nature portfolio

Corresponding author(s):	László G. Nagy
Last updated by author(s):	2022-11015

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 <u>Editorial Policies</u> and the <u>Editorial Policy Checklist</u>.

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

_				
C-	1	۱ + ۱	ıct	ics
			151	11 >

n/a	Confirmed
	$oxed{\boxtimes}$ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
\boxtimes	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.
\boxtimes	A description of all covariates tested
\boxtimes	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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
\boxtimes	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
\boxtimes	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
\boxtimes	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
	Our web collection on <u>statistics for biologists</u> contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection InterPro Scan (v.5.46-81.0), MAFFT, Tr

InterPro Scan (v.5.46-81.0), MAFFT, TrimAL, RAxML (v8.2.12), FigTree, IQ-Tree v2.0.3, MMseqs2 v13.45111, CAZY database, JGI Mycocosm

Data analysis COMPARE pipelir

COMPARE pipeline (available at https://github.com/zsmerenyi/compaRe), R packages used for analysis (topGO, phytools, ape, tidyr, phangorn, stat and tidyverse)

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 <u>data availability statement</u>. 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 <u>policy</u>

Accession codes, unique identifiers, or web links for publicly available datasets

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender

Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

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

Field-specific reporting

Please select the one belo	w that is the best fit for your research.	If you are not sure, read the appropriate sections before making your selection. $\label{eq:constraint}$
Life sciences	Rehavioural & social sciences	Fcological evolutionary & environmental sciences

For a reference copy of the document with all sections, see <u>nature.com/documents/nr-reporting-summary-flat.pdf</u>

Ecological, evolutionary & environmental sciences study design

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

Study description	In this paper we focused on the phylogenomic comparisons of Holomycota genomes with a diverse Amorphean dataset.
Research sample	published whole genomes of 123 species
Sampling strategy	Sampling of species based on taxonomic classification. We intend to sample all known (until 2022) phyla of the Holomycota clade.
Data collection	JGI mycocosm database
Timing and spatial scale	These are previously published genomes, therefore no data collection date is application. Versions of each species used in this study are mentioned in Supplementary Data 1.
Data exclusions	not applicable for this study
Reproducibility	To ensure reproducibility, the codes used in the study are made publicly available on github profile, mentioned in "code availability section of the manuscript.
Randomization	N/A
Blinding	N/A

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.

	=	÷
		ر
	Ω	ر
	7	+
	\subseteq	
	Ξ	₹
	a	Š
	٠,	1
	7	٦
	⊱	≺
	C	ע
	Ξ	٦
	⊆	1
	c	₹
	\leq	<u> </u>
	Ħ	
	\mathcal{C})
	`	
		٦
	n	C
	7	٦
	~	?
	C.)
		٦
	_	+
	Ε	₹
	H	?
(C	5
		ī
	U	า
	C	Ξ
	$\stackrel{\square}{=}$	=
	=	3
	=	3
	=	- 3 3
		3
	2 = 2	333
	2	3300

≥	
٥	
⇉	
S	
◺	
Ċ	
K	
=	₹

iviateriais & experimental systems		ivietnous	
n/a	Involved in the study	n/a	Involved in the study
\boxtimes	Antibodies	\boxtimes	ChIP-seq
\boxtimes	Eukaryotic cell lines	\times	Flow cytometry
\boxtimes	Palaeontology and archaeology	\boxtimes	MRI-based neuroimaging
\boxtimes	Animals and other organisms		
\boxtimes	Clinical data		
\boxtimes	Dual use research of concern		