

SynMapN: Comparing multiple genomes

Mingwei Li
mwli@email.arizona.edu

Andreina Castillo Siri
aicasti1@email.arizona.edu

Asher Keith Haug-Baltzell
ahaug@email.arizona.edu

Eric Lyons
elyons.uoa@gmail.com

Carlos Scheidegger
cscheid@email.arizona.edu

Background

With the decrease in DNA sequencing costs, researchers are generating and reusing exponentially more genomic data. However, existing tools do not easily scale to visualize multiple genomes.

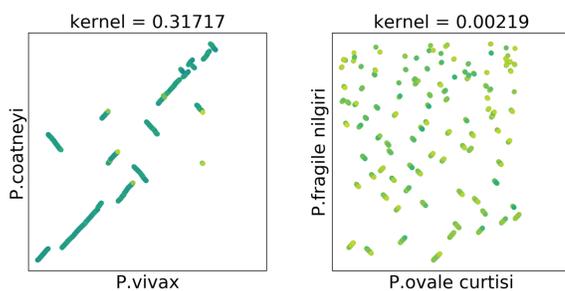
For example, *SynMap2* only depicts relationship between two genomes, while *SNP visualization with JBrowse* only displays SNPs data for a small number (~10) of individuals.

We are exploring scalable methods of visualizing both multiple genomes and individuals at the same time.

Goals

- (g1) Visualize multiple genomes and multiple individuals separately.
- (g2) Visualize multiple genomes and individuals at the same time.
- (g3) Select gene region of interest and recompute similarity within the selected region.
- (g4) Attribute genome affinity back to certain genes.
- (g5) Scale up to ~20 genomes x ~50 individuals

Method



$$kernel(A, B) = \frac{\sum_{i=1}^K e^{-k_i/\lambda}}{\sqrt{m \cdot n}}$$

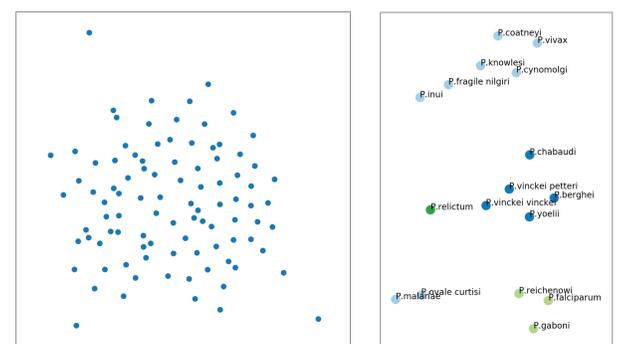
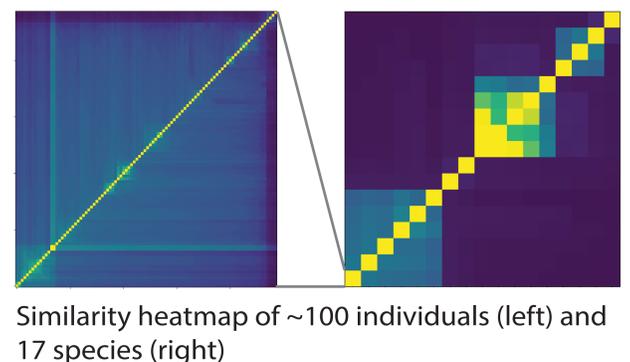
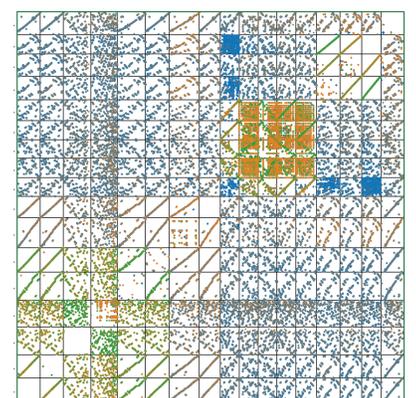
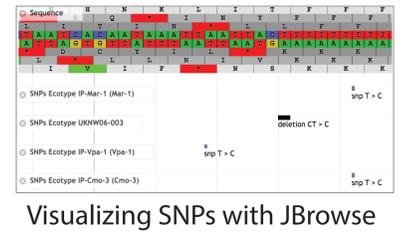
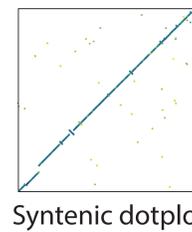
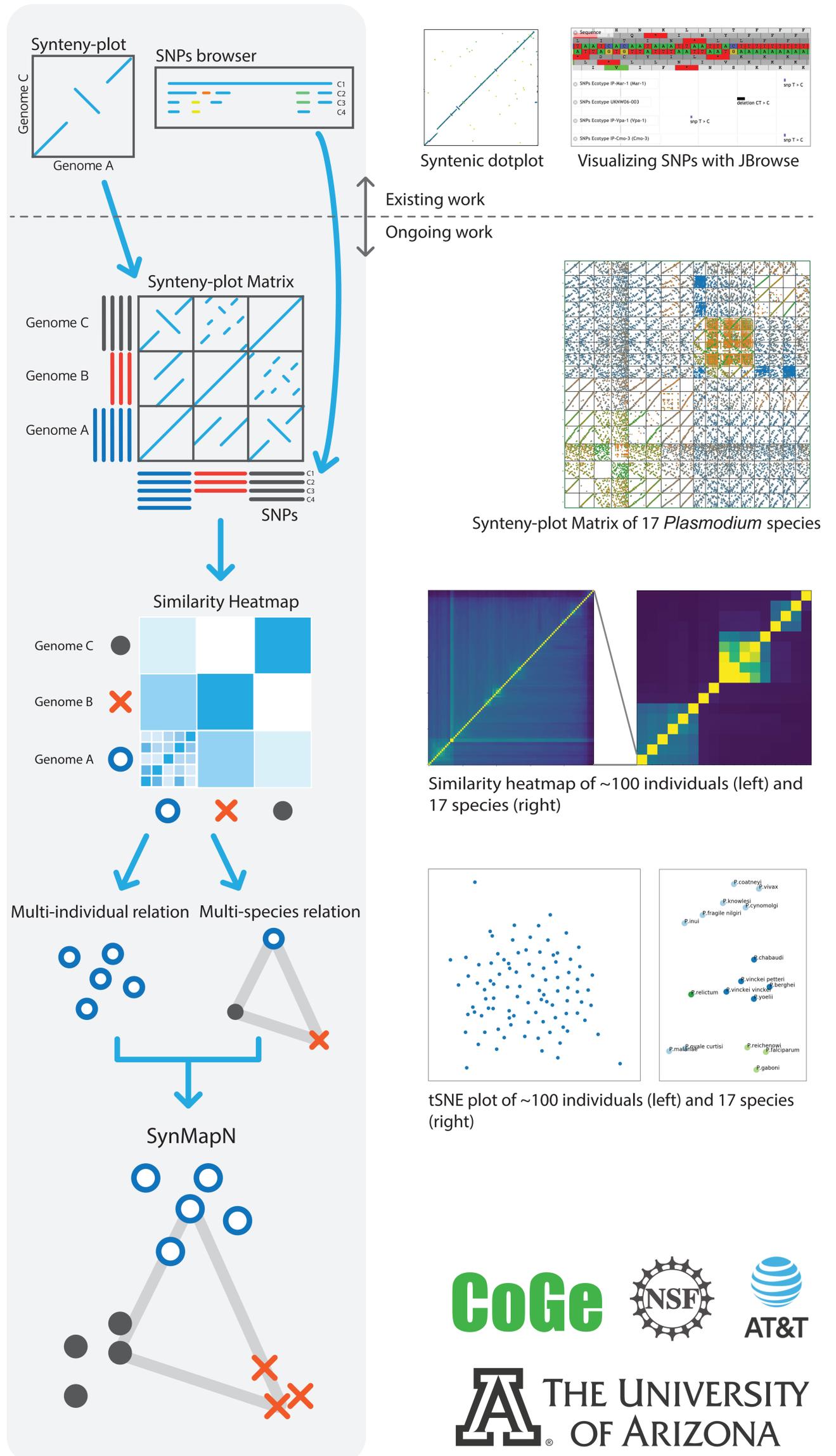
- Build inter-species kernel function from synonymous mutation rates of syntenic regions.
- Build intra-species kernel function among individuals from SNPs data.
- Standardize the two measurements through syntenicity.
- Create a link between Syntenicity matrix and SynMapN to create multi-scale zoom-and-filter views.

References

- SynMap2 & SynMap3D: Web-based whole-genome syntenicity browsers. Asher Haug-Baltzell, Sean Stephens, Sean Davey, Carlos Scheidegger, Eric Lyons. Bioinformatics DOI: 10.1093/bioinformatics/btx144 (2017)



<http://hdc.cs.arizona.edu>



Partially supported by: IIS-1513651