

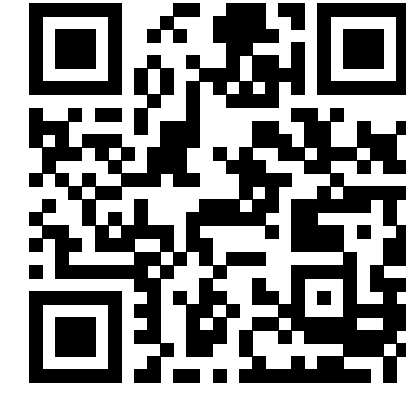
A STUDY OF FACTORS INFLUENCING THE PERFORMANCE OF THE RECONSTRUCTION OF TRANSMISSIONS IN DISEASE OUTBREAKS

Maryam Alamil¹, C. Bruchou¹, M. Ribaud¹, G. Thébaud², S. Soubeyrand¹

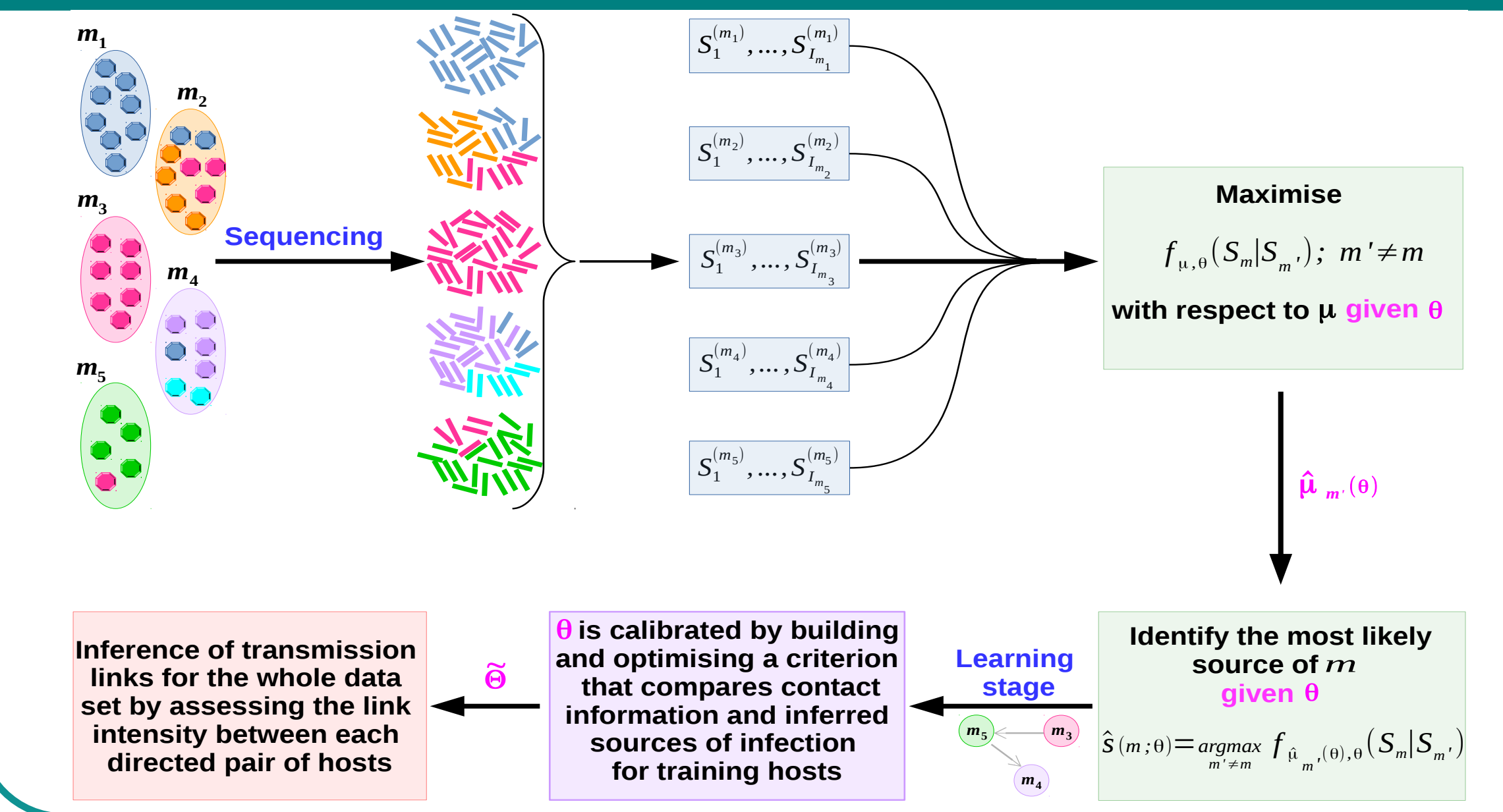
¹INRAE, BioSP, 84914 Avignon, France: maryam.alamil@inrae.fr; ²BGPI, INRAE, SupAgro, Cirad, Univ. Montpellier, Montpellier, France

Introduction

- To infer epidemiological links of infectious fast-evolving pathogens from deep sequencing data, we investigated a statistical learning approach called SLAFEEL⁽¹⁾.
- SLAFEEL is based on a pseudo-evolutionary model and grounded on statistical learning.
- To assess SLAFEEL performance, we applied it to simulated data, performing a sensitivity analysis of the relationship between SLAFEEL accuracy and the demo-genetic factors that may impact the virus evolution and transmission.
- Here, we present first a simulation stochastic model for the evolution and transmission of population of sequences, developed to generate data sets under multiple demographic and genetic settings.
- Then, we describe the procedure of the sensitivity analysis that we followed.
- Finally, we evaluate the impact of all factors on SLAFEEL accuracy using Sobol indices and we show how the most significant factors affect the genetic diversity and therefore the performance of our approach.

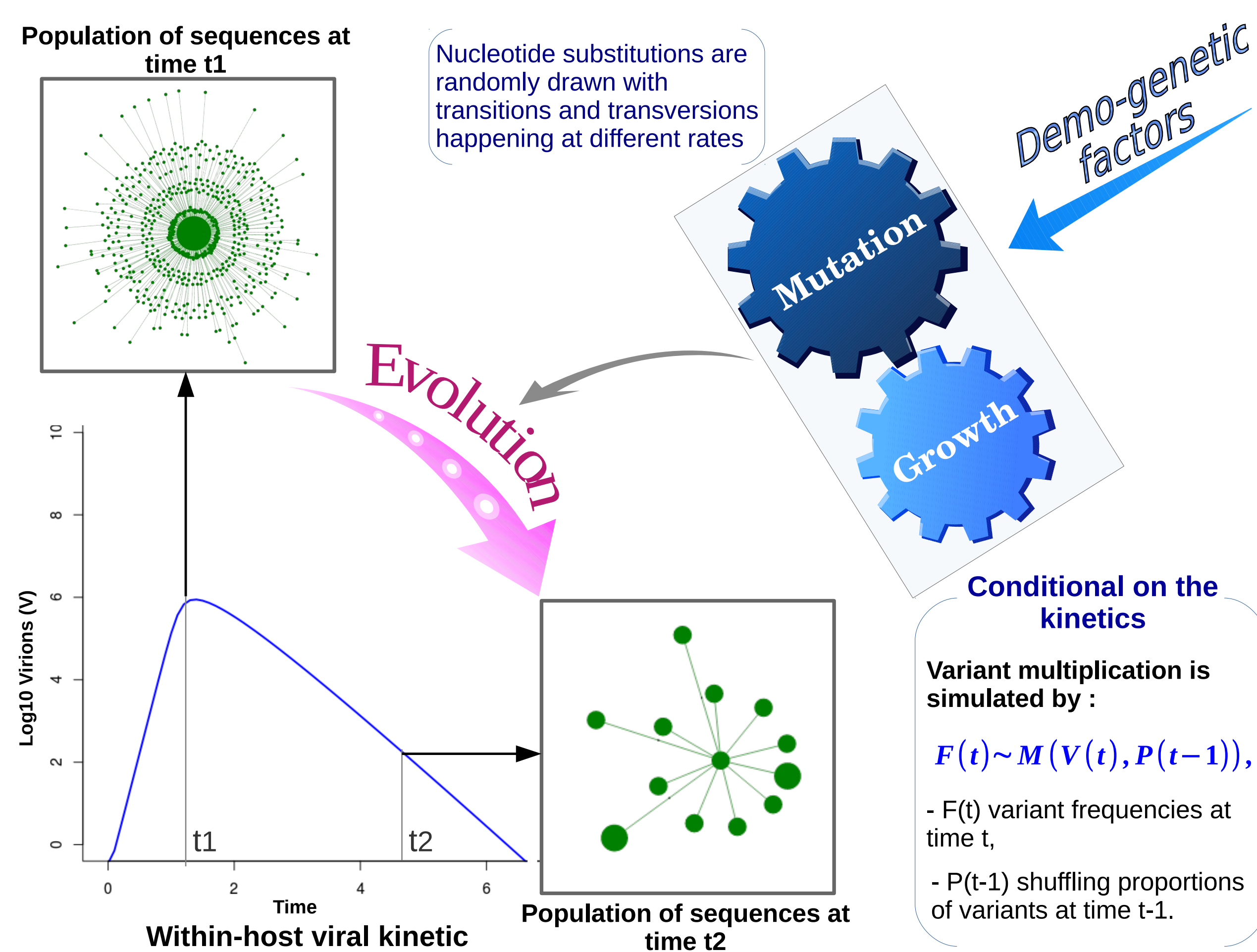


Graphical representation of SLAFEEL

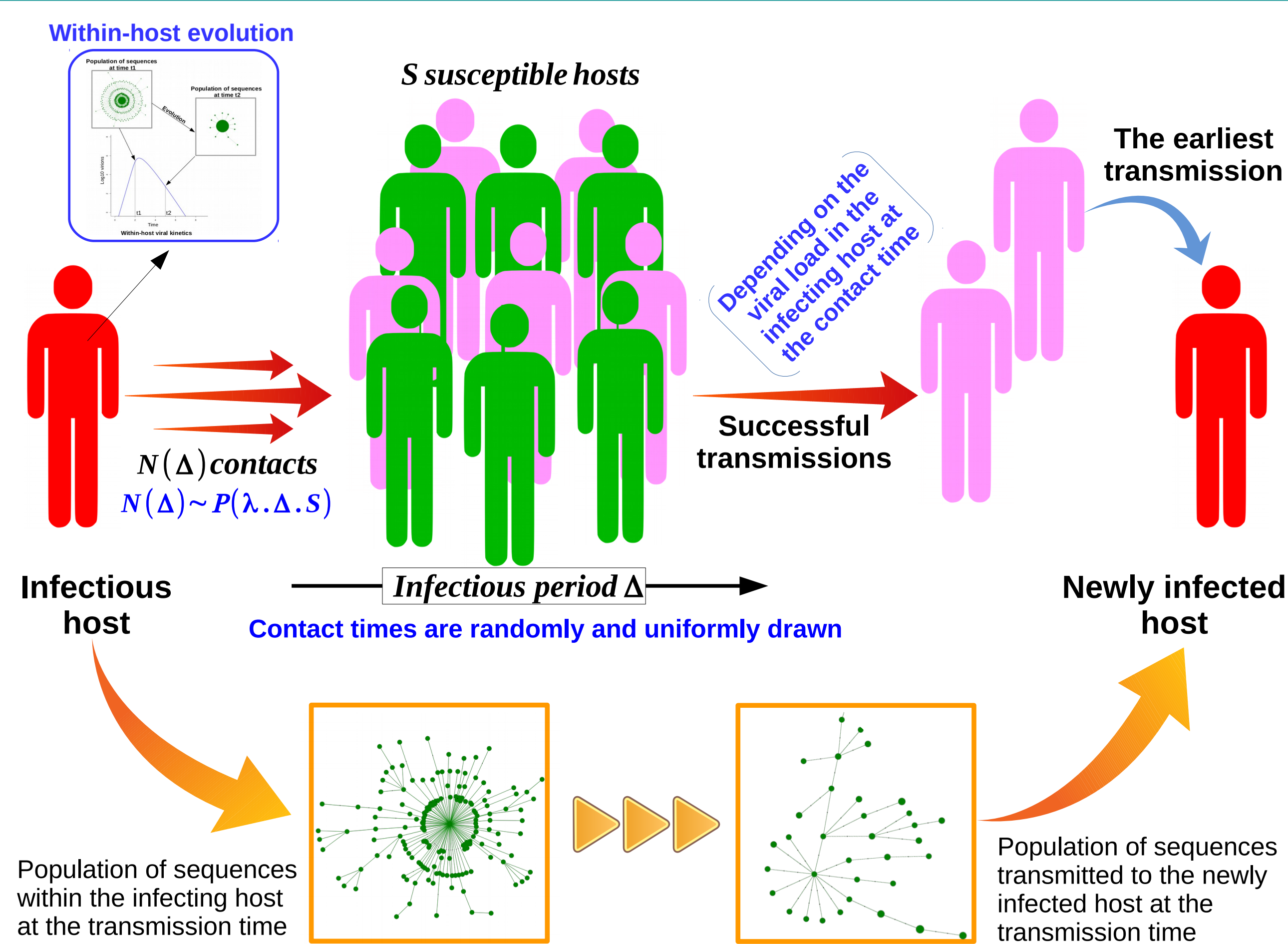


Evolution & transmission models

Simulation model for the within-host viral evolution



Simulation model for the inter-host viral transmission



Conclusion, perspectives & reference

Sensitivity analysis of the relationship between SLAFEEL accuracy and some demo-genetic factors shows that:

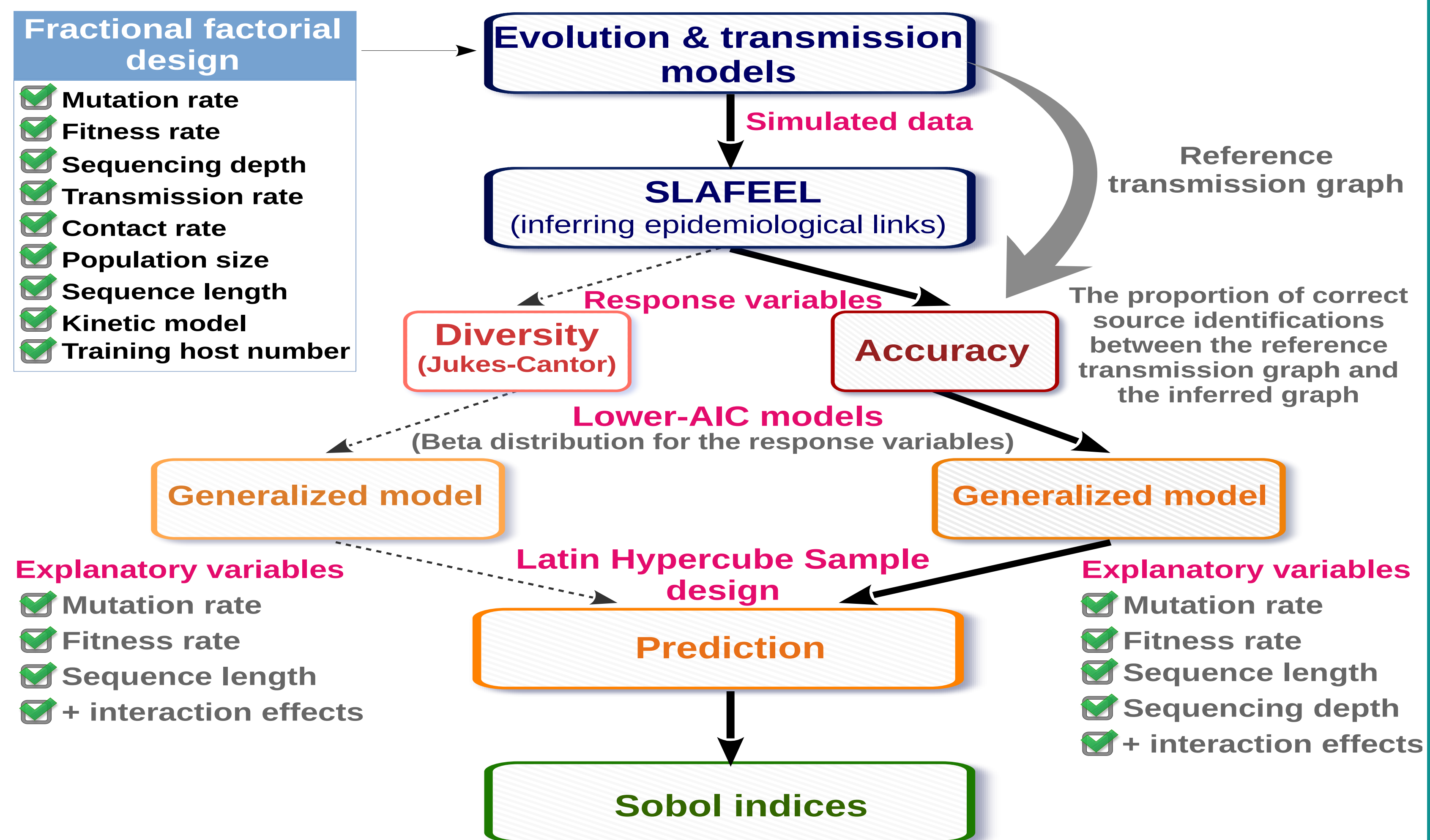
- the factors that have the greatest effect on SLAFEEL accuracy are the most important factors affecting the genetic diversity measured by Jukes-Cantor index.
- the lower the genetic diversity level, the better the performance of SLAFEEL.

To investigate deeply the relationship between SLAFEEL accuracy and the genetic diversity and the other factors:

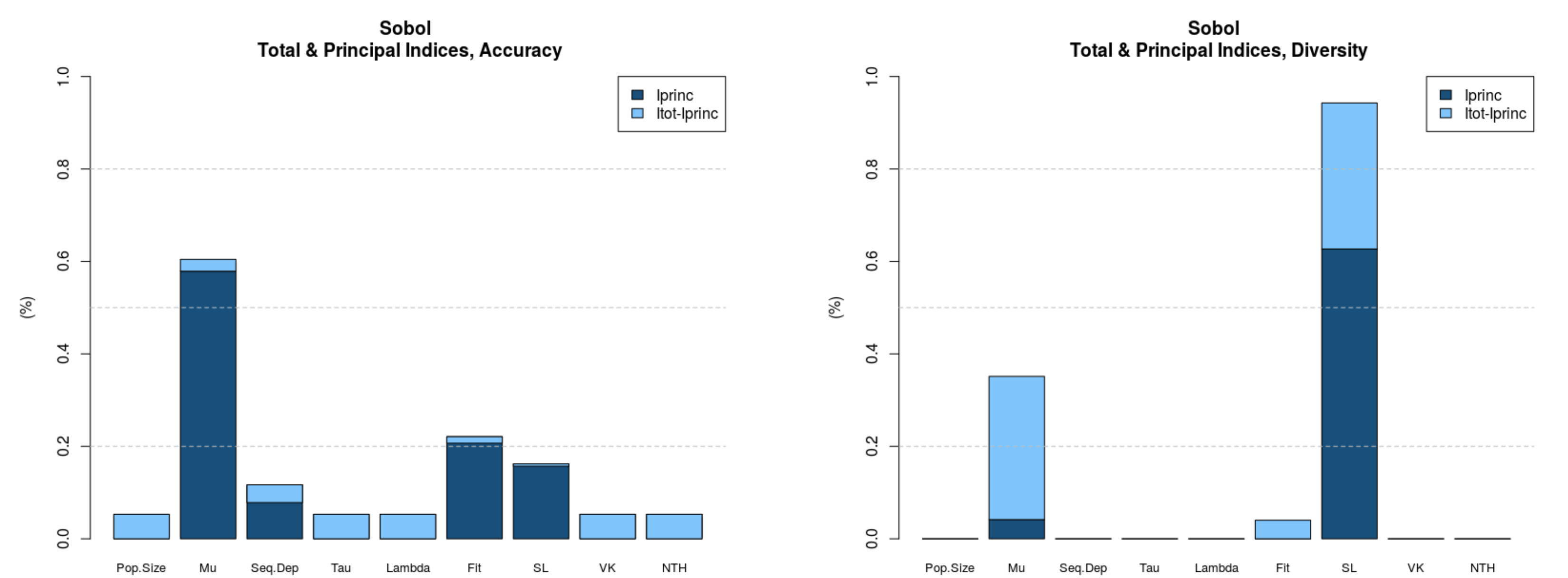
- we will exploit diversity data measured by other indices (e.g Shannon index, richness, ...)
- we will model differently the relation between SLAFEEL accuracy and the explanatory variables (e.g diverse probability distributions for the response variable, variable transformation, ...)

[1] M Alamil, Soubeyrand, and al. Inferring epidemiological links from deep sequencing data: a statistical learning approach for human, animal and plant diseases. *Philosophical Transactions of the Royal Society B*, 374(1775):20180258, 2019.

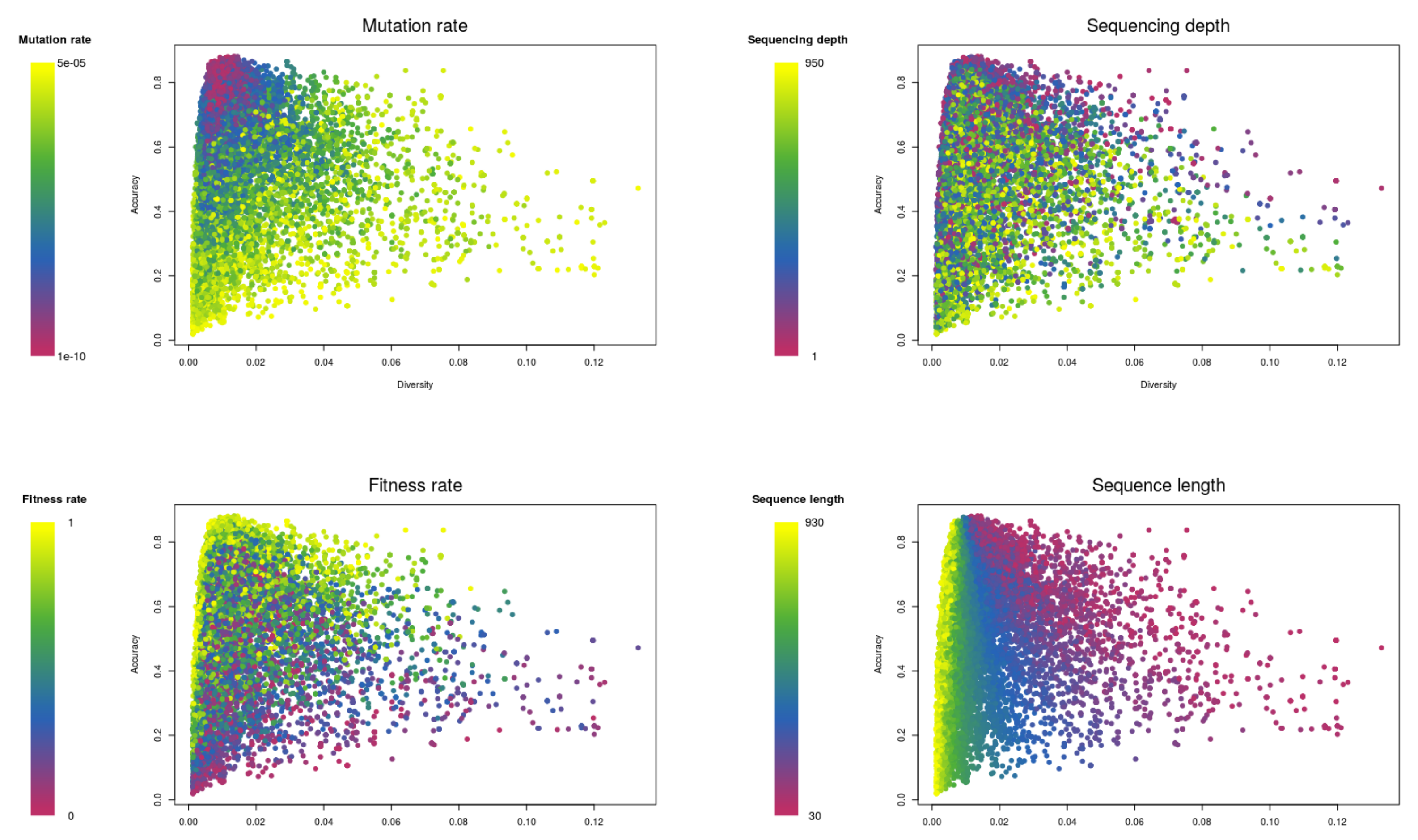
Sensitivity analysis



Sobol indices



Relationship between SLAFEEL performance (accuracy), genetic diversity & significant factors



Scatter plots of predicted SLAFEEL accuracy against the predicted genetic diversity with four factors: mutation rate, sequencing depth, fitness rate and sequence length. Each gradient colors represents a range of factor values. Purple and yellow colors indicate respectively the lower and higher values.