

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
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☐ | ☒ | 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
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ | 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 Data (sequences) were downloaded from online databases, with no specific software required.

Data analysis FastQC (v0.11.19) and Trimmomatic (v0.39) were used to curate the sequences.
GeneiousPrime (build 2020-11-05) was used to visualise the sequences.
RStudio v1.4.1103-4 was used to process the sequences, following the DADA2 SOP (<https://benjjneb.github.io/dada2/tutorial.html>).

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](#)

Run accession numbers for each sample are provided so that the original raw FASTQ files can be accessed online, where available. The six 16S rRNA gene libraries created (V1-V2, V3, V4, V5-V6, V7-V8, V4_p) will be made freely available in the near future (once unpublished sequences provided by various researchers have

been published), as accessible and easily utilized phyloseq objects in R, alongside cnidarian specific RDP annotation files.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research.](#)

Reporting on sex and gender	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

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 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	A systematic review was conducted into the previously published cnidarian 16S rRNA gene sequences. All sequences were reprocessed using the same pipeline to allow for cross comparison of data from older studies (published as OTUs) and newer studies (published as ASVs).
Research sample	16S rRNA gene amplicon sequences were downloaded from online databases (primarily Genbank), or acquired directly from authors.
Sampling strategy	Sampling was determined by the health status of the cnidarian sample, the amplified hypervariable region, and the sequencing chemistry.
Data collection	Sequences were downloaded, curated, and then processed.
Timing and spatial scale	Earliest included sequences were published in 2011, the most recent in 2021.
Data exclusions	During the DADA2 pipeline, sequencing reads that did not have sufficient quality, or read length, were removed for quality control.
Reproducibility	All analysis were conducted twice, using both rarefied and unrarefied data sets. Further, additional analysis were conducted only on a subset of samples, to reduce the potential influence of different methodologies for sampling/storage/processing.
Randomization	Groups were dependent on critical metadata, including host phylogeny ie. Cnidaria or Porifera, sequencing method ie. Mi-Seq, and ocean bodies ie, Pacific or Atlantic.
Blinding	Numerous factors resulted in a primarily blind study (barring some specific analyses) including; there were over 16,000 samples, each sample had an alphanumeric code, there was extensive metadata associated with each sample, and the samples were acquired from online databases and not personally collected.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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