

Long-Term Asymmetrical Acceleration of Protein Evolution after Gene Duplication

Oriol Pich i Roselló^{1,2,3} and Fyodor A. Kondrashov^{2,3,4,*}

¹Facultat de Medicina, Universitat de Barcelona (UB), Spain

²Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG), Barcelona, Spain

³Universitat Pompeu Fabra (UPF), Barcelona, Spain

⁴Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

*Corresponding author: E-mail: fyodor.kondrashov@crg.es.

Accepted: July 21, 2014

Abstract

Rapid divergence of gene copies after duplication is thought to determine the fate of the copies and evolution of novel protein functions. However, data on how long the gene copies continue to experience an elevated rate of evolution remain scarce. Standard theory of gene duplications based on some level of genetic redundancy of gene copies predicts that the period of accelerated evolution must end relatively quickly. Using a maximum-likelihood approach we estimate preduplication, initial postduplication, and recent postduplication rates of evolution that occurred in the mammalian lineage. We find that both gene copies experience a similar in magnitude acceleration in their rate of evolution. The copy located in the original genomic position typically returns to the preduplication rates of evolution in a short period of time. The burst of faster evolution of the copy that is located in a new genomic position typically lasts longer. Furthermore, the fast-evolving copies on average continue to evolve faster than the preduplication rates far longer than predicted by standard theory of gene duplications. We hypothesize that the prolonged elevated rates of evolution are determined by functional properties that were acquired during, or soon after, the gene duplication event.

Key words: gene duplication, evolution, selection.

Introduction

The study of gene duplications continues to be at the forefront of molecular evolution due to their likely role in the emergence of new functions (Bridges 1936; Ohno 1970; Conant and Wolfe 2008; Ponting 2008; Hahn 2009; Innan and Kondrashov 2010). New functions are thought to emerge in the process of sequence and regulatory divergence of the gene copies that result from a gene duplication event. Therefore, the efforts in theory and empirical observations of gene duplications remain focused on the rate of evolution after gene duplication. Models of gene duplication evolution suggest that either one (Ohno 1970; Hughes 1999) or both (Force et al. 1999) gene copies experience a brief period of accelerated evolution (Hahn 2009; Innan and Kondrashov 2010), an elevated ratio of nonsynonymous (dn) and synonymous (ds) rates, followed by a return to the preduplication levels (Walsh 1995; Stoltzfus 1999; Lynch and Conery 2003; Konrad et al. 2011; Proulx 2012). It is thought that within the initial period of evolution new functions are forged.

Several approaches have been used to detect the acceleration of evolution after a gene duplication. First, dn/ds ratios between duplicated pairs in the same genome were correlated with ds values, where ds is a proxy for time since duplication. The observation that for values of $ds > 0.1$ the dn/ds appear to be relatively constant was taken as evidence that the observed plateau must correspond to the preduplication levels (Lynch and Conery 2000; Vinogradov 2012). Second, the dn/ds ratios of duplicated and nonduplicated genes consistently show that recently duplicated genes evolve faster than that of nonduplicated genes (Kondrashov et al. 2002; Nembaware et al. 2002; Jordan et al. 2004; Yampolsky and Bouzinier 2014), although genes that duplicated a long time ago appear to be more conserved (Davis and Petrov 2004; Jordan et al. 2004). The observation that dn/ds ratios measured in paralogous comparisons were higher than when the dn/ds is measured between nonduplicated orthologs suggested an acceleration of evolution in gene copies.

© The Author(s) 2014. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

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

These two approaches, however, cannot be used to determine at which time point after the gene duplication event the gene duplications return to their preduplicated rates of evolution. The first approach suffers from a lack of a comparison to other genes, as it is not clear whether or not the plateau of $ds > 0.1$ (Lynch and Conery 2000) corresponds to the preduplication levels. The second approach compares different genes and as some genes are more likely to be duplicated than others (Kondrashov et al. 2002; Jordan et al. 2004; Kondrashov 2012) it may create an inherent bias. Furthermore, under both approaches the dn/ds ratio is measured between two paralogous sequences. Therefore, the dn/ds estimated across the time since the gene duplication and is expected to be elevated for all comparisons due to the inclusion of the initial fast period of evolution right after the gene duplication (fig. 1A). Thus, neither of these two approaches makes it possible to reliably address the issue of the long-term impact of gene duplication on the rate of evolution.

To measure the long-term effects of gene duplication on the rate of evolution, it is necessary to estimate the preduplication rate of evolution and the recent rate of evolution of the paralogs. For very recent gene duplications measuring the rate of recent evolution is trivial as an estimate of dn/ds between them is sufficient. It is feasible to estimate recent dn/ds values from an older gene duplication by comparing orthologs of the gene copies of two species that separated after the emergence of the gene duplication in question (fig. 1B).

The preduplication rate of evolution can be estimated as the dn/ds observed on the internal branch leading up to the gene duplication event (fig. 1B). This approach has been utilized by Pegueroles et al. (2013) and Cusack and Wolfe (2007). Briefly, Pegueroles et al. (2013) identified instances of gene duplication that were broadly characterized by a phylogenetic relationship in figure 1B. They studied whether or not the rates of duplicated genes measured as dn/ds between the mouse and rat orthologs of the duplicated genes were significantly different from the rate of evolution on the internal branch leading up to the gene duplication. They found that the younger gene duplications indeed evolve faster than the preduplication rates. However, the older gene duplications, those that occurred between approximately 70 and approximately 43 Ma, were found to evolve at a rate indistinguishable from that in the preduplication branch. Cusack and Wolfe (2007) utilized a congruent approach to study the long-term impact on the rate of evolution stemming from the whole-genome duplication (WGD) in yeast. The rate of preduplication evolution was similarly inferred from an internal branch leading up to the duplication event.

These two studies presented with slightly contradictory results. Pegueroles et al. (2013) suggested that gene duplications return to their preduplication rates of evolution relatively quickly, on the order of old world and new world primate divergence (Chatterjee et al. 2009). In contrast, Cusack and Wolfe (2007) observed that in yeast the surviving duplications

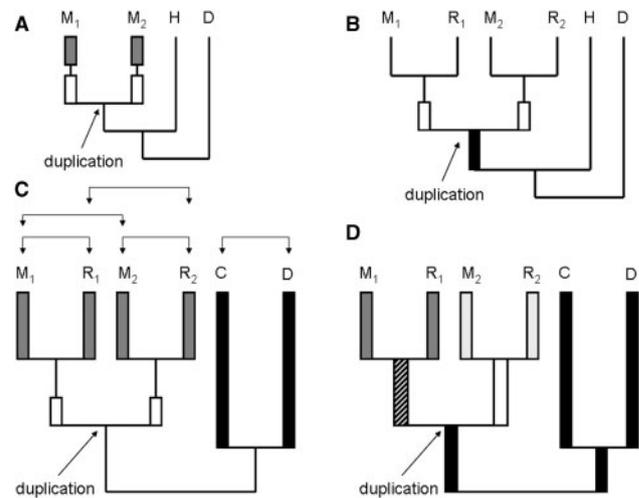


Fig. 1.—Phylogenetic topology of different approaches that may be used to study the evolution of gene duplications. When considering paralogs in only one species it is not possible to distinguish between recent evolution (gray box on the phylogeny) and the initial evolutionary acceleration (A). This issue can be resolved when considering dn/ds between orthologs of two species that have split after the gene duplication (B). However, if the preduplication rate of evolution is determined from the internal branch (black box) it may be difficult to resolve given the proximity to the expected acceleration of evolution immediately after the gene duplication (white boxes). In this study we avoid both of the compounding issues by focusing on five pairwise sequence comparisons, shown with brackets in (C). We can distinguish recent evolution of the duplicated genes (M_1-R_1 and M_2-R_2 , gray boxes) from the initial acceleration occurring immediately after the gene duplication event (white boxes). Similarly, we use the divergence between sequences from two species (C, D) that have diverged prior to the gene duplication event as a proxy for preduplication rates of evolution (black boxes). We date the gene duplication by estimating ds between the two paralogs in the two species (M_1-M_2 and R_1-R_2). (D) We estimated dn and ds on the phylogeny while restricting the model to estimating five different dn/ds values across the following branch segments indicated as follows: Black = preD, striped = ipostD₁, white = ipostD₂, dark gray = rpostD₁, light gray = rpostD₂.

from the WGD event continue to evolve faster than those genes that have lost their extra copy. Unfortunately, the Cusack and Wolfe (2007) study was focused on the remnants of the WGD event and used very long distances between some of the genes in the phylogeny, with $ds > 1$, reducing the reliability of the dn/ds estimates. An earlier study suggested that dn/ds on internal branches just prior to duplication may be accelerated in vertebrates (Johnston et al. 2007).

Here, we utilize two different methods to study the persistence of the acceleration of the rate of evolution after gene duplication. First, we employ a similar method to that used by previous studies (Cusack and Wolfe 2007; Johnston et al. 2007; Pegueroles et al. 2013) estimating the rate of evolution across different segments of the phylogeny that includes a recent gene duplication. Second, we investigate the rates of

evolution after gene duplications using data from gene duplications with tangible synonymous divergence distance and avoiding estimating the rate of evolution on internal phylogenetic branches.

Materials and Methods

We obtained protein-coding sequences for the mouse, rat, human, orangutan, dog, and cow from ENSEMBL (Vilella et al. 2009). We used the data in ENSEMBL to identify orthology relationships in these genomes. We selected the one-to-one dog–cow orthologs and for these orthologous pairs we identified instances of one-to-many orthology between dog and mouse (human) genes. For such mouse (human) genes, we selected those that showed a one-to-one orthology to the rat (orangutan) genes. This approach allowed us to identify a preliminary set of genes that were duplicated after the dog–human split but before the mouse–rat (human–orangutan) divergence, with the exception of those cases where a deletion of one copy occurred in the dog–cow lineage. Furthermore, these homologs included all cases when more than one duplication occurred along the aforementioned phylogenetic segment.

For all of these cases, we aligned all of the homologs using the protein sequence with MUSCLE (Edgar 2004) and reverse-translated to create a multiple nucleotide alignment. We then used the codeml package (Yang 2007) to estimate dn , ds , and dn/ds values across the phylogeny specifying five different areas of the phylogeny with independent dn/ds values with model = 2, clock = 0 and a user-defined tree that as shown in figure 1D. We also used the codeml program from the PAML package (Yang 2007) to estimate dn and ds between some pairwise comparisons of sequences from the sextuplet alignment.

For those homologous clusters that included more than one paralog in the mouse (human) genome, we estimated the order in which the duplications occurred. We obtained ds measurements for all pairwise mouse (human) paralogs. We then identified the pair with the smallest ds value as being the most recent gene duplication and created a set $((\text{Mouse}_1\text{--Rat}_1), (\text{Mouse}_2\text{--Rat}_2), (\text{Cow}, \text{Dog}))$ for that pair. We then removed at random one of the two paralogs and its rat (orangutan) ortholog and found the pair with the lowest ds among the remaining paralogs and the same set of six homologs was created. The procedure was then repeated recursively until only a single pair of mouse (human) paralogs was left. In the final data set, 25 out of 90 sets of six homologs originated from clusters representing more than one duplication between the dog–human common ancestor and the mouse–rat (human–orangutan) divergence.

To eliminate instances when the one-to-many orthology between dog and mouse (human) was caused by a gene loss in the dog–cow lineage, we eliminated all instances when the absolute ds values among the paralogous

comparisons were higher than expected to have originated after the dog–human split. Specifically, we eliminated all instances when either of the pairwise comparisons shown in figure 1 had a $ds > 1$. This approach must have also concurrently eliminated most instances of misalignment or erroneous ortholog assignment. We estimated δ by subtracting dn/ds of the dog–cow orthologs from the dn/ds value of each of the two mouse–rat (human–orangutan) orthologous comparisons.

Synteny analysis was performed by corresponding in ENSEMBL the neighboring genes of each dog gene to the neighboring genes of the paralogs in the mouse (human) genome. A mouse (human) homolog was identified as the original gene copy if its gene neighbors were orthologs to the gene neighbors of the dog gene. We excluded tandem duplications from this analysis.

Results and Discussion

We used mammalian genomes due to the availability of sequence data and an established phylogeny among major groups. We searched for gene duplications that have occurred after the split of Laurasiatheria (the group includes cow and dog) and Euarchontoglires (human and mouse) lineages but before the divergence of mouse–rat or human–orangutan lineages. Thus, we identified genes that were found in single copy in dog and cow, but were duplicated in mouse (human) and rat (orangutan) genomes (see Materials and Methods). Overall, we found 90 cases of such sextuplet homologs, with a gene duplication common to the mouse–rat (human–orangutan) lineage and single copy orthologs present in both the dog and the cow. We use a shorthand notation to label the cow (C), dog (D), mouse copy 1 (M_1), mouse copy 2 (M_2), rat copy 1 (R_1), rat copy 2 (R_2), human copy 1 (H_1), human copy 2 (H_2), orangutan copy 1 (O_1), and orangutan copy 2 (O_2) where $M_1\text{--}R_1$ ($H_1\text{--}O_1$) and $M_2\text{--}R_2$ ($H_2\text{--}O_2$) are one-to-one orthologs.

Due to the possible prevalence of gene conversion we first analyzed whether or not different rates of synonymous divergence are observed between gene copies different species. We correlated the estimated ds values obtained from $M_1\text{--}M_2$ ($H_1\text{--}H_2$) and $R_1\text{--}R_2$ ($O_1\text{--}O_2$) pairwise paralogous comparisons. We observe a very good correlation of the ds in paralogs of different species (fig. 2), indicating that gene conversion was not a large factor in the evolution of these genes. Of course, we cannot exclude the possibility of frequent gene conversion that occurred in both mouse (human) and rat (orangutan) lineages for a specific gene duplication. However, we are certain that this cannot affect much of our data as in that case we would have expected to see more paralogous sequence comparisons with a low ds .

Next, we introduce, δ , a measure of the difference between the preduplication and the recent postduplication rates of evolution. We measured these rates of evolution in

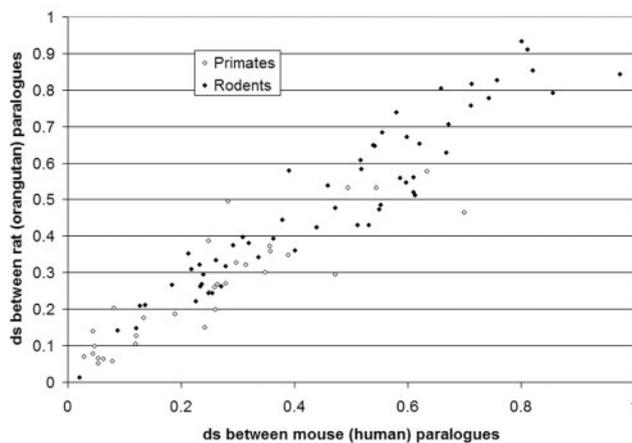


FIG. 2.—Relationship of synonymous divergence between gene duplications in sister species. The relationship between the synonymous divergence of paralogs in two species, M_1 – M_2 (H_1 – H_2) and R_1 – R_2 (O_1 – O_2) on the Y axis. The rodent and primate comparisons are shown with black and white circles, respectively.

two different ways. In a maximum-likelihood approach, we used PAML (see Materials and Methods) to estimate the preduplication (preD), two initial postduplication (ipostD₁ and ipostD₂), and two different recent postduplication (rpostD₁ and rpostD₂) rates of evolution or dn/ds (fig. 1D). We estimated the dn/ds separately in the two different copies (fig. 1) due to the possibility of asymmetrical evolution of the gene copies (Zhang et al. 2003; Cusack and Wolfe 2007; Han et al. 2009; Panchin et al. 2010; Pegueroles et al. 2013) and classify the gene copies into those with high and low dn/ds values, fast and slow rate evolution, respectively. We then measured $\delta_f = \max(\text{rpostD}_1, \text{rpostD}_2) - \text{preD}$ and $\delta_s = \min(\text{rpostD}_1, \text{rpostD}_2) - \text{preD}$, which represent a difference in the pre- and postduplication rate of evolution for fast- and slow-evolving gene copies, respectively.

We used the average ds summed across the branches separating the paralogous gene copies M_1 – M_2 and R_1 – R_2 (H_1 – H_2 and O_1 – O_2 ; see fig. 1D), as a proxy for the length of time since the origin of the gene duplication. This allowed us to access how long after the gene duplication event the rate of evolution (dn/ds) remains elevated. We found three patterns when we compared δ_f and δ_s in duplications of different age (fig. 3A). First, the slower-evolving gene copies appear to evolve at the preduplication rates regardless of how long ago the duplication event occurred. This observation is consistent with our selection of gene duplications that are older than the mouse–rat divergence such that we were expected to miss the initial phase of acceleration that may be affecting the slow copy. Second, the rate of evolution in the younger fast-evolving duplications appears to be higher than in older duplications. Finally, the rate of evolution in fast-evolving gene duplications appears to be elevated even for the oldest

duplications in our dataset with δ_f appearing to have reached a plateau at $\delta_f \sim 0.15$ for gene duplications older than $ds \sim 0.2$ (fig. 3A).

When considering perfectly symmetrical rates of evolution one copy is expected to evolve slower than the other due to the random variation of the dn/ds measurements. However, in a truly symmetrical case the increase in the rate of evolution in the fast-evolving copy will be of the same magnitude as the decrease in the rate of evolution of the slow copy. Our observation that $\delta_f > 0$ cannot be explained by the stochastic segregation of slow and fast gene copies because the observed increase of dn/ds in the fast copy is of much higher than the slight, insignificant decrease in the dn/ds of the slow copy.

The observation that the slow copy is evolving at the preduplication rate may be explained in two ways. First, the slow copy may have never experienced an acceleration in the rate of evolution. Second, the acceleration of evolution occurred but was short-lived such that it returns to normal within a modest timeframe ($ds < 0.2$). We therefore compared the rate of evolution of the initial postduplication branches leading up to the mouse–rat divergence (fig. 1D). We calculated δ_{if} and δ_{is} for the internal branch leading up to the slow copy and fast copy, respectively. Specifically, $\delta_{if} = \text{ipostD}_f - \text{preD}$, where ipostD_f is the dn/ds value from the branch leading up to the branch with δ_f . Conversely, $\delta_{is} = \text{ipostD}_s - \text{preD}$, where ipostD_s is the dn/ds value from the branch leading up to the branch with δ_s . We found that δ_{is} and δ_{if} are not significantly different even for gene duplications that have emerged recently (fig. 3B), indicating that the slow copy evolves faster immediately after gene duplication and subsequently returns to the preduplication levels.

It has been suggested that the old copy, the one that is not relocated in a gene duplication event, typically evolves slower than the new copy after the gene duplication (Cusack and Wolfe 2007; Han et al. 2009). We see the same pattern in our data. Due to the prevalence of tandem duplications in mammals it was often not possible to distinguish between the old and the new copies. For 26 cases, however, we used data on synteny between the rodent (primate) genome and that of dog and cow genomes to discriminate between the original and the new copy. For 21 out of 26 duplications, the fast-evolving copy was the novel copy (Fisher's exact test, $P = 0.02$). We hypothesize that the old copy is more likely to maintain the previous function and, therefore, more likely to have the preduplication levels of dn/ds .

Due to the decrease in selection pressure in primates caused by a smaller effective population size (Li 1997), it may be possible that the observed acceleration of the dn/ds in the duplicated genes is mostly influenced by the data from primate evolution. However, the fast-evolving gene copies are still significantly faster than the preduplication levels when we consider only rodent data, although the primate fast copies appear to have been accelerated to a greater degree (table 1). An increase in the strength of selection is anticipated in the

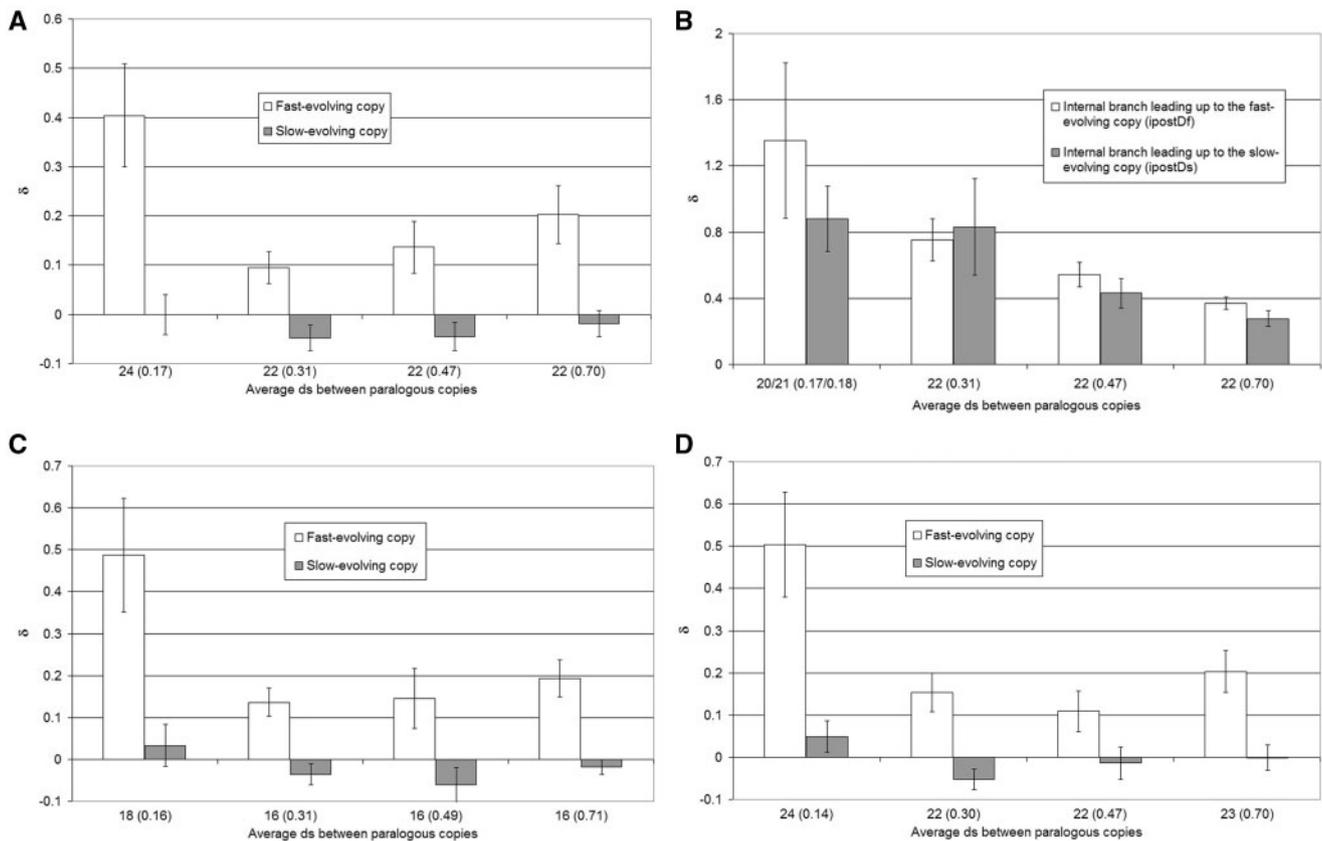


FIG. 3.—Difference in the rate of evolution in duplicated and nonduplicated orthologs. Average δ_f (white) and δ_s (gray) in bins. The number of duplications (average ds from the sum of branches connecting M_1 – M_2 [H_1 – H_2] and R_1 – R_2 [O_1 – O_2] nodes) is shown for each bin (A). The δ for internal branches (ipostD_f and ipostD_s) is shown in (B), branches of length 0 were removed from the analysis resulting in 20 (average ds =0.17) and 21 (average ds =0.18) duplications in one bin for ipostD_f and ipostD_s, respectively. The average δ_f (white) and δ_s (gray) excluding gene families are shown in (C). The average δ_f (white) and δ_s (gray) from pairwise comparisons are shown in (D), with average ds calculated similarly from pairwise comparisons of M_1 – M_2 (H_1 – H_2) and R_1 – R_2 (O_1 – O_2).

Table 1

Average dn/ds Values (Standard Error) for Slow and Fast-Evolving Gene Copies in Rodents and Primates

	Fast Copy	Slow Copy
Rodents	0.13 (± 0.027)	–0.026 (± 0.015)
Primates	0.46 (± 0.15)	0.038 (± 0.06)

rodent lineage relative to the cow–dog lineage (Li 1997), indicating that our observation of an increased δ_f is robust to lineage-specific changes in selection pressures.

In our data set, we include genes that have duplicated more than once since the divergence of rodents (primates) and the dog–cow common ancestor. We employ a scheme to separate such instances into individual duplication events (see Materials and Methods); however, it may be possible that some of the gene copies in our data set are influenced by more than one recently emerged copy. We therefore

performed the same analyses having retained only those genes that have experience only a single gene duplication event since the rodent (primate) and dog–cow divergence. We find the same patterns in the subset of genes with a single gene duplication (fig. 3C), indicating that multiple gene duplications do not substantially affect our results.

Finally, we have sought to replicate some of our results without attempting to reconstruct the rate of evolution in internal branches. We thus devised an approach that was based on five sequence comparisons among the orthologous and paralogous pairs of sequences estimating dn and ds (fig. 1C). We measured the rate of preduplication evolution $O = dn/ds$ observed in the C–D comparison. We then measured P_1 and P_2 as dn/ds values in the M_1 – R_1 (H_1 – O_1) and M_2 – R_2 (H_2 – O_2) orthologous comparisons, respectively, which estimate the recent rate of postduplication evolution. We then measured $\delta_f = \max(P_1, P_2) - O$ and $\delta_s = \min(P_1, P_2) - O$, which represent a difference in the pre- and postduplication rate of evolution for fast- and slow-evolving gene copies,

respectively. The average ds of paralogous comparisons, M_1-M_2 and R_1-R_2 (H_1-H_2 and O_1-O_2) was used as a proxy for the length of time since the origin of the gene duplication (fig. 1C).

The data obtained through pairwise sequence comparisons confirm our conclusions that 1) the slower-evolving gene copies appear to evolve at the preduplication rates regardless of how long ago the duplication occurred, 2) the rate of evolution is higher in younger fast-evolving duplications, and 3) the rate of evolution in fast-evolving gene duplications appears to be elevated even for the oldest duplications (fig. 3D). Due to the limitations of the pairwise comparisons using this approach we cannot confirm the observation of the initial acceleration in the rate of evolution of the slow-evolving gene copies.

Conclusions

Our results are consistent with the observations that asymmetrical evolution in gene duplication is common (Zhang et al. 2003; Cusack and Wolfe 2007; Han et al. 2009; Panchin et al. 2010; Pegueroles et al. 2013; Yampolsky and Bouzinier 2014) and with persistent, long-term faster evolution after a gene duplication in yeast (Scannell and Wolfe 2008). The observed persistence in the acceleration of evolution is surprising as it is not predicted by the more widespread theories of gene duplication. Most models of gene duplications that are based on a certain level of redundancy of gene duplications predict that the effect of the gene duplication event will be eliminated by accumulating mutations relatively quickly (Walsh 1995; Stoltzfus 1999; Lynch and Conery 2003; Konrad et al. 2011; Proulx 2012). Our data suggest that one of the gene copies seems to acquire some property other than redundancy, such as a new function (Notebaart et al. 2005; Assis and Bachtrog 2013), either at the duplication event or later, that results in a persistent faster rate of evolution than the ancestral single copy gene, or its ancestrally positioned paralog.

The following view on gene duplication evolution emerges from our data. First, immediately after gene duplication both gene copies experience an initial acceleration in their rate of evolution. Subsequently, one of the gene copies, which is typically the copy in the original genomic location, quickly returns to the preduplication levels in the timeframe of $ds \sim 0.15$. The second copy in the novel location eventually ($ds > 0.2$) achieves a stationary evolutionary rate. However, it does not appear to reach the preduplication levels of evolution even after a substantial period of time ($ds > 0.6$). Furthermore, the novel copy takes longer to reach the new stationary rate of evolution than is necessary for the old copy to return to the preduplication rates. Unfortunately, due to the scarcity of our data we cannot address the issue of deviation of specific gene copies from the described course of evolution. We believe that the functional characterization of individual cases may reveal

the properties of the new gene copies that result in their persistent accelerated evolution.

Acknowledgments

The work was supported by a grant of the HHMI International Early Career Scientist Program (55007424), the Spanish Ministry of Economy and Competitiveness (EUI-EURYIP-2011-4320) as part of the EMBO YIP program, two grants from the Spanish Ministry of Economy and Competitiveness, "Centro de Excelencia Severo Ochoa 2013–2017 (Sev-2012-0208)" and (BFU2012-31329), the European Union and the European Research Council under grant agreement 335980_EinME.

Literature Cited

- Assis R, Bachtrog D. 2013. Neofunctionalization of young duplicate genes in *Drosophila*. *Proc Natl Acad Sci U S A*. 110:43.
- Bridges CB. 1936. The bar "gene"—a duplication. *Science* 83:210–211.
- Chatterjee HJ, Ho SY, Barnes I, Groves C. 2009. Estimating the phylogeny and divergence times of primates using a supermatrix approach. *BMC Evol Biol*. 9:259.
- Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet*. 9:938–950.
- Cusack BP, Wolfe KH. 2007. Not born equal: increased rate asymmetry in relocated and retrotransposed rodent gene duplicates. *Mol Biol Evol*. 24:679–686.
- Davis JC, Petrov DA. 2004. Preferential duplication of conserved proteins in eukaryotic genomes. *PLoS Biol*. 2:E55.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32:1792–1797.
- Force A, et al. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.
- Hahn MW. 2009. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J Hered* 100:605–617.
- Han MV, Demuth JP, McGrath CL, Casola C, Hahn MW. 2009. Adaptive evolution of young gene duplicates in mammals. *Genome Res*. 19:859–867.
- Hughes AL. 1999. *Adaptive evolution of genes and genomes*. New York: Oxford University Press.
- Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet*. 11:97–108.
- Johnston CR, O'Dushlaine C, Fitzpatrick DA, Edwards RJ, Shields DC. 2007. Evaluation of whether accelerated protein evolution in chordates has occurred before, after, or simultaneously with gene duplication. *Mol Biol Evol*. 24:315–323.
- Jordan IK, Wolf YI, Koonin EV. 2004. Duplicated genes evolve slower than singletons despite the initial rate increase. *BMC Evol Biol*. 4:22.
- Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. 2002. Selection in the evolution of gene duplications. *Genome Biol*. 3. RESEARCH0008.
- Kondrashov FA. 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc Biol Sci*. 279:5048–5057.
- Konrad A, Teufel AI, Grahnen JA, Liberles DA. 2011. Toward a general model for the evolutionary dynamics of gene duplicates. *Genome Biol Evol*. 3:1197–1209.
- Li WH. 1997. *Molecular evolution*. Sunderland (MA): Sinauer.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
- Lynch M, Conery JS. 2003. The evolutionary demography of duplicate genes. *J Struct Funct Genomics*. 3:35–44.

- Nembaware V, Crum K, Kelso J, Seoighe C. 2002. Impact of the presence of paralogs on sequence divergence in a set of mouse-human orthologs. *Genome Res.* 12:1370–1376.
- Notebaart RA, Huynen MA, Teusink B, Siezen RJ, Snel B. 2005. Correlation between sequence conservation and the genomic context after gene duplication. *Nucleic Acids Res.* 33:6164–6171.
- Ohno S. 1970. *Evolution by gene duplication*. New York: Springer-Verlag.
- Panchin AY, Gelfand MS, Ramensky VE, Artamonova II. 2010. Asymmetric and non-uniform evolution of recently duplicated human genes. *Biol Direct.* 5:54.
- Pegueroles C, Laurie S, Albà MM. 2013. Accelerated evolution after gene duplication: a time-dependent process affecting just one copy. *Mol Biol Evol.* 30:1830–1842.
- Ponting CP. 2008. The functional repertoires of metazoan genomes. *Nat Rev Genet.* 9:689–698.
- Proulx SR. 2012. Multiple routes to subfunctionalization and gene duplicate specialization. *Genetics* 190:737–751.
- Scannell DR, Wolfe KH. 2008. A burst of protein sequence evolution and a prolonged period of asymmetric evolution follow gene duplication in yeast. *Genome Res.* 18:137–147.
- Stoltzfus A. 1999. On the possibility of constructive neutral evolution. *J Mol Evol.* 49:169–181.
- Vilella AJ, et al. 2009. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* 19:327–335.
- Vinogradov AE. 2012. Large scale of human duplicate genes divergence. *J Mol Evol.* 75:25–33.
- Walsh JB. 1995. How often do duplicated genes evolve new functions? *Genetics* 139:421–428.
- Yampolsky LY, Bouzinier MA. 2014. Faster evolving *Drosophila* paralogs lose expression rate and ubiquity and accumulate more non-synonymous SNPs. *Biol Direct.* 9:2.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
- Zhang P, Gu Z, Li WH. 2003. Different evolutionary patterns between young duplicate genes in the human genome. *Genome Biol.* 4:R56.

Associate editor: Eugene Koonin