BMC Bioinformatics

SOFTWARE Open Access

POInT$_{\text{browse}}$: orthology prediction and synteny exploration for paleopolyploid genomes

Mustafa Siddiqui[1] and Gavin C. Conant[1,2,3*]

*Correspondence: gconant@ncsu.edu

[1] Department of Biological Sciences, North Carolina State University, Raleigh, NC, USA
[2] Bioinformatics Research Center, North Carolina State University, Raleigh, NC, USA
[3] Program in Genetics, North Carolina State University, Raleigh, NC, USA

Abstract

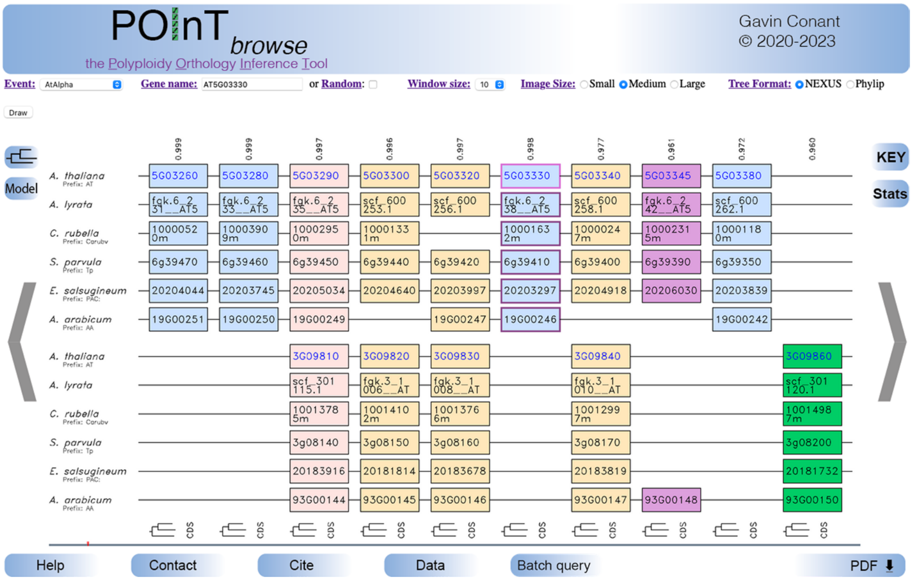
We describe POInT$_{\text{browse}}$, a web portal that gives access to the orthology inferences made for polyploid genomes with POInT, the Polyploidy Orthology Inference Tool. Ancient, or paleo-, polyploidy events are widely distributed across the eukaryotic phylogeny, and the combination of duplicated and lost duplicated genes that these polyploidies produce can confound the identification of orthologous genes between genomes. POInT uses conserved synteny and phylogenetic models to infer orthologous genes between genomes with a shared polyploidy. It also gives confidence estimates for those orthology inferences. POInT$_{\text{browse}}$ gives both graphical and query-based access to these inferences from 12 different polyploidy events, allowing users to visualize genomic regions produced by polyploidies and perform batch queries for each polyploidy event, downloading genes trees and coding sequences for orthologous genes meeting user-specified criteria. POInT$_{\text{browse}}$ and the associated data are online at https://wgd.statgen.ncsu.edu.

Keywords: POInT, Comparative genomics, Browser

Background

Ancient polyploidy events are widely distributed across the eukaryotic tree [1]. At the time of their formation, polyploid organisms have four (or more) complete sets of chromosomes in their nucleus [2], which can be thought of as a duplication of every gene in the genome (hence whole-genome duplication or WGD). This fully duplicated state is transitory and followed by the rapid loss of many of these duplicated genes [3]. Such losses may occasionally be due to selection [4] but probably most commonly occur through neutral processes [5, 6]. The losses can also occur both prior to or after speciation events among the taxa sharing the polyploidy. Losses result in a distinct pattern of double-conserved synteny (DCS) between the surviving genomes (Fig. 1), where the pre-polyploidy genome order can be reconstructed by merging the two duplicated regions, each of which preserves a fraction of the original gene content. Many of these events are *allopolyploidies*, meaning that the genomes that merged were not identical, making the event a combination of a hybridization and a genome doubling. For such events, it is

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.



Event: AtAlpha Gene name: AT5G03330 or Random: □ Window size: 10 Image Size: ○Small ●Medium ○Large Tree Format: ●NEXUS ○Phylip

Draw

	0.999	0.999	0.997	0.996	0.997	0.998	0.977	0.961	0.972	0.960
A. thaliana Prefix: AT	5G03260	5G03280	5G03290	5G03300	5G03320	5G03330	5G03340	5G03345	5G03380	
A. lyrata	fgk.6_2 31__AT5	fgk.6_2 33__AT5	fgk.6_2 35__AT5	scf_600 253.1	scf_600 256.1	fgk.6_2 38__AT5	scf_600 258.1	fgk.6_2 42__AT5	scf_600 262.1	
C. rubella Prefix: Carubv	1000052 0m	1000390 9m	1000295 0m	1000133 1m		1000163 2m	1000024 7m	1000231 5m	1000118 0m	
S. parvula Prefix: Tp	6g39470	6g39460	6g39450	6g39440	6g39420	6g39410	6g39400	6g39390	6g39350	
E. salsugineum Prefix: PAC:	20204044	20203745	20205034	20204640	20203997	20203297	20204918	20206030	20203839	
A. arabicum Prefix: AA	19G00251	19G00250	19G00249		19G00247	19G00246			19G00242	
A. thaliana Prefix: AT			3G09810	3G09820	3G09830		3G09840			3G09860
A. lyrata			scf_301 115.1	fgk.3_1 006__AT	fgk.3_1 008__AT		fgk.3_1 010__AT			scf_301 120.1
C. rubella Prefix: Carubv			1001378 5m	1001410 2m	1001376 6m		1001299 7m			1001498 7m
S. parvula Prefix: Tp			3g08140	3g08150	3g08160		3g08170			3g08200
E. salsugineum Prefix: PAC:			20183916	20181814	20183678		20183819			20181732
A. arabicum Prefix: AA			93G00144	93G00145	93G00146		93G00147	93G00148		93G00150

Fig. 1 An example visualization from POInT$_{\text{browse}}$. Shown is a region of ten duplicated regions ("pillars") with between 6 and 12 surviving genes from six species sharing the At-α polyploidy event. The view is centered on the selected gene from *Arabidopsis thaliana,* AT5G03330 (pink outline). Users control the visualization size, window size and the format of the downloaded gene trees from the top controls. The tree pictogram on the upper left gives the assumed species phylogeny, while the "Model" button gives a visualization of the duplicate loss model used for the event (including model parameters). At right, the "Key" button illustrates the color scheme while the "Stats" button gives the POInT$_{\text{browse}}$ version, the size of the current dataset and the sizes of all datasets currently in POInT$_{\text{browse}}$. Users can navigate ½ frame left or right with the arrows or re-center the frame on a pillar by clicking on it. Hovering over a gene gives its chromosomal coordinates and its common name (if known). Gene names shown in blue link to the corresponding model organism gene database entry for that gene. The location of the current frame relative to the full set of pillars is shown with the red region in the blue-gray bar at the bottom. This bar can also be used for coarse navigation within the pillars of an event. The upper panel with some blue genes shows the less fractionated subgenome, the bottom, the more fractionated one (green). Light pink genes are fully retained as duplicates and the darker pink pillar illustrates a reciprocal gene loss. Pillars with a mix of duplicated and single-copy genes are shown in tan. Numbers at the top of each pillar are POInT's confidence estimate (0.1) for the orthology relationships shown (see text). At the bottom of each pillar, the tree pictogram will download a gene tree with the corresponding orthology relationships for the genes in that pillar; the "CDS" button will download the coding sequences of the genes in question. A PDF version of the current window can be downloaded from the "PDF" button at right; the "Batch query" button opens a new window with the batch download interface

common to observe that one of the progenitor genomes is favored among the surviving single-copy genes, a pattern known as *biased fractionation* [7]. This pattern is illustrated in Fig. 1: the excess of blue columns relative to green ones is the result of duplicate losses more commonly coming from the lower subgenome than from the upper one.

Both duplicate losses and biased fractionation introduce complications for comparative genomics. Although DCS patterns are evident in any polyploid genome, it can be difficult to determine which region of any such genome is orthologous to a given region in a related genome [8]. For a genome duplication (tetraploidy) shared by n genomes, there are 2^n possible orthology relationships at each locus ("pillar" in Fig. 1). As shown in Fig. 2, the potential for independent duplicate gene losses in different genomes sharing a polyploid event can make identifying the "true" orthology relationship difficult. This difficulty can confound functional analyses, phylogenetics and studies in molecular evolution.

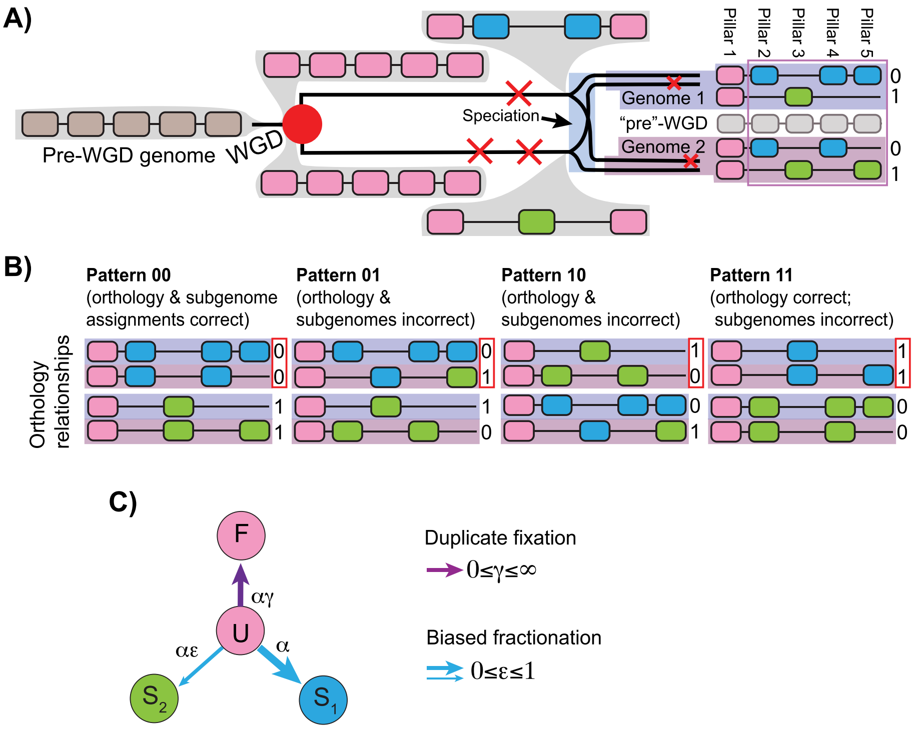


Fig. 2 Polyploidy, genome evolution and the POInT computation. **A** A polyploidy event followed by a speciation and duplicate gene losses produces regions of double-conserved synteny (DCS) in the two resulting genomes, raising the issue of phasing those regions with respect to each other. A region of five genes (brown) in the non-polyploid ancestor is duplicated (pink) and experiences three duplicate gene losses prior to a speciation event (red "X"s). After the speciation event, the two resulting species also experience independent losses, yielding the blocks of DCS in each genome. **B** There are $2^n = 4$ possible orthology relationships for the DCS blocks in these $n = 2$ genomes. These relationships are shown for the example (known) loss patterns from **A**. **C** For each orthology relationship in **B**, the likelihood of the observed presence/absence data at that pillar can be computed on the species phylogeny with a Markov model of duplicate losses. Those likelihoods can then be conditioned on the other pillars in the dataset. In this model, undifferentiated duplicate genes (*U*) can either be fixed (*F*) or lost from subgenome 2 (copy 1 or S_1 survives) or lost from subgenome 1 (S_2 survives). As the ε parameter ($0 \leq \varepsilon \leq 1$) approaches 0, subgenome 1 is increasingly favored over subgenome 2. Model parameters and tree branch lengths are estimated from the pillar data by maximum likelihood [9]. From this model, the relative likelihood of each of the orthology relationships in **B**, conditional on the full dataset, can be computed: these values are the confidence estimates at the top of the pillars in Fig. 1

To address this problem, we developed POInT (the Polyploidy Orthology Inference Tool), a phylogenetic modeling approach to studying shared polyploidies [8]. POInT uses a hidden Markov model to combine a phylogenetic model of duplicate loss after polyploidy with synteny information to infer which of these 2^n possible orthology relationships is most likely. The POInT computation has been described several times [8, 10, 11]. In Fig. 2, we give a cartoon overview. The polyploidy event leaves DCS as its hallmark. Duplicate gene losses leave "holes" in the DCS blocks that may be common to all species with the event or restricted to some clades (Fig. 2A). Since for real genomes we cannot know the true history (as we do for Fig. 2A), we employ a user-specified model of duplicate gene loss (Fig. 2C) to compute the likelihood of every possible orthology relationship (Fig. 2B) at every pillar, conditioned on all possible relationships at every other pillar and their syntenic relationships. At each pillar, the confidence in the inferred orthology relationship in Fig. 1 is then simply the likelihood of that orthology relationship at that pillar, conditional on every other pillar, over the total likelihood of the dataset. These confidence values are noted at the top of each pillar in Fig. 1.

Further development of POInT allowed us to model genome triplications (hexaploidy events) and biased fractionation [11, 12]. POInT now provides a statistical framework for testing hypotheses such as the presence and strength of biased fractionation and whether pairs of single-copy genes in different genomes are orthologs or are paralogs created by losses of alternative copies of the duplicate pair. Here we describe the $\text{POInT}_{\text{browse}}$ portal (wgd.statgen.ncsu.edu), which gives access to all of these data both for browsing and for download.

Construction and content

POInT is written in c++with dependencies on the LAPACK linear algebra libraries [13] and the GNU plotutils package; it is parallelized with OpenMP [14]. $\text{POInT}_{\text{browse}}$ is a c++CGI front-end that communicates with daemonized copies of POInT through UNIX interprocess communication. Hence, each running copy of POInT stores the computed orthology inferences for particular polyploidy event. When the CGI frontend sends a request for a browser frame from a given event, the appropriate POInT instance determines the best orthology relationship for each pillar in the requested window. It then creates the visualization in PNG format and returns that image to the browser. The generation of gene trees is handled in a similar manner.

To date, we have used POInT to analyze twelve polyploidy events, comprising 59 genomes and > 600,000 coding genes (Fig. 3), all available from $\text{POInT}_{\text{browse}}$. Of these twelve events, analyses of ten have been previously published, including the yeast WGD [15], the At-α event in *A. thaliana* and its relatives and the grass ρ event [11], the teleost genome duplication [16], hexaploidies in Brassiceae [12] and Solanaceae [17], a triploidy in parasitic nematodes [18] and WGD events in salmonids, paramecia and legumes [5]. The $\text{POInT}_{\text{browse}}$ documentation gives accession numbers and genome publication references for all twelve events.

$\text{POInT}_{\text{browse}}$ has three core functions. First, users can enter a gene identifier from one of the polyploid genomes and generate a visualization of the genomic region around that gene, including the corresponding orthologous and paralogous region(s) in the other polyploid genomes modeled (Fig. 1). Users can then step through the inferred regions with the provided arrows or recenter the view on a particular pillar by clicking on it. These interface details are borrowed from the Yeast Gene Order Browser [YGOB; 18]. The track at the very bottom of Fig. 1 illustrates the location of the current window relative to the entire set of pillars: clicking on this track allows the user to make larger jumps through the pillars. Any visualization generated can be downloaded as an Adobe PDF file for presentation.

$\text{POInT}_{\text{browse}}$'s second function is to allow users to download predicted gene trees and/or coding sequences for any selected pillar in a browser frame by clicking on the icons at the bottom of each pillar (Fig. 1). These gene trees are created by combining the assumed species tree for that polyploidy event (available from the button on the upper left) with POInT's orthology inferences. For example, in the case of a fully duplicated column, the gene tree returned will consist of two mirrored copies of the

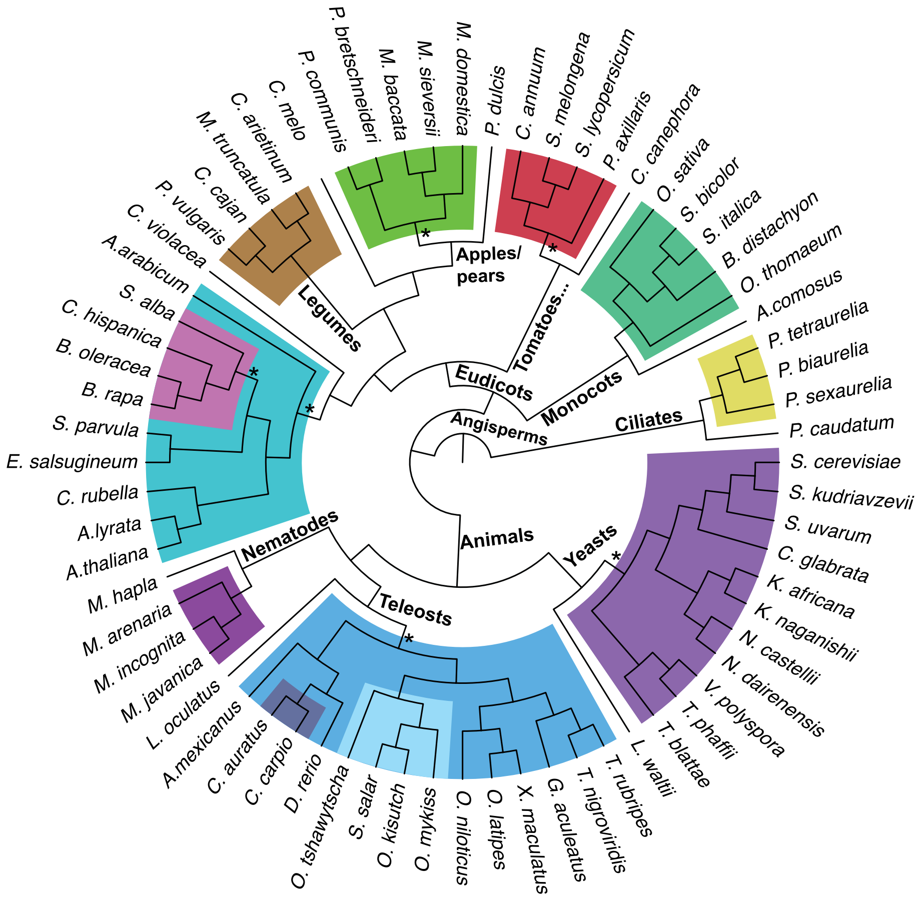


Fig. 3 Evolutionary relationships of the twelve polyploidy events in POInT$_{browse}$. The twelve polyploidy events currently displayed in POInT$_{browse}$ are shown as colored blocks with the assumed (labeled by "*") or computed (all others) species trees used by POInT depicted. We used a non-polyploid outgroup genome to infer the DCS blocks for each event, and these organisms are shown uncolored. References to our analyses of these events are given in the text and at the POInT$_{browse}$ website

species tree with the gene identifiers from the orthology predictions at the tips. In cases where duplicate losses have occurred, those tips are pruned from the species tree.

POInT$_{browse}$'s final capacity is a batch download feature, reached with the "Batch query" button (Fig. 1). This button opens a new window where the user selects a polyploidy event from which to download orthology inference sets. Pillars from that event can be selected based on orthology confidence combined with specifications for the number of duplicate genes required to be present (from fully duplicated to fully single-copy). Alternatively, the query can be restricted to single-copy orthologs. In each case, POInT returns a UNIX tar file containing CDS regions and gene trees meeting the selected criteria. Thus, when single-copy orthologs are requested, the download includes pillars where only a single gene survives from the polyploidy event in each genome and where POInT predicts all of these genes to be orthologs at the confidence level selected. In this case, the user can also request only orthologs from the less or more fractionated genome, again based on POInT's inferences.

Utility and discussion

POInT and POInT$_{browse}$ represent an advance on other polyploid-genome visualization tools [19, 20] for several reasons: in particular they allow hypothesis testing through differing models of duplicate loss [8, 16] and provide confidence estimates for their orthology inferences. Of course, as with any approach, there are limitations to the POInT framework. POInT assumes that duplicate losses are independent along a chromosome and follow an assumed species phylogeny, both of which may be violated in some cases [21]. Even if we accept POInT's modeling framework, datasets where the genomes considered are highly fragmented can result in generally low confidence in the orthology inferences, as is seen for the triploid nematodes [18].

Given these advantages and disadvantages, how can POInT$_{browse}$ help researchers? It is targeted to three groups: those studying processes associated with polyploidy, such as biased fractionation, those interested in phylogenomic questions, and users interested in molecular evolution more generally. The value of synteny-based orthology data is illustrated in each case by prior work using either data from POInT or from YGOB [19], which was the antecedent to POInT. As an example of the first case, namely the study of polyploidy, we used the synonymous divergence of conserved duplicates to assess the relative rate of duplicate loss immediately after polyploidy relative to the loss rate later in the history of those lineages. We found that many, but not all, polyploidy events were characterized by an especially rapid loss of duplicated genes immediately after the event [5]. Likewise, Marcet-Houben and Gabaldon [22] used data from YGOB, among other sources, to present phylogenetic evidence that the yeast genome duplication was an allopolyploidy. We have also used the inferred orthologs from POInT to test whether repetitive element distributions differed between the subgenomes of extant mesohexaploid vegetable crops [17].

In the case of phylogenetics, polyploidy causes at least two difficulties in tree inference. The mere presence of duplicated genes makes the problem of reducing gene trees to species trees complex [23]: the common solution to this problem is to use only single-copy genes in large-scale analyses [24]. However, even in this framework, the loss of duplicated genes, and in particular, the *reciprocal* loss of duplicated genes in different taxa (dark pink column in Fig. 1), can give rise to cases where single-copy genes in multiple genomes are not orthologous [3]. The rate of reciprocal gene loss varies considerably across polyploidy events but is a universal feature of post-polyploid evolution [5]. Since reciprocal gene loss has been shown to adversely affect the quality of phylogenies inferred for polyploid taxa [25], using synteny information to restrict analyses to true orthologs is a promising approach for phylogenetic analyses of paleopolyploid taxa [26]. POInT$_{browse}$ potentially provides a route around both of these problems, giving researchers access to any desired set of orthologous genes, single-copy or otherwise, from which to start the inference process.

The final utility of POInT$_{browse}$ is for more general questions regarding the molecular evolution of duplicated genes. Deluna et al., [27] have used YGOB data to explore how duplicated genes do or do not contribute to robustness to gene loss, while Gera et al., [28] used WGD-produced duplicated transcription factors (identified with YGOB) to explore the post-WGD divergence in their binding specificity. Understanding the paralogous structure of a genome using tools like YGOB has also been critical for detecting neofunctionalization: the appearance of novel functions through gene duplication [29]. Finally, we have used orthology data from POInT to study post-polyploidy gene conversion [30–32]. Because POInT provides high quality orthology inferences that are not dependent on gene trees inferred from the sequences involved, the orthology that is evident from the gene order can be contrasted with gene trees inferred from the sequences. In our case, we could show that paralogous ribosomal proteins showed evidence for very strong and recent gene conversion, such that those paralogs, created by the ancient genome duplication about one hundred million years ago [33], were more similar to each other than either was to its orthologous gene in a closely related yeast species, despite the much more recent split (a few million years) of those orthologs [31].

Conclusions

POInT$_{\text{browse}}$ is a freely available collection of orthology inferences for more than fifty polyploid genomes from across the eukaryotic tree of life. The syntenic regions, gene sequences and inferred gene trees can be useful for researchers studying polyploid genome evolution, systematics and molecular evolution more generally.

Abbreviations

WGD	Whole-genome duplication
DCS	Double-conserved synteny

Acknowledgements

We would like to thank K. Byrne and K. Wolfe for access to the YGOB system, C. Smith for assistance in deploying POInT$_{\text{browse}}$ and A. Coppage, M. Joshi, L. McRae, J. Naranjo, E. Oppold, I. Tenneti, J. Thorne, and Y. Yang for help with testing POInT$_{\text{browse}}$.

Author contributions

GCC developed the underlying POInT code. MS developed code for creating gene name aliases in POInT$_{\text{browse}}$. MS and GCC tested and debugged POInT$_{\text{browse}}$ and wrote the manuscript. Both authors read and approved the final manuscript.

Funding

This work was supported by U. S. National Science Foundation Grant NSF-DEB- 2241312.

Availability of data and materials

Project name: POInT$_{\text{browse}}$. ***Project home page:*** wgd.statgen.ncsu.edu. ***Operating system:*** Platform independent. ***Programming language:*** c++. ***Other requirements:*** Web browser. ***License:*** LGPL-3.0. ***Other restrictions:*** None. POInT$_{\text{browse}}$ is available at wgd.statgen.ncsu.edu; the full POInT software package (v1.61), including the browser code, is available at https://github.com/gconant0/POInT. All of the data distributed through POInT$_{\text{browse}}$ are also available for download directly from the POInT$_{\text{browse}}$ website.

Declarations

Ethics approval and consent to participate

No human subjects or animals were involved in the work described.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 1 February 2023 Accepted: 19 April 2023
Published online: 27 April 2023

References

1. Van de Peer Y, Mizrachi E, Marchal K. The evolutionary significance of polyploidy. Nat Rev Genet. 2017;18(7):411–24.
2. Otto SP. The evolutionary consequences of polyploidy. Cell. 2007;131(3):452–62.
3. Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature. 2006;440:341–5.

4. De Smet R, Adams KL, Vandepoele K, Van Montagu MC, Maere S, Van de Peer Y. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc Natl Acad Sci USA. 2013;110(8):2898–903.
5. Hao Y, Fleming J, Petterson J, Lyons E, Edger PP, Chris Pires J, Thorne JL, Conant GC. Convergent evolution of polyploid genomes from across the eukaryotic tree of life. G3 Genes|Genomes|Genetics. 2022. https://doi.org/10.1093/g3journal/jkac094.
6. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290:1151–5.
7. Freeling M. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu Rev Plant Biol. 2009;60:433–53.
8. Conant GC, Wolfe KH. Probabilistic cross-species inference of orthologous genomic regions created by whole-genome duplication in yeast. Genetics. 2008;179:1681–92.
9. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol. 1981;17:368–76.
10. Hao Y, Conant GC. POInT: a tool for modeling ancient polyploidies using multiple polyploid genomes. In: Pereira-Santana A, Gamboa-Tuz SD, Rodríguez-Zapata LC, editors. Plant comparative genomics. New York: Springer US; 2022. p. 81–91. https://doi.org/10.1007/978-1-0716-2429-6_6.
11. Marianne Emery M, Willis MS, Hao Y, Barry K, Oakgrove K, Peng Y, Schmutz J, Eric Lyons J, Pires C, Edger PP, Conant GC. Preferential retention of genes from one parental genome after polyploidy illustrates the nature and scope of the genomic conflicts induced by hybridization. PLoS Genet. 2018;14(3):e1007267. https://doi.org/10.1371/journal.pgen.1007267.
12. Hao Y, Mabry ME, Edger P, Freeling M, Zheng C, Jin L, VanBuren R, Colle M, An H, Abrahams RS, et al. The contributions of the allopolyploid parents of the mesopolyploid Brassiceae are evolutionarily distinct but functionally compatible. Genome Res. 2021;31:799–810.
13. Anderson E, Bai Z, Bischof C, Blackford S, Demmel J, Dongarra J, Du Croz J, Greenbaum A, Hammarling S, McKenney A, et al. LAPACK users' guide. 3rd ed. Philadelphia: Society for Industrial and Applied Mathematics; 1999.
14. Dagum L, Menon R. OpenMP: an industry standard API for shared-memory programming. Comput Sci Eng IEEE. 1998;5(1):46–55.
15. Conant GC. Comparative genomics as a time machine: How relative gene dosage and metabolic requirements shaped the time-dependent resolution of yeast polyploidy. Mol Biol Evol. 2014;31(12):3184–93.
16. Conant GC. The lasting after-effects of an ancient polyploidy on the genomes of teleosts. PLoS ONE. 2020;15(4):e0231356.
17. McRae L, Beric A, Conant GC. Hybridization order is not the driving factor behind biases in duplicate gene losses among the hexaploid Solanaceae. In: Proceedings of the royal society B, 2022.
18. Schoonmaker A, et al. A single shared triploidy in three species of parasitic nematodes. G3 Genes Genom Genet. 2020;10:225–33.
19. Byrne KP, Wolfe KH. The yeast gene order browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res. 2005;15(10):1456–61.
20. Lyons E, Freeling M. How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J. 2008;53(4):661–73.
21. Gordon JL, Armisen D, Proux-Wera E, OhEigeartaigh SS, Byrne KP, Wolfe KH. Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents. Proc Natl Acad Sci U S A. 2011;108(50):20024–9.
22. Marcet-Houben M, Gabaldon T. Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the Baker's yeast lineage. PLoS Biol. 2015;13(8):e1002220.
23. Chen K, Durand D, Farach-Colton M. NOTUNG: a program for dating gene duplications and optimizing gene family trees. J Comput Biol. 2000;7(3–4):429–47.
24. Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires JC, Leebens-Mack J, DePamphilis CW. Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels. BMC Evol Biol. 2010;10:61. https://doi.org/10.1186/1471-2148-1110-1161.
25. Xiong H, Wang D, Shao C, Yang X, Yang J, Ma T, Davis CC, Liu L, Xi Z. Species tree estimation and the impact of gene loss following whole-genome duplication. Syst Biol. 2022;71(6):1348–61.
26. Washburn JD, Schnable JC, Conant GC, Brutnell TP, Shao Y, Zhang Y, Ludwig M, Davidse G, Pires JC. Genome-guided phylo-transcriptomic methods and the nuclear phylogentic tree of the paniceae grasses. Sci Rep. 2017;7(1):13528.
27. Deluna A, Vetsigian K, Shoresh N, Hegreness M, Colon-Gonzalez M, Chao S, Kishony R. Exposing the fitness contribution of duplicated genes. Nat Genet. 2008;40:676.
28. Gera T, Jonas F, More R, Barkai N. Evolution of binding preferences among whole-genome duplicated transcription factors. Elife. 2022;11:e73225.
29. Penalosa-Ruiz G, Aranda C, Ongay-Larios L, Colon M, Quezada H, Gonzalez A. Paralogous ALT1 and ALT2 retention and diversification have generated catalytically active and inactive aminotransferases in Saccharomyces cerevisiae. 2012.
30. Casola C, Conant GC, Hahn MW. Very low rate of gene conversion in the yeast genome. Mol Biol Evol. 2012;29(12):3817–26.
31. Evangelisti AM, Conant GC. Nonrandom survival of gene conversions among yeast ribosomal proteins duplicated through genome doubling. Genome Biol Evol. 2010;2:826–34.
32. Scienski K, Fay JC, Conant GC. Patterns of Gene conversion in duplicated yeast histones suggest strong selection on a coadapted macromolecular complex. Genome Biol Evol. 2015;7(12):3249–58.
33. Wolfe KH, Shields DC. Molecular evidence for an ancient duplication of the entire yeast genome. Nature. 1997;387:708–13.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.