

# **Inferring Hierarchical Population Structure through a Consensus of Significant Evolutionary Events**

Mahdi Belcaid\*, Guylaine Poisson, David Schanzenbach, Kyungim Baek\*  
{mahdi, guylaine, davidls, kyungim}@hawaii.edu

Department of Information and Computer Sciences  
University of Hawai'i at Mānoa  
1680 East-West Road, POST 317  
Honolulu, HI 96822

Despite its paramount importance in a large number of bioinformatics analysis, estimating the structure of a population remains a complicated task, both from information theoretical and computational complexity perspectives. When reduced to a small number of individuals and a small set of SNPs, Ancestral Recombination Graphs are often used to gain insight about a population substructure. These graphs closely describe the putative hierarchical relationships, in terms of mutations, recombinations, and coalescence events, through which a set of extant sequences might have arisen.

Our method describes how information from partial Ancestral Recombination Graphs can be reconciled to generate a statistically more robust description of the relationships between individuals of a population. Our approach builds a consensus topology of the data by using empirically significant events, across multiple sub-regions of the input dataset. This technique is resilient to missing information and can also incorporate unphased SNPs.

We tested our approach on a dataset comprised of 525 individuals from 4 distinct populations, two of which considered to be admixed, and show that even with relatively small subsets, taking the information content into consideration allows us to capture the underlying population substructure, even for admixed populations.

---

\* Corresponding authors: kyungim@hawaii.edu (K. Baek), Mahdi@hawaii.edu (M. Belcaid)