*Article*

Estimation of Genetic Diversity in Seven Races of Native Maize from the Highlands of Mexico

Mario Rocandio-Rodríguez [1] , **Amalio Santacruz-Varela** [2,*] , **Leobigildo Córdova-Téllez** [2] , **Higinio López-Sánchez** [3] , **Aurelio Hernández-Bautista** [2] , **Fernando Castillo-González** [2] , **Ricardo Lobato-Ortiz** [2] , **J. Jesús García-Zavala** [2] and **Pedro Antonio López** [3]

[1]   Instituto de Ecología Aplicada, Universidad Autónoma de Tamaulipas, División del Golfo 356, Col. Libertad, Victoria 87019, Mexico; mrocandio@docentes.uat.edu.mx

[2]   Colegio de Postgraduados Campus Montecillo, km. 36.5. Carr. Mexico-Texcoco, Texcoco 56230, Mexico; lcordova@colpos.mx (L.C.-T.); aureliohb@hotmail.com (A.H.-B.); fcastill@colpos.mx (F.C.-G.); rlobato@colpos.mx (R.L.-O.); zavala@colpos.mx (J.J.G.-Z.)

[3]   Colegio de Postgraduados, Campus Puebla, km. 125.5. Carr. Fed. Mexico-Puebla, Santiago Momoxpan, Cholula, Puebla 72760, Mexico; higiniols@colpos.mx (H.L.-S.); palopez@colpos.mx (P.A.L.)

*   Correspondence: asvarela@colpos.mx

check for updates

Abstract: Characterizing the genetic diversity of maize (*Zea mays* L.) populations by their morphological and molecular attributes makes it possible to place populations into specific groups; thus, facilitating the design of procedures for their optimum and sustainable use. In this study, data from two lines of evidence were analyzed simultaneously to robustly classify maize populations and to determine their genetic relationships. Seven maize races of the central high plateau of Mexico were characterized using a combined analysis of 13 morphological traits and 31 microsatellite loci. The germplasm assessed included samples of 119 accessions held at Mexican germplasm banks. Cluster and principal component analyses were performed. Also, genetic and geographic relationships among the accessions were determined. Principal component analysis separated the different accessions into well-defined groups using first three principal components. The accessions of Arrocillo Amarillo and Elotes Cónicos races did not exhibit a grouping pattern, indicating greater genetic complexity. Better grounded grouping and phylogenetic relationships were obtained when traits of both lines of evidence were used simultaneously.

Keywords: combined analysis; evolution; genetic resources; germplasm; *Zea mays*

1. Introduction

The evolutionary history of maize (*Zea mays* L.), as well as its diversity, has been of interest for numerous studies [1], which have focused mainly on morphological variation [2,3]. In other classic studies, Cervantes, Goodman, Casas and Rawlings [4] used genetic effects and the interaction genotype × environment, while Sánchez, Goodman and Rawlings [5] involved the effects of the interaction genotype × environment along with stability parameters, which enabled to classify individuals more precisely into discrete units [6], although these parameters may not necessarily provide reliable information on possible phylogenetic relationships [7]. More recently, molecular tools, such as simple sequence repeats of DNA (SSRs) have been applied in studies of maize landraces [8–12], and the data and findings have added substantially to understanding maize evolution and diversification.

Some classification work on Mexican maize landraces has been based on phenetic analysis, using morphological traits [12,13], in combination, in some instances, with isozyme allelic frequencies [14,15], and combinations of morphological traits, isozymes and microsatellites have

also been used [16]. Morphological and agronomic characterization has been very useful for the study and assessment of genetic resources; however, morphological and agronomic traits are attributes that may or may not be inherited and may be controlled by one or multiple genes [17]. They may also have limitations that restrict recovering genetic information, such as pleiotropic effects, ignorance of the genetic base, type of inheritance, or high susceptibility to environmental influence [18]. For these reasons, in recent years and due to progress in molecular biology, the use of molecular markers such as microsatellites has complemented the information from morphological traits, making it possible to estimate neutral genetic diversity. Although assessment of molecular data is more complex than that of morphological traits, there is the advantage that environmental influence is negligible and comparisons can be made between individuals of the same or different species, kinship and phylogenetic relations can be established, and processes of migration and genetic drift can be revealed. Moreover, a large amount of genetic information on populations can be obtained [19].

The convenience of using sets of data separately and combined has been discussed. Avise [20] argue that separate analyses may result in conflicts in the resolution of the trees, or classifications, given that the traits may have different origins; that is, they may be affected to a greater or lesser degree by natural or artificial selection, or they may have a different evolutionary rate, among other reasons. Furthermore, [21,22] put forth the possibility of obtaining better grounded classification when traits of different nature are combined, although it is expected that some errors generated by a certain type of traits or marker can be minimized using other markers According to [23–27], one of the most accepted approaches for analyzing this information is that of simultaneous analysis, or total evidence, since it gives more precision to the classifications and maximizes descriptive efficiency and explanatory power of all the information.

In recent years, the genetic diversity of maize was analyzed by Rocandio-Rodríguez, Santacruz-Varela, Córdova-Téllez, López-Sánchez, Castillo-González, Lobato-Ortiz, García-Zavala and Ortega-Paczka [14], who assessed phenotypically 119 accessions of maize at the Mexican highlands. In other study Rocandio-Rodríguez, Santacruz-Varela, Córdova-Téllez, López-Sánchez, Castillo-González, Lobato-Ortiz, and García-Zavala [11], also studied the genetic structure and molecular genetic diversity within and among these same accessions along with the race Balsas of teosinte (*Zea mays* ssp. *parviglumis*). Based on the above, to understand maize diversity and obtain better grounded phylogenetic relations of the accessions within discrete units (races), this study aimed to classify accessions of the seven maize races most cultivated at the Mexican central high plateau, and examine similarity relations among populations and groups of populations according to the combined analysis of morphological traits and microsatellite allele frequencies.

2. Materials and Methods

2.1. Biological Material

One hundred and seven accessions, representative of seven maize races of the central highlands of Mexico, were characterized: 10 Arrocillo Amarillo, 11 Cacachuacintle, 22 Chalqueño, 23 Cónico, 14 Elotes Cónicos, 8 Palomero Toluqueño, 19 Purépecha, along with two commercial varieties used as reference. The seed used was acquired from the germplasm banks of the International Maize and Wheat Improvement Center (CIMMYT), and Mexican institutions as Chapingo Autonomous University, National Institute of Research on Forestry, Agriculture and Livestock, and Colegio de Postgraduados. The accessions were identified visually by expert classifiers as those that most closely exhibited the ear and plant characteristics of each race.

2.2. Field Assessments

The field experiments were set up during the 2010 Spring-Summer growing cycle in three environments of the high Mexican plateau: Ciudad Serdán (18°50′ latitude N, 97°28′ longitude W, altitude of 2570 masl) and San Mateo Capultitlán (19°08′ latitude N; 98°28′ longitude W, altitude

2260 masl), state of Puebla and Montecillo, State of Mexico (19°27′ latitude N, 98°54′ longitude W, altitude 2250 masl). Experimental design was complete randomized blocks with two replications in each site. The experimental unit or plot consisted of two rows 5-m long and 0.8-m wide in which 44 seeds were equidistantly planted.

2.3. Morphological Traits Measured

In each experiment phenotypic traits were recorded. We randomly chose five competitive plants per plot in order to collect data for phenologic, vegetative, tassel, ear and kernel traits. Also, some ratios among those traits were calculated (Table 1).

Table 1. Phenotypic traits assessed.

Traits
Phenological traits: days to male (MF) and female (FF) flowering; anthesis-silking interval (ASI).
Vegetative traits: number of tillers per plant (SP), plant height (PH), ear height (EH), length (LL) and width (LW) of the leaf at the primary ear in cm, total number of leaves (NL) and number of leaves above the ear (NLE).
Tassel traits: peduncle length (PLS), length of the branched part of the tassel (LBT), length of the central spike (LCS), total tassel length (TL) in cm, number of primary branches (PB).
Ear traits: length (EL) and ear diameter (ED); peduncle length (PLE); cob diameter (CD), in cm, number of kernel rows (KR), ear grain percentage (EGP = (grain weight/total weight of maize ear) × 100).
Kernel traits: length (KL), width (KW) and thickness (KT) in mm in 10 kernels of each of the five ears of the sampled plants, weight of 100 kernels (W100K) in g, volume of 100 kernels (V100K) in mL.
Ratios: EH/PH, ED/EL, LBT/TL, KW/KL, KT/KL and W100K/V100K.

2.4. Assessment of Microsatellite Polymorphism

The 107 accessions were analyzed using 31 microsatellite loci distributed on the ten maize chromosomes; there is ample information on these loci published in the Maize Genetics and Genomics Database (MaizeGDB), available on line at http://www.maizegdb.org/ssr.php# (Table 2). '

For the molecular analysis, genomic DNA was extracted from 100 mg of mesocotyl, coleoptile, and young leaf tissue individually from 25 plants per accession with a commercial DNA extraction kit (ChargeSwitch® gDNA Plant, Invitrogen, Thermo Fisher Scientific, Carlsbad, USA), using a DNA extraction and purification robot (King Fisher Flex™, ThermoScientific, Waltham, MA, USA). Amplification was performed individually by PCR in volumes of 25 µL that contained 10 mM nucleotides, 25 mM MgCl$_2$, 5× buffer, 100 ng template DNA, 1 unit DNA *Taq* polymerase, and 4 pmol of each primer. The protocol for PCR amplification consisted of initial denaturation for 4 min at 95 °C, followed by 25 cycles of 1 min at 95 °C, 2 min at 55 °C, 2 min at 72 °C, and one final extension of 60 min at 72 °C. PCR products were separated by capillary electrophoresis in a DNA sequencer (Genetic Analyzer ABI 3130™, Applied Biosystems, Foster City, CA, USA) and detection was based on the presence of the fluorescent labels 6-FAM, ROX or HEX at the 5′ extreme of the forward primers using LIZ-500 as the size standard internal marker. Data files were generated with the allelic profile of the markers for each of the populations with the software GeneMapper® V. 4.0 [28], and allele variability with POPGENE 1.31 [29] software.

2.5. Statistical Analysis

Using the averages of the morphological data from the three sites, a combined analysis of variance was performed with SAS V.9.0. [30]. The linear model used was: $Y_{ijk} = \mu + \alpha_i + \gamma_j + \delta_{ij} + B(\gamma)_{k(j)} + \varepsilon_{ijk}$, where Y_{ijk} is the observation of the *i*th accession in the *j*th environment of the *k*th block, μ is the general mean, α_i is the effect of the *i*th accession, γ_j is the effect of the *j*th environment, δ_{ij} is the interaction effect of the *i*th accession with the *j*th environment, $B(\gamma)_{k(j)}$ is the effect of the *k*th block nested into the *j*th environment, and ε_{ijk} is the random error associated with the experimental unit [31].

Table 2. Microsatellite loci and primers used for amplification of repeated simple sequences in maize landrace populations of the highlands of Mexico.

Locus	N. Bin.	Forward Primer//Reverse Primer
phi127	2.07	ROX-ATATGCATTGCCTGGAACTGGAAGGA//AATTCAAACACGCCTCCCGAGTGT
phi051	7.06	6-FAM-GCGAAAGCGAACGACAACAATCTT//ACATCGTCAGATTATATTGCAGACCA
phi115	8.03	HEX-GCTCCGTGTTTCGCCTGAA//ACCATCACCTGAATCCATCACA
phi015	8.08	HEX-GCAACGTACCGTACCTTTCCGA//ACGCTGCATTCAATTACCGGGAAG
phi033	9.02	6-FAM-ATCGAAATGCAGGCGATGGTTCTC//ATCGAGATGTTCTACGCCCTGAAGT
phi053	3.05	ROX-CTGCCTCTCAGATTCAGAGATTGAC//AACCCAACGTACTCCGGCAG
phi072	4.01	6-FAM-GTGCATGATTAATTTCTCCAGCCTT//GACAGCGCGCAAATGGATTGAACT
phi093	4.08	ROX-GTGCGTCAGCTTCATCGCCTACAAG//CCATGCATGCTTGCAACAATGGATACA
phi024	5.01	HEX-CTCCGCTTCCACTGTTCCA//TGTCCGCTGCTTCTACCCA
phi085	5.06	6-FAM-AGCAGAACGGCAAGGGCTACT//TTTGGCACACCACGACGA
phi034	7.02	HEX-TAGCGACAGGATGGCCTCTTCT//GGGGAGCACGCCTTCGTTCT
phi121	8.04	6-FAM-AGGAAAATGGAGCCGGTGAACCA//TTGGTCTGGACCAAGCACATACAC
phi056	1.01	ROX-ACTTGCTTGCCTGCCGTTAC//CGCACACCACTTCCCAGAA
phi064	1.11	HEX-CGAATTGAAATAGCTGCGAGAACCT//ACAATGAACGGTGGTTATCAACACGC
phi050	10.03	ROX-AACATGCCAGACACATACGGACAG//ATGGCTCTAGCGAAGCGTAGAG
phi96100	2.01	6-FAM-AGGAGGACCCCAACTCCTG//TTGCACGAGCCATCGTAT
phi101249	Unknown	6-FAM-TTCCTCCTCCACTGCCTC//AAGAACAGCGAAGCAGAGAAGG
phi109188	5.03	HEX-AAGCTCAGAAGCCGGAGC//GGTCATCAAGCTCTCTGATCG
phi029	3.04	ROX-TCTTTCTTCCTCCACAAGCAGCGAA//TTTCCAGTTGCCACCGACGAAGAACTT
phi073	3.05	HEX-GTGCGAGAGGCTTGACCAA//AAGGGTTGAGGGCGAGGAA
phi96342	10.02	6-FAM-GTAATCCCACGTCCTATCAGCC//TCCAACTTGAACGAACTCCTC
phi109275	1.03	6-FAM-CGGTTCATGCTAGCTCTGC//GTTGTGGCTGTGGTGGTG
phi427913	1.01	ROX-CAAAAGCTAGTCGGGGTCA//ATTGTTCGATGACACACTACGC
phi265454	1.11	6-FAM-CAAGCACCTCAACCTCTTCG//TCCACGCTGCTCACCTTC
phi402893	2.XX	HEX-GCCAAGCTCAGGGTCAAG//CACGAGCGTTATTCGCTGT
phi346482	1.XX	HEX-GCATCACACTTCACACAACAA//GTGGAATAGGAGGCGAGAGAGG
phi308090	4.04–4.05	6-FAM-CAGTCTGCCACGAAGCAA//CTGTCGGTTTCGGTCTTCTT
phi330507	5.02–5.06	ROX-GTAAAGTACGATGCGCCTCCC//CGGGGTAGAGGAGAGTTGTG
phi213398	4.01–4.04	6-FAM-GTGACCTAAACTTGGCAGACCC//CAAGAGGTACCTGCATGGC
phi339017	1.03	HEX-ACTGCTGTTGGGGTAGGG//GCAGCTTGAGCAGGAAGC
phi159819	6.00–6.08	6-FAM-GATGGGCCCTAGACCAGCTT//GCCTCTCCCATCTCTCGGT

Morphological traits were then selected using two statistical methods for additional analyses. The first method was based on repeatability coefficient (r), as suggested by Sánchez, Goodman and Rawlings [5], and the second one was based on the structure of the correlation matrix of independent variables distributed on Gabriel graphics [32]. The repeatability coefficient is a criterion proposed by Goodman and Paterniani [6], which is based on discriminating morphological variables based on the stability of the characters across the environments. When ≥1, the variable is considered stable and appropriate for classification.

Allele frequencies were selected based on the level of significance between populations for each of the alleles ($p \leq 0.05$) Also selected were those alleles that had a frequency above 2%. This was done to avoid problems of distancing between accessions and of interpretation when involving low frequency or single alleles in the grouping analyses.

With the morphological traits and allele frequencies selected from the 31 microsatellite loci detected in each of the individuals, a data matrix was constructed with 107 populations, using a total of 224 variables (13 morphological traits and frequency of 211 alleles previously selected). The variables were standardized by subtracting the mean and dividing by the standard deviation. The dataset was used to calculate Gower distances, which are recommended when variables of different nature are used [33] among populations; this was done with SAS V.9.0. [30] software. A phylogenetic tree was then constructed with the Neighbor-Joining method [34] in NTSYSpc V.2.2 software [35], and the population Sina-2 of the Chapalote race, considered one of the ancient maize landraces [2] was used as an outgroup.

Allele frequencies and morphological information were used to construct a correlation matrix among the traits. Each morphological trait and each allele were considered as an independent variable. The principal component analysis was based on the correlation matrix and processed with the statistical software SAS V.9.0. [30].

3. Results

3.1. Analysis of Variance

Analysis of variance evidenced highly significant differences among populations for all the measured characteristics (Table 3). For population-environment interaction, it was observed that traits as plant height, peduncle length of the tassel, length of branched part of tassel, ear length, kernel width, ear grain percentage, weight of 100 kernels, volume of 100 kernels, kernel thickness/kernel length, and weight of 100 kernels/volume of 100 kernels had non-significant interaction. Finally, estimators of variance components produced values of repeatability r ≥ 1 for 19 out of the 32 traits (Table 4).

Table 3. Mean squares of the combined analysis of variance across localities.

Trait	Populations	Localities	Pob × Loc	Error	Mean	Standard Deviation
Male flowering	932.5 **	26,311.7 **	39.4 **	13.5	100.64	15.79
Female flowering	813.2 **	34,724.2 **	33.0 **	15.7	93.46	15.71
Anthesis-silking interval	33.5 **	659.8 **	9.965 **	5.5	7.18	3.65
Shoots per plant	0.1 **	15.0 **	0.1 **	0.1	0.21	0.32
Plant height	10,455.7 **	421,681.3 **	330.9 ns	288.2	250.23	56.16
Ear height	5441.2 **	326,631.3 **	254.4 **	206.4	119.72	44.63
Ear height/plant height	0.0 **	1.2 **	0.0 **	0.0	0.46	0.09
Number of leaves	21.5 **	25.4 ns	1.2 **	0.6	12.06	2.08
Number of leaves above ear	2.1 **	10.6 **	0.1 **	0.1	4.81	0.69
Leaf length	393.7 **	57,622.6 **	57.5 **	29.2	75.08	16.07
Leaf width	7.7 **	955.6 **	1.7 *	1.4	8.81	2.28
Total length	155.3 **	210.7 ns	30.0 **	25.6	63.27	6.85
Peduncle length of the tassel	30.6 **	248.5 ns	12.9 ns	11.1	25.21	3.97
Length of central spike	64.6 **	265.0 **	10.6 **	7.7	31.58	4.34
Length of branched part of tassel	23.6 **	1190.5 **	2.4 ns	2.2	6.65	3.02
Length of branched part of tassel/total length	0.0 **	0.3 **	0.0 **	0.0	0.10	0.05
Number of primary branches	38.2 **	1180.2 **	6.2 **	4.8	8.17	3.75
Ear length	34.8 **	423.7 **	2.9 ns	2.7	14.54	3.05
Number of kernel rows	25.0 **	186.7 **	1.3 **	1.	14.48	2.37
Peduncle length of the ear	26.1 **	1549.9 **	8.5**	6.	12.01	3.83
Ear diameter	0.9 **	34.6 **	0.1 **	0.1	4.11	0.56
Ear diameter/ear length	0.0 **	0.0 ns	0.0 **	0.0	0.29	0.05
Cob diameter	0.6 **	12.3 **	0.1 **	0.1	1.95	0.45
Kernel length	783.1 **	14,837.6 **	97.8 **	64.1	12.67	1.54
Kernel width	826.7 **	1500.4 **	20.9 ns	19.0	7.73	1.25
Kernel thickness	156.9 **	138.3**	9.7 **	7.3	4.16	0.58
Ear grain percentage	56.6 **	223.1 **	14.4 ns	14.7	87.30	4.70
Weight of 100 kernels	248.1 **	1114.0 **	22.2 ns	18.8	27.79	7.81
Volume of 100 kernels	811.2 **	2990.6 **	50.1 ns	47.1	42.86	13.52
Kernel width/kernel length	0.1 **	0.1 **	0.0 **	0.0	0.62	0.11
Kernel thickness/kernel length	0.0 **	0.1 **	0.0 ns	0.0	0.33	0.07
Weight of 100 kernels/volume of 100 kernels	0.0 **	0.0 ns	0.0 ns	0.0	0.66	0.06

* Significance at 5%; ** significance at 1%; ns: non significance.

Table 4. Variance components for populations (σ^2_g), localities (σ^2_l), interaction populations × localities ($\sigma^2_{g \times l}$), error (σ^2_e), broad-sense heritability (H^2) and repeatability (r) of the traits studied.

Trait	σ^2_g	σ^2_l	$\sigma^2_{g \times l}$	σ^2_e	H^2	r
Female flowering †	148.84	108.4238	12.9696	13.488	84.91	1.2261
Male flowering	130.024	143.2363	8.6491	15.734	84.21	0.8561
Anthesis-silking interval	3.935	2.7019	2.1987	5.467	33.92	0.8030
Shoots per plant	0.001	0.0618	0.0098	0.052	1.59	0.0143
Plant height	1687.46	1734.8050	21.3440	288.216	84.50	0.9609
Ear height	864.46	1346.0031	24.0306	206.367	78.96	0.6310
Ear height/plant height	0.002	0.0050	0.0005	0.002	44.44	0.3706
Number of leaves †	3.388	0.0830	0.2757	0.6	79.46	9.4447
Number of leaves above ear †	0.325	0.0421	0.0193	0.108	71.85	5.3000
Leaf length	56.02	237.4093	14.1348	29.239	56.36	0.2227
Leaf width	0.985	3.9453	0.1478	1.439	38.30	0.2408
Total length †	20.883	0.2329	2.2085	25.596	42.89	8.5538
Peduncle length of the tassel †	2.949	0.8468	0.8712	11.149	19.70	1.7169
Length of central spike †	9.012	0.9987	1.4356	7.69	49.69	3.7021
Length of branched part of tassel	3.538	4.9074	0.1090	2.201	60.50	0.7055
Length of branched part of tassel/total length	0.0008	0.0011	0.0001	0.001	42.11	0.6358
Number of primary branches	5.335	4.8565	0.7163	4.79	49.21	0.9574
Ear length †	5.316	1.7033	0.1233	2.696	65.34	2.9107
Number of kernel rows †	3.932	0.7603	0.1303	1.044	77.00	4.4151
Peduncle length of the ear	2.923	6.3434	1.0325	6.46	28.06	0.3963

Table 4. *Cont.*

Trait	σ^2_g	σ^2_l	$\sigma^2_{g\times l}$	σ^2_e	H^2	r
Ear diameter	0.149	0.1415	0.0111	0.054	69.59	0.9795
Ear diameter/Ear length †	0.001	0.0001	0.0001	0.001	47.62	8.0939
Cob diameter †	0.085	0.0506	0.0093	0.074	50.51	1.4316
Kernel length †	114.213	60.0425	16.8775	64.056	58.53	1.4848
Kernel width †	134.291	6.1751	0.9700	18.99	87.06	18.7950
Kernel thickness †	24.545	0.5327	1.1700	7.33	74.28	14.4151
Ear grain percentage †	7.025	0.9063	−0.1294	14.699	32.53	9.0428
Weight of 100 kernels †	37.655	4.4013	1.733	18.755	64.76	6.1385
Volume of 100 kernels †	126.85	12.1132	1.5034	47.084	72.30	9.3161
Kernel width/kernel length †	0.008	0.0005	0.0005	0.003	69.57	9.2262
Kernel thickness/kernel length †	0.002	0.0005	0.0002	0.002	47.62	3.1202
Weight of 100 kernel/volume of 100 kernels †	0.001	0.0001	0.0001	0.002	32.26	10.2235

† Selected traits.

3.2. Cluster and Principal Component Analyses

To define the final set of traits, the correlation matrix was generated, and one trait of those pairs with a correlation coefficient above 0.7 was eliminated. In this way, 13 final traits (FF, LCS, NLE, KR, EL, ED/EL, KW, W100K, KW/KL, KT/KL, KT, KL and W100K/V100K) were chosen for the assessment of racial diversity and classification.

According to the cluster analysis, a total of seven groups were observed (Figure 1). Group I was integrated by accessions of the Purépecha, Palomero Toluqueño and Elotes Cónicos races. Group II, was placed in the upper part of the phylogram, it was represented by accessions Hgo-116, Mex-6, Pueb-618, Ver-537 and Tlax-255, belonging to the Elotes Cónicos race. Group III included nine populations of the Cónico race, originating from the states of Hidalgo and Tlaxcala, and three Palomero Toluqueño accessions from the states of Mexico and Tlaxcala. Group IV was formed by a mixed set of different races (Cacahuacintle, Elotes Cónicos, Arrocillo Amarillo, Chalqueño, Purépecha). Group V comprised eight Chalqueño populations, predominantly from the states of Durango, Guanajuato, Jalisco, Zacatecas and Querétaro, and two Cacahuacintle populations. Group VI was formed by Cónico accessions from the states of Mexico, Hidalgo and Tlaxcala as well as Arrocillo Amarillo accessions from states of Veracruz and Mexico. Group VII integrated populations of the races Cónico and Chalqueño, predominantly from the states of Mexico, Hidalgo Puebla and Tlaxcala, and one population (Pue-512) of the Elotes Cónicos race, distinguished for its recent formation.

Dispersion of the 107 maize accessions of the highlands of Mexico was represented in a three-dimensional space determined by the first three principal components (Figure 2), which exhibited broad variation within each racial group and groups defined by race. The first principal component explained 8.7% of the total variation, and the alleles that most contributed were *phi072-C*, *phi121-B*, *phi121-H*, *phi346482-J*, *phi093-E*, *phi015-K*, *phi050-H*, *phi024-Q*, *phi427913-F*, *phi346482-I*, *phi402893-G* and *phi127-B* toward the positive end, while the morphological traits that had the largest influence were ED/EL, EL and KT/KL. The second principal component explained 4.9% of the total variation and was influenced mostly by the high frequency of the alleles *phi033-L*, *phi115-C*, *phi265454-M*, *phi402893-B*, *phi115-A*, *phi051-D*, *phi053-G*, *phi96100-C*, *phi101249-B*, *phi96100-N* and *phi265454-P* at the positive end of the component. The third PC explained 4.6% of the total variation and was more closely associated with morphological traits (LCS, EL, KW, W100G, W100K/V100K, KT and FF) and with some allele frequencies (*phi050-G* and *phi427913-I*).
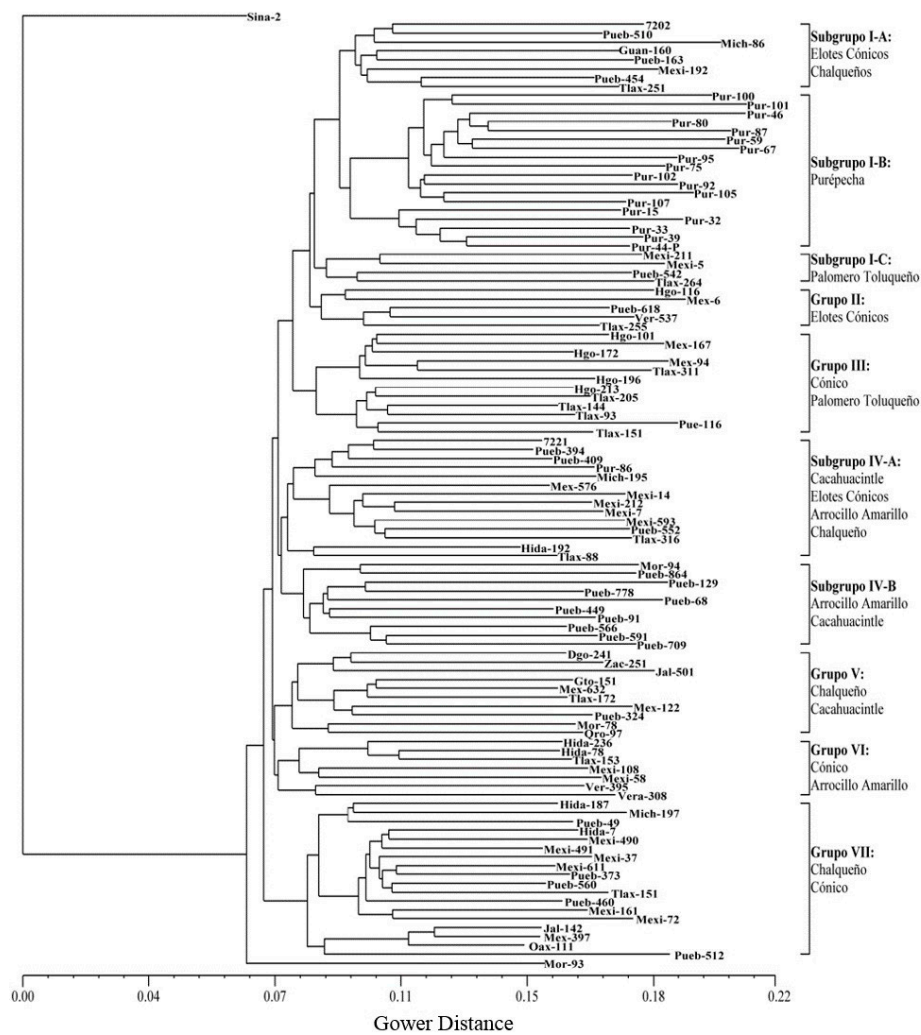
Figure 1. Dendrogram showing the hierarchical clustering produced by the Neighbor-Joining method from 107 maize accessions using Gower distances derived from 13 morphological variables and 211 SSR alleles.
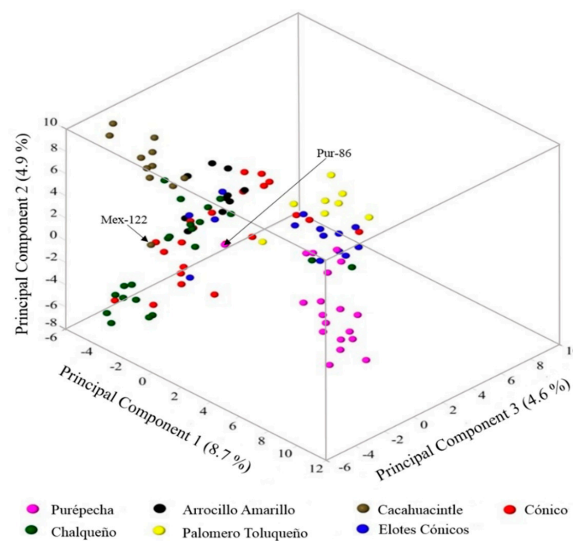

Figure 2. Dispersion of 107 accessions of Mexican maize based on the first three principal components derived from 13 morphological variables and 211 alleles of 31 SSR loci.

4. Discussion

The high degree of variation found suggests the existence of broad genetic diversity among genotypes. Based on the value of repeatability, a total of 32 traits were selected at the first stage. With the exception of PH and LBT, the selected traits in this stage were not statistically significant for the interaction genotypes × localities. This gives them properties that are desirable for classification. According to Sánchez, Godman and Rawlings [5], the traits that are less affected by environment are more useful for characterizing populations. Some of the traits detected here as appropriate for classification were also selected by [5], who indicated that ear, and in general reproductive traits are the most suitable for characterization of maize races. Among these NL, LCS, KW, ED/EL and KW/KL are included.

Thirteen (FF, LCS, NLE, KR, EL, ED/EL, KW, W100K, KW/KL, KT/KL, KT, KL and W100K/V100K) were selected for performing the cluster and PCA procedures, based on correlation analysis. Most of these traits had also been selected in other studies of maize genetic diversity [5,15,16,36,37].

The dendrogram differentiated seven groups, with subgroups within two of them. Group I had three subgroups. Subgroup I-A included six accessions of the Elotes Cónicos race and two of the Chalqueño race; Subgroup I-B comprised most of the populations of the Purépecha race, which formed a well-defined cluster, with poorly differentiated native populations within the Sierra Purépecha region (Salvador Escalante, Charapan, Tingambato, Nahuatzen and Paracho, Michoacán), which is isolated by several geographic barriers created by the mountainous orography; cultural identity is strong and there are semi-collective forms of land use [38]. The results suggest that these populations constituted a genetic group different from the Chalqueño group, and strongly justify their integration into the newly described race Purépecha, as it was proposed by Romero, Castillo and Ortega [39] and later confirmed by Subgroup I-C comprised four Palomero Toluqueño accessions.

The nine populations (Group III) of the Cónico race were grouped intermediately between Palomero Toluqueño populations (Subgroup I-C) and those of Cacahuacintle (Subgroup IV-A). This is in agreement with [2], who proposed that the Cónico race could be a product of hybridization between Palomero Toluqueño and Cacahuacintle races. In this group, the Cónico accession Pue-116 is outstanding as it appeared as a recently formed population.

Group IV is divided into two subgroups. Subgroup IV-A integrated six accessions of Cacahuacintle, three Elotes Cónicos, two Arrocillo Amarillo, two Chalqueño and one Purépecha population (Pur-86). Subgroup IV-B was defined by geographic origin, exhibiting populations from the state of Puebla (five Arrocillo Amarillo, three Cacahuacintle and one Cónico), as well as one population of the Chalqueño race from Totolapan, Morelos. The Cacahuacintle and Arrocillo Amarillo races are associated in both Subgroups (IV-A and IV-B), indicating a higher degree of diversity [2,13,40], since both are closely related and share similar traits.

The interspersing of the Cónico and Chalqueño races in Groups V, VI and VII indicates that the two races are closely related due to the constant gene flow among their populations. This did not allow placing the populations in well-defined groups because their geographic distribution is almost identical. Results reveal that the principal maize races of the Mexican highlands share a common origin and have a diffuse genetic background extending throughout the populations; this is also indicated by the presence of common alleles of the molecular markers and similar values of the morphological traits; therefore, it is inferred that differentiation of these races has been influenced mainly by natural and artificial selection pressure in relatively recent times.

The Cónico population Mor-93 was placed alone in the far inner part of the phylogram. It is likely that this population has not been in contact with those of other regions, or it has had influence from materials not included in this study.

Regarding the principal component analysis of the combined dataset (frequencies of 211 SSR alleles and 13 morphological traits), the first 20 principal components explained 55.1% of the variance. When allele frequencies are involved in the analysis, it is common that the percentage of explained variance decreases notably; in contrast, when only morphological characterization is used, a higher

percentage is reached with a smaller number of principal components [37,41]. Such reduction could be due to the rare alleles present in the population, effect of molecular markers in the genetic architecture of the traits, and the linkage between the molecular markers and major or minor genes explicating the phenotypic diversity present in the population [42].

The principal component analysis revealed grouping patterns very similar to those observed in the phylogenetic analysis, indicating that the definition of the associations is highly consistent. Races Cacahuacintle, Purépecha, Elotes Cónicos, Palomero Toluqueño constitute well-defined groups, while the Chalqueño race exhibis two groups, in which it was possible to distinguish accessions from the states of Durango, Zacatecas, Jalisco and Querétaro from those of the traditional core of this race located in the highlands of central Mexico. The Cónico accessions overlapped with the Chalqueño race, indicating genetic complexity influenced by neutral selective markers. Spatial distribution of populations observed on the principal component analysis was similar to that reported by Wellhausen, Roberts, Hernández and Mangelsdorf [2], who found that the race Cónico was localized between Palomero Toluqueño and Cacahuacintle. Such result suggested that the race Cónico was derived from the cross between both races, or that a close relationship does exist between the three races. In addition, a strong grouping was observed between the races Cónico and Elotes Cónicos, and some authors as Wellhausen, Roberts, Hernández and Mangelsdorf [2] even considered that Elotes Cónicos is only a pigmented sub-race of Cónico.

Spatial distribution of the Purépecha race, according to the three first principal components, shows that the populations are very similar, observed in the compact group toward the positive end of the first principal component. It can also be observed that populations of the Cónico race are found far apart, indicating that the variation into this race is very broad, or rather that the populations are not appropriately classified in the Mexican germplasm banks. Most of the maize accessions from Mexican germplasm banks were classified according to morphological variables. However, the morphological variables are often susceptible to phenotypic plasticity [43], which caused a bias in the classification of some populations of the Mexican germplasm banks. Our study demonstrated that this disadvantage can be solved with the combined use of molecular markers and morphological variables.

In the same context, it can be observed that accessions Mex-122 of the Cacahuacintle race and Pur-86 of the Purépecha race do not coincide with their respective groups. This suggests that they are perhaps not correctly classified or that they have had genetic influence from other races.

5. Conclusions

The molecular and morphological information analyzed together for racial classification provided strong elements for understanding diversification and evolutionary relationships that exist among the maize landraces of the highlands of Mexico. Results obtained in this study, in general, agree with the description in that the Pyramidal complex (races Cónico, Chalqueño, Palomero Toluqueño, Arrocillo Amarillo and Elotes Cónicos) are more closely related genetically, while the Cacahuacintle and Purépecha races are not closely associated with this complex.

References

1. Kato, Y.T.A.; Mapes, C.S.; Mera, L.M.O.; Serratos, J.A.H.; Bye, R.A.B. *Origen y Diversificación del Maíz: Una Revisión Analítica*; Universidad Nacional Autónoma de México, Comisión Nacional para el Conocimiento y Uso de la Biodiversidad: Mexico City, Mexico, 2009; p. 116.

2. Wellhausen, E.J.; Roberts, L.M.; Hernández, X.E.; Mangelsdorf, P.C. *Races of Maize in México, their Origin, Characteristics and Distribution*; The Bussey Institution of Harvard University: Cambridge, UK, 1952; p. 223.

3. Hernández, X.E.; Alanís, F.G. Morphological study of five new maize races of the Sierra Madre Occidental of Mexico. Phylogenetic and phytogeographic implications. *Agrociencia* **1970**, *5*, 3–30.

4. Cervantes, S.T.; Goodman, M.M.; Casas, D.E.; Rawlings, J.O. Use of genetic effects and genotype by environmental interactions for the classification of Mexican races of maize. *Genetics* **1978**, *90*, 339–348. [PubMed]

5. Sánchez, G.J.J.; Goodman, M.M.; Rawlings, J.O. Appropriate characters for racial classification in maize. *Econ. Bot.* **1993**, *47*, 44–59. [CrossRef]

6. Goodman, M.M.; Paterniani, E. The races of maize: III. Choices of appropriate characters for racial classification. *Econ. Bot.* **1969**, *23*, 265–273. [CrossRef]

7. Doebley, J.F.; Iltis, H.H. Taxonomy of *Zea* (Gramineae). I. A subgeneric classification with key to taxa. *Am. J. Bot.* **1980**, *67*, 982–993. [CrossRef]

8. Matsuoka, Y.; Vigoroux, Y.; Goodman, M.M.; Sánchez, G.J.; Buckler, E.; Doebley, J. A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6080–6084. [CrossRef] [PubMed]

9. Reif, J.C.; Waburton, M.L.; Xia, X.C.; Hoisington, D.A.; Crossa, J.; Taba, S.; Muminović, J.; Bohn, M.; Frisch, M.; Melchinger, A.E. Grouping of accessions of Mexican races of maize revisited with SSR markers. *Theor. Appl. Genet.* **2006**, *113*, 177–185. [CrossRef]

10. Vigouroux, Y.; Glaubitz, J.C.; Matsuoka, Y.; Goodman, M.M.; Sánchez, G.J.; Doebley, J. Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *Am. J. Bot.* **2008**, *95*, 1240–1253. [CrossRef]

11. Rocandio-Rodríguez, M.; Santacruz-Varela, A.; Córdova-Téllez, L.; López-Sánchez, H.; Castillo-González, F.; Lobato-Ortiz, R.; García-Zavala, J.J. Detection of genetic diversity of seven maize races from the high central valleys of Mexico using microsatellites. *Maydica* **2014**, *59*, 144–151.

12. Vega-Alvarez, I.; Santacruz-Varela, A.; Rocandio-Rodríguez, M.; Córdova-Téllez, L.; López-Sánchez, H.; Muñoz-Orozco, A.; Hernández-Bautista, A. Genetic diversity and structure of native maize races from Northwestern Mexico. *Pesq. Agropec. Bras.* **2017**, *52*, 1023–1032. [CrossRef]

13. Sánchez, G.J.J.; Goodman, M.M. Relationships among the Mexican races of maize. *Econ. Bot.* **1992**, *46*, 72–85. [CrossRef]

14. Rocandio-Rodríguez, M.; Santacruz-Varela, A.; Córdova-Téllez, L.; López-Sánchez, H.; Castillo-González, F.; Lobato-Ortiz, R.; García-Zavala, J.J.; Ortega-Paczka, R. Morphological and agronomic characterization of seven maize races from the highlands of México. *Rev. Fitotec. Mex.* **2014**, *37*, 351–361.

15. Sánchez, G.J.J.; Goodman, M.M.; Stuber, C.W. Isozymatic and morphological diversity in the races of maize of Mexico. *Econ. Bot.* **2000**, *54*, 43–59. [CrossRef]

16. Mijangos-Cortés, J.O.; Corona-Torres, T.; Espinosa-Victoria, D.; Muñoz-Orozco, A.; Romero-Peñaloza, J.; Santacruz-Varela, A. Differentiation among maize (*Zea mays* L.) landraces from the Tarasca Mountain Chain, Michoacan, Mexico and the Chalqueño complex. *Genet. Resour. Crop Evol.* **2007**, *54*, 309–325. [CrossRef]

17. Santacruz-Varela, A.; Widrlechner, M.P.; Ziegler, K.E.; Salvador, R.J.; Millard, M.J.; Bretting, P.K. Phylogenetic relationships among North American popcorns and their evolutionary links to Mexican and South American popcorns. *Crop. Sci.* **2004**, *44*, 1456–1467. [CrossRef]

18. Lowe, A.J.; Hanotte, O.; Guarino, L. Standardization of molecular genetic techniques for the characterization of germplasm collections: The case of random amplified polymorphic DNA (RAPD). *Plant Genet. Resour. Newsl.* **1996**, *107*, 50–54.

19. Pan, Y.B.; Burner, D.M.; Legendre, B.L.; Grisham, M.P.; White, W.H. An assessment of the genetic diversity within a collection of Saccharum spontaneum L. with RAPD-PCR. *Genet. Resour. Crop Evol.* **2004**, *51*, 895–903. [CrossRef]

20.　Avise, J.C. *Molecular Markers, Natural History and Evolution*, 2nd ed.; Sinauer Associates: Sunderland, MA, USA, 2004; p. 684.

21.　De Queiroz, A.; Donoghue, M.J.; Kim, J. Separate versus combined analysis of phylogenetic evidence. *Annu. Rev. Ecol. Evol. Syst.* **1995**, *26*, 657–681. [CrossRef]

22.　Chippindale, P.T.; Wiens, J.J. Weighting, partitioning, and combining characters in phylogenetic analysis. *Syst. Biol.* **1994**, *43*, 278–287. [CrossRef]

23.　Leal, A.A.; Mangolin, C.A.; do Amaral, A.T., Jr.; Gonçalves, L.S.A.; Scapim, C.A.; Mott, A.S.; Eloi, I.B.O.; Cordovés, V.; da Silva, M.F.P. Efficiency of RAPD versus SSR markers for determining genetic diversity among popcorn lines. *Genet. Mol. Res.* **2010**, *9*, 9–18. [CrossRef]

24.　Kluge, A.G. A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). *Syst. Zool.* **1989**, *38*, 7–25. [CrossRef]

25.　Kluge, A.G.; Wolf, A.J. Cladistics: What's in a word? *Cladistics* **1993**, *9*, 183–199. [CrossRef]

26.　Nixon, K.C.; Carpenter, J.M. On simultaneous analysis. *Cladistics* **1996**, *12*, 221–241. [CrossRef]

27.　Gulanbaier, T.; Seoung, B.M. Gower distance-based multivariate control charts for a mixture of continuous and categorical variables. *Expert Syst. Appl.* **2014**, *41*, 1701–1707.

28.　Applied, B. *GeneMapper® Software Version 4.0. Reference and Troubleshooting Guide*; Applied Biosystems Inc.: Foster City, CA, USA, 2005; p. 82.

29.　Yeh, F.C.; Yang, R.C.; Boyle, T. *POPGENE Version 1.31. Microsoft Window-based Freeware for Population Genetic Analysis, Quik User Guide*; University of Alberta and Center for International Forestry Research: Edmonton, AB, Canada, 1999; Available online: http://www.ualberta.ca/~{}fyeh/popgene.pdf (accessed on 23 March 2018).

30.　SAS, Institute. *SAS/STAT User's Guide, Software version 9.0.*; SAS Institute Inc.: Cary, NC, USA, 2002; p. 4424.

31.　Martínez, G.A. *Manual de diseño y análisis de los látices. Monografías y Manuales de Estadística y Cómputo. Vol. 8, Núm. 3*; Centro de Estadística y Cálculo, Colegio de Postgraduados: Chapingo, Mexico, 1989; p. 71.

32.　Pla, L.E. *Análisis Multivariado: Métodos de Componentes Principales. Programa Regional de Desarrollo Científico y Tecnológico*; Secretaria General de la Organización de los Estados Americanos: Washington, DC, USA, 1986; p. 94.

33.　Gower, J.C. A general coefficient of similarity and some of its properties. *Biometrics* **1971**, *27*, 857–874. [CrossRef]

34.　Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425.

35.　Rohlf, F.J. *NTSYSpc: Numerical Taxonomy System, Version 2.21c*; Exeter Software, Setauket: New York, NY, USA, 2009.

36.　Herrera, C.B.E.; Castillo, G.F.; Sánchez, G.J.J.; Ortega, P.R.; Goodman, M.M. Morphological traits to assess diversity among maize populations in a region: Chalqueño landrace case. *Rev. Fitotec. Mex.* **2000**, *23*, 335–354.

37.　López-Romero, G.; Santacruz-Varela, A.; Muñoz-Orozco, A.; Castillo-González, F.; Córdova-Téllez, L.; Vaquera-Huerta, H. Caracterización morfológica de poblaciones nativas de maíz del Istmo de Tehuantepec, México. *Interciencia* **2005**, *30*, 284–290.

38.　Maturana, M.S.; Sánchez, C.J. *Las comunidades de la Meseta Tarasca, Un Estudio Socioeconómico*; Centro de Investigaciones Agrarias: Mexico City, Mexico, 1970; p. 124.

39.　Romero, P.J.; Castillo, G.F.; Ortega, P.R. Crosses of native populations of maize from Chalqueño race: II. Genetic groups, genetic divergence and heterosis. *Rev. Fitotec. Mex.* **2002**, *25*, 107–115.

40.　Doebley, J.F.; Goodman, M.M.; Stuber, C.W. Isozyme variation in the races of maize from México. *Am. J. Bot.* **1985**, *72*, 629–639. [CrossRef]

41.　Ángeles-Gaspar, E.; Ortiz-Torres, E.; López, P.A.; López-Romero, G. Characterization and yield performance of native maize populations from Molcaxac, Puebla. *Rev. Fitotec. Mex.* **2010**, *33*, 287–296.

42. Myles, S.; Peiffer, J.; Brown, P.J.; Ersoz, E.S.; Zhang, Z.; Costich, D.E.; Buckler, E.S. Association mapping: Critical considerations shift from genotyping to experimental design. *Plant Cell* **2009**, *21*, 2194–2202. [CrossRef] [PubMed]

43. Govindaraj, M.; Vetriventhan, M.; Srinivasan, M. Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. *Genet. Res. Int.* **2015**, *2015*, 431487. [CrossRef] [PubMed]