Inference of historical demographic events from contemporary genomic sequence data has received a lot of attention in recent years. A particular focus has been on the recent exponential growth of population size in humans. This recent growth strongly impacts the distribution of rare alleles, which are of importance when studying disease related genetic variation. The popular method PSMC (Li and Durbin, 2011) is used to infer population sizes from a sample of two chromosomes. However, the small sample size severely limits the power of this method in the recent past.

To improve inference in the recent past, we extend the Coalescent Hidden Markov model approach of PSMC to larger sample sizes. Since using the full genealogical trees relating the sample at each locus is computationally prohibitive, we introduce a flexible mathematical framework to employ different representations of these local trees. In particular, we present the implementation of this framework using the height of the local trees (TMRCA), corresponding to PSMC for sample size 2, and using the total branch length of the local trees.

We evaluate the different representations in simulation studies and applications to genomic variation data from diverse human populations. We discuss potential extension of the framework to infer divergence times and migration rates in structured populations, and employing the posterior distribution of the local trees to detect regions under selection.