MOLECULAR EVOLUTION

The fitness landscape of a tRNA gene

Chuan Li, Wenfeng Qian,* Calum J. Maclean, Jianzhi Zhang†

Fitness landscapes describe the genotype-fitness relationship and represent major determinants of evolutionary trajectories. However, the vast genotype space, coupled with the difficulty of measuring fitness, has hindered the empirical determination of fitness landscapes. Combining precise gene replacement and next-generation sequencing, we quantified Darwinian fitness under a high-temperature challenge for more than 65,000 yeast strains, each carrying a unique variant of the single-copy $tRNA^{Arg}_{CCU}$ gene at its native genomic location. Approximately 1% of single point mutations in the gene were beneficial and 42% were deleterious. Almost half of all mutation pairs exhibited statistically significant epistasis, which had a strong negative bias, except when the mutations occurred at Watson-Crick paired sites. Fitness was broadly correlated with the predicted fraction of correctly folded transfer RNA (tRNA) molecules, thereby revealing a biophysical basis of the fitness landscape.

Fitness landscapes can provide information about the direction and magnitude of natural selection and can elucidate evolutionary trajectories (*1*), but their empirical determination requires quantifying the fitness of an astronomically large number of possible genotypes. Past studies were limited to relatively few genotypes (*2*, *3*). Next-generation DNA sequencing (NGS) has permitted the analysis of many more genotypes (*4*–*11*), but research has focused on biochemical functions (*4*, *6*, *8*–*12*) rather than fitness. In the few fitness landscapes reported, only a small fraction of sites or combinations of mutations per gene were examined (*5*–*7*, *9*).

Here, we combined gene replacement in *Saccharomyces cerevisiae* with an NGS-based fitness assay to determine the fitness landscape of a tRNA gene. tRNAs carry amino acids to ribosomes for protein synthesis, and mutations can cause diseases such as cardiomyopathy and deafness (*13*). tRNA genes are typically shorter than 90 nucleotides, allowing coverage by a single Illumina sequencing read. We focused on $tRNA^{Arg}_{CCU}$, which recognizes the arginine codon AGG via its anticodon 5′-CCU-3′. Because AGG is also recognizable by $tRNA^{Arg}_{UCU}$ via wobble pairing, $tRNA^{Arg}_{CCU}$ is known to be encoded by a single-copy nonessential gene in *S. cerevisiae* (*14*). Deleting the $tRNA^{Arg}_{CCU}$ gene (fig. S1 and table S1) reduces growth rates in both fermentable (YPD) and nonfermentable (YPG) yeast growth media, a problem exacerbated by high temperature (fig. S2).

We chemically synthesized the 72-nucleotide $tRNA^{Arg}_{CCU}$ gene with a mutation rate of 3% per site (1% per alternate nucleotide) at 69 sites; for technical reasons, we kept the remaining three sites invariant (*15*). Using these variants, we constructed a pool of $>10^5$ strains, each carrying a $tRNA^{Arg}_{CCU}$ gene variant at its native genomic location (Fig. 1 and fig. S1). Six parallel competitions of this strain pool were performed in YPD medium at 37°C for 24 hours. The $tRNA^{Arg}_{CCU}$ gene amplicons from the common starting population (T_0) and those from six replicate competitions (T_{24}) were sequenced with 100-nucleotide paired-end NGS (Fig. 1 and table S2). Genotype frequencies were highly correlated between two T_0 technical repeats (Pearson's correlation r = 0.99997; fig. S3A) and among six T_{24} biological replicates (average r = 0.9987; fig. S3B) (*15*). Changes in genotype frequencies between T_0 and T_{24} were used to determine the Darwinian fitness of each genotype relative to the wild type (*15*). For our fitness estimation, we considered 65,537 genotypes with read counts of ≥100 at T_0. In theory, a cell that does not divide has a fitness of 0.5 (*16*). Because $tRNA^{Arg}_{CCU}$ mutations are unlikely to be fatal, we set genotype fitness at 0.5 when the estimated fitness is <0.5 (due to stochasticity) (*15*). Fitness values from these en masse competitions agreed with those obtained from growth curve and pairwise competition (fig. S4), as reported previously (*16*). We observed strong fitness correlations across diverse environments for a subset of genotypes examined (fig. S5), which suggests that our fitness landscape is broadly relevant (*15*).

We estimated the fitness (f) of all 207 possible mutants that differ from the wild type by one point mutation (N1 mutants) and calculated the average mutant fitness at each site (Fig. 2A). Average fitness decreased to <0.75 by mutation at nine key sites, including all three anticodon positions (table S3), three TΨC loop sites, one D stem site, and two paired TΨC stem sites (Fig. 2A). The TΨC loop and stem sites are components of the B-box region of the internal promoter, with C55 essential for both TFIIIC transcription factor binding and polymerase III (Pol III) transcription (*17*). In addition, some sites such as T54 are ubiquitously posttranscriptionally modified (*18*). By contrast, the average mutant fitness is ≥0.95 at 30 sites (Fig. 2A). Overall, mutations in loops are more deleterious than in stems (P = 0.01, Mann-Whitney U test), although this difference becomes insignificant after excluding the anticodon (P = 0.09). Unsurprisingly, different mutations at a site have different fitness effects (fig. S6). For example, mutation C11T in the D stem is tolerated (f_{C11T} ± SE = 1.006 ± 0.036), but C11A and C11G are not (f_{C11A} = 0.676 ± 0.030 and f_{C11G} = 0.661 ± 0.035), likely because of G:U paring in RNA.

The fitness distribution of N1 mutants shows a mean of 0.89 and a peak at 1 (Fig. 2B). Only 1% of mutations are significantly beneficial (nominal P < 0.05; t test based on the six replicates), whereas 42% are significantly deleterious. We estimated the fitness of 61% of all possible genotypes carrying two mutations (N2 mutants) and observed a left-shifted distribution peaking at 0.50 and 0.67 (Fig. 2C). We also estimated the fitness of 1.6% of genotypes with three mutations (N3 mutants); they exhibited a distribution with only one dominant peak at 0.5, indicating that many triple mutations completely suppress yeast growth in the en masse competition (Fig. 2D).

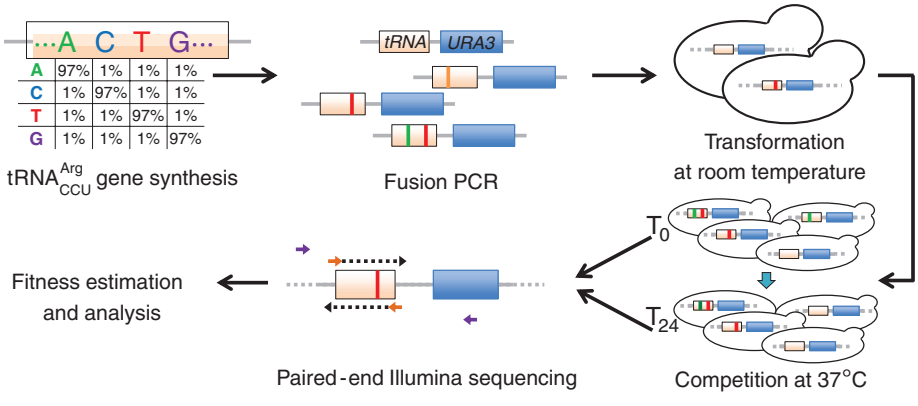

Fig. 1. Determining the fitness landscape of the yeast $tRNA^{Arg}_{CCU}$ gene. Chemically synthesized $tRNA^{Arg}_{CCU}$ gene variants are fused with the marker gene *URA3* before placement at the native $tRNA^{Arg}_{CCU}$ locus. The tRNA variant–carrying cells are competed with one another. The fitness of each $tRNA^{Arg}_{CCU}$ genotype relative to the wild type is calculated from the relative frequency change of paired-end sequencing reads covering the tRNA gene variant during competition (fig. S1) (*15*).

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA.
*Present address: Key Laboratory of Genetic Network Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. †Corresponding author. Email: jianzhi@umich.edu

The fitness distribution narrows and shifts further toward 0.5 in strains carrying more than three mutations (Fig. 2E).

Fitness landscapes allow prediction of evolution, because sites where mutations are on average more harmful should be evolutionarily more conserved. We aligned 200 nonredundant $tRNA^{Arg}_{CCU}$ gene sequences across the eukaryotic phylogeny (*15*). The percentage of sequences having the same nucleotide as yeast at a given site is negatively correlated with the average fitness upon mutation at the site (Spearman's $\rho = -0.61$, $P = 2 \times 10^{-8}$; Fig. 2F). Among N1 mutants, the number of times that a mutant nucleotide appears in the 200 sequences is positively correlated with the fitness of the mutant ($\rho = 0.51$, $P = 2 \times 10^{-15}$; Fig. 2G). Furthermore, mutations observed in other eukaryotes have smaller fitness costs in yeast than those unobserved in other eukaryotes ($P = 9 \times 10^{-6}$, Mann-Whitney *U* test).

Two mutations may interact with each other, creating epistasis (ε) with functional and evolutionary implications (*19*). We estimated ε within the tRNA gene from the fitness of 12,985 N2 mutants and 207 N1 mutants (Fig. 3A) (*15*). ε is negatively biased, with only 34% positive values ($P < 10^{-300}$, binomial test; Fig. 3B and figs. S7A and S8). Among the 45% of ε values that differ significantly from 0 (nominal $P < 0.05$, *t* test based on the six replicates), 86% are negative ($P < 10^{-300}$, binomial test; Fig. 3B and figs. S7A and S8). Consistent with the overall negative ε, the mean fitness of N2 mutants (0.75) is lower than that predicted from N1 mutants under the assumption of no epistasis (0.81) (Fig. 2E). Note that as the first mutation becomes more deleterious, the mean epistasis between this mutation and the next mutation becomes less negative and, in some cases, even positive (Fig. 3C and fig. S9), similar to between-gene epistasis involving an essential gene (*20*). Consequently, the larger the fitness cost of the first mutation, the smaller the mean fitness cost of the second mutation (Fig. 3D and fig. S10). Pairwise epistasis involving three or four mutations is also negatively biased (fig. S11). Consistently, N3 to N8 mutants all show lower average fitness than expected under the assumption of no epistasis (Fig. 2E).

The distribution of epistasis between mutations at paired sites is expected to differ from the above general pattern, because different Watson-Crick (WC) pairs may be functionally similar (*21*). We estimated the fitness of 71% of all possible N2 mutants at WC paired sites. Among the 41 cases that switched from one WC pair to another, 23 (56%) have positive ε (Fig. 3E). Among the 80 N2 mutants that destroyed WC pairing, 39 (49%) showed positive ε (Fig. 3F). The ε values are more positive for each of these two groups than for N2 mutants where the two mutations do not occur at paired sites ($P = 7 \times 10^{-6}$ and 2.6×10^{-3}, respectively; Mann-Whitney *U* test). Furthermore, ε is significantly more positive in the 41 cases with restored WC pairing than in the 80 cases with destroyed pairing ($P = 0.04$). These two trends also apply to cases with significant epistasis (corresponding to $P = 3 \times 10^{-5}$, 0.01, and 0.01, respectively; Fig. 3, E and F, and fig. S7, B and C). However, epistasis is not always positive between paired sites, likely because base pairing is not the sole function of the nucleotides at paired sites. We observed 160 cases of significant sign epistasis (*15*), which is of special interest because it may block potential paths for adaptation (*2*). We also detected ε with opposite signs in different genetic backgrounds, indicating a high-order epistasis (table S4).

A tRNA can fold into multiple secondary structures. We computationally predicted the proportion of $tRNA^{Arg}_{CCU}$ molecules that are potentially functional (i.e., correctly folded with no anticodon mutation) for each genotype (P_{func}). Raising P_{func} increases fitness ($\rho = 0.40$, $P < 10^{-300}$), albeit with diminishing returns (Fig. 4A), and this correlation holds after controlling for mutation number ($\rho = 0.26$, 0.37, and 0.24 for N1, N2, and N3 mutants, respectively). Because computational prediction of RNA secondary structures is only moderately accurate, the P_{func}-fitness correlation demonstrates an important role of P_{func} in shaping the tRNA fitness landscape. Nonetheless, after controlling for P_{func}, mutant fitness still correlates with mutation

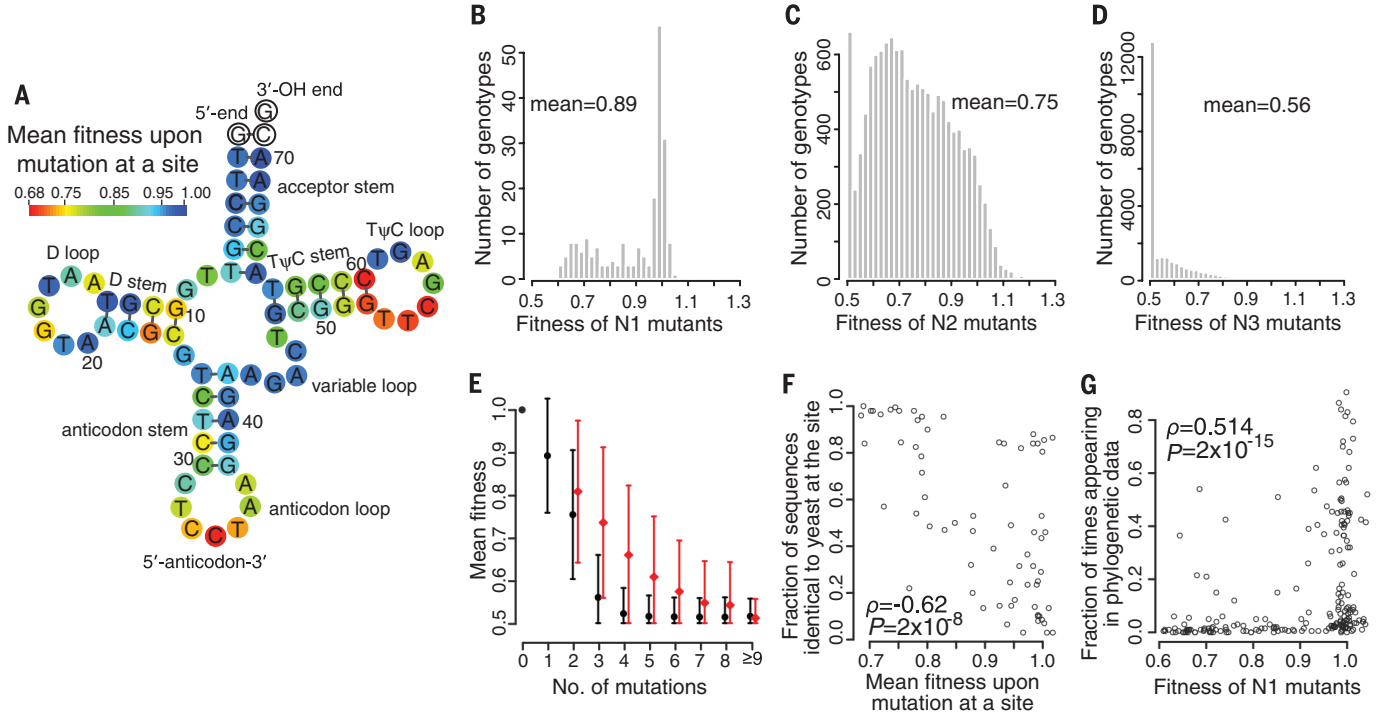

Fig. 2. Yeast $tRNA^{Arg}_{CCU}$ gene fitness landscape. (**A**) Average fitness upon a mutation at each site. White circles indicate invariant sites. (**B** to **D**) Fitness distributions of N1 (B), N2 (C), and N3 (D) mutants. (**E**) Mean observed fitness (black circles) decreases with mutation number. Red circles show mean expected fitness without epistasis (right-shifted for viewing). Error bars denote SD. (**F**) Fraction of the 200 eukaryotic $tRNA^{Arg}_{CCU}$ genes with the same nucleotide as yeast at a given site decreases with the average fitness upon mutation at the site in yeast. Each dot represents one of the 69 examined tRNA sites. (**G**) Fraction of times that a mutant nucleotide appears in the 200 sequences increases with the fitness of the mutant in yeast. Each dot represents a N1 mutant. In (F) and (G), ρ is the rank correlation coefficient; *P* values are from *t* tests.

number [ρ = −0.51, $P < 10^{-300}$; see also locally weighted polynomial regressions (LOESS) for N1, N2, and N3 mutants in Fig. 4B], which suggests that other factors also have an impact on fitness.

To investigate whether P_{func} explains epistasis, we computed epistasis using the fitness of N1 and N2 mutants predicted from their respective P_{func}-fitness regression curves (Fig. 4B) and observed a significant correlation between the predicted and observed epistasis (ρ = 0.04, $P = 2.7 \times 10^{-5}$). The weakness of this correlation is at least partly because epistasis is computed from three fitness measurements (or predictions) and is therefore associated with a considerable error. There is a similar bias in predicted epistasis toward negative values (Fig. 4C), but further analyses suggest that it probably arises from factors other than tRNA folding (*15*). These results regarding P_{func} and epistasis are not unexpected, given that a tRNA site can be involved in multiple molecular functions (*17*, *18*).

Our results clarify the in vivo fitness landscape of a yeast tRNA gene under a high-temperature

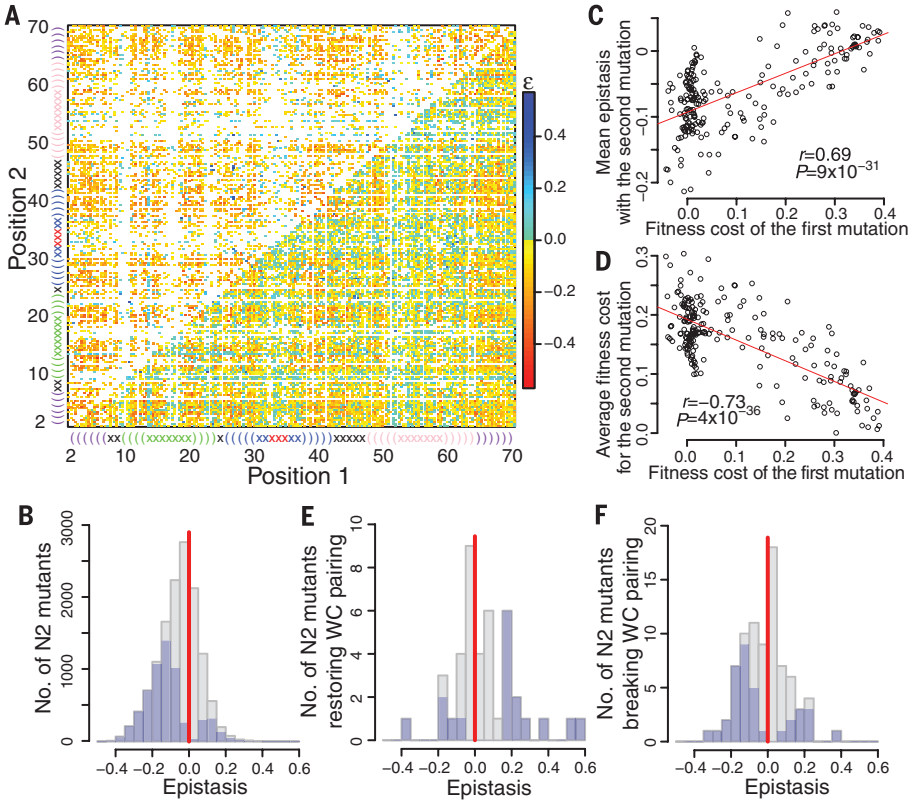

Fig. 3. Epistasis (ε) in fitness between point mutations in the $tRNA^{Arg}_{CCU}$ gene is negatively biased. (**A**) Epistasis between point mutations. Lower right triangle shows all pairwise epistasis (white = not estimated); upper left triangle shows statistically significant epistasis (white = no estimation or insignificant). $tRNA^{Arg}_{CCU}$ secondary structure is plotted linearly. Parentheses and crosses show stem and loop sites, respectively. Same color indicates sites in the same loop or stem. Each site has three mutations. (**B**) Distributions of pairwise epistasis (gray) and statistically significant pairwise epistasis (blue) among 12,985 mutation pairs. (**C**) Mean epistasis between first and second mutations increases with the fitness cost of the first mutation. (**D**) Mean fitness cost of the second mutation decreases with the fitness cost of the first mutation. In (C) and (D), the Pearson's correlation (*r*), associated *P* value, and linear regression (red) are shown. (**E** and **F**) Distributions of epistasis (gray) and statistically significant epistasis (blue) between pairs of mutations that convert a Watson-Crick (WC) base pair to another WC pair (E) or break a WC pair in stems (F). In (B), (E), and (F), the vertical red line shows zero epistasis.

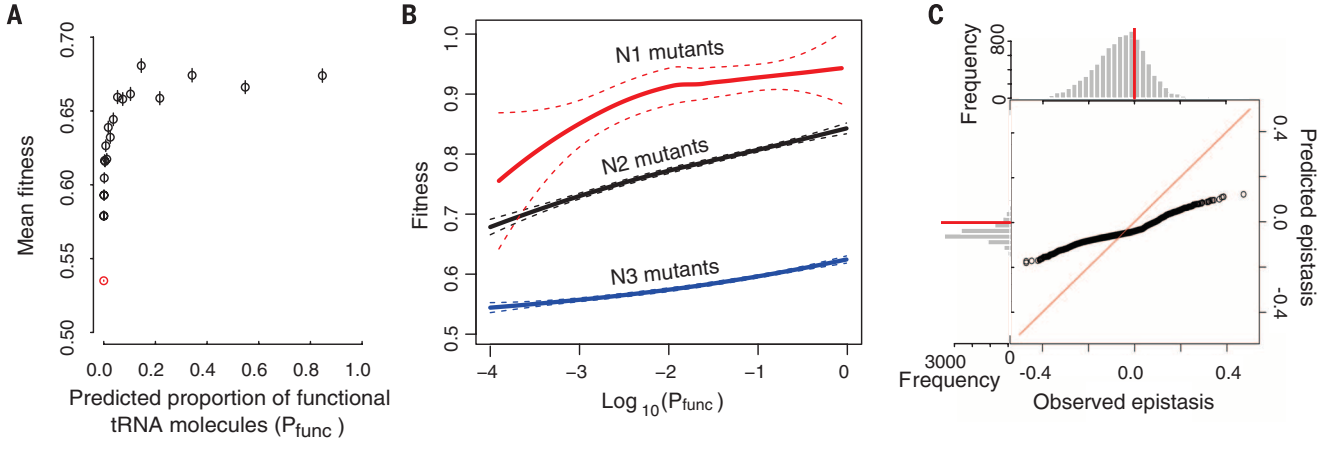

Fig. 4. $tRNA^{Arg}_{CCU}$ folding offers a mechanistic explanation of the fitness landscape. (**A**) Relationship between the predicted proportion of tRNA molecules that are functional (P_{func}) for a genotype and its fitness. Genotypes (with $P_{func} \geq 10^{-4}$) are ranked by P_{func} and grouped into 20 equal-size bins; mean P_{func} and mean fitness ± SE of each bin are shown. The red dot represents all variants with $P_{func} < 10^{-4}$. (**B**) LOESS regression curves between P_{func} and fitness for N1, N2, and N3 mutants, respectively, with dashed lines indicating 95% confidence intervals. (**C**) Quantile-quantile plot between epistasis predicted from P_{func} values using N1 and N2 LOESS curves and observed epistasis. The *i*th dot from the left shows the *i*th smallest predicted epistasis value (*y* axis) and *i*th smallest observed epistasis value (*x* axis). Red diagonal line shows the ideal situation of *y* = *x*. Above and left of the plot are frequency distributions of observed and predicted epistasis, respectively. Red horizontal and vertical lines indicate zero epistasis.

challenge. Broadly consistent with the neutral theory, beneficial mutations are rare (1%), relative to deleterious (42%) and (nearly) neutral mutations (57%). We found widespread intragenic epistasis between mutations, consistent with studies at smaller scales (*1*). Intriguingly, 86% of significant epistasis is negative, indicating that the fitness cost of the second mutation is on average greater than that of the first. A bias toward negative epistasis was also observed in protein genes (*7*, *10*, *11*, *22*); hence, this may be a general trend. Variation in fitness is partially explained by the predicted fraction of correctly folded tRNA molecules; this implies the existence of general principles underlying complex fitness landscapes. Our tRNA variant library provides a resource in which various mechanisms contributing to the tRNA's fitness landscape can be evaluated, and the methodology developed here is applicable to the study of fitness landscapes of longer genomic segments, including protein genes.

REFERENCES AND NOTES

1. J. A. de Visser, J. Krug, *Nat. Rev. Genet.* **15**, 480–490 (2014).
2. D. M. Weinreich, N. F. Delaney, M. A. Depristo, D. L. Hartl, *Science* **312**, 111–114 (2006).
3. P. A. Lind, O. G. Berg, D. I. Andersson, *Science* **330**, 825–827 (2010).
4. J. N. Pitt, A. R. Ferré-D'Amaré, *Science* **330**, 376–379 (2010).
5. R. T. Hietpas, J. D. Jensen, D. N. Bolon, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7896–7901 (2011).
6. G. M. Findlay, E. A. Boyle, R. J. Hause, J. C. Klein, J. Shendure, *Nature* **513**, 120–123 (2014).
7. C. Bank, R. T. Hietpas, J. D. Jensen, D. N. Bolon, *Mol. Biol. Evol.* **32**, 229–238 (2015).
8. M. P. Guy *et al*., *Genes Dev.* **28**, 1721–1732 (2014).
9. A. Melnikov, P. Rogov, L. Wang, A. Gnirke, T. S. Mikkelsen, *Nucleic Acids Res.* **42**, e112 (2014).
10. D. Melamed, D. L. Young, C. E. Gamble, C. R. Miller, S. Fields, *RNA* **19**, 1537–1551 (2013).
11. C. A. Olson, N. C. Wu, R. Sun, *Curr. Biol.* **24**, 2643–2651 (2014).
12. T. Hinkley *et al*., *Nat. Genet.* **43**, 487–489 (2011).
13. J. A. Abbott, C. S. Francklyn, S. M. Robey-Bond, *Front. Genet.* **5**, 158 (2014).
14. Z. Bloom-Ackermann *et al*., *PLOS Genet.* **10**, e1004084 (2014).
15. See supplementary materials on *Science* Online.
16. W. Qian, D. Ma, C. Xiao, Z. Wang, J. Zhang, *Cell Rep.* **2**, 1399–1410 (2012).
17. S. Hiraga, S. Botsios, D. Donze, A. D. Donaldson, *Mol. Biol. Cell* **23**, 2741–2754 (2012).
18. E. M. Phizicky, A. K. Hopper, *Genes Dev.* **24**, 1832–1860 (2010).
19. P. C. Phillips, *Nat. Rev. Genet.* **9**, 855–867 (2008).
20. X. He, W. Qian, Z. Wang, Y. Li, J. Zhang, *Nat. Genet.* **42**, 272–276 (2010).
21. M. V. Meer, A. S. Kondrashov, Y. Artzy-Randrup, F. A. Kondrashov, *Nature* **464**, 279–282 (2010).
22. S. Bershtein, M. Segal, R. Bekerman, N. Tokuriki, D. S. Tawfik, *Nature* **444**, 929–932 (2006).

ACKNOWLEDGMENTS

We thank S. Cho, W.-C. Ho, G. Kudla, and J.-R. Yang for valuable comments. Supported by NSF DDIG grant DEB-1501788 (J.Z. and C.L.) and NIH grant R01GM103232 (J.Z.). The NCBI accession number for the sequencing data is PRJNA311172.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6287/837/suppl/DC1
Materials and Methods
Figs. S1 to S11
Tables S1 to S4
References (*23*–*27*)

11 December 2015; accepted 23 March 2016
Published online 14 April 2016
10.1126/science.aae0568

MOLECULAR EVOLUTION

Network of epistatic interactions within a yeast snoRNA

Olga Puchta,[1] Botond Cseke,[2]* Hubert Czaja,[3] David Tollervey,[4,5] Guido Sanguinetti,[2,4] Grzegorz Kudla[1]†

Epistatic interactions play a fundamental role in molecular evolution, but little is known about the spatial distribution of these interactions within genes. To systematically survey a model landscape of intragenic epistasis, we quantified the fitness of ~60,000 *Saccharomyces cerevisiae* strains expressing randomly mutated variants of the 333-nucleotide-long U3 small nucleolar RNA (snoRNA). The fitness effects of individual mutations were correlated with evolutionary conservation and structural stability. Many mutations had small individual effects but had large effects in the context of additional mutations, which indicated negative epistasis. Clusters of negative interactions were explained by local thermodynamic threshold effects, whereas positive interactions were enriched among large-effect sites and between base-paired nucleotides. We conclude that high-throughput mapping of intragenic epistasis can identify key structural and functional features of macromolecules.

The effect of a mutation on phenotype may depend on the presence of additional mutations. This phenomenon, known as epistasis, explains synthetic lethal interactions, in which a combination of two individually viable mutations causes death, and compensatory interactions, in which the fitness cost of a mutation is reduced by a second mutation (*1*, *2*). Epistasis plays a major role in evolution; it determines the accessibility of mutational pathways (*3*) and thereby influences the rate of adaptation and the diversity and robustness of genetic variants (*4*, *5*). Although genome-wide studies have revealed a network of intergenic epistasis (*6*), it has been suggested that interactions within genes may be even more common (*7*–*11*). However, previous studies focused on relatively small networks of interactions, and the comprehensive pattern of epistasis has not yet been determined for any gene.

We used "doped" oligonucleotides to synthesize ~130,000 randomly mutated variants of the 333-nucleotide *Saccharomyces cerevisiae* gene *SNR17A*, which encodes the U3 small nucleolar RNA. U3 forms base pairs with the primary ribosomal RNA (rRNA) transcript (pre-rRNA), and this interaction is required for pre-rRNA cleavage and 18*S* rRNA biogenesis. Our mutagenesis approach ensures uniform coverage of mutations among positions 7 to 333, which encompass 98% of the gene, and prevents bias toward specific types of mutations (Fig. 1A and fig. S1). We generated two independent mutant libraries, which contained, on average, 3 and 10 single-nucleotide polymorphisms (SNPs) per allele, respectively. In addition to the SNPs, 43.6% of variants also contained short deletions [median length, 1 nucleotide (nt)] or insertions (median length, 1 nt). All 981 (3 × 327) possible point mutations were represented in the library, and 99.4% of the 53,301 (327 × 326/2) possible pairs of sites were jointly mutated, most of them in alleles that contained additional mutations. To facilitate unambiguous identification of variants by high-throughput sequencing, we tagged each variant with a unique 20-nt barcode (Fig. 1A) placed in a nontranscribed region downstream of the U3 gene to minimize interference with function.

To measure fitness, we used the D343 yeast strain, which contains a single copy of the wild-type U3 gene under the control of a galactose-inducible promoter (*12*). D343 cells can grow in galactose-containing medium, but shifting to glucose results in down-regulation of U3 and growth arrest. Transformation of wild-type U3 on a plasmid allows the cells to survive on glucose, but nonfunctional U3 mutants do not support growth (fig. S2). We transformed D343 cells with centromeric plasmids carrying the U3 mutant libraries and measured the frequency of each mutant during competitive growth on glucose (Fig. 1B). As expected, nonfunctional variants decreased in frequency during the competition, whereas the wild-type gene increased (Fig. 1C). Growth patterns were reproducible between four replicate experiments and across replicate U3 variants within an experiment (Fig. 1D and fig. S3).

We measured the logarithm of relative fitness (log fitness) of ~60,000 variants that passed quality filters by fitting exponential decay curves to the barcode count data (*13*). Log fitness of wild-type U3 was set to 0. We first focused on the effects of single mutations in an otherwise wild-type gene (*13*). In most positions, mutations were tolerated with minimal effect on fitness (Fig. 2A). The exceptions were the conserved protein binding sites known as boxes B, C, C′, and D, in which mutations are lethal or highly deleterious. In addition, a moderate fitness decrease

[1]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, Scotland, UK. [2]School of Informatics, University of Edinburgh, Edinburgh, Scotland, UK. [3]Scott Tiger SA, Warsaw, Poland. [4]SynthSys, Centre for Synthetic and Systems Biology, University of Edinburgh, Edinburgh, Scotland, UK. [5]Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh, Scotland, UK.
*Present address: Microsoft Research Cambridge, Cambridge, UK.
†Corresponding author. Email: gkudla@gmail.com