

# Generalized Viterbi algorithm for identifying the multiple most likely paths of hidden states in a HMM and application to detection of homozygosity by descent

## Authors:

Matthieu Pluntz<sup>2,3,\*</sup>, Sidonie Foulon<sup>1,2,3</sup>, Anne-Louise Leutenegger<sup>2</sup> et Hervé Perdry<sup>3</sup>

\*corresponding author: matthieu.pluntz@inserm.fr

## Affiliations:

1. Université Paris-Saclay, UVSQ, Inserm, Gustave Roussy, CESP, 94805, Villejuif, France
2. NeuroDiderot, Inserm, Université Paris Cité, UMR1141, 48 bd Sérurier, 75019, Paris, France
3. Université Paris-Saclay, UVSQ, Inserm, CESP, 94807, Villejuif, France

## Keywords:

Hidden Markov Models, Viterbi algorithm, best paths, homozygosity-by-descent

## Abstract:

In Hidden Markov models (HMM), inference of the unobserved Markov chain across hidden states is known as decoding. The Viterbi algorithm is an efficient decoding algorithm which finds the single most probable path of hidden states conditional on the observations. It contains a forward step which computes recursively for each state the maximal probability, conditional on the observations up to a step, of a path reaching that state at that step, as well as the previous state in that path (traceback). Then comes the backward reconstruction of the optimal path by concatenating tracebacks.

We propose a generalization of the Viterbi algorithm ( $\alpha$ -Viterbi) which, for a given probability ratio  $\alpha \in ]0,1]$ , computes the set of paths having a posterior probability at least  $\alpha$  times that of the optimal path computed by the classical Viterbi algorithm. This is done by computing in the forward part of the algorithm, instead of a single traceback, a probability ratio for each pair of consecutive states. Then we use this ratio to construct the set of paths in the backward stage, where we can limit the set to its  $k$  most probable paths.

In genomics, we use those decodings to detect simulated homozygous-by-descent (HBD) DNA segments in inbred individuals, where HBD status at a locus is unobserved and might randomly differ from HBD status at a neighbouring locus on the DNA chain due to the recombination that occurred since the common ancestor.