







## Fine-mapping of selective sweeps using allele frequency differentiation across populations

Florian J. Clemente<sup>1</sup>, Mathieu Gautier<sup>1,2</sup>, and Renaud Vitalis<sup>1,2</sup>

<sup>1</sup>Institut de Biologie Computationnelle, University of Montpellier, France <sup>2</sup>Centre de Biologie pour la Gestion des Populations, INRA, France

**Contact:** florian.clemente@univ-montp2.fr





revolution in sequence data generation has led to various new Ihe



We simulated four scenarios for

approaches trying to access the hidden information in these data. For example, [1] developed a model to detect selective sweeps based on allele differentiation to a neutral reference population. Recently, this model was extended to account for selection in the ancestral population of a 3-population tree [2].

Here, we present an extension of KimTree, a method for estimating divergence times among populations based on the Kimura diffusion approximation [3].

Following [1] and [2], but using a re-parameterised model in a hierarchical Bayesian framework, our approach may be used to detect the locus under selection for a previously found selection signal in any population tree.

![](_page_0_Picture_16.jpeg)

![](_page_0_Picture_17.jpeg)

We consider a model where the change of allele frequencies over generations for a given tree is described by the Kimura diffusion approximation for genetic drift [3] (Fig 1).

Figure 3: The boxplot shows 50 posterior means of the position under selection (red line indicates true value). The strength of selection was set to be s=0.1 with N<sub>e</sub>=1000 and sample size n=100.

Moreover, we applied our method to human data from [6], illustrating the well-known case of positive selection on SLC24A5 involved in skin pigmentation in Europeans and on CPT1A involved in the fat metabolism North-East Siberians (Fig 4 and 5).

a tree with three sampled populations using msms [5]. Each scenario consists of 50 data set replications with about 4000 neutral SNPs and one beneficial mutation in one of the populations. We used these data to evaluate the performance of our method in inferring the position under selection (Fig 3).

![](_page_0_Figure_22.jpeg)

Figure 1: Illustration of the KimTree model with selection. The lines represent the historical frequencies of an allele, which experienced a selective sweep in population 2. The model parameters are estimated from the observed SNP frequencies in all populations.

![](_page_0_Figure_24.jpeg)

Figure 2: Distortion due to linkage around the selected site. The joint effect of selection and recombination enter the model via the  $\delta$  parameter. The strongest deviation from neutrality occurs at the selected site and decreases in both directions along the chromosome. The amount of linkage is controlled by the  $\lambda$ parameter.

To model the impact of a selective sweep, allele frequencies are transformed according to the joint effect of selection and recombination [4]. This transformation is determined by the  $\delta$  parameter, which we define in terms of the distance between the position under selection and SNP *i*, and a scaling factor  $\lambda$ :

 $\delta_i := \exp(-\lambda |p_i - p_s|)$ 

A value of  $\delta=0$  indicates neutral evolution, whereas values of  $\delta > 0$ correspond to sites influenced by selection. Figure 2 shows the distortion of neutrality around a selected site for different values of  $\lambda$ .

We use a hierarchical Bayesian in a MCMC framework to model estimate the original KimTree model parameters together with the parameters of the selection model (position under selection and  $\lambda$ ).

![](_page_0_Figure_30.jpeg)

![](_page_0_Figure_31.jpeg)

![](_page_0_Picture_32.jpeg)

We provide a new method for estimating the locus under selection for a

tree of N populations. Our approach is based on allele frequency differentiation among these populations and takes advantage of relying on the accurate Kimura diffusion approximation to model their evolution.

## References

[1] Chen et al., (2010). Population differentiation as a test for selective sweeps. [2] Racimo, F., (2015). Testing for ancient selection using cross-population allele frequency differentiation. [3] Gautier, M. and Vitalis, R. (2013). Inferring population histories using genome-wide allele frequency data. [4] Smith J.M. and Haigh, J., (1974). The hitch-hiking effect of a favourable gene. [5] Ewing G. and Hermisson J., (2010). MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus.

[6] Clemente et al., (2014). A Selective Sweep on a Deleterious Mutation in CPT1A in Arctic Populations.

We show that our method works well on simulated data and is able to pick up a selection signal in real data sets. Current limitations are the sensitivity to local maxima, which can hinder the Markov chain in exploring the parameter space using a single chain.