nature portfolio

Corresponding author(s): G. Ozan Bozdag, Peter J. Yunker, William C. Ratcliff

Last updated by author(s): Mar 17, 2023

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the 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
☐	☒ A statement on 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 statistical parameters 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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Multicellular (or cluster) size, aspect ratio, and volume fraction data were collected using a Nikon Eclipse Ti inverted microscope (NIS-Elements version 4.30.01). 3D structure of the macroscopic snowflake yeast was resolved using a Serial Bulk Faced Scanning Electron Microscope (imaged on a Zeiss Sigma VP 3View system). Mechanical testing (presented in Fig3c) data were collected using a Puima Chiaro nanoindenter (Optics11, 19.5 um spherical glass probe; for ancestral clusters) or a Zwick Roell Universal Testing Machine (UTM; for macroscopic clusters). Whole genomes were sequenced using an Illumina HiSeq 2500 platform.
Data analysis	bwa-mem (BWA version 0.1.17), GATK4 HaplotypeCaller (v4.0.3.0), VCFTOOLS (version 0.1.17), SnpEff (v4.3T), Trimmomatic (v0.39), BCFtools-isec (v1.10), and Linux bash shell were used in sequence analysis and data parsing. Prism GraphPad version 9.5.0 for Mac OS was used for statistical analysis. Prism GraphPad version 9.5.0 for Mac OS was used for regression analyses. NIS-Elements version 4.30.01 and ImageJ (2.3.0) built-in macro functions were used for image analysis (measuring the cross-sectional area of snowflake yeast and aspect ratio of single cells; Fig 1e&f and Fig2d). An online tool (at yeast genome.org) was used to perform GO-term analysis (presented in Fig 4d). A custom Python code was used to calculate the average multicellular size presented in Fig 1e - we provide code used to generate plots in a public repository (GitHub). Mathematica 12 and Rhino6 were used to analyze Serial Bulk Faced Scanning Electron Microscope images. A new MATLAB code was provided as Supplementary File 1.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All source data, raw data, and code are available at https://github.com/ozanbozdag/de_novo_evolution_of_macroscopic_multicellularity. Microscopy images are archived at Ratcliff Lab Dropbox account and are available upon request.

Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

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

Study description	We conducted an evolution experiment over 600 days to select microscopic snowflake yeast clusters for large multicellular size. To investigate the effects of oxygen availability, we evolved these yeast under conditions with relatively low oxygen levels (PM1-5 and PO1-5 evolved in 25% PAL oxygen) or under anaerobic conditions (PA1-5), making a total of 3 x 5 = 15 replicate populations. All five replicate populations that were not constrained by limiting oxygen evolved macroscopic multicellular size, while populations evolving under a quarter of modern-day oxygen levels remained microscopic. This study examines this transition from microscopic to macroscopic size through phenotypic analysis, 3D microscopy, biophysical testing, and genome sequencing.
Research sample	We employed diploid Saccharomyces cerevisiae as our model organism. To initiate our experiment with a multicellular phenotype, we deleted the ACE2 transcription factor from the yeast genome. This organism is a well-established eukaryotic model system, making it genetically and phenotypically accessible.
Sampling strategy	We began our evolution experiment with five replicate populations and three metabolic treatment groups. To analyze evolutionary trends, we simply studied whole population samples. In that, to measure multicellular size we collected data from 1150 multicellular clusters per population, and to measure cellular aspect ratio, we collected data from an average of 453 individual cells per population, across 12 time points x 5 replicate populations. For genome sequencing, we isolated a single clonal snowflake yeast cluster from these populations and sequenced those isolates.
Data collection	GOB performed the long-term evolution experiment and created a -80 archive of each population at intervals of approximately 10-15 days. These populations can be revived for further examination. The data were collected by GOB, AZ, PK, TCD, DTL, and AJB without any manipulation by the experimenter. Genome sequencing data was generated using the Illumina HiSeq-2500 platform at the IBB sequencing core facility located at GA-Tech. Cluster size and aspect ratio data were collected using a Nikon Eclipse Ti inverted microscope at the Ratcliff Lab. Mechanical testing using UTM was performed at the Yunker Lab. AZ collected SEM images at the Core Facilities at the Carl R. Woese Institute for Genomic Biology for the SEM imaging. Microscopy images are available upon request (archived in our institute's Dropbox account).
Timing and spatial scale	Long-term evolution experiment (corresponding to the first 600 days) was performed between 2017-2018. Multicellular size, aspect ratio, SEM images, mechanical testing, and genomics data collection and analysis were performed in 2019-2021. Revision experiments (Fig1a&f, 4f, Extended data figures 2, 4, 8, 9, and 10) performed in 2022.
Data exclusions	No data were excluded from the analysis. All source data used to generate figures and all raw data used to calculate average values are shared in GOB's GitHub account.
Reproducibility	All five populations (PA1-5) evolved macroscopic size mainly driven by cell-level changes, such as cellular elongation. This parallel trend (5/5) not only demonstrates the reproducibility of our data but also provides confidence that, if these populations were to be re-evolved, the same outcome (macroscopic size) would arise. Mechanical testing was performed on 10 replicate clusters across various time points, and no failed attempts were encountered in the measurement or replication of these analyses. We have kept a frozen record of all those strain isolates and populations in our -80 freezer, which are available for further analysis, such as replicating our results. Therefore, frozen samples can be revived to reproduce the data presented in the paper.
Randomization	To ensure masked sample collection, we assigned random labels to samples at all stages of data collection, including cluster size and aspect ratio measurements, generation time measurements, and genomic DNA library preparation.

Blinding	Blinding was not necessary for our study as we did not need to make any decisions based on phenotypic analysis. We collected multicellular size and aspect ratio data by sampling a significant portion of the clusters from the original populations.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Other organisms are used: Saccharomyces cerevisiae - a non-pathogenic microbial eukaryote. Stain: Y55.
Wild animals	n/a
Field-collected samples	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.