Very Low Rate of Gene Conversion in the Yeast Genome

Claudio Casola,[1] Gavin C. Conant,[2,3] and Matthew W. Hahn*[,1,4]

[1]Department of Biology, Indiana University

[2]Division of Animal Sciences, University of Missouri

[3]Informatics Institute, University of Missouri

[4]School of Informatics and Computing, Indiana University

***Corresponding author:** E-mail: mwh@indiana.edu.

Associate editor: Helen Piontkivska

Abstract

Gene duplication is a major driver of organismal adaptation and evolution and plays an important role in multiple human diseases. Whole-genome analyses have shown similar and high rates of gene duplication across a variety of eukaryotic species. Most of these studies, however, did not address the possible impact of interlocus gene conversion (IGC) on the evolution of gene duplicates. Because IGC homogenizes pairs of duplicates, widespread conversion would cause gene duplication events that happened long ago to appear more recent, resulting in artificially high estimates of duplication rates. Although the majority of genome-wide studies (including in the budding yeast *Saccharomyces cerevisiae* [Scer]) point to levels of IGC between paralogs ranging from 2% to 18%, Gao and Innan (Gao LZ, Innan H. 2004. Very low gene duplication rate in the yeast genome. *Science* 306:1367–1370.) found that gene conversion in yeast affected >80% of paralog pairs. If conversion rates really are this high, it would imply that the rate of gene duplication in eukaryotes is much lower than previously reported. In this work, we apply four different methodologies—including one approach that closely mirrors Gao and Innan's method—to estimate the level of IGC in Scer. Our analyses point to a maximum conversion level of 13% between paralogs in this species, in close agreement with most estimates of IGC in eukaryotes. We also show that the exceedingly high levels of conversion found previously derive from application of an accurate method to an inappropriate data set. In conclusion, our work provides the most striking evidence to date supporting the reduced incidence of IGC among Scer paralogs and sets up a framework for future analyses in other eukaryotes.

Key words: gene duplication, *Saccharomyces cerevisiae*, interlocus gene conversion.

Introduction

Accurate estimates of gene duplication rates have major consequences for a number of important problems in biology, ranging from models of genome evolution to understanding the causes of human diseases, including cancer (Storchova and Pellman 2004), Alzheimer's disease (Rovelet-Lecrux et al. 2006), Parkinson's disease (Singleton et al. 2003), and Down syndrome (Korbel et al. 2009; Wiseman et al. 2009) among others (Bort et al. 1997; Girirajan et al. 2011). Duplications also play an important role in organismal adaptation (Conant and Wolfe 2008; Hahn 2009; Innan and Kondrashov 2010). Therefore, knowing the rate at which these mutations arise—and eventually fix—is key to understanding their role in evolution.

An early and very influential method for indirectly inferring duplication rates was introduced by Lynch and Conery (2000). In this article, the authors used the age distribution of paralogous pairs found within a genome to infer both the origination rate and loss rate of duplicates, based on a "demographic" model of duplicate gene life history. Lynch and Conery (2000) estimated the age distribution of paralogs using their synonymous divergence (e.g., the number of synonymous substitutions per synonymous site or d_S). Given the observed age distributions in a number of whole-genome sequences—especially the high numbers of very recent duplicates—Lynch and Conery inferred duplication rates of 0.0023–0.0208 per gene per millions of years (My).

Lynch and Conery (2000) noted two caveats in using d_S as a measure of paralog age: both a high rate of interlocus gene conversion (IGC) between paralogs and a high variance in d_S between genes can introduce biases in the age distribution. Interlocus gene conversion is the one-way transfer of genetic material between loci, with the donor locus sequence completely overwriting the acceptor locus sequence (Arnheim et al. 1980; Miyata et al. 1980; Scherer and Davis 1980; Slightom et al. 1980). Such gene conversion will cause two paralogs to appear more similar (as assessed by d_S or similar statistics) than their actual chronological age, consequently increasing the inferred rate of duplication in analyses based on the number of recent duplicates.

Although multiple articles using updated approaches based on d_S confirmed the apparent high rate of duplication (Gu et al. 2002; Lynch and Conery 2003), an intriguing study claimed that the duplication rate inferred by such methods may have vastly overestimated the true rate (Gao and Innan 2004). This study used an alternative method for assessing the age of duplicates, the logic of which is relatively unimpeachable: the phylogenetic distribution of genes can be used as independent evidence for their duplication time.

For instance, if two closely related species both have a pair of paralogs, then it is more parsimonious to assume that there was a single duplication event in their ancestor rather than two independent duplication events. Consequently, if the level of divergence between paralogs (as measured by d_S) is much lower than the level of divergence implied by their phylogenetic distribution, IGC is likely to have occurred. The same logic was applied in some of the original work describing general patterns of "concerted evolution" (Zimmer et al. 1980). Of course, the same pattern would result from parallel duplications in multiple lineages, especially if there is rapid turnover of duplicated genes (Baltimore 1981; Nei and Rooney 2005). Gao and Innan (2004) applied this phylogenetic test to 68 pairs of genes in the *Saccharomyces cerevisiae* (Scer) genome and showed that 55 of them (81%) were more similar than expected based on their phylogenetic distribution, presumably because of IGC. This high rate of gene conversion suggested that estimates of the duplication rate based on d_S are highly inflated.

Gao and Innan's work has been not only very influential but also perplexing, because it does not seem to be in accord with other data on the rate of either gene duplication or IGC. For instance, after accounting for IGC, Gao and Innan inferred 0.00001–0.00006 gene duplication events per gene per My in yeast. This value is in stark contrast to independent estimates of gene duplication that do not rely on d_S values between paralogs: these studies range from 0.0012 in *Drosophila* (Hahn et al. 2007), to 0.0016 in mammals (Demuth et al. 2006), to 0.0020 in yeast (Hahn et al. 2005). Alternative methods for estimating the influence of gene conversion have also produced much different numbers, showing that IGC affects from 2% to 15% of paralogs among yeast species (Drouin 2002; Morris and Drouin 2011), 8–10% in plants (Xu et al. 2008; Wang et al. 2009), 2% in the nematode *Caenorhabditis elegans* (Semple and Wolfe 1999), 13–19% in mammals (Ezawa et al. 2006; Hsu et al. 2009; McGrath et al. 2009), and 7–14% in *Drosophila* (Casola et al. 2010).

Here, we repeat the analysis of Gao and Innan using an expanded set of genes from the yeast genome, finding that the effect of gene conversion is more than five times lower than originally reported. Our approach mirrors the methodology developed by Gao and Innan but relies on a much larger set of genes (475 vs. 68 pairs of gene duplicates). In addition, we use three complementary approaches to estimate the proportion of genes affected by IGC. The results of these four analyses strongly indicate that only a small proportion of Scer paralogs are affected by IGC and support previous studies showing high rates of gene duplication in yeast.

Materials and Methods

S. cerevisiae Ohnolog Pairs and Their Orthologs in *S. castellii* and *Kluyveromyces polysporus*

Our first analysis is similar to that of Gao and Innan (2004), in which we compare synonymous divergence (d_S) between pairs of orthologs that are approximately the same age as pairs of "ohnologs" generated by a whole-genome duplication (WGD) event (fig. 1). Ohnologs are simply paralogous genes that originate from WGD events (Wolfe 2000). This term has been increasingly used in studies concerned with post-WGD gene evolution, not only in *S. cerevisiae* but also in teleosts (Postlethwait 2007), humans (Makino and McLysaght 2010), and plants (Schnable et al. 2011).

We first retrieved *S. cerevisiae* ohnolog pairs, *Sc1* and *Sc2*, for which we could identify clear orthologs in the genome of *Kluyveromyces polysporus* (Kpol) and/or *S. castellii* (Scas), using the set of 551 *S. cerevisiae* ohnologs from the Yeast Gene Order Browser (Byrne and Wolfe 2005). To do so, we used our previously described tool (Conant and Wolfe 2008) that estimates the probability that a Scer gene is orthologous (and not paralogous) to a gene in Kpol and/or Scas. The analysis begins with an inferred pre-WGD gene order (Gordon et al. 2009). We then use a model of post-WGD duplicate loss to estimate the probability of all possible orthology assignments. Note that these inferences rest on gene order information: sequence data are not considered. We required that the probability of this orthology inference be greater than 0.9 for Scer–Scas comparisons and greater than 0.75 for Scer–Kpol comparisons (fewer pairs of orthologs are shared between Scer and Kpol). We also required that the Scer genes hit the Kpol or Scas ortholog in a GenomeHistory (Conant and Wagner 2002) search with a BLAST *E*-value threshold of 10^{-4}. GenomeHistory was also used to calculate d_S values between these Scer–Kpol and Scer–Scas orthologs. To limit possible biases due to saturation of synonymous sites, we only used d_S values that pass a saturation test (Hahn et al. 2004). After this correction, we identified 108 Scer–Kpol orthologs and 263 Scer–Scas orthologs.

We also retrieved a set of *S. cerevisiae* genes that have no ohnolog in any of the five post-WGD genomes. From the study of Conant and Wolfe (2008), we obtained 766 *S. cerevisiae* genes with only a single ortholog in each of *S. bayanus*, *Candida glabrata*, *S. castellii*, and *K. polysporus* and a probability of orthology inference between these five genes of

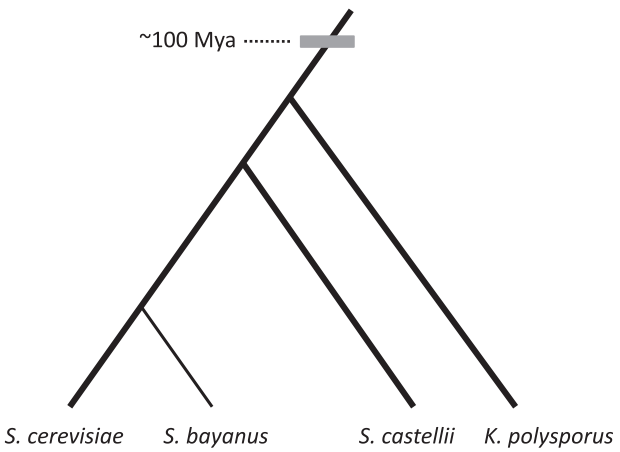

FIG. 1. Phylogenetic relationships of the species used in our analyses and timing of whole-genome duplication (WGD, gray bar; cf. Scannell et al. 2007). Lineages used in the comparison of ohnolog and ortholog d_S values are represented by thick lines; the lineage leading to *Saccharomyces bayanus* was only used in the approach relying on nonsynonymous divergence between ohnolog pairs.

greater than 0.9. We then filtered this data set by applying the same thresholds described earlier for GenomeHistory, as well as our saturation test, and found 267 Scer–Kpol single-copy orthologs and 288 Scer–Scas single-copy orthologs. Therefore, the total number of orthologs—with or without paralogs in Scer—between these species is 367 for Scer–Kpol and 551 for Scer–Scas.

From the same data set of *S. cerevisiae* pairs of ohnologs originally described by Byrne and Wolfe (2005), we also obtained a high-confidence set of 475 pairs of ohnologs formed by genes that share at least 30% amino acid identity and hit each other with a BLAST *E* value $\geq 10^{-4}$ (supplementary table S1, Supplementary Material online). These high-confidence pairs were used to infer gene conversion events. Pairs that did not pass our saturation test were kept in the data set to correctly assess the proportion of pairs with gene conversion among all pairs of ohnologs.

Sequence Data Sets and GENECONV Analysis

S. cerevisiae nucleotide coding sequences and protein sequences were retrieved from the fungal genomes research database (http://fungalgenomes.org/). To identify gene families, first we performed an all-against-all BLAST search on a data set of all protein sequences longer than 50 amino acids using default parameters, except a slightly more stringent expected value of 1×10^{-3}. The BLAST output was used to cluster *S. cerevisiae* proteins with the MCL program, version 09-308, under default settings except for the inflation value, $I = 6$ (Enright et al. 2002). Nucleotide alignments were obtained from the protein alignments using TransAlign (Bininda-Emonds 2005) implemented with MUSCLE (Edgar 2004). Alignments with fewer than three mismatches were removed, as well as alignments with regions of low identity according to a previously described method (Han et al. 2009). Gene conversion events were identified with the program GENECONV v.1.81 (Sawyer 1989), which employs permutation to determine whether possible gene conversion tracts (identical or nearly identical segments in the alignment) are statistically significant given the distribution of mismatches in the entire sequence alignment. When the alignment includes more than two sequences, GENECONV can assess both pairwise and global conversion tracts and calculates *P* values corrected for sequence length and, for global comparisons, also corrected for the number of aligned sequences. GENECONV analyses were performed with default settings except for the pairwise *P* values display option (–ListPair) and the "include monomorphic sites" option for alignments of only two sequences (–Include-monosites). Only gene conversion tracts with no mismatches and with $P < 0.05$ in the global or pairwise analysis were called significant.

Timing of Gene Duplication Events and Gene Conversion

Gene families affected by gene conversion tend to show phylogenies with more gene duplications toward the tips of their evolutionary tree because of the homogenization of paralogous genes' sequences (McGrath et al. 2009; Casola et al. 2010). We tested whether gene trees showed such gene conversion-driven bias by comparing the timing of duplications inferred through gene-tree/species-tree reconciliation using NOTUNG (Chen et al. 2000), which is sensitive to gene conversion bias, and through CAFE (Hahn et al. 2005), a method that employs only gene family sizes in current-day species to infer the timing of gene duplication and loss events across a species tree and is therefore immune to the effects of gene conversion in establishing the time of gene duplications. To assess gene duplications and losses with CAFE, we used gene families built with MCL as described earlier. Gene trees for the reconciliation analysis were obtained from data derived from the analysis of nine fungal genomes (Butler et al. 2009). We compared the gene duplication events according to the reconciliation method and the gene family evolution approach in 141 gene families that do not include ohnologs and have more than one gene in *S. cerevisiae* and in at least one other species.

S. cerevisiae Ohnolog Pairs and Their Orthologs in *S. bayanus* and Inference of Gene Conversion Using d_N

We applied a similar strategy to the one described earlier to identify ohnolog pairs, *Sc1* and *Sc2*, for which we could identify a clear ortholog to one of the two duplicates in the genome of *S. bayanus* based on synteny information (*Sb*). From this analysis, we retained triplets of three genes, *Sc1*, and *Sc2*, and the *S. bayanus* ortholog of *Sc1* (*Sb*). We required that the probability of this orthology inference be greater than 0.9 and that *Sc1* hit both *Sc2* and *Sb* in a GenomeHistory (Conant and Wagner 2002) search with a BLAST *E*-value threshold of 10^{-4}. We thus obtained 862 triplets of genes (*Sc1*, *Sc2*, and *Sb*), of which nine were removed due to the presence of premature stop codons in the coding sequences. Note that the same ohnolog pair could be analyzed twice if the corresponding genes in *S. bayanus* also survive as ohnologs. In this case, both sets of genes were tested for gene conversion and we report a potential event if either comparison suggests it.

To analyze sequence divergence in these triplets, we first aligned the protein sequences of each triplet using T-Coffee (Notredame et al. 2000) and inferred the corresponding nucleotide alignments. From those alignments we made maximum likelihood estimates (Conant and Wagner 2003) of the number of nonsynonymous substitutions per nonsynonymous site (d_N) for each of the three branches in the triplet. Define d_{N1} and d_{N2} as the d_N values for the branches leading to *Sc1* and *Sc2*, respectively, and d_{NB} as that leading to *Sb*. Because *Sc1* and *Sb* are more closely related than either is to *Sc2*, cases where $d_{NB} > d_{N1}, d_{N2}$ indicate potential gene conversion (see Evangelisti and Conant 2010 for details). To test the statistical support for inferences of conversion, we used a likelihood ratio test: we compared the likelihood of the sequence alignment under a model where all three d_N values were allowed to vary (lnLH0) to a constrained model where we required that $d_{N1} = d_{NB}$ (lnLHA). Statistical significance

was assessed by comparing twice the difference in log likelihood for the two models to a χ^2 distribution with one degree of freedom, as in Evangelisti and Conant (2010).

Results

We take four different approaches to inferring the amount of gene conversion between paralogs in yeast (fig. 2). In the first approach, we use the test outlined by Gao and Innan (2004), but we ask what proportion of all pairs of genes duplicated by a WGD show evidence for conversion. In the second, we compare reconciled gene trees estimated from the paralog sequences themselves to the maximum likelihood estimate of when duplication events took place among nine fungal genomes. In the third, we use the heterogeneity in divergence along the sequence of yeast paralogs to determine the fraction of gene pairs showing evidence for conversion. In the final approach, we use a likelihood-ratio test to compare the topology of trees generated using nonsynonymous divergence under models with and without IGC.

Comparison of Ohnolog and Ortholog d_S Values

Pairs of duplicates that experienced gene conversion are expected to have much lower levels of sequence divergence than orthologous single-copy genes of similar age; this is the basis of the test laid out by Gao and Innan. However, because parallel duplication events can obscure the true age of duplicates (i.e., by incorrectly implying that an ancestor had multiple gene copies), here we focus on paralogs generated by a WGD event in the ancestor of the yeasts. Because of the conserved syntenic relationships among these ohnologs, we can be highly confident that two duplicates in this data set originated at a specific point in the past (Wolfe and Shields 1997; Dietrich et al. 2004; Dujon et al. 2004; Kellis et al. 2004).

Both *K. polysporus* and *S. castellii* diverged from *S. cerevisiae* soon after the WGD (fig. 1) and represent the closest branching species to this event for which we have genome sequences (Scannell et al. 2007). Thus, we used the d_S values between *S. cerevisiae* genes and their orthologs in these two species to

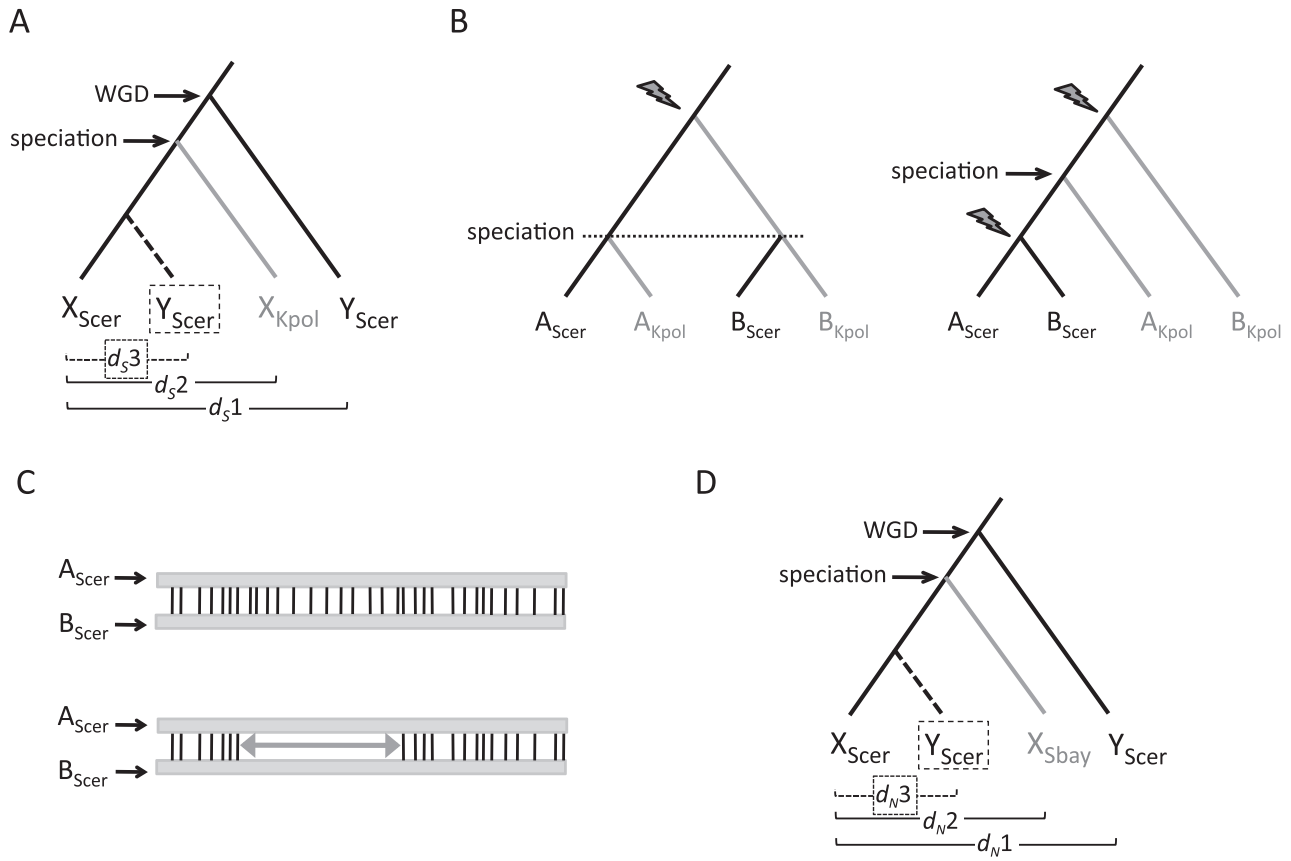

Fig. 2. The four strategies used to identify IGC events in *Saccharomyces cerevisiae* (Scer). (*A*) Comparison of ohnolog and ortholog d_S values. X and Y ohnologs in Scer originated after WGD (d_S1) and Scer and *Kluyveromyces polysporus* (Kpol) X orthologs diverged after speciation (d_S2). IGC between Scer ohnologs would lower the X–Y d_S value (d_S3) and result in a tree where the Y gene is moved closer to the X gene (dashed line). (*B*) Analysis of gene trees and gene duplication events. Left panel: gene tree/species tree reconciliation without IGC shows that a single gene duplication event (lightning bolt) formed the two paralogs A and B before Scer and Kpol diverged (this analysis excludes ohnologs). Right panel: phylogeny of the same four paralogs after IGC between Scer A and B. The reconciliation analysis performed on this tree would predict two gene duplication events. (*C*) GENECONV approach to detecting gene conversion. Upper panel: substitutions (black vertical bars) between Scer A and B paralogs (gray boxes) are uniformly distributed without IGC (this analysis includes both ohnolog and nonohnolog gene duplicates). Lower panel: regions of 100% identity between the two paralogs (gray double-headed arrows) are recognized using GENECONV as possible evidence of IGC. (*D*) Evidence for gene conversion among Scer ohnolog pairs using nonsynonymous divergence. This approach follows the strategy showed in (*A*) but relies on changes in d_N values in the presence of IGC between Scer X and Y ohnologs. In *A*, *B*, and *D*, the gray lines represent non-Scer genes.

estimate the expected sequence divergence between ohnologs in the absence of conversion. To detect an effect of IGC, we take the lower 95% confidence interval on d_S values between these orthologs as a minimum sequence divergence bound; ohnologs with d_S values below this bound are inferred to have been recently converted (cf. Gao and Innan 2004).

We found 375 Scer–Kpol orthologs and 551 Scer–Scas orthologs defined based on syntenic relationships between these genomes (see Materials and Methods). The average d_S values between orthologs were 4.77 for the Scer–Kpol comparison and 4.50 for the Scer–Scas comparison, with no significant difference in d_S between these two data sets ($P > 0.05$). However, we also found that orthologous genes with paralogs in the same genome showed significantly lower average d_S values than orthologs with no paralogs in both the Scer–Kpol and Scer–Scas comparison (4.12 vs. 5.04, $P = 7.3 \times 10^{-4}$; and 4.15 vs. 4.82, $P = 8.7 \times 10^{-5}$; respectively). Although gene conversion is not expected to decrease sequence divergence between orthologs—and in fact has been shown to sometimes increase divergence (Hurles et al. 2004)—to obtain the highest possible number of gene conversion events, we calculated the lower 95% confidence intervals using only orthologs without additional duplicates (the effect of which is to give a larger value of d_S as the lower bound). For the 267 Scer–Kpol orthologs and 288 Scer–Scas orthologs meeting these criteria, this corresponds to a cutoff of $d_S = 0.94$ and $d_S = 1.12$, respectively (fig. 3).

We were able to identify 475 high-confidence pairs of ohnologs in the *S. cerevisiae* genome (see Materials and Methods and supplementary table S1, Supplementary Material online). If none of these genes has experienced IGC—and assuming the rate of synonymous substitution is nearly the same between orthologs and ohnologs in the absence of IGC (see Discussion)—then only approximately 2.5% of ohnologs should have d_S values lower than these cutoffs. Instead, 13.7% (65/475) and 15.2% (72/475) of ohnolog pairs show d_S lower than 0.94 and 1.12, respectively (fig. 3). Given that we expected 2.5% of these to occur by chance, our results imply that the proportion of converted ohnolog pairs is maximally ~11.2–12.7%.

Analysis of Gene Trees and Gene Duplication Events

When paralogous genes undergo gene conversion, the gene duplication event through which they originated will appear to have occurred more recently than it actually did if this time is inferred from sequence divergence. This logic applies to all duplicates, whether or not they originate from WGD. To detect this possible hallmark in gene duplicates originating more recently than the ancestral WGD, we compared the timing of duplication events inferred using gene-tree/species-tree reconciliation with the timing inferred using a maximum-likelihood approach based on gene copy numbers that does not rely on sequence information and is therefore unaffected by gene conversion (Hahn et al. 2005). This general approach is similar in logic to that outlined in the previous section, but the timing of duplication is inferred from different data. Additionally, the duplicates used in this analysis may be of many different ages—necessitating separate comparisons of sequence-based inferences and copy-number-based inferences for each gene family—and parallel duplication events may be possible, which may increase the number of apparent conversion events.

We were able to identify 159 gene duplication events in *S. cerevisiae* that were not generated by the WGD event (i.e., they were not ohnologs). Of these, however, only 141 are informative for this analysis because they also have multiple copies in at least one of the eight other species in the Saccharomycetaceae group with sequenced genomes. For these 141 gene families, we constructed gene trees using all identified orthologs and paralogs from all nine

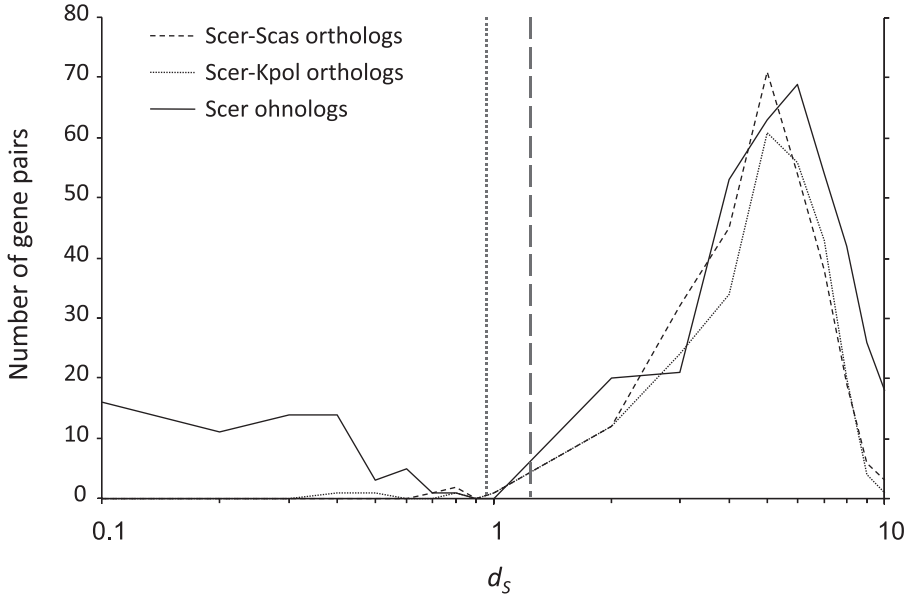

Fig. 3. Divergence at synonymous sites between *Saccharomyces cerevisiae* (Scer) ohnologs, Scer–*Kluyveromyces polysporus* (Kpol) orthologs, and Scer–*Saccharomyces castellii* (Scas) orthologs. The dashed vertical gray line and the dotted vertical gray line represent the lower boundary for a 95% confidence interval of Scer–Kpol and Scer–Scas ortholog d_S values, respectively.

yeast species and then carried out gene-tree/species-tree reconciliation using the program NOTUNG (Chen et al. 2000) to identify the timing of inferred gene duplication events. We also inferred the timing of duplication events using the program CAFE (De Bie et al. 2006), which minimizes the number of duplications and losses over the tree based on an analysis of gene family size (copy number) in each species (Hahn et al. 2005).

If IGC has affected pairs of duplicates, the inferred timing of duplication events in the tree reconciliation analysis will be more recent than in the copy number analysis. Thus, to find putative IGC events, we compared the timing of duplication in the two analyses to identify instances in which tree reconciliation implied a more recent duplication event than the copy number analysis (cf. McGrath et al. 2009; Casola et al. 2010). However, in none of the 141 analyzed gene families did we find this pattern. This implies that IGC has affected 0% of these gene trees.

GENECONV Approach to Detecting Gene Conversion

Even if IGC has not affected the value of d_S across a gene, or the gene tree inferred from whole-gene sequences, it can still leave a smaller signature on sequence divergence. In any gene conversion event, the conversion tract is the sequence that has been "copied and pasted" from a donor gene into the acceptor gene. Because the sequence corresponding to the conversion tract is consequently identical between the two paralogs, gene conversion events can be identified by looking for identical stretches of DNA embedded within more-diverged regions. Several statistical approaches have been developed to identify significant identical or highly similar DNA fragments between two or more sequences (Stephens 1985; DuBose et al. 1988; Sawyer 1989; Smith 1992; Sneath 1998; Worobey 2001). We applied the widely used method implemented in the program GENECONV (Sawyer 1989) to identify conversion tracts in alignments of *S. cerevisiae* paralogous genes.

To avoid possible biases due to poor alignment quality in large gene families, we analyzed both whole-family alignments that included all paralogs and pairwise alignments formed by pairs of paralogs. The whole-family alignments were searched using both the global and the pairwise options in GENECONV (see Materials and Methods). The level of gene conversion, measured as the proportion of gene pairs with evidence of conversion over all the examined gene pairs, was below 5% at $P < 0.05$ (table 1), implying that there is effectively no support for IGC. The whole-family data set consisted of 650 gene families, of which only 33 included ohnologs. Excluding gene families with ohnologs did not affect the proportion of gene pairs with gene conversion (data not shown). The pairwise alignment data set also showed a very low (<5% at $P < 0.05$) proportion of gene pairs with gene conversion, independently of the inclusion of ohnologs in the analysis (table 1).

We have previously shown that the power of GENECONV can be relatively low when paralogs have highly similar sequences ($d_S < 0.05$) or the conversion tracts are very short (McGrath et al. 2009). However, only 4% (234 of 5,812) of all *S. cerevisiae* pairs of paralogs analyzed here have $d_S < 0.05$ and only 5.6% of all pairs have $d_S < 0.1$. Thus, only a small proportion of *S. cerevisiae* paralogs could be converted without GENECONV detecting such events.

Evidence for Gene Conversion among *S. cerevisiae* Ohnolog Pairs Using Nonsynonymous Divergence

One further signature of IGC could be a subtle effect on only nonsynonymous substitutions (as measured by d_N). That is, if IGC is acting to maintain the functional similarity between a pair of paralogs, there may be a selective advantage to conversion events that homogenize any nonsynonymous differences that appear. We have previously found such a pattern of IGC among duplicated ribosomal proteins created by WGD in *S. cerevisiae*, using a comparison of d_N between ohnologs and orthologs (Evangelisti and Conant 2010). This approach is again conceptually similar to the one outlined earlier for all ohnologs but instead uses a likelihood-ratio test to compare tree topologies obtained using d_N with or without gene conversion for each triplet of genes (i.e., the pair of ohnologs and single ortholog). We have used this approach to look for gene conversion in the full set of ohnologs from *S. cerevisiae*.

We were able to find 862 gene triplets derived from 438 ohnolog pairs with high-confidence orthologs in *S. bayanus* identified on the basis of conserved gene order (see Materials and Methods and supplementary table S1, Supplementary Material online). Among the 438 ohnolog pairs analyzed, 35 (8%) had a signature of conversion using nonsynonymous divergence (d_N), but only 25 of these pairs (5.7%) showed statistically significant improvement when a model allowing gene conversion was used ($P < 0.05$; see Materials and Methods); once again, this is effectively the number of significant tests expected under the null model. Strikingly, of these 35 pairs, 26 encode ribosomal proteins and have already been identified as having undergone conversion (table 2; Evangelisti and Conant 2010). Moreover, of the remaining nine ohnolog pairs, five encode proteins associated with the ribosome (table 2). These genes are unlikely to be a random subset of all duplicates (see later).

Discussion

Our collective estimates of the number of paralogous pairs affected by interlocus gene conversion in *S. cerevisiae* (0–13%) are in stark contrast to those of Gao and Innan (81%), whereas they agree with other estimates of gene conversion in eukaryotes (Semple and Wolfe 1999; McGrath et al. 2009; Wang et al. 2009; Casola et al. 2010), including studies in *S. cerevisiae* (Drouin 2002; Morris and Drouin 2011). In addition, the very high level of gene conversion inferred by Gao and Innan seems unrealistic considering the implications of their results. For instance, as stated in the article by Gao and Innan, the gene duplication rate that would fit the estimated 81% of converted paralogs is approximately one duplication per gene per billion years, which is several orders of magnitude lower than gene duplication rates calculated using a variety of approaches in several model species (Hahn et al. 2005;

Table 1. GENECONV Estimates of Gene Conversion in *Saccharomyces cerevisiae* Gene Families.

	Whole-Family Alignments		Pairwise Alignments	
	Global Analysis	Pairwise Analysis	All Families	All Families No Ohnologs
Gene pairs with conversion	40/1,213	57/1,213	409/9,064	383/8,056
Percentage gene pairs with conversion	3.3	4.7	4.5	4.8

Table 2. Ohnolog Pairs with High-Confidence Orthologs in *Saccharomyces bayanus* Showing Signature of Conversion Using Nonsynonymous Divergence (d_N).

Sc1[a]	Annotation[b]	Sc2[c]	Annotation
EFT2	Elongation factor 2	EFT1	Elongation factor 2
ENO1	Enolase I	ENO2	Enolase II
HSC82	Hsp90-type chaperone	HSP82	Hsp90-type chaperone
IMD3	Inosine monophosphate dehydrogenase	IMD4	Inosine monophosphate dehydrogenase
MEP1	Ammonium permease	MEP3	Ammonium permease
RPL11B	Protein of the large ribosomal subunit	RPL11A	Protein of the large ribosomal subunit
RPL12B	Protein of the large ribosomal subunit	RPL12A	Protein of the large ribosomal subunit
RPL13B	Protein of the large ribosomal subunit	RPL13A	Protein of the large ribosomal subunit
RPL15A	Protein of the large ribosomal subunit	RPL15B	Protein of the large ribosomal subunit
RPL17B	Protein of the large ribosomal subunit	RPL17A	Protein of the large ribosomal subunit
RPL18A	Protein of the large ribosomal subunit	RPL18B	Protein of the large ribosomal subunit
RPL1B	Protein of the large ribosomal subunit	RPL1A	Protein of the large ribosomal subunit
RPL20A	Protein of the large ribosomal subunit	RPL20B	Protein of the large ribosomal subunit
RPL21A	Protein of the large ribosomal subunit	RPL21B	Protein of the large ribosomal subunit
RPL23A	Protein of the large ribosomal subunit	RPL23B	Protein of the large ribosomal subunit
RPL24A	Protein of the large ribosomal subunit	RPL24B	Protein of the large ribosomal subunit
RPL26B	Protein of the large ribosomal subunit	RPL26A	Protein of the large ribosomal subunit
RPL33B	Protein of the large ribosomal subunit	RPL33A	Protein of the large ribosomal subunit
RPL40A	Protein of the large ribosomal subunit	RPL40B	Protein of the large ribosomal subunit
RPL8A	Protein of the large ribosomal subunit	RPL8B	Protein of the large ribosomal subunit
RPS0A	Protein of the small ribosomal subunit	RPS0B	Protein of the small ribosomal subunit
RPS11B	Protein of the small ribosomal subunit	RPS11A	Protein of the small ribosomal subunit
RPS14A	Protein of the small ribosomal subunit	RPS14B	Protein of the small ribosomal subunit
RPS17B	Protein of the small ribosomal subunit	RPS17A	Protein of the small ribosomal subunit
RPS18A	Protein of the small ribosomal subunit	RPS18B	Protein of the small ribosomal subunit
RPS19B	Protein of the small ribosomal subunit	RPS19A	Protein of the small ribosomal subunit
RPS25A	Protein of the small ribosomal subunit	RPS25B	Protein of the small ribosomal subunit
RPS26B	Protein of the small ribosomal subunit	RPS26A	Protein of the small ribosomal subunit
RPS4B	Protein of the small ribosomal subunit	RPS4A	Protein of the small ribosomal subunit
RPS6B	Protein of the small ribosomal subunit	RPS6A	Protein of the small ribosomal subunit
RPS8A	Protein of the small ribosomal subunit	RPS8B	Protein of the small ribosomal subunit
SSB1	ATPase that is a ribosome-associated molecular chaperone	SSB2	ATPase that is a ribosome-associated molecular chaperone
SSF1	Constituent of 66 S preribosomal particles, required for ribosomal large subunit maturation	SSF2	Protein required for ribosomal large subunit maturation
TEF2	Translational elongation factor EF-1α	TEF1	Translational elongation factor EF-1α
TIF2	Translation initiation factor eIF4A	TIF1	Translation initiation factor eIF4A

[a]*Saccharomyces cerevisiae* ohnolog 1 (whose ortholog is the *Saccharomyces bayanus* gene *Sb*).
[b]SGD (Cherry et al. 1998) annotation of *Sc1*/*Sc2*.
[c]*Saccharomyces cerevisiae* ohnolog 2 (whose *paralog* is the *S. bayanus* gene *Sb*).

Demuth et al. 2006; Hahn et al. 2007). Furthermore, population studies have shown that copy number polymorphisms overlapping genes are common in *S. cerevisiae*, indicating that genes are duplicated and lost at a high rate in this species (Carreto et al. 2008).

Although our first approach closely followed the method used by Gao and Innan, we found fewer than 15% of ohnolog pairs with evidence for conversion using the same criteria, compared with the previous estimate of 81%. We, therefore, reasoned that this significant difference, rather than being

methodological, must stem from discrepancies between the data sets in this work and the original study. Indeed, we observed several biases in the data set used by Gao and Innan that were ultimately responsible for the exceedingly high level of gene conversion originally reported.

For example, although it is not mentioned in their original study, 50 of 68 pairs used by Gao and Innan are genes encoding ribosomal proteins; among the 55 pairs of genes identified as having undergone IGC by these authors, 47 were ribosomal protein pairs produced by WGD. Recently, we have found that at least 59% of *S. cerevisiae* ribosomal ohnologs showed signatures of gene conversion, compared with 3% of genes involved in metabolism (Evangelisti and Conant 2010). Given that gene conversion in *S. cerevisiae* can derive from an mRNA (or cDNA) donor (Derr and Strathern 1993; Storici et al. 2007), genes with high transcription levels—including ribosomal genes—could experience elevated conversion rates because of their large number of transcripts, as has been observed previously (Pyne et al. 2005; Sugino and Innan 2006). Therefore, the original data set of Gao and Innan did not represent a random assemblage of genes but rather one that was much more likely to show evidence for IGC.

In addition, as Lin et al. (2006) have shown, many pairs of *S. cerevisiae* ohnologs and their orthologs in other species show a very slow evolutionary rate of divergence due to coding-region conservation and codon-usage bias, which might be erroneously attributed to gene conversion. Lin et al. (2006) further point out that at least 57 pairs of genes analyzed by Gao and Innan show strong codon-usage bias. This result implies that the distribution of d_S values for all orthologs may not be an appropriate point of comparison for all ohnologs and may in fact lead to erroneous inferences of IGC.

Finally, for our main analysis we used 475 pairs of ohnologs so that we could analyze the effect of IGC on a "cohort" of genes of equivalent ages. In contrast, Gao and Innan generated their data set by using only pairs of paralogs with $d_S < 1.05$; this meant that their analysis was already strongly biased toward gene duplicates that were more likely to have been converted. In other words, the data set of Gao and Innan started by implicitly excluding many of the gene pairs that were not converted. To further demonstrate the effects of these inclusion criteria on their final results, we note that 49 of the 55 pairs of genes identified as having been converted by Gao and Innan are ohnologs (based on assignments in Byrne and Wolfe 2005). If—instead of comparing these 49 genes with all 68 pairs of genes with $d_S < 1.05$—Gao and Innan had compared them with the corresponding cohort of 475 pairs of currently existing ohnologs, they would have concluded that 10.3% of these paralogs show evidence for IGC, a result highly similar to ours.

The four methods we applied to detect IGC among *S. cerevisiae* paralogs have the advantage of relying on different evolutionary signals of this process. By using a data set of ohnolog pairs in our first and fourth methods, we controlled for independent, parallel gene duplications, which would have increased the rate of false positives in our analysis. This bias could affect estimates of gene conversion based on the comparison of duplication timing using gene trees and copy number data among species, although we did not find any evidence of conversion in the 141 *Saccharomyces* gene trees analyzed. On the other hand, genes duplicated by WGD may not be representative of IGC between all duplicates: because the total length of the duplicated sequence has a strong positive effect on the probability of pairing between paralogous regions and subsequent IGC (Hsu et al. 2009), ohnologs may experience more conversion than smaller-scale duplicates. The GENECONV approach provides estimates of gene conversion that are independent of both the inferred timing of gene duplication and the accurate identification of orthologs. Although GENECONV has limited power when paralog divergence is low, the gene conversion events missed by GENECONV will essentially involve donor and acceptor sequences that are identical or almost identical, with minimal, in any, functional consequences for the acceptor paralog (McGrath et al. 2009). Furthermore, GENECONV has been shown to be inaccurate only when there is rampant conversion (Mansai and Innan 2010), which is not observed in *S. cerevisiae*. Therefore, our results are completely consistent between methods, as would be expected with low levels of IGC.

A possible caveat with any analysis of gene conversion is the origin of paralogous genes. In fact, gene introgression from closely related species could erase evidence of gene conversion between paralogs by replacing converted genes with their orthologs. Because *S. cerevisiae* is known to undergo hybridization with congeneric species (Naumova et al. 2005; Nakao et al. 2009; Libkind et al. 2011), this phenomenon could lower the observed amount of gene conversion. However, we note that these reported hybridization events always involve *S. cerevisiae* as a donor of genomic material, not a recipient. Although we cannot exclude the possibility that some genes have been introduced into *S. cerevisiae* by ancient introgressions, there is no evidence that this process has played a major role in the evolution of the budding yeast genome.

Low levels of interparalog gene conversion have important evolutionary implications. A primary conclusion of our study is that the rate of gene duplication in *S. cerevisiae* is much higher than reported by Gao and Innan (2004) and is in fact much closer to the original estimate of Lynch and Conery (2000). Consequently, our results support models of rapid gene gain and loss as being primarily responsible for patterns of gene family evolution (Nei and Rooney 2005), rather than gene conversion. These results have major implications for, among other things, our understanding of evolution, adaptation, and the onset of disease.

Supplementary Material

Supplementary table S1 is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Mira Han for assistance with the tree reconciliation analysis and two anonymous reviewers for their

comments. This work was supported by a grant from the National Science Foundation (DBI-0845494) to M.W.H. and by Reproductive Biology Group of the Food for the 21st Century program at the University of Missouri (G.C.C.).

References

Arnheim N, Krystal M, Schmickel R, Wilson G, Ryder O, Zimmer E. 1980. Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and apes. *Proc Natl Acad Sci U S A.* 77:7323–7327.

Baltimore D. 1981. Gene conversion: some implications for immunoglobulin genes. *Cell* 24:592–594.

Bininda-Emonds OR. 2005. transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. *BMC Bioinformatics* 6:156.

Bort S, Martinez F, Palau F. 1997. Prevalence and parental origin of de novo 1.5-Mb duplication in Charcot-Marie-Tooth disease type 1A. *Am J Hum Genet.* 60:230–233.

Butler G, Rasmussen MD, Lin MF, et al. (51 co-authors). 2009. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature* 459:657–662.

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

Carreto L, Eiriz MF, Gomes AC, Pereira PM, Schuller D, Santos MA. 2008. Comparative genomics of wild type yeast strains unveils important genome diversity. *BMC Genomics* 9:524.

Casola C, Ganote CL, Hahn MW. 2010. Nonallelic gene conversion in the genus *Drosophila*. *Genetics* 185:95–103.

Chen K, Durand D, Farach-Colton M. 2000. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J Comput Biol.* 7:429–447.

Cherry JM, Adler C, Ball C, et al. (12 co-authors). 1998. SGD: *Saccharomyces* genome database. *Nucleic Acids Res.* 26:73–79.

Conant GC, Wagner A. 2002. GenomeHistory: a software tool and its application to fully sequenced genomes. *Nucleic Acids Res.* 30: 3378–3386.

Conant GC, Wagner A. 2003. Asymmetric sequence divergence of duplicate genes. *Genome Res.* 13:2052–2058.

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

De Bie T, Cristianini N, Demuth JP, Hahn MW. 2006. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22:1269–1271.

Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW. 2006. The evolution of mammalian gene families. *PLoS One* 1:e85.

Derr LK, Strathern JN. 1993. A role for reverse transcripts in gene conversion. *Nature* 361:170–173.

Dietrich FS, Voegeli S, Brachat S, et al. (14 co-authors). 2004. The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* 304:304–307.

Drouin G. 2002. Characterization of the gene conversions between the multigene family members of the yeast genome. *J Mol Evol.* 55: 14–23.

DuBose RF, Dykhuizen DE, Hartl DL. 1988. Genetic exchange among natural isolates of bacteria: recombination within the *phoA* gene of *Escherichia coli*. *Proc Natl Acad Sci U S A.* 85:7036–7040.

Dujon B, Sherman D, Fischer G, et al. (67 co-authors). 2004. Genome evolution in yeasts. *Nature* 430:35–44.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.

Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30: 1575–1584.

Evangelisti AM, Conant GC. 2010. Nonrandom survival of gene conversions among yeast ribosomal proteins duplicated through genome doubling. *Genome Biol Evol.* 2:826–834.

Ezawa K, Oota S, Saitou N. 2006. Proceedings of the SMBE Tri-National Young Investigators' Workshop 2005. Genome-wide search of gene conversions in duplicated genes of mouse and rat. *Mol Biol Evol.* 23: 927–940.

Gao LZ, Innan H. 2004. Very low gene duplication rate in the yeast genome. *Science* 306:1367–1370.

Girirajan S, Campbell CD, Eichler EE. 2011. Human copy number variation and complex genetic disease. *Annu Rev Genet.* 45: 203–226.

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 Genet.* 5: e1000485.

Gu Z, Cavalcanti A, Chen FC, Bouman P, Li WH. 2002. Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. *Mol Biol Evol.* 19:256–262.

Hahn MW. 2009. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J Hered.* 100:605–617.

Hahn MW, Conant GC, Wagner A. 2004. Molecular evolution in large genetic networks: does connectivity equal constraint? *J Mol Evol.* 58: 203–211.

Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N. 2005. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res.* 15:1153–1160.

Hahn MW, Han MV, Han SG. 2007. Gene family evolution across 12 *Drosophila* genomes. *PLoS Genet.* 3:e197.

Han MV, Demuth JP, McGrath CL, Casola C, Hahn MW. 2009. Adaptive evolution of young gene duplicates in mammals. *Genome Res.* 19: 859–867.

Hsu C-H, Zhang Y, Hardison R, Miller W. 2009. Whole-genome analysis of gene conversion events. *Lecture Notes Comput Sci.* 5817: 181–192.

Hurles ME, Willey D, Matthews L, Hussain SS. 2004. Origins of chromosomal rearrangement hotspots in the human genome: evidence from the *AZFa* deletion hotspots. *Genome Biol.* 5:R55.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet.* 11: 97–108.

Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428:617–624.

Korbel JO, Tirosh-Wagner T, Urban AE, et al. (33 co-authors). 2009. The genetic architecture of Down syndrome phenotypes revealed by high-resolution analysis of human segmental trisomies. *Proc Natl Acad Sci U S A.* 106:12031–12036.

Libkind D, Hittinger CT, Valerio E, Goncalves C, Dover J, Johnston M, Goncalves P, Sampaio JP. 2011. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc Natl Acad Sci U S A.* 108:14539–14544.

Lin YS, Byrnes JK, Hwang JK, Li WH. 2006. Codon-usage bias versus gene conversion in the evolution of yeast duplicate genes. *Proc Natl Acad Sci U S A.* 103:14412–14416.

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.

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.

Mansai SP, Innan H. 2010. The power of the methods for detecting interlocus gene conversion. *Genetics* 184:517–527.

McGrath CL, Casola C, Hahn MW. 2009. Minimal effect of ectopic gene conversion among recent duplicates in four mammalian genomes. *Genetics* 182:615–622.

Miyata T, Yasunaga T, Yamawaki-Kataoka Y, Obata M, Honjo T. 1980. Nucleotide sequence divergence of mouse immunoglobulin $\gamma 1$ and $\gamma 2b$ chain genes and the hypothesis of intervening sequence-mediated domain transfer. *Proc Natl Acad Sci U S A.* 77:2143–2147.

Morris RT, Drouin G. 2011. Ectopic gene conversions in the genome of ten hemiascomycete yeast species. *Int J Evol Biol.* 2011:970768.

Nakao Y, Kanamori T, Itoh T, Kodama Y, Rainieri S, Nakamura N, Shimonaga T, Hattori M, Ashikari T. 2009. Genome sequence of the lager brewing yeast, an interspecies hybrid. *DNA Res.* 16:115–129.

Naumova ES, Naumov GI, Masneuf-Pomarede I, Aigle M, Dubourdieu D. 2005. Molecular genetic study of introgression between *Saccharomyces bayanus* and *S. cerevisiae*. *Yeast* 22:1099–1115.

Nei M, Rooney AP. 2005. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet.* 39:121–152.

Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* 302: 205–217.

Postlethwait JH. 2007. The zebrafish genome in context: ohnologs gone missing. *J Exp Zool B Mol Dev Evol.* 308:563–577.

Pyne S, Skiena S, Futcher B. 2005. Copy correction and concerted evolution in the conservation of yeast genes. *Genetics* 170:1501–1513.

Rovelet-Lecrux A, Hannequin D, Raux G, et al. (13 co-authors). 2006. *APP* locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet.* 38:24–26.

Sawyer S. 1989. Statistical tests for detecting gene conversion. *Mol Biol Evol.* 6:526–538.

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. *Proc Natl Acad Sci U S A.* 104:8397–8402.

Scherer S, Davis RW. 1980. Recombination of dispersed repeated DNA sequences in yeast. *Science* 209:1380–1384.

Schnable JC, Springer NM, Freeling M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc Natl Acad Sci U S A.* 108:4069–4074.

Semple C, Wolfe KH. 1999. Gene duplication and gene conversion in the *Caenorhabditis elegans* genome. *J Mol Evol.* 48:555–564.

Singleton AB, Farrer M, Johnson J, et al. (22 co-authors). 2003. *α-Synuclein* locus triplication causes Parkinson's disease. *Science* 302:841.

Slightom JL, Blechl AE, Smithies O. 1980. Human fetal $G\gamma$- and $A\gamma$-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* 21:627–638.

Smith JM. 1992. Analyzing the mosaic structure of genes. *J Mol Evol.* 34: 126–129.

Sneath PH. 1998. The effect of evenly spaced constant sites on the distribution of the random division of a molecular sequence. *Bioinformatics* 14:608–616.

Stephens JC. 1985. Statistical methods of DNA sequence analysis: detection of intragenic recombination or gene conversion. *Mol Biol Evol.* 2: 539–556.

Storchova Z, Pellman D. 2004. From polyploidy to aneuploidy, genome instability and cancer. *Nat Rev Mol Cell Biol.* 5:45–54.

Storici F, Bebenek K, Kunkel TA, Gordenin DA, Resnick MA. 2007. RNA-templated DNA repair. *Nature* 447:338–341.

Sugino RP, Innan H. 2006. Selection for more of the same product as a force to enhance concerted evolution of duplicated genes. *Trends Genet.* 22:642–644.

Wang X, Tang H, Bowers JE, Paterson AH. 2009. Comparative inference of illegitimate recombination between rice and sorghum duplicated genes produced by polyploidization. *Genome Res.* 19:1026–1032.

Wiseman FK, Alford KA, Tybulewicz VL, Fisher EM. 2009. Down syndrome—recent progress and future prospects. *Hum Mol Genet.* 18: R75–R83.

Wolfe K. 2000. Robustness—it's not where you think it is. *Nat Genet.* 25: 3–4.

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

Worobey M. 2001. A novel approach to detecting and measuring recombination: new insights into evolution in viruses, bacteria, and mitochondria. *Mol Biol Evol.* 18:1425–1434.

Xu S, Clark T, Zheng H, Vang S, Li R, Wong GK, Wang J, Zheng X. 2008. Gene conversion in the rice genome. *BMC Genomics* 9:93.

Zimmer EA, Martin SL, Beverley SM, Kan YW, Wilson AC. 1980. Rapid duplication and loss of genes coding for the α chains of hemoglobin. *Proc Natl Acad Sci U S A.* 77:2158–2162.