Article

High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast

Alex N. Nguyen Ba[1,9], Ivana Cvijović[1,2,3,4,9], José I. Rojas Echenique[1,9], Katherine R. Lawrence[1,5], Artur Rego-Costa[1], Xianan Liu[6,7], Sasha F. Levy[6,7] & Michael M. Desai[1,3,4,8*]

In rapidly adapting asexual populations, including many microbial pathogens and viruses, numerous mutant lineages often compete for dominance within the population[1–5]. These complex evolutionary dynamics determine the outcomes of adaptation, but have been difficult to observe directly. Previous studies have used whole-genome sequencing to follow molecular adaptation[6–10]; however, these methods have limited resolution in microbial populations. Here we introduce a renewable barcoding system to observe evolutionary dynamics at high resolution in laboratory budding yeast. We find nested patterns of interference and hitchhiking even at low frequencies. These events are driven by the continuous appearance of new mutations that modify the fates of existing lineages before they reach substantial frequencies. We observe how the distribution of fitness within the population changes over time, and find a travelling wave of adaptation that has been predicted by theory[11–17]. We show that clonal competition creates a dynamical 'rich-get-richer' effect: fitness advantages that are acquired early in evolution drive clonal expansions, which increase the chances of acquiring future mutations. However, less-fit lineages also routinely leapfrog over strains of higher fitness. Our results demonstrate that this combination of factors, which is not accounted for in existing models of evolutionary dynamics, is critical in determining the rate, predictability and molecular basis of adaptation.

Rapidly adapting populations have complex evolutionary dynamics. In these systems, adaptation is not limited by the supply of mutations[18]. Instead, numerous beneficial mutations arise simultaneously and drive competing clonal expansions[7–9], often accompanied by neutral and deleterious hitchhiker mutations[6,19]. Studies have shown that this is the dominant mode of adaptation in many bacterial and viral pathogens[20–22], as well as in the somatic evolution of cancer[23] and immune repertoires[24]. In these contexts, clonal interference and hitchhiking have important consequences for the pace, outcomes and repeatability of evolution.
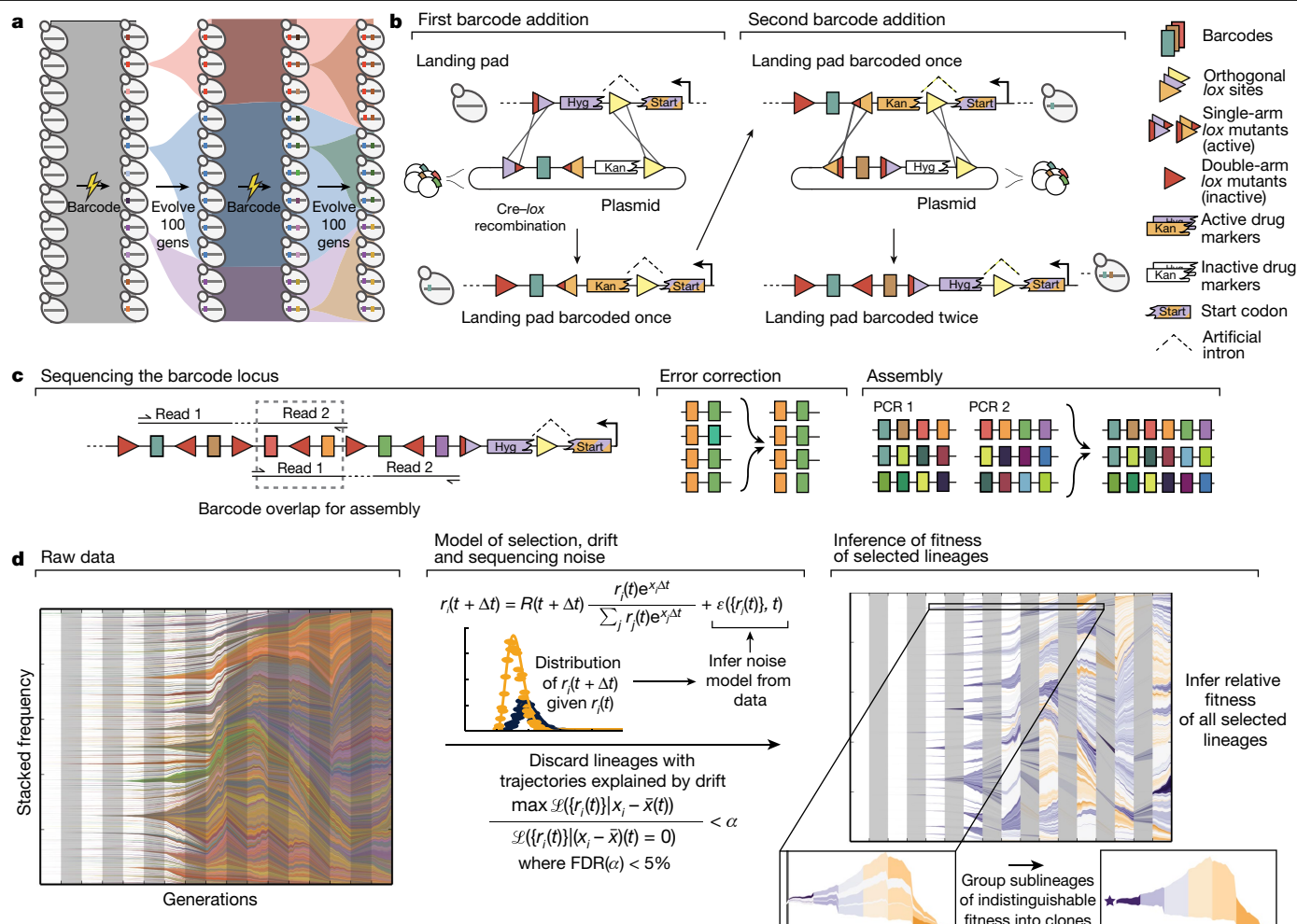
This mode of rapid adaptation cannot be described by classical evolutionary theory, because the fate of each mutation cannot be modelled in isolation[11,25,26]. Instead, selection acts on physically linked combinations of alleles, which leads to complex co-dependency between the fates of different mutations. This limits the efficiency of selection and renders evolution less predictable: strongly beneficial mutations can be outcompeted, whereas deleterious mutations in good genetic backgrounds can spread through the population[6,13,27].

Numerous studies have used whole-genome sequencing to investigate these effects in laboratory microbial populations[6–10], and have shown that clonal interference and hitchhiking are widespread. However, limitations on sequencing depth make it impractical to achieve a frequency resolution of higher than a few per cent, and barcoding-based methods[5,28,29] that offer better resolution are limited to short timescales. These limitations are critical in large microbial populations, in which theory suggests that the fates of mutations are often determined over long timescales by competition and hitchhiking among rare high-fitness lineages, and that the vast majority of driver mutations never reach detectable frequencies[11–17].

A renewable barcoding system

Here, we develop a renewable barcoding approach to observe evolutionary dynamics at high resolution on long timescales, by periodically adding new barcodes to split each tracked lineage into labelled 'sublineages' (Fig. 1a). Our approach uses three orthogonal *lox* sites: Cre-mediated recombination occurs between sites of the same type,

Fig. 1 | Renewable barcoding system and lineage dynamics. a, Experimental design. Diverse DNA barcodes are introduced into an initially clonal population, with each barcode labelling a small lineage. Every 100 generations (gens), we introduce new diverse barcodes immediately adjacent to existing barcodes, thereby subdividing each lineage into sublineages. **b**, Renewable barcoding system. The Cre–lox system consists of three orthogonal *lox* sites (coloured triangles), each of which can be modified with two arm disruptions (red) that are individually tolerated but jointly inactivating (Supplementary Information section 1). At each barcode addition, we combine arm disruptions to inactivate the old *lox* site, while adding a new orthogonal active *lox* site. Alternating *lox* orientations further limit undesired recombination. Drug markers (Kan, kanamycin (G418) resistance; Hyg, hygromycin B resistance) contain an intron 3′ splice-accepting site and must correctly integrate at the landing pad that contains the 5′ splice donor to be functional. **c**, When the barcode locus exceeds the length of an Illumina read, we use custom priming sites to sequence overlapping sets of four consecutive barcodes. After exploiting barcode diversity to identify and correct sequencing errors, we use these overlaps to unambiguously reconstruct the full barcode locus (Supplementary Information section 2). **d**, Inference pipeline. Left, raw barcode frequencies over time (left to right; colours chosen at random). For legibility, we only show lineages or sublineages with a frequency that exceeds 0.1% in at least one time point. Combined frequencies of lineages that do not individually reach 0.1% are shown as white space (or the colour of the parent when that parent exceeds a frequency of 0.1%). Middle, summary of the model used for identifying selected lineages (see Supplementary Information section 4 for details). In brief, we use the data to construct a parametric model for the strength of noise from genetic drift and sequencing and discard trajectories that are explained by noise alone, at a false discovery rate (FDR) of 5%. We then jointly infer the fitness of all remaining lineages and group lineages of indistinguishable fitness into clones (right).

but not between orthogonal types. Each site can be inactivated by two specific arm disruptions (one in each of the two Cre-binding regions), but retains high activity with only one disruption. We used this system to design barcoded plasmid libraries with complementary Cre–lox architecture, which we use to integrate barcodes at a designated genomic 'landing pad' locus (Fig. 1b, Supplementary Information section 1). At each barcode addition, Cre-mediated recombination combines arm disruptions to inactivate an old *lox* site, and adds a new orthogonal *lox* site with a single arm disruption to be used for the next barcode addition with a complementary plasmid library (Supplementary Information section 1). Each plasmid also contains an inactive drug marker that lacks a start codon; correct integration activates this marker by combining it with a start codon in the landing pad, separated by an artificial intron.

This system integrates new DNA barcodes immediately downstream of existing barcodes. Each individual thus acquires a string of barcodes that encode its ancestry, which can be read by sequencing. We read four barcodes per 150-bp paired-end Illumina read; when the barcode locus exceeds this length, we exploit overlapping fragments to assemble the complete locus (after using high barcode diversity to correct sequencing errors) (Fig. 1c, Supplementary Information sections 1.5, 2). This allows us to track the frequencies of all lineages and sublineages and hence trace the ancestry of the entire population.

Lineage tracking in evolving populations

We used this system to evolve two diploid yeast populations founded from identical clonal ancestors, each labelled with about 50,000 diverse
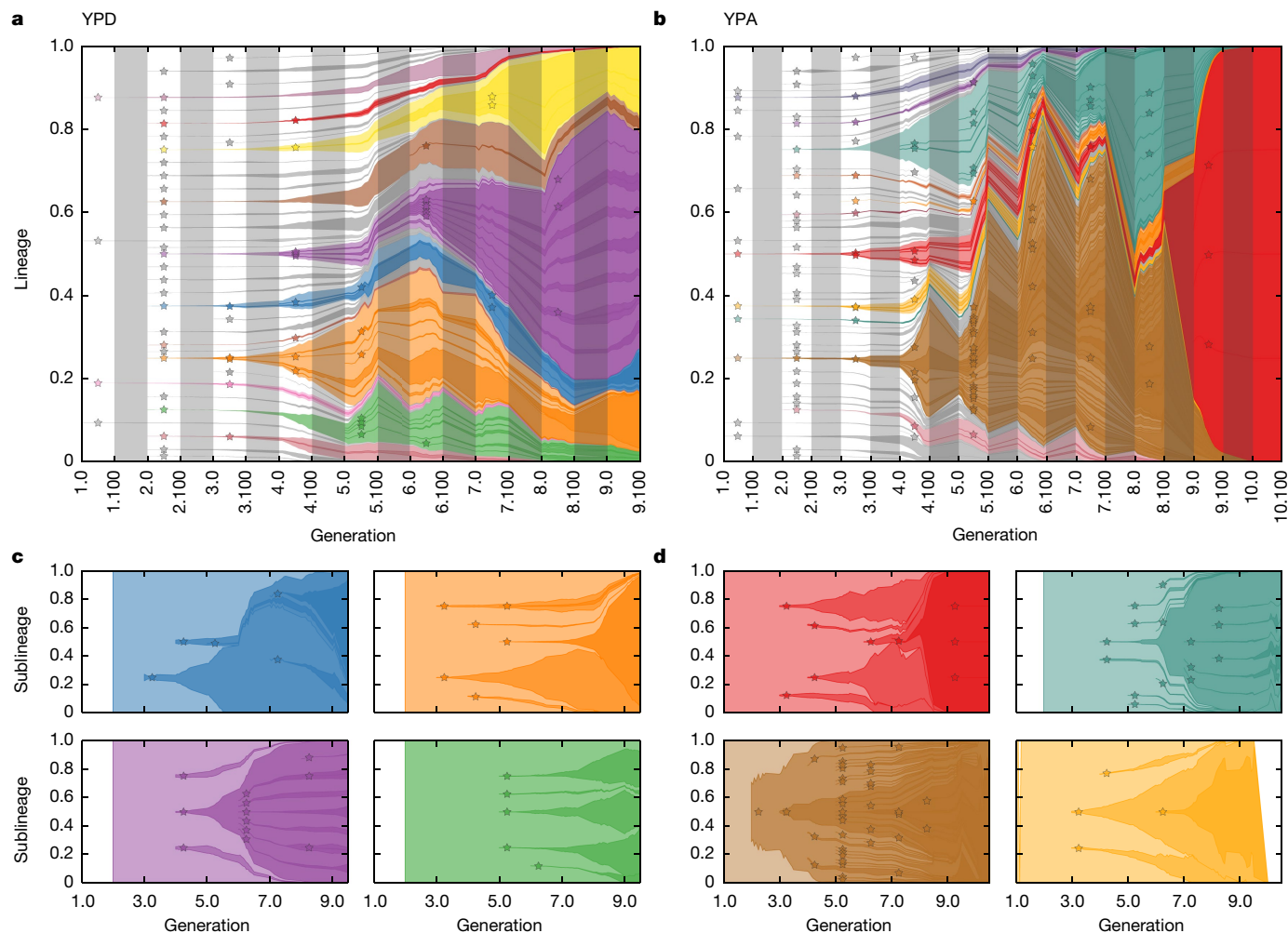
Fig. 2 | Inferred clonal dynamics. a, **b**, Muller diagrams showing dynamics of inferred beneficial mutations in YPD (**a**) and YPA (**b**) populations. Time is expressed in terms of epoch and generation (for example, 4.100 refers to generation 100 of epoch 4). Stars denote the establishment epoch of each new beneficial mutation (see Supplementary Information section 5). The opacity of the colours denotes the fitness of the corresponding lineage; mutant lineages that did not acquire additional beneficial mutations are grey. Grey bars denote barcoding intervals. **c**, **d**, Muller diagrams showing within-lineage dynamics in select lineages in the YPD (**c**) and YPA (**d**) populations. Colours are consistent with corresponding lineages in **a** and **b**. White space indicates periods during which the selected lineage was not observed.

barcodes. We maintained both populations in batch culture, with a 1:1,024 dilution every 24 h (10 generations per day with a bottleneck of about 500,000 cells; an effective size (N_e) of 5×10^6). An aliquot was frozen daily for analysis (Supplementary Information section 1.4). One population was maintained in rich medium (YPD) and the other in rich medium with 0.3% acetic acid (YPA), which leads to intracellular acidification that pilot studies have suggested leads to stronger selection pressures[30]. In studying these populations, our goal was to identify generic features of the evolutionary dynamics rather than details of differences between conditions. Our choice of environments maintains consistency with previous work, which indicates that these environments lead to rapid adaptation involving rich dynamics that could not be observed using earlier approaches[5–8].

We re-barcoded each population every 100 generations with about 50,000 additional unique barcodes. This diversity was chosen to ensure that barcoding does not introduce a substantial bottleneck; at 10% of the daily bottleneck every 10 days, it does not change the scale of genetic drift or the effective population size (Supplementary Information section 4.4). It also ensures that we can detect relevant selection pressures that act on lineages once those lineages become large enough that their dynamics are not dominated by drift (Supplementary Information section 4.4). However,

we note that although our barcoding procedure is designed to be minimally perturbative, it does involve propagation and selection steps. Thus, strictly speaking we are studying evolution in a fluctuating environment that alternates between 'evolution' and 'barcoding' conditions—although, as we see below, the role of these fluctuations is minor.

After 1,000 generations of evolution (ten 100-generation 'epochs'), we sequenced the barcode locus at a depth of around 10^5 reads in every frozen time point. This yielded 110 sequenced time points per population (11 time points per epoch at 10-generation intervals, although we excluded the final epoch of the YPD population owing to barcoding failure; see Supplementary Information section 1.4). We use this data to infer which lineages contain mutations that are beneficial in either evolution or barcoding conditions, and we exploit phylogeny to infer in which epoch each mutation was established (that is, within 100-generation resolution; Fig. 1d, Supplementary Information section 5). This allows us to group barcodes into 'clones', each founded by a new mutation. We then jointly infer the fitness effects of all mutations in evolution and barcoding conditions. Because we barcode frequently, the dynamics are determined by the average fitness across the two conditions (Supplementary Information section 6.2). We therefore use this average fitness for the analysis below (although simply neglecting
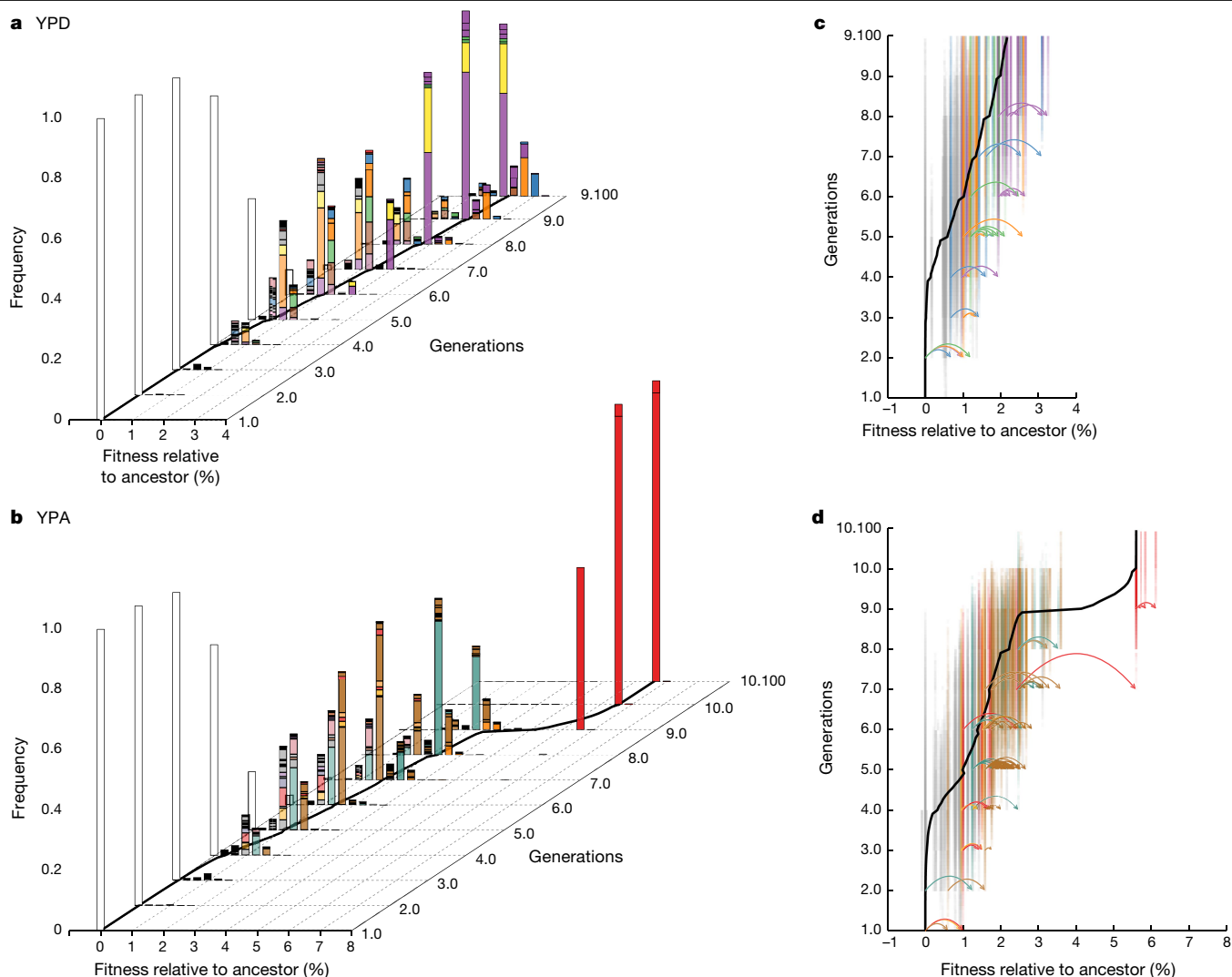
Fig. 3 | Travelling wave dynamics. a, **b**, Inferred distribution of fitness within the YPD (**a**) and YPA (**b**) populations over time. All fitness are the average across evolution and barcoding conditions (Supplementary Information 6.2). Each coloured bar denotes the frequency and fitness of a corresponding lineage in Fig. 2. White bars correspond to the ancestor. Black lines denote the inferred mean fitness of the population. **c**, **d**, Genealogical relationships among lineages in the YPD (**c**) and YPA (**d**) populations show frequent leapfrogging events. Each clonal lineage is shown at its corresponding fitness. The opacity of the colours indicates the frequency of the lineage. Colours of the lineages shown in Fig. 2c, d are consistent with that figure; all other lineages are grey. Mutational events within highlighted lineages are shown as arrows; each event arises in one clonal lineage and founds a new lineage at a new fitness.

the barcoding environment leads to qualitatively similar conclusions; see Supplementary Information section 6).

Our ability to detect mutations is limited primarily by genetic drift. We cannot identify mutations until they are common enough that their fitness effects lead to frequency changes larger than this stochastic force (which typically corresponds to lineages at frequencies greater than 10^{-4}). Because fitness inference requires sufficient time-course data, we are also unable to detect most mutations that arise in the final 100–200 generations of the experiment (Supplementary Information section 5.3). Our analysis thus only identifies a subset of beneficial mutations, and our clones are clonal only with respect to these.

We find that in both populations, many beneficial mutations arise early in the experiment, founding clones that compete for dominance (Fig. 2a, b). Some of these clones diversify through further beneficial mutations, and a handful obtain multiple mutations, which interfere with one another within the parent clone (Fig. 2c, d). In some cases we observe multiple nested interference events (Fig. 2c, d). All but the largest of these events are undetectable by metagenomic sequencing at approximately 25× depth (which corresponds to about the same total number of sequencing reads as our barcode data; Extended Data Fig. 1a, b).

We can also visualize how the fitness composition of the population changes over time (Fig. 3). The population initially diversifies as numerous beneficial mutations arise on the ancestral background, creating a distribution of fitness within the population (Fig. 3a, b). As these clones expand, the mean fitness of the population increases (Fig. 3, Extended Data Fig. 2), causing less-fit lineages to fall behind and begin to die out. However, diversity is maintained by new beneficial mutations, which continuously create clones of even higher fitness (Fig. 3c, d). This maintenance of diversity in the face of strong selection is an expected feature of rapid adaptation that has been predicted by theory[11–17] but not previously observed directly.

Determinants of lineage success

These dynamics lead to a complex picture of the determinants of success of individual lineages. In the absence of further mutations, the
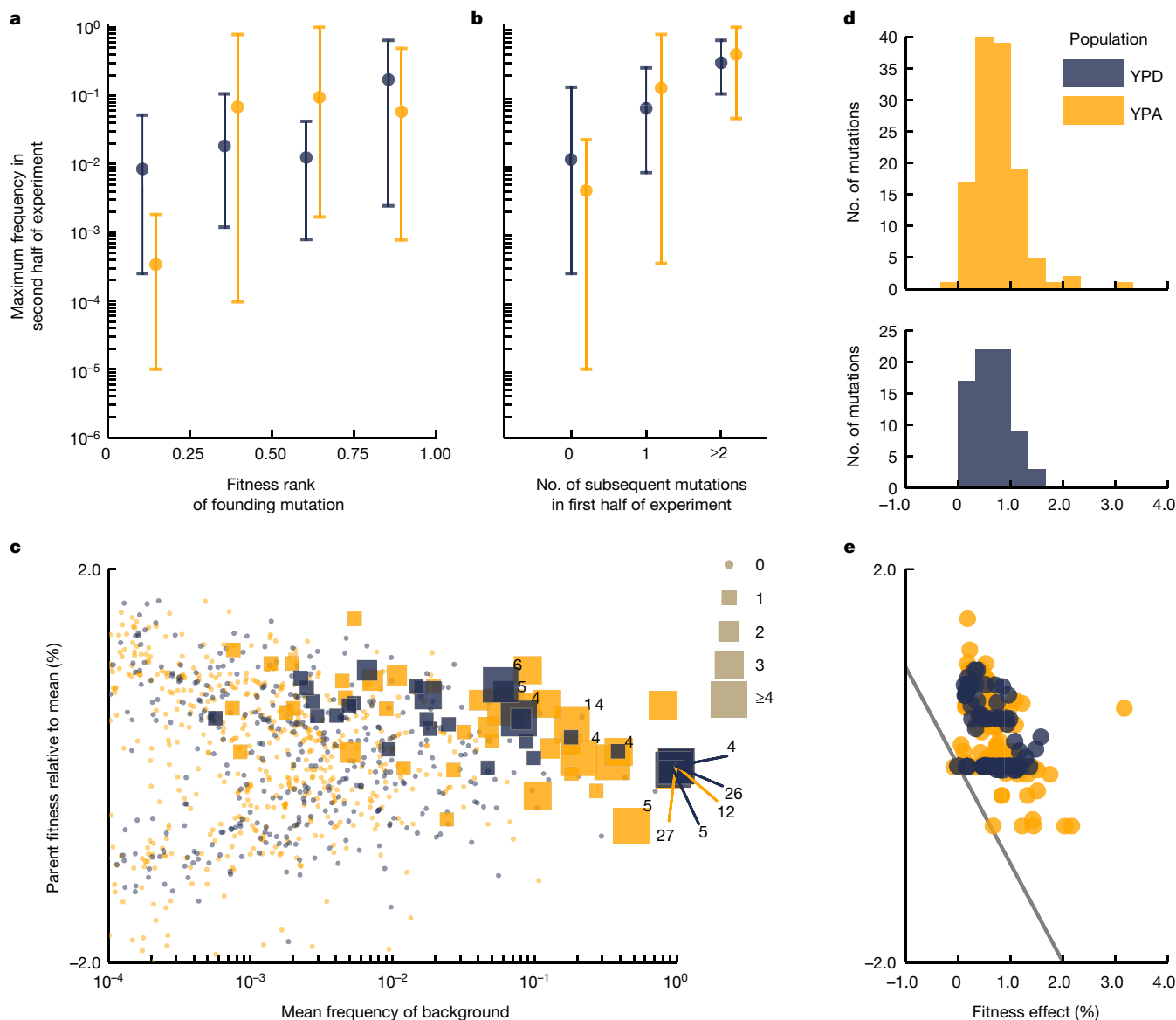
Fig. 4 | Travelling wave dynamics and factors determining the success of mutant lineages. a, Relationship between initial within-population fitness rank of a mutation that arises in the ancestral background and its maximum frequency in the second half of the experiment (using the second half avoids confounding axes in **b**). $n = 35$ and $n = 47$ unique lineages in YPD and YPA respectively. Dots represent the mean, and lines show the range of maximum frequencies in each founding fitness quantile. **b**, Relationship between the number of subsequent beneficial mutations landing on the founding clonal background of a lineage (in the first half of the experiment) and the eventual maximum frequency of that lineage (in the second half of the experiment). **c**,

Effect of lineage frequency and fitness on the likelihood of acquiring additional beneficial mutations. Each point represents the mean frequency and fitness of a lineage in a given 100-generation epoch; symbol size denotes how many additional beneficial mutations that lineage acquired (numbers indicate lineages that acquire more than four). **d**, Histograms of effect sizes of all inferred mutations. **e**, Effect sizes of mutations arising on parental backgrounds as a function of mean parental relative fitness in the epoch in which each mutation arose. The region below the grey line corresponds to mutations that would create lineages less fit than the current mean fitness.

fitness of a lineage should be the only predictor of its success. Yet we find that the initial fitness of a mutant lineage is only a modest predictor of its fate (Fig. 4a). Another key factor is whether a lineage acquires further beneficial mutations (Fig. 4b). Although this is influenced by fitness (see below), even high-fitness lineages that do not acquire further beneficial mutations are readily outcompeted, and lower-fitness lineages that acquire multiple mutations can succeed (Extended Data Fig. 3). The likelihood of a lineage acquiring further beneficial mutations is in turn affected by two main factors (Fig. 4c). First, larger lineages have more opportunities to acquire beneficial mutations. Second, the fitness of a lineage has a critical role: mutations that arise in a highly fit and hence rapidly expanding lineage will be less likely to be lost to genetic drift.

Thus, highly fit backgrounds can accumulate beneficial mutations of both strong and weak effect, whereas only rare strong mutations can establish on lower-fitness backgrounds. Consistent with this, our data show that high-fitness backgrounds acquire both weakly and strongly beneficial mutations, but low-fitness backgrounds only acquire the latter (Fig. 4d, e). This means that fitter backgrounds have access to a larger number of beneficial mutations, creating a rich-get-richer effect that can lead to bursts of mutations at the expanding front of the fitness wave. These bursts arise owing to dynamical considerations, and are not in themselves evidence of historical contingency as a result of mutator phenotypes or other modifiers of adaptability (Supplementary Information section 6.5).

These results are qualitatively consistent with recent theory that suggests that rapidly evolving populations can be described by 'travelling wave' models[11–17]. In this picture, mutations continuously generate variation in fitness while selection destroys it by eliminating less-fit genotypes, leading to a broad distribution of fitness around an increasing mean (a fitness wave). However, these models have only been analysed in parameter regimes in which the future common ancestor of a population is always one of the fittest lineages (although see one previous study that discusses scenarios in which this can be violated[17]). Instead, clonal competition in our experiment is characterized by routine 'leapfrogging' events, in which lineages of initially low relative fitness acquire strong beneficial mutations that pull them to prominence, causing dramatic reversals of fate. For example, in the YPD population (Fig. 3c) the green lineage—which is the fittest at the start—is leapfrogged by the orange, blue and purple backgrounds; the blue lineage then falls behind, only to later leapfrog all others. Similarly, in the YPA population (Fig. 3d), the brown lineage appears to outcompete the turquoise, red and yellow lineages, only to be leapfrogged by two strongly beneficial mutations in a red lineage that is initially much less fit (replay experiments validate this event; see Supplementary Information section 6.3).

Leapfrogging events not only alter the fates of individual lineages, but also cause fluctuations in the fitness distribution and modulate the pace of adaptation. Both within-population fitness and genetic variation increase during initial diversification before reaching a plateau as a travelling wave is established (Fig. 3, Extended Data Fig. 4). However, leapfrogging can cause fluctuations in this travelling wave: the creation of a lineage with anomalously high fitness can lead to a reduction in diversity at first, but at the same time enable rapid further diversification within this lineage that later re-establishes variation (Fig. 3c, d, Extended Data Fig. 4). These fluctuations affect the success of any individual mutation and the dynamics of the travelling wave, and hence have a major role in determining the outcomes of evolution.

Discussion

Previous theory has assumed that the effects of leapfrogging and fluctuations are occasional perturbations that can be largely ignored[11–17]. Our results suggest that they instead have a central role. Although our system involves microbial populations of modest size, the importance of these effects is expected to depend only weakly on population size and mutation rate (because relevant timescales only depend logarithmically on these quantities[12,17]). Thus our results suggest that leapfrogging and fluctuations may be routine in the evolution of a wide range of microorganisms and viruses. A new theoretical framework is essential to develop accurate models of evolution in these systems. The renewable barcoding approach we have introduced here offers the potential to test these models, and to observe evolutionary dynamics in a variety of contexts at sufficient resolution to investigate the role of other factors such as frequency-dependent selection or mutations that alter the adaptability of individual lineages.

Reporting summary

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

Data availability

Raw sequencing reads have been deposited in the NCBI BioProject database under accession number PRJNA559526. All associated metadata, as well as the source code for the sequencing pipeline, downstream analyses, and figure generation, are available at GitHub (https://github.com/icvijovic/lineage-tracking). Source Data for Figs. 2–4 are provided with the paper.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1749-3.

1. Gerrish, P. J. & Lenski, R. E. The fate of competing beneficial mutations in an asexual population. *Genetica* **102–103**, 127–144 (1998).
2. Desai, M. M., Fisher, D. S. & Murray, A. W. The speed of evolution and maintenance of variation in asexual populations. *Curr. Biol.* **17**, 385–394 (2007).
3. Miller, C. R., Joyce, P. & Wichman, H. A. Mutational effects and population dynamics during viral adaptation challenge current models. *Genetics* **187**, 185–202 (2011).
4. De Visser, J. A. G. M. et al. Diminishing returns from mutation supply rate in asexual populations. *Science* **283**, 404–406 (1999).
5. Levy, S. F. et al. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519**, 181–186 (2015).
6. McDonald, M. J., Rice, D. P. & Desai, M. M. Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature* **531**, 233–236 (2016).
7. Lang, G. I. et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **500**, 571–574 (2013).
8. Kvitek, D. J. & Sherlock, G. Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment. *PLoS Genet.* **9**, e1003972 (2013).
9. Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E. & Desai, M. M. The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45–50 (2017).
10. Tenaillon, O. et al. Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* **536**, 165–170 (2016).
11. Neher, R. A. Genetic draft, selective interference, and population genetics of rapid adaptation. *Annu. Rev. Ecol. Evol. Syst.* **44**, 195–215 (2013).
12. Desai, M. M. & Fisher, D. S. Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* **176**, 1759–1798 (2007).
13. Good, B. H., Rouzine, I. M., Balick, D. J., Hallatschek, O. & Desai, M. M. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proc. Natl Acad. Sci. USA* **109**, 4950–4955 (2012).
14. Rouzine, I. M., Brunet, E. & Wilke, C. O. The traveling-wave approach to asexual evolution: Muller's ratchet and speed of adaptation. *Theor. Popul. Biol.* **73**, 24–46 (2008).
15. Tsimring, L. S., Levine, H. & Kessler, D. A. RNA virus evolution via a fitness-space model. *Phys. Rev. Lett.* **76**, 4440–4443 (1996).
16. Hallatschek, O. The noisy edge of traveling waves. *Proc. Natl Acad. Sci. USA* **108**, 1783–1787 (2011).
17. Fisher, D. S. Asexual evolution waves: fluctuations and universality. *J. Stat. Mech.* **2013**, P01011 (2013).
18. Cvijović, I., Nguyen Ba, A. N. & Desai, M. M. Experimental studies of evolutionary dynamics in microbes. *Trends Genet.* **34**, 693–703 (2018).
19. Buskirk, S. W., Peace, R. E. & Lang, G. I. Hitchhiking and epistasis give rise to cohort dynamics in adapting populations. *Proc. Natl Acad. Sci. USA* **114**, 8330–8335 (2017).
20. Zanini, F. et al. Population genomics of intrapatient HIV-1 evolution. *eLife* **4**, e11282 (2015).
21. Lieberman, T. D. et al. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat. Genet.* **43**, 1275–1280 (2011).
22. Strelkowa, N. & Lässig, M. Clonal interference in the evolution of influenza. *Genetics* **192**, 671–682 (2012).
23. Nik-Zainal, S. et al. The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
24. Nourmohammad, A., Otwinowski, J., Luksza, M., Mora, T. & Walczak, A. M. Fierce selection and interference in B-cell repertoire response to chronic HIV-1. *Mol. Biol. Evol.* **36**, 2184–2194 (2019).
25. Muller, H. Some genetic aspects of sex. *Am. Nat.* **66**, 118–138 (1932).
26. Maynard Smith, J. Evolution in sexual and asexual populations. *Am. Nat.* **102**, 469–473 (1968).
27. Good, B. H. & Desai, M. M. Deleterious passengers in adapting populations. *Genetics* **198**, 1183–1208 (2014).
28. Blundell, J. R. & Levy, S. F. Beyond genome sequencing: lineage tracking with barcodes to study the dynamics of evolution, infection, and cancer. *Genomics* **104**, 417–430 (2014).
29. Blundell, J. R. et al. The dynamics of adaptive genetic diversity during the early stages of clonal evolution. *Nat. Ecol. Evol.* **3**, 293–301 (2019).
30. Ludovico, P., Sousa, M. J., Silva, M. T., Leão, C. & Côrte-Real, M. *Saccharomyces cerevisiae* commits to a programmed cell death process in response to acetic acid. *Microbiology* **147**, 2409–2415 (2001).

Article

Author contributions A.N.N.B., J.I.R.E. and M.M.D. designed the project; A.N.N.B. and J.I.R.E. constructed the barcoding system with assistance from X.L., S.F.L. and M.M.D.; A.N.N.B., J.I.R.E., K.R.L. and A.R.-C. conducted the experiments; A.N.N.B., J.I.R.E. and I.C. designed and conducted the bioinformatics analysis; I.C. developed theory and inference methods and analysed the data; I.C., M.M.D., A.N.N.B. and J.I.R.E. wrote the paper.

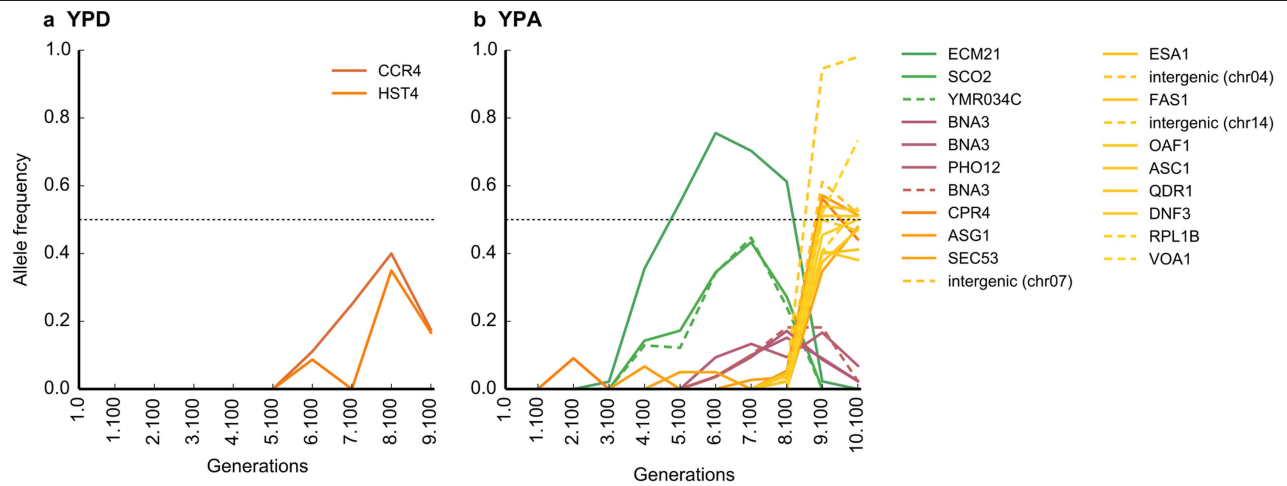**Competing interests** The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1749-3.
Correspondence and requests for materials should be addressed to M.M.D.
Peer review information *Nature* thanks David Gresham, Daniel Weinreich and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
**Reprints and permissions information** is available at http://www.nature.com/reprints.

a YPD

Legend: CCR4, HST4

b YPA

Legend:
ECM21, SCO2, YMR034C, BNA3, BNA3, PHO12, BNA3, CPR4, ASG1, SEC53, intergenic (chr07), ESA1, intergenic (chr04), FAS1, intergenic (chr14), OAF1, ASC1, QDR1, DNF3, RPL1B, VOA1

Extended Data Fig. 1 | Allele frequency trajectories in the two populations, as detected by metagenomic sequencing. a, **b**, In both the YPD (**a**) and the YPA (**b**) population, solid lines denote missense and nonsense mutations, and dotted lines denote synonymous mutations and those falling in intergenic regions. Lines are coloured according to the peak time of the trajectory. Note that a frequency of 50% (dotted line) corresponds to a mutation that fixes as a heterozygote.
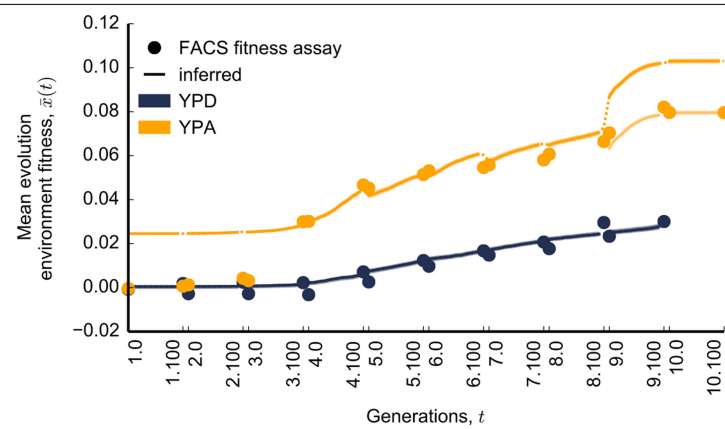
Extended Data Fig. 2 | Comparison of inferred and measured population mean fitness trajectories. All fitness measurements and inferences refer to the evolution environment only. Trajectories have been offset to agree with the fitness assay at time point 3.100. Dots denote barcoding intervals. Shaded regions around the trajectories denote estimates of 95% confidence intervals for the inferred mean fitness trajectory, which often do not exceed the width of the lines (Supplementary Information section 6.1). In the case of the YPA population, lighter colours denote mean fitness trajectories over the last two epochs, offset to agree with fitness assays in the last time point (see Supplementary Information section 6.6 for a discussion of potential reasons for these discrepancies) FACS, fluorescence-activated cell sorting.
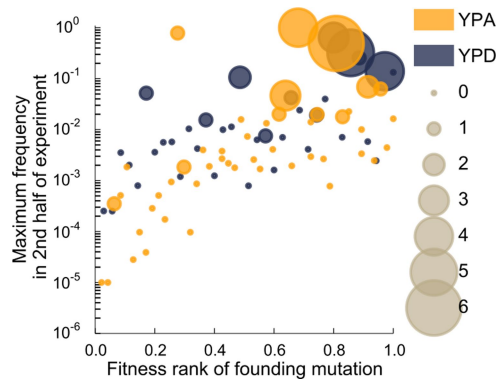
Extended Data Fig. 3 | Predictors of the success of lineages. The size of each dot denotes the number of later beneficial mutations that occur in the founding clonal background of a lineage (in the first half of the experiment).
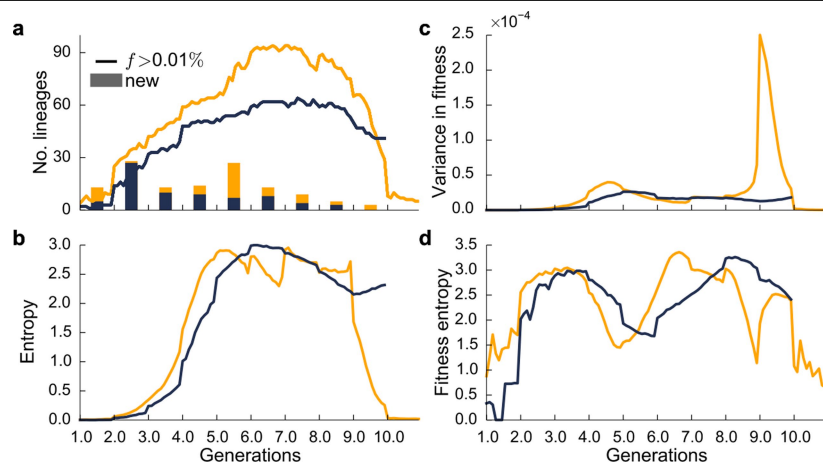
Extended Data Fig. 4 | Genetic variation over time. a, Total number of lineages above a threshold frequency (0.01%) over time. Bars denote the number of new lineages that arise in each 100-generation interval. **b**, Genetic diversity within each population over time, as measured by entropy (Supplementary Information section 6.4). **c**, Variance in fitness over time. **d**, Fitness diversity within each population over time, as measured by fitness entropy. Fitness entropy quantifies how fitness variance is distributed among lineages (Supplementary Information section 6.4).

nature research

Corresponding author(s): Michael M Desai

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	No software was used for data collection.
Data analysis	Freely available third-party software trimmomatic v0.35 and breseq v0.27.1b as well as custom scripts were used for data analysis. All custom algorithms are described in the SI Appendix; the source code is available on GitHub (https://github. com/icvijovic/lineage-tracking).

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

Raw sequencing reads have been deposited in the NCBI BioProject database under accession number PRJNA559526. All associated metadata, as well as the source code for the sequencing pipeline, downstream analyses, and figure generation, are available at GitHub (https://github. com/icvijovic/lineage-tracking).

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   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see 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	The sample size of a single population per evolutionary condition is sufficient, because our method makes it possible to detect a large number of mutations in each condition, allowing for statistical comparisons between their evolutionary dynamics and fates.
Data exclusions	The last 100 generations of evolution in the YPD environment were excluded from further analysis, because of a failure of the re-barcoding procedure, as described in the SI Appendix. The exclusion criteria were not pre-established. No other data were excluded.
Replication	No additional replication was performed.
Randomization	The clonal ancestors for the two evolution experiments were selected randomly from a single isogenic population.
Blinding	Blinding to the experimental environment was not relevant to our study because the two populations were subjected to identical data collection and data analysis procedures.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Obtaining unique materials	All unique materials are readily available from the authors.