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 strategies 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 essays 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:

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 (delta) (Gregorius and Roberts, 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 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 Fst 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 ($N_{em}$), were made with the following equation: $N_{em} = \frac{1}{2} F_{ST} N_{p}$ (Wright, 1951) where $N_{p}$ 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 (Wright, 1969) shows a linear relationship between $\log_{e}(p(1))$ and $\log_{e}(N_{em})$. Gene flow ($N_{em}$) 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 Maya sites 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 not 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 $N_{em}$ 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 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.
REFERENCES


