Experimental illumination of a fitness landscape

Ryan T. Hietpas[a], Jeffrey D. Jensen[b], and Daniel N. A. Bolon[a,1]

[a]Department of Biochemistry and Molecular Pharmacology and [b]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01605

Edited* by Robert T. Sauer, Massachusetts Institute of Technology, Cambridge, MA, and approved March 15, 2011 (received for review October 28, 2010)

The genes of all organisms have been shaped by selective pressures. The relationship between gene sequence and fitness has tremendous implications for understanding both evolutionary processes and functional constraints on the encoded proteins. Here, we have exploited deep sequencing technology to experimentally determine the fitness of all possible individual point mutants under controlled conditions for a nine-amino acid region of Hsp90. Over the past five decades, limited glimpses into the relationship between gene sequence and function have sparked a long debate regarding the distribution, relative proportion, and evolutionary significance of deleterious, neutral, and advantageous mutations. Our systematic experimental measurement of fitness effects of Hsp90 mutants in yeast, evaluated in the light of existing population genetic theory, are remarkably consistent with a nearly neutral model of molecular evolution.

The results of >150 y of biological research has demonstrated that selection pressures shape the evolution of organisms (1). The relationship between gene sequence and selective advantage/disadvantage provides the fundamental link between genotype and fitness. Until now, it had not been feasible to systematically measure this relationship because of the challenge of constructing and monitoring all possible genetic variants. Two classes of experiments have provided glimpses of the fitness landscape and inferences into the relationship between gene sequence and fitness: directed evolution (2–4) and microbial experimental evolution (5–7). In both of these approaches, the fitness landscape can only be inferred—either because the pool of starting mutations is unknown, or because mutational sampling is limited. Thus, the question remains: What does the fitness landscape look like for all possible point mutants?

Determining the fitness landscape of point mutations in a gene is conceptually simple: Measure the fitness of organisms with each possible point mutation in a specific gene in an otherwise identical genetic background. In practice, there are two technical challenges: generating high-quality systematic mutant libraries and measuring fitness in high throughput both accurately and with a large dynamic range. To address these challenges, we developed an approach that we call "extremely methodical and parallel investigation of randomized individual codons" (EMPIRIC) fitness (Fig. 1).

Results

We used the EMPIRIC approach to analyze yeast Hsp90, an essential chaperone in eukaryotes (8) required for the maturation of many kinases (9). Of note, analyzing an essential gene maximizes the potential fitness range of mutants and the signal of the analysis. The amino acid sequence of Hsp90 is highly conserved among eukaryotes with 45% of the amino acids identical between the human and *Saccharomyces cerevisiae* proteins. Based on the sequence and structure (10) of Hsp90, we focused on a nine-amino acid region that contains a diversity of different amino acids with positions that vary in both their level of phylogenetic conservation among diversely related eukaryotes and their physical environment (solvent exposed and buried) in the structure of Hsp90 (Fig. 2*A*). In addition, two solvent-exposed aromatic side chains (F583 and W585) were structurally intriguing for a chaperone based on their potential to bind to hydrophobic regions on binding partners. The randomization of this nine-amino acid region resulted in the parallel analyses of 180 amino acid substitutions and >500 different codon variants—a task that would be daunting by traditional approaches.

Our analysis method monitors plasmid abundance that we expect to parallel with cell growth such that selective pressure begins to impact plasmid abundance at about the same time that it impacts cell growth. When cells with null rescue plasmids were switched to nonpermissive conditions, growth began to retard noticeably after 8 h and was stably slowed after 12 h (Fig. 2*B*). At this time we also observed the effects of selective pressure on the relative abundance of our point mutant plasmid library as monitored by deep sequencing (Fig. 2 *C* and *D*). Starting at 12 h, the relative abundance of wild-type sequence reads starts to increase consistent with the wild-type sequence having better fitness relative to the average point mutant (Fig. 2*C*). Because we generate our libraries with mixtures of all four nucleotides at each codon position, stop codons are included in our library and provide an internal monitor of selection pressure. Stop codons at all of the positions that we analyzed rapidly decrease in relative abundance starting at 12 h in selective conditions (Fig. 2*D*), consistent with the known requirement of sequences C-terminal to this region for Hsp90 function (11). From these results, we conclude that our deep sequencing approach is an effective means to monitor selective pressure.

The inclusion of wild-type sequences in our libraries serves as an internal benchmark to calculate the competitive fitness of each mutant. Under our experimental conditions, the doubling time of a homogeneous culture of yeast harboring wild-type Hsp90 plasmid was 4 h (Fig. 2*B*). By measuring the change in the ratio of a mutant to wild-type sequence reads as a function of this wild-type generation time (Fig. 2*E*), we calculate the relative fitness of the mutant as a selection coefficients (s) (12). Because fitness is related to the change in abundances as a function of time, it does not require equal abundance of each variant at the beginning of the experiment. Thus, biases in the mutational process (i.e., from oligonucleotide synthesis) did not preclude the analysis of fitness of any mutants. The selection coefficients represent the difference in fitness between the mutant and wild type. For yeast, fitness is proportional to the inverse of the doubling time and, by definition, wild-type fitness is 1. Thus, a selection coefficient of zero (no change in mutant to wild-type ratio over time) means that a mutant is as fit as wild-type, a negative selection coefficient means that a mutant is less fit than wild-type (−1 if a mutant does not support any proliferation), and a positive selection coefficient means that a mutant is more fit than wild-type. We calculated selection coefficients from our EMPIRIC fitness measurements for each codon mutant in our library (Datasets S1 and S2). For six mutants, we compared the EMPIRIC measured fitness effects to those measured by traditional two strain competition using strains with different colored fluorescent proteins (Fig. S1). EMPIRIC

Author contributions: R.T.H. and D.N.A.B. designed research; R.T.H. performed research; R.T.H., J.D.J., and D.N.A.B. analyzed data; and R.T.H., J.D.J., and D.N.A.B. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: dan.bolon@umassmed.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1016024108/-/DCSupplemental.

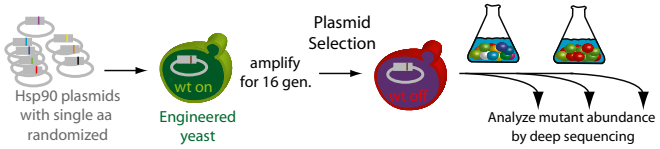

Fig. 1. EMPIRIC approach to experimentally determine fitness landscapes. Randomized individual codon libraries are introduced into a host cell whose only other copy of the gene is regulatable. The fitness of each individual codon mutation is determined by measuring its abundance in the mixed culture as a function of time under selective conditions.

and the bistrain competitions both parse WT-like and null-like mutants similarly, and for strains that persist in the cultures and that are therefore monitored with higher signal, we observe a strong positive correlation ($R^2 = 0.92$, $P = 0.003$ for two-tailed Student's t test) between EMPIRIC and biculture fitness measurements. Of note, one advantage of EMPIRIC measurements is that all mutants experience identical environmental conditions because they are physically located in the same flask compared with bistrain fitness competitions where each mutant is grown in separate flasks. To examine the reproducibility of EMPIRIC measurements, we repeated the EMPIRIC experiment (Fig. S1) and observe a strong correlation ($R^2 = 0.82$, $P = 10^{-67}$ two-tailed Student's t test). Thus, the EMPIRIC procedure is very reproducible.

The average selection coefficient for all codons from our EMPIRIC analyses was $s = -0.42$ ($P = 10^{-170}$ using two-tailed Student's t test compared with null hypothesis that the mean s = 0) indicating that the average mutation in this region is of deleterious effect. For comparison, the average selection coefficient of all stop codons was −0.75. The fitness of stop codons is greater than a true null ($s = -1$), indicating that the relative plasmid abundance of these nonsense mutations falls off rapidly with time in selective conditions, but that slow plasmid replication may persist in some fraction of these cells.

We examined selection coefficient differences between synonymous codons, which code for identical protein sequences. Synonymous codon substitutions among homologous genes are widely used in population genetic analysis as a measure of the neutral mutation rate (13) with the underlying assumption that these substitutions do not impact fitness and their dynamics are governed by genetic drift. This assumption is imperfect because species have distinct codon preferences within coding regions, indicating that selective pressure may in fact distinguish between synonymous codons (14). Our experimental data enabled us to analyze the variation in fitness between synonymous substitutions. We observe increased fitness variability between synonymous substitutions where the average synonym fitness was null-like (Fig. S2), which is likely caused by sampling noise (due to the low abundance of these codons in the competing culture). To minimize this noise, we calculated the variability among selection coefficients for synonymous codons with high fitness ($s > -0.05$, $n = 151$) as a root mean square deviation (rmsd = 0.018 compared with the synonym mean). For comparison we observe an rmsd of 0.35 when all possible substitutions including those that result in an amino acid change are considered. Thus, synonymous substitutions caused fitness changes that pale in magnitude to amino acid changing substitutions consistent with the expectation of neutrality at synonymous sites (15).

We averaged the observed EMPIRIC selection coefficient of synonymous codons to generate fitness profiles of each amino acid at each position (Fig. 3*A* and Fig. S3). The fitness of the hydrophobic amino acids exhibited less variability within a position than polar amino acids (rmsd of selection coefficients for VILMFYW of 0.15 compared with 0.26 for KRHDENQST). From a fitness perspective, the specific geometry of hydrophobic amino acids had a smaller impact compared with the varied physical properties of polar amino acids. To compare geometrical sensitivity among amino acids with similar physical properties, we compared selection coefficients at each position between the following pairs: D/E, K/R, N/Q, V/I, L/M, W/Y. The three polar pairs differ more than the three hydrophobic pairs (selection coefficient rmsd of 0.23 and 0.10, respectively), indicating that the fitness of polar amino acids is more sensitive to geometry than the fitness of hydrophobic amino acids. The relative insensitivity of amino acid substitutions between hydropho-

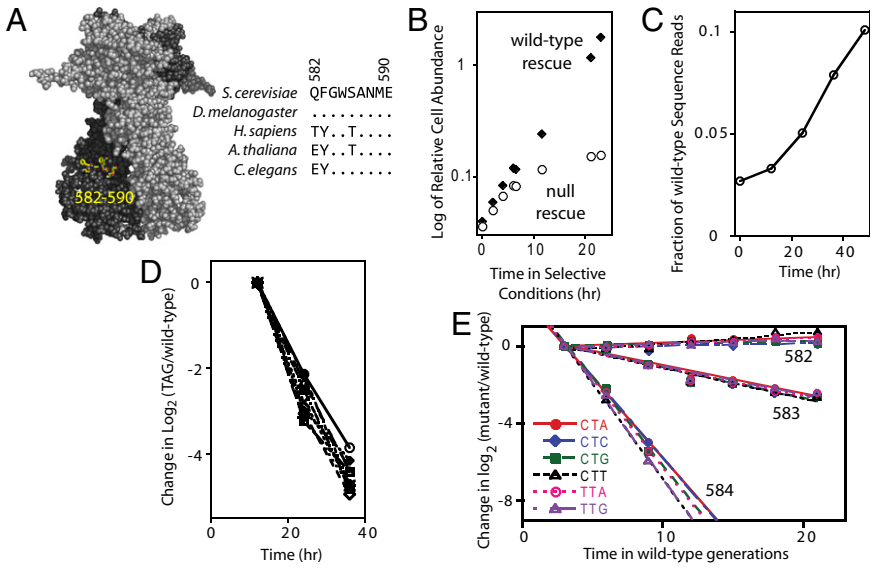

Fig. 2. Hsp90 region analyzed and application of selection pressure to point mutants of Hsp90 in yeast. (*A*) Positions 592–600 are highlighted in yellow in the dimeric structure of *S. cerevisiae* Hsp90. (*B*) Growth of an Hsp90 temperature-sensitive yeast strain at 36 °C is rescued with a wild-type Hsp90 plasmid. (*C* and *D*) Deep sequencing analysis of a library of single-codon mutants of Hsp90 from amino acids 582–590 grown in mixed culture. (*C*) Relative abundance of wild-type sequence as a function of time in selective conditions where the only other copy of Hsp90 is inactivated. (*D*) The ratio of TAG stop codons to wild-type codons decreases steeply over time in selective conditions. (*E*) Observed fitness of leucine synonyms at positions 582, 583, and 584.

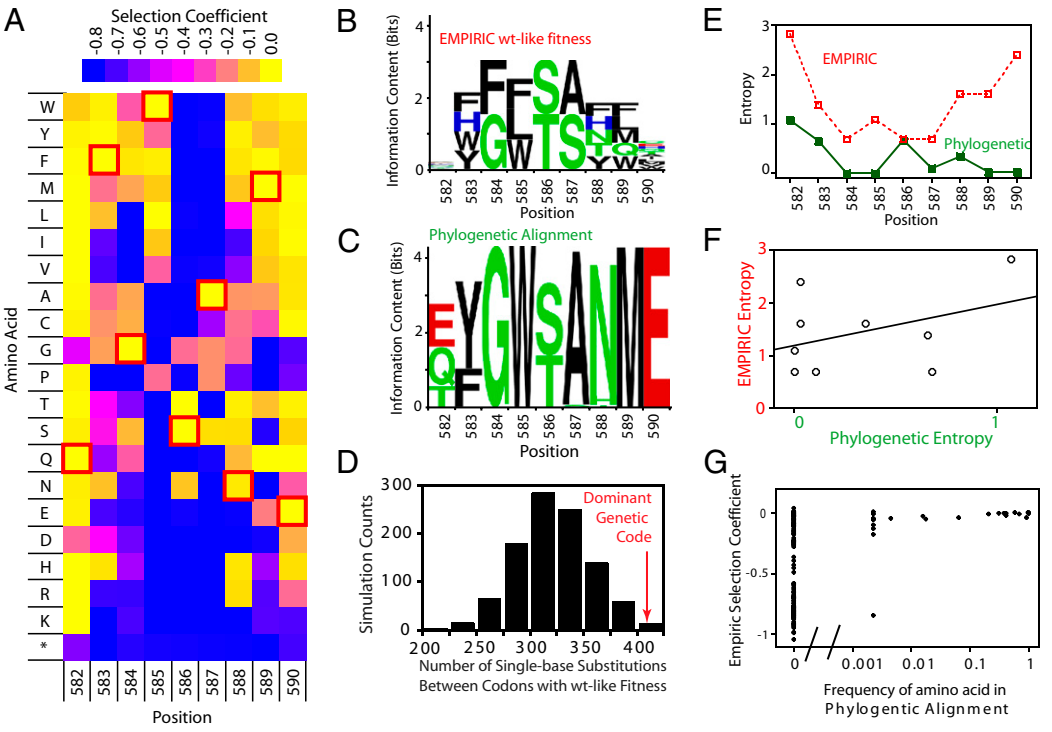

Fig. 3. Amino acid profile in phylogenetic alignment poorly predicts EMPIRIC fitness profile. (*A*) Heat map representation of the EMPIRIC fitness profile with the wild-type amino acids outlined in red. Information content logos generated from amino acids with WT-like EMPIRIC fitness (*B*) and a phylogenetic alignment of 448 Hsp90 protein sequences (*C*). (*D*) The dominant genetic code is optimized for single-base substitutions between codons with WT-like fitness compared with randomly simulated codes (+2.4σ). (*E*) Distribution of tolerated and phylogenetically observed amino acids expressed as an entropy where zero corresponds to a frozen position and 3 corresponds to unrestrained positions. (*F*) Relationship between tolerated amino acid profile from EMPIRIC fitness measurements and phylogenetic alignment. Linear regression indicates a very weak correlation with R^2 of 0.15. (*G*) EMPIRIC fitness analyzed as a function of amino acid prevalence in the phylogenetic alignment. Most amino acids observed in the phylogenetic alignment are well-tolerated when made in the yeast homolog.

bic amino acids is consistent with the finding that hydrophobic cores of proteins can be efficiently repacked with different hydrophobic sequences (16–18). All polar amino acids can form hydrogen bonds whose thermodynamic energy varies sharply with distance and angle (19), providing a rationale for the greater variability of the fitness of polar amino acids.

Based on the distribution of observed fitness effects, we classified mutants as WT-like if they had a selection coefficient within 5% of wild-type or better ($s > -0.05$). We chose this cutoff value because it is three times the rmsd between synonyms and, thus, represents a 99% confidence interval. We generated a logo of amino acids with WT-like fitness to analyze the patterns for underlying physical requirements for fitness at each position (Fig. 3*B*). Two of the nine positions analyzed exhibited a clear and consistent physical requirement for WT-like fitness: large hydrophobic side chains for position 585 and a γ-hydroxyl group for position 586. The physical properties for the preferred amino acids at these positions enable mechanistic predictions. The fitness preference for large hydrophobic amino acids of varied geometry (tryptophan, leucine, phenylalanine) at position 585, which is located on the surface of the Hsp90 structure, is consistent with involvement in loose contacts with hydrophobic partner molecules. The preference for only serine or threonine at position 586 indicates that the hydroxyl group common to both of these amino acids is important for function. In the structure of Hsp90, this hydroxyl group forms hydrogen bonds to two main-chain amide groups (Fig. S4). Although many hydrogen bonds are not important for protein function (20), our fitness measurements indicate that the hydrogen bonds formed at position 586 are critical for the function of Hsp90. Indeed, although tremendous strides have been made in understanding the relationship between protein structure and stability (17, 21), the ability to predict from structure the most important stabilizing contacts remains an unmet challenge. EMPIRIC fitness measurements provide a high-throughput approach to identify these important interactions experimentally and, hence, a route to develop and train predictive algorithms with improved accuracy.

Most of the other positions analyzed exhibit a preference for amino acids with varied physical properties. For example, at position 584, both glycine (the WT amino acid) and phenylalanine result in WT-like fitness. These amino acids differ dramatically in their physical properties: Phenylalanine is large and hydrophobic, and glycine is the smallest amino acid and imparts flexibility on the protein main chain. Despite their disparate physical properties, these two amino acids are clearly distinguished in fitness from all others. This type of physical plasticity illustrates the degenerate relationship between physics and biology: Biology is governed by physical interactions, but biological requirements can have multiple physical solutions.

The observed absence of phenylalanine at position 584 in a broad phylogenetic alignment (Fig. 3*C*) is consistent with the genetic code requiring two base substitutions to make this amino acid transition and the deleterious fitness effects of any of the single-base substitutions. Indeed, the fitness landscape combined with the genetic code may have broad impacts on evolutionary processes. The EMPIRIC approach provides a long-sought route (via larger datasets) to accurately examine the influence of the genetic code on evolution. For example, it makes it possible to determine whether the dominant genetic code is optimized for sampling evolutionarily neutral/favorable mutations. To address

this issue in our dataset of the fitness effects of >500 codon replacements, we counted the number of single-base substitutions that result in transitions between two codons with WT-like fitness for the dominant genetic code and for 1,000 randomly simulated genetic codes (Fig. 3*E*). We find that the genetic code is highly optimized (+2.4 σ) to favor single-base substitutions between codons with WT-like fitness compared with randomly generated codes as predicted from theoretical considerations of amino acid similarity (22). Thus, the genetic code generally permits single-base substitution pathways between codons with WT-like fitness.

To assess the EMPIRIC fitness profile against the evolutionary record, we compared our experimental results against the Hsp90 species tree (Fig. 3*C*). For almost every position, the amino acid entropy is higher for EMPIRIC fitness (Fig. 3*E*), indicating that more amino acid substitutions are compatible with high fitness in yeast Hsp90 than are observed in the phylogenetic alignment of Hsp90. Indeed, the relative amino acid entropy from the phylogenetic alignment was a poor predictor of the EMPIRIC entropy (Fig. 3*F*). The number and distribution of substitutions in the phylogenetic alignment did not accurately indicate the number of amino acids that would be compatible with high fitness experimentally. Many factors could contribute to this observation, including distinct fitness profiles under environmental conditions experienced in natural selection, and fitness differences beyond our ability to differentiate resulting in meaningful selection pressures in nature. Importantly, the genes in the phylogenetic alignment vary widely in their codon use and, hence, their nucleotide sequence (Fig. S5), indicating that mutational sampling occurred in this region and were subject to distinct evolutionary pressures in different organisms owing to varying selection intensities and/or effective population sizes.

Although the absence of a substitution in the phylogenetic alignment was a poor prognosticator of fitness effects, we find that all 17 amino acid substitutions observed at least twice in the phylogenetic alignment had WT-like ($s > -0.05$) experimental fitness (Fig. 3*G*). We do note that five amino acid substitutions that were observed only once in the phylogenetic alignment fall below this fitness cutoff, but only one is null-like ($s < -0.5$). In contrast, of the amino acid substitutions absent from the phylogenetic alignment, only 20% had WT-like experimental fitness (Fig. S6). Thus, the presence of an amino acid in a phylogenetic alignment was predictive that the corresponding point mutation in the yeast protein will be biochemically functional and evolutionarily nearly neutral. These observations indicate the important role of drift in the fixation of equivalent substitutions, and highlight the dominant role of purifying selection in suppressing deleterious fixations (23).

Discussion

The overall distribution of EMPIRIC fitness is bimodal (Fig. 4) with a clustering of amino acids with fitness similar to wild-type ($s \approx 0$), and a broader distribution of mutations of deleterious effect. Based on this distribution, we classified the first mode as "nearly neutral," and the second as "deleterious." Evaluating this directly measured distribution of fitness effects for >500 codon variants against the rich field of predictions from population genetics is of tremendous interest. Indeed, understanding this underlying distribution of selection coefficients has been a central focus of evolutionary biology over the past five decades (24). Contrary to recent inference made in *Drosophila* favoring models of frequent recurrent and strongly positive selection (25), but similar to inferences from genome-wide analyses of polymorphisms from *S. cerevisiae* and *S. paradoxus* (26), our direct observations in yeast are remarkably consistent with a nearly neutral model of molecular evolution (27), in which a large proportion of new mutations are strongly deleterious and are eliminated via purifying selection, whereas the great majority of

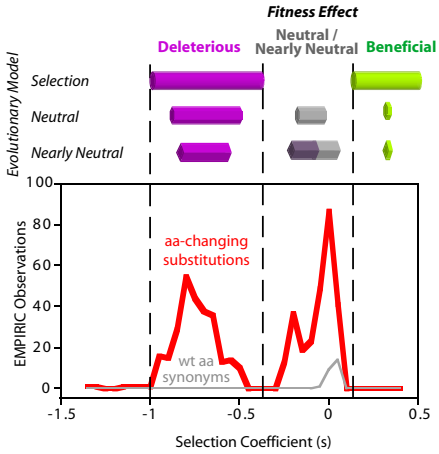

Fig. 4. Distribution of fitness effects of mutations from population genetic models and EMPIRIC measurement. For each model, the relative abundance of each type of fitness effect is illustrated by the length of the bar segment, and whereas the expected proportions of mutations in each class are unknown, the EMPIRIC approach provides an experimental measurement of class occupancy. The experimentally measured selection coefficients were binned in 0.05 increments.

remaining mutations are nearly neutral, with dynamics largely dictated by genetic drift (Fig. 4). Importantly, these initial results pertain to a conserved region of a highly conserved gene under a single growth condition. Examining and comparing the distribution of fitness effects for regions of variable levels of conservation, and under variable growth conditions, will be of extreme interest and should be a subject of future investigation.

As with the first techniques of protein electrophoresis allowing biologists to glimpse the extensive protein-level variation (28, 29) spurring the development of the neutral (30, 31) and nearly neutral theories of molecular evolution (27)—as well as the introduction of DNA sequencing technology allowing for inference to be drawn from nucleotide-level variation (32)—the EMPIRIC technique provides another layer of understanding, enabling direct measure of the distribution of selection coefficients by considering each possible point mutation at each site. In doing so, the EMPIRIC approach exposes a broad range of long-standing questions in population genetics to experimental examination including the effects of environmental conditions and genetic background on fitness landscapes.

Materials and Methods

Library Construction. In EMPIRIC fitness measurements, high-quality single-codon substitution libraries that avoid multiple mutations are important for two reasons: enabling fitness changes to be directly attributed to distinct mutations and providing a library size that can be accurately monitored in high-throughput. To generate these point mutant libraries, we optimized a cassette ligation strategy to rapidly generate plasmids containing single codons fully randomized to generate all 64 possibilities (Fig. S7).

We constructed a plasmid with a self-encoded removable fragment (SERF) composed of inverted BsaI restriction sites, such that treatment with this enzyme results in directional sticky ends and removal of the BsaI sites. To reduce background ligation products, an SphI site, introduced between the BsaI sites, was also digested in these vectors. We introduced a silent HpaII restriction site adjacent to the randomized region to facilitate adapter ligation required for deep sequencing. We generated the SERF vector by PCR from a yeast 417 shuttle plasmid containing a KanMX marker and the HSP82 (systematic name of yeast Hsp90) gene driven by a GPD promoter that expresses Hsp90 to endogenous levels (33). The HSP82 region of the SERF vector was fully sequenced to ensure the fidelity of the entire gene. Annealed oligonucleotide cassettes with a single codon randomized as NNN were ligated into the vector and transformed into *Escherichia coli*. Transformants were grown in mixed liquid culture from which plasmid DNA was isolated.

GENETICS

Growth Competition. We used iG170D *S. cerevisiae* cells (34) engineered with a temperature-sensitive chromosomal copy of Hsp90. This yeast strain grows robustly at the permissive temperature of 25 °C and rapidly slows growing at the nonpermissive temperature of 36 °C. Growth at the nonpermissive temperature is rescued with a plasmid bearing wild-type Hsp90 (Fig. 2*B*). Plasmid libraries for each randomized position were transformed into yeast by using the lithium acetate method. Transformants were grown in mixed liquid culture with G418 selection at 25 °C. After growing to saturation (2 d), cultures were outgrown overnight at 25 °C and an equal number of cells for each randomized position combined into a single culture. This culture was then heated in a water bath to 39 °C for 15 min to rapidly inactivate G170D Hsp90 and subsequently grown at 36 °C. These cultures were diluted every 8 h to maintain a culture density $<10^7$ cells per mL. Samples for analysis corresponding to $\approx 2 \times 10^8$ cells were harvested at different timepoints. All yeast growth was performed in synthetic dextrose media with 200 μg/mL G418 and 50 μg/mL ampicillin. Growth rates were determined for a strain with a wild-type Hsp90 rescue plasmid under identical conditions (doubling time of 4 h at 36 °C).

To validate our EMPIRIC approach, we experimentally determined the fitness effects of six point mutants through binary competition of strains fluorescently labeled with either CFP or YFP as described (35). Briefly, the CFP and YFP genes were chromosomally integrated into the iG170D parental yeast strain used in the EMPIRIC measurements. Plasmids containing either the wild-type yeast Hsp90 gene or a panel of seven mutants was introduced into each strain. Individually, CFP-labeled strains with either wild-type or one of the seven mutants were grown in competition with a YFP-labeled strain containing wild-type yeast Hsp90 under experiment conditions identical to the EMPIRIC experiment, and fluorescent measurements were made as a function of time.

DNA Preparation and Sequencing. Yeast pellets were lysed with zymolyase and total DNA was purified by using a silica column. A region containing all of the randomized codons was PCR amplified with primers that added a 3′ Illumina sequencing primer binding site. After purifying the PCR product on a silica column, a sticky end was created adjacent to the randomized region by digestion with the enzyme HpaII. This sticky end was ligated to an oligonucleotide cassette that included a three-base barcode with a Hamming distance (36) of two between any two codes (used to distinguish each time-point sample) and a 5′ Illumina sequencing primer-binding site. The ligation reactions for each time-point were column purified, combined and amplified in a single reaction with Illumina genomic sequencing primers. This PCR product was separated on an agarose gel and purified before 36-base Illumina sequencing.

Data Analysis. Illumina sequencing resulted in fastq file from which 2.6×10^7 reads were used for time-dependent analysis based on stringent accuracy requirements (>99% confidence across all 36 bases). The occurrence of each point mutant at each time-point was tabulated. Ten of the randomized codon sequences resulted in the formation of internal HpaII sites and were removed from further analysis. The ratio of each single-codon mutation relative to the wild-type sequence was calculated for each time point on a $\log_2$ scale. Selection coefficients (s) for each mutation were determined as the slope of this ratio to time in WT generations. Plots were thoroughly inspected for nonlinear relationships that would indicate a secondary genetic change during the time-course of the experiment, but none were observed. Selection coefficients for all stop codons were determined from the 12, 24, and 36 h time-points. Selection coefficients of mutants within three SDs of the stop codon mean ($s < -0.50$) were considered null-like and analyzed in the same manner. For all other mutations, selection coefficients were determined from the 12, 24, 36, 48, 60, 72, and 84 h time-points. To check for systematic influences of codon bias on fitness, we calculated the fitness difference between a codon and the average for all synonymous codons and compared this difference to the relative abundance of the codon in highly expressed yeast genes. For this analysis, we chose the 13 genes with the highest experimentally observed expression in *S. cerevisiae* (37). We averaged over all synonymous codons to calculate amino acid fitness and used the SD to estimate noise in our system. Amino acids were considered WT-like if their amino acid selection coefficient was greater than −0.05. Fitness logos of WT-like amino acids were generated by creating sequences with an equal number of each WT-like amino acid and the program weblogo (38). A similar logo was produced for the 448 sequences obtained by using BLASTP with the full-length yeast Hsp82 protein that aligned fully within this region.

Simulations of Alternate Genetic Codes. Genetic codes were chosen randomly for the 20 amino acids plus stop codons with the requirement that each of these 21 possible classes be encoded by at least one codon. The EMPIRIC fitness measurements were then searched by using these codes for all single-base substitutions between codons with WT-like fitness. The simulation was run for 1,000 iterations and compared with the dominant biological code.

ACKNOWLEDGMENTS. We thank P. Zamore for illuminating the potential of deep sequencing; O. Rando, A. Wong, and J. Bowie for thoughtful discussion; T. Pederson and T. Ohta for comments on the manuscript; and the Kishony laboratory for generously providing the CFP and YFP integration constructs. This work was supported by National Institutes of Health Grant R01-GM083038-01A, American Cancer Society Grant SG-08-17301-GMC (to D.N.A.B.), National Science Foundation Grant DEB-1002785, and a Worcester Foundation award (to J.D.J.).

1. Darwin C (1859) *On the Origin of Species by Means of Natural Selection* (John Murray, London).
2. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ (2007) Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 445:383–386.
3. Lunzer M, Miller SP, Felsheim R, Dean AM (2005) The biochemical architecture of an ancient adaptive landscape. *Science* 310:499–501.
4. Romero PA, Arnold FH (2009) Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol* 10:866–876.
5. Cooper TF, Rozen DE, Lenski RE (2003) Parallel changes in gene expression after 20,000 generations of evolution in Escherichia coli. *Proc Natl Acad Sci USA* 100: 1072–1077.
6. Lenski RE (1991) Quantifying fitness and gene stability in microorganisms. *Biotechnology* 15:173–192.
7. Lind PA, Berg OG, Andersson DI (2010) Mutational robustness of ribosomal protein genes. *Science* 330:825–827.
8. Borkovich KA, Farrelly FW, Finkelstein DB, Taulien J, Lindquist S (1989) hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. *Mol Cell Biol* 9:3919–3930.
9. Whitesell L, Mimnaugh EG, De Costa B, Myers CE, Neckers LM (1994) Inhibition of heat shock protein HSP90-pp60v-src heteroprotein complex formation by benzoquinone ansamycins: Essential role for stress proteins in oncogenic transformation. *Proc Natl Acad Sci USA* 91:8324–8328.
10. Ali MM, et al. (2006) Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* 440:1013–1017.
11. Wayne N, Bolon DN (2007) Dimerization of Hsp90 is required for in vivo function. Design and analysis of monomers and dimers. *J Biol Chem* 282:35386–35395.
12. Hegreness M, Shoresh N, Hartl D, Kishony R (2006) An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science* 311:1615–1617.
13. Nielsen R (2005) Molecular signatures of natural selection. *Annu Rev Genet* 39: 197–218.
14. Singh ND, et al. (2007) Patterns of mutation and selection at synonymous sites in Drosophila. *Mol Biol Evol* 24:2687–2697.
15. McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 351:652–654.
16. Lim WA, Sauer RT (1989) Alternative packing arrangements in the hydrophobic core of lambda repressor. *Nature* 339:31–36.
17. Dahiyat BI, Mayo SL (1997) De novo protein design: Fully automated sequence selection. *Science* 278:82–87.
18. Bowie JU, Reidhaar-Olson JF, Lim WA, Sauer RT (1990) Deciphering the message in protein sequences: Tolerance to amino acid substitutions. *Science* 247:1306–1310.
19. Mayo SL, Olafson BD, Goddard WA, 3rd (1990) DREIDING: A generic force field for molecular simulations. *J Phys Chem* 94:8897–8909.
20. Myers JK, Pace CN (1996) Hydrogen bonding stabilizes globular proteins. *Biophys J* 71: 2033–2039.
21. Kuhlman B, et al. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* 302:1364–1368.
22. Di Giulio M (1989) The extension reached by the minimization of the polarity distances during the evolution of the genetic code. *J Mol Evol* 29:288–293.
23. Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics* 134:1289–1303.
24. Sella G, Petrov DA, Przeworski M, Andolfatto P (2009) Pervasive natural selection in the Drosophila genome? *PLoS Genet* 5:e1000495.
25. Macpherson JM, Sella G, Davis JC, Petrov DA (2007) Genomewide spatial correspondence between nonsynonymous divergence and neutral polymorphism reveals extensive adaptation in Drosophila. *Genetics* 177:2083–2099.
26. Elyashiv E, et al. (2010) Shifts in the intensity of purifying selection: An analysis of genome-wide polymorphism data from two closely related yeast species. *Genome Res* 20:1558–1573.
27. Ohta T (1973) Slightly deleterious mutant substitutions in evolution. *Nature* 246: 96–98.
28. Lewontin RC, Hubby JL (1966) A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of Drosophila pseudoobscura. *Genetics* 54:595–609.
29. Harris H (1966) Enzyme polymorphisms in man. *Proc R Soc Lond B Biol Sci* 164: 298–310.

30. Kimura M (1968) Evolutionary rate at the molecular level. *Nature* 217:624–626.
31. King JL, Jukes TH (1969) Non-Darwinian evolution. *Science* 164:788–798.
32. Kreitman M (1983) Nucleotide polymorphism at the alcohol dehydrogenase locus of Drosophila melanogaster. *Nature* 304:412–417.
33. Chang HC, Lindquist S (1994) Conservation of Hsp90 macromolecular complexes in Saccharomyces cerevisiae. *J Biol Chem* 269:24983–24988.
34. Nathan DF, Lindquist S (1995) Mutational analysis of Hsp90 function: Interactions with a steroid receptor and a protein kinase. *Mol Cell Biol* 15:3917–3925.
35. DeLuna A, et al. (2008) Exposing the fitness contribution of duplicated genes. *Nat Genet* 40:676–681.
36. Hamming RW (1950) Error detecting and error correcting codes. *Bell Syst Tech J* 2: 147–160.
37. Ghaemmaghami S, et al. (2003) Global analysis of protein expression in yeast. *Nature* 425:737–741.
38. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Res* 14:1188–1190.