

PEQG - Pangenomics Workshop

Organizers: John Lovell, Avril Harder

Date: 9-June 2026 (Tuesday)

Time: 12:30-3:30p PT

Location: Fred Farr Room. Asilomar Conference Grounds, Pacific Grove, California

Conference Website: <https://genetics-gsa.org/peqg-2026/>

12:30-1:30p ... Part 1: Basics of pangenomics — alignments and annotations

12:30p Introductions and overview

John Lovell, HudsonAlpha Institute for Biotechnology

12:40p Whole genome alignments and graph construction: why and how of pangenome graphs

Avril Harder, HudsonAlpha Institute for Biotechnology

**Tutorial: [Summarizing and quality-controlling pangenome graphs](#)

1:05p Protein-coding genes: considerations for consistent annotations across pangenomes

Tomáš Brůna, DOE Joint Genome Institute

Tutorial: [Evaluating consistency of pan-genome annotations](#)

1:30-2:30p ... Part 2: Pan-gene analysis

1:30p Examining gene function and presence absence variation in Salmonidae

Airianna McGuire, University of Connecticut

Tutorial: [Identifying of gene presence/absence from alignments](#)

1:55p Synteny and structural variation in Carnivoran sex chromosomes

Ellie Armstrong, University of California, Riverside

Tutorial: [Mapping synteny with GENESPACE](#)

2:20p Group discussion and break

2:30-3:30p ... Part 3: Sequence diversity analysis in the context of pangenomes

2:30p HPRC2: A Near Complete Map of Common Genetic Variation

Prajna Hebbar, University of California, Santa Cruz

Tutorial: [Exploring UCSC Genome Browser resources and hosting your own assemblies](#)

2:55p Scalable Phylopangenomic Analysis Using PanMAMA and PanMAN

Russ Corbett-Detig, University of California, Santa Cruz

Tutorial: [Building pangenome trees and placing reads or samples onto them](#)

Alan Zhang, University of California, Santa Cruz

3:20 Conclusions / Airing of the grievances

**The tutorials are “walk-throughs”: output is printed and can be explored without actually running scripts. URLs here take you to repositories (those without hyperlinks will be updated once the tutorial is ready) containing all data to run through the tutorials on your own time. So, feel free to bring your computer and try stuff out. But there will not be a “hackathon” to troubleshoot things on your own machine. Note that access to power outlets in the Fred Farr Room may be limited.

Avril Harder — **Whole genome alignments and graph construction: why and how of pangenome graphs**

Graph-based tools are quickly expanding the investigative scope of multi-genome projects by increasing the ease and reliability of challenging tasks, such as genotyping structural variants from short-read data and mitigating reference bias. However, plant genomes pose particular challenges to applying these methods: many are polyploid and most are rife with repetitive sequences, rearrangements, and copy number variants. Highlighting diverse plant species, we identify the challenges of building pangenome graphs in these systems and the significant improvements that these graphs can provide over linear reference-based analyses. Taken together, our experiences highlight the gains that can be realized from leveraging multiple references simultaneously and the need to exercise both caution and creativity when applying graph-based approaches to complex systems.

Tomáš Brůna — **Protein-coding genes: considerations for consistent annotations across pangenomes**

Integrating protein-coding gene annotations into consistent pangene sets remains error-prone due to methodological inconsistencies and biological presence-absence variation. In this talk, I will present experimental findings from cotton and soybean pangenomes demonstrating how annotation pipeline differences, not biological variation, drive apparent gene PAV, and why careful integration is essential before any evolutionary or copy-number inferences can be made. The accompanying tutorial will cover practical approaches for evaluating annotation consistency across pangenome resources and strategies for achieving it.

Ellie Armstrong — **Synteny and structural variation in Carnivoran sex chromosomes**

Despite major advances in genome sequencing technologies, sex chromosomes, in particular the Y chromosome, remain poorly represented in most reference genome assemblies. Using long-read sequencing and HiC technology, we recovered near-complete Y chromosomes across a diverse set of carnivoran lineages, enabling one of the most comprehensive comparative assessments of Y-chromosome evolution in this clade to date. Comparative analyses of the resulting Y chromosomes reveal both conserved structural features and pronounced lineage-specific variation in size, repeat content, gene retention, and organization. Together, these patterns underscore the dynamic evolutionary history of the carnivoran Y chromosome and illustrate how improved assembly quality is reshaping our understanding of sex chromosome evolution.

Airianna McGuire — **Examining gene function, evolution and gene presence absence variation in Salmonidae**

Salmonids are an excellent system to study unique genomic features that underlie both phenotypic variation and adaptation. In some species, specific structural variants were associated with adaptation to water temperatures and differences in migratory behavior. Further, salmonids underwent a shared whole-genome duplication event (WGD) approximately 80 MYA; this autotetraploidization event is still evident in genomes today. This study leverages a pangenomic approach to assess structural variation, gene family evolution, repeat landscapes, and ohnolog function across 11 species representing all 5 genera and a variety of life-history strategies. Here, we describe gene families and their evolution across Salmonidae.

Prajna Hebbar — **HPRC2: A Near Complete Map of Common Genetic Variation**

A linear reference genome captures only a fraction of human diversity and cannot represent multi-copy gene families or expressed sequence variation. We present HPRC Release 2, containing 467 haplotypes from 233 individuals, a ~10-fold expansion over HPRC1, with telomere-to-telomere assemblies of most chromosomes and coverage of >99% of common variation in All of Us. Haplotype-resolved annotations recover ~20,000 protein-coding genes per assembly and reveal population-stratified expansions at hypervariable families including TBC1D3, AMY, and KIR. A matched long-read pantranscriptome from 206 individuals uncovers tens of thousands of reference-divergent transcripts and polymorphic TE-driven isoforms invisible to reference-based approaches.

Russ Corbett-Detig — **Scalable Phylopangenomic Analysis Using PanMAMA and PanMAN**

Phylopangenomic analyses require representations that jointly capture evolutionary history and genome-wide structural diversity, yet existing methods scale poorly with increasing sample size and complexity. We introduce PanMAMA and PanMAN, which are collectively a unified framework for large-scale phylopangenomics. PanMAN represents genomic variation as a single structure that simultaneously encodes both the pangenome and its underlying phylogeny, enabling coherent evolutionary inference across heterogeneous genomes. PanMAMA integrates sequencing data into this representation at scale. Together, these methods enable practical phylopangenomic representation and efficient read mapping for large-scale genomic datasets.