Empowering 21st Century Biology

GENE E. ROBINSON, JODY A. BANKS, DIANNA K. PADILLA, WARREN W. BURGGREN, C. SARAH COHEN, CHARLES F. DELWICHE, VICKI FUNK, HOPI E. HOEKSTRA, ERICH D. JARVIS, LORETTA JOHNSON, MARK Q. MARTINDALE, CARLOS MARTINEZ DEL RIO, MONICA MEDINA, DAVID E. SALT, SAURABH SINHA, CHELSEA SPECHT, KEVIN STRANGE, JOAN E. STRASSMANN, BILLIE J. SWALLA, AND LARS TOMANEK

Several lists of grand challenges in biology have been published recently, highlighting the strong need to answer fundamental questions about how life evolves and is governed, and how to apply this knowledge to solve the pressing problems of our times. To succeed in addressing the challenges of 21st century biology, scientists need to generate, have access to, interpret, and archive more information than ever before. But for many important questions in biology, progress is stymied by a lack of essential tools. Discovering and developing necessary tools requires new technologies, applications of existing technologies, software, model organisms, and social structures. Such new social structures will promote tool building, tool sharing, research collaboration, and interdisciplinary training. Here we identify examples of the some of the most important needs for addressing critical questions in biology and making important advances in the near future.

Keywords: ecology, genomics, bioinformatics, cell biology, biological infrastructure

Biology is confronted with the need to answer fundamental questions about how life and natural systems evolve, are governed, and respond to changing environments. We need to understand the basic biological processes that drive life on this planet—those common to all organisms as well as those that provide unique adaptation to different environments. We also urgently need to identify all the life forms on this planet and understand their interrelationships and geographic distributions.

Biology must also apply new and existing knowledge to solve the pressing problems of our times, which include the environmental crises of global climate change, ocean acidification, biodiversity loss and the introduction of nonnative species, serious concerns for human health, emerging and pandemic diseases, and critical needs for agricultural and biofuel production. The urgency of these fundamental and practical needs has prompted scientists to begin to identify sets of "grand challenges" in biology (Denver et al. 2009, NRC 2009, Satterlie et al. 2009, Schwenk et al. 2009).

To succeed in addressing the challenges of 21st century biology, scientists must generate, have access to, interpret, and archive more information than ever before. This effort will involve analyses that span scales of time and space, from decoding information from genomes to extracting information from the environment on how organisms survive and reproduce (NRC 2009). Scientists need to learn how complex biological systems work across levels of organization, from cells to ecosystems, and across time scales, from the millisecond response of neural systems to the long-term response of evolutionary change. We need to be able to trace the effects of changes in DNA sequence or epigenetic regulation on multiple organismal phenotypes, understand how these changes affect ecological relationships, and have sufficient examples of these to begin to articulate new theories of organismal function and evolution. Addressing the challenges of 21st century biology requires integrating approaches and results across different subdisciplines of biology, such as genetics, development, physiology, ecology, and evolution, as well as technologies, information, and approaches from other disciplines, such as engineering, computer science, physics, chemistry, mathematics, and the geological and atmospheric sciences (figure 1).

However, biologists do not have the tools required to achieve this vision. For many important questions in biology, progress is stymied by a lack of the essential instruments to make rapid advances. In some cases, certain devices or technologies exist in other fields but are currently unavailable to biologists. In other cases, we need tools that scientists have not yet imagined. Developing those tools may require new technologies, applications of existing technologies, software, model organisms, and social structures to promote tool building, tool sharing, research collaboration, and interdisciplinary training. This article presents examples of what we believe to be the most important needs for tools to address critical questions in biology. We focus on the tools and the social structures needed to enable such tools; for an in-depth treatment of biology's grand challenges, see Denver and colleagues (2009), the National Research Council report *A New Biology for the 21st Century* (2009), Satterlie and colleagues (2009), and Schwenk and colleagues (2009).

Tools

Researchers need tools to enable high-throughput acquisition and synthesis of information at all levels of the hierarchy of biological organization, and across all biologically relevant
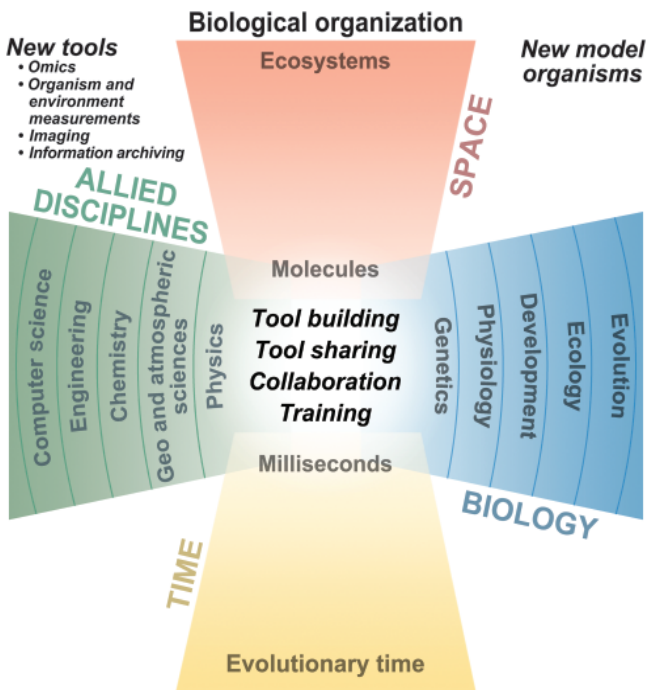
Figure 1. Tools for 21st century biology. To solve grand challenges, biology is becoming increasingly integrated across levels of organization, over different spatial and temporal scales, and it has become allied with other disciplines. Twenty-first century biology requires new tools that involve new technologies; new applications of existing technologies; new adaptations of tools from established model organisms to new models; new software; new model organisms; and new social structures to promote tool building, tool sharing, research collaboration, and interdisciplinary training.

spatial and temporal scales. These include technologies, software, and devices related to "omics"; informatics and systems biology; sensors and imaging; and information archiving.

Omics, informatics, and systems biology. The ability to sequence the genomes of microbes, plants, and animals has led to remarkable advances in biology. But this "first genomic revolution" has been based on the genome sequences of only a relatively small number of organisms: hundreds of microbes, and just a few dozen plant and animal species (*www.genomenewsnetwork. org/*). The relentless push to lower DNA sequencing costs for biomedical purposes will continue, and will soon make it possible to sequence the genomes of most species of interest for any biological question. Lower sequencing costs will usher in a "second genomic revolution," having a transformative effect on all areas of biology because genome sequence information can be used to illuminate questions at all levels of biological organization; we present just a few examples here.

DNA-based tools can have profound interdisciplinary impacts, beginning with faster and cheaper field identification of species and extending to assessments of genomewide patterns of genetic variation in different environments to determine what allows or limits the ability of individuals to

adapt to changing environments. Planning for the sequencing of 10,000 different vertebrate species' genomes has already begun (Haussler et al. 2009), and similar plans will certainly emerge for invertebrates and plants. In some cases, there also will be genome sequences for thousands of individuals belonging to the same species (Kuehn 2008, Weigel and Mott 2009). We envision many such projects for species that are important models in the laboratory or field, play particularly important ecological roles in different environments or that are situated at critical branch points in phylogeny.

Insights into the mechanisms and evolution of organisms can be gained with "ancient DNA" from specimens of archeological or historical importance (Millar et al. 2008), and from specimens collected over centuries and held in museums or natural science collections. Ancient DNA can be used to study major evolutionary patterns and diversification, extinction, and temporal changes in genetic variation—studies that compare Neanderthal and *Homo sapiens* genomes, for example (Green et al. 2006). Ancient DNA also could be used to understand molecular responses to past climate changes, and thus help predict potential responses in the future, but better tools are needed to facilitate the study of DNA. Improved DNA processing techniques that allow sequenceable DNA to be obtained from historic samples despite suboptimal preservation are needed to engender ever-more creative uses of ancient DNA in the future.

Metagenomics is revolutionizing the study of microbial ecology, from identifying new microbial species, strains, and genes to describing microbial communities associated with different parts of the human body. It is not far fetched to imagine the ability to extend this approach to eukaryotes, especially small ones. For example, strategically placed insect traps that feed into an automated metagenomic sequencing and informatics pipeline could be used to monitor outbreaks of agricultural pests or vectors of human disease.

New laboratory and computational tools also are required to leverage genome sequencing for a comprehensive omics revolution. Researchers need improved technologies for high-throughput interrogation of transcriptomes (including all forms of RNA transcripts) and novel methods for high-throughput *in situ* hybridization to precisely map changes in gene expression in order to illuminate our understanding of key biological processes.

To understand the complexity of biochemical processes in a living cell or organism, technologies to acquire comprehensive profiles of other molecules—such as metabolites (metabolomics), proteins (proteomics), elements (ionomics), and the stable isotopic composition of organisms—are just as important as genomic tools. The information obtained can be applied to a variety of problems, including the development of new drugs, improvements in human nutrition, and understanding the impacts of climate change.

Biologists need new bioinformatics methods to determine and compare genomes and functions across individuals and taxa, and to find and synthesize meaningful patterns within the floods of genomic data that will soon appear. New software

should be easily accessible to biologists, and computer programs should not require researchers to have extensive programming skills to use large, integrated data sets of genomic, phenotypic, phylogenetic, ecological, and environmental information for *in silico* hypothesis testing and discovery. There is a need for cyberinfrastructure composed of databases, communication protocols, and computational services designed to help make data and computational tools more usable for biologists (Giardine 2005, Stein 2008). And underlying this new cyber-infrastructure must be ever-more powerful computers to provide accessible high-end computing.

The impending omics revolution will require that traditional comparative genomics methods, which have served well in studies with few genes or genomes, give way to new visualization and analytical approaches that simultaneously can use thousands of genome sequences. The availability of thousands of genome sequences should make it possible to devise new algorithms to better detect orthologs in distantly related species in order to study the evolution of complex traits, such as flowering, development, and social behavior, or to discover new mechanisms of biofuel production.

Contemporary software is needed to extract and synthesize genomic data so that DNA sequences become, in essence, a new repository of biological information. Imagine an environmental biologist in need of a model species to study the risks of a certain type of environmental toxicant posing the following question: What species are particularly vulnerable to this environmental toxicant? This question prompts the software to perform an automatic knowledge-based search to identify genes from the literature known to encode the relevant detoxifying enzyme, using powerful text-mining algorithms in development (Muller et al. 2004, Ling et al. 2007). If an answer were available, then the program would conduct a Web search of all genome sequences to find and report on those species that lack the corresponding gene or genes, or those individuals within a species with particular single-nucleotide polymorphisms. Not all the information returned from such a search would be useful; for example, there is a variety of technical reasons a gene may be missing from an assembly of a particular genome. But the ability to move seamlessly between genome content and higher-level biological questions will provide biologists with a novel and powerful means of extracting information from genomes.

Systems analyses at all levels of biological organization promise to provide an innovative framework for understanding complex traits, including those at the molecular level. This will help illuminate how variation in genotype is related to variation in phenotype. Scientists will need new bioinformatic tools to integrate omics data into sophisticated models of gene and protein regulatory networks. To study how organisms adapt to environmental change, bioinformatics will need to be able to integrate the information derived from regulatory, signaling, and metabolic networks (Hyduke and Palsson 2010) with other types of phenotypic and population genetic data, ideally obtained across diverse environmental conditions. Automated methods of information analysis

are sorely needed, as are user-friendly software programs that facilitate the integration of data from multiple levels of biological organization across spatial and temporal scales, and that lead to predictive models for specific conditions. Biologists must be able to navigate by moving easily up and down the biological hierarchy from micro- to macroscales. Investigators also need user-friendly modeling tools with algorithms that can infer causal relationships from large data sets to help them develop hypotheses.

Scientific databases and literature must be more interactive and dynamic. Improved forms of text mining, already in development, employ statistical literature analyses to help identify new biomarkers for human disease (Shi et al. 2008) or explore how genes influence social behavior (Ling et al. 2007).

Sensors. Progress in many areas of biology depends on learning how genotypes generate specific phenotypes, how these relationships vary with environmental conditions, and how these relationships have evolved (Houle 2009). This requires new devices to measure organisms in their environment. Researchers need devices to enable acquisition of phenotypic and performance information that can be matched to genotypic and environmental information at all levels of the hierarchy of biological organization. Moreover, this information must be obtained under the vast array of natural and biological conditions in which organisms live and have evolved, and at biologically relevant spatial and temporal scales.

On a microscale, biologists need devices to continuously record the activity of cellular components as they interact naturally in living cells; on a macroscale, they need devices to continuously record the activity and performance of organisms and their component systems as they interact naturally in their environment. This instrumentation must be cost effective, miniaturized, and deployable in large numbers to continuously collect and transmit data in diverse environments, on small and large spatial scales. Automated image acquisition and shape-recognition software could permit the deployment of "smart" sensors that obtain information from specific organisms, both microscopic and macroscopic, and their environments, in real time. Stable isotopes already are being used as "natural sensors" because some enzymes (including rubisco and carbonic anhydrase) and biogeochemical processes (including temperature and precipitation) affect the isotopic profiles of organisms in characteristic ways. "Isoscapes," made from isotope analysis coupled with geographic information systems (GIS), are starting to reveal the relative importance of key physical and biological processes at continental scales (Bowen 2009). But isotopic analysis of biological materials is slow and labor intensive, and high-throughput methods are needed for this approach to realize its full potential. Advanced radio-tracking technologies will help with the study of dispersal and migration of small animals, such as bats, insects, and songbirds. It is expected that major innovations in radio-tracking technologies will make possible important new insights into conservation biology, climate change effects, and the spread of infectious disease (Wikelski et al. 2007).

Earth's diverse environments also need to be more intensely monitored if scientists are to appropriately contextualize new knowledge about an organism's physiological and behavioral responses. Researchers require new technology to cheaply monitor and measure many environmental parameters, including pH, temperature, conductivity, wind force, water flow rates and directions, dissolved oxygen, and mineral nutrient content, at biologically relevant scales, under controlled conditions and in the field, in real time. This information will help answer fundamental questions in biology that relate to genotype-phenotype linkages; it also is critical to understanding and predicting anthropogenic effects on natural resources and the impacts of climate change. It is hoped that the National Ecological Observatory Network (NEON; *www.neoninc.org*), funded by the US National Science Foundation (NSF), will provide large-scale terrestrial environmental data in the near future. Integrating NEON-type information with the above-mentioned measuring devices can then occur, enabling biologists to obtain information on individual organisms and to further study the impacts of climate change at all levels of biological organization.

Many forms of sophisticated sensory instrumentation and technology already exist, but they are not useful for discovering the linkages between genotypes, phenotypes, and the environment because of problems of scale. Many biological processes and environments are much smaller than current technology can measure. Miniaturization of instrumentation is critical for our understanding of basic life processes. Microfluidic devices are already starting to transform analyses of genomes, cells, and tissues in the laboratory (Whitesides 2006). These devices need to be adapted for wider use under an array of conditions, including in natural environments. For example, "mini–mass spectrometers" already exist that can be placed below sea level to study the adaptation of diverse forms of life to extreme environmental conditions (Bell et al. 2007). Further development of these technologies for a wide range of field-based applications could revolutionize real-time monitoring of organisms, populations, communities, and ecosystems.

Imaging. Biologists need handheld personal imaging systems that can be used or deployed in the field. Such imaging technologies would allow scientists to examine organisms in nature in detail, which is essential for making breakthroughs in fields such as sustainable agriculture, forestry, and conservation. These devices are now feasible, thanks to advances in real-time imaging. New methods, such as multilens cameras that allow post hoc adjustment of focus and depth of field and three-dimensional (3-D) image reconstruction (Bimber 2006), permit imaging in ways that are qualitatively different from what can be done with commonly used instruments. Deconvolution imaging makes it possible to overcome the limits of numeric aperture to produce images with resolution or depth of field that exceeds what is possible with a single image (Angel and Fugate 2000). Because these methods permit capture and integration of multiple traditional images into a single construct, they blur the line between image and computer model.

In microimaging, further development could make it possible to continuously monitor the 3-D structure of a developing organism, or to record the precise location and structure of every organism in an environment visible within the field of view. Raman spectroscopy is being used in the laboratory to provide detailed chemical analysis of specimens of ancient bone, shell, and teeth (Freudiger et al. 2008, Grant 2009). These and other imaging technologies need to be miniaturized, easy to use, portable, and cost effective. Such improvements would allow scientists to examine organisms in nature in detail, which is essential for making breakthroughs in fields such as sustainable agriculture, forestry, and conservation. It also would be possible to selectively sample organisms of interest and perform real-time, nondisruptive population monitoring, such as measuring the spread of invasive and pest species, including those that carry human disease. Discoveries could be made of the mechanisms that drive system and population resilience in the face of natural and anthropogenic disturbance or climate change.

In macroimaging, more user-friendly remote sensing and GIS would allow biologists to gain access to georeferenced data of all types. One widely used mapping tool is Google Maps, which has demonstrated the possibilities of using remote sensing and GIS technology to scientists for a variety of research purposes. Remote sensing and GIS technology can be improved in spatial resolution and image quality, and by the capability to integrate different types of data sets, including biological, geological, and topographical information (Makris et al. 2009). These innovations would increase the effectiveness of spatial modeling and habitat prediction algorithms to more accurately predict the spread of invasive or pathogenic species and the consequences of land-use change, for example, on local or global ecosystem scales. Improved GIS modeling and mapping could also help scientists predict when and where a new disease might emerge by revealing places where human hosts and certain animal vector species are in close proximity and at high densities. Better remote sensing and GIS predictive tools are especially needed to study the more remote regions of the world.

Information archiving. Even now, our ability to acquire biological data far outstrips our ability to store it in reliable and easily retrievable formats. This is true for all types of data, from genome sequence information to archived museum specimens, to the wealth of environmental data being collected. Biologists need modern methods of archiving, sharing, and accessing data. These needs will become even more acute with biology poised to acquire unprecedented amount of data, including reference data and reference specimens. In addition, researchers need easy access to the information, online or in person.

There is a need for improved software tools for deployment in online databases, seed banks, stock centers, museums, and other repositories of biological information. To be most useful, these repositories need to be curated, and must be replete with and searchable by different types of information (e.g., for organism specimens, DNA, species, time, and place of collection).

Such an update will require formalized ontologies for analytical data at all levels of biological organization (Ashburner et al. 2000), and formalized methods of recording metadata that describe the analytical data. Integrating and maintaining older legacy data poses other sets of challenges in the digital era.

New information technology is required to facilitate database creation. This software includes programs to facilitate uploading newly acquired data to centralized storage locations and programs that automate the process of creating and maintaining community-specific databases, such FlyBase (*www.flybase.org*) or WormBase (*www.wormbase.org*), which historically have required extensive, and increasingly prohibitively expensive, manual curation.

To effectively use different types of data to address a common problem, researchers need new tools for data integration across databases with different formats or that reside in different locations. For example, with the profusion of genome sequences expected to come in the near future, it is likely that an increasing number of sequences will reside only on the computer servers of individual laboratories rather than in centralized repositories like GenBank (*www.ncbi.nlm.nih.gov/Genbank/*), so it is imperative to develop software that can locate all the genome sequences in order to extract meaning from them. All of these needs again underscore the need for accessible high-end computing.

Model organisms

Over the past several decades, research efforts on plants and animals have increasingly focused on only a handful of model genetic organisms, especially thale cress (*Arabidopsis thaliana*), fruit fly (*Drosophila melanogaster*), worm (*Caenorhabditis elegans*), and mouse (*Mus musculus*). These species are especially useful for laboratory studies because they are relatively small bodied, have short generation times, can be maintained in the laboratory, and are comparatively easy to breed. An extensive array of genetic tools has been developed for these species, including finished genome sequences and advanced mutant and transgenic technology, such as transformation systems with control over spatial and temporal patterns of gene inactivation. However, these few species are not representative of the vast diversity of life. Twenty-first century biology would greatly benefit from a broader array of model organisms, including species not yet widely studied, to address a full range of important biological questions, especially those related to evolution and adaptation (Abzhanov et al. 2008, Brown et al. 2008, Behringer et al. 2009). But many organisms used to study evolution and adaptation lack the genetic and genomic tools necessary for the most rapid progress in scientific research.

It is now necessary to expand the number of model species so that a broader collection of questions can be studied effectively and efficiently. To accomplish this goal, scientists must use new model species to their fullest potential, using methods that work for a variety of species.

New ways are needed to facilitate the development of forward and reverse genetic techniques for the analysis of gene function. Viral vectors have been used to overexpress genes in a few animal species (Donaldson et al. 2008); innovations in viral modification could extend this technique more broadly. Gene knockdowns mediated by RNA interference (RNAi) are a powerful reverse genetics tool to analyze gene function, but this works better in some species and tissues than in others. The RNAi method also is limited by problems of delivery, knockdown efficiency, and artifacts resulting from off-target effects, and, in animals, innate immune responses. Enhancing the efficacy of RNAi across tissues and species, with methods that transfer easily across species, is an important goal. One critical component of empowering novel model organisms is the development of methods for genetic transformation that transfer easily from species to species. In animals, transposable elements such as PiggyBac show promise as vectors for transformation that can work on a broad variety of species, but additional research is required to build transformation systems that work efficiently for many species (Wu et al. 2007).

Improved proteomic and metabolomic technologies make it possible to use species as models for environmental-change research even without fully sequenced and annotated genomes (Epperson et al. 2004, Tomanek 2010). These tools would empower new model organisms for physiological, developmental, behavioral, ecological, and evolutionary research.

A key issue is whether new tools should be developed on a species-specific basis, or whether they can be applied broadly across taxa. Species-specific tools will always require significant investment, and involve careful justification of which species to choose (Mandoli and Olmstead 2000, Jenner and Wills 2007). We suggest that emphasis be placed on species that can be used to understand key evolutionary patterns and important biological phenomena that present the most immediately pressing need. Additionally, because we anticipate that some of the most significant advances in biological research in this century will involve integration across levels of biological organization and across scales of time and space, we also suggest a special focus to empower new model organisms that allow for such integration. It is hoped that as new technologies are developed, costs will lower, facilitating further development of an ever-increasing number of model organisms. If these goals are achieved, the designation of "model" organism will become less relevant over time, as more and more species will be accessible to investigation at multiple levels of biological organization.

People power

To make the most progress in 21st century biology, a scientific research culture that nurtures creativity, encourages and promotes tool building and sharing, and rewards scientists accordingly is essential. In addition, as biology becomes more and more interdisciplinary and information based, new training paradigms must prepare the next generation of biologists and tool builders. With the inter- and transdisciplinary development and use of tools comes the challenge of cross-disciplinary communication. This challenge is particularly vivid, for example, when computer scientists and biologists come together to develop and use bioinformatic tools. Bioinformatics must facilitate networking and the formation of virtual communities,

resources for those who can "translate" across disciplines, and institutional mechanisms that recognize and reward the work and time needed to bridge communication across disciplines and transfer knowledge and technology.

Tool building, tool sharing, and collaboration. Tools with transformational potential have been, and will continue to be, developed by inspired and highly motivated individual scientists and engineers. We advocate the creation of more collaborative mechanisms to enhance and facilitate the process of tool building, and offer a few suggestions here.

Many of us lament the limited opportunities for interactions on our own campuses between biologists and those in other fields, such as engineering. Engineers are frequently unaware of biologists' needs, and biologists do not always know what technologies engineers have already produced or invented. Similarly, engineers are not always familiar with the solutions that biological systems provide for a variety of engineering problems. Workshops that bring together engineers and biologists serve as a catalyst for innovative tool development and could help remedy this issue.

National or regional facilities could serve as tool-development incubators or tool-dissemination sources. For example, a center that brings together engineers and biologists could play a crucial role in the design, fabrication, testing, and use of microfluidic devices for both the laboratory and the field. Another center, involving a different mix of researchers, including perhaps geneticists, developmental biologists, and physiologists, could be formed to develop universal techniques of transgenesis. In some cases, innovation could arise from breaking down communication barriers between fields so that problems can be clearly seen from different disciplinary perspectives.

Interdisciplinary centers might be in physical locations, connected with universities or independent research institutes that serve as focal points to bring people together for periods of time. Examples of highly successful centers include the NSF-sponsored National Center for Ecological Analysis and Synthesis and the National Evolutionary Synthesis Center (NESCent) (Carpenter 2009). Field research stations and laboratories can be particularly effective venues for extended and informal cross-disciplinary interaction (Carpenter 2009).

Virtual interdisciplinary centers with geographically dispersed partnerships will help make transformations in biology. Virtual centers could be particularly useful for the development of some of the new bioinformatic tools outlined above. Funding mechanisms that encourage the development of virtual communities to address particular biological problems have seen strong success; for example, the National Institutes of Health–funded Glue Grants (*www.nigms.nih.gov/Initiatives/Collaborative/GlueGrants/*), NSF-funded Research Coordination Networks (*www.nsf.gov/funding/pgm_summ.jsp?pims_id=11691*), and the NSF iPlant initiative (*www.iplantcollaborative.org/*). iPlant is specifically designed to address the development of cyberinfrastructure to facilitate solutions to grand challenges in the plant sciences. The nanoHUB (*http://nanohub.org/*) is a virtual center that distributes newly developed computational tools through easily navigated interfaces to users at all levels of computer sophistication. It is easy to imagine additional virtual centers developing around new model organisms, technologies, or ways of integrating across levels of biological organization.

One highly successful social structure in science is the research community. Unlike hard disciplinary boundaries, research communities are self-assembling and dynamic, and often cross delineations of study. Clearly, a sense of community and wanting to belong is not just a human characteristic but also a desirable motivating force in science. Biological communities are often structured around organisms (e.g., model organisms), systems (e.g., plant communities), disciplines (e.g., comparative physiologists, biomechanics, and functional morphologists), or fields of interest (evolutionary development). Perhaps tool development can be facilitated by creating new research communities. Will important new advances be made if research communities are organized around tools or major problems, rather than the organisms or systems they study? This is an experiment worth trying.

New forms of networking in science, fueled by innovations in communication technology that operate on increasingly short time scales, can also contribute to changes in social structure to facilitate 21st century biology. Scientists use a growing number of networking tools for research and public outreach, including Google, as well as Drupal.org, Epernicus, Facebook, LinkedIn, MyExperiment.com, SciVee.tv, Skype, Twitter, and YouTube. These social-networking tools are showing promise in facilitating just the kinds of social processes that 21st biology requires. They provide friendly and informal environments that can aid the sharing of both tools and data. For example, Epernicus is a social-networking Web site and professional networking platform resource built by scientists to help scientists find the right people with the right expertise at the right time. Networking tools should also prove useful in developing new research communities and supporting new training opportunities.

New institutional incentives are also needed to enable building and sharing tools for biology. Although some tool-builders have been amply rewarded for their efforts, including with Nobel Prizes, the dominant motif in scientific research is discovery. The collective development of a new tool may not result in a publication in a high-impact journal, but a finding made with it might. Institutions need new ways of evaluating and recognizing collaborative efforts. Academic collaboration has long been valued in the abstract—most scientists recognize that the outcome of a successful collaboration is usually more than just the sum of individual parts, but traditional metrics of recognition favor individual achievement. Funding agencies and academic institutions have taken steps to incentivize scientists to form productive collaborative teams of researchers by establishing specific grants mechanisms for collaboration and creating interdisciplinary institutes, respectively. We expect these important paradigm-changing efforts to intensify in the future.

Training. Training paradigms are essential for preparing young biologists for the more extensive research collaboration that is needed for interdisciplinary work, and for providing them with the quantitative skills and broad perspectives necessary for success. Tomorrow's biologists must have training across genetics, development, biochemistry, physiology, ecology, and evolution, as well as experience working across different disciplines. More important, they also must have conceptual and quantitative training in mathematics and computer science to integrate these domains of knowledge using new computational tools. Wake (2008) emphasized the importance of training students early in their careers to not only think independently but also to work effectively in team-based scientific research.

Innovative training programs couple training with research in a team-based, problem-driven format. Undergraduate-focused universities provide opportunities for students to design and implement experiments and then analyze and present their results while receiving guidance and support from professors as well as peers. Innovative undergraduate training programs exist (Pevzner and Shamir 2009), and some undergraduates already are annotating genomes and using mass spectrometers for environmental metabolomic projects, for example. Investing more in research-intensive training programs at a variety of undergraduate-focused universities will increase the size of a diverse and highly motivated graduate student pipeline.

Conclusions

This article identifies some tools that are critically needed for biology to answer fundamental questions about how life evolves and is governed, as well as tools to apply this knowledge to solve the pressing problems of our times. We have tried to highlight possibilities for tools that integrate and affect disciplines and those that allow scientists to work across levels of biological organization—these will likely have the strongest influence on 21st century biology. We have outlined steps necessary to create the culture and social and educational structures that will facilitate and nurture tool development and toolmakers now and in the future. Scientists require more than new technologies, devices, and software; they also need to create and support a culture of science and education that stimulates and nurtures creativity, supports potential toolmakers, and trains the next generation of engineers. Many tools not yet imagined might make possible the next revolutionary biological discoveries; they might enable scientists to study remote areas of the world or reach and integrate underserved and underrepresented groups in science, thus encouraging progress toward common societal values for human health and the natural environment.

Acknowledgments

References cited

Abzhanov A, Extavour CG, Groover A, Hodges SA, Hoekstra HE, Kramer EM, Monteiro A. 2008. Are we there yet? Tracking the development of new model systems. Trends in Genetics 24: 353–360.

Angel R, Fugate B. 2000. Astronomy-adaptive optics. Science 288: 455–456.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H. 2000. Gene ontology: Tool for the unification of biology. Nature Genetics 25: 25–29.

Bell RJ, Short RT, van Amerom FHW, Byrne RH. 2007. Calibration of an *in situ* membrane inlet mass spectrometer for measurements of dissolved gases and volatile organics in seawater. Environmental Science and Technology 41: 8123–8128.

Behringer R, Johnson AD, Krumlauf RE, eds. 2009. Emerging Model Organisms: A Laboratory Manual. Cold Spring Harbor Laboratory Press.

Bimber O. 2006. Computational photography: The next big step. Computer 39: 28–29.

Bowen GJ, et al. 2009. Isoscapes to address large-scale Earth science challenges. Eos, Transactions of the American Geophysical Union 90: 109–110.

Brown FD, Prendergast A, Swalla BJ. 2008. Man is but a worm: Chordate origins. Genesis 46: 605–613.

Carpenter SR, et al. 2009. Accelerate synthesis in ecology and environmental sciences. BioScience 59: 699–701.

Denver RJ, Hopkins PM, McCormick SD, Propper CR, Riddiford L, Sower SA, Wingfield JC. 2009. Comparative endocrinology in the 21st century. Integrative and Comparative Biology 49: 339–348.

Donaldson ZR, Young LJ. 2008. Oxytocin, vasopressin, and the neurogenetics of sociality. Science 322: 900–904.

Epperson LE, Dahl TA, Martin SL. 2004. Quantitative analysis of liver protein expression during hibernation in the golden-mantled ground squirrel. Molecular and Cellular Proteomics 3: 920–933.

Freudiger CW, Min W, Saar BG, Lu S, Holtom GR, He C, Tsai JC, Kang JX, Xie XS. 2008. Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy. Science 322: 1857–1861.

Giardine B, et al. 2005. Galaxy: A platform for interactive large-scale genome analysis. Genome Research 15: 1451–1455.

Grant B. 2009. Sample, don't trample. The Scientist 23: 20–22.

Green RE, et al. 2006. Analysis of one million base pairs of Neanderthal DNA. Nature 444: 330–336.

Haussler D, et al. 2009. Genome 10K: A proposal to obtain whole-genome sequence for 10,000 vertebrate species. Journal of Heredity 100: 659–674.

Houle D. 2009. Numbering the hairs on our heads: The shared challenge and promise of phenomics. Proceedings of the National Academy of Sciences 107 (suppl. 1): 1793–1799.

Hyduke DR, Palsson BØ. 2010. Towards genome-scale signalling-network reconstructions. Nature Reviews Genetics 11: 297–307.

Jenner RA, Wills MA. 2007. The choice of model organisms in evo-devo. Nature Reviews Genetics 8: 311–319.

Kuehn BM. 2008. 1000 genomes project promises closer look at variation in human genome. Journal of the American Medical Association 300: 2715.

Ling X, Jiang J, He X, Mei Q, Zhai CX, Schatz B. 2007. Generating gene summaries from biomedical literature: A study of semi-structured summarization. Information Processing and Management 43: 1777–1791.

Makris NC, Ratilal P, Jagannathan S, Gong Z, Andrews M, Bertsatos I, Olav RG, Nero RW, Jeck JM. 2009. Critical population density triggers rapid formation of vast oceanic fish shoals. Science 323: 1734–1737.

Mandoli DF, Olmstead R. 2000. The importance of emerging model systems in plant biology. Journal of Plant Growth Regulation 19: 249–252.

Millar CD, Huynen L, Subramanian S, Mohandesan E, Lambert DM. 2008. New developments in ancient genomics. Trends in Ecology and Evolution 23: 386–393.

Muller HM, Kenny EE, Sternberg PW. 2004. Textpresso: An ontology-based information retrieval and extraction system for biological literature. PLoS Biology 2: e309.

[NRC] National Research Council. 2009. A New Biology for the 21st Century: Ensuring the United States Leads the Coming Biology Revolution. National Academies Press.

Pevzner P, Shamir R. 2009. Computing has changed biology: Biology education must catch up. Science 325: 541–542.

Satterlie RA, Pearse JS, Sebens KP. 2009. The black box, the creature from the black lagoon, August Krogh, and the dominant animal. Integrative and Comparative Biology 49: 89–92.

Schwenk K, Padilla DK, Bakken GS, Full RJ. 2009. Grand challenges in organismal biology. Integrative and Comparative Biology 49: 7–14.

Shi L, Perkins RG, Fanh H, Tong W. 2008. Reproducible and reliable microarray results through quality control: Good laboratory proficiency and appropriate data analysis practices are essential. Current Opinions Biotechnology 19: 10–18.

Stein LD. 2008. Wiki features and commenting: Towards a cyberinfrastructure for the biological sciences: Progress, visions and challenges. Nature Reviews Genetics 9: 678–688.

Tomanek L. 2010. Environmental proteomics: Changes in the proteome of marine organisms in response to environmental stress, pollutants, infection, symbiosis and development. Annual Review of Marine Science. (29 September 2010; *www.annualreviews.org/doi/pdf/10.1146/annurev-marine-120709-142729*)

Wake MH. 2008. Integrative biology: Science for the 21st century. BioScience 58: 349–353.

Weigel D, Mott R. 2009. The 1001 genomes project for *Arabidopsis thaliana*. Genome Biology 10: 107–111.

Whitesides GM. 2006. The origins and the future of microfluidics. Nature 442: 368–373.

Wikelski M, Kays RW, Kasdin J, Thorup K, Smith JA, Cochran WW, Swenson GW Jr. 2007. Going wild: What a global small-animal tracking system could do for experimental biologists. Journal of Experimental Biology 210: 181–186.

Wu S, Ying G, Qiang W, Capecchi MR. 2007. Toward simpler and faster genome-wide mutagenesis in mice. Nature Genetics 39: 922–930.

Gene E. Robinson (generobi@illinois.edu) is with the Department of Entomology, Neuroscience Program, and Institute for Genomic Biology, and Saurabh Sinha is with the Department of Computer Science, at the University of Illinois at Urbana-Champaign, in Urbana. Jody A. Banks (banksj@purdue.edu) is with the Department of Biology, and David E. Salt is with the Department of Horticulture and Landscape Architecture, at Purdue University, in West Lafayette, Indiana. Dianna K. Padilla (dianna.padilla@sunysb.edu) is with the Department of Ecology and Evolution, at Stony Brook University, in Stony Brook, New York. Warren W. Burggren is with Department of Biological Sciences at the University of North Texas, in Denton. C. Sarah Cohen is with the Tiburon Center and Department of Biology, at San Francisco State University, in Tiburon, California. Charles F. Delwiche is with the Department of Cell Biology and Molecular Genetics, at the University of Maryland, in College Park. Vicki Funk is with the Department of Botany, at the Smithsonian Institution in Washington, DC. Hopi E. Hoekstra is with the Department of Organismic and Evolutionary Biology, Museum of Comparative Zoology, at Harvard University, in Cambridge, Massachusetts. Erich D. Jarvis is with the Department of Neurobiology at Duke University, in Durham, North Carolina. Loretta Johnson is with the Division of Biology at Kansas State University, in Manhattan, Kansas. Mark Q. Martindale is with the Department of Zoology at the University of Hawaii, in Honolulu. Carlos Martinez del Rio is with the Department of Zoology and Physiology at the University of Wyoming, in Laramie. Monica Medina is with the School of Natural Sciences at the University of California, Merced. Chelsea Specht is with the Department of Plant and Microbial Biology at the University of California, Berkeley. Kevin Strange is with the Department of Anesthesiology at Vanderbilt University, in Nashville, Tennessee. Joan E. Strassmann is with the Department of Ecology and Evolutionary Biology at Rice University, in Houston, Texas. Billie J. Swalla is with the Department of Biology at the University of Washington, in Seattle. Lars Tomanek is with the Department of Biological Sciences at California Polytechnic State University, in San Luis Obispo.

"Lieberman finally got around to clearing off his desk and there they were."