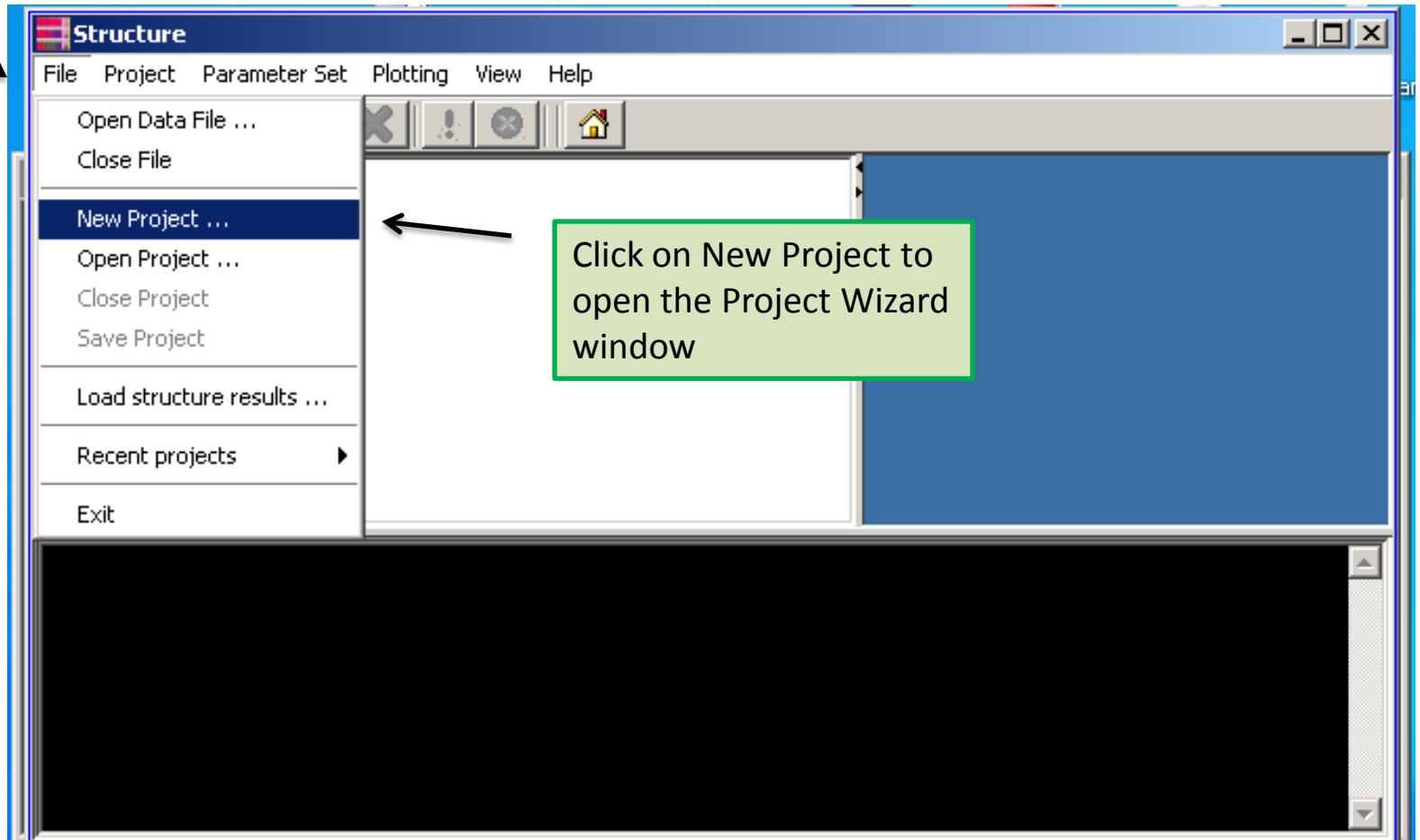


STRUCTURE EXAMPLE

STRUCTURE: Pritchard et al., 2000 (free ware from Jonathan Pritchard, University of Chicago,
<http://pritch.bsd.uchicago.edu/structure.html>)

Click on File to select
the New Project option

CLICK ON THE STRUCTURE ICON ON THE DESKTOP. THIS WINDOW OPENS.



You have now opened the Project Wizard at Step 1

The screenshot shows a dialog box titled "Step 1 of 4 - Project Wizard" with the subtitle "Step 1 of 4: Project information". It contains three input fields: "Name the project" with the value "FROG", "Select directory" with the value "DESKTOP" and a "Browse ..." button, and "Choose data file" with the value "FROG.TXT" and a "Browse ..." button. At the bottom are "Next>>" and "Cancel" buttons. Four green callout boxes with arrows provide instructions: "Enter an arbitrary name of your project" points to the project name field; "Select the directory and folder containing the dataset" points to the directory field and its browse button; "Select the dataset to be analyzed" points to the data file field and its browse button; and "Then click Next" points to the "Next>>" button.

Step 1 of 4 - Project Wizard

Step 1 of 4: Project information

Name the project

Select directory

Choose data file

Enter an arbitrary name of your project

Select the directory and folder containing the dataset

Select the dataset to be analyzed

Then click Next

The numbers and boxes checked in these two steps of the Project Wizard define the organization of the dataset used in this example and the boxes should be filled with numbers specific to the dataset.

Step 2 of 4: Information of input data set

Number of individuals: 2278 ←

Ploidy of data: 2 ←

Number of loci: 39 ←

Missing data value: -9 ←

Show data file format

<<Back Next>> Cancel

This screenshot shows the 'Step 2 of 4: Information of input data set' dialog box. It contains four input fields: 'Number of individuals' with the value 2278, 'Ploidy of data' with the value 2, 'Number of loci' with the value 39, and 'Missing data value' with the value -9. Each field has a black arrow pointing to it from the right. Below the fields is a 'Show data file format' button. At the bottom are three buttons: '<<Back', 'Next>>', and 'Cancel'. A black arrow points from the 'Next>>' button to a green box below the dialog.

Opens the next step of the Project Wizard

Step 3 of 4: Format of input data set

Please check box if data file contains following row(s):

→ ☒ Row of marker names

☐ Row of recessive alleles

☐ Map distances between loci

☐ Phase information

Special format

☐ Data file stores data for individuals in a single line

Show data file format

<<Back Next>> Cancel

This screenshot shows the 'Step 3 of 4: Format of input data set' dialog box. It contains a list of checkboxes under the heading 'Please check box if data file contains following row(s):'. The first checkbox, 'Row of marker names', is checked and has a black arrow pointing to it from the left. The other three checkboxes are unchecked. Below this is a section titled 'Special format' with one unchecked checkbox: 'Data file stores data for individuals in a single line'. There is a 'Show data file format' button below the checkboxes. At the bottom are three buttons: '<<Back', 'Next>>', and 'Cancel'. A black arrow points from the 'Next>>' button to a green box below the dialog.

Opens the next step of the Project Wizard

Step 4 of the Project Wizard finishes describing the format of the data set. The “Finish” button calls up a summary of the STRUCTURE project as you have set it up. The Confirmation page is an opportunity to return to previous steps by clicking on Go back to correct any incorrect entries in previous Steps

Step 4 of 4 - Project Wizard

Step 4 of 4: Format of input data set (cont'd)

Please check box if data file contains following column(s):

- ☒ Individual ID for each individual
- ☒ Putative population origin for each individual
- ☒ USEPOPINFO selection flag
- ☐ Phenotype information
- ☐ Other extra columns

Number of Extra Columns:

Show data file format

<<Back Finish Cancel

Closes the Project Wizard and opens a window that reviews all of your settings.

Confirmation

You are about to build a Structure project with following settings:

Project Name: FROG
Project Path: C:\Documents and Settings\kidd\Desktop
Data Source: C:\Documents and Settings\kidd\Desktop\FROG.TXT

Number of Individuals: 2278
Number of Loci: 39
Number of Ploidy: 2
Missing Value represented as: -9

Data File Contains Row of Marker Names
Data File Contains Individual Labels
Data File Contains Population Identifiers
Data File Contains USEPOPINFO Selection Flag
Number of Other Extra Columns: 0

Proceed Go back

When all entries are correct, “Proceed” will open the file within the STRUCTURE project

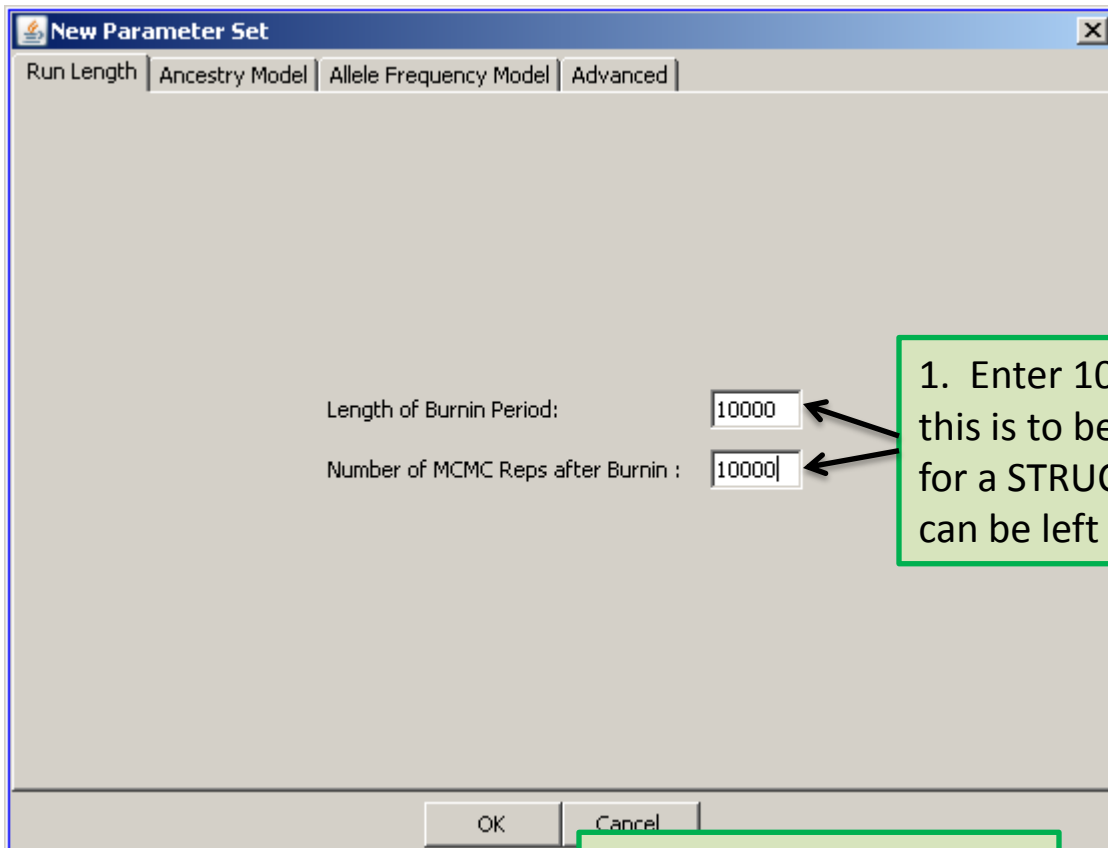
Click on Parameter Set to pull down the options for running STRUCTURE

This window shows the organization and contents of the dataset

The screenshot displays the STRUCTURE software interface. The 'Parameter Set' menu is open, showing options: 'Modify current set ...', 'New ...', 'Remove Parameter Set ...', 'Run', and 'Stop'. The 'New ...' option is highlighted. The 'Project Data' window is also open, showing a table with columns: Label, Pop ID, Flag, Locus 1, Locus 2, Locus 3, Locus 4, Locus 5, Locus 6, and Locus 7. The first row of data is highlighted in blue.

Label	Pop ID	Flag	Locus 1	Locus 2	Locus 3	Locus 4	Locus 5	Locus 6	Locus 7
JK0736	1	1	2	3	1	2	3	2	2
JK0736	1	1	2	3	1	2	3	2	2
JK0741	1	1	2	3	1	2	3	-9	2
JK0741	1	1	2	3	1	2	3	-9	2
JK1715	1	1	2	3	1	2	3	-9	2
JK0756	1	1	2	3	1	2	3	2	2
JK0756	1	1	2	3	1	2	3	2	2
JK1716	1	1	2	3	1	2	3	-9	2
JK1716	1	1	2	3	1	2	3	-9	2
JK1743	1	1	2	3	1	2	3	-9	2
JK1743	1	1	2	3	1	2	3	-9	2
JK1743	1	1	2	3	1	2	3	-9	2

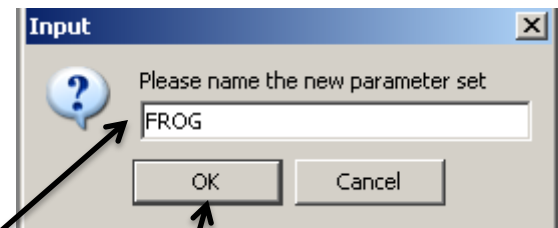
Click on "New" to open the parameter window.



In this window, specify the running parameters

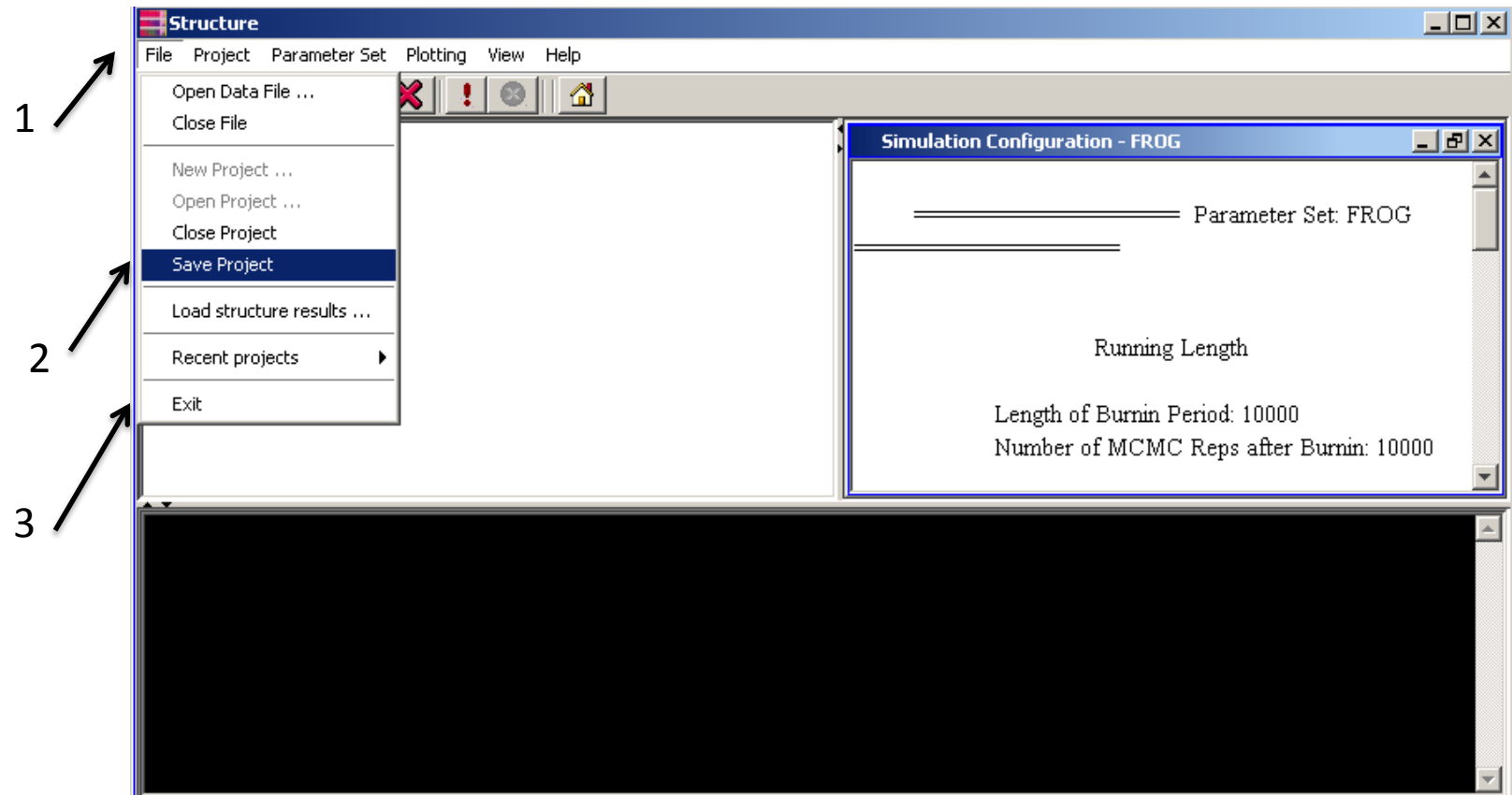
1. Enter 10000 in both of these windows; this is to be considered a minimum length for a STRUCTURE run. The remaining tabs can be left at default values for this example

2. Clicking OK opens the tiny name window; enter an arbitrary parameter name

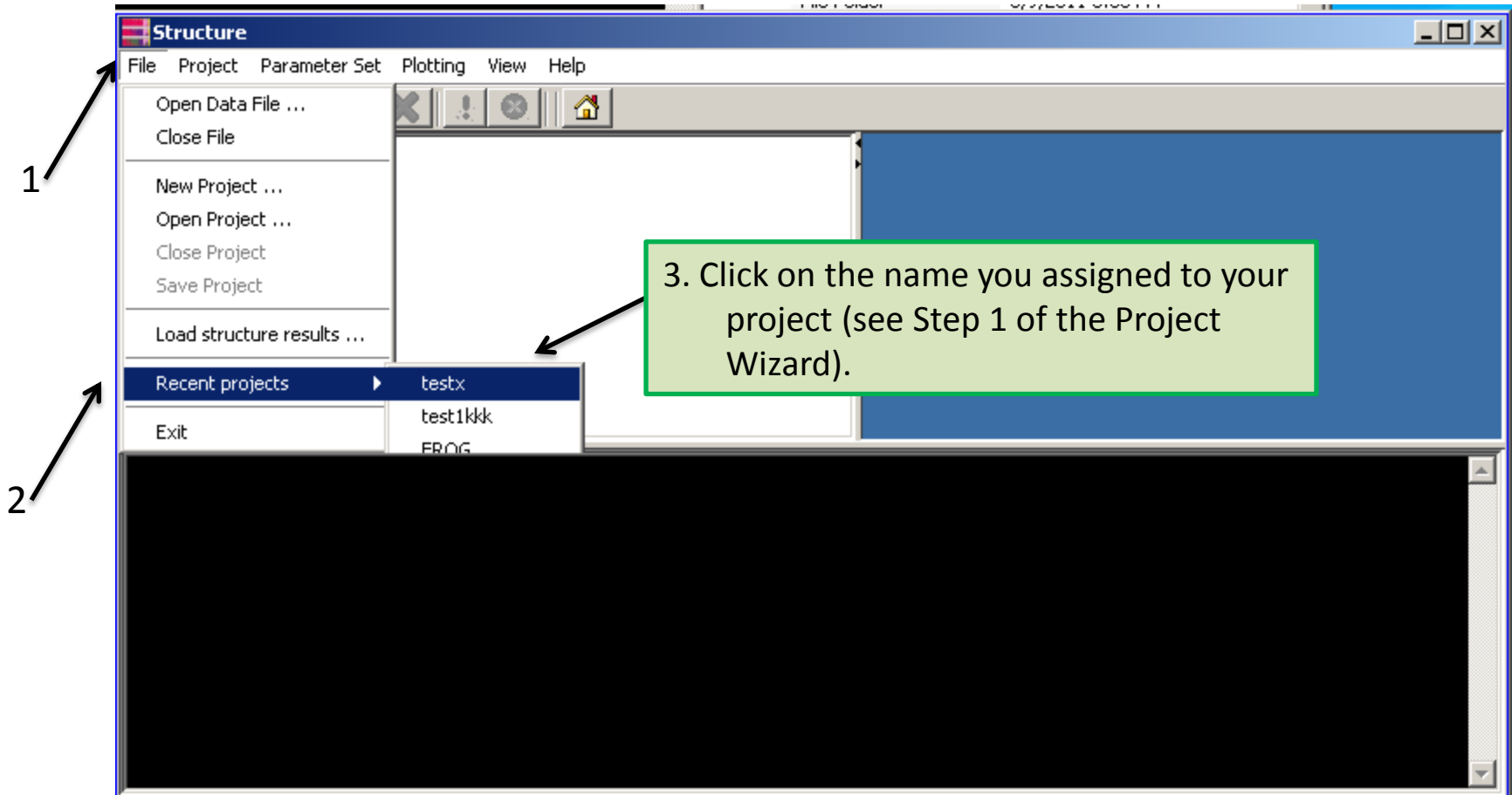


3. Returns to the main STRUCTURE page

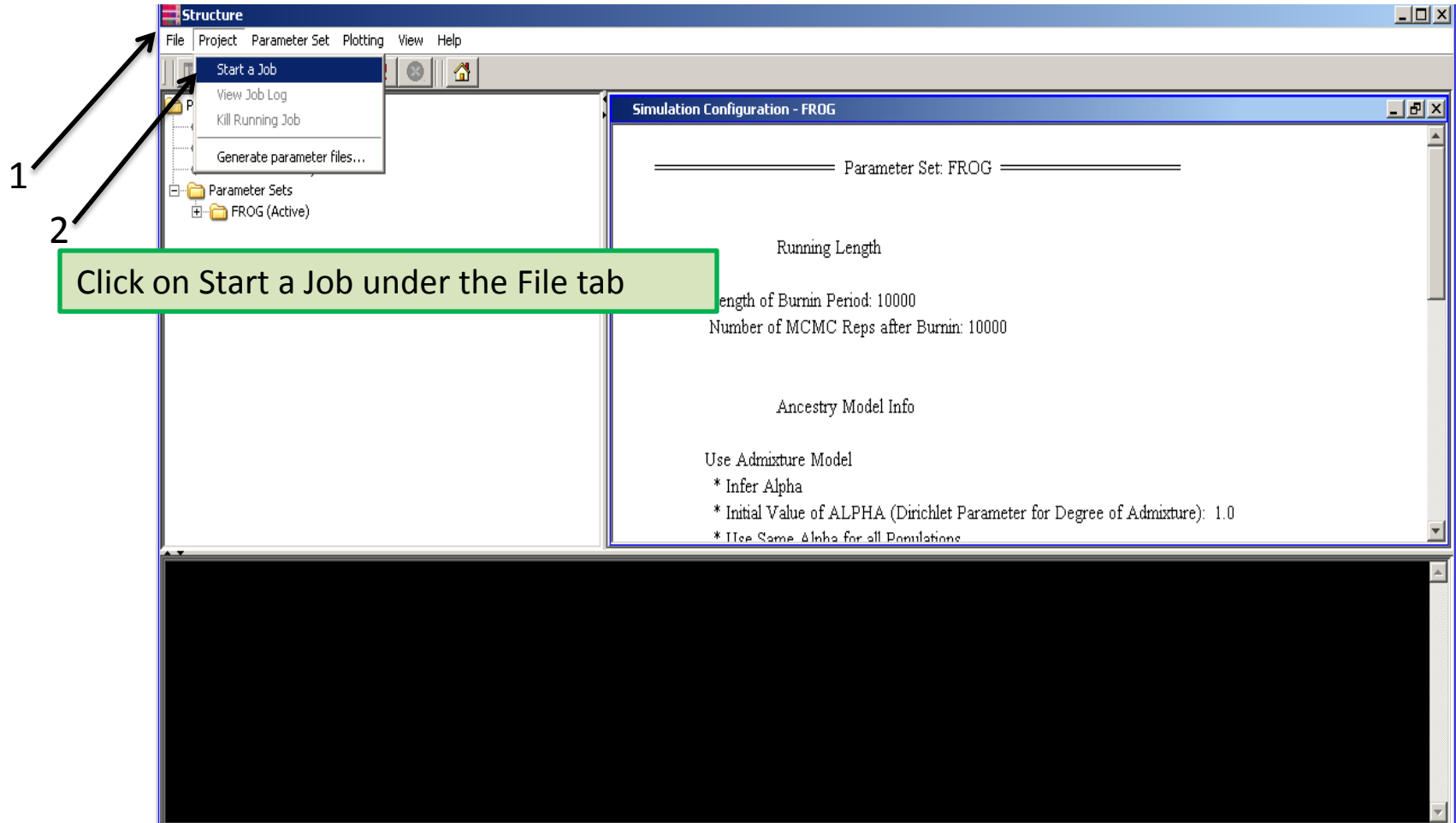
The STRUCTURE file must now be saved, closed, and reopened to run



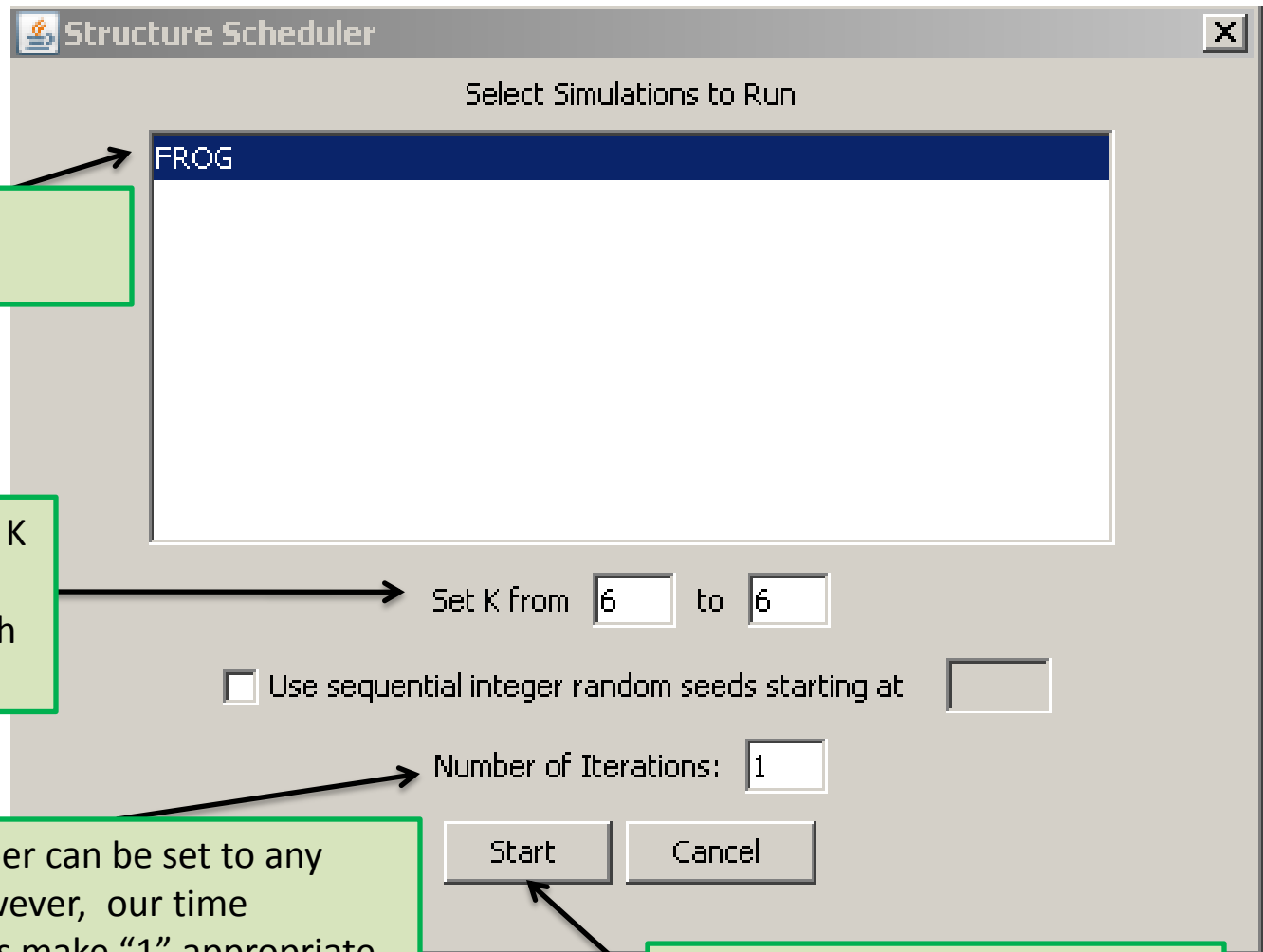
To reopen, click the STRUCTURE icon on the desktop, click on “File”, “Recent projects”, and select your project name.



Now you are ready to run your project!



These two steps move you to the select-a-job window



When the job is done, this will be your screen.

Structure

File Project Parameter Set Plotting View Help

Project - testx

- Project Data
- Project Information
- Simulation Summary
- Parameter Sets
 - FROG
 - Settings
 - Results
 - FROG_run_1 (K=6)

Project Data

Label	Pop ID	Flag	Locus 1	Locu
			1	2
JK0736	1	1	2	3
JK0736	1	1	2	3
JK0741	1	1	2	3
JK0741	1	1	2	3
JK0753	1	1	2	3
			2	3

Structure

Job is Completed !

OK

Structure Job Log

Start Job ...

Loading Simulation FROG ...

Set K = 6

Set Output Name as: FROG_run_1

Simulation FROG finished

Job is Completed!

OK

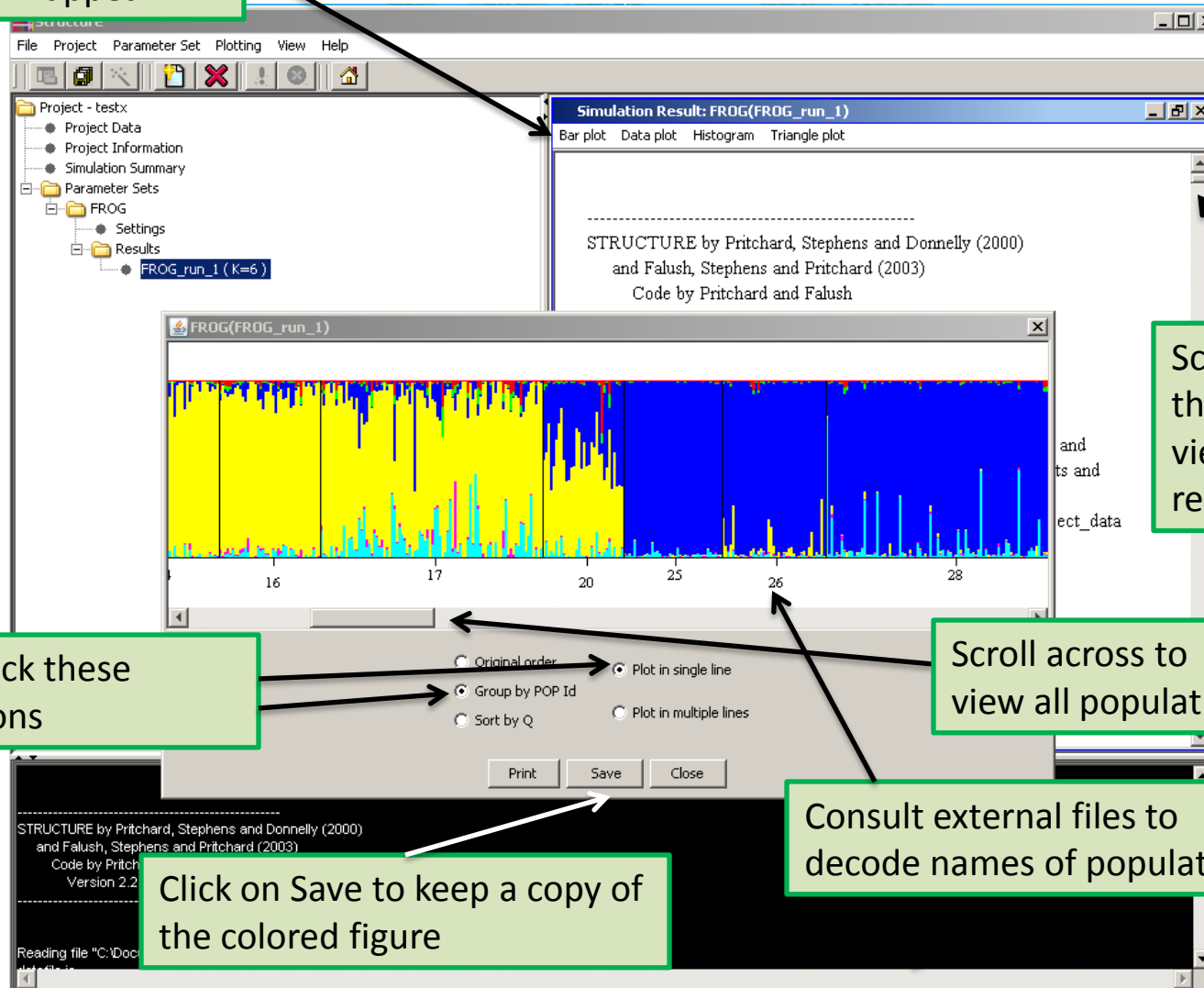
Click OK to close this window.

Click OK to close this window.

Final results printed to file C:\Documents and Settings\kidd\Desktop\testx\FROG\Results\FROG_

1. Click on "Bar plot";
figure will appear

To View the Graphic Representation of Run Results



Scroll down
this bar to
view statistical
results.

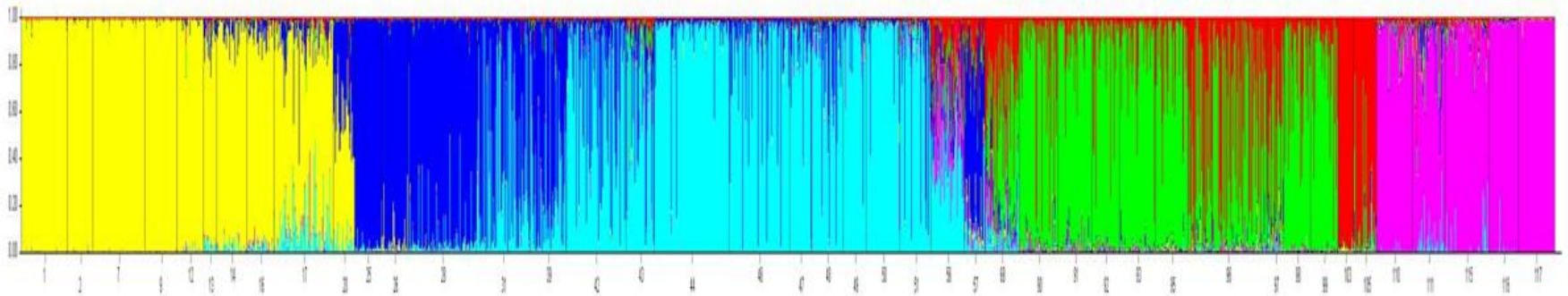
2. Click these
options

Scroll across to
view all populations

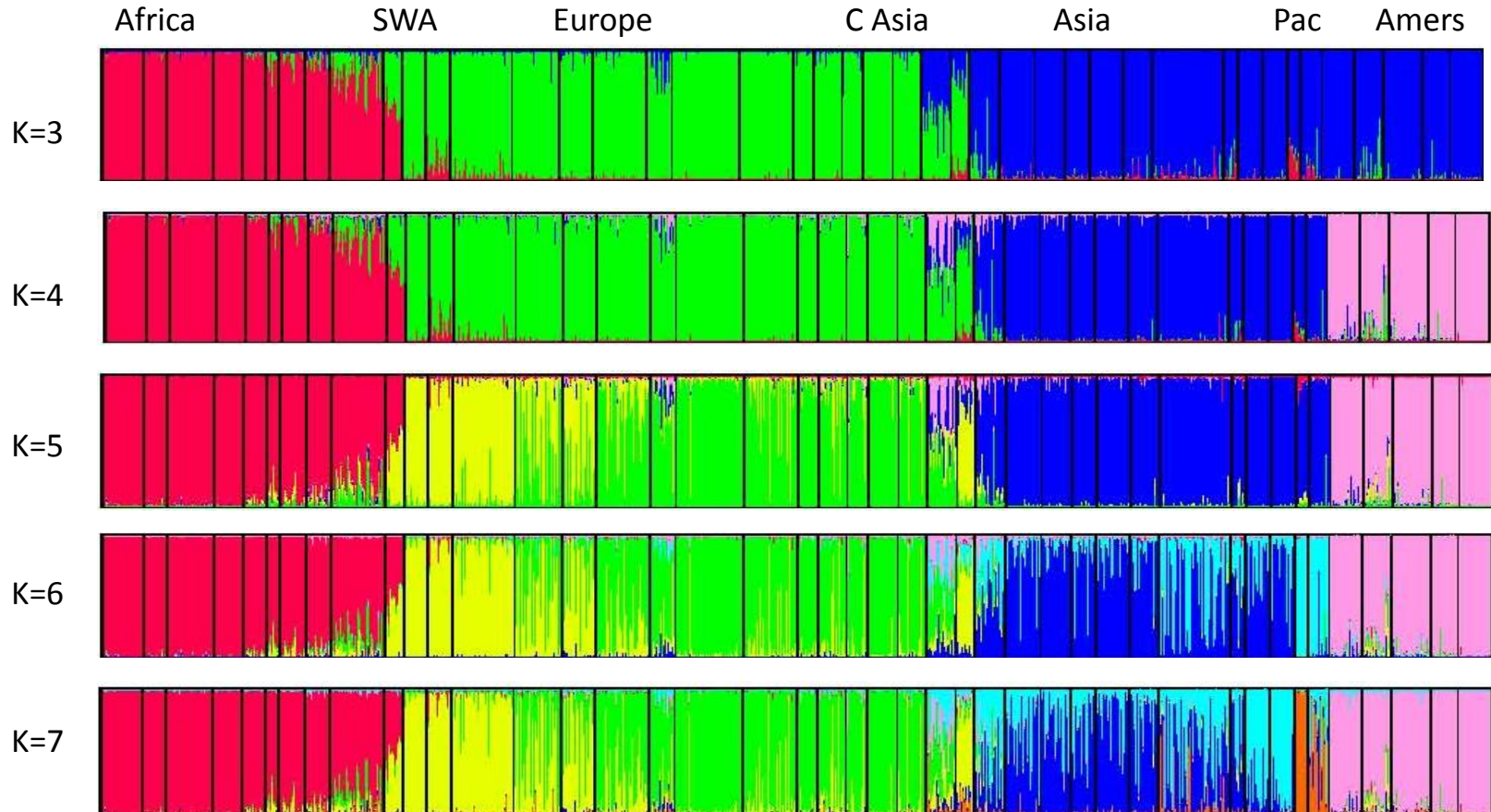
Consult external files to
decode names of populations

Click on Save to keep a copy of
the colored figure

The figure will be saved as a .jpg file which, in turn, can be inserted into a .docx file that can be labeled as desired (see next page).



The “best” D | K plots of selected 39 SNPs; 43 Populations, 2278 Individuals



Note: To co-ordinate the colors between these runs, we have used the freeware program DISTRUCT by N. Rosenberg (<http://rosenberglab.bioinformatics.med.umich.edu/distruct.html>)

Retrieving Statistical Results

1. Open the file automatically generated by STRUCTURE and carrying the name of your project. It is in the same folder as the data file

2. Open your job file

3. Open the Results file

The following table summarizes the file structure shown in the screenshots:

Folder/Path	Name	Size	Type
C:\Documents and Settings\kidd\Desktop\testx	FROG		File Folder
	project_data	409 KB	File
	testx.spj	1 KB	SPJ File
C:\Documents and Settings\kidd\Desktop\testx\FROG	PlotData		File Folder
	Results		File Folder
	.sim	1 KB	SIM File
	extraparams	1 KB	File
	mainparams	1 KB	File
C:\Documents and Settings\kidd\Desktop\testx\FROG\Results	FROG_run_1_f	161 KB	File

4. Open with a text editor your Run file

The Results are printed in a long, continuous file (broken into two and shown below). “ //... ” Indicates shortening of actual output file, leaving examples only of each part of the output file.

```

1 //
2 Run parameters:
3   2278 individuals
4   39 loci
5   6 populations assumed
6   10000 Burn-in period
7   10000 Reps
8
9
10 -----
11 Proportion of membership of each pre-defined
12 population in each of the 6 clusters
13
14 Given      Inferred Clusters
15 Pop        1      2      3      4      5      6      Number of
16                                     Individuals
17 1:      0.003  0.003  0.002  0.989  0.002  0.002      68
18 2:      0.004  0.003  0.003  0.986  0.002  0.002      38
19 7:      0.003  0.003  0.002  0.988  0.002  0.002      77
20 9:      0.003  0.003  0.002  0.987  0.002  0.002      48
21 12:     0.007  0.007  0.008  0.967  0.004  0.007      39
22 13:     0.014  0.008  0.064  0.876  0.006  0.033      20
23 14:     0.007  0.005  0.040  0.923  0.005  0.020      45
24 16:     0.013  0.008  0.045  0.899  0.010  0.025      40
25 // ...
26 Allele-freq. divergence among pops (Net nucleotide distance),
27 computed using point estimates of P.
28
29      1      2      3      4      5      6
30 1      -      0.0182  0.2105  0.2661  0.1255  0.2467
31 2      0.0182      -      0.2403  0.3025  0.1186  0.2849
32 3      0.2105  0.2403      -      0.3023  0.3185  0.0246
33 4      0.2661  0.3025  0.3023      -      0.4414  0.3739
34 5      0.1255  0.1186  0.3185  0.4414      -      0.3311
35 6      0.2467  0.2849  0.0246  0.3739  0.3311      -
36
37 Average distances (expected heterozygosity) between
38 individuals in same cluster:
39 cluster 1 : 0.2262
40 cluster 2 : 0.2060
41 cluster 3 : 0.2223
42 cluster 4 : 0.1077
43 cluster 5 : 0.1918
44 cluster 6 : 0.2040
45
46 -----
47

```

```

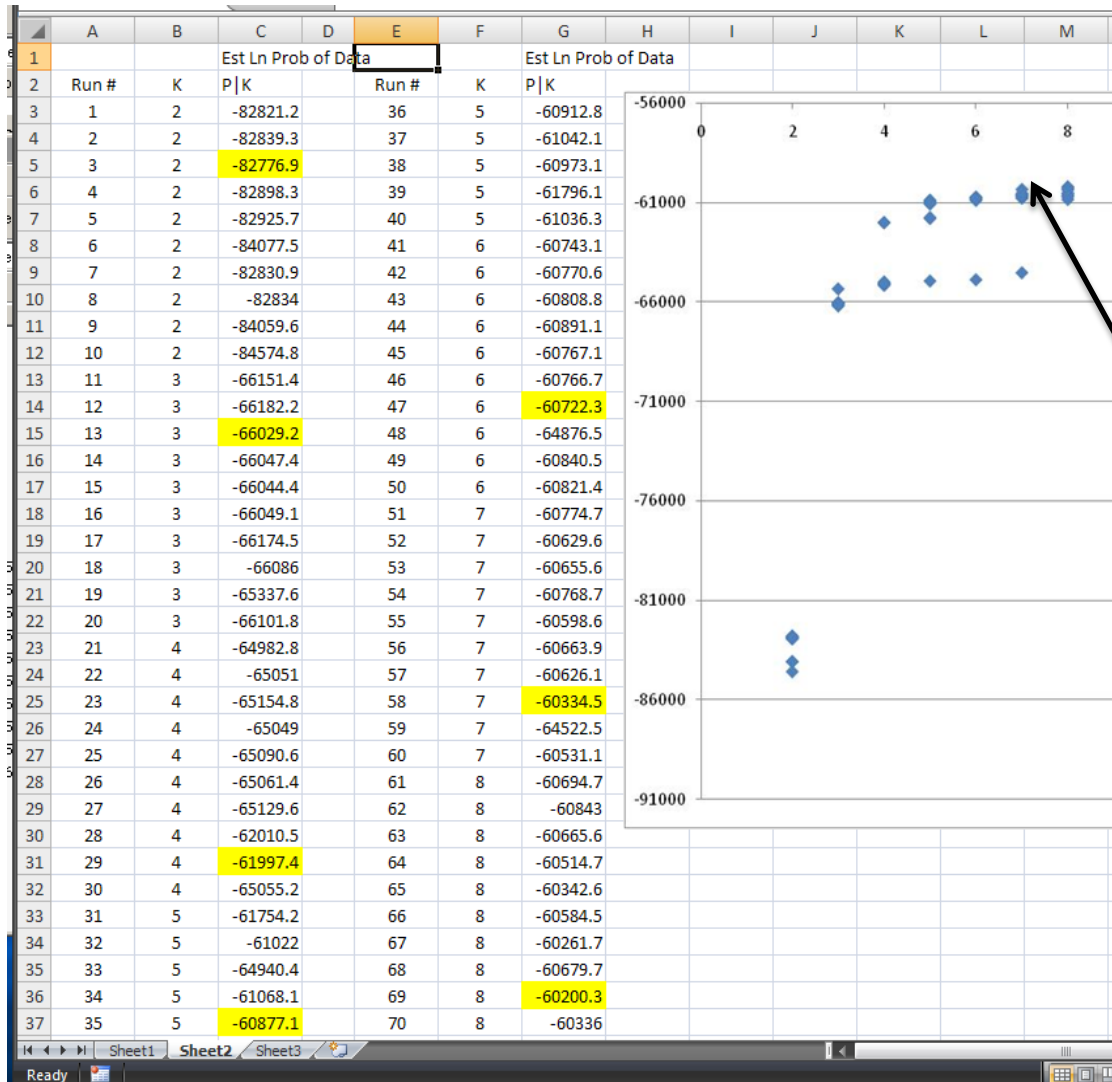
1
2 Estimated Ln Prob of Data      = -60732.5
3 Mean value of ln likelihood    = -59727.7
4 Variance of ln likelihood      = 2009.5
5 Mean value of alpha            = 0.0350
6
7 Mean value of Fst_1            = 0.4145
8 Mean value of Fst_2            = 0.4784
9 Mean value of Fst_3            = 0.3914
10 Mean value of Fst_4           = 0.6898
11 Mean value of Fst_5           = 0.5282
12 Mean value of Fst_6           = 0.4564
13 //
14
15 Inferred ancestry of individuals:
16      Label (%Miss) Pop:  Inferred clusters
17 1  JK0736 (0)  1 : 0.001 0.001 0.001 0.994 0.001 0.001
18 2  JK0741 (5)  1 : 0.003 0.002 0.002 0.990 0.001 0.002
19 3  JK0753 (0)  1 : 0.002 0.002 0.002 0.991 0.002 0.002
20 4  JK0787 (2)  1 : 0.002 0.002 0.001 0.993 0.001 0.001
21 5  JK1715 (5)  1 : 0.002 0.002 0.002 0.991 0.002 0.001
22 6  JK0756 (0)  1 : 0.001 0.001 0.001 0.994 0.001 0.001
23 7  JK1716 (5)  1 : 0.002 0.001 0.001 0.994 0.001 0.001
24 8  JK1743 (2)  1 : 0.018 0.019 0.013 0.933 0.008 0.009
25 9  JK0789 (7)  1 : 0.001 0.001 0.001 0.995 0.001 0.001
26 // ...
27 Estimated Allele Frequencies in each cluster
28 First column gives estimated ancestral frequencies
29
30 Locus 1 : 1
31 2 alleles
32 0.1% missing data
33 2 (0.097) 0.000 0.000 0.070 0.999 0.000 0.000
34 4 (0.903) 1.000 1.000 0.930 0.001 1.000 1.000
35
36 Locus 2 : 2
37 2 alleles
38 0.8% missing data
39 3 (0.526) 0.991 0.997 0.001 0.951 0.998 0.001
40 1 (0.474) 0.009 0.003 0.999 0.049 0.002 0.999
41
42 Locus 3 : 3
43 2 alleles
44 0.1% missing data
45 1 (0.772) 0.371 0.089 0.999 1.000 0.038 0.999
46 3 (0.228) 0.629 0.911 0.001 0.000 0.962 0.001
47 // ...
48

```

See next page

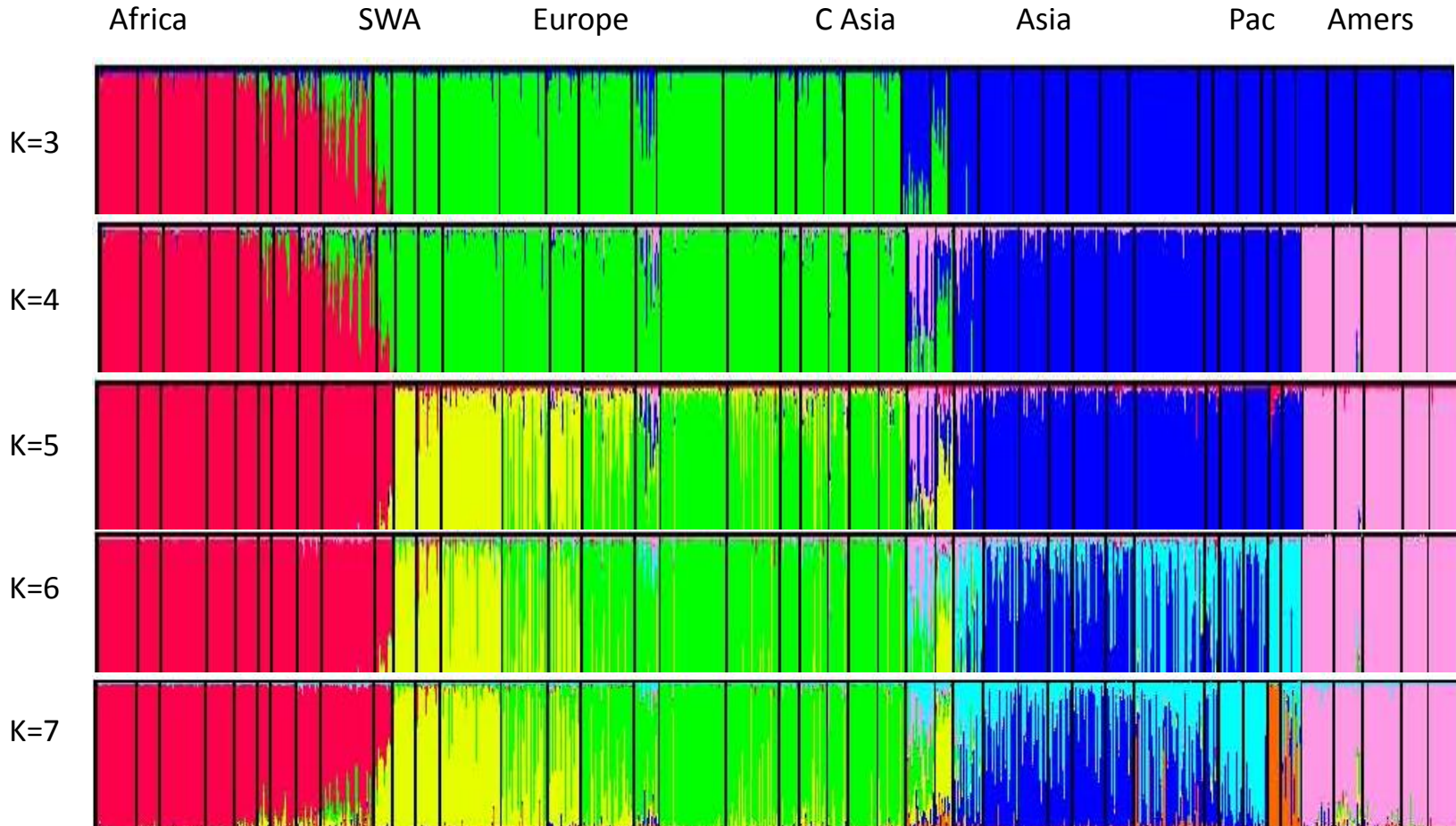
Evaluating Structure Runs

Structure Project: K=2-8, 10 replicates

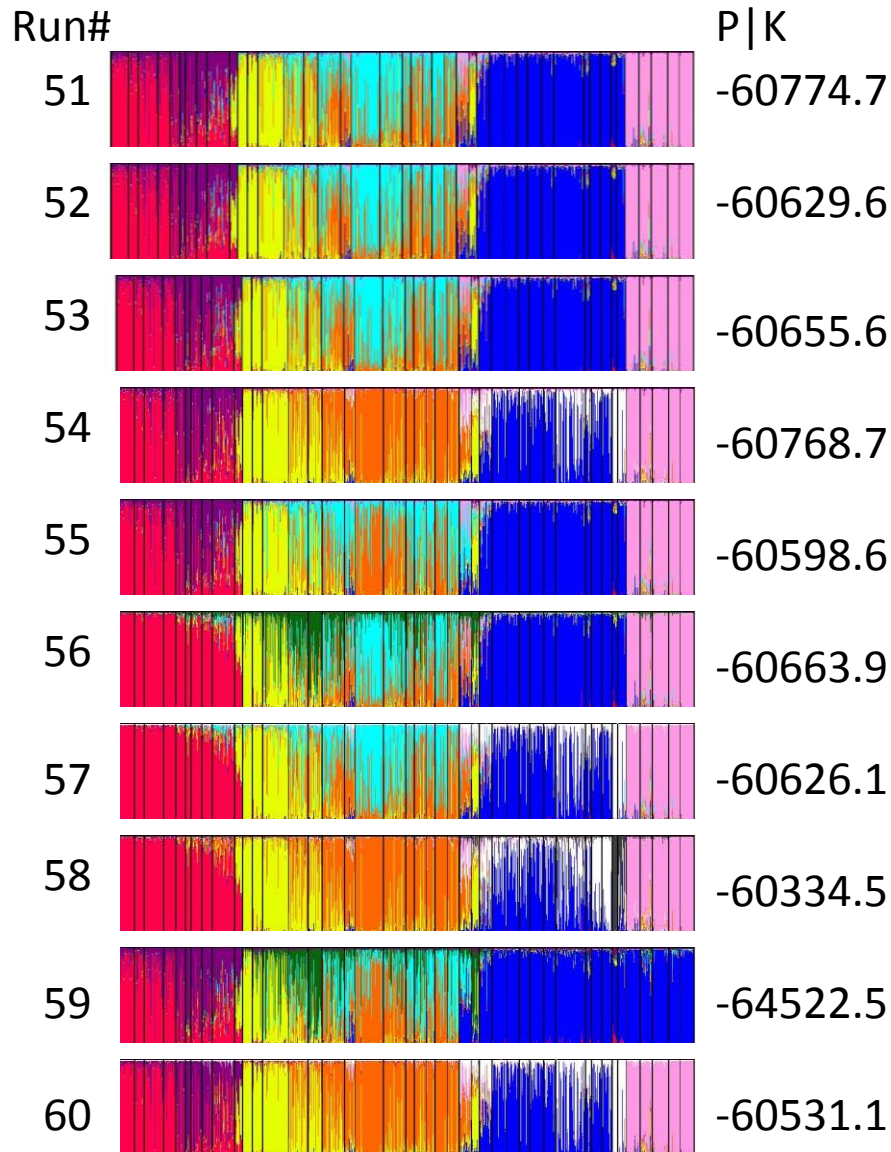


Plot the “Estimated Ln Probability of the Data” from each run against its K value. Consider the highest K value supported by the data to be that K at which the plot begins to plateau. In this example, the optimal K value is likely to be 7. There is no hard rule for the optimal K value.

The “best” D | K plots of selected 39 SNPs; 43 Populations, 2278 Individuals



Note: To co-ordinate the colors between these runs, we have used the freeware program DISTRUCT by N. Rosenberg (<http://rosenberglab.bioinformatics.med.umich.edu/distruct.html>)



Because there may well be several solutions to STRUCTURE analyses at the same K value with the same SNPS and samples, it may be instructive to plot each run at the optimal K level. In this example, run 58 is statistically the “best” solution; however, run 60 is also interesting.

NOTE: M. Jakobsson and N. A. Rosenberg have written a freeware program, CLUMP, that allows easy statistical comparison of different runs at the same K level.
<http://rosenberglab.bioinformatics.med.umich.edu/clumpp.html>.