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 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-HE)]
- Observed heterozygosity (H0)
- Expected heterozygosity (HE)
- Fixation Index (FIS)
- Within-population gene diversity (HI)
- Mean within-population gene diversity (Hs)
- Total diversity (HT)
- Coefficient of gene differentiation among populations (GST)
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.
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 = $\frac{14}{36} = 0.39$
Distance between subpopulation = $\frac{53}{81} = 0.65$

$F_{ST} = \frac{(0.65 - 0.39)}{0.65} = 0.4$
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).

\[
S_{nn} = \frac{1}{n} \sum_{j=1}^{n} X_j
\]

- where $X_j$ is 1 if the nearest neighbor was isolated from the same subpopulation or 0 otherwise.
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 $6/9$

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

$S_{nn} = \frac{14(6/9) + 4 (1/9)}{18} = 0.54$
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 \[ d = \frac{(1-f)}{2^{n-1}}. \]

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

\[ p_0 = f_{yy} + f_{nn} \]
\[ p_e = (f_{yy} + f_{yn})*(f_{yy} + f_{ny}) + (f_{nn} + f_{ny})*(f_{nn} + f_{yn}) \]
\[ \kappa = \frac{(p_0 - p_e)}{(1 - pe)} \]


<table>
<thead>
<tr>
<th>TABLE 2. Levels of agreement between methods as measured by pairwise ( \kappa ) scores</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comparison</strong></td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td><strong>Different-class methods</strong></td>
</tr>
<tr>
<td>SM vs ( F_{ST} )</td>
</tr>
<tr>
<td>SM vs ( S_{nn} )</td>
</tr>
<tr>
<td>SM vs AMOVA</td>
</tr>
<tr>
<td>AI vs ( F_{ST} )</td>
</tr>
<tr>
<td>AI vs ( S_{nn} )</td>
</tr>
<tr>
<td>AI vs AMOVA</td>
</tr>
<tr>
<td><strong>Same-class methods</strong></td>
</tr>
<tr>
<td>SM vs AI</td>
</tr>
<tr>
<td>( F_{ST} ) vs ( S_{nn} )</td>
</tr>
<tr>
<td>( F_{ST} ) vs AMOVA</td>
</tr>
<tr>
<td>( S_{nn} ) vs AMOVA</td>
</tr>
<tr>
<td>( r ) vs ( r_b )</td>
</tr>
</tbody>
</table>

\(^a\) 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$^a$

<table>
<thead>
<tr>
<th>Method</th>
<th>Proportion classified as compartmentalized when sample sizes were:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Equal ($n = 20$)</td>
</tr>
<tr>
<td>SM</td>
<td>0.98</td>
</tr>
<tr>
<td>$F_{ST}$</td>
<td>0.55</td>
</tr>
<tr>
<td>$S_{nn}$</td>
<td>0.99</td>
</tr>
<tr>
<td>AI</td>
<td>0.85</td>
</tr>
<tr>
<td>AMOVA</td>
<td>0.41</td>
</tr>
<tr>
<td>$r^b$</td>
<td>0.54</td>
</tr>
<tr>
<td>$r_b^c$</td>
<td>0.71</td>
</tr>
</tbody>
</table>

$^a$ 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).
Read the following data: 9 species: \{WC15\_cvl, WC15\_cvl\_2, WC15\_cvl\_3, WC15\_cvl\_4, WC15\_cvl\_5, C15\_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, WC5_cv1, 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_cvl, WC15_cvl_2, WC15_cvl_3, WC15_cvl_4, WC15_cvl_5, C15_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: cvl

Enter a regular expression to define the second clade: pl

Clade 1 includes 5 sequences:
  WC15_cvl
  WC15_cvl_2
  WC15_cvl_3
  WC15_cvl_4
  WC15_cvl_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: 0.5555556
Proportion of sequence in population 2: 0.4444444
$F_{ST}$ and $S_{nn}$

- Population characteristics:
  - Metapopulation diversity ($\pi_T$)
  - Mean subpopulation diversity ($\pi_S$)
  - Mean interpopulation diversity ($\pi_B$)

- $F_{ST}$

  - Hudson, Slatkin and Madison (Genetics 132:583-589)
  - Slatkin (Evolution 47:264-279)
  - Hudson ($S_{nn}$) (Genetics 155:2011-14)
Data set A

Population characteristics:
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 characteristics:
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 characteristics:
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 characteristics:
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

<table>
<thead>
<tr>
<th>Bootstrapped estimator statistics.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hudson, Slatkin and Madison (Genetics 132:583-589)</strong></td>
</tr>
<tr>
<td>Observed value : 0.843</td>
</tr>
<tr>
<td>Bootst. mean : 0.877</td>
</tr>
<tr>
<td>Bootst. median : 0.875</td>
</tr>
<tr>
<td>Bootst. st. dev.: 0.021</td>
</tr>
<tr>
<td>Bootst. 95% CI : 0.844 - 0.924</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Slatkin (Evolution 47:264-279)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed value : 0.729</td>
</tr>
<tr>
<td>Bootst. mean : 0.782</td>
</tr>
<tr>
<td>Bootst. median : 0.778</td>
</tr>
<tr>
<td>Bootst. st. dev.: 0.033</td>
</tr>
<tr>
<td>Bootst. 95% CI : 0.730 - 0.859</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hudson, Boos and Kaplan (Mol Bio Evol 9: 138-151)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed value : 0.726</td>
</tr>
<tr>
<td>Bootst. mean : 0.780</td>
</tr>
<tr>
<td>Bootst. median : 0.776</td>
</tr>
<tr>
<td>Bootst. st. dev.: 0.034</td>
</tr>
<tr>
<td>Bootst. 95% CI : 0.727 - 0.857</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hudson (S_nn) (Genetics 155:2011-14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed value : 1.000</td>
</tr>
<tr>
<td>Bootst. mean : 1.000</td>
</tr>
<tr>
<td>Bootst. median : 1.000</td>
</tr>
<tr>
<td>Bootst. st. dev.: 0.000</td>
</tr>
<tr>
<td>Bootst. 95% CI : 1.000 - 1.000</td>
</tr>
</tbody>
</table>

### Data set B

<table>
<thead>
<tr>
<th>Bootstrapped estimator statistics.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hudson, Slatkin and Madison (Genetics 132:583-589)</strong></td>
</tr>
<tr>
<td>Observed value : 0.130</td>
</tr>
<tr>
<td>Bootst. mean : 0.221</td>
</tr>
<tr>
<td>Bootst. median : 0.209</td>
</tr>
<tr>
<td>Bootst. st. dev.: 0.185</td>
</tr>
<tr>
<td>Bootst. 95% CI : -0.000 - 0.556</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Slatkin (Evolution 47:264-279)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed value : 0.070</td>
</tr>
<tr>
<td>Bootst. mean : 0.137</td>
</tr>
<tr>
<td>Bootst. median : 0.117</td>
</tr>
<tr>
<td>Bootst. st. dev.: 0.122</td>
</tr>
<tr>
<td>Bootst. 95% CI : -0.043 - 0.385</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hudson, Boos and Kaplan (Mol Bio Evol 9: 138-151)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed value : 0.066</td>
</tr>
<tr>
<td>Bootst. mean : 0.130</td>
</tr>
<tr>
<td>Bootst. median : 0.110</td>
</tr>
<tr>
<td>Bootst. st. dev.: 0.117</td>
</tr>
<tr>
<td>Bootst. 95% CI : -0.040 - 0.370</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hudson (S_nn) (Genetics 155:2011-14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed value : 0.750</td>
</tr>
<tr>
<td>Bootst. mean : 0.851</td>
</tr>
<tr>
<td>Bootst. median : 0.875</td>
</tr>
<tr>
<td>Bootst. st. dev.: 0.009</td>
</tr>
<tr>
<td>Bootst. 95% CI : 0.667 - 1.000</td>
</tr>
</tbody>
</table>
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).
Data set A

\[
\text{Prob } \{\text{Random } F_{\text{ST}} > \text{Observed } F_{\text{ST}}\}
\]

Hudson, Slatkin and Madison : 0
Slatkin : 0
Hudson, Boos and Kaplan : 0
Hudson, S\_nn : 0

Item Description
Randomly allocate sequences into subpopulations and tabulate the distribution of various F\_ST statistics.
Data set A

\[
\text{Prob } \{\text{Random } F_{ST} > \text{Observed } F_{ST}\}
\]

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

Data set B

\[
\text{Prob } \{\text{Random } F_{ST} > \text{Observed } F_{ST}\}
\]

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_cvl,(WC15_pl,((WC15_pl_2,WC15_pl_3),WC15_pl_4))),(WC15_cvl),)
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_cvl,WC15_cvl_3,WC15_cvl_4,WC15_cvl_2,WC15_cvl_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_cvl
   WC15_cvl_3
   WC15_cvl_4
   WC15_cvl_2
   WC15_cvl_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_cvl,(WC15_pl1,(WC15_pl2,WC15_pl3),WC15_pl4)),)),WC15_cvl
How many sequence types: (>=2): 2

Enter a reg exp used to define clade 1:pl
Matched: WC15_pl1, WC15_pl2, WC15_pl3, WC15_pl4
Matched: WC15_cvl, WC15_cvl3, WC15_cvl4, WC15_cvl2, WC15_cvl5

Set 1 (TYPE 1) includes 4 sequences:
  WC15_pl1
  WC15_pl2
  WC15_pl3
  WC15_pl4

Set 2 (TYPE 2) includes 5 sequences:
  WC15_cvl
  WC15_cvl3
  WC15_cvl4
  WC15_cvl2
  WC15_cvl5

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
Write a tree with branch partitions to:

Save As: wc15

DEVICES
- DALEK
- iDisk
- DALEK

PLACES
- Applications
- Desktop
- selene
- Documents
- Downloads

New Folder  Cancel  Save

Permutation Test

Skip
Most certainly

Item Description
Randomly allocate sequences into classes and tabulate the distribution of migration events.

1 required (1 chosen).  OK  Cancel
Data set A

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

Data set B

Prob\{as many or fewer migration events by chance\} = 0.054
You will need an alignment that includes a sequence that can be considered as an outgroup.
You will need an alignment that includes a sequence that can be considered as an outgroup.
Association Index

Data set A

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 resshufflings 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 B

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 resshufflings 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 A

Data set B
Correlation coefficients

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

Data set A

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_cvl
- WC15_cvl_3
- WC15_cvl_4
- WC15_cvl_2
- WC15_cvl_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 >= r_b observed} < 0.00899101
Prob{r random >= r observed} < 0.015984

Data set B

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_2_sp
- 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 >= r_b observed} < 0.167832
Prob{r random >= r observed} < 0.035964
Comparison between methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Data set A</th>
<th>Data set B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>F_{ST}</strong></td>
<td>✔</td>
<td>✘</td>
</tr>
<tr>
<td><strong>S_{nn}</strong></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td><strong>SM</strong></td>
<td>✔</td>
<td>✘</td>
</tr>
<tr>
<td><strong>AI</strong></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td><strong>r</strong></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td><strong>r_{b}</strong></td>
<td>✔</td>
<td>✘</td>
</tr>
</tbody>
</table>
Exercises
Exercises

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

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
<th>K</th>
<th>Q</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>FST (HSM)</td>
<td>0.074 p = 0.066</td>
<td>-0.033 p = 0.754</td>
<td>0.57 p = 0.12</td>
<td>0.83 p = 0</td>
<td>0.76 p = 0</td>
</tr>
<tr>
<td>FST (S)</td>
<td>0.039 p = 0.066</td>
<td>-0.016 p = 0.754</td>
<td>0.3 p = 0.12</td>
<td>0.70 p = 0</td>
<td>0.62 p = 0</td>
</tr>
<tr>
<td>FST (HBK)</td>
<td>0.039 p = 0.066</td>
<td>-0.016 p = 0.754</td>
<td>0.3 p = 0.12</td>
<td>0.70 p = 0</td>
<td>0.62 p = 0</td>
</tr>
<tr>
<td>Snn</td>
<td>0.64 p = 0.11</td>
<td>0.32 p = 0.952</td>
<td>0.72 p = 0.013</td>
<td>0.96 p = 0</td>
<td>1 p = 0</td>
</tr>
<tr>
<td>Al</td>
<td>0.71 boot = 85</td>
<td>1.38 boot = 1</td>
<td>0.63 boot = 89</td>
<td>0.23 boot = 100</td>
<td>1.2 x 10^{-8} boot = 100</td>
</tr>
<tr>
<td>SM</td>
<td>8 migrations p = 0.452</td>
<td>12 migrations p = 0.974</td>
<td>9 migrations p = 0.46</td>
<td>2 migrations p = 0</td>
<td>1 migration p = 0</td>
</tr>
<tr>
<td>r</td>
<td>-0.0012 P &lt; 0.27</td>
<td>-0.04 p &lt; 0.91</td>
<td>0.1 p &lt; 0.026</td>
<td>0.83 p &lt; 0.00099</td>
<td>0.95 p &lt; 0.00099</td>
</tr>
<tr>
<td>rb</td>
<td>0.021 p &lt; 0.42</td>
<td>-0.045 p &lt; 0.93</td>
<td>0.14 p &lt; 0.037</td>
<td>0.4729 p &lt; 0.00099</td>
<td>0.69 p &lt; 0.00099</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>D</td>
<td>K</td>
<td>Q</td>
<td>S</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Current CD4 (cells/mm$^3$)</td>
<td>312</td>
<td>55</td>
<td>221</td>
<td>68</td>
<td>32</td>
</tr>
<tr>
<td>Plasma RNA (log copies/ml)</td>
<td>5.7</td>
<td>5.9</td>
<td>5.7</td>
<td>5.1</td>
<td>6</td>
</tr>
<tr>
<td>CSF RNA (log copies/ml)</td>
<td>5.2</td>
<td>4</td>
<td>4.4</td>
<td>3.5</td>
<td>3.2</td>
</tr>
<tr>
<td>CSF WBC (cells/mm$^3$)</td>
<td>312</td>
<td>2</td>
<td>16</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Compartmentalized</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>