



FROM THE CLASSROOM TO AN OPINION NOTE: COMPLEMENTARY ANALYSIS OF THE GENETIC STRUCTURE OF THE NEOTROPICAL TREE *MANILKARA ZAPOTA* (L) P. ROYEN (SAPOTACEAE)

Barrandeguy, M. E.^{1,2,3}, D. J. Sanabria^{3,4}, M. V. García^{1,2,3}

¹ Laboratorio de Genética de Poblaciones y del Paisaje. Facultad de Ciencias Exactas, Químicas y Naturales. Universidad Nacional de Misiones; Posadas 3300 Misiones, Argentina.

² Instituto de Biología Subtropical - Nodo Posadas (UNaM – CONICET).

³ Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET - Argentina).

⁴ Laboratorio de Biología Molecular Aplicada. Facultad de Ciencias Exactas, Químicas y Naturales. Universidad Nacional de Misiones.

ABSTRACT

Here, we describe a learning strategy that results an excellent choice for a first approach of students to produce scientific knowledge that can be confronted in the scientific field as well as recognize in this knowledge the transferability to the natural resources management. Nowadays, the availability of several Population Genetics software together with public molecular database represents a valuable tool of great assistance for teachers of this discipline. In this way, we implemented a lecture where the students worked with empirical data set from a recent published article. The students joined theoretical concepts learned, computational software free available and empirical data set. The development of the activity comprised four steps: i) estimate population genetics parameters using software recommended by teachers, ii) understand results in a biological sense, iii) read the original manuscript from dataset authors and iv) compare both results in a comprehensive way. The students assumed the challenge under a reflective look and they kept a very fruitful discussion playing a role of population geneticists. Their exchange of ideas allowed them arrive to the conclusion that *Manilkara zapota* populations keep high levels of genetic diversity, although Ancient Maya left traces in the genetic makeup of these non-native populations with different management histories.

KEYWORDS: Learning strategy; Population Genetics; *Manilkara zapota*; Genetic structure; Bayesian analysis; Genetic representativeness

INTRODUCTION

The main aim of this work is to share the teaching experience in the undergraduate course of Population and Quantitative Genetics. This experience consisted in the implementation of special learning strategy for undergraduate students trying to help them to solve the gap between the unexciting theory and the exciting application field of Population Genetics.

The course of Population and Quantitative Genetics includes basic contents for the students of Bachelor in Genetics although this discipline annoys them because of it comprises a lot of mathematical and statistical concepts. Traditionally several equations from theory of Population Genetics are implemented for estimate population parameters. Nowadays, the availability of several Population Genetics software together with public molecular database represents a valuable tool of great assistance for teachers of this discipline. In this way, we implemented a lecture where the students worked with empirical data set from a recent published article. The students joined theoretical concepts learned, computational software free available and empirical data set during the proposed activity.

The development of the activity comprised four steps: i) estimate population genetics parameters using software recommended by teachers, ii) understand results in a biological sense, iii) read the original manuscript from dataset authors and iv) compare both results in a comprehensive way.

Following results of complementary data analysis are described in a way of an opinion note:

The original paper titled “Genetic variation and structure in the neotropical tree, *Manilkara zapota* (L) P. Royen (Sapotaceae) used by the ancient Maya” by Thompson et al. (2015) quantified patterns of genetic variation and structure in this species to discuss whether the patterns were consistent with the hypothesis that the ancient Maya cultivated *M. zapota* in gardens or managed forest patches, resulting in reduced levels of polymorphism and restricted gene flow. To test this hypothesis their objectives were to (1) compare levels of genetic variation in populations of *M. zapota* with distinct management histories, consisting of three forested ancient Maya sites, home gardens, and a set of clonally propagated cultivars, (2) characterize genetic differentiation among populations of *M. zapota* from Guatemala and Belize, compared to other similar tropical tree species. They examined these genetic patterns to assess whether they were consistent with ancient Maya influence or the natural history of the species. As conclusion of their study “they found no evidence for reduced genetic diversity and increased differentiation in current forest and garden populations that would have accompanied the active selection of fruit traits through cultivation events” and they also conclude that “while it is still possible that ancient Maya practices may have influenced the current genetic composition and structure of *M. zapota* the observed low levels of differentiation most likely derive from naturally high levels of gene flow in forests densely populated with *M. zapota*, as seen in other tropical tree species”.

In addition to the same analyses performed by Thompson et al. (2015) complementary analyses were performed. These complementary data analyses supported that the ancient Maya utilization of *M. zapota* left traces on the genetic structure even though it did not reduce the levels of genetic diversity.

MATERIALS AND METHODS

Molecular data

The raw genetic data from nine microsatellite markers available in DRYAD public database (Thompson et al., 2015b) were used to analyze population genetic structure and genetic representativeness of *M. zapota* populations.

Data analysis

The allelic frequencies for each locus were estimated using Genalex 6.4 (Peakall and Smouse, 2012). As results of large differences in the size of sampled populations allelic richness (R) was estimated and statistical differences among allelic richness of management units were estimated by the implementation of simulations. Both analyses were developed using FSTAT 2.9.3.2 (Goudet, 1995). Finally, an analysis of representativeness of gene pool genetic variation was performed in order to determine the difference between each population gene pool and its complement using Gregorius genetic differentiation index (*D*) (Gregorius, 1984).

Complement means the total gene pool defined by all populations without the gene pool of studied population. Also, a population differentiation index (δ) (Gregorius and Roberds, 1986) was estimated in order to show the average differentiation among populations. This analysis was performed using GSED 3.0 (Gillet, 2010).

The population genetic structure was described using the Bayesian theory. Within a given data set the proportion of individuals correctly assigned to each population can provide useful insights regarding to relative patterns of population genetic structure (Manel et al., 2005). This model-based method estimates the number of genetic clusters (K) and assigns the total number of individuals to these clusters (Pritchard et al., 2000). Bayesian analysis was performed using the admixture model with correlated allele frequencies between populations using the localization of sampled population as *a priori* information (LOCPRIOR). Cultivars were excluded because of the clonally origin of these individuals. The number of genetically different clusters (K) ranged from 1 to 5. The model was run with 10 independent simulations for each K using a *burn-in* length of 50,000 and a run length of 500,000 MCMC iterations. Other parameters were set to default values. This analysis was performed using STRUCTURE 2.3.3 (Pritchard et al., 2000). To infer the most likely value of K the *ad-hoc* ΔK statistics based on the second order of change in the log likelihood of data (ΔK) as a function of K calculated over 10 replicates (Evanno et al., 2005) was applied using the Web-based tool STRUCTURE HARVESTER (Earl and Von Holdt, 2012). Expected heterozygosity (*He*) and *F_{ST}* for each cluster were reported from STRUCTURE analysis results.

Two indirect methods were implemented to estimate gene flow. In the first indirect method, from Wright's fixation index (F_{ST}) the estimation of effective number of immigrants per generation (Nem), were made with the following equation $Nem = \frac{1}{4} (\frac{1}{F_{ST}})$ (Wright, 1951) where N_e is the effective number of individuals in the population and m is the immigration rate, while in the second indirect method gene flow was estimated using Slatkin's approach of private alleles $p(1)$ (Slatkin, 1980; Slatkin, 1981). This author defined $p(1)$ as the mean frequency of alleles found in only one sample (Slatkin, 1985). Using simulations under an island model, Slatkin (1985) shows a lineal relationship between $\log_{10}[P(1)]$ and $\log_{10}(Nem)$. Gene flow (Nem) was estimated from private allele method including a correction for sample size by implementing Genepop 4.0.10 (Raymond and Rousset, 1995).

RESULTS

Allele frequencies for each polymorphic locus are showed in **Figure 1**. There are shared alleles among populations and these alleles showed the highest frequencies. There were no statistically significant differences among allelic richness of different management units (**Table 1**). The snail graphics from genetic diversity representativeness are showed in **Figure 2**. Averages differentiation among populations were 0.0811 and 0.1098 for proportional to sample sizes and corrected for equal sample sizes analysis, respectively. From both analyses, ancient Maya sites show more representative gene pool than home gardens and cultivars.

The most likely cluster number was identify by Evanno method as $K = 2$ (Figure 3). Individual assignment tests over two genetic groups indicated a main cluster which includes a great number of individuals from Ancient Maya sites and all individuals from home gardens (**Table 2**). There was no coincidence between geographical origin of individuals and the cluster to which they were assigned. For both clusters expected heterozygosities were high and fixation index for each cluster indicated moderate genetic differentiation (Wright, 1978) (Table 2). Evanno method may be overly conservative in species with significant gene flow (Dewoody et al., 2015) as consequence of this barplots with partition results [$K = 2 - 5$] are showed in **Figure 4**.

Fixation index reached a value of 0.01 ($p < 0.001$) when all populations were analyzed and $F_{ST} = 0.06$ ($p < 0.001$) when the analysis included only the ancient Maya sites (Thompson et al., 2015). From these results gene flow was 17.61 and 3.92 among all populations and among ancient Maya sites, respectively. Gene flow estimated using private allele method reached a value of 9.11 and 17.90 among all populations and among ancient Maya sites, respectively. All results confirm the presence of high levels of gene flow among all the studied populations independently its management history.

DISCUSSION

The students assumed the challenge under a reflective look and they kept a very fruitful discussion playing a role of population geneticists. Their exchange of ideas allowed the following discussion:

Bayesian analysis using STRUCTURE software determines genetic structure and assigns individuals to homogenous genetic groups. These characteristics become it in the most useful and frequent software used in population genetics studies. The present analysis was performed as laboratory lecture for the Population Genetics undergraduate course 2016. Presence of ancient Mayas traces was determined on genetic structure of *M. zapota* populations from this new data treatment.

Despite Thompson et al. (2015) recognized gene flow as the principal micro-evolutionary process responsible for genetic variation distribution it was no estimated in their work. Allelic frequencies distribution reflects the homogenizing role of gene flow and this statement was no rejected from levels of gene flow estimated using two indirect methods. Values of $Nem > 1$ indicate that gene flow is sufficient to counteract the effects of genetic drift (Templeton, 2006). Biological traits of pollen and seed dispersal in *M. zapota* tree, high effective sizes of populations and the historical Maya management of the species allow enough exchange of alleles among populations which prevent genetic differentiation among populations.

Allelic richness allows the estimation of expected number of alleles if sampled populations were of equal size. No statistically significant differences were determined among different management units allowing conclude that there are no differences among the number of alleles in sampled populations (**Table 1**). The snails of the representativeness analysis showed that each ancient Maya site had a representative gene pool compared to its complement while home gardens and cultivars populations did not show the same gene pool compared to ancient Maya sites.

Bayesian analysis of genetic structure, following the Evanno's method indicated $K = 2$ as the most probable number of clusters. The assignment of sampled individuals to these clusters defined a main cluster that included 724 out of 782 individuals and a minor cluster that included 58 out of 782 individuals. Despite this pronounced difference in cluster's size, individuals were assigned to each cluster with high probability. Moderate genetic structure between clusters was detected (Table 2). Individuals from all sample sites were assigned to the main cluster.

The fact that all individuals from gardens were assigned to this main cluster shows that such cluster could be more ancient. Individuals from ancient Maya sites were assigned to the minor cluster therefore it could represent modern genetic diversity as consequence of the high mutation rate of microsatellite loci. Individuals from ancient Maya sites populations were assigned in all the clusters when its number reached $K = 4$ while individuals from Tikal defined a new cluster when $K = 5$ (Figure 4).

After these complementary analyses we can conclude that *M. zapota* populations keep high levels of genetic diversity although Ancient Maya left traces in the genetic makeup of these non-native populations with different management histories. In this way these results should be considered for making decisions for conservation and management of this natural resource. Also, this opinion note constitutes an example of the great value that public database of molecular data availability represents for the scientific community either to teach or to investigate.

Nowadays, Population Genetics represents the core of the knowledge for suitable management of natural resources. In this way, make aware of this value to the students is a challenge for the teachers. The main aim is that the students realize the unlimited potential of population genetics as a tool for understand micro-evolutionary processes and the role that this knowledge plays in the development of politics for the management of natural resources. Learning strategy as we described are an excellent choice for a first approach of students to production of scientific knowledge that can be confronted in the scientific field as well as recognize in this knowledge the transferability to the natural resources management.

FIGURE

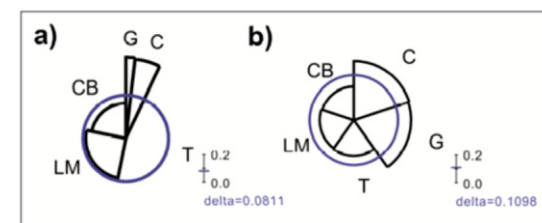
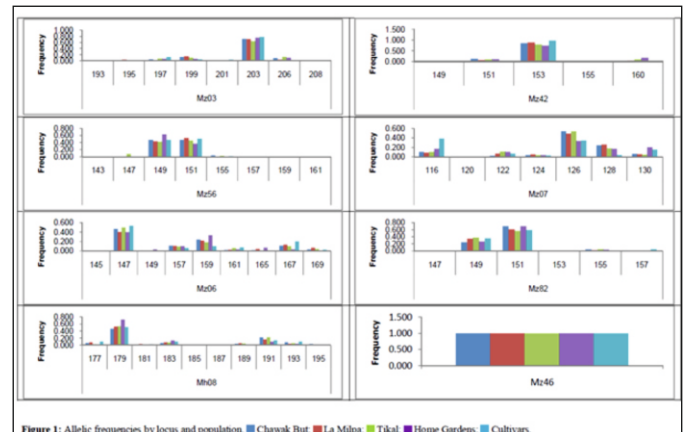


Figure 2: Snail graphic of genetic representativeness analysis, a) proportional to sample sizes and b) corrected for equal sample sizes. Populations: CB: Chawk But, LM: La Milpa, T: Tikal, G: Home Gardens; C: Cultivars.

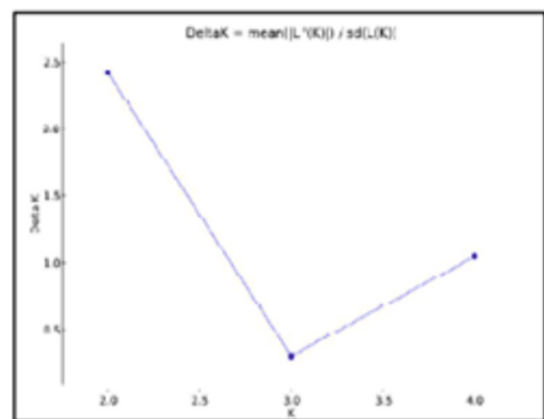


Figure 3: Estimation of K value

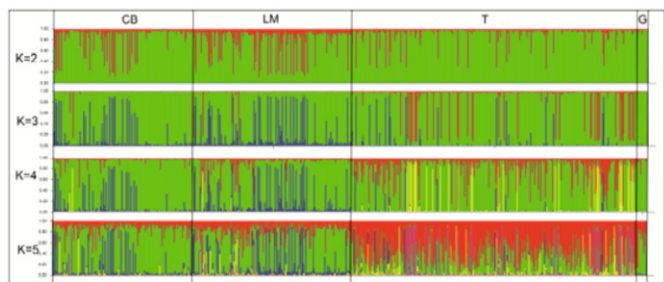


Figure 4: Barplots results of Bayesian analysis K= [1 - 5]. Populations: CB: Chawak But, LM: La Milpa, T: Tikal, G: Home Gardens

TABLE

Table 1: Allelic richness (R) by population and management unit. Statistical significance value (p) of differences among allelic richness of management units

Population	R	Management unit	R	p (95%)
Chawak But (CB)	4.07	Ancient Maya sites	3.35	0.435
La Milpa (LM)	4.17			
Tikal (T)	4.24			
Home Gardens (G)	3.89	Home Gardens	3.14	0.180
Cultivars (C)	3.56	Cultivars	2.95	0.325

Table 2: Proportion of membership of each pre-defined population, expected heterozygosity (He) and fixation index (F_{ST}) for each Bayesian cluster

Pre-define population	Inferred clusters		Number of individuals
	1	2	
CB	0.99	0.01	183
LM	0.98	0.02	209
T	0.90	0.10	375
G	1.00	0.00	15
He	0.50	0.50	-
F_{ST}	0.11	0.14	-

Populations: CB: Chawak But, LM: La Milpa, T: Tikal, G: Home Gardens, C: Cultivars

REFERENCES

- Dewoody J., Trewin H. and Taylor G (2015) Genetic and morphological differentiation in *Populus nigra* L.: isolation by colonization or isolation by adaptation? *Mol. Ecol.* 24, 2641–2655.
- Earl D.A. and vonHoldt B. M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Gen. Res.* 4(2), 359-361.
- Evanno G., Regnaut S. and Goudet J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *MolEcol* 14, 2611–2620.
- Gillet E. (2010) GSED Genetic Structures from Electrophoresis Version 3.0 Data User's Manual. Abt. Forstgenetik und Forstpflanzenzüchtung, Universität Göttingen, Göttingen.
- Goudet J. (1995) FSTAT (vers.2.9.3.2): A computer program to calculate F statistics. *Heredity* 86, 485–486.
- Gregorius, H. R. (1984) Measurement of genetic differentiation in plant populations. In: Gregorius, HR (ed.), *Population Genetics in Forestry*, pp. 276-285. Springer-Verlag, Berlin.
- Gregorius H. R. and Roberds J. H. (1986) Measurement of genetical differentiation among subpopulations. *Theor. Appl. Gen.* 71, 826-834.
- Manel S., Gaggiotti O.E., Waples R.S. (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends Ecol. Evol.* 20(3), 136-142.
- Peakall R. and Smouse P. E. (2012) GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research an update. *Bioinformatics* 28, 2537–2539.
- Pritchard J.K., Stephens M. and Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Raymond M. and Rousset F. (1995) GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Heredity* 86, 248–249.
- Slatkin M. (1980) The distribution of mutant alleles in a subdivided population. *Genetics* 95, 503–23.
- Slatkin M. (1981) Estimating levels of gene flow in natural populations. *Genetics* 99(2), 323-35.
- Slatkin M. (1985) Gene Flow in Natural Populations. *Annu. Rev. Ecoland Syst.* 16, 393-430.
- Templeton A. R. (2006) *Population genetics and microevolutionary theory*. Wiley-Liss Publication, New Jersey.
- Thompson K. M., Culley T. M., Zumberger A. M. and Lentz D.L. (2015) Genetic variation and structure in the neotropical tree, *Manilkara zapota* (L) P. Royen (Sapotaceae) used by the ancient Maya. *Tree Genet Genomes* 11, 40.
- Thompson K. M., Culley T. M., Zumberger A. M., Lentz D. L. (2015b) Data from: Genetic variation and structure in the neotropical tree, *Manilkara zapota* (L) P. Royen (Sapotaceae) used by the ancient Maya. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.244b8>

- Wright S. (1951) The genetical structure of populations. *Ann Eugen.* 15, 323–354.
- Wright S. (1978) *Evolution and the genetics of populations. Vol 4: Variability within and among natural populations.* University of Chicago press, Chicago.