Research article

Sudden shifts in expression after small-scale duplication in vertebrates and strong support for the ortholog conjecture

Tina Begum[1,2], **Pablo Duchen**[3,4], **Christabel Floi Bucao**[1,2], **Marc Robinson-Rechavi**[1,2]*

[1]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

[2]SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland

[3]Department of Computational Biology, University of Lausanne, Lausanne

[4]Institute for Organismal and Molecular Evolutionary Biology, Johannes Gutenberg University of Mainz, Germany

*Corresponding author email: marc.robinson-rechavi@unil.ch

Abstract

Gene duplication is a potential source of innovation, but the evolutionary dynamics of functional change are still poorly understood. Under the debated "ortholog conjecture", most functional change and innovation is assumed to follow duplication. Testing the ortholog conjecture allows to better understand and characterize the way in which gene function evolves. Most models of functional evolution assume continuous change, an assumption which we challenge here. We have applied a Lévy model of evolutionary trait jumps to the evolution of gene expression in vertebrates, with a special focus on duplication in teleost fishes. We show for the first time that trait jumps strongly affects paralogs, in addition to other modes of functional evolution. We find that at least 25% of teleost fish small-scale duplicates follow a rapid evolutionary rate shift model for both expression level and tissue-specificity, much more than after speciations. However, genome-wide duplicates (ohnologs) do not support such a trait jump model, and thus follow a different evolutionary dynamic. While there is some evidence for more positive selection at the protein-coding level after duplication, it is not strongly linked to jumps in expression. Finally, both small-scale paralogs and ohnologs strongly support the ortholog conjecture by contrasting speciation branches pre- and post-duplication to the duplication branches themselves, with trait jumps explaining much of the higher phylogenetic independent contrasts between small-scale paralogs.

Significance statement

The debate on the ortholog conjecture, i.e. that gene function changes little between orthologs but changes frequently between paralogs, provides a framework to understand better the evolution of

gene function. Here we add two pieces to the puzzle: a novel way to use phylogenetic contrasts to test the ortholog conjecture, by comparing not only duplication to speciation, but speciation according to whether they were preceded by a duplication; and a model of jumps rather than continuous change of gene function. We tested these on vertebrates, with emphasis on teleost fishes, distinguishing small-scale duplications and whole-genome duplication; in all cases we support strongly the ortholog conjecture. We find that trait jumps strongly affect small-scale paralogs but not genome duplication paralogs, providing an exciting new model for gene function evolution.

Introduction

Gene duplication is a potential source of functional innovation, and plays an important role in the evolution of genomes (Ohno 1970; Chen and Zhang 2012; Guschanski et al. 2017). Such innovation can lead to overall function gain, e.g. through neofunctionalization or subfunctionalization, or to specialization, and can be symmetric or not between paralogs (Assis and Bachtrog 2013; Assis and Bachtrog 2015; Braasch et al. 2016; Lien et al. 2016; Guschanski et al. 2017; Sandve et al. 2018). Paralogs can also be retained without any functional change, due to dosage constraints (Gout and Lynch 2015; Thompson et al. 2016). In this context, there has been a debate about the long standing hypothesis that paralogs diverge generally more in function than orthologs, the "ortholog conjecture" (Koonin 2005; Studer and Robinson-Rechavi 2009; Nehrt et al. 2011; Altenhoff et al. 2012; Chen and Zhang 2012; Gabaldón and Koonin 2013; Rogozin et al. 2014; Kryuchkova-Mostacci and Robinson-Rechavi 2016; Dunn et al. 2018; Fukushima and Pollock 2020; Stamboulian et al. 2020; Begum and Robinson-Rechavi 2021). This ortholog conjecture is of interest both in practice, for functional annotation of homologs, and in principle, to understand how gene function evolves. Thus contrasting the evolution of genes after duplication or speciation provides a clear framework to investigate the evolution of function.

Since the retention rates and evolutionary dynamics of duplicates vary among lineages and between genes, they need to be studied using a phylogenetic framework with proper care (Pagel 1999; Blomberg et al. 2003; Dunn et al. 2018; Begum and Robinson-Rechavi 2021). A phylogenetic approach also provides the opportunity to explicitly test different models of evolution. In most studies, the implicit model is that function evolves gradually, with changes in

evolutionary rate mapped to specific branches of the phylogeny. These approaches, while important, remain less explored than simpler pairwise comparisons.

There is abundant evidence for an asymmetric steady acceleration of sequence-level evolutionary rates (e.g. *d*N/*d*S) after duplication (Conant and Wagner 2003; Gu et al. 2005; Brunet et al. 2006; Kim and Yi 2006; Cusack and Wolfe 2007; Scannell and Wolfe 2008; Studer and Robinson-Rechavi 2009; Panchin et al. 2010; Pegueroles et al. 2013; Roselló and Kondrashov 2014; Braasch et al. 2016; Lien et al. 2016; Holland et al. 2017; Sandve et al. 2018). However, functional evolution following gene duplication might not be limited to such gradual modes of evolution; it might not even be the main mode of divergence. Studies of phenotype evolution have found that there can be a period of instantaneous bursts or jumps in trait values at the bases of clades that transitioned into new adaptive zones, due to the appearance of an evolutionary novelty (Simpson 1944; Bokma 2008; Landis et al. 2013; Duchen et al. 2017; Landis and Schraiber 2017). This pattern of evolutionary innovations is called "pulsed evolution", where the background trait evolutionary rates σ^2 remain unaltered following a jump with a strength λ (Duchen et al. 2017; Gao and Wu 2022). Such non-gradual evolutionary jumps were identified with a frequency of once-per-million-years in vertebrate evolution, using a pure-jump model (Uyeda et al. 2011; Landis et al. 2013). It is still an open question whether such pulsed evolution exists for gene and genome duplicates (Begum et al. 2021). We propose here to examine this question in light of the ortholog conjecture.

Since phenotypic evolution may arise through mutations that affect gene expression patterns (Guschanski et al. 2017), and that these are readily comparable between genes and species, we have performed large-scale phylogenetic comparative analyses considering two phenotypic traits which characterize gene expression: average expression levels, and tissue specificity (τ). We

focused on teleost fishes since they have undergone a whole genome duplication (the "FishWGD") around 320 million years ago (Mya), followed by a spectacular evolutionary radiation leading to more than half of all vertebrate species (Pasquier et al. 2017; Davesne et al. 2021; Parey et al. 2022). This allowed us to address the following questions using a phylogenetic framework: Do we find a global support for the ortholog conjecture for different types of duplicates and different expression traits? What is the contribution of instantaneous trait jump mechanisms to paralog evolution? We also explored whether changes in expression of paralogs are linked to positive selection at the coding-sequence level.

Results

Phylogenetic independent contrasts support the ortholog conjecture robustly

Phylogenetic Independent Contrasts is a phylogenetic comparative method which is well suited to test whether gene duplication promotes more functional divergence than evolution of genes in the absence of duplication, if care is taken in its application (Dunn et al. 2018; Begum and Robinson-Rechavi 2021). We used gene expression as a proxy for function in vertebrates (Fig. 1), through two quantitative traits: i) tissue specificity, τ (Yanai et al. 2005), and ii) average expression level. Gene expression is a common proxy for gene function because it is amenable to measurement over all genes in several species, and the level and tissue-specificity of expression patterns correlates well with biological function of genes (Kryuchkova-Mostacci and Robinson-Rechavi 2017). To compute average gene expression levels, we normalized gene expression levels such that the cross-species expression data clearly separates into clusters per organ (Supplementary figs. S1A and S1B)). This way gene expression is comparable across species and tissues or organs (Brawand et al. 2011; Fukushima and Pollock 2020).

The phylogenetic independent contrast of each node in a phylogenetic tree reflects independent changes in continuous traits, where higher contrasts capture more evolutionary changes in traits (Felsenstein 1985; Dunn et al. 2018; Begum and Robinson-Rechavi 2021; Begum et al. 2021). If duplication favors trait diversification, we expect higher Phylogenetic Independent Contrast values (PICs) for duplications than for speciations. This is indeed the case for both gene expression traits, and both for standardized phylogenies restricted to teleosts or including all 15 vertebrates of this study (Supplementary fig. S2). These results support the ortholog conjecture for vertebrates. In our analyses, since each data point is a node of a tree, it is possible that gene families with many

129 duplication nodes bias the ortholog conjecture test (Begum and Robinson-Rechavi 2021). Hence,
130 the difference observed could be solely due to a small number of gene families with large numbers
131 of duplications. To verify, we first computed the mean PIC per event for each tree to have one
132 datapoint per gene family, along with the frequency of each event. We then restricted our analysis
133 to gene trees with a maximum of five duplication events, with no limitation in speciation events
134 (Supplementary fig. S3A). This accounts for more than 94% of total gene trees for both the traits.
135 Our support for the ortholog conjecture still holds (Supplementary fig. S3B). This suggests that
136 our observation was not solely due to a small number of gene families with large numbers of
137 duplications.

138 Further analyses on synteny validated teleost third-round whole genome duplicates (FishWGDs)
139 or strict ohnologs, and on strict small-scale duplicates (SSDs) in teleosts (Figs. 2A-2D) also
140 support the ortholog conjecture for both gene expression traits. So do similar analyses on all 15
141 vertebrates (Supplementary figs. S4A-S4D). We observed a much weaker impact of genome
142 duplications than of small-scale duplications on gene expression trait divergence in terms of effect
143 size. This could be because the ohnologs are much older than the small-scale paralogs (median age
144 of ohnologs: 332 Mya, median age of SSDs: 103 Mya, Wilcoxon two-tailed tests $P < 2.2 \times 10^{-16}$),
145 or because of differences in evolutionary dynamics.

146 If the effect of duplications is time-dependent (Gu et al. 2005; Begum et al. 2021), evolution along
147 speciation branches following duplication branches might be impacted by the duplication. This
148 could impact testing of the ortholog conjecture if we simply compare all duplication branches to
149 all speciation branches. To account for this effect, we compared contrasts not only between
150 speciation and duplication nodes, but also between speciation nodes prior to duplications (pre-
151 duplication speciations or "ancestral" speciations) and speciation nodes immediately following

152 duplications (post-duplication speciations or "descendant" speciations) (Fig. 3A). Pre-duplication
153 speciation branches are devoid of any effect of duplication, and thus comparisons of pre-
154 duplication speciation contrasts to its subsequent duplication contrasts provide a more specific test
155 of the ortholog conjecture. Moreover, comparisons of pre- and post-duplication speciation
156 branches provide insight into the long term effect of duplications. We first restricted this
157 comparison to the oldest duplication node along a path in each tree.

158 The contrasts of duplications were always significantly higher than those of ancestral or
159 descendant speciations (Figs. 3B-3E, Supplementary fig. S5). In addition, there was a trend for
160 descendant speciations to have higher contrasts than ancestral speciations, indicating a longer-term
161 impact of the duplication on evolutionary rates. Depending on duplication type and species
162 sampling, this difference was sometimes but not always significant. This might be due in part to
163 small effect sizes which are difficult to test on small subsets of the data, e.g. only ~14% of the
164 4139 strict SSD trees fit the criteria of Fig. 3A; and in part to large divergence times between the
165 FishWGD event and the next speciations. To assess bias in the empirical data sets, we performed
166 randomization of traits (Begum and Robinson-Rechavi 2021). Randomized traits showed a pattern
167 distinct from that of real data in all cases (Supplementary figs. S6 and S7), indicating a biological
168 pattern rather than the sole result of a bias in the data (Begum and Robinson-Rechavi 2021). Indeed
169 after randomization the contrasts no longer follow a Brownian model, and become dependent on
170 phylogenetic parameters such as node depth or node age. Hence, the contrasts increase as the node
171 depths decrease towards present time along the randomized phylogenies (Supplementary figs. S6
172 and S7). Contrasts in randomized phylogenies increase with decreasing node depth, from pre-
173 duplication speciation to duplication to post-duplication speciation nodes, showing that bias is a
174 real concern. Yet unlike in real data, on randomized trees duplication PICs are not higher than

post-duplication speciation PICs, showing that the higher PIC of duplications is not due to such biases.

In case of the youngest duplication node along a path, time-dependent effects of prior duplication(s) might impact the contrasts of pre-duplication speciation branches of the youngest duplication branches. Our result was robust when we considered the youngest duplication node along a path whenever available (Supplementary figs. S8 and S9). This indicates that the pattern we observed was not strongly affected by the timing of duplication nor by the risk of a few wrongly annotated genome duplication nodes. Overall, we find that phylogenetic contrasts support the ortholog conjecture in vertebrate gene expression evolution.

Lévy's trait jump: sudden shifts in gene expression after small-scale duplication

Continuous traits can change in evolution in sudden shifts or "jumps" in values, without a change of the evolutionary rate outside of this jump. Although Gaussian processes (*i.e.*, Brownian motion (BM) and Ornstein–Uhlenbeck (OU) process) have been widely used to model continuous trait evolution in duplicates (Fukushima and Pollock 2020; Begum and Robinson-Rechavi 2021; Gillard et al. 2021), the relevance of rapid trait jumps to gene and genome duplicates has not been studied. If such jumps were common in gene evolution, this would impact characterization of the ortholog conjecture. Hence, we applied the method "levolution" (Duchen et al. 2017; Duchen et al. 2021) that couples a Brownian model with occasional jumps occurring as a Poisson process, to infer the locations and strength of evolutionary jumps for gene expression traits (Fig. 4A and 4B). Under the ortholog conjecture, our expectation is that there should be more jumps in traits after duplications than after speciations.

196 First, we do detect jumps in the evolution of gene expression, affecting more than half of all gene
197 trees (Figs. S10A and S10B), using a branch jump posterior threshold of 0.7. We used a threshold
198 of 0.7 to detect jumps with sufficient accuracy, because uncertainty increases when the posterior
199 probability value decreases (Lemmon and Moriarty 2004; Revell 2014; Pilmann Kotěrová et al.
200 2024). Second, there are indeed more than twice as many jumps on duplication branches than on
201 speciation branches for small scale duplications in teleosts (Figs. 4C and 4D). Third, there are
202 almost no jumps after whole genome duplication in teleosts (Figs. 4E and 4F). There are jumps on
203 the speciation and SSD branches of WGD trees, thus the lack of jumps on WGD branches is not
204 due to a bias in WGD genes. If different types of duplications are not distinguished, in a naive
205 testing of the ortholog conjecture, there is overall a significantly higher proportion of jumps per
206 branch for both the gene expression traits in teleosts and in vertebrates following duplications than
207 speciations (Supplementary figs. S10C-S10F). Such rapid trait shifts are slightly more asymmetric
208 for τ (asymmetric *vs.* symmetric jump branches: 51.44% *vs.* 48.56%, two-sample test for equality
209 of proportions, χ^2 test, $P = 4.05 \times 10^{-3}$) and are slightly more symmetric for average gene
210 expression levels (asymmetric *vs.* symmetric jump branches: 48.58% *vs.* 51.42%, two-sample test
211 for equality of proportions, χ^2 test, $P = 1.57 \times 10^{-7}$). Our results were robust after changing the
212 branch jump posterior threshold from 0.7 to 0.5 (Supplementary Tables S1 and S2).

213 When we applied PICs on the phylogenies supporting trait jumps, we observed a clear support for
214 the ortholog conjecture, with more than 21% higher contrasts for paralogs than orthologs
215 (Supplementary figs. S11A and S11B). This suggests that higher PICs could sometimes be due to
216 these rapid shifts in trait values. When we removed the contrasts of the nodes for which the
217 daughter branch(es) experience rapid shifts in their traits, the ortholog conjecture is no longer
218 supported for the traits (Supplementary figs. S11C and S11D). This confirms that trait jumps play

219 an important role in duplicate evolution. Interestingly, there is also support for the ortholog
220 conjecture for the phylogenies not supporting a trait jump model (Supplementary figs. S11E and
221 S11F), implying that these are genes which evolve after duplication in a more continuous manner.
222 Our results were robust when we replicated our analyses on teleosts only (Fig. 5; Supplementary
223 fig. S12). Of note, the nodes with high support for jumps are not always the same as nodes with
224 high PICs (Supplementary fig. S13). Elevated contrasts and high probability of trait jumps thus do
225 not describe the same process.

226 Tissue-specific trait jumps

227 We next performed our analyses separately on each tissue to investigate whether jumps in
228 expression following duplication were tissue-specific or shared between tissues. Since the trait
229 jump model was supported for SSD and speciation branches, but not genome duplication branches,
230 in the global analyses (Figs. 4 and 5; Supplementary figs. S11, S12 and S14), we restricted the
231 tissue-level analysis to strict SSD trees. Most of these trees supported the trait jump model for all
232 6 tissues (Fig. 6).

233 When we applied our PICs on the trusted SSD phylogenies supporting jumps in their traits for
234 individual tissues, we observed a strong support for the ortholog conjecture in brain, kidney and
235 testis (Supplementary figs. S15-S20). There is also support for the ortholog conjecture in heart,
236 but only in the teleost-only analysis (Supplementary fig. S16). Removal of nodes for which one or
237 of both the daughter branches experiences rapid shifts in their gene expression traits removes the
238 signal of the ortholog conjecture for these four tissues, including teleost-only heart (Supplementary
239 figs. S15C-S15F, S16C, S16F, S17C-S17F, S20C-S20F).

240 **Relation between coding sequence positive selection and evolutionary shifts in**
241 **expression**

242 To investigate a potential role of positive selection on genes which shift expression after
243 duplication, we turned to the detection of positively selected sites in the protein coding sequence.
244 We aim to test whether positive selection on the protein is associated with rapid shifts in expression
245 of small-scale duplicates.

246 Parsing the positive selection data for vertebrates from the Selectome database (Moretti et al. 2014)
247 and applying a false-discovery-rate correction (Storey and Tibshirani 2003; Anisimova and Yang
248 2007; Studer et al. 2009) cutoff of $q < 0.10$, we found less than 100 cases with both protein-level
249 positive selection and expression jumps. To gain power, we used the sum of log-likelihood ratios
250 (Daub et al. 2017), where higher log-likelihood ratio (ΔlnL) score indicates that the data have more
251 support for positive selection relative to a nearly-neutral null model. Summing these scores
252 provides the overall support for positive selection over a set of genes. Branches supporting rapid
253 shifts in expression traits have indeed higher ΔlnL scores for paralogs than for orthologs, thus more
254 evidence for positive selection (Supplementary figs. S21A and S21B). While the highest ΔlnL
255 scores were for paralogs with expression shifts, removal of branches supporting rapid jumps in
256 gene expression traits still showed higher ΔlnL scores for paralogs (Supplementary figs. S21C and
257 S21D). Our trends remain mostly unchanged when we repeated our analyses for teleosts only (Fig.
258 7), with smaller sample sizes. Thus there is overall more positive selection on protein-coding
259 sequences after duplication than speciation, and even more for genes which shift in expression.
260 This supports the ortholog conjecture regarding protein sequence evolutionary rates, as observed
261 by David et al. (2020) and Fukushima and Pollock (2020).

262

Discussion

Understanding the dynamics of gene function change after duplication is important both to our fundamental knowledge of gene evolution, and to the practical use of the ortholog conjecture (Koonin 2005; Studer and Robinson-Rechavi 2009; Nehrt et al. 2011; Altenhoff et al. 2012; Chen and Zhang 2012; Gabaldón and Koonin 2013; Rogozin et al. 2014; Kryuchkova-Mostacci and Robinson-Rechavi 2016; Dunn et al. 2018; Fukushima and Pollock 2020; Stamboulian et al. 2020; Begum and Robinson-Rechavi 2021). Indeed, the function of the majority of proteins is not experimentally determined, but rather is extrapolated from homologs, often specifically from orthologs. Although the ortholog conjecture remains debated (Nehrt et al. 2011; Dunn et al. 2018; Stamboulian et al. 2020), we previously found some support for the conjecture for a single trait, tissue specificity τ, from eight vertebrate species using a properly controlled phylogenetic independent contrasts method, as well as using phylogenetic modeling (Begum and Robinson-Rechavi 2021). In that study we did not distinguish types of duplication, and only tested for branch-specific changes in gradual rates.

Here we provide strong confirmation for the ortholog conjecture on gene expression in vertebrates: it holds both for average expression and for tissue-specificity; it holds with a use of phylogenetic independent contrasts which accounts for the long-term impact of duplication on the next speciation branch; and it holds for both small-scale and whole genome duplicates. Our results are in agreement with the results of pairwise-based analyses (Altenhoff et al. 2012; Chen and Zhang 2012; Gabaldón and Koonin 2013; Rogozin et al. 2014; Kryuchkova-Mostacci and Robinson-Rechavi 2016). While the usual interpretation of the orthologue conjecture has considered a change in sequence or trait evolutionary rates over a phylogenetic branch, there could also be sudden

285 changes of trait values without altering the evolutionary rates (Begum et al. 2021). And indeed,
286 we find evidence for such a rapid trait jump model for evolution after duplication.

287 Evolutionary jumps have been shown to affect the evolution of traits in adaptive landscapes
288 (Simpson 1944; Bokma 2008; Landis et al. 2013; Duchen et al. 2017; Duchen et al. 2021; Gao and
289 Wu 2022), but have been rarely applied to the evolution of gene function. Yet small-scale gene
290 duplication can be a rapid and potentially asymmetric phenomenon. Indeed, we find evidence for
291 the existence of a trait jump mechanism for small-scale duplicates. This can correspond to a sudden
292 change of function, e.g., from incomplete duplication of the coding portion of the progenitor locus
293 or of its regulatory regions or a change in chromatin environment (Lynch and Katju 2004; Dennis
294 et al. 2012; Dougherty et al. 2017; Guschanski et al. 2017). Paralogs can differ in structure from
295 their progenitor copy (Dougherty et al. 2017), and can acquire novel promoters (e.g., gene fusion),
296 or partially or completely lose regulatory elements (e.g., RNA-mediated duplication, Kaessmann
297 et al. 2009). These processes can lead to rapid functional changes, notably through altered
298 expression patterns. These mechanisms rapidly change the trait values rather than changing their
299 evolutionary rates over long evolutionary time. Tissue-specific support for trait jumps after
300 duplication appears to differ between organs, with strongest support in brain, testis and kidney
301 (Figs S15-S20). On the other hand, genome duplication does not seem to lead to sudden or very
302 asymmetric changes in gene expression, as shown by the lack of support for the jump model, which
303 is consistent with the short-term conservation of regulatory environment (Figs. 2 and 3,
304 Supplementary figs. S4 and S9). Interestingly, even for small-scale duplicates, those which do not
305 fit a jump model still support the ortholog conjecture, i.e. with a change in gradual evolutionary
306 rate. Thus the two models of divergence after duplication co-exist: rapid jumps after some small-
307 scale duplications, and gradual changes after whole genome and other small-scale duplications.

308 Positive selection on the coding sequence can contribute to the evolution of gene function
309 following duplication (Han et al. 2009; Pegueroles et al. 2013). We found indeed in general higher
310 evidence for positive selection after duplication. We also find slightly more evidence for positive
311 selection on the branches supporting rapid jumps in their traits following duplication than
312 speciation, but tests are not significant. This might be in part due to a lack of power, as sample
313 sizes for genes with evidence of both processes are small. Thus we cannot exclude a relation, but
314 if there is one it is weak and would need larger sample sizes to detect reliably.

315 In conclusion, we find strong support for rapid evolution of gene expressions after duplication,
316 both by sudden jumps and by long-term acceleration of evolutionary rates. Processes differ
317 between small-scale and genome-scale duplications, but in all cases support the ortholog
318 conjecture, i.e. stronger divergence of paralogs than of orthologs. Thus duplication is critical to
319 the evolution of genome function, and is best studied with a careful use of phylogenetic methods.

320

Materials and Methods

Processing gene trees

43491 empirical gene trees were parsed from Genomicus v.95.01 (Nguyen et al. 2022). This version includes the division of some large trees (> ~400 homologs) into subtrees, like Ensembl in later versions (Martin et al. 2023). Our analyses should not be affected by this, since the missing nodes in the split trees are mainly first round or second round WGD nodes at the base of vertebrates (Parey et al. 2020), which are not the focus of this study. Trees were processed to annotate "speciation", and "duplication" to the internal node events using available information from the tree data object ("D" = 0 as "speciation" and "D" = "Y" as "duplication"). We estimated tissue specificity, τ, and average gene expression levels from ten tetrapods (human, mouse, rat, macaque, cow, dog, ferret, opossum, rabbit, and chicken) and from five teleost fishes (zebrafish, medaka, cavefish, northern pike, and tilapia) across six common tissues (brain, heart, kidney, liver, muscle, and testis), and added those estimates to the tree data objects according to their corresponding Ensembl gene ID. After pruning the gene trees to remove tips with missing trait estimates we obtained 17103 trees. We excluded 148 gene trees with only duplication nodes and 7916 gene trees with no duplication nodes, and retained 9039 trees with at least one speciation and one duplication node for further processing. Cross-species normalized gene expression levels (see below) were then added to the tree data object, and all 9039 trees were retained after filtering for the eventuality of trees with genes that were not expressed in any of the six tissues or which had less than four tips.

The gene trees were then time calibrated using speciation clade ages from Timetree (Hedges et al. 2006). For time calibration, we used the "correlated" model of chronos() function from the R "ape"

343 package (Paradis et al. 2004; Dunn et al. 2018; Begum and Robinson-Rechavi 2021). Since the
344 focus of the study was teleost species, we considered tetrapods as the outgroups. We, therefore,
345 excluded trees with no teleost tip data for our analyses. We thus selected 6923 trees for which at
346 least three tips belonged to teleosts, and at least one tip contained trait data for an outgroup tetrapod
347 species.

348 ***Estimation of phylogenetic independent contrasts (PICs)***

349 For each individual trait, phylogenetic independent contrast or PIC was estimated for each internal
350 node of a gene tree using the pic() function of the R "ape" package (Paradis et al. 2004). PIC of a
351 parental node is an absolute ratio of changes in trait values for descendant nodes to the lengths of
352 its two descendant branches, *i.e.*, the expected variance (Felsenstein 1985; Dunn et al. 2018;
353 Begum and Robinson-Rechavi 2021), in millions of years (My).

354 ***Classification of gene trees based on types of duplications***

355 We identified synteny validated teleost-specific third round (relative to previous two rounds at the
356 base of vertebrates) whole genome duplicates (FishWGDs), or ohnologs, from the database
357 OHNOLOGS v.2 (Singh and Isambert 2020). Out of five teleosts, we obtained 3R ohnolog pairs
358 only for two (*e.g.*, zebrafish and medaka) from this database. We applied relaxed criteria (q-score
359 (outgroups) < 0.05 and q-score (self-comparison) < 0.3) as available in OHNOLOGS database
360 (Singh and Isambert 2020) to obtain a large dataset of ohnologs. Only "protein_coding" genes with
361 duplication time annotated as "FishWGD" were used. We annotated the most recent common
362 ancestral node of the identified ohnolog pairs in each gene tree as "FishWGD". Other duplication
363 nodes in the same trees were annotated as small-scale gene duplication or "SSD".

364 The teleost fish underwent a third round of whole genome duplication at the base of the
365 Osteoglossocephalai or the Clupeocephala clade, which has been dated 225 − 333 My ago
366 (Berthelot et al. 2014; Parey et al. 2020; Parey et al. 2022). This ambiguity in the position of the
367 teleost WGD is presumably because Genomicus v.95 trees were based on a different phylogeny
368 than that recently resolved in Parey et al. (2023). We thus considered "FishWGD" annotated to
369 Osteoglossocephalai or Clupeocephala clade descendants as "confident" FishWGD ohnologs. We
370 called the 1159 phylogenies with such confident ohnologs "strict WGD" trees. Other ohnolog-
371 annotated genes of the OHNOLOGS database were classified as "dubious 3RWGDs" when their
372 duplication time did not match with "FishWGD" or when the "FishWGD" annotation matched
373 with taxonomic levels other than Osteoglossocephalai or Clupeocephala. We thus obtained 628
374 "dubious WGD" trees.

375 To identify small scale duplication (SSD) trees, we first excluded the 1787 strict or dubious WGD
376 gene trees. From the remaining 5136 gene trees, we then eliminated phylogenies where duplication
377 was present at the Clupeocephala or Osteoglossocephalai taxonomic level. We called those 4139
378 trees "strict SSD" trees. The other 997 phylogenies were called "dubious SSD" trees, where
379 duplication occurred at the same Clupeocephala or Osteoglossocephalai clade as the 3R WGD. To
380 perform our phylogenetic analysis on dubious SSD trees, we removed contrasts (PICs) of those
381 Clupeocephala and Osteoglossocephalai clades.

382 ***Selection of contrasts standardized trees***

383 Since the performance of the PIC method primarily depends on the assumption of Brownian model
384 (BM) of trait evolution, we need to identify phylogenies for which contrasts (PICs) are adequately
385 standardized according to BM (Felsenstein 1985; Garland 1992; Diaz-Uriarte and Garland 1996;

386 Díaz-Uriarte and Garland 1998; Freckleton and Harvey 2006; Cooper et al. 2016; Begum and
387 Robinson-Rechavi 2021). Node contrasts (PICs) for those trees should be independent of internal
388 phylogenetic parameters, *i.e.*, logarithm of node age, expected standard deviations and node depth
389 in this case. We used diagnostic tests on phylogenies with duplications to draw unbiased inference
390 by PIC method, as recommended earlier (Garland 1992; Diaz-Uriarte and Garland 1996; Díaz-
391 Uriarte and Garland 1998; Freckleton and Harvey 2006; Cooper et al. 2016; Begum and Robinson-
392 Rechavi 2021). In these tests, we perform correlation between the absolute values of PICs and
393 each of the three internal phylogenetic parameters individually for each trait per phylogeny. Any
394 significant correlation ($P < 0.05$) indicates departure from BM, *i.e.*, phylogenetic dependence for
395 that tree. By eliminating trees showing a significant trend (positive or negative), we considered
396 sets of trees passing all three diagnostic tests for further analyses for each trait of our interest. We
397 estimated node age and node depth using branching.times() and node.depth() functions from the
398 R "ape" package, respectively, for each tree (Paradis et al. 2004). We applied the crunch() function
399 of the caper package to perform diagnostic tests (Purvis and Rambaut 1995; Cooper et al. 2016;
400 Orme 2018; Begum and Robinson-Rechavi 2021).

401 Following this, 746 strict FishWGD trees and 2482 strict SSD trees passed diagnostic tests for τ,
402 and 774 strict FishWGD trees and 2717 strict SSD trees passed diagnostic tests for average gene
403 expression levels. Overall, we obtained 4247 and 4603 trees passing diagnostic tests for τ and
404 average gene expression levels, respectively, including "dubious" FishWGD or SSD trees.

405 ***Randomization test***

406 To confirm biological effects vs. biases in data structure, we used randomization test of trait values
407 on different sets of contrasts standardized trees, following Begum and Robinson-Rechavi (2021).

408 To randomize, we permuted actual trait values of the tips without changing internal node events
409 for each tree. After randomization, we recalculated the node contrasts (PICs) for each tree using
410 permuted trait values of tips following above stated procedure.

411 ***Expression data***

412 We retrieved processed adult stage paired-end RNA-seq data for seven tetrapods (human:
413 GSE30611, mouse: GSE41367, rat: GSE41367, macaque: GSE41367, cow: GSE41367, opossum:
414 GSE30352, and chicken: GSE41367) and one teleost fish (zebrafish: SRP044781) from Bgee
415 database v.14 (Bastian et al. 2021). Pre-processed expression data for three other tetrapods (dog,
416 ferret, and rabbit) submitted to the NCBI Gene Expression Omnibus (GEO;
417 http://www.ncbi.nlm.nih.gov/geo/) by Chen et al. (Chen et al. 2019) under accession number
418 GSE106077 were directly downloaded from https://portals.broadinstitute.org/evee/. Paired-end
419 Illumina RNA-seq reads of three teleosts (medaka, cavefish, and northern pike) were retrieved
420 from the PhyloFish database (Pasquier et al. 2016). All the tissue data of PhyloFish came from
421 mature stage, except for testis, which came either from adult, or from unknown stage. We fetched
422 paired-end RNA-seq raw reads for tilapia fish directly from NCBI using accession number
423 SRR391690. Developmental stage information was unavailable for dog, ferret, rabbit, and tilapia.

424 Quality assessment of raw and trimmed reads was conducted using FastQC v.0.11.5 (Andrew
425 2019). Quality filtration was performed using Trimmomatic v.0.36 (Bolger et al. 2014) to discard
426 adaptor sequence or low-quality reads (settings: LEADING:4 TRAILING:4
427 SLIDINGWINDOW:4:20 HEADCROP:12 MINLEN:25). Transcript information was
428 downloaded from Ensemble v.95 (Martin et al. 2023) and was used as reference for fast and
429 accurate mapping against the cleaned reads with the aid of kallisto v.0.43.1 (Bray et al. 2016)
430 pseudoaligner. For rapid RNA-seq quantification, we applied quant argument of kallisto.

431 We used transcript abundances in transcripts per million as an estimate of gene expression levels,
432 with log transformation (log_2TPM). In case of tilapia, we had three biological replicates for each
433 tissue sample. We took average TPM for all the replicates against each tissue to obtain per tissue
434 gene expression level for each gene. TPM values from several sub parts of the tissues for tetrapods
435 from Bgee database (Bastian et al. 2021) are averaged to obtain gene expression levels of the main
436 organ or tissue. Teleosts specifically have an organ called "head kidney", which comprises about
437 20% of the anterior part of the kidney. Considering it as "kidney", we obtained six common tissues
438 (brain, heart, kidney, liver, muscle, and testis) for all the vertebrates of this study.

439 ***Normalization of gene expression data***

440 For a few species of this study, the developmental stages of tissues were unknown (see above).
441 Most PhyloFish samples, except testis, were extracted from females (Pasquier et al. 2016). On the
442 other hand, data from Bgee (Bastian et al. 2021) sometimes came from male, sometimes from
443 female, or also from mixed sex in some cases (e.g., heart and kidney expression of zebrafish). In
444 addition, these data come from different studies and different species. Thus we needed to normalize
445 gene expression levels to compare them across species and tissues.

446 For cross-species expression data normalization, we adopted a scaling procedure following
447 Brawand et al. (2011). Among the 1077 1:1 orthologs of all vertebrate genes with their expression
448 values in the interquartile range, we identified the top 500 conserved genes that have the maximum
449 conserved ranks among tissues. Median expression levels for those 500 genes were assessed in
450 each tissue for each species. We then derived scaling factors for each tissue of a species using
451 median of human tissues as references. Finally, we applied those scaling factors on the log
452 transformed corresponding tissue dataset of 15 species to obtain normalized gene expression level
453 for each gene of a species in that tissue.

454 ***Estimation of trait data***

455 Tissue specificity, τ, of a gene was computed across all the six tissues of each species using the
456 following equation (Yanai et al. 2005):

457 $$\tau = \frac{\sum_{i=1}^{N}(1-x_i)}{N-1},$$

458 where N is the total number of tissues or organs and x_i is the $\log_2$(TPM) value of gene x in i^{th}
459 tissue. τ close to 1 indicates higher tissue specific function of a gene, while more ubiquitous
460 expression is denoted by values close to 0 (Yanai et al. 2005; Kryuchkova-Mostacci and Robinson-
461 Rechavi 2016).

462 Average gene expression level was estimated by taking the mean of cross species scaling factor
463 normalized $\log_2$(TPM) values across all the six tissues for each gene per species.

464 ***Estimation of rapid trait jump***

465 Evolutionary jumps were estimated with the software "levolution" (Duchen et al. 2017). The input
466 consists of an ultrametric tree in Newick format, and a two-column file with trait data for each tip
467 of the tree. In our case the trees used were the gene trees described above, and the trait data were
468 the normalized gene expression for the different tissues, as well as for the τ and average gene
469 expression levels. "levolution" was then run separately for individual trait on contrast standardized
470 gene trees.

471 Briefly, "levolution" works by calculating the variance-covariance matrix of the traits, and
472 assigning a starting vector of jump configurations at every branch of the tree. Through a Markov
473 chain Monte Carlo (MCMC) method, jump configurations are then updated one branch at the time
474 with fixed hierarchical parameters, such as the jump rate, Brownian rate, jump strength, and root

475 state. Two steps take place during the MCMC run. First, an Expectation-Maximization approach
476 is used to infer maximum likelihood (ML) estimates of these hierarchical parameters. Then, the
477 position of jumps along the tree is inferred by using an empirical Bayes approach, where the
478 hierarchical parameters are fixed to their ML estimate, and the MCMC is then used to calculate
479 posterior probabilities of jumps at a particular branch (Duchen et al. 2017).

480 Mathematically, this jump model is depicted as follows. Given $\boldsymbol{x}$ as the set of traits at the tips of
481 the tree, the likelihood of $\boldsymbol{x}$ given the hierarchical parameters is $\mathrm{f}(\mathrm{x}|\text{hierarchical parameters}) =$
482 $\sum_{\mathrm{n}=1}^{\infty} \mathrm{P}(\nu = \mathrm{n})\phi(\mathrm{x}|\mathrm{n})$, where $\boldsymbol{n}$ is the number of jumps, $P(\nu = n)$ is the Poisson probability of
483 observing $\boldsymbol{n}$ jumps, and $\phi(x|n)$ is the multivariate normal probability of observing traits $\boldsymbol{x}$ given
484 that $\boldsymbol{n}$ jumps are present (Duchen et al. 2017, Eq. 4). The sum over all possible jump configurations
485 is taken over by the MCMC approach described above, and the entries in the variance covariance
486 matrix in $\phi(x|n)$ are given by the sum of the variance present along each branch and the variance
487 due to the jumps.

488 ***Analyses of positive selection***

489 We retrieved precomputed statistics on positive selection from Selectome (Moretti et al. 2014), a
490 database that applies the branch-site likelihood test for selection to Ensembl gene trees (release 98,
491 Martin et al. 2023). Each processed Genomicus tree (See 'Processing gene trees') was mapped, if
492 possible, to the corresponding Selectome tree by Ensembl gene IDs. Of 6923 processed Genomicus
493 trees, 6649 trees were successfully mapped to their corresponding tree in Selectome. Because the
494 Ensembl-based Selectome trees sample more taxa than we have included in our study, we retrieved
495 the likelihood-ratio test statistic only for nodes that have matching taxonomic labels between each
496 mapped Selectome and Genomicus tree, and excluded taxa that are not contained in the former.

There were 116 (respectively 402) strict SSD trees supporting both positive selection as well as jumps in τ (respectively in average expression levels), in their branches for the 15 vertebrates. For teleost-specific analyses, there were only 31 and 78 such trees.

Script and data availability

All the scripts and gene expression data files are available on GitHub: https://github.com/tbegum/Phylogenetic-modeling-of-evolutionary-trait-jumps. For reproducibility, "norm_exp_15spe_4tips_new.RData" file is archived at Zenodo (https://doi.org/10.5281/zenodo.10418895).

Details of other R packages used

We performed other analyses and plotting in R version 4.0.2 (R Core Team 2018) using rphylopic (Chamberlain 2018), phytools (Revell 2012), treeio (Guangchuang 2018), ggtree (Guangchuang et al. 2017), digest (Antoine Lucas et al. 2018), dplyr (Wickham et al. 2017), ggplot2 (Wickham 2016); gtools (Warnes et al. 2018), gridExtra (Auguie 2017), tidyverse (Wickham 2017), stringr (Wickham 2019), and png (Urbanek 2013) libraries.

Acknowledgements

We sincerely thank all the members of the Robinson-Rechavi and Salamin groups for their help and useful discussions. Parts of the computations were performed at the Vital-IT Center for high-performance computing of the SIB Swiss Institute of Bioinformatics, as well as using the Wally cluster of the University of Lausanne. Work supported by Swiss National Science Foundation grant 31003A_173048.

518 **References**

519

520 Altenhoff AM, Studer RA, Robinson-Rechavi M, Dessimoz C. 2012. Resolving the ortholog
521 conjecture: Orthologs tend to be weakly, but significantly, more similar in function than paralogs.
522 PLoS Comput Biol 8:e1002514. Available from:
523 https://www.ncbi.nlm.nih.gov/pubmed/22615551

524 Andrew S. 2019. FastQC: A quality control tool for high throughput sequence data 2010.

525 Anisimova M, Yang Z. 2007. Multiple hypothesis testing to detect lineages under positive
526 selection that affects only a few sites. Mol Biol Evol 24:1219–1228. Available from:
527 https://www.ncbi.nlm.nih.gov/pubmed/17339634

528 Antoine Lucas DE with contributions by, Tuszynski J, Bengtsson H, Urbanek S, Frasca M, Lewis
529 B, Stokely M, Muehleisen H, Murdoch D, Hester J, et al. 2018. Digest: Create compact hash
530 digests of r objects. Available from: https://CRAN.R-project.org/package=digest

531 Assis R, Bachtrog D. 2013. Neofunctionalization of young duplicate genes in drosophila. Proc
532 Natl Acad Sci U S A 110:17409–17414. Available from:
533 https://www.ncbi.nlm.nih.gov/pubmed/24101476

534 Assis R, Bachtrog D. 2015. Rapid divergence and diversification of mammalian duplicate gene
535 functions. BMC Evol Biol 15:138. Available from:
536 https://www.ncbi.nlm.nih.gov/pubmed/26173681

537 Auguie B. 2017. GridExtra: Miscellaneous functions for "grid" graphics. Available from:
538 https://CRAN.R-project.org/package=gridExtra

539 Bastian FB, Roux J, Niknejad A, Comte A, Fonseca Costa SS, Farias TM de, Moretti S, Parmentier
540 G, Laval VR de, Rosikiewicz M, et al. 2021. The Bgee suite: Integrated curated expression atlas
541 and comparative transcriptomics in animals. Nucleic Acids Res 49:D831–D847. Available from:
542 https://www.ncbi.nlm.nih.gov/pubmed/33037820

543 Beaulieu JM, Jhwueng DC, Boettiger C, O'Meara BC. 2012. Modeling stabilizing selection:
544 Expanding the ornstein-uhlenbeck model of adaptive evolution. Evolution 66:2369–2383.
545 Available from: https://www.ncbi.nlm.nih.gov/pubmed/22834738

546 Begum T, Robinson-Rechavi M. 2021. Special care is needed in applying phylogenetic
547 comparative methods to gene trees with speciation and duplication nodes. Mol Biol Evol 38:1614–
548 1626. Available from: https://www.ncbi.nlm.nih.gov/pubmed/33169790

549 Begum T, Serrano-Serrano ML, Robinson-Rechavi M. 2021. Performance of a phylogenetic
550 independent contrast method and an improved pairwise comparison under different scenarios of
551 trait evolution after speciation and duplication. Methods Ecol Evol 12:1875–1887.

552 Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, Bento P, Da Silva C, Labadie
553 K, Alberti A, et al. 2014. The rainbow trout genome provides novel insights into evolution after
554 whole-genome duplication in vertebrates. Nat Commun 5:3657. Available from:
555 https://www.ncbi.nlm.nih.gov/pubmed/24755649

556 Blomberg SP, Garland T, Ives AR. 2003. Testing for phylogenetic signal in comparative data:
557 Behavioral traits are more labile. Evolution 57:717–745. Available from:
558 https://www.ncbi.nlm.nih.gov/pubmed/12778543

559 Bokma F. 2008. Detection of "punctuated equilibrium" by bayesian estimation of speciation and
560 extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a
561 molecular phylogeny. Evolution 62:2718–2726. Available from:
562 https://www.ncbi.nlm.nih.gov/pubmed/18752617

563 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for illumina sequence
564 data. Bioinformatics 30:2114–2120. Available from:
565 https://www.ncbi.nlm.nih.gov/pubmed/24695404

566 Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T,
567 Batzel P, Catchen J, et al. 2016. The spotted gar genome illuminates vertebrate evolution and
568 facilitates human-teleost comparisons. Nat Genet 48:427–437. Available from:
569 https://www.ncbi.nlm.nih.gov/pubmed/26950095

570 Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A,
571 Aximu-Petri A, Kircher M, et al. 2011. The evolution of gene expression levels in mammalian
572 organs. Nature 478:343–348. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22012392

573 Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic rna-seq
574 quantification. Nat Biotechnol 34:525–527.

575 Brunet FG, Roest Crollius H, Paris M, Aury JM, Gibert P, Jaillon O, Laudet V, Robinson-Rechavi
576 M. 2006. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes.
577 Mol Biol Evol 23:1808–1816. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16809621

578 Chamberlain S. 2018. Rphylopic: Get’silhouettes’ of’organisms’ from’Phylopic’. R package
579 version 0.2. 0.

580 Chen J, Swofford R, Johnson J, Cummings BB, Rogel N, Lindblad-Toh K, Haerty W, Palma FD,
581 Regev A. 2019. A quantitative framework for characterizing the evolutionary history of
582 mammalian gene expression. Genome Res 29:53–63. Available from:
583 https://www.ncbi.nlm.nih.gov/pubmed/30552105

584 Chen X, Zhang J. 2012. The ortholog conjecture is untestable by the current gene ontology but is
585 supported by rna sequencing data. PLoS Comput Biol 8:e1002784. Available from:
586 https://www.ncbi.nlm.nih.gov/pubmed/23209392

587 Conant GC, Wagner A. 2003. Asymmetric sequence divergence of duplicate genes. Genome Res
588 13:2052–2058. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12952876

589 Cooper N, Thomas GH, FitzJohn RG. 2016. Shedding light on the 'dark side' of phylogenetic
590 comparative methods. Methods Ecol Evol 7:693–699. Available from:
591 https://www.ncbi.nlm.nih.gov/pubmed/27499839

592 Cusack BP, Wolfe KH. 2007. Not born equal: Increased rate asymmetry in relocated and
593 retrotransposed rodent gene duplicates. Mol Biol Evol 24:679–686. Available from:
594 https://www.ncbi.nlm.nih.gov/pubmed/17179139

595 Daub JT, Moretti S, Davydov II, Excoffier L, Robinson-Rechavi M. 2017. Detection of pathways
596 affected by positive selection in primate lineages ancestral to humans. Mol Biol Evol 34:391–
597 1402. Available from: https://doi.org/10.1093/molbev/msx083

598 Davesne D, Friedman M, Schmitt AD, Fernandez V, Carnevale G, Ahlberg PE, Sanchez S, Benson
599 RBJ. 2021. Fossilized cell structures identify an ancient origin for the teleost whole-genome

600 duplication. Proc Natl Acad Sci U S A 118. Available from:
601 https://www.ncbi.nlm.nih.gov/pubmed/34301898

602 David KT, Oaks JR, Halanych KM. 2020. Patterns of gene evolution following duplications and
603 speciations in vertebrates. PeerJ 8:e8813. Available from:
604 https://www.ncbi.nlm.nih.gov/pubmed/32266119

605 Dennis MY, Nuttle X, Sudmant PH, Antonacci F, Graves TA, Nefedov M, Rosenfeld JA, Sajjadian
606 S, Malig M, Kotkiewicz H, et al. 2012. Evolution of human-specific neural srgap2 genes by
607 incomplete segmental duplication. Cell 149:912–922. Available from:
608 https://www.ncbi.nlm.nih.gov/pubmed/22559943

609 Diaz-Uriarte R, Garland T. 1996. Testing hypotheses of correlated evolution using
610 phylogenetically independent contrasts: Sensitivity to deviations from brownian motion. Syst Biol
611 45:27–47.

612 Díaz-Uriarte R, Garland T. 1998. Effects of branch length errors on the performance of
613 phylogenetically independent contrasts. Syst Biol 47:654–672. Available from:
614 https://www.ncbi.nlm.nih.gov/pubmed/12066309

615 Dougherty ML, Nuttle X, Penn O, Nelson BJ, Huddleston J, Baker C, Harshman L, Duyzend MH,
616 Ventura M, Antonacci F, et al. 2017. The birth of a human-specific neural gene by incomplete
617 duplication and gene fusion. Genome Biol 18:49. Available from:
618 https://www.ncbi.nlm.nih.gov/pubmed/28279197

619 Duchen P, Alfaro ML, Rolland J, Salamin N, Silvestro D. 2021. On the effect of asymmetrical
620 trait inheritance on models of trait evolution. Syst Biol 70:376–388. Available from:
621 https://www.ncbi.nlm.nih.gov/pubmed/32681798

622 Duchen P, Leuenberger C, Szilágyi SM, Harmon L, Eastman J, Schweizer M, Wegmann D. 2017.
623 Inference of evolutionary jumps in large phylogenies using lévy processes. Syst Biol 66:950–963.
624 Available from: https://www.ncbi.nlm.nih.gov/pubmed/28204787

625 Dunn CW, Zapata F, Munro C, Siebert S, Hejnol A. 2018. Pairwise comparisons across species
626 are problematic when analyzing functional genomic data. Proc Natl Acad Sci U S A 115:E409–
627 E417. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29301966

628 Eastman JM, Alfaro ME, Joyce P, Hipp AL, Harmon LJ. 2011. A novel comparative method for
629 identifying shifts in the rate of character evolution on trees. Evolution 65:3578–3589. Available
630 from: https://www.ncbi.nlm.nih.gov/pubmed/22133227

631 Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution
632 39:783–791. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28561359

633 Freckleton R, Harvey P, Pagel M. 2002. Phylogenetic analysis and comparative data: A test and
634 review of evidence. Am Nat 160:712–726.

635 Freckleton RP, Harvey PH. 2006. Detecting non-brownian trait evolution in adaptive radiations.
636 PLoS Biol 4:e373. Available from: https://www.ncbi.nlm.nih.gov/pubmed/17090217

637 Fukushima K, Pollock DD. 2020. Amalgamated cross-species transcriptomes reveal organ-
638 specific propensity in gene expression evolution. Nat Commun 11:4459. Available from:
639 https://www.ncbi.nlm.nih.gov/pubmed/32900997

640 Gabaldón T, Koonin EV. 2013. Functional and evolutionary implications of gene orthology. Nat
641 Rev Genet 14:360–366. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23552219

642 Gao Y, Wu M. 2022. Microbial genomic trait evolution is dominated by frequent and rare pulsed
643 evolution. Sci Adv 8:eabn1916. Available from: DOI: 10.1126/sciadv.abn19

644 Garland T. 1992. Rate tests for phenotypic evolution using phylogenetically independent contrasts.
645 Am Nat 140:509–519. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19426053

646 Gillard, G.B., Grønvold, L., Røsæg, L.L. et al. 2021. Comparative regulomics supports pervasive
647 selection on gene dosage following whole genome duplication. Genome Biol 22:1-18.

648 Gout JF, Lynch M. 2015. Maintenance and loss of duplicated genes by dosage
649 subfunctionalization. Mol Biol Evol 32:2141–2148. Available from:
650 https://www.ncbi.nlm.nih.gov/pubmed/25908670

651 Grafen A. 1989. The phylogenetic regression. Philos Trans R Soc Lond B Biol Sci 326:119–157.
652 Available from: https://www.ncbi.nlm.nih.gov/pubmed/2575770

653 Gu X, Su Z. 2007. Tissue-driven hypothesis of genomic evolution and sequence-expression
654 correlations. Proc Natl Acad Sci U S A 104:2779–2784. Available from:
655 https://www.ncbi.nlm.nih.gov/pubmed/17301236

656 Gu X, Zhang Z, Huang W. 2005. Rapid evolution of expression and regulatory divergences after
657 yeast gene duplication. Proc Natl Acad Sci U S A 102:707–712. Available from:
658 https://www.ncbi.nlm.nih.gov/pubmed/15647348

659 Guangchuang Y. 2018. Treeio: Base classes and functions for phylogenetic tree input and output.
660 Available from: https://guangchuangyu.github.io/software/treeio

661 Guangchuang Y, David S, Huachen Z, Yi G, Tommy T-YL. 2017. Ggtree: An r package for
662 visualization and annotation of phylogenetic trees with their covariates and other associated data.
663 Methods Ecol Evol 8:28–36.

664 Guschanski K, Warnefors M, Kaessmann H. 2017. The evolution of duplicate gene expression in
665 mammalian organs. Genome Res 27:1461–1474. Available from:
666 https://www.ncbi.nlm.nih.gov/pubmed/28743766

667 Han MV, Demuth JP, McGrath CL, Casola C, Hahn MW. 2009. Adaptive evolution of young gene
668 duplicates in mammals. Genome Res 19:859–867. Available from:
669 https://www.ncbi.nlm.nih.gov/pubmed/19411603

670 Hansen TF. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution
671 51:1341–1351. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28568616

672 Hedges SB, Dudley J, Kumar S. 2006. TimeTree: A public knowledge-base of divergence times
673 among organisms. Bioinformatics 22:2971–2972. Available from:
674 https://www.ncbi.nlm.nih.gov/pubmed/17021158

675 Holland PW, Marlétaz F, Maeso I, Dunwell TL, Paps J. 2017. New genes from old: Asymmetric
676 divergence of gene duplicates and the evolution of development. Philos Trans R Soc Lond B Biol
677 Sci 372. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27994121

678 Kaessmann, H, Vinckenbosch N, Long M. 2009. RNA-based gene duplication: mechanistic and
679 evolutionary insights. Nat Rev Genet 10:19–31. Available from: https://doi.org/10.1038/nrg2487

680 Khabbazian M, Kriebel R, Rohe K, Ané C. 2016. Fast and accurate detection of evolutionary shifts
681 in ornstein-uhlenbeck models. Methods Ecol Evol 7:811–824.

682 Kim SH, Yi SV. 2006. Correlated asymmetry of sequence and functional divergence between
683 duplicate proteins of saccharomyces cerevisiae. Mol Biol Evol 23:1068–1075. Available from:
684 https://www.ncbi.nlm.nih.gov/pubmed/16510556

685 Koonin EV. 2005. Orthologs, paralogs, and evolutionary genomics. Annu Rev Genet 39:309–338.
686 Available from: https://www.ncbi.nlm.nih.gov/pubmed/16285863

687 Kryuchkova-Mostacci N, Robinson-Rechavi M. 2016. Tissue-specificity of gene expression
688 diverges slowly between orthologs, and rapidly between paralogs. PLoS Comput Biol
689 12:e1005274. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28030541

690 Kryuchkova-Mostacci N, Robinson-Rechavi M. 2017. A benchmark of gene expression tissue-
691 specificity metrics. Brief Bioinform 18:205–214. Available from:
692 https://doi.org/10.1093/bib/bbw008

693 Landis MJ, Schraiber JG. 2017. Pulsed evolution shaped modern vertebrate body sizes. Proc Natl
694 Acad Sci USA 114: 13224–13229. Available from:
695 https://www.ncbi.nlm.nih.gov/pubmed/29114046

696 Landis MJ, Schraiber JG, Liang M. 2013. Phylogenetic analysis using lévy processes: Finding
697 jumps in the evolution of continuous traits. Syst Biol 62:193–204. Available from:
698 https://www.ncbi.nlm.nih.gov/pubmed/23034385

699 Lemmon AR, Moriarty EC. 2004. The importance of proper model assumption in Bayesian
700 phylogenetics. Syst Biol 1:265-77.

701 Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley
702 DR, Zimin A, et al. 2016. The atlantic salmon genome provides insights into rediploidization.
703 Nature 533:200–205. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27088604

704 Lynch M, Katju V. 2004. The altered evolutionary trajectories of gene duplicates. Trends Genet
705 20:544–549. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15475113

706 Martins E, Hansen T. 1997. Phylogenies and the comparative method: A general approach to
707 incorporating phylogenetic information into the analysis of interspecific data. Am Nat 149:646–
708 667.

709 Martin FJ, Amode MR, Aneja A, Austine-Orimoloye O, Azov AG, *et al.* 2023. Ensembl 2023.
710 Nucleic Acids Res 51: D933–D941. Available from https://doi.org/10.1093/nar/gkac958

711 Moretti S, Laurenczy B, Gharib WH, Castella B, Kuzniar A, Schabauer H, Studer RA, Valle M,
712 Salamin N, Stockinger H, Robinson-Rechavi M. 2014. Selectome update: quality control and
713 computational improvements to a database of positive selection, Nucleic Acids Res 42:D917–
714 D921. https://doi.org/10.1093/nar/gkt1065

715 Nehrt NL, Clark WT, Radivojac P, Hahn MW. 2011. Testing the ortholog conjecture with
716 comparative functional genomic data from mammals. PLoS Comput Biol 7:e1002073. Available
717 from: https://www.ncbi.nlm.nih.gov/pubmed/21695233

718 Nguyen NTT, Vincens P, Dufayard JF, Roest Crollius H, Louis A. 2022. Genomicus in 2022:
719 Comparative tools for thousands of genomes and reconstructed ancestors. Nucleic Acids Res
720 50:D1025–D1031. Available from: https://www.ncbi.nlm.nih.gov/pubmed/34792170

721 Ohno S. 1970. Evolution by gene duplication. New York (EUA). Springer-Verlag.

722 O'Meara BC, Ané C, Sanderson MJ, Wainwright PC. 2006. Testing for different rates of
723 continuous trait evolution using likelihood. Evolution 60:922–933. Available from:
724 https://www.ncbi.nlm.nih.gov/pubmed/16817533

725 Orme D. 2018. The caper package: Comparative analysis of phylogenetics and evolution in R.
726 Available from: https://cran.r-project.org/web/packages/caper/vignettes/caper.pdf

727 Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

728 Panchin AY, Gelfand MS, Ramensky VE, Artamonova II. 2010. Asymmetric and non-uniform
729 evolution of recently duplicated human genes. Biol Direct 5:54. Available from:
730 https://www.ncbi.nlm.nih.gov/pubmed/20825637

731 Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R
732 language. Bioinformatics 20:289–290. Available from:
733 https://www.ncbi.nlm.nih.gov/pubmed/14734327

734 Parey E, Louis A, Cabau C, Guiguen Y, Roest Crollius H, Berthelot C. 2020. Synteny-guided
735 resolution of gene trees clarifies the functional impact of whole-genome duplications. Mol Biol
736 Evol 37:3324–3337. Available from: https://www.ncbi.nlm.nih.gov/pubmed/32556216

737 Parey E, Louis A, Montfort J, Guiguen Y, Roest Crollius H, Berthelot C. 2022. An atlas of fish
738 genome evolution reveals delayed rediploidization following the teleost whole-genome
739 duplication. Genome Res. Available from: https://www.ncbi.nlm.nih.gov/pubmed/35961774

740 Parey, E, Alexandra L, Jérôme M, et al. 2023. Genome structures resolve the early diversification
741 of teleost fishes. Science 379:572–3575. Available from:
742 https://www.ncbi.nlm.nih.gov/pubmed/36758078

743 Pasquier J, Braasch I, Batzel P, Cabau C, Montfort J, Nguyen T, Jouanno E, Berthelot C, Klopp
744 C, Journot L, et al. 2017. Evolution of gene expression after whole-genome duplication: New
745 insights from the spotted gar genome. J Exp Zool B Mol Dev Evol 328:709–721. Available from:
746 https://www.ncbi.nlm.nih.gov/pubmed/28944589

747 Pasquier J, Cabau C, Nguyen T, Jouanno E, Severac D, Braasch I, Journot L, Pontarotti P, Klopp
748 C, Postlethwait JH, et al. 2016. Gene evolution and gene expression after whole genome
749 duplication in fish: The PhyloFish database. BMC Genomics 17:368. Available from:
750 https://www.ncbi.nlm.nih.gov/pubmed/27189481

751 Pegueroles C, Laurie S, Albà MM. 2013. Accelerated evolution after gene duplication: A time-
752 dependent process affecting just one copy. Mol Biol Evol 30:1830–1842. Available from:
753 https://www.ncbi.nlm.nih.gov/pubmed/23625888

754 Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, Alfaro ME, Harmon
755 LJ. 2014. Geiger v2.0: An expanded suite of methods for fitting macroevolutionary models to
756 phylogenetic trees. Bioinformatics 30:2216–2218. Available from:
757 https://www.ncbi.nlm.nih.gov/pubmed/24728855

758 Pich I Roselló O, Kondrashov FA. 2014. Long-term asymmetrical acceleration of protein evolution
759 after gene duplication. Genome Biol Evol 6:1949–1955. Available from:
760 https://www.ncbi.nlm.nih.gov/pubmed/25070510

761 Pilmann Kotěrová A, Santos F, Bejdová Š. et al. 2024. Prioritizing a high posterior probability
762 threshold leading to low error rate over high classification accuracy: the validity of MorphoPASSE
763 software for cranial morphological sex estimation in a contemporary population. Int J Legal Med.
764 Available from: https://doi.org/10.1007/s00414-024-03215-1

765 Purvis A, Rambaut A. 1995. Comparative analysis by independent contrasts (caic): An apple
766 macintosh application for analysing comparative data. Comput Appl Biosci 11:247–251. Available
767 from: https://www.ncbi.nlm.nih.gov/pubmed/7583692

768 R Core Team. 2018. R: A language and environment for statistical computing. Vienna, Austria: R
769 Foundation for Statistical Computing. Available from: https://www.R-project.org/

770 Revell LJ. 2012. Phytools: An r package for phylogenetic comparative biology (and other things).
771 Methods Ecol Evol 3:217–223.

772 Revell LJ. 2014. Ancestral character estimation under the threshold model from quantitative
773 genetics, Evolution 68:743–759. Available from: https://doi.org/10.1111/evo.12300

774 Rogozin IB, Managadze D, Shabalina SA, Koonin EV. 2014. Gene family level comparative
775 analysis of gene expression in mammals validates the ortholog conjecture. Genome Biol Evol
776 6:754–762. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24610837

777 Rohlfs RV, Nielsen R. 2015. Phylogenetic ANOVA: The expression variance and evolution model
778 for quantitative trait evolution. Syst Biol 64:695–708. Available from:
779 https://www.ncbi.nlm.nih.gov/pubmed/26169525

780 Romero IG, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the evolution
781 of gene regulation. Nat Rev Genet 13:505–516. Available from:
782 https://www.ncbi.nlm.nih.gov/pubmed/22705669

783 Sandve SR, Rohlfs RV, Hvidsten TR. 2018. Subfunctionalization versus neofunctionalization after
784 whole-genome duplication. Nat Genet 50:908–909. Available from:
785 https://www.ncbi.nlm.nih.gov/pubmed/29955176

786 Scannell DR, Wolfe KH. 2008. A burst of protein sequence evolution and a prolonged period of
787 asymmetric evolution follow gene duplication in yeast. Genome Res 18:137–147. Available from:
788 https://www.ncbi.nlm.nih.gov/pubmed/18025270

789 Simpson G. 1944. Tempo and mode in evolution. A wartime book. New York: Columbia
790 University Press.

791 Singh PP, Isambert H. 2020. OHNOLOGS v2: A comprehensive resource for the genes retained
792 from whole genome duplication in vertebrates. Nucleic Acids Res 48:D724–D730. Available
793 from: https://www.ncbi.nlm.nih.gov/pubmed/31612943

794 Stamboulian M, Guerrero RF, Hahn MW, Radivojac P. 2020. The ortholog conjecture revisited:
795 The value of orthologs and paralogs in function prediction. Bioinformatics 36:i219–i226.
796 Available from: https://www.ncbi.nlm.nih.gov/pubmed/32657391

797 Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. Proc Natl Acad Sci
798 U S A 100:9440–9445. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12883005

799 Studer RA, Robinson-Rechavi M. 2009. How confident can we be that orthologs are similar, but
800 paralogs differ? Trends Genet 25:210–216. Available from:
801 https://www.ncbi.nlm.nih.gov/pubmed/19368988

802 Urbanek S. 2013. Png: Read and write png images. Available from: https://CRAN.R-
803 project.org/package=png

804 Uyeda J, Hansen T, Arnold S, Pienaar J. 2011. The million-year wait for macroevolutionary bursts.
805 Proc. Natl Acad. Sci. USA 108:15908–15913.

Uyeda JC, Harmon LJ. 2014. A novel bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. Syst Biol 63:902–918. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25077513

Warnes GR, Bolker B, Lumley T. 2018. Gtools: Various r programming tools. Available from: https://CRAN.R-project.org/package=gtools

Wickham H. 2016. Ggplot2: Elegant graphics for data analysis. Springer-Verlag New York Available from: https://ggplot2.tidyverse.org

Wickham H. 2017. Tidyverse: Easily install and load the 'tidyverse'. Available from: https://CRAN.R-project.org/package=tidyverse

Wickham H. 2019. Stringr: Simple, consistent wrappers for common string operations. Available from: https://CRAN.R-project.org/package=stringr

Wickham H, Francois R, Henry L, Muller K. 2017. Dplyr: A grammar of data manipulation. Available from: https://CRAN.R-project.org/package=dplyr

Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics 21:650–659. Available from: <Go to ISI>://WOS:000227241200012

824 # Figure captions

825

826 **Figure 1**: The vertebrate species phylogeny of this study. The 'brownish' color block highlights 5
827 teleost fishes, and the 'off-white' color block highlights the 10 tetrapods.

828 **Figure 2**: Tests of the ortholog conjecture by phylogenetic independent contrasts according to type
829 of duplication in teleosts. 746 and 774 trees with whole genome duplications were used in (A)-
830 (B), while 2482 and 2717 trees with small-scale gene duplications were used in (C)-(D). PICs:
831 Phylogenetic independent contrasts, FishWGD(s): Fish specific 3R whole genome duplicate(s) (or
832 ohnologs), SSD(s): small-scale duplicate(s). PICs of teleosts specific nodes were used for these
833 analyses. *P* values are from Wilcoxon two-tailed tests. Since trees with whole genome duplicates
834 may also contain SSDs apart from strict FishWGDs, both types of duplicates are used for analyses
835 in (A)-(B).

836 **Figure 3**: Tests of the ortholog conjecture by phylogenetic independent contrasts including
837 speciations before and after duplications in teleosts. (A) Illustration of how we can test the impact
838 of a duplication by comparing duplication nodes with speciation nodes before and after the
839 duplication. (B)-(E) 581, 610, 339 and 377 trees passing our criteria of Fig. 3A, respectively were
840 used for the analyses. We used the oldest duplication node along a path of a phylogeny for
841 identification of pre-duplication and post-duplication speciations. We considered the contrasts of
842 teleosts specific gene or genome duplication and descendant speciation nodes in (B)-(E) in absence
843 of teleost-specific pre-duplication speciation nodes contrasts. In many cases, we obtained two or
844 three post duplication speciation events following a single duplication event. Therefore, the
845 number of nodes used for comparisons in the plot increased relative to the number of gene trees.
846 PICs: Phylogenetic independent contrasts, exp: expression, Predup_speciation: pre-duplication

847 speciation, Postdup_speciation: post-duplication speciation, FishWGD: fish specific 3R whole
848 genome duplication, and SSD: small-scale duplication. Since contrasts are estimated for each tree,
849 paired Wilcoxon tests were used to compare them.

850 **Figure 4**: Trait jump models for tissues-specificity and for average gene expression levels
851 contrasting types of duplication in teleosts. FishWGD: fish specific 3R whole genome duplication,
852 SSD: small-scale duplication, and exp: average gene expression levels. (A) An example of a
853 FishWGD tree to demonstrate the branch-specific posterior probability of trait jumps as inferred
854 by levolution (Duchen et al. 2017; Duchen et al. 2021) in a phylogenetic tree. (B) Phylogenetic
855 simulation following the example tree of Fig. 4A. Colors show branches belonging to different
856 events as shown in Fig. 4A. Branches following jumps in their traits are shown by darker shades.
857 (C)-(E) Proportion of trait jumps according to type of duplications and to expression traits. We
858 used a strict branch jump posterior cutoff of ≥ 0.7 to consider a rapid shift in trait values. Only
859 teleosts-specific clades were used for the analyses. We used 1557 and 1593 strict SSD trees in (B)
860 and (C), and 435 and 398 strict FishWGD trees in (D) and (E), respectively showing at least one
861 branch jump with the above-mentioned threshold along a phylogeny for our analyses. For the
862 comparison of the proportions of trait jumps, we used trees with at least one teleost-specific
863 speciation and duplication. Since proportion of shifts in traits per branch of events is estimated for
864 each tree, paired Wilcoxon test is used to compare the difference.

865 **Figure 5**: Trait jump test of the ortholog conjecture on small-scale duplicate trees in teleosts. We
866 used a strict branch jump posterior cutoff of ≥ 0.7 as an evidence of trait jump. Only teleosts-
867 specific clades were used for the analyses. The ortholog conjecture tests on (A) 1557 calibrated
868 contrasts standardized strict SSD trees supporting rapid jumps in τ, and on (B) 1593 strict SSD
869 trees supporting jumps in average gene expression levels with the above-mentioned threshold. (C)-

870 (D) The ortholog conjecture tests after removing contrasts of nodes whose daughter branch(es)
871 experienced jumps in the corresponding trait for 1557 and 1593 strict SSD trees, respectively. (E)-
872 (F) The ortholog conjecture tests on 925 and 1124 strict SSD trees for which the trait jump model
873 is not supported for τ and for average gene expression levels by any of their branches, respectively.
874 *P* values are from Wilcoxon two-tailed tests.

875 **Figure 6**: Statistics of contrast standardized small-scale duplicate trees supporting jumps in each
876 tissue. Only strict SSD trees were used. For each tissue, there are more trees supporting trait jumps
877 than not (two-sample test for equality of proportions, χ^2 test, $P < 2.2 \times 10^{-16}$).

878 **Figure 7**: Positive selection and trait divergence analyses for contrast standardized strict SSD trees
879 in 5 teleost fishes. We only used phylogenies supporting an evolutionary trait jump, and then
880 considered teleosts-specific branches for these analyses. We used a strict posterior probability
881 cutoff of ≥ 0.7 to identify branches with rapid jumps in traits. exps: Gene expression levels, PICs:
882 phylogenetic independent contrasts, LRT: likelihood ratio test.

883

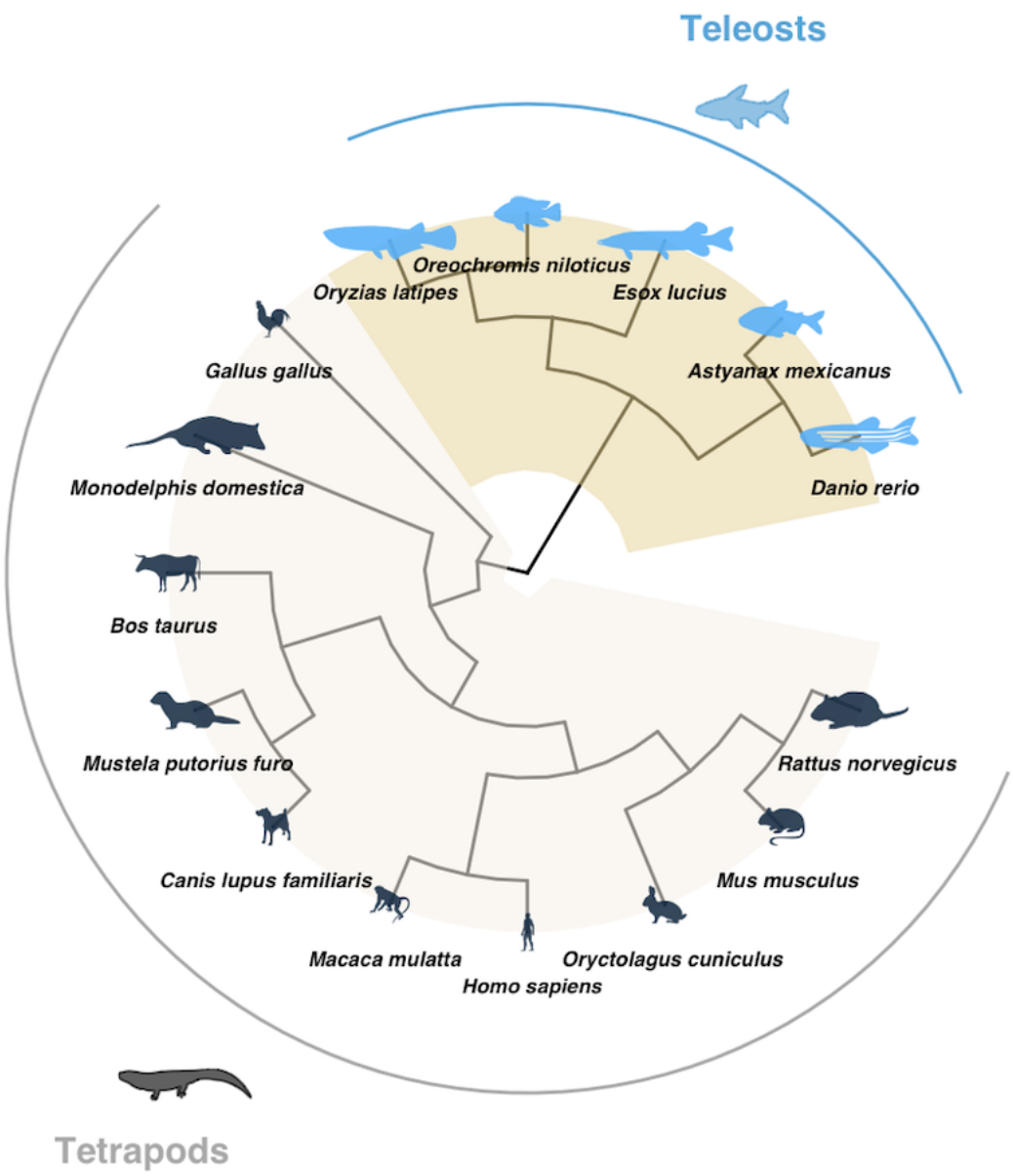

Teleosts
Oreochromis niloticus
Oryzias latipes
Esox lucius
Gallus gallus
Astyanax mexicanus
Monodelphis domestica
Danio rerio
Bos taurus
Mustela putorius furo
Rattus norvegicus
Canis lupus familiaris
Mus musculus
Macaca mulatta
Oryctolagus cuniculus
Homo sapiens
Tetrapods

Fig. 1

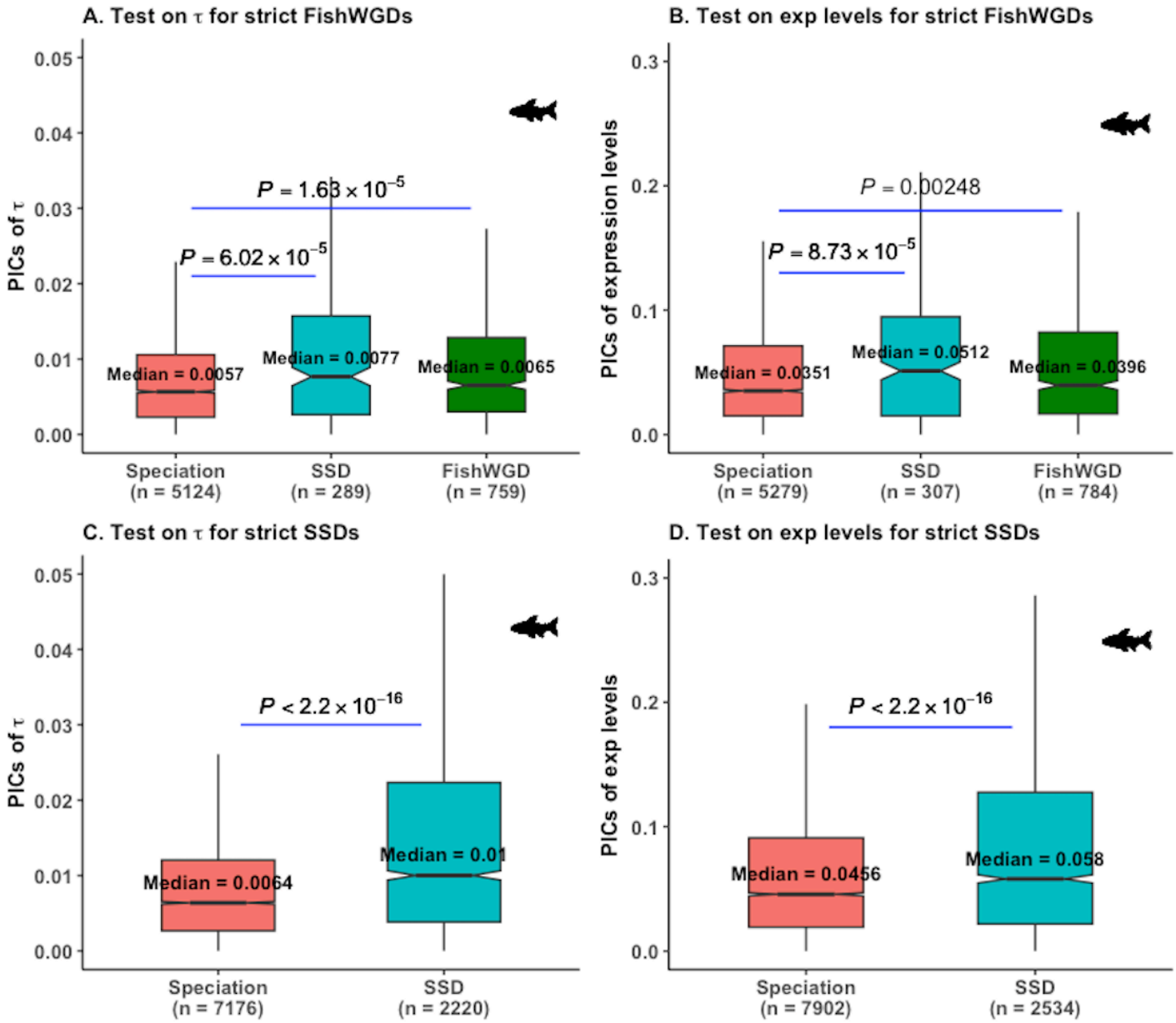

A. Test on τ for strict FishWGDs
PICs of τ
0.05
0.04
0.03
0.02
0.01
0.00
P = 1.63 × 10^−5
P = 6.02 × 10^−5
Median = 0.0057
Median = 0.0077
Median = 0.0065
Speciation (n = 5124)
SSD (n = 289)
FishWGD (n = 759)
B. Test on exp levels for strict FishWGDs
PICs of expression levels
0.3
0.2
0.1
0.0
P = 0.00248
P = 8.73 × 10^−5
Median = 0.0351
Median = 0.0512
Median = 0.0396
Speciation (n = 5279)
SSD (n = 307)
FishWGD (n = 784)
C. Test on τ for strict SSDs
PICs of τ
P < 2.2 × 10^−16
Median = 0.0064
Median = 0.01
Speciation (n = 7176)
SSD (n = 2220)
D. Test on exp levels for strict SSDs
PICs of exp levels
P < 2.2 × 10^−16
Median = 0.0456
Median = 0.058
Speciation (n = 7902)
SSD (n = 2534)

Fig. 2

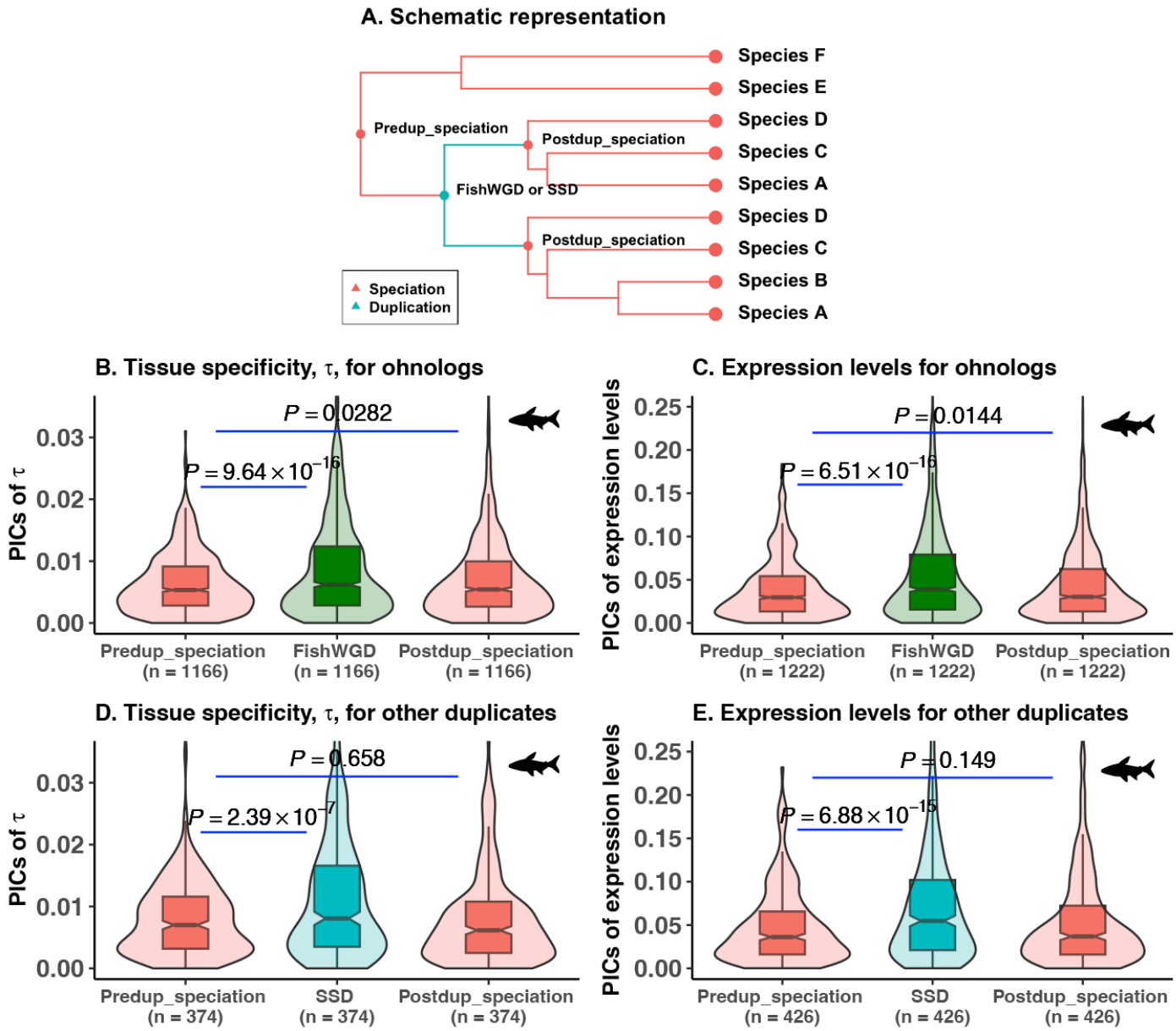

A. Schematic representation
Species F
Species E
Species D
Species C
Species A
Species D
Species C
Species B
Species A
Predup_speciation
Postdup_speciation
FishWGD or SSD
Postdup_speciation
Speciation
Duplication
B. Tissue specificity, τ, for ohnologs
PICs of τ
P = 0.0282
P = 9.64 × 10^−16
Predup_speciation (n = 1166)
FishWGD (n = 1166)
Postdup_speciation (n = 1166)
C. Expression levels for ohnologs
PICs of expression levels
P = 0.0144
P = 6.51 × 10^−16
Predup_speciation (n = 1222)
FishWGD (n = 1222)
Postdup_speciation (n = 1222)
D. Tissue specificity, τ, for other duplicates
PICs of τ
P = 0.658
P = 2.39 × 10^−7
Predup_speciation (n = 374)
SSD (n = 374)
Postdup_speciation (n = 374)
E. Expression levels for other duplicates
PICs of expression levels
P = 0.149
P = 6.88 × 10^−15
Predup_speciation (n = 426)
SSD (n = 426)
Postdup_speciation (n = 426)

Fig. 3

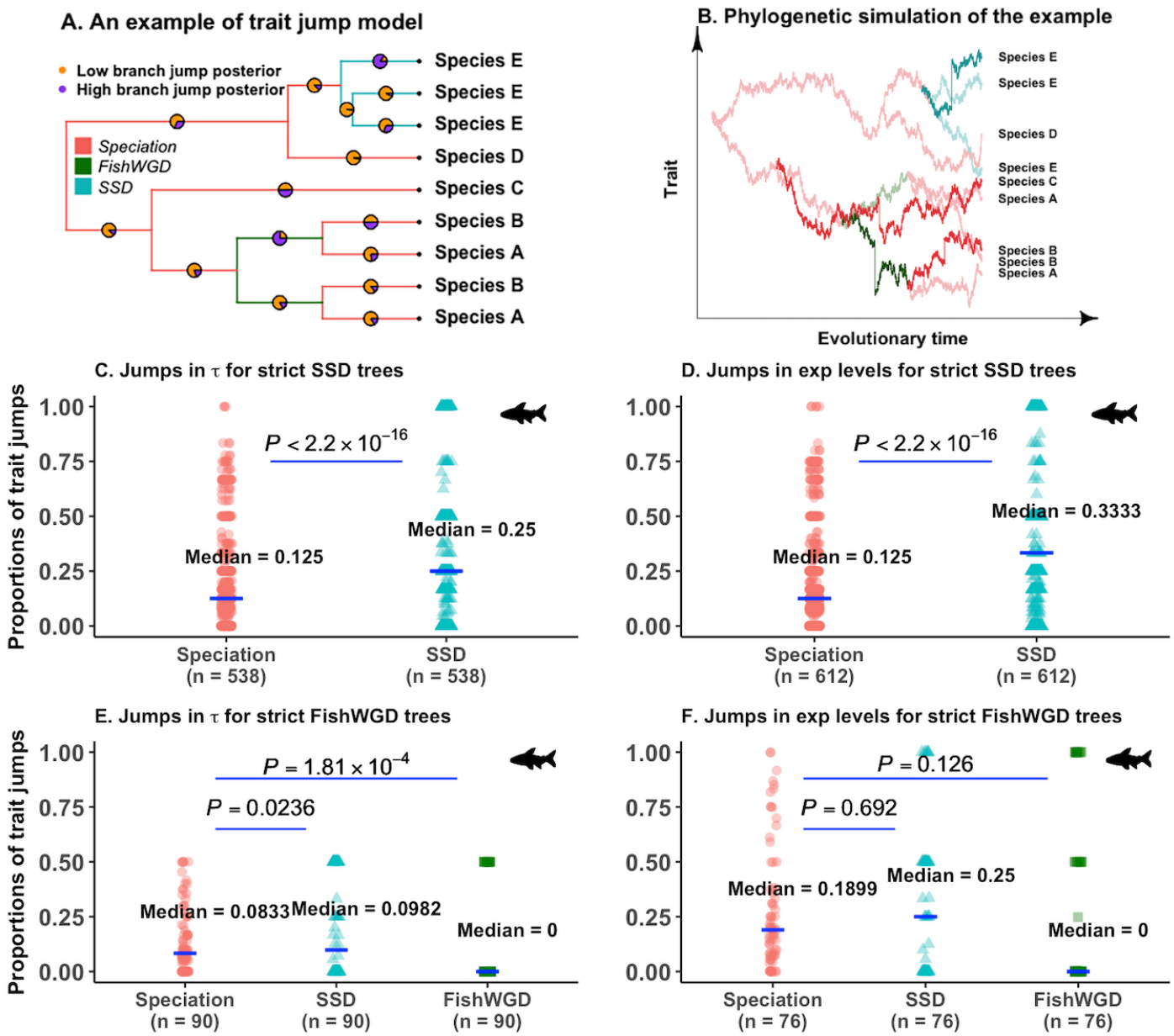

A. An example of trait jump model
Low branch jump posterior
High branch jump posterior
Speciation
FishWGD
SSD
Species E
Species E
Species E
Species D
Species C
Species B
Species A
Species B
Species A
B. Phylogenetic simulation of the example
Trait
Evolutionary time
Species E
Species E
Species D
Species E
Species C
Species A
Species B
Species B
Species A
C. Jumps in τ for strict SSD trees
Proportions of trait jumps
P < 2.2 × 10^−16
Median = 0.125
Median = 0.25
Speciation (n = 538)
SSD (n = 538)
D. Jumps in exp levels for strict SSD trees
P < 2.2 × 10^−16
Median = 0.125
Median = 0.3333
Speciation (n = 612)
SSD (n = 612)
E. Jumps in τ for strict FishWGD trees
P = 1.81 × 10^−4
P = 0.0236
Median = 0.0833
Median = 0.0982
Median = 0
Speciation (n = 90)
SSD (n = 90)
FishWGD (n = 90)
F. Jumps in exp levels for strict FishWGD trees
P = 0.126
P = 0.692
Median = 0.1899
Median = 0.25
Median = 0
Speciation (n = 76)
SSD (n = 76)
FishWGD (n = 76)

Fig. 4

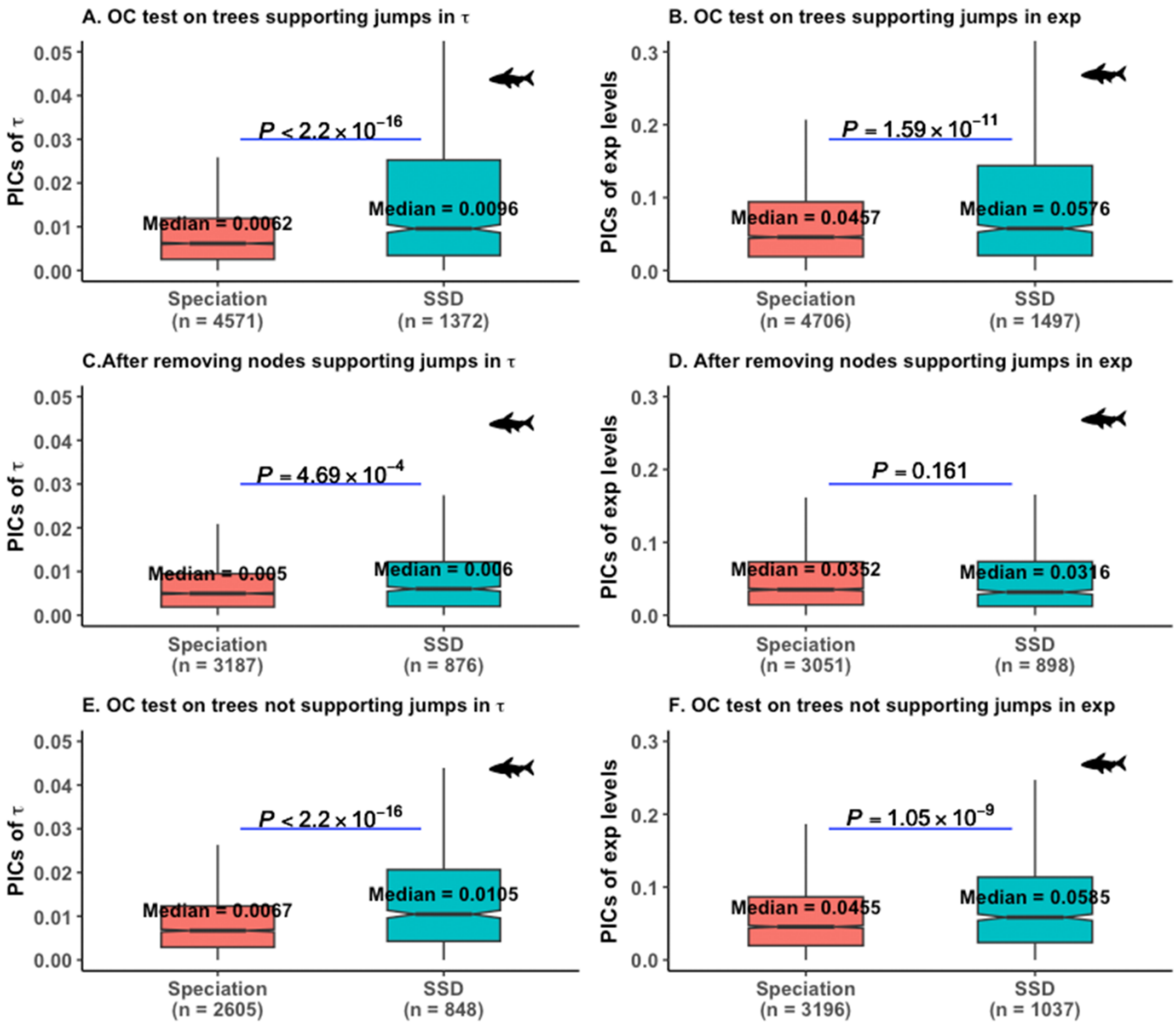

A. OC test on trees supporting jumps in τ
PICs of τ
P < 2.2 × 10^−16
Median = 0.0062
Median = 0.0096
Speciation (n = 4571)
SSD (n = 1372)
B. OC test on trees supporting jumps in exp
PICs of exp levels
P = 1.59 × 10^−11
Median = 0.0457
Median = 0.0576
Speciation (n = 4706)
SSD (n = 1497)
C.After removing nodes supporting jumps in τ
PICs of τ
P = 4.69 × 10^−4
Median = 0.005
Median = 0.006
Speciation (n = 3187)
SSD (n = 876)
D. After removing nodes supporting jumps in exp
PICs of exp levels
P = 0.161
Median = 0.0352
Median = 0.0316
Speciation (n = 3051)
SSD (n = 898)
E. OC test on trees not supporting jumps in τ
PICs of τ
P < 2.2 × 10^−16
Median = 0.0067
Median = 0.0105
Speciation (n = 2605)
SSD (n = 848)
F. OC test on trees not supporting jumps in exp
PICs of exp levels
P = 1.05 × 10^−9
Median = 0.0455
Median = 0.0585
Speciation (n = 3196)
SSD (n = 1037)

Fig. 5

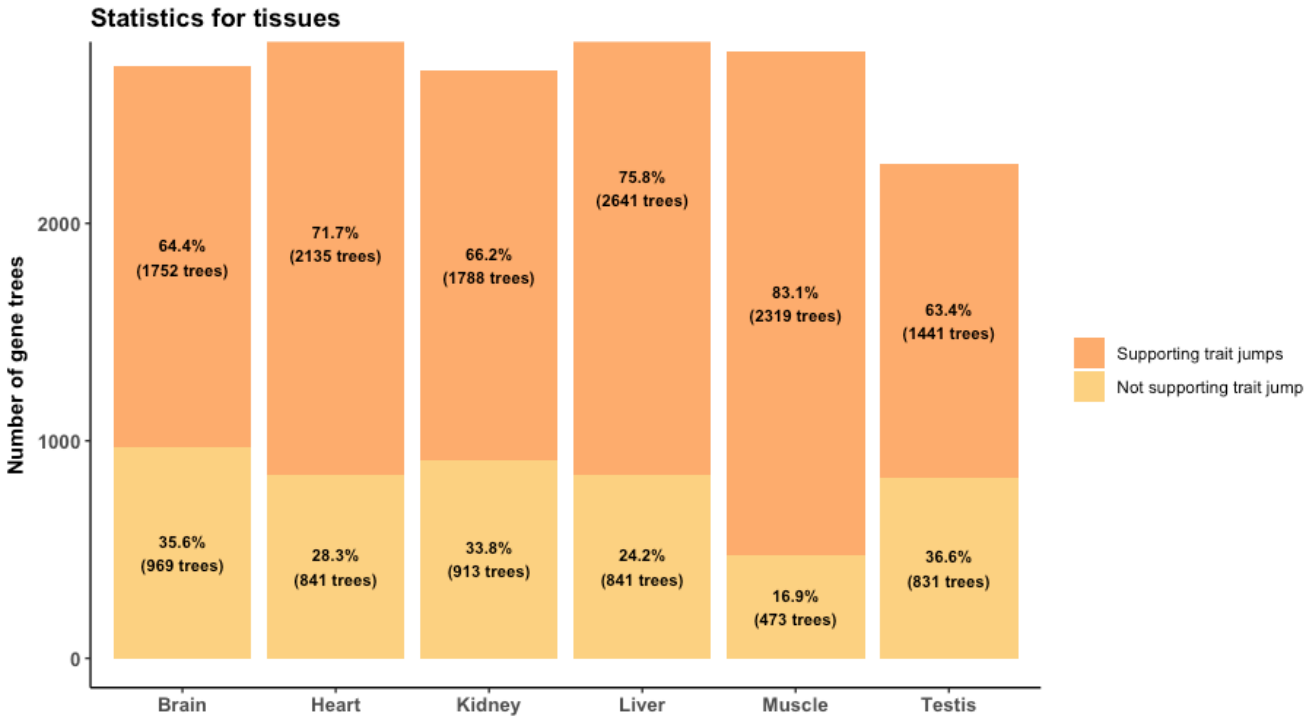

Statistics for tissues
Number of gene trees
2000
1000
0
64.4%
(1752 trees)
71.7%
(2135 trees)
66.2%
(1788 trees)
75.8%
(2641 trees)
83.1%
(2319 trees)
63.4%
(1441 trees)
35.6%
(969 trees)
28.3%
(841 trees)
33.8%
(913 trees)
24.2%
(841 trees)
16.9%
(473 trees)
36.6%
(831 trees)
Brain
Heart
Kidney
Liver
Muscle
Testis
Supporting trait jumps
Not supporting trait jump

Fig. 6

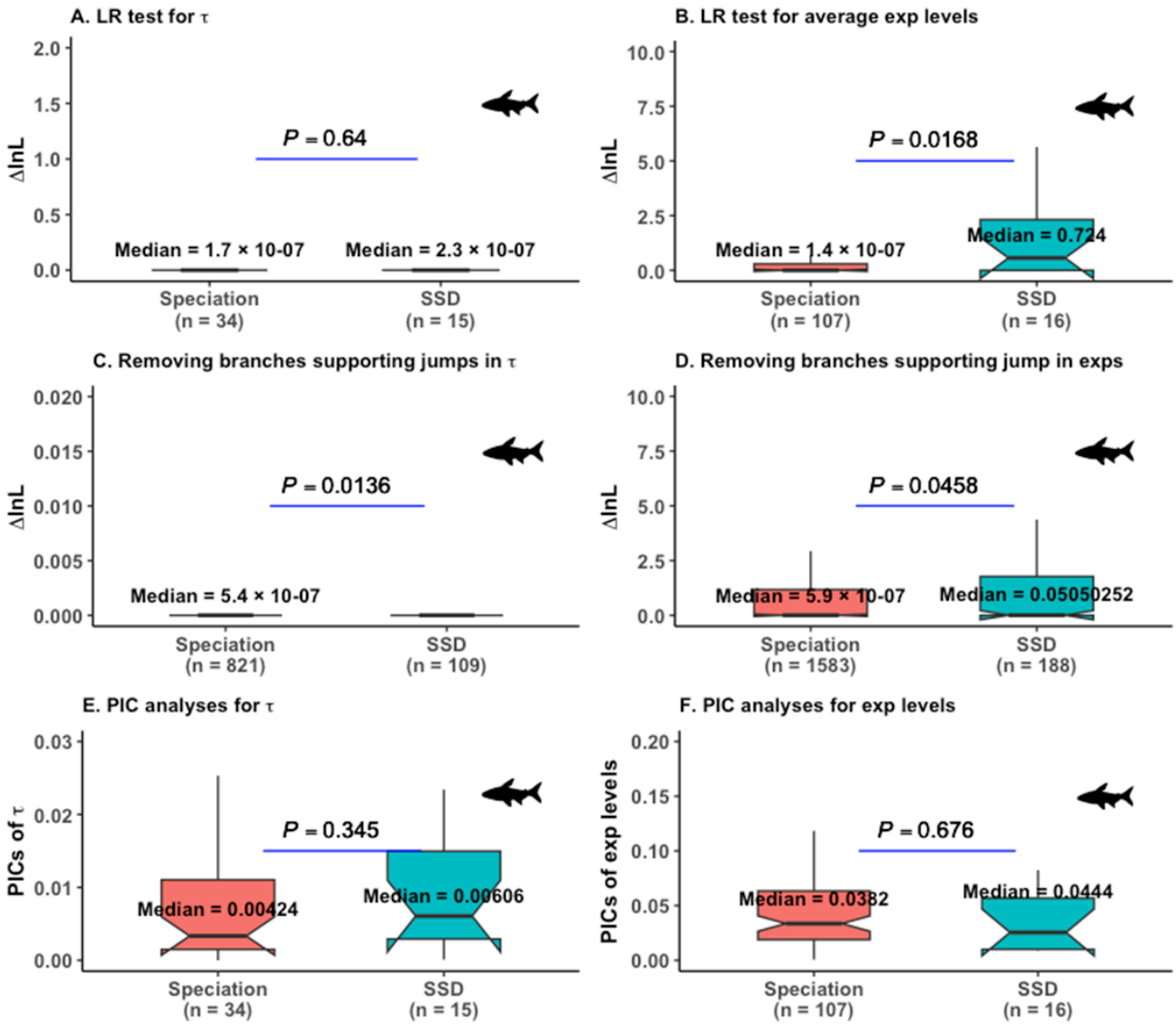

A. LR test for τ
ΔlnL
P = 0.64
Median = 1.7 × 10-07
Median = 2.3 × 10-07
Speciation (n = 34)
SSD (n = 15)
B. LR test for average exp levels
ΔlnL
P = 0.0168
Median = 1.4 × 10-07
Median = 0.724
Speciation (n = 107)
SSD (n = 16)
C. Removing branches supporting jumps in τ
ΔlnL
P = 0.0136
Median = 5.4 × 10-07
Speciation (n = 821)
SSD (n = 109)
D. Removing branches supporting jump in exps
ΔlnL
P = 0.0458
Median = 5.9 × 10-07
Median = 0.05050252
Speciation (n = 1583)
SSD (n = 188)
E. PIC analyses for τ
PICs of τ
P = 0.345
Median = 0.00424
Median = 0.00606
Speciation (n = 34)
SSD (n = 15)
F. PIC analyses for exp levels
PICs of exp levels
P = 0.676
Median = 0.0382
Median = 0.0444
Speciation (n = 107)
SSD (n = 16)

Fig. 7

Supporting Information

Figure S1: Principal component analysis (PCA) plots. (A) This plot uses normalized expression levels of 1077 1:1 orthologous gene set from 15 species. (B) We consider strict ohnologs or 3R whole genome duplicates (FishWGDs) from 5 teleosts in this plot. Normalized gene expression levels of 300 ohnolog set are used here. Duplicates in each teleost fish are classified based on their expression levels as "highly expressed" and "lowly expressed" ohnologs. Tissue colors for the highly expressed ohnolog copy and the lowly expressed ohnolog copy for a species are labeled as "High exp" and "Low exp". The plot shows that highly expressed ohnologs cluster separately from their lowly expressed copy for each tissue. Percentage of variations are indicated in each axis. PC1: Principal component 1, PC2: Principal component 2.

Figure S2: Phylogenetic independent contrast ortholog conjecture (OC) tests. PICs: Phylogenetic independent contrasts, exp: expression. Plots are based on 4247 and 4603 trees for τ and for average gene expression levels, respecively. For (A)-(B), we considered the PICs of teleosts specific speciation and duplication nodes for our analyses, while PICs of all the speciation and duplication nodes are used in (C)-(D). *P* values are from Wilcoxon two-tailed tests.

Figure S3: (A) Distributions of events for traits in teleosts and in vertebrates. (B) Phylogenetic independent contrast ortholog conjecture (OC) tests. Instead of using each node of a gene tree as a data point, we used the mean contrast per event for each tree to obtain one datapoint per gene family. Since we obtained one datapoint for speciation and one datapoint for duplication for each tree, paired Wilcoxon test was used to compare the difference. PICs: Phylogenetic independent

contrasts, exp: expression. For both (A) and (B), we only used gene trees with ≤5 duplication nodes.

Figure S4: The ortholog conjecture (OC) tests using calibrated contrasts standardized trees in 15 vertebrates. 746 and 774 trees with whole genome duplications were used in (A)-(B), while 2482 and 2717 trees with small-scale gene duplications were used in (C)-(D). PICs: Phylogenetic independent contrasts, FishWGD: Fish specific 3R whole genome duplication, 3RWGDs: ohnologs, SSDs: small-scale duplicates. Duplications in (A)-(B) refers to SSDs like (C)-(D). *P* values are from Wilcoxon two-tailed tests.

Figure S5: Modified phylogenetic approach to test the ortholog conjecture for teleosts and for vertebrates, respectively. We used the oldest duplication node along a path of a phylogeny for identification of pre-duplication and post-duplication speciations. For (A)-(D) plots are based on 1375 and 1493 trees for τ and for average gene expression levels, respecively. For (A) and (B), we consider the contrasts of teleosts specific duplication and descendant speciation nodes in absence of teleosts specific pre-duplication speciation nodes contrasts. Results of (C) and (D) are based on all 15 vertebrate species. PICs: Phylogenetic Independent Contrasts, exp: expression, Predup_speciation: Pre-duplication speciation, Postdup_speciation: Post-duplication speciation. Since contrasts are estimated for each tree, paired Wilcoxon test is used to compare the difference.

Figure S6: Randomization tests on our modified phylogenetic approach for teleosts and for vertebrates, respectively. Randomization of traits through tip shuffling allow us to use the same nodes used for the ortholog conjecture tests in Fig. S3. For (A)-(D) plots are based on randomization tests on 1375 and 1493 trees for τ and for average gene expression levels, respecively. We considered the contrasts of teleosts specific gene or genome duplication and descendant speciation nodes in (A)-(D) in absence of teleost-specific pre-duplication speciation

nodes contrasts. PICs: Phylogenetic Independent Contrasts, exp: expression, Predup_speciation: Pre-duplication speciation, Postdup_speciation: Post-duplication speciation. Since contrasts are estimated for each tree, paired Wilcoxon test is used to compare the difference.

Figure S7: Randomization test on our modified phylogenetic approach for strict 3RWGD and for strict SSD trees in teleosts, respectively. Randomization of traits through tip shuffling allow us to use the same nodes used for the ortholog conjecture tests in Fig. 4. (A)-(D): Plots are based on randomization tests for 581 and 610 strict 3R whole genome duplicate trees and for 339 and 377 strict small-scale duplicate trees for τ and average gene expression levels, respecively. We considered the contrasts of teleosts specific gene or genome duplication and descendant speciation nodes in (A)-(D) in absence of teleost-specific pre-duplication speciation nodes contrasts. PICs: Phylogenetic Independent Contrasts, exp: expression, Predup_speciation: Pre-duplication speciation, Postdup_speciation: Post-duplication speciation, FishWGD: fish specific 3R whole genome duplication, 3RWGDs: ohnologs, SSDs: small-scale duplicates. Since contrasts are estimated for each tree, paired Wilcoxon test is used to compare the difference.

Figure S8: Modified phylogenetic approach to test the ortholog conjecture for teleosts and for vertebrates, respectively. We used the youngest duplication node along a path of a phylogeny for identification of pre-duplication and post-duplication speciations. For (A) and (B), we considered the contrasts of teleosts specific duplication and post-duplication speciation nodes in absence of teleost-specific pre-duplication speciation nodes contrasts. (C) and (D) are based on all 15 vertebrate species. PICs: Phylogenetic Independent Contrasts, exp: expression, Predup_speciation: Pre-duplication speciation, Postdup_speciation: Post-duplication speciation. Since contrasts are estimated for each tree, paired Wilcoxon test is used to compare the difference.

67 **Figure S9**: The ortholog conjecture tests on phylogenies with genome and gene duplicates in
68 teleosts. 581, 610, 339 and 377 of trees passed our criteria of Fig. 4A in (A)-(D), respectively and
69 were used for the analyses. We here used only teleost-specific duplication and post-duplication
70 speciation nodes for contrasts analyses in absence of teleost-specific pre-duplication speciation
71 nodes contrasts. We considered the youngest duplication node along a path of a phylogeny for
72 identification of pre-duplication and post-duplication speciations. PICs: Phylogenetic independent
73 contrasts, exp: expression, Predup_speciation: pre-duplication speciation, Postdup_speciation:
74 post-duplication speciation, FishWGD: fish specific 3R whole genome duplications, 3RWGDs:
75 ohnologs, SSDs: small-scale duplicates. Since contrasts are estimated for each tree, paired
76 Wilcoxon test is used to compare the difference.

77 **Figure S10**: Analyses of trait jump models for τ and for average gene expression levels in 5
78 teleosts and in 15 vertebrates. We used a strict posterior probability cutoff of ≥ 0.7 as an evidence
79 of trait jump. (A)-(B) Statistics of calibrated Brownian trees supporting jumps for τ and for average
80 gene expression levels, respectively. Proportions of jumps in traits following speciations and
81 duplication events are shown in (C)-(D) for teleosts and in (E)-(F) for vertebrates. 2636 trees
82 supporting rapid jumps in τ and 2604 trees supporting jumps in average gene expression levels
83 with the above-mentioned threshold were used for analyses. Since proportion of shifts in traits per
84 branch of events is estimated for each tree, paired Wilcoxon test is used to compare the difference.
85 'exp': Average gene expression levels.

86 **Figure S11**: The ortholog conjecture (OC) tests on trees supporting rapid jump in τ and in average
87 gene expression levels for 15 vertebrates. We used a strict branch jump posterior cutoff of ≥ 0.7
88 as an evidence of trait jump. Only teleosts-specific clades were used for the analyses. The ortholog
89 conjecture tests on (A) 2636 calibrated contrast standardized trees supporting rapid jumps in τ,

and (B) 2604 trees supporting jumps in average gene expression levels with the above-mentioned threshold along a phylogeny for our analyses. (C)-(D) The ortholog conjecture tests after removing contrasts of nodes whose daughter branch(es) experienced jumps in the corresponding trait for 2636 and 2604 trees, respectively. (E)-(F) The ortholog conjecture tests on 1611 and 1999 trees for which the trait jump model is not supported for τ and for average gene expression levels, respectively. *P* values are from Wilcoxon two-tailed tests.

Figure S12: The ortholog conjecture (OC) tests on trees supporting rapid jump in τ and in average gene expression levels for 5 teleosts. We used a strict branch jump posterior cutoff of ≥ 0.7 as an evidence of trait jump. Only teleosts-specific clades were used for the analyses. (A) 2636 calibrated contrasts standardized trees supporting rapid jumps in τ, and (B) 2604 trees supporting jumps in average gene expression levels were used for the analyses. (C)-(D) The ortholog conjecture tests after removing contrasts of nodes whose daughter branch(es) experienced jumps in the corresponding trait for 2636 and 2604 trees, respectively. (E)-(F) The ortholog conjecture tests on 1611 and 1999 trees for which the trait jump model is not supported for τ and for average gene expression levels, respectively. *P* values are from Wilcoxon two-tailed tests.

Figure S13: Distribution of the phylogenetic independent contrasts (PICs) for all the trees supporting rapid jump in τ and in average gene expression levels for teleosts and for vertebrates. For these analyses, we considered PICs of the nodes for a tree having a branch jump posterior probability > 0. Numbers of trees supporting jumps increased from 2636 to 2675 for τ and from 2604 to 2647 for average expression levels by altering the PP threshold from 0.7 to 0. *P* values are from Wilcoxon two-tailed tests.

Figure S14: The ortholog conjecture (OC) tests on strict 3RWGD trees to analyze the jump results in τ and in average gene expression levels for teleosts. We used a strict branch jump posterior

113 cutoff of ≥ 0.7 as an evidence of trait jump. Only teleosts-specific clades were used for the
114 analyses. The ortholog conjecture tests on (A) 435 calibrated contrasts standardized strict 3RWGD
115 trees supporting rapid jumps in τ, and on (B) 398 strict 3RWGD trees supporting jumps in average
116 gene expression levels. (C)-(D) The ortholog conjecture tests after removing contrasts of nodes
117 whose daughter branch(es) experienced jumps in the corresponding trait for 435 and 398 strict
118 3RWGD trees, respectively. (E)-(F) The ortholog conjecture tests on 311 and 376 strict 3RWGD
119 trees for which the trait jump model is not supported for τ and for average gene expression levels,
120 respectively. FishWGD: ohnologs, other duplicates refer to small scale duplicates. *P* values are
121 from Wilcoxon two-tailed tests.

122 **Figure S15**: Analyses of rapid jumps in brain expression for strict SSD trees. We used a strict
123 posterior probability cutoff of ≥ 0.7 as an evidence of trait jump. (A)-(B) Proportions of jumps in
124 brain expressions following speciation and duplication events for teleosts and for vertebrates using
125 1752 trait jump supporting trees of Fig. 6A. Since proportion of shifts in traits per branch of events
126 is estimated for each tree, paired Wilcoxon test is used to compare the difference for (A)-(B). (C)-
127 (D) The ortholog conjecture (OC) tests on 1752 strict SSD trees to analyze the jump results in
128 brain expressions for teleosts and for 15 vertebrates, respectively. (E)-(F) The ortholog conjecture
129 tests after removing contrasts of nodes whose daughter branch(es) experienced jumps in the
130 corresponding trait for 1752 strict SSD trees for 5 teleosts and for 15 vertebrates, respectively. *P*
131 values are from Wilcoxon two-tailed tests for (C)-(F). 'exp': expression levels.

132 **Figure S16**: Analyses of rapid jumps in heart expression for strict SSD trees. We used a strict
133 posterior probability cutoff of ≥ 0.7 as an evidence of trait jump. (A)-(B) Proportions of jumps in
134 heart expressions following speciation and duplication events for teleosts and for vertebrates using
135 2135 trait jump supporting trees of Fig. 6B. Since proportion of shifts in traits per branch of events

is estimated for each tree, paired Wilcoxon test is used to compare the difference for (A)-(B). (C)-(D) The ortholog conjecture (OC) tests on 2135 strict SSD trees to analyze the jump results in heart expressions for teleosts and for 15 vertebrates, respectively. (E)-(F) The ortholog conjecture tests after removing contrasts of nodes whose daughter branch(es) experienced jumps in the corresponding trait for 2135 strict SSD trees for 5 teleosts and for 15 vertebrates, respectively. *P* values are from Wilcoxon two-tailed tests for (C)-(F). 'exp': expression levels.

Figure S17: Analyses of rapid jumps in kidney expression for strict SSD trees. We used a strict posterior probability cutoff of ≥ 0.7 as an evidence of trait jump. (A)-(B) Proportions of jumps in kidney expressions following speciation and duplication events for teleosts and for vertebrates using 1788 trait jump supporting trees of Fig. 6C. Since proportion of shifts in traits per branch of events is estimated for each tree, paired Wilcoxon test is used to compare the difference for (A)-(B). (C)-(D) The ortholog conjecture (OC) tests on 1788 strict SSD trees to analyze the jump results in kidney expressions for teleosts and for 15 vertebrates, respectively. (E)-(F) The ortholog conjecture tests after removing contrasts of nodes whose daughter branch(es) experienced jumps in the corresponding trait for 1788 strict SSD trees for 5 teleosts and for 15 vertebrates, respectively. *P* values are from Wilcoxon two-tailed tests for (C)-(F). 'exp': expression levels.

Figure S18: Analyses of rapid jumps in muscle expression for strict SSD trees. We used a strict posterior probability cutoff of ≥ 0.7 as an evidence of trait jump. (A)-(B) Proportions of jumps in muscle expressions following speciation and duplication events for teleosts and for vertebrates using 2319 trait jump supporting trees of Fig. 6D. Since proportion of shifts in traits per branch of events is estimated for each tree, paired Wilcoxon test is used to compare the difference for (A)-(B). (C)-(D) The ortholog conjecture (OC) tests on 2319 strict SSD trees to analyze the jump results in muscle expressions for teleosts and for 15 vertebrates, respectively. (E)-(F) The ortholog

159 conjecture tests after removing contrasts of nodes whose daughter branch(es) experienced jumps
160 in the corresponding trait for 2319 strict SSD trees for 5 teleosts and for 15 vertebrates,
161 respectively. *P* values are from Wilcoxon two-tailed tests for (C)-(F). 'exp': expression levels.

162 **Figure S19**: Analyses of rapid jumps in liver expression for strict SSD trees. We used a strict
163 posterior probability cutoff of ≥ 0.7 as an evidence of trait jump. (A)-(B) Proportions of jumps in
164 liver expressions following speciation and duplication events for teleosts and for vertebrates using
165 2641 trait jump supporting trees of Fig. 6E. Since proportion of shifts in traits per branch of events
166 is estimated for each tree, paired Wilcoxon test is used to compare the difference for (A)-(B). (C)-
167 (D) The ortholog conjecture (OC) tests on 2641 strict SSD trees to analyze the jump results in liver
168 expressions for teleosts and for 15 vertebrates, respectively. (E)-(F) The ortholog conjecture tests
169 after removing contrasts of nodes whose daughter branch(es) experienced jumps in the
170 corresponding trait for 2641 strict SSD trees for 5 teleosts and for 15 vertebrates, respectively. *P*
171 values are from Wilcoxon two-tailed tests for (C)-(F). 'exp': expression levels.

172 **Figure S20**: Analyses of rapid jumps in testis expression for strict SSD trees. We used a strict
173 posterior probability cutoff of ≥ 0.7 as an evidence of trait jump. (A)-(B) Proportions of jumps in
174 testis expressions following speciation and duplication events for teleosts and for vertebrates using
175 1441 trait jump supporting trees of Fig. 6F. Since proportion of shifts in traits per branch of events
176 is estimated for each tree, paired Wilcoxon test is used to compare the difference for (A)-(B). (C)-
177 (D) The ortholog conjecture (OC) tests on 1441 strict SSD trees to analyse the jump results in testis
178 expressions for teleosts and for 15 vertebrates, respectively. (E)-(F) The ortholog conjecture tests
179 after removing contrasts of nodes whose daughter branch(es) experienced jumps in the
180 corresponding trait for 1441 strict SSD trees for 5 teleosts and for 15 vertebrates, respectively. *P*
181 values are from Wilcoxon two-tailed tests for (C)-(F). 'exp': expression levels.

182 **Figure S21**: Positive selection and trait divergence analyses for contrast standardized strict SSD
183 trees in all 15 vertebrates. We only used phylogenies supporting an evolutionary trait jump model
184 for these analyses. We used a strict posterior probability cutoff of ≥ 0.7 to identify branches with
185 rapid jumps in traits. exp: Gene expression, PICs: phylogenetic independent contrasts, LRT:
186 likelihood ratio test.

187

188 **Table S1: Statistics on 1575 and 1619 strict SSD trees by using posterior probability (PP)**

189 **threshold of 0.5 for τ and for average expression levels, respectively.**

Events		Number of branches with evolutionary trait jumps (PP ≥ 0.5)	Number of branches with no inferred jump (PP < 0.5)	Proportions of trait jumps	Two-sample test for equality of proportions, χ^2 test with Yates correction
For τ in 5 teleost-specific clades	**Speciation**	865	3807	0.185	χ^2 = 74.21 on *df* = 1, $P < 1 \times 10^{-4}$
	SSD	581	1505	0.278	
For average expression levels in 5 teleost-specific clades	**Speciation**	1307	4177	0.238	χ^2 = 113.24 on *df* = 1, $P < 1 \times 10^{-4}$
	SSD	900	1654	0.352	
For τ in all 15 vertebrate clades	**Speciation**	8194	40152	0.169	χ^2 = 630.51on *df* = 1, $P < 1 \times 10^{-4}$
	SSD	2149	5219	0.292	
For average expression levels in all 15 vertebrate clades	**Speciation**	5245	22841	0.187	χ^2 = 597.56 on *df* = 1, $P < 1 \times 10^{-4}$
	SSD	1562	2947	0.346	

190 *Note*: Numbers of strict SSD trees with jumps increased from 1557 to 1575 and from 1593 to 1619

191 trees by altering the PP threshold from 0.7 to 0.5 for τ and for average expression levels,

192 respectively.

193

194 **Table S2: Statistics on 443 and 403 strict WGD trees by using posterior probability (PP)**

195 **threshold of 0.5 for τ and for average expression levels, respectively.**

Event		**Number of branches with evolutionary trait jumps (PP ≥ 0.5)**	**Number of branches with no inferred jump (PP < 0.5)**	**Proportions of trait jumps**	**Two-sample test for equality of proportions, χ^2 test with Yates correction**
For τ in 5 teleost-specific clades	Speciation *vs.* WGD	216 *vs.* 20	1162 *vs.* 170	0.157 *vs.* 0.105	χ^2 = 3.07 on *df* = 1, *P* = 0.08
	Speciation *vs.* SSD	216 *vs.* 62	1162 *vs.* 238	0.157 *vs.* 0.207	χ^2 = 4.09 on *df* = 1, *P* = 0.043
For average expression levels in 5 teleost-specific clades	Speciation *vs.* WGD	453 *vs.* 49	801 *vs.* 123	0.361 *vs.* 0.285	χ^2 = 3.54 on *df* = 1, *P* = 0.06
	Speciation *vs.* SSD	453 *vs.* 95	801 *vs.* 171	0.361 *vs.* 0.357	χ^2 = 0.003 on *df* = 1, *P* = 0.95
	Speciation *vs.* WGD	1067 *vs.* 58	8039 *vs.* 410	0.117 *vs.* 0.124	χ^2 = 0.14 on *df* = 1, *P* = 0.71

For τ in all 15 vertebrate clades	Speciation *vs.* SSD	1067 *vs.* 213	8039 *vs.* 709	0.117 *vs.* 0.231	χ^2 = 96.43 on *df* = 1, $P < 1 \times 10^{-4}$
For average expression levels in all 15 vertebrate clades	Speciation *vs.* WGD	1513 *vs.* 90	6399 *vs.* 328	0.191 *vs.* 0.215	χ^2 = 1.33 on *df* = 1, $P = 0.25$
	Speciation *vs.* SSD	1513 *vs.* 299	6399 *vs.* 595	0.191 *vs.* 0.334	χ^2 = 99.94 on *df* = 1, $P < 1 \times 10^{-4}$

Note: Numbers of strict WGD trees increased from 435 to 443 and from 398 to 403 trees by altering the PP threshold from 0.7 to 0.5 for τ and for average expression levels, respectively.

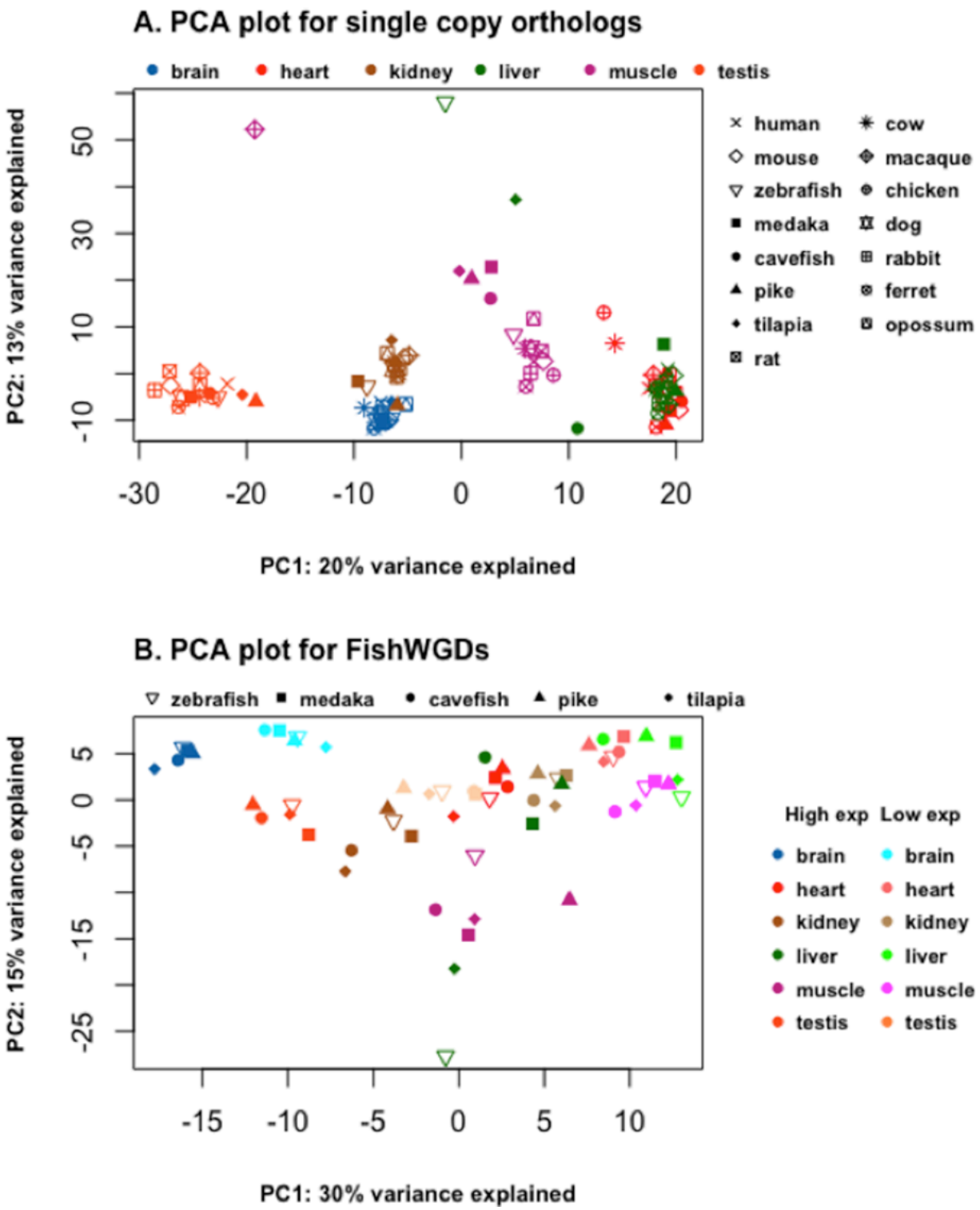

A. PCA plot for single copy orthologs
brain
heart
kidney
liver
muscle
testis
human
mouse
zebrafish
medaka
cavefish
pike
tilapia
rat
cow
macaque
chicken
dog
rabbit
ferret
opossum
PC2: 13% variance explained
PC1: 20% variance explained
B. PCA plot for FishWGDs
zebrafish
medaka
cavefish
pike
tilapia
High exp
Low exp
brain
heart
kidney
liver
muscle
testis
PC2: 15% variance explained
PC1: 30% variance explained

Fig. S1

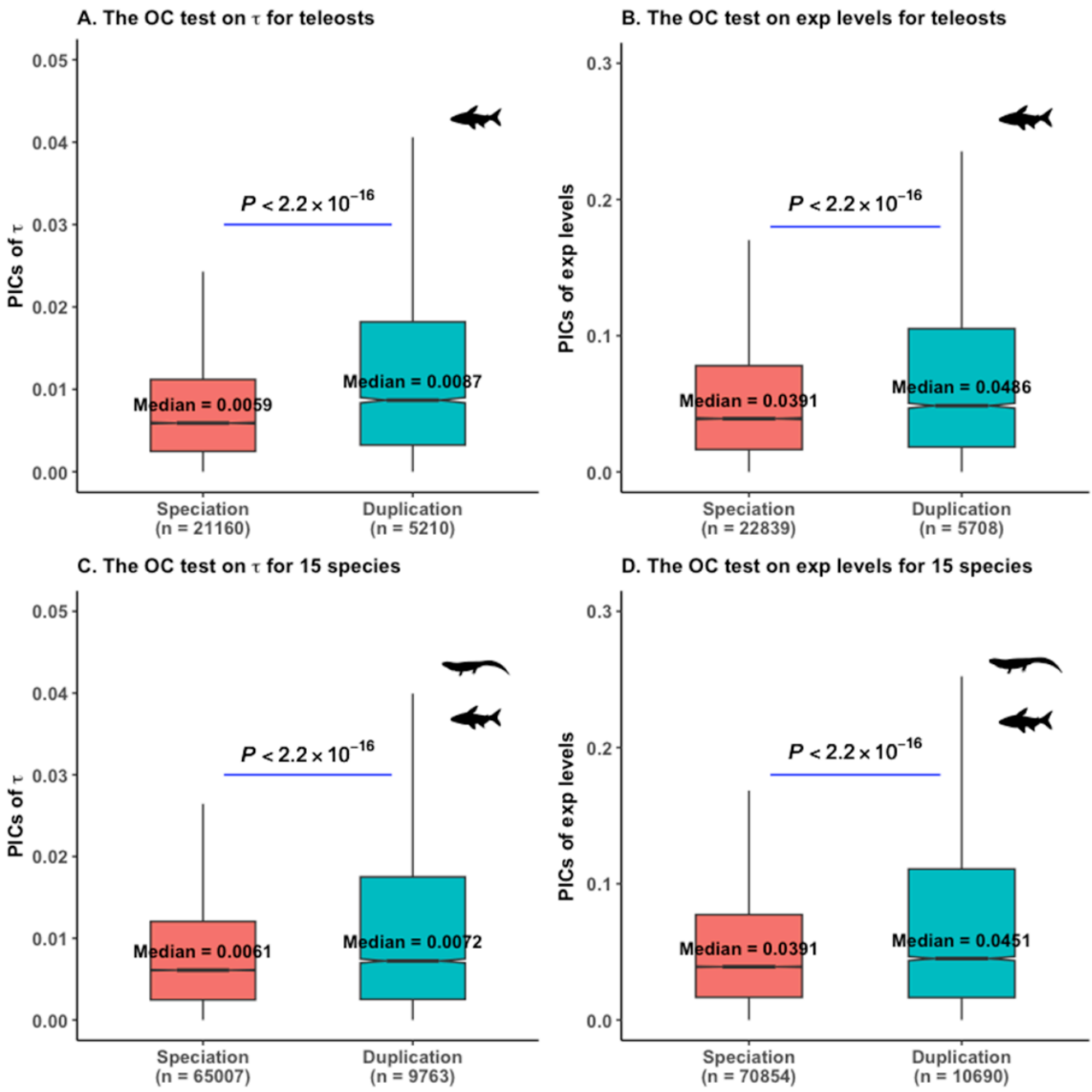

A. The OC test on τ for teleosts
PICs of τ
0.05
0.04
0.03
0.02
0.01
0.00
P < 2.2 × 10^−16
Median = 0.0059
Median = 0.0087
Speciation
(n = 21160)
Duplication
(n = 5210)
B. The OC test on exp levels for teleosts
PICs of exp levels
0.3
0.2
0.1
0.0
P < 2.2 × 10^−16
Median = 0.0391
Median = 0.0486
Speciation
(n = 22839)
Duplication
(n = 5708)
C. The OC test on τ for 15 species
PICs of τ
0.05
0.04
0.03
0.02
0.01
0.00
P < 2.2 × 10^−16
Median = 0.0061
Median = 0.0072
Speciation
(n = 65007)
Duplication
(n = 9763)
D. The OC test on exp levels for 15 species
PICs of exp levels
0.3
0.2
0.1
0.0
P < 2.2 × 10^−16
Median = 0.0391
Median = 0.0451
Speciation
(n = 70854)
Duplication
(n = 10690)

Fig. S2

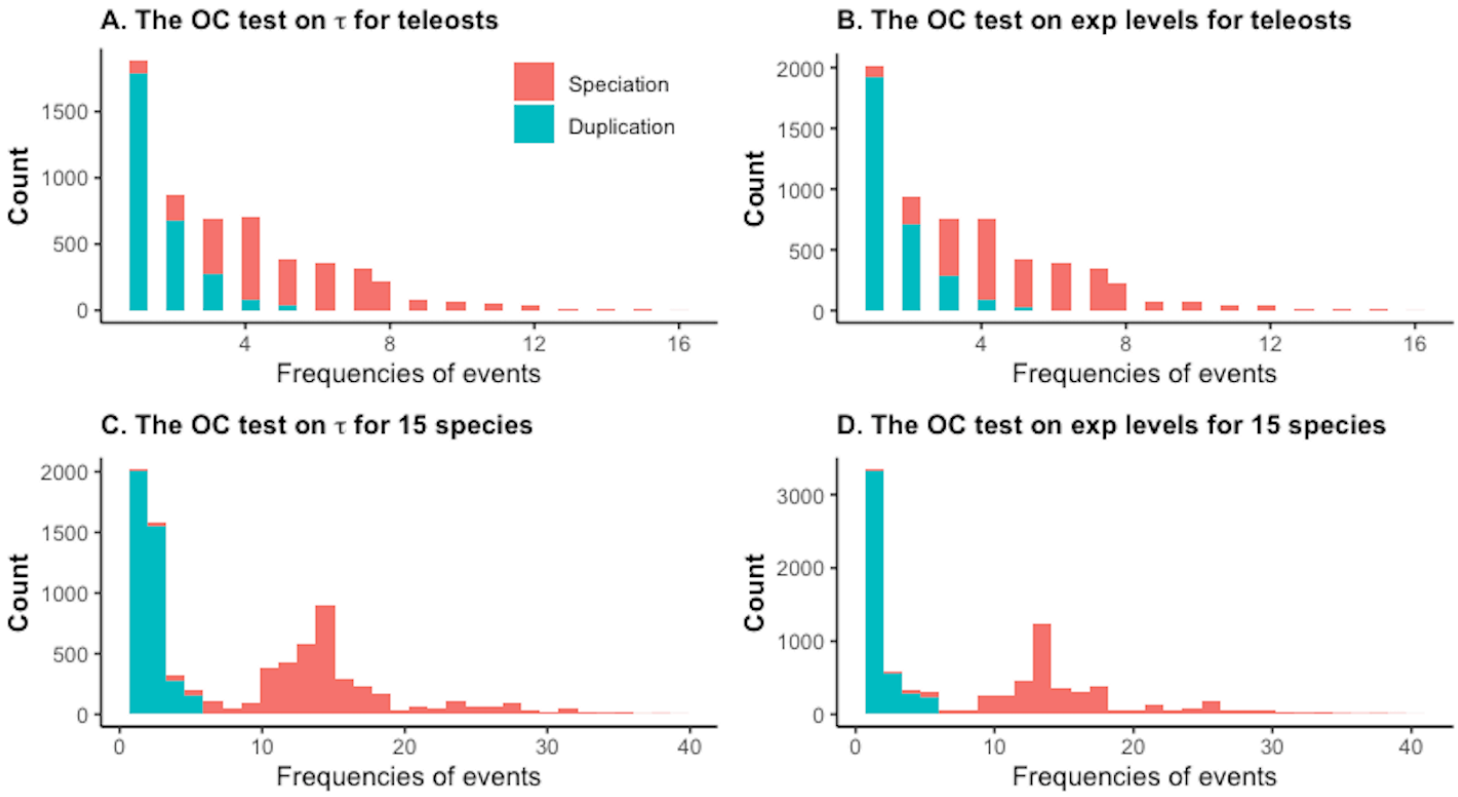

A. The OC test on τ for teleosts
Speciation
Duplication
Count
Frequencies of events
B. The OC test on exp levels for teleosts
Count
Frequencies of events
C. The OC test on τ for 15 species
Count
Frequencies of events
D. The OC test on exp levels for 15 species
Count
Frequencies of events

Fig. S3A

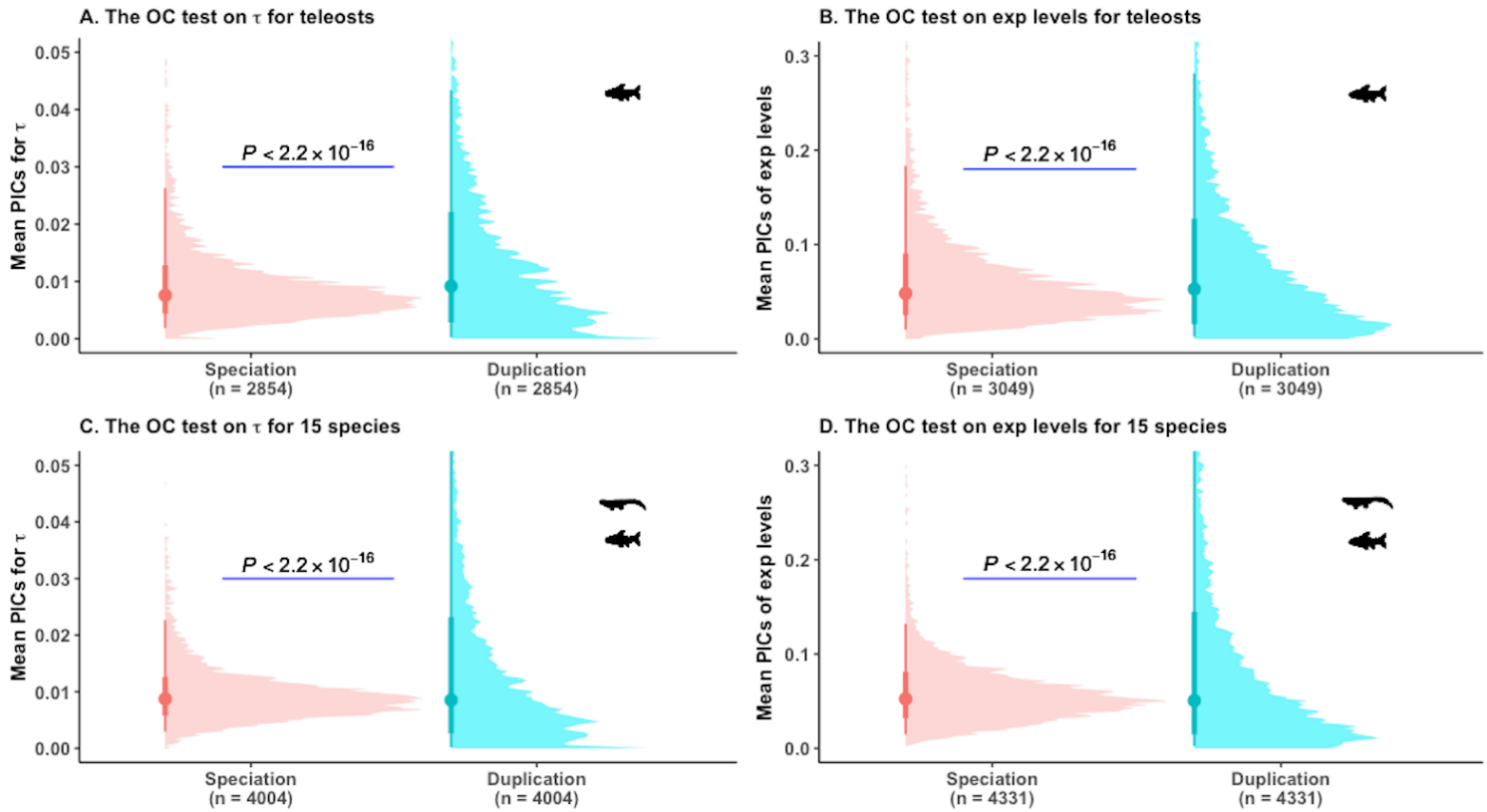

A. The OC test on τ for teleosts
Mean PICs for τ
0.05
0.04
0.03
0.02
0.01
0.00
P < 2.2 × 10^-16
Speciation
(n = 2854)
Duplication
(n = 2854)
B. The OC test on exp levels for teleosts
Mean PICs of exp levels
0.3
0.2
0.1
0.0
P < 2.2 × 10^-16
Speciation
(n = 3049)
Duplication
(n = 3049)
C. The OC test on τ for 15 species
Mean PICs for τ
0.05
0.04
0.03
0.02
0.01
0.00
P < 2.2 × 10^-16
Speciation
(n = 4004)
Duplication
(n = 4004)
D. The OC test on exp levels for 15 species
Mean PICs of exp levels
0.3
0.2
0.1
0.0
P < 2.2 × 10^-16
Speciation
(n = 4331)
Duplication
(n = 4331)

Fig. S3B

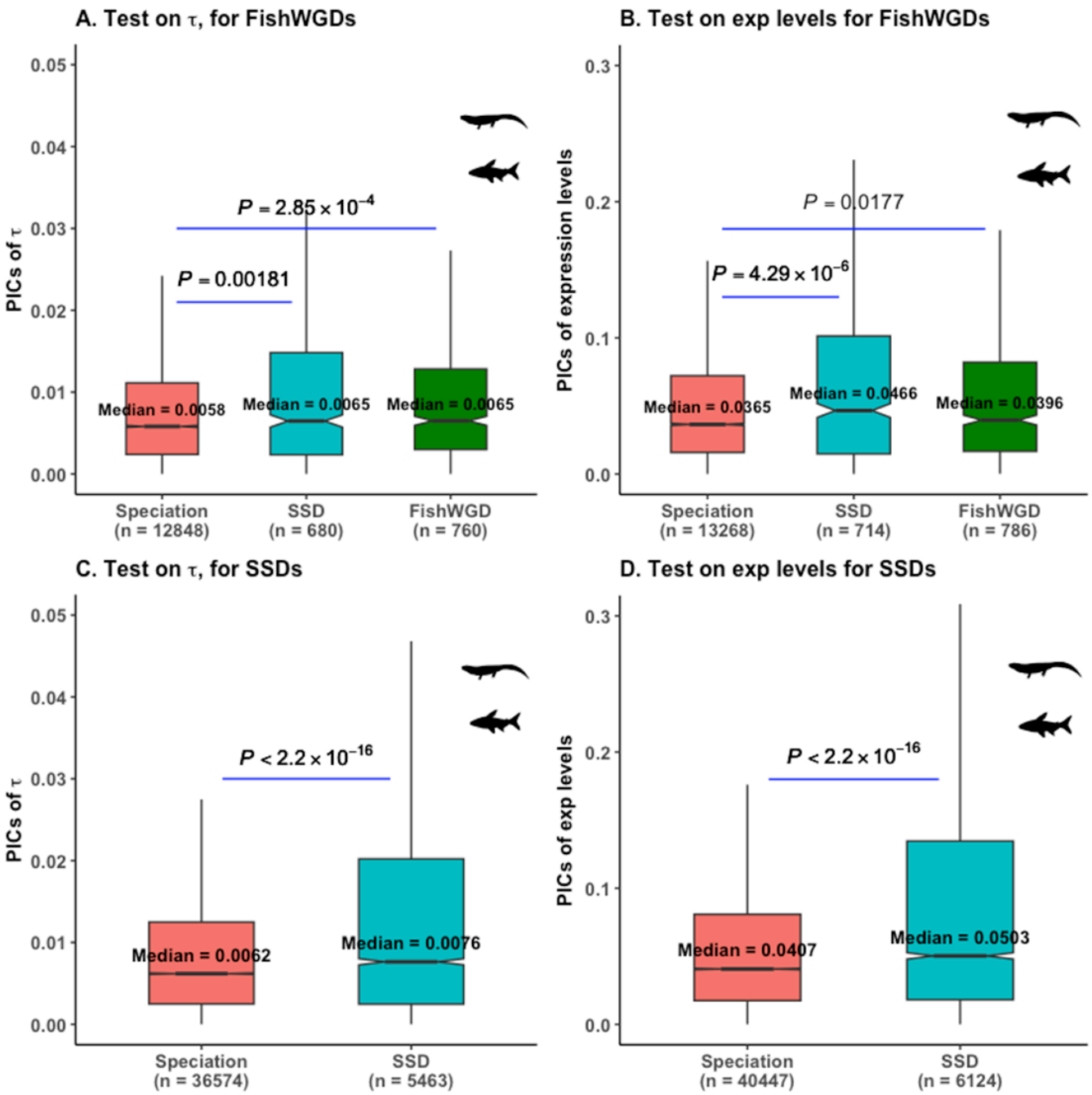

A. Test on τ, for FishWGDs
PICs of τ
0.05
0.04
0.03
0.02
0.01
0.00
P = 2.85 × 10^-4
P = 0.00181
Median = 0.0058
Median = 0.0065
Median = 0.0065
Speciation (n = 12848)
SSD (n = 680)
FishWGD (n = 760)
B. Test on exp levels for FishWGDs
PICs of expression levels
0.3
0.2
0.1
0.0
P = 0.0177
P = 4.29 × 10^-6
Median = 0.0365
Median = 0.0466
Median = 0.0396
Speciation (n = 13268)
SSD (n = 714)
FishWGD (n = 786)
C. Test on τ, for SSDs
PICs of τ
0.05
0.04
0.03
0.02
0.01
0.00
P < 2.2 × 10^-16
Median = 0.0062
Median = 0.0076
Speciation (n = 36574)
SSD (n = 5463)
D. Test on exp levels for SSDs
PICs of exp levels
0.3
0.2
0.1
0.0
P < 2.2 × 10^-16
Median = 0.0407
Median = 0.0503
Speciation (n = 40447)
SSD (n = 6124)

Fig. S4

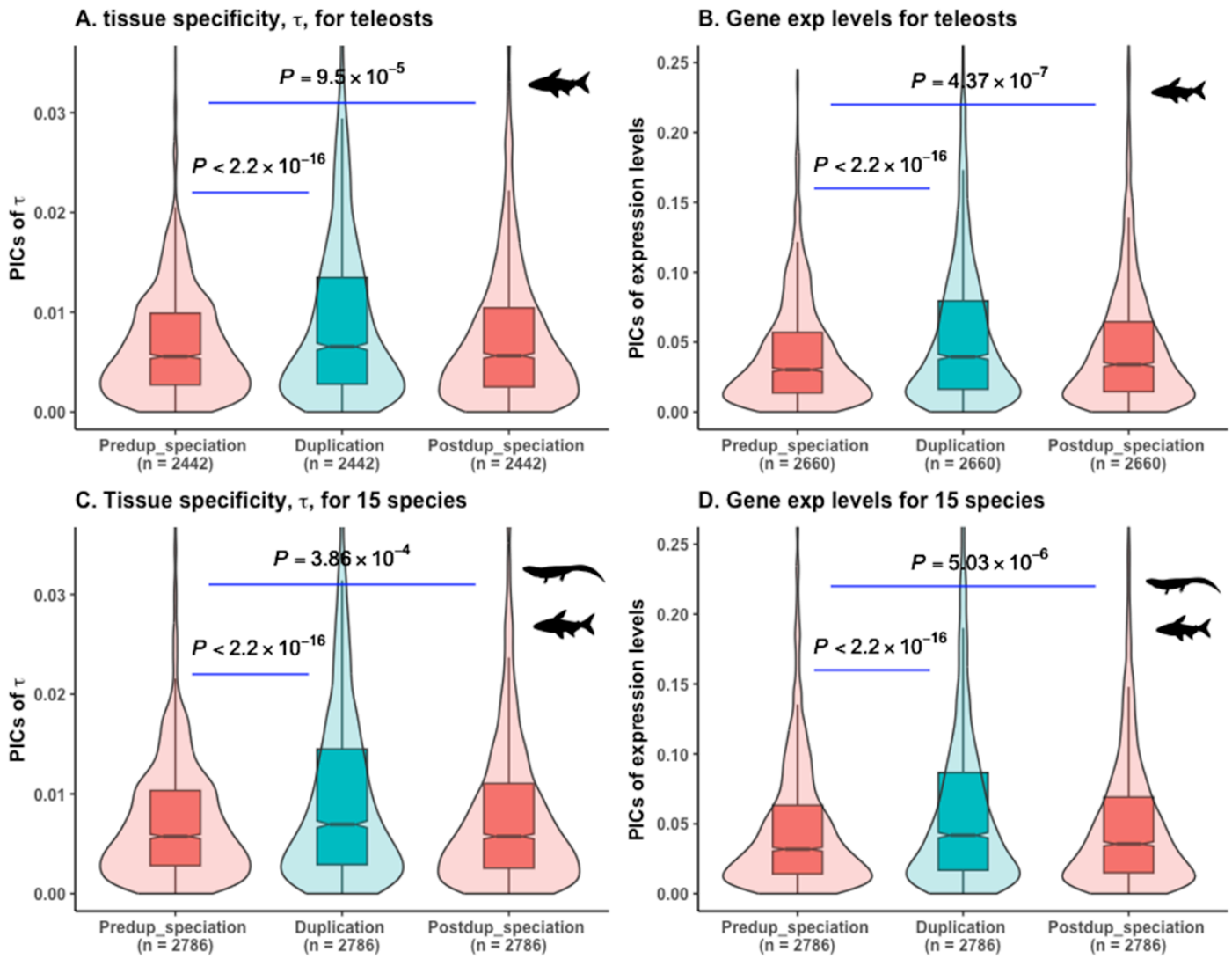

A. tissue specificity, τ, for teleosts
P = 9.5 × 10^−5
P < 2.2 × 10^−16
PICs of τ
0.00
0.01
0.02
0.03
Predup_speciation (n = 2442)
Duplication (n = 2442)
Postdup_speciation (n = 2442)
B. Gene exp levels for teleosts
P = 4.37 × 10^−7
P < 2.2 × 10^−16
PICs of expression levels
0.00
0.05
0.10
0.15
0.20
0.25
Predup_speciation (n = 2660)
Duplication (n = 2660)
Postdup_speciation (n = 2660)
C. Tissue specificity, τ, for 15 species
P = 3.86 × 10^−4
P < 2.2 × 10^−16
PICs of τ
0.00
0.01
0.02
0.03
Predup_speciation (n = 2786)
Duplication (n = 2786)
Postdup_speciation (n = 2786)
D. Gene exp levels for 15 species
P = 5.03 × 10^−6
P < 2.2 × 10^−16
PICs of expression levels
0.00
0.05
0.10
0.15
0.20
0.25
Predup_speciation (n = 2786)
Duplication (n = 2786)
Postdup_speciation (n = 2786)

Fig. S5

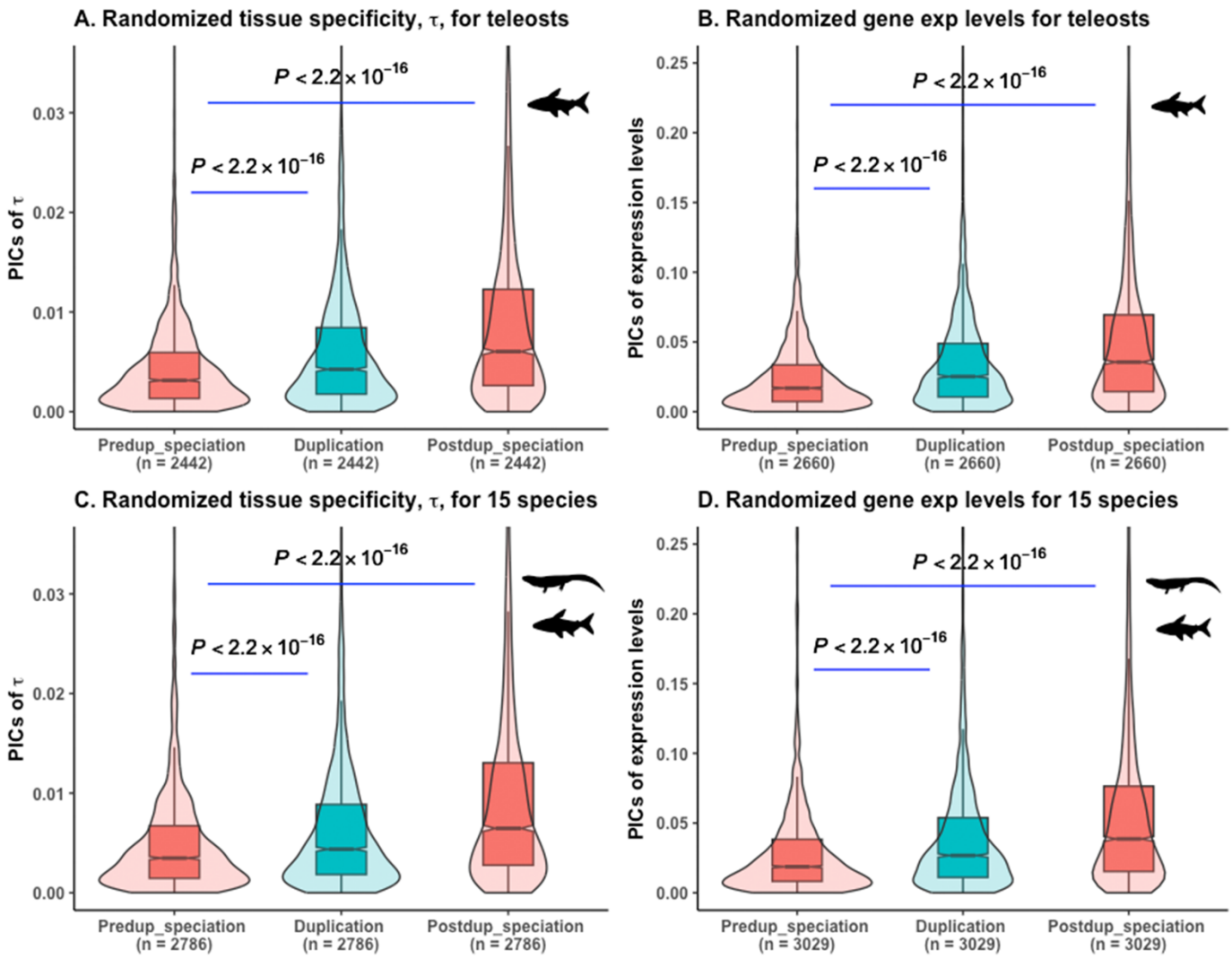

A. Randomized tissue specificity, τ, for teleosts
B. Randomized gene exp levels for teleosts
C. Randomized tissue specificity, τ, for 15 species
D. Randomized gene exp levels for 15 species
PICs of τ
PICs of expression levels
P < 2.2 × 10^−16
Predup_speciation (n = 2442)
Duplication (n = 2442)
Postdup_speciation (n = 2442)
Predup_speciation (n = 2660)
Duplication (n = 2660)
Postdup_speciation (n = 2660)
Predup_speciation (n = 2786)
Duplication (n = 2786)
Postdup_speciation (n = 2786)
Predup_speciation (n = 3029)
Duplication (n = 3029)
Postdup_speciation (n = 3029)

Fig. S6

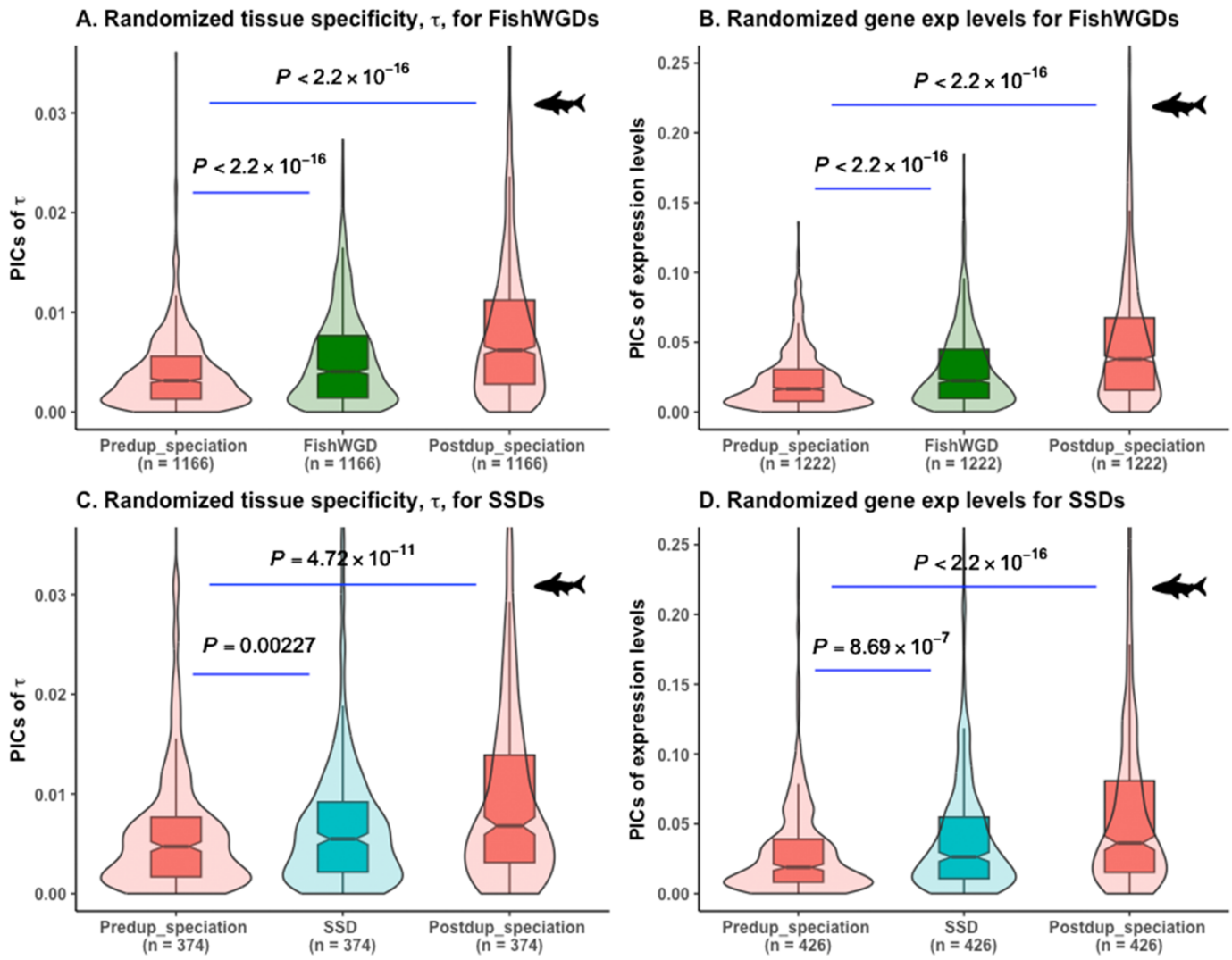
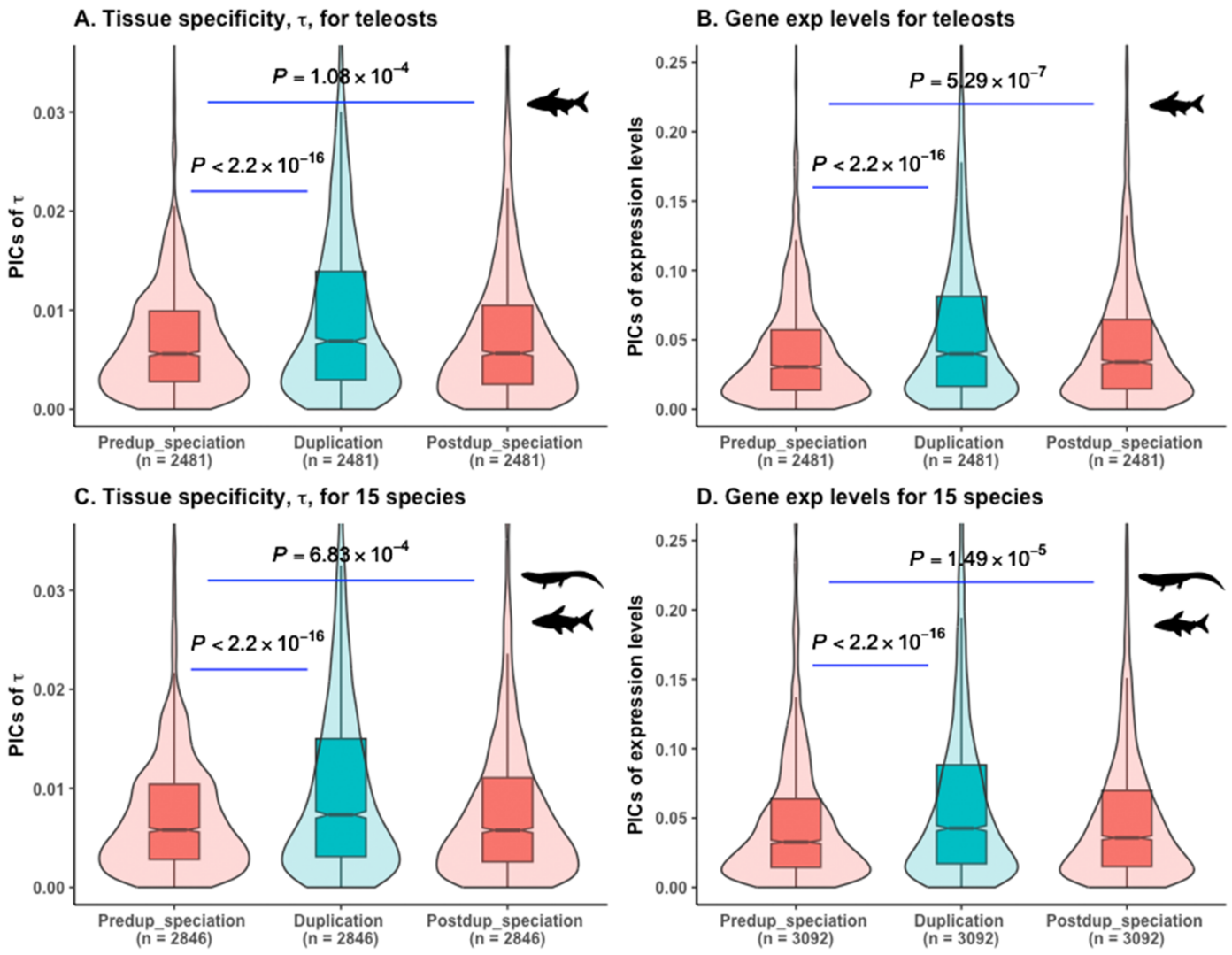

A. Randomized tissue specificity, τ, for FishWGDs
PICs of τ
P < 2.2 × 10^−16
P < 2.2 × 10^−16
Predup_speciation (n = 1166)
FishWGD (n = 1166)
Postdup_speciation (n = 1166)
B. Randomized gene exp levels for FishWGDs
PICs of expression levels
P < 2.2 × 10^−16
P < 2.2 × 10^−16
Predup_speciation (n = 1222)
FishWGD (n = 1222)
Postdup_speciation (n = 1222)
C. Randomized tissue specificity, τ, for SSDs
PICs of τ
P = 4.72 × 10^−11
P = 0.00227
Predup_speciation (n = 374)
SSD (n = 374)
Postdup_speciation (n = 374)
D. Randomized gene exp levels for SSDs
PICs of expression levels
P < 2.2 × 10^−16
P = 8.69 × 10^−7
Predup_speciation (n = 426)
SSD (n = 426)
Postdup_speciation (n = 426)

Fig. S7



228

229

230 **Fig. S8**

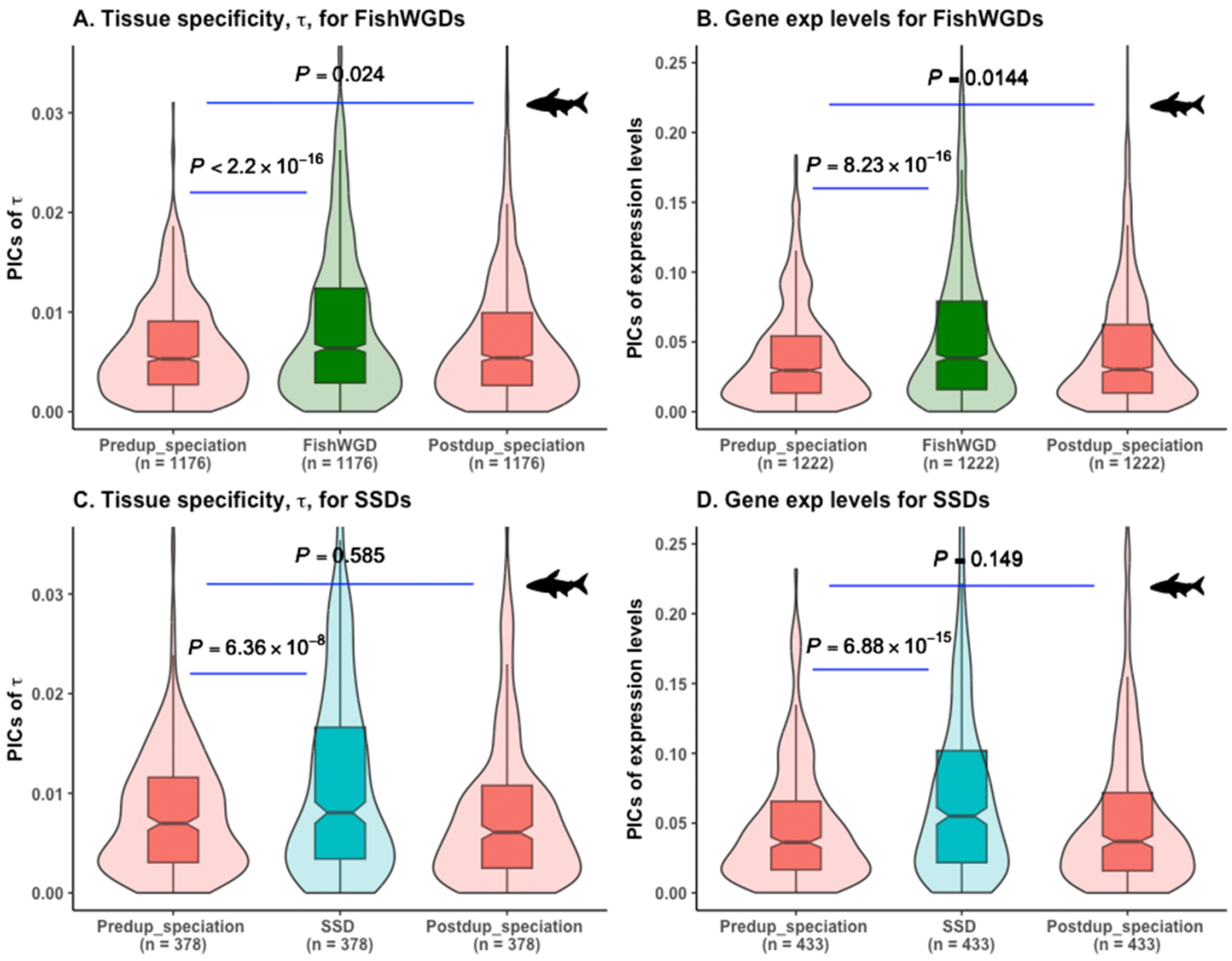
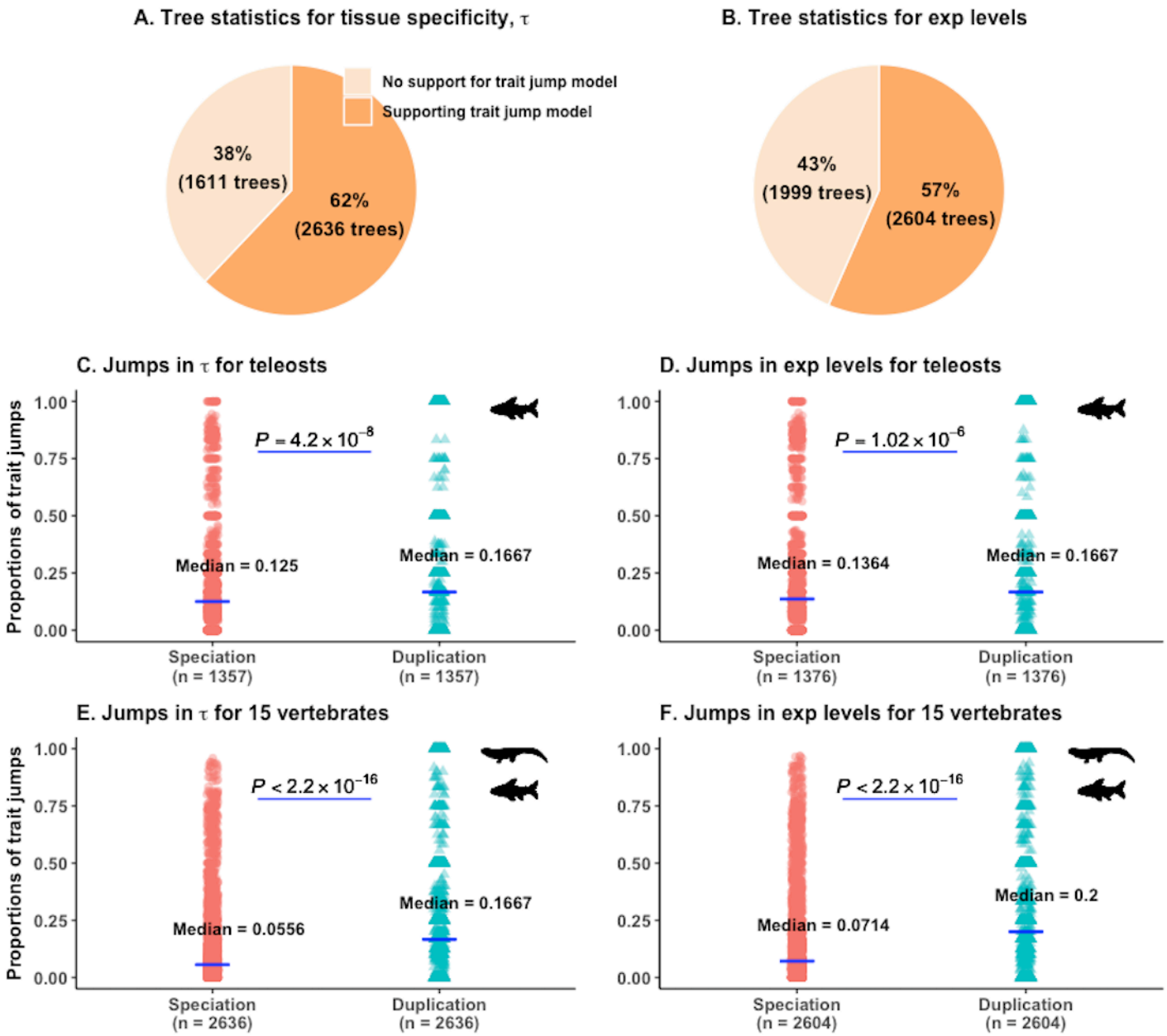

A. Tissue specificity, τ, for FishWGDs
P = 0.024
P < 2.2 × 10^−16
PICs of τ
0.00
0.01
0.02
0.03
Predup_speciation (n = 1176)
FishWGD (n = 1176)
Postdup_speciation (n = 1176)
B. Gene exp levels for FishWGDs
P = 0.0144
P = 8.23 × 10^−16
PICs of expression levels
0.00
0.05
0.10
0.15
0.20
0.25
Predup_speciation (n = 1222)
FishWGD (n = 1222)
Postdup_speciation (n = 1222)
C. Tissue specificity, τ, for SSDs
P = 0.585
P = 6.36 × 10^−8
PICs of τ
0.00
0.01
0.02
0.03
Predup_speciation (n = 378)
SSD (n = 378)
Postdup_speciation (n = 378)
D. Gene exp levels for SSDs
P = 0.149
P = 6.88 × 10^−15
PICs of expression levels
0.00
0.05
0.10
0.15
0.20
0.25
Predup_speciation (n = 433)
SSD (n = 433)
Postdup_speciation (n = 433)

Fig. S9


A. Tree statistics for tissue specificity, τ
No support for trait jump model
Supporting trait jump model
38%
(1611 trees)
62%
(2636 trees)
B. Tree statistics for exp levels
43%
(1999 trees)
57%
(2604 trees)
C. Jumps in τ for teleosts
Proportions of trait jumps
P = 4.2 × 10^-8
Median = 0.125
Median = 0.1667
Speciation
(n = 1357)
Duplication
(n = 1357)
D. Jumps in exp levels for teleosts
P = 1.02 × 10^-6
Median = 0.1364
Median = 0.1667
Speciation
(n = 1376)
Duplication
(n = 1376)
E. Jumps in τ for 15 vertebrates
P < 2.2 × 10^-16
Median = 0.0556
Median = 0.1667
Speciation
(n = 2636)
Duplication
(n = 2636)
F. Jumps in exp levels for 15 vertebrates
P < 2.2 × 10^-16
Median = 0.0714
Median = 0.2
Speciation
(n = 2604)
Duplication
(n = 2604)

.

Fig. S10

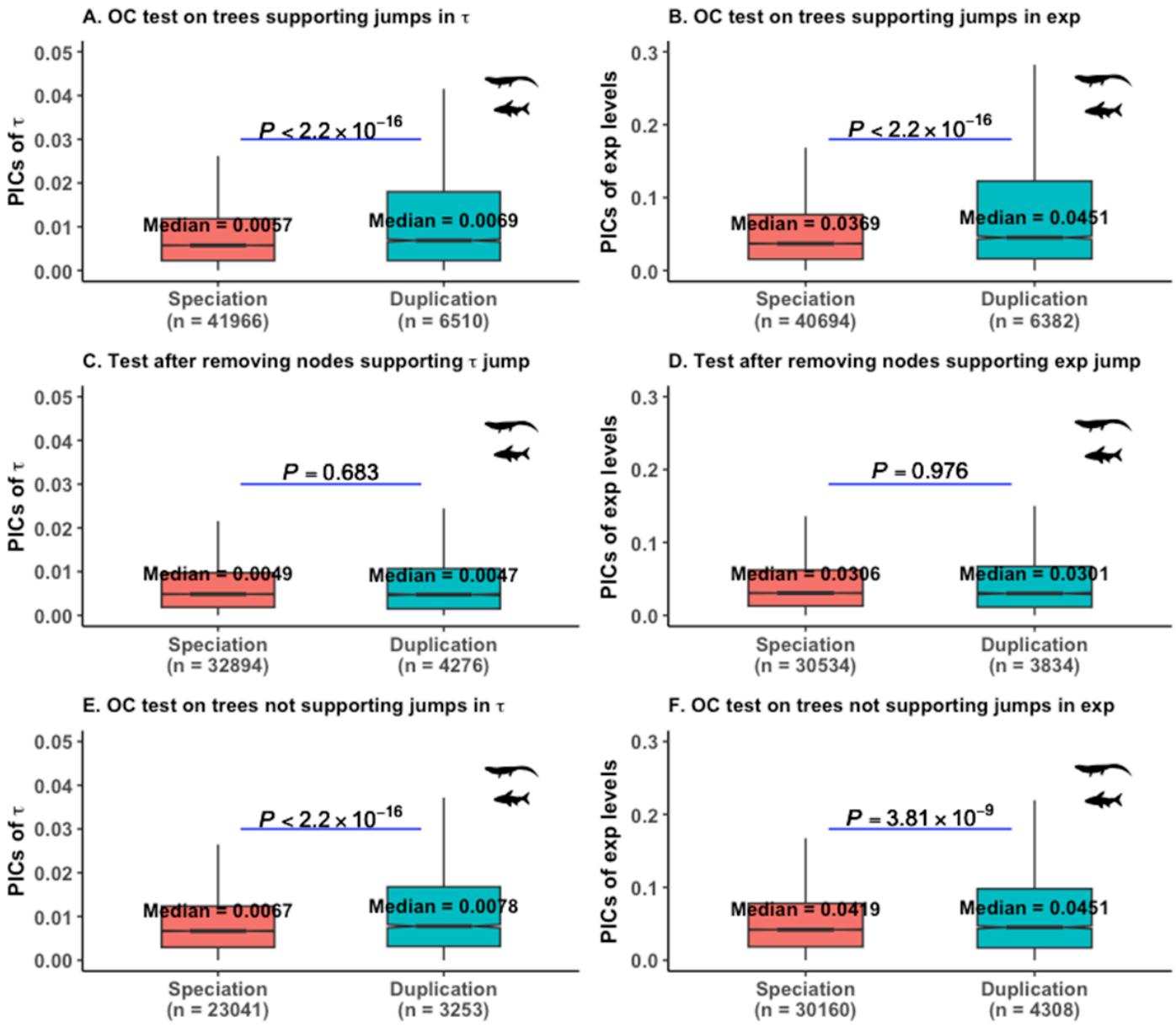

A. OC test on trees supporting jumps in τ
PICs of τ
P < 2.2 × 10^−16
Median = 0.0057
Median = 0.0069
Speciation (n = 41966)
Duplication (n = 6510)
B. OC test on trees supporting jumps in exp
PICs of exp levels
P < 2.2 × 10^−16
Median = 0.0369
Median = 0.0451
Speciation (n = 40694)
Duplication (n = 6382)
C. Test after removing nodes supporting τ jump
PICs of τ
P = 0.683
Median = 0.0049
Median = 0.0047
Speciation (n = 32894)
Duplication (n = 4276)
D. Test after removing nodes supporting exp jump
PICs of exp levels
P = 0.976
Median = 0.0306
Median = 0.0301
Speciation (n = 30534)
Duplication (n = 3834)
E. OC test on trees not supporting jumps in τ
PICs of τ
P < 2.2 × 10^−16
Median = 0.0067
Median = 0.0078
Speciation (n = 23041)
Duplication (n = 3253)
F. OC test on trees not supporting jumps in exp
PICs of exp levels
P = 3.81 × 10^−9
Median = 0.0419
Median = 0.0451
Speciation (n = 30160)
Duplication (n = 4308)

Fig. S11

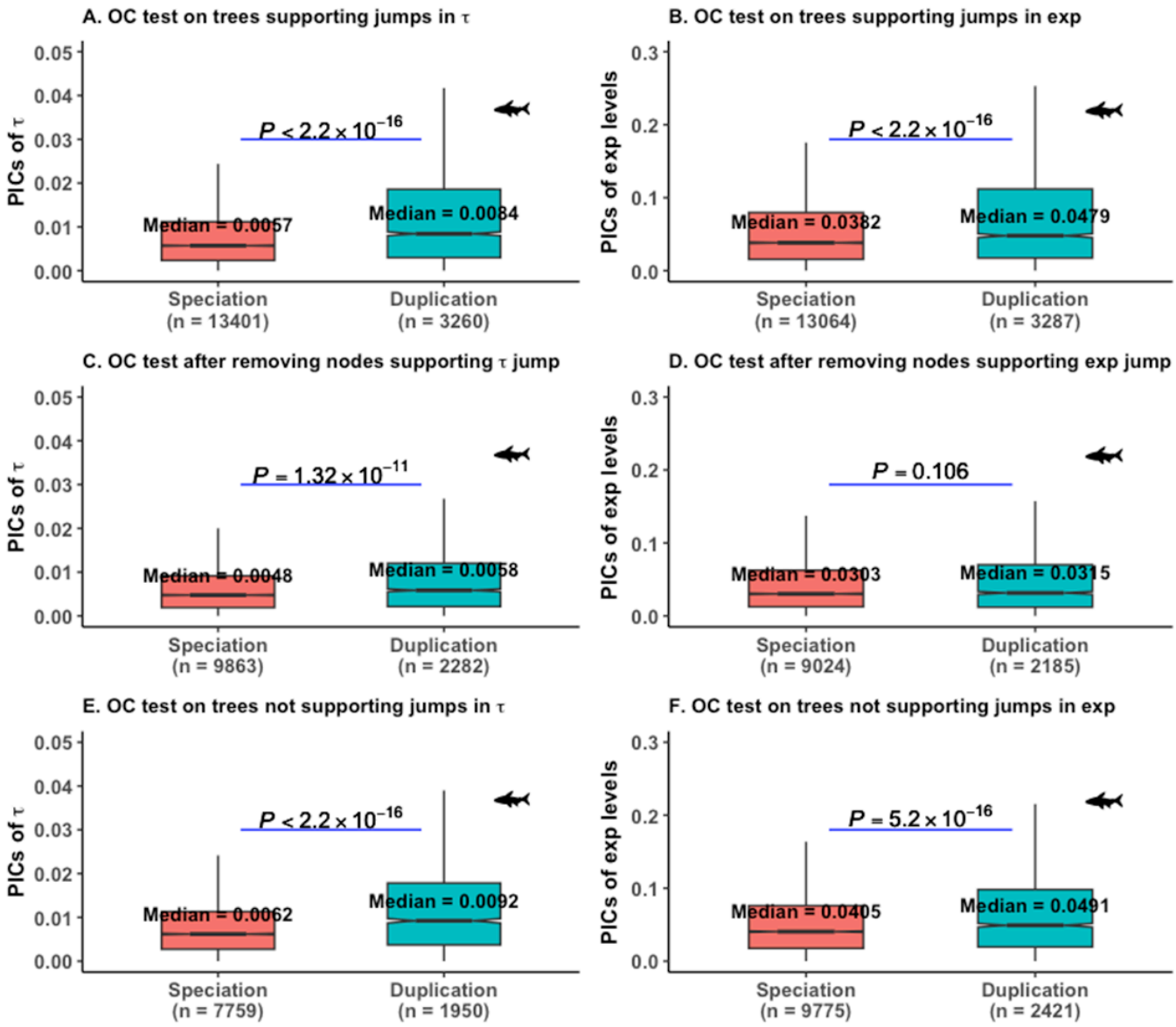

A. OC test on trees supporting jumps in τ
PICs of τ
P < 2.2 × 10^−16
Median = 0.0057
Median = 0.0084
Speciation (n = 13401)
Duplication (n = 3260)
B. OC test on trees supporting jumps in exp
PICs of exp levels
P < 2.2 × 10^−16
Median = 0.0382
Median = 0.0479
Speciation (n = 13064)
Duplication (n = 3287)
C. OC test after removing nodes supporting τ jump
PICs of τ
P = 1.32 × 10^−11
Median = 0.0048
Median = 0.0058
Speciation (n = 9863)
Duplication (n = 2282)
D. OC test after removing nodes supporting exp jump
PICs of exp levels
P = 0.106
Median = 0.0303
Median = 0.0315
Speciation (n = 9024)
Duplication (n = 2185)
E. OC test on trees not supporting jumps in τ
PICs of τ
P < 2.2 × 10^−16
Median = 0.0062
Median = 0.0092
Speciation (n = 7759)
Duplication (n = 1950)
F. OC test on trees not supporting jumps in exp
PICs of exp levels
P = 5.2 × 10^−16
Median = 0.0405
Median = 0.0491
Speciation (n = 9775)
Duplication (n = 2421)

Fig. S12

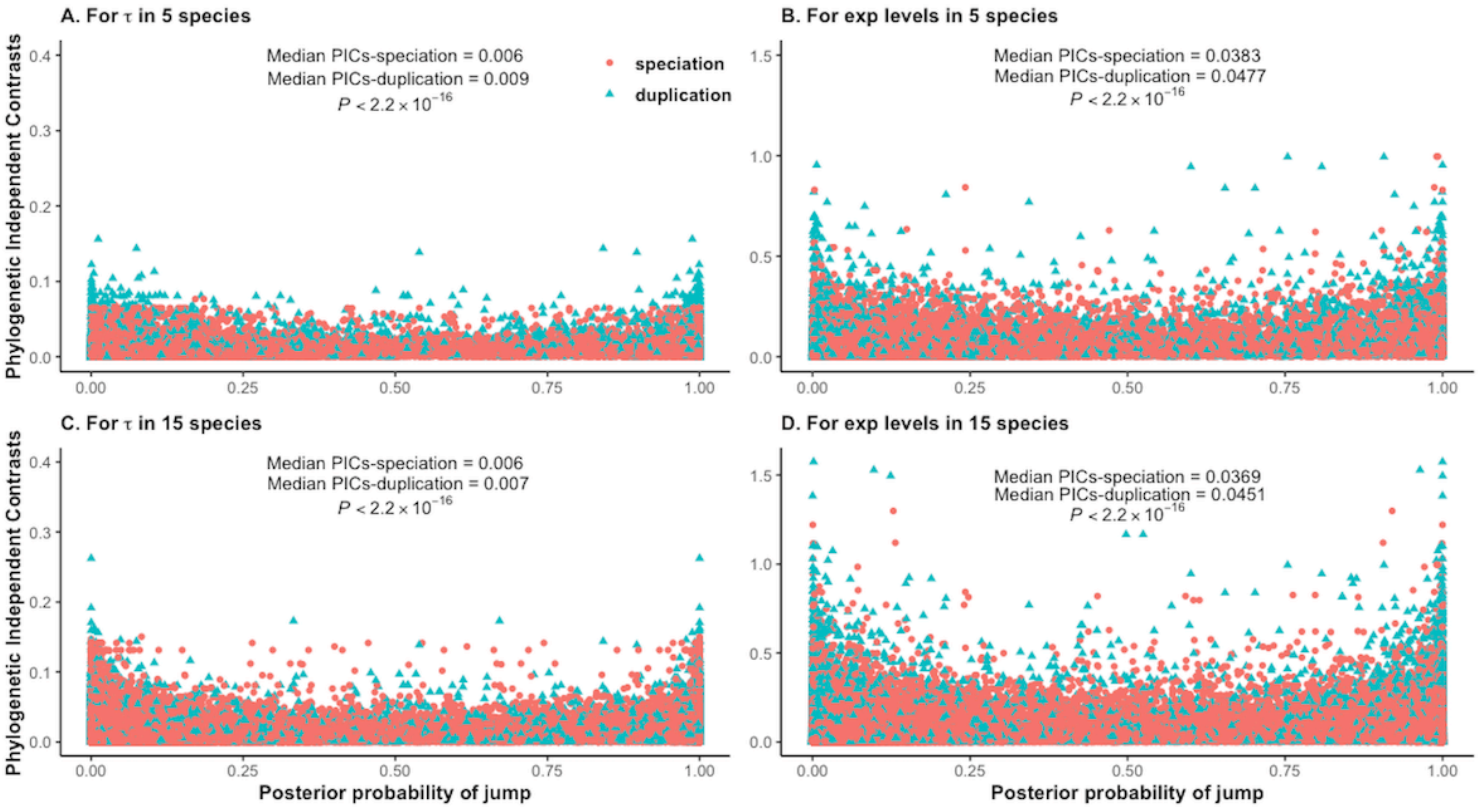

A. For τ in 5 species
Median PICs-speciation = 0.006
Median PICs-duplication = 0.009
P < 2.2 × 10^-16
speciation
duplication
B. For exp levels in 5 species
Median PICs-speciation = 0.0383
Median PICs-duplication = 0.0477
P < 2.2 × 10^-16
C. For τ in 15 species
Median PICs-speciation = 0.006
Median PICs-duplication = 0.007
P < 2.2 × 10^-16
D. For exp levels in 15 species
Median PICs-speciation = 0.0369
Median PICs-duplication = 0.0451
P < 2.2 × 10^-16
Phylogenetic Independent Contrasts
Posterior probability of jump

Fig. S13

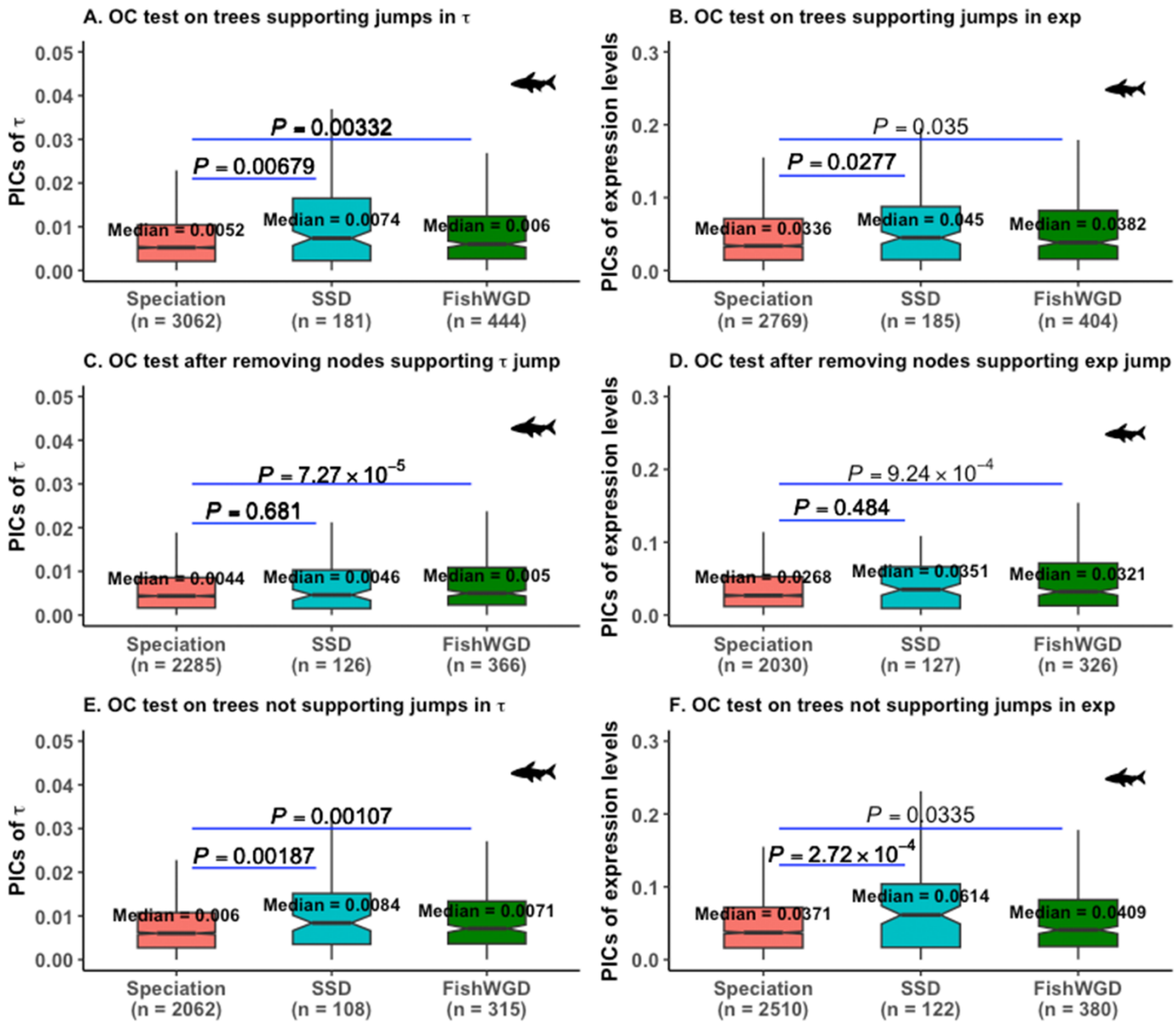

A. OC test on trees supporting jumps in τ
PICs of τ
P = 0.00332
P = 0.00679
Median = 0.0052
Median = 0.0074
Median = 0.006
Speciation (n = 3062)
SSD (n = 181)
FishWGD (n = 444)
B. OC test on trees supporting jumps in exp
PICs of expression levels
P = 0.035
P = 0.0277
Median = 0.0336
Median = 0.045
Median = 0.0382
Speciation (n = 2769)
SSD (n = 185)
FishWGD (n = 404)
C. OC test after removing nodes supporting τ jump
PICs of τ
P = 7.27 × 10^−5
P = 0.681
Median = 0.0044
Median = 0.0046
Median = 0.005
Speciation (n = 2285)
SSD (n = 126)
FishWGD (n = 366)
D. OC test after removing nodes supporting exp jump
PICs of expression levels
P = 9.24 × 10^−4
P = 0.484
Median = 0.0268
Median = 0.0351
Median = 0.0321
Speciation (n = 2030)
SSD (n = 127)
FishWGD (n = 326)
E. OC test on trees not supporting jumps in τ
PICs of τ
P = 0.00107
P = 0.00187
Median = 0.006
Median = 0.0084
Median = 0.0071
Speciation (n = 2062)
SSD (n = 108)
FishWGD (n = 315)
F. OC test on trees not supporting jumps in exp
PICs of expression levels
P = 0.0335
P = 2.72 × 10^−4
Median = 0.0371
Median = 0.0614
Median = 0.0409
Speciation (n = 2510)
SSD (n = 122)
FishWGD (n = 380)

Fig. S14

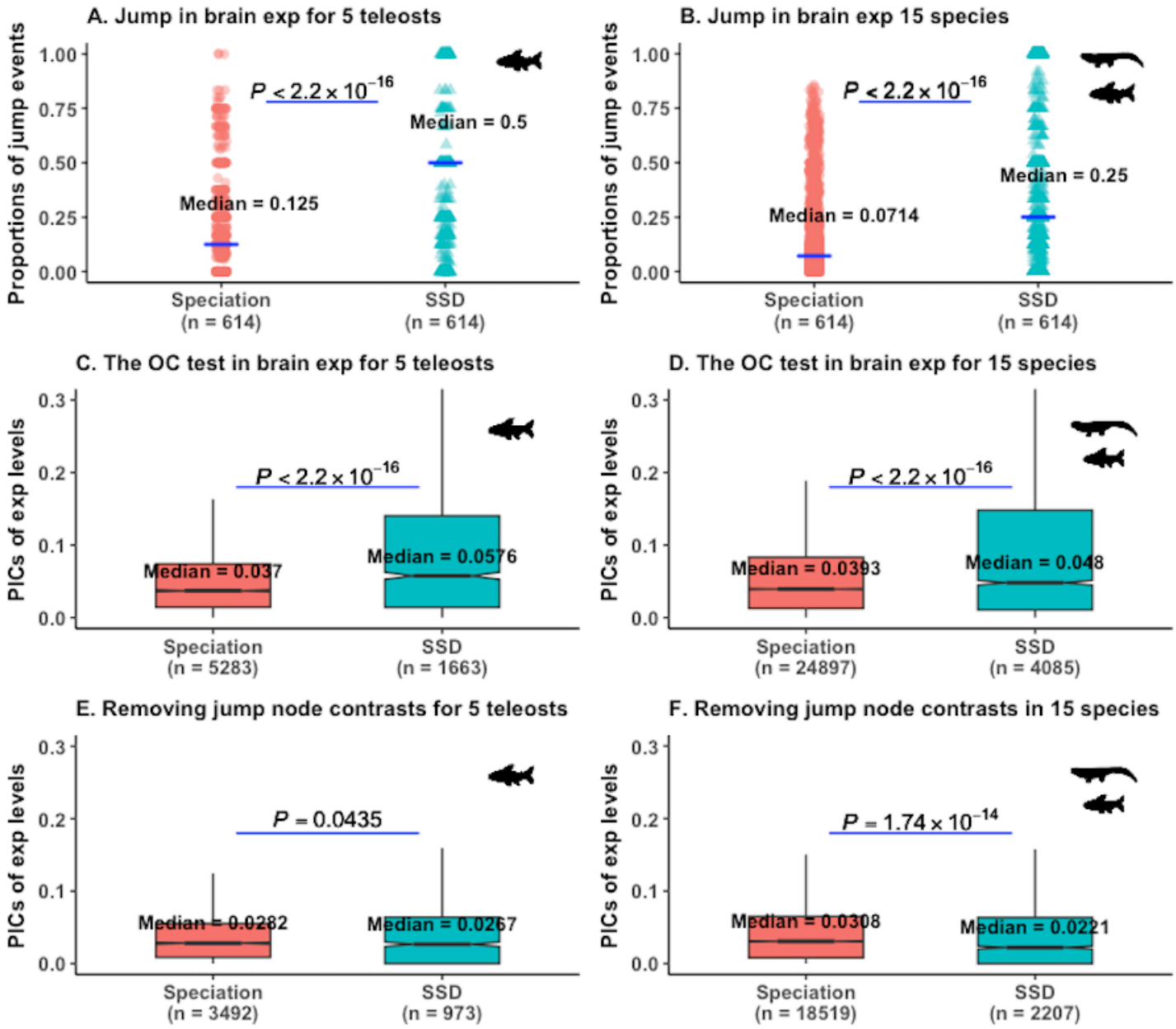

A. Jump in brain exp for 5 teleosts
Proportions of jump events
1.00
0.75
0.50
0.25
0.00
P < 2.2 × 10^-16
Median = 0.5
Median = 0.125
Speciation
(n = 614)
SSD
(n = 614)
B. Jump in brain exp 15 species
Proportions of jump events
1.00
0.75
0.50
0.25
0.00
P < 2.2 × 10^-16
Median = 0.25
Median = 0.0714
Speciation
(n = 614)
SSD
(n = 614)
C. The OC test in brain exp for 5 teleosts
PICs of exp levels
0.3
0.2
0.1
0.0
P < 2.2 × 10^-16
Median = 0.0576
Median = 0.037
Speciation
(n = 5283)
SSD
(n = 1663)
D. The OC test in brain exp for 15 species
PICs of exp levels
0.3
0.2
0.1
0.0
P < 2.2 × 10^-16
Median = 0.048
Median = 0.0393
Speciation
(n = 24897)
SSD
(n = 4085)
E. Removing jump node contrasts for 5 teleosts
PICs of exp levels
0.3
0.2
0.1
0.0
P = 0.0435
Median = 0.0267
Median = 0.0282
Speciation
(n = 3492)
SSD
(n = 973)
F. Removing jump node contrasts in 15 species
PICs of exp levels
0.3
0.2
0.1
0.0
P = 1.74 × 10^-14
Median = 0.0221
Median = 0.0308
Speciation
(n = 18519)
SSD
(n = 2207)

Fig. S15

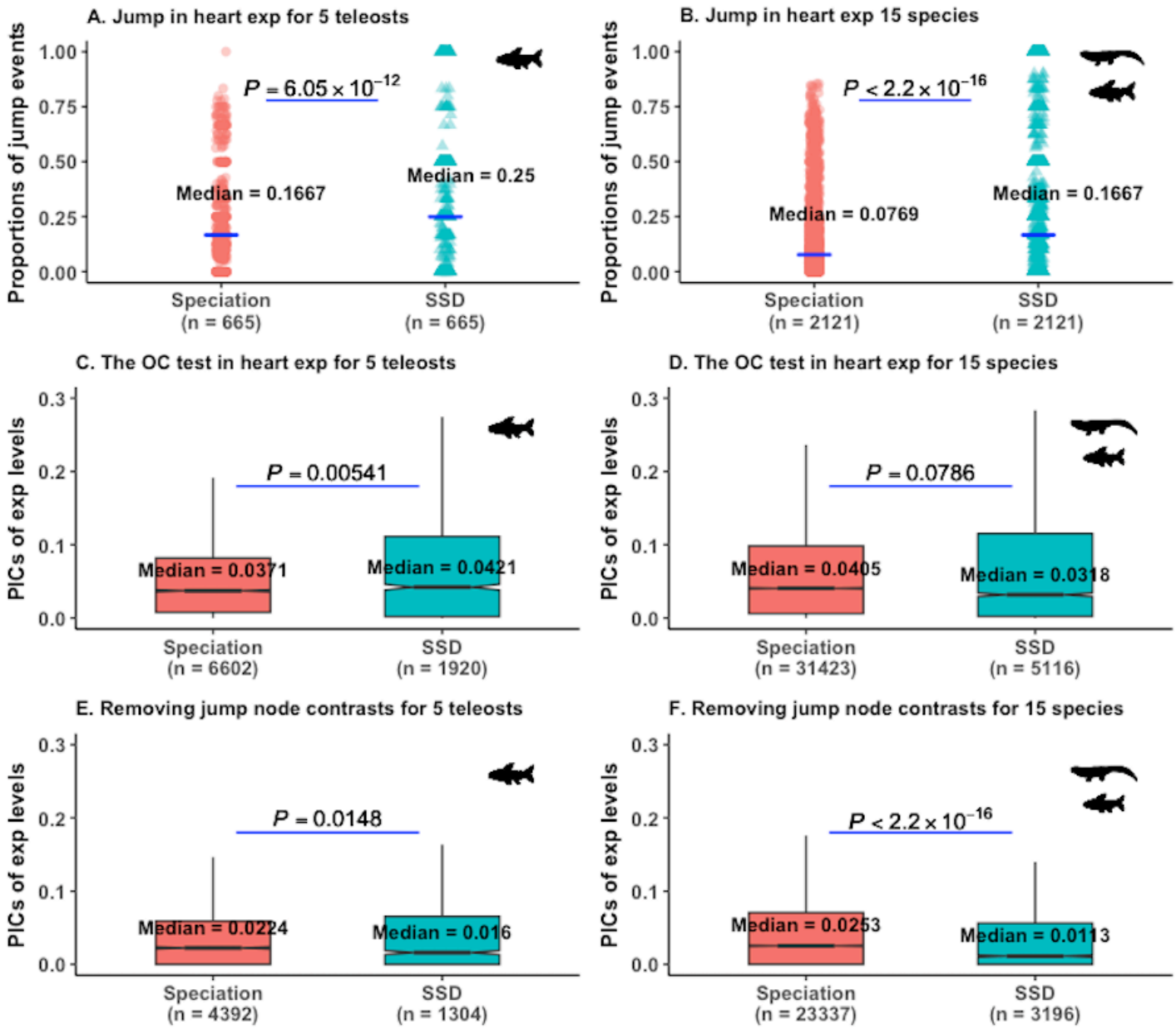

A. Jump in heart exp for 5 teleosts
Proportions of jump events
1.00
0.75
0.50
0.25
0.00
P = 6.05 × 10^−12
Median = 0.1667
Median = 0.25
Speciation
(n = 665)
SSD
(n = 665)
B. Jump in heart exp 15 species
Proportions of jump events
P < 2.2 × 10^−16
Median = 0.0769
Median = 0.1667
Speciation
(n = 2121)
SSD
(n = 2121)
C. The OC test in heart exp for 5 teleosts
PICs of exp levels
0.3
0.2
0.1
0.0
P = 0.00541
Median = 0.0371
Median = 0.0421
Speciation
(n = 6602)
SSD
(n = 1920)
D. The OC test in heart exp for 15 species
PICs of exp levels
P = 0.0786
Median = 0.0405
Median = 0.0318
Speciation
(n = 31423)
SSD
(n = 5116)
E. Removing jump node contrasts for 5 teleosts
PICs of exp levels
P = 0.0148
Median = 0.0224
Median = 0.016
Speciation
(n = 4392)
SSD
(n = 1304)
F. Removing jump node contrasts for 15 species
PICs of exp levels
P < 2.2 × 10^−16
Median = 0.0253
Median = 0.0113
Speciation
(n = 23337)
SSD
(n = 3196)

Fig. S16

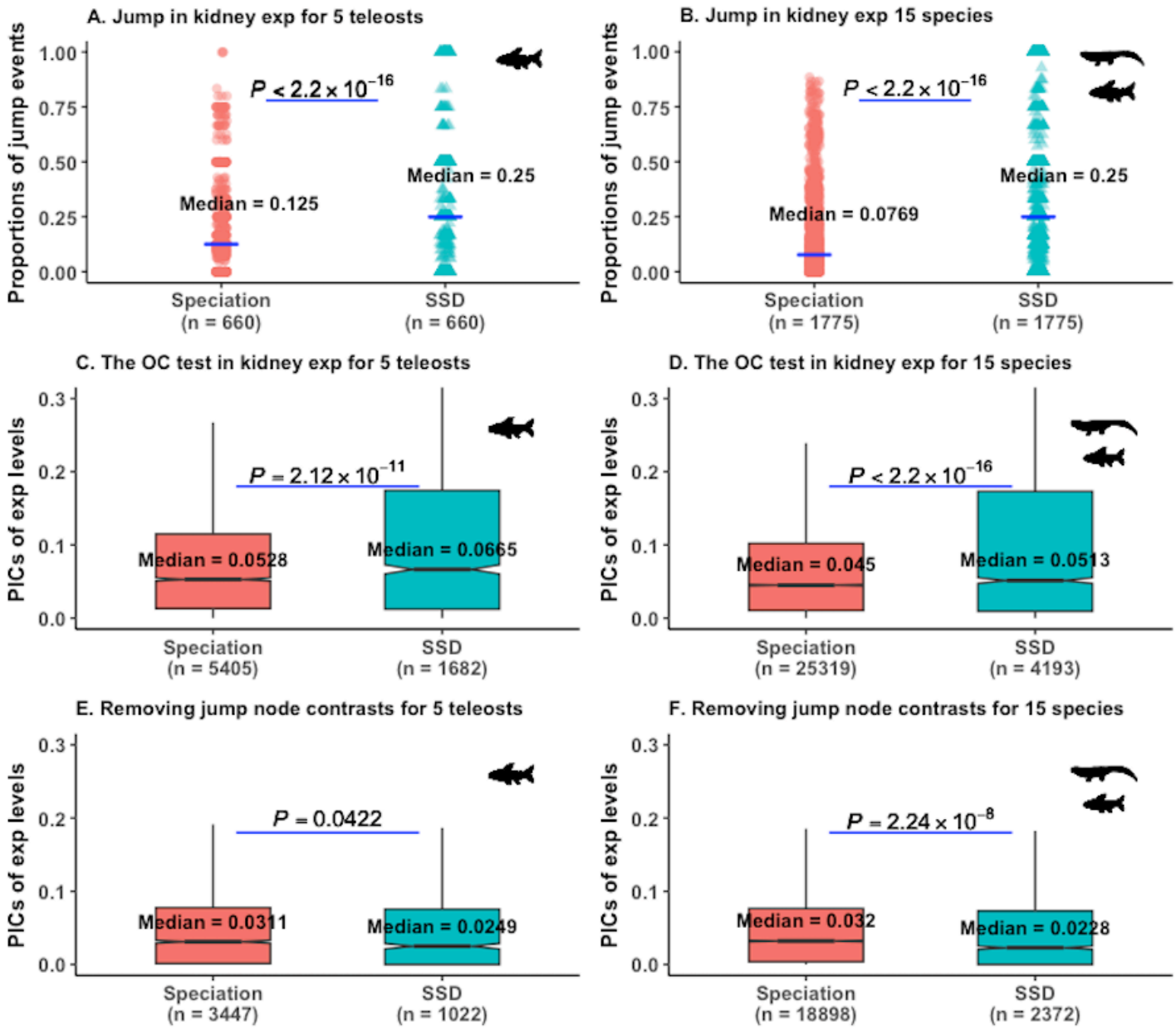

A. Jump in kidney exp for 5 teleosts
Proportions of jump events
1.00
0.75
0.50
0.25
0.00
P < 2.2 × 10^−16
Median = 0.125
Median = 0.25
Speciation
(n = 660)
SSD
(n = 660)
B. Jump in kidney exp 15 species
Proportions of jump events
P < 2.2 × 10^−16
Median = 0.0769
Median = 0.25
Speciation
(n = 1775)
SSD
(n = 1775)
C. The OC test in kidney exp for 5 teleosts
PICs of exp levels
0.3
0.2
0.1
0.0
P = 2.12 × 10^−11
Median = 0.0528
Median = 0.0665
Speciation
(n = 5405)
SSD
(n = 1682)
D. The OC test in kidney exp for 15 species
PICs of exp levels
P < 2.2 × 10^−16
Median = 0.045
Median = 0.0513
Speciation
(n = 25319)
SSD
(n = 4193)
E. Removing jump node contrasts for 5 teleosts
PICs of exp levels
P = 0.0422
Median = 0.0311
Median = 0.0249
Speciation
(n = 3447)
SSD
(n = 1022)
F. Removing jump node contrasts for 15 species
PICs of exp levels
P = 2.24 × 10^−8
Median = 0.032
Median = 0.0228
Speciation
(n = 18898)
SSD
(n = 2372)

Fig. S17

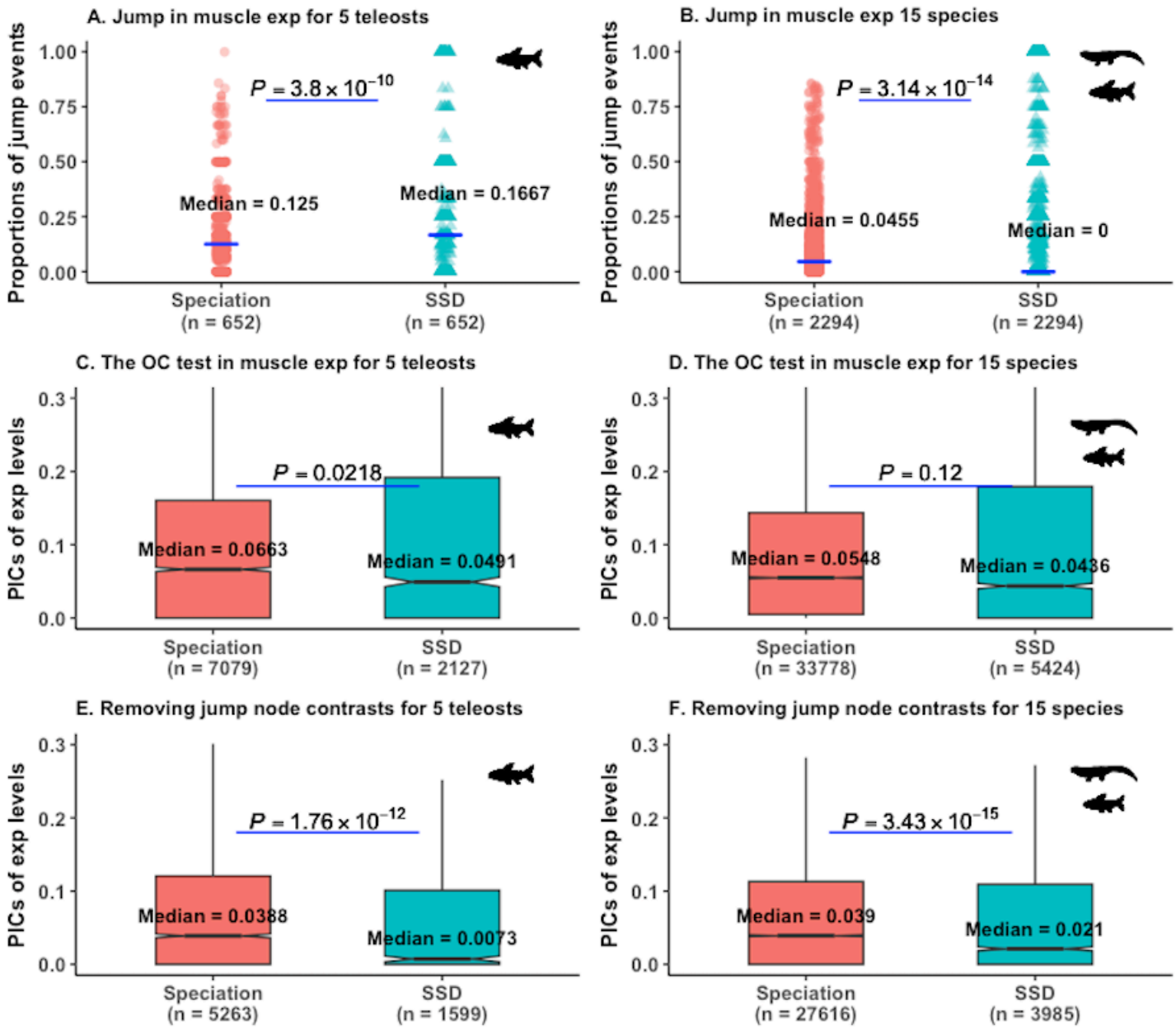

A. Jump in muscle exp for 5 teleosts
Proportions of jump events
1.00
0.75
0.50
0.25
0.00
P = 3.8 × 10^−10
Median = 0.125
Median = 0.1667
Speciation
(n = 652)
SSD
(n = 652)
B. Jump in muscle exp 15 species
Proportions of jump events
P = 3.14 × 10^−14
Median = 0.0455
Median = 0
Speciation
(n = 2294)
SSD
(n = 2294)
C. The OC test in muscle exp for 5 teleosts
PICs of exp levels
0.3
0.2
0.1
0.0
P = 0.0218
Median = 0.0663
Median = 0.0491
Speciation
(n = 7079)
SSD
(n = 2127)
D. The OC test in muscle exp for 15 species
PICs of exp levels
P = 0.12
Median = 0.0548
Median = 0.0436
Speciation
(n = 33778)
SSD
(n = 5424)
E. Removing jump node contrasts for 5 teleosts
PICs of exp levels
P = 1.76 × 10^−12
Median = 0.0388
Median = 0.0073
Speciation
(n = 5263)
SSD
(n = 1599)
F. Removing jump node contrasts for 15 species
PICs of exp levels
P = 3.43 × 10^−15
Median = 0.039
Median = 0.021
Speciation
(n = 27616)
SSD
(n = 3985)

Fig. S18

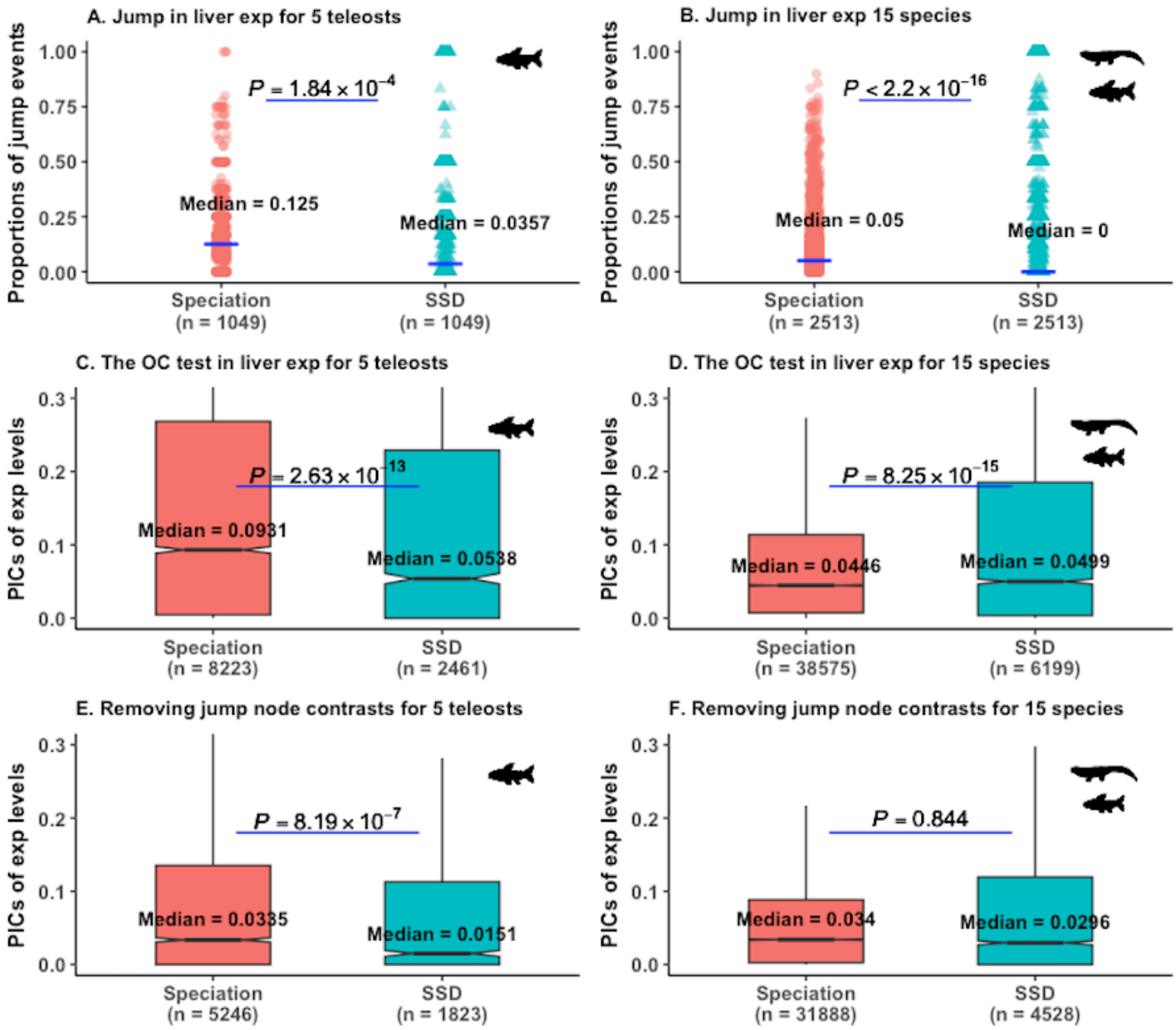

A. Jump in liver exp for 5 teleosts
Proportions of jump events
P = 1.84 × 10^−4
Median = 0.125
Median = 0.0357
Speciation (n = 1049)
SSD (n = 1049)
B. Jump in liver exp 15 species
P < 2.2 × 10^−16
Median = 0.05
Median = 0
Speciation (n = 2513)
SSD (n = 2513)
C. The OC test in liver exp for 5 teleosts
PICs of exp levels
P = 2.63 × 10^−13
Median = 0.0931
Median = 0.0538
Speciation (n = 8223)
SSD (n = 2461)
D. The OC test in liver exp for 15 species
P = 8.25 × 10^−15
Median = 0.0446
Median = 0.0499
Speciation (n = 38575)
SSD (n = 6199)
E. Removing jump node contrasts for 5 teleosts
P = 8.19 × 10^−7
Median = 0.0335
Median = 0.0151
Speciation (n = 5246)
SSD (n = 1823)
F. Removing jump node contrasts for 15 species
P = 0.844
Median = 0.034
Median = 0.0296
Speciation (n = 31888)
SSD (n = 4528)

Fig. S19

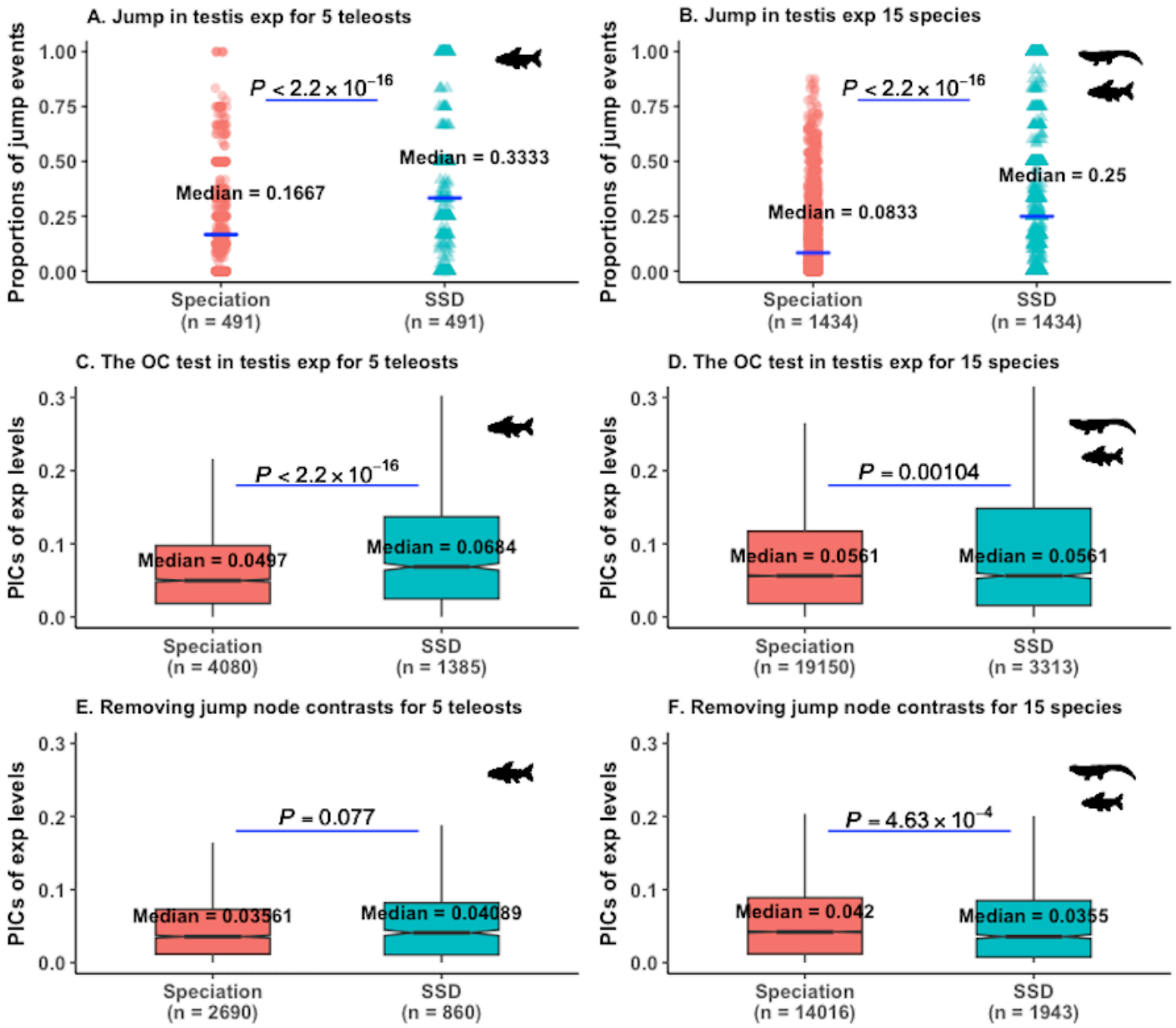

A. Jump in testis exp for 5 teleosts
Proportions of jump events
1.00
0.75
0.50
0.25
0.00
P < 2.2 × 10^-16
Median = 0.1667
Median = 0.3333
Speciation
(n = 491)
SSD
(n = 491)
B. Jump in testis exp 15 species
Proportions of jump events
P < 2.2 × 10^-16
Median = 0.0833
Median = 0.25
Speciation
(n = 1434)
SSD
(n = 1434)
C. The OC test in testis exp for 5 teleosts
PICs of exp levels
0.3
0.2
0.1
0.0
P < 2.2 × 10^-16
Median = 0.0497
Median = 0.0684
Speciation
(n = 4080)
SSD
(n = 1385)
D. The OC test in testis exp for 15 species
PICs of exp levels
P = 0.00104
Median = 0.0561
Median = 0.0561
Speciation
(n = 19150)
SSD
(n = 3313)
E. Removing jump node contrasts for 5 teleosts
PICs of exp levels
P = 0.077
Median = 0.03561
Median = 0.04089
Speciation
(n = 2690)
SSD
(n = 860)
F. Removing jump node contrasts for 15 species
PICs of exp levels
P = 4.63 × 10^-4
Median = 0.042
Median = 0.0355
Speciation
(n = 14016)
SSD
(n = 1943)

Fig. S20

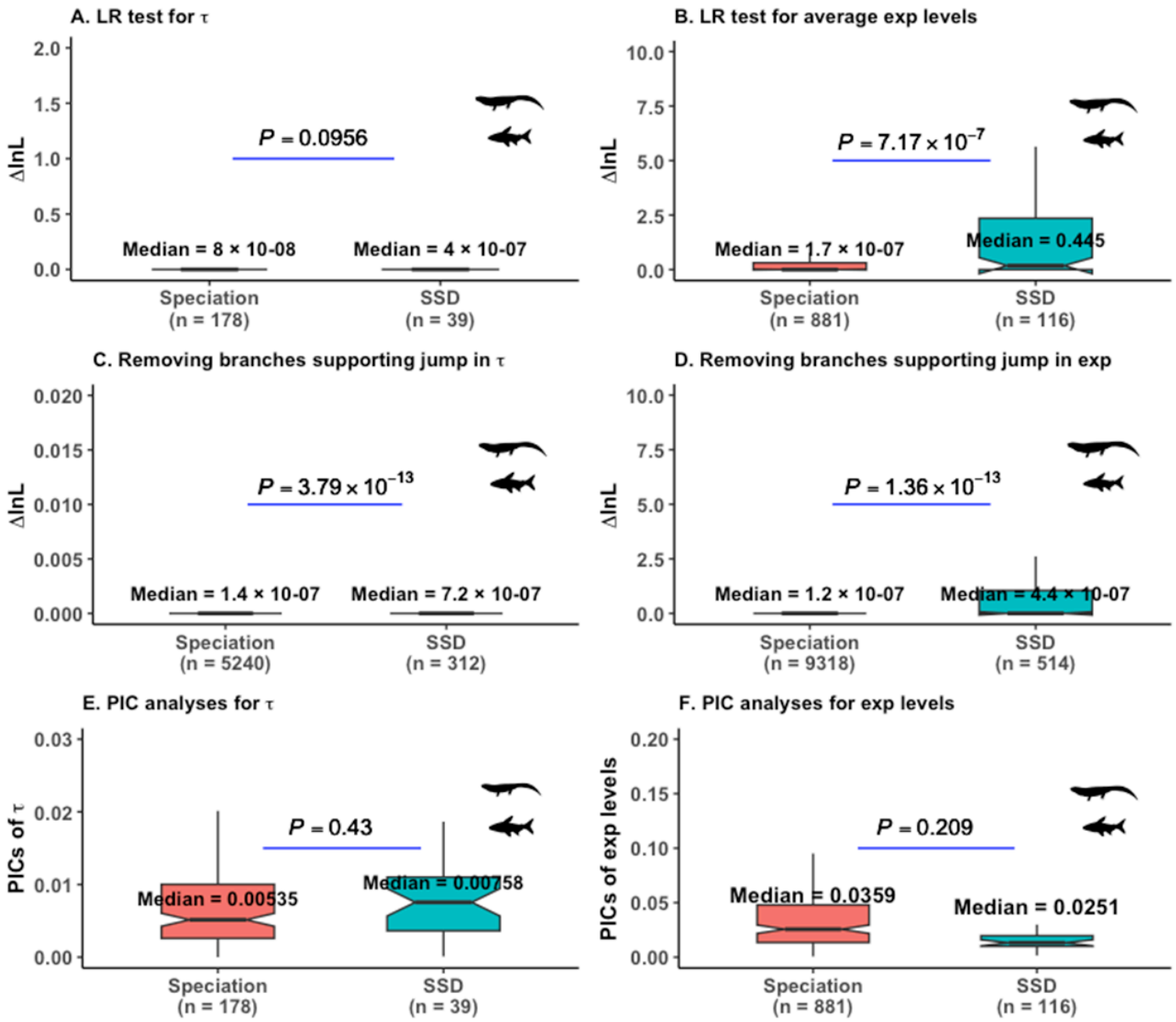

A. LR test for τ
ΔlnL
P = 0.0956
Median = 8 × 10-08
Median = 4 × 10-07
Speciation (n = 178)
SSD (n = 39)
B. LR test for average exp levels
ΔlnL
P = 7.17 × 10^-7
Median = 1.7 × 10-07
Median = 0.445
Speciation (n = 881)
SSD (n = 116)
C. Removing branches supporting jump in τ
ΔlnL
P = 3.79 × 10^-13
Median = 1.4 × 10-07
Median = 7.2 × 10-07
Speciation (n = 5240)
SSD (n = 312)
D. Removing branches supporting jump in exp
ΔlnL
P = 1.36 × 10^-13
Median = 1.2 × 10-07
Median = 4.4 × 10-07
Speciation (n = 9318)
SSD (n = 514)
E. PIC analyses for τ
PICs of τ
P = 0.43
Median = 0.00535
Median = 0.00758
Speciation (n = 178)
SSD (n = 39)
F. PIC analyses for exp levels
PICs of exp levels
P = 0.209
Median = 0.0359
Median = 0.0251
Speciation (n = 881)
SSD (n = 116)

Fig. S21