*Molecular Biology and Evolution*, 2025, **42**, 1–21
https://doi.org/10.1093/molbev/msaf107
Advance access publication 16 May 2025
**Discoveries**

MBE OXFORD

# Multiple Horizontal Transfers of Immune Genes Between Distantly Related Teleost Fishes

**Maxime Policarpo** [1,4,*] **Walter Salzburger** [1] **Florian Maumus** [2,5] **Clément Gilbert** [3,*]

[1]Zoological Institute, Department of Environmental Sciences, University of Basel, Basel, Switzerland
[2]URGI, INRAE, Université Paris-Saclay, Versailles 78026, France
[3]Université Paris-Saclay, CNRS, IRD, UMR Évolution, Génomes, Comportement et Écologie, Gif-sur-Yvette 91198, France
[4]Present address: Evolution of Sensory and Physiological Systems, Max Planck Institute for Biological Intelligence, Martinsried, Germany
[5]Present address: Université Paris-Saclay, INRAE, AgroParisTech, Institute Jean-Pierre Bourgin for Plant Sciences (IJPB), 78000, Versailles, France

*Corresponding authors: E-mails: maxime.policarpo@bi.mpg.de; clement.gilbert1@universite-paris-saclay.fr.
**Associate editor**: Irina Arkhipova

## Abstract

Horizontal gene transfer (HGT) is less frequent in eukaryotes than in prokaryotes, yet can have strong functional implications and was proposed as a causal factor for major adaptations in several eukaryotic lineages. Most cases of eukaryote HGT reported to date are inter-domain transfers, and few studies have investigated eukaryote-to-eukaryote HGTs. Here, we performed a large-scale survey of HGT among 242 species of ray-finned fishes. We found multiple lines of evidence supporting 19 teleost-to-teleost HGT events that involve 17 different genes in 11 teleost fish orders. The genes involved in these transfers show lower synonymous divergence than expected under vertical transmission, their phylogeny is inconsistent with that of teleost fishes, and they occur at non-syntenic positions in donor and recipient lineages. The distribution of HGT events in the teleost tree is heterogenous, with 8 of the 19 transfers occurring between the same two orders (Osmeriformes and Clupeiformes). Though we favor a scenario involving multiple HGT events, future work should evaluate whether hybridization between species belonging to different teleost orders may generate HGT-like patterns. Besides the previously reported transfer of an antifreeze protein, most transferred genes play roles in immunity or are pore-forming proteins, suggesting that such genes may be more likely than others to confer a strong selective advantage to the recipient species. Overall, our work shows that teleost-to-teleost HGT has occurred on multiple occasions, and it will be worth further quantifying these transfers and evaluating their impact on teleost evolution as more genomes are sequenced.

**Keywords:** horizontal gene transfers, immunity, teleost fishes

## Introduction

Horizontal gene transfer (HGT) is the passage of genes between organisms through ways other than reproduction (Soucy et al. 2015). In archaea and bacteria, where several mechanisms and vectors are specifically dedicated to such transfers of genetic material (Haudiquet et al. 2022), HGTs occur frequently and are key to their rapid adaptation to changing environmental conditions (Arnold et al. 2022). By contrast, the extent to which HGT has shaped eukaryotic genomes and their evolution remains debated (Martin 2017; 2018; Roger 2018; Leger et al. 2018). The frequency of HGT is much lower in eukaryotes than in prokaryotes. In unicellular eukaryotes, in which horizontal transfer has been more thoroughly studied, only one to a few percent of genes have been acquired through HGT (Van Etten and Bhattacharya 2020). The proportion of foreign genes is even lower in most multicellular eukaryotes investigated so far (but see Wilson et al. (2024)). That HGT is rarer in eukaryotes than in prokaryotes may be in part explained by differences in HGT mechanisms and selection regimes acting on horizontally transferred genes in the two domains of life (Keeling 2024). However, although rare, it is likely that HGT substantially impacted eukaryote evolution on several occasions, and was even instrumental to the adaptation of multiple lineages to new ecological niches (Keeling 2009; Danchin 2016). For example, in arthropods, HGT from bacteria and other sources likely facilitated the replicate evolution toward herbivory in mites and in species belonging to at least five insect orders (Xia et al. 2021; Gilbert and Maumus 2022; Kirsch et al. 2022; Wybouw et al. 2016; Wybouw et al. 2018). Major HGT episodes, again mainly from bacterial sources, may also have facilitated the colonization of land by plants (Ma et al. 2022). Another remarkable example is that of the interphotoreceptor retinoid-binding protein (IRBP) gene, which is involved in the formation of the vertebrate eye, and was acquired by the last common ancestor of vertebrates through HGT from bacteria (Kalluraya et al. 2023).

HGT detection is difficult to automate in eukaryotes, and problems linked to contamination have long plagued the field (Bemm et al. 2016; Salzberg 2017). In addition, hypotheses alternative to HGT such as incomplete lineage sorting, multiple gene losses, duplication followed by differential gene loss, and/or variation in evolutionary rates among lineages are difficult to exclude in many cases, especially when taxon sampling is sparse (Gabaldón 2020; Cote-L'Heureux et al. 2022). A good illustration of this is that several initial reports of HGT into the human genome have later been shown to be due to technical artifacts (Salzberg et al. 2001; Willerslev et al.

**Received:** December 17, 2024. **Revised:** April 29, 2025. **Accepted:** May 2, 2025

2002; Salzberg 2017). Other scenarios involving hybridization, possibly followed by introgression, might also sometimes be difficult to distinguish from HGT. Hybridization between closely related species is relatively common in nature, and dedicated methods exist to detect and quantify loci with a signal of hybridization-mediated interspecific gene flow (Edelman and Mallet 2021; Hibbins and Hahn 2022). Here, chromosomes of parental species may be sufficiently similar to allow chromosomal pairing and recombination to occur during meiosis in hybrids. Such hybrids may then backcross with parental species, yielding many introgressed genomic regions scattered along the genome, with conserved synteny between species (Fontaine et al. 2015; Jones et al. 2018; Edelman and Mallet 2021). Hybridization may also occur between more distantly related species, leading to the formation of allopolyploids containing one or more chromosome sets from each parental species. In cases where such hybrids are viable and fertile, homology between short chromosome regions may be sufficient to allow homeologous exchange through recombination, which can eventually be followed over multiple generations by reduction in chromosome number, chromosome mosaicism, and diploidization (Deb et al. 2023). Finally, cases of hybridization between species diverging dozens of millions of years ago have also been reported in some taxa, such as plants and fishes (Chen et al. 2018a). It has been proposed that such long-distance hybridization could lead to situations whereby two genomes from distant parental species are transiently in contact in a hybrid embryo, allowing some sequences to transfer from one to the other, before the chromosomes of one parental species are eliminated during early embryogenesis (Pereira et al. 2023). Should such a hybrid survive and be fertile, the transferred DNA sequences could be stably inherited, potentially mimicking a pattern resulting from HGT.

In spite of these limitations, several large-scale studies have reported hundreds of solid cases of HGT in multiple eukaryotic lineages such as microbial eukaryotes (Van Etten and Bhattacharya 2020; Cote-L'Heureux et al. 2022), plants (Mahelka et al. 2017; Hibdige et al. 2021; Ma et al. 2022), fungi (Marcet-Houben and Gabaldón 2010; Cote-L'Heureux et al. 2022; Sahu et al. 2023; Ciach et al. 2024), and insects (Li et al. 2022). However, most of these studies tended to focus on inter-domain HGT rather than within-eukaryote transfers, because the power to infer HGT (and discard alternative hypotheses) increases as the phylogenetic distance between donor and receiving lineages augments (Adato et al. 2015). Thus, although cases of eukaryote-to-eukaryote HGT have been inferred (Gasmi et al. 2015; Szöllősi et al. 2015; Mahelka et al. 2017; Hibdige et al. 2021; Gilbert and Maumus 2022; Mishina et al. 2023; Sahu et al. 2023; Ciach et al. 2024), this type of HGT is rarely the subject of large-scale, multi-taxa analyses.

To start filling this gap, we performed a systematic search for HGT events within a vertebrate lineage, ray-finned fishes. We focused on this group because vertebrates in general, and fishes in particular, have never been the subject of large-scale HGT surveys and numerous well-annotated fish genomes are available. Furthermore, the only known cases of vertebrate-to-vertebrate HGT reported so far involve two fish lineages. Phylogenetic and comparative genomics analyses revealed that the presence of a gene coding a type II antifreeze protein (AFP) in species from three fish orders (Osmeriformes, Clupeiformes, and Perciformes) was best explained by a scenario involving multiple HGT events (Graham and Davies 2021; Han et al. 2023). Evidence supporting HGT of this gene included similarity at the nucleotide level higher than expected under vertical transmission in coding and non-coding regions (84% to 95% over > 5 kb) between species separated by more than 200 million years (the Atlantic herring [*Clupea harengus*] and the rainbow smelt [*Osmerus mordax*]), as well as absence of synteny at the gene locus in the two species (Graham and Davies 2021 ). Remarkably, the presence of co-transferred transposable elements (TEs) in the flanking regions of the gene that are found hundreds of times in the herring genome but are absent elsewhere in the smelt genome allowed investigators to infer the direction of HGT, which likely occurred from herring to smelt (Graham and Davies 2021). In addition to this HGT case study, a large-scale survey of horizontal transfer of TEs unveiled no less than 975 independent such events among vertebrates, the vast majority of which involved teleost fishes (Zhang et al. 2020a). TEs can jump from one locus to another and generate many copies of themselves in a given genome, which might have increased their propensity to transfer compared to regular, static genes. In any case, these results can be taken as an indication that fish might be more prone to horizontal transfer (including HGT) than other vertebrates.

Here, we develop a semi-automated, multipronged approach to detect HGT among teleost fishes, excluding horizontal transfer of TEs, which has been previously studied (Kuraku et al. 2012; Zhang et al. 2020a). We show that, in addition to the AFP HGT (that we were able to recover), at least 16 different genes have undergone 18 teleost-to-teleost HGT events among 11 teleost fish orders separated by ≍ 140 to 260 million years of evolution. While we cannot totally exclude the possibility that some of these transfers occurred through some form of long-distance hybridization, we currently favor the hypothesis of HGT events involving some kind of viral vectors or sperm-mediated gene transfers, as proposed earlier for the AFP HGT (Graham and Davies 2021).

## Results

### Selecting Genes With Very Low Synonymous Divergence Between Fish Species

To identify HGT between teleosts, we took advantage of 225 complete or nearly complete ray-finned fishes' genomes (BUSCO score ≥ 80%) from 48 orders, with available annotations on RefSeq or GenBank. We further improved the taxonomic coverage of this dataset by annotating 17 additional genomes (Fig. 1, supplementary fig. S1, Supplementary Material online, supplementary files S1 and S2, Supplementary Material online), belonging to 14 orders for which no annotated genomes were available. We then applied a multipronged and highly conservative approach to this dataset, largely inspired from published HGT detection pipelines and metrics. Our objective was twofold:(i) to retrieve only HGT cases supported by multiple lines of evidence akin to that of the AFP gene (Graham and Davies 2021) and (ii) to use a series of filters producing a reasonable number of HGT candidates so that manual curation of the phylogenies of these candidates can be done (see next section). In agreement with the hypothesis proposed by Cote-L'heureux et al. (2022), we postulated that in instances where multiple gene losses mimic a pattern that may be erroneously interpreted as HGT, gene divergence should be consistent with vertical transmission. In contrast, horizontally transferred genes should show a lower divergence than most vertically transmitted genes.

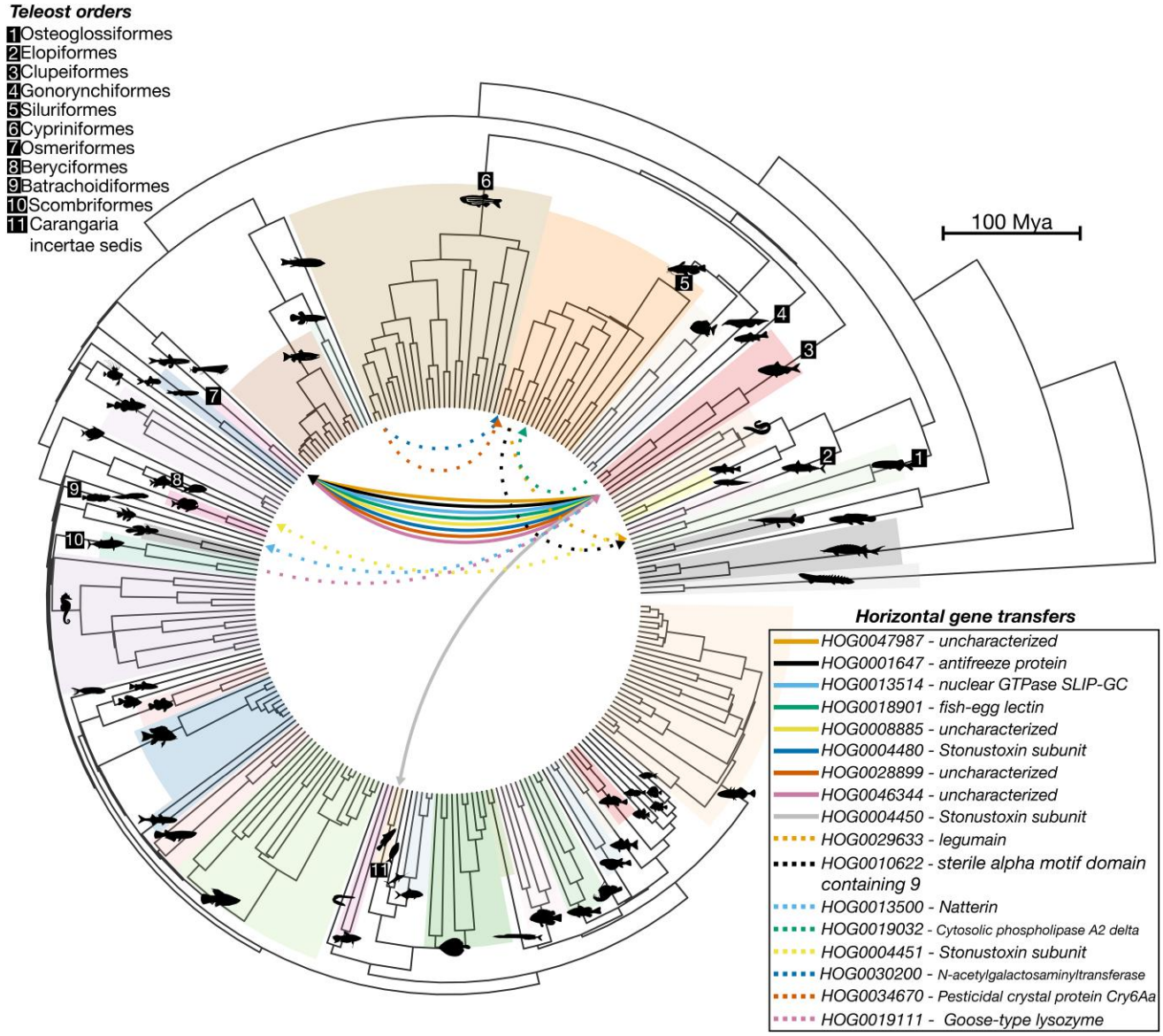


**Fig. 1.** HGTs between teleosts. Phylogeny of annotated ray-finned fishes inferred using ASTRAL-III (with 3,584 BUSCO gene phylogenies) and dated using the least squares dating method. The same species tree with species name and annotations sources can be found in supplementary fig. S1, Supplementary Material online. Alternating shades of color, and silhouettes, represent the different species orders. Orders involved in putative HGTs are annotated with numbers. Branches with no color shades represent orders with only one species in our dataset. Species silhouettes were retrieved in PhyloPic.org (supplementary file S2, Supplementary Material online). Arrows inside the tree represent putative HGTs and are drawn from one representative species of the donor clade to one representative species of the recipient clade. Arrows line type and colors are defined based on the HOG. HOG numbers and names (based on the best HHpred match) are indicated.

To identify the least diverged genes, we classified 5,859,286 genes extracted from the 242 fish genomes in 75,363 phylogenetic hierarchical orthologous groups (HOG, see Materials and Methods) (that is, 97% of the 6,035,924 annotated teleost genes, supplementary fig. S2, Supplementary Material online). Among each HOG, dS (number of synonymous substitutions per synonymous site) and coding sequence identities were computed between all possible pairs of sequences aligned over at least 100 codons. HOGs appeared to be accurate, as mean dS and quantile values computed from these were highly similar to the values computed using BUSCO groups, which include only single-copy and ultra-conserved genes, that are most likely always vertically transmitted (supplementary fig. S3-a,b, Supplementary Material online). Then, for each possible species pair, we produced a dS distribution, consisting of all dS values computed between the two species across all HOGs (supplementary fig. S3-c,d,e,f, Supplementary Material online). Considering that the dS values of horizontally transferred genes would be located on the left tails of these distributions, we extracted the quantile 0.5% (q0.5). As expected, mean and q0.5 dS values were strongly correlated to divergence times between species, but variances were high, likely reflecting heterogeneity in mutation rates across lineages (supplementary fig. S3-c,d, Supplementary Material online). At this stage, we kept 2,558,049 gene pairs, belonging to 6,367 distinct HOGs, which had a dS value below q0.5.

## Using Lack of Synteny to Support Horizontal Transfer

Next, following Adato et al. (2015), we reasoned that it would be highly unlikely for a horizontally transferred gene to be

located at the same locus, i.e. between the same orthologous genes, in the donor and recipient species. For each pair of species, and for each gene retained above based on low dS and belonging to the same HOG, we counted how many of the ten genes located upstream and the ten genes located downstream of the gene in question in the two species belonged to an identical HOG. This enabled us to compute a micro-syntenic score (akin to the synteny index of Adato et al. 2015), which represents the number of genes belonging to an identical HOG among the 20 flanking genes. A score of 0 indicates the complete lack of microsynteny for a given gene, while a score of 20 indicates complete microsynteny conservation (supplementary fig. S4-a, Supplementary Material online). As expected, the mean micro-syntenic score was highly correlated to divergence time, i.e. the more distantly related species are, the less conserved microsynteny blocks become (supplementary fig. S4-b, Supplementary Material online). The annotation method and assembly quality also had an impact: genomes with RefSeq annotations (representing high-quality annotations) had significantly higher mean micro-syntenic scores than other genomes (supplementary fig. S4-b, c, Supplementary Material online), and genomes with lower scaffold N50 had significantly lower scores (supplementary fig. S4-c, Supplementary Material online). Note that these micro-syntenic scores are slightly biased and tend to overestimate synteny, as two genes in the same HOG, which are orthologs at the teleost most recent common ancestor (MRCA) level, are not necessarily orthologs at the MRCA of two species. They could instead be paralogs, if one or more duplications occurred between the teleost MRCA and the MRCA of these two species. We discarded gene pairs that had non-null micro-syntenic scores, which narrowed down the list of candidate HGTs to 594,324 gene pairs, belonging to 2,381 distinct HOGs.

## Detection of HGT based on Alien Indexes

In parallel to this parametric method, we also used similarity-based approaches, in ways largely inspired by previous studies (Gladyshev et al. 2008; Boschetti et al. 2012; Koutsovoulos et al. 2022). For this, we performed all-versus-all gene BLASTN searches and the results were filtered using three strategies (detailed in Materials and Methods): (i) mismatch between a query corresponding species order and the best-match order (BLASTN-A), (ii) best-match divergence time higher than expected (BLASTN-B), and (iii) matches more similar to the query than expected, using the recently described HGTindex alien metrics (Yuan et al. 2023) (BLASTN-C). Again, after discarding gene pairs with a non-null micro-syntenic score, these methods retained 1,000, 1,177, and 522 gene pairs, belonging to 579, 612, and 283 distinct HOGs, respectively.

A total of 3,213 (non-redundant) HOGs showing putative evidence of HGT by any of the four methods described above (dS, BLAST-A, BLAST-B, or BLAST-C strategies) were retained for further analysis. Among them, 139 HOGs with at least one candidate transfer were retrieved by all three BLASTN strategies, while only 56 HOGs were found in common between our parametric method relying on dS values and all three BLASTN strategies (Fig. 2-a). The largest intersection was found between the BLASTN-A and the BLASTN-B methods, with 270 retained HOGs in common. This relatively low degree of overlap between the results of the different methods may be explained by the fact that these methods rely on different principles. Thus, they may retrieve different false positive HGT candidates, and they may identify different true positives that were not validated by manual curation of phylogenetic trees (see below).

## Phylogenetic Analysis and Manual Curation of Candidate Teleost-to-Teleost HGTs

As the phylogeny of horizontally transferred genes is generally inconsistent with that of the host taxa (Ravenhall et al. 2015), we computed maximum likelihood phylogenies for the 3,213 retained HOGs and we manually inspected the resulting trees, as is commonly done in HGT studies (Jaramillo et al. 2015; Shen et al. 2018; Li et al. 2022). We excluded ladder-shape trees (Jaramillo et al. 2015) and selected trees consistent with a HGT scenario, i.e. trees in which genes from paraphyletic species in the teleost phylogeny form a strongly supported monophyletic group (Fig. 3-b).

In total, we retained 18 HGTs involving 16 distinct HOGs (Fig. 2-b, Table 1, supplementary file S1, Supplementary Material online). We ensured that the tree topologies of the genes involved in transfer remained consistent with an HGT scenario when based on the same alignment but trimmed to keep the most conserved sites (supplementary file S1, Supplementary Material online). We further verified that the HGT scenario still holds when reconstructing a phylogeny based on an independent alignment, performed with the nearest sequences found in the UniProt database (The UniProt Consortium 2021) and in a database of ray-finned fishes' proteins (supplementary file S1, Supplementary Material online). These phylogenetic analyses allowed us to infer the direction of the transfer (Fig. 3; supplementary File S1, Supplementary Material online) based on the premise that the receiving clade should appear embedded in the donor clade in the gene tree. Many HGT candidates produced phylogenetic topologies that were potentially consistent with the HGT hypothesis but did not receive sufficient bootstrap support to be included in the final list of candidates considered to be solid. It is however noteworthy that the majority of HGT cases considered to be solid (15 out of 18) are retrieved by at least two of the methods described in the previous sections (Fig. 2-b, Table 1). Eight HGTs, in six distinct HOGs, were detected by all the methods. In addition to passing the q0.5 threshold established for all HOGs at the level of each pair of species (Fig. 3-a and supplementary File S1, Supplementary Material online), the dS of HGT candidates are also very low in the dS distribution of genes from the same HOG for all pairs of species (Fig. 3-d, supplementary File S1, Supplementary Material online). Thus, the lower-than-expected dS of these genes cannot be explained by a peculiar selection regime acting on some HOGs. Furthermore, all transferred genes were found either in two different species, for both the donor and recipient clades, or in two independent genome assemblies of the species involved. It is thus unlikely that the genes uncovered here result from contamination rather than HGT.

Among these HGTs, we retrieved the only vertebrate-to-vertebrate HGT event described to date: the transfer of an AFP from *Clupea harengus* (Clupeiformes) to Osmeriformes (*HOG0001647*) (Graham et al. 2008; Graham and Davies 2021). Surprisingly, eight HGTs—about half of all HGTs detected in this study—occurred between these two clades (Fig. 1, Fig. 3, Table 1, supplementary file S1, Supplementary Material online). There is no indication that these genes were transferred together, as they do not

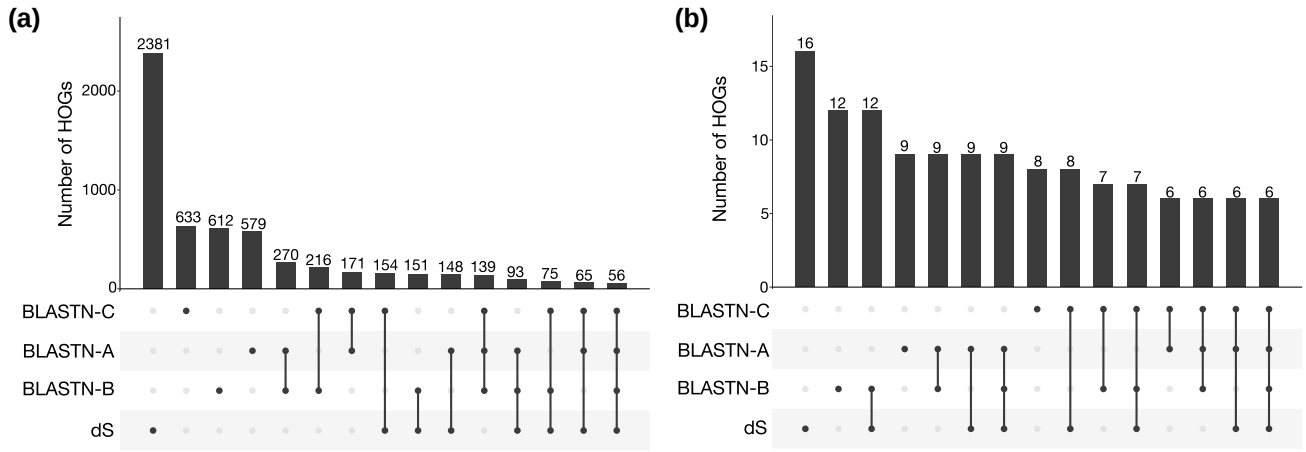


**Fig. 2.** HGTs detection methods. Upset plots showing the number of HOGs retained by the parametric procedure (based on dS) and by the three BLAST procedures (A, B, and C). a) Number of unique HOG with at least one candidate HGT retained before manual inspection. b) Number of unique HOG with at least one HGT discussed in this study (after manual curation).

colocalize in the donor and recipient clades (in the two representative highly continuous genomes *Clupea harengus* and *Hypomesus transpacificus*, respectively) (supplementary fig. S5, Supplementary Material online).

To further verify that these results were not due to contamination, we performed competitive mapping of the raw illumina reads from the two species against a set of mitochondrial genomes. We found that nearly all mapped reads from *H. transpacificus* (99.6% to 100% depending on the sequencing run) mapped to its mitochondrial genome, with no or very few reads mapping to *C. harengus* mitochondria (0% to 0.002%) (supplementary fig. S6, Supplementary Material online). The same observation was made using *C. harengus* reads, which nearly all mapped to its mitochondrial genome (96% to 100%), but not to *H. transpacificus* mitochondria (0% to 0.03%) (supplementary fig. S6, Supplementary Material online). In addition, for each Clupeiformes-to-Osmeriformes transfer, we found multiple long reads from *C. harengus* and *H. transpacificus* that spanned the entire transferred genes, including flanking regions, in each genome (supplementary fig. S7, Supplementary Material online, supplementary file S2, Supplementary Material online). For three out of the eight transfers, we could even find multiple long reads from *H. transpacificus* spanning the entire transferred genes when mapped to *C. harengus*, and vice-versa (supplementary fig. S7, Supplementary Material online). Similarity between these long reads was restricted to transferred genes and their immediately flanking regions and did not extend over their entire length (supplementary fig. S8, Supplementary Material online). This pattern mirrored corresponding regions of the genome assemblies (supplementary File S1, Supplementary Material online). Importantly, no long reads from a third species devoid of the transferred genes used as control (*Borostomias antarcticus*, closely related to Osmeriformes) spanned any of these genes on the genome of *C. harengus* and *H. transpacificus* (supplementary file S2, Supplementary Material online, supplementary fig. S7, Supplementary Material online).

Interestingly, the *HOG0019111* phylogeny topology suggested that Clupeiformes acquired this gene two times from Scombridae, once in the MRCA of *Alosa* and *Sardina* species, and once in *Clupea spp.*, before its subsequent transfer to Osmeriformes (supplementary File S1, Supplementary Material online, Table 1). Besides these candidate transfers between Osmeriformes and Clupeiformes, we found two HGTs from Cypriniformes to Pangasiidae, one from Clupeiformes to Carangaria, one from Clupeiformes to Batrachoididae, one from *Chanos spp.* to *Neoarius spp.*, and one from *Megalops* species to Berycidae species. In addition, we observed a putative transfer of *HOG0029633* from catfishes (Siluriformes) to *Paramormyrops* species (Osteoglossiformes). While investigating the flanking regions of this gene, we found a second likely co-transferred HOG, *HOG0046344*, for which the gene topology was also consistent with an HGT scenario, but pairwise dS values were slightly above the q0.5 threshold (< q0.6). Despite these two genes being next to each other in both the donor and recipient clades, null micro-syntenic scores were observed for *HOG0029633*, due to the absence of *HOG0046344* in several catfish species. Finally, one independent HGT (*HOG0010622*) was additionally retrieved between those two clades. Thus, a total of 19 HGTs (in 17 distinct HOGs) were found across teleost species.

Additional evidence supporting these HGTs was retrieved and is summarized for each HOG in supplementary file S1, Supplementary Material online, with the example of *HOG00 47987* shown in Fig. 3. First, for most HGTs (17 out of 19), we were able to observe high sequence similarity at the nucleotide level in flanking regions and/or introns, as reported for the AFP gene (Graham et al. 2008; Graham and Davies 2021). The length of regions showing similarity at the nucleotide level, and thus, the minimal length of transferred regions between the donors and recipients were between ≈2 kb and ≈15 kb (mean = 6.3 kb; median = 5 kb, supplementary table S1, Supplementary Material online). We used these non-coding regions to perform similarity searches (BLASTN) against all genome assemblies in our dataset, and obtained significant hits only to the corresponding donor and recipient clades. Thus, these regions do not correspond to teleost conserved non-coding elements (CNEs), in which one would expect high conservation in many species, through purifying selection. Furthermore, for seven HGTs (including the AFP HGT reported by Graham et al., here *HOG*0001647), we found that the donor and recipient regions harbored at least one common and highly similar TE copy. As reported for the AFP transfer, the TEs found in the vicinity of two additional HGTs (*HOG0 047987 and HOG0018901*) occurring between *Clupea harengus* and *Osmeriformes* further supported the Cluperiformes-to-Osmeriformes direction of the transfer initially inferred based on the HOG topology. Indeed, for each of these three cases, we

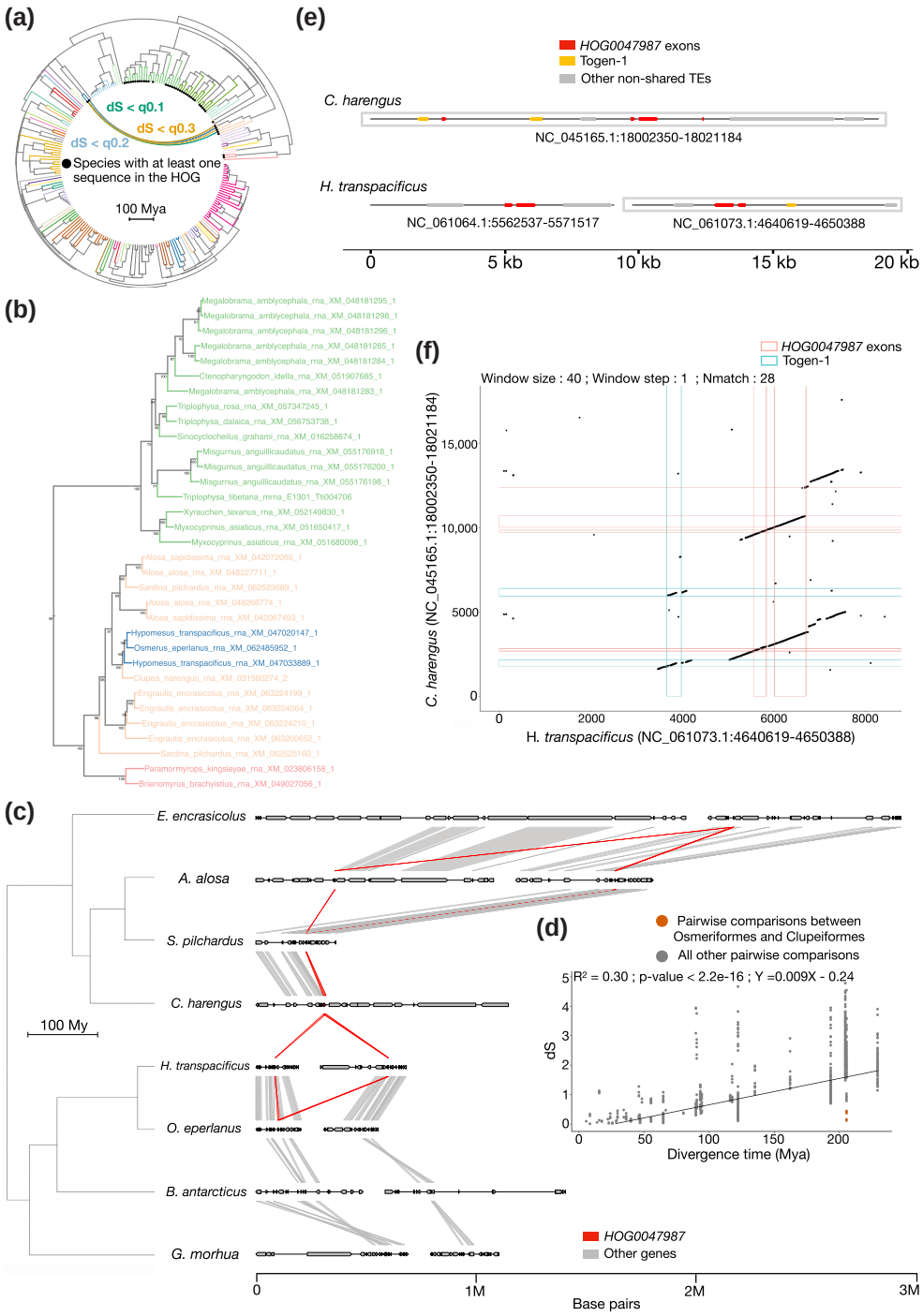


**Fig. 3.** Multiple lines of evidence supporting HGT in the orthogroup *HOG0047987*. a) Teleost species tree. Lines inside the tree connect two species if they share at least one gene which is less divergent than expected under vertical transmission, i.e. its dS falls below quantile thresholds of the dS distribution computed on all genes that could be aligned between the two species. b) Maximum likelihood phylogeny of the *HOG0047987* orthogroup. The tree was pruned to retain the clade with the transfer, as well as close branches. The complete tree can be found in supplementary file S1, Supplementary Material online. Branches are colored according to the species order. Bootstrap values are shown at each node. Here, the tree topology indicates a transfer from Clupea spp. to the ancestor of Osmeriformes, followed by a duplication, and a subsequent loss of one duplicate in *O. eperlanus*. c) Micro-synteny plot, showing the region of the transferred genes in the donor and recipient clades (Clupeiformes and Osmeriformes, respectively), as well as in closely related species. Red arrows represent the transferred genes, while gray arrows represent neighboring genes. Genes are linked between species if they belong to the same HOG. d) Relationship between dS values, computed between all *HOG0047987* pairs of genes, and divergence times between corresponding species. The linear regression is represented by a line, and its $R^2$, *P*-value, and equation are indicated. Orange dots correspond to pairwise comparisons between *C. harengus* and Osmeriformes genes. e) Zoom on the regions harboring the transferred gene in *C. harengus* and Hypomesus transpacificus (Osmeriformes). Transposable elements (TEs) found in only one of these regions are indicated by gray arrows, while those found in both species' regions are indicated by yellow arrows. Here, one TE, Togen-1, was found in common. Red arrows correspond to the transferred gene exons. Gray boxes indicate regions aligned in f. f) Dot-matrix plot between *C. harengus* and *H. transpacificus* regions. Black lines are indicative of conserved regions. Regions between red lines are *HOG0047987* exons, while the regions between blue lines correspond to Togen-1. Introns and flanking regions show strong conservation, and, when searched on all teleost genomes, only matched to Osmeriformes and Clupeiformes species. Such multiple lines of evidence, as well as additional analysis (e.g. correction of gene annotations, gene re-annotations, and inspection of alternative genome assemblies), are provided for all 19 HGT events discussed in this study, in supplementary file S1, Supplementary Material online.

**Table 1** Summary of HGTs between teleosts

| N5 HOG | Name (HHpred probability) | Method | dS | CDS identity [aln length (bp)] | Donor | Recipient | Additional evidence | Probability null microsynteny | Number of loss (alternative hypothesis) |
|---|---|---|---|---|---|---|---|---|---|
| *HOG0047987* | GCN1 (57.28%) | dS + A + B + C | q0.1 | 88.5% [972] | *Clupea spp.* | Osmeriformes | I + F + TE | 19.4% | 6 |
| *HOG0001647* | C-type lectin [AFP] (99.79%) | dS + A + B | q0.1 | 91.6% [438] | *Clupea spp.* | Osmeriformes | I + F + TE | 19.4% | 6 |
| *HOG0013514* | nuclear GTPase SLIP-GC (99.96%) | dS + A + B + C | q0.1 | 92.1% [1425] | *Clupea spp.* | Osmeriformes | I + F + TE | 19.4% | 6 |
| *HOG0018901* | Fish-egg lectin (99.86%) | dS + A + B | q0.3 | 91.7% [675] | *Clupea spp.* | Osmeriformes | I + F + TE | 19.4% | 6 |
| *HOG0008885* | Tax1-binding protein 1 (56.97%) | dS + A + B | q0.2 | 81.6% [1584] | *Clupea spp.* | Osmeriformes | F + TE | 19.4% | 6 |
| *HOG0004480* | Stonustoxin subunit alpha (100%) | dS | q0.1 | 95.2% [3804] | *Clupea spp.* | Osmeriformes | F | 19.4% | 6 |
| *HOG0028899* | De novo design protein (7.07%) | dS + B | q0.5 | 72.6% [336] | Clupeiformes | Osmeriformes | I + F | 19.4% | 5 |
| *HOG0004450* | Stonustoxin subunit alpha (100%) | dS | q0.1 | 82.1% [1851] | Clupeiformes | Carangaria | F | 17.4% | 17 |
| *HOG0029633* | Legumain (100%) | dS + B | q0.1 | 91.1% [585] | Siluriformes | *Paramormyrops* spp. | I + F | 18.9% | 9 |
| *HOG0046344* | MifM-stalling (18.43%) | None | q0.6 | 59.4% [591] | Siluriformes | *Paramormyrops* spp. | I + F | 18.9% | 9 |
| *HOG0010622* | Sterile alpha Motif domain containing 9 (99.97%) | dS + A + B +C | q0.1 | 84.8% [4461] | Siluriformes | *Paramormyrops* spp. | None | 18.9% | 9 |
| *HOG0013500* | Natterin (99.94%) | dS | q0.2 | 74.6% [1206] | *Clupea spp.* | Batrachoididae | None | 19% | 12 |
| *HOG0019032* | Cytosolic phospholipase A2 delta (100%) | dS + A + B + C | q0.1 | 84.7% [1272] | *Chanos spp.* | *Neoarius* spp. | I + F + TE | 12% | 8 |
| *HOG0004451* | Stonustoxin subunit alpha (100%) | dS + C | q0.1 | 81.2% [1413] | *Megalops spp.* | Berycidae | F | 10% | 11 |
| *HOG0030200* | N-Acetylgalactosaminyltransferase (99.76%) | dS + A + B + C | q0.1 | 93.8% [1446] | Cypriniformes | Pangasiidae | I + F | 13.7% | 8 |
| *HOG0034670* | Pesticidal crystal protein Cry6Aa (99.46%) | dS + B + C | q0.1 | 88.6% [1206] | Cypriniformes | Pangasiidae | I + F + TE | 13.7% | 8 |
| *HOG0019111* | Goose-type lysozyme (99.24%) | dS + A + B + C | q0.1 | 78.3% [456] | Scombridae | Clupeiformes | I + F | 21% | 16 |
| *HOG0019111* | Goose-type lysozyme (99.24%) | dS + A + B + C | q0.1 | 78.3% [456] | Scombridae | Clupeiformes | I + F | 21% | 16 |
| *HOG0019111* | Goose-type lysozyme (99.24%) | dS + A + B + C | q0.2 | 90.2% [471] | *Clupea spp.* | Osmeriformes | I + F | 21% | 16 |

Gene names are attributed based on the best HHpred match, along with the associated probability. For each hierarchical orthogroup (HOG), one representative sequence of the transfer was used as input in HHpred (see Materials and Methods). "Method" indicates with which method(s) the HGT was detected. "A", "B," and "C" correspond to the three BLASTN filtering strategies. dS: minimum dS quantile threshold below which pairwise dS values between donor and recipient clades were found. CDS identity: maximum nucleotide identity of the transferred gene between the donor and the recipient clades. The length (in base pairs, bp) of the alignment (denoted "aln") used to compute the coding sequence (CDS) identity, as well as the dS, is indicated between squared brackets. Additional evidence for a horizontal transfer, other than based on the coding sequence, is indicated. I = nucleotide-level similarity in intron(s); F = nucleotide-level similarity in flanking region(s). TE = presence of highly similar transposable elements in the gene flanking regions. The probability of null microsynteny is given for any gene between the donor and recipient species. It corresponds to the proportion of gene pairs with a micro-syntenic score of 0 between representative species of the donor and recipient clades. For example, there are 19.4% chances that one gene from the same HOG drawn from Clupea harengus and Osmeriformes has a null micro-syntenic score, i.e. with a complete lack of microsynteny. "Number of losses" indicates the minimum number of losses required in the species tree in the alternative scenario, assuming vertical transmission. *HOG0019111* is present three times, as two transfers were inferred from Scombriformes to Clupeiformes, in addition to one transfer from Clupea spp. to Osmeriformes.

found one TE (*crack-7*, *togen-1*, and *L1-19* respectively), present in many copies across the whole genome of *C. harengus*, but present as single copies and only next to the transferred genes in Osmeriformes.

We next estimated that scenarios alternative to horizontal transfer—that would involve vertical transmission of the 19 HGTs followed by repeated losses—would have required between five and sixteen independent losses across the teleost phylogeny, depending on the gene. These numbers were computed assuming that the gene appeared in the MRCA of the donor and recipient species and corresponded to the minimum number of losses needed to retrieve these species in a monophyletic gene clade, assuming zero duplication events, and without considering the rest of the HOG phylogeny.

Finally, we investigated the coherence of these HGTs in terms of species habitats. We found that HGTs always occurred between freshwater species or between marine species, but never between freshwater and marine species. Furthermore, there were always overlaps, of various sizes, in the distributions of donors and recipients (supplementary fig. S9, Supplementary Material online). One limitation to reliably pointing out the location where these transfers occurred is that these distributions correspond to extant species, and they may vary from those of ancestor lineages in which the HGTs occurred.

## Function of Putative Transferred Genes

We then investigated the functions of the 17 transferred genes, through inspection of their RefSeq names, by searching for known homologous sequences using BLASTP and HHpred (Table 1, supplementary table S2, Supplementary Material online), and by structure similarity searches (supplementary fig. S10, Supplementary Material online). We also assigned GO terms to each gene according to their best match in the UniProt database. We found that HGT genes were enriched (False discovery rate [FDR] < 0.05) for the GO terms "killing of cells of another organism" (GO:0031640), "extracellular region" (GO:0005576), and "interspecies interaction between organisms" (GO:0044419) (supplementary File S2, Supplementary Material online). These terms were assigned to three stonustoxin subunits found among transferred genes (*HOG0004480, HOG0004450, HOG0004451)*, which are pore-forming proteins (PFPs). Interestingly, two other genes were also PFPs, including one natterin (*HOG0013500*, Table 1, Fig. 4), and one Cry6Aa homolog (*HOG0034670*), the latter having been mainly studied in *Bacillus thuringiensis* (Dementiev et al. 2016 ). While stonustoxins and natterins have initially been described and studied in toxic species (Ghadessy et al. 1996; Magalhães et al. 2006), there is growing evidence for their role in defense against pathogens in fishes, probably preceding their use in venoms (Ellisdon et al. 2015 ; Lima et al. 2021 ; Seni-Silva et al. 2022 ), which is also the case for many other PFPs (Galinier et al. 2013; Xiang et al. 2014; Peraro and van der Goot 2016; Verma et al. 2021 ). Interestingly, some natterins have already been proposed to be horizontally transferred between two distant clades of eukaryotes, from fungi to corals (Gacesa et al. 2020 ). In addition, these GO terms were also associated with *HOG0019111*, homologous to the antibacterial enzyme goose-type lysozyme (LyzG) (Tullio et al. 2015; Liu et al. 2022), and to *HOG0010622*. This gene was homologous to SAMD9 (sterile alpha motif domain containing 9), playing roles in tumor suppressions and the response to viral infections in humans, and probably also in the zebrafish where it is an interferon-stimulated gene (Levraud et al. 2019; Nounamo et al. 2017; Ma et al. 2014b; Peng et al. 2022).

We found three additional genes with a main function related to vertebrate immunity: a fish-egg lectin (*FEL, HOG0018901)* which, in several teleost species, is mainly expressed in eggs and larvae, binds bacteria and promotes their phagocytosis by macrophages (Wang et al. 2016; Zhang et al. 2020b; Qiao et al. 2022); a nuclear GTPase SLIP-GC (*NUGGC, HOG0013514)* which enhances genome stability of B cells in mice and humans (Richter et al. 2009; 2012), but which is lacking any functional evidences in fishes; *HOG0008885* which was structurally similar to IL-40 (interleukin 40; probability = 0.66 ; e-value = 0.02), involved in B-cell development and immunoglobulin production (Catalan-Dibene et al. 2018; Dabbagh-Gorjani 2024). While no GO term was assigned by our method (best UniProt match) to *HOG0008885* and *HOG0018901*, *HOG0013514* was associated with the GO terms "somatic hypermutation of immunoglobulin genes" (GO:0016446) and "immune system process" (GO:0002376). The gene *HOG0029633,* a protease legumain, was also associated with "immune system process", with roles in human and zebrafish immunity (Manoury et al. 1998; Dall and Brandstetter 2016), but also in several other biological processes such as axon regeneration (Ma et al. 2014a) and bone mineralization (Jafari et al. 2017). The other transferred genes were involved in various biological processes (Table 1, supplementary table S2, Supplementary Material online, supplementary fig. S10, Supplementary Material online) such as the type II antifreeze protein (*HOG0001647*), allowing fish species to survive in icy seawater (DeVries 1971; Graham and Davies 2021).

Finally, to investigate whether horizontally transferred genes tend to undergo shifts in selective pressure after integration into the recipient genome, we computed a dN/dS ratio for every branches of HOGs containing HGTs and used branch-site models to assess the evolutionary regime of transferred genes (Fig. 4). First, when looking at dN/dS ratio distributions, we observed that transferred sequences did not appear to have particularly low or high ratios compared to other sequences from the same HOG. For most branches tested (28/41), no sign of diversifying or relaxed selection was found. For nine HOGs, at least one branch of the recipient clade was evolving under diversifying selection. Furthermore, across nine HOGs constituted of only one recipient branch, only four were evolving under diversifying selection. One gene (SAMD9; *HOG0010622*) was found to evolve under relaxed selection upon reception in the genome of *Paramormyrops kingsleyae (K* = 0.42 ; *P*-value = 3.5e-5*)*. Thus, these results indicate that there is, in some cases, an accelerated evolution of the gene upon reception, but this is far from being systematic.

## Pseudoparalogs and Post-transfer Duplications

While, among the 17 transferred genes, 14 were completely new to the recipient species, three (*HOG0019111*, *HOG0028899,* and *HOG0008885*) likely represented pseudoparalogs (Makarova et al. 2005), as the recipient species already possessed at least one other non-transferred gene in the same HOG. Also, in accordance with what has been described in other eukaryotes such as nematodes (Paganini et al. 2012) and algae (Schönknecht, Weber, and Lercher 2014), we inferred that half (9/19) of HGT events were followed by lineage-specific gene duplications in the recipient clade (supplementary File S1, Supplementary Material online).

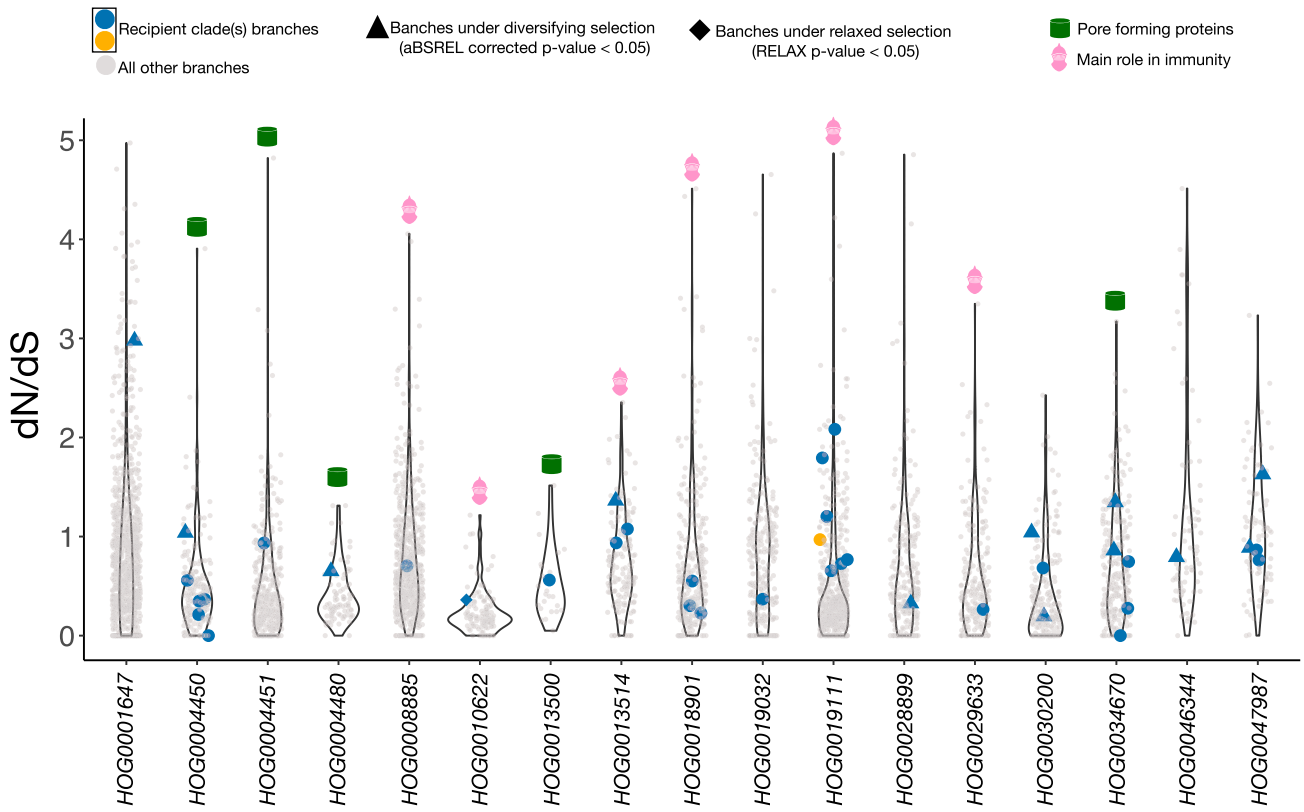


**Fig. 4.** Putative functional category and selective pressures acting on transferred genes. For each transferred HOG, dN/dS of every extant and ancestral sequences (terminal and internal branches, respectively) were computed with HyPhy. dN/dS values computed on branches corresponding to the recipient clades are shown with colored symbols. For *HOG0019111*, where three transfers occurred, dN/dS values on Clupeiformes branches are indicated in blue, while the unique Osmeriformes branch is shown in orange. Triangles represent sequences evolving under diversifying selection, while diamonds represent sequences evolving under relaxed selection. Symbols indicate genes which have a main function related to immunity or an immunity-related GO term, as well as pore-forming proteins. Note that for natterins (PFPs), immune functions have been suggested previously (Jia et al. 2016; Seni-Silva et al. 2022). Branches with a saturation of substitutions (dS or dN > 1) or not enough substitutions (dS < 0.01) were discarded.

This concerns three HGT events between Clupeiformes and Osmeriformes: (i) *HOG0047987* with one duplication in the MRCA of Osmeriformes (Fig. 3), (ii) *HOG0013514* with one duplication specific to *H. transpacificus* and two duplications specific to *Osmerus eperlanus*, and (iii) *HOG0018901* with three duplications specific to *H. transpacificus*. The *HOG0004450* transferred gene was also duplicated several times after the transfer, once in *Lates calcarifer* and two times in *Caranx melampygus*. In addition, there was one post-transfer duplication of *HOG0013500* in *Thalassophryne amazonica*, one of *HOG0030200* in *Pangasius djambal*, two of *HOG0019032* in *Neoarius graeffei*, two of *HOG0019111* in *Clupea harengus*, and three of *HOG0034670* in the MRCA of *Pangasianodon* species. While we here refer to specific species for which annotated genomes are available, some of these duplications may have occurred in common ancestors of these and other species not included in this study. These duplication events might have contributed to expanding the newly acquired beneficial functional gain (Schönknecht et al. 2014).

## Discussion

### Evidence Supporting Multiple Fish-to-fish HGTs

While increasing numbers of HGTs involving eukaryotes are being reported, few studies have investigated eukaryote-to-eukaryote gene transfers in general, and even fewer studies dealt with HGT between vertebrates. Here, we surveyed 242 ray-finned fish species and uncovered 19 fish-to-fish HGT events supported by multiple lines of evidence and involving 17 different genes and 11 fish lineages. Among the transferred genes is the AFP protein previously shown to have undergone an HGT event between Clupeiformes and Osmeriformes (Graham et al. 2008; Graham and Davies 2021). All transferred genes were found in multiple independent genomes from the donor and recipient fish clades, suggesting their presence in distantly related fishes is unlikely due to contamination. Phylogenetic analyses of these genes grouped distantly related fish species in strongly supported monophyletic groups, there was no conserved synteny around transferred genes between the donor and recipient lineages, and most transfers were also recovered using BLAST-based alien index approaches (Koutsovoulos et al. 2022; Yuan et al. 2023).

Manual curation revealed that in most instances, the high nucleotide identity in coding sequences between donor and recipient lineages extended to intronic and/or flanking sequences, which sometimes even included shared TEs. It is unlikely that these non-coding regions are CNEs evolving under purifying selection because they were not recovered in any other ray-finned fishes. We even found two cases where, as reported earlier for the AFP transfer (Graham and Davies 2021), the shared TE present in the gene flanking regions was present as a single copy in one species and in multiple copies in a second species. This pattern is consistent with the gene tree topology in suggesting that the second species is the

donor, in which the TE generated multiple copies of itself, while the first species is the recipient, in which the TE has not transposed from the horizontally acquired region. Finally, we took special care in examining scenarios alternative to HGT that would involve multiple gene losses. First, we found that between at least 5 and 17 independent such losses would be required, depending on the gene, to explain the topology of their phylogeny. These numbers are to be contrasted with the fact that genes showing high similarity between species at the sequence level (both in the coding sequence, introns, and flanking regions) are rather expected to show the lowest loss rates (Waterhouse et al. 2010). In fact, we show that the 17 transferred genes uncovered in this study are characterized by extremely low dS values, not only when compared to the distributions of dS values computed on all the genes between the species involved, but also when compared to dS values computed between all the genes among these HOGs (Fig. 3, supplementary fig. S1, Supplementary Material online). Thus, the lower-than-expected divergence of these genes is inconsistent with them being present in the common ancestors of fishes and enduring recurrent losses during fish evolution (Cote-L'Heureux et al. 2022).

## Mechanisms of HGT Between Fishes

The number of fish-to-fish HGT reported here is low (with only 0.02% of ray-finned fishes HOGs containing at least one HGT event), but finding that such transfers have occurred multiple times is interesting, begging among other questions that of the underlying mechanisms. It has previously been proposed that the mode of reproduction used by most fish species, namely external fertilization, may render their germline more susceptible to being colonized by foreign DNA (Huang 2013; Graham and Davies 2021). In particular, sperm cells have the ability to bind and internalize exogenous DNA and to act as vectors of this DNA in the oocyte during fertilization, a process called sperm-mediated gene transfer (SMGT) that has been used in the laboratory to generate multiple genetically modified animals (Lavitrano et al. 2013). Should SMGT occur under natural conditions, which is currently unknown, its frequency might be higher in species using external fertilization, which are expected to be exposed to more environmental DNA than species using internal fertilization (Smith and Spadafora 2005; Graham and Davies 2021). All species found involved in HGT in this study use external fertilization. However, at least 12 independent transitions from external to internal fertilization have occurred in fish (Benun Sutton and Wilson 2019), so that as more fish genomes become available, it will be possible to test whether transfers occur more often in external versus internal fertilizers. To increase the power of such an analysis, horizontal transfers of TEs, which seemingly occurred more frequently than HGT among fishes, should be included (Zhang et al. 2020). Beside sperm cells, another type of HGT vector regularly put forward in the literature are viruses (Gilbert and Cordaux 2017). Much like other animals, fishes are infected by a large diversity of viruses (Costa et al. 2024; Costa and Holmes 2024). These include large dsDNA viruses of the Iridoviridae family (Leiva-Rebollo et al. 2024), which are considered good candidate vectors of horizontal transfer because they can capture host genes (Filée et al. 2008; Yoxsimer et al. 2024) and TEs (Piégu et al. 2014; Loiseau et al. 2021), and they are able to infect species that diverged hundreds of millions of years ago (MYA) (Chuang et al. 2022). Other potentially good candidate HGT vectors are the Teratorn endogenous viral elements (EVEs), which result from the fusion of Alloherpesviridae genomes and piggyBac DNA transposons (Inoue and Takeda 2023). These elements are widespread in teleost fishes, are able to transpose and produce multiple copies of themselves in a given genome, and their phylogeny suggests they undergo recurrent cross-species transmission (Inoue and Takeda 2023). In any case, the HGT events inferred here suggest that the species involved in these transfers are or have been in direct or indirect contact (e.g. through shared viruses or other parasites), which is consistent with the fact that these species currently live in similar environments (marine or fresh-waters) and overlap, to some extent, in their geographical distribution (supplementary fig. S6, Supplementary Material online).

## Horizontal Transfer or Long-distance Hybridization?

Given that the mechanisms underlying HGT in most eukaryotes remain unclear, it is tempting to invoke hybridization as an alternative scenario to explain the distribution of genes that show a signal of interspecies gene flow among teleosts (Gabaldón 2020). Such a scenario could appear more parsimonious than HGT to account for the eight genes transferred from Clupeiformes to Osmeriformes, as one hybridization event instead of eight HGTs could be invoked.

Assuming that meiosis was possible in the resulting interspecific hybrids, crossing-overs between homeologous chromosomes could have led to the transfer of one or a few genomic regions containing the eight genes from one parental genome to the other. Considering one or a few crossing-overs per chromosome per meiosis, a single round of meiotic recombination would have led to megabase-sized exchanges between homeologous chromosomes. However, the horizontally transferred DNA sequences detected here are relatively short (2.2 to 15 kb). Such short introgressed segments could have resulted from repeated backcrosses in one parental lineage. The circumstances under which such a scenario would have generated sharp (single-gene) introgressed segments, as those we observe here, are unclear, given the high genetic distance between the parental chromosomes. While multiple cases of long-distance hybridization have been reported in fish species diverging by dozens of millions of years (Fig. 5) (Zhang et al. 2014), proper meiotic pairing and recombination have not been observed between chromosomes originating from species as divergent as those we report here to be involved in HGT (>124 MYA, Fig. 5). Would sufficient homology exist between divergent species, it is more likely to occur within gene bodies, which are more conserved than intergenic regions and introns (Zhang et al. 2020b). However, none of the transferred segments occur at syntenic positions in the donor and recipient fish species, which makes gene bodies unlikely to have provided the template for recombination. Nevertheless, one can envision a scenario in which transposition of TE copies from one parental genome to the other could distribute homologous segments between homeologous regions. Such TEs would provide the templates for ectopic recombination events bordering the observed exchanged segments. In this context, the TEs shown here to be shared between some donor and receiving HGT lineages may be viewed as relicts of such recombination triggers.

When parental species diverge by evolutionary distances as large as those inferred here to be involved in HGT, one would expect that hybridization produces allopolyploids (Deb et al.

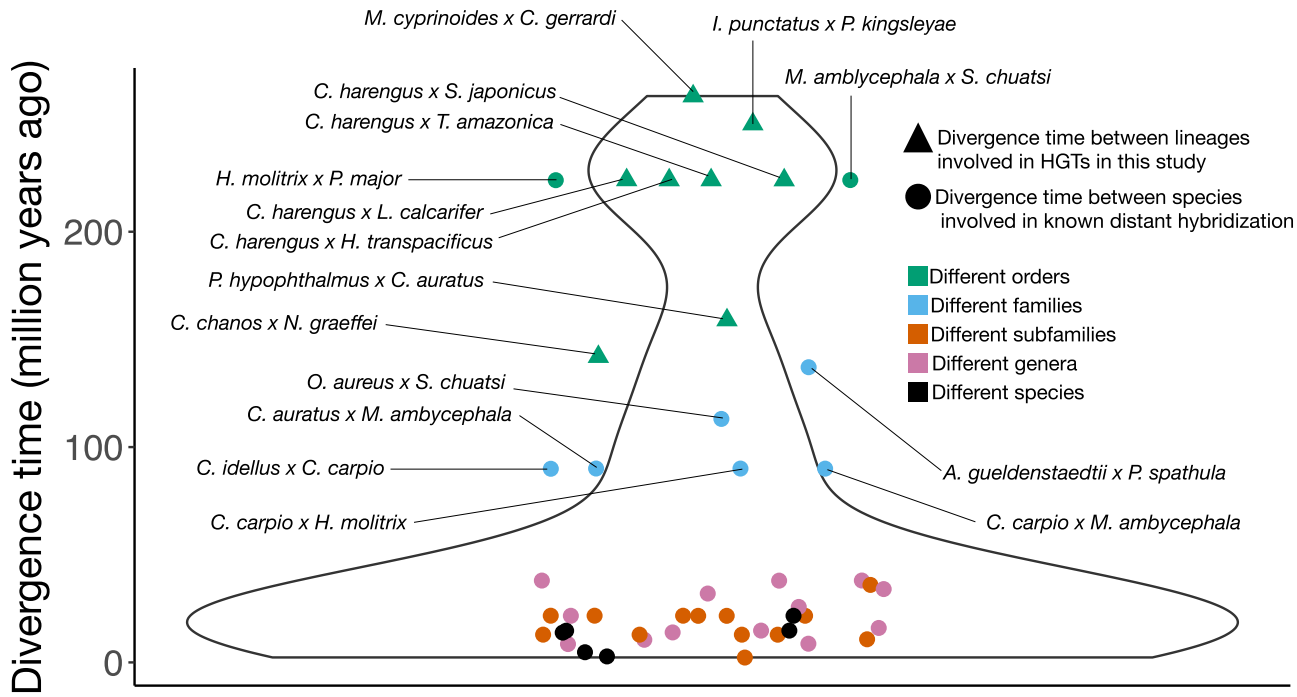


**Fig. 5.** Divergence between species involved in HGTs or in known distant hybridization. Violin plot showing the divergence times between (i) reported case of distant hybridization in the literature (circles [Chen et al. 2018; Zhang et al. 2014; Liu et al. 2020; 2021; Káldy et al. 2020; Wang et al. 2017; Wang et al. 2022]) or (ii) species with genes showing much higher similarity than expected in this study (triangles), and for which our main hypothesis is a horizontal transfer. Colors represent the taxonomic difference between the two species. For cases involving species from two different orders or families, the two species names are indicated.

2023). An illustration of this is the striking example of triploid and pentaploid hybrids resulting from crosses of the Russian sturgeon (*Acipenser gueldenstaedtii*) and the American Paddlefish (*Polyodon spathula*), which diverged 137 MYA (Fig. 5) (Káldy et al. 2020). However, hybrids resulting from such long-distance hybridization in fish are generally sterile. Their frequency in nature remains unknown, and their genomic architecture has not been investigated using whole genome sequencing. Furthermore, none of the HGT-receiving lineages in our study contain known polyploids (Comber and Smith 2004), and consistently, the numbers of annotated proteins are highly homogenous among the species involved in HGT (mean = 24,638; 21,690 to 29,714). In addition, the resulting mosaicism would have been extremely unbalanced to retain only a few short fragments from one of the parental DNA. However, one could invoke a less canonical event involving a transient polyploid bridge as proposed in (Pereira et al. 2023)(see also (Bartolić et al. 2024)).

Overall, we cannot exclude the possibility that some unconventional hybridization scenario would have led to a pattern that is difficult, if not impossible, to distinguish from vector-mediated HGT. Nonetheless, we tend to favor the HGT scenario over hybridization for two reasons. First, there is strong evidence that horizontal transfer of genes and TEs has occurred between eukaryotes (including between animals) (Moran and Jarvik 2010; Gasmi et al. 2015; Szöllősi et al. 2015; Gilbert and Feschotte 2018; Hibdige et al. 2021; Gilbert and Maumus 2022; Heisserer et al. 2023; Mishina et al. 2023; Widen et al. 2023), but so far, no study has shown that hybridization can yield HGT-like patterns such as those we report here in fish. Second, the vast majority of known long-distance hybridization cases in fish involve lineages separated by shorter divergence times than those among which we have inferred HGT (Fig. 5). Two cases of successful hybridization have been reported that involve lineages separated by divergence times comparable to those separating the lineages involved in HGT here (both about 224 MYA). The first one is a cross between *Hypophthalmichthys molitrix* and *Pagrosomus major*. It is listed in a review paper (Zhang et al. 2014), but the original article is not published in a peer-reviewed journal (Zhang et al. 2014). The second one involves crosses between *Megalobrama amblycephala* and *Siniperca chuatsi* (Wang et al. 2022). The morphology and karyotype of the resulting hybrid is identical to that of the maternal species and show no sign of paternal traits, the 5S rDNA gene is also identical to the maternal species and the only investigated microsatellite locus produces an unclear pattern (Wang et al. 2022). Furthermore, the study does not rule out an alternative scenario involving gynogenesis, a process of asexual reproduction known to occur naturally in fish, which requires the presence of sperm, without the DNA of sperm cells actually contributing to the egg cells (Schlupp 2005).

## Opportunity for Transfer, Retention Rate, and Functional Impact of HGT in Teleosts

According to Keeling (2024), opportunities for HGT may not be that rare in eukaryotes but it is rather the very low likelihood for a horizontally transferred gene to be retained in the receiving host that may largely explain the lower HGT rates observed in eukaryotes compared to prokaryotes. Such a low retention rate in eukaryotes could be due to the lower probability for a gene acquired through HGT to be beneficial in eukaryotes compared to prokaryotes. The number of HGTs that we can observe today in eukaryotes would thus correspond to a small fraction of a wider foreign gene pool, the majority of which would bring no benefit to the receiving host

and be lost through drift or purifying selection. In apparent agreement with this prediction, the number of teleost-to-teleost transfers of TEs (910 among 64 species; [Zhang et al. 2020a]) is much larger than that of genes (19 among 247 species; this study), indicating that the number of opportunities for transfer might be much higher than what the number of HGTs uncovered in this study suggests. While the capacity of TEs to transpose might, to some extent, result in a higher number of transfer opportunities for these elements compared to genes, there is no doubt that the capacity of TEs to generate multiple copies of themselves strongly increases their retention rate once transferred. In any case, following Keeling (2024)'s argument, horizontally acquired genes are either the result of a recent transfer and did not have time yet to decay (which could be the case of SAMD9, evolving under relaxed selective constraints in the unique recipient species, relative to other fishes), or they were retained because of the strong selective advantage they conferred to the recipient species. Here, we found that several transferred genes were likely involved in fishes' immunity. Indeed, we observed the transfer of six genes with established roles in the vertebrate immune system, as well as five PFPs, that might also be involved in the innate immunity of fishes (Fig. 4). This pattern suggests that immune and PFP genes may be preferentially retained after HGT, and thus more often beneficial to the host than other genes, enhancing and/or diversifying the species' defense repertoire. This is in line with many recent studies showing that, across vertebrates, genes under positive selection are often enriched for immune functions, highlighting their rapid evolution and impact on species' fitness (Limborg et al. 2012; Xiao et al. 2015; Vinkler et al. 2023; Rincon-Sandoval et al. 2024). This also fits well with previous HGT studies in eukaryotes, in which transferred genes were often found to be involved in the immune system (Ciach et al. 2024; Keeling 2024; Bai et al. 2016; Li et al. 2022; Tarnopol et al. 2025), but also in host–parasite interaction (Yang et al. 2016; Gasmi et al. 2021; 2015; Xia et al. 2021; Gilbert and Maumus 2023; Keeling 2024; Nowell et al. 2024).

### Uncertainties, Limitations, and Perspectives

Our results are consistent with previous studies of HGT in eukaryotes, which all suggest that the number of horizontally acquired genes is many folds lower than in prokaryotes (Danchin 2016; Martin 2017; Van Etten and Bhattacharya 2020; Keeling 2024). Yet, as a first step in evaluating HGT among relatively closely related eukaryote lineages (i.e. among vertebrates), we used stringent criteria to filter very large gene datasets and favored a manual curation to retain HGT cases that appeared supported by multiple lines of evidence. On one hand, we acknowledge that there is a possibility that rather than resulting from HGT, some genes reported here might result from unknown, biologically improbable events, or rare artifacts. On the other hand, it is also possible that the 19 HGTs we inferred only represent a fraction of the total number of gene transfers that have occurred among the fish species included in our dataset. Furthermore, we have not investigated HGT between fish and non-fish species. Our study included 233 out of the >30,000 known teleost species (Near et al. 2012), and the inclusion of many more genomes will allow us to more precisely constrain the timing of HGT events and the development of more powerful, unsupervised automatic methods to detect HGT. One area of improvement would be to automate, at least to some extent, the filtering of candidate HGT trees in order to assist the selection of solid HGT cases, which is so far done manually (this study, [Li et al. 2022]). This would allow screening of many more candidate HGT trees obtained by releasing the stringency of upstream filters, such as the one imposing strict absence of synteny between donor and receiving lineages. Such an approach might help assess whether, in some cases, larger genomic regions containing several genes have been transferred horizontally, either through bona fide HGT or via long-distance hybridization. The extent to which numbers of HGT among fish might inflate as more genomic data are available is difficult to predict, as the distribution of these transfers appears highly heterogeneous in the fish phylogeny. A striking pattern is that among the 58 teleost orders included in our study, only 11 were found to be involved in HGT and 8 of the 19 HGT cases occurred between two of these orders (Clupeiformes and Osmeriformes), with no evidence from the genomic location of transferred genes in favor of a single co-transfer (supplementary fig. S5, Supplementary Material online). To some extent, the unbalanced distribution of HGT cases among fishes is reminiscent of that observed in insects, in which horizontally transferred plant genes are only found in 14 out of 218 surveyed species, with the whitefly, *Bemisia tabaci* concentrating most of these genes (135 out of 156) (Gilbert and Maumus 2023). We have no good explanation for these patterns and can only speculate that shedding light on the mechanisms involved in HGT may help resolve them. Most relevant to HGT in fish, it will be interesting to evaluate the genomic consequences of long-distance hybridization in this group using whole genome sequencing to assess whether this process can produce HGT-like patterns like those observed here.

## Materials and Methods

### Genome Data and Annotations

About 230 annotated (annotations from RefSeq or GenBank) and 1,200 non-annotated ray-finned fishes' genomes were downloaded using genome_updater (https://github.com/pirovc/genome_updater) with the option -T "7898" (taxonomic identifier of ray-finned fishes), as of March 9, 2024. We used BUSCO v5.1.274 (Manni et al. 2021) to assess the completeness of each genome (supplementary File S2, Supplementary Material online), using the Actinopterygii odb10 gene database; 842 genomes with a BUSCO score ≥ 80% were retained, among which 225 had previously been annotated. The species order of these 842 genomes was extracted using Entrez (Maglott et al. 2007) and the NCBI taxonomy database (Schoch et al. 2020; supplementary File S2, Supplementary Material online). For each retained annotated genome, we used agat (Dainat 2022) to keep only the longest transcript per gene and to extract the corresponding coding sequence. Coding sequences with at least one internal stop codon were discarded. All coding sequences from these 225 species were combined and translated into proteins using EMBOSS transeq (Rice et al. 2000) to build a ray-finned fishes' protein database.

We identified 14 orders for which no annotated genomes were present, but which contained at least one non-annotated genome with a BUSCO score ≥ 80%. We annotated 17 representative species from these 14 orders as follows. First, for each of these species, a de novo repeat library was generated using RepeatModeler2 (Flynn et al. 2020). This library was combined with ray-finned fishes' repeats extracted from Dfam (Storer et al. 2021), and used as input to soft-mask

the corresponding genome assembly using RepeatMasker (Smit et al. 2013). In addition, we used the Sequence Read Archive (SRA) toolkit to download RNA-seq reads from twelve species with at least one RNA sequencing run present in the SRA database. Finally, these 17 soft-masked genome assemblies were annotated with BRAKER (Gabriel et al. 2024 ), using as input: (i) our ray-finned fishes' protein database, (ii) representative ray-finned fishes proteins extracted from OrthoDBv11 (Kuznetsov et al. 2022 ), (iii) RNA-seq reads when available (supplementary File S2, Supplementary Material online). Resulting gene models not supported by any hints were discarded using the "selectSupported Subsets.py" script provided in BRAKER. We further filtered BRAKER annotations by retaining only the longest transcript per gene and discarding genes with at least one internal stop codon. Thus, our final species dataset comprised 242 annotated ray-finned fishes, among which 233 were teleosts and nine were non-teleosts.

## Species Phylogeny

3,584 BUSCO proteins present in at least half of the annotated species (≥ 121) were aligned using MUSCLE v5.1 (Edgar 2022) and trimmed using trimAl (Capella-Gutiérrez et al. 2009) with the option "-automated1." A maximum likelihood phylogeny was reconstructed for each alignment using IQ-TREE2 (Minh et al. 2020), with the optimal model found by ModelFinder (Kalyaanamoorthy et al. 2017) and with 1000 ultrafast bootstraps to evaluate the robustness of the nodes (Hoang et al. 2018). Nodes with a low support (< 10% bootstrap support) were collapsed using the R package ape v5.0 (Paradis and Schliep 2019). The 3,584 unrooted gene trees were used as input in ASTRAL-III (Zhang et al. 2018) to compute an unrooted species tree. This tree was then rooted and dated using the least squares dating method (To et al. 2016) implemented in IQ-TREE2, with three calibration dates, on deep and strongly supported nodes (Near et al. 2012), retrieved on TimeTree.org (Kumar et al. 2022): (i) 250 million years for the split between *Danio rerio* and *Megalops atlanticus*; (ii) 396 million years between *Danio rerio* and *Polypterus senegalus*; and (iii) 224 million years between *Danio rerio* and *Takifugu bimaculatus*.

## Orthologous Groups Identification, Alignment and Pairwise Metrics

We used OrthoFinder (Emms and Kelly 2019) with (i) the proteomes of our 242 species (for a total of 6,263,181 ray-finned fishes' protein sequences), (ii) the rooted species tree, and (iii) non-teleost species as outgroups. HOG, as defined in this study, correspond to groups of genes which share a unique origin in the MRCA of teleost species (node 5 of the species tree, N5). For each HOG, protein sequences were aligned using MUSCLE v5.1, and the alignments were trimmed and back-translated to codon alignments using trimAl and the options "-automated1" and "-backtrans." We then computed four metrics, using the R (R Core Team 2018) packages "seqinr" (Charif and Lobry 2007), "bio3d" (Grant et al. 2006), and "MSA2dist" (Ullrich 2024), between every possible pairs of sequences among each orthogroup: (i) dS, (ii) dN, (iii) number of aligned nucleotides, and (iv) sequence identity. In addition, the same alignment procedure and metrics computation were used for BUSCO genes.

## Candidate Horizontal Gene Transfers Identification Using dS Values

For each possible teleost species pair (27,028 pairs), we computed a distribution of all dS values obtained between genes from these two species (only if the genes belonged to the same HOG), by first discarding (i) genes matching to any transposable element present in the RepeatPeps library of RepeatMasker or RepBase (Bao et al. 2015) (identified using BLASTP implemented in BLAST+ (Camacho et al. 2009) or Diamond (Buchfink et al. 2021)), (ii) genes for which less than 100 codons were aligned, and (iii) saturated dS values (value of 9.999). Using these distributions, we then computed the 0.5%, 0.4%, 0.3%, 0.2%, and 0.1% dS quantiles using the "stat" package in R (R Core Team 2018), as well as the mean and median dS values (supplementary fig. S11, Supplementary Material online). The same procedure was used to compute dS quantiles, mean, and median from BUSCO genes. 2,558,049 gene pairs with a dS value below q0.5 were kept, in a total of 6,367 distinct HOGs.

In addition, for each species pair, and for every gene belonging to the same HOG, we computed micro-syntenic scores, similar to the k-synteny index previously described by Adato et al. (2015). Considering the orthologous gene $G_i$ (defined by its presence in the same HOG) present in two species, $S_a$ and $S_b$, the micro-syntenic score represents the number of orthologous genes among the 20 flanking genes between $S_a$ and $S_b$ (ten genes upstream and ten genes downstream $G_i$). Thus, each pair of genes had a score between zero (if no common HOG were retrieved among flanking genes) and twenty (if all the flanking HOGs were in common). Microsynteny plots presented in this study were generated using gggenomes (Hackl et al. 2024).

We then removed gene pairs with a micro-syntenic score ≥ 1, as well as genes present on a scaffold containing only that gene. This led to a final dataset of 594,324 gene pairs belonging to 2,381 distinct orthogroups. Maximum likelihood phylogenies were computed for these 2,381 orthogroups, using IQ-TREE2, and the optimal model found by ModelFinder (using both non-trimmed and trimmed alignments). These trees were plotted using ggtree (Yu et al. 2017) and manually inspected to find topologies consistent with a horizontal gene transfer scenario. We also discarded HOGs for which multiple gene pairs were consistently showing low dS values and for which the phylogenetic tree was ladder-shaped, suggesting selection on synonymous sites and a lack of informative sites (supplementary fig. S12, Supplementary Material online, (Jaramillo et al. 2015)).

Finally, for each retained gene, we performed two BLASTP, using one protein sequence representative of the transfer as query, against (i) a database of all ray-finned fishes' proteins and (ii) the Uniprot database devoid of ray-finned fishes' proteins (UniProt Consortium 2021). The 50 and 10 best matches were retained, respectively. These protein sequences were aligned using MUSCLE v5, trimmed using trimAl, and a maximum likelihood phylogeny was computed with IQTREE2. This allowed us to ensure the robustness of the potential HGT topology, independently of prior OrthoFinder classifications, and computed from an independent alignment.

## Further Inspections of Putative Horizontal Gene Transfers

We further manually inspected putative horizontal gene transfers. First, we verified the quality of protein alignments using the "msaplot" function in ggtree. Then we ensured that among each orthogroup, where one transfer event was inferred, there

was a positive correlation between pairwise genes dS values and the divergence time between corresponding species. As expected, the dS values between transferred genes always appeared as outliers, i.e. showed low dS values compared to other pairwise comparisons with the same divergence time.

Furthermore, we selected one representative species from the two clades involved in the gene transfer. For these two species, the transferred gene(s) and flanking regions (5,000 bp upstream and 5,000 bp downstream) were extracted. These regions were aligned using dotter-plot matrices and the seqinr R package, allowing the identification of similarity at the nucleotide level in introns or in flanking regions. Any similar non-coding region identified on these dotter-plot matrices was extracted from the genomes using SAMtools (Danecek et al. 2021 ) and aligned using Needle (Rice et al. 2000). These regions were then searched in the genome assemblies of all teleost species using BLASTN. Transposable elements present in these regions were identified through a TBLASTN using a transposable element protein database (combination of RepeatMasker and RepBase databases) as a query. Redundancy of this TE database was first reduced by clustering similar proteins with cd-hit (Fu et al. 2012) and a similarity threshold of 80%. Finally, non-overlapping best-hit regions were extracted and used as queries in a BLASTX against the TE database, and these regions were assigned to a transposable element based on the best BLASTX match. TEs found in common between the donor and recipient species were used as queries in a TBLASTN against these species' genome assemblies, as well as against the genome assemblies of closely related species. Again, non-overlapping best-hit regions were extracted, and the best TE match against these regions was identified using BLASTX. Donor and recipient clades were assigned based on the HOG tree topology and were, in some cases, further supported by neighboring transposable elements, which were present in multiple copies around the donor genome while appeared only in one copy near the transferred gene in the recipient genome.

To rule out contamination as an alternative hypothesis to HGT, we ensured that transferred genes were present in at least two donor and two recipient species. If not, we ensured that these genes were present in another and independent genome assembly of these species. Finally, we systematically verified that the transferred genes were completely absent from species closely related to the donor and recipient clades, using a combination of TBLASTN and Exonerate (Slater and Birney 2005). In some cases, the same strategy was used to correct or complete partial annotations of genes. Detailed analysis results, for each HGT, can be found in supplementary file S1, Supplementary Material online.

Geographical distributions of species involved in HGT events were extracted from the OBIS and GBIF databases, accessed through the R packages "robis" (Provoost and Bosch 2022) and "rgbif" (Chamberlain and Boettiger 2017). Locations of transferred genes in the genome of *Clupea harengus* and *Hypomesus transpacificus* were represented using the R package "circlize" (Gu et al. 2014).

## Candidate Horizontal Gene Transfers Identification Using BLAST

We assessed if we could retrieve potential transfers identified using our dS procedure (see above) with another and independent method, relying on BLAST and which is commonly used to find horizontal transfers across the tree of life (supplementary fig. S13, Supplementary Material online; (Koutsovoulos et al. 2022; Dutton and Reiter 2023; Yuan et al. 2023)). First, each coding sequence of each teleost species was used as a query in a BLASTN against all coding sequences from other teleost species, and we discarded genes matching transposable elements. We then only retained matches if the BLAST alignment had a length $\geq 300$ nucleotides and $\geq 85\%$ of the query length.

For the first BLASTN filtering strategy (referred to as strategy A in this study), we retained 44,360 queries for which the best match (defined by the lowest *e*-value), did not belong to the same species order, which we only applied to orders with at least two species.

A second filtering strategy was applied (strategy B). For each species, only the best blast match (defined by the lowest e-value) per gene was retained. Then, for each species, we computed an inverse cumulative distribution of divergence times (which are pseudo-continuous) between this species and species corresponding to the best blast matches. We retained divergence time values for which $\leq 10\%$ of blast matches were found, allowing us to identify outliers, i.e. the best blast match for which the target species was more distant than expected from the query species.

Finally, for Strategy C, we adapted the HGTindex metric described in (Yuan et al. 2023), which was originally designed to detect HGT events between highly distant species, i.e. between life domains or between phyla. For each BLASTN query, we first retained the best match belonging to the same order, thereafter called $BM_{ingroup}$, as well as the best match belonging to another order, thereafter called $BM_{outgroup}$. Then, we computed the HGTindex per query, defined as the $BM_{outgroup}$ bitscore divided by the $BM_{ingroup}$ bitscore. For each species, we only retained queries with an HGTindex $\geq$ 0.85 and extracted the species corresponding to the $BM_{outgroup}$, thereafter called the outgroup species.

Then, for each query species, we computed the distributions of HGTindex values between this species and all retained outgroup species (one HGTindex distribution per outgroup species). Each of these HGTindex distributions was computed retaining the best matches belonging to the same order ($BM_{ingroups}$) and the best matches belonging to the corresponding outgroup species (i.e. constraining $BM_{outgroups}$ to belong to this species). As described above, HGT indexes were defined as the $BM_{outgroups}$ bitscores divided by the $BM_{ingroups}$ bitscores. Finally, genes were kept if meeting two criteria: (i) passed the initial HGTindex threshold ($\geq$ 0.85) and (ii) exceeded the 99.5th percentile of the HGTindex distribution for the corresponding outgroup species.

Again, for these three strategies, we discarded gene pairs (a query and its best match) with a micro-syntenic score $\geq 1$ as well as genes present on a scaffold containing only that gene. Maximum likelihood phylogenies of corresponding HOGs were computed with IQTREE2, as described above.

## Genes GC Content and Codon Usage Bias

We investigated whether we could use coding sequence compositions to infer transferred genes across teleost species, as in bacteria (Ravenhall et al. 2015; Dutton and Reiter 2023). For each teleost species, we computed the effective number of codons used per gene, while accounting for the nucleotide composition, thereafter called ENCprime (Novembre 2002). The ENCprime value per species was computed as the mean ENCprime across all genes, by first discarding coding

sequences smaller than 100 codons. If each amino acid is encoded by each of its corresponding codons at an equal frequency, ENCprime would have a value of 61. On the other hand, if each amino acid is encoded by a single codon, ENCprime would be equal to 20. It is generally assumed that ENCprime values above 35 indicate a weak codon usage preference, while ENCprime values below 35 indicate codon usage preferences (Gao et al. 2024; Zhang et al. 2024). We found that all teleost species have a very weak codon usage bias, with ENCprime values comprised between 52.4 and 58.3 (mean = 57 [supplementary fig. S14-b, Supplementary Material online]).

We then selected a set of seven representative teleost species across our species tree: *Clupea harengus, Paramormyrops kingsleyae, Megalops cyprinoides, Danio rerio, Hypomesus transpacificus, Pangasianodon hypophthalmus,* and *Oreochromis niloticus*. Five of these species were involved in transfers described in this study (detected with the dS and BLAST strategies), except the zebrafish, *D. rerio,* and the Nile tilapia, *O. niloticus*. For each species, we computed the GC3 content per gene using seqinr (GC content of third-codon position), by first discarding sequences smaller than 100 codons. We observed large overlaps in the GC content of genes from teleost species, preventing our ability to detect outlier genes in each genome (supplementary fig. S14-a, Supplementary Material online).

For each of these seven species, we also computed the relative synonymous codon usage (RSCU) of each codon (Sharp and Li 1986), using seqinr. This allowed us to build a codon usage bias profile per species, defined by 64 RSCU values. Codon usage bias profiles between species were compared using pairwise Pearson's correlations. We found that codon usage profiles, defined by RSCU values of all codons, were highly similar between species (supplementary fig. S14-c, Supplementary Material online), and thus not suited to find transferred genes between teleosts.

## Contamination Assessment Using Mitochondrial Genomes

The mitochondrial genome of *C. harengus, of H. transpacificus*, of nine control species (all with the same divergence time to *C. harengus* than *H. transpacificus*) and of the non-teleost fish *Lepisosteus oculatus* were retrieved from the RefSeq database (supplementary File S2, Supplementary Material online). We extracted the 13 mitochondrial genes from each of these mitochondrial genomes using RefSeq annotations. For *B. antarcticus*, as no RefSeq annotation was available for its mitochondrial genome, we manually annotated these 13 genes. Their locations were found through TBLASTN using the mitochondrial genes of the other species as queries, and their coding sequences were retrieved using EMBOSS getorf. The protein sequences of the 13 genes were aligned using MUSCLE v5.1, and these alignments were concatenated using AMAS (Borowiec 2016). A maximum likelihood phylogeny was computed from this concatenated protein alignment with IQ-TREE2, using empirical base frequencies, the mtVer matrix (Le, Dang, and Le 2017), and a discrete gamma model with four rate categories (mtVER + F + G4). The resulting phylogeny was rooted using *Lepisosteus oculatus* as outgroup.

The raw reads used to assemble the genome of *H. transpacificus* and of the ten control species were retrieved from the SRA database (supplementary File S2, Supplementary Material online). Short reads were cleaned using fastp (Chen et al. 2018b) and aligned, using bwa-mem2 (Vasimuddin et al. 2019) to a FASTA file containing the eleven mitochondrial genomes. The same procedure was followed for long reads, but using fastplong for cleaning and minimap2 for mapping (Li 2021). Resulting Sequence Alignment Map (SAM) files were sorted and converted to Binary Alignment Map (BAM) files using SAMtools *sort*, and the number of reads mapped to each mitochondrial genome was recorded using SAMtools *idxstats*.

## Mapping Long Reads From *H. transpacificus* and *C. harengus* to Validate Assemblies at Horizontally Transferred Gene Loci

Long reads data of *C. harengus*, *H. transpacificus,* and *B. antarcticus* were retrieved from the SRA database (supplementary File S2, Supplementary Material online). Reads were cleaned using fastplong and mapped to the genome assemblies of *C. harengus* and *H. transpacificus* using minimap2. We then used SAMtools *view* to extract the exact location of each mapped reads (scaffold, start and end of the alignment) and counted the number of long reads entirely spanning the horizontally transferred genes, including their species-specific duplicates (start of the alignment ≤ coding sequence start and end of the read alignment ≥ coding sequence stop, supplementary File S2, Supplementary Material online).

For each HGT, we extracted the longest read from *H. transpacificus* and *C. harengus* spanning the entire transferred gene. The genes' and exon locations on these two long reads were retrieved using EXONERATE, and the similarity between these two reads was assessed and visualized using a dotter plot.

## Genes Functions and Selection Analysis

For each transferred gene, we performed homology searches, through profile hidden Markov models, using HHpred (Söding 2005; Zimmermann et al. 2018) and three databases, PDB, Pfam, and SMART. One representative protein sequence of each transfer was used: *HOG0046344*, XM_023822671; *HOG0010622*, XM_023831520; *HOG0047987*, XM_047020147; *HOG0013514*, XM_047045767; *HOG0018901*, XM_047023826; *HOG0008885*, XM_047048755; *HOG0004480*, XM_062453720; *HOG0028899*, XM_047031800; *HOG0004450*, XM_042704194; *HOG0029633*, XM_046853148; *HOG0013500*, XM_034166331; *HOG0019032*, XM_060938829; *HOG0004451*, XM_036520634; *HOG0030200*, XM_026945431; *HOG0034670*, XM_053236679; *HOG0019111*, XM_047049897; and *HOG0001647*, XM_047032478. Gene names were attributed based on the best hit (highest probability). The 3D structure of the same proteins was predicted using AlphaFold3, on the AlphaFold server (Abramson et al. 2024). Resulting crystallographic information files (CIF) were used as input in FoldSeek (van Kempen et al. 2024) (AFDP-SwissProt and PDB100 databases) to find structurally similar proteins, and the best hit was retained (highest probability).

For each HOG, we used the standard MG94 fit program of HyPhy, with the coding sequence alignment and the unrooted maximum likelihood phylogeny as input, to compute one dN/dS value per branch (Kosakovsky Pond et al. 2019). Branches with a dN or dS saturation (≥1) or with too few synonymous mutations (dS < 0.01) were discarded. We then used aBSREL (Smith et al. 2015) to infer if internal and terminal branches of recipient clades were evolving under diversifying selection. Each branch for which aBSREL detected a signal of diversifying selection (Holm–Bonferroni corrected *P*-value < 0.05) were further assigned as a test branch in RELAX (Wertheim et al. 2015). A

branch was considered under relaxed selection if the RELAX *P*-value was inferior to 0.05 and if the parameter K (relaxation parameter) was inferior to 1.

### Gene Ontology Mining and Analysis

For each HOG, the longest sequence was extracted and used as a query in a BLASTP against the UniProt database. We retained the best match, if it had an *e*-value ≤ 1e-5, and extracted associated GO terms using the R package "UniprotR" (Soudy et al. 2020). We identified over-represented GO terms associated with at least two horizontally transferred genes using Fisher's Exact tests. As backgrounds, we used either all genes (every HOGs), or 5,616 genes passing the dS q0.5 filter, but having a non-null microsynteny score. A GO term was considered as over-represented when it had FDR-corrected *P*-value ≤ 0.05. The results using the two background gene lists were similar, with the same three significant GO terms (supplementary File S2, Supplementary Material online).

## Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

## Acknowledgments

We would like to thank Eric Jenczewski, Matthieu Haudiquet, as well as the members of the Salzburger laboratory, for enriching discussions about this study. We would also like to thank the anonymous reviewers for their comments and suggestions, which improved this study. All calculations were performed at sciCORE (http://scicore.unibas.ch/), the center of scientific computing at the University of Basel (with support from the SIB/Swiss Institute of Bioinformatics).

## Author Contributions

M.P., F.M., and C.G. conceived and designed this study, with input from W.S. M.P. performed all data analyses. C.G. and M.P. drafted the manuscript with input and revisions from F.M. and W.S.

## Funding

This work was funded by the Swiss National Science Foundation (SNSF; grants 189970 and 208002) and by Agence Nationale de la Recherche (project ANR-24-CE02-1004 VIRHOZFER).

## Conflicts of Interest

The authors declare no competing interest.

## Data Availability

All scripts and data needed to reproduce the results of this study are available on GitHub (https://github.com/MaximePolicarpo/HGT_Teleostei) and Zenodo (https://doi.org/10.5281/zenodo.14011660; https://doi.org/10.5281/zenodo.14012814; https://doi.org/10.5281/zenodo.14012963).

## References

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. 2024:630(8016):493–500. https://doi.org/10.1038/s41586-024-07487-w.

Adato O, Ninyo N, Gophna U, Snir S. Detecting horizontal gene transfer between closely related taxa. *PLoS Comput Biol*. 2015:11(10):e1004408. https://doi.org/10.1371/journal.pcbi.1004408.

Arnold BJ, Huang I-T, Hanage WP. Horizontal gene transfer and adaptive evolution in bacteria. *Nat Rev Microbiol*. 2022:20(4):206–218. https://doi.org/10.1038/s41579-021-00650-4.

Bai L, Qiao M, Zheng R, Deng C, Mei S, Chen W. Phylogenomic analysis of transferrin family from animals and plants. *Comp Biochem Physiol Part D: Genomics Proteomics*. 2016:17:1–8. https://doi.org/10.1016/j.cbd.2015.11.002.

Bao W, Kojima KK, Kohany O. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*. 2015:6(1):11. https://doi.org/10.1186/s13100-015-0041-9.

Bartolić P, Morgan EJ, Padilla-García N, Kolář F. Ploidy as a leaky reproductive barrier: mechanisms, rates and evolutionary significance of interploidy gene flow. *Ann Bot*. 2024:134(4):537–550. https://doi.org/10.1093/aob/mcae096.

Bemm F, Weiß CL, Schultz J, Förster F. Genome of a tardigrade: horizontal gene transfer or bacterial contamination? *Proc Natl Acad Sci U S A*. 2016:113(22):E3054–E3056. https://doi.org/10.1073/pnas.1525116113.

Benun Sutton F, Wilson AB. Where are all the moms? External fertilization predicts the rise of male parental care in bony fishes. *Evolution*. 2019:73(12):2451–2460. https://doi.org/10.1111/evo.13846.

Borowiec ML. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*. 2016:4:e1660. https://doi.org/10.7717/peerj.1660.

Boschetti C, Carr A, Crisp A, Eyres I, Wang-Koh Y, Lubzens E, Barraclough TG, Micklem G, Tunnacliffe A. Biochemical diversification through foreign gene expression in bdelloid rotifers. *PLoS Genet*. 2012:8(11):e1003035. https://doi.org/10.1371/journal.pgen.1003035.

Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*. 2021:18(4):366–368. https://doi.org/10.1038/s41592-021-01101-x.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009:10(1):421. https://doi.org/10.1186/1471-2105-10-421.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009:25(15):1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Catalan-Dibene J, McIntyre LL, Zlotnik A. Interleukin 30 to interleukin 40. *J Interferon Cytokine Res*. 2018:38(10):423–439. https://doi.org/10.1089/jir.2018.0089.

Chamberlain SA, Boettiger C. 2017. R Python, and Ruby Clients for GBIF Species Occurrence Data. bioRxiv 3304v1. https://doi.org/10.7287/peerj.preprints.3304v1, 29 September 2017, preprint: not peer reviewed.

Charif D, Lobry JR. Seqinr 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Bastolla U, Porto M, Eduardo Roman H, Vendruscolo M, editors. *Structural approaches to sequence evolution: molecules, networks, populations, biological and medical physics, biomedical engineering*. Berlin, Heidelberg: Springer; 2007. p. 207–232.

Chen J, Luo M, Li S, Tao M, Ye X, Duan W, Zhang C, Qin Q, Xiao J, Liu S. A comparative study of distant hybridization in plants and animals. *Sci China Life Sci*. 2018a:61(3):285–309. https://doi.org/10.1007/s11427-017-9094-2.

Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018b:34(17):i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Chuang H-C, Chu T-W, Cheng A-C, Chen N-Y, Lai Y-S. Iridovirus isolated from marine giant sea perch causes infection in freshwater

ornamental fish. *Aquaculture*. 2022:548:737588. https://doi.org/10.1016/j.aquaculture.2021.737588.
Ciach MA, Pawłowska J, Górecki P, Muszewska A. The interkingdom horizontal gene transfer in 44 early diverging fungi boosted their metabolic, adaptive, and immune capabilities. *Evol Lett*. 2024:8(4):526–538. https://doi.org/10.1093/evlett/qrae009.
Comber SC, Smith C. Polyploidy in fishes: patterns and processes. *Biol J Linn Soc Lond*. 2004:82(4):431–442. https://doi.org/10.1111/j.1095-8312.2004.00330.x.
Costa VA, Holmes EC. Diversity, evolution, and emergence of fish viruses. *J Virol*. 2024:98(6):e00118. https://doi.org/10.1128/jvi.00118-24.
Costa VA, Ronco F, Mifsud JCO, Harvey E, Salzburger W, Holmes EC. Host adaptive radiation is associated with rapid virus diversification and cross-Species transmission in African cichlid fishes. *Curr Biol*. 2024:34(6):1247–1257.e3. https://doi.org/10.1016/j.cub.2024.02.008.
Cote-L'Heureux A, Maurer-Alcalá XX, Katz LA. Old genes in new places: a taxon-rich analysis of interdomain lateral gene transfer events. *PLoS Genet*. 2022:18(6):e1010239. https://doi.org/10.1371/journal.pgen.1010239.
Dabbagh-Gorjani F. A comprehensive review on the role of interleukin-40 as a biomarker for diagnosing inflammatory diseases. *Autoimmune Dis*. 2024:2024(1):3968767. https://doi.org/10.1155/2024/3968767.
Dainat J. 2022. AGAT. Perl. NBIS—National Bioinformatics Infrastructure Sweden.
Dall E, Brandstetter H. Structure and function of legumain in health and disease. *Biochimie*. 2016:122:126–150. https://doi.org/10.1016/j.biochi.2015.09.022.
Danchin EGJ. Lateral gene transfer in eukaryotes: tip of the iceberg or of the ice cube? *BMC Biol*. 2016:14(1):101. https://doi.org/10.1186/s12915-016-0330-x.
Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, *et al*. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021:10(2):giab008. https://doi.org/10.1093/gigascience/giab008.
Deb SK, Edger PP, Chris Pires J, McKain MR. Patterns, mechanisms, and consequences of homoeologous exchange in allopolyploid angiosperms: a genomic and epigenomic perspective. *New Phytol*. 2023:238(6):2284–2304. https://doi.org/10.1111/nph.18927.
Dementiev A, Board J, Sitaram A, Hey T, Kelker MS, Xu X, Hu Y, Vidal-Quist C, Chikwana V, Griffin S, *et al*. The pesticidal Cry6Aa toxin from Bacillus Thuringiensis is structurally similar to HlyE-family alpha pore-forming toxins. *BMC Biol*. 2016:14(1):71. https://doi.org/10.1186/s12915-016-0295-9.
DeVries AL. Glycoproteins as biological antifreeze agents in antarctic fishes. *Science*. 1971:172(3988):1152–1155. https://doi.org/10.1126/science.172.3988.1152.
Dutton RJ, Reiter T. PreHGT: a scalable workflow that screens for horizontal gene transfer within and between kingdoms. *Arcadia Sci*. 2023. https://doi.org/10.57844/arcadia-jfbp-7p11.
Edelman NB, Mallet J. Prevalence and adaptive impact of introgression. *Annu Rev Genet*. 2021:55(1):265–283. https://doi.org/10.1146/annurev-genet-021821-020805.
Edgar RC. Muscle5: high-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat Commun*. 2022:13(1):6968. https://doi.org/10.1038/s41467-022-34630-w.
Ellisdon AM, Reboul CF, Panjikar S, Huynh K, Oellig CA, Winter KL, Dunstone MA, Hodgson WC, Seymour J, Dearden PK, *et al*. Stonefish toxin defines an ancient branch of the perforin-like superfamily. *Proc Natl Acad Sci U S A*. 2015:112(50):15360–15365. https://doi.org/10.1073/pnas.1507622112.
Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019:20(1):238. https://doi.org/10.1186/s13059-019-1832-y.
Filée J, Pouget N, Chandler M. Phylogenetic evidence for extensive lateral acquisition of cellular genes by nucleocytoplasmic large DNA viruses. *BMC Evol Biol*. 2008:8(1):320. https://doi.org/10.1186/1471-2148-8-320.
Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A*. 2020:117(17):9451–9457. https://doi.org/10.1073/pnas.1921046117.
Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, Jiang X, Hall AB, Catteruccia F, Kakani E, *et al*. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*. 2015:347(6217):1258524. https://doi.org/10.1126/science.1258524.
Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012:28(23):3150–3152. https://doi.org/10.1093/bioinformatics/bts565.
Gabaldón T. Patterns and impacts of nonvertical evolution in eukaryotes: a paradigm shift. *Ann N Y Acad Sci*. 2020:1476(1):78–92. https://doi.org/10.1111/nyas.14471.
Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. BRAKER3: fully automated genome annotation using RNA-Seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res*. 2024:34(5):769–777. https://doi.org/10.1101/gr.278090.123.
Gacesa R, Hung JY-h, Bourne DG, Long PF. Horizontal transfer of a natterin-like toxin encoding gene within the holobiont of the reef building coral Acropora digitifera (Cnidaria: Anthozoa: Scleractinia) and across multiple animal linages. *J Venom Res*. 2020:10:7.
Galinier R, Portela J, Moné Y, Allienne JF, Henri H, Delbecq S, Mitta G, Gourbal B, Duval D. Biomphalysin, a new β pore-forming toxin involved in Biomphalaria glabrata immune defense against Schistosoma mansoni. *PLoS Pathog*. 2013:9(3):e1003216. https://doi.org/10.1371/journal.ppat.1003216.
Gao W, Chen X, He J, Sha A, Luo Y, Xiao W, Xiong Z, Li Q. Intraspecific and interspecific variations in the synonymous codon usage in mitochondrial genomes of 8 pleurotus strains. *BMC Genomics*. 2024:25(1):456. https://doi.org/10.1186/s12864-024-10374-3.
Gasmi L, Boulain H, Gauthier J, Hua-Van A, Musset K, Jakubowska AK, Aury JM, Volkoff AN, Huguet E, Herrero S, *et al*. Recurrent domestication by lepidoptera of genes from their parasites mediated by bracoviruses. *PLoS Genet*. 2015:11(9):e1005470. https://doi.org/10.1371/journal.pgen.1005470.
Gasmi L, Sieminska E, Okuno S, Ohta R, Coutu C, Vatanparast M, Harris S, Baldwin D, Hegedus DD, Theilmann DA, *et al*. Horizontally transmitted parasitoid killing factor shapes insect defense to parasitoids. *Science*. 2021:373(6554):535–541. https://doi.org/10.1126/science.abb6396.
Ghadessy FJ, Chen D, Manjunatha Kini R, Chung MCM, Jeyaseelan K, Khoo HE, Yuen R. Stonustoxin is a novel lethal factor from stonefish (*Synanceja horrida*) venom. *J Biol Chem*. 1996:271(41):25575–-25581. https://doi.org/10.1074/jbc.271.41.25575.
Gilbert C, Cordaux R. Viruses as vectors of horizontal transfer of genetic material in eukaryotes. *Curr Opin Virol*. 2017:25:16–22. https://doi.org/10.1016/j.coviro.2017.06.005.
Gilbert C, Feschotte C. Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. *Curr Opin Genet Dev*. 2018:49:15–24. https://doi.org/10.1016/j.gde.2018.02.007.
Gilbert C, Maumus F. Multiple horizontal acquisitions of plant genes in the whitefly Bemisia tabaci. *Genome Biol Evol*. 2022:14(10):evac141. https://doi.org/10.1093/gbe/evac141.
Gilbert C, Maumus F. Sidestepping Darwin: horizontal gene transfer from plants to insects. *Curr Opin Insect Sci*. 2023:57:101035. https://doi.org/10.1016/j.cois.2023.101035.
Gladyshev EA, Meselson M, Arkhipova IR. Massive horizontal gene transfer in Bdelloid rotifers. *Science*. 2008:320(5880):1210–1213. https://doi.org/10.1126/science.1156407.
Graham LA, Davies PL. Horizontal gene transfer in vertebrates: a fishy tale. *Trends Genet*. 2021:37(6):501–503. https://doi.org/10.1016/j.tig.2021.02.006.

Graham LA, Lougheed SC, Vanya Ewart K, Davies PL. Lateral transfer of a lectin-like antifreeze protein gene in fishes. *PLoS One*. 2008:3(7):e2616. https://doi.org/10.1371/journal.pone.0002616.

Grant BJ, Rodrigues APC, ElSawy KM, Andrew McCammon J, Caves LSD. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*. 2006:22(21):2695–2696. https://doi.org/10.1093/bioinformatics/btl461.

Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in R. *Bioinformatics*. 2014:30(19):2811–2812. https://doi.org/10.1093/bioinformatics/btu393.

Hackl T, Ankenbrand MJ, van Adrichem B. Gggenomes. R [accessed 2024 Jan 5]. https://github.com/thackl/gggenomes.

Han Z, Xu S, Gao T. Unexpected complex horizontal gene transfer in teleost fish. *Curr Zool*. 2023:69(2):222–223. https://doi.org/10.1093/cz/zoac032.

Haudiquet M, de Sousa JM, Touchon M, Rocha EPC. Selfish, promiscuous and sometimes useful: how mobile genetic elements drive horizontal gene transfer in microbial populations. *Philos Trans R Soc B: Biol Sci*. 2022:377(1861):20210234. https://doi.org/10.1098/rstb.2021.0234.

Heisserer C, Muller H, Jouan V, Musset K, Periquet G, Drezen J-M, Volkoff A-N, Gilbert C. Massive somatic and germline chromosomal integrations of polydnaviruses in lepidopterans. *Mol Biol Evol*. 2023:40(3):msad050. https://doi.org/10.1093/molbev/msad050.

Hibbins MS, Hahn MW. Phylogenomic approaches to detecting and characterizing introgression. *Genetics*. 2022:220(2):iyab173. https://doi.org/10.1093/genetics/iyab173.

Hibdige SGS, Raimondeau P, Christin P-A, Dunning LT. Widespread lateral gene transfer among grasses. *New Phytol*. 2021:230(6):2474–2486. https://doi.org/10.1111/nph.17328.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018:35(2):518–522. https://doi.org/10.1093/molbev/msx281.

Huang J. Horizontal gene transfer in eukaryotes: the weak-link model. *BioEssays*. 2013:35(10):868–875. https://doi.org/10.1002/bies.201300007.

Inoue Y, Takeda H. Teratorn and its relatives—a cross-point of distinct mobile elements, transposons and viruses. *Front Vet Sci*. 2023:10:1158023. https://doi.org/10.3389/fvets.2023.1158023.

Jafari A, Qanie D, Andersen TL, Zhang Y, Chen L, Postert B, Parsons S, Ditzel N, Khosla S, Johansen HT, *et al*. Legumain regulates differentiation fate of human bone marrow stromal cells and is altered in postmenopausal osteoporosis. *Stem Cell Rep*. 2017:8(2):373–386. https://doi.org/10.1016/j.stemcr.2017.01.003.

Jaramillo VDA, Sukno SA, Thon MR. Identification of horizontally transferred genes in the genus Colletotrichum reveals a steady tempo of bacterial to fungal gene transfer. *BMC Genomics*. 2015:16(1):2. https://doi.org/10.1186/1471-2164-16-2.

Jia N, Liu N, Cheng W, Jiang YL, Sun H, Chen LL, Peng J, Zhang Y, Ding YH, Zhang ZH, *et al*. Structural basis for receptor recognition and pore formation of a zebrafish aerolysin-like protein. *EMBO Rep*. 2016:17(2):235–248. https://doi.org/10.15252/embr.201540851.

Jones MR, Scott Mills L, Alves PC, Callahan CM, Alves JM, Lafferty DJR, Jiggins FM, Jensen JD, Melo-Ferreira J, Good JM. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science*. 2018:360(6395):1355–1358. https://doi.org/10.1126/science.aar5273.

Káldy J, Mozsár A, Fazekas G, Farkas M, Fazekas DL, Fazekas GL, Goda K, Gyöngy Z, Kovács B, Semmens K, *et al*. Hybridization of Russian sturgeon (Acipenser gueldenstaedtii, Brandt and Ratzeberg, 1833) and American Paddlefish (Polyodon spathula, Walbaum 1792) and evaluation of their progeny. *Genes (Basel)*. 2020:11(7):753. https://doi.org/10.3390/genes11070753.

Kalluraya CA, Weitzel AJ, Tsu BV, Daugherty MD. Bacterial origin of a key innovation in the evolution of the vertebrate eye. *Proc Natl Acad Sci U S A*. 2023:120(16):e2214815120. https://doi.org/10.1073/pnas.2214815120.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017:14(6):587–589. https://doi.org/10.1038/nmeth.4285.

Keeling PJ. Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Curr Opin Genet Dev*. 2009:19(6):613–619. https://doi.org/10.1016/j.gde.2009.10.001.

Keeling PJ. Horizontal gene transfer in eukaryotes: aligning theory with data. *Nat Rev Genet*. 2024:25(6):416–430. https://doi.org/10.1038/s41576-023-00688-5.

Kempen Mv, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J, Steinegger M. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol*. 2024:42(2):243–246. https://doi.org/10.1038/s41587-023-01773-0.

Kirsch R, Okamura Y, Haeger W, Vogel H, Kunert G, Pauchet Y. Metabolic novelty originating from horizontal gene transfer is essential for leaf beetle survival. *Proc Natl Acad Sci U S A*. 2022:119(40):e2205857119. https://doi.org/10.1073/pnas.2205857119.

Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A, *et al*. Hyphy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. *Mol Biol Evol*. 2019:37(1):295–299. https://doi.org/10.1093/molbev/msz197.

Koutsovoulos GD, Noriot SG, Bailly-Bechet M, Danchin EGJ, Rancurel C. Avp: a software package for automatic phylogenetic detection of candidate horizontal gene transfers. *PLoS Comput Biol*. 2022:18(11):e1010686. https://doi.org/10.1371/journal.pcbi.1010686.

Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Blair Hedges S. TimeTree 5: an expanded resource for Species divergence times. *Mol Biol Evol*. 2022:39(8):msac174. https://doi.org/10.1093/molbev/msac174.

Kuraku S, Qiu H, Meyer A. Horizontal transfers of Tc1 elements between teleost fishes and their vertebrate parasites, lampreys. *Genome Biol Evol*. 2012:4(9):929–936. https://doi.org/10.1093/gbe/evs069.

Kuznetsov D, Tegenfeldt F, Manni M, Seppey M, Berkeley M, Kriventseva EV, Zdobnov EM. OrthoDB V11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res*. 2022:51(D1):D445–D451. https://doi.org/10.1093/nar/gkac998.

Lavitrano M, Giovannoni R, Cerrito MG. Methods for sperm-mediated gene transfer. In: Carrell DT, Aston KI, editors. *Spermatogenesis: methods and protocols*. Totowa, NJ: Humana Press; 2013. p. 519–529.

Le VS, Dang CC, Le QS. Improved mitochondrial amino acid substitution models for metazoan evolutionary studies. *BMC Evol Biol*. 2017:17(1):136. https://doi.org/10.1186/s12862-017-0987-y.

Leger MM, Eme L, Stairs CW, Roger AJ. Demystifying eukaryote lateral gene transfer (response to Martin 2017 DOI: 10.1002/Bies.201700115). *BioEssays*. 2018:40(5):1700242. https://doi.org/10.1002/bies.201700242.

Leiva-Rebollo R, Labella AM, Gémez-Mata J, Castro D, Borrego JJ. Fish iridoviridae: infection, vaccination and immune response. *Vet Res*. 2024:55(1):88. https://doi.org/10.1186/s13567-024-01347-1.

Levraud J-P, Jouneau L, Briolat V, Laghi V, Boudinot P. "IFN-stimulated genes in zebrafish and humans define an ancient arsenal of antiviral immunity. *J Immunol*. 2019:203(12):3361–3373. https://doi.org/10.4049/jimmunol.1900804.

Li H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics*. 2021:37(23):4572–4574. https://doi.org/10.1093/bioinformatics/btab705.

Li Y, Liu Z, Liu C, Shi Z, Pang L, Chen C, Chen Y, Pan R, Zhou W, Chen XX, *et al*. HGT is widespread in insects and contributes to male courtship in lepidopterans. *Cell*. 2022:185(16):2975–-2987.e10. https://doi.org/10.1016/j.cell.2022.06.014.

Lima C, Disner GR, Falcão MAP, Seni-Silva AC, Maleski ALA, Souza MM, Reis Tonello MC, Lopes-Ferreira M. The natterin proteins diversity: a review on phylogeny, structure, and immune function.

*Toxins (Basel)*. 2021:13(8):538. https://doi.org/10.3390/toxins13080538.

Limborg MT, Blankenship SM, Young SF, Utter FM, Seeb LW, Hansen MHH, Seeb JE. Signatures of natural selection among lineages and habitats in Oncorhynchus mykiss. *Ecol Evol*. 2012:2(1):1–18. https://doi.org/10.1002/ece3.59.

Liu Q, Liu J, Yuan L, Li L, Tao M, Zhang C, Qin Q, Chen B, Ma M, Tang C, *et al*. The establishment of the fertile fish lineages derived from distant hybridization by overcoming the reproductive barriers. *Reproduction*. 2020:159(6):R237–R249. https://doi.org/10.1530/REP-19-0576.

Liu Q, Zhang X, Liu J, Liu F, Shi F, Qin Q, Tao M, Tang C, Liu S. A new type of allodiploid hybrids derived from female Megalobrama amblycephala × male Gobiocypris rarus. *Front Genet*. 2021:12:685914. https://doi.org/10.3389/fgene.2021.685914.

Liu Y, Zha H, Yu S, Zhong J, Liu X, Yang H, Zhu Q. Molecular characterization and antibacterial activities of a goose-type lysozyme gene from Roughskin sculpin (*Trachidermus Fasciatus*). *Fish Shellfish Immunol*. 2022:127:1079–1087. https://doi.org/10.1016/j.fsi.2022.07.053.

Loiseau V, Peccoud J, Bouzar C, Guillier S, Fan J, Gueli Alletti G, Meignin C, Herniou EA, Federici BA, Wennmann JT, *et al*. Monitoring insect transposable elements in large double-stranded DNA viruses reveals host-to-virus and virus-to-virus transposition. *Mol Biol Evol*. 2021:38(9):3512–3530. https://doi.org/10.1093/molbev/msab198.

Ma J, Wang S, Zhu X, Sun G, Chang G, Li L, Hu X, Zhang S, Zhou Y, Song CP, *et al*. Major episodes of horizontal gene transfer drove the evolution of land plants. *Mol Plant*. 2022:15(5):857–871. https://doi.org/10.1016/j.molp.2022.02.001.

Ma L, Shen Y-Q, Khatri HP, Schachner M. The asparaginyl endopeptidase legumain is essential for functional recovery after spinal cord injury in adult zebrafish. *PLoS One*. 2014a:9(4):e95098. https://doi.org/10.1371/journal.pone.0095098.

Ma Q, Yu T, Ren Y-Y, Gong T, Zhong D-S. Overexpression of SAMD9 suppresses tumorigenesis and progression during non small cell lung cancer. *Biochem Biophys Res Commun*. 2014b:454(1):157–161. https://doi.org/10.1016/j.bbrc.2014.10.054.

Magalhães GS, Junqueira-de-Azevedo ILM, Lopes-Ferreira M, Lorenzini DM, Ho PL, Moura-da-Silva AM. Transcriptome analysis of expressed sequence tags from the venom glands of the fish Thalassophryne nattereri. *Biochimie*. 2006:88(6):693–699. https://doi.org/10.1016/j.biochi.2005.12.008.

Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez gene: gene-centered information at NCBI. *Nucleic Acids Res*. 2007:35(Database):D26–D31. https://doi.org/10.1093/nar/gkl993.

Mahelka V, Krak K, Kopecký D, Fehrer J, Šafář J, Bartoš J, Hobza R, Blavet N, Blattner FR. Multiple horizontal transfers of nuclear ribosomal genes between phylogenetically distinct grass lineages. *Proc Natl Acad Sci U S A*. 2017:114(7):1726–1731. https://doi.org/10.1073/pnas.1613375114.

Makarova KS, Wolf YI, Mekhedov SL, Mirkin BG, Koonin EV. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res*. 2005:33(14):4626–4638. https://doi.org/10.1093/nar/gki775.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021:38(10):4647–4654. https://doi.org/10.1093/molbev/msab199.

Manoury B, Hewitt EW, Morrice N, Dando PM, Barrett AJ, Watts C. An asparaginyl endopeptidase processes a microbial antigen for class II MHC presentation. *Nature*. 1998:396(6712):695–699. https://doi.org/10.1038/25379.

Marcet-Houben M, Gabaldón T. Acquisition of prokaryotic genes by fungal genomes. *Trends Genet*. 2010:26(1):5–8. https://doi.org/10.1016/j.tig.2009.11.007.

Martin WF. Too much eukaryote LGT. *BioEssays*. 2017:39(12):1700115. https://doi.org/10.1002/bies.201700115.

Martin WF. Eukaryote lateral gene transfer is Lamarckian. *Nat Ecol Evol*. 2018:2(5):754. https://doi.org/10.1038/s41559-018-0521-7.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020:37(5):1530–1534. https://doi.org/10.1093/molbev/msaa015.

Mishina T, Chiu MC, Hashiguchi Y, Oishi S, Sasaki A, Okada R, Uchiyama H, Sasaki T, Sakura M, Takeshima H, *et al*. Massive horizontal gene transfer and the evolution of nematomorph-driven behavioral manipulation of mantids. *Curr Biol*. 2023:33(22):4988–4994.e5. https://doi.org/10.1016/j.cub.2023.09.052.

Moran NA, Jarvik T. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science*. 2010:328(5978):624–627. https://doi.org/10.1126/science.1187113.

Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Leo Smith W. Resolution of ray-finned fish phylogeny and timing of diversification. *Proc Natl Acad Sci U S A*. 2012:109(34):13698–13703. https://doi.org/10.1073/pnas.1206625109.

Nounamo B, Li Y, O'Byrne P, Kearney AM, Khan A, Liu J. An interaction domain in human SAMD9 is essential for myxoma virus host-range determinant M062 antagonism of host anti-viral function. *Virology*. 2017:503:94–102. https://doi.org/10.1016/j.virol.2017.01.004.

Novembre JA. Accounting for background nucleotide composition when measuring codon usage bias. *Mol Biol Evol*. 2002:19(8):1390–1394. https://doi.org/10.1093/oxfordjournals.molbev.a004201.

Nowell RW, Rodriguez F, Hecox-Lea BJ, Mark Welch DB, Arkhipova IR, Barraclough TG, Wilson CG. Bdelloid rotifers deploy horizontally acquired biosynthetic genes against a fungal pathogen. *Nat Commun*. 2024:15(1):5787. https://doi.org/10.1038/s41467-024-49919-1.

Paganini J, Campan-Fournier A, Da Rocha M, Gouret P, Pontarotti P, Wajnberg E, Abad P, Danchin EG. Contribution of lateral gene transfers to the genome composition and parasitic ability of root-knot nematodes. *PLoS One*. 2012:7(11):e50875. https://doi.org/10.1371/journal.pone.0050875.

Paradis E, Schliep K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. 2019:35(3):526–528. https://doi.org/10.1093/bioinformatics/bty633.

Peng S, Meng X, Zhang F, Pathak PK, Chaturvedi J, Coronado J, Morales M, Mao Y, Qian SB, Deng J, *et al*. Structure and function of an effector domain in antiviral factors and tumor suppressors SAMD9 and SAMD9L. *Proc Natl Acad Sci U S A*. 2022:119(4):e2116550119. https://doi.org/10.1073/pnas.2116550119.

Peraro MD, van der Goot FG. Pore-forming toxins: ancient, but never really out of fashion. *Nat Rev Microbiol*. 2016:14(2):77–92. https://doi.org/10.1038/nrmicro.2015.3.

Pereira L, Christin P-A, Dunning LT. The mechanisms underpinning lateral gene transfer between grasses. *Plants People Planet*. 2023:5(5):672–682. https://doi.org/10.1002/ppp3.10347.

Piégu B, Guizard S, Yeping T, Cruaud C, Asgari S, Bideshi DK, Federici BA, Bigot Y. Genome sequence of a crustacean iridovirus, IIV31, isolated from the pill bug, Armadillidium vulgare. *J Gen Virol*. 2014:95(7):1585–1590. https://doi.org/10.1099/vir.0.066076-0.

Provoost P, Bosch S. 2022. Robis: Ocean Biodiversity Information System (OBIS) Client [accessed date 2024 Jan 08]. https://github.com/iobis/robis.

Qiao H, Wang Y, Zhang X, Lu R, Niu J, Nan F, Ke D, Zeng Z, Wang Y, Wang B. Cross-species opsonic activity of zebrafish fish-egg lectin on mouse macrophages. *Dev Comp Immunol*. 2022:129:104332. https://doi.org/10.1016/j.dci.2021.104332.

Ravenhall M, Škunca N, Lassalle F, Dessimoz C. Inferring horizontal gene transfer. *PLoS Comput Biol*. 2015:11(5):e1004095. https://doi.org/10.1371/journal.pcbi.1004095.

R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: Foundation for Statistical Computing; 2018.

Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends Genet*. 2000:16(6):276–277. https://doi.org/10.1016/S0168-9525(00)02024-2.

Richter K, Brar S, Ray M, Pisitkun P, Bolland S, Verkoczy L, Diaz M. Speckled-like pattern in the germinal center (SLIP-GC), a nuclear GTPase expressed in activation-induced deaminase-expressing lymphomas and germinal center B cells. *J Biol Chem*. 2009:284(44):30652–30661. https://doi.org/10.1074/jbc.M109.014506.

Richter K, Burch L, Chao F, Henke D, Jiang C, Daly J, Zhao M-L, Kissling G, Diaz M. Altered pattern of immunoglobulin hypermutation in mice deficient in *Slip-GC* protein*. *J Biol Chem*. 2012:287(38):31856–31865. https://doi.org/10.1074/jbc.M112.340661.

Rincon-Sandoval M, De-Kayne R, Shank SD, Pirro S, Ko'ou A, Abueg L, Tracey A, Mountcastle J, O'Toole B, Balacco J, *et al*. Ecological diversification of sea catfishes is accompanied by genome-wide signatures of positive selection. *Nat Commun*. 2024:15(1):10040. https://doi.org/10.1038/s41467-024-54184-3.

Roger AJ. Reply to 'eukaryote lateral gene transfer is Lamarckian. *Nat Ecol Evol*. 2018:2(5):755–755. https://doi.org/10.1038/s41559-018-0522-6.

Sahu N, Indic B, Wong-Bajracharya J, Merényi Z, Ke HM, Ahrendt S, Monk TL, Kocsubé S, Drula E, Lipzen A, *et al*. Vertical and horizontal gene transfer shaped plant colonization and biomass degradation in the fungal genus Armillaria. *Nat Microbiol*. 2023:8(9):1668–1681. https://doi.org/10.1038/s41564-023-01448-1.

Salzberg SL. Horizontal gene transfer is not a hallmark of the human genome. *Genome Biol*. 2017:18(1):85. https://doi.org/10.1186/s13059-017-1214-2.

Salzberg SL, White O, Peterson J, Eisen JA. Microbial genes in the human genome: lateral transfer or gene loss? *Science*. 2001:292(5523):1903–1906. https://doi.org/10.1126/science.1061036.

Schlupp I. The evolutionary ecology of gynogenesis. *Annu Rev Ecol Evol Syst*. 2005:36(1):399–417. https://doi.org/10.1146/annurev.ecolsys.36.102003.152629.

Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, Leipe D, Mcveigh R, O'Neill K, Robbertse B, *et al*. NCBI taxonomy: a comprehensive update on curation, resources and tools. *Database (Oxford)*. 2020:2020:baaa062. https://doi.org/10.1093/database/baaa062.

Schönknecht G, Weber APM, Lercher MJ. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. *BioEssays*. 2014:36(1):9–20. https://doi.org/10.1002/bies.201300095.

Seni-Silva AC, Maleski ALA, Souza MM, Falcao MAP, Disner GR, Lopes-Ferreira M, Lima C. Natterin-like depletion by CRISPR/cas9 impairs zebrafish (Danio rerio) embryonic development. *BMC Genomics*. 2022:23(1):123. https://doi.org/10.1186/s12864-022-08369-z.

Sharp PM, Li W-H. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J Mol Evol*. 1986:24(1-2):28–38. https://doi.org/10.1007/BF02099948.

Shen XX, Opulente DA, Kominek J, Zhou X, Steenwyk JL, Buh KV, Haase MAB, Wisecaver JH, Wang M, Doering DT, *et al*. Tempo and mode of genome evolution in the budding yeast subphylum. *Cell*. 2018:175(6):1533–1545.e20. https://doi.org/10.1016/j.cell.2018.10.023.

Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. 2005:6(1):31. https://doi.org/10.1186/1471-2105-6-31.

Smit AFA, Hubley R, Green P. 2013. RepeatMasker Open-4.0 [accessed date 2024 Feb 28]. http://www.repeatmasker.org.

Smith K, Spadafora C. Sperm-mediated gene transfer: applications and implications. *BioEssays*. 2005:27(5):551–562. https://doi.org/10.1002/bies.20211.

Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Pond SLK. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol*. 2015:32(5):1342–1353. https://doi.org/10.1093/molbev/msv022.

Söding J. Protein homology detection by HMM–HMM comparison. *Bioinformatics*. 2005:21(7):951–960. https://doi.org/10.1093/bioinformatics/bti125.

Soucy SM, Huang J, Gogarten JP. Horizontal gene transfer: building the web of life. *Nat Rev Genet*. 2015:16(8):472–482. https://doi.org/10.1038/nrg3962.

Soudy M, Anwar AM, Ahmed EA, Osama A, Ezzeldin S, Mahgoub S, Magdeldin S. Uniprotr: retrieving and visualizing protein sequence and functional information from universal protein resource (UniProt knowledgebase). *J Proteomics*. 2020:213:103613. https://doi.org/10.1016/j.jprot.2019.103613.

Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob DNA*. 2021:12(1):2. https://doi.org/10.1186/s13100-020-00230-y.

Szöllősi GJ, Davín AA, Tannier E, Daubin V, Boussau B. Genome-scale phylogenetic analysis finds extensive gene transfer among fungi. *Philos Trans R Soc B: Biol Sci*. 2015:370(1678):20140335. https://doi.org/10.1098/rstb.2014.0335.

Tarnopol RL, Tamsil JA, Cinege G, Ha JH, Verster KI, Ábrahám E, Magyar LB, Kim BY, Bernstein SL, Lipinszki Z, *et al*. Experimental horizontal transfer of phage-derived genes to Drosophila confers innate immunity to parasitoids. *Curr Biol*. 2025:35(3):514–529.e7. https://doi.org/10.1016/j.cub.2024.11.071.

To T-H, Jung M, Lycett S, Gascuel O. Fast dating using least-squares criteria and algorithms. *Syst Biol*. 2016:65(1):82–97. https://doi.org/10.1093/sysbio/syv068.

Tullio V, Spaccapelo R, Polimeni M. Lysozymes in the animal kingdom. In: Prato M, editors. *Human and mosquito lysozymes: old molecules for new approaches against malaria*. Cham: Springer International Publishing; 2015. p. 45–57.

Ullrich KK. 2024. MSA2dist: MSA2dist calculates pairwise distances between all sequences of a DNAStringSet or a AAStringSet using a custom score matrix and conducts codon based analysis [accessed date 2023 Dec 8]. https://gitlab.gwdg.de/mpievolbio-it/MSA2dist.

UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021:49(D1):D480–D489. https://doi.org/10.1093/nar/gkaa1100.

Van Etten J, Bhattacharya D. Horizontal gene transfer in eukaryotes: not if, but how much? *Trends Genet*. 2020:36(12):915–925. https://doi.org/10.1016/j.tig.2020.08.006.

Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-MEM for Multicore systems. In 2019 IEEE international parallel and distributed processing symposium (IPDPS), Rio de Janeiro, Brazil, 314–324.

Verma P, Gandhi S, Lata K, Chattopadhyay K. Pore-forming toxins in infection and immunity. *Biochem Soc Trans*. 2021:49(1):455–465. https://doi.org/10.1042/BST20200836.

Vinkler M, Fiddaman SR, Těšický M, O'Connor EA, Savage AE, Lenz TL, Smith AL, Kaufman J, Bolnick DI, Davies CS, *et al*. Understanding the evolution of immune genes in jawed vertebrates. *J Evol Biol*. 2023:36(6):847–873. https://doi.org/10.1111/jeb.14181.

Wang S, Ye X, Wang Y, Chen Y, Lin B, Yi Z, Mao Z, Hu F, Zhao R, Wang J, *et al*. A new type of homodiploid fish derived from the interspecific hybridization of female common carp × male blunt snout bream. *Sci Rep*. 2017:7(1):4189. https://doi.org/10.1038/s41598-017-04582-z.

Wang Y, Bu L, Yang L, Li H, Zhang S. Identification and functional characterization of fish-egg lectin in zebrafish. *Fish Shellfish Immunol*. 2016:52:23–30. https://doi.org/10.1016/j.fsi.2016.03.016.

Wang Y, Yao J, Liao A, Tan H, Luo Y, Wu P, Wang S, Zhang C, Qin Q, Tao M, *et al*. The formation of hybrid fish derived from hybridization of *Megalobrama Amblycephala* (♀) × *Siniperca Chuatsi* (♂). *Aquaculture*. 2022:548:737547. https://doi.org/10.1016/j.aquaculture.2021.737547.

Waterhouse RM, Zdobnov EM, Kriventseva EV. Correlating traits of gene retention, sequence divergence, duplicability and essentiality

in vertebrates, arthropods, and fungi. *Genome Biol Evol*. 2010:3: 75–86. https://doi.org/10.1093/gbe/evq083.

Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol*. 2015:32(3):820–832. https://doi.org/10.1093/molbev/msu400.

Widen SA, Bes IC, Koreshova A, Pliota P, Krogull D, Burga A. Virus-like transposons cross the Species barrier and drive the evolution of genetic incompatibilities. *Science*. 2023:380(6652):eade0705. https://doi.org/10.1126/science.ade0705.

Willerslev E, Mourier T, Hansen AJ, Christensen B, Barnes I, Salzberg SL. Contamination in the draft of the human genome masquerades as lateral gene transfer. *DNA Seq*. 2002:13(2):75–76. https://doi.org/10.1080/10425170290023392.

Wilson CG, Pieszko T, Nowell RW, Barraclough TG. Recombination in bdelloid Rotifer genomes: asexuality, transfer and stress. *Trends Genet*. 2024:40(5):422–436. https://doi.org/10.1016/j.tig.2024.02.001.

Wybouw N, Pauchet Y, Heckel DG, Van Leeuwen T. Horizontal gene transfer contributes to the evolution of arthropod herbivory. *Genome Biol Evol*. 2016:8(6):1785–1801. https://doi.org/10.1093/gbe/evw119.

Wybouw N, Van Leeuwen T, Dermauw W. A massive incorporation of microbial genes into the genome of Tetranychus urticae, a polyphagous arthropod herbivore. *Insect Mol Biol*. 2018:27(3):333–351. https://doi.org/10.1111/imb.12374.

Xia J, Guo Z, Yang Z, Han H, Wang S, Xu H, Yang X, Yang F, Wu Q, Xie W, *et al*. Whitefly hijacks a plant detoxification gene that neutralizes plant toxins. *Cell*. 2021:184(7):1693–1705.e17. https://doi.org/10.1016/j.cell.2021.02.014.

Xiang Y, Yan C, Guo X, Zhou K, Li S, Gao Q, Wang X, Zhao F, Liu J, Lee WH, *et al*. Host-derived, pore-forming toxin–like protein and trefoil factor Complex protects the host against microbial infection. *Proc Natl Acad Sci U S A*. 2014:111(18):6702–6707. https://doi.org/10.1073/pnas.1321317111.

Xiao J, Zhong H, Liu Z, Yu F, Luo Y, Gan X, Zhou Y. Transcriptome analysis revealed positive selection of immune-related genes in tilapia. *Fish Shellfish Immunol*. 2015:44(1):60–65. https://doi.org/10.1016/j.fsi.2015.01.022.

Yang Z, Zhang Y, Wafula EK, Honaas LA, Ralph PE, Jones S, Clarke CR, Liu S, Su C, Zhang H, *et al*. Horizontal gene transfer is more frequent with increased heterotrophy and contributes to parasite adaptation. *Proc Natl Acad Sci U S A*. 2016:113(45):E7010–E7019. https://doi.org/10.1073/pnas.1608765113.

Yoxsimer AM, Offenberg EG, Katzer AW, Bell MA, Massengill RL, Kingsley DM. Genomic sequence of the threespine stickleback iridovirus (TSIV) from wild Gasterosteus aculeatus in stormy lake, Alaska. *Viruses*. 2024:16(11):1663. https://doi.org/10.3390/v16111663.

Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. Ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol*. 2017:8(1): 28–36. https://doi.org/10.1111/2041-210X.12628.

Yuan L, Lu H, Li F, Nielsen J, Kerkhoven EJ. HGTphyloDetect: facilitating the identification and phylogenetic analysis of horizontal gene transfer. *Brief Bioinform*. 2023:24(2):bbad035. https://doi.org/10.1093/bib/bbad035.

Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time Species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*. 2018:19(S6):153. https://doi.org/10.1186/s12859-018-2129-y.

Zhang H-H, Peccoud J, Xu M-R-X, Zhang X-G, Gilbert C. Horizontal transfer and evolution of transposable elements in vertebrates. *Nat Commun*. 2020a:11(1):1362. https://doi.org/10.1038/s41467-020-15149-4.

Zhang K, Liu X, Li X, Liu Y, Yu H, Liu J, Zhang Q. Antibacterial functions of a novel fish-egg lectin from spotted knifejaw (*Oplegnathus punctatus*) during host defense immune responses. *Dev Comp Immunol*. 2020b:111:103758. https://doi.org/10.1016/j.dci.2020.103758.

Zhang K, Wang Y, Zhang Y, Shan X. Codon usage characterization and phylogenetic analysis of the mitochondrial genome in Hemerocallis citrina. *BMC Genom Data*. 2024:25(1):6. https://doi.org/10.1186/s12863-024-01191-4.

Zhang Z, Chen J, Li L, Tao M, Zhang C, Qin Q, Xiao J, Liu Y, Liu S. Research advances in animal distant hybridization. *Sci China Life Sci*. 2014:57(9):889–902. https://doi.org/10.1007/s11427-014-4707-1.

Zhang Z, Gou X, Xun H, Bian Y, Ma X, Li J, Li N, Gong L, Feldman M, Liu B, *et al*. Homoeologous exchanges occur through intragenic recombination generating novel transcripts and proteins in wheat and other polyploids. *Proc Natl Acad Sci U S A*. 2020b:117(25):14561–-14571. https://doi.org/10.1073/pnas.2003505117.

Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol*. 2018:430(15):2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007.