

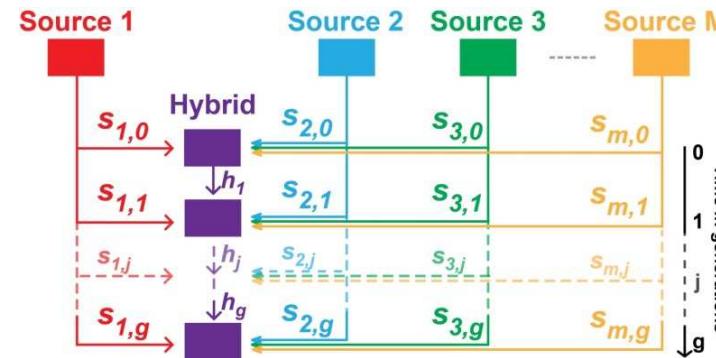
Complex genetic admixture histories reconstructed with Approximate Bayesian Computation

Paul Verdu

Lab: UMR7206 Eco-anthropologie
Institution: CNRS – MnHn - Université de Paris

Rosenberg Lab, Stanford University
February 2021

$$H_{i,g} = \begin{cases} 1 & \text{if } Y = S_i S_i \text{, with } P[Y = S_i S_i] = s_{i,g-1}^2 \\ \frac{H_{i,g-1} + 1}{2} & \text{if } Y = S_i H \text{, with } P[Y = S_i H] = 2s_{i,g-1} h_{g-1} \\ \frac{1}{2} & \text{if } Y = S_i S_j \text{, with } P[Y = S_i S_j] = 2s_{i,g-1} s_{j,g-1} \\ \frac{H_{i,g-1}^{(1)} + H_{i,g-1}^{(2)}}{2} & \text{if } Y = HH \text{, with } P[Y = HH] = h_{g-1}^2 \\ \frac{H_{i,g-1}}{2} & \text{if } Y = S_j H \text{, with } P[Y = S_j H] = 2s_{j,g-1} h_{g-1} \\ 0 & \text{if } Y = S_j S_j \text{, with } P[Y = S_j S_j] = s_{j,g-1}^2 \\ 0 & \text{if } Y = S_j S_i \text{, with } P[Y = S_j S_i] = 2s_{j,g-1} s_{i,g-1}. \end{cases}$$

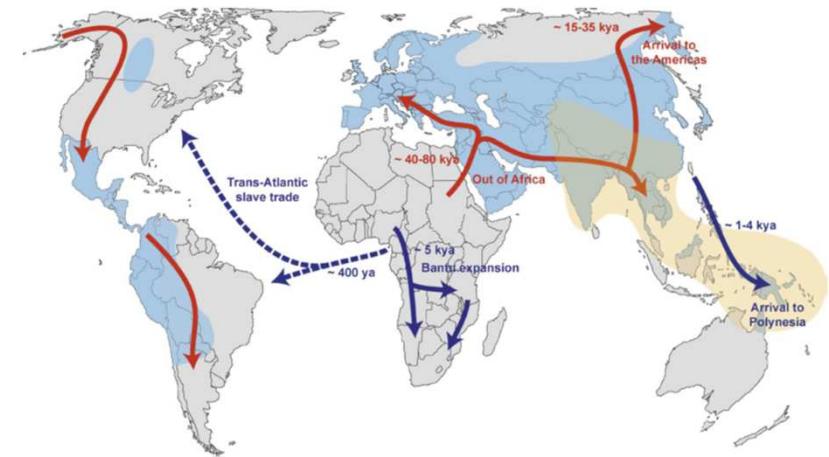
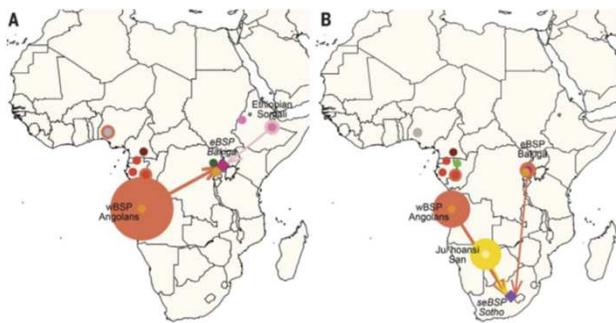
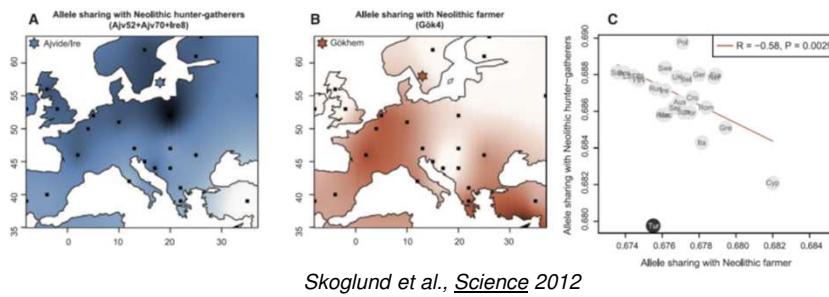
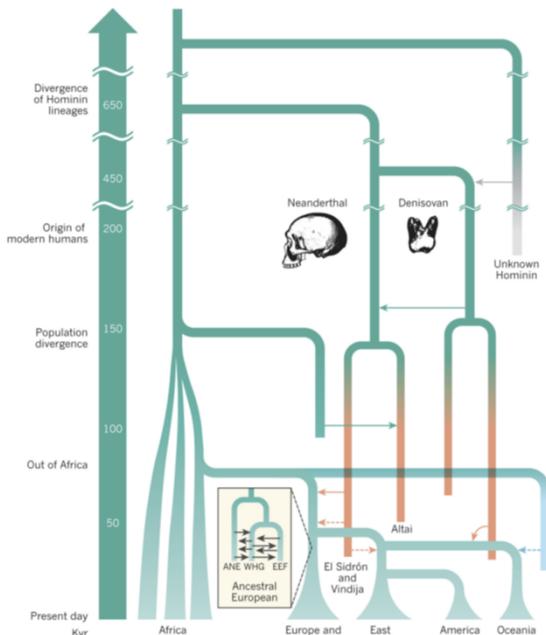


ANR



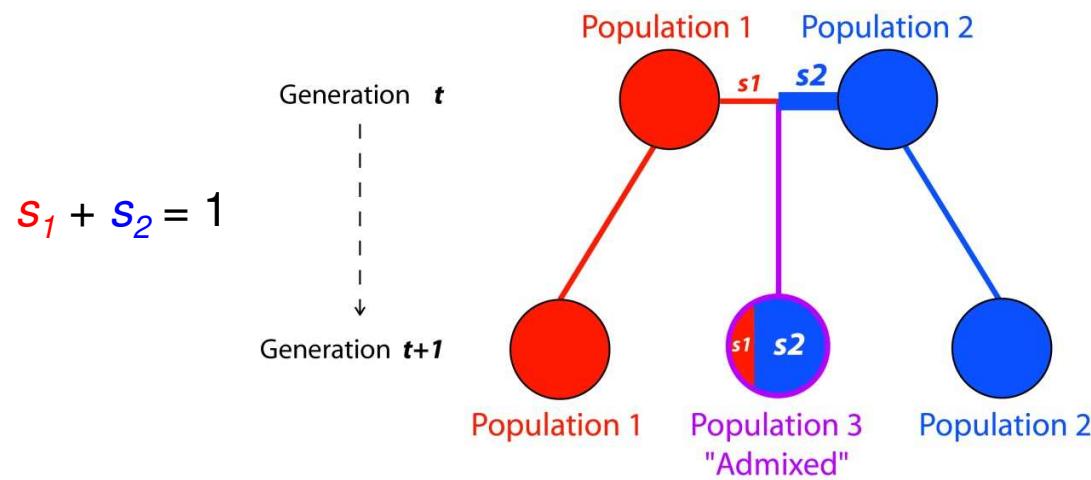
Admixture is ubiquitous in evolutionary history, through time and space

e.g. *Homo sapiens*



Defining Admixture

Exchanges of genes between two or more previously reproductively isolated populations

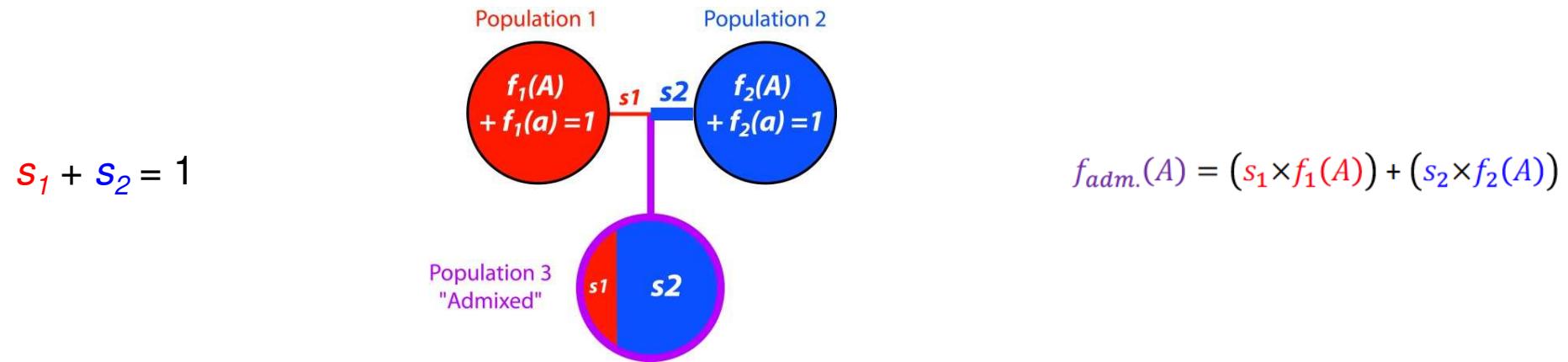


Admixed individuals have a fraction of their genome coming from each source population

Individuals in the “admixed population” are not necessarily all genetically admixed

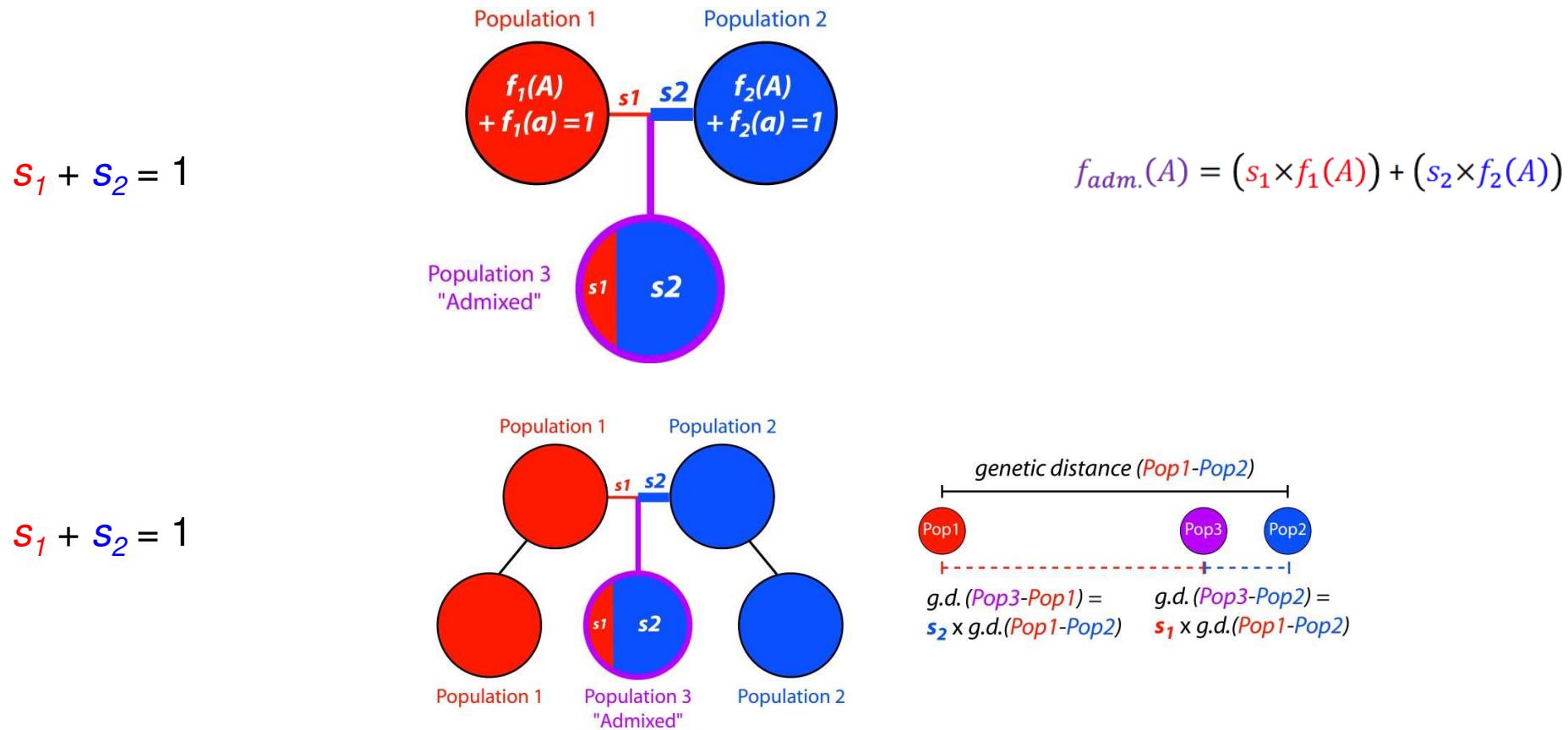
Estimating admixture from genetic data

Allele frequencies in an admixed population can be considered as **a linear combination of allele frequencies in the parental populations** (e.g. Bernstein 1931)



Estimating admixture from genetic data

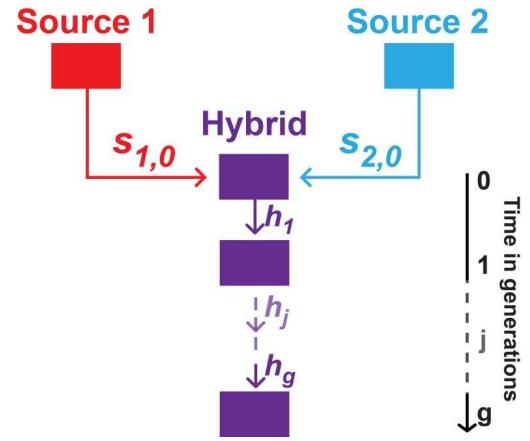
Allele frequencies in an admixed population can be considered as a **linear combination of allele frequencies in the parental populations** (e.g. Bernstein 1931)



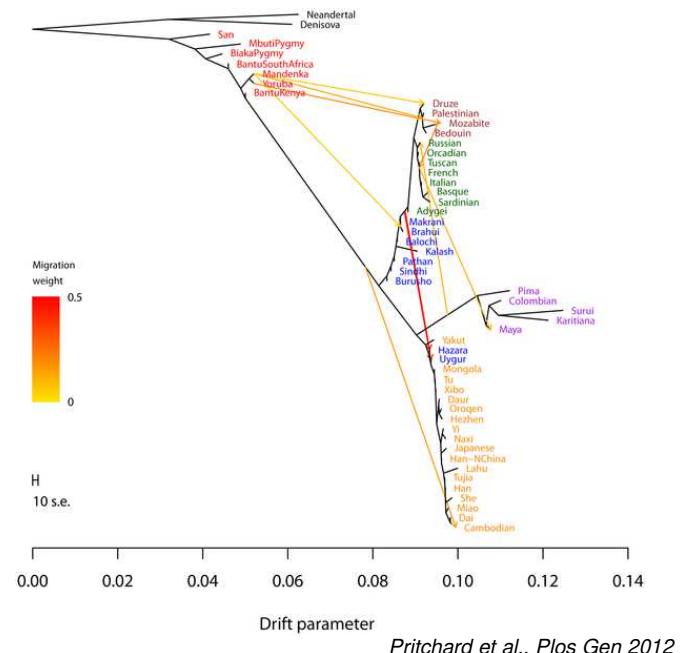
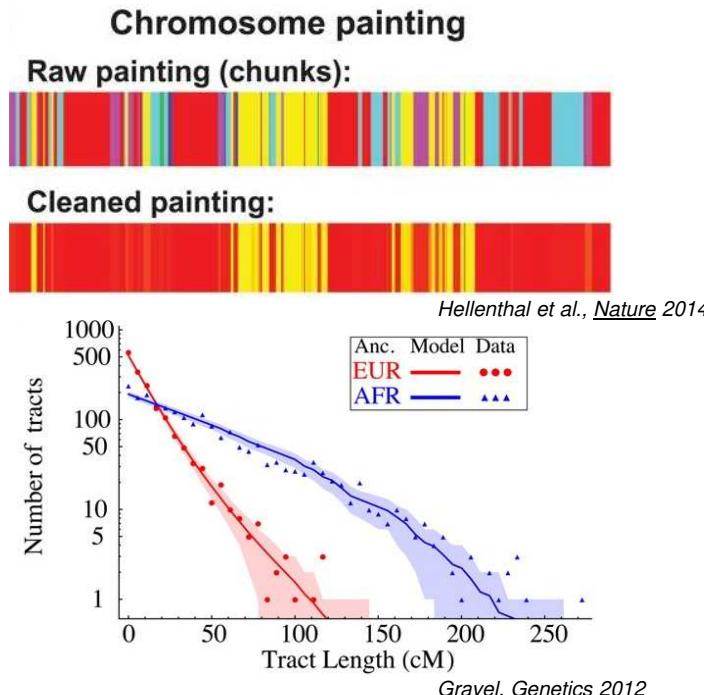
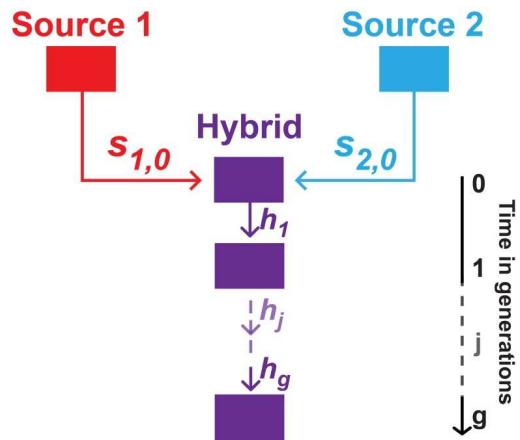
Numerous statistical approaches to detect and infer genetic admixture in a variety of plant and animal species have been developed

(e.g. Long, *Genetics* 1991; Bertorelle and Excoffier, *Mol Biol Evol* 2003; Falush et al., *Genetics* 2003; Alexander et al., *Gen Res* 2009; Patterson et al., *Genetics* 2012)

Inferring admixture history from genetic data



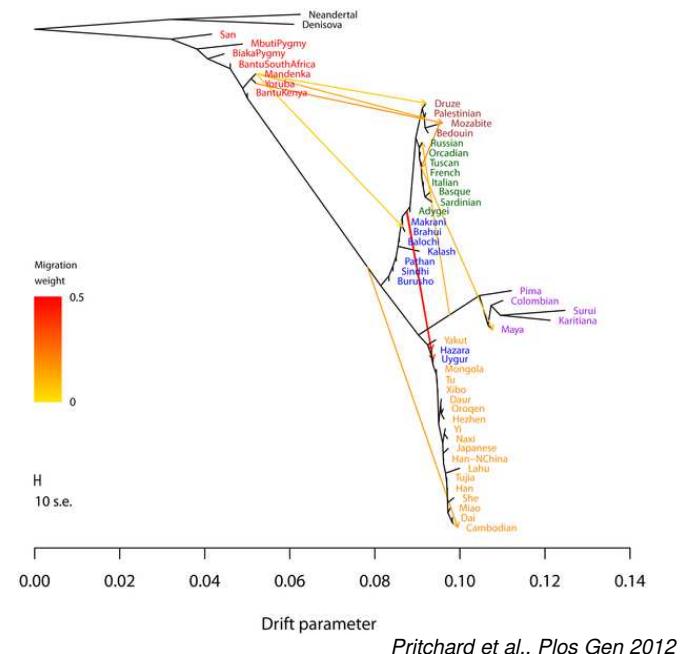
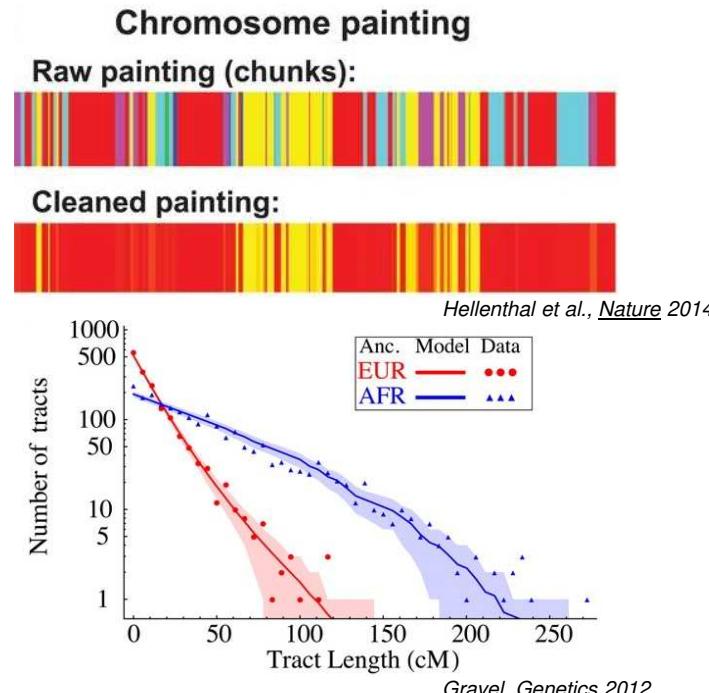
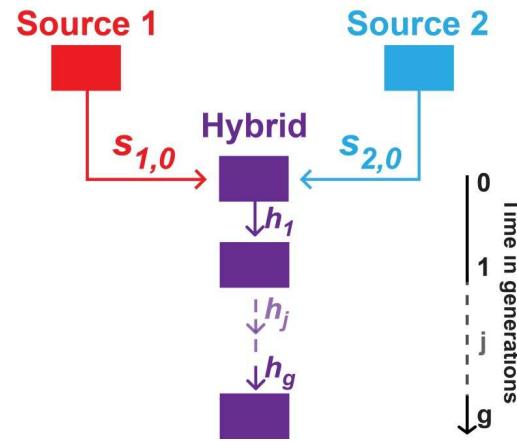
Infering admixture history from genetic data



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



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

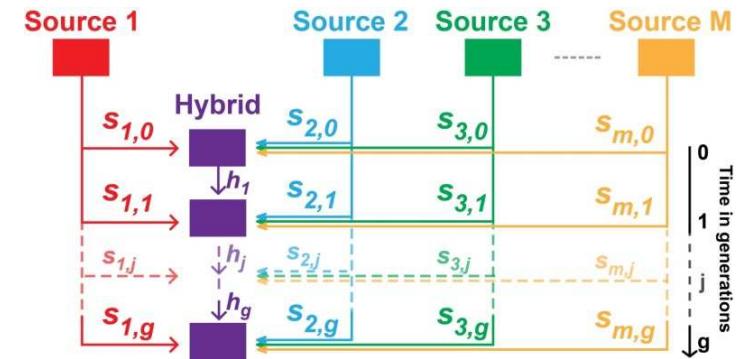
Infering admixture history from genetic data

ML-inference methods cannot operate

$$L(\text{ModelParams} | \text{Data}_{obs}) \propto P(\text{Data}_{obs} | \text{ModelParams}) P(\text{ModelParams})$$

for highly complex admixture models where likelihoods cannot be written

for highly complex admixture models where likelihoods are intractable



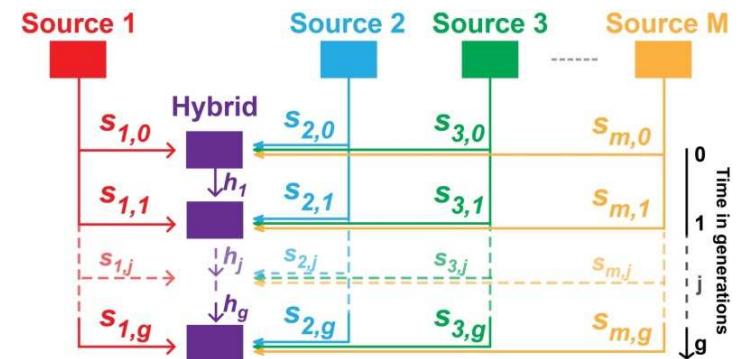
Infering admixture history from genetic data

ML-inference methods cannot operate

$$L(\text{ModelParams} | \text{Data}_{\text{obs}}) \propto P(\text{Data}_{\text{obs}} | \text{ModelParams}) P(\text{ModelParams})$$

for highly complex admixture models where likelihoods cannot be written

for highly complex admixture models where likelihoods are intractable



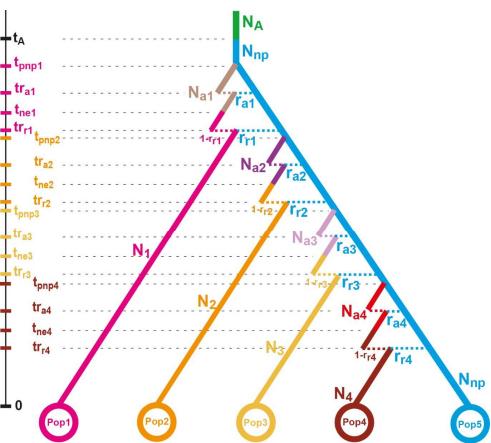
Approximate Bayesian Computation (Tavaré et al. 1997) may represent an alternative

$$L(\text{ModelParams} | \text{Data}_{\text{obs}}) \propto P(\text{Data}_{\text{obs}} | \text{ModelParams}) P(\text{ModelParams})$$

↙
↓
Approximation

$$L(\text{ModelParams} | \text{SumStat}_{\text{obs}}) \propto P(\text{SumStat}_{\text{obs}} | \text{ModelParams}) P(\text{ModelParams})$$

Approximate Bayesian Computation demographic inference



Parameter prior distributions

N_A Uniform [10 .. 1 000]

N_{np} Uniform [10 .. 100 000]

N₁ Uniform [10 .. 10 000]

...

r_{a1} Normal [0 , 1]

r_{r1} LogUniform [0 , 1]

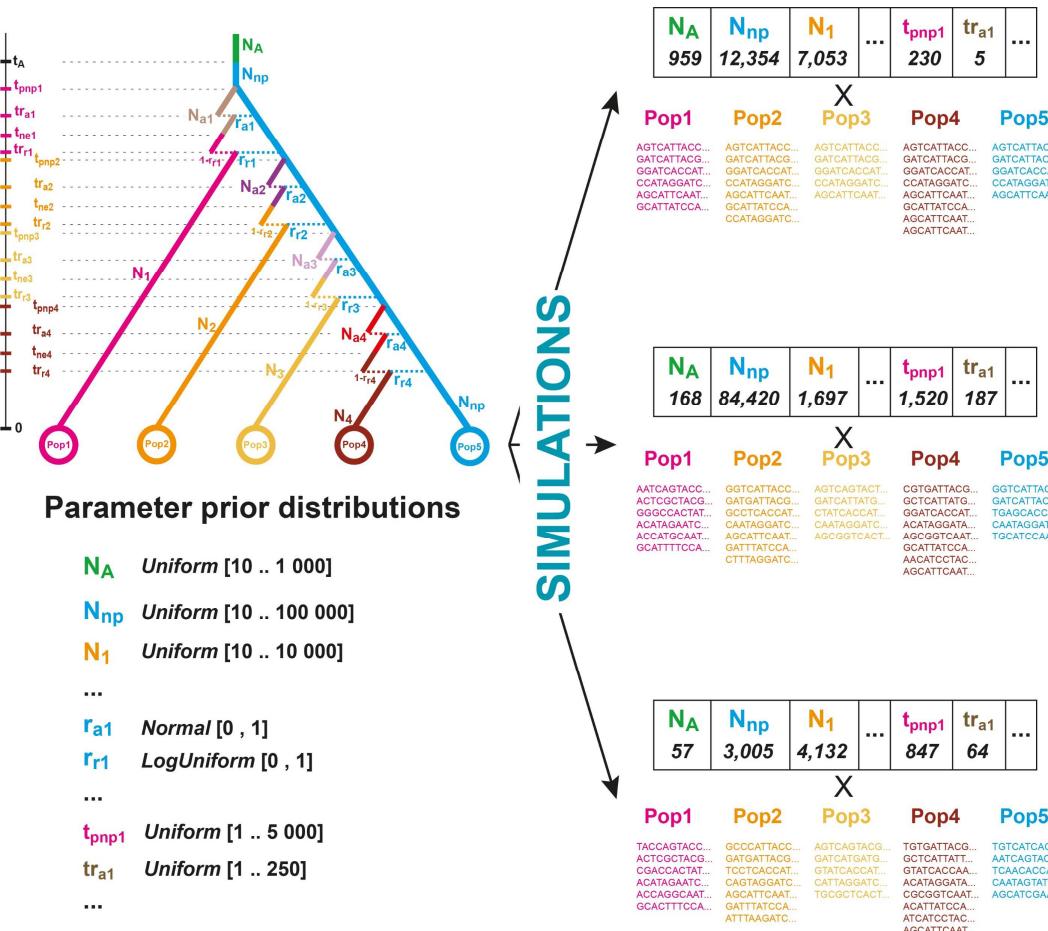
...

t_{pnp1} Uniform [1 .. 5 000]

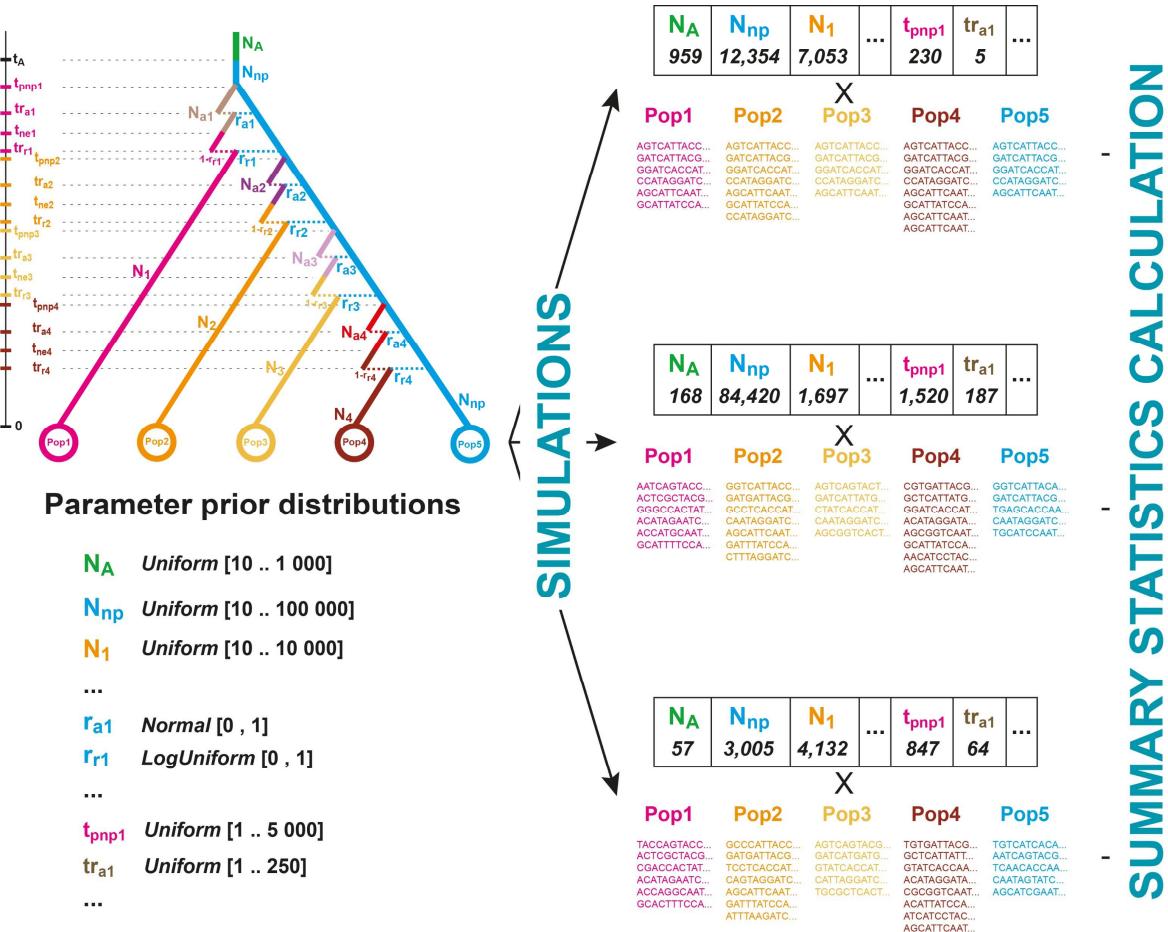
t_{r1} Uniform [1 .. 250]

...

Approximate Bayesian Computation demographic inference



Approximate Bayesian Computation demographic inference



N_A	N_{np}	N₁	...	t_{pnp1}	tr_{a1}	...
959	12,354	7,053		230	5	

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

N _A	N _{np}	N ₁	...	t _{pnp1}	tr _{a1}	...
168	84,420	1,697	...	1,520	187	...

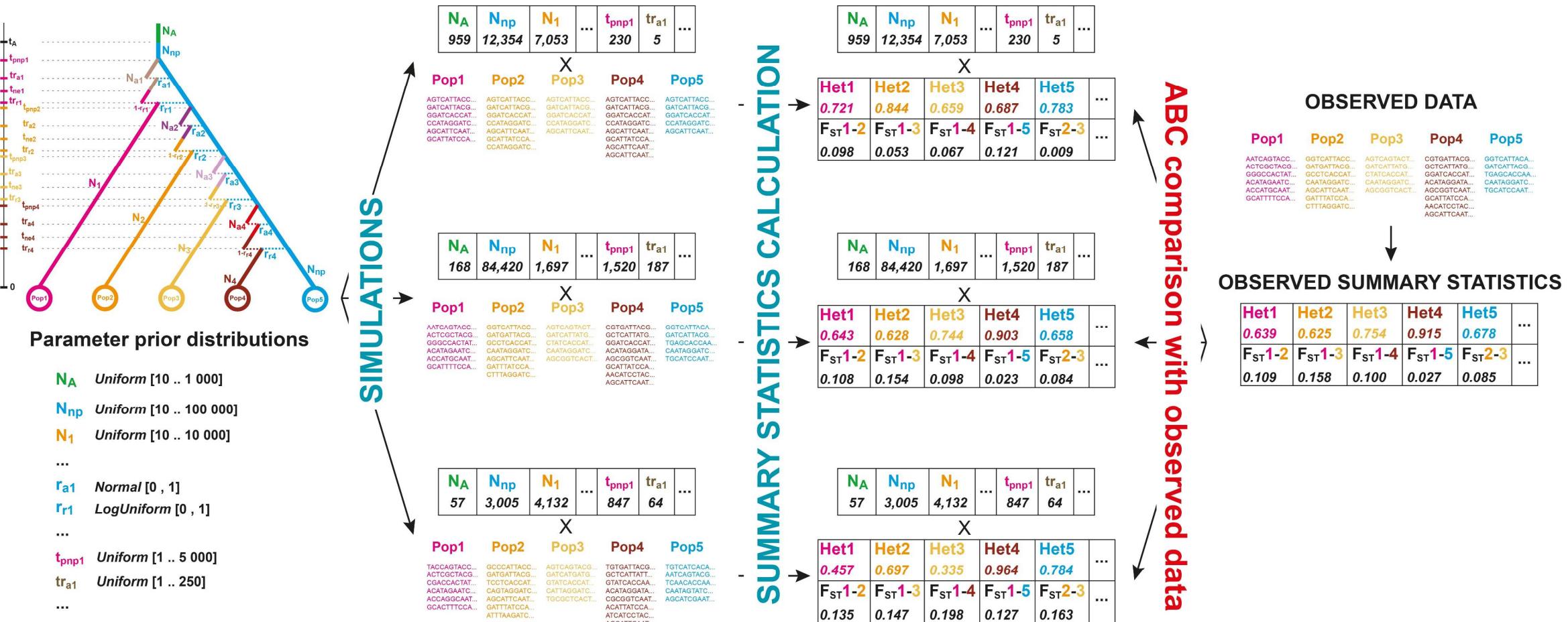
λ					
Het1	Het2	Het3	Het4	Het5	...
0.643	0.628	0.744	0.903	0.658	...
F_{ST} 1-2	F_{ST} 1-3	F_{ST} 1-4	F_{ST} 1-5	F_{ST} 2-3	...
0.108	0.151	0.098	0.023	0.081	...

N_A	N_{np}	N₁	...	t_{pnp1}	tr_{a1}	...
57	3.005	4.132	...	847	64	...

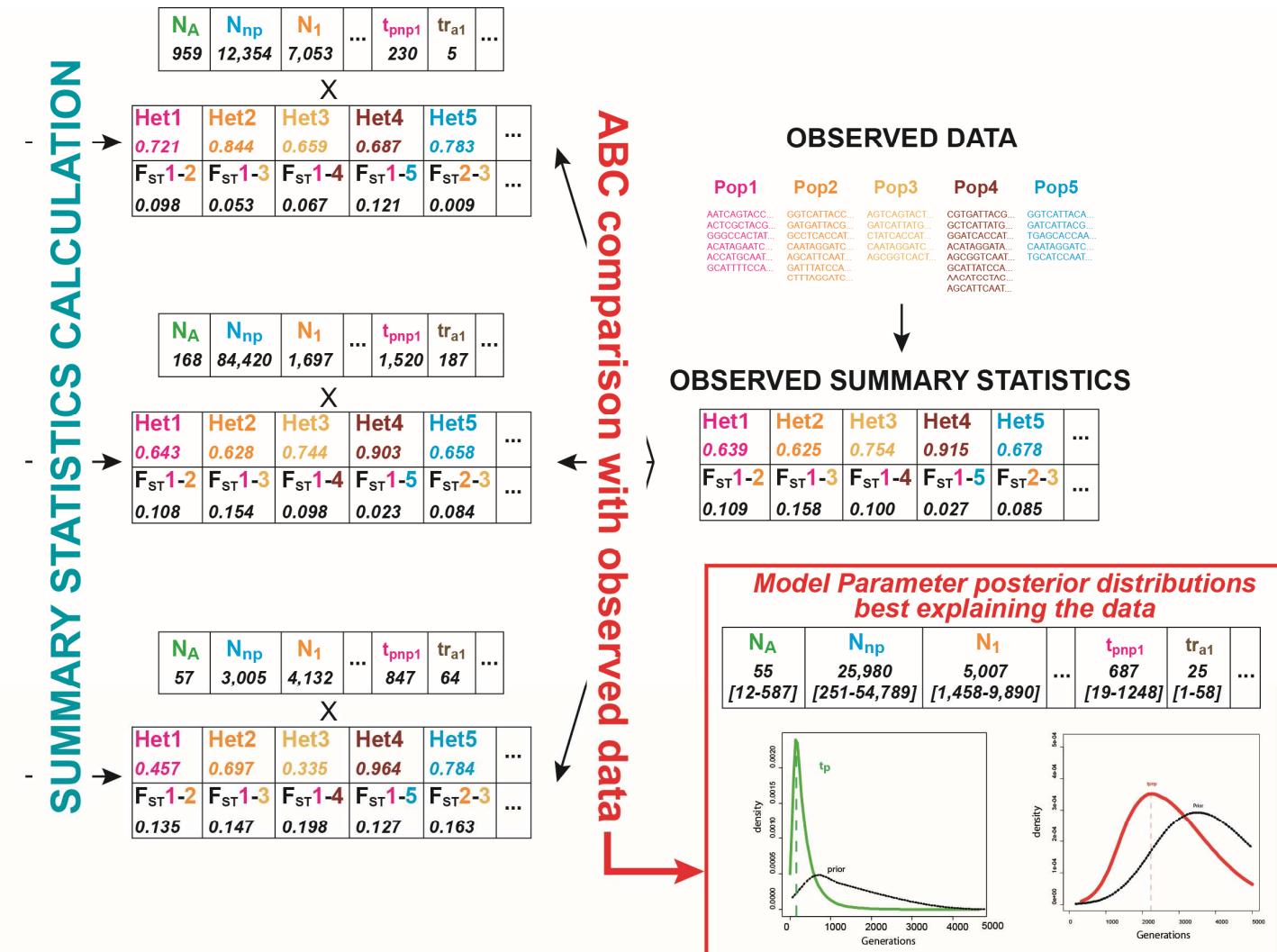
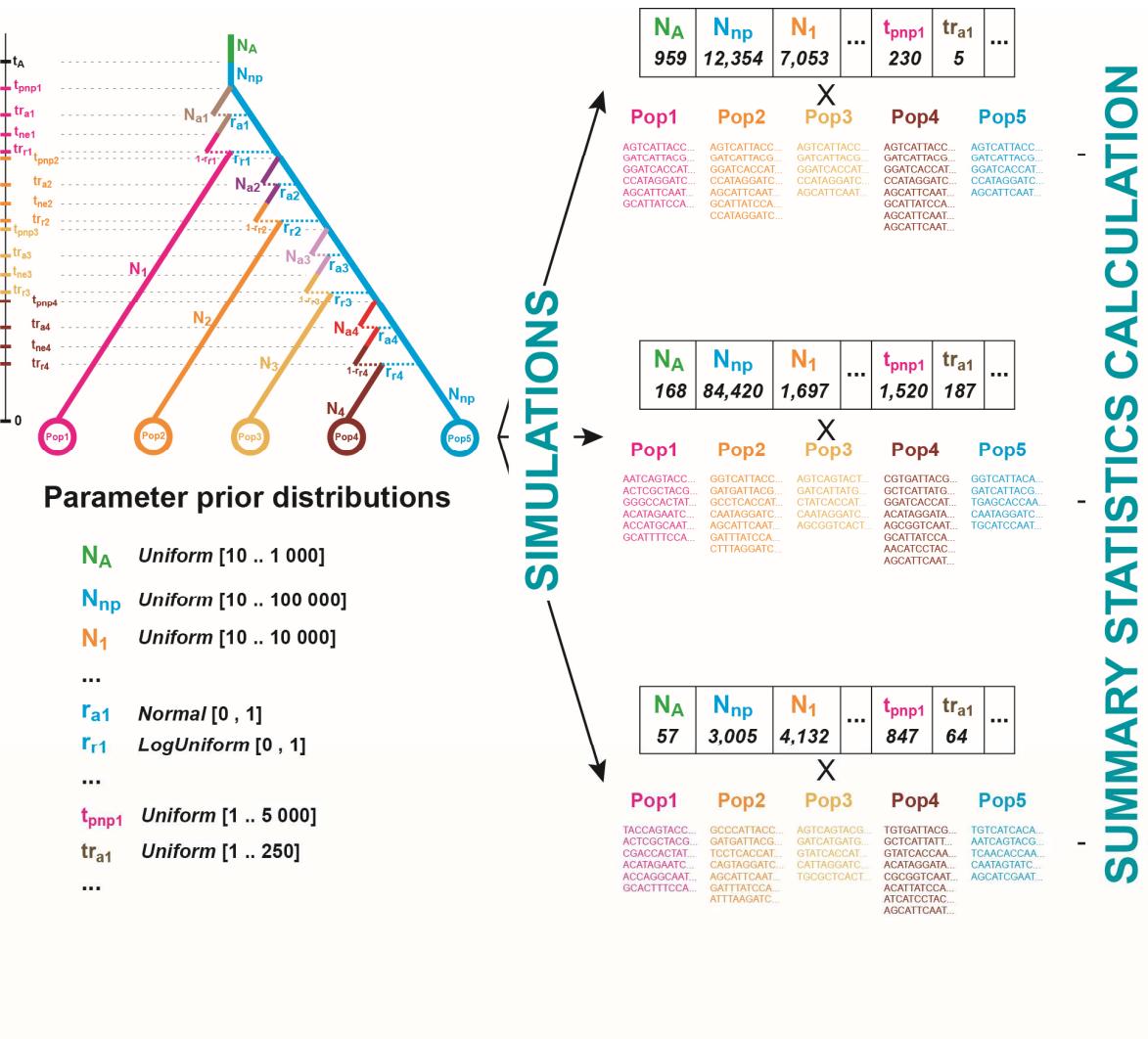
X	Het1 0.457	Het2 0.697	Het3 0.335	Het4 0.964	Het5 0.784	...
F _{ST}	F _{ST} 1-2 0.105	F _{ST} 1-3 0.117	F _{ST} 1-4 0.100	F _{ST} 1-5 0.107	F _{ST} 2-3 0.100	...

SUMMARY STATISTICS CALCULATION

Approximate Bayesian Computation demographic inference

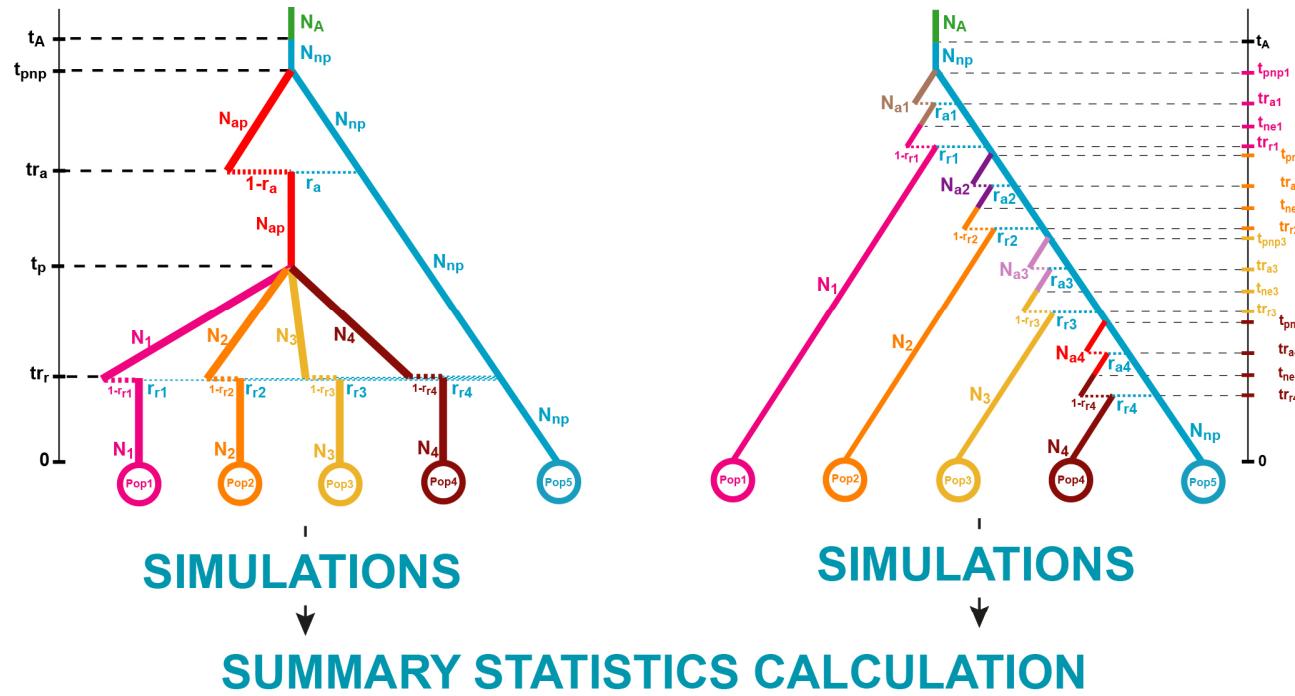


Approximate Bayesian Computation demographic inference



Approximate Bayesian Computation demographic inference

Formal model
comparison with
ABC

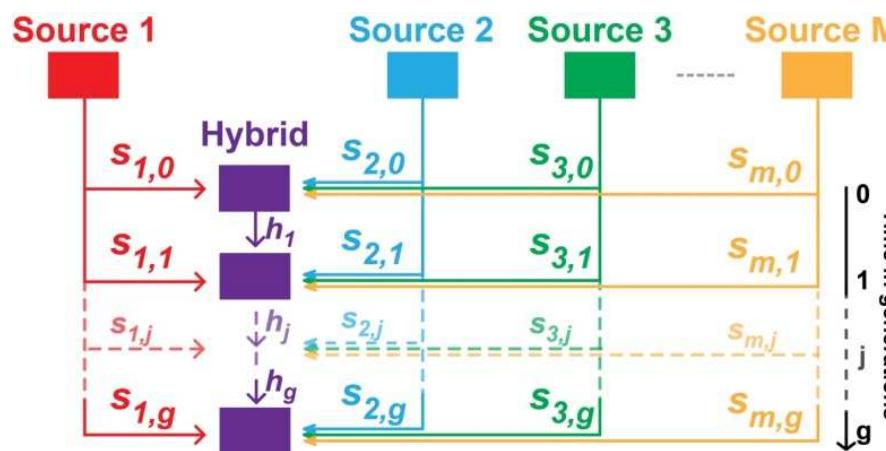


ABC model-choice:
Classify all scenarios in relative
distances to the observed data

Approximate Bayesian Computation for complex admixture history

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

$$L(\text{ModelParams} | \text{SumStat}_{obs}) \propto P(\text{SumStat}_{obs} | \text{ModelParams}) P(\text{ModelParams})$$



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

Complex admixture histories reconstructed with *MetHis-ABC*

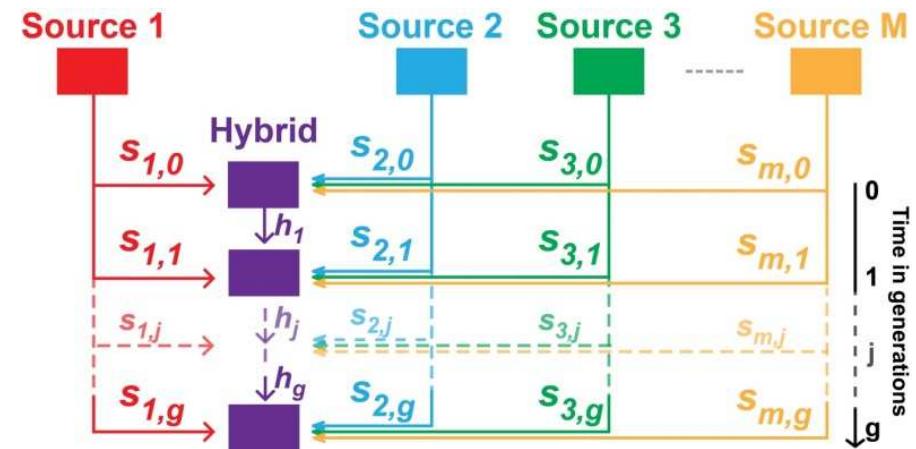
If simulations are feasible...

Complex admixture histories reconstructed with *MetHis*-ABC

If simulations are feasible...

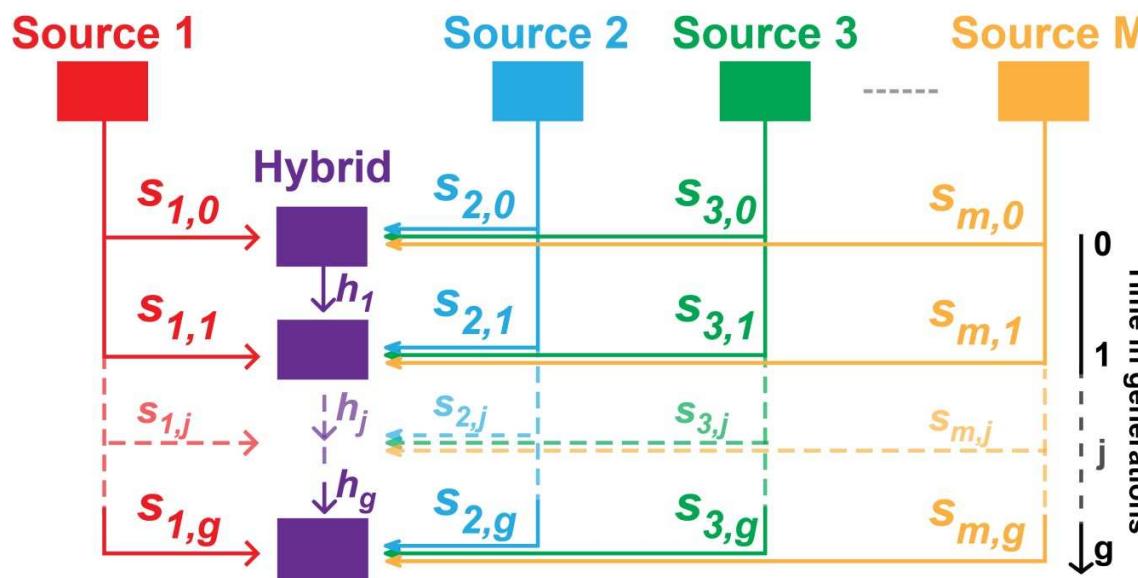
Simulating complex admixture histories under the coalescent is often not trivial:

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



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

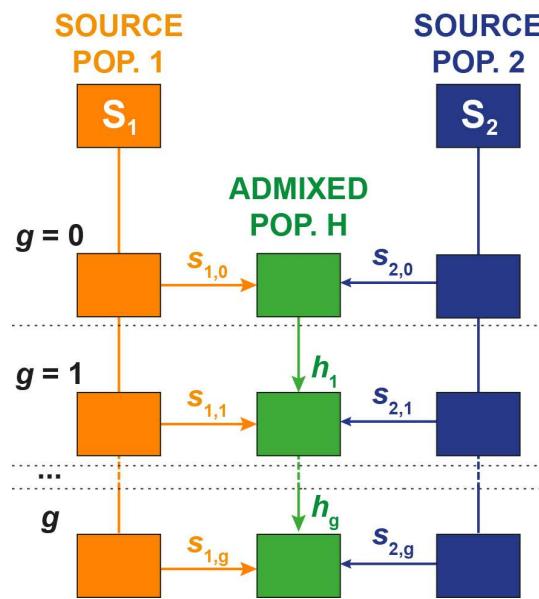
$$\sum_{m=1}^M s_{m,0} = 1$$

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

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

► Two source populations version

MetHis: genetic data simulator under complex admixture histories



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

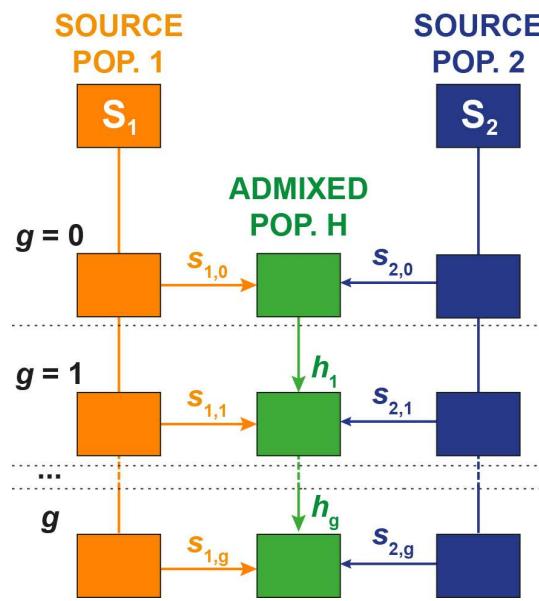
$\forall m \in \{1, \dots, M\}, s_{m,0} \in [0,1]$ such that

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

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

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

MetHis: genetic data simulator under complex admixture histories



MetHis simulates a **panmictic 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)

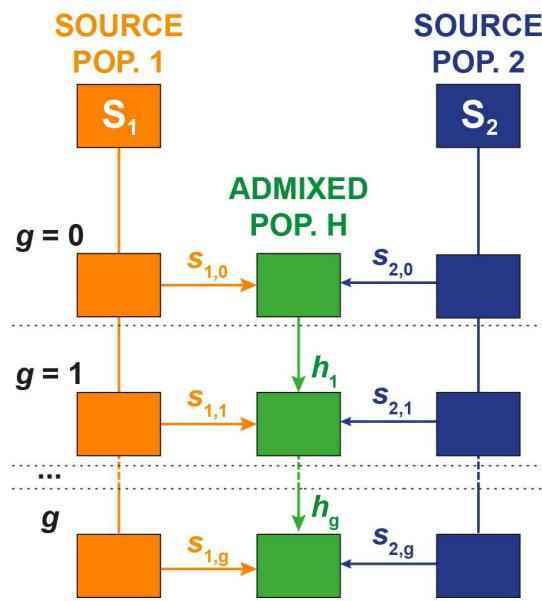
$$\forall m \in \{1, \dots, M\}, s_{m,0} \in [0,1] \text{ such that}$$

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

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

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

MetHis: genetic data simulator under complex admixture histories



MetHis simulates a **panmictic 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 S_1 , S_2 , and H in proportions $s_{1,g}$, $s_{2,g}$ and h_g

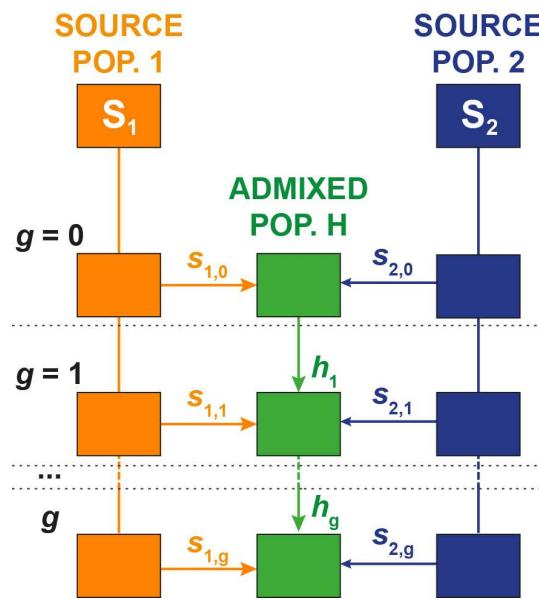
$\forall m \in \{1, \dots, M\}, s_{m,0} \in [0,1]$ such that

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

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

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

MetHis: genetic data simulator under complex admixture histories



MetHis simulates a **panmictic 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

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

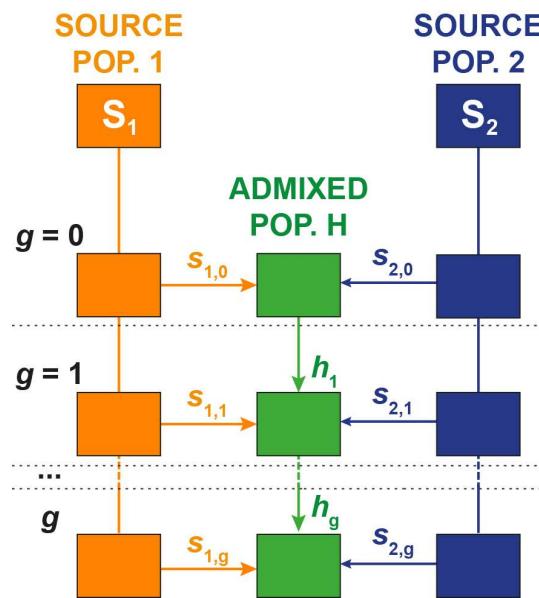
$$\sum_{m=1}^M s_{m,0} = 1$$

$$\forall m \in \{1, \dots, M\}, \forall j \in \{1, \dots, g\},$$

$$h_j \text{ and } s_{m,j} \in [0,1] \text{ such that}$$

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

MetHis: genetic data simulator under complex admixture histories



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

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

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

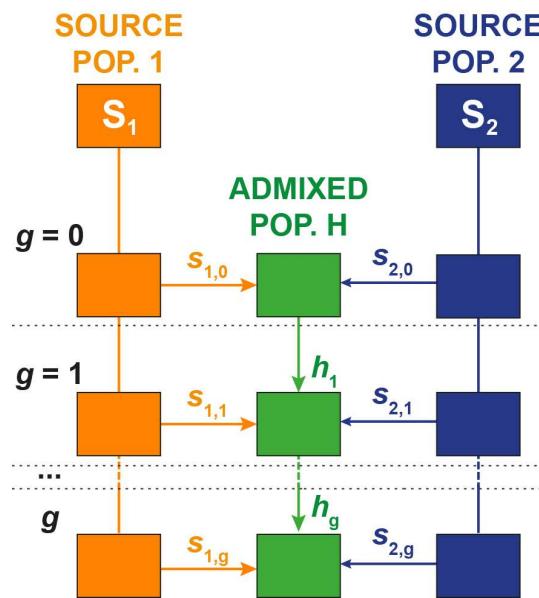
$$h_j + \sum_{m=1}^M s_{m,j} = 1$$

MetHis simulates a **panmictic 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 S_1 , S_2 , 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: genetic data simulator under complex admixture histories



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

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

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

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

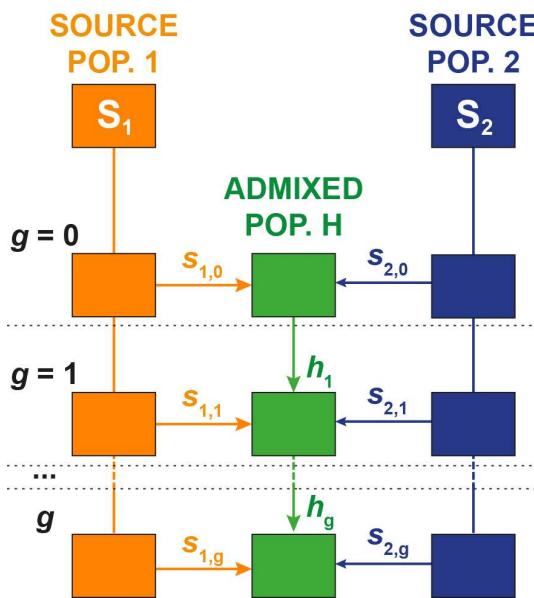
MetHis simulates a **panmictic admixed population of diploid size N_g** at generation g , **forward-in-time centered on individuals**

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 S_1 , S_2 , 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)

MetHis: genetic data simulator under complex admixture histories



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

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

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

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

MetHis simulates a **panmictic 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)

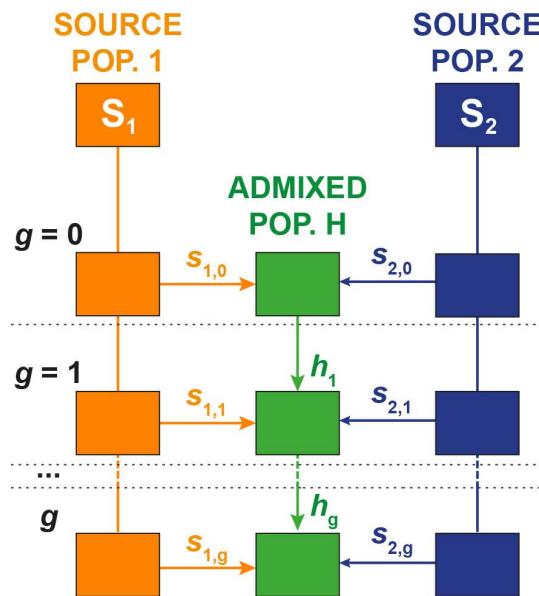
MetHis does not simulate source populations:

Simulate source populations using existing coalescent simulators

Simulate source populations by drawing alleles in the observed Allelic Frequency Spectrum

MetHis: genetic data simulator under complex admixture histories

MetHis full two-sources Verdu and Rosenberg 2011 Model



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

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

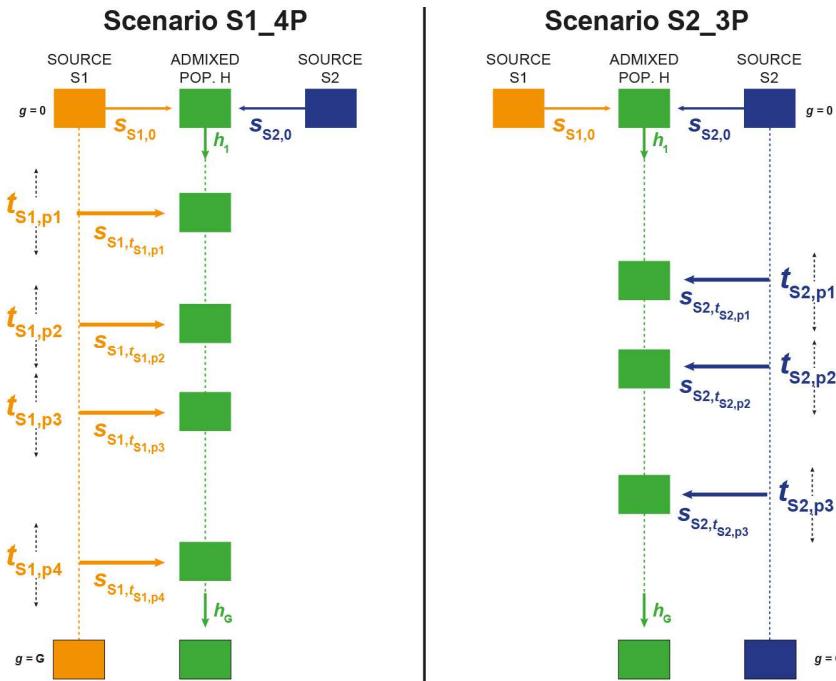
$\forall m \in \{1, \dots, M\}, \forall j \in \{1, \dots, g\},$
 h_j and $s_{m,j} \in [0,1]$ such that

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

**Input a model parameter
list at will, provided model
definition is respected**

MetHis: genetic data simulator under complex admixture histories

MetHis parameter generator tool *n-pulses of admixture*



For each source population separately:

1. Choose number of pulses
2. Choose time priors for each pulse
3. Choose priors for introgression rates for each pulse

MetHis parameter-generator tool builds parameters tables

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

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

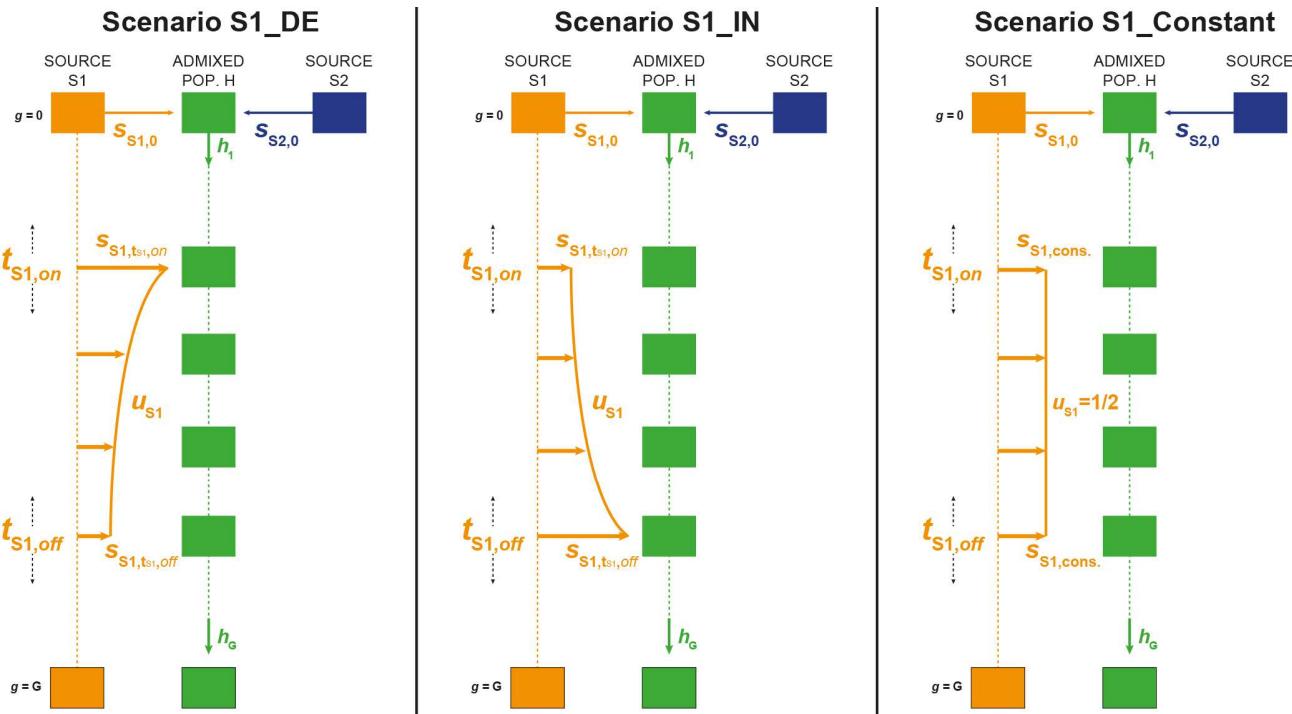
$\forall m \in \{1, \dots, M\}, \forall j \in \{1, \dots, g\}$,
 h_j and $s_{m,j} \in [0,1]$ such that

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

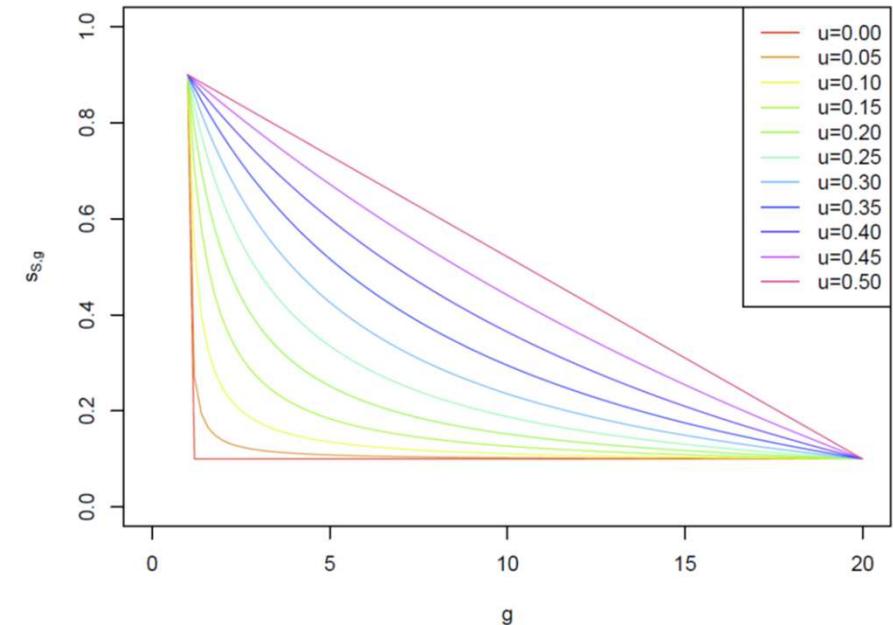
MetHis: genetic data simulator under complex admixture histories

MetHis parameter generator tool

Period of monotonically recurring admixture (rectangular hyperbola)



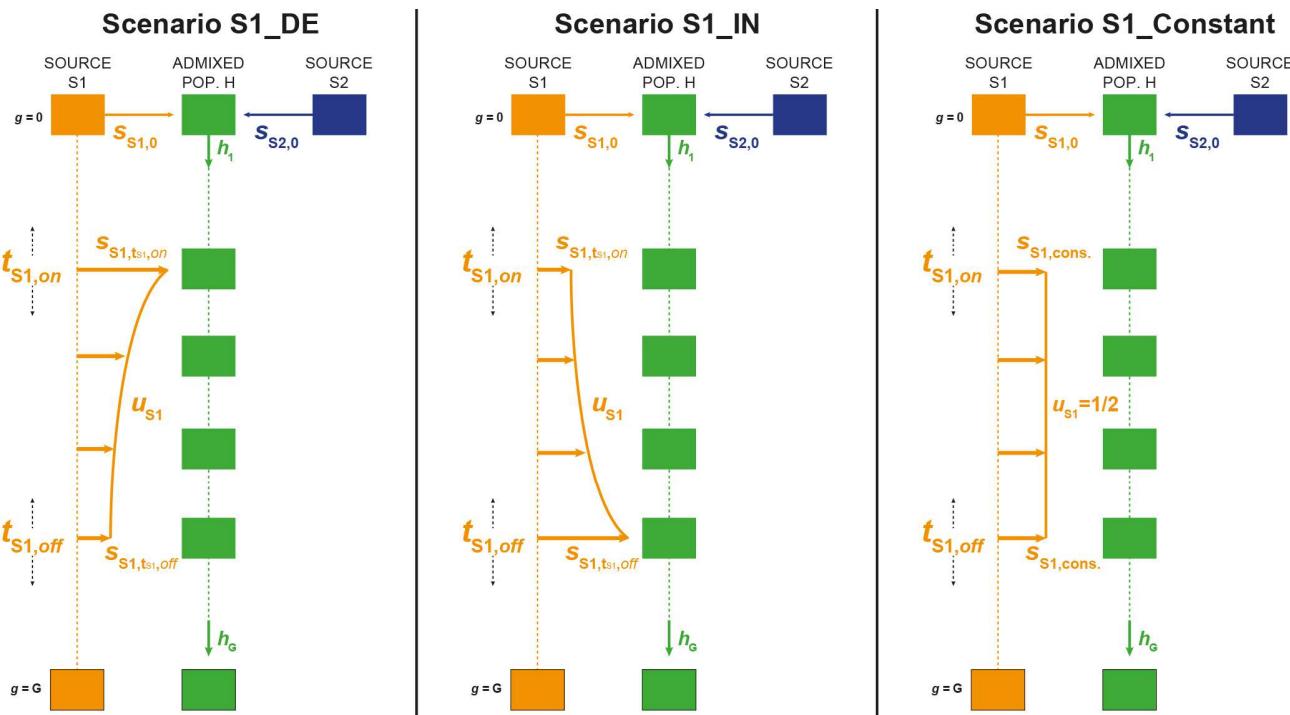
$$f(x) = \frac{a(1-x)}{a+x} \quad a = \frac{u^2}{1-2u}$$



MetHis: genetic data simulator under complex admixture histories

MetHis parameter generator tool

Period of monotonically recurring admixture (rectangular hyperbola)



For each source population separately:

1. Choose time priors for the beginning and the end of the admixture period
2. Choose priors for introgression rates at the beginning and the end of the admixture period
2. Choose prior for the shape of the scaled rectangular hyperbola

MetHis parameter-generator tool builds parameters tables

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

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

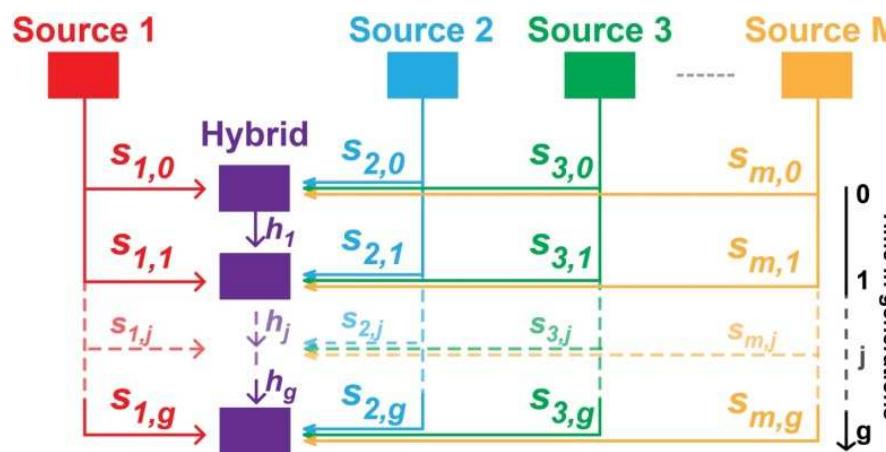
$$\forall m \in \{1, \dots, M\}, \forall j \in \{1, \dots, g\}, h_j \text{ and } s_{m,j} \in [0,1] \text{ such that}$$

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

Approximate Bayesian Computation for complex admixture history

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

$$L(\text{ModelParams} | \text{SumStat}_{obs}) \propto P(\text{SumStat}_{obs} | \text{ModelParams}) P(\text{ModelParams})$$



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

MetHis: summary-statistics calculator for ABC inference

If summary-statistics are informative about model-parameters...

MetHis: summary-statistics calculator for ABC inference

If summary-statistics are informative 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)

MetHis: summary-statistics calculator for ABC inference

If summary-statistics are informative 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)

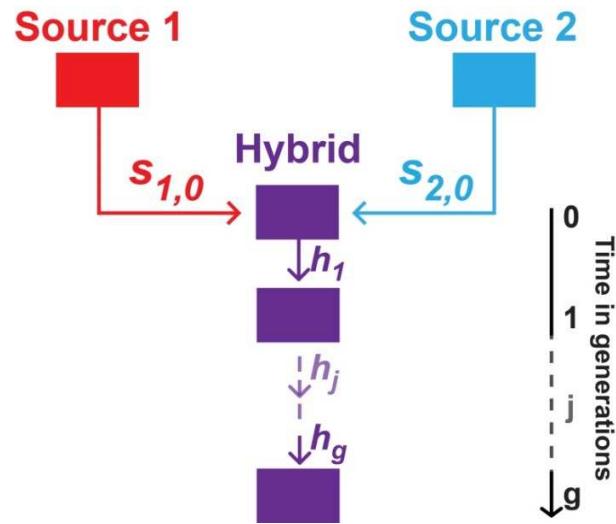
f_3 (Admixed; S1, S2) (Patterson et al., 2012)

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

What about individual admixture estimates ?

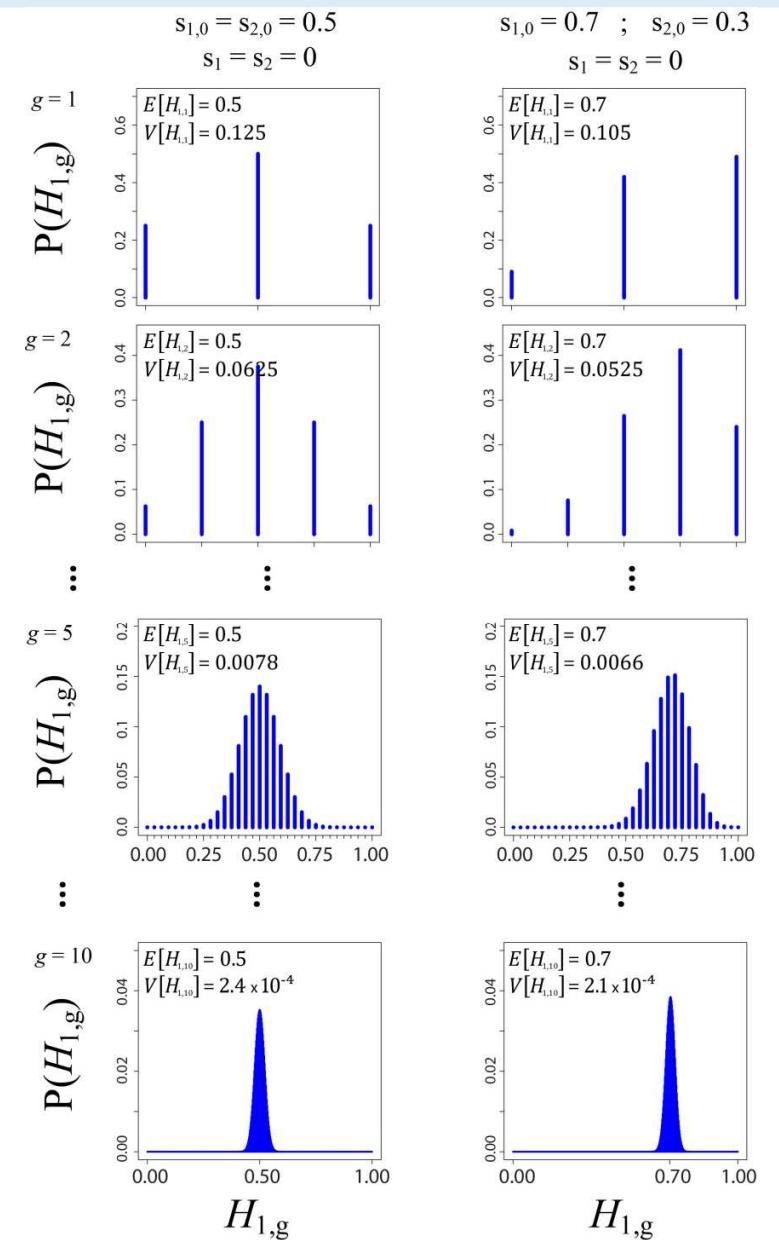
Methis: summary-statistics calculator for ABC inference

The moments of the distribution of admixture fractions in the admixed population are informative about the underlying admixture process



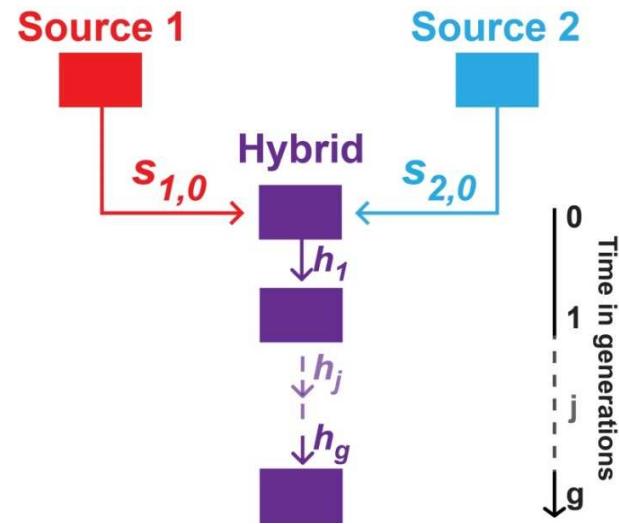
$$E[H_{1,g}] = s_{1,0}.$$

$$V[H_{1,g}] = \frac{s_{1,0}(1 - s_{1,0})}{2^g}.$$



Methis: summary-statistics calculator for ABC inference

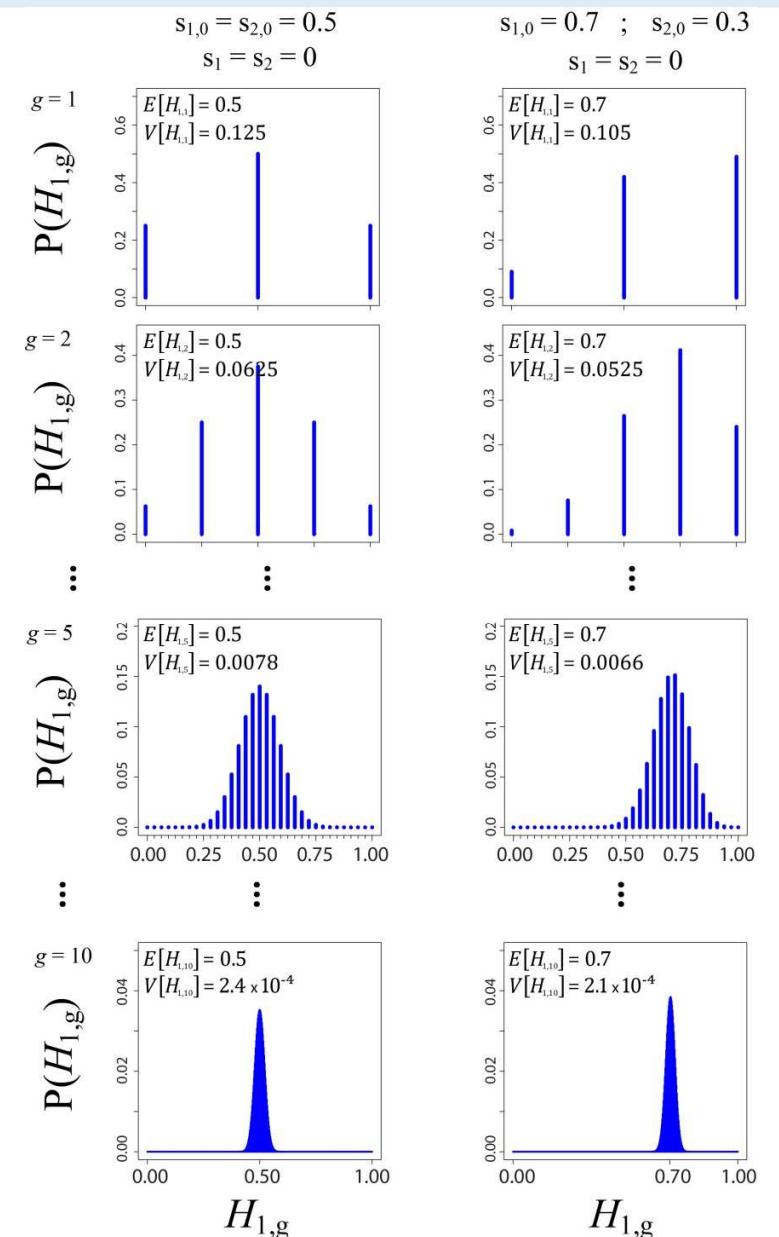
The moments of the distribution of admixture fractions in the admixed population are informative about the underlying admixture process



$$E[H_{1,g}] = s_{1,0}.$$

$$V[H_{1,g}] = \frac{s_{1,0}(1 - s_{1,0})}{2^g}.$$

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



Methis: summary-statistics calculator for ABC inference

ML methods to estimate admixture fractions (e.g ADMIXTURE, Alexander et al. 2009):

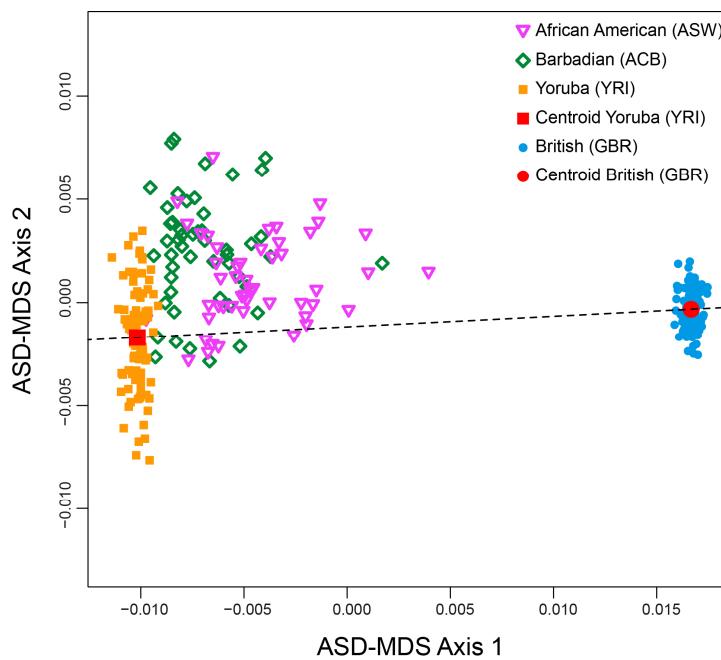
Computationally very intensive and difficult to compute for numerous simulations (>10K) for ABC inference

Methis: summary-statistics calculator for ABC inference

ML methods to estimate admixture fractions (e.g ADMIXTURE, Alexander et al. 2009):

Computationally very intensive and difficult to compute for numerous simulations (>10K) for ABC inference

Allele Sharing Dissimilarity (Bowcock et al. 1994)-MDS based individual admixture fraction estimation

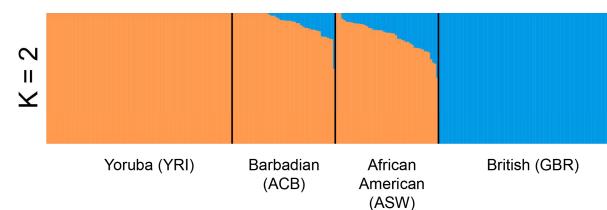
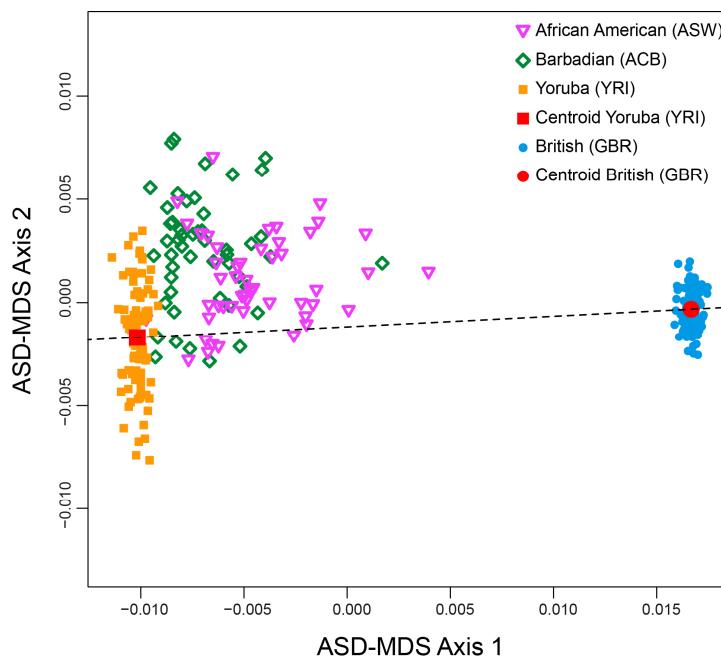


Methis: summary-statistics calculator for ABC inference

ML methods to estimate admixture fractions (e.g ADMIXTURE, Alexander et al. 2009):

Computationally very intensive and difficult to compute for numerous simulations (>10K) for ABC inference

Allele Sharing Dissimilarity (Bowcock et al. 1994)-MDS based individual admixture fraction estimation

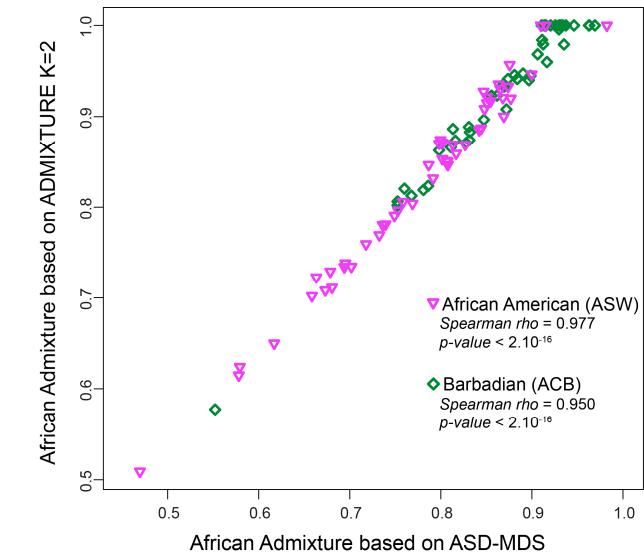
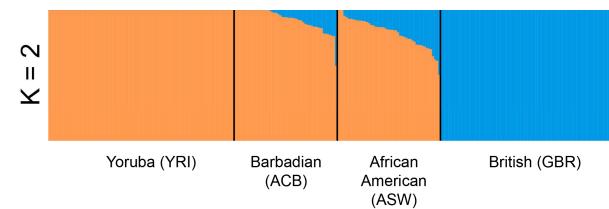
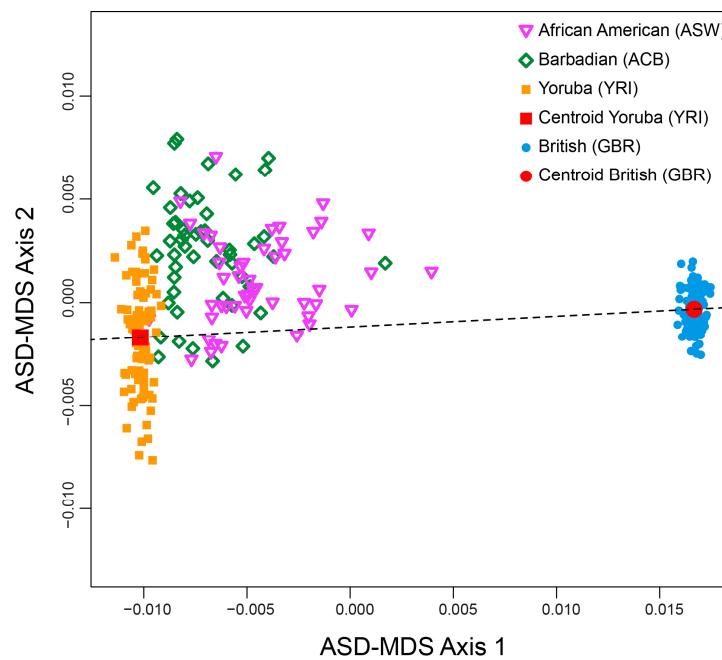


Methis: summary-statistics calculator for ABC inference

ML methods to estimate admixture fractions (e.g ADMIXTURE, Alexander et al. 2009):

Computationally very intensive and difficult to compute for numerous simulations (>10K) for ABC inference

Allele Sharing Dissimilarity (Bowcock et al. 1994)-MDS based individual admixture fraction estimation



MetHis: summary-statistics calculator for ABC inference

If **summary-statistics are informative 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)

ASD-MDS admixture fractions

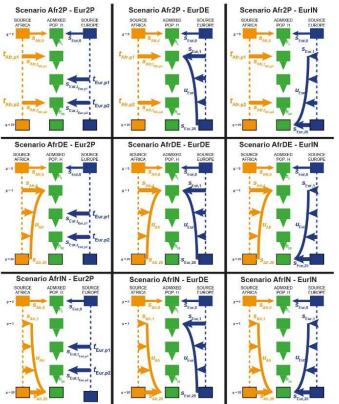
Min and **Max** admixture fractions

10% quantiles of individual admixture fractions

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

Complex admixture histories reconstructed with MetHis-ABC

Set by the user

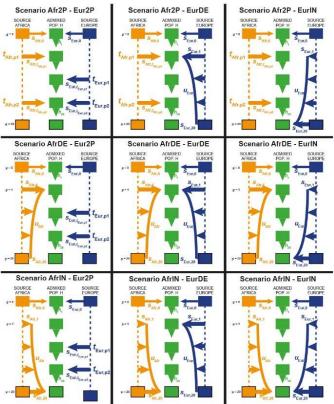


Parameter prior distributions

Pulses	$s_{Afr,0}$	Uniform [0,1]
	$s_{Afr,t_{Afr,p1}}$	$s_{Eur,t_{Eur,p1}}$ Uniform [0,1]
	$s_{Afr,t_{Afr,p2}}$	$s_{Eur,t_{Eur,p2}}$ Uniform [0,1]
	$t_{Afr,p1}$	$t_{Eur,p1}$ Uniform [1,20]
	$t_{Afr,p2}$	$t_{Eur,p2}$ Uniform [1,20]
	$t_{Afr,p1} \neq t_{Afr,p2}$	$t_{Eur,p1} \neq t_{Eur,p2}$
Decrease	$s_{Afr,0}$	Uniform [0,1]
	$s_{Afr,f}$	$s_{Eur,f}$ Uniform [0,1]
	$s_{Afr,20}$	$s_{Eur,20}$ Uniform [0, $s_{s,f}/3$]
	u_{Afr}	u_{Eur} Uniform [0,0.5]
Increase	$s_{Afr,0}$	Uniform [0,1]
	$s_{Afr,f}$	$s_{Eur,f}$ Uniform [0, $s_{s,20}/3$]
	$s_{Afr,20}$	$s_{Eur,20}$ Uniform [0,1]
	u_{Afr}	u_{Eur} Uniform [0,0.5]

Complex admixture histories reconstructed with MetHis-ABC

Set by the user



Parameter prior distributions

$$s_{\text{Afr},0} \quad \text{Uniform } [0,1]$$

$$s_{\text{Afr},t_{\text{Afr},p1}} \quad s_{\text{Eur},t_{\text{Eur},p1}} \quad \text{Uniform } [0,1]$$

$$s_{\text{Afr},t_{\text{Afr},p2}} \quad s_{\text{Eur},t_{\text{Eur},p2}} \quad \text{Uniform } [0,1]$$

$$t_{\text{Afr},p1} \quad t_{\text{Eur},p1} \quad \text{Uniform } [1,20]$$

$$t_{\text{Afr},p2} \quad t_{\text{Eur},p2} \quad \text{Uniform } [1,20]$$

$$t_{\text{Afr},p1} \neq t_{\text{Afr},p2} \quad t_{\text{Eur},p1} \neq t_{\text{Eur},p2}$$

$$s_{\text{Afr},0} \quad \text{Uniform } [0,1]$$

$$s_{\text{Afr},1} \quad s_{\text{Eur},f} \quad \text{Uniform } [0,1]$$

$$s_{\text{Afr},20} \quad s_{\text{Eur},20} \quad \text{Uniform } [0, s_{\text{s},1}/3]$$

$$u_{\text{Afr}} \quad u_{\text{Eur}} \quad \text{Uniform } [0,0.5]$$

$$s_{\text{Afr},0} \quad \text{Uniform } [0,1]$$

$$s_{\text{Afr},1} \quad s_{\text{Eur},f} \quad \text{Uniform } [0, s_{\text{s},20}/3]$$

$$s_{\text{Afr},20} \quad s_{\text{Eur},20} \quad \text{Uniform } [0,1]$$

$$u_{\text{Afr}} \quad u_{\text{Eur}} \quad \text{Uniform } [0,0.5]$$

**MetHis Tools
or
Set by the user**

Parameter Generator

Scenario Afr2P - Eur2P						
Sim1	$s_{\text{Afr},0}$ 0.35	$s_{\text{Afr},t_{\text{Afr},p1}}$ 0.003	$s_{\text{Eur},t_{\text{Eur},p1}}$ 0.124	$t_{\text{Afr},p1}$ 4	$t_{\text{Eur},p2}$ 11	...
Sim2	$s_{\text{Afr},0}$ 0.74	$s_{\text{Afr},t_{\text{Afr},p1}}$ 0.029	$s_{\text{Eur},t_{\text{Eur},p1}}$ 0.227	$t_{\text{Afr},p1}$ 12	$t_{\text{Eur},p2}$ 16	...
...						
SimN	$s_{\text{Afr},0}$ 0.51	$s_{\text{Afr},t_{\text{Afr},p1}}$ 0.326	$s_{\text{Eur},t_{\text{Eur},p1}}$ 0.048	$t_{\text{Afr},p1}$ 3	$t_{\text{Eur},p2}$ 4	...

→

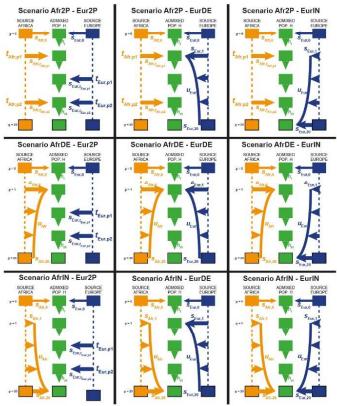
Scenario Afr2P - EurDE						
Sim1	$s_{\text{Afr},0}$ 0.22	$s_{\text{Afr},t_{\text{Afr},p1}}$ 0.047	$s_{\text{Eur},1}$ 0.874	$t_{\text{Afr},p1}$ 6	u_s 0.12	...
Sim2	$s_{\text{Afr},0}$ 0.02	$s_{\text{Afr},t_{\text{Afr},p1}}$ 0.721	$s_{\text{Eur},1}$ 0.356	$t_{\text{Afr},p1}$ 4	u_s 0.34	...
...						
SimN	$s_{\text{Afr},0}$ 0.10	$s_{\text{Afr},t_{\text{Afr},p1}}$ 0.344	$s_{\text{Eur},1}$ 0.089	$t_{\text{Afr},p1}$ 9	u_s 0.04	...

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

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

Complex admixture histories reconstructed with MetHis-ABC

Set by the user



Parameter prior distributions

$s_{\text{Afr},0}$	Uniform [0,1]
$s_{\text{Afr},t_{\text{Afr},p1}}$	$s_{\text{Eur},t_{\text{Eur},p1}}$ Uniform [0,1]
$s_{\text{Afr},t_{\text{Afr},p2}}$	$s_{\text{Eur},t_{\text{Eur},p2}}$ Uniform [0,1]
$t_{\text{Afr},p1}$	$t_{\text{Eur},p1}$ Uniform [1,20]
$t_{\text{Afr},p2}$	$t_{\text{Eur},p2}$ Uniform [1,20]
$t_{\text{Afr},p1} \neq t_{\text{Afr},p2}$	$t_{\text{Eur},p1} \neq t_{\text{Eur},p2}$
<hr/>	
$s_{\text{Afr},0}$	Uniform [0,1]
$s_{\text{Afr},1}$	$s_{\text{Eur},1}$ Uniform [0,1]
$s_{\text{Afr},20}$	$s_{\text{Eur},20}$ Uniform [0, $s_{\text{S},20}/3$]
u_{Afr}	u_{Eur} Uniform [0,0.5]
<hr/>	
$s_{\text{Afr},0}$	Uniform [0,1]
$s_{\text{Afr},1}$	$s_{\text{Eur},1}$ Uniform [0, $s_{\text{S},20}/3$]
$s_{\text{Afr},20}$	$s_{\text{Eur},20}$ Uniform [0,1]
u_{Afr}	u_{Eur} Uniform [0,0.5]

Increase Decrease

Parameter Generator

MetHis Tools or Set by the user

Scenario Afr2P - Eur2P						
Sim1	$s_{\text{Afr},0}$ 0.35	$s_{\text{Afr},t_{\text{Afr},p1}}$ 0.003	$s_{\text{Eur},t_{\text{Eur},p1}}$ 0.124	$t_{\text{Afr},p1}$ 4	$t_{\text{Eur},p2}$ 11	...
...						
Sim2	$s_{\text{Afr},0}$ 0.74	$s_{\text{Afr},t_{\text{Afr},p1}}$ 0.029	$s_{\text{Eur},t_{\text{Eur},p1}}$ 0.227	$t_{\text{Afr},p1}$ 12	$t_{\text{Eur},p2}$ 16	...
...						
SimN	$s_{\text{Afr},0}$ 0.51	$s_{\text{Afr},t_{\text{Afr},p1}}$ 0.326	$s_{\text{Eur},t_{\text{Eur},p1}}$ 0.048	$t_{\text{Afr},p1}$ 3	$t_{\text{Eur},p2}$ 4	...
Scenario Afr2P - EurDE						
Sim1	$s_{\text{Afr},0}$ 0.22	$s_{\text{Afr},t_{\text{Afr},p1}}$ 0.047	$s_{\text{Eur},1}$ 0.874	$t_{\text{Afr},p1}$ 6	u_s 0.12	...
Sim2	$s_{\text{Afr},0}$ 0.02	$s_{\text{Afr},t_{\text{Afr},p1}}$ 0.721	$s_{\text{Eur},1}$ 0.356	$t_{\text{Afr},p1}$ 4	u_s 0.34	...
...						
SimN	$s_{\text{Afr},0}$ 0.10	$s_{\text{Afr},t_{\text{Afr},p1}}$ 0.344	$s_{\text{Eur},1}$ 0.089	$t_{\text{Afr},p1}$ 9	u_s 0.04	...
...						

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

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

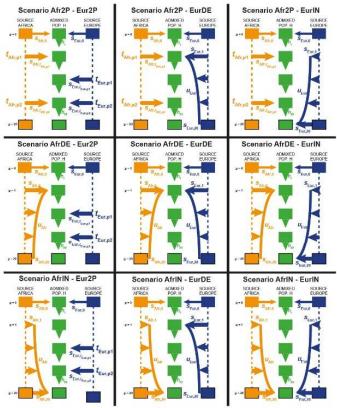
MetHis Simulation

Scenario Afr2P - Eur2P						
Sim1	$s_{\text{Afr},0}$ 0.35	$s_{\text{Afr},t_{\text{Afr},p1}}$ 0.003	$s_{\text{Eur},t_{\text{Eur},p1}}$ 0.124	$t_{\text{Afr},p1}$ 4	$t_{\text{Eur},p2}$ 11	...
X						
Afr						
Admix						
Eur						
...						
Sim28	$s_{\text{Afr},0}$ 0.35	$s_{\text{Afr},t_{\text{Afr},p1}}$ 0.003	$s_{\text{Eur},1}$ 0.874	$t_{\text{Afr},p1}$ 4	u_s 0.12	...
X						
Afr						
Admix						
Eur						
...						
Sim3	$s_{\text{Afr},0}$ 0.78	$s_{\text{Afr},t_{\text{Afr},p1}}$ 0.072	$s_{\text{Eur},1}$ 0.013	$t_{\text{Afr},p1}$ 6	u_s 0.38	...
X						
Afr						
Admix						
Eur						
...						

SIMULATIONS

Complex admixture histories reconstructed with MetHis-ABC

Set by the user



Parameter prior distributions

$s_{Afr,0}$ Uniform [0,1]

$s_{Afr,t_{Afr,p1}}$ $s_{Eur,t_{Eur,p1}}$ Uniform [0,1]

$s_{Afr,t_{Afr,p2}}$ $s_{Eur,t_{Eur,p2}}$ Uniform [0,1]

$t_{Afr,p1}$ $t_{Eur,p1}$ Uniform [1,20]

$t_{Afr,p2}$ $t_{Eur,p2}$ Uniform [1,20]

$t_{Afr,p1} \neq t_{Afr,p2}$ $t_{Eur,p1} \neq t_{Eur,p2}$

$s_{Afr,0}$ Uniform [0,1]

$s_{Afr,f}$ $s_{Eur,f}$ Uniform [0,1]

$s_{Afr,20}$ $s_{Eur,20}$ Uniform [0, $s_{s_1}/3$]

u_{Afr} u_{Eur} Uniform [0,0.5]

$s_{Afr,0}$ Uniform [0,1]

$s_{Afr,f}$ $s_{Eur,f}$ Uniform [0, $s_{s_2}/3$]

$s_{Afr,20}$ $s_{Eur,20}$ Uniform [0,1]

u_{Afr} u_{Eur} Uniform [0,0.5]

Parameter Generator

MetHis Tools or Set by the user

Scenario Afr2P - Eur2P					
Sim1	$s_{Afr,0}$	$s_{Afr,t_{Afr,p1}}$	$s_{Eur,t_{Eur,p1}}$	$t_{Afr,p1}$	$t_{Eur,p2}$
	0.35	0.003	0.124	4	11
...					
Sim2	$s_{Afr,0}$	$s_{Afr,t_{Afr,p1}}$	$s_{Eur,t_{Eur,p1}}$	$t_{Afr,p1}$	$t_{Eur,p2}$
	0.74	0.029	0.227	12	16
...					
SimN	$s_{Afr,0}$	$s_{Afr,t_{Afr,p1}}$	$s_{Eur,t_{Eur,p1}}$	$t_{Afr,p1}$	$t_{Eur,p2}$
	0.51	0.326	0.048	3	4

Scenario Afr2P - EurDE					
Sim1	$s_{Afr,0}$	$s_{Afr,t_{Afr,p1}}$	$s_{Eur,1}$	$t_{Afr,p1}$	u_s
	0.22	0.047	0.874	6	0.12
...					
Sim2	$s_{Afr,0}$	$s_{Afr,t_{Afr,p1}}$	$s_{Eur,1}$	$t_{Afr,p1}$	u_s
	0.02	0.721	0.356	4	0.34
...					
SimN	$s_{Afr,0}$	$s_{Afr,t_{Afr,p1}}$	$s_{Eur,1}$	$t_{Afr,p1}$	u_s
	0.10	0.344	0.089	9	0.04

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

SIMULATIONS

MetHis Simulation

Scenario Afr2P - Eur2P					
Sim1	$s_{Afr,0}$	$s_{Afr,t_{Afr,p1}}$	$s_{Eur,t_{Eur,p1}}$	$t_{Afr,p1}$	$t_{Eur,p2}$
	0.35	0.003	0.124	4	11
X					
Afr	AGTCATTACG...	GATCATTACG...	AGTCATTACG...	GATCATTACG...	AGTCATTACG...
Admix	GATCATTACG...	GGATCACCAT...	GGATCACCAT...	GGATCACCAT...	GGATCACCAT...
Eur	GGATCACCAT...	CCATAGGATC...	CCATAGGATC...	CCATAGGATC...	CCATAGGATC...

Scenario Afr2P - EurDE					
Sim1	$s_{Afr,0}$	$s_{Afr,t_{Afr,p1}}$	$s_{Eur,1}$	$t_{Afr,p1}$	u_s
	0.35	0.003	0.874	4	0.12
X					
Afr	GCTGATTACCG...	GATCATTACG...	GCTGATTACCG...	GATCATTACG...	GCTGATTACCG...
Admix	GCTGATTACCG...	GGCTCACCGAT...	GCTGATTACCG...	GGCTCACCGAT...	GGCTCACCGAT...
Eur	GGCTCACCGAT...	CAATAGGATC...	GGCTCACCGAT...	CAATAGGATC...	GGCTCACCGAT...

Scenario Afr2P - EurIN					
Sim1	$s_{Afr,0}$	$s_{Afr,t_{Afr,p1}}$	$s_{Eur,1}$	$t_{Afr,p1}$	u_s
	0.78	0.072	0.013	6	0.38
X					
Afr	GCGCAATTACG...	GATCATTACG...	GCGCAATTACG...	GATCATTACG...	GCGCAATTACG...
Admix	GCGCAATTACG...	TCTCTTACCAT...	GCGCAATTACG...	TCTCTTACCAT...	GCGCAATTACG...
Eur	TCTCTTACCAT...	CAGTAGGATC...	TCTCTTACCAT...	CAGTAGGATC...	TCTCTTACCAT...

MetHis Summary Statistics

Scenario Afr2P - Eur2P					
Sim1	$s_{Afr,0}$	$s_{Afr,t_{Afr,p1}}$	$s_{Eur,t_{Eur,p1}}$	$t_{Afr,p1}$	$t_{Eur,p2}$
	0.35	0.003	0.124	4	11
X					
HetAfr	0.457	HetAdmix	0.697	HetEur	0.335
f_{-3}	0.0087	F_{ST} Afr-Ad	0.147

Scenario Afr2P - EurDE					
Sim28	$s_{Afr,0}$	$s_{Afr,t_{Afr,p1}}$	$s_{Eur,1}$	$t_{Afr,p1}$	u_s
	0.35	0.003	0.874	4	0.12
X					
HetAfr	0.457	HetAdmix	0.697	HetEur	0.335
f_{-3}	0.0087	F_{ST} Afr-Ad	0.147

Scenario Afr2P - EurIN					
Sim3	$s_{Afr,0}$	$s_{Afr,t_{Afr,p1}}$	$s_{Eur,1}$	$t_{Afr,p1}$	u_s
	0.78	0.072	0.013	6	0.38
X					
HetAfr	0.457	HetAdmix	0.697	HetEur	0.335
f_{-3}	0.0087	F_{ST} Afr-Ad	0.147

SUMMARY STATISTICS CALCULATION

MetHis ref. tables for "abc" and "abcrf" R packages

Complex admixture histories reconstructed with MetHis-ABC

It is (about) time that we see whether all this works in reality !

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 îles Françaises de l'Amérique, & pour la Discipline & le Commerce des Nègres & Esclaves dans ledit Pays.

Donné à Versailles au mois de Mars 1685.

A V E C ,

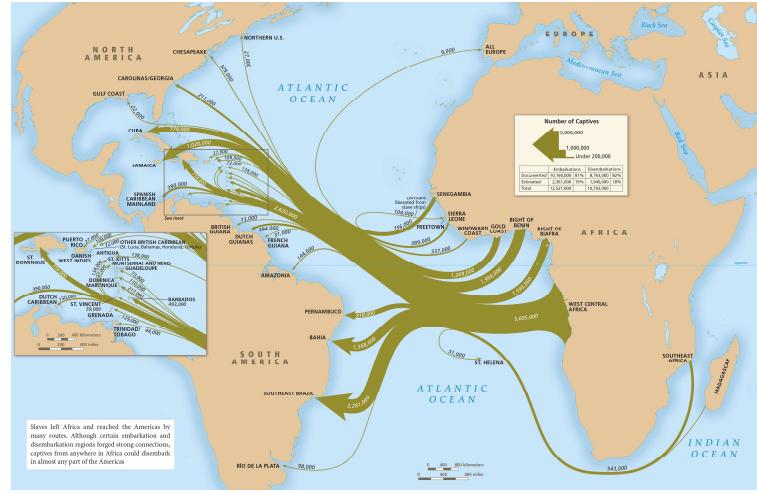
L'EDIT du mois d'Août 1685, portant établissement d'un Conseil Souverain & de quatre Sièges Royaux dans la Côte de l'île de S. Domingue.



A PARIS.

Chez la Veuve SAUVRAIN, à l'entrée du Quay de Gréves,
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

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 îles Françaises de l'Amérique, & pour la Discipline & le Commerce des Nègres & Esclaves dans ledit Pays.

Donné à Verfaille au mois de Mars 1685.

A V E C ,

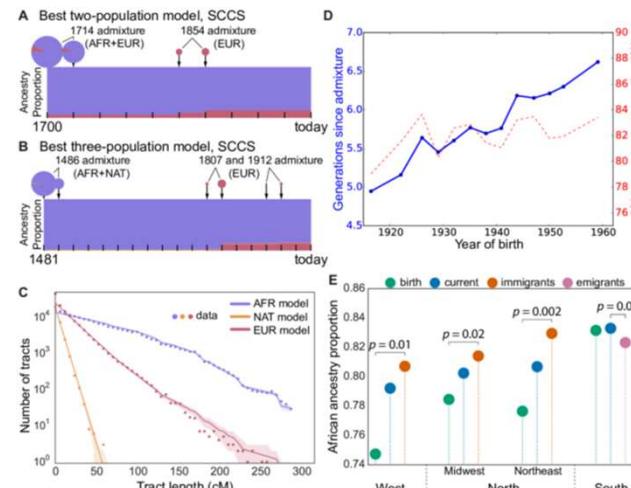
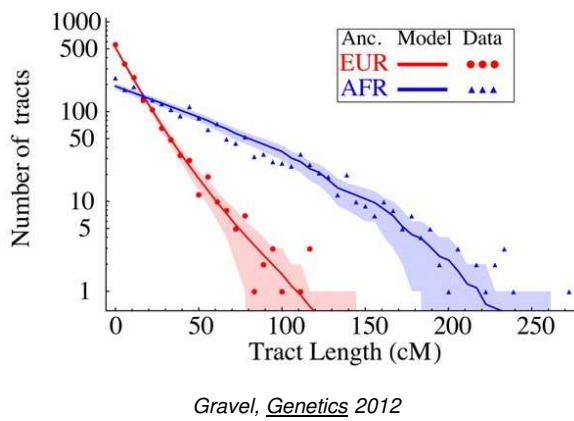
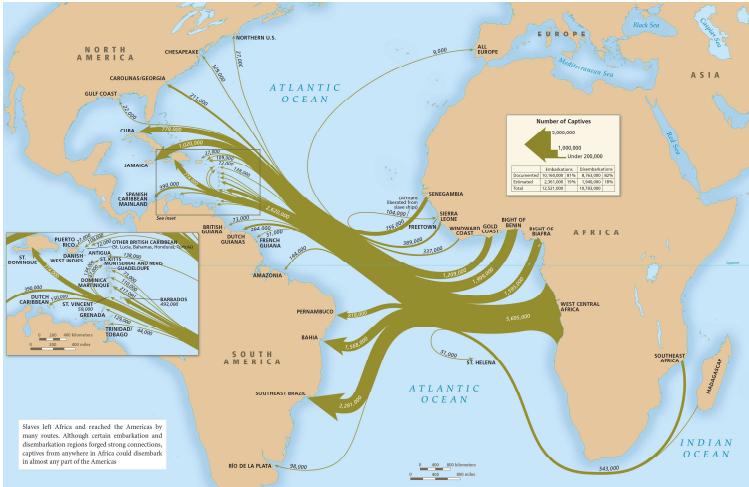
L'EDIT du mois d'Août 1685, portant établissement d'un Conseil Souverain & de quatre Sièges Royaux dans la Côte de l'île de S. Domingue.



A P A R I S .

Chez la Veuve SAUGRAIN, à l'entrée du Quay de Gréves,
du côté du Pont au Change, au Paradis.

M D C C X V I I I .



Transatlantic Slave Trade:

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

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.

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

Real Data Set

Previous studies found two proxy source populations for the admixture patterns in Afro-American and Barbadian populations: **British (GBR)** and **Yoruba (YRI)**

- ▶ Two source population admixture models

1000 Genomes Project Phase 3:

British (GBR) ▶ **89** unrelated individuals

Yoruba (YRI) ▶ **90** unrelated individuals

Afro-American (ASW) ▶ **50** unrelated individuals

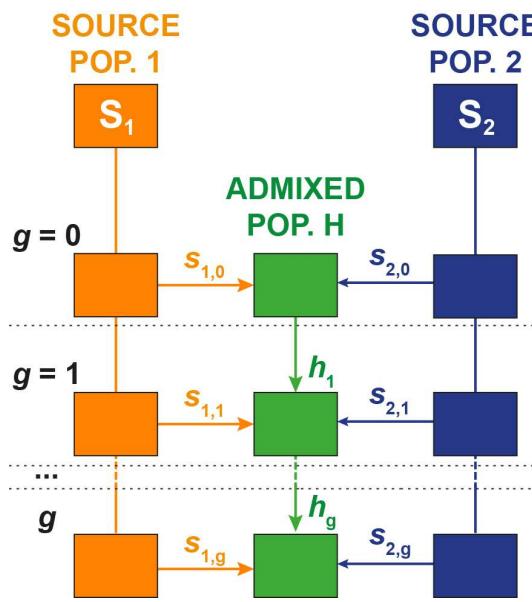
Barbadian (ACB) ▶ **50** unrelated individuals

- ▶ **100,000 independent SNPs** (no singleton, window = 100 SNPs, moving = 10 SNPs, $r^2 < 0.1$, random draw)

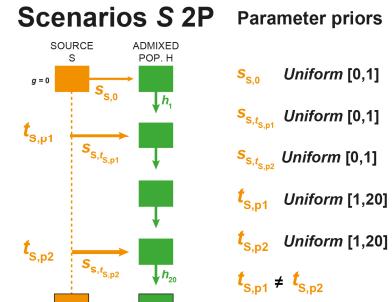
Nine competing models for the admixture history of the ASW and the ACB

21 generations before present ($g = 20$)

= Admixed population founded ~ 1500s



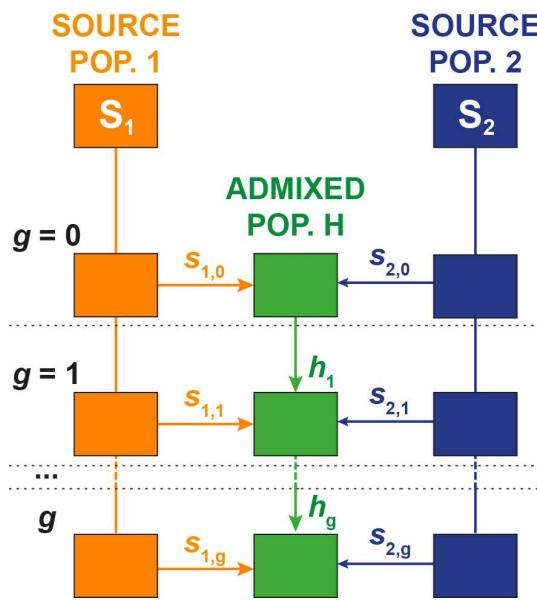
admixture
pulses



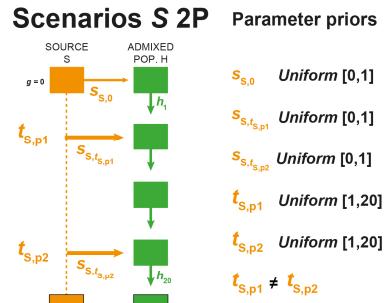
Nine competing models for the admixture history of the ASW and the ACB

21 generations before present ($g = 20$)

= Admixed population founded ~ 1500s



Monotonically decreasing \rightarrow 3 admixture pulses



The diagram illustrates Scenario S DE with two main components: SOURCE S and ADMIXED POP. H.

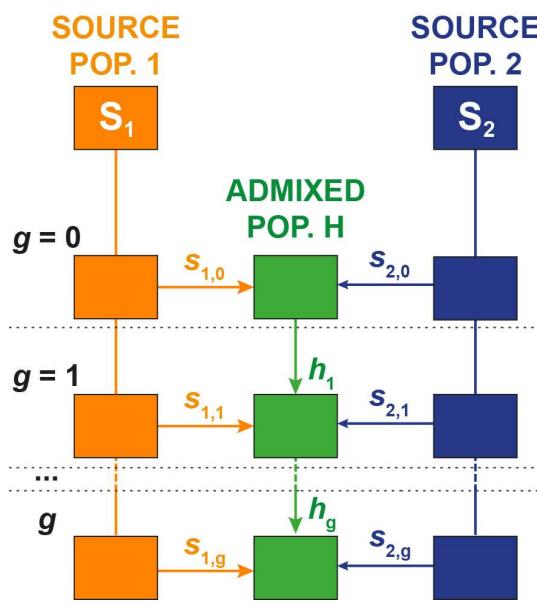
- SOURCE S:** Represented by a green box. A horizontal orange arrow labeled $s_{S,0}$ points from it to a green box labeled h_i . Below this, a vertical dashed orange arrow labeled $s_{S,1}$ points from the green box to a green box below it.
- ADMIXED POP. H:** Represented by a green box. A vertical orange arrow labeled $s_{S,20}$ points from the green box to a green box below it.

Parameter priors:

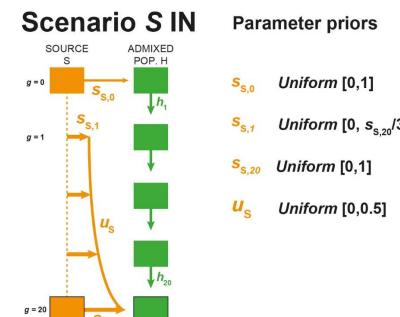
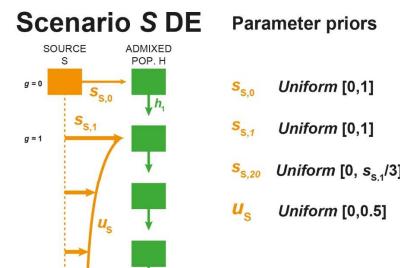
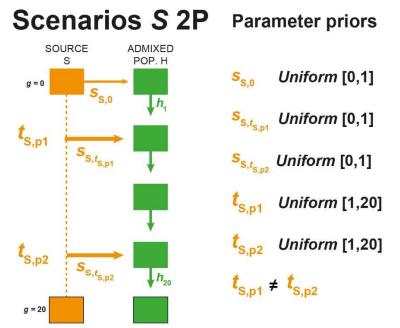
- $s_{S,0}$: Uniform [0,1]
- $s_{S,1}$: Uniform [0,1]
- $s_{S,20}$: Uniform [0, $s_{S,0}$]
- U_s : Uniform [0, 0.5]

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

21 generations before present ($g = 20$)
 = Admixed population founded ~ 1500s



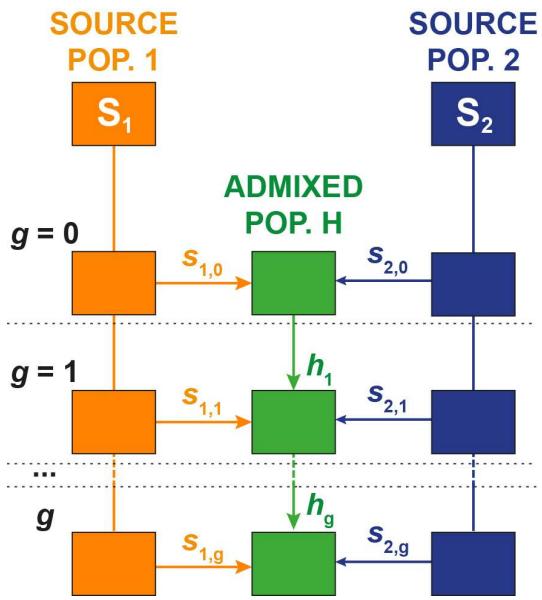
3 admixture pulses
 Monotonically decreasing
 Monotonically increasing



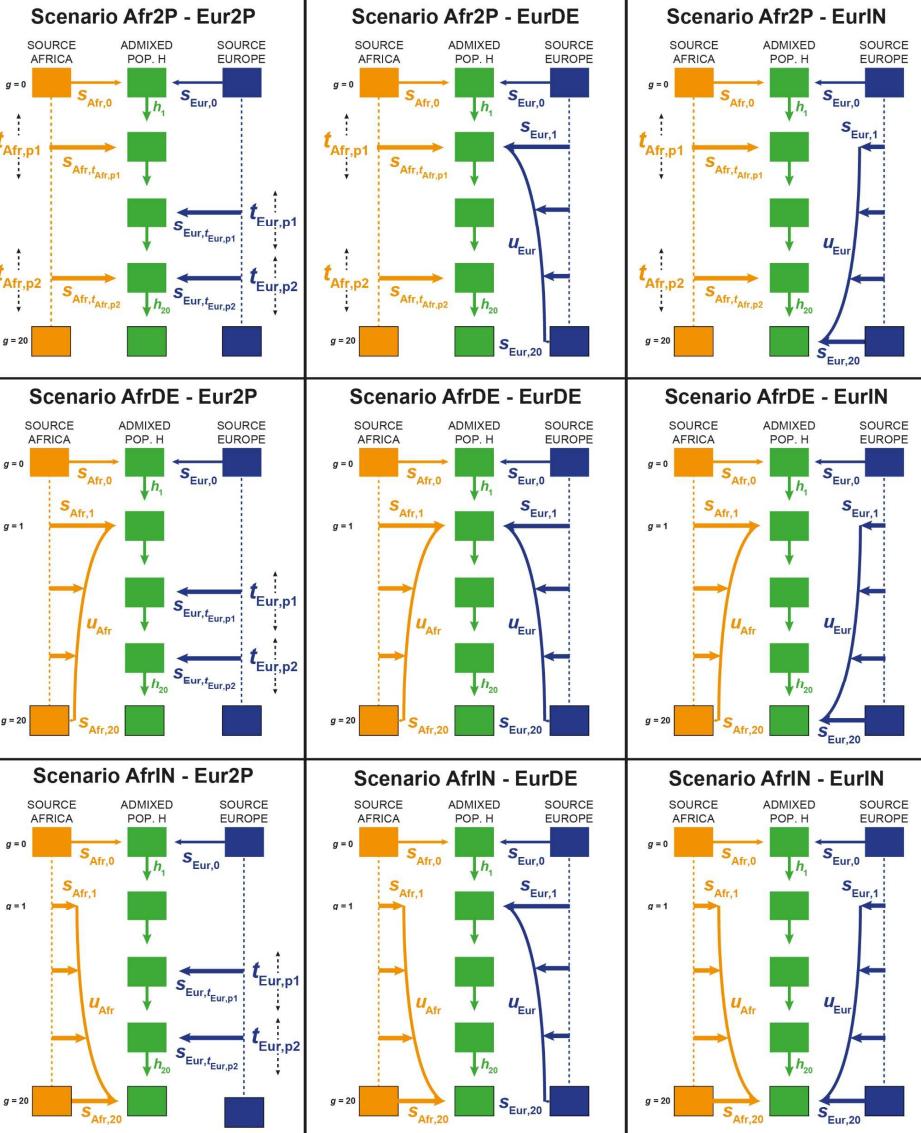
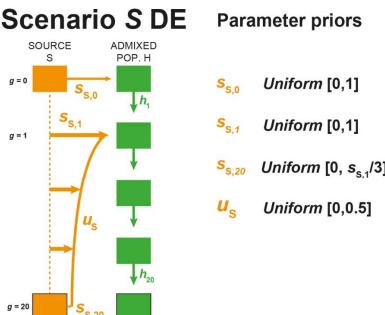
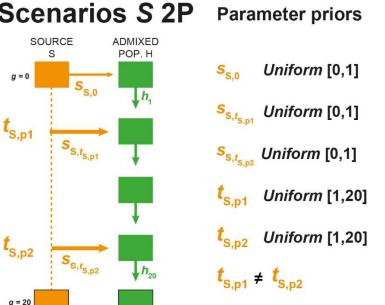
Flip the monotonic-decrease case !

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

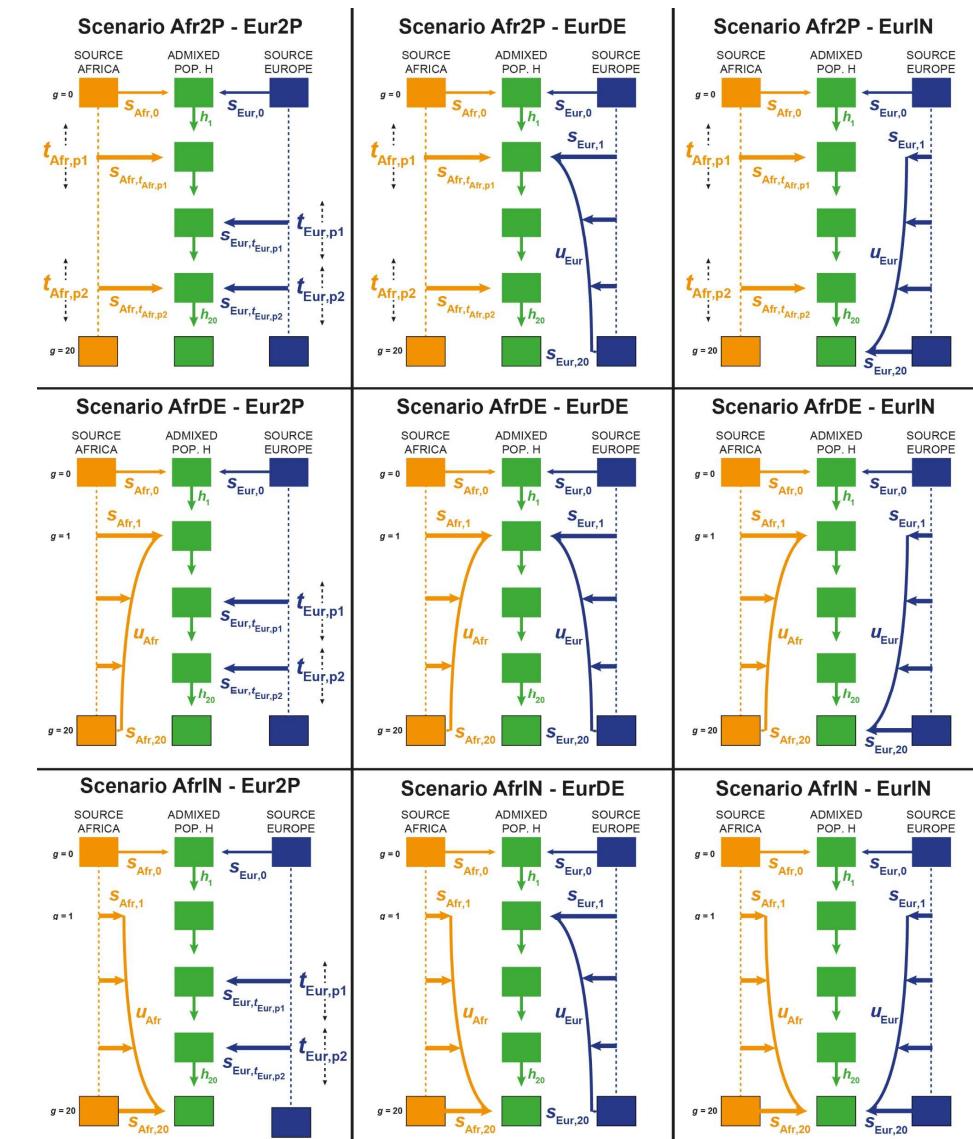
21 generations before present ($g = 20$)
 = Admixed population founded ~ 1500s



3 admixture pulses
 Monotonically decreasing
 Monotonically increasing

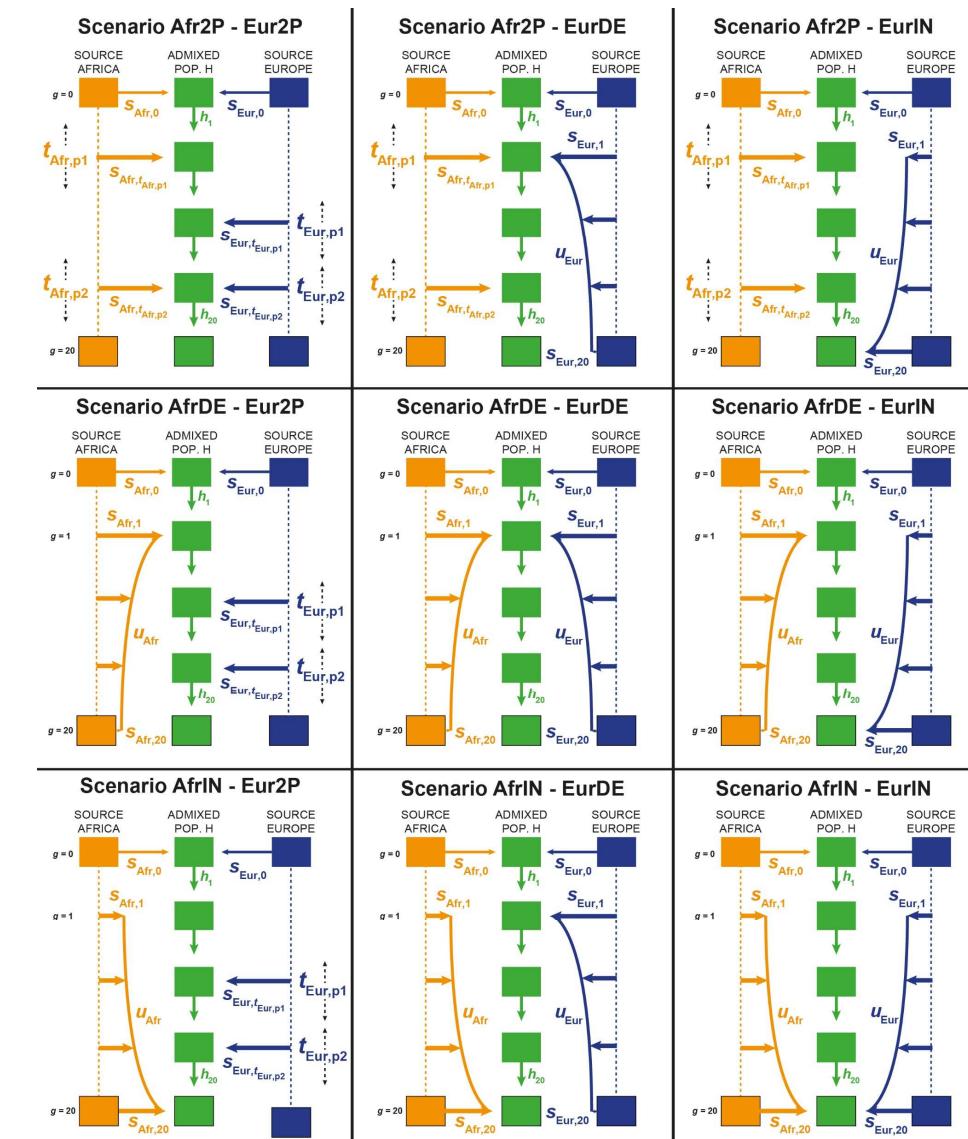


Nine competing models for the admixture history of the ASW and the ACB



- 100,000 independent SNPs
- N_g = constant 1,000 individuals (for simplicity here, *MetHis* option)
- No mutation (for simplicity, 21 generations process, *MetHis* option)
- 10,000 simulations per model = 90,000 simulations total
- Sample 90 Africans, 89 Europeans, and 50 individuals in the admixed population.
- No related individuals at the 2nd degree (*MetHis* option)

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

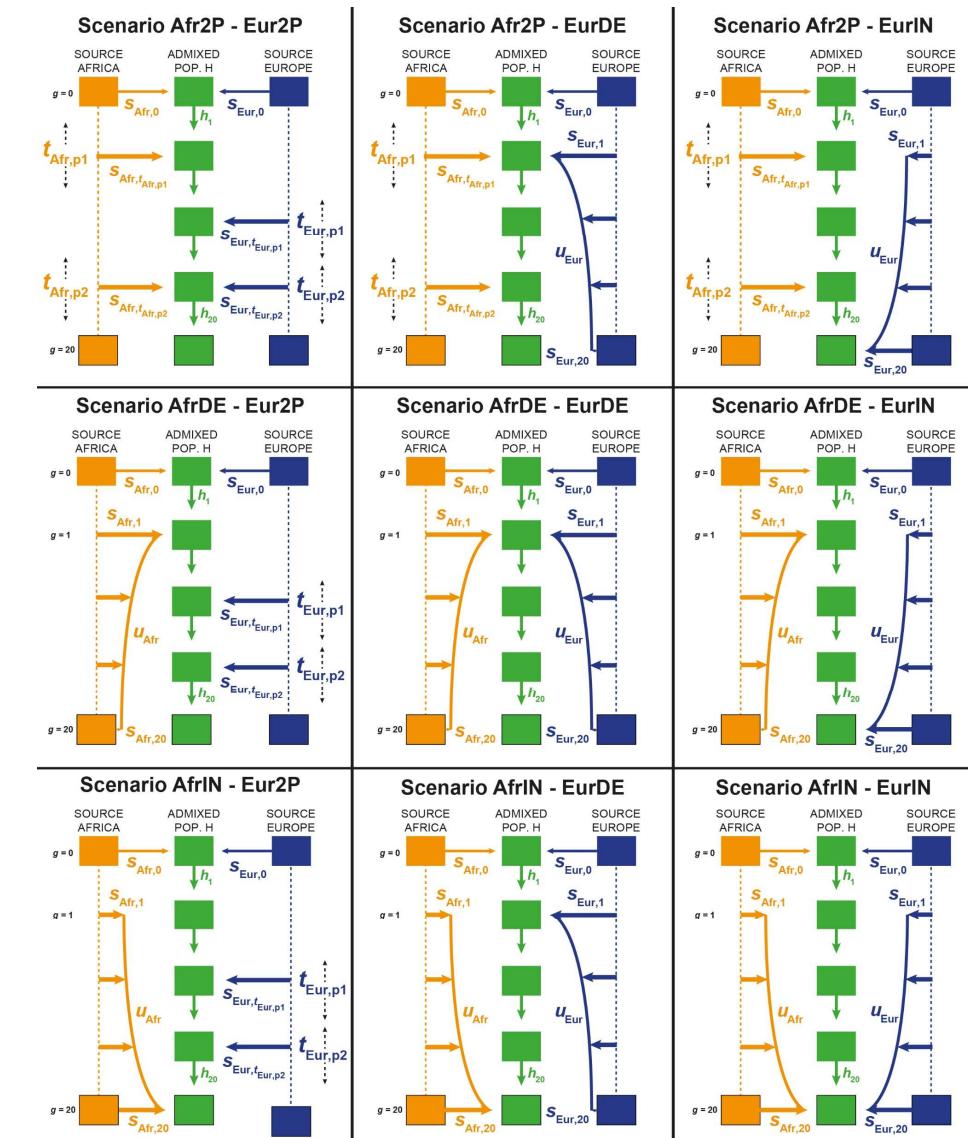


- 100,000 independent SNPs
- N_g = constant 1,000 individuals (for simplicity here, *Methis* option)
- No mutation (for simplicity, 21 generations process, *Methis* option)
- 10,000 simulations per model = 90,000 simulations total
- Sample 90 Africans, 89 Europeans, and 50 individuals in the admixed population.
- No related individuals at the 2nd degree (*Methis* option)

For each simulated data set, calculate 24 summary statistics:

Mean(Het), Var(Het), FST(Afr-H), FST(Eur-H), f3(H; Afr,Eur), Mean ASD, Var ASD,...
+ 16 statistics for admixture fraction distribution in population H.

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



- 100,000 independent SNPs
- N_g = constant 1,000 individuals (for simplicity here, *Methis* option)
- No mutation (for simplicity, 21 generations process, *Methis* option)
- 10,000 simulations per model = 90,000 simulations total
- Sample 90 Africans, 89 Europeans, and 50 individuals in the admixed population.
- No related individuals at the 2nd degree (*Methis* option)

For each simulated data set, calculate 24 summary statistics:

Mean(Het), Var(Het), FST(Afr-H), FST(Eur-H), f3(H; Afr,Eur), Mean ASD, Var ASD,...

+ 16 statistics for admixture fraction distribution in population H.

27 cores -> 3 days calculation total.

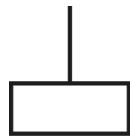
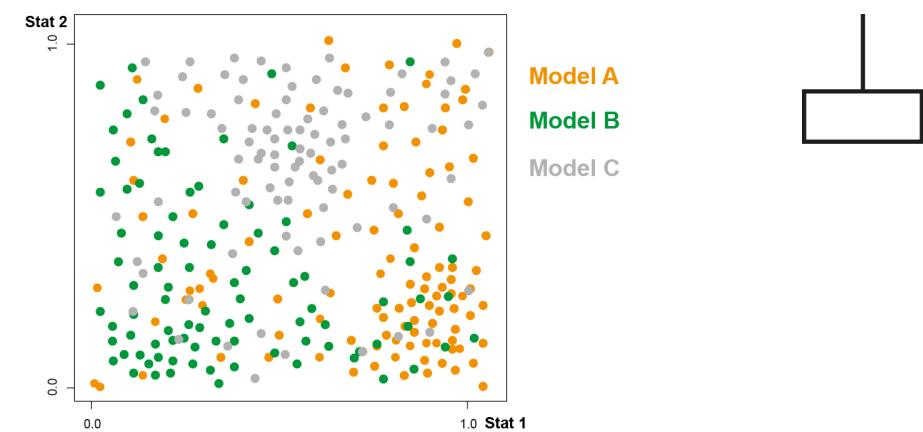
2/3 of this time = summary statistics calculation...

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

Random Forest algorithm for model choice (Breiman 2001)

Draw randomly Stats order

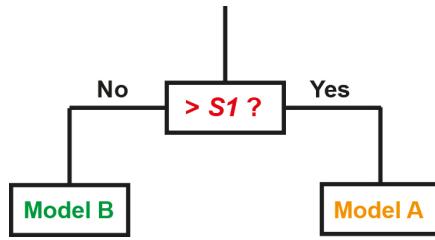
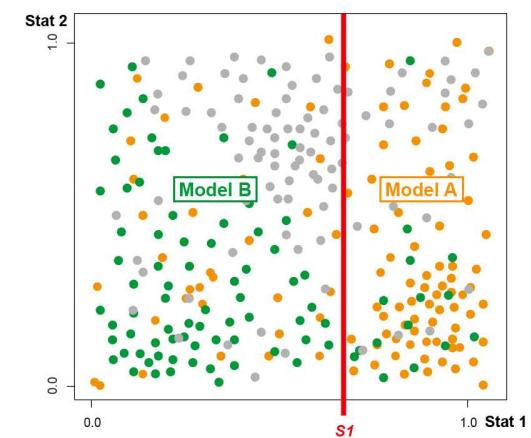
Build Decision Tree 1



Random Forest algorithm for model choice (Breiman 2001)

Draw randomly Stats order

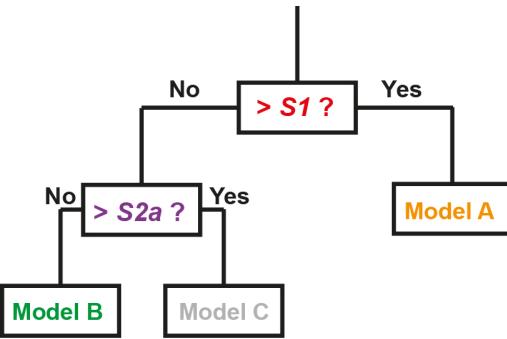
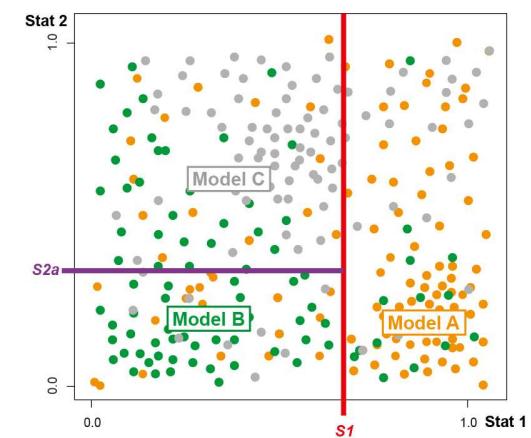
Build Decision Tree 1



Random Forest algorithm for model choice (Breiman 2001)

Draw randomly Stats order

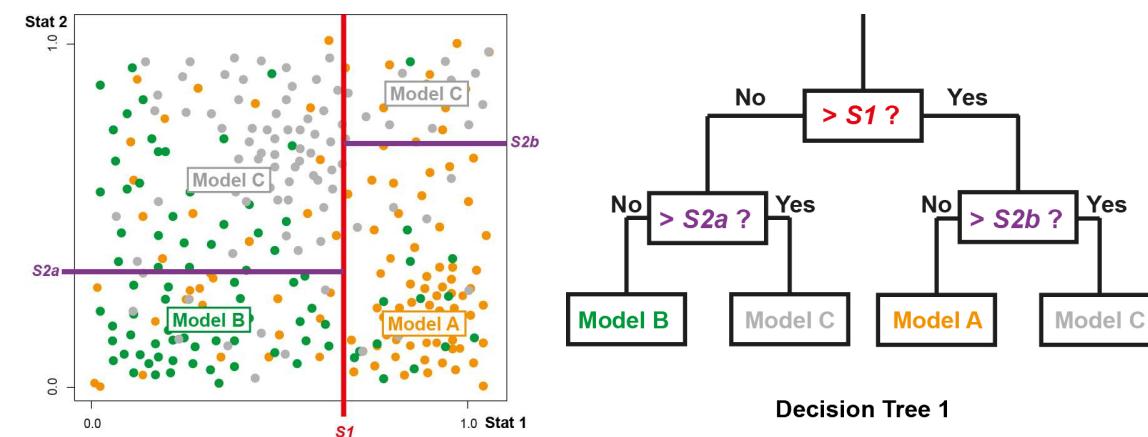
Build Decision Tree 1



Random Forest algorithm for model choice (Breiman 2001)

Draw randomly Stats order

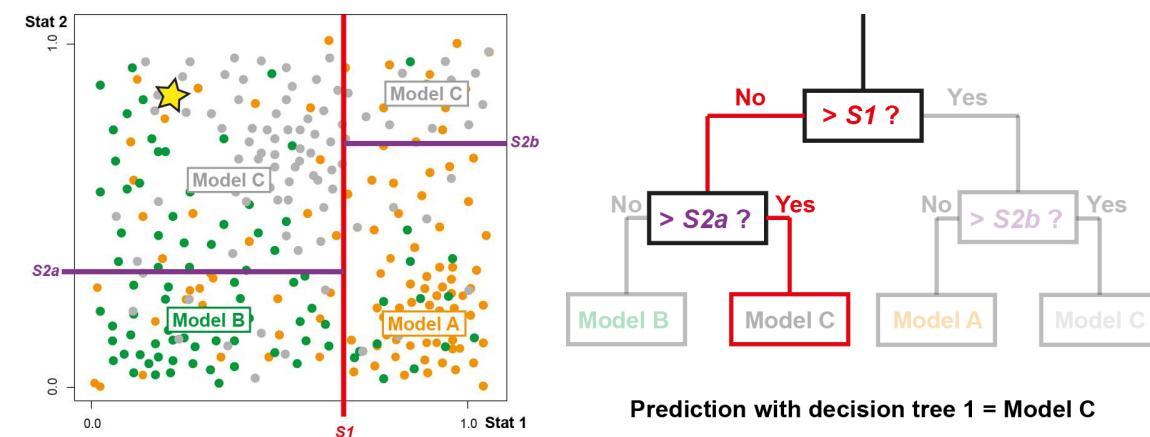
Build Decision Tree 1



Random Forest algorithm for model choice (Breiman 2001)

Draw randomly Stats order

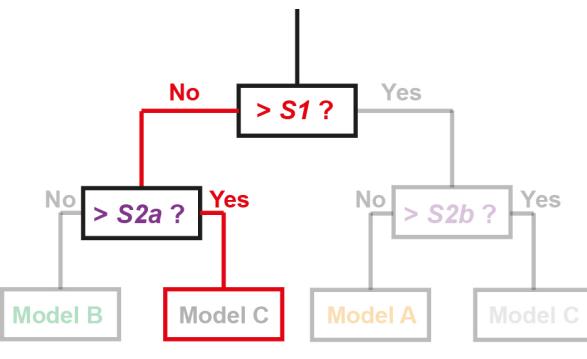
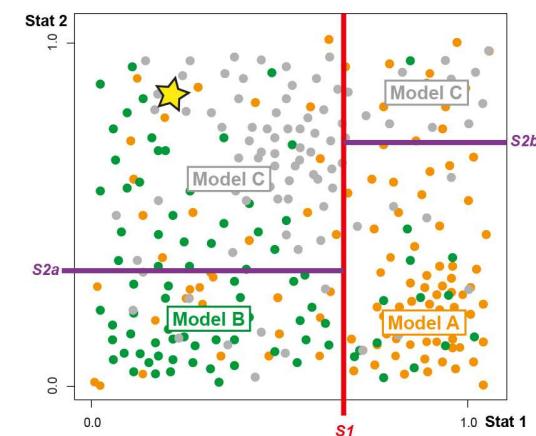
Build Decision Tree 1



Random Forest algorithm for model choice (Breiman 2001)

Draw randomly Stats order

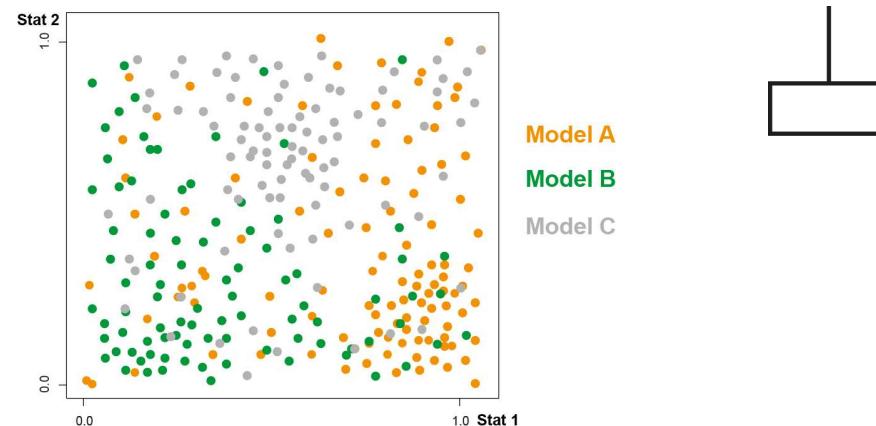
Build Decision Tree 1



Prediction with decision tree 1 = Model C

Draw randomly Stats order

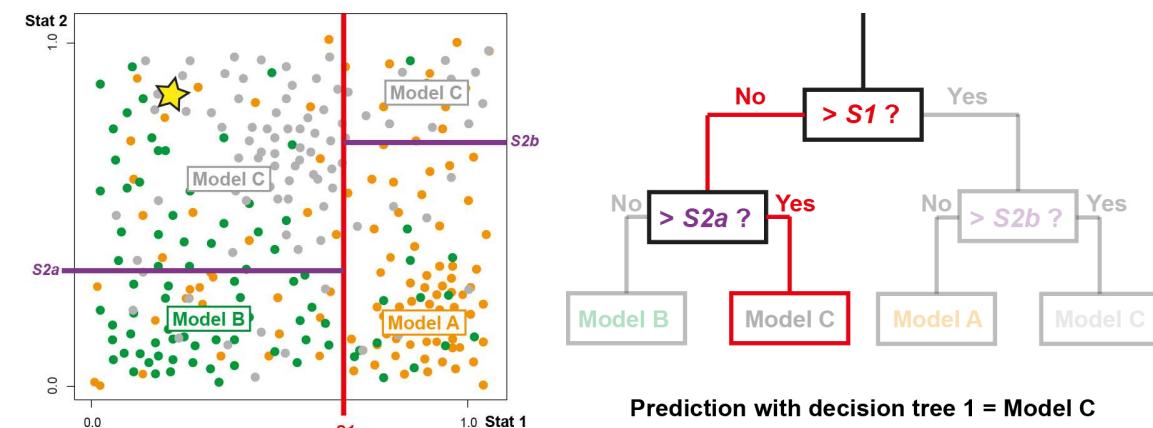
Build Decision Tree 2



Random Forest algorithm for model choice (Breiman 2001)

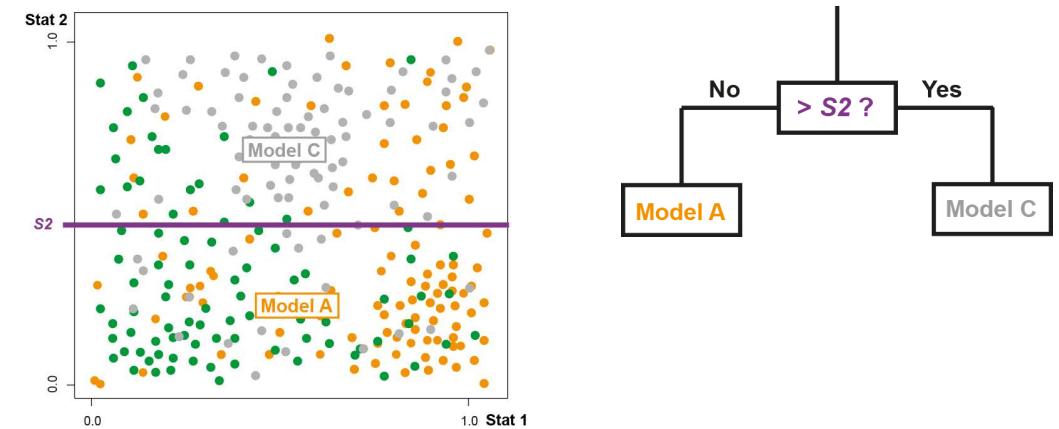
Draw randomly Stats order

Build Decision Tree 1



Draw randomly Stats order

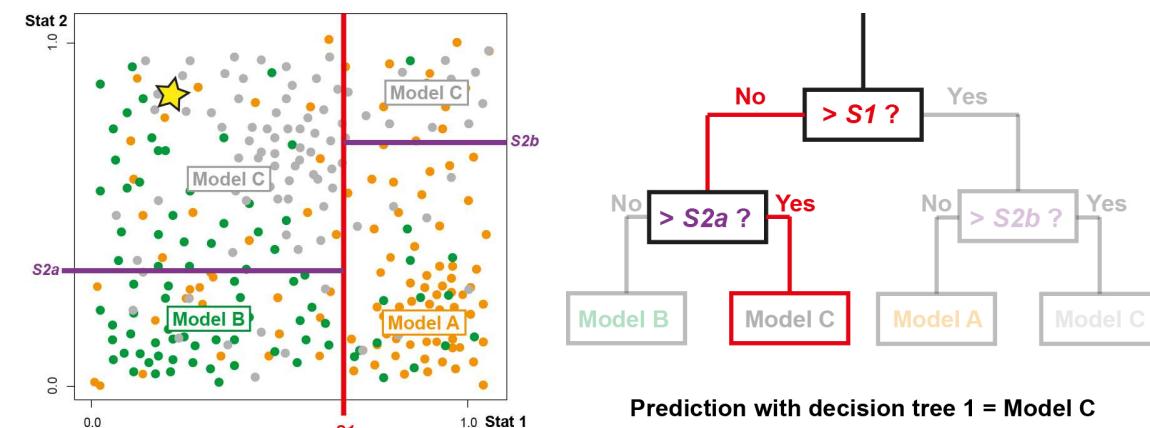
Build Decision Tree 2



Random Forest algorithm for model choice (Breiman 2001)

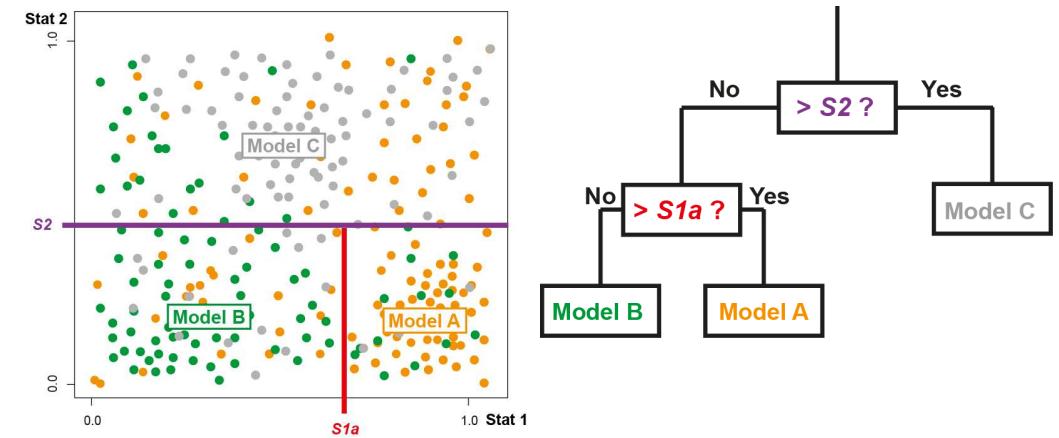
Draw randomly Stats order

Build Decision Tree 1



Draw randomly Stats order

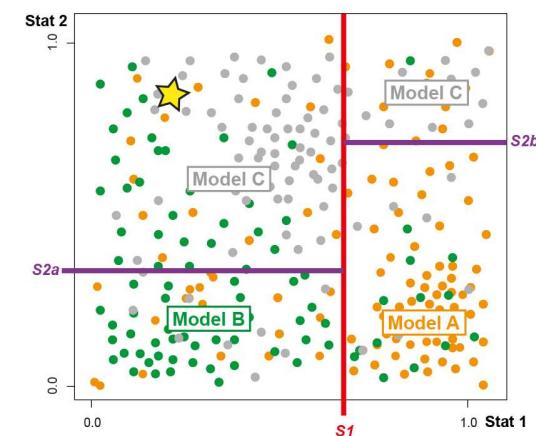
Build Decision Tree 2



Random Forest algorithm for model choice (Breiman 2001)

Draw randomly Stats order

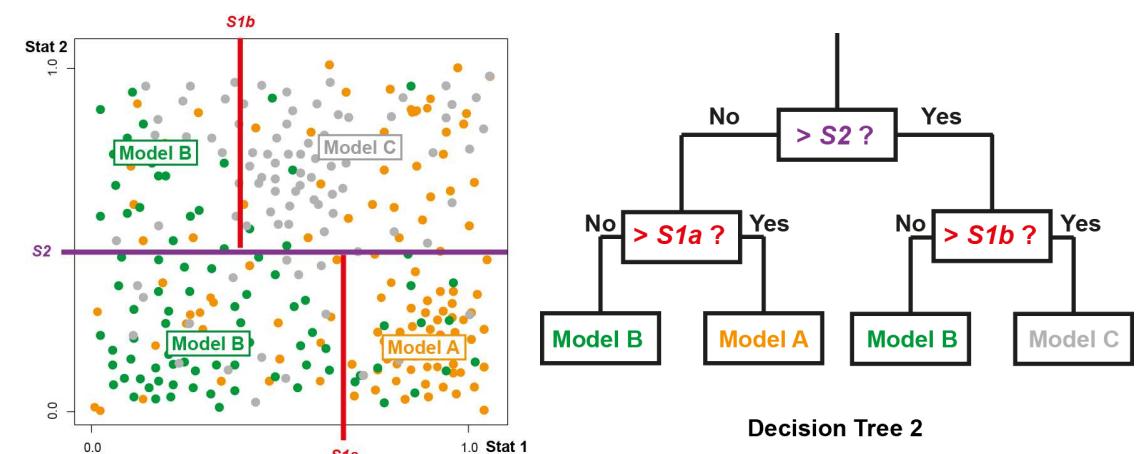
Build Decision Tree 1



Prediction with decision tree 1 = Model C

Draw randomly Stats order

Build Decision Tree 2

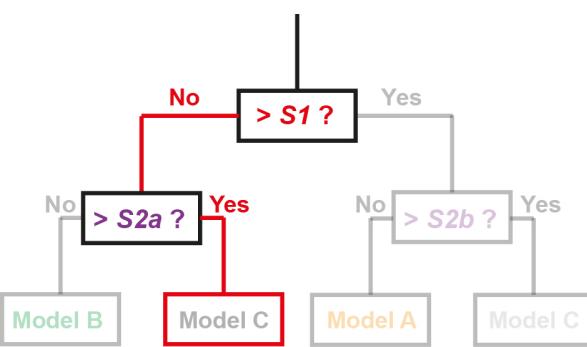
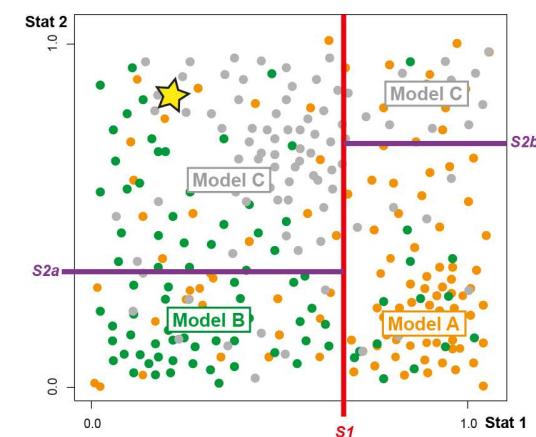


Decision Tree 2

Random Forest algorithm for model choice (Breiman 2001)

Draw randomly Stats order

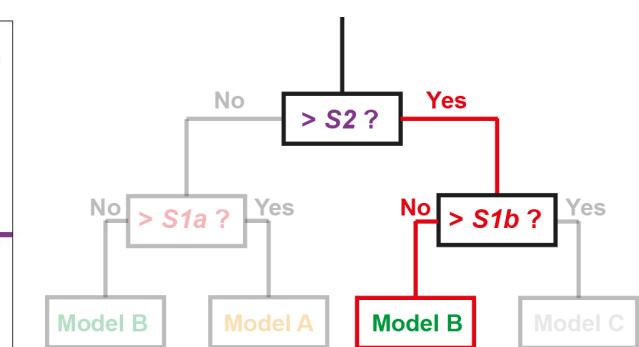
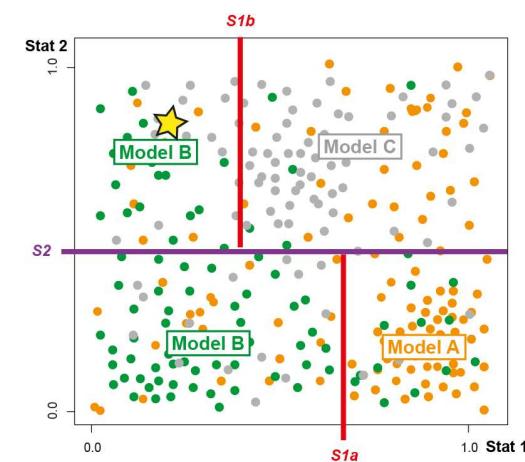
Build Decision Tree 1



Prediction with decision tree 1 = Model C

Draw randomly Stats order

Build Decision Tree 2



Prediction with decision tree 2 = Model B

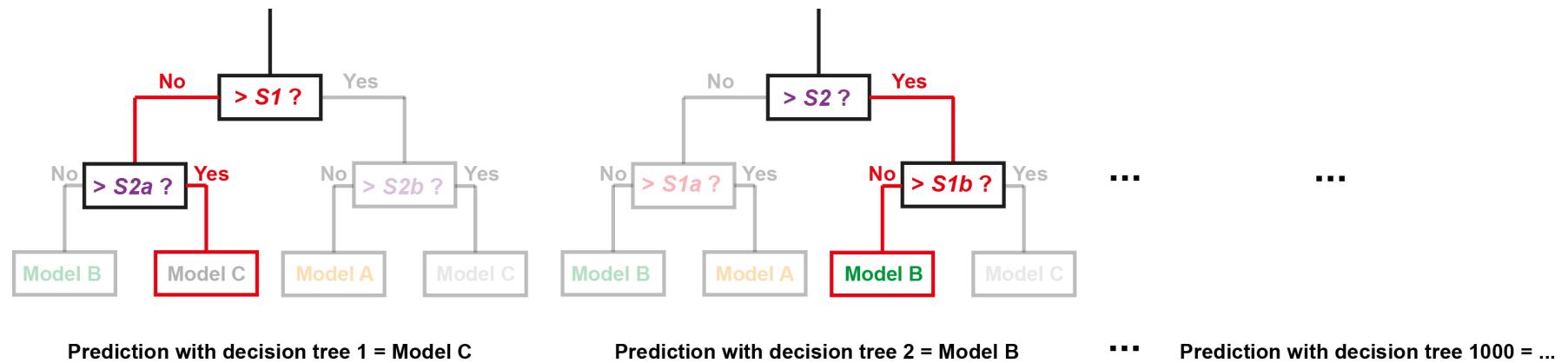
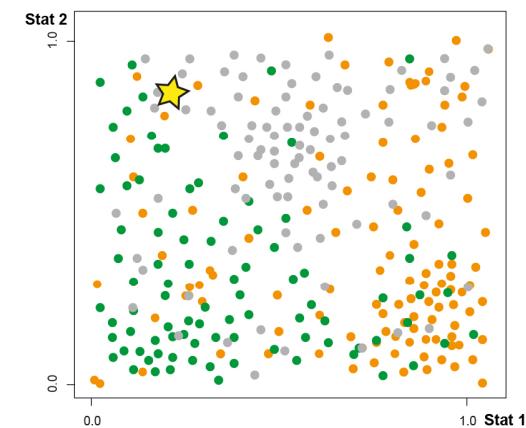
Random Forest algorithm for model choice (Breiman 2001)

Draw randomly Stats order

Build Decision Tree 1

Draw randomly Stats order

Build Decision Tree 2



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 a priori without observed data

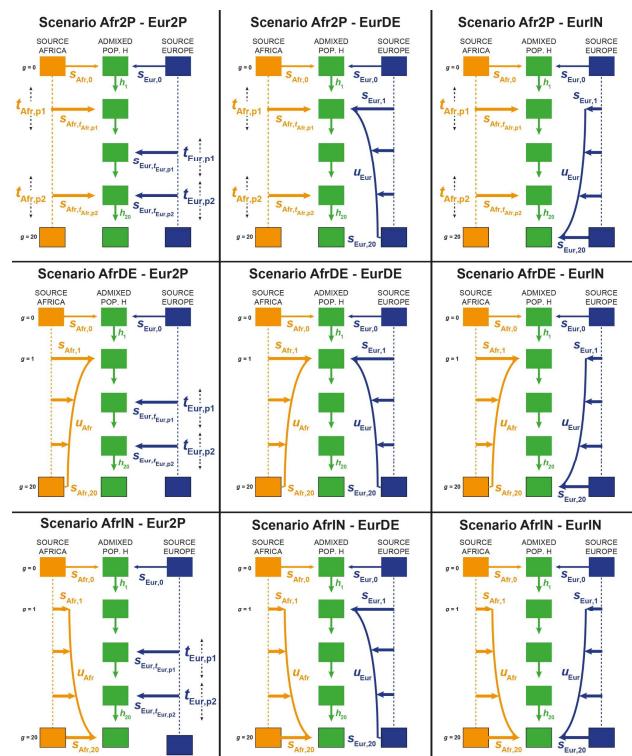
- Random Forest ABC (*abcrf* package in *R*, Pudlo et al. 2016) – 10,000 *MetHis* sims/scenario – 1,000 trees
 - Cross-validation all simulations
 - 100,000 independent SNPs
 - Sample sizes:

AfrIN - EurIN	1.2%	2.7%	2.9%	2.9%

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%

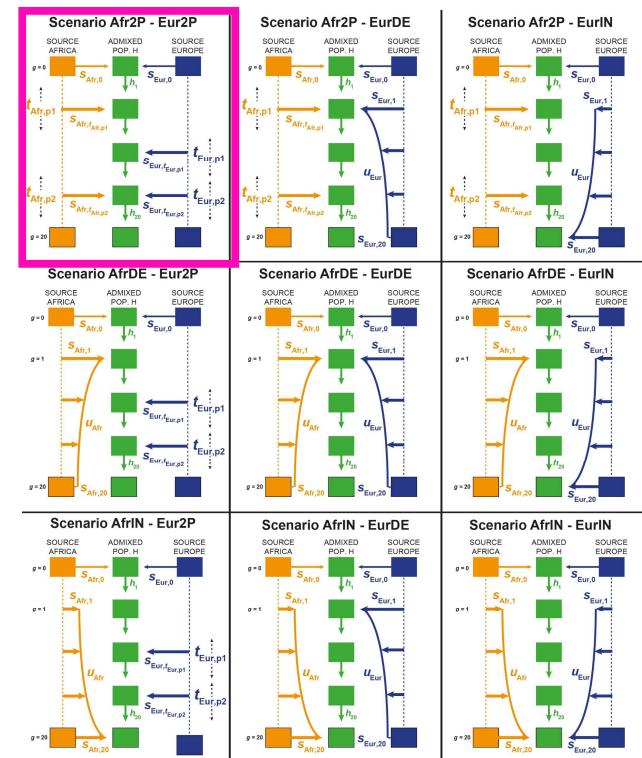
Prior error rate = 32.41%, Model-choice error a priori = 8/9 = 88.89%

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 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

	AfrIN - EurIN	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%

True model

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

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

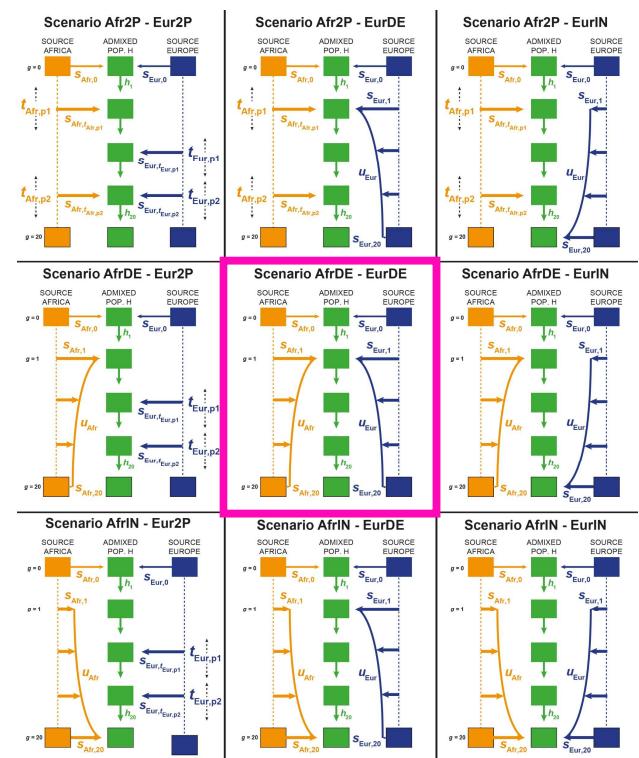
- Random Forest ABC (*abcrf* package in *R*, Pudlo et al. 2016) – 10,000 *MetHis* sims/scenario – 1,000 trees
 - Cross-validation all simulations
 - 100,000 independent SNPs
 - Sample sizes:

AfrIN - EurIN	1.2%	2.7%	2.9%	2.9%

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%

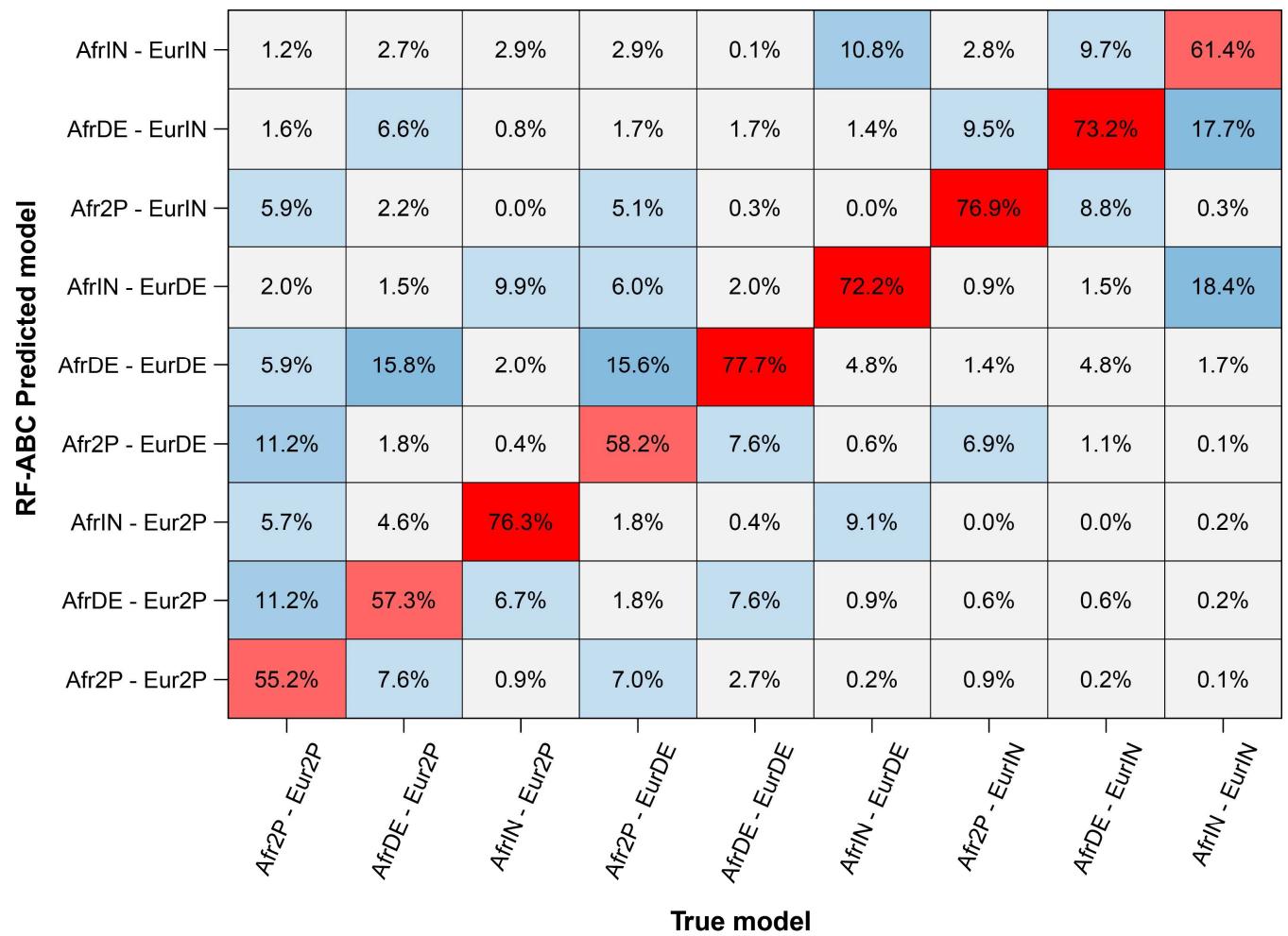
Results: *MetHis*-ABC model-choice with Random-Forest a priori without observed data

- Random Forest ABC (*abcrf* package in *R*, Pudlo et al. 2016) – 10,000 *MetHis* sims/scenario – 1,000 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)

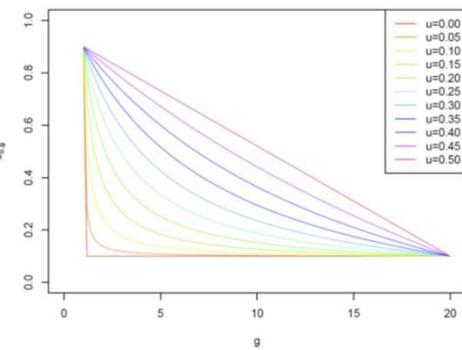
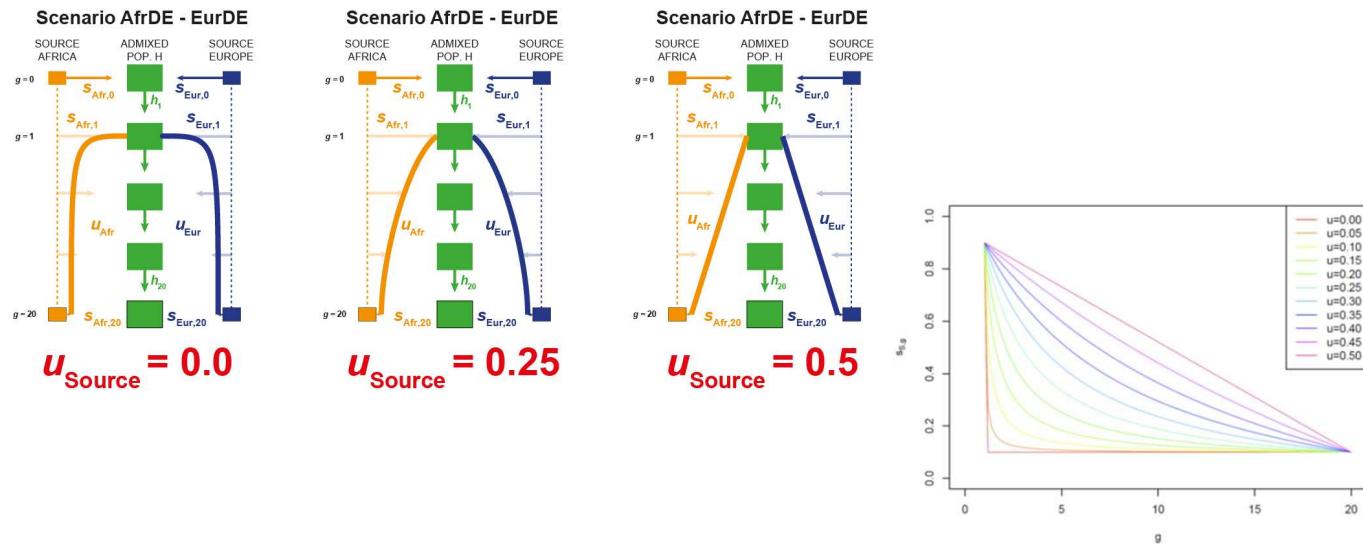
- Erroneous model-choice increased among scenarios qualitatively similar



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

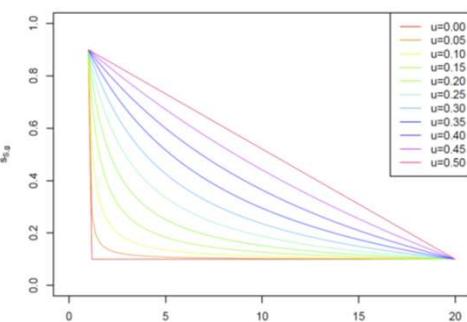
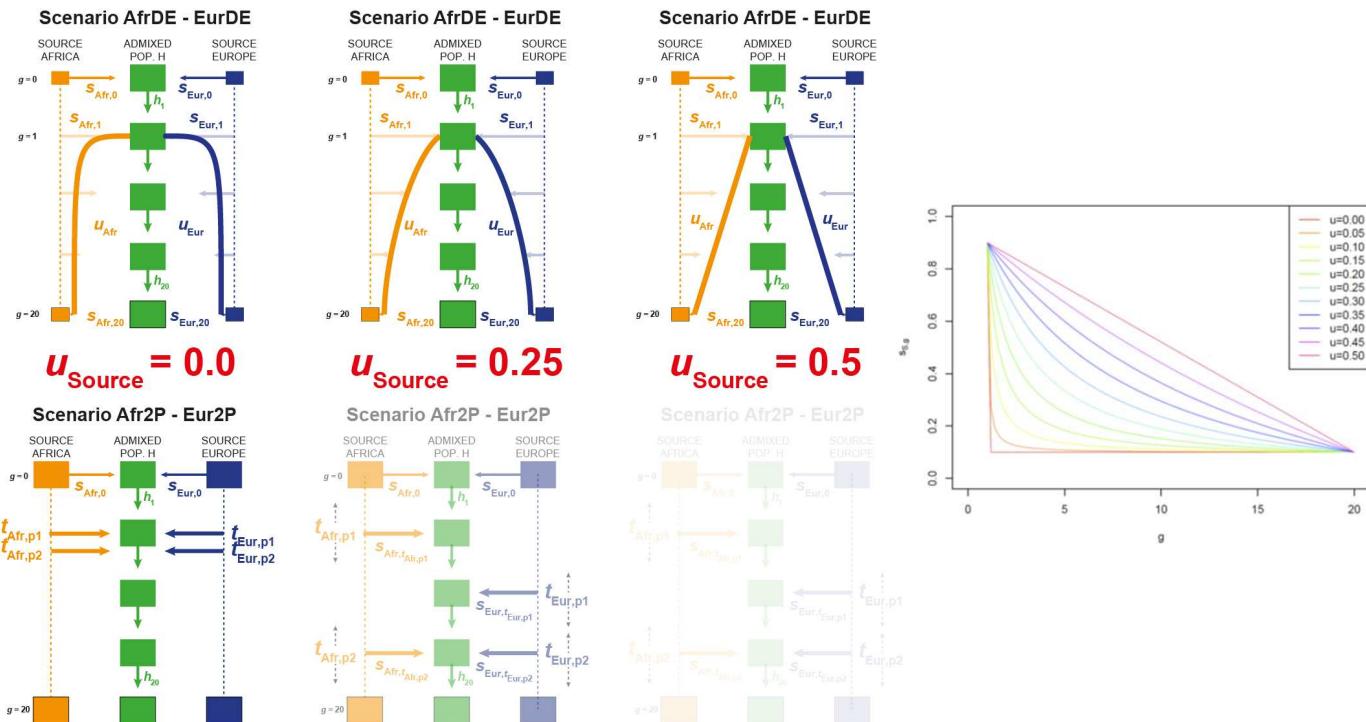
Results: MetHis-ABC model-choice with Random-Forest a priori without observed data

- 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 1000 simulations as pseudo-observed data under the AfrDE-EurDE model



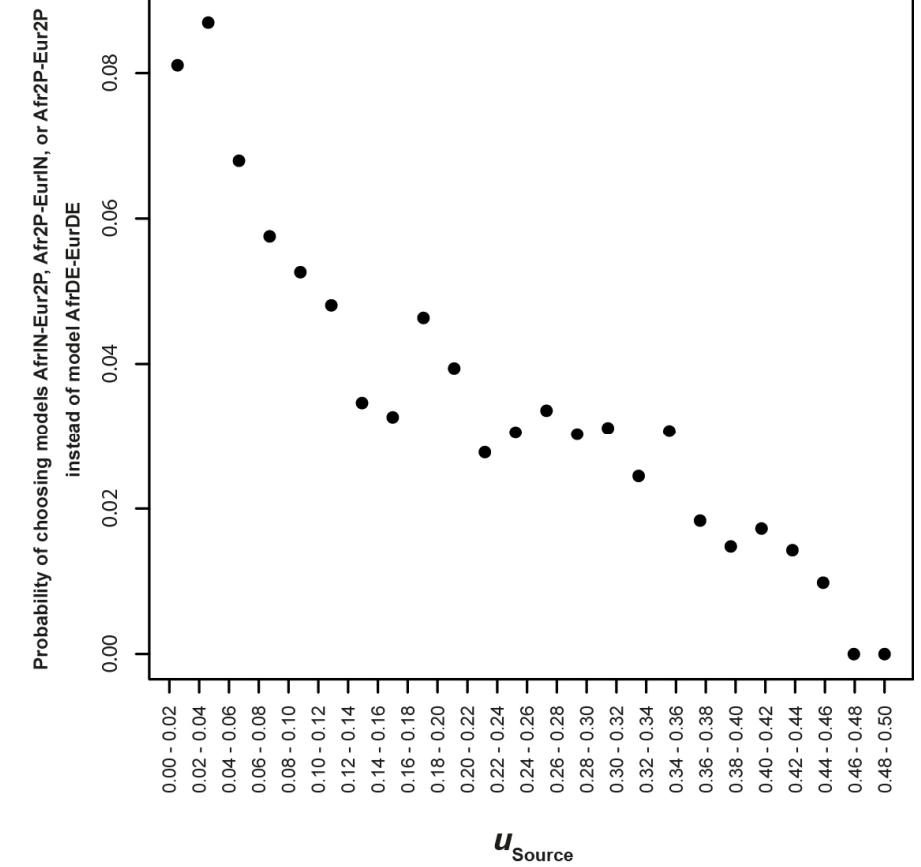
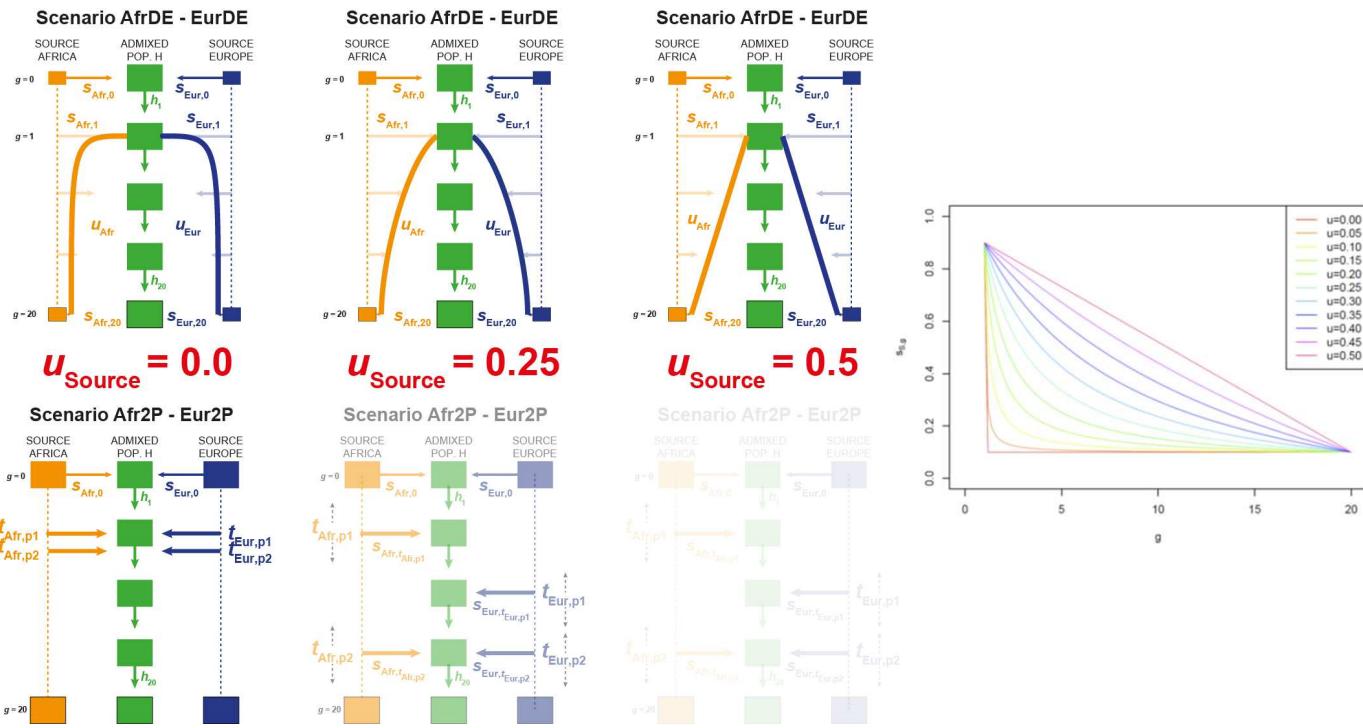
Results: MetHis-ABC model-choice with Random-Forest a priori without observed data

- 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 1000 simulations as pseudo-observed data under the AfrDE-EurDE model



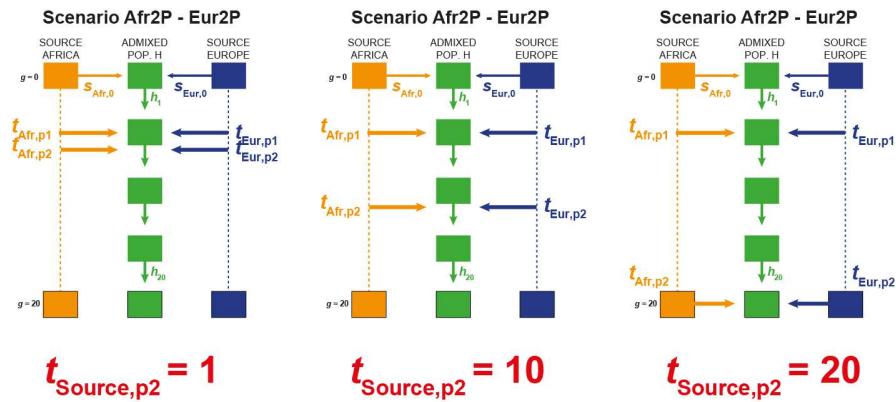
Results: MetHis-ABC model-choice with Random-Forest a priori without observed data

- 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 1000 simulations as pseudo-observed data under the AfrDE-EurDE model



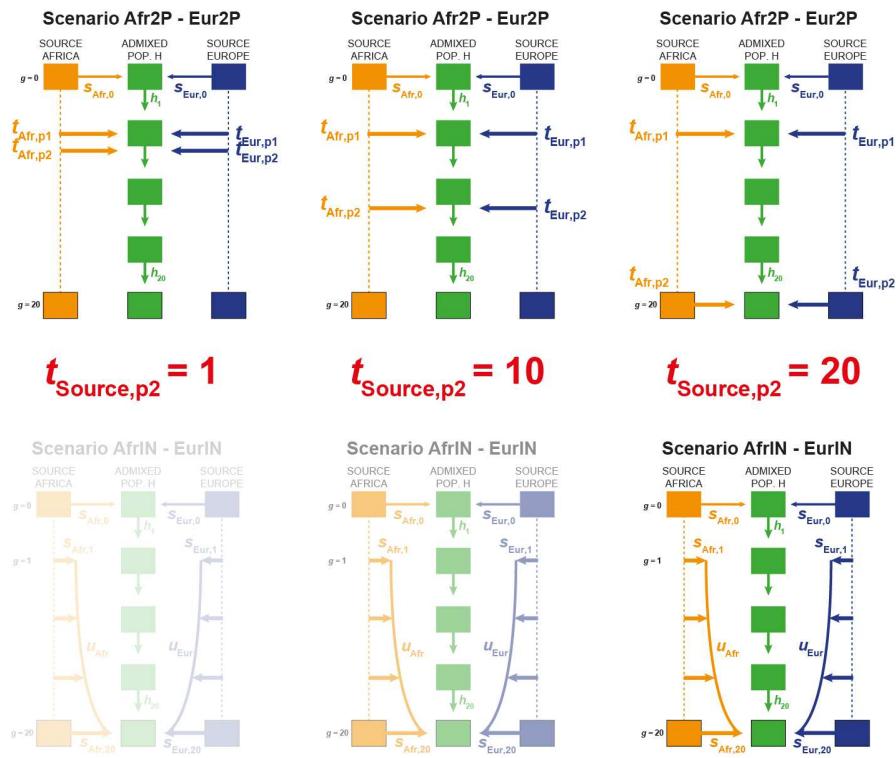
Results: *Methis*-ABC model-choice with Random-Forest a priori without observed data

- 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 1000 simulations as pseudo-observed data under the AfrDE-EurDE model



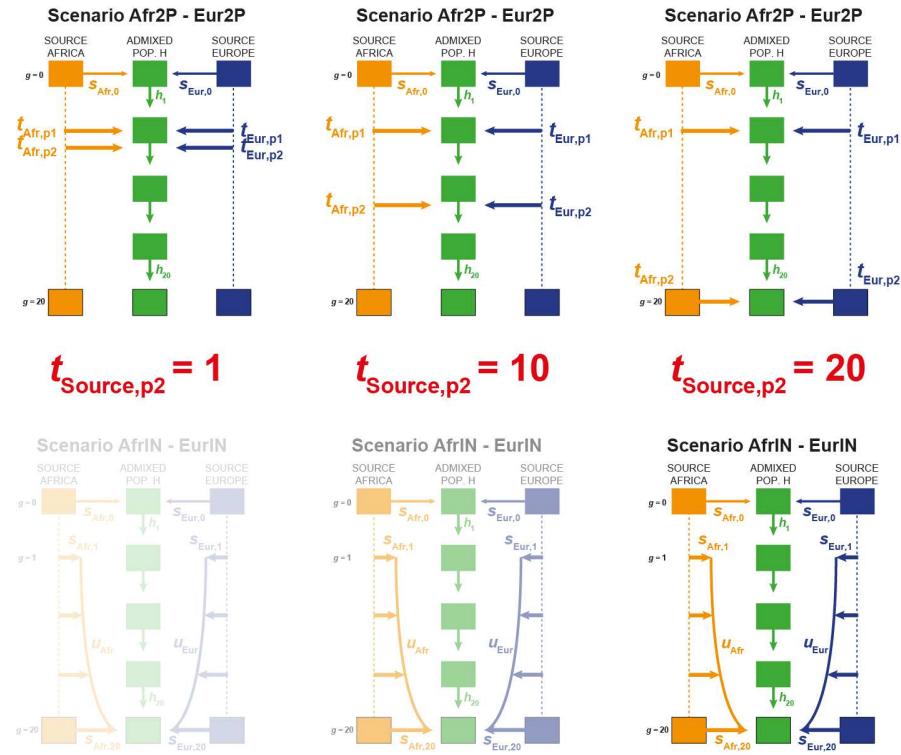
Results: MetHis-ABC model-choice with Random-Forest a priori without observed data

- 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 1000 simulations as pseudo-observed data under the AfrDE-EurDE model



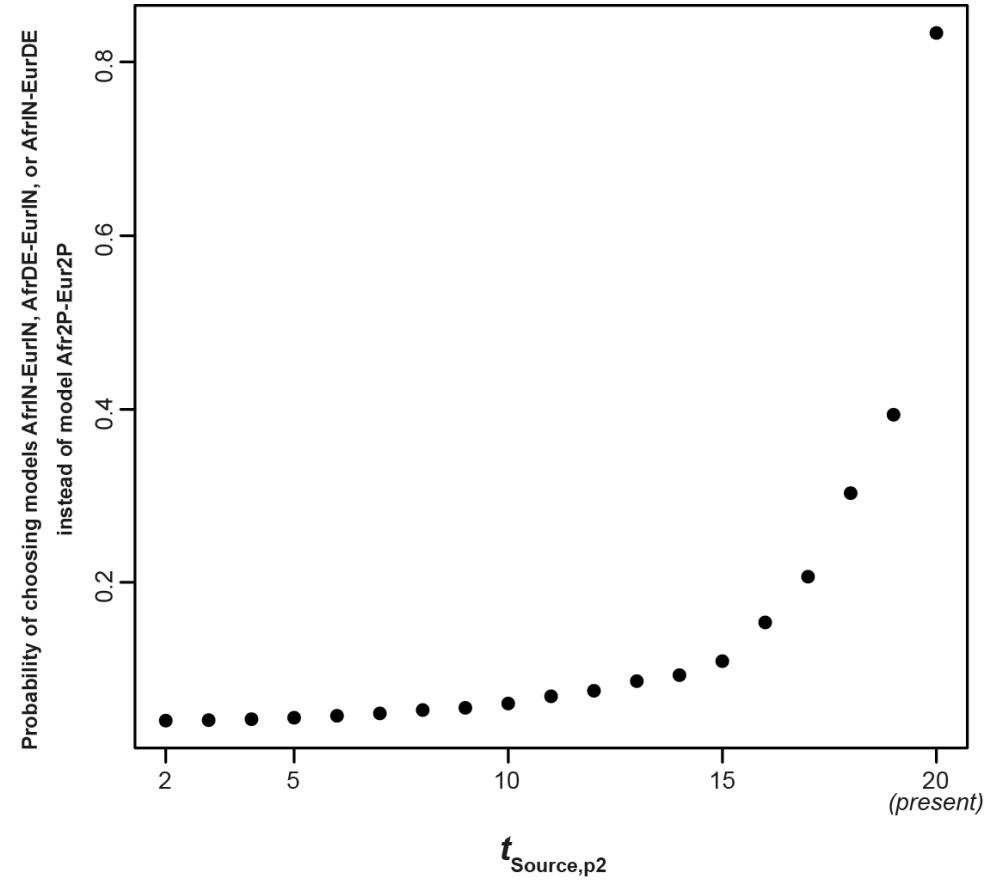
Results: MetHis-ABC model-choice with Random-Forest a priori without observed data

- 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 1000 simulations as pseudo-observed data under the AfrDE-EurDE model



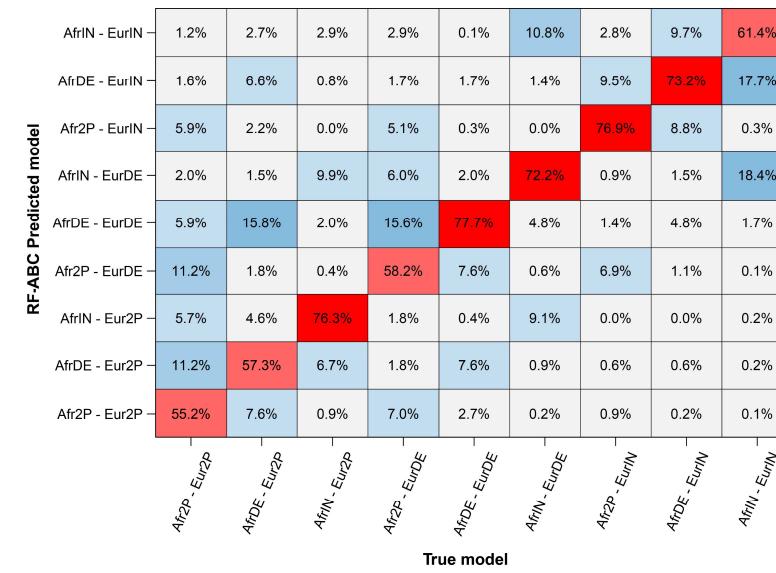
Model – Nestedness (Robert et al. 2011)

- Erroneous model-choice increases in the space of parameters where scenarios are highly nested and thus biologically similar



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



Prior error rate = 32.41%

Model-choice error a priori = 8/9 = 88.89%

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

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

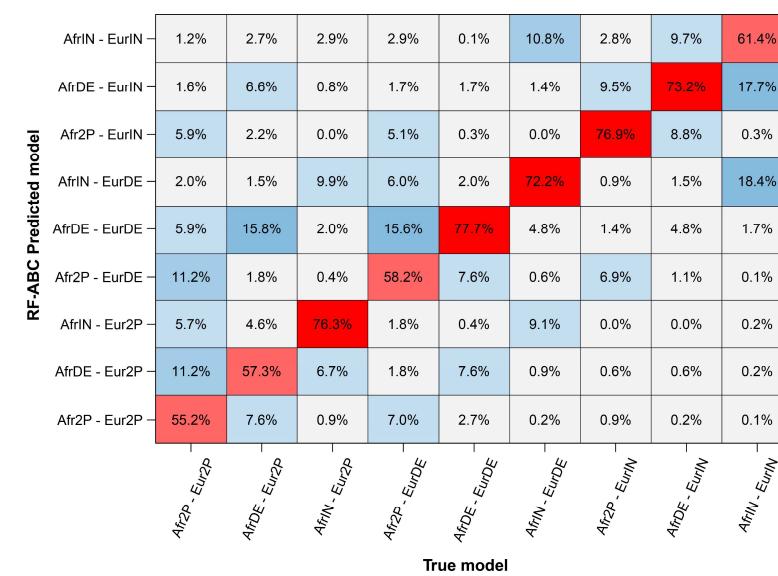
RF-ABC Predicted model	AfrIN - EurIN	AfrDE - EurIN	Afr2P - EurIN	AfrIN - EurDE	AfrDE - EurDE	Afr2P - EurDE	AfrIN - Eur2P	AfrDE - Eur2P	Afr2P - Eur2P
True model	1.4%	3.0%	2.9%	3.0%	0.1%	10.9%	3.0%	10.3%	62.9%
AfrIN - EurIN	2.0%	5.9%	0.8%	2.0%	2.4%	1.7%	10.2%	72.5%	17.3%
AfrDE - EurIN	6.2%	2.1%	0.0%	5.3%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

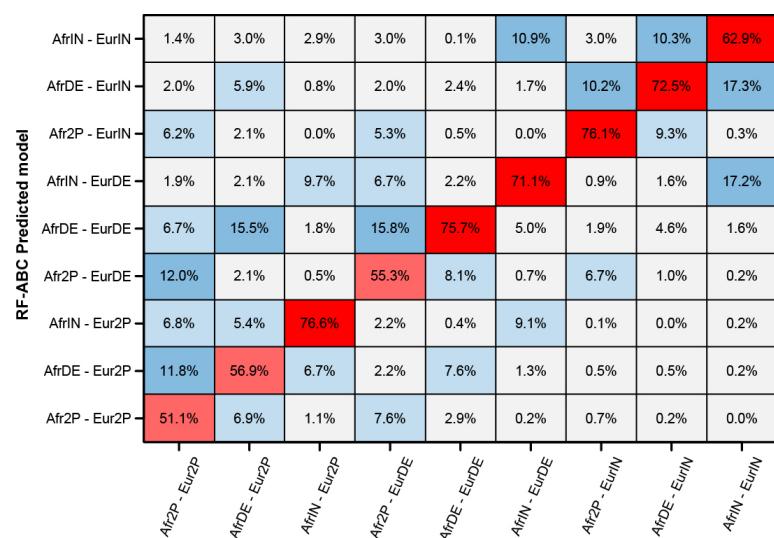
Results: MetHis-ABC model-choice with Random-Forest a priori without observed data

- 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
- 50,000 independent SNPs
- Sample sizes:
 - Afr Source : 90 indivs
 - Eur Source : 89 indivs
 - Admixed population H : 50 indivs
- 10,000 independent SNPs
- Sample sizes:
 - Afr Source : 90 indivs
 - Eur Source : 89 indivs
 - Admixed population H : 50 indivs



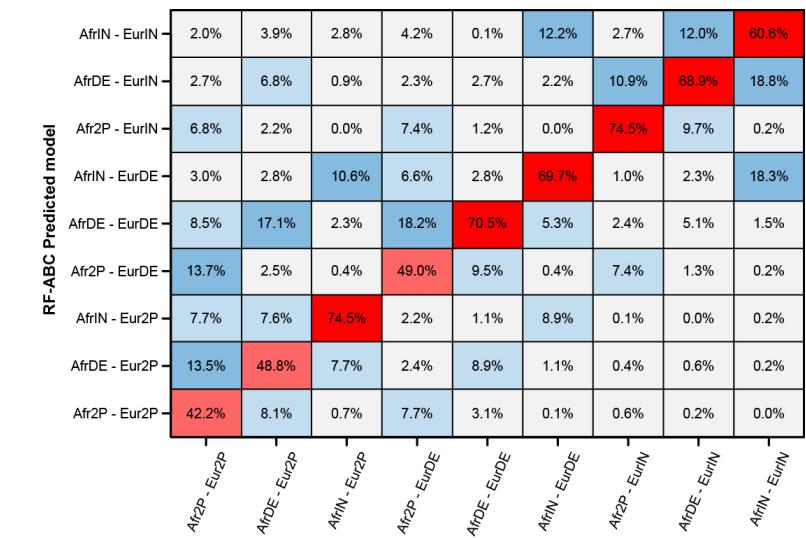
Prior error rate = 32.41%

Model-choice error a priori = 8/9 = 88.89%



Prior error rate = 33.53%

Model-choice error a priori = 8/9 = 88.89%



Prior error rate = 37.93%

Model-choice error a priori = 8/9 = 88.89%

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

- 100,000 independent SNPs
- Sample sizes:

Afr Source : 18 indivs

Eur Source : 18 indivs

Admixed population H : 10 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%

RF-ABC Predicted model		Afr2P - Eur2P	AfrDE - Eur2P	AfrIN - Eur2P	Afr2P - EurDE	AfrDE - EurDE	AfrIN - EurDE	Afr2P - EurIN	AfrDE - EurIN	AfrIN - EurIN
AfrIN - EurIN	1.6%	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

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

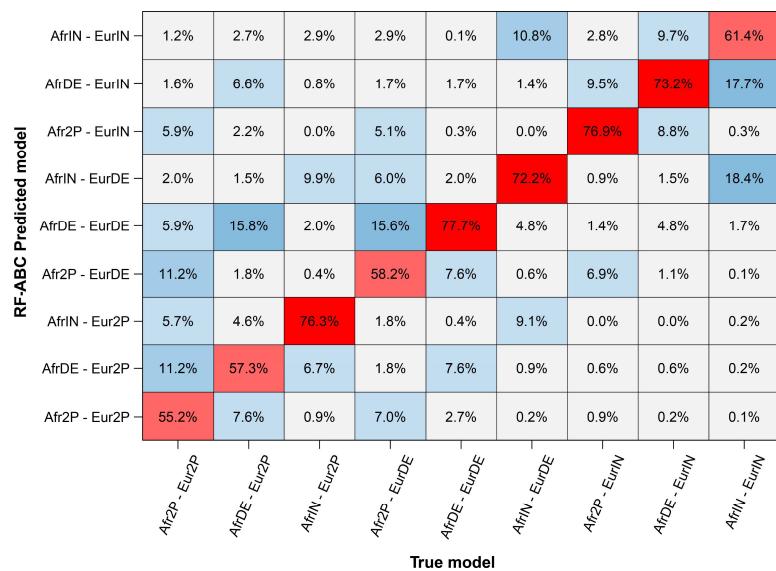
- 100,000 independent SNPs

- Sample sizes:

Afr Source : 90 indivs

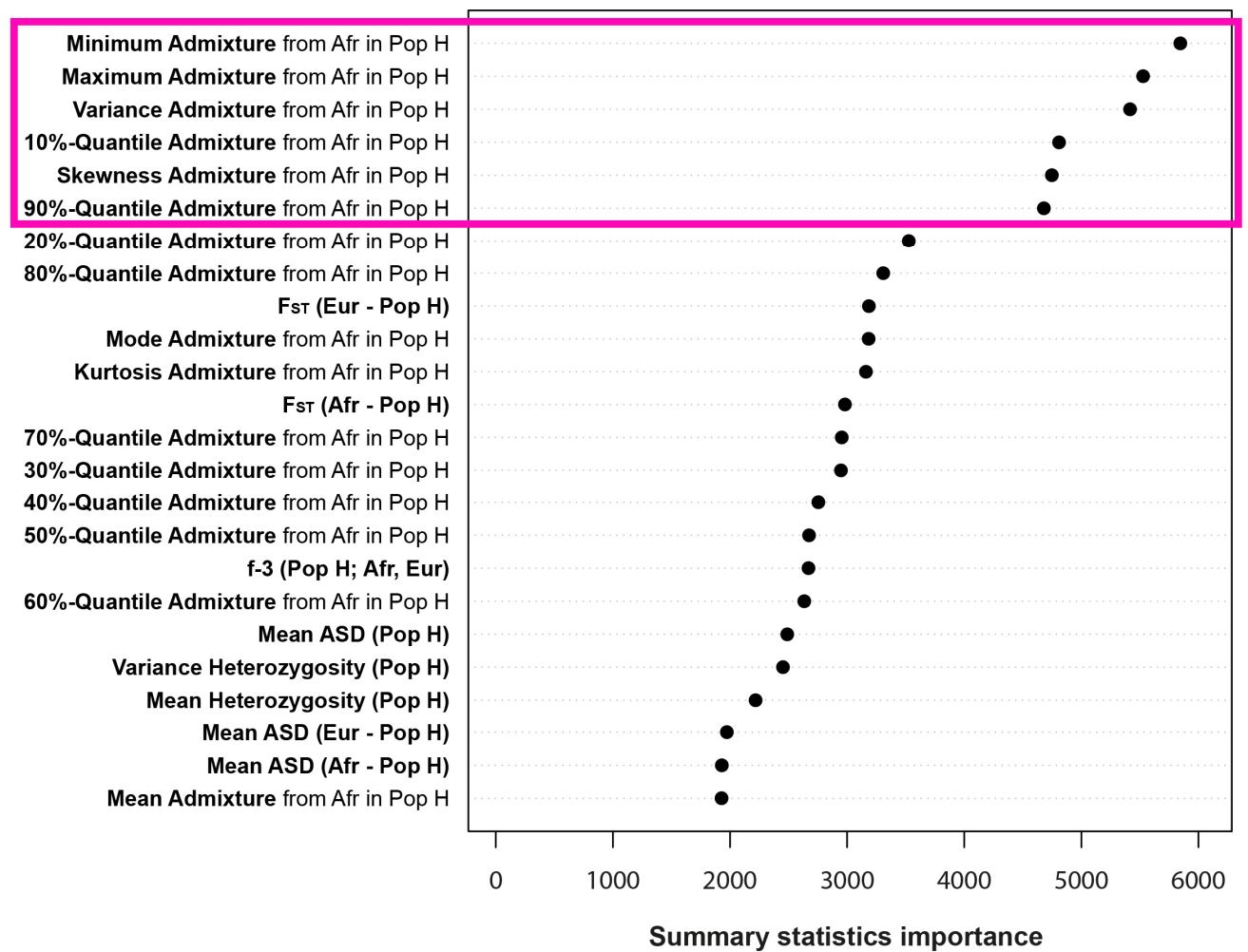
Eur Source : 89 indivs

Admixed population H : 50 indivs

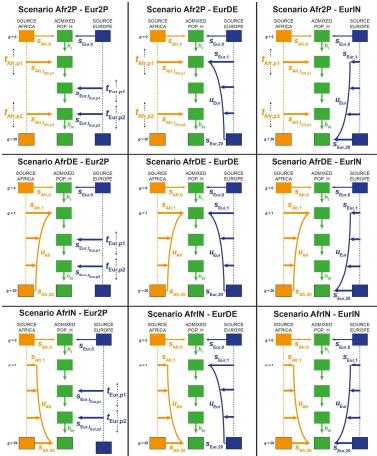


Prior error rate = 32.41%

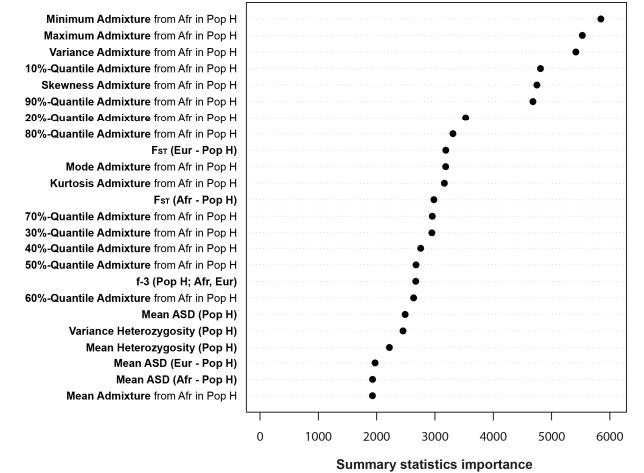
Model-choice error a priori = 8/9 = 88.89%



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

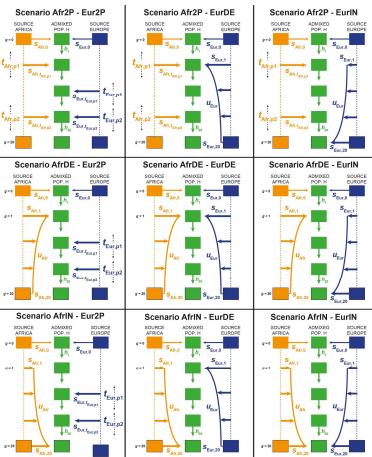


RF-ABC Predicted model	AfIN - EurIN	AfDE - EurIN	Af2P - EurIN	AfIN - EurDE	AfDE - EurDE	Af2P - EurDE	AfIN - Eur2P	AfDE - Eur2P	Af2P - Eur2P
AfIN - EurIN	1.2%	2.7%	2.9%	2.9%	0.1%	10.8%	2.8%	9.7%	61.4%
AfDE - EurIN	1.6%	6.6%	0.8%	1.7%	1.7%	1.4%	9.5%	73.2%	17.7%
Af2P - EurIN	5.9%	2.2%	0.0%	5.1%	0.3%	0.0%	76.0%	8.8%	0.3%
AfIN - EurDE	2.0%	1.5%	9.9%	6.0%	2.0%	72.2%	0.9%	1.5%	18.4%
AfDE - EurDE	5.9%	15.8%	2.0%	15.6%	77.7%	4.8%	1.4%	4.8%	1.7%
Af2P - EurDE	11.2%	1.8%	0.4%	58.2%	7.6%	0.6%	6.9%	1.1%	0.1%
AfIN - Eur2P	5.7%	4.6%	76.3%	1.8%	0.4%	9.1%	0.0%	0.0%	0.2%
AfDE - Eur2P	11.2%	57.3%	6.7%	1.8%	7.6%	0.9%	0.6%	0.6%	0.2%
Af2P - Eur2P	55.2%	7.6%	0.9%	7.0%	2.7%	0.2%	0.9%	0.2%	0.1%

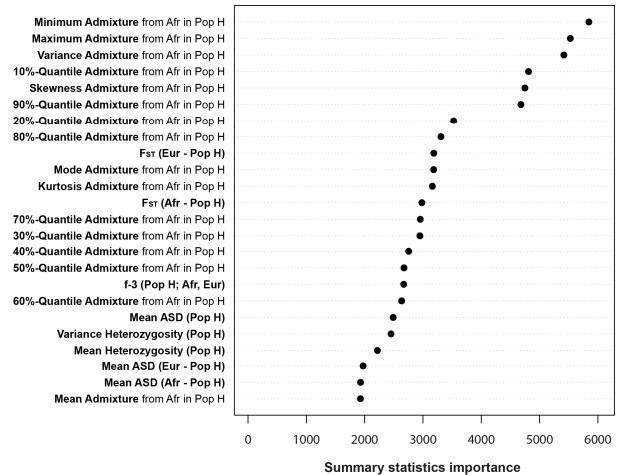


- MetHis-ABC Random-Forest model-choice a priori powerful to distinguish highly complex historical admixture models**

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

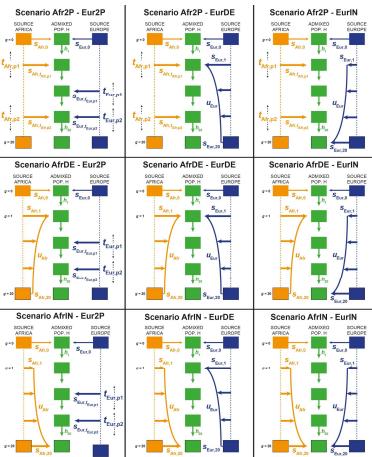


	AfriN - EurIN	AfriDE - EurIN	Afr2P - EurIN	AfrIN - EurDE	AfrDE - EurDE	Afr2P - EurDE	AfriN - Eur2P	AfrDE - Eur2P	Afr2P - Eur2P
RF-ABC Predicted model	1.2%	2.7%	2.9%	2.9%	0.1%	10.8%	2.8%	9.7%	61.4%
True model	AfriN - EurIN	AfriDE - EurIN	Afr2P - EurIN	AfrIN - EurDE	AfrDE - EurDE	Afr2P - EurDE	AfriN - Eur2P	AfrDE - Eur2P	Afr2P - Eur2P
AfriN - EurIN	1.2%	6.6%	0.8%	1.7%	1.7%	1.4%	9.5%	73.2%	17.7%
AfriDE - 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%	76.0%	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%



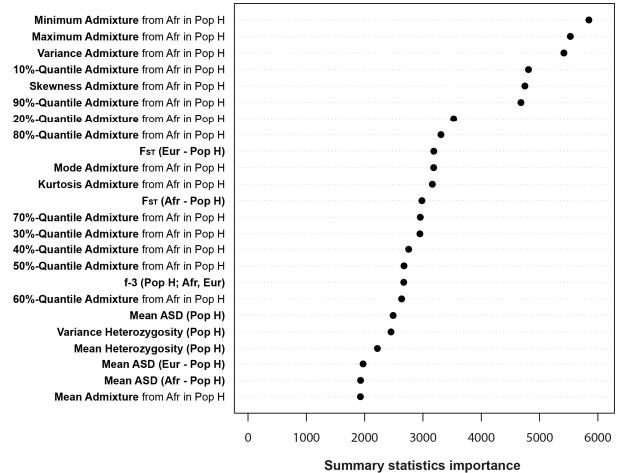
- **MetHis-ABC Random-Forest model-choice a priori powerful to distinguish highly complex historical admixture models**
- Errors are made in the parameter-space where models are highly nested and thus biologically similar

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



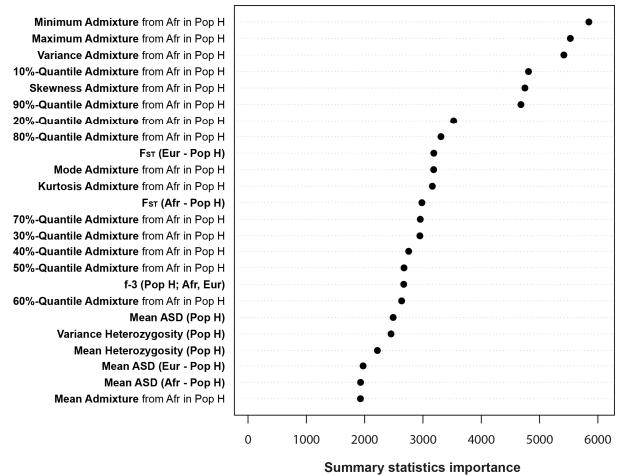
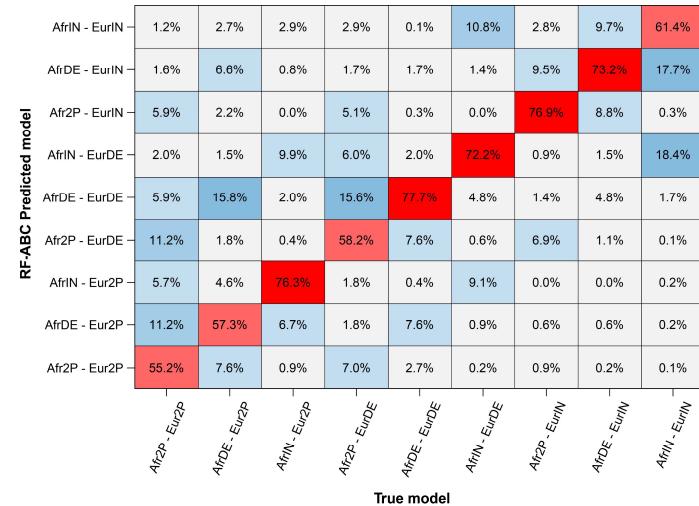
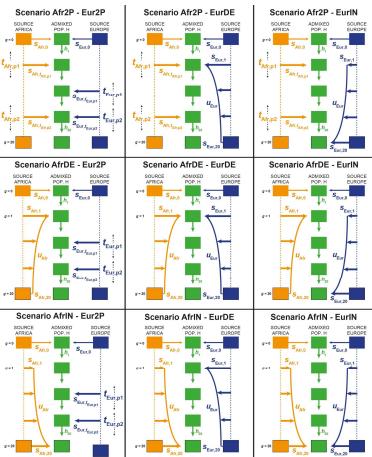
RF-ABC Predicted model	AfriN - EurIN	1.2%	2.7%	2.9%	2.9%	0.1%	10.8%	2.8%	9.7%	61.4%
	AfrDE - EurIN	1.6%	6.6%	0.8%	1.7%	1.7%	1.4%	9.5%	73.2%	17.7%
	Afr2P - EurIN	5.9%	2.2%	0.0%	5.1%	0.3%	0.0%	76.0%	8.8%	0.3%
	AfriN - EurDE	2.0%	1.5%	9.9%	6.0%	2.0%	72.2%	0.9%	1.5%	18.4%
	AfrDE - EurDE	5.9%	15.8%	2.0%	15.6%	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%

True model



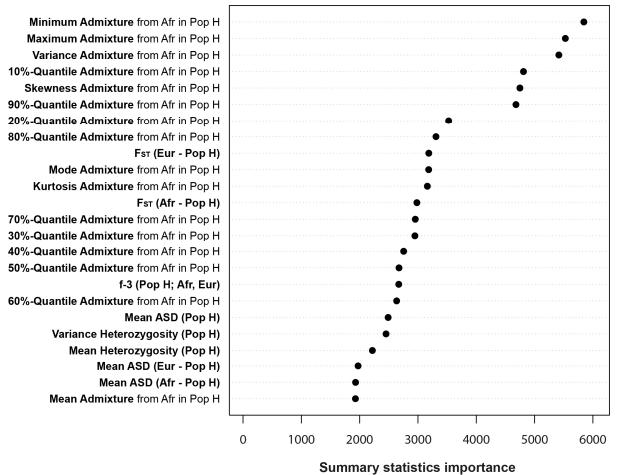
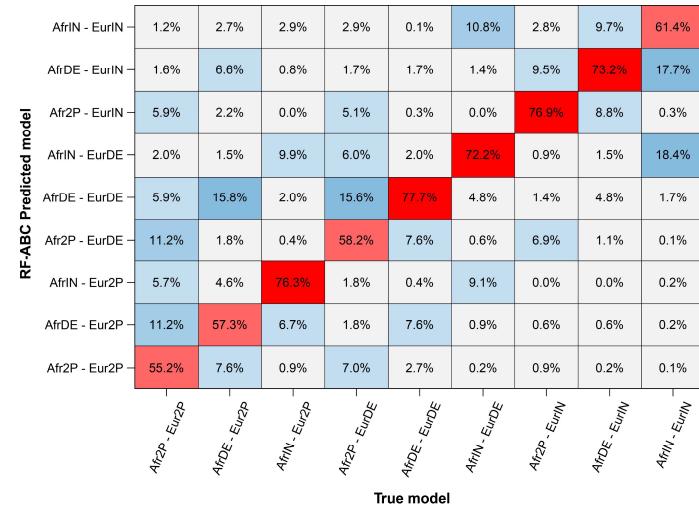
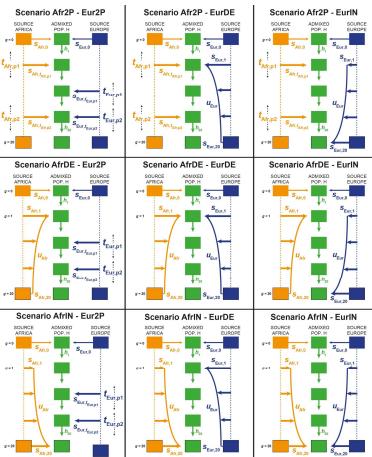
- *MetHis-ABC* Random-Forest model-choice *a priori* powerful to distinguish highly complex historical admixture models
- Errors are made in the parameter-space where models are highly nested and thus biologically similar
- *MetHis-ABC* Random-Forest model-choice performances are robust to reduced SNP and Sample sets

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



- **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

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

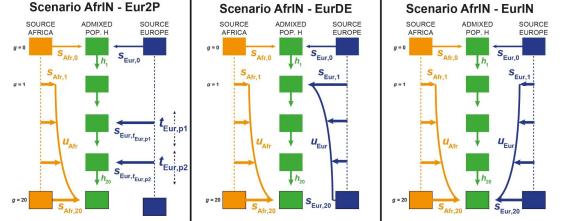
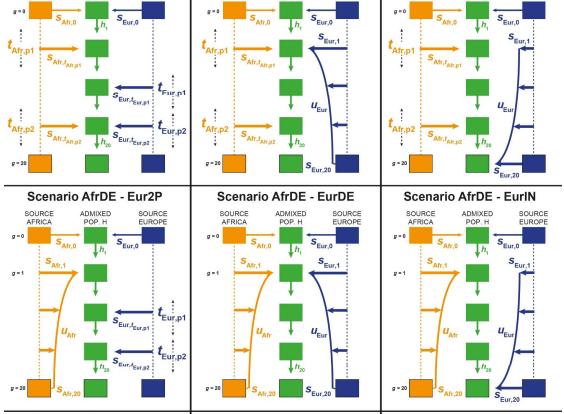
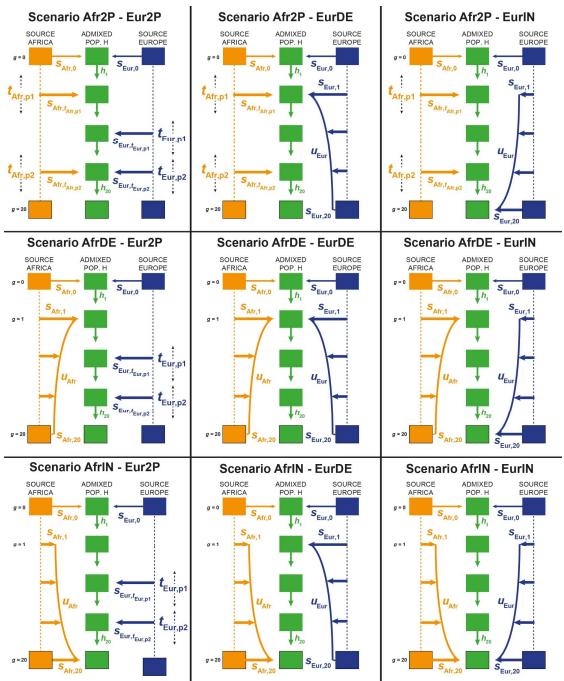


- **MetHis-ABC Random-Forest model-choice a priori powerful to distinguish highly complex historical admixture models**
- Errors are made in the parameter-space where models are highly nested and thus biologically similar
- **MetHis-ABC Random-Forest model-choice performances are robust to reduced SNP 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**

Results: *MetHis*-ABC complex admixture history of ACB and ASW populations

Simulating observed data with *MetHis* ?

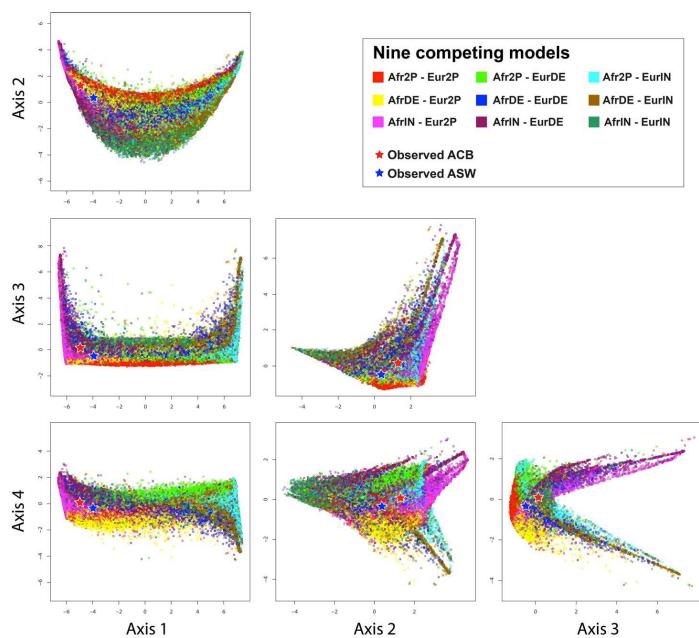
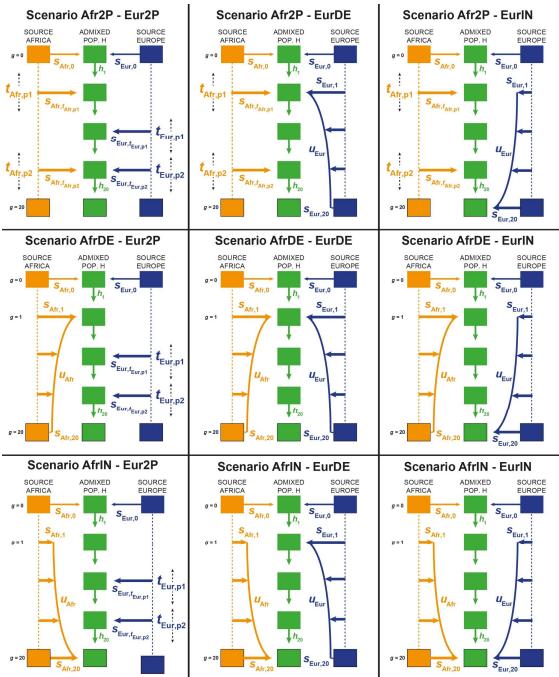
9 competing scenarios
 10,000 *MetHis* sims/scenario
 100,000 independent SNPs
 24 summary-statistics



Results: MetHis-ABC complex admixture history of ACB and ASW populations

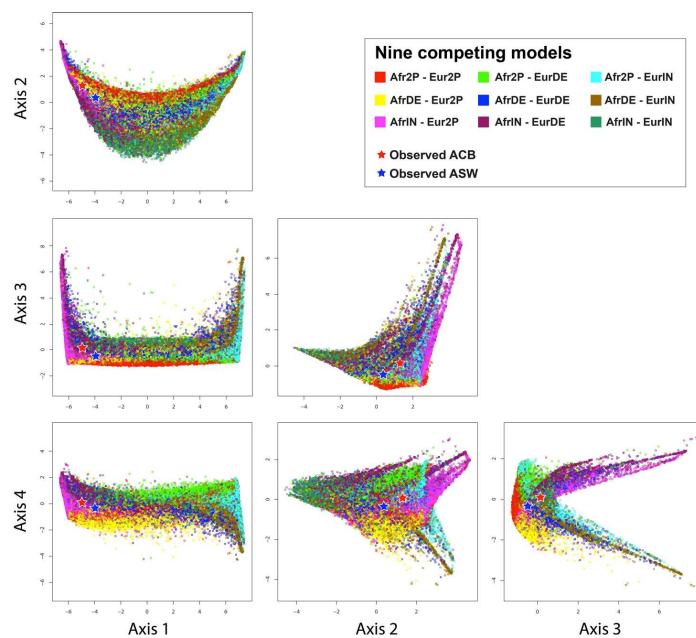
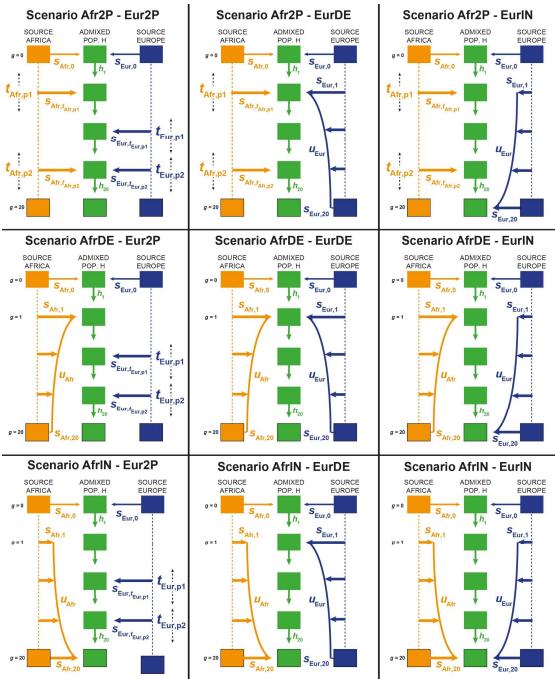
Simulating observed data with *MetHis* ?

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

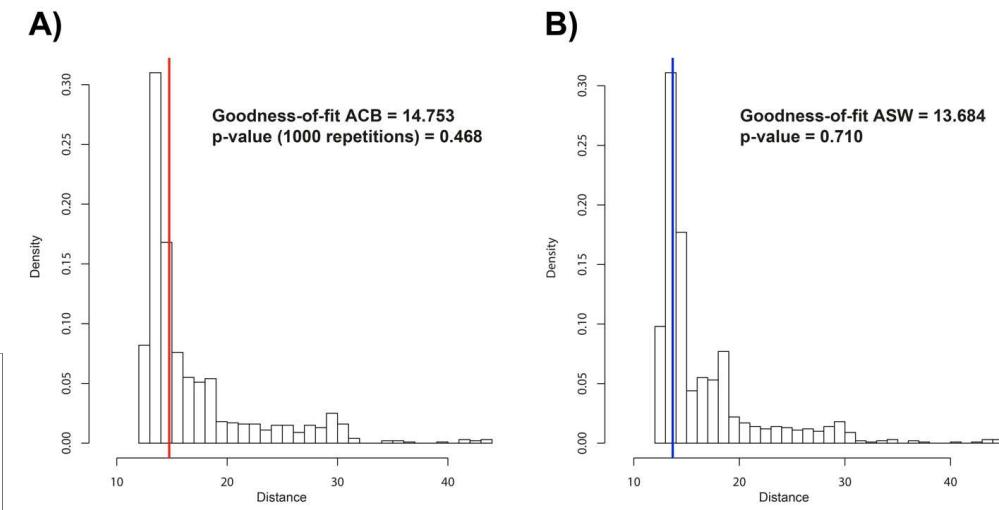


Results: *MetHis*-ABC complex admixture history of ACB and ASW populations

Simulating observed data with *MetHis* ?



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



Simulating admixed populations with *MetHis* provides summary statistics very reasonably fitting the observed data for ACB and ASW

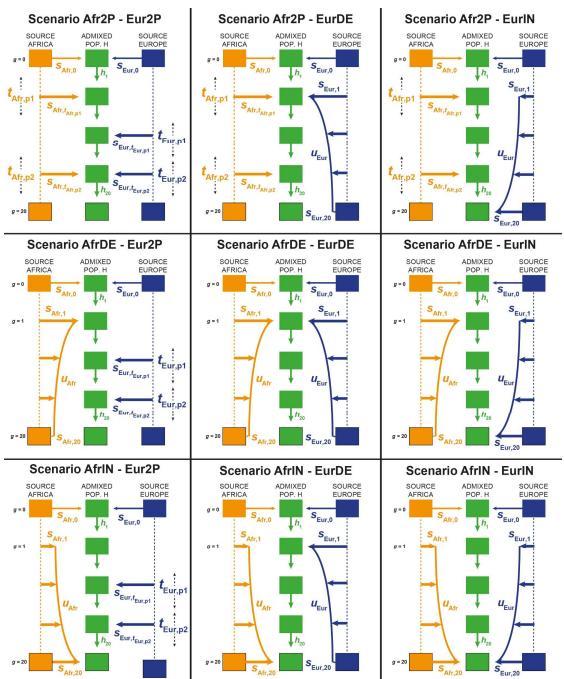
Results: MetHis-ABC complex admixture history of ACB and ASW populations

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

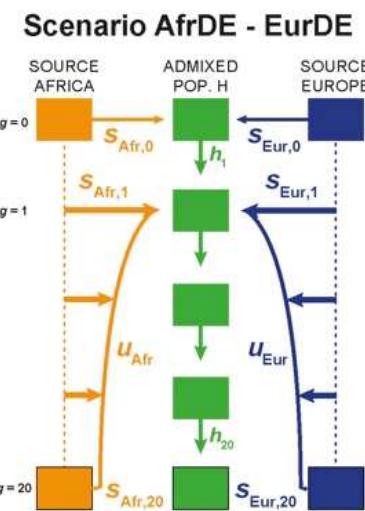
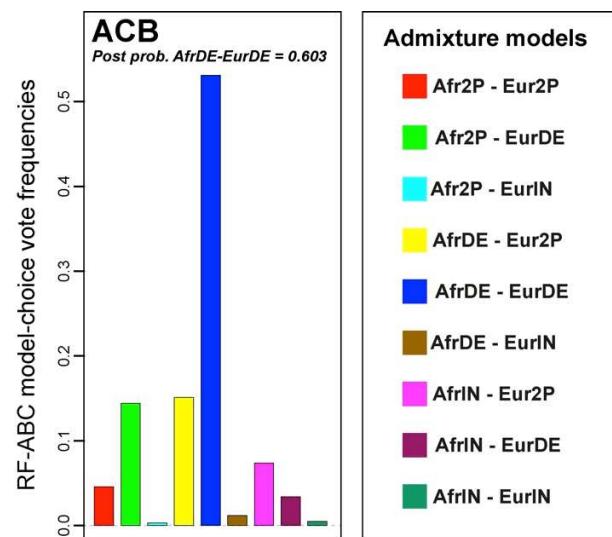
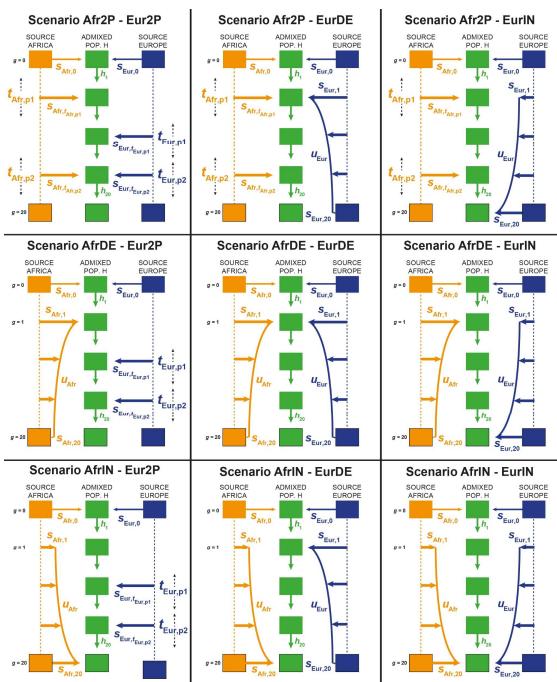
Results: MetHis-ABC complex admixture history of ACB and ASW populations

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

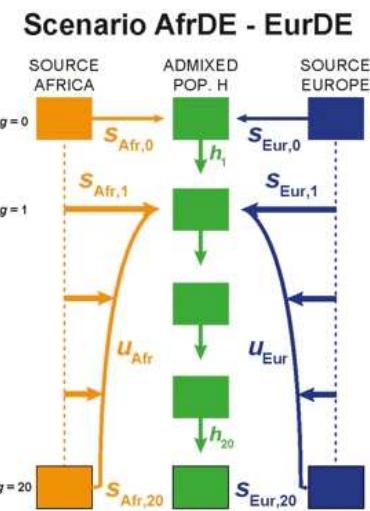
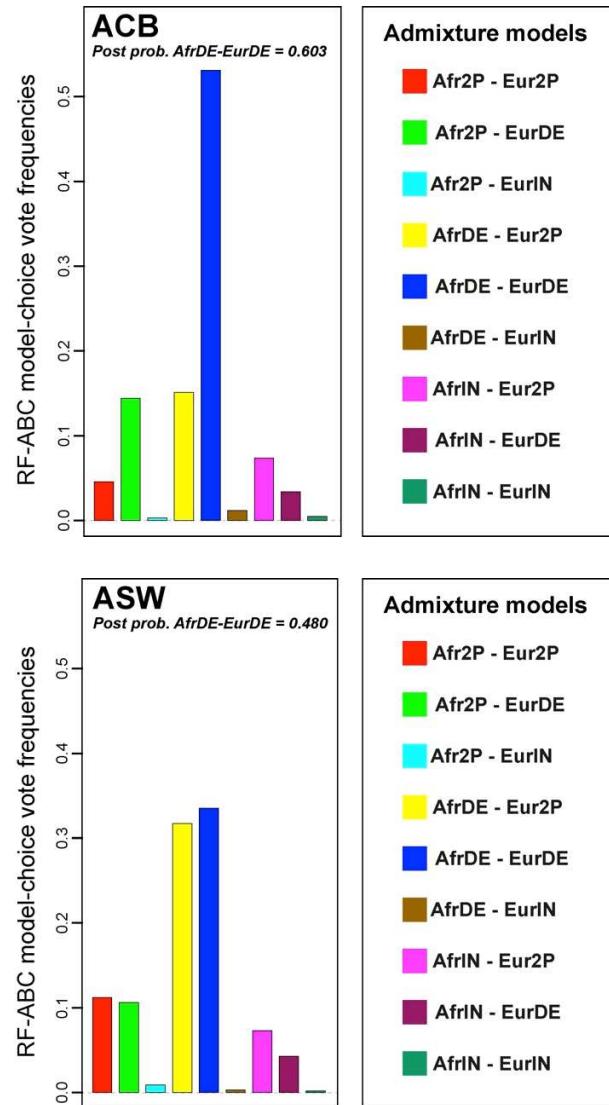
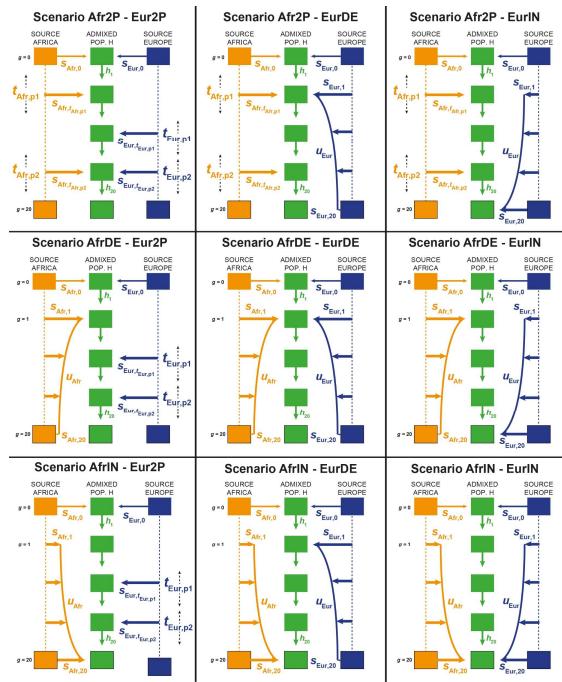
Results: MetHis-ABC complex admixture history of ACB and ASW populations

Model-choice

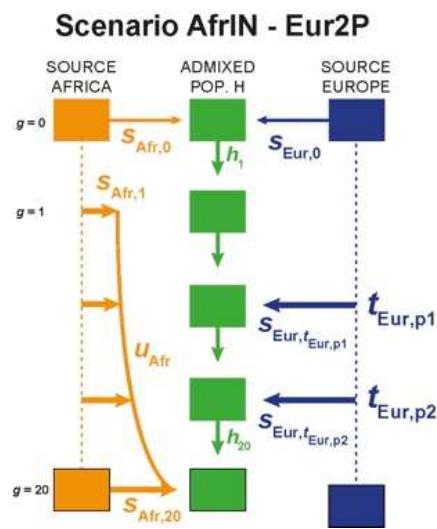
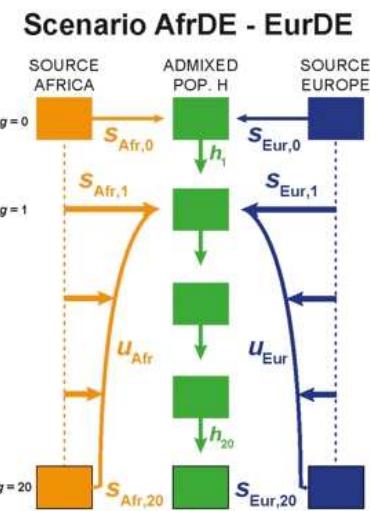
Random Forest ABC

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



Results: MetHis-ABC complex admixture history of ACB and ASW populations

Parameter-inference

Neural Network ABC

abc package in R (Csilléry et al. 2012)

Tolerance 1% (1,000 sims closest to obs.)

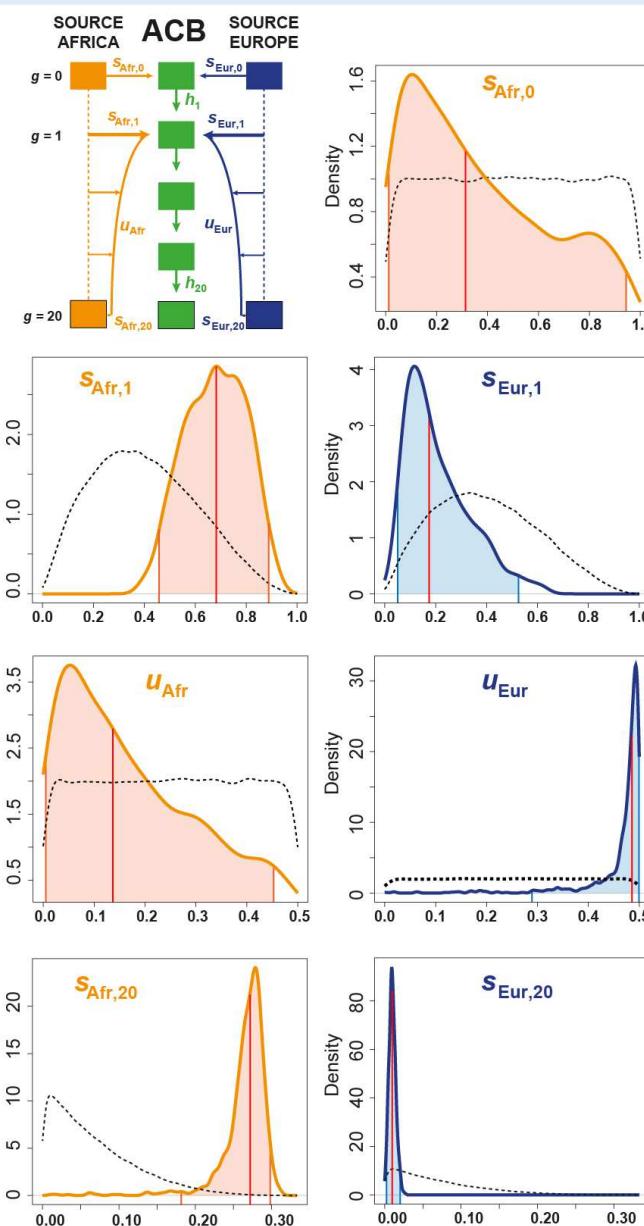
4 neurons in the hidden layer

1 best scenario

100,000 MetHis sims

100,000 independent SNPs

24 summary-statistics



Results: MetHis-ABC complex admixture history of ACB and ASW populations

Parameter-inference

Neural Network ABC

abc package in R (Csilléry et al. 2012)

Tolerance 1% (1,000 sims closest to obs.)

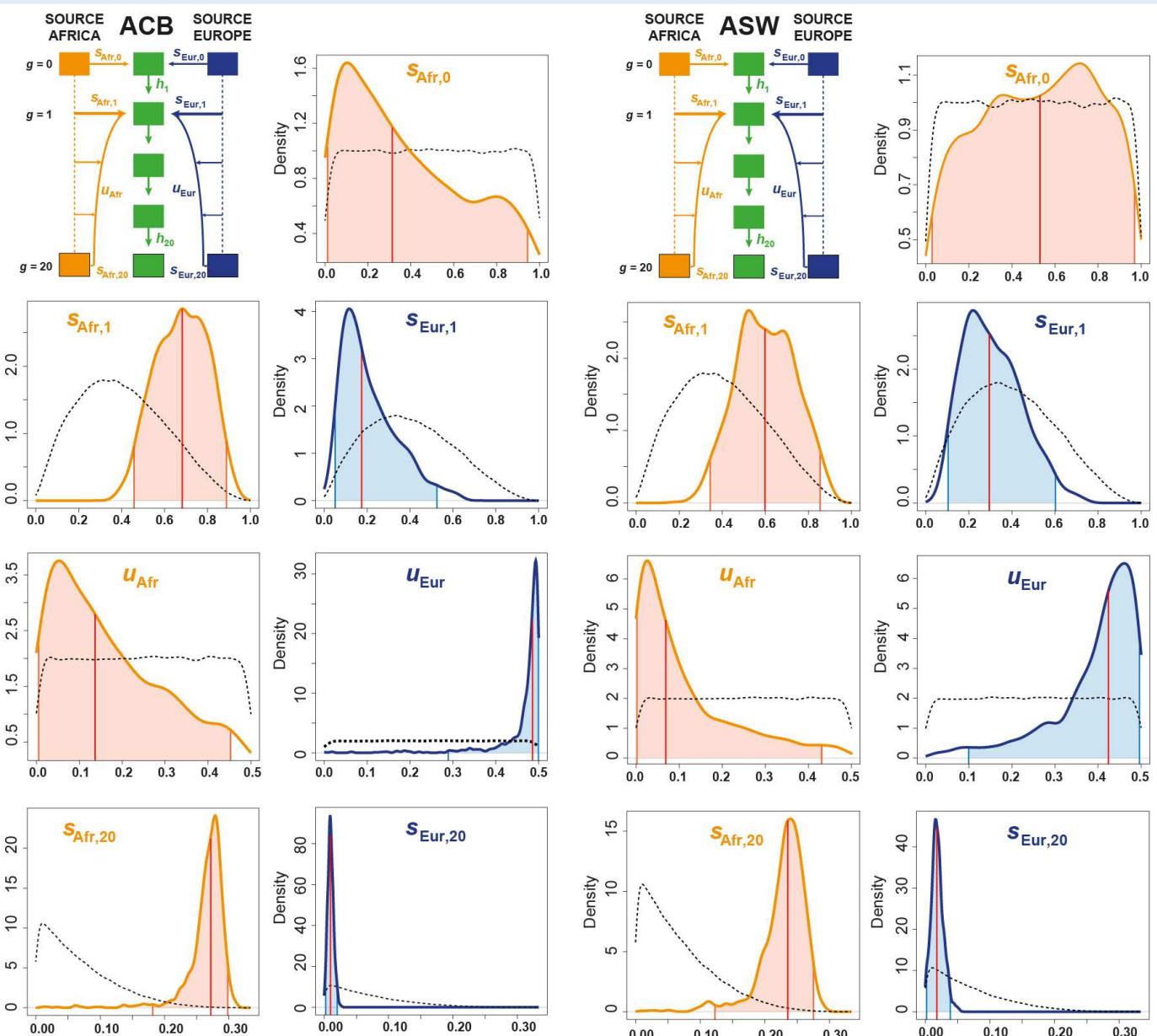
4 neurons in the hidden layer

1 best scenario

100,000 MetHis sims

100,000 independent SNPs

24 summary-statistics



Results: MetHis-ABC complex admixture history of ACB and ASW populations

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

1,000 cross-validation NN min av. posterior errors

1 best scenario

100,000 MetHis sims

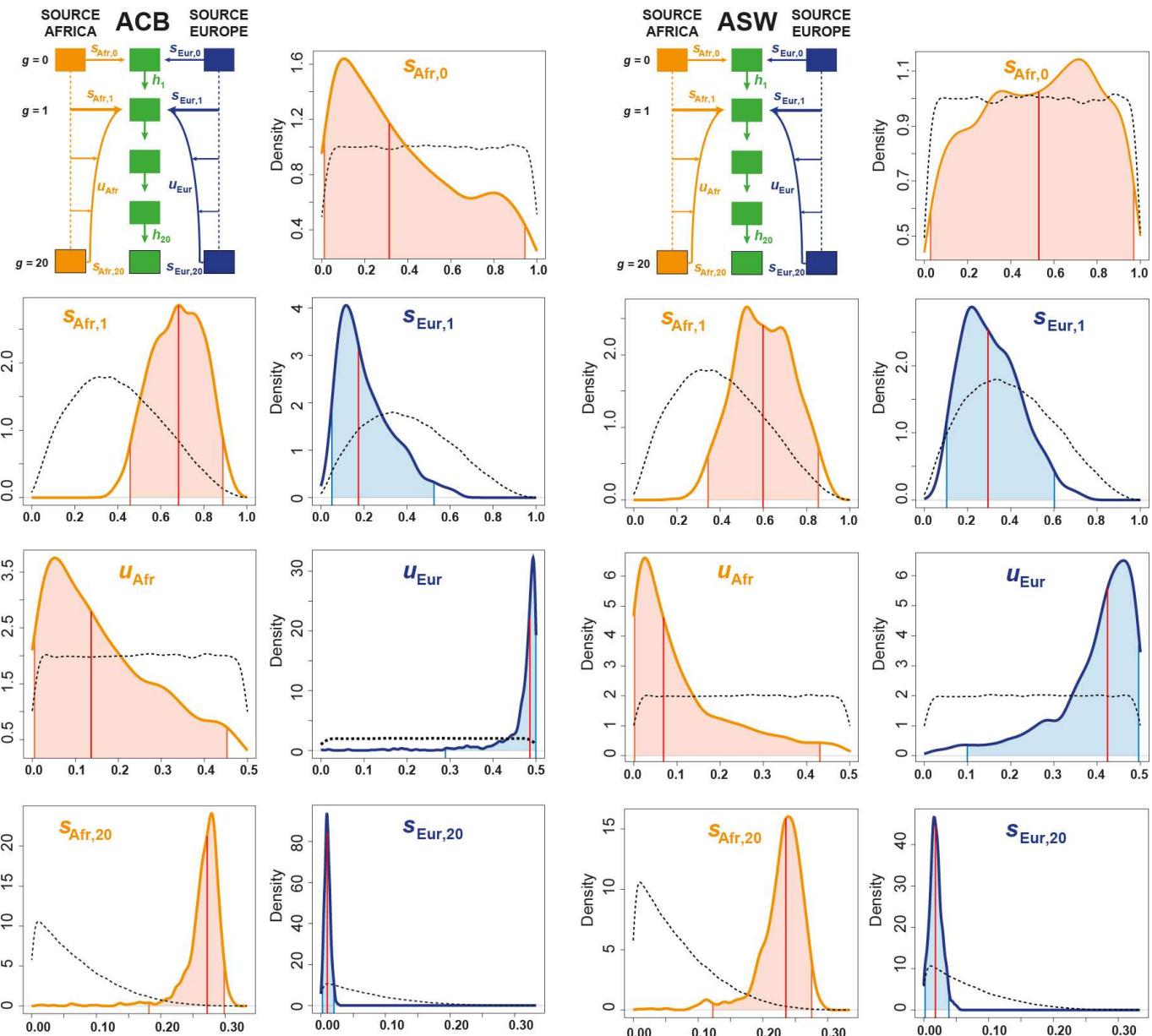
100,000 independent SNPs

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		ASW	
	Av. absolute Error	Mean-square Error / Var.	Av. absolute Error	Mean-square Error / Var.
$s_{Afr,0}$	0.2530	1.0070	0.2444	1.0081
$s_{Afr,1}$	0.1206	0.8533	0.1158	0.9259
$s_{Afr,20}$	0.0274	0.4162	0.0219	0.4773
u_{Afr}	0.1166	0.9974	0.1254	0.9757
$s_{Eur,1}$	0.0952	1.0526	0.1001	1.0152
$s_{Eur,20}$	0.0044	0.6452	0.0069	0.6623
u_{Eur}	0.1084	0.9431	0.1021	0.8036



Results: MetHis-ABC complex admixture history of ACB and ASW populations

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

1,000 cross-validation NN min av. posterior errors

1 best scenario

100,000 MetHis sims

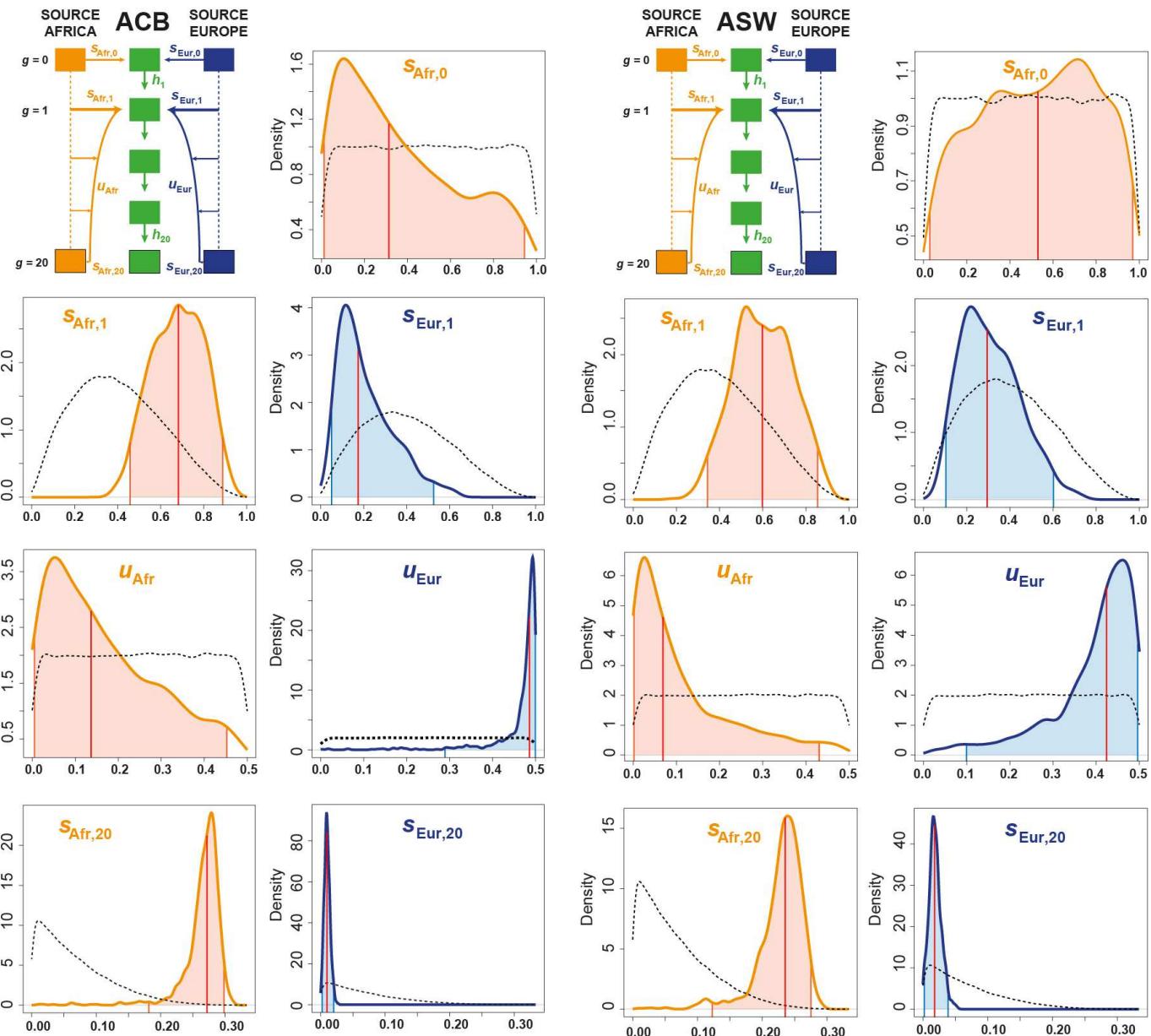
100,000 independent SNPs

24 summary-statistics

Cross-validation 95% CI accuracy

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

AfrDE-EurDE parameters	ACB				ASW			
	NN joint	NN indep.	RF indep.	Rejection indep.	NN joint	NN indep.	RF indep.	Rejection indep.
$S_{\text{Afr},0}$	0.956	0.934	0.929	0.952	0.952	0.931	0.937	0.950
$S_{\text{Afr},1}$	0.958	0.929	0.942	0.968	0.958	0.914	0.942	0.963
$S_{\text{Afr},20}$	0.964	0.926	0.956	0.971	0.963	0.928	0.960	0.978
u_{Afr}	0.953	0.932	0.930	0.950	0.944	0.914	0.925	0.945
$S_{\text{Eur},1}$	0.947	0.909	0.939	0.949	0.950	0.912	0.930	0.955
$S_{\text{Eur},20}$	0.944	0.908	0.930	0.957	0.952	0.919	0.929	0.968
u_{Eur}	0.941	0.919	0.927	0.943	0.947	0.928	0.936	0.952
Average credibility interval accuracy	0.951	0.922	0.936	0.955	0.952	0.920	0.937	0.958



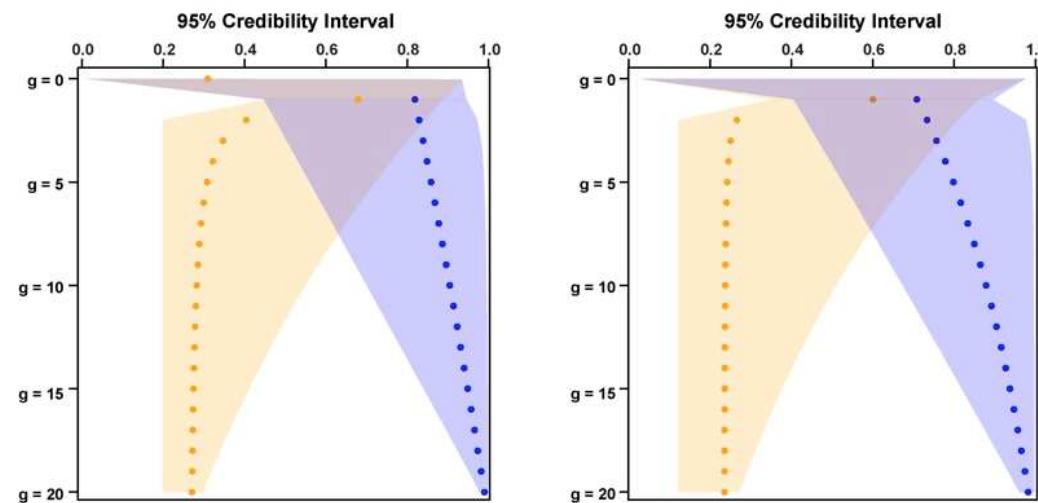
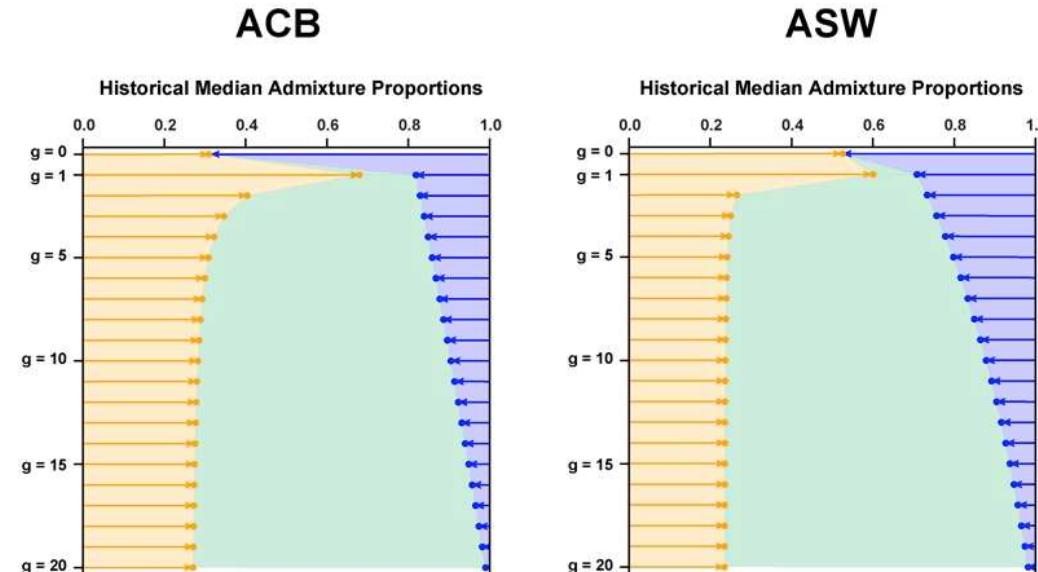
Results: MetHis-ABC complex admixture history of ACB and ASW populations

Rapid diminishing of European admixture

Less European admixture in ACB than in ASW

Strong continuing African admixture:

- post-slavery migrations
- assortative mating ?

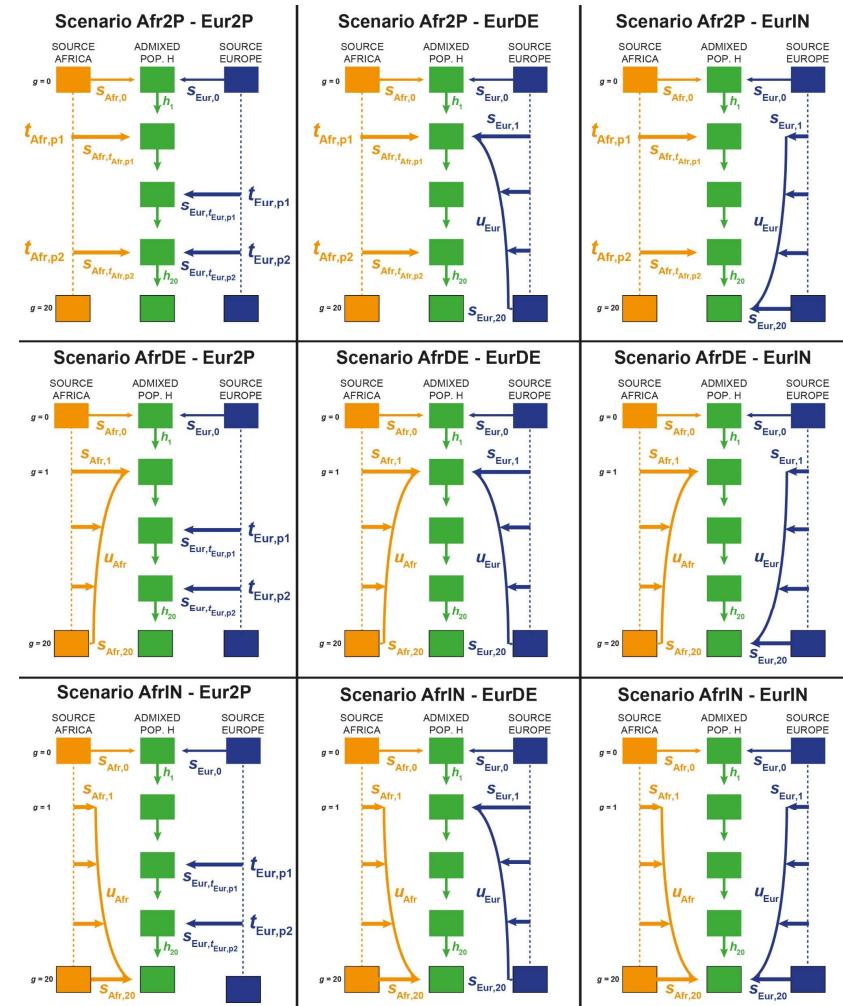
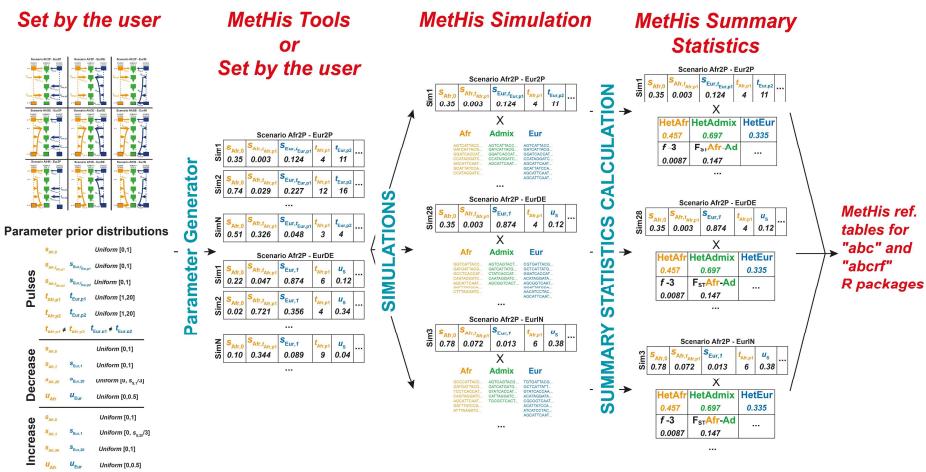


Conclusions

***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**



Conclusions

MOLECULAR ECOLOGY RESOURCES

RESOURCE ARTICLE |  Full Access

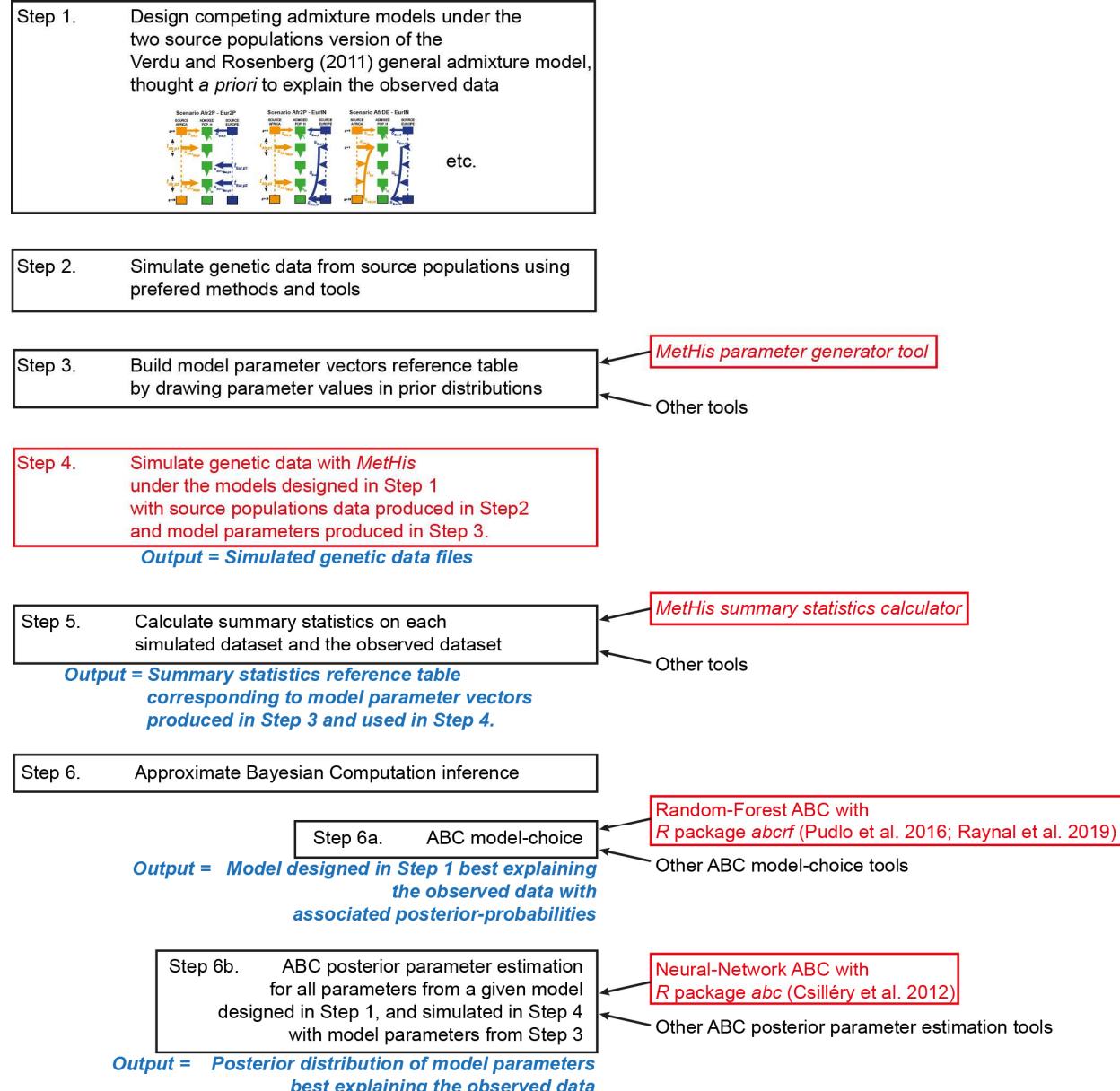
Complex genetic admixture histories reconstructed with Approximate Bayesian Computation

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>

Special issue *Machine Learning in Molecular Ecology*

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



Ongoing !

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



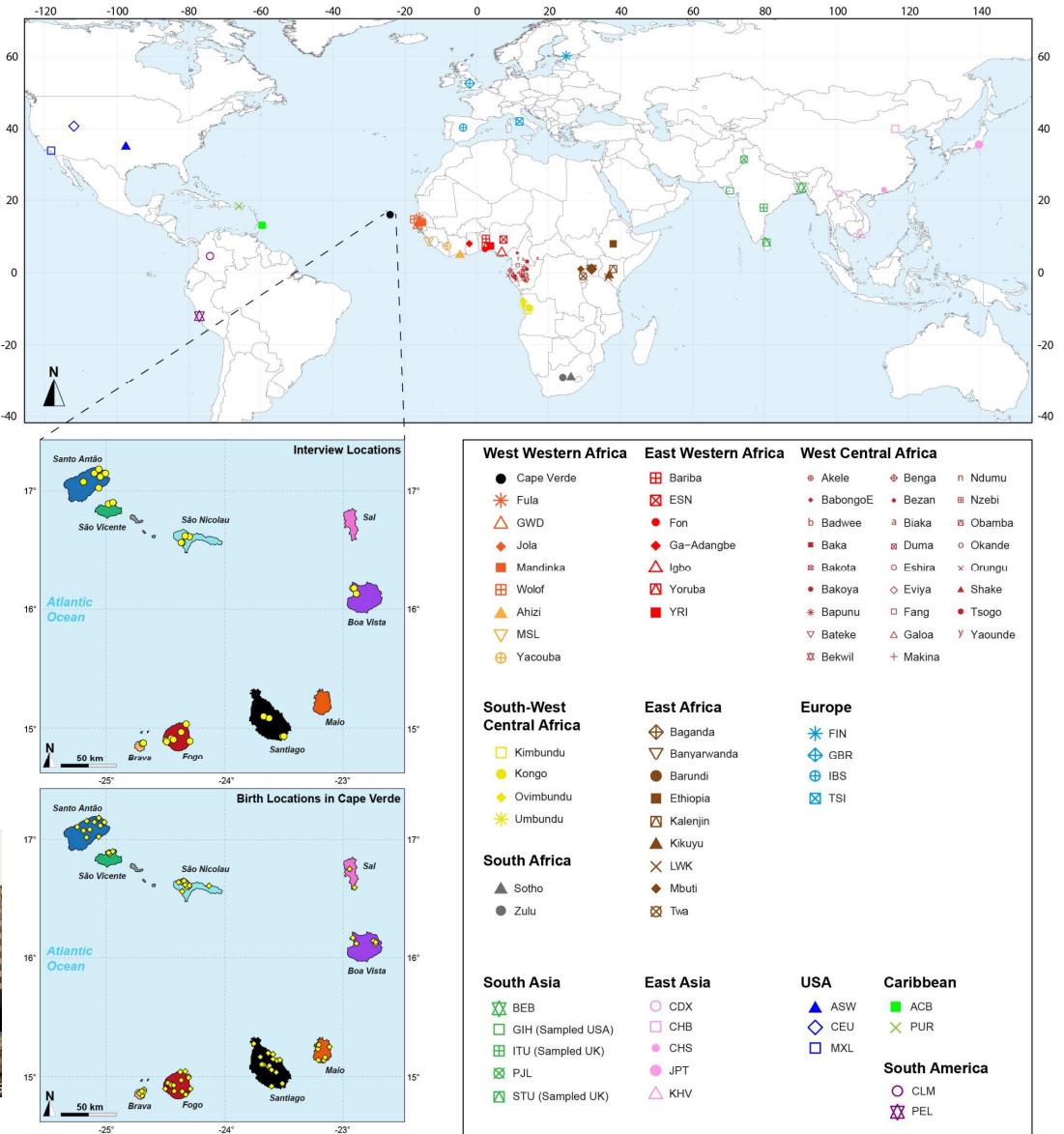
©Paul Verdu, Cape Verde 2010-2018



©Paul Verdu, Cape Verde 2010-2018

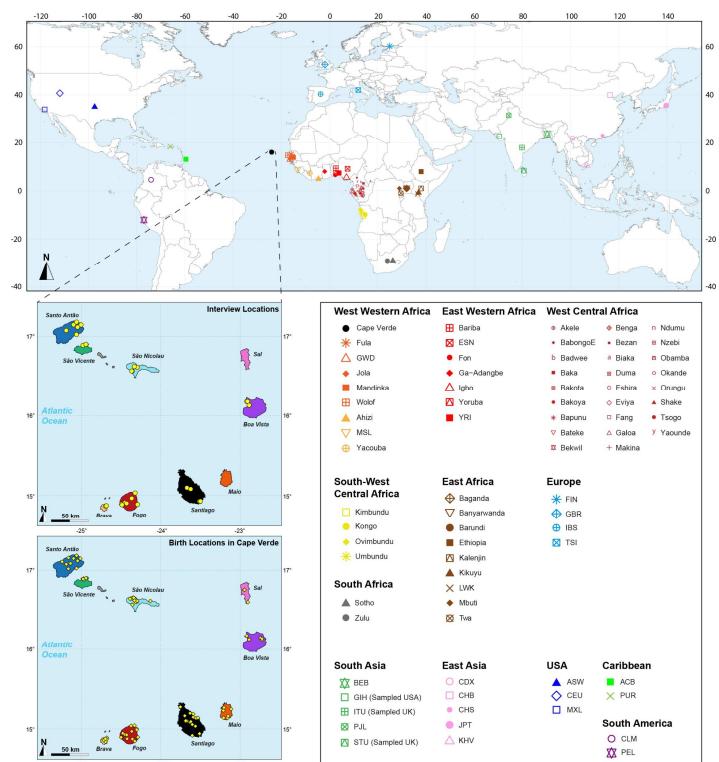
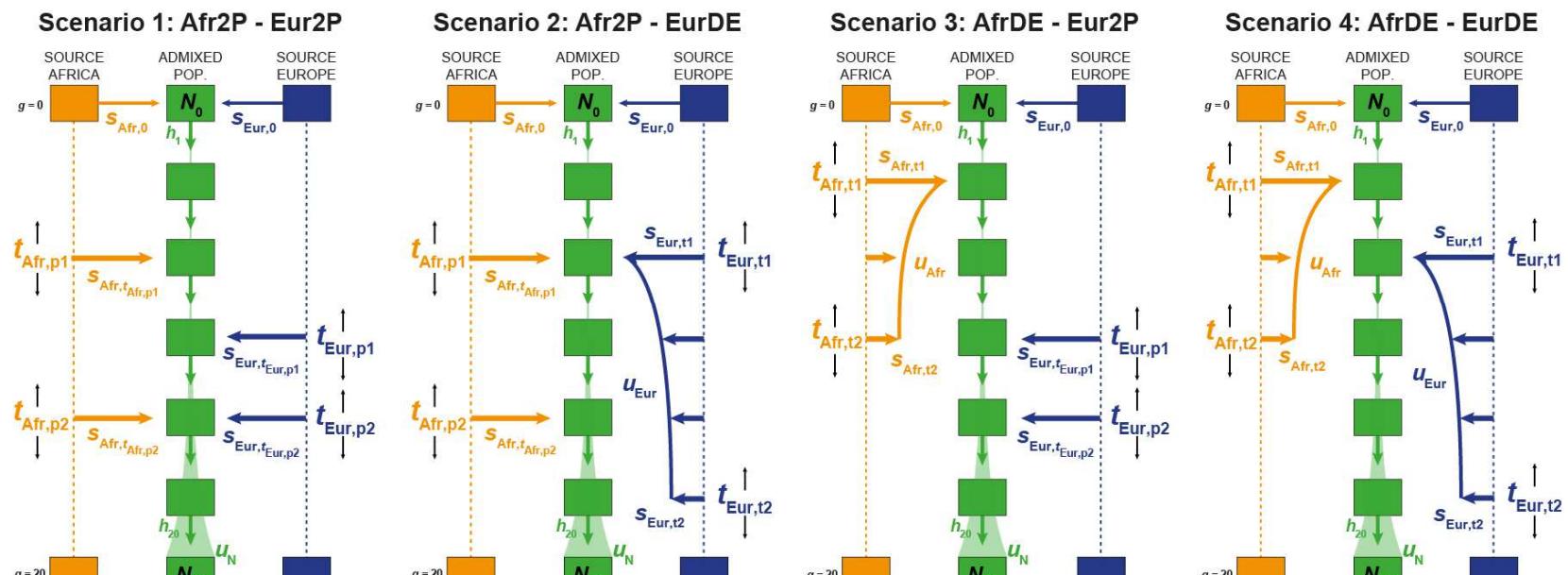


©Paul Verdu, Cape Verde 2010-2018



Ongoing !

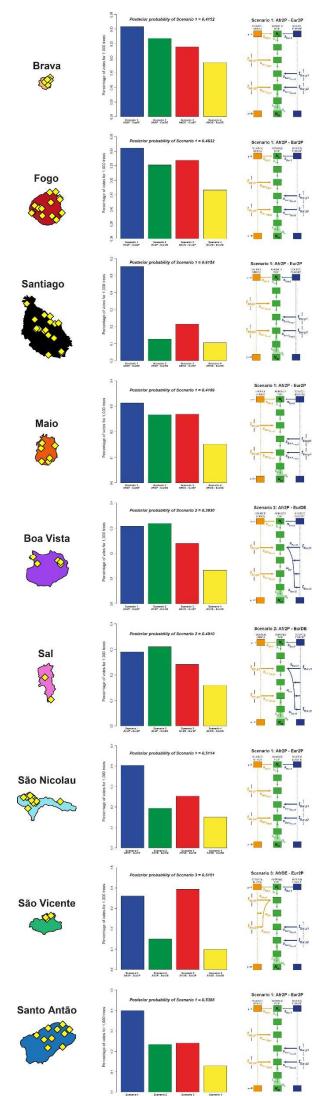
Reconstructing the admixture history of Cape Verde with Romain Laurent (UMR7206) Noah Rosenberg (Stanford University) and Marlyse Baptista (University of Michigan)



Ongoing !

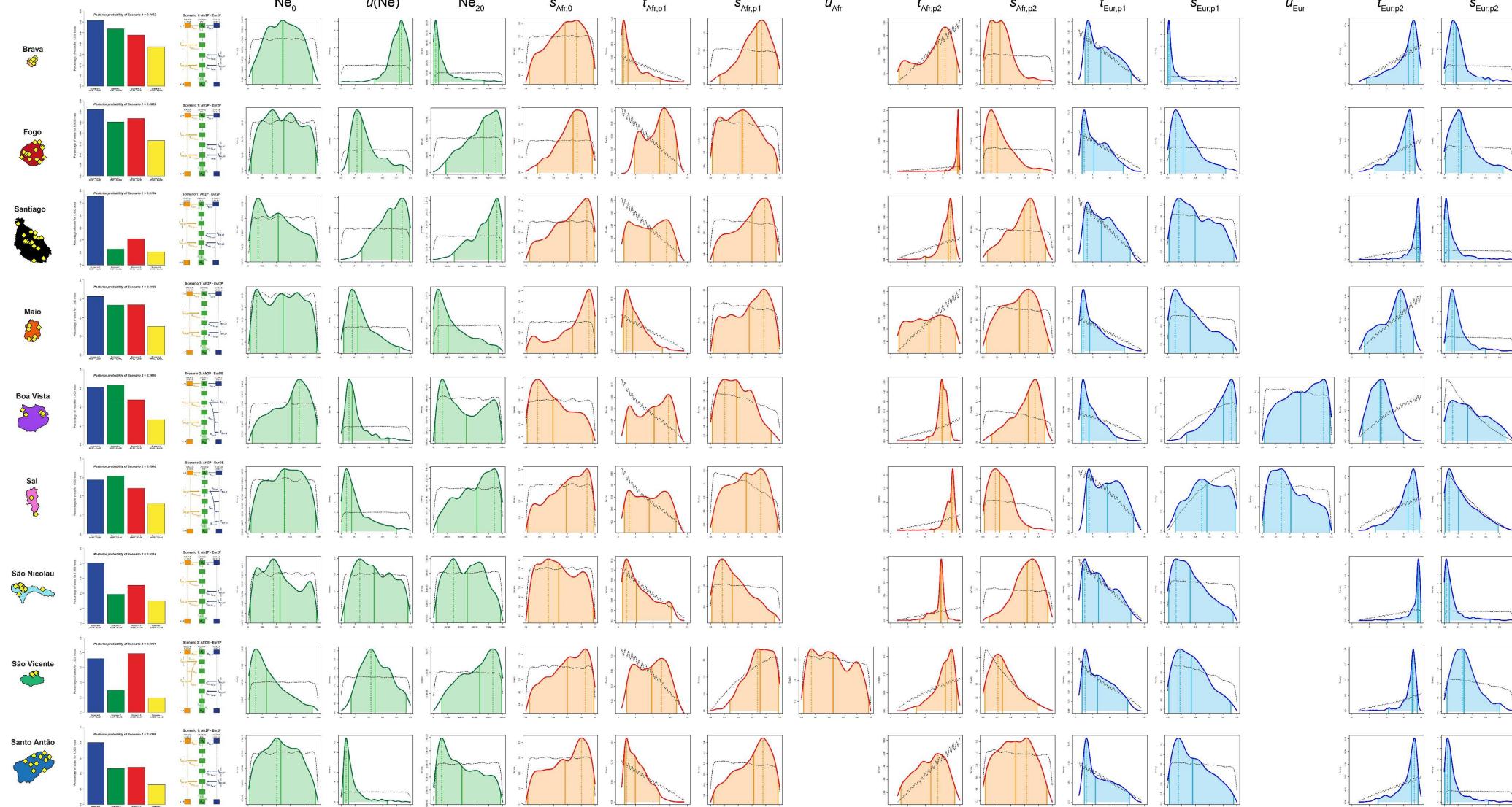
Reconstructing the admixture history of Cape Verde

with Romain Laurent (UMR7206), Noah Rosenberg (Stanford University) and Marlyse Baptista (University of Michigan)



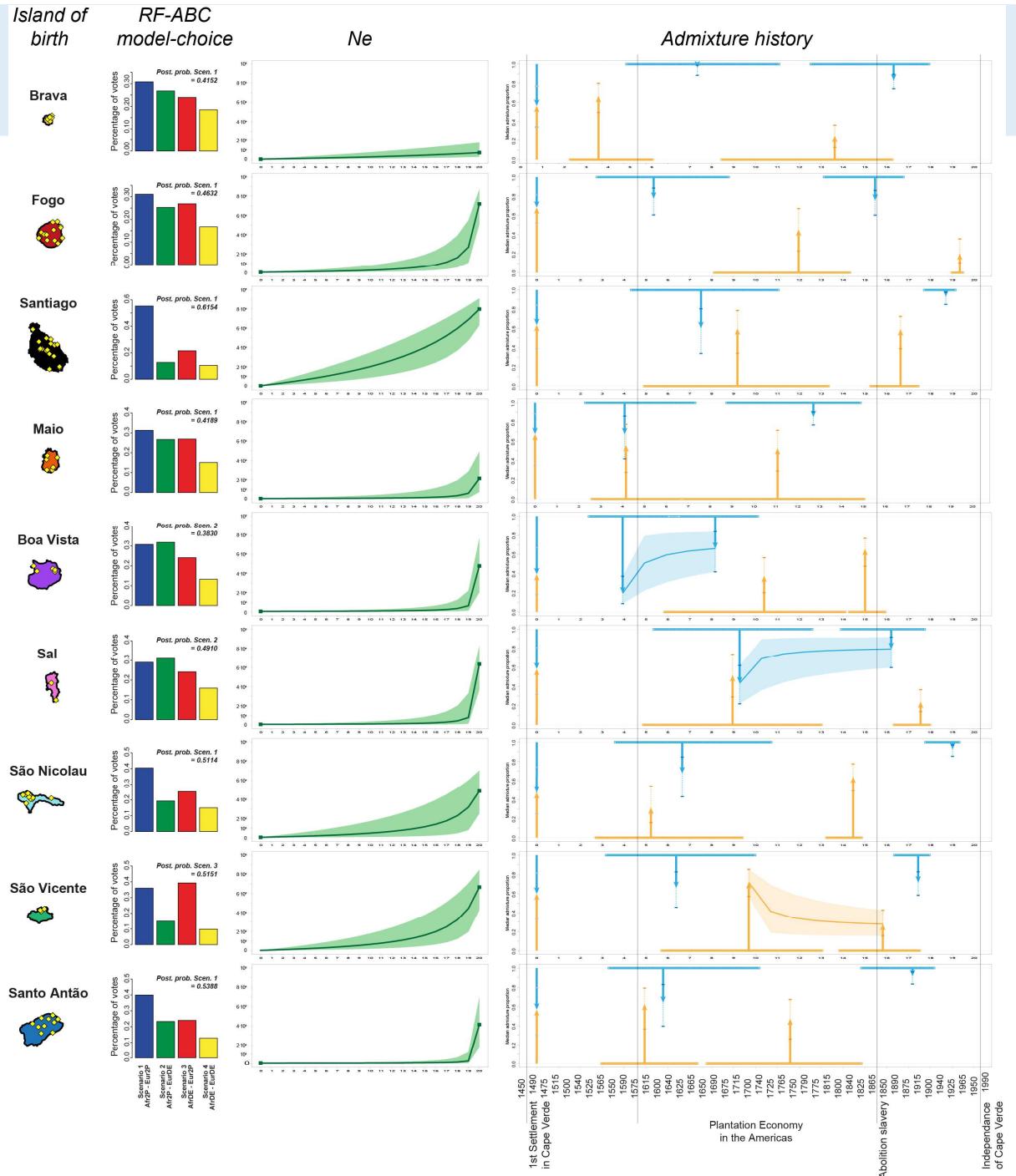
Ongoing !

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



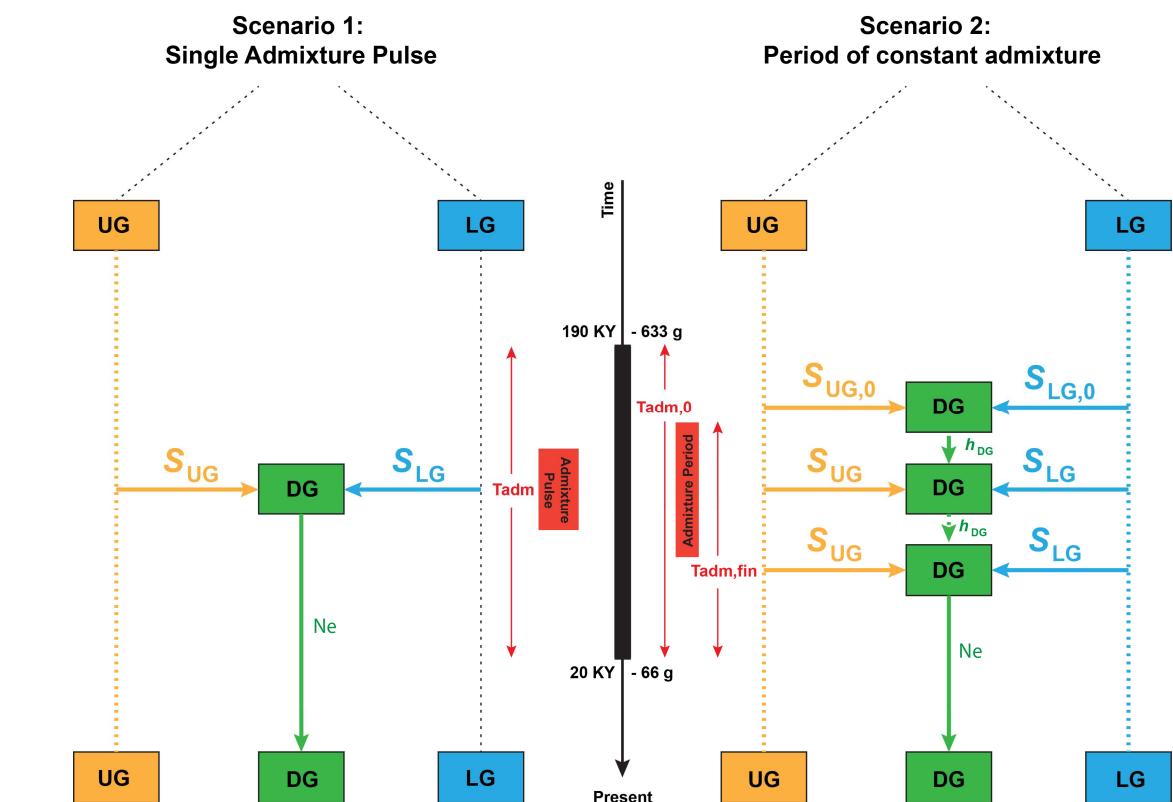
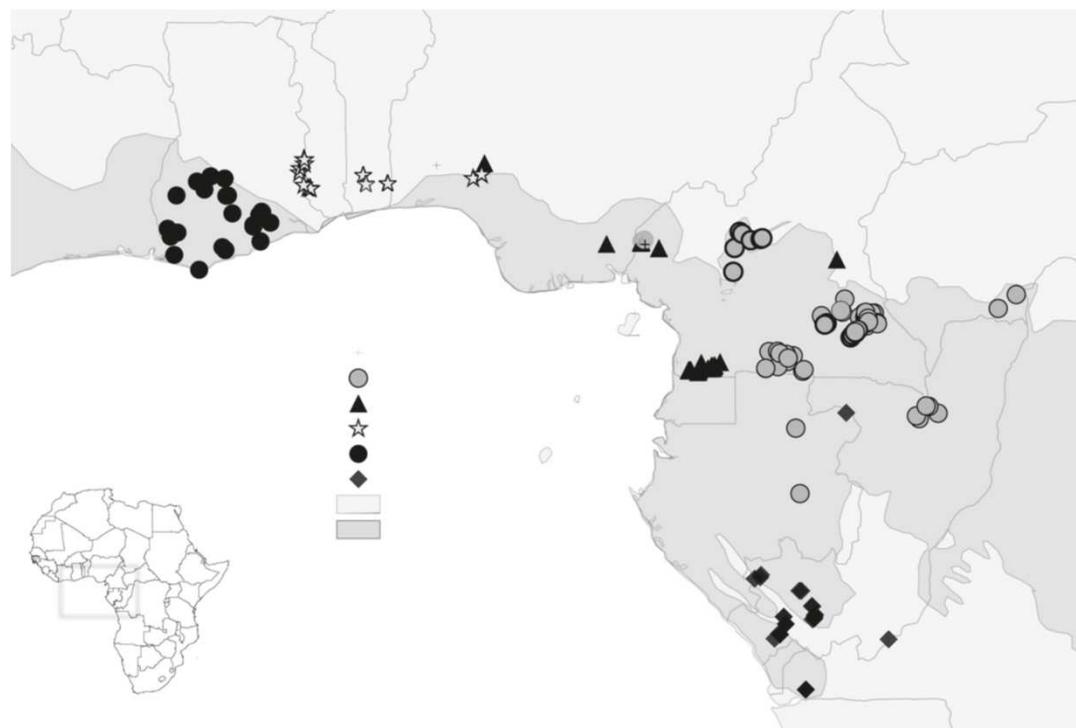
Ongoing !

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



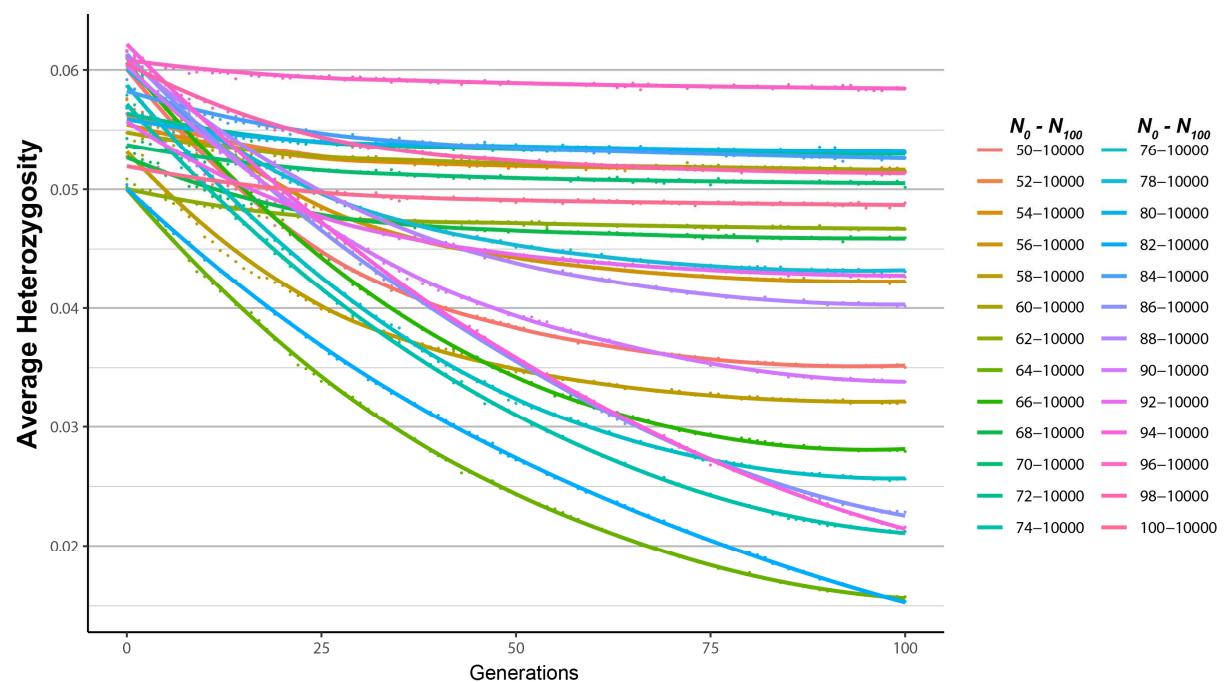
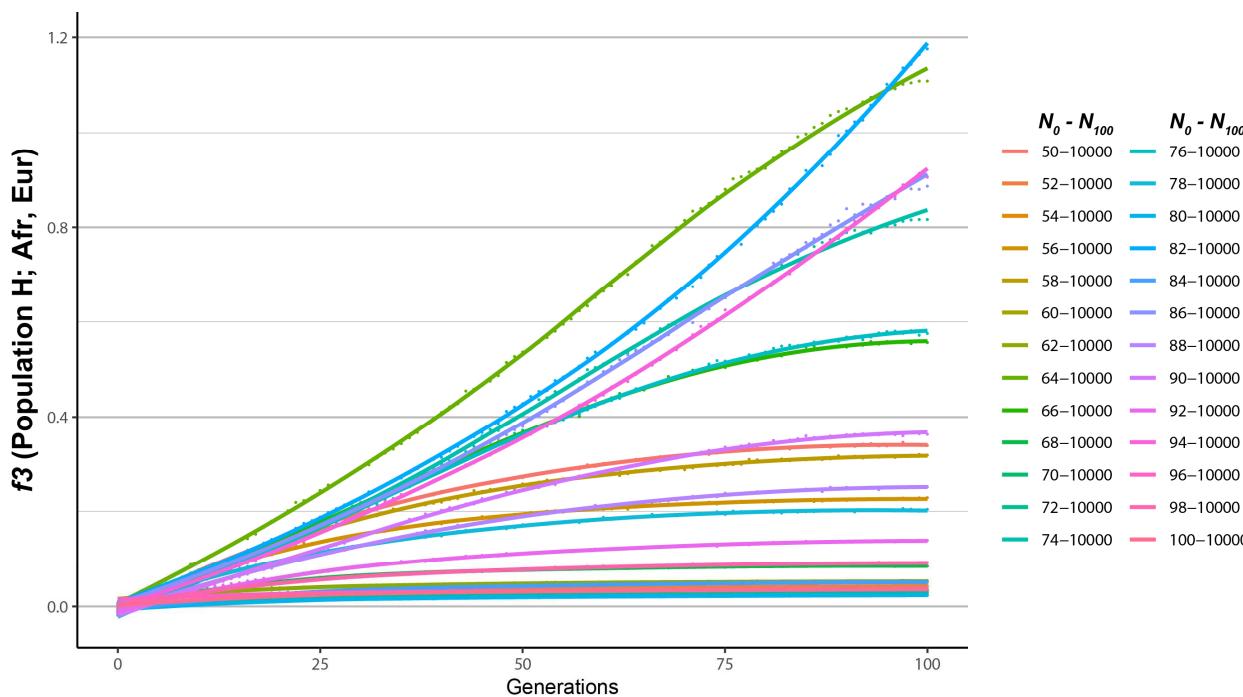
Ongoing !

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



Ongoing !

**Deterministic usage of *MetHis* for exploring admixture history influence on summary-statistics
with Maël Pretet (Master 2 student) and Romain Laurent (UMR7206)**



Ongoing !

Future developments of *Methis*:

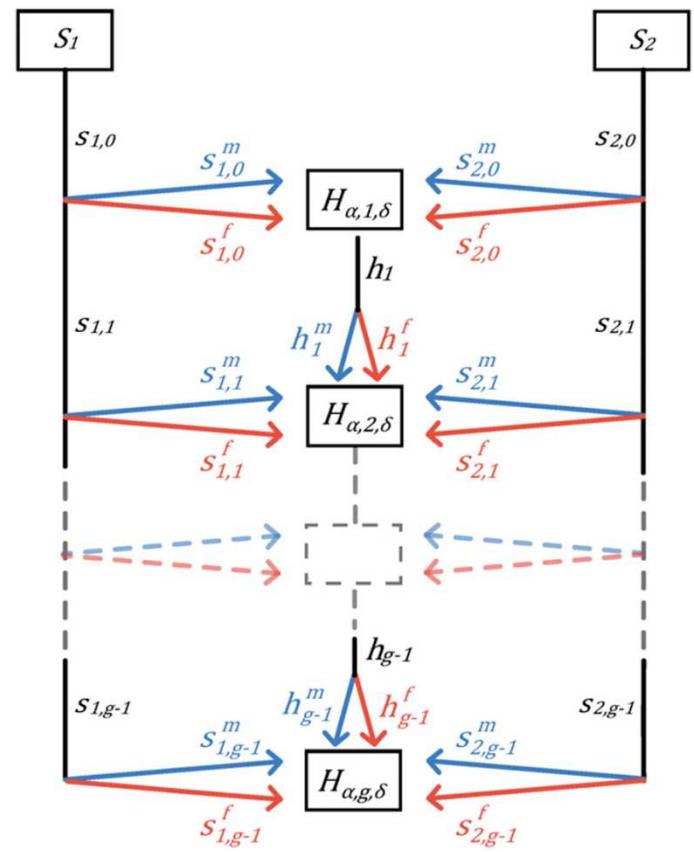
- Additional statistics (admixture LD, ROH...)
- Sequences simulations (and mutation...)
- M source populations (Verdu and Rosenberg 2011)
- sex-specific models and X, Y and mtDNA data and statistics

Marta Ciccarella, PhD student

dir. Paul Verdu and Jorge Rocha (Univ Porto)

-> Selection ?

Already in part feasible by parametrizing N_e changes accordingly



Great Many Thanks !

Co-authors !

Cesar Fortes-Lima (Uppsala University)

Romain Laurent (UMR7206)

Valentin Thouzeau (Univ. Paris-Dauphine)

Bruno Toupance (UMR7206)

Funding

ANR METHIS (15-CE32-0009-01)

Sven and Lilly Lawski's Foundation (N2019-0040)



Acknowledgements

Frédéric Austerlitz

Erkan O. Buzbas

Antoine Cools

Flora Jay

Olivier Hardy

Evelyne Heyer

Margueritte Lapierre

Guillaume Laval

Nina Marchi

Etienne Patin

Maël Pretet

Noah A. Rosenberg

Zachary A. Szpiech

MetHis – Simulations

MetHis simulates a **panmictic admixed population of diploid size N_g** at generation g , forward-in-time

MetHis does not simulate source populations, just the admixed population for the duration of the admixture process investigated

MetHis does not simulate mutation, yet...

Afro-American and Barbadian study cases:

- We neglect mutation -> recent admixture process
- Large source populations (Yoruba and British):
 - 20,000 gametes per source population
 - SNPs matching in frequencies observed 100,000 in each source
 - > fixed source populations gamete reservoir
 - Build randomly source population parents anew at each generation
- Independent SNPs -> we randomly build gametes and randomly pair them (avoiding selfing) according to admixture parameters at each generation
- For simplicity, we set a constant N_e at each generation in the admixed population of 1,000 diploid individuals

ABC-part of the *MetHis*-ABC framework

R packages:

- ***abcrf*** (Pudlo et al. 2016; Raynal et al. 2017): Random-Forest ABC
- ***abc*** (Csilléry et al. 2012): Neural-Network and Rejection ABC

Model Choice:

Random-Forest ABC with **10,000 simulations per scenario**

Parameter estimation:

Neural-Network with **100,000 simulations under the winning scenario**

Parameter estimation error:

Neural-Network with **100,000 simulations under the winning scenario**

1,000 closest simulations used in turn as pseudo-observed data

-> posterior parameter error

-> 95% Credibility Intervals accuracy

Results: MetHis simulations

