

Corresponding author(s): Michael S. RappéLast updated by author(s): Nov 14, 2025

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

No software was used for data collection.

Data analysis

The code used for analyses is available on GitHub (<https://github.com/kcfree/SAR11-genomes-from-the-tropical-Pacific>) linked to a Zenodo repository (<https://doi.org/10.5281/zenodo.17527686>). The URL <https://merenlab.org/data/sar11-phylogenomics/> serves a reproducible bioinformatics workflow with all the genomes used in our study to reproduce the phylogenomic tree or extend it with new genomes. Additionally, we used the following software/tools/algorithms/packages:

QIIME2 v2019.4.0, DADA2 v1.14, Silva rRNA v132, SPAdes 3.12.0 & SPAdes 3.13.0, Trim Galore! (<https://github.com/FelixKrueger/TrimGalore>), Unicycler v0.4.8, Bowtie2, SAMtools, anvi'o v8.0, Integrative Genomics Viewer (IGV) v2.8.4, Tablet v1.21.02.08, CheckM v1.1.2, GTDB-Tk, Muscle v3.8.1551, trimAL 1.3, IQ-Tree v2.1.2, ModelFinder, FigTree v1.4.4, phytools v1.2-0, R v4.2.1, FastANI v1.34, FastTree v2.1.10, PhyloRank v0.1.11 (<https://github.com/dparks1134/PhyloRank>), SRA toolkit v2.10.1, custom R code available at (<https://merenlab.org/data/sar11-saavs/>), vegan package v2.6-4 in R, ggplot2 v2_3.5.1, and plotly v4.10.4, dplyr v1.1.4 stringr, v1.5.2, maps v.3.4.3, sf 1.0-22, marmap v1.0.12, mapdata v2.3.1, ggspatial v1.1.10

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 data are available in the main text or the supplementary materials. The assembled sequence data for genomes reported here including raw sequencing reads are available at be found on NCBI associated with BioProject PRJNA1170004XXXXXX, and also available at FigShare at <https://doi.org/10.6084/m9.figshare.28087454.v1> in addition to full length 16S rRNA gene sequences (10.6084/m9.figshare.29983285).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Not applicable
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

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](https://www.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 collected seawater from Kāneʻohe Bay, Oʻahu, HI, USA to use for the cultivation of marine microbial isolates and sequenced genomes from relevant isolates of interest. We additionally used data our group generated published elsewhere to contextualize the data generated as a part of this cultivation strategy as well as available publicly deposited sequence data.
Research sample	We sampled surface seawater from approximately 2 meters depth from within and outside of Kāneʻohe Bay, Oʻahu, HI, USA.
Sampling strategy	We collected seawater at a single time point from two locations and further sampling strategy was not employed.
Data collection	Not applicable.
Timing and spatial scale	Not applicable.
Data exclusions	Not applicable.
Reproducibility	Not applicable.
Randomization	Not applicable.
Blinding	Not applicable.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	The samples were collected in July, there was no rain, temperatures of the air and water were standard for the time of year. We were working off of a small vessel on the water for sample collection.
Location	The two samples were collected from 2 meters below the sea surface at site SB within Kāneʻohe Bay (N 21° 26.181', W 157° 46.642) and site STO1 outside of Kāneʻohe Bay (N 21° 28.974, W 157° 45.978').
Access & import/export	This is not applicable to our study.
Disturbance	This is not applicable to our study.

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
☒	☐ Plants

Methods

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

Plants

Seed stocks	Not applicable to our study.
Novel plant genotypes	Not applicable to our study.
Authentication	Not applicable to our study.