

Basic concepts on simulation-based likelihood free methods

Program

Part I: Basic concepts on likelihood free methods.

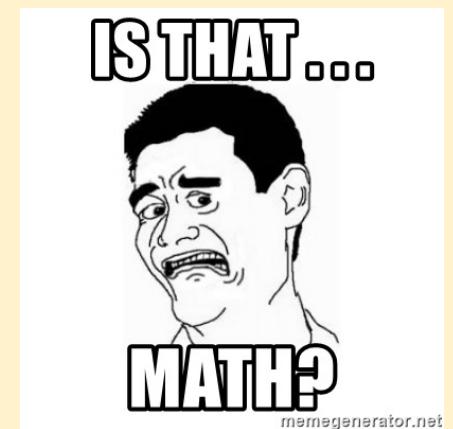
Likelihood-free methods

Complex evolutionary processes -> too many possible models and parameters to estimate the likelihood function (ML and Bayesian methods).

Post. Prob.

Likelihood Prior

$$\Pr(M|data) = \frac{\Pr(data|M) * \Pr(M)}{\Pr(data)}$$



Likelihood-free methods

Possible solutions:

Use a few (simple) models for all species/datasets
(IM; Migrate-n...)

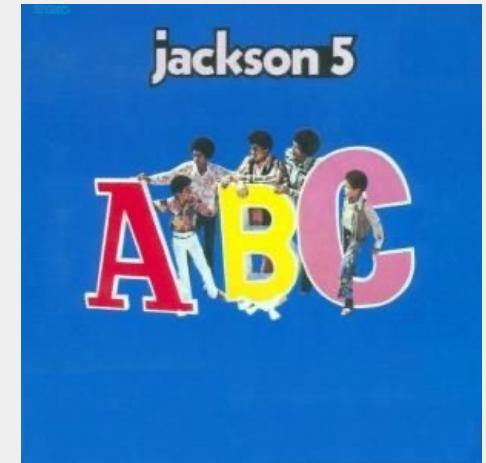
Likelihood-free methods

- ABC;
- Approximate Likelihoods;
- Deep Learning

1. Define Models

ABC

2. Sample parameters from a given distributions (prior)

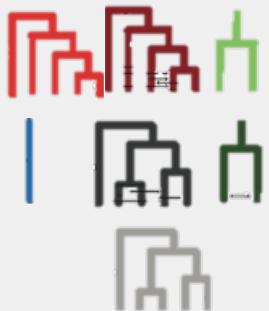


3. Simulate data for each model using the sampled parameter values

4. Calculate Summary Statistics (SuSt) from simulations and from the empirical data

5. Compare simulated and empirical data retaining only simulations that are more similar.

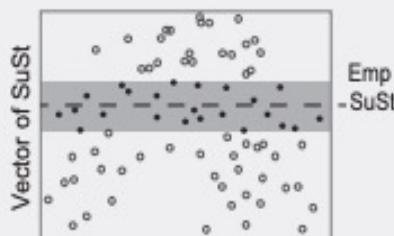
1) simulate each hypothesis



2b) extract SuSt from simulated data

SuSt
TT
S
D
θH
H
πW
πB

3b) retain only 20% more similar simulations



4b) approximate the posterior



Modified from
Perez et al. (2022) *Mol Ecol Res*

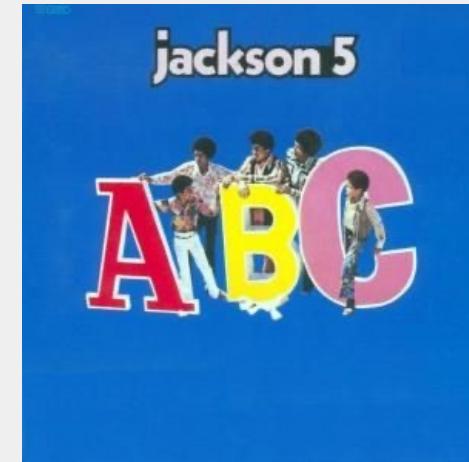
ABC

ABC implies 3 approximations:

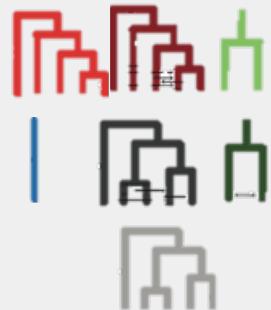
1. finite # simulations

2. informativeness of S

3. S don't match S^* exactly



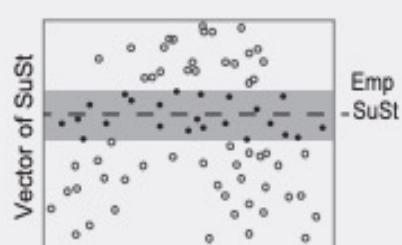
1) simulate each hypothesis



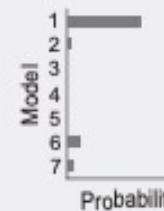
2b) extract SuSt from simulated data

T_T
S
D
θH
H
πW
πB

3b) retain only 20% more similar simulations

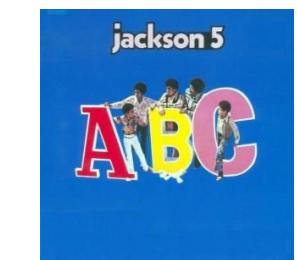


4b) approximate the posterior



Modified from
Perez et al. (2022) *Mol Ecol Res*

Why Deep Learning?

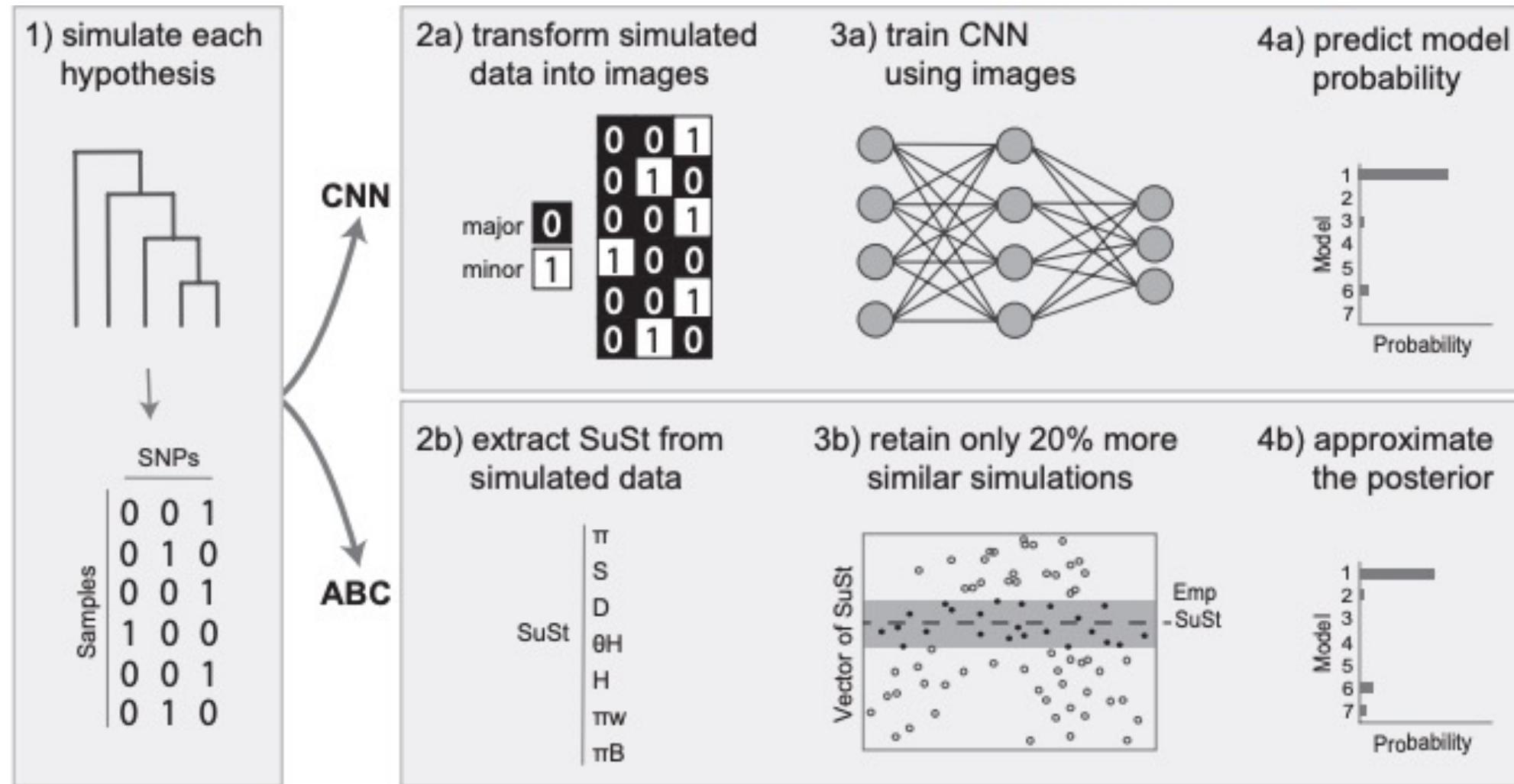


Deep Learning

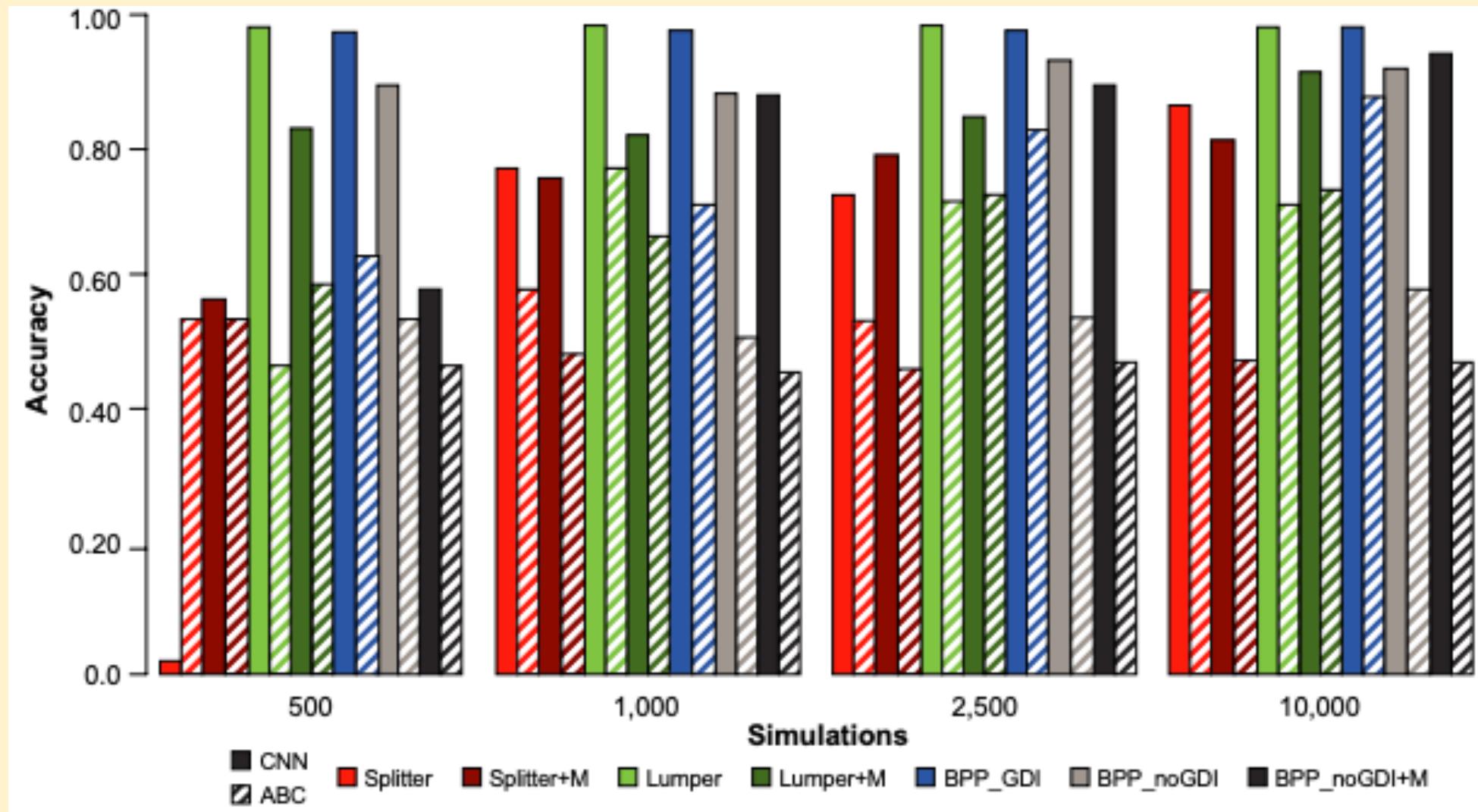


CJ Battey
Twitter 2018

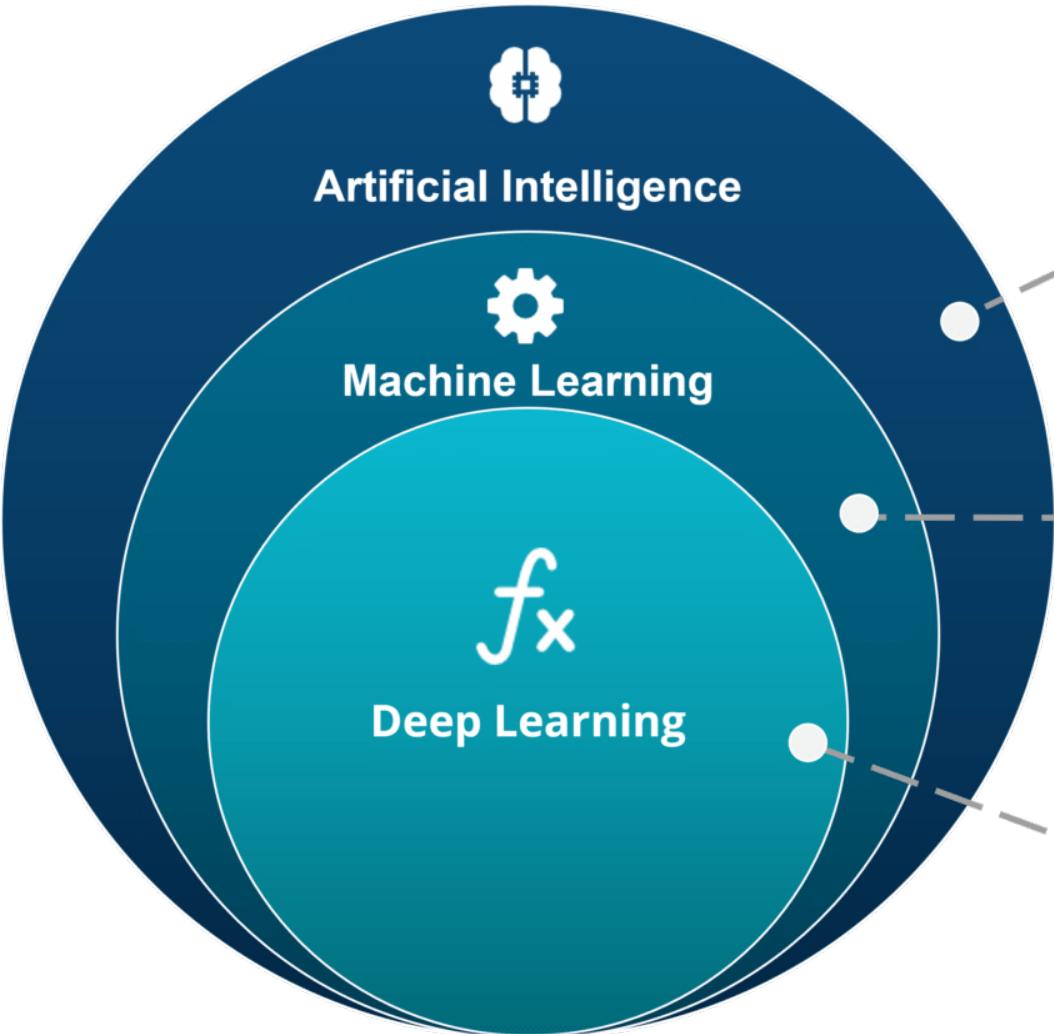
Why Deep Learning?



Why Deep Learning?



What is Deep Learning?

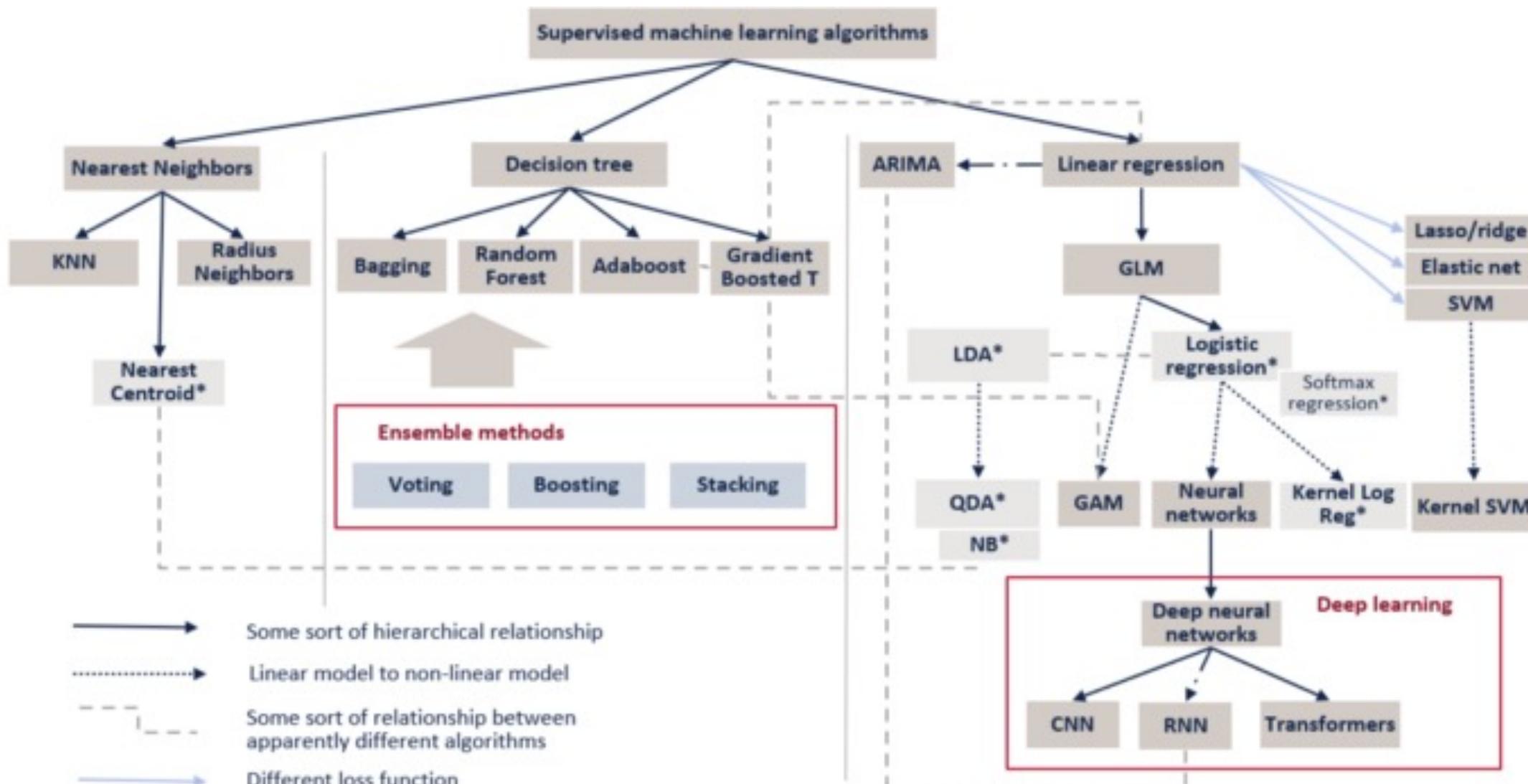


ARTIFICIAL INTELLIGENCE
A technique which enables machines
to mimic human behaviour

MACHINE LEARNING
Subset of AI technique which use
statistical methods to enable machines
to improve with experience

DEEP LEARNING
Subset of ML which make the
computation of multi-layer neural
network feasible

Some important definitions

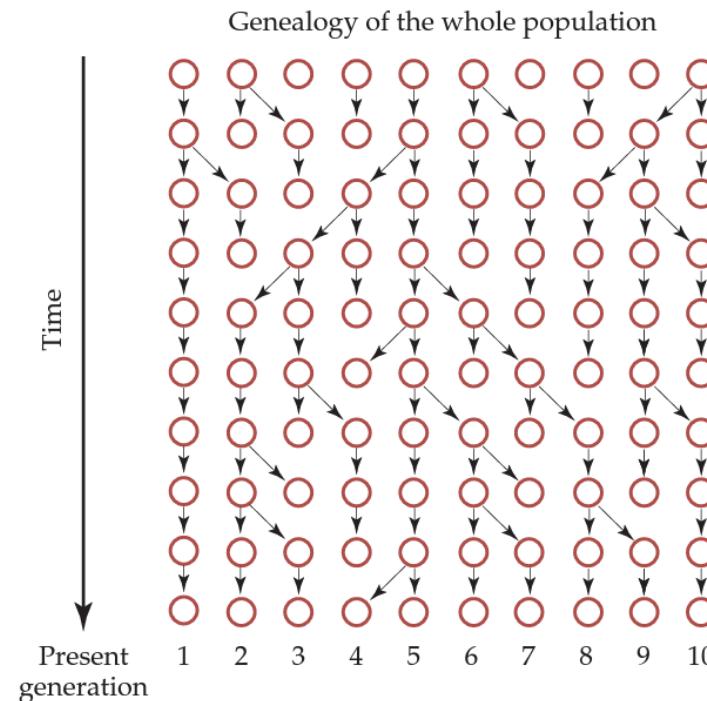


Part II: Simulating Genetic Data

Simulate Genetic Data

■ TABLE 6.1 Programs for simulating population samples of DNA sequences

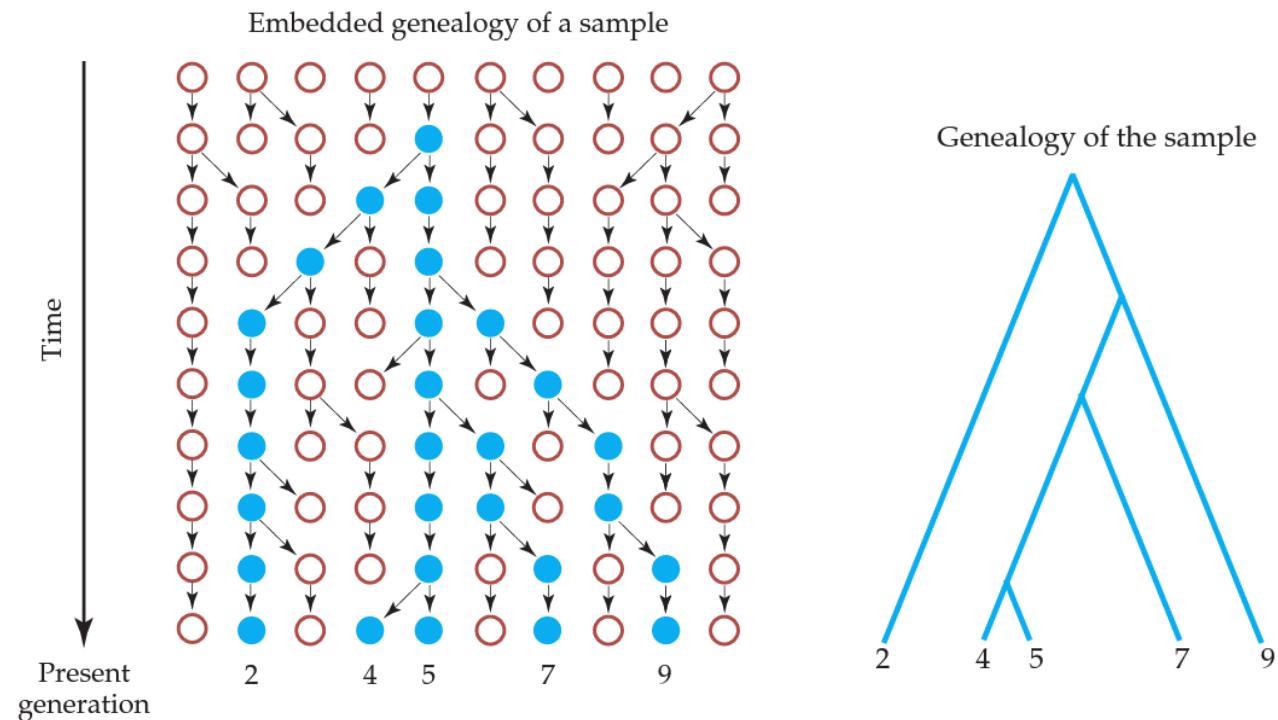
PROGRAM	SOURCE
COALESCENT ("BACKWARD") SIMULATION SOFTWARE	
<i>ms</i>	Hudson 1990; Hudson 2002
<i>GENOME</i>	Liang et al. 2007
<i>SIMCOAL/SIMCOAL 2.0</i>	Excoffier et al. 2000; Laval and Excoffier 2004
<i>CoaSim</i>	Mailund et al. 2005
<i>Recodon</i>	Arenas and Posada 2007
<i>discoal</i>	Kern and Schrider 2016
<i>msprime</i>	Kelleher et al. 2016
WRIGHT-FISHER ("FORWARD") SIMULATION SOFTWARE:	
<i>EASYPOL</i>	Baloux 2001
<i>simuPop</i>	Peng and Kimmel 2005
<i>Nemo</i>	Guillaume and Rougemont 2006
<i>FREGENE</i>	Hoggart et al. 2007; Chadeau-Hyam et al. 2008
<i>ForSim</i>	Lambert et al. 2008
<i>FORWSIM</i>	Padhukasahasram et al. 2008
<i>GENOMEPOP</i>	Carvajal-Rodriguez 2008
<i>SFS_CODE</i>	Hernandez 2008
<i>FFPopSim</i>	Zanini and Neher 2012
<i>fwdpp</i>	Thornton 2014
<i>SLiM/SLiM 2</i>	Messer 2013; Haller and Messer 2017
<i>ARGON</i>	Palamara 2016
APPROXIMATE COALESCENT ("SIDeways") SIMULATION SOFTWARE	
<i>FastCoal</i>	McVean and Cardin 2005; Marjoram and Wall 2006
<i>MaCS</i>	Chen et al. 2009
<i>fastsimcoal/fastsimcoal2</i>	(Excoffier and Foll 2011; Excoffier et al. 2013)



Simulate Genetic Data

■ TABLE 6.1 Programs for simulating population samples of DNA sequences

PROGRAM	SOURCE
COALESCENT ("BACKWARD") SIMULATION SOFTWARE	
<i>ms</i>	Hudson 1990; Hudson 2002
<i>GENOME</i>	Liang et al. 2007
<i>SIMCOAL/SIMCOAL 2.0</i>	Excoffier et al. 2000; Laval and Excoffier 2004
<i>CoaSim</i>	Mailund et al. 2005
<i>Recodon</i>	Arenas and Posada 2007
<i>discoal</i>	Kern and Schrider 2016
<i>msprime</i>	Kelleher et al. 2016
WRIGHT-FISHER ("FORWARD") SIMULATION SOFTWARE:	
<i>EASYPOP</i>	Baloux 2001
<i>simuPop</i>	Peng and Kimmel 2005
<i>Nemo</i>	Guillaume and Rougemont 2006
<i>FREGENE</i>	Hoggart et al. 2007; Chadeau-Hyam et al. 2008
<i>ForSim</i>	Lambert et al. 2008
<i>FORWSIM</i>	Padhukasahasram et al. 2008
<i>GENOMEPOP</i>	Carvajal-Rodriguez 2008
<i>SFS_CODE</i>	Hernandez 2008
<i>FFPopSim</i>	Zanini and Neher 2012
<i>fwdpp</i>	Thornton 2014
<i>SLiM/SLiM 2</i>	Messer 2013; Haller and Messer 2017
<i>ARGON</i>	Palamara 2016
APPROXIMATE COALESCENT ("SIDeways") SIMULATION SOFTWARE	
<i>FastCoal</i>	McVean and Cardin 2005; Marjoram and Wall 2006
<i>MaCS</i>	Chen et al. 2009
<i>fastsimcoal/fastsimcoal2</i>	(Excoffier and Foll 2011; Excoffier et al. 2013)



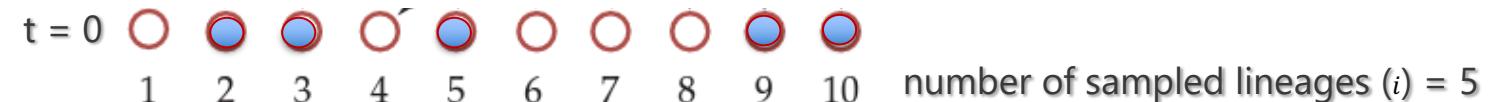
Simulate Genetic Data

$2N$ diploids = 10

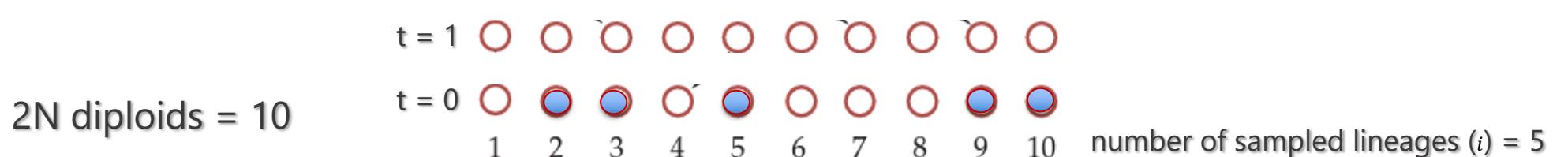


Simulate Genetic Data

$2N$ diploids = 10

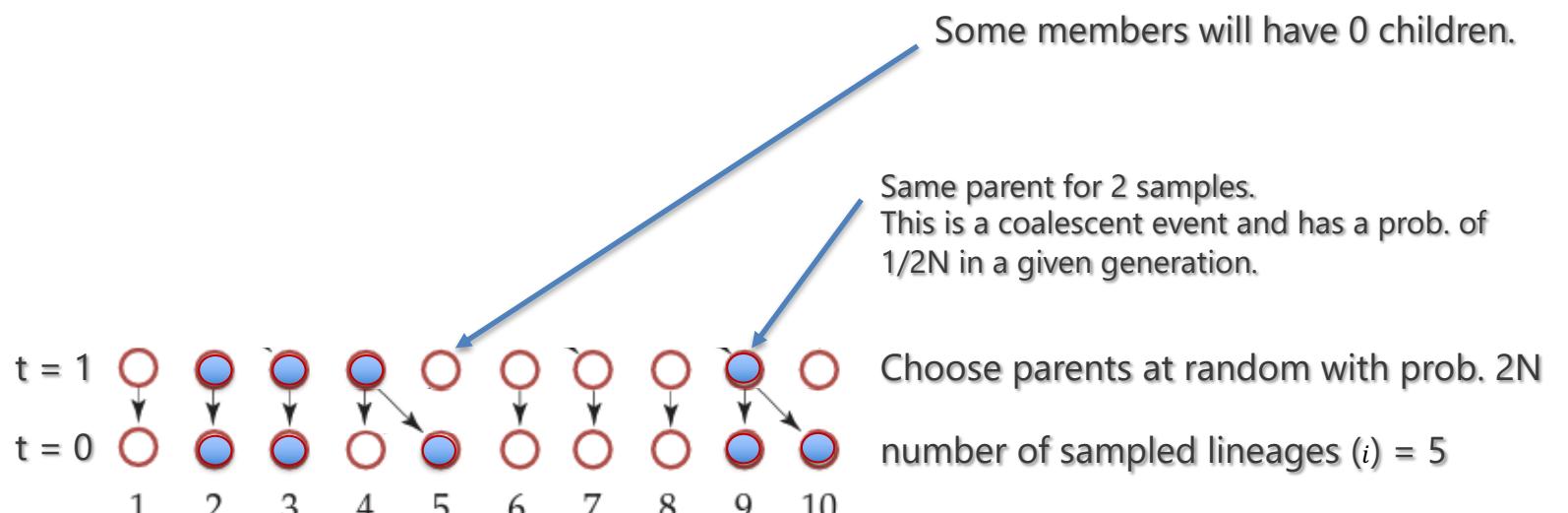


Simulate Genetic Data

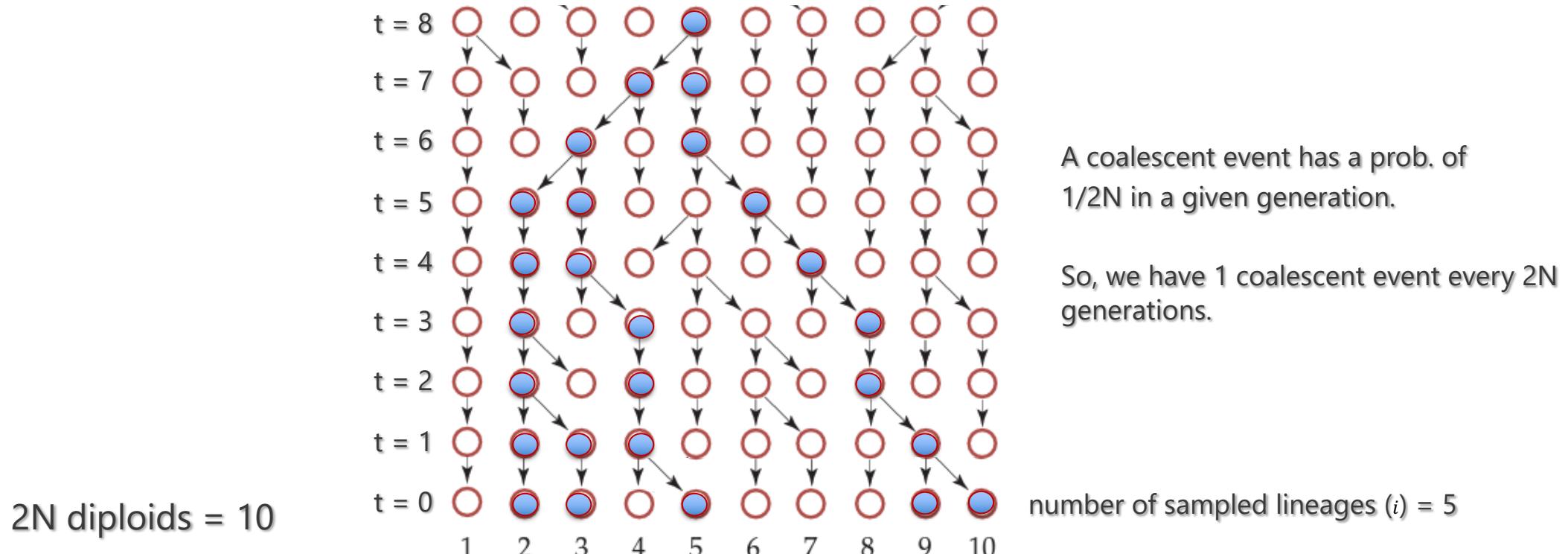


Simulate Genetic Data

$2N$ diploids = 10



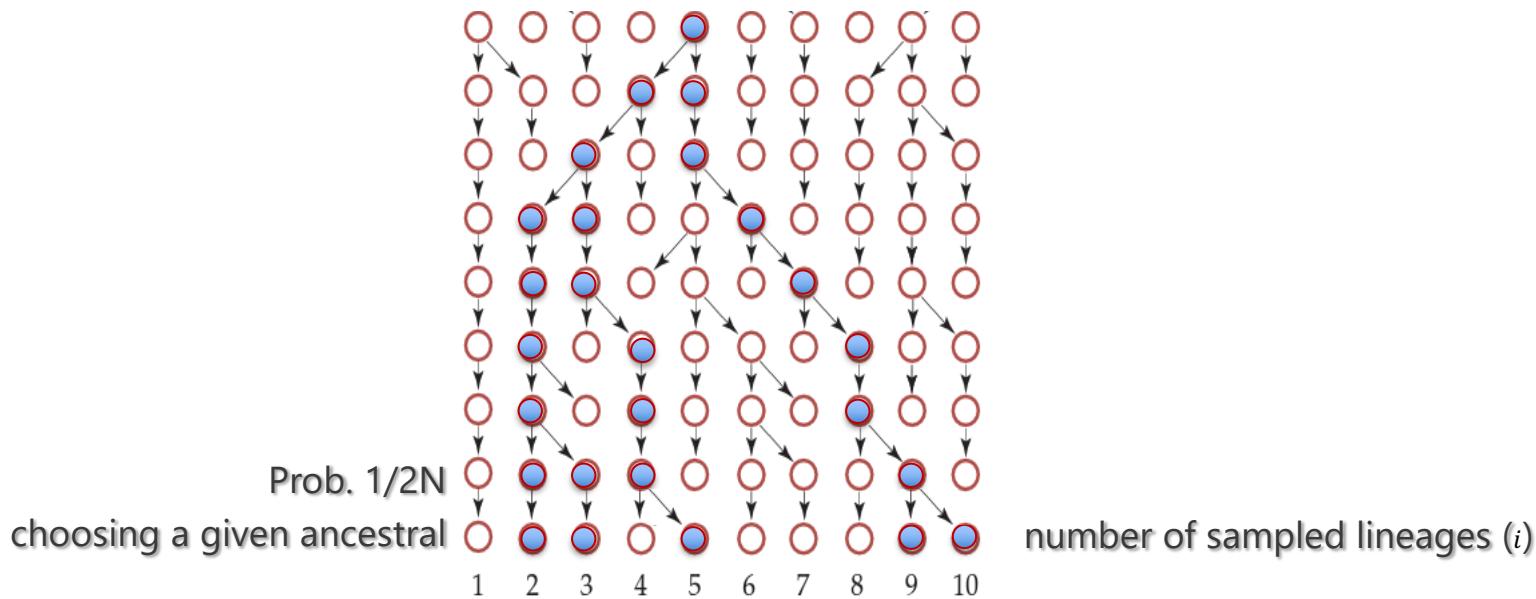
Simulate Genetic Data



A coalescent event has a prob. of $1/2N$ in a given generation.

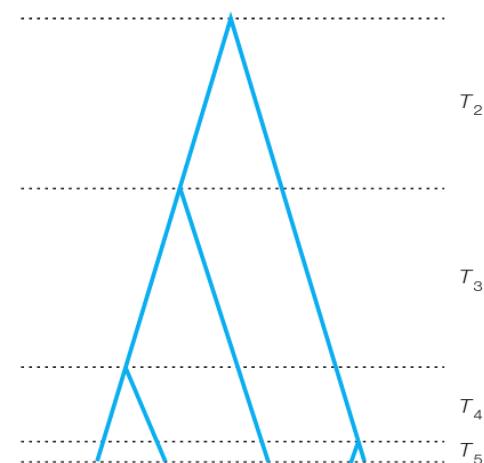
So, we have 1 coalescent event every $2N$ generations.

$2N$ diploids = 10



$$\binom{i}{2} = \frac{i!}{i! - 2!2!} = i(i-1)/2$$

to choose the same ancestral



Simulate Genetic Data

PROBABILITY OF A COALESCENT EVENT

$$P(i \rightarrow i-1) = \frac{i(i-1)}{4N}$$

TIME BETWEEN COALESCENT EVENTS

$$E(T_i) = \frac{4N}{i(i-1)}$$

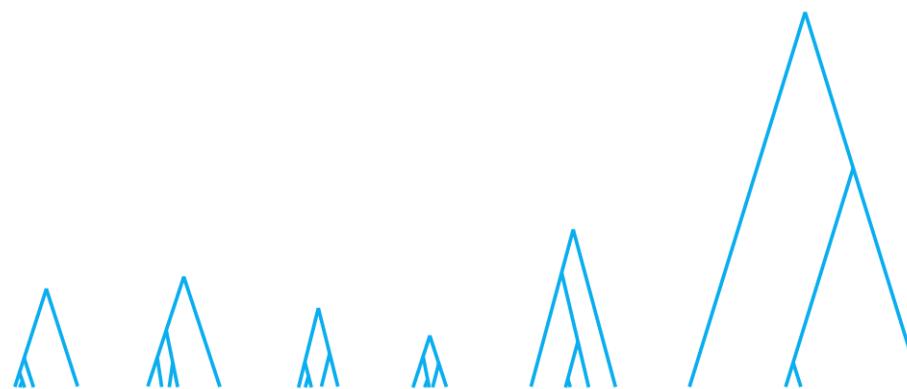
HEIGHT OF THE COALESCENT GENEALOGY

$$E(T_{\text{MRCA}}) = 4N \left(1 - \frac{1}{n}\right)$$

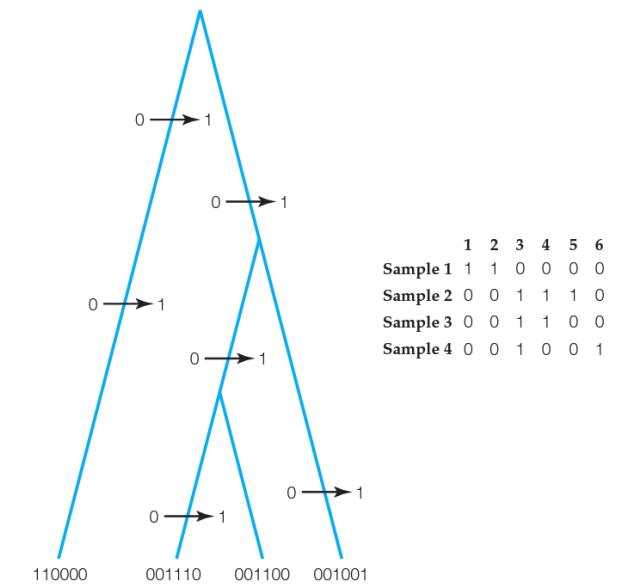
2N on average for 2 samples.
Approaches 4N When sample size gets larger.

Simulate Genetic Data

Mutations/branch = Poisson dist. with mean $\frac{t * \theta}{2}$

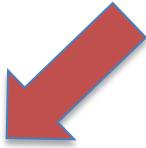


Evolutionary variance

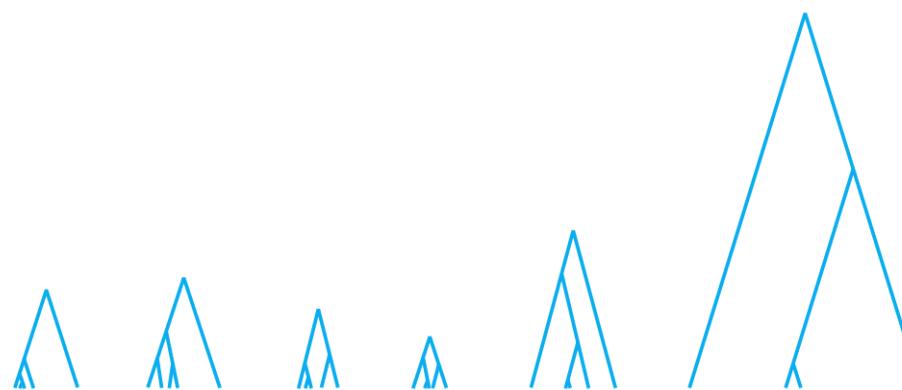


Sampling variance

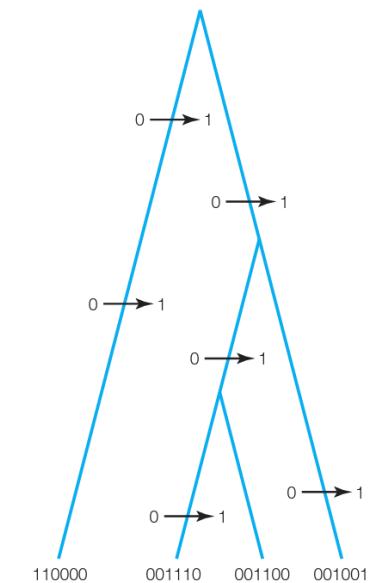
Simulate Genetic Data



Mutations/branch = Poisson dist. with mean $\frac{t * \theta}{2}$



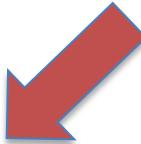
Evolutionary variance



Sampling variance

	1	2	3	4	5	6
Sample 1	1	1	0	0	0	0
Sample 2	0	0	1	1	1	0
Sample 3	0	0	1	1	0	0
Sample 4	0	0	1	0	0	1

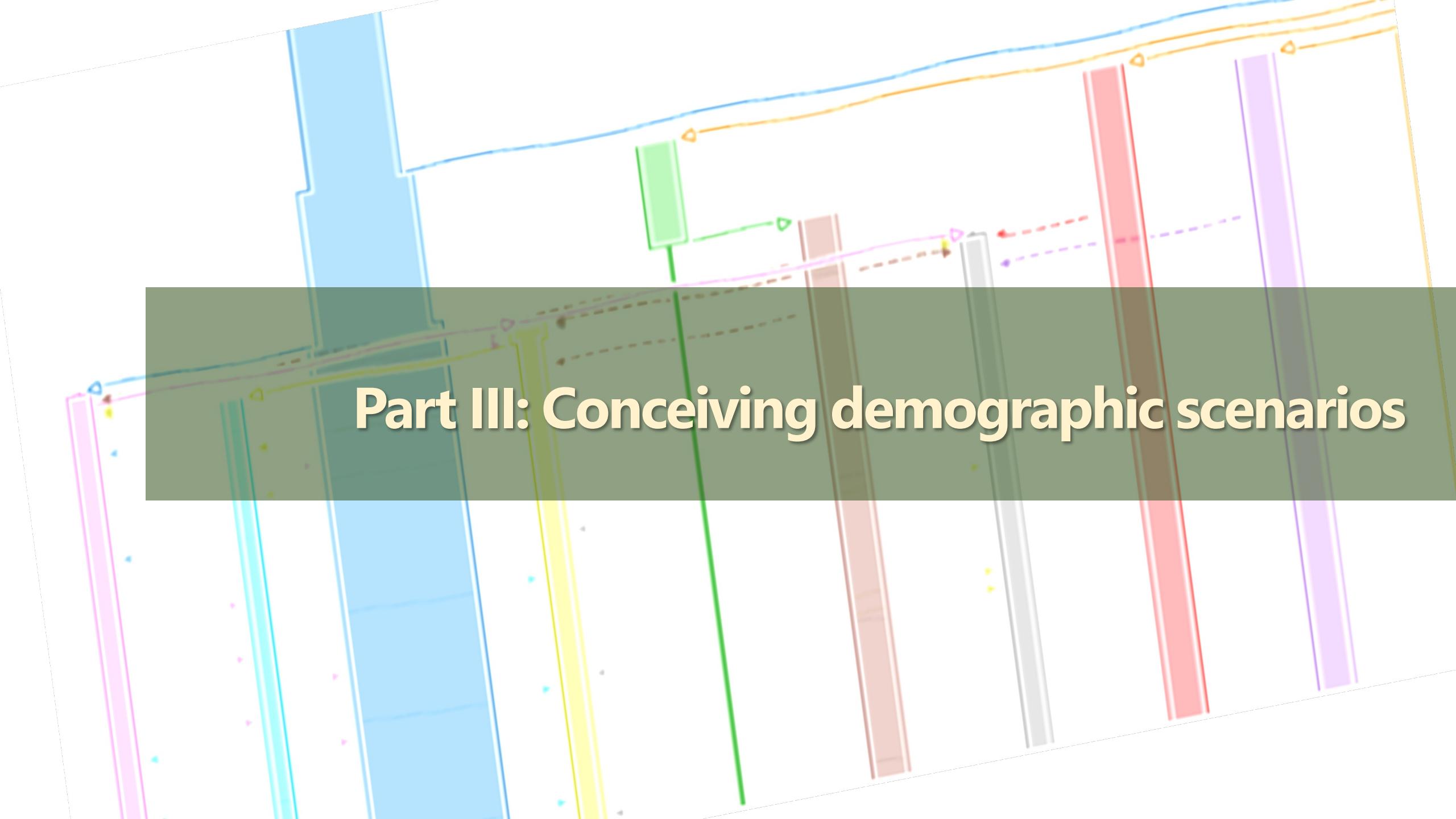
Simulate Genetic Data



Mutations/branch = Poisson dist. with mean $\frac{t * \theta}{2}$

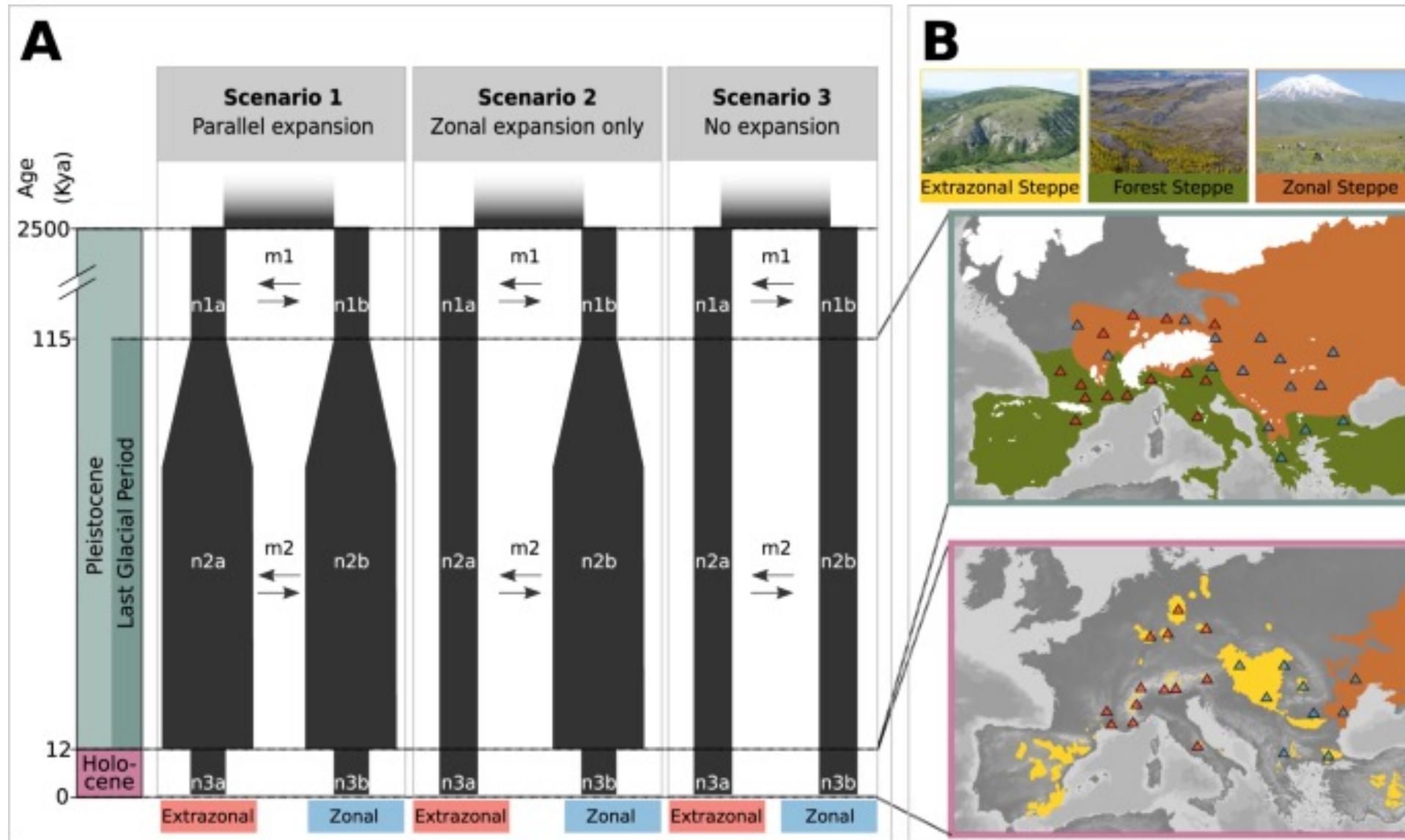
- Average time to the MRCA of any two randomly sampled sequences = $2Ne$ generations.
- Mutations occur with probability μ per generation.
- For diploids, we expect the number of differences between two sequences to be:

$$\theta = 2 * 2Ne * \mu = 4Ne\mu$$



Part III: Conceiving demographic scenarios

Simulate Genetic Data



Simulate Genetic Data – ms (Hudson 2002)

BIOINFORMATICS APPLICATIONS NOTE

Vol. 18 no. 2 2002
Pages 337–338



Generating samples under a Wright–Fisher neutral model of genetic variation

Richard R. Hudson

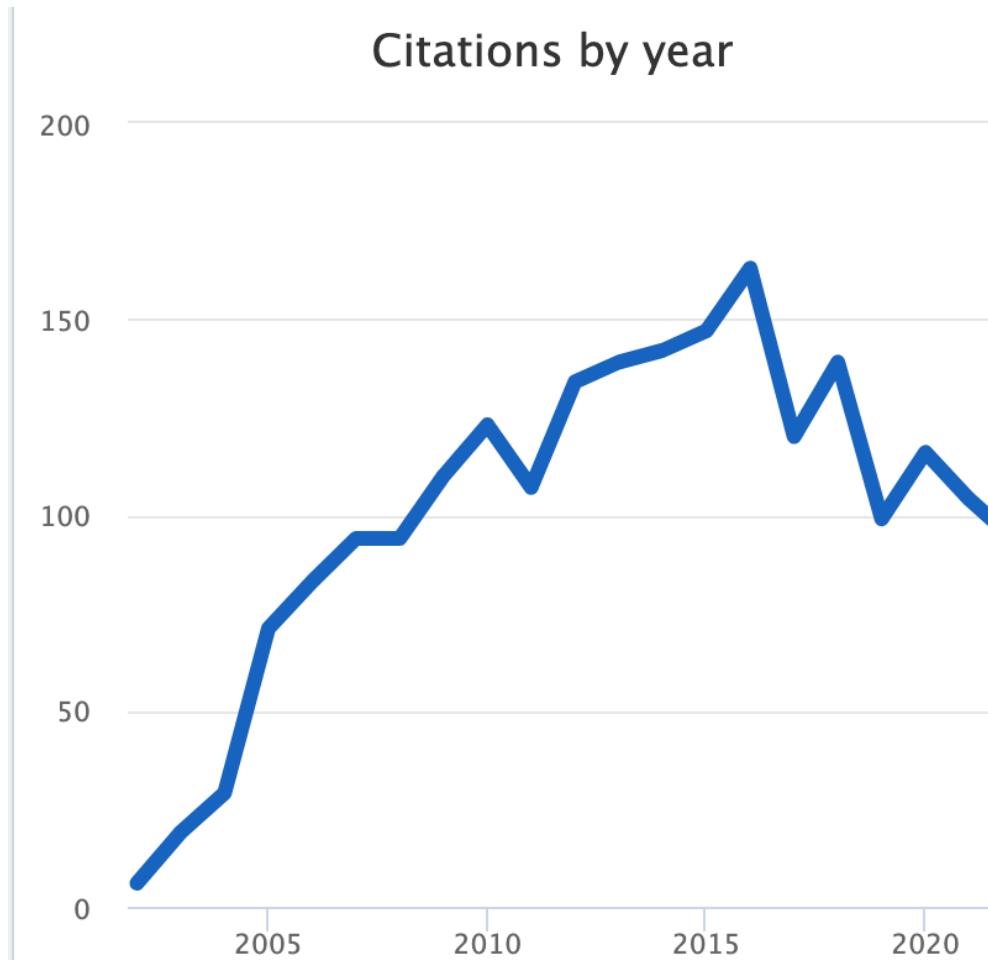
Department of Ecology and Evolution, University of Chicago, 1101 E. 57th Street,
Chicago, IL 60637, USA

Received on August 8, 2001; revised and accepted on September 13, 2001

- Seminal simulator.
- Allows to fix the number of segregating sites.

2,142 citations

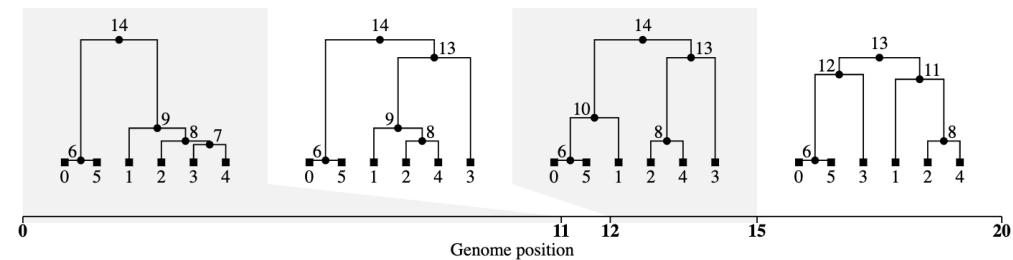
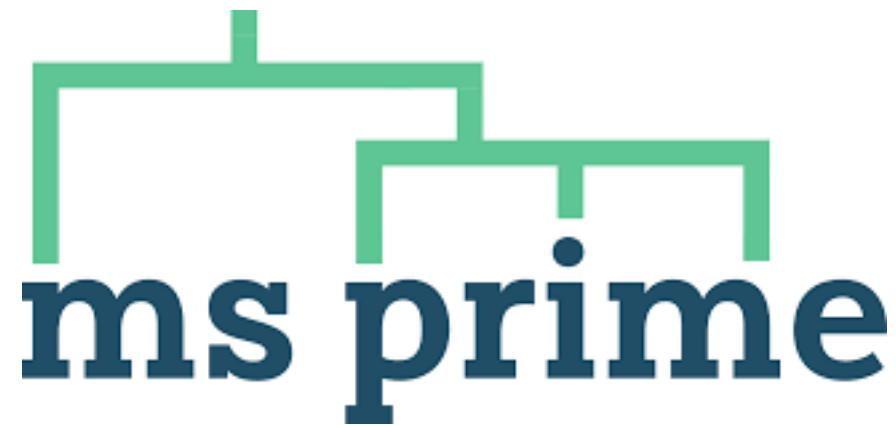
Citations by year



Simulate Genetic Data – other options

■ TABLE 6.1 Programs for simulating population samples of DNA sequences

PROGRAM	SOURCE
COALESCENT ("BACKWARD") SIMULATION SOFTWARE	
<i>ms</i>	Hudson 1990; Hudson 2002
<i>GENOME</i>	Liang et al. 2007
<i>SIMCOAL/SIMCOAL 2.0</i>	Excoffier et al. 2000; Laval and Excoffier 2004
<i>CoaSim</i>	Mailund et al. 2005
<i>Recodon</i>	Arenas and Posada 2007
<i>discoal</i>	Kern and Schrider 2016
<i>msprime</i>	Kelleher et al. 2016
WRIGHT-FISHER ("FORWARD") SIMULATION SOFTWARE:	
<i>EASYPOL</i>	Balloux 2001
<i>simuPop</i>	Peng and Kimmel 2005
<i>Nemo</i>	Guillaume and Rougemont 2006
<i>FREGENE</i>	Hoggart et al. 2007; Chadeau-Hyam et al. 2008
<i>ForSim</i>	Lambert et al. 2008
<i>FORWSIM</i>	Padhukasahasram et al. 2008
<i>GENOMEPOP</i>	Carvajal-Rodriguez 2008
<i>SFS_CODE</i>	Hernandez 2008
<i>FFPopSim</i>	Zanini and Neher 2012
<i>fwdpp</i>	Thornton 2014
<i>SLiM/SLiM 2</i>	Messer 2013; Haller and Messer 2017
<i>ARGON</i>	Palamara 2016
APPROXIMATE COALESCENT ("SIDeways") SIMULATION SOFTWARE	
<i>FastCoal</i>	McVean and Cardin 2005; Marjoram and Wall 2006
<i>MaCS</i>	Chen et al. 2009
<i>fastsimcoal/fastsimcoal2</i>	(Excoffier and Foll 2011; Excoffier et al. 2013)



Simulate Genetic Data – other options

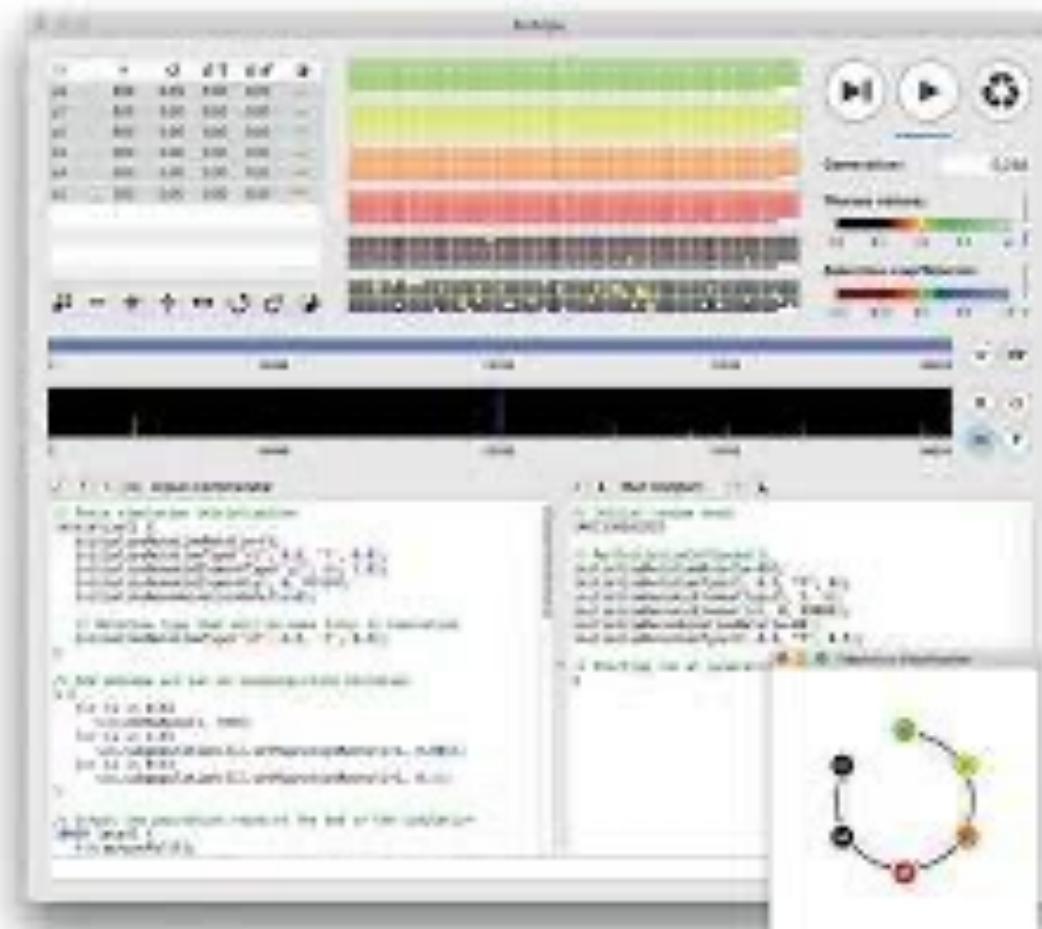
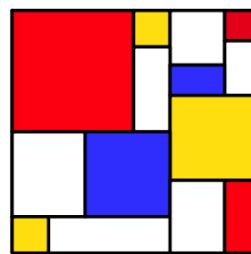
■ TABLE 6.1 Programs for simulating population samples of DNA sequences

PROGRAM	SOURCE
COALESCENT ("BACKWARD") SIMULATION SOFTWARE	
<i>ms</i>	Hudson 1990; Hudson 2002
<i>GENOME</i>	Liang et al. 2007
<i>SIMCOAL/SIMCOAL 2.0</i>	Excoffier et al. 2000; Laval and Excoffier 2004
<i>CoaSim</i>	Mailund et al. 2005
<i>Recodon</i>	Arenas and Posada 2007
<i>discoal</i>	Kern and Schrider 2016
<i>msprime</i>	Kelleher et al. 2016
WRIGHT-FISHER ("FORWARD") SIMULATION SOFTWARE:	
<i>EASYPOP</i>	Balloux 2001
<i>simuPop</i>	Peng and Kimmel 2005
<i>Nemo</i>	Guillaume and Rougemont 2006
<i>FREGENE</i>	Hoggart et al. 2007; Chadeau-Hyam et al. 2008
<i>ForSim</i>	Lambert et al. 2008
<i>FORWSIM</i>	Padhukasahasram et al. 2008
<i>GENOMEPOP</i>	Carvajal-Rodriguez 2008
<i>SFS_CODE</i>	Hernandez 2008
<i>FFPopSim</i>	Zanini and Neher 2012
<i>fwdpp</i>	Thornton 2014
<i>SLiM/SLiM 2</i>	Messer 2013; Haller and Messer 2017
<i>ARGON</i>	Palamara 2016
APPROXIMATE COALESCENT ("SIDeways") SIMULATION SOFTWARE	
<i>FastCoal</i>	McVean and Cardin 2005; Marjoram and Wall 2006
<i>MaCS</i>	Chen et al. 2009
<i>fastsimcoal/fastsimcoal2</i>	(Excoffier and Foll 2011; Excoffier et al. 2013)

SLiM: An Evolutionary Simulation Framework

Benjamin C. Haller and Philipp W. Messer

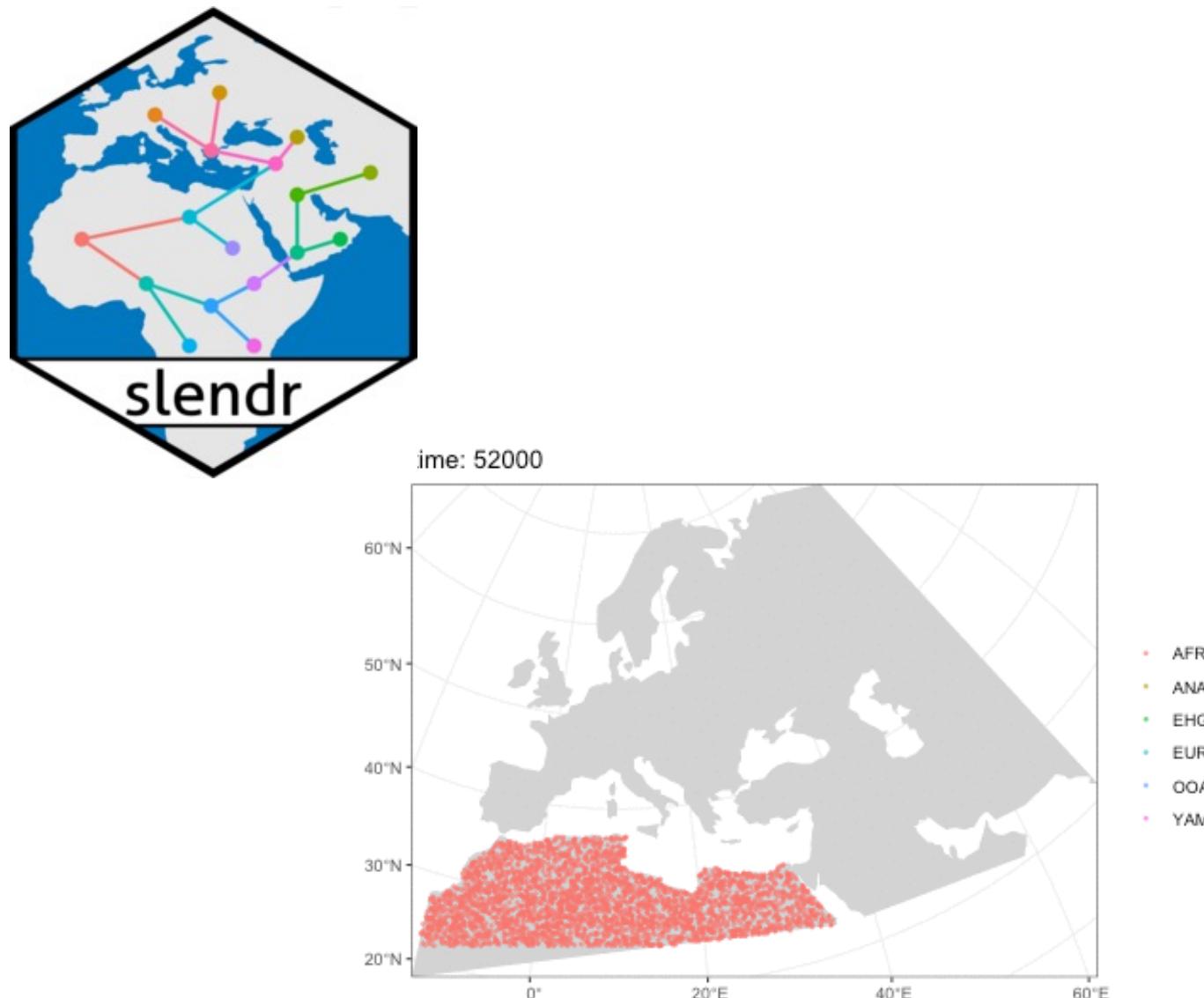
Dept. of Computational Biology
Cornell University, Ithaca, NY 14853
Correspondence: bhaller@benhaller.co



Simulate Genetic Data – other options

■ TABLE 6.1 Programs for simulating population samples of DNA sequences

PROGRAM	SOURCE
COALESCENT ("BACKWARD") SIMULATION SOFTWARE	
<i>ms</i>	Hudson 1990; Hudson 2002
<i>GENOME</i>	Liang et al. 2007
<i>SIMCOAL/SIMCOAL 2.0</i>	Excoffier et al. 2000; Laval and Excoffier 2004
<i>CoaSim</i>	Mailund et al. 2005
<i>Recodon</i>	Arenas and Posada 2007
<i>discoal</i>	Kern and Schrider 2016
<i>msprime</i>	Kelleher et al. 2016
WRIGHT-FISHER ("FORWARD") SIMULATION SOFTWARE:	
<i>EASYPOP</i>	Balloux 2001
<i>simuPop</i>	Peng and Kimmel 2005
<i>Nemo</i>	Guillaume and Rougemont 2006
<i>FREGENE</i>	Hoggart et al. 2007; Chadeau-Hyam et al. 2008
<i>ForSim</i>	Lambert et al. 2008
<i>FORWSIM</i>	Padhukasahasram et al. 2008
<i>GENOMEPOP</i>	Carvajal-Rodriguez 2008
<i>SFS_CODE</i>	Hernandez 2008
<i>FFPopSim</i>	Zanini and Neher 2012
<i>fwdpp</i>	Thornton 2014
<i>SLiM/SLiM 2</i>	Messer 2013; Haller and Messer 2017
<i>ARGON</i>	Palamara 2016
APPROXIMATE COALESCENT ("SIDeways") SIMULATION SOFTWARE	
<i>FastCoal</i>	McVean and Cardin 2005; Marjoram and Wall 2006
<i>MaCS</i>	Chen et al. 2009
<i>fastsimcoal/fastsimcoal2</i>	(Excoffier and Foll 2011; Excoffier et al. 2013)





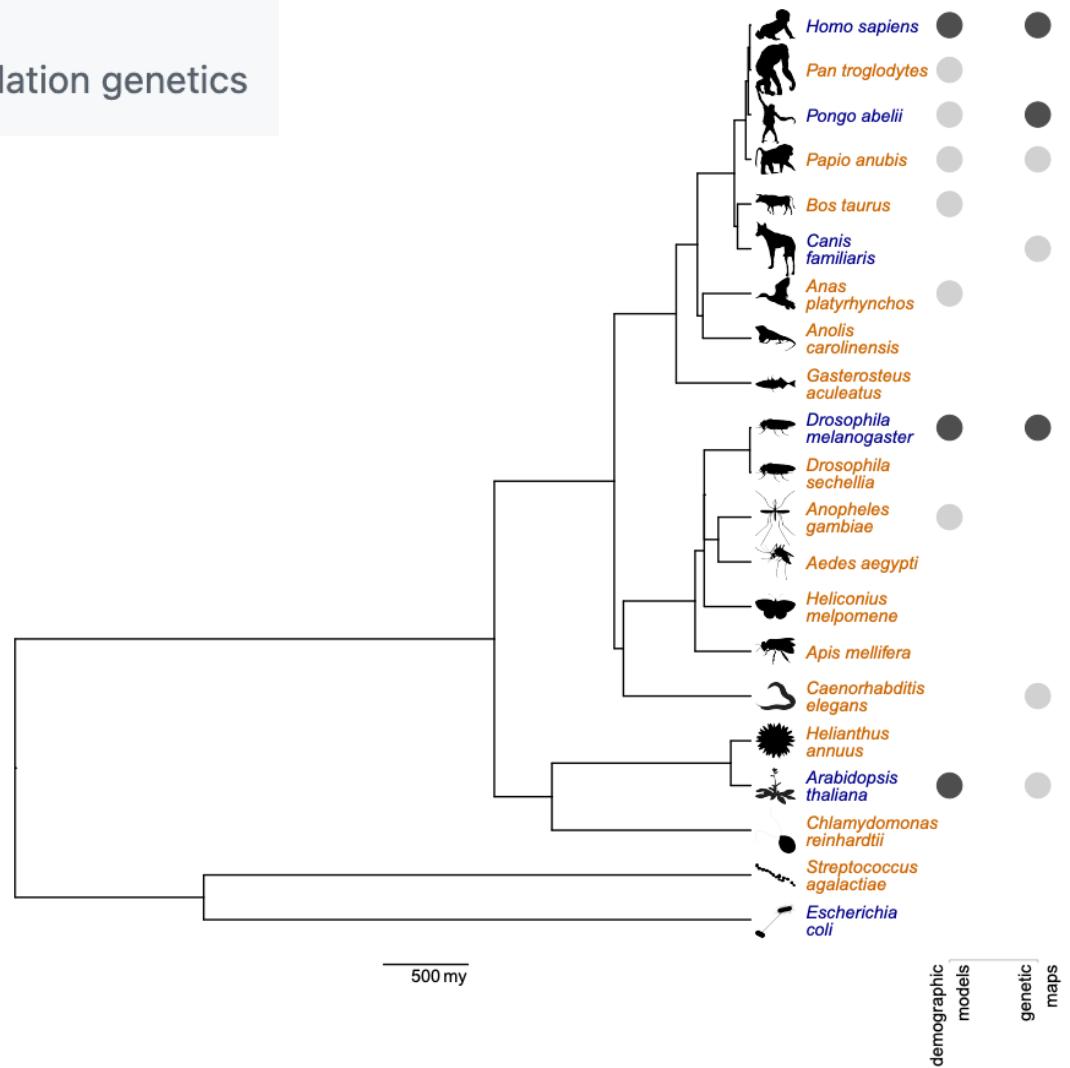
Simulate Genetic Data – other options

PopSim Consortium

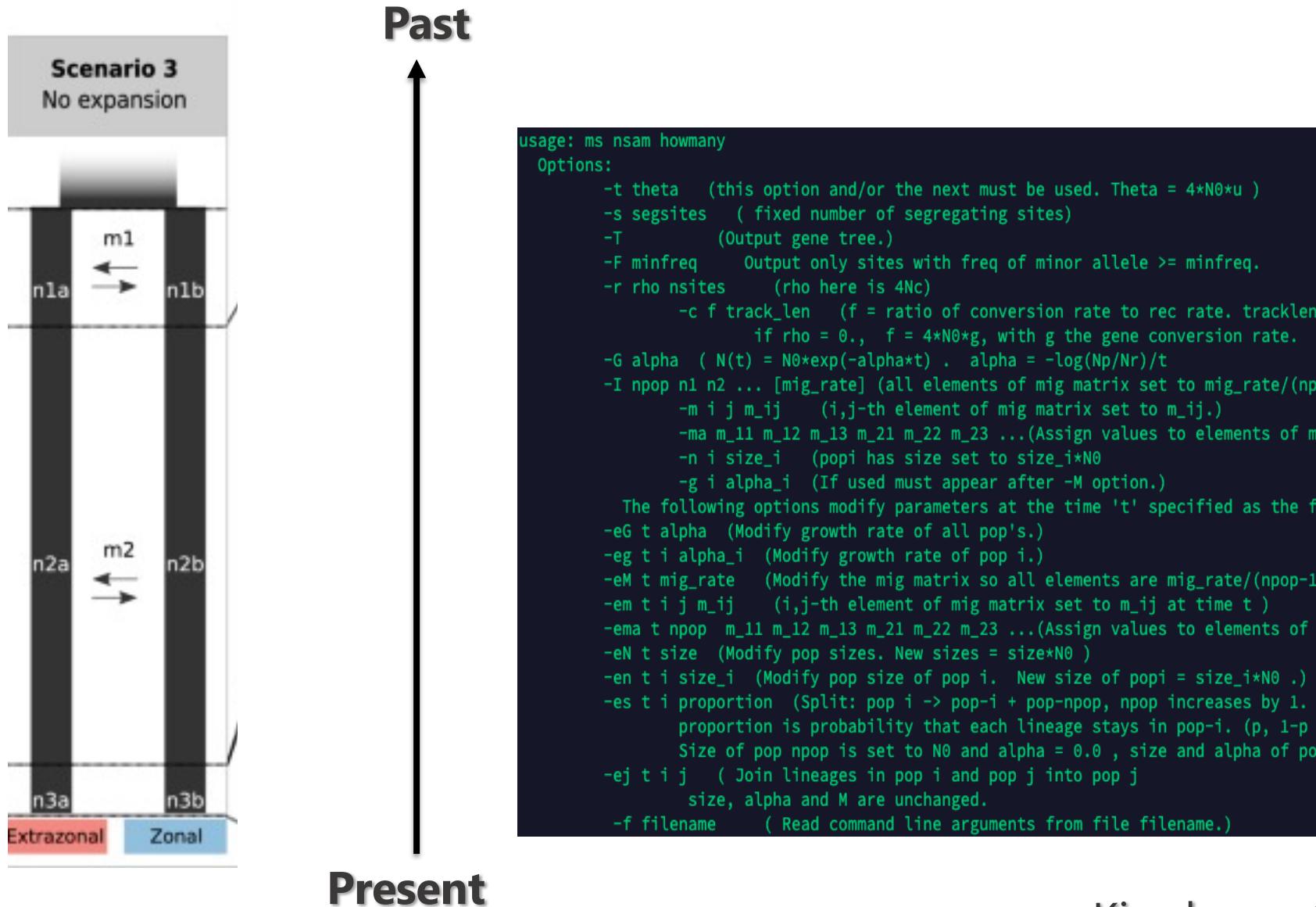
A community-driven effort to standardize population genetics

Expanding the `stdpopsim` species catalog, and lessons learned for realistic genome simulations

M. Elise Lauterbur^{1,+}, Maria Izabel A. Cavassim^{2,*}, Ariella L. Gladstein^{3,*}, Graham Gower^{4,*}, Nathaniel S. Pope^{5*}, Georgia Tsambos^{6,*}, Jeff Adrion^{5,7}, Saurabh Belsare⁵, Arjun Biddanda⁸, Victoria Caudill⁵, Jean Cury⁹, Ignacio Echevarria¹⁰, Benjamin C. Haller¹¹, Ahmed R. Hasan^{12,13}, Xin Huang^{14,15}, Leonardo Nicola Martin Iasi¹⁶, Ekaterina Noskova¹⁷, Jana Obšteter¹⁸, Vitor Antonio Corrêa Pavinato¹⁹, Alice Pearson^{20,21}, David Peede^{22,23}, Manolo F. Perez²⁴, Murillo F. Rodrigues⁵, Chris C. R. Smith⁵, Jeffrey P. Spence²⁵, Anastasia Teterina⁵, Silas Tittes⁵, Per Unneberg²⁶, Juan Manuel Vazquez²⁷, Ryan K. Waples²⁸, Anthony Wilder Wohns²⁹, Yan Wong³⁰, Franz Baumdicker³¹, Reed A. Cartwright³², Gregor Gorjanc³³, Ryan N. Gutenkunst³⁴, Jerome Kelleher³⁰, Andrew D. Kern⁵, Aaron P. Ragsdale³⁵, Peter L. Ralph^{5,36}, Daniel R. Schrider³⁷, and Ilan Gronau^{38,+}



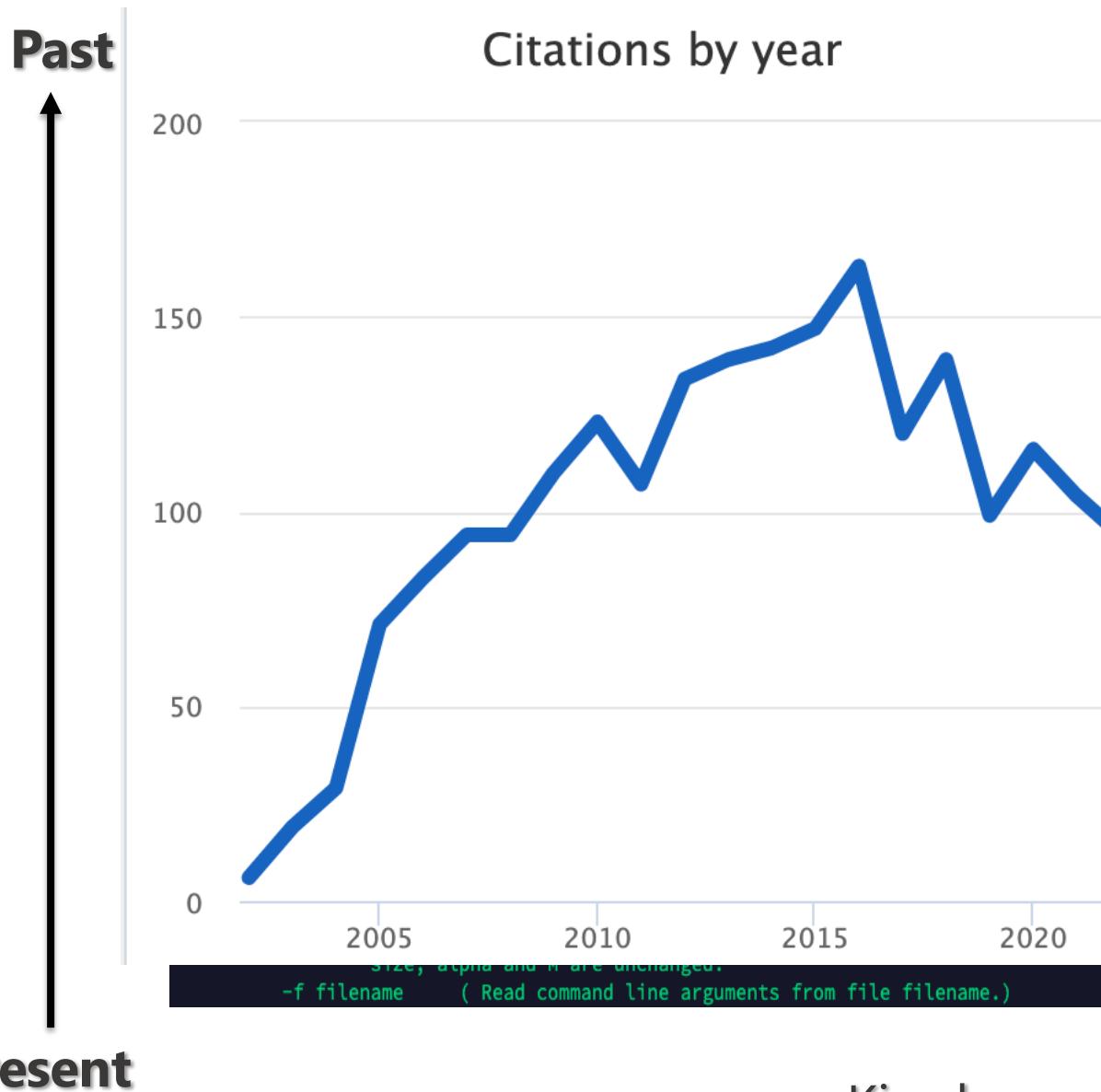
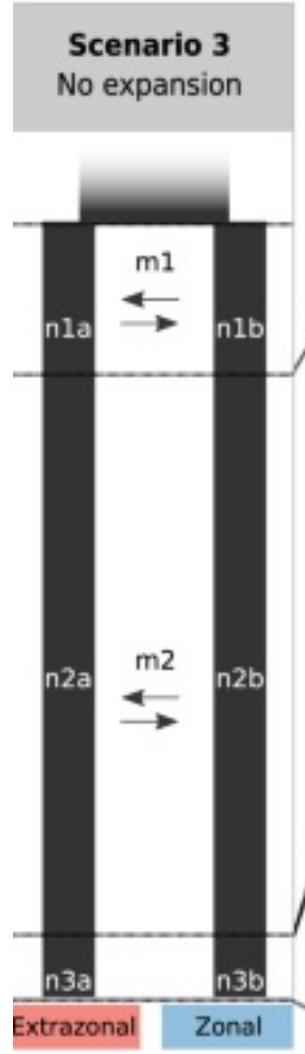
Simulate Genetic Data – ms (Hudson 2002)



A lot of options!
(not very human friendly)

- **PopPlanner** can help

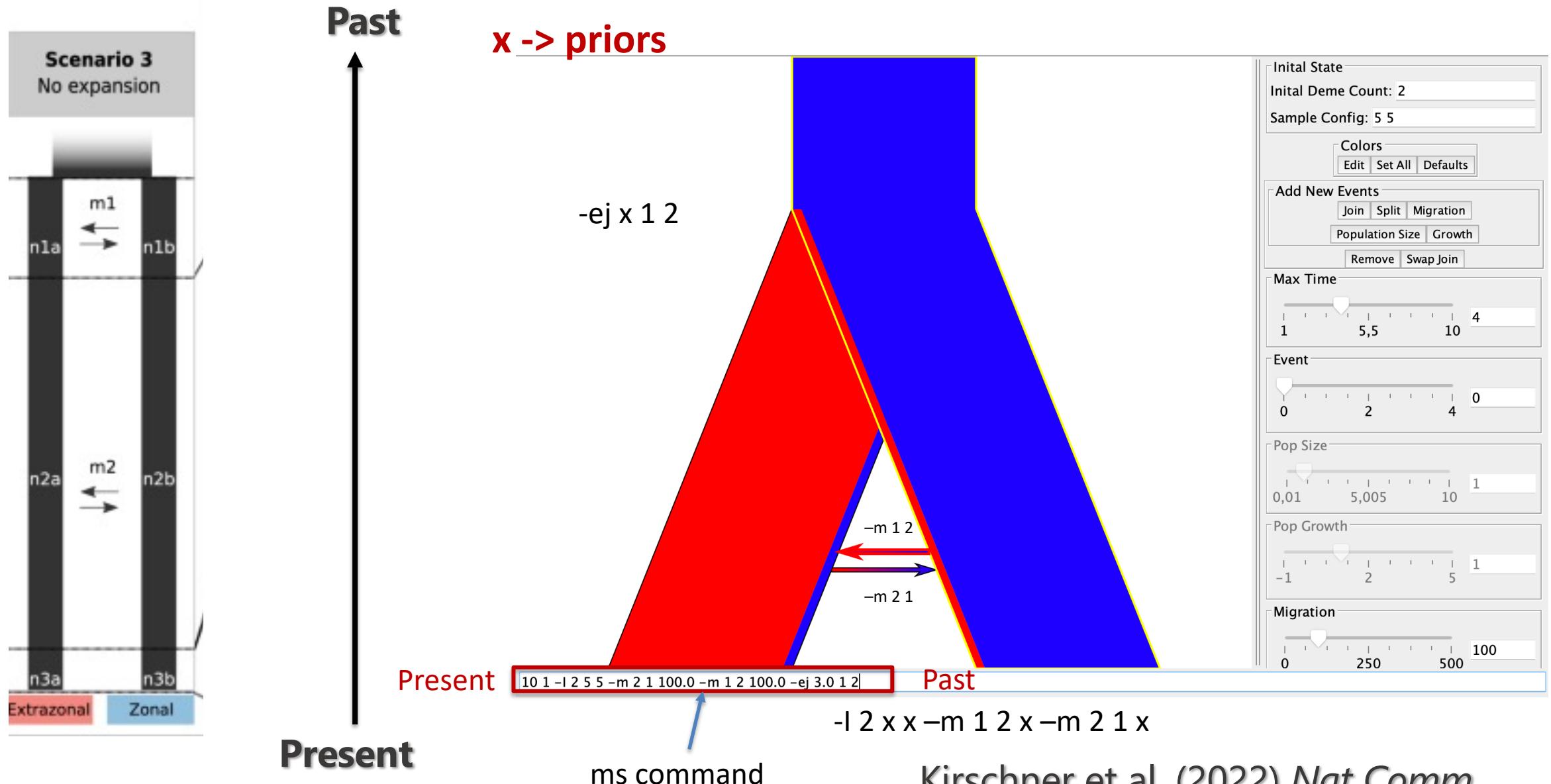
Simulate Genetic Data – ms (Hudson 2002)



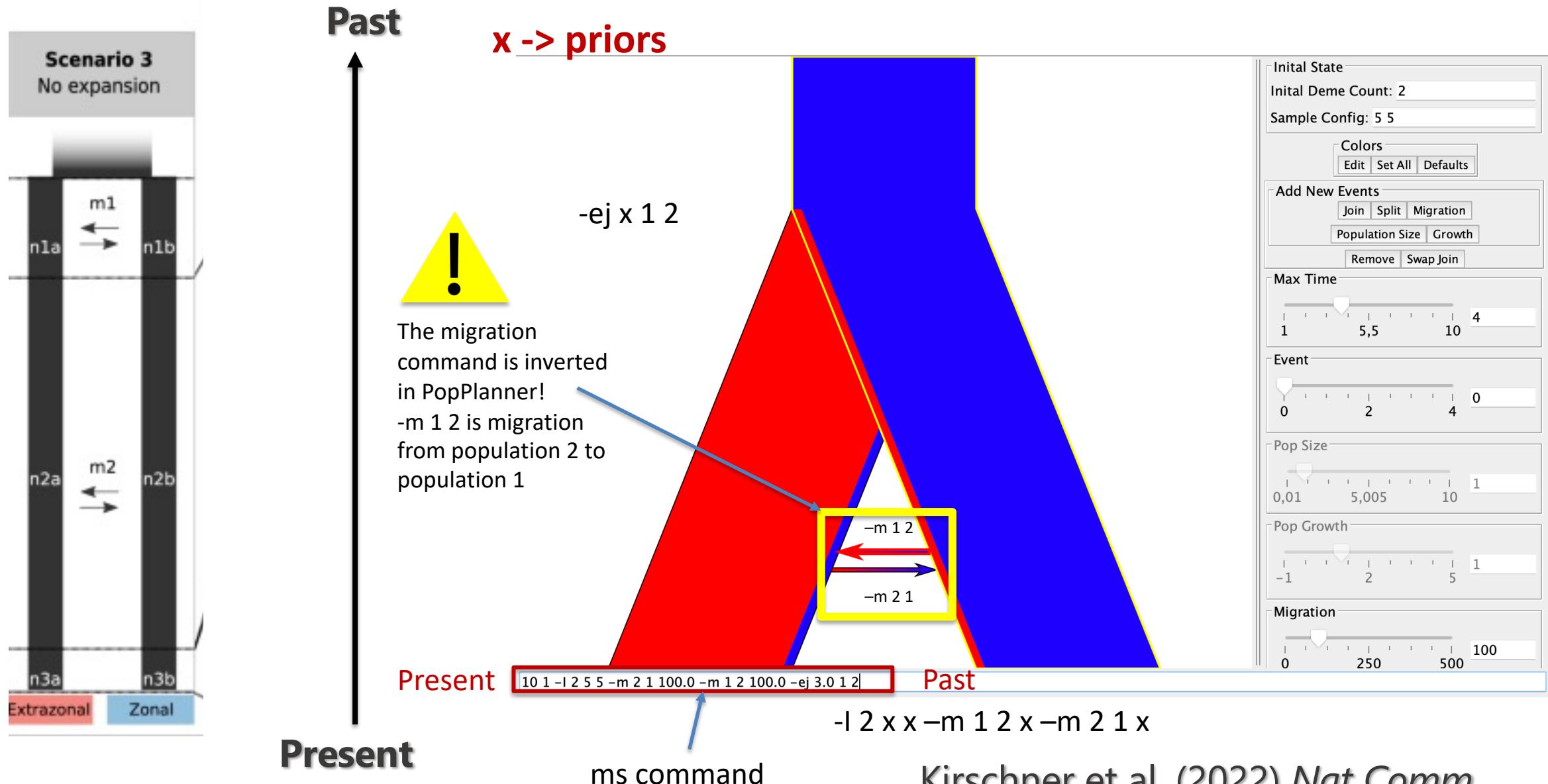
A lot of options!
(not very human friendly)

- **PopPlanner** can help

Simulate Genetic Data – ms (Hudson 2002)



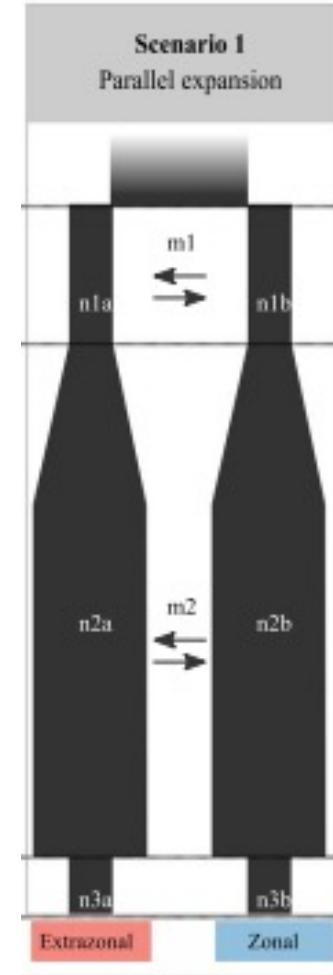
Simulate Genetic Data – ms (Hudson 2002)



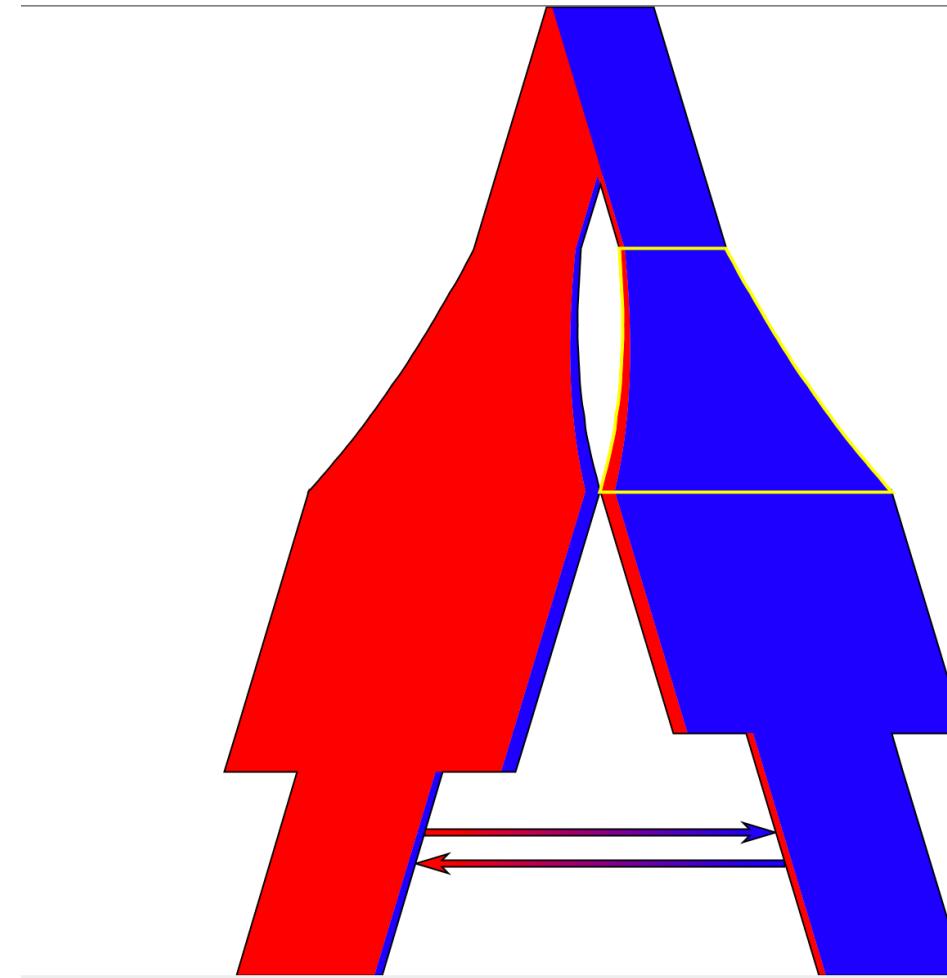
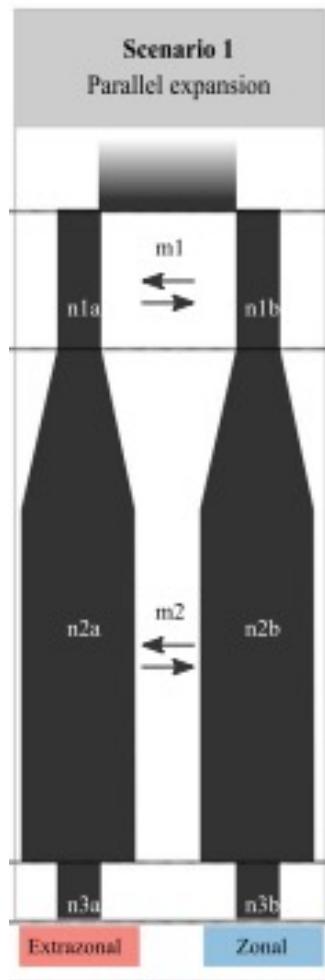
Simulate Genetic Data – ms (Hudson 2002)

- **Practical Exercise 1:**

- Explore all options for demographic events in PopPlanner.
- Try to build scenario 1 (the most complex) in PopPlanner using a single migration rate starting in the present.
- Build possible scenarios for the species you are working on. Export your scenarios as PNG images.
- Discuss your models and your thoughts/difficulties in groups.



Simulate Genetic Data – ms (Hudson 2002)



Initial State
Initial Deme Count: 2

Sample Config: 5 5

Colors
Edit Set All Defaults

Add New Events
Join Split Migration
Population Size Growth
Remove Swap Join

Max Time
1 5,5 10 4

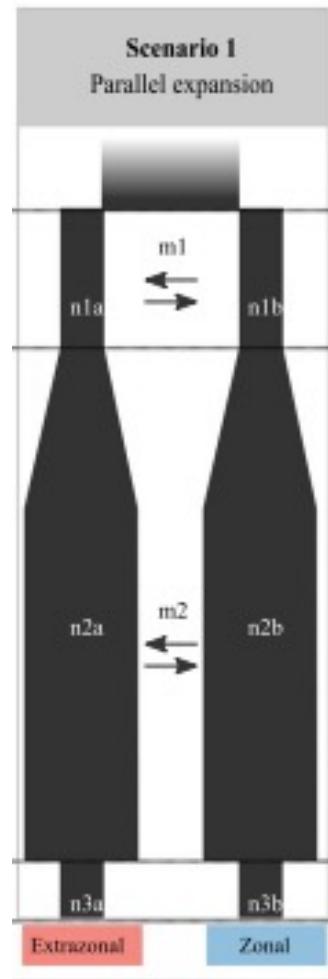
Event
0 2 4 0

Pop Size
0,01 5,005 10 2

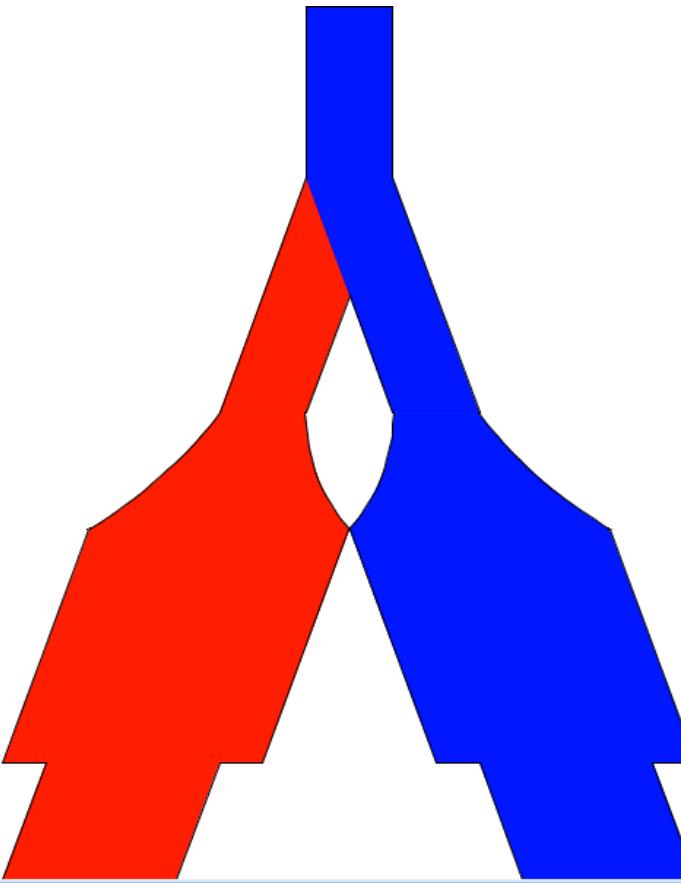
Pop Growth
-1 2 5 1

Migration
0 250 500 100

Simulate Genetic Data – ms (Hudson 2002)



x → priors



10 1 -l 2 5 5 -en 0.5 1 1.5 -en 0.5 2 1.5 -eg 1.5 1 2.222 -eg 1.501745 2 2.258 -en 2.0 2 0.5 -en 2.0 1 0.5 -ej 3.0 1 2

Initial State
Initial Deme Count: 2
Sample Config: 5 5

Colors
Edit Set All Defaults

Add New Events
Join Split Migration
Population Size Growth
Remove Swap Join

Max Time
1 5,5 10 3,745

Event
0 1,873 3,745 1,5

Pop Size
0,01 5,005 10 1

Pop Growth
-1 2 5 2,222

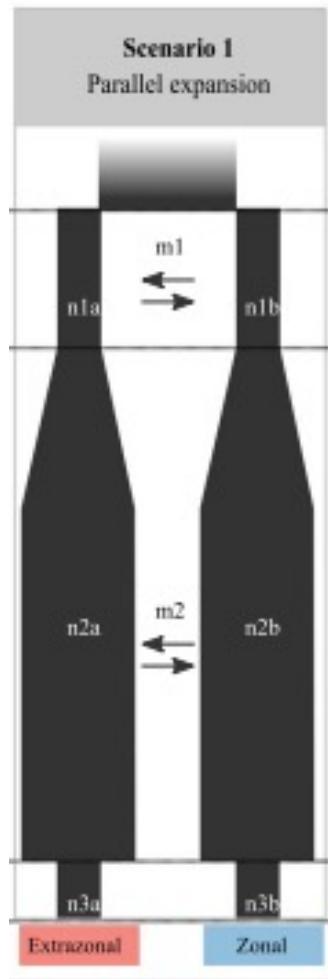
Migration
0 250 500 1

Seg.
reps sites 2 pops Size change Size growth stop growth stop divergence

Present . /ms x 1000 -s 1 -t x -l 2 x x -en x 1 x -en x 2 x -eg x 1 x -eg x 2 x -eg x 1 0 -eg x 2 0 -ej x 1 2 Past

theta

Simulate Genetic Data – ms (Hudson 2002)



Past

divergence

-ej x 1 2

Stop
growth

-eg x 1 0 -eg x 2 0

growth

-eg x 1 x -eg x 2 x

Size
change

-en x 1 x -en x 2 x

2 pops

-l 2 x x

Present

Initial State
Initial Deme Count: 2
Sample Config: 5 5

Colors
Edit Set All Defaults

Add New Events
Join Split Migration
Population Size Growth
Remove Swap Join

Max Time
1 5,5 10 3,745

Event
0 1,873 3,745 1,5

Pop Size
0,01 5,005 10 1

Pop Growth
-1 2 5 2,222

Migration
0 250 500 1

10 1 -l 2 5 5 -en 0.5 1 1.5 -en 0.5 2 1.5 -eg 1.5 1 2.222 -eg 1.5 01745 2 2.258 -en 2.0 2 0.5 -en 2.0 1 0.5 -ej 3.0 1 2

Simulating Genetic Data - Priors

Choosing the **Priors** is a very important step to guide simulations
(Biological meaningful values).

Information from previous work can be used (Ne, divergence time, migration rates).

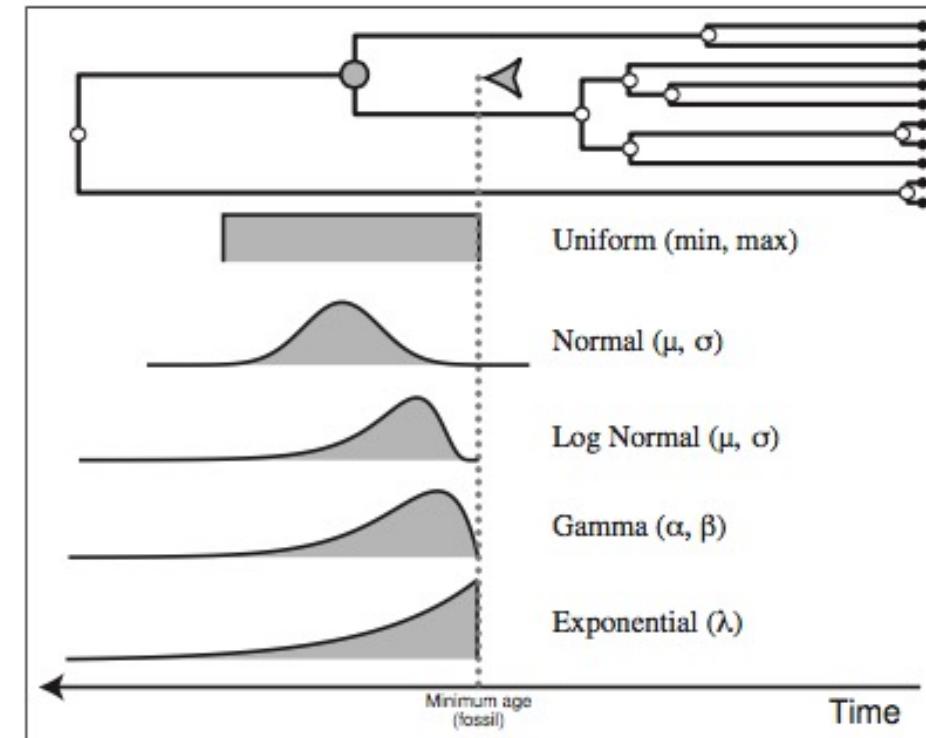
Use **broad intervals** can be a good start – **informativeness of the data.**

Simulating Genetic Data - Priors

The priors are sampled from a distribution containing all possible values

Sufficient simulations must be conducted

Selecting an informative distribution is also important



Simulating Genetic Data - Priors

Which priors are mandatory for my species?

- mutation rate per generation

$$\theta = 4N_e\mu$$

- Effective population size

- generation time

➤ Times in ms are in units of $4N_e$ generations:

- divergence times

=Time in years/(generation*4*N_e)

Simulating Genetic Data

- **Practical Exercise 2:**
- Use PopPlanner to conceive scenarios with 30 samples divided into either one, two or three populations.
- Visualize at least 2 simulations from each scenario. For that, copy the command and add to the end of it: -T >> trees.tre
- Open the created file (or export it to the terminal with the command "cat trees.tre" and copy the trees in FigTree.
- Are the trees in each simulation and in each scenario visually different? How? Now change the Θ values and see how the number of segregating sites behaves.
- Discuss your results and your thoughts/difficulties in groups.

Part III: Script to simulate genetic data under demographic scenarios

```
#!/usr/bin/python3
## in order to use this code you have to
## ms can be freely downloaded from:
## http://home.uchicago.edu/rhudson1/source/mksam
## import all required modules.
## import random
import os
import math
import shlex, subprocess
import numpy as np
##define a function to read ms's simulations and transform them into a NumPy
def ms2numpy(xfile):
    g = list(xfile)
    k = [idx for idx,i in enumerate(g) if i.startswith('k')]
    f = []
    for i in k:
        L = g[i+A:i+nDNAnsam+A]
        q = []
        for j in L:
            i = int(j)
            array(i)
            q.append(array(i))
        f.append(q)
    return f
```

Simulating Genetic Data - Script

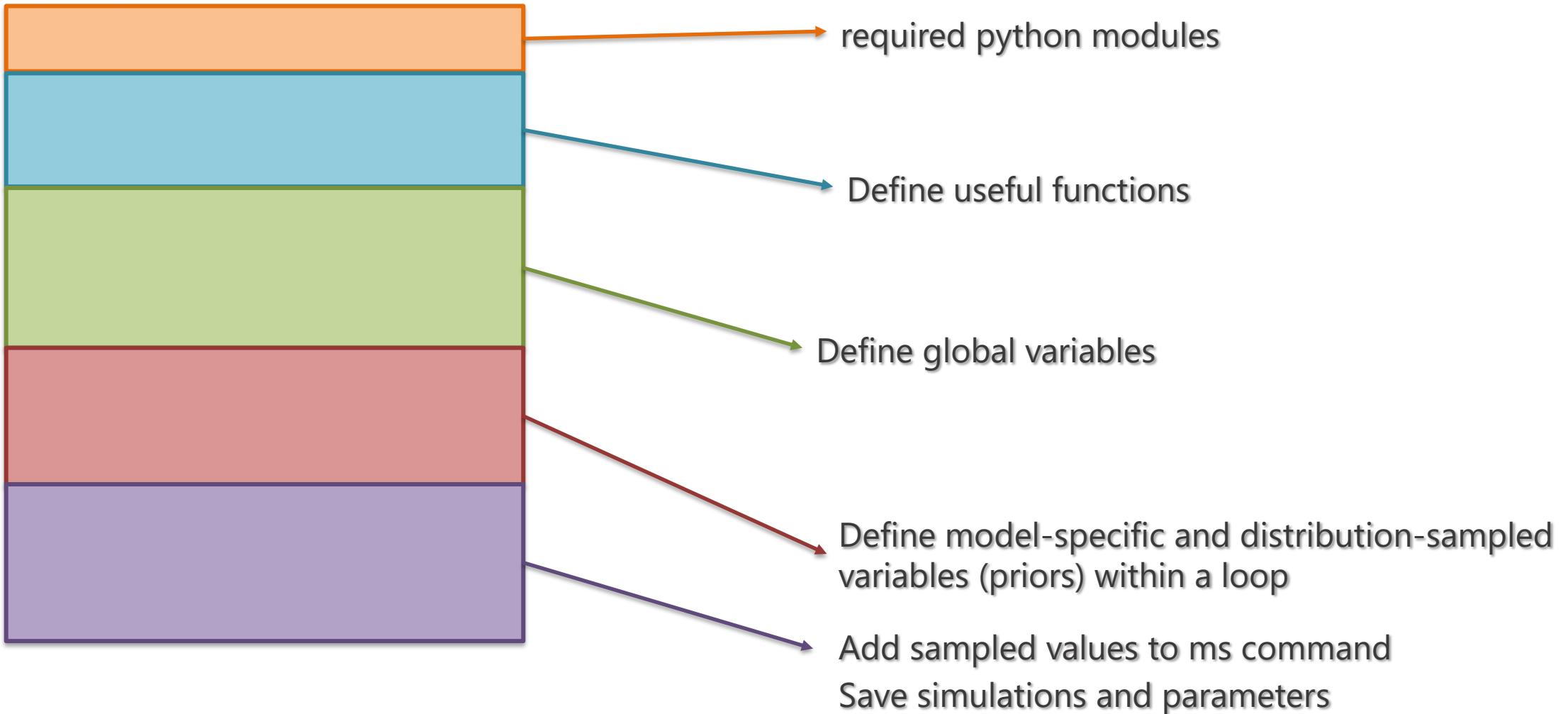
What do we need for simulating demographic scenarios?

Simulating Genetic Data - Script

What do we need for simulating demographic scenarios?

- Define scenarios.
- Define priors and parameters matching our empirical dataset.
- Repeat simulations several times with different prior values (at least 10,000 per model for Deep Learning).
- Save simulations and parameters in a proper format for Deep Learning.

Simulating Genetic Data - Script



Simulating Genetic Data - Script

```
#!/usr/bin/python3

## in order to use this code you have to have ms installed on your computer
## ms can be freely downloaded from:
## http://home.uchicago.edu/rhudson1/source/mksamples.html

## import all required modules
import random
import os
import math
import shlex, subprocess
import numpy as np

##define a function to read ms' simulations and transform then into a NumPy array.
def ms2nparray(xfile):
    g = list(xfile)
    k = [idx for idx,i in enumerate(g) if len(i) > 0 and i.startswith(b'//')]
    f = []
    for i in k:
        L = g[i+4:i+nDNANsam+4]
        q = []
        for i in L:
            i = i = [int(j) for j in list(i.decode('utf-8'))]
            i = np.array(i)
            q.append(i)
        q = np.array(q)
        q = q.astype("int8")
        f.append(np.array(q))
    return f

### variable declarations.

#define the number of simulations.
Priorsize = 10000

## sample size for Extrazonal.
nExZ = 160
## sample size for Zonal.
nZ = 110
## Combines sample size.
nDNANsam = nExZ + nZ

##mutation rate.
mutrate = 7.0E-9

simModel1 = []
simModel2 = []
simModel3 = []

## create a file to store parameters and one to store the models.
parameters1 = open("parameters1.txt","w")
parameters2 = open("parameters2.txt","w")
parameters3 = open("parameters3.txt","w")
```

required python modules

Define a function to transform simulations in the proper format

Define general variables (fixed)

- Number of simulations
- Sample sizes (number of alleles per loci)
- Mutation rate

Simulating Genetic Data - Script

Simulate models

Define model-specific and distribution-sampled variables (priors)

- Generation length
 - Ne
 - used to calculate $\Theta = 4 * N_e * \mu$
 - Times
 - Coalescent units (4Ne generations): $\text{coalTime} = \text{Time} / (\text{generation} * 4 * N_e)$
 - Relative Population sizes
 - Growth
 - Migration

Add sampled values to ms command
Save simulations and parameters

Simulating Genetic Data - Script

Examples of ms commands from the script

- Scenario 1:

```
./ms 270 1000 -s 1 -t 0.280081 -l 2 160 110 -n 2 1.875624 -en 0.001289 1 35.544203 -en 0.001289 2 79.531376 -em  
0.001289 1 2 1.263896 -em 0.001289 2 1 1.919980 -eg 0.017061 1 38.360445 -eg 0.017061 2 43.970864 -em  
0.017061 1 2 0.301733 -em 0.017061 2 1 3.967323 -eg 0.127240 1 0 -eg 0.127240 2 0 -ej 0.293251 1 2
```

- Scenario 2:

```
./ms 270 1000 -s 1 -t 0.197474 -l 2 160 110 -n 2 1.881221 -en 0.000138 2 13.295751 -em 0.000138 1 2 3.076617 -  
em 0.000138 2 1 3.641901 -eg 0.010041 2 23.017965 -em 0.010041 1 2 2.368179 -em 0.010041 2 1 0.699033 -eg  
0.130037 2 0 -ej 0.756109 1 2
```

Scenario 3:

```
./ms 270 1000 -s 1 -t 0.388039 -l 2 160 110 -n 2 1.378919 -em 0.000606 1 2 3.543538 -em 0.000606 2 1 1.884716 -  
em 0.006555 1 2 1.994695 -em 0.006555 2 1 3.124491 -ej 0.237054 1 2
```

Simulating Genetic Data - Script

- **Practical Exercise 3:**
- Copy the commands in PopPlanner and try to visualize what is being simulated (you can change the Max Time if you need).
- Is it difficult to see the scenario in PopPlanner? Do you have any thoughts about why simulations show like that?
- Now visualize the segregating sites and the trees using -T >> trees.tre. Can you see a difference in the output of these simulations compared to the simulations from Exercise 2?
- Add trees from different scenarios into FigTree and see if you can see any similarities or differences.