

Impact of linked selection on demographic inference: insights from the Inverse Instantaneous Coalescence Rate (IICR)

Simon Boitard¹, Lounès Chikhi^{2,3}, Olivier Mazet⁴

1: INRAE, Centre de Biologie et de Gestion des Populations (CBGP), Montpellier, France - 2: CNRS, Evolution et Diversité Biologique (EDB), Toulouse, France - 3: Instituto Gulbenkian de Ciência (IGC), Oeiras, Portugal - 4: INSA, Institut de Mathématiques de Toulouse (IMT), Toulouse, France

Virtual conference *Probabilistic Modeling In Genomics*, April 14 - 16, 2021

CONTEXT

- **Linked selection is pervasive** (Elyashiv *et al*, 2016; Pouyet *et al*, 2018) and **biases demographic inference** (Ewing and Jensen, 2016; Schrider *et al*, 2016; Pouyet *et al*, 2018; Johri *et al* 2021).
- Can be modelled approximately by a local reduction (background selection & sweeps) or increase (balancing selection) of effective population size N_e (Hill and Robertson, 1966)
⇒ **Variable levels of N_e genome-wide** (Gossmann *et al*, 2011) reflecting the variations of recombination rate or gene density.
- **Study the genome-wide distribution of pairwise coalescence times (T_2) for models with variable genomic N_e to predict the impact of linked selection on PSMC (Li and Durbin, 2011).**

LINKED SELECTION UNDER PANMIXIA

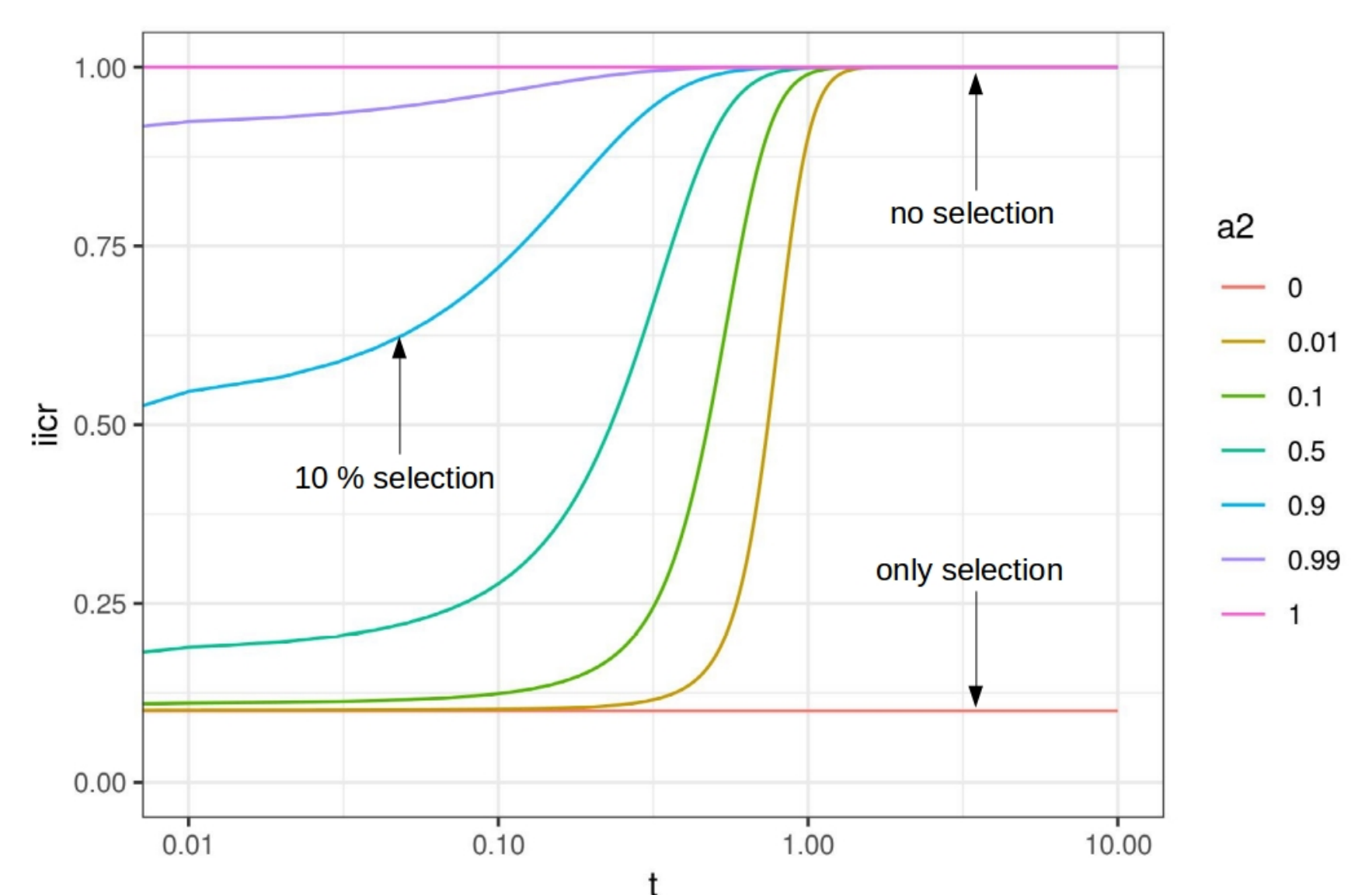
- K genomic classes with relative proportion a_i .
- Class i evolves under the WF model with $\lambda_i N$ diploids.

$$IICR(t) = \frac{\sum_{i=1}^K a_i \mathbb{P}(T_2^i \geq t)}{\sum_{i=1}^K a_i d\mathbb{P}(T_2^i = t)} = \frac{\sum_{i=1}^K a_i e^{-\mu_i t}}{\sum_{i=1}^K a_i \mu_i e^{-\mu_i t}}, \quad \mu_i = \frac{1}{\lambda_i}$$

General results

Example: $K = 2$, $\lambda_1 = 0.1$, $\lambda_2 = 1$ (neutral)

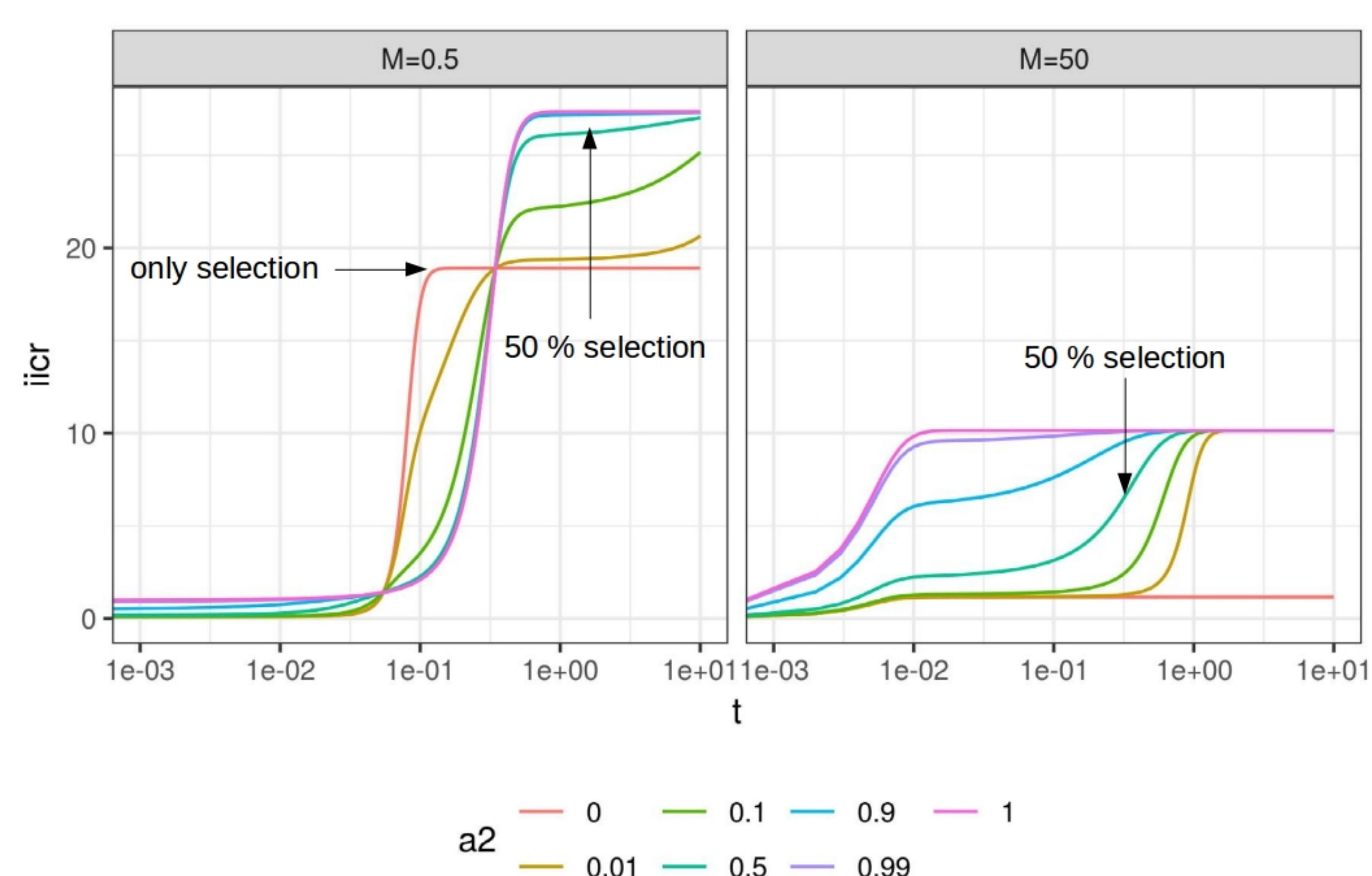
- $IICR'(t) > 0$
spurious signal of N_e decline
- $IICR(0) = \frac{1}{\sum_{i=1}^K \frac{a_i}{\lambda_i}}$
- $IICR(t) \rightarrow \lambda_{max}$ as $t \rightarrow +\infty$



LINKED SELECTION AND STRUCTURE

Class i evolves under an **island model** with n demes, scaled migration rate $M = 4Nm$ (not affected by selection) and **deme size** $\lambda_i N$.

Example for $n = 10$, $K = 2$, $\lambda_1 = 0.1$, $\lambda_2 = 1$:



- Close to panmixia (see poster left column) for large M .
- IICR plateau \uparrow as $M \downarrow$ (general property of structured models).
- **Selection effect weaker than under panmixia** for $M \leq 1$.
- Counter-intuitive effects: IICR sometimes \uparrow as selection proportion \uparrow .

THE IICR (MAZET *et al*, 2016)

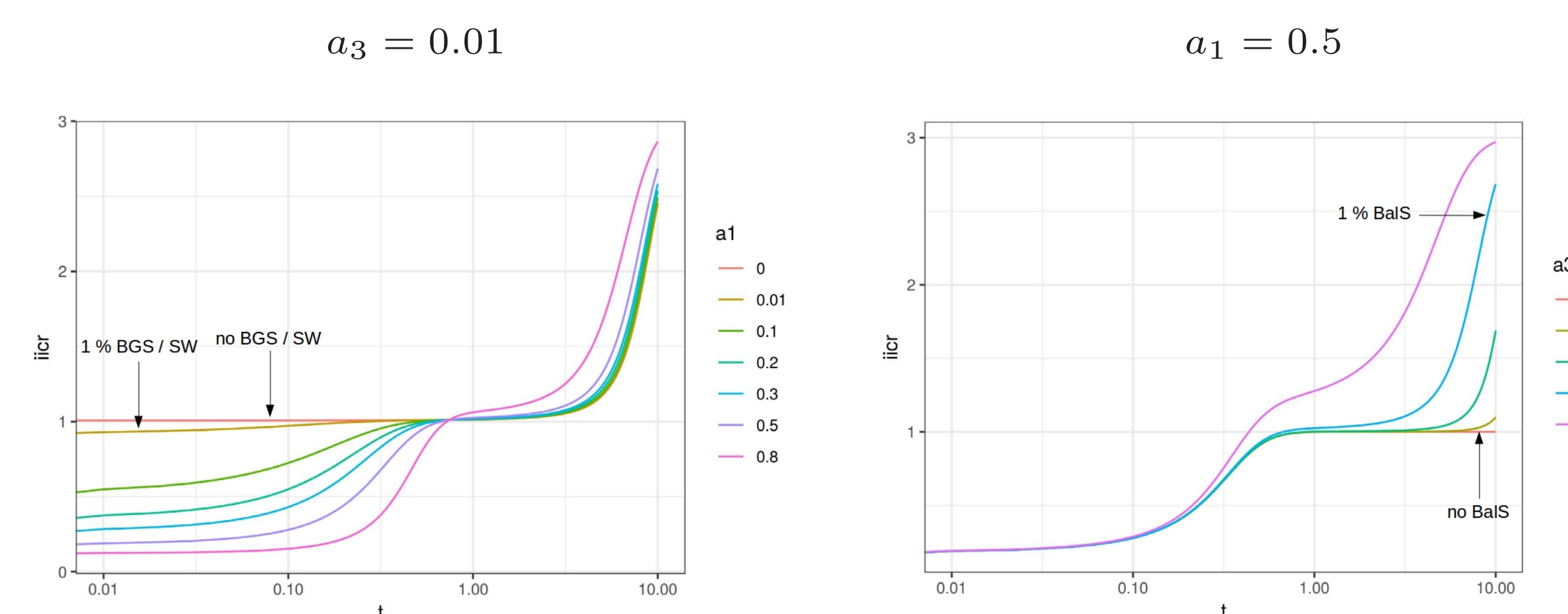
- For a given evolution model, the IICR is a function $\lambda()$ defined by

$$\frac{1}{\lambda(t)} = \frac{d\mathbb{P}(T_2 = t)}{\mathbb{P}(T_2 \geq t)}, \quad t \geq 0 \text{ in } 2N_e \text{ units}$$

- **The IICR $\lambda()$ is the quantity estimated by PSMC.**
- It corresponds to the **temporal trajectory of N_e if and only if the population considered has always evolved under panmixia.**

COMBINING VARIOUS FORMS OF SELECTION

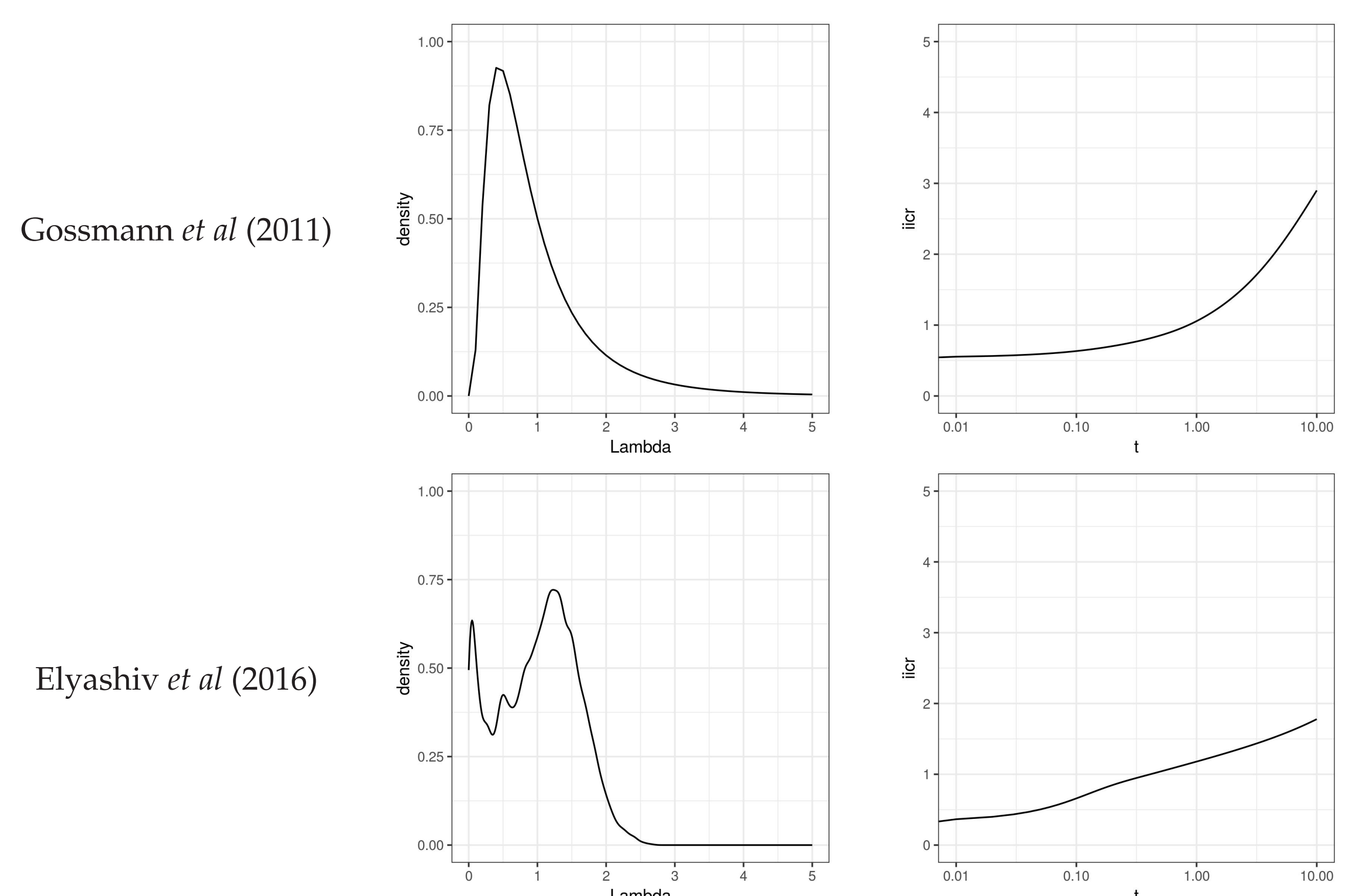
$K = 3$: $\lambda_1 = 0.1$ (BGS & sweeps), $\lambda_2 = 1$ (neutral), $\lambda_3 = 3$ (balancing).



- BGS & sweeps (resp. balancing sel.) affect recent (resp. ancient) IICR
- **Stronger effect of balancing selection for the same proportion.**

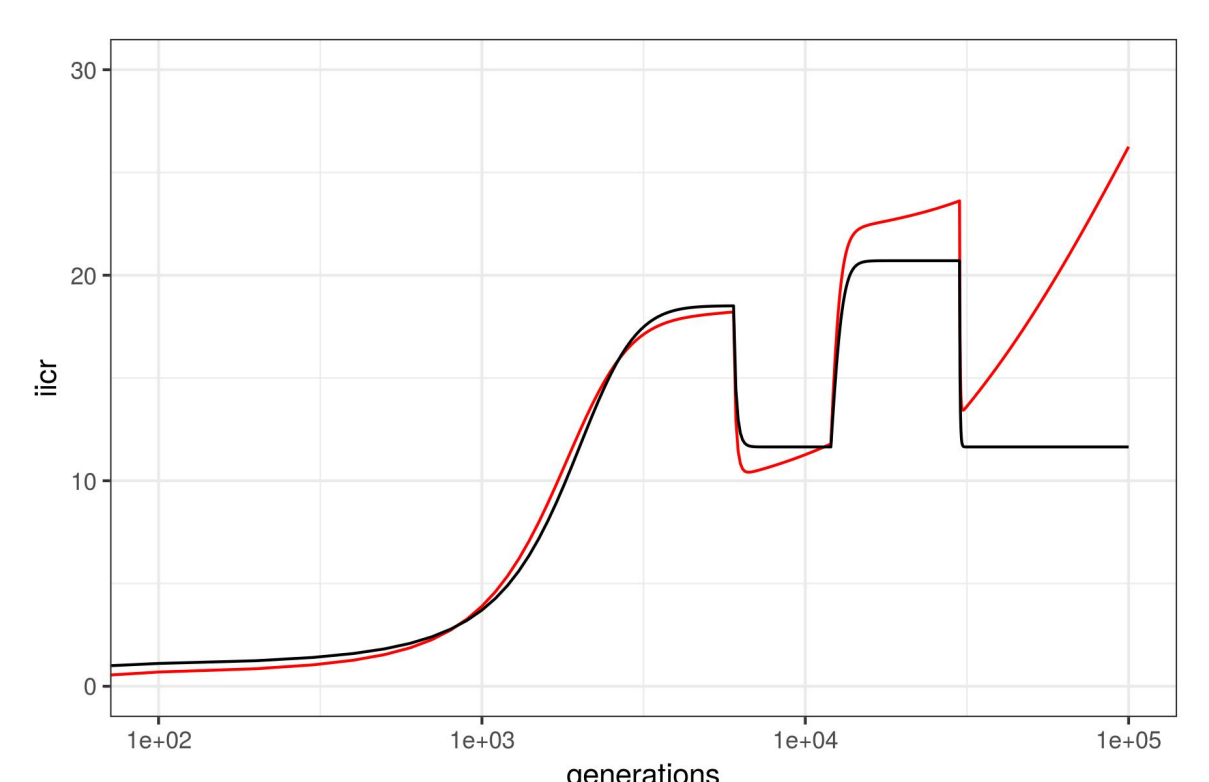
REALISTIC N_e DISTRIBUTIONS

Estimated for *Drosophila melanogaster* from polymorphism and divergence data, discretized into $K = 25$ classes.



LINKED SELECTION, HUMAN STRUCTURE & IICR

- **Demographic model (black):** Island model with **no population size change** but 4 changes of M over time **mimicking the human PSMC** (Mazet *et al* 2016).
- **With selection (red):** $\lambda()$ distribution from Gossmann *et al* (2011).



- **Limited effect of selection except in ancient past.**

FUNDING