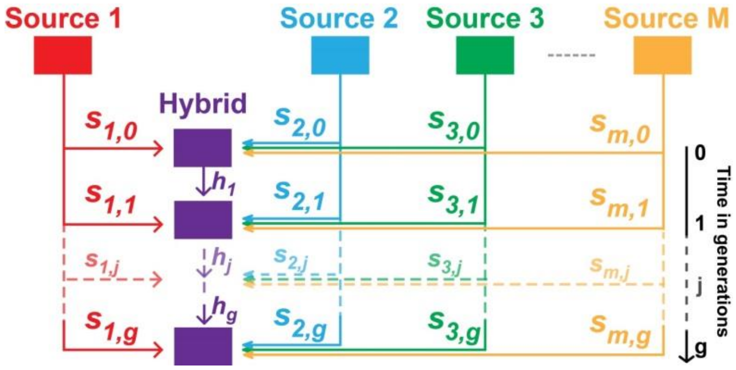
Reconstructing highly complex admixture histories using genetic data

Paul Verdu

Lab: UMR7206 Eco-anthropologie
Institution: CNRS – MNHN - Université Paris Cité

Rencontres de la Chaire MMB, Ecole Polytechnique
25 Avril 2022

$$H_{\mathrm{i,g}} = \begin{cases} 1 & \text{if } Y = \mathrm{S_iS_i}\text{, with } P[Y = \mathrm{S_iS_i}] = s^2_{\mathrm{i,g-1}} \\ \dfrac{H_{\mathrm{i,g-1}} + 1}{2} & \text{if } Y = \mathrm{S_iH}\text{, with } P[Y = \mathrm{S_iH}] = 2s_{\mathrm{i,g-1}}h_{\mathrm{g-1}} \\ \dfrac{1}{2} & \text{if } Y = \mathrm{S_iS_j}\text{, with } P\left[Y = \mathrm{S_iS_j}\right] = 2s_{\mathrm{i,g-1}}s_{\mathrm{j,g-1}} \\ \dfrac{H^{(1)}_{\mathrm{i,g-1}} + H^{(2)}_{\mathrm{i,g-1}}}{2} & \text{if } Y = \mathrm{HH}\text{, with } P[Y = \mathrm{HH}] = h^2_{\mathrm{g-1}} \\ \dfrac{H_{\mathrm{i,g-1}}}{2} & \text{if } Y = \mathrm{S_jH}\text{, with } P\left[Y = \mathrm{S_jH}\right] = 2s_{\mathrm{j,g-1}}h_{\mathrm{g-1}} \\ 0 & \text{if } Y = \mathrm{S_jS_j}\text{, with } P\left[Y = \mathrm{S_jS_j}\right] = s^2_{\mathrm{j,g-1}} \\ 0 & \text{if } Y = \mathrm{S_jS_l}\text{, with } P\left[Y = \mathrm{S_jS_l}\right] = 2s_{\mathrm{j,g-1}}s_{\mathrm{l,g-1}}. \end{cases}$$



ANR

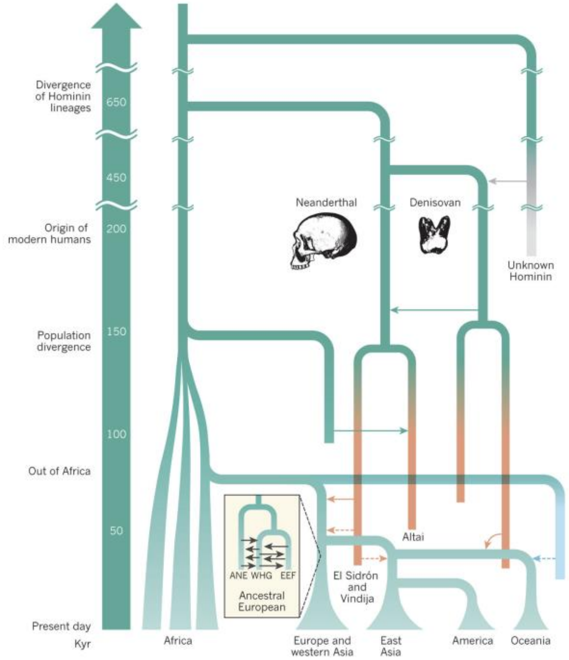
cnrs

MUSÉE DE L'HOMME

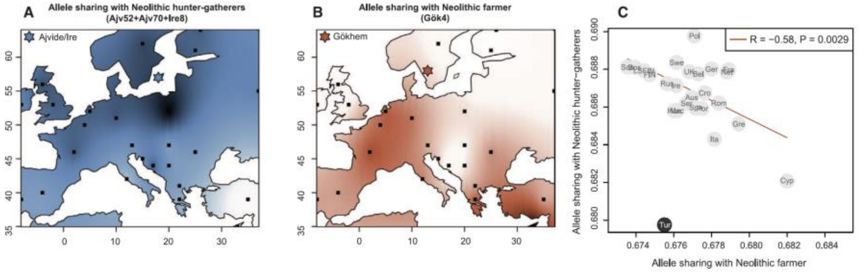
Université Paris Cité

Admixture is ubiquitous in evolutionary history, through time and space

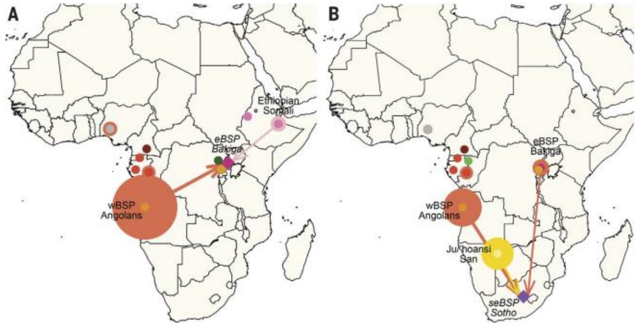
e.g. *Homo sapiens*



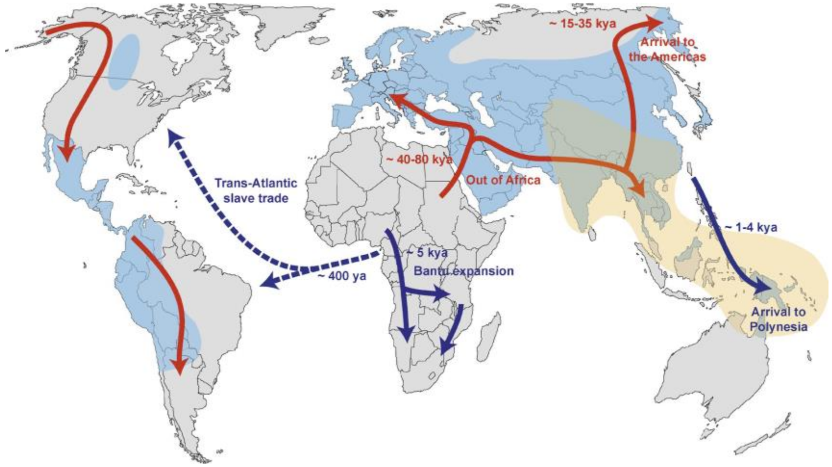
e.g. Nielsen et al., Nature 2018



Skoglund et al., Science 2012

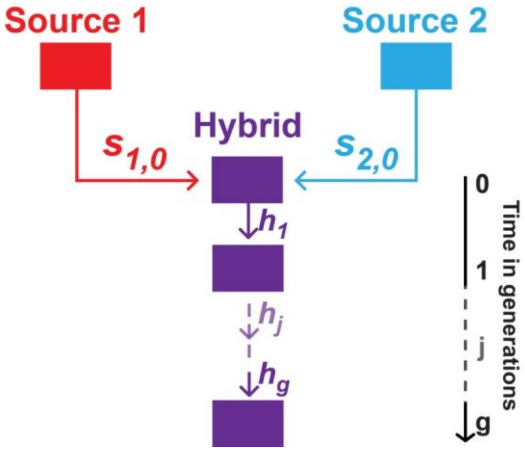


Patin et al., Science 2017

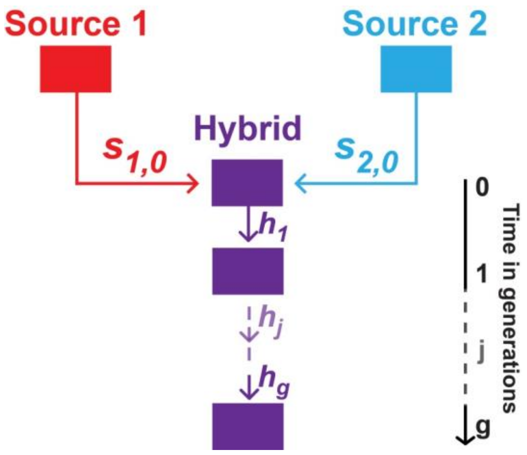


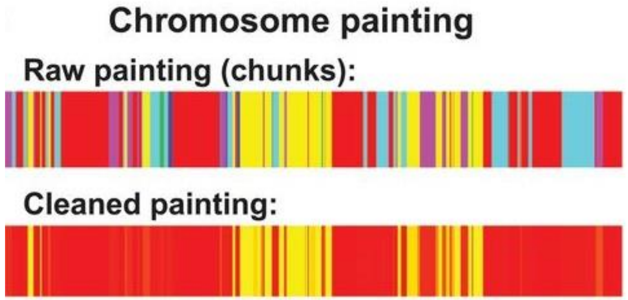
Quach and Quintana-Murci, JEM 2017

Inferring admixture history from genetic data

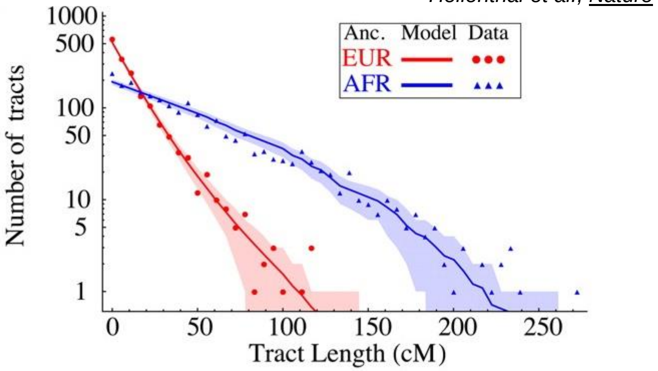


Inferring admixture history from genetic data

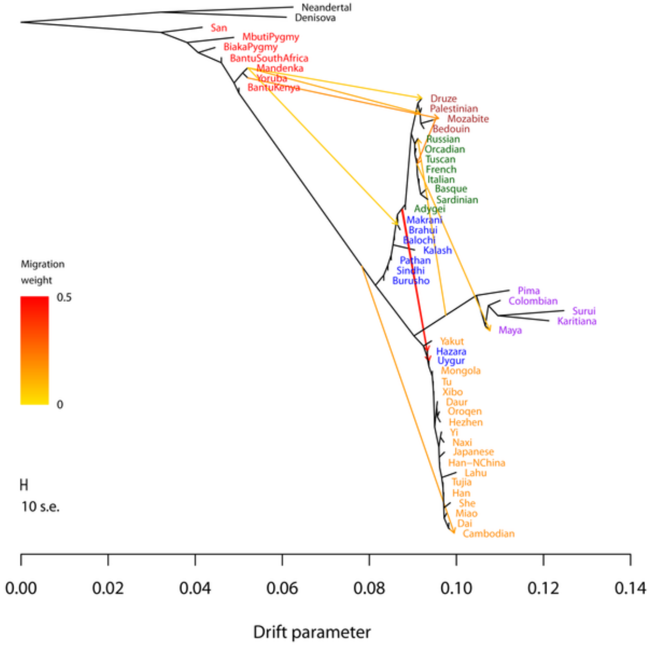




Hellenthal et al., Nature 2014



Gravel, Genetics 2012



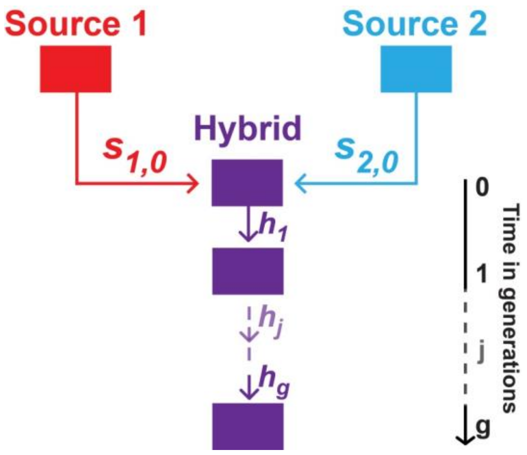
Pritchard et al., Plos Gen 2012

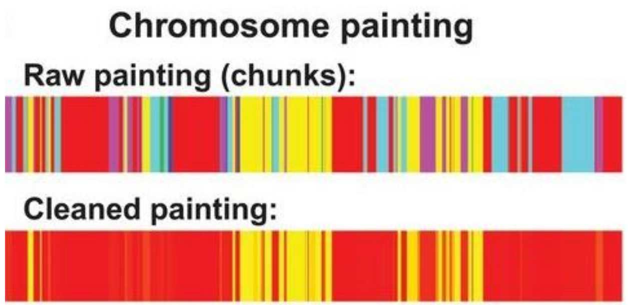
Maximum likelihood approaches relying on:

► Admixture Linkage-Disequilibrium distributions - e.g. *TRACTS* (Gravel 2012), *GLOBETROTTER* (Hellenthal et al. 2014)

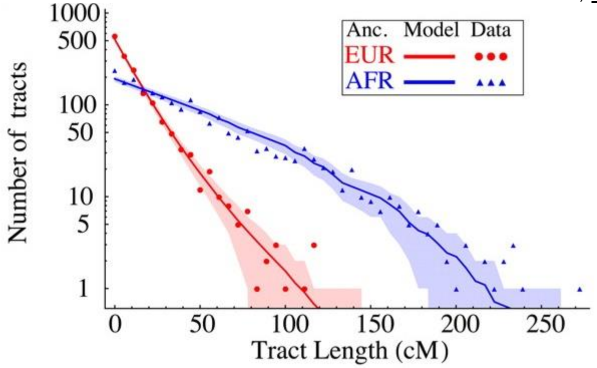
► Moments of allelic frequency spectrum divergences - e.g. *M/ALDER* (Loh et al. 2013), *TreeMix* (Pritchard et al. 2012)

Inferring admixture history from genetic data

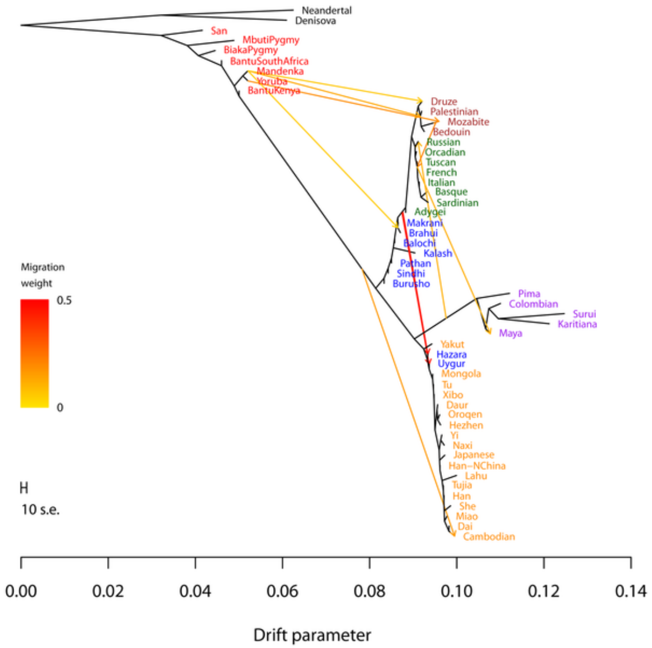




Hellenthal et al., Nature 2014



Gravel, Genetics 2012



Pritchard et al., Plos Gen 2012

Maximum likelihood approaches relying on:

► Admixture Linkage-Disequilibrium distributions - e.g. *TRACTS* (Gravel 2012), *GLOBETROTTER* (Hellenthal et al. 2014)

► Moments of allelic frequency spectrum divergences - e.g. *M/ALDER* (Loh et al. 2013), *TreeMix* (Pritchard et al. 2012)

Limited by:

► **“Simple” admixture models: one or two admixture pulses per source population**

► **No formal model-choice**

► **Massive genomic data** and **accurate phasing** for admixture-LD methods

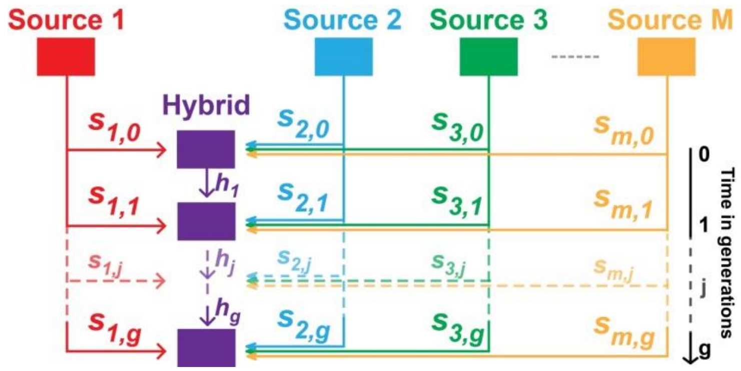
Inferring admixture history from genetic data

ML-inference methods cannot operate

$$L(ModelParams|Data_{obs}) \propto P(Data_{obs}|ModelParams)\ P(ModelParams)$$

for highly complex admixture models where likelihoods cannot be written

for highly complex admixture models where likelihoods are intractable



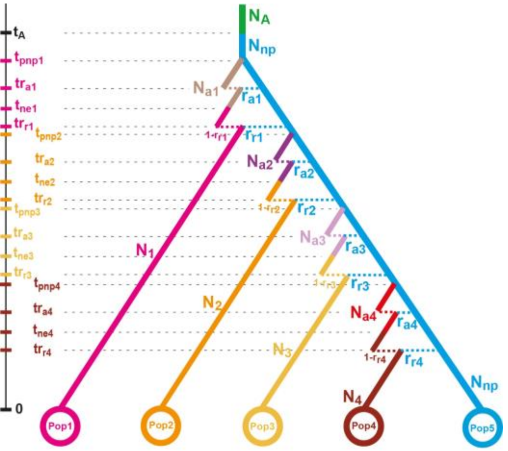
Approximate Bayesian Computation (Tavaré et al. 1997) may represent an alternative

$$L(ModelParams|Data_{obs}) \propto P(Data_{obs}|ModelParams)\ P(ModelParams)$$

$\wr$ Approximation

$$L(ModelParams|SumStat_{obs}) \propto P(SumStat_{obs}|ModelParams)\ P(ModelParams)$$

Approximate Bayesian Computation demographic inference

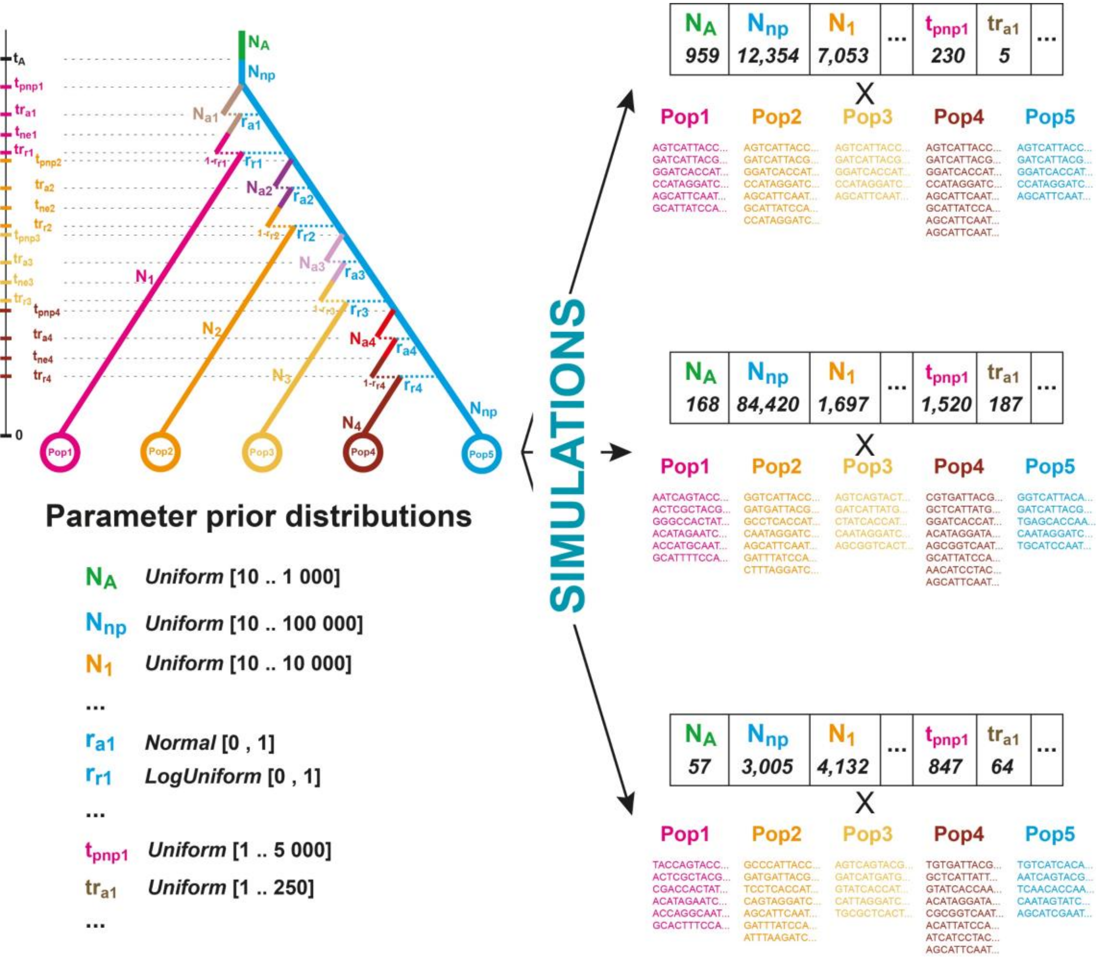


Parameter prior distributions

N_A	*Uniform* [10 .. 1 000]
N_{np}	*Uniform* [10 .. 100 000]
N_1	*Uniform* [10 .. 10 000]
...	
r_{a1}	*Normal* [0 , 1]
r_{r1}	*LogUniform* [0 , 1]
...	
t_{pnp1}	*Uniform* [1 .. 5 000]
tr_{a1}	*Uniform* [1 .. 250]
...	

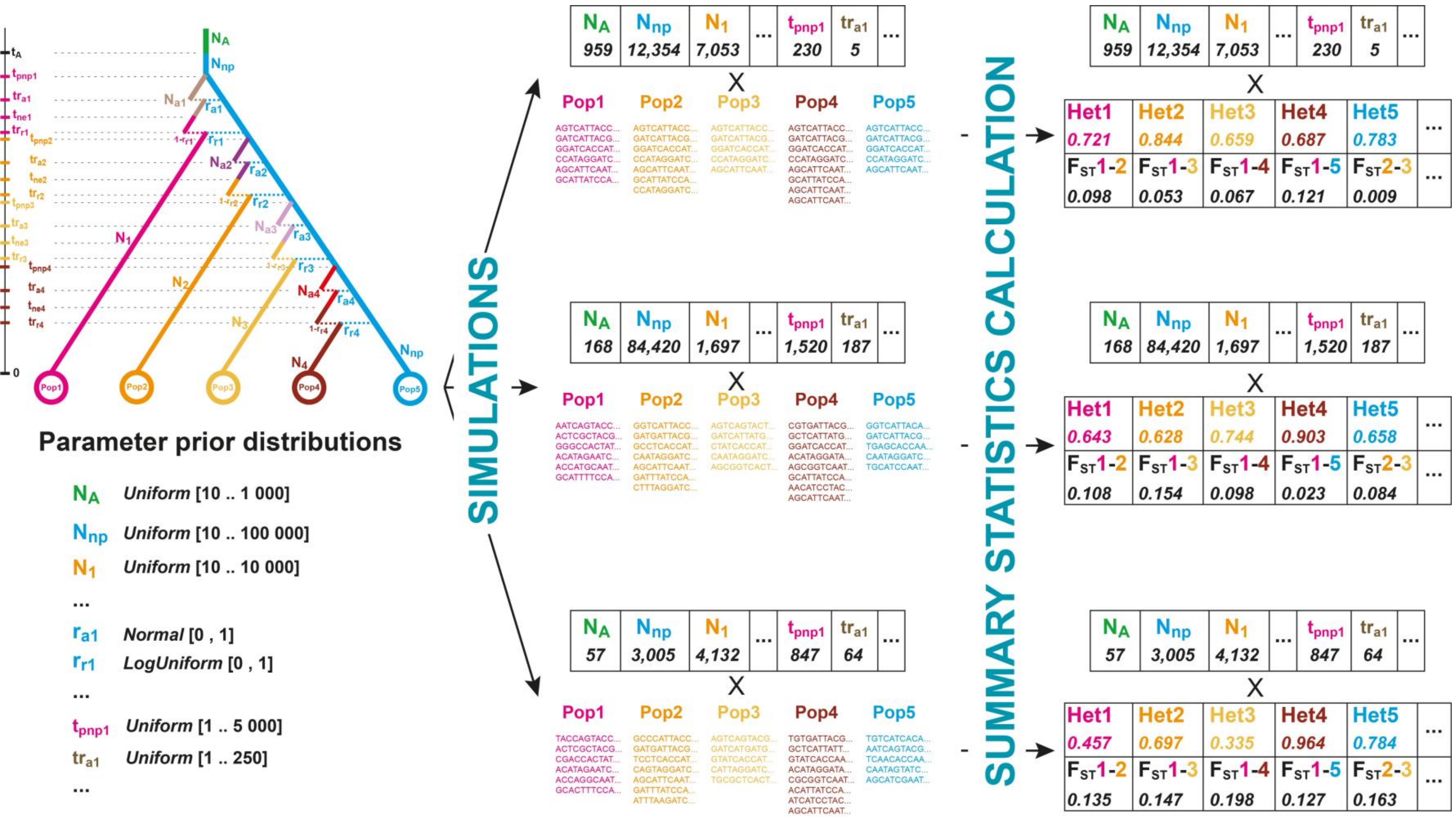
Tavaré et al., Plos Gen 1997
Pritchard et al., Genetics 2000
Beaumont et al., Genetics 2002

Approximate Bayesian Computation demographic inference


t_A
t_{pnp1}
tr_{a1}
t_{ne1}
tr_{r1}
t_{pnp2}
tr_{a2}
t_{ne2}
tr_{r2}
t_{pnp3}
tr_{a3}
t_{ne3}
tr_{r3}
t_{pnp4}
tr_{a4}
t_{ne4}
tr_{r4}
0
N_A
N_{np}
N_{a1}
r_{a1}
$1-r_{r1}$
r_{r1}
N_{a2}
r_{a2}
$1-r_{r2}$
r_{r2}
N_{a3}
r_{a3}
$1-r_{r3}$
r_{r3}
N_{a4}
r_{a4}
$1-r_{r4}$
r_{r4}
N_1
N_2
N_3
N_4
N_{np}
Pop1
Pop2
Pop3
Pop4
Pop5
Parameter prior distributions
N_A Uniform [10 .. 1 000]
N_{np} Uniform [10 .. 100 000]
N_1 Uniform [10 .. 10 000]
...
r_{a1} Normal [0 , 1]
r_{r1} LogUniform [0 , 1]
...
t_{pnp1} Uniform [1 .. 5 000]
tr_{a1} Uniform [1 .. 250]
...
SIMULATIONS
N_A | N_{np} | N_1 | ... | t_{pnp1} | tr_{a1} | ...
959 | 12,354 | 7,053 | | 230 | 5 |
X
Pop1: AGTCATTACC... GATCATTACG... GGATCACCAT... CCATAGGATC... AGCATTCAAT... GCATTATCCA...
Pop2: AGTCATTACC... GATCATTACG... GGATCACCAT... CCATAGGATC... AGCATTCAAT... GCATTATCCA... CCATAGGATC...
Pop3: AGTCATTACC... GATCATTACG... GGATCACCAT... CCATAGGATC... AGCATTCAAT...
Pop4: AGTCATTACC... GATCATTACG... GGATCACCAT... CCATAGGATC... AGCATTCAAT... GCATTATCCA... AGCATTCAAT... AGCATTCAAT...
Pop5: AGTCATTACC... GATCATTACG... GGATCACCAT... CCATAGGATC... AGCATTCAAT...
N_A | N_{np} | N_1 | ... | t_{pnp1} | tr_{a1} | ...
168 | 84,420 | 1,697 | | 1,520 | 187 |
X
Pop1: AATCAGTACC... ACTCGCTACG... GGGCCACTAT... ACATAGAATC... ACCATGCAAT... GCATTTTCCA...
Pop2: GGTCATTACC... GATGATTACG... GCCTCACCAT... CAATAGGATC... AGCATTCAAT... GATTTATCCA... CTTTAGGATC...
Pop3: AGTCAGTACT... GATCATTATG... CTATCACCAT... CAATAGGATC... AGCGGTCACT...
Pop4: CGTGATTACG... GCTCATTATG... GGATCACCAT... ACATAGGATA... AGCGGTCAAT... GCATTATCCA... AACATCCTAC... AGCATTCAAT...
Pop5: GGTCATTACA... GATCATTACG... TGAGCACCAA... CAATAGGATC... TGCATCCAAT...
N_A | N_{np} | N_1 | ... | t_{pnp1} | tr_{a1} | ...
57 | 3,005 | 4,132 | | 847 | 64 |
X
Pop1: TACCAGTACC... ACTCGCTACG... CGACCACTAT... ACATAGAATC... ACCAGGCAAT... GCACTTTCCA...
Pop2: GCCCATTACC... GATGATTACG... TCCTCACCAT... CAGTAGGATC... AGCATTCAAT... GATTTATCCA... ATTTAAGATC...
Pop3: AGTCAGTACG... GATCATGATG... GTATCACCAT... CATTAGGATC... TGCGCTCACT...
Pop4: TGTGATTACG... GCTCATTATT... GTATCACCAA... ACATAGGATA... CGCGGTCAAT... ACATTATCCA... ATCATCCTAC... AGCATTCAAT...
Pop5: TGTCATCACA... AATCAGTACG... TCAACACCAA... CAATAGTATC... AGCATCGAAT...

Tavaré et al., Plos Gen 1997
Pritchard et al., Genetics 2000
Beaumont et al., Genetics 2002

Approximate Bayesian Computation demographic inference



Parameter prior distributions

N_A	*Uniform* [10 .. 1 000]
N_{np}	*Uniform* [10 .. 100 000]
N_1	*Uniform* [10 .. 10 000]
...	
r_{a1}	*Normal* [0 , 1]
r_{r1}	*LogUniform* [0 , 1]
...	
t_{pnp1}	*Uniform* [1 .. 5 000]
tr_{a1}	*Uniform* [1 .. 250]
...	

N_A	N_{np}	N_1	...	t_{pnp1}	tr_{a1}	...
959	*12,354*	*7,053*		*230*	*5*	

Pop1	Pop2	Pop3	Pop4	Pop5
AGTCATTACC...	AGTCATTACC...	AGTCATTACC...	AGTCATTACC...	AGTCATTACC...
GATCATTACG...	GATCATTACG...	GATCATTACG...	GATCATTACG...	GATCATTACG...
GGATCACCAT...	GGATCACCAT...	GGATCACCAT...	GGATCACCAT...	GGATCACCAT...
CCATAGGATC...	CCATAGGATC...	CCATAGGATC...	CCATAGGATC...	CCATAGGATC...
AGCATTCAAT...	AGCATTCAAT...	AGCATTCAAT...	AGCATTCAAT...	AGCATTCAAT...
GCATTATCCA...	GCATTATCCA...		GCATTATCCA...	
	CCATAGGATC...		AGCATTCAAT...	
			AGCATTCAAT...	

Het1	Het2	Het3	Het4	Het5	...
0.721	*0.844*	*0.659*	*0.687*	*0.783*	
F_{ST}1-2	F_{ST}1-3	F_{ST}1-4	F_{ST}1-5	F_{ST}2-3	...
0.098	*0.053*	*0.067*	*0.121*	*0.009*	

N_A	N_{np}	N_1	...	t_{pnp1}	tr_{a1}	...
168	*84,420*	*1,697*		*1,520*	*187*	

Pop1	Pop2	Pop3	Pop4	Pop5
AATCAGTACC...	GGTCATTACC...	AGTCAGTACT...	CGTGATTACG...	GGTCATTACA...
ACTCGCTACG...	GATGATTACG...	GATCATTATG...	GCTCATTATG...	GATCATTACG...
GGGCCACTAT...	GCCTCACCAT...	CTATCACCAT...	GGATCACCAT...	TGAGCACCAA...
ACATAGAATC...	CAATAGGATC...	CAATAGGATC...	ACATAGGATA...	CAATAGGATC...
ACCATGCAAT...	AGCATTCAAT...	AGCGGTCACT...	AGCGGTCAAT...	TGCATCCAAT...
GCATTTTCCA...	GATTTATCCA...		GCATTATCCA...	
	CTTTAGGATC...		AACATCCTAC...	
			AGCATTCAAT...	

Het1	Het2	Het3	Het4	Het5	...
0.643	*0.628*	*0.744*	*0.903*	*0.658*	
F_{ST}1-2	F_{ST}1-3	F_{ST}1-4	F_{ST}1-5	F_{ST}2-3	...
0.108	*0.154*	*0.098*	*0.023*	*0.084*	

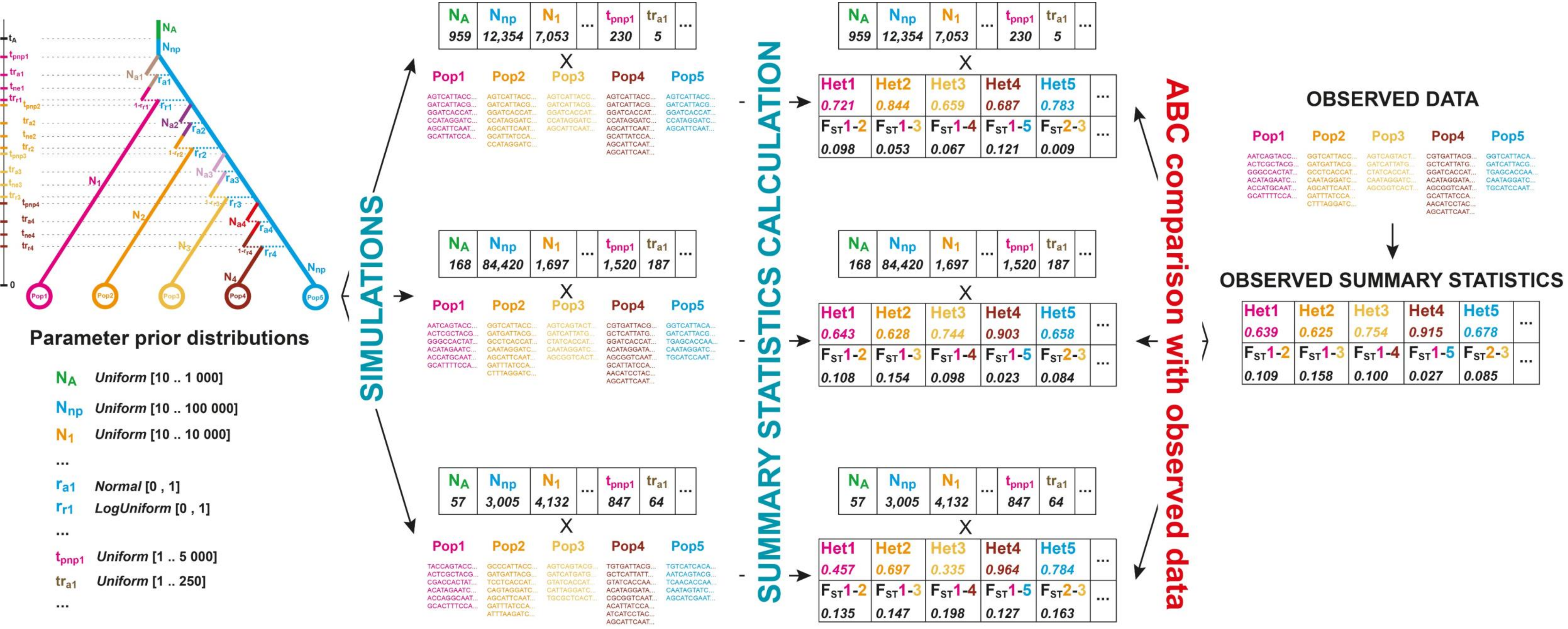
N_A	N_{np}	N_1	...	t_{pnp1}	tr_{a1}	...
57	*3,005*	*4,132*		*847*	*64*	

Pop1	Pop2	Pop3	Pop4	Pop5
TACCAGTACC...	GCCCATTACC...	AGTCAGTACG...	TGTGATTACG...	TGTCATCACA...
ACTCGCTACG...	GATGATTACG...	GATCATGATG...	GCTCATTATT...	AATCAGTACG...
CGACCACTAT...	TCCTCACCAT...	GTATCACCAT...	GTATCACCAA...	TCAACACCAA...
ACATAGAATC...	CAGTAGGATC...	CATTAGGATC...	ACATAGGATA...	CAATAGTATC...
ACCAGGCAAT...	AGCATTCAAT...	TGCGCTCACT...	CGCGGTCAAT...	AGCATCGAAT...
GCACTTTCCA...	GATTTATCCA...		ACATTATCCA...	
	ATTTAAGATC...		ATCATCCTAC...	
			AGCATTCAAT...	

Het1	Het2	Het3	Het4	Het5	...
0.457	*0.697*	*0.335*	*0.964*	*0.784*	
F_{ST}1-2	F_{ST}1-3	F_{ST}1-4	F_{ST}1-5	F_{ST}2-3	...
0.135	*0.147*	*0.198*	*0.127*	*0.163*	

Tavaré et al., Plos Gen 1997
Pritchard et al., Genetics 2000
Beaumont et al., Genetics 2002

Approximate Bayesian Computation demographic inference



Parameter prior distributions

N_A	*Uniform* [10 .. 1 000]
N_{np}	*Uniform* [10 .. 100 000]
N_1	*Uniform* [10 .. 10 000]
...	
r_{a1}	*Normal* [0 , 1]
r_{r1}	*LogUniform* [0 , 1]
...	
t_{pnp1}	*Uniform* [1 .. 5 000]
tr_{a1}	*Uniform* [1 .. 250]
...	

N_A	N_{np}	N_1	...	t_{pnp1}	tr_{a1}	...
959	12,354	7,053		230	5	

Het1	Het2	Het3	Het4	Het5	...
0.721	0.844	0.659	0.687	0.783	
F_{ST}1-2	F_{ST}1-3	F_{ST}1-4	F_{ST}1-5	F_{ST}2-3	...
0.098	0.053	0.067	0.121	0.009	

N_A	N_{np}	N_1	...	t_{pnp1}	tr_{a1}	...
168	84,420	1,697		1,520	187	

Het1	Het2	Het3	Het4	Het5	...
0.643	0.628	0.744	0.903	0.658	
F_{ST}1-2	F_{ST}1-3	F_{ST}1-4	F_{ST}1-5	F_{ST}2-3	...
0.108	0.154	0.098	0.023	0.084	

N_A	N_{np}	N_1	...	t_{pnp1}	tr_{a1}	...
57	3,005	4,132		847	64	

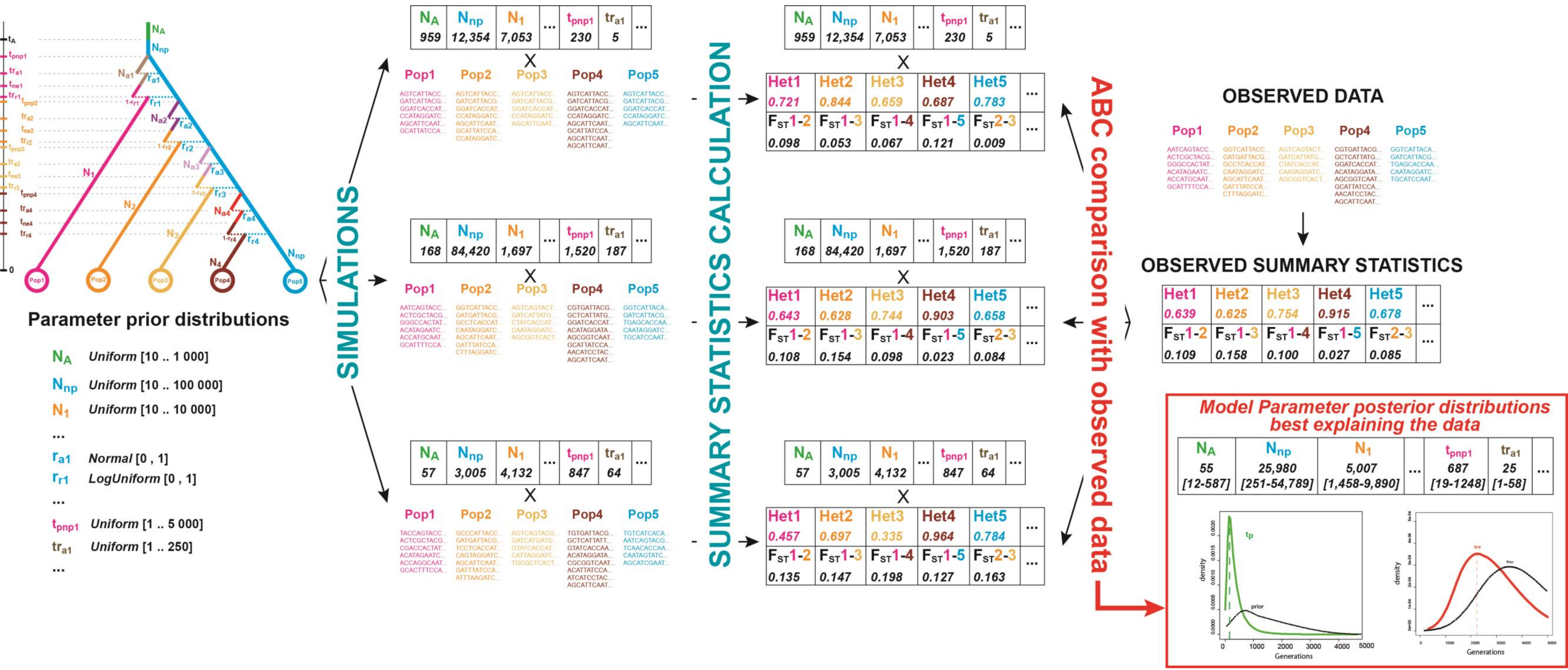
Het1	Het2	Het3	Het4	Het5	...
0.457	0.697	0.335	0.964	0.784	
F_{ST}1-2	F_{ST}1-3	F_{ST}1-4	F_{ST}1-5	F_{ST}2-3	...
0.135	0.147	0.198	0.127	0.163	

OBSERVED SUMMARY STATISTICS

Het1	Het2	Het3	Het4	Het5	...
0.639	0.625	0.754	0.915	0.678	
F_{ST}1-2	F_{ST}1-3	F_{ST}1-4	F_{ST}1-5	F_{ST}2-3	...
0.109	0.158	0.100	0.027	0.085	

Tavaré et al., Plos Gen 1997
Pritchard et al., Genetics 2000
Beaumont et al., Genetics 2002

Approximate Bayesian Computation demographic inference



Parameter prior distributions

N_A	*Uniform* [10 .. 1 000]
N_{np}	*Uniform* [10 .. 100 000]
N_1	*Uniform* [10 .. 10 000]
...	
r_{a1}	*Normal* [0 , 1]
r_{r1}	*LogUniform* [0 , 1]
...	
t_{pnp1}	*Uniform* [1 .. 5 000]
tr_{a1}	*Uniform* [1 .. 250]
...	

N_A	N_{np}	N_1	...	t_{pnp1}	tr_{a1}	...
959	12,354	7,053		230	5	

Het1	Het2	Het3	Het4	Het5	...
0.721	0.844	0.659	0.687	0.783	
F_{ST}1-2	F_{ST}1-3	F_{ST}1-4	F_{ST}1-5	F_{ST}2-3	...
0.098	0.053	0.067	0.121	0.009	

N_A	N_{np}	N_1	...	t_{pnp1}	tr_{a1}	...
168	84,420	1,697		1,520	187	

Het1	Het2	Het3	Het4	Het5	...
0.643	0.628	0.744	0.903	0.658	
F_{ST}1-2	F_{ST}1-3	F_{ST}1-4	F_{ST}1-5	F_{ST}2-3	...
0.108	0.154	0.098	0.023	0.084	

N_A	N_{np}	N_1	...	t_{pnp1}	tr_{a1}	...
57	3,005	4,132		847	64	

Het1	Het2	Het3	Het4	Het5	...
0.457	0.697	0.335	0.964	0.784	
F_{ST}1-2	F_{ST}1-3	F_{ST}1-4	F_{ST}1-5	F_{ST}2-3	...
0.135	0.147	0.198	0.127	0.163	

OBSERVED SUMMARY STATISTICS

Het1	Het2	Het3	Het4	Het5	...
0.639	0.625	0.754	0.915	0.678	
F_{ST}1-2	F_{ST}1-3	F_{ST}1-4	F_{ST}1-5	F_{ST}2-3	...
0.109	0.158	0.100	0.027	0.085	

Model Parameter posterior distributions best explaining the data

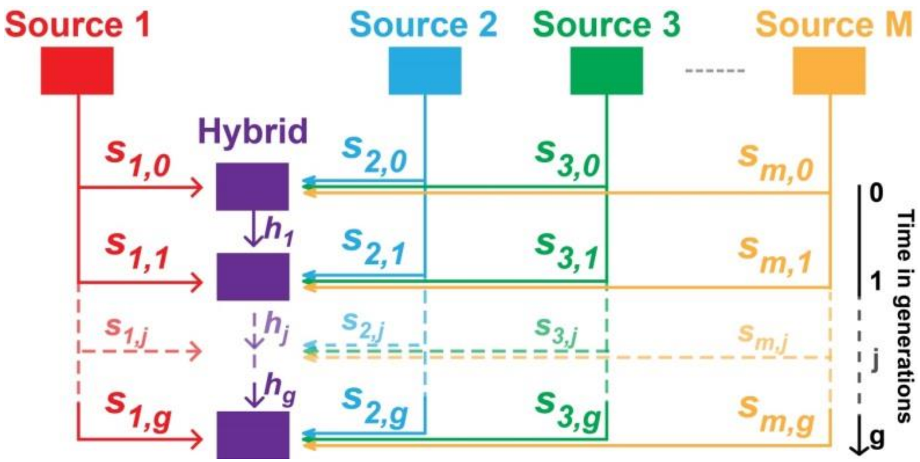
N_A	N_{np}	N_1	...	t_{pnp1}	tr_{a1}	...
55 [12-587]	25,980 [251-54,789]	5,007 [1,458-9,890]		687 [19-1248]	25 [1-58]	

Tavaré et al., Plos Gen 1997
Pritchard et al., Genetics 2000
Beaumont et al., Genetics 2002

Approximate Bayesian Computation for complex admixture history

Approximate Bayesian Computation (Tavaré et al. 1997) represent an alternative ?

$$L(ModelParams|SumStat_{obs}) \propto P(SumStat_{obs}|ModelParams)\ P(ModelParams)$$



If <u>simulations are feasible</u>

If <u>summary-statistics are informative</u> about model-parameters,

► ABC inferences may be successful for the reconstruction of complex admixture histories from genetic data

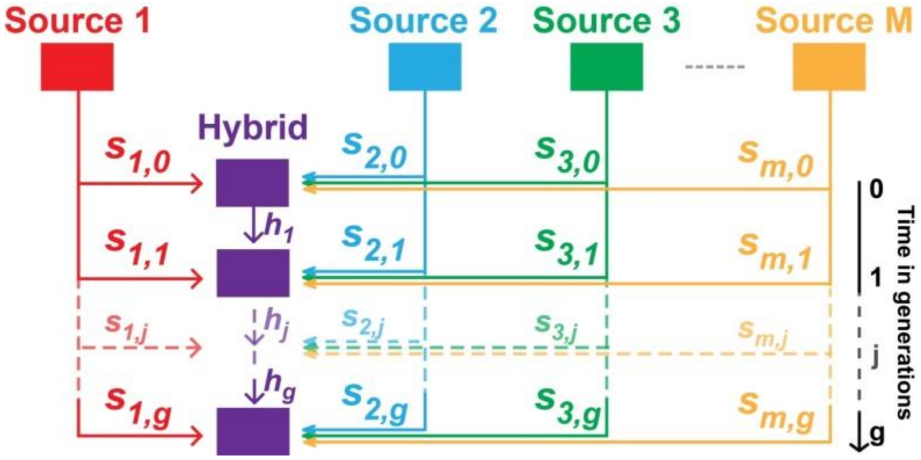
Tavaré et al., <u>Plos Gen</u> 1997
Pritchard et al., <u>Genetics</u> 2000
Beaumont et al., <u>Genetics</u> 2002

Complex admixture histories reconstructed with *MetHis*-ABC

If simulations are feasible…

Simulating complex admixture histories under the coalescent is often not trivial:

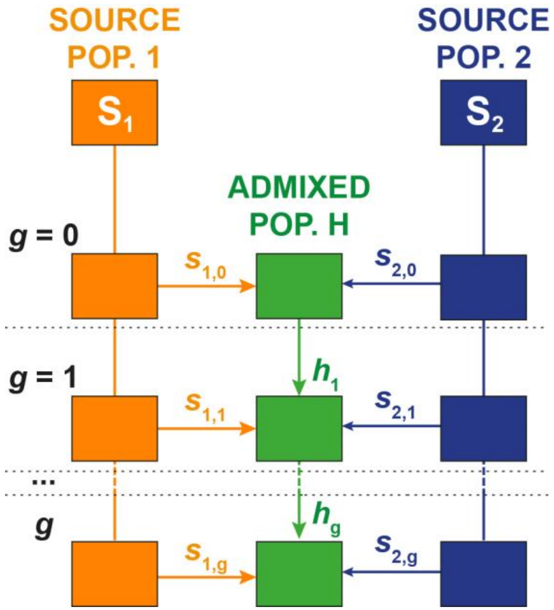
Different pedigree for each independent locus instead of a single pedigree having, in reality, produced all observed gene genealogies (see Wakeley et al. 2012)



MetHis

genetic data simulator under complex admixture models

MetHis: genetic data simulator under complex admixture histories



$\forall\, m \in \{1,..,M\}, s_{m,0} \in [0,1]$ such that

$$\sum_{m=1}^{M} s_{m,0} = 1$$

$\forall\, m \in \{1,..,M\}, \forall\, j \in \{1,..,g\},$
h_j and $s_{m,j} \in [0,1]$ such that

$$h_j + \sum_{m=1}^{M} s_{m,j} = 1$$

MetHis simulates a **random-matting admixed population of diploid size N_g** at generation *g,* **forward-in-time centered on individuals**

MetHis simulates any number of **autosomal independent** genetic loci (SNPs or microsatellites)

1. For each N_{g+1} individuals in population H at generation g+1:
Draw randomly parents in populations S1, S2, and H in proportions $s_{1,g}$, $s_{2,g}$ and h_g

2. Create haploid gametes for each parent by randomly drawing alleles at each loci

3. Pair gametes to create a new individual in population H, avoiding selfing

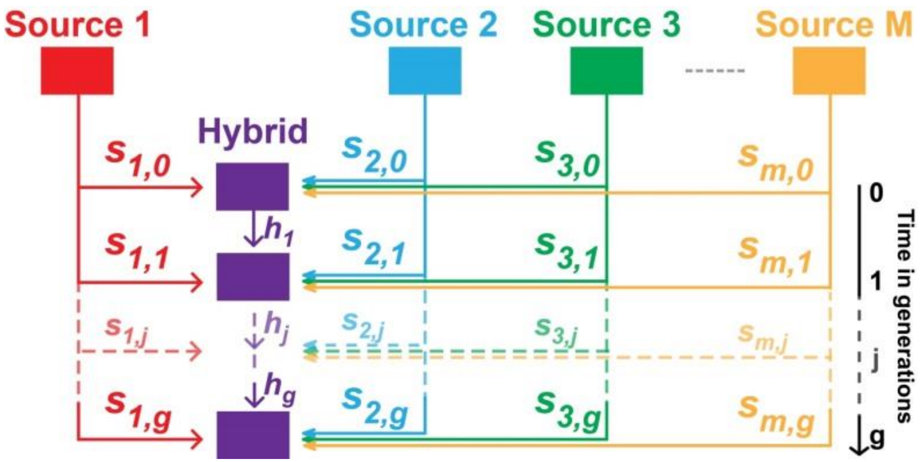
MetHis simulates a mutation model for microsatellites (GSMM with in/del)

Fortes-Lima et al., Mol Ecol Res 2021

Approximate Bayesian Computation for complex admixture history

Approximate Bayesian Computation (Tavaré et al. 1997) represent an alternative ?

$$L(ModelParams|\color{red}{SumStat_{obs}}) \propto P(\color{red}{SumStat_{obs}}|ModelParams)\ P(ModelParams)$$



If simulations are feasible

If **summary-statistics are informative** about model-parameters,

► ABC inferences may be successful for the reconstruction of complex admixture histories from genetic data

Tavaré et al., Plos Gen 1997
Pritchard et al., Genetics 2000
Beaumont et al., Genetics 2002

MetHis: summary-statistics calculator for ABC inference

If <u>summary-statistics are informative</u> about model-parameters…

“Classical” population genetics summary-statistics

Expected Heterozygosity (Nei, 1978)

Inbreeding coefficient $\boldsymbol{F}$ (Danecek et al. 2011)

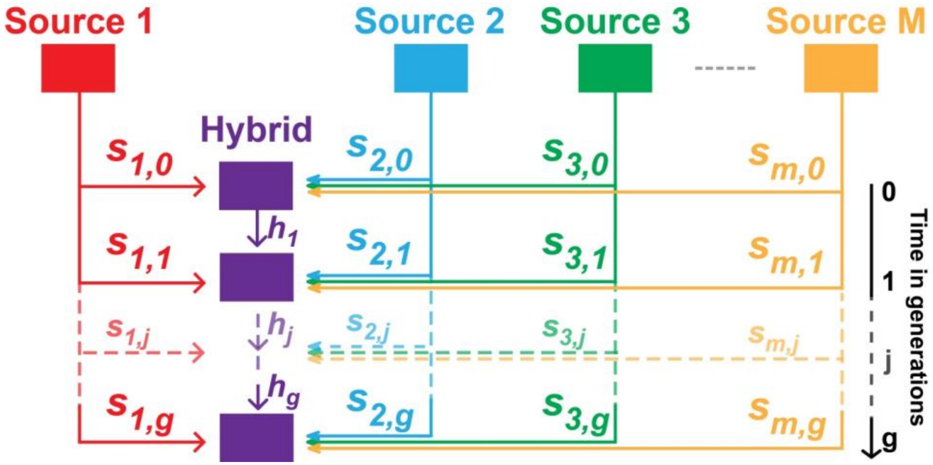
Multilocus pairwise $\boldsymbol{F_{ST}}$ (Weir and Cockerham, 1984)

$\boldsymbol{f_3}$ **(Admixed; S1, S2)** (Patterson et al., 2012)

Individual pairwise **Allele Sharing Dissimilarities** (Bowcock et al., 1994)

Fortes-Lima et al., <u>Mol Ecol Res</u> 2021

MetHis: Distribution of admixture fractions informative about model-parameters



$$\forall\, m \in \{1,..,M\}, s_{m,0} \in [0,1] \text{ such that}$$
$$\sum_{m=1}^{M} s_{m,0} = 1$$

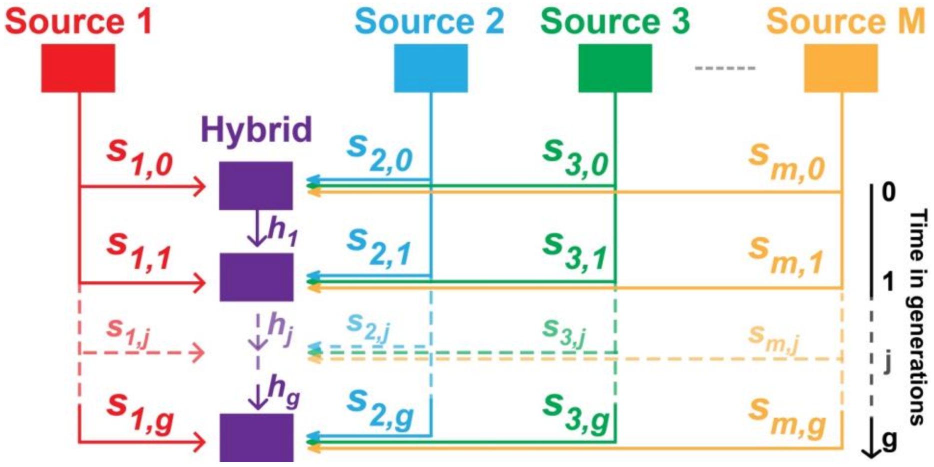
$$\forall\, m \in \{1,..,M\}, \forall\, j \in \{1,..,g\},$$
$$h_j \text{ and } s_{m,j} \in [0,1] \text{ such that}$$
$$h_j + \sum_{m=1}^{M} s_{m,j} = 1$$

For a random locus in a random individual in the hybrid population at generation g > 0,
let $\boldsymbol{H_{i,g}}$ be the probability that this locus originally derives from Source Population *i* (with *i* in {1,.., *M*}):

$$H_{\mathrm{i},1} = \begin{cases} 1 & \text{if } Y = \mathrm{S_i S_i}, \text{with } P[Y = \mathrm{S_i S_i}] = s_{\mathrm{i},0}^2 \\ \frac{1}{2} & \text{if } Y = \mathrm{S_i S_j}, \text{with } P\left[Y = \mathrm{S_i S_j}\right] = 2 s_{\mathrm{i},0} s_{\mathrm{j},0} \\ 0 & \text{if } Y = \mathrm{S_j S_j}, \text{with } P\left[Y = \mathrm{S_j S_j}\right] = s_{\mathrm{j},0}^2 \\ 0 & \text{if } Y = \mathrm{S_j S_l}, \text{with } P\left[Y = \mathrm{S_j S_l}\right] = 2 s_{\mathrm{j},0} s_{\mathrm{l},0}, \end{cases}$$

Verdu and Rosenberg, Genetics 2011

MetHis: Distribution of admixture fractions informative about model-parameters



$\forall\, m \in \{1,..,M\}, s_{m,0} \in [0,1]$ such that

$$\sum_{m=1}^{M} s_{m,0} = 1$$

$\forall\, m \in \{1,..,M\}, \forall\, j \in \{1,..,g\}$,
h_j and $s_{m,j} \in [0,1]$ such that

$$h_j + \sum_{m=1}^{M} s_{m,j} = 1$$

For a random locus in a random individual in the hybrid population at generation g > 0,
let $H_{i,g}$ be the probability that this locus originally derives from Source Population *i* (with *i* in {1,.., *M*}):

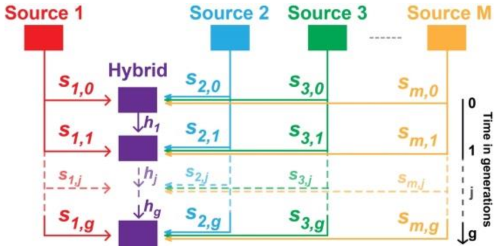
$$H_{\mathrm{i},1} = \begin{cases} 1 & \text{if } Y = \mathrm{S_iS_i}, \text{with } P[Y = \mathrm{S_iS_i}] = s_{\mathrm{i},0}^2 \\ \dfrac{1}{2} & \text{if } Y = \mathrm{S_iS_j}, \text{with } P\left[Y = \mathrm{S_iS_j}\right] = 2s_{\mathrm{i},0}s_{\mathrm{j},0} \\ 0 & \text{if } Y = \mathrm{S_jS_j}, \text{with } P\left[Y = \mathrm{S_jS_j}\right] = s_{\mathrm{j},0}^2 \\ 0 & \text{if } Y = \mathrm{S_jS_l}, \text{with } P\left[Y = \mathrm{S_jS_l}\right] = 2s_{\mathrm{j},0}s_{\mathrm{l},0}, \end{cases}$$

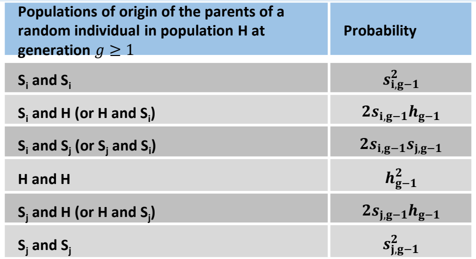
and

$$H_{\mathrm{i,g}} = \begin{cases} 1 & \text{if } Y = \mathrm{S_iS_i}, \text{with } P[Y = \mathrm{S_iS_i}] = s_{\mathrm{i,g-1}}^2 \\ \dfrac{H_{\mathrm{i,g-1}} + 1}{2} & \text{if } Y = \mathrm{S_iH}, \text{with } P[Y = \mathrm{S_iH}] = 2s_{\mathrm{i,g-1}}h_{\mathrm{g-1}} \\ \dfrac{1}{2} & \text{if } Y = \mathrm{S_iS_j}, \text{with } P\left[Y = \mathrm{S_iS_j}\right] = 2s_{\mathrm{i,g-1}}s_{\mathrm{j,g-1}} \\ \dfrac{H_{\mathrm{i,g-1}}^{(1)} + H_{\mathrm{i,g-1}}^{(2)}}{2} & \text{if } Y = \mathrm{HH}, \text{with } P[Y = \mathrm{HH}] = h_{\mathrm{g-1}}^2 \\ \dfrac{H_{\mathrm{i,g-1}}}{2} & \text{if } Y = \mathrm{S_jH}, \text{with } P\left[Y = \mathrm{S_jH}\right] = 2s_{\mathrm{j,g-1}}h_{\mathrm{g-1}} \\ 0 & \text{if } Y = \mathrm{S_jS_j}, \text{with } P\left[Y = \mathrm{S_jS_j}\right] = s_{\mathrm{j,g-1}}^2 \\ 0 & \text{if } Y = \mathrm{S_jS_l}, \text{with } P\left[Y = \mathrm{S_jS_l}\right] = 2s_{\mathrm{j,g-1}}s_{\mathrm{l,g-1}}. \end{cases}$$

Verdu and Rosenberg, Genetics 2011

MetHis: Distribution of admixture fractions informative about model-parameters





Populations of origin of the parents of a random individual in population H at generation $g \geq 1$	Probability
S_i and S_i	$s^2_{i,g-1}$
S_i and H (or H and S_i)	$2s_{i,g-1}h_{g-1}$
S_i and S_j (or S_j and S_i)	$2s_{i,g-1}s_{j,g-1}$
H and H	h^2_{g-1}
S_j and H (or H and S_j)	$2s_{j,g-1}h_{g-1}$
S_j and S_j	$s^2_{j,g-1}$

$$H_{i,g} = \begin{cases} 1 & \text{if } Y = S_iS_i, \text{with } P[Y = S_iS_i] = s^2_{i,g-1} \\ \dfrac{H_{i,g-1}+1}{2} & \text{if } Y = S_iH, \text{with } P[Y = S_iH] = 2s_{i,g-1}h_{g-1} \\ \dfrac{1}{2} & \text{if } Y = S_iS_j, \text{with } P[Y = S_iS_j] = 2s_{i,g-1}s_{j,g-1} \\ \dfrac{H^{(1)}_{i,g-1}+H^{(2)}_{i,g-1}}{2} & \text{if } Y = HH, \text{with } P[Y = HH] = h^2_{g-1} \\ \dfrac{H_{i,g-1}}{2} & \text{if } Y = S_jH, \text{with } P[Y = S_jH] = 2s_{j,g-1}h_{g-1} \\ 0 & \text{if } Y = S_jS_j, \text{with } P[Y = S_jS_j] = s^2_{j,g-1} \\ 0 & \text{if } Y = S_jS_l, \text{with } P[Y = S_jS_l] = 2s_{j,g-1}s_{l,g-1}. \end{cases}$$

Distribution of admixture fractions

$$\boldsymbol{q} \in \boldsymbol{Q}_{\mathbf{g}} = \left\{\mathbf{0}, \frac{\mathbf{1}}{\mathbf{2}^{\boldsymbol{g}}}, \ldots, \frac{\mathbf{2}^{\boldsymbol{g}}-\mathbf{1}}{\mathbf{2}^{\boldsymbol{g}}}, \mathbf{1}\right\}$$

$\forall\, i \geq 1,$

$$P\left(H_{1,1} = q\right) = \begin{cases} s^2_{1,0} & \text{if } q = 1 \\ 2s_{1,0}s_{2,0} & \text{if } q = 1/2 \\ s^2_{2,0} & \text{if } q = 0. \end{cases}$$

$\forall\, i \geq 1, \forall\, g \geq 2$

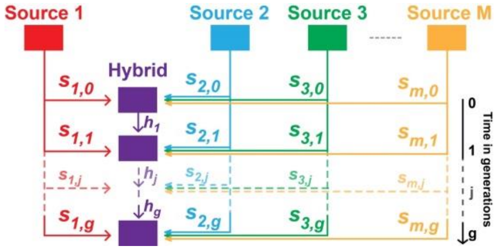
$$P\left(H_{1,g} = q\right) = h^2_{g-1}\sum_{r=0}^{2^{g-1}}\left[P\left(H_{1,g-1} = \frac{r}{2^{g-1}}\right)P\left(H_{1,g-1} = \frac{2^g q - r}{2^{g-1}}\right)\right] + 2s_{2,g-1}h_{g-1}P\left(H_{1,g-1} = 2q\right) + 2s_{1,g-1}h_{g-1}P\left(H_{1,g-1} = 2q-1\right) + I_g(q),$$

with

$$I_g(q) = \begin{cases} s^2_{1,g-1} & \text{if } q = 1 \\ 2s_{1,g-1}s_{2,g-1} & \text{if } q = 1/2 \\ s^2_{2,g-1} & \text{if } q = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Verdu and Rosenberg, Genetics 2011

MetHis: Distribution of admixture fractions informative about model-parameters



Populations of origin of the parents of a random individual in population H at generation $g \geq 1$	Probability
S_i and S_i	$s_{i,g-1}^2$
S_i and H (or H and S_i)	$2s_{i,g-1}h_{g-1}$
S_i and S_j (or S_j and S_i)	$2s_{i,g-1}s_{j,g-1}$
H and H	h_{g-1}^2
S_j and H (or H and S_j)	$2s_{j,g-1}h_{g-1}$
S_j and S_j	$s_{j,g-1}^2$

$$H_{i,g} = \begin{cases} 1 & \text{if } Y = S_iS_i\text{, with } P[Y = S_iS_i] = s_{i,g-1}^2 \\ \frac{H_{i,g-1}+1}{2} & \text{if } Y = S_iH\text{, with } P[Y = S_iH] = 2s_{i,g-1}h_{g-1} \\ \frac{1}{2} & \text{if } Y = S_iS_j\text{, with } P[Y = S_iS_j] = 2s_{i,g-1}s_{j,g-1} \\ \frac{H_{i,g-1}^{(1)}+H_{i,g-1}^{(2)}}{2} & \text{if } Y = HH\text{, with } P[Y = HH] = h_{g-1}^2 \\ \frac{H_{i,g-1}}{2} & \text{if } Y = S_jH\text{, with } P[Y = S_jH] = 2s_{j,g-1}h_{g-1} \\ 0 & \text{if } Y = S_jS_j\text{, with } P[Y = S_jS_j] = s_{j,g-1}^2 \\ 0 & \text{if } Y = S_jS_l\text{, with } P[Y = S_jS_l] = 2s_{j,g-1}s_{l,g-1}. \end{cases}$$

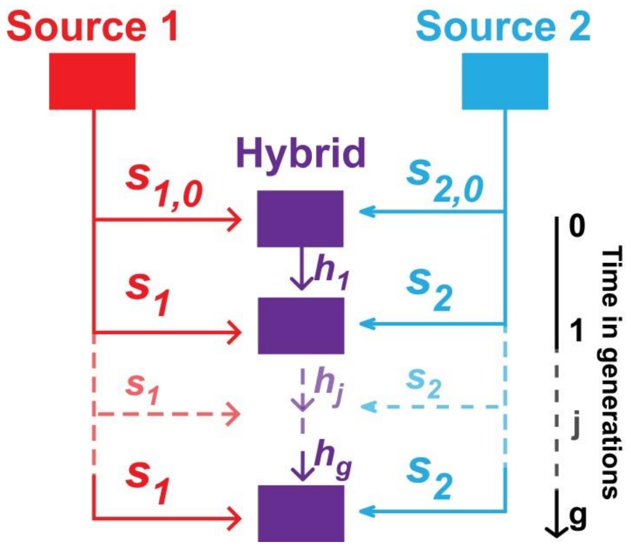
<u>**Law of total expectation**</u>: $$E[H_{i,g}] = E_Y\left[E[H_{i,g}|Y]\right] = \sum_{y\in A} P(Y=y)E[H_{i,g}|\,Y=y].$$

k^{th} - moment

$$\begin{cases} \forall\, k \geq 1, \forall\, i \geq 1, & E[H_{i,1}^k] = s_{i,0}^2 + \frac{s_{i,0}}{2^{k-1}}\sum_{\substack{j=1\\ j\neq i}}^{m} s_{j,0}\,. \\ \forall\, k \geq 1, \forall\, i \geq 1, \forall\, g\ \geq 2 & E[H_{i,g}^k] = s_{i,g-1}^2 + \frac{s_{i,g-1}h_{g-1}}{2^{k-1}}\left(\sum_{r=0}^{k}\binom{k}{r}E[H_{i,g-1}^r]\right) + \frac{s_{i,g-1}}{2^{k-1}}\sum_{\substack{j=1\\ j\neq i}}^{m} s_{j,g-1} + \frac{h_{g-1}^2}{2^k}\left(\sum_{r=0}^{k}\binom{k}{r}E[H_{i,g-1}^r]E[H_{i,g-1}^{k-r}]\right) + \left(\frac{h_{g-1}}{2^{k-1}}\sum_{\substack{j=1\\ j\neq i}}^{m} s_{j,g-1}\right)E[H_{i,g-1}^k]. \end{cases}$$

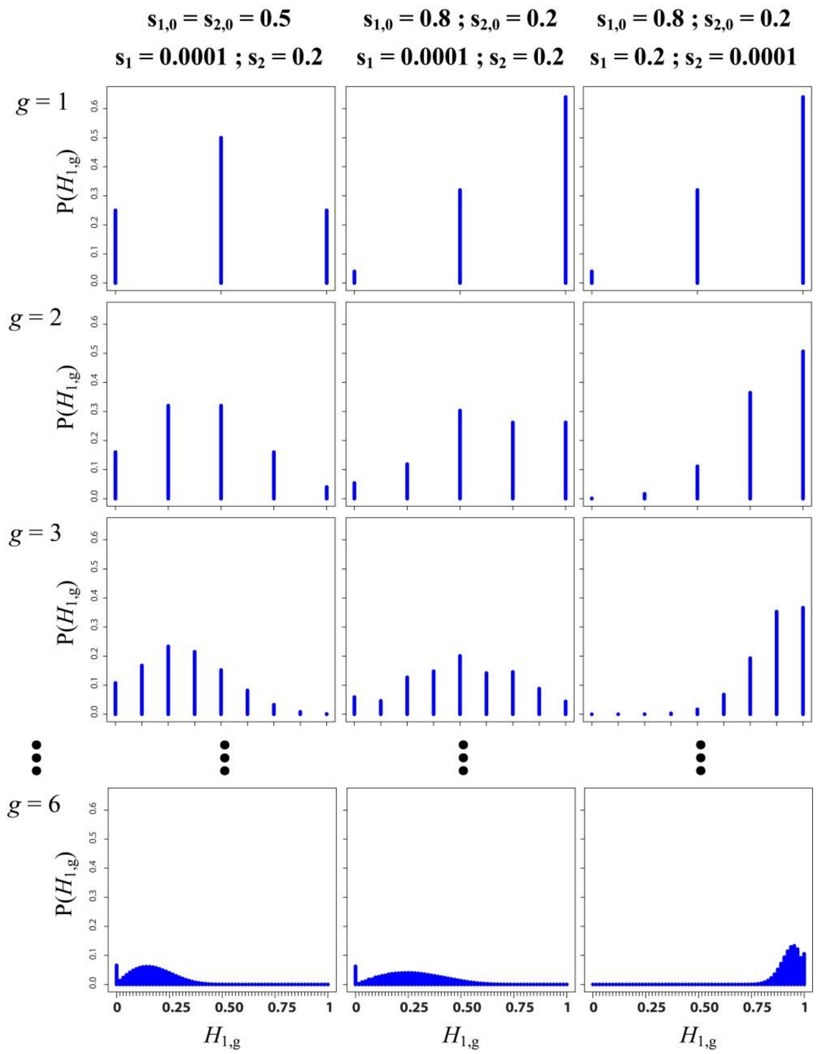
Verdu and Rosenberg, <u>Genetics</u> 2011

MetHis: Distribution of admixture fractions informative about model-parameters



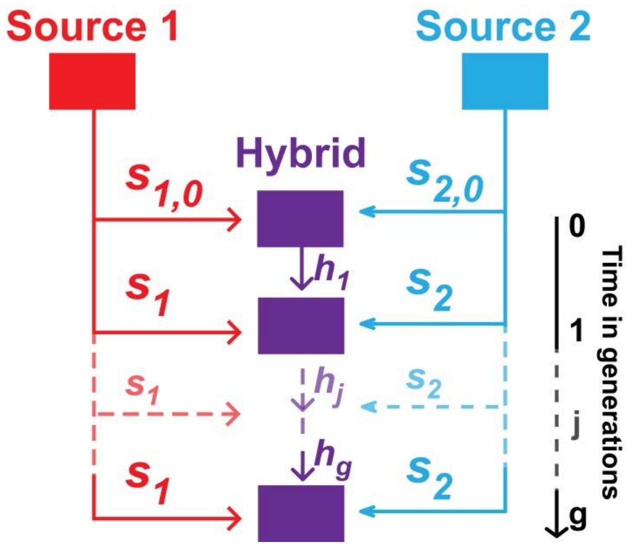
$$E\left[H_{1,\mathrm{g}}\right] = \begin{cases} s_{1,0}, & g = 1 \\ s_{1,0}h^{\mathrm{g}-1} + s_1 \dfrac{1-h^{\mathrm{g}-1}}{1-h}, & g \geq 2 \ . \end{cases}$$

$$V\left[H_{1,\mathrm{g}}\right] = \begin{cases} \dfrac{s_{1,0}\left(1-s_{1,0}\right)}{2}, & g = 1 \\ A_1 + A_2 h^{\mathrm{g}-1} + A_3\left(\dfrac{h}{2}\right)^{\mathrm{g}-1} + A_4\left(\dfrac{h}{2}\right)^{\mathrm{g}-1}\displaystyle\sum_{\mathrm{i}=1}^{\mathrm{g}-1}(2h)^{\mathrm{i}} - \left(s_{1,0}h^{\mathrm{g}-1} + s_1\dfrac{1-h^{\mathrm{g}-1}}{1-h}\right)^2, & g \geq 2. \end{cases}$$



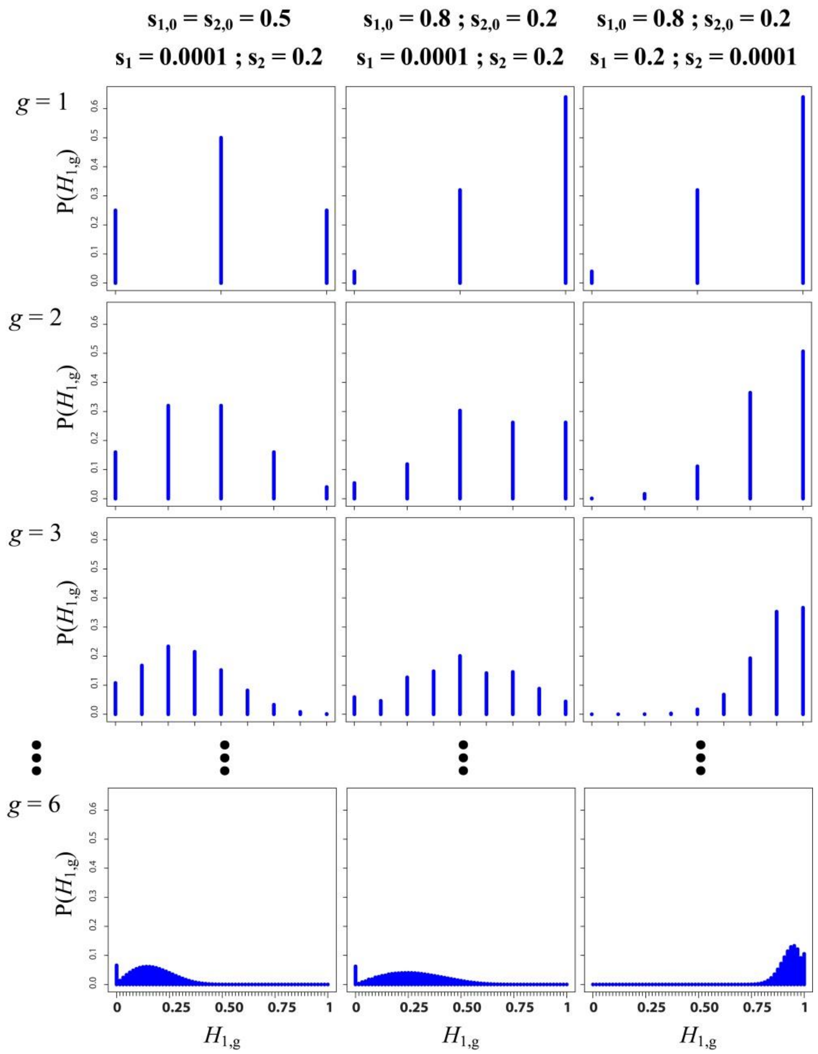
Verdu and Rosenberg, Genetics 2011

MetHis: Distribution of admixture fractions informative about model-parameters



$$E\left[H_{1,\mathrm{g}}\right] = \begin{cases} s_{1,0}, & g = 1 \\ s_{1,0}h^{\mathrm{g}-1} + s_1 \dfrac{1 - h^{\mathrm{g}-1}}{1 - h}, & g \geq 2\ . \end{cases}$$

$$V\left[H_{1,\mathrm{g}}\right] = \begin{cases} \dfrac{s_{1,0}\left(1 - s_{1,0}\right)}{2}, & g = 1 \\ A_1 + A_2 h^{\mathrm{g}-1} + A_3 \left(\dfrac{h}{2}\right)^{\mathrm{g}-1} + A_4 \left(\dfrac{h}{2}\right)^{\mathrm{g}-1} \displaystyle\sum_{\mathrm{i}=1}^{\mathrm{g}-1} (2h)^{\mathrm{i}} - \left(s_{1,0}h^{\mathrm{g}-1} + s_1 \dfrac{1 - h^{\mathrm{g}-1}}{1 - h}\right)^2, & g \geq 2. \end{cases}$$



Use the distribution of admixture fractions as an ABC-informative summary statistics !

Verdu and Rosenberg, Genetics 2011

MetHis: summary-statistics calculator for ABC inference

If <u>summary-statistics are informative</u> about model-parameters…

“Classical” population genetics summary-statistics

Expected Heterozygosity (Nei, 1978)

Inbreeding coefficient ***F*** (Danecek et al. 2011)

Multilocus pairwise F_{ST} (Weir and Cockerham, 1984)

f_3 **(Admixed; S1, S2)** (Patterson et al., 2012)

Individual pairwise **Allele Sharing Dissimilarities** (Bowcock et al., 1994)

Individual admixture fractions

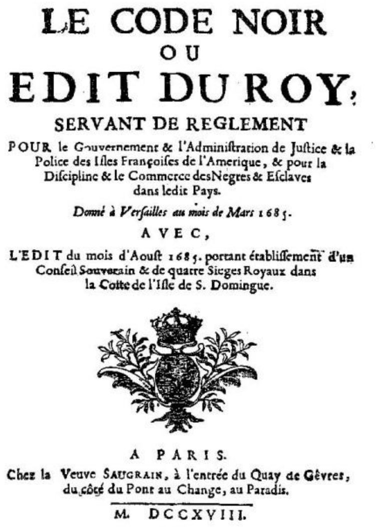
Min and **Max** admixture fractions

10% quantiles of individual admixture fractions

Mode, **Mean**, **Variance**, **Kurtosis**, **Skewness** of individual admixture fractions

Fortes-Lima et al., Mol Ecol Res 2021

Case study: Admixture history of the Afro-American and Barbadian populations from the 1KGP



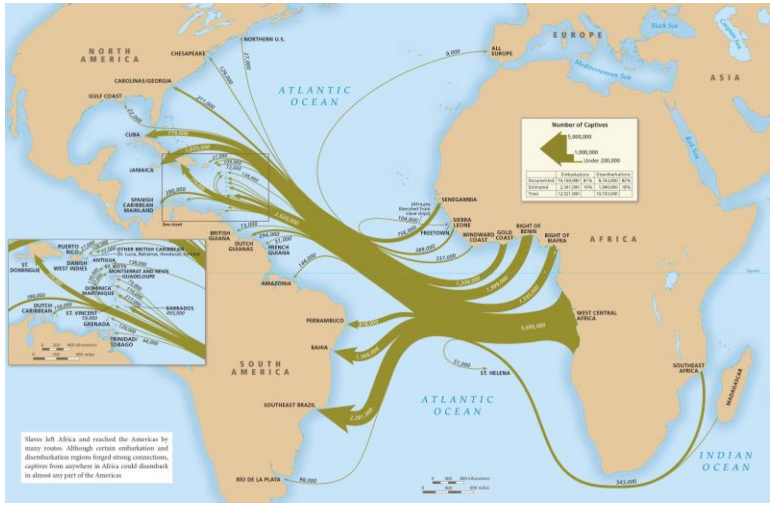
LE CODE NOIR
OU
EDIT DU ROY,
SERVANT DE REGLEMENT
POUR le Gouvernement & l'Adminiſtration de Juſtice & la
Police des Iſles Françoiſes de l'Amerique, & pour la
Diſcipline & le Commerce des Negres & Eſclaves
dans ledit Pays.

Donné à Verſailles au mois de Mars 1685.

AVEC,

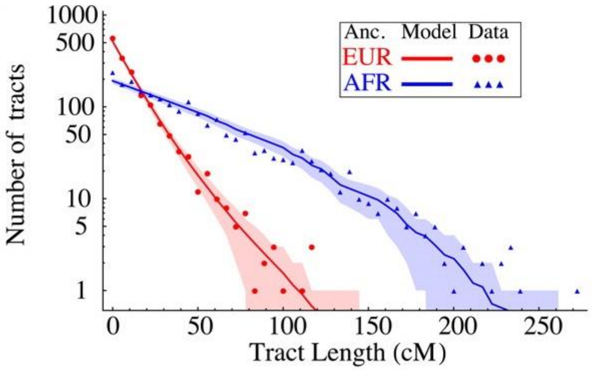
L'EDIT du mois d'Aouſt 1685. portant établiſſement d'un
Conſeil Souverain & de quatre Sieges Royaux dans
la Cotte de l'Iſle de S. Domingue.

A PARIS.
Chez la Veuve SAUGRAIN, à l'entrée du Quay de Gêvres,
du côté du Pont au Change, au Paradis.
M. DCCXVIII.

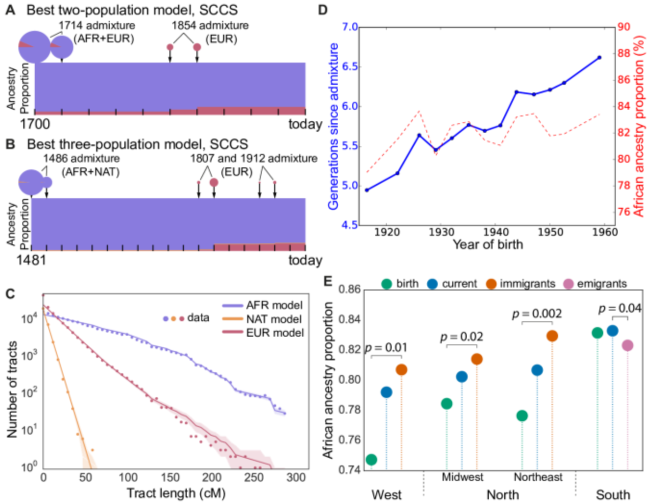


Transatlantic Slave Trade:

- ► Recent admixture history (~20 generations)
- ► Variable migration histories to the Americas
- ► Variable slavery histories in the Americas



Gravel, Genetics 2012



Baharian et al., Plos Gen 2016

Previous studies tested 2 admixture pulses at most, **with ML methods using >> 1 million SNPs**

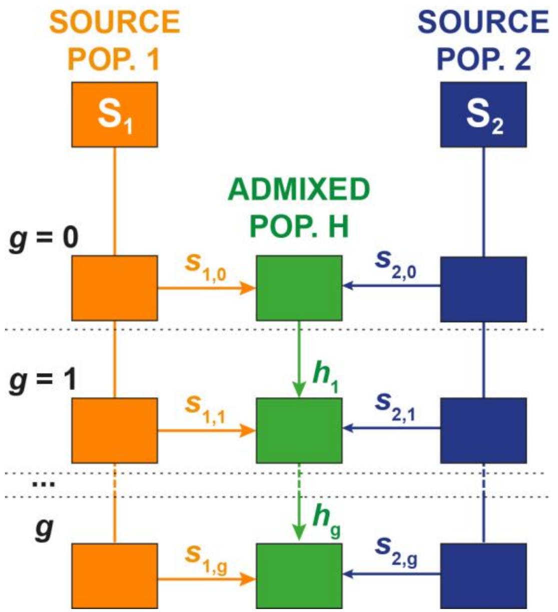
They found that 2 admixture pulses from Europe were most likely… but cannot try more complex models.

Fortes-Lima and Verdu, Hum Mol Gen 2021

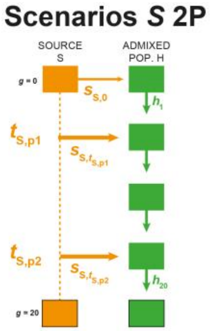
Nine competing models for the admixture history of the ASW and the ACB

21 generations before present (*g* = 20)

= Admixed population founded ~ 1500s



3 admixture pulses



Parameter priors

$s_{S,0}$	*Uniform* [0,1]
$s_{S,t_{S,p1}}$	*Uniform* [0,1]
$s_{S,t_{S,p2}}$	*Uniform* [0,1]
$t_{S,p1}$	*Uniform* [1,20]
$t_{S,p2}$	*Uniform* [1,20]

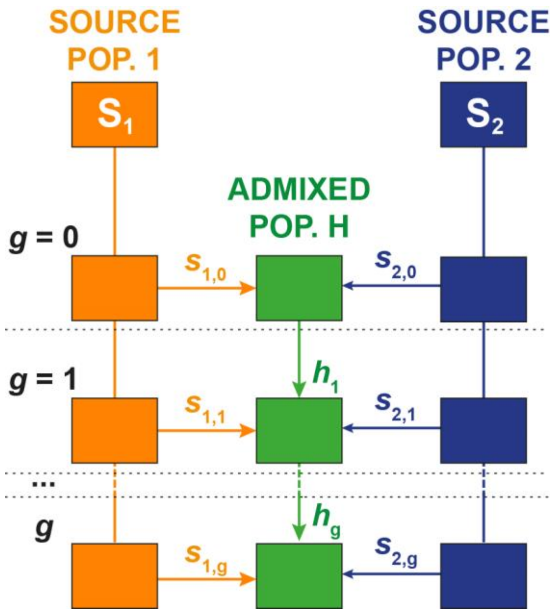
$t_{S,p1} \neq t_{S,p2}$

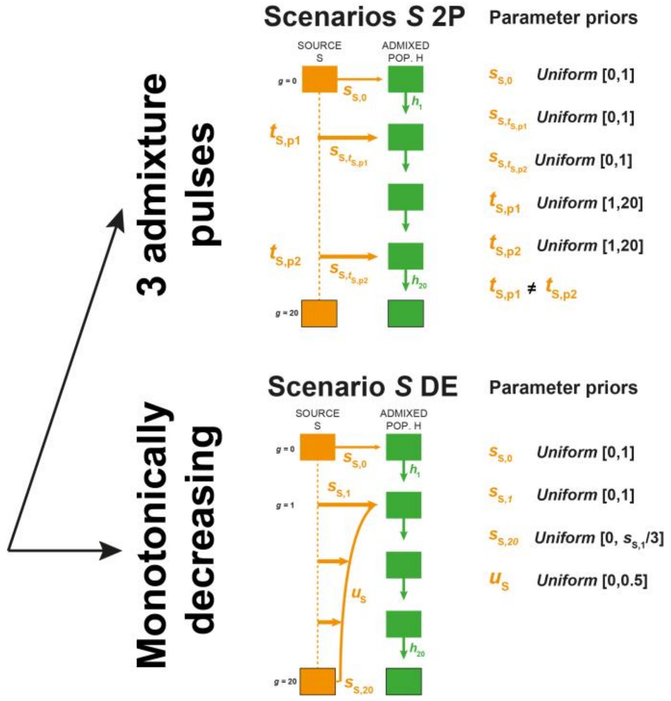
Fortes-Lima et al., Mol Ecol Res 2021

Nine competing models for the admixture history of the ASW and the ACB

21 generations before present (*g* = 20)

= Admixed population founded ~ 1500s





Scenarios *S* 2P

Parameter priors	
$s_{S,0}$	*Uniform* [0,1]
$s_{S,t_{S,p1}}$	*Uniform* [0,1]
$s_{S,t_{S,p2}}$	*Uniform* [0,1]
$t_{S,p1}$	*Uniform* [1,20]
$t_{S,p2}$	*Uniform* [1,20]
$t_{S,p1} \neq t_{S,p2}$	

Scenario *S* DE

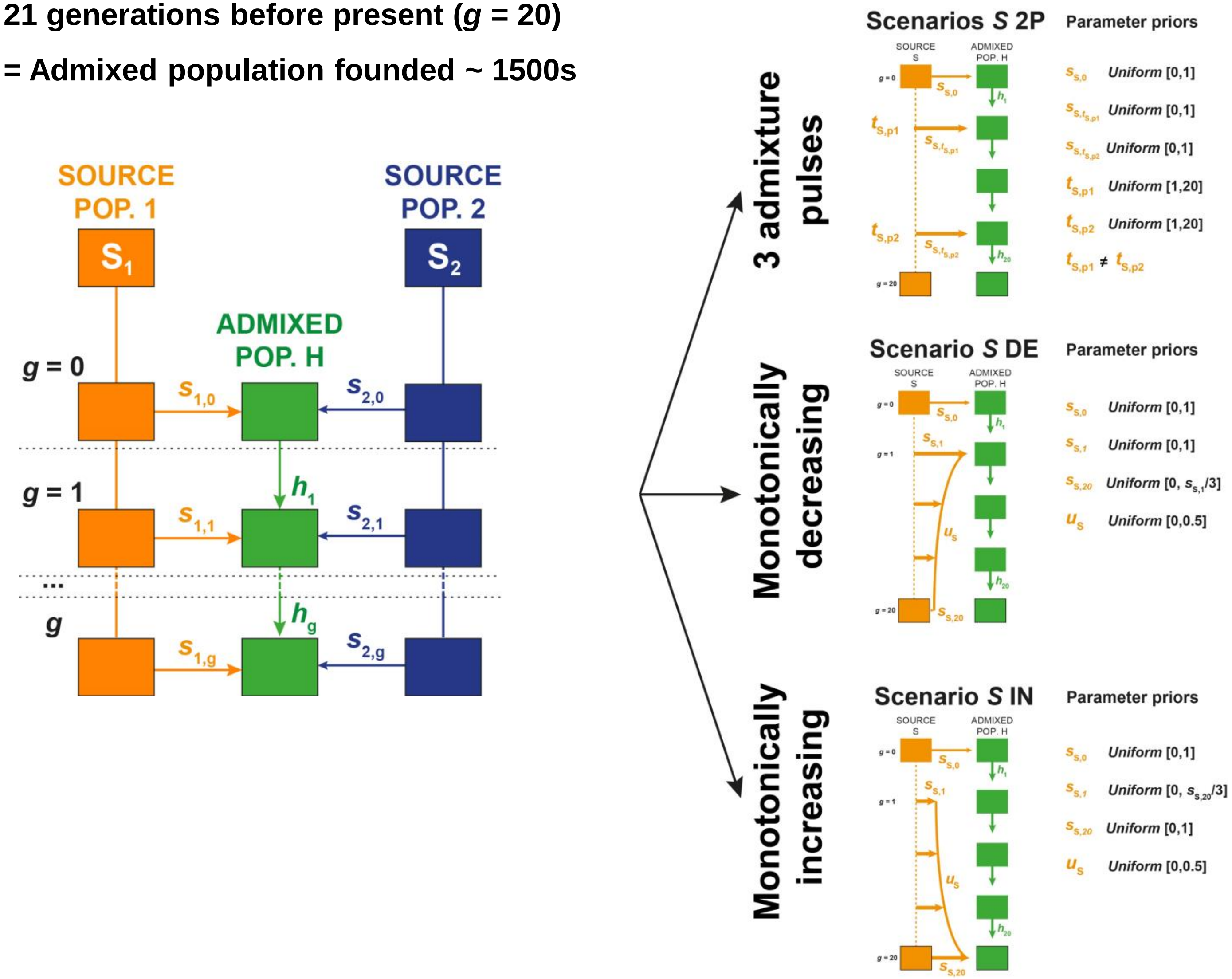
Parameter priors	
$s_{S,0}$	*Uniform* [0,1]
$s_{S,1}$	*Uniform* [0,1]
$s_{S,20}$	*Uniform* [0, $s_{S,1}$/3]
u_S	*Uniform* [0,0.5]

Fortes-Lima et al., Mol Ecol Res 2021

Nine competing models for the admixture history of the ASW and the ACB

21 generations before present (*g* = 20)

= Admixed population founded ~ 1500s



Scenarios *S* 2P — Parameter priors

$s_{S,0}$ *Uniform* [0,1]

$s_{S,t_{S,p1}}$ *Uniform* [0,1]

$s_{S,t_{S,p2}}$ *Uniform* [0,1]

$t_{S,p1}$ *Uniform* [1,20]

$t_{S,p2}$ *Uniform* [1,20]

$t_{S,p1} \neq t_{S,p2}$

Scenario *S* DE — Parameter priors

$s_{S,0}$ *Uniform* [0,1]

$s_{S,1}$ *Uniform* [0,1]

$s_{S,20}$ *Uniform* [0, $s_{S,1}$/3]

u_S *Uniform* [0,0.5]

Scenario *S* IN — Parameter priors

$s_{S,0}$ *Uniform* [0,1]

$s_{S,1}$ *Uniform* [0, $s_{S,20}$/3]

$s_{S,20}$ *Uniform* [0,1]

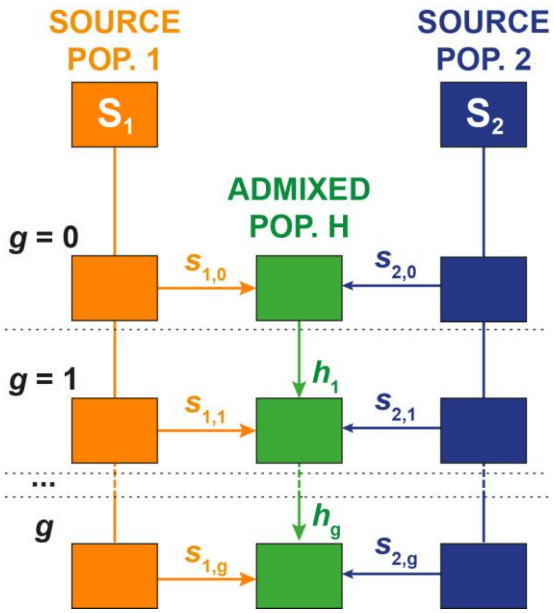
u_S *Uniform* [0,0.5]

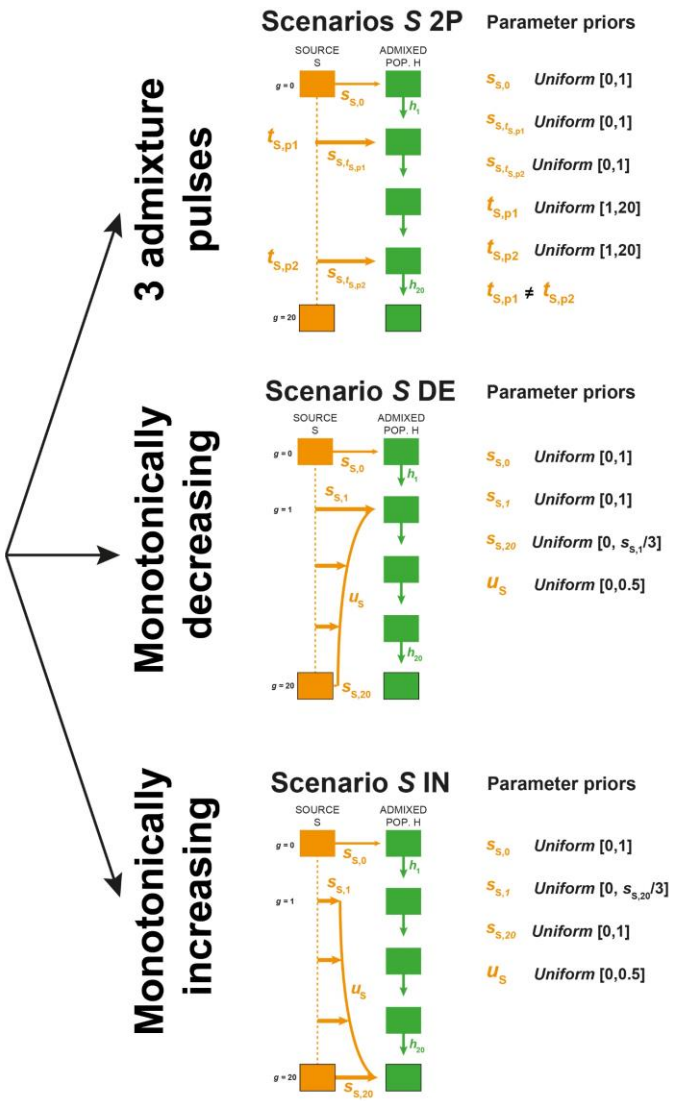
Fortes-Lima et al., Mol Ecol Res 2021

Nine competing models for the admixture history of the ASW and the ACB

21 generations before present (*g* = 20)

= Admixed population founded ~ 1500s





Parameter priors (Scenarios *S* 2P)

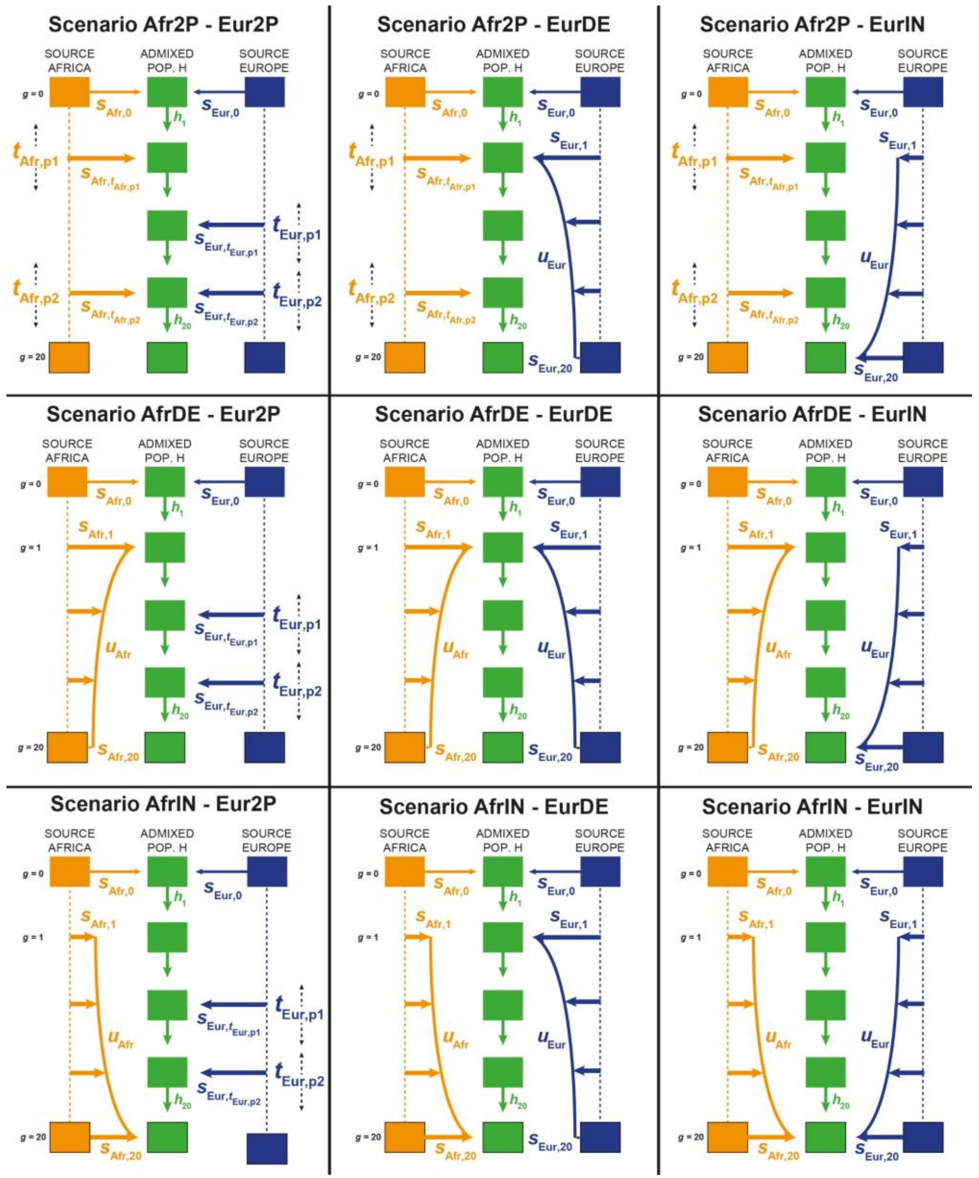
$s_{S,0}$	*Uniform* [0,1]
$s_{S,t_{S,p1}}$	*Uniform* [0,1]
$s_{S,t_{S,p2}}$	*Uniform* [0,1]
$t_{S,p1}$	*Uniform* [1,20]
$t_{S,p2}$	*Uniform* [1,20]
$t_{S,p1} \neq t_{S,p2}$	

Parameter priors (Scenario *S* DE)

$s_{S,0}$	*Uniform* [0,1]
$s_{S,1}$	*Uniform* [0,1]
$s_{S,20}$	*Uniform* [0, $s_{S,1}$/3]
u_S	*Uniform* [0,0.5]

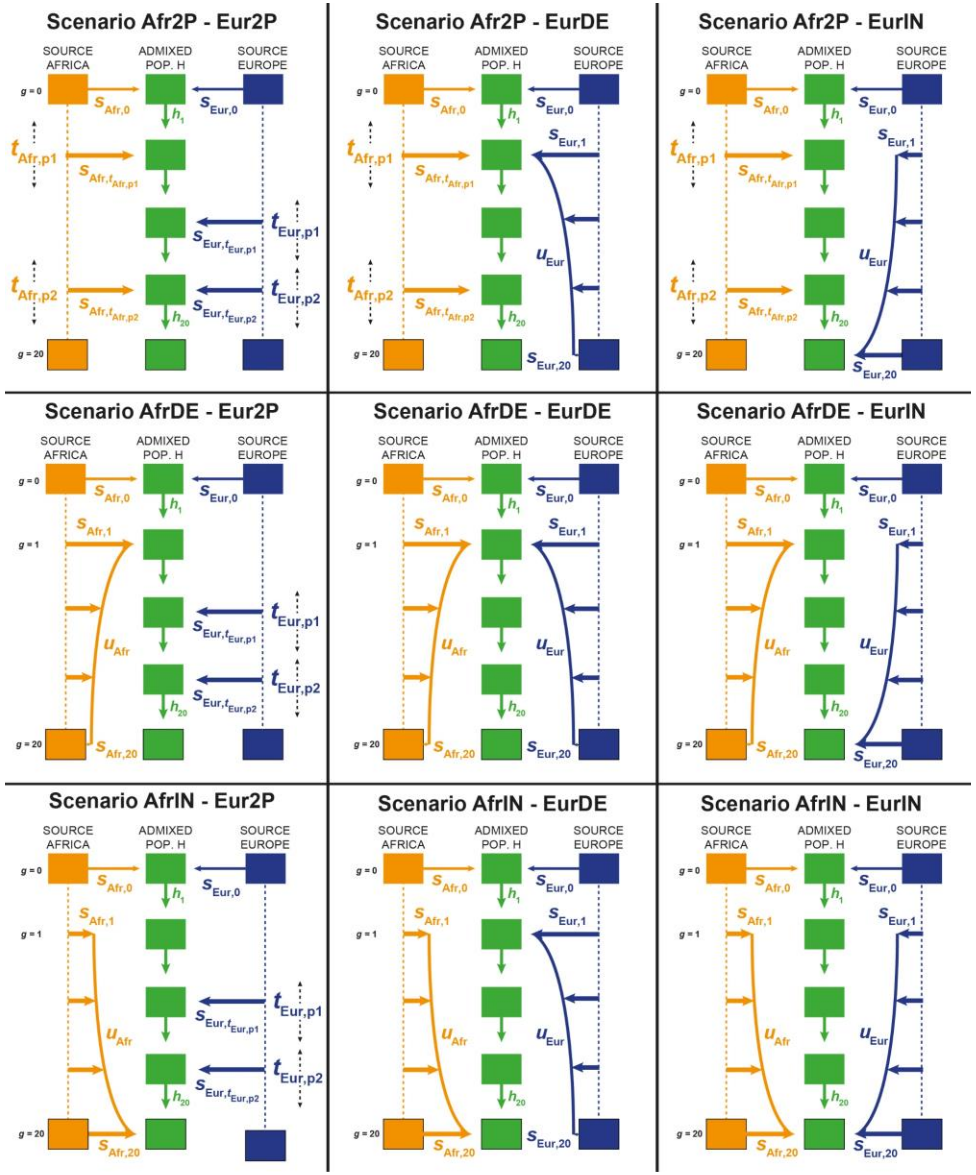
Parameter priors (Scenario *S* IN)

$s_{S,0}$	*Uniform* [0,1]
$s_{S,1}$	*Uniform* [0, $s_{S,20}$/3]
$s_{S,20}$	*Uniform* [0,1]
u_S	*Uniform* [0,0.5]



Fortes-Lima et al., Mol Ecol Res 2021

Nine competing models for the admixture history of the ASW and the ACB



Fortes-Lima et al., Mol Ecol Res 2021

- **100,000 independent SNPs**
- ***N_g* = constant 1,000 individuals** (for simplicity here, *MetHis* option)
- **10,000 simulations per model = 90,000 simulations total**
- **Sample 90 Africans, 89 Europeans, and 50 individuals in the admixed population.**

For each simulated data set, calculate 24 population genetics summary statistics

27 cores -> 3 days calculation total.

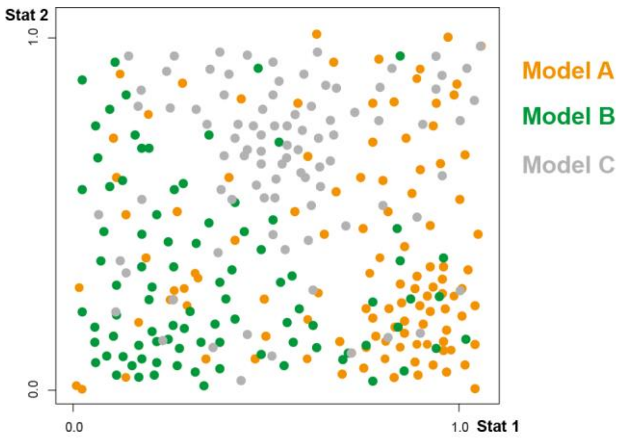
2/3 of this time = summary statistics calculation…

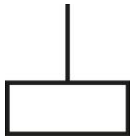
Simulation time increases with Ne much more rapidly than with g.

Random Forest algorithm for model choice (Breiman 2001)

Draw randomly stats and order them by variance explained

Build Decision Tree 1

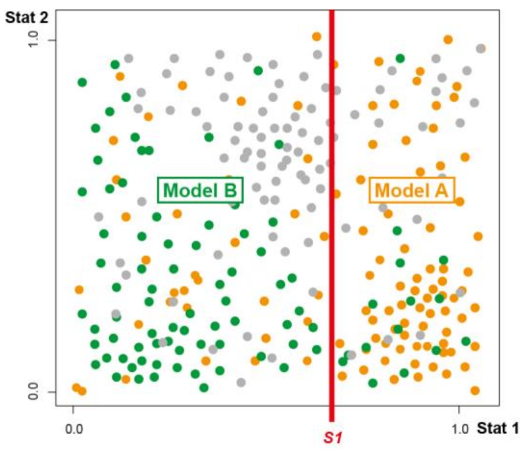


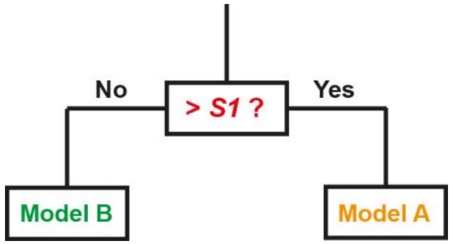


Random Forest algorithm for model choice (Breiman 2001)

Draw randomly stats and order
them by variance explained

Build Decision Tree 1

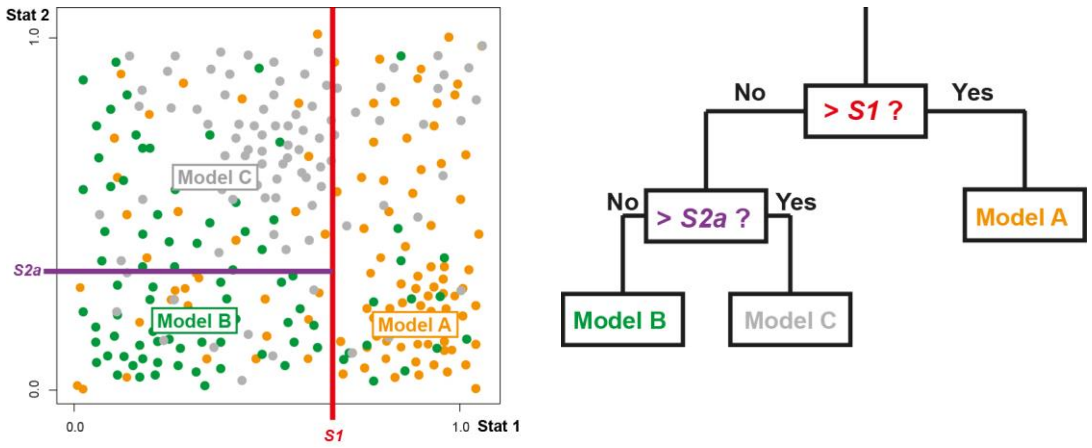




Random Forest algorithm for model choice (Breiman 2001)

Draw randomly stats and order
them by variance explained

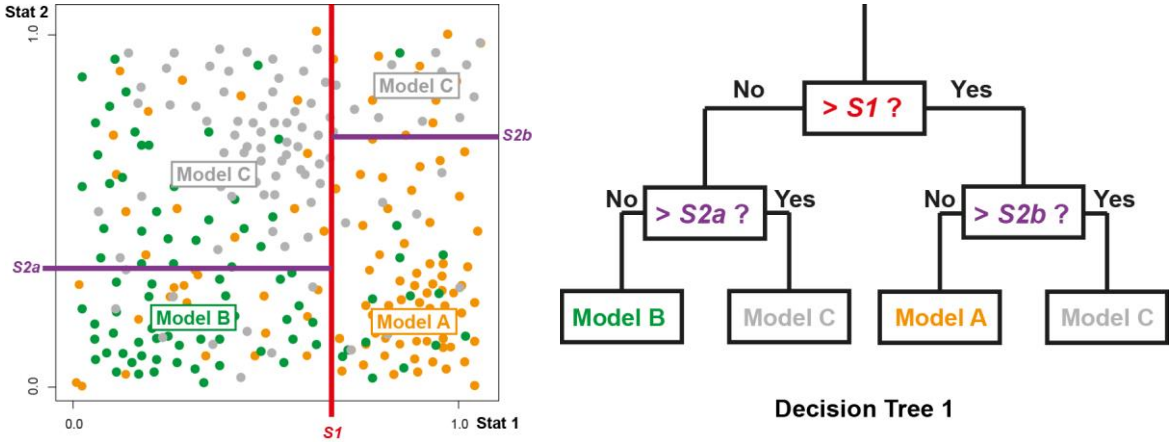
Build Decision Tree 1



Random Forest algorithm for model choice (Breiman 2001)

Draw randomly stats and order
them by variance explained

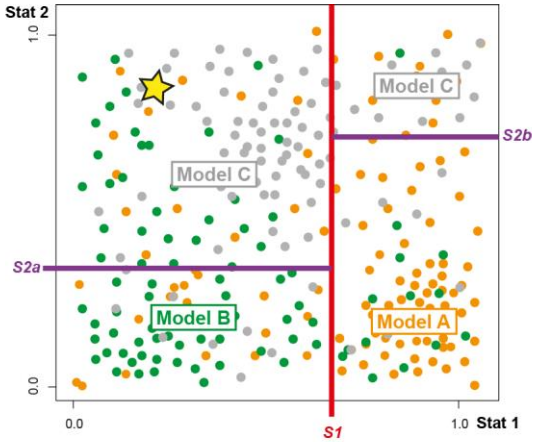
Build Decision Tree 1

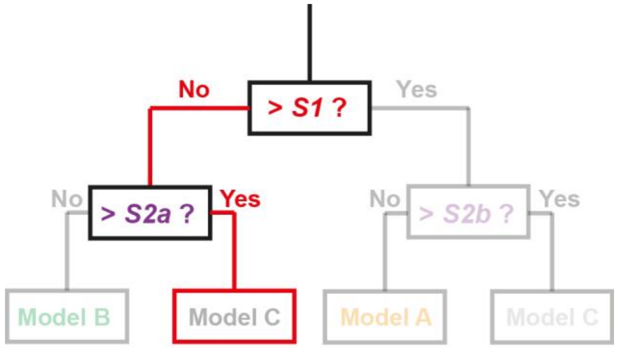


Random Forest algorithm for model choice (Breiman 2001)

Draw randomly stats and order
them by variance explained

Build Decision Tree 1



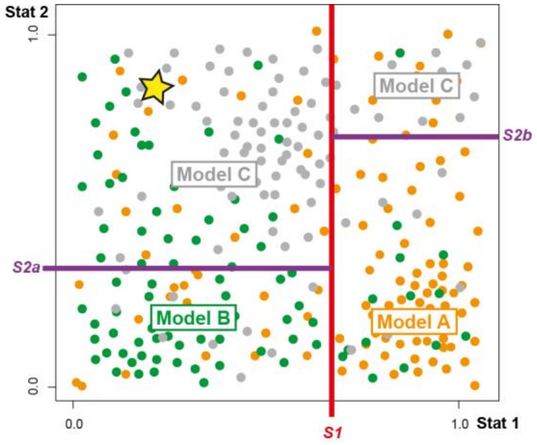


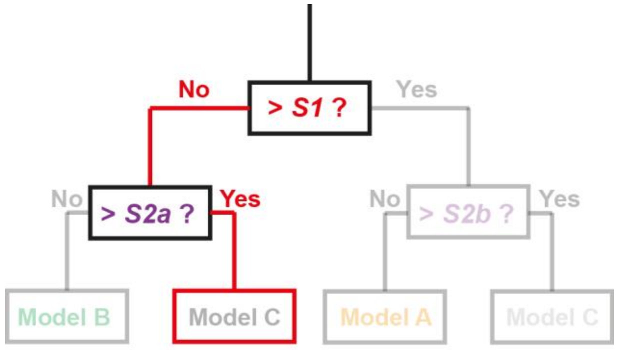
Prediction with decision tree 1 = Model C

Random Forest algorithm for model choice (Breiman 2001)

Draw randomly stats and order them by variance explained

Build Decision Tree 1

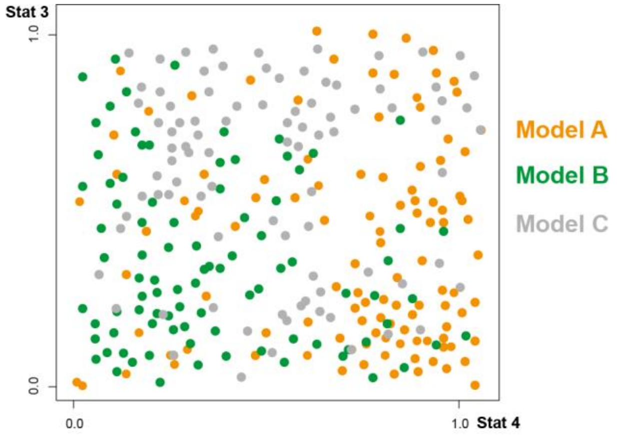


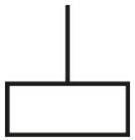


Prediction with decision tree 1 = Model C

Draw randomly other stats and order them by variance explained

Build Decision Tree 2





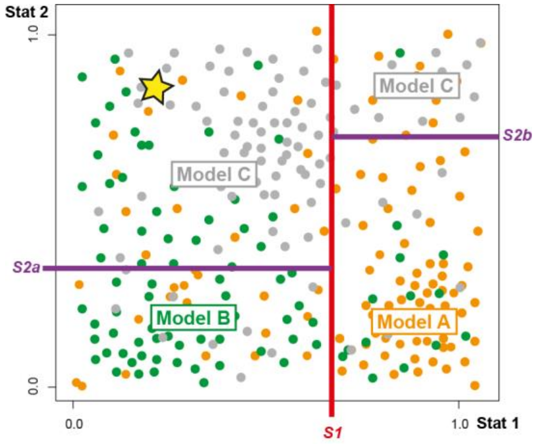
Random Forest algorithm for model choice (Breiman 2001)

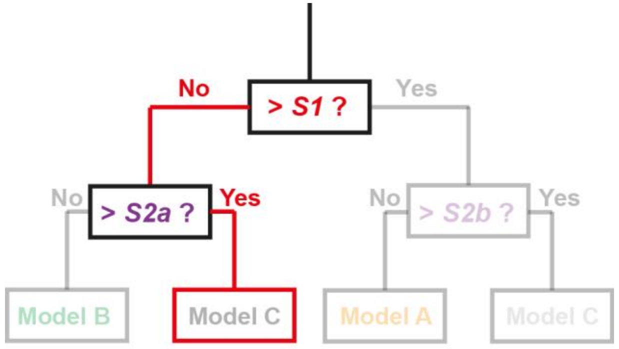
Draw randomly stats and order them by variance explained

Build Decision Tree 1

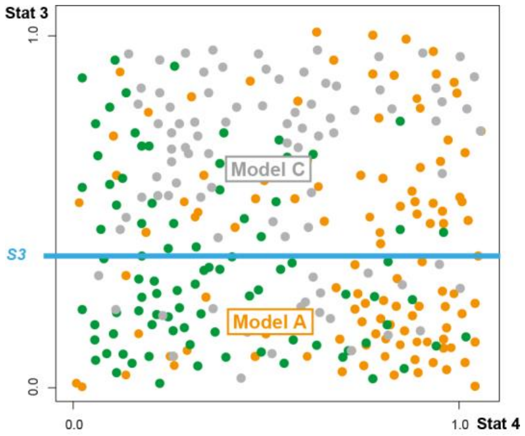
Draw randomly other stats and order them by variance explained

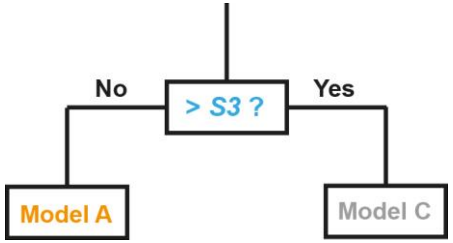
Build Decision Tree 2





Prediction with decision tree 1 = Model C





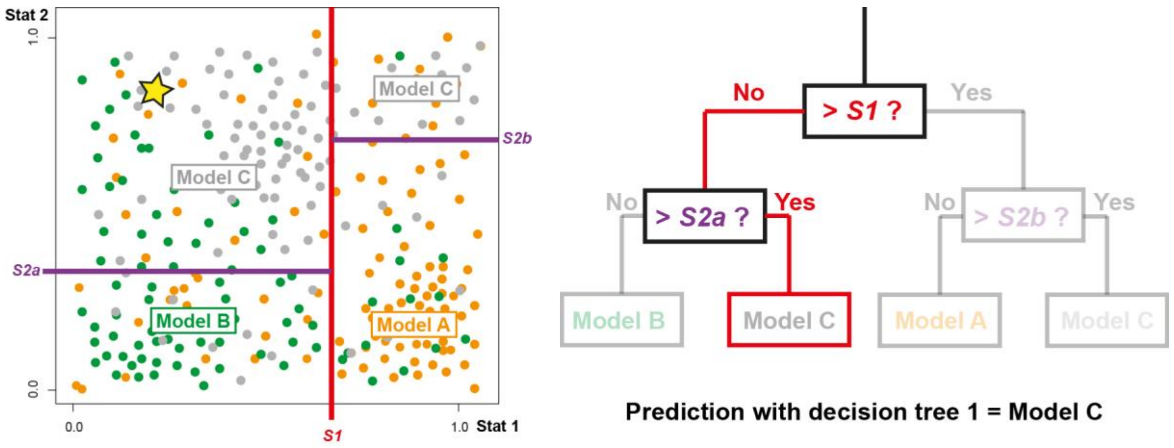
Random Forest algorithm for model choice (Breiman 2001)

Draw randomly stats and order them by variance explained

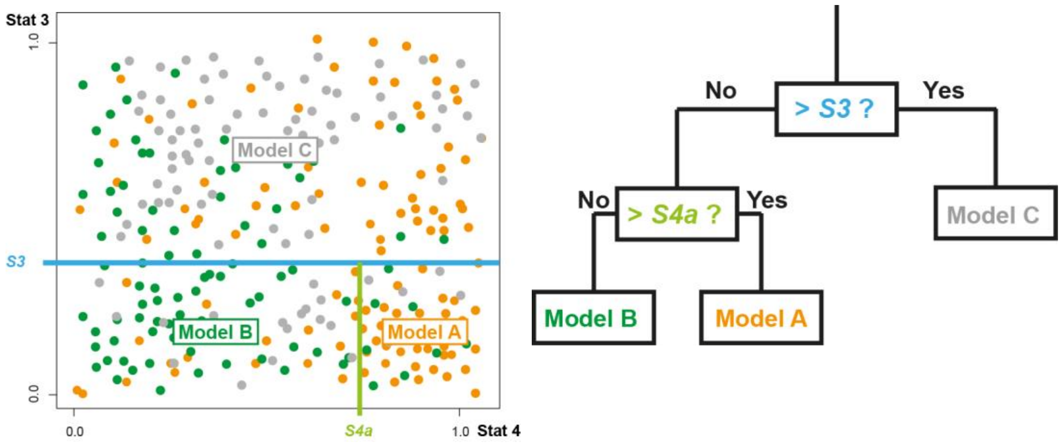
Build Decision Tree 1

Draw randomly other stats and order them by variance explained

Build Decision Tree 2



Prediction with decision tree 1 = Model C



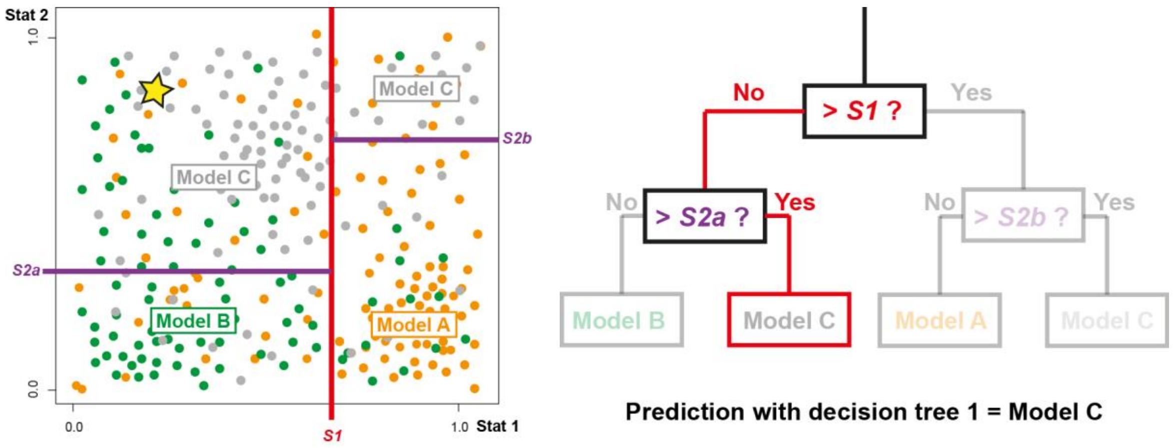
Random Forest algorithm for model choice (Breiman 2001)

Draw randomly stats and order them by variance explained

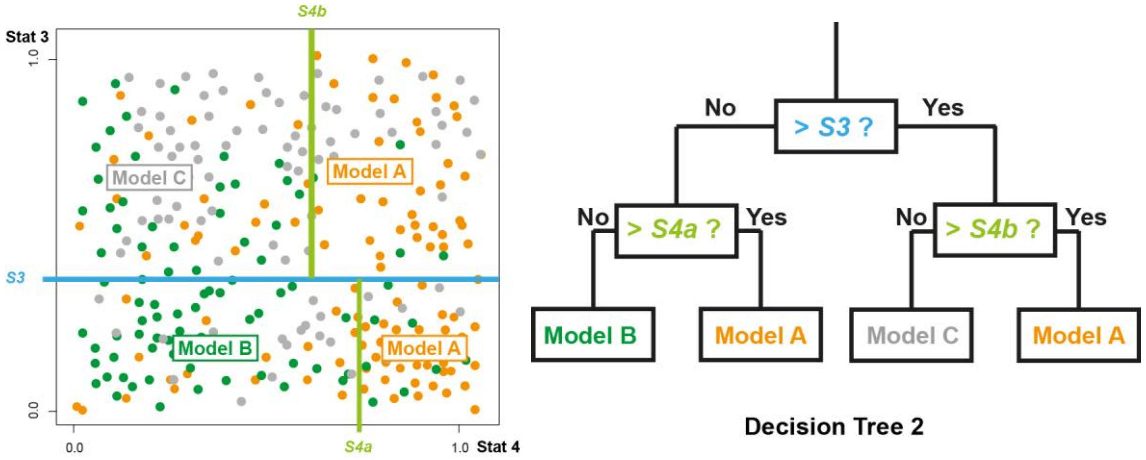
Build Decision Tree 1

Draw randomly other stats and order them by variance explained

Build Decision Tree 2



Prediction with decision tree 1 = Model C



Decision Tree 2

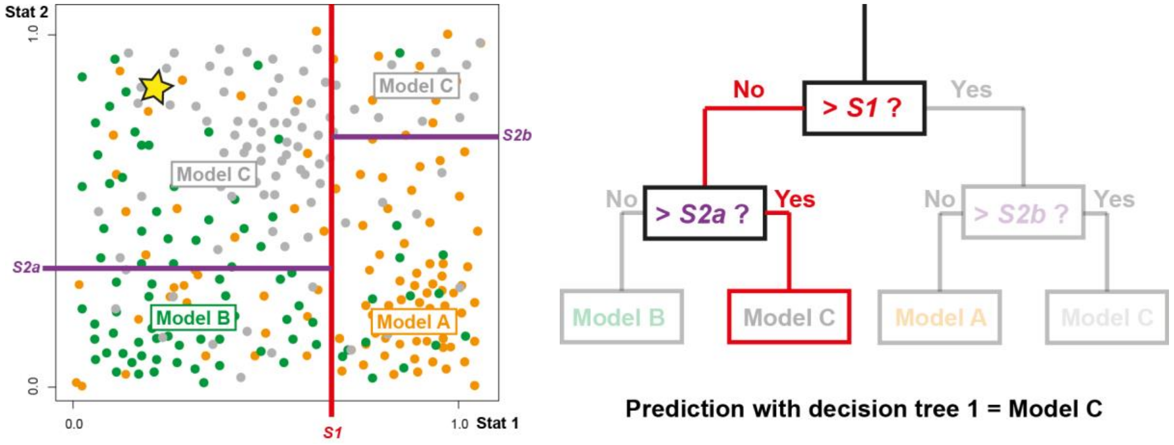
Random Forest algorithm for model choice (Breiman 2001)

Draw randomly stats and order them by variance explained

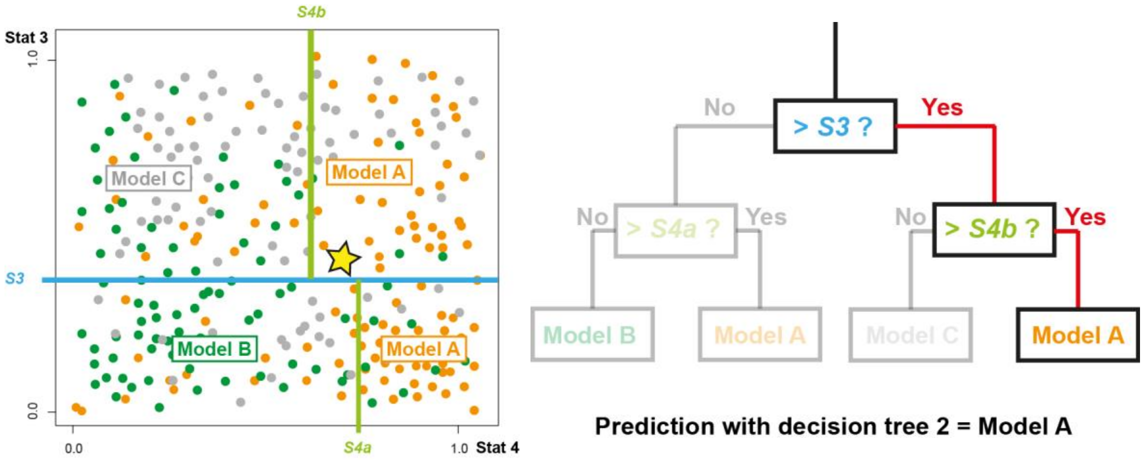
Build Decision Tree 1

Draw randomly other stats and order them by variance explained

Build Decision Tree 2



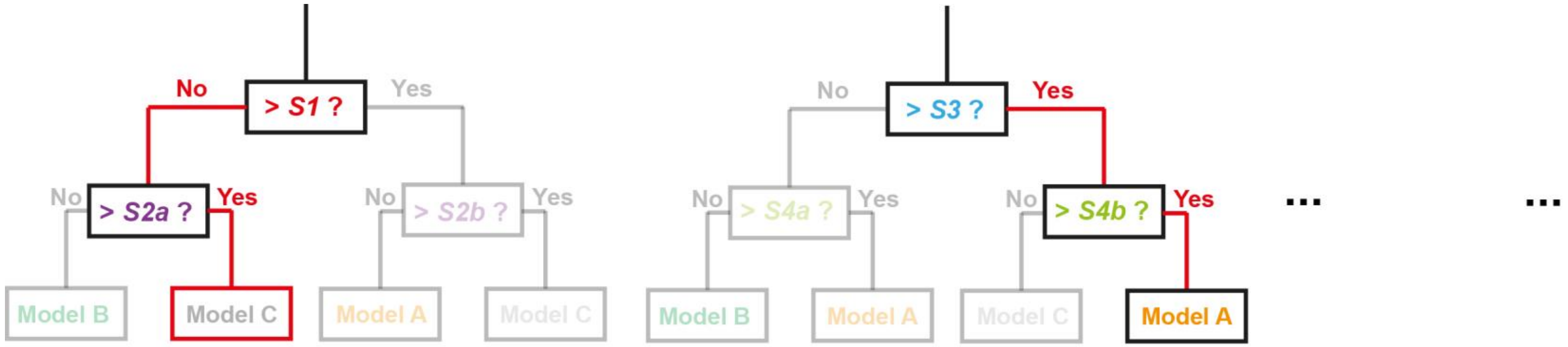
Prediction with decision tree 1 = Model C



Prediction with decision tree 2 = Model A

Random Forest algorithm for model choice (Breiman 2001)

Repeat and build a random-forest of decision trees



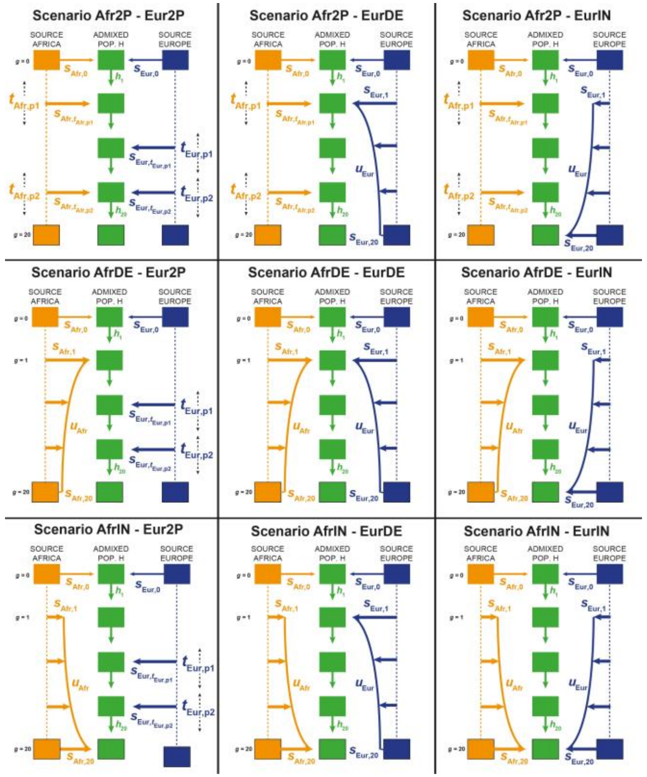
Prediction with decision tree 1 = Model C Prediction with decision tree 2 = Model A ... Prediction with decision tree 1000 = ...

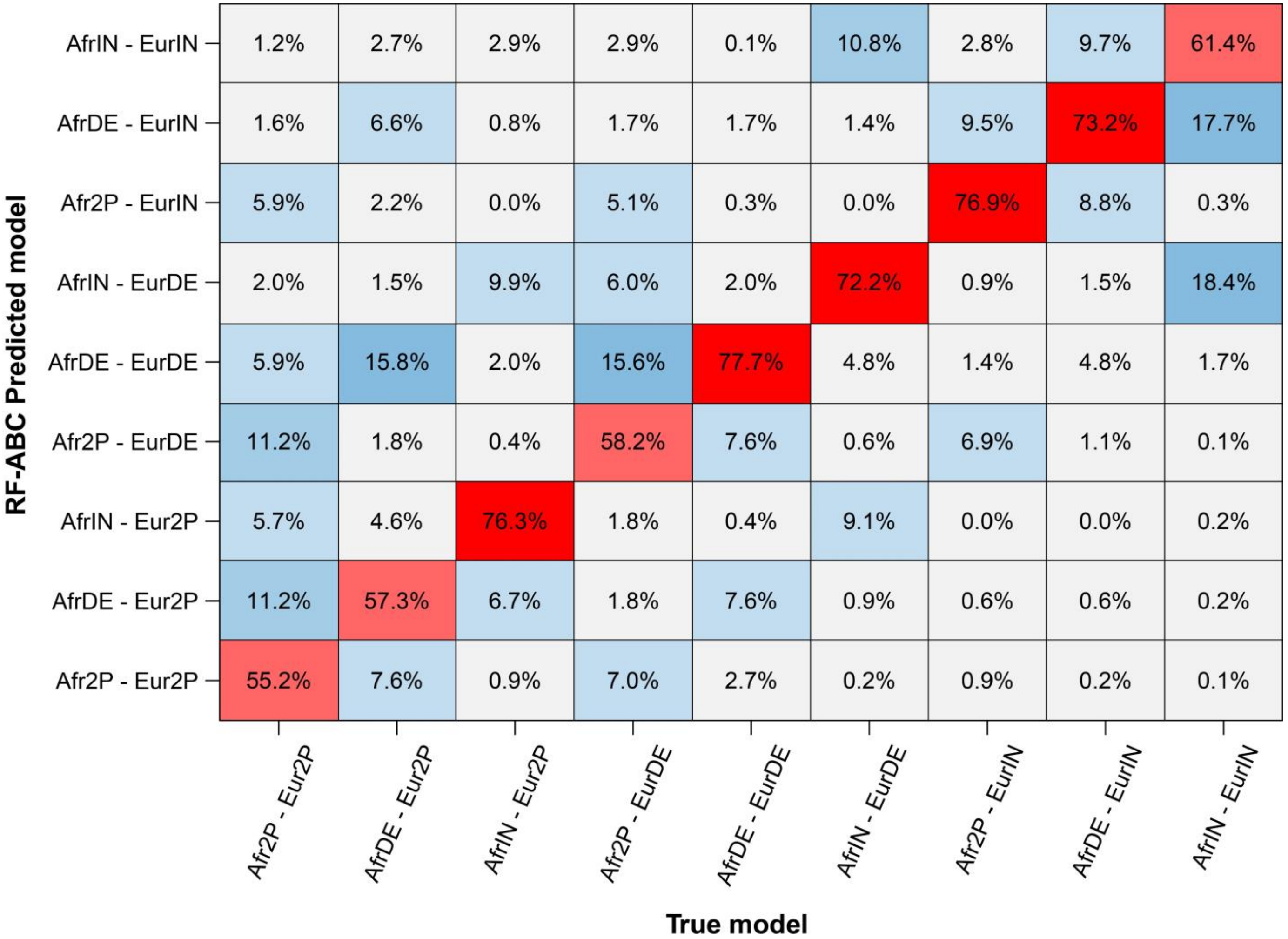
Prediction for the observed data is, e.g., the majority of votes across 1000 trees in the random forest

Results: *MetHis*-ABC model-choice with Random-Forest <u>a priori without observed data</u>

- **Random Forest ABC *(abcrf* package in *R, Pudlo et al. 2016)* – 10,000 *MetHis* sims/scenario – 1,000 trees**
- **Cross-validation all simulations**
- **100,000 independent SNPs**
- **Sample sizes:**

  Afr Source : 90 indivs
  Eur Source : 89 indivs
  Admixed population H : 50 indivs





RF-ABC Predicted model \ True model	Afr2P - Eur2P	AfrDE - Eur2P	AfrIN - Eur2P	Afr2P - EurDE	AfrDE - EurDE	AfrIN - EurDE	Afr2P - EurIN	AfrDE - EurIN	AfrIN - EurIN
AfrIN - EurIN	1.2%	2.7%	2.9%	2.9%	0.1%	10.8%	2.8%	9.7%	61.4%
AfrDE - EurIN	1.6%	6.6%	0.8%	1.7%	1.7%	1.4%	9.5%	73.2%	17.7%
Afr2P - EurIN	5.9%	2.2%	0.0%	5.1%	0.3%	0.0%	76.9%	8.8%	0.3%
AfrIN - EurDE	2.0%	1.5%	9.9%	6.0%	2.0%	72.2%	0.9%	1.5%	18.4%
AfrDE - EurDE	5.9%	15.8%	2.0%	15.6%	77.7%	4.8%	1.4%	4.8%	1.7%
Afr2P - EurDE	11.2%	1.8%	0.4%	58.2%	7.6%	0.6%	6.9%	1.1%	0.1%
AfrIN - Eur2P	5.7%	4.6%	76.3%	1.8%	0.4%	9.1%	0.0%	0.0%	0.2%
AfrDE - Eur2P	11.2%	57.3%	6.7%	1.8%	7.6%	0.9%	0.6%	0.6%	0.2%
Afr2P - Eur2P	55.2%	7.6%	0.9%	7.0%	2.7%	0.2%	0.9%	0.2%	0.1%

Prior error rate = 32.41%, Model-choice error a priori = 8/9 = 88.89%

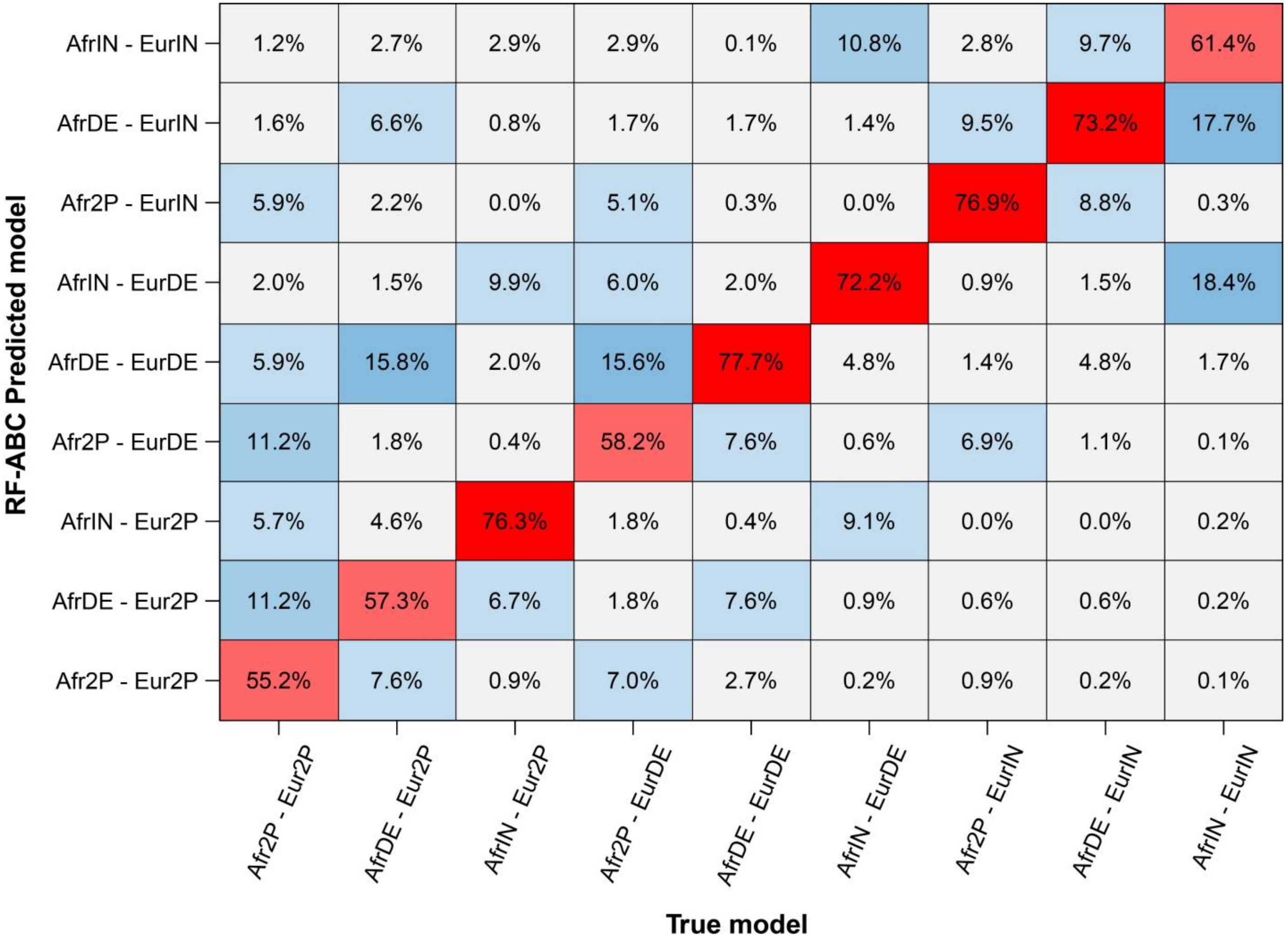
Fortes-Lima et al., <u>Mol Ecol Res</u> 2021

Results: *MetHis*-ABC model-choice with Random-Forest <u>a priori without observed data</u>

- **Random Forest ABC *(abcrf* package in *R, Pudlo et al. 2016)* – 10,000 *MetHis* sims/scenario – 1,000 trees**
- **Cross-validation all simulations**
- **100,000 independent SNPs**
- **Sample sizes:**
  - **Afr Source : 90 indivs**
  - **Eur Source : 89 indivs**
  - **Admixed population H : 50 indivs**

Model – Nestedness (Robert et al. 2011)

- **Erroneous model-choice increased among scenarios qualitatively similar**



	Afr2P - Eur2P	AfrDE - Eur2P	AfrIN - Eur2P	Afr2P - EurDE	AfrDE - EurDE	AfrIN - EurDE	Afr2P - EurIN	AfrDE - EurIN	AfrIN - EurIN
AfrIN - EurIN	1.2%	2.7%	2.9%	2.9%	0.1%	10.8%	2.8%	9.7%	61.4%
AfrDE - EurIN	1.6%	6.6%	0.8%	1.7%	1.7%	1.4%	9.5%	73.2%	17.7%
Afr2P - EurIN	5.9%	2.2%	0.0%	5.1%	0.3%	0.0%	76.9%	8.8%	0.3%
AfrIN - EurDE	2.0%	1.5%	9.9%	6.0%	2.0%	72.2%	0.9%	1.5%	18.4%
AfrDE - EurDE	5.9%	15.8%	2.0%	15.6%	77.7%	4.8%	1.4%	4.8%	1.7%
Afr2P - EurDE	11.2%	1.8%	0.4%	58.2%	7.6%	0.6%	6.9%	1.1%	0.1%
AfrIN - Eur2P	5.7%	4.6%	76.3%	1.8%	0.4%	9.1%	0.0%	0.0%	0.2%
AfrDE - Eur2P	11.2%	57.3%	6.7%	1.8%	7.6%	0.9%	0.6%	0.6%	0.2%
Afr2P - Eur2P	55.2%	7.6%	0.9%	7.0%	2.7%	0.2%	0.9%	0.2%	0.1%

Prior error rate = 32.41%, Model-choice error a priori = 8/9 = 88.89%

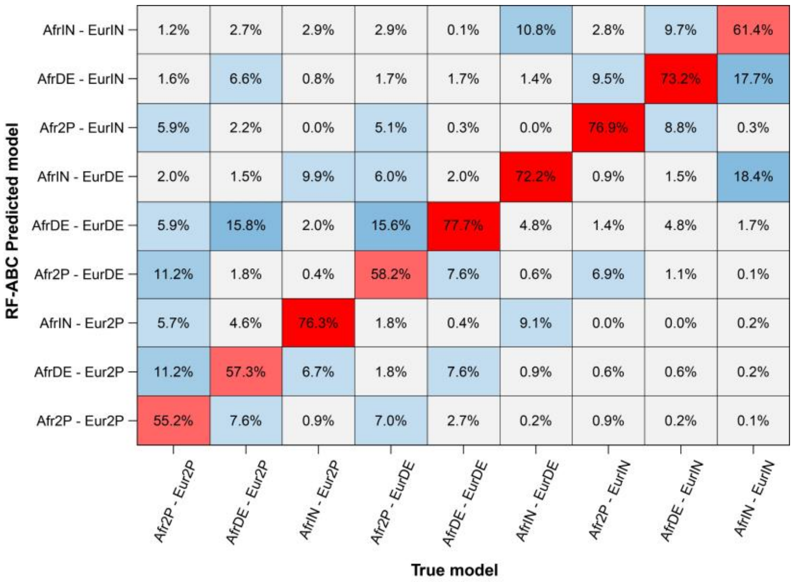
Fortes-Lima et al., <u>Mol Ecol Res</u> 2021

Results: *MetHis*-ABC model-choice with Random-Forest <u>a priori without observed data</u>

- **Random Forest ABC *(abcrf* package in *R, Pudlo et al. 2016)* – 10,000 *MetHis* sims/scenario – 1,000 decision trees in the random forest**
- **Cross-validation all simulations**

- **100,000 independent SNPs**
- **Sample sizes:**

  Afr Source : 90 indivs
  Eur Source : 89 indivs
  Admixed population H : 50 indivs



RF-ABC Predicted model	Afr2P - Eur2P	AfrDE - Eur2P	AfrIN - Eur2P	Afr2P - EurDE	AfrDE - EurDE	AfrIN - EurDE	Afr2P - EurIN	AfrDE - EurIN	AfrIN - EurIN
AfrIN - EurIN	1.2%	2.7%	2.9%	2.9%	0.1%	10.8%	2.8%	9.7%	61.4%
AfrDE - EurIN	1.6%	6.6%	0.8%	1.7%	1.7%	1.4%	9.5%	73.2%	17.7%
Afr2P - EurIN	5.9%	2.2%	0.0%	5.1%	0.3%	0.0%	76.9%	8.8%	0.3%
AfrIN - EurDE	2.0%	1.5%	9.9%	6.0%	2.0%	72.2%	0.9%	1.5%	18.4%
AfrDE - EurDE	5.9%	15.8%	2.0%	15.6%	77.7%	4.8%	1.4%	4.8%	1.7%
Afr2P - EurDE	11.2%	1.8%	0.4%	58.2%	7.6%	0.6%	6.9%	1.1%	0.1%
AfrIN - Eur2P	5.7%	4.6%	76.3%	1.8%	0.4%	9.1%	0.0%	0.0%	0.2%
AfrDE - Eur2P	11.2%	57.3%	6.7%	1.8%	7.6%	0.9%	0.6%	0.6%	0.2%
Afr2P - Eur2P	55.2%	7.6%	0.9%	7.0%	2.7%	0.2%	0.9%	0.2%	0.1%

Prior error rate = **32.41%**
Model-choice error a priori = 8/9 = 88.89%

Fortes-Lima et al., <u>Mol Ecol Res</u> 2021

Results: *MetHis*-ABC model-choice with Random-Forest a priori without observed data

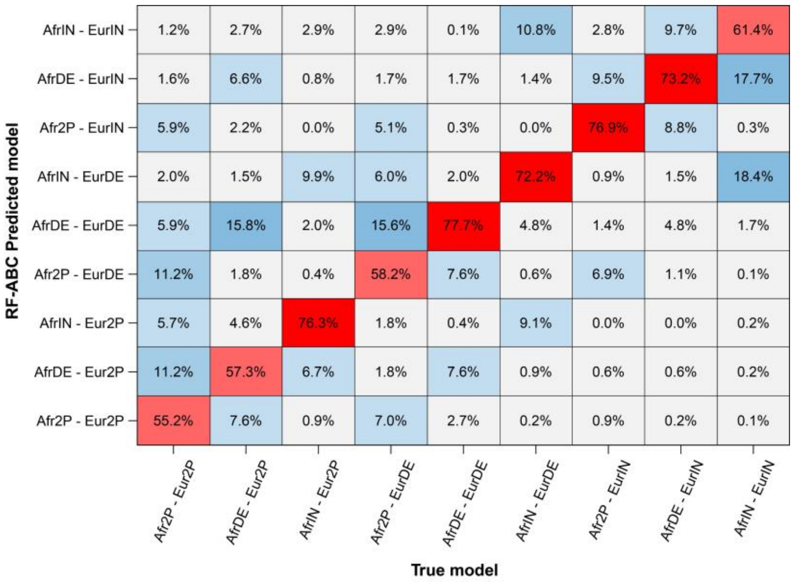
- Random Forest ABC *(abcrf* package in *R, Pudlo et al. 2016)* – 10,000 *MetHis* sims/scenario – 1,000 decision trees in the random forest
- Cross-validation all simulations

Left:

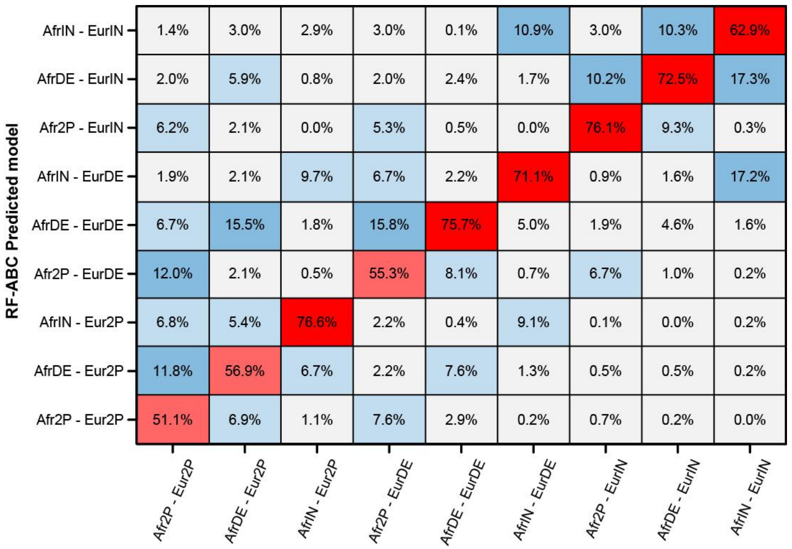
- 100,000 independent SNPs
- Sample sizes:
  Afr Source : 90 indivs
  Eur Source : 89 indivs
  Admixed population H : 50 indivs

Right:

- 50,000 independent SNPs
- Sample sizes:
  Afr Source : 90 indivs
  Eur Source : 89 indivs
  Admixed population H : 50 indivs



RF-ABC Predicted model \ True model	Afr2P - Eur2P	AfrDE - Eur2P	AfrIN - Eur2P	Afr2P - EurDE	AfrDE - EurDE	AfrIN - EurDE	Afr2P - EurIN	AfrDE - EurIN	AfrIN - EurIN
AfrIN - EurIN	1.2%	2.7%	2.9%	2.9%	0.1%	10.8%	2.8%	9.7%	61.4%
AfrDE - EurIN	1.6%	6.6%	0.8%	1.7%	1.7%	1.4%	9.5%	73.2%	17.7%
Afr2P - EurIN	5.9%	2.2%	0.0%	5.1%	0.3%	0.0%	76.9%	8.8%	0.3%
AfrIN - EurDE	2.0%	1.5%	9.9%	6.0%	2.0%	72.2%	0.9%	1.5%	18.4%
AfrDE - EurDE	5.9%	15.8%	2.0%	15.6%	77.7%	4.8%	1.4%	4.8%	1.7%
Afr2P - EurDE	11.2%	1.8%	0.4%	58.2%	7.6%	0.6%	6.9%	1.1%	0.1%
AfrIN - Eur2P	5.7%	4.6%	76.3%	1.8%	0.4%	9.1%	0.0%	0.0%	0.2%
AfrDE - Eur2P	11.2%	57.3%	6.7%	1.8%	7.6%	0.9%	0.6%	0.6%	0.2%
Afr2P - Eur2P	55.2%	7.6%	0.9%	7.0%	2.7%	0.2%	0.9%	0.2%	0.1%



RF-ABC Predicted model	Afr2P - Eur2P	AfrDE - Eur2P	AfrIN - Eur2P	Afr2P - EurDE	AfrDE - EurDE	AfrIN - EurDE	Afr2P - EurIN	AfrDE - EurIN	AfrIN - EurIN
AfrIN - EurIN	1.4%	3.0%	2.9%	3.0%	0.1%	10.9%	3.0%	10.3%	62.9%
AfrDE - EurIN	2.0%	5.9%	0.8%	2.0%	2.4%	1.7%	10.2%	72.5%	17.3%
Afr2P - EurIN	6.2%	2.1%	0.0%	5.3%	0.5%	0.0%	76.1%	9.3%	0.3%
AfrIN - EurDE	1.9%	2.1%	9.7%	6.7%	2.2%	71.1%	0.9%	1.6%	17.2%
AfrDE - EurDE	6.7%	15.5%	1.8%	15.8%	75.7%	5.0%	1.9%	4.6%	1.6%
Afr2P - EurDE	12.0%	2.1%	0.5%	55.3%	8.1%	0.7%	6.7%	1.0%	0.2%
AfrIN - Eur2P	6.8%	5.4%	76.6%	2.2%	0.4%	9.1%	0.1%	0.0%	0.2%
AfrDE - Eur2P	11.8%	56.9%	6.7%	2.2%	7.6%	1.3%	0.5%	0.5%	0.2%
Afr2P - Eur2P	51.1%	6.9%	1.1%	7.6%	2.9%	0.2%	0.7%	0.2%	0.0%

Prior error rate = **32.41%**
Model-choice error a priori = 8/9 = 88.89%

Prior error rate = **33.53%**
Model-choice error a priori = 8/9 = 88.89%

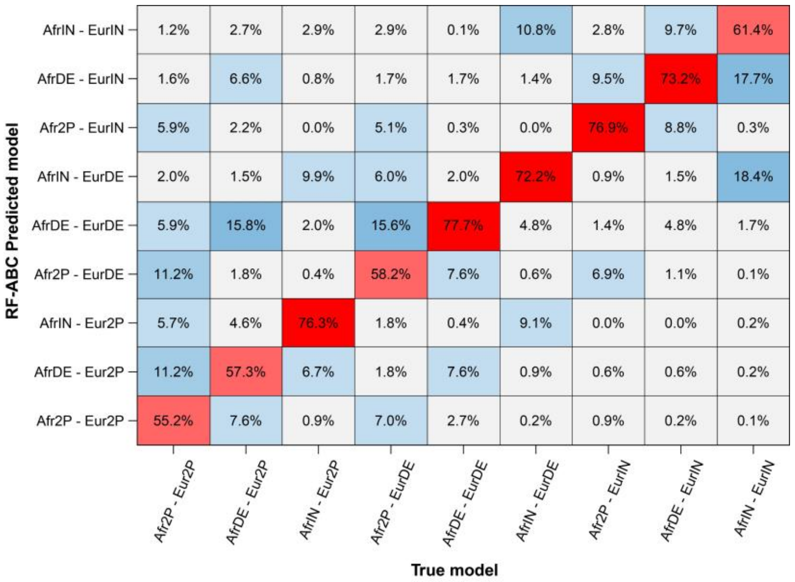
Fortes-Lima et al., Mol Ecol Res 2021

Results: *MetHis*-ABC model-choice with Random-Forest a priori without observed data

- Random Forest ABC *(abcrf* package in *R, Pudlo et al. 2016)* – 10,000 *MetHis* sims/scenario – 1,000 decision trees in the random forest
- Cross-validation all simulations

- 100,000 independent SNPs
- Sample sizes:

  Afr Source : 90 indivs
  Eur Source : 89 indivs
  Admixed population H : 50 indivs

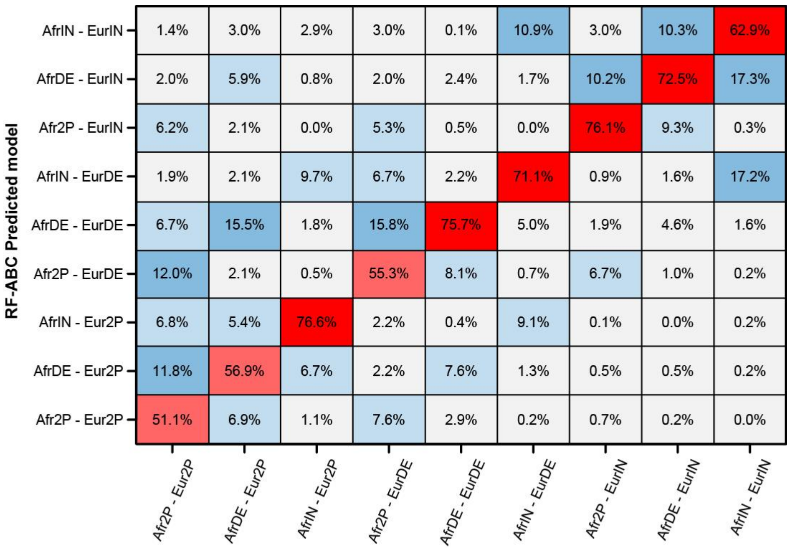


	Afr2P - Eur2P	AfrDE - Eur2P	AfrIN - Eur2P	Afr2P - EurDE	AfrDE - EurDE	AfrIN - EurDE	Afr2P - EurIN	AfrDE - EurIN	AfrIN - EurIN
AfrIN - EurIN	1.2%	2.7%	2.9%	2.9%	0.1%	10.8%	2.8%	9.7%	61.4%
AfrDE - EurIN	1.6%	6.6%	0.8%	1.7%	1.7%	1.4%	9.5%	73.2%	17.7%
Afr2P - EurIN	5.9%	2.2%	0.0%	5.1%	0.3%	0.0%	76.9%	8.8%	0.3%
AfrIN - EurDE	2.0%	1.5%	9.9%	6.0%	2.0%	72.2%	0.9%	1.5%	18.4%
AfrDE - EurDE	5.9%	15.8%	2.0%	15.6%	77.7%	4.8%	1.4%	4.8%	1.7%
Afr2P - EurDE	11.2%	1.8%	0.4%	58.2%	7.6%	0.6%	6.9%	1.1%	0.1%
AfrIN - Eur2P	5.7%	4.6%	76.3%	1.8%	0.4%	9.1%	0.0%	0.0%	0.2%
AfrDE - Eur2P	11.2%	57.3%	6.7%	1.8%	7.6%	0.9%	0.6%	0.6%	0.2%
Afr2P - Eur2P	55.2%	7.6%	0.9%	7.0%	2.7%	0.2%	0.9%	0.2%	0.1%

Prior error rate = **32.41%**
Model-choice error a priori = 8/9 = 88.89%

- 50,000 independent SNPs
- Sample sizes:

  Afr Source : 90 indivs
  Eur Source : 89 indivs
  Admixed population H : 50 indivs

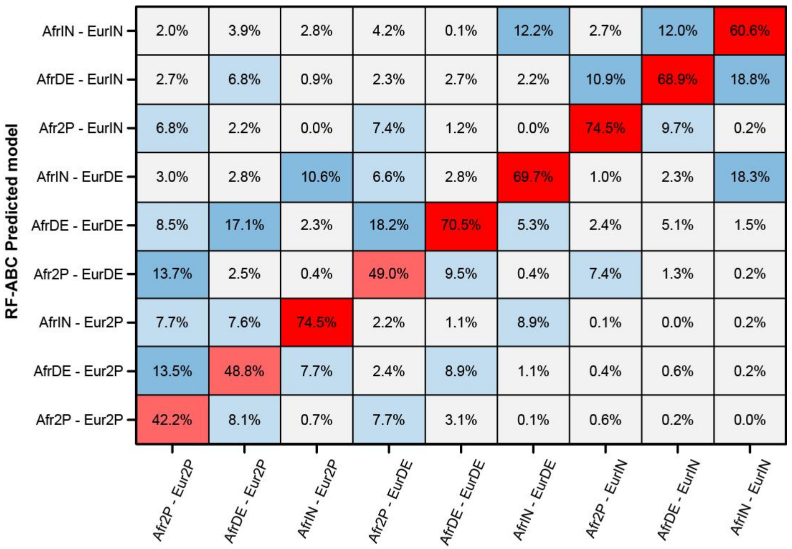


	Afr2P - Eur2P	AfrDE - Eur2P	AfrIN - Eur2P	Afr2P - EurDE	AfrDE - EurDE	AfrIN - EurDE	Afr2P - EurIN	AfrDE - EurIN	AfrIN - EurIN
AfrIN - EurIN	1.4%	3.0%	2.9%	3.0%	0.1%	10.9%	3.0%	10.3%	62.9%
AfrDE - EurIN	2.0%	5.9%	0.8%	2.0%	2.4%	1.7%	10.2%	72.5%	17.3%
Afr2P - EurIN	6.2%	2.1%	0.0%	5.3%	0.5%	0.0%	76.1%	9.3%	0.3%
AfrIN - EurDE	1.9%	2.1%	9.7%	6.7%	2.2%	71.1%	0.9%	1.6%	17.2%
AfrDE - EurDE	6.7%	15.5%	1.8%	15.8%	75.7%	5.0%	1.9%	4.6%	1.6%
Afr2P - EurDE	12.0%	2.1%	0.5%	55.3%	8.1%	0.7%	6.7%	1.0%	0.2%
AfrIN - Eur2P	6.8%	5.4%	76.6%	2.2%	0.4%	9.1%	0.1%	0.0%	0.2%
AfrDE - Eur2P	11.8%	56.9%	6.7%	2.2%	7.6%	1.3%	0.5%	0.5%	0.2%
Afr2P - Eur2P	51.1%	6.9%	1.1%	7.6%	2.9%	0.2%	0.7%	0.2%	0.0%

Prior error rate = **33.53%**
Model-choice error a priori = 8/9 = 88.89%

- 10,000 independent SNPs
- Sample sizes:

  Afr Source : 90 indivs
  Eur Source : 89 indivs
  Admixed population H : 50 indivs



	Afr2P - Eur2P	AfrDE - Eur2P	AfrIN - Eur2P	Afr2P - EurDE	AfrDE - EurDE	AfrIN - EurDE	Afr2P - EurIN	AfrDE - EurIN	AfrIN - EurIN
AfrIN - EurIN	2.0%	3.9%	2.8%	4.2%	0.1%	12.2%	2.7%	12.0%	60.6%
AfrDE - EurIN	2.7%	6.8%	0.9%	2.3%	2.7%	2.2%	10.9%	68.9%	18.8%
Afr2P - EurIN	6.8%	2.2%	0.0%	7.4%	1.2%	0.0%	74.5%	9.7%	0.2%
AfrIN - EurDE	3.0%	2.8%	10.6%	6.6%	2.8%	69.7%	1.0%	2.3%	18.3%
AfrDE - EurDE	8.5%	17.1%	2.3%	18.2%	70.5%	5.3%	2.4%	5.1%	1.5%
Afr2P - EurDE	13.7%	2.5%	0.4%	49.0%	9.5%	0.4%	7.4%	1.3%	0.2%
AfrIN - Eur2P	7.7%	7.6%	74.5%	2.2%	1.1%	8.9%	0.1%	0.0%	0.2%
AfrDE - Eur2P	13.5%	48.8%	7.7%	2.4%	8.9%	1.1%	0.4%	0.6%	0.2%
Afr2P - Eur2P	42.2%	8.1%	0.7%	7.7%	3.1%	0.1%	0.6%	0.2%	0.0%

Prior error rate = **37.93%**
Model-choice error a priori = 8/9 = 88.89%

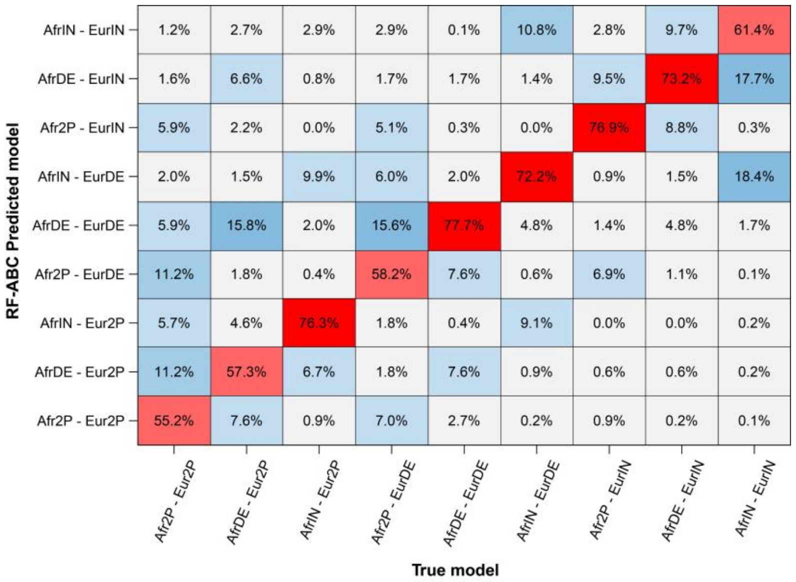
Fortes-Lima et al., Mol Ecol Res 2021

Results: *MetHis*-ABC model-choice with Random-Forest a priori without observed data

- Random Forest ABC *(abcrf* package in *R, Pudlo et al. 2016)* – 10,000 *MetHis* sims/scenario – 1,000 decision trees in the random forest
- Cross-validation all simulations

- 100,000 independent SNPs
- Sample sizes:

  Afr Source : 90 indivs
  Eur Source : 89 indivs
  Admixed population H : 50 indivs

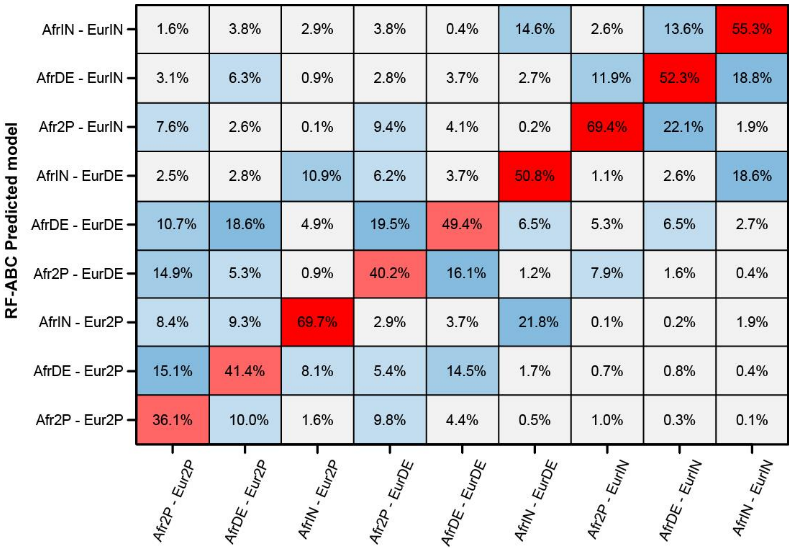


	Afr2P - Eur2P	AfrDE - Eur2P	AfrIN - Eur2P	Afr2P - EurDE	AfrDE - EurDE	AfrIN - EurDE	Afr2P - EurIN	AfrDE - EurIN	AfrIN - EurIN
AfrIN - EurIN	1.2%	2.7%	2.9%	2.9%	0.1%	10.8%	2.8%	9.7%	61.4%
AfrDE - EurIN	1.6%	6.6%	0.8%	1.7%	1.7%	1.4%	9.5%	73.2%	17.7%
Afr2P - EurIN	5.9%	2.2%	0.0%	5.1%	0.3%	0.0%	76.9%	8.8%	0.3%
AfrIN - EurDE	2.0%	1.5%	9.9%	6.0%	2.0%	72.2%	0.9%	1.5%	18.4%
AfrDE - EurDE	5.9%	15.8%	2.0%	15.6%	77.7%	4.8%	1.4%	4.8%	1.7%
Afr2P - EurDE	11.2%	1.8%	0.4%	58.2%	7.6%	0.6%	6.9%	1.1%	0.1%
AfrIN - Eur2P	5.7%	4.6%	76.3%	1.8%	0.4%	9.1%	0.0%	0.0%	0.2%
AfrDE - Eur2P	11.2%	57.3%	6.7%	1.8%	7.6%	0.9%	0.6%	0.6%	0.2%
Afr2P - Eur2P	55.2%	7.6%	0.9%	7.0%	2.7%	0.2%	0.9%	0.2%	0.1%

Prior error rate = **32.41%**
Model-choice error a priori = 8/9 = 88.89%

- 100,000 independent SNPs
- Sample sizes:

  Afr Source : 18 indivs
  Eur Source : 18 indivs
  Admixed population H : 10 indivs



	Afr2P - Eur2P	AfrDE - Eur2P	AfrIN - Eur2P	Afr2P - EurDE	AfrDE - EurDE	AfrIN - EurDE	Afr2P - EurIN	AfrDE - EurIN	AfrIN - EurIN
AfrIN - EurIN	1.6%	3.8%	2.9%	3.8%	0.4%	14.6%	2.6%	13.6%	55.3%
AfrDE - EurIN	3.1%	6.3%	0.9%	2.8%	3.7%	2.7%	11.9%	52.3%	18.8%
Afr2P - EurIN	7.6%	2.6%	0.1%	9.4%	4.1%	0.2%	69.4%	22.1%	1.9%
AfrIN - EurDE	2.5%	2.8%	10.9%	6.2%	3.7%	50.8%	1.1%	2.6%	18.6%
AfrDE - EurDE	10.7%	18.6%	4.9%	19.5%	49.4%	6.5%	5.3%	6.5%	2.7%
Afr2P - EurDE	14.9%	5.3%	0.9%	40.2%	16.1%	1.2%	7.9%	1.6%	0.4%
AfrIN - Eur2P	8.4%	9.3%	69.7%	2.9%	3.7%	21.8%	0.1%	0.2%	1.9%
AfrDE - Eur2P	15.1%	41.4%	8.1%	5.4%	14.5%	1.7%	0.7%	0.8%	0.4%
Afr2P - Eur2P	36.1%	10.0%	1.6%	9.8%	4.4%	0.5%	1.0%	0.3%	0.1%

Prior error rate = **48.39%**
Model-choice error a priori = 8/9 = 88.89%

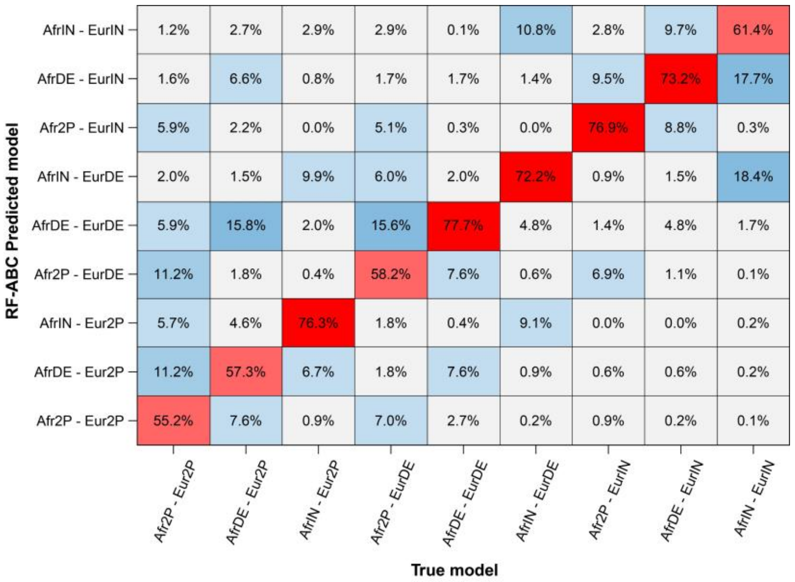
ABC relies on summary-statistics informativeness rather than absolute amount of data

Fortes-Lima et al., Mol Ecol Res *2021*

Results: *MetHis*-ABC model-choice with Random-Forest a priori without observed data

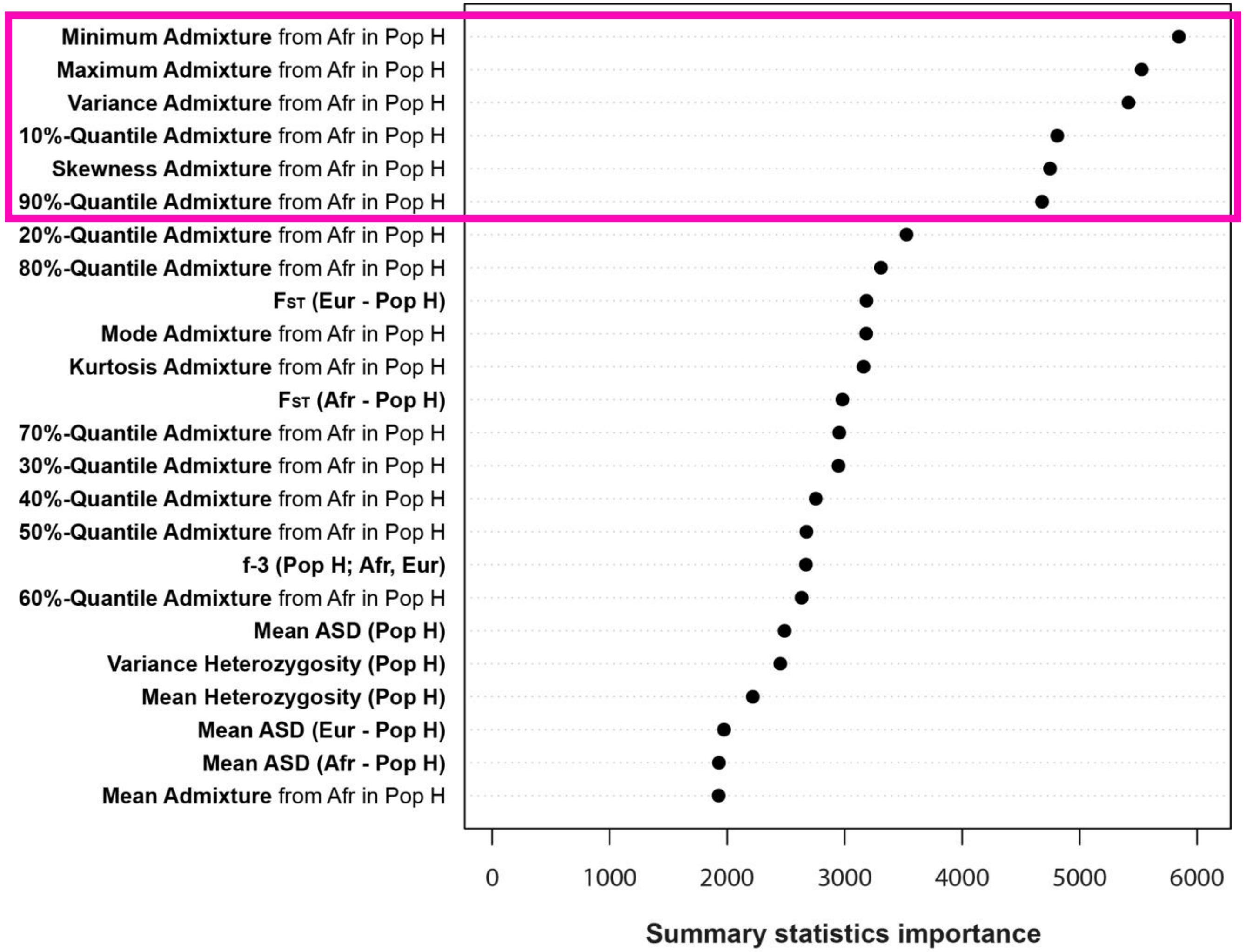
- Random Forest ABC *(abcrf* package in *R, Pudlo et al. 2016)* – 10,000 *MetHis* sims/scenario – 1,000 decision trees in the random forest
- Cross-validation all simulations
  - 100,000 independent SNPs
  - Sample sizes:

    Afr Source : 90 indivs
Eur Source : 89 indivs
Admixed population H : 50 indivs



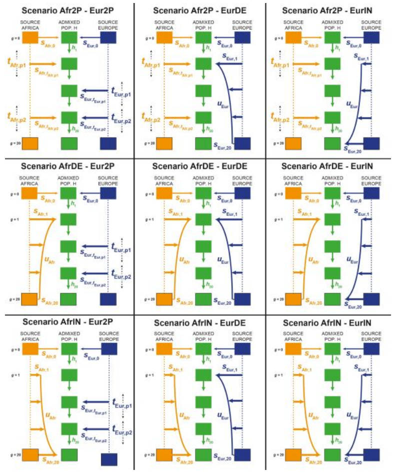
RF-ABC Predicted model \ True model	Afr2P - Eur2P	AfrDE - Eur2P	AfrIN - Eur2P	Afr2P - EurDE	AfrDE - EurDE	AfrIN - EurDE	Afr2P - EurIN	AfrDE - EurIN	AfrIN - EurIN
AfrIN - EurIN	1.2%	2.7%	2.9%	2.9%	0.1%	10.8%	2.8%	9.7%	61.4%
AfrDE - EurIN	1.6%	6.6%	0.8%	1.7%	1.7%	1.4%	9.5%	73.2%	17.7%
Afr2P - EurIN	5.9%	2.2%	0.0%	5.1%	0.3%	0.0%	76.9%	8.8%	0.3%
AfrIN - EurDE	2.0%	1.5%	9.9%	6.0%	2.0%	72.2%	0.9%	1.5%	18.4%
AfrDE - EurDE	5.9%	15.8%	2.0%	15.6%	77.7%	4.8%	1.4%	4.8%	1.7%
Afr2P - EurDE	11.2%	1.8%	0.4%	58.2%	7.6%	0.6%	6.9%	1.1%	0.1%
AfrIN - Eur2P	5.7%	4.6%	76.3%	1.8%	0.4%	9.1%	0.0%	0.0%	0.2%
AfrDE - Eur2P	11.2%	57.3%	6.7%	1.8%	7.6%	0.9%	0.6%	0.6%	0.2%
Afr2P - Eur2P	55.2%	7.6%	0.9%	7.0%	2.7%	0.2%	0.9%	0.2%	0.1%

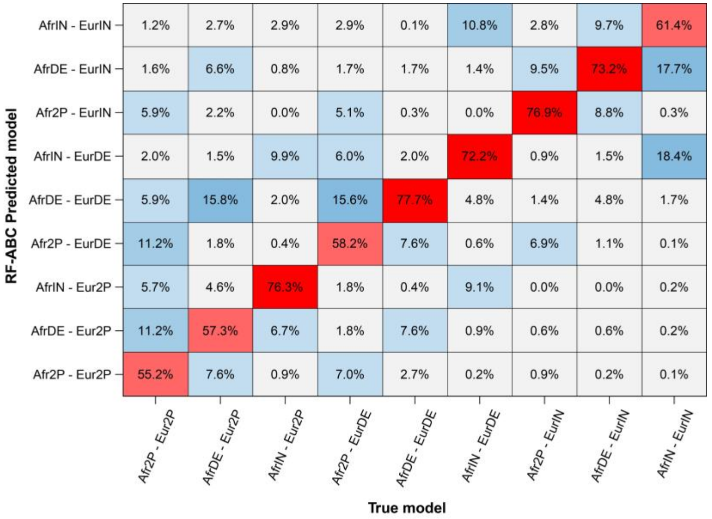
Prior error rate = **32.41%**
Model-choice error a priori = 8/9 = 88.89%



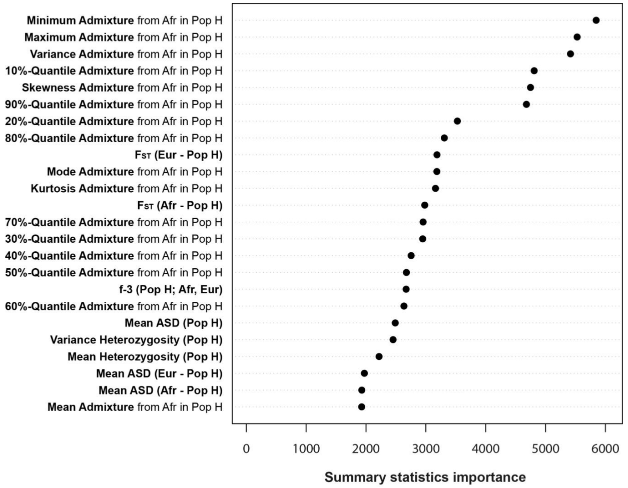
Fortes-Lima et al., Mol Ecol Res 2021

Results: *MetHis*-ABC model-choice with Random-Forest a priori without observed data





RF-ABC Predicted model \ True model	Afr2P - Eur2P	AfrDE - Eur2P	AfrIN - Eur2P	Afr2P - EurDE	AfrDE - EurDE	AfrIN - EurDE	Afr2P - EurIN	AfrDE - EurIN	AfrIN - EurIN
AfrIN - EurIN	1.2%	2.7%	2.9%	2.9%	0.1%	10.8%	2.8%	9.7%	61.4%
AfrDE - EurIN	1.6%	6.6%	0.8%	1.7%	1.7%	1.4%	9.5%	73.2%	17.7%
Afr2P - EurIN	5.9%	2.2%	0.0%	5.1%	0.3%	0.0%	76.9%	8.8%	0.3%
AfrIN - EurDE	2.0%	1.5%	9.9%	6.0%	2.0%	72.2%	0.9%	1.5%	18.4%
AfrDE - EurDE	5.9%	15.8%	2.0%	15.6%	77.7%	4.8%	1.4%	4.8%	1.7%
Afr2P - EurDE	11.2%	1.8%	0.4%	58.2%	7.6%	0.6%	6.9%	1.1%	0.1%
AfrIN - Eur2P	5.7%	4.6%	76.3%	1.8%	0.4%	9.1%	0.0%	0.0%	0.2%
AfrDE - Eur2P	11.2%	57.3%	6.7%	1.8%	7.6%	0.9%	0.6%	0.6%	0.2%
Afr2P - Eur2P	55.2%	7.6%	0.9%	7.0%	2.7%	0.2%	0.9%	0.2%	0.1%



- ***MetHis*-ABC Random-Forest model-choice a priori powerful to distinguish highly complex historical admixture models**
- **Errors are made in the parameter-space where models are highly nested and thus biologically similar**
- ***MetHis*-ABC Random-Forest model-choice performances are robust to reduced SNP and Sample sets** -> ABC relies on summary-statistics informativeness rather than absolute amount of data
- **Distribution of admixture fractions is highly informative for admixture history inference**

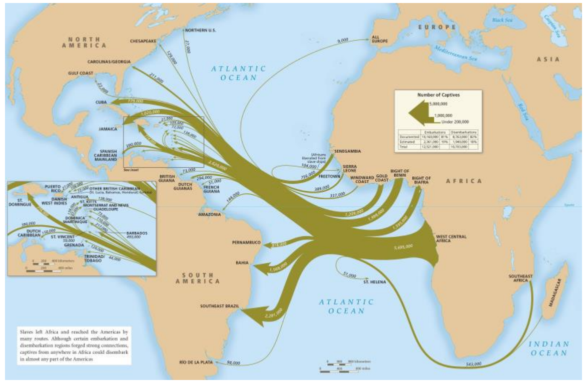
Fortes-Lima et al., Mol Ecol Res 2021

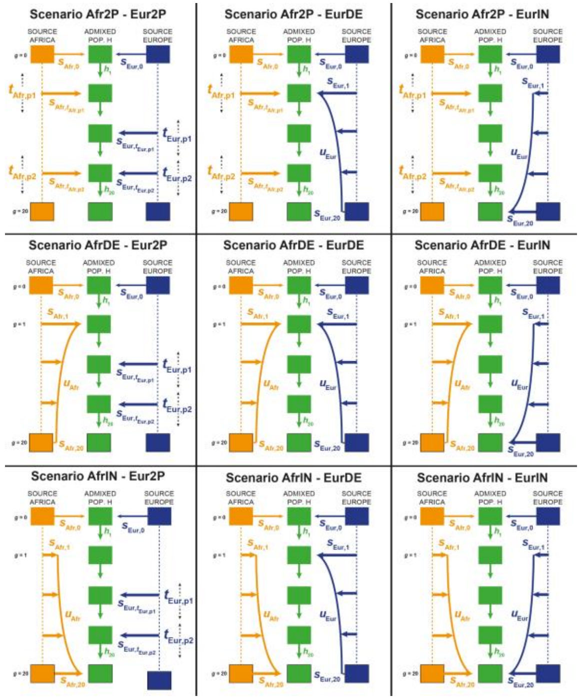
Results: *MetHis*-ABC complex admixture history of Barbadian (ACB) population

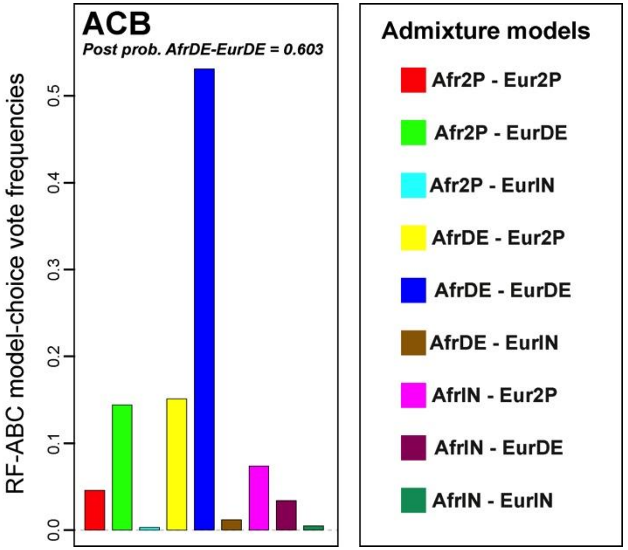
Model-choice
Random Forest ABC
abcrf package in *R* (Pudlo et al. 2016)

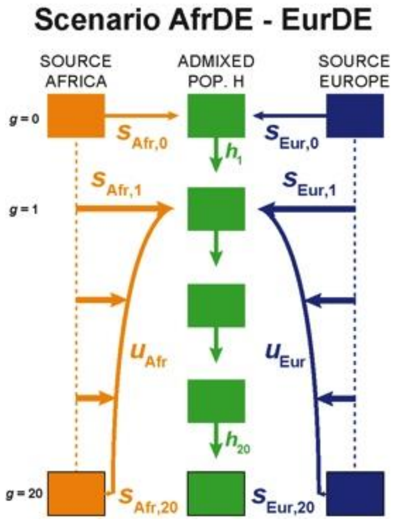
1,000 decision trees in the forest

9 competing scenarios
10,000 *MetHis* sims/scenario
100,000 independent SNPs
24 summary-statistics









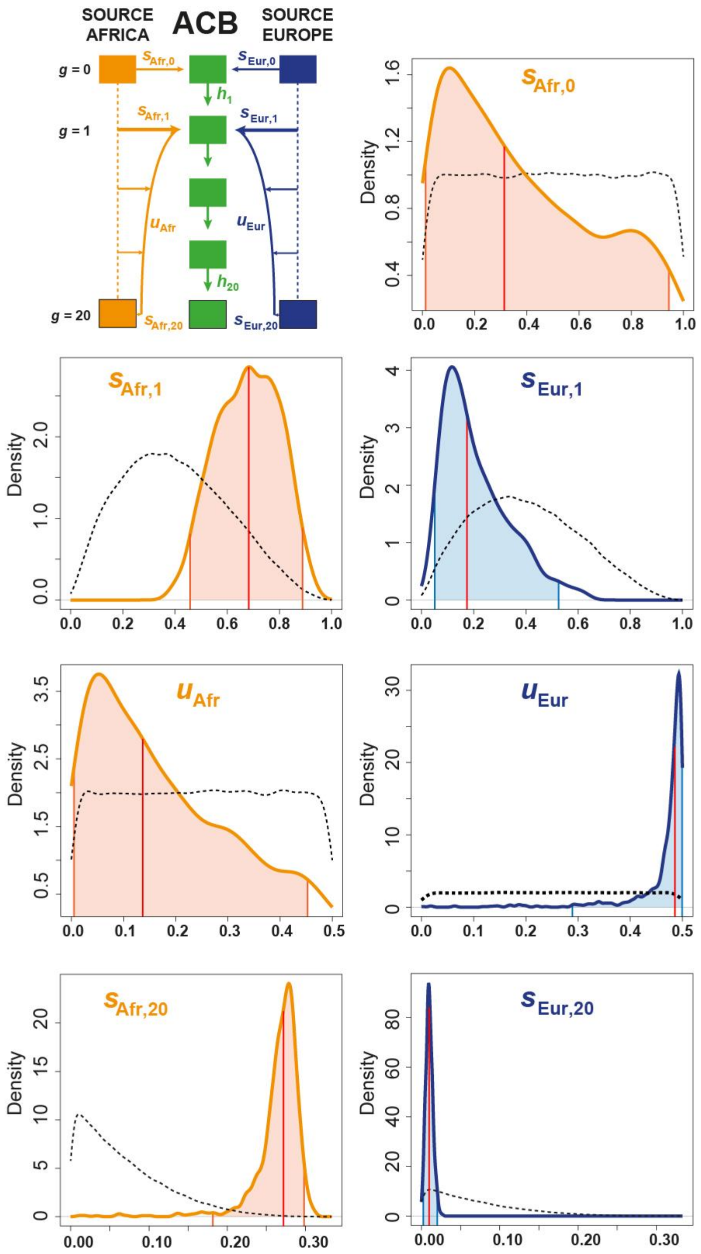
Fortes-Lima et al., Mol Ecol Res 2021

Results: *MetHis*-ABC complex admixture history of Barbadian (ACB) population

Parameter-inference
Neural Network ABC
abc package in *R* (Csilléry et al. 2012)

Tolerance 1% (1,000 sims closest to obs.)
4 neurons in the hidden layer

1 best scenario
100,000 *MetHis* sims
100,000 independent SNPs
24 summary-statistics



Fortes-Lima et al., Mol Ecol Res 2021

Results: *MetHis*-ABC complex admixture history of Barbadian (ACB) population

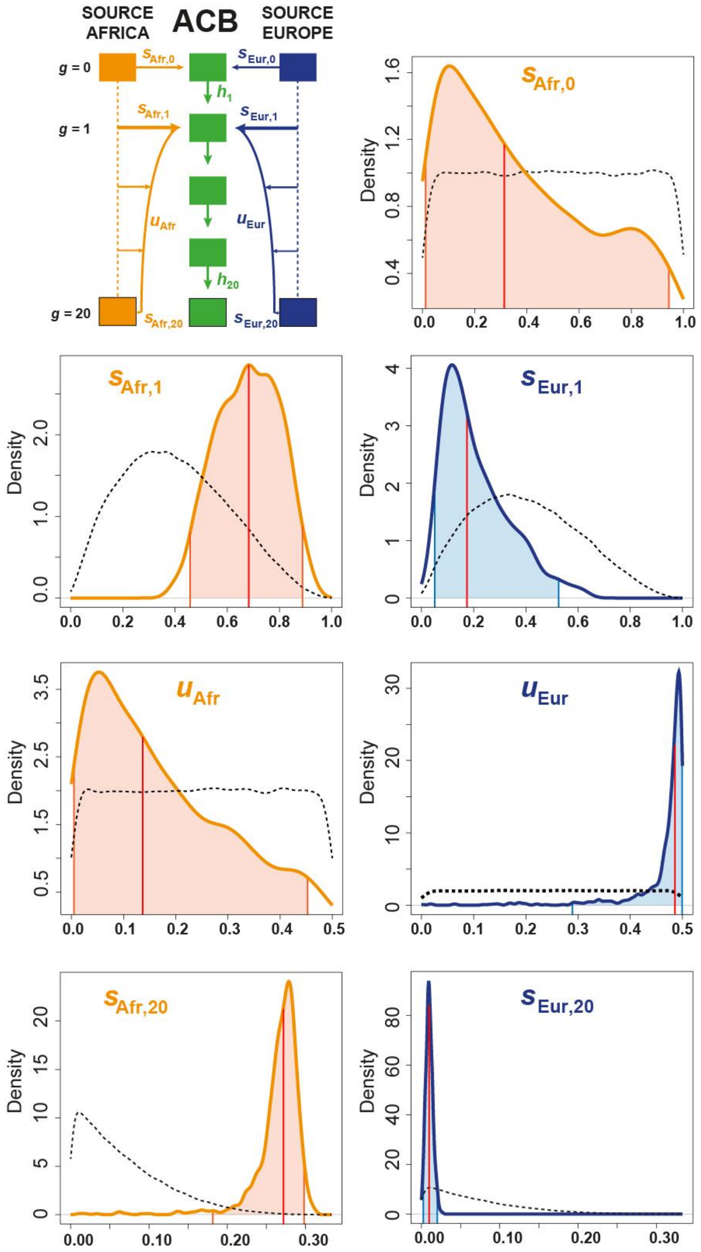
Parameter-inference
Neural Network ABC
abc package in *R* (Csilléry et al. 2012)

Tolerance 1% (1,000 sims closest to obs.)
4 neurons in the hidden layer

1 best scenario
100,000 *MetHis* sims
100,000 independent SNPs
24 summary-statistics

Cross-validation post param error
1,000 closest simulations in turn used as controlled pseudo-observed data for NN-ABC param inference

AfrDE-EurDE parameters	ACB	
	Av. absolute Error	Mean-square Error / Var.
$s_{Afr,0}$	0.2530	1.0070
$s_{Afr,1}$	**0.1206**	0.8533
$s_{Afr,20}$	**0.0274**	0.4162
u_{Afr}	0.1166	0.9974
$s_{Eur,1}$	**0.0952**	1.0526
$s_{Eur,20}$	**0.0044**	0.6452
u_{Eur}	0.1084	0.9431



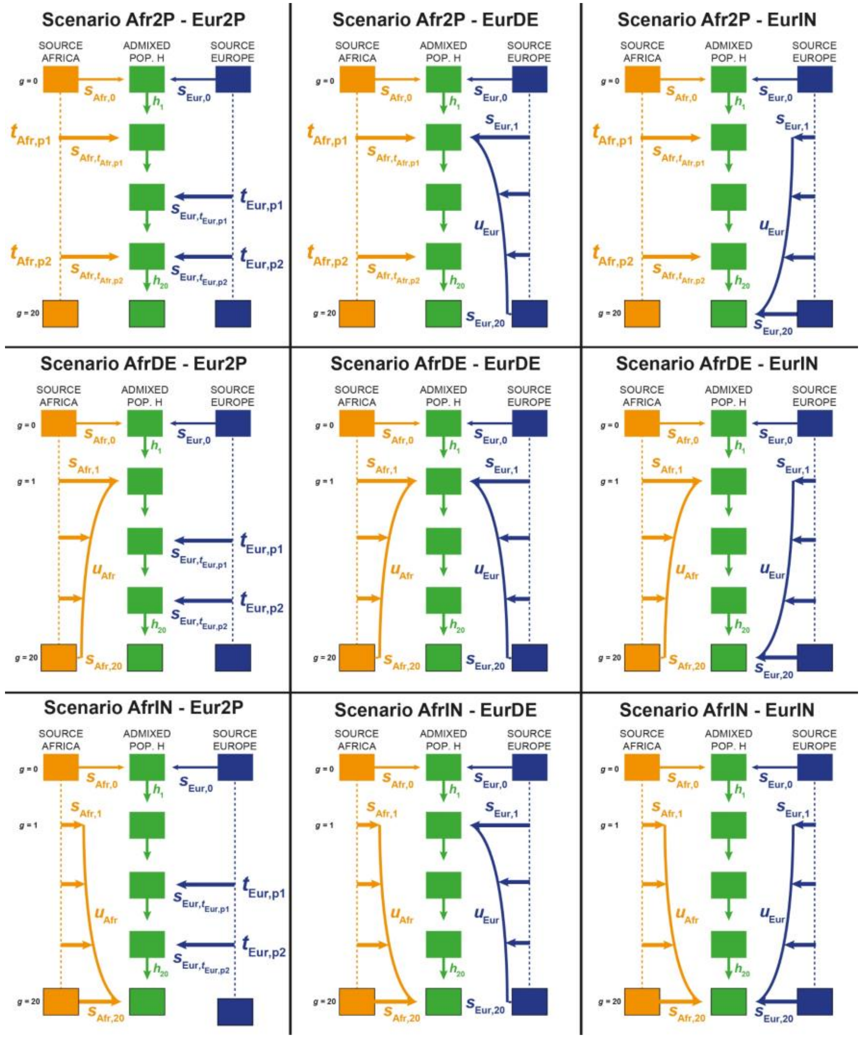
Fortes-Lima et al., Mol Ecol Res 2021

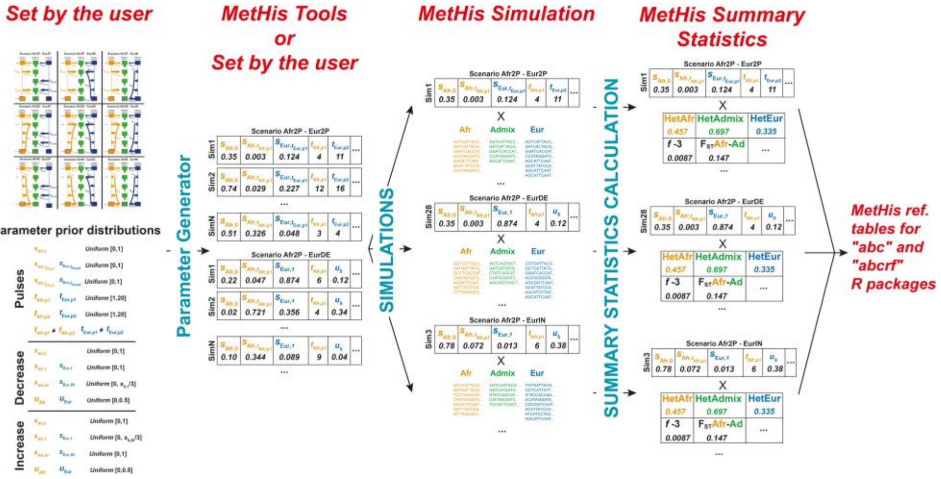
Conclusions

MetHis **– RF-ABC is successful in distinguishing a priori among competing highly complex admixture models**

Admixture distribution is highly informative for ABC model choice as expected theoretically

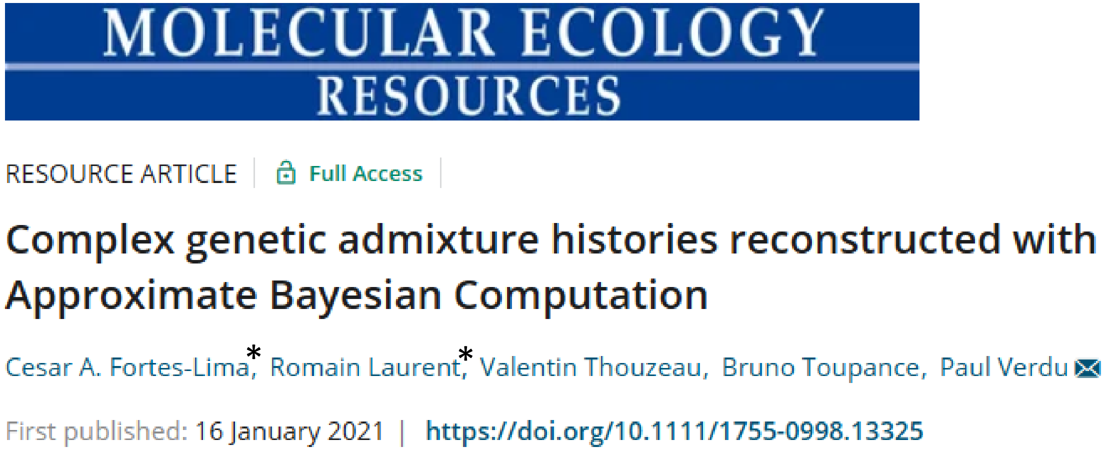
MetHis **– NN-ABC produces accurate posterior parameter estimation and relatively conservative 95% CI inference**





Fortes-Lima et al., Mol Ecol Res 2021

Conclusions



MOLECULAR ECOLOGY RESOURCES

RESOURCE ARTICLE | Full Access

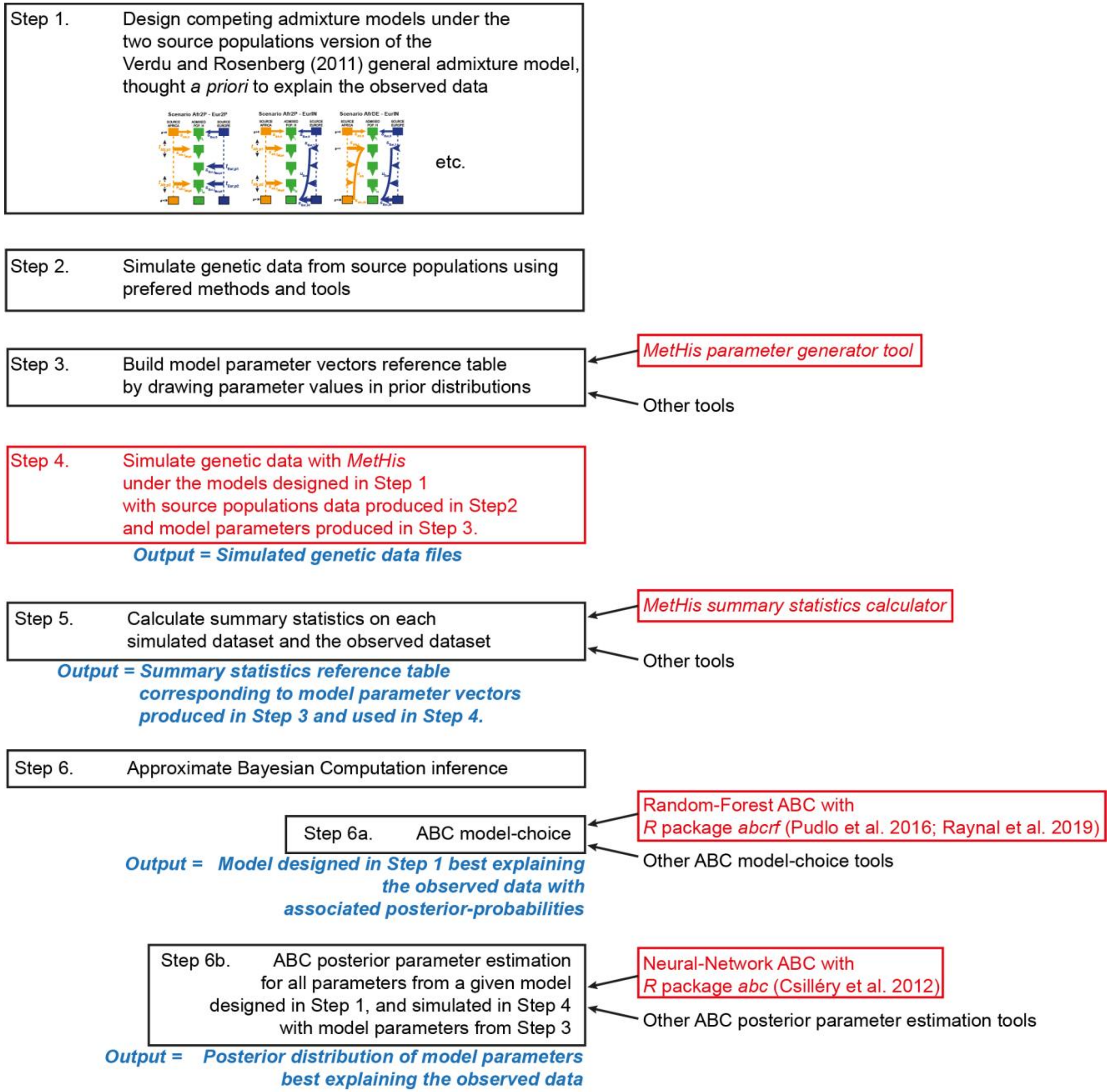
Complex genetic admixture histories reconstructed with Approximate Bayesian Computation

Cesar A. Fortes-Lima*, Romain Laurent*, Valentin Thouzeau, Bruno Toupance, Paul Verdu

First published: 16 January 2021 | https://doi.org/10.1111/1755-0998.13325

<u>Special issue *Machine Learning in Molecular Ecology*</u>

https://github.com/romain-laurent/MetHis



Step 1. Design competing admixture models under the two source populations version of the Verdu and Rosenberg (2011) general admixture model, thought *a priori* to explain the observed data

etc.

Step 2. Simulate genetic data from source populations using prefered methods and tools

Step 3. Build model parameter vectors reference table by drawing parameter values in prior distributions

MetHis parameter generator tool

Other tools

Step 4. Simulate genetic data with *MetHis* under the models designed in Step 1 with source populations data produced in Step2 and model parameters produced in Step 3.

Output = Simulated genetic data files

Step 5. Calculate summary statistics on each simulated dataset and the observed dataset

Output = Summary statistics reference table corresponding to model parameter vectors produced in Step 3 and used in Step 4.

MetHis summary statistics calculator

Other tools

Step 6. Approximate Bayesian Computation inference

Step 6a. ABC model-choice

Output = Model designed in Step 1 best explaining the observed data with associated posterior-probabilities

Random-Forest ABC with *R* package *abcrf* (Pudlo et al. 2016; Raynal et al. 2019)

Other ABC model-choice tools

Step 6b. ABC posterior parameter estimation for all parameters from a given model designed in Step 1, and simulated in Step 4 with model parameters from Step 3

Output = Posterior distribution of model parameters best explaining the observed data

Neural-Network ABC with *R* package *abc* (Csilléry et al. 2012)

Other ABC posterior parameter estimation tools

Fortes-Lima et al., <u>Mol Ecol Res</u> 2021

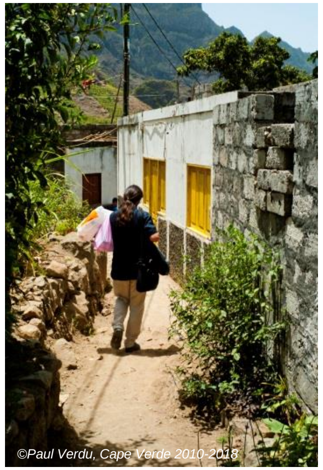
The admixture histories of Cabo Verde
Laurent et al. *Nature Communications* (under revision), BioRxiv https://doi.org/10.1101/2022.04.11.487833

Reconstructing the admixture history of Cabo Verde with Romain Laurent (UMR7206), Noah Rosenberg (Stanford University) and Marlyse Baptista (University of Michigan)

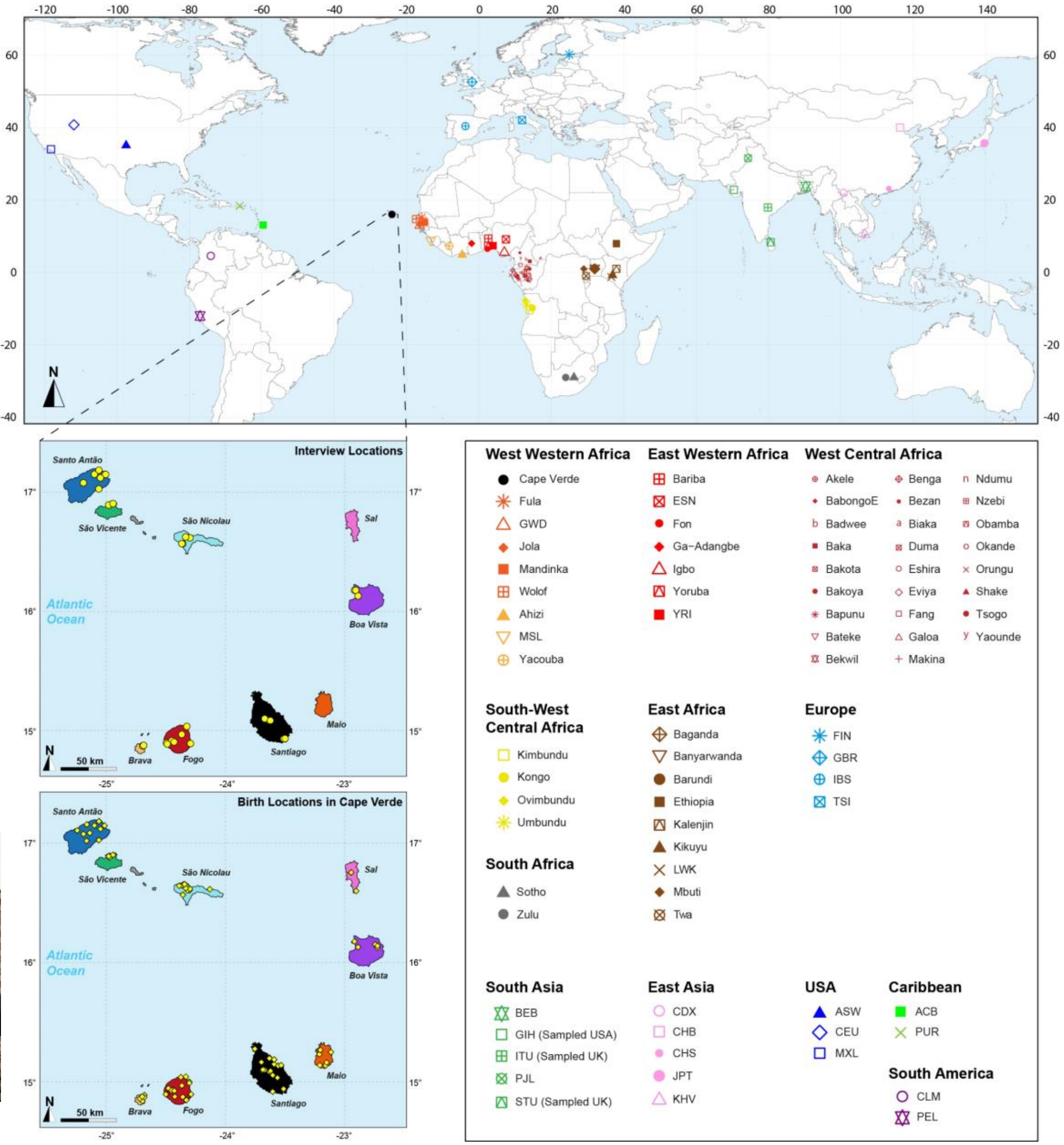
©Paul Verdu, Cape Verde 2010-2018

©Paul Verdu, Cape Verde 2010-2018

©Paul Verdu, Cape Verde 2010-2018

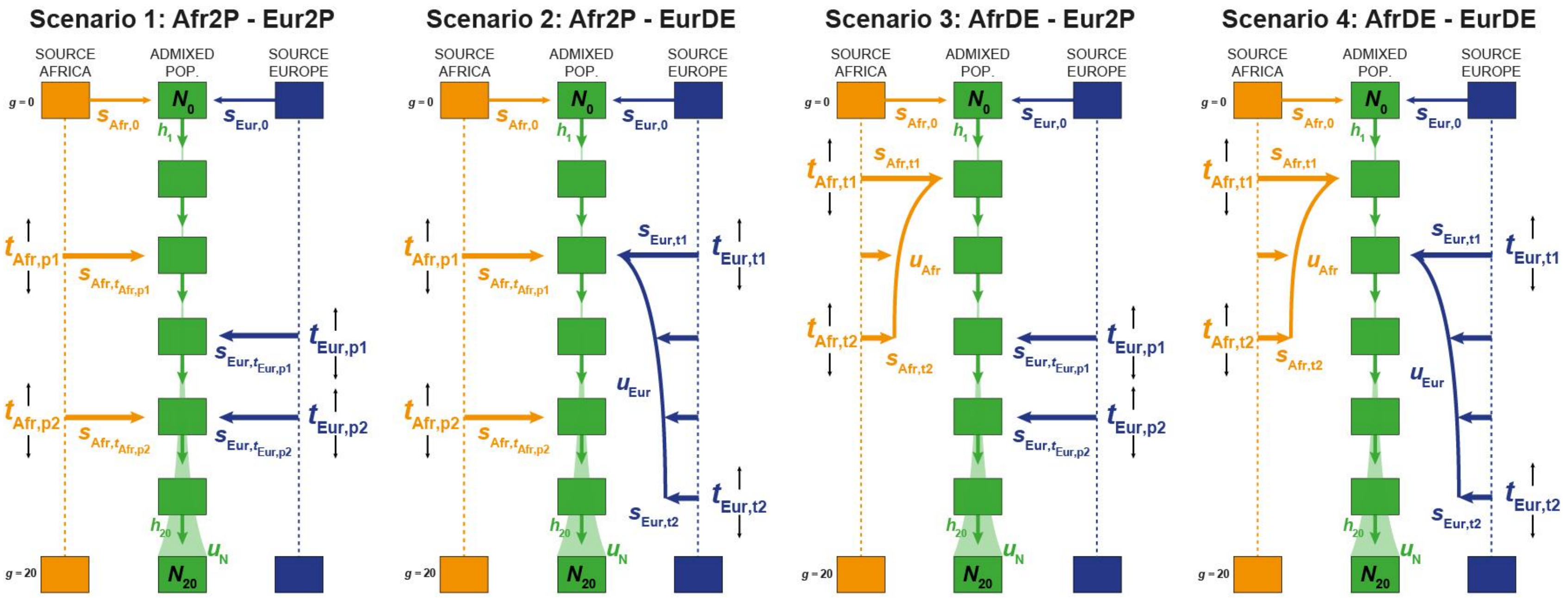

©Paul Verdu, Cape Verde 2010-2018

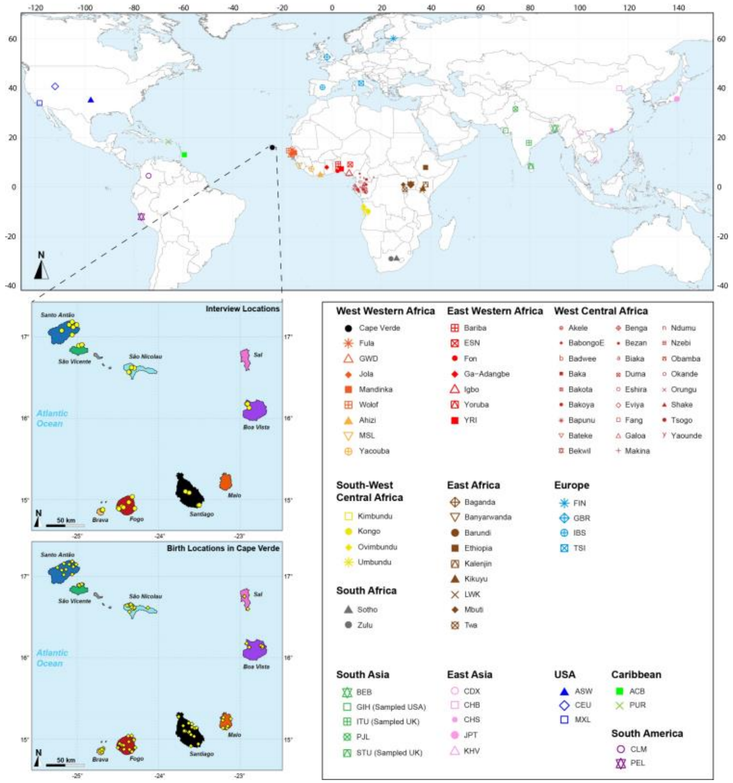
©Paul Verdu, Cape Verde 2010-2018



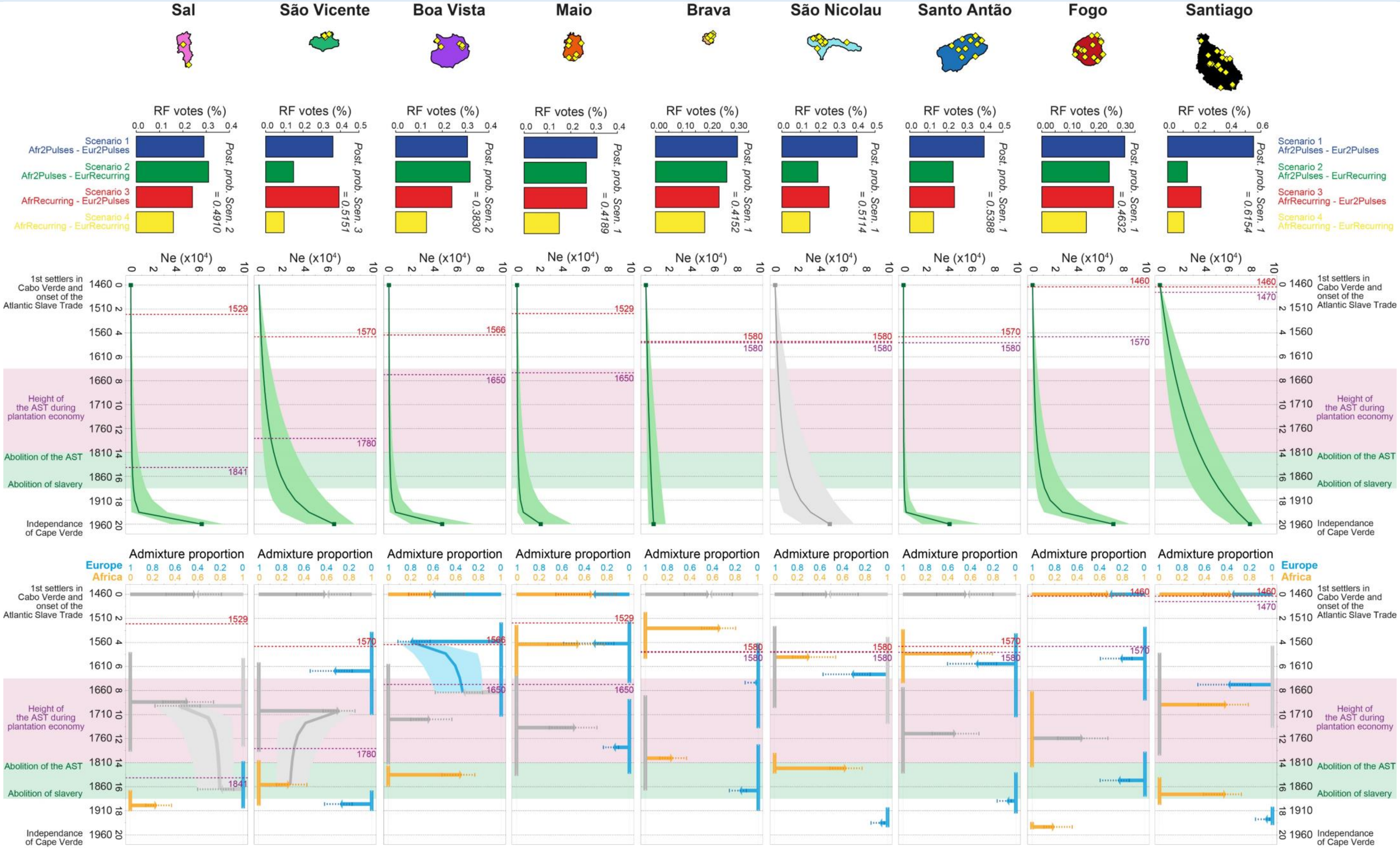
The admixture histories of Cabo Verde

Laurent et al. *Nature Communications* (under revision), BioRxiv https://doi.org/10.1101/2022.04.11.487833



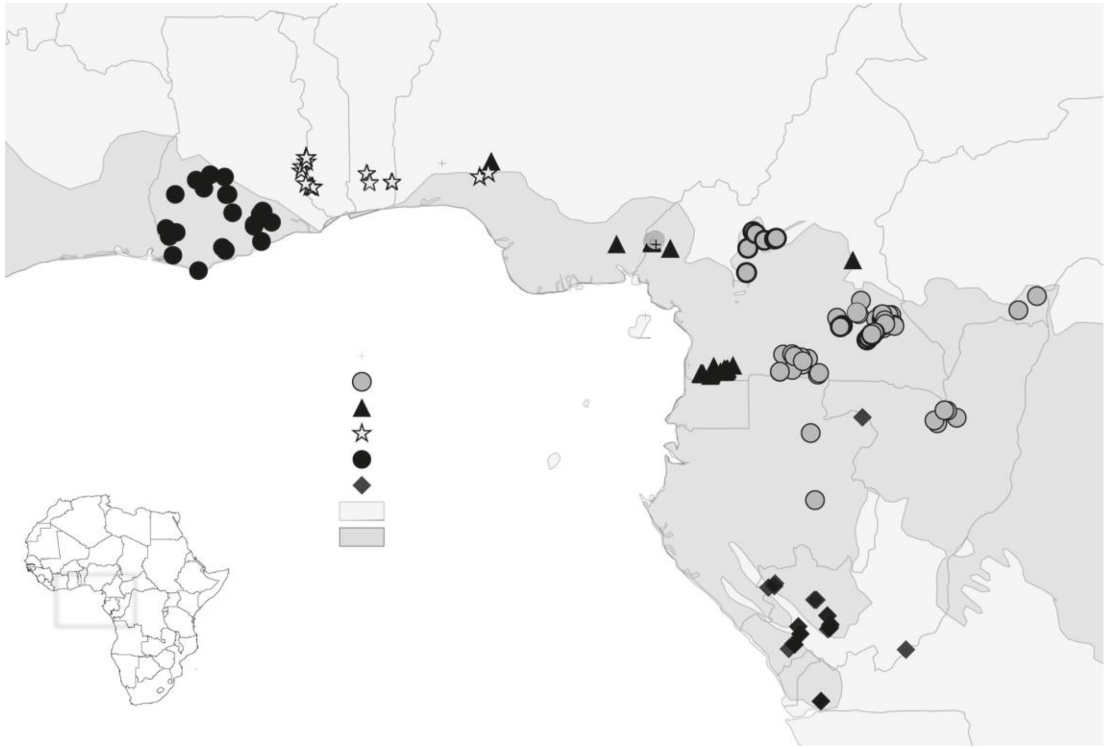


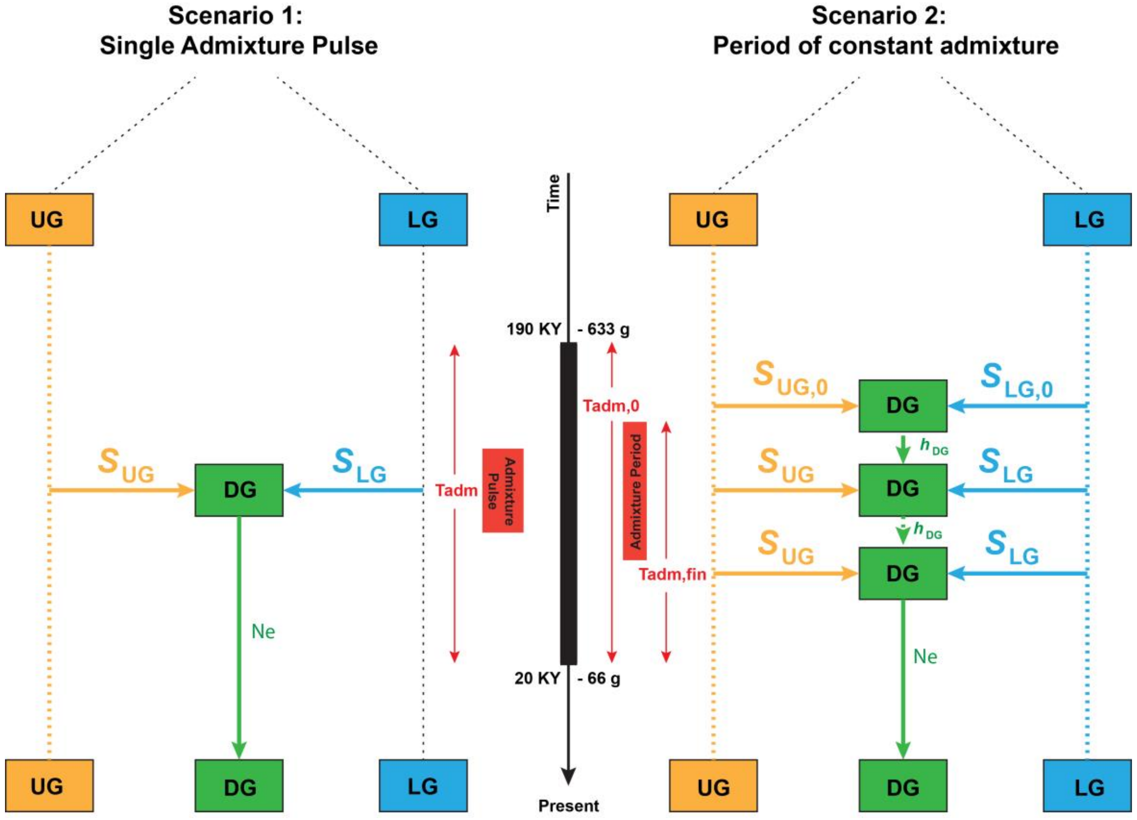
The admixture histories of Cabo Verde
Laurent et al. *Nature Communications* (under revision), BioRxiv https://doi.org/10.1101/2022.04.11.487833



Ongoing !

Microsatellites and *Terminalia superba* hybridization zone in the « Dahomey Gap » (West Africa)
With Romain Laurent (UMR7206) and Olivier Hardy (Univ. Bruxelles)





Great Many Thanks !

Projects co-authors

Romain Laurent *(CNRS-MNHN-Université Paris Cité)*

Bruno Toupance *(CNRS-MNHN-Université Paris Cité)*

Zachary A. Szpiech *(Penn State University)*

Sérgio da Costa *(assoc. UMR7206)*

Valentin Thouzeau *(ENS-Univ. Paris Dauphine)*

Cesar A. Fortes-Lima *(Uppsala University)*

Françoise Dessarps-Freichey *(CNRS-MNHN-Université Paris Cité)*

Laure Lémée *(Institut Pasteur)*

José Utgé *(CNRS-MNHN-Université Paris Cité)*

Marlyse Baptista *(University of Michigan)*

Noah A. Rosenberg *(Stanford University)*

Acknowledgements

Ethan M. Jewett

Samuel Pavard

Marta Ciccarella

Jorge Rocha

Angelo Barbosa

Elisabeth Lima

Emanuel de Pina

P2GM technical platform of the MNHN

France-Stanford Center for Interdisciplinary Study

UMR7206 Eco-anthropology

ANR-METHIS *(15-CE32-0009-01)*

ANR

cnrs dépasser les frontières

MUSÉUM NATIONAL D'HISTOIRE NATURELLE MUSÉE DE L'HOMME

Université Paris Cité