Author's Proof

Before checking your proof, please see the instructions below.

• Carefully read the entire proof and mark all corrections in the appropriate place, using the Adobe Reader commenting tools (Adobe Help).
• Provide your corrections in a single PDF file or post your comments in the Production Forum making sure to reference the relevant query/line number. Upload or post all your corrections directly in the Production Forum to avoid any comments being missed.
• We do not accept corrections in the form of edited manuscripts nor via email.
• Before you submit your corrections, please make sure that you have checked your proof carefully as once you approve it, you won’t be able to make any further corrections.
• To ensure the timely publication of your article, please submit the corrections within 48 hours. After submitting, do not email or query asking for confirmation of receipt.

Do you need help? Visit our Production Help Center for more information. If you can’t find an answer to your question, contact your Production team directly by posting in the Production Forum.

Quick Check-List

☐ Author names - Complete, accurate and consistent with your previous publications.
☐ Affiliations - Complete and accurate. Follow this style when applicable: Department, Institute, University, City, Country.
☐ Tables - Make sure our formatting style did not change the meaning/alignment of your Tables.
☐ Figures - Make sure we are using the latest versions.
☐ Funding and Acknowledgments - List all relevant funders and acknowledgments.
☐ Conflict of Interest - Ensure any relevant conflicts are declared.
☐ Supplementary files - Ensure the latest files are published and that no line numbers and tracked changes are visible. Also, the supplementary files should be cited in the article body text.
☐ Queries - Reply to all typesetters queries below.
☐ Content - Read all content carefully and ensure any necessary corrections are made.

Author Queries Form

<table>
<thead>
<tr>
<th>Query No.</th>
<th>Details Required</th>
<th>Author's Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Kindly check with edit made all author affiliation and its affiliation number is fine.</td>
<td></td>
</tr>
</tbody>
</table>
Consensus Guidelines for Advancing Coral Holobiont Genome and Specimen Voucher Deposition

Christian R. Voolstra1*, Kate M. Quigley2, Sarah W. Davies3, John Everett Parkinson4, Raquel S. Peixoto5, Manuel Aranda2, Andrew C. Baker4, Adam R. Barno6, Daniel J. Barshis7, Francesca Benzioni8, Victor Bonito9, David G. Boume3,9, Carol Buitrago-López6, Thom C. L. Bridge10,11, Cheong Xin Chan12, David J. Combes13, Jamie Crapps14,15, Jörg C. Frommlet16, Santiago Herrera17, Andrea M. Quattrini18, Till Röthig19, James D. Reimer20, Esther Rubio-Portillo21, David J. Suggett22, Helena Villela4, Maren Ziegler5 and Michael Sweet12

1 Department of Biology, University of Konstanz, Konstanz, Germany, 2 Australian Institute of Marine Science, Townsville, QLD, Australia, 3 Department of Biology, Boston University, Boston, MA, United States, 4 Department of Integrative Biology, University of South Florida, Tampa, FL, United States, 5 Red Sea Research Center, Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, 6 Department of Marine Biology and Ecology, Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL, United States, 7 Department of Biological Sciences, Old Dominion University, Norfolk, VA, United States, 8 Coral Coast Conservation Center, Fiji, Fiji, 9 College of Science and Engineering, James Cook University and Australian Institute of Marine Science, Townsville, QLD, Australia, 10 Biodiversity and Geosciences Program, Museum of Tropical Queensland, Queensland Museum Network Townsville, Townsville, QLD, Australia, 11 Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD, Australia, 12 Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD, Australia, 13 Marine Laboratory, University of Guam, Mangilao, GU, United States, 14 Horniman Museum and Gardens, London, United Kingdom, 15 Aquatic Research Facility, Environmental Sustainability Research Centre, University of Derby, Derby, United Kingdom, 16 Centre for Environmental and Marine Studies (CESAM), Department of Biology, University of Aveiro, Aveiro, Portugal, 17 Department of Biological Sciences, Lehigh University, Bethlehem, PA, United States, 18 Department of Invertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC, United States, 19 Branch for Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Giessen, Germany, 20 Molecular Invertebrate Systematics and Ecology Lab, Faculty of Science, University of the Ryukyus, Nishihara, Japan, 21 Department of Physiology, Genetics and Microbiology, University of Alicante, Alicante, Spain, 22 Faculty of Science, Climate Change Cluster, University of Technology Sydney, Ultimo, NSW, Australia, 23 Department of Animal Ecology and Systematics, Justus Liebig University Giessen, Giessen, Germany

Coral research is being ushered into the genomic era. To fully capitalize on the potential discoveries from this genomic revolution, the rapidly increasing number of high-quality genomes requires effective pairing with rigorous taxonomic characterizations of specimens and the contextualization of their ecological relevance. However, to date there is no formal framework that genologists, taxonomists, and coral scientists can collectively use to systematically acquire and link these data. Spurred by the recently announced “Coral symbiosis sensitivity to environmental change hub” under the “Aquatic Symbiosis Genomics Project” – a collaboration between the Wellcome Sanger Institute and the Gordon and Betty Moore Foundation to generate gold-standard genome sequences for coral animal hosts and their associated Symbiodiniaceae microalgae (among the sequencing of many other symbiotic aquatic species) – we outline consensus guidelines to reconcile different types of data. The metaorganism nature of the coral holobiont provides a particular challenge in this context and is a key factor to consider for developing a framework to consolidate genomic, taxonomic, and ecological (meta)data. Ideally, genomic data should be accompanied by taxonomic references, i.e.,
INTRODUCTION

The rapid development of sequencing technologies and the ever-decreasing cost has led to a discrepancy between the generation of primary sequencing data (sequence reads) and their assembly, annotation, and curation (genomes, genes, etc.): we are producing more data than we can “consume” (Richards, 2015; Voolstra et al., 2017a). This inconsistency is highlighted by the now routinely required provisioning of primary sequencing data to a public database (NCBI nr, EMBL ENA, and DDBJ) prior to publication vs. the provisioning of assembled and annotated sequencing data (the type of data that most people work with), which currently is not a strict requirement (Liew et al., 2016; Voolstra et al., 2017a). Indeed, accessibility to assembled sequencing data is generally provided on a voluntary basis, and more often than not, relies on secondary databases, such as reefgenomics.org (Liew et al., 2016) or symporal.org (Hume et al., 2019) in the marine/coral reef domain. These secondary outlets often lack funding (or the availability of funding schemes that support such endeavors), rendering their continued upkeep financially challenging, e.g., CnidBase that is now no longer accessible (Ryan and Finnerty, 2003) or GeoSymbio which is no longer updated (Franklin et al., 2012). Even when processed sequencing data are available, another problem is version control, i.e., access to and documentation of previous transcriptome or genome versions (assemblies), which in some instances are critical to reproduce results. Public databases often put restrictions in place for the upload of genome/transcriptome assemblies or gene sets, resulting in different versions used for analysis, relative to those that are published with the respective study. This disparity is further complicated by the circumstance that sequencing databases often produce their “own” version of an uploaded genome based on a standardized analytical framework. In the case of the Aiptasia (Exaiptasia diaphana) genome (Baumgarten et al., 2015), for instance, a comparison of the submitted GenBank version (PRJNA261862) to the RefSeq version (PRJNA386175) reveals different lengths and numbers of protein-coding genes. The same can be observed for the genome of the coral Stylophora pistillata (Voolstra et al., 2017b) with the author-published version featuring 25,769 genes, the corresponding submitted GenBank version harboring 24,140 of these genes, and the associated RefSeq version featuring 33,239 genes with no corresponding gene mapping file to cross reference the different genes and identifiers.

Large-scale sequencing projects often prioritize the generation of genomic and transcriptomic data over comprehensive formal descriptions of samples and their environmental/ecological setting (i.e., metadata). This is true even for species with high intraspecific variation in heritable functional traits, such as scleractinian corals, for which ecological and environmental context matters greatly (Ziegler et al., 2014; Sawall et al., 2015; Röthig et al., 2017; Thomas et al., 2018; Bongaerts et al., 2020; Kavousi et al., 2020). Underlying this problem is that most molecular databases focus largely on sequencing data deposition and do not provide a comprehensive framework for the deposition of associated metadata (Riginos et al., 2020). The association between sequencing data and contextual, environmental (meta) data makes interpretation of these data more meaningful and allows the alignment of molecular patterns with phenotypes (Hume et al., 2020; Voolstra et al., 2020; Grottoli et al., 2021). The recently established Genomic Observatories Metadatabase (GEOME) aims to expedite and improve deposition and retrieval of molecular data and metadata for biodiversity research (Deck et al., 2017; Riginos et al., 2020). Here, we address a specific key issue relevant to this aim: the importance of accurate taxonomic descriptions of sequenced coral holobiont specimens and the deposition of specimen vouchers to provide a formal taxonomic framework for sequencing data, coupled with the ability to update existing descriptions. The absence of a proper taxonomic treatment

skeletal vouchers as formal morphological references for corals and strain specimens in the case of microalgal and bacterial symbionts (cultured isolates). However, exhaustive taxonomic characterization of all coral holobiont member species is currently not feasible simply because we do not have a comprehensive understanding of all the organisms that constitute the coral holobiont. Nevertheless, guidelines on minimal, recommended, and ideal-case descriptions for the major coral holobiont constituents (coral animal, Symbiodiniaceae microalgae, and prokaryotes) will undoubtedly help in future referencing and will facilitate comparative studies. We hope that the guidelines outlined here, which we will adhere to as part of the Aquatic Symbiosis Genomics Project sub-hub focused on coral symbioses, will be useful to a broader community and their implementation will facilitate cross- and meta-data comparisons and analyses.

Keywords: coral reef, coral holobiont, scleractinia, symbiodiniaceae, prokaryotes, genome sequencing, taxonomy, genomics


3http://aiptasia.reefgenomics.org/download/
4http://spis.reefgenomics.org/download/
associated with sequenced specimens makes cross-referencing and meta-analyses challenging and, in the worst case, can confound analyses due to taxonomic misclassification of sequence data (Bonito et al., 2021). Simply put, while everyone agrees on the value of properly curated specimens and associated sequencing data, what is missing is a guide or reference that details what should be provided when sequencing a genome.

Here we were motivated to provide such consensus guidelines as we embark on a new initiative to substantially improve the number and quality of genomes available from scleractinian corals and their associated Symbiodiniaceae microalgae (Supplementary Table 1). The “Coral symbiosis sensitivity to environmental change hub” is embedded in a phylogenetically broader effort to survey genomes across a wide variety of marine organisms and their microbial symbionts (octocorals, sponges, clams, nudibranchs, etc.) entitled the “Aquatic Symbiosis Genomics Project”, which is jointly funded by the Wellcome Sanger Institute and the Gordon and Betty Moore Foundation.

We aim to provide consensus guidelines on the “minimal taxonomic information” that should be provided to maximize the utility of the generated sequence data. We further expand these guidelines to also include coral-associated prokaryotic genomes due to recent efforts in describing and collating the culturable fraction of the prokaryotic community of the coral holobiont (Sweet et al., 2021). We advocate for the provision of taxonomic information for the most important (i.e., best understood, most commonly researched) coral holobiont entities: the coral animal host, the Symbiodiniaceae microalgae, and the associated prokaryotes (bacteria and archaea). Although the focus of the guidelines is aimed toward shallow-water stony corals (Scleractinia), they are broadly applicable to all coral taxa, and we incorporate specific considerations for temperate, cold-/deep-water corals as well as octocorals (Octocorallia), black corals (Antipatharia), and other hexacorals (Hexacorallia) where applicable.

**CONSENSUS GUIDELINES – ASSESSMENT AND RECOMMENDATIONS**

Standardized morphological and molecular taxonomic practices are not equally available for all coral holobiont entities, nor equally well tried-and-tested. For instance, coral skeletal-based taxonomy has a long history (Veron, 2000), but is not without discrepancies if compared against molecular-based analyses (Fukami et al., 2004; Kitahara et al., 2016; Terraneo et al., 2019a; Cowman et al., 2020). But therein lies the conundrum: while molecular analyses commonly achieve superior taxonomic resolution, they rely on initial expert review and annotation to prevent error-propagation through incorrect phylogenetic annotations of sequence database entries (Tripp et al., 2011).

It is important to acknowledge that taxonomic identification is challenging because morphological characteristics that differentiate species in one genus may not be applicable to other genera, and the same is true for molecular markers (Veron, 2000; Shearer et al., 2002; Stolarski et al., 2021). In the case of many coral lineages, species-level molecular markers are simply not (yet) available (Quattrini et al., 2018; Cowman et al., 2020; Erickson et al., 2021), partially due to ongoing taxonomic revisions, but also due to corals exhibiting low levels of congeneric divergence for commonly employed (mitochondrial) gene markers, effectively hampering species-level resolutions (Shearer et al., 2002; Supplementary Table 2).

Both circumstances support the necessity of skeletal voucher specimens as a reference to validate, synchronize, or update ascribed taxonomic annotations and allow later re-evaluation in case of taxonomic revisions. Importantly, specimens should be identified with reference to the original type specimens and descriptions, and not the most recent or most easily accessible revision, unless these provide a formal re-description (or illustration) of type material (or neotype specimen where applicable). Nevertheless, for most sequenced coral genomes to date, such information is not or not easily accessible (Supplementary Table 3). With most museums placing emphasis on digitizing collections, it should become easier to access photographs of type specimens and original descriptions—a major step forward from even a decade ago.

Museum curators and collection managers can also facilitate this process by providing access to specimens (including digitized versions) in their collections—a valuable service to the broader scientific community.

By comparison, formal descriptions of Symbiodiniaceae are rather recent, with the vast majority established or formalized after molecular data began to be integrated (LaJeunesse et al., 2012; Wham et al., 2017; Lee et al., 2020). The updated taxonomy provided an overdue revision of this group of microalgal symbionts, acknowledging their substantial genetic divergence and discouraging the use of informal clade designations as auxiliary constructs (LaJeunesse et al., 2018).

The majority of sequenced genomes are currently available from the genus *Symbiodinium*, with many genera not yet having genome assemblies available (Supplementary Table 4). Rather, Symbiodiniaceae associations are commonly described through means of marker gene elucidation using a range of different methodologies (Sampayo et al., 2009; LaJeunesse et al., 2018; Hume et al., 2019; Grottoli et al., 2021). Common markers that are sometimes used in conjunction include ITS, ITS2, psbA, SSU, LSU, and cp23S, which are utilized along with morphological data and host associations (Sampayo et al., 2009; LaJeunesse et al., 2018; Hume et al., 2019).

For coral-associated prokaryotes, much work remains to be done (Supplementary Table 5), but the recent assembly and genome-level description of bacteria associated with *Porites lutea* Milne Edwards and Haine (1851) (Robbins et al., 2019) and the cataloging of cultured bacterial coral isolates (Sweet et al., 2021) provide a groundwork to build upon. Given that coral genomics is a nascent field, any guidelines put forward here must be considered provisional, and indeed current limitations should be a motivation rather than a barrier to begin to work on formulating the types of information that are most important to provide alongside sequencing data. While it is evident that multiple challenges are associated with taxonomy at all levels of the coral holobiont, we begin with a set of guidelines focusing...
on what should be provided when generating reference genomic data for the coral animal host, Symbiodiniaceae microalgae, and those prokaryotes that are either cultured or for which a full-length 16S rRNA gene reference sequence or a well-assembled (meta)genome is available (Supplementary Material). Our recommendations are not prescribed for metabarcoding, gene expression, or metagenomic/-transcriptomic surveys per se, as they may become overburdening for these latter types of studies. Although providing metadata descriptors for these data types in as comprehensive a manner as possible is desirable, they typically do not represent “reference datasets” because multiple studies are typically available for these types of sequencing data for any given species (e.g., 16S metagenome datasets exist for the same species from multiple locations). We further advocate establishing a well-curated set of specimen vouchers associated with primary reference sequencing data, which then allows alignment of samples against that reference. This should minimize misannotation and curtail error propagation caused by annotating tertiary sequencing data against secondary sequencing data.

The Coral Animal Host

To date, more than 9,000 nominal coral species (coral defined as animals in the cnidarian classes Anthozoa and Hydrozoa that secrete calcareous or proteinaceous skeletons sensu Cairns (Cairns, 2007) have been described (WoRMS Editorial Board, 2020). These include 5,941 scleractinian coral species of which 1,627 are currently considered valid (Hoeksema and Cairns, 2020). Accordingly, the boundaries and classification of these animals can be blurred by the great morphological plasticity of the skeletal features traditionally used for their identification (Veron, 2000), their hybridization potential (Vollmer and Palumbi, 2002; Willis et al., 2006; Richards and Hobbs, 2015; Quattrini et al., 2019), as well as widespread cryptic speciation (Todd, 2008; Forsman et al., 2009; Herrera and Shank, 2016; Bongaerts et al., 2020; Gómez-Corralas and Prada, 2020). To obtain a more precise taxonomic classification, coral taxonomists have started to use genetic/genomic data to identify phylogenetically informative morphological characters, which can be incorporated into identification keys (Terraneo et al., 2019b; Arrigoni et al., 2020). To this end, several mitochondrial and nuclear markers have been developed to resolve the taxonomy of corals to reflect their actual evolutionary relationships (Supplementary Table 2). With the advent of sequencing technologies becoming more affordable, genome-wide information (e.g., single nucleotide polymorphisms, ultraconserved elements, exons) can now also be incorporated into coral classification methodologies (Arrigoni et al., 2020), although the cost of sequencing still remains a hurdle for many researchers. Moreover, the sequencing and assembly of coral genomes provide a further important source of information to complement previous identification efforts (Shinzato et al., 2021).

Records of the sampling location, depth, and specimen phenotypic traits (including field images and the collection of a specimen/skeletal voucher) are important to inform accurate species identification, but are not always provided. Likewise, taxonomic identification (genotyping) based on specific molecular markers/barcodes and/or whole mitochondrial genome comparison is desirable (e.g., Buitrago-López et al., 2020). Notably, the vast majority of genome reports have deposited the raw sequences in publicly available sequencing databases. Although we recognize that sequencing genomes typically aligns to research projects in a given region (or even reef), ideally specimens should be collected from the type locality for the species of interest, or at least compared (genetically and morphologically) with a specimen from the type locality to ensure the specimen represents the species of interest. Likewise, the specimen to be sequenced should be selected based on morphological comparison to the name-bearing type specimen and the original description. Selecting specimens closely resembling the original type specimen from the type locality significantly reduces the chances of applying an incorrect taxonomic name to the genome, even when the species is the subject of subsequent taxonomic revision. Collecting from the type locality is particularly important given the extensive geographic and depth structure reported in many putatively widespread coral species that may well represent distinct species (e.g., Richards et al., 2016; Sheets et al., 2018; Bongaerts et al., 2021). Collection of high-quality field images and specimen/skeletal vouchers enables comparison of detailed skeletal morphology to the type specimen and informs on genome-to-morphotype correlations. While some specimens may be transported to aquaria, it is important to ensure that a voucher is taken of the original colony in the field, as coral morphology can change dramatically under aquarium conditions.

Since we recognize that coral taxonomy is a “moving target”, there is a need to bridge efforts for genomics to reconcile with the constantly evolving species classification. To this end, we suggest somewhat flexible taxon-description guidelines for coral genomic researchers (Table 1 and Figure 1), which attempt to avoid errors that have been commonly made in the past when assigning a species name to a genome, most notably the failure to maintain a specimen/skeletal voucher to ensure comparison with type material morphology. These guidelines are more fully described in the Supplement (Supplementary Methods). Implementing this practice will become fundamental as more genomes are sequenced, more cryptic species are identified, and novel morphological tools and techniques are developed to assign taxonomic status and identity. Without a reference specimen voucher, it becomes impossible to independently evaluate and update the taxonomy of a specimen and we are left relying only on the genome sequence and its associated metadata for taxonomic assignment. Having voucher specimens will allow the processes of genome sequencing and taxonomic assignment to be iterative, and mistakes can be corrected over time as new data emerge and taxonomic assignments are modified accordingly. This process will be facilitated by biologists and genomics working together with taxonomists, and it
TABLE 1 | Consensus guidelines regarding associated metadata deposition for coral specimen collection targeted for genome sequencing.

<table>
<thead>
<tr>
<th>Metadata provision guideline</th>
<th>Coral genome from sperm</th>
<th>Coral genome from holobiont sample (colony fragment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>– High quality DNA voucher material from sperm isolation</td>
<td>– High quality DNA voucher material from holobiont isolation</td>
</tr>
<tr>
<td></td>
<td>– Common phylogenetic marker sequences (e.g., COI, ITS, 18S, mtMutS, 28S)</td>
<td>– Common phylogenetic marker sequences (e.g., COI, ITS, 18S, mtMutS, 28S)</td>
</tr>
<tr>
<td></td>
<td>– Voucher photograph of live parent colony from which sperm was collected; photographs should include close-ups of skeletal structures</td>
<td>– Voucher photograph of live coral colony from which specimen was collected; photographs should include close-ups of skeletal structures</td>
</tr>
<tr>
<td></td>
<td>– Comprehensive metadata: GPS location, sampling date, depth, temperature, (provisional) taxon ID</td>
<td>– Comprehensive metadata: GPS location, sampling date, depth, temperature, (provisional) taxon ID</td>
</tr>
<tr>
<td></td>
<td>– Reference to the original species description</td>
<td>– Reference to the original species description</td>
</tr>
<tr>
<td>Recommended (in addition to Minimum)</td>
<td>– Cryopreserved sperm sample</td>
<td>– Cryopreserved holobiont sample</td>
</tr>
<tr>
<td></td>
<td>– High quality DNA voucher material from the (holobiont) parent colony</td>
<td>– High quality DNA voucher material from the (holobiont) coral colony</td>
</tr>
<tr>
<td></td>
<td>– Parent colony specimen deposited and registered in a museum with a collection code</td>
<td>– Skeletal and (holobiont) coral colony specimen deposited and registered in a museum with a collection code</td>
</tr>
<tr>
<td>Talid (in addition to Recommended)</td>
<td>– Ramets of the parental colony should be maintained long-term in (public) aquariums/research facilities, preferably across multiple locations in case of mortality (Zoccola et al., 2020)</td>
<td>– Ramets of the parental colony should be maintained long-term in (public) aquariums/research facilities, preferably across multiple locations in case of mortality (Zoccola et al., 2020)</td>
</tr>
<tr>
<td></td>
<td>– In situ tagging of colony from which sperm was collected for long-term resampling and photographing</td>
<td>– In situ tagging of colony that was sequenced for long-term resampling and photographing</td>
</tr>
<tr>
<td></td>
<td>– Complete formal taxonomic description published, if not available prior (including name, type specimen, museum registration code)</td>
<td>– Complete formal taxonomic description published, if not available prior (including name, type specimen, museum registration code)</td>
</tr>
</tbody>
</table>

Relatively pure coral DNA can be collected from coral sperm, but requires sample collection during spawning, whereas DNA obtained from a colony fragment contains a mix from many different organisms, most notably “contaminating” DNA from the endosymbiotic Symbiodiniaceae.

FIGURE 1 | Overview of consensus guidelines regarding metadata deposition for coral, Symbiodiniaceae, and prokaryotic specimen collections targeted for (meta)genomic sequencing (further details in Tables 1–3).

constitutes an ongoing process rather than a singular event (Buckner et al., 2021).

The Microalgal Symbiont (Symbiodiniaceae)

The primary eukaryotic symbionts of shallow-water corals belong to the family Symbiodiniaceae, a taxonomically, ecologically, and genetically diverse group of dinoflagellate microalgae (LaJeunesse et al., 2018). Symbiodiniaceae have wide-ranging physiological tolerances to light, temperature, salinity, and nutrient preferences, which impact coral health and resilience (Rowan et al., 1997; Baker, 2003; Sampayo et al., 2008; Suggett et al., 2015, 2017; Ochsenkühn et al., 2017; Morris et al., 2019). Like other dinoflagellates, their genomes are large and complex (LaJeunesse et al., 2005; Lin, 2011) and they can be
TABLE 2 | Consensus guidelines regarding associated metadata deposition for Symbiodiniaceae specimen collection targeted for genome sequencing.

<table>
<thead>
<tr>
<th>Metadata provision guideline</th>
<th>Symbiodiniaceae genome from available culture</th>
<th>Symbiodiniaceae genome from holobiont sample (colony fragment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>– High quality DNA voucher material from microalgal culture isolation&lt;br&gt;– Common phylogenetic marker sequences (e.g., LSU, ITS2, cob, cp23S, psbA cops; the optimal combination will vary by species)&lt;br&gt;– Light microscopy images (for cell sizes as rough morphological feature)&lt;br&gt;– Comprehensive metadata: (coral) host species, GPS location, sampling date, depth, temperature, (provisional) taxon ID&lt;br&gt;– Indicator whether culture is the dominant symbiont of the “host” it was isolated from</td>
<td>– High quality DNA voucher material from holobiont isolation&lt;br&gt;– Common phylogenetic marker sequences (e.g., LSU, ITS2, cob, cp23S, psbA cops; these would only represent the numerically dominant, eco-physiologically relevant, and temporally stable primary symbiont)&lt;br&gt;– Light microscopy images (for cell sizes as rough morphological feature)&lt;br&gt;– Comprehensive metadata: (coral) host species, GPS location, sampling date, depth, temperature, (provisional) taxon ID&lt;br&gt;– ITS2 defining intragenomic variant (DIV) profiles or denaturing gradient gel electrophoresis (DGGE) profiles of all symbionts in the host (useful for assessing community members in mixed samples and identifying the dominant species and potential contaminants, while acknowledging that without correction for ITS2 copy number they won’t necessarily reflect relative abundance accurately)</td>
</tr>
<tr>
<td>Recommended (in addition to Minimum)</td>
<td>– Cryopreserved stock&lt;br&gt;– ITS2 defining intragenomic variant (DIV) profiles of the culture from amplicon sequencing (useful for monoclonal strains to generate genetic fingerprints to be used as reference for other studies)</td>
<td>– Cryopreserved stock (will have background symbiont and host contamination, which should be indicated)</td>
</tr>
<tr>
<td>Ideal (in addition to Recommended)</td>
<td>– Live culture stock started from single-cell isolation and deposition in a recognized culture collection (e.g., ANACC, CCAP, NCMA)&lt;br&gt;– SEM/TEM images (including deposition of SEM stubs as reference for other studies)</td>
<td>– SEM/TEM images (including deposition of SEM stubs as holotype with a museum or public collection/herbarium; notably, it may be difficult to determine if a given cell is the appropriate species in a mixed community)&lt;br&gt;– Complete formal taxonomic description published, if not available prior (including name, type specimen, museum registration code)</td>
</tr>
</tbody>
</table>

Highly divergent (LaJeunesse et al., 2005; Lin, 2011; Aranda et al., 2016; González-Pech et al., 2019, 2021; Nand et al., 2021; Supplementary Table 4). Initially, all Symbiodiniaceae were thought to comprise a single species (Freudenthal, 1962; Kevin et al., 1969; Taylor, 1974), but the accumulation of molecular data has led to our current understanding that there are likely hundreds of species spread across tens of genera within this microalgal family (LaJeunesse et al., 2018). Most await formal description with only ~40 valid Symbiodiniaceae taxa currently formally described. Such descriptions will be needed to map the microalgal symbionts to their coral host distributions, to define their relevant units for conservation and protection, and to understand the extent to which their functional variation translates into acclimatory and adaptive potential for the coral holobiont (Howells et al., 2012, 2020; Hume et al., 2016, 2020; Thornhill et al., 2017; Torda et al., 2017; Voolstra et al., 2021). Due to the cryptic morphology of these organisms, their taxonomic recognition relies on molecular evidence, necessitating new tools to resolve diversity, e.g., SymPortal (Hume et al., 2019) and new approaches to link genomic data to voucher specimens.

The intracellular nature of the coral-Symbiodiniaceae symbiosis complicates genome sequencing because it can be difficult to obtain pure Symbiodiniaceae (or conversely coral) DNA. Consequently, many Symbiodiniaceae genomic resources are “contaminated” with DNA from their coral hosts and vice versa (Celis et al., 2018). The potential presence of cells from multiple Symbiodiniaceae species in the same host adds further complexity. Therefore, the isolation of individual symbiont cells to establish clonal cultures is an important step for targeted sequencing (Nitschke et al., 2020). Most ecologically important symbionts have yet to be cultured, and many may ultimately prove unculturable given their narrow growth requirements (Krueger and Gates, 2012). In addition, cultured cells are not necessarily representative of their in hospite counterparts, both genetically and functionally (Santos et al., 2001; Maruyama and Weis, 2021). To resolve the complex diversity of Symbiodiniaceae (LaJeunesse et al., 2018), a combined approach of sequencing in hospite cells from holobiont tissue samples as well as cells from independent isoclonal cultures will be needed. This is the strategy pursued in the “Coral symbiosis sensitivity to environmental change hub”. Additionally, flow cytometry and fluorescent-activated cell sorting (FACS) with subsequent sequencing may be employed (Rosental et al., 2017; Levy et al., 2021). Whether the microalgae are sourced from mixed holobiont tissue or pure cultures, the “minimal taxonomic information” for sequencing Symbiodiniaceae genomes (Table 2 and Figure 1) should include the deposition of cryo-preserved DNA, genetic characterization with standard phylogenetic markers, light microscopy images of cells for morphological characterization, and metadata describing coral host identity, the coral host’s symbiont population composition, and the environment from...
which microalgal cells were isolated (Supplementary Methods). Whenever possible, additional useful steps would include generating amplicon sequencing data, establishing live cultures, and publishing a formal taxonomic description of the species in advance of or alongside the genome. However, we are keenly aware that Symbiodiniaceae taxonomy is in its infancy, that the number of undescribed species is staggering, and that formal descriptions require a tremendous amount of work and funding. While all Symbiodiniaceae species should eventually be formally named, we recognize that in the near future many genomes will need to be published for undescribed or not fully characterized specimens. Following the consensus guidelines outlined here should maximize the potential for creating unambiguous genomic information associated with a given specimen and minimize errors, while the Symbiodiniaceae taxonomy continues to be resolved. Although deep-sea corals lack Symbiodiniaceae symbionts, they can host other eukaryotic microbes in their tissues, e.g., apicomplexans (Vohsen et al., 2020a). Similarly, there are numerous additional soft-bodied anthozoan taxa, many in symbioses with Symbiodiniaceae symbionts, they can host other eukaryotic microbes in their tissues, e.g., apicomplexans (Vohsen et al., 2020a). Similarly, there are numerous additional soft-bodied anthozoan taxa, many in symbioses with Symbiodiniaceae symbionts. This is because in many cases the availability of direct full-length sequencing of 16S rRNA genes (Carrade-de et al., 2020). Fewer studies have utilized shotgun metagenomic sequencing to obtain prokaryotic genomes via metagenome-assembled genomes (MAGs) (Neave et al., 2017a; Cárdenas et al., 2018; Robbins et al., 2019). As outlined above, it is desirable to provide both the raw sequencing data and the assembled genomes, as well as the bioinformatic pipelines used for assembly and annotation (Mende et al., 2020; Sweet et al., 2021; Cárdenas and Voolstra, 2021). If available, culture-based methods are valuable because they directly align a 16S rRNA gene sequence or genome with a cultured isolate that can then be subjected to further study and experimental validation (Neave et al., 2014, 2017a). Despite these advantages, microbial culturing is challenging. This is because in many cases the biotic and abiotic conditions necessary to obtain microbial growth are unknown or hard to mimic in a laboratory context (Bodor et al., 2020), on top of the difficulties associated with taxonomic identification of cultured strains (Varghese et al., 2015). In addition, the incorporation of genomic information into the hierarchical system of classification for prokaryotes

The Prokaryotic Community

Bacteria are pivotal members of the coral holobiont contributing to metabolism, health, and stress tolerance (Rosenberg et al., 2007; Ziegler et al., 2017; Robbins et al., 2019; Voolstra and Ziegler, 2020; Peixoto et al., 2021). Coral-associated bacterial communities are complex and highly variable, which must be considered in the implementation of consensus guideline approaches (Roder et al., 2015; Williams et al., 2015; Röthig et al., 2017; Sweet et al., 2017; Vohsen et al., 2020b; Voolstra and Ziegler, 2020). While historically bacteria (host-associated and free-living) were characterized employing culturing methods, this has been largely replaced by sequencing-based approaches that are more affordable and higher throughput, although the two different approaches are complementary in scope and insight (Sweet et al., 2021). Here, we discuss methods best suited to characterize prokaryotic associates and provide suggestions to “standardize” coral microbiome work for enhanced comparability and meta-analysis.

Many studies feature 16S rRNA gene amplicon sequencing to describe the microbiome of corals. Large datasets, such as obtained for the Earth Microbiome Project, maximize the comparability among studies (Thompson et al., 2017), but the employed primers are prone to misamplification in corals and provide limited coverage of some taxonomic groups (Bayer et al., 2013; Robbins et al., 2019; van de Water et al., 2020). Such methodological constraints may resolve in the near future with the availability of direct full-length sequencing of 16S rRNA genes (Carrade-de et al., 2020). Fewer studies have utilized shotgun metagenomic sequencing to obtain prokaryotic genomes via metagenome-assembled genomes (MAGs) (Neave et al., 2017a; Cárdenas et al., 2018; Robbins et al., 2019). As outlined above, it is desirable to provide both the raw sequencing data and the assembled genomes, as well as the bioinformatic pipelines used for assembly and annotation (Mende et al., 2020; Sweet et al., 2021; Cárdenas and Voolstra, 2021). If available, culture-based methods are valuable because they directly align a 16S rRNA gene sequence or genome with a cultured isolate that can then be subjected to further study and experimental validation (Neave et al., 2014, 2017a). Despite these advantages, microbial culturing is challenging. This is because in many cases the biotic and abiotic conditions necessary to obtain microbial growth are unknown or hard to mimic in a laboratory context (Bodor et al., 2020), so that the two different approaches are complementary in scope and insight (Sweet et al., 2021). Here, we discuss methods best suited to characterize prokaryotic associates and provide suggestions to “standardize” coral microbiome work for enhanced comparability and meta-analysis.

### TABLE 3 | Consensus guidelines regarding associated metadata deposition for prokaryotic specimen collection targeted for (metagenome sequencing.

<table>
<thead>
<tr>
<th>Metadata provision guideline</th>
<th>Cultured bacteria</th>
<th>Uncultured bacteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Minimum</strong></td>
<td>DNA sequence available in public database</td>
<td>DNA sequence available in public database</td>
</tr>
<tr>
<td></td>
<td>DNA extraction methods, sequencing platform</td>
<td>DNA extraction methods, sequencing platform</td>
</tr>
<tr>
<td></td>
<td>Host description, photos, and culturing methods</td>
<td>Host description (including environmental conditions, health state) and photos</td>
</tr>
<tr>
<td></td>
<td>Comprehensive metadata: (coral) host species, GPS location, sampling date, depth, temperature, (provisional) taxon ID</td>
<td>Comprehensive metadata: (coral) host species, GPS location, sampling date, depth, temperature, (provisional) taxon ID</td>
</tr>
<tr>
<td><strong>Recommended (in addition to Minimum)</strong></td>
<td>Cryopreserved stock</td>
<td>Host voucher</td>
</tr>
<tr>
<td></td>
<td>Information on growth rates and conditions</td>
<td>Detailed bioinformatic methods and assembled sequence data</td>
</tr>
<tr>
<td></td>
<td>Detailed bioinformatic methods and assembled sequence data</td>
<td>Photos of bacterial colonies, information on color, border, shape of cultured cells</td>
</tr>
<tr>
<td></td>
<td>Photographs of bacterial colonies, information on color, border, shape of cultured cells</td>
<td>Photographs of bacterial colonies, information on color, border, shape of cultured cells</td>
</tr>
<tr>
<td><strong>Ideal (in addition to Recommended)</strong></td>
<td>Live culture stock started from single-cell isolation and deposition in a recognized culture collection (e.g., ANACCC, CCAP, NCMA)</td>
<td>High quality DNA voucher material</td>
</tr>
<tr>
<td></td>
<td>SEM/TEM images (including deposition of SEM stubs as holotype with a museum or public collection/herbarium)</td>
<td>Tagging of coral colony that was sequenced for long-term resampling and photographing</td>
</tr>
<tr>
<td></td>
<td>Complete formal taxonomic description published, if not available prior (including name, type specimen, museum registration code)</td>
<td>Complete formal taxonomic description published, if not available prior (including name, type specimen, museum registration code)</td>
</tr>
</tbody>
</table>
has been proven to be challenging below the genus level.

Arguably, resolving species- and strain-level differences are critical to understand ecologically and physiologically relevant distinctions, and alternative prokaryotic taxa classifications have been proposed to amend these issues (Staley, 2006; Neave et al., 2017a; Parks et al., 2018; Van Rossum et al., 2020; Yan et al., 2020). Given the current classification “fluidity”, a comprehensive assessment and description of obtained microbial cultures associated with host metadata is therefore required to facilitate contextualization of results from different studies, enable cross-comparability, and allow for reproducibility (Table 3, Figure 1, and Supplementary Methods).

**DISCUSSION AND PERSPECTIVE**

The sequencing era has the potential to unlock the complexity of the coral holobiont by means of highly resolved genomic interrogation of its member species (i.e., coral animal host, Symbiodiniaceae microalgae, associated prokaryotes, etc.). While initially the focus was on sequencing “one genome at a time”, e.g., the *Stylophora pistillata* Esper, 1797 holobiont genomics studies (Bayer et al., 2013; Aranda et al., 2016; Neave et al., 2017a,b; Voolstra et al., 2017b), there is now a suite of efforts to target the coordinated sequencing of all (or the most common) holobiont member species (Robbins et al., 2019). One of these efforts is the “Aquatic Symbiosis Genomics Project”. To maximize the utility of the generated data, a common commitment to formulate and adhere to consensus guidelines within a defined taxonomic framework is required. Here, we lay out the guidelines that the “Coral symbiosis sensitivity to environmental change hub” will follow to facilitate meta-analyses, cross-comparisons, and back-tracking of samples, with the intent that other initiatives can join and adopt this approach. Our first step was to decide on the coral species that would be part of this project (Supplementary Table 1). To do this, we collated the current state of play of coral genomes and assessed which key species were missing or suffered from incomplete and/or fragmented genome assemblies. We then compared where our selected corals were initially described from, that is the country of origin of the type specimen (type locality).

Samples are currently in the process of being collected by in-country scientists and experts who are in charge of sampling and archiving specific types of metadata for each specimen. Without such data, type specimens and previous data collections cannot be ground-truthed or revised (Blom, 2021; Buckner et al., 2021), which ultimately limits the usefulness of -omics data for current and future analyses.

One additional barrier is the lack of a central repository that integrates (i) several or all types of data (genomic, taxonomic, physiological, chemico-physical, etc.) from (ii) multiple coral holobiont entities (cross-kingdom) with (iii) the inclusion of version control and access to “derived” data products. Recent initiatives aim to provide centrally available open-access databases that integrate primary data and some associated metadata (Box 1). While the broad centralized integration of data is meaningful, our point is not to suggest a single database to hold all data, as this is likely to affect implementation, focus, and usability. Rather, the coordination of efforts into a few collective and linked databases is desirable to avoid duplication of efforts. The power of a consensus framework was recently outlined for coral bleaching experimentation, which detailed response variables to increase comparability and hasten scientific insight (Grottoli et al., 2021). Given the pervasive lack of long-term funding for data centralization (including logistics, sorting, and collection), the alternative bottom-up, community-driven model is a more realistic goal to attain, and will be particularly valuable if it manages to incentivize data and meta-data deposition. Arguably, the burden to follow through with comprehensive data deposition lies with the individual researcher and is typically done “after the fact” (after publication). However, free-of-charge repositories, such as zenodo.org or figshare.org, provide digital object identifiers (DOIs) and by that a mechanism of citing and acknowledging well-curated data, ultimately incentivizing such efforts. For the “Aquatic Symbiosis Genomics Project”, all sequence data will be openly accessible. All raw and assembled sequence data will be deposited in the European Nucleotide Archive (ENA) database which is part of the International Nucleotide Sequence Database Collaboration that also entails the DNA Database of Japan (DDBJ) and the GenBank at NCBI, which exchange data on a daily basis. Further, our intention is to rapidly publish all submitted genome assemblies alongside their associated meta-data as Wellcome Open Research Data Notes, which can be cited. It is now up to us (the scientific community) to further foster these endeavors through proper

---

**BOX 1 | Open access databases that integrate primary data and associated metadata and provide tools for standardization for the genomic interrogation of (coral) holobionts.**

*Genomic Observatories Metadatabase at geome-db.org (Deck et al., 2017);* database that captures metadata about biological samples and associated genetic sequences.

*Reefgenomics at reefgenomics.org (Liew et al., 2016);* repository for curated marine genomics data.

*Coral trait database at coraltraits.org (Madin et al., 2016);* community-driven compilation of observations and measurements of scleractinian corals at the individual and contextual level.

*SymPortal at symportal.org (Hume et al., 2019);* analytical framework and platform for Symbiodiniaceae next-generation-sequencing (NGS) ITS2 profiling with integrated curated, public database.


*Brazilian Microbiome Project at bmmicrobiome.org;* aims to assemble a Brazilian Metagenomic Consortium/Database across taxonomic groups.

*Global Ocean Microbiome Project at ocean-microbiome.embl.de/companion.html (Sunagawa et al., 2015);* data portal of Tara Oceans global microbiome analysis products (processed sequencing data, focus not on corals).

*Earth Microbiome Project at earthmicrobiome.org (Thompson et al., 2017);* ongoing collaborative effort to characterize global microbial taxonomic and functional diversity across taxonomic groups, provides links to metadata, other results, and sequencing data.

*Coral/Symbiont Genomes and Transcriptomes Resource Database at bit.ly/coralsymgenomes: living spreadsheet for tracking which genomic resources are available or under development for corals, Symbiodiniaceae, and related marine organisms.*
acknowledgement and citation of non-traditional publication outlets. We hope that the consensus guidelines detailed here provide a path to broaden our understanding of coral holobionts, to accelerate discovery, and to facilitate novel solutions to mitigate coral degradation, which becomes ever more pertinent as we witness the continuous loss of reef ecosystems globally.

AUTHOR CONTRIBUTIONS

CV, KQ, SD, JP, RP, MA, ACB, CC, MZ, and MS conceived the idea. CV, KQ, SD, JP, RP, CB-L, TB, CC, DC, JE, TR, DS, MZ, and MS wrote the manuscript. CC conceived and created Figure 1. All authors collected the data, provided input, and contributed to the writing of the manuscript.

FUNDING

CV acknowledges funding from the German Research Foundation (DFG), grants 433042944 and 458901010. Open Access publication fees are covered by an institutional agreement of the University of Konstanz with Frontiers.

ACKNOWLEDGMENTS

We would like to acknowledge and thank the “Aquatic Symbiosis Genomics Project”, which is led by the Tree of Life Program at the Wellcome Sanger Institute. It is funded by the Gordon and Betty Moore Foundation through their “Symbiosis in Aquatic Systems Initiative” and by the Wellcome Trust through core funding to the Sanger Institute.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.701784/full#supplementary-material

REFERENCES


Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Voolstra, Quigley, Davies, Parkinson, Peixoto, Aranda, Baker, Barno, Barnish, Benzi, Bonito, Bourne, Buitrago-López, Bridge, Chan, Combosch, Craggs, Frommlet, Herrera, Quattrini, Röthig, Reimr, Rubio-Portillo, Suggett, Villela, Ziegler and Sweet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.