




OPEN LETTER

**REVISED** **The Aquatic Symbiosis Genomics Project: probing the evolution of symbiosis across the Tree of Life**

[version 2; peer review: 4 approved, 1 approved with reservations]


Victoria McKenna<sup>1</sup>, John M. Archibald<sup>2</sup>, Roxanne Beinart<sup>3</sup>, Michael N. Dawson<sup>4</sup>, Ute Hentschel <sup>5</sup>, Patrick J. Keeling<sup>6</sup>, Jose V. Lopez <sup>7</sup>, José M. Martín-Durán<sup>8</sup>, Jillian M. Petersen<sup>9</sup>, Julia D. Sigwart <sup>10</sup>, Oleg Simakov<sup>11</sup>, Kelly R. Sutherland <sup>12</sup>, Michael Sweet<sup>13</sup>, Nicholas J. Talbot <sup>14</sup>, Anne W. Thompson <sup>15</sup>, Sara Bender<sup>16</sup>, Peter W. Harrison <sup>17</sup>, Jeena Rajan<sup>17</sup>, Guy Cochrane <sup>17</sup>, Matthew Berriman <sup>1</sup>, Mara K.N. Lawniczak <sup>1</sup>, Mark Blaxter <sup>1</sup>







- <sup>1</sup>Tree of Life, Wellcome Sanger Institute, Cambridge, CB10 1SA, UK
- <sup>2</sup>Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, B3H 4R2, Canada
- <sup>3</sup>Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island, 02882, USA
- <sup>4</sup>Life & Environmental Sciences, University of California, Merced, Merced, California, 95343, USA
- <sup>5</sup>Marine Symbioses Research Unit, GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany
- <sup>6</sup>Department of Botany, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada
- <sup>7</sup>Halmos College of Arts and Sciences, Nova Southeastern University, Dania Beach, Florida, 33004, USA
- <sup>8</sup>School of Biological and Chemical Sciences, Queen Mary University of London, London, E1 4NA, UK
- <sup>9</sup>Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, A-1090, Austria
- <sup>10</sup>Marine Zoology Department, Senckenberg Research Institute, Frankfurt, 60325, Germany
- <sup>11</sup>Department for Neurosciences and Developmental Biology, University of Vienna, Vienna, 1010, Austria
- <sup>12</sup>Department of Biology, University of Oregon, Eugene, Oregon, 97403-1210, USA
- <sup>13</sup>Aquatic Research Facility, Environmental Sustainability Research Centre, University of Derby, Derby, DE22 1GB, UK
- <sup>14</sup>The Sainsbury Laboratory, Norwich, NR4 7UH, UK
- <sup>15</sup>Department of Biology, Portland State University, Portland, Oregon, 97201, USA
- <sup>16</sup>Gordon and Betty Moore Foundation, Palo Alto, California, 94304, USA
- <sup>17</sup>European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, CB10 1SD, UK

**v2** **First published:** 06 Oct 2021, **6**:254  
<https://doi.org/10.12688/wellcomeopenres.17222.1>  
**Latest published:** 10 Jun 2024, **6**:254  
<https://doi.org/10.12688/wellcomeopenres.17222.2>

**Abstract**  
We present the Aquatic Symbiosis Genomics Project, a global collaboration to generate high quality genome sequences for a wide range of eukaryotes and their microbial symbionts. Launched under the Symbiosis in Aquatic Systems Initiative of the Gordon and Betty Moore Foundation, the ASG Project brings together researchers from across the globe who hope to use these reference genomes to augment and extend their analyses of the dynamics, mechanisms and environmental importance of symbioses. Applying large-scale, high-throughput sequencing and assembly technologies, the ASG

**Open Peer Review**

**Approval Status** 

	1	2	3	4	5
<b>version 2</b>					
(revision)					
10 Jun 2024			<a href="#">view</a>	<a href="#">view</a>	<a href="#">view</a>
<b>version 1</b>					
06 Oct 2021	<a href="#">view</a>	<a href="#">view</a>			
<b>1. James Davis Reimer</b>  , University of the Ryukyus, Nishihara, Japan					





collaboration will assemble and annotate the genomes of 500 symbiotic organisms – both the “hosts” and the microbial symbionts with which they associate. These data will be released openly to benefit all who work on symbioses, from conservation geneticists to those interested in the origin of the eukaryotic cell.

## Keywords

Symbiosis, Marine, Freshwater, Genome Sequencing, Collaboration, Open Science



This article is included in the [Tree of Life gateway](#).

2. **James Bernot** , Smithsonian National Museum of Natural History, Washington, USA
3. **Alison L. Gould** , Temple University, Philadelphia, USA
4. **Raúl A. González-Pech** , The Pennsylvania State University - University Park Campus, University Park, USA
5. **Senjie Lin** , University of Connecticut, Storrs, USA

Any reports and responses or comments on the article can be found at the end of the article.

**Corresponding author:** Mark Blaxter ([mark.blaxter@sanger.ac.uk](mailto:mark.blaxter@sanger.ac.uk))

**Author roles:** **McKenna V:** Conceptualization, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Archibald JM:** Conceptualization, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Beinart R:** Conceptualization, Writing – Review & Editing; **Dawson MN:** Conceptualization, Writing – Review & Editing; **Hentschel U:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Keeling PJ:** Conceptualization, Writing – Review & Editing; **Lopez JV:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Martín-Durán JM:** Conceptualization, Writing – Review & Editing; **Petersen JM:** Conceptualization, Writing – Review & Editing; **Sigwart JD:** Conceptualization, Writing – Review & Editing; **Simakov O:** Conceptualization, Writing – Review & Editing; **Sutherland KR:** Conceptualization, Writing – Review & Editing; **Sweet M:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Talbot NJ:** Conceptualization, Writing – Review & Editing; **Thompson AW:** Conceptualization, Writing – Review & Editing; **Bender S:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Harrison PW:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Rajan J:** Conceptualization, Writing – Review & Editing; **Cochrane G:** Conceptualization, Funding Acquisition, Writing – Original Draft Preparation, Writing – Review & Editing; **Berriman M:** Conceptualization, Funding Acquisition, Writing – Original Draft Preparation, Writing – Review & Editing; **Lawniczak MKN:** Conceptualization, Funding Acquisition, Writing – Original Draft Preparation, Writing – Review & Editing; **Blaxter M:** Conceptualization, Funding Acquisition, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was supported by Wellcome [206194] and the Gordon and Betty Moore Foundation [GBMF8897]. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2024 McKenna V *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** McKenna V, Archibald JM, Beinart R *et al.* **The Aquatic Symbiosis Genomics Project: probing the evolution of symbiosis across the Tree of Life [version 2; peer review: 4 approved, 1 approved with reservations]** Wellcome Open Research 2024, 6:254 <https://doi.org/10.12688/wellcomeopenres.17222.2>

**First published:** 06 Oct 2021, 6:254 <https://doi.org/10.12688/wellcomeopenres.17222.1>

**REVISED Amendments from Version 1**

The text of the article has been edited in response to reviewer comments. For example:

- In the abstract and elsewhere in the article: we have changed the term “symbiosis” to the plural “symbioses” where appropriate.
- In the “Genomics of symbiosis” section, we have added an additional reference as recommended.
- We have rephrased slightly and added a reference for the text concerning “create biodiversity hotspots which house upwards of 25% of all described ocean species”.
- In the legend for [Figure 1](#) we have added an explanation for using red and green fonts to indicate the taxa with primary plastids that subsequently spread to other taxa.
- “In the section “The ASG project will transform symbiosis research”, the third paragraph here (starting with “The hub partners...”) needs elaboration.” We have added a hyperlink to a description of the hub partners to clarify the intention here.
- In [Table 1](#) we have replaced the genus name “Symbiodinium” with the family Symbiodiniaceae referring to the family.
- We have rephrased and provided references for the text “Many of the fish that throng around coral reefs are open spawners, ...”.
- Also in this paragraph, we have corrected the word provides to ‘provide’ in “Much like a healthy reef, our hope is that the high-quality genomes we produce will generate the chatter that attracts new researchers and provides a foundation for growth of fundamental ...”

**Any further responses from the reviewers can be found at the end of the article**

**Disclaimer**

The views expressed in this article are those of the author(s). Publication in Wellcome Open Research does not imply endorsement by Wellcome.

**The genomics of symbiosis**

Symbiosis, the living together of distinct organisms ([Archibald, 2014](#); [Oulhen et al., 2016](#)), describes a spectrum of relationships from mutualistic to parasitic, and from obligate to temporary. Symbiosis has been and is fundamental to the evolution of life on Earth, from the deep origins of the eukaryotic cell and photosynthetic eukaryotes, through to the recent emergence of new partnerships. The power of symbiosis arises from the ability of the joint organism to draw from the independent, billion-year evolutionary histories of both partners. Symbiosis is a fact of life – it has arisen many, many times and new symbioses are constantly evolving ([Figure 1](#)). In this era of rapid climate change and biodiversity loss, many key-stone symbiotic systems are threatened, and their loss imperils the ecosystems they support.

Well-known mutualist symbioses permit colonisation of otherwise inaccessible habitats, are critical to ecosystem functioning, and support marine and freshwater diversity. For example, coral reefs, built through a photosymbiotic association between cnidarians and dinoflagellate algae

([LaJeunesse et al., 2018](#); [Weis, 2019](#)), create biodiversity hotspots which house upwards of 25% of all described ocean species ([Fisher et al., 2015](#)). The dominant animals colonising deep-sea hydrothermal vents are nutritionally dependent on chemosymbiotic associations with bacteria ([Roeselers & Newton, 2012](#)), allowing them to thrive in the food-limited dark ocean. For these symbioses, the biological fitness consequences are largely understood, but in many less well-known symbioses, such as those between sponges and their bacterial collaborators, or partnerships in the diverse world of single celled eukaryotes, the basis of the relationships are not known in any detail.

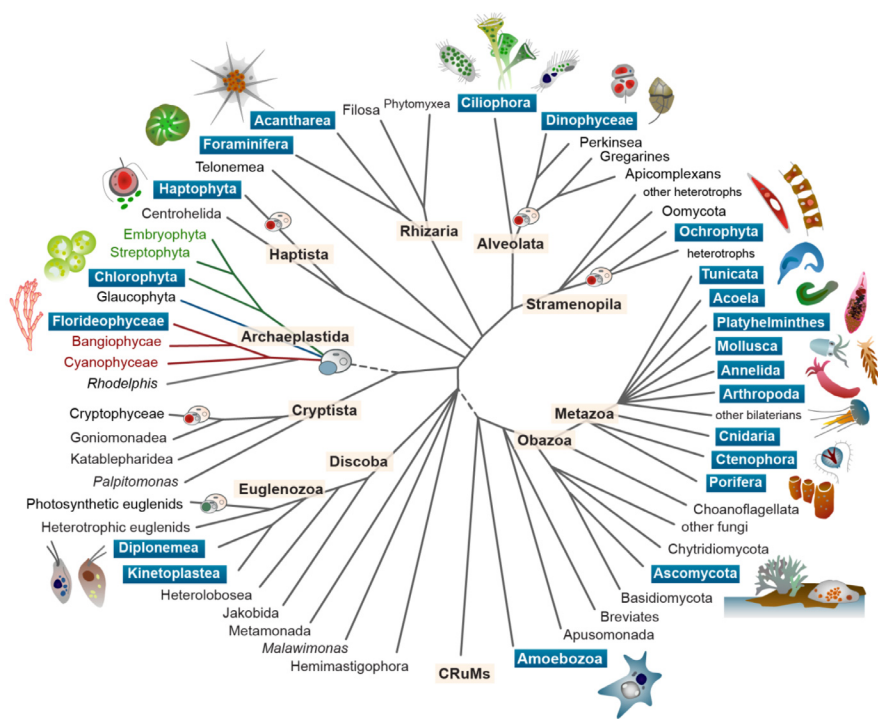
**The aquatic symbiosis genomics project will transform symbiosis research**

The Gordon and Betty Moore Foundation has created a major funding initiative focused on investigating the biology of symbiosis in marine and freshwater ecosystems (see [Symbiosis in Aquatic Systems Initiative](#)). To support this global initiative, the Aquatic Symbiosis Genomics project (ASG; see [Aquatic Symbiosis Genomics Project – Wellcome Sanger Institute](#)) plans to generate high-quality genome sequences from a wide range of symbiotic systems. Our focus is on symbioses involving at least one eukaryotic partner, and where there is likely to be co-evolving interplay between the species involved.

Like a symbiotic organism, the ASG project is more than the simple sum of its parts. ASG will merge the decades of ecological, evolutionary, taxonomic, and experimental expertise of researchers from diverse backgrounds with the decades of genomics experience of the Wellcome Sanger Institute. ASG works on a hub and spokes model, where communities of researchers nucleated on specific questions and/or species systems have come together as hubs to propose sets of taxa for sequencing ([Table 1](#)). These (currently) total ~450 distinct symbiotic organisms from the open ocean, the deep sea, coastal, littoral, and freshwater ecosystems, which are expected to include over 1000 nominal species of hosts and symbionts. The ASG target list includes species representing many phyla of animals, protists, algae and fungi, and encompasses ancient and recently-evolved partnerships.

The [hub partners](#) have defined the major scientific questions they wish to explore, and will source and identify specimens that will deliver answers. ASG follows an ethical code of sampling practice, avoiding overcollection and respecting local and international laws and protocols, especially as ASG will be sampling from endangered ecosystems and in some cases endangered species. The project participants are fully committed to the Convention on Biological Diversity Nagoya Protocols on Access and Benefit Sharing, and only samples where express permission has been obtained will be sourced and sequenced. Samples may come from the wild, from mesocosms and aquaria, from explant lab cultures or from culture collections.

Genome sequencing and assembly will be delivered by the Tree of Life programme at the [Sanger Institute](#) using pipelines being developed for the [Darwin Tree of Life](#) and other major biodiversity genomics projects. Genomes will be assembled,



**Figure 1. The phylogenetic diversity of eukaryotic symbioses.** Symbiotic taxa, and Aquatic Symbiosis Genomics target species, are found across the diversity of the eukaryotic Tree of Life. Taxa highlighted with blue boxes include ASG targets. Within the tree, the small cartoons indicate the major event of plastid acquisition through symbiosis with a cyanobacterium (in the Archaeplastida; blue cell engulfed) and the several events of secondary and tertiary plastid acquisition in other lineages. The taxa containing primary plastids are shown in green and red. Illustration by John Archibald and Mark Blaxter.

**Table 1. Aquatic symbiosis genomics project hubs.**

Lead researcher*	Project Title (short)	Major taxa represented	
		Hosts	Symbionts
Archibald	New symbioses in single-celled eukaryotes	Amoebozoa, Dinophyceae, Diplonemea (Euglenozoa), Haptophyta, Ochrophyta	Bacteria, Kinetoplastea, Ochrophyta
Beinart, Petersen, Sigwart	Molluscan symbioses	Mollusca	Arthropoda, Bacteria, Chlorophyta, Cnidaria, Dinophyceae, Platyhelminthes, Florideophyceae
Dawson, Sutherland, Thompson	Pelagic symbioses	Acoela, Ctenophora, Cnidaria, Tunicata	Bacteria, Chlorophyta, Dinophyceae, other Alveolata
Hentschel	Sponge symbioses	Porifera	Bacteria, Archaea, Viruses, Symbiodiniaceae (Dinophyceae) and others
Keeling	Symbiosis in ciliates	Ciliophora	Archaea, Bacteria, Chlorophyta, Ciliophora, Dinophyceae
Lopez	Metazoan photosymbioses	Acoela, Cnidaria, Mollusca, Porifera, Tunicata	Bacteria, Chlorophyta, Cnidaria, Dinophyceae, Haptophyta, Myxozoa
Martín-Durán	Annelid chemosymbioses	Annelida	Bacteria, Archaea
Simakov	Cephalopod symbioses	Mollusca	Bacteria, Archaea
Sweet	Coral symbioses	Cnidaria	Symbiodiniaceae (Dinophyceae)
Talbot	Marine lichens	Fungi	Bacteria, Chlorophyta, ascomycete Fungi, Ochrophyta

\* see author list for affiliations.

annotated and released openly through the European Bioinformatics Institute (EMBL-EBI).

### Sequencing symbionts: from sample to openly accessible genome assembly

Each ASG Hub (Table 1) has defined a set of taxa that it will sample for sequencing. We will sequence from single eukaryotic host specimens or clonal cultures rather than bulk samples whenever possible. While this can limit the mass of DNA and RNA available for sequencing, it has the very strong benefit of reducing allelic sequence complexity and enabling assembly. Importantly, we do not require that the symbiotic partners are separated before sequencing, as we will separate the host and symbiont genomes bioinformatically during assembly (Challis *et al.*, 2020).

Each sample is formally identified and associated with rich metadata describing its collection location and other environmental features. We collate and validate these metadata through the [COPO biodiversity data brokering system](#). Samples are shipped to the Sanger Institute for long DNA and RNA extraction and sequencing, with particular focus on low-input methods. We are generating a combination of long read and long range genomic data. For long reads we primarily use the Pacific Biosciences Sequel IIe circular consensus sequencing approach to generate high fidelity (HiFi) reads in the 15 to 20 kilobase range, and include Oxford Nanopore Technologies long reads where needed. For long range data we use chromatin conformation capture sequencing (known as Hi-C). These long range data generate important information that link sequences within chromosomes and organelles in the multi-kilobase to megabase range and will allow us to disentangle genomes from different species. The joint transcriptome of the symbioses will be sampled using RNA-Seq, both on Illumina short read and Pacific Biosciences long read platforms.

We have strong expectations about what we should find in the sequence data, and what we should be assembling, but biology is full of exceptions and surprises and organisms taken from the wild are frequently found in association with other cobionts. Each symbiosis contains a community of genomes that can be viewed as a low complexity metagenome: the “host” genome and the genomes of its organelles (mitochondrion and in some cases plastid), the symbiont genome (which if it is eukaryotic contains one or more organellar genomes) and the genomes of other commensals and cobionts. We separate data into presumed organismal and organellar subsets and assemble each independently. First we identify taxonomically informative marker loci, such as small subunit ribosomal RNAs (organellar 12S, prokaryotic 16S and eukaryotic 18S), cytochrome oxidase I genes, and ribulose-1,5-bisphosphate carboxylase-oxygenase genes, in the HiFi reads and primary assembly. These tell us which taxa are likely to be present and thus which genomes we should expect to assemble. To separate the data we use intrinsic features (GC and tetranucleotide composition, read coverage, coding capacity), sequence similarity to known genomes, and Hi-C linkage information. Binning contigs

and their constituent reads into distinct subsets facilitates complete assembly of each organismal and organellar genome (Challis *et al.*, 2020; Kumar & Blaxter, 2011). We aim to automate this cobiont identification and binning process, as it will be of utility in analyses of all Tree of Life genomes: many specimens harbour parasitic and other cobionts. Given 25- to 30-fold genome coverage in HiFi reads for each symbiont partner, we expect to generate primary assemblies with contig N50s in the multi-megabase range. The Hi-C data are used to scaffold these contigs into near-chromosomal pseudomolecules.

For each symbiotic system we will then curate the assemblies to improve accuracy (Howe *et al.*, 2021) with particular attention to correct scaffolding of nuclear chromosomes and circularisation of organellar and prokaryotic genomes, and identification of remaining complex and unresolvable repetitive regions (such as ribosomal RNA and centromeric repeats). We aim to achieve or exceed the latest Earth BioGenome Project (Lewin *et al.*, 2018) assembly standards. Curated assemblies and all raw data will be submitted to the European Nucleotide Archive (ENA) (Harrison *et al.*, 2021) and from there to the rest of the International Nucleotide Sequence Database Consortium for immediate open release. The genomes will be annotated using the RNA-Seq transcriptomic data binned by species, and the annotations released openly. We have developed an [ASG-specific data portal](#) that collates all of the data generated by the project and promotes analysis. The Aquatic Symbiosis Genomics project relies on engagement and support from the whole of the Tree of Life production genomics team and of many colleagues who are participants in the ten Hubs. Each symbiotic system will be the subject of an open access publication, a Genome Note, that credits the full team that generated the assemblies, from collectors to annotators (Threlfall & Blaxter, 2021).

### Building an aquatic symbiosis genomics community

The ASG project aims to generate a lasting resource in terms of the ~1000 genomes involved in ~500 symbiotic systems. To ensure this resource results in a flourishing ecosystem of postgenomic research, we are building community and expertise through a parallel programme of training and mentoring in genomics and bioinformatics. In collaboration with [Wellcome Connecting Science and The Carpentries](#), the ASG project will deliver intensive and extensive collaborative training and investigative informatic analysis of symbiont genomes, to build collective genomics and bioinformatics capacity in the symbiosis community. Training will include core informatics, coding, and reproducible science, as well as deeper analytical dives into co-evolving genomes, detailed genome annotation, and prediction of the metabolic underpinnings of symbiotic cooperation.

Just as reefs built by corals and their symbiotic algae allow an exuberant and diverse ecology to thrive, the ASG project will build a lasting genomic foundation for flourishing and diverse analyses of symbioses. Many bony coral reef fish species have a



pelagic early life history, their larvae spending their first weeks in the open ocean (Leis & McCormick, 2002). These may be recruited back to the reef because they can hear and smell it: the chatter generated by a healthy reef attracts, recruits, and builds the reef community (Gordon *et al.*, 2019). Much like a healthy reef, our hope is that the high-quality genomes we produce will generate the chatter that attracts new researchers and provides a foundation for growth of fundamental research on the nature of symbiosis and conservation of habitats where symbioses abound.

## Data availability

No data are associated with this article. ASG data will be released openly in the European Nucleotide Archive.

## Acknowledgements

We thank Jonathan Threlfall for assistance with manuscript editing. This research was funded in part by Wellcome [grant 206194]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

## References

- Archibald J: **One plus one equals one: symbiosis and the evolution of complex life.** Oxford University Press, USA. 2014.  
[Reference Source](#)
- Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit - interactive quality assessment of genome assemblies.** *G3 (Bethesda)*. 2020; **10**(4): 1361–74.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fisher R, O'Leary RA, Low-Choy S, *et al.*: **Species richness on coral reefs and the pursuit of convergent global estimates.** *Curr Biol*. 2015; **25**(4): 500–505.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Gordon TAC, Radford AN, Davidson IK, *et al.*: **Acoustic enrichment can enhance fish community development on degraded coral reef habitat.** *Nat Commun*. 2019; **10**(1): 5414.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Harrison PW, Ahamed A, Aslam R, *et al.*: **The European Nucleotide Archive in 2020.** *Nucleic Acids Res*. 2021; **49**(D1): D82–85.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience*. 2021; **10**(1): gaa153.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kumar S, Blaxter ML: **Simultaneous genome sequencing of symbionts and their hosts.** *Symbiosis*. 2011; **55**(3): 119–26.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lajeunesse TC, Parkinson JE, Gabrielson PW, *et al.*: **Systematic revision of symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts.** *Curr Biol*. 2018; **28**(16): 2570–2580.e6.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Leis JM, McCormick MI: **The biology, behavior, and ecology of the pelagic, larval stage of coral reef fishes.** In: *Coral Reef Fishes*. Elsevier, 2002; 171–199.  
[Publisher Full Text](#)
- Lewin HA, Robinson GE, Kress WJ, *et al.*: **Earth biogenome project: sequencing life for the future of life.** *Proc Natl Acad Sci U S A*. 2018; **115**(17): 4325–33.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Oulhen N, Schulz BJ, Carrier TJ: **English translation of Heinrich Anton de Bary's 1878 speech, 'Die Erscheinung Der Symbiose' ('De La Symbiose').** *Symbiosis*. 2016; **69**: 131–139.  
[Publisher Full Text](#)
- Roeselers G, Newton ILG: **On the evolutionary ecology of symbioses between chemosynthetic bacteria and bivalves.** *Appl Microbiol Biotechnol*. 2012; **94**(1): 1–10.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Threlfall J, Blaxter M: **Launching the Tree of Life Gateway [version 1; peer review: not peer reviewed].** *Wellcome Open Res*. 2021; **6**: 125.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Weis VM: **Cell biology of coral symbiosis: foundational study can inform solutions to the coral reef crisis.** *Integr Comp Biol*. 2019; **59**(4): 845–55.  
[PubMed Abstract](#) | [Publisher Full Text](#)

# Open Peer Review

Current Peer Review Status:



Version 2

Reviewer Report 27 May 2025

<https://doi.org/10.21956/wellcomeopenres.24685.r122604>

© 2025 Lin S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Senjie Lin 

Department of Marine Sciences, University of Connecticut, Storrs, Groton, USA

The Letter effectively captures the scope of the Aquatic Symbiosis Genome Project, the diversity of symbiotic systems involved, and the anticipated outcomes. My main suggestion is to include more detail on the expected range of data quality or completeness. For example, it would be helpful to note that for species where high-molecular-weight (HMW) DNA suitable for long-read sequencing is not available, only raw or fragmented data may be produced, whereas for others, near-complete chromosomal assemblies will be achieved. Additionally, offering guidance on how readers should interpret the results and suggestions for improving dataset quality would further enhance the manuscript's utility.

I also have a couple of specific comments:

- **Regarding Table 1:** While I understand that listing all organisms involved may not be feasible, I recommend including as complete a list of **phyla** as possible. I noticed that a euglenid host with bacterial endosymbionts and a dinoflagellate with algal endosymbionts appear to be missing from the hub led by Joe Lopez.
- **Data access:** Adding links to data repositories mentioned in the Letter would be highly beneficial for readers interested in accessing and exploring the datasets.

**Is the rationale for the Open Letter provided in sufficient detail?**

Yes

**Does the article adequately reference differing views and opinions?**

Yes

**Are all factual statements correct, and are statements and arguments made adequately supported by citations?**

Yes

**Is the Open Letter written in accessible language?**

Yes

**Where applicable, are recommendations and next steps explained clearly for others to follow?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Ecological genomics of algae and protists

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 27 May 2025

<https://doi.org/10.21956/wellcomeopenres.24685.r122598>

© 2025 González-Pech R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Raúl A. González-Pech** 

The Pennsylvania State University - University Park Campus, University Park, Pennsylvania, USA

This open letter presents the ambitious effort being done by the Aquatic Symbiosis Genomics Project to generate high-quality genomic data for about 500 systems. There is no doubt about the value of the resources this initiative is generating, which will advance the field of symbiosis to new frontiers.

I have personally been tracking this project over the years and have even been tangentially involved with the sequencing of a few samples by providing material and advising in the data generation. I congratulate the team for pushing such a huge enterprise and for their achievements thus far. While I do not have any major comments, there are two minor concerns I have that might be worth expanding on in the text:

1. Taking into consideration idiosyncratic genomic features of each of the systems for tailored data generation and bioinformatic analyses. For example, dinoflagellates (including coral symbionts) have a specific gene architecture (refer 1) that needs to be considered when doing gene prediction (refer 2). Because of this, people from ASG should consult with experts of each system to ensure optimal genomic data generation. I know this has been done to certain degree, but it could be taken further. A few words about this might be worth mentioning.
2. Some of the symbionts might have complex systematics and their taxonomy (e.g., species designation) might not be fully established. It would be useful to know how ASG is handling these cases.

More punctual suggestions:

- The "The genomics of symbiosis" section doesn't really touch on genomics. I suggest either changing the name of the section to something more generic (like "Background") or expanding on how genomics play an important role in understanding biology and evolution of symbiotic



systems.

- Following up on the previous point, while the authors mention a few examples of aquatic symbiotic systems, questions and directions that could be explored with the data being generated remain vague. I suggest, as previous reviewers have, mentioning a few more specific examples of topics that could be explored. A useful resource for this could be a review we recently published on marine holobionts (refer 2).
- Single celled eukaryotes in the last sentence of the genomics of symbiosis sections should have a hyphen (single-celled eukaryotes).
- Consider changing "Sequencing symbionts: from sample to openly accessible genome assembly" to "Sequencing symbionts: from sample to openly accessible genomic data".

Other than those minor comments, I endorse indexing.

## References

1. Shoguchi E, Shinzato C, Kawashima T, Gyoja F, et al.: Draft assembly of the Symbiodinium minutum nuclear genome reveals dinoflagellate gene structure. *Curr Biol*. 2013; **23** (15): 1399-408 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Chen Y, González-Pech R, Stephens T, Bhattacharya D, et al.: Evidence That Inconsistent Gene Prediction Can Mislead Analysis of Dinoflagellate Genomes. *Journal of Phycology*. 2020; **56** (1): 6-10 [Publisher Full Text](#)
3. González-Pech RA, Li VY, Garcia V, Boville E, et al.: The Evolution, Assembly, and Dynamics of Marine Holobionts. *Ann Rev Mar Sci*. 2024; **16**: 443-466 [PubMed Abstract](#) | [Publisher Full Text](#)

## Is the rationale for the Open Letter provided in sufficient detail?

Yes

## Does the article adequately reference differing views and opinions?

Partly

## Are all factual statements correct, and are statements and arguments made adequately supported by citations?

Partly

## Is the Open Letter written in accessible language?

Yes

## Where applicable, are recommendations and next steps explained clearly for others to follow?

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Symbiodiniaceae, coral symbiosis, coral holobiont, comparative genomics, evolutionary genomics

**I confirm that I have read this submission and believe that I have an appropriate level of**

**expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 27 May 2025

<https://doi.org/10.21956/wellcomeopenres.24685.r122603>

© 2025 Gould A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Alison L. Gould** 

Temple University, Philadelphia, Pennsylvania, USA

This article presents an exciting global collaboration to generate high quality reference genomes for representative aquatic symbioses across the tree of life. With respect to the questions above, I found the article adequately references differing views and opinions, all factual statements are correct, the statements and arguments made are adequately supported by citations, and it is written in accessible language.

I do think the rationale for the Open Letter is provided in sufficient detail, but your overall motivation for producing these reference genomes is somewhat lacking. You say this initiative will "Transform symbiosis research" in the header of one section but don't provide much support for how it will do so. This section instead provides details on the logistics and design of your hub. It would be helpful to include here some more details on how will the research community will benefit from having access to these genomic resources. In what ways is this going to "transform" symbiosis research?

Other than this comment, I have a few minor suggestions.

The figure legend could use more detail. I see that you updated the branch/text colors in the Archaeplastida to green and red but don't explain what the two different colors indicate. I suggest clarifying this point in the legend text.

There is some inconsistency in the tense used throughout, specifically in the Sequencing Symbionts section. The section begins as if this is work to be done and then switches to present tense in the binning sequences paragraph. I recommend choosing one tense and maintaining throughout.

**Is the rationale for the Open Letter provided in sufficient detail?**

Partly

**Does the article adequately reference differing views and opinions?**

Yes

**Are all factual statements correct, and are statements and arguments made adequately supported by citations?**

Yes

**Is the Open Letter written in accessible language?**

Yes

**Where applicable, are recommendations and next steps explained clearly for others to follow?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Microbial symbiosis, marine science

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

### Version 1

Reviewer Report 15 December 2021

<https://doi.org/10.21956/wellcomeopenres.19032.r47364>

© 2021 Bernot J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**James Bernot** 

Department of Invertebrate Zoology, Smithsonian National Museum of Natural History, Washington, DC, USA

This open letter presents a global genomics project focused on sequencing the genomes of 500 aquatic organisms comprising multiple lineages of protists, algae, fungi, and invertebrates, along with their symbionts. Collectively these hosts and symbionts encompass an enormous swath of the tree of life. It is therefore an ambitious and challenging project which nonetheless holds great promise, especially in light of its large, international team with diverse expertise in a variety of organismal systems, genomic techniques, and bioinformatic approaches. This project is of particular interest given that increased attention on microbiome research in recent years has revealed diverse interactions between host and symbionts. Clearly, genome biology is key to understanding the complexities of this interacting holobiome.

In this paper, the authors describe their target hosts and symbionts, which they group into 10 project hubs, each with one to three project leads as currently designed. The manuscript presents the planned sequencing methods, which combine Illumina short reads, PacBio and Nanopore long reads, and Hi-C chromatin conformation data – the best practices in modern genome sequencing. The authors state they will couple the genomic sequencing with short and long read RNA-Seq,

which will be crucial for annotating the genomes and holds promise for uncovering functional interactions between members of each host-symbiont community. The paper also briefly outlines the author's bioinformatic approach to separate and assemble the host-symbiont metagenomes, which has several notable strengths including the ability to assemble difficult-to-culture organisms and characterize previously unknown symbionts.

Like other large-scale genome sequencing initiatives, the major test of this project will be whether or not it can maintain the sustained, long-term effort needed to complete such an ambitious goal. If successful, this work will be of great interest to the biological science community, not only to those studying the target organisms, but also anyone interested in the microbiome, the evolution of symbioses, and the evolution of eukaryotic life itself.

I have a few minor comments the authors may want to consider:

- While the project is clearly exploratory by design, the paper could be improved by the addition of some research questions the team wants to explore or hypotheses they hope to test.
- The hubs each present their major host and symbiont taxa, but what approaches will be used to select other symbiont taxa to characterize and analyze in detail, especially in cases of diverse host-symbiont associations? While characterizing known, ecologically important symbionts appears to be the priority of the project, it could be valuable to establish objective thresholds for targeting other symbionts to focus efforts on (for example, perhaps the number of gDNA reads as a proxy for symbiont abundance, or the number of RNA-Seq reads aligning to a genome as evidence of transcriptomic activity). Some objective guidelines could help avoid biasing the project towards those symbionts with better characterized host interactions and increase the possibility of elucidating new interactions.
- Work on mycorrhizal fungi, coral and squid symbionts, and others has shown that host-symbiont interactions can vary drastically across environmental conditions – the same host-symbiont species pair can be commensal, mutualistic, or parasitic depending on the environment. While the proposed framework will build a valuable foundation for understanding host-symbiont genomics, a more holistic view of these interactions requires understanding how relationships may change in different ecological contexts. Have the authors considered exploring any of these relationships under different environmental conditions? The possibility of drastically different interactions under varying ecological conditions lends additional importance to the author's proposal to collect rich environmental metadata, especially because it may not be initially apparent which environmental variables are important to the host-symbiont interactions.  
Some examples of varying host symbiont-interactions that might be of interest: Bronstein (2001)<sup>1</sup> and Grman (2012)<sup>2</sup>
- The proposed PacBio gDNA and RNA-Seq may pose potential challenges. First, while sequencing costs continue to decrease, is it currently feasible to complete the proposed PacBio gDNA and RNA sequencing for each host, especially in the beginning of the project when these costs are still quite high (or is this more so an ideal the project aspires to)? Second, are the PacBio input requirements feasible for most of the proposed taxa? Perhaps the host taxa will have enough genetic material, but I can imagine challenges obtaining reasonable read depth for symbionts when the community is diverse and input material

limited. Have the authors made estimates of necessary read depth for symbiont genome and transcriptome assembly?

- Coordinated, large scale efforts such as those described here hold great promise for pushing science forward in major ways, but also pose a number of challenges such as accountability for project goals, timely reporting requirements, transparency in progress, and objective measures of success. Towards this end, I credit the authors for having already developed a data portal and status tracking system to increase transparency and expedite access to research data. Any additional efforts to expound on timelines, publication strategies, progress updates, etc. would be beneficial.

In summary, this is a large-scale genomics project with ambitious goals that holds great promise for improving our understanding of host-symbiont interactions across extremely diverse aquatic taxa. It is of broad interest to the scientific community, and I for one look forward to seeing future publications on the proposed work.

## References

1. Bronstein J: The exploitation of mutualisms. *Ecology Letters*. 2001; **4** (3): 277-287 [Publisher Full Text](#)
2. Grman E: Plant species differ in their ability to reduce allocation to non-beneficial arbuscular mycorrhizal fungi. *Ecology*. 2012; **93** (4): 711-8 [PubMed Abstract](#) | [Publisher Full Text](#)

## Is the rationale for the Open Letter provided in sufficient detail?

Yes

## Does the article adequately reference differing views and opinions?

Yes

## Are all factual statements correct, and are statements and arguments made adequately supported by citations?

Yes

## Is the Open Letter written in accessible language?

Yes

## Where applicable, are recommendations and next steps explained clearly for others to follow?

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Evolutionary biology, genomics, bioinformatics, phylogenetics, taxonomy

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 18 October 2021

<https://doi.org/10.21956/wellcomeopenres.19032.r46287>

© 2021 Reimer J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**James Davis Reimer**

Graduate School of Engineering and Science, University of the Ryukyus, Nishihara, Japan

### Overview:

A solid statement paper explaining the rationale and general outline of the ambitious Aquatic Symbiosis Genomics Project, this letter is well put together and provides a general overview of the framework and research to be done in this project. I have only some minor comments that need to be addressed. Some of my comments are scientific, while others are related to the writing style and grammar.

### Major concerns:

None.

### Minor comments:

1. Please note for the question "Does the article adequately reference differing views and opinions?" I have answered "no" but do not think that this question is relevant for this paper; our overall understanding of the importance of symbioses is quite clear.
2. Regarding the question "Are all factual statements correct, and are statements and arguments made adequately supported by citations?" I have answered partly, as there are one or two areas where references or additional references are needed (see below for details).
3. This comment is more at the discretion of the authors, but in the Abstract, the term "symbiosis" is used as a singular noun in two locations ("environmental importance of symbiosis" and "all who work on symbiosis") where I would instead use the plural "symbioses". I imagine this is a matter of style, and leave the final choice to the authors, but feel plural serves these sentences in question better. A similar comment can be made on page 5, in the sentence "...flourishing and diverse analyses of symbiosis."
4. In the "genomics of symbiosis" section, the authors reference Weis (2019) when discussing the photosymbiotic association between cnidarians and Symbiodiniaceae, but think there are both previous and wider (from the viewpoint of ecology or coral reef science) references available. Authors do not need to replace this reference, but I would add at least one more here, the reference chosen is focused more on cell biology.
5. Following the comment directly above, the second half of this sentence needs a reference. ("create biodiversity hotspots which house upwards of 25% of all described species in the oceans"). I am also not certain if hotspots is the best term, or "hotspot" describing the Coral



Triangle. I suppose if the authors also wish to emphasize biodiversity centers such as the Red Sea and the Caribbean then plural is OK here.

6. Figure 1 – why are the taxa Embryophyta and Streptophyta written in green? No explanation for this is given in the legend.
7. In the section “aquatic symbiosis genomics project will transform symbiosis research”, the third paragraph here (starting with “The hub partners...”) needs elaboration. I would at the least add in here active collaboration with local and regional collaborators, and also the deposit of specimens in appropriate museums or collections that have public access for all researchers.
8. Table 1: In two instances, the authors list “*Symbiodinium*”, but I think they are referring to Symbiodiniaceae and not just the genus (see Lajeunesse *et al.* 2018 Curr Biol).<sup>1</sup>
9. Page 5 (of the PDF): I am not an expert on fish, but is “open spawners” the best term here? “Many of the fish that throng around coral reefs are open spawners, ...”.
10. As well, in the sentence immediately following the one above, you state “They are recruited back to the reef because they can hear and smell it: ...”. Is this true or can this be said for all of the “many of the fish that throng around coral reefs”? You may need to qualify this sentence to some degree.
11. I would change the “provides” in this sentence “Much like a healthy reef, our hope is that the high quality genomes we produce will generate the chatter that attracts new researchers and provides a foundation ...” to “provide”.

## References

1. Lajeunesse T, Parkinson J, Gabrielson P, Jeong H, et al.: Systematic Revision of Symbiodiniaceae Highlights the Antiquity and Diversity of Coral Endosymbionts. *Current Biology*. 2018; **28** (16): 2570-2580.e6 [Publisher Full Text](#)

## Is the rationale for the Open Letter provided in sufficient detail?

Yes

## Does the article adequately reference differing views and opinions?

No

## Are all factual statements correct, and are statements and arguments made adequately supported by citations?

Partly

## Is the Open Letter written in accessible language?

Yes

## Where applicable, are recommendations and next steps explained clearly for others to

follow?

Yes

**Competing Interests:** I am collaborating in a large project with many other members with Michael Sweet.

**Reviewer Expertise:** Symbiodiniaceae, coral reefs, cnidarians, field and molecular ecology

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 08 May 2024

**Tree of Life Team Sanger**

**Response from the author team:** Thank you for the helpful comments. We have made the following changes to address your concerns:

- In the abstract and elsewhere in the article: we have changed the term “symbiosis” to the plural “symbioses” where appropriate.
- In the “**Genomics of symbiosis**” section, we have added an additional reference as recommended.
- We have rephrased slightly and added a reference for the text concerning “create biodiversity hotspots which house upwards of 25% of all described ocean species”.
- In the legend for Figure 1 we have added an explanation for using red and green fonts to indicate the taxa with primary plastids that subsequently spread to other taxa.
- “In the section “The ASG project will transform symbiosis research”, the third paragraph here (starting with “The hub partners...”) needs elaboration.” We have added a link to a description of the hub partners.
- In Table 1 we have replaced the genus name “Symbiodinium” with the family Symbiodiniaceae referring to the family.
- We have rephrased and provided references for the text “Many of the fish that throng around coral reefs are open spawners, ...”.
- Also in this paragraph, we have corrected the word provides to ‘provide’ in “Much like a healthy reef, our hope is that the high-quality genomes we produce will generate the chatter that attracts new researchers and provides a foundation for growth of fundamental ...”

**Competing Interests:** No competing interests were disclosed.