

ms - a program for generating samples under neutral models

Hudson (2002) Bioinformatics 18:337-8

<http://home.uchicago.edu/~rhudson1/source/mksamples.html>

ms

- Generates samples under a variety of assumptions about migration, recombination rate and population size
- Samples are generated using the standard coalescent approach:
 - 1st Generate the genealogy
 - 2nd Add mutations to the genealogical tree
- Assumes the small sample approximations of the coalescent ($n \ll N$)
- Assumes infinite sites model for mutations
 - BUT, when used in conjunction with other programs, other mutation models can be used

The ms command line

For a constant population size, with random mating, no recombination and infinite site mutation model the command is:

ms nsam nreps -t $4N_0\mu$

Number of gene
copies sampled

Number of
repetitions
(simulations)

Scaled mutation parameter:
 N_0 **diploid** population size
 μ mutation rate for the entire locus

EASY!!

The ms output

For instance, for the following command line

```
ms 10 1 -t 4 > ms.out
```

Population with effective size
1000 and mutation rate
0.001

The output is stored in the file ms.out

```
ms 10 1 -t 4
```

```
60711 7473 11225
```

→ **Command line**

→ **Random number generator seeds**

```
//
```

```
segsites: 17
```

→ **number segregating sites**

```
positions: 0.1690 0.2248 0.3120 0.3346 0.3662 0.3935 0.4243
```

```
0.4814 0.5587 0.5735 0.5753 0.6957 0.7958 0.8226 0.8943 0.9272
```

```
0.9404
```

→ **Position of segregating sites**

```
01010111000010000
```

```
01010111001010000
```

```
01011111010010000
```

```
01010111000010000
```

```
11010111000010000
```

```
01010111000110000
```

```
01010111000010000
```

```
01010111000110000
```

```
01010111000010000
```

```
00100000100001111
```

→ **Sample for 10 gene copies and 17 segregating sites**

0 indicates ancestral state

1 indicates derived state

10 gene copies (5 diploids)

17 seg sites

Population Expansion or Collapse

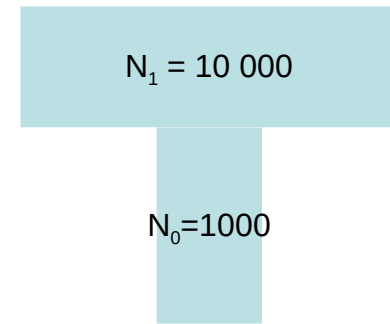
Sudden Population Collapse

ms 100 1 -t 4 -eN 0.5 10

$4N_0\mu$
Event
Change N

Scaled time
 $t=T/4N_0$

Relative N
 $x=N_1/N_0$



$T=2000$ gen
 $\mu=10^{-3}$

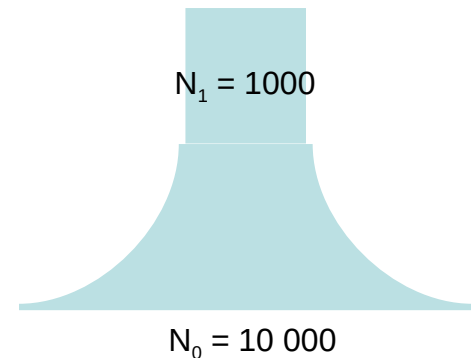
Exponentially growing population

ms 100 1 -t 40 -G 46.05 -eN 0.05 0.1

$4N_0\mu$
Set Growth rate
Positive: expansion
Negative: decrease

Growth rate α

$N_1=N_0*\exp(-\alpha t)$, where $t=T/4N_0$



$T=2000$ gen
 $\mu=10^{-3}$

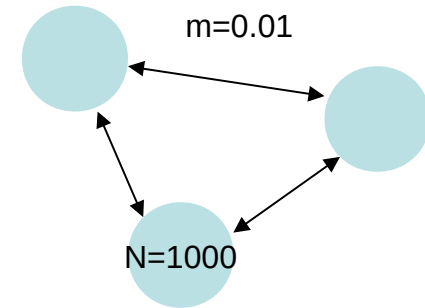
Population Structure

Island Model

ms 30 1 -t 4 -I 3 10 10 10 40

Number of pop
Sample from pop1, pop2, etc

Scaled migration rate
 $4N_0m$



Population Split

ms 40 1 -t 4 -I 2 10 30 -n 1 5

N3 as reference pop
 $N_0 = N3$

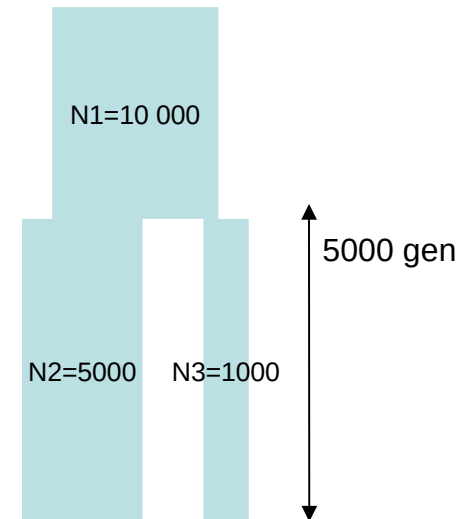
Number of present pop

Set relative N for pop1

-ej 1.25 2 1 -en 1.25 1 10

Event of joining lineages (Split)
Scaled time = $5000/4N_0 = 1.25$
All lineages of pop2 go to pop1

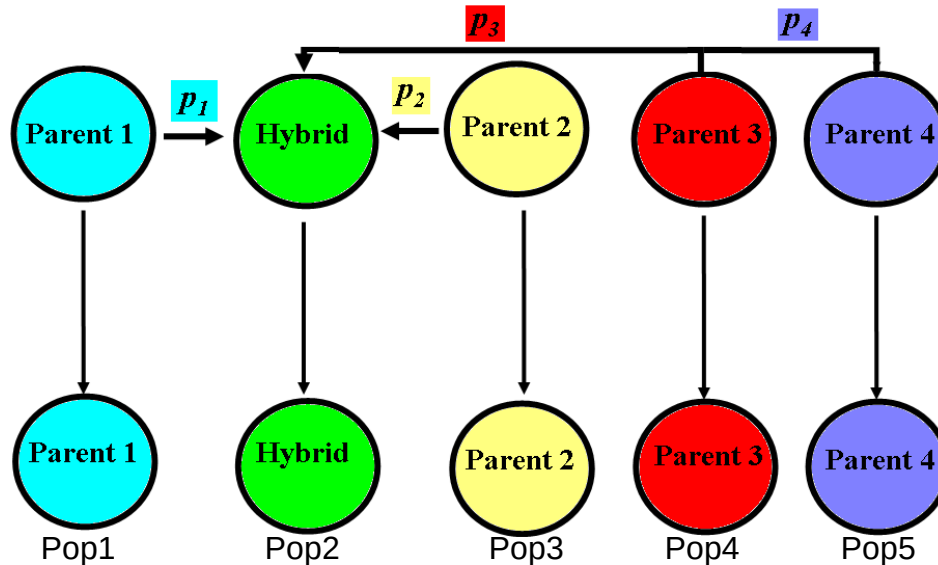
Event of changing size
Scaled time = $5000/4N_0 = 1.25$
Pop1 is 10 times N_0



Remember

- N_0 is the reference effective size, and it can be chosen arbitrarily
- Mutation rate, migration rate and time of the events are scaled by the chosen effective size N_0

One example of complex ms command



$N1=100$ $N2=100$ $Nh=100$ $N3=100$ $N4=100$

$Nanc=100$

Mutation rate=0.0001

$Tadm=100$ $Tsplit=10000$

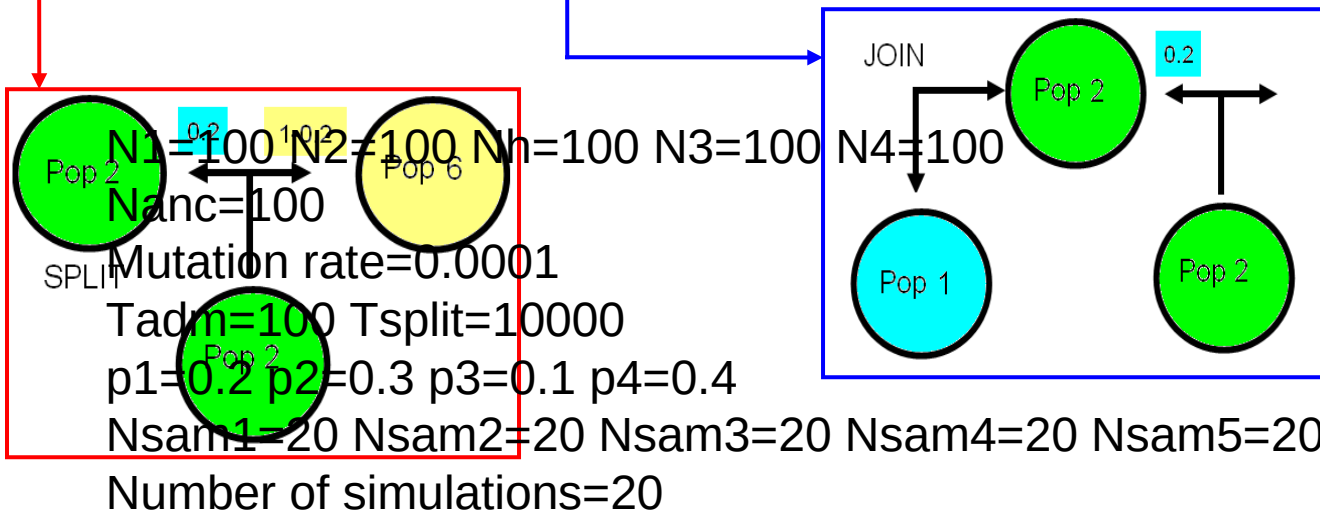
$p1=0.2$ $p2=0.3$ $p3=0.1$ $p4=0.4$

$Nsam1=20$ $Nsam2=20$ $Nsam3=20$ $Nsam4=20$ $Nsam5=20$

Number of simulations=20

One example of ms command

```
ms 100 20 -t 0.04 -l 5 20 20 20 20 20 -n 1 1 -n 2 1 -n 3 1 -n 4 1 -n 5 1
-es 0.25 2 0.2 -ej 0.25 2 1 -es 0.25 6 0.375 -ej 0.25 6 3 -es 0.25 7 0.2 -ej
0.25 7 4 -ej 0.25 8 5 -en 25 1 1 -ej 25 3 1 -ej 25 4 1 -ej 25 5 1 | microsat
> ms_result_msat.txt
```



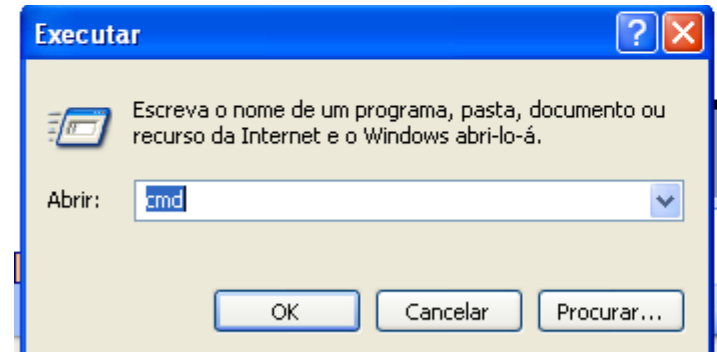
- Some issues:
 - Population sizes and time are scaled
 - long commands for complex models

Example

- Simulate a sample of 50 gene copies with 5000 nucleotides, from a constant population with effective size 10 000 and mutation rate 10^{-8} per nucleotide per generation
- Open the command line. If you are on Windows do: START>Run>cmd
- Go to the folder with ms.exe
- Compute $4N_0\mu = 4 \cdot 10000 \cdot 5000 \cdot 10^{-8} = 20$

ms 50 1 -t 20 > ms_stable.out

Have a look at the file ms_stable.out with a text editor (TextPad, WordPad, NotePad, etc)



```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows XP [Versão 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:\Documents and Settings\Proprietário>cd c:\ms
C:\ms>cd msdir
C:\ms\msdir>ms 50 1 -t 20 > ms_stable.out
C:\ms\msdir>
```