



**HAL**  
open science

# Genes encoding teleost orthologs of human haploinsufficient and monoallelically expressed genes remain in duplicate more frequently than the whole genome

Floriane Picolo, Anna Grandchamp, Benoit Piegu, Antoine D. Rolland, Reiner A. Veitia, Philippe Monget

## ► To cite this version:

Floriane Picolo, Anna Grandchamp, Benoit Piegu, Antoine D. Rolland, Reiner A. Veitia, et al.. Genes encoding teleost orthologs of human haploinsufficient and monoallelically expressed genes remain in duplicate more frequently than the whole genome. *International journal of genomics*, 2021, 2021, pp.1-6. 10.1155/2021/9028667 . hal-03341375

**HAL Id: hal-03341375**

**<https://hal.science/hal-03341375>**

Submitted on 10 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## Research Article

# Genes Encoding Teleost Orthologs of Human Haploinsufficient and Monoallelically Expressed Genes Remain in Duplicate More Frequently Than the Whole Genome

Floriane Picolo,<sup>1</sup> Anna Grandchamp,<sup>1</sup> Benoît Piégu,<sup>1</sup> Antoine D. Rolland ,<sup>2</sup>  
Reiner A. Veitia,<sup>3,4,5</sup> and Philippe Monget <sup>1</sup>

<sup>1</sup>PRC, UMR85, INRAE, CNRS, IFCE, Université de Tours, F-37380 Nouzilly, France

<sup>2</sup>Univ Rennes, Inserm, EHESP, Irset (Institut de Recherche en Santé, Environnement et Travail)-UMR S 1085, F-35000 Rennes, France

<sup>3</sup>Université de Paris, F-75006 Paris, France

<sup>4</sup>Université de Paris, CNRS, Institut Jacques Monod, F-75006 Paris, France

<sup>5</sup>Université Paris-Saclay, Institut de Biologie F. Jacob, Commissariat à l'Energie Atomique, Fontenay aux Roses, France

Correspondence should be addressed to Philippe Monget; [philippe.monget@inrae.fr](mailto:philippe.monget@inrae.fr)

Received 14 April 2021; Accepted 5 July 2021; Published 30 July 2021

Academic Editor: Mohamed Salem

Copyright © 2021 Floriane Picolo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Gene dosage is an important issue both in cell and evolutionary biology. Most genes are present in two copies or alleles in diploid eukaryotic cells. The most outstanding exception is monoallelic gene expression (MA) that concerns genes localized on the X chromosome or in regions undergoing parental imprinting in eutherians, and many other genes scattered throughout the genome. In diploids, haploinsufficiency (HI) implies that a single functional copy of a gene in a diploid organism is insufficient to ensure a normal biological function. One of the most important mechanisms ensuring functional innovation during evolution is whole genome duplication (WGD). In addition to the two WGDs that have occurred in vertebrate genomes, the teleost genomes underwent an additional WGD, after their divergence from tetrapods. In the present work, we have studied on 57 teleost species whether the orthologs of human MA or HI genes remain more frequently in duplicates or returned more frequently in singleton than the rest of the genome. Our results show that the teleost orthologs of HI human genes remained more frequently in duplicate than the rest of the genome in all of the teleost species studied. No signal was observed for the orthologs of genes mapping to the human X chromosome or subjected to parental imprinting. Surprisingly, the teleost orthologs of the other human MA genes remained in duplicate more frequently than the rest of the genome for most teleost species. These results suggest that the teleost orthologs of MA and HI human genes also undergo selective pressures either related to absolute protein amounts and/or of dosage balance issues. However, these constraints seem to be different for MA genes in teleost in comparison with human genomes.

## 1. Introduction

Gene dosage effects are an important phenomenon in cell biology that has evolutionary consequences. Indeed, in diploid eukaryotic cells, most genes are present in two copies that are transcribed and produce functional proteins. However, there are exceptions. The most outstanding exception is the case of monoallelic gene expression (MA). This is so for the majority of genes that are present on the X chromosome of eutherian

mammals, genes that present a parental imprinting in eutherians, and genes encoding immunoglobulins and olfactory receptors [1]. Monoallelic expression of genes is under an epigenetic control that is not well understood. For these genes, dysregulation of the mechanism(s) underlying monoallelic expression can lead to expression of both alleles and to overexpression of the corresponding protein and thus to severe pathologies [2]. An abnormal situation concerns haploinsufficiency. Haploinsufficiency is a biological phenomenon

responsible for the fact that a single functional copy of a gene in a diploid organism is insufficient to ensure a normal biological function. Haploinsufficiency is detected more frequently in essential genes than in nonessential genes in yeast [3]. Two nonmutually exclusive theories have been proposed to explain the cause of haploinsufficiency: the “insufficient amounts” hypothesis and the gene dosage balance hypothesis (GDBH). The “insufficient amounts” hypothesis states that haploinsufficiency is the consequence of a reduced protein amount due to the loss of function of one allele, this amount being insufficient to ensure its biological function [4]. This hypothesis does not explain why haploinsufficiency persisted over evolutionary time. The GDBH suggests that the phenotype caused by changes of protein level in a biological process is due to stoichiometric imbalances in protein complexes or cellular circuits involved in cellular functions [5, 6]. This hypothesis predicts that haploinsufficient genes is responsible for a biological defect when the amount of proteins is halved (such as A in a complex A-B-A) but also in excess in particular cases (such as B in the same complex) [6]. In contrast to the “insufficient amounts” hypothesis, this hypothesis proposes an explanation for the conservation of haploinsufficiency during evolution.

One of the most important mechanisms ensuring functional innovation during evolution is gene duplication or the duplication of entire genome [7, 8]. Whole genome duplication (WGD) events have been observed in all taxonomic groups: bacteria [9], unicellular eukaryotes [10], and plants [11]. In vertebrates, there have been two rounds of duplication of the ancestral deuterostome genome [12]. One of the striking features that characterize the teleost genomes is that they underwent an additional WGD, also called the teleost-specific genome duplication (TGD), after the divergence from tetrapod [13]. This specific WGD event provided important additional genetic material, which strongly contributed to the radiation of teleost fishes [14]. Teleost constitutes a monophyletic group of ray finned fishes and is the largest and most diverse group of vertebrates [15–18]. The high diversity of fish species combined with a recent complete duplication makes Clupeocephala a group of great interest for the study of complete genome duplication in the animal kingdom.

Unlike single-gene duplication events, a WGD provides all at once a large number of new genetic material, promoting an increased inter- and intraspecific diversity [19, 20]. Interestingly, after WGD, all genes do not remain in duplicate with the same probability. Most models predict a rapid return of part of the duplicates to a singleton state [21], the extra-copies being rapidly pseudogenized [22]. In particular for the rainbow trout, whose genome has duplicated one more time than that of the teleost about 100 my ago, it is estimated that about 48% of the genome remained in duplicate, when the remaining 52% of the genome quickly returned to a singleton state [23].

Understanding the rules explaining why certain genes remain in duplicate when others return to singleton is a challenging issue. It has been shown that certain families of genes are more likely to remain as duplicates in all taxonomic groups studied. This is the case for transcription factors, pro-

tein kinases, enzymes, and transporters [24]. Recently, we showed that this is also the case for genes encoding membrane receptors and their ligands [25]. The first explanation that has been put forward to explain the fact that genes are more often kept in duplicate is that these molecules are involved in key functions common to all organisms. Their quantitative increase would favor these key functions because of an increase in the number of molecules produced (selection for an absolute dosage increase) and/or because of a compensation of a potential loss of function mutation of one of both copies. Another explanation is based on the respect of gene dosage balance. This is particularly so for proteins whose function is heavily dependent on interactions with partners.

In the present work, we have studied on 57 teleost species whether the orthologs of human genes known to present a monoallelic (MA) expression or to be haploinsufficient (HI) in human remain more frequently in duplicate or returned more frequently to as singleton state than the whole genome in fish species or not.

## 2. Results and Discussion

We found a mean number of 13882 human genes on 22836 (60.8%) that possess at least one ortholog in at least one teleost genome. Among them, an average of 9854 (ranging from 3530 to 10868) have returned in singleton, an average of 3135 (ranging from 2323 to 7066) remained in duplicate, and an average of 893 (ranging from 337 to 4772) are in triplicate or more copies.

Concerning the 312 human HI genes, 299 (95.8%) possessed at least one ortholog in at least one teleost genome. Among them, an average of 172 (ranging from 47 to 199 depending on the studied species) have returned to singleton, an average of 85 (ranging from 68 to 122 depending on the species) remained in duplicate, and an average of 19 (ranging from 3 to 140) are in triplicate or more copies. A total of 285 genes remained in duplicate (or more) in at least one species among the 57 teleost species studied (Figure 1 and Suppl. Data (available here)). In comparison with the whole genome, this higher percentage of genes returned to singleton and remained in duplicate or more is significantly different for 55 species out of 57 ( $\chi^2$  analysis,  $p$  value ranging from  $0.058$  to  $4.2E - 6$ ) and for the 57 species studied (according to a hypergeometric test,  $p$  value ranging from  $0.034$  to  $8.5E - 6$ ; Suppl. data). Moreover, in comparison with the whole genome as well, the higher percentage of genes that are in triplicate or more copies is significantly higher in the genomes of rainbow trout, brown trout, Atlantic salmon, huchen, and common carp ( $p$  value ranging from  $1.3E - 8$  to  $8.1E - 4$ ) but not in the genome of the other teleosts. These results suggest that the teleost orthologs of HI human genes are also subjected to selective pressures either related to absolute protein amounts and/or of dosage balance issues. This suggests that HI genes in humans undergo similar constraints in teleost.

Among the 285 genes that remained in duplicate in at least one teleost species, 76 genes remained in duplicate or more in at least 80% (45) of the species. These genes encode

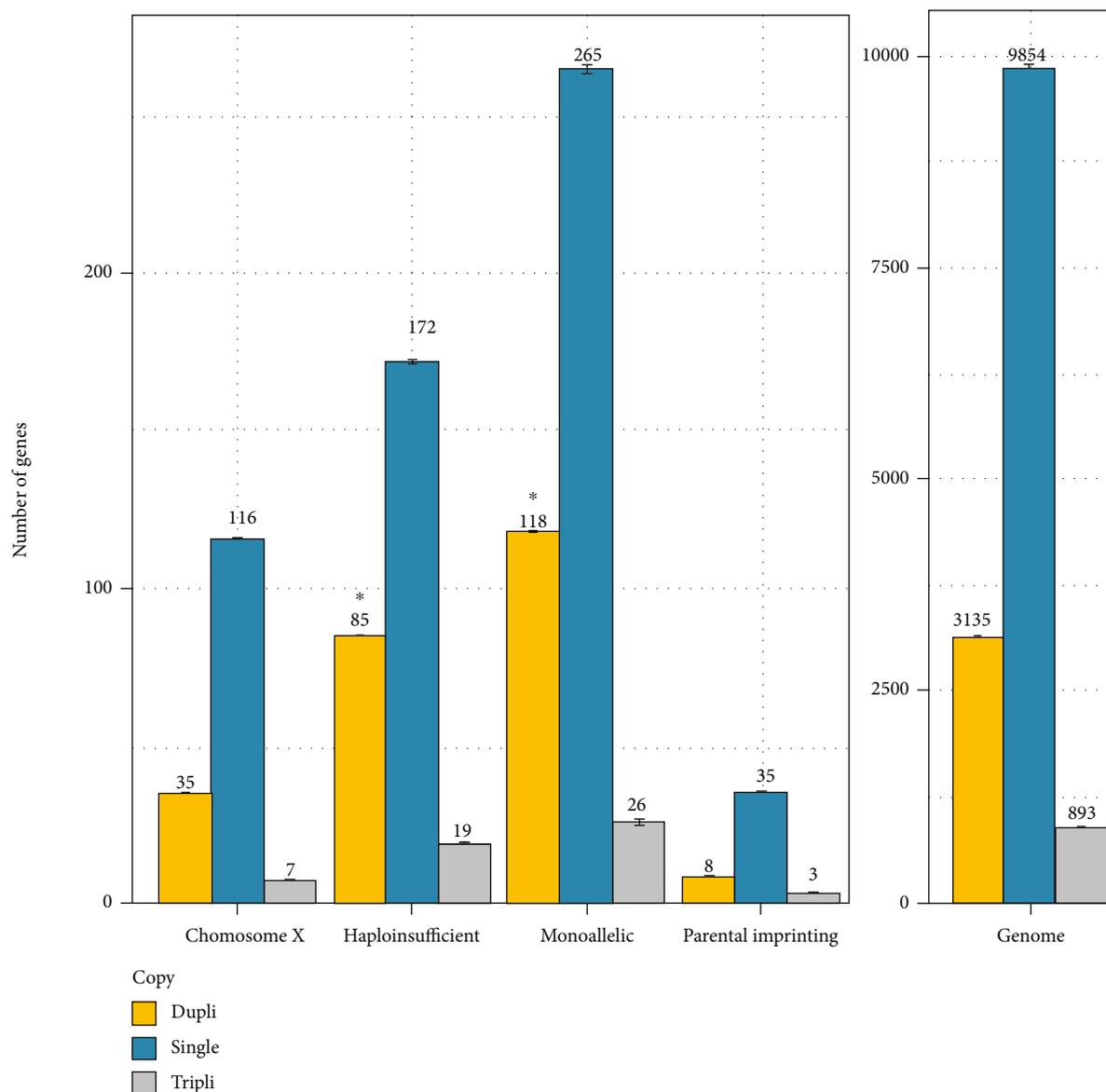


FIGURE 1: Barplot of the global distribution of the genes in each category: teleost orthologs of human genes mapping to the X chromosome, of human haploinsufficient (HI) genes, of human genes of monoallelic expression (MA, except genes that present a parental imprinting and localized on the X chromosome), and of human genes that present a parental imprinting. Right: teleost orthologs of human genes of the whole genome. The yellow bars correspond to the genes that remained in duplicate; the blue bars correspond to the genes returned to singleton. The grey bars correspond to the genes in triplicate or more. The results are presented as the mean  $\pm$  SEM. \* indicates a significant difference compared with the whole genome ( $p < 0.05$ ).

more (from 3 to 38 more times) transcription factors than the rest of the genome: bHLH transcription factor binding (Gene Ontology/GO:0043425); RNA polymerase II activating transcription factor binding (GO:0001102); activating transcription factor binding (GO:0033613); transcription factor binding (GO:0008134); DNA-binding transcription factor binding (GO:0140297); DNA-binding transcription factor activity and RNA polymerase II-specific (GO:0000981); and DNA-binding transcription factor activity (GO:0003700). This enrichment of GO terms is completely in accordance with previously reported findings [6]. There was no particular representative GO among the genes retained as triplicates in the genome of teleost species. These results are compatible

both with direct insufficiency of a transcription factor and with balance issues (as they are often multisubunit complexes). Threshold effects can also be at play because of the strongly nonlinear relationships (sigmoidal or S-shaped) produced by the cooperative binding of a transcription factor to a cis-regulatory sequence and the transcriptional response. Thus, depending on the concentration of transcription factor, a halved dosage may not be sufficient to cross the threshold required for a normal transcriptional response [6].

Concerning the 206 X-linked human genes, 176 (82.6%) possessed at least one ortholog in at least one teleost genome. Among them, an average of 116 (ranging from 32 to 132 depending on the studied species) have returned to singleton,

an average of 35 (ranging from 23 to 79 depending on the species) remained in duplicate, and an average of 7 (ranging from 0 to 54) are in triplicate or more copies (Figure 1 and Suppl. Data). Concerning the 90 imprinted genes, 51 (56.7%) had at least one ortholog in at least one teleost genome. Among them, an average of 35 (ranging from 12 to 41 depending on the studied species) have returned to singleton, an average of 8 (ranging from 3 to 23 depending on the species) remained in duplicate, and an average of 3 (ranging from 0 to 20) are in triplicate or more copies (Figure 1 and Suppl. Data). Thus, the teleost orthologs of human genes subjected to genetic imprinting or located on the human X chromosome returned to singleton or remained in duplicate (or remain present as triplicates or more copies), in the same proportions than the rest of the genome.

Concerning the 580 human MA genes that are not on the X chromosome and that are not subjected to parental imprinting, 469 (80.9%) had at least one ortholog in at least one teleost genome. Among them, an average of 265 (ranging from 87 to 296) have returned to singleton, an average of 118 (ranging from 87 to 193) remained in duplicate, and an average of 26 (ranging from 4 to 160) were found in triplicate or more copies. A total of 437 genes remained in duplicate in at least one species among the 57 teleost species studied (Figure 1 and Suppl. Data). In comparison with the whole genome, the difference of percentage of genes remained in duplicate or more is significantly higher for 47 species on 57 ( $\chi^2$  analysis,  $p$  value ranging from 0.055 to 6.5 4) and for 50 species on 57 (hypergeometric test,  $p$  value ranging from 0.044 to 6.2 4; Suppl. data). Moreover, in comparison with the whole genome as well, the difference of percentage of genes that are in triplicate or more copies is significantly higher in the genomes of rainbow trout, brown trout, Atlantic salmon, huchen, and common carp ( $p$  value ranging from 0.056 to 5.3 3), not in the genome of the other teleosts. Whether these results are generalizable to other salmonids needs further investigation. We found this result surprising. Indeed, one would have hypothesized that the teleost orthologs of MA human genes returned more frequently to singleton than the whole genome. This suggests that the regulation (epigenetic mechanism) of monoallelic expression is not likely to occur for these genes in teleosts. Moreover, this suggests that the constraints to express only one allele in the human do not exist for these genes in teleosts. That being said, MA is a complex regulatory process that has evolved perhaps due to parent-offspring conflict. It might be possible that genes that need such control in mammals would be those with a general need for dosage balance in other species like teleost. Unlike the HI genes, there was no particularly representative GO among the MA genes.

### 3. Material and Methods

We studied 57 species of fish: Amazon molly (*Poecilia formosa*), Atlantic herring (*Clupea harengus*), Atlantic salmon (*Salmo salar*), ballan wrasse (*Labrus bergylta*), barramundi perch (*Lates calcarifer*), blue tilapia (*Oreochromis aureus*), blunt-snouted clingfish (*Gouania willdenowi*), brown trout (*Salmo trutta*), Burton's mouthbrooder (*Haplochromis bur-*

*toni*), channel bull blenny (*Cottoperca gobio*), channel catfish (*Ictalurus punctatus*), climbing perch (*Anabas testudineus*), cod (*Gadus morhua*), common carp (*Cyprinus carpio* common\_carp\_genome), denticle herring (*Denticeps clupeoides*), Eastern happy (*Astatotilapia calliptera*), electric eel (*Electrophorus electricus*), European seabass (*Dicentrarchus labrax*), fugu (*Takifugu rubripes*), gilthead seabream (*Sparus aurata*), greater amberjack (*Seriola dumerili*), guppy (*Poecilia reticulata*), huchen (*Hucho hucho*), Indian glassy fish (*Parambassis ranga*), Indian medaka (*Oryzias melastigma*), Japanese medaka HdrR (*Oryzias latipes* ASM223467v1), Japanese medaka HNI (*Oryzias latipes* ASM223471v1), Japanese medaka HSOK (*Oryzias latipes* ASM223469v1), jewelled blenny (*Salarias fasciatus*), large yellow croaker (*Larimichthys crocea*), live sharksucker (*Echeneis naucrates*), lyretail cichlid (*Neolamprologus brichardi*), Makobe Island cichlid (*Pundamilia nyererei*), Mexican tetra (*Astyanax mexicanus* *Astyanax\_mexicanus*-2.0), Midas cichlid (*Amphilophus citrinellus*), mummichog (*Fundulus heteroclitus*), Nile tilapia (*Oreochromis niloticus*), northern pike (*Esox lucius*), orbiculate cardinalfish (*Sphaeramia orbicularis*), Pachon cavefish (*Astyanax mexicanus* *Astyanax\_mexicanus*-1.0.2), pinecone soldierfish (*Myripristis murdjan*), rainbow trout (*Oncorhynchus mykiss*), red-bellied piranha (*Pygocentrus nattereri*), sailfin molly (*Poecilia latipinna*), sheepshead minnow (*Cyprinodon variegatus*), shortfin molly (*Poecilia mexicana*), siamese fighting fish (*Betta splendens*), stickleback (*Gasterosteus aculeatus*), swamp eel (*Monopterus albus*), tetraodon (*Tetraodon nigroviridis*), tiger tail seahorse (*Hippocampus comes*), tongue sole (*Cynoglossus semilaevis*), turbot (*Scophthalmus maximus*), yellowtail amberjack (*Seriola lalandi dorsalis*), zebra mbuna (*Maylandia zebra*), zebrafish (*Danio rerio*), and zig-zag eel (*Mastacembelus armatus*).

These fish species diverged after complete TGD. The human genes were retrieved from ENSEMBL. The ortholog copy for each gene was established in every one of the 57 fish species. Then, in each species, the fate (singleton vs duplicate) of the entirety of the human gene orthologs was studied. Moreover, a total of 312 human genes known to be haploinsufficient were recovered from Clingen (<https://www.ncbi.nlm.nih.gov/projects/dbvar/clingen/>), 580 human genes were known to be monoallelic [26], 206 X human chromosome genes were recovered for GeneImprint (<http://www.geneimprint.com/site/genes-by-species>), and there are 90 genetic imprinting genes [27], and the fate of their fish orthologs was recovered. A list of human genes (GRCh38.p13) was generated using BioMart from Ensembl Genes 101. The set of human genes encoding a protein (protein\_coding) is selected from the gene type filter. The selected attributes in the homologous category are the different species of teleostans listed in ENSEMBL. Only stable gene IDs were selected. A list of 22836 human genes encoding a protein is listed.

We got between 12,918 (tetraodon) and 14,626 (brown trout) orthologous genes by fish species (average: 13,882). This does not represent the entire genome of each fish but allowed us to make strong statistics. Moreover, we compared the global evolution of the whole human genome that had orthologs in fishes with the specific evolution of human

MA and HI genes in fish species. We studied whether these fish orthologs of MA and HI genes remained as a duplicate copy or had return to singleton in the same proportion as whole human ortholog genes.

Both the  $\chi^2$  test statistical analysis and hypergeometric analysis with Benjamini-Hochberg correction were used to test the hypothesis that teleost genes that are orthologs of human MA and HI genes remained more in duplicate than the whole genome. All the statistical tests conducted in our study were performed in R. Moreover, the Panther database (<http://www.pantherdb.org/>) was used to study the gene ontology of teleost genes that are orthologs to human HI genes, and Fisher's test with Benjamini-Hochberg correction was used to classify genes according to the family.

### Data Availability

The underlying data supporting the results of this study can essentially be found in Supplemental data and can be verified on ENSEMBL (<http://www.ensembl.org/index.html>).

### Disclosure

This manuscript has been submitted as a preprint at <https://www.biorxiv.org/content/10.1101/2021.01.12.426466v1>.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

The present study was supported by a fellowship from the French Ministry of Research, by the INRAE institute, and by a thesis scholarship funded by the University of Tours (France). Thanks are also due to Alexandra Louis and Hugues Roest Crolius for helpful discussion and technical assistance.

### Supplementary Materials

Tables of statistic tests for each species of teleost and for each category. Each category is represented in each of the four Suppl. Data (HI, MA, X chromosome, and parental imprinting), and these four tables have the same construction. By column (category HI for example): (A) species; (B) total number of teleost genes returned in singleton; (C) total number of teleost genes remained in duplicate; (D) total number of teleost genes in triplicate or more copies; (E) total number of teleost genes with a human ortholog; (F) number of teleost orthologs of HI human genes returned in singleton; (G) number of teleost orthologs of HI human genes remained in duplicate; (G) number of teleost orthologs of HI human genes in triplicate or more copies; (I) total number of teleost orthologs to human HI gene; (J)  $\chi^2$  value of the repartition of HI orthologs in singleton, in duplicate or more copies in comparison with the whole genome; (K)  $p$  value of  $\chi^2$  test; (L)  $\chi^2$  false discovery rate (FDR) by Benjamini Hochberg (BH) procedure. (M)  $p$  value of hypergeometric test between singleton and duplicate/more copies; (N) hypergeometric

FDR by BH procedure; (O)  $p$  value of hypergeometric test between triplicate or more copies and duplicate or less copies; (P) hypergeometric FDR by BH procedure. The same organization of columns is used for the other categories (MA, X chromosome, and imprinted genes). Concerning for HI and MA categories, the  $\chi^2$  test is significant for 55/57 and 47/57 species, respectively, and the hypergeometric test is significant for 57/57 and 50/57 species, respectively; i.e., these orthologs remain more frequently in duplicate than the whole genome. For comparison between triplicate (or more copies) and duplicate (or less copies), the hypergeometric test is significant for 5/57 (salmonids and carp) and 3/57 species, respectively (salmonid and carp as well). (*Supplementary Materials*)

### References

- [1] A. Chess, "Monoallelic gene expression in mammals," *Annual Review of Genetics*, vol. 50, no. 1, pp. 317–327, 2016.
- [2] B. Horsthemke, "Mechanisms of imprint dysregulation," *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, vol. 154C, no. 3, pp. 321–328, 2010.
- [3] S. Ohnuki and Y. Ohya, "High-dimensional single-cell phenotyping reveals extensive haploinsufficiency," *PLoS Biology*, vol. 16, no. 5, article e2005130, 2018.
- [4] A. M. Deutschbauer, D. F. Jaramillo, M. Proctor et al., "Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast," *Genetics*, vol. 169, no. 4, pp. 1915–1925, 2005.
- [5] B. Papp, C. Pál, and L. D. Hurst, "Dosage sensitivity and the evolution of gene families in yeast," *Nature*, vol. 424, no. 6945, pp. 194–197, 2003.
- [6] R. A. Veitia, "Exploring the etiology of haploinsufficiency," *Bio Essays: News and Reviews in Molecular, Cellular and Developmental Biology*, vol. 24, no. 2, pp. 175–184, 2002.
- [7] H. Innan and F. Kondrashov, "The evolution of gene duplications: classifying and distinguishing between models," *Nature Reviews Genetics*, vol. 11, no. 2, pp. 97–108, 2010.
- [8] S. Ohno, U. Wolf, and N. B. Atkin, "Evolution from fish to mammals by gene duplication," *Hereditas*, vol. 59, no. 1, pp. 169–187, 1968.
- [9] M. Kuroda, T. Ohta, I. Uchiyama et al., "Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*," *Lancet (London, England)*, vol. 357, no. 9264, pp. 1225–1240, 2001.
- [10] M. Kellis, B. W. Birren, and E. S. Lander, "Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*," *Nature*, vol. 428, no. 6983, pp. 617–624, 2004.
- [11] K. L. Adams and J. F. Wendel, "Polyploidy and genome evolution in plants," *Current Opinion in Plant Biology*, vol. 8, no. 2, pp. 135–141, 2005.
- [12] B. K. Mable, M. A. Alexandrou, and M. I. Taylor, "Genome duplication in amphibians and fish : an extended synthesis," *Journal of Zoology*, vol. 284, no. 3, pp. 151–182, 2011.
- [13] S. M. K. Glasauer and S. C. F. Neuhaus, "Whole-genome duplication in teleost fishes and its evolutionary consequences," *Molecular Genetics and Genomics: MGG*, vol. 289, no. 6, pp. 1045–1060, 2014.
- [14] V. Ravi and B. Venkatesh, "Rapidly evolving fish genomes and teleost diversity," *Current Opinion in Genetics & Development*, vol. 18, no. 6, pp. 544–550, 2008.

- [15] A. Christoffels, E. G. L. Koh, J. Chia, S. Brenner, S. Aparicio, and B. Venkatesh, "Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes," *Molecular Biology and Evolution*, vol. 21, no. 6, pp. 1146–1151, 2004.
- [16] M. Robinson-Rechavi, O. Marchand, H. Escriva, and V. Laudet, "An ancestral whole-genome duplication may not have been responsible for the abundance of duplicated fish genes," *Current Biology: CB*, vol. 11, no. 12, pp. R458–R459, 2001.
- [17] J. S. Taylor, I. Braasch, T. Frickey, A. Meyer, and Y. Van de Peer, "Genome duplication, a trait shared by 22000 species of ray-finned fish," *Genome Research*, vol. 13, no. 3, pp. 382–390, 2003.
- [18] J. S. Taylor and J. Raes, "Duplication and divergence : the evolution of new genes and old ideas," *Annual Review of Genetics*, vol. 38, no. 1, pp. 615–643, 2004.
- [19] Y. Van de Peer, S. Maere, and A. Meyer, "The evolutionary significance of ancient genome duplications," *Genetics*, vol. 10, no. 10, pp. 725–732, 2009.
- [20] Y. Van de Peer, E. Mizrahi, and K. Marchal, "The evolutionary significance of polyploidy," *Genetics*, vol. 18, no. 7, pp. 411–424, 2017.
- [21] S. Maere, S. De Bodt, J. Raes et al., "Modeling gene and genome duplications in eukaryotes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 15, pp. 5454–5459, 2005.
- [22] D. Sankoff, C. Zheng, and Q. Zhu, "The collapse of gene complement following whole genome duplication," *BMC Genomics*, vol. 11, no. 1, p. 313, 2010.
- [23] C. Berthelot, F. Brunet, D. Chalopin et al., "The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates," *Nature Communications*, vol. 5, no. 1, p. 3657, 2014.
- [24] G. Conant and K. Wolfe, "Turning a hobby into a job : how duplicated genes find new functions," *Nature reviews Genetics*, vol. 9, pp. 938–950, 2008.
- [25] A. Grandchamp, B. Piégu, and P. Monget, "Genes encoding teleost fish ligands and associated receptors remained in duplicate more frequently than the rest of the genome," *Genome Biology and Evolution*, vol. 11, no. 5, pp. 1451–1462, 2019.
- [26] A. Nag, S. Vigneau, V. Savova, L. M. Zwemer, and A. A. Gimelbrant, "Chromatin signature identifies monoallelic gene expression across mammalian cell types," *G3 (Bethesda, Md.)*, vol. 5, no. 8, pp. 1713–1720, 2015.
- [27] L. Carrel and H. F. Willard, "X-inactivation profile reveals extensive variability in X-linked gene expression in females," *Nature*, vol. 434, no. 7031, pp. 400–404, 2005.