D906–D912 *Nucleic Acids Research, 2023, Vol. 51, Database issue* *Published online 26 August 2022*
https://doi.org/10.1093/nar/gkac718

DupScan: predicting and visualizing vertebrate genome duplication database

Jianguo Lu [1,2,3,*], Peilin Huang[1], Jialiang Sun[4] and Jian Liu [4,5,*]

[1]School of Marine Sciences, Sun Yat-sen University, Zhuhai 519082, China, [2]Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519080, China, [3]Guangdong Provincial Key Laboratory of Marine Resources and Coastal Engineering, Guangzhou 510275, China, [4]College of Computer Science, Nankai University, Tianjin 300350, China and [5]Centre for Bioinformatics and Intelligent Medicine, Nankai University, Tianjin 300350, China

Received July 12, 2022; Revised July 24, 2022; Editorial Decision July 26, 2022; Accepted August 10, 2022

ABSTRACT

Duplicated genes prevail in vertebrates and are important in the acquisition of new genes and novelties. Whole genome duplication (WGD) is one of the sources of duplicated genes. It can provide raw materials for natural selection by increasing the flexibility and complexity of the genome. WGDs are the driving force for the evolution of vertebrates and contribute greatly to their species diversity, especially in fish species with complicated WGD patterns. Here, we constructed the DupScan database (https://dupscan.sysumeg.com/) by integrating 106 chromosomal-level genomes, which can analyze and visualize synteny at both the gene and genome scales, visualize the *Ka, Ks,* and 4DTV values, and browse genomes. DupScan was used to perform functional adaptation for the intricate WGD investigation based on synteny matching. DupScan supports the analysis of five WGD rounds (R): VGD2 (vertebrate genome duplication 2), Ars3R (*Acipenser-ruthenus*-specific 3R), Pss3R (*Polyodon-spathula*-specific 3R), Ts3R (teleost-specific duplication 3R), Ss4R (salmonid-specific 4R), and Cs4R (carp-specific 4R). DupScan serves as one-stop analysis platform for synteny and WGD research in which users can analyze and predict synteny and WGD patterns across 106 species of whole genome sequences. This further aided us in elucidating genome evolutionary patterns across over 60,000 vertebrate species with synteny and WGD events.

INTRODUCTION

Duplicated genes prevail across organisms (1). They provide genomes with the genetic backup that is important to the origin of novelties (2,3). Duplicated genes can be generated not only from single gene duplication but also from small-scale duplication (SSD) and whole-genome duplication (WGD) events (4). Notably, WGDs reshape genomes with much greater intensity than SSDs and single gene duplications because of their acquisition of a whole set of chromosomes. WGDs have been considered a significant contributor to speciation and species diversity in vertebrates (5,6).

Vertebrates, including mammals, have undergone 2R (two-round) WGD events (7). As the largest vertebrate group, Teleostei species then experienced another one to three rounds of WGD (6,8–16), leading to an astonishing diversity of morphology, physiology, and behavior, with ∼30 000 recorded species. This suggests that WGDs can provide Teleostei with a competitive edge over their diploid progenitors. However, the mechanisms of WGDs in evolution remain discussible. Additionally, teleosts and mammals possess a similar gene repertoire, and their basic body plan and developmental programs are similar. Therefore, understanding how vertebrates, especially teleosts, benefit from WGDs will help elucidate the molecular mechanisms underlying mammalian development and diseases.

Synteny, the degree to which homologous genes are retained on the corresponding chromosomes, is informative of the genome evolutionary pattern after WGD (17). However, it requires whole-genome alignments and further large-scale comparisons. Thus, WGD-targeted databases are necessary. To our knowledge, several databases related to synteny and WGDs have been established. For example, the Plant Genome Duplication Database (PGDD) (http://chibba.agtec.uga.edu/duplication/) (18) for plant WGDs was created in 2012. It is currently inaccessible due to lack of maintenance. The PhyloFish database (http://phylofish.sigenae.org/) (19) and the OHNOLOGS (http://ohnologs.curie.fr./) (20) were established as well. PhyloFish allows users to search and browse the gene families for 23 teleost species with Ts3R. OHNOLOGS consisted of 23 verte-

*To whom correspondence should be addressed. Tel: +86 756 3668927; Fax: +86 756 3668927; Email: lujianguo@mail.sysu.edu.cn
Correspondence may also be addressed to Jian Liu. Email: jianliu@nankai.edu.cn

brates with VGD2 and 4 fish with Ts3R. These current WGD databases do not support the analysis of WGD events of Ars3R, Pss3R, Ss4R and Cs4R in other vertebrates, which prevents us from investigating the subsequent process after WGDs. In addition, specialized visualization tools for WGDs are not implemented in these databases.

Here, we collected the chromosomal-level genomes of 106 vertebrate species and constructed the DupScan database (https://dupscan.sysumeg.com/). DupScan can analyze and visualize synteny at both the gene and genome scales, visualize the *Ka*, *Ks* and 4DTV values, and browse the genomes. In particular, we developed WGD-related features. DupScan covers vertebrates with five different WGD patterns (VGD2, Ars3R, Pss3R, Ts3R, Ss4R and Cs4R), so it presents comprehensive aspects of WGDs under the same evolutionary background. Visual tools are specifically designed in DupScan for analyzing and predicting WGD events. Given these features, the DupScan database can serve as an instrumental platform for genome evolutionary pattern research across over 60 000 vertebrate species.

MATERIALS AND METHODS

Data collection

We collected a total of 106 genomes of living organisms and their annotation information from NCBI (21) and Ensembl (22) (Figure 1, Supplementary Table S1). The species included 89 common teleost fish, 4 basal organisms of Actinopterygii, and 13 important vertebrates of mammals, amphibians and sharks with chromosome-level assemblies.

Data processing

The MCScanX toolkit (v0.8) (23) was implemented to define synteny blocks based on the MCScan algorithm (24), which was also used to build the PGDD (18). The scanning process of MCScanX was based on BLASTP (v2.10.0) alignment result (25). A total of 5 130 978 protein sequences were used to identify intraspecies and interspecies homology. There were 2 239 380 145 hits and 10 751 982 synteny blocks after the selection of the top five hits. Then, the synteny block redundancy was merged using an in-house script. For gene/genome visualization, JBrowse (v1.16.9) was implemented in the DupScan database (26). The *Ka* (the number of nonsynonymous substitutions per nonsynonymous site) and *Ks* (the number of synonymous substitutions per synonymous site) values of the synteny blocks were calculated with the YN estimation model using KaKs_Calculator Toolbox (v2.0) (27,28). The 4DTV (transversion rates on 4-fold degenerated sites) was calculated with HKY substitution models (29) using an in-house script (https://github.com/xiaolinfrank/batch_4DTV_calculation).

The phylogenomic tree of the 106 species (Figure 1B) was reconstructed based on gene trees identified by OrthoFinder (v2.5.2) (30). The common ancestor of *Scyliorhinus canicula* and *Carcharodon carcharias* was manually defined as the root. Subsequently, r8s (v1.81) was used to infer the species divergence time for each node in the species tree (Figure 1B) according to previous fossil-calibrated and data-deduced times (31–33). The positions of WGD events on the tree were defined based on previous studies that reported the occurrence of WGDs (6,8–16) and were calibrated by our divergence time inference. The controversy about the time of the Ars3R event was clarified based on our 4DTV analysis.

Website implementation

Our website was built using the Django framework based on Python and deployed using the uWSGI server and Nginx reverse proxy server. We used the MVC framework to decouple codes and separate codes into the controller layer, the business logic layer, and the DAO layer on the server side. The client-side webpage was implemented using HTML, CSS, JavaScript and jQuery.

RESULTS

Database description

DupScan is a database for visualizing and predicting vertebrate synteny and genome duplication. It provides specialized functions to analyze computation-consuming WGD events by visualizing macro-synteny/micro-synteny, chromosome structures, gene/gene family evolution patterns and *Ka*/*Ks*/4DTV distributions (Figure 1A). DupScan now incorporates 106 species, and their WGD patterns are diverse (Table 1, Figure 1B, Supplementary Table S1). A total of 15 species have experienced only 2R WGD events (VGD1 and VGD2), including mammals, amphibians, sharks, and ancient actinopterygian fish. Two Acipenseriformes species had Ars3R and Pss3R duplication patterns. A total of 89 teleost fish species have experienced an additional WGD event (i.e. Ts3R). Interestingly, ten salmonids and two carps evolved from another round of WGD after Ts3R, which are called Ss4R and Cs4R, respectively.

A massive amount of genetic data was integrated into the DupScan database, including 5 130 978 CDSs, 6 195 741 synteny blocks and 85 841 593 pairs of homologs. Moreover, the *Ka*, *Ks*, *Ka*/*Ks* and 4DTV metadata for over 67 million homogeneous gene pairs were calculated and deposited in the DupScan database for further duplication pattern evolutionary analysis (Table 1).

From macro-synteny to micro-syteny across genomes

The users can browse the whole genome-wide homologs across all chromosomes with colored curves using the Circos Plot function (Figure 2A). To analyze the genome-wide synteny, the Dot Plot function aims to visualize the interchromosomal homologous gene pairs (Figure 2B). The Dot Plot function was integrated with the Genome Browser function for users to accurately locate the specific synteny area. The dot plots are interactive and can be zoomed in or viewed in Genome Browser by simply dragging the mouse. The range of gene pairs can be adjusted by the '*Ks* filter' option.

Case 1. Figure 2 shows a case study conducted from macro-synteny to micro-synteny. Sterlets (*Acipenser ruthenus*) experienced one more round of WGD than humans. After quick selection, the homology between sterlets

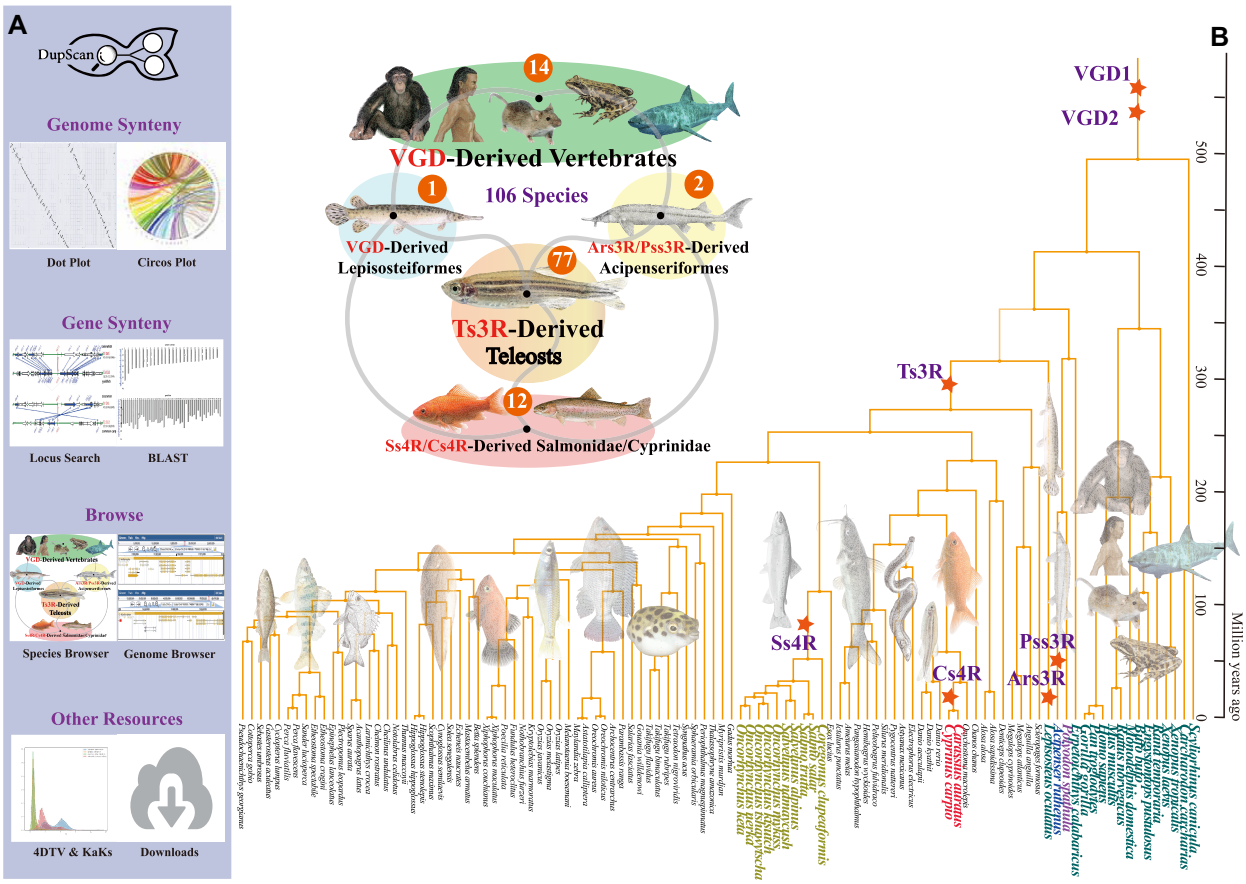

Figure 1. Overview of the DupScan database. (**A**) Main functions of DupScan. (**B**) Phylogenetic tree for all recorded species. The tree was reconstructed based on the orthogroups and orthologs identified by OrthoFinder. Each pentagram mark: VGD: vertebrate-specific WGD; Ars3R: *Acipenser-ruthenus*-specific 3R; Pss3R: *Polyodon-spathula*-specific 3R; Ts3R: teleost-specific 3R; Ss4R: salmonid-specific 4R; Cs4R: carp-specific 4R; R: round of WGD event.

Table 1. DupScan statistics summary

Feature	Number
Order/family/genus/species	42/65/86/106
2R/3R/4R species	15/83/8
Total genome size	141.93 GB
Coding DNA sequences	5 130 978
Synteny blocks	6 195 741
Sequences in synteny blocks	3 469 841
Homolog pairs in synteny blocks	85 841 593
Homolog pairs with Ka/Ks/KaKs	67 369 341
Homolog pairs with 4DTV	67 223 684

and humans at the chromosomal level can be displayed in a Circos plot (Figure 2A). In a dot plot, this relationship is shown as dots (Figure 2B). By dragging the left mouse, we observed a zoomed-in plot (Figure 2B–D). The Genome Browser windows are displayed by dragging the right mouse button (Figure 2E), providing additional features such as browse and search functions. Users can click the 'View synteny blocks' links to locate the genes in other blocks (Figure 2F). In addition, the Genome Browser can be accessed as a stand-alone feature to compare many chromosomes from 106 species with a multitrack view. The users can identify more patterns of WGDs from the DupScan database.

Detection of potential WGD-derived paralogs

The local BLAST tool was implemented for gene duplication detection based on sequence similarity. We developed a feature of gene duplication events by showing the BLAST hits not only in tables but also in chromosome bar plots (Figure 3A). It can quickly display the gene either by clicking 'View Genome Browser' or clicking 'View synteny blocks' to switch to the Locus Search. The duplication analysis tool Locus Search takes a sequence ID as input and searches all synteny blocks containing the query. Two paths can be used to retrieve the sequence IDs by the 'Browse - Sequence Browser' function or by the quick links in other functions.

Case 2. BLAST and Locus Search were designed to scan duplicates originating from duplications across species. The key point for these two functions is to elucidate the specific gene duplication pattern across all vertebrate species with their evolutionary timeline. As an example, users can search for the orthologs of the *sox1* gene with the BLAST and Locus Search function (Figure 3). The BLAST hits on milkfish (*Chanos chanos*) are indicated on the chromosome bar plot (Figure 3A). Then, click the quick link to search for the best hit of the *sox1* gene in the Locus Search (Figure 3B). Four hits were shown on river trout (*Salmo trutta*) chromosomes

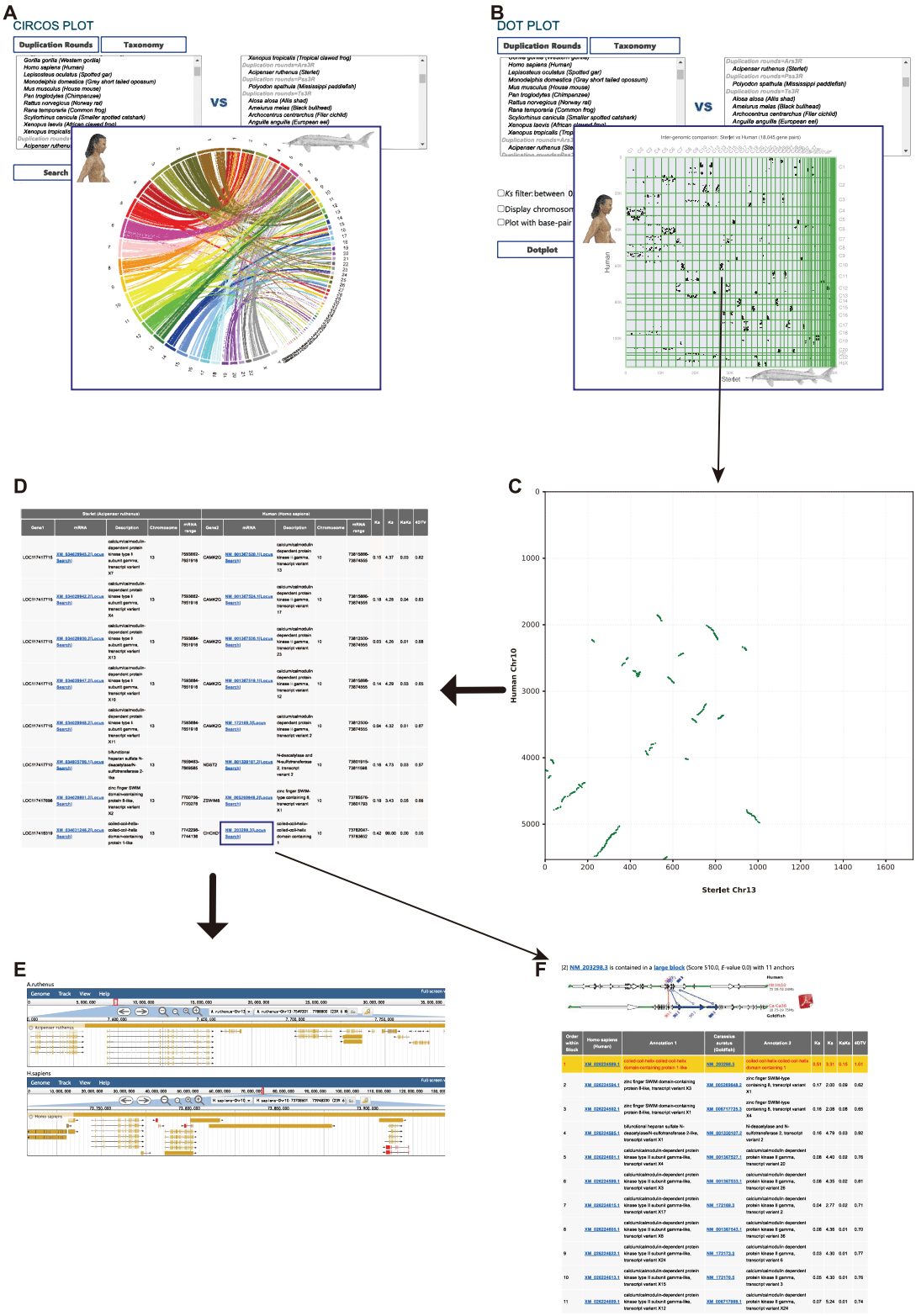

Figure 2. The case of macro-synteny. Sterlets (*Acipenser ruthenus*) experienced one more round of WGD than humans. (**A**) A Circos plot example. The curves connect orthologs between the chromosomes of humans and sterlets. (**B**) An example of a dot plot. X and Y axes represent chromosomes. Each black dot represents a pair of homologous genes and is plotted with gene ranks by default. Plotting with base-pair distance is optional. (**C**) Enlarged view after dragging the mouse on the dot plot. (**D**) Genes are shown with the dot plot being further enlarged. (**E**) The JBrowse windows are displayed below to visualize the corresponding gene. (**F**) Click on 'Locus Search' to quickly search micro-synteny blocks. The pictures of sterlet and human were edited afterward.

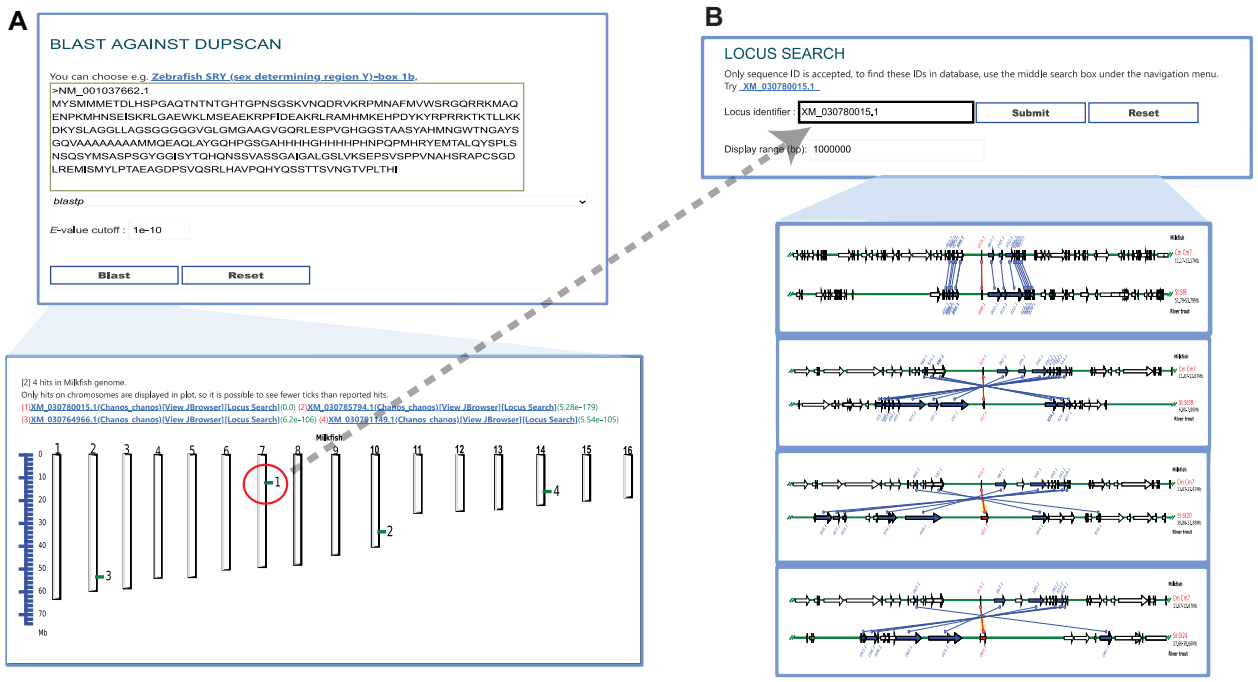

Figure 3. The case of micro-synteny for the *sox1* gene. (**A**) The BLAST result on milkfish (*Chanos chanos*) chromosomes for the zebrafish *sox1b* gene. (**B**) The Locus Search result on river trout (*Salmo trutta*) for the best hit in (A).

6, 39, 20 and 24. We speculated that these four hits might be evidence of the Ts3R and Ss4R WGD events.

Scanning and predicting the WGD signals

Users can visualize the distribution patterns of *Ka*, *Ks*, *Ka*/*Ks* and 4DTV across multiple species. In particular, the values of *Ks* and 4DTV can be used for WGD pattern research. As an example of the 4DTV distribution pattern, several genome duplication events can be shown in a single plot (Figure 4). Each major peak on the plot is related to a genome duplication event. The paralogs' 4DTV distributions can be affected by WGD because it produces paralogs with zero 4DTV values. Consequently, the occurrence of WGD is indicated as a peak on the curve (the detailed method can be found on the DupScan Document on the Help web page).

Case 3. The 4DTV distribution pattern can indicate signals of WGD events. The detailed method can be found in the DupScan Document on the Help page of the database. The Ars3R duplication event occurred at 21.3 Mya (16) or at 180 Mya (15). To resolve the controversy of estimation time, all the 4DTV distribution patterns of Ars3R, Pss3R, and Ss4R in Acipenseriformes and Salmoniformes are shown in Figure 4. The peaks of Ars3R appear to the left of Ss4R and Pss3R, suggesting that Ars3R must have occurred later than Ss4R and Pss3R. According to the 4DTV values of the Ars3R and Pss3R peaks, as well as the time of Pss3R, the time of Ars3R should be 17.6 Mya.

DISCUSSION

DupScan is not only a synteny visualization platform but also an enlightening and explorable research tool for the complicated WGD patterns in vertebrates. It presents comprehensive aspects of WGDs of the VGD1/2, Ars3R, Pss3R, Ts3R, Ss4R and Cs4R events under the same WGD evolutionary background. The timing of WGD events varies greatly within different studies because the background value of evolutionary time was missing. For example, we predicted the more precise evolutionary times of Ars3R using the DupScan database. WGDs have been considered to contribute significantly to vertebrate species diversity, especially for Ts3R-derived fish species. The DupScan database can provide fundamental assistance by supporting the analysis of recent WGD events, such as Ss4R.

The occurrence of WGD events is hard to define based on the number of gene families. That is why we need genome-level synteny. An essential role of the dot plot is to help us detect evidence of WGD events at the genome level, especially for ancient events. Nevertheless, the biggest advantage of our database is that the analysis is well supported for recently occurring and reported WGD events. For example, Ss4R is a proper case to study the significant contribution of WGDs to species diversity. In the past, Ts3R-derived fish species were the focus of many studies. However, the Ts3R event is too ancient, so weak genetic changes might be ignored. In the DupScan database, ten species radiated rapidly after Ss4R, making it possible for our users to contribute more to this topic. For novel WGDs, the corresponding method was standardized to scan and predict the WGD signals based on 4DTV distribution curves.

Currently, vertebrate genomes are being updated and produced rapidly. So we are also developing a pipeline to include more and the latest assemblies. We will continue to follow the progress of vertebrate genome research and update the DupScan database annually. DupScan will serve as

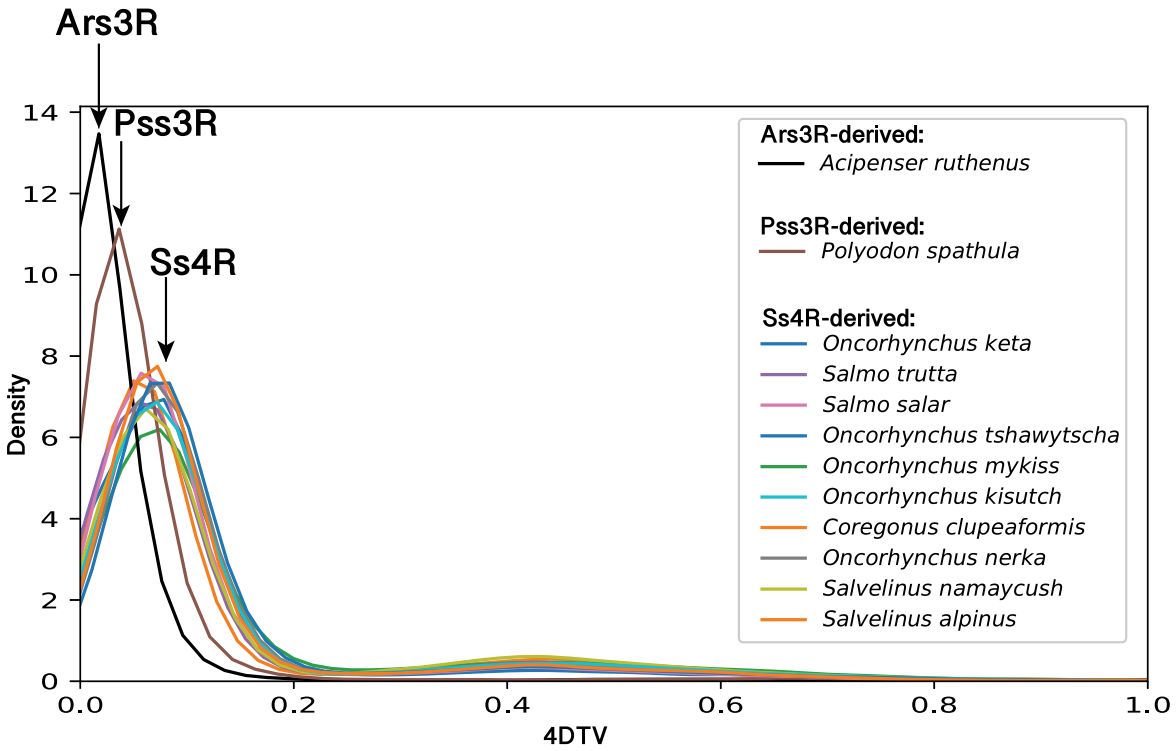

Figure 4. The case of the paralogs' 4DTV distribution curves, whose peaks denote the WGD events. The black arrows, tags, and labels were edited afterward. Ars3R: *Acipenser-ruthenus*-specific 3R; Pss3R: *Polyodon-spathula*-specific 3R; Ss4R: salmonid-specific 4R; R: round of WGD event.

a fundamental analytical platform for genomic and evolutionary studies of over 60,000 vertebrate species.

DATA AVAILABILITY

DupScan is freely accessible at https://dupscan.sysumeg.com/.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Thanks Elsevier for the language editing services.

FUNDING

National Key R&D Program of China [2021YFC2100800, 2021YFC2100801, 2020YFA0908700, 2020YFA0908702]; National Natural Science Foundation of China [91858208, 32002366, 31902427]; Guangzhou Science and Technology Project [201803020017]; R&D Project for Jinwan Yellowfin Seabream Breeding System Construction [K20-42000-018]; Science and Technology Project of Zhanjiang [2019A03011]; Project supported by Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) [311020005]. Funding for open access charge: National Natural Science Foundation of China [91858208].

Conflict of interest statement. None declared.

REFERENCES

1. Lynch,M. and Conery,J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science*, **290**, 1151–1155.
2. Conant,G.C. and Wolfe,K.H. (2008) Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.*, **9**, 938–950.
3. Spillane,C., Schmid,K.J., Laoueille-Duprat,S., Pien,S., Escobar-Restrepo,J.M., Baroux,C., Gagliardini,V., Page,D.R., Wolfe,K.H. and Grossniklaus,U. (2007) Positive darwinian selection at the imprinted MEDEA locus in plants. *Nature*, **448**, 349–352.
4. Kuzmin,E., Taylor,J.S. and Boone,C. (2022) Retention of duplicated genes in evolution. *Trends Genet.*, **38**, 59–72.
5. Van de Peer,Y., Mizrachi,E. and Marchal,K. (2017) The evolutionary significance of polyploidy. *Nat. Rev. Genet.*, **18**, 411–424.
6. Van de Peer,Y., Maere,S. and Meyer,A. (2009) The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.*, **10**, 725–732.
7. Albalat,R. and Canestro,C. (2016) Evolution by gene loss. *Nat. Rev. Genet.*, **17**, 379–391.
8. Ohno,S., Wolf,U. and Atkin,N.B. (1968) Evolution from fish to mammals by gene duplication. *Hereditas*, **59**, 169–187.
9. 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.
10. Kasahara,M., Naruse,K., Sasaki,S., Nakatani,Y., Qu,W., Ahsan,B., Yamada,T., Nagayasu,Y., Doi,K., Kasai,Y. *et al.* (2007) The medaka draft genome and insights into vertebrate genome evolution. *Nature*, **447**, 714–719.
11. 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.
12. Lien,S., Koop,B.F., Sandve,S.R., Miller,J.R., Kent,M.P., Nome,T., Hvidsten,T.R., Leong,J.S., Minkley,D.R., Zimin,A. *et al.* (2016) The Atlantic salmon genome provides insights into rediploidization. *Nature*, **533**, 200–205.

13. Xu,P., Xu,J., Liu,G., Chen,L., Zhou,Z., Peng,W., Jiang,Y., Zhao,Z., Jia,Z., Sun,Y. *et al.* (2019) The allotetraploid origin and asymmetrical genome evolution of the common carp cyprinus carpio. *Nat. Commun.*, **10**, 4625.
14. Xu,P., Zhang,X., Wang,X., Li,J., Liu,G., Kuang,Y., Xu,J., Zheng,X., Ren,L., Wang,G. *et al.* (2014) Genome sequence and genetic diversity of the common carp, cyprinus carpio. *Nat. Genet.*, **46**, 1212–1219.
15. Du,K., Stock,M., Kneitz,S., Klopp,C., Woltering,J.M., Adolfi,M.C., Feron,R., Prokopov,D., Makunin,A., Kichigin,I. *et al.* (2020) The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. *Nat. Ecol. Evol.*, **4**, 841–852.
16. Cheng,P., Huang,Y., Lv,Y., Du,H., Ruan,Z., Li,C., Ye,H., Zhang,H., Wu,J., Wang,C. *et al.* (2021) The american paddlefish genome provides novel insights into chromosomal evolution and bone mineralization in early vertebrates. *Mol. Biol. Evol.*, **38**, 1595–1607.
17. Parey,E., Louis,A., Cabau,C., Guiguen,Y., Roest Crollius,H. and Berthelot,C. (2020) Synteny-Guided resolution of gene trees clarifies the functional impact of whole-genome duplications. *Mol. Biol. Evol.*, **37**, 3324–3337.
18. Lee,T.H., Tang,H., Wang,X. and Paterson,A.H. (2013) PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res.*, **41**, D1152–D1158.
19. Pasquier,J., Cabau,C., Nguyen,T., Jouanno,E., Severac,D., Braasch,I., Journot,L., Pontarotti,P., Klopp,C., Postlethwait,J.H. *et al.* (2016) Gene evolution and gene expression after whole genome duplication in fish: the phylofish database. *BMC Genom*, **17**, 368.
20. Singh,P.P. and Isambert,H.J.N.a.r. (2020) OHNOLOGS v2: a comprehensive resource for the genes retained from whole genome duplication in vertebrates. *Nucleic Acids Res.*, **48**, D724–D730.
21. Sayers,E.W., Beck,J., Bolton,E.E., Bourexis,D., Brister,J.R., Canese,K., Comeau,D.C., Funk,K., Kim,S., Klimke,W. *et al.* (2021) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **49**, D10–D17.
22. Cunningham,F., Allen,J.E., Allen,J., Alvarez-Jarreta,J., Amode,M.R., Armean,I.M., Austine-Orimoloye,O., Azov,A.G., Barnes,I., Bennett,R. *et al.* (2022) Ensembl 2022. *Nucleic Acids Res.*, **50**, D988–D995.
23. Wang,Y., Tang,H., Debarry,J.D., Tan,X., Li,J., Wang,X., Lee,T.H., Jin,H., Marler,B., Guo,H. *et al.* (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.*, **40**, e49.
24. Tang,H., Wang,X., Bowers,J.E., Ming,R., Alam,M. and Paterson,A.H. (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.*, **18**, 1944–1954.
25. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinform.*, **10**, 421.
26. Buels,R., Yao,E., Diesh,C.M., Hayes,R.D., Munoz-Torres,M., Helt,G., Goodstein,D.M., Elsik,C.G., Lewis,S.E., Stein,L. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.
27. Yang,Z. and Nielsen,R. (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.*, **17**, 32–43.
28. Wang,D., Zhang,Y., Zhang,Z., Zhu,J. and Yu,J. (2010) KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics*, **8**, 77–80.
29. Hasegawa,M., Kishino,H. and Yano,T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.*, **22**, 160–174.
30. Emms,D.M. and Kelly,S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.*, **20**, 238.
31. Peng,Z., Ludwig,A., Wang,D., Diogo,R., Wei,Q. and He,S. (2007) Age and biogeography of major clades in sturgeons and paddlefishes (Pisces: acipenseriformes). *Mol. Phylogenet. Evol.*, **42**, 854–862.
32. Near,T.J., Eytan,R.I., Dornburg,A., Kuhn,K.L., Moore,J.A., Davis,M.P., Wainwright,P.C., Friedman,M. and Smith,W.L. (2012) Resolution of ray-finned fish phylogeny and timing of diversification. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 13698–13703.
33. Chen,Z., Omori,Y., Koren,S., Shirokiya,T., Kuroda,T., Miyamoto,A., Wada,H., Fujiyama,A., Toyoda,A., Zhang,S. *et al.* (2019) De novo assembly of the goldfish (Carassius auratus) genome and the evolution of genes after whole-genome duplication. *Sci. Adv.*, **5**, eaav0547.