D724–D730 *Nucleic Acids Research, 2020, Vol. 48, Database issue* *Published online 15 October 2019*
doi: 10.1093/nar/gkz909

OHNOLOGS v2: a comprehensive resource for the genes retained from whole genome duplication in vertebrates

Param Priya Singh* **and Hervé Isambert***

Institut Curie, Research Center, CNRS UMR168, PSL Research University, 26 rue d'Ulm, 75005, Paris, France

Received July 30, 2019; Revised September 20, 2019; Editorial Decision October 01, 2019; Accepted October 10, 2019

ABSTRACT

All vertebrates including human have evolved from an ancestor that underwent two rounds of whole genome duplication (2R-WGD). In addition, teleost fish underwent an additional third round of genome duplication (3R-WGD). The genes retained from these genome duplications, so-called ohnologs, have been instrumental in the evolution of vertebrate complexity, development and susceptibility to genetic diseases. However, the identification of vertebrate ohnologs has been challenging, due to lineage specific genome rearrangements since 2R- and 3R-WGD. We previously identified vertebrate ohnologs using a novel synteny comparison across multiple genomes. Here, we refine and apply this approach on 27 vertebrate genomes to identify ohnologs from both 2R- and 3R-WGD, while taking into account the phylogenetically biased sampling of available species. We assemble vertebrate ohnolog pairs and families in an expanded OHNOLOGS v2 database. We find that teleost fish have retained more 2R-WGD ohnologs than mammals and sauropsids, and that these 2R-ohnologs have retained significantly more ohnologs from the subsequent 3R-WGD than genes without 2R-ohnologs. Interestingly, species with fewer extant genes, such as sauropsids, have retained similar or higher proportions of ohnologs. OHNOLOGS v2 should allow deeper evolutionary genomic analysis of the impact of WGD on vertebrates and can be freely accessed at http://ohnologs.curie.fr.

INTRODUCTION

Gene duplication provides raw material for the evolution of new gene functions (1). Duplication of single genes or genomic segments is a continuous evolutionary process that creates diversity in terms of copy number variations across individuals, and paralogs across species. In addition, dramatic evolutionary accidents corresponding to whole genome duplication (WGD) have also occurred in the evolutionary past of most eukaryotic lineages including plans, fungi and animals (2–4). For example, all extant vertebrates have experienced two rounds of WGDs (2R-WGD) in their evolutionary past (5–8). In addition, a third round of genome duplication has also occurred in the teleost fish lineage (3R-WGD) (9–11). 2R-WGDs likely played important roles in the evolution and diversification of vertebrate specific innovations such as neural crest cells, placodes and a complex brain (12,13). Many key genes implicated in the development of these structures can be traced back to 2R-WGD. Similarly, 3R-WGD likely played an important role in the expansion of the diversity of teleost fish lineage making it the most species rich vertebrate group (14–17). Hence, the genes retained from these three WGD events have been instrumental in the evolution of vertebrates (18).

The genes originated from these ancient polyploidy (paleo-polyploidy) events are now called ohnologs after Susumu Ohno who first hypothesized the two rounds of WGD events in vertebrate ancestors (1,5,19). Ohnologs are known to have distinct evolutionary, genomic and functional properties that distinguish them from small-scale duplicates and singletons (20–23). They also show greater association with diseases and cancer than non-ohnolog genes (24–29), and have been suggested to be dosage balanced (24), which was subsequently argued to be indirectly mediated by their high susceptibility to dominant mutations (25,28), as supported by quantitative population genetics models (27) and by a global inference approach assessing direct *versus* indirect causal relationships across multiple genomic properties (30).

Given the specific impact WGDs have had on the evolution of vertebrates, a comprehensive database of vertebrate ohnologs is highly desirable. While there are some useful resources available for comparison of synteny across

*To whom correspondence should be addressed. Tel: +33 1 56 24 64 74; Email: herve.isambert@curie.fr
Correspondence may also be addressed to Param Priya Singh. Tel: + 1 408 680 9845; Email: param@stanford.edu
Present address: Param Priya Singh, Stanford University, Department of Genetics, 300 Pasture Drive, Stanford, CA 94305, USA.
Disclaimer: The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

species (31–34) there is no database that reliably identifies ohnologs from both vertebrate 2R-WGDs and fish 3R-WGD. To start filling this gap, we developed in 2015, OHNOLOGS, a repository of ohnologs retained from the 2R-WGD in six amniote vertebrates (human, mouse rat, pig, dog and chicken) (34). OHNOLOGS is based on a novel comparative macro-synteny approach that reliably identifies ohnologs, despite lineage specific genome rearrangement, gene loss and small scale duplication events, by combining macro-synteny information (gene content regardless of exact order) across multiple outgroups and vertebrate genomes (34).

Here, we expand this multiple genome synteny comparison approach to 27 vertebrate species including four teleost fish species. We further improve the statistical confidence assessment of each ohnolog pair with a weighted quantitative confidence score (q-score) taking into account the phylogenetically biased sampling of available vertebrate species. In addition, we uncover ohnologs, including in non-protein coding RNA gene classes, from both 2R-WGD in early vertebrates (2R-ohnologs) and 3R-WGD in teleost fish (3R-ohnologs). The expanded OHNOLOGS database is the most comprehensive repository of ohnologs in vertebrates. Using the new OHNOLOGS database we show that on average 25% of extant genes are 2R-ohnologs in vertebrates and that 18% of extant genes are 3R-ohnologs in teleost fish. Sauropsids show the highest lineage-specific loss of 2R-ohnologs, and teleost fish show the highest lineage specific retention of 2R-ohnologs. We also found that 2R-ohnologs are significantly more likely to retain 3R-ohnologs in teleost fish, in agreement with earlier reports (35). OHNOLOGS v2 should facilitate deeper evolutionary analysis of the unique properties of ohnologs, and their lineage-specific retention and loss in different vertebrates.

RESULTS

Data collection and processing

OHNOLOGS v2 includes 2R-ohnolog pairs and families in 27 vertebrates that have a chromosome level assembly with a majority of their genes anchored on chromosomes in Ensembl version 84 (36). This includes 18 mammals, 4 sauropsids (lizards and birds), 4 teleost fish and spotted gar. In addition, we also included 3R-ohnolog pairs and families in four teleost fish genomes. We used five non-vertebrate outgroups to identify 2R-ohnologs and seven vertebrate outgroups to identify fish specific 3R-ohnologs (Figure 1 and Supplementary Table S1).

We collected genes (protein coding, micro-RNA, miscellaneous RNA, rRNA, snRNA and snoRNA) for all these organisms from Ensembl v84 using biomaRt (37,38). These six classes of genes were chosen because they have information on orthologs and paralogs across many vertebrates. Orthologs, paralogs and relative duplication node for all the genes were obtained from Ensembl comparative genomics resource (39). These homology relationships and their relative duplication time have been computed by reconciling gene based phylogenetic trees with the species phylogeny for each Ensembl gene family (40). To identify duplication time of paralogs consistently, we took the consensus timing across 7 Ensembl versions (v80–v86). Genes with a lot of small-scale duplications (>30), which inflate the synteny calculations, were excluded from analysis. Genome data for Amphioxus was obtained from JGI and amphioxus orthologs with other organisms were identified using BLASTp (8).

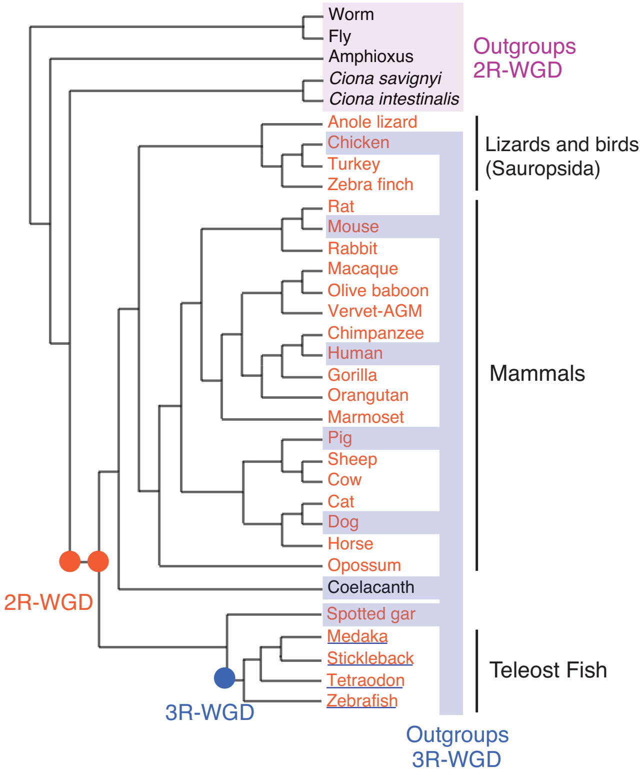


Figure 1. A schematic phylogeny (not scaled) of the organisms in the OHNOLOGS v2 database. Vertebrates analysed for 2R-WGD are in orange, and teleost fish species analysed for 3R-WGD are underlined. Outgroup species used to identify 2R- and 3R-ohnologs have been highlighted.

We adapted the macro-synteny comparison approach, previously developed in (34), to identify ohnologs retained from both 2R-WGD (2R-ohnologs) and 3R-WGD (3R-ohnologs). Briefly, for each pair of outgroup and paleo-polyploid organisms, we first identified blocks of conserved macro-synteny using windows ranging from 100 to 500 genes (outgroup comparison). These macro-synteny blocks have a pattern of doubly conserved synteny, where a window in the outgroup genome shares orthology with at least two other windows in the paleo-polyploid genome. The paralogs residing on these windows and duplicated at the time of 2R- or 3R-WGD are candidates for being 2R- or 3R-ohnologs, respectively. Similarly, we also identified syntenic windows by comparing each paleo-polyploid genome to itself (self comparison).

To refine these ohnologs further and eliminate spurious synteny patterns, we computed a quantitative score (called q-score) to assess the probability that any ohnolog pair could be identified by chance, following the approach developed in (34). In brief, all q-scores from different windows and outgroups were combined to give a global q-score for each ohnolog pair from outgroup comparison. Using multiple outgroups allowed us to identify ohnologs that may have moved to non-syntenic locations in some of the outgroup genomes. Similarly we obtained a q-score for self comparison to assess the chance of spurious association. In addition, while we used a simple geometric average of q-scores in (34), which cannot capture the gain of statistical power expected from the integration of multiple vertebrate genomes, here we developed a refined weighting scheme of species, which also takes into account the strong phylogenetically biased sampling of included species by using different weights for each vertebrate genome depending on its shared homology with other included genomes (see Supplementary Methods for details, Supplementary Tables S2 and 3).

Using both self and outgroup weighted q-scores, we generated three sets of ohnologs (corresponding to strict, intermediate and relaxed criteria) and combined them into ohnolog families. At last, we compiled both the 2R- and 3R-ohnolog pairs and ohnolog families for each organism in the interactive OHNOLOGS v2 database using Apache, CGI, Perl, Bootstrap and jQuery.

Navigating the OHNOLOGS database

The home page lists all the organisms that are included in OHNOLOGS for 2R and 3R-WGD along with an introduction on ohnologs and WGDs. The search page allows a user to search for a gene symbol, Ensembl Id GO term or any keyword (Figure 2A). The search page also allows one to generate ohnolog families for any user-defined q-score criteria for a given organism. Upon a keyword or GO term query, all matching genes will be displayed along with their ohnolog status (Figure 2B). If a queried gene is an ohnolog, its ohnolog family will be displayed on the result page (for both 2R and 3R WGD for teleost fish) (Figure 2C and D). We show families for our strict q-score filter, and display the intermediate and relaxed families only if additional ohnologs are identified upon relaxing the q-score filter. The result page also includes links to pair page that has all ohnolog pairs that went into constructing that family (Figure 2E). The family result pages also links to the orthologous genes and ohnolog families in other vertebrates, to study the conservation of ohnolog families in other vertebrates.

The ohnolog pairs and families for our three pre-defined q-score filters can be explored and downloaded from the Browse/Download pages (Figure 2F). We link the genes on the browse pages to external databases including Ensembl, NCBI gene, GeneCards (for human), MGI (for mouse) and ZFIN (for zebrafish). The details of our approach, family descriptions and more details on q-score have also been included on the help page.

Summary of the contents of the OHNOLOGS database

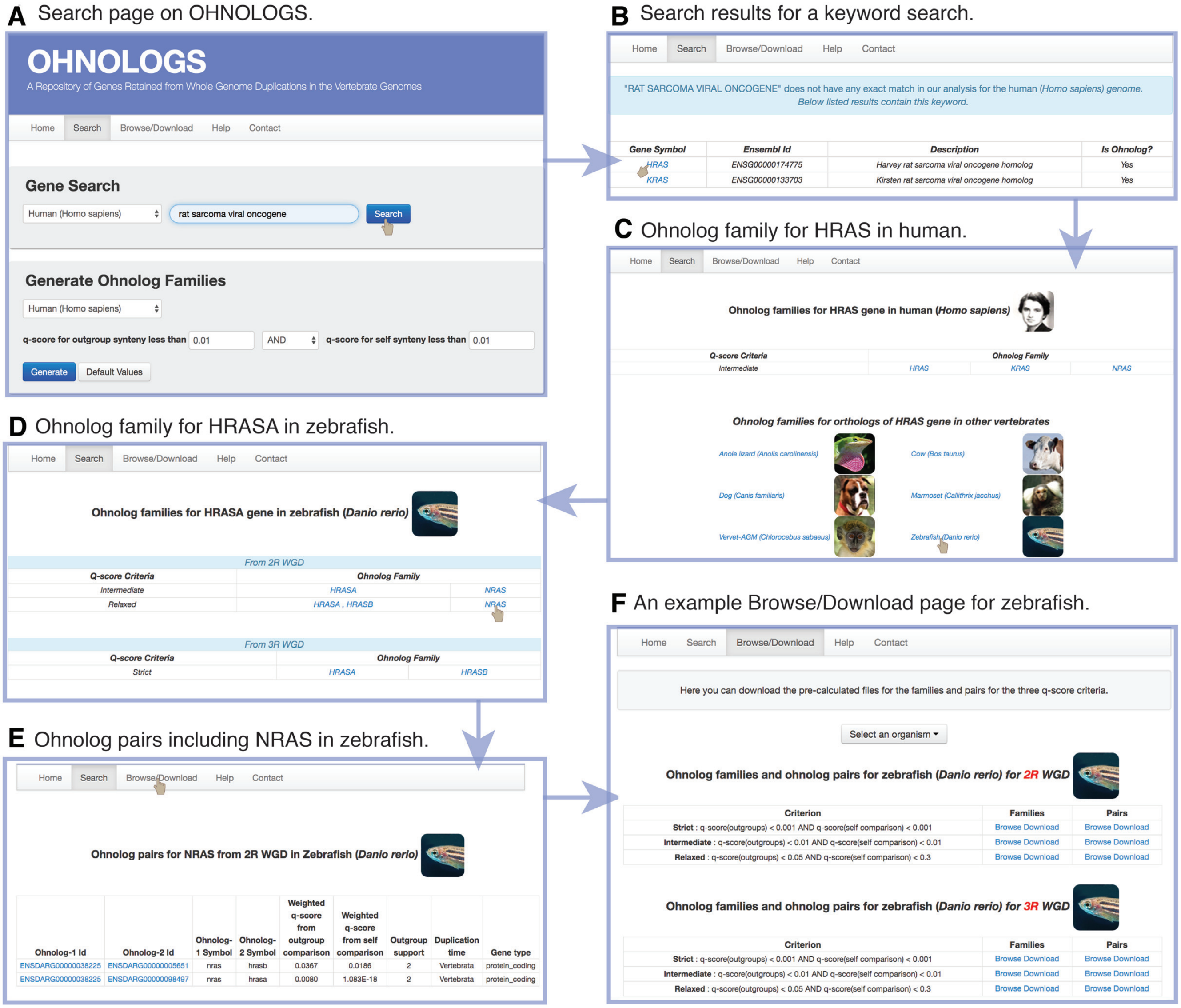
Using the expanded OHNOLOGS database we assessed the retention and loss of ohnologs across different vertebrates. We found that on average 25% of extant genes are 2R-ohnologs in vertebrates (intermediate criterion), which include two rounds of WGD, and that 18% of extant genes are 3R-ohnologs in teleost fish, which include an additional WGD (Figure 3A and B; Supplementary Table S4). Teleost fish have also retained more 2R-ohnologs in both absolute numbers (Figure 3A) and relative proportion of extant genes (32% on average). Interestingly, while sauropsids have usually fewer extant genes and 2R-ohnologs than other vertebrates (Figure 3A), they have retained similar or higher proportions of 2R-ohnologs in their genomes (28% on average). Similarly, at the level of individual species, we observe that more compact genomes, such as turkey and tetraodon, which typically contain also fewer genes, have retained about the same numbers and thus larger proportions of ohnologs than other birds or fish, respectively (Supplementary Table S4). This enhanced conservation of ohnologs in individual species or clades with fewer extant genes is consistent with their proposed retention mechanism through purifying selection in paleo-polyploid species (25,27–28).

A vast majority of retained ohnologs consists of protein-coding genes, while non-protein coding genes represent only a small fraction of ohnologs (Supplementary Table S5). For example, in human, out of the 7358 2R-ohnolog pairs from the relaxed criterion only 28 (0.4%) are mi-RNA ohnolog pairs and 2 (0.02%) are sno-RNA ohnolog pairs (Supplementary Table S5).

Remarkably, for all analysed vertebrates the size of 2R-ohnolog families rarely exceeds four ohnologs (Figure 3C), as expected for two rounds of WGD events. Similarly, virtually all 3R-ohnolog families are of size 2, as they are derived from just a single WGD event (Figure 3D). These family sizes also suggest a low rate of small-scale duplications and genome rearrangements following both 2R and 3R-WGD as previously noticed (24).

We then assessed whether teleost fish with their additional 3R-WGD event had further expanded the same ohnolog families as from the previous 2R-WGD events. Indeed, we found that in all four analysed teleost species, 2R-ohnologs tend to retain significantly more 3R-ohnologs (Figure 3E), in agreement with earlier reports (35). The retention of 3R-ohnologs is even higher for 2R-ohnologs that have retained three or four family members, and for the 2R-ohnologs that have been retained in all the 27 vertebrates (Figure 3E). For example zebrafish 2R-ohnologs from the intermediate criteria that have been also retained in all the analysed vertebrates are twice as likely to retain their 3R-ohnologs compared to genome-wide expectation ($P = 5e\text{-}88$, Chi-square test). This suggests that the evolutionary mechanism for the expansion of specific gene families through the retention of 2R-ohnologs (25,27–28) might also explain the biased retention of 3R-ohnologs.

We next compared the new OHNOLOGS v2 database (this study) with the 2015 version (v1) (34) to quantify the changes due to the improved pipeline. We noticed that the majority of ohnologs are shared between the two versions for all the six species included in v1 (Figure 4A and B). For example, using the relaxed criterion, 87% of individual ohnologs (Figure 4A) and 65% of ohnolog pairs (Figure 4B) in human were already present in the 2015 version. These differences are due to the improved weighted q-score taking into account the phylogenetically biased sampling of species, a broader taxonomic range and changes in ortholog/paralog relations in the recent Ensembl versions. Indeed, out of 3090 ohnolog pairs not identified in the up-



A Search page on OHNOLOGS.

OHNOLOGS

A Repository of Genes Retained from Whole Genome Duplications in the Vertebrate Genomes

Gene Search

Human (Homo sapiens) | rat sarcoma viral oncogene | Search

Generate Ohnolog Families

Human (Homo sapiens)

q-score for outgroup synteny less than 0.01 AND q-score for self synteny less than 0.01

B Search results for a keyword search.

"RAT SARCOMA VIRAL ONCOGENE" does not have any exact match in our analysis for the human (*Homo sapiens*) *genome. Below listed results contain this keyword.*

Gene Symbol	*Ensembl Id*	*Description*	*Is Ohnolog?*
HRAS	*ENSG00000174775*	*Harvey rat sarcoma viral oncogene homolog*	*Yes*
KRAS	*ENSG00000133703*	*Kirsten rat sarcoma viral oncogene homolog*	*Yes*

C Ohnolog family for HRAS in human.

Ohnolog families for HRAS gene in human (*Homo sapiens*)

Q-score Criteria	*Ohnolog Family*		
Intermediate	*HRAS*	*KRAS*	*NRAS*

Ohnolog families for orthologs of HRAS gene in other vertebrates

Anole lizard (Anolis carolinensis)
Cow (Bos taurus)
Dog (Canis familiaris)
Marmoset (Callithrix jacchus)
Vervet-AGM (Chlorocebus sabaeus)
Zebrafish (Danio rerio)

D Ohnolog family for HRASA in zebrafish.

Ohnolog families for HRASA gene in zebrafish (*Danio rerio*)

From 2R WGD

Q-score Criteria	*Ohnolog Family*	
Intermediate	*HRASA*	*NRAS*
Relaxed	*HRASA , HRASB*	*NRAS*

From 3R WGD

Q-score Criteria	*Ohnolog Family*	
Strict	*HRASA*	*HRASB*

E Ohnolog pairs including NRAS in zebrafish.

Ohnolog pairs for NRAS from 2R WGD in Zebrafish (*Danio rerio*)

Ohnolog-1 Id	Ohnolog-2 Id	Ohnolog-1 Symbol	Ohnolog-2 Symbol	Weighted q-score from outgroup comparison	Weighted q-score from self comparison	Outgroup support	Duplication time	Gene type
ENSDARG00000038225	ENSDARG00000005651	nras	hrasb	0.0367	0.0186	2	Vertebrata	protein_coding
ENSDARG00000038225	ENSDARG00000098497	nras	hrasa	0.0080	1.083E-18	2	Vertebrata	protein_coding

F An example Browse/Download page for zebrafish.

Here you can download the pre-calculated files for the families and pairs for the three q-score criteria.

Ohnolog families and ohnolog pairs for zebrafish (*Danio rerio*) *for* 2R WGD

Criterion	Families	Pairs
Strict : q-score(outgroups) < 0.001 AND q-score(self comparison) < 0.001	Browse Download	Browse Download
Intermediate : q-score(outgroups) < 0.01 AND q-score(self comparison) < 0.01	Browse Download	Browse Download
Relaxed : q-score(outgroups) < 0.05 AND q-score(self comparison) < 0.3	Browse Download	Browse Download

Ohnolog families and ohnolog pairs for zebrafish (*Danio rerio*) *for* 3R WGD

Criterion	Families	Pairs
Strict : q-score(outgroups) < 0.001 AND q-score(self comparison) < 0.001	Browse Download	Browse Download
Intermediate : q-score(outgroups) < 0.01 AND q-score(self comparison) < 0.01	Browse Download	Browse Download
Relaxed : q-score(outgroups) < 0.05 AND q-score(self comparison) < 0.3	Browse Download	Browse Download

Figure 2. Navigating the OHNOLOGS database. (**A**) Screenshot of the search page. (**B**) Result page for a keyword search of 'rat sarcoma viral oncogene' shows the matching genes in human. (**C**) Ohnolog family page for HRAS gene in the human genome. (**D**) From the family page, users can navigate to ortholog families in other vertebrates, e.g. zebrafish HRASA. (**E**) Ohnolog pair page for zebrafish for NRAS gene. (**F**) Browse/Download page for zebrafish showing both 2R and 3R-ohnolog pairs and families for all the three criteria.

dated v2 version for human using relaxed criterion (Figure 4B), 36% are filtered out due to poor q-score, 38% due to duplication timing not being at the base of vertebrates, 25% due to changes in orthologs with outgroup genome(s) and 1% due to other Ensembl version related changes. We also compared ohnologs from the current study with ohnologs from Makino *et al*. (24) and Sacerdot *et al*. (41) studies that used different methodological approaches. All these datasets share a significant overlap between them. For example, using the relaxed criterion, 75% of v2 ohnologs are common to the three ohnolog datasets (Figure 4C). We noticed that out of 1089 ohnolog pairs identified by both Makino *et al*. and Sacerdot *et al*. studies but excluded by our analysis (Figure 4D), 56% are filtered out due to poor q-scores and 44% due to changes in ortholog/paralog relationships or Ensembl version related differences. These comparisons suggest that in addition to synteny, identification of the correct timing of duplication and homology relationships are also critical for accurate identification of ohnologs.

At last, the OHNOLOGS v2 database can be used to analyze the branch-specific loss and retention of ohnologs. For instance, we found that 1316 out of 2373 ohnolog families with relaxed confidence criterion in human had an identical size in nearly all the 18 mammals (i.e. corresponding to a variance over mean size ratio lower than 0.1 across all 18 mammals, where ohnolog family sizes are not affected by additional small-scale duplicates). Then, out of these 1316 conserved 2R-ohnolog families in mammals, 702 have an identical size in teleost fish, including 396 families which also share the same size in sauropsids while the remaining 306 families correspond mainly to ad-

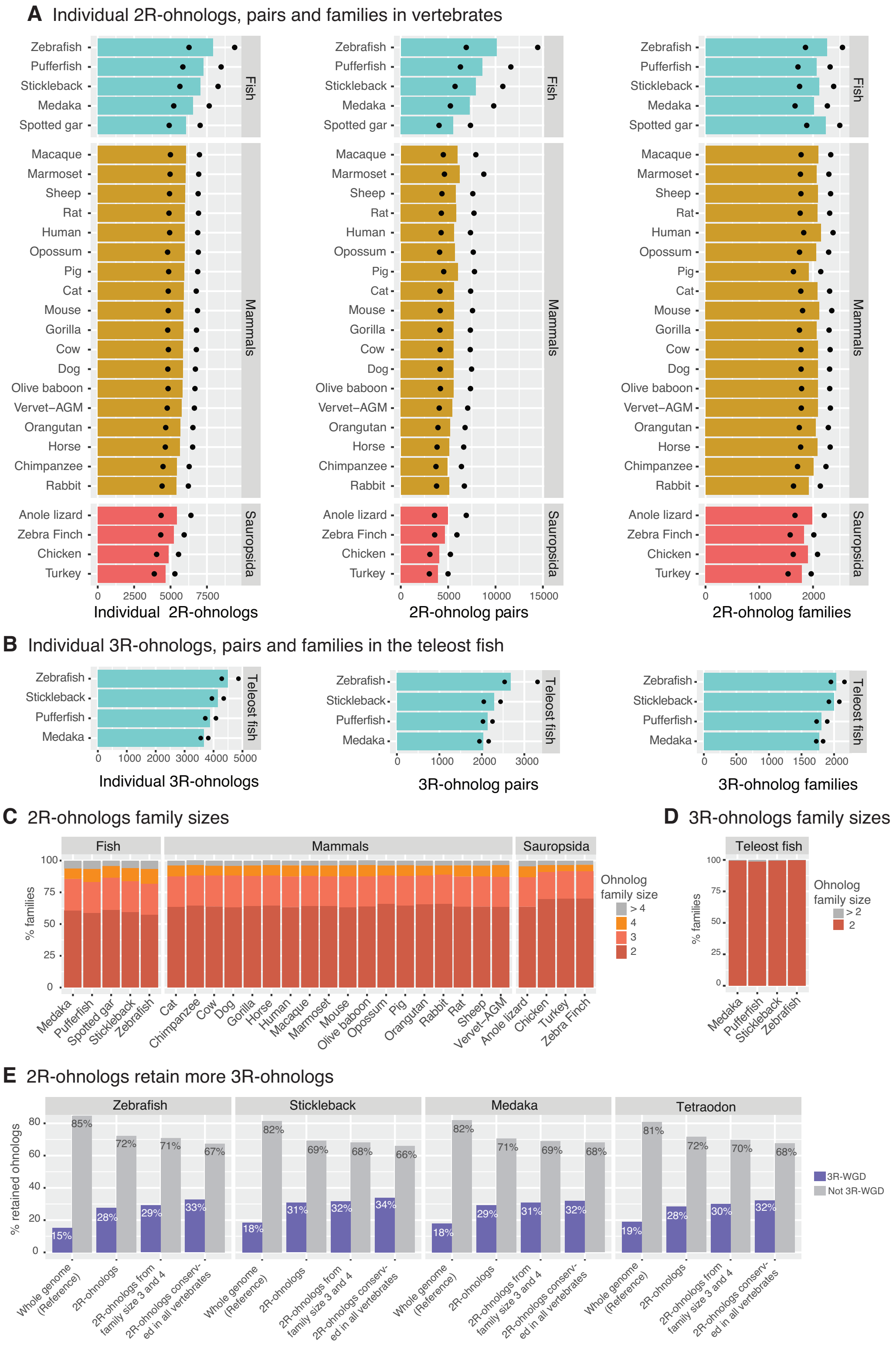


Figure 3. Description of the ohnolog genes, pairs and families in the database. (**A**) Number of retained individual 2R-ohnolog genes, pairs and families in all the 27 vertebrates. Bars represent the numbers from the intermediate criterion. Ohnologs from strict and relaxed criteria are indicated by dots. (**B**) Number of retained individual 3R-ohnolog genes, pairs and families in the four teleost fish species. Bars represent the numbers from the intermediate criterion. Ohnologs from strict and relaxed criteria are indicated by dots. (**C**) Size of the 2R-ohnolog families from the intermediate criterion in vertebrates. Note that a vast majority of the families are of size 2, 3 or 4. (**D**) Sizes of the 3R-ohnolog families from the intermediate criterion in the teleost fish hardly exceed size two. (**E**) The 2R-ohnologs are significantly more likely to retain 3R-ohnologs, compared to genome-average. The retention of 3R-ohnologs is even higher for the 2R-ohnologs that belong to family size 3 or 4, and for 2R-ohnologs conserved in all the 27 vertebrates. All the *P*-values are <1e-41, Chi-square test. Family counts are from the intermediate criterion.

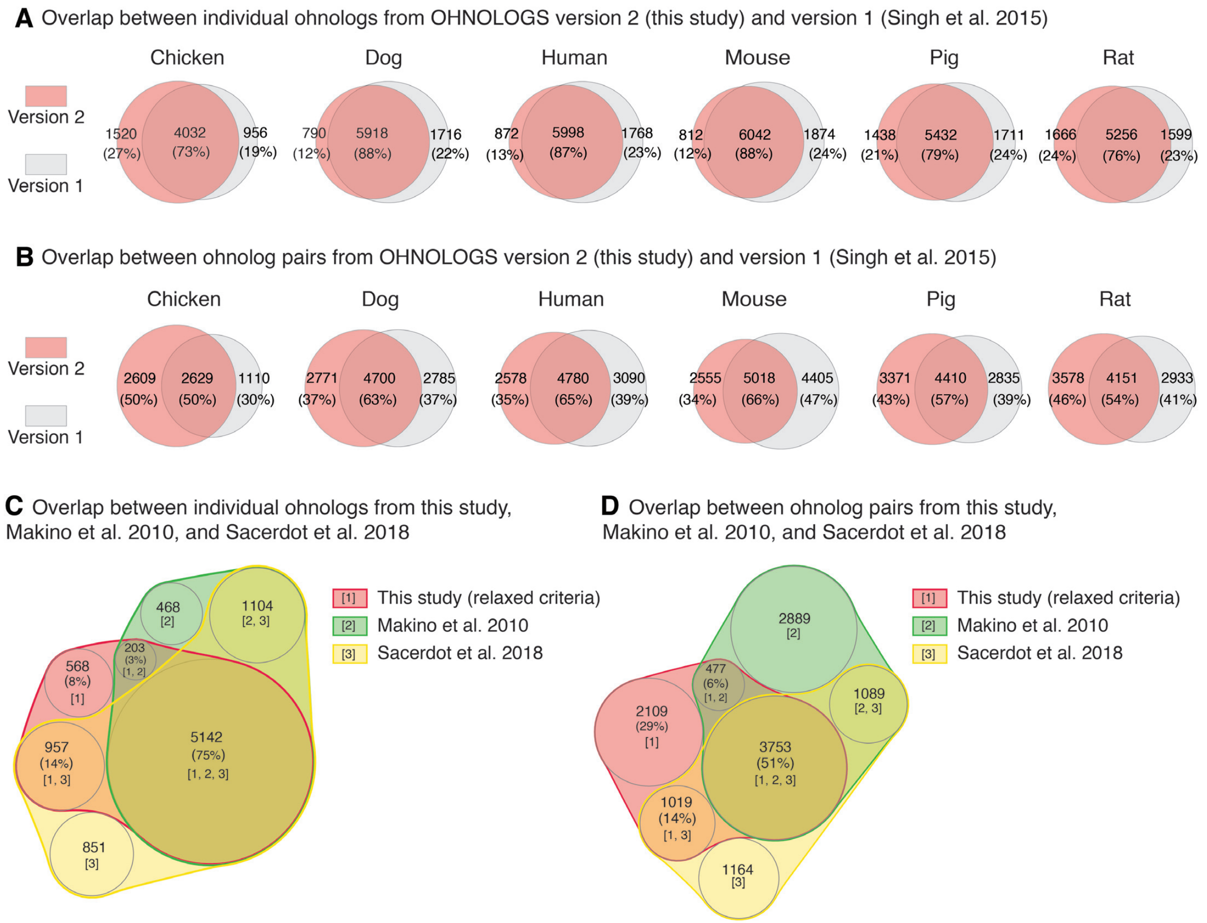


Figure 4. Comparison of ohnologs across different studies. (**A**) Comparison of individual ohnologs from OHNOLOGS v2 (this study) with v1 (34) for the six vertebrates already included in v1. (**B**) Comparison of ohnolog pairs from OHNOLOGS v2 (this study) and v1 (34). The majority of individual ohnologs and pairs are shared between both versions. (**C**) Overlap among individual ohnologs from this study, Makino *et al.* (24) and Sacerdot *et al.* (41). (**D**) Overlap among ohnolog pairs from this study, Makino *et al.* (24) and Sacerdot *et al.* (41). The majority of individual ohnologs and pairs are shared across the three studies. The venn diagram between three sets have been generated using nVenn (42).

ditional 2R-ohnolog losses in sauropsids; 119 families are larger in teleost fish and contain fish-specific 2R-ohnologs, while 86 families are smaller in teleost fish and correspond to 29 amniota-specific, 49 mammalia-specific and only 8 sauropsida-specific retentions of 2R-ohnologs.

CONCLUSION

The updated OHNOLOGS v2 database is a comprehensive resource for the genes retained from WGDs across 27 vertebrates. It includes ohnologs from both ancestral vertebrate 2R-WGDs and teleost fish specific 3R-WGD. It is based on a robust pipeline that downloads and processes datasets automatically using Ensembl, which makes it amenable to easy updates. We plan to expand and update OHNOLOGS periodically. Algorithmically, it is based on a quantitative comparative macro-synteny approach, which also takes into account the phylogenetically biased sampling of available vertebrate species. This approach assesses the confidence in each ohnolog pair and robustly identifies ohnologs, despite lineage specific genome rearrangement, gene loss and small-scale duplication events. Using the datasets in OHNOLOGS we show a greater lineage-specific ohnolog loss in sauropids compared to other vertebrate groups, and a high retention of 2R-ohnologs in subsequent 3R-WGD in teleost fish. In the light of the evolutionary significance of ancient WGDs and ohnologs for vertebrate evolution, the expanded and improved OHNOLOGS database should facilitate deeper comparative, evolutionary, genomic and functional analyses of the ohnolog genes in vertebrates.

DATA AVAILABILITY

All the data and code used to construct OHNOLOGS is available at http://ohnologs.curie.fr and its associated GitHub repository at https://github.com/param-p-singh/Ohnologs-v2.0.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We acknowledge technical support from the service informatique of Institut Curie for hosting and maintaining the server infrastructure. We thank Vincent Cabeli and Marcel Ribeiro Dantas for help in updating the server.

FUNDING

Erasmus Mundus (UPMC) PhD Fellowship (to P.P.S.); La Ligue Contre le Cancer fellowship (to P.P.S.). Funding for open access charge: Institutional Grant.

Conflict of interest statement. None declared.

REFERENCES

1. Ohno,S. (1970) *Evolution by Gene Duplication*. Springer-Verlag, NY.
2. Van de Peer,Y., Maere,S. and Meyer,A. (2009) The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.*, **10**, 725–732.
3. Van de Peer,Y., Mizrachi,E. and Marchal,K. (2017) The evolutionary significance of polyploidy. *Nat. Rev. Genet.*, **18**, 411–424.
4. Schwager,E.E., Sharma,P.P., Clarke,T., Leite,D.J., Wierschin,T., Pechmann,M., Akiyama-Oda,Y., Esposito,L., Bechsgaard,J., Bilde,T. *et al.* (2017) The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.*, **15**, 62.
5. Ohno,S., Wolf,U. and Atkin,N.B. (1968) Evolution from fish to mammals by gene duplication. *Hereditas*, **59**, 169–187.
6. Abi-Rached,L., Gilles,A., Shiina,T., Pontarotti,P. and Inoko,H. (2002) Evidence of en bloc duplication in vertebrate genomes. *Nat. Genet.*, **31**, 100–105.
7. Dehal,P. and Boore,J.L. (2005) Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.*, **3**, e314.
8. Putnam,N.H., Butts,T., Ferrier,D.E., Furlong,R.F., Hellsten,U., Kawashima,T., Robinson-Rechavi,M., Shoguchi,E., Terry,A., Yu,J.K. *et al.* (2008) The amphioxus genome and the evolution of the chordate karyotype. *Nature*, **453**, 1064–1071.
9. Christoffels,A., Koh,E.G., Chia,J.M., Brenner,S., Aparicio,S. and Venkatesh,B. (2004) Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol. Biol. Evol.*, **21**, 1146–1151.
10. Jaillon,O., Aury,J.M., Brunet,F., Petit,J.L., Stange-Thomann,N., Mauceli,E., Bouneau,L., Fischer,C., Ozouf-Costaz,C., Bernot,A. *et al.* (2004) Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature*, **431**, 946–957.
11. Meyer,A. and Van de Peer,Y. (2005) From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays*, **27**, 937–945.
12. Cañestro,C. (2012) *Two Rounds of Whole-Genome Duplication: Evidence and Impact on the Evolution of Vertebrate Innovations*. Springer, Berlin, Heidelberg
13. Holland,L.Z. (2009) Chordate roots of the vertebrate nervous system: expanding the molecular toolkit. *Nat. Rev. Neurosci.*, **10**, 736–746.
14. Hoegg,S., Brinkmann,H., Taylor,J.S. and Meyer,A. (2004) Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J. Mol. Evol.*, **59**, 190–203.
15. Semon,M. and Wolfe,K.H. (2007) Reciprocal gene loss between Tetraodon and zebrafish after whole genome duplication in their ancestor. *Trends Genet.*, **23**, 108–112.
16. Glasauer,S.M. and Neuhauss,S.C. (2014) Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genomics*, **289**, 1045–1060.
17. Taylor,J.S., Braasch,I., Frickey,T., Meyer,A. and Van de Peer,Y. (2003) Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Res.*, **13**, 382–390.
18. Marletaz,F., Firbas,P.N., Maeso,I., Tena,J.J., Bogdanovic,O., Perry,M., Wyatt,C.D.R., de la Calle-Mustienes,E., Bertrand,S., Burguera,D. *et al.* (2018) Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature*, **564**, 64–70.
19. Wolfe,K. (2000) Robustness–it's not where you think it is. *Nat. Genet.*, **25**, 3–4.
20. Huminiecki,L. and Heldin,C.H. (2010) 2R and remodeling of vertebrate signal transduction engine. *BMC Biol.*, **8**, 146.
21. Roux,J., Liu,J. and Robinson-Rechavi,M. (2017) Selective constraints on coding sequences of nervous system genes are a major determinant of duplicate gene retention in vertebrates. *Mol. Biol. Evol.*, **34**, 2773–2791.
22. Brunet,F.G., Volff,J.N. and Schartl,M. (2016) Whole genome duplications shaped the receptor tyrosine kinase repertoire of jawed vertebrates. *Genome Biol. Evol.*, **8**, 1600–1613.
23. Guo,B. (2017) Complex genes are preferentially retained after whole-genome duplication in teleost fish. *J. Mol. Evol.*, **84**, 253–258.
24. Makino,T. and McLysaght,A. (2010) Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 9270–9274.
25. Singh,P.P., Affeldt,S., Cascone,I., Selimoglu,R., Camonis,J. and Isambert,H. (2012) On the expansion of 'dangerous' gene repertoires by whole-genome duplications in early vertebrates. *Cell Rep.*, **2**, 1387–1398.
26. Dickerson,J.E. and Robertson,D.L. (2012) On the origins of Mendelian disease genes in man: the impact of gene duplication. *Mol. Biol. Evol.*, **29**, 61–69.
27. Malaguti,G., Singh,P.P. and Isambert,H. (2014) On the retention of gene duplicates prone to dominant deleterious mutations. *Theor. Popul. Biol.*, **93**, 38–51.
28. Singh,P.P., Affeldt,S., Malaguti,G. and Isambert,H. (2014) Human dominant disease genes are enriched in paralogs originating from whole genome duplication. *PLoS Comput. Biol.*, **10**, e1003754.
29. Tinti,M., Dissanayake,K., Synowsky,S., Albergante,L. and MacKintosh,C. (2014) Identification of 2R-ohnologue gene families displaying the same mutation-load skew in multiple cancers. *Open Biol.*, **4**, 140029.
30. Verny,L., Sella,N., Affeldt,S., Singh,P.P. and Isambert,H. (2017) Learning causal networks with latent variables from multivariate information in genomic data. *PLoS Comput. Biol.*, **13**, e1005662.
31. Catchen,J.M., Conery,J.S. and Postlethwait,J.H. (2009) Automated identification of conserved synteny after whole-genome duplication. *Genome Res.*, **19**, 1497–1505.
32. Muffato,M., Louis,A., Poisnel,C.E. and Roest Crollius,H. (2010) Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes. *Bioinformatics*, **26**, 1119–1121.
33. Jandzik,D., Garnett,A.T., Square,T.A., Cattell,M.V., Yu,J.K. and Medeiros,D.M. (2015) Evolution of the new vertebrate head by co-option of an ancient chordate skeletal tissue. *Nature*, **518**, 534–537.
34. Singh,P.P., Arora,J. and Isambert,H. (2015) Identification of ohnolog genes originating from whole genome duplication in early vertebrates, based on synteny comparison across multiple genomes. *PLoS Comput. Biol.*, **11**, e1004394.
35. Berthelot,C., Brunet,F., Chalopin,D., Juanchich,A., Bernard,M., Noel,B., Bento,P., Da Silva,C., Labadie,K., Alberti,A. *et al.* (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat. Commun.*, **5**, 3657.
36. Zerbino,D.R., Achuthan,P., Akanni,W., Amode,M.R., Barrell,D., Bhai,J., Billis,K., Cummins,C., Gall,A., Giron,C.G. *et al.* (2018) Ensembl 2018. *Nucleic Acids Res.*, **46**, D754–D761.
37. Durinck,S., Moreau,Y., Kasprzyk,A., Davis,S., De Moor,B., Brazma,A. and Huber,W. (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439–3440.
38. Durinck,S., Spellman,P.T., Birney,E. and Huber,W. (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.*, **4**, 1184–1191.
39. Herrero,J., Muffato,M., Beal,K., Fitzgerald,S., Gordon,L., Pignatelli,M., Vilella,A.J., Searle,S.M., Amode,R., Brent,S. *et al.* (2016) Ensembl comparative genomics resources. *Database*, **2016**, bav096.
40. Vilella,A.J., Severin,J., Ureta-Vidal,A., Heng,L., Durbin,R. and Birney,E. (2009) EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.*, **19**, 327–335.
41. Sacerdot,C., Louis,A., Bon,C., Berthelot,C. and Roest Crollius,H. (2018) Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biol.*, **19**, 166.
42. Perez-Silva,J.G., Araujo-Voces,M. and Quesada,V. (2018) nVenn: generalized, quasi-proportional Venn and Euler diagrams. *Bioinformatics*, **34**, 2322–2324.