



**HAL**  
open science

# SelNeTime: a new method inferring demography and selection from genomic time series data

Simon Boitard, Mathieu Uhl, Miguel de Navascués, Bertrand Servin

## ► To cite this version:

Simon Boitard, Mathieu Uhl, Miguel de Navascués, Bertrand Servin. SelNeTime: a new method inferring demography and selection from genomic time series data. *Mathematical and Computational Evolutionary Biology*, Jun 2024, Hameau de l'étoile, France. hal-04659093

**HAL Id: hal-04659093**

**<https://hal.inrae.fr/hal-04659093>**

Submitted on 22 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# SelNeTime : a new method inferring demography and selection from genomic time series data

Simon Boitard<sup>1</sup>, Mathieu Uhl<sup>1,2</sup>, Miguel de Navascués<sup>1</sup>,  
Bertrand Servin<sup>3</sup>

<sup>1</sup> INRAE, CBGP, Montpellier, France

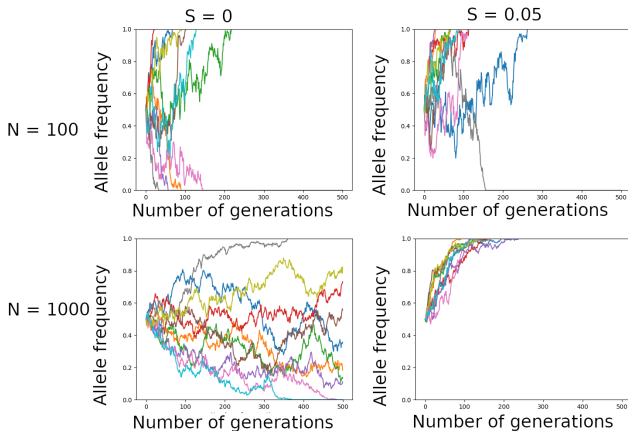
<sup>2</sup> CNRS, CEFE, Montpellier, France

<sup>3</sup> INRAE, GenPhySE, Toulouse, France

MCEB 2024

# Genomic time series

- Demographic and selective effects often difficult to disentangle from genomic data sampled at a single time.
- Temporal trajectories of allele frequencies more informative.



Various contexts and temporal scales:



Experimental  
evolution

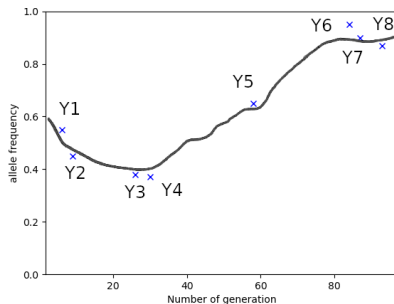
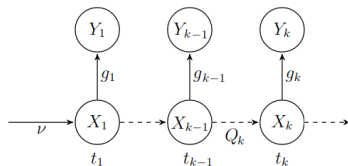


Monitoring of  
wild populations



Ancient  
DNA

# Hidden Markov Model (HMM) (Bollback 2008)



- $X_k$  population allele frequency at time  $t_k$  (hidden)
- $Y_k$  sampled allele frequency at time  $t_k$  (observed)
- $Q_k$  transition matrix from time  $t_{k-1}$  to time  $t_k$

# Transition matrix

- Computed under a Wright-Fisher model.
- Depends on  $N$ ,  $s$  and  $t_{k-1} - t_k$ .
- Example for  $N = 4$ :

$$Q = \begin{array}{c} \\ 0/4 \\ 1/4 \\ 2/4 \\ 3/4 \\ 4/4 \end{array} \begin{array}{ccccc} & 0/4 & 1/4 & 2/4 & 3/4 & 4/4 \\ \left( \begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ 0.32 & 0.42 & 0.21 & 0.05 & 0.04 \\ 0.06 & 0.25 & 0.38 & 0.25 & 0.06 \\ 0.004 & 0.05 & 0.21 & 0.42 & 0.32 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right)$$

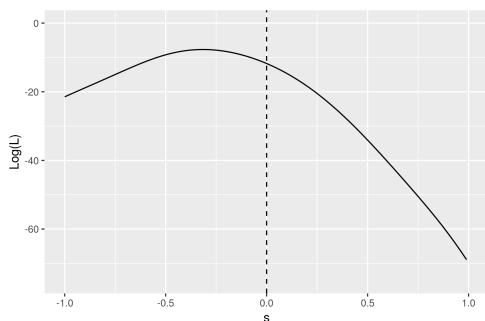
# Demography and Selection Inference

Exact (and fast) computation of the likelihood

$$P(\bar{Y}|N, s) = P(Y_1, \dots, Y_n|Q_1, \dots, Q_n)$$

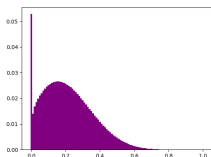
for any values of  $N$  and  $s$

- 1 Inference of  $N$**  : consider  $p$  independent loci and optimize  $P(\bar{Y}_1|N, s_1 = 0)P(\bar{Y}_2|N, s_2 = 0) \dots P(\bar{Y}_p|N, s_p = 0)$  over  $N$ .
- 2 Inference of  $s$**  : for each locus  $i$ , optimize  $P(\bar{Y}_i|\hat{N}, s_i)$  over  $s_i$ .

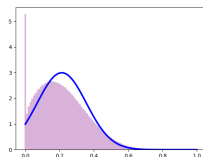


# Wright-Fisher approximations

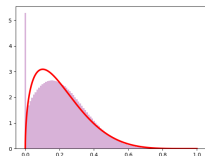
- Wright-Fisher model limited to  $N \approx 500$  for numerical reasons ( $Q$  of size  $N \times N$ ).
- Continuous approximations



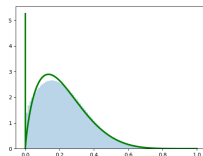
Wright-Fisher



Gaussian



Beta

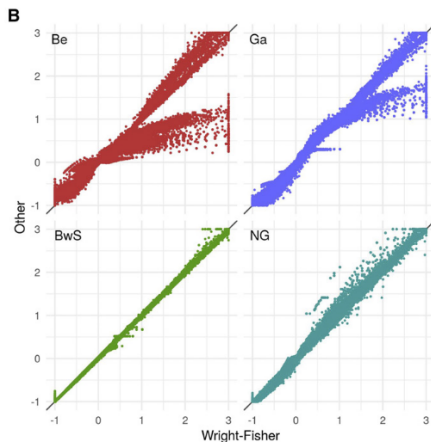


Beta with Spikes



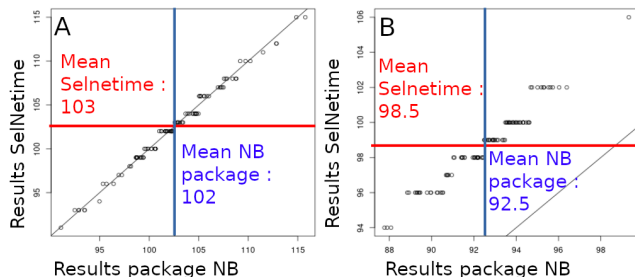
# Wright-Fisher approximations

The Beta with Spikes distribution (Tataru *et al* 2019) is a very good approximation (Paris *et al*, 2019).

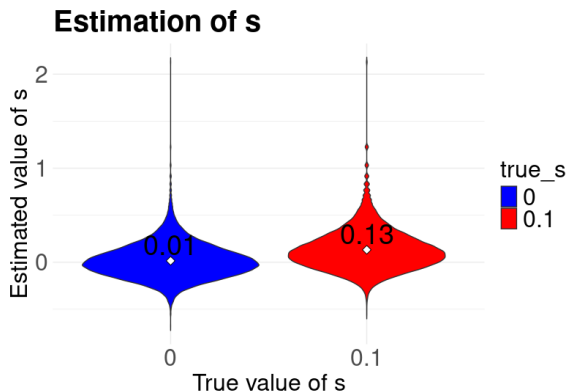


- **Models** Beta-with-Spikes and Wright-Fisher transitions.
- **Infers**  $N$  assuming  $s = 0$  and / or  $s$  given  $N$ .
- **Computation time** For  $t = 1 \dots 10$  and 1000 loci:  
8.3s to infer  $N$ , 26s to infer all  $s$   
(using one core)
- **Availability** <https://pypi.org/project/selnetime/>

# Estimation of $N$



- $t = 1 \dots 10$  (A) or  $t = 1, 10, \dots, 100$  (B),  $s = 0, 1000$  loci.
- Better estimation with the BwS than with the Beta model (Hui and Burt 2015) for large  $\delta_t$ .



- $t = 1 \dots 10$ ,  $N = 100$ , BwS model.
- Unbiased estimation of  $s$ , as in Paris *et al* (2019).

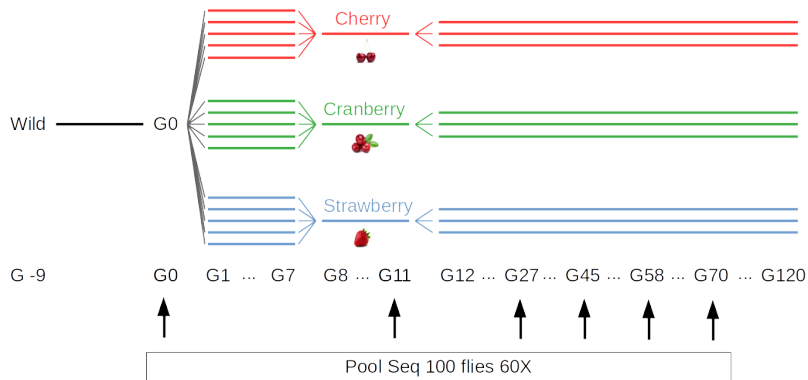
# *Drosophila suzukii*



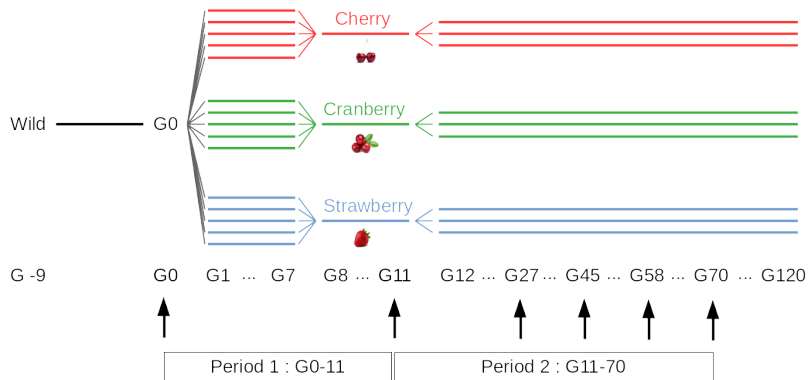
From Aspen et al. 2015 J. Pest. Sci.



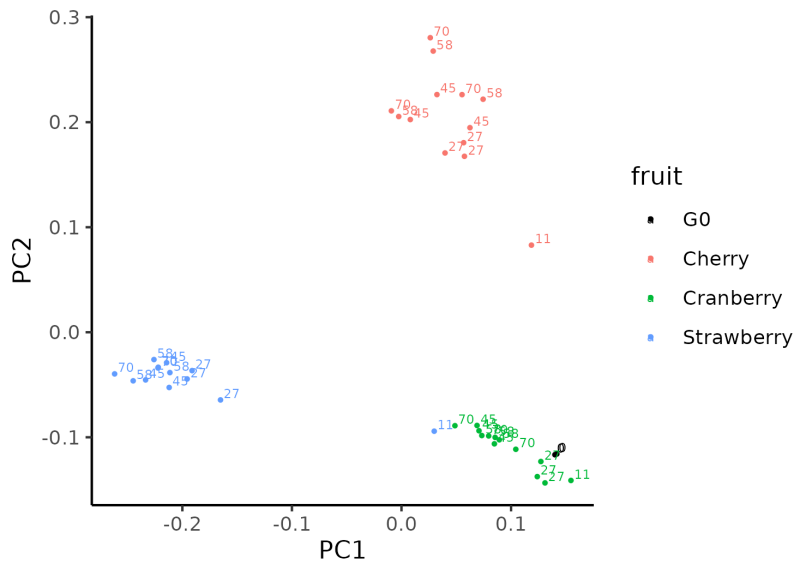
# Evolve & Resequence experiment



# Evolve & Resequense experiment

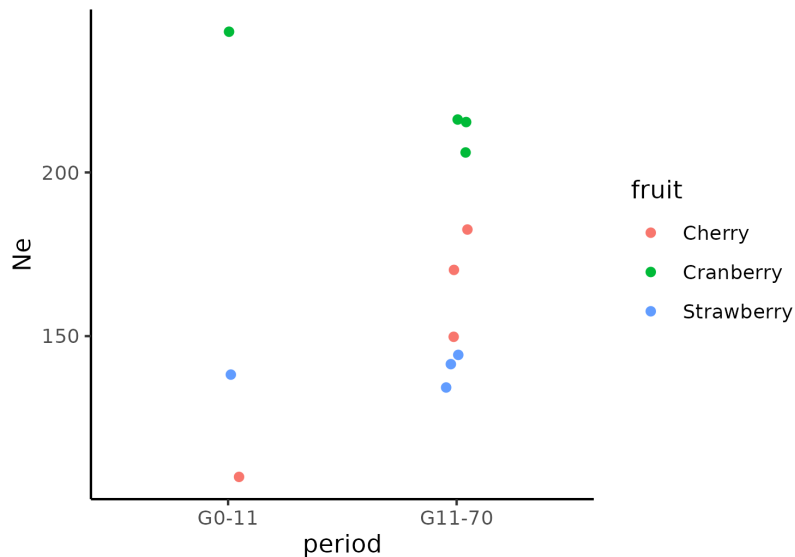


# Genetic diversity structuring

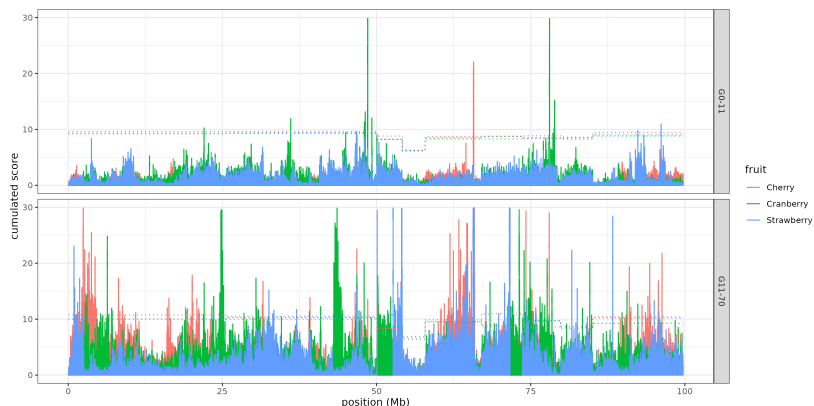




# Inferred $N$



# Candidate regions under selection ( $\hat{s} \neq 0$ )



p-values obtained from the HMM and 'cumulated' using a local score approach (Fariello *et al*, 2017).

- Linked selection **pervasive** (Elyashiv *et al*, 2016).
- **Biases demographic inference** (Schridder *et al*, 2016; Pouyet *et al*, 2018; Johri *et al* 2021, Boitard *et al* 2022).

→ **Joint estimation** of demography and selection from genomic time series.

- PhD / postdoc position available.

## **Fly experiment:**

- Candice Deschamps, Mathieu Gautier, Laure Olazcuagua & Nicolas Rode, CBGP

## **DNA preparation:**

- Anne Loiseau, CBGP

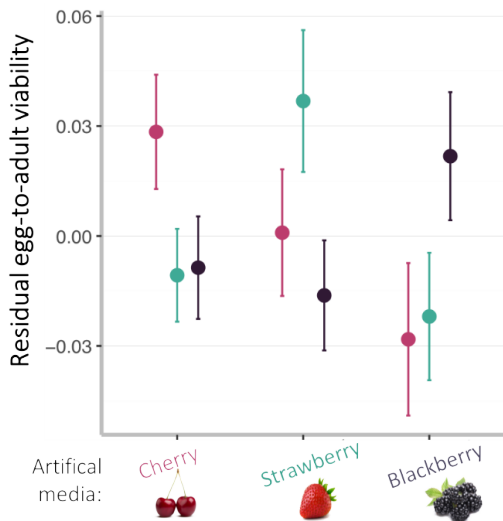
## **Read mapping and variant calling:**

- Mathieu Gautier

## **Time series methodology:**

- Miguel de Navasués & Mathieu Uhl, CBGP
- Cyriel Paris & Bertrand Servin, INRAE, GenPhySE, Toulouse, France

# Local adaptation to host plant (Olazcuagua *et al*, 2022)



Local adaptation test:  
 $P = 0.005$

Fly population from:

- Cherry 
- Strawberry 
- Blackberry 

# Inferred $N$

