Integrative and Comparative Biology

Integrative and Comparative Biology, volume 60, number 2, pp. 397–401
doi:10.1093/icb/icaa059

Society for Integrative and Comparative Biology

SYMPOSIUM

Recommendations for Advancing Genome to Phenome Research in Non-Model Organisms

Karen G. Burnett,[1,*] David S. Durica,[†] Donald L. Mykles,[‡] Jonathon H. Stillman[§] and Carl Schmidt[¶]

*Grice Marine Laboratory, College of Charleston, 205 Fort Johnson Rd, Charleston, SC 29412, USA; [†]Department of Biology, University of Oklahoma, 730 Van Vleet Oval, Norman, OK 73019, USA; [‡]Department of Biology, Colorado State University, 1878 Campus, Fort Collins, CO 80523, USA; [§]Department of Biology, San Francisco State University, San Francisco, CA 94123, USA; [¶]Department of Animal and Food Sciences, University of Delaware, Newark, DE 19716, USA

From the Symposium "Building Bridges from Genome to Phenome: Molecules, Methods and Models" presented at the annual meeting of the Society for Integrative and Comparative Biology January 3–7, 2020 at Austin, Texas.

[1]E-mail: burnettk@cofc.edu

Synopsis The 2020 SICB Society-wide Symposium "Building Bridges from Genome to Phenome: Molecules, Methods and Models" brought together a diverse group of scientists to discuss recent progress in linking phenotype plasticity to changes at the level of the genome, epigenome, and proteome, while exploring the boundaries between variation and speciation. In a follow-up workshop, participants were asked to assess strengths and weaknesses of current approaches, to identify common barriers inhibiting their progress, and to outline the resources needed to overcome those barriers. Discussion groups generally recognized the absence of any overarching theoretical framework underlying current genome to phenome research and, therefore, called for a new emphasis on the development of conceptual models as well as the interdisciplinary collaborations needed to create and test those models. Participants also recognized a critical need for new and improved molecular and bioinformatic approaches to assist in describing function/phenotypes across phylogeny. Additionally, like all scientific endeavors, progress in genome to phenome research will be enhanced by improvements in science education and communication both within and among working groups.

Introduction

The Animal Genome to Phenome Research Coordination Network (AG2P RCN, https://ag2p.net) was created in 2015 and supported by the National Science Foundation (NSF) as a means of fostering the development of mechanistic approaches and infrastructure for understanding how changes at the genomic level are linked to physiological processes in non-model organisms. The AG2P RCN provides organizational support for collaboration and networking among investigators faced with the challenge of linking "-omics" (genomics, transcriptomics, and proteomics) data to whole-organism phenotype and performance, primarily by way of supporting workshops and meetings. From its beginning, the effort included both model and non-model organisms with the intent that the experience obtained studying model organisms could hasten discovery in non-model organisms. In this context, a "non-model" is defined as an organism for which a genome sequence is lacking or incomplete or is not amenable to classical mutational analysis of phenotype.

The AG2P RCN co-sponsored its first symposium and workshop at the January 2016 meeting of the Society for Integrative and Comparative Biology in Portland, OR. Four members of the AG2P RCN executive committee (Burnett, Durica, Mykles, and Stillman) organized the all-Society symposium, "Tapping the Power of Crustacean Transcriptomics to Address Grand Challenges in Comparative Biology." In a complementary workshop, participants were asked to identify some of the impediments that limited the use of transcriptomics by

Advance Access publication June 4, 2020

integrative organismal biologists. The resulting white paper based on that workshop (Mykles et al. 2016) emphasized the need to standardize and lower the barriers for analysis of transcriptomic data, so that scientists from many disciplines and research environments could develop consensus approaches for new advances linking genomic with phenotypic diversity. Notably, ongoing development of the Cyverse (https://www.cyverse.org/) platform and outreach by its facilitators to a wide variety of communities in the biological sciences has provided a powerful resource for analysis and sharing of transcriptomic and other large datasets for non-model organisms.

Four years later, at the January 2020 Annual Meeting of SICB, the Building Bridges from Genome to Phenome Symposium brought together investigators representing diverse professional rank, experience, as well as gender and ethnicity, to review the status of genome to phenome research. Our considerations engendered not just a retrospective view of progress to date or lessons learned, but also sought to identify the leading edges of, and main barriers to, genome to phenome research. Following 2 days of invited and complementary research presentations and discussion, >40 participants gathered at an on-site workshop to synthesize their perspectives on (1) major advances that came to light at the symposium, (2) major impediments to advancing the field, and (3) questions or efforts that should be prioritized for future support. It came as no surprise that the answers to each of these general questions often represented different sides of the same coin; major advances often lead to the identification of new impediments, which may be singled out for future support. In this white paper, we present major recommendations identified by the workshop participants along with some of the research advances and key barriers leading to those recommendations.

Recommendations: Focus areas for advancing genome to phenome research

Theoretical frameworks

A fundamental concern expressed repeatedly in group discussions was the lack of common, testable hypotheses, or conceptual frameworks for modeling how an organism develops or manifests a phenotype. While new and ever more powerful tools are becoming available to generate and analyze massive "-omics" datasets, there is a dearth of unifying theoretical constructs or models against which these new tools and bioinformatics processing power can be tested for their ability to predict or drive phenotypes. Among workshop participants, there was widespread and repeated emphasis on the need to support the development of such theoretical models, likely by collaborations involving biologists and mathematicians with expertise in machine learning, statistics, and mathematical modeling. Such frameworks for modeling phenotype would need to consider (1) genetic/epigenetic factors that contribute to phenotype, (2) plastic versus immutable phenotypes, (3) polygenic nature of phenotypes, and (4) systems approaches that simplify the axis of phenotype. Efforts to model how a cell works or how an organism functions by integrating modules of co-expressed genes or traits in networks, referred to as systems or synthetic biology, might serve as a platform for the derivation of a fundamental "central dogma" for the emergence of phenotype from the genome. The Systems and Synthetic Biology Cluster in the NSF Biological Sciences Directorate (BIO) offers relevant grant support for collaborations that aim for such a comprehensive understanding of complex interactions within biological systems across different scales (https://www.nsf.gov/funding/pgm_summ.jsp? pims_id=504863&org=MCB). Solving the puzzle of how genomes give rise to phenomes has also been identified as a question fundamental to the Reintegrating Biology Initiative—an NSF supported effort to integrate research methods and perspectives across different subdisciplines of biology in order to fully understand and realize the potential of biological systems (https://reintegratingbiology.org/). Short of the goal of developing a genome to phenome "central dogma," workshop participants saw a need for broadly aimed efforts to identify shared or fundamental mechanisms of phenotype emergence. For example, what commonalities in networking might govern gene function in different tissues of the same species (paralogies) and/or in the same tissue among different species (orthologies)? At the least, there is a need for a common understanding of the goals or a central framework for the genome to phenome question to facilitate collaboration and cooperation of researchers from across biological disciplines as well as from other fields.

Collaborations

Workshop participants repeatedly emphasized that diverse perspectives and expertise are vital to addressing genome to phenome questions. Collaborations in two major areas should be fostered by funding agencies—concept development and training/technical development.

Concept development

The kinds of systems and network analyses that are needed to develop theoretical frameworks for genome to phenome research require collaborations among biologists, mathematicians, modelers, and others, beginning at the earliest stages of experimental design. Such interdisciplinary groups are crucial to share ideas and construct new research concepts but are often difficult to form. Such conceptual collaborations would be encouraged by (1) creating and supporting more virtual and physical spaces for extemporary ideas to arise, such as think tanks or research retreats, (2) promoting networking mechanisms to identify and bring together an amalgam of statisticians, biologists, modelers, and other scientists who otherwise might not know each other or understand how their expertise might synergize, and (3) developing funding models that discourage unnecessary rivalry among laboratories and promote real or virtual multi-disciplinary research centers.

Training/technical development

Each step that is needed to bridge the genome to phenome gap is technically challenging. For scientists at all professional levels (students, postdoctoral fellows, junior, and senior investigators) new skills need to be learned on an on-going basis. However, the complexity of data analysis and the time required to become skilled in each new type of analysis can slow the pace of progress. For example, analysis of transcriptomic, genomic, and proteomic data requires experts who know the best tools and have developed the complex workflows that are required to complete and analyze the results of these data. Resources such as Cyverse (www.cyverse.org) offer training, tools, and workflows for data analysis but still require a high level of expertise to understand and troubleshoot outcomes. A construct is needed to establish ways to work collaboratively without physical contact to take advantage of existing expertise/researchers/teams that have already developed these complex workflows and to avoid "reinventing the wheel" in individual research laboratories.

Better resources for describing function/phenotypes across organisms

Workshop participants emphasized the need to continue asking genome to phenome questions in a wide variety of non-model organisms whose diverse life histories and natural genetic variation give rise to interesting and variable phenotypes. This allows investigators to ask a broad range of questions so that orthologous concepts can be explored across biological scales and phylogenies. The molecular and bioinformatic tools to address these questions in non-model organisms are increasingly becoming available and support for enhanced biological data storage and analysis and software development tools is encouraged. Website platforms where bioinformatics software has been developed and deployed, such as GitHub (www.github.com) and Cyverse, are also extremely valuable and their development should be ongoing. Several groups noted that interactions between users and programmers/informatics developers through the cloud are not always optimal. Greater feedback and interaction between users, as well as between users and developers, would broaden and improve the utility of existing applications and assist in the identification of bugs.

The "Building Bridges" workshop groups recognized that moving from big data ("-omics") to define phenome/function is still difficult in non-model organisms. For example, the Kyoto Encyclopedia of Genes and Genomes (KEGG) was developed to assign function from large-scale molecular datasets in model organisms. Most of the Symposium workshop groups expressed a need for "custom KEGGs" that are not just superimposed on non-model organisms but explore commonalities and diversity within gene network pathways being uncovered by comparative biologists, for example, pathways for molting and regenerating systems in invertebrates. Workshop participants called for new approaches to place function into a phylogenetic context. Pairings of statisticians and biologists might be able to develop such gene pathways/sets for biological functions in non-model organisms. This effort might be facilitated by developing links between data and analysis websites for related organisms (e.g., website for all tunicate genomes and access to tools for analysis) or for related biological processes across organisms (e.g., function at animal level). More fundamentally, participants saw methodological and database gaps in their ability to move beyond individual organism genome/phenome analyses toward an ability to examine genome to phenome interactions in an adaptive and evolutionary context.

Other key improvements in research tools and methods to support genome to phenome research

While access to and quality of transcriptome data sequencing, analysis, and interpretation was a major concern at the 2016 SICB Crustacean Transcriptomics workshop (Mykles et al. 2016), progress in this area has been rapid, such that transcriptomic data are now widely published for a

broad diversity of organisms. Notably, single cell genome-wide patterns of gene expression of non-model organisms are now being reported in the literature (e.g., Northcutt et al. 2019). Workshop participants indicated that a repository for such single cell data would be a useful resource for the genome to phenome research community.

More recently, concerns about access and quality have shifted into the arena of proteomic analyses. Approaches to proteomic analysis vary among laboratories, usually requiring expensive and complex instrumentation and software tools, which are not widely accessible. There is a critical need for standardized and widely available tools for proteomic analysis as well as data repositories for the broad range of non-model and model organisms being used for genome to phenome research.

Clustered Regularly Interspaced Short Palindromic Repeats associated protein 9 (CRISPR/Cas9) genome editing has recently become one of the most powerful molecular tools available to support genomic and phenotypic manipulations in both model and a limited number of non-model organisms (Knott and Doudna 2018). While it can be a powerful way to manipulate gene expression in subtle and relevant ways, CRISPR/Cas9 is still very difficult to develop in non-model organisms (e.g., federally protected species) and may be an inappropriate tool for addressing the question of interest (e.g., examining impacts of environmental perturbation on phenotypic variation in natural populations). New techniques and approaches must be developed that can be used to identify and quantify the links between genotype and phenotype. Such approaches might include, but not be limited to, the establishment of cell culture systems, improved techniques for extracting viable samples, and even single cells from complex tissues.

Support for smarter science and communication

With greater accessibility to techniques and resources to conduct "-omics" research comes the temptation to use these big data approaches to address every question. Workshop participants repeatedly emphasized that mechanisms should be identified to encourage junior and senior investigators to conduct focused, hypothesis-driven research framed in a genome to phenome context and to critically consider experimental design, power, and sample sizes when launching "-omics"-based experiments. Phenotype should drive the research question, not the scale or popularity of a technique.

Workshop groups commonly agreed that funding and laboratory "safe spaces" should be more readily available to try out unproven techniques that might yield novel, paradigm-shifting outcomes. If such novel approaches fail, then there should be more avenues for communication and encouragement to share negative results, so that failures are not repeated.

While multiple viewpoints in the areas of science education and communication arose during the workshop, these discussions gave rise to two major suggestions. First, it is increasingly evident that the discipline of computer science and coding is melding with biology. This should be reflected in educational programs at the undergraduate and graduate levels. Second, workshop participants noted that written, oral or poster presentations based on "-omics" datasets are often (1) narrowly focused on one biological pathway in response to phenotypic change, while ignoring the rest of the dataset or (2) make broad sweeping conclusions about the dataset, while offering only shallow interpretations. More effective ways to present the outcomes of research involving massive and complex "-omics" datasets should be developed and widely adopted, benefiting communication among scientists and with the public.

Conclusions

Despite having made great strides in handling and analyzing "-omics datasets, workshop discussion groups recognized that large gaps remain in our understanding of the linkages between genotypic and phenotypic variation. Moving forward, emphasis should be placed on the establishment of robust collaborative groups and real/virtual working spaces aimed at developing and testing conceptual frameworks that can bridge the gap between genome and phenome. This effort would benefit from advancements in laboratory methods, software platforms, and bioinformatic analyses, as well as from improvements in education and communication among scientific working groups. Such advancements should be aimed at enhancing our ability to link genotypic and phenotypic variation in the context of adaptation and evolution.

Acknowledgments

The authors thank the invited speakers and all those who participated in the complementary oral and poster presentations and workshop activities of the Building Bridges Symposium. The authors also extend their sincere appreciation to Program Officer Susan Williams and to the Burk & Associates staff for their assistance in organizing the symposium.

Funding

This work was supported by the National Science Foundation [Symposium Award IOS-1927470] and Animal Genome to Phenome Research Coordination Network [Award IOS-1456942]. We gratefully acknowledge the support of the Society for Integrative and Comparative Biology, the SICB Divisions of Comparative Physiology and Biochemistry, Comparative Endocrinology, Evolutionary Developmental Biology, Ecoimmunology and Disease Ecology, Ecology and Evolution, Invertebrate Zoology, Phylogenetics and Comparative Biology as well as The Crustacean Society and American Microscopical Society.

References

Knott GJ, Doudna JA. 2018. CRISPR-Cas guides the future of genetic engineering. Science (New York, N.Y.) 361:866–9.

Mykles DL, Burnett KG, Durica DS, Joyce BL, McCarthy FM, Schmidt CJ, Stillman JH. 2016. Resources and recommendations for using transcriptomics to address grand challenges in comparative biology. Integr Comp Biol 56:1183–91.

Northcutt AJ, Kick DR, Otopalik AG, Goetz BM, Harris RM, Santin JM, Hofmann HA, Marder E, Schulz DJ. 2019. Molecular profiling of single neurons of known identity in two ganglia from the crab *Cancer borealis*. Proc Natl Acad Sci U S A 116:26980–90.