ANNUAL REVIEWS

*Annual Review of Animal Biosciences*

# Multiple Facets of Marine Invertebrate Conservation Genomics

Jose V. Lopez,[1] Bishoy Kamel,[2] Mónica Medina,[3] Timothy Collins,[4] and Iliana B. Baums[3]

[1]Department of Biological Sciences, Halmos College of Natural Sciences and Oceanography, Nova Southeastern University, Dania Beach, Florida 33004, USA; email: joslo@nova.edu

[2]Department of Biology, Center for Evolutionary and Theoretical Immunology, University of New Mexico, Albuquerque, New Mexico 87131, USA; email: bishoyh@unm.edu

[3]Department of Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA; email: mum55@psu.edu, baums@psu.edu

[4]Department of Biological Sciences, Florida International University, Miami, Florida 33199, USA; email: collinst@fiu.edu

Annu. Rev. Anim. Biosci. 2019. 7:473–97

First published as a Review in Advance on November 28, 2018

The *Annual Review of Animal Biosciences* is online at animal.annualreviews.org

https://doi.org/10.1146/annurev-animal-020518-115034





## Keywords

marine invertebrates, biodiversity, endangered, invasive, genomics

## Abstract

Conservation genomics aims to preserve the viability of populations and the biodiversity of living organisms. Invertebrate organisms represent 95% of animal biodiversity; however, few genomic resources currently exist for the group. The subset of marine invertebrates includes the most ancient metazoan lineages and possesses codes for unique gene products and possible keys to adaptation. The benefits of supporting invertebrate conservation genomics research (e.g., likely discovery of novel genes, protein regulatory mechanisms, genomic innovations, and transposable elements) outweigh the various hurdles (rare, small, or polymorphic starting materials). Here we review best conservation genomics practices in the laboratory and in silico when applied to marine invertebrates and also showcase unique features in several case studies of acroporid corals, crown-of-thorns starfish, apple snails, and abalone. Marine conservation genomics should also address how diversity can lead to unique marine innovations, the impact of deleterious variation, and how genomic monitoring and profiling could positively affect broader conservation goals (e.g., value of baseline data for in situ/ex situ genomic stocks).

## INTRODUCTION

The study of conservation genomics encompasses multiple practices and crosses over several disciplinary boundaries, similar to epidemiology or the study of cancer. During the present time of accelerating environmental change and imminent extinctions in the Anthropocene epoch (1), conservation practitioners may take on the zeal and urgency of first responders at a burning house. By definition, conservation genomics aims to apply modern genomic analysis to the preservation of the viability of populations and the biodiversity of living organisms. Thus, the field has an origin and strong and continuing links with population genetics, which measures levels of genetic diversity within populations, demographic history, and effective population size. Modern genomic methods can address classification via molecular taxonomy and phylogenomics, looking to characterize mechanisms of change and the potential for and degrees of hybridization. The field has been clearly affected by the current revolution and advances in next-generation or high-throughput DNA sequencing (HTS) methods (2). Several excellent reviews on conservation genomics exist in the literature (3–5), but few have focused specifically on marine invertebrates (6). This is somewhat startling when considering that invertebrates compose the majority of all extant animal phyla (7, 8). Thus, the charge of this article is to persuade the reader of the importance of invertebrates while also adequately explaining the basic framework and application of conservation genomics, as well as any peculiar perspectives relating to this unique clade of animals. The article starts broadly and eventually narrows to specific case studies and the details of genomics and bioinformatics used to implement and execute overall conservation goals. We first address the fundamental topic of biological diversity, genomics, and its conservation with respect to marine invertebrates.

### What Are We Conserving, and Why?

The genetic blueprint of DNA encodes most of the instructions needed for an organism to develop to its full potential (9). Genes underlie the organismal phenotype, and the invertebrates as an artificial clade present many examples of unique adaptations, bizarre morphologies, and astounding behaviors that can stretch belief (8, 10). Staying within the marine realm, for example, decorator crabs (*Loxorhynchus crispatus*) adorn their shells by hooking fellow sessile anemones, sponges, and bryozoans to their setae (Velcro-like bristles) as if to impress, but the desired outcome results in a protective camouflage that blurs their presence within the habitat. At the other end of the size scale, the giant squid (*Architeuthis dux*) captured not only the fears and imagination of ancient mariners but also prey more than 10 m away by extending two feeding tentacles tipped with serrated suckers. Going deeper into the sunless depths below 200 m (meso- and bathypelagic), multiple invertebrate faunas have evolved to create or harness their own light sources through bioluminescence. This biological production of light at night or in deeper aphotic zones can be used for intraspecific communication (mate attraction), or as interspecific prey avoidance via counterillumination or alarm warnings. Martini & Haddock (11) provide an illuminating review that summarizes the prevalence of this surprisingly common deep-sea trait among invertebrates (but which can vary with depth): 94.2% of Appendicularia (larvaceans), 92.9% of Polychaeta, 91.8% Ctenophora (comb jellies), and >97% of pelagic Cnidaria (Siphonophora, Hydromedusae, and Scyphozoa) appear to be bioluminescent. Of course, some of the phenotypes described above can also be affected and modified by nascent and current environmental variables in which an organism or individual may find itself. Thus, the term biological diversity (or biodiversity) itself has a rich, complicated, and multilayered origin and context that defy even a single paragraph for definition. Biodiversity encompasses variation from the molecular to the organism to the ecosystem level (12).

Genetic aspects of biodiversity (e.g., genetic variation) stem from various sources, such as mutation, drift, and selection, which is why studying population-level dynamics remains central to this field.

Assigning value to biological commodities such as diversity can be a tricky endeavor, yet the concept of conservation in itself implicitly implies some type of value (to humans, biology, and evolution). Energy and effort are expended to maintain some sort of stability in genomes while at the same time permitting gene diversification in certain vital systems and loci (immunity, mate recognition) (13). One of the first fully sequenced invertebrate genomes, from the sea urchin (*Strongylocentrotus purpuratus*), revealed an innate immune system with a surprising variation that includes over 200 Toll receptors and another 200 Nod-like NACHT domain–leucine-rich repeat proteins and scavenger receptor cysteine-rich proteins (13–15). Thus, to the individual, there is clear value in gene diversity.

More broadly, Pearson (16) explains that varying levels of biodiversity may not have equal value and should be prioritized. So, the conservation genomics field has advanced the combination of the two concepts of conservation and biodiversity. The pairing can be viewed as natural and unambiguous for most biologists who value diversity. Yet invertebrate diversity has functional as well as intellectual utility. For example, many marine invertebrates contribute to large-scale commercial fisheries, which feed millions of people worldwide. Invertebrates also have pivotal roles as critical ecosystem components, such as scleractinian corals on tropical reefs and mussels at deep-sea hydrothermal vents.

Semantically, genome (and its component gene) conservation does not fully equate to conservation genomics per se, although portions of each overlap and are related. We can briefly explore this from a reductionist molecular biology viewpoint. Many individual genes and their DNA/protein sequences display biological conservation on their own, simply by long outliving the individual host carriers and species and even higher Linnaean groupings to which these genes/sequences belong. Concrete examples can be found easily in sequence databases: Molecules that catalyze and compose essential cellular metabolism and structure include actin, tubulin, and glycolytic enzymes such as thymidine kinase. The corresponding genes encode conserved proteins that have changed little over hundreds of millions of years, with few amino acid sequence substitutions from early-radiating to the most-derived organisms (17). Comparison of tubulin amino acid sequences between a protozoan and humans, which diverged more than 1.7 billion years ago, shows roughly 85% identity. This exemplifies gene conservation at the most fundamental level. It also corroborates previous assertions that "gene orthologies [are the rule, not the exception] and thus comparative genomics are unifying forces of biology" (18). For example, hundreds of clusters of orthologous groups have been documented across the growing number of reference genomes as part of an effort to define the gene set of the last universal common ancestor (19).

In some ways, this also harkens to the provocative axiom that "the individual is a survival machine built by short-lived confederations of long-lived genes" (20). Thus, genes could be the fundamental unit of selection. New data, possibly from future advanced conservation genomics studies, can confirm or refute how long the confederations actually last (e.g., by measuring linkage disequilibrium across genomes) and how the gene linkages may benefit populations. These highly conserved sequences could derive from marine invertebrates and appear throughout the genome as unexpected syntenic groups or in various, unrelated metabolic pathways or even within noncoding genetic expanses (see below) where the conserved sequences are maintained broadly by mechanisms such as stabilizing selection.

Long-lived conserved sequences may also fit into emerging theoretical frameworks, such as gene essentiality, which refers to those genes that are required for organismal survival (21, 22). It would be worthwhile to fully explore the tie between the evolutionary conservation of genome

sequences and their biological necessity. This underscores another long-standing question of how to bridge genotypes to phenotypes. As recently stated by the Earth BioGenome Project (EBP), "The evolutionary history of point mutations, duplications, deletions, insertions, translocations, inversions, fusions, and fissions is crucial to our understanding of the relationships between genotype and phenotype and the changes in genomic architecture that led to multicellularity and organismal complexity" (23, p. 4325; 24).

We also know that certain molecules, typically ones that are not nucleic acids in nature, command more value from the commercial market than others. For example, the cost for one week of treatment with the chemotherapeutic compound Taxol (paclitaxel), derived from the Pacific yew tree (*Taxus brevifolia*), starts at US$6,000. This is more than three orders of magnitude higher than the cost for the well-known analgesic aspirin ($2–3), which also stems from a plant source. Contrasts between different biomolecules can be taken further by comparing the price for generic pure nucleic acids (a gram of human or salmon DNA can be purchased from Sigma for approximately $120). Sigma-Aldrich currently charges $171 for 10 μg of the protein kinase C modulator bryostatin, derived from the bryozoan *Bugula neritina* (**Figure 1*a***).

However, macromolecular coding sequences underlie the metabolic pathways that enable the biological creation of these unique biochemical compounds. This points to an origin and source of tangible value. Genotypic codes represent a potential for value, wealth, and phenotypic expression.



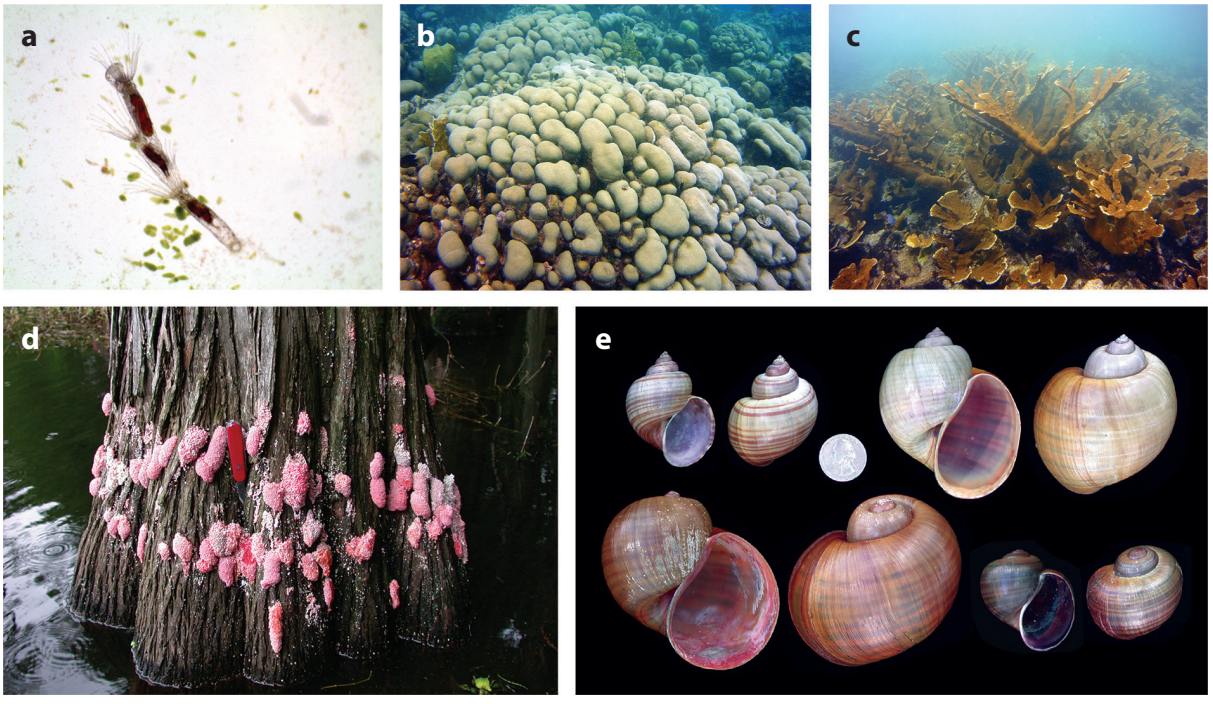


**Figure 1**

Example of invertebrate species in which conservation genomic approaches can be applied. (*a*) Colonial moss animal, bryozoan *Bugula neritina* (photo courtesy of Grace Lim-Fong). (*b*) Massive Caribbean reef-building coral, *Orbicella annularis* (photo by Iliana Baums). (*c*) Branching Caribbean reef-building coral, *Acropora palmata* (photo by Iliana Baums). (*d*) Egg masses of invasive *Pomacea maculata* on cypress trunk in Lake Munson, Florida (photo by Tim Collins). (*e*) Some native and nonnative apple snails in the United States: (*upper left*) nonnative *Pomacea diffusa*, (*lower left*) nonnative *Pomacea maculata*, (*upper right*) nonnative *Pomacea haustrum*, and (*lower right*) native *Pomacea paludosa* (photos by Tim Collins and Tim Rawlings). US quarter for scale.

In recent decades, this promise ushered in a rush of bioprospecting to harness biodiversity at the sequence level with the hope of producing more value downstream.

We can ask many questions in relation to conservation: Should we conserve only specific marker genes that are useful for microevolutionary studies (25), or are there are other sequences with currently unknown function (and value) that also need focus as possible conservation targets? Are the most important features to conserve novel, newly arisen sequences, or sequences that have been maintained across evolutionary time and diverse lineages? How much should the field invest in and rely on the massively parallel HTS approach (26, 27)? Does the sequencing of complete organismal genomes provide useful biological knowledge, allowing conservation biologists to capture all primary coding information? (Answer: epigenetically, no.) Yet perhaps conserving the complete genome sequences would allow society to apply novel techniques in the future to unravel as-yet-unknown gene functions and assess their value even when the individuals and species carrying these genomes are lost.



## Unique Genome Landscapes of Invertebrates

Conserved noncoding sequences (CNSs), also known as conserved noncoding elements (CNEs), ultraconserved elements (UCEs), or ultraconserved regions, should be considered for their potential value and roles in organizing genome structure and gene regulation (28–31). Although many CNEs/CNSs do not produce protein products, CNEs/CNSs can hold genes encoding transcription factors, regulatory genes, or *cis*-regulatory elements/enhancers that drive vertebrate development. Value again is in the eye of the beholder. Fewer examples of CNSs exist within invertebrate genomes at present, so the depauperate data set hinders in-depth comparisons and hypothesis testing. The number and length of UCEs decrease as evolutionary distances between species pairs increase (32). The few studies that have uncovered CNSs stem mostly from model organisms such as *Drosophila* and *Caenorhabditis elegans*, which further points to the need for exploring wider invertebrate biodiversity. Quattrini & Faircloth (33) recently applied targeted bait capture to obtain more UCEs from octocorals. Moreover, many conserved sequences appear to be influenced by strong negative or purifying selection (34).

Conservation genomics may also help identify value in more unique situations, such as nonmodel species or gene regions. Traditional phylogenetics, which plays a large part in conservation studies, typically choose relatively conserved, universally required coding genes, such as ribosomal RNA or cytochrome oxidase, or noncoding intron sequences that all evolve neutrally, to reconstruct phylogenies of taxa with deep evolutionary roots. Recent phylogenomics efforts encompass larger data sets based on whole-genome sequence (WGS) but reveal telling omissions of invertebrate taxa among genomes (8, 35). The pace of invertebrate genome sequencing appears to steadily grow, with at least 578 (or 210 when insects and chelicerates are excluded) genomes now completed (**Table 1**). If there is value in identifying unique populations, substructure has a direct underpinning in the genomes that encode each individual member of the population. Going further, perhaps conservation genomics needs to also develop methods that go beyond the study of orthologous gene systems. The uniqueness of the most interesting species often stems from a new mutation, gene arrangement, or atypical horizontal gene transfer. Genomic novelties should be addressed and distinguished from traditional ortholog analyses recognizing that lack of evolutionary conservation does not reliably indicate lack of function, since certain genome segments evolve rapidly in contrast to CNEs (36).

These examples show that the evolutionary dynamics of genome segments or gene elements can potentially influence important, traditional conservation genomics parameters, namely organismal fitness, adaptation, selection (e.g., disruptive versus stabilizing), genetic drift, and informative

**Table 1 Different phyla and the number of currently available public genome sequences from the National Center for Biotechnology Information**

| Phylum | Number of sequenced genomes[a] |
|---|---|
| Brachiopoda | 2 |
| Ctenophora | 2 |
| Placozoa | 2 |
| Porifera | 2 |
| Annelida | 5 |
| Rotifera | 5 |
| Chordata[b] | 10 |
| Echinodermata | 11 |
| Cnidaria | 15 |
| Mollusca | 22 |
| Platyhelminthes | 22 |
| Nematoda | 91 |
| Arthropoda[c] | 389 |

[a]Data retrieved on September 9, 2018.
[b]This number excludes vertebrates.
[c]Note that Hexapoda and Chelicerata make up 368 genomes of the total 389 arthropod genomes.



neutral markers. Conservation genomics should also address the impact of deleterious variation, how diversity can lead to adaptation, and how genomic monitoring and characterizations could positively affect broader conservation goals (e.g., baseline data for in situ/ex situ genome stocks).

Answers to the questions posed above may lie in present and future methods that allow rapid assessment of diversity at the genome level. For example, physical platforms that recruit marine larvae and cryptic meiofauna have been used to sample diverse ecosystems and provide biomass for HTS and barcoding/metabarcoding (37). Continuing refinement of each sequencing method's hardware and accuracy, choice of genetic markers and relevance, and analytical approaches has accompanied the field's growth since inception (18, 38–40). A detailed inventory of the Earth's total biodiversity will now advance dramatically with the most modern molecular tools.

Conservation goals can advance when we have a better idea of which taxa are present and may need the greatest attention. In this regard, autonomous reef monitoring structures (ARMS) are artificial structures that allow recruitment of larval and encrusting reef organisms. ARMS have allowed sampling of cryptic species on increasingly endangered and deteriorating reef ecosystems. For example, with two-thirds of the total operational taxonomic units, ARMS indicate that the smallest fractions (100–500 um) in two benthic communities (temperate and subtropical) held the highest diversity. This approach has detailed profiles that show "enormous numbers of marine animal species that remain genetically unanchored to conventional taxonomy" (41). The ARMS approach also highlights how new genomic methods can facilitate the characterization and preservation of genomic diversity.

Coral reefs hold some of the highest levels of species biodiversity on the planet (42–44). Further, a single resident marine sponge species can harbor on average 20,000–40,000 operational taxonomic units of prokaryotes (45). Cnidarian and other benthic taxon microbiome estimates of bacterial diversity follow closely behind (46). The recent Anna Karenina principle (whereby "All happy families are alike; each unhappy family is unhappy in its own way") applied to the coral

ecosystem suggests that the dynamics and beta diversity of the symbiotic microbiome in corals mirror the health status of the holobiont (host + symbionts) (47). For symbiotic reef organisms, it is possible that the documented loss of one visible macroorganismal species, owing to habitat degradation or other factors, may actually underestimate the total loss of species richness and genomic diversity contributed by the associated microbial symbionts. Consideration of the holobiont and its integrative nature should hold an imperative position for conservation genomics (48). We are not at the stage of advocating for the conservation of all microbial genomes, though this notion could appreciate as sequencing costs continue to fall and rare biosphere values rise.

In line with these studies, HTS efforts have also begun on meiofauna habitats (49, 50). Research coordination networks (RCN) such as those focusing on eukaryotic biodiversity (e.g., RCN EukHiTS) research using HTS promote precise morphological identifications of intact organisms under the microscope, in conjunction with high-throughput genome and metagenomics sequencing to be carried out later. Regional and local workshops can have resounding effects, benefits, and impacts (51). The Global Invertebrate Genomics Alliance (GIGA) was formed in 2013 as a grassroots community of scientists with specific aims to promote invertebrate genomics studies, many with an implicit focus on conservation (8, 52). The sheer taxonomic breadth and ancient evolutionary depth of invertebrate animals justified the establishment of such a community. As the pace of HTS increases, a precise count of invertebrate genomes may become more difficult, but we have attempted to show the current status in **Figure 2** (and see 53). Well-curated, focused inventories and collaborative efforts can provide clearer project targets, help avoid duplicative efforts, and spur other projects that need further attention.

## TECHNOLOGICAL INNOVATIONS FOR INVERTEBRATE GENOMICS

### Genome Size and Sequencing Platform Considerations

Invertebrate genomes span a wide range of sizes and complexities (54, 55) (see also **http://www.genomesize.com**). To choose the right method for establishing a genome project, one must first determine the genome size of the target species. Many methods are available; typically flow cytometry in comparison to a reference species or Feulgen densitometry (55, 56) is used for genome size estimation. Other methods utilize K-mer coverage curves with either preliminary data (57, 58) or simply information from a closely related species. Empirical measurement using fluorescent dyes and analysis of K-mer information from short-read data usually do not result in the same estimates, but both provide sufficient rough estimates to inform a genome project (59, 60).

A wide range of technologies are now available, which makes even the largest and most difficult-to-assemble genomes tractable (61, 62). For example, short-read technologies, such as Illumina, enable very high coverage of genomic regions with high accuracy. Longer-read technologies with higher error rates, such as PacBio's SMRT, create genome assemblies to span repeat regions and, when combined with Illumina short-read data or other technologies, can produce very high-quality assemblies for large genomes (63). The first step in producing high-quality assemblies, independent of the sequencing technology applied, is the isolation of sufficient amounts of high–molecular weight DNA (HMW DNA). This allows us to obtain the various large size fragments needed for producing high-quality assemblies: shearing long stretches of DNA (in the megabase range) into smaller fragments (from 500 bp to 2–35 Kbp) and much longer fragments for long insert libraries (mate pairs) for Illumina or for producing long reads using PacBio or Nanopore technologies (64, 65). Although in principle a simple procedure, HMW DNA is not always easily obtained from many marine taxa owing to the mechanical treatments or harsh chemicals needed to remove inhibitors present in the tissues (66). This can be circumvented by using DNA isolated from gametes or gonad tissues, which usually contain a higher ratio of DNA to mass, thus resulting

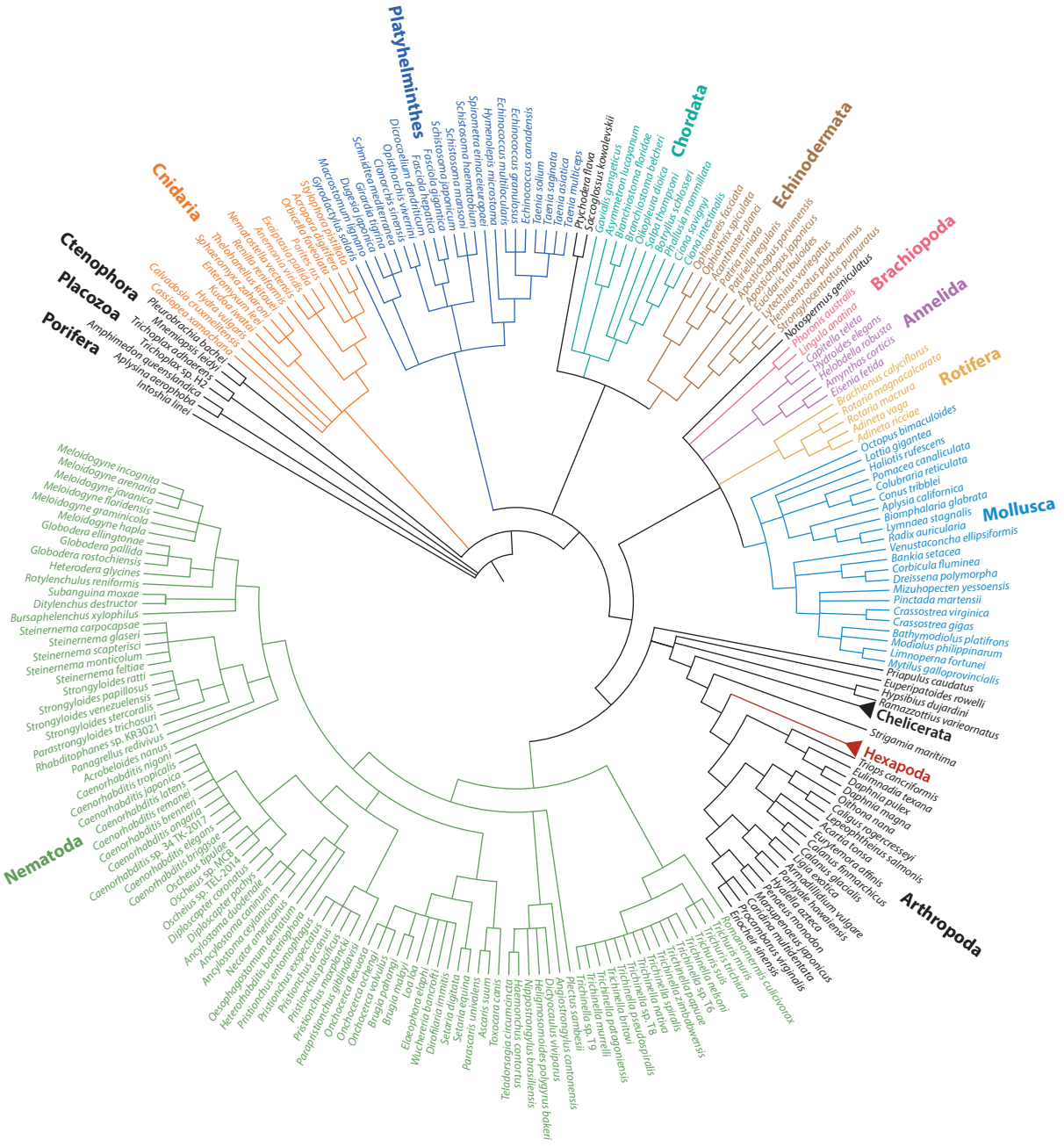


**Figure 2**

Currently sequenced taxa available from the National Center of Biotechnology Information (NCBI). The tree clearly shows the uneven phylogenetic representation among currently available genomes. The tree is constructed according to NCBI taxonomic data. The information was retrieved from **ftp://ftp.ncbi.nlm.nih.gov/genomes/GENOME_REPORTS/eukaryotes.txt** and parsed into a Newick format using the ETE3 package. For other current metazoan trees or alternative phylogenies, we refer the reader to References 8, 23, and 35. Due to the overrepresentation of certain groups, some leaves have been collapsed. Raw data for the construction of this tree are available in **Supplemental Table 1** with additional details. Figure adapted from original construction using the iTOL online tool (**https://itol.embl.de/**). The colors on the branches do not have taxonomic information and are shown only for clarity.



Supplemental Material >

in better-quality starting material (66). With foresight, investigators can also isolate intact HMW DNA with long-standing pulsed-field gel electrophoresis methods (67), and unused gel plugs of DNA may be safely stored for future use.

In planning for a genome sequencing project, sufficient amounts of DNA must be secured from a single individual, enough to enable the construction of libraries for all the stages of the sequencing project and to enable the use of future technologies that might develop after starting to sequence a particular taxon. Currently it is possible to sequence entire genomes using long-read methods such as PacBio and NanoPore; however, the genomes still require multiple steps of polishing and error correction to be accurate enough for most useful endeavors. Thus, it is still important, especially for large metazoan genomes, to sequence at sufficient depth using highly accurate short-read methods to be able to complement and correct the long-read assemblies (61, 68, 69). The optimization of how much to sequence using short reads in addition to long reads is dependent on the genome size in question and the budget available. For example, the minimum coverage needed for a PacBio assembly is approximately 35×. However, owing to the presence of heterozygous alleles in the data of sexually reproducing animals (unless specifically haploid DNA was obtained), the amount of PacBio data needed for self-error correction, i.e., by aligning all the PacBio reads to themselves, tends to be much higher (70, 71). Some of these issues may be eliminated with future advances in both the sequencing technology itself and the algorithms used in the assembly.

## Addressing Ploidy

Many invertebrate species, in addition to having large, complex, AT-rich genomes, also are polyploid (72). In addition, some of these taxa are parthenogenetic; thus, multiple variants of ploidy can exist within a population (73, 74). The first step in identifying whether an animal is polyploid starts with a count of the number of chromosomes in the species compared with the chromosome numbers of related species. This can be achieved using both flow cytometry methods and karyotyping techniques. These techniques are straightforward for large animals but problematic with smaller non-model species, which pose challenges in determining the correct ploidy or genome size (74). Indeed, karyotyping, the counting and visualization of an organism's complete chromosome complement, has become a lost art in the age of molecular genetics. The number of genetics papers studying karyotypes appeared to peak in the 1980s (75, 76) (**Figure 3**). Although not as common as plants, many metazoan lineages do have well-documented polyploid species. Many bivalve and gastropod species are polyploid (77, 78), for example, in the genus *Corbicula*. Different levels of ploidy up to hexaploidy are reported in the marine clams of the genus *Lasaea* (79). Some freshwater clam species, e.g., *Sphaerium corneum*, have well-documented polyploidy (80). In gastropods, most documented polyploidy cases come from the Pulmonata, such as *Gyraulus circumstriatus* (81), and other taxa within the Planorbidae, such as *Bulinus* (77). Other groups, such as annelids and crustaceans, also possess many polyploid lineages (82, 83).

Complex polyploid genomes present a challenge for genome assembly, known as the minimum error correction problem (84–86). This issue stems from the fact that the assembler software must be able to differentiate between sequencing errors in the data and true haplotypes. In the case of diploid genomes, there are only two haplotypes. However, in a polyploid situation, there may be three or more different haplotypes for every genomic locus. Thus, enough sequences must be generated to ensure accurate identification of all haplotypes. Additionally, if aggressive error correction is used to collapse haplotypes, the assembly generated will likely resemble a mosaic of unique regions in the genome with contigs generated from multiple chromosome regions (85). Assembly of large polyploid genomes has been improving with the advent of long-read sequencing

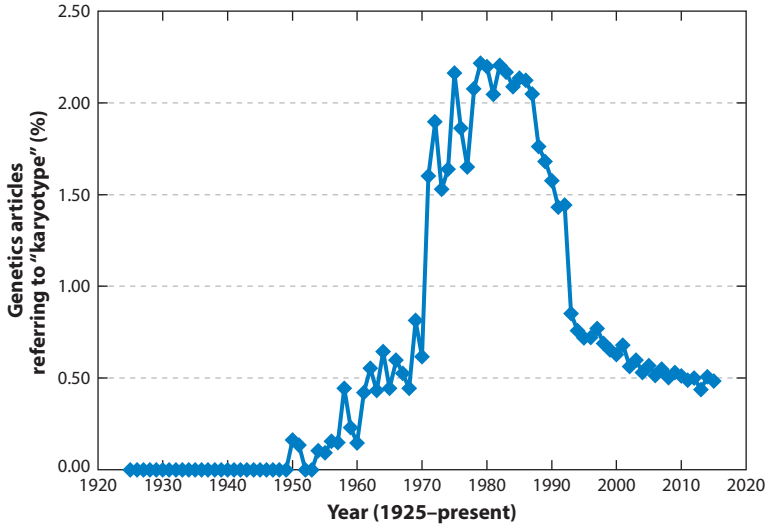


**Figure 3**

Percentage of genetics articles per year citing organismal chromosome numbers or performing "karyological" studies from 1925 to the present. Based on a Web of Science literature survey of 159,103,651 total genetics articles.

technologies and is currently an active area of research in algorithm development. Whereas little attention has been given to polyploid animal genomes, many advances have been made in plant genomics, where polyploidy is more common (87, 88).

However, encountering polyploid animal genomes is likely to become more frequent as more invertebrate genomes are sequenced. To date, mostly small invertebrate genomes have been sequenced, with limited taxonomic coverage (8, 35, 89).

Several methods have been developed to sequence highly polyploid genomes. Currently, most involve using long-read methods combined with optical mapping and other long-range scaffolding methods (87, 90, 91). Other employed methods make use of population-level data to generate dense genetic maps. For example, the software POPSEQ can arrange contigs according to an anchored genetic map, which can lead to a chromosome-level assembly of complex and repetitive genomes (92, 93). Population data in many cases might not be available or easy to obtain. Methods such as optical mapping require very HMW DNA because they are based on building high-resolution restriction maps from a single molecule of DNA. Optical mapping methods are starting to be more common not only in aiding the assembly of complex polyploid genomes but also in verifying correct scaffolding and ordering of contigs in assembled genomes (94). For some of the hardest-to-assemble genomes, a combination of all mentioned methods is needed: population-level data, long-read sequencing, short-read high coverage, and optical mapping with long-range scaffolding (95).

## New Advances in Scaffolding Methods

After a genome assembly is generated, the next challenge is to scaffold and stitch the smaller contigs into larger scaffolds and eventually chromosomes. Most assemblers contain a scaffolding step, which can use paired-end information to overcome small gaps (96). However, additional

scaffolding must be done with additional long reads or mate-pair libraries that can span longer repeats or polymorphic regions. There are few specialized scaffolding software programs compared with assemblers (97, 98), and new advances have been made in generating specialized libraries that enable seamless scaffolding of contigs. For example, the Molecule approach, which Illumina eventually acquired, allows the construction of long reads from many small reads via the addition of sequential indexes currently available as the TruSeq Synthetic Long-Read DNA library (99). Usually the reads are approximately 30–50 kb or more. A similar approach is 10X, which uses a combination of emulsion PCR to introduce barcodes from a very large pool of barcodes onto gel-coated beads containing single sheared DNA molecules. After mass sequencing of the short reads with the barcode sequences, they are stitched into high-quality 50-kbp fragments, which can later be assembled or used to scaffold other assemblies (100–102). Other synthetic long-read methods continue to emerge as well and are based on the same principle of stringing together many short reads with a known order reflecting the original DNA sequence (100, 101). Synthetic long-read technologies compete with true long-read technologies, such as PacBio and Oxford Nanopore, which generate long reads as their de facto output. Long-read technologies, however, tend to have high error rates, which requires more coverage to generate useful data that can be used in the error-correction process (69). Although in the future true long-read technologies will continue to mature and improve sequencing accuracy (103), synthetic long-read methods are gaining traction and offer some advantages over other methods owing to the low cost of the equipment involved, which builds upon widespread short-read sequencers, compared with more expensive technologies such as PacBio and Nanopore (if the same amount of output is desired). Conversely, other methods work by inferring long-range information about the DNA. For example, the Hi-C method, which is based on proximity ligation, establishes a chromosome contact map used for investigating the 3D structure of the chromatin (104) in genome assembly (105), currently available through in-house-made protocols or as kits such as the Proximo kit. The Hi-C method can also be made available as a commercial service.

Tissues compose the starting material for Hi-C approaches, instead of the HMW DNA used in other methods, which might make them attractive for certain applications where HMW DNA is hard to obtain (95). Similarly, the Chicago libraries, together with the HighRise software, can be used to generate long-range information and correct genome scaffolds by using in vitro–reconstituted chromatin instead of whole chromosomes, as in Hi-C (106). This method can produce chromosome-level assemblies in some cases, depending on the state of pre-scaffold assembly. The methods have been published but are also available as a commercial service through Dovetail Inc. The advantage of the Chicago method stems from its application of HMW DNA without having to isolate intact chromosomes. A critical limiting factor for all of these methods is the availability of high-quality HMW DNA or intact tissues.

## Hard-to-Get Samples, Low-Quality DNA, and the Promise of Single-Cell Sequencing

In many cases, HMW DNA can be difficult to obtain. This may be due to the absence of optimized protocols for the isolation of DNA from a target species or the presence of inhibitors and unique secondary metabolites in the tissues of these animals, leading to the employment of harsh chaotropic agents or multiple extraction steps that can compromise DNA integrity. However, because technologies such as Illumina can use very short reads (e.g., as low as 35 bp), sequencing genomes with degraded DNA becomes possible. Advances in this area come from sequencing efforts with museum samples (107, 108), preserved medical tissues (109, 110), and ancient DNA methodologies (111, 112). The ability to generate genomic resources from formaldehyde- and

ethanol-preserved museum samples opens the door to asking important questions about extinct and inaccessible species. Usually some additional steps are needed to repair the DNA, and kits for this are available (113). After conventional short-read sequencing, rather than assembly, the DNA is typically mapped to a closely related species to build a reference-based assembly (111). However, in many cases, if the DNA is not damaged, de novo assembly is possible and can be scaffolded to close relatives. In most low-quality DNA samples, contamination also tends to be an issue. Intracellular symbionts that cannot be mechanically separated from target tissues lead to a mixed DNA sample, with symbiont sequences sometimes greatly outnumbering the target sequence (114). Potential methods to counter contamination are the use of flow cytometry and single-cell sequencing (115). In one modification, flow cytometry can sort nuclei rather than cells, segregating host nuclei and their DNA apart from intracellular symbionts (116–118). Other capture-based methods for enriching a target DNA population are also available, with varying degrees of success (119).

## Importance of Transcriptomics in Informing Genome Sequencing and Annotation

Although genome sequencing can provide ample information about a target species, in most cases genome projects rely heavily on transcriptome data to verify the completeness of assemblies and to annotate genomes (120). This is very important in non-model systems in which data from closely related species are usually absent or incomplete (121), thus limiting the applicability of using public databases to interrogate the coding regions in the genome. Therefore, it is of utmost importance to be able to fully annotate a genome and conduct transcriptome sequencing. A wide range of transcriptome samples from various stress and developmental conditions can ideally provide the most comprehensive coverage of genome coding regions (122). When such high-quality data sets are combined with ab initio gene-prediction methods, high-quality annotations will result (123). Using transcriptomic data trains gene-prediction software, in addition to verifying the accuracy of predictions. In addition, transcriptomic data can improve the annotations for areas of the genome that might be misassembled owing to high repeat content, limited coverage, or gaps (124).

## Genome Finishing

After generating a draft genome, scientists may opt to produce a finished genome. A finished genome is one in which the remaining gaps and difficult-to-assemble regions are sequenced manually with various methods, such as bacterial artificial chromosome (BAC) library sequencing or genome walking approaches (125). Few genome projects invest time in this task, because it can be time consuming and result in little new information. However, as genome sequencing technologies improve, more draft genomes are nearing closure. By combining long-read methods (true and synthetic) with long-range mapping, such as Hi-C, optical mapping, and Chicago libraries, it is possible to finish genomes without the need for manual genome walking approaches. This reduces the cost and labor needed to obtain high-quality complete genomes (63, 95, 126).

## Upgrading Current Genomes Using New Methods

These new methods allow for high-quality scaffolding and assembly and also provide the opportunity to improve the genomes of many non-model species that were sequenced initially using a combination of Sanger-based methods and early short-read technologies. In addition, high-quality genomes from closely related species can be used to improve the assemblies of previously sequenced species, producing new, improved assemblies for a particular lineage. More contiguous

assemblies can be achieved by comparative assembly approaches. Briefly, syntenic regions are identified between closely related taxa and are used to create an assembly graph, which resolves contig order or creates longer scaffolds according to conserved blocks across the related genomes (127–129). In the case of sister species or various ecotypes of a particular taxon for which genome resources are available, it is possible to assemble a *pangenome*. The pangenome encompasses all of the variations present in the different genomes and the core genome, combining all the genomic regions that are shared between the species within a particular group, while also highlighting ecotype or species-specific genomic regions (130). This information can then be linked to various phenotypic traits or other metadata. Pangenomes have been widespread in bacteria (131, 132) and in larger genomes of plants (133) but in only a few animals. However, as new data start to accumulate from many closely related taxa, we will likely see more efforts toward invertebrates. Specifically, pangenomes might highlight the need to conserve certain ecotypes of a particular species owing to the uniqueness of their genetic makeup.

## Selecting a Representative Genome

Although future plummeting of sequencing costs will enable clade- and population-level genome sequencing, currently most projects focus on a single individual. Choosing a particular individual for genome sequencing is an important step in the early planning stages. In many cases, parameters revolve around the scientific questions being asked with regard to a specific phenotype. However, other technical reasons for why a particular individual might be a better candidate may be considered. For example, if a population is known to be highly outbred, potential inbreeding for some generations may be implemented to produce individuals that are more amenable to genome sequencing (122). However, more biologically relevant genome information is likely to come from sequencing more than one individual from a population or from combining a single individual genome with other genetic information from RNA sequencing data, exome capture, or restriction site–associated sequencing (RAD-Seq) data from multiple individuals (134).

## Hosting and Data Release

The final step in genome sequencing, which usually does not receive sufficient foresight, is planning the data release and how to present various relevant data about a genome to the wider community. Most journals request the submission of sequenced genomes to a public data repository, such as the National Center for Biotechnology Information (NCBI) or European Molecular Biology Laboratory of the European Bioinformatics Institute (EMBL-EBI). In addition to the completed assemblies, raw data usually must be deposited as well. However, some data, such as optical maps, might not be accepted at some repositories and must be disseminated through other methods. Although NCBI and EMBL-EBI provide adequate access to the assembled genomes and even provide annotation pipelines, there is still a need for integrated and functional portals that link multiple data sources to the genomic data, such as expression data, population genetic data, or other phenotypic traits (135). Many large efforts seek to accomplish this, such as the GMOD database, which provides the GBROWSE (136) and JBROWSE (137) genome browsers, in addition to Galaxy (138) and Intermine (139). These platforms aim to provide a user-friendly database system that can be deployed by a community of scientists to provide a specialized interface to the genomic data generated that might not be available via NCBI or EMBL standard features. GIGA has partnered with Compagen.org and Reefgenomics.org to realize these goals (8, 140). One downside can be obsolescence, as databases can age or require curators and, owing to the lack of funding, become outdated. Potentially a more centralized funded effort, such as NCBI,

EMBL, or Japan's GenomeNet, can be constructed to provide access to these specialized platforms through container-based systems (141). These will provide potential longevity to multiple projects with a solid single technical platform (142). This reduces the cost for individual scientists (143) while still giving the flexibility to deploy and develop specific software features that can be useful for a particular taxon (144).

## CASE STUDIES WITH INVERTEBRATE CONSERVATION GENOMICS

In the following sections, we describe specific invertebrate species in which genomics approaches provide insights concerning the population structure and population history of species of conservation interest and clarify host–symbiont interactions that impact the holobiont species. We further suggest ways in which genomic approaches, in combination with other emerging techniques, may be useful for controlling populations of invasive species.

### Coral Conservation Genomics

Populations of Caribbean reef-building corals, such as *Acropora palmata* (**Figure 1*c***), *Acropora cervicornis*, and the *Orbicella* species complex, have declined in recent decades owing to anthropogenic impacts, disease, and temperature-induced bleaching events (145, 146; see sidebar titled *Acanthaster planci*), leading to their current status as a federally listed threatened species under the US Endangered Species Act. It is critical for the conservation of these species to maintain and enhance their capacity to adapt to changing environments. In long-lived corals, this capacity depends on the presence of sufficient levels of standing genetic diversity in functional traits in the population on which selection can act.

Genetic diversity of a population varies with life history traits, population size, and history. Genetic diversity can be defined on several levels in organisms that have sexual and asexual reproductive modes, such as corals and plants (147). Genotypic diversity refers to the number of genets in a population as defined via multilocus genotyping (148). Genets are the result of sexual reproduction. Each genet may consist of many ramets (colonies) that were the result of asexual processes such as fragmentation (149). Genetic diversity (or gene diversity) refers to the amount of variation on the level of individual genes in a population. Genetic diversity may be expressed as heterozygosity or allelic richness.

### *ACANTHASTER PLANCI*

One of the most alarming predators of reef-building corals in the Indo-Pacific is the crown-of-thorns (COTS) starfish (*Acanthaster planci*) (183). This species can reproduce and spread quite rapidly, leading to outbreaks that can decimate vast areas of healthy coral in a matter of days. The sequencing of two individuals from different locations (Japan and Australia) produced a high-quality reference genome with 99% sequence similarity (184), which can now be used to evaluate local adaptation in different subpopulations during regional outbreaks (185). The genome of this species is conserved relative to other deuterostome genomes; nevertheless, two groups of ependymin-related and G-protein-coupled receptor proteins that act as signaling and chemoreceptor molecules have undergone gene family expansions (185). These COTS-specific molecules may prove to be important in regulating behavioral cues related to reproduction and predator avoidance, making these gene families ideal target candidates to design mitigation or eradication strategies for COTS outbreak control.

Single-nucleotide polymorphisms (SNPs) are ubiquitous throughout the genome, yet each locus has a maximum of four alleles (the four bases). This is in contrast to microsatellite loci that consist of tandem repeats, in which allelic variation is determined by the number of tandem repeats, and thus allelic variation can be large. The limited number of alleles at each SNP locus requires a larger number of loci to be assayed to achieve the same ability to detect population genetic structure as a panel of microsatellite loci (150–152). The advent of reduced representation sequencing methods has made it possible to develop and assay a large number of single-nucleotide variant (SNV) loci at a reasonable cost (153). However, some of these methods (collectively called RAD-tag) have the disadvantage that missing data are common. With the advent of next-generation sequencing and the availability of genomic resources, it is now possible to design high-resolution gene chips that provide highly standardized and genome-wide genotype information, similar to the gene chips commonly applied in human health and crop sciences (154–156).

Genome-wide SNP data can be used to resolve species, define populations, identify genets, and interrogate coding and noncoding portions of the genome for allelic variation; all these data can inform the design of conservation and restoration programs. SNPs have higher power to detect subtler levels of population differentiation—this was the case when we turned to SNP genotyping in *A. palmata* to uncover additional population structure compared with previous microsatellite-based analyses (157, 158). These findings are directly applicable to the definition of propagule transfer zones.

***Orbicella* spp.** The *Orbicella* coral species complex is composed of three species (*Orbicella annularis*, *Orbicella faveolata*, and *Orbicella franksi*) (**Figure 1*b***). These species diverged in the Pliocene (159) and are ecologically dominant on tropical Western Atlantic reefs (160). Throughout their range, *Orbicella* spp. are distributed along a depth gradient, with *O. annularis* generally found in shallower areas of the reef, *O. franksi* at greater depths, and *O. faveolata* spanning the depth range of the other two (160–163). Because *Orbicella* spp. are major reef-building taxa in the Caribbean, their fossil record has been studied extensively (159, 162). Genomic data from the three species have enabled the reconstruction of demographic fluctuations in each lineage, corroborating paleontological observations (164). Coalescent analysis of whole genomes from the three species revealed that *O. annularis* and *O. faveolata* underwent major population expansions in shallow areas when the pillar coral *Orbicella nancyi* went extinct approximately 100,000 years ago (164). Light niche partitioning in *Orbicella* species is linked to their associated microbiome (i.e., *Symbiodinium* and bacteria), and therefore genotype-by-genotype interactions probably define local adaptation (F.J. Pollock, C. Prada, T. López-Londoño, S. Roitman, D. Levitan, N. Knowlton, R. Iglesias-Prieto, and M. Medina, manuscript in preparation). Thus, the use of genome data from the members of a coral holobiont will be most helpful when defining conservation strategies of locally adapted genotypes.

## Caribbean Acroporids: *Acropora palmata* and *Acropora cervicornis*

The genus *Acropora* contains more than 100 species of reef-building corals, and most of this diversity is found in the Pacific, where species occur in sympatry over large geographic areas. In the western North Atlantic and Caribbean, diversity is reduced to only two species, *A. palmata* and *A. cervicornis*, both of which are listed as threatened under the US Endangered Species Act. Acroporids are broadcast spawners that often reproduce in mass-spawning events, providing opportunities for hybridization. The best-studied *Acropora* hybrid system is the Caribbean species pair that forms early-generation hybrids. Recently, genomes of both parent species have become available (**www.baumslab.org**), enabling in-depth studies of the hybridization dynamics in this system (156). Of particular conservation concern is determining the frequency and directionality

of hybridization because of the potential for genetic swamping of one of the parent species (6). However, hybrids have a wide range of novel growth morphologies and grow over large depth ranges, including very warm and shallow waters, and might provide an avenue of evolutionary rescue (166). Haplotype block analysis has proven to be a powerful method for determining hybrid ancestry in non-model species without pedigree data. Several computational tools, including SABER (167), HPMIX (168), and a program written for the Galaxy web server (169), produce individual hybridization maps showing the ancestral makeup of chromosomal regions. Most allow SNP data to be unphased, as is the case for most non-model species. The power of this approach to accurately estimate hybridization history is understood in considerable detail (170). Genome-enabled studies of coral hybridization dynamics in this and other coral species complexes will provide essential information for conservation (171, 172).

***Acropora millepora.*** Standing genetic variation that confers thermal tolerance along a latitudinal thermal gradient has been revealed by population genomic analysis in the Great Barrier Reef *Acropora millepora* (173). These genomic data, along with dispersal biophysical models, have been used to predict the future of this species under increasing sea-surface temperatures. The study demonstrates that there is a high chance for migration of thermal-tolerant individuals to enable local population adaptation to climate change. However, these populations will also become more sensitive to extreme random fluctuations in temperature. The use of CRISPR/Cas9 has also been reported in the coral *A. millepora* (174). Single-guide RNA was microinjected in one-cell zygotes after mass spawning. Three genes encoding fibroblast growth factor 1a, green fluorescent protein, and red fluorescent protein were successfully mutated, and the mutations were maintained after multiple cell divisions. This new technology may eventually lead to genetically engineered organisms that may be able to cope with different environmental insults, such as those brought on by climate change.

## Apple Snails

Apple snails in the genus *Pomacea* are large, primarily herbivorous freshwater snails native to South and Central America and the Caribbean, with one species native to North America (175) (**Figure 1*d*,*e***). Two species, *Pomacea maculata* and *Pomacea canaliculata*, are of particular interest as troublesome invasive species, listed in the world's top 100 worst invasive species in the Global Invasive Species Database for the threat the snails pose to native ecosystems, species, and agriculture (176). In some Southeast Asian countries, for example, apple snails are the number-one pest of rice crops. In addition, *P. maculata* is an intermediate host to human parasites such as the nematode lungworm *Parastrongylus* (= *Angiostrongylus*) *cantonensis* and digenetic trematode parasites, including *Schistosoma*. All of this suggests that developing genomic methods to control nonnative populations of these species would be of great value. Apple snails are from a clade that is most closely related to the clade of most modern marine snails, but its genome is atypically small, at 600 Mbp. Many snails have much larger genomes, some owing to genome duplications, and have genome sizes that range from 1.5 to 5 Gbp. Moreover, an online transcriptome database includes eight species of apple snails, including *P. canaliculata* and *P. maculata* (177). The combination of the compact size of the apple snail genome and the available transcriptome database should make apple snails an excellent model system in which to develop CRISPR/gene-driven approaches to controlling populations of a troublesome invasive species and human parasite vector. Apple snails are centrally located on the broader gastropod phylogeny, so the features determined from this genome would likely prove useful for comparative purposes with other model gastropods with much larger genomes, such as *Biomphalaria*, *Aplysia*, and *Conus*.

## Abalone

Abalone are marine gastropods in the genus *Haliotis*, with approximately 50+ species. This is an economically important seafood delicacy around the world. The intensive harvesting of natural populations in the twentieth century has triggered a tightly controlled fishery and increased aquaculture farming for many species. RAD-Seq analysis of populations of the Northeastern Pacific *Haliotis fulgens* revealed a panmictic species throughout its range, suggesting that local broodstocks and limited translocation might not be necessary for restoring natural populations or the restocking of natural genotypes in hatchery populations (178). However, these authors (179) point out the need to still evaluate genetic variants for local adaptation. A study of the commercially important greenlip abalone (*Haliotis laevigata*) in Australia provided evidence of local adaptive divergence in a widespread population of a broadcast spawning species. This study, also based on RAD-Seq, found that whereas the majority of the more than 9,000 SNPs in their data set appeared to be selectively neutral, 323 were potentially adaptive (179, 180). These potentially adaptive SNP loci were correlated with water temperature and oxygen concentration, and gene annotation showed that three-quarters of these SNPs were associated with genes related to high temperature and low oxygen tolerance. These results have significant implications both for managing this species and in considering the potential effects of changing climate for this species and other similarly distributed abalone. Finally, a 300× coverage draft of the 1.8-Gb Asian abalone *Haliotis discus hannai* genome has now been published (181), and endangered white abalone (*Haliotis sorenseni*) genome sequences may not be far behind (K.S. Aquilino & S. Boles, unpublished results). This draft will be a great resource for abalone fisheries and aquaculture management.

Although daunting challenges remain in understanding and preserving invertebrate biodiversity, we have at our disposal increasingly powerful genomic tools to help us understand the biology, physiology, and population structure of species in ways that could not be imagined for non-model organisms just a few years ago (6, 18, 24, 38, 182). The accelerating pace of advances in sequencing technology and methods for genetic manipulation may also enable us to address issues with invasive species that have largely been considered hopeless. Perhaps the largest impediments remaining at this point are the needs for further development of the analytical tools and pipelines to help make sense of this flood of data, sustaining traditional taxonomic training, and the coordination of multidisciplinary research communities to synthesize lasting knowledge.

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

We wish to thank Grace Lim-Fong for her photo of *Bugula neritina*, and an anonymous reviewer for many helpful comments. J.V.L. would like to thank Nova Southeastern University for an academic sabbatical that allowed completion of the manuscript.

## LITERATURE CITED

1. Crutzen PJ. 2002. Geology of mankind. *Nature* 415:23
2. Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, et al. 2017. DNA sequencing at 40: past, present and future. *Nature* 550:345–53
3. O'Brien SJ. 1994. A role for molecular genetics in biological conservation. *PNAS* 91:5748–55

4. Allendorf FW, Hohenlohe PA, Luikart G. 2010. Genomics and the future of conservation genetics. *Nat. Rev. Genet.* 11:697
5. Allendorf FW, Luikart G, Aitken SN. 2012. *Conservation and the Genetics of Populations*. Hoboken, NJ: John Wiley & Sons. 2nd ed.
6. Baums IB. 2008. A restoration genetics guide for coral reef conservation. *Mol. Ecol.* 17:2796–811
7. Zhang Z-Q. 2011. *Animal Biodiversity: An Outline of Higher-Level Classification and Survey of Taxonomic Richness*. Longwood, FL: Magnolia
8. GIGA Community Sci. 2013. The Global Invertebrate Genomics Alliance (GIGA): developing community resources to study diverse invertebrate genomes. *J. Hered.* 105:1–18
9. Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R. 2013. *Molecular Biology of the Gene*. London: Pearson Educ.
10. Knowlton N. 2010. *Citizens of the Sea: Wondrous Creatures from the Census of Marine Life*. Washington, DC: Natl. Geogr.
11. Martini S, Haddock SH. 2017. Quantification of bioluminescence from the surface to the deep sea demonstrates its predominance as an ecological trait. *Sci. Rep.* 7:45750
12. World Resour. Inst., World Conserv. Union, UN Environ. Progr. 1992. *Global Biodiversity Strategy: Guidelines for Action to Save, Study, and Use Earth's Biotic Wealth Sustainably and Equitably*. Washington, DC: World Resour. Inst., World Conserv. Union, UN Environ. Progr.
13. Loker ES, Adema CM, Zhang SM, Kepler TB. 2004. Invertebrate immune systems—not homogeneous, not simple, not well understood. *Immunol. Rev.* 198:10–24
14. Rast JP, Smith LC, Loza-Coll M, Hibino T, Litman GW. 2006. Genomic insights into the immune system of the sea urchin. *Science* 314:952–56
15. Sodergren E, Weinstock GM, Davidson EH, Cameron RA, Gibbs RA, et al. 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314:941–52
16. Pearson RG. 2016. Reasons to conserve nature. *Trends Ecol. Evol.* 31:366–71
17. Fothergill-Gilmore LA, Michels PA. 1993. Evolution of glycolysis. *Prog. Biophys. Mol. Biol.* 59:105–235
18. Richards S. 2015. It's more than stamp collecting: how genome sequencing can unify biological research. *Trends Genet*. 31:411–21
19. Goldman AD, Bernhard TM, Dolzhenko E, Landweber LF. 2012. LUCApedia: a database for the study of ancient life. *Nucleic Acids Res*. 41:D1079–D82
20. Dawkins R. 1976. *The Selfish Gene*. New York: Oxford Univ. Press
21. Yu H, Greenbaum D, Xin Lu H, Zhu X, Gerstein M. 2004. Genomic analysis of essentiality within protein networks. *Trends Genet*. 20:227–31
22. Rancati G, Moffat J, Typas A, Pavelka N. 2018. Emerging and evolving concepts in gene essentiality. *Mol. Ecol. Resour.* 19:34–49
23. Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, et al. 2018. Earth BioGenome Project: sequencing life for the future of life. *PNAS* 115:4325–33
24. Kumar S, Dudley JT, Filipski A, Liu L. 2011. Phylomedicine: an evolutionary telescope to explore and diagnose the universe of disease mutations. *Trends Genet*. 27:377–86
25. Avise J. 2004. *Molecular Markers, Natural History and Evolution*. Sunderland, MA: Sinauer Assoc. 2nd ed.
26. Heather JM, Chain B. 2016. The sequence of sequencers: the history of sequencing DNA. *Genomics* 107:1–8
27. Gong J, Pan K, Fakih M, Pal S, Salgia R. 2018. Value-based genomics. *Oncotarget* 9:15792–815
28. Katzman S, Kern AD, Bejerano G, Fewell G, Fulton L, et al. 2007. Human genome ultraconserved elements are ultraselected. *Nat. Biotechnol.* 317:915
29. Vavouri T, Lehner B. 2009. Conserved noncoding elements and the evolution of animal body plans. *BioEssays* 31:727–35
30. Polychronopoulos D, Weitschek E, Dimitrieva S, Bucher P, Felici G, Almirantis Y. 2014. Classification of selectively constrained DNA elements using feature vectors and rule-based classifiers. *Genomics* 104:79–86
31. Polychronopoulos D, King JWD, Nash AJ, Tan G, Lenhard B. 2017. Conserved non-coding elements: Developmental gene regulation meets genome organization. *Nucleic Acids Res*. 45:12611–24

32. Ryu T, Seridi L, Ravasi T. 2012. The evolution of ultraconserved elements with different phylogenetic origins. *BMC Evol. Biol.* 12:236
33. Quattrini AM, Faircloth BC. 2018. Universal target-enrichment baits for anthozoan (Cnidaria) phylogenomics: new approaches to long-standing problems. *Mol. Ecol. Resour.* 18:281–95
34. Casillas S, Barbadilla A, Bergman CM. 2007. Purifying selection maintains highly conserved noncoding sequences in *Drosophila*. *Mol. Biol. Evol*. 24:2222–34
35. Dunn CW, Ryan JF. 2015. The evolution of animal genomes. *Curr. Opin. Genet. Dev.* 35:25–32
36. Pang KC, Frith MC, Mattick JS. 2006. Rapid evolution of noncoding RNAs: Lack of conservation does not mean lack of function. *Trends Genet*. 22:1–5
37. Plaisance L, Knowlton N, Paulay G, Meyer C. 2009. Reef-associated crustacean fauna: biodiversity estimates using semi-quantitative sampling and DNA barcoding. *Coral Reefs* 28:977–86
38. Mardis ER. 2017. DNA sequencing technologies: 2006–2016. *Nat. Protoc.* 12:213–18
39. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, et al. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551:457–63
40. Parker J, Helmstetter AJ. 2017. Field-based species identification of closely-related plants using real-time nanopore sequencing. *Sci. Rep.* 7:8345
41. Leray M, Knowlton N. 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *PNAS* 112:2076–81
42. Reaka-Kudla ML. 1997. The global biodiversity of coral reefs: a comparison with rain forests. In *Biodiversity II: Understanding and Protecting Our Biological Resources*, ed. ML Reaka-Kudla, DE Wilson, EO Wilson, pp. 83–108. Washington, DC: Natl. Acad. Press
43. Knowlton N, Brainard RE, Fisher R, Moews M, Plaisance L, Caley MJ. 2010. Coral reef biodiversity. In *Life in the World's Oceans: Diversity, Distribution, and Abundance*, ed. AD McIntyre, pp. 65–74. Oxford, UK: Wiley-Blackwell
44. Hughes TP, Kerry JT, Baird AH, Connolly SR, Dietzel A, et al. 2018. Global warming transforms coral reef assemblages. *Nature* 556:492–96
45. Webster NS, Taylor MW, Behnam F, Lucker S, Rattei T, et al. 2010. Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environ. Microbiol.* 12:2070–82
46. Sunagawa S, Woodley CM, Medina M. 2010. Threatened corals provide underexplored microbial habitats. *PLOS ONE* 5:e9554
47. Zaneveld JR, McMinds R, Vega Thurber R. 2017. Stress and stability: applying the Anna Karenina principle to animal microbiomes. *Nat. Microbiol.* 2:17121
48. Peixoto RS, Rosado PM, Leite DC, Rosado AS, Bourne DG. 2017. Beneficial Microorganisms for Corals (BMC): proposed mechanisms for coral health and resilience. *Front. Microbiol.* 8:341
49. Brannock PM, Halanych KM. 2015. Meiofaunal community analysis by high-throughput sequencing: comparison of extraction, quality filtering, and clustering methods. *Mar. Genom.* 23:67–75
50. Carugati L, Corinaldesi C, Dell'Anno A, Danovaro R. 2015. Metagenetic tools for the census of marine meiofaunal biodiversity: an overview. *Mar. Genom.* 24(Pt. 1):11–20
51. Bik HM. 2017. Let's rise up to unite taxonomy and technology. *PLOS Biol*. 15:e2002231
52. Voolstra CR, GIGA Community Sci., Wörheide G, Lopez JV. 2017. Advancing genomics through the Global Invertebrate Genomics Alliance (GIGA). *Invert. Syst*. 31:1–7
53. Wikipedia. 2013. List of sequenced animal genomes. *Wikipedia.org.* **https://en.wikipedia.org/wiki/List_of_sequenced_animal_genomes**
54. Fielman KT, Marsh AG. 2005. Genome complexity and repetitive DNA in metazoans from extreme marine environments. *Gene* 362:98–108
55. Gregory TR, ed. 2005. *The Evolution of the Genome*. Burlington, MA: Elsevier Acad.
56. Kocot KM, Jeffery NW, Mulligan K, Halanych KM, Gregory TR. 2015. Genome size estimates for Aplacophora, Polyplacophora and Scaphopoda: small solenogasters and sizeable scaphopods. *J. Molluscan Stud.* 82:216–19
57. Hardie DC, Gregory TR, Hebert PD. 2002. From pixels to picograms: a beginners' guide to genome quantification by Feulgen image analysis densitometry. *J. Histochem. Cytochem.* 50:735–49

58. Liu B, Shi Y, Yuan J, Hu X, Zhang H, et al. 2013. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. arXiv:1308.2012 [q-bio.GN]
59. Sun H, Ding J, Piednoël M, Schneeberger K. 2018. findGSE: estimating genome size variation within human and *Arabidopsis* using k-mer frequencies. *Bioinformatics* 34:550–57
60. Guo LT, Wang SL, Wu QJ, Zhou XG, Xie W, Zhang YJ. 2015. Flow cytometry and K-mer analysis estimates of the genome sizes of *Bemisia tabaci* B and Q (Hemiptera: Aleyrodidae). *Front. Physiol.* 6:144
61. Evans T, Johnson A, Loose M. 2018. Virtual Genome Walking across the 32 Gb *Ambystoma mexicanum* genome; assembling gene models and intronic sequence. *Sci. Rep.* 8:618
62. Nowoshilow S, Schloissnig S, Fei J-F, Dahl A, Pang AW, et al. 2018. The axolotl genome and the evolution of key tissue formation regulators. *Nature* 554:50
63. Minio A, Lin J, Gaut BS, Cantu D. 2017. How single molecule real-time sequencing and haplotype phasing have enabled reference-grade diploid genome assembly of wine grapes. *Front. Plant Sci.* 8:826
64. English AC, Richards S, Han Y, Wang M, Vee V, et al. 2012. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLOS ONE* 7:e47768
65. van Heesch S, Kloosterman WP, Lansu N, Ruzius F-P, Levandowsky E, et al. 2013. Improving mammalian genome scaffolding using large insert mate-pair next-generation sequencing. *BMC Genom.* 14:257
66. Panova M, Aronsson H, Cameron RA, Dahl P, Godhe A, et al. 2016. DNA extraction protocols for whole-genome sequencing in marine organisms. *Methods Mol. Biol.* 1452:13–44
67. Sambrook J, Russell DW. 2001. *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harb. Lab. Press
68. Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC, McCombie WR. 2015. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res.* 25:1750–56
69. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27:722–36
70. Xiao C-L, Chen Y, Xie S-Q, Chen K-N, Wang Y, et al. 2017. MECAT: fast mapping, error correction, and *de novo* assembly for single-molecule sequencing reads. *Nat. Methods* 14:1072–74
71. Zimin AV, Puiu D, Luo M-C, Zhu T, Koren S, et al. 2017. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res.* 27:787–92
72. Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annu. Rev. Genet.* 34:401–37
73. Wilson JT, Dixon DR, Dixon LR. 2002. Numerical chromosomal aberrations in the early life-history stages of a marine tubeworm, *Pomatoceros lamarckii* (Polychaeta: Serpulidae). *Aquat. Toxicol.* 59:163–75
74. Gregory TR, Mable BK. 2005. Polyploidy in animals. In *The Evolution of the Genome*, ed. TR Gregory, pp. 427–517. Cambridge, MA: Academic
75. Filipski A, Kumar S. 2005. Comparative genomics in eukaryotes. In *The Evolution of the Genome*, ed. TR Gregory, pp. 521–83. Cambridge, MA: Academic
76. O'Brien SJ, Menninger JC, Nash WG. 2006. *An Atlas of Mammalian Chromosomes*. Hoboken, NJ: John Wiley & Sons
77. Goldman MA, LoVerde PT, Chrisman CL. 1983. Hybrid origin of polyploidy in freshwater snails of the genus *Bulinus* (Mollusca: Planorbidae). *Evolution* 37:592–600
78. Weiss M, Weigand H, Weigand AM, Leese F. 2018. Genome-wide single-nucleotide polymorphism data reveal cryptic species within cryptic freshwater snail species—the case of the *Ancylus fluviatilis* species complex. *Ecol. Evol.* 8:1063–72
79. Foighil DÓ, Smith MJ. 1995. Evolution of asexuality in the cosmopolitan marine clam *Lasaea*. *Evolution* 49:140–50
80. Lee T. 1999. Polyploidy and meiosis in the freshwater clam *Sphaerium striatinum* (Lamarck) and chromosome numbers in the Sphaeriidae (Bivalvia, Veneroida). *Cytologia* 64:247–52
81. Burch JB, Jung Y. 1993. Polyploid chromosome numbers in the Torquis group of the freshwater snail genus *Gyraulus* (Mollusca: Pulmonata: Planorbidae). *Cytologia* 58:145–49

82. Marotta R, Crottini A, Raimondi E, Fondello C, Ferraguti M. 2014. Alike but different: the evolution of the *Tubifex tubifex* species complex (Annelida, Clitellata) through polyploidization. *BMC Evol. Biol.* 14:73
83. Tweeten KA, Morris SJ. 2016. Flow cytometry analysis of DNA ploidy levels and protein profiles distinguish between populations of *Lumbriculus* (Annelida: Clitellata). *Invert. Biol.* 135:385–99
84. Cilibrasi R, Van Iersel L, Kelk S, Tromp J. 2005. *On the complexity of several haplotyping problems*. Presented at the International Workshop on Algorithms in Bioinformatics, Mallorca, Spain, Oct. 3–6
85. Bansal V, Bafna V. 2008. HapCUT: an efficient and accurate algorithm for the haplotype assembly problem. *Bioinformatics* 24:i153–i59
86. Wang R-S, Wu L-Y, Li Z-P, Zhang X-S. 2005. Haplotype reconstruction from SNP fragments by minimum error correction. *Bioinformatics* 21:2456–62
87. Jiao W-B, Schneeberger K. 2017. The impact of third generation genomic technologies on plant genome assembly. *Curr. Opin. Plant Biol.* 36:64–70
88. Ravi V, Venkatesh B. 2018. The divergent genomes of teleosts. *Annu. Rev. Anim. Biosci.* 6:47–68
89. Voolstra CR, Worheide G, Lopez JV. 2017. Advancing genomics through the Global Invertebrate Genomics Alliance (GIGA). *Invertebr. Syst.* 31:1–7
90. Schatz MC, Witkowski J, McCombie WR. 2012. Current challenges in *de novo* plant genome sequencing and assembly. *Genome Biol.* 13:243
91. Tang H, Lyons E, Town CD. 2015. Optical mapping in plant comparative genomics. *GigaScience* 4:3
92. Mascher M, Muehlbauer GJ, Rokhsar DS, Chapman J, Schmutz J, et al. 2013. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J.* 76:718–27
93. Gardiner LJ, Bansept-Basler P, Olohan L, Joynson R, Brenchley R, et al. 2016. Mapping-by-sequencing in complex polyploid genomes using genic sequence capture: a case study to map yellow rust resistance in hexaploid wheat. *Plant J.* 87:403–19
94. Udall JA, Dawe RK. 2018. Is it ordered correctly? Validating genome assemblies by optical mapping. *Plant Cell* 30:7–14
95. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, et al. 2017. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat. Genet.* 49:643
96. Zhang W, Chen J, Yang Y, Tang Y, Shang J, Shen B. 2011. A practical comparison of *de novo* genome assembly software tools for next-generation sequencing technologies. *PLOS ONE* 6:e17915
97. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2010. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–79
98. Yeo S, Coombe L, Warren RL, Chu J, Birol I. 2017. ARCS: scaffolding genome drafts with linked reads. *Bioinformatics* 34:725–31
99. McCoy RC, Taylor RW, Blauwkamp TA, Kelley JL, Kertesz M, et al. 2014. Illumina TruSeq synthetic long-reads empower de novo assembly and resolve complex, highly-repetitive transposable elements. *PLOS ONE* 9:e106689
100. Eisenstein M. 2015. Startups use short-read data to expand long-read sequencing market. *Nat. Biotechnol.* 33:433–35
101. Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17:333
102. Jones SJ, Haulena M, Taylor GA, Chan S, Bilobram S, et al. 2017. The genome of the northern sea otter (*Enhydra lutris kenyoni*). *Genes* 8:379
103. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA. 2018. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 36:338–45
104. Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, et al. 2012. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* 9:999
105. Selvaraj S, Dixon JR, Bansal V, Ren B. 2013. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nat. Biotechnol.* 31:1111
106. Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, et al. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* 26:342–50

107. Staats M, Erkens RH, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, et al. 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PLOS ONE* 8:e69189
108. Tin MM-Y, Economo EP, Mikheyev AS. 2014. Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. *PLOS ONE* 9:e96793
109. Bagnall RD, Ingles J, Yeates L, Berkovic SF, Semsarian C. 2017. Exome sequencing–based molecular autopsy of formalin-fixed paraffin-embedded tissue after sudden death. *Genet. Med.* 19:1127
110. Chin S-F, Santoja A, Grzelak M, Ahn S, Sammut S-J, et al. 2017. Shallow whole genome sequencing for robust copy number profiling of formalin-fixed paraffin-embedded breast cancers. *Exp. Mol. Pathol.* 104:161–69
111. Prüfer K, de Filippo C, Grote S, Mafessoni F, Korlević P, et al. 2017. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* 358:655–58
112. Bhattacharya S, Li J, Sockell A, Kan MJ, Bava FA, et al. 2018. Whole-genome sequencing of Atacama skeleton shows novel mutations linked with dysplasia. *Genome Res.* 28:423–31
113. Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, et al. 2013. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat. Biotechnol.* 31:1023
114. Geniez S, Foster JM, Kumar S, Moumen B, LeProust E, et al. 2012. Targeted genome enrichment for efficient purification of endosymbiont DNA from host DNA. *Symbiosis* 58:201–7
115. Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, et al. 2017. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* 14:955
116. Groot TV, Breeuwer JA. 2006. *Cardinium* symbionts induce haploid thelytoky in most clones of three closely related *Brevipalpus* species. *Exp. Appl. Acarol.* 39:257–71
117. Lin K, Limpens E, Zhang Z, Ivanov S, Saunders DG, et al. 2014. Single nucleus genome sequencing reveals high similarity among nuclei of an endomycorrhizal fungus. *PLOS Genet.* 10:e1004078
118. Guérin F, Arnaiz O, Boggetto N, Wilkes CD, Meyer E, et al. 2017. Flow cytometry sorting of nuclei enables the first global characterization of *Paramecium* germline DNA and transposable elements. *BMC Genom.* 18:327
119. Jones MR, Good JM. 2016. Targeted capture in evolutionary and ecological genomics. *Mol. Ecol.* 25:185–202
120. Saha S, Sparks AB, Rago C, Akmaev V, Wang CJ, et al. 2002. Using the transcriptome to annotate the genome. *Nat. Biotechnol.* 20:508
121. Cahais V, Gayral P, Tsagkogeorga G, Melo-Ferreira J, Ballenghien M, et al. 2012. Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Mol. Ecol. Resour.* 12:834–45
122. Zhang G, Fang X, Guo X, Li L, Luo R, et al. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490:49
123. Abugessaisa I, Kasukawa T, Kawaji H. 2017. Genome annotation. In *Bioinformatics*, Vol. 1: *Data, Sequence Analysis, and Evolution*, ed. JM Keith, pp. 107–21. New York: Springer
124. Markelz RC, Covington MF, Brock MT, Devisetty UK, Kliebenstein DJ, et al. 2017. Using RNA-seq for genomic scaffold placement, correcting assemblies, and genetic map creation in a common *Brassica rapa* mapping population. *G3* 7:2259–70
125. Celniker SE, Wheeler DA, Kronmiller B, Carlson JW, Halpern A, et al. 2002. Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biol.* 3:research0079.1–14
126. O'Neill K, Hills M, Gottlieb M, Borkowski M, Karsan A, Lansdorp PM. 2017. Assembling draft genomes using contiBAIT. *Bioinformatics* 33:2737–39
127. Anselmetti Y, Berry V, Chauve C, Chateau A, Tannier E, Bérard S. 2015. Ancestral gene synteny reconstruction improves extant species scaffolding. *BMC Genom.* 16:S11
128. Assour L, Emrich S. 2015. *Multi-genome synteny for assembly improvement*. Paper presented at the 7th International Conference on Bioinformatics and Computational Biology, Honolulu, HI, March 9–11

129. Tang H, Zhang X, Miao C, Zhang J, Ming R, et al. 2015. ALLMAPS: robust scaffold ordering based on multiple maps. *Genome Biol.* 16:3
130. Eggertsson HP, Jonsson H, Kristmundsdottir S, Hjartarson E, Kehr B, et al. 2017. Graphtyper enables population-scale genotyping using pangenome graphs. *Nat. Genet.* 49:1654
131. Galardini M, Pini F, Bazzicalupo M, Biondi EG, Mengoni A. 2015. Pangenome evolution in the symbiotic nitrogen fixer *Sinorhizobium meliloti*. In *Biological Nitrogen Fixation*, ed. FJ de Bruijn, pp. 265–70. New York: John Wiley & Sons
132. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, et al. 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* 359:eaar4120
133. Golicz AA, Bayer PE, Barker GC, Edger PP, Kim H, et al. 2016. The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat. Commun.* 7:13390
134. Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. 2011. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the Teleost genome duplication. *Genetics* 188:799–808
135. Kalderimis A, Lyne R, Butano D, Contrino S, Lyne M, et al. 2014. InterMine: extensive web services for modern biology. *Nucleic Acids Res.* 42:W468–W72
136. Donlin MJ. 2009. Using the generic genome browser (GBrowse). *Curr. Protoc. Bioinform.* 17:9.9.1–24
137. Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, et al. 2016. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* 17:66
138. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, et al. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* 15:1451–55
139. Smith RN, Aleksic J, Butano D, Carr A, Contrino S, et al. 2012. InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. *Bioinformatics* 28:3163–65
140. Liew YJ, Aranda M, Voolstra CR. 2016. Reefgenomics.org: a repository for marine genomics data. *Database* 2016:baw152
141. Carrere S, Gouzy J. 2017. myGenomeBrowser: building and sharing your own genome browser. *Bioinformatics* 33:1255–57
142. Challis RJ, Kumar S, Stevens L, Blaxter M. 2017. GenomeHubs: simple containerized setup of a custom Ensembl database and web server for any species. *Database* 2017:bax039
143. Fischer J, Tuecke S, Foster I, Stewart CA. 2015. Jetstream: a distributed cloud infrastructure for underresourced higher education communities. In *Proceedings of the 1st Workshop on the Science of Cyberinfrastructure: Research, Experience, Applications and Models*, ed. S Jha, pp. 53–61. New York: Assoc. Comput. Mach.
144. Stewart CA, Cockerill TM, Foster I, Hancock D, Merchant N, et al. 2015. Jetstream: a self-provisioned, scalable science and engineering cloud environment. In *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*, ed. GD Peterson, 29:1–29.8. New York: Assoc. Comput. Mach.
145. Dudgeon SR, Aronson RB, Bruno JF, Precht WF. 2010. Phase shifts and stable states on coral reefs. *Mar. Ecol. Prog. Ser.* 413:201–16
146. Bruckner AW. 2002. *Proceedings of the Caribbean Acropora Workshop: potential application of the U.S. Endangered Species Act as a conservation strategy.* Tech. Memo., NMFS-OPR-24, Natl. Ocean. Atmos. Adm., Silver Spring, MD. 184 pp.
147. Toro MA, Caballero A. 2005. Characterization and conservation of genetic diversity in subdivided populations. *Philos. Trans. R. Soc. B* 360:1367–78
148. Arnaud-Haond S, Duarte CM, Alberto F, Serrao EA. 2007. Standardizing methods to address clonality in population studies. *Mol. Ecol.* 16:5115–39
149. Highsmith RC. 1982. Reproduction by fragmentation in corals. *Mar. Ecol. Prog. Ser.* 7:207–26
150. Fischer MC, Rellstab C, Leuzinger M, Roumet M, Gugerli F, et al. 2017. Estimating genomic diversity and population differentiation—an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. *BMC Genom.* 18:69

151. Hess JE, Matala AP, Narum SR. 2011. Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Mol. Ecol. Resour.* 11:137–49
152. Hauser L, Baird M, Hilborn RAY, Seeb LW, Seeb JE. 2011. An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Mol. Ecol. Resour.* 11:150–61
153. Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, et al. 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 407:513–16
154. Rasheed A, Hao Y, Xia X, Khan A, Xu Y, et al. 2017. Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. *Mol. Plant* 10:1047–64
155. Cui F, Zhang N, Fan X-L, Zhang W, Zhao C-H, et al. 2017. Utilization of a Wheat660K SNP array-derived high-density genetic map for high-resolution mapping of a major QTL for kernel number. *Sci. Rep.* 7:3788
156. Kitchen S, Ratan A, Bedoya-Reina O, Burhans R, Fogarty N, et al. 2018. Genomic variants among threatened Acropora corals. bioRxiv:349910
157. Devlin-Durante MK, Baums IB. 2017. Genome-wide survey of single-nucleotide polymorphisms reveals fine-scale population structure and signs of selection in the threatened Caribbean elkhorn coral, *Acropora palmata*. *PeerJ* 5:e4077
158. Baums IB, Miller MW, Hellberg ME. 2005. Regionally isolated populations of an imperiled Caribbean coral, *Acropora palmata*. *Mol. Ecol.* 14:1377–90
159. Budd AF, Klaus JS. 2001. The origin and early evolution of the *Montastraea "annularis"* species complex (Anthozoa: Scleractinia). *J. Paleontol.* 75:527–45
160. Knowlton N, Weil E, Weigt LA, Guzmán HM. 1992. Sibling species in *Montastraea annularis*, coral bleaching, and the coral climate record. *Science* 255:330–33
161. Weil E, Knowton N. 1994. A multi-character analysis of the Caribbean coral *Montastraea annularis* (Ellis and Solander, 1786) and its two sibling species, *M. faveolata* (Ellis and Solander, 1786) and *M. franksi* (Gregory, 1895). *Bull. Mar. Sci.* 55:151–75
162. Pandolfi JM, Lovelock CE, Budd AF. 2002. Character release following extinction in a Caribbean reef coral species complex. *Evolution* 56:479–501
163. Pandolfi JM, Budd AF. 2008. Morphology and ecological zonation of Caribbean reef corals: the *Montastraea "annularis"* species complex. *Mar. Ecol. Prog. Ser.* 369:89–102
164. Prada C, Hanna B, Budd AF, Woodley CM, Schmutz J, et al. 2016. Empty niches after extinctions increase population sizes of modern corals. *Curr. Biol.* 26:3190–94
165. Deleted in proof
166. Fogarty ND. 2012. Caribbean acroporid coral hybrids are viable across life history stages. *Mar. Ecol. Prog. Ser.* 446:145–59
167. Tang K, Thornton KR, Stoneking M. 2007. A new approach for using genome scans to detect recent positive selection in the human genome. *PLOS Biol.* 5:1587–602
168. Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, et al. 2009. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLOS Genet.* 5(6):e1000519
169. Bedoya-Reina O, Ratan A, Burhans R, Kim H, Giardine B, et al. 2013. Galaxy tools to study genome diversity. *GigaScience* 2:17
170. Verdu P, Rosenberg NA. 2011. A general mechanistic model for admixture histories of hybrid populations. *Genetics* 189:1413–26
171. Willis BL, van Oppen MJH, Miller DJ, Vollmer SV, Ayre DJ. 2006. The role of hybridization in the evolution of reef corals. *Annu. Rev. Ecol. Evol. Syst.* 37:489–517
172. Van Oppen MJH, Gates RD, Blackall LL, Cantin N, Chakravarti LJ, et al. 2017. Shifting paradigms in restoration of the world's coral reefs. *Glob. Change Biol.* 23:3437–48
173. Matz MV, Treml EA, Aglyamova GV, Bay LK. 2018. Potential and limits for rapid genetic adaptation to warming in a Great Barrier Reef coral. *PLOS Genet.* 14:e1007220
174. Cleves PA, Strader ME, Bay LK, Pringle JR, Matz MV. 2018. CRISPR/Cas9-mediated genome editing in a reef-building coral. *PNAS* 115:5235–40

175. Cowie RH, Thiengo SC. 2003. The apple snails of the Americas (Mollusca: Gastropoda: Ampullariidae: Asolene, Felipponea, Marisa, Pomacea, Pomella): a nomenclatural and type catalog. *Malacologia* 45:41–100
176. Int. Union Conserv. Nat. 2018. *100 of the world's worst invasive alien species*. Glob. Invasive Species Database, Int. Union Conserv. Nat., Gland, Switz. **http://www.iucngisd.org/gisd/100_worst.php**
177. Ip JCH, Mu HW, Chen Q, Sun J, Ituarte S, et al. 2018. AmpuBase: a transcriptome database for eight species of apple snails (Gastropoda: Ampullariidae). *BMC Genom*. 19:9
178. Gruenthal KM, Witting DA, Ford T, Neuman MJ, Williams JP, et al. 2014. Development and application of genomic tools to the restoration of green abalone in southern California. *Conserv. Genet*. 15:109–21
179. Sandoval-Castillo J, Robinson NA, Hart AM, Strain LW, Beheregaray LB. 2018. Seascape genomics reveals adaptive divergence in a connected and commercially important mollusc, the greenlip abalone (*Haliotis laevigata*), along a longitudinal environmental gradient. *Mol. Ecol.* 27:1603–20
180. Lampert KP. 2018. Oxygen concentration drives local adaptation in the greenlip abalone (*Haliotis laevigata*). *Mol. Ecol.* 27:1521–23
181. Nam BH, Kwak W, Kim YO, Kim DG, Kong HJ, et al. 2017. Genome sequence of Pacific abalone (*Haliotis discus hannai*): the first draft genome in family Haliotidae. *GigaScience* 6:1–8
182. Renard E, Leys SP, Wörheide G, Borchiellini C. 2018. Understanding animal evolution: the added value of sponge transcriptomics and genomics: the disconnect between gene content and body plan evolution. *BioEssays* 40(9):1700237
183. De'ath G, Fabricius KE, Sweatman H, Puotinen M. 2012. The 27-year decline of coral cover on the Great Barrier Reef and its causes. *PNAS* 109:17995–99
184. Hall MR, Kocot KM, Baughman KW, Fernandez-Valverde SL, Gauthier ME, et al. 2017. The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. *Nature* 544:231
185. Vogler C, Benzie J, Barber PH, Erdmann MV, Ambariyanto, et al. 2012. Phylogeography of the crown-of-thorns starfish in the Indian Ocean. *PLOS ONE* 7(8):e43499

Annual Review of
Animal Biosciences

Volume 7, 2019

# Contents

**Errata**

An online log of corrections to *Annual Review of Animal Biosciences* articles may be found at http://www.annualreviews.org/errata/animal