# Molecular Chaperones Accelerate the Evolution of Their Protein Clients in Yeast

David Alvarez-Ponce[1,2,*,†], José Aguilar-Rodríguez[3,4,†], and Mario A. Fares[2,5,‡]

[1]Biology Department, University of Nevada, Reno

[2]Instituto de Biología Molecular y Celular de Plantas, CSIC-UPV, Valencia, Spain

[3]Department of Biology, Stanford University, CA

[4]Department of Chemical and Systems Biology, Stanford University School of Medicine, CA

[5]Smurfit Institute of Genetics, University of Dublin, Trinity College Dublin, Ireland

[†]These authors contributed equally to this work.

[‡]Posthumous author.

*Corresponding author: E-mail: dap@unr.edu.

## Abstract

Protein stability is a major constraint on protein evolution. Molecular chaperones, also known as heat-shock proteins, can relax this constraint and promote protein evolution by diminishing the deleterious effect of mutations on protein stability and folding. This effect, however, has only been stablished for a few chaperones. Here, we use a comprehensive chaperone–protein interaction network to study the effect of all yeast chaperones on the evolution of their protein substrates, that is, their clients. In particular, we analyze how yeast chaperones affect the evolutionary rates of their clients at two very different evolutionary time scales. We first study the effect of chaperone-mediated folding on protein evolution over the evolutionary divergence of *Saccharomyces cerevisiae* and *S. paradoxus*. We then test whether yeast chaperones have left a similar signature on the patterns of standing genetic variation found in modern wild and domesticated strains of *S. cerevisiae*. We find that genes encoding chaperone clients have diverged faster than genes encoding non-client proteins when controlling for their number of protein–protein interactions. We also find that genes encoding client proteins have accumulated more intraspecific genetic diversity than those encoding non-client proteins. In a number of multivariate analyses, controlling by other well-known factors that affect protein evolution, we find that chaperone dependence explains the largest fraction of the observed variance in the rate of evolution at both evolutionary time scales. Chaperones affecting rates of protein evolution mostly belong to two major chaperone families: Hsp70s and Hsp90s. Our analyses show that protein chaperones, by virtue of their ability to buffer destabilizing mutations and their role in modulating protein genotype–phenotype maps, have a considerable accelerating effect on protein evolution.

**Key words:** protein evolution, molecular chaperones, mutational robustness, $d_N/d_S$.

## Introduction

Proteins within the proteome of any organism evolve at very different rates: whereas some proteins remain largely unaltered during long evolutionary periods, others can undergo fast evolutionary changes (Zuckerkandl and Pauling 1965; Zuckerkandl 1976; Li et al. 1985). The reasons for this diversity in rates of protein evolution are still a subject of intense debate (Rocha 2006; Alvarez-Ponce 2014; Zhang and Yang 2015). A number of factors have been shown to affect rates of evolution, including gene expression levels (Pál et al. 2001;

Drummond et al. 2005), expression breadth in multicellular organisms (Duret and Mouchiroud 2000; Wright et al. 2004; Zhang and Li 2004; Alvarez-Ponce and Fares 2012), essentiality (Hurst and Smith 1999; Jordan et al. 2002; Alvarez-Ponce et al. 2016; Aguilar-Rodríguez and Wagner 2018), duplicability (Nembaware et al. 2002; Yang et al. 2003; Pegueroles et al. 2013), and the number of protein–protein interactions (Fraser et al. 2002; Hahn and Kern 2005; Alvarez-Ponce and Fares 2012). However, a comprehensive understanding of which factors affect rates of protein evolution, their relative

impacts on rates of evolution, and the molecular mechanisms underlying these impacts, is lacking.

Molecular chaperones (Ellis 1987) help other proteins achieve their functional and 3D native conformations, prevent protein aggregation, and restore the native conformation of proteins destabilized by environmental perturbations (Hartl and Hayer-Hartl 2009; Hartl et al. 2011). As such, they can render neutral certain amino acid substitutions that would otherwise (in the absence of chaperones) be deleterious (or at least diminish their negative fitness effects) (Tokuriki and Tawfik 2009). Chaperones thus represent an extrinsic source of protein robustness: They can increase the tolerance of a protein phenotype (e.g., protein structure responsible for the protein function) against mutational insults. Therefore, chaperones can be not only a source of environmental robustness but also of mutational robustness (Jarosz et al. 2010; Lauring et al. 2013; Fares 2015; Payne and Wagner 2019). That is, chaperones can effectively buffer certain types of mutations in proteins, and thus are expected to contribute to the accumulation of genetic variation, and to increase the rates of evolution of their clients.

This increased rate of protein evolution of the clients of certain chaperones has been detected at the genomic level in a number of studies. Comparative analysis of bacterial genomes shows that the GroEL/ES chaperonin system can increase the evolutionary rate of its client proteins: after controlling for confounding factors, proteins that are clients of the system evolve faster on average than those that are not clients (Bogumil and Dagan 2010; Williams and Fares 2010). The bacterial DnaK also accelerates the rate of evolution of its clients (Aguilar-Rodríguez et al. 2016; Kadibalban et al. 2016). In yeast, Hsp90 clients evolve faster than their non-client paralogs (Lachowiec et al. 2013), and distinct groups of proteins interacting with different chaperones evolve at different rates (Bogumil et al. 2012). In mammals, kinases with higher binding affinity to Hsp90 evolve faster than kinases with lower binding affinity (Lachowiec et al. 2015). It has also been shown that both co- and posttranslationally acting chaperones can promote nonconservative amino acid substitutions, more likely destabilizing mutations, in their clients (Pechmann and Frydman 2014).

However, most studies so far have focused on individual chaperones and species, and the effect of most chaperones on protein evolution remains unknown. In this study, we evaluate the effect of all yeast protein chaperones on the evolution of their protein clients. We conducted a comprehensive analysis of the chaperone–client interaction network of 35 chaperones in yeast (Gong et al. 2009). This network was established with TAP-tag pulldown assays followed by both liquid chromatography tandem mass spectrometry (LC-MS/MS) and by matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF). We used this high-quality network to evaluate whether chaperone clients evolve faster in yeast, and also to measure the contribution of

different chaperone families to this acceleration of the rate of protein evolution. We show that many chaperones accelerate not only the rates of evolution of their clients but also their levels of nonsynonymous polymorphism.

## Materials and Methods

### Rates of Protein Evolution

The *S. cerevisiae* and *S. paradoxus* protein and coding (CDS) sequences were obtained from the *Saccharomyces* Genome Database (Cherry et al. 2012). Each *S. cerevisiae* protein sequence was used as query in a BLASTP search ($E$ value cutoff $= 10^{-10}$) against the *S. paradoxus* proteome. Similarly, each *S. paradoxus* protein was used in a BLASTP search against the *S. cerevisiae* proteome. Pairs of best reciprocal hits were considered to be encoded by pairs of orthologs. For each pair of orthologs, protein sequences were aligned using ProbCons (Do et al. 2005), and the resulting alignments were used to guide the alignment of the corresponding CDSs. PAML version 4.4d (codeml program, M0 model; Yang 2007) was used to estimate $d_N$, $d_S$, and $d_N/d_S$ values.

### Positive Selection Analyses

The *Saccharomyces cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, and *S. bayanus* protein and CDS sequences were obtained from the *Saccharomyces* Genome Database (Cherry et al. 2012). Each *S. cerevisiae* protein sequence was used as query in a BLASTP search ($E$ value cutoff $= 10^{-10}$) against the *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, and *S. bayanus* proteomes. Similarly, each *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, and *S. bayanus* protein was used in a BLASTP search against the *S. cerevisiae* proteome. Pairs of best reciprocal hits were considered to be encoded by pairs of orthologs. Only genes with putative orthologs in *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, and *S. bayanus* were retained for analysis. For each groups of orthologs, protein sequences were aligned using ProbCons (Do et al. 2005), and the resulting alignments were used to guide the alignment of the corresponding CDSs. Alignments were filtered as in a previous study (Luisi et al. 2015).

The filtered alignments were used in tests of positive selection using PAML version 4.4d (codeml program, M8 vs. M7 test; Yang et al. 2000). Twice the difference in the log-likelihood of both models was assumed to follow a $\chi^2$ distribution with two degrees of freedom. Genes with a $P$ value $<0.05$ and a fraction of codons with $d_N/d_S >1$ were assumed to be under positive selection. All computations were run using three starting $d_N/d_S$ values (0.04, 0.4, and 4) in order to alleviate the problem of local optima. The alignments corresponding to genes with signatures of positive selection were visualized using BioEdit version 7.2.5 in order to discard alignment or annotation errors.
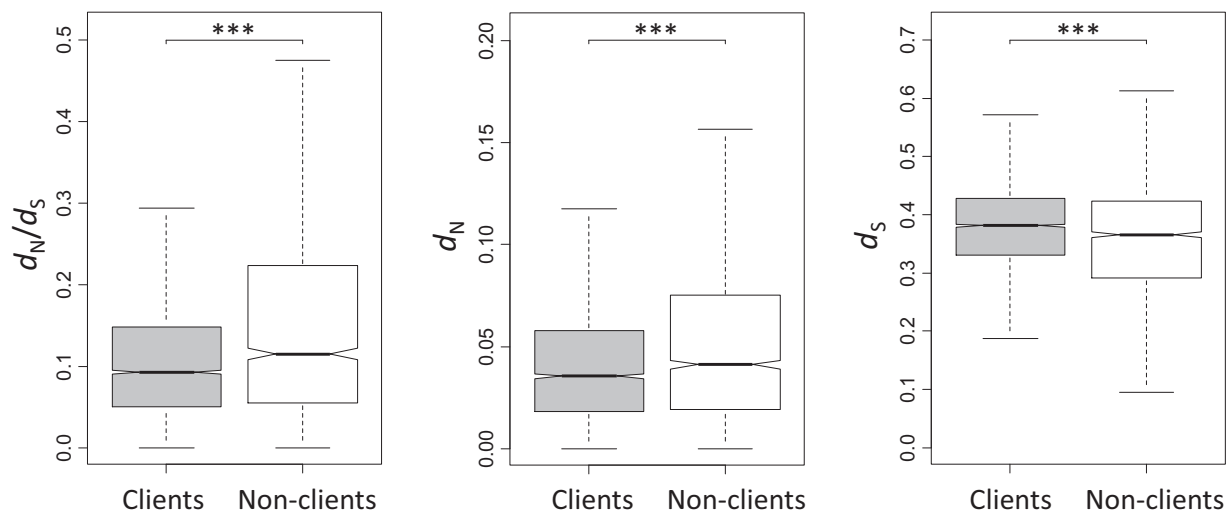
Fig. 1.—Rates of evolution of yeast chaperone clients and non-clients. Outliers (those above the 90th and below the 10th percentiles) are not shown. Significance levels: $*P < 0.05$, $**P < 0.001$, and $***P < 10^{-5}$.

**Table 1**

Comparison between Yeast Chaperone Clients and Non-clients

| | Chaperone Clients | | | Non-clients | | | P value |
|---|---|---|---|---|---|---|---|
| | *n* | **Mean** | **Median** | *n* | **Mean** | **Median** | |
| $d_N/d_S$ | 3,958 | 0.1165 | 0.0930 | 1,574 | **0.2563** | **0.1149** | $9.48 \times 10^{-22}$*** |
| $d_N$ | 3,958 | 0.0432 | 0.0355 | 1,574 | **0.0653** | **0.0411** | $5.66 \times 10^{-12}$*** |
| $d_S$ | 3,958 | **0.3795** | **0.3817** | 1,574 | 0.3722 | 0.3655 | $2.80 \times 10^{-11}$*** |
| Number of protein–protein interactions | 3,875 | **30.3130** | **16** | 1,265 | 18.7107 | 8 | $3.62 \times 10^{-53}$*** |
| Expression level | 3,434 | **71.1133** | **23** | 1,184 | 69.5845 | 20 | $1.71 \times 10^{-5}$*** |
| Protein length | 3,958 | **553.5682** | **462** | 1,574 | 327.5172 | 269 | $3.10 \times 10^{-138}$*** |

Note.—For each pair of clients versus non-client values, the higuest value is shown in bold face. *P* values correspond to the Mann–Whitney test.
$***P < 10^{-5}$.

## Chaperone Client Data

Chaperone–client interaction data were obtained from Gong et al. (2009). Their study included 35 chaperones and 29 co-chaperones. For each chaperone, we obtained a list of clients from their supplementary table 2.

## Additional Information

For each *S. cerevisiae* gene, the following information was gathered from different sources. The nonsynonymous to synonymous polymorphism ratio was obtained from Peter et al. (2018). For each gene, the average $d_N/d_S$ across all pairs of genomes was used (YN00). We obtained gene expression data for *S. cerevisiae* grown in rich media (YPAD) at 30 °C to mid exponential phase, where gene expression levels are measured as number of RNA-seq reads per gene length (Nagalakshmi et al. 2008). The number of protein–protein interactions (degree centrality) was obtained from the BioGRID database, version v3.2.101. Only physical, nonredundant interactions among *S. cerevisiae* proteins were included

in the analysis. Degrees were recomputed on a high-quality subnetwork, including those interactions determined by low-throughput studies or by more than one high-throughput study. A list of paralogs was obtained from Ensembl's Biomart (Kinsella et al. 2011), and genes with at least one paralog were classified as duplicates. A list of genes essential for growth in rich glucose media was obtained from Giaever et al. (2002).

## Statistical Analyses

Statistical analyses were conducted using the R package (R Core Team 2014). Partial correlation analyses were conducted using the "pcor.test" function (Kim 2015). We used the package "pls" to carry out the principal component regression analysis. We carried out base-10 logarithmic transformations of the continuous variables when such transformations led to a higher $R^2$. If a continuous variable contained values equal to zero, we added a small constant (0.001) to all its values to
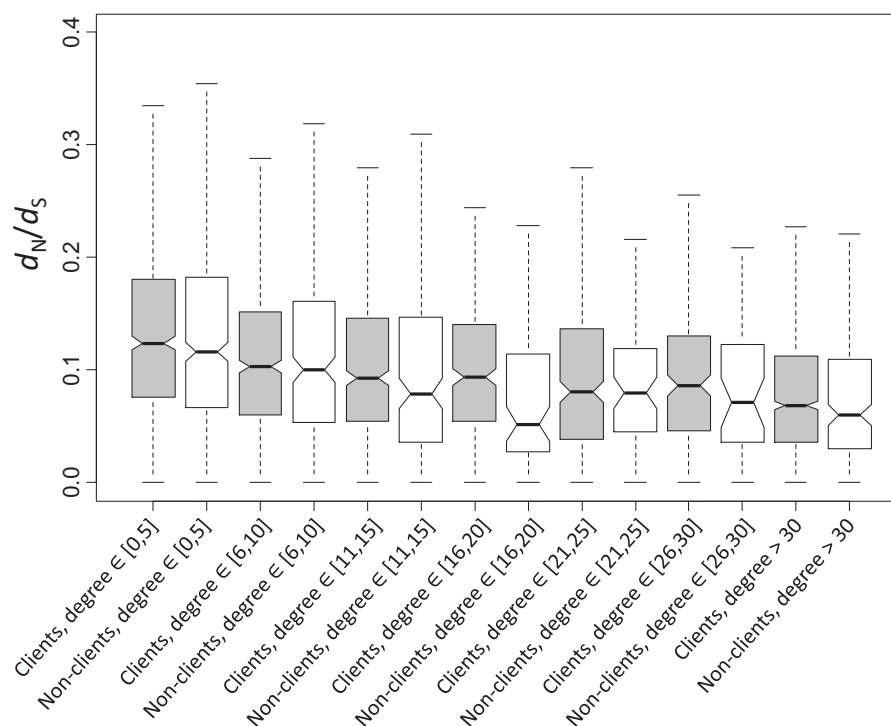
Fig. 2.—Comparison of the rate of evolution of clients and non-clients with different numbers of protein-protein interactions. Clients are represented in gray and non-clients in white. Outliers (those above the 90th and below the 10th percentiles) are not shown.
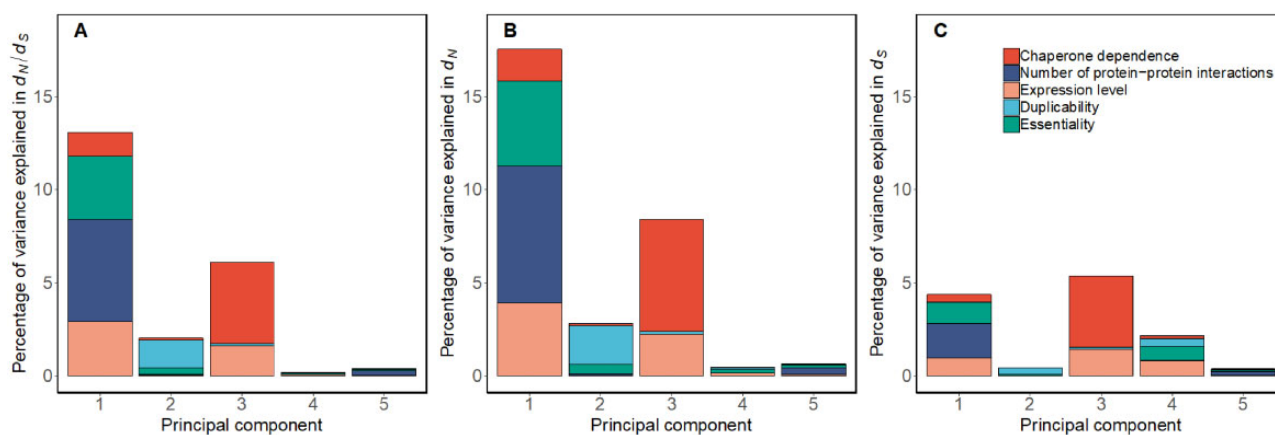


Fig. 3.—Principal component regression on (A) $d_N/d_S$, (B) $d_N$, and (C) $d_S$ calculated using divergence data between *Saccharomyces cerevisiae* and *S. paradoxus* for 5,532 yeast genes. For each principal component, the height of the bar represents the percent of variance in the rate of evolution explained by the component. The relative contribution of each variable to a principal component is represented with different colors. Table 3 contains the numerical data used to draw this figure.

allow its logarithmic transformation. We scaled the independent variables to zero mean and unit variance.

## Results

### Yeast Chaperone Clients Evolve Slower than Non-clients

We classified all *Saccharomyces cerevisiae* proteins into three classes: chaperones ($n = 35$), co-chaperones ($n = 29$), and others ($n = 6,653$), using the chaperone and co-chaperone list by Gong et al. (2009). The latter class was further classified into chaperone clients (those that interact with any of the chaperones according to the data set of Gong et al. 2009; $n = 4,209$) and non-clients (all remaining proteins, $n = 2,444$).

For each *S. cerevisiae* gene, the most likely ortholog in *S. paradoxus* was identified using a best-reciprocal-hit approach (see Materials and Methods), and the rate of protein evolution

was measured from the nonsynonymous to synonymous divergence ratio ($d_N/d_S$). These species diverged from a common ancestor ~5–10 Ma ago (Dori-Bachash et al. 2011). Orthologs could be identified for 5,603 of the *S. cerevisiae* genes. Values of $d_N/d_S > 8$ were removed, as they probably represent artifacts (ten genes were removed). The mean $d_N/d_S$ value was 0.1553, and the median was 0.0970, consistent with prior results (e.g., Alvarez-Ponce et al. 2017). After applying these filters, a total of 3,958 clients and 1,574 non-clients were available for analysis. All remaining genes were excluded from further analyses.

Clients exhibit substantially lower $d_N/d_S$ values (median: 0.0930) than non-clients (median: 0.1149; Mann–Whitney $U$ test, $P$ value $= 9.48 \times 10^{-22}$; fig. 1 and table 1). They also exhibit lower $d_N$ and higher $d_S$ values (fig. 1 and table 1). Next, we considered whether the number of chaperones of which each protein is client correlates with its rate of evolution. Among the 3,958 genes that have an ortholog in *S. paradoxus* and are clients of at least one chaperone, $d_N/d_S$ negatively correlates with the number of chaperones (Spearman's rank correlation coefficient, $\rho = -0.0784$, $P = 7.79 \times 10^{-7}$). The number of chaperones also correlates with $d_N$ ($\rho = -0.0596$, $P = 0.0002$) and, to a lesser extent, with $d_S$ ($\rho = 0.0323$, $P = 0.0422$).

We next considered whether chaperone clients may be enriched in proteins encoded by genes under positive selection. For each *S. cerevisiae* gene, we identified its most likely orthologs in another four species of the genus *Saccharomyces* (*S. paradoxus*, *Saccharomyces mikatae*, *Saccharomyces kudriavzevii*, and *Saccharomyces bayanus*). Only genes with a putative ortholog in all species ($n = 2,047$) were included in this analysis. The M8 versus M7 test (Yang 2000) was used to identify signatures of positive selection (see Materials and Methods). Among chaperone clients, 19 genes (3.40%) were encoded by genes under positive selection. Among non-clients, 72 (4.84%) were encoded by genes with signatures of positive selection. The fraction of genes under positive selection was not significantly different between clients and non-clients (Fisher's exact test, $P = 0.0967$).

## The Low Rate of Evolution of Chaperone Clients Is Not Due to Their Expression Levels, Essentiality, or Duplicability

Rates of protein evolution are affected by a number of factors, including expression levels (Pál et al. 2001; Drummond et al. 2005), gene essentiality (Hurst and Smith 1999; Jordan et al. 2002; Alvarez-Ponce et al. 2016), gene duplicability (Nembaware et al. 2002; Yang et al. 2003; Pegueroles et al. 2013), and number of protein–protein interactions (Fraser

**Table 2**

Multiple Linear Regression of Divergence Data

| | $d_N/d_S$ | $d_N$ | $d_S$ |
|---|---|---|---|
| Chaperone dependence | 0.17*** | 0.14*** | 0.06*** |
| Number of protein–protein interactions | −0.16*** | −0.13*** | −0.01* |
| Expression level | −0.34*** | −0.31*** | −0.09*** |
| Duplicability | −0.38*** | −0.34*** | −0.02* |
| Essentiality | −0.35*** | −0.30*** | −0.01 |

Note.—Regression coefficients are shown.
*$P < 0.05$ and ***$P < 10^{-5}$.

**Table 4**

Total Variance Explained by Each Variable in the Principal Component Regression Analysis of Divergence Data

| | $d_N/d_S$ | $d_N$ | $d_S$ |
|---|---|---|---|
| Chaperone dependence | 5.77% | 7.91% | 4.48% |
| Number of protein–protein interactions | 5.75% | 7.78% | 2.08% |
| Expression level | 4.72% | 6.47% | 3.27% |
| Duplicability | 1.68% | 2.36% | 0.86% |
| Essentiality | 3.93% | 5.37% | 2.03% |

**Table 3**

Results from the Principal Component Regression Analysis of Divergence Data

| | Principal Components | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | All |
| Percentage of explained variance in | | | | | | |
| $d_N/d_S$ | 13.09*** | 2.04*** | 6.11*** | 0.22*** | 0.39*** | 21.85 |
| $d_N$ | 17.55*** | 2.82*** | 8.41*** | 0.46*** | 0.66*** | 29.89 |
| $d_S$ | 4.36*** | 0.44*** | 5.38*** | 2.16*** | 0.38*** | 12.72 |
| Percent contributions of each variable | | | | | | |
| Chaperone dependence | 0.10 | 0.05 | **0.71** | 0.08 | 0.06 | |
| Number of protein–protein interactions | **0.42** | 0.02 | 0.00 | 0.02 | **0.54** | |
| Expression level | **0.22** | 0.02 | **0.27** | **0.37** | 0.12 | |
| Duplicability | 0.00 | **0.74** | 0.02 | 0.19 | 0.05 | |
| Essentiality | **0.26** | 0.18 | 0.00 | **0.34** | **0.22** | |

Note.—We indicate in bold the contributions of a predictor to a component when >20%.
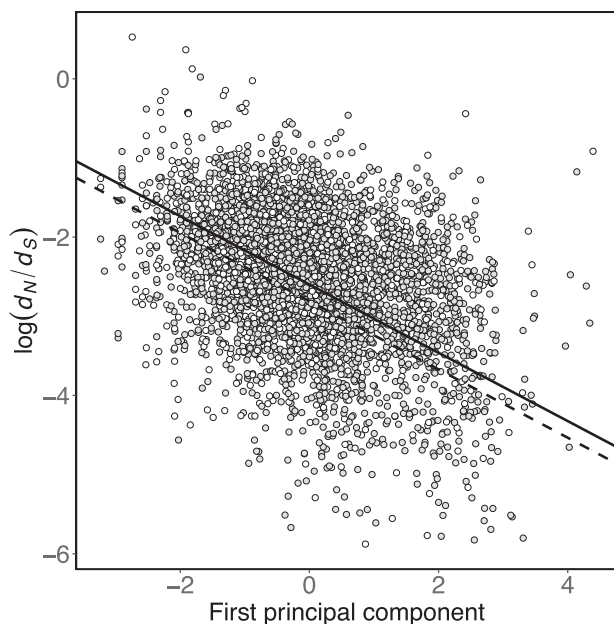***$P < 10^{-5}$.

Fig. 4.—ANCOVA. Chaperone clients (gray points, continuous line) evolve 23% above the genome average rate (light points, dashed line) when considering divergence data between *Saccharomyces cerevisiae* and *S. paradoxus*.

et al. 2002; Hahn and Kern 2005; Alvarez-Ponce and Fares 2012) (for review, see Rocha 2006; Alvarez-Ponce 2014; Zhang and Yang 2015). Clients and non-clients differ in all these parameters (table 1), and thus it is conceivable that the observed differences in the rates of evolution of clients and non-clients (fig. 1 and table 1) might be a byproduct of differences in these factors. In order to discard this possibility, we conducted a number of controls.

Expression level seems to be a major determinant of protein's rates of evolution, with highly expressed genes tending to be more selectively constrained (Pál et al. 2001; Drummond et al. 2005, 2006). In agreement with prior results, we observed a negative correlation between expression levels and $d_N/d_S$ ($\rho = -0.4138$, $P = 1.73 \times 10^{-190}$). Chaperone clients are more highly expressed than non-clients (median expression level for clients: 23; median expression level for non-clients: 20; Mann–Whitney test, $P = 1.71 \times 10^{-5}$). This raises the possibility that the lower rates of evolution of clients might be a byproduct of clients being more highly expressed. However, partial correlation analysis shows that the relationship between "chaperone dependence" (a dummy variable taking the value of 1 if the protein is client of at least one chaperone, and 0 otherwise) and $d_N/d_S$ is independent of expression level (partial Spearman's rank correlation coefficient, $\rho = -0.0414$, $P = 0.0049$). Furthermore, among chaperone clients, the partial correlation between $d_N/d_S$ and number of chaperones controlling for expression level is significantly negative ($\rho = -0.0643$, $P = 0.00016$).

Proteins encoded by essential genes tend to be more constrained than those encoded by nonessential genes (Hurst and Smith 1999; Alvarez-Ponce et al. 2016). Among the 3,958 chaperone clients with $d_N/d_S$ information, 831 (i.e., 21%) are essential. Among the 1,574 non-clients, only 228 (14.5%) are essential. Thus, clients are enriched in essential genes (Fisher's exact test, $P < 10^{-6}$), which could potentially explain their low evolutionary rates. To discard this possibility, we analyzed essential and nonessential genes separately, and in both cases clients exhibited a lower $d_N/d_S$. Among essential genes, the median $d_N/d_S$ was 0.0692 for clients and 0.0913 for non-clients (Mann–Whitney test, $P = 0.0016$). Among nonessential genes, the median $d_N/d_S$ was 0.0990 for clients and 0.1179 for non-clients (Mann–Whitney test, $P = 2.06 \times 10^{-16}$).

Proteins encoded by duplicated genes tend to evolve slower than those encoded by singleton genes (Nembaware et al. 2002; Yang et al. 2003), in spite of the fact that gene duplication transiently accelerates protein evolution (Han et al. 2009; Pegueroles et al. 2013). Among clients, 1,684 (42.54%) are encoded by duplicated genes, and among non-clients, 547 (34.75%) are encoded by duplicated genes; that is, clients are enriched in proteins encoded by duplicated genes (Fisher's exact test, $P < 10^{-6}$), which might account for their slow evolution. To discard this possibility, we analyzed singleton and duplicated genes separately. Among singletons, clients exhibit lower $d_N/d_S$ values (median = 0.1048) than non-clients (median = 0.1462; Mann–Whitney $U$ test, $P = 9.45 \times 10^{-27}$). Among the less numerous duplicates, clients also exhibited lower $d_N/d_S$ values, but the differences were not significant (median for clients: 0.0752, median for non-clients: 0.0786, $P = 0.3210$). In addition, among clients, the number of chaperones significantly correlates with $d_N/d_S$, among both singletons ($\rho = -0.0705$, $P = 0.0008$) and duplicates ($\rho = -0.0745$, $P = 0.0022$). These results indicate that the lower rates of evolution of chaperone clients are not due to their enrichment in proteins encoded by duplicated genes.

## Controlling for Number of Physical Interactions Reveals That Chaperone Dependence Accelerates Protein Evolution

The number of protein–protein interactions with which a protein interacts (degree centrality) negatively correlates with its rate of evolution (Fraser et al. 2002; Hahn and Kern 2005; Alvarez-Ponce and Fares 2012), a pattern that was also apparent in our data set ($\rho = -0.2788$, $P = 2.14 \times 10^{-92}$). This, together with the fact that chaperone clients tend to exhibit more protein–protein interactions (median = 16) than non-clients (median = 8; Mann–Whitney $U$ test, $P = 3.62 \times 10^{-53}$), might account for the low rates of evolution of chaperone clients.

Indeed, the partial correlation between $d_N/d_S$ and chaperone dependence while controlling for degree is significantly

**Table 5**

Comparison between the Rates of Evolution of Clients of Different Yeast Chaperones and Proteins That Are Not Clients of Any Chaperone

| Class | Chaperone | Clients | | | Non-clients | | | Mann–Whitney | | Partial Correlation | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | n | Median | Mean | n | Median | Mean | P value | Q value | $\rho$ | P value | Q value |
| CCTs | Cct2 | 120 | 0.0807 | 0.0953 | 1,574 | 0.1149 | 0.2563 | $1.64 \times 10^{-6}$*** | $2.61 \times 10^{-6}$*** | 0.0289 | 0.2822 | 0.4490 |
| CCTs | Cct3 | 119 | 0.0800 | 0.0993 | 1,574 | 0.1149 | 0.2563 | $3.41 \times 10^{-5}$** | $4.26 \times 10^{-5}$** | 0.0311 | 0.2481 | 0.4135 |
| CCTs | Cct4 | 158 | 0.0861 | 0.1051 | 1,574 | 0.1149 | 0.2563 | $5.52 \times 10^{-5}$** | $6.66 \times 10^{-5}$** | 0.0560 | 0.0345* | 0.1098 |
| CCTs | Cct5 | 34 | 0.0603 | 0.0769 | 1,574 | 0.1149 | 0.2563 | 0.0002** | 0.0002** | −0.0225 | 0.4177 | 0.5629 |
| CCTs | Cct6 | 92 | 0.0689 | 0.0822 | 1,574 | 0.1149 | 0.2563 | $5.00 \times 10^{-8}$*** | $1.17 \times 10^{-7}$*** | −0.0190 | 0.4850 | 0.5853 |
| CCTs | Cct7 | 40 | 0.0461 | 0.0611 | 1,574 | 0.1149 | 0.2563 | $1.49 \times 10^{-7}$*** | $2.90 \times 10^{-7}$*** | −0.0542 | 0.0502 | 0.1255 |
| CCTs | Cct8 | 179 | 0.0848 | 0.1016 | 1,574 | 0.1149 | 0.2563 | $6.07 \times 10^{-7}$*** | $1.12 \times 10^{-6}$*** | 0.0390 | 0.1389 | 0.2749 |
| CCTs | Tcp1 | 46 | 0.0550 | 0.0776 | 1,574 | 0.1149 | 0.2563 | $7.91 \times 10^{-6}$*** | $1.15 \times 10^{-5}$** | −0.0241 | 0.3832 | 0.5588 |
| Hsp70s | Ecm10 | 64 | 0.0633 | 0.0799 | 1,574 | 0.1149 | 0.2563 | $2.08 \times 10^{-6}$*** | $3.17 \times 10^{-6}$*** | −0.0215 | 0.4342 | 0.5629 |
| Hsp70s | Kar2 | 68 | 0.0694 | 0.1000 | 1,574 | 0.1149 | 0.2563 | 0.0002** | 0.0002** | 0.0274 | 0.3181 | 0.4841 |
| Hsp70s | Lhs1 | 78 | 0.0782 | 0.0837 | 1,574 | 0.1149 | 0.2563 | $1.28 \times 10^{-6}$*** | $2.13 \times 10^{-6}$*** | 0.0042 | 0.8781 | 0.8781 |
| Hsp70s | Ssa1 | 2,385 | 0.0903 | 0.1098 | 1,574 | 0.1149 | 0.2563 | $6.43 \times 10^{-22}$*** | $1.13 \times 10^{-20}$*** | 0.0606 | 0.0003** | 0.0013* |
| Hsp70s | Ssa2 | 1,828 | 0.0907 | 0.1104 | 1,574 | 0.1149 | 0.2563 | $4.69 \times 10^{-19}$*** | $4.10 \times 10^{-18}$*** | 0.0742 | $3.86 \times 10^{-5}$** | 0.0004** |
| Hsp70s | Ssa3 | 304 | 0.0949 | 0.1129 | 1,574 | 0.1149 | 0.2563 | $2.70 \times 10^{-5}$*** | $3.50 \times 10^{-5}$** | 0.0955 | 0.0002** | 0.0010* |
| Hsp70s | Ssa4 | 436 | 0.0918 | 0.1088 | 1,574 | 0.1149 | 0.2563 | $6.75 \times 10^{-8}$*** | $1.39 \times 10^{-7}$*** | 0.0965 | 0.0001** | 0.0006** |
| Hsp70s | Ssb1 | 3,109 | 0.0916 | 0.1148 | 1,574 | 0.1149 | 0.2563 | $1.51 \times 10^{-21}$*** | $1.76 \times 10^{-20}$*** | 0.0604 | $7.02 \times 10^{-5}$** | 0.0005** |
| Hsp70s | Ssb2 | 1,167 | 0.0881 | 0.1116 | 1,574 | 0.1149 | 0.2563 | $1.18 \times 10^{-17}$*** | $6.88 \times 10^{-17}$*** | 0.0833 | $4.01 \times 10^{-5}$*** | 0.0004** |
| Hsp70s | Ssc1 | 191 | 0.0603 | 0.0827 | 1,574 | 0.1149 | 0.2563 | $5.22 \times 10^{-15}$*** | $2.61 \times 10^{-14}$*** | −0.0165 | 0.5285 | 0.5967 |
| Hsp70s | Sse1 | 1,862 | 0.0913 | 0.1098 | 1,574 | 0.1149 | 0.2563 | $5.32 \times 10^{-18}$*** | $3.72 \times 10^{-17}$*** | 0.0742 | $3.45 \times 10^{-5}$** | 0.0004** |
| Hsp70s | Sse2 | 234 | 0.0812 | 0.0990 | 1,574 | 0.1149 | 0.2563 | $3.13 \times 10^{-8}$*** | $7.83 \times 10^{-8}$*** | 0.0409 | 0.1138 | 0.2489 |
| Hsp70s | Ssq1 | 91 | 0.0621 | 0.0766 | 1,574 | 0.1149 | 0.2563 | $1.82 \times 10^{-9}$*** | $5.79 \times 10^{-9}$*** | −0.0318 | 0.2413 | 0.4135 |
| Hsp70s | Ssz1 | 636 | 0.0948 | 0.1099 | 1,574 | 0.1149 | 0.2563 | $5.16 \times 10^{-9}$*** | $1.51 \times 10^{-8}$*** | 0.1001 | $1.19 \times 10^{-5}$** | 0.0004** |
| Hsp90s | Hsc82 | 421 | 0.0810 | 0.0986 | 1,574 | 0.1149 | 0.2563 | $1.97 \times 10^{-13}$*** | $8.62 \times 10^{-13}$*** | 0.0482 | 0.0480* | 0.1255 |
| Hsp90s | Hsp82 | 832 | 0.0886 | 0.1157 | 1,574 | 0.1149 | 0.2563 | $2.58 \times 10^{-13}$*** | $1.00 \times 10^{-12}$*** | 0.0696 | 0.0015* | 0.0058* |
| Hsp100s | Hsp78 | 773 | 0.0779 | 0.0958 | 1,574 | 0.1149 | 0.2563 | $4.45 \times 10^{-24}$*** | $1.56 \times 10^{-22}$*** | 0.0398 | 0.0730 | 0.1703 |
| Hsp100s | Hsp104 | 358 | 0.0877 | 0.1079 | 1,574 | 0.1149 | 0.2563 | $5.36 \times 10^{-8}$*** | $1.17 \times 10^{-7}$*** | 0.0719 | 0.0037* | 0.0130* |
| Small | Hsp31 | 98 | 0.0848 | 0.0987 | 1,574 | 0.1149 | 0.2563 | 0.0004** | 0.0004** | 0.0319 | 0.2385 | 0.4135 |
| Small | Hsp32 | 2 | 0.0635 | 0.0639 | 1,574 | 0.1149 | 0.2563 | 0.3045 | 0.3045 | −0.0100 | 0.7214 | 0.7651 |
| Small | Hsp33 | 3 | 0.0363 | 0.0454 | 1,574 | 0.1149 | 0.2563 | 0.0739 | 0.0783 | −0.0221 | 0.4315 | 0.5629 |
| Small | Sno4 | 3 | 0.0261 | 0.0622 | 1,574 | 0.1149 | 0.2563 | 0.1723 | 0.1773 | −0.0061 | 0.8284 | 0.8528 |
| Other | Hsp12 | 91 | 0.0731 | 0.0930 | 1,574 | 0.1149 | 0.2563 | $2.21 \times 10^{-5}$** | $2.98 \times 10^{-5}$** | 0.0180 | 0.5076 | 0.5922 |
| Other | Hsp26 | 85 | 0.0651 | 0.0806 | 1,574 | 0.1149 | 0.2563 | $1.60 \times 10^{-8}$*** | $4.31 \times 10^{-8}$*** | −0.0205 | 0.4523 | 0.5654 |
| Other | Hsp42 | 361 | 0.0839 | 0.1044 | 1,574 | 0.1149 | 0.2563 | $8.57 \times 10^{-10}$*** | $3.00 \times 10^{-9}$*** | 0.0493 | 0.0468* | 0.1255 |
| Other | Hsp60 | 95 | 0.0777 | 0.0917 | 1,574 | 0.1149 | 0.2563 | $1.59 \times 10^{-5}$** | $2.23 \times 10^{-5}$** | 0.0159 | 0.5573 | 0.6095 |
| Other | Mcx1 | 41 | 0.0511 | 0.0718 | 1,574 | 0.1149 | 0.2563 | $1.27 \times 10^{-6}$*** | $2.13 \times 10^{-6}$*** | −0.0407 | 0.1414 | 0.2749 |

Note.—For each pair of clients versus non-clients, the highest $d_N/d_S$ values are shown in bold face. Partial correlations correspond to the Spearman correlation between chaperone dependence and $d_N/d_S$, controlling for number of protein–protein interactions. Q values were computed using the Benjamini–Hochberg approach (Benjamini and Hochberg 1995).
*$P < 0.05$, **$P < 0.001$, and ***$P < 10^{-5}$.

**Table 6**

Comparison between the Rates of Evolution of Clients of Different Chaperone Families and Proteins That Are Not Clients of Any Chaperone

| Class | Clients | | | Non-clients | | | Mann–Whitney | | Partial Correlation | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | Median | Mean | $n$ | Median | Mean | P value | Q value | $\rho$ | P value | Q value |
| CCTs | 614 | 0.0794 | 0.0967 | 1,574 | **0.1149** | **0.2563** | $2.55 \times 10^{-20}$*** | $4.25 \times 10^{-20}$*** | 0.0349 | 0.1310 | 0.1638 |
| Hsp70s | 3,783 | 0.0932 | 0.1156 | 1,574 | **0.1149** | **0.2563** | $2.66 \times 10^{-21}$*** | $6.65 \times 10^{-21}$*** | 0.0550 | 0.0001** | 0.0005** |
| Hsp90s | 1,101 | 0.0861 | 0.1115 | 1,574 | **0.1149** | **0.2563** | $1.27 \times 10^{-17}$*** | $1.59 \times 10^{-17}$*** | 0.0615 | 0.0028* | 0.0070* |
| Hsp100s | 1,004 | 0.0824 | 0.1005 | 1,574 | **0.1149** | **0.2563** | $2.91 \times 10^{-23}$*** | $1.46 \times 10^{-22}$*** | 0.0537 | 0.0106* | 0.0176* |
| Small | 104 | 0.0809 | 0.0966 | 1,574 | **0.1149** | **0.2563** | 0.0001** | 0.0001** | 0.0285 | 0.2926 | 0.2926 |

NOTE.—For each pair of clients versus non-clients, the highest $d_N/d_S$ values are shown in bold face. Partial correlations correspond to the Spearman correlation between chaperone dependence and $d_N/d_S$, controlling for number of protein–protein interactions. Q values were computed using the Benjamini–Hochberg approach (Benjamini and Hochberg 1995).

*$P < 0.05$, **$P < 0.001$, and ***$P < 10^{-5}$.

**Table 7**

Multiple Linear Regression of Different Chaperone Families

| | $d_N/d_S$ | $d_N$ | $d_S$ |
|---|---|---|---|
| HSP70 dependence | 0.15*** | 0.12*** | 0.04** |
| HSP90 dependence | 0.10* | 0.06 | 0.02 |
| HSP100 dependence | 0.04 | 0.05 | 0.05*** |
| CTT dependence | −0.03 | −0.04 | −0.01 |
| SMALL dependence | −0.10 | 0.02 | 0.03 |
| Number of protein–protein interactions | −0.16*** | −0.13*** | −0.01* |
| Expression level | −0.34*** | −0.31*** | −0.09*** |
| Duplicability | −0.38*** | −0.34*** | −0.02* |
| Essentiality | −0.35*** | −0.30*** | −0.01 |

*$P < 0.05$, **$P < 0.001$, and ***$P < 10^{-5}$.

positive ($\rho = 0.0507$, $P = 0.0003$), as is the partial correlation between the $d_N/d_S$ values of clients and their number of chaperones while controlling for degree ($\rho = 0.0181$, $P = 2.71 \times 10^{-6}$). These results indicate that chaperones accelerate the rates of evolution of their clients.

We repeated these analyses using degree values computed from a subset of protein–protein interactions of high quality (interactions identified either by low-throughput screens or by two or more high-throughput screens). This reduced the number of genes for which available network data were available from 5,140 to 4,011. The partial correlation between $d_N/d_S$ and chaperone dependence while controlling for degree remains significantly positive ($\rho = 0.0405$, $P = 0.0104$), while the correlation between $d_N/d_S$ and the number of chaperones with which clients interact was not significant ($\rho = 0.0055$, $P = 0.7546$).

To further validate our results, we binned proteins into seven degree classes: 1–5 interactions (744 clients and 480 non-clients), 6–10 interactions (681 clients and 246 non-clients), 11–15 interactions (499 clients and 158 non-clients), 16–20 interactions (330 clients and 80 non-clients), 21–25 interactions (280 clients and 71 non-clients), 26–30 interactions (203 clients and 40 non-clients), and >30 interactions

(1,138 clients and 190 non-clients). Within each of the classes, chaperone clients exhibited a higher median $d_N/d_S$ than non-clients (fig. 2), with significant differences in the classes of degree 15–20 (one-tailed Mann–Whitney test, $P = 4.30 \times 10^{-5}$) and degree >30 ($P = 0.0385$). In addition, the observation that in all seven categories clients have a higher median $d_N/d_S$ is not expected at random (binomial test, $P = 0.0156$).

## Multivariate Analyses Confirm the Accelerating Effect of Chaperones on the Evolution of Their Clients

We performed a multivariate regression analysis to study the relative influence of all the studied factors (chaperone dependence, expression level, number of protein–protein interactions, duplicability, and essentiality) simultaneously. We regressed $d_N/d_S$ against the five biological factors, and found that all make a significant contribution to the regression and that the overall $R^2$ is 0.219 (table 2). Chaperone dependence was the only factor with a positive coefficient, indicating that chaperone dependence increases protein evolutionary rates. Multivariate regression assumes that the predictor variables are statistically independent. To evaluate if our predictors intercorrelate, we used the variance inflation factor (VIF) to quantify the degree of collinearity. We found VIF values for each of the predictor variables that range from 1.03 to 1.26, which indicates that while collinearity is present in our model, it is rather low. Nevertheless, multivariate regression can produce spurious results in the presence of both collinearity and noise (Drummond et al. 2006), and our variables are affected by noisy measurements. Therefore, we also performed a principal component regression analysis, which is an established method to study the relative contributions of different determinants of protein evolutionary rates (Drummond et al. 2006), although it is not entirely insensitive to noise (Plotkin and Fraser 2007). Principal component regression finds new variables, called principal components, which are linear combinations of the original predictor variables, and then regresses the response variable against all of them. We performed principal component regression using the same
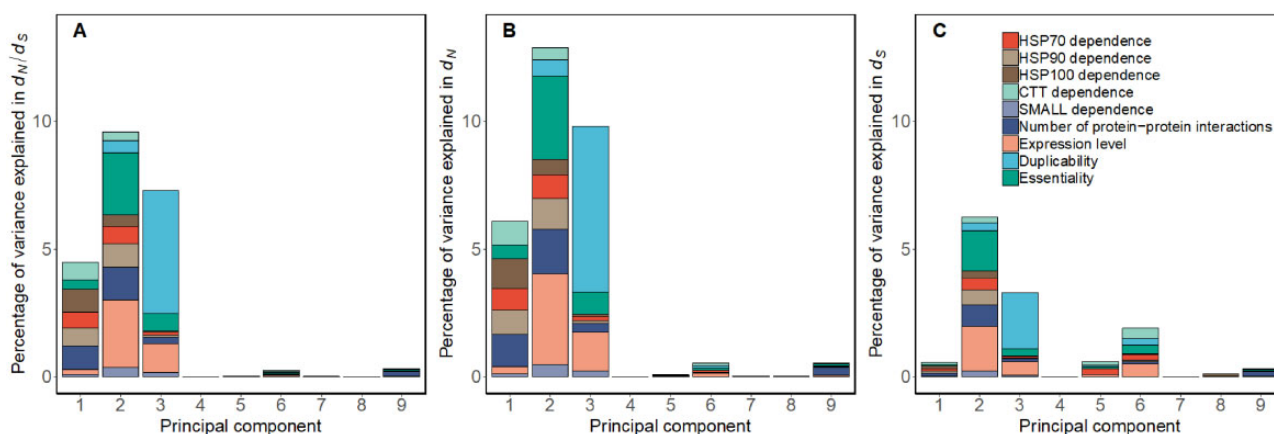
Fig. 5.—Principal component regression on (A) $d_N/d_S$, (B) $d_N$, and (C) $d_S$ calculated using divergence data between *Saccharomyces cerevisiae* and *S. paradoxus* for 5,532 yeast genes. For each principal component, the height of the bar represents the percent of variance in the rate of evolution explained by the component. The relative contribution of each variable to a principal component is represented with different colors. Table 8 contains the numerical data used to draw this figure.

predictor variables as above. Table 3 shows numerical data from the analysis, while figure 3 shows these data graphically.

For evolutionary rates measured as $d_N/d_S$, we found a principal component with a ~70% contribution of chaperone dependence and ~30% of expression level. This component explained a modest 6% of the variance with high significance (table 2 and fig. 3A). Another significant principal component explains 13% of the variance. This component is mainly determined by the number of protein–protein interactions, essentiality, and expression level. A component explaining just ~2% of the variance was mainly determined by duplicability. The other two significant components explained in combination <1% of the variance. In summary, we found that chaperone dependence was the biological factor explaining the largest fraction of the total variance in the rate of evolution measured as $d_N/d_S$ (5.77%) (table 4). It explained a larger fraction of the total variance than expression level (4.72%), and similar to the fraction explained by the number of protein–protein interactions (5.75%). Similar results were observed for $d_N$ (tables 3 and 4; fig. 3). For $d_S$, chaperone dependence was still the main factor explaining the total variance in the rate of evolution, with a contribution of 4.48%—still above that of expression level (3.27%) (table 4). Indeed, it was the main determinant (~70%) of the principal component explaining the largest fraction of the variance (5.38%) (table 3).

Finally, we performed an analysis of covariance (ANCOVA), which is a category-based analysis in which we evaluated the effect of chaperone dependence on the rate of protein evolution measured as $d_N/d_S$ while controlling for the effect of the most important predictors: number of protein–protein interactions, expression level, and essentiality. We used the principal component of these three variables (principal component 1 in table 3 and fig. 3) as the continuous variable in

the ANCOVA. We found that chaperone clients evolve on average 23% faster than all proteins ($P = 8.6 \times 10^{-7}$) (fig. 4).

## Separate Analysis of the Clients of Individual Chaperones

Thus far, we have aggregated the clients of all chaperones into a single group. However, different chaperones may affect the rates of protein evolution in different ways. We thus considered the clients of each chaperone separately. For each chaperone, we compared the clients of the chaperone against the proteins that are not clients of any chaperone. We again found that in all 35 cases clients exhibit a lower median and average $d_N/d_S$, with significant differences in 32 cases (Mann–Whitney $U$ test, $P < 0.05$; table 5). However, partial correlations between $d_N/d_S$ and chaperone dependence using degree as controlling variable were positive in 23 cases (significantly positive in 13 cases) and negative in 12 cases (significantly negative in 0 cases). This approach has the limitations that some chaperones have very few known clients, and that clients of the chaperone of interest may also be clients of other chaperones.

## Analysis of the Clients of Different Groups of Chaperones

We next grouped chaperones into five groups: small Hsps (Hsp31, Hsp32, Hsp33, and Sno4), Hsp70s (Kar2, Ssb1, Sse1, Sse2, Ssa1, Ssa2, Ssa3, Ssa4, Ssb2, Ecm10, Ssc1, Ssq1, Ssz1, and Lhs1), Hsp90s (Hsp82 and Hsc82), Hsp100s (Hsp78 and Hsp104), and CCTs (Tcp1, Cct4, Cct8, Cct2, Cct3, Cct5, Cct6, and Cct7), and investigated the rates of evolution of the clients of each group. Single-family chaperones (Hsp26, Hsp42, Hsp12, Mcx1, and Hsp60) were not included in this analysis.

For each group of chaperones, we compared the rates of evolution of proteins that are clients of any of the chaperones

**Table 8**

Results from the Principal Component Regression of Different Chaperone Families

| | Principal Components | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | All |
| Percentage of explained variance in | | | | | | | | | | |
| $d_N/d_S$ | 4.48*** | 9.58*** | 7.28*** | 0.01 | 0.04 | 0.26*** | 0.02 | 0.01 | 0.32*** | 21.99 |
| $d_N$ | 6.08*** | 12.87*** | 9.80*** | 0.00 | 0.09* | 0.54*** | 0.04 | 0.04 | 0.55*** | 29.99 |
| $d_S$ | 0.56*** | 6.25*** | 3.28*** | 0.00 | 0.60*** | 1.90*** | 0.01 | 0.12* | 0.31*** | 13.03 |
| Percent contributions of each variable | | | | | | | | | | |
| HSP70 dependence | 0.14 | 0.07 | 0.02 | 0.03 | **0.37** | 0.12 | | | | |
| HSP90 dependence | 0.16 | 0.10 | 0.01 | 0.01 | 0.00 | 0.03 | | | | |
| HSP100 dependence | **0.20** | 0.05 | 0.01 | 0.00 | 0.00 | 0.01 | | | | |
| CTT dependence | 0.15 | 0.04 | 0.00 | 0.00 | **0.24** | **0.22** | | | | |
| SMALL dependence | 0.02 | 0.04 | 0.02 | 0.90 | 0.00 | 0.01 | | | | |
| Number of protein–protein interactions | **0.21** | 0.13 | 0.03 | 0.00 | 0.00 | 0.05 | | | | |
| Expression level | 0.05 | **0.28** | 0.00 | 0.00 | 0.14 | **0.25** | | | | |
| Duplicability | 0.00 | 0.05 | 0.04 | 0.04 | 0.07 | 0.13 | | | | |
| Essentiality | 0.26 | **0.25** | 0.01 | 0.01 | 0.17 | 0.19 | | | | |

NOTE.—We indicate in bold the contributions of a predictor to a component when >20%.

*$P < 0.05$ and ***$P < 10^{-5}$.

**Table 9**

Total Variance Explained by Each Variable in the Principal Component Regression Analysis of Different Chaperone Families

| | $d_N/d_S$ | $d_N$ | $d_S$ |
|---|---|---|---|
| HSP70 dependence | 1.48% | 2.04% | 1.05% |
| HSP90 dependence | 1.71% | 2.32% | 0.80% |
| HSP100 dependence | 1.43% | 1.95% | 0.54% |
| CTT dependence | 1.10% | 1.54% | 0.88% |
| SMALL dependence | 0.64% | 0.84% | 0.35% |
| Number of protein–protein interactions | 2.66% | 3.66% | 1.35% |
| Expression level | 4.07% | 5.55% | 2.86% |
| Duplicability | 5.36% | 7.25% | 2.79% |
| Essentiality | 3.55% | 4.84% | 2.43% |

of the group, against proteins that are not clients of any chaperone. In all five cases, clients had a significantly lower $d_N/d_S$. However, partial correlations between the dependence of each group and $d_N/d_S$ controlling for degree were always positive, and significant for the three chaperone classes with more clients (Hsp70s, Hsp90s, and Hsp100s) (table 6). This approach has the limitation that clients of one group of chaperones may also be clients of chaperones outside that group.

Next, in order to tease apart the effects of the different chaperone groups on rates of protein evolution while controlling for possible confounding factors, we performed two different multivariable analyses. We first performed a multiple linear regression analysis regressing $d_N/d_S$ against the four confounding biological factors we consider here (number of protein–protein interactions, expression level, essentiality, and duplicability), and dependence of the five chaperone families (Hsp70s, Hsp90s, Hsp100s, CTTs, and small Hsps). We found that among the chaperone families only Hsp70s and Hsp90s

make a significant contribution to the regression and that the overall $R^2$ is 0.220 (table 7). Hsp70s and Hsp90s dependence were the only factors with a positive coefficient, indicating that dependence on these two major chaperone groups increases protein evolutionary rates. The contribution of Hsp90s was lost when regressing $d_N$ or $d_S$ instead of $d_N/d_S$ (table 7).

We then performed a principal component regression analysis using the same predictor variables as above. Table 8 shows numerical data from the analysis, while figure 5 shows these data graphically. Neither Hsp70s dependence nor Hsp90s dependence contributed individually >20% to any significant principal component, but in combination they determine 30% of a component explaining 4.48% of the variance in $d_N/d_S$ (table 8). In combination, Hsp70s and Hsp90s dependence contribute 3.19% to the total variance in the rate of evolution, which is above the contribution of the number of protein–protein interactions, but below the contributions of expression level, essentiality, or duplicability (table 9).

## Chaperones Increase the Ratio of Nonsynonymous to Synonymous Polymorphism Ratio

For each *S. cerevisiae* gene, we obtained the nonsynonymous to synonymous polymorphism ratio ($d_N/d_S$) from Peter et al. (2018). Chaperone clients exhibit a significantly lower $d_N/d_S$ ratio (median for clients: 0.2352, median for non-clients: 0.2642, Mann–Whitney U test, $P = 2.96 \times 10^{-10}$). Partial correlation between $d_N/d_S$ and chaperone dependence controlling for expression level was nonsignificant ($\rho = -0.0015$, $P = 0.9159$), and the partial correlation between $d_N/d_S$ and chaperone dependence controlling for network degree was significantly positive ($\rho = 0.0612$, $P = 10^{-5}$).

**Table 10**
Comparison between the Nonsynonymous to Synonymous Polymorphism Ratio of Clients of Different Yeast Chaperones and Proteins That Are Not Clients of Any Chaperone

| Class | Chaperone | Clients | | | Non-clients | | | Mann–Whitney | | Partial Correlation | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | n | Median | Mean | n | Median | Mean | P value | Q value | $\rho$ | P value | Q value |
| CCTs | Cct2 | 119 | 0.2198 | 0.2508 | 2,152 | 0.2642 | 0.3537 | 0.0053** | 0.0103* | 0.0504 | 0.0538 | 0.0942 |
| CCTs | Cct3 | 120 | 0.2428 | 0.2488 | 2,152 | 0.2642 | 0.3537 | 0.0157* | 0.0239* | 0.0534 | 0.0406* | 0.0748 |
| CCTs | Cct4 | 162 | 0.2567 | 0.2925 | 2,152 | 0.2642 | 0.3537 | 0.5048 | 0.5354 | 0.1136 | $9.06 \times 10^{-6}$*** | $3.52 \times 10^{-5}$** |
| CCTs | Cct5 | 35 | 0.2447 | 0.2589 | 2,152 | 0.2642 | 0.3537 | 0.1807 | 0.2181 | 0.0296 | 0.2712 | 0.4157 |
| CCTs | Cct6 | 92 | 0.1751 | 0.2021 | 2,152 | 0.2642 | 0.3537 | $3.65 \times 10^{-6}$*** | $1.54 \times 10^{-5}$** | −0.0203 | 0.4413 | 0.6031 |
| CCTs | Cct7 | 41 | 0.2006 | 0.2065 | 2,152 | 0.2642 | 0.3537 | 0.0022** | 0.0048** | −0.0031 | 0.9072 | 0.9622 |
| CCTs | Cct8 | 177 | 0.2423 | 0.2604 | 2,152 | 0.2642 | 0.3537 | 0.0059** | 0.0109* | 0.0603 | 0.0185* | 0.0381* |
| CCTs | Tcp1 | 46 | 0.2231 | 0.2340 | 2,152 | 0.2642 | 0.3537 | 0.0118* | 0.0200* | 0.0140 | 0.6008 | 0.7510 |
| Hsp70s | Ecm10 | 64 | 0.1925 | 0.1891 | 2,152 | 0.2642 | 0.3537 | $1.92 \times 10^{-5}$** | $6.72 \times 10^{-5}$** | −0.0245 | 0.3574 | 0.5212 |
| Hsp70s | Kar2 | 67 | 0.2074 | 0.2436 | 2,152 | 0.2642 | 0.3537 | 0.0120* | 0.0200* | 0.0345 | 0.1938 | 0.3230 |
| Hsp70s | Lhs1 | 76 | 0.1968 | 0.2165 | 2,152 | 0.2642 | 0.3537 | 0.0004** | 0.0010** | 0.0099 | 0.7096 | 0.8564 |
| Hsp70s | Ssa1 | 2,380 | 0.2375 | 0.2714 | 2,152 | 0.2642 | 0.3537 | $1.79 \times 10^{-8}$*** | $2.09 \times 10^{-7}$*** | 0.0815 | $7.13 \times 10^{-7}$*** | $3.12 \times 10^{-6}$*** |
| Hsp70s | Ssa2 | 1,818 | 0.2397 | 0.2735 | 2,152 | 0.2642 | 0.3537 | $2.05 \times 10^{-6}$*** | $1.03 \times 10^{-5}$** | 0.0978 | $4.05 \times 10^{-8}$*** | $2.95 \times 10^{-7}$*** |
| Hsp70s | Ssa3 | 300 | 0.2429 | 0.2782 | 2,152 | 0.2642 | 0.3537 | 0.0361* | 0.0505 | 0.1010 | $3.91 \times 10^{-5}$** | 0.0001** |
| Hsp70s | Ssa4 | 433 | 0.2453 | 0.2771 | 2,152 | 0.2642 | 0.3537 | 0.0208* | 0.0303* | 0.1193 | $4.17 \times 10^{-7}$*** | $2.09 \times 10^{-6}$*** |
| Hsp70s | Ssb1 | 3,102 | 0.2355 | 0.2730 | 2,152 | 0.2642 | 0.3537 | $5.26 \times 10^{-9}$*** | $1.84 \times 10^{-7}$*** | 0.0780 | $2.38 \times 10^{-7}$*** | $1.39 \times 10^{-6}$*** |
| Hsp70s | Ssb2 | 1,160 | 0.2444 | 0.2734 | 2,152 | 0.2642 | 0.3537 | $9.19 \times 10^{-5}$** | $2.47 \times 10^{-4}$** | 0.1200 | $1.73 \times 10^{-9}$*** | $2.23 \times 10^{-8}$*** |
| Hsp70s | Ssc1 | 190 | 0.1807 | 0.2279 | 2,152 | 0.2642 | 0.3537 | $1.62 \times 10^{-6}$*** | $1.13 \times 10^{-6}$*** | 0.0162 | 0.5245 | 0.6799 |
| Hsp70s | Sse1 | 1,845 | 0.2330 | 0.2688 | 2,152 | 0.2642 | 0.3537 | $2.44 \times 10^{-8}$*** | $2.14 \times 10^{-7}$*** | 0.0732 | $3.86 \times 10^{-5}$** | 0.0001** |
| Hsp70s | Sse2 | 229 | 0.2547 | 0.2786 | 2,152 | 0.2642 | 0.3537 | 0.1279 | 0.1599 | 0.1057 | $2.51 \times 10^{-5}$** | $8.79 \times 10^{-5}$** |
| Hsp70s | Ssq1 | 93 | 0.1681 | 0.1936 | 2,152 | 0.2642 | 0.3537 | $2.38 \times 10^{-7}$*** | $1.39 \times 10^{-6}$*** | −0.0289 | 0.2732 | 0.4157 |
| Hsp70s | Ssz1 | 628 | 0.2533 | 0.2798 | 2,152 | 0.2642 | 0.3537 | 0.0393* | 0.0529 | 0.1362 | $1.05 \times 10^{-9}$*** | $2.23 \times 10^{-8}$*** |
| Hsp90s | Hsc82 | 419 | 0.2185 | 0.2547 | 2,152 | 0.2642 | 0.3537 | $3.95 \times 10^{-6}$*** | $1.54 \times 10^{-5}$** | 0.0672 | 0.0047** | 0.0103* |
| Hsp90s | Hsp82 | 828 | 0.2525 | 0.2831 | 2,152 | 0.2642 | 0.3537 | 0.0150* | 0.0239* | 0.1282 | $1.91 \times 10^{-9}$*** | $2.23 \times 10^{-8}$*** |
| Hsp100s | Hsp78 | 759 | 0.2255 | 0.2556 | 2,152 | 0.2642 | 0.3537 | $1.52 \times 10^{-8}$*** | $2.09 \times 10^{-7}$*** | 0.0805 | 0.0002** | $5.00 \times 10^{-4}$** |
| Hsp100s | Hsp104 | 358 | 0.2381 | 0.2647 | 2,152 | 0.2642 | 0.3537 | 0.0021* | 0.0048* | 0.0908 | 0.0002** | $5.00 \times 10^{-4}$** |
| Small | Hsp31 | 99 | 0.2311 | 0.2862 | 2,152 | 0.2642 | 0.3537 | 0.3822 | 0.4180 | 0.0759 | 0.0038* | 0.0089* |
| Small | Hsp32 | 2 | 0.2390 | 0.2390 | 2,152 | 0.2642 | 0.3537 | 0.7342 | 0.7558 | 0.0084 | 0.7591 | 0.8856 |
| Small | Hsp33 | 3 | 0.1244 | 0.1769 | 2,152 | 0.2642 | 0.3537 | 0.3003 | 0.3390 | −0.0037 | 0.8916 | 0.9622 |
| Small | Sno4 | 3 | **0.3315** | 0.2547 | 2,152 | 0.2642 | 0.3537 | 0.7703 | 0.7703 | 0.0207 | 0.4480 | 0.6031 |
| Other | Hsp12 | 88 | 0.2476 | 0.2645 | 2,152 | 0.2642 | 0.3537 | 0.0886 | 0.1149 | 0.0599 | 0.0232* | 0.0451* |
| Other | Hsp26 | 84 | 0.1676 | 0.2215 | 2,152 | 0.2642 | 0.3537 | $6.75 \times 10^{-5}$** | $1.97 \times 10^{-4}$** | −0.0013 | 0.9594 | 0.9865 |
| Other | Hsp42 | 364 | 0.2520 | 0.2854 | 2,152 | 0.2642 | 0.3537 | 0.2499 | 0.2916 | 0.1316 | $4.22 \times 10^{-8}$*** | $2.95 \times 10^{-7}$*** |
| Other | Hsp60 | 95 | 0.1943 | 0.2150 | 2,152 | 0.2642 | 0.3537 | $6.04 \times 10^{-5}$** | $1.92 \times 10^{-4}$** | −0.0032 | 0.9043 | 0.9622 |
| Other | Mcx1 | 42 | 0.1655 | 0.2165 | 2,152 | 0.2642 | 0.3537 | 0.0034* | 0.0070* | 0.0005 | 0.9865 | 0.9865 |

Note.—For each pair of clients versus non-clients, the highest $d_N/d_S$ values are shown in bold face. Partial correlations correspond to the Spearman correlation between chaperone dependence and $d_N/d_S$, controlling for number of protein–protein interactions. Q values were computed using the Benjamini–Hochberg approach (Benjamini and Hochberg 1995).
*$P < 0.05$, **$P < 0.001$, and ***$P < 10^{-5}$.

**Table 11**

Comparison between the Nonsynonymous to Synonymous Polymorphism Ratio of Clients of Different Chaperone Families and Proteins That Are Not Clients of Any Chaperone

| Class | Clients | | | Non-clients | | | Mann–Whitney | | Partial Correlation | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | Median | Mean | $n$ | Median | Mean | $P$ value | $Q$ value | $\rho$ | $P$ value | $Q$ value |
| CCTs | 487 | 0.2436 | 0.2694 | 2,152 | **0.2642** | **0.3537** | 0.0004** | 0.0001** | 0.0942 | $5.36 \times 10^{-5}$** | 0.0020* |
| Hsp70s | 836 | 0.2416 | 0.2907 | 2,152 | **0.2642** | **0.3537** | 0.0009** | 0.0201* | 0.0504 | 0.0201* | 0.0352* |
| Hsp90s | 947 | 0.2433 | 0.2731 | 2,152 | **0.2642** | **0.3537** | 0.0001** | $4.60 \times 10^{-6}$*** | 0.1024 | $9.20 \times 10^{-7}$*** | 0.0003** |
| Hsp100s | 863 | 0.2272 | 0.2592 | 2,152 | **0.2642** | **0.3537** | $1.82 \times 10^{-8}$*** | 0.0007** | 0.0752 | 0.0004** | 0.0120* |
| Small | 103 | 0.2479 | 0.2875 | 2,152 | **0.2642** | **0.3537** | 0.4503 | 0.0020* | 0.0825 | 0.0016* | 0.0066* |

Note.—For each pair of clients versus non-clients, the highest $d_N/d_S$ values are shown in bold face. Partial correlations correspond to the Spearman correlation between chaperone dependence and $d_N/d_S$, controlling for number of protein–protein interactions. $Q$ values were computed using the Benjamini–Hochberg approach (Benjamini and Hochberg 1995).

*$P < 0.05$, **$P < 0.001$, and ***$P < 10^{-5}$.
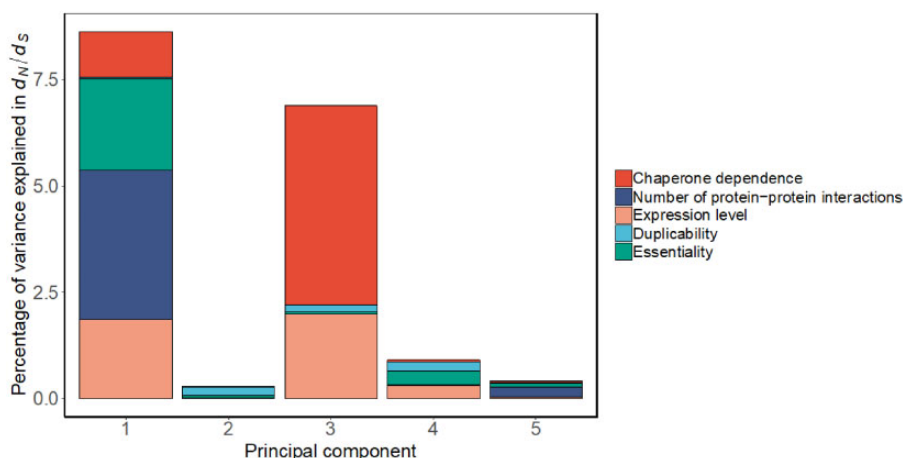


FIG. 6.—Principal component regression on $d_N/d_S$ calculated using genetic variants segregating in *Saccharomyces cerevisiae* for 6,132 yeast genes. For each principal component, the height of the bar represents the percent of variance in the rate of evolution explained by the component. The relative contribution of each variable to a principal component is represented with different colors. Table 13 contains the numerical data used to draw this figure.

For each chaperone, we compared the rates of evolution of their clients ($n$ ranged from 2 to 3,102) against the rates of evolution of non-clients (proteins that are not clients of any chaperone, $n = 2,152$). In all 35 cases, clients exhibited a lower average $d_N/d_S$, and in 34 of the cases they also exhibited a lower median $d_N/d_S$, with significant differences in 26 cases (Mann–Whitney $U$ test, $P < 0.05$; table 10). Partial correlations between $d_N/d_S$ and chaperone dependence controlling for degree were positive in 28 cases (significantly positive in 19 cases) and negative in seven cases (significantly negative in 0 cases).

Next, for each group of chaperones (small Hsps, Hsp70s, Hsp90s, Hsp100s, and CCTs), we compared the rates of nonsynonymous to synonymous polymorphism of the clients of any of the group ($n$ ranged from 103 to 947) against those of proteins that are not clients of any chaperone ($n = 2,152$). In all five cases, clients exhibited lower median and mean $d_N/d_S$, with significant differences (Mann–Whitney $U$ test, $P < 0.05$) in all cases except for the clients of small Hsps (the smallest group; table 11). However, partial correlations between $d_N/d_S$

and chaperone dependence controlling for network degree was always significantly positive.

Finally, we performed a multivariable analysis to study the effect of chaperone dependence on $d_N/d_S$ at the intrapopulation level controlling simultaneously for all the studied variables, as we did previously for the divergence data. The results are very similar. We first regressed $d_N/d_S$ against the five biological factors, and found that all make a significant contribution to the regression and that the overall $R^2$ is 0.17 (table 12). Chaperone dependence was the only factor with a positive coefficient, indicating that chaperone dependence also increases $d_N/d_S$ within yeast populations. We also performed a principal component regression analysis using the same predictor variables as above. Table 13 shows numerical data from the analysis, while figure 6 shows these data graphically.

As with divergence data, we found a principal component with a 70% contribution of chaperone dependence and 30% expression level. This component explained ~7% of the variance of $d_N/d_S$ (table 13 and fig. 6). Another significant

**Table 12**

Multiple Linear Regression of Polymorphism Data

| | $d_N/d_S$ |
|---|---|
| Chaperone dependence | 0.16*** |
| Number of protein–protein interactions | −0.06*** |
| Expression level | −0.23*** |
| Duplicability | −0.06* |
| Essentiality | −0.20*** |

*$P < 0.05$ and ***$P < 10^{-5}$.

**Table 14**

Total Variance Explained by Each Variable in the Principal Component Regression Analysis of Polymorphism Data

| | $d_N/d_S$ |
|---|---|
| Chaperone dependence | 5.87% |
| Number of protein–protein interactions | 3.76% |
| Expression level | 4.23% |
| Duplicability | 0.62% |
| Essentiality | 2.66% |

**Table 13**

Results from the Principal Component Regression Analysis of Polymorphism Data

| | Principal Components | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | All |
| Percentage of explained variance in | | | | | | |
| $d_N/d_S$ | 8.63 | 0.30 | 6.88 | 0.91 | 0.41 | 17.13 |
| Percent contributions of each variable | | | | | | |
| Chaperone dependence | 0.13 | 0.06 | **0.68** | 0.05 | 0.08 | |
| Number of protein–protein interactions | **0.41** | 0.02 | 0.00 | 0.02 | **0.56** | |
| Expression level | **0.22** | 0.04 | **0.29** | **0.34** | 0.12 | |
| Duplicability | 0.00 | **0.71** | 0.02 | **0.24** | 0.03 | |
| Essentiality | **0.25** | 0.18 | 0.01 | **0.36** | **0.22** | |

NOTE.—We indicate in bold the contributions of a predictor to a component when >20%.

principal component explains 8.6% of the variance. This component is mainly determined by the number of protein–protein interactions, essentiality, and expression level. The other three significant components explained in combination <2% of the variance. In summary, we also found that chaperone dependence was the biological factor explaining the largest fraction of the total variance in the rate of evolution measured as $d_N/d_S$ (5.87%), explaining a larger fraction of the total variance than expression level (4.23%) and the number of protein–protein interactions (3.76%) (table 14).

Finally, we performed an ANCOVA to evaluate the effect of chaperone dependence on the rate of protein evolution while controlling for the effect of the number of protein–protein interactions, expression level, and essentiality. As the continuous variable in the ANCOVA, we used the principal component of these three variables (principal component 1 in table 13 and fig. 6). We found that chaperone clients evolve on average 19.2% faster than the proteome average ($P = 3.6 \times 10^{-11}$) (fig. 7).

## Discussion

We study how the different yeast chaperones affect the evolutionary rate of their protein clients. In particular, we analyze
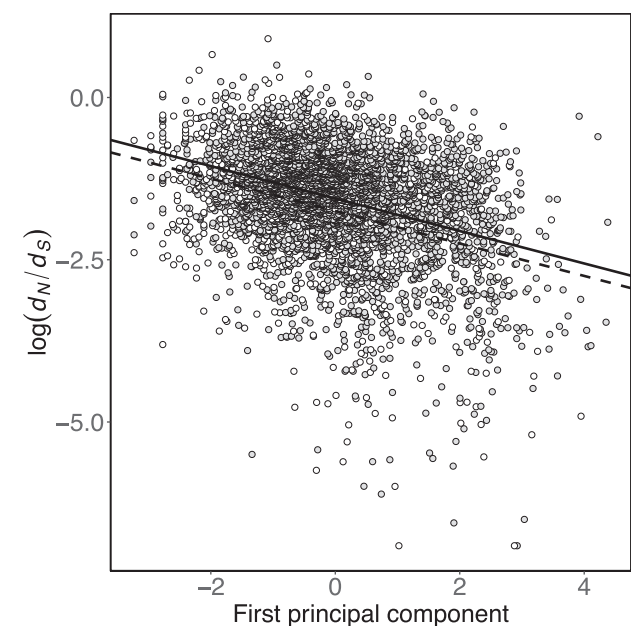


FIG. 7.—ANCOVA. Chaperone clients (gray points, continuous line) evolve 19.2% above the genome average rate (light points, dashed line) when considering genetic variants segregating in *Saccharomyces cerevisiae*.

the effect of chaperone dependence on protein evolution at two very different evolutionary time scales. We first study how chaperone-mediated folding has affected protein evolution over the evolutionary divergence of *S. cerevisiae* and *S. paradoxus*. We then study if the same process has left a signature on the patterns of standing genetic variation found in modern wild and domesticated strains of *S. cerevisiae* (Peter et al. 2018). We find that chaperone-mediated buffering has indeed left a trace on the protein-coding regions of the yeast genome, such that genes encoding chaperone clients ("client genes") have diverged faster than genes encoding non-client proteins ("non-client genes") when controlling for their number of protein–protein interactions. We also find that client genes have accumulated more genetic diversity than non-client genes among natural strains of *S. cerevisiae*. In a principal component regression analysis, we find that chaperone dependence explains the largest fraction of the observed

variance in the rate of evolution at both evolutionary time scales. This contribution of chaperone-mediated folding to the variations on the rate of protein evolution is well above the fraction of the variance explained by other well-known factors that affect protein evolution such as expression level or protein–protein interactions (Pál et al. 2001; Fraser et al. 2002; Drummond et al. 2005).

Cost-benefit trade-offs are common in evolution, including protein evolution. Proteins are marginally stable (DePristo et al. 2005) and soluble (Tartaglia et al. 2007) inside a cell and their native structure is sensitive to mutations. Protein stability is a major constraint on protein evolution (Bloom et al. 2006; Zeldovich et al. 2007). Most nonsynonymous mutations diminish protein stability or solubility, and are therefore deleterious (Dobson 1999). Moreover, neofunction-alizing mutations that confer new protein functions, including new protein–protein interactions, tend to be highly destabiliz-ing (Tokuriki et al. 2008; Soskine and Tawfik 2010). Therefore, in the absence of chaperone buffering, the cost of a neofunctionalizing mutation may be larger than its ben-efit (Tokuriki and Tawfik 2009). Chaperones, by diminishing the negative effect of mutations on protein stability and fold-ing, can promote protein evolution, and potentiate the regu-latory or metabolic effect of a protein mutation (Taipale et al. 2010). Our finding that yeast chaperones can accelerate pro-tein evolution is in line with previous observations that chap-erones can act as evolutionary capacitors (Queitsch et al. 2002; Rutherford 2003; Jarosz and Lindquist 2010), buffer the destabilizing effect of mutations (Tokuriki and Tawfik 2009), facilitate the divergence of gene duplicates (Lachowiec et al. 2013), and ultimately allow proteins to ex-plore a larger fraction of their sequence space (Williams and Fares 2010; Pechmann and Frydman 2014; Aguilar-Rodríguez et al. 2016; Kadibalban et al. 2016). However, it is important to notice that chaperones do not just modify the effects of mutations affecting protein stability or folding. A chaperone can also modify (either buffer or potentiate) the fitness or phenotypic effects of mutations in proteins that do not have a direct functional relationship with it. For example, many of the protein clients of Hsp90 are transcription factors and signaling proteins (Taipale et al. 2010; Zabinsky et al. 2018). Therefore, the modifying effect of Hsp90 can perco-late throughout the molecular networks of the cell affecting mutations in many genes that do not have a direct physical or functional relationship with Hsp90. Even if chaperone-buffered genetic variants are only rarely acquired, they could be enriched in a population if stabilizing selection does not remove them (because their deleterious phenotypic conse-quences are masked by a chaperone). A recent study has found evidence for this hypothesis among Hsp90-dependent variants in *S. cerevisiae* that affect cell size and shape (Geiler-Samerotte et al. 2016). We also find evidence for this enrichment of cryptic genetic variants within client genes among genetic variants segregating in *S. cerevisiae*.

Nonsynonymous mutations that allow the establishment of new physical interactions with other proteins are a class of neofunctionalizing mutations that can be highly destabilizing (Pechmann and Frydman 2014). Therefore, some chaperones can also buffer mutations that rewire protein interactions, thus promoting the evolution of protein networks (Pechmann and Frydman 2014), and perhaps explaining why chaperone clients tend to be well-connected in such net-works, as we observed here. Furthermore, chaperones and their clients coevolve in a process where sequence changes in the chaperone may lead to compensatory changes in their clients and further rewiring of the protein networks they form (Koubkova-Yu et al. 2018).

In a multivariable statistical analysis, we find that the chaperones affecting rates of protein evolution belong to two major chaperone families: Hsp70s and Hsp90s. While there is ample evidence that Hsp90s can accelerate the rate of protein evolution in other eukaryotic species (Lachowiec et al. 2013, 2015; Pechmann and Frydman 2014), the evidence for eukaryotic Hsp70 chaperones having a similar effect is not so abundant. A previous study found that the ribosome-associated Hsp70 SSB chaperone that preferentially binds long and disordered nascent polypeptide chains accelerates the rate of accu-mulation of mutations likely to be destabilizing among weakly interacting clients (Pechmann and Frydman 2014). In a previous study, we found that bacterial DnaK, which belongs to the same major chaperone fam-ily, also accelerates protein evolution using a combination of experimental and comparative genomics approaches (Aguilar-Rodríguez et al. 2016). While it has been shown before that the chaperonin GroEL accelerates protein evo-lution (Bogumil and Dagan 2010; Williams and Fares 2010), we do not find good evidence here that the eu-karyotic chaperonin system CCT, present in eukarya and archaea but absent from bacteria, has the same effect on protein evolution. We find that the chaperone Hsp104 from the Hsp100 family accelerates the evolution of its protein clients when controlling for number of protein–protein interactions. This could be the first observation that this important chaperone could affect protein evolutionary rates. However, we do not observe any effect of the family Hsp100 (Hsp78 and Hsp104) when control-ling for possible confounding variables in a multiple linear regression and in a principal component regression anal-ysis. Finally, we do not detect any significant effect of small heat shock proteins in the rate of evolution of their clients.

In summary, we analyzed the evolution of proteins that are subjected to folding assisted by different chaperones in the complex yeast chaperone network over two different evolu-tionary time scales. Our comparative approach indicates that chaperone-assisted folding increases the rate of protein evo-lution when properly controlling for confounding factors at

both time scales. We show how protein chaperones, by virtue of their role in modulating protein genotype–phenotype maps, have a disproportionate effect on the evolution of the protein-coding regions of a genome. Our results highlight the importance of integrating different cellular factors when studying protein sequence evolution.

## Acknowledgments

## Literature Cited

Aguilar-Rodríguez J, et al. 2016. The molecular chaperone DnaK is a source of mutational robustness. Genome Biol Evol. 8(9):2979–2991.

Aguilar-Rodríguez J, Wagner A. 2018. Metabolic determinants of enzyme evolution in a genome-scale bacterial metabolic network. Genome Biol Evol. 10:3076–3088.

Alvarez-Ponce D. 2014. Why proteins evolve at different rates: the determinants of proteins' rates of evolution. In: Fares MA, editor. Natural selection: methods and applications. London: CRC Press (Taylor & Francis). p. 126–178.

Alvarez-Ponce D, Fares MA. 2012. Evolutionary rate and duplicability in the *Arabidopsis thaliana* protein-protein interaction network. Genome Biol Evol. 4(12):1263–1274.

Alvarez-Ponce D, Feyertag F, Chakraborty S. 2017. Position matters: network centrality considerably impacts rates of protein evolution in the human protein–protein interaction network. Genome Biol Evol. 9(6):1742–1756.

Alvarez-Ponce D, Sabater-Muñoz B, Toft C, Ruiz-González MX, Fares MA. 2016. Essentiality is a strong determinant of protein rates of evolution during mutation accumulation experiments in *Escherichia coli*. Genome Biol Evol. 8(9):2914–2927.

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

Bloom JD, Labthavikul ST, Otey CR, Arnold FH. 2006. Protein stability promotes evolvability. Proc Natl Acad Sci U S A. 103(15):5869–5874.

Bogumil D, Dagan T. 2010. Chaperonin-dependent accelerated substitution rates in prokaryotes. Genome Biol Evol. 2: 602–608.

Bogumil D, Landan G, Ilhan J, Dagan T. 2012. Chaperones divide yeast proteins into classes of expression level and evolutionary rate. Genome Biol Evol. 4(5):618–625.

Cherry JM, et al. 2012. Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. 40(Database issue):D700–705.

DePristo MA, Weinreich DM, Hartl DL. 2005. Missense meanderings in sequence space: a biophysical view of protein evolution. Nat Rev Genet. 6(9):678–687.

Do CB, Mahabhashyam MS, Brudno M, Batzoglou S. 2005. ProbCons: probabilistic consistency-based multiple sequence alignment. Genome Res. 15(2):330–340.

Dobson CM. 1999. Protein misfolding, evolution and disease. Trends Biochem Sci. 24(9):329–332.

Dori-Bachash M, Shema E, Tirosh I. 2011. Coupled evolution of transcription and mRNA degradation. PLoS Biol. 9(7):e1001106.

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly expressed proteins evolve slowly. Proc Natl Acad Sci U S A. 102(40):14338–14343.

Drummond DA, Raval A, Wilke CO. 2006. A single determinant dominates the rate of yeast protein evolution. Mol Biol Evol. 23(2):327–337.

Duret L, Mouchiroud D. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. Mol Biol Evol. 17(1):68–74.

Ellis J. 1987. Proteins as molecular chaperones. Nature 328(6129):378–379.

Fares MA. 2015. The origins of mutational robustness. Trends Genet. 31(7):373–381.

Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary rate in the protein interaction network. Science 296(5568):750–752.

Geiler-Samerotte KA, Zhu YO, Goulet BE, Hall DW, Siegal ML. 2016. Selection transforms the landscape of genetic variation interacting with Hsp90. PLoS Biol. 14(10):e2000465.

Giaever G, et al. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. Nature 418(6896):387–391.

Gong Y, et al. 2009. An atlas of chaperone–protein interactions in *Saccharomyces cerevisiae*: implications to protein folding pathways in the cell. Mol Syst Biol. 5: 275.

Hahn MW, Kern AD. 2005. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol. 22(4):803–806.

Han MV, Demuth JP, McGrath CL, Casola C, Hahn MW. 2009. Adaptive evolution of young gene duplicates in mammals. Genome Res. 19(5):859–867.

Hartl FU, Bracher A, Hayer-Hartl M. 2011. Molecular chaperones in protein folding and proteostasis. Nature 475(7356):324.

Hartl FU, Hayer-Hartl M. 2009. Converging concepts of protein folding in vitro and in vivo. Nat Struct Mol Biol. 16(6):574.

Hurst LD, Smith NG. 1999. Do essential genes evolve slowly? Curr Biol. 9(14):747–750.

Jarosz DF, Lindquist S. 2010. Hsp90 and environmental stress transform the adaptive value of natural genetic variation. Science 330(6012):1820–1824.

Jarosz DF, Taipale M, Lindquist S. 2010. Protein homeostasis and the phenotypic manifestation of genetic diversity: principles and mechanisms. Annu Rev Genet. 44:189–216.

Jordan IK, Rogozin IB, Wolf YI, Koonin EV. 2002. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. Genome Res. 12(6):962–968.

Kadibalban AS, Bogumil D, Landan G, Dagan T. 2016. DnaK-dependent accelerated evolutionary rate in prokaryotes. Genome Biol Evol. 8(5):1590–1599.

Kim S. 2015. ppcor: an R package for a fast calculation to semi-partial correlation coefficients. Commun Stat Appl Methods. 22:665–674.

Kinsella RJ, et al. 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database (Oxford) 2011:bar030.

Koubkova-Yu T-T, Chao J-C, Leu J-Y. 2018. Heterologous Hsp90 promotes phenotypic diversity through network evolution. PLoS Biol. 16(11):e2006450.

Lachowiec J, et al. 2013. The protein chaperone HSP90 can facilitate the divergence of gene duplicates. Genetics 193(4):1269–1277.

Lachowiec J, Lemus T, Borenstein E, Queitsch C. 2015. Hsp90 promotes kinase evolution. Mol Biol Evol. 32(1):91–99.

Lauring AS, Frydman J, Andino R. 2013. The role of mutational robustness in RNA virus evolution. Nat Rev Microbiol. 11(5):327.

Li WH, Wu CI, Luo CC. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol. 2(2):150–174.

Luisi P, et al. 2015. Recent positive selection has acted on genes encoding proteins with more interactions within the whole human interactome. Genome Biol Evol. 7(4):1141–1154.

Nagalakshmi U, et al. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320(5881):1344–1349.

Nembaware V, Crum K, Kelso J, Seoighe C. 2002. Impact of the presence of paralogs on sequence divergence in a set of mouse-human orthologs. Genome Res. 12(9):1370–1376.

Pál C, Papp B, Hurst LD. 2001. Highly expressed genes in yeast evolve slowly. Genetics 158(2):927–931.

Payne JL, Wagner A. 2019. The causes of evolvability and their evolution. Nat Rev Genet. 20(1):24–38.

Pechmann S, Frydman J. 2014. Interplay between chaperones and protein disorder promotes the evolution of protein networks. PLoS Comput Biol. 10(6):e1003674.

Pegueroles C, Laurie S, Alba MM. 2013. Accelerated evolution after gene duplication: a time-dependent process affecting just one copy. Mol Biol Evol. 30(8):1830–1842.

Peter J, et al. 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556(7701):339.

Plotkin JB, Fraser HB. 2007. Assessing the determinants of evolutionary rates in the presence of noise. Mol Biol Evol. 24(5):1113–1121.

Queitsch C, Sangster TA, Lindquist S. 2002. Hsp90 as a capacitor of phenotypic variation. Nature 417(6889):618–624.

R Core Team. 2014. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: http://www.R-project.org/.

Rocha EP. 2006. The quest for the universals of protein evolution. Trends Genet. 22(8):412–416.

Rutherford SL. 2003. Between genotype and phenotype: protein chaperones and evolvability. Nat Rev Genet. 4(4):263–274.

Soskine M, Tawfik DS. 2010. Mutational effects and the evolution of new protein functions. Nat Rev Genet. 11(8):572.

Taipale M, Jarosz DF, Lindquist S. 2010. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. Nat Rev Mol Cell Biol. 11(7):515.

Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M. 2007. Life on the edge: a link between gene expression levels and aggregation rates of human proteins. Trends Biochem Sci. 32(5):204–206.

Tokuriki N, Stricher F, Serrano L, Tawfik DS. 2008. How protein stability and new functions trade off. PLoS Comput Biol. 4(2):e1000002.

Tokuriki N, Tawfik DS. 2009. Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature 459(7247):668–673.

Williams TA, Fares MA. 2010. The effect of chaperonin buffering on protein evolution. Genome Biol Evol. 2: 609–619.

Wright SI, Yau CB, Looseley M, Meyers BC. 2004. Effects of gene expression on molecular evolution in Arabidopsis thaliana and Arabidopsis lyrata. Mol Biol Evol. 21(9):1719–1726.

Yang J, Gu Z, Li WH. 2003. Rate of protein evolution versus fitness effect of gene deletion. Mol Biol Evol. 20(5):772–774.

Yang Z. 2000. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. J Mol Evol. 51(5):423–432.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.

Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155(1):431–449.

Zabinsky RA, Mason GA, Queitsch C, Jarosz DF. 2018. It's not magic – Hsp90 and its effects on genetic and epigenetic variation. Semin Cell Dev Biol. 88:21–35.

Zeldovich KB, Chen P, Shakhnovich EI. 2007. Protein stability imposes limits on organism complexity and speed of molecular evolution. Proc Natl Acad Sci U S A. 104(41):16152–16157.

Zhang J, Yang JR. 2015. Determinants of the rate of protein sequence evolution. Nat Rev Genet. 16(7):409–420.

Zhang L, Li WH. 2004. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. Mol Biol Evol. 21(2):236–239.

Zuckerkandl E. 1976. Evolutionary processes and evolutionary noise at the molecular level. I. Functional density in proteins. J Mol Evol. 7(3):167–183.

Zuckerkandl E, Pauling L. 1965. Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel H, editors. Evolving genes and proteins New York: Academic Press. p. 97–166.

Associate editor: Balazs Papp