Patterns of gene conversion in duplicated yeast histones suggest strong selection on a coadapted macromolecular complex

Kathy Scienski  
*University of Missouri Columbia*

Justin C. Fay  
*Washington University School of Medicine*

Gavin C. Conant  
*University of Missouri Columbia*

Follow this and additional works at: [https://digitalcommons.wustl.edu/open_access_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

**Recommended Citation**  
Scienski, Kathy; Fay, Justin C.; and Conant, Gavin C., "Patterns of gene conversion in duplicated yeast histones suggest strong selection on a coadapted macromolecular complex." *Genome Biology and Evolution*. 7,12. 3249-3258. (2015).  
[https://digitalcommons.wustl.edu/open_access_pubs/5337](https://digitalcommons.wustl.edu/open_access_pubs/5337)

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Patterns of Gene Conversion in Duplicated Yeast Histones Suggest Strong Selection on a Coadapted Macromolecular Complex

Kathy Scienski1,2, Justin C. Fay3,4,*, and Gavin C. Conant1,5,*

1Division of Animal Sciences, University of Missouri, Columbia
2Present address: Genetics Graduate Program, Texas A&M University, College Station, TX.
3Department of Genetics, Washington University
4Center for Genome Sciences and Systems Biology, Washington University
5Informatics Institute, University of Missouri, Columbia

*Corresponding author: E-mail: conantg@missouri.edu; jfay@genetics.wustl.edu.

Abstract

We find evidence for interlocus gene conversion in five duplicated histone genes from six yeast species. The sequences of these duplicated genes, surviving from the ancient genome duplication, show phylogenetic patterns inconsistent with the well-resolved orthology relationships inferred from a likelihood model of gene loss after the genome duplication. Instead, these paralogous genes are more closely related to each other than any is to its nearest ortholog. In addition to simulations supporting gene conversion, we also present evidence for elevated rates of radical amino acid substitutions along the branches implicated in the conversion events. As these patterns are similar to those seen in ribosomal proteins that have undergone gene conversion, we speculate that in cases where duplicated genes code for proteins that are a part of tightly interacting complexes, selection may favor the fixation of gene conversion events in order to maintain high protein identities between duplicated copies.

Key words: genome duplication, gene conversion, histones.

Introduction

Protein interactions underlie many cellular functions (Bork et al. 2004; Gavin et al. 2006; Krogan et al. 2006). Because these interactions appear to be well conserved (Qian et al. 2011), the evolution of protein complexes occurs in two dimensions, with the protein sequences of the interacting partners being under selection to maintain the interaction (Mintseris and Weng 2005; Codoner and Fares 2008; Liberles et al. 2012), while at the same time complexes can diversify or subfunctionalize through gene or genome duplications (Wagner 2001, 2003; Kunin et al. 2004; Li et al. 2006).

As it happens, there has been considerable discussion as to how the presence of one or more protein interactions alters the propensity for a gene to undergo duplication in the first place (He and Zhang 2006; Li et al. 2006; Pérez-Bercoff et al. 2010; Zhu et al. 2013). The evidence to date suggests that duplications of single genes are biased toward genes with fewer interactions than average, but a genome duplication will tend to preserve duplicates with larger numbers of interactions (Hakes et al. 2007; Pérez-Bercoff et al. 2010). This observation would be in accord with recent theory on the importance of maintaining the proper balance of genes and proteins in cells (Veitia et al. 2008). The “dosage balance hypothesis” postulates that genes whose functions involve precise interactions with other genes’ products will be under selection against large changes in dosage, hence the rarity of single gene duplications among such genes. Because whole genome duplications (WGDs) double all genes simultaneously, dosage balance would predict that highly interacting genes would instead be common among the surviving WGD duplicates, which is exactly the pattern observed (Papp et al. 2003; Edger and Pires 2009; Freeling 2009; Makino and McLysaght 2010; Birchler and Veitia 2012; Veitia et al. 2013).

Although a WGD transiently doubles the entire genome, many of the resulting duplicates are lost through “fractionation” (Sémon and Wolfe 2007; Woodhouse et al. 2010).
Dosage balance is a critical force in this process, with genes coding for proteins involved in processes such as transcription or complex formation being maintained in duplicate (Blanc and Wolfe 2004; Maere et al. 2005; Carretero-Paulet and Fares 2012) to maintain a balance in dosage. Indeed, we were recently able to show that not only were genes whose encoded proteins have many interactions overretained in duplicate after WGD, but that, when they were lost, it was more likely that interacting pairs would be lost together (Conant 2014).

Although surviving duplicates may have initially been preserved because the members of a complex or group of interacting genes were under selection to remain in dosage balance, it is not expected that such selection will be active indefinitely, nor that the functions of the retained duplicates will remain constant (Conant et al. 2014). Instead, the surviving duplicates have a number of potential fates, including retained redundancy (Keane et al. 2014), functional innovation (Boles et al. 1997; Özcan and Johnston 1999; Rodríguez et al. 2001), and subfunctionalization (van Hoof 2005; Hittinger and Carroll 2007). We have previously studied a group of duplicates with apparent redundancy. The yeast ribosomal proteins (Planta and Mager 1998; Kellis et al. 2004; Kim et al. 2009), which in many cases consist of duplicate pairs with identical amino acid sequences. We found that these duplicates, created by WGD, were undergoing recurring gene conversions (GCs; Evangelisti and Conant 2010). Indeed, these conversions were so numerous that they gave rise to the mistaken impression that conversion was rampant in yeast (Gao and Innan 2004). In fact, they are the exception: GC is otherwise rare in bakers’ yeast (Casola et al. 2012).

GC occurs when one piece of DNA is “overwritten” by a second (Arnheim et al. 1980; Miyata et al. 1980; Scherer and Davis 1980). Mechanisms may include DNA/DNA-based recombination between homologous sequences (Chen et al. 2007) in mitosis or meiosis or events involving an RNA intermediate (Derr and Strathern 1993; Storici et al. 2007). It may be observed between tandemly duplicated DNA or between homologous regions on different chromosomes, as was the case for the ribosomal proteins (Scherer and Davis 1980; Chen et al. 2007).

Here we present a second example of GC among WGD-produced duplicates, namely that of histone genes. Histones have been known to undergo conversion for some time (Maxson et al. 1983; Taylor et al. 1986; Matsuo and Yamazaki 1989; DeBry and Marzluff 1994; Wang, Krasikov, et al. 1996; Wang, Tisovec, et al. 1996; Baldo et al. 1999; Liao 1999; but see Piontkivska et al. 2002; Rooney et al. 2002). The bakers’ yeast *Saccharomyces cerevisiae* has duplicate copies of the four histones whose origins predate the yeast WGD (Byrne and Wolfe 2005). These genes are somewhat diverged in sequence (ranging from $K_s = 0.08$, $K_a \approx 0$ for the *HHF1/HHF2* pair to $K_s = 0.49$, $K_a = 0.03$ for *HTB1* and *HTB2*) and conversion among them was not initially suspected (Maxson et al. 1983). However, a more recent analysis suggested conversion at least between the *HHF12* pair (Kellis et al. 2004). The picture is more complicated in other yeast genomes. In addition to the old duplicates (where rejecting the null hypothesis of no GC is more challenging; Casola et al. 2012), these other species have histone duplicates produced by the yeast WGD itself, and, for some of those duplicates, we find clear evidence of conversion.

The existence of a second example of frequent conversion post-WGD potentially helps explain the evolutionary force underlying the fixation of these conversion events. Redundancy in duplicate genes is difficult to sustain by natural selection (Cooke et al. 1997; Nowak et al. 1997; Wagner 2000; Qian et al. 2010), but ribosomal proteins and histones may be exceptions to this rule due to the requirement for high expression of these types of genes (Kondrashov FA and Kondrashov AS 2006; Ihmels et al. 2007; Qian et al. 2010). Such selection on expression magnitude, in combination with selection to maintain dosage balance, would explain the survival of the WGD duplicates for these two classes of genes. GC, then, would have the secondary role of keeping the sequences of the duplicates similar enough that both copies function equally well in the ribosome and the nucleosome, both of which are tightly interacting and essential complexes.

**Methods**

**Data Collection**

Gene sequences from 8 histones (2 ancient duplicates each of histones 2A, 2B, 3, and 4) from 12 post-WGD yeasts were obtained from YGOB (Yeast Genome Order Browser Project; Byrne and Wolfe 2005). A histone pair was categorized as having undergone conversion if the two WGD-produced paralogs from a species had higher protein sequence identity to each other than either did to any homolog in its nearest relative. Histone genes *HTB2*, *HHF1*, and *HHF2* were not found to have conversion events and were not further analyzed. The remaining five genes (table 1) had duplicates with evidence of conversion in one or more yeasts.

**Orthology Inference Using Polyploidy Orthology Inference Tool to Establish Expected Gene Relationships**

Recall that we have here the special case of GCs after WGD. Thus, paralogs produced by WGD should be more distantly related to each other than to their orthologs in other genomes sharing the WGD. If two WGD-produced paralogs are found to be more closely related to each other than either is to its respective ortholog in another post-WGD yeast, that is evidence of GC. We used POInT (Polyploidy Orthology Inference Tool) to estimate, for each of the potentially converted histone genes, its ortholog in its nearest neighboring genome. If one genome was missing both copies of that histone, we removed that species from our orthology inferences.
Patterns of Histone Protein Sequence Identity and Gene Phylogenies Provide Evidence for Gene Conversion among Duplicated Histones

Table 1

| Saccharomyces cerevisiae gene* | Gene type | Gene IDs | Dist(D1,D2) | Min(Dist(D1,D2),Dist(C0,D2),Dist(C0,D1)) | lnL_{app} | lnL_{GC} | lnL_{phyML} |
|-------------------------------|-----------|----------|-------------|-------------------------------------|-----------|-----------|-------------|
| HTA1                          | D2       | TPHA0L1110 | 0.008       | 0.038                               | -1,586    | -1,534    | -1,524      |
|                              | D1       | TPHA0C02050 | 0.008       | 0.038                               | -1,586    | -1,534    | -1,524      |
|                              | O        | Kpol_1031.53 | 0.008       | 0.038                               | -1,586    | -1,534    | -1,524      |
| HTA2                          | D1       | KAFRO600780 | 0.0         | 0.015                               | -1,357    | -1,324    | -1,316      |
|                              | D2       | KAFRO602490 | 0.0         | 0.015                               | -1,357    | -1,324    | -1,316      |
|                              | O        | KNAGO601430 | 0.0         | 0.015                               | -1,357    | -1,324    | -1,316      |
| HTB2                          | D1       | KAFRO600770 | 0.030       | 0.091                               | -1,294    | -1,259    | -1,258      |
|                              | D2       | KAFRO602480 | 0.030       | 0.091                               | -1,294    | -1,259    | -1,258      |
|                              | O        | KNAGO601420 | 0.030       | 0.091                               | -1,294    | -1,259    | -1,258      |
| HHF1                          | D1       | CAGL0C04136g | 0.0         | 0.010                               | -1,251    | -1,174    | -1,146      |
|                              | D2       | CAGL0H09834g | 0.0         | 0.010                               | -1,251    | -1,174    | -1,146      |
|                              | O        | YBR009C (HHF1) | 0.0         | 0.010                               | -1,251    | -1,174    | -1,146      |
| HHF2                          | D1       | NDA0803480 | 0.0         | 0.010                               | -1,011    | -963      | -962        |
|                              | D2       | NDA0G00750 | 0.0         | 0.010                               | -1,011    | -963      | -962        |
|                              | O        | NCAS0806180 | 0.0         | 0.010                               | -1,011    | -963      | -962        |
|                              | D1       | NCAS0G03710 | 0.0         | 0.010                               | -1,011    | -963      | -962        |
|                              | D2       | NDA0803480 | 0.0         | 0.010                               | -1,011    | -963      | -962        |

*Saccharomyces cerevisiae histone gene name. Note that S. cerevisiae has no surviving histone duplicates from the WGD, making these names unambiguous.

Proportion of amino acid difference between the two paralogs (D1 and D2) created by WGD.

Minimum of the proportion of amino acid difference between one of the two orthologs (D1 or D2) and the nearest homolog in its nearest species relative (O).

In-likelihood of the full sequence alignment fit to the assumed species tree (ln\_phyML), the gene conversion tree (ln\_GC), or the phylogeny estimated by PhyML (ln\_app). See Methods for details.

Relationship between two paralogs hypothesized to have undergone gene conversion (D1 and D2) and an assumed ortholog of D1, O (see table 2 for precise orthology inferences).

**Table 1**

| Patterns of Histone Protein Sequence Identity and Gene Phylogenies Provide Evidence for Gene Conversion among Duplicated Histones |
|-------------------------------------------------------------------------------------------------------------------------------|

**Triplet Tests for Gene Conversion**

There are several signatures that can be used to infer GC events. For arbitrary sequences, the GENECONV program (Sawyer 1989) identifies runs of sequence similarity between pairs of sequences that are unexpectedly long given the overall distribution of similar bases in a sequence alignment, while controlling for the structure of the genetic code. The structure of this computation illustrates some of the difficulties in testing for GC. Such conversion events violate two key assumptions of the standard models of molecular evolution. First, a GC event can result in local regions of a sequence that do not follow either the overall gene tree for that gene or the species tree. More seriously, GC events, if they result in a “track” of converted bases, also violate the assumption of independence of sites in an alignment.

These difficulties make it difficult to explicitly account for GC in evolutionary models. Instead, using the WGD as our baseline, we have chosen to test for GC by seeking to reject a null model that does not include such events. As we did in our previous analyses of ribosomal proteins (Evangelisti and Conant 2010), we used a triplet-based test to compare two duplicated histone genes (D1 and D2) to the nearest ortholog (O) of D1, identified using POInT. Note that D1 and O are expected to be phylogenetically much closer (separated by a recent speciation) than are D1 and D2 (which last shared a common ancestor at the WGD).

We first aligned the three sequences using T-Coffee (Notredame et al. 2000). For each of the three branches of the tree (D1, D2, and O), we made maximum-likelihood estimates for the number of nonsynonymous (K_a) and synonymous (K_s) substitutions per site (Conant and Wagner 2003). Using a likelihood-ratio test, we then assessed the statistical support for an inference of GC between genes (D1 and D2) created by WGD.

We calculated the likelihood of the alignment allowing for the three values of K_a (or K_s) leading to D1 was constrained to be no less than that leading to O using a chi-square distribution with 1 degree of freedom (table 2).

**Gene Tree Tests of Conversion**

We analyzed alignments of all post-WGD histones for each of the five genes showing evidence of conversion. After T-Coffee alignment, we estimated maximum-likelihood gene trees for the alignments with PhyML (Guindon and Gascuel 2003). Using POInT and the species tree topology from YGOb (Byrne and Wolfe 2005), we created an expected species tree for the histone genes and putative conversion trees as described in the Results section. For all three trees (PhyML...
Triplet-based Relative Rate Tests Coupled to Orthology Predictions Show Evidence for Gene Conversion at Synonymous Sites of Duplicated Histones

Table 2

| Saccharomyces cerevisiae gene | Gene type | Species-specific genes | Probabilities of orthology relationship | $K_s$ | $P^d$ | $K_s^*$ | $P^d$ |
|-------------------------------|-----------|------------------------|----------------------------------------|-------|------|--------|------|
| HTA1                          | $D_1^*$   | TPA04000110            | >0.99                                  | 0.004 | 0.15 | 0.062  | 0.016|
|                              | $D_2^*$   | TPA04002050            | ≥0                                     | 0     | 0.039|        |      |
|                              | $O^a$     | Kpol_1031.53           | ≤0.015                                 | 0.015 | 0.186|        |      |
| HTA2                          | $D_1^*$   | KAF0000788             | ≥0.99                                  | 0     | 0.02 | 0      | <0.001|
|                              | $D_2^*$   | KAF0002490             | ≤0.015                                 | 0.189 |      |        |      |
|                              | $O^a$     | KNAG001430             | ≤0.012                                 | 0.551 |      |        |      |
| HTB2                          | $D_1^*$   | KAF0000770             | ≥0.99                                  | 0.008 | 0.08 | 0.156  | 0.011|
|                              | $D_2^*$   | KAF0002480             | ≤0.015                                 | 0.098 |      |        |      |
|                              | $O^a$     | KNAG001420             | ≤0.027                                 | 0.411 |      |        |      |
| HHF1                          | $D_1^*$   | KAF0000700             | ≥0.97                                  | 0     | 0.07 | 0.011  | <0.001|
|                              | $D_2^*$   | KAFR0001280            | ≤0.010                                 | 0.124 |      |        |      |
|                              | $O^a$     | KNAG001060             | ≤0.010                                 | 0.459 |      |        |      |
|                              | $D_1^*$   | CAGL004136g            | 0.97                                   | ≤0     | 0.24 | 0.028  | <0.001|
|                              | $D_2^*$   | CAGL004136g            | ≤0                                     | 0     | 0.001|        |      |
|                              | $O^a$     | NCAS000078             | ≤0                                     | 0.426 |      |        |      |
| HHF2                          | $D_1^*$   | NDAG003480             | ≥0.99                                  | 0     | 0.20 | 0.054  | <0.001|
|                              | $D_2^*$   | NDAG003480             | ≤0                                     | 0.076 |      |        |      |
|                              | $O^a$     | NCAS006180             | ≤0                                     | 0.315 |      |        |      |
|                              | $D_1^*$   | NCAS006180             | ≥0.99                                  | 0     | 0.21 | 0.071  | 0.005|
|                              | $D_2^*$   | NCAS006180             | ≤0                                     | 0     | 0.088|        |      |
|                              | $O^a$     | NCAS006180             | ≤0                                     | 0.297 |      |        |      |

$^a$Saccharomyces cerevisiae histone gene name (see table 1).

$^b$Estimated probability of the full set of orthology relationships used for this and later analyses from POInT. Thus, of all possible orthology relationship, what proportion of the probability is apportioned to the one described.

$^c$Using our triplet-based likelihood approach (Conant and Wagner 2003), we estimated for each of the three branches (corresponding to the three genes) the number of nonsynonymous ($K_a$) and synonymous ($K_s$) substitutions per site.

$^d$P value for the hypothesis test of equal values of $K_a$ (or $K_s$) for $D_1$ and $O$. This condition corresponds to the hypothesis of no gene conversion: $D_1$ and its ortholog $O$ are equally distant from paralog $D_2$. The test is based on a likelihood-ratio test of a null model where all values of $K_a$ (or $K_s$) are free to an alternative model where the $K_a$ (or $K_s$) values of $D_1$ and $O$ are forced to be equal. The $P$ value was computed by comparing twice the difference in In-likelihood to a chi-square distribution with one degree of freedom. Values shown in bold are significant at $P^c=0.05$.

$^e$Relationship between two paralogs hypothesized to have undergone gene conversion ($D_1$ and $D_2$) and the orthology of $D_1$, $O$.

Results

GC can produce, in an ancient shared gene duplication, a situation where two paralogs from one species are more similar to each other in sequence than either is to its ortholog in the other species. Unfortunately, such a pattern of sequence similarity could also result from independent gene duplications in the two lineages. As we have previously discussed, the special case of a genome duplication allows us to avoid this confound, because the independent signal of shared gene order, or synteny, allows us to infer orthology between duplicated genes regardless of patterns of sequence evolution (Casola et al. 2012).

Here, we have applied a synteny-based approach to the analysis of several WGD-produced histone duplicates that are identical or nearly identical at the amino acid level. Importantly, these paralogs show lower identity to the most closely related of their homologs in their nearest relative than to each other (fig. 2 and table 1). However, for a strict test of GC, we need to compare each paralog with its ortholog in its nearest relative. (The full list of duplicated histone genes in these 12 taxa is available as supplementary data, Supplementary Material online.) Unfortunately, although identifying paralogs shared from the WGD in pairs of yeast genomes is now relatively straightforward (Byrne and Wolfe 2005), assigning orthology between those genes is more
difficult (fig. 2A). We have developed a software tool called POInT that uses synteny (Gordon et al. 2009) and a maximum-likelihood phylogenetic model of gene loss to make probabilistic estimates of which homologous genes in two species sharing a WGD are actually orthologs (Conant and Wolfe 2008; Conant 2014). This program’s inferences are illustrated in figure 2A: The numbers above each column give POInT’s estimate of the probability of the orthology relationship shown as compared with all other $2^{n-1}-1$ possible assignments. Here $n$ is the number of genomes. Each duplicated gene in a particular genome could be assigned as the ortholog of one of the two genes (or potentially a position where a gene was inferred to be lost—gaps in fig. 2A) in each of the other $n-1$ genomes. The synteny data strongly support the hypothesis that these histone duplicates last shared a common ancestor at the WGD event and not after the more recent speciation event. Thus, for HTA2, even though the amino acid sequences of the two duplicates from Kazachstania africana are identical, one of those two genomic loci last shared a common ancestor with its Kazachstania naganishii ortholog more recently than with its WGD-produced naganishii ortholog. This pattern of sequence identity can either be explained by random fixations giving rise to a misleading gene tree when comparing the gene sequences (akin to lineage sorting) or by GC acting on those sequences.

We adopted two tests to choose between these hypotheses. First, as in our previous analysis, we compared the sequence divergence of the putatively converted genes with one of their orthologs using a triplet-based analysis of asymmetric evolution (Conant and Wagner 2003; Evangelisti and Conant 2010). With this approach, we compute branch-specific values for the nonsynonymous ($K_a$) or synonymous ($K_s$) divergence for three genes: Two WGD-produced paralogs $D_1$ and $D_2$ and $O$, the ortholog of $D_1$ (table 2). If GC has not occurred, the sum of the $K_a$ values for $O$ and $D_1$ should be less than the sum for $D_1$ and $D_2$. Our observation of GC implies instead that the sum of the $K_s$ values for $D_1$ and $D_2$ is less than that for $D_2$ and $O$. Thus, rejecting the null hypothesis that $K_a$ for $D_1$ is equal to that for $O$ is equivalent to rejecting the null hypothesis of no GC (see Methods). We suspected that the high amino acid identities of the histones would make this test relatively insensitive. And, indeed, as table 2 shows, we were unable to reject the hypothesis of equal divergence in $K_a$ between $D_1$ and $D_2$ in six of the seven cases. However, in all cases we saw significant evidence of conversion in $K_s$ values ($P \leq 0.015$, likelihood-ratio test; Methods).

Because this first test was inconclusive, we next placed the putative conversion events onto the species phylogeny postulated by YGOB (Byrne and Wolfe 2005; Gordon et al. 2009). For the five putative cases of conversion in table 1, we assigned orthology for the duplicated genes using our orthology inference tool POInT (see Methods). The In-likelihood of the species phylogeny for each alignment was compared with all possible GC-type gene trees (table 1), using a codon model
of evolution (see Methods). These trees were created by taking the species phylogeny and moving one of the putatively gene converted sequences to be sister to its WGD-produced paralog (fig. 2). From this set of possible rearranged trees, we retained the one with the highest ln-likelihood. In all cases, this best GC tree had a higher ln-likelihood than the species tree. To assess if this higher ln-likelihood was statistically significant, we simulated 1,000 alignments on the species tree under the same codon model. For each simulation, we then compared the ln-likelihood of the species tree with that of the GC tree found to have the highest likelihood for that simulation (fig. 1). In no case did the simulated data sets have an improvement ln-likelihood from the optimal GC tree as large as the improvement seen in the real data (P < 0.001; dashed line in fig. 1). We therefore conclude that these sequences show significant evidence for GC, as the only difference between the species tree and the GC trees is the position of the putatively converted paralogs.

Unusual Substitution Patterns among Histone Genes

To further explore the evolution of these converted genes, we applied an SG (Similarity Groups) model (Conant et al. 2007) to each histone alignment used above. The SG model separates the amino acids into polar and nonpolar residues and allows one selective constraint (Rr) for substitutions within the same group and a second (Rc) for substitutions between groups. The parameter estimates for the MG/GY and SG models are given in supplementary table S1, Supplementary Material online. The supplementary figure, Supplementary Material online, illustrates the location of substitutions in the converted histone genes relative to the S. cerevisiae structural model (White et al. 2001), while supplementary table S2,
Supplementary Material online, gives the locations of all substitutions relative to the crystal structure. In general, most genes show $R_r < R_c$, because substitutions that do not change polarity should be less drastic and hence less likely to provoke purifying selection (Zhang 2000). Unexpectedly, however, all the histone genes with at least one instance of conversion showed $R_r > R_c$ (fig. 3). Even in sequences that are simply drifting, observing $R_r > R_c$ in five of the five cases is unexpected (i.e., under the null hypothesis that $R_r = R_c$ in 50% of the cases; $P = 0.03$, binomial test). However, after a false-discovery rate correction, we cannot reject the null hypothesis of $R_r = R_c$ for any of those five alignments ($P > 0.05$, likelihood-ratio test with FDR correction; Benjamini and Hochberg 1995). Note however that, in three cases, the genes inferred to have undergone GC showed at least one radical amino acid substitution and no conservative substitutions (fig. 3). These patterns are again unexpected if only strong purifying selection is acting on histone genes and may suggest the presence of coevolution among histones.

**Discussion**

With a second group of genes that have undergone GC since their most recent speciation event, we may have the beginnings of a pattern. In particular, the two groups of genes share two important features: They are well conserved, and they function as part of essential macromolecular complexes. The dosage balance hypothesis predicts that such genes are likely to be retained in duplicate after the WGD due to their many interactions (Veitia et al. 2008). But why might GC occur? We speculate that the tight nature of the complexes may be the reason. The fixation of a GC might be beneficial in the presence of coevolution between a pair of duplicates and their interaction partner(s). If a change in the interaction partner has resulted in a compensating mutation in one of the duplicates, the transfer of that change to the second duplicate by GC would then be beneficial because it allows the complex to be maintained no matter the duplicate copy used. This idea is speculative because we do not yet have a test that would indicate whether a gene converting mutation became fixed through selection. However, there are several suggestive points. Among the ribosomal proteins, the signal of conversion was much stronger for nonsynonymous substitutions than for synonymous ones, a fact difficult to reconcile with drift. Here, although we did not find such a difference in the two types of substitution, we do note that the duplicated genes are not identical at the nucleotide level despite showing a signal of conversion in synonymous substitutions (table 2). Similarly, the amino acid substitutions that have occurred actually appear biased toward more radical amino acid changes (fig. 3). However, the observation of $R_r > R_c$ should not be taken as “classical” positive selection in this instance and not merely because the statistical evidence for rejecting $R_r = R_c$ is lacking. (As an aside, we note that while the use of $R_r$ and $R_c$ has been criticized, our likelihood-based estimates of these parameters do not suffer from many of the biases seen with older estimation methods; Dagan et al. 2002.) The reason for caution in interpreting these values is that, in these histone genes, there are forces at work beyond those of mutation and selection that form the basis of the standard models. In addition to the mutational process generating variants in the duplicate copies, there is a GC process that may either copy a variant into the other duplicate or eliminate it through the alterative conversion. Selection in turn operates on both types of event, while at the same time there is a potential for selection driving coevolution between interacting partners in the nucleosome. With all these forces at work, our intuition as to how evolution is operating is likely to be rather poor: We simply argue that our observations here are not consistent with drift or the simplest form of purifying selection.

We also note that, because GC often occurs in tracks of multiple bases, the observation of conversion at synonymous positions might then be a hitchhiking effect of the selective
preservation of nonsynonymous conversions. We also previously showed with the ribosomal proteins that expression level alone was not sufficient to explain the frequency of GC in these genes, despite the existence of RNA-based GC mechanisms in yeast (Derr and Strathern 1993; Storici et al. 2007).

These results also fit into the larger picture of the yeast genome duplication. An elegant analysis by Marce-Houben and Gabaldon (2015) used gene trees to infer that the yeast WGD was very likely an allopolyploid event, e.g., the merging of the “diploid” genomes of two distinct species. Interestingly, however, the extant duplicated genomes are not an equal mix of these two source genomes: One genome seems to have come to dominance, possibly by a combination of biased gene losses and GCs (Tang et al. 2012; Wolfe 2015). We speculate that the process of GC observed here for the histone and ribosomal protein genes may have started with selection acting for this sort of genome dominance. In that view, cassettes of critical genes, such as those for histones, ribosomal proteins, or DNA repair enzymes, would have come from both parents, but might not function interchangeably. Indeed, such a combination might set the stage for dominant negative interactions, where the presence of an alternative version of a particular gene caused a reduction in fitness (De Smet et al. 2013). We previously observed one potential solution to this conundrum: The duplicate copies of genes for mitochondrially targeted proteins and DNA repair enzymes were rapidly lost after the yeast WGD (Conant 2014). GC is a second solution, where two genetic loci are retained but both contain sequences originating from a single parent.

Duplicate genes continue to surprise us with their multifaceted evolutionary patterns (Hahn 2009). In so doing, they justify Ohno’s interest not only from an evolutionary perspective but also because the response of a biological complex to the duplication of its members reveals a good deal about its function.

Supplementary Material

Supplementary figure, data, tables S1 and S2 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

We would like to thank K. Byrne, J. Gordon, and K. Wolfe for access to genomic data. We also appreciate helpful discussions with J. C. Pires and computational assistance from M. Becchi. This work was supported by National Science Foundation grant NSF-IOS-1339156 to the University of Missouri.

Literature Cited

Arnheim N, et al. 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.

Baldo AM, Les DH, Strausbaugh LD. 1999. Potentials and limitations of histone repeat sequences for phylogenetic reconstruction of Sophophora. Mol Biol Evol. 16:1511–1520.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B. 57:289–300.

Birchler JA, Veitia RA. 2012. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proc Natl Acad Sci U S A. 109:14746–14753.

Blanc G, Wolfe KH. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16:1679–1691.

Boles E, et al. 1997. Characterization of a glucose-repressed pyruvate kinase (Pyk2p) in Saccharomyces cerevisiae that is catalytically insensitive to fructose-1,6-biphosphate. J Bacteriol. 179:2987–2993.

Bork P, et al. 2004. Protein interaction networks from yeast to human. Curr Opin Struct Biol. 14:292–299.

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

Carretero-Paulet L, Fares MA. 2012. Evolutionary dynamics and functional specialization of plant paralogs formed by whole and small-scale genome duplications. Mol Biol Evol. 29:3541–3551.

Casola C, Conant GC, Hahn MW. 2012. Very low rate of gene conversion in the yeast genome. Mol Biol Evol. 29:3817–3826.

Chen JM, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP. 2007. Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet. 8:762–775.

Codoner FM, Fares MA. 2008. Why should we care about molecular coevolution? Evol Bioinform Online 4:29–38.

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, 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. 2003. Asymmetric sequence divergence of duplicate genes. Genome Res. 13:2052–2058.

Conant GC, Wagner GP, Stadler PF. 2007. Modeling amino acid substitution patterns in orthologous and paralogous genes. Mol Phylogenet Evol. 42:298–307.

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

Cooke J, Nowak MA, Boerlijst M, Maynard-Smith J. 1997. Evolutionary origins and maintenance of redundant gene expression during metazoan development. Trends Genet. 13:360–364.

Dagan T, Talmor Y, Graur D. 2002. Ratios of radical to conservative amino acid replacement are affected by mutational and compositional factors and may not be indicative of positive Darwinian selection. Mol Biol Evol. 19:1022–1025.

De Smet R, et al. 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.

DeBry RW, Marzluff WF. 1994. Selection on silent sites in the rodent H3 histone gene family. Genetics 138:191–202.
Gene Conversion and Selection on Macromolecular Complexes

Derr, LK; Strathern, JN. 1993. A role for reverse transcripts in gene convert-
sion. Nature 361:170–173.

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.

Evangelist, AM; Conant, GC. 2010. Nonrandom survival of gene conve-
sions among yeast ribosomal proteins duplicated through genome
doubling. Genome Biol. Evol. 2:826–834.

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

Gao, L.; Inman, H. 2004. Very low gene duplication rate in the yeast
genome. Science 306:1367–1370.

Gavin, AC, et al. 2006. Proteome survey reveals modularity of the yeast cell
machinery. Nature 440:631–636.

Goldman, N; Yang, Z. 1994. A codon-based model of nucleotide substitu-
tion for protein-coding DNA sequences. Mol Biol Evol. 11:725–736.

Gordon, JL; Byrne, KP; Wolfe, KH. 2009. Additions, losses and rearrange-
ments on the evolutionary route from a reconstructed ancestor to the
modern Saccharomyces cerevisiae genome. PLoS Genet. 5:e1000485.

Guindon, S; Gascuel, O. 2003. A simple, fast, and accurate algorithm to
estimate large phylogenies by maximum likelihood. Syst Biol. 52:696–704.

Hahn, MW. 2009. Distinguishing among evolutionary models for the main-
tenance of gene duplicates. J Hered. 100:605–617.

Hakes, L; Pinney, JV; Lovell, SC; Oliver, SG; Robertson, DL. 2007. All dupli-
cates are not equal: the difference between small-scale and genome
duplication. Genome Biol. 8:R209.

He, X; Zhang, J. 2006. Higher duplicability of less important genes in yeast
genomes. Mol Biol Evol. 23:144–151.

Hittinger, CT; Carroll, SB. 2007. Gene duplication and the adaptive evolu-
tion of a classic genetic switch. Nature 449:677–681.

Ihmels, J; Collins, SR; Schuldiner, M; Krogan, NJ; Weissman, JS. 2007. Backup
without redundancy: genetic interactions reveal the cost of duplicate
gene loss. Mol Syst Biol. 3:86.

Kim, TY; Ha, CW; Huh, WK. 2009. Differential subcellular localization of
ribosomal protein L7 paralogs in Saccharomyces cerevisiae. Mol Cell.
21:539–546.

Kondrashov, FA; Kondrashov, AS. 2006. Role of selection in fixation of gene
duplications. J Theor Biol. 239:141–151.

Kroga, MJ; Krogan, NJ; et al. 2006. Global landscape of protein complexes in the yeast
Saccharomyces cerevisiae. Nature 440:637–643.

Kunin, V; Pereira-Leal, JB; Ouzounis, CA. 2004. Functional evolution of the yeast
protein interaction network. Mol Biol Evol. 21:1171–1176.

Li, L; Huang, Y; Xia, X; Sun, Z. 2006. Preferential duplication in the sparse part
of yeast protein interaction network. Mol Biol Evol. 23:2467–2473.

Liao, D. 1999. Concerted evolution: molecular mechanism and biological
implications. Am J Hum Genet. 64:24–30.

Liberman, DA; et al. 2012. The interface of protein structure, protein bio-
physics, and molecular evolution. Protein Sci. 21:769–785.

Maere, S; et al. 2005. Modeling gene and genome duplications in eukary-
otes. Proc Natl Acad Sci U S A. 102:5454–5459.

Makino, T; McLysaght, A. 2010. Ohnologs in the human genome are
preservation of genetic and regulatory robustness in ancient gene du-
plexing without redundancy: genetic interactions reveal the cost of duplicate
gene loss. Mol Syst Biol. 3:86.

Matsuo, Y; Yamazaki, T. 1989. Nucleotide variation and divergence in the
histone multigene family in Drosophila melanogaster. Genetics
122:87–97.

Masson, R; Cohn, R; Kedes, L; Mohun, T. 1983. Expression and organization of
histone genes. Annu Rev Genet. 17:239–277.

Mintseris, J; Weng, Z. 2005. Structure, function, and evolution of transient and
obligate protein-protein interactions. Proc Natl Acad Sci U S A.
102:10930–10935.

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 se-
quence-mediated domain transfer. Proc Natl Acad Sci U S A.
77:2143–2147.

Muse, SV; Gaut, BS. 1994. A likelihood approach for comparing synonomous and non-synonomous nucleotide substitution
rates, with application to the chloroplast genome. Mol Biol Evol.
11:715–724.

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.

Novak, MA; Boerlijst, MC; Cooke, J; Maynard-Smith, J. 1997. Evolution of
genetic redundancy. Nature 388:167–171.

Ozcan, S; Johnston, M. 1999. Function and regulation of yeast hexose transporters. Microbiol Mol Biol Rev. 63:554–569.

Papp, B; Pal, C; Hurst, LD. 2003. Dosage sensitivity and the evolution of gene
families in yeast. Nature 424:194–197.

Pérez-Bercoff, A; Makino, T; McLysaght, A. 2010. Duplicability of self-
interacting human genes. BMC Evol Biol. 10:160.

Plontikvskia, H; Rooney, AP; Nei, M. 2002. Purifying selection and birth-
and-death evolution in the histone H4 gene family. Mol Biol Evol.
19:689–697.

Planta, RJ; Mayer, WH. 1998. The list of cytoplasmic ribosomal proteins of
Saccharomyces cerevisiae. Yeast 14:471–477.

Qian, W; He, X; Chan, E; Xu, H; Zhang, J. 2011. Measuring the evolutionary rate of protein-protein interaction. Proc Natl Acad Sci U S A.
108:8725–8730.

Qian, W; Liao, BY; Chang, AYF; Zhang, J. 2010. Maintenance of duplicate
genes and their functional redundancy by reduced expression. Trends
Genet. 26:425–430.

Rodríguez, A; de la Cera, T; Herrero, P; Moreno, F. 2001. The hexokinase 2
protein regulates the expression of the GLK1, HXK1 and HXK2 genes of
Saccharomyces cerevisiae. Biochem J. 355:625–631.

Rooney, AP; Plontikvskia, H; Nei, M. 2002. Molecular evolution of the non-
tandemly repeated genes of the histone 3 multigene family. Mol Biol
Evol. 19:68–75.

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

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

Séron, M; Wolfe, KH. 2007. Consequences of genome duplication. Curr
Opin Genet Dev. 17:505–512.

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

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

Tang, H; et al. 2012. Altered patterns of fractionation and exon deletions in
Brassica rapa support a two-step model of paleohexaploidy. Genetics
190:1563–1574.

Taylor, JD; Wellman, SE; Marzluff, WF. 1986. Sequences of four mouse
histone H3 genes: implications for evolution of mouse histone
genes. J Mol Evol. 23:242–249.

van Hoof, A. 2005. Conserved functions of yeast genes support the dupli-
cation, degeneration and complementation model for gene duplica-
tion. Genetics 171:1455–1461.
Veitia RA, Bottani S, Birchler JA. 2008. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. Trends Genet. 24:390–397.

Veitia RA, Bottani S, Birchler JA. 2013. Gene dosage effects: nonlinearities, genetic interactions, and dosage compensation. Trends Genet. 29:385–393.

Wagner A. 2000. The role of population size, pleiotropy, and fitness effects of mutations in the evolution of overlapping gene function. Genetics 154:1389–1401.

Wagner A. 2001. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. Mol Biol Evol. 18:1283–1292.

Wagner A. 2003. How the global structure of protein interaction networks evolves. Proc Biol Sci. 270:457–466.

Wang ZF, Krasikov T, et al. 1996. Characterization of the mouse histone gene cluster on chromosome 13: 45 histone genes in three patches spread over 1Mb. Genome Res. 6:688–701.

Wang ZF, Tisovec R, et al. 1996. Characterization of the 55-kb mouse histone gene cluster on chromosome 3. Genome Res. 6:702–714.

White CL, Suto RK, Luger K. 2001. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. EMBO J. 20:5207–5218.

Wolfe KH. 2015. Origin of the yeast whole-genome duplication. PloS Biol. 13:e1002221.

Woodhouse MR, et al. 2010. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. PloS Biol. 8:e1000409.

Zhang J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J Mol Evol. 50:56–68.

Zhu Y, Lin Z, Nakhleh L. 2013. Evolution after whole-genome duplication: a network perspective. G3 (Bethesda) 3:2049–2057.

**Associate editor:** Rebecca Zufall