

**U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
NATIONAL INSTITUTES OF HEALTH (NIH)  
NATIONAL LIBRARY OF MEDICINE (NLM)  
BOARD OF REGENTS (BOR) COMPARATIVE GENOMICS RESOURCE (CGR)  
WORKING GROUP (WG) MEETING  
FEBRUARY 1, 2023**

**MEMBERS PRESENT**

Kristi L. Holmes, PhD, Northwestern University, Chair

**EXTERNAL MEMBERS PRESENT**

Alejandro Sanchez Alvarado, PhD, Stowers Institute for Medical Research

Hannah Carey, PhD, University of Wisconsin-Madison

Wayne Frankel, PhD, Columbia University Medical Center

Ani W. Manichaikul, PhD, University of Virginia School of Medicine

Len Pennacchio, PhD, Lawrence Berkeley National Laboratory

Valerie Schneider, PhD, DHHS/NIH/NLM/NCBI, Executive Secretary

Kenneth Stuart, PhD, Seattle Children's Research Institute

Tandy Warnow, PhD, University of Illinois, Champaign-Urbana

Rick Woychik, PhD, NIEHS & NTP, NIH CGR Steering Committee Liaison

**EXTERNAL MEMBERS NOT PRESENT**

Cathy Wu, PhD, University of Delaware

**OTHERS PRESENT**

Anne Ketter, Sr. Product Manager, NCBI

Sarah Kinling, Strategic Communications Principal, MITRE

Danny McLean, Strategic Communication Specialist, MITRE

Kim Pruitt, Chief, Information Engineering Branch, NCBI

## **I. WELCOME AND INTRODUCTIONS**

*Kristi Holmes, PhD, Chair*

*Valerie Schneider, PhD, Executive Secretary*

Dr. Holmes greeted Working Group members and thanked them for their continued participation in the Working Group. She emphasized the value of members' expertise and perspective in both understanding the needs of the scientific community and guiding the development of CGR as an effective discovery tool to address those needs.

## **II. CGR PROGRESS UPDATE**

*Valerie Schneider, PhD, Executive Secretary*

*Working Group Members*

Dr. Schneider reviewed recent progress of the CGR project with a focus on maximizing its potential impact for scientific research communities. There have been several recent updates to the products which make up CGR. The Foreign Contamination Screen Genome Cross-Species Aligner (FCS-GX) tool has been integrated into the GenBank screening process. The NCBI Datasets command line tools and web-based Gene tables with ortholog links were also updated. In addition, other CGR data analysis tools have been updated with additional features and capabilities. Dr. Schneider encouraged Working Group members to promote the new Basic Local Alignment Search Tool (BLAST) testing community, particularly for testing of the CGR-associated ClusteredNR protein database.

Dr. Schneider noted that NCBI has completed over 200 engagement activities since the inception of the project, with support from MITRE. She highlighted some of the activities which have taken place since September 2022. An informational webinar providing an overview of CGR was held with a recording available online via YouTube. Multiple feedback sessions have been held, as well as conference presentations to the scientific community. Dr. Schneider reviewed an overview of engagement measures based on prior Working Group member input to engagement metrics and targets. Fewer training activities were held in Q1 of FY23 than targeted, but Dr. Schneider anticipated that the number of training activities will increase over the course of the year.

In addition to engagement activities, NCBI, with support from MITRE, is also working to develop tangible example case studies to demonstrate the potential impact of CGR on genomic research. Case studies will include a brief summary, as well as an extended description of the research project including examples of potential improvements enabled by using CGR. Over time, the aim is to derive user testimonials and example case studies through ongoing engagement with researchers. Dr. Schneider noted that per recommendation by the NIH CGR Steering Committee to contact authors of case studies and involve them in this outreach, 2 studies have been identified for which the NCBI plans to do so.

### **III. CGR TRADEOFFS EXERCISE**

*Valerie Schneider, PhD, Executive Secretary  
Working Group Members*

Dr. Schneider facilitated a Tradeoffs Exercise for Working Group members to provide feedback and guidance on key areas of CGR product development and project-level decision-making. The exercise provided a brief description of NCBI's current approach, as well as rationales for the possible directions for each area.

NCBI currently focuses on developing its own tools, rather than adapting existing third-party tools, in cases where existing tools do not meet user needs and/or the goals of CGR. Working Group members were asked to provide input on whether NCBI should continue to focus primarily on developing internal tools or identifying and adapting third-party tools for CGR. Members noted the potential for increasing community engagement through the incorporation of third-party tools. Members also discussed the potential role of NCBI as a curator or indexer of third-party analysis tools, in addition to adapting them for use with CGR. Based on Working Group member responses, it was agreed that NCBI will continue to focus slightly more on developing tools internally than on adapting external tools.

NCBI also currently prioritizes public adoption of genome screening and annotation tools, rather than the development of NCBI's capacity to run them as a public service on behalf of public users. Working Group members were asked to provide input on whether NCBI should run the tools as a public service or prioritize adoption and use by data submitters. Dr. Schneider noted that 65% of the eukaryotic assembled genomes submitted to GenBank are provided by the 6% of submitters who submit more than 10 assemblies and may be considered bulk submitters. She also noted that documentation for public use of screening and annotation tools can be developed. Per Working Group members' concerns whether prospective users would have adequate resources to use these tools, Dr. Schneider noted that the tools are being developed so as not to be cost prohibitive for public users; she also noted that some of these tools may be run in the cloud and would therefore not require users to have their own intensive computing power. Working Group members emphasized the importance of adequate support and access for public users, as well as consideration of the long-term goals of CGR within the capabilities and resources of NCBI.

Dr. Schneider noted that the Working Group members will be consulted on additional tradeoffs after the meeting.

### **IV. NCBI DATASETS DEMO AND DISCUSSION**

*Valerie Schneider, PhD, Executive Secretary  
Working Group Members*

Dr. Schneider presented a more detailed overview of the new NCBI Datasets resource, which provides intuitive web interfaces to interact with gene, transcript, genome sequences,

annotation, and metadata data from across NCBI databases. The resource supports the CGR project by facilitating access to eukaryotic genomic data. The tool delivers coherent data packages which adhere to Findability, Accessibility, Interoperability, and Reuse of digital assets (FAIR) principles and support data science. Downloads can also be configured to include multiple genomes, additional annotations, or additional metadata.

Dr. Schneider demonstrated the search for and download of a genome data package using the NCBI Datasets resource. The newly designed genome assembly webpage was highlighted, which includes statistics about the genome assembly for a particular organism as well as Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis of gene content based on NCBI annotations. The resource presents clear access to download package options, allowing users to select from various filetypes and files to include in the final data package. The “Browse Taxonomy” feature enables users to view the number of available genomes at each taxonomic level; Dr. Schneider welcomed Working Group member input on any additional information that may be helpful to be integrated into the feature. The command line tool, accessible on GitHub and available for use via the Conda package management system, was also presented and demonstrated.

Ongoing work for the NCBI Datasets resource will include the addition of organelle genomes, as well as the conversion of legacy genome record pages to Datasets pages. A manuscript detailing the tool and its features is planned. Dr. Schneider encouraged Working Group members to familiarize themselves with the resource and promote its use through outreach to the scientific community.

## **V. ENGAGEMENT COMMITMENTS**

*Kristi Holmes, PhD, Chair*

*Valerie Schneider, PhD, Executive Secretary*

*Working Group Members*

Dr. Holmes presented a review of the Working Group’s engagement commitments, highlighting engagements completed since September 2022 and reviewing the targets set by the Working Group for each engagement type. Dr. Holmes noted 1 completed opinion piece since September 2022, facilitated by materials provided by NCBI, developed with MITRE support. Per Dr. Schneider, NCBI is also currently developing a White Paper about the CGR project with MITRE contributions. Working Group members agreed to move forward with an opinion piece, coauthored by the Working Group; the publication to which the piece will be submitted remains to be determined. In addition, Working Group members noted the potential for coauthoring with members of research organizations or professional societies, and for publication in community newsletters in addition to academic journals.

Regarding community engagement through attending events, Dr. Holmes noted that extensive presentation materials about CGR are available for Working Group use. Per a member’s recommendation, Dr. Schneider confirmed that information about CGR will be compiled in a poster format for display at future events. Regarding connections with

professional societies and research organizations, Dr. Holmes requested that Working Group members notify NCBI when a new connection is in progress so that support can be provided through the preparation or adaptation of informational materials.

Dr. Schneider noted a new web form under development through which Working Group members can submit information about new engagement activity for tracking purposes. Per Dr. Holmes, Working Group members will be consulted after the meeting for partnership on community engagement efforts.

## **VI. NEXT STEPS**

*Valerie Schneider, PhD, Executive Secretary*

In addition to completing the remaining Tradeoff Exercises and further discussing engagement commitments, Dr. Schneider noted that Working Group members will be polled regarding the comprehensiveness of the support materials provided by NCBI to facilitate community engagement.