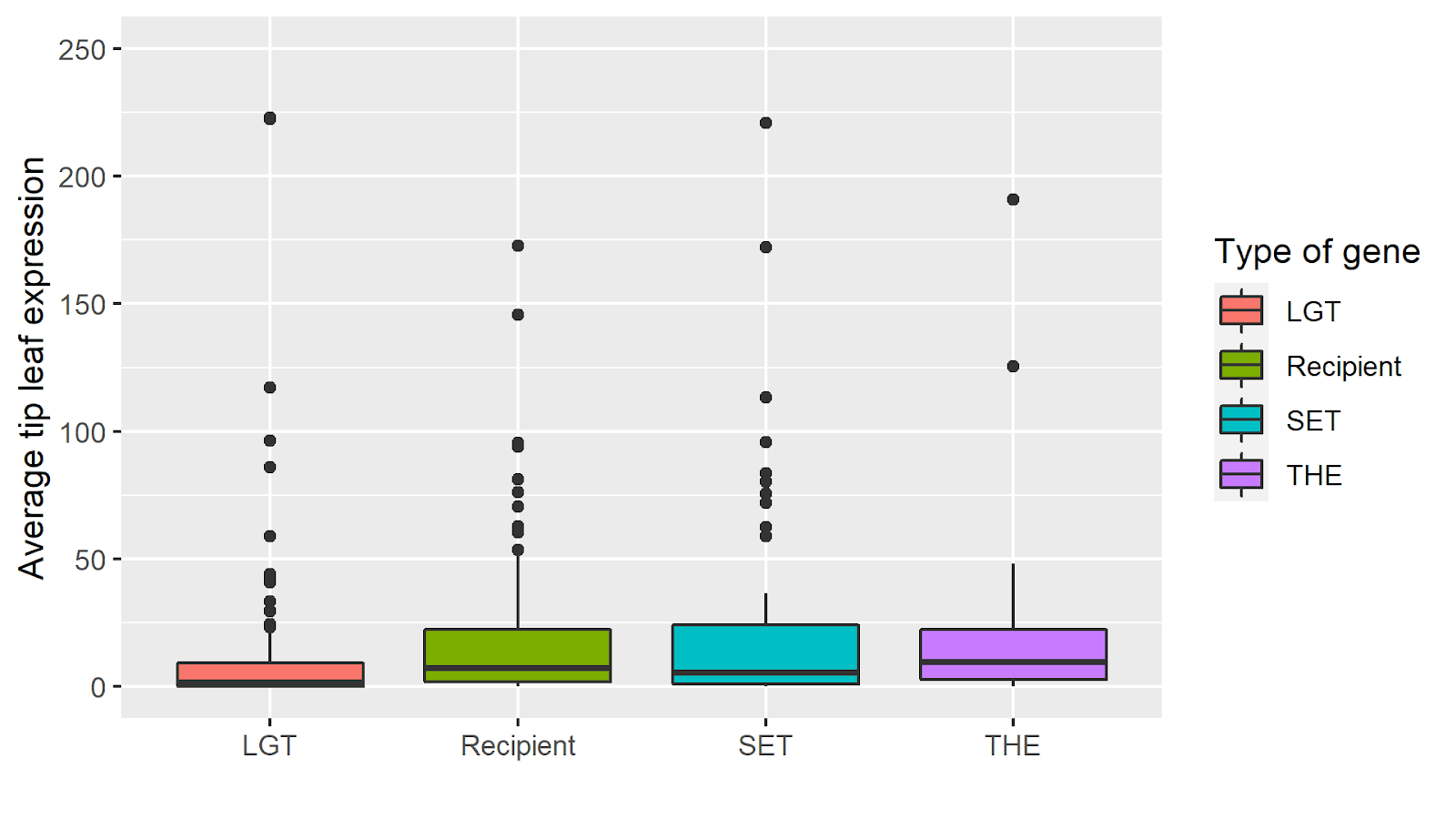
*LGT expression is consistently lower than native gene expression.*

LGTs were found to have significantly lower expression than all three native orthologues for both the roots (Wilcoxon signed-rank test: *A. semialata* - V = 2097, p = 0.001; *S. italica* - V = 1046, p = 0.002; *T. triandra* - V = 125, p = < 0.001) and leaves (Wilcoxon signed-rank test: *A. semialata -* W = 2030, p < 0.001; *S. italica -* V = 1185, p = 0.017*; T. triandra -* V = 159, p = 0.004), with a higher percentage of non-expressed genes than equivalent native orthologues (Table 2). However, this trend is not absolute as various LGTs exceed the expression of their associated native orthologues (Fig 4). Lastly, there was no significant expression differences between donors and recipients in the roots (Wilcoxon signed-rank test: V = 3820, p = 0.987) or tip leaf (Wilcoxon signed-rank test: V = 3524, p = 0.668).

**A**



**B**



**Figure 4.** Average root **(A)** and tip leaf **(B)** expression (Y axis) for LGTs, recipient genes and donors (*S. italica* and *T. triandra*) (X axis). Colours indicate the type of gene/orthologue. A total of nine genes had expression values >250 FPKM but were not included in the figures to improve clarity. Boxplots indicate median values and interquartile ranges, with points indicating values above the upper quartile range.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 2.** Patterns of gene expression (FPKM) in orthologues shared by LGTs and native genes. | | | | |
| Source of gene | Root expression (IQR) | Tip leaf expression (IQR) | Percentage of genes with zero root expression | Percentage of genes with zero tip leaf expression |
| *A. semialata* (recipient) | 30.32 | 21.67 | 6.8% | 7.6% |
| LGT (Andropogoneae) | 9.91 | 11.36 | 19.2% | 19.2% |
| *T. triandra* | 33.75 | 21.04 | 7% | 11.6% |
| LGT (Cenchrinae) | 12.09 | 8.01 | 27.1% | 30.2% |
| *S. italica* | 26.62 | 23.27 | 9.6% | 11% |

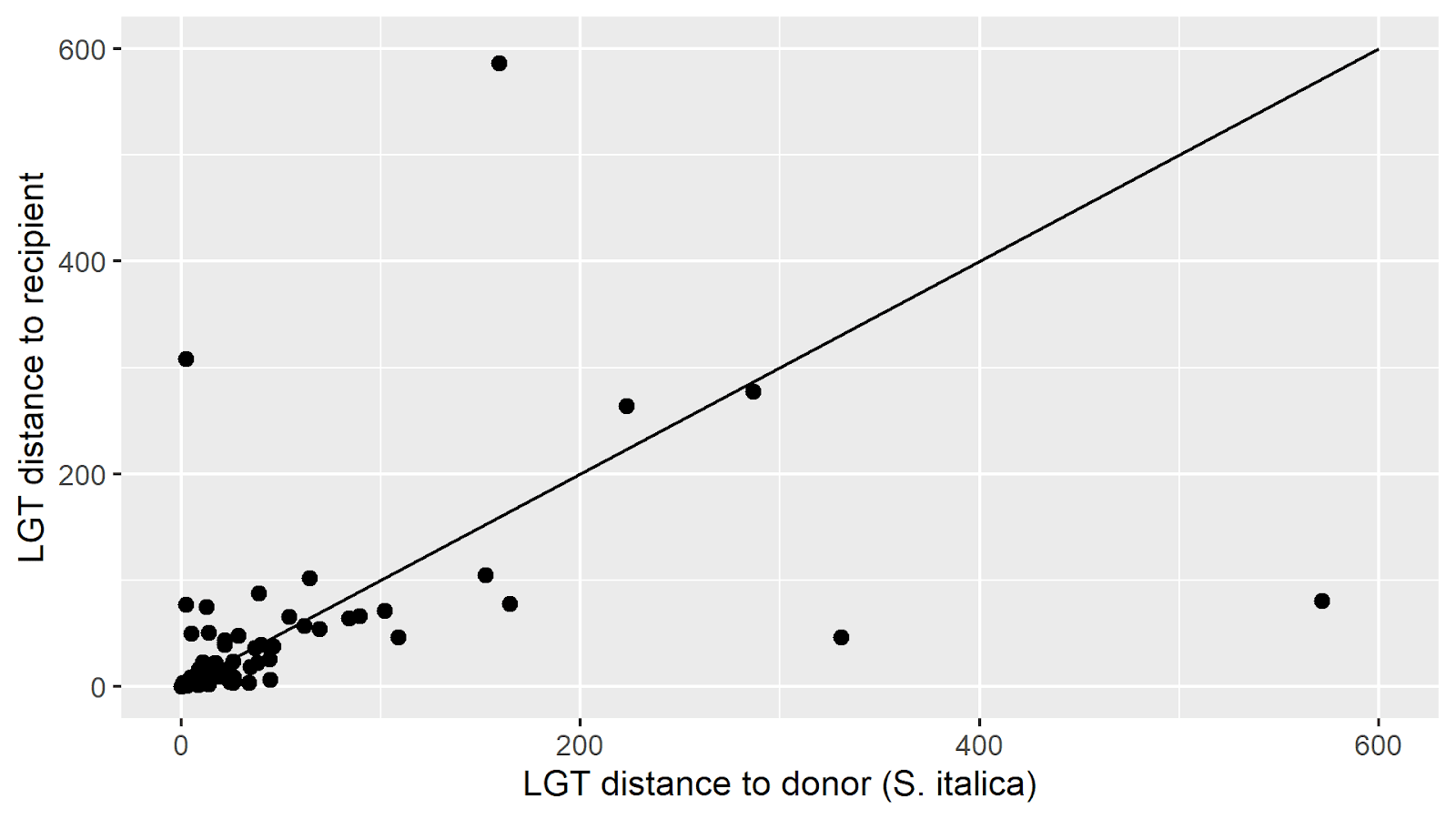
Note – Recipient *A. semialata* genes were filtered to only account for orthologues which also had expression data for both LGTs and donors (n = 118). LGTs were filtered twice, first for genes derived from Andropogoneae (n = 48) which had orthologous genes detected in *T. triandra* (n = 43) and secondly for genes derived from Cenchrinae (n = 96) which had orthologous genes detected in *S. italica* (n = 73). The total number of LGTs belonging to an orthologue was sometimes higher than donors, as some LGTs were found in multiple accessions.

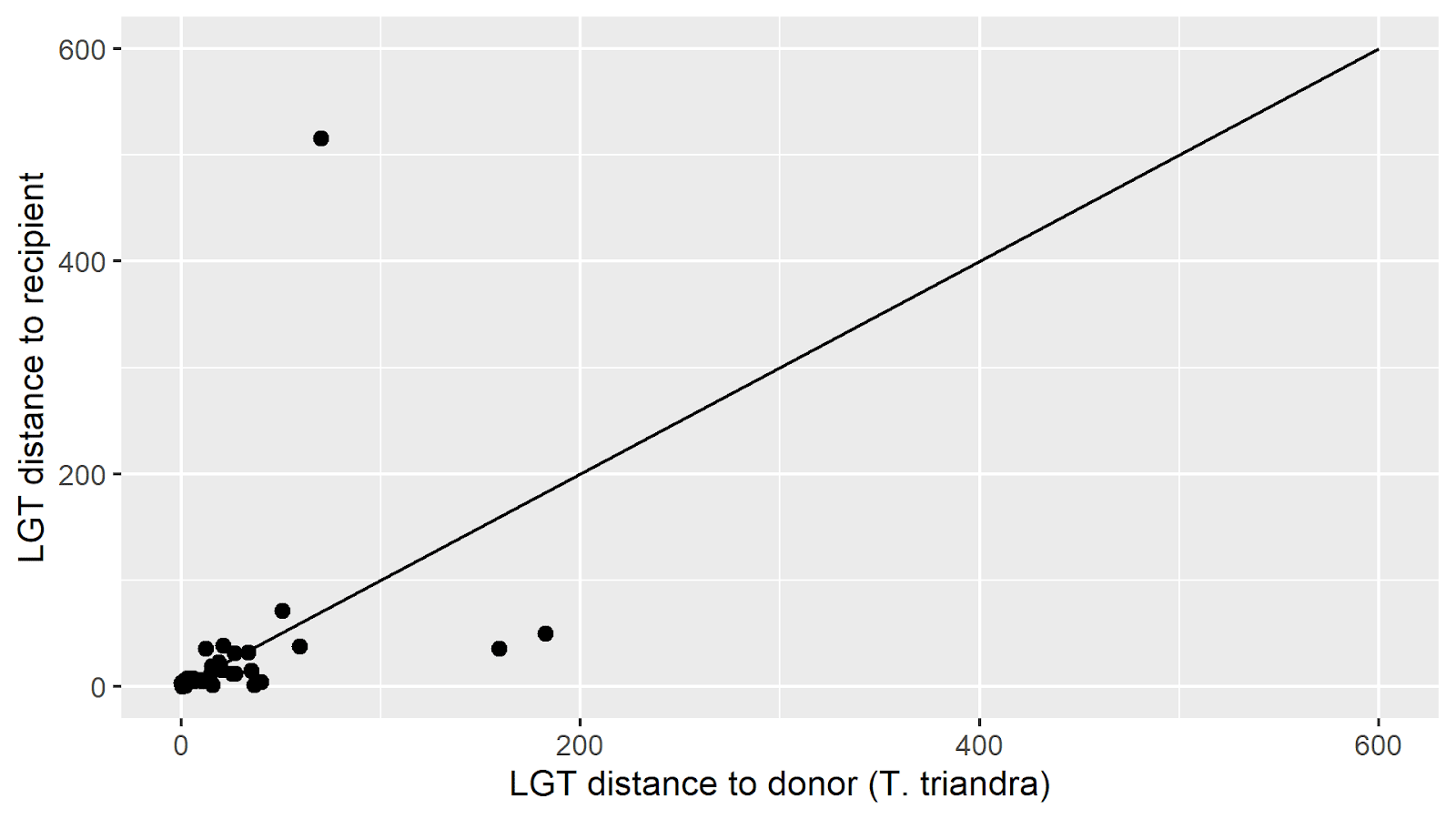
*LGT expression patterns are not more similar to the recipient or donor.*

When comparing the expression distance (FPKM) of LGTs to native *A. semialata* (recipient) and *S. italica* orthologues (n = 74), three LGTs were equally distant from both *A. semialata* and *S. italica* (two of which had zero expression across all orthologues), 41 LGTs were closer in expression to the native genes of *A. semialata* and 30 LGTs were closer in expression to *S. italica*. This difference was not significant (Fig 5A; Wilcoxon signed-rank test: V = 1189, p = 0.612). However, when comparing the expression distance of either the recipient or *S. italica* donor orthologue with the orthogroup’s combined average distance to the LGT, we found only 39% of LGTs had distances to either orthologue within 20% of the total average distance. In contrast 16% of LGTs were more than twice as close in expression to their respective *S. italica* orthologues, while 20% of LGTs were more than twice as close to their respective native *A. semialata* orthologues (Table 3). Similar patterns were found in comparisons of native orthologues of *A. semialata* and *T. triandra* (n = 35), where 19 LGTs were closer to *A. semialata* and 16 LGTs closer to *T. triandra*. Again, overall distances were not significantly different (Fig 5B; Wilcoxon signed-rank test: V = 255, p = 0.334), despite the fact only 34% of LGTs had a distance to either orthologue within 20% of their average distance to both. Conversely, 23% of LGTs were at least twice as close to their respective donor orthologue than recipient, and 34% of LGTs were at least twice as close to their recipient orthologue (Table 3).

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 3.** Expression distances of LGTs to donor and native *A. semialata* (recipient) orthologues. | | | |
| Distance comparisons (recipient x donor) | LGTs distance to recipient/donors within 20% of total average distance. | LGTs more than twice as close to donor than recipient | LGTs more than twice as close to recipient than donor |
| *A. semialata x S. italica* (n = 74) | 39% | 16% | 20% |
| *A. semialata x T. triandra* (n = 35) | 34% | 23% | 34% |

Note – percentage values represent the proportion of orthogroups which fit the associated pattern.

**A****B**

****

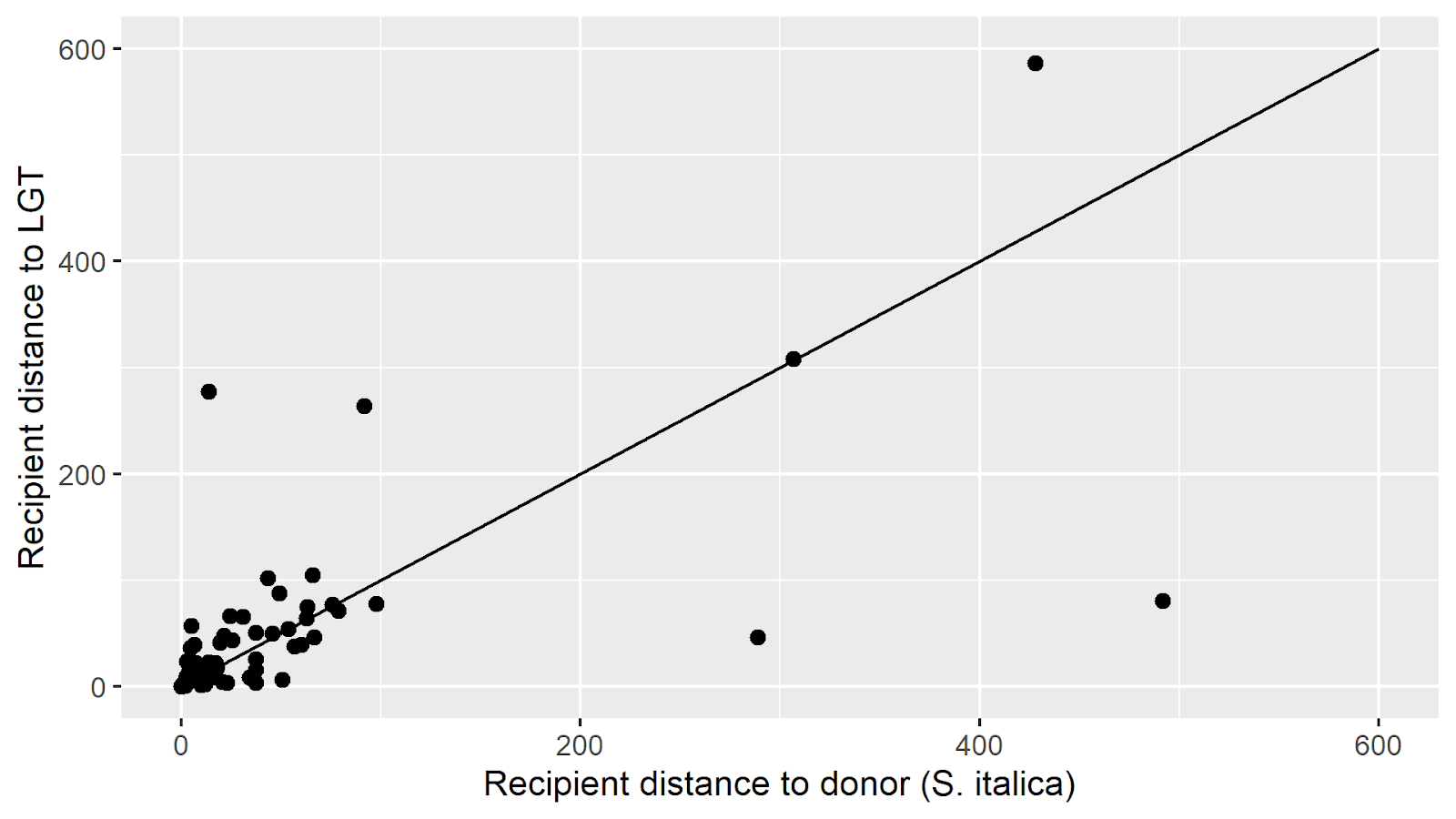
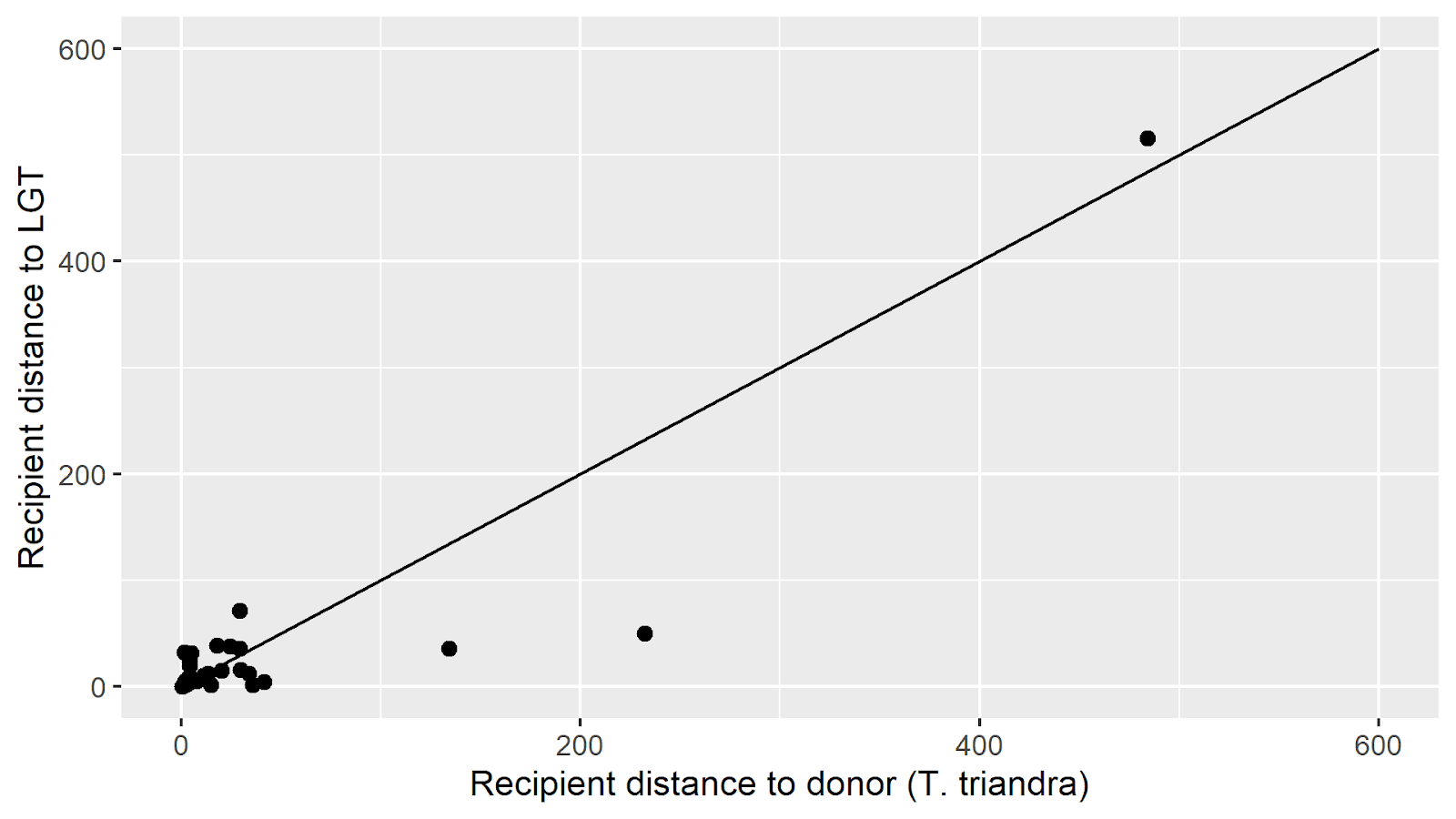
**Figure 5.** The distance of LGTs to recipients (Y axis) and donor orthologues (X axis) associated with **(A)** *S. italica* and **(B)** *T. triandra*. The line shows equidistance to donor and recipient orthologs. Values below the line indicate the LGT is closer in expression to the recipient, while values above the line indicate the LGT is closer in expression to the donor.

*The expression of LGTs is not uniquely diverged when compared to donor orthologues.*

As LGTs only began diverging from donor orthologues at the time of their transfer, distance plots were made to assess whether the LGTs were still as different from the native *A. semialata* orthologues as donor genes were in terms of expression (FPKM). In *A. semialata* x *S. italica* comparisons (n = 74) a total of 46 recipient genes were closer to donor orthologues, while 25 recipient genes were closer to their LGTs in terms of expression. This difference was not significant (Fig 6A; Wilcoxon signed-rank test: V = 1663, p = 0.0505). Although, when comparing the expression distance of either LGTs or *S. italica* donor orthologues with the orthogroup’s combined average distance to the recipient gene, we found that only 22% of recipient genes had distances to either orthologue within 20% of the total average distance. In contrast, 30% of recipients across orthogroups were at least twice as close in expression to their associated donor than their associated LGT, and 16% of recipients were at least twice as close in expression to their associated LGT than donor (Table 4). Comparatively, in *A. semialata* x *T. triandra* (n = 35), a total of 16 recipient genes were closer to donor orthologues, while 18 recipient genes were closer to LGTs in expression. This difference was not significant (Fig 6B; Wilcoxon signed-rank test; V = 283, p = 0.8108), despite the fact only 29% of recipient orthologues had an expression distance to LGTs/donors within 20% of their average distance to both. Meanwhile, in 23% of orthologues, recipients were at least twice as close in expression to their associated donor than LGT, while 20% of orthologues were twice as close in expression to their respective LGT than their respective donor (Table 4).

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 4.** Expression distances of native *A. semialata* (recipient) genes to LGTs and donor orthologues. | | | |
| Distance comparisons (LGT x donor) | Recipient distance to LGTs and donor orthologues within 20% of total average distance. | Recipient at least twice as close to donor than LGT | Recipient at least twice as close to LGT than donor. |
| *A. semialata x S. italica* (n = 74) | 22% | 30% | 16% |
| *A. semialata x T. triandra* (n = 35) | 29% | 23% | 20% |

Note – percentage values represent the proportion of orthogroups which fit the associated pattern.

**A****B**

**Figure 6.** The distance of recipient orthologues to LGTs (Y axis) and donors (X axis) associated with **(A)** *S. italica* and **(B)** *T. triandra*. The line shows equidistance of recipient genes to donor and LGT orthologues. Values below the line indicate the recipient expression is closer to the LGT, while values above the line indicate the recipient expression is closer to the donor.

**Discussion**

Here we quantified and compared the expression of LGTs and their orthologous recipient (*A. semialata*) and donor (*S. italica* for Cenchrinae and *T. triandra* for Andropogoneae) genes to understand the regulatory dynamics of LGT in grasses.

We first found that on average the expression of LGTs was not more similar to either recipient or donor orthologues. This implies for most LGTs neither cis nor trans-regulatory elements play a more influential role in regulating the expression of LGTs (Fig 5). While the novelty of this study means we cannot directly compare our results to past research, studies investigating the expression of genes in interspecific hybrids of *Arabidopsis* (Shi *et al.,* 2012), cotton (Bao *et al.,* 2019) and *Drosophila* (Landry *et al.,* 2005) have shown that some genes can be subject to compensatory cis and trans-regulatory effects. This can lead to expression mismatches for hybrid genes, relative to parent orthologues (Landry *et al.,* 2005), which could be a potential explanation for our results. Although, despite these general observations, 57% of Andropogoneae-derived LGTs and 36% of Cenchrinae-derived LGTs are distinctly closer to either the recipient or donor orthologue (Table 3). This diversity in expression could be explained as the contribution of cis and trans-elements to expression is known to vary across genes, and at times can provide opposing effects on regulation (Landry *et al.,* 2005; Gordon and Ruvinsky, 2012; Coolon *et al.,* 2014; Osada *et al.,* 2017; Mattioli *et al.,* 2020). For example, cis-regulatory elements are known to facilitate expression via promoters and enhancers (Merli *et al.,* 1996; Mitsuhara *et al.,* 1996, Zheng *et al.,* 2015), or reduce expression via silencers (van der Meer *et al.,* 1992; Liu *et al.,* 2020). In addition, the complexity of cis-trans regulation could also be dependent on the size of the genomic fragment transferred through LGT, as genomic fragments in *A.* semialata often contain non-coding regions where cis-regulatory elements are found (Dunning *et al.,* 2019). However, as the dynamics of hybridisation vastly differ from LGT, further studies are required to identify regulatory biases and potential cis/trans compensatory effects in the context of grass LGTs.

Our second aim was to assess whether LGTs were differently diverged relative to donor orthologues. We investigated this by comparing the expression distance of LGTs and donors to recipient orthologues, finding that neither LGTs nor donors were closer to recipients (Fig 6). This suggests that the expression divergence of LGTs is not unique, although as more LGTs will likely be identified in future studies, repeating this experiment may prove valuable for Cenchrinae-derived orthogroups as we calculated a p value of 0.505. Regardless of general observations however, recipient genes in 43% of Andropogoneae-associated orthogroups and 46% of Cenchrinae-associated orthogroups were distinctly closer to either LGTs or donor orthologues (Table 4). Divergences in the expression of functional genes have been previously identified in grasses (Davidson et al., 2012; Assis, 2018), so one possibility is that functionally important LGTs could be more likely to show unique expression patterns. Although, as LGTs tend to be less expressed than native orthologues (Fig 4) and more donors are closer to recipients, some divergence could be explained by the fact LGTs in *A. semialata* have increased rates of loss and gain (Raimondeau *et al.,* *unpublished*). This is relevant as LGTs in bacteria have been suggested to be initially slightly deleterious (Park and Zhang, 2012), while recent LGTs are thought to undergo faster rates of evolutionary change (Hao and Golding, 2006; Marri, Hao and Golding, 2006). Negative selection pressures and increased rates of evolutionary change could therefore lead to a higher tendency of LGTs becoming pseudogenes. Overall, both our results indicate that the regulatory dynamics of LGTs in grasses can vary considerably across certain orthogroups and understanding the expression of LGTs may often require gene-specific context.

*Accession-specific regulation*

The three accessions used in our study were from three geographically distinct populations, and due to this we tested whether it was suitable to use data from all three accessions as one data set. We found that patterns of gene expression were consistent across our accessions (Fig 1), however it should be noted that certain orthologous genes found in multiple accessions varied considerably in expression (Fig 2).

The highest number of LGTs was found in the ZAM15-05 accession. This accession is near the centre of origin of the species, in the ‘Zambezian’ region of Africa (Lundgren *et al.,* 2015; Bianconi *et al.,* 2020) where *A. semialata* is known to form multispecies clumps which could drive LGT (Raimondeau *et al., unpublished*). Additionally, as ZAM15-05 was found to have the highest rates of gaining and losing LGTs (Raimondeau *et al., unpublished*), it could be expected that it would also have the highest proportion of pseudogenes. This appears to be true as 22% of ZAM15-05 genes were not expressed in either the root or tip leaf, compared to 14% of KWT3 genes and 11% of AUS1 genes. Additionally, when comparing orthologous LGTs found in two or more accessions, expression was lowest in ZAM15-05, with AUS1 having the highest expression followed by KWT3. This could be explained in part due to differential niche requirements (Lundgren *et al.,* 2015). Alternatively, the higher rate of LGT in ZAM15-05 is known to increase the size of its pangenome more than AUS1 or KWT3 (Raimondeau *et al.,* *unpublished*). This could possibly lead to a higher chance of functional genes becoming pseudogenes if they are replaced by a more efficient gene, as differently derived LGTs can fill the similar niches. For example, the photosynthetically important phosphoenolpyruvate carboxylase gene (*ppc*) used by South African *A. semialata* derives from Cenchrinae, while the *ppc* gene used by Australian accessions derives from Andropogoneae (Christin *et al.,* 2012a; Olofsson *et al.,* 2016). Alternatively, some LGTs in *A. semialata* are known to begin as pseudogenes and later become functionally relevant (Olofsson *et al.,* 2019), and this could potentially occur for some accessions but not all. Lastly, as the South African and Australian *A. semialata* accessions dispersed from the Zambezian region in different photosynthetic lineages (Lundgren *et al.,* 2015; Bianconi *et al.,* 2020), it’s unsurprising that the AUS1 and KWT3 accessions shared the least amount of LGTs in this study and using other accessions from Australia/Asia may be useful in drawing parallels to AUS1. Overall, while few LGT appear to be shared and patterns of LGT expression are consistent, understanding accession-specific regulatory differences could provide context to why related LGTs may succeed in some populations but not others.

*Donor-specific regulation*

In our study we used *S. italica* to represent Cenchrinae and *T. triandra* to represent Andropogoneae. As these two species are differently diverged, it was not suitable to merge the data. Cenchrinae had the most LGT events (96) and members of this subtribe are known to co-occur with *A. semialata* in regions of South Africa (Christin *et al.,* 2012a) and can also be found across Africa and Australia (Tsai et al., 2016). This means LGTs derived from this tribe have a high potential to occur in all three of our accessions. Comparatively, Andropogoneae had fewer detected LGTs (47) with only a single LGT detected in KWT3 (KWT3-10003 of reference genome). This orthogroup also had LGTs in both AUS1 and ZAM15-05, suggesting it may derive from an ancient LGT event before the dispersal of *A. semialata* from the Zambezian region. Although members of Andropogoneae (including *T. triandra*) are found in South Africa (Dunning *et al.,* 2017) some geographical restrictions could possibly have caused the fewer gain of Andropogoneae LGTs in KWT3. Correspondingly, LGTs shared between AUS1 and ZAM15-05 likely are restricted to LGT events before the divergence of *A. semialata* to Asia/Oceania ~426-450 kya (Olofsson *et al.,* 2019). In terms of expression LGTs from the two donors were reasonably similar, however LGTs derived from Andropogoneae were the only group of genes in our study which shown higher tip leaf than root expression (Table 2). This could be a coincidence due to low sample size, but nonetheless highlights a divergence from the expression patterns of donor orthologues. Lastly, it should be noted that as RNAseq data was missing for the roots and tip leaf of one *T. triandra* sample, it’s possible that LGTs derived from this group could be underestimated in our expression quantification.

*Highly expressed genes*

As adaptive significance is a focal topic concerning LGT, we attempted to identify the function of highly expressed genes (> 250 FPKM). In total 8 orthogroups had at least one gene which was highly expressed in the roots and/or tip leaf. Three of these orthogroups included highly expressed LGTs, two included highly expressed *S. italica* genes, two included highly expressed *T. triandra* genes and six included highly expressed native *A. semialata* genes. Orthogroup 25 was the most highly expressed orthogroup in our study, and among orthologues the ZAM LGT (ZAM-43373 and ZAM-44371 of the reference genome) had by far the highest expression found in our study at 3772 FPKM. When this gene was searched for using NCBI BLAST (Sayers *et al.,* 2021), the closest match was a laterally transferred PEP carboxykinase (*pck*) gene previously found in a South African population of *A. semialata* (Christin *et al.,* 2012a). The *pck* gene acquired by South African populations is thought to derive from a LGT event prior to the divergence of *A. semialata* and *A. angusta*, as well as fulfil a key role in C4 photosynthesis (Christin *et al.,* 2012a). Therefore, it’s unsurprising that this LGT had the highest expression in our study, and that its expression was almost entirely restricted to the leaves. This leaf-specific expression was also found in then native orthologues of *S. italica* with a measurement of 867 FPKM in the roots, indicating that the LGT fulfils a similar role to its donor. Comparatively, native ZAM15-05 orthologues were also moderately expressed in the root (52 FPKM) and leaves (94 FPKM), and AUS1 orthologues were highly expressed in the root (733 FPKM) and leaves (641). This co-expression likely indicates the orthologous LGTs and *A. semialata* native genes now fulfil different roles and that LGTs do not need to replace their native orthologue to become functionally significant. For example, *pck* is prevalent across flowering plants (Leegood and Walker, 2003), with potential roles in gluconeogenesis (Malone *et al.,* 2007), nitrogen metabolism (Delgado-Alvarado *et al.,* 2007) and the catabolism of citrate or malate (Famiani et al., 2005) among others. Lastly, as this gene supposedly was transferred prior to the origination of *A. semialata* and has been previously identified in South African accessions (Christin *et al.,* 2012a) we would expect this LGT to be found in both AUS1 and KWT3. As it was not, it indicates that even highly adaptive LGTs may be underrepresented in our study.

The second most expressed LGT found in the ZAM15-05 accession (ZAM-42050 in the reference genome) of orthogroup 178. This gene derives from Andropogoneae and was highly expressed in both the roots (301 FPKM) and leaves (838 FPKM). Searching for this gene using NCBI BLAST (Sayers *et al.,* 2021) identified a close relative in *Sorghum bicolor*, which acted as a macrophage migration inhibitory factor (MIF). Proteins associated with MIFs are known to have a role in the immune system of humans (Lue *et al.,* 2002; Calandra and Roger, 2003) and have also been identified in *Arabidopsis* (Panstruga, Baumgarten and Brenhagen *et al.,* 2015) and wheat (Zhao *et al*., 2021), giving them a putative role in plant immunity and disease resistance. This orthogroup was moderate-highly expressed in the roots (192 FPKM) and leaves (330 FPKM) of ZAM15-05 recipients and highly expressed in the roots (352 FPKM) and leaves (787 FPKM) of native *T. triandra*. As these genes are co-expressed, its plausible that they could function in slightly different, yet beneficial roles similar to *pck* genes.

The third highly expressed LGT was also found in the ZAM15-05 accession (ZAM-14689 in reference genome) and is found in orthogroup 83. This gene derives from Cenchrinae and was highly expressed in the roots (430 FPKM) and moderately expressed in the leaves (117 FPKM). Searching for this gene using NCBI BLAST (Sayers *et al.,* 2021) identified a close match to a *S. viridis* gene associated with a BTB/POZ domain-containing protein. In *Arabidopsis* these proteins have been linked with roles in plant-pathogen interactions, and are also involved in the regulation of root and leaf development (Couzigou *et al.,* 2012). Similar roles in leaf development have also been identified in barley (Tavakol *et al.,* 2015). Once again, these genes are co-expressed with recipient genes, although it should be noted that the ZAM15-05 LGT is much higher expressed in the roots than either donor or recipient orthologue, and very similarly expressed in tip leaf expression to the *S. italica* native gene.

However, it should be noted that not all highly expressed orthogroups had highly expressed LGTs, as LGTs had negligible expression in orthogroups 53, 123, 167, 124 and 96 despite high expression in donor orthologues, and sometimes recipients. When examined in BLAST (Sayers *et al.,* 2021) these genes were closely related to genes associated with phosphatase (53), esterase/lipase (123, 167), electron transfer (124) and oxidoreductases (96). The most notable example is orthogroup 96, where ZAM15-05 LGT expression was below 2 FPKM in both the roots and leaves, yet at 251 FPKM in the roots of both native (recipient) ZAM15-05 genes and 1069 FPKM in *T. triandra*. This shows that not all highly functional genes will be similarly expressed after transfer events.

**Conclusions**

Overall, we found that LGTs were in general not more similarly expressed to donor or recipient orthologues and were also not uniquely diverged relative to donors. However, various LGTs did not follow this pattern and instead at times vastly exceeded the expression of either orthologue. Out of the three most highly expressed LGTs, each had more similar expression to their donors than recipients, however, they were still more highly expressed than their donor orthologue. This makes it plausible that increased functional significance could be a valid explanation of increased expression, rather than expression patterns being influenced by cis-regulatory elements. Correspondingly, as function can differ between the same recipient and donor orthologue, the presence of a highly expressed LGT does not always mean a replacement of the native gene. Compensatory cis and trans-regulatory effects have been identified in hybrid studies however, so similar studies in the context of LGT could still prove beneficial. But as not all highly expressed donor orthologues were also highly expressed as LGTs, the largest driver of our highest expressed LGTs appears to be functional significance.

**References:**

Andrews, S. (2022) ‘*FastQC:* *A Quality Control Tool for High Throughput Sequence Data*’. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed: 7th June 2022).

Assis, R. (2018) ‘Lineage-specific expression divergence in grasses is associated with male reproduction, host-pathogen defense, and domestication’, *Genome Biology and Evolution*, 11(1), pp. 207-219. doi: https://doi.org/10.1093/gbe/evy245

Bao, Y., Hu, G., Grover, C., Conover, J., Yuan, D., and Wendel, J. (2019) ‘Unraveling cis and trans regulatory evolution during cotton domestication’, *Nature Communications*, 10(1). doi: https://doi.org/10.1038/s41467-019-13386-w

Bennetzen J. L. *et al.* (2012) ‘Reference genome sequence of the model plant *Setaria*’, *Nature Biotechnology,* 30(6), pp. 555-561. doi: https://doi.org/10.1038/nbt.2196

Bianconi, M. *et al.* (2020) ‘Contrasted histories of organelle and nuclear genomes underlying physiological diversification in a grass species’, *Proceedings of the Royal Society B: Biological Sciences*, 287(1938). doi: https://doi.org/10.1098/rspb.2020.1960

Biłas, R., Szafran, K., Hnatuszko-Konka, K., and Kononowicz, A.(2016) ‘Cis-regulatory elements used to control gene expression in plants’, *Plant Cell, Tissue and Organ Culture*, 127(2), pp.269-287. doi: https://doi.org/10.1007/s11240-016-1057-7

Bolger, A., Lohse, M., and Usadel, B. (2014) ‘Trimmomatic: a flexible trimmer for Illumina sequence data’, *Bioinformatics*, 30(15), pp. 2114-2120. doi: <https://doi.org/10.1093/bioinformatics/btu170>

Calandra, T., and Roger T. (2003) ‘Macrophage migration inhibitory factor: a regulator of innate immunity’, *Nature Reviews Immunology*, 3, pp. 791-800. doi: https://doi.org/10.1038/nri1200

Camino, E., Butts, J., Ordway, A., Vellky, J., Rebeiz, M., and Williams, T*.* (2015) ‘The Evolutionary Origination and Diversification of a Dimorphic Gene Regulatory Network through Parallel Innovations in cis and trans’, *PLOS Genetics*, 11(4), p.e1005136. doi: https://doi.org/10.1371/journal.pgen.1005136

Christin, P. A., Edwards, E. J., Besnard, G., Boxall, S. F., Gregory, R., Kellogg, E. A., Hartwell, J., and Osborne, C. P. (2012a) ‘Adaptive evolution of C 4 photosynthesis through recurrent lateral gene transfer’, *Current Biology*, 22(5), pp. 445–449. doi: https://doi.org/10.1016/j.cub.2012.01.054

Christin, P. A., Wallace, M., Clayton, H., Edwards, E. J., Furbank, R., Hattersley, P., Sage, R., Macfarlane, T., and Ludwig, M. (2012b) ‘Multiple photosynthetic transitions, polyploidy, and lateral gene transfer in the grass subtribe Neurachninae’, *Journal of Experimental Botany*, 63(17), pp. 6297-6308. doi: https://doi.org/10.1093/jxb/ers282

Coolon, J. D., McManus, C. J., Stevenson, K. R., Graveley, B. R., and Wittkopp, P. J. (2014) ‘Tempo and mode of regulatory evolution in *Drosophila*’, *Genome Research*, 24(5), pp. 797–808. doi: https://doi.org/10.1101/gr.163014.113

Couzigou, J-M. *et al*. (2012) ‘*NODULE ROOT* and *COCHLEATA* Maintain Nodule Development and Are Legume Orthologs of *Arabidopsis BLADE-ON-PETIOLE* Genes’, *The Plant Cell*, 24(11), pp. 4498–4510. doi: https://doi.org/10.1105/tpc.112.103747

Danchin, E. (2016) ‘Lateral gene transfer in eukaryotes: tip of the iceberg or of the ice cube?’, *BMC Biology*, 14(101). doi: https://doi.org/10.1186/s12915-016-0330-x

Davidson, R., Gowda, M., Moghe, G., Lin, H., Vaillancourt, B., Shiu, S., Jiang, N., and Robin Buell, C. (2012) ‘Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution’. *The Plant Journal*, 71(3), pp. 492-502. doi: https://doi.org/10.1111/j.1365-313X.2012.05005.x

Delgado-Alvarado A., Walker R. P., and Leegood R., C. (2007) ‘Phospho*enol*pyruvate carboxykinase in developing pea seeds is associated with tissues involved in solute transport and is nitrogen-responsive’, *Plant Cell & Environment,* 30(2), pp. 225-235. doi: https://doi.org/10.1111/j.1365-3040.2006.01622.x

Deng, Y., Xu, H., Su, Y., Liu, S., Xu, L., Guo, Z., Wu, J., Cheng, C., and Feng, J. (2019) ‘Horizontal gene transfer contributes to virulence and antibiotic resistance of *Vibrio harveyi* 345 based on complete genome sequence analysis’, *BMC Genomics*, 20. doi: https://doi.org/10.1186/s12864-019-6137-8

Dunning, L. T. *et al.* (2017) ’The recent and rapid spread of *Themeda triandra*’, *Botany Letters*, 164(4), pp. 327–337. doi: https://doi.org/10.1080/23818107.2017.1391120

Dunning, L. T. *et al.* (2019) ‘Lateral transfers of large DNA fragments spread functional genes among grasses’. *Proceedings of the National Academy of Sciences*, 116(10), pp. 4416-4425. doi: https://doi.org/10.1073/pnas.1810031116

Famiani, F., Cultrera, N., Battistelli, A., Casulli, V., Proietti, P., Standardi, A., Chen, Z., Leegood, R., and Walker, R. (2005) ‘Phosphoenolpyruvate carboxykinase and its potential role in the catabolism of organic acids in the flesh of soft fruit during ripening’. *Journal of Experimental Botany*, 56(421), pp. 2959-2969. doi: https://doi.org/10.1093/jxb/eri293

Fan, X., Qiu, H., Han, W., Wang, Y., Xu, D., Zhang, X., Bhattacharya, D., and Ye, N. (2020) ‘Phytoplankton pangenome reveals extensive prokaryotic horizontal gene transfer of diverse functions’. *Science Advances*, 6(18). doi: https://doi.org/10.1126/sciadv.aba0111

Freschi, L., Vincent, A.T., Jeukens, J., Emond-Rheault, J.-G., Kukavica-Ibrulj, I., Dupont, M.-J., Charette, S.J., Boyle, B., and Levesque, R.C. (2019) ‘The *Pseudomonas aeruginosa* Pan-Genome Provides New Insights on Its Population Structure, Horizontal Gene Transfer, and Pathogenicity’, *Genome Biology and Evolution*, 11(1), pp. 109–120. doi: https://doi.org/10.1093/gbe/evy259

Gibbs-Russell G. E. (1983) ‘The taxonomic position of C3 and C4 *Alloteropsis semialata* (Poaceae) in southern Africa, *Bothalia*, 14, pp. 205 -213. doi: https://doi.org/10.4102/ABC.V14I2.1160

Gilad, Y., Rifkin, S., and Pritchard, J. (2008) ‘Revealing the architecture of gene regulation: the promise of eQTL studies’, *Trends in Genetics*, 24(8), pp. 408-415. doi: https://doi.org/10.1016/j.tig.2008.06.001

Gordon, K. L., and Ruvinsky, I. (2012) ‘Tempo and Mode in Evolution of Transcriptional Regulation’, *PLoS Genetics*, *8*(1). doi: https://doi.org/10.1371/journal.pgen.1002432

Hao, W., and Golding, G. (2006) ‘The fate of laterally transferred genes: Life in the fast lane to adaptation or death’, *Genome Research*, 16(5), pp. 636-643. doi: https://doi.org/10.1101/gr.4746406

Hibdige, S. G. S., Raimondeau, P., Christin, P. A., and Dunning, L. T. (2021) ‘Widespread lateral gene transfer among grasses’, *New Phytologist*, *230*(6), pp. 2474–2486. doi: https://doi.org/10.1111/nph.17328

Kaessmann, H. (2010) ‘Origins, evolution, and phenotypic impact of new genes’, *Genome Research*, 20(10), pp. 1313-1326. doi: https://doi.org/10.1101/gr.101386.109

Kamal, S., Simpson, D., Wang, Z., Gänzle, M., and Römling, U. (2021) ‘Horizontal Transmission of Stress Resistance Genes Shape the Ecology of Beta- and Gamma-Proteobacteria’, *Frontiers in Microbiology*, 12. doi: https://doi.org/10.3389/fmicb.2021.696522

Kaplinsky, N., Braun, D., Penterman, J., Goff, S., and Freeling, M. (2002) ‘Utility and distribution of conserved noncoding sequences in the grasses’, *Proceedings of the National Academy of Sciences*, 99(9), pp. 6147-6151. doi: https://doi.org/10.1073/pnas.052139599

Landry, C. R., Wittkopp, P. J., Taubes, C. H., Ranz, J. M., Clark, A. G., and Hartl, D. L. (2005) ‘Compensatory *cis-trans* Evolution and the Dysregulation of Gene Expression in Interspecific Hybrids of *Drosophila*’, *Genetics*, 171(4), pp. 1813–1822. doi: https://doi.org/10.1534/genetics.105.047449

Langmead, B., and Salzberg, S. (2012) ‘Fast gapped-read alignment with Bowtie 2’, *Nature Methods*, 9(4), pp. 357-359. doi: https://doi.org/10.1038/nmeth.1923

Leegood, R., and Walker, R. (2003) ‘Regulation and roles of phosphoenolpyruvate carboxykinase in plants’, *Archives of Biochemistry and Biophysics*, 414(2), pp. 204-210. doi: https://doi.org/10.1016/s0003-9861(03)00093-6

Li, F. *et al.* (2014) ‘Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns’, *Proceedings of the National Academy of Sciences*, 111(18), pp. 6672-6677. doi: https://doi.org/10.1073/pnas.1319929111

Liu, R., Zou, X., Wang, Y., Long, Q., and Pei., Y. (2020) ‘A 100 bp GAGA motif-containing sequence in AGAMOUS second intron is able to suppress the activity of CaMV35S enhancer in vegetative tissues’, *PLOS ONE* 15(3). doi: https://doi.org/10.1371/journal.pone.0230203

López-Maury, L., Marguerat, S., and Bähler, J. (2008) ‘Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation’, *Nature Reviews Genetics* 9, pp. 583-593. doi: https://doi.org/10.1038/nrg2398

Lue H., Kleemann R., Calandra T., Roger T., and Bernhagen J. (2002) ‘Macrophage migration inhibitory factor (MIF): mechanisms of action and role in disease’, *Microbes Infect.* 4(4), pp. 449-460. doi: https://doi.org/10.1016/s1286-4579(02)01560-5

Lundgren, M. *et al.* (2015) ‘Photosynthetic innovation broadens the niche within a single species’, *Ecology Letters*, 18(10), pp. 1021-1029. doi: https://doi.org/10.1111/ele.12484

Lundgren, M. *et al.* (2019) ‘C4 anatomy can evolve via a single developmental change’, *Ecology Letters*, *22*(2), pp. 302–312. doi: https://doi.org/10.1111/ele.13191

Mahelka, V., Krak, K., Kopecký, D., Fehrer, J., Šafář, J., Bartoš, J., and Blattner, F. R. (2017). ‘Multiple horizontal transfers of nuclear ribosomal genes between phylogenetically distinct grass lineages’, *Proceedings of the National Academy of Sciences, USA*, 114(7), pp. 1726– 1731. doi: https://doi.org/10.1073/pnas.1613375114

Mahelka, V., Krak, K., Fehrer, J., Caklová, P., Nejedlá. M. N., Čegan, R., Kopecký, D., and Šafář, J. (2021) ‘A panicum-derived chromosomal segment captured by Hordeum a few million years ago preserves a set of stress-related genes’*, The Plant Journal* 105, pp. 1141– 1164. doi: https://doi.org/10.1111/tpj.15167

Malone, S., Chen, Z., Bahrami, A., Walker, R., Gray, J., and Leegood, R. (2007) ‘Phospho*enol*pyruvate Carboxykinase in Arabidopsis: Changes in Gene Expression, Protein and Activity during Vegetative and Reproductive Development’, *Plant and Cell Physiology*, 48(3), pp. 441-450. doi: https://doi.org/10.1093/pcp/pcm014

Marri, P., Hao, W. and Golding, G. (2006). ‘Gene Gain and Gene Loss in Streptococcus: Is It Driven by Habitat?’, *Molecular Biology and Evolution*, 23(12), pp. 2379-2391. doi: https://doi.org/10.1093/molbev/msl115

Mattioli, K., Oliveros, W., Gerhardinger, C., Andergassen, D., Maass, P. G., Rinn, J. L., and Melé, M. (2020) ‘Cis and trans effects differentially contribute to the evolution of promoters and enhancers’, *Genome Biology*, 21(1). doi: https://doi.org/10.1186/s13059-020-02110-3

McCormick, R. F. *et al.* (2017) ‘The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization’, *The Plant Journal*, *93*(2), pp. 338–354. doi: https://doi.org/10.1111/tpj.13781

Merli, C., Bergstrom, D. E., Cygan, J. A., and Blackman, R. K. (1996) ‘Promoter specificity mediates the independent regulation of neighboring genes’, *Genes & Development*, 10(10), pp. 1260–1270. doi: https://doi.org/10.1101/gad.10.10.1260

Mitsuhara, I. *et al.* (1996) ‘Efficient Promoter Cassettes for Enhanced Expression of Foreign Genes in Dicotyledonous and Monocotyledonous Plants’, *Plant and Cell Physiology*, *37*(1), pp. 49–59. doi: https://doi.org/10.1093/oxfordjournals.pcp.a028913

Moreno-Villena, J., Dunning, L. T., Osborne, C. P., and Christin, P. A. (2018) ‘Highly Expressed Genes Are Preferentially Co-Opted for C4 Photosynthesis’, *Molecular Biology and Evolution*, 35(1), pp. 94-106. doi: https://doi.org/10.1093/molbev/msx269

Ochman, H., Lawrence, J. and Groisman, E., (2000) ‘Lateral gene transfer and the nature of bacterial innovation’, *Nature*, 405(6784), pp. 299-304. doi: https://doi.org/10.1038/35012500

Olofsson, J. *et al.* (2016) ‘Genome biogeography reveals the intraspecific spread of adaptive mutations for a complex trait’, *Molecular Ecology*, 25(24), pp. 6107-6123. doi: https://doi.org/10.1111/mec.13914

Olofsson, J. *et al.* (2019) ‘Population-Specific Selection on Standing Variation Generated by Lateral Gene Transfers in a Grass’, *Current Biology*, 29(22), pp. 3921-3927.e5. doi: https://doi.org/10.1016/j.cub.2019.09.023

Osada, N., Miyagi, R., and Takahashi, A. (2017) ‘Cis- and Trans-regulatory Effects on Gene Expression in a Natural Population of *Drosophila melanogaster*’, *Genetics*, 206(4), pp. 2139–2148. doi: https://doi.org/10.1534/genetics.117.201459

Panstruga, R., Baumgarten, K., and Bernhagen, J. (2015) ‘Phylogeny and evolution of plant macrophage migration inhibitory factor/D-dopachrome tautomerase-like proteins’, *BMC Evolutionary Biology*, 15(1). doi: https://doi.org/10.1186/s12862-015-0337-x

Park, C. and Zhang, J. (2012) ‘High Expression Hampers Horizontal Gene Transfer’, *Genome Biology and Evolution*, 4(4), pp. 523-532. doi: https://doi.org/10.1093/gbe/evs030

Paterson, A. *et al.* (2009) ‘The *Sorghum bicolor* genome and the diversification of grasses’, *Nature* 457, pp.551–556. doi: https://doi.org/10.1038/nature07723

Phansopa, C., Dunning, L. T., Reid, J. and Christin, P. A. (2020) ‘Lateral Gene Transfer Acts As an Evolutionary Shortcut to Efficient C4 Biochemistry’, *Molecular Biology and Evolution*, 37(11), pp. 3094-3104. doi: https://doi.org/10.1093/molbev/msaa143

Prentice, H. C., Li, Y., Lönn, M., Tunlid, A., Ghatnekar, L. (2015) ‘A horizontally transferred nuclear gene is associated with microhabitat variation in a natural plant population’, *Proceedings of the Royal Society B: Biological Sciences* 282(1821). doi: https://doi.org/10.1098/rspb.2015.2453

R Core Team (2021) ‘*R: A language and environment for statistical computing’*. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/ (Accessed: 7th June 2022).

Raimondeau, P., Bianconi, M. E., Pereira, L., Parisod, C., Christin, P. A., and Dunning L., T. (Unpublished) ‘High turnover of lateral gene transfers diversifies a grass pangenome’

Reynolds, H., Vijayakumar, V., Gluck-Thaler, E., Korotkin, H., Matheny, P. and Slot, J. (2018) ‘Horizontal gene cluster transfer increased hallucinogenic mushroom diversity’, *Evolution Letters*, 2(2), pp.88-101. doi: https://doi.org/10.1002/evl3.42

Roberts, A., and Pachter, L. (2013) ‘Streaming fragment assignment for real-time analysis of sequencing experiments’, *Nature Methods*, 10(1), pp.71-73. doi: https://doi.org/10.1038/nmeth.2251

Rossoni, A., Price, D., Seger, M., Lyska, D., Lammers, P., Bhattacharya, D. and Weber, A. (2019) ‘The genomes of polyextremophilic cyanidiales contain 1% horizontally transferred genes with diverse adaptive functions’. *eLife*, 8. doi: https://doi.org/10.7554/eLife.45017

Sayers, E. W. *et al.* (2021) ‘Database resources of the National Center for Biotechnology Information’, *Nucleic acids research*, 49(D1), pp. D10–D17. doi: https://doi.org/10.1093/nar/gkaa892

Shi, X., Ng, D., Zhang, C., Comai, L., Ye, W. and Chen, J. Z. (2012) ‘Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids’, *Nature Communications*, 3. doi: https://doi.org/10.1038/ncomms1954

Soanes, D., and Richards, T. (2014) ‘Horizontal Gene Transfer in Eukaryotic Plant Pathogens’, *Annual Review of Phytopathology*, 52(1), pp. 583-614. doi: https://doi.org/10.1146/annurev-phyto-102313-050127

Tavakol, E. *et al.* (2015) ‘The Barley *Uniculme4* Gene Encodes a BLADE-ON-PETIOLE-Like Protein That Controls Tillering and Leaf Patterning’, *Plant Physiology*, 168(1), pp.164–174. doi: https://doi.org/10.1104/pp.114.252882

Tsai, K., Lu, M., Yang, K., Li, M., Teng, Y., Chen, S., Ku, M. and Li, W. (2016) ‘Assembling the *Setaria italica* L. Beauv. genome into nine chromosomes and insights into regions affecting growth and drought tolerance’, *Scientific Reports*, 6(1). doi: https://doi.org/10.1038/srep35076

Vallenback, P., Jaarola, M., Ghatnekar, L., Bengtsson, B. O. (2008) ‘Origin and timing of the horizontal transfer of a *PgiC* gene from *Poa* to *Festuca ovina*’, *Molecular Phylogenetics and Evolution* 46, pp: 890– 896. doi: https://doi.org/10.1016/j.ympev.2007.11.031

van der Meer, I., Brouwer, M., Spelt, C., Mol, J., and Stuitje, A. (1992) ‘The TACPyAT repeats in the chalcone synthase promoter of *Petunia hybrida* act as a dominant negative cis-acting module in the control of organ-specific expression’, *The Plant Journal*, *2*(4), pp. 525–535. doi: https://doi.org/10.1046/j.1365-313x.1992.t01-23-00999.x

Vogel, A. *et al.* (2018) ‘Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris*’, *Nature Communications*, 9(1). doi: https://doi.org/10.1038/s41467-018-04344-z

Wickell, D. and Li, F. (2019) ‘On the evolutionary significance of horizontal gene transfers in plants’, *New Phytologist*, 225(1), pp. 113-117. doi: https://doi.org/10.1111/nph.16022

Wickham *et al*. (2019) ‘Welcome to the tidyverse’, *Journal of Open Source Software*, 4(43), 1686. doi: <https://doi.org/10.21105/joss.01686>

Xia, J. *et al.* (2021) ‘Whitefly hijacks a plant detoxification gene that neutralizes plant toxins’, *Cell*, 184(13), pp. 1693-1705.e17. doi: https://doi.org/10.1016/j.cell.2021.02.014

Yang, Z. *et al.* (2016) ‘Horizontal gene transfer is more frequent with increased heterotrophy and contributes to parasite adaptation’, *Proceedings of the National Academy of Sciences*, 113(45). doi: https://doi.org/10.1073/pnas.1608765113

Yang, Z. *et al.* (2019) ‘Convergent horizontal gene transfer and cross-talk of mobile nucleic acids in parasitic plants’, *Nature Plants*, 5(9), pp.991-1001. doi: https://doi.org/10.1038/s41477-019-0458-0

Zhao, M., Chang, Q., Liu, Y., Sang, P., Kang, Z. and Wang, X. (2021) ‘Functional Characterization of the Wheat Macrophage Migration Inhibitory Factor TaMIF1 in Wheat-Stripe Rust (*Puccinia striiformis*) Interaction’, *Biology*, 10(9), 878. doi: https://doi.org/10.3390/biology10090878

Zheng, L., McMullen, M., Bauer, E., Schön, C., Gierl, A. and Frey, M. (2015) ‘Prolonged expression of the BX1 signature enzyme is associated with a recombination hotspot in the benzoxazinoid gene cluster in Zea mays’, *Journal of Experimental Botany*, 66(13), pp. 3917-3930. doi: https://doi.org/10.1093/jxb/erv192