Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 4658506, 12 pages
http://dx.doi.org/10.1155/2016/4658506

Research Article

Positive Selection and Centrality in the Yeast and Fly Protein-Protein Interaction Networks

Sandip Chakraborty and David Alvarez-Ponce

Department of Biology, University of Nevada, Reno, NV 89557, USA

Correspondence should be addressed to David Alvarez-Ponce; dap@unr.edu

Received 29 December 2015; Accepted 7 March 2016

Academic Editor: Luoying Zhang

Proteins within a molecular network are expected to be subject to different selective pressures depending on their relative hierarchical positions. However, it is not obvious what genes within a network should be more likely to evolve under positive selection. On one hand, only mutations at genes with a relatively high degree of control over adaptive phenotypes (such as those encoding highly connected proteins) are expected to be "seen" by natural selection. On the other hand, a high degree of pleiotropy at these genes is expected to hinder adaptation. Previous analyses of the human protein-protein interaction network have shown that genes under long-term, recurrent positive selection (as inferred from interspecific comparisons) tend to act at the periphery of the network. It is unknown, however, whether these trends apply to other organisms. Here, we show that long-term positive selection has preferentially targeted the periphery of the yeast interactome. Conversely, in flies, genes under positive selection encode significantly more connected and central proteins. These observations are not due to covariation of genes' adaptability and centrality with confounding factors. Therefore, the distribution of proteins encoded by genes under recurrent positive selection across protein-protein interaction networks varies from one species to another.

1. Introduction

Scientists have been fascinated for decades by the emergence and fixation of advantageous alleles by positive selection [1, 2]. Occasionally, a new mutation is beneficial or an existing mutation becomes beneficial due to a change in the environment. Under certain conditions, as individuals carrying such mutations have an increased fitness, these mutations can quickly spread through the population, leaving a characteristic footprint in the patterns of DNA variability [3, 4]. Certain genes are more likely than others to undergo positive selection, and understanding the reasons is essential to understand adaptation. The propensity of genes to undergo positive selection depends on the balance between the potential beneficial and deleterious effects of mutations at these genes [5]. On one hand, only genes whose variability has a considerable impact on the organism's fitness (i.e., genes with a high degree of control over advantageous traits) will be able to respond to natural selection [6, 7]. On the other hand, highly pleiotropic genes (at which mutations have a high likelihood of being deleterious) will less frequently respond to positive selection [1, 8–10]. Genes do not act in isolation; instead, they often function as parts of molecular pathways and networks. Both the importance and the degree of pleiotropy of genes are affected by their position within such networks, and therefore, a network framework may enable a better understanding of genes' different propensities to be targeted by positive selection.

Proteins within a molecular pathway or network have different relative impacts on the final output of the system (the phenotype, and ultimately fitness): alteration of certain key proteins profoundly impacts the behavior of the system, whereas alteration of other, less important proteins has only marginal effects [11]. The relative importance of proteins depends not only on their intrinsic properties (e.g., their kinetic properties), but also on the position that they occupy within the network. For instance, genes acting at the upstream part, or at bifurcating points of metabolic pathways, tend to have a great influence on metabolic flux [12–14], and proteins involved in many protein-protein interactions are often essential [15–17]. Proteins' degree of pleiotropy also depends on network position, with highly connected proteins, and those involved in a high number of pathways, being often highly pleiotropic. Therefore, genes acting at different parts of a network are expected to have different propensities to undergo positive selection, but it is often not obvious what parts of the network should be targeted more often by positive selection. Adaptive evolution is expected to target genes acting at key network positions (as less important proteins will rarely be seen by natural selection), particularly if the network is far from its adaptive optimum, but adaptation may be hindered by pleiotropy at these positions. Indeed, even though multiple studies have shown that genes' propensity to undergo adaptive evolution depends on their network position, clear rules have not emerged.

Population genetics studies based on a handful of well-defined metabolic and signaling pathways have so far suggested that positive selection often targets genes with relatively important pathway positions. For instance, positive selection acted on (i) genes encoding enzymes that act at bifurcating points of the *Drosophila melanogaster* pathways involved in glucose metabolism [18] and the human *N*-glycosylation pathway [19]; (ii) the gene encoding the first enzyme of the *Arabidopsis thaliana* glucosinolate pathway [13]; and (iii) genes encoding the most connected proteins in the human insulin/TOR pathway [20]. Beneficial mutations at genes acting at such key pathway positions may lead to rapid evolutionary change. Simulation analyses of evolving pathways suggest that, at the beginning of the adaptation process, when pathways are far away from their optimum, positive selection preferentially occurs at upstream genes, and at those acting at branch points; however, once pathways approach their optimum, upstream genes are highly constrained and downstream genes are the ones that undergo positive selection [12, 14].

In the last years, a considerable amount of genomic and functional data has accumulated, allowing evolutionary biologists to study the distribution of genes under positive selection in different kinds of large-scale networks. Genes under positive selection in honey bees, as inferred from the McDonald-Kreitman test [21], are lowly connected in the gene coexpression network [22]. In the *D. melanogaster* metabolic network, genes under positive selection, as inferred from the comparison of six *Drosophila* genomes, do not exhibit any particular network position; however, very few genes under positive selection were found in this study, which may have resulted in limited statistical power [23].

The relationship between genes' propensity to exhibit signatures of positive selection and the number of physical protein-protein interactions in which the encoded product is involved has only been addressed in humans, and contrasting trends have been observed depending on the evolutionary timescale considered. When recurring, long-term positive selection was inferred from comparison of the human and chimpanzee genomes [24], or from comparison of 10 mammalian genomes [16] (including 9 placentals and one marsupial, which diverged 157–170 million years ago [25]), using tests based on the nonsynonymous to synonymous divergence ratio ($\omega = d_N/d_S$), positive selection was found to target preferentially genes acting at the periphery of the human protein-protein interaction network (i.e., genes encoding lowly connected proteins). Conversely, when recent selective sweeps were inferred from comparison of hundreds of human genomes, it was found that genes under positive selection were significantly more connected than genes with no signatures of positive selection [16, 26].

Are the trends observed thus far in the human protein-protein interaction network common to all organisms? Here, we characterize the distribution of genes under recurrent positive selection in the interactomes of *Saccharomyces cerevisiae* and *D. melanogaster*. For that purpose, we infer long-term positive selection events by comparing the genomes of five *Saccharomyces* and six *Drosophila* species. We find that, similar to what was previously observed in humans [16, 24], genes under positive selection act at the periphery of the *S. cerevisiae* protein-protein interaction network. Conversely, in *D. melanogaster*, genes under positive selection are significantly more connected than genes with no signatures of positive selection.

2. Materials and Methods

2.1. Tests of Positive Selection. For each *S. cerevisiae* gene, the longest encoded protein was selected for analysis, and orthologs were identified in another 4 *Saccharomyces* genomes using a best reciprocal hit approach. Each *S. cerevisiae* longest protein was used as a query in BLASTP search (*E*-value cut-off: 10^{-10}) against the proteomes of *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, and *S. bayanus*. The best hit in each proteome was used as a query in a second BLASTP search (*E*-value < 10^{-10}) against the *S. cerevisiae* proteome. If the best hit identified in the second search was the original *S. cerevisiae* protein, then the encoding genes were considered to be orthologs. Only *S. cerevisiae* genes with identifiable orthologs in all four *Saccharomyces* species were used in our analyses. The same strategy was adopted to identify orthologs of all *D. melanogaster* genes in the genomes of *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, and *D. ananassae*. We did not include more distant species in our analyses in order to (i) avoid saturation of synonymous sites; (ii) maintain a high number of analyzable genes (genes with orthologs in all considered species); and (iii) minimize problems resulting from alignment of highly divergent sequences.

Groups of orthologous protein sequences were aligned using ProbCons [27]. Given that gene annotations of non-model organisms are performed using automatic methods, which often produce imperfect gene models [28, 29], and that tests of positive selection are highly sensitive to such errors [30–33], we stringently filtered our protein sequence alignments using a three-step procedure (as in [16]). First, Gblocks version 0.91b [34] was used to eliminate nonalignable and poorly alignable regions. Second, a sliding window approach was used to identify alignment regions of 15 amino acids in which one of the sequences presented 10 or more singleton amino acids (amino acids that are unique to one sequence), and regions of 5 amino acids in which all were singletons in one of the species; such regions are unlikely to be correctly annotated. The original (unfiltered) sequence alignments were combined with the coding sequences (CDSs) and the results of the two first filtering steps to produce CDS alignments using an in-house pipeline. The resulting alignments were used in a test of positive selection using the M8 versus M7 test (see below). Third, for genes inferred to be under positive selection, CDS sequence alignments were visualized and erroneously annotated regions were manually removed using BioEdit version 7.2.5 [35] before rerunning the test of positive selection.

For each alignment, the presence of a set of codons with $\omega > 1$ was inferred using the M8 versus M7 test [36]. The likelihood of alignment under the M8 and M7 models was estimated using the codeml program of the PAML package version 4.4d [37]. In order to alleviate the problem of local optima, all computations were repeated using three starting ω values ($\omega = 0.04$, 0.4, and 4). Both models assume that codons' ω values follow a beta distribution, with values ranging from 0 to 1. The M8 model allows for an additional class of codons with $\omega_s > 1$. The fit of both models was compared using a likelihood ratio test: twice the difference in the log-likelihood of both models $[2\Delta\ell = 2 \times (\ell_{M8} - \ell_{M7})]$ was assumed to follow a χ^2 distribution with two degrees of freedom [38]. Genes with a P value lower than 0.05 and ω_s higher than 1 were considered to be under positive selection. Analyses were repeated using a more stringent P value ($P < 0.01$ and $\omega_s > 1$), and controlling the false discovery rate associated with multiple testing using the Benjamini and Hochberg approach ($q < 0.1$ and $\omega_s > 1$) [39]. Unless stated otherwise, genes considered to be under positive selection throughout this study correspond to those with $P < 0.05$ and $\omega_s > 1$.

2.2. Network Data and Analyses. Protein-protein interaction data for *S. cerevisiae* and *D. melanogaster* were obtained from the BioGRID database version 3.4.129 [40]. This database contains only experimentally determined interactions. Only physical nonredundant interactions between proteins from the same organism were used in our analyses. Additional analyses were conducted using interaction data from the STRING database version 10 [41]. This database contains data from both experimentally determined and computationally predicted (based on coexpression, phylogenetic profiles, etc.) interactions. Only interactions with a confidence score ≥40% were used in our analyses.

For each protein and network, degree was computed as the number of other proteins with which the protein interacts, betweenness was computed as the number of shortest paths among other proteins that pass through the protein [42], and closeness was computed as one divided by the average distance (number of steps) between the protein and all other proteins. Betweenness and closeness computations were conducted using Pajek version 4.05 [43].

2.3. Protein Abundance and Gene Expression Data. Protein abundance data for *S. cerevisiae* and for the whole body of *D. melanogaster* adults was obtained from the PaxDB database version 4 [44]. Messenger RNA abundance data for *S. cerevisiae* was obtained from [45]. Messenger RNA abundance data for the whole *D. melanogaster* adult and 16 adult nonredundant tissues/organs were obtained from the FlyAtlas database [46]. Probes were mapped to genes using the Affymetrix annotation file "Drosophila 2" version 35. Probes matching multiple genes were not used in our analyses. For genes matching multiple probes, the probe with the highest mRNA abundance in the whole fly was used. The expression breadth of each *D. melanogaster* gene was computed as the number of tissues/organs in which the gene is expressed. The considered tissues were brain, head, eye, thoracicoabdominal ganglion, salivary gland, crop, midgut, tubule, hindgut, heart, fat body, ovary, testis, male accessory glands, virgin spermatheca, and carcass. A gene was considered to be expressed in a tissue/organ if the database reported presence in at least 3 out of the 4 biological replicates.

2.4. Number of Publications. For each *S. cerevisiae* gene, the number of publications in which it is referred was obtained from the *Saccharomyces* Genome Database [47]. The number of publications related to each *D. melanogaster* gene was obtained from FlyBase [48]. These data were obtained in February 2016.

3. Results

3.1. Positive Selection Acted Preferentially at the Periphery of the Yeast Protein-Protein Interaction Network. We identified the orthologs of all *S. cerevisiae* genes in the genomes of another four *Saccharomyces* genomes. A total of 2071 *S. cerevisiae* genes had identifiable orthologs in all four genomes. Sequence alignments were filtered using highly stringent criteria, and the filtered alignments were used in a maximum likelihood test of positive selection [36]. A total of 91 genes exhibited signatures of positive selection according to our initial criteria ($P < 0.05$ and $\omega_s > 1$). This number is moderately higher than that resulting from a scan based on three *Saccharomyces* genomes [49].

We reconstructed the yeast protein-protein interaction network from the experimentally determined physical protein-protein interactions recorded in the BioGRID database [40]. The network contained a total of 5864 nonredundant proteins and 81,040 nonredundant interactions (Table S1 in Supplementary Material available online at http://dx.doi.org/10.1155/2016/4658506). Out of the 5864 genes encoding the proteins represented in the network, 89 exhibited signatures of positive selection, 1956 did not exhibit signatures of positive selection, and in the remaining genes the test could not be performed, as orthologs were not identified in all yeast species.

Genes with signatures of positive selection encode proteins that exhibit a significantly lower number of interactions (average for genes under positive selection: 18.01; average for genes with no signatures of positive selection: 27.23; Mann-Whitney U test, $P = 0.016$) and a significantly lower closeness centrality (mean for genes under positive selection: 0.387; mean for genes without signatures of positive selection: 0.400; $P = 0.008$). Genes under positive selection also exhibit

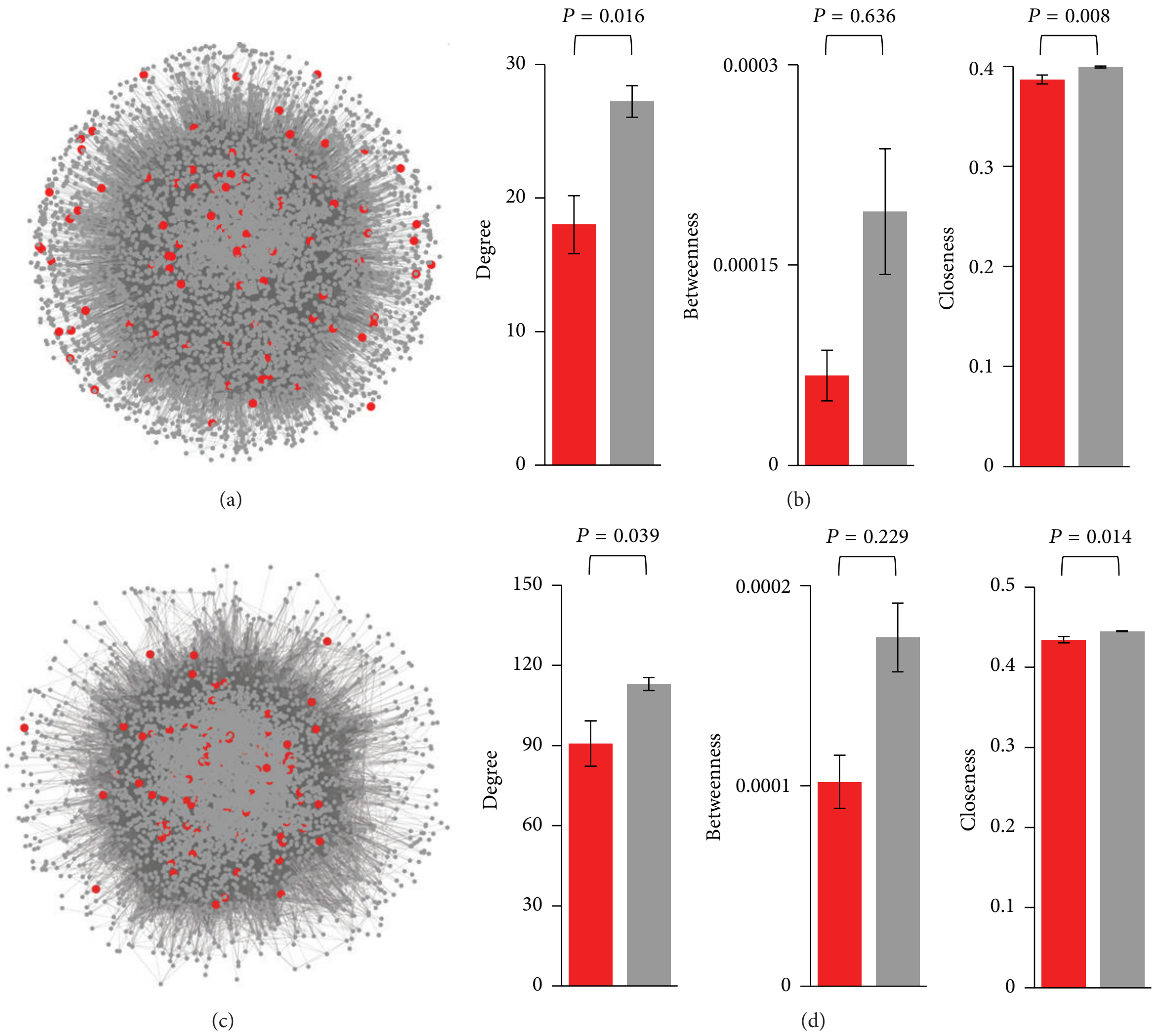

Figure 1: Distribution of proteins encoded by genes under positive selection in the *S. cerevisiae* protein-protein interaction network. (a) BioGRID network. (b) Average network centrality metrics calculated from the BioGRID network. (c) STRING network. (d) Average network centrality metrics calculated from the STRING network. In panels (a) and (c), proteins encoded by genes under positive selection are represented in red, and the rest of the proteins are represented in gray. In panels (b) and (d), genes under positive selection are represented in red, and genes with no signatures of positive selection are represented in gray. Error bars correspond to the standard error of the mean. Genes were considered to be under positive selection if they exhibited $P < 0.05$ and $\omega_s > 1$. For analyses based on more stringent criteria, see Tables S3 and S4. P values represented in the figure correspond to the Mann-Whitney U test.

a substantially but not significantly lower betweenness centrality (average for genes under positive selection: 6.73×10^{-5}; average for genes with no signatures of positive selection: 1.90×10^{-4}; $P = 0.636$) (Figures 1(a) and 1(b); Table S2). When a more stringent P value cut-off was applied in our tests of positive selection ($P < 0.01$), only 31 network genes were considered to be under positive selection. When the results of the tests of positive selection were corrected for multiple testing ($q < 0.1$), only 5 of these genes remained significant. Both gene sets encoded proteins with a substantially lower degree and betweenness, but differences were not significant, probably due to limited statistical power resulting from the small sample sizes (Tables S3 and S4).

We repeated our network analyses using a denser network obtained from the data recorded in the STRING database, which contains not only experimentally determined but also computationally predicted protein and gene interactions [41] (Table S1). Similar results were obtained: proteins encoded by genes under positive selection exhibit a significantly lower degree and closeness and a substantially, but not significantly, lower betweenness (Figures 1(c) and 1(d); Table S2). No significant differences were observed when more stringent criteria were used in the tests of positive selection ($P < 0.01$ or $q < 0.1$; Tables S3 and S4).

We next considered whether our observations might be due to covariation of both gene adaptability and network centrality with different potentially confounding factors, rather than to a direct link between adaptability and centrality. Previous results in yeasts and other organisms have shown that central genes tend to be highly expressed [50–58], and

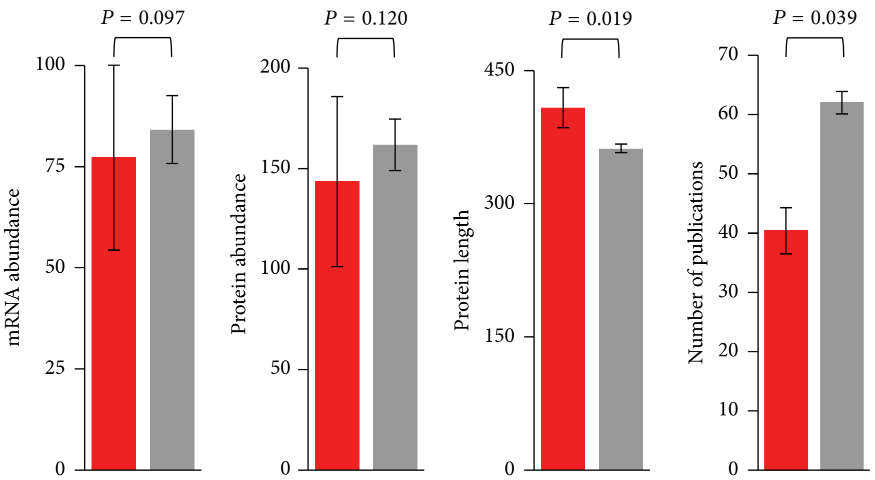

Figure 2: Differences in mRNA abundance, protein abundance, protein length, and number of publications between positively selected genes (in red) and genes without signatures of positive selection (in gray) in *S. cerevisiae*. Genes were considered to be under positive selection if they exhibited $P < 0.05$ and $\omega_s > 1$. P values represented in the figure correspond to the Mann-Whitney U test.

previous works in humans have shown that highly expressed genes are unlikely to undergo adaptive evolution [24, 59]. Combined, these observations raise the possibility that our observations could merely be a byproduct of the distribution of expression levels in the network. We found a positive correlation between degree and both mRNA abundance (BioGRID network: Spearman's rank correlation coefficient, $\rho = 0.294$, $P = 9.6 \times 10^{-37}$; STRING network: $\rho = 0.320$, $P = 7.1 \times 10^{-44}$) and protein abundance (BioGRID network: $\rho = 0.452$, $P = 2.1 \times 10^{-103}$; STRING network: $\rho = 0.441$, $P = 5.8 \times 10^{-99}$). However, we found no differences between the expression levels and protein abundances of genes under positive selection and genes with no signatures of positive selection in yeast (Figure 2), which allowed us to discard these factors as the reason underlying our observations.

Also consistent with previous results in yeasts [60], we observed a positive correlation between proteins' length and number of protein-protein interactions (BioGRID network: $\rho = 0.076$, $P = 6.3 \times 10^{-4}$; STRING network: $\rho = 0.078$, $P = 4.1 \times 10^{-4}$). In line with previous results in mammals [16], we found that yeast genes under positive selection encode significantly longer proteins than those with no signatures of positive selection (Figure 2), which is consistent with the power of the test depending on the number of codons analyzed [61]. Combined, these observations indicate that our observation that genes under positive selection tend to encode peripheral genes cannot be due to covariation with protein length.

Interactomic datasets are known to be subjected to a number of biases [62]. In particular, such datasets tend to include a disproportionally high number of interactions involving proteins that have been studied in great detail (e.g., because of their particular importance or interest), as more resources have been devoted to study them. Indeed, we observed that protein degrees positively correlate with the number of publications mentioning the proteins (BioGRID network: $\rho = 0.651$, $P < 10^{-6}$; STRING network: $\rho = 0.553$, $P < 10^{-6}$) and that genes under positive selection tend to be mentioned in a lower number of publications ($P = 0.039$; Figure 2). Nonetheless, this bias is unlikely to explain our observations: a partial correlation analysis shows that, in the STRING network, $2\Delta\ell$ correlates with degree, even when controlling for the number of publications ($\rho = -0.066$, $P = 0.003$). In addition, when we used a subnetwork of the BioGRID network containing only the interactions determined by high-throughput techniques (which are expected to be less prone to this kind of bias) the difference between the degree of proteins encoded by genes under positive selection (mean = 15.31, median = 9) and the degree of the proteins encoded by genes with no signatures of positive selection (mean = 22.18, median = 11) remains substantially and marginally significantly different ($P = 0.058$).

Another known problem of currently available interactomes is their high rate of false positives [63–65]. To alleviate this problem, we generated two highly stringent subnetworks of our BioGRID and STRING networks. The first subnetwork was generated by considering only those protein-protein interactions determined by low-throughput techniques (which are expected to produce more reliable results than high-throughput techniques). In this case, the difference between the degree of proteins encoded by genes under positive selection (mean = 7.62, median = 5) and the degree of proteins encoded by genes with no signatures of positive selection (mean = 9.69, median = 5) remained substantial; however, the differences were not statistically significant ($P = 0.808$), probably owing to the reduced statistical power resulting from filtering the network. The second subnetwork was obtained by considering only the interactions described in the STRING database with a confidence score ≥50%. The degrees of proteins encoded by genes under positive selection were significantly lower (genes under positive selection: mean = 65.87, median = 43; genes with no signatures of positive selection: mean = 80.93, median = 54; $P = 0.047$). When an even more stringent cut-off was applied (score ≥90%), the differences were even more marked, but nonsignificant (positively selected: mean = 22.05, median = 9.5; non-positively selected: mean = 27.92, median = 12; P = 0.177), probably due to reduced statistical power.

3.2. Positive Selection Acted Preferentially at the Center of the Fly Protein-Protein Interaction Network. We performed a scan of positive selection using the genomes of six *Drosophila* species. Orthologs in all species were found for 10,340 *D. melanogaster* genes, and signatures of positive selection were detected in 533 of these genes using a P value threshold of 0.05. This number is smaller than those resulting from previous scans of positive selection in *Drosophila* [31, 66], as expected from the fact that we applied highly stringent criteria, including manual inspection and editing of the alignments in which positive selection was detected (Section 2). The protein-protein interaction network, constructed from the data available at the BioGRID database [40], consisted of 7968 nonredundant proteins and 36,589 nonredundant interactions (Table S1). Out of the 7968 genes whose encoded products are represented in the network, positive selection was inferred in 350, no signatures of positive selection were found in 6171, and the positive selection test could not be performed on the rest because orthologs were not present in some of the genomes.

Remarkably, we found that, contrary to what was found in yeast (Figure 1; Table S2) and humans [16, 24], genes under positive selection encoded proteins with a significantly higher degree (average for genes under positive selection: 10.49; average for genes with no signatures of positive selection: 9.19; $P = 0.049$), betweenness (average for genes under positive selection: 4.5×10^{-4}; average for genes with no signatures of positive selection: 3.8×10^{-4}; $P = 0.008$), and closeness (average for genes under positive selection: 0.242; average for genes with no signatures of positive selection: 0.239; $P = 0.045$) (Figures 3(a) and 3(b), Table S5). Similar results were obtained when a more stringent P value cut-off was used ($P < 0.01$), with the only exception that the differences are not significant, albeit substantial, for betweenness (Table S6). When the results of the positive selection test were corrected for multiple testing by controlling the false discovery rate ($q < 0.1$), only 50 network genes retained signatures of positive selection. These genes exhibit a higher degree, betweenness, and closeness, even though the differences are not statistically significant, probably due to a reduced statistical power due to the small genes under positive selection (Table S7). Similar results were also obtained when the network was assembled from the contents of the STRING database [41]; in this case, the differences are statistically significant for degree and closeness when a P value cut-off of $P < 0.05$ was used to detect positive selection (Figures 3(c) and 3(d), Table S5), and for degree, betweenness, and closeness when the more stringent cut-off of $P < 0.01$ was used (Table S6) or when correction for multiple testing was applied ($q < 0.1$; Table S7).

Consistent with previous results in *Drosophila* and other organisms [50–58], protein degree positively correlates with mRNA abundance (BioGRID network: $\rho = 0.097$, $P = 1.1 \times 10^{-81}$; STRING network: $\rho = 0.338$, $P = 8.2 \times 10^{-114}$). We also found degree to positively correlate with protein abundance (BioGRID network: $\rho = 0.120$, $P = 4.3 \times 10^{-10}$; STRING network: $\rho = 0.288$, $P = 1.8 \times 10^{-83}$) and with expression breadth (BioGRID network: $\rho = 0.179$, $P = 2.2 \times 10^{-47}$; STRING network: $\rho = 0.036$, $P = 0.017$). However, none of these parameters significantly differs between genes under positive selection and genes with no signatures of positive selection (Figure 4). In addition, genes with different expression breadths do not differ in their propensity to exhibit signatures of positive selection (correlation between expression breadth and the fraction of genes under positive selection: $\rho = 0.018$, $P = 0.948$; Figure 5). These observations allow us to discard the possibility that the observed higher centrality of genes under positive selection (Figure 3; Tables S5–S7) could be due to covariation of adaptability and centrality with these expression parameters. Similar to what is observed in yeast (Figure 2) and humans [16], genes under positive selection tend to encode long proteins in *Drosophila* (Figure 4). However, protein length does not correlate with number of interactions (BioGRID network: $\rho = 0.028$, $P = 0.133$; STRING network: $\rho = 0.009$, $P = 0.555$), indicating that our observations are not due to covariation with protein length either.

Protein degrees were found to positively correlate with the number of publications mentioning each protein (BioGRID network: $\rho = 0.249$, $P < 10^{-6}$; STRING network: $\rho = 0.251$, $P < 10^{-6}$). However, three lines of evidence demonstrate that this has not biased our results. First, the average number of publications does not significantly differ between genes under positive selection and genes with no signatures of positive selection ($P = 0.161$; Figure 4). Second, we repeated our analyses using the *D. melanogaster* protein-protein interaction network generated by Giot et al. [67]. This network is the result of a large-scale experiment in which virtually every possible interaction was tested, and therefore it is expected to be unbiased. Similar to our analyses on the entire BioGRID network, we observed that proteins encoded by genes under positive selection exhibited a higher degree (genes under positive selection: mean = 6.50, median = 3; genes with no signatures of positive selection: mean = 5.81, median = 3; $P = 0.224$). Third, when we repeated our analyses on a subnetwork containing only those interactions determined by high-throughput techniques, we observed significant differences between proteins encoded by genes under positive selection and those encoded by proteins with no signatures of positive selection (positively selected: mean = 10.42, median = 4; non-positively selected: mean = 8.97, median = 4; $P = 0.046$).

Finally, we repeated our analyses on two highly stringent subnetworks of our BioGRID and STRING networks. When we considered only those protein-protein interactions determined by low-throughput experiments, proteins encoded by genes under positive selection remained substantially more central; however, the differences were not statistically significant, probably due to reduced statistical power resulting from filtering our network (positively selected: mean = 2.91, median = 1; non-positively selected: mean = 2.56, median = 1;

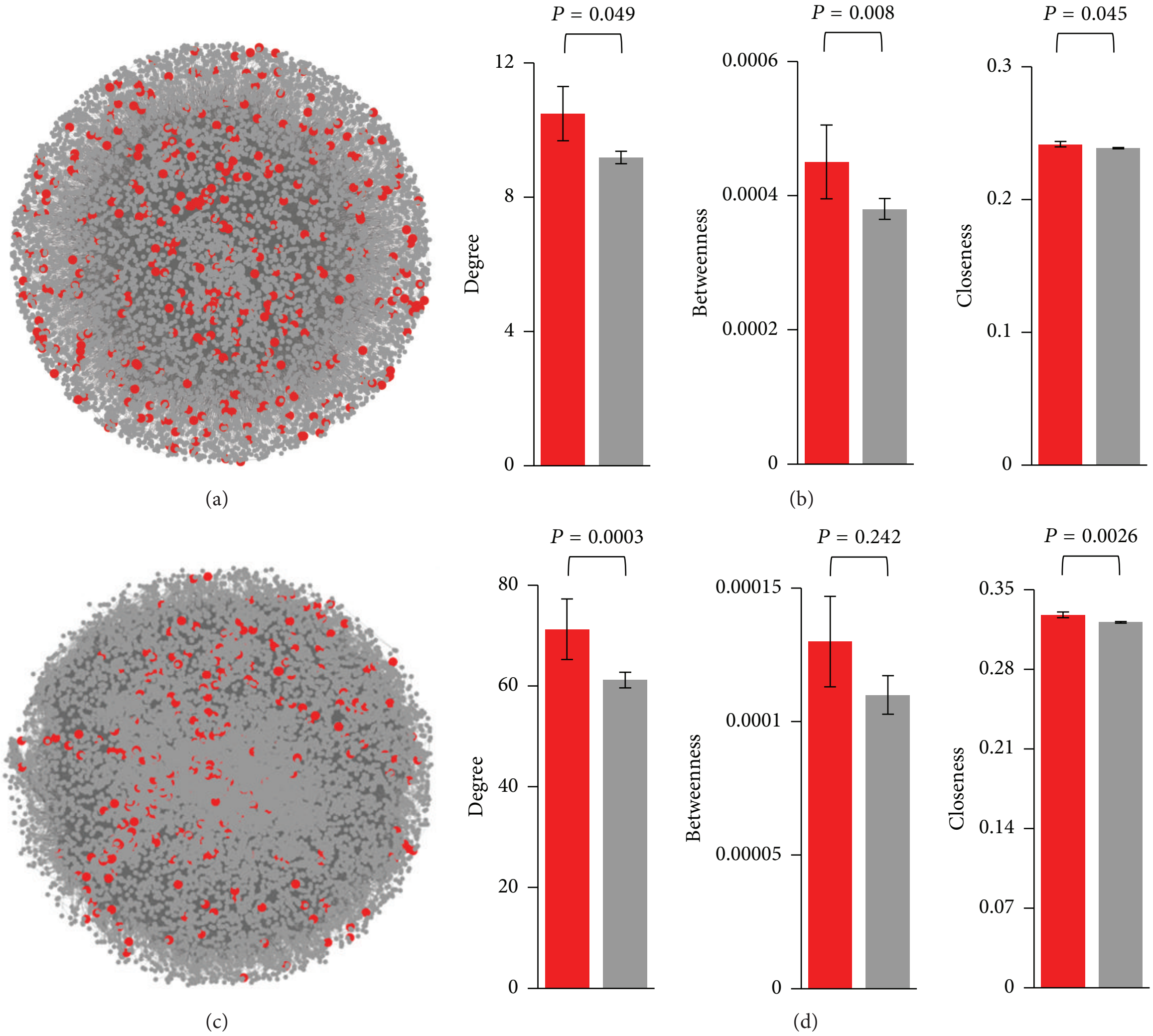

Figure 3: Distribution of proteins encoded by genes under positive selection in the *D. melanogaster* protein-protein interaction network. (a) BioGRID network. (b) Average network centrality metrics calculated from the BioGRID network. (c) STRING network. (d) Average network centrality metrics calculated from the STRING network. In panels (a) and (c), proteins encoded by genes under positive selection are represented in red, and the rest of the proteins are represented in gray. In panels (b) and (d), genes under positive selection are represented in red, and genes with no signatures of positive selection are represented in gray. Error bars correspond to the standard error of the mean. Genes were considered to be under positive selection if they exhibited $P < 0.05$ and $\omega_s > 1$. For analyses based on more stringent criteria, see Tables S6 and S7. P values represented in the figure correspond to the Mann-Whitney U test.

$P = 0.708$). When we considered only the interactions described in the STRING database with a confidence score ≥50% or ≥60%, proteins encoded by genes under positive selection remained significantly more central ($P = 2.3 \times 10^{-3}$ and 8.1×10^{-3}, resp.). When only interactions with a score ≥90% were considered, the differences were even more marked, but nonsignificant (genes under positive selection: mean = 15.15, median = 4; genes with no signatures of positive selection: mean = 12.89, median = 4; $P = 0.756$), probably due to reduced statistical power.

4. Discussion

We have performed two scans of positive selection by comparing the genomes of five *Saccharomyces* and six *Drosophila* species and investigated the position of the proteins encoded by genes under positive selection in the protein-protein interaction networks of *S. cerevisiae* and *D. melanogaster*. Consistent with previous results in humans [16, 24], we found that genes under positive selection encode significantly less connected proteins in the interactome of *S. cerevisiae*. However, the opposite was observed in *Drosophila*: proteins encoded by genes under positive selection are significantly more connected than those encoded by genes with no signatures of positive selection. These observations were not due to covariation of network centrality and positive selection with protein abundance, expression level, protein length, or, in the case of *Drosophila*, expression breadth. Equivalent results were obtained when considering betweenness and closeness (two descriptors of the global position of proteins

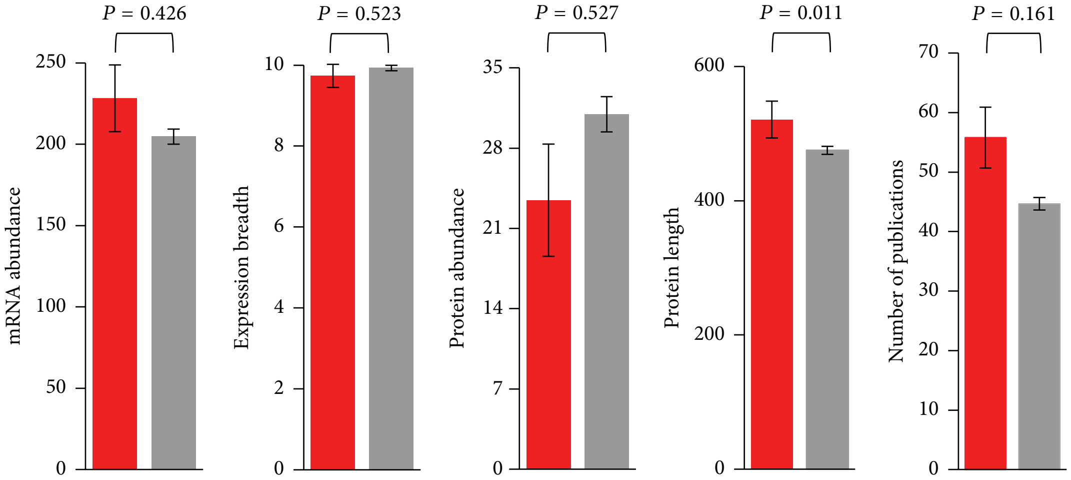

FIGURE 4: Differences in mRNA abundance, expression breadth, protein abundance, protein length, and number of publications between positively selected genes (in red) and genes without signatures of positive selection (in gray) in *D. melanogaster*. Genes were considered to be under positive selection if they exhibited $P < 0.05$ and $\omega_s > 1$. P values represented in the figure correspond to the Mann-Whitney U test.

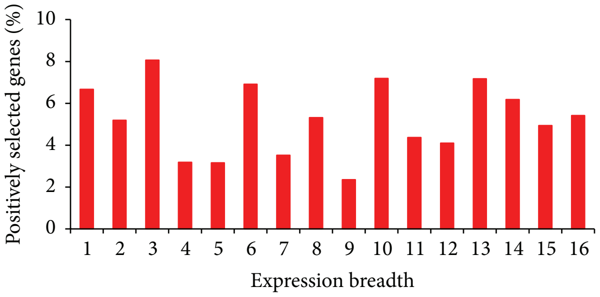

FIGURE 5: Percent of genes under positive selection among genes with different expression breadths in *D. melanogaster*. Genes were considered to be under positive selection if they exhibited $P < 0.05$ and $\omega_s > 1$.

within the network, which take into account not only their direct interactors, but rather their potential role in connecting different parts of the network [42, 68]). Equivalent results were also obtained by analyzing the STRING database [41], which contains both experimentally determined and computationally predicted gene interactions.

Our observations imply that interactome position has an impact on the propensity of genes to undergo adaptive evolution. In other words, genes under positive selection do not distribute randomly in protein-protein interaction networks. Previous studies in humans have suggested that the relationship between network centrality and the propensity of genes to undergo positive selection depends on the timescale considered: genes that underwent positive selection recurrently during long evolutionary times, as revealed from comparison of the genomes of different species, act at the periphery of the human interactome [16, 24], whereas genes that underwent positive selection recently, as inferred from comparison of human genomes, encode highly interactive proteins [16, 26]. Our study shows that the distribution of genes under positive selection in the protein-protein interaction network also varies from one species to another.

The test of positive selection that we used [36] can detect adaptation events that affected the protein sequence in a recurring manner during long evolutionary periods (e.g., as a result of an arms race dynamics [69]), and not adaptation events in the regulatory region. Therefore, our study focused on the adaptation at the protein sequence level (likely protein function), rather than at the regulatory level. This was also the case of the study by Kim et al. [24], and the interspecific analysis conducted by Luisi et al. [16]. In contrast, studies in which positive selection was inferred from the SNP frequency spectrum, estimated from comparison of DNA sequences of alleles of the same species (e.g., the population genomics studies conducted by Luisi et al. [16] and by Qian et al. [26]), may have captured recent adaptation events, at both the regulatory and the protein sequence level.

Genes within a network have different hierarchical positions and, therefore, a different relative potential to affect adaptive phenotypes. In the context of protein-protein interaction networks, centrality is a proxy for this relative importance. Mutations affecting proteins involved in a high number of protein-protein interactions, or those with a high global centrality, are expected to have a high influence on network dynamics, and to have highly pleiotropic effects (indeed, highly pleiotropic genes tend to encode highly interactive proteins [70]). Consistently, genes encoding central proteins are often essential [15–17] and highly constrained by purifying selection [15, 16, 50, 71, 72].

At least two opposing forces may determine the direction of the relationship between genes' adaptability and network centrality. On one hand, beneficial mutations at genes encoding the "key" proteins of the network (e.g., those involved in a high number of protein-protein interactions) are likely to have a great impact on phenotypes and fitness, whereas beneficial mutations at genes encoding less important proteins (those whose variability does not impact much the final output of the network) will rarely be seen by natural selection. This is expected to result in positive selection targeting preferentially genes acting at the center of the network. On the other hand, pleiotropy is thought to constrain the adaptation of protein sequences. Mutations at genes involved in many biological processes, or in many interactions, are expected to affect a high number of phenotypes, thus making it unlikely that such genes can experience drastic changes at the protein sequence level [1, 8–10]. In addition, purifying selection may rapidly remove most nonsynonymous mutations from these genes, further hindering adaptation at the protein level. This is expected to result in positive selection acting preferentially at the network periphery. Nonetheless, compensatory mutations may promote recurring adaptation events at highly pleiotropic proteins. Adaptation of one aspect of the function of a protein may have negative side effects on other aspects of its function, which can be ameliorated or restored by subsequent adaptation events [73, 74].

It is possible that the balance between both forces is different in yeasts and *Drosophila* and that it is also different when considering the long-term evolutionary history of mammals *versus* the recent evolutionary history of humans. However, it is not obvious why the balance might be different in different organisms and/or timescales. One factor affecting the relative importance of both forces may be the point of the adaptive landscape in which a population is. Adaptive walks often proceed by an initial period of fixation of large-effect adaptive mutations, followed by fixation of mutations of smaller effects [1, 75, 76]. Populations that are poorly adapted to the environment (e.g., due to an environmental change) might undergo big adaptive leaps by fixing mutations at highly central genes. Conversely, populations that are near their adaptive optimum may undergo adaptation preferentially at the network periphery. Consistent with this model of diminishing returns, simulation of the adaptation of hypothetical, randomly generated metabolic pathways has shown that the first steps of adaptation (when pathways are far away from their optimal function) take place through positive selection acting at upstream genes, and those acting at branch points (the ones with a higher degree of control over the pathway flux), whereas at the end of the simulations (when pathways are near the optimum) pathways are fine-tuned by positive selection at downstream genes (which have a smaller influence on flux) [12, 14]. It is unclear, however, why the *Drosophila* network would be far away from its optimal functioning compared to the yeast and mammalian networks.

Effective population size (N_e) may be another key modulator of the centrality of genes under positive selection. In organisms with large N_e, the efficacy of natural selection is high, and even mutations with small selection coefficients will be fixed or removed by natural selection. This is expected to result in genes under positive selection at both the center and the periphery of the network. Conversely, in organisms with small N_e, genetic drift can outpower natural selection, and only mutations with large effects are expected to be fixed/removed by natural selection [6], which is expected to result in positive selection mostly at the center of the network. Nonetheless, this is unlikely to explain the different trends observed in yeasts, *Drosophila*, and humans, as *D. melanogaster* is thought to have N_e higher than humans and lower than yeast (e.g., see [77–80]).

Another potentially important consideration is the so-called "cost of complexity." Large-size mutations will more often be disruptive in complex organisms (those with many characters) than in simple ones [1, 81]. This may promote adaptation at the periphery of the networks of complex organisms. However, it is again unlikely that this factor has caused the observed differences between *Drosophila* and yeasts and mammals, as *Drosophila* exhibits an intermediate complexity between yeasts and mammals [82, 83]. The cost of complexity might be partially reduced by network modularity, as it may significantly reduce the pleiotropic effects of adaptive mutations by containing genes in small areas of influence [84, 85]. Nonetheless, there is no reason to think that the *Drosophila* interactome is more modular than those of yeast and human [86, 87], and *Drosophila* genes under positive selection exhibit high betweenness and closeness centralities, which is not compatible with their being confined in modules.

The five *Saccharomyces* species used in our analysis (all belonging to the *Saccharomyces sensu stricto* complex) diverged 10–20 million years ago [88]. The six *Drosophila* species analyzed in this study (all belonging to the *melanogaster* group) diverged ~30 million years ago [89]. Kim et al. [24] inferred positive selection from comparison of the human and chimpanzee genomes, which diverged ~6 million years ago [90, 91], and the 10 mammalian genomes studied by Luisi et al. [16] diverged 157–170 million years ago [25]. Therefore, the divergence time considered in our scan of positive selection in *Drosophila* falls within the range of the divergence times for the species used in the scans for the other taxa, indicating that the peculiar distribution of genes under positive selection within the *Drosophila* network is not due to divergence times.

Therefore, it is unclear why the distribution of genes under recurrent positive selection is different in the *Drosophila* interactome and in the human and yeast interactomes. In any case, our observations imply that even though network position is a key factor determining genes' propensity to undergo positive selection, the relationship between both factors is complex and lineage-specific.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The authors are grateful to Felix Feyertag for assistance with handling *D. melanogaster* gene expression data. This work has been supported by funds from the University of Nevada, Reno.

References

[1] R. A. Fisher, *The Genetical Theory of Natural Selection*, Oxford University Press, Oxford, UK, 1930.

[2] H. J. Muller, "Some genetic aspects of sex," *The American Naturalist*, vol. 66, no. 703, pp. 118–138, 1932.

[3] J. J. Vitti, S. R. Grossman, and P. C. Sabeti, "Detecting natural selection in genomic data," *Annual Review of Genetics*, vol. 47, pp. 97–120, 2013.

[4] Z. Yang and J. R. Bielawski, "Statistical methods for detecting molecular adaptation," *Trends in Ecology and Evolution*, vol. 15, no. 12, pp. 496–503, 2000.

[5] C. F. Olson-Manning, M. R. Wagner, and T. Mitchell-Olds, "Adaptive evolution: evaluating empirical support for theoretical predictions," *Nature Reviews Genetics*, vol. 13, no. 12, pp. 867–877, 2012.

[6] M. Kimura, *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge, UK, 1983.

[7] M. Kimura, T. Maruyama, and J. F. Crow, "The mutation load in small populations," *Genetics*, vol. 48, no. 10, pp. 1303–1312, 1963.

[8] D. L. Stern and V. Orgogozo, "The loci of evolution: how predictable is genetic evolution?" *Evolution*, vol. 62, no. 9, pp. 2155–2177, 2008.

[9] S. B. Carroll, "Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution," *Cell*, vol. 134, no. 1, pp. 25–36, 2008.

[10] P. Beldade, K. Koops, and P. M. Brakefield, "Modularity, individuality, and evo-devo in butterfly wings," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 22, pp. 14262–14267, 2002.

[11] H. Kacser and J. A. Burns, "The control of flux," *Symposia of the Society for Experimental Biology*, vol. 27, pp. 65–104, 1973.

[12] K. M. Wright and M. D. Rausher, "The evolution of control and distribution of adaptive mutations in a metabolic pathway," *Genetics*, vol. 184, no. 2, pp. 483–502, 2010.

[13] C. F. Olson-Manning, C.-R. Lee, M. D. Rausher, and T. Mitchell-Olds, "Evolution of flux control in the glucosinolate pathway in *Arabidopsis thaliana*," *Molecular Biology and Evolution*, vol. 30, no. 1, pp. 14–23, 2013.

[14] M. D. Rausher, "The evolution of genes in branched metabolic pathways," *Evolution*, vol. 67, no. 1, pp. 34–48, 2013.

[15] M. W. Hahn and A. D. Kern, "Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks," *Molecular Biology and Evolution*, vol. 22, no. 4, pp. 803–806, 2005.

[16] P. Luisi, D. Alvarez-Ponce, M. Pybus, M. A. Fares, J. Bertranpetit, and H. Laayouni, "Recent positive selection has acted on genes encoding proteins with more interactions within the whole human interactome," *Genome Biology and Evolution*, vol. 7, no. 4, pp. 1141–1154, 2015.

[17] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, "Lethality and centrality in protein networks," *Nature*, vol. 411, no. 6833, pp. 41–42, 2001.

[18] J. M. Flowers, E. Sezgin, S. Kumagai et al., "Adaptive evolution of metabolic pathways in *Drosophila*," *Molecular Biology and Evolution*, vol. 24, no. 6, pp. 1347–1354, 2007.

[19] G. M. Dall'Olio, H. Laayouni, P. Luisi, M. Sikora, L. Montanucci, and J. Bertranpetit, "Distribution of events of positive selection and population differentiation in a metabolic pathway: the case of asparagine N-glycosylation," *BMC Evolutionary Biology*, vol. 12, article 98, 2012.

[20] P. Luisi, D. Alvarez-Ponce, G. M. Dall'Olio, M. Sikora, J. Bertranpetit, and H. Laayouni, "Network-level and population genetics analysis of the insulin/TOR signal transduction pathway across human populations," *Molecular Biology and Evolution*, vol. 29, no. 5, pp. 1379–1392, 2012.

[21] J. H. McDonald and M. Kreitman, "Adaptive protein evolution at the Adh locus in Drosophila," *Nature*, vol. 351, no. 6328, pp. 652–654, 1991.

[22] W. C. Jasper, T. A. Linksvayer, J. Atallah, D. Friedman, J. C. Chiu, and B. R. Johnson, "Large-scale coding sequence change underlies the evolution of postdevelopmental novelty in honey bees," *Molecular Biology and Evolution*, vol. 32, no. 2, pp. 334–346, 2015.

[23] A. J. Greenberg, S. R. Stockwell, and A. G. Clark, "Evolutionary constraint and adaptation in the metabolic network of *Drosophila*," *Molecular Biology and Evolution*, vol. 25, no. 12, pp. 2537–2546, 2008.

[24] P. M. Kim, J. O. Korbel, and M. B. Gerstein, "Positive selection at the protein network periphery: evaluation in terms of structural constraints and cellular context," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 51, pp. 20274–20279, 2007.

[25] J. E. Tarver, M. Dos Reis, S. Mirarab et al., "The interrelationships of placental mammals and the limits of phylogenetic inference," *Genome Biology and Evolution*, vol. 8, no. 2, pp. 330–344, 2016.

[26] W. Qian, H. Zhou, and K. Tang, "Recent coselection in human populations revealed by protein-protein interaction network," *Genome Biology and Evolution*, vol. 7, no. 1, pp. 136–153, 2014.

[27] C. B. Do, M. S. P. Mahabhashyam, M. Brudno, and S. Batzoglou, "ProbCons: probabilistic consistency-based multiple sequence alignment," *Genome Research*, vol. 15, no. 2, pp. 330–340, 2005.

[28] Q. Tu, R. A. Cameron, K. C. Worley, R. A. Gibbs, and E. H. Davidson, "Gene structure in the sea urchin *Strongylocentrotus purpuratus* based on transcriptome analysis," *Genome Research*, vol. 22, no. 10, pp. 2079–2087, 2012.

[29] D. Devos and A. Valencia, "Intrinsic errors in genome annotation," *Trends in Genetics*, vol. 17, no. 8, pp. 429–431, 2001.

[30] A. Schneider, A. Souvorov, N. Sabath, G. Landan, G. H. Gonnet, and D. Graur, "Estimates of positive darwinian selection are inflated by errors in sequencing, annotation, and alignment," *Genome Biology and Evolution*, vol. 1, pp. 114–118, 2009.

[31] P. Markova-Raina and D. Petrov, "High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila genomes*," *Genome Research*, vol. 21, no. 6, pp. 863–874, 2011.

[32] G. Jordan and N. Goldman, "The effects of alignment error and alignment filtering on the sitewise detection of positive selection," *Molecular Biology and Evolution*, vol. 29, no. 4, pp. 1125–1139, 2012.

[33] W. Fletcher and Z. Yang, "The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection," *Molecular Biology and Evolution*, vol. 27, no. 10, pp. 2257–2267, 2010.

[34] J. Castresana, "Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis," *Molecular Biology and Evolution*, vol. 17, no. 4, pp. 540–552, 2000.

[35] T. A. Hall, "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT," *Nucleic Acids Symposium Series*, vol. 41, pp. 95–98, 1999.

[36] Z. Yang, "Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A," *Journal of Molecular Evolution*, vol. 51, no. 5, pp. 423–432, 2000.

[37] Z. Yang, "PAML 4: Phylogenetic analysis by maximum likelihood," *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1586–1591, 2007.

[38] S. Whelan and N. Goldman, "Distributions of statistics used for the comparison of models of sequence evolution in phylogenetics," *Molecular Biology and Evolution*, vol. 16, no. 9, pp. 1292–1299, 1999.

[39] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society—Series B. Methodological*, vol. 57, no. 1, pp. 289–300, 1995.

[40] A. Chatr-Aryamontri, B.-J. Breitkreutz, R. Oughtred et al., "The BioGRID interaction database: 2015 update," *Nucleic Acids Research*, vol. 43, no. 1, pp. D470–D478, 2015.

[41] D. Szklarczyk, A. Franceschini, S. Wyder et al., "STRING v10: protein-protein interaction networks, integrated over the tree of life," *Nucleic Acids Research*, vol. 43, no. 1, pp. D447–D452, 2015.

[42] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[43] W. de Nooy, V. Batagelj, and A. Mrvar, *Exploratory Social Network Analysis with Pajek*, Cambridge University Press, Cambridge, UK, 2005.

[44] M. Wang, C. J. Herrmann, M. Simonovic, D. Szklarczyk, and C. von Mering, "Version 4.0 of PaxDb: protein abundance data, integrated across model organisms, tissues, and cell-lines," *Proteomics*, vol. 15, no. 18, pp. 3163–3168, 2015.

[45] U. Nagalakshmi, Z. Wang, K. Waern et al., "The transcriptional landscape of the yeast genome defined by RNA sequencing," *Science*, vol. 320, no. 5881, pp. 1344–1349, 2008.

[46] V. R. Chintapalli, J. Wang, and J. A. T. Dow, "Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease," *Nature Genetics*, vol. 39, no. 6, pp. 715–720, 2007.

[47] J. M. Cherry, E. L. Hong, C. Amundsen et al., "Saccharomyces Genome Database: the genomics resource of budding yeast," *Nucleic Acids Research*, vol. 40, no. 1, pp. D700–D705, 2012.

[48] H. Attrill, K. Falls, J. L. Goodman et al., "FlyBase: establishing a Gene Group resource for *Drosophila melanogaster*," *Nucleic Acids Research*, vol. 44, no. D1, pp. D786–D792, 2016.

[49] Y.-D. Li, H. Liang, Z. Gu et al., "Detecting positive selection in the budding yeast genome," *Journal of Evolutionary Biology*, vol. 22, no. 12, pp. 2430–2437, 2009.

[50] D. Alvarez-Ponce and M. A. Fares, "Evolutionary rate and duplicability in the *Arabidopsis thaliana* protein-protein interaction network," *Genome Biology and Evolution*, vol. 4, no. 12, pp. 1263–1274, 2012.

[51] B. Lemos, B. R. Bettencourt, C. D. Meiklejohn, and D. L. Hartl, "Evolution of proteins and gene expression levels are coupled in *Drosophila* and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions," *Molecular Biology and Evolution*, vol. 22, no. 5, pp. 1345–1354, 2005.

[52] J. D. Bloom and C. Adami, "Apparent dependence of protein evolutionary rate on number of interactions is linked to biases in protein-protein interactions data sets," *BMC Evolutionary Biology*, vol. 3, article 21, 2003.

[53] J. D. Bloom and C. Adami, "Evolutionary rate depends on number of protein-protein interactions independently of gene expression level: response," *BMC Evolutionary Biology*, vol. 4, article 14, 2004.

[54] H. B. Fraser, D. P. Wall, and A. E. Hirsh, "A simple dependence between protein evolution rate and the number of protein-protein interactions," *BMC Evolutionary Biology*, vol. 3, article 11, 2003.

[55] H. B. Fraser and A. E. Hirsh, "Evolutionary rate depends on number of protein-protein interactions independently of gene expression level," *BMC Evolutionary Biology*, vol. 4, article 13, 2004.

[56] H. B. Fraser, "Modularity and evolutionary constraint on proteins," *Nature Genetics*, vol. 37, no. 4, pp. 351–352, 2005.

[57] J. B. Plotkin and H. B. Fraser, "Assessing the determinants of evolutionary rates in the presence of noise," *Molecular Biology and Evolution*, vol. 24, no. 5, pp. 1113–1121, 2007.

[58] B. Lemos, C. D. Meiklejohn, and D. L. Hartl, "Regulatory evolution across the protein interaction network," *Nature Genetics*, vol. 36, no. 10, pp. 1059–1060, 2004.

[59] C. Kosiol, T. Vinař, R. R. da Fonseca et al., "Patterns of positive selection in six Mammalian genomes," *PLoS Genetics*, vol. 4, no. 8, Article ID e1000144, 2008.

[60] D. Ekman, S. Light, Å. K. Björklund, and A. Elofsson, "What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*?" *Genome Biology*, vol. 7, no. 6, article R45, 2006.

[61] M. Anisimova, J. P. Bielawski, and Z. Yang, "Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution," *Molecular Biology and Evolution*, vol. 18, no. 8, pp. 1585–1592, 2001.

[62] M. H. Schaefer, L. Serrano, and M. A. Andrade-Navarro, "Correcting for the study bias associated with protein-protein interaction measurements reveals differences between protein degree distributions from different cancer types," *Frontiers in Genetics*, vol. 6, article 260, 2015.

[63] E. Sprinzak, S. Sattath, and H. Margalit, "How reliable are experimental protein-protein interaction data?" *Journal of Molecular Biology*, vol. 327, no. 5, pp. 919–923, 2003.

[64] J. S. Bader, A. Chaudhuri, J. M. Rothberg, and J. Chant, "Gaining confidence in high-throughput protein interaction networks," *Nature Biotechnology*, vol. 22, no. 1, pp. 78–85, 2004.

[65] E. J. Deeds, O. Ashenberg, and E. I. Shakhnovich, "A simple physical model for scaling in protein-protein interaction networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 2, pp. 311–316, 2006.

[66] A. M. Larracuente, T. B. Sackton, A. J. Greenberg et al., "Evolution of protein-coding genes in *Drosophila*," *Trends in Genetics*, vol. 24, no. 3, pp. 114–123, 2008.

[67] L. Giot, J. S. Bader, C. Brouwer et al., "A protein interaction map of *Drosophila melanogaster*," *Science*, vol. 302, no. 5651, pp. 1727–1736, 2003.

[68] H. Yu, P. M. Kim, E. Sprecher, V. Trifonov, and M. Gerstein, "The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics," *PLoS Computational Biology*, vol. 3, no. 4, pp. 713–720, 2007.

[69] M. D. Daugherty and H. S. Malik, "Rules of engagement: molecular insights from host-virus arms races," *Annual Review of Genetics*, vol. 46, pp. 677–700, 2012.

[70] X. He and J. Zhang, "Toward a molecular understanding of pleiotropy," *Genetics*, vol. 173, no. 4, pp. 1885–1891, 2006.

[71] H. B. Fraser, A. E. Hirsh, L. M. Steinmetz, C. Scharfe, and M. W. Feldman, "Evolutionary rate in the protein interaction network," *Science*, vol. 296, no. 5568, pp. 750–752, 2002.

[72] D. Alvarez-Ponce, "The relationship between the hierarchical position of proteins in the human signal transduction network and their rate of evolution," *BMC Evolutionary Biology*, vol. 12, article 192, 2012.

[73] D. L. Hartl and C. H. Taubes, “Compensatory nearly neutral mutations: selection without adaptation,” *Journal of Theoretical Biology*, vol. 182, no. 3, pp. 303–309, 1996.

[74] J. Charlesworth and A. Eyre-Walker, “The other side of the nearly neutral theory, evidence of slightly advantageous back-mutations,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 43, pp. 16992–16997, 2007.

[75] H. A. Orr, “The genetic theory of adaptation: a brief history,” *Nature Reviews Genetics*, vol. 6, no. 2, pp. 119–127, 2005.

[76] H. A. Orr, “The population genetics of adaptation: the adaptation of DNA sequences,” *Evolution*, vol. 56, no. 7, pp. 1317–1330, 2002.

[77] D. A. Skelly, J. Ronald, C. F. Connelly, and J. M. Akey, “Population genomics of intron splicing in 38 *Saccharomyces cerevisiae* genome sequences,” *Genome Biology and Evolution*, vol. 1, pp. 466–478, 2009.

[78] J. A. Shapiro, W. Huang, C. Zhang et al., “Adaptive genic evolution in the *Drosophila* genomes,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 7, pp. 2271–2276, 2007.

[79] N. Yu, M. I. Jensen-Seaman, L. Chemnick, O. Ryder, and W.-H. Li, “Nucleotide diversity in gorillas,” *Genetics*, vol. 166, no. 3, pp. 1375–1383, 2004.

[80] M. Lynch, *The Origins of Genome Architecture*, Sinauer Associates, Sunderland, Mass, USA, 2007.

[81] H. A. Orr, “Adaptation and the cost of complexity,” *Evolution*, vol. 54, no. 1, pp. 13–20, 2000.

[82] T. Cavalier-Smith, *The Evolution of Genome Size*, John Wiley & Sons, New York, NY, USA, 1985.

[83] M. W. Hahn and G. A. Wray, “The g-value paradox,” *Evolution and Development*, vol. 4, no. 2, pp. 73–75, 2002.

[84] J. J. Welch and D. Waxman, “Modularity and the cost of complexity,” *Evolution*, vol. 57, no. 8, pp. 1723–1734, 2003.

[85] G. P. Wagner and J. Zhang, “The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms,” *Nature Reviews Genetics*, vol. 12, no. 3, pp. 204–213, 2011.

[86] J.-D. J. Han, N. Berlin, T. Hao et al., “Evidence for dynamically organized modularity in the yeast protein-protein interaction network,” *Nature*, vol. 430, no. 6995, pp. 88–93, 2004.

[87] I. W. Taylor, R. Linding, D. Warde-Farley et al., “Dynamic modularity in protein interaction networks predicts breast cancer outcome,” *Nature Biotechnology*, vol. 27, no. 2, pp. 199–204, 2009.

[88] M. Kellis, N. Patterson, M. Endrizzi, B. Birren, and E. S. Lander, “Sequencing and comparison of yeast species to identify genes and regulatory elements,” *Nature*, vol. 423, no. 6937, pp. 241–254, 2003.

[89] L. Ometto, A. Cestaro, S. Ramasamy et al., “Linking genomics and ecology to investigate the complex evolution of an invasive *Drosophila* pest,” *Genome Biology and Evolution*, vol. 5, no. 4, pp. 745–757, 2013.

[90] A. Scally, J. Y. Dutheil, L. W. Hillier et al., “Insights into hominid evolution from the gorilla genome sequence,” *Nature*, vol. 483, no. 7388, pp. 169–175, 2012.

[91] K. E. Langergraber, K. Prüfer, C. Rowney et al., “Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 39, pp. 15716–15721, 2012.