Inference of recent fragmentation

Rémi Tournebize, Simon Boitard, Olivier Mazet, Lounès Chikhi

▶ To cite this version:

Rémi Tournebize, Simon Boitard, Olivier Mazet, Lounès Chikhi. Inference of recent fragmentation. 6th European Conference of Tropical Ecology - GTOE2023, The Czech Society for Ecology, Jun 2023, Ceske Budejovice, Czech Republic. hal-04659155

HAL Id: hal-04659155

https://hal.inrae.fr/hal-04659155v1

Submitted on 22 Jul 2024

inference of recent fragmentation

Rémi Tournebize
Simon Boitard
Olivier Mazet
Lounès Chikhi

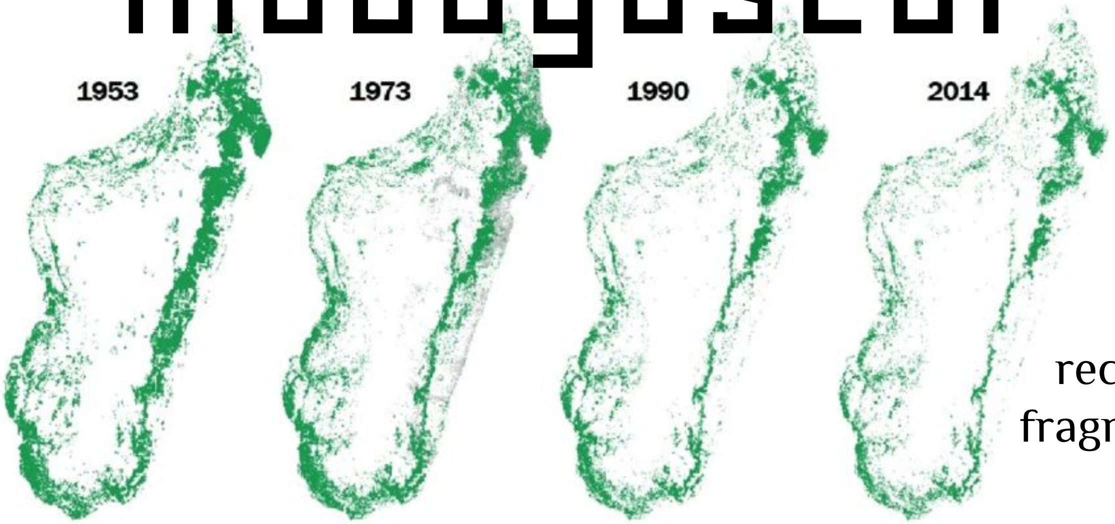
UMR EDB
Toulouse
France

- GTOE -
20 VI 2023

Populations

populations

Madagascar



recent
fragmentation

forest cover

-44%

Rakotoarinivo *et al.* (2020)

when?

did it **occur**

how much?

did it **affect**

CONSERVATION

how fragmentation **will affect** the genomic composition of populations?

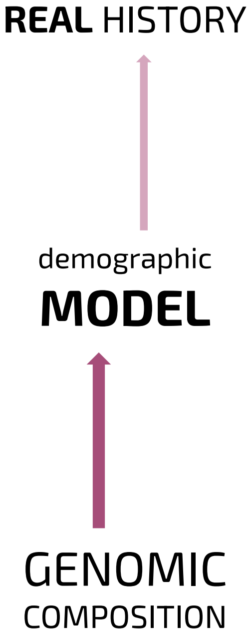
REAL HISTORY

demographic
MODEL

GENOMIC
COMPOSITION

THEORETICAL

which **genomic statistics** will best help us **infer** the demographic model?



REAL HISTORY

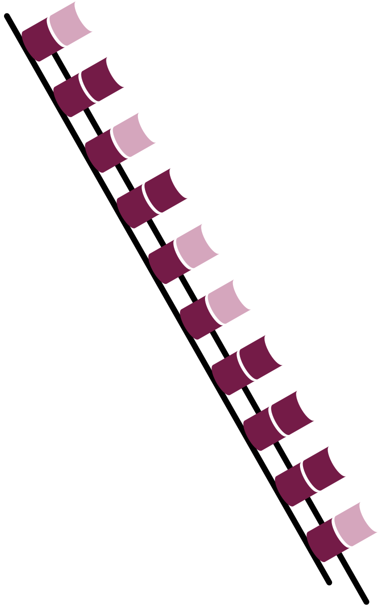
demographic

MODEL

GENOMIC

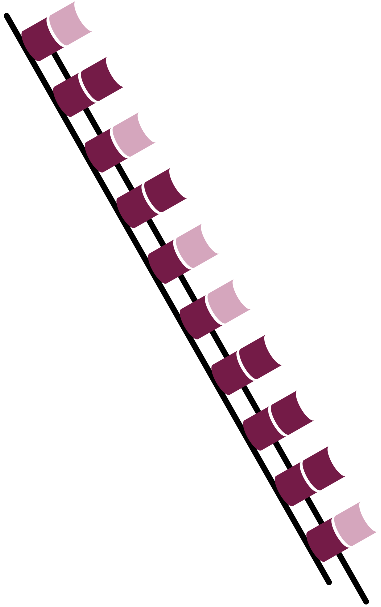
COMPOSITION

Types of genomic data

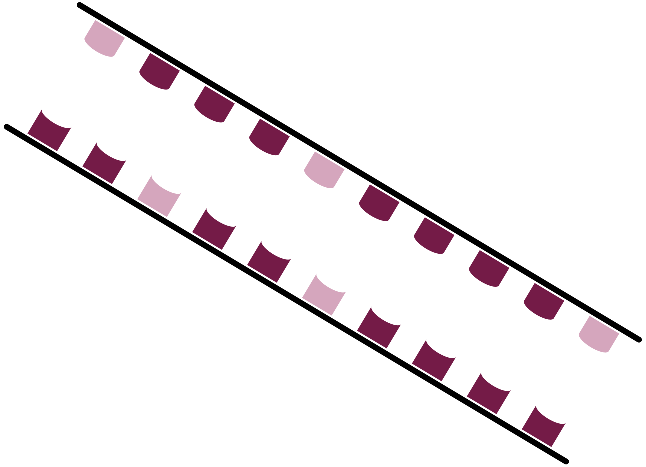


Types of genomic <u>data</u>

unphased

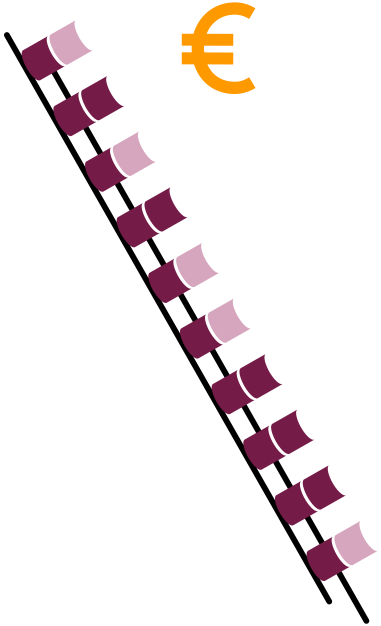


phased

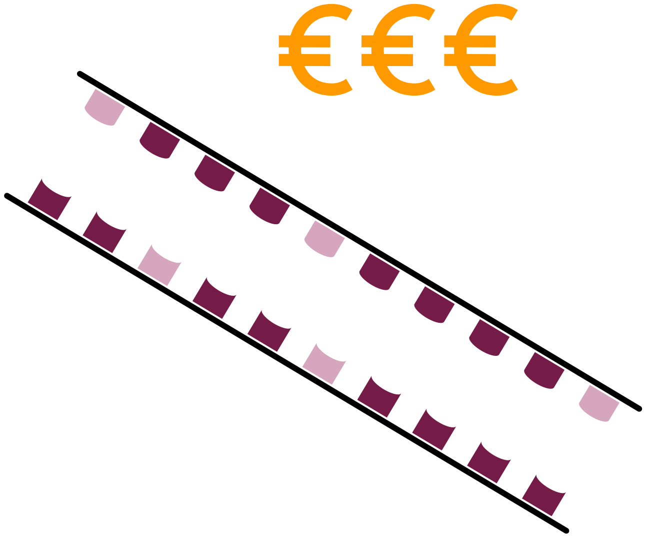


Types of genomic data

unphased

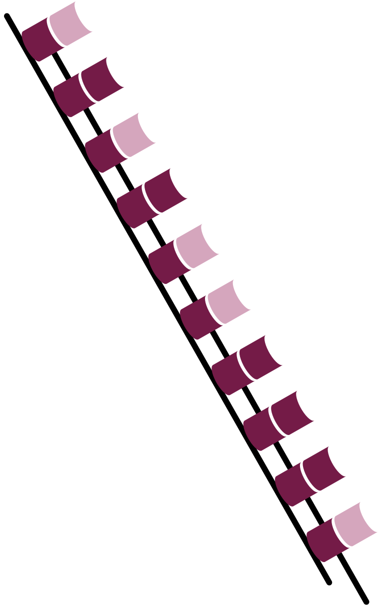


phased

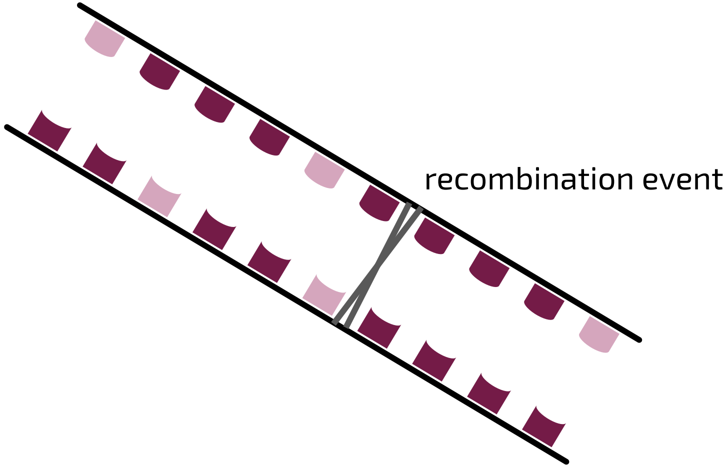


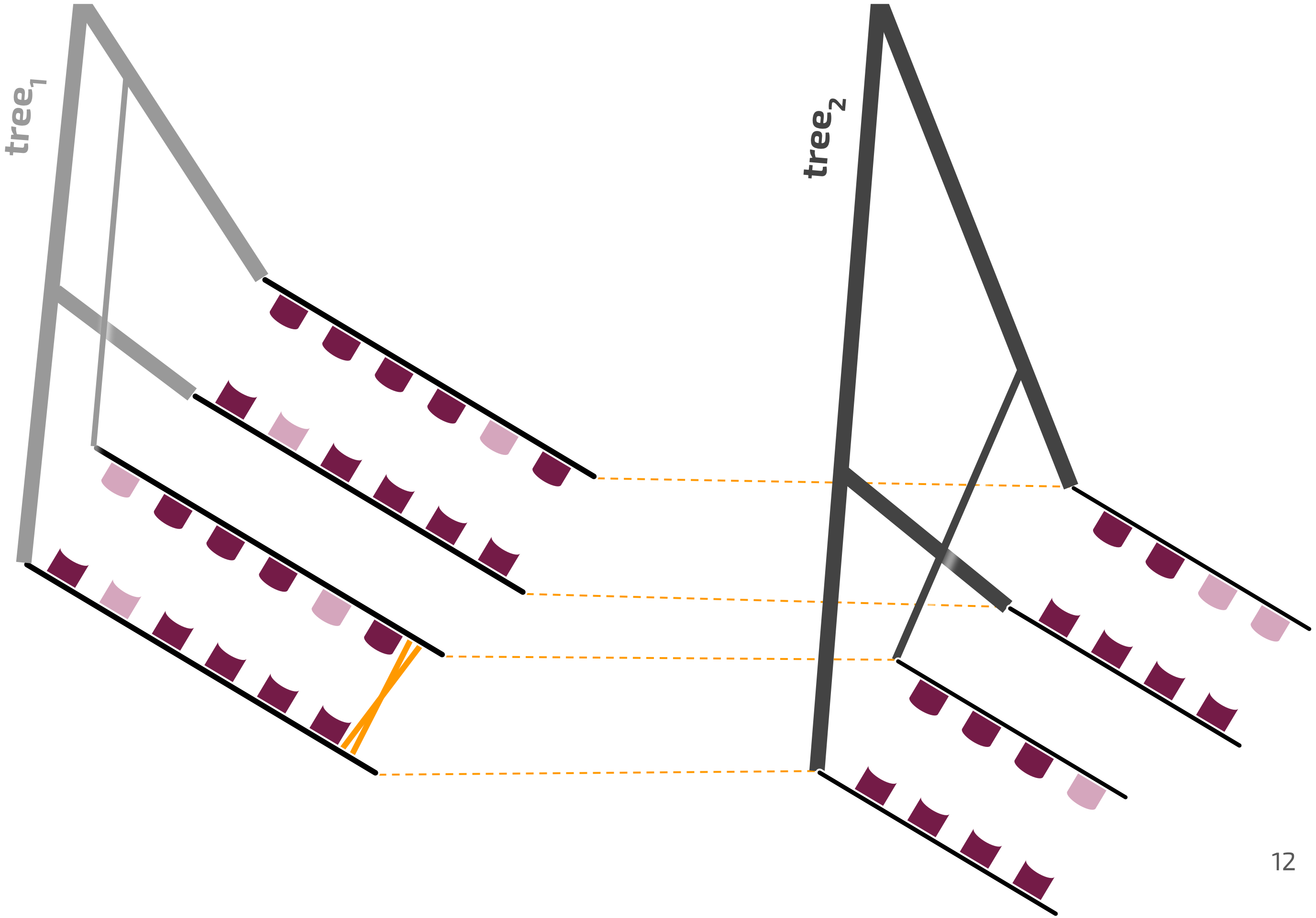
Types of genomic data

unphased



phased

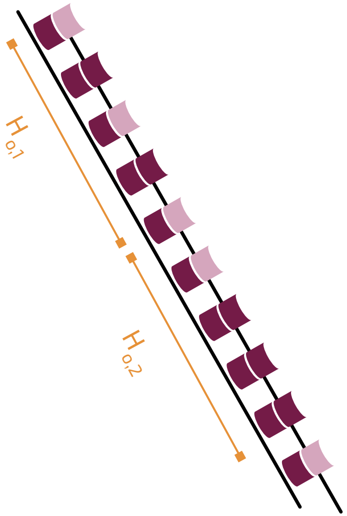



$tree_1$
$tree_2$

Types of genomic statistics

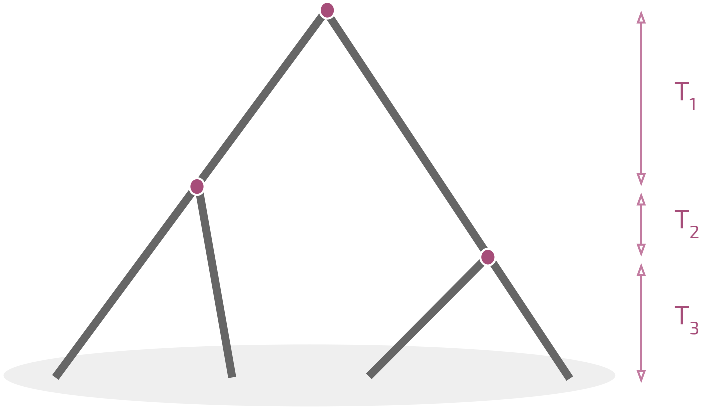
genotype-based

unphased genetic data

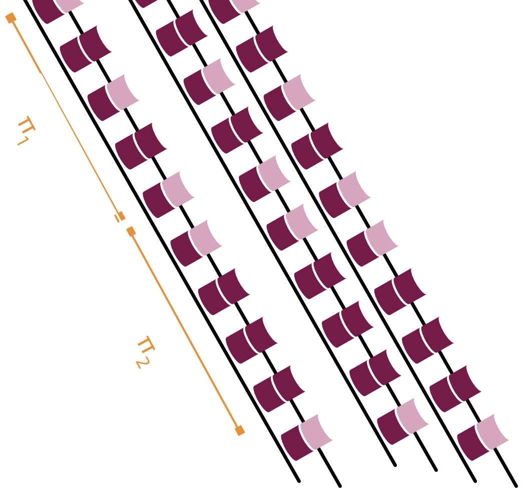


genealogy-based

phased genetic data



Genotype-based statistics

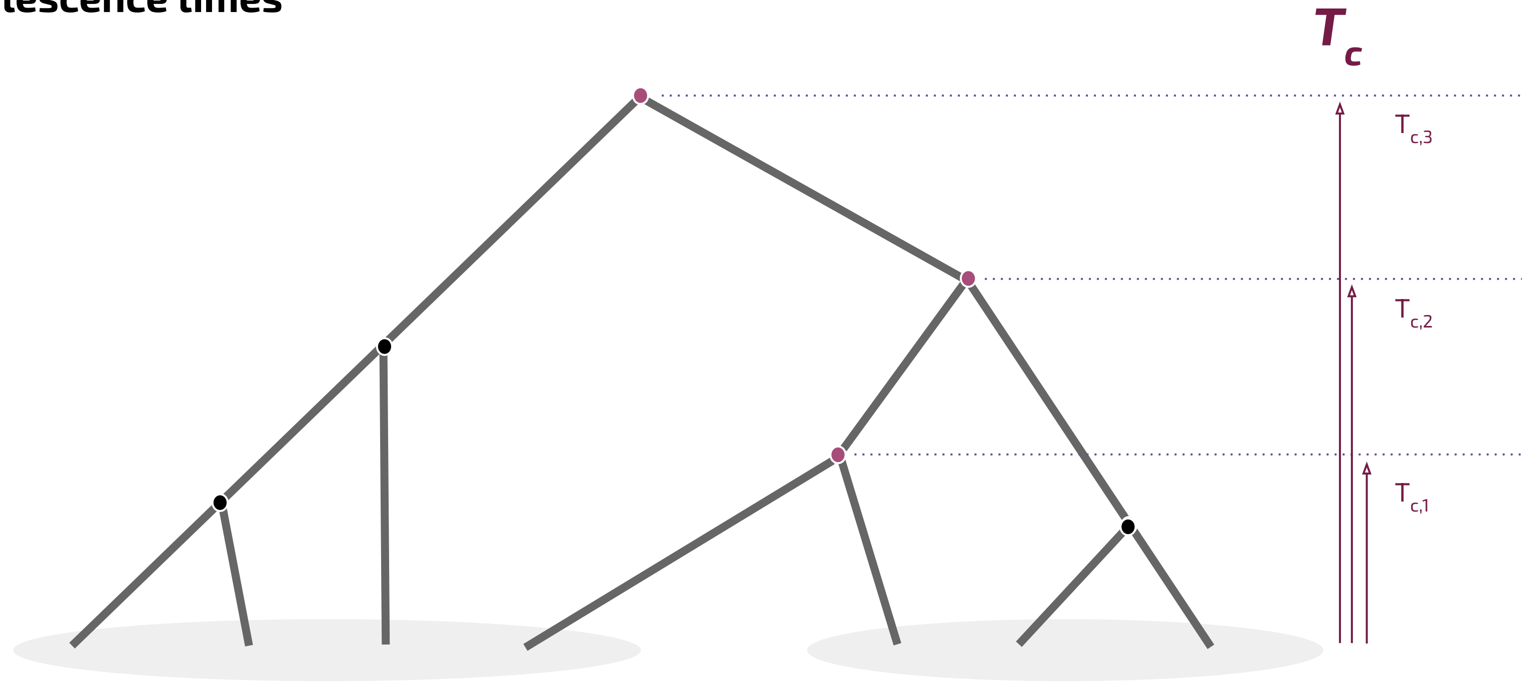


within-population	***between-populations***
nucleotide diversity	nucleotide divergence
θ_W	F_{ST}
D_{Tajima}	ΔD_{Tajima}
$ASCEND_{\mathrm{w}}$	$ASCEND_{\mathrm{b}}$

Tournebize et al. (2022) *PLoS Genetics*

Genealogy-based statistics

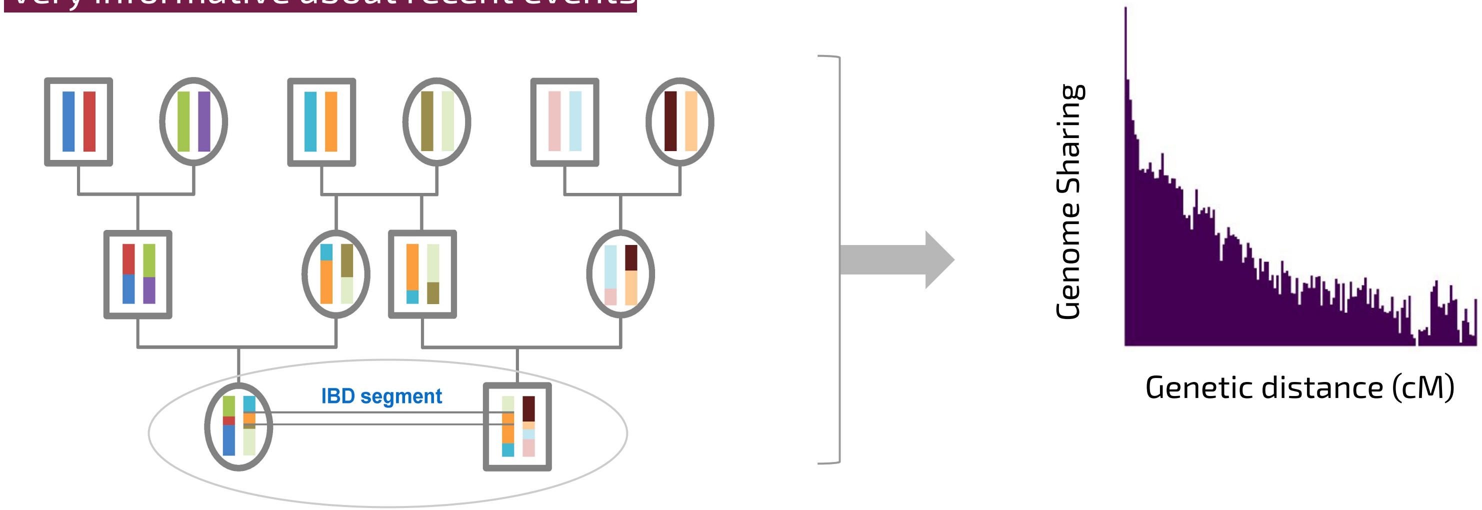
coalescence times



Genealogy-based statistics

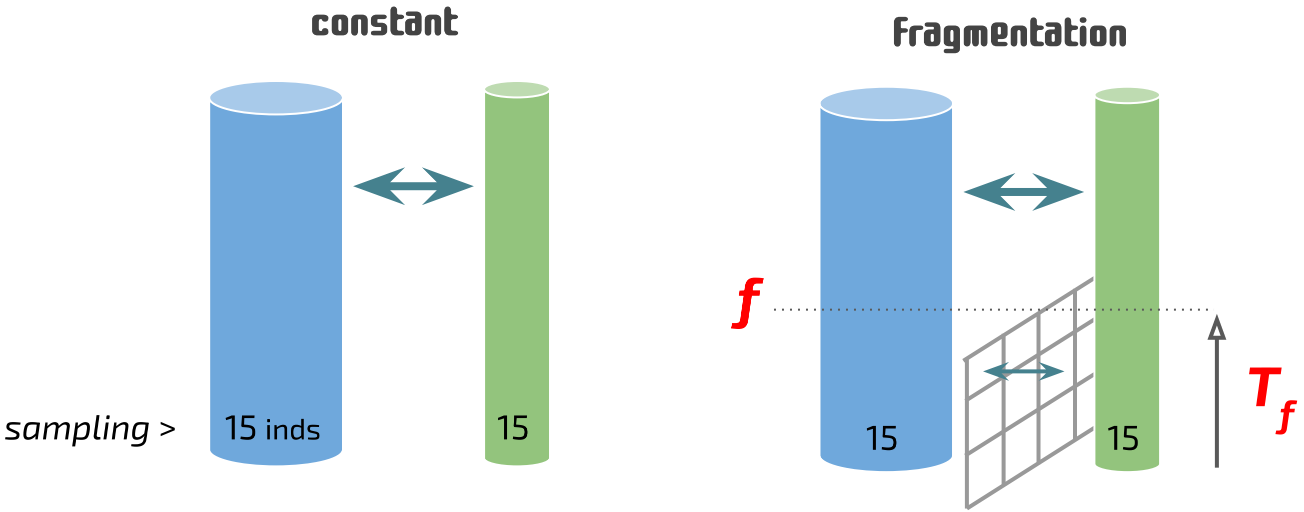
lengths of the **identity-by-descent** [**IBD**] segments

very informative about recent events



Study by simulations

2,500 simulations
of **two** models
using coalescent theory

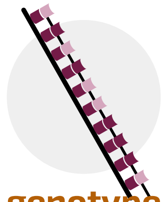


inference in **ABC** with **random forests**

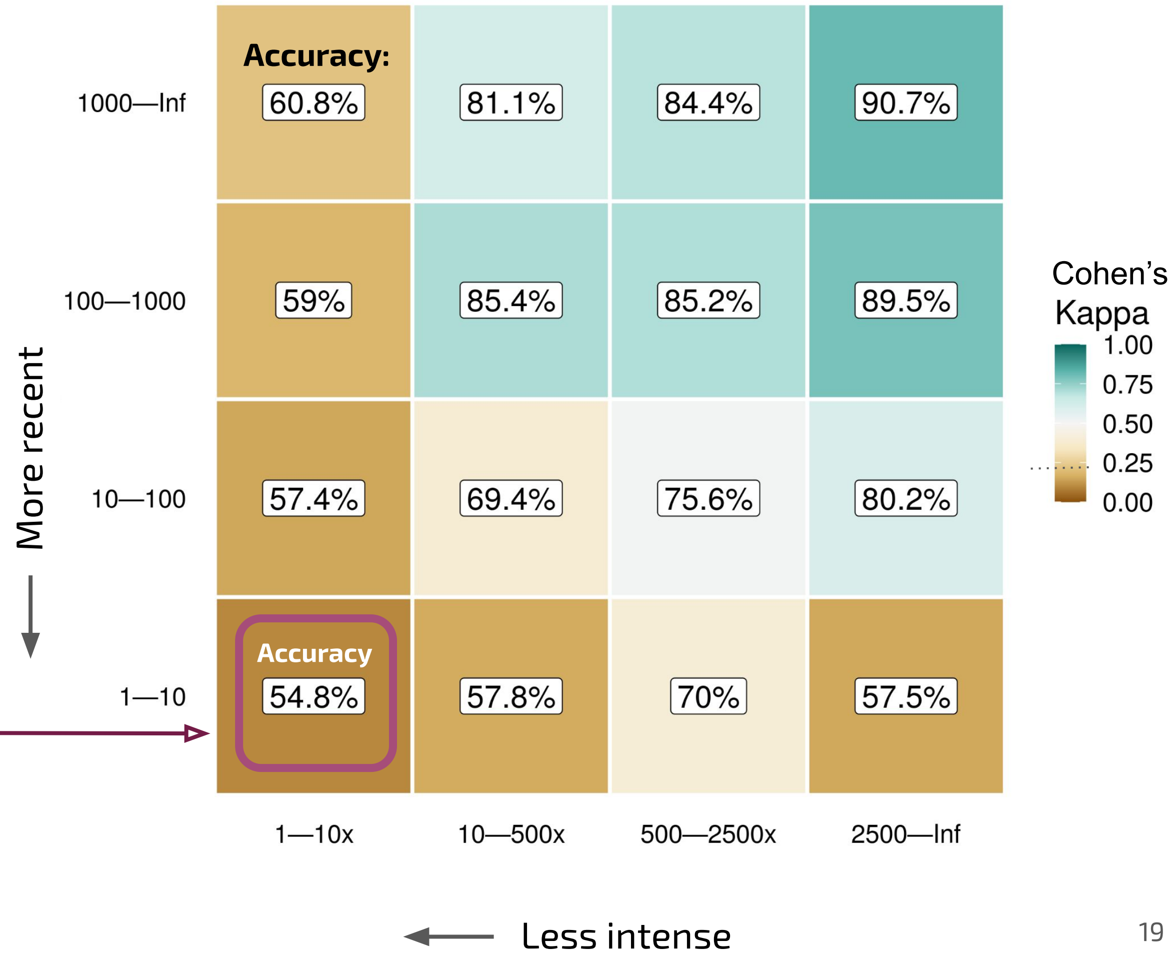
$$T_f \sim Stat_1 + Stat_2 + Stat_3 + \ldots$$

How well can we *detect* fragmentation?

Detection power



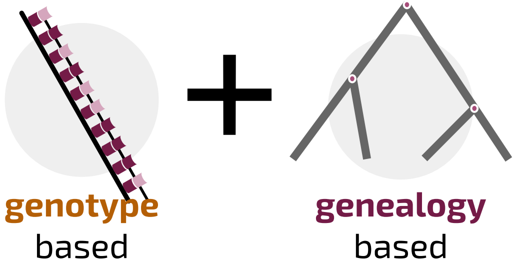
genotype
based



	1—10x	10—500x	500—2500x	2500—Inf
1000—Inf	60.8%	81.1%	84.4%	90.7%
100—1000	59%	85.4%	85.2%	89.5%
10—100	57.4%	69.4%	75.6%	80.2%
1—10	54.8%	57.8%	70%	57.5%

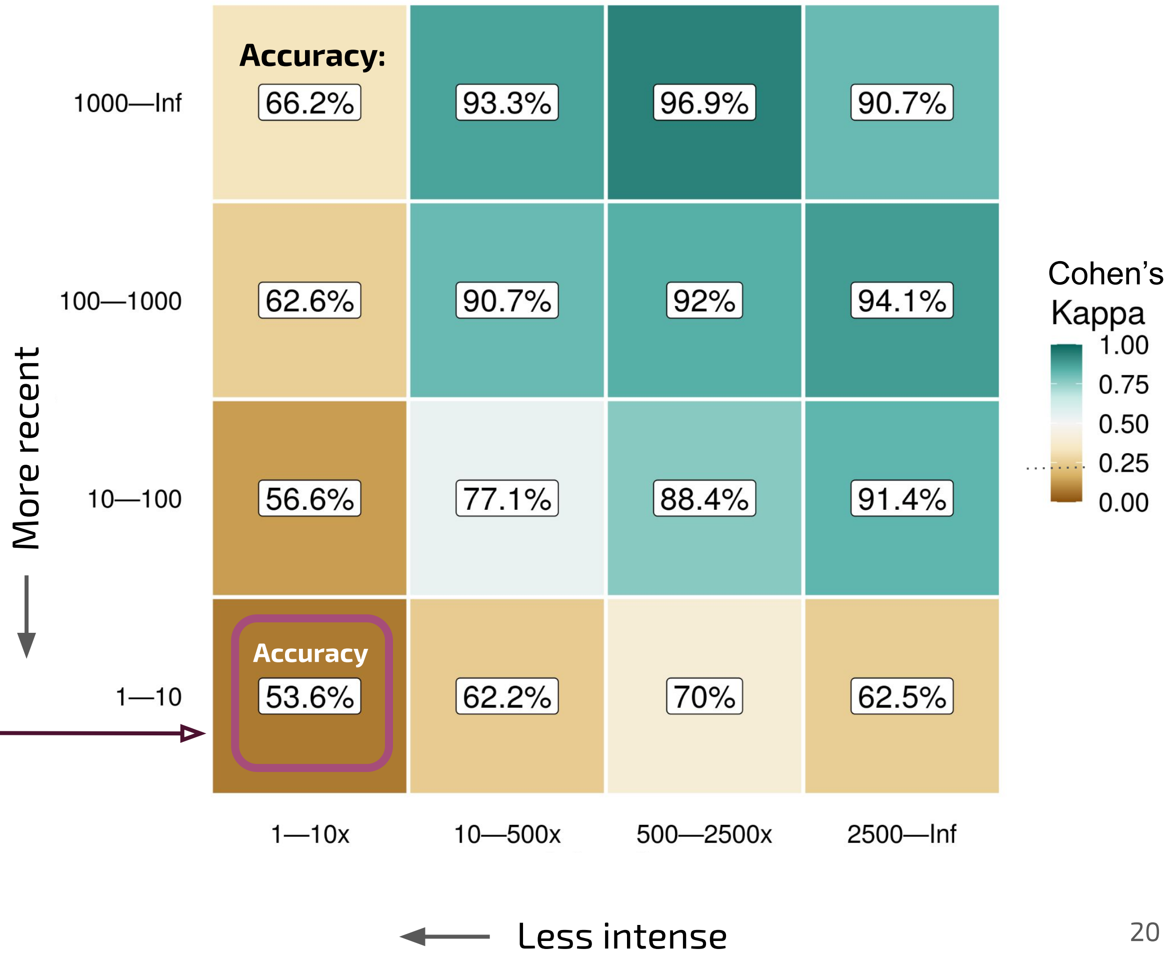
limited power to detect **recent & weak** fragmentation

Detection power



more detection **power** in general

still **limited** power



	1—10x	10—500x	500—2500x	2500—Inf
1000—Inf	66.2%	93.3%	96.9%	90.7%
100—1000	62.6%	90.7%	92%	94.1%
10—100	56.6%	77.1%	88.4%	91.4%
1—10	53.6%	62.2%	70%	62.5%

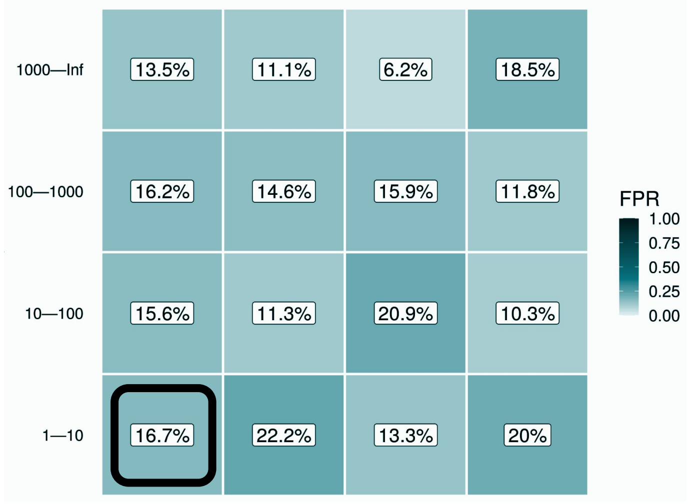
Detection conundrum

We are more likely to **miss** a recent fragmentation that **occurred**

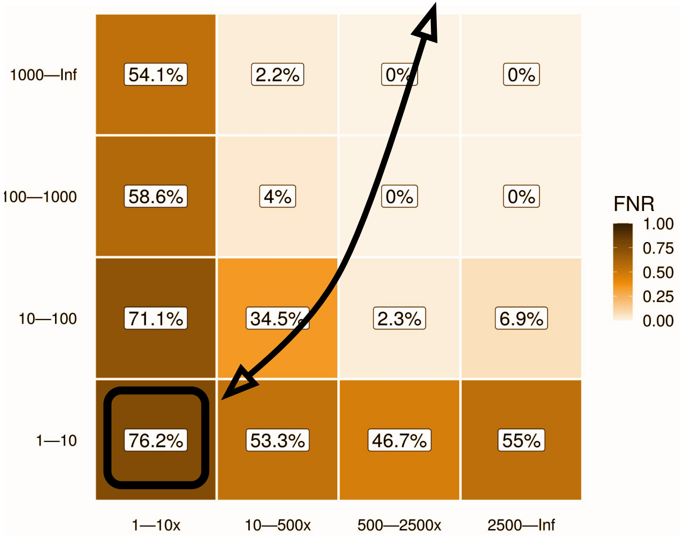
than to **detect** a recent fragmentation when it **did not occur**!

underdetection of recent fragmentations

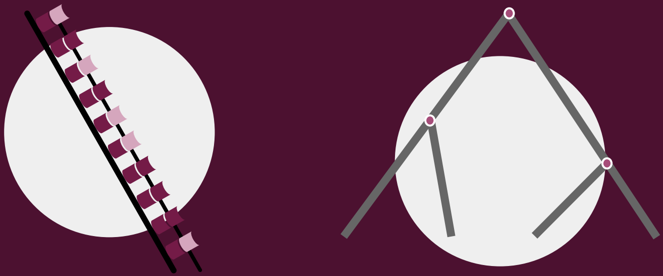
False Positive



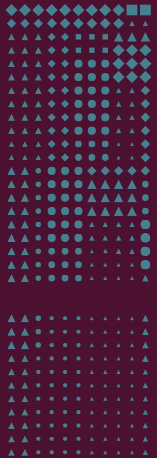
False Negative



Conclusion

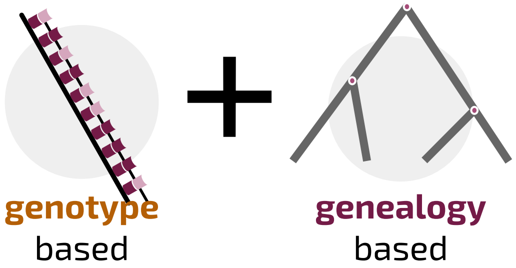


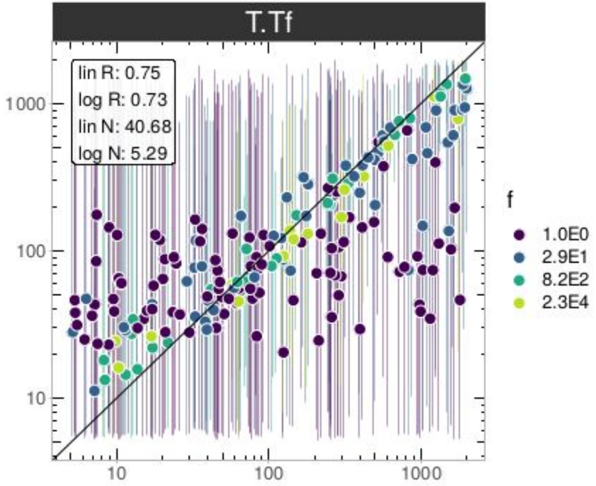
Population genetics may need to be **complemented** with **demographic** studies to *detect* **recent** & **weak** fragmentation



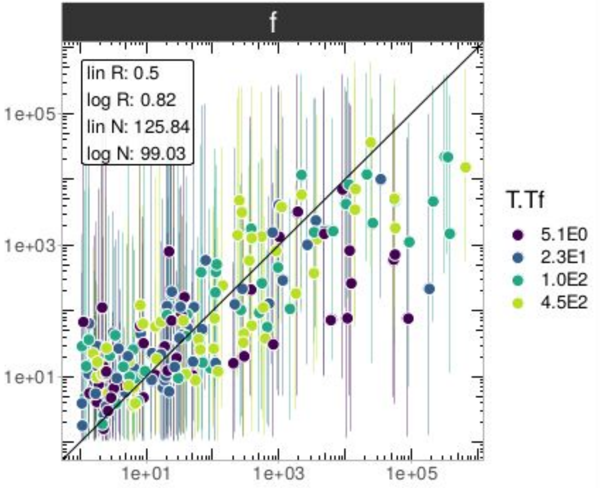
How well can we *characterize* fragmentation?

Estimation power



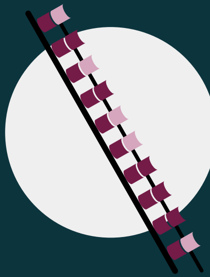


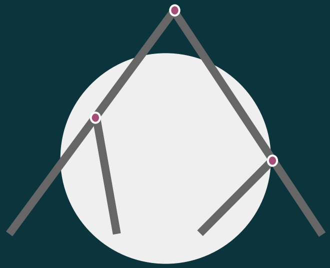
less intense



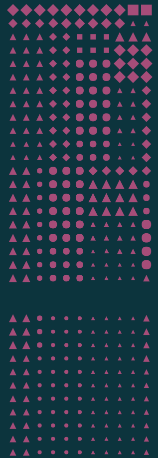
more recent

Conclusion





Genomic statistics **have** power to **estimate the <u>parameters</u> of recent** & **ancient** fragmentation



Take Home Message

Population **gen**etics:

likely to **underestimate** **fra**gmentation **pre**valence

Genomic statistics:

useful to **infer** **fra**gmentation **par**ameters

Thank you for your attention!

Population & Conservation Genetics Group

Lisbon & Toulouse

Inês Carvalho

Hugo Lainé

Gabriele Sgarlata

Beatriz Mourato

Lounes Chikhi

Filipa Borges

Barbara Parreira

Tânia Minhós

Barbara Le Pors

Margarida Henrique

Rémi Tournebize

Ravi Vishwakarma

Marine Ha-Shan

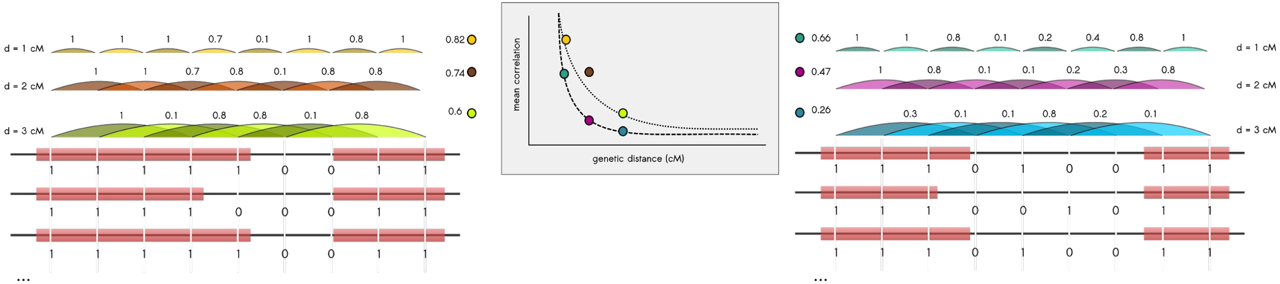
Alexane Jouniaux

Clément Couloigner

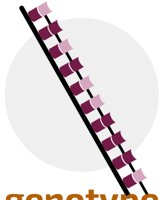
Appendix

ASCEND: Allele Sharing Correlation

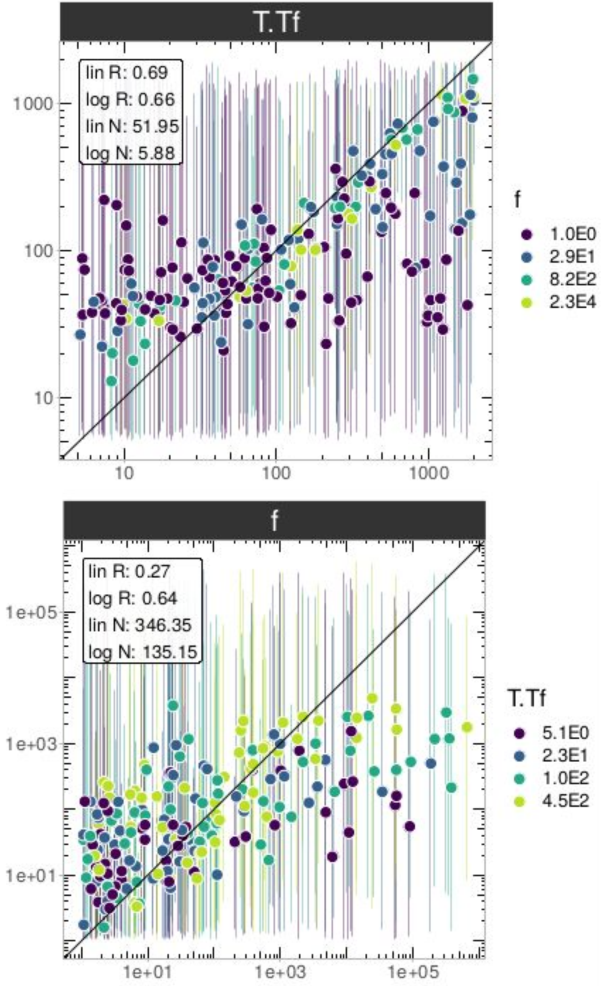
proxy of the *IBD* segment length distribution



Estimation power



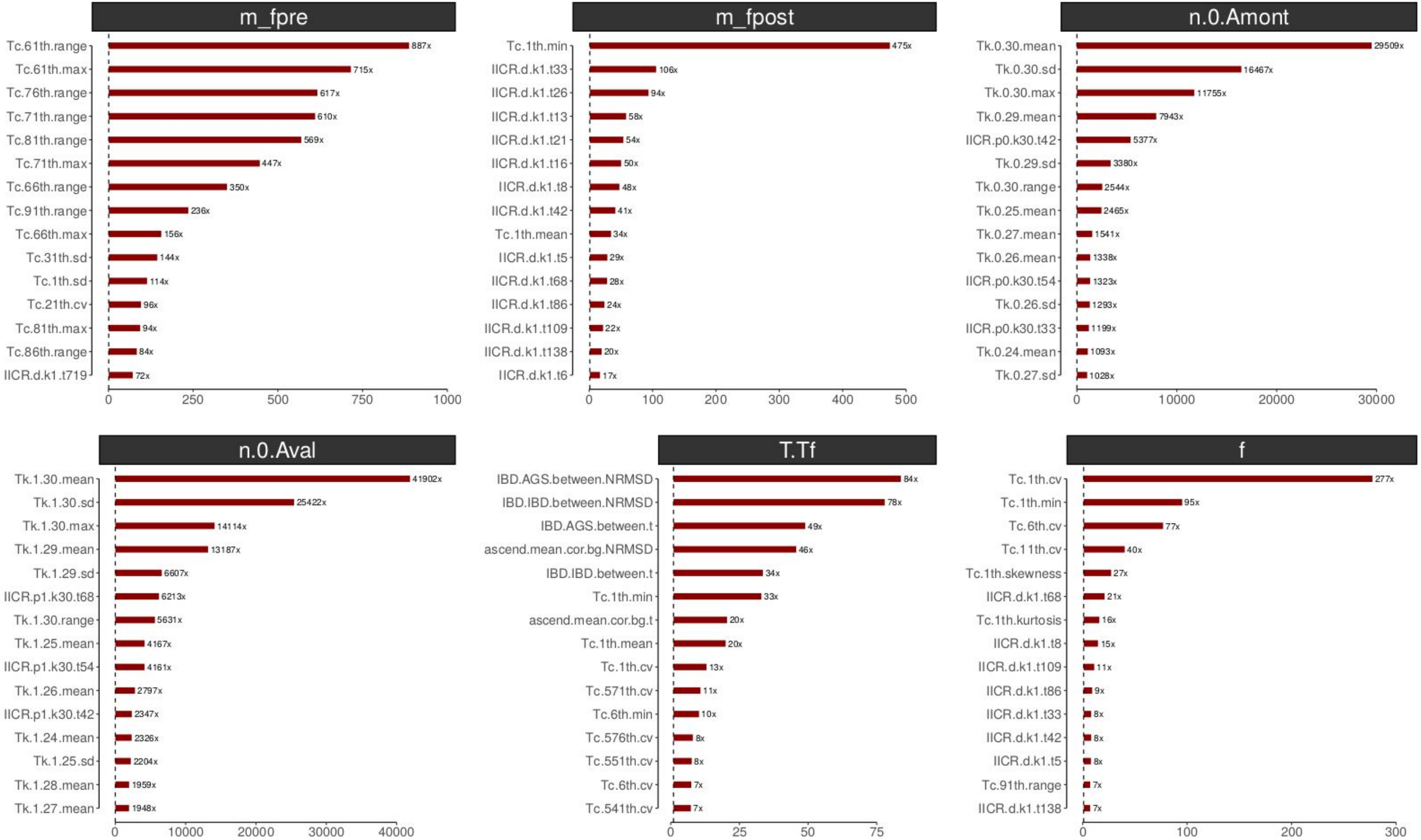
genotype based



less intense

more recent

limited power to estimate
recent T_f
& **extreme** $\boldsymbol{f}$


m_fpre
Tc.61th.range 887x
Tc.61th.max 715x
Tc.76th.range 617x
Tc.71th.range 610x
Tc.81th.range 569x
Tc.71th.max 447x
Tc.66th.range 350x
Tc.91th.range 236x
Tc.66th.max 156x
Tc.31th.sd 144x
Tc.1th.sd 114x
Tc.21th.cv 96x
Tc.81th.max 94x
Tc.86th.range 84x
IICR.d.k1.t719 72x
0
250
500
750
1000
m_fpost
Tc.1th.min 475x
IICR.d.k1.t33 106x
IICR.d.k1.t26 94x
IICR.d.k1.t13 58x
IICR.d.k1.t21 54x
IICR.d.k1.t16 50x
IICR.d.k1.t8 48x
IICR.d.k1.t42 41x
Tc.1th.mean 34x
IICR.d.k1.t5 29x
IICR.d.k1.t68 28x
IICR.d.k1.t86 24x
IICR.d.k1.t109 22x
IICR.d.k1.t138 20x
IICR.d.k1.t6 17x
0
100
200
300
400
500
n.0.Amont
Tk.0.30.mean 29509x
Tk.0.30.sd 16467x
Tk.0.30.max 11755x
Tk.0.29.mean 7943x
IICR.p0.k30.t42 5377x
Tk.0.29.sd 3380x
Tk.0.30.range 2544x
Tk.0.25.mean 2465x
Tk.0.27.mean 1541x
Tk.0.26.mean 1338x
IICR.p0.k30.t54 1323x
Tk.0.26.sd 1293x
IICR.p0.k30.t33 1199x
Tk.0.24.mean 1093x
Tk.0.27.sd 1028x
0
10000
20000
30000
n.0.Aval
Tk.1.30.mean 41902x
Tk.1.30.sd 25422x
Tk.1.30.max 14114x
Tk.1.29.mean 13187x
Tk.1.29.sd 6607x
IICR.p1.k30.t68 6213x
Tk.1.30.range 5631x
Tk.1.25.mean 4167x
IICR.p1.k30.t54 4161x
Tk.1.26.mean 2797x
IICR.p1.k30.t42 2347x
Tk.1.24.mean 2326x
Tk.1.25.sd 2204x
Tk.1.28.mean 1959x
Tk.1.27.mean 1948x
0
10000
20000
30000
40000
T.Tf
IBD.AGS.between.NRMSD 84x
IBD.IBD.between.NRMSD 78x
IBD.AGS.between.t 49x
ascend.mean.cor.bg.NRMSD 46x
IBD.IBD.between.t 34x
Tc.1th.min 33x
ascend.mean.cor.bg.t 20x
Tc.1th.mean 20x
Tc.1th.cv 13x
Tc.571th.cv 11x
Tc.6th.min 10x
Tc.576th.cv 8x
Tc.551th.cv 8x
Tc.6th.cv 7x
Tc.541th.cv 7x
0
25
50
75
f
Tc.1th.cv 277x
Tc.1th.min 95x
Tc.6th.cv 77x
Tc.11th.cv 40x
Tc.1th.skewness 27x
IICR.d.k1.t68 21x
Tc.1th.kurtosis 16x
IICR.d.k1.t8 15x
IICR.d.k1.t109 11x
IICR.d.k1.t86 9x
IICR.d.k1.t33 8x
IICR.d.k1.t42 8x
IICR.d.k1.t5 8x
Tc.91th.range 7x
IICR.d.k1.t138 7x
0
100
200
300