Frontiers in Genomics: Insights into Protist Evolutionary Biology, University of Iowa, May 19–21, 2004

DEBASHISH BHATTACHARYA\textsuperscript{a} and LAURA A. KATZ\textsuperscript{b}

\textsuperscript{a}Department of Biological Sciences and Roy J. Carver Center for Comparative Genomics, 312 Biology Building, University of Iowa, Iowa City, Iowa 52242, USA, and

\textsuperscript{b}Department of Biological Sciences, Smith College, Northampton, Massachusetts 01063, USA

ABSTRACT. Protists constitute the bulk of eukaryotic diversity yet their genomes remain relatively unexplored. To address this issue, a workshop entitled, “Frontiers in Genomics: Insights into Protist Evolutionary Biology”, was convened at the University of Iowa on June 19–21, 2004. The specific aims of the workshop were to define the role of genomes in the eukaryotic tree of life, to identify challenges in characterizing protist (i.e. microbial eukaryote) genomes, and in proposing specific solutions to these challenges. The findings of the workshop are presented here and in a white paper that provide a set of guidelines for organizing the protist community and for planning and executing a protist genome project.

Key Words. ESTs, genomics, microbial eukaryotes, protists, workshop.

We are now at a unique juncture in time when comparative genomics can be coupled with available sequencing technologies and high-throughput sequencing capacity to explore the genomes of eukaryotic microbes (protists). Presently over 60 complete (or nearly complete) nuclear genomes have been determined from eukaryotes (e.g. nine plants, over 30 fungi, 10 apicomplexans, 14 metazoans, see http://www.ncbi.nlm.nih.gov/Genomes) and target taxa are increasingly being chosen from free-living protists in the tree of life (e.g. \textit{Thalassiosira pseudonana} [stramenopile], \textit{Tetrahymena thermophila}, \textit{Paramecium tetraurelia} [ciliate], \textit{Emiliania huxleyi} [haptophyte], \textit{Ostreococcus} spp. [prasinophyte green alga], \textit{Cyanidioschyzon merolae}, \textit{Galdieria sulphuraria} [Cyanidiales red alga]). In addition, over a hundred protist expressed sequence tags (EST) projects are complete or currently underway.

To explore this synergy and to further invigorate genomic studies of microbial eukaryotes, a workshop entitled “Frontiers in Genomics: Insights into Protist Evolutionary Biology” was convened at the University of Iowa in Iowa City on May 19–21, 2004. The workshop was co-organized by D. B. and L. A. K. and funded by the National Science Foundation (NSF). This report summarizes the proceedings of the workshop, which provided a framework for logically prioritizing microbial eukaryotes for future genome sequencing projects and for coordinating efforts among the diverse members of the scientific community who study eukaryotic microbes. A more detailed account of the workshop proceedings is available as a White Paper that can be downloaded at: http://www.biology.uiowa.edu/workshop/Genomics_of_Eukaryotic_Microbes.php.

The workshop participants came from many different backgrounds spanning protist biology to genomics and bioinformatics (see the workshop website for details, http://www.biology.uiowa.edu/workshop/). Four break-out groups each met in four separate sessions and reached a consensus on the activities, listed below, that are necessary to enhance genomic studies of eukaryotic microbes.

Workshop recommendations.

(i) Organize the community and demonstrate the critical nature of eukaryotic microbes. This report and the white paper are necessary steps to increase awareness in the broad scientific community of the tremendous diversity and importance of eukaryotic microbes. Furthermore, an initial steering committee was established to coordinate future efforts (Table 1). This committee was formed with the goal of capturing diversity on many of its axes, including scientific interest, location, and type of the researchers’ institution, as well as gender of participants. The steering committee will take an active role in forwarding efforts described in this report. Success in these efforts will be achieved most effectively if the international community is coordinated to take advantage of the diversity of existing funding and sequencing resources. At the same time, we believe that new programs need to be developed to explicitly target the genomes of eukaryotic microbes. To facilitate interactions, a listserv has been established that is focused specifically on genomic studies of eukaryotic microbes (further information can be obtained by sending an email to genomes@smith.edu).

(ii) Provide a model for selecting eukaryotic microbes for genomic approaches. Unambiguous criteria are required for choosing among the many possible microbial eukaryotes for further genomic studies. These criteria focus on a variety of factors, including the role of the species on our planet and its relevance to our lives. These criteria must be balanced with issues such as genome size, the ability to bring the species into culture, and genome complexity. Applying these specific criteria (Table 2), the working groups established a list of possible high priority organisms (see http://www.biology.uiowa.edu/workshop/Genomics_of_Eukaryotic_Microbes.php).

(iii) Promote relevant activities for outreach and education. The final recommendation is that education and outreach efforts relevant to understanding eukaryotic microbes need to be strengthened. Such efforts include support for existing collections (e.g. ALGOBANK [http://www.unicaen.fr/unicaen/util/lbbm/algobank/], ATCC [http://www.atcc.org/], CCAP [http://www.ife.ac.uk/ccap/], CCMP [http://ccmp.biggelow.org/], SAG [http://www.epsag.uni-goettingen.de/html/sag.html], UTEX [http://www.utexas.edu/research/utex/]) and courses (e.g. “Living in the Microbial World” [http://www.mbl.edu/education/]), as well as the development of new courses and workshops. One essential component is supporting the development of workshops dedicated to training researchers in the identification and cultivation of microbial eukaryotes. Such training is critical for providing the context for genomic studies over both the short and long term. Because there is already a serious lack of classical protistology workers in the field and thus, even fewer being trained, we have to more actively encourage the use of protists as educational tools in the classroom. We should also encourage the development of existing resources such as micro*scope (http://microscope.mbl.edu) to broaden public knowledge of microbial eukaryotes. Finally, as grants are announced for sequencing of genomes from eukaryotic microbes, Principal Investigators will be encouraged to request resources for this type of activity.

\textsuperscript{Corresponding Author:} Debashish Bhattacharya—Telephone number: 319-335-1977; \textsuperscript{FAX number:} 319-335-1069; \textsuperscript{E-mail:} debashishbhattacharya@uiowa.edu; \textsuperscript{Website:} http://www.biology.uiowa.edu/debweb/
Choosing the approach and funding the research. The workshop participants spent a considerable amount of time discussing how to effectively design a protist genome project and where to seek funding. It was agreed that applying a variety of genome technologies to eukaryotic microbes will generate a wealth of data and the choice of technology should be driven by the specific questions being addressed. Complete genome sequencing is essential for full knowledge of the biology of microbial eukaryotes. Such data will allow the discovery of novel genes and physiological pathways while providing comprehensive information on the genetic constitution of the organisms. At the same time, surveys of expressed genes using ESTs will be an important tool for the characterization of a greater number of microbial eukaryotes, in particular those with extremely large genomes and for supporting initial functional genomic studies. Of particular interest was the choice of funding agency. The recommendations that arose from this discussion are as follows:

(i) Investigators should target the organism to the appropriate funding agency. The National Science Foundation is a major source of funds for microbial genome sequencing projects that fundamentally advance basic biological concepts (http://www.nsf.gov/pubs/2005/nsf05512/nsf05512.htm). Other examples of appropriate agencies are the US Department of Energy Microbial Genome Program (http://microbialgenome.org/index.shtml) for genomic analysis of organisms that play a key role in environmental processes such as waste remediation, carbon management, and energy production, the NHGRI Comparative Sequencing Program at the National Institutes of Health (http://www.genome.gov/30002389), and the DOE Joint Genome Institute Community Sequencing Program (http://www.jgi.doe.gov/CSP/index.html), which provides the opportunity for sequencing for a variety of microbial eukaryotes.

(ii) Genomic data will also fit into the recommendations made by the US Commission on Ocean Policy (available at http://oceancommission.gov/welcome.html) when, for example, they address coral reef protection, microbial phylogenetic and genomic diversity, or the connection between release of anthropogenic wastes and the growth of pathogenic marine organisms.

(iii) It was strongly urged that researchers should seek alternative funding sources; i.e. tie in to agriculture, industry, and food production by demonstrating that target organism genome sequencing will benefit society. A further opportunity would be to provide incentives for multiple agencies to jointly fund large projects.

Data release. This was a controversial area of discussion among workshop participants and one that escaped a blanket solution. It became clear that different rules will apply to different situations and thus across-the-board standards are hard to define. For example, data release timing and manner depends on the aims of the project; i.e. is it meant to support a community (e.g. the Tetrahymena research community) which may entail ongoing release of data as it is produced or is it meant for a smaller-scale comparative genomics initiative? It may, for example, be appropriate for small labs to have more time to analyze and publish their own data to maintain a competitive edge, while still releasing their data in a timely manner (‘raw data released as it is available’ rather than on a 6-month embargo). Immediate release is favored for large research communities. It was felt that a good rapport with the community is essential; investigators should be inclusive and provide clear statements of their policies when projects are initiated.

Final thoughts. For many protist researchers, the goal of generating high-throughput genomic data from their model organism is a daunting task. To facilitate this process, we have assembled a set of guidelines that can be used to assess the feasibility of a genome project. Most importantly, the workshop participants felt that it was necessary to define the community for any particular microbial eukaryote genome program and to develop a concerted approach to working on projects through national and international collaborations. A number of other key considerations of a protist genome project included partnership with a genome sequencing center, clarification of the bioinformatics component before
sequencing, maintenance of cultures, and post-sequencing annotation. In general, the workshop provided an informal and convivial atmosphere to freely discuss issues associated with genome projects with experts on hand to answer questions and propose guidelines. These guidelines are discussed in more detail in the white paper.

An important follow-up to the workshop was a NSF-funded one-day symposium at the 2004 Annual Meeting of the Society of Protozoologists entitled “Genome Evolution in Microbial Eukaryotes”, co-organized by L.A.K. and D.B. The proceedings of this symposium will be published in an upcoming volume edited by the organizers that will include papers on genomic methods and genome evolution of microbial eukaryotes.

In summary, both the workshop and the symposium provided significant background material and inspiration to participants to initiate genome projects with microbial eukaryotes. It is the greatest desire of the organizers that the protist community will be spurred on by this initiative to form partnerships with genome centers and to seek funding to commence genome projects with their model taxa. Such studies will make even more apparent the critical importance of protists to understanding the breadth of eukaryotic evolution, and will ultimately lead to a more diverse set of genetic model systems to spur future biological research.

Received 01/10/05; accepted 01/10/05