**Perl: instructions for software installation**

- Download and install ActivePerl using default settings
  http://www.activestate.com/activeperl/

- Open a command line window. Go to Start >>Run, type `cmd` and press return.

- At the command prompt, type
  `perl --version`

  press return, and note the version of perl

- Install Bioperl. Go to

http://www.bioperl.org/wiki/Installing_Bioperl_on_Windows

see 'Installation using the Perl Package Manager' (briefly explained here)

1) Start the Perl Package Manager Graphical User Interface from the Start menu.

2) Go to Edit >> Preferences and click the Repositories tab. 'Add Repository:' referring to the table by copying and pasting the Name and Location (URL) for your version of perl, then pressing 'add'. When you're done click 'OK'.

3) Select View >> All Packages

4) In the search box type `bioperl`

5) Right-click the most recent version of Bioperl available and choose 'install'. This will mark packages for installation.

6) Click the green arrow 'Run marked actions'. Check the Status tab to make sure the installation is complete.

- Install Tie-IxHash using the Perl Package Manager
  In the search box type `Tie-IxHash`

  Right-click the most recent version of Tie-IxHash available and choose 'install'. This will mark it for installation.

  Click the green arrow 'Run marked actions'. Check the Status tab to make sure the installation is complete. Exit the Perl Package Manager.

**Perl: instructions for running**

- To run the Perl script, create a working directory on the C drive (e.g., tomatoPCA) containing your data file in FASTA format (infile.fasta) and the perl script (AlleleCoder.pl). [Note: the data used to generate Fig. S1 are included in supplementary file 'FileS3-infile.fasta'. This can be used to test successful installation of Perl, R are R Commander. 'FileS3-infile.fasta' corresponds to raw data provided in supplementary file 'FileS4-SNP-matrix.xls'.]
- On Windows desktop go to start > Run
- In the 'Open:' field, type cmd and press 'OK'
- A command line window will appear. Change to the working directory by typing the name of the drive followed by a colon

  ```
  C:
  ```

  and press enter. Now, type

  ```
  cd tomatoPCA
  ```

  then enter. Then, at the prompt type

  ```
  AlleleCoder.pl infile.fasta > 123.txt
  ```

  Press enter. The output '123.txt' file will contain the coded genotypic data for input to R commander for PCA.

**R: instructions for software installation**

- Download and install R using default settings
  http://www.r-project.org/

- Launch R from the Start Menu

- Install R Commander (the graphical user interface). Type

  ```
  install.packages("Rcmdr", dependencies=TRUE)
  ```

  at the prompt and press return. Installation is complete.

**R: instructions for running**

- Launch R from the Start Menu

- Load R commander. This must be done each time you run it. Type

  ```
  library(Rcmdr)
  ```

  at the prompt and press return. A window will open that contains a Script Window, Output Window, and Messages.

- Select Data >> Import data >> from text file. The 'Read Text Data From File' dialog box will pop up, click 'OK' to accept the default settings. Browse for 123.txt created by the Perl script. Click the 'View data set' button to review your input file. Make sure it appears as you expect (note: numbering starts from zero), then close it.

- Select Statistics >> Dimensional analysis >> Principal-components analysis
  Select the variables to include (usually all). Uncheck all boxes and click 'OK'.

  (Note: if the dataset is sufficiently small, you can use this output from princomp and skip using prcomp. If it is rejected, use prcomp).

- Edit the text in the 'Script Window' to change princomp to prcomp. Highlight the text in the script window from .PC <— to ...data=Dataset) then click the 'Submit' button.

- In the script window highlight and submit

  ```
  summary(.PC)
  ```

  This outputs the standard deviation, proportion of variance and cumulative proportion of variance for each PC.

- In the script window type in (or copy and paste), highlight and submit

  ```
  .PC$x
  ```

  This outputs the contribution of each PC to each genotype. This is what will be plotted. Highlight and copy the first three or more PCs including the headers and row names and paste into a Wordpad document. Save as a Text Document file named 'scores.txt'.

- In the script window type in (or copy and paste), highlight and submit

  ```
  .PC$rotation[,1:3]
  ```

This outputs the contribution of each marker to the first three PCs. [Note: if there are more than three PCs, change :3 to :n where n is the number of PCs].

- Select Data >> Import data >> from text file. The 'Read Text Data From File' dialog box will pop up, click 'OK' to accept the default settings, 'OK' to overwrite current data file, and browse for scores.txt created above.

- Now you are ready to graph the results. Select Graphs >> 3D graph >> 3D scatterplot. Select response variable PC1 and explanatory variables PC2, PC3. Uncheck all boxes and click 'OK'.

- Some predefined sets of options for graph colors have been coded. In WordPad open 'FileS5-Rcommander-options.txt'. This file contains predefined sets of parameter options for viewing the PCA 3D graphical display. Copy one of the lines (without the >) and paste it into the script window. Highlight and submit.

- To visualize groups of individuals (e.g. populations), using a spreadsheet, insert a column of group identities into the scores.txt file. Save as 'groups-scores.txt'

- Select Data >> Import data >> from text file and browse for groups-scores.txt. Click the 'View data set' button to review your input file.

- Select Graphs >> 3D graph >> 3D scatterplot. Select response variable PC1 and explanatory variables PC2, PC3. Uncheck all boxes, click 'plot by groups' and click 'OK'.