Unveiling the Functional Fate of Duplicated Genes Through Expression Profiling and Structural Analysis

Alex Warwick Vesztrocy[1,2], Natasha Glover[1], Paul D. Thomas[3], Christophe Dessimoz[1,2], Irene Julca[1,2*]

1. SIB Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
2. Department of Computational Biology, University of Lausanne, 1015 Lausanne, Switzerland
3. Department of Population and Public Health Sciences, University of Southern California, Los Angeles, USA

*Corresponding author: irene.julca@unil.ch

Abstract

Gene duplication is a major evolutionary source of functional innovation. Following duplication events, gene copies (paralogues) may undergo various fates, including retention with functional modifications (such as silencing, sub-functionalisation, or neo-functionalisation) or loss. When paralogues are retained, this results in complex orthology relationships, including one-to-many or many-to-many. In such cases, determining which one-to-one pair is more likely to have conserved functions, can be challenging. It has been proposed that, following gene duplication, the copy that diverges more slowly in sequence is more likely to maintain the ancestral function – referred to here as "the least diverged orthologue (LDO) conjecture". This study explores this conjecture, using a novel method to identify asymmetric evolution of paralogues and apply it to all gene families across the Tree of Life in the PANTHER database. Structural data for over 1 million proteins and expression data for 16 animals and 20 plants were then used to investigate functional divergence following duplication. This analysis, the most comprehensive to date, revealed that whilst the majority of paralogues display similar rates of sequence evolution, significant differences in branch lengths following gene duplication can be correlated with functional divergence. Overall, the results support the least diverged orthologue conjecture, suggesting that the least diverged orthologue (LDO) tends to retain the ancestral function, whilst the most diverged orthologue (MDO) may acquire a new, potentially specialised, role.

Keywords: least divergent orthologue conjecture, duplication, structure, expression

Introduction

Gene duplication is widely recognised as a key mechanism driving genome evolution (Ohno 1970). Indeed, redundant gene copies provide a *"temporary escape from the relentless pressure of natural selection"* (Ohno 1972), allowing the evolution of new functions. These genes that result from duplication events are referred to as "paralogues", in contrast to "orthologues" – which descend from a common ancestral gene (Fitch 1970).

The role of duplication in the emergence of new gene functions has been extensively discussed (Hurles 2004). It is widely acknowledged that orthologues are typically more conserved in sequence, structure and function compared to paralogues – a concept known as the "orthologue conjecture" (Nehrt et al. 2011; Altenhoff et al. 2012; Gabaldón and Koonin 2013). Following a duplication event, the two gene copies can be retained with functional redundancy, diverge (with silencing, sub- or neo-functionalization) or become lost (Sémon and Wolfe 2007; Leitch and Leitch 2008). The most common phenomenon is that one copy accumulates deleterious mutations and thereby is silenced and eventually pseudogenised. Alternatively, mutations can lead to functional divergence between the duplicates, resulting in subfunctionalization (each copy may retain part of the ancestral function), or neofunctionalization (one copy may retain the ancestral function, while the other acquires a new function) (Lynch and Force 2000; Adams and Wendel 2005; Sehrish et al. 2014). Thus, gene duplication can provide the raw material for the evolution of functional innovation.

Asymmetric sequence evolution of paralogues can correlate with asymmetric functional divergence. It is generally accepted that following duplication, the two gene copies typically undergo a period of accelerated evolution (Huminiecki and Wolfe 2004), which may be apparent only in one copy (Huerta-Cepas et al. 2011; Pegueroles et al. 2013; Pich I Roselló and Kondrashov 2014). This asymmetry has been interpreted as supporting the Ohno model of evolution (Scannell and Wolfe 2008; Pegueroles et al. 2013), which hypothesises that one copy ("slow" copy, shorter branch in a gene tree) maintains the ancestral rate of evolution and function, while the other copy ("fast" copy, longer branch) may acquire a novel function (Ohno 1970).

The orthologue conjecture has motivated the notion of "least diverged orthologues" (LDO) in the PANTHER database of phylogenetic trees (Mi et al. 2010). Branch lengths are compared for each duplicate lineage following duplication, choosing only the shortest branch. In this way, a single orthologous pair can be chosen even in the presence of post-speciation duplication events, with the pair being expected to be more likely than other pairs to retain the ancestral function. The advantage of this approach is that least diverged orthologues are distinguishable in all cases, except those where the branch lengths following duplication are identical. The

disadvantage, however, is that even small amounts of stochastic variation in the distance estimation will be conflated with actual differences in evolutionary rates.

Here, a generalisation of this approach is used: in instances where gene duplications arise following a speciation event, orthologous relationships can take the form of one-to-many or many-to-many. In this study, the copy with the shorter branch in the gene tree is termed the "least diverged orthologue" (LDO), while the copy with the longer branch is called the "most diverged orthologue" (MDO). Further, the assumption in PANTHER that their LDO gene pairs are more likely to retain the ancestral function can also be extended to this framework: after duplication, the gene copy with the shorter branch (LDO) is more likely to maintain the ancestral function than that with the longer branch (MDO). Whilst not described using the same terminology, the conservation of the function of the LDO is a topic of ongoing debate and is termed here as "The Least Diverged Orthologue Conjecture".

Although not explicitly described in these terms, several studies have investigated aspects of the LDO conjecture empirically. These studies mainly focused on the "orthologue conjecture", but also investigated the asymmetrical evolution of paralogues. They have identified a distinct rate of sequence evolution between paralogues, indicative of functional differences (Huminiecki and Wolfe 2004; Cusack and Wolfe 2007; Scannell and Wolfe 2008; Pegueroles et al. 2013; Pich I Roselló and Kondrashov 2014). To assess functional divergence between paralogous genes, various datasets have been employed, including Gene Ontology terms (Nehrt et al. 2011; Altenhoff et al. 2012), protein-protein interactions (Kim and Yi 2006), and expression patterns (Gu et al. 2005; Li et al. 2005; Wang et al. 2012; Assis and Bachtrog 2013). Additionally, the fast-evolving copy has been associated with increased levels of tissue specificity (Scannell and Wolfe 2008; Huerta-Cepas et al. 2011). Furthermore, some studies have demonstrated that even the slower evolving copy shows evidence of a burst of protein sequence evolution immediately after duplication (Scannell and Wolfe 2008). However, the majority of these studies have relied on sequence differences (Ka/Ks, dN/dS) to differentiate between the fast- and slow-evolving copies, along with expression data from a restricted set of species and anatomical samples.

Aiming to more reliably capture actual differences in selective pressures, this study introduces a novel method to identify significant rate differences between branches immediately following gene duplication events. This approach estimates gene family-specific evolutionary rates and subsequently computes statistically significant rate shifts post-duplication, measured in the branch lengths. To validate this approach, it has been applied to all gene families across the Tree of Life in the PANTHER database. After identifying statistically significant rate shifts, duplication events were placed into six categories: i) normal-normal: both branches are not

111 significantly different, ii) short-short: both branches are significantly shorter than expected, iii)
112 long-long: both branches are significantly longer than expected, iv) normal-long: only one
113 branch is significantly longer than expected, v) short-normal: only one branch is significantly
114 shorter than expected, vi) short-long: one branch is significantly shorter than expected and the
115 other is significantly longer than expected.

116 These six categories represent the two evolutionary models: symmetric (normal-normal, short-
117 short, long-long) and asymmetric (normal-long, normal-short, short-long). Then, in order to
118 explore the differences between the two evolutionary models, we used two distinct datasets
119 as related to function: protein 3D structure, and mRNA expression. Structural data was
120 obtained for more than 1 million proteins (Varadi et al. 2022). Additionally, expression data
121 was incorporated from two separate sources: the Bgee database (Bastian et al. 2021),
122 encompassing 16 animals, as well as The Plant Expression Omnibus (Koh et al. 2023),
123 representing 20 plant species. Furthermore, using the same expression framework, the
124 specialisation hypothesis was investigated – that following duplication, the least diverged
125 orthologue tends to maintain the ancestral function, while the other copy is free to develop a
126 new, more highly diverged function.

127

128 ## Results

129 ### Least diverged orthologue (LDO) conjecture analysis across the Tree of Life

130 This study investigated the least diverged orthologue (LDO) conjecture, which implies that the
131 gene copy with the shortest branch in the gene tree tends to retain the ancestral function after
132 a duplication event. Analysis was performed on a dataset comprising 143 organisms
133 representing different lineages of the Tree of Life, across 15,693 gene trees. This conjecture
134 relies solely on observed branch length and classifies gene copies with the shorter branch as
135 the least diverged orthologue (LDO) and those with the longer branch as the most diverged
136 orthologue (MDO).

137 However, when gene families have different evolutionary rates it is not clear whether small
138 differences in branch length estimates can accurately classify gene copies into these two
139 categories (LDO / MDO). Here, in order to detect distinct evolutionary rates on each of the
140 branches, a statistical method was developed that allows for these family-specific evolutionary
141 rates.

142 The first step of this method is to estimate gene family-specific relative rates of evolution
143 (family-specific factors), for each gene tree (Fig. 1A). The distribution of these factors has an

average value of one, suggesting that most gene trees evolve at a uniform rate compared to the species tree. However, some gene families exhibit relatively high rates, indicating fast-evolving genes (Supplemental Fig. S1, Supplemental Table S1).

Using the family-specific factors, the expected branch length was computed for all branches in every gene tree (Fig. 1A), before computing the z-score of the differences between the observed and expected branch lengths. Interestingly, a significant increase in evolutionary rate was observed between branches with no duplication (average -0.02) and those with duplications (0.07) ($p<0.001$ Mann-Whitney U, Fig. 1B)—suggesting that paralogues tend to evolve faster than orthologues and demonstrate the sensitivity of this method in detecting differences in branch lengths. Similar results were observed when each major lineage was analysed independently (Supplemental Fig. S2).

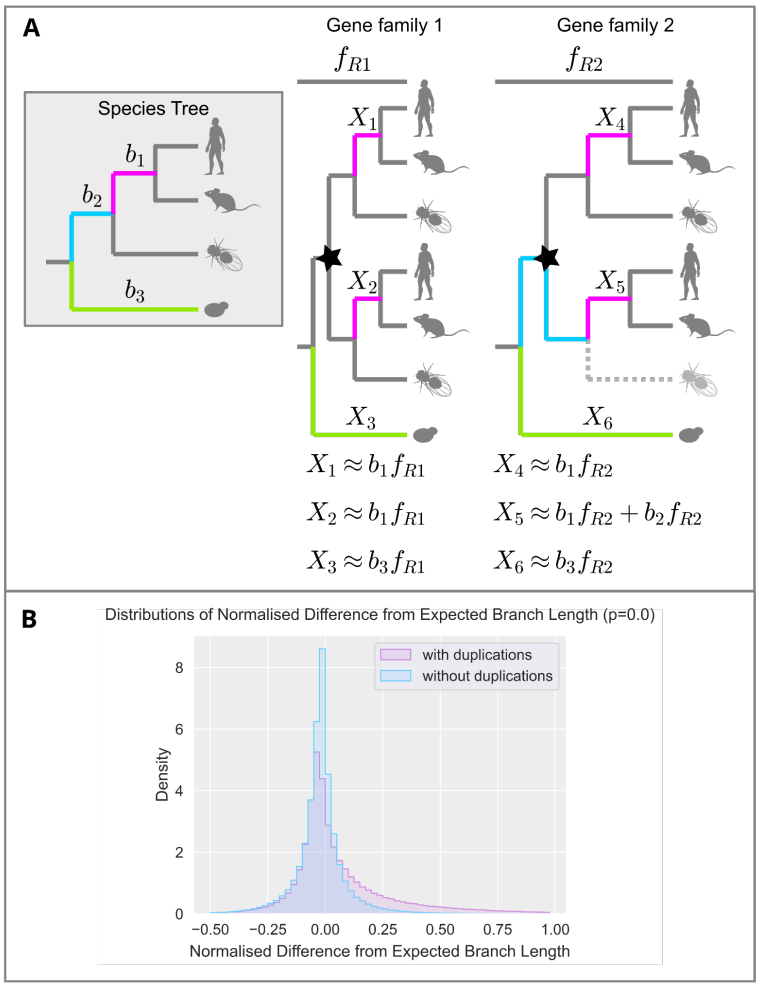

Figure 1. Analysis and calculation of gene tree expected branch lengths. (A) Graphical illustration of the calculation of the expected branch lengths (X). On the right, a species tree depicts the phylogenetic relationships of four species with three branches highlighted ($b1$, $b2$ and $b3$). On the left, two gene trees show a duplication event (node marked with a star). The first gene tree is complete, whilst the second tree has a loss on one side following the duplication. f_R represents the family-specific rate. When no losses occur, there is an exact correspondence between the speciation branch in the gene tree and the species tree (X_1, X_2, X_3, X_4, X_6). When there is a loss, the speciation branch in the gene tree represents the sum of the corresponding branches in the species tree (X_5). These examples highlight a few branches; however, the model incorporates all branches in the species tree along with their corresponding observations in the gene tree. **(B)** Distributions of normalised differences from expected branch lengths in branches with duplications (pink) and without duplications (light blue). The p-value for the Mann–Whitney U test is shown between brackets (p<0.01).

Following this analysis, duplication events were identified across all gene trees and classified into six categories (Fig. 2A, Supplemental Table S2). This gave a total of 926,814 duplication events, with an average of 68 duplications per-family. When examining the major lineages

173 (Fig. 2B, grey bars), the Viridiplantae lineage exhibits the highest percentage of duplications
174 (68%, Fig. 2B, Supplemental Table S3), followed by Deuterostomia (16%) and Protostomia
175 (4%). Internal nodes within this tree contributed only 5% of the total number of duplications
176 (Supplemental Table S3). These findings highlight the significant role of gene duplication in
177 the evolutionary history of plants, in agreement with previous studies (Jiao et al. 2011; Soltis
178 et al. 2015).

179 Further analysis of the distribution of duplications across the six categories within each major
180 lineage (Fig. 2B, coloured bars) revealed that a significant percentage of duplications fell into
181 the normal-normal group (~82%). The second most abundant category was the normal-long
182 group (~13%), which includes paralogues with significantly different branch lengths. The
183 remaining four categories (short-short, long-long, short-normal and short-long) accounted for
184 only ~6% of duplications. When grouping these categories based on their evolutionary pattern,
185 86% displayed symmetric evolution (both branches following similar rates: normal-normal,
186 short-short and long-long); with 14% exhibiting asymmetric evolution (one significantly longer:
187 normal-long, short-normal, short-long). These results indicate that the majority of genes
188 following duplication events have similar evolutionary rates, despite observed variations in
189 absolute branch lengths.

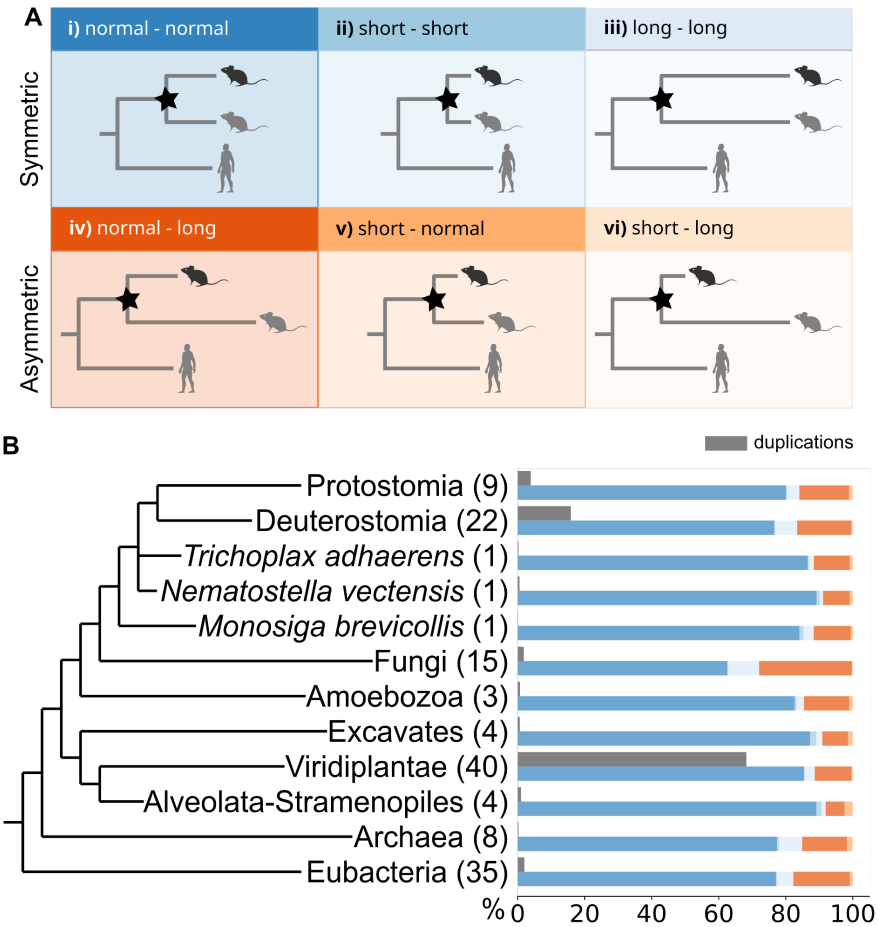

Figure 2. Branch length analysis following duplication events. (A) Depiction of the six categories indicating differences in branch length following a duplication event. The two rows divide the symmetric and asymmetric evolution of paralogues following duplication. Dark colour animal silhouette (black mouse) indicates the least diverged orthologue (LDO) and the grey silhouette (grey mouse) indicates the most diverged orthologue (MDO). **(B)** Species tree showing the major lineages analysed. The numbers between brackets indicate the number of species included in the lineage. The grey bars show the percentage of duplication and the coloured bars show the percentages of duplication within each category (A) from the total duplications in the lineage.

Gene duplications in internal branches represent more ancient duplications compared to species-specific duplications (terminal branches). To investigate any differences, the distribution of duplications between these two types of duplications was examined within each major lineage subtree, excluding three major lineages which consist of only one species (*Trichoplax adhaerens*, *Nematostella vectensis*, *Monosiga brevicollis*).

206 Across all cases, a large proportion of duplication events were present in terminal branches
207 (~70%, Supplemental Fig. S3A), which was similar in all major lineages (Supplemental Fig.
208 S3B). When analysing the distribution of the duplications between the two evolutionary
209 models, a larger proportion of duplications followed symmetric evolution (Supplemental Fig.
210 S3C) in all major lineages (Supplemental Fig. S3D). Interestingly, the proportion of
211 duplications following asymmetric evolution was higher in internal than in terminal branches,
212 suggesting that ancestral duplications may have more time to evolve divergent gene copies.

213

214 **Shorter branches show more similar protein structures**

215 Since protein structure is associated with protein function (Orengo et al. 1999), structural
216 similarity can be used to test for functional divergence. This was performed between extant
217 paralogues (Fig. 3A), as between each paralogue and an outgroup orthologous gene (Fig.
218 3B).

219 Paralogues following symmetric evolution exhibited higher Foldseek LDDT (Local Distance
220 Difference Test) values than those following asymmetric evolution (Fig. 3A). These results
221 suggest that paralogues undergoing symmetric evolution tend to have more similar structures,
222 implying similar functions.

223

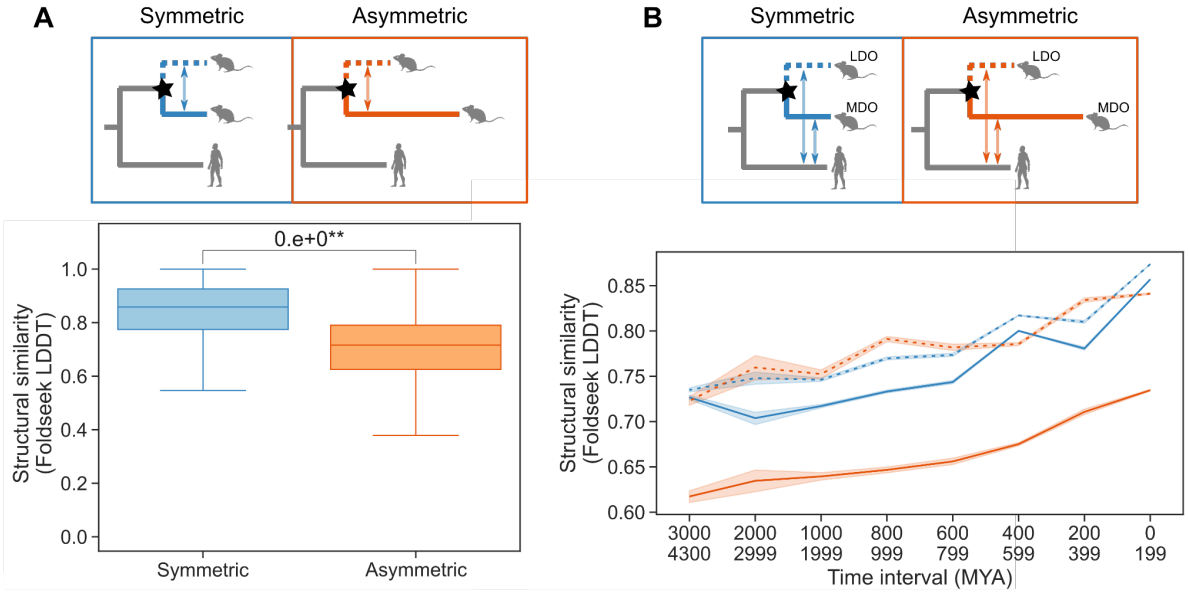

Figure 3. Structural similarity between paralogues and paralogues and orthologues. (A) Structural similarity between paralogues following symmetric and asymmetric duplications (represented as a star). The p-value is indicated on top of the figure, for the Mann–Whitney U test (** p<0.01). **(B)** Structural similarity between each paralogue and its orthologue in symmetric (blue) and asymmetric (orange) duplications. Dashed lines represent the comparison of the least diverged orthologue (LDO), while solid lines represent the comparison of the most diverged orthologue (MDO). The x-axis shows the divergence time of the species analysed in time intervals. Above each panel, the comparisons used are depicted.

To test the least diverged orthologue (LDO) conjecture, structural similarity was computed between each paralogous gene copy and their corresponding co-orthologous outgroup gene. The analysis of the distribution of LDDT scores, as a proxy for functional similarity, revealed differences in protein structure between paralogues in both cases of symmetric and asymmetric evolution (p<0.01 Mann–Whitney U test, Fig. 3B). However, the most striking difference was observed in paralogues following asymmetric duplications, where the copy from the longer branch exhibited much lower LDDT values with the outgroup (average LDDT=0.72), whilst the other branch followed a distribution similar to that of the two copies from symmetric duplications (average LDDT=0.83, Fig. 3B). Additionally, when comparing the LDDT values over time, a general decrease in the structural similarity scores was observed for all comparisons, with no clear differences between symmetric and asymmetric duplications. Similar results were observed when analysing the lineages independently (Supplemental Fig. S4). These results suggest that sequence divergence correlates with structural differences.

248 This supports the hypothesis that functional divergence occurs in the longest branch following
249 duplication, whilst the shortest branch maintains a more similar function to their shared
250 ancestor.

251

252 **Differences in branch length support differences in expression profiles**

253 To test functional differences across genes, gene expression profiles were also used as a
254 proxy for functional similarity. For animals, expression atlases of 16 species across 32
255 anatomical entities were obtained from the Bgee database (Bastian et al. 2021). For plants,
256 publicly available expression atlases from 20 species across 72 anatomical samples were
257 used (Julca et al. 2021; Koh et al. 2023) (Supplemental Table S4). In order to assess if the
258 observed differences in branch length following gene duplication have functional implications,
259 gene expression profiles between paralogues resulting from symmetric and asymmetric
260 evolution were compared using Pearson's correlation coefficient (PCC). Subsequently, the
261 direction of change was investigated by comparing the expression profiles of each paralogue
262 to their co-orthologous outgroup gene.

263 When analysing pairs of paralogues in each category, within both animals and plants,
264 paralogues from the asymmetric category showed lower correlation values compared to those
265 from the symmetric category (Fig. 4A). Furthermore, since most duplications were located in
266 terminal branches and represent more recent duplication events, their gene expression
267 profiles were evaluated separately (Supplemental Fig. S5). In both cases, terminal and internal
268 branches present consistent results, with paralogues in the symmetric category showing
269 slightly higher expression similarity. Notably, symmetrically evolving paralogues arising from
270 duplications on the terminal branches within plants exhibited the highest expression similarity
271 (Supplemental Fig. S5). This observation may suggest that recent duplications have
272 undergone less evolutionary divergence due to a shorter time frame. However, the same
273 pattern is not observed in animals.

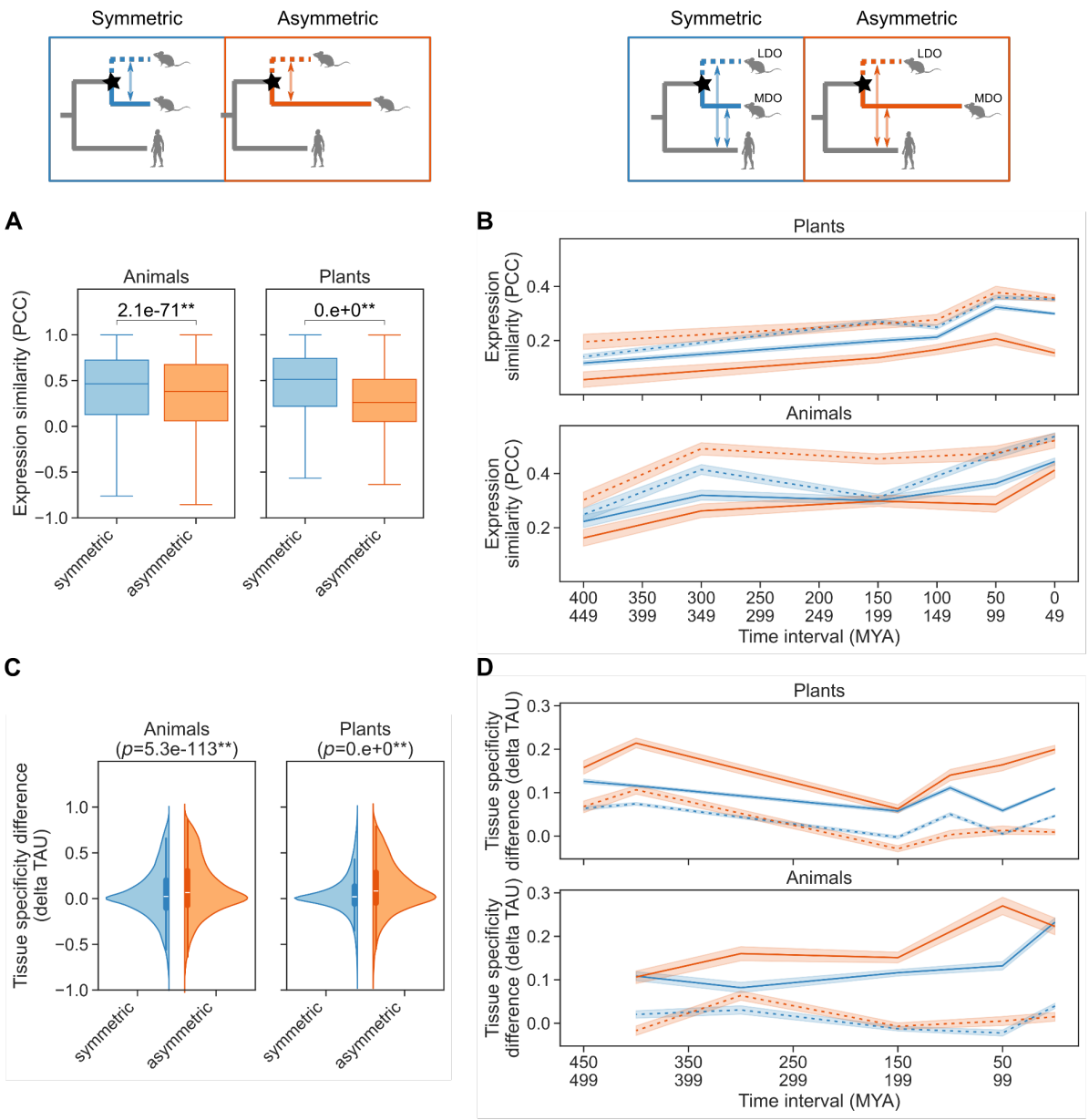

Figure 4. Functional analysis of paralogues using expression data. (A) Expression profile similarity of the paralogues following symmetric and asymmetric duplications. Expression similarities were calculated using Pearson's correlation coefficient (PCC). The p-values are indicated on top of the figures, for the Mann–Whitney U test (* p<0.05 and ** p<0.01). **(B)** Expression similarity (PCC) between the LDO (dashed lines) and MDO (solid lines) genes and their closest outgroup co-orthologue following symmetric (blue) and asymmetric (orange) duplications. **(C)** Violin plot showing the tissue specificity differences (delta Tau) between the paralogues following symmetric and asymmetric duplications. **(D)** Tissue specificity differences (delta Tau) between the LDO (dashed lines) and MDO (solid lines) genes and their closest orthologues following symmetric (blue) and asymmetric (orange) duplications. In plots b) and d), the x-axis indicates the divergence time between the species compared in time intervals. Plant and animal datasets were compared independently. Above the panels, the comparisons used are depicted.

287 Different species may influence the results due to their evolutionary history (for example,
288 polyploid plants). Therefore, each species was analysed independently and showed
289 consistent results (Supplemental Fig. S6). Amongst the animals, 13 species showed similar
290 patterns, indicating lower expression similarity between paralogues within the asymmetric
291 category. However, three species, *Felis catus*, *Drosophila melanogaster* and *Gorilla gorilla*
292 *gorilla*, did not show significant differences between the two categories. For plants, 19 species
293 supported the previous pattern, whilst only one (*Amborella trichopoda*) did not. Interestingly,
294 the polyploid plants (*Brassica napus*, tetraploid; *Triticum aestivum*, hexaploid; *Glycine max*,
295 tetraploid) exhibited major differences between symmetric and asymmetric paralogs. In
296 summary, these results reveal a general trend where paralogues following asymmetric
297 evolution exhibit less similar expression profiles, suggesting that sequence divergence
298 between paralogues correlates with differences in gene expression. Nevertheless,
299 discrepancies observed in a few species highlight the importance of including many different
300 species in this kind of study.

301 To investigate the direction of functional changes following duplication – whether the least
302 (LDO) or most diverged orthologue (MDO) conserves the ancestral gene function – expression
303 profiles were compared between each paralogous gene copy and their co-orthologous
304 outgroup gene. For plants, 112,015 duplications were analysed, with 99,995 symmetric and
305 12,020 asymmetric duplications. For animals, 22,885 duplications were examined, of which
306 16,273 were symmetric and 6,612 were asymmetric duplications. Then, within each category,
307 gene expression similarity was computed (PCC) between each copy and the outgroup gene
308 (Fig. 4B). Differences were observed in both categories, symmetric and asymmetric, with
309 greater differences observed between the asymmetric category (Fig. 4B). Interestingly, in both
310 datasets – animals and plants – the gene following the longest branch (MDO in the asymmetric
311 category), consistently displays low expression similarity (Fig. 4B), suggestive of a lower
312 functional similarity with the ancestral gene. Similar results were observed by analysing each
313 species independently (Supplemental Fig. S7).

314 In order to study the effect of time since duplication, expression similarity was compared
315 across species divergence time (Fig. 4B). In both animals and plants, more ancient paralogues
316 exhibited a slight decrease in expression similarity. In summary, these results suggest that
317 paralogous genes descending from duplications with significantly different branch lengths
318 have more distinct expression profiles. In particular, the most diverged orthologues (MDO)
319 display the greatest change in gene expression, thus suggesting that they have a greater
320 functional divergence from the ancestral gene.

321

322 **Most divergent orthologues (MDO) show more functional specificity**

323 Previous studies have shown that after gene duplication, paralogues tend to increase the
324 levels of tissue specificity, which occurs promptly after the duplication event (Huminiecki and
325 Wolfe 2004; Huerta-Cepas et al. 2011). To test whether the least diverged orthologue (LDO,
326 paralogue with a shorter branch) or the most diverged orthologue (MDO, paralogue with a
327 longer branch) have different tissue specificity, the Tau score (defined in Methods) was
328 computed for each gene in every species.

329 To compare each category, the difference in tissue specificity was calculated (delta Tau)
330 between the most diverged orthologue (MDO) and the least diverged orthologue (LDO) in
331 each group (symmetric and asymmetric duplications, Fig. 4C). A value of zero indicates no
332 difference in sample-specificity between the compared genes, while positive values indicates
333 that the MDO exhibits more specificity and negative values that the LDO is more specific. Our
334 results show that paralogues within the asymmetric category, in both animals and plants, have
335 slightly higher positive values, indicating that the MDOs tend to have a more tissue-specific
336 pattern of expression (Fig. 4C). Similar results were observed when analysing duplication
337 events separately for internal and terminal branches, with terminal branches exhibiting more
338 pronounced differences (Supplemental Fig. S8) and independently within each species
339 (Supplemental Fig. S9). These results suggest that the gene copy which exhibits greater
340 sequence divergence tends to have greater tissue specificity.

341 To determine whether the most diverged orthologue (MDO) also displays increased tissue
342 specificity compared to the ancestral gene, the difference was computed between the Tau
343 values for each paralogous gene copy and their co-orthologous outgroup gene (Fig. 4D). A
344 zero value indicates no difference in tissue specificity, whilst positive values suggest increased
345 specificity of the paralogous gene. In both categories, there are differences between the
346 paralogues, with the MDO genes exhibiting higher delta Tau values. Specifically, in both plants
347 and animals, the MDO gene of the asymmetric category consistently showed the greatest
348 increase in tissue specificity over time (Fig. 4D). Moreover, similar results were observed when
349 analysing individual species (Supplemental Fig. S10). Interestingly, the delta Tau in animals
350 showed a slight decrease over time, whilst in plants this decrease was not apparent (Fig. 4D).
351 This result indicates that tissue specificity in animals may decrease over time, with recent
352 duplications exhibiting higher specificity. To summarise, these findings suggest that when one
353 gene copy is under selective pressure to retain the ancestral function (LDO), the other with
354 the longer branch (MDO) tends to acquire a more specialised role, which may differ over time.

355

Discussion

To investigate the least diverged orthologue (LDO) conjecture, we developed a method to identify paralogues with significantly divergent branch lengths, representing asymmetric evolution. This enabled the exploration of functional divergence between genes following duplication using gene trees from the PANTHER database, covering 143 species, along with structural data for more than one million proteins and expression atlases from 16 animal and 20 plant species. This study provides the most comprehensive analysis, using the largest dataset to date for investigating functional divergence following gene duplication events. Furthermore, it avoids the pitfalls associated with comparing Gene Ontology (GO) functional annotations across different species (Altenhoff et al. 2012; Thomas et al. 2012; Gaudet and Dessimoz 2017).

In the PANTHER database, pairs of extant genes are classified as least diverged orthologues (LDOs) based on their gene tree branch lengths immediately following each post-speciation duplication event (Mi et al. 2010). However, this definition uses only the observed branch length, with no test being employed to determine the significance of any differences between the branch lengths following duplication. To address this gap, this study provides a generalisation of the definition (least and most diverged orthologues – LDO and MDO), as well as a method to evaluate gene tree branch lengths and identify post-speciation duplication events with significant rate differences between branches. The results show that despite observed differences in branch lengths between paralogues, most are not statistically significant (symmetric duplications). These results were consistent across all lineages included in this analysis. Previous studies have suggested that paralogues evolve at similar rates (Kondrashov et al. 2002) and approximately only 25% of duplications within a species display asymmetric evolution (Conant and Wagner 2003). The results presented here are consistent with previous studies, indicating that genes following a duplication event tend to evolve symmetrically. However, this observation does not preclude the possibility of functional changes occurring between paralogues.

If the least diverged orthologue conjecture holds true, it suggests that the least divergent copy (LDO) tends to retain the ancestral function, whilst the most divergent orthologue (MDO) is free to evolve a new function (Ohno 1970). Notably, the MDO copy in asymmetric duplications exhibited more structural differences compared to the ancestral gene than the LDO. It has been observed that the function of a protein strongly depends on its structure (Orengo et al. 1999; Sotomayor-Vivas et al. 2022) and that an increase in structural similarity corresponds to an increase in functional similarity (Barrio-Hernandez et al. 2023). Considering this, the

390 results of this study suggest that sequence divergence, as reflected in gene tree branch
391 lengths, correlates with functional divergence.

392 Genes with highly similar expression profiles are often members of similar biological
393 processes and frequently share similar functions (Eisen et al. 1998; Hughes et al. 2000). In
394 this context expression profiles were used to test the LDO conjecture, showing that paralogous
395 genes which have shorter branches (LDO) tend to have more similar expression profiles to
396 the orthologous gene, with larger differences observed in genes following asymmetric
397 evolution.

398 Previous studies which focused on the models of gene duplication retention have proposed
399 that either one (Gu et al. 2005; Brunet et al. 2006; Panchin et al. 2010; Pegueroles et al. 2013;
400 Pich I Roselló and Kondrashov 2014) or both copies (Huminiecki and Wolfe 2004; Scannell
401 and Wolfe 2008) undergo a rapid period of accelerated evolution, followed by a return to pre-
402 duplication levels. Therefore, it has been suggested that expression divergence between
403 paralogues can occur rapidly after a duplication event (Gu et al. 2002; Guschanski et al. 2017).
404 Additionally, it has been observed that the expression of duplicated genes tends to evolve
405 asymmetrically, with one copy maintaining the ancestral expression profile (Gu et al. 2005;
406 Panchin et al. 2010; Pegueroles et al. 2013) and that asymmetric sequence divergence
407 correlates with asymmetric functional divergence (Blanc and Wolfe 2004; Zhang et al. 2004;
408 Kim and Yi 2006). Overall, this study strongly corroborates these previous works, supporting
409 a rapid change in expression profile following duplication. Furthermore, results highlight that
410 the expression profile of the copy with the shortest branch (LDO) tends to correlate with the
411 ancestral gene expression profile, supporting the least diverged orthologue conjecture.

412 Assuming that the LDO conjecture holds and the MDO copy has a more divergent function,
413 one hypothesis proposes that the new function tends to be more specific (Li et al. 2005). To
414 test this “specialisation hypothesis”, the tissue-specificity of each gene copy was analysed.
415 This revealed that following either symmetric or asymmetric evolution, the MDO copy tends to
416 show higher tissue specificity, particularly notable in the MDO of asymmetric duplications.
417 Additionally, for animals, the levels of tissue-specificity in paralogs, when compared with their
418 co-orthologs, appear to change over time, with the MDO tending to have a more specialized
419 function in recent duplications. These results are consistent with previous studies, which
420 showed that gene duplication leads to a rapid increase in tissue specificity (Huminiecki and
421 Wolfe 2004; Huerta-Cepas et al. 2011) and that expression divergence between paralogues
422 results in increased tissue specificity (Assis and Bachtrog 2015). Furthermore, young
423 paralogues were reported to be highly tissue-specific and become more ubiquitous with
424 divergence time (Kryuchkova-Mostacci and Robinson-Rechavi 2016).

425 In conclusion, this comprehensive study using expression and structural aspects of function
426 corroborates the least divergent orthologue (LDO) conjecture: the gene copy with the shortest
427 branch after a duplication event tends to retain the ancestral function. The LDO gene
428 maintains a similar structure, expression profile, and tissue specificity to the ancestral gene,
429 highlighting its functional conservation over time. Furthermore, the differences in structure and
430 gene expression observed in paralogues following symmetric evolution suggested a rapid
431 change occurring after gene duplication. Overall, these results significantly contribute to the
432 understanding of gene duplication dynamics and their effect on gene function.

433

434 **Methods**

435 ***Gene tree processing***

436 To study the least diverged orthologue conjecture, the species tree for 143 organisms and
437 15,693 gene trees were downloaded from PANTHER v18.0 (Thomas et al. 2022). PANTHER
438 gene trees are annotated with evolutionary information in their internal nodes, including
439 speciation, gene duplication, horizontal gene transfer, or unknown events (nodes for which
440 the type cannot be reliably inferred). When extracting duplication events, some were discarded
441 from the analysis. These include: duplications at the root of the gene tree (no speciation event
442 to compare); those with horizontal gene transfer or “unknown” events annotated on either side
443 of the duplication; as well as those with gene losses immediately following the duplication.
444 Branches with poorly fitting distances (arbitrarily set to 2.0 substitutions/site in PANTHER)
445 were also removed. When a polytomy was observed (i.e. duplication nodes with more than
446 two descendant branches), all branches were analysed independently. Finally, a total of
447 15,577 gene trees were used for further analysis (Supplemental Table S2).

448

449 ***Calculation of the gene tree expected branch lengths***

450 Expected branch lengths were computed, using a simple evolutionary model: the expected
451 length between two speciation events in a gene tree is equal to the corresponding branch in
452 the species tree scaled by a gene family-specific rate. In the case of gene loss, the model
453 assumes that the branch length in the gene tree represents the sum of all the branches
454 corresponding to the branch in the species tree scaled by the family-specific rate (Fig. 1A).

455 By expressing this as a system of linear equations, it is possible to optimise the family rates
456 and species tree branch lengths, constrained by the observed branch lengths within the many

457 gene families. That is, $X \approx \sum_{i \in X} b_i f_R$, where X represents the branch length in the species
458 tree, b is the branch length in the gene tree, and f_R stands for the family-specific rate. Using
459 this method, we calculated the expected branch length in the gene trees and identified
460 statistically significant rate shifts (Supplemental Table S2). The Python scripts generated for
461 this method are available on GitHub (https://github.com/DessimozLab/ldo_study).

462

463 ***Identification of branch length differences after duplication***

464 Since all paralogues resulting from a duplication event can be classified as either least
465 diverged orthologues (LDO, shorter branch) or most diverged orthologues (MDO, longer
466 branch), a statistical method was used, to determine whether the differences observed in the
467 branch length post-duplication were significant.

468 All observations of branches between all speciation events were extracted from gene trees
469 and compared to the expected branch length (calculated under the model described in the
470 previous section). To compare different branches in the species tree, these differences were
471 transformed into z-scores. The resulting distributions of the normalised differences between
472 branches with and without duplication were compared using the Mann–Whitney U test from
473 the Python package SciPy v1.13.0 (Virtanen et al. 2020).

474 All further analysis filtered out gene trees without duplication events, which resulted in 13,505
475 gene trees and 553,821 duplication events distributed across different lineages of the Tree of
476 Life.

477 For each duplication event, the two branches were compared using their z-score of the
478 difference between observed and expected, by computing *p*-values using a two-tailed test.
479 Based on this comparison, duplication events were placed into six categories (Fig. 2A): i)
480 normal-normal: both branches are not significantly different; ii) short-short: both branches are
481 significantly shorter than expected; iii) long-long: both branches are significantly longer than
482 expected; iv) normal-long: only one branch is significantly longer than expected; v) short-
483 normal: only one branch is significantly shorter than expected; vi) short-long: one branch is
484 significantly shorter than expected and the other is significantly longer than expected. Using
485 this information, the dataset was divided into two duplication models: symmetric (normal-
486 normal, short-short, long-long) and asymmetric evolution (normal-long, short-normal, short-
487 long).

488

489 ***Pairwise structural analysis***

490 Structural data for 1,829,120 proteins out of 1,968,858 of the PANTHER dataset were
491 downloaded from the AlphaFold Protein Structure Database v4 (Varadi et al. 2022). The
492 139,738 proteins didn't have structural data available.

493 Pairwise structural alignments were computed between pairs of paralogues, as well as each
494 paralogue and their co-orthologous outgroup gene, using Foldseek version 8.ef4e960 (van
495 Kempen et al. 2024). Outgroup co-orthologous genes were selected from the closest outgroup
496 species before the duplication event in the gene tree. Then, the closest gene in that species
497 (least diverged copy) was selected.

498 The Foldseek LDDT (Local Distance Difference Test) score, which is the average LDDT of the
499 alignment, was used to evaluate similarity in structure – as a proxy for functional similarity.
500 LDDT scores express the percentage of inter-atomic distances and range from 0 (not
501 conserved distances) to 1 (perfect model) (Mariani et al. 2013). Distributions of the different
502 datasets were compared using the Mann–Whitney U test.

503

504 ***Compilation of gene expression atlases***

505 As another functional characteristic of paralogous genes, expression atlases of 16 animals
506 and 20 plants were used. For animals, RNA-Seq gene expression data and anatomical
507 annotations (UBERON terms) were retrieved from Bgee 15.1 (Bastian et al. 2021). This
508 consisted of 2,639 RNA-seq experiments across 32 different anatomical samples. For plants,
509 expression atlases were taken from previous studies (The Plant Expression Omnibus) (Julca
510 et al. 2021; Koh et al. 2023) and anatomical entities were obtained from the sample names of
511 RNA-seq experiments in the NCBI Biosample database. Manual curation was conducted to
512 define anatomical entity names (for example, leaf, flowers), resulting in 72 different anatomical
513 samples for plants. All species included, for both animals and plants, had at least four different
514 anatomical samples (Supplemental Table S4). Also, only genes with TPM >=2 in at least one
515 sample were retained. Then, the TPM data of each gene was log-transformed using the
516 hyperbolic arcsine function (Johnson and Krishnan 2022): $arcsinh(x) = ln(x + \sqrt{x^2 + 1})$.
517 Finally, the average value was computed for samples coming from the same anatomical entity.

518

519 ***Analysis of gene expression profiles***

520 To test for functional differences between pairs of paralogues, as well as each paralogue and
521 the closest orthologous outgroup, expression profiles were compared by computing Pearson's
522 correlation coefficient (PCC). To facilitate comparison across different species, only equivalent
523 anatomical entities were used (for example, leaf for Arabidopsis and tomato) (Supplemental
524 Table S4). When selecting an orthologous outgroup to compare the paralogous genes to, the
525 closest species was used for which data was available (Supplemental Table S5). Additionally,
526 species were removed from the outgroup analysis when there were too few orthologous genes
527 (<50 genes). Finally, again, distributions of the different datasets were compared using the
528 Mann–Whitney U test.

529

530 ***Analysis of gene expression specificity***

531 Sample-specificity of genes, based on expression data, was calculated using the Tau score
532 (Yanai et al. 2005; Kryuchkova-Mostacci and Robinson-Rechavi 2017). Tau values range from
533 0 to 1, where 0 indicates a gene is broadly expressed and 1 denotes tissue-specific
534 expression. Gene pairwise comparisons were performed by calculating the difference in tissue
535 specificity ("delta Tau") between the two genes. Comparisons were performed for the pairs of
536 extant paralogous genes, as well as independently for each gene copy with the chosen
537 orthologous outgroup gene. Differences in their distribution were assessed using the Mann–
538 Whitney U test.

539

540 **Estimation of divergence times**

541 Divergence times (adjusted times) between species were obtained from the TimeTree
542 database (Kumar et al. 2022) and are expressed in millions of years ago (MYA). These time
543 estimates were used to plot the differences between genes in the structure, expression profile,
544 and sample specificity analyses.

545

546 **Data access**

547 All the scripts generated in this study are available via GitHub
548 (https://github.com/DessimozLab/Ido_study).

549

550

551 **Competing interest statement**

552 The authors declare no competing interests.

553

554 **Acknowledgments**

555 This work was funded by the SNSF grant 205085. We thank all members of the Comparative
556 Genomics lab and Riccardo Delli Ponti for their valuable insights.

557

558 Author Contributions: CD and PDT conceived the project idea. CD, AWV, NG, and IJ designed
559 the methodology. AWV developed the scripts to analyse the data. AWV and IJ analysed the
560 data. IJ and AWV wrote the manuscript with comments from all authors. All authors read and
561 approved the final manuscript.

562

563

564 **References**

565

566 Adams KL, Wendel JF. 2005. Novel patterns of gene expression in polyploid plants. *Trends*
567 *Genet* **21**: 539–543.

568 Altenhoff AM, Studer RA, Robinson-Rechavi M, Dessimoz C. 2012. Resolving the ortholog
569 conjecture: orthologs tend to be weakly, but significantly, more similar in function than
570 paralogs. *PLoS Comput Biol* **8**: e1002514.

571 Assis R, Bachtrog D. 2013. Neofunctionalization of young duplicate genes in Drosophila.
572 *Proc Natl Acad Sci USA* **110**: 17409–17414.

573 Assis R, Bachtrog D. 2015. Rapid divergence and diversification of mammalian duplicate
574 gene functions. *BMC Evol Biol* **15**: 138.

575 Barrio-Hernandez I, Yeo J, Jänes J, Mirdita M, Gilchrist CLM, Wein T, Varadi M, Velankar S,
576 Beltrao P, Steinegger M. 2023. Clustering predicted structures at the scale of the known
577 protein universe. *Nature* **622**: 637–645.

578 Bastian FB, Roux J, Niknejad A, Comte A, Fonseca Costa SS, de Farias TM, Moretti S,
579 Parmentier G, de Laval VR, Rosikiewicz M, et al. 2021. The Bgee suite: integrated
580 curated expression atlas and comparative transcriptomics in animals. *Nucleic Acids Res*
581 **49**: D831–D847.

582 Blanc G, Wolfe KH. 2004. Functional divergence of duplicated genes formed by polyploidy
583 during Arabidopsis evolution. *Plant Cell* **16**: 1679–1691.

584 Brunet FG, Roest Crollius H, Paris M, Aury J-M, Gibert P, Jaillon O, Laudet V, Robinson-
585 Rechavi M. 2006. Gene loss and evolutionary rates following whole-genome duplication
586 in teleost fishes. *Mol Biol Evol* **23**: 1808–1816.

587 Conant GC, Wagner A. 2003. Asymmetric sequence divergence of duplicate genes.
588 *Genome Res* **13**: 2052–2058.

589 Cusack BP, Wolfe KH. 2007. Not born equal: increased rate asymmetry in relocated and
590 retrotransposed rodent gene duplicates. *Mol Biol Evol* **24**: 679–686.

591 Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of
592 genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**: 14863–14868.

593 Fitch WM. 1970. Distinguishing homologous from analogous proteins. *Syst Zool* **19**: 99–113.

594 Gabaldón T, Koonin EV. 2013. Functional and evolutionary implications of gene orthology.
595 *Nat Rev Genet* **14**: 360–366.

596 Gaudet P, Dessimoz C. 2017. Gene ontology: pitfalls, biases, and remedies. *Methods Mol*
597 *Biol* **1446**: 189–205.

598 Guschanski K, Warnefors M, Kaessmann H. 2017. The evolution of duplicate gene
599 expression in mammalian organs. *Genome Res* **27**: 1461–1474.

600 Gu X, Zhang Z, Huang W. 2005. Rapid evolution of expression and regulatory divergences
601 after yeast gene duplication. *Proc Natl Acad Sci USA* **102**: 707–712.

602 Gu Z, Nicolae D, Lu HH-S, Li WH. 2002. Rapid divergence in expression between duplicate
603 genes inferred from microarray data. *Trends Genet* **18**: 609–613.

604 Huerta-Cepas J, Dopazo J, Huynen MA, Gabaldón T. 2011. Evidence for short-time
605 divergence and long-time conservation of tissue-specific expression after gene
606 duplication. *Brief Bioinformatics* **12**: 442–448.

607 Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA,
608 Coffey E, Dai H, He YD, et al. 2000. Functional discovery via a compendium of
609 expression profiles. *Cell* **102**: 109–126.

610 Huminiecki L, Wolfe KH. 2004. Divergence of spatial gene expression profiles following
611 species-specific gene duplications in human and mouse. *Genome Res* **14**: 1870–1879.

612 Hurles M. 2004. Gene duplication: the genomic trade in spare parts. *PLoS Biol* **2**: E206.

613 Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP,
614 Hu Y, Liang H, Soltis PS, et al. 2011. Ancestral polyploidy in seed plants and
615 angiosperms. *Nature* **473**: 97–100.

616 Johnson KA, Krishnan A. 2022. Robust normalization and transformation techniques for
617 constructing gene coexpression networks from RNA-seq data. *Genome Biol* **23**: 1.

618 Julca I, Ferrari C, Flores-Tornero M, Proost S, Lindner A-C, Hackenberg D, Steinbachová L,
619 Michaelidis C, Gomes Pereira S, Misra CS, et al. 2021. Comparative transcriptomic
620 analysis reveals conserved programmes underpinning organogenesis and reproduction
621 in land plants. *Nat Plants*.

622 Kim S-H, Yi SV. 2006. Correlated asymmetry of sequence and functional divergence
623 between duplicate proteins of Saccharomyces cerevisiae. *Mol Biol Evol* **23**: 1068–1075.

624 Koh E, Goh W, Julca I, Villanueva E, Mutwil M. 2023. PEO: Plant Expression Omnibus - a
625 comparative transcriptomic database for 103 Archaeplastida. *Plant J*.

626 Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. 2002. Selection in the evolution of gene
627 duplications. *Genome Biol* **3**: RESEARCH0008.

628 Kryuchkova-Mostacci N, Robinson-Rechavi M. 2017. A benchmark of gene expression
629 tissue-specificity metrics. *Brief Bioinformatics* **18**: 205–214.

630 Kryuchkova-Mostacci N, Robinson-Rechavi M. 2016. Tissue-Specificity of Gene Expression
631 Diverges Slowly between Orthologs, and Rapidly between Paralogs. *PLoS Comput Biol*

632 **12**: e1005274.

633 Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB.
634 2022. Timetree 5: an expanded resource for species divergence times. *Mol Biol Evol* **39**.

635 Leitch AR, Leitch IJ. 2008. Genomic plasticity and the diversity of polyploid plants. *Science*
636 **320**: 481–483.

637 Li W-H, Yang J, Gu X. 2005. Expression divergence between duplicate genes. *Trends Genet*
638 **21**: 602–607.

639 Lynch M, Force A. 2000. The probability of duplicate gene preservation by
640 subfunctionalization. *Genetics* **154**: 459–473.

641 Mariani V, Biasini M, Barbato A, Schwede T. 2013. lDDT: a local superposition-free score for
642 comparing protein structures and models using distance difference tests. *Bioinformatics*
643 **29**: 2722–2728.

644 Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. 2010. PANTHER version 7:
645 improved phylogenetic trees, orthologs and collaboration with the Gene Ontology
646 Consortium. *Nucleic Acids Res* **38**: D204-10.

647 Nehrt NL, Clark WT, Radivojac P, Hahn MW. 2011. Testing the ortholog conjecture with
648 comparative functional genomic data from mammals. *PLoS Comput Biol* **7**: e1002073.

649 Ohno S. 1972. An argument for the genetic simplicity of man and other mammals. *J Hum*
650 *Evol* **1**: 651–662.

651 Ohno S. 1970. *Evolution by gene duplication*. Springer Berlin Heidelberg, Berlin, Heidelberg.

652 Orengo CA, Todd AE, Thornton JM. 1999. From protein structure to function. *Curr Opin*
653 *Struct Biol* **9**: 374–382.

654 Panchin AY, Gelfand MS, Ramensky VE, Artamonova II. 2010. Asymmetric and non-uniform
655 evolution of recently duplicated human genes. *Biol Direct* **5**: 54.

656 Pegueroles C, Laurie S, Albà MM. 2013. Accelerated evolution after gene duplication: a
657 time-dependent process affecting just one copy. *Mol Biol Evol* **30**: 1830–1842.

658 Pich I Roselló O, Kondrashov FA. 2014. Long-term asymmetrical acceleration of protein
659 evolution after gene duplication. *Genome Biol Evol* **6**: 1949–1955.

660 Scannell DR, Wolfe KH. 2008. A burst of protein sequence evolution and a prolonged period
661 of asymmetric evolution follow gene duplication in yeast. *Genome Res* **18**: 137–147.

662 Sehrish T, Symonds VV, Soltis DE, Soltis PS, Tate JA. 2014. Gene silencing via DNA
663 methylation in naturally occurring Tragopogon miscellus (Asteraceae) allopolyploids.
664 *BMC Genomics* **15**: 701.

665 Sémon M, Wolfe KH. 2007. Consequences of genome duplication. *Curr Opin Genet Dev* **17**:
666 505–512.

667 Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. 2015. Polyploidy and genome evolution
668 in plants. *Curr Opin Genet Dev* **35**: 119–125.

669 Sotomayor-Vivas C, Hernández-Lemus E, Dorantes-Gilardi R. 2022. Linking protein
670 structural and functional change to mutation using amino acid networks. *PLoS ONE* **17**:
671 e0261829.

672 Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou L-P, Mi H. 2022. PANTHER:
673 Making genome-scale phylogenetics accessible to all. *Protein Sci* **31**: 8–22.

674 Thomas PD, Wood V, Mungall CJ, Lewis SE, Blake JA, Gene Ontology Consortium. 2012.
675 On the Use of Gene Ontology Annotations to Assess Functional Similarity among
676 Orthologs and Paralogs: A Short Report. *PLoS Comput Biol* **8**: e1002386.

677 van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J,
678 Steinegger M. 2024. Fast and accurate protein structure search with Foldseek. *Nat*
679 *Biotechnol* **42**: 243–246.

680 Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O,
681 Wood G, Laydon A, et al. 2022. AlphaFold Protein Structure Database: massively
682 expanding the structural coverage of protein-sequence space with high-accuracy
683 models. *Nucleic Acids Res* **50**: D439–D444.

684 Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E,
685 Peterson P, Weckesser W, Bright J, et al. 2020. SciPy 1.0: fundamental algorithms for
686 scientific computing in Python. *Nat Methods* **17**: 261–272.

687 Wang Y, Wang X, Paterson AH. 2012. Genome and gene duplications and gene expression
688 divergence: a view from plants. *Ann N Y Acad Sci* **1256**: 1–14.

689 Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-
690 Saban S, Safran M, Domany E, et al. 2005. Genome-wide midrange transcription
691 profiles reveal expression level relationships in human tissue specification.
692 *Bioinformatics* **21**: 650–659.

693 Zhang Z, Gu J, Gu X. 2004. How much expression divergence after yeast gene duplication
694 could be explained by regulatory motif evolution? *Trends Genet* **20**: 403–407.

695

696

697 **Supplemental Tables**

698

699 **Supplemental Table S1.** Family-specific factors of each gene tree. The columns in order
700 indicate: the gene tree ID as in PANTHER, the gene family name as in PANTHER, the number
701 of branches in the gene tree, and the family-specific factor.

702

703 **Supplemental Table S2.** Table showing duplication and speciation events, branch lengths,
704 expected branch lengths, and statistical measures.

705

706 **Supplemental Table S3.** Number of duplications per major lineage and category. The first
707 column shows the lineage name, the second to the seventh columns represent the duplication
708 categories (normal-normal, short-short, long-long, normal-short, normal-long, short-long), the
709 eighth column indicates the total number of duplications per lineage, and the ninth shows the
710 percentage of duplication from the total number.

711

712 **Supplemental Table S4.** Number of RNA-seq experiments used per anatomical entity in
713 plants and animals. The first column shows the species name, the second, is the mnemonic,
714 and the following columns are the anatomical entity name.
715
716 **Supplemental Table S5.** Table presenting expression data comparisons between paralogs
717 (MDO-LDO), paralogous and co-orthologous. The following measures are shown: Pearson
718 correlation between paralogs, Pearson correlation between paralogs and orthologs, specificity
719 measure of each paralogous and the ortholog, and tissue specificity difference (delta_tau).