Organ-specific propensity drives patterns of gene expression evolution

Kenji Fukushima[1,2*] and David D. Pollock[1,*]

[1] Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO 80045, USA.

[2] Present address: Institute for Molecular Plant Physiology and Biophysics, University of Würzburg

* Correspondence to: kenji.fukushima@uni-wuerzburg.de (K.F.) and david.pollock@cuanschutz.edu (D.D.P.)

Keywords: evolutionary transcriptomics, preadaptation, vertebrate evolution

Abstract:

The origins of multicellular physiology are tied to evolution of gene expression. Genes can shift expression as organisms evolve, but how ancestral expression influences altered descendant expression is not well understood. To examine this, we amalgamated 1,903 RNA-seq datasets from 182 research projects, including 6 organs in 21 vertebrate species. Quality control eliminated project-specific biases, and expression shifts were reconstructed using gene-family-wise phylogenetic Ornstein–Uhlenbeck models. Expression shifts following gene duplication result in more drastic changes in expression properties. The expression properties were tightly coupled with protein evolutionary rate, depending on whether and how gene duplication occurred. Fluxes in expression patterns among organs were nonrandom, forming modular connections which were reshaped by gene duplication. Thus, if expression shifted, ancestral expression in some organs induces a strong propensity for expression in particular organs in descendants. This supports a major role for what might be termed "preadaptive" pathways of gene expression evolution.

Introduction:

Vertebrate organs organize physiological activities, and their identities and functions are determined by the diverse expression patterns of thousands of genes. Because of this, the evolution of gene expression patterns plays a central role in organismal evolution. The degree of organ expression specificity correlates to how fast amino acids substitute (Zhang and Li, 2004), how rapidly they change expression levels (Liao and Zhang, 2006), and patterns of histone modifications (She et al., 2009). Major organ-altering evolutionary events such as development of the hominoid brain are also associated with gene expression shifts (Chen et al., 2013; Kaessmann, 2010; Zhang and Long, 2014; Zhang et al., 2011). However, although gene duplication is well-known to play an important role in expression pattern shifts (see e.g. the "ortholog conjecture" (Altenhoff et al., 2012; Castillo-Davis et al., 2004; Chen and Zhang, 2012; Rogozin et al., 2014)), the evolutionary dynamics of expression patterns with and without gene duplication remain poorly understood. An important question is whether long-term expression in one organ predisposes genes to be subsequently utilized in other organs.

A theoretical basis for such predisposition may be the idea that certain pre-existing adapted states are more conducive to evolution of specific new traits than other pre-existing states. This is known as preadaptation, and when a trait makes such a shift it is referred to as exaptation (Gould and Vrba, 1982). Evidence for preadaptation was long ago found in phenotypic traits (Budd, 2007), and recently in molecular traits such as protein sequences during *de novo* gene birth (Wilson et al., 2017) or during functional innovations (Starr et al., 2017). Protein sequence evolution generally involves highly epistatic interactions and context-dependent changes (Podgornaia and Laub, 2015; Starr et al., 2017) that affect preadaptation, but the modular nature of expression regulation (de-Leon and Davidson, 2007) makes it unclear whether preexisting expression patterns constrain evolutionary outcomes. By analyzing such influences on a genome-wide scale in this study, we aim to contribute some evidence towards this question.

Evolution of gene expression has been studied at genome-wide scales mainly using two distinct approaches: phylogenetic and pairwise analyses. Phylogenetic approaches model gene expression dynamics and infer ancestral expression patterns in the context of gene phylogenies. For example, Brownian motion models embody purely neutral expression evolution (Bedford and Hartl, 2009), whereas Ornstein–Uhlenbeck models are designed to detect purifying selection and adaptive evolution along with neutral

63 fluctuation (Brawand et al., 2011; Chen et al., 2019; Rohlfs et al., 2014). Although each
64 gene family has a distinct evolutionary history, a species phylogeny is often used for the
65 sake of simplicity. Because such approximations cannot be applied to gene families with
66 lineage-specific gene duplications and losses, its application has mostly been limited to
67 single-copy genes. In contrast, pairwise analysis compares gene expression between
68 paralogs in single species (Assis and Bachtrog, 2015; Lan and Pritchard, 2016) or
69 between orthologs or paralogs in pairs of species (Chen and Zhang, 2012; Guschanski et
70 al., 2017; Kryuchkova-Mostacci et al., 2016; Rogozin et al., 2014; Warnefors and
71 Kaessmann, 2013). Although pairwise approaches can evaluate the effect of gene
72 duplications, ancestral expression cannot be inferred.
73 To infer the adaptive evolution of gene expression in diverse gene families, we
74 applied Ornstein–Uhlenbeck models for complex gene family phylogenies containing
75 gene duplications and losses, without assuming species phylogeny. We also developed a
76 curation pipeline to amalgamate large amounts of transcriptome data from many studies
77 for a better phylogenetic resolution. Using these methods, we ask how expression changes
78 link to gene duplication and the rate of protein evolution, and whether previous patterns
79 of gene expression constrain later evolution, what we will call the "preadaptation
80 hypothesis".
81

82 **Results:**

83

84 **Duplication-permissive genome-wide analysis of gene families.** To allow evolutionary
85 expression analysis on a broad set of genes, we used a phylogenetic approach that deals
86 with the complex history of gene family trees with duplications and losses, and applied it
87 to 21 tetrapod genomes (Fig. S1). A major challenge in using gene trees was divergence
88 time estimation, a prerequisite for applying phylogenetic comparative methods. We
89 overcame this problem by incorporating phylogeny reconciliation in estimating
90 divergence time of gene trees. Gene divergence nodes were constrained by the
91 corresponding divergence times in a known species tree, and duplication nodes were
92 constrained by ancestral and descendant speciation events (see Materials and Methods for
93 details). Because we estimated individual gene phylogenies rather than using a single
94 species phylogeny, we could analyze gene families that included many lineage-specific
95 gene duplications and losses, making our study less biased towards conserved genes with
96 slow gene turnovers (Guschanski et al., 2017). Use of gene family trees also allowed us
97 to include many organ-specific genes that are enriched in lineage-specific and young
98 duplications (Guschanski et al., 2017). There were only 1,377 single-copy orthologs for
99 which the species phylogeny was applicable, but we were able to include 15,475 genes
100 per species on average (including 20,873 human genes, merged into 15,280 gene families).
101 This approach eliminates problems with pairwise analyses that ignore phylogenetic tree
102 structure, and allowed us to infer expression at ancestral nodes in the tree.

103

104 **Transcriptome amalgamation.** To attain high resolution in our analyses, we
105 amalgamated 1,903 RNA-seq experiments from 182 research projects (i.e., 182
106 BioProject IDs in the NCBI SRA database) and generated a dataset covering 6 organs
107 from 21 vertebrate species without missing data (Tables S1–S2; Fig. S1). In comparison,
108 other recent comparative transcriptomic analyses of vertebrates (Barbosa-Morais et al.,
109 2012; Breschi et al., 2016; Carelli et al., 2016; Chen et al., 2019; Cortez et al., 2014;
110 Guschanski et al., 2017; Julien et al., 2012; Necsulea et al., 2014; Warnefors and
111 Kaessmann, 2013) often used the same dataset containing 131 RNA-seq experiments
112 from 6 organs and 10 species (Brawand et al., 2011), with some additional data in
113 different studies. RNA-seq reads were first mapped to corresponding reference genomes
114 and then the expression level was quantified by two metrics: transcripts per million

115 (TPM) and fragments per kilobase million normalized by trimmed mean of M-values
116 (Robinson and Oshlack, 2010) (TMM-FPKM). To reduce the among-species variation,
117 the TMM normalization was applied across all 1,903 samples using the 1,377 single-copy
118 orthologs.

119 To allow rapid integrated analysis of datasets, we employed automated multi-aspect
120 quality controls, including metadata curation (Table S3), sequence read filtering (Fig. S2),
121 and iterative removal of anomalous RNA-seq samples by monitoring correlations
122 between and within data categories (Fig. 1A; Fig. S3). The metadata curation enabled us
123 to select appropriate samples from the NCBI SRA database. Data that were not
124 compatible with those from other research projects (defined by BioProject ID) were
125 removed in the correlation analysis by implementing a majority rule (Table S3), resulting
126 in a cleaned dataset. This filtering step was designed to fulfill the assumption that any
127 samples from the same organ should correlate better than samples from different organs
128 within species. When anomalous data were detected, all samples belonging to the same
129 research projects (i.e., the same BioProject ID) were discarded.

130 Finally, we applied surrogate variable analysis (SVA) (Leek and Storey, 2007) to
131 detect and correct hidden biases likely originating from heterogeneous sampling
132 conditions and sequencing procedures among experiments in both $\log_2$-scaled TPM and
133 TMM-FPKM (SVA-log-TPM and SVA-log-TMM-FPKM, respectively; Fig. S4A–B).
134 This correction greatly improved the correlation of expression levels within organs from
135 the same species, even when data were derived from different research projects (Fig. 1B–
136 C; Fig. S4C–F). Among surrogate variables, BioProject IDs tend to show a high
137 predictive power, suggesting project-specific sources of bias (Fig. 1D; Fig. S4G–H).
138 Although the inclusion of many species from phylogenetically diverse lineages makes it
139 difficult to extract organ-wise characteristics from the limited number of single-copy
140 orthologs, a principal component analysis produced moderate organ-wise segregation in
141 the multispecies comparison (Fig. 1E; Fig. S4I–K), further indicating that the curated
142 dataset is sufficiently reliable for use in cross-species expression pattern analyses. The
143 previously-reported uniqueness of testis transcriptomes (Brawand et al., 2011) was partly
144 resolved as the third principal component (Fig. S4K).

145 In addition to the better phylogenetic coverage, greater accuracy of estimated
146 expression levels is another possible advantage of integrating many RNA-seq datasets.
147 This idea is supported by subsampling analysis on a housekeeping gene GAPDH

148 (glyceraldehyde-3-phosphate dehydrogenase), where, as more data are used, estimated
149 expression levels tend to quickly converge to a final value (Fig. 1F; Fig. S4L).
150
151 **Modeling expression evolution.** We next used the amalgamated transcriptomes to
152 evaluate how expression evolved along 15,280 maximum-likelihood gene family
153 phylogenies (Table S4), employing multi-optima Ornstein–Uhlenbeck (OU) models
154 (Hansen, 1997) to allow for possible adaptive shifts of optimal expression levels and
155 neutral fluctuations (Brawand et al., 2011; Chen et al., 2019; Rohlfs et al., 2014). This
156 modeling identified statistically supported expression 'regime shifts' (Hansen, 1997;
157 Khabbazian et al., 2016) on each gene tree (Fig. 2A), which were then analyzed in the
158 context of preceding duplication events. 'Speciation' events (S node; Fig. 2B) with no
159 duplication were considered the 'baseline' mode of expression evolution because
160 regulatory environments and expression patterns are more preserved among orthologous
161 genes in comparison with paralogous genes produced by gene duplication (Altenhoff et
162 al., 2012; Castillo-Davis et al., 2004; Chen and Zhang, 2012; Rogozin et al., 2014).
163 Because OU shift detection has been applied for gene expression by assuming species
164 tree phylogeny in single-copy genes, shifts in S branches are equivalent to those
165 characterized previously (Brawand et al., 2011; Chen et al., 2019) but also includes many
166 more speciation events in duplication-prone gene families. Gene tree nodes associated
167 with preceding duplication events were categorized as DNA-based duplication or
168 retrotransposition (D or R nodes, respectively) depending on intron losses (Fig. 2B).
169 Organ-wise means of the two expression values, SVA-log-TPM and SVA-log-
170 TMM-FPKM, were separately used to model expression evolution with OU processes.
171 The two analyses resulted in similar numbers and characteristics of expression regime
172 shifts (Fig. S5), but the shift locations were sometimes inconsistent. S branches were a
173 major source of apparently inconsistent regime shifts, whereas branches following
174 duplication (D and R) showed largely consistent detection between the two metrics (Fig.
175 2C). Although the inclusion of inconsistently detected branches did not change the results,
176 we retained only consistently detected regime shifts for all downstream analysis to draw
177 a more robust conclusion. While SVA-log-TMM-FPKM values were reported in the main
178 text unless otherwise mentioned, the comparisons with SVA-log-TPM-based analyses are
179 available as Supplementary Information.

180 As an example of this analysis, an orthogroup of phosphoglycerol kinases (PGKs),
181 containing all three categories of branching events followed by expression shifts (S, D,
182 and R), is shown in Fig. 2D. This protein catalyzes the first ATP-generating step in the
183 glycolytic pathway and is required for most cell types including sperms (Danshina et al.,
184 2010; Liu et al., 2015). PGK1, the original copy on the X chromosome, is known to have
185 duplicated independently in eutherians and marsupials to produce the autosomal
186 retrocopy PGK2 that compensates the protein activity during X-inactivation (Boer et al.,
187 1987; McCarrey and Thomas, 1987; Potrzebowski et al., 2008). Our automated analysis
188 correctly recovered both retrotranspositions as well as the subsequent gains of testis-
189 specific expression in eutherian and marsupial lineages. This illustrates that our
190 automated genome-wide analysis can recover evolutionary trajectories that are
191 compatible with focused single gene family analyses (see Supplementary Data for
192 individual gene trees).
193
194 **Duplication-specific effects in expression evolution.** Across gene trees, expression
195 regime shifts were relatively rare in S branches, for a probability of 2.2% per branch, and
196 an average rate of 5.1×10^{-4} shifts/MY (million years) (Fig. 3A–B). In agreement with
197 the idea that gene duplication tends to free genes up for functional divergence and
198 enhance long-term retention of duplicated copies (Assis and Bachtrog, 2015; Huerta-
199 Cepas et al., 2011), the frequency of regime shifts in D branches was four times as much
200 per branch (9.0% per branch, at a rate of 1.9×10^{-3} shifts/MY across all genes and all D
201 branches). Thus, although far fewer branches are preceded by DNA-based duplication
202 events (72,476 branches) than speciation events (556,240 branches), D branches account
203 for over 33% of all regime shifts consistently detected by the two expression measures.
204 We note that this result reinforces previous results on the "ortholog conjecture", the idea
205 that duplicated gene copies (paralogs) are more prone to expression shifts than orthologs
206 (Altenhoff et al., 2012; Castillo-Davis et al., 2004; Chen and Zhang, 2012; Rogozin et al.,
207 2014). R branches were far more likely to result in expression regime shifts (37.3% per
208 branch, at a rate of 7.6×10^{-3} shifts/MY), but with only 2,985 R branches, this resulted in
209 only 1,107 shifts (5.6% of the total). Translocated genes are more likely to shift
210 expression than those that do not (Fig. S6), in line with previous observations from the
211 human genome (Lan and Pritchard, 2016). While the expression shift frequency in S and
212 D branches varies slightly across the phylogeny, R branches showed much stronger

213 among-lineage heterogeneity, and had a particularly high frequency in the mammalian
214 lineage (Fig. 3B). These retrotransposition-related expression changes may be related to
215 the variation in the retrotransposition rate itself, which is known to vary across lineages
216 (Marques et al., 2005; Yu et al., 2007). Among-species heterogeneity in gene prediction
217 quality may also be attributed to this pattern. In the absence of regime shifts, expression
218 levels varied most in ovary and testes, which had significantly higher average stationary
219 variances than the other four organs on the basis of tree-wise stationary variance (Fig.
220 3C). This supports previously observed high variation of gene expression in testes
221 (Brawand et al., 2011) and extends it to the reproductive organs of both sexes.
222 If expression regime shifts are due to functional divergence, it is highly relevant to
223 characterize how expression properties changed from ancestral to derived regimes. To do
224 this, we examined changes in organ expression specificity, maximum expression level,
225 and organ expression complementarity. The specificity measure τ ranges from 0 for
226 uniformly expressed genes to 1 for genes with highly specific expression (Yanai et al.,
227 2005). The distributions of regime shifts in D and R branches are shifted towards greater
228 organ specificity compared to shifts in S branches, with R branches creating the most
229 specific expression (Fig. 3D). To characterize the 'on state' transcriptional activity, we
230 analyzed the maximum fitted expression levels among the six organs (μ_{max}). D and R
231 branches appear to be enriched for down-regulation compared to S branches (Fig. 3E).
232 Complementarity of organ expression patterns was measured to evaluate the
233 differentiation between a pair of sister branches. We used a metric on the fitted organ-
234 wise expression levels (μ) called 'TEC', which ranges from 0 for completely overlapping
235 expression to 1 for mutually exclusive patterns (Huerta-Cepas et al., 2011), Nearly all
236 branches with regime shifts (95%) had complementarity values greater than 0.5,
237 indicating that most regime shifts detected involve differentiation of expression patterns,
238 rather than overall up- or down-regulation. Regime shifts in D and R branches often had
239 more complementarity expression than those in S branches (Fig. 3F), further supporting
240 the role of gene duplication in functional differentiation. The more drastic effect in R
241 branches probably reflects the regular loss of regulatory elements in retrotranspositions,
242 whereas DNA-based duplication can more often retain regulatory regions (Babushok et
243 al., 2007). Jointly, these results indicate that, compared with the baseline from speciation-
244 associated shifts, gene duplication tend to produce more organ-specific, more often down-
245 regulated, and more differentiated expression patterns. Although the down regulation

246 may be explainable by a tendency to need less of the newly functional expression regime,
247 it may also be explained by either recent nonfunctionalization (Balakirev and Ayala,
248 2003; Mighell et al., 2000) or specialized expression in organs that were not part of this
249 analysis.
250
251 **Context-dependent change in the rate of protein evolution.** Change in gene expression
252 can be accompanied by accelerated or decelerated protein evolution, which may be
253 detected by change in the ratio of nonsynonymous/synonymous substitutions (*dN/dS* or
254 ω) along branches. In D and R branches, median ω values are more than double the
255 baseline seen in S branches (Fig. 4A; Fig. S7A), again supporting the ortholog conjecture
256 and the idea that gene duplication tends to free genes up for functional divergence (Assis
257 and Bachtrog, 2015; Huerta-Cepas et al., 2011). Within each of S, D, and R branch
258 categories, branches with expression regime shifts accompany an increased ω compared
259 to sister branches (Fig. 4A). The increased rate was quite small in S branches (median ω,
260 0.096 in branches with shifts versus 0.102 in sister branches), much bigger in D branches
261 (0.182 versus 0.244), and huge in R branches (0.036 versus 0.394). If the changes in
262 expression and the rate of protein evolution are due to functional changes, this may
263 indicate that functional divergence is sometimes effected by joint changes in expression
264 and accelerated protein evolution.
265 Although 52.3% of all branches with expression regime shifts had higher ω relative
266 to sister branches (ω ratio), 27.5% of sister branch pairs are relatively undifferentiated,
267 with differences in ω within ±5%, and 42.5% had lower ω in the branches with
268 expression shifts. This finding led us to hypothesize that the direction of the rate changes
269 in protein evolution is linked to how, rather than whether, expression is changed.
270 Strikingly, we found a context-dependent association between protein and expression
271 evolution (Fig. 4B; Fig. S7B). Increased ω ratio was linked strongly to increased, rather
272 than decreased, organ expression specificity in S and D branches, potentially reflecting
273 adaptive evolution coupled with specialized expression. However, it was in turn highly
274 associated with decreased specificity in R branches, which may be explained by frequent
275 gene decay in unsuccessful retro-copies. The change in maximum expression level was
276 overall negatively correlated with ω ratio, but this link was stronger in down-regulation
277 compared with up-regulation, except for D branches where the ω ratio was smaller when

278 $\Delta \mu_{max}$ was larger (Fig. 4B). It has been reported that high expression slows protein
279 evolution (Zhang and Li, 2004), and our results suggest that DNA-based duplication
280 creates such constraints when accompanied by up-regulation. The organ expression
281 complementarity between sister lineages was positively correlated with ω ratio (Fig. 4B),
282 and its association was strongest in R branches, suggesting that protein evolution
283 accelerates as gene expression patterns differentiate from their ancestral state.
284 Collectively, these results suggest that protein evolution rate is linked to changes in
285 expression properties through a complicated association, which masks their relationships
286 in a global, unstratified analysis, and potentially explain a previous report of no strong
287 relation (Warnefors and Kaessmann, 2013).
288
289 **Organ-specific propensity in gene expression evolution.** The preadaptation hypothesis
290 predicts that the ancestral organ expression prior to the shifts will affect which organs are
291 likely to become the target of newly specific expression. To assess this, we tested whether
292 expression shifts are random with respect to change in expression from one organ to
293 another, by characterizing the organ in which genes are most highly expressed (PEO,
294 primary-expressed organ).
295 Across vertebrates, switching from one PEO to another was detected in 6,886, 3,586,
296 and 746 regime shifts in S, D, and R branches, respectively. The gain/loss ratios are
297 heterogeneous among organs (Fig. 5A), suggesting that vertebrate organs serve as both
298 sources and sinks in expression evolution, but that their relative contributions are organ-
299 specific. Although S, D, and R branches shared a global trend of relatively abundant
300 testis-related PEO shifts, their distributions are largely different. D branches were
301 moderately similar to both S and R branches (Spearman's $\rho \sim 0.6$), but S and R branches
302 were dissimilar ($\rho = 0.28$). This result suggests a role for gene duplication, including by
303 retrotranspositions, in remodeling the among-organ flow of expressed genes.
304 Controlling the total number of shifts from and to each PEO, some PEO shifts are
305 significantly different from the random expectation (Fig. 5B). There are clear patterns of
306 evolutionary transitions that are statistically supported by independent OU modeling of
307 the two expression metrics. In S branches, the pairs of brain–testis and testis–ovary
308 showed strong connections, indicating a solid exchange module. Kidney and liver also
309 donate genes to one another, forming a separate module from brain–testis–ovary. D and
310 R branches showed a pronounced acceleration of PEO shifts between testis and ovary.

311 PEO shifts in S and D branches were moderately symmetric in the flow between pairs of
312 organs (Fig. 5C), meaning two organs tend to donate comparable numbers of expressed
313 genes each other. In contrast, R branches were more asymmetric. The lower symmetry
314 may have perturbed the evolutionary dynamics of gene expression.

315 To check the robustness of our analysis, we analyzed high-confidence subsets of
316 expression regime shifts. The most drastic expression changes were characterized by
317 introducing a cutoff of organ expression specificity ($\tau > 0.5$) to define organ-specific
318 genes. Although some previously significant trends were not recovered due to small
319 sample sizes, the result was largely consistent with the broader analysis (Fig. S8). Because
320 tree inference errors can bias the downstream analysis including OU modeling, we also
321 analyzed expression regime shifts found in clades which have a high support in tree
322 inference (>99% bootstrap support). Again, the results were largely consistent (Fig. S8).
323 Especially, the brain–testis–ovary and kidney–liver modules in S branches and the testis–
324 ovary connection in D and R branches were always reproduced in the analyses with above
325 thresholds in combinations with the two expression metrics (SVA-log-TMM-FPKM and
326 SVA-log-TPM; Fig. S8), demonstrating the robustness of our conclusion.

327 To obtain insight into the biological relevance of the among-organ modules, we
328 characterized human genes involved in PEO shifts from all branch categories using
329 the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis.
330 The genes descended from the testis–ovary PEO shifts enriched only one KEGG pathway
331 term "Cell cycle" (Table S5; adjusted P value < 0.05), likely reflecting their function in
332 meiosis. Although the adjusted P value was not statistically significant, it is noteworthy
333 that the top-ranked term for the brain–testis connection was "Endocrine and other factor-
334 regulated calcium reabsorption" annotated to four genes including GNAQ, which has
335 been implicated to tumor formation in neuronal tissues (Gessi et al., 2013; Van
336 Raamsdonk et al., 2009). In the kidney–liver module, 15 terms were significantly
337 enriched, many of which appear to be related to the functions and diseases in those organs,
338 for example, "Bile secretion", "Phagosome", "Lysosome", "ABC transporters",
339 "Sphingolipid metabolism", and "Type I diabetes mellitus" (Table S5). These results
340 suggest that the among-organ modules in the PEO shifts played a role in supplying
341 functionally important genes.

342 Because ancestral expression was shown to orient new expression by the analysis of
343 PEO shifts, we concluded that there was prevalent organ-specific propensity, which
344 supports the presence of preadaptation in gene expression.
345
346 **Discussion:**
347 Our results suggest that the landscape of expression evolution is strongly shaped by
348 mechanisms of gene birth. Patterns of PEO shifts depend on ancestral states, suggesting
349 preadaptation in the evolution of vertebrate gene expression. However, the mechanisms
350 responsible for the propensity are unknown. Part of the propensity can be explained by
351 the "out of the testis" hypothesis, which explains the accelerated gains of testis expression
352 based on the permissive chromatin state, abundant transcriptional machinery, relatively
353 simple promoters required for the expression in spermatogenic cells and following gains
354 of new expression patterns (Kaessmann, 2010; Kleene, 2005). This theory fits to the
355 accelerated testis-related PEO shifts, but the other detected patterns (e.g., kidney–liver
356 module) require a different view.
357 One potential cause is a 'regulatory' mechanism whereby frequent gene-exchanging
358 organs use similar sets of regulatory elements. Altered expression between such organs
359 could occur with relatively few mutations in regulatory sequences. *Cis*-regulation is
360 indeed a major source of expression evolution, as it explains a certain fraction of
361 expression variability, for example, in budding yeast (30% in duplicates and 19% in
362 singletons) and undergoes a more rapid divergence than *trans*-regulation (Dong et al.,
363 2011). Another hypothetical 'coding' mechanism is that frequently interacting organs
364 may have similar environmental requirements for expressed proteins, such that few amino
365 acid substitutions tend to be required to adjust biochemical properties. Protein reusability
366 may be determined by cellular environments such as pH and temperature or by functional
367 categories of proteins. Our analysis indicates that the regime shifts that drastically
368 differentiate the expression tend to be coupled with accelerated protein evolution (Fig.
369 4B), and this result can be viewed as a support for the act of 'coding' mechanism. The
370 'regulatory' and 'coding' hypotheses are not mutually exclusive, and both factors may
371 contribute to generate preadaptive patterns of gene expression.
372 In this study, we established a method to standardize RNA-seq data from disparate
373 research projects and developed a pathway for data amalgamation. By reconstructing
374 gene expression in tailored gene family phylogenies, our analysis revealed non-

375 randomness and directionality of expression evolution. This suggests prevalent
376 preadaptation in gene expression, that adaptation to expression in certain organs is more
377 conducive to future expression in other organs, and shows how gene duplication helped
378 to reshape the dynamics of expression evolution that contributed to the vertebrate
379 diversification.
380

381 **Materials and Methods:**

382

383 **Species selection.** A total of 105 species included in the Ensembl release 91 (Yates et al.,
384 2016) were searched for data availability in the NCBI SRA database (Leinonen et al.,
385 2011) (final search on May 1, 2018) and 22 species were found to have RNA-seq data for
386 6 organs: brain, heart, kidney, liver, ovary, and testis. *Lepisosteus oculatus* was excluded
387 due to an insufficient quality of available expression data, and therefore remaining 21
388 species were selected for further analysis.

389

390 **Species tree.** The dated species tree for the 21 species was retrieved from TimeTree
391 (Hedges et al., 2006) (downloaded on March 15, 2018; Supplementary Data). Some
392 species were unavailable in the database and therefore they were temporarily replaced by
393 closely related species to obtain the dated species tree.

394

395 **Transcriptome metadata curation.** We developed an automated python program for
396 SRA metadata curation (Table S3; Supplementary Data). RNA-seq data were selected
397 from the NCBI SRA database by keyword searches limited to the 21 species, the 6 organs,
398 and Illumina sequencing platforms. Orthographical variants of annotations were
399 standardized with keyword libraries created by manually checking the original
400 annotations. Prenatal or unhealthy samples and small-scale sequencing samples (<5
401 million reads) were excluded. Data for non-messenger RNA sequencings were also
402 removed. In treatment and control RNA-seq pairs, only control experiments were
403 included.

404

405 **Transcriptome quantification.** Fastq files were extracted from downloaded SRA files
406 using parallel-fastq-dump 0.6.2 (https://github.com/rvalieris/parallel-fastq-dump) with
407 the minimum read length of 25 bp and the quality filter (-E option) (Leinonen et al., 2011).
408 The fastq sequences were then subjected to a quality filtering by fastp 0.12.3 (Chen et al.,
409 2018). The filtered reads were mapped to genomic features annotated as non-messenger
410 RNAs in the Ensembl GTF files using bowtie2 2.3.4 (Langmead and Salzberg, 2012) and
411 resultant unmapped reads were used for expression level quantification using kallisto
412 0.43.1 with the sequence-based bias correction (Bray et al., 2016). Samples were removed
413 if 20% or smaller percentages of reads were mappable (Fig. S2). Estimated mapped read

414 counts and transcript lengths were used to calculate TPM and FPKM values. For the latter,
415 the "trimmed mean of M-values (TMM)" normalization method was applied (Robinson
416 and Oshlack, 2010). Sample-wise TMM scaling factors were obtained across all RNA-
417 seq samples using the FPKM values of 1,377 single-copy orthologs. Because TMM and
418 TPM transformations are incompatible with each other, the TMM scaling was applied
419 only to FPKM values. TPM and TMM-FPKM values were subsequently transformed to
420 $log(N + 1)$ values (log-TPM and log-TMM-FPKM, respectively). Paralogous genes that
421 haven't diverged in their nucleotide sequences could not be distinguished well in the
422 quantification step. Although our scope is to characterize gene expression in the timescale
423 of vertebrate evolution, this difficulty likely leads to an underestimation of expression
424 regime shifts in young duplicates.
425
426 **Iterative anomalous sample removal coupled with surrogate variable analysis.**
427 Anomalous RNA-seq samples were iteratively removed by a correlation analysis coupled
428 with an expression bias correction. In each iteration, unwanted biases in expression level
429 were removed by surrogate variable analysis (SVA, 'sva' function in an R package 'sva')
430 (Leek and Storey, 2007). SVA analysis was applied at the beginning of each iteration so
431 that it is not influenced by anomalous samples removed in previous iterations.
432 Subsequently, Pearson's correlation coefficients were calculated for every RNA-seq data
433 against mean expression level in each organ generated by averaging all other data
434 excluding those from the same BioProject (Fig. S3). We assume that the sample's
435 correlation coefficient against the same organ is higher than any of the values against the
436 other organs, and we removed all samples from the same BioProject when violations were
437 found. These steps were repeated until no violations were left and SVA-corrected
438 expression levels were finally reported (SVA-log-TMM-FPKM and SVA-log-TPM). The
439 curation steps were skipped if multiple samples were unavailable in the species and hence
440 SVA analysis was inapplicable. The final dataset was comprised of 1,903 RNA-seq
441 experiments from 182 BioProjects that cover 6 organs from 21 vertebrate species without
442 missing data (Tables S1–S2).
443
444 **Orthogroup classification.** Orthogroups, which contain all genes descended from one
445 gene in the common ancestor, were inferred from coding sequences of the 21 species
446 using OrthoFinder 2.1.2 (Emms and Kelly, 2015) guided by the species tree. In total,

447 17,896 orthogroups were generated. The largest orthogroup, which is comprised of 7,893
448 olfactory receptor genes, was removed from the analysis because of computational
449 burden. After sequence alignment processing (see "Multiple sequence alignment"), we
450 removed small orthogroups which retained less than four genes and orthogroups that
451 showed no parsimony informative sites, because phylogenetic relationships cannot be
452 inferred. As the result of filtering, 15,280 orthogroups were left for OU modeling, with
453 the largest one containing 3,796 zinc-finger proteins (Table S4).
454
455 **Multiple sequence alignment.** Multiple fasta files containing coding sequences were
456 generated for each orthogroup. Stops and ambiguous codons were masked as gaps (for
457 implementation, see https://github.com/kfuku52/cdskit). In-frame multiple codon
458 sequence alignments were generated using MAFFT 7.394 with the 'auto' option (Katoh
459 and Standley, 2013) and tranalign in EMBOSS 6.5.7.0 (Rice et al., 2000). Anomalous
460 genes were excluded by MaxAlign (Gouveia-Oliveira et al., 2007) which decreased the
461 largest orthogroup size from 3,796 to 2,382 genes. Spurious codons were removed in-
462 frame using pgtrimal in Phylogears2-2.0.2016.09.06
463 (https://www.fifthdimension.jp/products/phylogears/) with the 'gappyout' option
464 (Capella-Gutierrez et al., 2009).
465
466 **Gene tree reconstruction.** Maximum-likelihood trees were reconstructed using IQ-
467 TREE 1.6.5 (Nguyen et al., 2015) with the best-fit nucleotide substitution models selected
468 by ModelFinder with the Bayesian Information Criterion (Kalyaanamoorthy et al., 2017).
469 Larger orthogroups and longer genes tended to fit more complex substitution matrices
470 and larger numbers of categories of rate heterogeneity (Fig. S9A–B; Table S4). Ultrafast
471 bootstrapping with 1,000 replicates was performed to evaluate the credibility of tree
472 topology (Minh et al., 2013) with a further optimization of each bootstrapping tree (-bnni
473 option) (Hoang et al., 2018).
474
475 **Reconciliation-assisted gene tree rooting.** Candidate rooting positions were inferred
476 with different methods. Using the dated species tree, all rooting branches with the
477 minimum duplication-loss score were identified using the rooting mode of NOTUNG 2.9
478 with the default parameters (duplication score = 1.5, loss score = 1.0) (Chen et al., 2000).
479 The midpoint of the longest path (Farris, 1972) and the position with the minimal ancestor

480 deviations (MAD) (Tria et al., 2017) were also considered as candidates. The final rooting
481 position was reported based on overlaps among those rooting positions (Figs. S9C,E;
482 Table S4).
483
484 **Reconciliation-assisted divergence time estimation.** To prepare dated gene trees, we
485 first matched species tree nodes with corresponding gene tree nodes using the
486 reconciliation mode of NOTUNG 2.9 (Chen et al., 2000) and created time constraints of
487 speciation nodes (Fig. S9F). Duplication nodes were constrained with the upper and lower
488 age limits derived from corresponding speciation nodes. If the root node is a duplication
489 node and is not covered by the range of the species tree, the upper age limit was set to
490 1,105 million years ago, which corresponds to the split of animals and fungi (Hedges et
491 al., 2006). Divergence time was then estimated by a penalized likelihood method
492 (Sanderson, 2002) implemented in an R package 'ape' ('chronos' function with 'discrete'
493 model) (Popescu et al., 2012) with time constraints on speciation, duplication, and root
494 nodes. When reasonable initial parameters were not found after 1,000 trials, the above
495 constraints were partly relaxed (Fig. S9D; Table S4). The implementation is provided on
496 GitHub (https://github.com/kfuku52/RADTE).
497
498 **Modeling and shift detection of expression evolution.** Using the dated gene trees and
499 organ-wise mean values of SVA-log-TMM-FPKM and SVA-log-TPM, regime shifts in
500 gene expression were detected as shifts of optimal trait values in Ornstein–Uhlenbeck
501 (OU) models determined by a Lasso-based model selection with AICc in an R package
502 '*l*1ou' (Khabbazian et al., 2016). Because there is no available software to handle within-
503 species variation in phylogenetic OU shift detection, we used mean expression level as
504 the input. In the model, α and σ^2 parameters were assumed unchanged in the tree
505 (Khabbazian et al., 2016), and therefore only the global, rather than branch- or clade-wise,
506 stationary variance (γ) were obtained. Expression levels in the six organs were treated as
507 multivariate traits where α and σ^2 were estimated for each organ but regime shifts were
508 assumed to occur jointly in the same set of branches (Khabbazian et al., 2016). The
509 phylogenetic mean (expression level at the root node) was estimated with the
510 "OUfixedRoot" model. To handle gene trees recalcitrant to this analysis (especially those
511 with a large number of genes), we skimmed gene trees by collapsing clades with small

512 changes in expression level (Fig. S10). Specifically, we first calculated all-vs-all
513 Pearson's correlation coefficients of gene expression level among all genes that belong
514 the clade. The clade was collapsed into a single tip if the expression patterns were almost
515 identical (minimum correlation coefficient between genes > 0.99). Phylogenetic means
516 of the collapsed clade were calculated by assuming the Brownian motion and were used
517 as expression level at the new tip. The upper limit of regime shifts was set as
518 $\max\left[\min(N/2, 100), \sqrt[2]{N}\right]$, except for the largest tree with 2,382 genes where the upper
519 limit was decreased to 10 to cope with an unrealistically large number of branch
520 combinations to consider. The number of detected regime shifts was always smaller than
521 the upper limit in the SVA-log-TMM-FPKM analysis, whereas three out of 15,820 trees
522 reached the upper limit in the SVA-log-TPM analysis (Fig. S5A).

523

524 **Analysis of expression patterns.** We characterized expression patterns of extant and
525 ancestral genes by calculating different metrics from fitted values (μ) in the OU models.
526 Organ specificity was measured by τ (Yanai et al., 2005), which outperformed other
527 methods in a benchmark for tissue specificity (Kryuchkova-Mostacci and Robinson-
528 Rechavi, 2016). Expression complementarity between sister lineages was measured by
529 the metrics called 'TEC' (Huerta-Cepas et al., 2011). Because μ was estimated from log-
530 transformed expression levels, these expression metrics were calculated with unlog-
531 transformed μ values. PEOs were defined as the organ in which the highest expression
532 levels were observed among the six organs we analyzed.

533

534 **Estimation of protein evolution rate.** Parameters for codon substitution matrix, shape
535 parameter of discrete gamma distribution for rate heterogeneity (α), equilibrium
536 transition/transversion rate ratio (κ), equilibrium nonsynonymous/synonymous
537 substitution rate ratio (ω) were estimated using IQ-TREE 1.6.5 (Nguyen et al., 2015) by
538 fitting GY+F3X4+G4 models to each gene tree. Equilibrium base composition (θ) was
539 estimated from empirical codon state frequencies which are calculated from the alignment
540 by counting. To obtain θ at the root node, we calculated θ at subroot nodes by taking
541 advantage of IQ-TREE's empirical Bayesian method for ancestral sequence
542 reconstruction. Considering subroot branch length, a weighted average of the subroot θ

543 values were calculated as the θ value at the root node. Using all those parameters, branch-
544 wise nonsynonymous/synonymous substitution rate ratios were estimated by stochastic
545 substitution mapping (*mapdNdS*) (Guéguen and Duret, 2018) using bio++ library
546 (Guéguen et al., 2013). To examine the robustness of the *mapdNdS*-based ω estimation,
547 we compared the results with those obtained by maximum-likelihood ω estimation by
548 fitting MG94W9 models in HyPhy 2.3.11 (Pond et al., 2005). The two methods yielded
549 consistent results on the effect of branching events and expression shifts (Fig. 4A; Fig.
550 S7A), suggesting methodological robustness. We reported *mapdNdS*-based results in the
551 main text.
552
553 **Analysis of gene structure and location.** The number of introns and chromosomal
554 location were obtained for each gene from the Ensembl gene models (GFF3 files). The
555 intron numbers were subsequently converted to binary values that represent intronless
556 and intron-containing states. Chromosomal locations were categorized into autosome, X
557 chromosome, and Y chromosome. Genes from non-therian species were treated as
558 missing data because mechanisms of their sex determination are not homologous to the
559 mammalian XY system (Veyrunes et al., 2008). Genes from *Chinchilla lanigera* were
560 also treated as missing data because their sequenced genomes are not anchored to
561 chromosomes in the Ensembl release 91. The posterior probabilities of ancestral character
562 states were inferred by the stochastic character mapping of discrete traits (Bollback, 2006)
563 implemented in an R package 'phytools' (Revell, 2012). Because functional
564 retrotranspositions (Jun et al., 2009; Marques et al., 2005) and inter-chromosomal
565 duplications (Pace et al., 2009) are rare events relative to the timescale of the vertebrate
566 evolution on the per-gene basis, we set the transition rate parameters to a sufficiently
567 small value (1×10^{-3} per gene per million years). Since intron gain occurs few orders of
568 magnitude less frequently than its loss caused by retrotransposition and others (Roy and
569 Gilbert, 2005), the rate of intron gain is set to be lower (1×10^{-4} per gene per million
570 years).
571
572 **Analysis of branching events.** Speciation and duplication nodes (S and D/R nodes,
573 respectively) were classified by a species-overlap method (Huerta-Cepas et al., 2007) and
574 were mapped to the species tree on the basis of species coverages of the gene tree clades.
575 A transition from intron-containing to intronless states with a posterior probability greater

than 0.5 was classified as a retrotransposition event (R node). The branches that correspond to the original copy of a retrotransposition event were not included in R branches. Although our classification cannot detect retrotranspositions from originally intronless genes, we expect such situations would be rare because most vertebrate genes contain at least one intron (e.g., 20,160/21,242 human genes). Inter-chromosomal translocation was detected by considering chromosomal locations with the highest posterior probability as the ancestral states. Because of the difficulty in determining rooting positions of deep phylogenies, gene tree nodes older than the root node of the species tree were removed from the analysis.

KEGG pathway enrichment analysis. Human genes that descend from the shift branches were pooled for each specific PEO shift. The gene lists were examined for enrichment against an Enrichr library "KEGG_2019_Human" (Kuleshov et al., 2016) using a python package GSEApy (https://github.com/zqfang/GSEApy). Statistically significant (adjusted P value < 0.05) KEGG pathway terms were reported in Table S5.

Data visualization. Phylogenetic trees were visualized using a python package 'ETE 3' (Huerta-Cepas et al., 2016) and an R package 'ggtree' (Yu et al., 2017). A part of animal silhouettes in Fig. 3A were obtained from PhyloPic (http://phylopic.org). Box plot elements of all figures are defined as follows: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers. Box-plot outliers are suppressed in Figs. 3, 4, S7, and S8.

Code availability. Scripts and parameter values used in this study are available as Supplementary Data (Mendeley Data ID available on publication).

Acknowledgements:

We acknowledge the following sources for funding: MEXT/JSPS KAKENHI 18J00178 (K.F.), Sofja Kovalevskaja programme by the Alexander von Humboldt Foundation (K.F.), and NIH R01 GM083127 (D.D.P.). Computations were partially performed on the NIG supercomputer. The authors declare no competing financial interests.

Author Contributions:

K.F. and D.D.P. jointly designed the study and wrote the paper. K.F. designed and wrote all programs and performed data analysis.

614 **References:**

615 Altenhoff, A.M., Studer, R.A., Robinson-Rechavi, M., Dessimoz, C., and Couto, F.
616 (2012). Resolving the ortholog conjecture: orthologs tend to be weakly, but
617 significantly, more similar in function than paralogs. PLoS Comput. Biol. *8*, e1002514.

618 Assis, R., and Bachtrog, D. (2015). Rapid divergence and diversification of mammalian
619 duplicate gene functions. BMC Evol. Biol. *15*, 138.

620 Babushok, D. V., Ostertag, E.M., and Kazazian, H.H. (2007). Current topics in genome
621 evolution: Molecular mechanisms of new gene formation. Cell. Mol. Life Sci. *64*, 542–
622 554.

623 Balakirev, E.S., and Ayala, F.J. (2003). Pseudogenes: Are they "junk" or functional
624 DNA? Annu. Rev. Genet. *37*, 123–151.

625 Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J.,
626 Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., et al. (2012). The evolutionary
627 landscape of alternative splicing in vertebrate species. Science *338*, 1587–1593.

628 Bedford, T., and Hartl, D.L. (2009). Optimization of gene expression by natural
629 selection. Proc. Natl. Acad. Sci. U. S. A. *106*, 1133–1138.

630 Boer, P.H., Adra, C.N., Lau, Y.F., and McBurney, M.W. (1987). The testis-specific
631 phosphoglycerate kinase gene pgk-2 is a recruited retroposon. Mol. Cell. Biol. *7*, 3107–
632 3112.

633 Bollback, J. (2006). SIMMAP: Stochastic character mapping of discrete traits on
634 phylogenies. BMC Bioinformatics *7*, 88.

635 Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier,
636 M., Liechti, A., Aximu-Petri, A., Kircher, M., et al. (2011). The evolution of gene
637 expression levels in mammalian organs. Nature *478*, 343–348.

638 Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic
639 RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

640 Breschi, A., Djebali, S., Gillis, J., Pervouchine, D.D., Dobin, A., Davis, C.A., Gingeras,
641 T.R., and Guigó, R. (2016). Gene-specific patterns of expression variation across organs
642 and species. Genome Biol. *17*, 151.

643 Brunner, E., and Munzel, U. (2000). The nonparametric Behrens-Fisher problem:
644 asymptotic theory and a small-sample approximation. Biom. J. *42*, 17–25.

645 Budd, G.E. (2007). On the origin and evolution of major morphological characters.
646 Biol. Rev. *81*, 609–628.

647 Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009). trimAl: a tool for
648 automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*,
649 1972–1973.

650 Carelli, F.N., Hayakawa, T., Go, Y., Imai, H., Warnefors, M., and Kaessmann, H.
651 (2016). The life history of retrocopies illuminates the evolution of new mammalian
652 genes. Genome Res. *26*, 301–314.

653 Castillo-Davis, C.I., Hartl, D.L., and Achaz, G. (2004). cis-Regulatory and protein
654 evolution in orthologous and duplicate genes. Genome Res. *14*, 1530–1536.

655 Chen, X., and Zhang, J. (2012). The ortholog conjecture is untestable by the current
656 gene ontology but is supported by RNA sequencing data. PLoS Comput. Biol. *8*,
657 e1002784.

658 Chen, J., Swofford, R., Johnson, J., Cummings, B.B., Rogel, N., Lindblad-Toh, K.,
659 Haerty, W., Palma, F. di, and Regev, A. (2019). A quantitative framework for
660 characterizing the evolutionary history of mammalian gene expression. Genome Res.
661 *29*, 53–63.

662 Chen, K., Durand, D., and Farach-Colton, M. (2000). NOTUNG: a program for dating
663 gene duplications and optimizing gene family trees. J. Comput. Biol. *7*, 429–447.

664 Chen, S., Krinsky, B.H., and Long, M. (2013). New genes as drivers of phenotypic
665 evolution. Nat. Rev. Genet. *14*, 645–660.

666 Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ
667 preprocessor. Bioinformatics *34*, i884–i890.

668 Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P.D.,
669 Grützner, F., and Kaessmann, H. (2014). Origins and functional evolution of Y
670 chromosomes across mammals. Nature *508*, 488–493.

671 Danshina, P. V., Geyer, C.B., Dai, Q., Goulding, E.H., Willis, W.D., Kitto, G.B.,
672 McCarrey, J.R., Eddy, E.M., and O'Brien, D.A. (2010). Phosphoglycerate kinase 2
673 (PGK2) is essential for sperm function and male fertility in mice. Biol. Reprod. *82*,
674 136–145.

675 de-Leon, S.B.-T., and Davidson, E.H. (2007). Gene regulation: gene control network in
676 development. Annu. Rev. Biophys. Biomol. Struct. *36*, 191–212.

677 Dong, D., Yuan, Z., and Zhang, Z. (2011). Evidences for increased expression variation
678 of duplicate genes in budding yeast: from *cis*- to *trans*- regulation effects. Nucleic Acids
679 Res. *39*, 837–847.

680 Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole
681 genome comparisons dramatically improves orthogroup inference accuracy. Genome
682 Biol. *16*, 157.

683 Farris, J.S. (1972). Estimating phylogenetic trees from distance matrices. Am. Nat. *106*,
684 645–668.

685 Gessi, M., Hammes, J., Lauriola, L., Dörner, E., Kirfel, J., Kristiansen, G., Muehlen, A.
686 zur, Denkhaus, D., Waha, A., and Pietsch, T. (2013). GNA11 and N-RAS mutations:
687 alternatives for MAPK pathway activating GNAQ mutations in primary melanocytic
688 tumours of the central nervous system. Neuropathol. Appl. Neurobiol. *39*, 417–425.

689 Gould, S.J., and Vrba, E.S. (1982). Exaptation—a missing term in the science of form.
690 Paleobiology *8*, 4–15.

691 Gouveia-Oliveira, R., Sackett, P.W., and Pedersen, A.G. (2007). MaxAlign:
692 maximizing usable data in an alignment. BMC Bioinformatics *8*, 312.

693 Guéguen, L., and Duret, L. (2018). Unbiased estimate of synonymous and
694 nonsynonymous substitution rates with nonstationary base composition. Mol. Biol.
695 Evol. *35*, 734–742.

696 Guéguen, L., Gaillard, S., Boussau, B., Gouy, M., Groussin, M., Rochette, N.C., Bigot,
697 T., Fournier, D., Pouyet, F., Cahais, V., et al. (2013). Bio++: efficient extensible
698 libraries and tools for computational molecular evolution. Mol. Biol. Evol. *30*, 1745–
699 1750.

700 Guschanski, K., Warnefors, M., and Kaessmann, H. (2017). The evolution of duplicate
701 gene expression in mammalian organs. Genome Res. *27*, 1461–1474.

702 Hansen, T.F. (1997). Stabilizing selection and the comparative analysis of adaptation.
703 Evolution *51*, 1341–1351.

704 Hedges, S.B., Dudley, J., and Kumar, S. (2006). TimeTree: a public knowledge-base of
705 divergence times among organisms. Bioinformatics *22*, 2971–2972.

706 Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018).
707 UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518–
708 522.

709 Huerta-Cepas, J., Dopazo, H., Dopazo, J., Gabaldón, T., McPherson, J., Marra, M.,
710 Hillier, L., Waterston, R., Chinwalla, A., Wallis, J., et al. (2007). The human phylome.
711 Genome Biol. *8*, 934–941.

712 Huerta-Cepas, J., Dopazo, J., Huynen, M.A., and Gabaldón, T. (2011). Evidence for
713 short-time divergence and long-time conservation of tissue-specific expression after
714 gene duplication. Brief. Bioinform. *12*, 442–448.

715 Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: Reconstruction, analysis, and
716 visualization of phylogenomic data. Mol. Biol. Evol. *33*, 1635–1638.

717 Julien, P., Brawand, D., Soumillon, M., Necsulea, A., Liechti, A., Schütz, F., Daish, T.,
718 Grützner, F., and Kaessmann, H. (2012). Mechanisms and evolutionary patterns of
719 mammalian and avian dosage compensation. PLoS Biol. *10*, e1001328.

720 Jun, J., Ryvkin, P., Hemphill, E., Mandoiu, I., and Nelson, C. (2009). The birth of new
721 genes by RNA- and DNA-mediated duplication during mammalian evolution. J.
722 Comput. Biol. *16*, 1429–1444.

723 Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes.
724 Genome Res. *20*, 1313–1326.

725 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S.
726 (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat.
727 Methods *14*, 587–589.

728 Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software
729 version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

730 Khabbazian, M., Kriebel, R., Rohe, K., and Ané, C. (2016). Fast and accurate detection
731 of evolutionary shifts in Ornstein-Uhlenbeck models. Methods Ecol. Evol. *7*, 811–824.

732 Kleene, K.C. (2005). Sexual selection, genetic conflict, selfish genes, and the atypical
733 patterns of gene expression in spermatogenic cells. Dev. Biol. *277*, 16–26.

734 Kryuchkova-Mostacci, N., and Robinson-Rechavi, M. (2016). A benchmark of gene
735 expression tissue-specificity metrics. Brief. Bioinform. *44*, bbw008.

736 Kryuchkova-Mostacci, N., Robinson-Rechavi, M., Shmoish, M., Chalifa-Caspi, V.,
737 Shklar, M., and Ophir, R. (2016). Tissue-specificity of gene expression diverges slowly
738 between orthologs, and rapidly between paralogs. PLOS Comput. Biol. *12*, e1005274.

739 Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z.,
740 Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a
741 comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids
742 Res. *44*, W90–W97.

743 Lan, X., and Pritchard, J.K. (2016). Coregulation of tandem duplicate genes slows
744 evolution of subfunctionalization in mammals. Science *352*, 1009–1013.

745 Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2.
746 Nat. Methods *9*, 357–359.

747 Leek, J.T., and Storey, J.D. (2007). Capturing heterogeneity in gene expression studies
748 by surrogate variable analysis. PLoS Genet. *3*, e161.

749 Leinonen, R., Sugawara, H., and Shumway, M. (2011). The Sequence Read Archive.
750 Nucleic Acids Res. *39*, D19–D21.

751 Liao, B.-Y., and Zhang, J. (2006). Low rates of expression profile divergence in highly
752 expressed genes and tissue-specific genes during mammalian evolution. Mol. Biol.
753 Evol. *23*, 1119–1128.

754 Liu, X.-X., Zhang, H., Shen, X.-F., Liu, F.-J., Liu, J., and Wang, W.-J. (2015).
755 Characteristics of testis-specific phosphoglycerate kinase 2 and its association with
756 human sperm quality. Hum. Reprod. *126*, dev301.

757 Marques, A.C., Dupanloup, I., Vinckenbosch, N., Reymond, A., and Kaessmann, H.
758 (2005). Emergence of young human genes after a burst of retroposition in primates.
759 PLoS Biol. *3*, e357.

760 McCarrey, J.R., and Thomas, K. (1987). Human testis-specific PGK gene lacks introns
761 and possesses characteristics of a processed gene. Nature *326*, 501–505.

762 Mighell, A.J., Smith, N.R., Robinson, P.A., and Markham, A.F. (2000). Vertebrate
763 pseudogenes. FEBS Lett. *468*, 109–114.

764 Minh, B.Q., Nguyen, M.A.T., and von Haeseler, A. (2013). Ultrafast approximation for
765 phylogenetic bootstrap. Mol. Biol. Evol. *30*, 1188–1195.

766 Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., Baker,
767 J.C., Grützner, F., and Kaessmann, H. (2014). The evolution of lncRNA repertoires and
768 expression patterns in tetrapods. Nature *505*, 635–640.

769 Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A
770 fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies.
771 Mol. Biol. Evol. *32*, 268–274.

772 Pace, J.K., Sen, S.K., Batzer, M.A., and Feschotte, C. (2009). Repair-mediated
773 duplication by capture of proximal chromosomal DNA has shaped vertebrate genome
774 evolution. PLoS Genet. *5*, e1000469.

775 Podgornaia, A.I., and Laub, M.T. (2015). Pervasive degeneracy and epistasis in a
776 protein-protein interface. Science *347*, 673–677.

777 Pond, S.L.K., Frost, S.D.W., and Muse, S. V. (2005). HyPhy: hypothesis testing using
778 phylogenies. Bioinformatics *21*, 676–679.

779 Popescu, A.-A., Huber, K.T., and Paradis, E. (2012). ape 3.0: New tools for distance-
780 based phylogenetics and evolutionary analysis in R. Bioinformatics *28*, 1536–1537.

781 Potrzebowski, L., Vinckenbosch, N., Marques, A.C., Chalmel, F., Jégou, B., and
782 Kaessmann, H. (2008). Chromosomal gene movements reflect the recent origin and
783 biology of therian sex chromosomes. PLoS Biol. *6*, e80.

784 Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and
785 other things). Methods Ecol. Evol. *3*, 217–223.

786 Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular
787 Biology Open Software Suite. Trends Genet. *16*, 276–277.

788 Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for
789 differential expression analysis of RNA-seq data. Genome Biol *11*, 2010–2011.

790 Rogozin, I.B., Managadze, D., Shabalina, S.A., and Koonin, E. V (2014). Gene family
791 level comparative analysis of gene expression in mammals validates the ortholog
792 conjecture. Genome Biol. Evol. *6*, 754–762.

793 Rohlfs, R. V., Harrigan, P., and Nielsen, R. (2014). Modeling Gene Expression
794 Evolution with an Extended Ornstein–Uhlenbeck Process Accounting for Within-
795 Species Variation. Mol. Biol. Evol. *31*, 201–211.

796 Roy, S.W., and Gilbert, W. (2005). Rates of intron loss and gain: implications for early
797 eukaryotic evolution. Proc. Natl. Acad. Sci. U. S. A. *102*, 5773–5778.

798 Sanderson, M.J. (2002). Estimating absolute rates of molecular evolution and
799 divergence times: a penalized likelihood approach. Mol. Biol. Evol. *19*, 101–109.

800 She, X., Rohl, C.A., Castle, J.C., Kulkarni, A. V, Johnson, J.M., and Chen, R. (2009).
801 Definition, conservation and epigenetics of housekeeping and tissue-enriched genes.
802 BMC Genomics *10*, 269.

803 Soubrier, J., Steel, M., Lee, M.S.Y., Der Sarkissian, C., Guindon, S., Ho, S.Y.W., and
804 Cooper, A. (2012). The influence of rate heterogeneity among sites on the time
805 dependence of molecular rates. Mol. Biol. Evol. *29*, 3345–3358.

806 Starr, T.N., Picton, L.K., and Thornton, J.W. (2017). Alternative evolutionary histories
807 in the sequence space of an ancient protein. Nature *549*, 409–413.

808 Tria, F.D.K., Landan, G., and Dagan, T. (2017). Phylogenetic rooting using minimal
809 ancestor deviation. Nat. Ecol. Evol. *1*, 0193.

810 Van Raamsdonk, C.D., Bezrookove, V., Green, G., Bauer, J., Gaugler, L., O'Brien,
811 J.M., Simpson, E.M., Barsh, G.S., and Bastian, B.C. (2009). Frequent somatic
812 mutations of GNAQ in uveal melanoma and blue naevi. Nature *457*, 599–602.

813 Veyrunes, F., Waters, P.D., Miethke, P., Rens, W., McMillan, D., Alsop, A.E.,
814 Grutzner, F., Deakin, J.E., Whittington, C.M., Schatzkamer, K., et al. (2008). Bird-like
815 sex chromosomes of platypus imply recent origin of mammal sex chromosomes.
816 Genome Res. *18*, 965–973.

817 Warnefors, M., and Kaessmann, H. (2013). Evolution of the Correlation between
818 Expression Divergence and Protein Divergence in Mammals. Genome Biol. Evol. *5*,
819 1324–1335.

820 Wilson, B.A., Foy, S.G., Neme, R., and Masel, J. (2017). Young genes are highly
821 disordered as predicted by the preadaptation hypothesis of de novo gene birth. Nat.
822 Ecol. Evol. *1*, 0146.

823 Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-
824 Even, A., Horn-Saban, S., Safran, M., Domany, E., et al. (2005). Genome-wide

825 midrange transcription profiles reveal expression level relationships in human tissue
826 specification. Bioinformatics *21*, 650–659.

827 Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences
828 with variable rates over sites: Approximate methods. J. Mol. Evol. *39*, 306–314.

829 Yang, Z. (1995). A space-time process model for the evolution of DNA sequences.
830 Genetics *139*, 993–1005.

831 Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D.,
832 Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., et al. (2016). Ensembl 2016. Nucleic
833 Acids Res. *44*, D710–D716.

834 Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.-Y. (2017). ggtree : an R package
835 for visualization and annotation of phylogenetic trees with their covariates and other
836 associated data. Methods Ecol. Evol. *8*, 28–36.

837 Yu, Z., Morais, D., Ivanga, M., and Harrison, P.M. (2007). Analysis of the role of
838 retrotransposition in gene evolution in vertebrates. BMC Bioinformatics *8*, 308.

839 Zhang, L., and Li, W.-H. (2004). Mammalian housekeeping genes evolve more slowly
840 than tissue-specific genes. Mol. Biol. Evol. *21*, 236–239.

841 Zhang, Y.E., and Long, M. (2014). New genes contribute to genetic and phenotypic
842 novelties in human evolution. Curr. Opin. Genet. Dev. *29*, 90–96.

843 Zhang, Y.E., Landback, P., Vibranovski, M.D., Long, M., and Stark, A. (2011).
844 Accelerated recruitment of new brain development genes into the human genome. PLoS
845 Biol. *9*, e1001179.

846

847

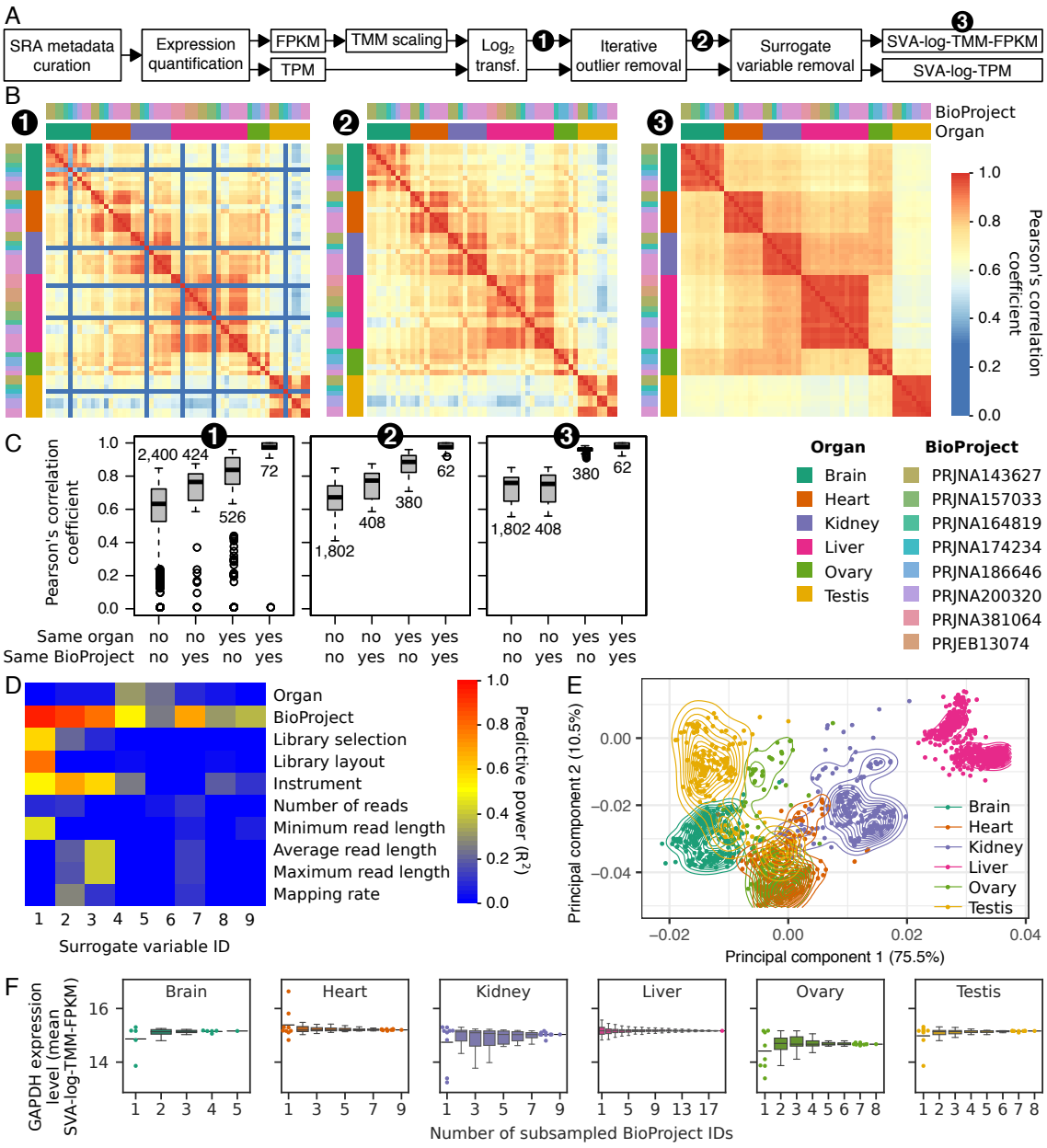


Figure 1. Transcriptome amalgamation to integrate heterogeneous RNA-seq samples. (**A**) A simplified flow chart of the transcriptome amalgamation. The full chart is available in Fig. S1. (**B–D**) Transcriptome curation within species. Data from *Monodelphis domestica* with SVA-log-TMM-FPKM metrics are shown as an example. The heatmaps show Pearson's correlation coefficients among RNA-seq samples (**B**). Each row and column correspond to one RNA-seq sample. The expression levels of all genes were used to calculate the correlation coefficients. Note that anomalous samples contaminated in the curated metadata (low correlation samples in 1) are successfully removed, and that project-specific correlations visible in the uncorrected data (marked 2)

858 are absent in the corrected data (marked 3). The box plots show distinct distributions of
859 the correlation coefficients depending on whether a pair of samples are the same organ or
860 whether they are from the same research project (**C**). The numbers of comparisons are
861 provided in the plot. The correlation coefficients are largely improved in within-organ
862 comparisons when surrogate variables are removed, while within-project biases are
863 attenuated. In this species, nine surrogate variables were detected against 52 RNA-seq
864 data from eight projects (**D**). Analysis of those variables by linear regression highlights
865 the BioProject feature as the strongest source of removed biases. For full description of
866 predictors, see Fig. S4. (**E**) A principal component analysis using expression levels of
867 1,377 single-copy orthologs from 21 species. Points correspond to RNA-seq samples.
868 Curves show the estimated kernel density. Explained variations in percentages are
869 indicated in each axis. (**F**) Estimated organ-wise expression levels of a housekeeping gene.
870 Since data from relatively many BioProjects are available, glyceraldehyde-3-phosphate
871 dehydrogenase gene (GAPDH, Ensembl gene ID: ENSGALG00000014442) in *Gallus*
872 *gallus* is shown as an example. Points corresponds to the average expression level
873 calculated by random subsampling. All data points and the median value (bar), rather than
874 a box plot, are shown if the number of subsampled BioProject combinations is less than
875 10.

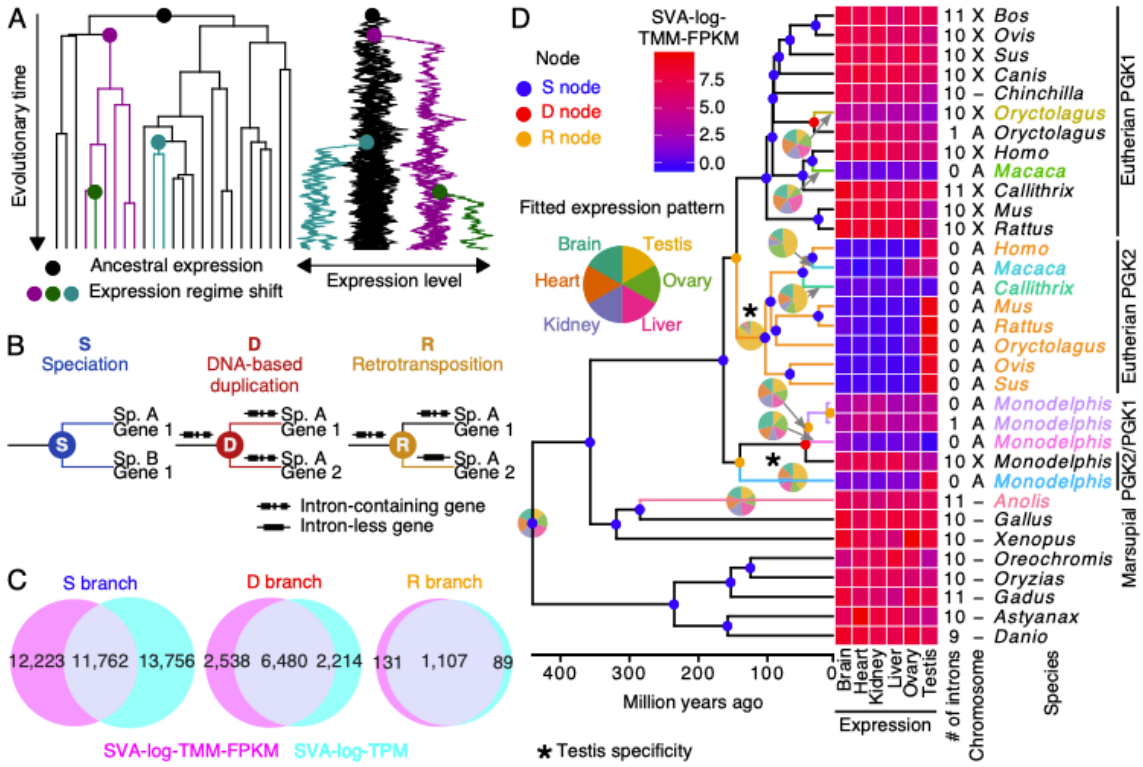


Figure 2. Expression evolution in a complex history of gene family evolution. (**A**) Modeling expression evolution with multi-optima Ornstein–Uhlenbeck process. A phylogenetic simulation is shown. Colors show branches belonging to different regimes. Regime shifts (change of color) appear as a substantial change in optimal trait values. (**B**) Nodes and branches of gene family phylogeny were categorized into S, D, or R based on the branching events, i.e., speciation, DNA-based duplication, or retrotransposition, respectively. (**C**) Venn diagrams of expression regime shifts. Circles represent the sets of branches where regime shifts were detected with SVA-log-TMM-FPKM or SVA-log-TPM. (**D**) The gene tree of phosphoglycerol kinases (orthogroup ID: OG0002332) is shown as an example. This orthogroup shows ortholog-specific expression patterns as well as regime shifts after speciation and lineage-specific gene duplication. Tips correspond to genes. The colors of branches and tip labels indicate expression regimes. Node colors match to the categorization in **B**. The heatmap shows expression levels and among-organ expression patterns are shown as a pie chart for each regime. To the right, the number of introns and located chromosomes (A, autosome; X, X chromosome; Y, Y chromosome) are indicated. For full information including complete tip labels and bootstrap branch supports, see Supplementary Data.

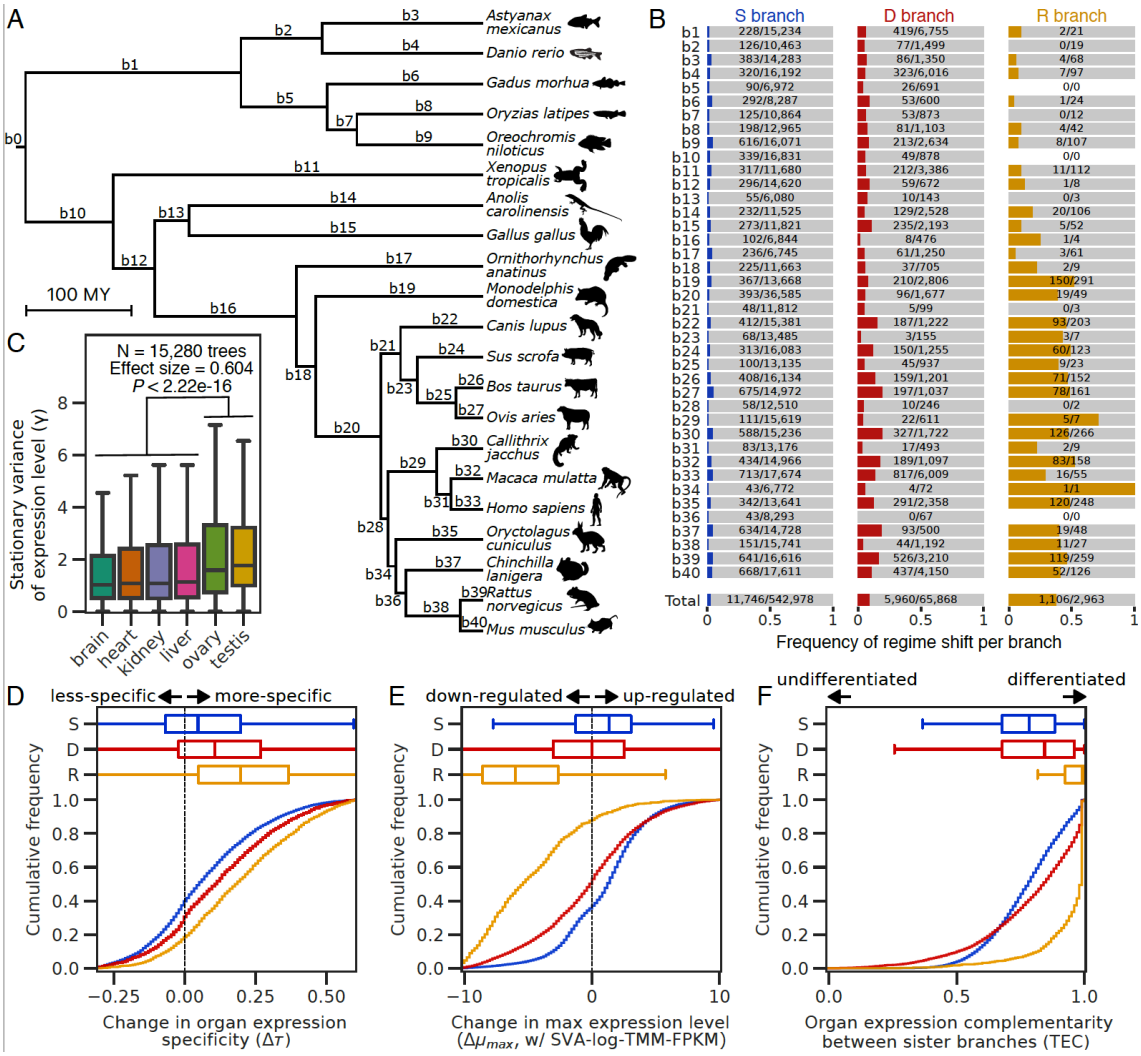


Figure 3. Characteristics of expression shifts in 15,280 gene trees. (**A**) The species tree showing analyzed genomes and their divergence time. (**B**) Mapping of 18,812 expression shifts in the species tree. The number and proportion of expression regime shifts in S, D, and R branches are shown. Corresponding branches in the species tree are indicated in **A**. (**C**) Organ-specific stationary variances (γ) of expression level evolution in vertebrates. The distribution of γ between reproductive and non-reproductive organs were compared by a Brunner-Munzel test. (**D–F**) Cumulative frequency of change in organ expression specificity (**D**), change in maximum expression level (**E**), and expression complementarity between sister lineages (**F**) among detected expression shifts. Number of analyzed regime shifts are shown in **B**.

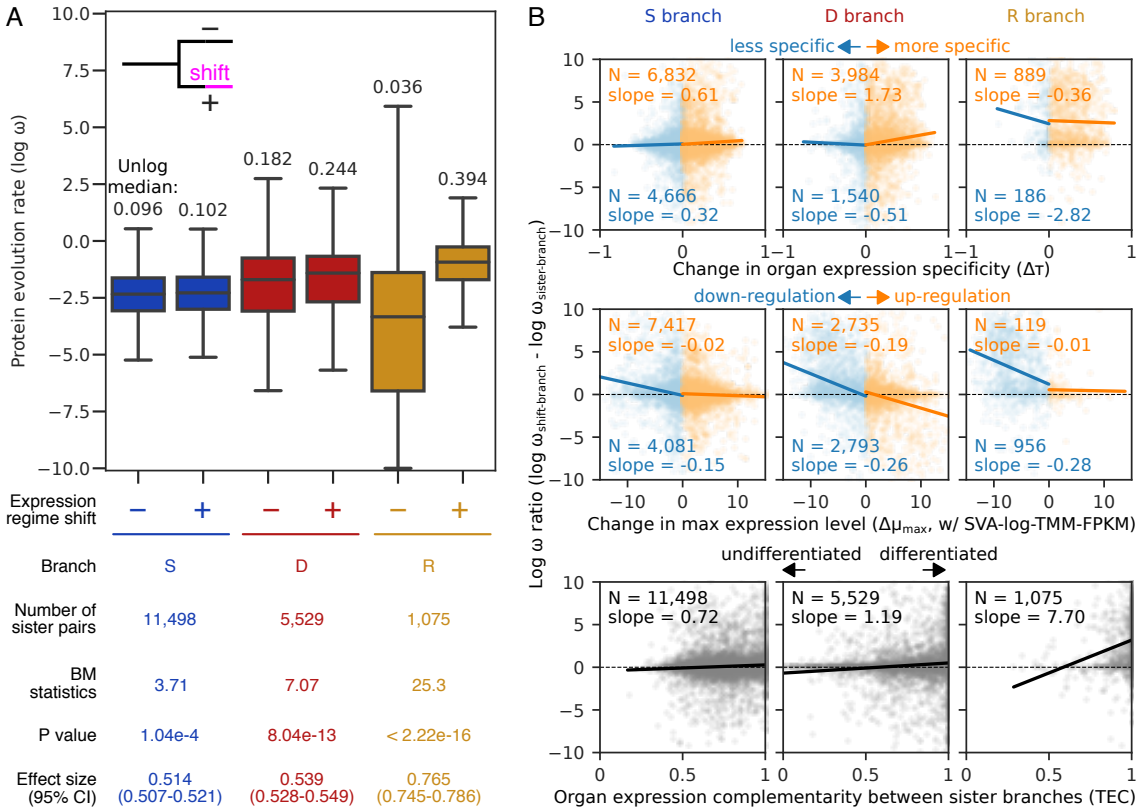


Figure 4. Context-dependent changes of protein evolution rate coupled with expression regime shifts. (**A**) Distribution of ω values. A plus (+) indicates branches with expression shifts, whereas minus (−) branches are sisters to the 'plus' branches. Statistical differences between pairs of distributions were tested using a two-sided Brunner–Munzel test . Non-log-transformed median values are shown above the boxplots. For visualization purpose, extreme values exceeding ± 10 were clipped. (**B**) Relationships between protein evolution rate and change in expression properties. Points correspond to expression regime shifts. Dashed lines indicate no between-branch difference in ω. Solid lines show a linear regression. Its slope and number of regime shifts are provided in the plot. Regime shifts with negative and positive changes were separately analyzed for organ expression specificity (upper) and maximum expression level (middle).

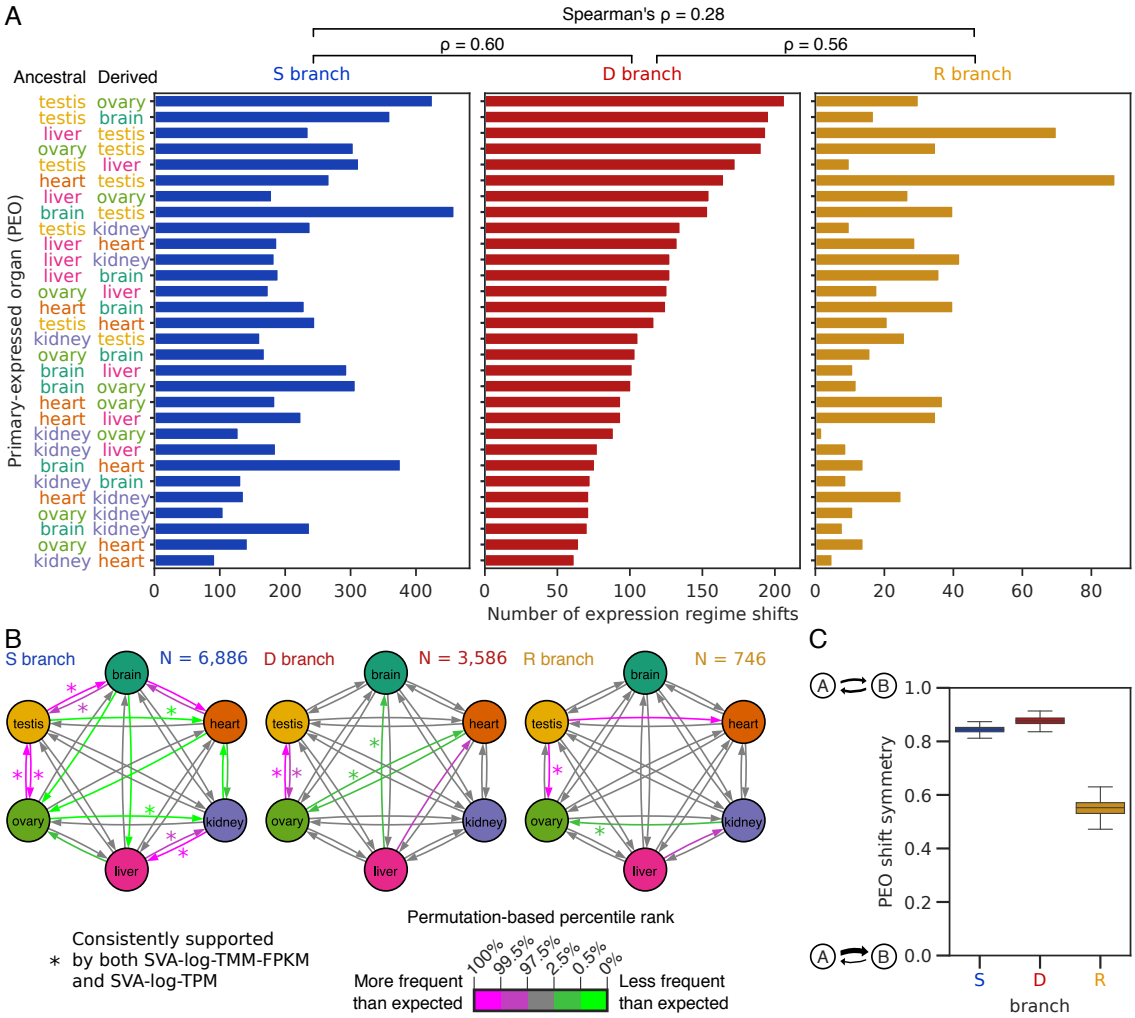


Figure 5. Evolutionary dynamics of gene expression. (**A**) Shift distributions of primary-expressed organs (PEOs). Y axis was sorted by abundance in D branches. Spearman's correlation coefficients among S, D, and R branches are shown above the plots. (**B**) Preadaptation networks in organ expression. Arrows represent transitions from ancestral PEOs to derived PEOs, and its color shows statistical significance based on 10,000 permutations. The results were obtained with SVA-log-TMM-FPKM, and an asterisk (*) indicates the statistical significance supported also by SVA-log-TMM-based analysis (Fig. S8). (**C**) The global polarity of PEO shifts. The global polarity is defined by the scaled sum of differences between two opposite PEO shifts. Boxplots show the distribution estimated by 1,000 bootstrap resampling.

931 Supplementary Materials for

932

933 **Organ-specific propensity drives patterns of gene expression evolution**

934

935 Kenji Fukushima and David D. Pollock

936

937

938 **List of Supplementary Materials:**

939

940 Tables S1–S5 (separate file)

941 Figs. S1–S10

942 Supplementary Data (separate file)

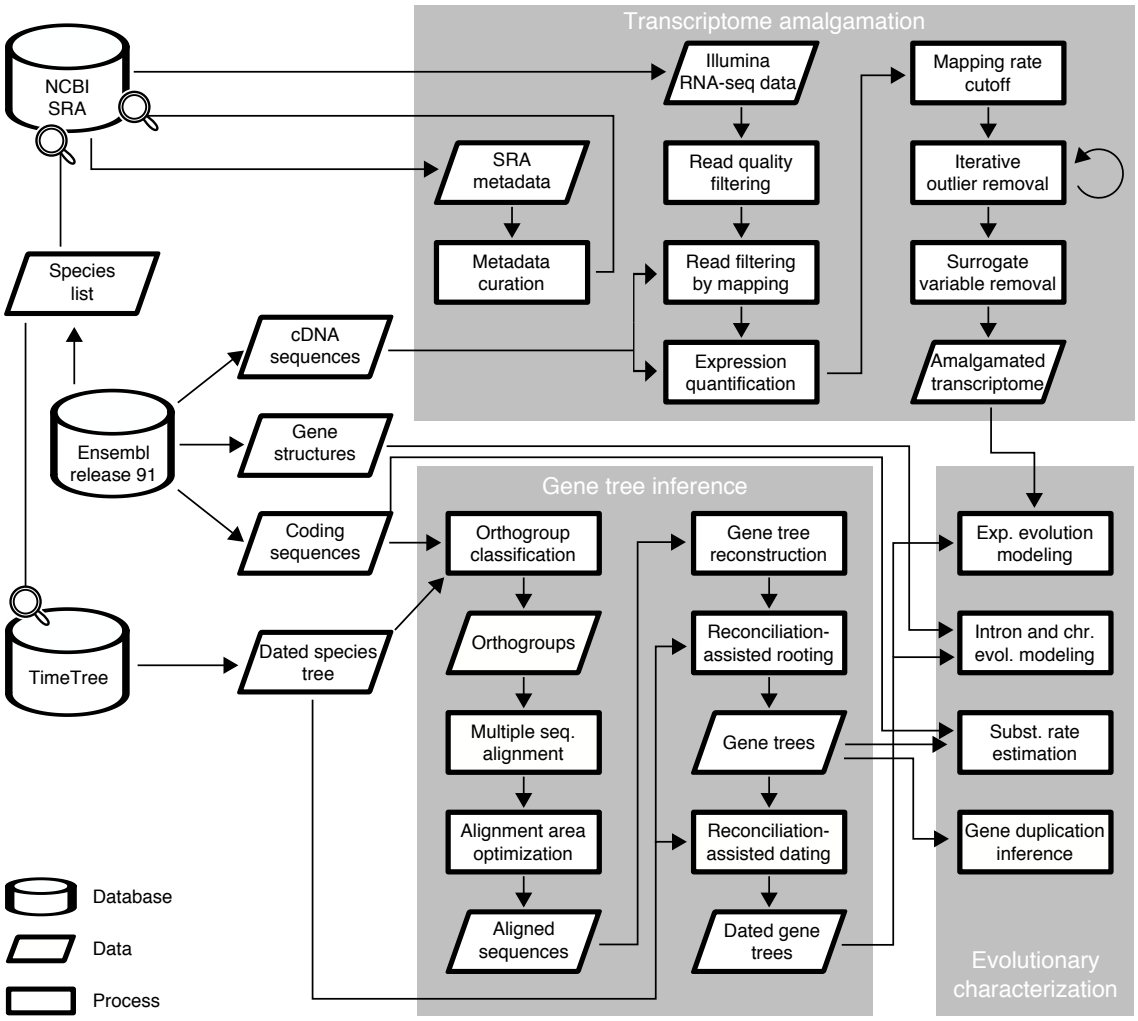


Fig. S1. A flow-chart of transcriptome amalgamation, gene tree inference, and evolutionary characterization in this study.

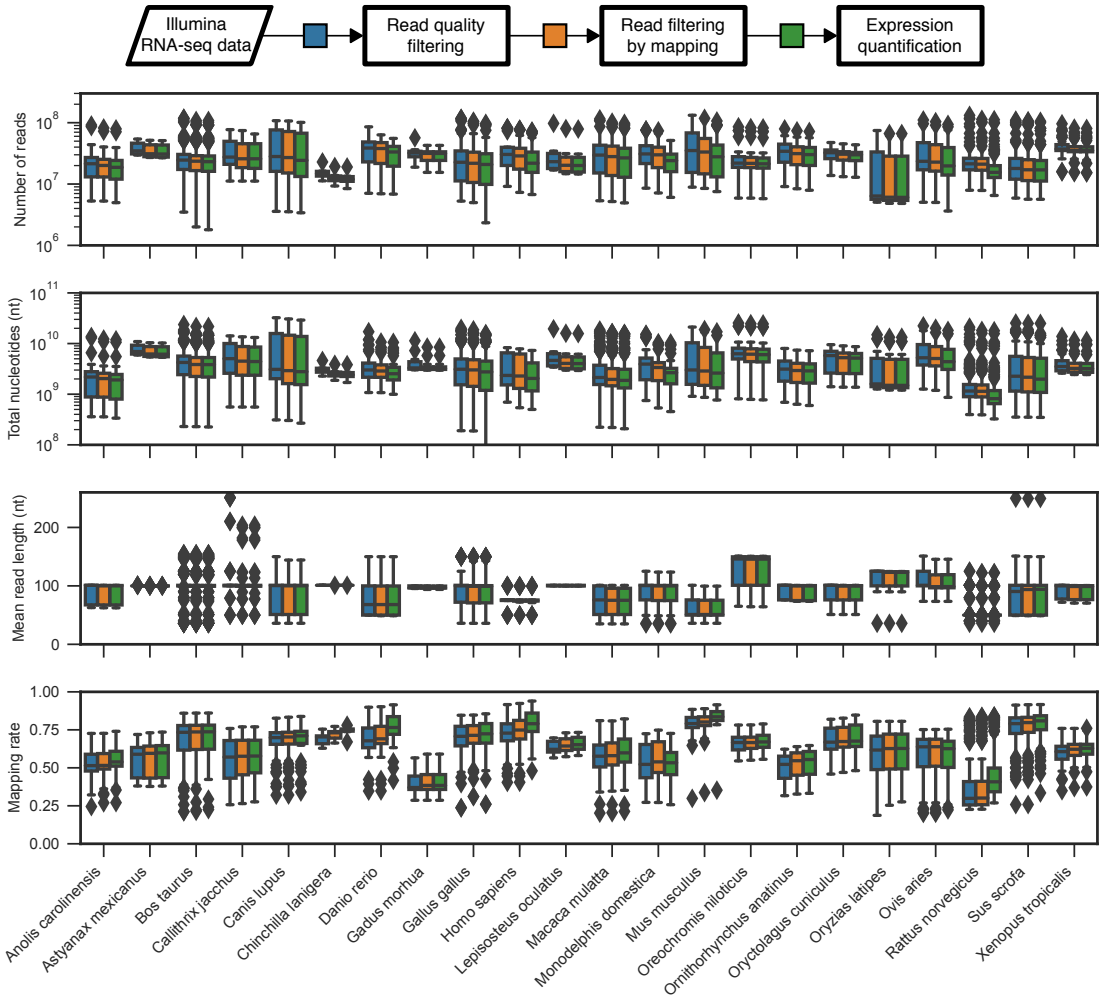


Fig. S2. Changes in number of reads, total nucleotide length, mean read length, and mapping rate by RNA-seq read filtering. Mapping rates were, in almost every case, improved by both read quality filtering and read filtering by mapping to miscellaneous genomic features.

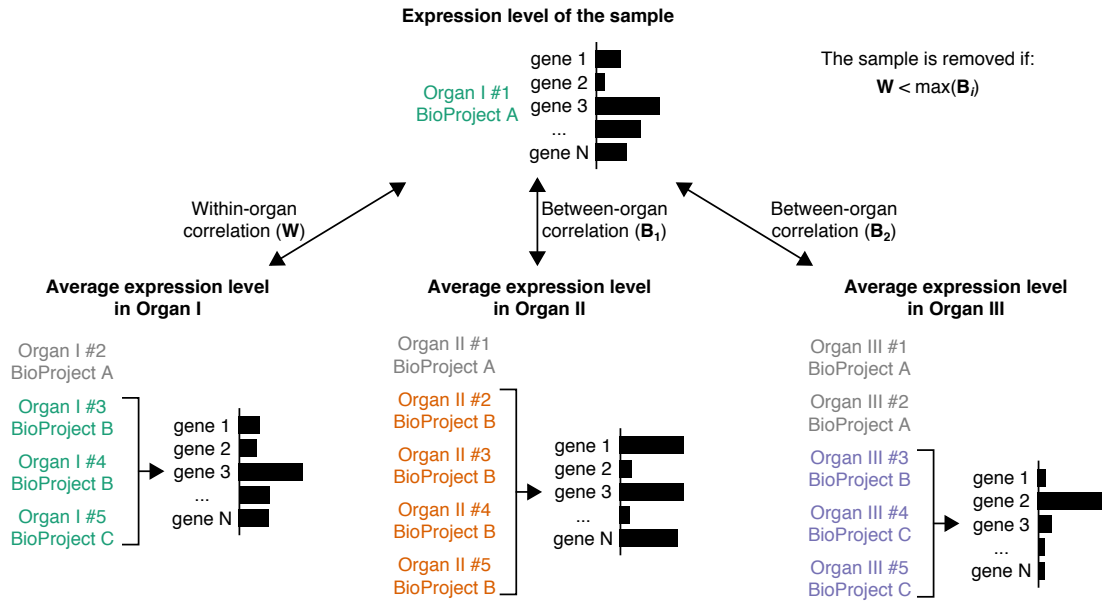


Fig. S3. A correlation analysis for the detection and removal of anomalous RNA-seq samples. Expression levels of all genes were compared between the sample and the organ averages. A sample was removed if any between-organ comparisons yielded a correlation coefficient higher than the within-organ comparison.

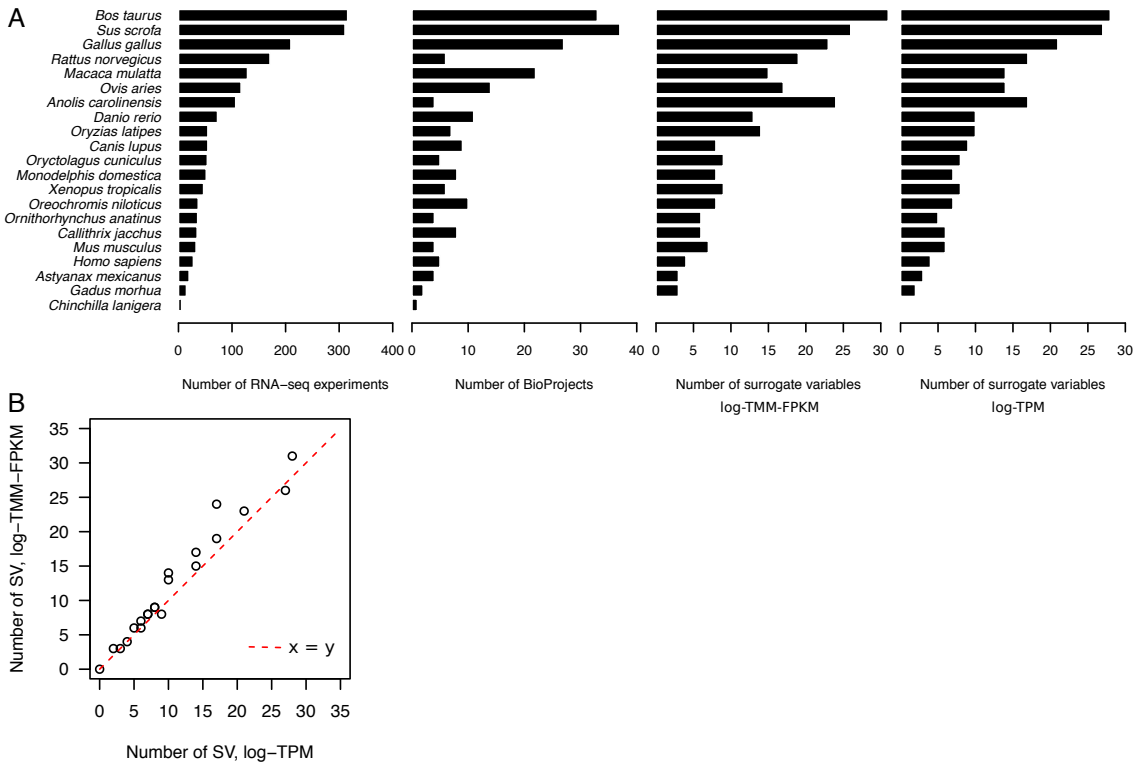


Fig. S4. Characteristics of amalgamated transcriptome. (A) The number of RNA-seq experiments, BioProjects, and detected surrogate variables in each species. The counts were derived from the final dataset. The numbers of surrogate variables (SV) are correlated with the numbers of RNA-seq samples and BioProjects. (**B**) Relationships of the number of SVs with log-TPM and log-TMM-FPKM. Points correspond to species. (**C–D**) Correlation heatmaps of corrected transcriptomes. See Table S1 for full descriptions including RNA-seq sample IDs and BioProject IDs. (**E–F**) Distinct distributions of Pearson's correlation coefficients depending on whether a pair of RNA-seq samples are the same organ or whether they are from the same research project. (**G–H**) Predictor analysis of detected surrogate variables. The predictive power was analyzed by linear regression using different properties of RNA-seq experiments: organ (brain, heart, kidney, liver, ovary, and testis), BioProject (e.g., PRJNA176589), library selection (e.g., cDNA and polyA), library layout (single and paired), instrument (e.g., Illumina HiSeq 2500 and NextSeq 550), number of read (e.g., 91,641,467 reads), % lost, fastp (percentage of reads that are removed by fastp; e.g., 5%), % lost, misc feature (percentage of reads that are mapped to non-nuclear-mRNA features and are removed from the analysis; e.g., 5%), minimum read length (e.g., 25 nt), average read length (e.g., 70 nt), maximum read length (e.g., 75 nt), and mapping rate (e.g., 80%). The predictors are

summarized in Table S1. (**I–J**) Multispecies correlation analysis of averaged organ expression. Corrected expression levels of 1,377 single-copy orthologs were used to calculate pairwise Pearson's correlation coefficients. (**K**) A principal component analysis using expression levels of 1,377 single-copy orthologs from 21 species. Points correspond to RNA-seq samples. Curves show the estimated kernel density. Explained variations in percentages are indicated in each axis. (**L**) Estimated organ-wise expression levels of a housekeeping gene. Since data from relatively many BioProjects are available, glyceraldehyde-3-phosphate dehydrogenase gene (GAPDH, ENSGALG00000014442) in *Gallus gallus* is shown. Points correspond to the average expression level calculated by a random subsampling. All data points and the median value (bar), rather than a box plot, are shown if the number of subsampled BioProject combinations is less than 10.

C SVA-log-TMM-FPKM

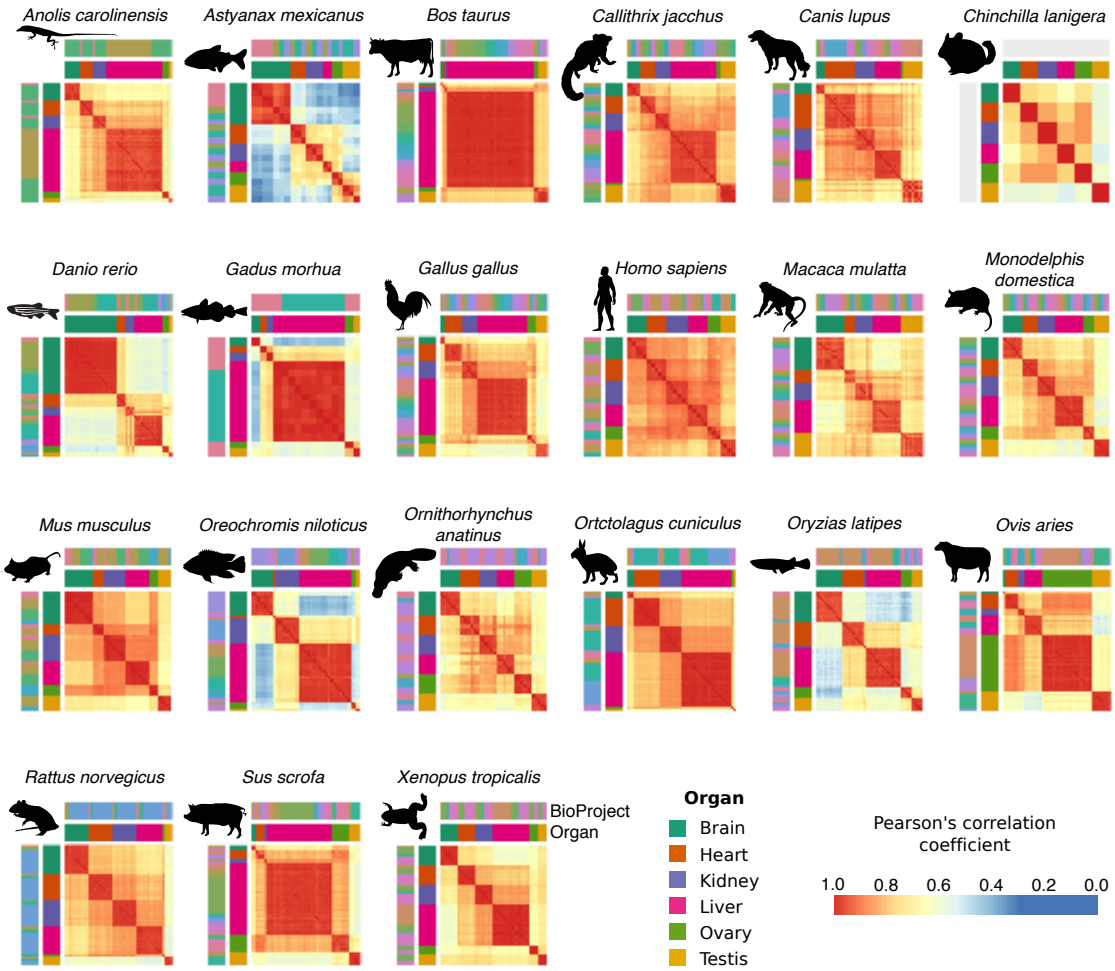


Fig. S4 (continued)

D SVA-log-TPM

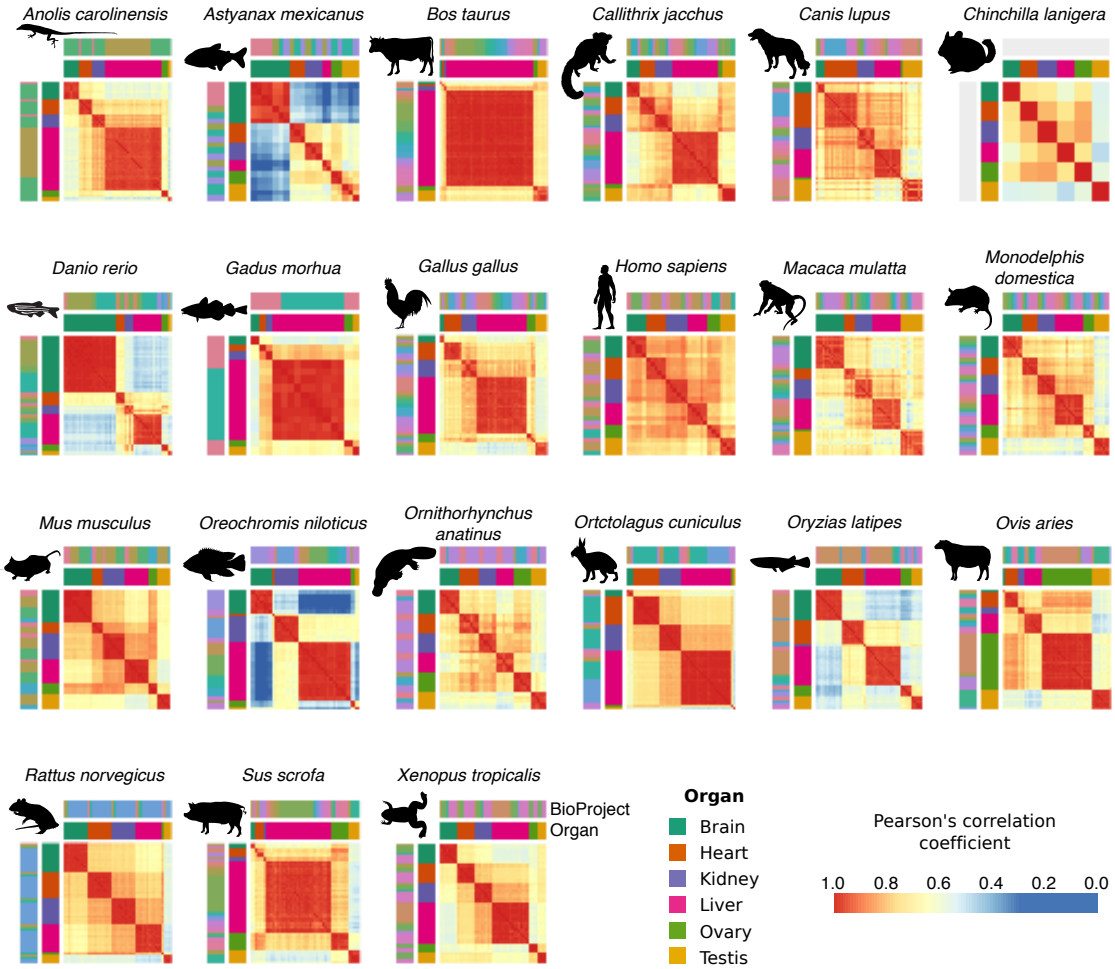


Fig. S4 (continued)

E SVA-log-TMM-FPKM

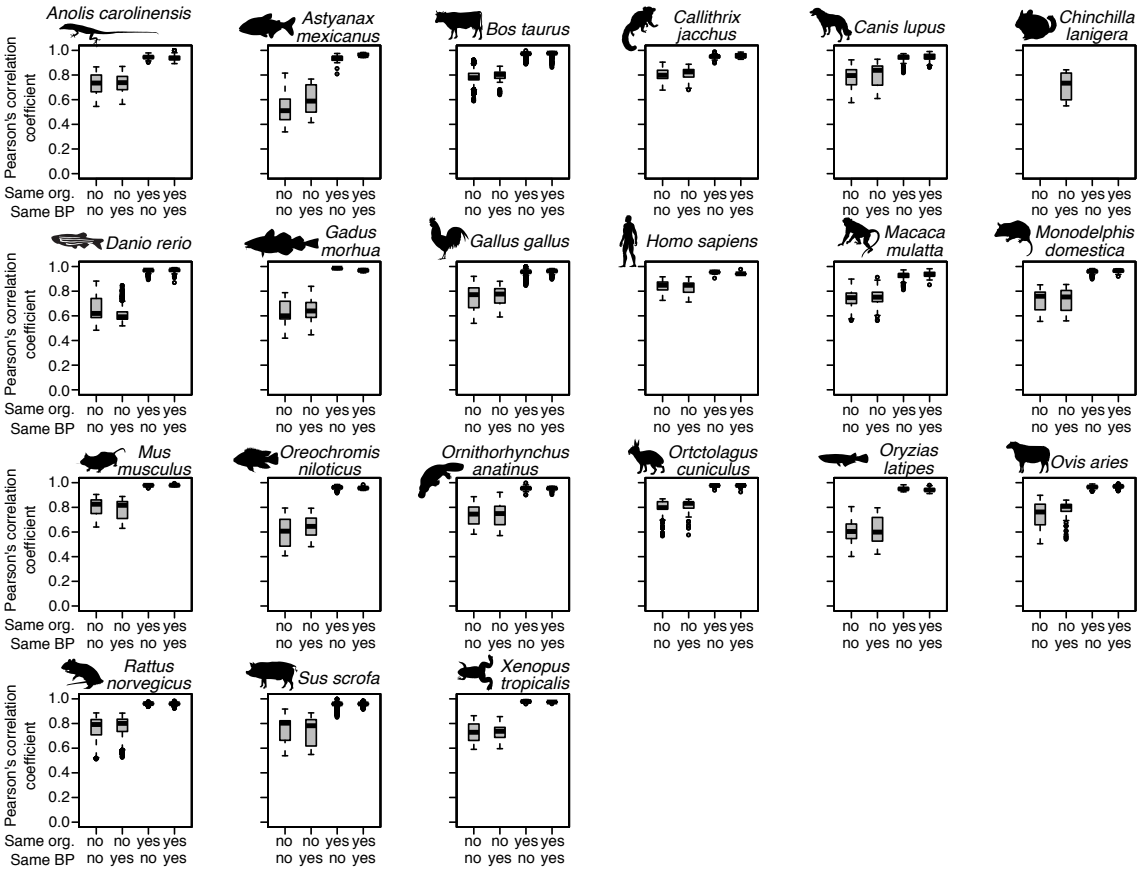


Fig. S4 (continued)

F SVA-log-TPM

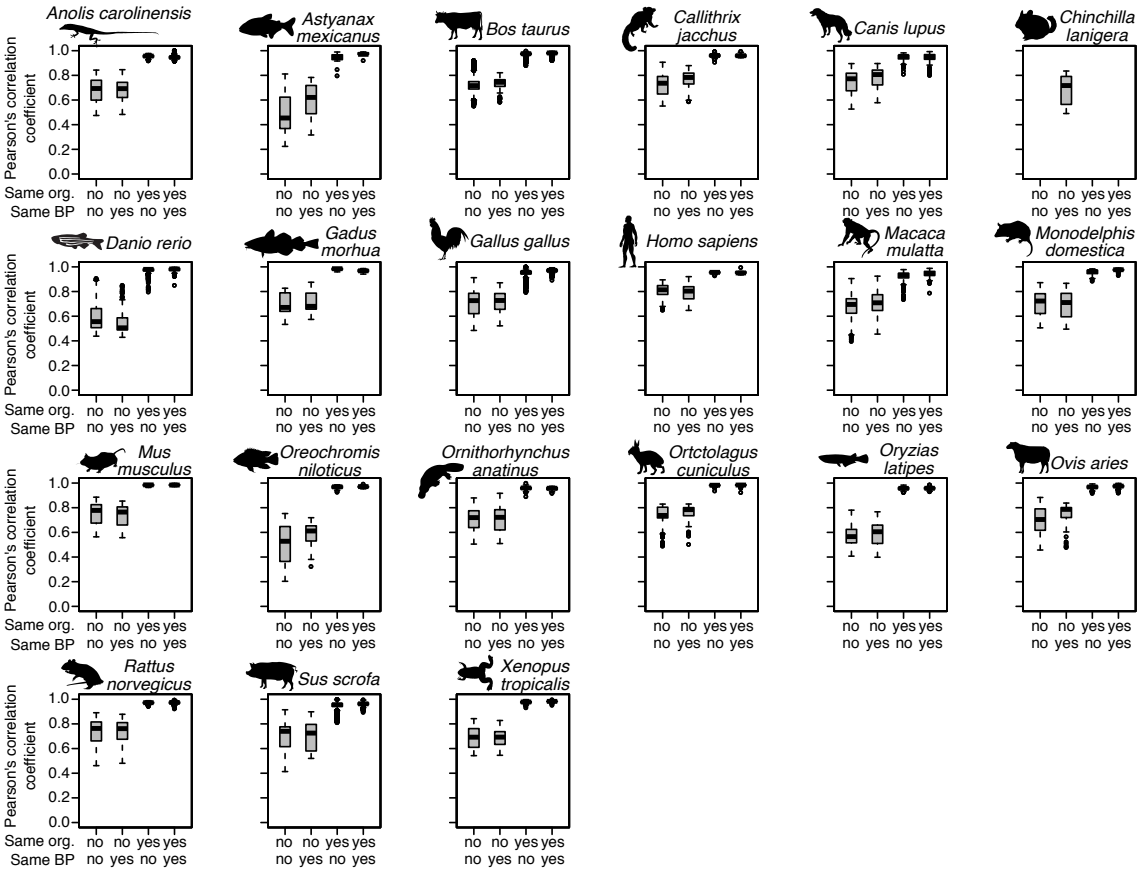


Fig. S4 (continued)

G SVA-log-TMM-FPKM

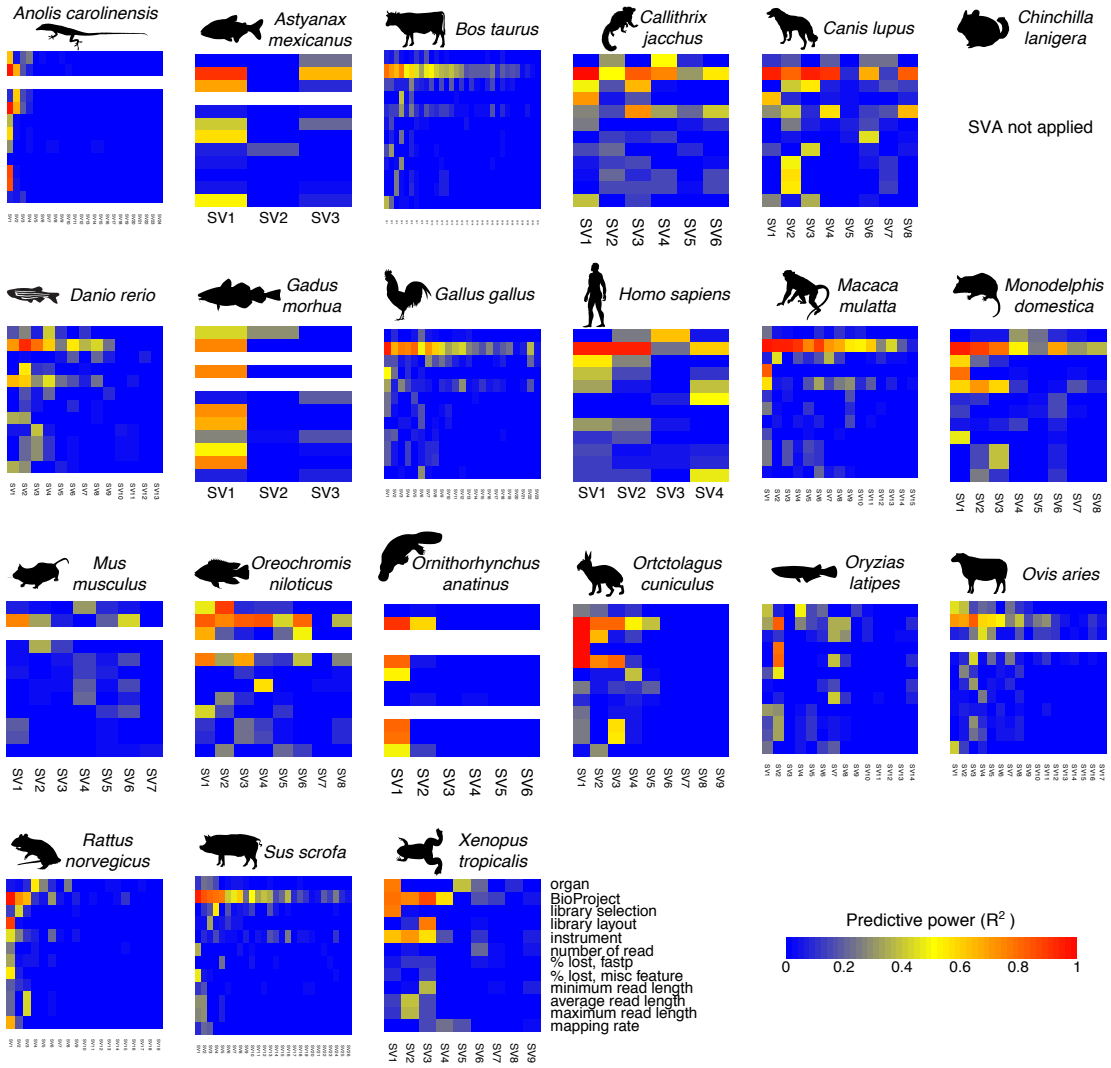


Fig. S4 (continued)

H SVA-log-TPM

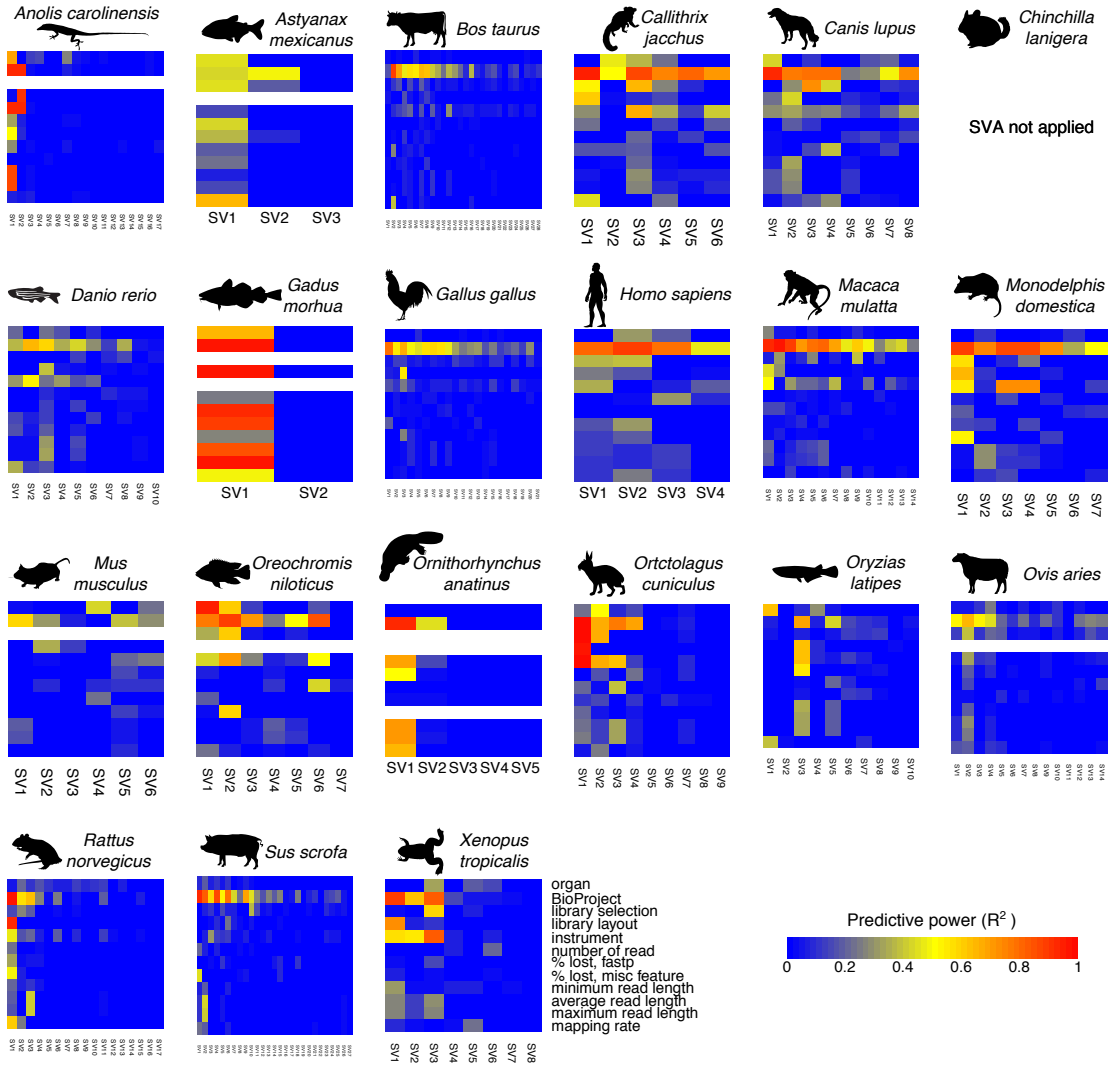
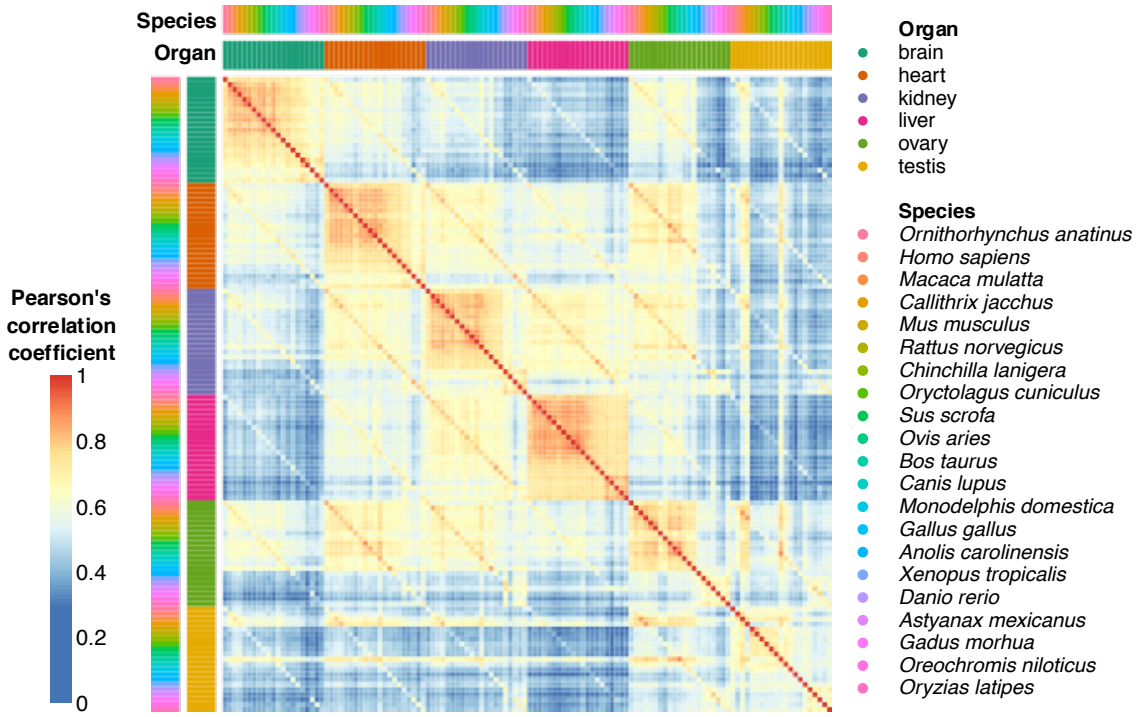


Fig. S4 (continued)

I SVA-log-TMM-FPKM



J SVA-log-TPM

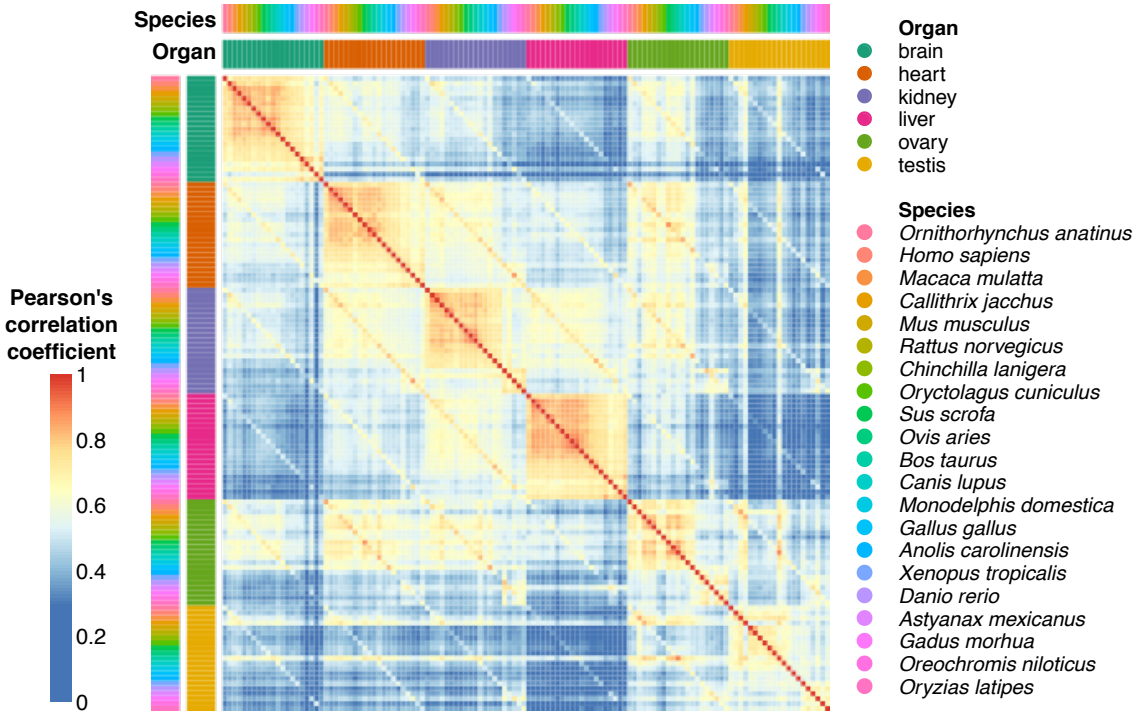
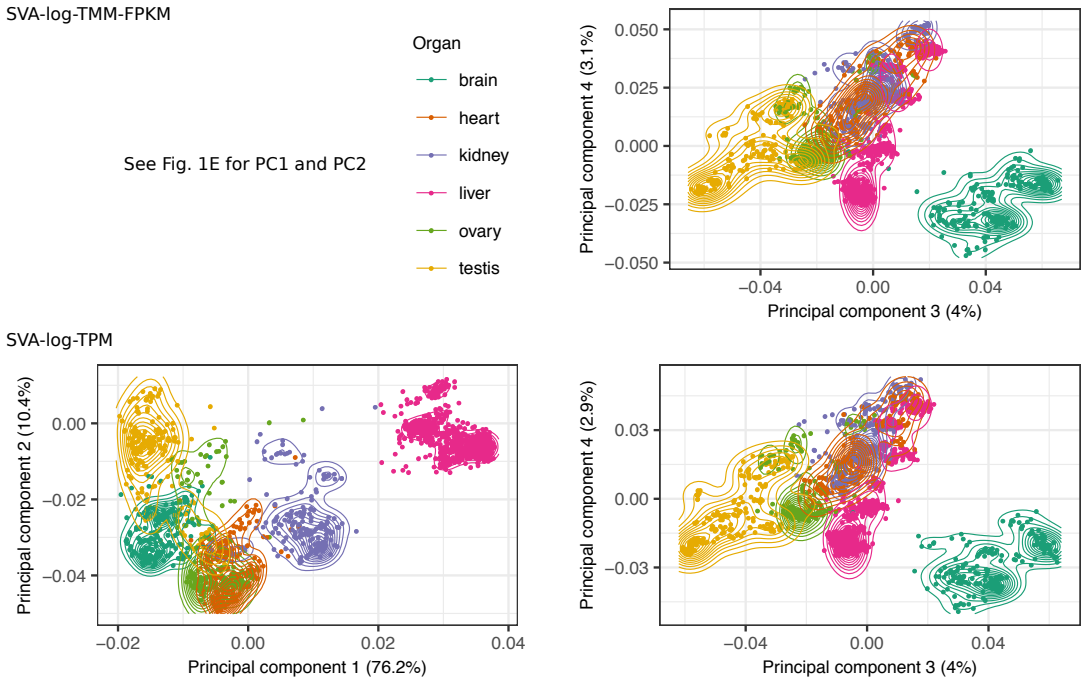


Fig. S4 (continued)

K



L

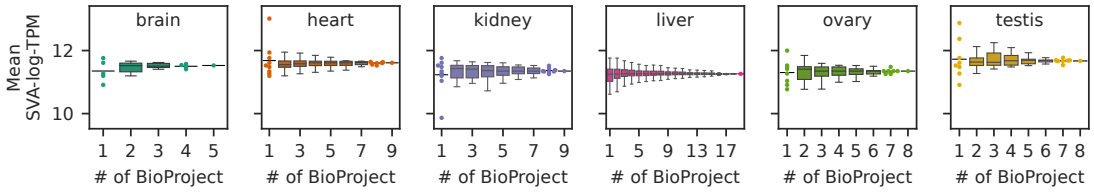


Fig. S4 (continued)

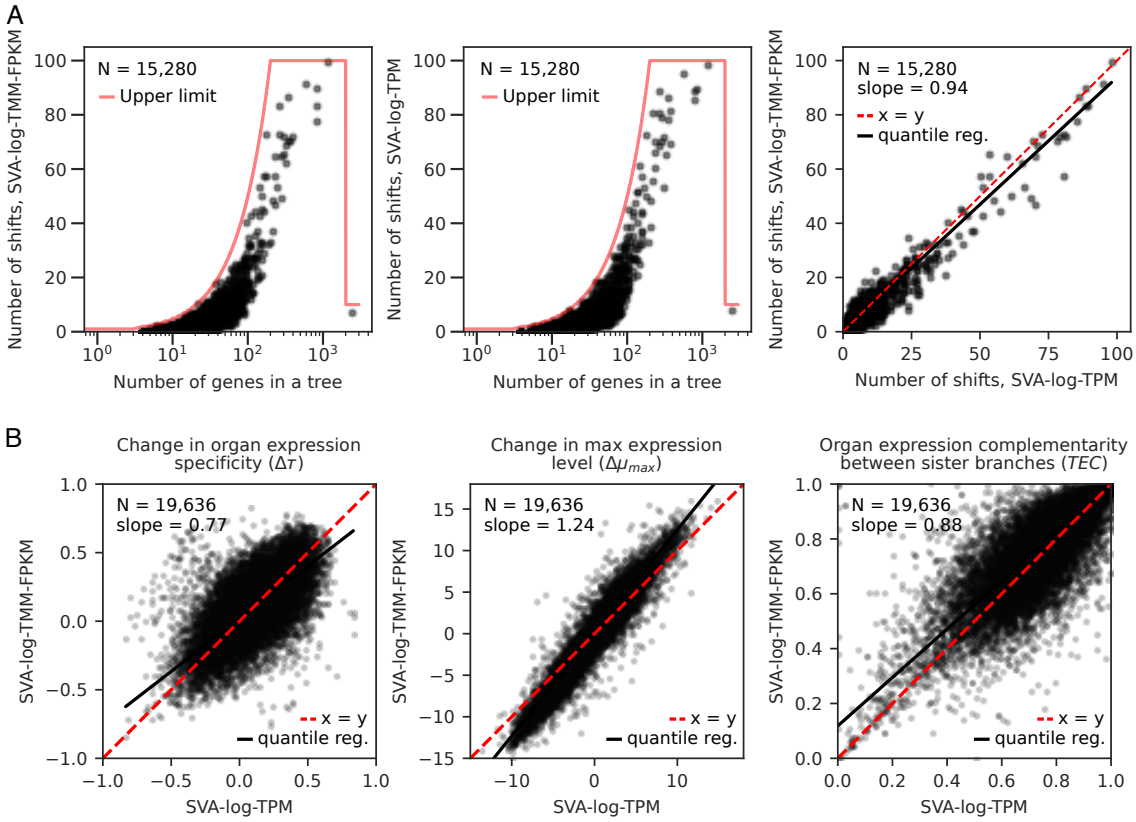


Fig. S5. Comparison of SVA-log-FPKM and SVA-log-TPM in expression regime shift detection. (**A**) The numbers of expression regime shifts. Each point corresponds to a gene tree. The solid red lines show the upper limits in the regime shift search. The numbers of shifts detected with SVA-log-TMM-FPKM (left) and SVA-log-TMM (center) are well correlated (right). The black line is a quantile regression, and the dashed red line shows a slope of 1. (**B**) Expression properties. Each point corresponds to an expression regime shift which is consistently detected by SVA-log-TMM-FPKM and SVA-log-TPM. (**C–D**) The branch-wise numbers of detected shifts in the species tree. The shifts were categorized by primary-expressed organs (PEOs) in ancestral and derived states. The order of the species tree branches (y axis) corresponds to the total number of corresponding gene tree branches in the dataset. See Fig. 3A for branch ids.

C SVA-log-TMM-FPKM

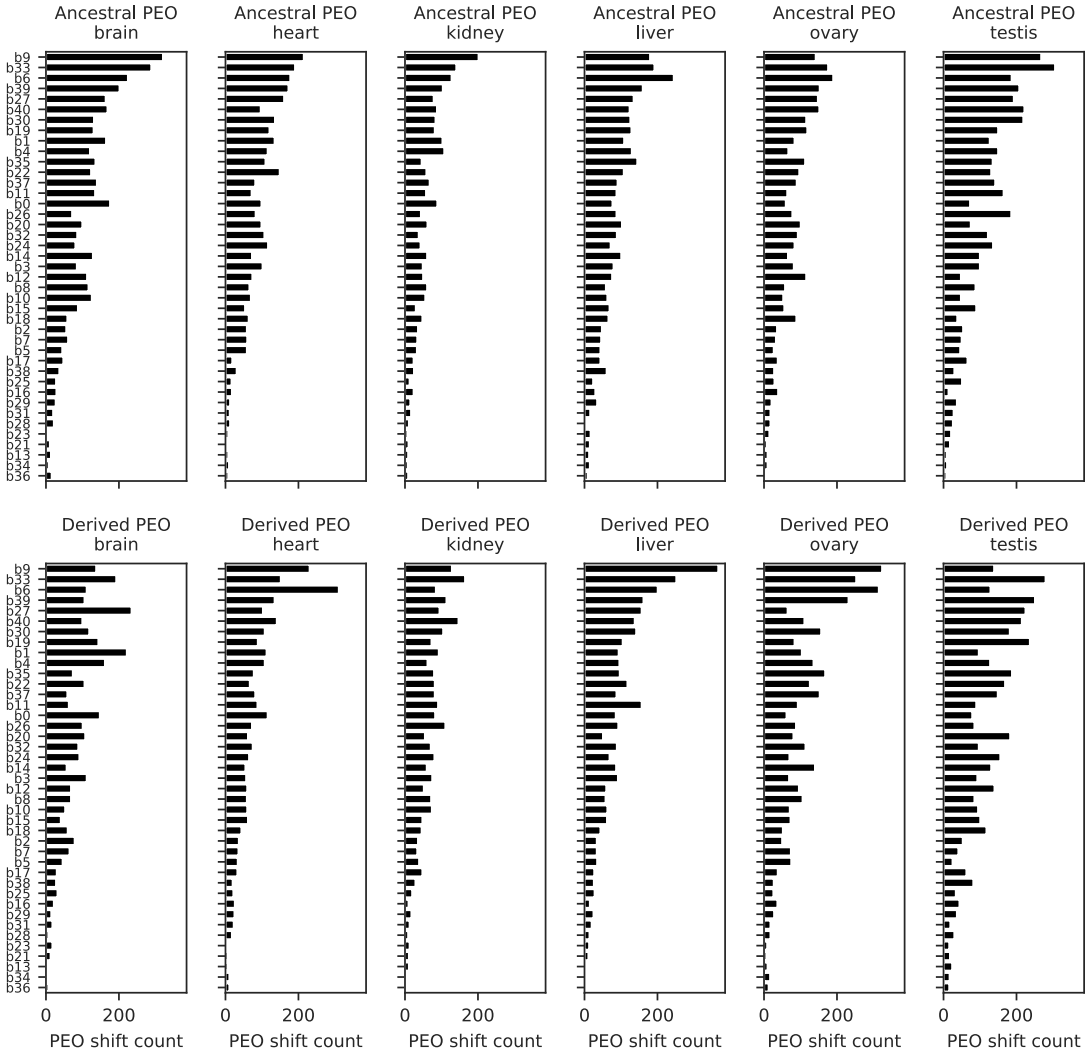


Fig. S5 (continued)

D SVA-log-TPM

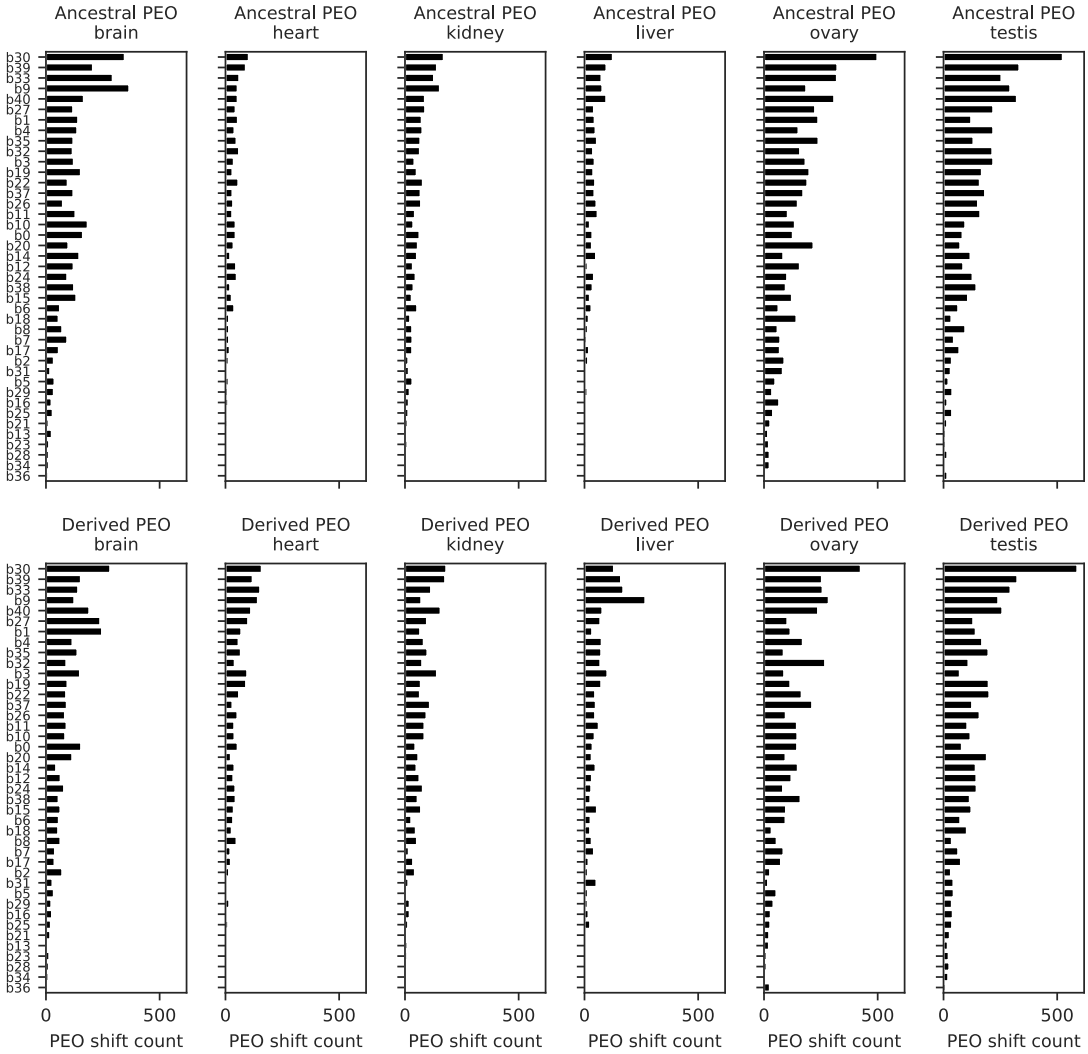


Fig. S5 (continued)

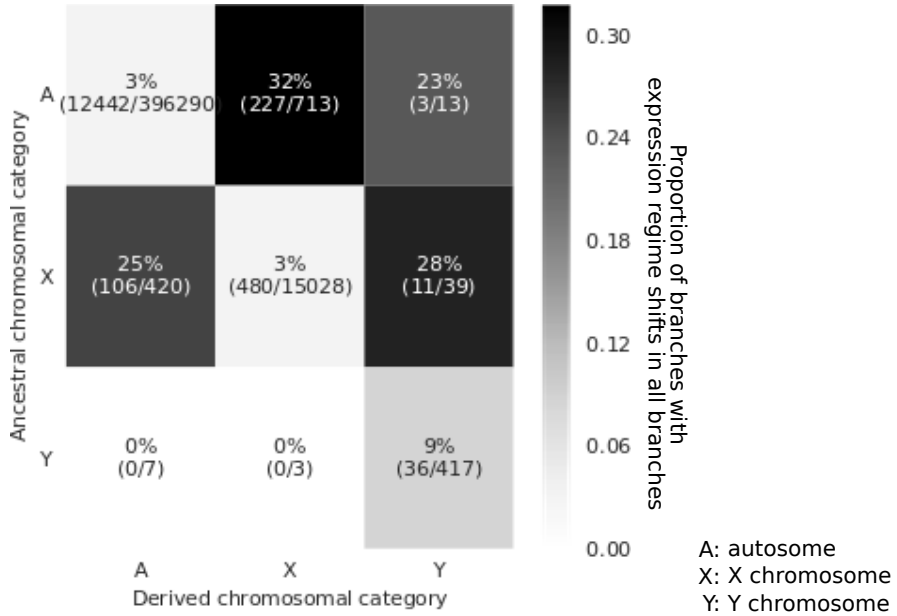


Fig. S6. The relationships between expression shifts and chromosomal location. The heatmap shows the frequency of expression shifts observed among the branches with or without a change in the chromosomal category (non-diagonal or diagonal, respectively). Chromosomal locations were categorized into autosomes (A), X chromosome (X) and Y chromosome (Y), and the ancestral locations were inferred by stochastic character mapping.

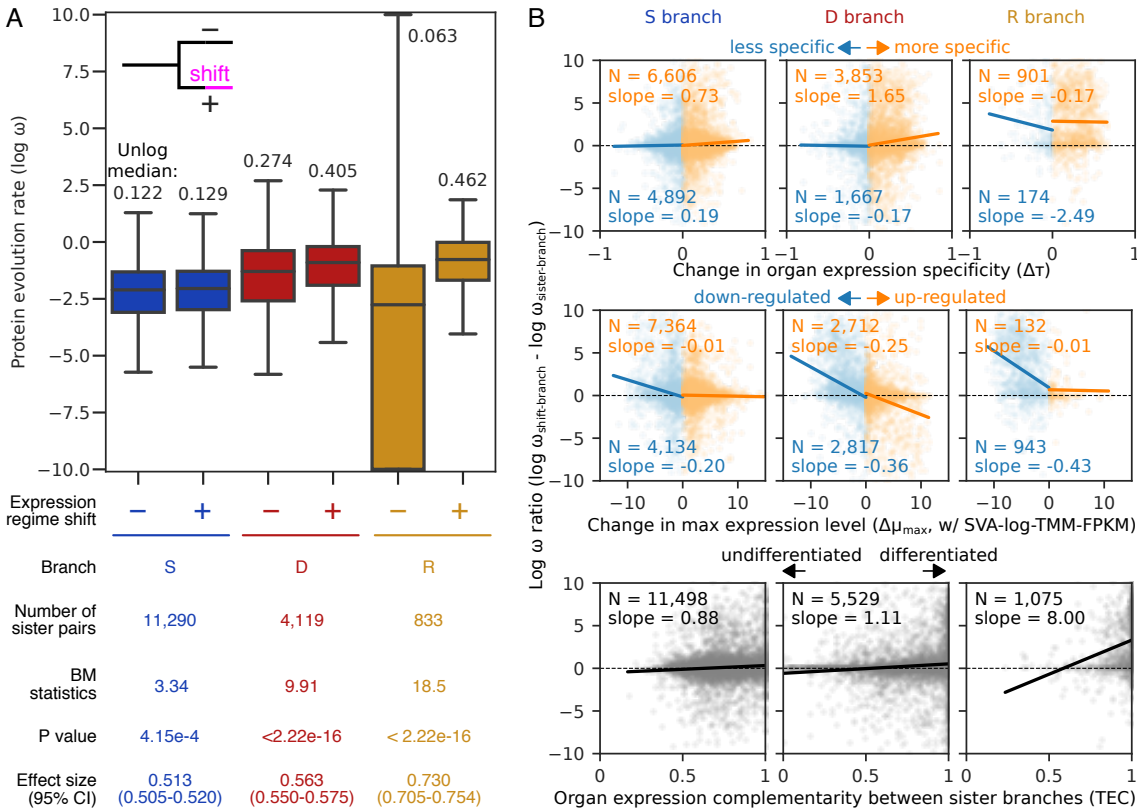


Fig. S7. Alternative analysis supporting non-linear change in protein evolution rate in correlation with expression regime shifts. (**A**) Distribution of ω values. While the results from stochastic character mapping (map*dNdS*) are reported in the main text, results from maximum-likelihood estimation (HyPhy) are shown here. A plus (+) indicates branches with expression shifts, whereas minus (−) branches are sisters to the 'plus' branches. Statistical differences between pairs of distributions were tested using a two-sided Brunner–Munzel test (Brunner and Munzel, 2000). Non-log-transformed median values are shown above the boxplots. For visualization purpose, extreme values exceeding ±10 were clipped. (**B**) Relationships between protein evolution rate and change in expression properties. While SVA-log-TMM-FPKM-based analysis is reported in the main text, results from SVA-lot-TPM-based analysis is shown here. Stochastic character mapping was used to obtain branch-wise ω values. Points correspond to expression regime shifts (log ω ratio = 0). Dashed lines indicate no between-branch difference in protein evolution rate. Solid lines show a linear regression. Its slope and number of regime shifts are also provided. Regime shifts with negative and positive changes were separately analyzed for organ specificity (upper) and expression level (middle).

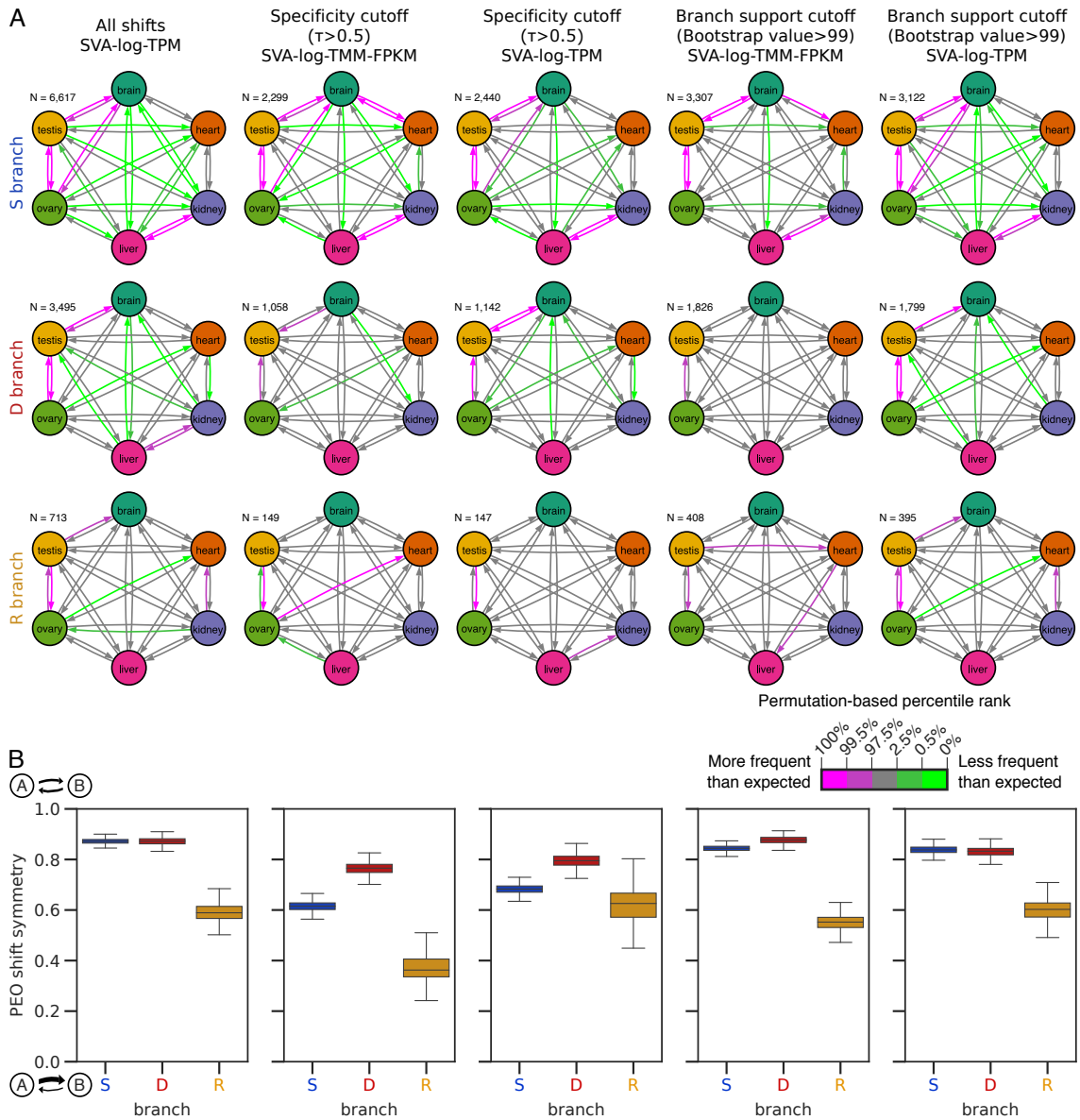


Fig. S8. Evolutionary dynamics of gene expression analyzed with conservative datasets. Both SVA-log-TPM and SVA-log-TMM FPKM values were analyzed. Organ-specific genes were examined by selecting PEO shifts with high ancestral and derived organ expression specificity ($\tau > 0.5$). To examine the effect of inference errors in gene tree reconstruction, branches with high support values (ultrafast bootstrap percentage > 99) were analyzed. **A** and **B** correspond to Fig. 5B and C, respectively.

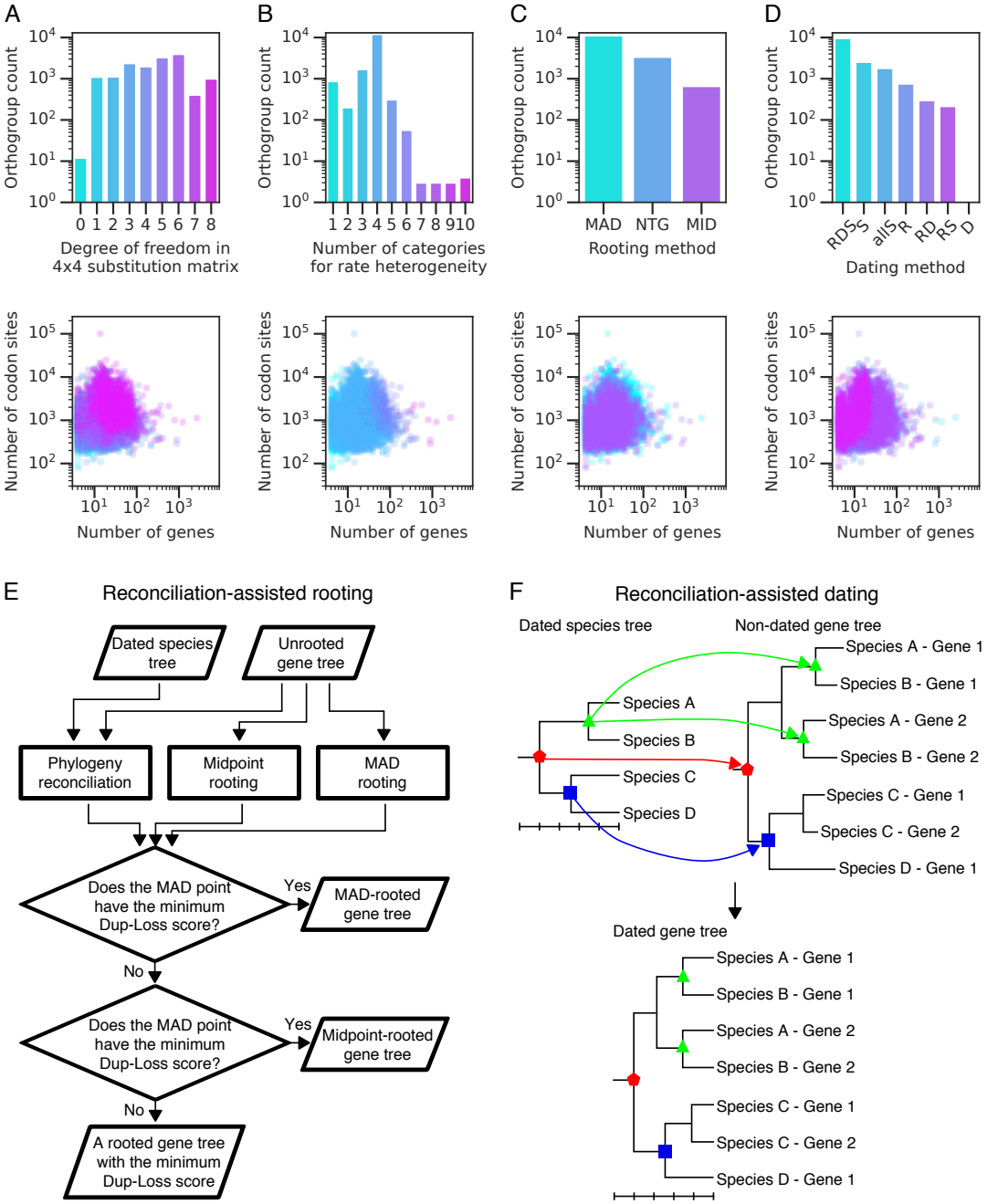


Fig. S9. Gene tree reconstruction. (**A**) Complexity of best-fit nucleotide substitution matrices. (**B**) Complexity of rate heterogeneity among nucleotide sites. Both discrete Gamma models (Yang, 1994) and FreeRate models (Soubrier et al., 2012; Yang, 1995) were included to count the number of categories for rate heterogeneity. (**C**) Selected rooting positions in reconciliation-assisted gene tree rooting. MAD, minimal ancestor deviation; NTG, 'rooting mode' of NOTUNG; MID, midpoint between the longest path. (**D**) Time-constrained nodes in tree dating. R, root node; S, speciation node; D, duplication node. All available constraints (RDS) are used in the first trial and then

1094 successively relaxed if estimation fails. In the category 'allS', all nodes are speciation
1095 nodes and therefore no divergence time estimation was performed for the trees. (**E**)
1096 Reconciliation-assisted gene tree rooting. Rooting points were estimated with two
1097 different methods: the minimum ancestor deviation (MAD) method and midpoint rooting.
1098 If they were compatible with the event parsimony involving gene duplication and loss,
1099 the MAD- or midpoint-rooted tree was reported in sequence. If not, one of event
1100 parsimony trees was reported. (**F**) Reconciliation-assisted gene tree dating. Speciation
1101 nodes in the dated species tree were mapped onto the non-dated gene tree by phylogeny
1102 reconciliation. By using those nodes as calibration points, the other node ages were
1103 estimated using the penalized likelihood method.
1104

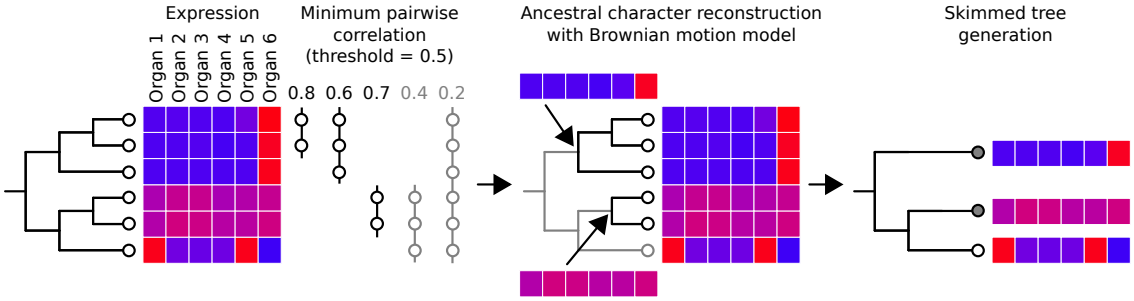


Fig. S10. Gene tree skimming in phylogenetic comparative analysis. Gene tree clades are collapsed if character states are highly correlated. Resultant trees contain a smaller number of leaves than the original trees while preserving drastic changes in character evolution. Note that, unlike this example, we used extremely stringent threshold in the analysis (Pearson's correlation coefficient > 0.99).