

# **An atlas of fish genome evolution reveals delayed rediploidization following the teleost whole-genome duplication**

Elise Parey<sup>1,2</sup>, Alexandra Louis<sup>1</sup>, Jérôme Montfort<sup>2</sup>, Yann Guiguen<sup>2</sup>, Hugues Roest Crollius<sup>1\*</sup>, Camille Berthelot<sup>1,3\*</sup>

<sup>1</sup> Institut de Biologie de l'Ecole normale supérieure (IBENS), Département de Biologie, Ecole normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France.

<sup>2</sup> INRAE, LPGP, 35000, Rennes, France.

<sup>3</sup> *Current address*: Institut Pasteur, Université de Paris, CNRS UMR 3525, INSERM UA12, Comparative Functional Genomics group, F-75015 Paris, France.

\* Corresponding authors

## **Supplementary Material**

### **Supplementary Tables**

Supplementary Table S1: List of the 101 genome assemblies used in this study.

Supplementary Table S2: Ohnologs retention across the 74 teleost genomes.

Supplementary Table S3: Gene ontology terms enriched in LORe interspersed

Supplementary Table S4: KEGG pathway enriched in LORe interspersed

Supplementary Table S5: Total gene retention on homeologous chromosomes in the post-duplication teleost ancestor Osteoglossocephalai.

Supplementary Table S6: Conservative estimate of total gene retention on homeologous chromosomes in Medaka.

### **Supplementary Figures**

Supplementary Figure S1: Species tree of the 101 genomes used in this study.

Supplementary Figure S2: Illustration of the key steps in the comparative atlas workflow.

Supplementary Figure S3: Noise robustness of the comparative atlas pipeline.

Supplementary Figure S4: Impact of SCORPiOs on the establishment of the comparative atlas.

Supplementary Figure S5: Karyotypes of medaka, stickleback and tetraodon with 'a' and 'b' gene copies annotations.

Supplementary Figure S6: Species tree of the 6 teleosts for the rediploidisation timing analysis.

Supplementary Figure S7: Gene retention on anciently duplicated chromosome copies in medaka, using the bowfin genome as a proxy for ancestral gene order.

Supplementary S8: Gene retention on anciently duplicated chromosome copies using different ingroup and outgroup genomes.

Supplementary Figure S9: Zebrafish ohnologs gene expression across homeologs, averaged across tissues.

Supplementary Figure S10-S11: Ohnologs gene expression across homeologs in 11 tissues, respectively in Medaka and Zebrafish.

| Species                            | Source  | Data        |
|------------------------------------|---------|-------------|
| <i>Acanthochromis polyacanthus</i> | ensembl | ensembl V95 |
| <i>Amphilophus citrinellus</i>     | ensembl | ensembl V95 |
| <i>Amphiprion ocellaris</i>        | ensembl | ensembl V95 |
| <i>Amphiprion percula</i>          | ensembl | ensembl V95 |
| <i>Anabas testudineus</i>          | ensembl | ensembl V95 |
| <i>Anolis carolinensis</i>         | ensembl | ensembl V95 |
| <i>Astatotilapia calliptera</i>    | ensembl | ensembl V95 |
| <i>Astyanax mexicanus</i>          | ensembl | ensembl V95 |
| <i>Bos taurus</i>                  | ensembl | ensembl V95 |
| <i>Caenorhabditis elegans</i>      | ensembl | ensembl V95 |
| <i>Canis lupus familiaris</i>      | ensembl | ensembl V95 |
| <i>Chrysemys pictabellii</i>       | ensembl | ensembl V95 |
| <i>Ciona intestinalis</i>          | ensembl | ensembl V95 |
| <i>Cynoglossus semilaevis</i>      | ensembl | ensembl V95 |
| <i>Cyprinodon variegatus</i>       | ensembl | ensembl V95 |
| <i>Danio rerio</i>                 | ensembl | ensembl V95 |
| <i>Drosophila melanogaster</i>     | ensembl | ensembl V95 |
| <i>Eptatretus burgeri</i>          | ensembl | ensembl V95 |
| <i>Esox lucius</i>                 | ensembl | ensembl V95 |
| <i>Fundulus heteroclitus</i>       | ensembl | ensembl V95 |
| <i>Gadus morhua</i>                | ensembl | ensembl V95 |
| <i>Gallus gallus</i>               | ensembl | ensembl V95 |
| <i>Gambusia affinis</i>            | ensembl | ensembl V95 |
| <i>Gasterosteus aculeatus</i>      | ensembl | ensembl V95 |
| <i>Haplochromis burtoni</i>        | ensembl | ensembl V95 |
| <i>Hippocampus comes</i>           | ensembl | ensembl V95 |
| <i>Homo sapiens</i>                | ensembl | ensembl V95 |
| <i>Ictalurus punctatus</i>         | ensembl | ensembl V95 |
| <i>Kryptolebias marmoratus</i>     | ensembl | ensembl V95 |
| <i>Labrus bergylta</i>             | ensembl | ensembl V95 |
| <i>Latimeria chalumnae</i>         | ensembl | ensembl V95 |
| <i>Lepisosteus oculatus</i>        | ensembl | ensembl V95 |
| <i>Loxodonta africana</i>          | ensembl | ensembl V95 |
| <i>Mastacembelus armatus</i>       | ensembl | ensembl V95 |
| <i>Maylandia zebra</i>             | ensembl | ensembl V95 |
| <i>Mola mola</i>                   | ensembl | ensembl V95 |
| <i>Monodelphis domestica</i>       | ensembl | ensembl V95 |
| <i>Monopterus albus</i>            | ensembl | ensembl V95 |
| <i>Mus musculus</i>                | ensembl | ensembl V95 |
| <i>Neolamprologus brichardi</i>    | ensembl | ensembl V95 |

| Species                              | Source  | Data                                                                            |
|--------------------------------------|---------|---------------------------------------------------------------------------------|
| <i>Oreochromis niloticus</i>         | ensembl | ensembl V95                                                                     |
| <i>Oryzias latipes</i>               | ensembl | ensembl V95                                                                     |
| <i>Oryzias melastigma</i>            | ensembl | ensembl V95                                                                     |
| <i>Paramormyrops kingsleyae</i>      | ensembl | ensembl V95                                                                     |
| <i>Petromyzon marinus</i>            | ensembl | ensembl V95                                                                     |
| <i>Poecilia formosa</i>              | ensembl | ensembl V95                                                                     |
| <i>Poecilia latipinna</i>            | ensembl | ensembl V95                                                                     |
| <i>Poecilia mexicana</i>             | ensembl | ensembl V95                                                                     |
| <i>Poecilia reticulata</i>           | ensembl | ensembl V95                                                                     |
| <i>Pundamilia nyererei</i>           | ensembl | ensembl V95                                                                     |
| <i>Pygocentrus nattereri</i>         | ensembl | ensembl V95                                                                     |
| <i>Scleropages formosus</i>          | ensembl | ensembl V95                                                                     |
| <i>Scophthalmus maximus</i>          | ensembl | ensembl V95                                                                     |
| <i>Seriola dumerili</i>              | ensembl | ensembl V95                                                                     |
| <i>Seriola lalandi dorsalis</i>      | ensembl | ensembl V95                                                                     |
| <i>Stegastes partitus</i>            | ensembl | ensembl V95                                                                     |
| <i>Takifugu rubripes</i>             | ensembl | ensembl V95                                                                     |
| <i>Tetraodon nigroviridis</i>        | ensembl | ensembl V95                                                                     |
| <i>Xenopus tropicalis</i>            | ensembl | ensembl V95                                                                     |
| <i>Xiphophorus couchianus</i>        | ensembl | ensembl V95                                                                     |
| <i>Xiphophorus maculatus</i>         | ensembl | ensembl V95                                                                     |
| <i>Acanthopagrus schlegelii</i>      | gigaDB  | <a href="http://gigadb.org/dataset/100409">http://gigadb.org/dataset/100409</a> |
| <i>Betta splendens</i>               | gigaDB  | <a href="http://gigadb.org/dataset/100433">http://gigadb.org/dataset/100433</a> |
| <i>Channa argus</i>                  | gigaDB  | <a href="http://gigadb.org/dataset/100279">http://gigadb.org/dataset/100279</a> |
| <i>Eleginops maclovinus</i>          | gigaDB  | <a href="http://gigadb.org/dataset/102163">http://gigadb.org/dataset/102163</a> |
| <i>Hippocampus erectus</i>           | gigaDB  | <a href="http://gigadb.org/dataset/100298">http://gigadb.org/dataset/100298</a> |
| <i>Lateolabrax maculatus</i>         | gigaDB  | <a href="http://gigadb.org/dataset/100458">http://gigadb.org/dataset/100458</a> |
| <i>Protosalanx hyalocranius</i>      | gigaDB  | <a href="http://gigadb.org/dataset/100262">http://gigadb.org/dataset/100262</a> |
| <i>Alligator mississippiensis</i>    | NCBI    | GCF_000281125.3                                                                 |
| <i>Austrofundulus limnaeus</i>       | NCBI    | GCF_001266775.1                                                                 |
| <i>Boleophthalmus pectinirostris</i> | NCBI    | GCF_000788275.1                                                                 |
| <i>Branchiostoma belcheri</i>        | NCBI    | GCF_001625305.1                                                                 |
| <i>Branchiostoma floridae</i>        | NCBI    | GCF_000003815.1                                                                 |
| <i>Carassius auratus</i>             | NCBI    | GCF_003368295.1                                                                 |
| <i>Clupea harengus</i>               | NCBI    | GCF_000966335.1                                                                 |
| <i>Columba livia</i>                 | NCBI    | GCF_000337935.1                                                                 |
| <i>Coturnix japonica</i>             | NCBI    | GCF_001577835.1                                                                 |
| <i>Cyprinus carpio</i>               | NCBI    | GCF_000951615.1                                                                 |
| <i>Electrophorus electricus</i>      | NCBI    | GCF_003665695.1                                                                 |
| <i>Falco peregrinus</i>              | NCBI    | GCF_000337955.1                                                                 |
| <i>Larimichthys crocea</i>           | NCBI    | GCF_000972845.1                                                                 |

| Species                             | Source | Data            |
|-------------------------------------|--------|-----------------|
| <i>Lates calcarifer</i>             | NCBI   | GCF_001640805.1 |
| <i>Nothobranchius furzeri</i>       | NCBI   | GCF_001465895.1 |
| <i>Notothenia coriiceps</i>         | NCBI   | GCF_000735185.1 |
| <i>Oncorhynchus kisutch</i>         | NCBI   | GCF_002021735.1 |
| <i>Oncorhynchus mykiss</i>          | NCBI   | GCF_002163495.1 |
| <i>Oncorhynchus tshawytscha</i>     | NCBI   | GCF_002872995.1 |
| <i>Pangasianodon hypophthalmus</i>  | NCBI   | GCF_003671635.1 |
| <i>Paralichthys olivaceus</i>       | NCBI   | GCF_001970005.1 |
| <i>Rhincodon typus</i>              | NCBI   | GCF_001642345.1 |
| <i>Salmo salar</i>                  | NCBI   | GCF_000233375.1 |
| <i>Salvelinus alpinus</i>           | NCBI   | GCF_002910315.1 |
| <i>Sinocyclocheilus anshuiensis</i> | NCBI   | GCF_001515605.1 |
| <i>Sinocyclocheilus grahami</i>     | NCBI   | GCF_001515645.1 |
| <i>Sinocyclocheilus rhinoceros</i>  | NCBI   | GCF_001515625.1 |
| <i>Xenopus laevis</i>               | NCBI   | GCF_001663975.1 |
| <i>Callorhynchus milii</i>          | NCBI   | GCF_000165045.1 |
| <i>Amia calva</i>                   | NCBI   | GCA_017591415.1 |
| <i>Arapaima gigas</i>               | NCBI   | GCA_007844225.1 |
| <i>Oryzias javanicus</i>            | NCBI   | GCA_003999625.1 |
| <i>Perca flavescens</i>             | NCBI   | GCA_004354835.1 |

**Supplementary Table S1: List of the 101 genome assemblies used in this study.** The dataset contains genomic resources from Ensembl version 95, NCBI and GigaDB.

| Species                             | Cohort            | Order             | Family          | Genes in the atlas | Ohnologs | Ohnolog fraction |
|-------------------------------------|-------------------|-------------------|-----------------|--------------------|----------|------------------|
| <i>Paramormyrops kingsleyae</i>     | Osteoglossomorpha | Osteoglossiformes | Mormyridae      | 20,022             | 6,277    | 31%              |
| <i>Scleropages formosus</i>         | Osteoglossomorpha | Osteoglossiformes | Osteoglossidae  | 19,757             | 6,464    | 33%              |
| <i>Arapaima gigas</i>               | Osteoglossomorpha | Osteoglossiformes | Osteoglossidae  | 17,221             | 5,119    | 30%              |
| <i>Clupea harengus</i>              | Otomorpha         | Clupeiformes      | Clupeidae       | 20,037             | 5,137    | 26%              |
| <i>Sinocyclocheilus anshuiensis</i> | Otomorpha         | Cypriniformes     | Cyprinidae      | 36,445             | 9,185    | 25%              |
| <i>Sinocyclocheilus grahami</i>     | Otomorpha         | Cypriniformes     | Cyprinidae      | 38,342             | 9,762    | 25%              |
| <i>Sinocyclocheilus rhinoceros</i>  | Otomorpha         | Cypriniformes     | Cyprinidae      | 37,522             | 9,408    | 25%              |
| <i>Cyprinus carpio</i>              | Otomorpha         | Cypriniformes     | Cyprinidae      | 38,952             | 9,326    | 24%              |
| <i>Carassius auratus</i>            | Otomorpha         | Cypriniformes     | Cyprinidae      | 43,853             | 11,014   | 25%              |
| <i>Danio rerio</i>                  | Otomorpha         | Cypriniformes     | Danionidae      | 21,496             | 5,362    | 25%              |
| <i>Astyanax mexicanus</i>           | Otomorpha         | Characiformes     | Characidae      | 20,943             | 5,282    | 25%              |
| <i>Pygocentrus nattereri</i>        | Otomorpha         | Characiformes     | Serrasalminidae | 21,456             | 5,450    | 25%              |
| <i>Ictalurus punctatus</i>          | Otomorpha         | Siluriformes      | Ictaluridae     | 19,725             | 4,414    | 22%              |
| <i>Pangasianodon hypophthalmus</i>  | Otomorpha         | Siluriformes      | Pangasiidae     | 18,783             | 4,179    | 22%              |
| <i>Electrophorus electricus</i>     | Otomorpha         | Gymnotiformes     | Gymnotidae      | 19,059             | 4,639    | 24%              |
| <i>Esox lucius</i>                  | Euteleosteomorpha | Esociformes       | Esocidae        | 20,069             | 5,491    | 27%              |
| <i>Salmo salar</i>                  | Euteleosteomorpha | Salmoniformes     | Salmonidae      | 34,617             | 9,769    | 28%              |
| <i>Salvelinus alpinus</i>           | Euteleosteomorpha | Salmoniformes     | Salmonidae      | 29,053             | 7,175    | 25%              |
| <i>Oncorhynchus mykiss</i>          | Euteleosteomorpha | Salmoniformes     | Salmonidae      | 30,196             | 7,989    | 26%              |
| <i>Oncorhynchus kisutch</i>         | Euteleosteomorpha | Salmoniformes     | Salmonidae      | 29,352             | 7,489    | 26%              |
| <i>Oncorhynchus tshawytscha</i>     | Euteleosteomorpha | Salmoniformes     | Salmonidae      | 32,997             | 8,316    | 25%              |
| <i>Protosalanx hyalocranius</i>     | Euteleosteomorpha | Osmeriformes      | Salangidae      | 15,729             | 3,268    | 21%              |
| <i>Gadus morhua</i>                 | Euteleosteomorpha | Gadiformes        | Gadidae         | 17,167             | 3,257    | 19%              |

| Species                              | Cohort            | Order              | Family          | Genes in the atlas | Ohnologs | Ohnolog fraction |
|--------------------------------------|-------------------|--------------------|-----------------|--------------------|----------|------------------|
| <i>Hippocampus comes</i>             | Euteleosteomorpha | Syngnathiformes    | Syngnathidae    | 18,089             | 3,705    | 20%              |
| <i>Hippocampus erectus</i>           | Euteleosteomorpha | Syngnathiformes    | Syngnathidae    | 14,727             | 2,567    | 17%              |
| <i>Boleophthalmus pectinirostris</i> | Euteleosteomorpha | Gobiiformes        | Oxudercidae     | 18,598             | 3,993    | 21%              |
| <i>Channa argus</i>                  | Euteleosteomorpha | Anabantiformes     | Channidae       | 16,985             | 3,558    | 21%              |
| <i>Anabas testudineus</i>            | Euteleosteomorpha | Anabantiformes     | Anabantidae     | 20,561             | 4,787    | 23%              |
| <i>Betta splendens</i>               | Euteleosteomorpha | Anabantiformes     | Osphronemidae   | 19,023             | 4,375    | 23%              |
| <i>Monopterus albus</i>              | Euteleosteomorpha | Synbranchiformes   | Synbranchidae   | 19,031             | 4,102    | 22%              |
| <i>Mastacembelus armatus</i>         | Euteleosteomorpha | Synbranchiformes   | Mastacembelidae | 19,776             | 4,448    | 22%              |
| <i>Lates calcarifer</i>              | Euteleosteomorpha | Carangaria         | Latidae         | 21,349             | 4,997    | 23%              |
| <i>Paralichthys olivaceus</i>        | Euteleosteomorpha | Pleuronectiformes  | Paralichthyidae | 19,690             | 4,528    | 23%              |
| <i>Scophthalmus maximus</i>          | Euteleosteomorpha | Pleuronectiformes  | Scophthalmidae  | 18,658             | 4,302    | 23%              |
| <i>Cynoglossus semilaevis</i>        | Euteleosteomorpha | Pleuronectiformes  | Cynoglossidae   | 18,814             | 4,001    | 21%              |
| <i>Seriola lalandi dorsalis</i>      | Euteleosteomorpha | Carangiformes      | Carangidae      | 20,996             | 4,969    | 24%              |
| <i>Seriola dumerili</i>              | Euteleosteomorpha | Carangiformes      | Carangidae      | 20,009             | 4,698    | 23%              |
| <i>Cyprinodon variegatus</i>         | Euteleosteomorpha | Cyprinodontiformes | Cyprinodontidae | 19,635             | 4,231    | 22%              |
| <i>Fundulus heteroclitus</i>         | Euteleosteomorpha | Cyprinodontiformes | Fundulidae      | 19,471             | 4,257    | 22%              |
| <i>Xiphophorus maculatus</i>         | Euteleosteomorpha | Cyprinodontiformes | Poeciliidae     | 19,912             | 4,421    | 22%              |
| <i>Xiphophorus couchianus</i>        | Euteleosteomorpha | Cyprinodontiformes | Poeciliidae     | 17,278             | 3,501    | 20%              |
| <i>Gambusia affinis</i>              | Euteleosteomorpha | Cyprinodontiformes | Poeciliidae     | 18,936             | 4,191    | 22%              |
| <i>Poecilia formosa</i>              | Euteleosteomorpha | Cyprinodontiformes | Poeciliidae     | 20,316             | 4,519    | 22%              |
| <i>Poecilia atipinna</i>             | Euteleosteomorpha | Cyprinodontiformes | Poeciliidae     | 20,189             | 4,397    | 22%              |
| <i>Poecilia mexicana</i>             | Euteleosteomorpha | Cyprinodontiformes | Poeciliidae     | 20,470             | 4,477    | 22%              |
| <i>Poecilia reticulata</i>           | Euteleosteomorpha | Cyprinodontiformes | Poeciliidae     | 19,441             | 4,273    | 22%              |
| <i>Kryptolebias marmoratus</i>       | Euteleosteomorpha | Cyprinodontiformes | Rivulidae       | 19,079             | 4,081    | 21%              |

| Species                            | Cohort            | Order              | Family           | Genes in the atlas | Ohnologs | Ohnolog fraction |
|------------------------------------|-------------------|--------------------|------------------|--------------------|----------|------------------|
| <i>Austrofundulus limnaeus</i>     | Euteleosteomorpha | Cyprinodontiformes | Rivulidae        | 17,243             | 3,445    | 20%              |
| <i>Nothobranchius furzeri</i>      | Euteleosteomorpha | Cyprinodontiformes | Nothobranchiidae | 18,308             | 3,800    | 21%              |
| <i>Oryzias javanicus</i>           | Euteleosteomorpha | Beloniformes       | Adrianichthyidae | 17,511             | 3,513    | 20%              |
| <i>Oryzias latipes</i>             | Euteleosteomorpha | Beloniformes       | Adrianichthyidae | 20,230             | 4,220    | 21%              |
| <i>Oryzias melastigma</i>          | Euteleosteomorpha | Beloniformes       | Adrianichthyidae | 19,983             | 4,292    | 21%              |
| <i>Oreochromis niloticus</i>       | Euteleosteomorpha | Cichliformes       | Cichlidae        | 18,512             | 4,073    | 22%              |
| <i>Haplochromis burtoni</i>        | Euteleosteomorpha | Cichliformes       | Cichlidae        | 19,880             | 4,519    | 23%              |
| <i>Pundamilia nyererei</i>         | Euteleosteomorpha | Cichliformes       | Cichlidae        | 19,877             | 4,526    | 23%              |
| <i>Maylandia zebra</i>             | Euteleosteomorpha | Cichliformes       | Cichlidae        | 22,113             | 4,969    | 22%              |
| <i>Astatotilapia calliptera</i>    | Euteleosteomorpha | Cichliformes       | Cichlidae        | 21,333             | 4,736    | 22%              |
| <i>Neolamprologus brichardi</i>    | Euteleosteomorpha | Cichliformes       | Cichlidae        | 19,861             | 4,467    | 22%              |
| <i>Amphilophus citrinellus</i>     | Euteleosteomorpha | Cichliformes       | Cichlidae        | 19,981             | 4,531    | 23%              |
| <i>Amphiprion ocellaris</i>        | Euteleosteomorpha | Ovalentaria        | Pomacentridae    | 20,281             | 4,593    | 23%              |
| <i>Amphiprion percula</i>          | Euteleosteomorpha | Ovalentaria        | Pomacentridae    | 20,274             | 4,687    | 23%              |
| <i>Acanthochromis polyacanthus</i> | Euteleosteomorpha | Ovalentaria        | Pomacentridae    | 20,523             | 4,696    | 23%              |
| <i>Stegastes partitus</i>          | Euteleosteomorpha | Ovalentaria        | Pomacentridae    | 19,908             | 4,651    | 23%              |
| <i>Labrus bergylta</i>             | Euteleosteomorpha | Labriformes        | Labridae         | 22,332             | 5,065    | 23%              |
| <i>Notothenia coriiceps</i>        | Euteleosteomorpha | Notothenioidei     | Nototheniidae    | 19,343             | 3,942    | 20%              |
| <i>Eleginops maclovinus</i>        | Euteleosteomorpha | Notothenioidei     | Eleginopidae     | 16,843             | 3,468    | 21%              |
| <i>Gasterosteus aculeatus</i>      | Euteleosteomorpha | Gasterosteoidei    | Gasterosteidae   | 17,979             | 3,722    | 21%              |
| <i>Perca flavescens</i>            | Euteleosteomorpha | Percoidei          | Percidae         | 18,541             | 3,933    | 21%              |
| <i>Lateolabrax maculatus</i>       | Euteleosteomorpha | Acropomatiformes   | Lateolabracidae  | 16,876             | 3,164    | 19%              |
| <i>Takifugu rubripes</i>           | Euteleosteomorpha | Tetraodontiformes  | Tetraodontidae   | 18,086             | 3,646    | 20%              |
| <i>Tetraodon nigroviridis</i>      | Euteleosteomorpha | Tetraodontiformes  | Tetraodontidae   | 17,619             | 3,625    | 21%              |



| Species                         | Cohort           | Order             | Family     | Genes in the atlas | Ohnologs | Ohnolog fraction |
|---------------------------------|------------------|-------------------|------------|--------------------|----------|------------------|
| <i>Mola mola</i>                | Euteleostomorpha | Tetraodontiformes | Molidae    | 18,631             | 4,197    | 23%              |
| <i>Acanthopagrus schlegelii</i> | Euteleostomorpha | Eupercaria        | Sparidae   | 16,040             | 3,220    | 20%              |
| <i>Larimichthys crocea</i>      | Euteleostomorpha | Acanthuriformes   | Sciaenidae | 20,033             | 4,712    | 24%              |

**Supplementary Table S2: Duplicate gene (ohnolog) retention across the 74 teleost genomes.** Ohnolog fractions are given as the number of ohnologs identified in the comparative atlas divided by the total number of genes annotated in the atlas for that species. Species are ordered based on their phylogenetic position. Cohort, Order and Family are given according to the DeepFin classification.

| GO ID      | GO name                                          | Enrichment | P-value               | FDR                   | Genes                                                                                                                                                                                                                                                              |
|------------|--------------------------------------------------|------------|-----------------------|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| GO:0051338 | regulation of transferase activity               | 5.724      | $1.00 \times 10^{-6}$ | $2.84 \times 10^{-4}$ | ENSDARG00000038524, ENSDARG00000060316, ENSDARG00000038095, ENSDARG00000077226, ENSDARG00000005941, ENSDARG00000038139, ENSDARG00000086778, ENSDARG00000063583, ENSDARG00000077177, ENSDARG00000015803, ENSDARG00000070360, ENSDARG00000070404                     |
| GO:0051174 | regulation of phosphorus metabolic process       | 3.524      | $6.80 \times 10^{-5}$ | 0.0097                | ENSDARG00000038524, ENSDARG00000060316, ENSDARG00000038095, ENSDARG00000077226, ENSDARG00000005941, ENSDARG00000038139, ENSDARG00000086778, ENSDARG00000063583, ENSDARG00000077177, ENSDARG00000015803, ENSDARG00000070360, ENSDARG00000070404, ENSDARG00000062277 |
| GO:1902531 | regulation of intracellular signal transduction  | 3.286      | $1.38 \times 10^{-4}$ | 0.0131                | ENSDARG00000060316, ENSDARG00000038095, ENSDARG00000089873, ENSDARG00000077226, ENSDARG00000038139, ENSDARG00000086778, ENSDARG00000063583, ENSDARG00000077177, ENSDARG00000014465, ENSDARG00000015803, ENSDARG00000003313, ENSDARG00000002816, ENSDARG00000075054 |
| GO:0007167 | enzyme linked receptor protein signaling pathway | 3.686      | $3.52 \times 10^{-4}$ | 0.0250                | ENSDARG00000038524, ENSDARG00000060316, ENSDARG00000038095, ENSDARG00000031751, ENSDARG00000005941, ENSDARG000000104100, ENSDARG00000010785, ENSDARG00000038139, ENSDARG00000086778, ENSDARG00000038067                                                            |

**Supplementary Table S3: Gene ontology terms enriched in LORe interspersed.** Biological process gene ontology enrichment results for the set of n=215 LORe interspersed zebrafish genes.

| KEGG ID  | KEGG pathway               | Enrichment | p-value               | FDR    | genes                                                                                                                                                        |
|----------|----------------------------|------------|-----------------------|--------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| dre04350 | TGF-beta signaling pathway | 5.465      | $2.63 \times 10^{-4}$ | 0.0426 | ENSDARG00000104100,<br>ENSDARG00000102742,<br>ENSDARG00000010785,<br>ENSDARG00000100666,<br>ENSDARG00000061108,<br>ENSDARG00000038067,<br>ENSDARG00000075226 |

**Supplementary Table S4: KEGG pathway enriched in LORe interspersed.** KEGG pathway enrichment results for the set of n=215 LORe interspersed zebrafish genes.

| Ancestral chromosome                               | 1    | 2                                                  | 3                                                  | 4                                                  | 5    | 6                                                 | 7                                                  | 8                                                  | 9                                                 | 10   | 11                                                 | 12   | 13                                                 |
|----------------------------------------------------|------|----------------------------------------------------|----------------------------------------------------|----------------------------------------------------|------|---------------------------------------------------|----------------------------------------------------|----------------------------------------------------|---------------------------------------------------|------|----------------------------------------------------|------|----------------------------------------------------|
| Total pre-duplication gene families <sup>(1)</sup> | 1526 | <b>1769</b>                                        | <b>1440</b>                                        | <b>1144</b>                                        | 1954 | <b>682</b>                                        | <b>664</b>                                         | <b>1498</b>                                        | <b>1903</b>                                       | 1320 | <b>855</b>                                         | 1975 | <b>2502</b>                                        |
| Genes retained on copy 'a' <sup>(2)</sup>          | 961  | <b>996</b>                                         | <b>1057</b>                                        | <b>883</b>                                         | 1326 | <b>480</b>                                        | <b>523</b>                                         | <b>1054</b>                                        | <b>1264</b>                                       | 870  | <b>650</b>                                         | 1280 | <b>1815</b>                                        |
| Genes retained on copy 'b' <sup>(3)</sup>          | 966  | <b>1234</b>                                        | <b>827</b>                                         | <b>570</b>                                         | 1317 | <b>433</b>                                        | <b>371</b>                                         | <b>843</b>                                         | <b>1143</b>                                       | 888  | <b>450</b>                                         | 1287 | <b>1450</b>                                        |
| Fisher's test p-value <sup>(4)</sup>               | 0.88 | <b>2.87</b><br><b><math>\times 10^{-16}</math></b> | <b>7.25</b><br><b><math>\times 10^{-19}</math></b> | <b>1.91</b><br><b><math>\times 10^{-41}</math></b> | 0.88 | <b>1.17</b><br><b><math>\times 10^{-2}</math></b> | <b>1.41</b><br><b><math>\times 10^{-18}</math></b> | <b>2.74</b><br><b><math>\times 10^{-15}</math></b> | <b>8.83</b><br><b><math>\times 10^{-5}</math></b> | 0.63 | <b>2.06</b><br><b><math>\times 10^{-23}</math></b> | 0.88 | <b>1.51</b><br><b><math>\times 10^{-26}</math></b> |

**Supplementary Table S5: Total gene retention on homeologous chromosomes in the post-duplication teleost ancestor Osteoglossocephalai.** The table shows the total number of genes retained on each ancestrally duplicated chromosome copy in Osteoglossocephalai. Homeologs with a significant bias in genes retention between the two copies are shown in bold (p-value < 0.05, Fisher's test corrected for multiple testing using the Benjamini-Hochberg procedure).

<sup>(1)</sup> Total Osteoglossocephalai gene families for each pre-duplication chromosome

<sup>(2), (3)</sup> Gene copies retained on chromosomes 'a' and 'b' respectively

<sup>(4)</sup> Fisher's test p-value for biased gene retention, corrected for multiple testing with the Benjamini-Hochberg procedure.

We note that Fisher's test assumes independence between observations, which may not be verified here if several genes are affected by a single large deletion. The window-based approach in Figure 3B attempts to limit this potential bias.

| Ancestral chromosome                                                   | 1     | 2                                                  | 3                                                  | 4                                                  | 5     | 6     | 7                                                  | 8                                                  | 9                                                 | 10    | 11                                                 | 12    | 13                                                 |
|------------------------------------------------------------------------|-------|----------------------------------------------------|----------------------------------------------------|----------------------------------------------------|-------|-------|----------------------------------------------------|----------------------------------------------------|---------------------------------------------------|-------|----------------------------------------------------|-------|----------------------------------------------------|
| Total pre-duplication gene families <sup>(1)</sup>                     | 1277  | <b>1519</b>                                        | <b>1143</b>                                        | <b>955</b>                                         | 1615  | 584   | <b>618</b>                                         | <b>1217</b>                                        | 1618                                              | 1098  | <b>725</b>                                         | 1661  | <b>2099</b>                                        |
| Genes retained on copy 'a' <sup>(2)</sup>                              | 718   | <b>728</b>                                         | <b>775</b>                                         | <b>706</b>                                         | 972   | 349   | <b>440</b>                                         | <b>808</b>                                         | 963                                               | 640   | <b>509</b>                                         | 988   | <b>1423</b>                                        |
| Genes retained on copy 'b' <sup>(3)</sup>                              | 697   | <b>974</b>                                         | <b>490</b>                                         | <b>368</b>                                         | 920   | 304   | <b>224</b>                                         | <b>525</b>                                         | 828                                               | 577   | <b>294</b>                                         | 890   | <b>923</b>                                         |
| Non-annotated genes <sup>(4)</sup>                                     | 110   | <b>173</b>                                         | <b>119</b>                                         | <b>80</b>                                          | 169   | 282   | <b>156</b>                                         | <b>112</b>                                         | 134                                               | 96    | <b>67</b>                                          | 139   | <b>185</b>                                         |
| Fisher's test p-value <sup>(5)</sup>                                   | 0.426 | <b>5.21</b><br><b><math>\times 10^{-19}</math></b> | <b>9.71</b><br><b><math>\times 10^{-33}</math></b> | <b>1.30</b><br><b><math>\times 10^{-54}</math></b> | 0.074 | 0.011 | <b>1.17</b><br><b><math>\times 10^{-34}</math></b> | <b>2.19</b><br><b><math>\times 10^{-30}</math></b> | <b>3.45</b><br><b><math>\times 10^{-6}</math></b> | 0.010 | <b>9.61</b><br><b><math>\times 10^{-30}</math></b> | 0.001 | <b>5.48</b><br><b><math>\times 10^{-54}</math></b> |
| Fisher's test p-value corrected for non-annotated genes <sup>(6)</sup> |       | <b>0.017</b>                                       | <b>8.38</b><br><b><math>\times 10^{-11}</math></b> | <b>1.79</b><br><b><math>\times 10^{-29}</math></b> |       |       | <b>0.004</b>                                       | <b>8.38</b><br><b><math>\times 10^{-11}</math></b> | 1.00                                              |       | <b>3.41</b><br><b><math>\times 10^{-13}</math></b> | 0.204 | <b>3.39</b><br><b><math>\times 10^{-20}</math></b> |

**Supplementary Table S6: Conservative estimate of total gene retention on homeologous chromosomes in Medaka.** The table shows the total number of genes retained on each ancestrally duplicated chromosome copy in medaka, using conservative homeologous copy assignments to medaka genes for which the pre-duplication chromosome is known (Nakatani and McLysaght, 2017) but not the post-duplication chromosome. Ancestrally duplicated chromosomes whose retention remained significantly biased towards the same copy after assigning non-annotated medaka genes to the copy with lower retention are shown in bold (p-value < 0.05, Fisher's test corrected for multiple testing using the Benjamini-Hochberg procedure).

<sup>(1)</sup> Total medaka gene families for each pre-duplication chromosome.

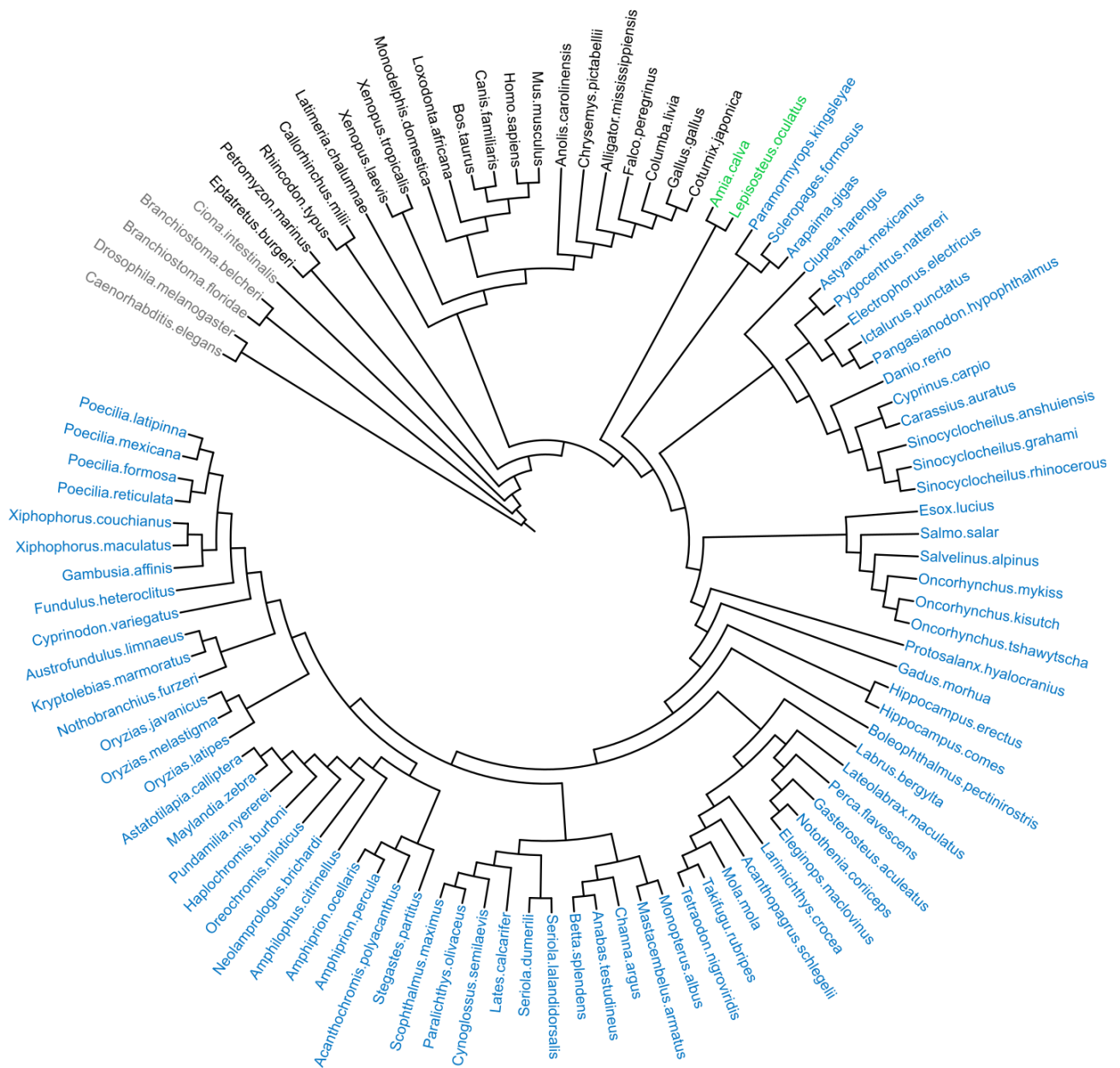
<sup>(2), (3)</sup> Medaka gene copies retained on chromosomes 'a' and 'b' respectively.

<sup>(4)</sup> Non-annotated medaka genes: genes where the post-duplication chromosome is unknown.

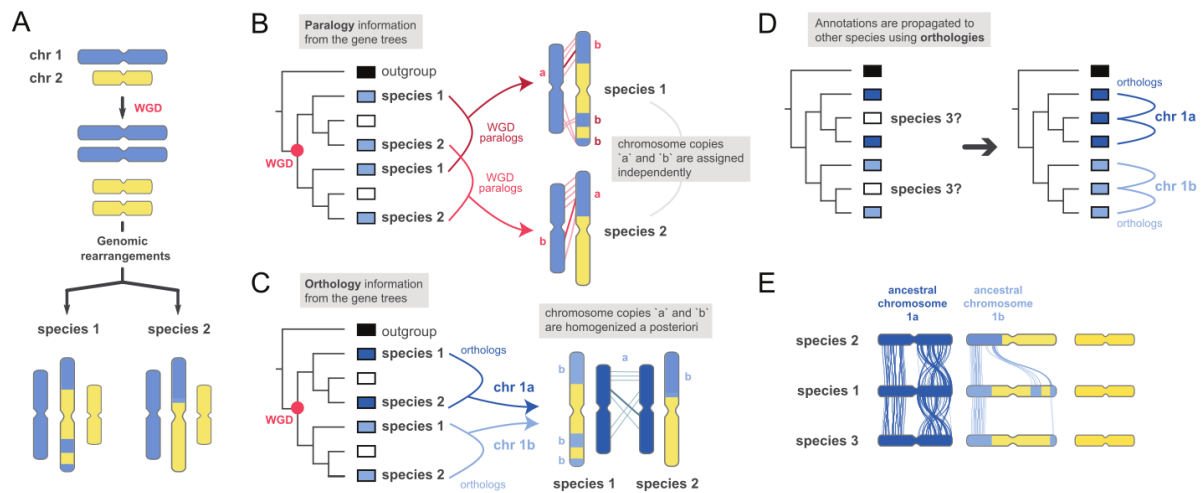
<sup>(5)</sup> Fisher's test p-value for biased gene retention, corrected for multiple testing with the Benjamini-Hochberg procedure, but without accounting for non-annotated genes.

<sup>(6)</sup> Fisher's test p-value for biased gene retention, corrected for multiple testing with the Benjamini-Hochberg procedure and corrected for non-annotated genes. Cells in bold show chromosomes for which the correction did not affect the retention imbalance. Cells are left blank for unbiased chromosomes and for chromosomes where accounting for non-annotated genes changes the direction of the imbalance.

We note that Fisher's test assumes independence between observations, which may not be verified here if several genes are affected by a single large deletion. The window-based approach in Figure 3B attempts to limit this potential bias. We note that Fisher's test assumes independence between observations, which may not be verified here if several genes are affected by a single large deletion. The window-based approach in Figure 3B attempts to limit this potential bias.

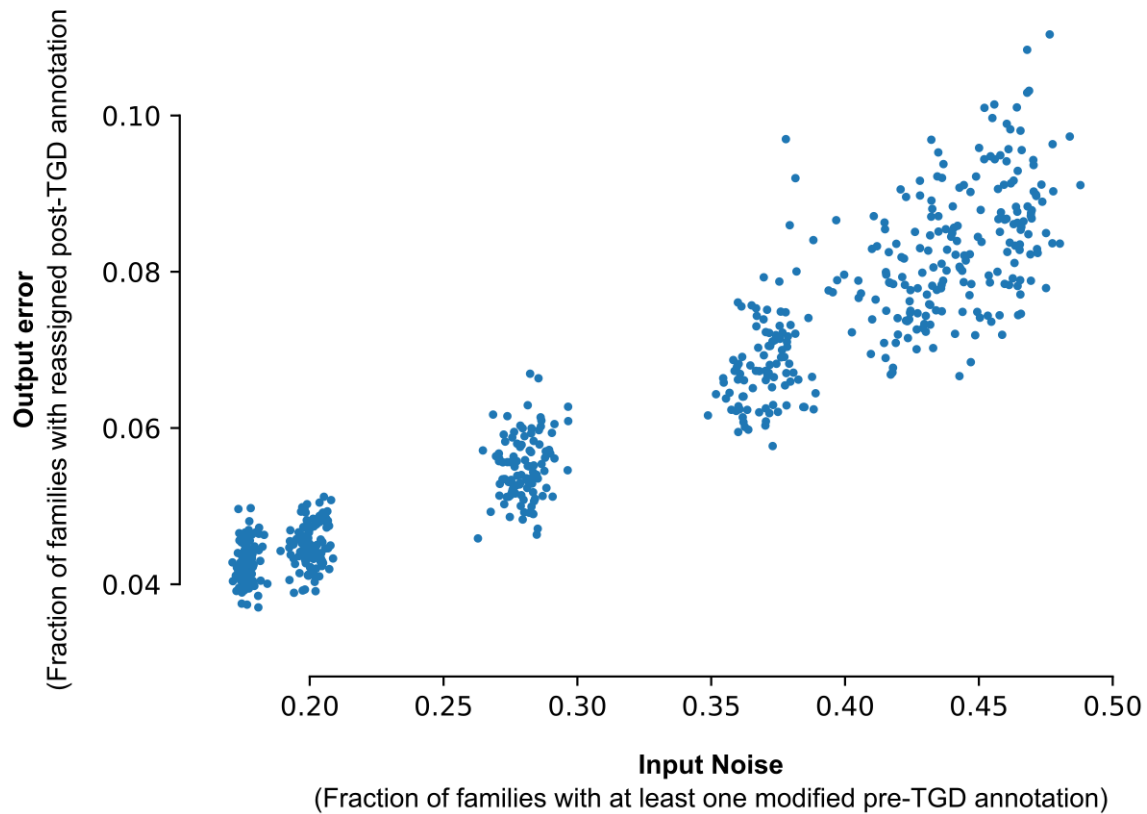


**Supplementary Figure S1: Species tree of the 101 genomes used in this study.** The 74 teleost species are shown in blue, non-duplicated fish outgroups in green, other vertebrates in black and non-vertebrate outgroups in grey. Branch lengths are not to scale.

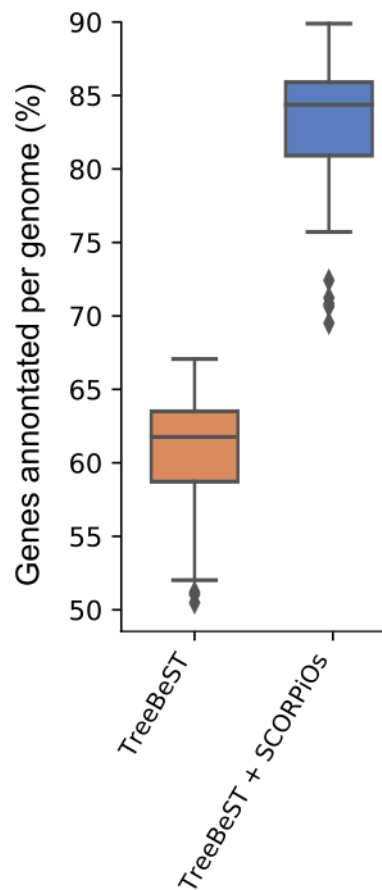


**Supplementary Figure S2: Illustration of the key steps in the comparative atlas workflow. A.** Schematic representation of karyotype histories for two species with a common whole-genome duplication followed by a chromosomal fusion event. **B.** Identification of WGD-duplicated regions using paralogy relationships inferred from gene trees (in red). This identification is performed in each species independently. **C.** Identification of orthologous regions across species inferred from gene trees (in blue). This information is used to homogenize 'a' and 'b' ancestral chromosome assignments across species. **D.** Propagation of duplicated regions annotations to a non-reference species. Here annotation from reference species 1 and 2 are propagated to species 3 through gene orthologies. **E.** Schematic representation of the comparative atlas. Dark blue and light blue regions correspond to WGD-inherited paralogy regions across all species. Links represent orthologous genes.

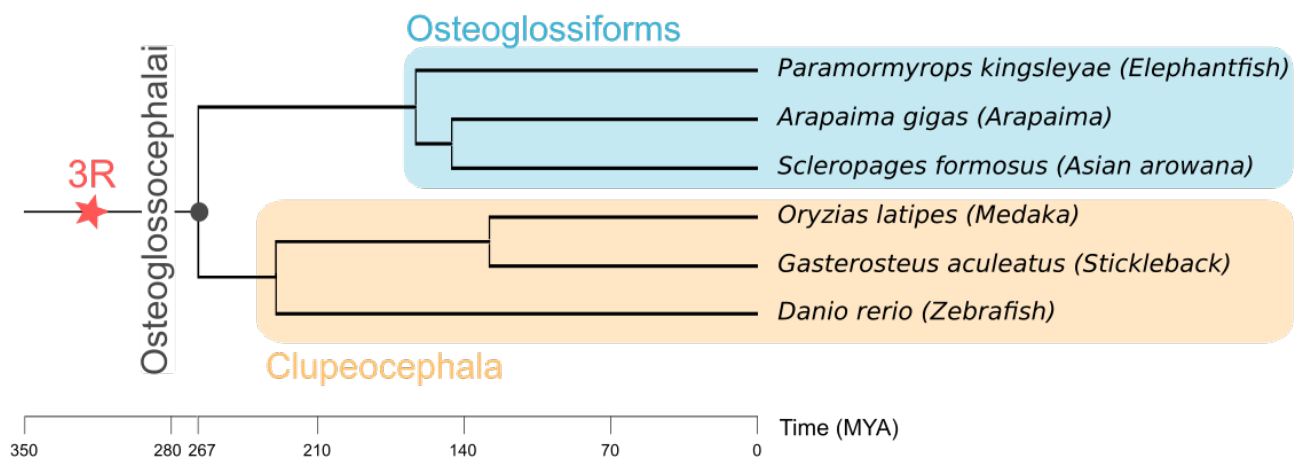




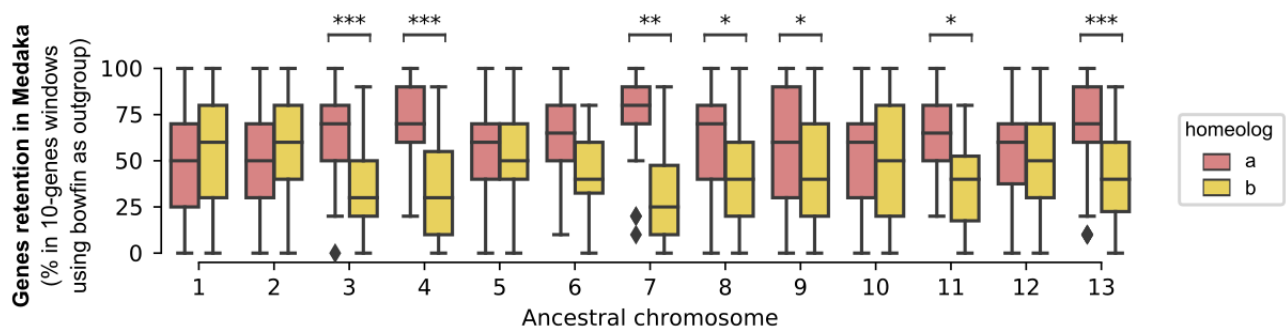
**Supplementary Figure S3: Noise robustness of the comparative atlas.** Proportion of gene families with post-TGD ancestral chromosome reassignments for different input noise settings (Methods). The input noise represents the proportion of families with at least one reference gene having a modified pre-TGD chromosome annotation due to an ancestral chromosome boundary shift in the simulation. The output error represents the fraction of gene families consequently assigned to a different post-TGD ancestral chromosome in the simulation. In the comparative atlas workflow, the majority vote procedure ensures that the majority of these individual errors in inputs are not propagated to the output.



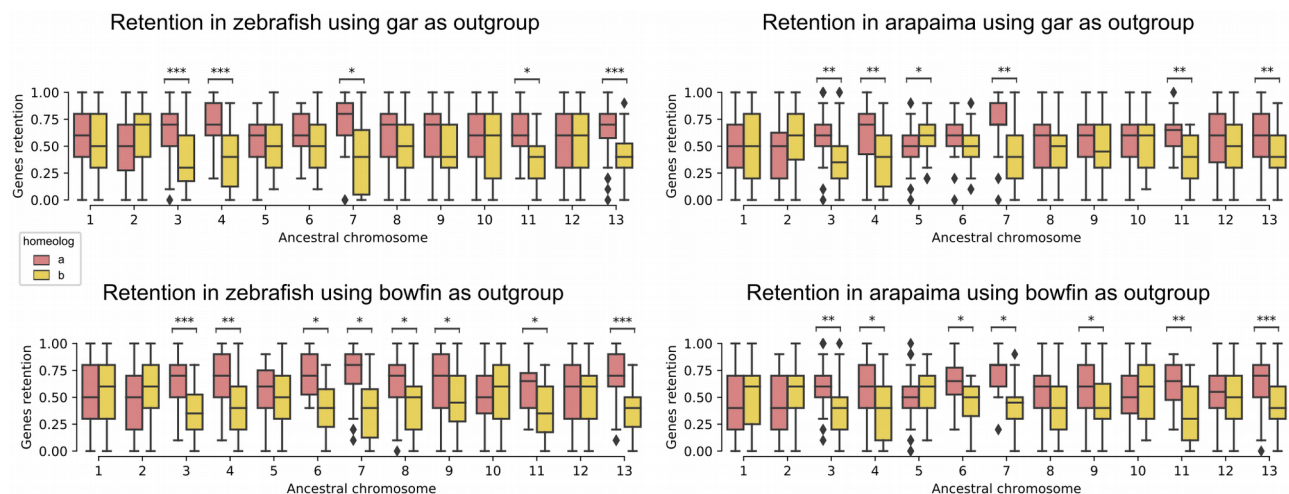
**Supplementary Figure S4: Impact of SCORPiOs on the establishment of the comparative atlas.** Distribution of the proportion of extant genes assigned to an ancestral chromosome in the comparative atlas before (orange) and after (blue) SCORPiOs correction across species. Each boxplot summarizes the distribution of 74 points (one per genome).



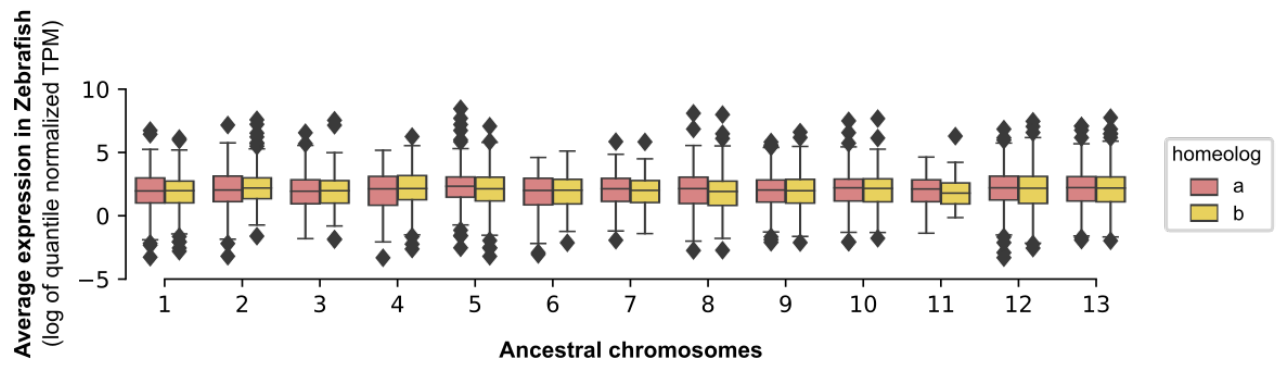
**Supplementary Figure S5: Species tree of the 6 teleosts for the rediploidisation timing analysis.** Divergence times were extracted from TimeTree (Kumar et al. 2017). The teleost duplication is indicated by a red star.



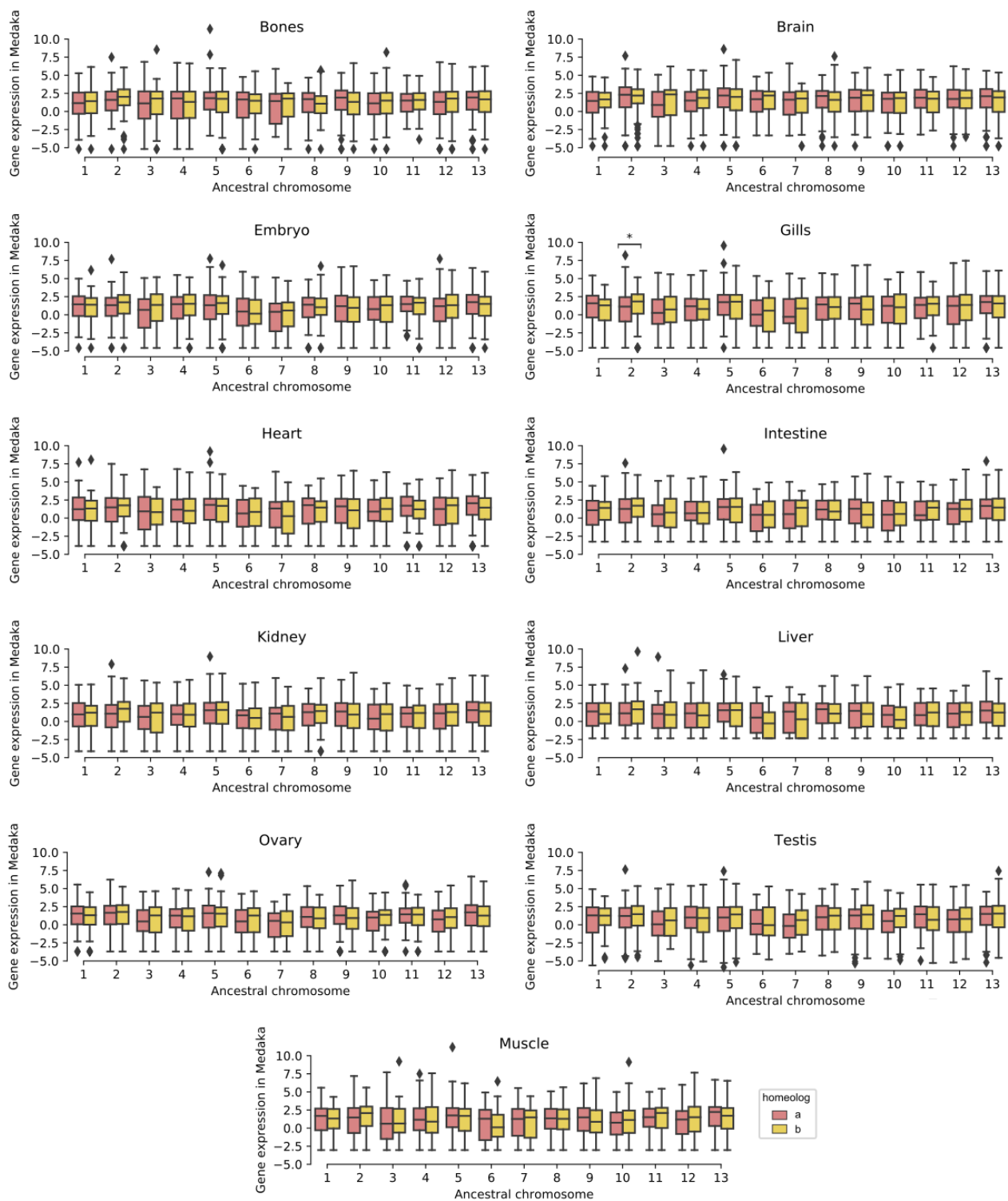
**Supplementary Figure S6: Gene retention on anciently duplicated chromosome copies in medaka, using the bowfin genome as a proxy for ancestral gene order.** Ancestral chromosomes with a significant bias in gene retention on one of the two copies are highlighted (\*\* p-values < 0.01, \* p-values < 0.05, Wilcoxon paired tests corrected for multiple testing using the Benjamini-Hochberg procedure).



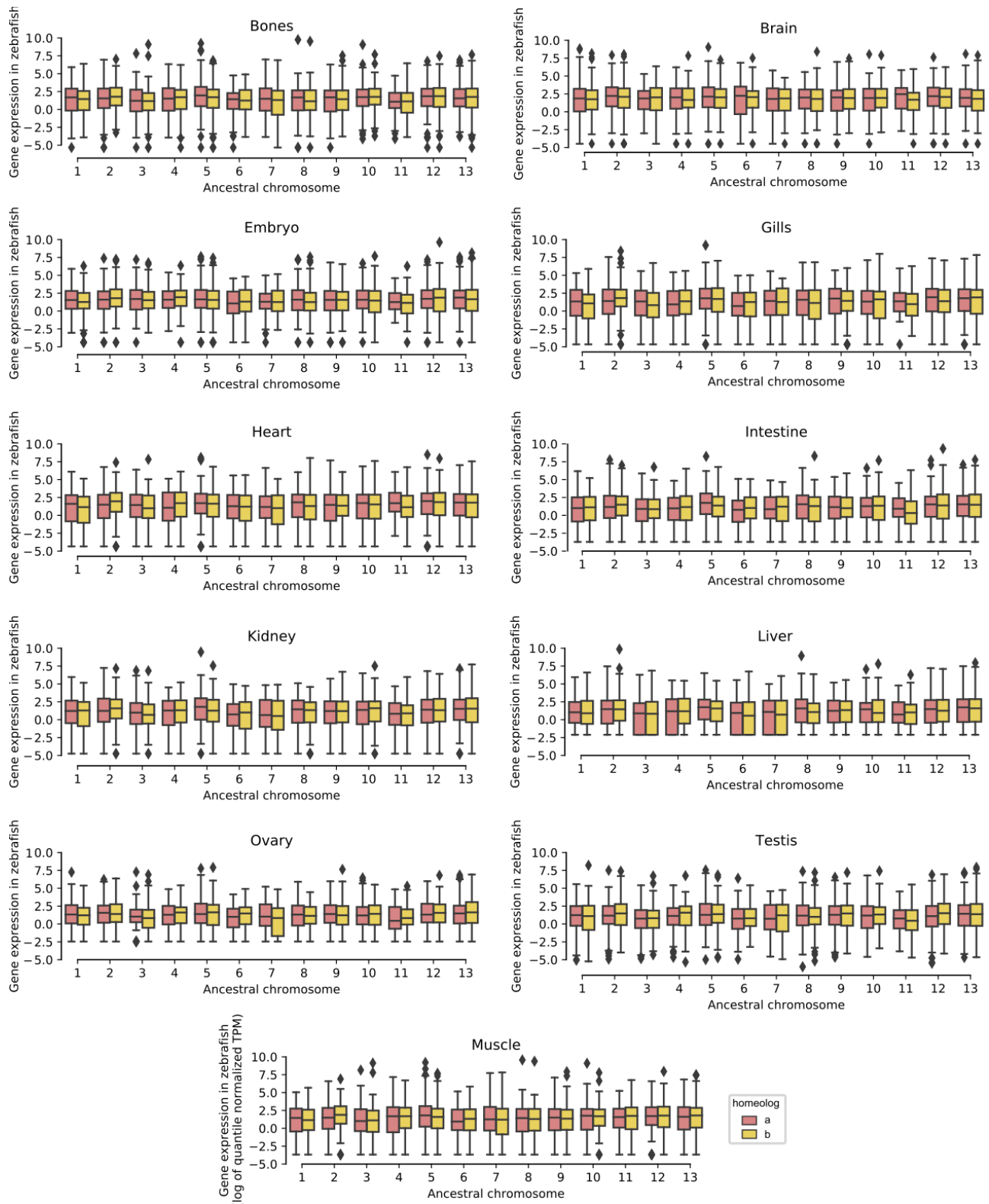
**Supplementary Figure S7: Gene retention on anciently duplicated chromosome copies in zebrafish and arapaima, using the bowfin and the spotted gar genomes as proxies for ancestral gene order.** Ancestral chromosomes with a significant bias in gene retention on one of the two copies are highlighted (\*\*\*) p-values < 0.001, \*\* p-values < 0.01, \* p-values < 0.05, Wilcoxon paired tests corrected for multiple testing using the Benjamini-Hochberg procedure).



**Supplementary Figure S8: Zebrafish ohnologs gene expression across homeologs, averaged across tissues.** There is no significant difference in expression between genes of anciently duplicated chromosome copies 'a' and 'b' (Wilcoxon paired tests corrected for multiple testing using the Benjamini-Hochberg procedure, at  $\alpha=0.05$ ).

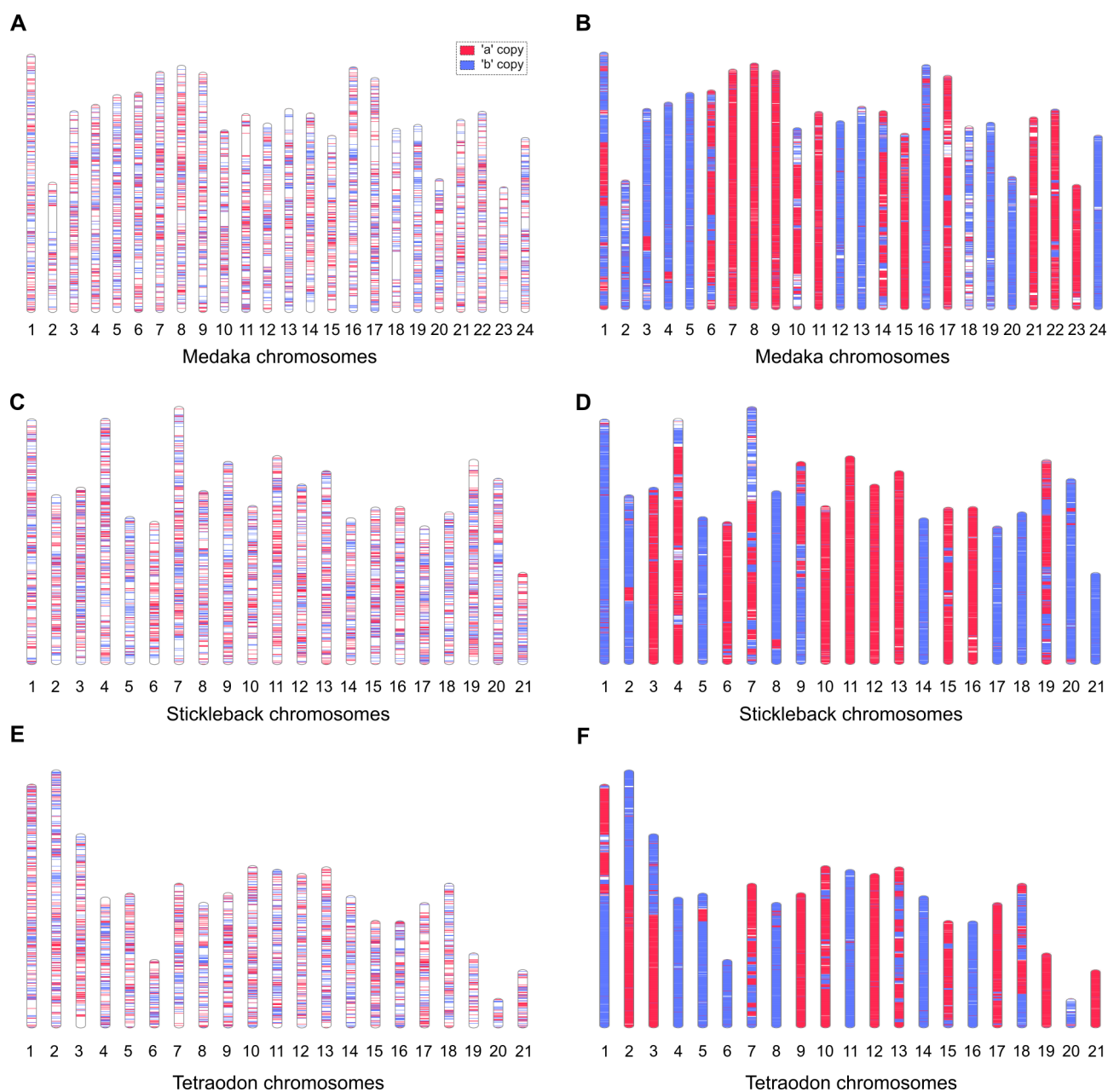


**Supplementary Figure S9: Medaka ohnologs gene expression across homeologs in 11 tissues.** Only for anciently duplicated chromosome 2, in gills, is ohnologous gene expression significantly higher for genes on homeologous copy 'b' (\* p-value < 0.05, Wilcoxon paired tests corrected for multiple testing using the Benjamini-Hochberg procedure).



**Supplementary Figure S10: Zebrafish ohnologs gene expression across homeologs in 11 tissues.** There is no significant difference in expression between genes of anciently duplicated chromosome copies 'a' and 'b', in any tissue (Wilcoxon paired tests corrected for multiple testing using the Benjamini-Hochberg procedure, at  $\alpha=0.05$ ).





**Supplementary Figure S11: Karyotypes of medaka, stickleback and tetraodon with 'a' and 'b' gene copies annotations.** **A.** Visualization of medaka 'a' and 'b' copies from ensembl (transfer of zebrafish ZFIN names) on the genome. **B.** Full re-annotation of medaka 'a' and 'b' copies using the comparative atlas. **C.** Stickleback gene copies, as in A. **D.** Re-annotation of stickleback gene copies, as in B. **E.** Tetraodon gene copies, as in A. **F.** Re-annotation of tetraodon gene copies, as in B.