Evolutionary Bioinformatics

Libertas Academica
FREEDOM TO RESEARCH

ORIGINAL RESEARCH

Evolutionary Analysis of Sequence Divergence and Diversity of Duplicate Genes in *Aspergillus fumigatus*

Ence Yang[1], Amanda M. Hulse[2] and James J. Cai[1,2]

[1]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas, USA. [2]Interdisciplinary Program in Genetics, Texas A&M University, College Station, Texas USA. Corresponding author email: jcai@tamu.edu

Abstract: Gene duplication as a major source of novel genetic material plays an important role in evolution. In this study, we focus on duplicate genes in *Aspergillus fumigatus*, a ubiquitous filamentous fungus causing life-threatening human infections. We characterize the extent and evolutionary patterns of the duplicate genes in the genome of *A. fumigatus*. Our results show that *A. fumigatus* contains a large amount of duplicate genes with pronounced sequence divergence between two copies, and approximately 10% of them diverge asymmetrically, i.e. two copies of a duplicate gene pair diverge at significantly different rates. We use a Bayesian approach of the McDonald-Kreitman test to infer distributions of selective coefficients $\gamma (=2N_e s)$ and find that (1) the values of γ for two copies of duplicate genes co-vary positively and (2) the average γ for the two copies differs between genes from different gene families. This analysis highlights the usefulness of combining divergence and diversity data in studying the evolution of duplicate genes. Taken together, our results provide further support and refinement to the theories of gene duplication. Through characterizing the duplicate genes in the genome of *A. fumigatus*, we establish a computational framework, including parameter settings and methods, for comparative study of genetic redundancy and gene duplication between different fungal species.

Keywords: duplicate gene, *Aspergillus fumigatus*, positive selection, sequence diversity

Evolutionary Bioinformatics 2012:8 623–644

doi: 10.4137/EBO.S10372

This article is available from http://www.la-press.com.

Introduction

Eukaryotic genomes are characterized by the presence of numerous multigene families. In a typical eukaryotic species, more than a third of its proteins are encoded in genes that belong to multigene families formed by duplication of a single original gene.[1–3] Gene duplication is believed to be a major evolutionary event that provides a source for genetic innovation.[4] Duplicate genes may facilitate the development of phenotypic diversity in organismal evolution and are likely to play a prominent role in the adaptive evolution of eukaryotes.[5,6] Computational analysis based on molecular evolution models may provide valuable information about gene evolution. By examining the extent of gene duplication, which is manifested by the frequency and magnitude of gene duplication events and the consequent evolutionary fates of gene pairs following the duplication events,[7,8] we may gain novel insights into the evolution of organismal adaptation.[5,9,10] One of the evolutionary patterns related to gene duplication is the difference in evolutionary rates between the two copies of duplicate genes, which has attracted great interest.[9–16] Understanding the evolutionary forces underlying the different evolutionary rates of duplicated genes requires examination of this phenomenon in more organisms. With the plethora of genomic data, it is possible to study the features of gene duplication, including the asymmetric divergence of the duplicate genes, on a genome-wide scale. In the present study, we examined the extent and evolutionary patterns of duplicate genes in several fungal species.

Aspergillus fumigatus is a ubiquitous filamentous fungus and one of the most important opportunistic fungal pathogens. It plays an essential role in recycling environmental carbon and nitrogen by growing on organic debris in soil, its natural ecological niche.[17] Due to the widespread distribution of small airborne conidia, *A. fumigatus* is inevitably inhaled into the airways and the lungs of human beings.[18] These conidia are cleared by the innate immune system of healthy individuals but may cause invasive infection in immunosuppressed individuals. Aspergillosis represents a major cause of morbidity and mortality in patients receiving immunosuppressive therapy for autoimmune or neoplastic disease, in organ transplant recipients, and in AIDS patients.[19,20] For many years, *A. fumigatus* was not thought to reproduce sexually, as neither mating nor meiosis had ever been observed.[21] Recently, *A. fumigatus* was shown to possess a fully functional sexual reproductive cycle, and it was accordingly renamed *Neosartorya fumigata*.[22] Considering that it has been known as an anamorph for the majority of its research history, we use *A. fumigatus* hereafter.

This study focused on examining the extent and evolutionary patterns of duplicate genes in *A. fumigatus* using evolutionary bioinformatics approaches. To gain a comprehensive and broad view across species, we compared the analytical results obtained in *A. fumigatus* with those obtained in four other fungal species, which were selected from a diverse phyletic background. The four species were *Cryptococcus neoformans*, *Neurospora crassa,*[23] *Saccharomyces cerevisiae,*[24] and *Schizosaccharomyces pombe*[25] (Table 1). These fungi have distinct life styles and diverse phenotypic characteristics. The brewer's yeast *S. cerevisiae* and the fission yeast *S. pombe* have a life cycle characterized by a unicellular thallus that reproduces by budding and fission respectively. *N. crassa* is a filamentous ascomycetes growing hyphae apically and branching laterally. *C. neoformans* is a dimorphic fungus that is able to grow in either a yeast or hyphal mode in response to certain environmental conditions.[26,27] We used one species within the *Aspergillus* genus, *A. fischeri*,[28] as the closely related outgroup species for evolutionary analysis of *A. fumigatus* genes. Finally we used polymorphic data ascertained in 12 different strains of *A. fumigatus* to infer the population-scale selection

Table 1. Fungal species and sources of genomic and proteomic sequences.

Species	Strain	Reference	Source*
A. fumigatus	Af293	66	AspGD
C. neoformans var. *grubii*	H99	*C. neoformans* sequencing project†	FGI
N. crassa	OR74A	23	FGI
S. cerevisiae	S288c	24	NCBI
S. pombe	972h-	25	NCBI
A. fischeri	NRRL181	28	NCBI

Note: †*Cryptococcus neoformans* var. *grubii* H99 Sequencing Project, Broad Institute of Harvard and MIT (http://www.broadinstitute.org/). AspGD, *Aspergillus* Genome Database, http://www.aspgd.org/; FGI, The Fungal Genome Initiative of Broad Institute, http://www.broadinstitute.org/scientific-community/science/projects/fungal-genome-initiative; NCBI, http://www.ncbi.nlm.nih.gov/genome.

coefficient for duplicate genes using a McDonald-Kreitman (MK) type analysis.[29,30]

Materials and Methods

Identification of duplicate genes

The computer program BLASTCLUST (BLAST score-based single-linkage clustering) was used to automatically and systematically cluster protein sequences. (For documentation on its use, see ftp://ftp.ncbi.nlm.nih.gov/blast/documents/blastclust.html.) Briefly, BLASTCLUST clustering is based on pairwise matches between protein sequences found using the BLAST algorithm. BLASTCLUST uses the default values for the BLAST parameters, including the matrix BLOSUM62, gap opening cost 11, gap extension cost 1, no low-complexity filtering, and e-value threshold 1e-6 for protein sequences. For each pair of sequences, the top-scoring alignment is evaluated according to the minimum length coverage (L = 0.0 to 1.0) and a similarity threshold (ie, the percent of identical residues, S = 3% to 100%) to determine whether the pair of sequences should be linked to each other, providing the base for clustering by the single-linkage method. Different combinations of L and S values influence the results of clustering for the same protein sequences. In order to obtain the best combination of L and S for most accurately clustering protein sequences in our study, we iterated a full range of possible values of the two parameters and ran BLASTCLUST to cluster two test sets of protein families (Supplementary Tables S1 and S2). The two test sets of protein families were constructed by randomly selecting proteins from *A. fumigatus* and *C. neoformans* proteomes. For each fungus, a total of 50 protein families were manually constructed using BLAST search against the fungus' own proteome sequences and visual inspection. These protein families varied in size: 10 families contained 5 proteins or more, and 10 families each contained 4, 3, 2, and 1 protein(s). The two sets of protein families constructed for the two fungal species were used as two test sets to search for the optimized values for parameters L and S. During the search, L was set from 10% to 100% in intervals of 1% and S was set from 25% to 45% with a 1% interval. For each combination of parameters, the accuracy of the BLASTCLUST result was estimated using the percentage of protein families correctly clustered. For each family, correct clustering meant the complete match of gene members between BLASTCLUST clustering results and clusters in the original manually constructed training sets.

Estimation of divergence rates

Protein sequences were aligned using CLUSTALW (version 1.82) with the default parameters. To obtain the alignments of codons, the corresponding nucleotide-sequence alignments were derived by substituting the respective coding sequences from the protein sequences. The number of synonymous substitutions per synonymous site, d_S, and the number of nonsynonymous substitutions per nonsynonymous site, d_N, were calculated using the maximum-likelihood method implemented in the codeml program of the PAML package.[31] For each pair of genes, we repeated the computation of d_S, d_N and d_N/d_S 1000 times and took the median of the 1000 repeats as the final values. Pairs with $d_S \geq 5.0$ were eliminated because such high sequence divergence is often associated with problems such as difficulty in alignment, different codon usage biases, or nucleotide compositions in different sequences. The genetic distance between proteins d_{WAG} based on the empirical WAG model[32] were computed using MBEToolbox.[33] Protein pairs with $d_{WAG} \geq 2.0$ were excluded from further analysis.

Test for asymmetric evolution

A total of 202 pairs of *A. fumigatus* duplicate genes with $d_S \leq 5.0$ between copies were used in this test. These genes' orthologs were identified using reciprocal BLASTP search in *A. fischeri*. We found that 44 *A. fumigatus* duplicate gene pairs in which two copies of genes had the same orthologous gene in *A. fischeri*. In these cases, two copies of *A. fumigatus* duplicate genes and their *A. fischeri* orthologs formed sequence triplets (Fig. 1, left panel). There were 158 *A. fumigatus* gene pairs in which two copies of genes had two distinct orthologous genes in *A. fischeri*. In these cases, two copies of *A. fumigatus* duplicate genes and their corresponding *A. fischeri* orthologs formed sequence quadruplets (Fig. 1, right panel). The test for the asymmetric evolution was constituted as a relative rate test between a pair of *A. fumigatus* duplicate genes on an unrooted tree. *A. fischeri* orthologous sequence(s) were used as outgroup(s). The statistical tests were

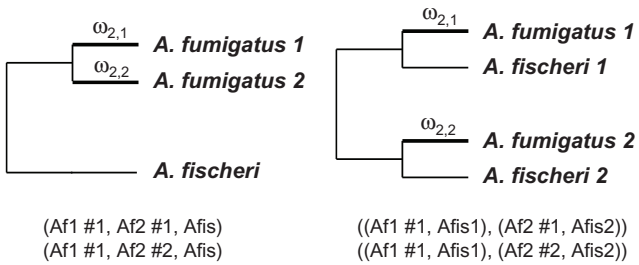

Figure 1. Outline of branch-site models used in the study.
Notes: The phylogenies illustrate the cases of two *A. fumigatus* duplicate genes with one *A. fisheri* ortholog (left) and two *A. fumigatus* duplicate genes with one *A. fisheri* orthologs (right). The two branches leading to the *A. fumigatus* duplicate genes are labeled with class 2 selective pressure measures $\omega_{2,1}$ and $\omega_{2,2}$, respectively. Phylogenies in Newick format are given under the trees. The labels #1 and #2 specify the two models in which $\omega_{2,1} = \omega_{2,2}$ and $\omega_{2,1} \neq \omega_{2,2}$, respectively. Af *A. fumigatus*, Afis, *A. fisheri*.

conducted with a codon-based branch-site model using Codeml program of the PAML package.[31] We used the clade model C, which allows for two branch types (clades) and assumes three site classes: site class 0 of conserved sites with $\omega_0 < 1$, site class 1 of neutral sites with $\omega_1 = 1$, and site class 2 with different selective pressures ($\omega_{2,1}$ and $\omega_{2,2}$) in the two clades. We compared $\omega_{2,1}$ and $\omega_{2,2}$ between two copies of a pair of duplicate gene to detect the difference in the proportion of selected sites in the two clades.[34] Likelihood ratio (LR) test was used to test for significance. To do this, two models were applied to the data: model 1 ($\omega_{2,1} = \omega_{2,2}$) constrains the two ω_2 values to be equal on the two sequences, and model 2 ($\omega_{2,1} \neq \omega_{2,2}$) estimates the two ω_2 values as free parameters. Collected maximum likelihood values ML_1 and ML_2 from the two models were used to calculate the likelihood ratio, $LR = 2(lnML_1 - lnML_2)$. LR is then compared against the χ^2 distribution with one degree of freedom.

Estimation of selection coefficients

We used the MKPRF test[29] to estimate γ of duplicate genes in *A. fumigatus*. The default values of initial parameters as given in the web service of the program at http://cbsuapps.tc.cornell.edu/MKPRF.aspx were taken. Notably, we used the hierarchical model option FIXED_VARIANCE = 0 and the standard deviation (σ) of the Gaussian prior of γ at 8.0. Given that results of MKPRF may be sensitive to some initial values of parameters,[35] we repeated the analysis using different values of σ at 1, 4, and 16. No qualitatively different results were produced in estimation of the means or 95% CIs of γ for duplicate genes in *A. fumigatus*. The coding SNPs in *A. fumigatus* genes were obtained from the *A. fumigatus* genome sequencing project at J. Craig Venter Institute (JCVI) in collaboration with the University of Manchester. These SNPs were ascertained from genome sequences of 12 strains Af293, A1163, AF10, AF210, AFB62, F11628, F11698, F12865, F14946, F15767, F15861, and F16867 using a SNP calling pipeline developed at JCVI. The released project data can be retrieved from the NCBI BioProject database (http://www.ncbi.nlm.nih.gov/bioproject) with IDs 14003, 18733, 46347, 52783, 9521, and 67101.

Results

Extent of duplicate genes in fungal species

To compare the genome-wide extent of gene duplication across species, we obtained the protein sequences of complete proteomes of *A. fumigatus*, *C. neoformans*, *N. crassa*, *S. cerevisiae*, and *S. pombe* from various sources (Table 1). We computationally identified multigene families in each of those fungi by clustering proteins into family groups based on the sequence similarity between protein pairs. It is known that the clustering process is sensitive to the statistical criteria used in determining sequence homologs.[36,37] For instance, when sequence homologs are determined by using the e-value of BLAST algorithm alone, without specifying the proportion of alignable regions, two non-homologous proteins are likely to be grouped into the same family as homologs due to domain sharing.[38]

We adapted the program BLASTCLUST that takes two key parameters, sequence similarity (*S*) and the proportion of alignable regions (*L*), to ascertain the homologous relationship between a pair of protein sequences. To determine the optimized values of *L* and *S*, we manually created two sizable sets of gene families and used them as training data sets. The best combination of values of *L* and *S* were those values that produced the most accurate clustering results for the training data sets. That is to say, the results of automatic clustering by using BLASTCLUST with these *L* and *S* values were most similar to the results of manual clustering (Materials and Methods). The results of BLASTCLUST are given in Figure 2. It is noteworthy that although the test protein sets were constructed separately for *A. fumigatus* and *C. neoformans*, the two most diverged species in our analysis, highly similar values of optimized parameters were obtained: For *A.*

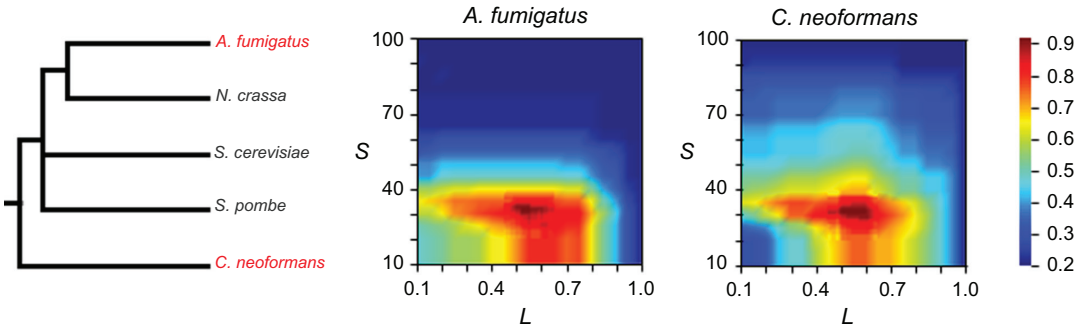

Figure 2. Optimization of the values of *L* and *S* parameters for BLASTCLUST in *A. fumigatus* and *C. neoformans*.
Notes: Phylogenetic tree of the five fungal species included in this study, *A. fumigatus*, *N. crassa*, *S. cerevisiae*, *S. pombe,* and *C. neoformans*. The tree topology is taken from James et al.[67] Heat maps show the percentage of accurately clustered gene families as a function of *L* and *S* values in the two tested species, *A. fumigatus* and *C. neoformans*.

fumigatus, the optimized alignable region between two proteins was between 52% and 55% of the longer protein and the optimized amino-acid similarity was between 31% and 32%, while for *C. neoformans*, the values were between 53% and 61% and between 30% and 32%, respectively (Fig. 2). Accordingly, we set our criteria of homologous sequences as the alignable region between two proteins to be at least 53% of the longer protein and the alignable region contain more than 31% amino-acid identities.

The BLASTCLUST results showed that 25.9% of *A. fumigatus* proteins (2565 of 9887) belong to multigene families (including at least two genes). The percentages for *C. neoformans*, *N. crassa*, *S. cerevisiae*, and *S. pombe* are 18.1%, 15.4%, 29.5%, and 23.1%, respectively. *S. cerevisiae* showed a higher percentage of duplicate genes compared with *A. fumigatus*. Notably, *S. cerevisiae* showed a higher percentage of duple duplicate families than *A. fumigatus* and others, which is probably due to the whole genome duplication of *S. cerevisiae*.[39] For triple duplicate families, *A. fumigatus* is higher than others. For tetra duplicate families, *A. fumigatus* is as large as *S. pombe* and higher than the others. For larger families, the percentage of *A. fumigatus* is also higher than that of other species. Taken together, *A. fumigatus* exhibits more proteins that belong to multigene families containing more than two genes than the other fungi in consideration (Table 2).

Our clustering results for *S. cerevisiae* were comparable to those obtained in previous studies,[40,41] but were higher than those obtained by Kondrashov et al,[10] who used BLASTCLUST with alignments of at least 95% of their lengths and with a score density of 1.5 bits per position, which approximately corresponds to 75% identity. We found this setting is too strict to produce a sufficient number of multigene families for the non-*S. cerevisiae* fungal species we analyzed.

Age distribution of duplicate genes in fungal species

To obtain the age distribution of duplicate genes, we constructed the distribution of synonymous substitution rate (d_S) for all fungal species, as d_S increases approximately linearly with divergence

Table 2. Distributions of protein-coding genes in singleton and multigene families.

Species	*A. fumigatus*	*C. neoformans*	*N. crassa*	*S. cerevisiae*	*S. pombe*
Number of total genes	9,887	6,968	9,733	5,863	5,010
Number of total families	8,077	6,106	8,742	4,733	4,256
Number and % of multigene families					
$n \geq 5$	100 (1.24%)	46 (0.75%)	54 (0.62%)	47 (0.99%)	33 (0.78%)
$n = 4$	59 (0.73%)	28 (0.46%)	25 (0.29%)	27 (0.57%)	31 (0.73%)
$n = 3$	145 (1.80%)	65 (1.06%)	84 (0.96%)	70 (1.48%)	53 (1.25%)
$n = 2$	451 (5.58%)	261 (4.27%)	341 (3.90%)	455 (9.61%)	285 (6.70%)
Number and % of singletons					
$n = 1$	7,322 (90.65%)	5,706 (93.45%)	8,238 (94.23%)	4,134 (87.34%)	3,854 (90.55%)

time. We first computed d_S between all pairs of duplicate genes in the same multigene family for all gene families. We then excluded gene pairs with $d_S \geqq 5.0$ to avoid the problem of d_S saturation. We followed the procedure described by Zhang et al[8] to pick representative duplicate gene pair(s) from each gene family. The procedure guaranteed that the distribution and the mean of d_S was not determined by those gene families with extremely large numbers of genes. From each family we picked the gene pair with smallest d_S and excluded the pair from the gene family. We then repeated this process for the remaining genes within the same family until the last pair. A total of 228 pairs of duplicate genes in *A. fumigatus* were selected and analyzed. We plotted the relative frequency of gene pairs as a function of d_S (Fig. 3).

All fungal species except *N. crassa* displayed similar patterns in the distribution of d_S between two copies of duplicate genes. The extremely young duplicate genes (with d_S ranges between 0 and 0.25) were proportionally more abundant than duplicate genes of other ages. *S. cerevisae* showed the strongest degree of this enrichment: more than 40% of its duplicate genes belong to the extremely young duplicate genes, which may be attributed to the recent genome duplication of this fungus.[39,42] The distributions of d_S also showed that the frequency of slightly older duplicate genes (with $d_S > 0.25$) dropped quickly with the increase of duplication age (ie, d_S value). The similar shape of d_S distributions in all non-*N. crassa* fungi suggested that the frequency of gene duplication, as a basic evolutionary parameter, may be constant in diverse fungal species. Unlike in other fungi, few duplicate genes in *N. crassa* had small d_S, forming the distinct pattern shown in Fig. 3.

The abnormally low number of duplicate genes whose two copies are highly similar to each other in *N. crassa* may be attributed to the influence of the repeat-induced point mutation (RIP), which acts as a defense against selfish and mobile DNA by detecting and mutating both copies of the duplicated sequence.[43,44] It has been proposed that RIP is distributed not only in *Neurospora* but also in other filamentous ascomycetes including *A. fumigatus*, but only one gene so far has been shown to be specifically essential for RIP in *N. crassa*.[45]

Selective pressure and sequence divergence of duplicate genes

To estimate the selective pressure acting on duplicate genes, we computed d_N and d_N/d_S ratio between pairs of duplicate genes (Fig. 3). The ratio d_N/d_S

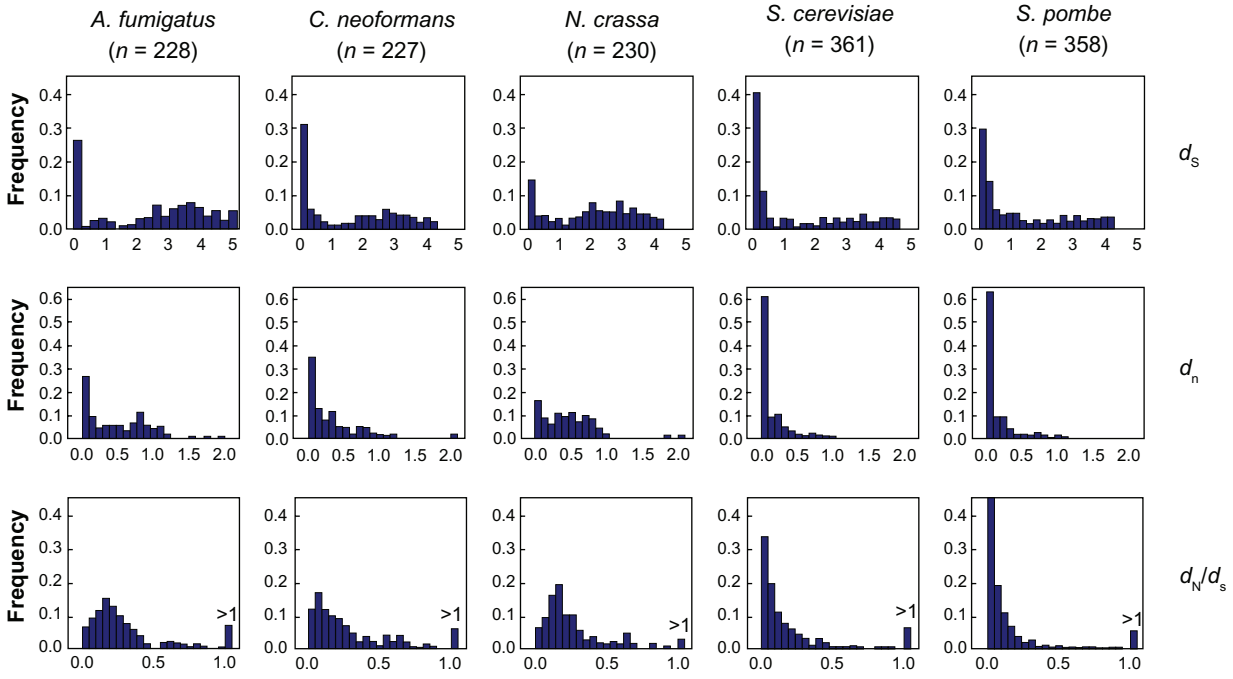

Figure 3. Distributions of synonymous substitution rate, d_S, nonsynonymous substitution rate, d_N, and the ratio between nonsynonymous and synonymous substitution rates, d_N/d_S, versus the frequency of total genes in *A. fumigatus*, *C. neoformans, N. crassa, S. cerevisiae,* and *S. pombe*.
Note: The numbers of gene pairs included in analysis are given in parentheses.

measures the selection pressure to which a gene pair is subjected.[46] Generally speaking, a $d_N/d_S = 1$ indicates that the duplicate genes are under few or no selective constraints. A $d_N/d_S > 1$ is strong evidence for positive selection (ie, replacement substitutions occur at a rate higher than expected by chance, so advantageous mutations have occurred during sequence divergence). In contrast, a $d_N/d_S < 1$ indicates purifying selection (ie, amino-acid replacement substitutions have been purged by natural selection because of their deleterious effects). As shown in Figure 3, all fungal species contained 2% to 7% duplicate genes with $d_N/d_S > 1$, suggesting that positive selection drives sequence divergence of duplicate genes to some extent. Now, considering only gene pairs with $d_N/d_S < 1$, the medians of d_N/d_S of two filamentous fungi, *A. fumigatus* and *N. crassa*, were significantly higher than those of duplicate genes in two yeasts, *S. cerevisiae* and *S. pombe* ($P < 0.001$ in all comparisons between filamentous fungi and yeasts, Kolmogorov-Smirnov [K-S] test). This indicates that purifying selection constraining the sequence divergence between two copies of duplicate genes is more relaxed in the two filamentous fungi than in the two yeasts. A more relaxed pattern of evolution in filamentous fungi, notably *A. fumigatus*, may be related to a recent reduction in the effective population size of the species possibly due to a lowered frequency (or, a loss) of sexual reproduction.[47] A reduced population size would lead to a larger effect produced by genetic drift, which may have allowed additional duplicate copies to be maintained in *A. fumigatus* and to have evolved a subfunction or neofunction in opportunistic pathogenicity where gene copies would ultimately be maintained by positive selection. This pattern in *A. fumigatus* becomes more intriguing when considering that proportionally more duplicate genes in the two unicellular yeasts have extremely small d_S, which is more likely to numerically inflate d_N/d_S values. The dimorphic fungus, *C. neoformans*, showed an intermediate d_N/d_S median. It is worth noting that although comparing the patterns of the distributions of evolutionary parameters gives an impression of the relationships between the fungal species under consideration, these relationships are not necessarily consistent with their evolutionary relationships (such as in Fig. 2) or morphological groups.

In the above analyses, we excluded gene pairs with highly diverged sequences. In order to get a full picture of genetic divergence between duplicate genes, we went back and included those diverged duplicate genes in the analysis. We computed the evolutionary distance (d_{WAG}) between protein sequences of pairs of duplicate genes using the WAG model of protein divergence.[32] We used the same procedure described above to pick pairs of duplicate genes with the smallest d_{WAG} from each gene family. We included all sampled duplicate gene pairs with $d_{WAG} < 2$ and plotted the frequency distribution of pairs of duplicate genes as a function of d_{WAG} (Fig. 4). By using the protein distance, we included nearly twice as many gene pairs in *S. cerevisiae* and at least more than three times more gene pairs in other fungal species than in the previous analysis. In all fungal species, most gene pairs have a d_{WAG} that ranges from 0.5 to 1.5. The distribution of d_{WAG} in the two yeasts showed a bimodal pattern with a main peak at 0.5–1.5 and an extra peak at 0–0.2. The extra peaks of smaller d_{WAG} in yeasts correspond to the same duplicate genes with small d_N and d_N/d_S. The lack of the extra peaks in filamentous fungi suggests that large amounts of diverged duplicated genes are present in their genomes, which may be due to a high retention rate of duplicate genes and weak purifying selection. *C. neoformans* displayed a small d_{WAG}

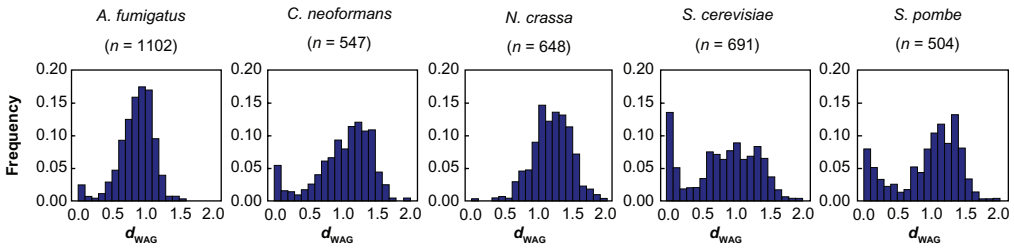

Figure 4. Distributions of evolutionary distance, d_{WAG}, versus frequency of total genes in *A. fumigatus*, *C. neoformans, N. crassa, S. cerevisiae,* and *S. pombe*.
Note: The numbers of gene pairs included in analysis are given in parentheses.

peak, which is higher than those of filamentous fungi but lower than those of yeasts.

Asymmetric evolution of duplicate genes in *A. fumigatus*

To assess the asymmetric evolution between two copies of duplicate genes in *A. fumigatus*, we adapted a codon-based test (Materials and Methods). Codon-based tests take into account the ratio between the rate of nonsynonymous and synonymous substitution and give a more direct measure of the strength of selection and functional constraint in the genes. It is believed that codon-based tests are more sensitive than nucleotide- and amino acid-based tests.[48,49] Among the studied duplicate gene pairs in *A. fumigatus*, 202 pairs had unambiguous orthlogous gene(s) identified in *A. fischeri*, a close relative of *A. fumigatus*.[28]

We used the 202 gene pairs in the test for asymmetric evolution and used the clade model C implemented in Codeml to compute the values of selective pressure ω for both branches of two copies of duplicate genes in *A. fumigatus*.[34] In this branch-site model, the parameter of selective pressure ω is classified into three categories: ω_0, ω_1, and ω_2. Among them, ω_2 is allowed to vary in value during model optimization. Codeml estimates the value for ω_2 and simultaneously assigns the portions of sites that belong to each of the ω classes. We used a likelihood ratio (LR) test to determine the asymmetric evolution between two copies of duplicate genes; we then examined the values of ω_2 at two branches for the two copies of duplicate genes on gene trees (Fig. 1).

The LR test suggested that 24 out of 202 (11.8%) duplicate gene pairs in *A. fumigatus* have a significant P value < 0.05 (χ_2 test), indicating that they evolved at significantly different rates, that is, under asymmetric evolution (Table 3). Among them, 18 gene pairs had at least one copy with $\omega_2 > 1$, and the remaining 6 gene pairs had no copy with $\omega_2 > 1$. These results indicated that a small portion of duplicate genes are under asymmetric evolution and most asymmetrically evolved genes (18/24 = 75%) are driven by positive selection acting on one copy of the duplicate genes. The portion of ω_2 sites ranged widely from 4% to 78.5% among gene pairs (Table 3). Among the 24 gene pairs, 17 pairs remained significant after applying the Bonferroni correction at the 5% level, which suggests that 8.4% gene pairs (17 out of 202) represents a lower boundary for the fraction of duplicate genes evolving by an asymmetric means. Again, most of these highly significant gene pairs (15 out of 17) contain at least one copy of genes with $\omega_2 > 1$.

We hypothesized that asymmetric evolution of duplicate genes is likely to be associated with the functional divergence between two copies of the genes. If true, the difference in gene function-related measures (such as the level of gene expression) between two copies of asymmetrically divergent duplicate genes should be greater than that in non-asymmetrically divergent duplicate genes. To test this, we obtained the gene expression data of *A. fumigatus* in shake cultures from the RNA sequencing (RNA-seq) study by Gibbons et al.[50] We log-transformed the RNA-seq RPKM values and quantile-normalized them across samples, then we computed the absolute value of the difference between two copies of each pair of duplicate genes, $|\Delta e|$. We found that the values of $|\Delta e|$ for the 24 asymmetrically diverged duplicate genes are higher than those for the other 178 pairs of duplicate genes ($P = 0.03$, K-S test). This result suggests that the asymmetric sequence divergence of *A. fumigatus* duplicate genes may be associated with the expression divergence. In most asymmetrically diverged duplicate genes (17 out of 24), the fast evolving copy has a higher expression level compared with the slowly evolving copy. Although this portion is not significant statistically ($P = 0.053$, Fisher's exact one-sided test), the correlated divergence of sequence and expression seems consistent with the theoretical expectation.

Distribution of selection coefficients of duplicate genes in *A. fumigatus*

To estimate the distribution of selection intensities among duplicate genes in *A. fumigatus*, we used the MKPRF test[29,30] to compare the number of synonymous and nonsynonymous polymorphisms within 12 *A. fumigatus* strains and the number of synonymous and nonsynonymous fixed differences between *A. fumigatus* and *A. fischeri*. The MKPRF program uses a Markov chain Monte Carlo algorithm to sample for the posterior distribution of parameters in the models based on Poisson random field (PRF) theory.[29,51,52] For each gene, we used the program to estimate the value of population-effective selection coefficient γ ($=2N_e s$, where N_e is the effective population size and s is the selection coefficient

Table 3. Asymmetrically evolved duplicate gene pairs in *A. fumigatus*. $\omega_{2,1}$ and $\omega_{2,2}$ are variables of the selective pressure, class 2 of d_N/d_S, for two copies of duplicate genes in the branch-site model.

Gene pair	$\omega_{2,1}$	Systematic name	Functional description	$\omega_{2,2}$	Systematic name	Functional description	Portion of ω_2 sites	*P*-value‡
1	>10	Afu1g10880	P-type calcium ATPase	>10	Afu3g10690	Calcium-translocating P-type ATPase	65.6%	<0.0001
2	>10	Afu1g00650†	Extracellular alpha-1,3-glucanase/mutanase	0.0263	Afu7g08510*	Extracellular alpha-1,3-glucanase/mutanase	61.8%	<0.0001
3	>10	Afu4g02720*	GPI anchored glycosyl hydrolase	0.00001	Afu3g00340	Glycosyl hydrolase	5.4%	<0.0001
4	>10	Afu8g04160	Folylpolyglutamate synthetase	0.00001	Afu6g09440	Tetrahydrofolylpolyglutamate synthase	4.0%	<0.0001
5	>10	Afu4g00860	Cell surface protein	0.00001	Afu6g12180	Conserved hypothetical protein	5.4%	<0.0001
6	>10	Afu1g00810*	Allergen Asp F7-like	0.00001	Afu4g06670	Allergen Asp f 7 Precursor	28.2%	<0.0001
7	>10	Afu1g05720	C-14 sterol reductase	0.2819	Afu1g03150	C-14 sterol reductase	4.0%	<0.0001
8	>10	Afu6g02240	ORF, Uncharacterized	0.1749	Afu3g03740*	Protein kinase	9.8%	<0.0001
9	5.7	Afu7g05110	Hypothetical protein	0.00001	Afu6g09410*	Conserved hypothetical protein	15.1%	<0.0001
10	5.1	Afu3g02420	ThiJ/PfpI family transcriptional regulator	0.00001	Afu4g01400	ThiJ/PfpI family protein	10.3%	<0.0001
11	3.4	Afu2g10840†	RTA1 domain protein	0.0183	Afu6g11810†	RTA1 domain protein	31.6%	<0.0001
12	4.1	Afu5g14920*	Hypothetical protein	0.0001	Afu5g07980	Hypothetical protein	19.8%	<0.0001
13	2.0	Afu7g08250*	C6 finger domain protein	0.00001	Afu2g08040	C6 finger domain protein	30.3%	<0.0001
14	1.2	Afu6g09300	C2H2 finger domain protein	0.0001	Afu4g14400†	C2H2 finger domain protein	18.5%	<0.0001
15	1.1	Afu1g01610	Conserved hypothetical protein	0.00001	Afu4g02880	Conserved hypothetical protein	34.5%	<0.001
16	0.0295	Afu5g11240	Oxidoreductase, short chain dehydrogenase/reductase family	0.0001	Afu6g09140	Oxidoreductase, short-chain dehydrogenase/reductase family	15.1%	<0.0001
17	0.0001	Afu3g15280	Methyltransferase	0.0000	Afu2g04380	LaeA-like methyltransferase	78.5%	<0.0001
18	0.0001	Afu5g14930*	Conserved hypothetical protein	0.0000	Afu1g16030*	Conserved hypothetical protein	8.5%	0.002
19	0.0001	Afu5g02410	ATP-dependent RNA helicase fal1	0.0000	Afu3g07200	DEAD/DEAH box RNA helicase	69.2%	0.004
20	0.0001	Afu4g10240	Small nuclear ribonucleoprotein SmD3	0.0000	Afu2g12020*	U6 snRNA-associated Sm-like protein LSM4	53.7%	0.007
21	1.8	Afu8g07150	Arsenic resistance protein ArsH	0.3139	Afu1g16120†	Arsenic resistance protein ArsH, flavoprotein	21.9%	0.009
22	0.2531	Afu6g07320	MFS multidrug transporter	0.1189	Afu6g09110	MFS multidrug transporter	31.1%	0.027
23	2.2	Afu5g14230*	C6 transcription factor	0.00001	Afu1g11290	Transcription regulator	6.2%	0.028
24	>10	Afu5g03930	Alcohol dehydrogenase	0.0001	Afu1g11020†	L-arabinitol 4-dehydrogenase	32.8%	0.047

Notes: The symbol *indicates negatively selected genes identified by using MKPRF; the symbol † indicates positively selected genes identified by using MKPRF. *P* values were obtained by likelihood ratio test between model $\omega_{2,1} = \omega_{2,2}$ and $\omega_{2,1} \neq \omega_{2,2}$ (Materials and Methods). ‡Duplicate genes with $P < 0.0001$ remain significant after Bonferroni correction.

in a Wright-Fisher genic selection model) and the 95% confidence interval (CI) of γ for each gene. If a gene has its 95% CI of γ completely above 0, the gene appears to have been evolving under positive selection. On the other hand, if the 95% CIs of γ are completely below 0, the gene appears to be under negative selection.

The means of γ varied among genes in different families, as well as between two copies of duplicate genes in the same family (Fig. 5). Among *A. fumigatus*' 202 duplicate gene pairs, 38 (18.8%) contained a gene with CI of $\gamma > 0$ indicating positive selection, and 49 (24.2%) contained a gene with CI of $\gamma < 0$ indicating negative selection (supplementary Table S3). For comparison, in the study of human protein-coding genes, MKPRF analysis discovered 304 (9.0%) out of 3377 tested loci showing evidence of positive selection and 813 (13.5%) out of 6033 loci showing evidence of negative selection.[30] Similarly, most of the genes, in either *A. fumigatus* or human, showed no evidence of selection according to MKPRF with a 5% cutoff, indicating weak negative selection and/or balancing selection operating on mutations at these genes. Four pairs contained one gene under negative selection and the other gene under positive selection. Despite the existence of these pairs with gene(s) under selection, the overall correlation between γ values of two copies of duplicate genes was strong and significant (Spearman's $\rho = 0.405$, $P = 3.1 \times 10^{-9}$). This indicates that two copies of duplicate genes had significantly more similar selection intensities than randomly selected pairs of genes overall.

The percentage of genes under positive selection identified using MKPRF was slightly higher than that identified using the codon-based LR test (18.8% versus 11.8%). Interestingly we found inconsistency between the MKPRF estimation of γ and the codon-based estimation of ω_2 among genes. In Table 3, we marked those positively and negatively selected genes ascertained by using MKPRF. Among those asymmetrically evolved gene pairs identified using the LR test, all copies with a smaller ω_2 than the other copies were under negative selection according to the LR test (as indicated by $\omega_2 < 1$). However, there were several cases, in which these negatively selected genes, such as Afu1g11020 and Afu6g11810, were found to be under positive selection using MKPRF. On the other hand, several genes, such as Afu4g02720 and Afu1g00810, with a greater ω_2 than their duplicated copies due to positive selection as indicated by

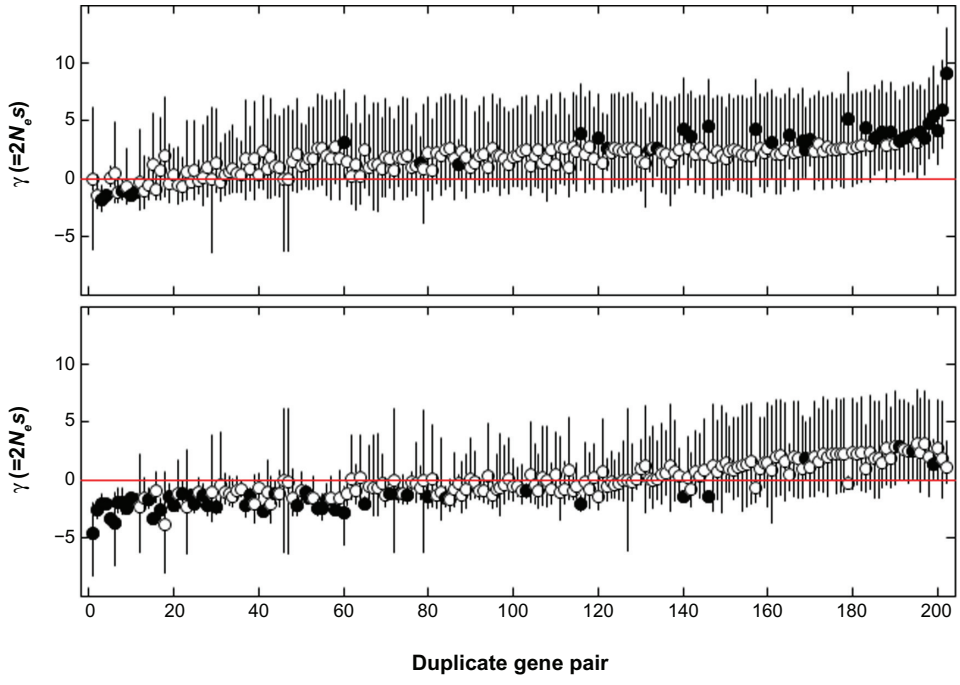

Figure 5. Means of the posterior distributions of the selection coefficient γ (dots) and the 95% CIs (vertical lines) for duplicate genes in *A. fumigatus*.
Notes: For each pair of duplicate genes, the copy with larger γ is plotted in the upper panel and the other copy with smaller γ is plotted in the lower panel. Genes that are under positive (95% CIs of $\gamma > 0$) or negative selection (95% CIs of $\gamma < 0$) are indicated with filled circles. Duplicate gene pairs (x-axis) are ranked by the average values of γ for the two copies of duplicate genes.

$\omega_2 > 1$ were reported to be under negative selection by MKPRF (Table 3).

Discussion

The completion of genome sequencing and the discovery of the sexual cycle in *A. fumigatus* have placed the foundations for the fungus as an emerging model organism for studying the biology, ecology, and pathogenicity of filamentous fungi. Despite these advents, we still know very little about the function of most of *A. fumigatus* genes. Given the slow pace associated with the experimental determination of gene functions, computer-based analysis serves as an initial screening in the characterization of roles of genes. In this study, we systematically assessed the extent of duplicate genes in the genomes of *A. fumigatus* and four other fungal species. We also systematically studied the molecular evolutionary forces associated with the divergence of duplicate genes in *A. fumigatus*. We focused on the role of natural selection on newly created and long-established gene pairs, as well as the role of ongoing selection in shaping nucleotide diversity of duplicate genes in population-genetic samples, which provides insights into the microevolutionary dynamics of these loci.

In examining the extent of gene duplication, we searched for the optimal parameter values for the clustering algorithm we employed. The optimization was performed against sizable sets of gene families that were manually constructed. The optimized values of two key parameters, *L* and *S*, for *A. fumigatus* and *C. neoformans*, were found to be nearly identical. It was, therefore, justified to apply the same criteria determined by the two optimized parameters to all fungal species considered in our study. This guaranteed the clustering of individual proteins into multigene families by a consistent and objective means, making it feasible to compare the size of multigene families across species. Our results revealed complex patterns in differences in the family size distributions among different fungal species. More than a quarter of *A. fumigatus* genes were found to be members of multigene families compared with nearly 30% for baker's yeast *S. cerevisiae* and about 15% for another filamentous fungus *N. crassa*. The other two fungal species, *C. neoformans* and *S. pombe*, showed intermediate ratios. There is no apparent link between the extent of gene duplication in the genomes of species and the features in life style and morphology (eg, unicellar yeasts versus multicellar hypha) of these fungal species.

The extremely high ratio for *S. cerevisiae* may be attributed to the whole genome duplication ca. 10 million years ago.[39,42] The previous comparative genomic analysis on several other yeast species, inducing *Candida glabrata*, *Kluyveromyces lactis*, *Debaryomyces hansenii*, *and Yarrowia lipolytica*, also revealed the influence of other evolutionary mechanisms, tandem gene repeat formation, segmental duplication, and extensive gene loss on the formation of gene duplication patterns.[42] On the other hand, the extremely low ratio for *N. crassa* may be attributed to the strong influence of RIP mutation.[43,44] As mentioned, RIP may be widely present in many other fungal species.[45] However, the impact of RIP on duplicate genes seems more pronounced in *Neurospora* than other fungi in question. Examining the age distribution of duplicate genes gives a sense of the average rate of duplication and the scale of duplication events.[53,54] Our results showed that the genomes of yeasts have been shaped by genome duplication or large-scale gene duplications in the recent evolution, whereas the genome of *A. fumigatus* contains more functionally divergent genes that may be resulted from ancient gene duplications.

Increasing evidence indicates that two copies of duplicate genes can assume unequal roles in divergence.[11] Although the functional significance of asymmetric divergence is still unclear, it has been argued that some form of evolutionary asymmetry is required for functional diversification of duplicate genes.[55] The study of asymmetric evolution of duplicate genes is important for determining the evolutionary processes that have occurred in the genome in order to obtain clues as to the development of the unequal roles of two copies of the same genes. Previous studies on asymmetric evolution between two copies of duplicate genes led to inconsistent conclusions (compare the studies of Kondrashov et al,[10] Hughes and Hughes,[12] and Van de Peer[13] with the study by Conant and Wagner[11]). For example, Kondrashov et al[10] found that no duplicate genes ($n = 15$) in *S. cerevisiae* showed signs of asymmetric evolution, or differential evolutionary rates between duplicate genes, and concluded that both copies of duplicate genes typically evolved at the same rates. Conant and Wagner[11] found

21% ($n = 14$) showed significant signatures of asymmetric evolution in the same species, and suggested asymmetric divergence between two copies of duplicate genes is not uncommon. The discrepancy may be due to the small sample size and the sensitivity of methods used in different studies (eg, Kondrashov et al[10] used a distance-based method, while Conant and Wagner[11] used a codon-based model approach), as well as the way in which outgroup sequences were selected.

We conducted our test in *A. fumigatus* genes using the codon-based LR test with a branch-site model. We used orthologous sequences from the closely related species *A. fischeri* as outgroups. Our LR test identified asymmetric evolution in 12% of duplicate genes in *A. fumigatus*. This result is consistent with that of Conant and Wagner.[11] Our result supports the general conclusion of several previous theoretical and empirical studies,[56–58] indicating that positive selection plays a role in the evolutionary histories of duplicate genes. The fate of duplicate gene pairs may be determined in the initial phases of duplicate gene evolution during the period of reinforced asymmetric evolution. If true, young duplicate genes should be more likely to be driven by positive selection to fixation. In other words, positively selected duplicate genes should be younger. However, our result did not provide evidence supporting this aspect of the theory. Positively selected duplicate genes were not younger than other duplicate genes in *A. fumigatus,* as shown in that the distribution of d_S between two copies of asymmetrically evolved duplicate genes did not differ from that between two copies of randomly selected duplicate genes ($P > 0.05$, K-S test). Our results could suggest that positive selection can occur both in the short period of asymmetric evolution directly following duplication but can also occur after the period of asymmetric evolution, which correlates with the models advocating neofunctionalization, in which one copy develops a novel function, as well as subfunctionalization, in which the copies share the function of the original unduplicated copy.[59]

A close examination of functions of those duplicate genes suggested to be under asymmetric evolution cover a broad range of functionalities, including dehydrogenases, ATPases, glucanases, mutanases, glycosyl hydrolases, as well as genes that encode cell surfaces proteins, arsenic resistance proteins, and transcriptional factors. Recent studies have shown that many of these protein classes are important to the evolved opportunistic pathogenic nature of *A. fumigatus* and are important virulence factors. Kumar et al[60] examined the secretory proteins of *A. fumigatus* showing that glucanases, mutanases, and hydrolases are associated with virulence. Hydrolases have also been shown to be involved in ergot alkaloid synthesis, a complex family of mycotoxins with a variety of pathogenicity functions.[61] However, functions of most of these genes are putatively assigned and need further experimental validation; also many genes encode hypothetical gene products or proteins with unknown function. These again underscore the need of functional analysis of duplicate genes in *A. fumigatus*.

Taking advantage of the availability of sequence polymorphisms ascertained among multiple *A. fumigatus* strains, we applied the MKPRF method, an extension of the MK test, to infer the selection coefficients for individual genes. The MK framework allows examining the levels and patterns of nucleotide polymorphisms and increases the sensitivity of detecting natural selection in protein-coding genes compared with methods using the d_N/d_S ratio alone.[62] It is noteworthy that MKPRF estimates γ also using the divergence between genes and their orthologs but not using the divergence between two copies of duplicate genes. We obtained the selection coefficients γ for individual genes in duplicate gene pairs and found that two copies of duplicate genes have highly correlated γ. This result suggests that two copies of duplicate genes are under selection of largely equal strength. The strength of selection is characteristic of individual gene families, determined by the functions of the families. This is consistent with the conclusion of a previous study that gene duplication (and loss) is highly constrained by the functional properties of genes.[63] MKPRF test also showed that 18.8% of duplicate genes are under positive selection, and 24.2% are under negative selection. These figures are supported by multiple independent repetition of MKPRF with different initial parameters. Strikingly, there are marked discrepancies between the results of MKPRF and LR tests. Many genes that were detected to be under positive selection in one test were not detected or were detected to be under negative selection in the other. We believe that these discrepancies

are most likely rooted from the methodological difference between the MKPRF and the LR test.[35] If there is a biological reason behind the discrepancies, it would suggest a high turnover rate among positive, negative, and neutral selections that act on duplicate genes during different periods of evolution. The two tests happen to be more sensitive to different spectrums of the selective signal generated during the complex evolutionary process. Nevertheless, caution should be taken when interpreting the MKPRF results. MKPRF requires several assumptions in order to apply PRF theory. Some of these assumptions might not hold with our polymorphism data. For example, the sample size ($n = 12$) of *A. fumigatus* strains from which SNPs were ascertained is not large enough to allow low-frequency SNPs to be discovered. Also the influence of gene conversion between two copies of duplicate genes on the results is unclear. In addition, these strains are clinical isolates that may have experienced severe bottlenecks of population size during transmission and strain establishment. Bottleneck-induced drift may result in an elevated rate of fixation of slightly deleterious mutations, which in turn can lead to the biased results of MK test.[64,65] In the future, repeating this analysis using SNPs discovered in more environmental stains of *A. fumigatus* is desired.

In summary, we conducted a systematic examination on the extent of duplicate genes in *A. fumigatus* and showed the difference in the size of multiple gene families between *A. fumigatus* and other fungal species. The established bioinformatics procedure and optimized parameters for the clustering program are ready to be adapted for other studies. We used *A. fumigatus* genes as examples to refine the theories of gene duplication and showed that negative selection contributes to the fixation and persistence of the duplicate genes, while positive selection may also play a role in sequence and functional divergence of duplicate genes.

Acknowledgements

The authors thank Natalie Fedorova, William Nierman, and Paul Bowyer for providing the polymorphism data of *A. fumigatus*, and Suman Pakala for technical assistance. The authors acknowledge the Texas A&M Supercomputing Facility (http://sc.tamu.edu/) for providing computing resources useful in conducting the research reported in this paper.

Author Contributions

Conceived and designed the experiments: JC and EY. Analyzed the data: EY and JC. Wrote the first draft of the manuscript: EY and JC. Contributed to the writing of the manuscript: AH. Agree with manuscript results and conclusions: EY, AH, and JC. Jointly developed the structure and arguments for the paper: AH and JC. Made critical revisions and approved final version: JC, AH, and EY. All authors reviewed and approved of the final manuscript.

Funding

EY is partially supported by CVM Postdoctoral Trainee Research Grant (02-144002-03504) at Texas A&M University.

Competing Interests

Author(s) disclose no potential conflicts of interest.

Disclosures and Ethics

As a requirement of publication author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

References

1. Grant D, Cregan P, Shoemaker RC. Genome organization in dicots: Genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. *P Natl Acad Sci U S A*. 2000;97:4168–73.
2. Sidow A. Gen(om)e duplications in the evolution of early vertebrates. *Curr Opin Genet Dev*. 1996;6:715–22.
3. Wolfe KH, Shields DC. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*. 1997;387:708–13.
4. Ohno S. *Evolution by Gene Duplication*. New York, NY: Springer-Verlag; 1970.
5. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science*. 2000;290:1151–5.
6. Han MV, Demuth JP, McGrath CL, Casola C, Hahn MW. Adaptive evolution of young gene duplicates in mammals. *Genome Res*. 2009;19:859–67.
7. Meyer A, Schartl M. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol*. 1999;11:699–704.

8. Zhang P, Gu Z, Li WH. Different evolutionary patterns between young duplicate genes in the human genome. *Genome Biol*. 2003;4:R56.
9. Dermitzakis ET, Clark AG. Differential selection after duplication in mammalian developmental genes. *Mol Biol Evol*. 2001;18:557–62.
10. Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. Selection in the evolution of gene duplications. *Genome Biol*. 2002;3:RESEARCH0008.
11. Conant GC, Wagner A. Asymmetric sequence divergence of duplicate genes. *Genome Res*. 2003;13:2052–8.
12. Hughes MK, Hughes AL. Evolution of Duplicate Genes in a Tetraploid Animal, *Xenopus laevis*. *Mol Biol Evol*. 1993;10:1360–9.
13. Robinson-Rechavi M, Laudet V. Evolutionary rates of duplicate genes in fish and mammals. *Mol Biol Evol*. 2001;18:681–3.
14. Van de Peer Y, Taylor JS, Braasch I, Meyer A. The ghost of selection past: Rates of evolution and functional divergence of anciently duplicated genes. *J Mol Evol*. 2001;53:436–46.
15. Gaut BS, Zhang LQ, Vision TJ. Patterns of nucleotide substitution among simultaneously duplicated gene pairs in *Arabidopsis thaliana*. *Mol Biol Evol*. 2002;19:1464–73.
16. Li WH, Zhang P, Gu ZL. Different evolutionary patterns between young duplicate genes in the human genome. *Genome Biol*. 2003;4(9):R56.
17. Latge JP. Aspergillus fumigatus and aspergillosis. *Clin Microbiol Rev*. 1999;12:310–50.
18. Segal BH. Medical Progress Aspergillosis. *New Engl J Med*. 2009;360: 1870–84.
19. Ben-Ami R, Lewis RE, Kontoyiannis DP. Enemy of the (immunosuppressed) state: an update on the pathogenesis of *Aspergillus fumigatus* infection. *British Journal of Haematology*. 2010;150:406–17.
20. Hohl TM, Feldmesser M. *Aspergillus fumigatus*: Principles of pathogenesis and host defense. *Eukaryot Cell*. 2007;6:1953–63.
21. Dyer PS, Paoletti M. Reproduction in *Aspergillus fumigatus*: sexuality in a supposedly asexual species? *Med Mycol*. 2005;43:S7–14.
22. O'Gorman CM, Fuller HT, Dyer PS. Discovery of a sexual cycle in the opportunistic fungal pathogen Aspergillus fumigatus. *Nature*. 2009;457: 471–4.
23. Galagan JE, Calvo SE, Borkovich KA, et al. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature*. 2003;422:859–68.
24. Goffeau A, Barrell BG, Bussey H, et al. Life with 6000 genes. *Science*. 1996;274:546, 563–47.
25. Wood V, Gwilliam R, Rajandream MA, et al. The genome sequence of *Schizosaccharomyces pombe*. *Nature*. 2002;415:871–80.
26. Wickes BL, Mayorga ME, Edman U, Edman JC. Dimorphism and haploid fruiting in Cryptococcus neoformans: Association with the alpha-mating type. *P Natl Acad Sci U S A*. 1996;93:7327–31.
27. Nielsen K, Cox GM, Wang P, Toffaletti DL, Heitman J. Sexual cycle of *Cryptococcus neoformans* var. *grubii* and virulence of congenic a and alpha isolates. *Infect Immun*. 2003;71:4831–41.
28. Fedorova ND, Khaldi N, Joardar VS, et al. Genomic islands in the pathogenic filamentous fungus Aspergillus fumigatus. *PLoS Genet*. 2008;4:e1000046.
29. Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD, Hartl DL. The cost of inbreeding in Arabidopsis. *Nature*. 2002;416:531–4.
30. Bustamante CD, Fiedel-Alon A, Williamson S, et al. Natural selection on protein-coding genes in the human genome. *Nature*. 2005;437:1153–7.
31. Yang ZH. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 1997;13:555–6.
32. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. 2001;18:691–9.
33. Cai JJ, Smith DK, Xia X, Yuen KY. MBEToolbox 2.0: An enhanced version of a MATLAB toolbox for Molecular Biology and Evolution. *Evol Bioinform*. 2006;2:179–82.
34. Bielawski JP, Yang Z. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J Mol Evol*. 2004;59:121–32.
35. Li YF, Costello JC, Holloway AK, Hahn MW. "Reverse ecology" and the power of population genomics. *Evolution*. 2008;62:2984–94.
36. Rubin GM, Yandell MD, Wortman JR, et al. Comparative genomics of the eukaryotes. *Science*. 2000;287:2204–15.
37. Hughes AL, Friedman R. Gene duplication and the structure of eukaryotic genomes. *Genome Res*. 2001;11:373–81.
38. Li WH, Gu ZL, Cavalcanti A, Chen FC, Bouman P. Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. *Mol Biol Evol*. 2002;19:256–62.
39. Wolfe KH, Seoighe C. Extent of genomic rearrangement after genome duplication in yeast. *P Natl Acad Sci U S A*. 1998;95:4447–52.
40. Gu Z, Cavalcanti A, Chen FC, Bouman P, Li WH. Extent of gene duplication in the genomes of Drosophila, nematode, and yeast. *Mol Biol Evol*. 2002;19:256–62.
41. Friedman R, Hughes AL. Gene duplication and the structure of eukaryotic genomes. *Genome Res*. 2001;11:373–81.
42. Dujon B, Sherman D, Fischer G, et al. Genome evolution in yeasts. *Nature*. 2004;430:35–44.
43. Nelson MA, Kang S, Braun EL, et al. Expressed sequences from conidial, mycelial, and sexual stages of Neurospora crassa. *Fungal Genet Biol*. 1997; 21:348–63.
44. Galagan JE, Calvo SE, Borkovich KA, et al. The genome sequence of the filamentous fungus Neurospora crassa. *Nature*. 2003;422:859–68.
45. Clutterbuck AJ. Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. *Fungal Genet Biol*. 2011;48:306–26.
46. Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol*. 2000;15:496–503.
47. Paoletti M, Rydholm C, Schwier EU, et al. Evidence for sexuality in the opportunistic fungal pathogen Aspergillus fumigatus. *Curr Biol*. 2005;15: 1242–8.
48. Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*. 1994;11:725–36.
49. Muse SV, Gaut BS. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol*. 1994;11:715–24.
50. Gibbons JG, Beauvais A, Beau R, McGary KL, Latgé JP, Rokas A. Global transcriptome changes underlying colony growth in the opportunistic human pathogen Aspergillus fumigatus. *Eukaryot Cell*. 2012;11:68–78.
51. Barrier M, Bustamante CD, Yu JY, Purugganan MD. Selection on rapidly evolving proteins in the *Arabidopsis* genome. *Genetics*. 2003;163: 723–33.
52. Sawyer SA, Hartl DL. Population genetics of polymorphism and divergence. *Genetics*. 1992;132:1161–76.
53. Lynch M, Conery JS. The evolutionary demography of duplicate genes. *Journal of Structural and Functional Genomics*. 2003;3:35–44.
54. Blanc G, Wolfe KH. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *The Plant Cell*. 2004;16: 1667–78.
55. Krakaue DC, Nowak MA. Evolutionary preservation of redundant duplicated genes. *Seminars in Cell and Developmental Biology*. 1999;10: 555–9.
56. Han MV, Demuth JP, McGrath CL, Casola C, Hahn MW. Adaptive evolution of young gene duplicates in mammals. *Genome Res*. 2009;19:859–67.
57. Hahn MW. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J Hered*. 2009;100:605–17.
58. Kondrashov FA. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc R Soc B*. Sep 2012. [Epub ahead of print.]
59. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*. 1999;151:1531–45.
60. Kumar A, Ahmed R, Singh PK, Shukla PK. Identification of virulence factors and diagnostic markers using immunosecretome of Aspergillus fumigatus. *J Proteomics*. 2011;74:1104–12.
61. Coyle CM, Panaccione DG. An ergot alkaloid biosynthesis gene and clustered hypothetical genes from Aspergillus fumigatus. *Appl Environ Microb*. 2005;71:3112–8.
62. Kryazhimskiy S, Plotkin JB. The population genetics of dN/dS. *PLoS Genet*. 2008;4:e1000304.
63. Wapinski I, Pfeffer A, Friedman N, Regev A. Natural history and evolutionary principles of gene duplication in fungi. *Nature*. 2007;449:54–61.

64. Eyre-Walker A, Keightley PD. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol Biol Evol*. 2009;26:2097–108.
65. Charlesworth J, Eyre-Walker A. The McDonald-Kreitman test and slightly deleterious mutations. *Mol Biol Evol*. 2008;25:1007–15.
66. Nierman WC, Pain A, Anderson MJ, et al. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature*. 2005;438:1151–6.
67. James TY, Kauff F, Schoch CL, et al. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*. 2006;443:818–22.

Supplementary Data

Table S1. Lists of gene families in *A. fumigatus* used for parameter optimization of BLASTCLUST.

Family	No of Genes	Gene ID
1	21	Afu8g06290, Afu2g18080, Afu8g01040, Afu8g00310, Afu6g00780, Afu3g09430, Afu4g02640, Afu4g14870, Afu6g00790, Afu6g09470, Afu6g14660, Afu7g08200, Afu4g00820, Afu5g00270, Afu6g09480, Afu6g14650, Afu3g09440, Afu4g14880, Afu3g15360, Afu4g14370, Afu7g06935
2	10	Afu6g09480, Afu6g14650, Afu2g18080, Afu8g01040, Afu8g00310, Afu8g06290, Afu3g09440, Afu4g14880, Afu3g15360, Afu5g00270
3	9	Afu6g03030, Afu2g01580, Afu6g03490, Afu2g15520, Afu6g11480, Afu1g13660, Afu2g00140, Afu3g02620, Afu3g03280
4	7	Afu7g04180, Afu3g00680, Afu1g13440, Afu3g14590, Afu5g07360, Afu5g01470, Afu7g08470
5	6	Afu7g04080, Afu4g10950, Afu1g12650, Afu2g11350, Afu6g14200, Afu8g04000
6	5	Afu2g00530, Afu3g07550, Afu5g13060, Afu8g01120, Afu3g00670
7	5	Afu2g11610, Afu2g11620, Afu7g06380, Afu8g07070, Afu3g07380
8	5	Afu2g00710, Afu4g10130, Afu2g03230, Afu3g00900, Afu2g13460
9	5	Afu8g07260, Afu2g17320, Afu3g07220, Afu3g00170, Afu6g00520
10	5	Afu1g10910, Afu7g00250, Afu1g02550, Afu2g14990, Afu1g13390
11	4	Afu1g11350, Afu5g09940, Afu3g03500, Afu4g01360
12	4	Afu6g11890, Afu4g00540, Afu3g13580, Afu7g08580
13	4	Afu1g14170, Afu3g00380, Afu1g16700, Afu6g06660
14	4	Afu3g01440, Afu4g03340, Afu5g07620, Afu3g13940
15	4	Afu1g10790, Afu7g04720, Afu6g13760, Afu5g10520
16	4	Afu6g03890, Afu3g02270, Afu2g18030, Afu2g00200
17	4	Afu5g00410, Afu3g00240, Afu2g15490, Afu5g00380
18	4	Afu6g12950, Afu2g04010, Afu4g03190, Afu5g14300
19	4	Afu1g12400, Afu5g03110, Afu1g12190, Afu5g02850
20	4	Afu4g00730, Afu8g01530, Afu2g14320, Afu3g13130
21	3	Afu2g11270, Afu3g00910, Afu1g15440
22	3	Afu3g10910, Afu3g07030, Afu6g09910
23	3	Afu1g17570, Afu1g04870, Afu6g03200
24	3	Afu2g05880, Afu1g10930, Afu5g11020
25	3	Afu8g00770, Afu7g05550, Afu2g09450
26	3	Afu2g09520, Afu6g11600, Afu6g07480
27	3	Afu3g01170, Afu7g01590, Afu5g13810
28	3	Afu7g00420, Afu8g06100, Afu8g06590
29	3	Afu2g00120, Afu2g03820, Afu2g00540
30	3	Afu6g00500, Afu8g00930, Afu4g01290
31	2	Afu3g07860, Afu4g14070
32	2	Afu1g04780, Afu1g12910
33	2	Afu2g07940, Afu6g10990
34	2	Afu6g07640, Afu1g04460
35	2	Afu1g11290, Afu5g14230
36	2	Afu7g04800, Afu2g12640
37	2	Afu8g06080, Afu4g03410
38	2	Afu2g00920, Afu2g12770
39	2	Afu7g05740, Afu6g05210

(Continued)

Table S1 (*Continued*)

Family	No of Genes	Gene ID
40	2	Afu1g12050, Afu4g09540
41	1	Afu1g01350
42	1	Afu4g01470
43	1	Afu5g04090
44	1	Afu2g04580
45	1	Afu1g06000
46	1	Afu5g06690
47	1	Afu5g08030
48	1	Afu2g11880
49	1	Afu3g12050
50	1	Afu1g14980

Table S2. Lists of gene families in *C. neoformans* used for parameter optimization of BLASTCLUST.

Family	No of Genes	Gene ID
1	9	CNAG_00792T0, CNAG_00823T0, CNAG_01575T0, CNAG_02262T0, CNAG_02430T0, CNAG_02977T0, CNAG_03450T0, CNAG_03503T0, CNAG_07781T0
2	7	CNAG_02883T0, CNAG_03130T0, CNAG_03315T0, CNAG_05348T0, CNAG_05968T0, CNAG_05998T0, CNAG_06606T0
3	6	CNAG_00099T0, CNAG_03341T0, CNAG_03962T0, CNAG_04052T0, CNAG_05825T0, CNAG_06182T0
4	6	CNAG_00550T0, CNAG_00770T0, CNAG_01642T0, CNAG_01916T0, CNAG_02682T0, CNAG_05201T0
5	6	CNAG_00859T0, CNAG_02018T0, CNAG_06944T0, CNAG_07002T0, CNAG_07753T0, CNAG_07893T0
6	6	CNAG_01495T0, CNAG_05320T0, CNAG_05321T0, CNAG_05329T0, CNAG_06530T0, CNAG_07626T0
7	5	CNAG_00308T0, CNAG_03420T0, CNAG_03960T0, CNAG_04141T0, CNAG_04988T0
8	5	CNAG_01500T0, CNAG_01714T0, CNAG_03389T0, CNAG_06249T0, CNAG_06876T0
9	5	CNAG_04474T0, CNAG_06536T0, CNAG_06537T0, CNAG_06538T0, CNAG_06539T0
10	5	CNAG_05369T0, CNAG_06931T0, CNAG_06936T0, CNAG_06985T0, CNAG_07707T0
11	4	CNAG_00122T0, CNAG_01940T0, CNAG_02189T0, CNAG_05264T0
12	4	CNAG_00575T0, CNAG_04981T0, CNAG_05015T0, CNAG_05256T0
13	4	CNAG_00789T0, CNAG_00980T0, CNAG_03477T0, CNAG_05911T0
14	4	CNAG_00862T0, CNAG_05319T0, CNAG_05376T0, CNAG_05383T0
15	4	CNAG_00863T0, CNAG_02475T0, CNAG_04561T0, CNAG_07389T0
16	4	CNAG_01373T0, CNAG_01740T0, CNAG_02196T0, CNAG_05925T0
17	4	CNAG_01968T0, CNAG_03797T0, CNAG_07447T0, CNAG_07613T0
18	4	CNAG_01840T0, CNAG_03787T0, CNAG_04948T0, CNAG_06914T0
19	4	CNAG_02217T0, CNAG_03326T0, CNAG_06487T0, CNAG_07499T0
20	4	CNAG_05276T0, CNAG_01699T0, CNAG_05690T0, CNAG_05563T0
21	3	CNAG_00063T0, CNAG_04828T0, CNAG_06745T0
22	3	CNAG_00919T0, CNAG_01040T0, CNAG_02966T0
23	3	CNAG_01635T0, CNAG_02016T0, CNAG_04225T0
24	3	CNAG_01681T0, CNAG_04276T0, CNAG_04982T0
25	3	CNAG_02552T0, CNAG_06172T0, CNAG_07445T0

(*Continued*)

Table S2 (*Continued*)

Family	No of Genes	Gene ID
26	3	CNAG_02958T0, CNAG_06241T0, CNAG_07865T0
27	3	CNAG_03277T0, CNAG_05316T0, CNAG_06623T0
28	3	CNAG_04326T0, CNAG_06374T0, CNAG_06638T0
29	3	CNAG_05937T0, CNAG_05941T0, CNAG_07703T0
30	3	CNAG_06297T0, CNAG_06298T0, CNAG_06388T0
31	2	CNAG_01487T0, CNAG_01511T0
32	2	CNAG_00005T0, CNAG_02012T0
33	2	CNAG_02086T0, CNAG_02087T0
34	2	CNAG_02241T0, CNAG_05453T0
35	2	CNAG_02899T0, CNAG_03007T0
36	2	CNAG_02896T0, CNAG_03311T0
37	2	CNAG_03355T0, CNAG_05590T0
38	2	CNAG_05450T0, CNAG_05454T0
39	2	CNAG_06010T0, CNAG_06018T0
40	2	CNAG_06524T0, CNAG_06821T0
41	1	CNAG_01563T0
42	1	CNAG_03763T0
43	1	CNAG_04049T0
44	1	CNAG_04610T0
45	1	CNAG_05721T0
46	1	CNAG_06603T0
47	1	CNAG_06620T0
48	1	CNAG_06828T0
49	1	CNAG_07687T0
50	1	CNAG_07766T0

Table S3. Gamma and 59% CIs of MKPRF results for 202 duplicate gene pairs in *A. fumigatus*.

Gene 1			Gene 2		
Gene ID	gamma	95% CI	Gene ID	gamma	95% CI
Afu2g12020	–4.66	[1.64, –8.35]	Afu4g10240	–0.09	[3.15, –6.15]
Afu5g04060	–3.80	[2.14, –8.08]	Afu3g12850	1.95	[2.25, –1.52]
Afu5g02180	–3.66	[1.54, –7.36]	Afu2g07620	0.41	[1.80, –2.04]
Afu5g14920	–3.39	[0.15, –3.69]	Afu5g07980	1.23	[1.84, –1.33]
Afu1g00350	–3.38	[0.25, –3.90]	Afu8g06210	0.05	[0.45, –0.67]
Afu5g05510	–2.88	[1.26, –5.66]	Afu7g03750	3.19	[2.02, 0.03]
Afu6g11710	–2.65	[0.28, –3.19]	Afu7g08440	2.37	[2.11, –0.90]
Afu3g02800	–2.55	[0.46, –3.42]	Afu6g02480	0.67	[1.82, –1.82]
Afu2g17900	–2.53	[0.41, –3.30]	Afu2g00160	–1.43	[0.77, –2.58]
Afu1g16080	–2.53	[0.27, –3.04]	Afu5g14990	2.78	[2.06, –0.45]
Afu5g00280	–2.43	[0.27, –2.94]	Afu2g05190	–0.61	[1.22, –2.18]
Afu7g01690	–2.41	[0.35, –3.07]	Afu8g05220	2.61	[2.09, –0.65]
Afu3g00670	–2.39	[0.59, –3.44]	Afu5g13060	2.60	[2.05, –0.60]
Afu6g12820	–2.37	[2.06, –6.42]	Afu2g12200	0.65	[1.79, –1.77]
Afu6g02420	–2.35	[2.02, –6.26]	Afu6g13170	–0.25	[1.80, –2.65]
Afu2g17430	–2.33	[0.37, –3.01]	Afu5g09600	1.36	[1.89, –1.24]
Afu6g09410	–2.17	[0.16, –2.47]	Afu7g05110	0.38	[0.89, –0.87]
Afu4g08240	–2.17	[0.58, –3.18]	Afu2g13270	1.75	[2.26, –1.84]

(*Continued*)

Table S3 (*Continued*)

Gene 1			Gene 2		
Gene ID	gamma	95% CI	Gene ID	gamma	95% CI
Afu2g00920	–2.16	[0.49, –3.04]	Afu2g12770	2.10	[2.14, –1.22]
Afu1g17580	–2.15	[0.48, –3.02]	Afu3g15040	0.98	[1.85, –1.53]
Afu2g08910	–2.10	[0.44, –2.88]	Afu1g02350	–1.82	[0.59, –2.78]
Afu8g06580	–2.10	[0.37, –2.77]	Afu8g04910	0.66	[1.80, –1.76]
Afu1g16030	–2.09	[0.14, –2.35]	Afu5g14930	–1.39	[0.20, –1.76]
Afu8g06630	–2.07	[0.34, –2.68]	Afu1g16115	3.90	[1.96, 0.73]
Afu6g14200	–2.06	[1.04, –3.71]	Afu1g12650	1.87	[2.25, –1.64]
Afu8g01400	–2.05	[0.31, –2.61]	Afu1g09980	2.48	[2.12, –0.80]
Afu5g09580	–2.04	[1.04, –3.71]	Afu1g17250	1.71	[2.24, –1.77]
Afu4g00570	–1.99	[0.57, –2.96]	Afu3g02060	–1.09	[0.46, –1.83]
Afu1g03280	–1.91	[0.58, –2.88]	Afu6g11920	–1.18	[0.56, –2.04]
Afu2g17360	–1.70	[0.59, –2.63]	Afu4g02750	1.93	[2.18, –1.48]
Afu5g11080	–1.67	[0.45, –2.42]	Afu2g15140	–0.57	[0.97, –1.86]
Afu5g00370	–1.66	[0.17, –1.97]	Afu5g00930	–1.14	[0.42, –1.83]
Afu8g04110	–1.63	[0.75, –2.76]	Afu1g00440	2.58	[2.10, –0.70]
Afu5g02870	–1.60	[1.06, –3.13]	Afu5g09290	1.47	[1.83, –1.11]
Afu5g14230	–1.59	[0.30, –2.13]	Afu1g11290	1.77	[1.83, –0.88]
Afu8g01560	–1.55	[0.74, –2.68]	Afu2g11250	–1.03	[1.17, –2.58]
Afu2g13020	–1.53	[0.78, –2.68]	Afu6g07710	1.77	[2.25, –1.78]
Afu6g09370	–1.52	[0.14, –1.78]	Afu1g00150	–1.41	[0.13, –1.66]
Afu4g02720	–1.52	[0.45, –2.27]	Afu3g00340	2.45	[2.13, –0.91]
Afu7g06620	–1.50	[0.77, –2.64]	Afu4g00600	1.70	[1.86, –0.89]
Afu3g12770	–1.50	[0.98, –2.89]	Afu1g14210	1.78	[2.25, –1.78]
Afu1g00950	–1.49	[0.19, –1.84]	Afu7g07090	4.31	[1.93, 1.24]
Afu4g10000	–1.48	[0.52, –2.33]	Afu2g03090	–0.37	[1.29, –2.02]
Afu5g03930	–1.46	[0.76, –2.61]	Afu1g11020	3.52	[1.66, 0.93]
Afu5g01230	–1.45	[0.62, –2.40]	Afu3g01030	2.26	[2.16, –1.11]
Afu7g08510	–1.42	[0.17, –1.73]	Afu1g00650	4.47	[1.85, 1.48]
Afu4g14510	–1.37	[0.63, –2.37]	Afu2g11120	0.81	[1.51, –1.17]
Afu3g10960	–1.37	[0.76, –2.49]	Afu2g03670	2.21	[1.83, –0.47]
Afu1g16040	–1.36	[0.15, –1.63]	Afu5g14940	–0.33	[0.30, –0.84]
Afu8g06640	–1.34	[0.47, –2.11]	Afu3g02400	1.99	[2.23, –1.50]
Afu6g10820	–1.34	[0.81, –2.48]	Afu2g17560	2.58	[2.09, –0.67]
Afu8g02500	–1.33	[0.64, –2.31]	Afu1g17010	1.91	[2.22, –1.58]
Afu1g01670	–1.32	[0.47, –2.08]	Afu7g01980	0.11	[1.84, –2.33]
Afu3g11480	–1.31	[1.12, –2.81]	Afu3g07560	2.32	[2.15, –1.03]
Afu8g06930	–1.27	[0.26, –1.73]	Afu6g12160	0.88	[1.45, –1.09]
Afu1g01260	–1.24	[0.55, –2.09]	Afu5g09470	1.07	[1.82, –1.45]
Afu6g02790	–1.24	[0.66, –2.23]	Afu3g13680	1.55	[1.59, –0.62]
Afu2g00880	–1.24	[0.44, –1.95]	Afu1g12450	1.87	[1.84, –0.72]
Afu7g08250	–1.19	[0.21, –1.58]	Afu2g08040	1.79	[1.81, –0.81]
Afu5g01040	–1.18	[1.12, –2.70]	Afu6g09970	–0.61	[1.25, –2.20]
Afu5g01160	–1.17	[0.65, –2.17]	Afu8g01180	0.57	[1.44, –1.32]
Afu8g04470	–1.16	[0.66, –2.13]	Afu6g11320	1.79	[2.31, –1.80]
Afu7g00390	–1.13	[0.49, –1.91]	Afu3g01120	–0.66	[0.98, –1.94]
Afu5g00980	–1.13	[0.82, –2.31]	Afu1g10370	0.97	[1.82, –1.53]
Afu7g06660	–1.11	[0.56, –1.99]	Afu6g13800	–0.20	[0.76, –1.27]
Afu7g06750	–1.10	[1.13, –2.61]	Afu1g06590	0.83	[1.79, –1.63]

(*Continued*)

Table S3 *(Continued)*

Gene 1			Gene 2		
Gene ID	gamma	95% CI	Gene ID	gamma	95% CI
Afu6g03980	–1.06	[0.67, –2.05]	Afu2g16330	0.35	[1.37, –1.47]
Afu3g03740	–1.05	[0.30, –1.57]	Afu6g02240	1.17	[1.35, –0.63]
Afu5g01440	–1.01	[1.81, –3.52]	Afu6g12500	1.94	[2.22, –1.50]
Afu6g09140	–1.01	[1.82, –3.52]	Afu5g11240	2.80	[2.07,–0.45]
Afu6g04790	–1.00	[1.85, –3.53]	Afu5g02150	–0.07	[3.18, –6.33]
Afu3g13650	–0.98	[0.69, –2.00]	Afu8g05805	–0.94	[0.59, –1.85]
Afu4g02700	–0.97	[0.69, –2.00]	Afu4g13750	1.05	[1.50, –0.99]
Afu1g00810	–0.97	[0.31, –1.50]	Afu4g06670	2.54	[2.07, –0.68]
Afu6g00410	–0.94	[0.53, –1.76]	Afu7g00440	0.44	[1.20, –1.12]
Afu7g01920	–0.94	[0.71, –1.97]	Afu1g02460	2.15	[2.19, –1.21]
Afu7g07000	–0.90	[0.52, –1.73]	Afu3g08140	2.73	[2.05, –0.47]
Afu8g01190	–0.89	[0.48, –1.66]	Afu1g17630	2.05	[1.80, –0.57]
Afu5g09100	–0.88	[0.90, –2.10]	Afu1g12940	1.24	[2.43, –2.75]
Afu6g09850	–0.87	[0.87, –2.09]	Afu6g10310	2.37	[2.12, –0.96]
Afu8g04080	–0.87	[1.18, –2.43]	Afu1g00470	2.49	[2.10, –0.79]
Afu4g08410	–0.84	[1.22, –2.41]	Afu1g13280	1.86	[2.26, –1.68]
Afu7g00300	–0.83	[0.38, –1.48]	Afu7g00260	0.34	[1.14, –1.19]
Afu4g14830	–0.81	[0.90, –2.04]	Afu4g14810	2.53	[2.12, –0.76]
Afu8g06040	–0.81	[0.91, –2.05]	Afu2g02390	2.60	[2.01, –0.54]
Afu3g12790	–0.80	[1.21, –2.35]	Afu4g11390	1.75	[2.26, –1.85]
Afu2g07710	–0.80	[1.21, –2.39]	Afu1g09240	2.76	[2.08, –0.45]
Afu4g01400	–0.78	[1.20, –2.34]	Afu3g02420	2.03	[1.84, –0.63]
Afu2g12790	–0.75	[1.20, –2.34]	Afu6g09880	3.65	[1.70, 1.01]
Afu1g09930	–0.71	[1.79, –3.13]	Afu6g10260	1.96	[2.20, –1.41]
Afu5g01550	–0.69	[0.56, –1.54]	Afu4g14720	0.38	[0.80, –0.74]
Afu7g01470	–0.69	[1.81, –3.08]	Afu1g10340	1.25	[2.46, –2.73]
Afu6g10480	–0.63	[1.77, –3.02]	Afu8g03890	1.20	[2.43, –2.79]
Afu5g01210	–0.62	[0.75, –1.70]	Afu3g02600	1.39	[0.93, 0.03]
Afu3g03310	–0.62	[0.97, –1.91]	Afu5g05640	4.22	[1.95, 1.07]
Afu1g03150	–0.56	[0.96, –1.83]	Afu1g05720	1.93	[1.78, –0.69]
Afu8g02310	–0.54	[0.96, –1.85]	Afu4g01550	2.63	[1.22, 0.78]
Afu3g01610	–0.52	[1.29, –2.15]	Afu4g09300	1.02	[1.83, –1.46]
Afu1g15120	–0.50	[0.67, –1.48]	Afu6g12910	1.98	[1.81, –0.62]
Afu3g00450	–0.49	[0.80, –1.61]	Afu1g12460	2.33	[2.14, –1.01]
Afu2g07550	–0.47	[1.79, –2.85]	Afu1g13600	–0.27	[1.31, –1.95]
Afu6g09440	–0.44	[1.29, –2.06]	Afu8g04160	1.79	[1.84, –0.82]
Afu5g02410	–0.36	[1.28, –2.02]	Afu3g07200	1.96	[1.83, –0.67]
Afu2g16320	–0.36	[0.82, –1.5]	Afu4g06050	2.59	[2.09, –0.65]
Afu7g08290	–0.31	[0.35, –0.89]	Afu7g07030	1.28	[0.87, 0.02]
Afu5g09950	–0.30	[1.01, –1.66]	Afu1g10900	0.87	[1.83, –1.59]
Afu7g05650	–0.30	[1.32, –1.99]	Afu7g04870	1.06	[1.85, –1.46]
Afu8g00680	–0.26	[1.02, –1.64]	Afu3g03620	2.51	[2.16, –0.81]
Afu6g09300	–0.24	[0.26, –0.69]	Afu4g14400	5.15	[1.75, 2.31]
Afu6g00270	–0.22	[1.33, –1.91]	Afu1g11260	2.14	[2.18, –1.23]
Afu5g01340	–0.21	[1.02, –1.57]	Afu4g08720	1.35	[1.83, –1.23]
Afu5g00460	–0.19	[0.41, –0.86]	Afu6g13390	0.96	[1.50, –1.06]
Afu4g03550	–0.19	[1.04, –1.58]	Afu6g04370	2.28	[2.20, –1.14]
Afu8g02530	–0.17	[1.30, –1.86]	Afu4g04010	2.68	[2.09, –0.58]

(*Continued*)

Table S3 *(Continued)*

Gene 1			Gene 2		
Gene ID	gamma	95% CI	Gene ID	gamma	95% CI
Afu4g04810	–0.13	[3.17, –6.33]	Afu2g15570	–0.02	[3.17, –6.24]
Afu1g13780	–0.09	[3.16, –6.21]	Afu2g13860	–0.07	[3.16, –6.21]
Afu1g04950	–0.08	[3.15, –6.19]	Afu5g06700	2.59	[2.12, –0.68]
Afu5g06580	–0.07	[1.32, –1.79]	Afu4g11780	2.39	[2.11, –0.89]
Afu7g05950	–0.06	[3.17, –6.21]	Afu6g14240	0.85	[2.64, –3.79]
Afu6g03520	–0.05	[3.19, –6.20]	Afu2g01400	0.67	[1.84, –1.79]
Afu5g09210	–0.04	[1.82, –2.44]	Afu7g04930	2.66	[2.14, –0.66]
Afu5g11430	0.02	[1.81, –2.36]	Afu1g02090	1.71	[2.32, –1.93]
Afu3g08960	0.07	[1.31, –1.68]	Afu4g03260	2.60	[1.61, 0.19]
Afu4g01040	0.09	[1.41, –1.70]	Afu6g02990	0.22	[1.82, –2.22]
Afu3g11280	0.10	[1.83, –2.33]	Afu8g01410	0.73	[1.84, –1.77]
Afu3g02780	0.12	[1.36, –1.66]	Afu5g00420	0.95	[0.70, –0.10]
Afu2g11420	0.12	[1.34, –1.64]	Afu2g12500	1.25	[1.83, –1.29]
Afu4g08230	0.14	[0.57, –0.74]	Afu2g02950	2.37	[1.81, –0.32]
Afu4g11060	0.15	[1.37, –1.61]	Afu3g15380	2.44	[1.24, 0.62]
Afu5g00760	0.16	[1.37, –1.59]	Afu4g04180	2.67	[2.07, –0.60]
Afu8g00410	0.17	[1.43, –1.66]	Afu2g01750	0.26	[1.81, –2.17]
Afu2g01170	0.23	[1.80, –2.19]	Afu2g05340	1.41	[1.80, –1.14]
Afu3g10690	0.25	[1.14, –1.26]	Afu1g10880	1.74	[1.83, –0.88]
Afu8g05750	0.29	[1.00, –1.05]	Afu3g01040	2.51	[2.14, –0.84]
Afu4g03460	0.31	[1.81, –2.11]	Afu1g17060	0.92	[1.56, –1.11]
Afu6g03320	0.33	[1.42, –1.52]	Afu1g12620	2.60	[2.07, –0.61]
Afu6g11560	0.44	[1.82, –2.01]	Afu2g15440	2.24	[2.18, –1.18]
Afu2g04070	0.49	[1.02, –0.91]	Afu6g03720	3.75	[1.81, 0.91]
Afu4g00860	0.50	[1.45, –1.40]	Afu6g12180	1.26	[1.78, –1.29]
Afu2g04380	0.51	[1.81, –1.90]	Afu3g15280	1.08	[1.81, –1.43]
Afu3g00920	0.55	[1.44, –1.33]	Afu6g14140	2.13	[2.20, –1.28]
Afu4g03321	0.66	[1.82, –1.82]	Afu7g06080	2.10	[2.18, –1.29]
Afu2g10140	0.72	[1.46, –1.23]	Afu3g14010	2.18	[2.16, –1.22]
Afu4g00800	0.74	[1.80, –1.71]	Afu4g00990	2.37	[2.12, –0.95]
Afu5g06230	0.76	[1.82, –1.74]	Afu4g03370	1.32	[1.81, –1.21]
Afu2g11600	0.82	[1.82, –1.65]	Afu1g13320	0.99	[1.83, –1.53]
Afu5g13290	0.82	[1.82, –1.68]	Afu2g15730	2.52	[1.86, –0.18]
Afu3g11560	0.88	[1.79, –1.58]	Afu2g13050	2.12	[2.21, –1.32]
Afu6g02630	0.89	[2.64, –3.68]	Afu1g15620	3.14	[2.00, 0.06]
Afu6g07320	0.95	[1.47, –1.04]	Afu6g09110	2.39	[2.13, –0.91]
Afu6g08550	0.95	[1.49, –1.05]	Afu6g00120	4.37	[1.50, 1.96]
Afu4g11890	0.99	[1.82, –1.47]	Afu6g02840	2.62	[2.11, –0.68]
Afu8g07150	0.99	[1.84, –1.52]	Afu1g16120	3.28	[2.01, 0.14]
Afu8g05740	1.04	[1.01, –0.39]	Afu2g00420	1.48	[1.84, –1.07]
Afu5g10180	1.05	[1.79, –1.46]	Afu4g01530	3.34	[1.52, 1.07]
Afu4g01340	1.06	[1.81, –1.38]	Afu1g11830	2.33	[2.15, –1.05]
Afu8g02350	1.07	[0.95, –0.29]	Afu3g02570	9.08	[1.86, 5.79]
Afu5g05480	1.20	[2.41, –2.77]	Afu3g10740	1.99	[2.22, –1.47]
Afu2g10100	1.21	[2.46, –2.83]	Afu1g06830	1.36	[2.36, –2.45]
Afu3g01340	1.24	[2.45, –2.71]	Afu2g15290	1.44	[2.37, –2.37]
Afu5g11230	1.25	[2.46, –2.74]	Afu5g12130	1.83	[2.25, –1.71]
Afu8g04650	1.27	[2.43, –2.75]	Afu2g17450	2.14	[2.17, –1.22]
Afu8g06160	1.35	[0.89, 0.06]	Afu1g01050	5.42	[1.96, 2.18]
Afu6g05040	1.41	[2.34, –2.37]	Afu5g01870	2.05	[2.19, –1.37]
Afu1g14380	1.45	[2.37, –2.34]	Afu7g00840	2.58	[2.11, –0.66]
Afu4g12010	1.46	[1.53, –0.66]	Afu1g00490	4.07	[1.97, 0.91]

(*Continued*)

Table 3 (*Continued*)

Gene 1			Gene 2		
Gene ID	gamma	95% CI	Gene ID	gamma	95% CI
Afu4g13310	1.49	[1.57, –0.69]	Afu5g12770	2.20	[2.17, –1.20]
Afu1g00540	1.54	[1.78, –1.00]	Afu8g04060	1.74	[1.85, –0.86]
Afu2g10920	1.57	[2.32, –2.09]	Afu3g03410	1.95	[2.24, –1.52]
Afu8g04070	1.66	[1.85, –0.92]	Afu1g00480	2.98	[2.01, –0.19]
Afu3g07830	1.67	[2.27, –1.88]	Afu1g06710	2.55	[2.08, –0.69]
Afu2g10840	1.81	[1.29, 0.01]	Afu6g11810	2.47	[1.61, 0.10]
Afu4g12050	1.84	[2.22, –1.62]	Afu1g00530	3.55	[1.99, 0.39]
Afu4g09700	1.89	[2.21, –1.61]	Afu6g07000	3.57	[1.84, 0.71]
Afu8g04090	1.91	[2.22, –1.58]	Afu1g00460	2.36	[2.16, –0.99]
Afu8g04020	1.93	[1.85, –0.71]	Afu1g00610	3.99	[1.95, 0.88]
Afu3g01560	1.93	[2.22, –1.55]	Afu8g02200	5.98	[1.94, 2.69]
Afu2g01230	1.97	[2.22, –1.46]	Afu8g02270	4.73	[1.84, 1.78]
Afu3g13130	2.00	[2.19, –1.45]	Afu8g01530	2.15	[2.21, –1.29]
Afu8g04100	2.00	[2.19, –1.45]	Afu1g00450	2.27	[2.12, –1.05]
Afu4g07330	2.01	[2.19, –1.46]	Afu2g11750	2.43	[2.14, –0.87]
Afu1g05270	2.02	[2.21, –1.44]	Afu4g09890	2.72	[1.87, –0.07]
Afu1g01990	2.10	[2.19, –1.37]	Afu6g12970	2.11	[2.16, –1.27]
Afu1g01610	2.22	[1.83, –0.41]	Afu4g02880	2.80	[2.04, –0.42]
Afu6g07070	2.24	[2.13, –1.11]	Afu6g11610	2.61	[2.07, –0.63]
Afu1g00310	2.26	[1.61, –0.07]	Afu3g07160	2.51	[2.18, –0.85]
Afu2g14590	2.26	[2.17, –1.13]	Afu2g02110	2.53	[1.87, –0.19]
Afu7g06400	2.29	[2.16, –1.08]	Afu8g05910	2.38	[2.08, –0.87]
Afu3g03080	2.29	[2.12, –1.05]	Afu6g14540	2.57	[2.12, –0.76]
Afu1g11560	2.34	[2.15, –1.04]	Afu3g14820	4.04	[1.80, 1.19]
Afu8g04130	2.39	[2.12, –0.88]	Afu1g00410	2.92	[2.04, –0.26]
Afu1g16090	2.41	[2.08, –0.83]	Afu5g15000	2.58	[2.06, –0.63]
Afu8g04120	2.43	[2.13, –0.86]	Afu1g00420	2.94	[2.03, –0.26]
Afu8g04040	2.46	[1.81, –0.25]	Afu1g00590	3.69	[1.95, 0.59]
Afu7g00280	2.48	[1.66, 0.04]	Afu8g05090	3.76	[1.99, 0.62]
Afu5g07560	2.68	[1.84, –0.06]	Afu4g14360	3.45	[1.99, 0.31]
Afu2g17160	2.73	[1.81, –0.03]	Afu3g13630	4.11	[1.80, 1.25]
Afu5g15010	2.78	[2.04, –0.39]	Afu1g16100	2.88	[2.02, –0.28]
Afu6g14450	2.84	[1.78, 0.11]	Afu3g01970	3.24	[1.54, 0.89]
Afu5 g15020	2.86	[2.09, –0.37]	Afu1g16110	3.05	[2.03, –0.14]
Afu8 g06380	3.10	[2.01, –0.04]	Afu7g08610	3.50	[1.92, 0.39]
Afu5 g15040	3.14	[2.06, –0.04]	Afu1g16170	3.16	[2.04, –0.03]