

# Pairwise and higher-order genetic interactions during the evolution of a tRNA

Júlia Domingo<sup>1</sup>, Guillaume Diss<sup>1</sup> & Ben Lehner<sup>1,2,3\*</sup>

**A central question in genetics and evolution is the extent to which the outcomes of mutations change depending on the genetic context in which they occur<sup>1–3</sup>. Pairwise interactions between mutations have been systematically mapped within<sup>4–18</sup> and between<sup>19</sup> genes, and have been shown to contribute substantially to phenotypic variation among individuals<sup>20</sup>. However, the extent to which genetic interactions themselves are stable or dynamic across genotypes is unclear<sup>21,22</sup>. Here we quantify more than 45,000 genetic interactions between the same 87 pairs of mutations across more than 500 closely related genotypes of a yeast tRNA. Notably, all pairs of mutations interacted in at least 9% of genetic backgrounds and all pairs switched from interacting positively to interacting negatively in different genotypes (false discovery rate < 0.1). Higher-order interactions are also abundant and dynamic across genotypes. The epistasis in this tRNA means that all individual mutations switch from detrimental to beneficial, even in closely related genotypes. As a consequence, accurate genetic prediction requires mutation effects to be measured across different genetic backgrounds and the use of higher-order epistatic terms.**

Genetic (epistatic) interactions have been extensively mapped between pairs of mutations within individual genes<sup>4–18</sup>, and also between individual alleles of many different genes<sup>19</sup>. However, the pairwise mapping of interactions only provides a limited view of genotype space, which has a vast combinatorial size<sup>22</sup>. Interactions between genes have been reported as only poorly or moderately conserved between species<sup>21</sup>. Moreover, analyses of the effects of combinations of mutations within individual genes have pointed to the importance of higher-order epistasis<sup>22–25</sup>, in which mutations interact beyond pairwise interactions to determine mutation effect.

To directly test the extent to which the effects of mutations and the interactions between mutations are stable or change depending upon the genotype in which they occur, we designed an experiment in which mutation effects and interactions are quantified across a large number of closely related genetic backgrounds. As a model system, we used the single-copy arginine-CCU tRNA (tRNA-Arg(CCU)) gene that is conditionally required for the growth of budding yeast (Extended Data Fig. 1a) and for which pairwise interactions have been previously mapped in one genetic background<sup>15</sup>. The small size of the gene allowed us to design a library that covered all 5,184 ( $2^6 \times 3^4$ ) genotypes containing the 14 nucleotide substitutions observed in ten positions in post-whole-genome-duplication yeast species<sup>26</sup> (Fig. 1a, b). Each genotype therefore varies from zero to a maximum of ten nucleotides divergence from the *Saccharomyces cerevisiae* tRNA sequence (Extended Data Fig. 1b). After transformation of the library into *S. cerevisiae*, we performed six selection experiments in parallel to quantify the relative fitness of each of the 5,184 variants under restrictive conditions (high temperature and 1 M NaCl) (Fig. 1c). The fitness of each genotype was quantified as the change in its abundance in each culture between the beginning and end of the competition period determined using deep sequencing with a hierarchical error model and normalized in log scale to the fitness of the *S. cerevisiae*

genotype (henceforth ‘fitness’). After filtering, we obtained fitness measurements for 4,176 variants (Supplementary Table 1) that correlated well across replicates (Fig. 1d). The median fitness declines as the number of mutations increases but there are still many combinations of mutations with high fitness amongst genotypes that are far from the reference genotype (Fig. 1e).

We first examined the fitness consequences of single mutations and how these change across different genetic backgrounds (Fig. 2a). In the *S. cerevisiae* genotype, six of the 14 individual mutations were detrimental (Fig. 2b). However, when the same 14 mutations were made in the tRNA genotypes of the other six extant species (these alternative ‘wild-type’ tRNAs have fitness very close to the *S. cerevisiae* tRNA when expressed in *S. cerevisiae*, Supplementary Table 2), their effects changed substantially (Fig. 2b). For example, the mutation C66A had no effect in the *S. cerevisiae* background but became detrimental in the *Candida glabrata* tRNA, which only differs by two substitutions (paired *t*-test,  $q = 0.006$ ,  $n = 6$ ). Indeed, 11 out of 14 mutations had effects that changed across these seven tRNAs from different species (Extended Data Fig. 2a, false discovery rate (FDR) < 0.1).

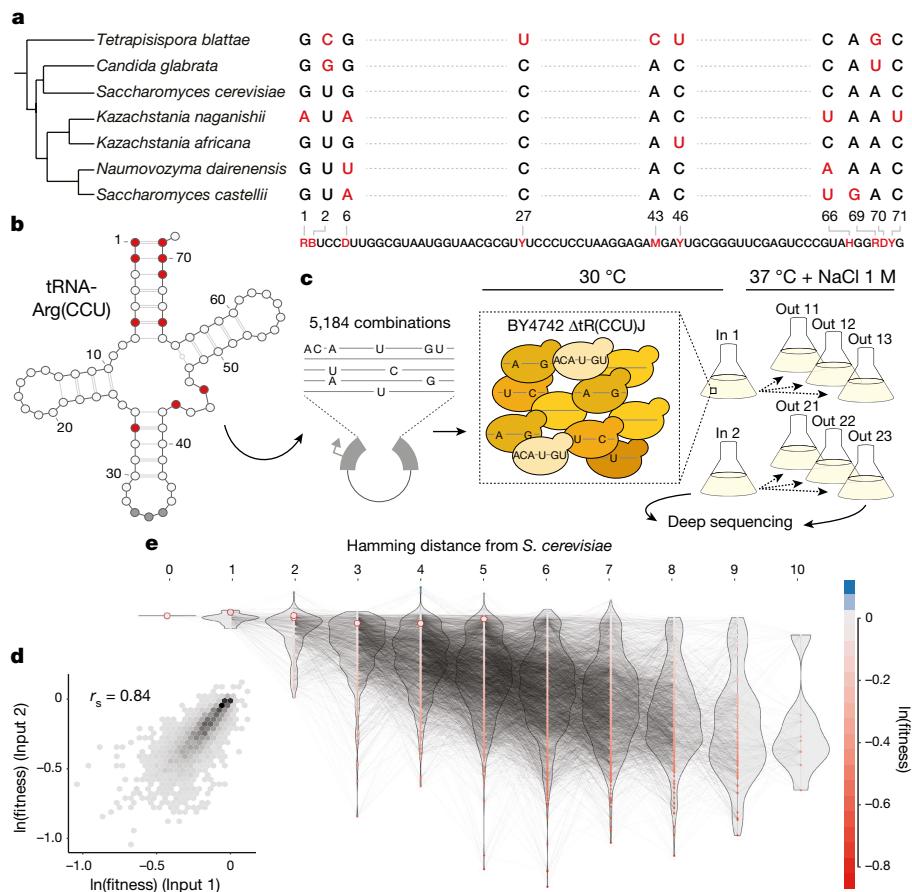
We next compared the effects of the single mutations across the complete set of genetic backgrounds in the library. In total, we tested each mutation in a median of 1,449 genetic backgrounds (minimum = 1,088, maximum = 1,993, Extended Data Fig. 1c, d). Notably, we found that every mutation was both detrimental and beneficial in a substantial number of genetic backgrounds (Fig. 2b, c, median number of backgrounds in which the less frequent sign was observed = 6.4%; minimum = 3.4%; maximum = 11.9% across all 14 mutations, FDR < 0.1,  $n = 21,450$ ). Restricting the analyses to background genotypes with high or intermediate fitness, to genotypes with high input read counts, or to genotypes with few mutations did not change this conclusion (Extended Data Fig. 2b). Thus, all mutations have effects that switch from beneficial to detrimental in closely related genotypes.

To investigate the interactions between mutations that underlie these changes in mutation effects, we first quantified pairwise genetic interactions between the 14 mutations, which is a total of 87 pairs in any genotype. We define epistasis as the difference between the fitness of each double mutant and the sum of the fitness of the two corresponding individual mutations. Consistent with previous results<sup>15</sup>, in the *S. cerevisiae* genotype, many pairs of mutations (40.2%, 35 out of 87) had combined fitness effects that were more detrimental than expected (negative epistasis) and only a few had effects that were less detrimental than expected (positive epistasis, 5.7%, 5 out of 87, FDR < 0.1, Fig. 3a). However, these interactions changed when they were tested in tRNAs from different species (Fig. 3b, c, Extended Data Fig. 3), with 83 out of the 87 interactions differing across the species ( $n = 1,000$  paired *t*-tests, FDR < 0.1, Extended Data Fig. 4).

We next analysed how the 87 interactions changed across all the genetic backgrounds in the library. Each interaction was quantified in a median of 506 genetic backgrounds (minimum = 240, maximum = 946, Extended Data Fig. 1d). Notably, all 87 interactions switched from positive to negative in a substantial proportion of the

<sup>1</sup>Systems Biology Program, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain. <sup>2</sup>Universitat Pompeu Fabra (UPF), Barcelona, Spain.

<sup>3</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. \*e-mail: ben.lehner@crg.eu

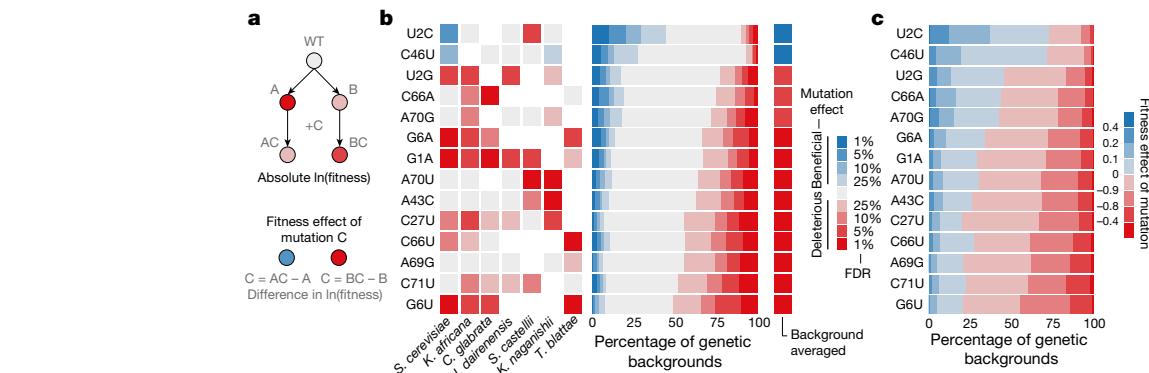


**Fig. 1 | Combinatorially complete fitness landscape of a tRNA.**

a, Species phylogenetic tree<sup>26</sup> and multiple sequence alignment of the tRNA-Arg(CCU) orthologues. Variable positions across the seven yeast species with the synthesized library are shown below: R (A or G);

b, Secondary structure of *S. cerevisiae* tRNA-Arg(CCU) (varied positions indicated in red). c, Selection experiment and structure of the replicates. From each independent yeast transformation (input) three independent

genetic backgrounds (Fig. 3a). Restricting our analyses to genetic backgrounds with high or intermediate fitness, to combinations with high expected fitness or to genotypes with high input read counts did not change this conclusion (Extended Data Fig. 5b). Across all genetic backgrounds, positive and negative interactions were similarly prevalent (11.4% and 10.3% for positive and negative epistasis respectively, FDR < 0.1,  $n = 47,649$ ).

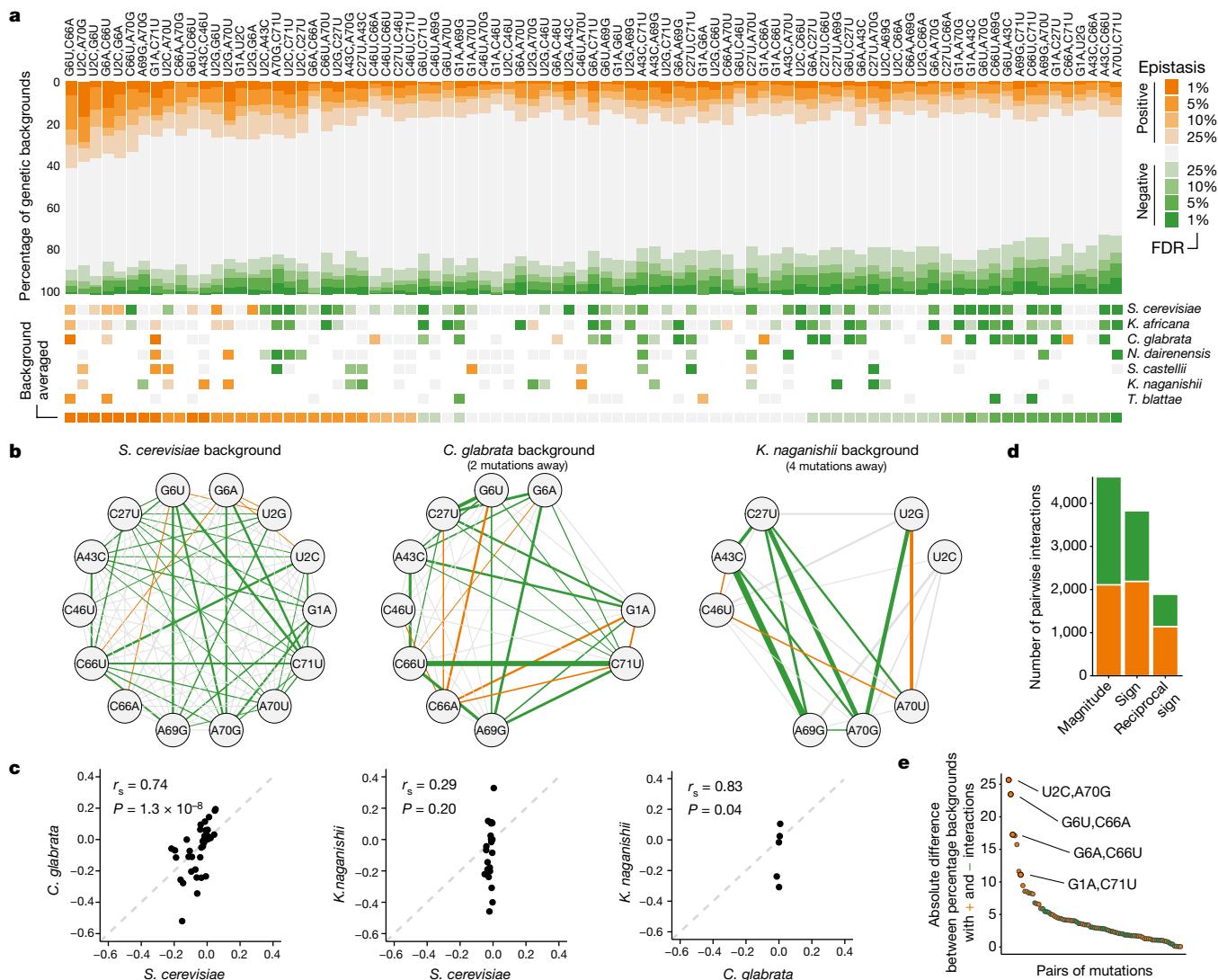


**Fig. 2 | All single mutations switch sign from detrimental to beneficial in different genetic backgrounds.** a, The same mutation can have different fitness consequences depending on the genetic background. b, Significance of beneficial (blue) or detrimental (red) mutation effects in

selection experiments were performed. d, Correlation between weighted-averaged input replicates ( $r_s$ , Spearman correlation coefficient;  $n = 4,176$  genotypes). e, Fitness landscape of the tRNA-Arg(CCU) genotypes (nodes). Colour indicates ln(fitness) relative to the *S. cerevisiae* tRNA. Edges connect genotypes differing by a single substitution. Genotypes and the distribution of fitness values (violins) are arranged on the x-axis according to the total number of substitutions from the *S. cerevisiae* tRNA. Highlighted nodes indicate the genotypes of the seven extant species.

Changes in base pairing only partially explained changes in the sign and magnitude of the effect of single mutations (Extended Data Fig. 6). The four pairs of mutations that restore Watson–Crick base pairs (WCBPs) were amongst the most robust positive interactions (Fig. 3e). However, even these combinations interacted negatively in a large fraction of backgrounds (5.9–8.4%). This is consistent with the presence of non-WCBP nucleotides in these positions in the tRNAs from other

the backgrounds of each species (left) and across all genetic backgrounds (right).  $n = 21,450$  backgrounds. c, Proportion of genetic backgrounds in which each mutation has beneficial (blue) or detrimental (red) effects.



**Fig. 3 | Genetic interactions between all pairs of mutations switch from positive to negative epistasis in different genetic backgrounds.** **a**, Proportion of backgrounds (top) and species (middle) in which each pair of mutations interacts positively (orange) or negatively (green) at different FDRs ( $n = 47,649$  backgrounds). Bottom, background-averaged epistasis ( $n = 87$  pairs of mutations). **b**, Interaction networks for three species (other species are shown in Extended Data Fig. 4b). Edge colours indicate epistasis sign ( $\text{FDR} < 0.1$ ) and widths indicate the strength of the

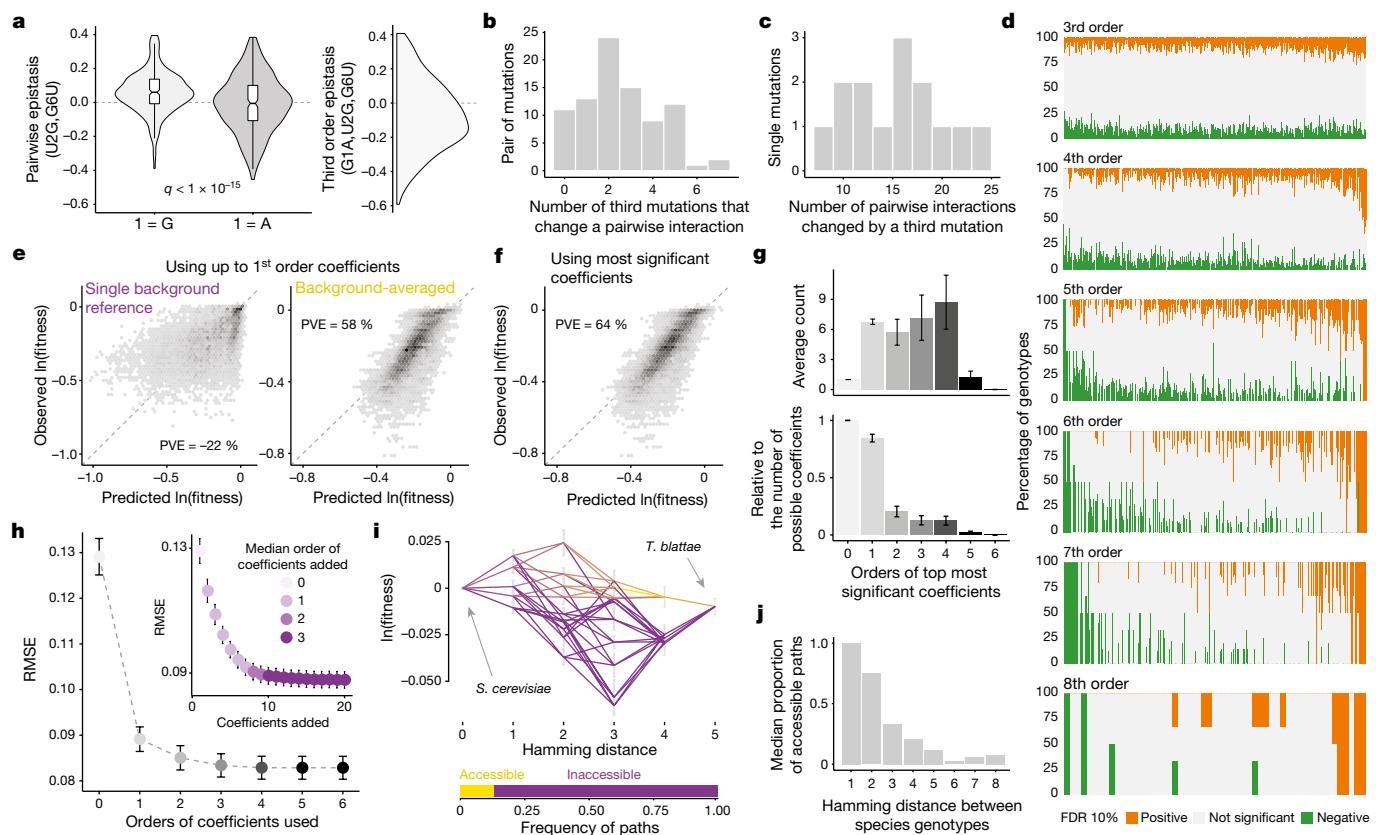
species<sup>27</sup> (Extended Data Fig. 5c). Double mutants in the same RNA strand of the acceptor stem were enriched for negative epistasis (odds ratio (OR) = 1.23, Fisher's exact test  $P = 2.15 \times 10^{-6}$ , Extended Data Fig. 5d, e) and the restoration of a WCBP was also more likely to result in a negative interaction when the stem harboured multiple additional mutations in a single strand (Extended Data Fig. 5f). This suggests that other mechanisms, for example stacking interactions, are also important determinants of tRNA function.

We next tested whether pairwise interactions changed in backgrounds containing each additional single mutation (Fig. 4a, Extended Data Fig. 7a). Notably, when averaging across genetic backgrounds, a total of 138 out of 316 possible third-order interactions were found (Extended Data Fig. 7b,  $\text{FDR} < 0.1$ ), meaning that 76 out of 87 pairwise interactions were altered by the presence of a single additional mutation in the background (Fig. 4b). All 14 individual mutations altered at least eight pairwise interactions (median = 16.5, maximum = 24, Fig. 4c). As with second-order interactions, third-order interactions were enriched amongst proximal mutations and mutations found on the same strand (Extended Data Fig. 7c, d).

interaction. **c**, Comparison of epistasis scores between these three species ( $n = 43, 22$  and  $6$  comparisons from left to right, respectively). **d**, Number of positive (orange) or negative (green) magnitude, sign or reciprocal sign pairwise epistasis ( $n = 10,330$  significant interactions from 47,649 tested). **e**, Consistency of each interaction quantified as the absolute difference between the percentage of backgrounds in which the interaction is positive or negative at  $\text{FDR} < 0.1$ . Colour indicates the predominant sign. The four pairs that restore WCBPs are highlighted.

However, as for pairwise interactions, all third-order interactions (316 out of 316) also switched from positive to negative across different genetic backgrounds, indicating the presence of even higher-order epistasis (Fig. 4d). 260 out of 316 third-order interactions changed in the presence of a fourth mutation ( $\text{FDR} < 0.1, n = 740$ ). Indeed, interactions can be detected in this dataset up to the eighth order (Extended Data Fig. 7b, a total of 763 background-averaged epistatic interactions from 3,961 possible interactions tested from order one to eight,  $\text{FDR} < 0.1$ ). Consistent with the behaviour of the lower-order interactions, the signs of many higher-order interactions also switch from positive to negative as the genetic background changes (Fig. 4d, 1,981 out of 3,691 interactions in the total dataset interact both positively and negatively in different genetic backgrounds at  $\text{FDR} < 0.1$ ).

Finally, we evaluated the extent to which epistasis affected our ability to predict phenotypes from genotypes. We quantified the accuracy of genetic prediction in the 76 complete di-allelic sub-landscapes of eight mutations using models restricted to a single genetic background as a reference or models that averaged epistatic terms across multiple backgrounds (see Methods section 'Genetic prediction'). Although



**Fig. 4 | Averaging coefficients across genetic backgrounds and using higher order epistatic terms is important for genetic prediction.**

**a**, Changes in the distribution of pairwise epistasis when the genetic backgrounds contain or do not contain the indicated mutation (left) and the distribution of the corresponding third-order epistasis values (right). **b**, Distribution of pairwise interactions that are altered by a third mutation. **c**, Distribution of single mutations that are involved in a third order interaction. **d**, Proportion of genetic backgrounds in which each combination of mutations from third to eighth order interact positively (orange) or negatively (green) at a FDR < 0.1. **e**, Agreement between observed and predicted fitness values of all eighth order complete sub-landscapes ( $n = 19,456$  genotypes, 76 sub-landscape with 256 genotypes each) when using up to first order epistatic coefficients, relative to a single background genotype (left) or averaged across backgrounds (right, tenfold cross-validation). **f**, Agreement between observed and predicted

individual mutation effects quantified in a single genetic background provide very poor prediction (Fig. 4e), the average effect of each mutation across all genotypes within a sub-landscape improves the prediction (Fig. 4e, percentage of variance explained, PVE = 58% on held-out data, tenfold cross-validation). Including a limited number of significant interaction terms further improves the prediction (Fig. 4f, Extended Data Fig. 8a, PVE = 64%). The best models evaluated by cross-validation contain first and second order coefficients, but also higher-order interactions (Fig. 4g) that progressively reduce the prediction error (Fig. 4h). However, these models contain a relatively small number of coefficients (20 out of 256 coefficients on average across sub-landscapes, Extended Data Fig. 8b), suggesting that although pairwise and higher-order epistasis is important, reasonably sparse models can provide good genetic predictions when coefficients are measured across different genetic backgrounds.

Taken together, our results show that even single steps in sequence space substantially change the effects of both individual mutations and how these mutations combine to alter fitness. By a range of metrics, the combinatorially complete tRNA fitness sub-landscapes are most similar to rugged theoretical fitness landscapes<sup>28</sup> that constrain evolution (Extended Data Fig. 9). Indeed, the abundance of sign epistasis (Fig. 3d) limits the number of accessible evolutionary paths<sup>29</sup>, for example, paths

fitness values for all complete eighth order sub-landscapes for the best models incorporating epistatic coefficients according to the rank of their significance and evaluated by cross-validation (an average of 20 out of 256 epistatic coefficients per model). **g**, Mean orders of the most significant epistatic coefficients for the models used in **f** (bottom, relative to the possible number of coefficients per order; top, absolute counts). Error bars are 95% confidence intervals. **h**, Mean root-mean-square error (RMSE) across the 76 eighth order sub-landscapes when cumulatively adding the most significant coefficients determined by cross-validation (inset, colour indicates the median order of the coefficient added across the 76 sub-landscapes) or all significant coefficients from the same order (main). Error bars are 95% confidence intervals. **i**, Example of shortest paths between two extant species (top) and the accessible proportion (bottom). **j**, Average frequency of accessible paths between species.

between the genotypes of extant species (Fig. 4i, j, Extended Data Fig. 10). These results add to a growing body of evidence<sup>2</sup> that evolution is highly contingent at the molecular level. As a consequence, models that use coefficients averaged across different genetic backgrounds and that incorporate higher-order epistatic terms provide more accurate genetic prediction.

### Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0170-7>.

Received: 29 May 2017; Accepted: 9 April 2018;  
Published online 30 May 2018.

- Lehner, B. Genotype to phenotype: lessons from model organisms for human genetics. *Nat. Rev. Genet.* **14**, 168–178 (2013).
- de Visser, J. A. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* **15**, 480–490 (2014).
- Starr, T. N. & Thornton, J. W. Epistasis in protein evolution. *Protein Sci.* **25**, 1204–1218 (2016).
- Fowler, D. M. et al. High-resolution mapping of protein sequence–function relationships. *Nat. Methods* **7**, 741–746 (2010).

5. Araya, C. L. et al. A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proc. Natl Acad. Sci. USA* **109**, 16858–16863 (2012).
6. Diss, G., Lehner, B. The genetic landscape of a physical interaction. *eLife* **7**, e32472 (2018).
7. Melamed, D., Young, D. L., Gamble, C. E., Miller, C. R. & Fields, S. Deep mutational scanning of an RRM domain of the *Saccharomyces cerevisiae* poly(A)-binding protein. *RNA* **19**, 1537–1551 (2013).
8. Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2**, e00631 (2013).
9. Olson, C. A., Wu, N. C. & Sun, R. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr. Biol.* **24**, 2643–2651 (2014).
10. Gong, L. I. & Bloom, J. D. Epistatically interacting substitutions are enriched during adaptive protein evolution. *PLoS Genet.* **10**, e1004328 (2014).
11. Bank, C., Hietpas, R. T., Jensen, J. D. & Bolon, D. N. A systematic survey of an intragenic epistatic landscape. *Mol. Biol. Evol.* **32**, 229–238 (2015).
12. Hayden, E. J., Bendixsen, D. P. & Wagner, A. Intramolecular phenotypic capacitance in a modular RNA molecule. *Proc. Natl Acad. Sci. USA* **112**, 12444–12449 (2015).
13. Bank, C., Matuszewski, S., Hietpas, R. T. & Jensen, J. D. On the (un)predictability of a large intragenic fitness landscape. *Proc. Natl Acad. Sci. USA* **113**, 14085–14090 (2016).
14. Puchta, O. et al. Network of epistatic interactions within a yeast snoRNA. *Science* **352**, 840–844 (2016).
15. Li, C., Qian, W., Maclean, C. J. & Zhang, J. The fitness landscape of a tRNA gene. *Science* **352**, 837–840 (2016).
16. Julien, P., Miñana, B., Baéza-Centurion, P., Valcárcel, J. & Lehner, B. The complete local genotype–phenotype landscape for the alternative splicing of a human exon. *Nat. Commun.* **7**, 11558 (2016).
17. Sarkisyan, K. S. et al. Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397–401 (2016).
18. Guy, M. P. et al. Identification of the determinants of tRNA function and susceptibility to rapid tRNA decay by high-throughput in vivo analysis. *Genes Dev.* **28**, 1721–1732 (2014).
19. Costanzo, M. et al. A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, <https://doi.org/10.1126/science.aaf1420> (2016).
20. Forsberg, S. K., Bloom, J. S., Sadhu, M. J., Kruglyak, L. & Carlberg, Ö. Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. *Nat. Genet.* **49**, 497–503 (2017).
21. Tischler, J., Lehner, B. & Fraser, A. G. Evolutionary plasticity of genetic interaction networks. *Nat. Genet.* **40**, 390–391 (2008).
22. Weinreich, D. M., Lan, Y., Wylie, C. S. & Heckendorf, R. B. Should evolutionary geneticists worry about higher-order epistasis? *Curr. Opin. Genet. Dev.* **23**, 700–707 (2013).
23. Palmer, A. C. et al. Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. *Nat. Commun.* **6**, 7385 (2015).
24. Sailer, Z. R. & Harms, M. J. Detecting high-order epistasis in nonlinear genotype–phenotype maps. *Genetics* **205**, 1079–1088 (2017).
25. Wu, N. C., Dai, L., Olson, C. A., Lloyd-Smith, J. O. & Sun, R. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* **5**, 16965 (2016).
26. Marçet-Houben, M. & Gabaldón, T. Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biol.* **13**, e1002220 (2015).
27. Hopf, T. A. et al. Mutation effects predicted from sequence co-variation. *Nat. Biotechnol.* **35**, 128–135 (2017).
28. Ferretti, L. et al. Measuring epistasis in fitness landscapes: The correlation of fitness effects of mutations. *J. Theor. Biol.* **396**, 132–143 (2016).
29. Weinreich, D. M., Watson, R. A. & Chao, L. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165–1174 (2005).

**Acknowledgements** We thank J. Schmiedel for statistical guidance. This work was supported by a European Research Council Consolidator grant (616434), the Spanish Ministry of Economy and Competitiveness (BFU2011-26206 and SEV-2012-0208), the AXA Research Fund, the Bettencourt Schueller Foundation, Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR), the EMBL-CRG Systems Biology Program, and the CERCA Program/Generalitat de Catalunya. Deep sequencing was performed in the EMBL Heidelberg GeneCore Genomics Core Facility.

**Reviewer information** Nature thanks Z. Blount, D. Marks, A. Wagner and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Author contributions** J.D. performed all experiments and analyses. J.D., G.D., and B.L. designed the experiments and analyses. B.L. and J.D. wrote the manuscript.

**Competing interests** The authors declare no competing interests.

#### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-018-0170-7>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-018-0170-7>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to B.L.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

**Library design.** tRNAs orthologous to *S. cerevisiae* tRNA-Arg(CCU) (encoded by *HSX1*) were collected from the Genomic tRNA Database<sup>30</sup> or extracted from the genome of each species using BLAST<sup>31</sup> (blastall 2.2.25). The sequences were aligned with Clustal Omega<sup>32</sup>. Across the 12 species closest to *S. cerevisiae*, only the six species shown in Fig. 1a had substitutions in the gene, with a total of 14 substitutions in ten positions. Allowing all of these substitutions to co-occur results in a total library size of 5,184 ( $2^6 \times 3^4$ ) possible mutation combinations.

**Plasmid library construction.** An oligonucleotide of 115 nucleotides containing 72 nucleotides of tRNA flanked by 21 and 22 nucleotides of the yeast endogenous promoter and terminator was synthesized by IBA Lifesciences. At ten of the 72 positions of the tRNA, two or three different nucleotides were mixed in equal proportions during synthesis. For example, position one can be G or A, but position two can be T, G or C.

The oligonucleotide was amplified using PCR for ten cycles (Q5 Hot Start High-Fidelity DNA Polymerase, NEB), purified using an E-gel electrophoresis system (E-Gel SizeSelect Agarose Gel 2%) with column purification (MinElute PCR Purification Kit, Qiagen). Subsequently, the purified oligonucleotide was cloned into a version of the yeast centromeric plasmid pRS413 (HIS3 marker)<sup>33</sup> that contained the *HSX1* gene flanked by 218 bp of upstream and 202 bp of downstream genomic sequences (pJD001). pJD001 was linearized from the *HSX1* flanking regions (excluding the *HSX1* sequence) using PCR (Q5 Hot Start High-Fidelity DNA Polymerase, NEB) and then purified using gel extraction (QIAquick Gel Extraction Kit, Qiagen). The library of oligonucleotides was cloned into 400 µg of linearized pJD001 substituting the wild-type *HSX1* gene using a Gibson reaction (prepared in house) at 50 °C for 12 h with a ratio 5:1 of insert:vector. After dialysing the reaction with 0.025 µm VSWP membrane filters (Merck Millipore) for 1.5 h, the product was concentrated 4× using speed-vac. Six microlitres of the concentrated reaction was transformed into 100 µl of electrocompetent *Escherichia coli* (NEB 10-beta Electrocompetent *E. coli*, NEB) according to the manufacturer's protocol. Cells were allowed to recover in SOC (NEB 10-beta/Stable Outgrowth Medium) for 30 min and later transferred to 150 ml of LB medium with ampicillin 4× overnight. The total number of transformants was estimated to be  $\sim 9.59 \times 10^6$ . Given the complexity of the library, each variant was therefore represented  $\sim 1,849$  times on average. 50 ml of *E. coli* saturated culture was harvested to extract the plasmid library using plasmid midi prep (QIAfilter Plasmid Midi Kit, Qiagen).

**Selection experiment.** Yeast strain and conditional growth defect in different environmental conditions. The *HSX1* deletion strain was obtained by replacing the *HSX1* gene with a nourseothricin resistance cassette in the haploid laboratory strain BY4742 (MAT $\alpha$  his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 *HSX1*::natMX4) and later confirmed using colony PCR. The deletion of the single copy tRNA-Arg(CCU) (*HSX1*) in yeast was previously reported to lead to a conditional growth defect when the temperature is raised from 30 °C to 37 °C<sup>15</sup>. We found that a similar growth defect is observed if the growth medium contains high salt concentrations (1 M NaCl), and that a combination of high temperature and high salt gives an even stronger defect (Extended Data Fig. 1a). Synthetic complete medium lacking histidine (SC-HIS) 1 M NaCl at 37 °C was therefore used as the selective condition for the library selection experiment.

**Large-scale yeast transformation.** The high-efficiency yeast-transformation protocol was derived from a previously described method<sup>7</sup>. Two pre-cultures of the tRNA deletion strain were grown independently in 25 ml standard YPDA at 30 °C overnight. The next morning, the cultures were diluted into 175 ml of fresh YPDA to OD<sub>600 nm</sub> = 0.3. The two cultures were incubated at 30 °C for 4 h ( $\sim 2\text{--}3$  generations). After the growth period, the cells were harvested and centrifuged for 5 min at 3,000g, washed in sterile water and later in SORB (100 mM LiOAc, 10 mM Tris pH 8.0, 1 mM EDTA, 1 M sorbitol). The cells were re-suspended in 8.6 ml of SORB and incubated at room temperature for 30 min. After incubation, 175 µl of 10 mg ml<sup>-1</sup> boiled salmon sperm DNA (Agilent Genomics) was added to each tube of cells, as well as 3.5 µg of plasmid library. After 10 min of gentle shaking at room temperature, 35 ml of Plate Mixture (100 mM LiOAc, 10 mM Tris-HCl pH 8, 1 mM EDTA/NaOH, pH 8, 40% PEG3350) was added to the cells and incubated at room temperature for 30 more min. 3.5 ml of DMSO was added to each tube and the cells were then heat shocked at 42 °C for 20 min (inverting tubes from time to time to ensure homogenous heat transfer). After heat shock, each independent tube of cells was centrifuged and re-suspended in 350 ml of YPD + 0.5M Sorbitol and allowed to recover for 1 h at 30 °C. The cells were then centrifuged, washed twice with SC-HIS medium and re-suspended in 350 ml SC-HIS. The two independent transformations were grown at 30 °C for  $\sim 60$  h until saturation. For the two independent transformations,  $1.5 \times 10^6$  and  $1.1 \times 10^6$  transformants were obtained, which ensured that each variant of the library was on average represented  $\sim 250$  times<sup>34</sup>.

**Competition assay.** The competition experiment had two different phases. In phase one, the environment had minimal selection on the tRNA functionality (SC-HIS at 30 °C), allowing the pool of variants to be amplified and the cells to enter the

exponential growth phase (input library)<sup>34</sup>. In the second stage, the medium was changed to a condition (SC-HIS 1 M NaCl medium at 37 °C) in which non-functional tRNA variants would lead to a severe growth defect phenotype (output library). The assay was performed immediately after yeast transformation to avoid recovering cells from frozen glycerol stocks. Once the two independently transformed cultures reached saturation ( $\sim 60$  h after plasmid transformation), they were inoculated at an OD<sub>600 nm</sub> of 0.08 in 500 ml of SC-HIS medium and grown for four generations at 30 °C ( $\sim 11$  h). When exponential phase was reached after four generations of growth, the cells were harvested and washed with selection medium (warm SC-HIS NaCl 1 M) and then inoculated in 500 ml of selection medium at an OD<sub>600 nm</sub> of 0.015. The remainder of the cells was harvested and stored at  $-20^{\circ}\text{C}$  for later DNA extraction of the input libraries. Each independent input library was divided into three different output libraries (six replicates in total). Cells were grown in selective conditions for  $\sim 6.5$  generations ( $\sim 26.5$  h). This number of generations was chosen so that the average read coverage in the input would be of  $\sim 150$  reads per variant and that null alleles, which grow  $\sim 0.18$  generations every 3 h, would be detected in the output after sequencing. After 6.5 generations, the cells were harvested and the cell pellets stored at  $-20^{\circ}\text{C}$  for later DNA extraction of the output libraries.

**DNA extraction and quantification.** Cell pellets (eight tubes, two inputs and six outputs) were re-suspended in 1.5 ml extraction buffer (2% Triton-X, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl pH 8, 1 mM EDTA pH 8), frozen using an dry ice–ethanol bath and incubated at 62 °C in a water bath twice. Subsequently, 1.5 ml of phenol-chloro isoamyl alcohol (25:24:1 ratio, equilibrated in 10 mM Tris-HCl, 1 mM EDTA, pH 8) was added, together with 1.5 g of glass beads and the samples were vortexed for 10 min. Samples were centrifuged at room temperature for 30 min at 3,200g and the aqueous phase was transferred into new tubes. The same step was repeated twice. 0.15 ml of NaOAc 3 M and 3.3 ml of cold ethanol 100% were added to the aqueous phase. The mix was incubated at  $-20^{\circ}\text{C}$  for 30 min and then centrifuged for 30 min at full speed at 4 °C to precipitate the DNA. The ethanol was removed and the DNA pellet allowed to dry overnight at room temperature. DNA pellets were re-suspended in 900 µl TE 1× and treated with RNaseA (10 mg ml<sup>-1</sup>, Thermo Scientific) for 30 min at 37 °C. To desalt and concentrate the DNA solutions, a QIAEX II Gel Extraction Kit was used (75 µl of QIAEX II beads suspension). The samples were washed three times with PE buffer and eluted twice in 375 µl of 10 mM Tris-Cl buffer, pH 8.5.

**Sequencing library preparation.** The plasmid concentration in each total DNA sample was quantified in triplicate by real-time quantitative PCR, using primers that had homology to the origin of replication region of the pJD001 plasmid backbone (Supplementary Table 3). On average, we obtained  $\sim 3.5 \times 10^6$  plasmid molecules per µl of DNA sample.

A two-step PCR using high fidelity Q5 Hot Start High-Fidelity DNA Polymerase (NEB) was used to amplify the input and output libraries for sequencing. For each sample,  $\sim 150$  million plasmid molecules were amplified for ten cycles using primers with overhang homology to Illumina sequencing adapters (Supplementary Table 3). The samples were then treated with ExoSAP (Affymetrix) and cleaned using bead purification with a QIAEX II kit (10 µl pf QIAEX II beads suspension). The whole eluates, corresponding to the entire first PCR reactions, were used for the second PCR reactions (15 cycles), in which the rest of the Illumina adaptor was added as overhangs on the primers, in addition to sample-specific indexes. The DNA concentration of each individual second PCR was quantified by fluorometric quantitation (Quant-iT PicoGreen dsDNA Assay Kit) and pooled together at an equimolar ratio. Finally, the pooled sequencing library was gel purified (QIAEX II Gel Extraction Kit) and subjected to 125 bp paired-end sequencing on an Illumina HiSeq 2500v5 sequencer at the EMBL Genomics Core Facility (Heidelberg, Germany).

**From sequencing reads to fitness values.** The sequencing reads of each sample (two inputs and six outputs) were processed and filtered independently. Each sequencing read covered the entire tRNA. The 5' and 3' constant regions of the read (primers annealing sites) were removed with the 'cutadapt' software<sup>35</sup>. The forward and reverse reads were merged using PEAR<sup>36</sup> and sequences that were either not assembled owing to low quality or unexpected length were discarded. Unique genotypes were called and quantified with custom Python scripts. Genotypes with less than nine input reads in any input replicate, unexpected nucleotide substitutions (sequencing or PCR errors) or zero reads in the outputs were discarded. After filtering, we ended up with a total of 4,176 sequence genotypes quantified in all inputs and outputs.

To obtain accurate fitness and error estimates for each variant we took into account the hierarchical structure of the replicates<sup>37</sup> as well as sampling error owing to the low number of read counts<sup>38</sup>. Input and output frequencies for each genotype from each of the independent competition experiments were first calculated and then these were combined into a single output measurement for each input replicate. The number of cells expressing each genotype in each input ( $f_{\text{in},gi}$ ) and output replicate ( $f_{\text{out},gi}$ ) was calculated using the following formulae:

$$f_{\text{in}_{gi}} = \text{OD}_{\text{in}_{ij}} \times \frac{\text{counts}_{\text{in}_{gi}}}{\sum_{g=1}^l \text{counts}_{\text{in}_{gi}}}$$

$$f_{\text{out}_{gi}} = \text{OD}_{\text{out}_{ij}} \times \frac{\text{counts}_{\text{out}_{gi}}}{\sum_{g=1}^l \text{counts}_{\text{out}_{gi}}}$$

in which  $g$  is the genotype (from 1 to  $l$ , with  $l$  being the total number of genotypes after filtering),  $i$  is the number of input replicates (1 or 2) and  $j$  is the number of output replicates per input replicate (1 to 3).

These formulae assume that each read derives from an individual cell, so that by multiplying the frequency of reads in the output with the final ( $\text{OD}_{\text{out}}$ ) and initial culture density ( $\text{OD}_{\text{in}}$ ) we can estimate the number of cells for a particular genotype at the beginning ( $f_{\text{in}}$ ) and end ( $f_{\text{out}}$ ) of the competition experiment.

Each input and output frequency is associated to a Poisson variance given the number of read counts of each genotype and the total read count<sup>38</sup>:

$$\sigma_{\text{in}_{gi}} = \sqrt{\frac{1}{\text{counts}_{\text{in}_{gi}}} + \frac{1}{\sum_{g=1}^l \text{counts}_{\text{in}_{gi}}}}$$

$$\sigma_{\text{out}_{gi}} = \sqrt{\frac{1}{\text{counts}_{\text{out}_{gi}}} + \frac{1}{\sum_{g=1}^l \text{counts}_{\text{out}_{gi}}}}$$

We calculated a single output frequency score for each input replicate using a weighted average in which the weight of each score ( $f_{\text{out}_{gi}}$ ) is the inverse of the variance of the genotype ( $\sigma_{\text{out}_{gi}}^2$ ):

$$f_{\text{out}_{gi}} = \frac{\sum_{j=1}^3 f_{\text{out}_{gij}} \times \frac{1}{\sigma_{\text{out}_{gij}}^2}}{\sum_{j=1}^3 \frac{1}{\sigma_{\text{out}_{gij}}^2}}$$

The output frequency errors of each replicate were then combined to yield an overall output frequency error:

$$\sigma_{\text{out}_{gi}} = \sqrt{\frac{1}{\sum_{j=1}^3 \sigma_{\text{out}_{gij}}^{-2}}}$$

The number of generations  $n_{gi}$  was then calculated as the log2 ratio of the normalized input and output frequencies:

$$n_{gi} = \log_2 \left( \frac{f_{\text{out}_{gi}}}{f_{\text{in}_{gi}}} \right)$$

with an associated error of:

$$\sigma_{n_{gi}} = \frac{1}{\ln(2)} \times \sqrt{\sigma_{\text{out}_{gi}}^2 + \sigma_{\text{in}_{gi}}^2}$$

The number of generations in each input replicate ( $n_{g1}$  and  $n_{g2}$ ) was combined using a weighted average as before to obtain a single growth measurement and an error for each genotype:

$$n_g = \frac{\sum_{i=1}^2 f_{\text{out}_{gij}} \times \frac{1}{\sigma_{n_{gj}}^2}}{\sum_{i=1}^2 \frac{1}{\sigma_{n_{gj}}^2}}$$

$$\sigma_{n_g} = \sqrt{\frac{1}{\sum_{i=1}^2 \sigma_{n_{gj}}^{-2}}}$$

Finally, fitness values (in log-scale) relative to the *S. cerevisiae* wild type and the propagated error were calculated as follows:

$$\omega_g = \ln \left( \frac{n_g}{n_{\text{wt}}} \right)$$

$$\sigma_{\omega_g} = \sqrt{\left( \frac{\sigma_{n_g}}{n_g} \right)^2 + \left( \frac{\sigma_{n_{\text{wt}}}}{n_{\text{wt}}} \right)^2}$$

In log-space, if a particular genotype grew faster or slower than the wild type, the  $\ln(\text{fitness})$  value would be  $>0$  or  $<0$ , respectively.

**Single mutation effects, pairwise genetic interactions and higher order epistasis.** On a log-scale, the fitness effect of a mutation A on a genetic background X was calculated as the relative fitness gain of the variant AX respect to X:

$$\varepsilon_{A|X}^1 = \omega_{AX} - \omega_X$$

This fitness effect of a mutation can also be referred to as the first order epistatic term ( $\varepsilon^1$ )<sup>39</sup>.

A pairwise epistatic interaction between two mutations was defined as the difference between the observed fitness of the double mutant AB and the expected fitness obtained by the addition of the two single mutant fitness values (A and B). The fitness effects of the mutations A, B and AB can be calculated on each genetic background X by subtracting the fitness of X itself from the fitness of AX, BX and ABX, as described above. Pairwise epistasis (or second-order epistasis  $\varepsilon^2$ ) is then the change in the effect of each single mutation in the presence of the second mutation:

$$\varepsilon_{AB|X}^2 = (\omega_{ABX} - \omega_X) - ((\omega_{AX} - \omega_X) + (\omega_{BX} - \omega_X))$$

$$= \omega_{ABX} - \omega_{AX} - \omega_{BX} + \omega_X$$

$$= \varepsilon_{A|BX}^1 - \varepsilon_{A|X}^1 = \varepsilon_{B|AX}^1 - \varepsilon_{B|X}^1$$

This same analysis can be expanded to higher order terms<sup>22,39</sup>. For example, a third-order interaction ( $\varepsilon^3$ ) is the degree to which second-order epistasis is different when a third mutation is present in the background:

$$\varepsilon_{ABC|X}^3 = \varepsilon_{AB|CX}^2 - \varepsilon_{AB|X}^2 = \varepsilon_{AC|BX}^2 - \varepsilon_{AC|X}^2 = \varepsilon_{BC|AX}^2 - \varepsilon_{BC|X}^2$$

$$= \omega_{ABCX} - \omega_{ABX} - \omega_{ACX} - \omega_{BCX} + \omega_{AX} + \omega_{BX} + \omega_{CX} - \omega_X$$

Higher order terms follow the same principle, so we can calculate any  $n^{\text{th}}$ -order term using the formula<sup>39</sup>:

$$\begin{aligned} \varepsilon^n &= (-1)^0 \sum \omega^n + (-1)^1 \sum \omega^{n-1} \\ &\quad + (-1)^2 \sum \omega^{n-2} + \dots + (-1)^n \sum \omega^{n-n} \\ &= \sum_{i=0}^n \left( (-1)^{n-i} \sum \omega^i \right) \end{aligned}$$

in which  $\omega^n$  are all fitness terms of order  $n$  in a specific genetic background. It is important to note that an epistatic term of any order  $n$  can only be calculated if the genotype space is complete (that is, that the fitness of all genotypes from order 0 to  $n$  were quantified in the experiment). In our dataset, higher-order epistasis was quantified up to order eight (76 cases in this dataset), which was the highest order in which the fitness of a combinatorially-complete set of genotypes could be quantified after data filtering (Extended Data Fig. 1d).

To quantify how many epistatic terms were significantly positive or negative across all the backgrounds in which they were tested, a one-sample  $t$ -test was performed (using the epistatic term and its respective propagated error). The FDR was adjusted across all the tests performed (a total of 203,240 tests for all interactions of all orders across all backgrounds) using the Benjamini–Hochberg method<sup>40</sup>.

**Controlling for background fitness, sequence divergence and the number of input sequencing reads.** Across all the data, there was a weak correlation between the fitness of the genetic background and both the fitness effect of the single mutations and pairwise epistasis (Extended Data Figs. 2c, 5a). We therefore repeated all of the analyses on the subset of the genetic backgrounds with fitness close to the wild-type *S. cerevisiae* ( $-0.15 < \text{fitness} < 0.15$ ,  $n = 1,479$  library genotypes) and also on genetic backgrounds with moderate fitness decreases ( $-0.3 < \text{fitness} < -0.15$ ,  $n = 1,577$ ). We also repeated all of the analyses on the genetic backgrounds that were closest to the *S. cerevisiae* sequence (one to four mutations away,  $n = 1,040$ ) or excluding all variants with a mean input frequency of less than 100 reads ( $n = 1,315$ ). With each of these filters we excluded approximately two thirds of the original number of variants in the library.

**Classifying pairwise epistasis.** Significant pairwise interactions in the dataset ( $n = 10,330$  out of 47,649 tested) were classified into three categories: magnitude, sign, and reciprocal sign epistasis<sup>41</sup>. Pairwise epistasis was thus classified as follows. When the fitness effect of both single mutants differs in magnitude but not in sign in the presence of the other mutation, the epistatic interaction was classified as

magnitude epistasis. For sign epistasis, the sign of one of the individual fitness has effects on changes in the presence of a second mutation. Finally, if the sign of effect changes for both individual mutations, the interaction was classified as reciprocal sign epistasis. The way a single mutation effect changes in the presence of another mutation can be inferred if the fitness effect and sign of the single mutations (A and B) and the fitness of the double mutant (AB) are known. For instance, if the two single mutations A and B have significantly beneficial (positive) effects and the double mutant has higher fitness than both single mutants, then none of the single mutations are changing sign, so this interaction would be classified as magnitude. However, if the double mutant has a fitness value lower than both single mutations, then this interaction would be classified as reciprocal sign (both single mutations are changing sign in the presence of the other). Otherwise, this interaction will be classified as sign (fitness of the double is lower than only one of the singles).

The sign of each of the single mutants in the dataset ( $n = 21,450$ ) was assigned after performing a one-sample  $t$ -test (Benjamini–Hochberg FDR controlled across all tested interactions of all orders from one to eight,  $n = 203,240$  as described in the Methods section ‘Single mutation effects, pairwise genetic interactions and higher order epistasis’). Single mutants with  $q \geq 0.1$  were assigned as neutral (or not-significant) and the rest as positive (beneficial) or negative (deleterious) when the fitness effect of the mutation was more or less than 0 respectively.

Exceptional interactions between two mutations in which both single mutations had a neutral category (no significant fitness effect at  $FDR < 0.1$ ) were classified as magnitude epistasis (either positive or negative). When only one of the single mutations had a neutral category they were then classified as sign or magnitude epistasis depending on whether the other single mutation changed sign or not. Whenever both single mutations had either positive or negative categories, epistasis was classified as explained above.

**Background-averaged epistatic interactions.** We quantified the background-averaged epistatic interaction of a particular mutation combination (ranging from order one to eight) by averaging all epistatic coefficients of that same combination of mutations across all backgrounds in which it was found. To assess the significance of the average epistatic coefficient, the errors of all individual fitness terms were propagated and a one-sample  $t$ -test was performed. The  $P$  value was adjusted for all tests performed from order one to eight (a total of 3,691 tests) using the Benjamini–Hochberg FDR method<sup>40</sup>.

After identifying those mutations that interacted significantly when averaging across backgrounds (at  $FDR < 0.1$ ), we counted the number of times the interactions between two mutations changed owing to another mutation in the background, or calculated the number of times a single mutation was able to change a pairwise interaction (Fig. 4b, c).

**Genetic prediction.** As described in the section ‘Single mutation effects, pairwise genetic interactions and higher order epistasis’, epistatic terms were calculated as linear combinations of the fitness values of genotypes of different orders. This system of linear combination can be represented in a matrix form, which allows the epistatic coefficients to be calculated from fitness values, and fitness values back from epistasis<sup>39</sup>.

In a complete  $n$  loci di-allelic genotype space, in which each locus can harbour two different nucleotides, epistatic terms can be calculated as follows:

$$\bar{\varepsilon} = G\bar{\omega}$$

in which  $\bar{\omega}$  corresponds to a vector with the fitness values of the  $2^n$  genotypes from order 0 to  $n$ ,  $\bar{\varepsilon}$  is a vector with all the corresponding epistatic terms and  $G$  is a matrix that defines the linear mapping between  $\bar{\omega}$  and  $\bar{\varepsilon}$  for all orders.  $G$  can be recursively constructed as follows:

$$G_{n+1} = \begin{pmatrix} G_n & 0 \\ -G_n & G_n \end{pmatrix} \text{ with } G_0 = 1$$

In this case, epistatic terms are calculated relative to a single background (0<sup>th</sup> order genotype or ‘wt’). However, within a complete landscape, epistatic terms can be calculated across many different backgrounds. For instance, in a di-allelic landscape of three loci, the same single mutation effect (epistasis term of order one) can be measured four times from four different backgrounds. To obtain epistatic coefficients averaged amongst backgrounds we can use a similar version of the previous equation:

$$\bar{e} = VH\bar{\omega}$$

In this case, the  $\bar{e}$  vector corresponds to the background average epistatic coefficients.  $H$  (the Walsh–Hadamard transform<sup>22,39</sup>) defines the mapping from fitness to epistatic coefficients and can be recursively constructed as follows:

$$H_{n+1} = \begin{pmatrix} H_n & H_n \\ H_n & -H_n \end{pmatrix} \text{ with } H_0 = 1$$

The coefficient obtained by multiplying  $H$  by  $\bar{\omega}$  would correspond to the sum of the same coefficient across backgrounds, not the average. Moreover, coefficients of odd orders would have an opposite sign. The  $V$  matrix weights the coefficients by averaging and corrects the sign of odd orders depending on the order of each term.

$$V_{n+1} = \begin{pmatrix} \frac{1}{2}V_n & 0 \\ 0 & -V_n \end{pmatrix} \text{ with } V_0 = 1$$

Fitness values can be obtained by a linear combination of epistatic coefficients using the inverse mapping, for either relative or background-averaged epistatic coefficients:

$$\bar{\omega} = G^{-1}\bar{\varepsilon}$$

$$\bar{\omega} = (VH)^{-1}\bar{e}$$

For an overview and extended definitions, we refer the reader to the previously published description<sup>39</sup>.

**Cross-validation.** To detect model over-fitting, we used a tenfold cross-validation approach in which the background-averaged epistatic coefficients were quantified using 90% of the genotypes (training set) within each of the 76 eight-loci tRNA sub-landscapes with the remaining 10% used for evaluation (test set). With 10% of genotypes missing, computation of seventh or eighth order coefficients is no longer possible. Coefficients of other orders were averaged across backgrounds for which all intermediate genotypes were available. To assess the significance of each epistatic coefficient, the estimates of fitness errors were propagated accordingly and the  $t$ -statistic for a one sample  $t$ -test was calculated. Within each of the ten training sets for each complete sub-landscape, the coefficients were ranked by their absolute  $t$ -statistic and cumulatively used to predict fitness of the held-out test set genotypes (least significant coefficients were iteratively set to zero before predicting fitness values) using the inverse of the Walsh–Hadamard transform as described above (using a weighting matrix  $V$  in which the weights correspond to the number of backgrounds each coefficients had been averaged across). The best predictive model for each of the ten training sets of each sub-landscape was selected as the model that gave the lowest prediction error on the corresponding test set (Extended Data Fig. 8).

The accuracy of all the above predictions was quantified using root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{\text{SS}_{\text{res}}}{n}}$$

in which  $\text{SS}_{\text{res}}$  is the residual sum of squares and  $n$  is the total number of predicted genotypes. To calculate the percentage of variance explained (PVE) we used the formula:

$$\text{PVE} = \left(1 - \frac{\text{SS}_{\text{res}}}{\text{SS}_{\text{total}}}\right) \times 100$$

in which  $\text{SS}_{\text{total}}$  is the total sum of squares.

**Comparisons to theoretical fitness landscapes.** We used three different landscape statistics ( $\gamma$  statistic<sup>28</sup>, roughness-to-slope ratio<sup>42</sup> and the proportion of epistasis types<sup>42</sup>) to compare the tRNA fitness landscape to theoretical landscapes. To estimate the robustness of these measurements, all the statistics were calculated for all possible di-allelic (two possible nucleotide substitutions per position) complete tRNA sub-landscapes from three to eight loci that started from the wild-type *S. cerevisiae* genotype ( $n = 293, 568, 638, 403, 132, 18$  landscapes with three to eight loci respectively).

**Generation of theoretical landscapes.** We generated five different model landscapes using the software package MAGELLAN (<http://wwwabi.snv.jussieu.fr/public/magellan/Magellan.main.html>): the additive model (fitness effect of each mutation is independent of the genetic background), the House of Cards model (HOC, fitness values of different genotypes are independent and identically distributed random variables), the Rough Mount Fuji model (RMF has both additive and HOC components), the Kauffman NK model (in which each locus interacts with  $K$  other loci in the landscape) and the egg box model (maximally epistatic, anti-correlated fitness landscape, in which neighbouring fitness changes systematically from low to high, or vice versa, between genetic backgrounds one step apart). Further descriptions of the models can be found in previously published works<sup>2,13,28,42</sup>. We simulated 250 di-allelic landscapes of each theoretical model of size  $n$  ( $n = 3–8$ ) with an average fitness value and associated error similar to the tRNA landscape (average fitness effect of 0.04 and an associated standard error of 0.012). The RMF landscape

was modelled with a mix of 50% additive and 50% HoC and the  $K$  parameter of the NK model (each locus interacts with  $K$  loci) was set to  $K = n/2$ . These parameters were selected as they resulted in landscape statistics most similar to those of the tRNA sub-landscapes (data not shown).

**$\gamma$  statistic: correlation of fitness effects.** The  $\gamma$  statistic was recently introduced<sup>28</sup> and extended by others<sup>13</sup>.  $\gamma$  quantifies the correlation of fitness effects of the same mutation in single-mutant neighbours. It measures how the effect of a focal mutation is altered by another mutation at another locus in the background, averaged across the whole landscape. The statistic is bounded between  $-1$  and  $1$ . In a scenario without epistasis (the effect of a mutation is completely independent of the background),  $\gamma = 1$ . The  $\gamma$  measure gives information on the amount of epistasis in a combinatorially-complete landscape, but does not discriminate between different landscape topographies (two landscapes that differ in structure can have the same  $\gamma$  value). As with  $\gamma$ ,  $\gamma_d$  (the decay of correlation of fitness effects with mutational distance) can be defined as the correlation of fitness effects of mutations between genotypes that are  $1, 2, 3, \dots, d$  mutations away.  $\gamma_d$  gives extra information about the structure of the landscape, as it describes the cumulative epistatic effect of  $d$  mutations<sup>13, 28</sup>. In a completely additive landscape,  $\gamma_d$  is always  $1$  because the effect of a mutation is independent of the background genotype that is  $1, 2, 3$  or  $d$  mutations away. However, in a maximally rugged fitness landscape (in which the effect of a mutation depends entirely on its genetic background)  $\gamma_1$  is  $0$  and  $\gamma_d$  is  $0$  for all values of  $d$ . The behaviour of  $\gamma_d$  as a function of  $d$  varies for different theoretical landscape models<sup>13, 28</sup> (Extended Data Fig. 9a).

We calculated  $\gamma_d$  values for all possible complete di-allelic tRNA sub-landscapes of three to eight mutations combinations that contained the *S. cerevisiae* genotype using the software MAGELLAN (eight being the maximum number of loci in which a complete genotype space is available in the dataset). We later compared the statistic to the values for the theoretical landscapes. As a measure of similarity, we calculated the Euclidean distance between the  $\gamma_d$  of all tRNA sub-landscapes and the  $\gamma_d$  of the theoretical models (each tRNA landscape,  $n = 73,250, 142,000, 159,500, 100,750, 33,000$  and  $4,500$  for tRNA landscapes from three to eight mutations respectively, was compared to the 250 simulations of each theoretical landscape).

**Other quantitative measures of landscape ruggedness.** In addition to the  $\gamma$  statistic, for all complete tRNA and theoretical sub-landscapes from three to eight loci, we also calculated the roughness-to-slope ratio ( $r/s$  ratio) and characterized the local pairwise epistatic interactions. The  $r/s$  ratio measures how well the landscape can be described by a linear model, which corresponds to the purely additive limit<sup>42</sup>. The roughness is given by the variance of the residuals from the linear model and the slope is given by the average of the absolute values of the linear coefficients. The higher the  $r/s$ , the higher the deviation from the linear model and the more epistasis is present (in a non-epistatic scenario,  $r/s = 0$ ). To characterize the local interactions of each landscape we calculated the fraction of magnitude, sign or reciprocal sign pairwise epistasis within each landscape. We used the software MAGELLAN to calculate all the described statistics.

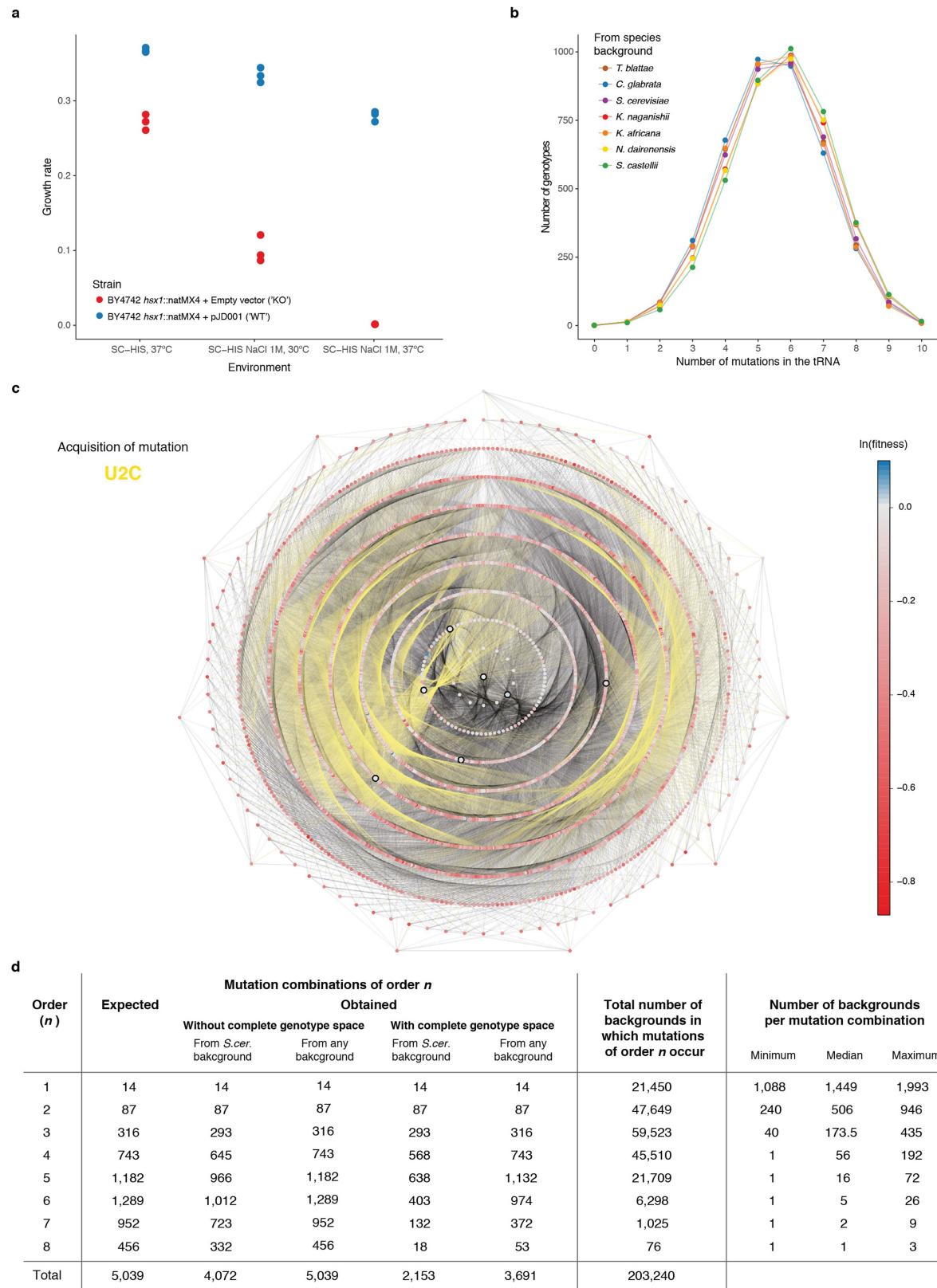
**Accessible paths between extant species.** An accessible path between two genotypes in the landscape was defined as a mutation trajectory in which none of the

intermediate genotypes has significantly lower fitness than both the initial and final genotypes that they connect ( $t$ -test between all the intermediate genotypes against the origin and end-point genotypes,  $n = 1–8$  tests). A path that had at least one deleterious intermediate genotype ( $P < 0.05$ ) was classified as inaccessible. We measured the number of accessible direct (shortest) paths between 20 pairwise comparisons of the extant genotypes in the landscape using the R package igraph. **Statistical analyses.** All statistical analyses were performed in R (v.3.3.3) and figures were made using the R package ggplot2. Lower and upper hinges of box plots correspond to the first and third quartiles (25<sup>th</sup> and 75<sup>th</sup> percentiles). The upper and lower whiskers extend from the hinge to the largest and lower value no further than  $1.5 \times \text{IQR}$  (inter-quartile range) respectively. Higher or lower points (outliers) are plotted individually (or not plotted in those cases where the box plot is plotted together with a violin plot). Notches give roughly 95% confidence interval for comparing the medians.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

**Data availability.** The complete dataset is available as Supplementary Table 1. Custom code used in this study is available from the authors upon request. Raw sequencing data has been submitted to GEO (accession number GSE99418).

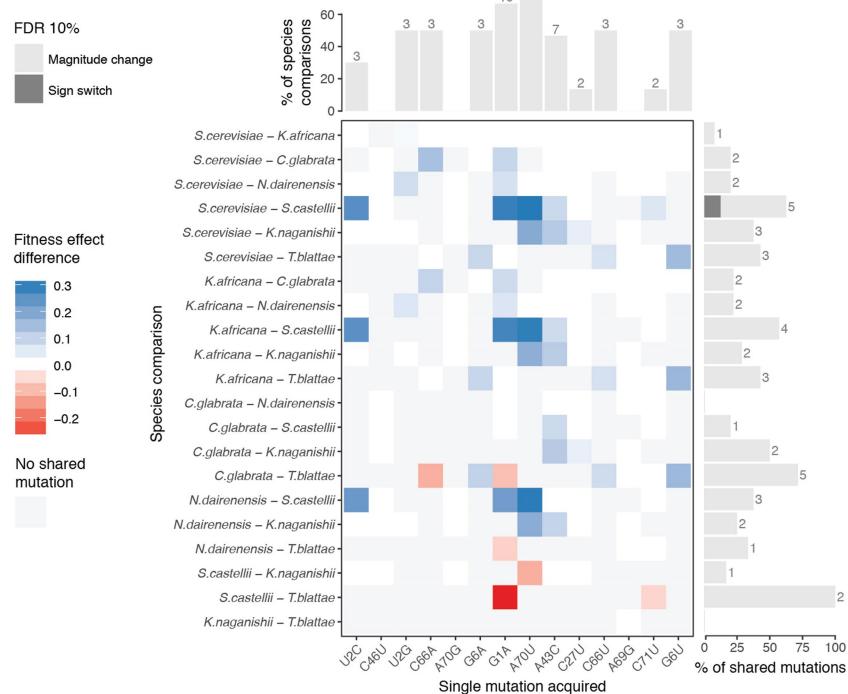
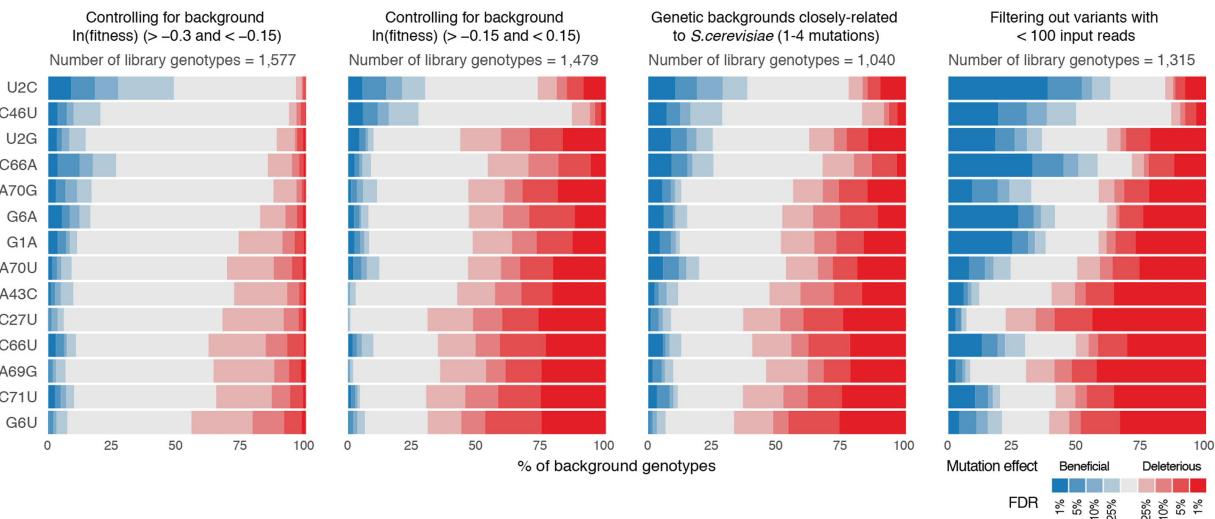
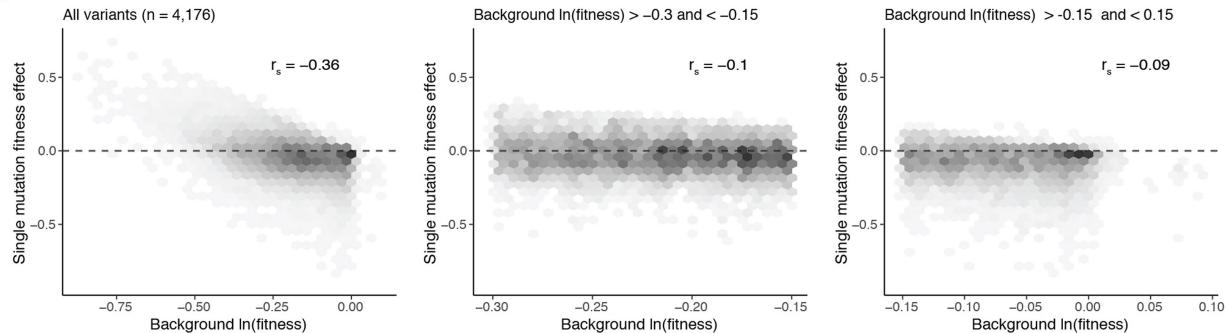
30. Chan, P. P. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* **37**, D93–D97 (2009).
31. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
32. McWilliam, H. et al. Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res.* **41**, W597–W600 (2013).
33. Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19–27 (1989).
34. Matuszewski, S., Hildebrandt, M. E., Ghenu, A. H., Jensen, J. D. & Bank, C. A statistical guide to the design of deep mutational scanning experiments. *Genetics* **204**, 77–87 (2016).
35. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12, (2011).
36. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd merger. *Bioinformatics* **30**, 614–620 (2014).
37. Crawley, M. J. *The R Book*. (Wiley, Chichester, 2007).
38. Rubin, A. F. et al. A statistical framework for analyzing deep mutational scanning data. *Genome Biol.* **18**, 150 (2017).
39. Poelwijk, F. J., Krishna, V. & Ranganathan, R. The context-dependence of mutations: a linkage of formalisms. *PLOS Comput. Biol.* **12**, e1004771 (2016).
40. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
41. Poelwijk, F. J., Kiviet, D. J., Weinreich, D. M. & Tans, S. J. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445**, 383–386 (2007).
42. Szendro, I. G., Schenck, M. F., Franke, J., Krug, J. & de Visser, J. A. Quantitative analyses of empirical fitness landscapes. *J. Stat. Mech.* **2013**, P01005 (2013).



Extended Data Fig. 1 | See next page for caption.

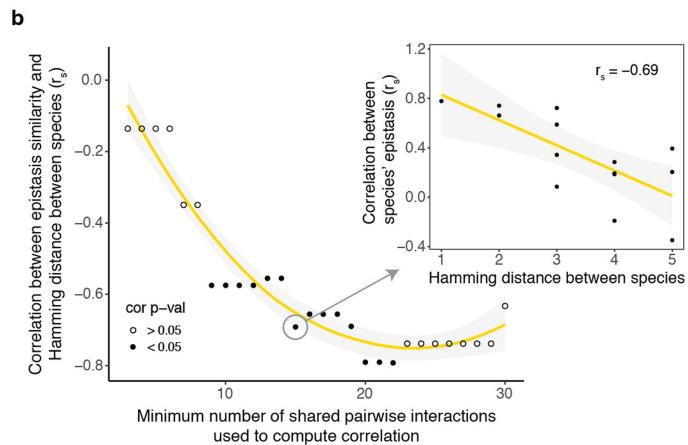
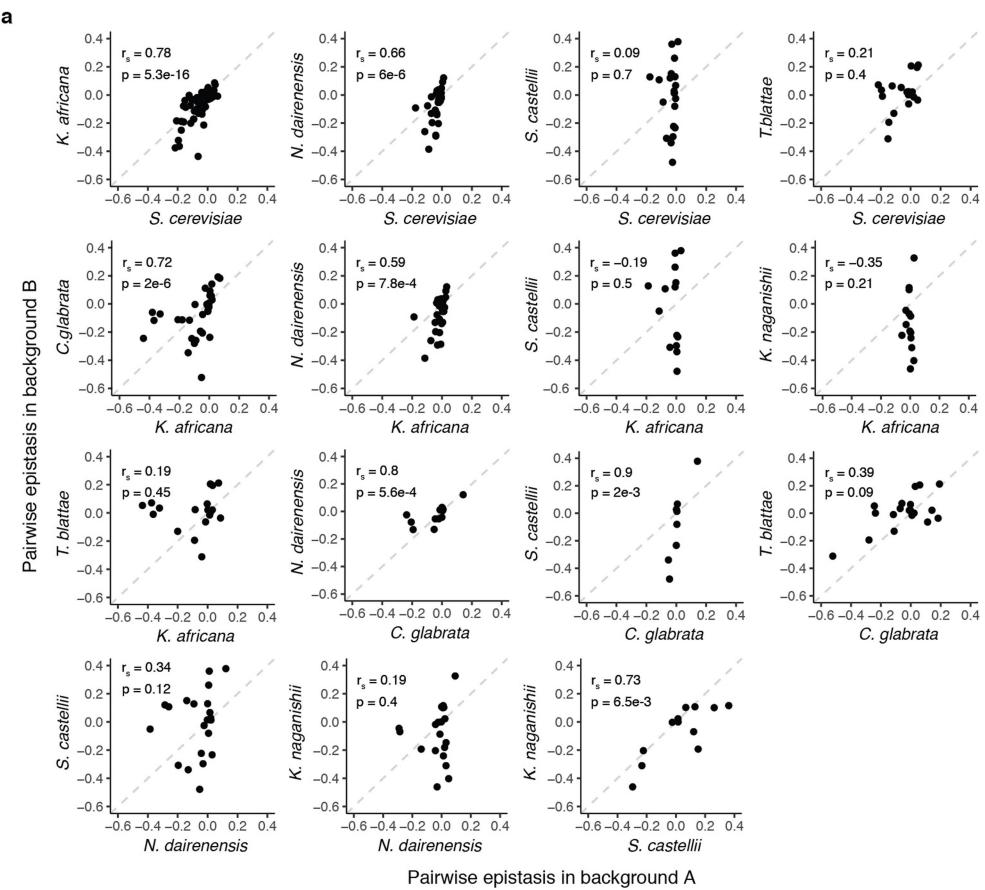
**Extended Data Fig. 1 | Experimental design.** **a**, Maximum growth rate (measured in a plate reader using spectrophotometry) of tRNA-Arg(CCU) (*HSX1*) deletion strain carrying either an empty plasmid (red) or a single-copy plasmid expressing wild-type tRNA-Arg(CCU) (blue) at high temperature, high salt, and high temperature with high salt ( $n = 3$  independent colonies from the plasmid transformation). **b**, Distribution of number of mutations per genotype in the library relative to the sequence of the tRNA from each species. **c**, Genotype network of the 4,176 tRNA-Arg(CCU) variants. Each node is one genotype. Colour indicates the  $\ln(\text{fitness})$  relative to *S. cerevisiae*. Edges connect genotypes differing by a single substitution, acquisition of a U2C mutation is highlighted in yellow as example. Genotypes are arranged in concentric

circles according to the total number of substitutions (one to ten) from the *S. cerevisiae* tRNA, which is the central node. Highlighted nodes indicate the genotypes of the seven extant species. **d**, Table showing the possible number of mutation combinations from order one to eight, with or without a complete genotype space (whether all intermediate genotypes are measured in the library or not) when using *S. cerevisiae* as a reference or any other background (the effect of a given combination of mutations can be measured from at least one genetic background). The total number of unique backgrounds is also indicated, together with the minimum, median and maximum number of backgrounds in which these mutations can be found.

**a****b****c****Extended Data Fig. 2 |** See next page for caption.

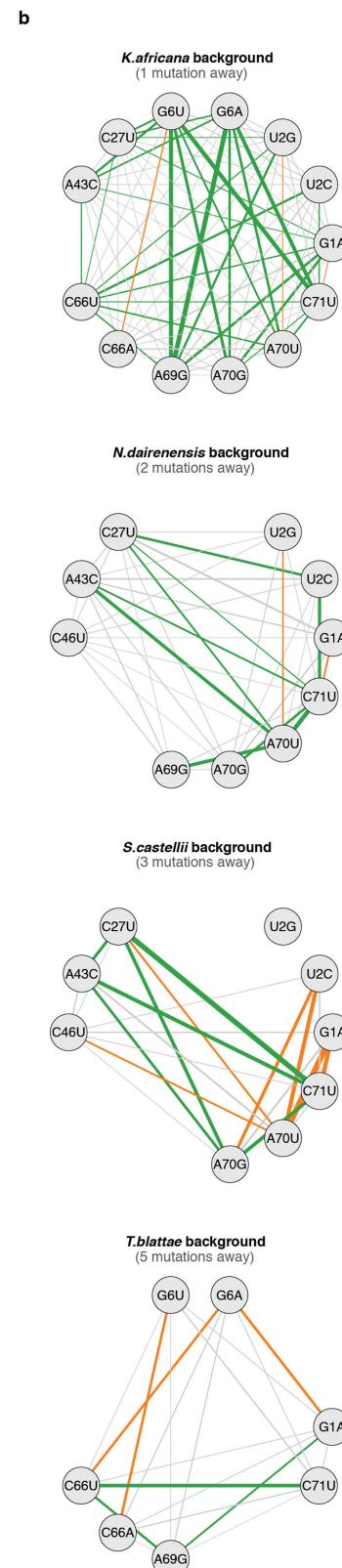
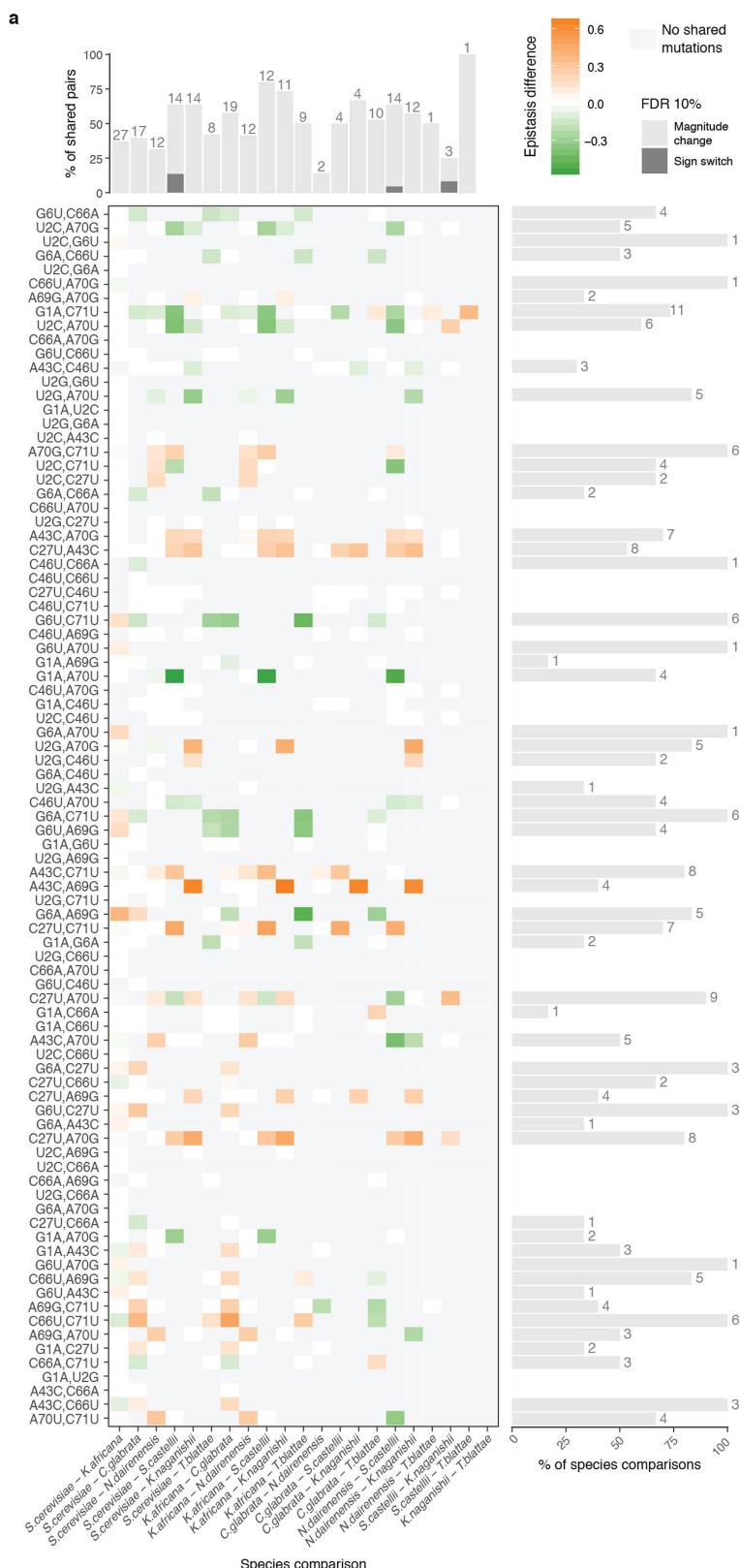
**Extended Data Fig. 2 | Mutations have varying fitness effects in different backgrounds.** **a**, Single mutations (columns) have effects that differ significantly between genetic backgrounds from different species (rows). Paired two-sided *t*-test between fitness effects of mutations of tRNAs from different species (145 tests of  $n=6$ ). Significant fitness effects differences ( $\text{FDR} < 0.1$ ) shown in blue (positive) or red (negative), non-significant differences ( $\text{FDR} \geq 0.1$ ) coloured in white. Mutations that were not shared are coloured in grey (that is, a substitution that would result in a mutation in one species but is part of the wild-type background in another). Bar plots show the percentage (absolute numbers on top) of species comparisons or shared mutations between species in which the effect of the mutation significantly changes in magnitude (light grey) or switches sign (dark grey). **b**, Proportion of genetic backgrounds in which

each mutation has a beneficial (blue) or detrimental (red) fitness effect at different FDRs for backgrounds with  $-0.3 < \ln(\text{fitness}) < -0.15$  (left), backgrounds with  $-0.15 < \ln(\text{fitness}) < 0.15$  (middle left), genotypes with no more than four mutations from the *S. cerevisiae* sequence (middle right) and genotypes with average input read counts of more than 100 (right). *q* values were obtained after adjusting for FDR across the total number of single mutations with unique background after filtering ( $n = 10,746, 6,129, 3,568, 6,338$  tests respectively). **c**, Fitness effect of single mutations plotted against the  $\ln(\text{fitness})$  of the backgrounds in which the mutation are made; for all genetic backgrounds (left), backgrounds with  $-0.3 < \ln(\text{fitness}) < -0.15$  (middle) and backgrounds with  $-0.15 < \ln(\text{fitness}) < 0.15$  (right).



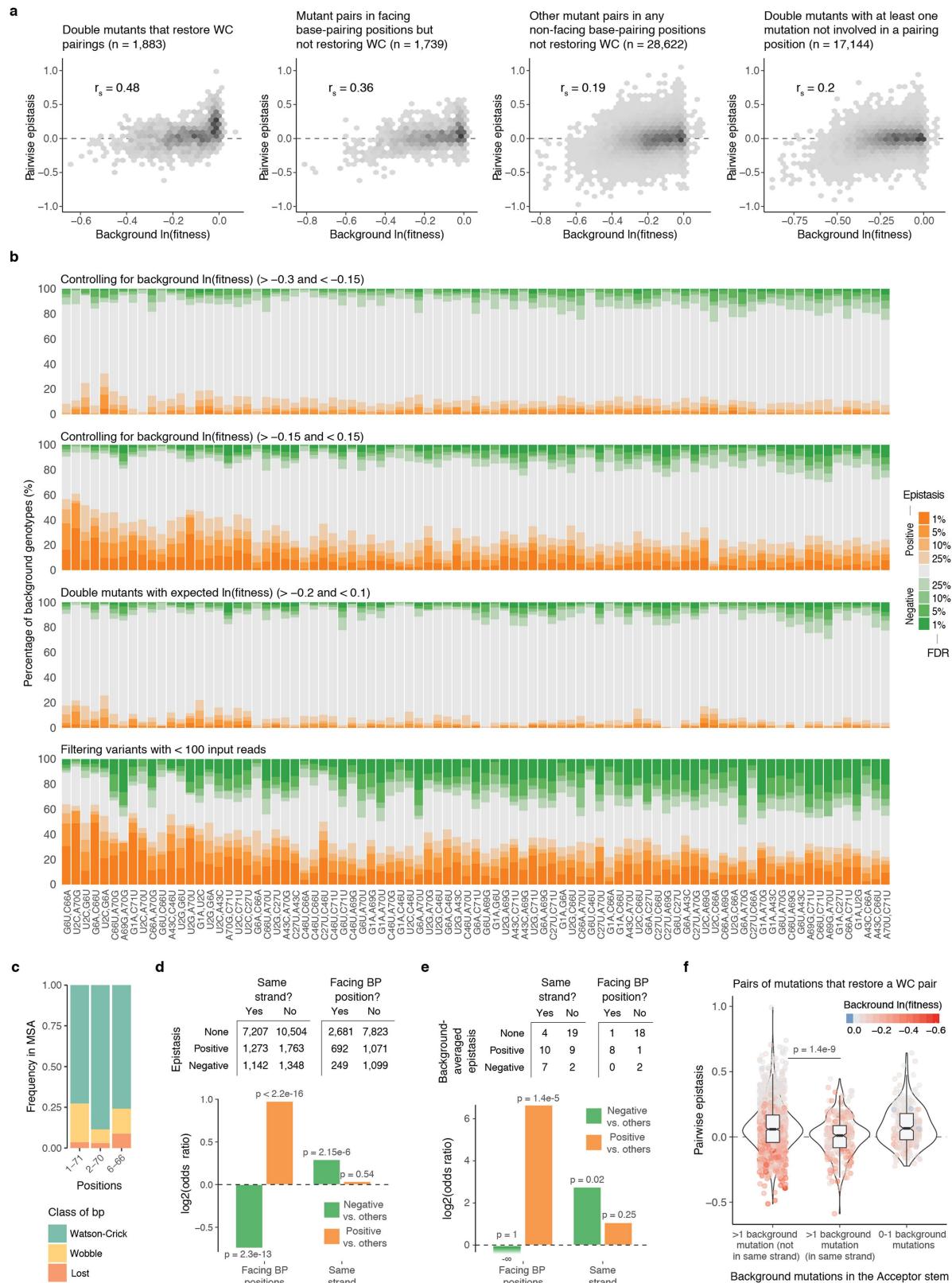
**Extended Data Fig. 3 | Comparison of epistasis scores between all pairs of species.** **a**, Comparison of epistasis scores for species pairs not shown in Fig. 3c. Pairs of species that share less than three mutations are not shown. **b**, Decline of correlation between epistasis scores and Hamming distance

between the tRNA genotypes from different species (inset). The left plot shows how this negative correlation holds when restricting the minimum number of shared pairs of mutations between the two species to compute the correlation.



**Extended Data Fig. 4 | Changes in pairwise epistasis between mutations across the seven extant species.** **a**, Comparison of pairwise epistasis (rows) between different species (columns) (1,000 paired two-sided *t*-tests of  $n=6$ ). Differences in epistasis are only shown for comparisons with FDR < 0.1 in orange or green for positive or negative differences respectively. Comparisons with FDR ≥ 0.1 are coloured in white. Pairs of mutations that are not shared between species are coloured in grey. Bar

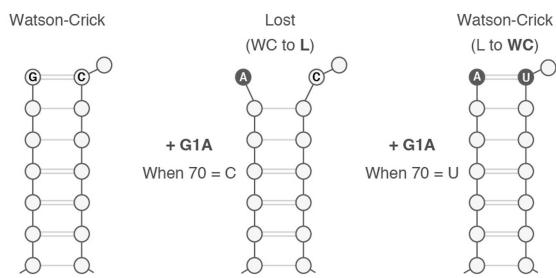
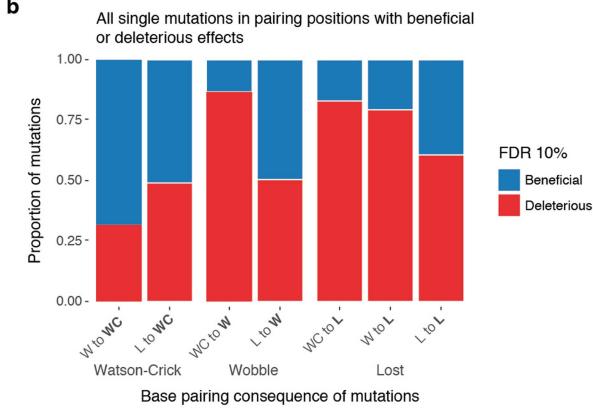
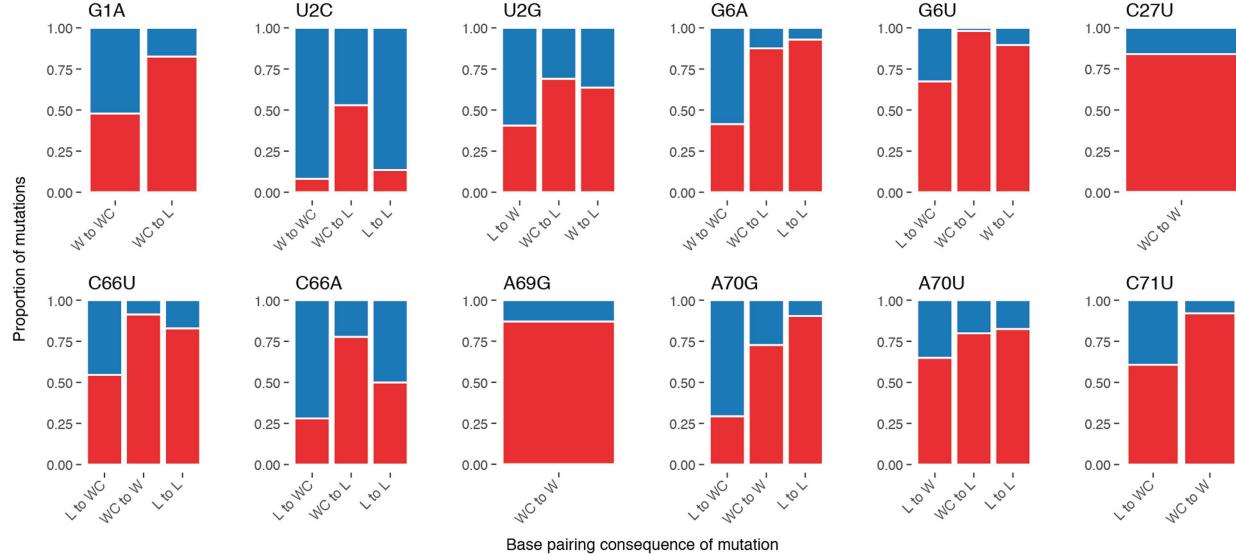
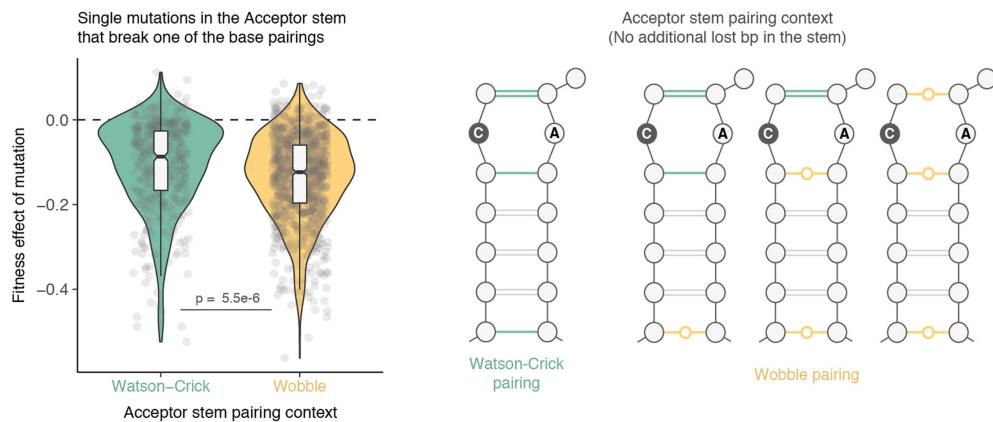
plots show the percentage of species comparisons (right) or shared pairs of mutations between species (top) that significantly change (light grey) or switch (dark grey). **b**, Interaction networks of four extant species not shown in Fig. 3b. Colours indicate epistasis sign (orange for positive, green for negative and grey for not significant at FDR < 0.1) and edge width indicates epistasis magnitude.



Extended Data Fig. 5 | See next page for caption.

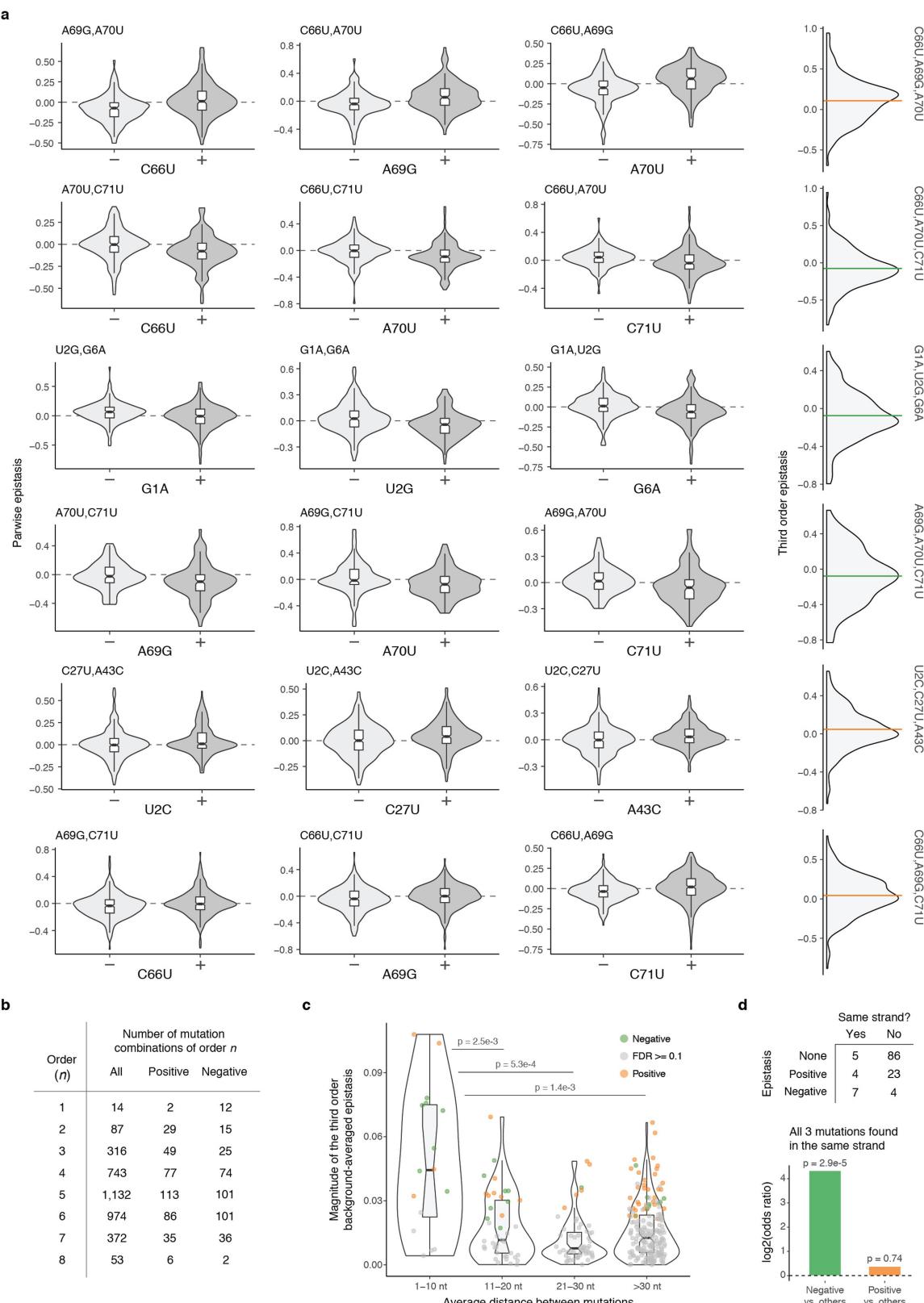
**Extended Data Fig. 5 | Pairwise epistatic interactions switch from positive to negative.** **a**, Epistasis scores between pairs of mutations plotted against the  $\ln(\text{fitness})$  of the genetic background. Scatter plots are divided into double mutants that restore WCBPs (left,  $n = 1,883$ ), other double mutants in which both mutation are in facing base pair positions (middle left,  $n = 1,739$ ), in base pair positions but not facing each other (middle right,  $n = 28,622$ ), and the rest (right,  $n = 17,144$ ). **b**, Proportion of genetic backgrounds in which each pair of mutations interacts with positive (orange) or negative (green) epistasis at different FDRs restricted to genetic backgrounds with  $-0.3 < \text{fitness} < -0.15$  (top), with  $-0.15 < \text{fitness} < 0.15$  (top middle), with additive expected fitness outcome greater than  $-0.2$  and less than  $0.1$  (middle bottom) or when excluding all genotypes with average input counts less than 100 (bottom). 23,128, 23,652, 29,628 and 15,306 one sample two-sided  $t$ -tests ( $n = 6$ ). **c**, A small fraction of tRNA-Arg(CCU) from other eukaryotic species have lost the base pairing in positions 1–71, 2–70 and 6–66 of the tRNA (multiple sequence alignment (MSA) across 1,614 species was taken from previously published work<sup>27</sup>; sequences with indels were excluded).

**d**, Number of positive, negative or not significant pairwise interactions at  $\text{FDR} < 0.1$  within the acceptor stem of the tRNA ( $n = 23,237$ ) when both mutations are found in the same helix strand or when each mutation is located in a different strand ( $n = 13,615$ ). log<sub>2</sub> odds ratio shown below together with two-sided Fisher's exact test  $P$  values. **e**, Number of positive, negative and non-significant background-averaged pairwise interactions between pairs of mutations in the acceptor stem that are found in the same RNA strand and between mutations that are in positions that base pair with each other. log<sub>2</sub> odds ratio and two-sided Fisher's exact test  $P$  values are shown below. **f**, Distribution of pairwise epistasis values of mutation pairs that restore a canonical WCBP depending on the location of their background mutations in the acceptor stem ( $P$  values from Welch's two-sided  $t$ -test,  $n = 263$  or  $n = 1,368$  when more than one background mutations are in the same strand or not, respectively). The same result is obtained when epistasis values are corrected for the  $\ln(\text{fitness})$  of the background (residuals of a linear model using background  $\ln(\text{fitness})$  to predict epistasis, data not shown).

**a****b****c****d**

**Extended Data Fig. 6 | Changes in base pairing partially explain the consequences on fitness of single mutations.** **a**, A single mutation can either disrupt or restore a canonical WCBP depending on the background context. **b**, Percentage of deleterious or beneficial single mutations (at  $FDR < 0.1$ ) that restore or disturb a canonical WCBP in any base pairing position of the tRNA. From a total of 4,300 mutations that restore WCBP, 721 are beneficial and 498 deleterious. 13,195 mutations result in the loss of a canonical pair ( $n = 6,806$  mutations that create a wobble base pair and  $n = 6,389$  that completely break the base pair interaction), of these 3,030 and 721 have significant deleterious and beneficial effects,

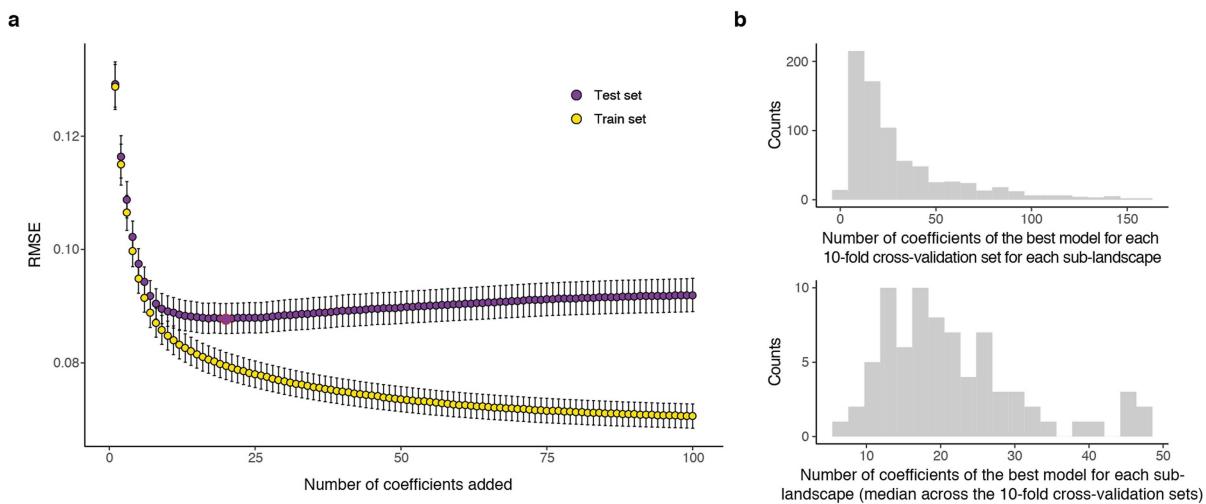
respectively. WC, Watson–Crick, W, wobble and L, lost base pair. **c**, Same as **b** but split by mutation identity. **d**, Distribution of the effects of mutations in the tRNA acceptor stem that break a base pairing (left,  $n = 1,356$  single mutations with higher background fitness than  $-0.15$ ) have more deleterious effects when the neighbour base-pairing positions are composed of one or more wobble interactions ( $n = 921$ ), instead of all canonical WCBP ( $n = 435$ , average fitness effect difference = 0.028, Welch's two-sided  $t$ -test  $P$  value shown). Right plot illustrates the context of the base pairing of the stem.



Extended Data Fig. 7 | See next page for caption.

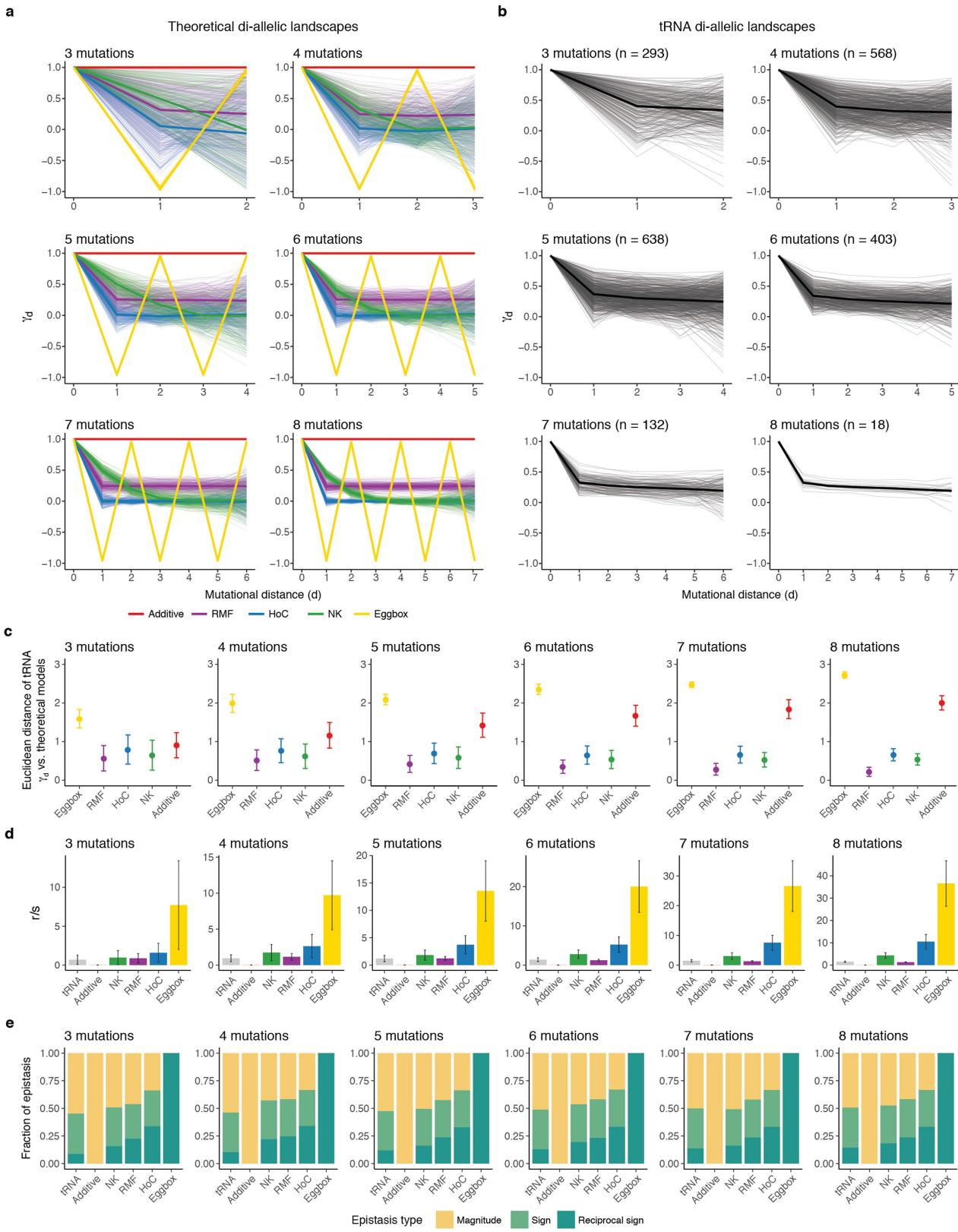
**Extended Data Fig. 7 | Background-averaged third and higher-order interactions.** **a**, The most significant background-averaged third-order interactions (8 out of 74, FDR < 0.1,  $n = 3,691$  tests for all interactions across all orders). The first three plots of each row show how the distribution of pairwise epistasis of two mutations across different genetic backgrounds (each double mutation can be found in a median of 506 different genetic backgrounds) changes in the presence or absence of a third mutation. The paired differences between pairwise interactions in those three cases correspond to third order epistatic coefficients. Distributions of third-order epistasis for the same three mutations are shown to the right. Horizontal lines correspond to the background-averaged third-order epistatic term, coloured by sign (orange or green for positive or negative respectively). **b**, Number of significantly positive and negative background-averaged epistatic interactions of order one

to eight (at FDR < 0.1). **c**, Distribution of the absolute magnitude of averaged third-order interactions plotted against the mean nucleotide distance between the three mutations ( $n = 316$  triple mutations). Welch's two-sided  $t$ -test  $P$  values for differences between the groups are shown. Significant interactions (one-sample two-sided  $t$ -test at FDR < 0.1) are coloured in orange or green for positive or negative epistasis respectively. **d**, Top, Number of positive, negative or non-significant background-averaged third-order interactions (FDR < 0.1) within the acceptor stem of the tRNA when both mutations are found in the same helix strand or not ( $n = 129$ ). Bottom, the log<sub>2</sub> odds ratios (when all three mutations are found in the same strand of the tRNA acceptor stem) of significantly positive interactions versus others (negative or not significant interactions) and significantly negative interactions versus other double mutants.  $P$  values reported from the two-sided Fisher's exact test.



**Extended Data Fig. 8 | Genetic prediction.** **a**, Mean RMSE of the fitness prediction for tenfold cross-validation held-out genotypes (purple, test set) or genotypes included in the training set (yellow) for each of the eight-mutation sub-landscapes when progressively adding the 100 most significant epistatic coefficients out of the 256 possible coefficients. Highlighted in red is the average number of epistatic coefficients to

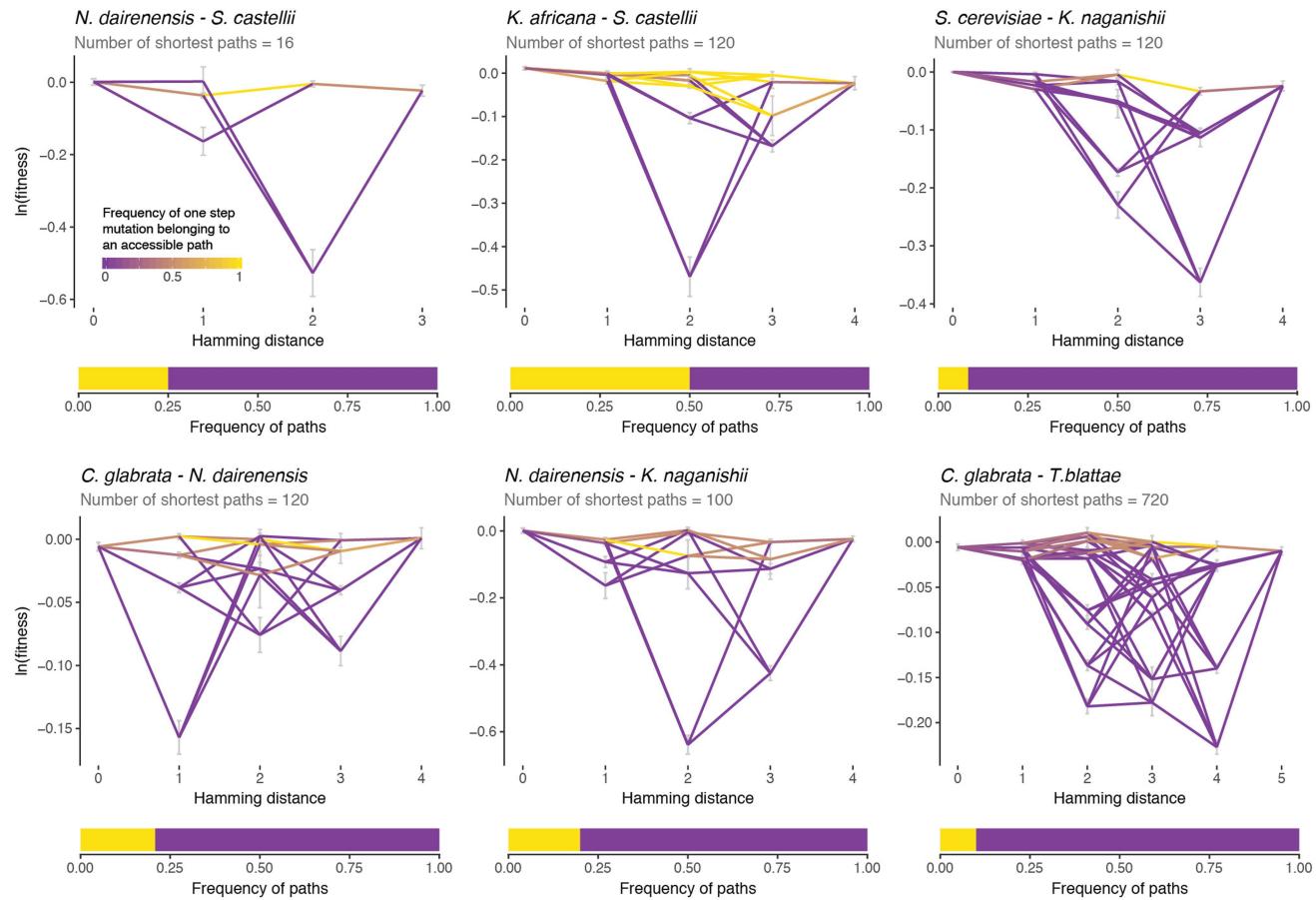
obtain the lowest RMSE across all the sub-landscapes. **b**, Histogram of the minimum number of epistatic coefficients that give the minimum RMSE when predicting the fitness of the test genotypes by tenfold cross-validation in all complete eight-mutation sub-landscapes (top). Histogram of the median number of coefficients for each sub-landscape (bottom).



Extended Data Fig. 9 | See next page for caption.

**Extended Data Fig. 9 | Comparison of the combinatorially-complete tRNA sub-landscapes to theoretical fitness landscapes.** **a**, Expected pattern of the average correlation of fitness effects  $\gamma_d$  at different mutational distances for theoretical di-allelic fitness landscapes with three to eight mutated positions. The average  $\gamma_d$  behaviour is highlighted in bold for each theoretical landscape ( $n = 250$  simulated landscapes for each theoretical model). The NK landscape was modelled with  $K = L/2$  ( $L$ , number of mutated positions) and the RMF as a mixture of 50% additive and 50% HoC. **b**, Decay of  $\gamma_d$  with mutational distance for all tRNA complete di-allelic sub-landscapes containing the *S. cerevisiae* parental genotype of three to eight loci (mean behaviour of  $\gamma_d$  in bold).

**c**, Mean euclidean distance between the  $\gamma_d$  for the tRNA sub-landscapes and the  $\gamma_d$  of theoretical landscapes (each tRNA landscape was compared to the 250 simulations of each theoretical landscape,  $n = 73,250, 142,000, 159,500, 100,750, 33,000$  and  $4,500$  for tRNA landscapes from three to eight mutations respectively). **d, e**, Mean roughness-to-slope ratio ( $r/s$ ) (**d**) and epistasis classes (**e**) for all combinatorially-complete tRNA di-allelic landscapes from three to eight mutations, as well as for all theoretical landscape models ( $n = 250$  for each theoretical landscape models and  $293, 568, 638, 403, 132$  and  $18$  tRNA landscapes from three to eight mutations respectively). Error bars are s.d.



**Extended Data Fig. 10 | Direct paths accessibility between extant species.** Shortest paths between some pairs of extant species (top) together with the proportion of them that are accessible (bottom; yellow, accessible; purple, inaccessible). Nodes are the  $\ln(\text{fitness})$  of the species genotypes and the intermediate genotypes between them. Edge colours indicate the

frequency at which a one-step mutation belongs to an accessible path (completely accessible, yellow; completely inaccessible, purple). Error bars are  $\ln(\text{fitness})$  s.e.m. of each genotype (propagated error from the  $n = 6$  replicates).

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  - Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted
  - Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated
- Clearly defined error bars
  - State explicitly what error bars represent (e.g. SD, SE, CI)*

*Our web collection on [statistics for biologists](#) may be useful.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

All the data analysed in this study was produced in this study. The sequencing reads of each sample (two inputs and six outputs) were processed and filtered independently. Each sequencing read covered the entire tRNA. The 5' and 3' constant regions of the read (primers annealing sites) were removed with the 'cutadapt' software. The forward and reverse reads were merged using 'PEAR' and sequences that were either not assembled due to low quality or unexpected length were discarded. Unique genotypes were called and quantified with custom python scripts. Genotypes with less than nine input reads in any input replicate, unexpected nucleotide substitutions (sequencing or PCR errors) or 0 reads in the outputs were discarded.

Data analysis

All data analysis were performed in R (version 3.3.3). We used the software 'MAGELLAN' to generate theoretical fitness landscapes and calculate the gamma statistic to compare the tRNA landscape to theoretical models.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The complete dataset is available as Supplementary Table 1. Custom code used in this study is available upon request. Raw sequencing data has been submitted to GEO (accession number GSE99418).

## Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](http://nature.com/authors/policies/ReportingSummary-flat.pdf)

## Life sciences

### Study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was determined by the number substitutions to co-occur in the evolution of the Arginine tRNA which results in a total library size of 5,184 ( $=2^6 \times 3^4$ ) possible mutation combinations.
Data exclusions	Sequencing data was filtered with the following criteria before being analyzed: Sequences that were either not assembled, due to low quality or unexpected length, were discarded. Variants with less than 9 input reads, unexpected nucleotide substitutions (sequencing or PCR errors) or 0 reads in the output were discarded. After filtering, we ended up with a total of 4,176 sequence variants quantified in all input and outputs.
Replication	The study included in total 6 replicates. 2 independent transformations (inputs) with each split into 3 independent selection experiments (outputs). All attempts of replications were successful.
Randomization	Samples were grouped by replicates and no other grouping or randomization of samples were done.
Blinding	There was no blinded data in this study. The only group of data during the analysis were the replicates.

### Materials & experimental systems

Policy information about [availability of materials](#)

n/a	Involved in the study
<input checked="" type="checkbox"/>	Unique materials
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Research animals
<input checked="" type="checkbox"/>	Human research participants

### Method-specific reporting

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	Magnetic resonance imaging