FEVER: An interactive web-based resource for evolutionary transcriptomics across fishes

Documentation

FEVER is free and open to all users and there is no login requirement.

Introduction

Gene duplication and subsequent expression changes are thought to underlie many phenotypic differences across animal species. Teleost fish represent a unique opportunity to study this phenomenon as they represent half vertebrates and exhibit an outstanding phenotypic diversity due to their adaption to a wide range of ecological niches. Notably, their ancestor underwent a Whole Genome Duplication event (~300 mya) that provided new raw genetic material for natural selection to act on. Since then, intense genomic reshuffling occurred which probably facilitated the diversification of species and organs. Additionally, fish species such as zebrafish and medaka represent major models in various research fields such as developmental biology, biomedical research, ecology and evolution. Recent sequencing endeavors provided high-quality genomes for species covering the main fish evolutionary lineages. However, transcriptomic data across fish species and organs are still scarce and have not been integrated with newly sequenced genomes making gene expression quantification and comparative analyses particularly challenging. Thus, tools allowing the exploration of gene evolutionary history and their expression profiles across species and organs are still lacking. Here, we present FEVER, a web-based resource allowing evolutionary genomics and transcriptomics across species and tissues. First, based on query genes, FEVER reconstructs gene trees providing orthologous and paralogous relationships across 13 species covering the major fish lineages, and 4 model species as evolutionary outgroups. Second, it provides unbiased gene expression across 11 tissues using up-to-date genomes. Finally, genomic and transcriptomic data are combined together allowing the exploration of gene expression evolution following speciation and duplication events. FEVER is implemented in R and is accessible at https://fever.sk8.inrae.fr/ with all major browsers.

Gene tree reconstruction

To investigate the evolutionary history of extant genes across fishes, we reconstruct reconciled trees providing orthologous and paralogous relationships across 13 species covering the major fish lineages (rainbow trout, brown trout, eastern mudminnow, northern pike, medaka, Atlantic cod, Mexican tetra, striped catfish, zebrafish, allis shad, European eel, spotted gar, and bowfin) and 4 model species as evolutionary outgroups (human, mouse, drosophila and round worm).
First, the longest transcripts of all genes from the 17 aforementioned species are translated into proteins. Next, the proteins are aligned to each other (diamond (1), v2.0.8, parameters: -k 0 --outfmt 6 --evalue 1e-5) and clustered together based on alignment scores (hcluster_sg, http://treesoft.svn.sourceforge.net/viewvc/treesoft/branches/lh3/hcluster,v0.5.1-2, parameters: -m 750 -w 0 -s 0.34) to build families composed of homologous genes (orthologous and paralogous). Multiple alignments are then performed between all proteins composing each family (T-coffee (2), v11.00, parameters: -type=PROTEIN -method mafftgins_msa, muscle_msa,kalign_msa). Next, TreeBeSt is used to perform protein-guided alignments based on previous multiple protein alignments and coding sequences (TreeBeSt, https://github.com/Ensembl/treebest, v1.9.2, parameters: backtrans). Based on those alignments and the species tree, TreeBeSt (https://github.com/Ensembl/treebest, v1.9.2, parameters: best) reconstructs gene trees that are bootstrapped (100 times), reconciled with the

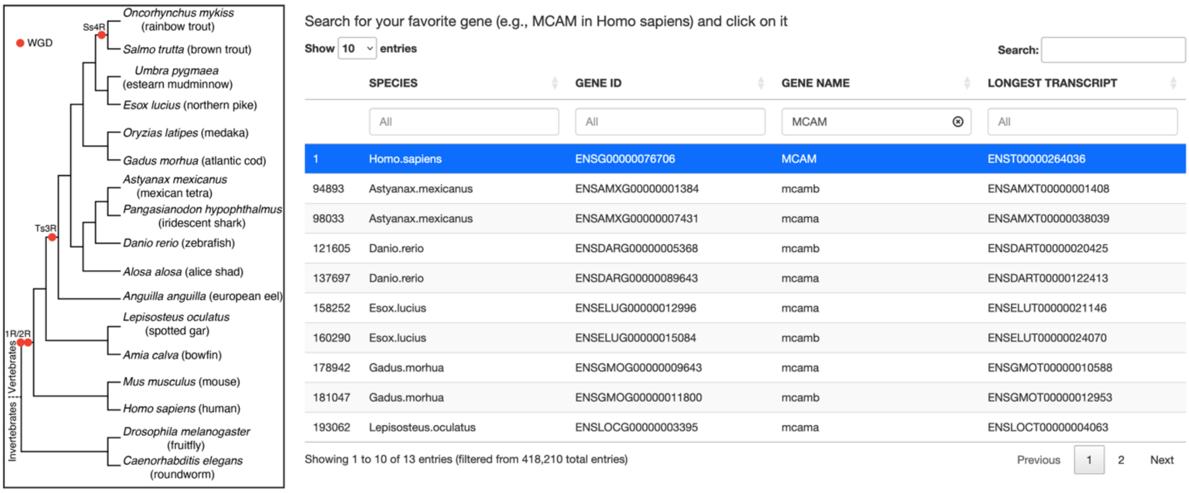
species tree and rooted by minimizing the number of duplications and losses. Finally, those trees provide orthologous and paralogous genes across 17 species as well as their evolutionary history (speciation and duplication events). This methodology is inspired from landmark resources (3–6).

Gene expression

RNA-seq datasets across 11 tissues (brain, gills, heart, muscle, liver, kidney, bones, intestine, embryo, ovary, and testis) and 13 species (rainbow trout, brown trout, eastern mudminnow, northern pike, medaka, atlantic cod, mexican tetra, striped catfish, zebrafish, allis shad, european eel, spotted gar, and bowfin) were downloaded from a previous study (7). Raw reads with known 3' adaptor and low-quality bases (Phred score < 20) were trimmed with TrimGalore (v.0.6.6) (https://github.com/FelixKrueger/TrimGalore) (parameters: -r_clip 13 -three_prime_clip 2). Next, we mapped the trimmed reads from each library against reference genomes and annotated transcripts using Salmon (v.1.8) (8) (parameters: defaults). Gene-expression levels were measured in transcripts per kilobase million (TPM), a unit which corrects for both feature length and sequencing depth. Tissue-specificity indexes are based on the Tau metric of tissue specificity (9) ranging from 0 (broad expression) to 1 (restricted expression). Tau is calculated for each gene across brain, gills, heart, muscle, liver, kidney, bones, intestine, embryo, ovary and testis. Tau is not calculated when at least one expression value is missing. A gene is considered tissue-specific when its Tau is greater than 0.9.

Usage and Interpretation

To investigate the evolutionary history of a given gene across fishes, the user needs to provide a gene of interest via its gene name, ID, or its longest transcript ID in a given species (example in Figure1).


WGD
Ss4R
Ts3R
1R/2R
Invertebrates
Vertebrates
Oncorhynchus mykiss (rainbow trout)
Salmo trutta (brown trout)
Umbra pygmaea (estearn mudminnow)
Esox lucius (northern pike)
Oryzias latipes (medaka)
Gadus morhua (atlantic cod)
Astyanax mexicanus (mexican tetra)
Pangasianodon hypophthalmus (iridescent shark)
Danio rerio (zebrafish)
Alosa alosa (alice shad)
Anguilla anguilla (european eel)
Lepisosteus oculatus (spotted gar)
Amia calva (bowfin)
Mus musculus (mouse)
Homo sapiens (human)
Drosophila melanogaster (fruitfly)
Caenorhabditis elegans (roundworm)
Show 10 entries
Search:
Previous
1
2
Next

Search for your favorite gene (e.g., MCAM in Homo sapiens) and click on it

	SPECIES	GENE ID	GENE NAME	LONGEST TRANSCRIPT
	All	All	MCAM	All
1	Homo.sapiens	ENSG00000076706	MCAM	ENST00000264036
94893	Astyanax.mexicanus	ENSAMXG00000001384	mcamb	ENSAMXT00000001408
98033	Astyanax.mexicanus	ENSAMXG00000007431	mcama	ENSAMXT00000038039
121605	Danio.rerio	ENSDARG00000005368	mcamb	ENSDART00000020425
137697	Danio.rerio	ENSDARG00000089643	mcama	ENSDART00000122413
158252	Esox.lucius	ENSELUG00000012996	mcama	ENSELUT00000021146
160290	Esox.lucius	ENSELUG00000015084	mcamb	ENSELUT00000024070
178942	Gadus.morhua	ENSGMOG00000009643	mcama	ENSGMOT00000010588
181047	Gadus.morhua	ENSGMOG00000011800	mcamb	ENSGMOT00000012953
193062	Lepisosteus.oculatus	ENSLOCG00000003395	mcama	ENSLOCT00000004063

Showing 1 to 10 of 13 entries (filtered from 418,210 total entries)

Figure 1: Example of gene entry.

To illustrate the features of FEVER, we provide and interpret the results of FEVER using *MCAM* as a query gene that is involved in cell adhesion and in cohesion of the endothelial monolayer at intercellular junctions in vascular tissue. FEVER revealed that this gene was duplicated by the teleost-specific WGD (Ts3R) and maintained in two copies (*MCAMA* and *MCAMB*) in most of them (Figure 2A). At the expression level, we note that *MCAMA* is specific to brain and *MCAMB* to heart in most teleost fish species, while the pre-Ts3R *MCAM* gene was likely to be particularly expressed in heart (Figure 2B), as inferred from *MCAM* expression in bowfin and mammals (from external data: https://apps.kaessmannlab.org/evodevoapp/). Additionally, *MCAMA* and *MCAMB* were duplicated once again by the salmonid-specific WGD (Ss4R) resulting in four *MCAM* copies in the brown trout for instance. Interestingly, one post-Ss4R *MCAMA* paralog became specific to testis while the other one kept the pre-Ss4R *MCAMA* brain specificity, and the two post-Ss4R *MCAMB* paralogs maintained the pre-Ss4R *MCAMB* heart specificity. For an easier comparison across species and tissues, bar plots showing expression values (TPM) are also provided (Figure 2C). Expression values can be downloaded from the web service (csv format). This example illustrates that gene specificity may change following WGD events and that FEVER is a useful tool to assess gene evolutionary dynamics and potential sub/neo-functionalization.

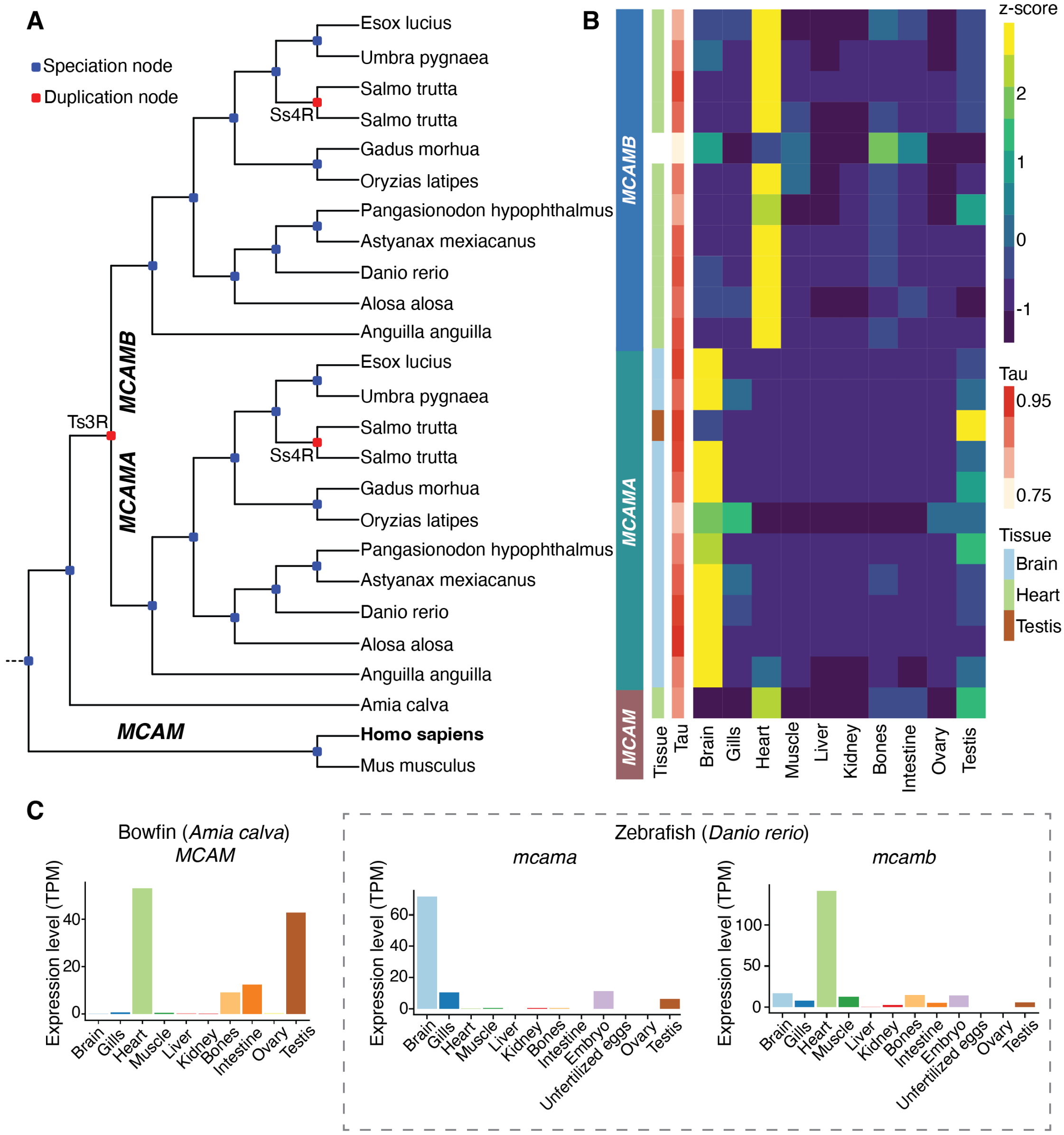


Figure 2: Example of the output generated by FEVER based on a query gene, *MCAM* (melanoma cell adhesion molecule). (A**)** FEVER displays the gene tree containing the query gene with duplication and speciation nodes annotated, in red and blue, respectively. (B**)** The heatmap shows gene expression across species and tissues as well as tissue-specificity indexes (Tau) and the tissue the genes are specific to, when applicable. (C**)** The bar plots show expression values (TPM) across tissues and organs for genes belonging to the same tree of the query gene. Expression values for *MCAM* in bowfin and *mcama*/*mcamb* in zebrafish are shown as a case study.

References

1. Buchfink,B., Xie,C. and Huson,D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*, **12**, 59–60.
2. Notredame,C., Higgins,D.G. and Heringa,J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, **302**, 205–17.
3. Muffato,M., Louis,A., Poisnel,C.-E. and Roest Crollius,H. (2010) Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes. *Bioinformatics*, **26**, 1119–21.
4. Nguyen,N.T.T., Vincens,P., Dufayard,J.F., Roest Crollius,H. and Louis,A. (2022) Genomicus in 2022: comparative tools for thousands of genomes and reconstructed ancestors. *Nucleic Acids Res.*, **50**, D1025–D1031.
5. Martin,F.J., Amode,M.R., Aneja,A., Austine-Orimoloye,O., Azov,A.G., Barnes,I., Becker,A., Bennett,R., Berry,A., Bhai,J., *et al.* (2023) Ensembl 2023. *Nucleic Acids Res.*, **51**, D933–D941.
6. Harrison,P.W., Amode,M.R., Austine-Orimoloye,O., Azov,A.G., Barba,M., Barnes,I., Becker,A., Bennett,R., Berry,A., Bhai,J., *et al.* (2023) Ensembl 2024. *Nucleic Acids Res.*, 10.1093/nar/gkad1049.
7. Pasquier,J., Cabau,C., Nguyen,T., Jouanno,E., Severac,D., Braasch,I., Journot,L., Pontarotti,P., Klopp,C., Postlethwait,J.H., *et al.* (2016) Gene evolution and gene expression after whole genome duplication in fish: The PhyloFish database. *BMC Genomics*, **17**, 1–10.
8. Patro,R., Duggal,G., Love,M.I., Irizarry,R.A. and Kingsford,C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*, **14**, 417–419.
9. Yanai,I., Benjamin,H., Shmoish,M., Chalifa-Caspi,V., Shklar,M., Ophir,R., Bar-Even,A., Horn-Saban,S., Safran,M., Domany,E., *et al.* (2005) Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, **21**, 650–659.