

**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
September 11, 2023**

MEMBERS PRESENT

Kristi L. Holmes, PhD, Northwestern University, Chair

EXTERNAL MEMBERS PRESENT

Hannah Carey, PhD, University of Wisconsin-Madison

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

Cathy Wu, PhD, University of Delaware

EXTERNAL MEMBERS NOT PRESENT

Alejandro Sanchez Alvarado, PhD, Stowers Institute for Medical Research

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

OTHERS PRESENT

Anne Ketter, Sr. Product Manager, NCBI

Sarah Kinling, Strategic Communications Principal, MITRE

Aron Marchler-Bauer, PhD, Lead Scientist, NCBI

Danny McLean, Strategic Communication Specialist, MITRE

Terence Murphy, Staff Scientist, NCBI

Kim Pruitt, Chief, Information Engineering Branch, NCBI

Ekaterina Sukharnikov, PMP, Project Manager, NCBI

I. WELCOME AND INTRODUCTIONS

Kristi Holmes, PhD, Chair

Dr. Holmes greeted Working Group members and thanked them for their continued participation in the Working Group. She introduced the agenda topics for the meeting and welcomed the Working Group members' continuing feedback on how best to engage the scientific community to promote CGR.

II. CGR PROJECT UPDATES

Kristi Holmes, PhD, Chair

Valerie Schneider, PhD, Executive Secretary

Working Group Members

Dr. Schneider reviewed the recent development of the CGR project and its potential impacts for scientific research communities. She noted the recent submission of two manuscripts, one regarding the Foreign Contamination Screen (FCS) tool and the other a white paper relevant to the entire CGR project. The CGR review white paper, which discusses the impact of comparative genomics on human health, has been accepted and will be published soon. In addition, manuscripts for two other CGR products on NCBI Datasets and the Comparative Genome Viewer (CGV) visualization tool are in preparation.

Dr. Schneider discussed the progress of the NCBI Eukaryotic Genome Annotation Pipeline, including a first Proof of Concept (POC) for the public annotation tool scheduled for Alpha testing in January 2024. Feedback from researchers interested in annotations for genome assemblies will also be used to develop this tool. Regarding NCBI Datasets, the legacy NCBI Assembly, Genome, and Record webpages have been retired. Instead, users will now be directed to the new NCBI Datasets webpages, which give a more seamless and modern experience in obtaining genomic data.

Dr. Schneider shared several ways that both Working Group members and CGR users can stay abreast of CGR activities and upcoming projects. The CGR Public Roadmap, available on the CGR website and updated quarterly, outlines project releases and projects anticipated for the future. The CGR Newsletter, distributed to more than 7,000 subscribers monthly, provides ongoing updates and an overview of upcoming CGR activities. Working Group members can also email CGR directly, visit the CGR website, subscribe to the CGR mailing list, or monitor the CGR BOR Working Group Microsoft SharePoint team site to learn more about ongoing CGR development.

The development of CGR outreach content and materials is also ongoing, focusing on collaborating with the community to enable comparative genomics through the NCBI tools and resources. User-specific language was created to help users understand the benefits of CGR products. Additional clusters have been added to the Basic Local Alignment Search Tool (BLAST) ClusteredNR database to help users explore taxonomic diversity and identify

valuable sequences for comparative analyses. Engagement with researchers and other stakeholders is also ongoing to standardize outreach and connections with different communities.

Dr. Schneider highlighted the 385 engagement activities since the CGR project's inception (as of June 30, 2023), which have contributed to updates to CGR messaging and the introduction of CGR products to various communities and public domain users. NCBI has met with the Alliance of Genome Resources to align ortholog and gene ontology terminology for CGR products, and with the National Human Genome Research Institute (NHGRI) Genomic Data Science Analysis, Visualization, and Informatics Lab-space (AnVIL) to discuss collaboration opportunities for implementing CGR products. NCBI will also begin to meet with the Vertebrate Genomes Project (VGP) every other month to share comparative genomics data and tools. In addition, NCBI has engaged with medical librarians through webinars, published nine blogs via the CGR Newsletter, held interactive events, and completed its first set of Impact Spotlights, which include literature-based case studies to enhance researcher engagement.

III. WORKING GROUP UPDATES

Hannah Carey, PhD, University of Wisconsin-Madison

Kristi Holmes, PhD, Chair

Valerie Schneider, PhD, Executive Secretary

Working Group Members

Dr. Schneider reviewed the documentation and tracking of engagement efforts using various tools such as NCBI Outreach Tracker, CGR User Research Log, CGR Marketing Tracker, NCBI Slack conversations, demo scripts, meeting notes, and the BOR Working Group Outreach Tracker. Overall, outreach metric goals have been fulfilled; however, Dr. Schneider noted that the implementation of the target of three training activities is still pending.

Dr. Schneider encouraged Working Group members to provide feedback on their outreach efforts and obstacles in sharing CGR tools with the scientific research community using the BOR Working Group Outreach Tracker, which is used to monitor the overall success of the CGR project. This feedback is critical for improving how NCBI can support members in connecting with the community about CGR products.

Dr. Hannah Carey, a Working Group member, described her experience interacting with biomedical researchers at a research conference, recognizing the potential to promote CGR in the context of cross-disciplinary genomic research. Dr. Schneider will continue to collaborate with Working Group members to foster additional user group-oriented community engagement channels.

IV. TRADE-OFFS: ENGAGEMENT PRIORITIES

Valerie Schneider, PhD, Executive Secretary

Working Group Members

Dr. Schneider presented the CGR Tradeoffs Exercise, allowing Working Group members to provide feedback on crucial aspects of CGR product development and project-level decision-making. The exercise provided an overview of NCBI's current stance on proposed topics and rationale for potential modifications in each area.

NCBI currently prioritizes outreach and engagement activities for CGR contamination screening and annotation tools to high-volume submitters (>10 submissions), who accounted for over 65% of the eukaryotic genome assemblies to GenBank in 2022. Working Group members were asked to provide input on whether to prioritize high-volume submitters or smaller scale submitters, who typically have less expertise in genome assembly and annotation. Based on member feedback, NCBI will continue to prioritize high-volume submitters, which will benefit the CGR project by promoting strong working relationships and obtaining high-value feedback from these entities to help ensure that CGR component tools meet the needs of users.

NCBI also currently focuses on developing training materials to introduce NCBI Toolkit tools and data to trainees, rather than to educators. Dr. Schneider emphasized that trainees learn about tools from peers and mentors, so promoting understanding of the CGR tool's benefits and values is crucial. However, educators, who often reuse material, should also be informed on how to include CGR in their own training materials to increase awareness of its tools and data.

The Working Group highlighted the importance of targeting trainees for CGR engagements as they are typically younger individuals who may be more receptive to new innovations and therefore more inclined to use CGR tools. They also expressed concern about smaller universities' inability to obtain CGR products and tools due to a lack of funding. Dr. Schneider recommended soliciting feedback from librarians at various institutions during upcoming NLM webinars to assess accessibility barriers for students.

V. FEEDBACK FROM 1:1 CONVERSATIONS

Kristi Holmes, PhD, Chair

Valerie Schneider, PhD, Executive Secretary

Working Group Members

Dr. Holmes reviewed the outcomes of the one-on-one conversations with Working Group members, which were aimed at fostering relationships, gauging members' interest in and capacity to assist the CGR project, identifying areas for effective outreach, and soliciting suggestions for efficiency. Key takeaways included focusing outreach efforts on students and trainees, as well as more targeted outreach to specific members of research communities. Members recommended tailoring messaging to be more user-specific and expressed interest in assisting in the development of use cases but were uncertain on how to help promote CGR in their own research communities. Members also proposed

broadening outreach efforts beyond the wet lab community to include other communities such as developers. Finally, Working Group members emphasized that quarterly meetings should continue to focus on outreach priorities and CGR impacts.

Dr. Schneider suggested collaborating with smaller groups amongst Working Group members to continue improving on outreach efforts. Dr. Holmes also highlighted additional emergent themes based on member feedback, including engaging with communities through domain meetings and training sessions, and ensuring CGR product information is conveyed effectively to grant administrators and trainees. Following the one-on-one conversations, the CGR team continued research community engagements per Working Group recommendations.

VI. PROTEIN DOMAINS CURATION DEMO

*Aron Marchler-Bauer, PhD, Lead Scientist, NCBI
Working Group Members*

Dr. Marchler-Bauer presented a detailed overview regarding the role of proteins and protein domains in comparative analysis of genomics. He presented several use cases including a detailed analysis and understanding of molecular function and phenotype at the level of protein and protein domains, as well as resources to help assemble data collections for specific purposes such as drug targets and drug repurposing. Dr. Marchler-Bauer reviewed a range of resources for protein domain curation, including tools for sequence analysis and annotation, domain architecture evolution, and molecular visualization.

Dr. Marchler-Bauer discussed the significance of the Conserved Domain Database (CDD) and the Subfamily Protein Architecture Labeling Engine (SPARCLE) in protein domain curation data collection. The CDD provides sequence alignment models for ancient domains and full-length proteins, while SPARCLE aids in functional characterization and labeling of protein sequences based on conserved domain architecture. Dr. Marchler-Bauer noted that 10,000 models for novel domain families from the protein 'dark matter' are being added to these databases to help expand domain architecture coverage and annotation.

Both long term and short-term goals for protein domain classification and protein domain architecture curation were shared. Long-term aims included combining data and tool feeds into annotation pipelines, which may contribute to improved quality of submissions. The curation infrastructure will also aid future high-value target activities that will follow the CGR project. Short-term goals include increasing consumer access to reagents for more accurate and comprehensive analysis, improving interoperability of classification resources through domain architecture attribution, and adding novel domain models to close functional annotation gaps for assemblies outside the immediate CGR scope.

VII. WRAP UP

Kristi Holmes, PhD, Chair

Valerie Schneider, PhD, Executive Secretary

Working Group Members

Working Group members evaluated, via polling, their experience and NCBI's support engaging their research communities to promote CGR. Dr. Holmes thanked Working Group members for their continued participation and for their valuable feedback. NCBI will also continue to provide Working Group members with resources and materials to help them promote CGR.