

Compartmentalization detection

Selene Zárate

Viruses and compartmentalization

- ❖ Virus infection may establish itself in a variety of the different organs within the body and can form somewhat separate viral populations, driven to adapt to their particular environments and subjected to different selective pressures
- ❖ Virus populations can become isolated from each other, if trafficking and gene flow between viral subpopulations is significantly restricted, then each subpopulation can become genetically distinct from others, i.e., compartmentalized.

Compartmentalization

- ❖ Compartmentalization has been defined in different ways:
 - ❖ as genetic heterogeneity between subpopulations
 - ❖ as the result of independent micro-evolution
 - ❖ as the result of restricted viral gene flow
 - ❖ as the presence of distinct but phylogenetically related genotypes.
- ❖ In HIV, compartmentalized viral populations have been shown to possess distinct phenotypic characteristics, such as cellular tropism, drug resistance, and level of pathogenesis.

How to determine population structure?

- ❖ A population is considered structured if:
 - ❖ genetic drift is occurring in some of its subpopulations
 - ❖ migration does not happen uniformly throughout the population
 - ❖ mating is not random throughout the population.
- ❖ A population's structure affects the extent of genetic variation and its patterns of distribution.

Standard Genetic Parameters for Population Diversity Analysis

- ❖ Allele richness (A)
- ❖ Effective number of alleles [$AE = 1 / (1 - H_E)$]
- ❖ Observed heterozygosity (H_0)
- ❖ Expected heterozygosity (H_E)
- ❖ Fixation Index (FIS)
- ❖ **Within-population gene diversity (H_I)**
- ❖ **Mean within-population gene diversity (H_s)**
- ❖ **Total diversity (H_T)**
- ❖ **Coefficient of gene differentiation among populations (G_{ST})**

Methods used to detect virus compartmentalization

- ❖ Distance-based: F_{ST} , nearest neighbor.
- ❖ Tree-based: Slatkin-Maddisson, Association Index, Correlation coefficients

Distance based methods

Wright's measure of population subdivision: F_{ST}

- ❖ Compares the mean pairwise genetic distance between two sequences sampled from different compartments to the mean distance between sequences sampled from the same compartment.
- ❖ Statistical significance is derived via a population-structure randomization test.

F_{ST} score

$$F_{ST} = \frac{\Pi_{\text{Between}} - \Pi_{\text{Within}}}{\Pi_{\text{Between}}}$$

When the differences between compartments is much larger than the differences within compartments, the values of F_{ST} approaches 1. Therefore values of F_{ST} close to 1 indicate compartmentalization



Distance within subpopulation = $14/36 = 0.39$

Distance between subpopulation = $53/81 = 0.65$

$$F_{ST} = (0.65 - 0.39) / 0.65 = 0.4$$

Nearest-neighbor statistic (S_{nn})

- ❖ Is a measure of how often the nearest neighbors of each sequence were isolated from the same or different compartments. The distance between sequences is measured using the TN93 metric (not the number of sites in which two sequences differ, as in the original description).

$$S_{nn} = \sum_{j=1}^n X_j / n.$$

- ❖ where X_j is 1 if the nearest neighbor was isolated from the same subpopulation or 0 otherwise.



$$S_{nn} = \sum_{j=1}^n X_j / n.$$

Sequence A was isolated from one subpopulation, 6 of its nearest neighbors are from the same population and 2 are not, its contribution to S_{nn} is $6/9$

$$S_{nn} = (14(6/9) + 4(1/9)) / 18 = 0.54$$

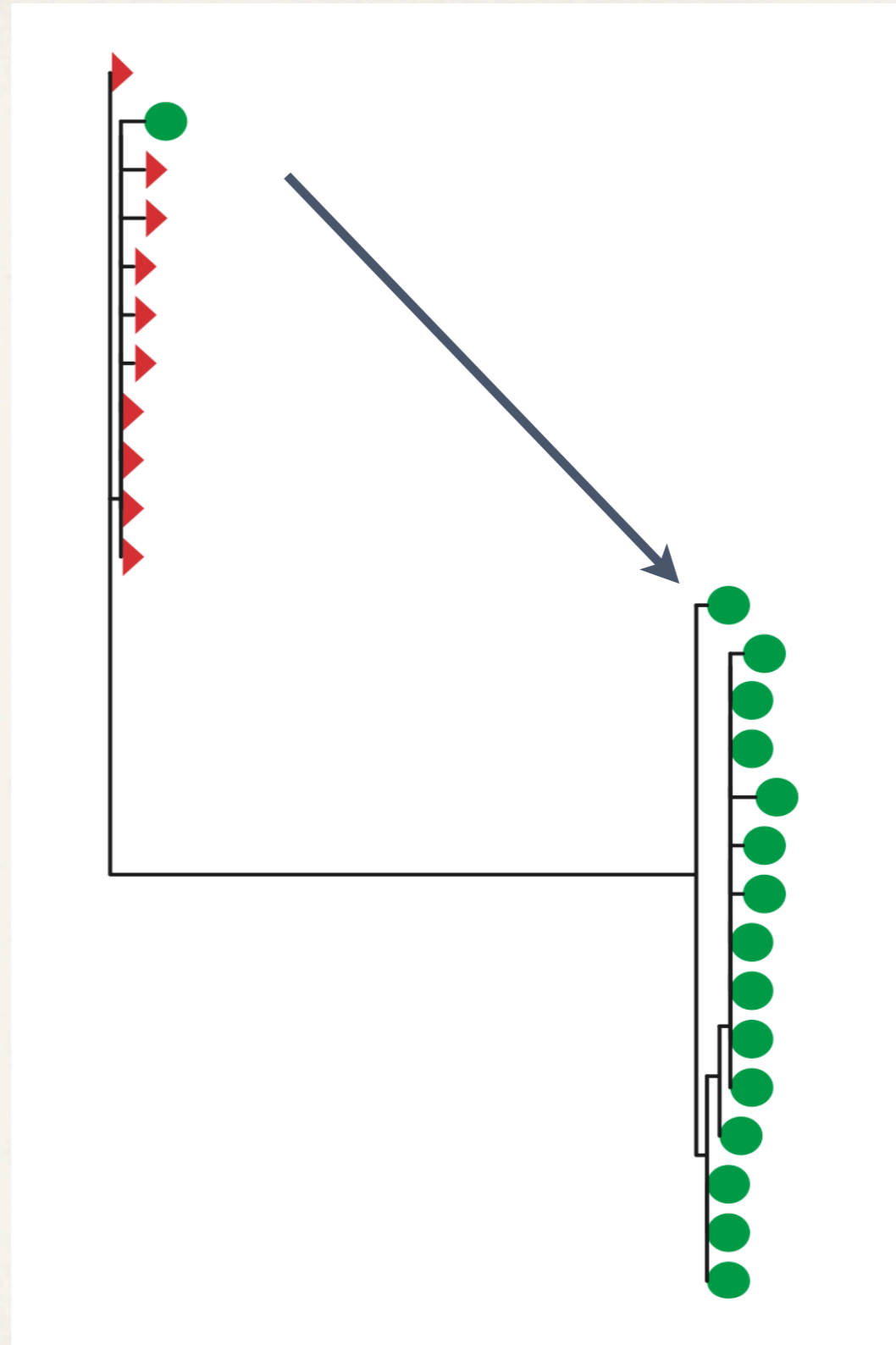
Tree-based methods

Slatkin-Maddison (SM)

- ❖ Determines the minimum number of **migration events** between the separated populations consistent with the structure of the reconstructed phylogenetic tree.
- ❖ Statistical support is based on the number of migration events that would be expected in a randomly structured population, derived by permuting sequences between compartments.

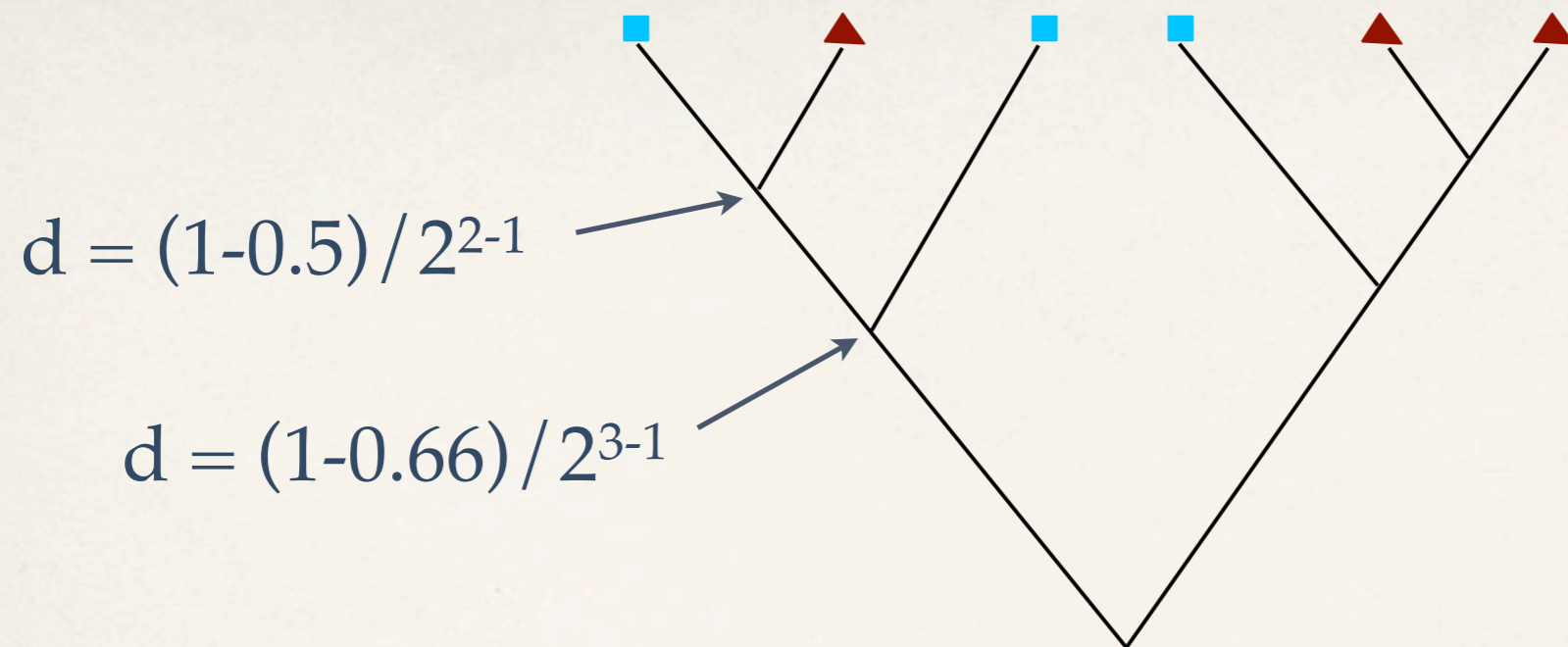
In the phylogeny shown here, one migration event explains the distribution of the sequences in the topology.

The more migration events needed to explain the distribution of sequences the compartmentalization hypothesis becomes less likely



Simmonds association index (AI)

- ❖ Assesses the degree of population structure in the phylogenetic tree by weighting the contribution of each internal node based on its depth in the tree (progressively less for nodes near the root) and evaluating the significance of the observed value using a bootstrap sample both over the structure of the population and the shape of the phylogenetic tree.



At each node determine the number of sequences below it (n), and the frequency of the most frequent variant (f), and calculate

$$d = (1-f) / 2^{n-1}.$$

The AI is calculated as the ratio between the mean score of 100 bootstrap replicates, and the mean of 10 sample reassigned controls

The smaller the score, the more likely the population is compartmentalized

Correlation coefficients (r , r_b)

- ❖ Correlation coefficients are a way to correlate distances between two sequences in a phylogenetic tree with the information about whether or not they were isolated from the same compartment.
- ❖ The distance between two sequences can be either the number of tree branches separating the sequences (r_b) or the cumulative genetic distance between the sequences (r).
- ❖ To assess whether the computed coefficient was statistically significant, we estimated the distribution of these coefficients by permuting sequences between compartments. A P value of 0.05 or less was considered statistically significant.

How these methods compare

$$p_0 = f_{yy} + f_{nn}$$

$$p_e = (f_{yy} + f_{yn}) * (f_{yy} + f_{ny}) \\ + (f_{nn} + f_{ny}) * (f_{nn} + f_{yn})$$

$$\kappa = (p_0 - p_e) / (1 - p_e)$$

TABLE 2. Levels of agreement between methods as measured by pairwise κ scores

Comparison	κ		
	CNS	FGT ^a	Simulations
Different-class methods			
SM vs F_{ST}	0.62	0.35	0.46
SM vs S_{nn}	0.66	0.52	0.77
SM vs AMOVA	0.09	0.14	0.36
AI vs F_{ST}	0.56	0.29	0.44
AI vs S_{nn}	0.58	0.44	0.67
AI vs AMOVA	0.05	0.17	0.35
Same-class methods			
SM vs AI	0.48	0.44	0.67
F_{ST} vs S_{nn}	0.48	0.35	0.43
F_{ST} vs AMOVA	0.03	0.11	0.80
S_{nn} vs AMOVA	0.05	0.14	0.43
r vs r_b	0.67	0.66	0.68

^a FGT, female genital tract.

Biased sample sizes

TABLE 5. Proportion of simulated data sets classified as compartmentalized when equal and different sample sizes are drawn from the compartments^a

Method	Proportion classified as compartmentalized when sample sizes were:	
	Equal ($n = 20$)	Skewed ($n = 5$ and 20)
SM	0.98	0.27
F_{ST}	0.55	0.33
S_{nn}	0.99	0.59
AI	0.85	0.37
AMOVA	0.41	0.24
r^b	0.54	0.04
r_b^c	0.71	0.19

^a A migration rate of 0.0005 migrations per generation was used to simulate both data sets.

^b r , correlation coefficient by length of branches.

^c r_b , correlation coefficient by number of branches.

A tour around HYPHY

A quick example:

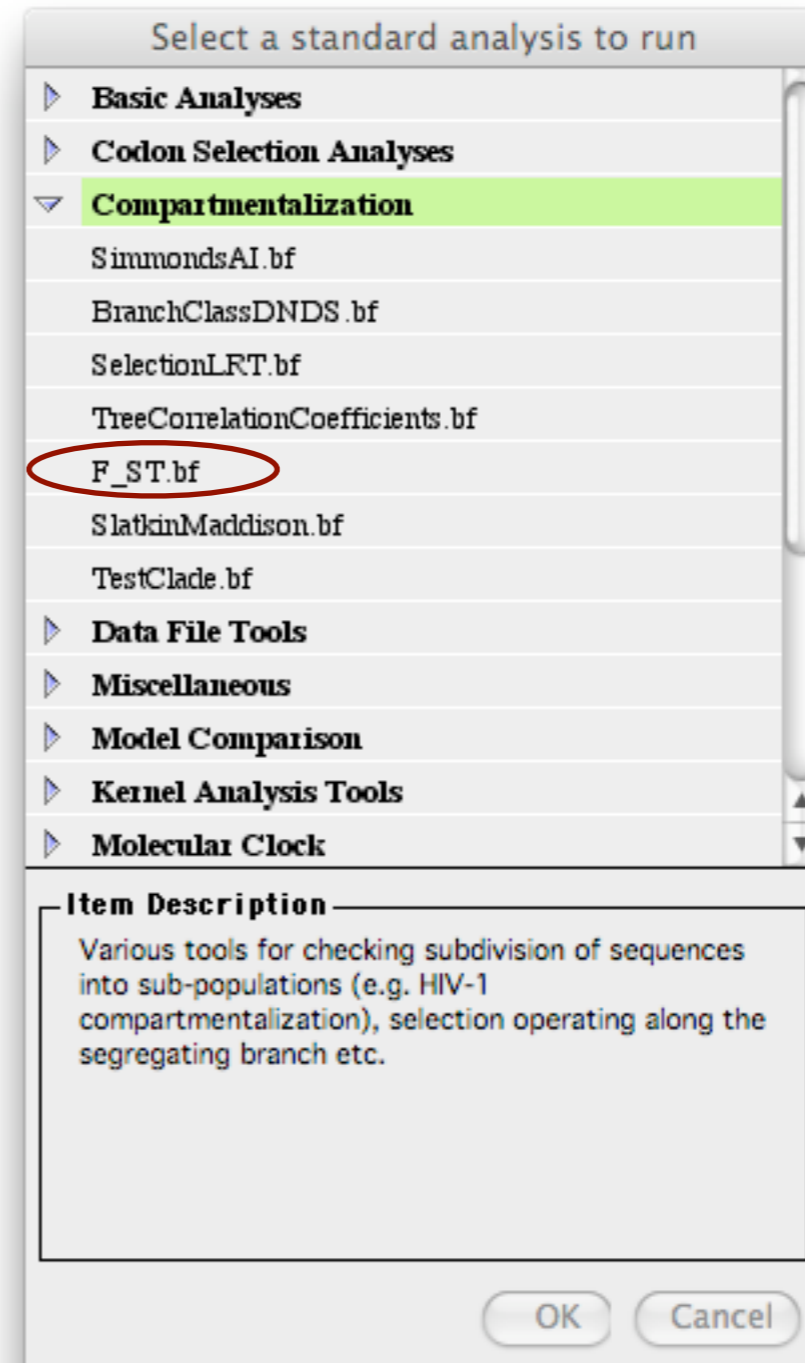
- ❖ Lets take two sets of data, one group of HIV sequences* derived from either plasma or female genital tract (A), and a second data set⁺, with samples derived from plasma or CSF (B).
- ❖ We aligned the sequences and reconstructed the phylogeny in order to carry out the compartmentalization analysis

*Kemal, PNAS:100(22).

⁺Gatanaga, Arch. Virol. :144(1).

Starting an analysis

- ❖ Open the “Standard analysis” menu and select **Compartmentalization**.
- ❖ We will start with an F_{ST} analysis.
- ❖ You will need:
 - ❖ A sequence alignment
 - ❖ A distinct label in the sequence name for each compartment
 - ❖ A substitution model



Distance Computation

Distance formulae

Full likelihood

Load Matrix

Item Description

Use one of the predefined distance measures based on data comparisons. Fast.

1 required (1 chosen).

OK

Cancel

Distance Computation

Distance formulae

Full likelihood

Load Matrix

Item Description

Use one of the predefined distance measures based on data comparisons. Fast.

1 required (1 chosen).

OK

Cancel

Data type

Nucleotide/Protein

Codon

Item Description

Nucleotide or amino-acid (protein).

1 required (1 chosen).

OK

Cancel

Distance Computation

- Distance formulae
- Full likelihood
- Load Matrix

Item Description

Use one of the predefined distance measures based on data comparisons. Fast.

1 required (1 chosen). OK Cancel

Data type

- Nucleotide/Protein
- Codon

Item Description

Nucleotide or amino-acid (protein).

1 required (1 chosen)

Please choose a nucleotide or amino-acid data file:

Kemal2003 search

Name	Date Modified
WC27_aln	Jul 31, 2006 12:47 PM
WC26_aln	Jul 31, 2006 12:47 PM
WC25_aln	Jul 31, 2006 12:47 PM
WC23_aln	Jul 31, 2006 12:47 PM
WC22_aln	Jul 31, 2006 12:47 PM
WC21_aln	Jul 31, 2006 12:47 PM
WC19_aln	Jul 31, 2006 12:47 PM
WC18_aln	Jul 31, 2006 12:47 PM
WC17_aln	Jul 31, 2006 12:47 PM
WC16_aln	Jul 31, 2006 12:47 PM
WC15_aln	Jul 31, 2006 12:47 PM
WC28_tree	Feb 14, 2006 1:54 PM
WC27_tree	Feb 14, 2006 1:53 PM

Enable: All Documents

Cancel Open

Read the following data:9 species:{WC15_cv1,WC15_cv1_2,WC15_cv1_3,WC15_cv1_4,WC15_cv1_5,WC15_pl,WC15_pl_2,WC15_pl_3,WC15_pl_4};
Total Sites:1269;
Distinct Sites:41

Enter a regular expression to define the first clade:

Read the following data:9 species:{WC15_cv1,WC15_cv1_2,WC15_cv1_3,WC15_cv1_4,WC15_cv1_5,WC15_pl,WC15_pl_2,WC15_pl_3,WC15_pl_4};
Total Sites:1269;
Distinct Sites:41

Enter a regular expression to define the first clade:

Enter a regular expression to define the first clade:cv1

Enter a regular expression to define the second clade:pl

Clade 1 includes 5 sequences:

WC15_cv1
WC15_cv1_2
WC15_cv1_3
WC15_cv1_4
WC15_cv1_5

Clade 2 includes 4 sequences:

WC15_pl
WC15_pl_2
WC15_pl_3
WC15_pl_4

Is this partitioning correct (y/n)y

Proportion of sequence in population 1: 0.555556

Proportion of sequence in population 2: 0.444444

Read the following data: 9 species: {WC15_cv1, WC15_cv1_2, WC15_cv1_3, WC15_cv1_4, WC15_cv1_5, WC15_pl, WC15_pl_2, WC15_pl_3, WC15_pl_4};
Total Sites: 1269;
Distinct Sites: 41

Enter a regular expression to define the first clade:

Enter a regular expression to define the first clade: cv1

Enter a regular expression to define the second clade: pl

Clade 1 includes 5 sequences:

WC15_cv1
WC15_cv1_2
WC15_cv1_3
WC15_cv1_4
WC15_cv1_5

Clade 2 includes 4 sequences:

WC15_pl
WC15_pl_2
WC15_pl_3
WC15_pl_4

Is this partitioning correct (y/n)y

Proportion of sequence in population 1: 0.555556

Proportion of sequence in population 2: 0.444444

Nucleotide based distance formula.

JC69
K2P
K2P_RV
p_Distance
T3P
TN84
TN93
TN93_RV
Unaligned_LZ
Unaligned_LZ_FR

Item Description

Tamura-Nei (93) distance (unequal character frequencies, A->G, C->T and transversional bias corrections).

1 required (1 chosen).

F_{ST} and S_{nn}

- ❖ Population characteristics:
 - ❖ Metapopulation diversity (π_T)
 - ❖ Mean subpopulation diversity (π_S)
 - ❖ Mean interpopulation diversity (π_B)
- ❖ F_{ST}
 - ❖ Hudson, Slatkin and Madison (Genetics 132:583-589)
 - ❖ Slatkin (Evolution 47:264-279)
 - ❖ Hudson, Boos and Kaplan (Mol Bio Evol 9: 138-151)
 - ❖ Hudson (S_{nn}) (Genetics 155:2011-14):

Data set A

```
Population characterisitcs:  
Metapopulation diversity (pi_T)      = 0.0222394  
Mean subpopulation diversity (pi_S)  = 0.00609144  
Mean interpopulation diversity (pi_B) = 0.0387911  
  
F_ST  
  
Hudson, Slatkin and Madison (Genetics 132:583-589): 0.842968  
Slatkin (Evolution 47:264-279) : 0.728561  
Hudson, Boos and Kaplan (Mol Bio Evol 9: 138-151) : 0.726097  
Hudson (S_nn) (Genetics 155:2011-14): 1
```

Data set B

```
Population characterisitcs:  
Metapopulation diversity (pi_T)      = 0.0477232  
Mean subpopulation diversity (pi_S)  = 0.0445892  
Mean interpopulation diversity (pi_B) = 0.051275  
  
F_ST  
  
Hudson, Slatkin and Madison (Genetics 132:583-589): 0.13039  
Slatkin (Evolution 47:264-279) : 0.0697419  
Hudson, Boos and Kaplan (Mol Bio Evol 9: 138-151) : 0.0656693  
Hudson (S_nn) (Genetics 155:2011-14): 0.75
```

Data set A

```
Population characterisitcs:  
Metapopulation diversity (pi_T)      = 0.0222394  
Mean subpopulation diversity (pi_S)  = 0.00609144  
Mean interpopulation diversity (pi_B) = 0.0387911  
  
F_ST  
  
Hudson, Slatkin and Madison (Genetics 132:583-589): 0.842968  
Slatkin (Evolution 47:264-279) : 0.728561  
Hudson, Boos and Kaplan (Mol Bio Evol 9: 138-151) : 0.726097  
Hudson (S_nn) (Genetics 155:2011-14): 1
```

Bootstrap Estimators

Skip
Sure

Item Description
Resample with replacement within populations to estimate sampling properties of the estimators.

1 required (1 chosen).

Data set B

```
Population characterisitcs:  
Metapopulation diversity (pi_T)      = 0.0477232  
Mean subpopulation diversity (pi_S)  = 0.0445892  
Mean interpopulation diversity (pi_B) = 0.051275  
  
F_ST  
  
Hudson, Slatkin and Madison (Genetics 132:583-589): 0.13039  
Slatkin (Evolution 47:264-279) : 0.0697419  
Hudson, Boos and Kaplan (Mol Bio Evol 9: 138-151) : 0.0656693  
Hudson (S_nn) (Genetics 155:2011-14): 0.75
```


Data set A

Bootstrapped estimator statistics.

Hudson, Slatkin and Madison (Genetics 132:583-589)

Observed value : 0.843
Bootst. mean : 0.877
Bootst. median : 0.875
Bootst. st. dev.: 0.021
Bootst. 95% CI : 0.844 - 0.924

Slatkin (Evolution 47:264-279)

Observed value : 0.729
Bootst. mean : 0.782
Bootst. median : 0.778
Bootst. st. dev.: 0.033
Bootst. 95% CI : 0.730 - 0.859

Hudson, Boos and Kaplan (Mol Bio Evol 9: 138-151)

Observed value : 0.726
Bootst. mean : 0.780
Bootst. median : 0.776
Bootst. st. dev.: 0.034
Bootst. 95% CI : 0.727 - 0.857

Hudson (S_{nn}) (Genetics 155:2011-14)

Observed value : 1.000
Bootst. mean : 1.000
Bootst. median : 1.000
Bootst. st. dev.: 0.000
Bootst. 95% CI : 1.000 - 1.000

Data set B

Bootstrapped estimator statistics.

Hudson, Slatkin and Madison (Genetics 132:583-589)

Observed value : 0.130
Bootst. mean : 0.221
Bootst. median : 0.209
Bootst. st. dev.: 0.185
Bootst. 95% CI : -0.090 - 0.556

Slatkin (Evolution 47:264-279)

Observed value : 0.070
Bootst. mean : 0.137
Bootst. median : 0.117
Bootst. st. dev.: 0.122
Bootst. 95% CI : -0.043 - 0.385

Hudson, Boos and Kaplan (Mol Bio Evol 9: 138-151)

Observed value : 0.066
Bootst. mean : 0.130
Bootst. median : 0.110
Bootst. st. dev.: 0.117
Bootst. 95% CI : -0.040 - 0.370

Hudson (S_{nn}) (Genetics 155:2011-14)

Observed value : 0.750
Bootst. mean : 0.851
Bootst. median : 0.875
Bootst. st. dev.: 0.099
Bootst. 95% CI : 0.667 - 1.000

Permutation Test

Skip

But of course

Item Description

Randomly allocate sequences into subpopulations and tabulate the distribution of various F_{ST} statistics.

1 required (1 chosen).

OK

Cancel

Permutation Test

Skip

But of course

Item Description

Randomly allocate sequences into subpopulations and tabulate the distribution of various F_{ST} statistics.

1 required (1 chosen).

OK

Cancel

Data set A

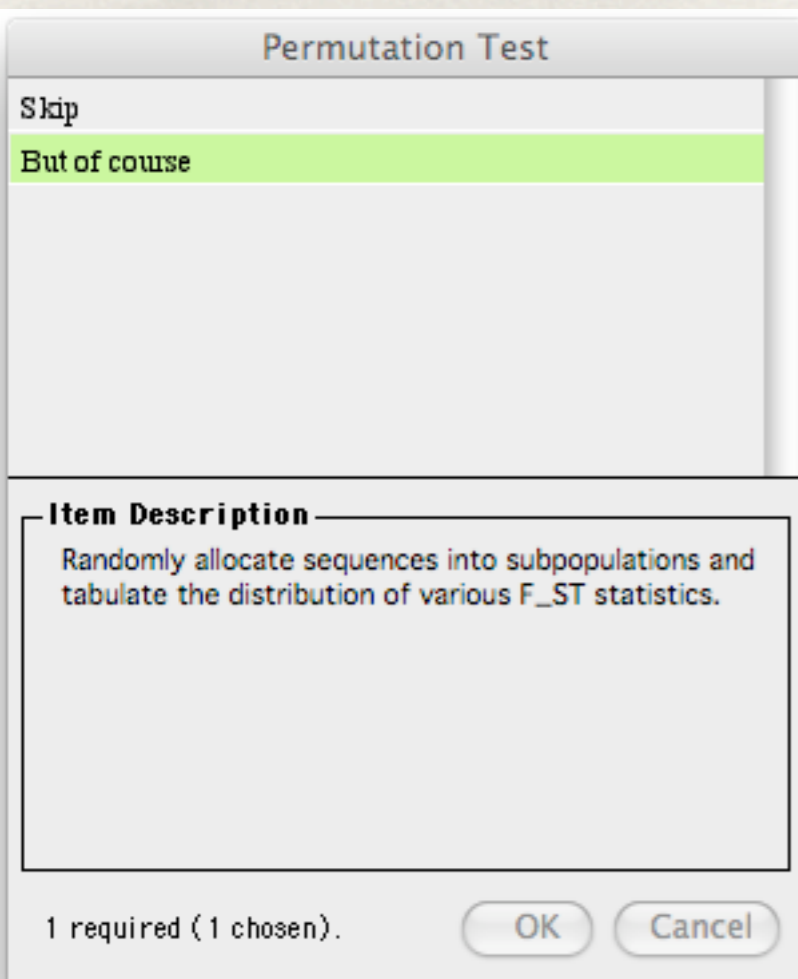
Prob {Random F_{ST} > Observed F_{ST} }

Hudson, Slatkin and Madison : 0

Slatkin : 0

Hudson, Boos and Kaplan : 0

Hudson, S_{nn} : 0



Data set A

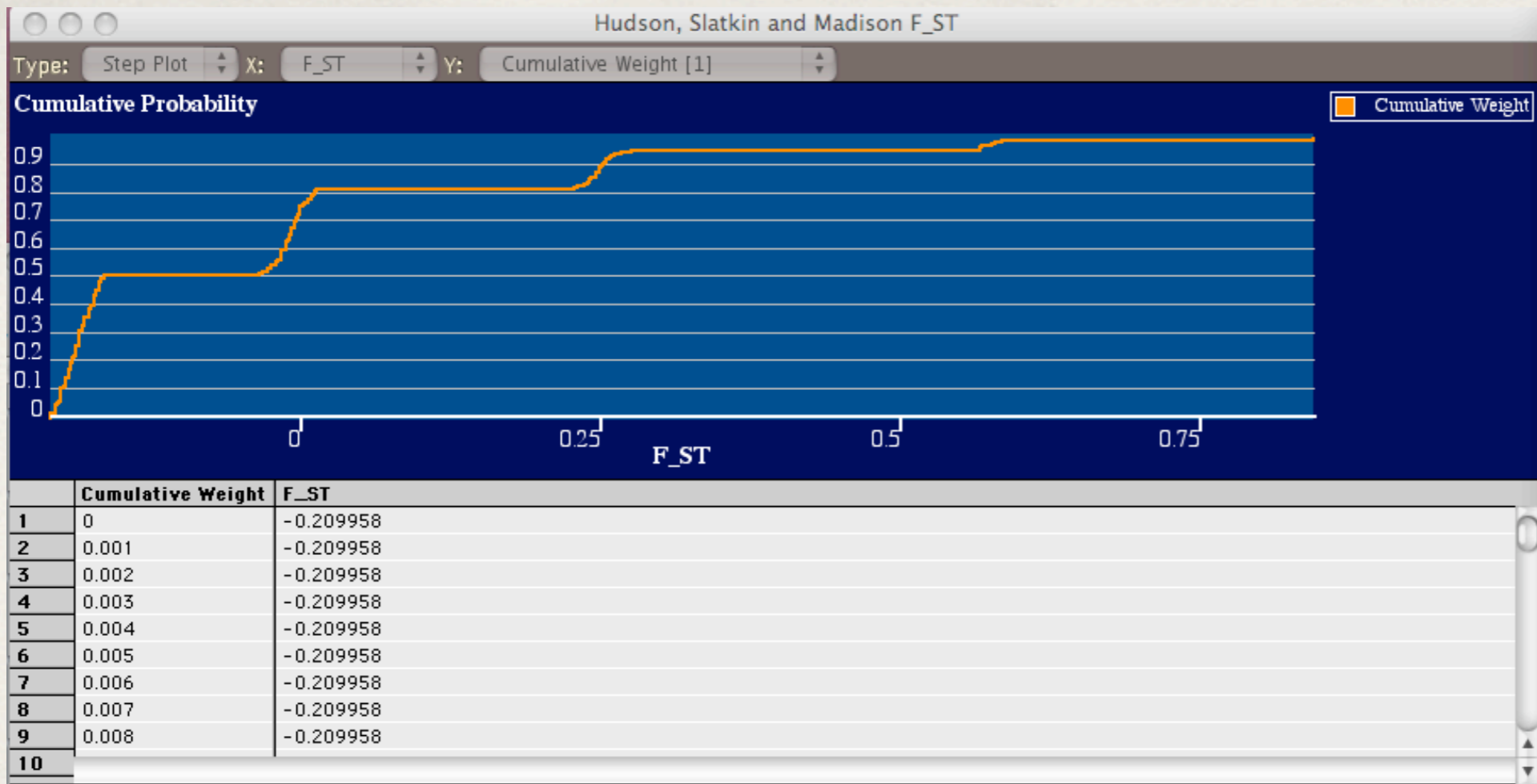
Prob {Random F_{ST} > Observed F_{ST} }

Hudson, Slatkin and Madison	: 0
Slatkin	: 0
Hudson, Boos and Kaplan	: 0
Hudson, S _{nn}	: 0

Data set B

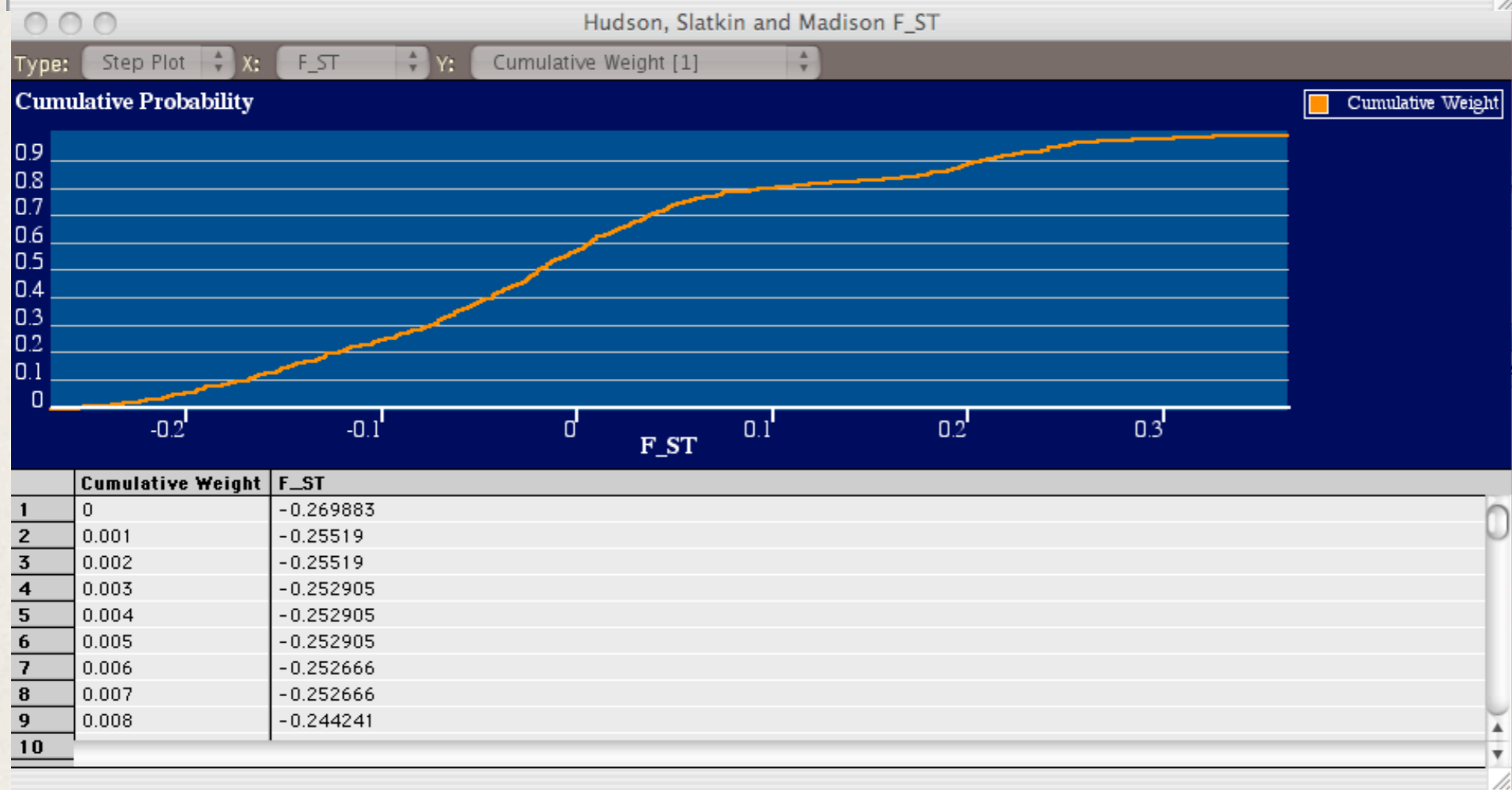
Prob {Random F_{ST} > Observed F_{ST} }

Hudson, Slatkin and Madison	: 0.173
Slatkin	: 0.173
Hudson, Boos and Kaplan	: 0.173
Hudson, S _{nn}	: 0.023



Data set A

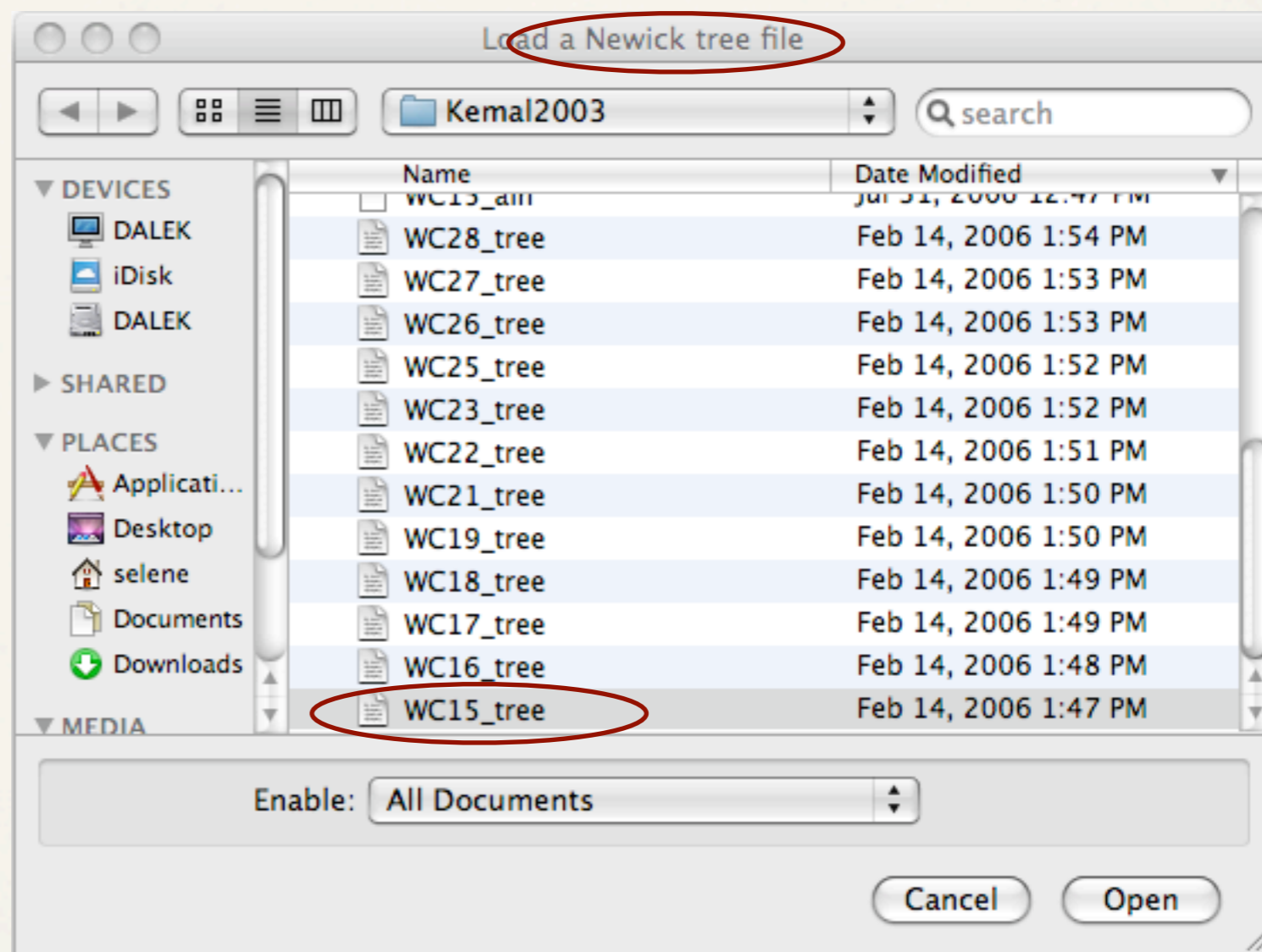
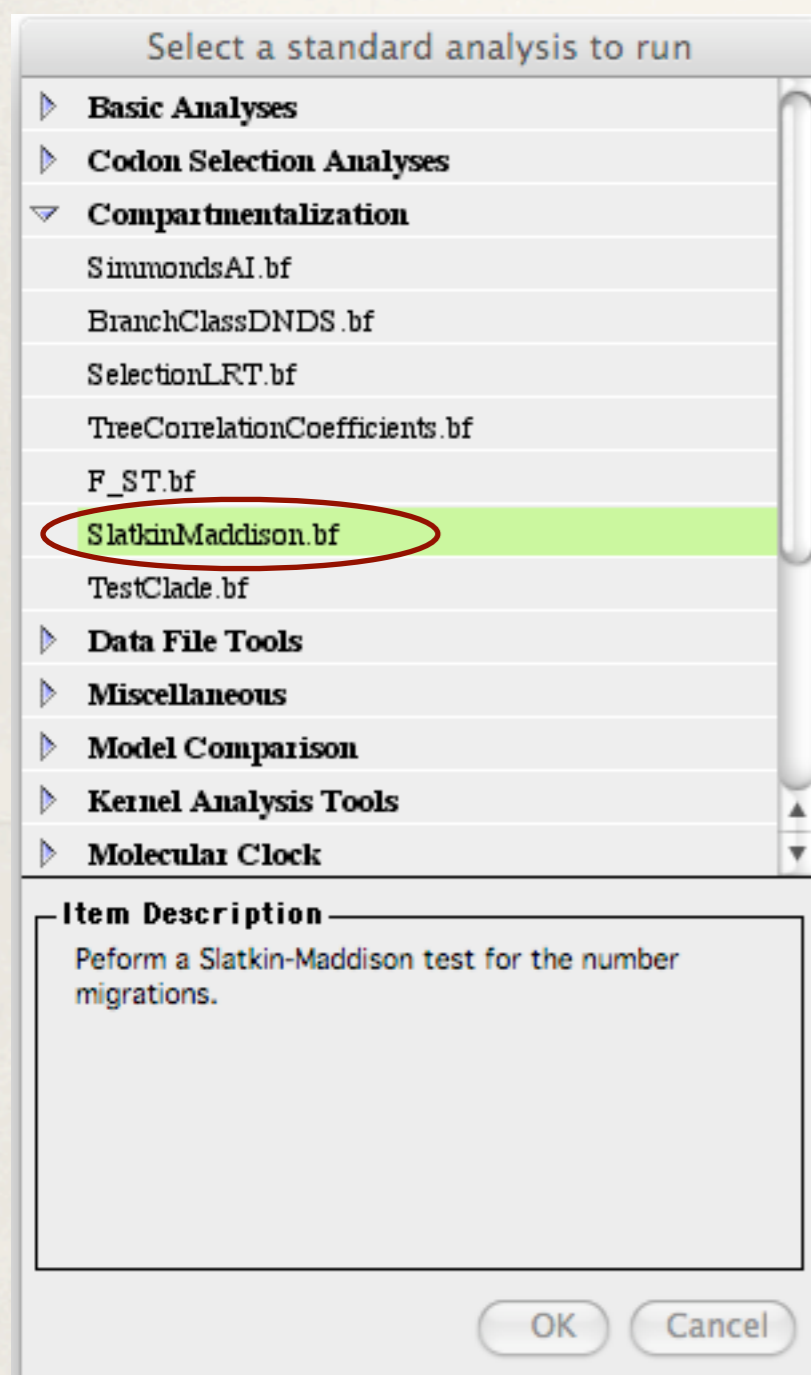
$$F_{ST} = 0.84$$



Data set B

$$F_{ST} = 0.13$$

Slatkin-Maddison test



You will need a phylogeny to carry out this analysis

Read tree: (((WC15_cv1,(WC15_pl,((WC15_pl_2,WC15_pl_3),WC15_pl_4))),WC15_c

How many sequence types: (>=2):2

Enter a reg exp used to define clade 1:pl

Matched: WC15_pl,WC15_pl_2,WC15_pl_3,WC15_pl_4

Matched: WC15_cv1,WC15_cv1_3,WC15_cv1_4,WC15_cv1_2,WC15_cv1_5

Set 1 (TYPE 1) includes 4 sequences:

- WC15_pl
- WC15_pl_2
- WC15_pl_3
- WC15_pl_4

Set 2 (TYPE 2) includes 5 sequences:

- WC15_cv1
- WC15_cv1_3
- WC15_cv1_4
- WC15_cv1_2
- WC15_cv1_5

```
Read tree: (((WC15_cv1,(WC15_pl,((WC15_pl_2,WC15_pl_3),WC15_pl_4))),WC15_c
How many sequence types: (>=2):2

Enter a reg exp used to define clade 1:pl
Matched: WC15_pl,WC15_pl_2,WC15_pl_3,WC15_pl_4
Matched: WC15_cv1,WC15_cv1_3,WC15_cv1_4,WC15_cv1_2,WC15_cv1_5

Set 1 (TYPE 1) includes 4 sequences:
  WC15_pl
  WC15_pl_2
  WC15_pl_3
  WC15_pl_4

Set 2 (TYPE 2) includes 5 sequences:
  WC15_cv1
  WC15_cv1_3
  WC15_cv1_4
  WC15_cv1_2
  WC15_cv1_5
```

Data set A

```
Is this partitioning correct (y/n)y
Please enter a descriptive name for TYPE 1 sequences:plasma

Proportion of plasma sequences: 0.444444
Please enter a descriptive name for TYPE 2 sequences:cv

Proportion of cv sequences: 0.555556

Inferred 1 migration events
```

```
The following branches have migration events:
```

```
cv --> plasma:
Node5
```



```

Read tree: (((WC15_cv1,(WC15_pl,((WC15_pl_2,WC15_pl_3),WC15_pl_4))),WC15_c
How many sequence types: (>=2):2

Enter a reg exp used to define clade 1:pl
Matched: WC15_pl,WC15_pl_2,WC15_pl_3,WC15_pl_4
Matched: WC15_cv1,WC15_cv1_3,WC15_cv1_4,WC15_cv1_2,WC15_cv1_5

Set 1 (TYPE 1) includes 4 sequences:
  WC15_pl
  WC15_pl_2
  WC15_pl_3
  WC15_pl_4

Set 2 (TYPE 2) includes 5 sequences:
  WC15_cv1
  WC15_cv1_3
  WC15_cv1_4
  WC15_cv1_2
  WC15_cv1_5

```

Data set A

```

Is this partitioning correct (y/n)y
Please enter a descriptive name for TYPE 1 sequences:plasma

Proportion of plasma sequences: 0.444444
Please enter a descriptive name for TYPE 2 sequences:cv

Proportion of cv sequences: 0.555556

Inferred 1 migration events

```

The following branches have migration events:

```

cv --> plasma:
Node5

```

Data set B

```

Proportion of brain sequences: 0.375
Please enter a descriptive name for TYPE 2 sequences:sp

Proportion of sp sequences: 0.625

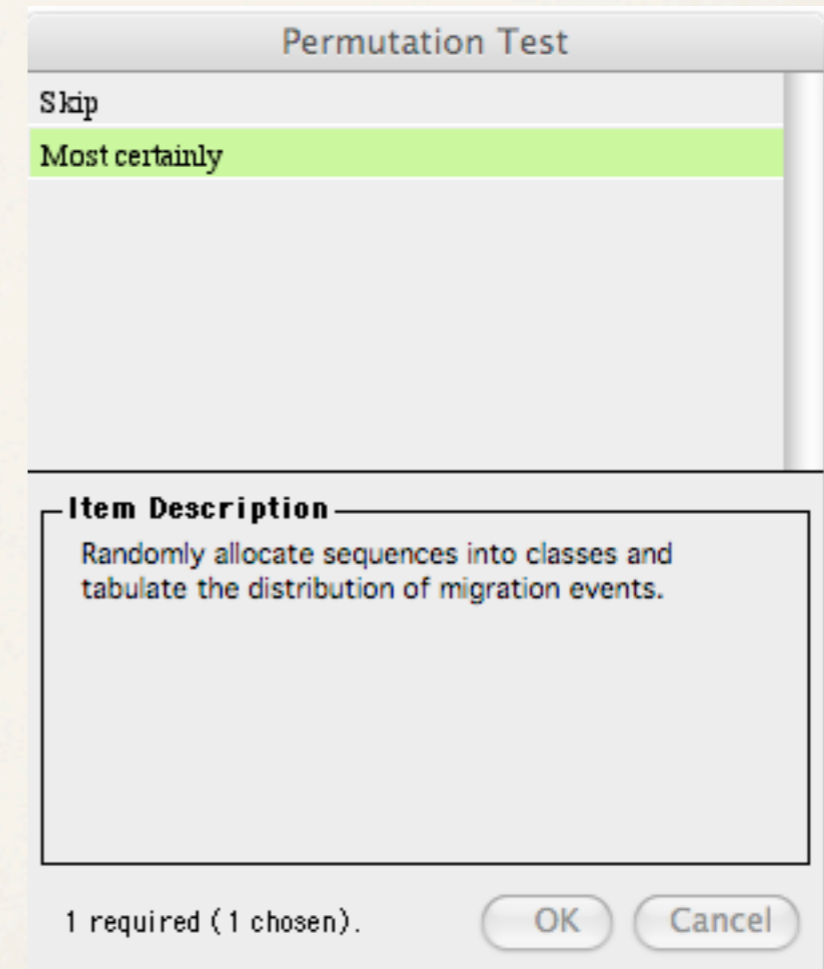
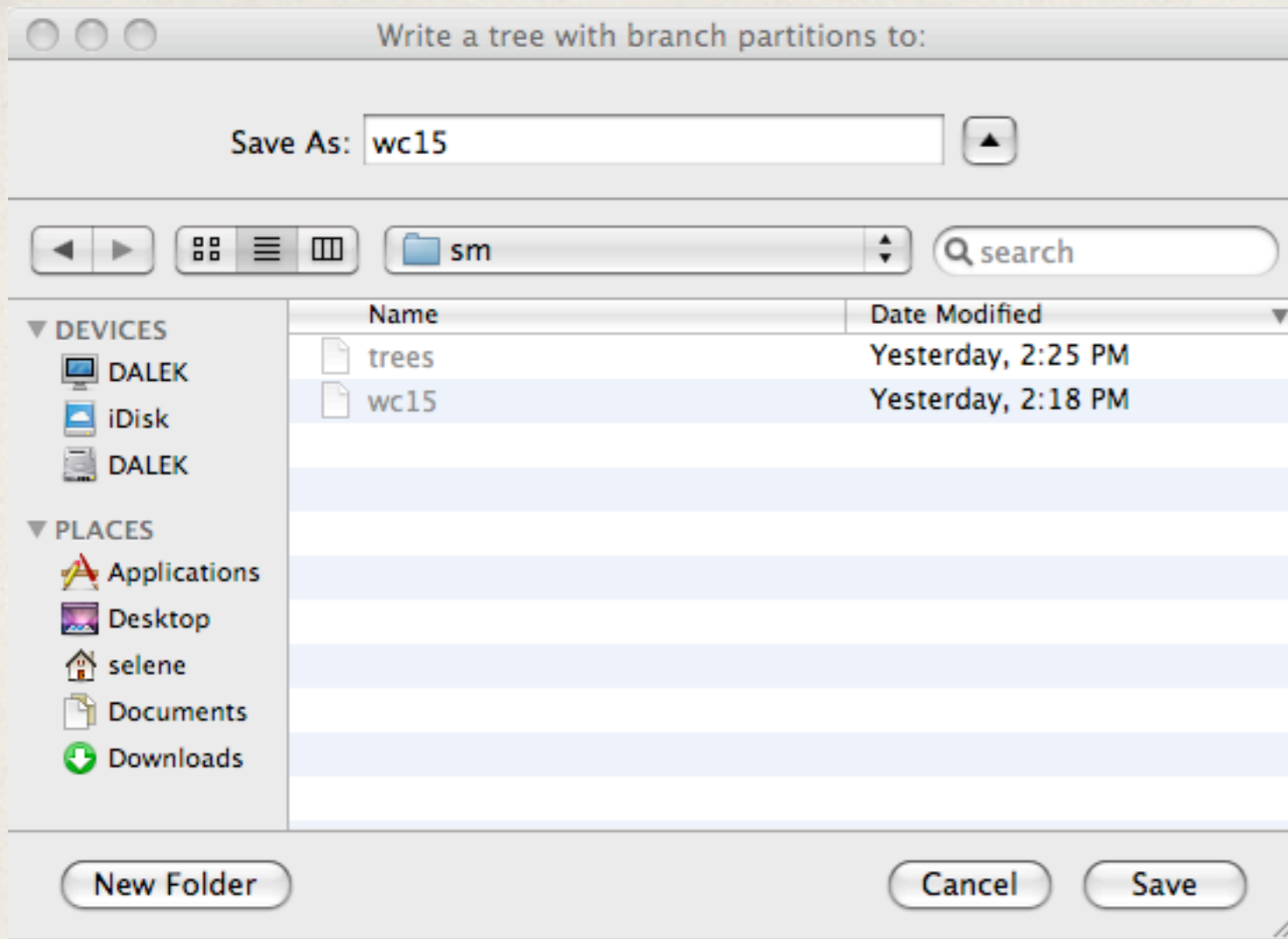
Inferred 3 migration events

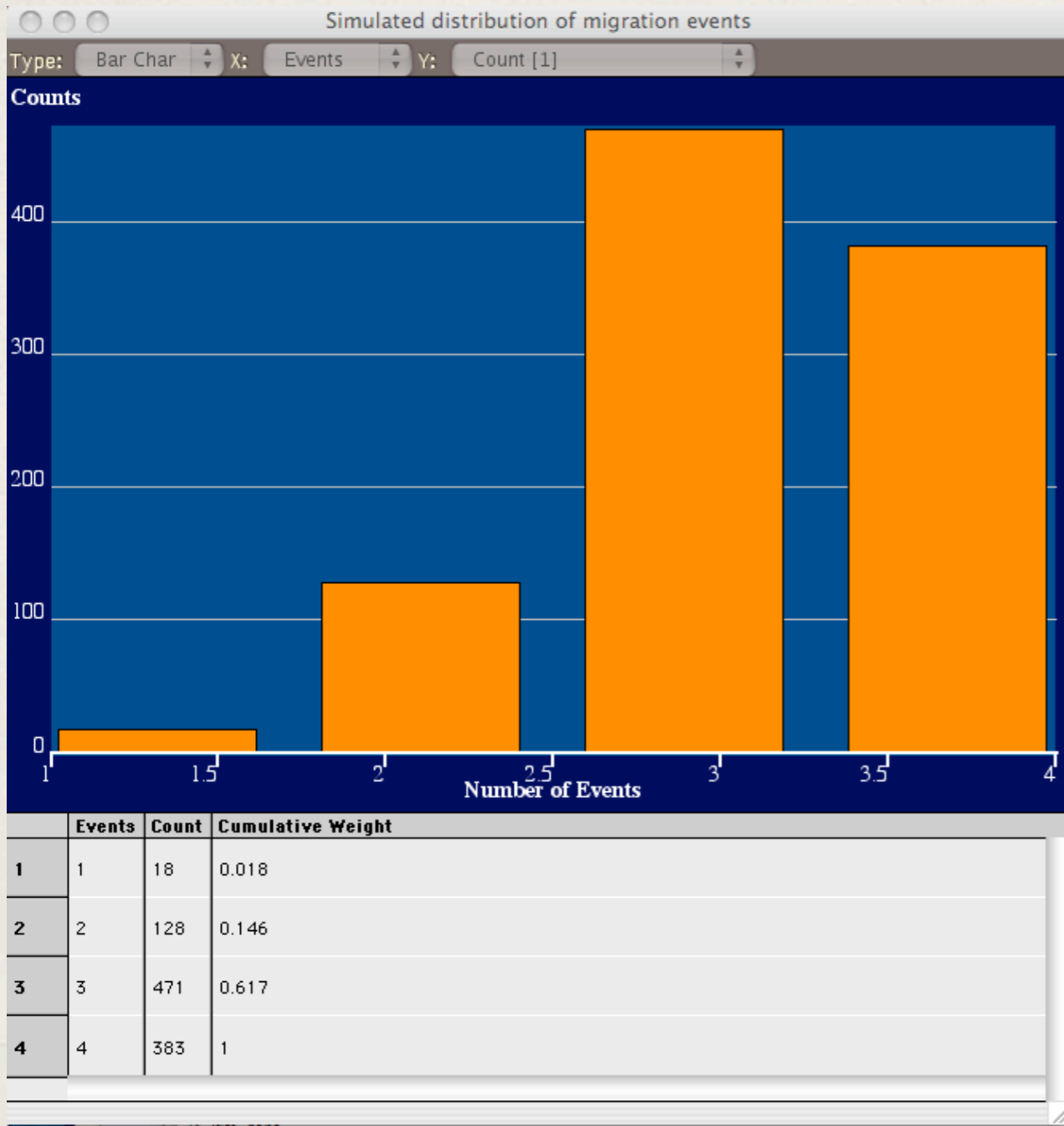
```

```

brain --> sp:
Node1
B_JP_X_SUBJECT_2_9sp
B_JP_X_SUBJECT_2_10sp

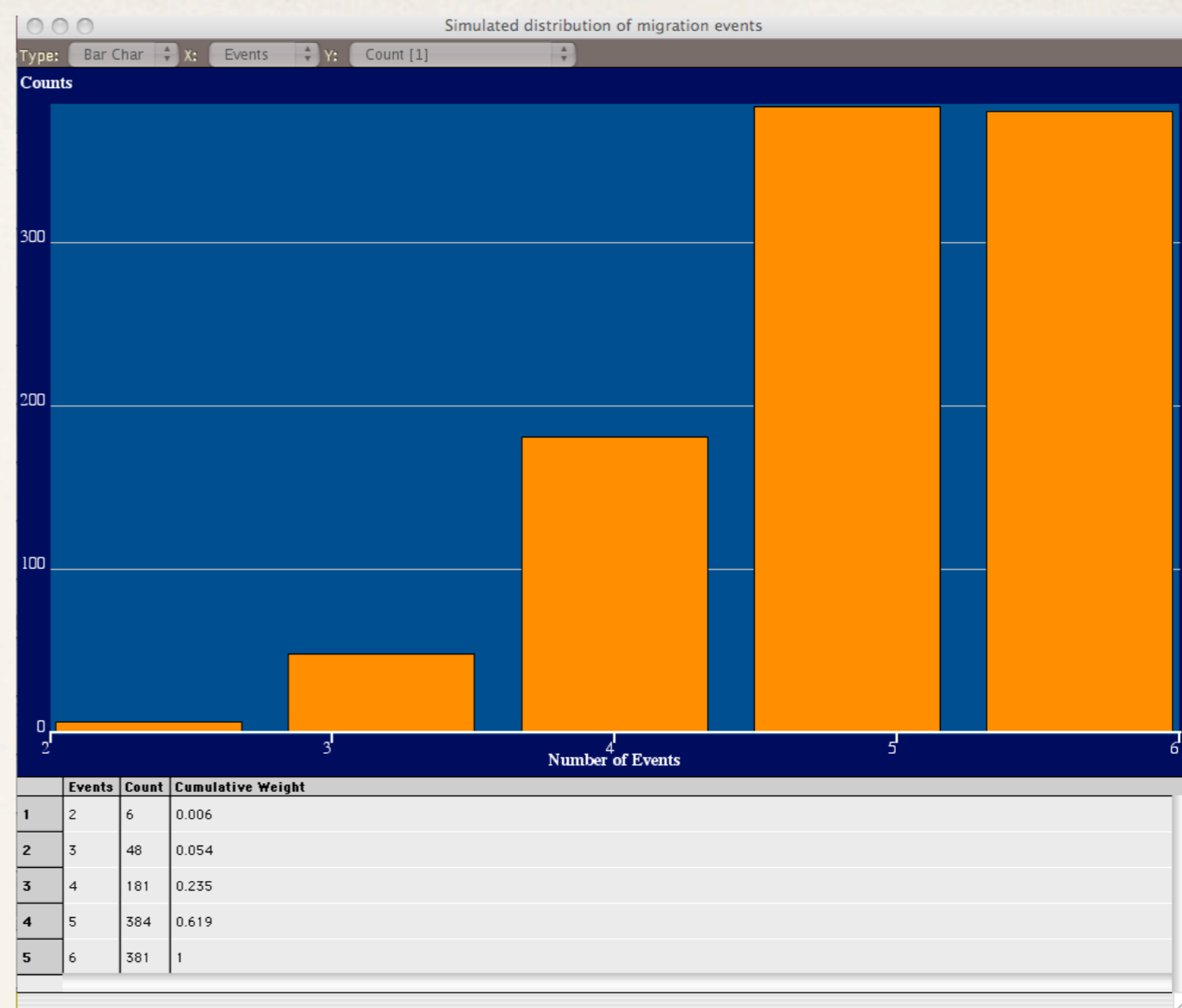
```





Prob{as many or fewer migration events by chance} = 0.018

Data set A

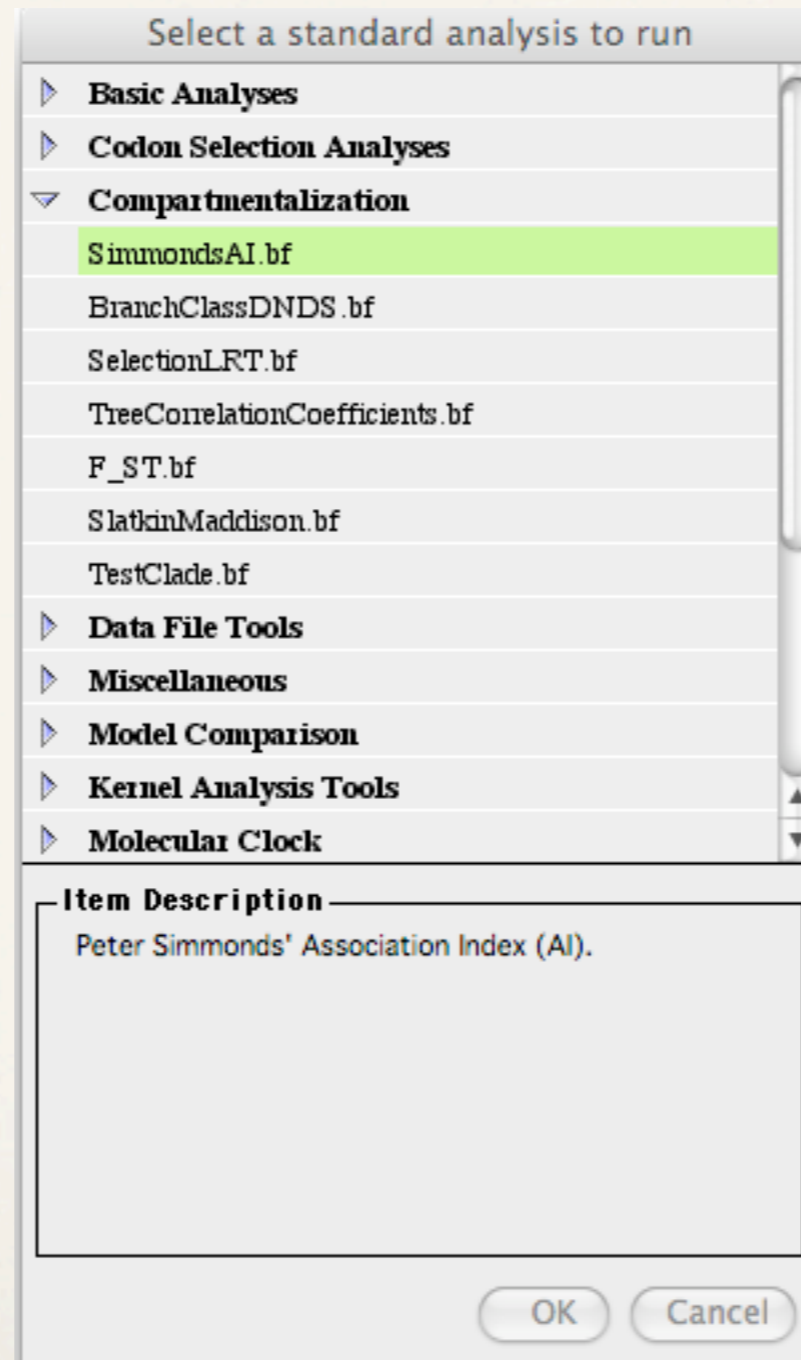


Prob{as many or fewer migration events by chance} = 0.054

Data set B

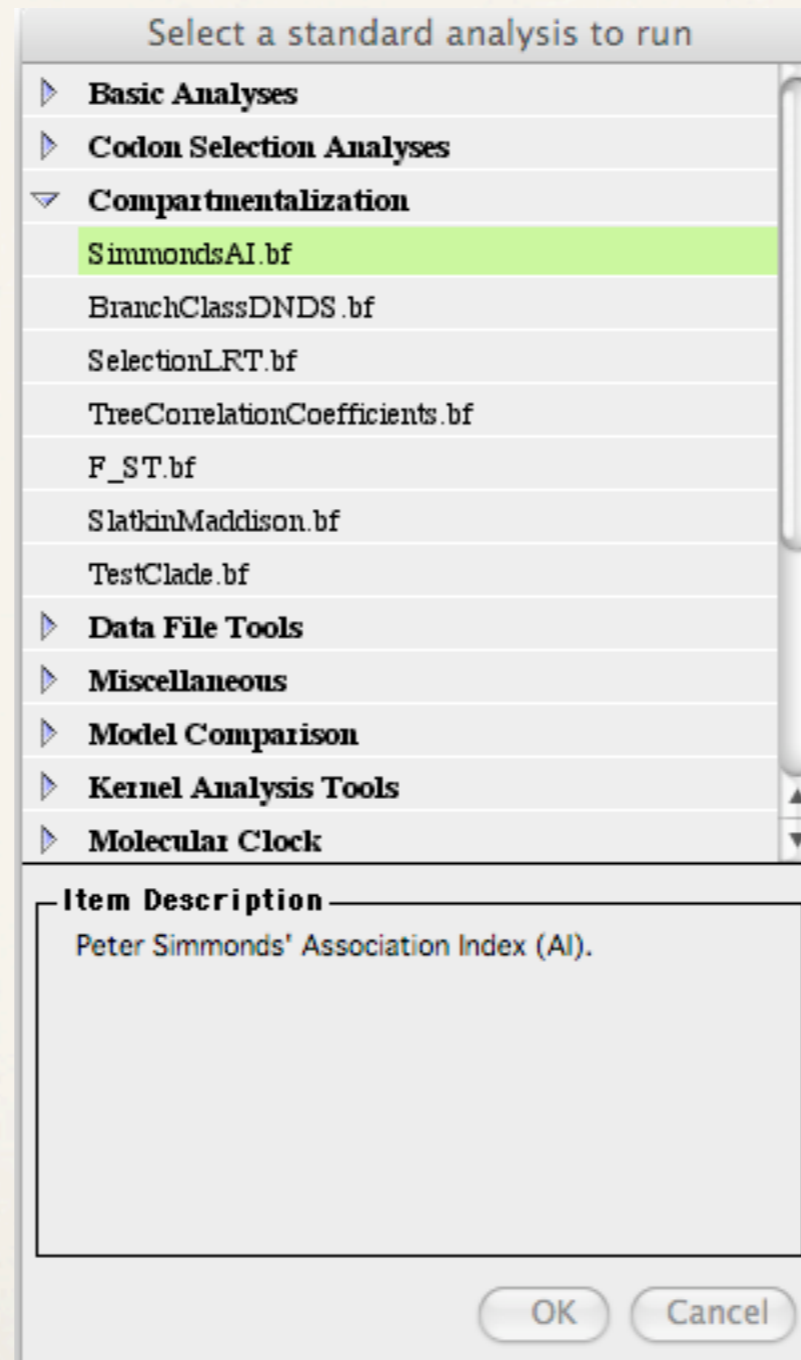
Association Index

- ❖ You will need an alignment that includes a sequence that can be considered as an outgroup



Association Index

- ❖ You will need an alignment that includes a sequence that can be considered as an outgroup



Association Index

```
Proportion of sequences in group 0: 0.4
Proportion of sequences in group 1: 0.6
How many relabelings per sample (default 10):?
How many tree bootstrap samples (default 100):?
Proportion of reshufflings less associated than the sample needed for significance (default 2/3)?
Using 100 tree bootstraps and 100 relabelings per sample with significance called at 0.666667

Baseline d = 0.00173611
Running tree simulations...

Association Index: 0.0142258
Bootstrap significance :100/100
```

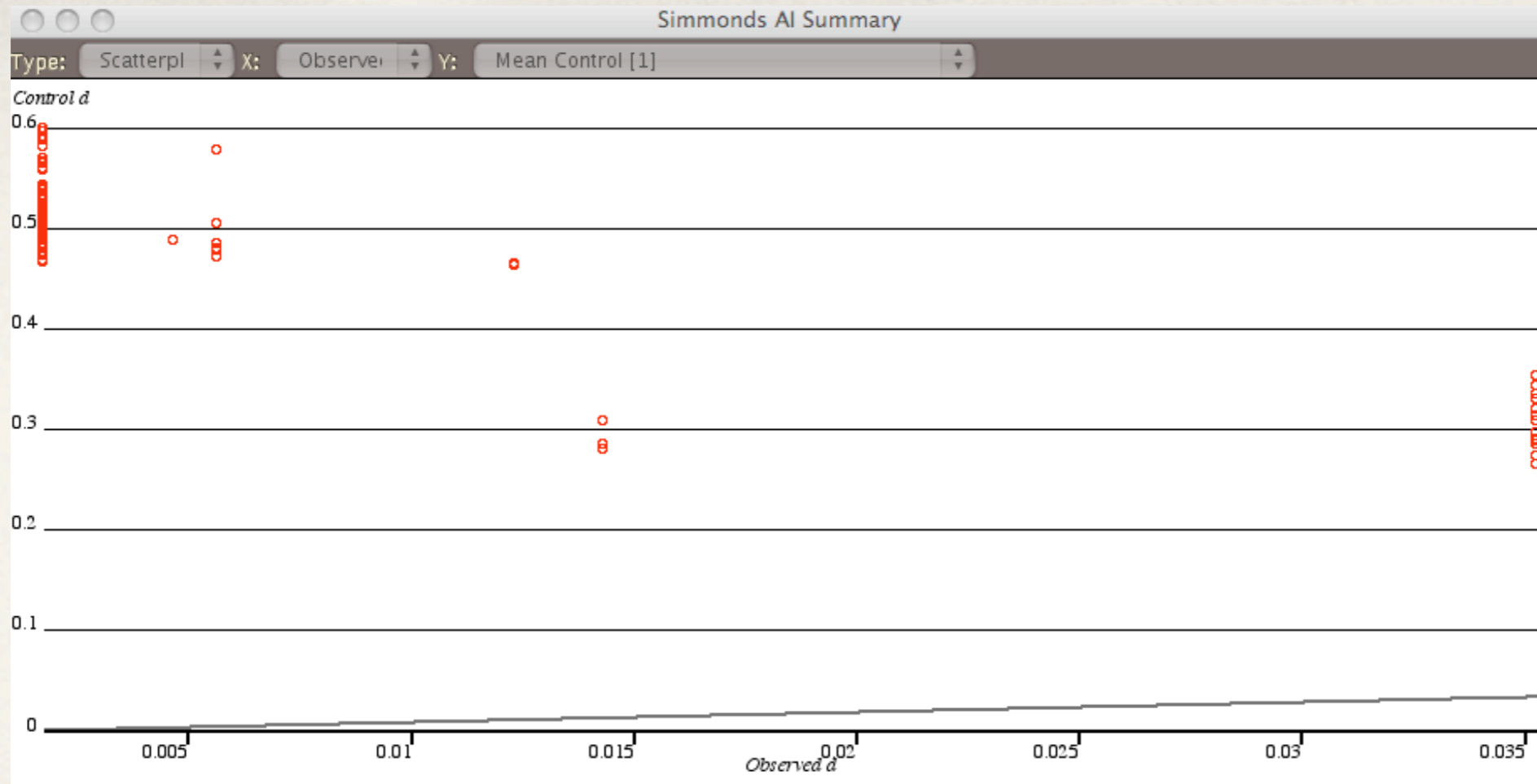
Data set A

```
Proportion of sequences in group 0: 0.352941
Proportion of sequences in group 1: 0.647059
How many relabelings per sample (default 10):?
How many tree bootstrap samples (default 100):?
Proportion of reshufflings less associated than the sample needed for significance (default 2/3)?
Using 100 tree bootstraps and 100 relabelings per sample with significance called at 0.666667

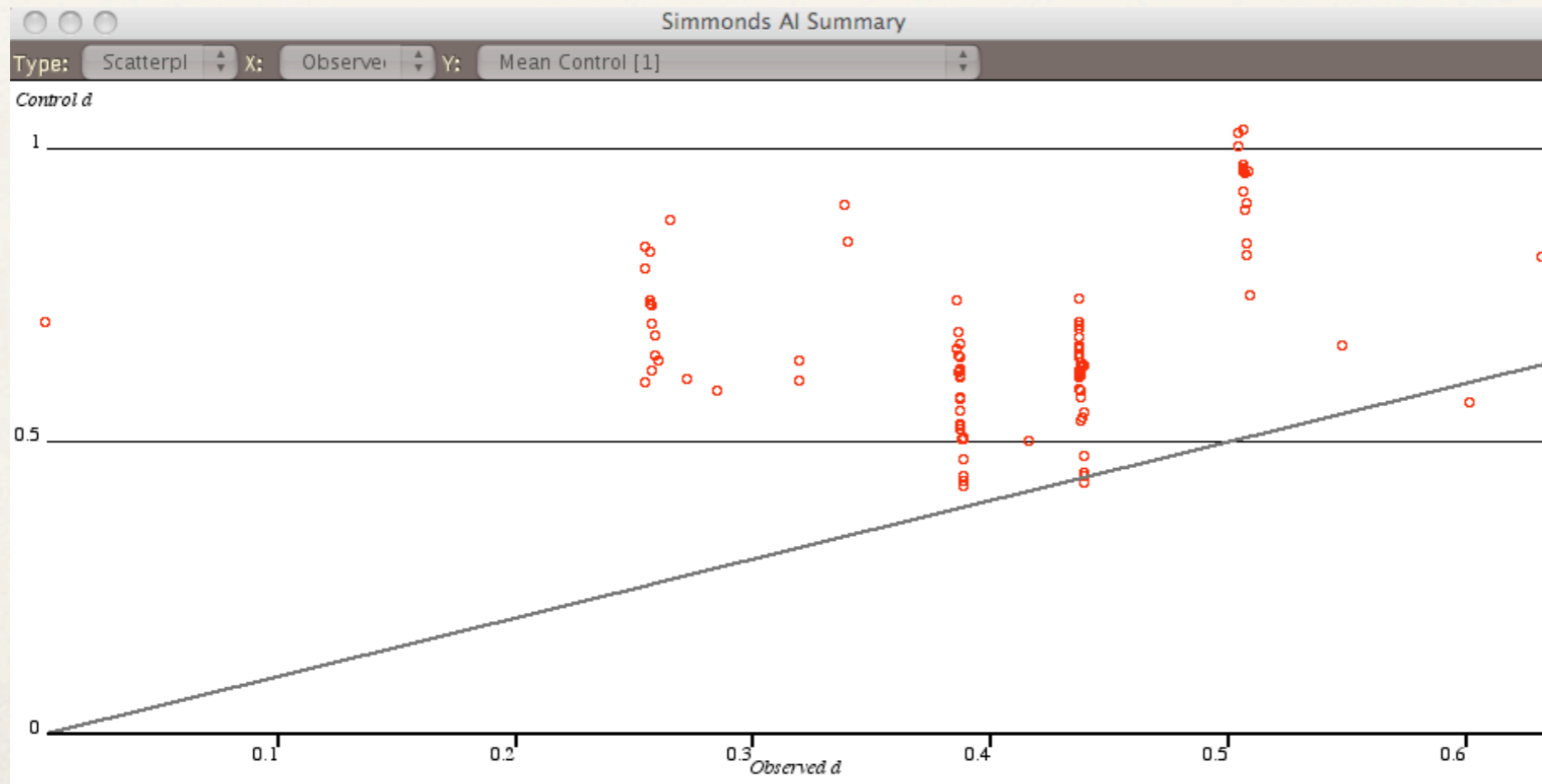
Baseline d = 0.44045
Running tree simulations...

Association Index: 0.601938
Bootstrap significance :87/100
```

Data set B



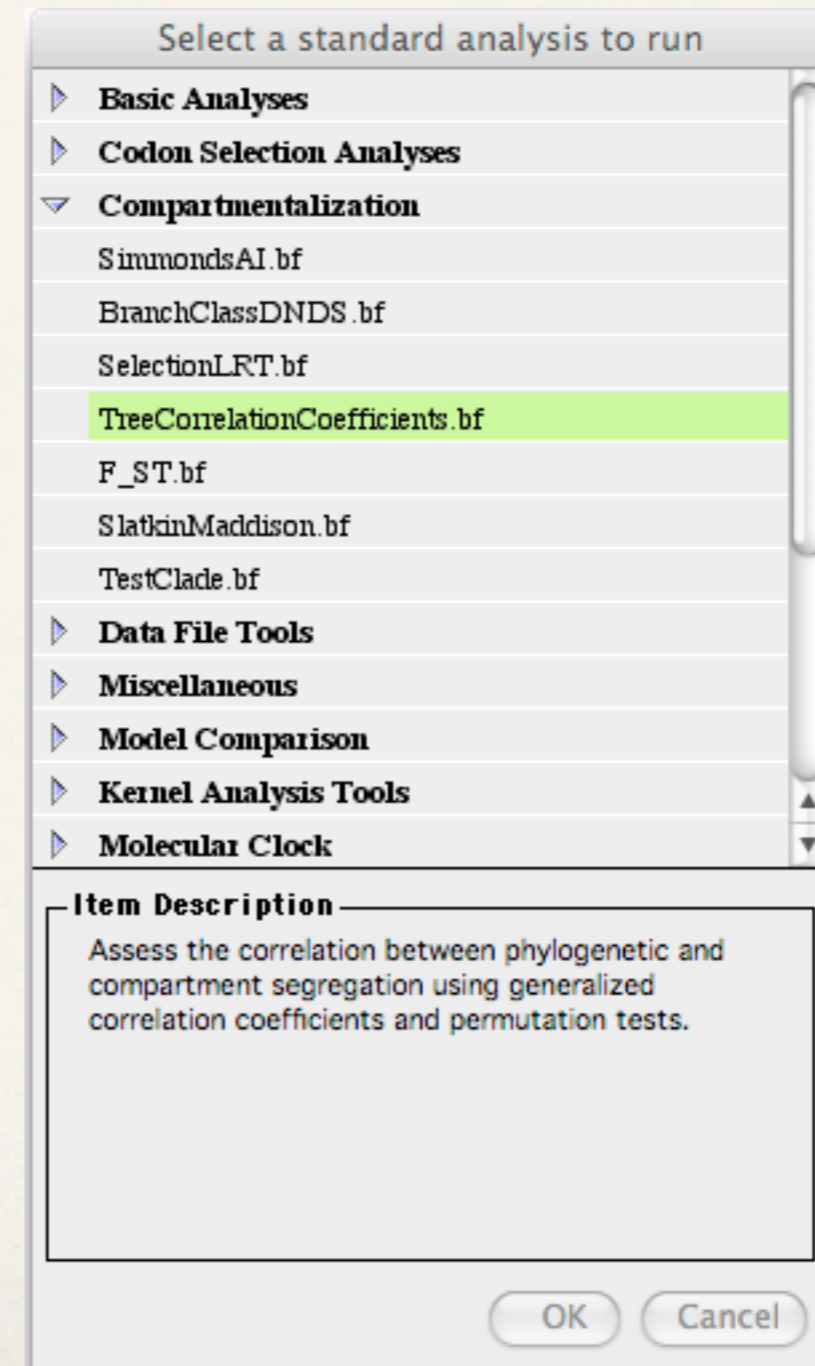
Data set A



Data set B

Correlation coefficients

- ❖ You will need to load a phylogeny to carry out this analysis.



Correlation Coefficients

Set 1 (TYPE 1) includes 4 sequences:

WC15_pl
WC15_pl_2
WC15_pl_3
WC15_pl_4

Set 2 (TYPE 2) includes 5 sequences:

WC15_cv1
WC15_cv1_3
WC15_cv1_4
WC15_cv1_2
WC15_cv1_5

Is this partitioning correct (y/n)y

Correlation coefficients:

Branch counts (r_b) : 0.73093
Path lengths (r) : 0.993191

Prob{ r_b random $\geq r_b$ observed} < 0.00899101
Prob{ r random $\geq r$ observed } < 0.015984

Data set A

Set 1 (TYPE 1) includes 6 sequences:

B_JP_X_SUBJECT_2_13br
B_JP_X_SUBJECT_2_11br
B_JP_X_SUBJECT_2_12br
B_JP_X_SUBJECT_2_14br
B_JP_X_SUBJECT_2_15br
B_JP_X_SUBJECT_2_16br

Set 2 (TYPE 2) includes 10 sequences:

B_JP_X_SUBJECT_2sp
B_JP_X_SUBJECT_2_2sp
B_JP_X_SUBJECT_2_3sp
B_JP_X_SUBJECT_2_7sp
B_JP_X_SUBJECT_2_6sp
B_JP_X_SUBJECT_2_4sp
B_JP_X_SUBJECT_2_5sp
B_JP_X_SUBJECT_2_8sp
B_JP_X_SUBJECT_2_9sp
B_JP_X_SUBJECT_2_10sp

Is this partitioning correct (y/n)y

Correlation coefficients:

Branch counts (r_b) : 0.233241
Path lengths (r) : 0.110911

Prob{ r_b random $\geq r_b$ observed} < 0.167832
Prob{ r random $\geq r$ observed } < 0.035964

Data set B

Comparison between methods

	Data set A	Data set B
F_{ST}	✓	✗
S_{nn}	✓	✓
SM	✓	✗
AI	✓	✓
r	✓	✓
r_b	✓	✗

Exercises

Exercises

- * Follow the instructions to determine if sequences from patients C, D, K, Q and S show evidence of compartmentalization

Results

	C	D	K	Q	S
FST (HSM)	0.074 p = 0.066	-0.033 p = 0.754	0.57 p = 0.12	0.83 p = 0	0.76 p = 0
FST (S)	0.039 p = 0.066	-0.016 p = 0.754	0.3 p = 0.12	0.70 p = 0	0.62 p = 0
FST (HBK)	0.039 p = 0.066	-0.016 p = 0.754	0.3 p = 0.12	0.70 p = 0	0.62 p = 0
Snn	0.64 p = 0.11	0.32 p = 0.952	0.72 p = 0.013	0.96 p = 0	1 p = 0
AI	0.71 boot = 85	1.38 boot = 1	0.63 boot = 89	0.23 boot = 100	1.2 x 10 ⁻⁸ boot = 100
SM	8 migrations p = 0.452	12 migrations p = 0.974	9 migrations p = 0.46	2 migrations p = 0	1 migration p = 0
r	-0.0012 P < 0.27	-0.04 p < 0.91	0.1 p < 0.026	0.83 p < 0.00099	0.95 p < 0.00099
rb	0.021 p < 0.42	-0.045 p < 0.93	0.14 p < 0.037	0.4729 p < 0.00099	0.69 p < 0.00099

Patient data

	C	D	K	Q	S
Current CD4 (cells/mm ³)	312	55	221	68	32
Plasma RNA (log copies/ml)	5.7	5.9	5.7	5.1	6
CSF RNA (log copies/ml)	5.2	4	4.4	3.5	3.2
CSF WBC (cells/mm ³)	312	2	16	2	3
Compartmentalized					