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-04659093v1

Submitted on 22 Jul 2024

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

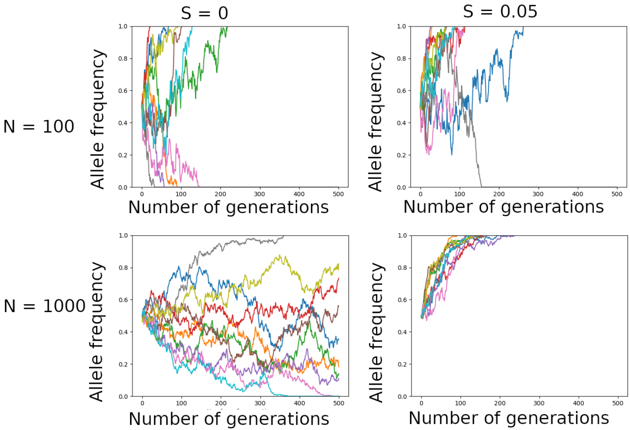
Simon Boitard[1], Mathieu Uhl[1,2], Miguel de Navascués[1],
Bertrand Servin[3]

[1] INRAE, CBGP, Montpellier, France
[2] CNRS, CEFE, Montpellier, France
[3] 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.



Genomic time series

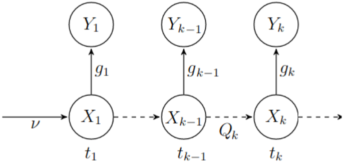
Various contexts and temporal scales:

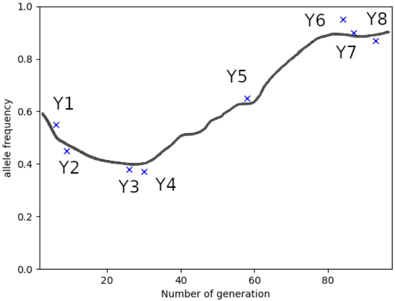
Experimental
evolution

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 freequency 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 $N = 4$:

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

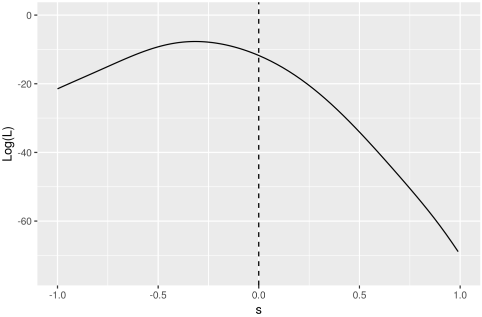
Demography and Selection Inference

Exact (and fast) computation of the likelihood

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

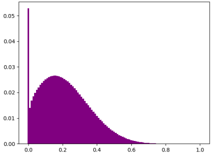
for any values of N and s

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

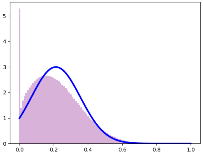


Wright-Fisher approximations

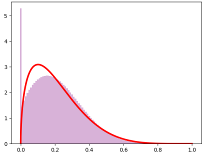
- Wright-Fisher model limited to $N \approx 500$ for numerical reasons (Q of size $N \times N$).
- Continuous approximations



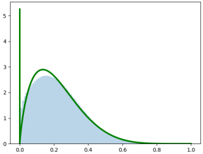
Wright-Fisher



Gaussian



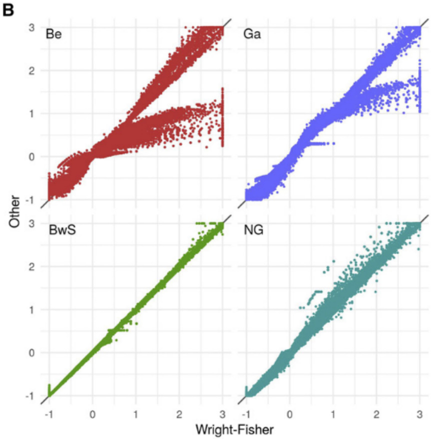
Beta



Beta with Spikes

Wright-Fisher approximations

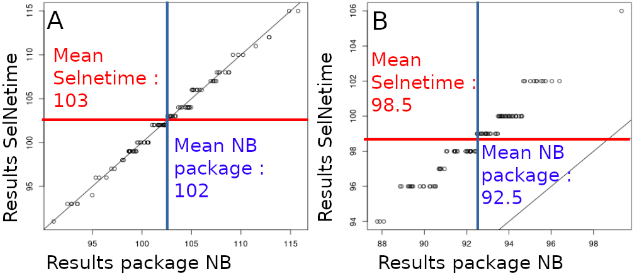
The Beta with Spikes distribution (Tataru *et al* 2019) is a very good approximation (Paris *et al*, 2019).



python package selnetime

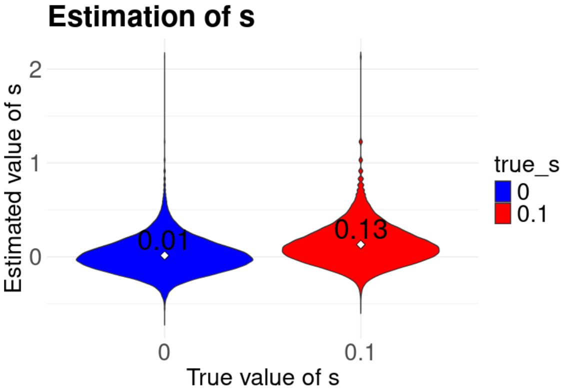
- **Models** Beta-with-Spikes and Wright-Fisher transitions.
- **Infers** N assuming $s = 0$ and / or s given N.
- **Computation time** For $t = 1 \dots 10$ and 1000 loci:
  8.3s to infer N, 26s to infer all s
  (using one core)
- **Availability** https://pypi.org/project/selnetime/

Estimation of N



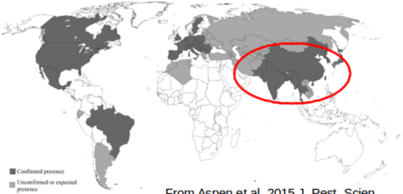
- $t = 1 \ldots 10$ (A) or $t = 1, 10, \ldots, 100$ (B), $s = 0$, 1000 loci.
- Better estimation with the BwS than with the Beta model (Hui and Burt 2015) for large δ_t.

Estimation of s



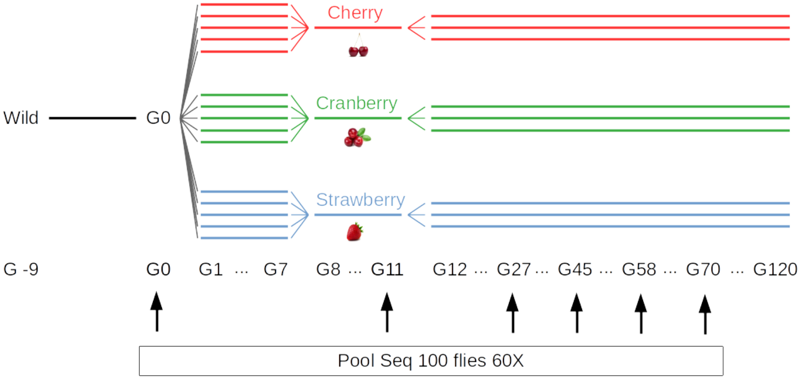
- $t = 1 \ldots 10$, $N = 100$, BwS model.
- Unbiased estimation of s, as in Paris *et al* (2019).

Drosophila suzukii

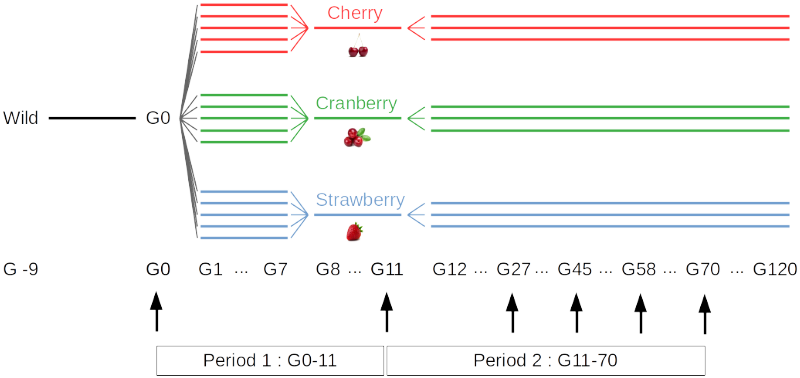


From Aspen et al. 2015 J. Pest. Scien.

Evolve & Resequence experiment



Evolve & Resequence experiment



Wild

G0

Cherry

Cranberry

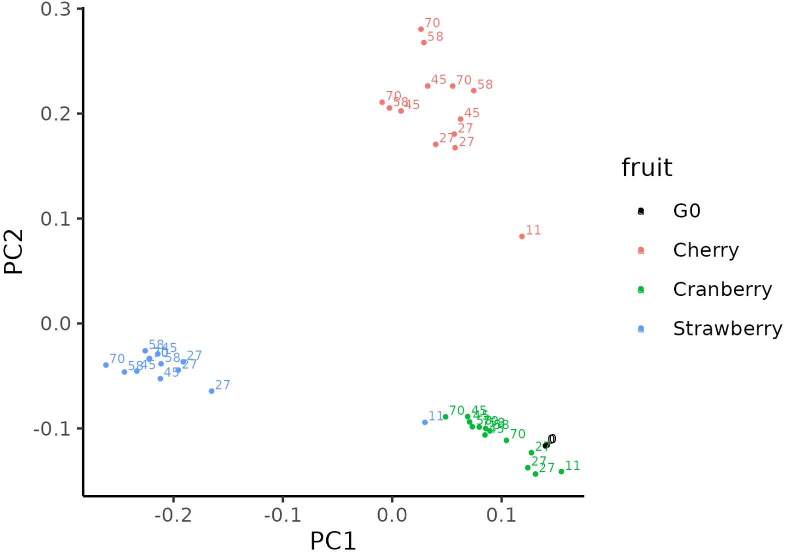
Strawberry

G -9 G0 G1 ... G7 G8 ... G11 G12 ... G27 ... G45 ... G58 ... G70 ... G120

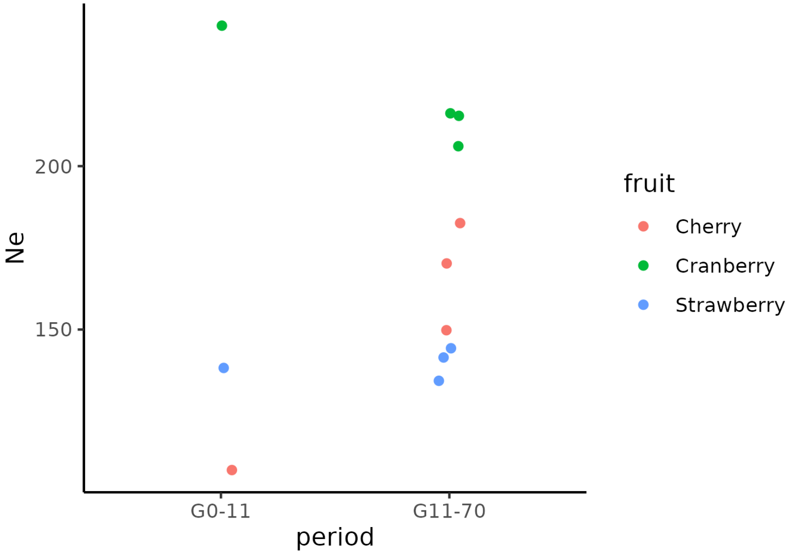
Period 1 : G0-11

Period 2 : G11-70

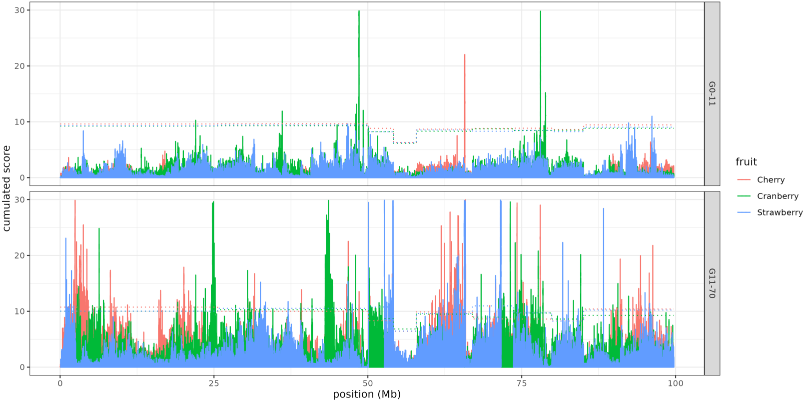
Genetic diversity structuring



Inferred *N*


Ne
200
150
G0-11
G11-70
period
fruit
Cherry
Cranberry
Strawberry

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).

Perspectives

- Linked selection **pervasive** (Elyashiv *et al*, 2016).
- **Biases demographic inference** (Schrider *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.

Acknowledgements

Fly experiment:

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

DNA preparation:

- Anne Loiseau, CBGP

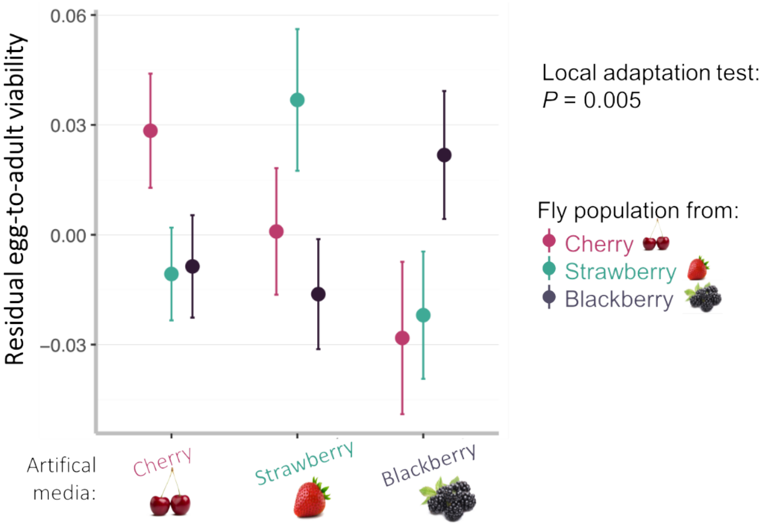
Read mapping and variant calling:

- Mathieu Gautier

Time series methodology:

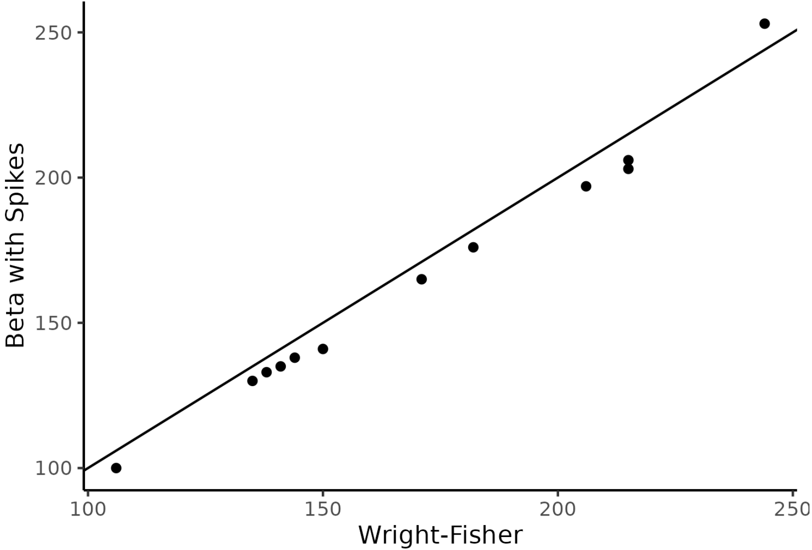
- Miguel de Navasués & Mathieu Uhl, CBGP
- Cyriel Paris & Bertrand Servin, INRAE, GenPhySE, Toulouse, France

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



Local adaptation test:
$P = 0.005$

Inferred N


250
200
150
100
Beta with Spikes
100
150
200
250
Wright-Fisher