QUANTITATIVE AND SYSTEMS BIOLOGY COLLOQUIUM:
Uncovering the diversity and evolutionary history of structural variation using long-read haplotype-resolved genome assemblies

Runyang Nicolas Lou
University of California, Berkeley

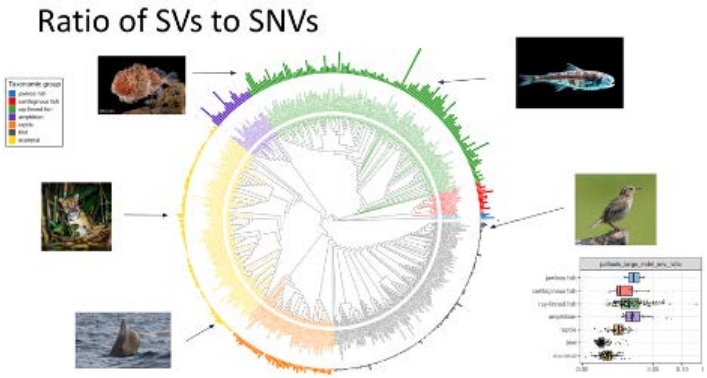
Date:
4/17/2025

Time:
10:30 AM - 11:45 AM

Location:
COB 1 114

About the Speaker:

R. Nicolas Lou is an evolutionary biologist broadly interested in population genetics, conservation genetics, and bioinformatics. Nicolas completely his PhD in 2022 in the Therkildsen lab at Cornell University. His PhD thesis explores the power and shortcomings of low-coverage whole genome sequencing and its applications in fisheries management and biodiversity conservation. He then started a postdoc in the Sudmant lab at Univeristy of California, Berkeley, where he studied structural variation evolution in diverse systems including human, chimpanzee, rockfish, and other vertebrate species. He is now an Assistant Project Scientist in the Sudmant lab.



Abstract:

Structural variants (SVs) contribute substantially to genetic variation and play vital roles in fundamental biological processes such as adaptation and disease. However, due to their complex nature, SVs have been less studied compared to single nucleotide variants (SNVs). In this talk, I will share two stories where we took advantage of recently generated long-read haplotype-resolved genome assemblies to study this vital component of genetic variation that has so far been elusive. In the first story, we focus on the amylase locus in the human genome, which harbors three amylase genes that are vital for starch digestion. All three amylase genes exhibit extensive copy number variation, and previous research have shown that agricultural populations tend to harbor more copies of these genes. Puzzlingly, this locus does not exhibit typical signatures of recent selection. Using 94 genome assemblies from diverse human populations, we identify 28 distinct amylase structural architectures and demonstrate that nearly identical structures have arisen recurrently on different haplotype backgrounds throughout recent human history. This recurrent evolution makes neighboring SNVs poor predictors of the haplotype structure, explaining the failure of SNV-based selection scans. Instead of relying on neighboring SNVs, we develop a pangenome-based method to directly infer haplotype structures from short-read data. Applying this method on 533 ancient human genomes, we find that duplication-containing haplotypes (with more gene copies than the ancestral haplotype) have r0apidly increased in frequency over the past 12,000 years in West Eurasians, consistent with positive selection following human's adoption of agriculture. In the second story, we broadly survey the diversity of SVs and SNVs across the vertebrate tree of life using genome assemblies from ~500 vertebrate species. We identify SVs and SNVs that segregate across two representative haplotypes in each genome assembly, and by analyzing SV length profiles and performing de-novo repeat annotations on SVs, we characterize lineage-specific expansions of transposable element activities across different taxonomic groups (e.g. Alu and L1 elements in primates, LTR elements in birds). While the levels of diversity in SVs and SNVs are generally positively correlated, we find that fish and amphibians tend to have more SVs than amniotes given the same number of SNVs. In addition, endangered or threatened species tend to have lower levels of diversity in both SVs and SNVs. SVs are more likely to be deleterious than SNVs, and species harboring less genetic diversity tend to have a higher proportion of deleterious variants. Together, these stories highlight the unique characteristics of SVs and their important roles in evolution and biodiversity conservation.

For more information, contact: Bailey Carlson
bcarlson4@ucmerced.edu