

# A Class of Allopolyploidy Showing High Duplicate Retention and Continued Homoeologous Exchanges

Abbey Coppage <sup>1</sup>, Esha Bhatnagar <sup>1</sup>, Mitali Joshi <sup>1</sup>, Mustafa Siddiqui<sup>1</sup>, Logan McRae<sup>1</sup>, Gavin C. Conant <sup>1,2,3,\*</sup>

<sup>1</sup>Department of Biological Sciences, North Carolina State University, Raleigh, NC, USA

<sup>2</sup>Genetics and Genomics Academy, North Carolina State University, Raleigh, NC, USA

<sup>3</sup>Bioinformatics Research Center, North Carolina State University, Raleigh, NC, USA

\*Corresponding author: E-mail: gconant@ncsu.edu.

Accepted: March 16, 2025

## Abstract

We describe four ancient polyploidy events where the descendant taxa retain many more duplicated gene copies than has been seen in other paleopolyploidies of similar ages. Using POInT (the Polyploidy Orthology Inference Tool), we modeled the evolution of these four events, showing that they do not represent recent independent polyploidies despite the rarity of shared gene losses. We find that these events have elevated rates of interlocus gene conversion and that these gene conversion events are spatially clustered in the genomes. Regions of gene conversion also show very low synonymous divergence between the corresponding paralogous genes. We suggest that these genomes have experienced a delay in the return to a diploid state after their polyploidies. Under this hypothesis, homoeologous exchanges between the duplicated regions created by the polyploidy persist to this day, explaining the high rates of duplicate retention. Genomes with these characteristics arguably represent a new class of paleopolyploid taxa because they possess evolutionary patterns distinct from the more common and well-known paradigm of the rapid loss of many of the duplicated pairs created by polyploidy.

**Key words:** polyploidy, ohnolog retention, evolutionary model.

## Significance

We describe four ancient polyploidies with unusually high levels of duplicate gene survival, as well as high levels of recent gene conversion. We argue that delayed rediploidization has retained and homogenized the duplicates in these genomes, in contrast to the more common pattern of rapid post-polyploidy duplicate gene loss.

## Introduction

Ancient polyploidy events have been found across the eukaryotic tree (Van de Peer et al. 2017; Hao et al. 2022) and are particularly abundant in the flowering plants and the ray-finned fishes. Although a newly formed tetraploid genome effectively contains four copies of every gene, the majority of the paleopolyploid genomes examined to date show very extensive *fractionation* (Thomas et al. 2006), meaning that most of the ohnologs (i.e. duplicate gene copies produced by polyploidy; Wolfe 2000) have been

subsequently lost, due either to genetic drift or natural selection (Lynch and Conery 2000; Blanc and Wolfe 2004; De Smet et al. 2013). Many of these losses are likely to have occurred quite quickly relative to the timescale of speciation events, given the large number of shared ohnolog losses typically observed in the different genomes descending from the same polyploidy (Scannell et al. 2007; Hao et al. 2022).

Polyploidy events are commonly divided into allopolyploidies and autopolyploidies. Allopolyploidy occurs through the merging of genomes from two different, albeit related,

© The Author(s) 2025. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

species, meaning that it is in effect a genome doubling coupled to a hybridization. Autopolyploidy, on the other hand, occurs when the two genome copies derive from the same species (Stebbins 1947). Allopolyploidies appear to be more common among the surviving ancient polyploidy events (Barker et al. 2016; Hao et al. 2022). One piece of evidence for this claim is the fact that it is common to find that the ohnolog losses after polyploidy are unbalanced, with one of the progenitor genomes experiencing more losses than the other. This pattern is known as *biased* fractionation (Thomas et al. 2006; Woodhouse et al. 2010; Emery et al. 2018). Its presence is highly suggestive of the event in question being an allopolyploidy (Garsmeur et al. 2014), since it is difficult to understand how two initially identical genomes could produce loss biases (but see Makino and McLysaght 2012). The most often-invoked mechanism for biased fractionation is differential expression silencing of the subgenome with an excess of transposable elements. Such expression biases are then predicted to have the knock-on effect of favoring ohnolog loss from the suppressed genome (Freeling et al. 2012; Alger and Edger 2020). However, other mechanisms have recently been proposed: that the inherent kinetics of gene expression may give rise to expression and loss biases (An et al. 2024) or that the maternal subgenome may dominate in expression and gene preservation (Xu et al. 2023).

A prerequisite for any ohnolog losses after polyploidy is probably the suppression of homologous exchanges between the duplicated regions of the polyploid genome. Upon formation, at least for autopolyploids and allopolyploids of closely-related species, the similarities between the resulting homoeologous chromosomes will result in genetic exchanges between them at meiosis (Furlong and Holland 2002; Mandáková and Lysak 2018). This phase of polyploid evolution can be complex, with events such as entire chromosomes from one allopolyploid progenitor being replaced with the homoeologous copy from the other (Gaeta et al. 2007). Once such meiotic exchanges have ceased, we can refer to all or part of the polyploid genome as having become *diploidized*, such that each chromosome or chromosomal region undergoes recombination only with its sister chromosome (Robertson et al. 2017; Redmond et al. 2023) and not with the other progenitor subgenome.

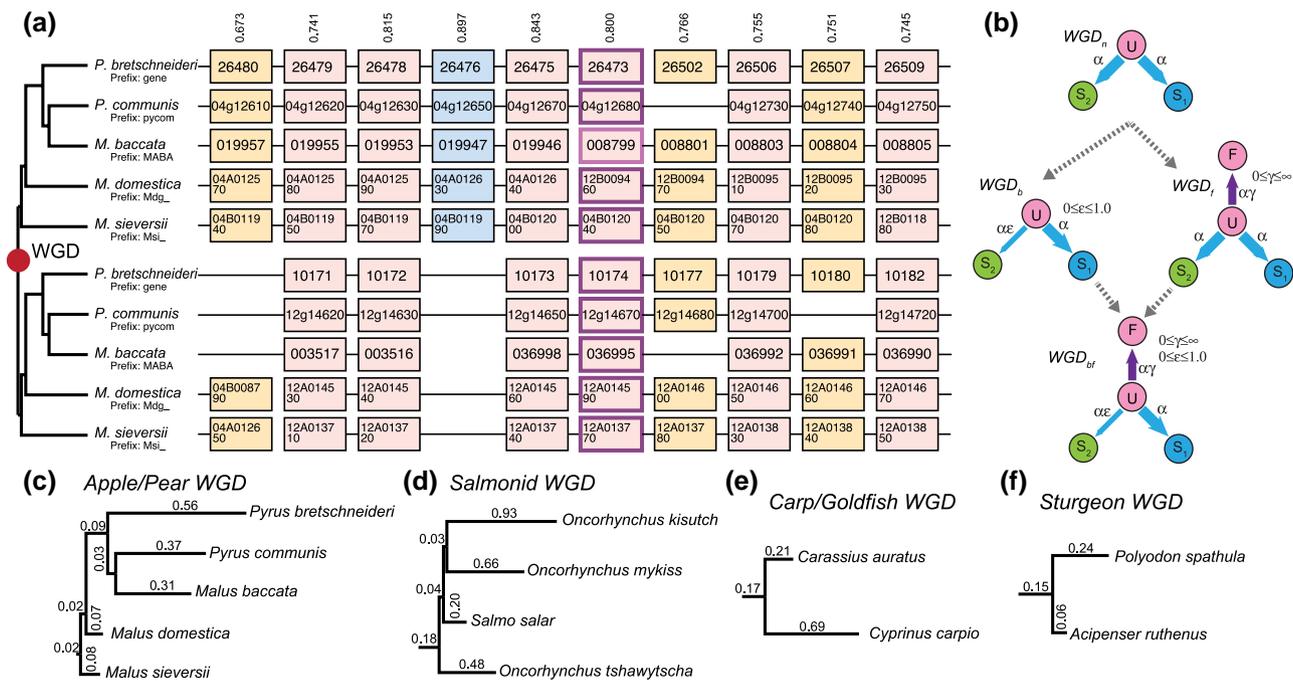
In the absence of homoeologous exchanges, gene trees inferred from the ohnolog sequences from species sharing a polyploidy are expected to resolve to mirrored copies of the species tree (Fig. 1a). This pattern is seen because both ohnologs have experienced the same pattern of post-polyploidy speciation events. However, if speciation events occur prior to diploidization, any post-speciation recombination events will result in ohnologs that do not yield mirrored species trees (Robertson et al. 2017). On the basis of such conflicting phylogenetic patterns, authors have argued that the whole genome duplication (WGD) events in the salmonids and sturgeons, as well as the ancient teleost

WGD, were followed by a prolonged period of homoeologous exchanges (Robertson et al. 2017; Parey et al. 2022; Redmond et al. 2023), meaning that diploidization was considerably delayed after these events. Indeed, in the case of the sturgeon WGD, prior work had argued for at least two independent genome duplications due to the conflicting gene trees (Crow et al. 2012; Cheng et al. 2021). However, a careful phylogenetic analysis resolved the corresponding gene trees as resulting from independent diploidization after a shared polyploidy in the lineages in question (Redmond et al. 2023).

The salmonid WGD is arguably the clearest case of “late diploidization” because extant species possessing it still show tetravalent pairing for some chromosomes (Phillips and Rab 2001; Braasch and Postlethwait 2012). Moreover, ohnologs found in these tetrasomically-inherited regions are less diverged from each other than are other ohnologs (Campbell et al. 2019). Since genetic exchanges between such homoeologous regions could also preserve ohnologs, such exchanges are also a plausible explanation for the excess of surviving ohnologs in the salmonids.

A clear open question is thus the frequency with which polyploidies with late diploidization are found in the tree of life. Using gene tree inferences to address this question is complicated by the fact that independent genome duplications can yield gene trees with topologies similar to those produced by late diploidization (Redmond et al. 2023). In this work, we sidestep the difficulties with using gene trees by using gene order, or *synteny*, data to detect late diploidization in polyploid genomes (Byrne and Wolfe 2005). We do so using our software package POInT (the Polyploidy Orthology Inference Tool; Conant and Wolfe 2008), which phylogenetically models the resolution of a polyploidy event using shared gene order and ohnolog losses (Fig. 1b). We have previously used POInT to analyze ten paleopolyploidy events, finding that neither the salmonid WGD nor the *Paramecium* octoploidy showed the extensive levels of ohnolog loss seen in the other eight “typical” polyploidy events (Hao et al. 2022). We also found elevated rate of interlocus gene conversion for both events. These conversion events “overwrite” (part of) the sequence of one ohnolog with the sequence from the other (Chen et al. 2007), possibly due to homoeologous recombination events. Because our analysis of the *Paramecium* genomes involved a phasing of a more recent and a more ancient tetraploidy (Hao et al. 2022), we were reluctant to draw overly firm conclusions from those observations at the time. Nonetheless, they are consistent with late diploidization for these two events and a more thorough exploration of its frequency is warranted.

Ancient genome duplications are known from the apples and pears, carp and goldfish, and the sturgeons (Van de Peer et al. 2017). Since our prior work, new genomes have become available from members of these three



**Fig. 1.** Modeling the evolution of four WGD events. a) An example region of ten ancestral genes that was duplicated in a WGD and is now found in two copies in five species of apples and pears. The upper and lower panels represent the two groups of orthologous regions (or *subgenomes*) formed by the WGD (abbreviated gene names are given in the boxes). This subgenome inference is more probable than the other  $2^5-1$  possible subgenome assignments that we could make by swapping the paired regions from one or more species. The model’s confidence  $c$  in the depicted subgenome assignments (as opposed to the other  $2^5-1$  possible ones) is given by the numbers above the column ( $0 \leq c \leq 1$ ). In the absence of homoeologous exchanges, fully retained ohnologs (columns without missing boxes; light pink) are expected to evolve under the mirrored species tree at left. A case of shared gene loss where all of the genes in the lower panel are missing is shown (blue), while other cases of individual losses have some duplicated and some single-copy genes (tan). b) The subgenome assignments in a) are made using the gene presence/absence data shown and evolutionary models of ohnolog loss. Four nested loss models that differ with respect to the presence of ohnolog pair fixation (rate  $\gamma$ ) and biased fractionation (rate  $\epsilon$ ) are illustrated. Ohnolog pairs all start in state  $U$  (Undifferentiated ohnolog) immediately post-WGD. Loss of the copy from subgenome 2 moves the pair to state  $S_1$  (only copy #1 retained) or symmetrically to  $S_2$ .  $WGD_n$  is a null model without biased losses or duplicate fixation. The instantaneous ohnolog loss rate to both subgenomes is then equal ( $\alpha$ ).  $WGD_f$  allows for fixation of ohnologs which occur at a relative rate  $\gamma$  ( $0 \leq \gamma < \infty$ ): Ohnologs in  $F$  cannot undergo loss.  $WGD_b$  introduces biased losses through the  $\epsilon$  ( $0 \leq \epsilon \leq 1$ ) parameter, making subgenome 1 increasingly favored as  $\epsilon$  decreases.  $WGD_{br}$  combines the  $WGD_b$  and  $WGD_f$  models. c to f) Inferred maximum likelihood topologies for the four events with the estimated branch lengths for each. Note that the branch lengths are not shown in the same scale across the panels.

groups, allowing us to analyze them with POInT. Comparisons of the shared WGD in the carp and goldfish genomes have shown that some of the ohnolog pairs in these genomes are more similar to each other than they are to their orthologs, suggesting recent gene conversion (Li et al. 2021). There is also phylogenetic evidence for homoeologous exchanges between subgenomes in these lineages (Li et al. 2021) (note that despite these patterns, phylogenetics still supports a shared ancient event in these lineages; Xu et al. 2023). Our preliminary analyses, meanwhile, found elevated ohnolog survival rates across all three of these events. We therefore analyzed them, as well as the salmonid WGD, to assess whether they displayed hallmarks of late diploidization. We found that high ohnolog retention characterizing these four events reduces the phylogenetic loss signal available to POInT, but that sufficient signal remains for the robust detection of shared WGD events. We further found that all four

events show spatial clustering in their retained ohnologs. They also show overly frequent and spatially clustered patterns of interlocus gene conversion, as well as very low synonymous divergence in these converted genes. We suggest that these four polyploidy events reflect a distinct class of paleopolyploid genome where diploidization is still incomplete even after tens or hundreds of millions of years of evolution.

## Results

### Modeling Post-Tetraploidy Ohnolog Losses Using Synteny Data and POInT

We inferred the regions of double-conserved synteny (DCS) created by ancient tetraploidy events in the salmonids, the apples and pears, carp and goldfish, and the sturgeons (Fig. 1a, Materials and Methods). For each event, we used POInT to infer the optimal phylogenetic topology from the DCS data

**Table 1** Models of post-WGD ohnolog loss

Event	WGD <sub>n</sub> lnL <sup>a</sup>	WGD <sub>b</sub> lnL <sup>b</sup>	<i>P</i> (WGD <sub>n</sub> → WGD <sub>b</sub> ) <sup>c</sup>	WGD <sub>f</sub> lnL <sup>d</sup>	<i>P</i> (WGD <sub>n</sub> → WGD <sub>f</sub> ) <sup>c</sup>	WGD <sub>bf</sub> lnL <sup>e</sup>	<i>P</i> (WGD <sub>b</sub> → WGD <sub>bf</sub> ) <sup>c</sup>	<i>P</i> (WGD <sub>f</sub> → WGD <sub>bf</sub> ) <sup>c</sup>
Salmon WGD	-42,326.95	-42,245.03	<10 <sup>-10</sup>	-41,578.1	<10 <sup>-10</sup>	-41,451.01 <sup>f</sup>	<10 <sup>-10</sup>	<10 <sup>-10</sup>
Apple/pear WGD	-37,496.10	-36,979.41	<10 <sup>-10</sup>	-37,482.16	1 × 10 <sup>-7</sup>	-36,929.74 <sup>f</sup>	<10 <sup>-10</sup>	<10 <sup>-10</sup>
Goldfish WGD	-17,609.54	-17,460.92 <sup>f</sup>	<10 <sup>-10</sup>	-17,609.52	<0.5	NA	NA	NA
Sturgeon WGD	-20,829.79 <sup>f</sup>	-20,667.29 <sup>f</sup>	<10 <sup>-10</sup>	-20,802.28	<10 <sup>-10</sup>	-20,666.71	0.28	<10 <sup>-10</sup>

<sup>a</sup>ln-Likelihood of the observed ohnolog loss data under a model without either biased fractionation or duplicate fixation (null model; Fig. 1b). <sup>b</sup>ln-Likelihood of the observed ohnolog loss data under the model allowing for biased fractionation but not duplicate fixation (Fig. 1b). <sup>c</sup>*P*-value for the test of no improvement in fit from the first to second model: i.e. twice the difference in ln-likelihood compared to a chi-square distribution, 1 df in all cases. <sup>d</sup>ln-Likelihood of the observed ohnolog loss data under the model allowing for ohnolog fixation but not biased fractionation (Fig. 1b). <sup>e</sup>ln-Likelihood of the observed ohnolog loss data under the model allowing for both biased fractionation and ohnolog fixation (Fig. 1b). <sup>f</sup>Optimal model.

using exhaustive tree search (Fig. 1c to f, Materials and Methods). Our datasets had 8,849, 10,513, 13,304, and 14,489 *pillars*, for the carp, apple/pear, sturgeon, and salmonid WGD events, respectively. Each pillar corresponds to an ancestral gene duplicated at the WGD and retained in at least one copy in every genome analyzed (Fig. 1a).

Using these optimal topologies, POInT allows the comparison of nested models of the losses of ohnolog copies along the phylogeny to test for both the presence of biased fractionation and of ohnolog fixation (Fig. 1b). All four events show bias toward one subgenome in the ohnolog losses (biased fractionation), consistent with these events having been allopolyploidies. For the apple/pear and salmonid events, there is also significant evidence for duplicate fixation (Table 1). However, once biased fractionation is accounted for, there is no evidence for fixation in the carp and sturgeon events, likely because only two genomes were studied in each case, too small a sample for the robust detection of fixation events.

### An Excess of Retained Ohnologs After Four WGD Events

In our previous work (Hao et al. 2022), the majority of the WGD events we studied retained relatively few of the ohnolog pairs created by the WGD, with percentages of ohnologs retained across all genomes ranging from about 12% (the yeast WGD) to 27% (At-α). In contrast, the events analyzed here retained 48% (carp WGD), 58% (salmonid WGD), 65% (sturgeon WGD), or even 70% (apple/pear WGD) of their ohnologs. The paramecium WGD also showed a high retention rate (49%) in our prior work: We do not analyze it here because we used an octoploid deconvolution approach when studying it that would be difficult to compare to these events (Hao et al. 2022).

### Confirmation of Single Polyploidy Events in Each of the Four Clades

One possible explanation for the apparent excess of ohnologs would be that we have mistaken several recent and independent WGD events for one ancient one, such that there has been less time for losses to occur than we assumed. We can test this

hypothesis with POInT (Conant and Wolfe 2008). Using the goldfish and carp as an example, if these two species had experienced *independent* WGD events, there would be no true shared ohnolog losses between them, although by chance a few genes might have experienced parallel losses. In the context of POInT's models, such independent events could be modeled on a phylogenetic tree where the root branch is of zero length, corresponding to no shared losses.

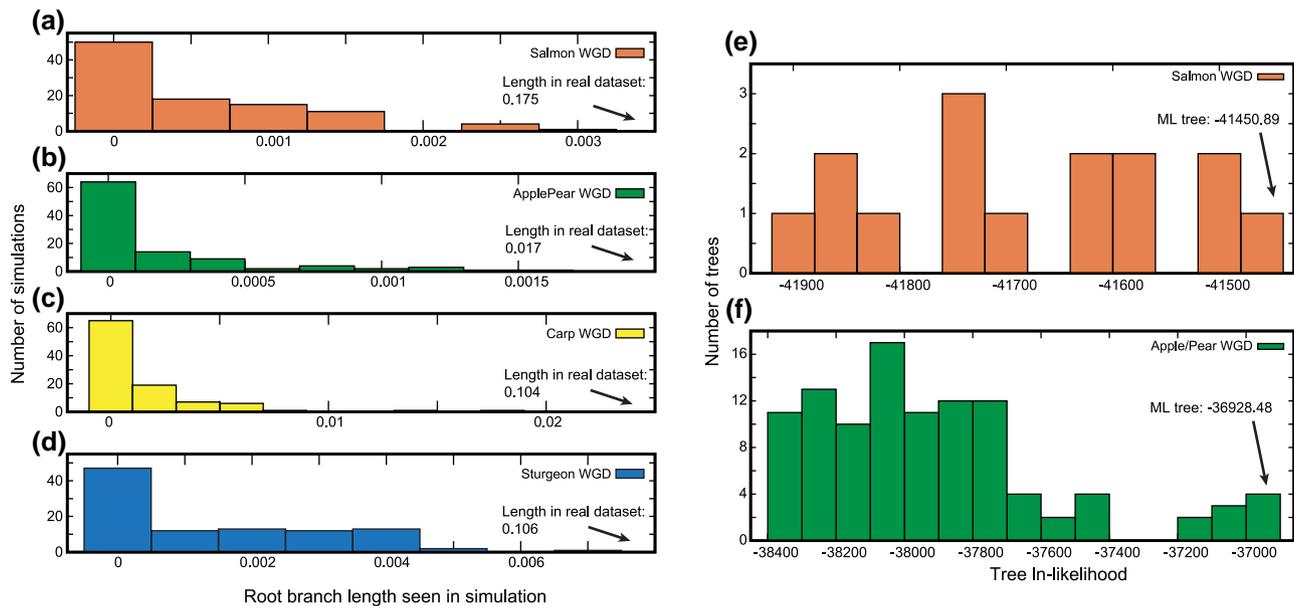
We can therefore fit a model with a zero-length root branch to the DCS pillars and then use the resulting model parameters to *simulate* the evolution of a number of genomes under the assumption of independent polyploidies (Materials and Methods). If we fit to the resulting simulated polyploid genomes a model that allows the root branch to have a non-zero length, we can estimate a distribution of the apparent length of the root branch seen for cases where the true length was known to be zero (the apparent length will generally be nonzero because of parallel losses). If we compare that distribution to the length of the root branch for the real DCS data, we can compute a *P*-value for the test of the null hypothesis of a zero-length root branch having generated the observed root branch lengths in the real datasets.

For all four events, the apparent lengths of the root branch for the simulated datasets are much smaller than that seen in the actual data (*P* < 0.01 in all cases; Fig. 2a to d), confirming that these are indeed shared ancient WGD events.

One could instead argue that, despite the speciation events observed, these polyploidies with high retention rates are so young that there was little time for ohnolog losses. However, their age is not noticeably younger than are other polyploidies with many more losses (Materials and Methods; supplementary figure, Supplementary Material online): In particular, the sturgeon event is the second oldest we have analyzed to date.

### The DCS Regions Contain Phylogenetic Signal Despite Yielding Unexpected Topologies

The topologies shown in Fig. 1c and d are not particularly similar to the expected relationships for the species in question



**Fig. 2.** Four shared polyploidy events with phylogenetic signals of ohnolog loss. Panels a) to d) show 100 simulations of independent polyploidies fit to trees with the assumption of shared losses. In other words, the root branch is simulated as having zero length, corresponding to two independent polyploidies, but those simulated data are then fit with a model that allows the estimated length to be nonzero. Plotted on x is the estimated length of the root branch of the phylogeny for these simulations with a model allowing a nonzero root, showing the level of apparent shared ancestry expected for independent polyploidies. The arrows indicate the length of the root branch for the real data. Note that the  $WGD_n$  model was used for the sturgeon and goldfish/carp WGD events because the  $\epsilon$  parameter in the  $WGD_b$  model can inflate the estimates of shared losses when only two genomes are analyzed (data not shown). Panels e) and f) illustrate the distribution of In-likelihoods for all possible phylogenetic topologies for the salmonid and Apple/Pear WGD events, with the maximum likelihood tree indicated.

(i.e. the members of the various genera are not monophyletic). One might therefore suspect that POInT's evolutionary models cannot recover any signal from these data and are hence inappropriate. However, we do not believe this to be the case. In Fig. 2e and f, we show the distribution of In-likelihoods for all of the possible phylogenetic trees for the salmonids and apples and pears, respectively. In the absence of signal, we would expect all of the topologies to show roughly similar likelihoods: Instead, in both cases, there is a large group of topologies that show poor fit to the DCS data and a smaller number of topologies with higher likelihoods. Hence, while these datasets do not show ideal phylogenetic performance, there is still signal to be found.

### The Apple/Pear and Salmonid WGD Events Show Elevated and Spatially-Structured Patterns of Ohnolog Fixation

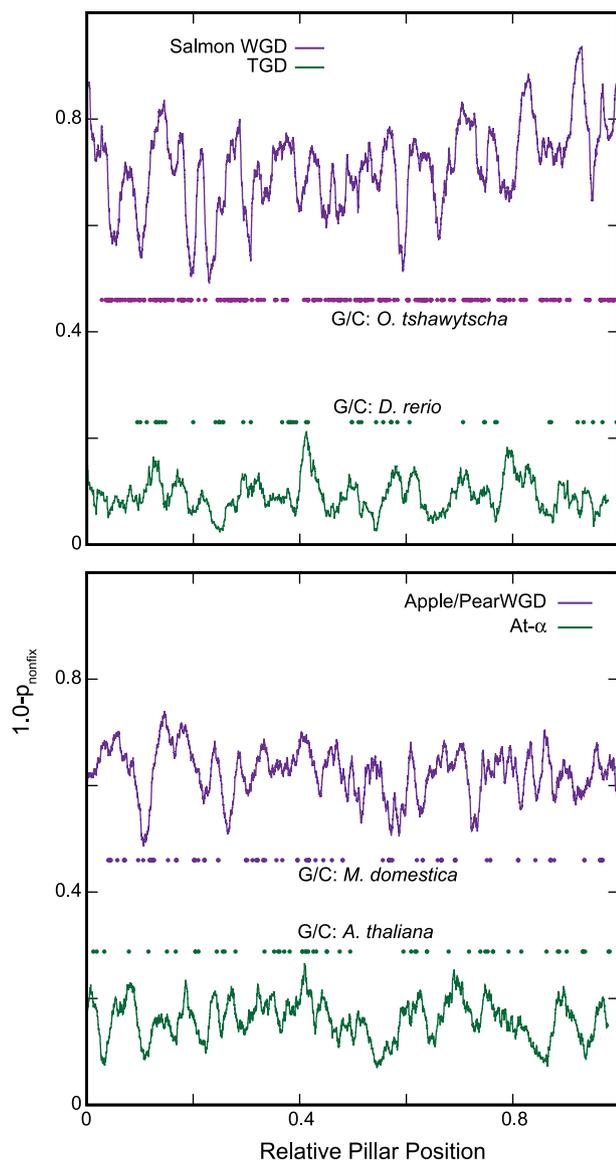
The  $WGD_{bf}$  model allows an ohnolog pair to become fixed along a branch of the phylogeny, such that that pair never undergoes loss later in the tree. For each pillar in the apple/pear and salmonid WGD events, we computed  $p_{nonfix}$ : the probability that no such fixation had occurred along any branch for that pillar (Materials and Methods). If we compute  $1.0 - p_{nonfix}$ , we thus have the probability of at least one fixation event at that pillar. As shown in Fig. 3, both

events are characterized by more duplicate fixation than other WGD events. A sliding window approach also suggests that these fixation rates are likely nonuniform across the genome, with hot and cold spots of fixation.

### These Polyploidy Events Show Elevated Rates of Recent Gene Conversion in the Coding Regions of Their Ohnologs

Using our previously described approach (Evangelisti and Conant 2010), we tested each surviving ohnologous gene pair with a clear ortholog for evidence of recent gene conversion. This test asks whether the two ohnolog sequences are more similar to each other than one is to its ortholog (Materials and Methods). Since the ohnologs had a common ancestor at the ancient WGD event while the orthologs diverged at a more recent speciation event, such a pattern is evidence of gene conversion since that speciation.

These genomes generally show a higher relative frequency of gene conversion than is observed in comparisons for the teleost genome duplication (TGD) or the recent WGD in *Arabidopsis thaliana* and its relatives. Moreover, unlike those events, the gene conversions in these genomes show spatial clustering (supplementary table S1, Supplementary Material online).



**Fig. 3.** High and spatially-structured patterns of ohnolog fixation in the salmonid and apple/pear genomes. At each pillar, we computed the probability that no ohnolog pairs had been fixed (Materials and Methods). On the y axis is 1.0 minus that probability: in other words, the probability of at least one fixation at that pillar. The values shown are averaged across a sliding window of 2% of the size of the overall dataset (in terms of the number of pillars). Also shown are the position of all the significant gene conversion events for the genome with the most such events (dots colored by event; Materials and Methods and supplementary table S1, Supplementary Material online). On the x axis is the scaled pillar position (0 to 1) so that events with differing numbers of pillars can be compared. a) The salmonid WGD compared to the teleost genome duplication (TGD). b) The apple/pear WGD compared to the At- $\alpha$  WGD.

To get a sense of whether these gene conversion events are recent, we compared the synonymous divergence of the paralogs ( $K_s$ , number of synonymous substitutions per synonymous site) between pairs with evidence of gene

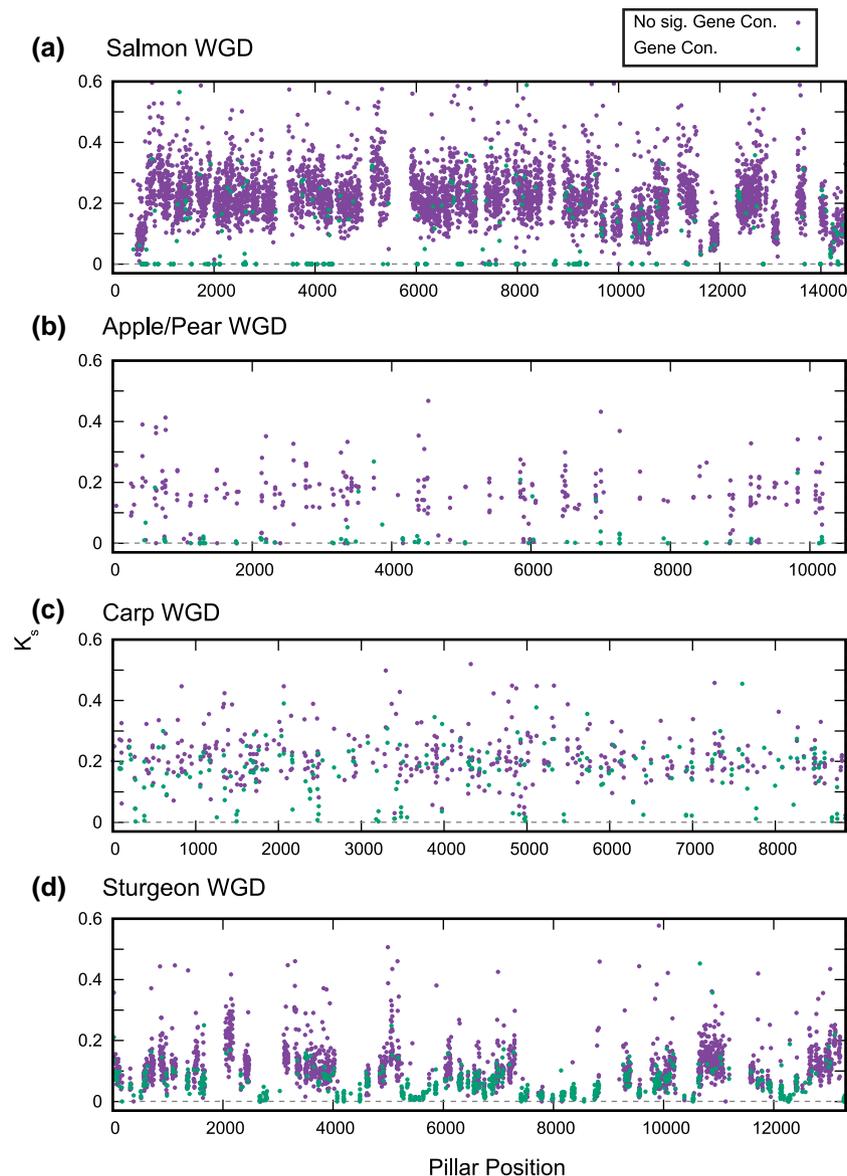
conversion to those without. As shown in Fig. 4, many of the paralog pairs with evidence for recent gene conversion show synonymous divergence approaching 0, consistent with an ongoing process of gene conversions in these taxa.

## Discussion

In most new-formed polyploids, recombinations and genetic exchanges between homoeologous chromosomes are extensive (Gaeta et al. 2007). Exceptions include the hexaploid common wheat, where the *Ph1* gene suppresses such recombinations, giving rise to “diploid-like” meiotic pairing (Riley and Chapman 1958; Zhang et al. 2020). Among the much older paleopolyploidies, homoeologous exchanges appear to have largely ceased, as recent gene conversion in these taxa is markedly less frequent and most of the duplicated genes have commonly been lost (Casola, Conant, and Hahn 2012; Hao et al. 2022; Yang et al. 2023). These facts suggest that such genomes long ago experienced full diploidization.

Here, we describe four paleopolyploidy events where the descendent genomes have three unusual features: (i) They have high rates of ohnolog retention (supplementary figure, Supplementary Material online), (ii) those retentions are spatially structured along chromosomes (Fig. 3), and (iii) they exhibit high rates of gene conversion (supplementary table S1, Supplementary Material online). We propose that a process of spatially heterogeneous and incomplete diploidization after these polyploidies can explain all three of these observations, justifying our claim of these events representing a new class of paleopolyploid genome. Figure 5 presents a cartoon view of these two models of post-polyploidy evolution.

In support of this hypothesis, we note that such continued tetrasomic pairing is already known for salmonids (Phillips and Rab 2001; Braasch and Postlethwait 2012), where some regions of the chromosomes undergo such pairing and others do not. When such pairings occur, they will tend to locally both preserve ohnolog pairs and homogenize their sequences, accounting for the high rates of ohnolog retention and gene conversion among the salmonids. We also note that in synthetic wheat tetraploids where homoeologous recombination does occur, it tends to be confined to gene bodies, in contrast to homologous pairing that is concentrated in other chromosomal regions (Zhang et al. 2020), an interesting parallel to the gene conversion events seen here. Furthermore, looking at Fig. 4a and d, we see suggestions of local regions of lower and higher  $K_s$ , which could be explained by a more recent or more ancient loss of tetrasomic pairing. This observation would be consistent with the spatial structuring seen in the gene conversions and ohnolog fixations. We therefore propose that these genomes preserve in their gene loss and conversion patterns a history of slow rediploidization. How slow? One of our more striking results is that, for all four events, we find regions marked by gene conversion that

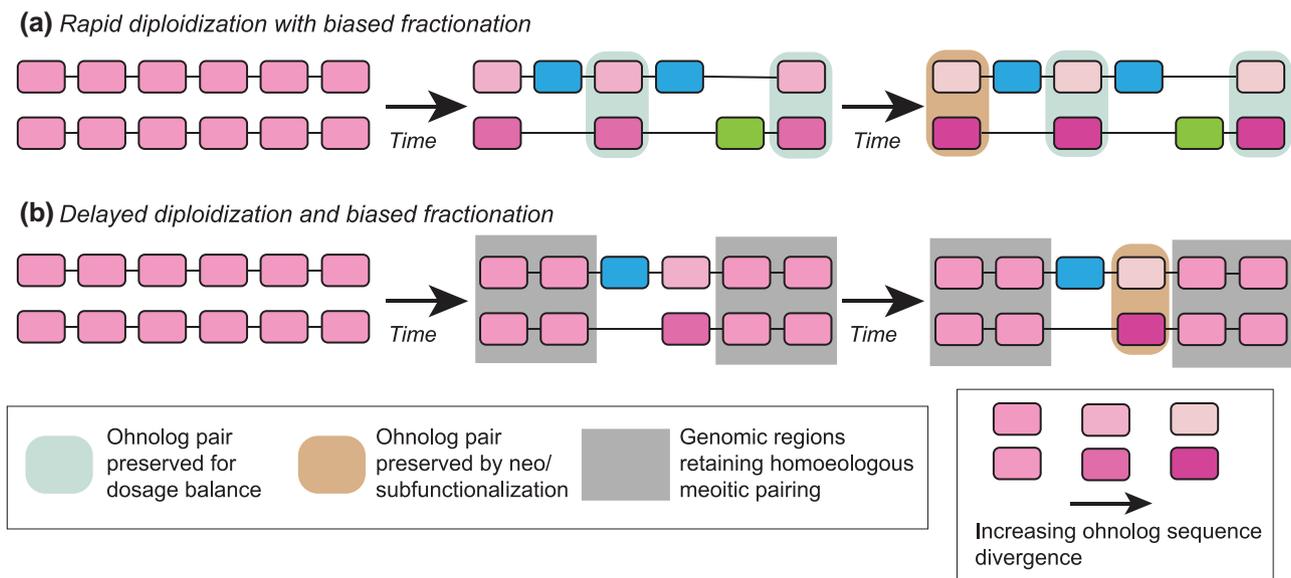


**Fig. 4.** Low synonymous divergence between paralogs with evidence of recent gene conversion. For each of the four WGD events, we selected the genome with the largest number of cases of gene conversion (supplementary table S1, Supplementary Material online). We then computed the pairwise synonymous divergence ( $K_s$ ) for all paralogs used in the gene conversion analysis and plotted these values for paralog pairs with ( $P < 0.05$ , supplementary table S1, Supplementary Material online; green) and without ( $P > 0.05$ , supplementary table S1, Supplementary Material online; purple) evidence of recent gene conversion.  $K_s$  often approaches 0 (dashed line) for pairs with evidence for conversion. The x axis represents the relative position of the ohnolog pair in our ancestral genome order reconstruction. Cases where  $K_s > 0.6$  are omitted for clarity. a to d) Plots for the four WGD events considered.

also include ohnolog pairs with little to no synonymous or nonsynonymous divergence between them (supplementary table S2, Supplementary Material online). We are only able to explain this pattern on the hypothesis that diploidization is in fact still ongoing in these lineages, with gene conversion events between the ohnologs continuing essentially until the present.

The evolution of “conventional” paleopolyploid genomes appears to be governed by the degenerative effects of genetic drift eliminating ohnologs on the one hand

(Li 1980; Lynch and Conery 2000) and preservation of those ohnologs to maintain dosage balance (Freeling 2009; Birchler and Veitia 2012) on the other (see cartoon in Fig. 5). This second force maintains ohnolog pairs because the loss of one copy can change the corresponding protein’s expression level relative to other gene products that it might physically interact with, regulate, be regulated by or otherwise functionally associate with (Veitia et al. 2013; Pires and Conant 2016). That such relative dosage imbalances can be detrimental is shown by the correlation



**Fig. 5.** Two models for post-polyploidy genome evolution. a) In a “typical” tetraploidy, many or most ohnolog pairs are reduced to single-copy by genetic drift fairly quickly after the event. Ohnologs may be protected from loss by a need for dosage balance (teal) or, over longer timescales, by functional divergence (tan). One of the two progenitor subgenomes may be favored in the losses (blue versus green, i.e. biased fractionation). b) For the polyploidies considered here, some regions of the genome (gray) maintain meiotic pairing between the progenitor subgenomes. As a result, ohnolog pairs in these regions are maintained and experience limited sequence divergence (observable in our data as gene conversions). Other regions of the genome may follow patterns more similar to standard tetraploidies. The difference in shading between ohnolog pairs illustrates their sequence divergence (or lack thereof).

between a human copy-number variant’s propensity to be pathogenic and the tendency of the gene containing it to have maintained its copy number status over evolutionary time (Rice and McLysaght 2017). The new class of paleopolyploidy just described is presumably still subject to the constraints of dosage balance. It appears, however, that the propensity of drift to remove duplicated material is now opposed by recombination with homoeologous chromosomes (Fig. 5).

From a methodological perspective, these patterns will have the unfortunate consequence of confounding phylogenetic inferences made using either the ohnolog losses (as for POInT) or the sequences themselves. In the case of POInT, the low levels of ohnolog loss yield little information for topology inference, and indeed, the topologies in Fig. 1 are not particularly sensible. At least among the surviving ohnologs, traditional phylogenetic methods will also tend to be misled due to the overwriting of the phylogenetic signal by gene conversion. We do note, however, that there are now sequence evolution models available that include gene conversions (Ji et al. 2016; Yang et al. 2023).

It is also worth speculating about the larger consequences of this new type of polyploidy. It has long been recognized that extending the “half-life” of a duplicated gene pair gives a relatively greater chance for functional innovations to occur in those duplicates (Rastogi and Liberles 2005; He and Zhang 2006). Of course, gene conversions can limit functional divergence since the duplicates maintain very similar sequences.

However, this limitation is not absolute, as, in yeast, there are duplicates with amino acid sequences that are kept identical by gene conversion that are nonetheless functionally distinct due to differing expression patterns (Ni and Snyder 2001; Komili et al. 2007). Whether these patterns of retention and conversion will persist indefinitely is also unclear, although the salmonid genome duplication is believed to be many tens of millions of years old (Braasch and Postlethwait 2012), while the sturgeon event is even older than this (Redmond et al. 2023). Polyploidy events continue to surprise us with their complexities, and there is no reason to believe that more surprises do not await us as new genomes allow us to consider new events.

## Materials and Methods

In addition to the salmonid WGD already described (Hao et al. 2022), we analyzed shared WGD events in the apples and pears (Van de Peer et al. 2017), sturgeons (Redmond et al. 2023), and the event shared between common carp and goldfish (Chen et al. 2019b) using POInT (Conant and Wolfe 2008; Emery et al. 2018).

## Genomes

We included five genomes in our analysis of apples and pears: those of *Malus domestica* (domestic apple; Daccord et al. 2017), *Malus baccata* (Chinese wild apple; Chen et al. 2019a), *Malus sieversii* (central Asian wild apple; Sun et al.

2020), *Pyrus bretschneideri* (Chinese white pear; Xue et al. 2018), and *Pyrus communis* (European pear; Chagné et al. 2014). The almond genome (*Pyrus dulcis*) was used as the nonpolyploid outgroup reference (Alioto et al. 2020). These six genomes were downloaded from the Genome Database for Rosaceae (Jung et al. 2019). For the carp/goldfish event, we used the genomes of *Cyprinus carpio* (common carp; Xu et al. 2014) and *Carassius auratus* (goldfish; Chen et al. 2019b). The genome of *Cyprinodon variegatus* (sheepshead minnow; Lencer et al. 2017) was used as a non-tetraploid genome reference. All three genome sequences were obtained from Ensembl release 107 (Cunningham et al. 2022). For the sturgeon/paddlefish event, we compared the genomes of *Acipenser ruthenus* (sterlet sturgeon; Du et al. 2020) and *Polyodon spathula* (paddlefish; Cheng et al. 2021) to the genome of the outgroup *Lepisosteus oculatus* (spotted gar; Braasch et al. 2016). The sturgeon and paddlefish genomes were both obtained from NCBI's GenBank (Sayers et al. 2022); the gar genome was taken from Ensembl release 98.

#### Inference of Blocks of Shared DCS Produced by Paleotetraploidies

For all three polyploidy events above, we used a three-step method to identify the blocks of DCS that survive from the ancient tetraploidies. The first step is a homology search of each of the nine tetraploid genomes against the unduplicated outgroup (i.e. *P. dulcis*, *C. variegatus*, or *L. oculatus*). This search uses GenomeHistory (v. 2.0; Conant and Wagner 2002) and BLASTP (v. 2.7.1; Altschul et al. 1997) to identify homologous pairs of genes from the outgroup and tetraploid genomes: these pairs are then filtered, aligned, and their synonymous and nonsynonymous divergence computed (Li 1997). For the comparison of apple and pear genomes to almond, we required a BLAST  $E \leq 10^{-10}$  and 70% or greater amino acid identity to retain a homolog pair. The corresponding figures for the carp/goldfish and the sturgeon events were  $E \leq 10^{-8}$  and 60% or greater identity.

The second step of the analysis was a search for DCS blocks between each tetraploid genome and its respective outgroup (Hao and Conant 2022). This search uses simulated annealing to maximize the size of the inferred DCS blocks. Successively longer runs of these searches were made until this DCS score converged.

The third and final step was to merge the five sets of DCS blocks inferred from the apple and pear genomes and the two sets of DCS blocks from the carp/goldfish and sturgeon/paddlefish into a single set of blocks each. We then optimized the order of the homoeologous pillars within these three merged sets of DCS blocks, again using simulated annealing. These ordered pillars could then be analyzed with POInT (Hao and Conant 2022).

#### Modeling of Ohnologous Gene Losses With POInT

For all four WGDs, we modeled the loss of ohnologous genes using four related models of ohnolog loss (Fig. 1b). For each model, we use POInT (v. 1.61) to search for the maximum likelihood phylogenetic topology using an exhaustive tree search. In other words, we sought the topology of highest likelihood for the five apple/pear species from among the 115 possible rooted topologies and similarly for the 15 possible topologies for the four salmonid species. The topologies of the goldfish and carp and of the sturgeons and paddlefish are trivial. We computed the significance of the improvement in model fit for more complex models using a likelihood ratio test (Sokal and Rohlf 1995). These inferences and DCS blocks can be visualized at the POInT<sub>browse</sub> portal ([wgd.statgen.ncsu.edu](http://wgd.statgen.ncsu.edu); Siddiqui and Conant 2023).

#### Dating the Four Polyploidy Events Considered

Dating polyploidy events can be contentious. As a rough approximation, we used the TimeTree package (Kumar et al. 2017) to compare the earliest estimated divergence date between any of the pairs of species in each of our four datasets. We compared these divergence values to those from eight other polyploidies we had previously analyzed (Hao et al. 2022; McRae et al. 2022). We then computed the minimum, maximum, and average proportion of homoeologs that retained both/all copies of the genes in question for all of these events (supplementary figure, Supplementary Material online). We note that the date for the salmonid WGD in particular is likely an underestimate, as the whitefishes and graylings are thought to share this event (Phillips and Rab 2001; Braasch and Postlethwait 2012): including them in the TimeTree comparisons pushes the date of this event back to roughly 55 MYA.

#### Test for a Single, Shared Polyploidy Event

To test whether the four WGD events considered might represent two independent WGD events in each lineage, we fit a model to each dataset that forced the length of the root branch of the optimal topology to have zero length. This restriction corresponds to the assumption of no shared gene losses between the lineages separated by the first speciation event in the tree. We then simulated 100 sets of genomes from this topology and model parameters. To each simulated dataset, we fit a model with a zero-length root branch (Model 1) and a model where the root branch was allowed to have an arbitrary length (Model 2). We then computed the difference in ln-likelihood between these two models (Conant and Wolfe 2008). We also compared the resulting distribution of the length of the root branch in the simulations analyzed under Model 2 to the observed

length of the root branch in the actual datasets (Fig. 2a). Using either the distribution of differences in log-likelihood between Model 2 and Model 1 or the distribution of estimated root branch length from the simulations resulted in the same rejection of the null hypothesis of no improvement in fit for the real data when allowing a nonzero length root (Fig. 2).

### Analysis of Fixation Frequency Across the Genome

To assess whether the patterns of ohnolog fixation were uniform throughout the genome, we adopted a sliding window measure of fixation frequency. At each pillar, we computed  $p_{\text{fix}}^i$ : the probability that the ohnolog gene pair at that pillar was fixed for taxa  $i$ , using the state conditional probabilities for that pillar (Emery et al. 2018). We then computed  $p_{\text{nonfix}}$ : the probability of no fixation events across that pillar as

$$p_{\text{nonfix}} = \prod_i (1 - p_{\text{fix}}^i). \quad (1)$$

In other words,  $p_{\text{nonfix}}$  gives the probability of no fixation events across any genomes at that pillar.

### Tests for Gene Conversion

We tested ohnologous gene pairs for evidence of recent gene conversion events with an approach similar to our previous work (Evangelisti and Conant 2010). For a given genome with a WGD, we first identified pairs of ohnologous genes  $G_1$  and  $G_2$  for which we could identify the syntenic ortholog  $O$  of  $G_1$  in a close polyploid relative with high confidence  $c$ . The value of  $c$  is taken from the orthology confidence estimates in POInT; we used a value of  $c \geq 0.9$  for the TGD, At- $\alpha$ , salmonid, and sturgeon events and  $c \geq 0.8$  for the carp and apple/pear WGD events. We used lower cutoffs for the latter two events because they demonstrated low orthology confidence values overall. Results from using a cutoff of  $c \geq 0.9$  for the apple/pear events were qualitatively similar, but three of the five comparisons did not show significant clustering because of the small number of ohnologs tested. As an aside, we note that the low orthology confidences are due to the small number of shared gene losses for these events: that small number of losses is in turn due to the high ohnolog retention rates discussed here.

We aligned the translated coding sequences of these three genes ( $G_1$ ,  $G_2$ , and  $O$ ) with T-Coffee v. 13.45 (Notredame et al. 2000) and deduced the corresponding nucleotide alignment. We then fit the Muse and Gaut/Goldman and Yang model of codon evolution (Goldman and Yang 1994; Muse and Gaut 1994) to the sequence triplet as previously described (Conant and Wagner 2003). The result of that analysis is a branch-specific value of  $K_a$  (number of nonsynonymous substitutions per nonsynonymous site)

for each of the three genes. The signal of gene conversion for such a triplet is when  $K_a$  for  $O$  ( $K_a^O$ ) is greater than  $K_a$  for both  $G_1$  and  $G_2$  ( $K_a^{G_1}$  and  $K_a^{G_2}$ ). Recall that  $G_1$  and  $O$  last shared a common ancestor at their (recent) speciation event, while  $G_1$  and  $G_2$  last shared a common ancestor at the (more ancient) WGD event. Hence, in the absence of gene conversion,  $K_a^O$  should be less than  $K_a^{G_2}$ . When  $K_a^O > K_a^{G_1}$ ,  $K_a^{G_2}$ , we fit an alternative model that forced  $K_a^O = K_a^{G_2}$ . We computed the significance of the inferred gene conversion by comparing the log-likelihoods of the original model and the constrained model where  $K_a^O = K_a^{G_2}$  using a likelihood ratio test with 1 df (Sokal and Rohlf 1995).

Among the pillars with significant evidence for a gene conversion event ( $P \leq 0.05$ ), we tested for spatially clustering as follows. For each such event, we computed the distance to the closest other pillar with a significant gene conversion event and took the average of this value across all significant events. We then conducted 1,000 randomizations where we laid the same number of events uniformly on the pillars and computed the same distance statistic. We then computed the proportion of the simulations with a smaller mean distance between gene conversion events than the observed distance for the real gene conversions (supplementary table S1, Supplementary Material online). For all ohnolog pairs in the gene conversion analysis, we also computed their pairwise synonymous divergence ( $K_s$ ) again using the codon model of Muse and Gaut/Goldman and Yang (Goldman and Yang 1994; Muse and Gaut 1994). The list of all significant cases of gene conversion is available from POInT<sub>browse</sub> ([https://wgd.statgen.ncsu.edu/Downloads/HighDupl\\_Poly\\_sig\\_GC\\_list.txt](https://wgd.statgen.ncsu.edu/Downloads/HighDupl_Poly_sig_GC_list.txt)).

### Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

### Acknowledgments

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

### Data Availability

DCS blocks, coding region sequences, optimal phylogenetic trees, and loss models are available from POInT<sub>browse</sub> (<https://wgd.statgen.ncsu.edu>). A list of all identified, significant cases of gene conversion for the 13 genomes considered here is also available from POInT<sub>browse</sub> ([https://wgd.statgen.ncsu.edu/Downloads/HighDupl\\_Poly\\_sig\\_GC\\_list.txt](https://wgd.statgen.ncsu.edu/Downloads/HighDupl_Poly_sig_GC_list.txt)). POInT (v. 1.61) is available from GitHub (<https://github.com/gconant0/POInT>). Software for the analysis of sequence

triplets is also available at GitHub ([https://github.com/gconant0/like\\_tri\\_test](https://github.com/gconant0/like_tri_test)).

## Literature Cited

- Alger EI, Edger PP. One subgenome to rule them all: underlying mechanisms of subgenome dominance. *Curr Opin Plant Biol*. 2020;54:108–113. <https://doi.org/10.1016/j.pbi.2020.03.004>.
- Alioto T, Alexiou KG, Bardil A, Barteri F, Castanera R, Cruz F, Dhingra A, Duval H, Fernández I Martí Á, Frias L, et al. Transposons played a major role in the diversification between the closely related almond and peach genomes: results from the almond genome sequence. *Plant J*. 2020;101(2):455–472. <https://doi.org/10.1111/tpj.14538>.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new-generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389–3402. <https://doi.org/10.1093/nar/25.17.3389>.
- An H, Pires JC, Conant GC. Gene expression bias between the subgenomes of allopolyploid hybrids is an emergent property of the kinetics of expression. *PLoS Comput Biol*. 2024;20(1):e1011803. <https://doi.org/10.1371/journal.pcbi.1011803>.
- Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. On the relative abundance of autopolyploids and allopolyploids. *New Phytol*. 2016;210(2):391–398. <https://doi.org/10.1111/nph.13698>.
- Birchler JA, Veitia RA. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc Natl Acad Sci U S A*. 2012;109(37):14746–14753. <https://doi.org/10.1073/pnas.1207726109>.
- Blanc G, Wolfe KH. Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *Plant Cell*. 2004;16(7):1679–1691. <https://doi.org/10.1105/tpc.021410>.
- Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J, et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet*. 2016;48(4):427–437. <https://doi.org/10.1038/ng.3526>.
- Braasch I, Postlethwait JH. Polyploidy in fish and the teleost genome duplication. In: *Polyploidy and genome evolution*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012, p. 341–383. [https://doi.org/10.1007/978-3-642-31442-1\\_17](https://doi.org/10.1007/978-3-642-31442-1_17).
- 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):1456–1461. <https://doi.org/10.1101/gr.3672305>.
- Campbell MA, Hale MC, McKinney GJ, Nichols KM, Pearse DE. Long-term conservation of ohnologs through partial tetrasomy following whole-genome duplication in Salmonidae. *G3 (Bethesda)*. 2019;9(6):2017–2028. <https://doi.org/10.1534/g3.119.400070>.
- Casola C, Conant GC, Hahn MW. Very low rate of gene conversion in the yeast genome. *Mol Biol Evol*. 2012;29(12):3817–3826. <https://doi.org/10.1093/molbev/mss192>.
- Chagné D, Crowhurst RN, Pindo M, Thrimawithana A, Deng C, Ireland H, Fiers M, Dzierzon H, Cestaro A, Fontana P, et al. The draft genome sequence of European pear (*Pyrus communis* L. 'Bartlett'). *PLoS One*. 2014;9(4):e92644. <https://doi.org/10.1371/journal.pone.0092644>.
- Chen JM, Cooper DN, Chuzhanova N, Férec C, Patrinos GP. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet*. 2007;8(10):762–775. <https://doi.org/10.1038/nrg2193>.
- Chen X, Li S, Zhang D, Han M, Jin X, Zhao C, Wang S, Xing L, Ma J, Ji J, et al. Sequencing of a wild apple (*Malus baccata*) genome unravels the differences between cultivated and wild apple species regarding disease resistance and cold tolerance. *G3 (Bethesda)*. 2019a;9(7):2051–2060. <https://doi.org/10.1534/g3.119.400245>.
- Chen Z, Omori Y, Koren S, Shirokiya T, Kuroda T, Miyamoto A, Wada H, Fujiyama A, Toyoda A, Zhang S, et al. De novo assembly of the goldfish (*Carassius auratus*) genome and the evolution of genes after whole-genome duplication. *Sci Adv*. 2019b;5(6):eaav0547. <https://doi.org/10.1126/sciadv.aav0547>.
- Cheng P, Huang Y, Lv Y, Du H, Ruan Z, Li C, Ye H, Zhang H, Wu J, Wang C, et al. The American paddlefish genome provides novel insights into chromosomal evolution and bone mineralization in early vertebrates. *Mol Biol Evol*. 2021;38(4):1595–1607. <https://doi.org/10.1093/molbev/msaa326>.
- Conant GC, Wagner A. GenomeHistory: a software tool and its application to fully sequenced genomes. *Nucleic Acids Res*. 2002;30(15):3378–3386. <https://doi.org/10.1093/nar/gkf449>.
- Conant GC, Wagner A. Asymmetric sequence divergence of duplicate genes. *Genome Res*. 2003;13(9):2052–2058. <https://doi.org/10.1101/gr.1252603>.
- Conant GC, Wolfe KH. Probabilistic cross-species inference of orthologous genomic regions created by whole-genome duplication in yeast. *Genetics*. 2008;179(3):1681–1692. <https://doi.org/10.1534/genetics.107.074450>.
- Crow KD, Smith CD, Cheng JF, Wagner GP, Amemiya CT. An independent genome duplication inferred from Hox paralogs in the American paddlefish—a representative basal ray-finned fish and important comparative reference. *Genome Biol Evol*. 2012;4(9):937–953. <https://doi.org/10.1093/gbe/evs067>.
- Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R, et al. Ensembl 2022. *Nucleic Acids Res*. 2022;50(D1):D988–D995. <https://doi.org/10.1093/nar/gkab1049>.
- Daccord N, Celton J-M, Linsmith G, Becker C, Choise N, Schijlen E, van de Geest H, Bianco L, Micheletti D, Velasco R, et al. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nat Genet*. 2017;49(7):1099–1106. <https://doi.org/10.1038/ng.3886>.
- 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 U S A*. 2013;110(8):2898–2903. <https://doi.org/10.1073/pnas.1300127110>.
- Du K, Stöck M, Kneitz S, Klopp C, Woltering JM, Adolphi MC, Feron R, Prokopov D, Makunin A, Kichigin I, et al. The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. *Nat Ecol Evol*. 2020;4(6):841–852. <https://doi.org/10.1038/s41559-020-1166-x>.
- Emery M, Willis MMS, Hao Y, Barry K, Oakgrove K, Peng Y, Schmutz J, Lyons E, Pires JC, Edger PP, et al. 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>.
- Evangelisti AM, Conant GC. Nonrandom survival of gene conversions among yeast ribosomal proteins duplicated through genome doubling. *Genome Biol Evol*. 2010;2:826–834. <https://doi.org/10.1093/gbe/evq067>.
- 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(1):433–453. <https://doi.org/10.1146/annurev.arplant.043008.092122>.
- Freeling M, Woodhouse MR, Subramaniam S, Turco G, Lisch D, Schnable JC. Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Curr Opin Plant Biol*. 2012;15(2):131–139. <https://doi.org/10.1016/j.pbi.2012.01.015>.

- Furlong RF, Holland PW. Were vertebrates octoploid? *Philos Trans R Soc Lond B Biol Sci.* 2002;357(1420):531–544. <https://doi.org/10.1098/rstb.2001.1035>.
- Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell.* 2007;19(11):3403–3417. <https://doi.org/10.1105/tpc.107.054346>.
- Garsmeur O, Schnable JC, Almeida A, Jourda C, D'Hont A, Freeling M. Two evolutionarily distinct classes of paleopolyploidy. *Mol Biol Evol.* 2014;31(2):448–454. <https://doi.org/10.1093/molbev/mst230>.
- Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol.* 1994;11(5):725–736. <https://doi.org/10.1093/oxfordjournals.molbev.a040153>.
- Hao Y, Conant GC. POInT: a tool for modeling ancient polyploidies using multiple polyploid genomes. *Methods Mol Biol.* 2022;2512:81–91. [https://doi.org/10.1007/978-1-0716-2429-6\\_6](https://doi.org/10.1007/978-1-0716-2429-6_6).
- Hao Y, Fleming J, Petterson J, Lyons E, Edger PP, Pires JC, Thorne JL, Conant GC. Convergent evolution of polyploid genomes from across the eukaryotic tree of life. *G3 (Bethesda).* 2022;12(6):jkac094. <https://doi.org/10.1093/g3journal/jkac094>.
- He X, Zhang J. Higher duplicability of less important genes in yeast genomes. *Mol Biol Evol.* 2006;23(1):144–151. <https://doi.org/10.1093/molbev/msj015>.
- Ji X, Griffing A, Thorne JL. A phylogenetic approach finds abundant interlocus gene conversion in yeast. *Mol Biol Evol.* 2016;33(9):2469–2476. <https://doi.org/10.1093/molbev/msw114>.
- Jung S, Lee T, Cheng CH, Buble K, Zheng P, Yu J, Humann J, Ficklin SP, Gasic K, Scott K, et al. 15 years of GDR: new data and functionality in the genome database for Rosaceae. *Nucleic Acids Res.* 2019;47(D1):D1137–D1145. <https://doi.org/10.1093/nar/gky1000>.
- Komili S, Farny NG, Roth FP, Silver PA. Functional specificity among ribosomal proteins regulates gene expression. *Cell.* 2007;131(3):557–571. <https://doi.org/10.1016/j.cell.2007.08.037>.
- Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol.* 2017;34(7):1812–1819. <https://doi.org/10.1093/molbev/msx116>.
- Lencer ES, Warren WC, Harrison R, McCune AR. The *Cyprinodon variegatus* genome reveals gene expression changes underlying differences in skull morphology among closely related species. *BMC Genomics.* 2017;18(1):424. <https://doi.org/10.1186/s12864-017-3810-7>.
- Li J-T, Wang Q, Huang Yang M-D, Li Q-S, Cui M-S, Dong Z-J, Wang H-W, Yu J-H, Zhao Y-J, Yang C-R, et al. Parallel subgenome structure and divergent expression evolution of allo-tetraploid common carp and goldfish. *Nat Genet.* 2021;53(10):1493–1503. <https://doi.org/10.1038/s41588-021-00933-9>.
- Li W-H. Rate of gene silencing at duplicate loci: a theoretical study and interpretation of data from tetraploid fish. *Genetics.* 1980;95(1):237–258. <https://doi.org/10.1093/genetics/95.1.237>.
- Li W-H. *Molecular evolution.* Sunderland (MA): Sinauer Associates; 1997.
- Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science.* 2000;290(5494):1151–1155. <https://doi.org/10.1126/science.290.5494.1151>.
- Makino T, McLysaght A. Positionally biased gene loss after whole genome duplication: evidence from human, yeast, and plant. *Genome Res.* 2012;22(12):2427–2435. <https://doi.org/10.1101/gr.131953.111>.
- Mandáková T, Lysak MA. Post-polyploid diploidization and diversification through dysploid changes. *Curr Opin Plant Biol.* 2018;42:55–65. <https://doi.org/10.1016/j.pbi.2018.03.001>.
- McRae L, Beric A, Conant GC. Hybridization order is not the driving factor behind biases in duplicate gene losses among the hexaploid Solanaceae. *Proc Biol Sci.* 2022;289(1985):20221810. <https://doi.org/10.1098/rspb.2022.1810>.
- Muse SV, Gaut BS. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol.* 1994;11(5):715–724. <https://doi.org/10.1093/oxfordjournals.molbev.a040152>.
- Ni L, Snyder M. A genomic study of the bipolar bud site selection pattern in *Saccharomyces cerevisiae*. *Mol Biol Cell.* 2001;12(7):2147–2170. <https://doi.org/10.1091/mbc.12.7.2147>.
- Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* 2000;302(1):205–217. <https://doi.org/10.1006/jmbi.2000.4042>.
- Parey E, Louis A, Montfort J, Guiguen Y, Crollius HR, Berthelot C. An atlas of fish genome evolution reveals delayed rediploidization following the teleost whole-genome duplication. *Genome Res.* 2022;32(9):1685–1697. <https://doi.org/10.1101/gr.276953.122>.
- Phillips R, Rab P. Chromosome evolution in the Salmonidae (Pisces): an update. *Biol Rev.* 2001;76(1):1–25. <https://doi.org/10.1111/j.1469-185X.2000.tb00057.x>.
- Pires JC, Conant GC. Robust yet fragile: expression noise, protein misfolding, and gene dosage in the evolution of genomes. *Annu Rev Genet.* 2016;50(1):113–131. <https://doi.org/10.1146/annurev-genet-120215-035400>.
- Rastogi S, Liberles DA. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol Biol.* 2005;5:28. <https://doi.org/10.1186/1471-2148-5-28>.
- Redmond AK, Casey D, Gundappa MK, Macqueen DJ, McLysaght A. Independent rediploidization masks shared whole genome duplication in the sturgeon-paddlefish ancestor. *Nat Commun.* 2023;14(1):2879. <https://doi.org/10.1038/s41467-023-38714-z>.
- Rice AM, McLysaght A. Dosage sensitivity is a major determinant of human copy number variant pathogenicity. *Nat Commun.* 2017;8:14366. <https://doi.org/10.1038/ncomms14366>.
- Riley R, Chapman V. Genetic control of the cytologically diploid behaviour of hexaploid wheat. *Nature.* 1958;182:713–715.
- Robertson FM, Gundappa MK, Grammes F, Hvidsten TR, Redmond AK, Lien S, Martin SAM, Holland PWH, Sandve SR, Macqueen DJ. Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biol.* 2017;18(1):111. <https://doi.org/10.1186/s13059-017-1241-z>.
- Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Schoch CL, Sherry ST, Karsch-Mizrachi I. GenBank. *Nucleic Acids Res.* 2022;50(D1):D161–D164. <https://doi.org/10.1093/nar/gkab1135>.
- Scannell DR, Frank AC, Conant GC, Byrne KP, Woolfit M, Wolfe KH. Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proc Natl Acad Sci U S A.* 2007;104(20):8397–8402. <https://doi.org/10.1073/pnas.0608218104>.
- Siddiqui M, Conant GC. POInTbrowse: orthology prediction and synteny exploration for paleopolyploid genomes. *BMC Bioinformatics.* 2023;24(1):174. <https://doi.org/10.1186/s12859-023-05298-w>.
- Sokal RR, Rohlf FJ. *Biometry.* 3rd ed. New York: W. H. Freeman and Company; 1995.
- Stebbins GL Jr. Types of polyploids; their classification and significance. *Adv Genet.* 1947;1:403–429. [https://doi.org/10.1016/S0065-2660\(08\)60490-3](https://doi.org/10.1016/S0065-2660(08)60490-3).
- Sun X, Jiao C, Schwaninger H, Chao CT, Ma Y, Duan N, Khan A, Ban S, Xu K, Cheng L, et al. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat Genet.* 2020;52(12):1423–1432. <https://doi.org/10.1038/s41588-020-00723-9>.
- Thomas BC, Pedersen B, Freeling M. Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes.

- Genome Res. 2006;16(7):934–946. <https://doi.org/10.1101/gr.4708406>.
- Van de Peer Y, Mizrahi E, Marchal K. The evolutionary significance of polyploidy. *Nat Rev Genet.* 2017;18(7):411–424. <https://doi.org/10.1038/nrg.2017.26>.
- Veitia RA, Bottani S, Birchler JA. Gene dosage effects: nonlinearities, genetic interactions, and dosage compensation. *Trends Genet.* 2013;29(7):385–393. <https://doi.org/10.1016/j.tig.2013.04.004>.
- Wolfe KH. Robustness: it's not where you think it is. *Nat Genet.* 2000;25(1):3–4. <https://doi.org/10.1038/75560>.
- Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. *PLoS Biol.* 2010;8(6):e1000409. <https://doi.org/10.1371/journal.pbio.1000409>.
- Xu MR, Liao ZY, Brock JR, Du K, Li GY, Chen ZQ, Wang YH, Gao ZN, Agarwal G, Wei KH, et al. Maternal dominance contributes to sub-genome differentiation in allopolyploid fishes. *Nat Commun.* 2023;14(1):8357. <https://doi.org/10.1038/s41467-023-43740-y>.
- Xu P, Zhang X, Wang X, Li J, Liu G, Kuang Y, Xu J, Zheng X, Ren L, Wang G, et al. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nat Genet.* 2014;46(11):1212–1219. <https://doi.org/10.1038/ng.3098>.
- Xue H, Wang S, Yao JL, Deng CH, Wang L, Su Y, Zhang H, Zhou H, Sun M, Li X, et al. Chromosome level high-density integrated genetic maps improve the *Pyrus bretschneideri* 'DangshanSuli' v1.0 genome. *BMC Genomics.* 2018;19(1):833. <https://doi.org/10.1186/s12864-018-5224-6>.
- Yang Y, Xu T, Conant G, Kishino H, Thorne JL, Ji X. Interlocus gene conversion, natural selection, and paralog homogenization. *Mol Biol Evol.* 2023;40(9):msad198. <https://doi.org/10.1093/molbev/msad198>.
- Zhang Z, Gou X, Xun H, Bian Y, Ma X, Li J, Li N, Gong L, Feldman M, Liu B, et al. Homoeologous exchanges occur through intragenic recombination generating novel transcripts and proteins in wheat and other polyploids. *Proc Natl Acad Sci U S A.* 2020;117(25):14561–14571. <https://doi.org/10.1073/pnas.2003505117>.

Associate editor: Christopher Wheat