The continuing impact of an ancient polyploidy on the genomes of teleosts

Gavin C. Conant[1-4,*]

[1]Department of Biological Sciences, [2]Bioinformatics Research Center, and [3]Program in Genetics, North Carolina State University, Raleigh, NC, U.S.A. and [4]Division of Animal Sciences, University of Missouri, Columbia, MO, U.S.A.

*Correspondence: G. Conant, gconant@ncsu.edu

Running Head: Lasting effects of the teleost genome duplication

Keywords: polyploidy, evolution of development, evolutionary model, sensory evolution, evolution of visual systems

Abbreviations: WGD: whole-genome duplication, TGD: teleost genome duplication, DCS: double conserved synteny, POInT: Polyploid Orthology Inference Tool

Abstract:

The ancestor of most teleost fishes underwent a whole-genome duplication event three hundred million years ago. Despite its antiquity, the effects of this event are evident both in the structure of teleost genomes and in how those genes still operate to drive form and function. I describe the inference of a set of shared syntenic regions that survive from the teleost genome duplication (TGD) using eight teleost genomes and the outgroup gar genome (which lacks the TGD). I phylogenetically modeled the resolution of the TGD via shared and independent gene losses, concluding that it was likely an allopolyploidy event due to the biased pattern of these gene losses. Duplicate genes surviving from this duplication in zebrafish are less likely to function in early embryo development than are genes that have returned to single copy. As a result, surviving ohnologs function later in development, and the pattern of which tissues these ohnologs are expressed in and their functions lend support to recent suggestions that the TGD was the source of a morphological innovation in the structure of the teleost retina. Surviving duplicates also appear less likely to be essential than singletons, despite the fact that their single-copy orthologs in mouse are no less essential than other genes. Nonetheless, the surviving duplicates occupy central positions in the zebrafish metabolic network.

Introduction

The study of doubled genomes (or polyploids) has a long history in genetics (Kuwada 1911; Clausen and Goodspeed 1925; Ohno 1970; Taylor and Raes 2004; Garsmeur, et al. 2013), but it was the advent of complete genome sequencing that most dramatically confirmed the role of polyploidy in shaping modern eukaryote genomes (Van de Peer, et al. 2017). The remnants of ancient genome duplications have been found across the eukaryotic phylogeny, from flowering plants (Soltis, et al. 2009) and yeasts (Wolfe and Shields 1997) to ciliates (Aury, et al. 2006) vertebrates (Ohno 1970; Kasahara 2007; Makino and McLysaght 2010) nematodes (Blanc-Mathieu, et al. 2017) and arachnids (Schwager, et al. 2017).

While flowering plants may be the "champions" of polyploidy (Soltis, et al. 2009), genome duplication has also extensively shaped the evolution of the teleost fishes (Chenuil, et al. 1999; Alves, et al. 2001; Braasch and Postlethwait 2012; Yang, et al. 2015). Polyploidies ranging in age from recent (<1Mya) hybridizations to very old events are known, including events shared among clades in the salmonids, carps, and sturgeons. The event considered here is a very old one that occurred between 320 and 400 Mya in the ancestor of most ray-fined fishes: the *teleost genome duplication* (TGD; Christoffels, et al. 2004; Hoegg, et al. 2004; Vandepoele, et al. 2004; Crow, et al. 2005; Sémon and Wolfe 2007). Evidence for this event started to accumulate in the late 1990s and became effectively irrefutable with the sequencing of the first teleost genomes (Aparicio, et al. 2002; Jaillon, et al. 2004; Van de Peer 2004).

Several evolutionary changes are associated with the TGD, including molecular divergence in vitamin receptors (Kollitz, et al. 2014), circulatory system genes (Moriyama, et al. 2016) and in the structure of core metabolism (Steinke, et al. 2006). Indeed, the classic example of duplicate gene divergence by subfunctionalization involves two zebrafish *ohnologs* (duplicates that are the products of a WGD; Wolfe 2000) *eng1a* and *eng1b* from the TGD (Force, et al. 1999). At the genome scale, the TGD probably increased the genome rearrangement rate for a period (Sémon and Wolfe 2007), as well as increasing the rate of sequence insertions and deletions in surviving ohnologs (Guo, et al. 2012). Likewise, extant teleost genomes show evidence for *reciprocal gene loss* between alternative copies of homologous genes created by the TGD, a pattern that can induce reproductive isolation between populations possessing it (Naruse, et al. 2004; Scannell, et al. 2006; Semon and Wolfe 2007).

A phylogenomic study of the TGD was undertaken by Inoue and colleagues (2015), who concluded that, as with other WGD events, the TGD was followed by an initial period of very rapid duplicate gene loss (Scannell, et al. 2007; McGrath, et al. 2014). However, their study was limited because it used gene tree/species tree reconciliation to identify the relics of the TGD, an approach which is known to have limitations relative to methods based on the identification and analysis of blocks of double-conserved synteny (DCS; Nakatani and McLysaght 2019; Zwaenepoel, et al. 2019). As a result, Inoue et al., could not completely phase the surviving genes in each genome relative to each other, making the process of estimating loss timings more challenging and suggesting that a new analysis of the TGD is appropriate. This opportunity is particularly important because zebrafish's role as a developmental model gives us a unique opportunity to explore the effects of WGD on developmental evolution. That its effects may be important has long been hypothesized, with one example being the suspected role of Hox gene duplication in creating plasticity in body-plan evolution (Hoegg, et al. 2007). Moreover, much of the work on the "rules" of evolution after genome duplication have been performed using relatively recent events in flowering plants and yeasts, with less understanding of the very-long term effects of polyploidy. These proposed rules include the tendency of more highly interacting genes to remain as ohnolog pairs longer after WGD, a pattern explained by the *dosage balance hypothesis* (DBH; Freeling and Thomas 2006; Birchler and Veitia 2007; Edger and Pires 2009; Freeling 2009; Makino and McLysaght 2010; Conant, et al. 2014). The DBH argues that cellular interactions, governed as they are by biochemical kinetics (Veitia and Potier 2015), can be sensitive to imbalances in the concentrations of the interacting entities, driving those interacting genes to be maintained in similar dosages (e.g., as ohnologs after WGD). Importantly, the effects of the DBH may not preserve ohnologs indefinitely (Conant, et al. 2014), and the TGD is old enough to explore this question. A second rule of polyploidy is that the two parent genomes that contribute to an allopolyploid are not generally equal: instead *biased fractionation* is observed, whereby one of the two parent genomes retains more genes than does the other (Thomas, et al. 2006; Sankoff, et al. 2010; Schnable, et al. 2011; Tang, et al. 2012; Bird, et al. 2018; Emery, et al. 2018; Wendel, et al. 2018). The role of biased fractionation in the resolution of the TGD has also not, to my knowledge, been explored.

Using POInT, the Polyploid Orthology Inference Tool (Conant and Wolfe 2008), I analyzed the resolution of the TGD, based on nearly 5600 duplicated loci that are detectable in

eight teleost genomes. I find that the surviving ohnologs produced by the TGD are distinct in their character even after more than 300 million years of evolution. They are rarely expressed in the earliest phases of development, are less likely to be essential in zebrafish, and they occupy more central positions in the zebrafish metabolic network. In addition, there are suggestions that the TGD helped shape a key innovation in the teleost visual system.

Results

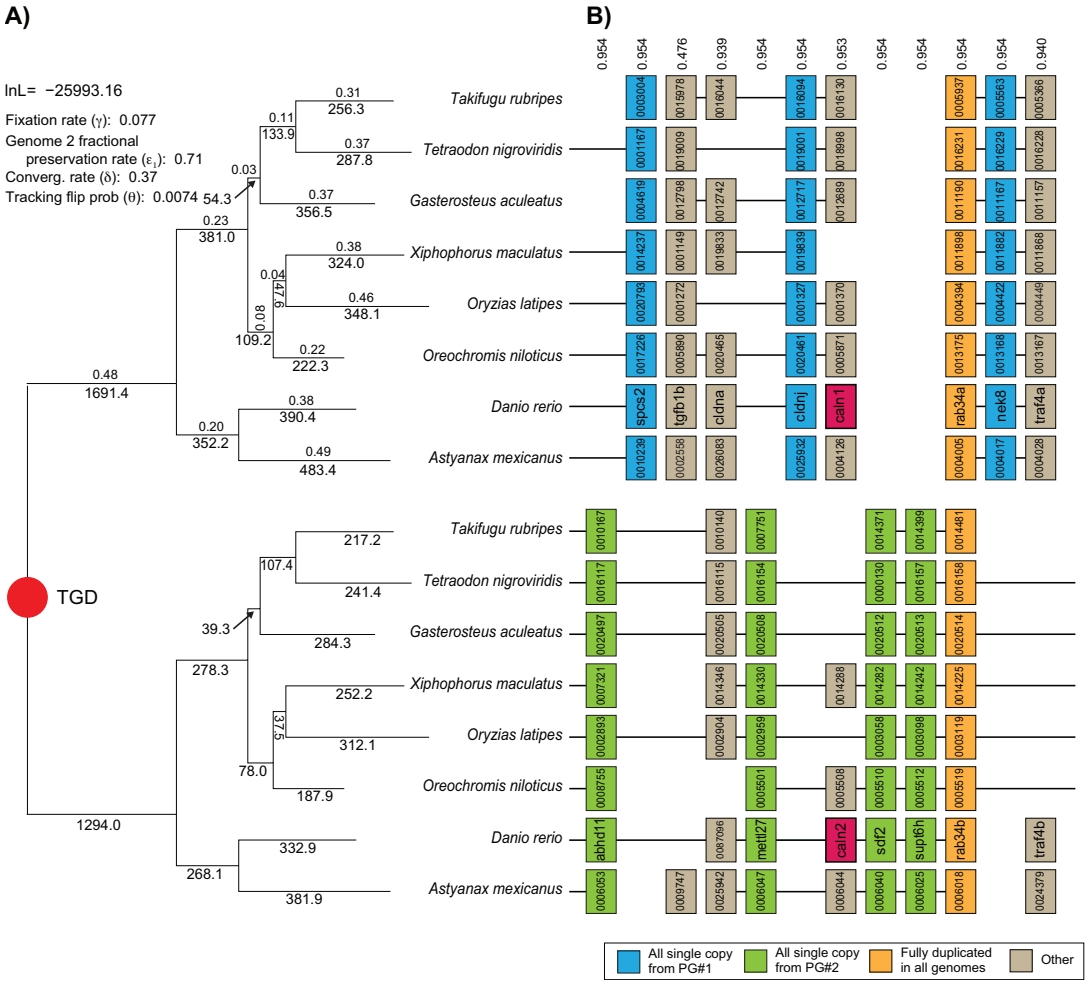


Figure 1: Resolution of the TGD through ohnolog losses. **A)** Shown is the assumed phylogeny of the eight species analyzed (see *Methods*). The TGD induces two mirrored gene trees, corresponding to the genes from the less fractionated parental genome (top) and the more fractionated genome (bottom). Below the branches in each tree are POInT's predicted number of gene losses along that branch for the parental genome in question. Above the branches in the upper tree are POInT's branch length estimates, namely the α parameter in Figure 2 and corresponding to the overall estimated level of gene loss on that branch (larger α implies a

greater number of losses relative to the total number of surviving ohnologs at the start of the branch). In the upper left are POInT's parameter estimates (γ,ε_1,δ) for the WGD-$bc^{nbn}f$ model (see Figure 2). **B)** An example region of the eight genomes, showing the blocks of DCS. For all species except zebrafish, truncated Ensembl gene identifiers are given; for zebrafish gene names are shown. The numbers above each column gives POInT's confidence in the orthology relationship shown, relative to the 2^8-*1*(=255) other possible orthology relationships. Genes are color-coded based on the pattern of ohnolog survival in the eight genomes. A pair of ohnologs expressed in the zebrafish retina are shown in magenta.

Identifying the relics of the TGD in eight teleost genomes.

Using a pipeline that places homologs of genes from gar (which lacks the TGD) into double-conserved synteny (DCS) blocks, I identified 5589 ancestral loci in eight teleost genomes surviving from the TGD (Figure 1). Each such locus was duplicated in the TGD and now has either one or both surviving genes from that event in each of the 8 genomes (c.f., Figure 1). I analyzed these blocks with POInT (Conant and Wolfe 2008), our tool for modeling the evolution of polyploid genomes. Because POInT infers orthologous chromosome segments based on a common gene order and shared gene losses along a phylogeny, it requires an estimate of the order of these loci in the ancestral genome immediately prior to the TGD (e.g., as was previously done for yeast; Gordon, et al. 2009). The TGD is considerably older and the genomes involved more rearranged than was the case for the yeast, *Arabidopsis* and grass polyploidies we previously analyzed (Conant and Wolfe 2008; Conant 2014; Emery, et al. 2018). Hence, I explored several means for estimating this order (*Methods*): while the number of synteny breaks in the optimal order estimated for the TGD was larger in proportion to the number of loci than was the case for our previous work, POInT's estimates of the model parameters are relatively insensitive to the exact order used (Supplemental Table 1). Similarly, the use of stringent homology criteria linked to requirements for synteny yield a set of DCS blocks that are very well supported across all eight genomes and represent a conservative set of loci with which to study the resolution of the TGD (*Methods*).

Ohnolog fixation, biased fractionation and convergent losses are all observed after the TGD.

POInT uses the copy-number status of each DCS locus in each genome, which is either duplicated (states U, F or C_1/C_2 in Figure 2) or single-copy (states S_1 and S_2), to model the

resolution of a WGD along a phylogeny in a manner analogous to models of DNA sequence evolution. This computation simultaneously considers phylogenetic topology, orthology relations and copy-number states, and, unlike all other model-based approaches to gene family evolution, uses synteny data, meaning that the orthology inferences for linked loci are mutually reinforcing. Note that POInT does not use the DNA sequences of the genes in question in its computations once the homology estimation step is complete. The model parameters are instead estimated via maximum likelihood from the observed set of single-copy and duplicated loci and their synteny relationships as previously described (Conant and Wolfe 2008). As part of this computation, POInT phases the synteny blocks into to groups of orthologs (top and bottom blocks in Figure 1) by summing over all 2^n possible orthology states at each DCS locus (where *n* is the number of genomes), resulting in inferences such as those shown in Figure 1. Note that this orthology inference procedure accounts for the reciprocal gene losses that can create single-copy paralogs in taxa sharing a WGD (Scannell, et al. 2006); it is entirely distinct from more common orthology inference approaches that do not account for shared polyploidies (Li, et al. 2003; Jensen, et al. 2008).

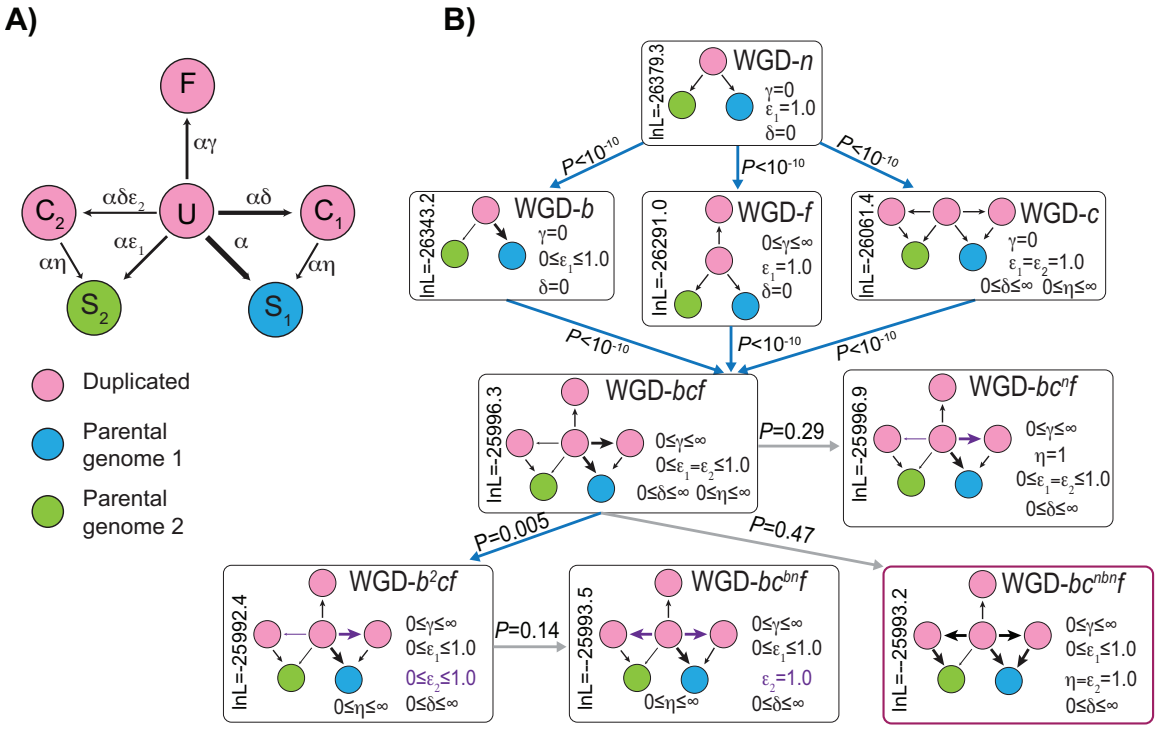


Figure 2: Testing nested models of post-WGD ohnolog evolution. **A)** Model states and parameter definitions for the set of models considered. **U**, $\mathbf{C_1}$, $\mathbf{C_2}$ and **F** are duplicated states,

while $\mathbf{S_1}$ and $\mathbf{S_2}$ are single-copy states (see *Methods*). $\mathbf{C_1}$ and $\mathbf{S_1}$ are states where the gene from the less-fractionated parental genome will be or are preserved, and $\mathbf{C_2}$ and $\mathbf{S_2}$ the corresponding states for the more-fractionated parental genome. The fractionation rate ε (the probability of the loss of a gene from the less fractionated genome relative to the more fractionted one) can either be the same for conversions to $\mathbf{C_1}$ and $\mathbf{C_2}$ as it is for $\mathbf{S_1}$ and $\mathbf{S_2}$ ($\varepsilon_1=\varepsilon_2$) or it can differ (see **B**). **B)** Testing nested models of WGD resolution. The most basic model (top) has neither biased fractionation nor duplicate fixation nor convergent losses. Adding any of these three processes improves the model fit (second row; $P<10^{-10}$). Adding the remaining two processes also improves the fit in all three cases (WGD-*bcf* model in the third row; $P<10^{-10}$). However, there is no significant evidence that the ε_2 parameter is significantly different from 1.0 (WGD-$bc^{bn}f$ verses WGD-b^2cf or WGD-$bc^{nbn}f$ verses WGD-*bcf*), implying no biased fractionation in the transitions to states $\mathbf{C_1}$ and $\mathbf{C_2}$. Likewise, there is no evidence that the η parameter is significantly different from 1.0 (WGD-$bc^{n}f$ verses WGD-*bcf* or WGD-$bc^{nbn}f$ verses WGD-*bcf*), meaning that losses from $\mathbf{C_1}$ and $\mathbf{C_2}$ occur at similar rates as do losses from **U**. Hence, the WGD-$bc^{nbn}f$ model is best supported by these data and is used for the remaining analyses.

I fit nested models of WGD evolution to the estimated ancestral order with the highest ln-likelihood under the WGD-$bc^{nbn}f$ model (Figure 2) in order to assess which of three processes observed after other WGD events were also detected after the TGD. The first process is duplicate fixation, meaning that some ohnolog pairs persist across the phylogeny longer than would be expected. The second process is biased fractionation, meaning that ohnolog losses favor one of the two parental genomes (defined as "Parental Genome 1" in Figure 2), and the third is the presence of convergent losses. These losses represent overly frequent parallel losses of the same member of the ohnolog pair on independent branches of the phylogeny. No matter what the order that these three phenomena are added to the duplicate loss model, all three are independently statistically significant ($P<10^{-10}$; Figure 2). From these analyses, I obtained lists of surviving ohnolog pairs from zebrafish (*Dr_Ohno_all* and *Dr_Ohno_POInT*, for all zebrafish ohnologs and zebrafish ohnologs also found syntenically in other genomes, respectively; see *Methods)*, the corresponding single-copy gene sets (*Dr_Sing_all* or *Dr_Sing_POInT*) and a set of early and late ohnolog losses (e.g., losses along the root and zebrafish tip branches of Figure 1A: *POInT_RootLosses* and *POInT_DrLosses,* respectively, see *Methods*).

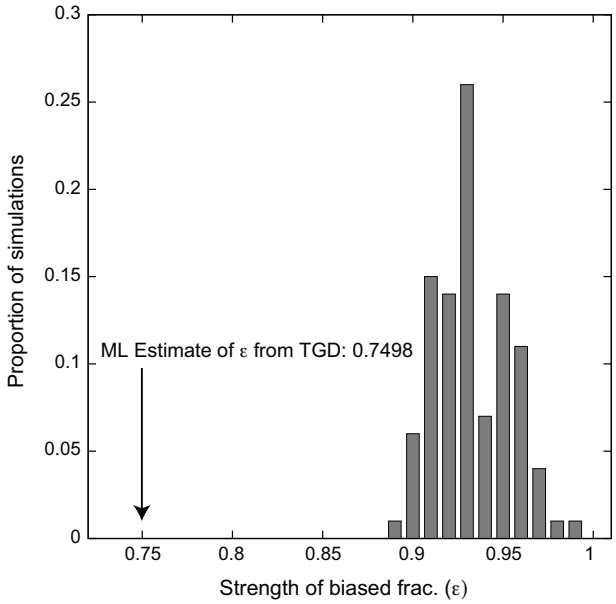


Figure 3: The estimated value of the biased fractionation parameter ε in the real teleost genomes (WGD-*bf* model, arrow, see *Methods*) is significantly different than those estimated from simulated genomes where biased fractionation was explicitly not included in the model (e.g., simulated ε=1.0, bars). Estimates of ε from these 100 simulations are always less than 1.0 because the model fits stochastic variations in the preservation patterns as potential biased fractionation. However, this stochastic variation never yields estimates of ε as small as seen in the real dataset (P<0.01).

The results in Figure 2 notwithstanding, because each synteny block will have some variation in loss patterns, it is possible that the estimates of the strength of biased fractionation were artifacts of stochastic variation in the loss patterns in those different blocks. In our previous work (Emery, et al. 2018), we visually assessed and rejected this possibility, but there was a degree of subjectivity to that approach. Here, I instead used POInT to simulate sets of 8 genomes under a model without biased fractionation (WGD-*f*). For each simulation, I then estimated the value of the ε parameter under a model with such a bias (WGD-*bf*, see *Methods*), allowing me to assess what degree of spurious bias might be induced by our approach. The level of biased fractionation seen after the TGD is inconsistent with purely stochastic variation (P<0.01, Figure 3), strongly supporting the conclusion that biased fractionation occurred after the TGD.

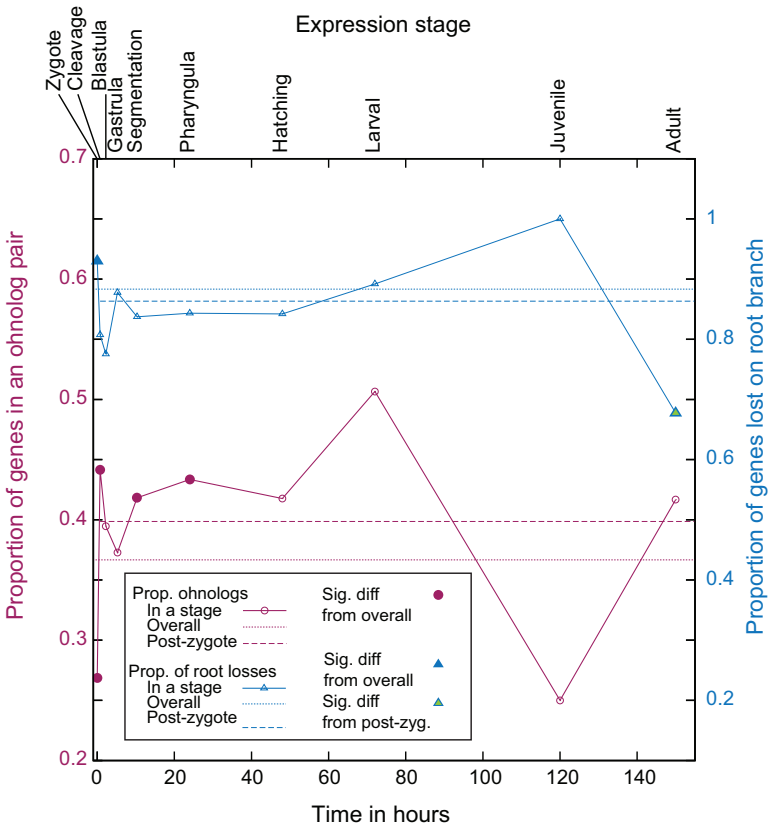


Figure 4: Timing of gene expression in development and patterns of ohnolog loss and retention. On the *x*-axis is a timeline of zebrafish development from ZFIN (Howe, Bradford, et al. 2013), with the relevant stage names indicated at the top. The trendline in red indicates the proportion of zebrafish genes with an ohnolog partner expressed at that stage (relative to total number of zebrafish genes analyzed with POInT expressed at that stage). The dotted red line is the overall proportion of genes with an ohnolog partner in the POInT dataset, while the dashed line is this proportion excluding any genes expressed in the zygote (see *Methods*). Open points show no statistically distinguishable difference from the overall proportion (chi-square test with an FDR correction, P>0.05; Benjamini and Hochberg 1995). Red-filled points are significantly different from this overall mean ($P \leq 0.05$). Trendlines in blue show similar values comparing the set of genes that POInT predicts were returned to single copy along the root branch of Figure 1 (confidence ≥ 0.85) to those only returned to single-copy along the tip branch leading to zebrafish. Hence, the right *y*-axis gives the proportion of losses that occurred along the root branch (relative to the sum of that number and the number of losses along the zebrafish branch). The dotted blue line is the overall proportion of genes returned to single-copy on the root branch (scaled as just described) while the dashed line is this proportion excluding any genes expressed in the zygote (see *Methods*). Open points are not statistically different from the overall proportion (chi-square test with an FDR correction, P>0.05; Benjamini and Hochberg 1995). Blue-filled points are significantly different from this mean ($P \leq 0.05$), while green filled points are also different from the mean seen when zygotic-expressed genes are excluded ($P \leq 0.05$).

Ohnolog pairs are unusually rare amongst genes expressed in the earliest stages of development.

As mentioned, the TGD affords the opportunity to study the effects of WGD on the evolution of developmental pathways. We had speculated that genes used in the earliest stages of development might be overly likely to be preserved in duplicate after WGD because the noise buffering effects of gene duplication might be beneficial at such times (Raser and O'Shea 2005; Pires and Conant 2016). However, such is not the case: genes with mRNAs present in the zygote were much *less* likely to be preserved as ohnolog pairs than genes first expressed later in development (Figure 4). I wondered if this observation might be driven by a dearth of ohnolog pairs among those genes where mRNA transcribed from the material genome is used in the early embryo (maternal mRNAs), since such sex-biased expression patterns might favor early gene losses. Aanes et al., (2011) have partitioned the mRNAs present in the earliest stages of zebrafish development into three groups: maternal transcripts and those seen prior to and after the mid-blastula transition. As Table 1 shows, there is some deficit of ohnologs amongst the maternally-expressed genes, but the significance of this deficit depends on the ohnolog set used, and there is no excess of early duplicate losses among this set. In contrast, the genes expressed from the embryonic genome prior to the MBT are strongly depleted in ohnologs and the single-copy genes in question are more likely to have returned to single copy along the root branch than expected (Table 1). There is then relatively little signal of ohnolog excess or deficit amongst the genes expressed later in development (post-MTB).

GO analyses show similar patterns of ohnolog loss and retention as seen in other ancient polyploids.

I used the PANTHER classification system (Mi, et al. 2017) to look for over and under-represented functions among the surviving ohnologs (and among the early ohnolog losses) in zebrafish. Supplemental Table 3 gives the complete list of significantly over and under-represented GO terms across the three hierarchies (molecular function, biological process, and cellular compartment). Here I discuss some notable results from the *Dr_Ohno_POInT* to *Dr_Sing_POInT* comparison.

Some of the Molecular Function terms over-represented among surviving ohnologs mirror results from other polyploids, such as "translation regulator activity" (P=0.01), "kinase activity" (P=0.008) and "sequence-specific DNA binding transcription factor activity," (P=0.02). Many of the Biological Process terms found to be over-represented involve various aspects of nervous system function: "nervous system development" ($P<10^{-5}$), "neuron-neuron synaptic transmission" ($P<10^{-5}$), "synaptic vesicle exocytosis" (P=0.0014) and "sensory perception" ($P<10^{-4}$).

I was particularly interested to see if the *under*-represented terms might shed any light on the relative absence of ohnologs among the mRNAs present in the earliest stages of development. And indeed, the four most statistically under-represented Biological Process (excepting "Unclassified") terms among the surviving ohnologs were for "DNA metabolic process," "translation," "tRNA metabolic process" and "RNA metabolic process" ($P<10^{-4}$ for all), while the four most significantly under-represented Molecular function terms (again excepting "Unclassified") were "methyltransferase activity," "structural constituent of ribosome," "nuclease activity" and "nucleotidyltransferase activity" ($P<10^{-3}$ for all). Since the earliest cell divisions in the embryo do not involve cell-type differentiation, the over-abundance of single-copy genes with roles in basic cellular processes (which would be needed even prior to such differentiation) are in accord with the expression timing results above.

Ohnolog pairs are unusually abundant in certain nervous and sensory tissues.

Using ZFIN data (Howe, Bradford, et al. 2013) on the anatomical locations of gene expression, I asked whether any embryological tissues had more or fewer members of ohnolog pairs expressed in them than expected, given the number of single-copy genes active in these same locations. Relative to the corresponding single copy genes (*Dr_Sing_POInT*), ohnologs (*Dr_Ohno_POInT*) are excessively likely to be expressed in the brain, diencephalon and epiphysis of the segmentation stage, (10.33-24 hours) and in the olfactory epithelium, retinal ganglion cell layer, and the retinal inner nuclear layer of the pharyngula stage (24-48 hours, P<0.05, chi-square test with FDR multiple test correction; Benjamini and Hochberg 1995). All of these locations except the olfactory epithelium also showed a significant excess of expressed ohnologs relative to single copy genes when the full set of zebrafish ohnologs were used

(*Dr_Ohno_all* verses *Dr_Sing_all,* Supplemental Table 2). One concern with this analysis might be that the data in ZFIN are biased toward surviving ohnologs due to the genes researchers have sought to localize in the embryo: however this does not appear to be the case: 58% of ohnologs (*Dr_Ohno_POInT*) were identified in at least one anatomical location, which is actually less than the 61% of the single-copy genes (*Dr_Sing_POInT*) so identified.

The TGD and the organization of the teleost retina.

The overrepresentation of ohnologs in genes expressed in parts of the retina was intriguing because there are suggestions in the literature that the complex mosaic organization of teleost retinae (Lyall 1957; Engström 1960; Stenkamp and Cameron 2002) might be an innovation due to the TGD (Braasch and Postlethwait 2012; Sukeena, et al. 2016). I conducted a GO analysis of all ohnologs and single-copy genes expressed in either the ganglion or inner cell layers of the retina at the pharyngula stage of development. No terms associated with biological process were over-represented in either tissue, and no terms associated with molecular function were over-represented in the inner cell layer. However, for the ganglion layer, the term "transmembrane transporter activity" was significantly overabundant among the surviving ohnologs (P=0.044 after FDR correction). Interestingly, the expression of duplicated genes from the TGD in these locations are probably not specific to zebrafish: the *only* two GO biological process terms that are *under* represented among the genes returned to single-copy along the root branch of Figure 1 (e.g., genes that survived in duplicate at least to the first post-TGD speciation) are "synaptic transmission" and "cell-cell signaling," while the single Cellular Compartment term similarly under represented is "neuron projection" (Supplemental Table 3). Collectively, these results suggest that the duplicated genes created by the TGD were likely involved in subsequent evolution changes in neuronal development, accounting for their retention as ohnologs across the teleost phylogeny.

Surviving TGD ohnologs are less likely to be essential.

I compared the proportion of phenotyped genes with surviving ohnologs judged to be

essential in zebrafish to the same proportion among those genes without surviving ohnologs: the genes with ohnologs show a reduced propensity to be essential (Table 2, see *Methods* for details). Importantly, this effect does not appear to be a result of any intrinsic features of these genes: when examining the two groups in the unduplicated outgroup mouse, I find that that single-copy mouse orthologs of the duplicated and the unduplicated zebrafish genes have similar essentiality in that animal. However, I also note that this effect is not a strong one: when I examined the smaller set of ohnologs with support across the eight genomes (*Dr_Ohno_POInT* verses *Dr_Sing_POInT*), the proportions shown in Table 2 are nearly identical, but the effect is non-significant due to the smaller sample size (P=0.14, chi-square test).

TGD ohnologs lie in connected parts of the zebrafish metabolic network.

I examined the position of the ohnolog pairs in the published zebrafish metabolic network (Bekaert 2012). Ohnologs are more likely to be members of this network than are single copy genes (P=0.0005 and P=0.025 for *Dr_Ohno_all* verses *Dr_Sing_all* and *Dr_Ohno_POInT* verse *Dr_Sing_POInT,* respectively). Ohnolog pairs also occupy more connected parts of this network (e.g., they have higher mean degree; Table 3), though they do not different from the single-copy genes with respect to other network statistics.

Discussion

Polyploidies of widely differing ages are ubiquitous across the tree of life (Van de Peer, et al. 2017), yet many of the detailed studies of its genome-wide effects have focused on recent events. Thus, while we know quite a bit about the fate of individual ohnolog pairs surviving from events such as the TGD and the vertebrate 2R events (Force, et al. 1999; McLysaght, et al. 2002; Steinke, et al. 2006; Makino and McLysaght 2010; Guo, et al. 2012; Kollitz, et al. 2014; Moriyama, et al. 2016; Xie, et al. 2016), we do not currently know as much about whether the patterns of genome evolution, such as adherence to the DBH and the occurrence of biased fractionation, seen after recent polyploidies, also apply to these more ancient events. Existing data should also be interpreted with caution as the methods used to identify the relics of ancient WGDs are subject to bias. Hence, Inoue et al.,'s estimates of the timing of ohnolog losses after the TGD differ from those presented here (Inoue, et al. 2015), with their estimates of the

proportion of losses along the root branch being >1.5 greater than that estimated with POInT, with an average of only 21% as many proportional losses inferred along the tip branches as POInT predicts. The reason for the discrepancy is likely that Inoue et al.,'s method does not phase post-WGD orthologs. Without such phasing, independent losses in different lineages will be mistaken for shared losses, leading to the type of over-estimates of initial losses that was apparently observed.

The data shown here support a role for the DBH in resolving the TGD: the location of ohnologs in the zebrafish metabolic network is similar to the pattern seen in the network of the polyploid plant *Arabidopsis thaliana* (Bekaert, et al. 2011) and the classes of ohnologs retained follow the predictions of the DBH (Freeling 2009; Conant, et al. 2014). However, further work will be needed to assess whether these surviving ohnologs with high interaction degree are still be maintained by selection on relative dosage or if some other force is now at work (Conant, et al. 2014).

Likewise, the TGD appears to have been an allopolyploidy, because there is strong evidence for biased fractionation and no known mechanism by which autopolyploidy could generate such biases. This result is also consistent with studies in yeast and plants (Garsmeur, et al. 2013; Marcet-Houben and Gabaldon 2015; Emery, et al. 2018). Perhaps more unexpectedly, the pattern of early gene losses after the TGD, namely among genes for the core of the cellular machinery, recall patterns seen in plants and yeast, where processes such as DNA repair were rapidly returned to single copy after polyploidy (De Smet, et al. 2013; Conant 2014). Indeed, "DNA repair" is a highly under-represented term ($P<10^{-3}$) among the zebrafish TGD ohnologs, though not one of the top 4 listed above. De Smet et al., have argued that these loss patterns suggest selection to return genes with these types of function to single copy. What was less clear from previous work is that the dearth of ohnologs among these basic processes would correspond to an excess of them involved in other processes such as multicellular development. Hence, polyploidy in multicellular organisms might concentrate its effects in such developmental processes (Holland, et al. 1994).

In this vein, the over-abundance of ohnologs expressed in the developing retina is very interesting because recent work in the spotted gar strongly suggests that the mosaic organization of teleost retainae (Lyall 1957; Engström 1960; Stenkamp and Cameron 2002) represents a morphological innovation whose evolutionary appearance was coincident with the TGD

(Sukeena, et al. 2016). Not only are ohnologs over-represented in genes expressed in some of the retinal layers, but a GO analysis suggested that many of these duplicated genes function as transmembrane transporters. Several analyses have suggested that cell-to-cell communication in the early stages of retinal development may drive the mosaic organization (Stenkamp and Cameron 2002; Raymond and Barthel 2004), and such transmembrane proteins are obvious candidates for such communication. Moreover, there may be other neurological innovations due to the TGD, if the excess of ohnologs in nervous tissues and with nervous system functions is a guide.

While duplicate genes can provide a "backup" for each other in response to gene knockout, this effect is expected to degrade as the duplicate pair ages (Gu, et al. 2003), making the apparent rarity of essential genes among the ohnolog pairs a bit surprising, given the TGD's age. However, gene essentiality and gene duplications interact with each other in a complex way. On the one hand, a gene's propensity to duplicate is associated with whether or not it is essential: small scale duplications favor less essential genes (Woods, et al. 2013), but post-WGD evolution appears to neither favor nor disfavor the retention of (formerly) essential genes after WGD (Wapinski, et al. 2007; Deluna, et al. 2008). Gene duplication then apparently imparts the (partial) redundancy seen in studies of yeast, nematodes and mice (Gu, et al. 2003; Conant and Wagner 2004; Makino, et al. 2009). I suspect that the combined observation of reduced essentiality of zebrafish ohnologs with no reduction in the essentiality of their single-copy mouse orthologs mostly likely represents surviving shared functions between ohnolog pairs that were preserved in duplicate due to other selective pressures.

The most general message apparent from these analyses is that polyploidy continues to shape the evolutionary trajectories of its possessors over very long time scales, both through first-order effects such as genetic robustness, and, more importantly, through the appearance of duplication-driven evolutionary innovations. Examples such as the changes in retinal structure just discussed are particularly important because they appear to be a class of innovations requiring changes in many genes at once, meaning that they may have only been feasible with the large number of duplicates induced by polyploidy. Though relatively few examples of such innovations are currently known (Conant and Wolfe 2007; Merico, et al. 2007; van Hoek and Hogeweg 2009; Edger, et al. 2015), as our knowledge of both polyploidy and the systems biology of the cell increases, it is very likely more will be found.

Methods

Identifying the relics of the TGD from double-conserved synteny blocks

We have developed a pipeline for inferring shared blocks of double-conserved synteny (DCS, Figure 1) from a group of genomes sharing a WGD and a reference genome from an unduplicated relative (Emery, et al. 2018). I applied this tool to eight polyploid fish genomes, taken from the Ensembl database (Aken, et al. 2017): *Astyanax mexicanus* (Cave fish; McGaugh, et al. 2014), *Danio rerio* (Zebrafish; Howe, Clark, et al. 2013), *Takifugu rubripes* (Fugu; Aparicio, et al. 2002), *Oryzias latipes* (Medaka; Kasahara, et al. 2007) *Xiphophorus maculates* (Platyfish; Schartl, et al. 2013), *Gasterosteus aculeatus* (Stickleback; Jones, et al. 2012), *Tetraodon nigroviridis* (Jaillon, et al. 2004) and *Oreochromis niloticus* (Tilapia; Brawand, et al. 2014). The genome of *Lepisosteus oculatus* (spotted gar; Braasch, et al. 2016) was used as the unduplicated outgroup.

The pipeline has three steps. First, I performed a homology search of each polyploid genome against the gar genome with GenomeHistory (Conant and Wagner 2002). I defined a gene from a polyploid genome to be a homolog of a gar gene if it had a BLAST E-value (Altschul, et al. 1997) of 10^{-8} or smaller and was 60% or more identical to that gene at the amino acid level. I further required that the length of the genes' pairwise alignment be 65% or more of their mean length and that the pair have nonsynonymous divergence (K_a) less than 0.6. These parameters result in good coverage of the genomes involved: between 70% and 80% of gar genes have a homolog in each of the TGD-possessing genomes, and between 70% and 82% of genes in those genomes have a gar homolog. Nonetheless, the parameters do not overly merge gene families, with 58% to 60% of the gar genes showing only a single homolog in the TGD-possessing genomes.

This set of homologs was then the input to the second step of the pipeline: the inference of DCS blocks in each polyploid genome. This step determines which of the potentially many homologs of a given gene in gar are the ohnologs from the TGD by maximizing the number of homologs placed in the DCS blocks. We have referred to each such gar gene and its corresponding two potential products of the TGD as a "pillar" (Byrne and Wolfe 2005; Emery, et

al. 2018): the resulting set of these *n* pillars is denoted $A_1..A_n$. Each pillar has associated with it a set of homologous genes from the polyploid genome $h_1...h_h$. At most two of these homologs can be assigned to the pillar's ohnolog positions, denoted $A_i(p_1)$ and $A_i(p_2)$. We define $A_{O(i)}$ to be the i^{th} pillar in the reordered version of this dataset. It is necessary to estimate the $A_{O(i)}$s because the teleost genomes have undergone rearrangements since the TGD (Nakatani and McLysaght 2017). Using simulated annealing (Kirkpatrick, et al. 1983; Conant and Wolfe 2006), I sought the combination of homolog assignments and pillar order that maximizes the number of pillars where the genes in neighboring pillars are also neighbors in their genome (Emery, et al. 2018). Precisely, I maximized the score *s* of such a combination of homolog assignments and pillar orders:

$$s = \sum_{i=1}^{n} \sum_{k=1}^{2} \begin{cases} 1 & A_{O(i)}(p_k) \text{ and } A_{O(i+j)}(p_k) \text{ are neighbors} \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

Once those inferences were complete for each of the eight polyploid genomes, I merged them, mapping between polyploid genomes using the gar genes as references. Taking an extremely conservative approach, I retained pillars only if each assigned homolog from every genome had synteny support in at least one direction. The result was 5589 pillars with at least one syntenic gene from each polyploid genome. I then again used simulated annealing to infer the optimal pillar order over all eight genomes. Because of the high degree of rearrangement, I made inferences of the optimal ordering under three different criteria. First, as previously (Emery, et al. 2018), I started with the order of the gar reference genes and sought orderings with the fewest total synteny breaks (*Naïve_Opt*). Second, I used an initial greedy search to place pillars with many neighboring genes in the eight extant teleost genomes near to each other, which reduced the number of initial breakpoints by about 30%. I then again sought an order with minimal breaks (*Greedy_Opt*). Finally, I sought an ordering that maximized the number of neighboring pillars having no synteny breaks between them in any genome and, after using this optimization criterion for several iterations, again applied the standard search for the fewest total breaks (*Global_Break_Opt*). I then used the inferred order that gave the highest likelihood of observing that WGD data under the WGD-$bc^{nbn}f$ gene loss model (Supplemental Table 2; see *Modeling the evolution of the TGD* below) for all further analyses.

Quality of inferred double-conserved synteny blocks

Given the ancient nature of the TGD, it is reasonable to ask if this DCS inference protocol is sufficient for these genomes. However, despite the divergence, the mapping between the genomes possessing the TGD and spotted gar is less difficult than might be expected, with 69-71% of the genes in the teleost genomes in our final dataset having only a single gar homolog (and where that gar gene matches at most 2 genes in the genome with the TGD; Supplemental Table 1). Although I required every analyzed gene to be in synteny in Step 2 of the pipeline, the estimate of a global ancestral order requires breaking some of these synteny blocks. But this problem is not serious: >94% of the genes across all the genomes with the TGD that I analyzed are in synteny blocks in the estimated ancestral order used, with the large majority in blocks of 5 or more genes (Supplemental Table 1). I provide the synteny relationships under the inferred order for the eight genomes as supplemental data.

For comparative purposes, I also explored whether a gene tree/species tree reconciliation approach might offer improved accuracy relative to synteny-based methods. Of the 132 loci in the dataset where all eight species share preserved ohnolog pairs, there are nine loci where all of the 16 genes that are members of these ohnolog pairs show clear syntenic associations in both directions. Such positions represent the best-case scenario for gene tree-based methods, as there is no gene loss to confound the inference process. I extracted the (9x8x2=144) genes in question and made codon-preserving alignments with T-Coffee (Notredame, et al. 2000) for each of the nine loci. Using phyml (Guindon and Gascuel 2003), I inferred maximum likelihood trees from these alignments under the GTR model with 4 categories of substitution rates that followed an estimated discrete gamma distribution. For none of the nine loci was the expected pair of mirrored species trees inferred (see Figure 1). In fact, of the 18 gene trees inferred (two per ohnolog pair), only 3 matched the assumed species tree, and no other topology was seen more frequently. Hence, while it is possible to infer products of the TGD by reconciling such gene trees with a proposed species tree using tools such as NOTUNG (Chen, et al. 2000), there is little reason to believe that such an approach would perform as well as the one employed. Moreover there are good theoretical reasons to believe that genes tree inference from these genomes should be challenging: most obviously, the relationships in question are characterized by long branches

and may experience significant gene conversion post-WGD (Felsenstein 1978; Evangelisti and Conant 2010; McGrath, et al. 2014; Scienski, et al. 2015).

Modeling the evolution of the TGD

I analyzed the DSC blocks from these genomes using POInT (Conant and Wolfe 2008; Conant 2014) under several models of post-WGD duplicate loss. These models have four to six states (Figure 2): **U** (undifferentiated duplicated genes), **F** (fixed duplicate genes), $\mathbf{S_1}$ and $\mathbf{S_2}$ (single copy states) and the converging states $\mathbf{C_1}$ and $\mathbf{C_2}$. These last two states model the potential for the independent parallel losses first seen in yeast (Scannell, et al. 2007; Conant and Wolfe 2008). I used likelihood ratio tests to identify the combination of these factors best fitting the data (Figure 2; Sokal and Rohlf 1995). POInT's orthology inferences for all pillars are given as Supplemental Data.

Simulating genome evolution under a model where no biased fractionation occurs.

We have previously described using POInT as a simulation engine for generating sample genome duplications (Conant and Wolfe 2008). Briefly, I started from a set of completely duplicated loci and the assumed gene order previously estimated. In locations where gene losses in one genome had generated a synteny break (e.g., after *caln1* in Figure 1), I extended the left contig to include the introduced duplicates. Then, using the maximum likelihood estimates of the model parameters and branch lengths under the WGD-*f* model, I generated a new set of post-WGD duplicate losses along the phylogeny of Figure 1. Finally, I applied the "Tracking flip prob." parameter noted in Figure 1 to model POInT's estimated errors in orthology inference, introducing new synteny breaks in the simulated genomes whenever a uniform random number was drawn with a value less than this parameter. I analyzed 100 such simulated sets of genomes with POInT under the WGD-*bf* model (e.g., biased fractionation and fixation allowed, but the δ parameter in Figure 2 set to 0) and extracted the value of ε, which is plotted in Figure 3. No simulated dataset had a value of ε as small as seen in the real dataset (P<0.01).

The TGD and the teleost phylogeny.

As our estimate for the phylogenetic relationships of these 8 species, we used the phylogeny of Near et al., (2012). While exhaustively examining all possible topologies was not computationally feasible, I was able to examine 4 trees that were near topological neighbors to that of Near et al,: all gave lower likelihoods of observing the genomic data than did that of Near et al., (Supplemental Figure 1). I note that the TimeTree package estimates that the first split between the eight taxa studied here occurred between 230 and 315 million years ago (Kumar, et al. 2017), consistent with the TGD's estimated age.

Zebrafish ohnolog and single-copy gene sets

Based on the inferences above, I defined two sets of zebrafish ohnologs and corresponding single copy genes. *Dr_Ohno_all* is the set of all ohnolog pairs that are part of DCS blocks found in the pairwise comparison of *D. rerio* to gar; *Dr_Sing_all* is the corresponding WGD loci that have returned to single copy. *Dr_Ohno_POInT* corresponds to the set of ohnologs from zebrafish for which the WGD locus/pillar in question was also identified in the other seven polyploid teleost genomes, with *Dr_Sing_POInT* being the corresponding single copy set. I also defined a pair of gene sets consisting of genes that POInT predicts with high confidence ($P \geq 0.85$) to have been returned to single copy on the shared root branch of the phylogeny in Figure 1 (*POInT_RootLosses*) and a corresponding set predicted with the same confidence to have been lost only on the branch leading to the extant *D. rerio* (e.g., after the split of zebrafish and cavefish; *POInT_DrLosses*).

Gene expression timing and WGD

I extracted the earliest developmental stage at which each zebrafish gene's transcript has been observed and the corresponding time of expression in terms of hours post-fertilization from the ZFIN database (Howe, Bradford, et al. 2013). From the same database, I also extracted all non-adult anatomical locations at which each gene's transcript had been detected. For each developmental stage and location, I used a chi-square test with a false-discovery rate correction

(Benjamini and Hochberg 1995) to test for differences in the proportion of ohnologs and non-ohnologs (*Dr_Ohno_all* vs *Dr_Sing_all* and *Dr_Ohno_POInT vs Dr_Sing_POInT*) expressed at that location. I similarly compared the proportion of single copy genes in each location and stage that were early and late losses (*POInT_RootLosses* verses *POInT_DrLosses*). For the anatomical tests, any gene expressed in the zygote was omitted from the analysis to avoid having the strong bias against ohnologs in this stage give rise to spurious associations.

Aanes et al., (2011) have partitioned gene expression in the early zebrafish embryo into three groups: genes expressed from inherited maternal transcripts, genes expressed from the embryo's genome prior to the midblastula transition (pre-MTB) and genes expressed first in the zygotic stage (e.g., post-MTB). Using these gene lists, I compared the frequency of ohnologs and single-copy (*Dr_Ohno_all* vs *Dr_Sing_all* and *Dr_Ohno_POInT vs Dr_Sing_POInT*) genes in each, again using a chi-square test, as well as the proportion of root losses and tip losses (*POInT_RootLosses* vs *POInT_DrLosses*) with the same approach (Supplemental Table 3).

GO analyses

To understand how gene function may have shaped the resolution of the TGD, I used the Gene List Analysis tool from the PANTHER classification system (version 13.1; Mi, et al. 2017) to find over or under-represented Gene Ontology (GO) terms associated with the surviving ohnologs (*Dr_Ohno_all* compared to *Dr_Sing_all* and *Dr_Ohno_POInT* to *Dr_Sing_POInT*) and the early verses late ohnolog losses (*POInT_RootLosses* compared to *POInT_DrLosses*). In each case, I asked whether there were any biological processes, molecular functions, or cellular component ontology terms that were significantly over or under-represented on the first list, using Fisher's exact test with an FDR multiple test correction (Mi, et al. 2013; Mi, et al. 2017). Lists of all significantly enriched terms for any comparison are given as Supplemental Table 4.

Gene essentiality and the TGD

From ZFIN (Howe, Bradford, et al. 2013), I extracted all genes with known phenotypes, as well as the subset of those genes with phenotypes described as "lethal," "dead" or "inviable:"

hereafter I note this second set as the "essential genes." I compared the proportion of phenotyped ohnologs in the essential list to the same proportion among the single copy genes. For comparative purposes, I obtained a list of essential mouse genes from the International Mouse Phenotyping Consortium (Koscielny, et al. 2013; Dickinson, et al. 2016). Using our orthology inference pipeline, I inferred the gar orthologs of these mouse genes (Conant 2009; Bekaert and Conant 2011), retrieving 10,644 gar genes with a mouse ortholog. For each gar gene with phenotype data in a mouse ortholog, we compared the proportion of genes with a surviving ohnolog in zebrafish that were essential when knocked out in mouse to the proportion of genes without a surviving zebrafish ohnolog pair that were essential (Table 2; other phenotype classes such as "subviable" were excluded).

The TGD and the zebrafish metabolic network.

I extracted an enzyme-centered metabolic network from the reconstruction of zebrafish metabolism published by Bekaert (2012). In this network nodes are biochemical reactions and edges connect pairs of nodes with a common metabolite. The 13 currency metabolites given by Bekaert (2012) were excluded from the edge computation. Each reaction was linked to one or more Ensembl gene identifiers corresponding to genes encoding enzymes catalyzing that reaction.

To test for differences in network position between the products of ohnologs and single-copy genes, I compared the two groups for their average degree (number of edges), clustering coefficient (indicating the propensity of connected nodes to have common neighbors; Watts and Strogatz 1998), and betweenness-centrality (the number of the network's shortest paths passing through a given node; Hahn and Kern 2005). I then used randomization to assess the statistical significance of the differences. To maintain the structure introduced by the WGD, all ohnolog pairs were reduced to a single entity, which was then assigned to all nodes that products of either of the two ohnologs appeared in. These merged ohnolog products were then randomized along with the products of the single copy genes, and the differences in the three statistics for each randomized network recomputed. If less than 5% of the randomized networks had a difference as large as that observed for the real data, I concluded that there was evidence for a difference between duplicated and unduplicated genes.

Acknowledgements:

I would like to thank J. C. Pires, J. Thorne and X. Ji for helpful discussions and K. Dudley for computational assistance. This work was supported by the United States National Science Foundation (grant numbers NSF-IOS-1339156 and NSF-CCF-1421765).

Table 1. Expression timing and fate of TGD-produced ohnologs

Expression cluster[a]	1st gene set	2nd gene set	Prop. of 1st set in cluster[b]	Prop. of 2nd set in cluster[b]	*P*[c]
Maternal transcripts[d]	*Dr_Ohno_all*[e]	*Dr_Sing_all*[e]	0.03 (116/4279)	0.04 (484/11616)	**2.4x10^{-5}**
	Dr_Ohno_POInT[f]	*Dr_Sing_POInT*[f]	0.03 (81/2552)	0.04 (193/4408)	**0.015**
	POInT_RootLosses[g]	*POInT_DrLosses*[g]	0.05 (103/1894)	0.03 (8/250)	0.18
Pre-MBT transcripts[h]	*Dr_Ohno_all*	*Dr_Sing_all*	0.10 (435/4279)	0.15 (1709/11616)	**1.2x10^{-13}**
	Dr_Ohno_POInT	*Dr_Sing_POInT*	0.11 (284/2552)	0.17 (762/4408)	**3.9x10^{-12}**
	POInT_RootLosses	*POInT_DrLosses*	0.19 (351/1894)	0.12 (31/250)	**0.022**
Zygotic transcripts[i]	*Dr_Ohno_all*	*Dr_Sing_all*	0.06 (250/4279)	0.05 (573/11616)	**0.024**
	Dr_Ohno_POInT	*Dr_Sing_POInT*	0.06 (142/2552)	0.05 (216/4408)	0.25
	POInT_RootLosses	*POInT_DrLosses*	0.05 (92/1894)	0.05 (12/250)	>0.95

a: Three expression clusters drawn from Aanes et al., (Aanes, et al. 2011).

b: Proportion of all genes in the 1st set (see left) that were observed to be expressed in the cluster in question, with the total number of expressed genes over the total number of genes in that set given in parentheses.

c: *P*-value for the hypothesis test of equal proportion of genes in both sets falling into the expression cluster (chi-square test with 1 degree of freedom)

d: Genes determined by Aanes et al., to have been expressed in the developing embryo from maternally-derived transcripts.

e: Comparison of all identified zebrafish ohnologs to all zebrafish single-copy (with respect to the TGD) genes. See *Methods* for further details.

f: Comparison of all zebrafish ohnolog pairs found in the 8-species POInT analysis to the corresponding zebrafish single-copy (with respect to the TGD) genes. See *Methods* for further details.

g: Comparison of zebrafish single-copy genes inferred to have been lost on the common root branch of Figure 1 to zebrafish single-copy genes inferred by POInT to have been lost after the zebrafish/cavefish split (inference confidence = 0.85 in both cases). See *Methods* for further details.

h: Genes determined by Aanes et al., to have been expressed in the developing embryo prior to the mid-blastula transition (<3.5 hours post-fertilization).

i: Genes determined by Aanes et al., to have been expressed in the developing embryo only after the mid-blastula transition (>3.5 hours post-fertilization).

Table 2. Essentiality and the TGD

Essentiality data[a]	**Prop. of phenotyped genes with an ohnolog that are essential**[b]	**Prop. of phenotyped genes without an ohnolog that are essential**[c]	***P***[d]
Zebrafish[e]	0.062 (6/97)[f]	0.145 (46/318)[f]	**0.048**
Mouse[g]	0.556 (42/72)[h]	0.506 (161/318)[h]	0.53

a: We compared the essentiality of genes with and without surviving TGD-produced ohnologs in zebrafish for their essentiality in zebrafish and the essentiality of their single-copy orthologs from mouse.

b: Proportion of genes with a ZFIN phenotype with a surviving TGD-produced ohnolog that are essential in the dataset in question (set *Dr_Ohno_all*).

c: Proportion of of genes with a ZFIN phenotype without a surviving TGD-produced ohnolog that are essential in the dataset in question (set *Dr_Sing_all*).

d: *P*-value for the hypothesis test of equal proportion of essential genes in *Dr_Ohno_all* vs *Dr_Sing_all*.

e: Essentiality defined as genes in the ZFIN database (Howe, Bradford, et al. 2013) phenotyped as "lethal," "dead" or "inviable."

f: Numbers in the parenthesis give the number of essential genes over the total number in the set.

g: Essentiality defined by the International Mouse Phenotyping Consortium's list of essential mouse genes (Koscielny, et al. 2013; Dickinson, et al. 2016).

h: Numbers in the parenthesis give the number of essential genes over the total number in the set. Note that ohnolog pairs in zebrafish are by definition single-copy in gar and mouse, accounting for the smaller number of comparisons in the ohnolog column for the mouse analysis.

Table 3. The TGD and the zebrafish metabolic network

Network statistic[a]	Ohnolog datasets compared[b]	Mean ohnolog value[c]	Mean single-copy gene value[d]	*P*[e]
Node degree[f]	*Dr_Ohno_all/Dr_Sing_all*[g]	30.9	21.4	**0.002**
	Dr_Ohno_POInT Dr_Sing_POInT[h]	31.5	23.2	**0.032**
Avg. clustering coeff.[i]	*Dr_Ohno_all/Dr_Sing_all*[g]	0.78	0.77	>0.5
	Dr_Ohno_POInT Dr_Sing_POInT[h]	0.77	0.76	>0.5
Mean # shortest paths[j]	*Dr_Ohno_all/Dr_Sing_all*[g]	18020	12904	0.07
	Dr_Ohno_POInT Dr_Sing_POInT[h]	19280	14132	0.18

a: Network statistic for which the ohnologs and single copy genes were compared.

b: Derivation of the set of ohnologs and single-copy genes.

c: Mean value of the statistic in question for the ohnolog pairs (ohnolog pairs were merged and averaged prior to computing the global average).

d: Mean value of the statistic in question for the single-copy genes.

e: *P*-value for the hypothesis test of equal mean statistic value for the ohnologs and single-copy genes (Network randomization test; *Methods*).

f: Number of edges per network node.

g: Comparison of all identified zebrafish ohnologs to all zebrafish single-copy (with respect to the TGD) genes. See *Methods* for further details.

h: Comparison of all zebrafish ohnolog pairs used in the 8 species POInT analysis to the corresponding zebrafish single-copy (with respect to the TGD) genes. See *Methods* for further details.

i: Ratio of the number of edges between each triplet of nodes to the maximum number of such connections possible (Watts and Strogatz 1998).

j: The mean of the number of shortest paths through the network that pass through a given node, also known as betweenness-centrality (Hahn and Kern 2005).

Supplemental Information:

Supplemental Figure 1: Alternative phylogenetic topologies and the TGD. I tested 4 other topologies in addition to that of Near et al., (Near, et al. 2012) using the optimal DCS block order and the WGD-$bc^{nbn}f$ model in POInT: shown are the induced branch lengths for these topologies as well as the corresponding model parameter estimates. For reference, the topology at the bottom is that of Near et al., used for all other analyses.

Supplemental Table 1: Quality of inferred DCS blocks across the eight genomes.

Supplemental Table 2: Effects of the estimated ancestral pillar order on parameter estimates from POInT.

Supplemental Table 3: Developmental time points and anatomical regions with more or fewer ohnologs present than expected.

Supplemental Table 4: Differentially abundant GO terms for Ohnologs and single-copy genes from the TGD.

Supplemental Data: A compressed tar file (http://conantlab.org/data/TGD/Supplemental_data.tar.gz) consisting of nine files: 1) POInT's orthology inferences for the 5589 pillars analyzed (tab-delimited text), giving the estimated probability of all 2^8=256 possible orthology relationships at every pillar. 2-9) For each genome with the TGD, I give the synteny relationships seen in the estimated optimal ancestral order (tab-delimited text).

References:

Aanes H, Winata CL, Lin CH, Chen JP, Srinivasan KG, Lee SG, Lim AY, Hajan HS, Collas P, Bourque G, et al. 2011. Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. Genome Res 21:1328-1338.

Aken BL, Achuthan P, Akanni W, Amode MR, Bernsdorff F, Bhai J, Billis K, Carvalho-Silva D, Cummins C, Clapham P, et al. 2017. Ensembl 2017. Nucleic Acids Res 45:D635-D642.

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997. Gapped Blast and Psi-Blast : A new-generation of protein database search programs. Nucleic Acids Research 25:3389-3402.

Alves M, Coelho M, Collares-Pereira M. 2001. Evolution in action through hybridisation and polyploidy in an Iberian freshwater fish: a genetic review. Genetica 111:375-385.

Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, et al. 2002. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science 297:1301-1310.

Aury JM, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Segurens B, Daubin V, Anthouard V, Aiach N, et al. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature 444:171-178.

Bekaert M. 2012. Reconstruction of Danio rerio metabolic model accounting for subcellular compartmentalisation. PloS one 7:e49903.

Bekaert M, Conant GC. 2011. Copy number alterations among mammalian enyzmes cluster in the metabolic network. Molecular Biology and Evolution 28:1111-1121.

Bekaert M, Edger PP, Pires JC, Conant GC. 2011. Two-phase resolution of polyploidy in the *Arabidopsis* metabolic network gives rise to relative followed by absolute dosage constraints. The Plant Cell 23:1719-1728.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B (Methodological) 57:289-300.

Birchler JA, Veitia RA. 2007. The gene balance hypothesis: from classical genetics to modern genomics. Plant Cell 19:395-402.

Bird KA, VanBuren R, Puzey JR, Edger PP. 2018. The causes and consequences of subgenome dominance in hybrids and recent polyploids. New Phytologist.

Blanc-Mathieu R, Perfus-Barbeoch L, Aury J-M, Da Rocha M, Gouzy J, Sallet E, Martin-Jimenez C, Bailly-Bechet M, Castagnone-Sereno P, Flot J-F. 2017. Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. PLoS genetics 13:e1006777.

Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J, et al. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat Genet 48:427-437.

Braasch I, Postlethwait JH. 2012. Polyploidy in fish and the teleost genome duplication. In. Polyploidy and genome evolution: Springer. p. 341-383.

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezault E, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513:375-381.

Byrne KP, Wolfe KH. 2005. The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. Genome research 15:1456-1461.

Chen K, Durand D, Farach-Colton M. 2000. NOTUNG: a program for dating gene duplications and optimizing gene family trees. Journal of Computational Biology 7:429-447.

Chenuil A, Galtier N, Berrebi P. 1999. A test of the hypothesis of an autopolyploid vs. allopolyploid origin for a tetraploid lineage: application to the genus Barbus (Cyprinidae). Heredity 82:373.

Christoffels A, Koh EG, Chia J-m, Brenner S, Aparicio S, Venkatesh B. 2004. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. Molecular biology and evolution 21:1146-1151.

Clausen R, Goodspeed T. 1925. Interspecific hybridization in Nicotiana. II. A tetraploid glutinosa-tabacum hybrid, an experimental verification of Winge's hypothesis. Genetics 10:278.

Conant GC. 2014. Comparative genomics as a time machine: How relative gene dosage and metabolic requirements shaped the time-dependent resolution of yeast polyploidy. Mol Biol Evol 31:3184-3193.

Conant GC. 2009. Neutral evolution on mammalian protein surfaces. Trends in genetics 25:377-381.

Conant GC, Birchler JA, Pires JC. 2014. Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. Curr Opin Plant Biol 19:91-98.

Conant GC, Wagner A. 2004. Duplicate genes and robustness to transient gene knockouts in *Caenorhabditis elegans*. Proceedings of the Royal Society, Biological Sciences 271:89-96.

Conant GC, Wagner A. 2002. GenomeHistory: A software tool and its application to fully sequenced genomes. Nucleic Acids Research 30:3378-3386.

Conant GC, Wolfe KH. 2006. Functional partitioning of yeast co-expression networks after genome duplication. PLoS Biology 4:e109.

Conant GC, Wolfe KH. 2007. Increased glycolytic flux as an outcome of whole-genome duplication in yeast. Molecular Systems Biology 3:129.

Conant GC, Wolfe KH. 2008. Probabilistic cross-species inference of orthologous genomic regions created by whole-genome duplication in yeast. Genetics 179:1681-1692.

Crow KD, Stadler PF, Lynch VJ, Amemiya C, Wagner GnP. 2005. The “fish-specific” Hox cluster duplication is coincident with the origin of teleosts. Molecular Biology and Evolution 23:121-136.

De Smet R, Adams KL, Vandepoele K, Van Montagu MC, Maere S, Van de Peer Y. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc Natl Acad Sci U S A 110:2898-2903.

Deluna A, Vetsigian K, Shoresh N, Hegreness M, Colon-Gonzalez M, Chao S, Kishony R. 2008. Exposing the fitness contribution of duplicated genes. Nature Genetics.

Dickinson ME, Flenniken AM, Ji X, Teboul L, Wong MD, White JK, Meehan TF, Weninger WJ, Westerberg H, Adissu H. 2016. High-throughput discovery of novel developmental phenotypes. Nature 537:508.

Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, Tang M. 2015. The butterfly plant arms-race escalated by gene and genome duplications. Proceedings of the National Academy of Sciences 112:8362-8366.

Edger PP, Pires JC. 2009. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res 17:699-717.

Emery M, Willis MMS, Hao Y, Barry K, Oakgrove K, Peng Y, Schmutz J, Lyons E, Pires JC, Edger PP, et al. 2018. Preferential retention of genes from one parental genome after polyploidy illustrates

the nature and scope of the genomic conflicts induced by hybridization. PLoS Genetics 14:e1007267em.

Engström K. 1960. Cone types and cone arrangement in the retina of some cyprinids. Acta Zoologica 41:277-295.

Evangelisti AM, Conant GC. 2010. Nonrandom survival of gene conversions among yeast ribosomal proteins duplicated through genome doubling. Genome Biology and Evolution 2:826-834.

Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematic Zoology 27:401-410.

Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531-1545.

Freeling M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annual Review of Plant Biology 60:433-453.

Freeling M, Thomas BC. 2006. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome research 16:805-814.

Garsmeur O, Schnable JC, Almeida A, Jourda C, D'Hont A, Freeling M. 2013. Two Evolutionarily Distinct Classes of Paleopolyploidy. Molecular Biology and Evolution 31:448-454.

Gordon JL, Byrne KP, Wolfe KH. 2009. Additions, losses and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. PLoS Genetics 5:e1000485.

Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li W-H. 2003. Role of duplicate genes in genetic robustness against null mutations. Nature 421:63-66.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52:696-704.

Guo B, Zou M, Wagner A. 2012. Pervasive indels and their evolutionary dynamics after the fish-specific genome duplication. Molecular biology and evolution 29:3005-3022.

Hahn MW, Kern AD. 2005. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Molecular Biology and Evolution 22:803-806.

Hoegg S, Boore JL, Kuehl JV, Meyer A. 2007. Comparative phylogenomic analyses of teleost fish Hox gene clusters: lessons from the cichlid fish Astatotilapia burtoni. BMC genomics 8:317.

Hoegg S, Brinkmann H, Taylor JS, Meyer A. 2004. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. Journal of molecular evolution 59:190-203.

Holland PWH, Garciafernandez J, Williams NA, Sidow A. 1994. Gene Duplications and the Origins of Vertebrate Development. Development:125-133.

Howe DG, Bradford YM, Conlin T, Eagle AE, Fashena D, Frazer K, Knight J, Mani P, Martin R, Moxon SA, et al. 2013. ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics. Nucleic Acids Res 41:D854-860.

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, et al. 2013. The zebrafish reference genome sequence and its relationship to the human genome. Nature 496:498-503.

Inoue J, Sato Y, Sinclair R, Tsukamoto K, Nishida M. 2015. Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling. Proc Natl Acad Sci U S A 112:14918-14923.

Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, et al. 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature 431:946-957.

Jensen LJ, Julien P, Kuhn M, von Mering C, Muller J, Doerks T, Bork P. 2008. eggNOG: automated construction and annotation of orthologous groups of genes. Nucleic Acids Research 36:D250-254.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55-61.

Kasahara M. 2007. The 2R hypothesis: an update. Curr Opin Immunol 19:547-552.

Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, et al. 2007. The medaka draft genome and insights into vertebrate genome evolution. Nature 447:714-719.

Kirkpatrick S, Gelatt CDJ, Vecchi MP. 1983. Optimization by simulated annealing. Science 220:671-680.

Kollitz EM, Hawkins MB, Whitfield GK, Kullman SW. 2014. Functional diversification of vitamin D receptor paralogs in teleost fish after a whole genome duplication event. Endocrinology 155:4641-4654.

Koscielny G, Yaikhom G, Iyer V, Meehan TF, Morgan H, Atienza-Herrero J, Blake A, Chen C-K, Easty R, Di Fenza A. 2013. The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. Nucleic acids research 42:D802-D809.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol Biol Evol 34:1812-1819.

Kuwada Y. 1911. Maiosis in the Pollen Mother Cells of Zea Mays L.(With Plate V.). 植物学雑誌 25:163-181.

Li L, Stoeckert CJ, Jr., Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome research 13:2178-2189.

Lyall A. 1957. Cone arrangements in teleost retinae. Journal of Cell Science 3:189-201.

Makino T, Hokamp K, McLysaght A. 2009. The complex relationship of gene duplication and essentiality. Trends Genet 25:152-155.

Makino T, McLysaght A. 2010. Ohnologs in the human genome are dosage balanced and frequently associated with disease. Proc Natl Acad Sci U S A 107:9270-9274.

Marcet-Houben M, Gabaldon T. 2015. Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage. PLoS Biol 13:e1002220.

McGaugh SE, Gross JB, Aken B, Blin M, Borowsky R, Chalopin D, Hinaux H, Jeffery WR, Keene A, Ma L, et al. 2014. The cavefish genome reveals candidate genes for eye loss. Nat Commun 5:5307.

McGrath CL, Gout J-F, Johri P, Doak TG, Lynch M. 2014. Differential retention and divergent resolution of duplicate genes following whole-genome duplication. Genome research 24:1665-1675.

McLysaght A, Hokamp K, Wolfe KH. 2002. Extensive genomic duplication during early chordate evolution. Nature Genetics 31:200-204.

Merico A, Sulo P, Piškur J, Compagno C. 2007. Fermentative lifestyle in yeasts belonging to the *Saccharomyces* complex. FEBS Journal 274:976-989.

Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res 45:D183-D189.

Mi H, Muruganujan A, Casagrande JT, Thomas PD. 2013. Large-scale gene function analysis with the PANTHER classification system. Nat Protoc 8:1551-1566.

Moriyama Y, Ito F, Takeda H, Yano T, Okabe M, Kuraku S, Keeley FW, Koshiba-Takeuchi K. 2016. Evolution of the fish heart by sub/neofunctionalization of an elastin gene. Nat Commun 7:10397.

Nakatani Y, McLysaght A. 2017. Genomes as documents of evolutionary history: a probabilistic macrosynteny model for the reconstruction of ancestral genomes. Bioinformatics 33:i369-i378.

Nakatani Y, McLysaght A. 2019. Macrosynteny analysis shows the absence of ancient whole-genome duplication in lepidopteran insects. Proceedings of the National Academy of Sciences 116:1816-1818.

Naruse K, Tanaka M, Mita K, Shima A, Postlethwait J, Mitani H. 2004. A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. Genome research 14:820-828.

Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith WL. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proceedings of the National Academy of Sciences 109:13698-13703.

Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of Molecular Biology 302:205-217.

Ohno S. 1970. Evolution by gene duplication. New York: Springer.

Pires JC, Conant GC. 2016. Robust Yet Fragile: Expression Noise, Protein Misfolding and Gene Dosage in the Evolution of Genomes. Annual Review of Genetics 50:113–131.

Raser JM, O'Shea EK. 2005. Noise in gene expression: origins, consequences, and control. Science 309:2010-2013.

Raymond PA, Barthel LK. 2004. A moving wave patterns the cone photoreceptor mosaic array in the zebrafish retina. Int J Dev Biol 48:935-945.

Sankoff D, Zheng C, Zhu Q. 2010. The collapse of gene complement following whole genome duplication. BMC genomics 11:313.

Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature 440:341-345.

Scannell DR, Frank AC, Conant GC, Byrne KP, Woolfit M, Wolfe KH. 2007. Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. Proceedings of the National Academy of Sciences, U.S.A. 104:8397-8402.

Schartl M, Walter RB, Shen Y, Garcia T, Catchen J, Amores A, Braasch I, Chalopin D, Volff JN, Lesch KP, et al. 2013. The genome of the platyfish, Xiphophorus maculatus, provides insights into evolutionary adaptation and several complex traits. Nat Genet 45:567-572.

Schnable JC, Springer NM, Freeling M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proceedings of the National Academy of Sciences 108:4069-4074.

Schwager EE, Sharma PP, Clarke T, Leite DJ, Wierschin T, Pechmann M, Akiyama-Oda Y, Esposito L, Bechsgaard J, Bilde T, et al. 2017. The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. BMC Biol 15:62.

Scienski K, Fay JC, Conant GC. 2015. Patterns of Gene Conversion in Duplicated Yeast Histones Suggest Strong Selection on a Coadapted Macromolecular Complex. Genome biology and evolution 7:3249-3258.

Semon M, Wolfe KH. 2007. Reciprocal gene loss between Tetraodon and zebrafish after whole genome duplication in their ancestor. Trends in genetics 23:108-112.

Sémon M, Wolfe KH. 2007. Rearrangement rate following the whole-genome duplication in teleosts. Molecular biology and evolution 24:860-867.

Sokal RR, Rohlf FJ. 1995. Biometry: 3rd Edition. New York: W. H. Freeman and Company.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, dePamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and angiosperm diversification. Am. J. Bot. 96:336-348.

Steinke D, Hoegg S, Brinkmann H, Meyer A. 2006. Three rounds (1R/2R/3R) of genome duplications and the evolution of the glycolytic pathway in vertebrates. BMC Biology 4:16.

Stenkamp DL, Cameron DA. 2002. Cellular pattern formation in the retina: retinal regeneration as a model system. Mol Vis 8:280-293.

Sukeena JM, Galicia CA, Wilson JD, McGinn T, Boughman JW, Robison BD, Postlethwait JH, Braasch I, Stenkamp DL, Fuerst PG. 2016. Characterization and evolution of the Spotted Gar retina. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 326:403-421.

Tang H, Woodhouse MR, Cheng F, Schnable JC, Pedersen BS, Conant G, Wang X, Freeling M, Pires JC. 2012. Altered patterns of fractionation and exon deletions in Brassica rapa support a two-step model of paleohexaploidy. Genetics 190:1563-1574.

Taylor JS, Raes J. 2004. Duplication and divergence: The evolution of new genes and old ideas. Annual Review of Genetics 38:615-643.

Thomas BC, Pedersen B, Freeling M. 2006. Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome Res 16:934-946.

Van de Peer Y. 2004. Tetraodon genome confirms Takifugu findings: most fish are ancient polyploids. Genome Biol 5:250.

Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. Nature Reviews Genetics 18:411-424.

van Hoek MJ, Hogeweg P. 2009. Metabolic adaptation after whole genome duplication. Molecular Biology and Evolution 26:2441-2453.

Vandepoele K, De Vos W, Taylor JS, Meyer A, Van de Peer Y. 2004. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc Natl Acad Sci U S A 101:1638-1643.

Veitia RA, Potier MC. 2015. Gene dosage imbalances: action, reaction, and models. Trends in biochemical sciences 40:309-317.

Wapinski I, Pfeffer A, Friedman N, Regev A. 2007. Natural history and evolutionary principles of gene duplication in fungi. Nature 449:54-61.

Watts DJ, Strogatz SH. 1998. Collective dynamics of 'small-world' networks. Nature 393:440-442.

Wendel JF, Lisch D, Hu G, Mason AS. 2018. The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. Current opinion in genetics & development 49:1-7.

Wolfe KH. 2000. Robustness: It's not where you think it is. Nature Genetics 25:3-4.

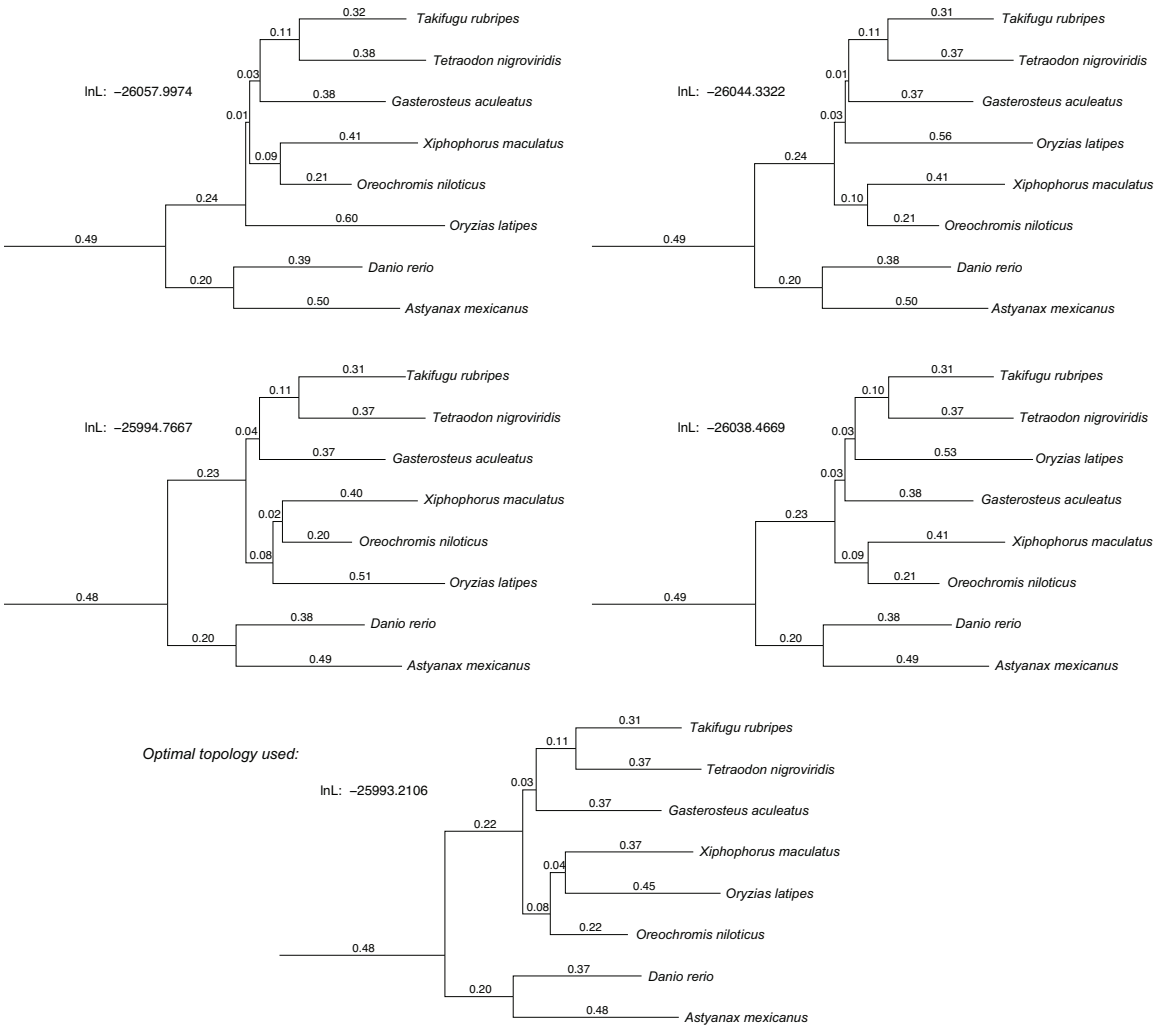
Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387:708-713.

Woods S, Coghlan A, Rivers D, Warnecke T, Jeffries SJ, Kwon T, Rogers A, Hurst LD, Ahringer J. 2013. Duplication and retention biases of essential and non-essential genes revealed by systematic knockdown analyses. PLoS Genetics 9:e1003330.

Xie T, Yang Q-Y, Wang X-T, McLysaght A, Zhang H-Y. 2016. Spatial colocalization of human ohnolog pairs acts to maintain dosage-balance. Molecular biology and evolution 33:2368-2375.

Yang L, Sado T, Hirt MV, Pasco-Viel E, Arunachalam M, Li J, Wang X, Freyhof J, Saitoh K, Simons AM. 2015. Phylogeny and polyploidy: resolving the classification of cyprinine fishes (Teleostei: Cypriniformes). Molecular Phylogenetics and Evolution 85:97-116.

1 Zwaenepoel A, Li Z, Lohaus R, Van de Peer Y. 2019. Finding evidence for whole genome duplications: a
2 reappraisal. Molecular plant 12:133-136.
3



Supplemental Figure 1: Alternative phylogenetic topologies and the TGD. I tested 4 other topologies in addition to that of Near et al., (Near et al. 2012) using the optimal DCS block order and the WGD-$bc^{nbn}f$ model in POInT: shown are the induced branch lengths for these topologies as well as the corresponding model parameter estimates. For reference, the topology at the bottom is that of Near et al., used for all other analyses.

Supplemental Table 1: Inference of Double-Conserved Synteny (DCS) blocks from eight teleost genomes by reference to the spotted gar genome.

Species	Total genes in dataset[a]	% of genes with 1:1/1:2 homology to gar[b]	% genes in synteny blocks[c]	% genes in synteny blocks w/ >5 genes[d]	# of genes not in synteny blocks with ambig. homology[e]
Astyanax mexicanus	6707	71.4	94.0	80.2	116
Danio rerio	6849	69.3	95.3	85.1	97
Gasterosteus aculeatus	6799	69.6	99.2	98.0	22
Oreochromis niloticus	6936	69.6	99.4	97.8	13
Oryzias latipes	6601	69.9	99.3	97.3	12
Takifugu rubripes	6725	70.4	99.0	96.6	34
Tetraodon nigroviridis	6660	70.0	98.7	96.4	41
Xiphophorus maculatus	6685	70.1	98.0	93.4	41

a: Number of genes from the genome placed into the shared DCS blocks used for POInT analyses (see *Methods*).

b: Percent of cases in the genome in question where genes in the final dataset from that genome match one and only one gene in the gar genome and the gar gene in question matches at most two genes in that genome.

c: Percentage of genes from the genome in question where that gene is in a synteny block with at least one other gene using the optimal ancestral order from Supplemental Table 2.

d: Percentage of genes from the genome in question where that gene is in a synteny block with at least four other genes (blocks of size 5) using the optimal ancestral order from Supplemental Table 2.

e: Number of genes in the dataset analyzed which are not in synteny blocks in the optimal order that also have some ambiguity in their homology relationships: either a teleost gene with multiple gar homologs or where the gar gene in question has >2 homologs in this genome.

Supplemental Table 2: Effects of the estimated ancestral pillar order on parameter estimates from POInT.

Optimization Method[a]	Breaks[b]	lnL[c]	γ[d]	ε[d]	δ[d]	θ[d]
Global_Break_Opt	14342	-26240.049	0.075	0.676	0.342	0.0089
	14693	-26170.596	0.074	0.669	0.339	0.0085
	14787	-26188.226	0.075	0.684	0.343	0.0084
	14965[e]	*-26080.462*	*0.074*	*0.657*	*0.341*	*0.0079*
	15148	-26178.494	0.076	0.683	0.352	0.0076
	15966	-26272.204	0.076	0.728	0.382	0.0076
	16323	-26272.011	0.075	0.673	0.361	0.0073
	16766	-26323.416	0.074	0.680	0.379	0.0074
	17523	-26265.175	0.076	0.723	0.407	0.0066
	21656	-27092.183	0.071	0.634	0.416	0.0052
Greedy_Opt	13852	-26226.784	0.077	0.701	0.370	0.0097
	13898	-26203.132	0.078	0.711	0.370	0.0087
	13937	-26103.825	0.077	0.685	0.356	0.0084
	13974	-26169.419	0.077	0.704	0.361	0.0087
	14086	-26136.803	0.079	0.751	0.383	0.008
	14370[f]	***-25993.112***	***0.077***	***0.701***	***0.371***	***0.0074***
	14698	-26114.818	0.076	0.681	0.370	0.0084
	14847	-26147.026	0.076	0.695	0.377	0.0086
	15105	-26207.243	0.073	0.634	0.354	0.0079
	15253	-26175.667	0.075	0.680	0.369	0.008
Naïve_Opt	13658	-26362.214	0.074	0.61	0.326	0.01
	13695	-26290.942	0.076	0.662	0.343	0.0091
	13772	-26314.732	0.078	0.687	0.357	0.0091
	13897	-26324.514	0.079	0.669	0.345	0.0083
	13912	-26323.322	0.081	0.722	0.365	0.0084
	13945	-26238.085	0.080	0.699	0.358	0.0081
	13978	-26257.832	0.077	0.661	0.344	0.0088
	14002	-26279.572	0.079	0.708	0.363	0.0086
	14010[g]	*-26161.701*	*0.078*	*0.674*	*0.351*	*0.0084*
	14122	-26336.625	0.076	0.6388	0.333	0.0088

a: Three methods for estimating the ancestral pillar order at the TGD were used: See *Methods* for details.

b: Number of synteny breaks in the estimated order (see *Methods*).

c: ln-likelihood of the pillar data under the assumed order and the WGD-$bc^{nbn}f$ model (Figure 2).

d: Parameter estimates for the WGD-$bc^{nbn}f$ model: γ=fixation rate, ε=strength of biased fractionation, δ=rate of duplicates entering the converging state, θ=estimated pillar-to-pillar probability of a change in orthology assignment. See Figure 2.

e: Order with the largest lnL, Global_Break_Opt

f: Order with the **overall** largest lnL across all three optimization methods.

g: Order with the largest lnL, Naïve_Opt

Supplemental Table 3: Developmental time points and anatomical regions with more or fewer ohnologs present than expected.

Anatomical Region	Dr_Ohno_POInT: #Ohnologs (Expected)	FDR-corrected P-value	Dr_Ohno_All: #Ohnologs (Expected)	FDR-corrected P-value
Blastula_EVL	10 (9.45)	1.000	14 (13.69)	1.00
Blastula_YSL	20 (25.98)	0.738	34 (39.12)	0.77
Blastula_margin	4 (5.51)	1.000	6 (7.82)	0.97
Blastula_proliferativeregion	2 (1.57)	1.000	4 (4.47)	1.00
Gastrula_adaxialcell	19 (17.32)	1.000	28 (25.43)	0.97
Gastrula_anterioraxialhypoblast	3 (5.12)	0.992	6 (6.99)	1.00
Gastrula_axis	11 (12.6)	1.000	18 (17.04)	1.00
Gastrula_centralnervoussystem	63 (59.05)	1.000	*93 (70.13)*	*0.02*
Gastrula_ectoderm	3 (3.15)	1.000	5 (5.59)	1.00
Gastrula_endoderm	3 (3.94)	1.000	7 (7.26)	1.00
Gastrula_forebrainneuralkeel	3 (2.76)	1.000	7 (5.87)	1.00
Gastrula_forerunnercellgroup	4 (3.94)	1.000	7 (8.94)	0.97
Gastrula_head	11 (11.02)	1.000	14 (14.53)	1.00
Gastrula_hindbrainneuralplate	6 (6.3)	1.000	11 (9.22)	0.97
Gastrula_integument	19 (15.75)	1.000	29 (22.35)	0.38
Gastrula_mesoderm	10 (8.27)	1.000	18 (13.97)	0.59
Gastrula_midbrainneuralkeel	4 (5.51)	1.000	4 (6.15)	0.85
Gastrula_neuralkeel	0 (1.18)	1.000	1 (2.51)	0.86
Gastrula_neuralplate	22 (22.44)	1.000	31 (31.01)	1.00
Gastrula_notochord	38 (35.43)	1.000	57 (48.06)	0.42
Gastrula_opticprimordium	6 (4.72)	1.000	9 (7.82)	1.00
Gastrula_oticplacode	16 (17.32)	1.000	28 (30.18)	1.00
Gastrula_paraxialmesoderm	6 (4.72)	1.000	10 (6.43)	0.42
Gastrula_periderm	16 (15.75)	1.000	26 (25.98)	1.00
Gastrula_polster	20 (17.72)	1.000	28 (25.15)	0.97
Gastrula_segmentalplate	8 (10.63)	1.000	17 (17.32)	1.00
Gastrula_tailbud	13 (14.17)	1.000	21 (19.56)	1.00
Gastrula_trunk	2 (3.15)	1.000	4 (4.75)	1.00
Gastrula_ventralmesoderm	18 (18.11)	1.000	32 (23.19)	0.15
Segmentation_Kupffersvesicle	2 (4.33)	0.850	8 (12.29)	0.51
Segmentation_alarplatemidbrainregion	10 (16.14)	0.608	15 (22.91)	0.24
Segmentation_artery	7 (6.3)	1.000	13 (8.94)	0.43
Segmentation_axialvasculature	9 (5.12)	0.548	13 (8.1)	0.24
Segmentation_basalplatemidbrainregion	13 (13.78)	1.000	22 (19.28)	0.97
Segmentation_blood	8 (9.45)	1.000	13 (11.74)	1.00
Segmentation_brain	*106 (82.28)*	*0.029*	*147 (113.16)*	*0.00*
Segmentation_cardiovascularsystem	7 (4.72)	0.878	11 (5.87)	0.11

Segmentation_cerebellum	14 (12.99)	1.000	22 (18.72)	0.86
Segmentation_cranialganglion	36 (29.13)	0.698	*53 (37.44)*	*0.03*
Segmentation_diencephalon	*67 (49.6)*	*0.035*	*100 (69.29)*	*0.00*
Segmentation_dorsalaorta	5 (3.54)	1.000	8 (5.03)	0.50
Segmentation_epiphysis	*56 (40.16)*	*0.035*	*82 (54.76)*	*0.00*
Segmentation_fin	6 (8.27)	1.000	14 (11.74)	0.97
Segmentation_floorplate	8 (8.27)	1.000	14 (15.93)	1.00
Segmentation_forebrain	22 (26.77)	0.878	41 (38)	0.97
Segmentation_gut	19 (25.98)	0.673	38 (41.91)	0.97
Segmentation_hatchinggland	11 (11.42)	1.000	19 (17.88)	1.00
Segmentation_headmesenchyme	7 (6.3)	1.000	12 (10.62)	1.00
Segmentation_heartrudiment	10 (9.45)	1.000	17 (12.57)	0.50
Segmentation_hindbrain	67 (62.59)	1.000	*118 (87.73)*	*0.00*
Segmentation_hypochord	5 (4.33)	1.000	8 (7.54)	1.00
Segmentation_hypophysis	15 (11.02)	0.771	24 (17.6)	0.31
Segmentation_hypothalamus	14 (13.78)	1.000	35 (24.31)	0.09
Segmentation_immatureeye	24 (29.13)	0.853	33 (38.84)	0.65
Segmentation_innerear	4 (5.12)	1.000	6 (6.71)	1.00
Segmentation_intermediatecellmassofmesoderm	9 (7.09)	1.000	11 (7.82)	0.59
Segmentation_laterallinesystem	15 (11.42)	0.811	21 (15.65)	0.42
Segmentation_lateralplatemesoderm	0 (3.15)	0.548	1 (3.35)	0.58
Segmentation_medianfinfold	4 (3.54)	1.000	7 (8.1)	1.00
Segmentation_midbrain	43 (44.09)	1.000	68 (63.7)	0.97
Segmentation_midbrainhindbrainboundary	13 (14.96)	1.000	20 (17.6)	0.97
Segmentation_musculaturesystem	14 (13.78)	1.000	24 (16.49)	0.15
Segmentation_myotome	40 (42.91)	1.000	61 (57.28)	0.97
Segmentation_neuralcrest	12 (10.24)	1.000	21 (13.97)	0.15
Segmentation_neuralrod	9 (7.09)	1.000	14 (8.94)	0.24
Segmentation_neuraltube	33 (29.13)	1.000	49 (42.47)	0.59
Segmentation_neuron	37 (31.1)	0.773	50 (37.16)	0.09
Segmentation_olfactoryplacode	33 (27.56)	0.775	44 (44.15)	1.00
Segmentation_opticcup	7 (3.15)	0.232	8 (4.75)	0.41
Segmentation_opticvesicle	21 (16.53)	0.773	32 (26.54)	0.59
Segmentation_oticvesicle	32 (32.28)	1.000	51 (53.09)	1.00
Segmentation_pharyngealarch3-7	2 (3.15)	1.000	9 (7.54)	1.00
Segmentation_pharyngealarch3-7skeleton	20 (21.65)	1.000	30 (30.46)	1.00
Segmentation_pharyngealarch	25 (20.08)	0.773	46 (32.69)	0.06
Segmentation_pharynx	6 (9.45)	0.775	11 (17.04)	0.35
Segmentation_post-ventregion	4 (1.97)	0.738	4 (2.51)	0.86
Segmentation_posteriorcardinalvein	5 (3.54)	1.000	6 (5.03)	1.00
Segmentation_presumptivediencephalon	2 (3.15)	1.000	3 (5.59)	0.64
Segmentation_presumptiveneuralretina	1 (1.18)	1.000	2 (2.51)	1.00
Segmentation_primitivehearttube	8 (7.87)	1.000	11 (9.22)	0.97

Segmentation_pronephricduct	38 (36.22)	1.000	55 (56.72)	1.00
Segmentation_pronephros	8 (8.66)	1.000	14 (15.65)	1.00
Segmentation_rhombomere5	6 (3.15)	0.636	12 (5.59)	0.03
Segmentation_rhombomere	9 (10.24)	1.000	17 (13.41)	0.66
Segmentation_solidlensvesicle	31 (23.23)	0.548	46 (31.01)	0.03
Segmentation_somite	56 (55.11)	1.000	92 (76.28)	0.15
Segmentation_telencephalon	54 (49.6)	1.000	85 (68.73)	0.11
Segmentation_trigeminalganglion	12 (7.48)	0.548	*18 (10.06)*	*0.05*
Segmentation_trigeminalplacode	18 (15.35)	1.000	29 (19.28)	0.09
Segmentation_vein	9 (8.66)	1.000	18 (11.18)	0.12
Pharyngula_caudalfin	2 (2.36)	1.000	5 (4.47)	1.00
Pharyngula_epidermis	12 (12.99)	1.000	25 (25.98)	1.00
Pharyngula_eye	67 (59.05)	0.773	102 (82.98)	0.09
Pharyngula_heart	49 (37.79)	0.318	*72 (53.37)*	*0.03*
Pharyngula_hearttube	11 (12.2)	1.000	18 (15.93)	0.98
Pharyngula_intestinalbulb	7 (11.81)	0.674	16 (20.96)	0.59
Pharyngula_intestine	23 (20.86)	1.000	33 (31.57)	1.00
Pharyngula_lens	35 (27.56)	0.636	66 (40.51)	0.00
Pharyngula_liver	53 (50.78)	1.000	86 (86.06)	1.00
Pharyngula_neuromast	6 (8.66)	0.977	10 (14.25)	0.59
Pharyngula_olfactorybulb	17 (12.6)	0.738	23 (15.09)	0.11
Pharyngula_olfactoryepithelium	*17 (9.05)*	*0.035*	24 (15.09)	0.08
Pharyngula_optictectum	54 (50.78)	1.000	72 (65.66)	0.78
Pharyngula_pancreaticbud	7 (8.27)	1.000	9 (9.5)	1.00
Pharyngula_pectoralfin	28 (22.05)	0.704	51 (38.28)	0.10
Pharyngula_pectoralfinbud	7 (7.09)	1.000	20 (12.29)	0.09
Pharyngula_pectoralfinmusculature	25 (27.56)	1.000	43 (40.51)	1.00
Pharyngula_peripheralolfactoryorgan	35 (31.1)	1.000	46 (46.66)	1.00
Pharyngula_retina	91 (72.44)	0.090	*123 (95.28)*	*0.01*
Pharyngula_retinalganglioncelllayer	*57 (32.28)*	*0.000*	*73 (36.6)*	*0.00*
Pharyngula_retinalinnernuclearlayer	*36 (22.44)*	*0.022*	*44 (26.82)*	*0.00*
Pharyngula_retinalphotoreceptorlayer	13 (9.05)	0.713	15 (12.29)	0.86
Pharyngula_spinalcord	59 (50)	0.674	89 (73.48)	0.15
Pharyngula_tegmentum	34 (28.34)	0.773	56 (41.35)	0.07
Pharyngula_ventricularzone	15 (14.96)	1.000	21 (20.96)	1.00
Hatching_ciliarymarginalzone	5 (3.15)	0.952	8 (4.47)	0.30
Hatching_habenula	12 (9.05)	0.878	16 (9.78)	0.13
Hatching_pancreas	6 (5.91)	1.000	10 (9.22)	1.00
Hatching_swimbladder	1 (4.33)	0.636	5 (6.15)	1.00
Larval_kidney	22 (16.53)	0.674	28 (27.1)	1.00
Larval_ovary	16 (16.53)	1.000	29 (29.9)	1.00
Larval_testis	23 (21.26)	1.000	38 (32.13)	0.59

Supplemental Table 4: Differentially abundant GO terms for Ohnologs and single-copy genes from the TGD

Comparison[a]	GO Hierarchy	GO Term	# in Ohnolog set[b]	Ohnolog enrichment[c]	P-value[d]
Dr_Ohno_all/ Dr_Sing_all[e]	Molecular Function	channel regulator activity	4	9.21	1.32E-04
		carbohydrate phosphatase activity	2	7.89	2.85E-02
		phosphatase activity	115	1.67	3.98E-03
		cyclic nucleotide-gated ion channel activity	2	7.89	2.79E-02
		ligand-gated ion channel activity	78	2.8	3.95E-09
		ion channel activity	186	2.43	1.26E-14
		transmembrane transporter activity	468	1.82	7.59E-15
		transporter activity	555	1.73	1.33E-14
		voltage-gated calcium channel activity	8	7.56	1.49E-06
		cation channel activity	58	2.86	2.94E-07
		cation transmembrane transporter activity	139	1.65	1.96E-03
		voltage-gated ion channel activity	51	2.48	1.27E-04
		glutamate receptor activity	12	7.45	3.70E-09
		receptor activity	610	1.32	4.79E-04
		phosphatase inhibitor activity	4	5.26	2.19E-02
		antigen binding	13	4.05	7.02E-04
		binding	2653	1.18	1.14E-06
		gap junction channel activity	14	3.19	9.12E-03
		voltage-gated potassium channel activity	37	3.06	1.26E-05
		kinase inhibitor activity	14	2.63	4.10E-02
		kinase activity	347	1.42	8.31E-04
		lipid transporter activity	25	2.53	9.15E-03
		ATPase activity, coupled to transmembrane movement of substances	38	2.35	2.67E-03
		actin binding	72	2.34	2.06E-05
		cytoskeletal protein binding	194	1.56	1.21E-03
		protein binding	1496	1.3	1.62E-08
		calmodulin binding	51	2.32	4.23E-04
		adenylate cyclase activity	66	1.99	2.12E-03
		sequence-specific DNA binding RNA polymerase II transcription factor activity	137	1.81	1.36E-04
		sequence-specific DNA binding transcription factor activity	410	1.35	2.41E-03
		calcium ion binding	117	1.62	9.13E-03
		signal transducer activity	524	1.54	3.14E-08
		GTPase activity	220	1.41	1.33E-02
		protein kinase activity	227	1.38	2.09E-02
		receptor binding	422	1.27	1.94E-02
		Unclassified	5234	0.89	2.77E-08
		oxidoreductase activity	345	0.71	1.07E-02
		RNA binding	232	0.5	7.27E-05
		cysteine-type peptidase activity	71	0.48	4.94E-02
		isomerase activity	66	0.36	9.09E-03
		nuclease activity	100	0.29	9.51E-05
		helicase activity	48	0.27	9.10E-03
		single-stranded DNA binding	31	0.25	4.35E-02
		structural constituent of ribosome	89	0.18	3.35E-06
		nucleotidyltransferase activity	47	0.06	6.82E-05
		RNA methyltransferase activity	25	< 0.01	2.74E-03
		methyltransferase activity	72	0.22	1.68E-04
		hydro-lyase activity	28	< 0.01	1.20E-03
		aminoacyl-tRNA ligase activity	32	< 0.01	5.47E-04
		ligase activity	148	0.57	1.10E-02

		peroxidase activity	16	< 0.01	3.59E-02
		damaged DNA binding	21	< 0.01	1.02E-02
Dr_Ohno_POInT/ Dr_Sing_POInT[f]	Molecular Function	channel regulator activity	10	8.35	1.40E-02
		glutamate receptor activity	31	7.39	2.10E-06
		voltage-gated calcium channel activity	15	6.26	4.85E-03
		cation channel activity	49	3.27	1.98E-05
		ion channel activity	116	2.52	2.03E-08
		transmembrane transporter activity	215	1.75	2.50E-07
		transporter activity	244	1.7	1.87E-07
		cation transmembrane transporter activity	59	1.64	3.52E-02
		voltage-gated ion channel activity	27	2.25	3.10E-02
		translation elongation factor activity	11	4.59	3.03E-02
		RNA binding	21	0.41	1.43E-03
		binding	739	1.15	7.64E-03
		translation regulator activity	25	2.78	1.33E-02
		ATPase activity, coupled to transmembrane movement of substances	24	3.08	8.81E-03
		ligand-gated ion channel activity	63	2.84	7.31E-06
		adenylate cyclase activity	39	2.03	1.74E-02
		actin binding	38	1.98	2.35E-02
		protein binding	479	1.27	3.80E-04
		sequence-specific DNA binding RNA polymerase II transcription factor activity	59	1.76	1.78E-02
		sequence-specific DNA binding transcription factor activity	133	1.41	1.97E-02
		signal transducer activity	157	1.61	2.88E-04
		kinase activity	146	1.46	7.74E-03
		Unclassified	993	0.88	1.62E-04
		oxidoreductase activity	55	0.65	2.91E-02
		isomerase activity	4	0.23	1.38E-02
		structural constituent of ribosome	4	0.16	3.45E-04
		nuclease activity	2	0.1	6.77E-04
		RNA methyltransferase activity	0	< 0.01	1.93E-02
		methyltransferase activity	2	0.1	2.15E-04
		hydro-lyase activity	0	< 0.01	1.50E-02
		aminoacyl-tRNA ligase activity	0	< 0.01	1.87E-02
		peroxidase activity	0	< 0.01	2.79E-02
		antioxidant activity	0	< 0.01	6.03E-03
		single-stranded DNA binding	0	< 0.01	1.43E-02
		nucleotidyltransferase activity	0	< 0.01	1.40E-03
		DNA helicase activity	0	< 0.01	1.80E-02
		helicase activity	2	0.16	2.27E-02
POInT_RootLosses /POInT_DrLosses[g]	Molecular Function	No Significantly differentially abundant terms			
Dr_Ohno_all/ Dr_Sing_all[e]	Biological Process	gluconeogenesis	11	4.82	6.36E-03
		primary metabolic process	933	0.89	2.01E-03
		asymmetric protein localization	8	4.21	3.43E-02
		localization	500	1.13	4.48E-02
		blood circulation	16	4.21	2.62E-03
		system process	353	2	5.00E-21
		single-multicellular organism process	549	1.93	6.42E-31
		multicellular organismal process	550	1.93	6.30E-31
		complement activation	20	4.05	5.98E-04
		response to stimulus	699	1.32	2.42E-09
		muscle organ development	20	4.05	5.82E-04
		system development	163	2.06	1.27E-10
		developmental process	483	1.48	1.63E-11
		B cell mediated immunity	20	4.05	5.67E-04
		muscle contraction	79	3.71	8.17E-13
		cell recognition	24	3.51	4.85E-04
		cellular process	2002	1.14	4.72E-10
		defense response to bacterium	20	3.29	2.61E-03
		spermatogenesis	17	3.19	7.33E-03
		neuron-neuron synaptic transmission	54	2.9	8.22E-07
		synaptic transmission	180	1.93	5.20E-10
		cell-cell signaling	240	1.93	3.51E-13

		cell communication	840	1.47	6.24E-20
		response to biotic stimulus	23	2.88	2.70E-03
		synaptic vesicle exocytosis	31	2.4	3.13E-03
		neurotransmitter secretion	45	1.97	4.05E-03
		neurological system process	289	1.84	4.15E-14
		embryo development	50	2.16	5.88E-04
		behavior	27	2.15	1.35E-02
		nervous system development	130	2.12	3.71E-09
		negative regulation of apoptotic process	41	2	4.94E-03
		regulation of biological process	715	1.42	2.34E-14
		biological regulation	877	1.42	5.40E-18
		phagocytosis	48	1.97	2.83E-03
		endocytosis	107	1.62	6.53E-04
		protein phosphorylation	38	1.92	1.11E-02
		cyclic nucleotide metabolic process	50	1.85	4.53E-03
		nucleobase-containing compound metabolic process	498	0.82	2.59E-04
		G-protein coupled receptor signaling pathway	116	1.84	5.77E-06
		cell surface receptor signaling pathway	335	1.51	1.33E-08
		signal transduction	714	1.45	7.14E-16
		MAPK cascade	126	1.81	3.46E-06
		intracellular signal transduction	393	1.58	4.71E-12
		cellular component morphogenesis	157	1.78	3.48E-07
		anatomical structure morphogenesis	53	1.62	2.46E-02
		cellular component organization	445	1.14	4.60E-02
		cytokinesis	50	1.73	1.33E-02
		cell differentiation	185	1.7	2.48E-07
		mesoderm development	105	1.68	2.94E-04
		sensory perception	61	1.57	2.18E-02
		regulation of phosphate metabolic process	173	1.52	9.47E-05
		phosphate-containing compound metabolic process	462	1.27	6.57E-05
		regulation of catalytic activity	126	1.52	1.32E-03
		regulation of molecular function	152	1.53	3.01E-04
		response to endogenous stimulus	89	1.47	1.49E-02
		locomotion	94	1.43	2.05E-02
		anion transport	81	1.41	4.51E-02
		ion transport	128	1.43	5.61E-03
		homeostatic process	101	1.41	2.25E-02
		cytoskeleton organization	109	1.36	3.10E-02
		Unclassified	1403	0.83	3.01E-13
		response to stress	83	0.66	2.63E-03
		protein targeting	22	0.57	4.64E-02
		DNA replication	14	0.47	2.35E-02
		DNA metabolic process	26	0.34	2.77E-08
		purine nucleobase metabolic process	6	0.37	4.53E-02
		RNA catabolic process	5	0.35	4.67E-02
		RNA metabolic process	230	0.69	6.20E-07
		cell-matrix adhesion	6	0.3	1.10E-02
		regulation of sequence-specific DNA binding transcription factor activity	2	0.19	3.55E-02
		DNA repair	5	0.12	7.42E-09
		rRNA metabolic process	3	0.11	1.58E-06
		tRNA metabolic process	3	0.1	2.99E-07
		tRNA aminoacylation for protein translation	1	0.08	2.88E-03
		translation	11	0.22	6.60E-08
		cytokine production	0	< 0.01	4.66E-02
		mRNA polyadenylation	0	< 0.01	4.68E-02
		mRNA 3'-end processing	0	< 0.01	8.68E-03
		mRNA processing	33	0.6	2.34E-02
		pteridine-containing compound metabolic process	0	< 0.01	4.62E-02
		nitrogen compound metabolic process	482	0.84	2.35E-03
		mitochondrial translation	0	< 0.01	3.12E-02
		mitochondrion organization	8	0.34	7.45E-03
		fatty acid beta-oxidation	0	< 0.01	4.56E-02

Dr_Ohno_POInT/ Dr_Sing_POInT[f]	Biological Process	blood circulation	10	5.56	2.82E-02
		system process	254	2.27	7.43E-17
		single-multicellular organism process	383	2.15	5.67E-24
		multicellular organismal process	384	2.16	7.49E-24
		muscle contraction	58	5.1	4.18E-10
		synaptic vesicle exocytosis	23	3.84	1.44E-03
		neurotransmitter secretion	35	2.54	3.16E-03
		neurological system process	214	2.13	1.62E-12
		neuron-neuron synaptic transmission	47	3.41	6.88E-06
		synaptic transmission	130	2.09	1.98E-07
		cell-cell signaling	161	1.93	1.33E-07
		cell communication	564	1.55	1.34E-14
		cellular process	1323	1.16	1.51E-07
		sensory perception	54	2.65	5.74E-05
		nervous system development	98	2.27	1.21E-06
		system development	119	2.31	3.87E-08
		developmental process	329	1.46	1.71E-06
		G-protein coupled receptor signaling pathway	65	2.21	1.81E-04
		cell surface receptor signaling pathway	218	1.63	1.81E-06
		signal transduction	490	1.58	2.35E-13
		cellular calcium ion homeostasis	33	2.12	2.02E-02
		homeostatic process	68	1.77	5.85E-03
		biological regulation	531	1.45	1.73E-10
		cyclic nucleotide metabolic process	39	1.97	2.10E-02
		nucleobase-containing compound metabolic process	338	0.85	3.53E-02
		primary metabolic process	620	0.88	1.58E-02
		transmembrane receptor protein tyrosine kinase signaling pathway	40	1.91	2.46E-02
		cellular component morphogenesis	104	1.89	7.12E-05
		cell differentiation	132	1.88	5.44E-06
		locomotion	66	1.87	3.09E-03
		MAPK cascade	95	1.76	7.73E-04
		intracellular signal transduction	281	1.67	6.47E-09
		regulation of catalytic activity	85	1.73	2.19E-03
		regulation of molecular function	105	1.79	2.53E-04
		mesoderm development	70	1.65	1.75E-02
		endocytosis	66	1.62	2.32E-02
		cellular component movement	89	1.62	7.40E-03
		regulation of phosphate metabolic process	123	1.58	1.63E-03
		phosphate-containing compound metabolic process	335	1.29	1.50E-03
		regulation of biological process	422	1.45	6.26E-08
		regulation of nucleobase-containing compound metabolic process	94	1.47	2.64E-02
		Unclassified	761	0.84	7.28E-06
		response to stress	38	0.52	1.66E-03
		response to stimulus	411	1.38	2.66E-06
		mRNA processing	17	0.46	1.62E-02
		RNA metabolic process	146	0.66	5.56E-05
		DNA replication	7	0.34	2.50E-02
		DNA metabolic process	15	0.29	7.57E-06
		DNA repair	5	0.2	6.51E-04
		cell-matrix adhesion	3	0.18	6.40E-03
		RNA catabolic process	2	0.16	1.89E-02
		rRNA metabolic process	2	0.1	2.12E-04
		tRNA aminoacylation for protein translation	0	< 0.01	1.53E-02
		translation	8	0.22	1.36E-05
		protein metabolic process	194	0.78	1.29E-02
		tRNA metabolic process	0	< 0.01	3.36E-05
POInT_RootLosses /POInT_DrLosses[g]	Biological Process	synaptic transmission	26	0.27	3.74E-02
		cell-cell signaling	33	0.28	2.03E-02
Dr_Ohno_all/ Dr_Sing_all[e]	Cellular Compartment	plastid	4	10.52	4.55E-02
		organelle	687	0.8	1.44E-07
		basal part of cell	5	6.58	4.30E-02
		cell part	1106	0.89	2.59E-04

		postsynaptic membrane	24	4.51	2.12E-05
		membrane	694	1.42	9.64E-14
		synapse	63	2.55	1.42E-06
		axon	32	3.66	1.05E-05
		neuron projection	122	1.92	5.43E-07
		cell projection	147	1.59	5.02E-05
		immunoglobulin complex	20	3.51	7.23E-04
		macromolecular complex	406	0.88	4.42E-02
		presynaptic membrane	12	2.63	4.28E-02
		actin cytoskeleton	96	2.53	2.40E-09
		cytoskeleton	160	1.34	7.54E-03
		dendrite	53	2.02	4.93E-04
		plasma membrane	583	1.71	1.41E-23
		integral to membrane	342	1.46	1.20E-07
		extracellular space	128	1.42	4.43E-03
		extracellular region	166	1.25	4.27E-02
		nucleoplasm	55	0.7	3.56E-02
		nucleus	288	0.7	1.21E-07
		cytosol	81	0.64	5.10E-04
		cytoplasm	653	0.89	7.75E-03
		intracellular	1013	0.86	4.74E-06
		mitochondrion	52	0.57	5.38E-04
		chromosome	16	0.41	1.11E-03
		ribosome	8	0.21	2.79E-06
		ribonucleoprotein complex	26	0.26	5.49E-14
		nuclear envelope	4	0.21	1.06E-03
		nucleolus	3	0.1	2.39E-07
Dr_Ohno_POInT/ Dr_Sing_POInT[f]	Cellular Compartment	postsynaptic membrane	23	4.8	2.68E-04
		membrane	448	1.41	1.89E-07
		synapse	55	2.96	9.61E-06
		presynaptic membrane	10	4.17	3.65E-02
		axon	22	4.08	1.06E-03
		neuron projection	78	2.1	1.08E-04
		cell projection	91	1.65	2.59E-03
		cell part	745	0.91	2.68E-02
		dendrite	40	2.9	2.34E-04
		actin cytoskeleton	57	2.44	1.22E-04
		organelle	468	0.82	3.44E-04
		plasma membrane	368	1.68	5.47E-12
		integral to membrane	217	1.5	8.00E-05
		intracellular	690	0.88	3.23E-03
		nucleoplasm	41	0.63	3.76E-02
		nucleus	202	0.74	7.21E-04
		mitochondrial inner membrane	6	0.33	2.96E-02
		mitochondrion	42	0.64	3.80E-02
		ribosome	6	0.23	7.64E-04
		ribonucleoprotein complex	16	0.23	1.39E-09
		macromolecular complex	260	0.81	9.44E-03
		nucleolus	2	0.08	1.13E-05
POInT_RootLosses /POInT_DrLosses[g]	Cellular Compartment	neuron projection	14	0.21	4.59E-02

a: Ohnolog and single-copy gene sets used for comparisons (See *Methods*).

b: Number of genes in the ohnolog set with the specified Go term (see **e** and **f** for totals).

c: Enrichment of the specified term in the ohnolog set relative to the single-copy set (values less than 1.0 represent GO terms seen less frequently among the surviving ohnolog pairs).

d: FDR-corrected *P*-value for the test of the hypothesis of equal proportions of genes in the ohnolog and single-copy gene sets annotated with the term in question.

e: See *Methods* for set definitions: *Dr_Ohno_all* includes 3994 genes and *Dr_Sing_all* contains 11,097 genes.

f: See *Methods* for set definitions: *Dr_Ohno_POInT* includes 2520 genes and *Dr_Sing_POInT* contains 4329 genes.

g: See *Methods* for set definitions: *POInT_RootLosses* includes 1880 genes and *POInT_DrLosses* contains 245 genes.