

# TESS Ad-Mixer: A novel program for visualizing TESS $Q$ matrices

Matthew W. Mitchell · Brian Rowe ·  
Paul R. Sesink Clee · Mary Katherine Gonder

Received: 31 March 2013 / Accepted: 19 June 2013 / Published online: 10 August 2013  
© Springer Science+Business Media Dordrecht 2013

**Abstract** TESS Ad-Mixer (available at <http://www.albany.edu/faculty/gonder/Lab/index.html>), is a Windows program we developed to display spatial interpolations of  $Q$  matrices generated by the program TESS. TESS Ad-Mixer provides an easy way to create two-dimensional representation of the  $Q$  matrices for two or more genetic clusters. The program uses a pixel based, vector algorithm that allows the user to specify a color scheme. The program generates spatially accurate, high-resolution files that are ideal for data display.

**Keywords** Population structure · Population genetics · TESS · Visualization · Spatial ·  $Q$  matrix

## Introduction

Deciphering patterns of population structure and individual ancestry remain fundamental problems of population genetics (Endler 1977; Cavalli-Sforza et al. 1994; François and Durand 2010). Bayesian methods are widely used to infer population structure, which are implemented in number of programs (reviewed in François and Durand 2010). Many of these approaches use multi-locus genotype correlations and solve, using a Markov Chain Monte Carlo (MCMC) method (with the exception of BAPS, which uses a maximization algorithm (Corander et al. 2008)), for the most likely proportions of an individual's genetic ancestry ( $q$ ) in two or more ( $k$ ) genetic clusters. For all of the

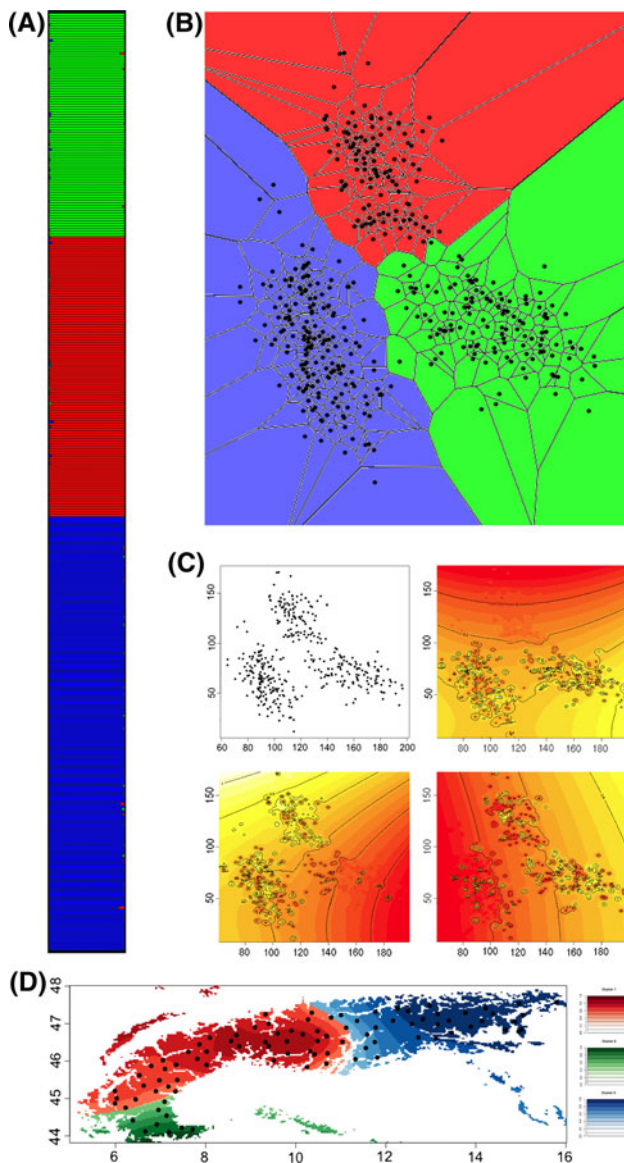
individuals in a given study, these values can be represented in a matrix called the  $Q$  matrix.

Several of these programs incorporate geospatial data from individuals or sampling localities, for example GENELAND (Guillot et al. 2005), BAPS5 (Corander et al. 2008), POPS (Jay 2011), and TESS (Chen et al. 2007; Durand et al. 2009). These programs allow the user to output their results as a bar plot (Fig. 1a) and to use geospatial data to map population structure across a study area. However, these programs offer limited options for displaying multiple clusters in a single image or to display overlap between them. For example, the TESS hard-clustering method represents individuals into cells, which are shaded a single color (Fig. 1b) representing their membership in a single cluster. This hard-clustering method makes it impossible to represent overlap between clusters. In addition, Universal Kriging analysis (Ripley 1981) has been used to interpolate population structure across sampled and unsampled parts of a study area (*R* script for Universal Kriging method available at: [http://membres-timc.imag.fr/Olivier.Francois/admix\\_display.html](http://membres-timc.imag.fr/Olivier.Francois/admix_display.html)). Problems with this Universal Kriging method include: (1) interpolated fractions of ancestry for each cluster that are plotted on separate maps; and (2) color displays cannot be changed (Fig. 1c). The program POPS also provides methods to summarize maps based on Universal Kriging interpolation scripts available as an *R* script (Jay et al. 2012) (Fig. 1d).

The program TESS also contains an option to create ASCII based posterior predictive maps of admixture proportions, but offers no way to display them. TESS generates a single ASCII file for each cluster specified by the user during any particular TESS run (Fig. 2a). These ASCII files are two-dimensional predictions of the  $Q$  matrix given a simulated coordinate space of  $n$  pixels.

---

M. W. Mitchell (✉) · B. Rowe · P. R. Sesink Clee ·  
M. K. Gonder  
University at Albany, State University of New York,  
Albany, NY, USA  
e-mail: mwmitchell@albany.edu



**Fig. 1** Graphical representation of example data included with the TESS (a–c) and POPS (d) distribution packages. **a** A bar plot assuming three populations created using DISTRUCT (Rosenberg 2004). **b** Color-coded Voronoi cells from TESS. **c** Four images created using an R script using Universal Kriging to display the spatial distribution of three different clusters. **d** Map of admixture coefficients using POPS example data, created using Universal Kriging interpolation scripts in R

For  $k$  clusters assumed by TESS, the posterior predictive mapping function outputs  $k$  files, each with predictions of probable ancestry for each pixel on the map. In this paper, we present a new post hoc method for mapping the structure of multiple clusters simultaneously based on individuals'  $q$  values and predictive maps. This method is

implemented in the program, TESS Ad-Mixer, which is a Clojure program that runs on a Java virtual machine (JVM).

## Functionality description

### User inputs

User inputs in the program include providing  $k$  ASCII files (representing spatial interpolations of the  $Q$  matrix) and  $k$  color codes (represented by RGB values) that will visually distinguish each of  $k$  clusters determined by TESS. The  $k$  ASCII files will be notated as  $Q_1, Q_2, \dots, Q_k$ . Spatially predicted  $q$  values, inferred from the  $Q$ -matrix, are denoted as  $Q_i(x, y)$ , where  $i$  is a value between 1 and  $k$ , and  $(x, y)$  ranges over the study area.

### Data normalization

TESS Ad-Mixer layers  $k$  ASCII grids onto a single image, while also mixing the colors of corresponding cells from each of  $k$  ASCII grids into proportional amounts that accurately reflect their  $q$  values. Each of the cells,  $Q_i(x, y)$ , from each of the  $k$  ASCII grids, contains a  $q$  value pertinent to a single particular geographic location,  $(x, y)$ . For each  $(x, y)$ , a vector of the normalized associated data,  $q_{\text{norm}}^{\rightarrow}$ , is calculated so that each cell displays a proportional amount of  $k$  colors specified by the user.

$$q_{\text{norm}}^{\rightarrow}(x, y) = \frac{1}{\sum_{i=1}^k Q_i(x, y)} \begin{pmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_k \end{pmatrix} \quad (1)$$

In order to avoid dividing by 0, if:

$$\sum_{i=1}^k Q_i(x, y) < \epsilon \quad (2)$$

where,  $\epsilon$  is a near 0 value, then:

$$q_{\text{norm}}^{\rightarrow} = 0 \quad (3)$$

### Color computation

In order to generate the output image,  $q_{\text{norm}}^{\rightarrow}$  and the RGB code values specified for each cluster are integrated for each pixel.

$$R = q_{\text{norm}}^{-\rightarrow T} \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_k \end{pmatrix} \tag{4}$$

$$G = q_{\text{norm}}^{-\rightarrow T} \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_k \end{pmatrix} \tag{5}$$

$$B = q_{\text{norm}}^{-\rightarrow T} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix} \tag{6}$$

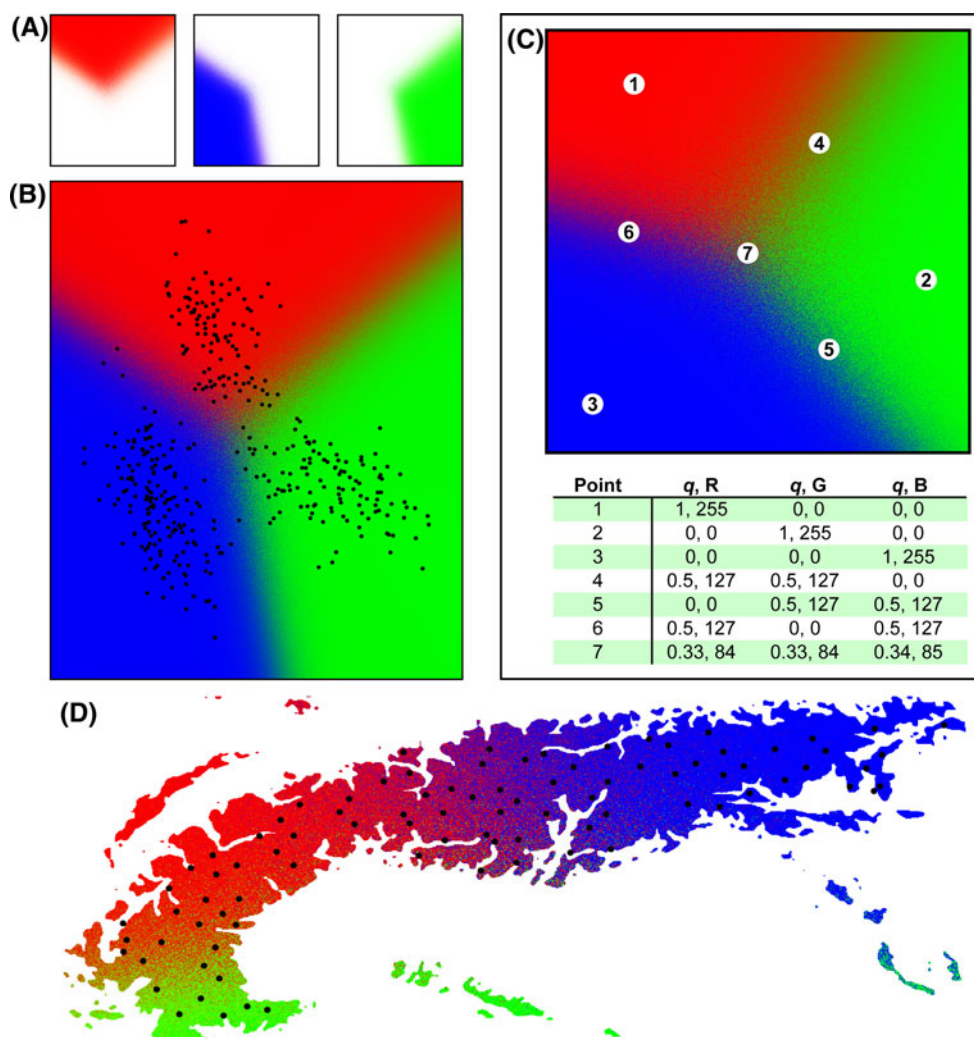
where each  $r_i, g_i, b_i$ , are the color components (red, green and blue, respectively) previously specified by the user, designated for the  $i$ th cluster, and  $R, G, B$  are the color components for this pixel in the final image. The program computes  $R, G, B$  for every pixel designated by the input

ASCII grids. These pixel values are used to construct a PNG image using Java’s BufferedImage class. The resulting PNG image (Fig. 2b) is a complete representation of the spatially interpolated  $Q$  matrix as generated by TESS.

### Advantages of using TESS Ad-Mixer

TESS Ad-Mixer improves upon existing methods for representing the spatial distribution of population structure. Specifically, TESS Ad-Mixer: (1) allows users to choose colors for different clusters; (2) generates a single image of multiple clusters from TESS posterior predictive ASCII files; and (3) shows areas of cluster overlap from  $q$  values, which are represented by color mixing the pixels in question (Fig. 2c). Color mixing is especially useful for displaying adjacent clusters with large degrees of overlap (Fig. 2d) compared to methods based on Universal Kriging (Fig. 1d). The single output image generated by TESS Ad-

**Fig. 2** The process of using TESS Ad-Mixer. **a** Visual representations of admixture proportions of the TESS posterior predictive maps. **b** Output image combining the posterior predictive maps into a single image using TESS Ad-Mixer (sample points were added using ArcMap). **c** Close up view of the TESS Ad-Mixer output file showing *example points* and a *corresponding table*. The table shows each example point’s expected  $q$  value for each of the three clusters, and its corresponding  $R, G,$  and  $B$  values. **d** Output image combining the posterior predictive maps into a single image using POPS example data (Jay 2011) for  $K = 3$ . Map extent was clipped using a shapefile from Jay et al. (2012), and sample points were added using ArcMap). For comparison to POPS,  $R$  spatial interpolation for these data see Fig. 1d



Mixer can be georeferenced to maps using programs, such as ArcMap (ESRI Corp.).

**Acknowledgments** We thank Eric Durand for providing assistance in understanding the functionality of TESS, and the option to create posterior predictive maps of admixture proportions. We thank the University at Albany—State University of New York for hosting the program on their web server. NSF awards 0755823 and 1243524 to MKG supported this work.

## References

- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton
- Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Mol Ecol Notes* 7(5):747–756. doi:[10.1111/j.1471-8286.2007.01769.x](https://doi.org/10.1111/j.1471-8286.2007.01769.x)
- Corander J, Marttinen P, Sirén J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* 9:539. doi:[10.1186/1471-2105-9-539](https://doi.org/10.1186/1471-2105-9-539)
- Durand E, Jay F, Gaggiotti OE, François O (2009) Spatial inference of admixture proportions and secondary contact zones. *Mol Biol Evol* 26(9):1963–1973. doi:[10.1093/molbev/msp106](https://doi.org/10.1093/molbev/msp106)
- Ender JA (1977) Geographic variation, speciation, and clines. Princeton University Press, Princeton
- François O, Durand E (2010) Spatially explicit Bayesian clustering models in population genetics. *Mol Ecol Resour* 10(5):773–784. doi:[10.1111/j.1755-0998.2010.02868.x](https://doi.org/10.1111/j.1755-0998.2010.02868.x)
- Guillot G, Mortier F, Estoup A (2005) GENELAND: a computer package for landscape genetics. *Mol Ecol Notes* 5(3):712–715. doi:[10.1111/j.1471-8286.2005.01031.x](https://doi.org/10.1111/j.1471-8286.2005.01031.x)
- Jay F (2011) POPS: prediction of population genetic structure—program documentation and tutorial. University Joseph Fourier, Grenoble
- Jay F, Manel S, Alvarez N, Durand E, Thuiller W, Holderegger R, Taberlet P, François O (2012) Forecasting changes in population genetic structure of alpine plants in response to global warming. *Mol Ecol* 21(10):2354–2368. doi:[10.1111/j.1365-294X.2012.05541.x](https://doi.org/10.1111/j.1365-294X.2012.05541.x)
- Ripley BD (1981) Spatial statistics. Wiley, New York
- Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* 4(1):2. doi:[10.1046/j.1471-8286.2003.00566.x](https://doi.org/10.1046/j.1471-8286.2003.00566.x)