Advancing genomics through the Global Invertebrate Genomics Alliance (GIGA)

Christian R. Voolstra[A], *GIGA Community of Scientists (COS)*[1], *Gert Wörheide*[B,D]
and Jose V. Lopez[C,D]

[A]Red Sea Research Center, Division of Biological and Environmental Science and Engineering (BESE),
 King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia.
[B]Department of Earth and Environmental Sciences, Palaeontology & Geobiology, and GeoBio-Center,
 Ludwig-Maximilians-Universität München, Richard-Wagner Str. 10, 80333, SNSB – Bavarian State
 Collections of Palaeontology and Geology, Richard-Wagner Str. 10, 80333 Munich, Germany.
[C]Halmos College of Natural Sciences and Oceanography, Nova Southeastern University, 8000 North Ocean Drive,
 Dania Beach, FL 33004, USA.
[D]Corresponding authors. Email: woerheide@lmu.de; joslo@nova.edu

Abstract. The Global Invertebrate Genomics Alliance (GIGA), a collaborative network of diverse scientists, marked its second anniversary with a workshop in Munich, Germany in 2015, where international attendees focused on discussing current progress, milestones and bioinformatics resources. The community determined the recruitment and training of talented researchers as one of the most pressing future needs and identified opportunities for network funding. GIGA also promotes future research efforts to prioritise taxonomic diversity and create new synergies. Here, we announce the generation of a central and simple data repository portal with a wide coverage of available sequence data, via the compagen platform, in parallel with more focused and specialised organism databases to globally advance invertebrate genomics. This article serves the objectives of GIGA by disseminating current progress and future prospects in the science of invertebrate genomics with the aim of promotion and facilitation of interdisciplinary and international research.

Introduction

Genomic research will likely exhibit exponential growth for years to come (Stephens *et al.* 2015). In anticipation of this, the Global Invertebrate Genomics Alliance (GIGA) formed to address the multiple needs and opportunities of the expanding invertebrate genomics community (GIGA COS 2014). GIGA complements taxonomically similar but differently focused efforts, such as the arthropod-focused Genome 10K (Genome 10K COS 2009; Koepfli *et al.* 2015) and the Arthropod Genomics Consortium (i5k) (Robinson *et al.* 2011), and interacts with other initiatives, such as the Genomic Observatories Network (Davies *et al.* 2014), the Global Genome Initiative (GGI, http://naturalhistory.si.edu/ggi/), and the Ocean Genome Legacy (OGL, http://www.northeastern.edu/cos/marinescience/ogl/). These latter initiatives were created to archive diverse arrays of specimens, tissues, nucleic acid samples, voucher material and supporting data in a global effort to help coordinate and disseminate these data and materials across the broader scientific community.

GIGA has adopted a 'bottom up' philosophy to advance its community goals, with voluntary and vigorous contributions of members building a consensus approach to avoiding duplication of research efforts through effective coordination within the community (see Fig. S1 for the distribution of member countries, available as Supplementary material). GIGA was launched with an initial workshop in 2013 and an inaugural white paper (GIGA COS 2014). The second GIGA workshop (March 2015 at the Ludwig-Maximilians-Universität, Munich, Germany) reviewed the progress of this growing initiative. More than 80 scientists from 23 countries participated in the 3-day event, which explored aspects of project implementation and comparative analyses across invertebrates as well as training and funding efforts and opportunities. We briefly report here on the major outcomes of these discussions.

Current status of invertebrate genome and transcriptome sequencing efforts

Genomes, transcriptomes and 'priority species'

The first GIGA white paper (GIGA COS 2014) and two workshops identified many major hurdles hindering large-scale

---

[1]Authors are listed in Appendix 1.

genome research, which have started to be addressed by concerted efforts from the community. For example, GIGA is tallying the growing number of completed invertebrate genomes from individual laboratories (Dunn and Ryan 2015), tracking new advances in sequencing and bioinformatics and monitoring the investment of specific research consortia. To keep abreast of these ongoing efforts, the updated GIGA homepage (www.giga-cos. org) features improved access to member projects and relevant publications, and now provides phylogenetically sorted links to organismal genomes and curated transcriptome resources (see Box 1). In addition, the GIGA listserver (https://lists.lrz. de/mailman/listinfo/giga) was set up to facilitate information exchange within the community.

While GIGA supports the focused efforts of single laboratories, the consortium acknowledges the need to collectively target priority species for sequencing due to their economic, evolutionary and/or conservation significance (see Box 2). These priority species (not exhaustive) should occupy a distinct space in the GIGA framework, hopefully becoming the focus of future concerted community efforts (Table 1).

Distributed effort model and ambassador framework

The GIGA II workshop identified 'divide-and-conquer' as the best strategy for the distribution of efforts concerning the coordination of genome and transcriptome sequencing and knowledge exchange in different taxonomic and phylogenetic groups and scientific communities. For example, clade-specific 'ambassadors' (see http://giga-cos.org/index.php/members) were chosen to recruit within their own smaller communities and discuss candidate taxa for future genomics. The ambassador framework will also effectively distribute the efforts to provide a complete and comprehensive overview of current and future invertebrate genomic and transcriptomic resources, and GIGA members will have direct 'go-to' experts for the organisms of interest.

Developing online resources for sharing GIGA data

Sequence data archives

High-throughput sequencing technologies have boosted the field of comparative genomics (Green *et al.* 2015; Koepfli *et al.* 2015). The volume of sequence data available through public databases such as the International Nucleotide Sequence

Database Collaboration (INSDC) portal and its members, GenBank (at the National Center for Biotechnology Information, NCBI), European Nucleotide Archive (ENA) (at European Bioinformatics Institute) and DNA Data Bank of Japan (DDBJ), has increased exponentially (Baxevanis 2011; Stephens *et al.* 2015). As overall sequencing costs continue to decrease, individual laboratories and community efforts can generate ever larger amounts of novel genomic and transcriptomic data from non-model organisms. Aggregative databases such as Compagen (http://compagen.org) (Hemmrich & Bosch 2008) or Reefgenomics (http://reefgenomics.org) (Liew *et al.* 2016) coordinate data presentation and analyses for particular species or ecosystem groups. While the development of a single database covering data from a wide variety of invertebrate species would likely have wide appeal, the programming, curatorial and funding burden of producing and maintaining a centralised resource goes beyond the reach of GIGA and similar consortia at present. Many researchers only need simple queries (e.g. BLAST-based sequence similarity searches), while a minority needs more complex, multivariate queries (e.g. synteny-based orthology inference). We propose an approach that will cater to both needs by providing: (1) a central sequence database that focuses on simple queries accessing a large range of data, in combination with (2) more focused, de-centralised species databases that satisfy complex queries and more specific requirements (Parkhill *et al.* 2010). The GIGA homepage will serve as a hub from which to access these platforms.

Compagen.org as a general data portal for GIGA transcriptomes and genomes

Compagen.org, originally designed to cater to the early branching (non-bilaterian) animal community, will be extended to serve as our central sequence database providing access to processed datasets such as assembled transcriptomes/genomes

---

Box 1.   **The GIGA homepage as a community hub for invertebrate genomics data**

The GIGA homepage (http://giga-cos.org) provides an informative, interactive and mobile-friendly portal to relevant invertebrate genomic research, data and news. A sign-up button welcomes new members to register, while another link invites members to upload current publications. Importantly, GIGA provides a landing page to relevant transcriptome and genome 'links' with the opportunity for users to add their own assembled sequence data. The GIGA website is not meant to be a database in itself. Rather, the community nature of this homepage is emphasised by its open access to data. GIGA strongly supports contributions from the community (bottom-up philosophy) to share their new publications, links to datasets and news items via web forms that can be filled without any prior knowledge on coding languages. Our anticipation is that the usefulness of the webpage will grow with the community.

Box 2.   **Criteria for sequencing candidacy – priority species for sequencing**

Several detailed biological criteria for species selection were discussed previously (see GIGA COS 2014), including:
- keystone, endangered or threatened and invasive species status;
- species representing deep, isolated phylogenetic branches;
- species with contrasting life spans (particularly long or very short);
- species with specific physiological abilities such as regeneration or anhydrobiosis; and
- species inhabiting 'extreme' environments.

Practical constraints include technological, technical, financial and cultural obstacles. Sample availability, individual researcher motivations, collaborations and availability of genome size estimates must also be considered.

Major technical criteria for taxon selection includes:
- the availability of appropriate sampling permits (following CITES, Rio Biodiversity and Nagoya Protocol guidelines (UNEP/CBD/COP 2002); https://www.cbd.int/abs/);
- availability of metadata;
- proper vouchering;
- the species being model and/or popular;
- availability of transcriptome data (or plans to generate them); and
- agreements for data sharing and access.

Table 1.   Proposed priority species for whole genome sequencing

N/A, not available; TBD, to be determined. We are limited in the number of taxa we can list, so this table only shows a small cross section of priority taxa

Species	Common name	Phylum/Class	Genome size (Gb)	Significance
Alitta (*Nereis*) *virens*	King ragworm or sandworm	Annelida/Polychaeta	TBD	Ecological model
Magelona pitelkai	N/A	Annelida/Polychaeta	TBD	Early branching annelid
Owenia fusiformis	Spindle-shaped tubeworm	Annelida/Polychaeta	TBD	Early branching annelid
Platynereis massiliensis	N/A	Annelida/Polychaeta	0.3	Ecological/developmental model
Spirobranchus lamarcki	Keel worm	Annelida/Polychaeta	1.0	Ecological/developmental model
Streblospio benedicti	N/A	Annelida/Polychaeta	1.4	Developmental/ecological model
Artemia franciscana	Brine shrimp	Arthropoda/Branchiopoda	0.9	Commercial, model
Acanthephyra purpurea	Mid-water shrimp	Arthropoda/Malacostraca	TBD	Ecological importance, diurnal migrator
Asellus aquaticus	Waterlouse	Arthropoda/Malacostraca	1.9	Research (ecology, evolution)
Barbouria cubensis	Cave shrimp	Arthropoda/Malacostraca	TBD	Cave conservation importance
Glyptonotus spp.	Antarctic isopod	Arthropoda/Malacostraca	0.5	Model for adaptation (cold)
Scylla olivacea	Orange mud crab	Arthropoda/Malacostraca	TBD	Commercial
Sergia robusta	White shrimp	Arthropoda/Malacostraca	3	Ecological, commercial
Systellaspis debilis	Mid-water shrimp	Arthropoda/Malacostraca	TBD	Bioluminescent model species
Petrolisthes cinctipes	Porcelain crab	Arthropoda/Maxillopoda	TBD	Ecophysiology model intertidal species
Amphibalanus amphitrite	Striped barnacle	Arthropoda/Maxillopoda	TBD	Commercial, pest
Pollicipes polymerus	Gooseneck Barnacle	Arthropoda/Maxillopoda	TBD	Commercial, adaptation (intertidal)
Bugula neritina	Brown bryozoan	Bryozoa/Gymnolaemata	0.2	Pharmaceutical
Cerianthus membranaceus	Cylinder anemone	Cnidaria/Anthozoa	0.5	Sister group to *aiptasia* and corals
Eguchipsammia fistula	Deep sea coral	Cnidaria/Anthozoa	0.5	Deep sea ecology, adaptation (cold, depth)
Eunicella cavolini	Yellow sea whip	Cnidaria/Anthozoa	0.5	Foundation species for temperature marine ecosystems
Heliopora coerulea	Blue coral	Cnidaria/Anthozoa	TBD	Research
Montipora digitata	Finger coral	Cnidaria/Anthozoa	TBD	Research
Orbicella annularis	Boulder star coral	Cnidaria/Anthozoa	TBD	Research, taxonomy
Ricordea florida	Florida false coral	Cnidaria/Anthozoa	TBD	
Tubipora musica	Organ pipe coral	Cnidaria/Anthozoa	TBD	Research
Millepora dichotoma	Net fire coral	Cnidaria/Hydrozoa	TBD	Research
Cassiopea frondosa	Upside-down jellyfish	Cnidaria/Scyphozoa	0.3	Research
Gephyrocrinus messingi	Sea lily	Echinodermata/Crinoidea	TBD	Sister to remaining echinoderms
Himerometra robustipinna	Feather star	Echinodermata/Crinoidea	TBD	Research
Pontiometra andersoni	Feather star	Echinodermata/Crinoidea	TBD	Research, taxonomy
Bankia setacea	Feathery shipworm	Mollusca/Bivalvia	1.5	Physiology, invasive
Bathymodiolus azoricus.	N/A	Mollusca/Bivalvia		Deep sea ecology
Corbicula fluminea	Asian clam or golden clam	Mollusca/Bivalvia	1.5	Physiology, invasive
Dreissena polymorpha	Zebra mussel	Mollusca/Bivalvia	1.7	Invasive, pest
Limnoperna fortunei	Golden mussel	Mollusca/Bivalvia	TBD	Invasive
Mya arenaria	Soft-shell clam	Mollusca/Bivalvia	1.4	Invasive, commercial
Ruditapes decussatus	Grooved carpet shell	Mollusca/Bivalvia	1.8	Commercial, ecological
Ruditapes philippinarum	Manila clam	Mollusca/Bivalvia	1.9	Commercial, invasive, model for mitochondrial biology
Cranchia scabra	Glass squid	Mollusca/Cephalopoda	TBD	Deep sea ecology
Graneledone verrucosa	Atlantic longarm octopus	Mollusca/Cephalopoda	TBD	Benthic deep sea species
Japetella diaphana	Pelagic bolitaenid octopod	Mollusca/Cephalopoda	TBD	Pelagic deep sea species
Loligo pealeii	Longfin inshore squid	Mollusca/Cephalopoda	2.7	Cellular neurobiology, fishery
Macrotritopus defillipi	N/A	Mollusca/Cephalopoda	TBD	Research
Conus bullatus	Bubble cone	Mollusca/Gastropoda	2.7	Natural products
Pomacea maculata	Apple snail	Mollusca/Gastropoda	0.6	Invasive species, biomedical research
Cellana sandwicensis	Hawaiian limpet	Mollusca/Gastropoda	TBD	Model for sympatric diversification
Aplysina cauliformis	Row pore rope sponge	Porifera/Demospongiae	TBD	Research
Cinachyrella alloclada	Golfball sponge	Porifera/Demospongiae	TBD	Research
Haliclona oculata	N/A	Porifera/Demospongiae	0.1–0.2	Research
Tedania ignis	Fire sponge	Porifera/Demospongiae		
Themiste lageniformis	Peanut worm	Sipuncula/Sipunculidea	0.5	Taxonomy, developmental biology
Milnesium tardigradum	Limno-terrestrial tardigrade	Tartigrada/Eutardigrada	TBD	Taxonomy
Xenoturbella profundus	N/A	Xenacoelomorpha/Xenoturbellida	0.6	Early diverging bilaterian clade

and predicted peptides, thereby complementing sequence datasets available from public domains (e.g. NCBI Sequence Read Archive (SRA), NCBI dbEST (GenBank's sequence data on cDNA sequences or expressed sequence tags) and ENA trace archives). Compagen.org provides a critical extension to public databases due to the disconnect between what is submitted in the form of primary sequence data and what is available in the form of assembled and annotated data (Fig. 1). Having compagen. org as a central repository will clearly facilitate provision of assembled invertebrate sequence data as well as retrieval of data for analyses. This present choice does not preclude the additional development of other (taxon or feature specific) portals to GIGA-related sequences in the future.

Databases for GIGA-relevant taxa with a larger research community

For taxon groups studied by larger research communities, it is realistic (and tractable) to develop or coordinate independent, researcher or laboratory-curated, taxon-specific databases, such as http://www.spongebase.net for sponges or http://reefgenomics.org for coral reef invertebrates. These databases can accommodate the needs of the specific research groups focusing on specific taxa or groups of taxa, emphasising genomics, transcriptomics, variation or multi-species interaction as required. In connection with those, several single-species databases already exist, including the *Aiptasia* genome platform (http://aiptasia.reefgenomics.org/) (Baumgarten *et al.* 2015) and the *Hypsibius dujardini* genome server (http://www.tardigrades.org) (Koutsovoulos *et al.* 2016). From these efforts, cross-species databases can arise, such as http://comparative.reefgenomics.org, which provides a comparative platform for coral transcriptomes focusing on orthologous genes that can be used to interrogate coral species-specific genes (Bhattacharya *et al.* 2016). Community-managed data repositories should be extended to link genomic efforts to physical specimens deposited in natural history collections or to genetic resources deposited in genetic repositories in natural history museums or in other institutions. Overall, the accessibility of novel sequence data enhanced with metadata will open multiple doors to new hypotheses and crossover collaborations, thereby propelling broad scientific, technological and societal advances (i5K Consortium 2013; Dubilier *et al.* 2015; Green *et al.* 2015; Koepfli *et al.* 2015).

Community opportunities

Collaboration

The GIGA II workshop highlighted the common goals shared among i5k and GIGA initiatives, primarily with regards to help in structuring and advising the community for handling large genome project initiatives. Other key shared goals include funding, expert genome annotation and dissemination of results. There will be additional opportunities for collaboration between research groups involved in these and other large invertebrate genomics and transcriptomics consortia (such as 1Kite or the nematode.net) due to the shared methodological and bioinformatic approaches and training opportunities.
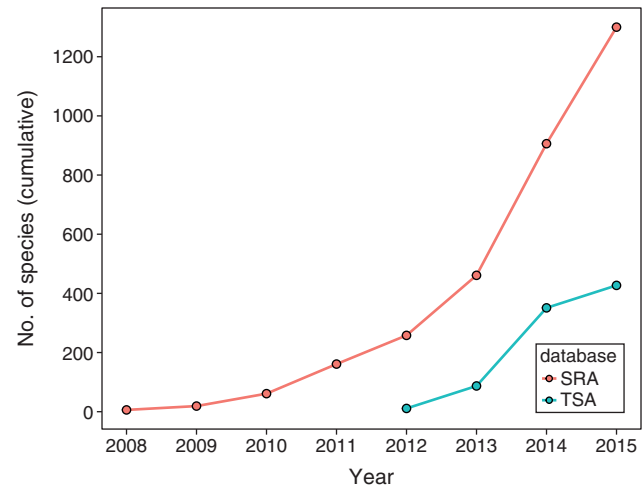

Fig. 1. Disconnect between primary (Sequence Read Archive, SRA) and assembled (Transcriptome Shotgun Assembly, TSA) high throughput sequence data available for invertebrates at the National Center for Biotechnology Information (NCBI). The plot shows a strong difference between the number of species for which primary transcriptomic sequence data are available (SRA) in comparison with the number of species for which assembled transcriptomes are available (TSA). One aim of GIGA is to bridge this gap by providing a hub for assembled transcriptome and genome data currently not accessible from public databases.

Training

Achieving the broad goals of GIGA will require new commitment and investment in training and educating the next generation of genome scientists (Fig. 2). Computational biology skills are needed to allow students to navigate, customise and populate increasingly complex databases, but additionally, training in aspects of organismal and evolutionary biology, ecology and biodiversity legislation that regulates the exchange of genetic resources across national boundaries needs to be provided. Besides a sweeping view and comparison of various Big Data issues, recent reviews provide a compelling argument for the genomics community to prepare an educational foundation to handle the deluge of genomic data and its associated metadata (Stephens *et al.* 2015; Dougherty *et al.* 2016). In general, current genomics and bioinformatics curricula appear relatively staid, while independent bioinformatics workshops (e.g. evomics.org, and those listed at http://meetings.cshl.edu/courseshome.aspx) can provide cutting edge approaches and showcase innovation; however, scheduling may be more uncertain and irregular for the latter and those specialist courses do not cover all aspects needed for the holistic training needed to cater for future job markets in (invertebrate and/or biomedical) genomics.

Funding agencies for invertebrate genomes

Currently, targeted funding for invertebrate genomes pales in comparison to the investments being made in vertebrate genomics (Couzin-Frankel 2012; AAAS 2015), so we compiled a limited selection of some of the agencies that we thought might contribute funding to further GIGA efforts (Table 2). Realising these will eventually expire, individual scientists, small collaborative cohorts or larger groups will have to forge imagination, fortitude
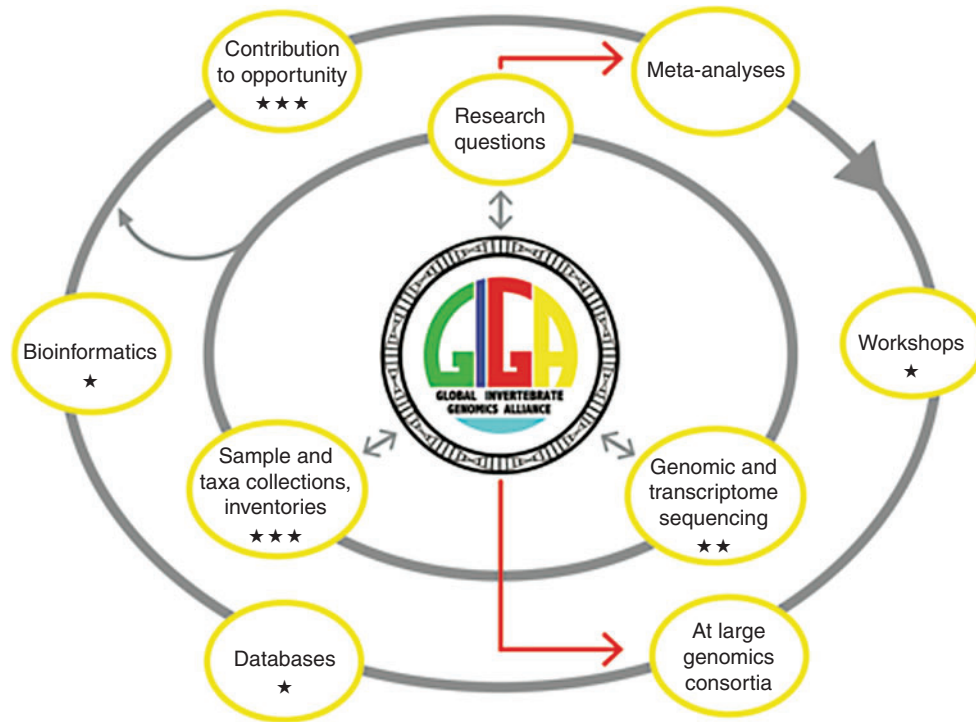
Fig. 2. Proposed GIGA deliverables and funding opportunities. Funding opportunities intersect with GIGA activities. Various activities have cumulative and synergistic effect. Specific areas for funding opportunities and requests are indicated by stars, with the number of stars proportional to estimated relative costs (e.g. workshops and conferences cost less than current sample collection or bioinformatics training – see Stephens *et al.* 2015)). 'At Large' genomics consortia can represent new and old groups with similar or complementary goals (Genome 10K 2009; Koepfli *et al.* 2015).

Table 2.   Potential funding agencies/opportunities
JGI, Joint Genomics Institute; NSF, National Science Foundation

Program	URL
NSF Genealogy of Life (GoLife)	http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5129
Marie Skłodowska-Curie Innovative Training Networks (EC Horizon 2020 Program)	http://ec.europa.eu/research/mariecurieactions/about-msca/actions/itn/index_en.htm
JGI Community Science Program	http://jgi.doe.gov/collaborate-with-jgi/community-science-program/
Earth Cube	http://earthcube.org/home
Alfred P. Sloan Foundation	http://www.sloan.org/apply-for-grants/the-grant-application-process/
Gordon and Betty Moore Foundation	https://www.moore.org/grants

and the scientific foundations and infrastructure discussed herein to persuade funding entities on the values of each particular project. However, during GIGA II the current dearth of international, multilateral funding programs that allow researchers in different continents (e.g. EU, Asia and Americas) to collaborate became strikingly obvious.

Concluding statement

Many benefits derived from single or multi-species genome sequencing projects have been described, realised and projected in recent years (GIGA COS 2014). Discoveries will continue to accrue with further innovations. Efforts will be further enhanced through the GIGA framework, which coordinates and promotes the concerted contributions of diverse experts. In the words of

Immanuel Kant by way of Durant (1938), 'Science is organized knowledge; wisdom is organized life'. In a similar spirit, GIGA will facilitate scientific organisation, and consequently the collective progress (wisdom) of humanity.

References

i5K Consortium (2013). The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *Journal of Heredity* **104**, 595–600. doi:10.1093/jhered/est050

AAAS (2015). Report: Research & Development series and analysis of FY 2016 budget request. (American Association for the Advancement of Science (AAAS): Washington, DC.) Available at https://www.aaas.org/page/aaas-report-xxxix-research-and-development-fy-2015

Baumgarten, S., Simakov, O., Esherick, L. Y., Liew, Y. J., Lehnert, E. M., Michell, C. T., Li, Y., Hambleton, E. A., Guse, A., Oates, M. E., Gough, J., Weis, V. M., Aranda, M., Pringle, J. R., and Voolstra, C. R. (2015). The genome of *Aiptasia*, a sea anemone model for coral symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 11893–11898. doi:10.1073/pnas.1513318112

Baxevanis, A. D. (2011). The importance of biological databases in biological discovery. In 'Current Protocols in Bioinformatics', Chapter 1, Unit 1.1. (John Wiley & Sons: New York.) doi:10.1002/0471250953.bi0101s34

Bhattacharya, D., Agrawal, S., Aranda, M., Baumgarten, S., Belcaid, M., Drake, J. L., Erwin, D., Foret, S., Gates, R. D., Gruber, D. F., Kamel, B., Lesser, M. P., Levy, O., Liew, Y. J., MacManes, M., Mass, T., Medina, M., Mehr, S., Meyer, E., Price, D. C., Putnam, H. M., Qiu, H., Shinzato, C., Shoguchi, E., Stokes, A. J., Tambutté, S., Tchernov, D., Voolstra, C. R., Wagner, N., Walker, C. W., Weber, A. P. M., Weis, V., Zelzion, E., Zoccola, D., and Falkowski, P. G. (2016). Comparative genomics explains the evolutionary success of reef-forming corals. *eLife* **5**, e13288. doi:10.7554/eLife.13288

Couzin-Frankel, J. (2012). Research philanthropy. Billionaire restores funding for novel biomedical network. *Science* **338**, 873–874. doi:10.1126/science.338.6109.873

Davies, N., Field, D., Amaral-Zettler, L., Clark, M. S., Deck, J., Drummond, A., Faith, D. P., Geller, J., Gilbert, J., Glockner, F. O., Hirsch, P. R., Leong, J. A., Meyer, C., Obst, M., Planes, S., Scholin, C., Vogler, A. P., Gates, R. D., Toonen, R., Berteaux-Lecellier, V., Barbier, M., Barker, K., Bertilsson, S., Bicak, M., Bietz, M. J., Bobe, J., Bodrossy, L., Borja, A., Coddington, J., Fuhrman, J., Gerdts, G., Gillespie, R., Goodwin, K., Hanson, P. C., Hero, J. M., Hoekman, D., Jansson, J., Jeanthon, C., Kao, R., Klindworth, A., Knight, R., Kottmann, R., Koo, M. S., Kotoulas, G., Lowe, A. J., Marteinsson, V. T., Meyer, F., Morrison, N., Myrold, D. D., Pafilis, E., Parker, S., Parnell, J. J., Polymenakou, P. N., Ratnasingham, S., Roderick, G. K., Rodriguez-Ezpeleta, N., Schonrogge, K., Simon, N., Valette-Silver, N. J., Springer, Y. P., Stone, G. N., Stones-Havas, S., Sansone, S. A., Thibault, K. M., Wecker, P., Wichels, A., Wooley, J. C., Yahara, T., Zingone, A., and GOs-COS (2014). The founding charter of the Genomic Observatories Network. *GigaScience* **3**, 2. doi:10.1186/2047-217X-3-2

Dougherty, M. J., Wicklund, C., and Johansen Taber, K. A. (2016). Challenges and opportunities for genomics education: insights from an Institute of Medicine roundtable activity. *The Journal of Continuing Education in the Health Professions* **36**, 82–85. doi:10.1097/CEH.0000000000000019

Dubilier, N., McFall-Ngai, M., and Zhao, L. (2015). Microbiology: create a global microbiome effort. *Nature* **526**, 631–634. doi:10.1038/526631a

Dunn, C. W., and Ryan, J. F. (2015). The evolution of animal genomes. *Current Opinion in Genetics & Development* **35**, 25–32. doi:10.1016/j.gde.2015.08.006

Durant, W. (Ed.) (1938). Immanuel Kant and German Idealism: Transcendental Analytic. In 'The Story of Philosophy. p. 253. (Garden City Publishing Company: Garden City, NY.)

Genome 10K (2009). Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species. *Journal of Heredity* **100**, 659–674. doi:10.1093/jhered/esp086

GIGA COS (2014). The Global Invertebrate Genomics Alliance (GIGA): developing community resources to study diverse invertebrate genomes. *The Journal of Heredity* **105**, 1–18. doi:10.1093/jhered/est084

Green, E. D., Watson, J. D., and Collins, F. S. (2015). Human Genome Project: twenty-five years of big biology. *Nature* **526**, 29–31. doi:10.1038/526029a

Hemmrich, G., and Bosch, T. C. (2008). Compagen, a comparative genomics platform for early branching metazoan animals, reveals early origins of genes regulating stem-cell differentiation. *BioEssays* **30**, 1010–1018.

Koepfli, K. P., Paten, B., Genome, K. C. O. S., and O'Brien, S. J. (2015). The Genome 10K Project: a way forward. *Annual Review of Animal Biosciences* **3**, 57–111. doi:10.1146/annurev-animal-090414-014900

Koutsovoulos, G., Kumar, S., Laetsch, D. R., Stevens, L., Daub, J., Conlon, C., Maroon, H., Thomas, F., Aboobaker, A. A., and Blaxter, M. (2016). No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 5053–5058. doi:10.1073/pnas.1600338113

Liew, Y.J., Aranda, M., and Voolstra, C.R. (2016). Reefgenomics.Org – a repository for marine genomics data. *Database* **2016**, baw152. doi:10.1093/database/baw152

Parkhill, J., Birney, E., and Kersey, P. (2010). Genomic information infrastructure after the deluge. *Genome Biology* **11**, 402. doi:10.1186/gb-2010-11-7-402

Robinson, G. E., Hackett, K. J., Purcell-Miramontes, M., Brown, S. J., Evans, J. D., Goldsmith, M. R., Lawson, D., Okamuro, J., Robertson, H. M., and Schneider, D. J. (2011). Creating a buzz about insect genomes. *Science* **331**, 1386. doi:10.1126/science.331.6023.1386

Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., Iyer, R., Schatz, M. C., Sinha, S., and Robinson, G. E. (2015). Big Data: astronomical or genomical? *PLoS Biology* **13**, e1002195. doi:10.1371/journal.pbio.1002195

UNEP/CBD/COP (2002). Report of the sixth meeting of the conference of the parties to the convention on biological diversity: Decision VI/7: Identification, monitoring, indicators and assessments (UNEP/CBD/COP: The Hague: Netherlands.) Available at https://www.cbd.int/doc/meetings/cop/cop-06/official/cop-06-20-en.pdf

Appendix 1. The GIGA Community of Scientists (COS)

Marcin Adamski (University of Bergen, Bergen, Norway), Shobhit Agrawal (King Abdullah University of Science and Technology, Thuwal, Saudi Arabia), Agostinho Antunes (Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Portugal), Manuel Aranda (King Abdullah University of Science and Technology, Thuwal, Saudi Arabia), Andy Baxevanis (National Institutes of Health, Bethesda, MD, USA), Mark Blaxter (University of Edinburgh, Edinburgh, Scotland), Thomas Bosch (University of Kiel, Kiel, Germany), Heather Bracken-Grissom (Florida International University, Miami, FL, USA), Jonathan Coddington (Smithsonian Institution, Washington, DC, USA), Timothy Collins (Florida International University, Miami, FL, USA), Keith Crandall (George Washington University, Washington, DC, USA), Mikael Dahl (The Arctic University of Tromsø, Norway), Daniel Distel (Ocean Genome Legacy, Nahant, MA, USA), Casey Dunn (Brown University, Providence RI, USA), Michael Eitel (Ludwig-Maximilians-Universität München, München, Germany), Jean-Francois Flot (Université Libre de Bruxelles, Belgium), Gonzalo Giribet (Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA), Fabrizio Ghiselli (University of Bologna, Bologna, Italy), Steven H. D. Haddock (Monterey Bay Aquarium Research Institute, Monterey, CA, USA), Kevin Kocot (University of Alabama, Tuscaloosa, AL, USA), Yi Jin Liew (King Abdullah University of Science and Technology, Thuwal, Saudi Arabia), Audrey Majeske (University of Puerto Rico Mayagüez, Puerto Rico), Bernhard Misof (Zoological Research Museum Alexander Koenig, Leibniz, Germany), Christopher Mungal (Lawrence Berkeley National Laboratory, Berkeley, CA, USA), Stephen J. O'Brien (Nova South-eastern University, Dania Beach, FL, USA), Timothy Ravasi (King Abdullah University of Science and Technology, Thuwal, Saudi Arabia), Mauro de Freitas Rebelo (Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil), Ana Riesgo (Natural History Museum, London, UK), Christine Schnitzler (National Institutes of Health, Bethesda, MD, USA), Anja Schulze (Texas A&M University at Galveston, TX, USA), Billie Swalla (Friday Harbor Laboratory, WA, USA), Denis Tagu (French Institute for Agricultural Research, Paris, France), Robert Toonen (University of Hawaii at Manoa, HI, USA), Marcela Uliano da Silva (Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil), Sergio Vargas (Ludwig-Maximilians-Universität München, München, Germany), Xin Zhou (Beijing Genomics Institute, Shenzhen, China).