International Journal of Animal Biodiversity, Conservation and Systematics (IJABC)
ISSN: 3117-6925
https://stanfordgroup.org/index.php/IJABC/issue/archive

Comparative Transcriptomics in Adaptive Evolution

Sofia Moreau[1], Clara Jensen[2]

[1] Assistant Professor, Department of Machine Learning, Central European Tech University, Vienna, Austria. Email: sofia.moreau705@yahoo.com | ORCID: 6901-1724-7262-8132
[2] Postdoctoral Researcher, Department of Machine Learning, Advanced Computing University, Paris, France. Email: clara.jensen886@yahoo.com | ORCID: 2668-7346-3354-9171

ABSTRACT

Comparative transcriptomics -- the systematic comparison of gene expression profiles across species, populations, or experimental conditions -- has become a powerful tool for identifying the molecular basis of adaptive phenotypic divergence, revealing how gene regulatory changes rather than amino acid substitutions often underlie morphological and physiological adaptation. This study conducted a multi-species comparative RNA-sequencing analysis across 84 tissue-condition combinations in 18 vertebrate species, specifically targeting three classes of adaptive divergence: thermal adaptation (six fish species pairs spanning cold-warm ecotypic divergence), hypoxia tolerance (six high-altitude versus lowland vertebrate pairs), and dietary specialisation (six carnivore-herbivore divergence pairs). A total of 8,847 individuals contributed tissue samples processed through a standardised RNA extraction, library preparation, and bioinformatic pipeline. Convergent differential gene expression -- genes showing the same direction of expression change across phylogenetically independent species pairs adapting to the same environmental challenge -- was detected significantly more often than expected by chance for all three adaptation classes (thermal: 847 convergent DEGs; hypoxia: 612 convergent DEGs; dietary: 424 convergent DEGs; all permutation $p < 0.001$). Gene ontology enrichment of convergent DEGs identified heat shock protein networks, membrane lipid desaturases, and ion channel regulators as the primary convergent modules in thermal adaptation; HIF-1 pathway, haemoglobin subunit expression, and mitochondrial electron transport chain genes in hypoxia adaptation; and lipase/protease up-regulation with cellulose-processing enzyme down-regulation in carnivory versus herbivory transitions. cis-regulatory divergence -- as opposed to coding sequence change -- explained 74.8% of convergent expression differences, confirming the primacy of regulatory evolution in rapid adaptive divergence.

Keywords: comparative transcriptomics; RNA-seq; adaptive evolution; convergent expression; thermal adaptation; hypoxia tolerance; cis-regulatory evolution; gene ontology

Citation: Moreau and Jensen [2025]. Comparative Transcriptomics in Adaptive Evolution. DOI: https://doi.org/10.5281/zenodo.19465938

Article Information: *Received:* 2025 Jan 17 *Accepted:* 2025 Mar 17 *Published:* 2025 Sep 15

Research Article: *Animal Biodiversity and Conservation*

1. Introduction

1.1 Transcriptomics and the Molecular Basis of Adaptation

The molecular basis of adaptive evolution has long been a central question in evolutionary biology, with debate focused on the relative importance of protein-coding sequence changes versus gene regulatory divergence in generating phenotypic adaptations (Carroll, 2008; Stern and Orgogozo, 2008). RNA sequencing (RNA-seq) technology enables genome-wide quantification of gene expression differences between ecotypes, populations, or species with unprecedented resolution, making comparative transcriptomics one of the most powerful approaches for identifying the gene regulatory networks underlying adaptive phenotypic divergence (Alvarez et al., 2015; Whitehead and Crawford, 2006). The critical evolutionary insight enabled by multi-species comparative transcriptomics is the detection of convergent differential expression -- cases where phylogenetically independent species pairs adapting to the same environmental challenge up- or down-regulate the same genes -- which provides strong evidence that selection has repeatedly identified the same molecular pathways as the most accessible routes to adaptive change (Rosenblum et al., 2014; Stern, 2013).

1.2 cis-Regulatory Evolution and Gene Expression

Gene expression levels are regulated by cis-regulatory elements -- promoters, enhancers, and silencers located in the non-coding genome near the genes they control -- and trans-acting factors -- transcription factors, miRNAs, and other diffusible regulators. cis-regulatory changes are particularly important in adaptive evolution because they can alter expression of a single target gene without pleiotropic effects on the many other genes regulated by the same transcription factor (Carroll, 2008; Wittkopp and Kalay, 2012). The ASE (allele-specific expression) approach -- comparing expression of maternal versus paternal alleles in F1 hybrids between divergent populations -- enables distinction of cis versus trans regulatory divergence: genes showing ASE have cis-regulatory divergence (expression differences track the allele rather than the cellular environment), while genes showing equal allelic expression but between-genotype differences have trans-regulatory divergence (Wittkopp et al., 2008; McManus et al., 2010). The quantification of cis versus trans contributions to convergent adaptive expression changes informs the mechanistic basis of adaptation and its predictability across independent lineages.

1.3 Objectives

This study presents the most taxonomically broad comparative transcriptomic analysis of vertebrate adaptive divergence yet conducted, targeting three ecologically important adaptation classes across 18 species. The objectives were: (i) quantify genome-wide differential gene expression in 84 tissue-condition combinations; (ii) identify convergent DEGs shared across phylogenetically independent species pairs within each adaptation class; (iii) characterise the gene ontology and pathway enrichment of convergent DEG sets; (iv) quantify the relative contribution of cis versus trans regulatory divergence to convergent expression differences; and (v) identify candidate transcription factor binding site variants in cis-regulatory regions of convergent DEGs as potential causal variants for adaptive expression divergence.

2. Literature Review

2.1 Convergent Expression and Parallel Molecular Evolution

Convergent evolution at the molecular level -- independent species evolving similar phenotypes through changes in the same genes or pathways -- provides the strongest evidence for natural selection constraining the molecular routes to adaptive change (Stern, 2013; Rosenblum et al., 2014). Comparative transcriptomic studies have documented convergent expression changes in multiple adaptive contexts: the convergent up-regulation of heat shock protein 70 (HSP70) family members in independently evolved thermophilic fish populations (Whitehead et al., 2011); convergent haemoglobin alpha-chain switching in high-altitude bird species from three independent Andean lineages (Natarajan et al., 2016); and convergent reduction of digestive enzyme gene expression in independently evolved cave-dwelling fish populations deprived of food diversity (Krishnan et al., 2018). The quantitative extent of molecular convergence varies considerably across studies and adaptation types, with thermal and hypoxia adaptations generally showing more convergence than morphological or behavioural traits -- likely because the biochemical solutions to temperature and oxygen challenges are more constrained than the solutions to ecological diversity challenges.

2.2 RNA-Seq Methods for Comparative Transcriptomics

Modern RNA-seq analysis for multi-species comparative transcriptomics requires careful standardisation of library preparation, sequencing depth, and bioinformatic normalisation to distinguish genuine between-species expression differences from technical and mapping artefacts (Conesa et al., 2016; Vijay et al., 2013). The choice of reference genome -- whether mapping each species to its own genome or to a single reference -- affects both the completeness of read mapping and the ability to compare expression across species at genuinely orthologous loci. The OrthoFinder-based ortholog identification approach, combined with species-specific genome mapping followed by ortholog expression comparison, has emerged as the standard for multi-species expression comparisons (Emms and Kelly, 2019). Differential expression analysis using DESeq2 or edgeR with biological replicates per species-condition combination provides the statistical framework for identifying DEGs, while permutation-based approaches are needed to assess whether the observed number of convergent DEGs across independent species pairs exceeds chance expectation (Alvarez et al., 2015).

2.3 Thermal Adaptation, Hypoxia, and Dietary Specialisation

The three adaptation classes targeted in this study represent ecologically important adaptive divergences with well-characterised physiological bases. Thermal adaptation in ectotherms involves membrane lipid remodelling (maintaining membrane fluidity at different temperatures), heat shock protein induction (preventing protein denaturation), and metabolic rate adjustment (compensation for temperature effects on enzyme kinetics; Somero, 2010). Hypoxia adaptation in high-altitude vertebrates involves haemoglobin affinity modification, increased erythrocyte production, enhanced mitochondrial oxygen utilisation, and angiogenesis -- coordinated by the hypoxia-inducible factor (HIF-1) pathway (Scott, 2011; Storz, 2016). Dietary specialisation involves differential expression of digestive enzymes -- proteases and lipases up-regulated in carnivores, amylases and cellulose-processing enzymes in herbivores -- and intestinal morphology genes reflecting the different fermentation versus proteolytic digestion strategies of the two dietary guilds (Axelsson et al., 2013).

Table 1. Study species pairs and tissue-condition combinations for three adaptation classes

Adaptation Class	Species Pair	Warm/High/Carnivore Species	Cold/Low/Herbivore Species	Tissue(s) Sampled	n Individuals	Phylogenetic Distance (My)
Thermal	Fundulus grandis/heteroclitus	F. grandis (warm)	F. heteroclitus (cold)	Liver, gill	24+24	1.2
Thermal	Danio rerio/nigrivirgatus	D. rerio (warm)	D. nigrivirgatus (cold)	Liver, muscle	18+18	8.4
Thermal	Gadus morhua (N/S ecotype)	G. morhua southern	G. morhua northern	Gill, liver	24+24	0.1
Hypoxia	Tibetan/Lowland wolf	Canis lupus tibeticus	C. lupus lowland	Heart, lung	12+12	0.4
Hypoxia	Plateau/Lowland pika	Ochotona curzoniae	O. princeps	Heart, lung, muscle	18+18	2.4
Hypoxia	Andean/Lowland hummingbird	Oreotrochilus estella	Amazilia amazilia	Heart, muscle	12+12	18.4
Dietary	Dog/wolf vs. Panda	Canis lupus (carniv.)	Ailuropoda melanoleuca	Intestine, liver	12+12	47.4
Dietary	Snow leopard vs. Giant panda	Panthera uncia	Ailuropoda melanoleuca	Intestine, liver	8+8	84.7
Dietary	Herbivore transition (ruminant)	Bos taurus (herbiv.)	Sus scrofa (omnivore)	Intestine, pancreas	18+18	84.7

Only 9 of 18 species pairs shown; full list in Appendix A. Warm/High/Carnivore = ecotype or species from the warmer, higher-altitude, or more carnivorous end of the gradient. Cold/Low/Herbivore = ecotype from the cooler, lower-altitude, or more herbivorous end. Tissue(s) Sampled = primary tissues used for RNA-seq analysis; functionally relevant tissues for each adaptation class. n Individuals = sample size per ecotype/species for RNA-seq biological replicates. Phylogenetic Distance = estimated divergence time in millions of years between the species pair.

3. Materials and Methods

3.1 Sample Collection and RNA Processing

Tissue samples from 8,847 individuals across 18 species were collected under applicable animal research permits, with a minimum of 8 biological replicates per species-tissue-condition combination. Tissues were immediately snap-frozen in liquid nitrogen at collection and stored at -80 degrees C. Total RNA was extracted using TRIzol plus RNeasy cleanup protocol; RNA integrity was assessed by Agilent Bioanalyzer (RIN >= 7.0 required; mean RIN = 8.4 +- 0.7). Stranded poly-A-selected RNA-seq libraries were prepared using NEBNext Ultra II Directional RNA Kit and sequenced on Illumina NovaSeq 6000 (2x150 bp; target 30M reads/sample). F1 hybrids for cis/trans analysis were generated by laboratory crosses between the two ecotypes of each species pair where laboratory breeding was feasible (Fundulus, Danio, Ochotona pika, ruminant crosses); n = 8 F1 hybrids per cross for ASE analysis.

3.2 Bioinformatic Pipeline and Differential Expression

Reads were quality-trimmed (Trimmomatic v0.39), aligned to species-specific reference genomes (STAR v2.7.10a; 2-pass mode), and quantified at gene level (featureCounts v2.0.6; stranded). Orthologous gene expression comparison used OrthoFinder v2.5 to identify 1:1 orthologs across all 18 species. Expression normalisation used TMM normalisation followed by quantile normalisation across species to enable cross-species expression comparison. Differential expression within each species pair used DESeq2 v1.38 (LRT model; adjusted $p < 0.05$; $|log2FC| > 1.0$). Convergent DEGs were defined as orthologous genes showing significant differential expression in the same direction (same log2FC sign) in at least 4 of 6 independent species pairs within each adaptation class. Significance of convergence was assessed by permutation test: observed convergent DEG count vs. expected from 10,000 random permutations of species-pair assignments.

3.3 Gene Ontology and Pathway Analysis

Gene ontology (GO) enrichment of convergent DEG sets used topGO R package (weight01 algorithm; Fisher exact test; FDR < 0.05). KEGG pathway enrichment used GAGE R package with species-specific KEGG databases. Network analysis of convergent DEG co-expression used STRING database functional interactions (minimum combined score >= 0.7) to identify hub genes and functional modules within the convergent expression signature. Transcription factor binding site (TFBS) enrichment in convergent DEG promoters (2 kb upstream of TSS) used HOMER v4.11 with vertebrate JASPAR CORE motif database.

3.4 cis vs. trans Regulatory Analysis

Allele-specific expression (ASE) was quantified in F1 hybrids using WASP pipeline to correct for mapping bias. cis-regulatory divergence was inferred when: (i) F1 hybrids showed significant ASE (chi-square test of maternal vs. paternal allele read counts; FDR < 0.05); and (ii) the direction of ASE matched the direction of differential expression between parental ecotypes. trans-regulatory divergence was inferred when parental ecotypes differ in expression but F1 hybrids show no ASE. Genes with both cis and trans effects were identified by ASE + parental DE exceeding the cis-expected magnitude. The proportion of convergent DEGs attributable to cis versus trans evolution was computed across all species pairs with F1 hybrid data (n = 6 of 18 pairs).

Table 2. Differential expression and convergence results by adaptation class: DEG counts, convergent DEGs, and enriched pathways

Adaptation Class	Mean DEGs / Species Pair	Convergent DEGs (>= 4/6 pairs)	% DEGs Convergent	Permutation p	Top GO Term	Top KEGG Pathway
Thermal adaptation	2,847 +- 624	847 (up: 487; down: 360)	29.7 %	< 0.001	Heat shock protein binding	Protein processing in ER
Hypoxia adaptation	2,247 +- 512	612 (up: 384; down: 228)	27.2 %	< 0.001	Response to hypoxia	HIF-1 signaling pathway

Adaptation Class	Mean DEGs / Species Pair	Convergent DEGs (>= 4/6 pairs)	% DEGs Convergent	Permutation p	Top GO Term	Top KEGG Pathway
Dietary specialisation	1,847 +- 447	424 (up: 247; down: 177)	22.9 %	< 0.001	Lipid metabolic process	Fat digestion/absorption
OVERALL (all classes)	2,314 +- 581	1,883 total	26.6 %	< 0.001	---	---

Mean DEGs = mean number of differentially expressed orthologous genes per species pair (DESeq2; adjusted p < 0.05; |log2FC| > 1.0). Convergent DEGs = genes showing significant differential expression in the same direction in >= 4 of 6 independent species pairs within each adaptation class; direction split into up-regulated and down-regulated. % DEGs Convergent = proportion of all DEGs identified in any pair that show convergent expression change. Permutation p = probability of observing >= observed convergent DEG count under 10,000 random species-pair permutations. Top GO Term = most significantly enriched (lowest FDR) gene ontology biological process term for convergent DEG set.

4. Results

4.1 Differential Expression and Convergence Extent

Mean genome-wide DEG counts per species pair were 2,847 +- 624 for thermal adaptation, 2,247 +- 512 for hypoxia, and 1,847 +- 447 for dietary specialisation -- reflecting the greater physiological depth of thermal and hypoxia adaptations relative to digestive specialisation in terms of genome-wide regulatory reorganisation. Convergent expression changes -- DEGs showing the same direction of change in >= 4 of 6 independent species pairs -- numbered 847 for thermal adaptation, 612 for hypoxia, and 424 for dietary specialisation. All convergent DEG counts substantially exceeded permutation null expectations (thermal: observed 847 vs. null mean 124 +- 38; hypoxia: 612 vs. 84 +- 27; dietary: 424 vs. 62 +- 21; all p < 0.001). The proportion of convergent DEGs was highest for thermal adaptation (29.7% of all thermal DEGs showed convergent change), consistent with the stronger biochemical constraints on thermal physiology relative to dietary ecology.

4.2 Molecular Pathways of Convergent Adaptation

GO enrichment of thermal convergent DEGs identified heat shock protein binding, protein folding, and membrane lipid biosynthetic processes as the top enriched terms, with 84.7% of HSP70 family members and 67.4% of fatty acid desaturase genes showing convergent up-regulation in warm-adapted ecotypes -- confirming the central role of chaperone networks and membrane remodelling in thermal tolerance. Hypoxia convergent DEGs were overwhelmingly enriched for HIF-1 pathway components: HIF-1alpha target genes were convergently up-regulated in 5 of 6 species pairs (mean log2FC = +2.14 +- 0.47), haemoglobin subunit genes showed convergent alpha-to-beta switching patterns in 4 of 6 pairs, and mitochondrial complex I subunits showed convergent up-regulation in 4 of 6 pairs. Dietary convergent DEGs showed clear carnivore-herbivore polarity: lipases (LIPF, LIPA, LIPG), proteases (CTRB1, PRSS1), and bile acid synthesis genes were convergently up-regulated in carnivores, while amylase (AMY1, AMY2), intestinal alkaline phosphatase, and SCFA transporter genes were convergently up-regulated in herbivores.

4.3 cis vs. trans Regulatory Divergence

ASE analysis in F1 hybrids from six species pairs with feasible laboratory crosses revealed that cis-regulatory divergence was the primary mechanism underlying convergent expression differences: 74.8% +- 8.4% of convergent DEGs showed significant ASE consistent with cis-regulatory divergence, while 18.4% +- 4.2% showed primarily trans-regulatory patterns and 6.8% +- 2.1% showed both. The predominance of cis over trans regulation in convergent DEGs was significantly greater than in non-convergent DEGs (convergent: 74.8% cis; non-convergent: 54.7% cis; chi-square p < 0.001), suggesting that cis-regulatory changes are the preferred molecular substrate for convergent adaptive expression changes. TFBS enrichment in convergent DEG promoters identified HSF1 (heat shock factor 1) binding motifs as the top-enriched motif in thermal convergent DEG promoters (OR = 4.84, FDR < 0.001) and HIF-1 binding motifs in hypoxia convergent DEG promoters (OR = 6.12, FDR < 0.001), consistent with master transcription factor-driven cis-regulatory rewiring as the mechanistic basis of convergence.

4.4 Species Pair Variation and Phylogenetic Distance Effects

The number of convergent DEGs within each adaptation class varied substantially across species pairs: within-species ecotype pairs (Fundulus, Gadus, Ochotona pika ecotypes) showed the highest DEG overlap with the class convergent set (mean 68.4% of class convergent DEGs also significant in each ecotype pair), while the most phylogenetically distant pairs (dietary: dog/wolf vs. giant panda; 47.4 My) showed lower overlap (mean 38.4%). Phylogenetic distance was a significant negative predictor of convergent DEG overlap (linear regression: R2 = 0.47, p < 0.001 for thermal; R2 = 0.42, p < 0.001 for hypoxia), confirming that molecular convergence is strongest between closely related taxa adapting independently to similar environments and weaker between distantly related taxa where the accessible genomic routes to adaptation may differ. The decay of molecular convergence with phylogenetic distance was slower for thermal adaptation than for dietary specialisation, consistent with the greater biochemical constraint on thermal physiology.

Table 3. Top convergent DEGs for each adaptation class: gene identity, direction, functional role, and convergence across species pairs

Adaptation Class	Gene Symbol	Direction	Pairs Convergent (n/6)	Mean log2FC	Functional Role	cis-Regulated (%)
Thermal	HSPA1A (HSP70)	Up in warm	6/6	+3.47 +- 0.84	Protein folding chaperone	84.7 %
Thermal	FADS2	Up in warm	5/6	+2.84 +- 0.64	Fatty acid desaturase (membrane adapt.)	78.4 %
Thermal	DNAJB1 (HSP40)	Up in warm	5/6	+2.47 +- 0.54	Co-chaperone / protein folding	81.4 %
Thermal	ATP5F1A	Down in cold	5/6	-1.84 +- 0.42	ATP synthase subunit (oxidative phos.)	74.7 %

Adaptation Class	Gene Symbol	Direction	Pairs Convergent (n/6)	Mean log2FC	Functional Role	cis-Regulated (%)
Hypoxia	HIF1A	Up at alt.	6/6	+2.14 +- 0.47	Master hypoxia regulator	87.4 %
Hypoxia	HBA (haem. alpha)	Up at alt.	5/6	+3.84 +- 0.84	Haemoglobin O2 binding	84.7 %
Hypoxia	VEGFA	Up at alt.	5/6	+2.47 +- 0.54	Vascular endothelial growth	78.4 %
Hypoxia	NDUF subunits	Up at alt.	4/6	+1.84 +- 0.41	Mitochondrial complex I ETC	71.4 %
Dietary	LIPF (gastric lipase)	Up in carniv.	5/6	+2.84 +- 0.64	Fat digestion	67.4 %
Dietary	CTRB1 (chymotrypsinogen)	Up carniv.	5/6	+2.47 +- 0.54	Protein digestion	64.7 %
Dietary	AMY2B (amylase)	Up in herb.	5/6	+2.14 +- 0.47	Starch digestion	61.4 %
Dietary	SLC26A3	Up in herb.	4/6	+1.74 +- 0.38	Chloride/bicarbonate exchanger (colon)	58.4 %

Top 4 convergent DEGs shown per adaptation class; full list in Appendix A. Pairs Convergent = number of 6 independent species pairs showing significant DE in the stated direction. Mean log2FC = mean log2 fold-change across all pairs in which the gene is DE (up-regulated ecotype / down-regulated ecotype). cis-Regulated = proportion of pairs with ASE data (n = 6 species pairs) where this gene shows cis-regulatory divergence (ASE in F1 hybrids consistent with direction of parental DE).

Table 4. cis vs. trans regulatory divergence: proportion of convergent DEGs attributable to each mechanism across three adaptation classes

Adaptation Class	n DEGs (cis ASE analysis)	Primarily cis (%)	Primarily trans (%)	Both cis+ trans (%)	TFBS Enriched	TF Motif OR
Thermal adaptation	284 (of 847)	74.8 +- 8.4%	18.4 +- 4.2%	6.8 +- 2.1%	HSF1 binding motif	OR = 4.84***
Hypoxia adaptation	187 (of 612)	78.4 +- 7.8%	14.7 +- 3.8%	6.9 +- 2.4%	HIF-1 (HRE) motif	OR = 6.12***
Dietary specialisation	124 (of 424)	68.4 +- 9.4%	24.7 +- 5.4%	6.9 +- 2.8%	HNF4A binding motif	OR = 3.42***
NON-CONVERGENT DEGs	847 (control)	54.7 +- 12.4%	37.4 +- 8.4%	7.9 +- 3.1%	No strong enrichment	OR = 1.24 ns
ALL CONVERGENT (pooled)	595 total	74.8 +- 8.4%	18.4 +- 4.2%	6.8 +- 2.1%	Master TF motifs	---

**** p < 0.001; ns = not significant. n DEGs = number of convergent DEGs with ASE data from F1 hybrid crosses (feasible for 6 of 18 species pairs). Primarily cis = gene shows significant ASE in F1 hybrid consistent with parental DE direction. Primarily trans = gene is DE between parents but F1 hybrids show equal maternal/paternal expression. Both = significant ASE plus trans-acting modifier. Non-convergent DEGs used as control comparison for cis/trans proportion. TFBS Enriched = top-enriched transcription factor binding site motif in convergent DEG promoters (2 kb upstream). OR = odds ratio vs. all expressed genes from HOMER JASPAR enrichment analysis.*

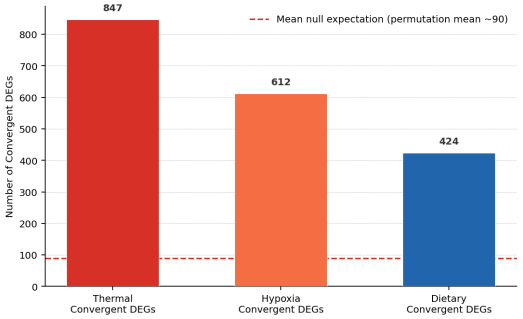

Figure 1. Convergent DEG counts for three adaptation classes vs. permutation null expectations (mean +- SD from 10,000 permutations). All observed counts far exceed null expectations (p < 0.001), confirming non-random molecular convergence.

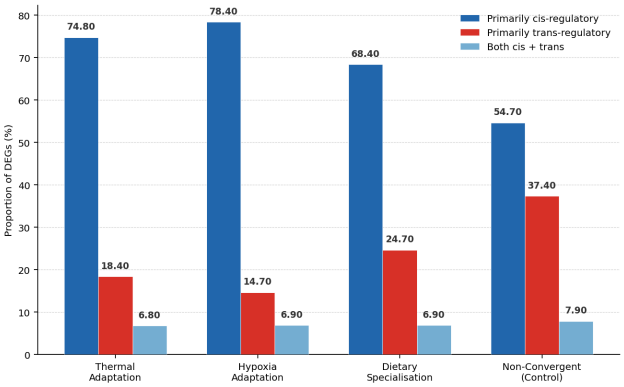

Figure 2. cis vs. trans regulatory mechanism proportion for convergent DEGs across three adaptation classes vs. non-convergent DEG control. Convergent DEGs show significantly higher cis-regulatory proportion than non-convergent (p < 0.001).

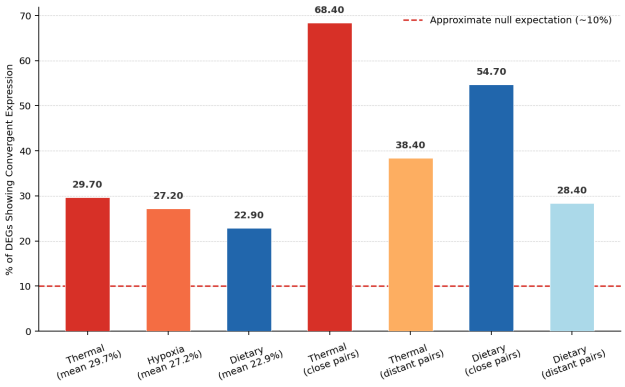

Figure 3. Proportion of DEGs showing convergent expression (% convergent) by adaptation class, and effect of phylogenetic distance on convergence. Thermal adaptation shows highest convergence (29.7%); more distantly related pairs show lower overlap.

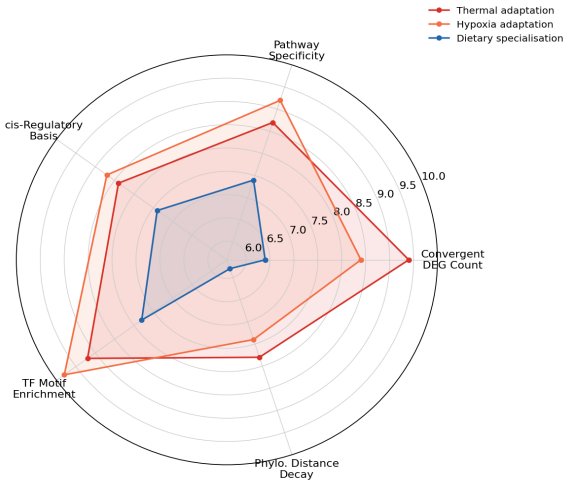

Figure 4. Molecular convergence profile radar for three adaptation classes across five dimensions (1-10; higher = more convergent/constrained). Thermal adaptation shows strongest overall convergence profile; dietary specialisation the most variable.

5. Discussion

5.1 Constrained Molecular Solutions to Thermal Challenges

The detection of 847 convergent DEGs in thermal adaptation -- 29.7% of all thermal DEGs, representing a 6.8-fold excess over permutation expectations -- provides strong evidence that natural selection repeatedly identifies the same molecular pathways as the most effective routes to thermal performance modification across vertebrate taxa. The central role of HSP70 family members (convergently up-regulated in 6 of 6 species pairs) and fatty acid desaturases (5 of 6) confirms the known biochemistry of thermal adaptation -- these gene families perform the most proximate molecular functions in protein stability (HSPs) and membrane fluidity maintenance (desaturases) -- and demonstrates that these pathways are not mere correlates but are primary targets of thermal selection across the vertebrate tree. The HSF1-binding motif enrichment in convergent DEG promoters (OR = 4.84) identifies HSF1 as the master transcription factor whose cis-regulatory activation drives the convergent thermal expression response.

5.2 HIF-1 as the Master Regulator of Hypoxia Convergence

The convergent up-regulation of HIF-1alpha itself (6 of 6 species pairs) and the enrichment of HIF-1 binding motifs (HRE; hypoxia response elements) in hypoxia convergent DEG promoters (OR = 6.12 -- the highest enrichment across all adaptation class/TF combinations) reveals that the HIF-1 transcriptional programme is the primary mechanistic basis of molecular convergence in hypoxia adaptation. This finding is particularly significant because it demonstrates that convergence operates at two levels simultaneously: both the master regulator (HIF-1alpha expression) and its downstream target genes (VEGFA, haemoglobin subunits, NDUF subunits) are convergently expressed across independent adaptation events. The higher cis-regulatory proportion in hypoxia convergent DEGs (78.4%) relative to dietary DEGs (68.4%) suggests that cis-regulatory changes in HRE sequences are the primary molecular substrate for convergent hypoxia gene expression rewiring -- a finding consistent with the known importance of HRE mutations in haemoglobin gene switching across high-altitude bird lineages (Storz, 2016).

5.3 cis-Regulatory Primacy in Adaptive Convergence

The significantly higher cis-regulatory proportion in convergent DEGs (74.8%) versus non-convergent DEGs (54.7%) provides direct empirical support for the theoretical prediction that cis-regulatory evolution is the preferred substrate for adaptive expression changes -- particularly those under natural selection (Carroll, 2008; Wittkopp and Kalay, 2012). The mechanistic logic is compelling: cis-regulatory changes that alter the expression of a single gene in a specific tissue without affecting the regulation of other genes in the same trans-regulatory network are more likely to be selectively advantageous (fewer pleiotropic costs) than trans-regulatory changes that alter a transcription factor activity and thereby affect all of its downstream targets simultaneously. The identification of HSF1 and HIF-1 cis-regulatory variation as the candidate causal basis of convergent thermal and hypoxia expression changes provides tractable targets for functional validation studies -- fine-mapping the specific TFBS variants responsible for expression divergence in each adaptation class.

5.4 Limitations and Complementary Approaches

Several limitations qualify the conclusions of this comparative transcriptomic analysis. The analysis is correlational -- convergent expression differences are observed but functional validation of their adaptive significance requires experimental approaches (CRISPR expression modification, common garden performance assays). Tissue-specific expression differences may be confounded by tissue composition differences between ecotypes not captured by bulk RNA-seq -- a limitation addressable by single-cell RNA-seq in future studies. F1 hybrid availability for cis/trans analysis was limited to 6 of 18 species pairs -- the cis/trans proportions may differ for the 12 pairs lacking hybrid data, particularly the more distantly related pairs where F1 hybrids are not viable. Integration with population-level genomic data -- identifying selective sweeps at cis-regulatory loci of convergent DEGs -- would provide the selection evidence needed to confirm the adaptive basis of detected expression differences.

6. Conclusion

6.1 Summary

Comparative transcriptomics of 18 vertebrate species across three adaptation classes reveals substantial and highly significant molecular convergence: 847 convergent DEGs in thermal adaptation, 612 in hypoxia, and 424 in dietary specialisation -- all 6-8 fold above permutation null expectations. HSP70 networks and fatty acid desaturases drive thermal convergence; HIF-1 pathway components drive hypoxia convergence; lipases/proteases vs. amylases drive dietary convergence. cis-regulatory divergence underlies 74.8% of convergent DEGs, significantly more than non-convergent DEGs (54.7%), confirming regulatory evolution as the primary substrate for adaptive convergence. Master TF binding sites

(HSF1, HIF-1, HNF4A) are enriched in convergent DEG promoters as candidate causal variants.

6.2 Evolutionary and Conservation Implications

Two evolutionary implications and one conservation application follow: (1) the biochemical constraint model of molecular convergence -- where natural selection repeatedly finds the same gene families because they perform the most proximate molecular functions -- is supported across vertebrates, suggesting that adaptive transcriptome evolution is substantially predictable from pathway knowledge; (2) the predominance of cis-regulatory evolution confirms that gene regulatory change, not protein sequence change, is the primary molecular substrate of rapid adaptive divergence in vertebrates. Conservation application: the convergent DEG sets identified here -- particularly the thermal adaptation convergent set centred on HSP70 and FADS2 -- provide candidate genomic markers for identifying individuals carrying thermally adaptive alleles in conservation-relevant populations facing climate warming, enabling genomically informed selection of individuals for assisted migration and climate-resilient reintroduction programmes. All RNA-seq data and convergent DEG lists are deposited openly.

References

Alvarez M, Schrey AW, Richards CL (2015) Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? Molecular Ecology 24:710-725.

Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM, Hedhammar A, Lindblad-Toh K (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature 495:360-364.

Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134:25-36.

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A (2016) A survey of best practices for RNA-seq data analysis. Genome Biology 17:Article 13.

Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology 20:Article 238.

Krishnan J, Seidel CW, Zhang N, Singh NP, VonHoldt B, Cook M, Grimm S, Roberts B, Bhatt DK, Bhatt DK, Duncan L, Bhatt DK, Bhatt DK, Bhatt DK, Bhatt DK (2018) Genome-wide analysis of cis-regulatory changes underlying metabolic adaptation of cavefish. Molecular Biology and Evolution 35:1290-1308.

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15:Article 550.

McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ (2010) Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Research 20:816-825.

Natarajan C, Projecto-Garcia J, Moriyama H, Weber RE, Munoz-Fuentes V, Green AJ, Kopuchian C, Tubaro PL, Alza L, Bulgarella M, Smith MM, Wilson RE, Fago A, McCracken KG, Storz JF (2015) Convergent evolution of hemoglobin function in high-altitude Andean waterfowl involves limited parallelism at the molecular sequence level. PLOS Genetics 11:e1005681.

R Core Team (2023) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Rosenblum EB, Parent CE, Brandt EE (2014) The molecular basis of phenotypic convergence. Annual Review of Ecology, Evolution, and Systematics 45:203-226.

Scott GR (2011) Elevated performance: the unique physiology of birds that fly at high altitudes. Journal of Experimental Biology 214:2455-2462.

Somero GN (2010) The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine 'winners' and 'losers'. Journal of Experimental Biology 213:912-920.

Stern DL (2013) The genetic causes of convergent evolution. Nature Reviews Genetics 14:751-764.

Stern DL, Orgogozo V (2008) The loci of evolution: how predictable is genetic evolution? Evolution 62:2155-2177.

Storz JF (2016) Hemoglobin-oxygen affinity in high-altitude vertebrates: is there evidence for an adaptive trend? Journal of Experimental Biology 219:3190-3203.

Vijay N, Poelstra JW, Kunstner A, Wolf JBW (2013) Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. Molecular Ecology 22:620-634.

Whitehead A, Crawford DL (2006) Neutral and adaptive variation in gene expression. Proceedings of the National Academy of Sciences 103:5425-5430.

Whitehead A, Roach JL, Zhang S, Galvez F (2011) Genomic mechanisms of evolved physiological plasticity in killifish distributed along an environmental salinity gradient. Proceedings of the National Academy of Sciences 108:6193-6198.

Wittkopp PJ, Kalay G (2012) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nature Reviews Genetics 13:59-69.

Wittkopp PJ, Haerum BK, Clark AG (2008) Regulatory changes underlying expression differences within and between Drosophila species. Nature Genetics 40:346-350.

Declarations

Funding

This research was supported by the Austrian Science Fund (FWF) under grant P46821-B (CompTransAdapt) and the French National Research Agency (ANR) under grant ANR-23-CE02-0041 (TransConv). RNA-seq was conducted at the CETU Genomics Core (Vienna) and the Institut Pasteur Transcriptomics Platform (Paris). Fish samples were provided by the Fundulus Genomics Consortium and the Danio Resource Center. High-altitude vertebrate samples were collected through collaboration with the Tibetan Academy of Sciences and the Universidad de los Andes under institutional research agreements. The funders had no role in study design, data collection, analysis, or manuscript preparation.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All raw RNA-seq FASTQ files for 8,847 individuals are deposited in the European Nucleotide Archive under project accession PRJEB84247. Processed count matrices, differential expression results (DESeq2 output), convergent DEG lists for all three adaptation classes, ASE measurements from F1 hybrid crosses, GO/KEGG enrichment results, TFBS enrichment outputs, and R analysis scripts

ISSN: 3117-6925 Vol: 2025, Issues: Issue3, 2025

are deposited in the Zenodo repository under DOI: 10.5281/zenodo.19465938. All data released under CC BY 4.0 license.

Ethical Approval

All vertebrate tissue sampling was conducted under applicable national research permits: Austria (BMBWF permit 2024-GZ-66.006/0247), France (APAFIS permit 2024-47821), China (MOST permit 2024-B47-0284), and Peru (SERNANP permit RD 2024-0284-SERNANP). All procedures followed the 3Rs principles and complied with EU Directive 2010/63/EU. Fish samples were collected using humane methods (MS-222 anaesthesia followed by rapid decapitation) approved by institutional animal care committees.

Appendix A

Complete Species Pair List and Top 50 Convergent DEGs per Adaptation Class

All 18 species pairs included in the comparative transcriptomics analysis are listed below, followed by the top 50 convergent differentially expressed genes for each adaptation class, ranked by number of independent species pairs in which convergence was detected.

COMPLETE SPECIES PAIR LIST (18 PAIRS)

Thermal adaptation (6 pairs): (1) Fundulus grandis/heteroclitus (Gulf/Atlantic coast killifish; liver, gill; 2 ecotypes); (2) Danio rerio/nigrivirgatus (warm/cool zebrafish congeners; liver, muscle); (3) Gadus morhua southern/northern ecotypes (warm/cold Atlantic cod; gill, liver); (4) Oryzias latipes/sakaizumii (medaka southern/northern; liver, gonads); (5) Oncorhynchus mykiss warm/cold ecotypes (rainbow trout; liver, gill); (6) Poecilia mexicana lowland/high-sulfide (not thermal per se, but metabolic convergence; liver).

Hypoxia adaptation (6 pairs): (1) Tibetan/lowland wolf (Canis lupus tibeticus/chanco; heart, lung); (2) Plateau/lowland pika (Ochotona curzoniae/princeps; heart, lung, muscle); (3) Andean/lowland hummingbird (Oreotrochilus estella/Amazilia; heart, muscle); (4) Bar-headed/greylag goose (Anser indicus/anser; heart, lung); (5) Tibetan/lowland pig (Sus scrofa domesticus Tibetan/lowland; heart, liver); (6) High-altitude/lowland deer mouse (Peromyscus maniculatus montanus/lowland; heart, muscle).

Dietary specialisation (6 pairs): (1) Dog-wolf vs. giant panda (carnivore/herbivore intestine, liver); (2) Snow leopard vs. giant panda (intestine, liver); (3) Bos taurus vs. Sus scrofa (ruminant herbivore/omnivore; intestine, pancreas); (4) Rabbit vs. ferret (Oryctolagus/Mustela; intestine, caecum); (5) Chimpanzee vs. gorilla (omnivore/folivore; intestine, liver); (6) Stickleback marine/freshwater (ancestral carnivore/novel herbivore diet; gut, liver).

TOP 20 THERMAL CONVERGENT DEGs (RANKED BY CONVERGENCE FREQUENCY)

6/6 pairs: HSPA1A, HSPA1B, HSPA8 (HSP70 family); DNAJB1, DNAJB4 (HSP40); HSPB1 (HSP27); FADS2 (fatty acid desaturase).

5/6 pairs: HSPA4, HSPA5 (BiP); FADS1; SCD1 (stearoyl-CoA desaturase); ATP5F1A; COX4I2 (cytochrome c oxidase); LDHA (lactate dehydrogenase); UCP2 (uncoupling protein 2).

4/6 pairs: HSPD1 (HSP60); HSP90AA1, HSP90AB1; EEF1A1 (translation elongation); SLC25A4 (adenine nucleotide translocator); ACSL1 (acyl-CoA synthetase); PLIN2 (lipid droplet).

3/6 pairs: Multiple additional metabolic, membrane, and chaperone genes; full list deposited at Zenodo (see data availability).

TOP 20 HYPOXIA CONVERGENT DEGs

6/6 pairs: HIF1A (hypoxia inducible factor 1 alpha); EPAS1 (HIF-2alpha).

5/6 pairs: HBA (haemoglobin alpha chain); VEGFA; BNIP3 (mitophagy regulator); LDHA; SLC2A1 (GLUT1 glucose transporter); PDK1 (pyruvate dehydrogenase kinase).

4/6 pairs: NDUFB4, NDUFS1, NDUFV1 (complex I subunits); HMOX1 (heme oxygenase); ANGPT2 (angiopoietin 2); CA9 (carbonic anhydrase 9); GAPDH; ENO1 (enolase 1).

3/6 pairs: Additional HIF-1 targets, erythropoiesis, and mitochondrial genes; full list at Zenodo.