Inter-individual gene expression variability implies stable regulation of brain-biased genes across organs

Christabel F. Bucao[1,2*], Consolée Aletti[1], Sébastien Moretti[1,2], Andrew W. Thompson[3,4,5], Brett L. Racicot[4], Catherine A. Wilson[6], Julien Bobe[7], Ingo Braasch[4,5], Yann Guiguen[7], John H. Postlethwait[6], 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 Biological Sciences, Western Michigan University, Kalamazoo, Michigan, USA
[4] Department of Integrative Biology, Michigan State University, East Lansing, Michigan, USA
[5] Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, Michigan, USA
[6] Institute of Neuroscience, University of Oregon, Eugene, Oregon, USA
[7] INRAE, LPGP, Rennes 35000, France

[*]Corresponding author: christabelfloi.bucao@unil.ch

Abstract

Phenotypic variation among individuals plays a key role in evolution, since variation provides the material on which natural selection can act. One important link between genetic and phenotypic variation is gene expression. As for other phenotypes, the range of accessible expression variation is limited and biased by different evolutionary and developmental constraints. Gene expression variability broadly refers to the tendency of a gene to vary in expression (i.e., between individuals or cells) due to stochastic fluctuations or differences in genetic, epigenetic, or environmental factors, separately from the differences between e.g. organs. Variability due to biomolecular stochasticity (transcriptional 'noise') and cell-to-cell heterogeneity has been well-studied in isogenic populations of unicellular organisms such as bacteria and yeasts. However, for more complex organisms with multiple cells, tissues, and organs sharing the same genetic background, the interplay between inter-individual expression variability, gene and organ function, and gene regulation remains an open question. In this study, we used highly multiplexed 3'-end Bulk RNA Barcoding and sequencing (BRB-seq) to generate transcriptome profiles spanning at least nine organs in outbred individuals of three ray-finned fish species: zebrafish, Northern pike, and spotted gar. For each condition, we measured expression variation per gene independent of mean expression level. We observed that lowly variable genes are enriched in cellular housekeeping functions whereas highly variable genes are enriched in stimulus-response functions. Furthermore, genes with highly variable expression between individuals evolve under weaker purifying selection at the coding sequence level, indicating that intra-species gene expression variability predicts inter-species protein sequence divergence. Genes that are broadly expressed across organs tend to be both highly expressed and lowly variable between individuals, whereas organ-biased genes are typically highly variable within their top organ of expression. For genes with organ-biased expression profiles, we inferred differences in selective pressure on gene regulation depending on their top organ. We

found that genes with peak expression in the brain have low inter-individual expression variability across non-nervous organs, suggesting stabilizing selection on regulatory evolution of brain-biased genes. Conversely, liver-biased genes have highly variable expression across organs, implying weaker regulatory constraints. These patterns show that gene regulatory mechanisms evolved differently based on constraints on the primary organ.

Introduction

Phenotypic variation among individuals is crucial for evolutionary processes, as it serves as the substrate for natural selection. Gene expression is a key mediator of this phenotypic diversity. Similar to how phenotypic variation is shaped and constrained by evolutionary and developmental factors, variation in gene expression among individuals is also subject to comparable limitations and biases, with some genes having a wider range of expression than others. Variation in gene expression is influenced by various 'intrinsic' (gene-specific) and 'extrinsic' factors spanning biological scales, including transcriptional or translational stochasticity [1], cellular heterogeneity [2], genetic variation [3], epigenetic modifications [2,4], as well as micro- and macro-environmental cues [5]. Observed expression variation is often used as a proxy measure for expression *variability*, broadly defined as the *propensity* of a gene to vary its expression level in response to stochastic, genetic, or environmental perturbations [6].

Focusing on transcriptional or translational stochasticity ('expression noise') in genetically identical cells, many studies on isogenic populations of unicellular organisms (e.g., *Escherichia coli* [7–11] and *Saccharomyces* [12–18]) have probed the relationship between expression noise and gene promoter features [8–11], as well as properties of the gene regulatory network [19–21]. Evidence suggests that natural selection can shape expression noise according to gene function [8,12,13,15,22] and that constraints on expression noise are mediated by environmental context [23]. Furthermore, a correlation between gene expression noise in isogenic yeast cells and broader expression variation in non-isogenic populations has been reported [24], suggesting that some shared regulatory mechanisms underlie both intra-genotype expression noise and inter-genotype expression variability [25]. However, little is known about the relationship among expression variability, gene function, and gene regulation in animals, which have functionally specialized organs with distinct expression profiles despite a common genetic background. Although organ-specific enhancers exist to facilitate spatiotemporal control of gene expression [26], we expect some level of pleiotropy of regulation of gene expression in different organs. Thus selection on gene expression in different organs is not totally independent, and the expression variability of a gene within any given organ may potentially be constrained by antagonistic organ-specific expression optima.

Studies on gene expression variability in animals have primarily focused on select cell types or organs in model organisms [27–29], with particular interests in the role of stochasticity in cell differentiation [30] (including shifts in inter-cell [31,32] or inter-embryo expression variability over development [33]), as well as expression variability in human organs and its association with disease [34–40]. (Of note, although most

75 studies compare expression in different organs, the terminology of 'tissues' and 'tissue-specificity' is
76 frequently used in the literature instead.) Although the functional relevance of expression variability is
77 increasingly recognized [41–43], how such variability is modulated by organ context remains largely
78 unexplored, even in studies that sample multiple organs [25,37,39,40]. Few studies have been published on
79 variability in outbred, non-model animal populations. This is partly due to heavy resource requirements such
80 as the need for many replicates for statistical significance, as well as the challenges associated with
81 disentangling different sources of expression variation, whether in bulk or at the single-cell level.
82 Recently, the advent of low-cost high-throughput sequencing methods has made it increasingly
83 feasible to generate sufficient transcriptome data to measure within-organ gene expression variation across
84 several individuals and estimate within-organ expression *variability*. In this study, we present the first analysis
85 of inter-individual gene expression variability across multiple organs, using newly generated 3′-end Bulk RNA
86 Barcoding and sequencing (BRB-seq) [44] data in three ray-finned fish species. We confirmed that gene
87 expression variability between biological replicates is a functionally relevant trait. Genes that are broadly
88 expressed across organs have lower within-organ variability in expression between individuals compared to
89 organ-biased genes, reflecting pleiotropic constraints on expression. We further found that genes with
90 organ-biased expression profiles tend to be variably, and thus dynamically, expressed in their top ('focal')
91 organ, and that moreover their variability across other organs depends on their focal organ. Investigating
92 gene expression variability within and between animal organs provides a new framework for understanding
93 gene regulation and the evolution of gene function in complex multicellular organisms.
94

95 **Results**

96 **Organ transcriptome profiles of ray-finned fishes**

97 To investigate inter-individual gene expression variability within and between organs, we generated
98 transcriptomic data across multiple organs using 3′-end Bulk RNA Barcoding and sequencing (BRB-seq) [44]
99 for three species of ray-finned fishes (Actinopterygii): two teleost species (*Danio rerio* (zebrafish) and *Esox*
100 *lucius* (Northern pike)) and one non-teleost outgroup (*Lepisosteus oculatus* (spotted gar)). Zebrafish and
101 Northern pike represent two of the major teleost clades (Otomorpha and Euteleosteomorpha, respectively)
102 and neither species have undergone additional lineage-specific genome duplications after the teleost
103 whole-genome duplication [45], while the spotted gar is a slowly evolving outgroup to the teleosts [46]. For
104 each species, we sampled multiple outbred biological replicates spanning several organs in both sexes (when
105 known) (**Figure 1A**) and applied a two-step filter to remove low quality samples based on sequencing quality
106 and within-organ expression correlation (**Figure S1**, Materials and Methods). Hierarchical clustering of
107 zebrafish samples based on pairwise Pearson correlation shows that transcriptome profiles cluster into five
108 major groups by related organ systems (**Figure 1B**), with reproductive organs (cluster 1) having the most
109 divergent expression profiles. For non-reproductive organs, the following clusters can be observed:
110 gastrointestinal and urinary systems (kidney, intestine, and liver; cluster 2); nervous system (brain and eye;

cluster 3); external organs (pectoral fin, skin, and gills; cluster 4); and lastly, muscular system (heart and muscle; cluster 5). Apart from the gonads, we do not observe strong clustering by sex except for the liver and pectoral fin. Although some differences can be observed in Northern pike and spotted gar, clustering of organs of the nervous system, muscular system, and external organs are generally consistent among species (**Figure S2**).

A

	D. rerio		*E. lucius*		*L. oculatus*
	♀	♂	♀	♂	♀ / ♂
blood			2	3	6
brain	13	11	10	9	11
eye	11	11	6	5	10
gills	7	9	9	9	6
gonads	11	9	7	5	5
heart	1	2	10	9	9
intestine	11	12	8	8	15
kidney	4	6	9	8	
liver	10	9	4	10	8
muscle	1	1	7	8	7
pectoral fin	10	12	10	10	8
skin	10	11			12
swim bladder			3	5	7
Σ	89	93	85	89	104

B

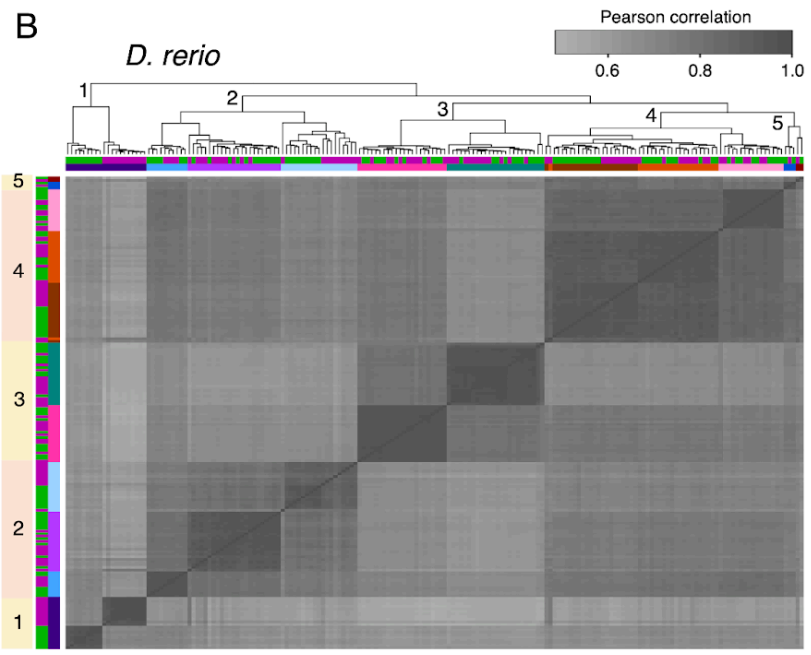


Figure 1. Study design for multi-organ analysis of inter-individual gene expression variability in fishes. (A) Summary of BRB-seq dataset. Table shows the number of biological replicates per organ (rows) and sex (columns) within each species, after removal of low-quality samples. There is no information on sex for spotted gar (*L. oculatus*). Organ-sex conditions with fewer than 4 replicates (light gray) were excluded from expression variability analyses. **(B)** Hierarchical clustering heatmap of organ transcriptomes (n = 182) in zebrafish (*D. rerio*). Samples cluster into 5 groups by related organ systems: reproductive system (1), gastrointestinal and urinary system (2), nervous system (3), external organs (4), and muscular system (5). Pairwise distances between samples are computed based on Pearson correlation of log-transformed, TMM-normalized counts per million ($\log_2$ TMM-CPM) of protein-coding genes (n = 19,466). The color keys for organ (inner color bars) and sex (outer color bars) are shown in (A).

Gene expression variability is estimated independently of expression level

Here, we use 'variation' to refer to a measurable statistical feature of an ensemble (i.e., statistical dispersion), and 'variability' to refer to a dispositional property inferred from such measurements [6]. Measuring expression variation as a proxy for quantifying variability has not been standardized in the literature, as it is not well-studied beyond unicellular organisms or model cell lines. Different measures of statistical dispersion (e.g., standard deviation, variance, median absolute deviation, or coefficient of variation (CV)) are known to be strongly anti-correlated with mRNA abundance (**Figure 2A, Figure S4**) [47,48]. To correct for this mean-dispersion relationship, many studies use regression to model variation as a function of expression level [48] and use either the ratio [28,33,39] or difference [25,38] (i.e., model residual) of observed to expected variation as an adjusted estimate of expression variability. Although this removes the global correlation between expression mean and variation, this does not correct for the non-homogenous distribution of variation across the range of expression levels [27,37]. To address this, other studies apply percentile rank normalization (e.g., 'local coefficient of variation' (LCV) [37,49] or 'relative noise rank' [27]), in which the effect of mean expression is removed by ranking the expression variation of each gene relative to a window of genes with similar expression levels. Here, we used the local (residual) coefficient of variation as our estimate of expression variability because it is globally uncorrelated with mean expression level, as in regression-based approaches, and it also imposes a uniform expression variability distribution across expression levels. Although the latter is a very strong correction, by forcing a uniform distribution we maximally remove mean expression as an explanatory factor for downstream analyses.

Since actual expression variation may be skewed by samples with more divergent expression profiles, we computed the LCV for each gene within an organ (for unsexed spotted gar) or organ-sex condition (for sexed Northern pike and zebrafish) using a jackknife resampling or 'leave-one-out' procedure to mitigate the effect of outliers on expression variability estimates. For each subset of n - 1 replicates, we first computed the log-transformed, squared coefficient of variation ($\log(CV)^2$, where CV is the ratio of standard deviation to the mean) and fitted a local polynomial regression (LOESS) to model $\log(CV)^2$ as a function of mean expression (**Figure 2B**). Second, we ranked the residual $\log(CV)^2$ ('residual CV') of each gene within sliding windows of 100 neighboring genes sorted by expression level, such that genes with rank 0.00 have the lowest dispersion whereas genes with rank 1.00 have the highest dispersion within a local gene neighborhood (**Figure 2C**). Lastly, we calculated the mean LCV over all n - 1 subsets within a condition and

used this metric, which we call 'variability rank', as our estimate of gene expression variability (**Figures S5-S7**).

Because variability ranks are computed relative to the set of genes expressed within a specific condition, it is not an absolute measure of variability that is directly comparable between conditions. A gene with a high variability rank in the brain is considered highly variable only with respect to other genes expressed in the brain (e.g., a gene with a variability rank of 0.80 in the brain is not necessarily more variable, in the absolute sense, than a gene with a variability rank of 0.60 in the liver). By its construction, variability ranks are thus generally organ-specific, with weak to moderate correlation across organs (**Figure S8-S10**).

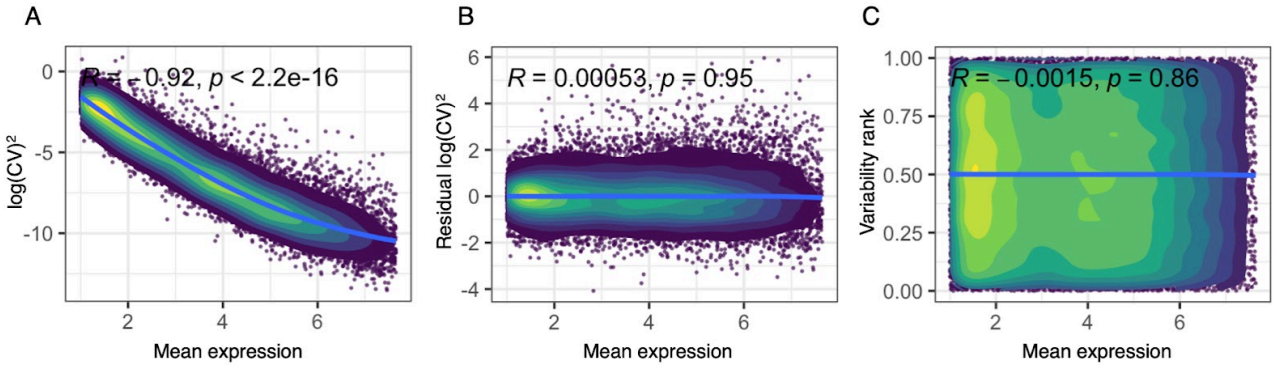


Figure 2. Measures of expression variation as a function of expression level. Data from female zebrafish brain (n = 13 samples). All values are means estimated from jackknife (n - 1) resampling. Mean expression refers to log-transformed, TMM-normalized counts per million ($\log_2$ TMM-CPM). **(A)** Log-transformed (squared) coefficient of variation (CV, defined as the ratio of standard deviation to the mean). **(B)** Residuals of $\log_2$ (squared) CV. **(C)** Local percentile rank of residual $\log(CV)^2$ ('variability rank') within sliding windows of 100 neighboring genes sorted by expression level. Blue lines are LOESS regression curves and R is Pearson's correlation coefficient.

One limitation of the variability rank metric is that it does not describe the modality of the underlying expression distribution. Overdispersion may be due to either a broader range of expression centered at a single mean (i.e., a unimodal distribution) or the presence of two or more expression peaks (a multimodal distribution), the latter of which may be indicative of hidden substructures within the population [43]. To investigate this possibility, we assessed whether more bimodal expression profiles can be observed for sets of increasingly more variable genes (**Figure S11**) (Supplementary Results; Supplementary Methods). Overall, we find weak evidence for bimodality in our dataset, only in some conditions, and only at very high variability ranks. Thus, we infer that higher variability is likely primarily driven by a wider range of expression centered around a single peak, rather than by bimodality of expression.

Gene expression variability is associated with gene function

After applying a strong correction to remove the effect of mean expression, we verified that our estimate of gene expression variability retains a functionally relevant signal. For this, we performed a Gene Ontology (GO) enrichment analysis of GO biological process terms repeatedly represented ($n \geq 3$ conditions) in genes

with lowly or highly variable expression. We observed that sets of lowly (respectively highly) variable genes show a substantial overlap in enriched GO biological process categories across several conditions. Underdispersed genes (variability rank ≤ 0.20) are consistently enriched in core cellular (housekeeping) functions such as protein localization and transport, protein modification, RNA processing, and organelle organization (**Figure 3A**). Conversely, overdispersed genes (variability rank ≥ 0.80) are enriched in functions such as signaling, chemotaxis, response to stimuli, and immune response (**Figure 3B**). Functional categories that are under-represented in lowly variable genes also tend to be over-represented in highly variable genes, and vice versa (**Figure S18**). We also performed a GO enrichment analysis for moderately dispersed genes (0.40 ≤ variability rank ≤ 0.60) but observed virtually no over-represented terms, and under-representation of terms related to defense response, immune response, and muscle movement (**Figure S19**). These findings are consistent with previous studies in humans [39,40] and *Drosophila* [25] which have shown that lowly and highly variable genes are enriched in distinct functional categories.

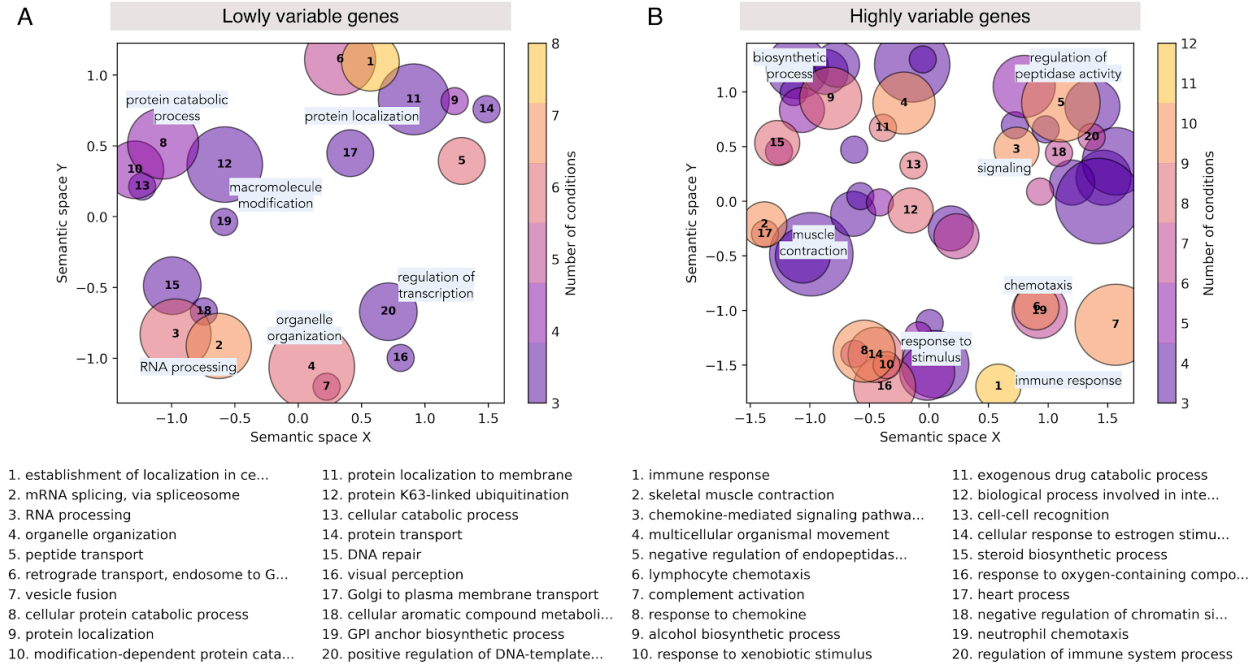


Figure 3. Semantic similarity bubble plots of enriched Gene Ontology (GO) biological process terms for (A) underdispersed (variability rank ≤ 0.20) and (B) overdispersed (variability rank ≥ 0.80) genes in zebrafish organs. Over-representation analysis of GO terms was performed separately for the set of under- and overdispersed genes in each organ-sex condition using the conditional hypergeometric test in *GOstats* **[50]**. Terms that are significant (unadjusted *p*-value < 0.01) in at least three organ-sex conditions were retained. Redundancy reduction and visualization of terms was performed using *GO-Figure!* **[51]**. Each circle represents a group of similar GO terms collapsed into a representative term. Representative terms are plotted in a semantic similarity space such that more similar terms are positioned closer together. The top 20 representative terms are listed and sorted by the number of organ-sex conditions in which the representative is significant. Circle color indicates the number of conditions in which a representative term is significant while circle size is scaled by the number of related terms summarized in a cluster.

Variably expressed genes evolve under weaker purifying selection

We assessed whether inter-individual gene expression variability is associated with the strength of natural selection on protein sequence by comparing evolutionary rates of genes with consistently low, moderate, or high expression dispersion across multiple organ- or organ-sex conditions within a species. We retrieved precomputed selection statistics from Selectome [52,53], a database of conservative branch-site likelihood tests for positive selection conducted on vertebrate gene trees (Materials and Methods). We found that, for all three fish species, our estimate of gene expression variability is inversely related to the strength of purifying selection, i.e., frequently underdispersed genes (variability rank $\leq$ 0.20) have evolved under stronger negative selection, whereas frequently overdispersed genes (variability rank $\geq$ 0.80) have evolved under more relaxed selective constraints (**Figure 4**). In addition, we observed a statistically significant association between expression variability categories and the proportion of branches with evidence for positive selection in zebrafish (Pearson's chi-squared test, χ^2= 21.866, p = 1.786 x 10^{-5}) and Northern pike (χ^2 = 48.122, p = 3.552 x 10^{-11}), but not in spotted gar (χ^2= 1.4127, p = 0.4935). Specifically, we found depletion of branches under positive selection for lowly variable genes in both zebrafish and pike, as well as enrichment of positive selection for highly variable genes in pike (**Table S4**). Thus, higher gene expression variability between individuals is associated with faster protein-coding sequence evolution between species, both due to relaxed negative selection and to increased positive selection.

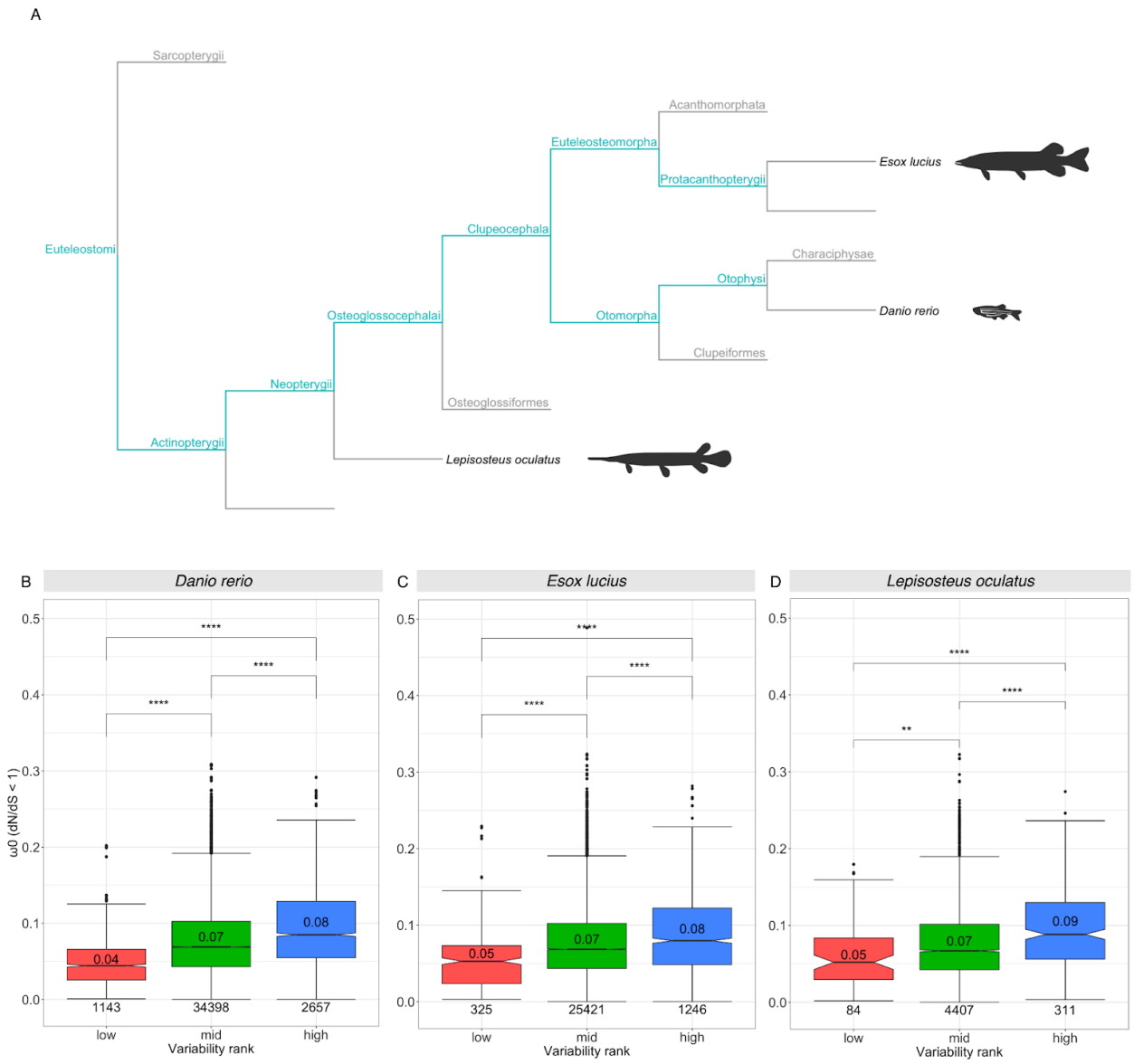


Figure 4. Strength of purifying selection of genes classified by estimated expression variability for zebrafish, Northern pike, and spotted gar. Selection statistics based on conservative branch-site likelihood tests were retrieved from the Selectome database **[52,53]**. **(A)** Species phylogeny with tested internal branches highlighted (turquoise). Tested branches included in all three species (B,C,D) are: Euteleostomi, Actinopterygii, Neopterygii; for teleosts, i.e. zebrafish and pike, only (A,B): Osteoglossocephalai, Clupeocephala; for zebrafish only (A): Otomorpha, Otophysi; and for pike only (B): Euteleosteomorpha, Protacanthopterygii. Icons retrieved from PhyloPic: *Danio rerio* by Jake Warner (CC0 1.0), *Esox lucius* by Timothy Knepp and T. Michael Keesey (PDM 1.0), *Lepisosteus oculatus* by Ingo Braasch (CC0 1.0). Edge lengths are not scaled by divergence time. **(B-D)** Boxplots show the distribution of ratios of nonsynonymous substitutions (dN) to synonymous substitutions (dS) of sites that have evolved under purifying selection (ω_0 ratio, for dN/dS < 1), for sets of genes with low (red), medium ('mid'; green), and high (blue) expression dispersion across multiple organ or organ-sex conditions for **(B)** zebrafish, **(C)** Northern pike, and **(D)** spotted gar. The median ω_0 is indicated within each boxplot and the number of gene - internal branch pairs included are annotated below each. Variability rank ('rank') cutoffs for each category are as follows: rank ≤ 0.20 for low, rank ≥ 0.80 for high, and 0.20 < rank < 0.80 for moderate expression variability. Pairwise comparisons were performed using Wilcoxon rank-sum tests (*p*-value significance levels: '****': $p \leq 0.0001$, '***': $p \leq 0.001$, '**': $p \leq 0.01$, '*': $p \leq 0.05$, 'ns': $p > 0.05$).

Brain-biased genes have low expression variability across non-nervous organs

We have shown that genes which are lowly variable across individuals in multiple conditions are enriched in housekeeping functions, while highly variable genes are enriched in stimulus-sensitive functions. Given that housekeeping genes are, by definition, ubiquitously expressed across conditions (but see [54] for a discussion), it follows that the association between expression variability and gene function also holds in the reverse direction, i.e., we expect genes that are broadly expressed across organs to have low within-organ expression variability across individuals, reflecting stronger pleiotropic constraints. To test this hypothesis, we quantified expression breadth per gene using the expression specificity metric tau (τ), with range [0,1], in which genes with τ close to 0 are uniformly expressed across organs whereas genes with τ close to 1 are highly organ-specific [55,56] (Materials and Methods). Importantly, τ only considers mean expression differences between conditions and does not take within-condition expression dispersion into account, thus it is independent of our estimate of variability. Across conditions in each species, results consistently show a positive association between variability rank and organ specificity (**Figures S20-S22**). Based on the lower quartile of the τ distribution averaged across species, we designated genes with $\tau \leq 0.3$ as 'broadly expressed' and genes with $\tau > 0.3$ as 'organ-biased'. Note that we explicitly use the term 'organ-biased' rather than 'organ-specific' because our relaxed τ cutoff necessarily includes genes which are preferentially expressed in one organ but are still expressed in other organs, albeit at lower levels. Per organ, broadly expressed genes are generally underdispersed relative to organ-biased genes (**Figure 5, Figures S23-S25, Figures S26-S28**). These results are consistent with previous studies which suggest that expression variability is associated with both tissue and cell-type-specific expression [57,58].

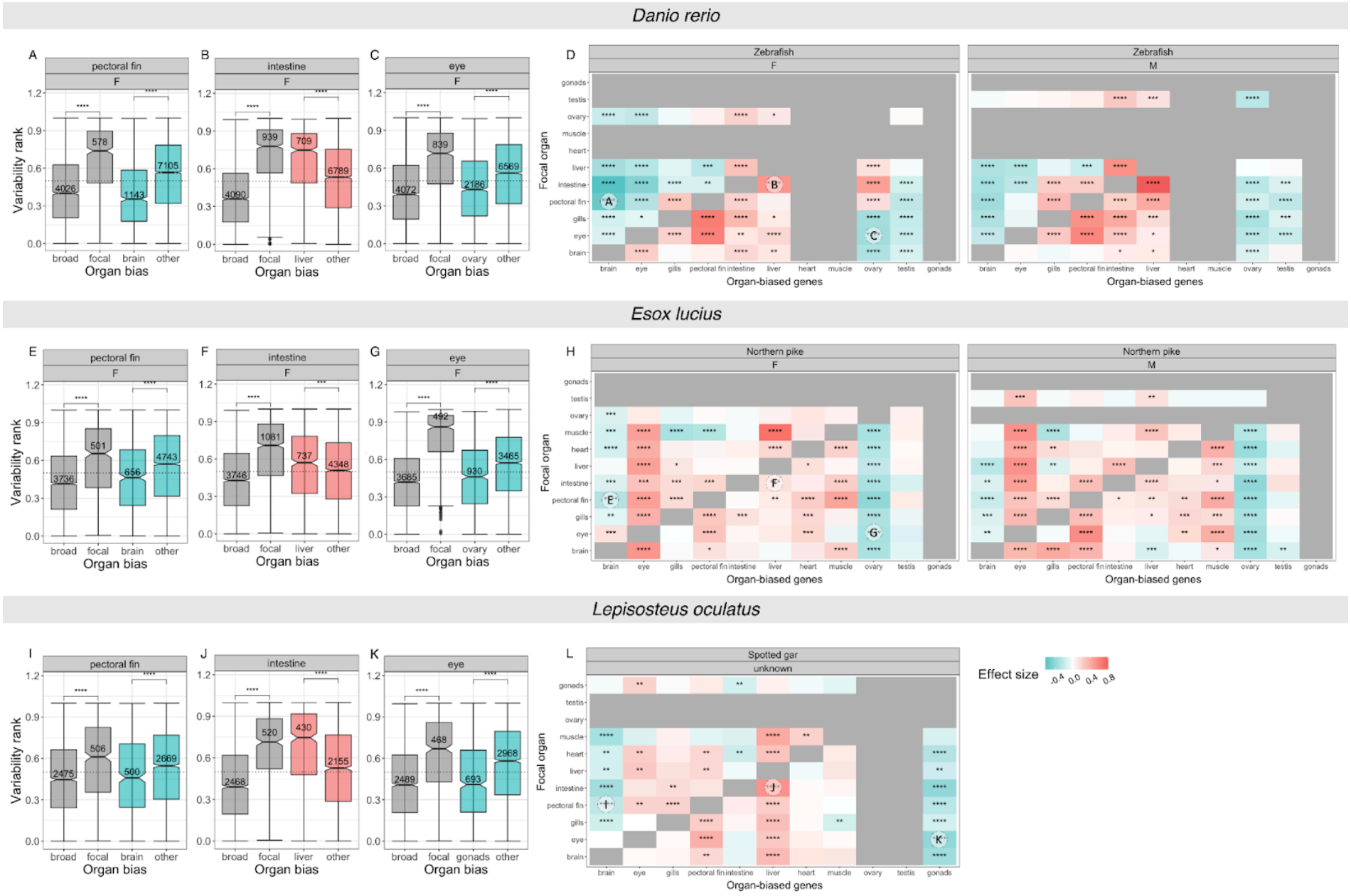
Since organ-biased genes are likely to have evolved under selection for an organ-specific function, one prediction is that these genes would have a narrower expression range within their top (focal) organ of expression, indicating more fine-tuned regulation for its putative primary function. On the contrary, we observed significant overdispersion of organ-biased genes within their focal organ (**Figure 5A-C, 5E-G, 5I-K; Figures S23-S25**, **Figures S26-S28, Figures S29-S31**). For example, genes which are most highly expressed in the pectoral fin are overdispersed within the pectoral fin, relative both to broadly expressed genes as well as to other organ-biased genes expressed in this organ (**Figure 5A, 5E, 5I**). This is consistent with a previous claim that highly variable genes within a cell type population are typically indicative of cell type function [59]. This pattern is lost when the top organ is assigned randomly, such that focal-biased genes have dispersion levels that are indistinguishable from the background of other organ-biased genes (**Figures S32-S34; Figure S35**). We infer that expression of organ-biased genes within their focal organ tends to be highly plastic in response to genetic or environmental cues, and that context-dependent upregulation of expression of organ-biased genes may be mediated by *trans*-acting factors that amplify variation.

Suboptimal expression can be more or less deleterious in different organs, due to differences in physiological function, cellular composition and turnover, and sensitivity to perturbations [60]. Thus, we expect differences in selective pressure on gene regulation depending on the focal organ of each gene. In

279 particular, it has been suggested that neural tissues are highly sensitive to protein misfolding, and that this
280 sensitivity contributes to strong evolutionary constraints of genes expressed in the brain [61]. We confirm
281 that protein-coding sequences of brain-biased genes have evolved under stronger purifying selection not
282 only compared to other organ-biased genes, but also to broadly expressed genes **(Figure S36)**, which is
283 consistent with reports of strong negative selection on the mammalian brain relative to other organs
284 [62–66]. Interestingly, we find that brain-biased genes have consistently underdispersed expression across
285 non-nervous system organs relative to other organ-biased genes (**Figure 5D, 5H, 5L; Figures S23-S25**). This
286 suggests that genes involved in brain function may have evolved *cis*-regulatory mechanisms that facilitate
287 highly stable expression downregulation across other organs, independent of the organ's cell type
288 composition. By contrast, liver-biased genes, which evolve under weaker purifying selection at the
289 protein-coding sequence level (**Figure S36**) [65], tend to be overdispersed across multiple organs
290 (**Figures S26-Figure S28**), implying weaker global expression constraints. Although mammalian testis-specific
291 genes have been shown to be faster-evolving compared to other organ-specific genes [65], we observed
292 moderate rates of purifying selection on both testis- and ovary-biased genes in fishes (**Figure S36**).
293 Furthermore, we found that genes preferentially expressed in the gonads, especially the ovary, tend to be
294 underdispersed across organs (**Figures 5D, 5H, 5L**) **(Figures S29-S31).** We observed qualitatively similar
295 organ-dependent patterns of expression dispersion after controlling for differences in transcriptome
296 complexity (i.e., number of genes expressed) across organs [67–69] (**Figure S37**), with particularly robust
297 results in zebrafish (**Figure S38D**), although the effect is weaker in Northern pike (**Figure S38H**) and spotted
298 gar (**Figure S38L**). No pattern is recovered in any of the species when the top organ is assigned randomly
299 (**Figures S39-S41**), confirming that these results are biological and not due to biases in the data.

300 Overall, our findings show a generally positive association between inter-species protein sequence
301 divergence and intra-species expression variability (**Figure 4**), suggesting that evolutionary constraints at the
302 protein sequence level co-vary with organ-dependent patterns of expression variability. One interesting
303 exception involves eye-biased genes, which evolve under strong purifying selection across the three fish
304 species (**Figure S36**) but which tend to have overdispersed expression in pike and spotted gar (**Figure 5H, 5L**).
305 We suggest that the differential variability of eye-biased genes may be an interesting subject for future
306 research, particularly given the remarkable diversity of visual opsin gene repertoires in fishes and its relation
307 to differences in light environment and ecology [70].

308 Overall, we infer that (1) organ-biased genes are highly variable within their focal organ, likely due to
309 context-sensitive expression, but that (2) their global variability reflects organ-dependent evolutionary
310 constraints on both protein-coding and *cis*-regulatory sequences. We found that these organ-dependent
311 patterns of expression variability are largely consistent in all three sampled species, thus constituting an
312 evolutionarily conserved signal across at least 320 million years of evolution.


Danio rerio
Esox lucius
Lepisosteus oculatus
Variability rank
Organ bias
Focal organ
Organ-biased genes
Zebrafish
Northern pike
Spotted gar
Effect size

313

Figure 5. Relationship between expression variability and organ-biased expression for (A-D) zebrafish, (E-H) Northern pike, and (I-L) spotted gar. (A) Boxplot of expression variability ranks of genes expressed in female ('F') zebrafish pectoral fin, for sets of genes classified by organ bias and with an emphasis on brain-biased genes. Broadly expressed genes ('broad') have a measured organ expression specificity index (tau) of $\tau \leq 0.3$; genes with $\tau > 0.30$ are considered organ-biased. Among the set of organ-biased genes, genes are further classified into three categories: focal-biased genes ('focal') are preferentially expressed in the focal organ (i.e., pectoral fin), brain-biased genes ('brain') are preferentially expressed in the brain, and the remaining ('other') are preferentially expressed in other organs. The number of genes included are annotated for each category. Comparisons between broadly expressed and focal-biased genes, and between brain-biased and other organ-biased genes were performed using Wilcoxon rank-sum tests (*p*-value significance levels: '****': $p \leq 0.0001$, '***': $p \leq 0.001$, '**': $p \leq 0.01$, '*': $p \leq 0.05$, 'ns': $p > 0.05$). The dashed line indicates a variability rank of 0.5. **(B)** Boxplot of variability ranks of genes expressed in female zebrafish intestine, with genes classified by organ bias and with an emphasis on liver-biased genes. **(C)** Boxplot of variability ranks of genes expressed in female zebrafish eye, with genes classified by organ bias and with an emphasis on ovary-biased genes. **(D)** Heatmap summarizing pairwise comparisons of expression variability ranks between subsets of organ-biased genes for female ('F', left) and male ('M', right) zebrafish. Rows correspond to focal organ whereas columns correspond to subsets of organ-biased genes**.** Pairwise comparisons were performed using Wilcoxon rank-sum tests with Benjamini-Hochberg ('BH') correction for multiple testing (adjusted *p*-value < 0.05). Effect size was computed using Glass's delta (δ), with color scale shown in (L). Cells in gray are not included in the dataset. Three cells are annotated, corresponding to the following comparisons in boxplots (A-C): (A) brain-biased vs. other organ-biased genes expressed in female zebrafish pectoral fin, (B) liver-biased vs. other organ-biased biased genes expressed in female zebrafish intestine, and (C) ovary-biased vs. other organ-biased genes expressed in female zebrafish eye. A negative effect size entails lower expression variability relative to other organ-biased genes (A,C) (turquoise), whereas a positive effect size entails higher variability (B) (salmon). **(E-H)** Corresponding plots for Northern pike, with cells annotated corresponding to comparisons in boxplots (E-G). **(I-L)** Corresponding plots for unsexed spotted gar, with cells annotated corresponding to comparisons in boxplots (I-K).

Discussion

Gene expression variability is a functionally relevant trait that has been implicated in disease [34–36] and drug efficacy [37], aging [38], circadian rhythm [71], cellular differentiation [30–32], embryogenesis [33,72], developmental canalization [49], and other biological processes. Most gene expression studies have focused on testing for differences in average expression levels between groups of interest, largely to the exclusion of expression variation and variability. Here, we conducted a large-scale comparative analysis of inter-individual gene expression variability across multiple adult organs in three species of ray-finned fishes. We confirm that genes which are lowly variable among individuals are typically enriched in housekeeping functions, whereas highly variable genes are enriched in stimulus-sensitive functions, as reported previously in other species [25,39,40,58]. This biologically intuitive result reaffirms that expression variability is a functionally relevant trait independent of average expression. There is a growing body of literature that recognizes the functional relevance of expression variability [41–43] and how expression variability is associated with promoter characteristics such as sequence, core promoter motifs, and transcriptional start site architecture [25,58,73]. However, there has been little attempt thus far to characterize how gene expression variability in animals is systematically modulated by the interplay between gene function and organ context, even in studies that sample diverse organs. To our knowledge, this is the first study that investigates how gene expression variability is influenced both by expression specificity and organ context.

Our cross-organ approach helps clarify the connection between two previously reported findings. First, although lowly and highly variable genes show distinct functional signatures, some studies report that gene expression variability tends to be tissue-specific (i.e., organ-specific) [40], with little overlap in the set of most overdispersed genes between organs [38,39] or cell types [59] (but see [25]). Second, organ- or

358 cell-type-specific genes are found to have highly variable expression [57–59]. We verify that organ-biased
359 genes are more variably expressed relative to broadly expressed genes, but in addition, we have also found
360 that organ-biased genes are typically highly variable within their primary (focal) organ. Since the most
361 variable genes within an organ are typically genes which are biased or specific to it, we think it is unsurprising
362 to find few highly variable genes shared between multiple conditions. We infer that plastic expression within
363 a focal organ is a functional feature of more specialized genes, and that expression is dynamically induced by
364 context-specific regulatory cues, such as organ-specific transcription factor binding to enhancer elements or
365 enhancer-promoter interactions. Promoters of specialized genes (e.g., developmental genes) have been
366 shown to interact with different classes of enhancers compared to those of housekeeping genes [74]. In
367 addition, such specialized genes tend to have more enhancer-promoter interactions than housekeeping
368 genes [26,75], and the number of interacting enhancers has been found to have an additive effect on gene
369 expression levels [26,76,77]. Thus, to augment transcription of organ-biased genes in the relevant context,
370 we expect an increase in regulatory inputs, which introduces more sources of expression variation. By
371 contrast, outside of their focal organ, organ-biased genes have basal expression levels that are relatively
372 more stable among individuals.

373 Furthermore, we find striking evidence across species that the degree to which biased genes are
374 stably downregulated across non-focal organs reflects organ-dependent evolutionary constraints. Most
375 notably, we observe consistently underdispersed expression of brain-biased genes across non-nervous
376 organs, despite these organs having very different transcriptional repertoires, cell-type compositions, and
377 local micro-environments. Low variability within a condition suggests stabilizing selection on expression,
378 which is consistent with strong purifying selection on *cis*-regulatory sequences, and/or with buffering of
379 expression variation by compensatory *cis-trans* effects [78]. Given that we observe consistent
380 underdispersion across multiple organs having very different local micro-environments, we reason that stable
381 basal expression of brain-biased genes is likely to be predominantly driven by strong purifying selection at
382 the *cis*-regulatory level rather than by context-dependent *trans* effects.

383 We recognize that comparing gene expression variability between organs is challenging because any
384 measured changes in dispersion may simply be due to technical variation (e.g., sampling design, sequencing
385 depth or RNA quality [37]) or 'extrinsic' sources of biological variation (e.g., cell type composition [57],
386 transcriptome complexity [67,68]). Indeed, a limitation of bulk RNA-seq is that it does not allow us to
387 distinguish between multiple sources of variation within an organ, whether it is expression variation within a
388 cell type population, expression variation between cell type populations, or heterogeneity in cell type
389 composition between individuals. Observing a consistent signal of underdispersion across conditions with
390 heterogeneous backgrounds is thus particularly informative and suggests that our method captures a signal
391 of stable *cis*-mediated downregulation that is independent of organ-context. We infer that brain-biased
392 genes have evolved under strong purifying selection, not only at the protein-coding sequence level, but also
393 at the *cis*-regulatory level. From this, we hypothesize that genes selected for function in the brain may have

394 co-evolved with more stringent 'enhancer grammar' (e.g., number, type, order, affinity, spacing, or
395 orientation of transcription factor binding sites) [79,80] relative to other organ-biased genes. Our results
396 highlight how analysis of gene expression variability across organs can yield additional insights into gene
397 regulatory strategies in animals. Extending on these findings, we propose that intra-specific expression
398 variability can potentially be informative of broader macro-evolutionary trends, such as patterns of duplicate
399 gene retention and functional divergence after whole-genome duplication.
400

401 Materials and Methods

402 Sample preparation

403 *Sample collection*

404 Zebrafish

405 Thirteen female and twelve male zebrafish (*Danio rerio*), strain AB, were raised at the University of Oregon
406 Aquatic Animal Care Services. Animals were raised in E5 medium [81] and fed rotifers from 4-10dpf [82].
407 After 10dpf, fish were moved to a circulating fish water system and fed Zeigler Adult Zebrafish Diet.
408 Three-month old adults were euthanized by cold shock [81], and organs were dissected and immediately
409 stored in RNAlater. Procedures were approved by the University of Oregon IACUC protocol #18-13.
410

411 Northern pike

412 Ten female and ten male one-year-old Northern pike (*Esox lucius*) were sourced from a commercial hatchery
413 (Aqua 2B, Pouancé, France). The fish were euthanized using a lethal dose of tricaine methanesulfonate (MS
414 222, 400 mg/L), supplemented with 150 mg/L of sodium bicarbonate. Following euthanasia, organs were
415 immediately dissected and preserved in RNAlater. All animal procedures were conducted in full compliance
416 with French and European legislation (French decree 2013-118 and European Directive 2010-63-UE),
417 governing the ethical use and care of laboratory animals in scientific research.
418

419 Spotted gar

420 Spotted gars (*Lepisosteus oculatus*) were obtained as embryos from hormone-induced spawns of wild-caught
421 broodstock from bayous near Thibodaux, Louisiana and raised at Michigan State University for 23 or 31
422 months (April 2018 - March/November 2020) in 150 to 300 gallon tanks on a 12h light : 12h dark cycle and a
423 diet starting from *Artemia* brine shrimp to feeder fish (mostly fathead minnows). Twenty-two (22) unsexed,
424 pre-mature individuals of 20.1 - 23.3 cm standard length (tip of snout to caudal peduncle) and 24.8 - 29.0cm
425 total length (tip of snout to tip of caudal fin) were euthanized in 300 mg/L MS-222 (Sigma). Tissues were
426 dissected and stored in RNAlater (Invitrogen) at -80°C until shipped frozen to the University of Lausanne. All
427 spotted gar experiments were approved by the Institutional Animal Care & Use Committee (IACUC) of
428 Michigan State University (animal use protocol PROTO201900309).
429

430 *RNA extraction*

431 Tissue samples were thawed on ice, removed from RNA*later* RNA Stabilization Solution (Invitrogen, catalog
432 no. AM7024) and transferred to new Eppendorf tubes containing 1 mL TRIzol Reagent (Invitrogen, ref.
433 15596018) and about 10 - 15 SiLi beads (Type ZS, 1.4 - 1.6 mm) (Sigmund Lindner, ref. 9315-33). Excluding
434 blood, samples were homogenized at 4°C using PreCellys Evolution (Bertin Technologies, ref. P000062
435 _PEV0) for at least 2 cycles at 6600 rpm for 30 s. Phase separation of homogenized samples was performed
436 following the TRIzol Reagent user guide (MAN0001271). For Northern pike and spotted gar, RNA extraction
437 with TRIzol Reagent was carried out following the same manual. After washing and resuspending RNA,
438 samples were treated with RNase-free DNase I (Thermo Scientific, catalog no. EN0525) to remove DNA in
439 accordance with the user guide (MAN0012000). For zebrafish, column-based RNA extraction with TRIzol was
440 performed using the PureLink RNA Mini Kit (Invitrogen, ref. 12183018A), following manufacturer's
441 instructions (MAN0000406). DNA digestion during on-column RNA purification was done using the PureLink
442 DNase Set (Invitrogen, ref. 12185010). All samples were eluted in 50 - 60 µl RNase-free water. Following
443 extraction, RNA concentration was quantified with the QuantiFluor RNA System (Promega, ref. E3310). A
444 subset of samples was also assessed for RNA quality using the 5200 Fragment Analyzer System (Agilent
445 Technologies, M5310AA) with PROSize 3.0.1.6 data analysis software (Agilent Technologies). Samples with
446 low RNA concentration or signs of degradation were removed from subsequent steps. RNA preparations
447 were diluted to a minimum concentration of 50 ng/µl, transferred to low-binding 96-well plates, and stored
448 in -80°C prior to sequencing.

449

450 **Data generation**

451 *Bulk RNA barcoding and sequencing (BRB-seq)*

452 Samples were sent to Alithea Genomics (Lausanne, CH) for library preparation and sequencing using highly
453 multiplexed 3′-end bulk RNA barcoding and sequencing (BRB-seq) [44] over three phases (corresponding to
454 projects AG0012, AMP0027, and AMP0020) (**Table S1**) (Supplementary Data). BRB-seq libraries were
455 prepared using the MERCURIUS BRB-seq library preparation kit for Illumina and following the manufacturer's
456 manual (Alithea Genomics, #18013). For project AG0012 (Northern pike, spotted gar), libraries were
457 sequenced using Illumina NovaSeq 6000, generating multiplexed, asymmetric paired-end reads, where read
458 R1 (25 bp) contains the sample barcode and UMI and read R2 (80 bp) contains the cDNA fragment sequence
459 to be aligned to the reference genome. For projects AMP0027 (spotted gar) and AMP0020 (zebrafish),
460 libraries were paired-end sequenced using Illumina NextSeq 550 (R1 = 21 bp, R2 = 55 bp).

461

462 *Read alignment and quantification*

463 Read trimming, alignment, and quantification steps were performed by Alithea Genomics using STARsolo
464 (preprint) [83] from STAR 2.7.9a [84]. R2 reads were trimmed ("--clipAdapterType CellRanger4") and aligned
465 to the following reference genomes: for spotted gar, LepOcu1 (Ensembl 103 [85]); zebrafish, GRCz11

466 (Ensembl 101 [85]); Northern pike, Eluc_v4 (Ensembl 101). To generate both raw and UMI-deduplicated
467 counts, the following parameter was used: "--soloUMIdedup NoDedup 1MM_Directional". UMI-deduplicated
468 count matrices were used for downstream analyses.
469
470 **Data analysis**
471 For subsequent steps, data analyses were performed using R 4.1.0 [86], Bioconductor 3.13 [87], and RStudio
472 2023.03.0+386 [88] unless specified otherwise.
473
474 *Sample filtering*
475 Sample libraries were filtered in two main steps independently for each species. First, samples with low
476 sequencing quality were removed based on a minimum number of uniquely mapped reads or number of
477 genes with nonzero counts ("detected genes"), following Liu et al. [33]. For example, in zebrafish, samples
478 with less than 500,000 uniquely mapped reads or less than 10,000 detected genes were excluded from
479 downstream analysis. To account for differences in reference genome annotation quality, we selected
480 species-specific minimum uniquely mapped reads and detected genes. (See **Table S2** for filtering criteria.) For
481 each species, we visually inspected the plot of detected genes against uniquely mapped reads to estimate
482 the point in which the relationship between the two variables becomes nonlinear and set this point as the
483 minimum threshold (**Figure S1A-C**). We verified that samples which do not meet specified minimum criteria
484 are visible outliers on a principal component analysis (PCA) (**Figure S1D-F**).
485 Second, samples that were poorly correlated with other samples within its annotated organ were
486 removed. We performed iterative removal of lowly correlated samples based on a protocol described by
487 Fukushima & Pollock [89]. For each iteration, we first generated the all-against-all pairwise Pearson
488 correlation matrix of retained samples based on gene-filtered and normalized expression data in the form of
489 log-transformed, TMM-normalized counts per million ($\log_2$ TMM-CPM) (Gene filtering and normalization).
490 Then, for each sample, we computed its mean Pearson correlation with each organ. We removed any sample
491 which is more highly correlated with another organ than the one in which it is annotated to. This process was
492 repeated until no more samples fit removal criteria. Considering the set of protein-coding genes in each
493 species, we verified that all retained samples cluster by organ based on hierarchical clustering of the pairwise
494 Pearson correlation matrix (**Figure 1B, Figure S2**) and PCA of gene-filtered and normalized expression
495 (**Figure S3**). Gene biotype information was retrieved for each species from Ensembl 105 [90].
496 For computing organ expression specificity (Organ expression specificity), we required a minimum of
497 1 replicate per organ-sex condition for Northern pike and zebrafish, and 2 replicates per organ (only)
498 condition for spotted gar (**Table S2**). For estimating expression variability (Gene expression variability) and
499 testing its association with gene function (Gene Ontology (GO) statistical over- and under-representation
500 analysis) and coding-sequence selection (Selection analysis), we required a minimum of 4 replicates per
501 organ (for spotted gar) or organ-sex condition. Any conditions which did not meet the minimum number of

502 replicates after quality filtering were excluded from downstream analysis (**Table S3**). See **Tables S2-S3** for
503 samples retained after each step of the sample filtering pipeline and **Figure 1A** for a complete breakdown of
504 the number of replicates retained per condition.
505
506 *Gene filtering and normalization*
507 At the start of each step of the sample filtering pipeline (Sample Filtering), we removed non-expressed genes
508 from the full raw count matrix by retaining only genes with both mean and median counts per million (CPM)
509 greater than 1.0 in at least one organ. From the gene-filtered raw count matrix, we performed library size
510 normalization over all conditions using the trimmed mean of M-values (TMM) method in *edgeR* (3.34.1) [91],
511 then re-converted counts to CPM and transformed CPM to $\log_2$-scale ($\log_2$ TMM-CPM). To remove negative
512 values, $\log_2$ TMM-CPM was vertically offset by the absolute value of the global expression minimum such
513 that the minimum value in the normalized expression matrix is always 0.
514
515 *Gene expression variability estimation*
516 To estimate gene expression variability within an organ (for spotted gar) or organ-sex condition (for Northern
517 pike and zebrafish), we considered only the set of conditions with at least 4 replicates after sample filtering.
518 Gene filtering and normalization steps were performed as described in the previous section, but with each
519 condition treated separately. As an additional filter to remove lowly expressed genes specific to a condition,
520 we removed genes with mean or median $\log_2$ TMM-CPM < 1.0. Since actual expression variation may be
521 sensitive to outliers, we used a jackknife resampling ('leave-one-out') procedure to compute expression
522 dispersion statistics and generate a less biased estimate of gene expression variability. That is, for a given
523 condition with n replicates, we formed all n possible subsets of n - 1 replicates and computed expression
524 dispersion metrics for each subset, then computed the mean for each statistic over all subsets. We initially
525 considered different measures of expression dispersion (Supplementary Methods), with the aim to correct
526 for the strong covariation between expression variation and average expression (**Figure S4**). From the metrics
527 we considered, we selected the local (residual) coefficient of variation (LCV) [27,37] as our estimate of
528 expression variability, as it maximally removes the association between mean expression level and
529 expression dispersion (**Figure 2C; Figures S5-S7**). The LCV was computed in two steps then averaged over all
530 n - 1 subsets, as described below.
531
532 <u>Residual coefficient of variation</u>
533 The coefficient of variation (CV) is the ratio of the standard deviation to the mean. First, we computed the
534 squared CV and applied a $\log_2$ transformation ($\log_2(CV)^2$) for each gene, following Faure et al. [27]. Next, using
535 the 'loess' function from the *stats* package in R, we fit a LOESS regression curve (with span = 0.6 and default
536 parameters) to model $\log_2(CV)^2$ as a function of mean expression level. The difference between observed and

537 predicted $\log_2(CV)^2$ is the residual $\log_2(CV)^2$ ('residual CV', for brevity) after adjusting for expression level.
538 Genes with residual CV > 0 overdispersed, while genes genes with residual CV < 0 are underdispersed.
539
540 Local (residual) coefficient of variation (LCV)
541 Given residual CVs for each gene, we performed a percentile rank normalization step following the 'local
542 coefficient of variation' [37,49] or 'relative noise rank' [27] metric recommended in previous publications.
543 First, genes were sorted by mean expression level. Next, we defined a sliding window of 100 neighboring
544 genes and a step size of 1, with a focal gene at the center of each window. At each iteration of the sliding
545 window, the residual CV of the focal gene was ranked against the residual CVs of genes within the window, on
546 a percentile rank scale of 0.00 (lowest dispersion) to 1.00 (highest dispersion). We computed the LCV statistic
547 only for complete windows; as such, for each *n* - 1 subset, we do not have variability estimates for the 50
548 most lowly and 50 most highly expressed genes.
549
550 Summarizing jackknife results
551 Since transcriptome-level expression distributions are positively skewed, with the most highly expressed
552 genes spanning a wide expression range, expression variability tends to be poorly estimated for outlier genes
553 with very high expression. As such, for each condition, we removed the top 5% most highly expressed genes
554 based on jackknifed mean expression level. Then, we computed the mean LCV over all *n* - 1 subsets and used
555 this metric, which we call 'variability rank', as our estimate of expression variability.
556
557 *Gene Ontology (GO) statistical over- and under-representation analysis*
558 For each organ-sex condition in zebrafish, we retrieved three sets of genes categorized by expression
559 variability: genes with variability rank ≥ 0.80 (highly variable), genes with rank ≤ 0.20 (lowly variable), and
560 genes with 0.40 ≤ rank ≤ 0.60 (moderately variable). For each variability category per condition, we used
561 *GOstats* (2.58.0) [50] to conduct conditional hypergeometric tests for statistical over- and
562 under-representation of GO biological process terms, based on the org.Dr.eg.db (3.13.0) zebrafish genome
563 annotation [92]. Per test, the background gene set was defined as the set of genes which have expression
564 variability estimates within the organ-sex condition. For each variability category, we retained GO terms that
565 were significantly over-represented (or under-represented) (unadjusted *p*-value < 0.01) in at least three
566 conditions. Redundancy reduction and visualization of GO terms was done using the Python package
567 *GO-Figure!* (1.0.0) [51] (**Figure 3, Figures S18-S19**).
568
569 *Coding-sequence selection analysis*
570 Given a list of protein-coding genes (Ensembl 105 [90]) and taxonomic branches for each species, we
571 retrieved precomputed selection statistics from Selectome (https://selectome.org/), a database of evidence
572 for positive selection based on a conservative branch-site likelihood test applied to *Euteleostomi* gene trees

573 (Ensembl 98) [52,53]. For each gene - internal branch (i.e., taxonomic clade), the Selectome dataset provides
574 a measure of selective pressure for sites under purifying selection (ω_0 ratio for sites with dN/dS < 1), as well
575 as a likelihood-ratio test statistic for positive selection. The tested internal branches for each species are as
576 follows. For spotted gar: Euteleostomi, Actinopterygii, Neopterygii. For zebrafish: Euteleostomi,
577 Actinopterygii, Neopterygii, Osteoglossocephalai, Clupeocephala, Otomorpha, Otophysi. For Northern pike:
578 Euteleostomi, Actinopterygii, Neopterygii, Osteoglossocephalai, Clupeocephala, Euteleosteomorpha,
579 Protacanthopterygii (**Figure 4A**). We included only distinct gene - internal branch entries in our downstream
580 analysis, as duplicate gene - internal branch entries may be a result of uncertainties in the gene tree.
581 For each species, we formed a background gene set consisting of protein-coding genes that are
582 included in the Selectome dataset, and which have expression variability data in at least half of the available
583 conditions for the species. Within each organ (for spotted gar) or organ-sex (for Northern pike and zebrafish)
584 condition, genes were grouped based on the following variability rank thresholds: rank ≤ 0.20 for lowly
585 variable genes, rank ≥ 0.80 for highly variable genes, and 0.20 < rank < 0.80 for moderately variable genes.
586 From the background gene set, we identified genes with consistently lowly, highly, or moderately variable
587 expression in more than half of the conditions in which the particular gene has expression variability data.
588 Then, we compared the ω_0 distribution of the three gene categories to test for any association between gene
589 expression variability and purifying selection on protein-coding sequences (**Figure 4**). In addition, for each
590 gene category, we compared the proportion of gene - internal branches with evidence for positive selection
591 based on a *q*-value cutoff of $q < 0.05$ (**Table S4**) on the branch-site likelihood test. To check for an association
592 between variability category and positive selection, we conducted a randomization test by permuting the
593 *q*-values on the background gene list ($n = 2000$ permutations), in addition to a Pearson's chi-squared test.
594
595 *Organ expression specificity*
596 Data preprocessing
597 To compute organ expression specificity within a species, we required at least 2 biological replicates per
598 organ for spotted gar, and at least 1 biological replicate per organ-sex condition for Northern pike and
599 zebrafish. Furthermore, to make expression specificity measurements comparable between species, we
600 considered only the set of organs with sufficient samples in every species, i.e., brain, eye, intestine, liver,
601 heart, muscle, gills, pectoral fin, and gonads. After excluding organs with missing data (blood, kidney, skin,
602 swim bladder), we repeated our gene filtering and normalization step starting from raw read counts (Gene
603 filtering and normalization).
604
605 Expression specificity index
606 For each retained gene per species, organ expression specificity was measured using the tau (τ) index, which
607 is defined as:

$$\tau = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}\left(1 - \frac{\bar{x}_i}{\bar{x}_{max}}\right)$$

where n is the number of organs, $\bar{x}_i$ is the mean expression of gene x in organ i, and $\bar{x}_{max}$ is the maximum mean expression of gene x within an organ $1 \leq i \leq n$ [55,56]. The tau index gives a single measure of organ specificity per gene, ranging from broadly expressed across all organs ($\tau \rightarrow 0$) to specifically expressed in a single organ ($\tau \rightarrow 1$). For species with differentiated sexes (Northern pike and zebrafish), the ovary and testis were considered separate anatomical entities. However, for non-reproductive organs in the same species, we did not consider each sex separately.

Organ-biased expression and gene expression variability

Per species, genes with tau (τ) ≤ 0.30 were classified as broadly expressed, while genes with tau above this threshold were classified as organ-biased. For genes which are organ-biased, the 'focal organ' is defined as the organ in which the gene is maximally expressed.

<u>Multiple pairwise comparisons</u>

For each species (spotted gar) or species-sex (Northern pike and zebrafish), let S_1 be the set of m organs which (a) have been used in expression specificity computations, and (b) have gene expression variability estimates (**Figure 1**):

$$S_1 = \{1, 2, \ldots, m\}.$$

Let S_2 be the set of $n \geq m$ organs which satisfy (a) but not (b):

$$S_2 = \{1, 2, \ldots, m, \ldots, n\}.$$

For each organ $k \in S_1$, we compared the variability rank distribution within k for broadly expressed genes ($\tau \leq 0.30$) versus genes which are organ-biased ($\tau > 0.30$) and with top expression in focal organ k. For example, given the pectoral fin as the focal organ, we compared the variability ranks of broadly expressed genes against pectoral-fin-biased genes (**Figure 5A, 5E, 5I**).

Next, still given k, we iterated through the set of other (non-focal) organs in S_1 which are not k:

$$\{i \in S_1 \mid i \neq k\}.$$

For each iteration of $i \neq k$, we defined a set of i-biased genes, which are organ-biased and with top expression in non-focal organ i, but still expressed in focal organ k. Then, we compared the variability rank distribution within k for genes which are i-biased versus other organ-biased genes with top expression in the set J of (non-focal) organs in S_2 which are neither k nor i:

$$J = \{j \in S_2 \mid j \neq k, j \neq i\}.$$

As an example, for expression variability in the pectoral fin, we compared the variability ranks of brain-biased (or liver-biased, ovary-biased, etc.) genes expressed in the pectoral fin against the variability ranks of the set

641 of other organ-biased genes that are neither pectoral-fin-biased nor brain-biased (or liver-biased,
642 ovary-biased, etc.), and are also expressed in the pectoral fin (**Figure 5A, 5E, 5I**).

643 For each pairwise comparison of variability ranks of i-biased versus (other-biased) J-biased genes,
644 iterated for each organ $k \in S_1$, we tested for a difference in group means using Wilcoxon-rank sum tests, and
645 applied a Benjamini-Hochberg ('BH') correction for multiple testing. In addition, we calculated the effect size
646 using Glass's δ, using the set of J-biased genes as the reference group:

647 $$\delta_{k,i} = \frac{\bar{v}_{k,i} - \bar{v}_{k,J}}{s_{k,J}}$$

648 where $\bar{v}_{k,i}$ is the mean variability rank of i-biased genes in focal organ k, $\bar{v}_{k,J}$ is the mean variability rank of J
649 -biased genes, and $s_{k,J}$ is the sample standard deviation of the variability ranks of J-biased genes. We
650 visualized the distribution of effect sizes using a heatmap, where rows correspond to focal organ k and the
651 columns correspond to sets of i-biased genes (**Figure 5D, 5H, 5L**). We showed pairwise comparisons that are
652 considered statistically significant after BH multiple testing correction (adjusted *p*-value < 0.05).

653

654 <u>Analysis on randomized data</u>

655 Given the gene x organ mean expression matrix of a species, we permuted the values of the mean expression
656 vector per gene so that the focal organ of each gene is assigned randomly without changing tau. Since tau is
657 unchanged, the same set of genes are classified as broadly expressed ($\tau \leq 0.30$) and organ-biased, but the
658 subclassifications of organ-biased genes (e.g., brain-biased, liver-biased, etc.) differ. We pre-set two seeds
659 (12345, 67890) for our sample randomizations. Multiple pairwise comparisons of expression variability ranks
660 for subsets of organ-biased genes were performed as previously described (**Figure S35**).

661

662 <u>Analysis on nonzero expression matrix</u>

663 To control for differences in number of genes expressed among organs (**Figure S37**), we repeated the same
664 analysis (i.e., gene expression normalization, expression variability estimation, comparisons of expression
665 variability between subsets of genes) as previously described, considering only genes that have nonzero
666 counts across all samples within a species (**Figure S38**).

667

668 *Organ-biased expression and coding-sequence selection*

669 Per species, we retrieved expression specificity measurements for each gene (Organ expression specificity)
670 and filtered selection statistics for each gene - internal branch (Coding-sequence selection analysis), as
671 described in previous sections. We annotated each gene from the Selectome dataset with its top (focal)
672 organ of expression (i.e., organ bias). If a gene has $\tau \leq 0.30$, its organ bias is labeled 'broad'. We compared
673 the ω_0 distribution of genes across organ bias categories to check for any association between focal organ
674 and strength of purifying selection at the coding-sequence level (**Figure S36**).

Author Contributions

M.R.-R., I.B., Y.G., and J.H.P. conceived the experimental design. C.F.B. and M.R.-R. conceived the design of data analysis. A.W.T., B.L.R., and I.B. sourced and dissected the spotted gar samples; C.A.W. and J.H.P., the zebrafish samples; J.B. and Y.G., the Northern pike samples. C.A. performed RNA extraction. S.M. provided Selectome data for positive selection analysis. C.F.B. performed the data analysis and generated the figures with input from M.R.-R. C.F.B. wrote the original draft of the manuscript with input from M.R.-R; C.F.B., A.W.T., J.B., I.B., Y.G., J.H.P., and M.R.-R wrote the final version of the manuscript. C.F.B., A.W.T., J.B., I.B., Y.G., J.H.P., and M.R.-R. contributed to the interpretation and discussion of results.

Acknowledgements

We would like to thank Allyse Ferrara and Quenton Fontenot (Nicholls State University, USA) for providing gar embryos and members of the Braasch Lab for help with gar husbandry. We acknowledge having used ChatGPT (GPT-4) (OpenAI, 2023, https://chat.openai.com) to improve the clarity of the manuscript. Any errors or omissions are our own.

Funding

This work was supported by the Swiss National Science Foundation (SNSF, grants 173048 and 207853) to M.R.-R.; the Agence Nationale de la Recherche, France (ANR) on the GenoFish project, 2016-2021 (grant No. ANR-16-CE12-003) to J.B., J.H.P., M.R.-R., and Y.G.; the National Institute of Health (NIH) (award #R01OD011116) to J.H.P.; and the National Science Foundation (NSF) (award #202916) to I.B..

Competing Interests

The authors declare no competing interests.

Data Availability

Demultiplexed FASTQ files generated via BRB-seq for zebrafish (192 samples), Northern pike (252 samples), and spotted gar (177 samples) will be made available in the NCBI Sequence Read Archive (SRA) prior to peer review. This manuscript is already being made available for rapid communication on the topic, and will be updated with the SRA information as soon as possible.

Data analysis scripts are available on GitHub:
https://github.com/christabel-bucao/fish-variability-across-organs

Additionally, scripts and supplementary data files are available on Zenodo:
https://doi.org/10.5281/zenodo.14063787

References

1. Kærn M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: From theories to phenotypes. Nat Rev Genet. 2005;6: 451–464. doi:10.1038/nrg1615
2. Carter B, Zhao K. The epigenetic basis of cellular heterogeneity. Nat Rev Genet. 2021;22: 235–250. doi:10.1038/s41576-020-00300-0
3. Hill MS, Vande Zande P, Wittkopp PJ. Molecular and evolutionary processes generating variation in gene expression. Nat Rev Genet. 2020. doi:10.1038/s41576-020-00304-w
4. Choi JK, Kim YJ. Epigenetic regulation and the variability of gene expression. Nat Genet. 2008;40: 141–147. doi:10.1038/ng.2007.58
5. López-Maury L, Marguerat S, Bähler J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nat Rev Genet. 2008;9: 583–593. doi:10.1038/nrg2398
6. Wagner GP, Altenberg L. Complex adaptations and the evolution of evolvability. Evolution. John Wiley & Sons, Ltd; 1996. pp. 967–976. doi:10.1111/j.1558-5646.1996.tb02339.x
7. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002;297: 1183–1186. doi:10.1126/science.1070919
8. Silander OK, Nikolic N, Zaslaver A, Bren A, Kikoin I, Alon U, et al. A genome-wide analysis of promoter-mediated phenotypic noise in Escherichia coli. PLoS Genet. 2012;8: 1002443. doi:10.1371/journal.pgen.1002443
9. Jones DL, Brewster RC, Phillips R. Promoter architecture dictates cell-to-cell variability in gene expression. Science. 2014;346: 1533–1536. doi:10.1126/science.1255301
10. Wolf L, Silander OK, van Nimwegen E. Expression noise facilitates the evolution of gene regulation. eLife. 2015;4. doi:10.7554/eLife.05856
11. Urchueguía A, Galbusera L, Chauvin D, Bellement G, Julou T, van Nimwegen E. Genome-wide gene expression noise in Escherichia coli is condition-dependent and determined by propagation of noise through the regulatory network. Balaban N, editor. PLOS Biol. 2021;19: e3001491. doi:10.1371/journal.pbio.3001491
12. Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB. Noise minimization in eukaryotic gene expression. PLoS Biol. 2004;2: e137. doi:10.1371/journal.pbio.0020137
13. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, et al. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature. 2006;441: 840–846. doi:10.1038/nature04785
14. Zhang Z, Qian W, Zhang J. Positive selection for elevated gene expression noise in yeast. Mol Syst Biol. 2009;5: 1–12. doi:10.1038/msb.2009.58
15. Lehner B. Selection to minimise noise in living systems and its implications for the evolution of gene expression. Mol Syst Biol. 2008;4: 170. doi:10.1038/msb.2008.11
16. Lehner B. Conflict between noise and plasticity in yeast. PLoS Genet. 2010;6: e1001185. doi:10.1371/journal.pgen.1001185
17. Bódi Z, Farkas Z, Nevozhay D, Kalapis D, Lázár V, Csörgő B, et al. Phenotypic heterogeneity promotes adaptive evolution. PLoS Biol. 2017;15: e2000644. doi:10.1371/journal.pbio.2000644
18. Duveau F, Hodgins-Davis A, Metzger BPH, Yang B, Tryban S, Walker EA, et al. Fitness effects of altering gene expression noise in saccharomyces cerevisiae. eLife. 2018;7. doi:10.7554/eLife.37272
19. Kittisopikul M, Süel GM. Biological role of noise encoded in a genetic network motif. Proc Natl Acad Sci. 2010;107: 13300–13305. doi:10.1073/pnas.1003975107
20. Chalancon G, Ravarani CNJ, Balaji S, Martinez-Arias A, Aravind L, Jothi R, et al. Interplay between gene expression noise and regulatory network architecture. Trends Genet. 2012;28: 221–232. doi:10.1016/j.tig.2012.01.006
21. Munsky B, Neuert G, van Oudenaarden A. Using Gene Expression Noise to Understand Gene Regulation. Science. 2012;336: 183–187. doi:10.1126/science.1216379
22. Richard M, Yvert G. How does evolution tune biological noise? Front Genet. 2014;5: 374. doi:10.3389/fgene.2014.00374
23. LaBoone PA, Assis R. Stress-Induced Constraint on Expression Noise of Essential Genes in E. coli. J Mol Evol. 2024 [cited 21 Oct 2024]. doi:10.1007/s00239-024-10211-x
24. Dong D, Shao X, Deng N, Zhang Z. Gene expression variations are predictive for stochastic noise. Nucleic

Acids Res. 2011;39: 403–413. doi:10.1093/nar/gkq844
25. Sigalova OM, Shaeiri A, Forneris M, Furlong EE, Zaugg JB. Predictive features of gene expression variation reveal mechanistic link with differential expression. Mol Syst Biol. 2020;16. doi:10.15252/msb.20209539
26. Schoenfelder S, Fraser P. Long-range enhancer–promoter contacts in gene expression control. Nat Rev Genet. 2019;20: 437–455. doi:10.1038/s41576-019-0128-0
27. Faure AJ, Schmiedel JM, Lehner B. Systematic Analysis of the Determinants of Gene Expression Noise in Embryonic Stem Cells. Cell Syst. 2017;5: 471-484.e4. doi:10.1016/j.cels.2017.10.003
28. Barroso GV, Puzovic N, Dutheil JY. The evolution of gene-specific transcriptional noise is driven by selection at the pathway level. Genetics. 2018;208: 173–189. doi:10.1534/genetics.117.300467
29. Foreman R, Wollman R. Mammalian gene expression variability is explained by underlying cell state. Mol Syst Biol. 2020;16. doi:10.15252/msb.20199146
30. Simon CS, Hadjantonakis AK, Schröter C. Making lineage decisions with biological noise: Lessons from the early mouse embryo. Wiley Interdiscip Rev Dev Biol. 2018;7: e319. doi:10.1002/WDEV.319
31. Ohnishi Y, Huber W, Tsumura A, Kang M, Xenopoulos P, Kurimoto K, et al. Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. Nat Cell Biol. 2014;16: 27–37. doi:10.1038/ncb2881
32. Richard A, Boullu L, Herbach U, Bonnafoux A, Morin V, Vallin E, et al. Single-Cell-Based Analysis Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a Differentiation Process. PLOS Biol. 2016;14: e1002585. doi:10.1371/journal.pbio.1002585
33. Liu J, Frochaux M, Gardeux V, Deplancke B, Robinson-Rechavi M. Inter-embryo gene expression variability recapitulates the hourglass pattern of evo-devo. BMC Biol. 2020;18: 1–12. doi:10.1186/s12915-020-00842-z
34. Ho JWK, Stefani M, Dos Remedios CG, Charleston MA. Differential variability analysis of gene expression and its application to human diseases. Bioinformatics. 2008;24: i390–i398. doi:10.1093/bioinformatics/btn142
35. Li J, Liu Y, Kim TH, Min R, Zhang Z. Gene expression variability within and between human populations and implications toward disease susceptibility. PLoS Comput Biol. 2010;6: e1000910. doi:10.1371/journal.pcbi.1000910
36. Mar JC, Matigian NA, Mackay-Sim A, Mellick GD, Sue CM, Silburn PA, et al. Variance of Gene Expression Identifies Altered Network Constraints in Neurological Disease. PLOS Genet. 2011;7: e1002207. doi:10.1371/journal.pgen.1002207
37. Simonovsky E, Schuster R, Yeger-Lotem E. Large-scale analysis of human gene expression variability associates highly variable drug targets with lower drug effectiveness and safety. Bioinformatics. 2019;35: 3028–3037. doi:10.1093/bioinformatics/btz023
38. Bashkeel N, Perkins TJ, Kærn M, Lee JM. Human gene expression variability and its dependence on methylation and aging. BMC Genomics. 2019;20: 941. doi:10.1186/s12864-019-6308-7
39. Alemu EY, Carl JW, Bravo HC, Hannenhalli S. Determinants of expression variability. Nucleic Acids Res. 2014;42: 3503–3514. doi:10.1093/nar/gkt1364
40. Wolf S, Melo D, Garske KM, Pallares LF, Lea AJ, Ayroles JF. Characterizing the landscape of gene expression variance in humans. PLOS Genet. 2023;19: e1010833. doi:10.1371/journal.pgen.1010833
41. Geiler-Samerotte K, Bauer C, Li S, Ziv N, Gresham D, Siegal M. The details in the distributions: why and how to study phenotypic variability. Curr Opin Biotechnol. 2013;24: 752–759. doi:10.1016/j.copbio.2013.03.010
42. De Jong TV, Moshkin YM, Guryev V. Gene expression variability: The other dimension in transcriptome analysis. Physiol Genomics. 2019;51: 145–158. doi:10.1152/physiolgenomics.00128.2018
43. Mar JC. The rise of the distributions: why non-normality is important for understanding the transcriptome and beyond. Biophys Rev. 2019;11: 89–94. doi:10.1007/s12551-018-0494-4
44. Alpern D, Gardeux V, Russeil J, Mangeat B, Meireles-Filho ACA, Breysse R, et al. BRB-seq: Ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. Genome Biol. 2019;20: 1–15. doi:10.1186/s13059-019-1671-x
45. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Maucell E, et al. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature. 2004;431: 946–957. doi:10.1038/nature03025
46. Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, et al. The spotted gar genome

illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat Genet. 2016;48: 427–437. doi:10.1038/ng.3526
47. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, et al. Noise in protein expression scales with natural protein abundance. Nat Genet 2006 386. 2006;38: 636–643. doi:10.1038/ng1807
48. Eling N, Morgan MD, Marioni JC. Challenges in measuring and understanding biological noise. Nat Rev Genet. 2019;20: 536–548. doi:10.1038/s41576-019-0130-6
49. Horta-Lacueva QJ-B, Jónsson ZO, Thorholludottir DAV, Hallgrímsson B, Kapralova KH. Rapid and biased evolution of canalization during adaptive divergence revealed by dominance in gene expression variability during Arctic charr early development. Commun Biol. 2023;6: 1–12. doi:10.1038/s42003-023-05264-5
50. Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. Bioinformatics. 2007;23: 257–258. doi:10.1093/bioinformatics/btl567
51. Reijnders MJMF, Waterhouse RM. Summary Visualizations of Gene Ontology Terms With GO-Figure! Front Bioinforma. 2021;1: 6. doi:10.3389/fbinf.2021.638255
52. Proux E, Studer RA, Moretti S, Robinson-Rechavi M. Selectome: a database of positive selection. Nucleic Acids Res. 2009;37: D404–D407. doi:10.1093/nar/gkn768
53. Moretti S, Laurenczy B, Gharib WH, Castella B, Kuzniar A, Schabauer H, et al. Selectome update: quality control and computational improvements to a database of positive selection. Nucleic Acids Res. 2014;42. doi:10.1093/nar/gkt1065
54. Joshi CJ, Ke W, Drangowska-Way A, O'Rourke EJ, Lewis NE. What are housekeeping genes? PLOS Comput Biol. 2022;18: e1010295. doi:10.1371/journal.pcbi.1010295
55. Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics. 2005;21: 650–659. doi:10.1093/bioinformatics/bti042
56. Kryuchkova-Mostacci N, Robinson-Rechavi M. A benchmark of gene expression tissue-specificity metrics. Brief Bioinform. 2017;18: 205–214. doi:10.1093/bib/bbw008
57. Fair BJ, Blake LE, Sarkar A, Pavlovic BJ, Cuevas C, Gilad Y. Gene expression variability in human and chimpanzee populations share common determinants. eLife. 2020;9: 1–26. doi:10.7554/eLife.59929
58. Einarsson H, Salvatore M, Vaagensø C, Alcaraz N, Bornholdt J, Rennie S, et al. Promoter sequence and architecture determine expression variability and confer robustness to genetic variants. Eyras E, Perry GH, editors. eLife. 2022;11: e80943. doi:10.7554/eLife.80943
59. Osorio D, Yu X, Zhong Y, Li G, Serpedin E, Huang JZ, et al. Single-Cell Expression Variability Implies Cell Function. Cells. 2020;9: 14. doi:10.3390/cells9010014
60. Gu X, Su Z. Tissue-driven hypothesis of genomic evolution and sequence-expression correlations. Proc Natl Acad Sci. 2007;104: 2779–2784. doi:10.1073/pnas.0610797104
61. Drummond DA, Wilke CO. Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. Cell. 2008;134: 341–352. doi:10.1016/j.cell.2008.05.042
62. Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, et al. Parallel Patterns of Evolution in the Genomes and Transcriptomes of Humans and Chimpanzees. Science. 2005;309: 1850–1854. doi:10.1126/science.1108296
63. Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, et al. The evolution of gene expression levels in mammalian organs. Nature. 2011;478: 343–348. doi:10.1038/nature10532
64. Necsulea A, Kaessmann H. Evolutionary dynamics of coding and non-coding transcriptomes. Nat Rev Genet. 2014;15: 734–748. doi:10.1038/nrg3802
65. Kryuchkova-Mostacci N, Robinson-Rechavi M. Tissue-specific evolution of protein coding genes in human and mouse. PLoS ONE. 2015;10. doi:10.1371/journal.pone.0131673
66. Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y, et al. Gene expression across mammalian organ development. Nature. 2019;571: 505–509. doi:10.1038/s41586-019-1338-5
67. Ramsköld D, Wang ET, Burge CB, Sandberg R. An Abundance of Ubiquitously Expressed Genes Revealed by Tissue Transcriptome Sequence Data. PLOS Comput Biol. 2009;5: e1000598. doi:10.1371/journal.pcbi.1000598
68. Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, et al. Cellular Source and Mechanisms of High Transcriptome Complexity in the Mammalian Testis. Cell Rep. 2013;3: 2179–2190. doi:10.1016/j.celrep.2013.05.031

69. Darbellay F, Necsulea A. Comparative Transcriptomics Analyses across Species, Organs, and Developmental Stages Reveal Functionally Constrained lncRNAs. Mol Biol Evol. 2020;37: 240–259. doi:10.1093/molbev/msz212
70. Musilova Z, Salzburger W, Cortesi F. The Visual Opsin Gene Repertoires of Teleost Fishes: Evolution, Ecology, and Function. Annu Rev Cell Dev Biol. 2021;37: 441–468. doi:10.1146/annurev-cellbio-120219-024915
71. Cortijo S, Aydin Z, Ahnert S, Locke JC. Widespread inter-individual gene expression variability in Arabidopsis thaliana. Mol Syst Biol. 2019;15: e8591. doi:10.15252/msb.20188591
72. Hasegawa Y, Taylor D, Ovchinnikov DA, Wolvetang EJ, de Torrenté L, Mar JC. Variability of Gene Expression Identifies Transcriptional Regulators of Early Human Embryonic Development. PLoS Genet. 2015;11: e1005428. doi:10.1371/journal.pgen.1005428
73. Schor IE, Degner JF, Harnett D, Cannavò E, Casale FP, Shim H, et al. Promoter shape varies across populations and affects promoter evolution and expression noise. Nat Genet. 2017;49: 550–558. doi:10.1038/ng.3791
74. Zabidi MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, et al. Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. Nature. 2015;518: 556–559. doi:10.1038/nature13994
75. Furlong EEM, Levine M. Developmental enhancers and chromosome topology. Science. 2018;361: 1341–1345. doi:10.1126/science.aau0320
76. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell. 2016;167: 1369-1384.e19. doi:10.1016/j.cell.2016.09.037
77. Schoenfelder S, Furlan-Magaril M, Mifsud B, Tavares-Cadete F, Sugar R, Javierre B-M, et al. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res. 2015;25: 582–597. doi:10.1101/gr.185272.114
78. Signor SA, Nuzhdin SV. The Evolution of Gene Expression in cis and trans. Trends Genet. 2018;34: 532–544. doi:10.1016/j.tig.2018.03.007
79. Long HK, Prescott SL, Wysocka J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell. 2016;167: 1170–1187. doi:10.1016/j.cell.2016.09.018
80. Jindal GA, Farley EK. Enhancer grammar in development, evolution, and disease: dependencies and interplay. Dev Cell. 2021;56: 575–587. doi:10.1016/j.devcel.2021.02.016
81. Westerfield M. The zebrafish book; a guide for the laboratory use of zebrafish (Brachydanio rerio). Eugene: University of Oregon Press; 1993.
82. Best J, Adatto I, Cockington J, James A, Lawrence C. A Novel Method for Rearing First-Feeding Larval Zebrafish: Polyculture with Type L Saltwater Rotifers (Brachionus plicatilis). Zebrafish. 2010;7: 289–295. doi:10.1089/zeb.2010.0667
83. Kaminow B, Yunusov D, Dobin A. STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. bioRxiv. 2021; 2021.05.05.442755. doi:10.1101/2021.05.05.442755
84. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinforma Oxf Engl. 2013;29: 15–21. doi:10.1093/bioinformatics/bts635
85. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. Nucleic Acids Res. 2020;48: D682–D688. doi:10.1093/nar/gkz966
86. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. p. https://www.R-project.org. Available: http://www.r-project.org
87. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods. 2015;12: 115–121. doi:10.1038/nmeth.3252
88. RStudio Team. RStudio: Integrated Development Environment for R. Boston, MA: RStudio, PBC; 2021. p. http://www.rstudio.com/. Available: http://www.rstudio.com/
89. Fukushima K, Pollock DD. Amalgamated cross-species transcriptomes reveal organ-specific propensity in gene expression evolution. Nat Commun. 2020;11: 1–14. doi:10.1038/s41467-020-18090-8
90. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Ridwan Amode M, et al. Ensembl 2021. Nucleic Acids Res. 2021;49: D884–D891. doi:10.1093/nar/gkaa942
91. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression

926 analysis of digital gene expression data. Bioinformatics. 2009;26: 139–140.
927 doi:10.1093/bioinformatics/btp616
928 92. Carlson M. org.Dr.eg.db: Genome wide annotation for Zebrafish. R package version 3.13.0. 2021.
929 93. Wang J, Wen S, Symmans WF, Pusztai L, Coombes KR. The Bimodality Index: A Criterion for Discovering
930 and Ranking Bimodal Signatures from Cancer Gene Expression Profiling Data. Cancer Inform. 2009;7:
931 199–216.
932