Genetic Inference of Recent Fragmentation

Rémi Tournebize, Simon Boitard, Olivier Mazet, Lounès Chikhi

To cite this version:

Rémi Tournebize, Simon Boitard, Olivier Mazet, Lounès Chikhi. Genetic Inference of Recent Fragmentation. Conservation Genomics Paris, Jun 2024, Paris, France. hal-04659072

HAL Id: hal-04659072

https://hal.inrae.fr/hal-04659072v1

Submitted on 22 Jul 2024

genetic inference of recent fragmentation


MidJourney

Rémi Tournebize
Simon Boitard
Olivier Mazet
Lounès Chikhi

05 VI 2024

wild
populations
are
often
structured

MidJourney

barriers to
gene flow
can **reduce**
or **disrupt**
their
connectivity

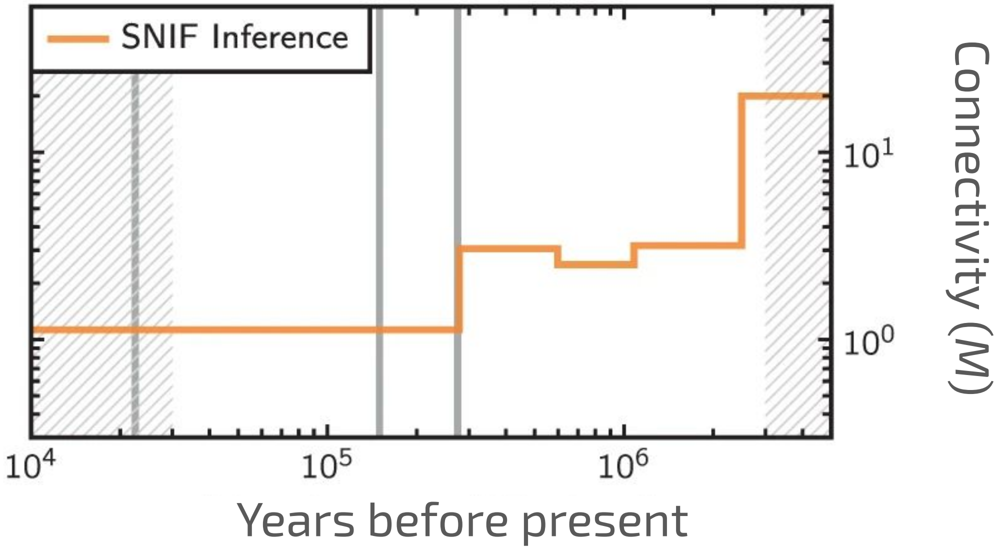
MidJourney

in the long run, this can lead to increased **divergence**

MidJourney

Most current genetic methods infer only **ancient** changes in connectivity

Using a French human genome



Arredondo et al. (2021)
Steux et al. (2024) (*in prep.*)

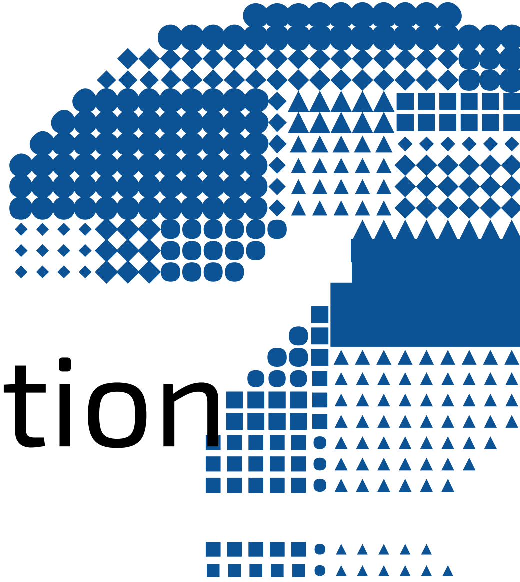
but **human activities** have created **very recent** *possible* barriers to gene flow (< 10 generations ago)

how well DNA

can help us

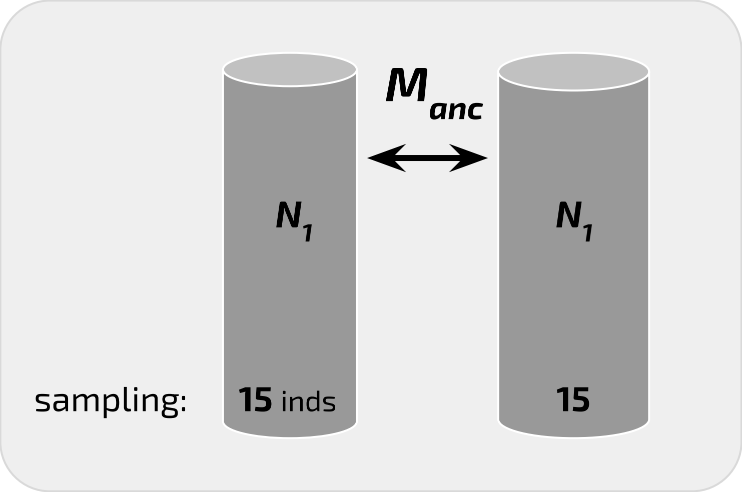
detect

very
recent
fragmentation

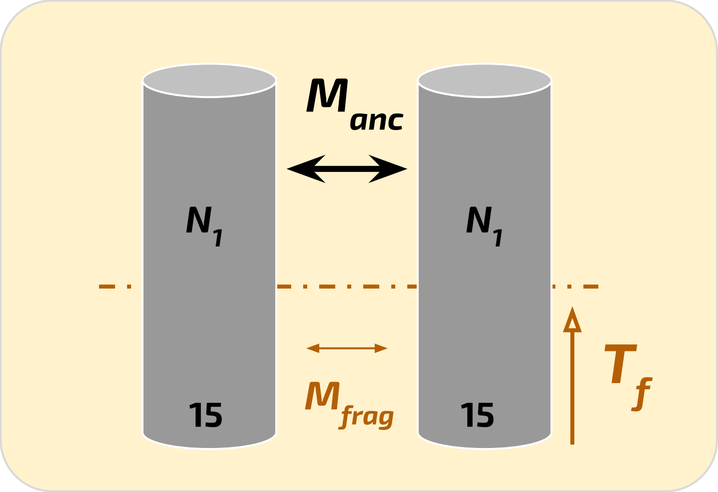


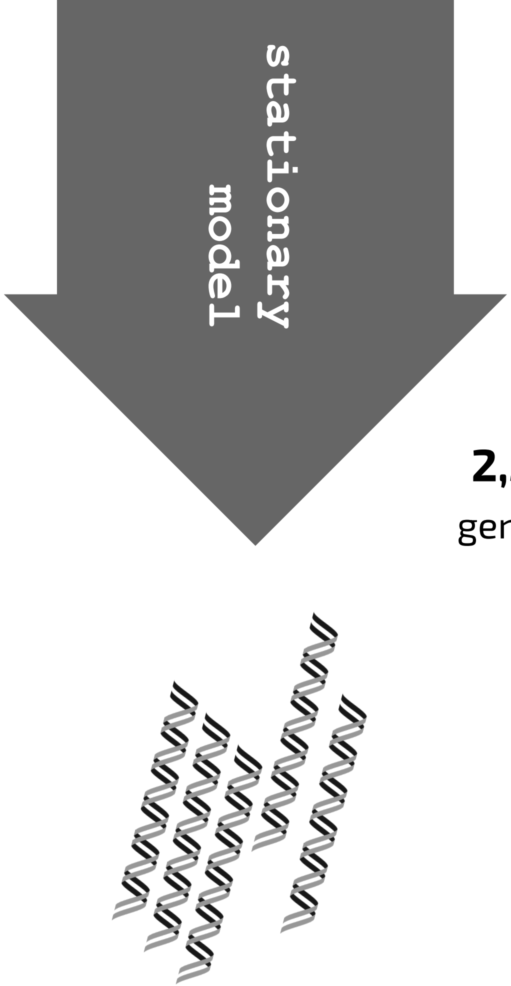
A study using **simulations**

stationary
model

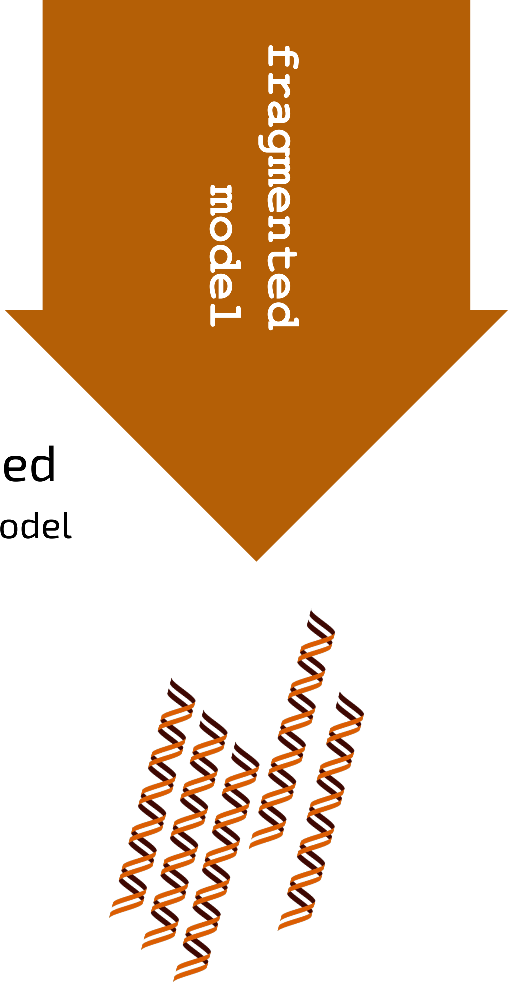


fragmentation
model





`stationary model`



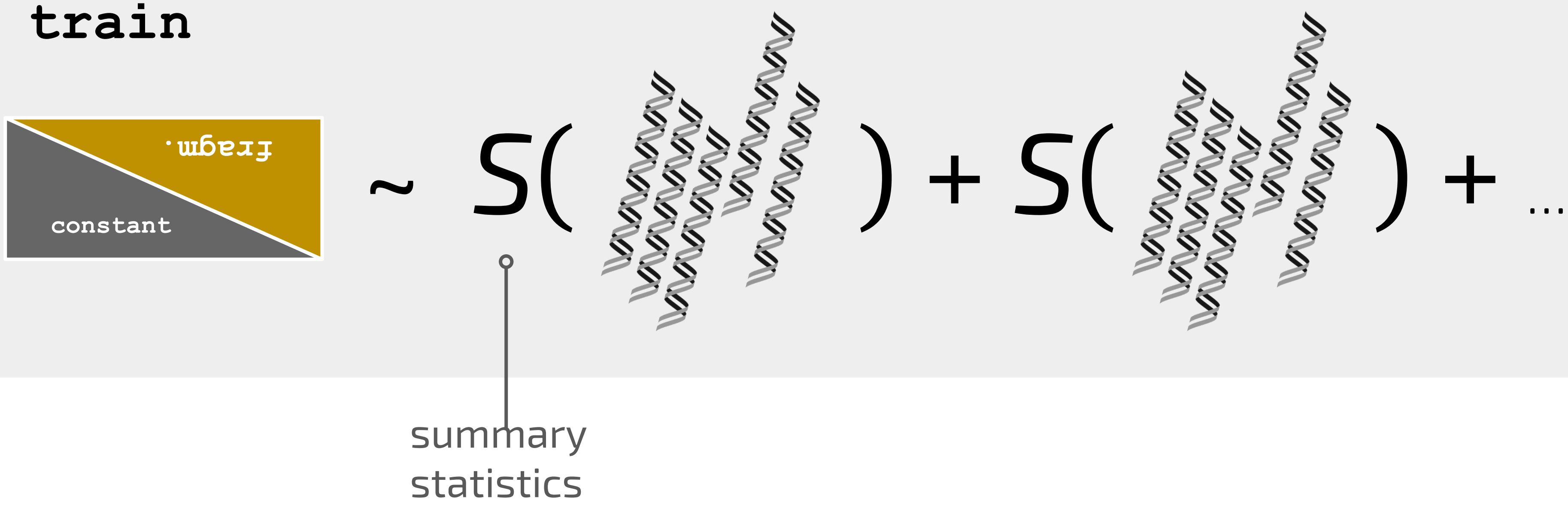
`fragmented model`

2,500 simulated

genetic data per model

G=5×25 Mbp

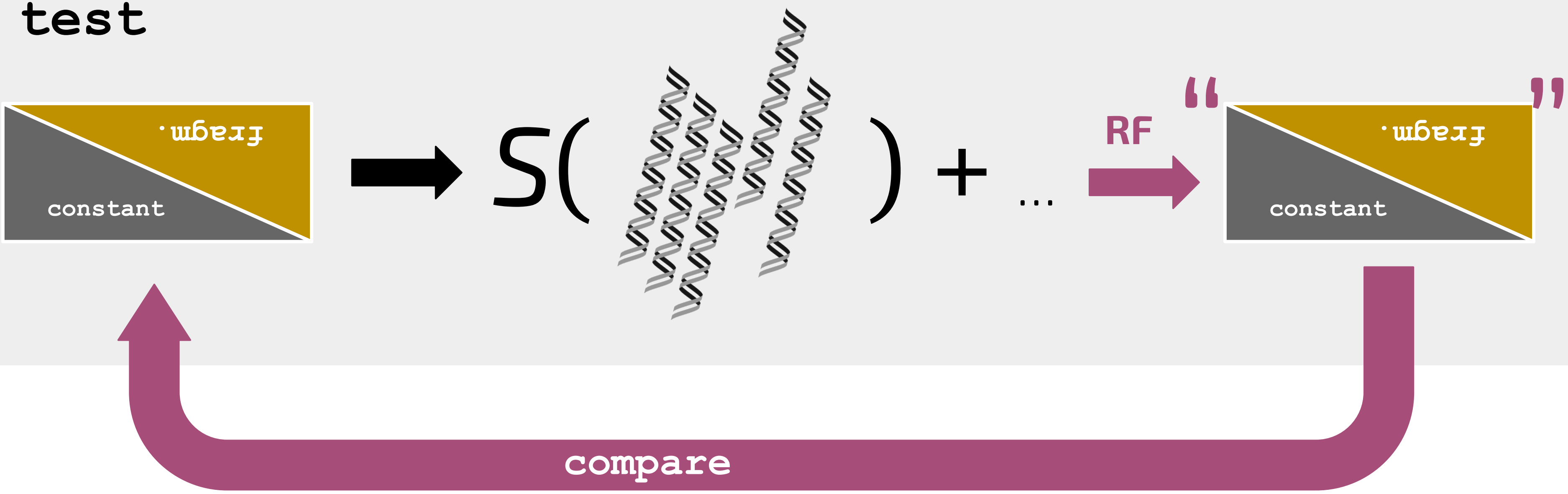
Test the performance to infer underlying models using **ABC** with **random forests**



`train`

summary statistics

Test the performance to infer underlying models using **ABC** with **random forests**



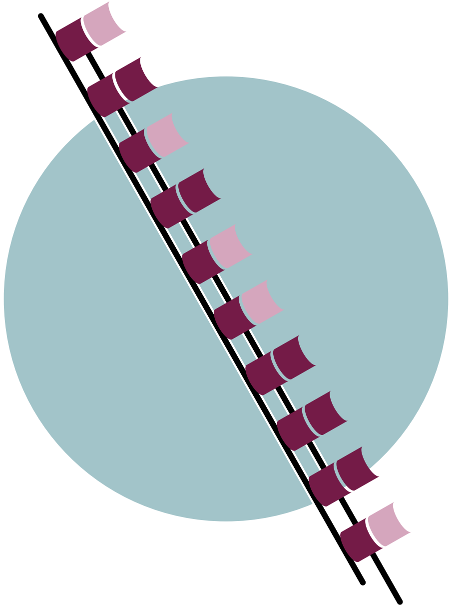
using **different**

types of

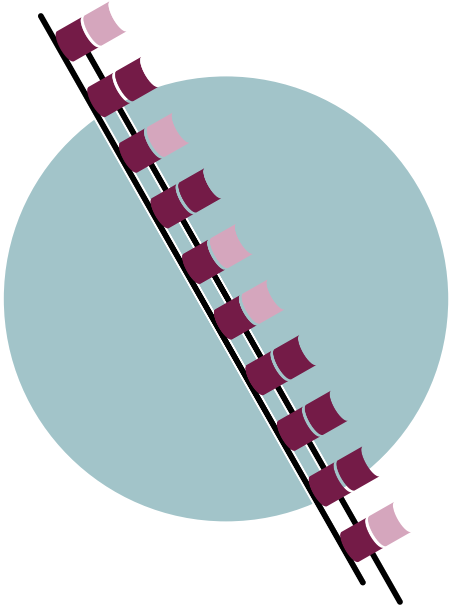
genomic data

types of genomic **data**

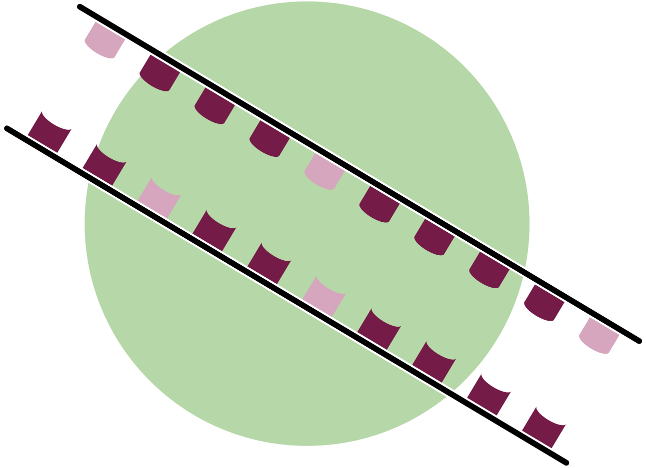
`unphased`



unphased

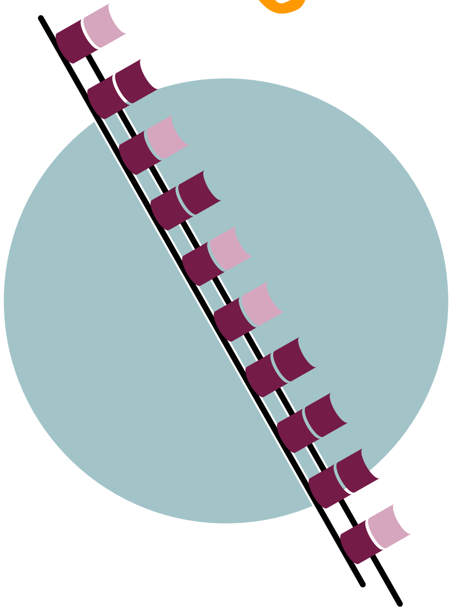


phased



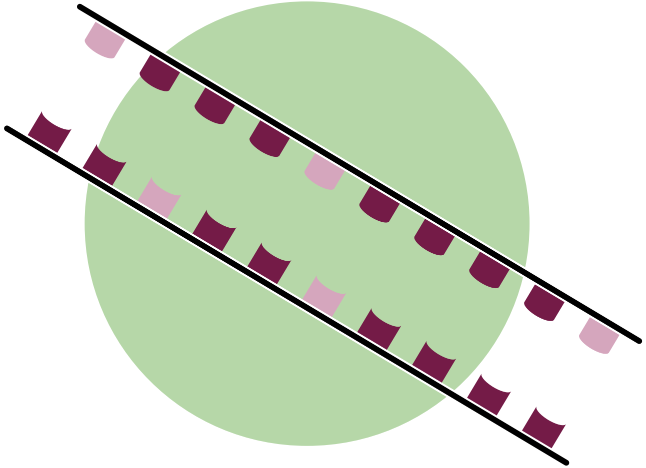
unphased

€

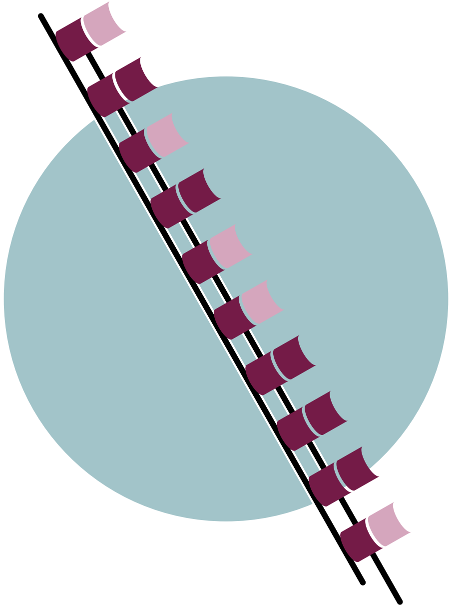


phased

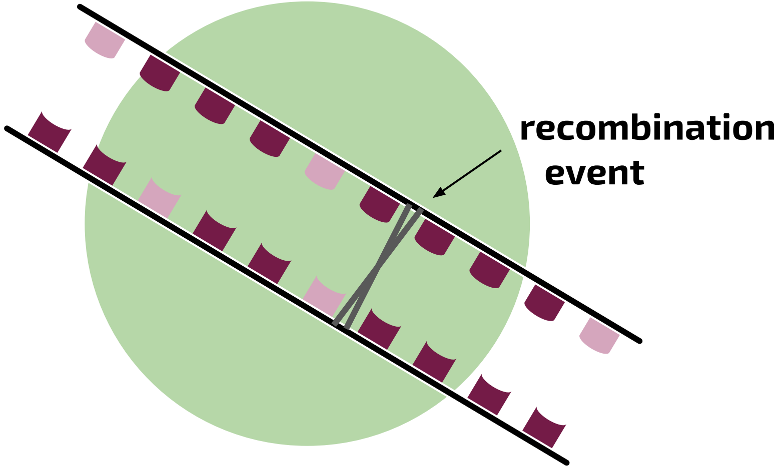
€€€

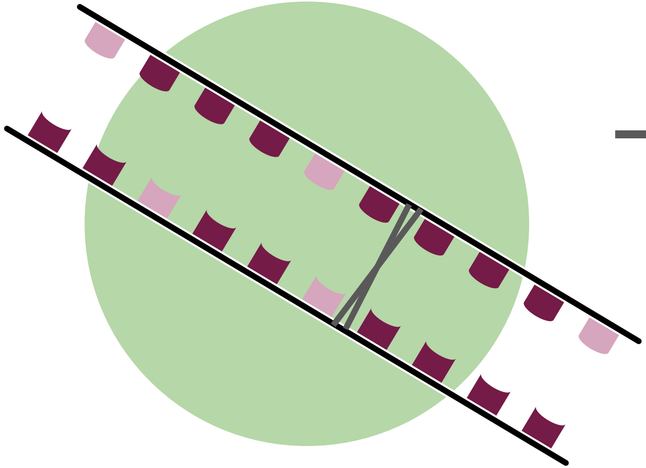


unphased

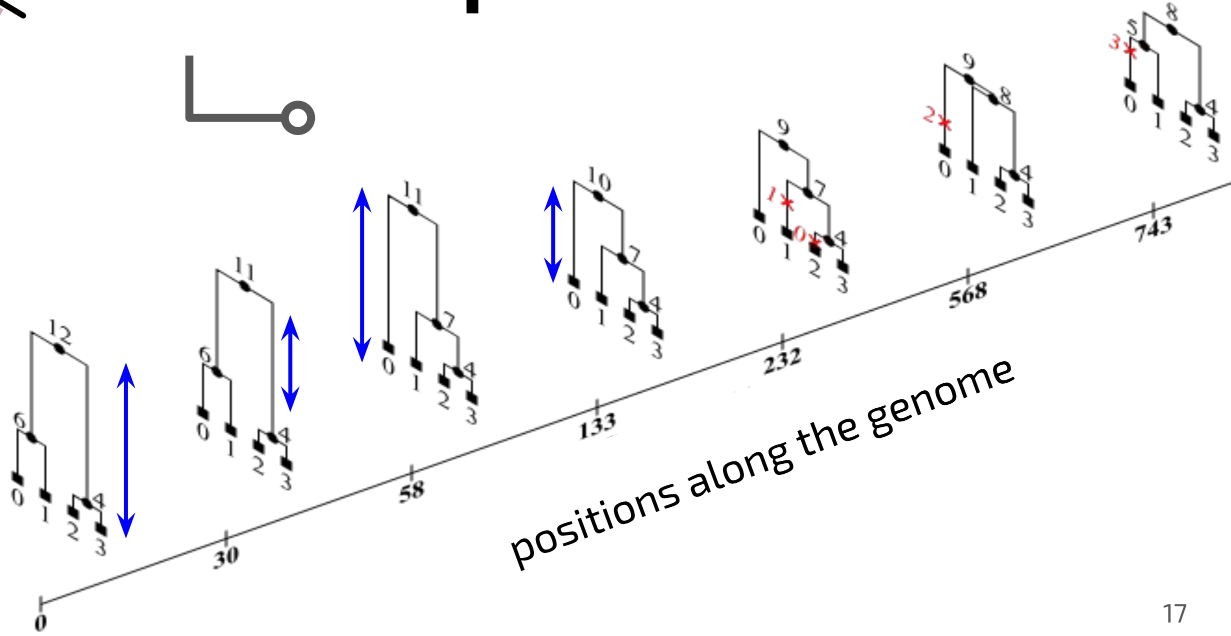


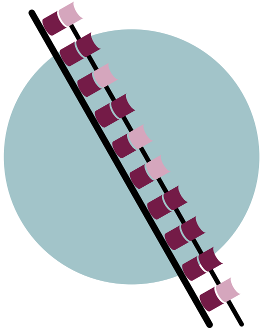
phased





tree sequence

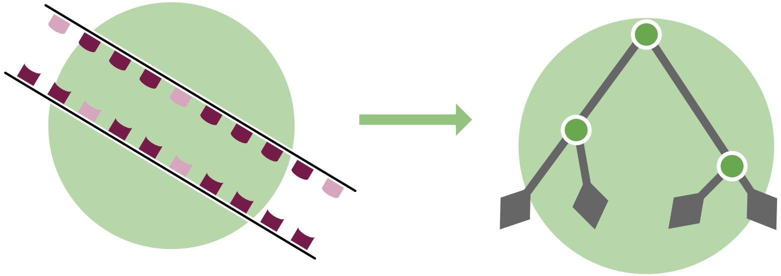




genotypic
statistics

within-pop ——— **between**

- **2D-jAFS** (Allele Frequency Spectrum)
- **Tajima's *D***
- **Nucleotide diversity/divergence**
- **F_{ST}**



genealogical
statistics

within-pop ——— **between**

- **Distributions of $(T_k)_k$**
- **Distributions of $(T_c)_k$**
- **PLFT** (Proportion of Lineages as a Function of Time)

Maruvka et al. (2011)

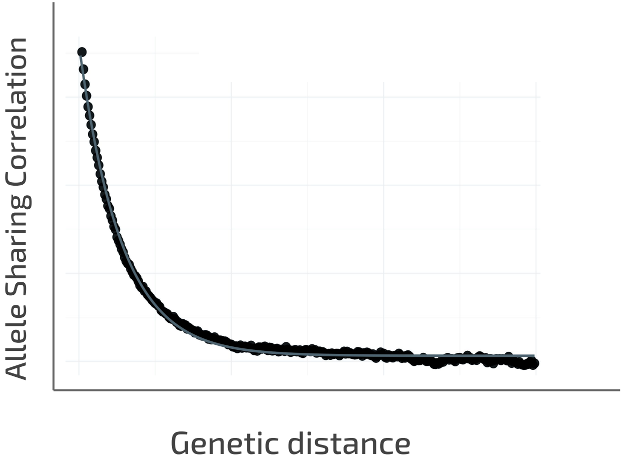
+ subtracted
between - E(within1, within2)

genotypic statistics

within-pop —— between

➔ ***ASCEND***

Tournebize, Chu, Moorjani (2022)

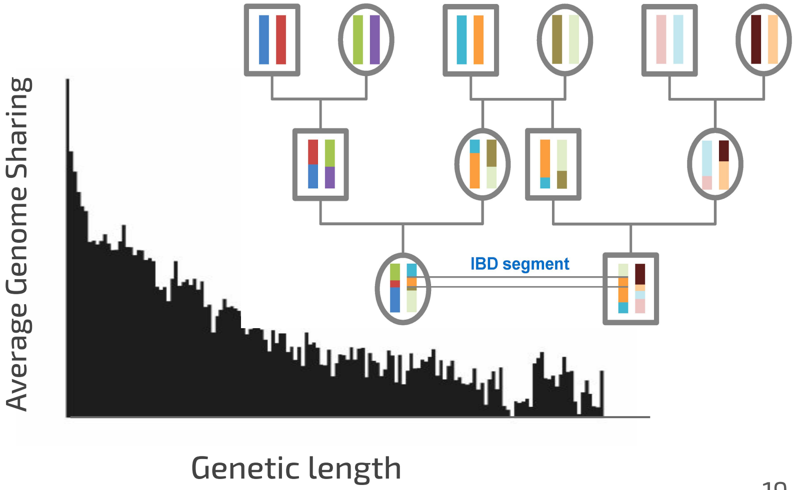


genealogical statistics

within-pop —— between

➔ **Genomic correlations in pairwise coalescences + T_k + T_c + PLFT**

➔ **Distribution of IBD lengths**



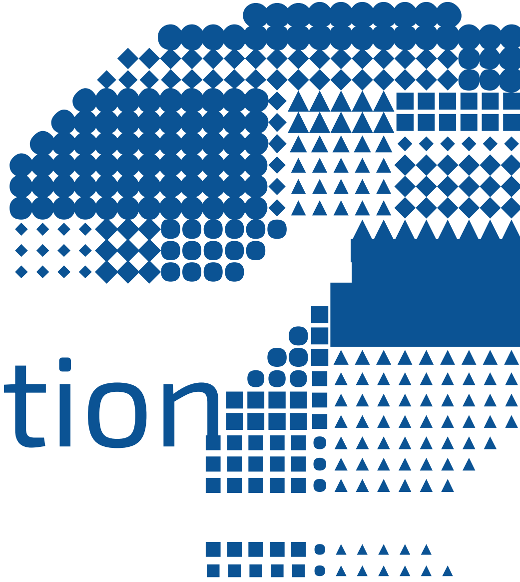
+ subtracted

how well ≠ summary statistics

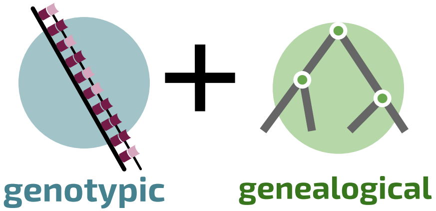
can help us

detect

very recent fragmentation



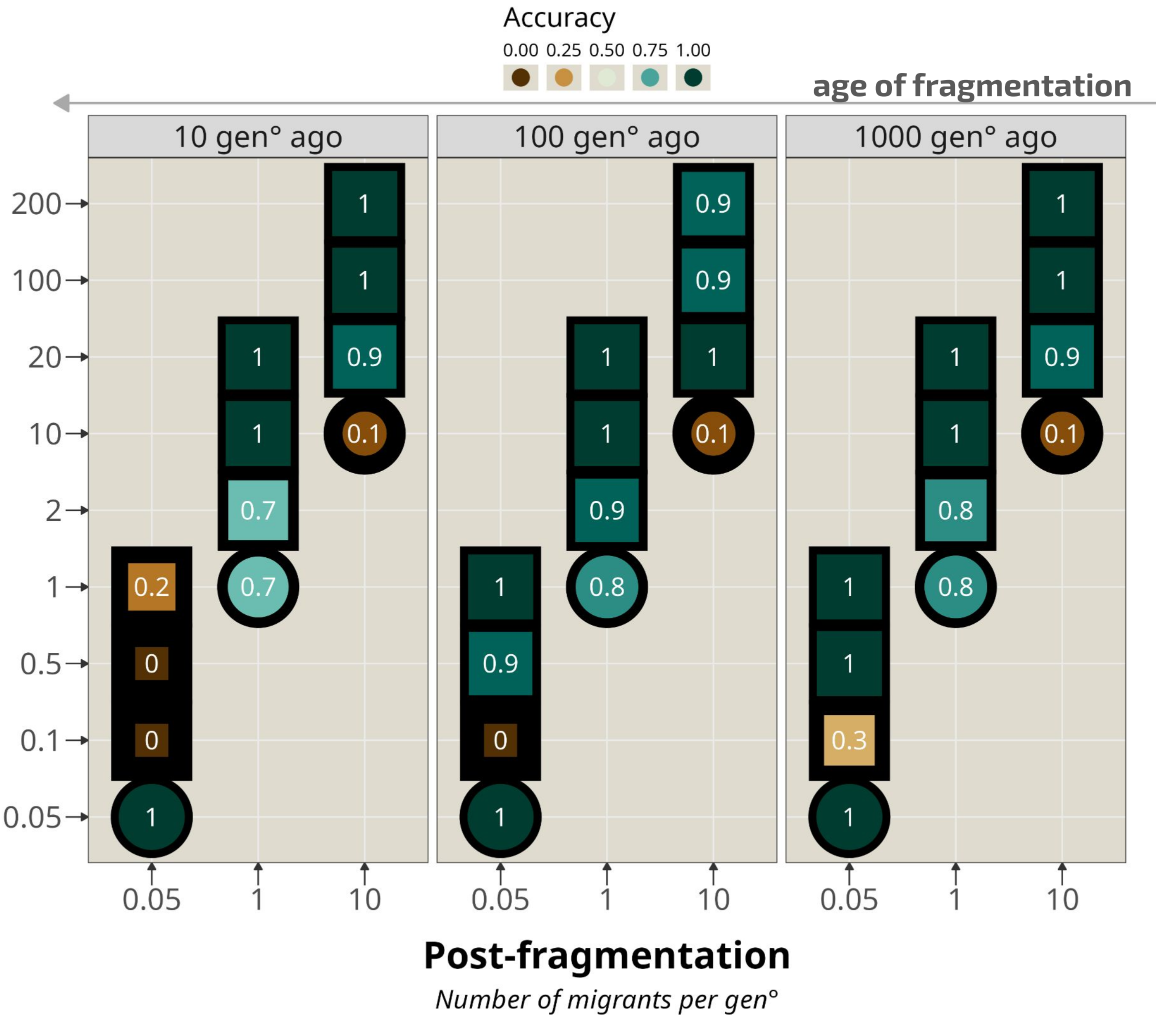
Detection power



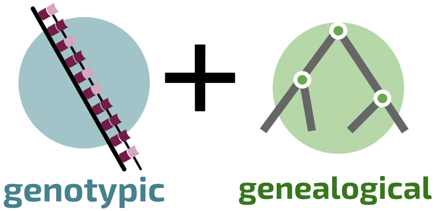
generally a great **power** to detect **fragmentation**

☐ fragmentation

○ stationary



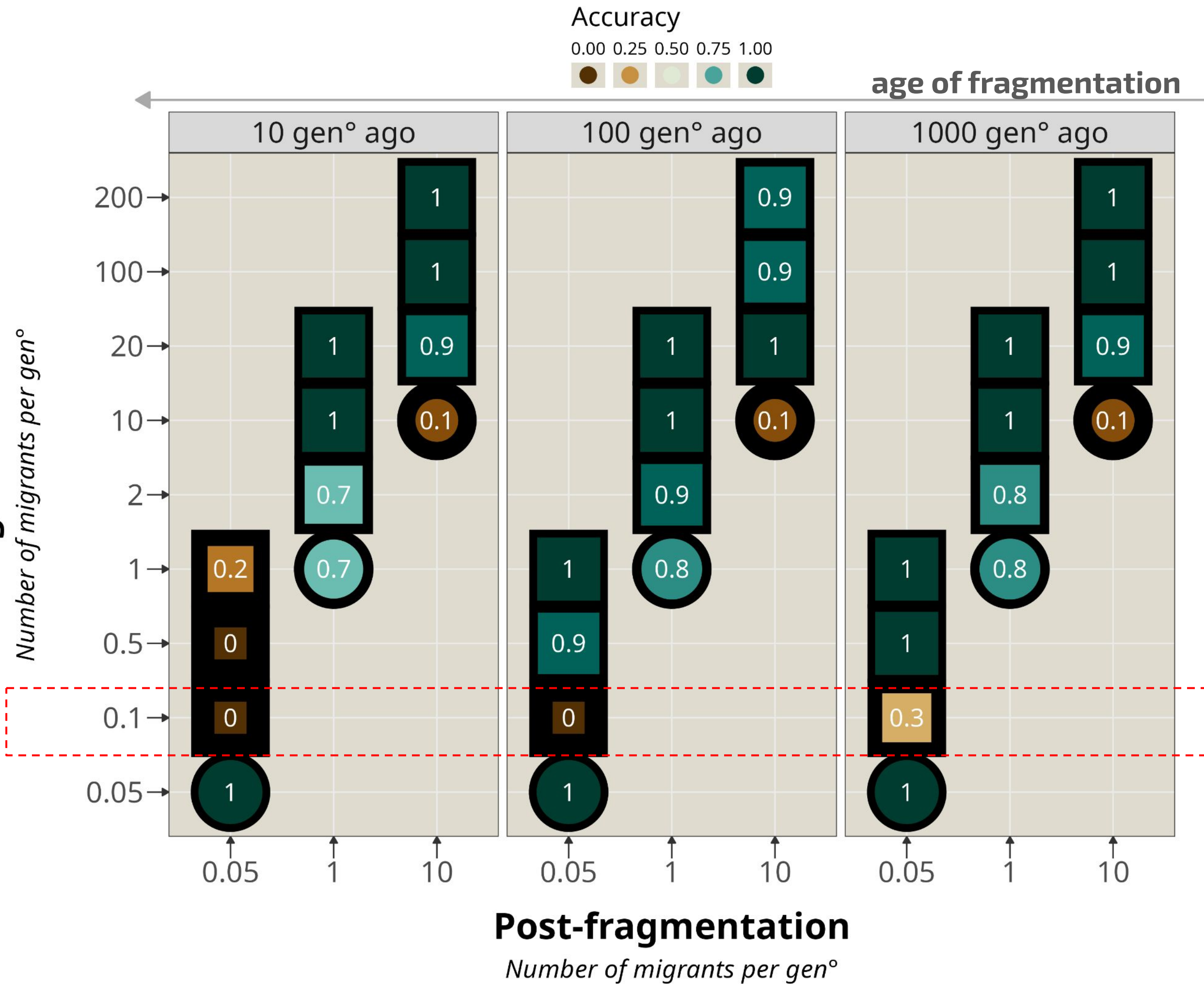
Detection power



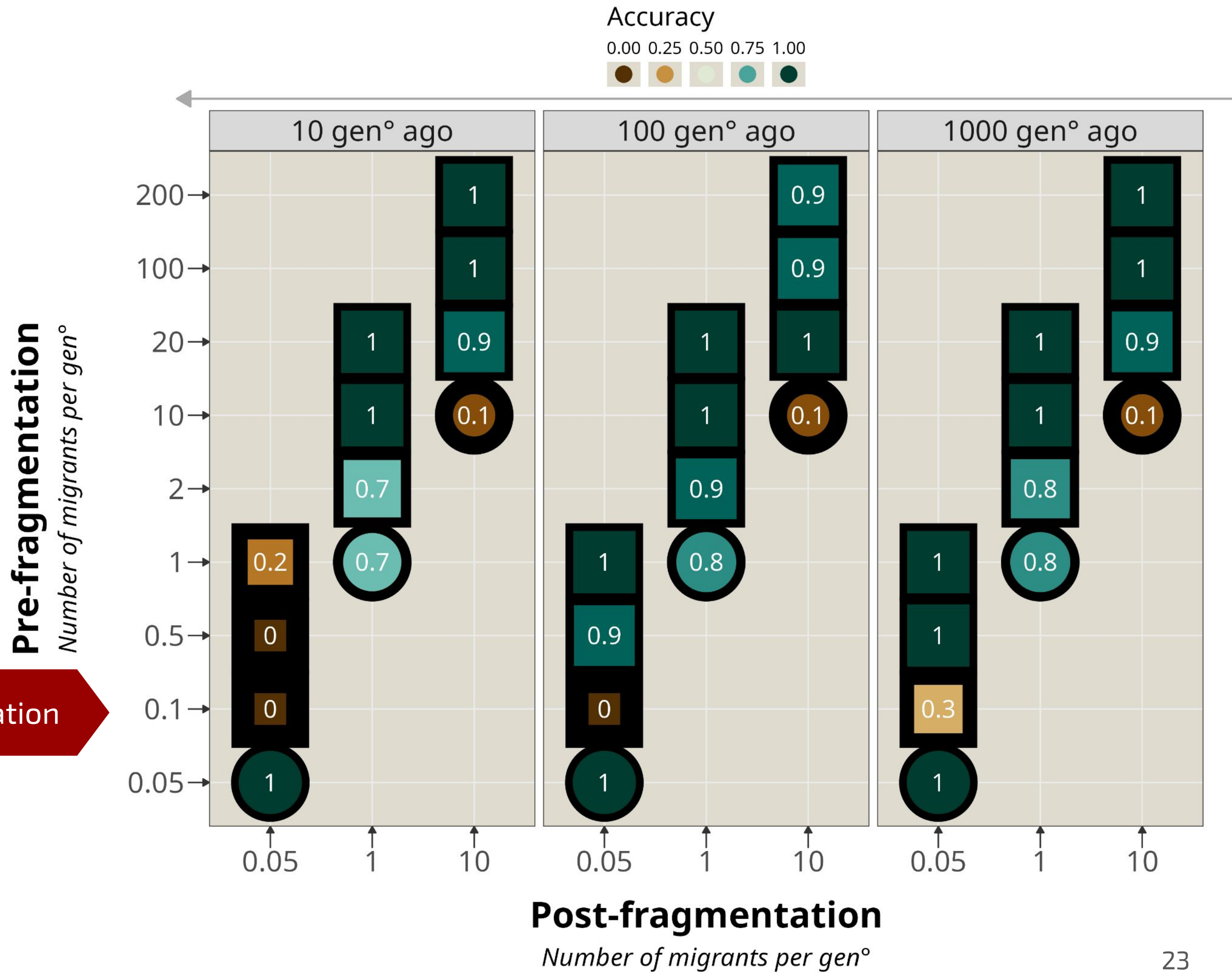
limited power to detect **weak** reduction when pops **already weakly connected**

☐ fragmentation

○ stationary

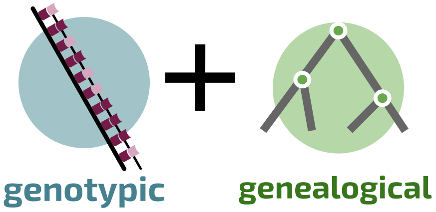


A *conservation genomic* concern

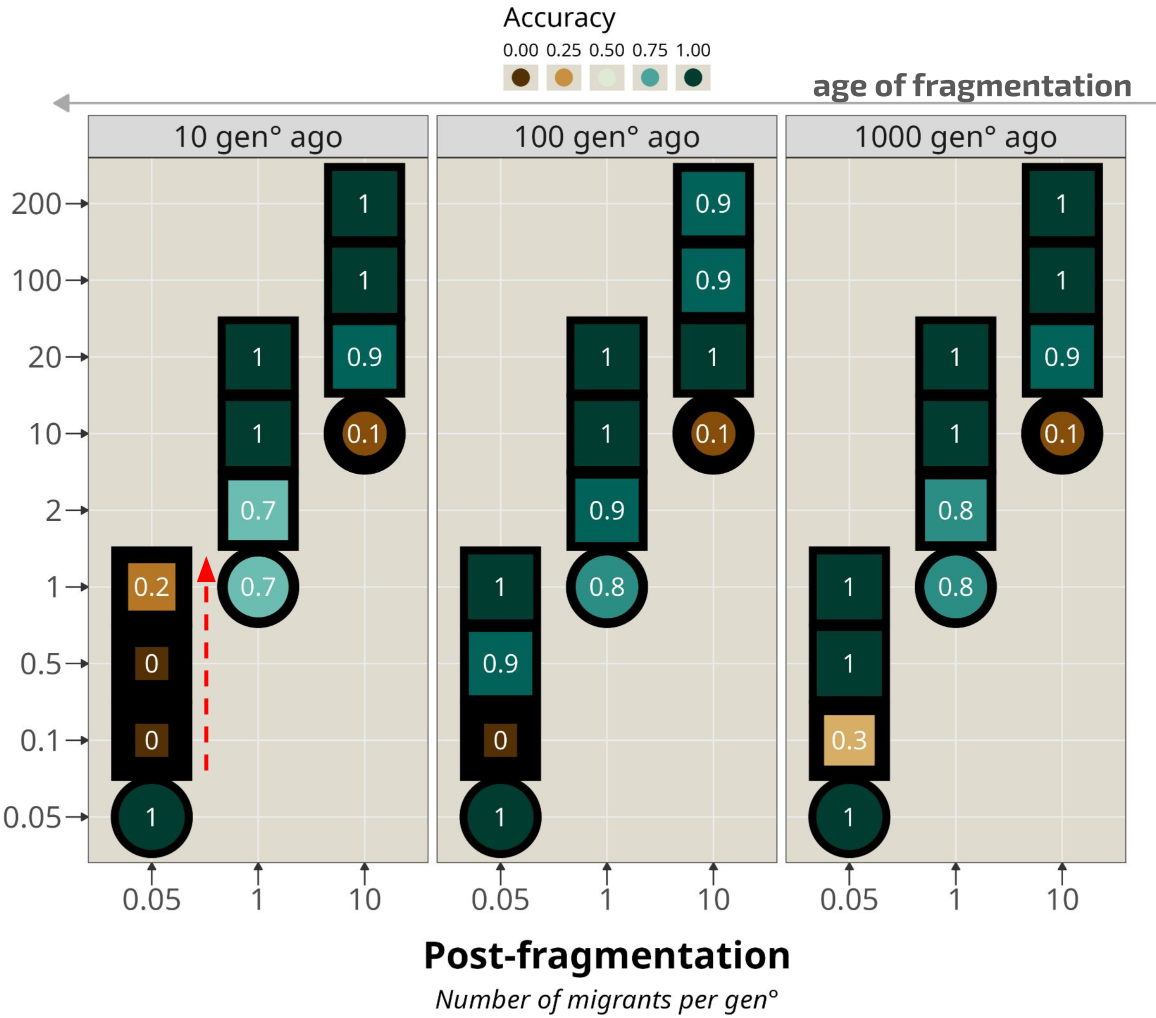


likely to **underdetect** weak fragmentation

Detection power



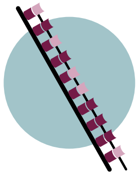
yet **power** increases as fragmentation **intensity increases**



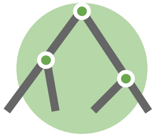
☐ `fragmentation`

○ `stationary`

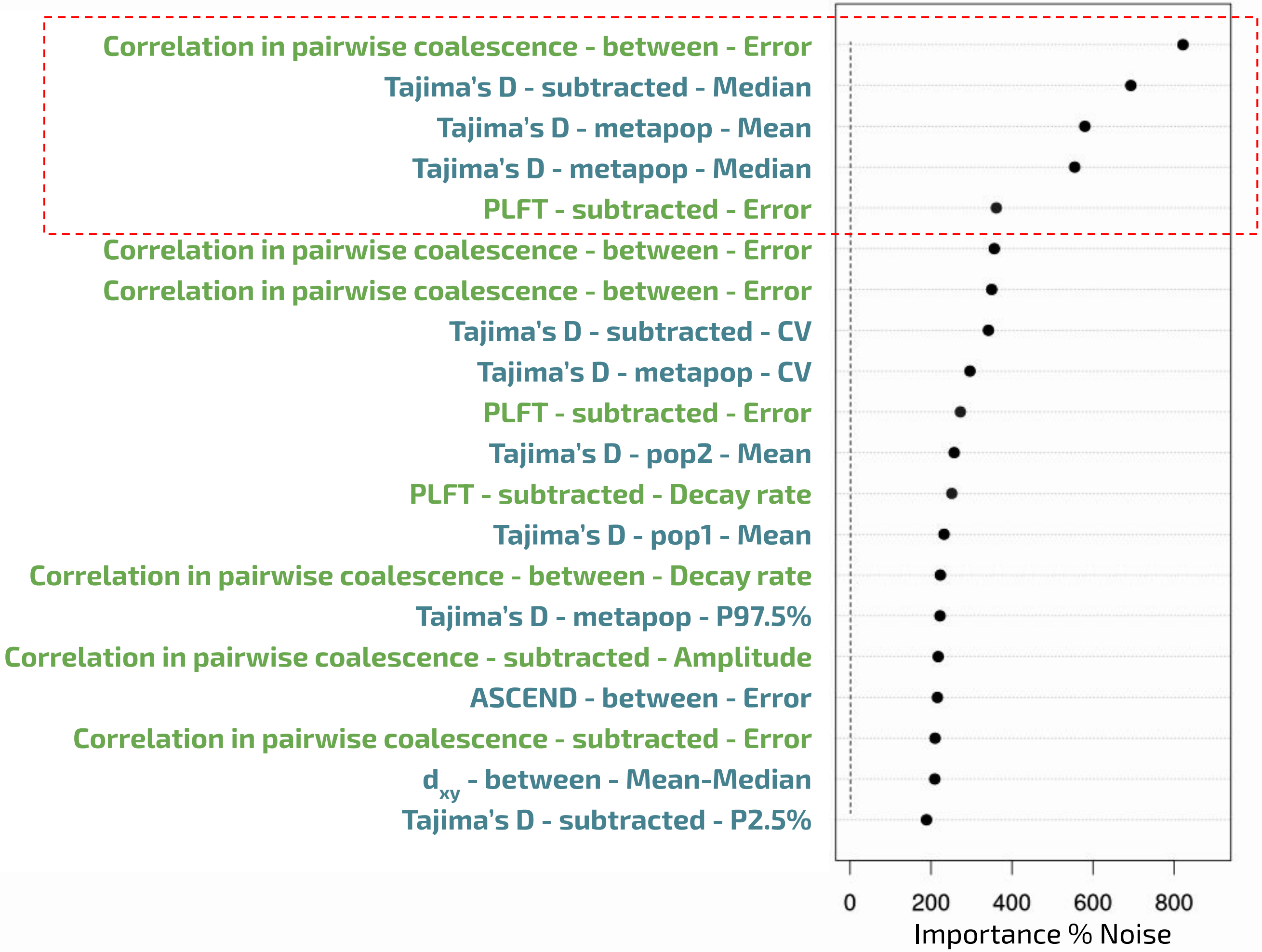
Variable importance



genotypic



genealogical



Conclusions

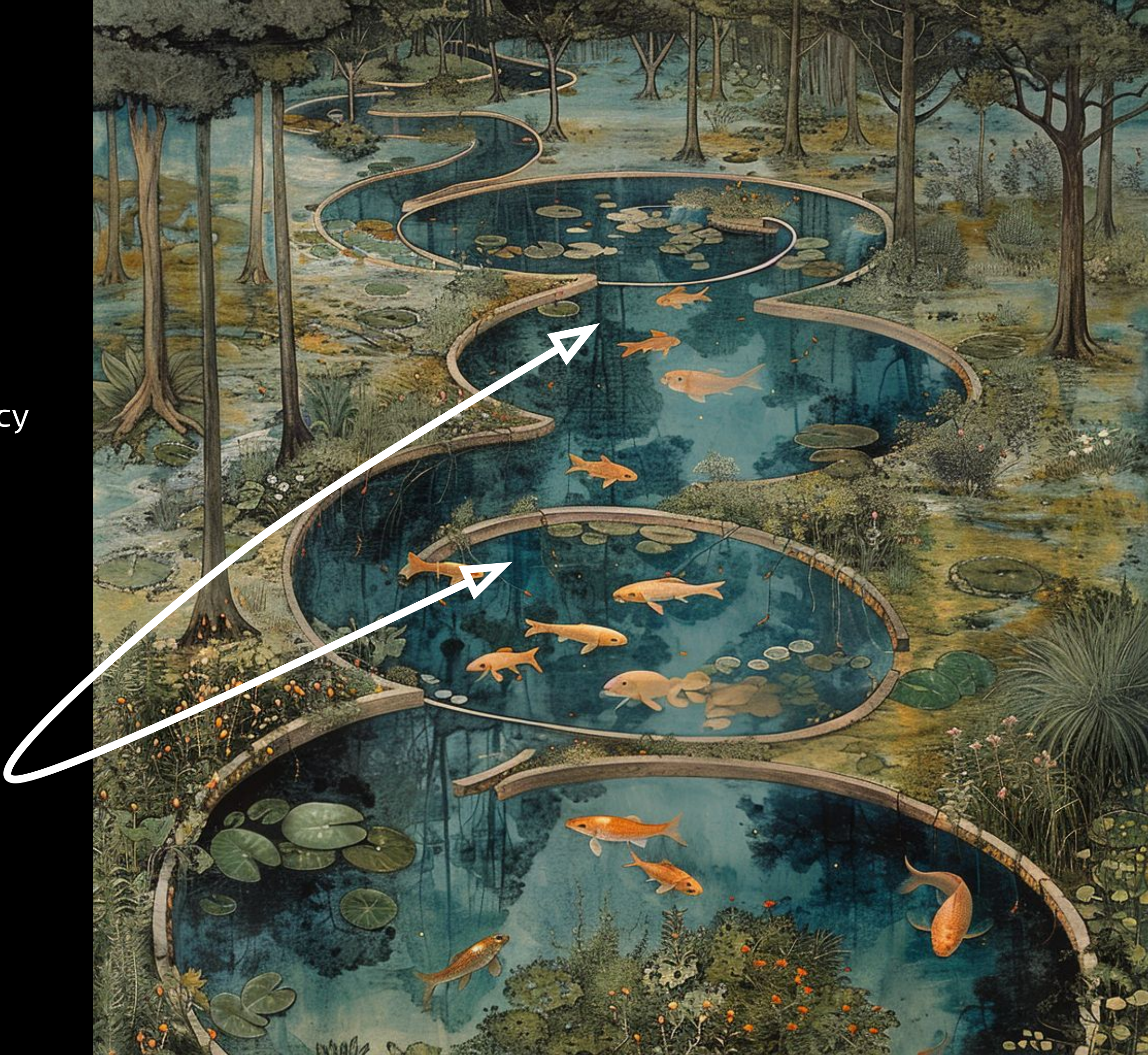
Genomic data **useful** to detect recent fragm°

Genotypic statistics already **sufficient**

Genealogical data **improve** detection accuracy

Still problematic to identify fragmentation if:
weak &
very recent (< 10 gen°) &
pops were **already poorly connected**

But alternatives possible… i.e. *M* in **other** non-fragmented pops as a proxy to M_{anc}



Thank you!

Simon Boitard
Olivier Mazet
Lounès Chikhi
& the DevOcGen team members

BiodivOc
Biodiversité Occitanie

cnrs
dépasser les frontières

UNIVERSITÉ TOULOUSE III PAUL SABATIER
Université de Toulouse

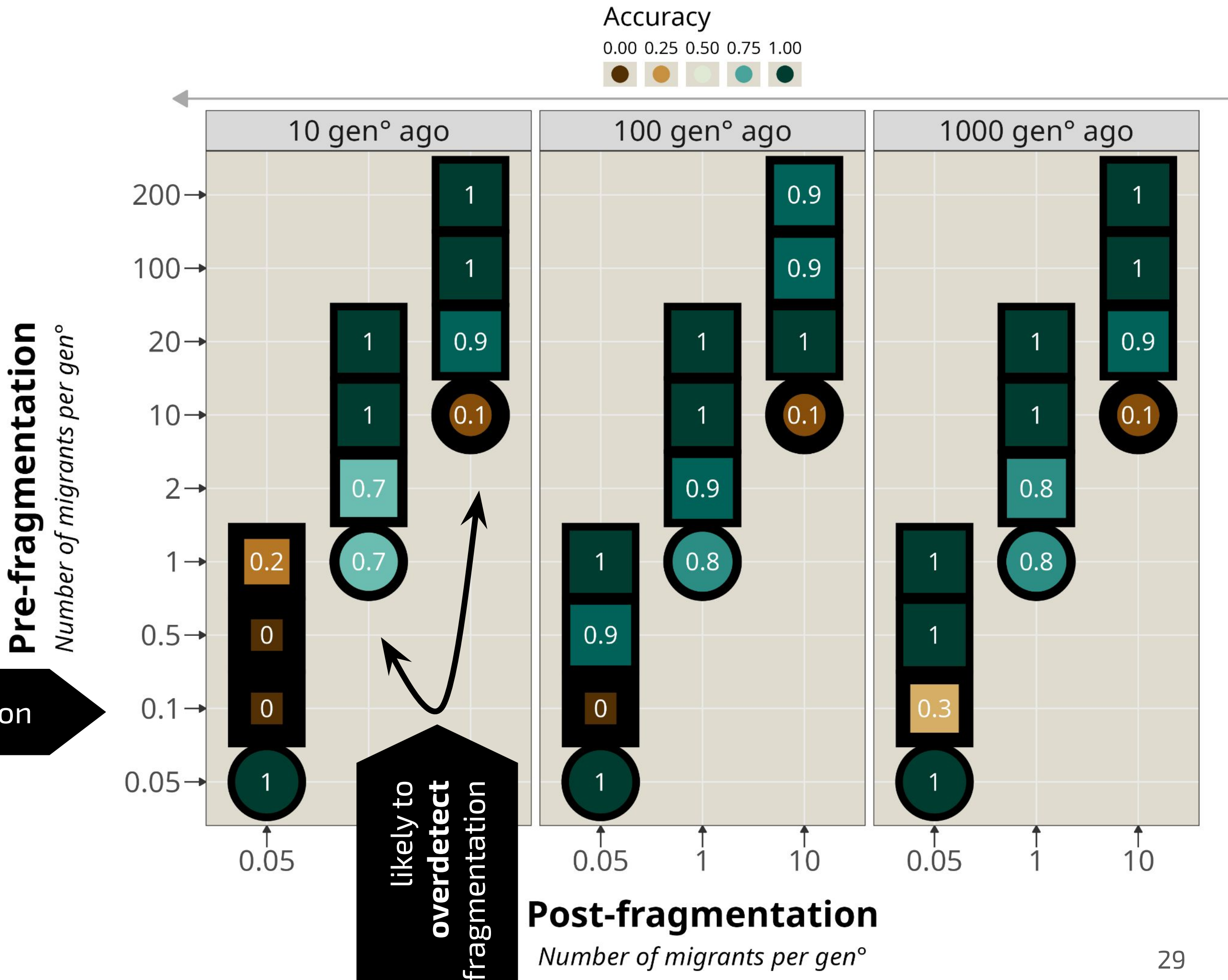
INSTITUT DE MATHÉMATIQUES DE TOULOUSE

INRAE

IRD
Institut de Recherche pour le Développement
FRANCE

Appendix

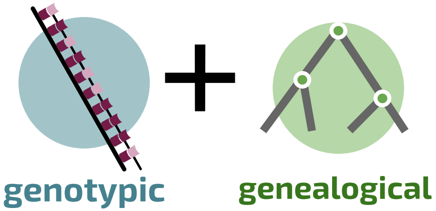
A *conservation genomic* concern



likely to **underdetect** weak fragmentation

likely to **overdetect** fragmentation

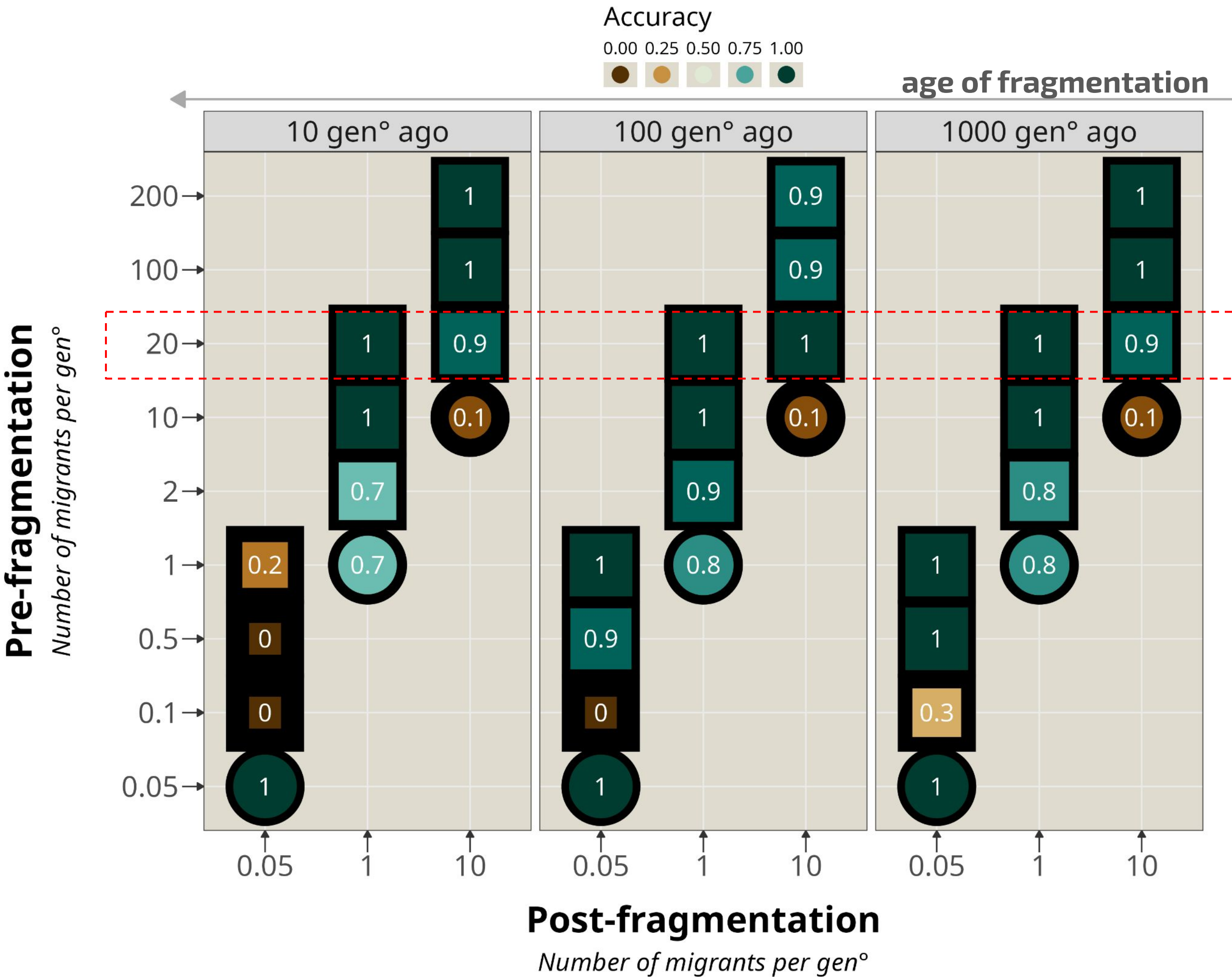
Detection power



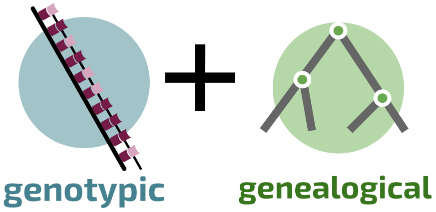
power tends to **increase** with founder **intensity**

☐ `fragmentation`

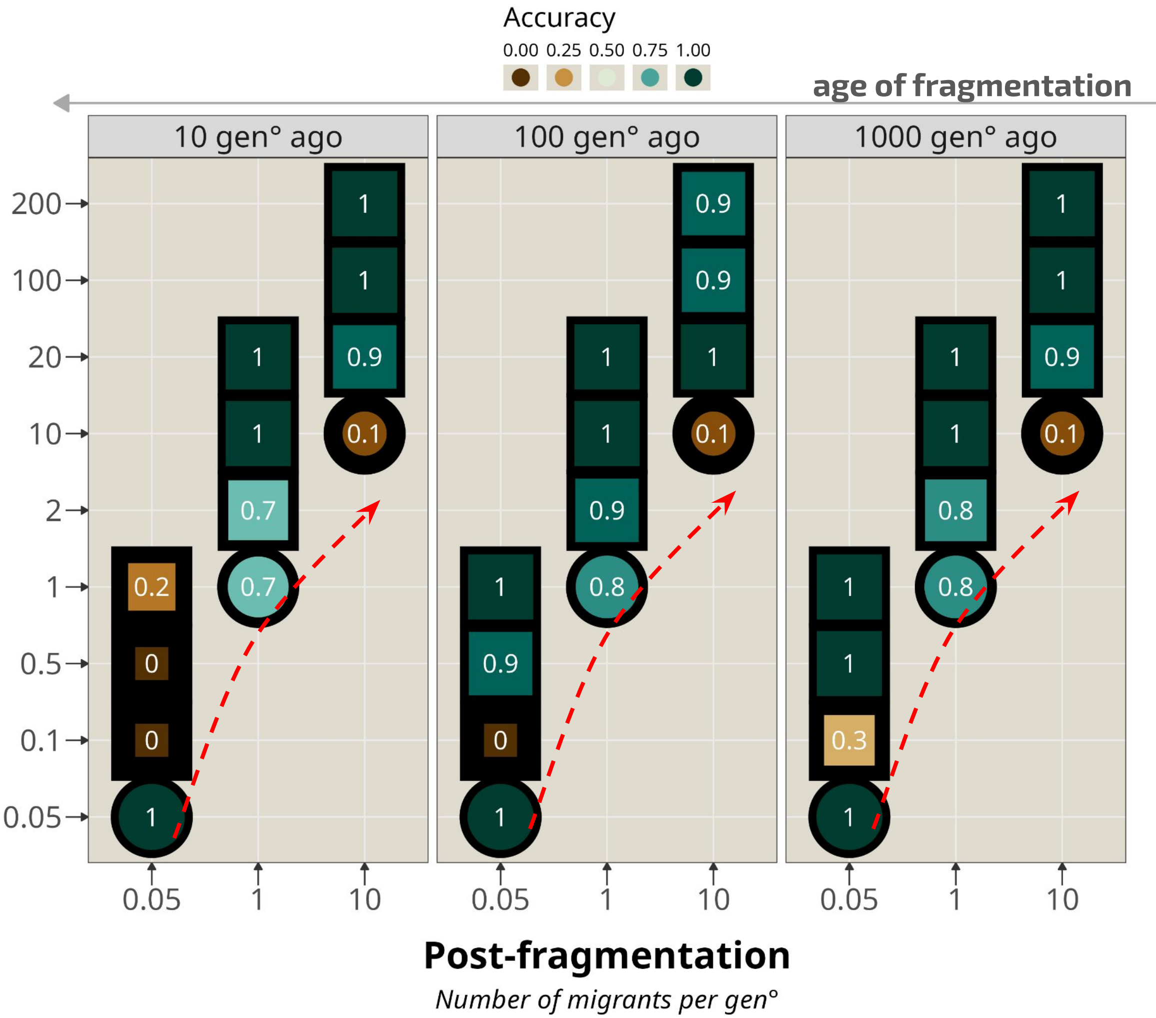
○ `stationary`



Detection power



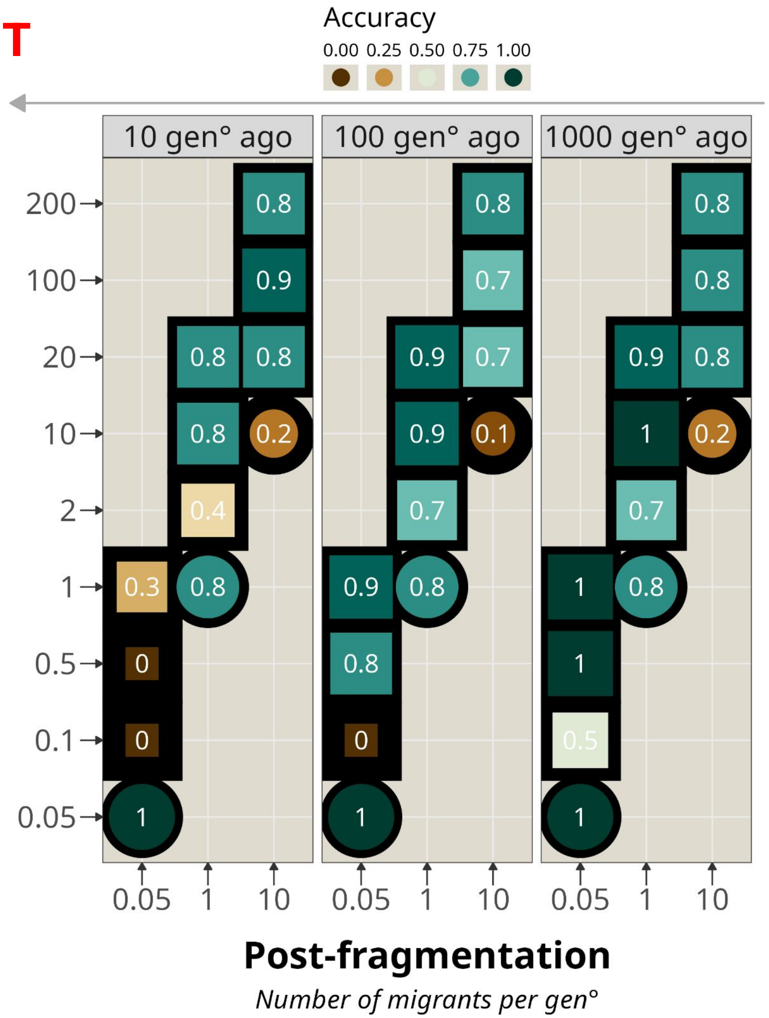
power to detect **stationary** model **decreases** with **migration rate**



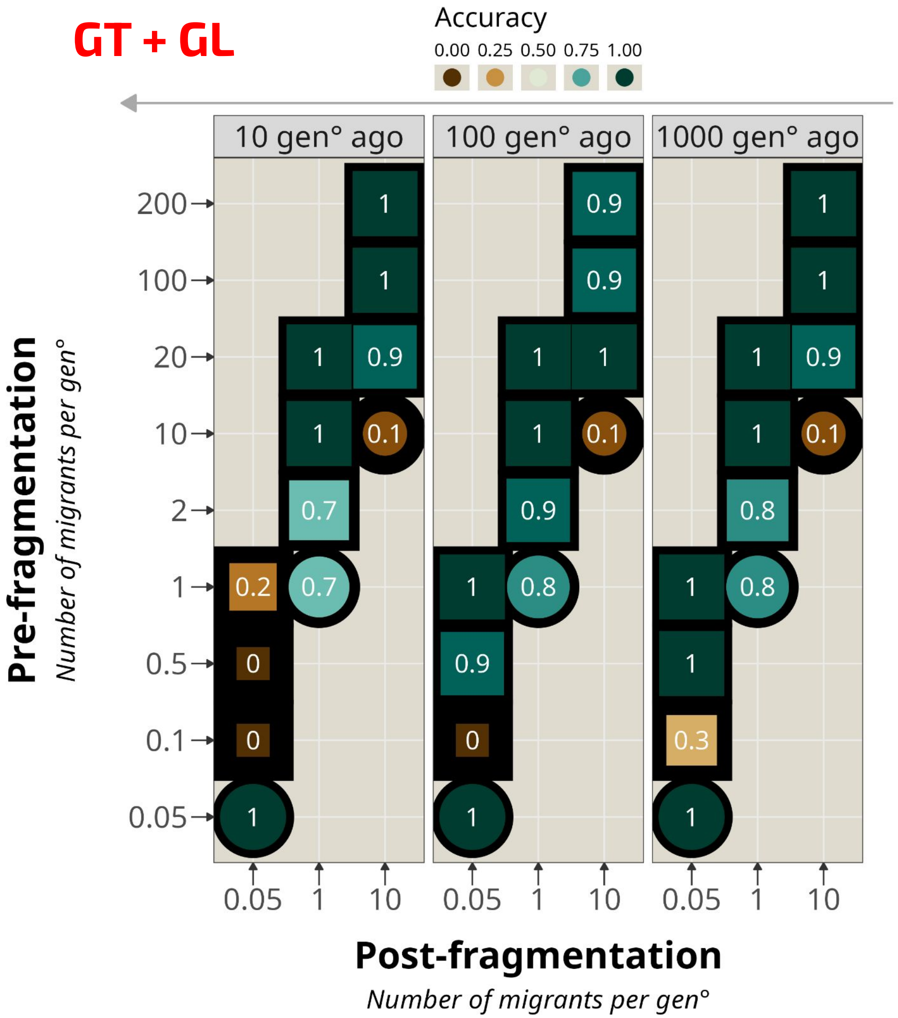
☐ fragmentation

○ stationary

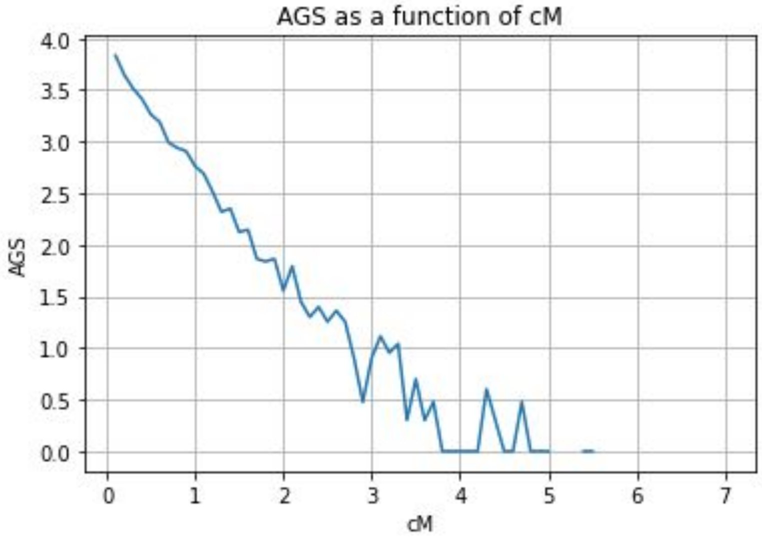
GT


Accuracy
0.00
0.25
0.50
0.75
1.00
10 gen° ago
100 gen° ago
1000 gen° ago
200
100
20
10
2
1
0.5
0.1
0.05
0.8
0.9
0.8
0.8
0.8
0.2
0.4
0.3
0.8
0
0
1
0.8
0.7
0.9
0.7
0.9
0.1
0.7
0.9
0.8
0.8
0
1
0.8
0.8
0.9
0.8
1
0.2
0.7
1
0.8
1
0.5
1
0.05
1
10
0.05
1
10
0.05
1
10
Pre-fragmentation
Number of migrants per gen°
Post-fragmentation
Number of migrants per gen°

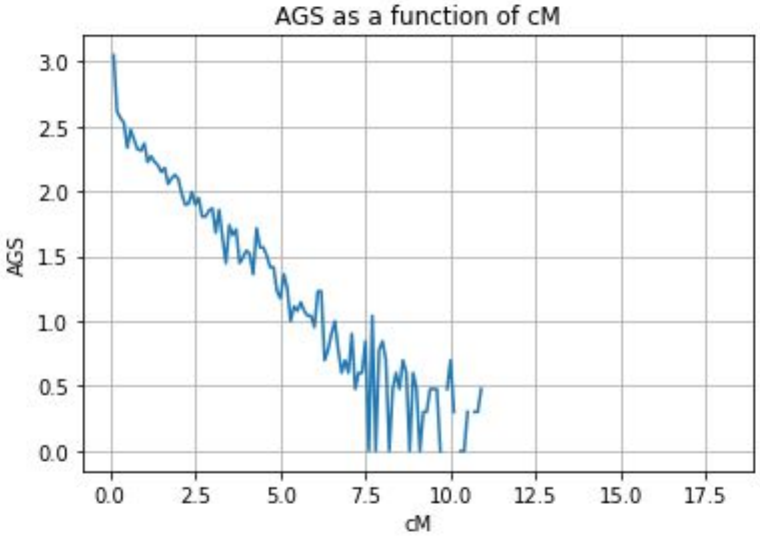
GT + GL


Accuracy
0.00
0.25
0.50
0.75
1.00
10 gen° ago
100 gen° ago
1000 gen° ago
200
100
20
10
2
1
0.5
0.1
0.05
1
1
1
0.9
1
0.1
0.7
0.2
0.7
0
0
1
0.9
0.9
1
1
1
0.1
0.9
1
0.8
0.9
0
1
1
1
1
0.9
1
0.1
0.8
1
0.8
1
0.3
1
0.05
1
10
0.05
1
10
0.05
1
10
Pre-fragmentation
Number of migrants per gen°
Post-fragmentation
Number of migrants per gen°

Hudson


AGS as a function of cM
AGS
cM
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
5
6
7

DTWF


AGS as a function of cM
AGS
cM
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5