

# Genetic structure in cultivated quinoa (*Chenopodium quinoa* Willd.), a reflection of landscape structure in Northwest Argentina

S. M. Costa Tártara · M. M. Manifesto ·  
S. J. Bramardi · H. D. Bertero

Received: 21 May 2011 / Accepted: 28 March 2012 / Published online: 19 April 2012  
© Springer Science+Business Media B.V. 2012

**Abstract** Quinoa (*Chenopodium quinoa* Willd.), one of the main crops domesticated in the Andean highlands 1,000 of years ago, played an important role as a protein source. 35 germplasm accessions collected along the Northwest Argentina (NWA) region were studied using 22 microsatellite (SSR) markers. Results showed a great level of genetic diversity, differing from previous reports about the geographical distribution of quinoa variability. All SSR loci analysed were highly polymorphic detecting a total of 354 alleles among all populations, with an average of 16 alleles per locus. Cluster analyses grouped the accessions into four main clusters at the average genetic distance level (0.80), each of which represented a different environment of the NWA region: Puna ( $U_{He} = 0.42, \pm 0.07$  SE), Dry

Valleys ( $U_{He} = 0.27, \pm 0.05$  SE), Eastern Humid Valleys ( $U_{He} = 0.16, \pm 0.04$  SE) and a transition area with high altitudes between the last two environments ( $U_{He} = 0.25, \pm 0.03$  SE). An eastward decreasing genetic diversity gradient was found. AMOVA analyses showed a strong genetic structure: a high population subdivision relative to the grouping by region ( $F_{sr} = 0.47$ ) together with a high genetic differentiation among populations ( $F_{st} = 0.58$ ) and a heterozygous defect ( $F_{is} = 0.63$ ) in each of them. The variability structure, a reflection of the structure of the NWA landscapes, is discussed in connection with environmental variables.

**Keywords** Genetic diversity · Quinoa · Germplasm · Microsatellites · Northwest Argentina · Cluster analysis

S. M. Costa Tártara (✉) · M. M. Manifesto  
Instituto de Recursos Biológicos CIRN-INTA, N Repetto y Los  
Reseros s/n, 1686 Hurlingham, Buenos Aires, Argentina  
e-mail: scosta@cni.inta.gov.ar

S. M. Costa Tártara · H. D. Bertero  
Consejo Nacional de Investigaciones Científicas y Técnicas, Av.  
Rivadavia 1917, 1033 Buenos Aires, Argentina

S. J. Bramardi  
Facultad de Ciencias Agrarias, Universidad Nacional del  
Comahue, Ruta 151, Km 12, 8303 Cinco Saltos, Río Negro,  
Argentina

S. J. Bramardi  
Facultad de Ciencias Agrarias y Forestales, Universidad  
Nacional de La Plata, Av. 60 s/n La Plata, 1900 Buenos Aires,  
Argentina

H. D. Bertero  
Cátedra de Producción Vegetal, Facultad de Agronomía,  
Universidad de Buenos Aires, Av. San Martín 4453, 1417  
Buenos Aires, Argentina

## Introduction

Quinoa (*Chenopodium quinoa* Willd.), one of the crops domesticated in the South American Andean highlands 1,000 of years ago, played a significant role in its inhabitants' diet and culture because of its nutritional value and high tolerance to water stress, low temperatures and high salinity (Tapia et al. 1980). The species is distributed from Colombia to the South of Chile, from sea level (Chilean coast) up to more than 4,000 m.a.s.l., and its importance is higher in certain areas of the Bolivian Southern Altiplano, where it is the only seed crop under rain-fed conditions. Quinoa is an allotetraploid ( $2n = 4x = 36$ ) species, exhibiting disomic inheritance for most genetic markers, both morphological and molecular (Bonifacio 1990; Ward 2000; Maughan et al. 2004), and its reproductive system is primarily (~95 %) autogamous (Risi and Galwey 1984). In Andean quinoa, cultivated forms were defined by

Wilson (1990) as *C. quinoa* spp. *quinoa*, and weedy types, known as *ajaras*, as *C. quinoa* spp. *melanospermum*.

Quinoa shows a wide genetic variability with high adaptability to an extensive range of environments. In the Northwest Argentina (NWA) region, quinoa is currently cultivated in small plots, between latitude 22° and 27.5°S and between longitude 65° and 67.5°W, as a sole crop or intercropped with maize or potatoes (Brizuela 2010). The region presents contrasting environments and landscapes. On the West, the sub-region called Puna is formed by high plateaus with an average altitude of 3,500 m.a.s.l. It presents an arid climate with large thermal amplitude, frequent frosts and low rainfall (from no rain in the SW up to 300–400 mm per year in the NE). The Puna limits to the east with the “Cordillera Occidental”, which consists of two mountain chains: the Occidental chain with high peaks, some of them with permanent snow and an arid climate, and the Oriental chain with lower altitudes and a more humid climate. They are separated by valleys, being the most important one the Quebrada de Humahuaca, which is the main communication between the Bolivian highlands and the Eastern lowlands. The average monthly temperature in this area ranges from 7.5 to 16 °C and seasonal rainfall from 150 to 200 mm per year. To the South, there is a second dry valley called Valles Calchaquíes, which is similar to the Quebrada de Humahuaca in terms of climate (Lorenzini et al. 1999; weather section in DiPPEC, <http://www.dippec.jujuy.gov.ar/clima.html>). Native quinoa accessions are distributed in all these types of environment: to the west, in the arid elevations of the Puna; to the east, in the slopes of the Oriental chain, and between the two extremes in the Quebrada de Humahuaca and the Valles Calchaquíes (Curti et al. 2012).

During the last decades, the characterization of the species and the study of genetic relationships among *C. quinoa* populations from different South American regions were performed using morphological descriptors first and molecular markers later. Gandarillas (1979a) proposed classifying Andean quinoa from Ecuador, Peru and Bolivia based on their botanical characteristics and environment of origin. Quinoa varieties were then grouped according to the five major ecosystems where they are cultivated: valley (from the Interandean valleys), Altiplano (from the highland plateau in Bolivia and Peru), Salares (from the salt flats of Southern Bolivia), Sea level (from Central Chile) and Subtropical (from the Bolivian Yungas) types (Tapia et al. 1980). Biochemical markers were the first molecular tool used to characterize quinoa germplasm (Wilson 1988). The variation in electrophoretic patterns of 21 isozymes loci together with the morphometric data were used to compare 98 quinoa populations from South America. This work revealed two main elements: a Coastal type from south-central Chile and an Andean type distributed at elevations above 1,800 m from NWA to southern Colombia. This last group was also divided into northern and southern Andean quinoa. Ruas et al.

(1999) and del Castillo et al. (2007) used Random Amplified Polymorphic DNAs (RAPDs) not only to study the genetic relationships among different accessions of *C. quinoa* and related species but also to assess the genetic structure of *C. quinoa* populations from the Bolivian Altiplano respectively. Then, Anabalón-Rodríguez and Thomet-Isla (2009) used AFLPs markers together with morphological descriptors to characterize local quinoa varieties from North and Central Chile. Since the development of the first set of quinoa Simple Sequence Repeats (SSR or microsatellites) markers (Mason et al. 2005), more detailed studies of genetic variability could be performed. These markers are the best option for population studies, because of their codominant nature and their capacity to detect a high level of polymorphism. They have been largely used in several species to assess genetic diversity and population structure (i.e. Mondini et al. 2010; Asfaw et al. 2009; Naghavi et al. 2010). More recently, genetic diversity of the USDA (US Department of Agriculture) and CIP-FAO (International Potato Center) international nursery quinoa collection—which contains accessions from Peru, Bolivia, Ecuador, Argentina and Chile was assessed using SSR markers (Christensen et al. 2007). The accessions were clustered in two main groups: one of them included accessions from the lowlands of Chile together with a set of USDA accessions, while the other grouped those from the Andean highlands of Peru, Bolivia, Ecuador, Argentina and Chile. Fuentes et al. (2008) utilized highly polymorphic SSR loci and classified Chilean quinoa accessions into two discrete groups: from the southern lowlands and from the northern highlands.

*Chenopodium quinoa* spp. *quinoa* from the NWA region has been underrepresented in other germplasm collections and previous studies. Besides, this region represents the species' southern extreme of distribution within the central Andes and could provide novel adaptive characteristics (Wilson 1988; National Research Council 1989). The main objective of this study was to measure the genetic diversity level and to analyse the structure of that diversity in Argentinean germplasm using SSR markers, so as to generate useful information for genetic conservation and future breeding programs. The level of genetic diversity found in each population is discussed against information about its environment of origin, and results are compared with other reports of quinoa variability and structure using molecular markers (del Castillo et al. 2007; Christensen et al. 2007; Fuentes et al. 2008).

## Materials and methods

### Plant material

Thirty-five accessions of *C. quinoa* spp. *quinoa* collected in NWA were selected from a collection of 90 accessions

**Table 1** Passport data on *C. quinoa* spp. *quinoa* accessions native from the Northwest of Argentina, used for the analysis

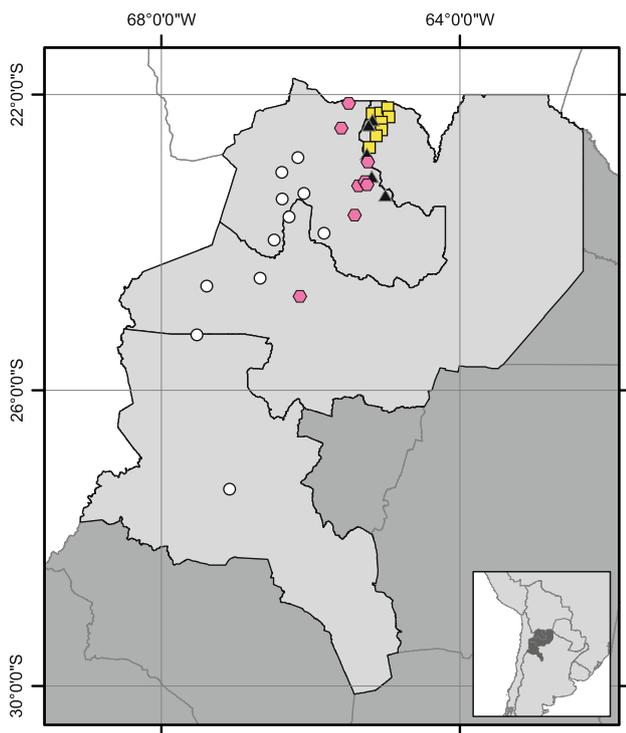
Donor ID	Passport/department origin	Province	Longitude	Latitude	Altitude
CHEN183	QQ 101-NSL 106396, Yavi	Jujuy	−65.47	−22.12	3,457
CHEN458	Morro de Pucará, Santa Victoria	Salta	−64.97	−22.18	2,645
CHEN451	Rodeo Pampa, Santa Victoria	Salta	−65.03	−22.23	2,898
CHEN482	Lizoite, Santa Victoria	Salta	−65.16	−22.26	3,310
CHEN212	San Felipe-Santa Victoria, Santa Victoria	Salta	−64.97	−22.27	2,507
CHEN466	San José del Aguilar, Santa Victoria	Salta	−65.17	−22.34	3,960
CHEN456	Trigo Huaico, Santa Victoria	Salta	−65.05	−22.36	3,230
CHEN465	Santa Cruz del Aguilar, Santa Victoria	Salta	−65.18	−22.38	3,955
CHEN434	Cangrejillos, Yavi	Jujuy	−65.58	−22.43	3,583
CHEN461	Poscaya, Santa Victoria	Salta	−65.08	−22.45	3,208
CHEN463	Aguadas, Santa Victoria	Salta	−65.11	−22.53	3,073
CHEN213	Chiyayoc, Iruya	Salta	−65.20	−22.73	3,200
CHEN256	Campo Luján, Iruya	Salta	−65.22	−22.78	3,000
CHEN438	Rachaite, Cochinoca	Jujuy	−66.15	−22.85	3,640
CHEN60	Abraite de Colanzulí, Iruya	Salta	−65.23	−22.90	3,711
CHEN431	Susques, Susques	Jujuy	−66.36	−23.04	3,619
CHEN446	La consulta 660-Aparzo, Humahuaca	Jujuy	−65.18	−23.08	3,823
CHEN58	Coctaca, Humahuaca	Jujuy	−65.28	−23.15	3,215
CHEN275	1485 Coctaca, Humahuaca	Jujuy	−65.28	−23.15	3,215
CHEN182	QQ 95-NSL 106394, Humahuaca	Jujuy	−65.33	−23.20	2,939
CHEN231	Ocumaso, Humahuaca	Jujuy	−65.25	−23.20	3,000
CHEN274	1484, Humahuaca	Jujuy	−65.33	−23.20	2,939
CHEN211	Barrancas, Cochinoca	Jujuy	−66.08	−23.33	3,500
CHEN468	Santa Ana, Valle Grande	Jujuy	−64.99	−23.36	3,379
CHEN184	QQ 87-NSL 101697, Susques	Jujuy	−66.36	−23.40	3,619
CHEN252	Maimará, Tilcara	Jujuy	−65.40	−23.62	2,334
CHEN432	Cobres, La Poma	Salta	−66.28	−23.65	3,591
CHEN269	1479, La Poma	Jujuy	−65.82	−23.85	3,480
CHEN272	1482, La Poma	Jujuy	−65.82	−23.85	3,480
CHEN427	Puesto Sey, Susques	Jujuy	−66.48	−23.95	4,012
CHEN426	Santa Rosa de los Pastos Grandes, Los Andes	Salta	−66.66	−24.47	3,939
CHEN422	Tolar Grande, Los Andes	Salta	−67.38	−24.58	3,533
CHEN414	La Poma, La Poma	Salta	−66.12	−24.72	3,016
CHEN420	Antofallita, Los Andes	Salta	−67.52	−25.25	3,498
BL	Barranca Larga, Belén	Catamarca	−67.07	−27.32	1,900

conserved at the Seed Bank at IRB-INTA Castelar (Table 1; Fig. 1). This collection includes materials from systematic field collection during 2006 and 2007, plus Argentinean accessions retrieved from collections in Mendoza, Argentina (INTA La Consulta), Bolivia (Proinpa Foundation), Peru (Universidad Nacional del Altiplano) and the Germplasm Bank at the USDA (National Plant Germplasm System, <http://www.ars-grin.gov/cgi-bin/npgs/html>). Seed samples were donated by farmers who reported their repetitive use year after year, and even for generations. Each donation was considered an accession for the Seed Bank and a population for this study. The criterion

used to select them was to cover different environments of the NWA region. Every accession was represented by ten individual plants in order to assess the genetic variation within populations.

#### Genomic DNA extraction

Five or six young leaves per plant were harvested and freeze-dried. Genomic DNA was extracted using a protocol described by Dellaporta et al. (1983), modified for rice (McCouch et al. 1988) and adapted for a micro-extraction in this study. DNA quality and concentration was checked



**Fig. 1** Geographical location of the 35 populations of *C. quinoa* spp. *quinoa* assessed in the present study. Different symbols show the four clusters of populations obtained after the grouping analyses representing four distinct environments of the NWA region: (circles) Puna; (hexagons) Dry Valleys; (triangles) Transition Zone and (squares) Eastern Humid Valleys

by 0.8 % agarose gel electrophoresis with UV transillumination.

#### PCR amplification

From the primers developed by Mason et al. (2005) and Jarvis et al. (2008), 35 and 4 SSR primers pairs, respectively, were selected according to the values of Heterozygosity (which ranged from 0.44 to 0.87), the Observed Number of Alleles reported (which ranged from 2 to 11), the chromosomal location and simple pattern amplification. Amplification reactions were carried out in a 7  $\mu$ l reaction mixture. Each reaction contained 10  $\mu$ M of each primer, 2.5 mM MgCl<sub>2</sub>, 0.200  $\mu$ M dNTPs, 0.5U *Taq* polymerase, and 30 ng of genomic DNA as template (Mason et al. 2005). Amplifications were performed with the following protocol: 94 °C for 1 min, as an initial denaturing, followed by 35 cycles at 94 °C for 30 s, annealing temperature (depending on the primer, between 50 and 61 °C) for 30 s, 72 °C for 1 min, and a final extension step at 72 °C for 10 min.

PCR amplification products were separated on a 6 % (w/v) polyacrylamide gel run in 0.5X/1X TBE at 50 W for 2–3 h (Life Technologies-Gibco BRL S2 Model) and

visualized with silver staining (Silver Sequence Promega Biotech, Madison, WI). The size of each band was estimated simultaneously by means of a 10 bp DNA Ladder using the program GEL (J Dubcovsky, unpublished) based on the reciprocal method (Elder and Southern 1987). In order to standardize allelic identity across gels, some samples from other populations were used as controls.

#### Data analysis

Informativeness is defined as the probability that a marker will distinguish between two randomly selected individuals in a population. In order to measure the marker's informative values (i) heterozygosity ( $H_e$ ) (ii) number of alleles per locus ( $A$ ) and (iii) Information Index ( $I$ ) were calculated for each locus. Heterozygosity, a measure of allelic diversity (Nei 1978), was estimated as follows:

$$H_e = 1 - \sum_{i=1}^k p_i^2$$

where  $p_i$  is the frequency of the  $i$ -th allele for the total and  $p_i^2$  is the sum of the squared total allele frequencies (Peakall and Smouse 2006). Information Index, equivalent to the Shannon and Weaver Index used in ecological studies, was also used to evaluate allelic diversity detected in each locus analysed:

$$I = - \sum [p_i * \ln(p_i)]$$

where  $p_i$  is the frequency of the  $i$ -th allele in every locus analysed (Peakall and Smouse 2006). This Index is based on the frequency of alleles in polymorphic loci and characterizes the locus (Nassiry et al. 2009), being the detected level of diversity higher as the value of  $I$  gets closer to 1. It is useful to choose the most informative SSR loci for their use in further genetic studies.

Population genetic diversity was estimated using the Unbiased Expected Heterozygosity ( $UHe$ ) measure at Hardy–Weinberg equilibrium, which is an indirect measure of genetic heterogeneity and estimates the probability that two alleles extracted at random from the set of genes can be different. The  $UHe$  is relative to population's size ( $N$ ), and it is recommended for small sample size ( $N < 50$ , Nei 1978). Since it ignores alleles in low frequencies and the percentage of Polymorphic Loci, genetic heterogeneity becomes less affected by sampling error (Nei and Chesser 1983).

$$UHe = \left( \frac{2N}{2N - 1} \right) * H_e$$

another measure at the population level was allelic richness, corresponding to the total number of alleles present in a population (Frankel et al. 1995) and the Percentage of

Polymorphic Loci. The presence of private alleles (exclusive to a single population) or rare alleles (present in a frequency lower than 0.05) was calculated using GenAlex 6.2 (Peakall and Smouse 2006).

The genetic structure of a plant species is defined as the non-random distribution of alleles or genotypes in space or time (Loveless and Hamrick 1984) which is influenced by their spatial structure or actual physical distribution. Using the Analysis of Molecular Variance (AMOVA) (Excoffier et al. 1992), the genetic structure of quinoa population was assessed by means of studying genetic differentiation through the partitioning of variance at different hierarchical levels. Deviations in Expected Heterozygosity ( $H_e$ ) in each level were quantified using the Fixation Index resulting from AMOVA (Wright 1921): within individuals ( $F_{it}$ ,  $F_{is}$ ), among populations ( $F_{st}$ ) and among regions ( $F_{rt}$ ), using GenAlex 6.2 software (Peakall and Smouth 2006).

Allelic frequencies per population were used to construct the genetic distance matrix according to Prevosti coefficient (Wright 1978). Cluster analysis using the unweighted pair group method with arithmetic mean (UPGMA) was performed with the NTSYS pc v-2.0 computer programme (Rohlf 1989). Correlations between the distance matrix and the cophenetic matrix were calculated by Pearson's product-moment and the significance of the correlations evaluated by Mantel's test (Mantel, 1967) (NTSYS program, MXCOMP module).

## Results

### Assessment of genetic diversity

From the 35 SSR markers selected for the study, only 22 were used according to their success in PCR amplification together with the clarity in the visualized pattern and the level of polymorphism detected (Table 2). All loci analysed amplified simple monogenic banding patterns, as it is shown in Fig. 2, and were easily scored in a codominant fashion (Maughan et al. 2004; Mason et al. 2005). Ten individuals of each population were run together and samples from other populations were used as controls in order to standardize allelic identity across gels. The total number of alleles detected in the 35 quinoa populations for the 22 SSR loci was 354, with a mean number of alleles per locus of 16 ranging from 6 (QCA067) to 29 (QAAT022). All SSR loci were highly polymorphic with Heterozygosity ( $H_e$ ) values above 0.70, except for the QAAT084 locus which presented a  $H_e$  value of 0.58. Values of Information Index ( $I$ ) ranged from 0.24 (QCA067) to 0.71 (QAAT022) showing more variability across loci. Ninety-two private alleles were found. These came from almost all loci

analysed with a frequency of 0.05 as a minimum (data not shown), so none of them were considered rare alleles.

Patterns of allelic diversity were calculated for each population (Table 3). Arithmetic mean values of genetic diversity ( $U_{He}$ ) and percentage of polymorphic loci were  $0.29 \pm 0.04$  and  $61.30 \pm 4.94$  % respectively. There was a wide range of variation in genetic diversity: from a population without polymorphism (CHEN274) in all loci analysed to highly diverse ones (CHEN272,  $U_{He} = 0.71$ ). Allelic richness presented a wide range of values, with an average value of  $49.71 \pm 4.21$  SE among populations. Accessions CHEN 427, 432, 272 and 269 showed the highest allelic richness together with the highest  $U_{He}$  values and 100 % of polymorphic loci. Population CHEN 446 presented polymorphism for every loci studied although other parameters were comparatively lower (allelic richness = 60,  $U_{He} = 0.38 \pm 0.01$ ).

### Cluster analysis

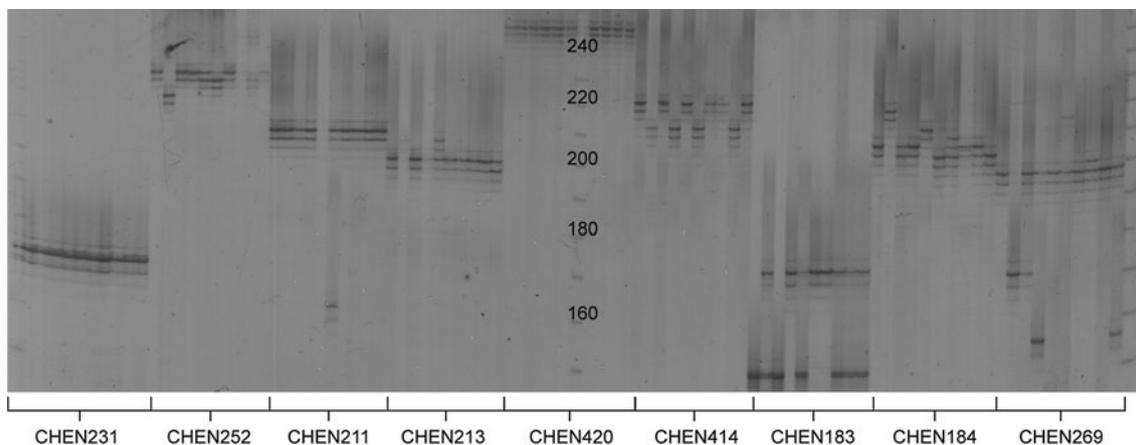
The average distance between all populations was  $0.80 \pm 0.01$  with a range from 0.17 (between CHEN 465 and 466) to 0.99 (between CHEN BL and 256). Cluster analysis separated the populations into four distinct groups, which were congruent with the different environments of origin (Fig. 3). Cluster one (C1), at the deepest node of the dendrogram, at a distance of 0.88, consisted of five accessions coming from a Transition area between the Quebrada de Humahuaca and the Oriental mountain chain, characterized by high altitudes. Cluster 2 (C2), the next to join at a 0.86 distance, consisted of twelve accessions collected in the Puna (the highland Plateau). At the last join, clusters 3 (C3) and 4 (C4), rooted together at a 0.84 distance, were composed of accessions from two different environments. C3 consisted of eight accessions from Santa Victoria and Iruya Departments (an Eastern region with humid valleys) and C4 was composed of ten accessions from the drier Quebrada de Humahuaca and Valles Calchaquíes. The dendrogram showed good cophenetic correlation with the corresponding genetic distance matrix ( $r = 0.90$ ,  $p = 0.001$  based on 1,000 permutations), revealing an accurate and useful graphic representation.

The level of genetic diversity contained in each cluster was measured. C2 was the most diverse ( $U_{He} = 0.42 \pm 0.07$  SE) and contained the highest quantity of private alleles (41). The levels of genetic diversity for C4 and C1 were quite similar ( $U_{He} = 0.27 \pm 0.05$  SE and  $U_{He} = 0.25 \pm 0.03$  SE respectively) but private alleles detected in C4 (27) were two times those in C1 (13). C3 was the least diverse ( $U_{He} = 0.16 \pm 0.04$  SE) presenting only 11 private alleles (Table 4). In addition, the highest percentage of polymorphic loci was detected in C2 corresponding with its genetic diversity level, while the lowest

**Table 2** SSR Loci, heterozygosity value (He), number of alleles (A), number of private alleles, Allelic size range (bp) and mean arithmetic value of Information Index (I) per locus, across populations

Locus	He	A	Private alleles	Allelic size range (bp)	I
QAAT074	0.82	14	4	167–207	0.536 (0.094)
QAAT088	0.90	21	3	112–200	0.706 (0.101)
QAAT001	0.89	19	5	134–215	0.581 (0.082)
QAAT097	0.89	18	4	167–218	0.627 (0.104)
QAAT076	0.92	20	4	142–215	0.582 (0.092)
QAAT087	0.91	19	6	166–235	0.510 (0.086)
QAAT112	0.78	11	3	185–220	0.409 (0.073)
QAAT050	0.84	16	6	184–219	0.397 (0.077)
QAAT027	0.87	20	3	147–187	0.586 (0.085)
QAAT062	0.90	18	1	155–206	0.594 (0.082)
QAAT106	0.75	12	4	287–319	0.398 (0.068)
KGA03	0.73	9	2	137–167	0.286 (0.062)
QAAT022	0.93	29	9	146–252	0.709 (0.098)
KGA020	0.89	15	5	156–199	0.383 (0.075)
QGA002	0.62	10	1	184–208	0.374 (0.079)
QAAT024	0.84	19	6	193–247	0.505 (0.078)
QAAT078	0.80	14	5	178–253	0.540 (0.075)
QAAT084	0.58	9	2	147–177	0.324 (0.077)
QGA03	0.77	20	10	129–197	0.498 (0.078)
QCA067	0.70	6	0	197–214	0.239 (0.067)
QAAT026	0.84	14	5	151–206	0.367 (0.072)
QAAT071	0.93	21	4	138–257	0.473 (0.087)
Total		354	92		
Arithmetic mean	0.82	16			

Numbers between brackets correspond to standard error (SE) value



**Fig. 2** Example of polymorphic microsatellite markers in *C. quinoa*. The figure reports the DNA amplification from accessions CHEN 231, 252, 211, 213, 420, 414, 183, 184 and 269 with QAAT022. Ten individuals of each population were run together. The size of each

band was estimated simultaneously by means of a 10 bp DNA Ladder (in *central* and *outside lanes*). Although only nine populations are presented in the image, this pattern was observed for all populations in every loci analyzed so bands were scored in a codominant fashion

value was in C3. Although in C4 there were ten populations in contrast with C1 (five), the latter presented a higher percentage of polymorphic loci showing more polymorphic populations.

As it was described before, the four clusters represented different environments or sub-regions (Figs. 1 and 4). From West to East, the Puna (C2), the Dry valleys (C4), the Transition zone (C1) and the Eastern humid valleys (C3)

**Table 3** Allelic pattern by population. Total number of alleles (A), arithmetic mean across loci of Unbiased expected Heterozygosity (Uhe)\*\* and Percentage of Polymorphic Loci (% P)

Population	A	Mean UHe	% P
CHEN274	20	0.00 (0.06)	0.00
CHEN463*	23	0.02 (0.03)	4.55
CHEN458	24	0.03 (0.04)	9.09
CHEN212*	26	0.03 (0.06)	18.18
CHEN434	30	0.07 (0.05)	36.36
CHEN060*	28	0.09 (0.02)	27.27
CHEN420	31	0.10 (0.06)	31.82
CHEN466*	34	0.11 (0.02)	45.45
CHEN213*	33	0.12 (0.02)	45.45
CHENBL*	32	0.14 (0.05)	45.45
CHEN456	32	0.16 (0.03)	40.91
CHEN211*	36	0.16 (0.06)	54.55
CHEN183*	37	0.16 (0.04)	50.00
CHEN451*	36	0.18 (0.03)	59.09
CHEN422*	38	0.21 (0.00)	59.09
CHEN256*	39	0.21 (0.05)	50.00
CHEN431*	45	0.24 (0.04)	63.64
CHEN465*	44	0.25 (0.06)	59.09
CHEN231*	41	0.27 (0.04)	59.09
CHEN252*	44	0.29 (0.02)	59.09
CHEN468*	43	0.30 (0.04)	63.64
CHEN414*	39	0.34 (0.05)	68.18
CHEN482*	52	0.35 (0.05)	81.82
CHEN461*	51	0.39 (0.05)	72.73
CHEN446*	60	0.39 (0.01)	100.00
CHEN058*	60	0.42 (0.05)	86.36
CHEN426	60	0.45 (0.03)	86.36
CHEN182*	57	0.45 (0.05)	81.82
CHEN438*	74	0.50 (0.05)	95.45
CHEN184*	73	0.53 (0.06)	95.45
CHEN275*	91	0.58 (0.06)	95.45
CHEN269*	98	0.61 (0.02)	100.00
CHEN427*	89	0.65 (0.04)	100.00
CHEN432*	96	0.70 (0.04)	100.00
CHEN272*	124	0.71 (0.05)	100.00
Arithmetic mean	49.71 (4.21)	0.29 (0.04)	61.30 (4.94)

\* Populations revealed private alleles

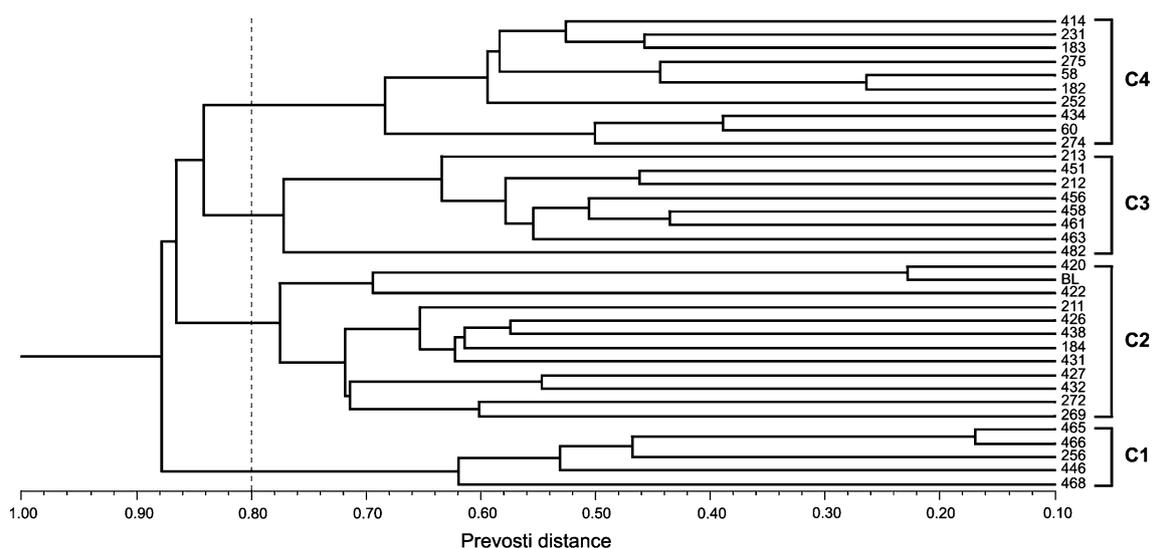
\*\* Numbers between brackets correspond to standard error (SE) value

show a decreasing eastward gradient of genetic diversity and a reduction in the number of private alleles (Table 4). In order to analyse the relationship between environmental variables and genetic diversity, a multiple linear regression (MLR) of genetic diversity (UHe) on 19 climatic variables related to temperature and precipitation from the Bioclim database (Hijmans et al. 2005), and to altitude from the Shuttle Radar Topography Mission (SRTM, Farr et al. 2007) was performed using JMP Base v. 9.0.0 (SAS Institute Inc. 2010). Software SAGA GIS v. 2.0.7 (Böhner et al. 2006) was used to extract each variable’s value according to each population’s geographic position. Using

the stepwise method, annual precipitation and precipitation of the wettest quarter showed a statistically significant effect ( $p < 0.01$ ). The  $R^2$  was moderate (0.49) and negative.

#### Genetic structure

Based on the four clusters resulting from cluster analysis, we performed an AMOVA analysis to consider the partition of total genetic diversity variance and the significance of its structure. Results showed that 20 % of total variance was partitioned among regions, 38 % among populations,



**Fig. 3** Dendrogram showing 35 quinoa accessions in the UPGMA cluster analysis. Populations were represented in four major clusters: (C1) Transition (C2) Puna (C3) Eastern Humid Valleys and (C4) Dry Valleys

**Table 4** Number of population (N), mean Genetic diversity (UHe)<sup>a</sup>, mean Percentage of Polymorphic Loci (% P)<sup>a</sup>, Number of Alleles (A) and Private alleles for the different Sub-regions

Cluster	N	UHe	% P	A	Private alleles
C1 (Transition)	5	0.25 (0.03)	63.64 (0.10)	220	13
C2 (Puna)	12	0.42 (0.07)	77.65 (0.07)	796	41
C3 (Eastern humid valleys)	8	0.16 (0.04)	41.48 (0.10)	277	11
C4 (Dry valleys)	10	0.27 (0.05)	56.36 (0.09)	447	27

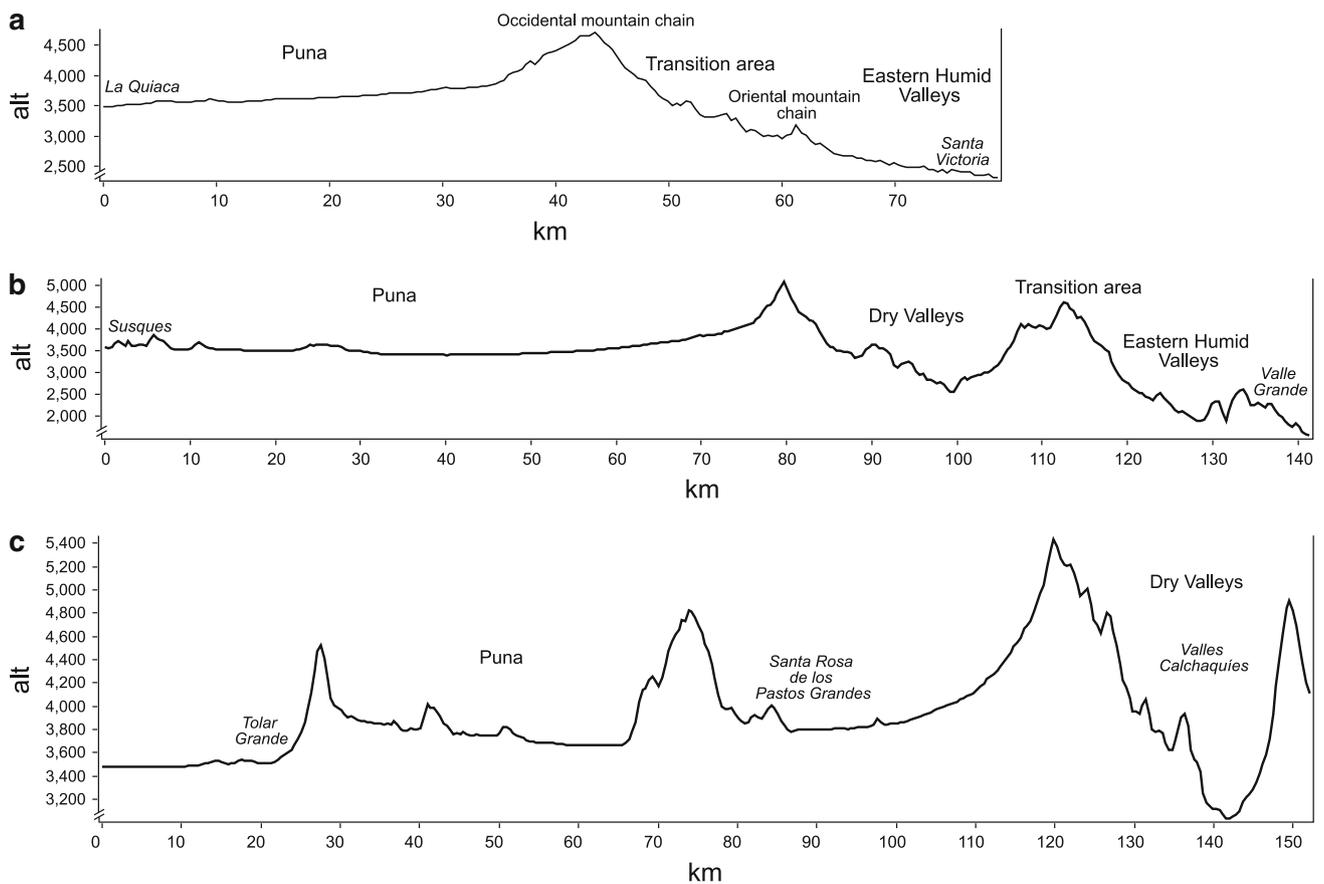
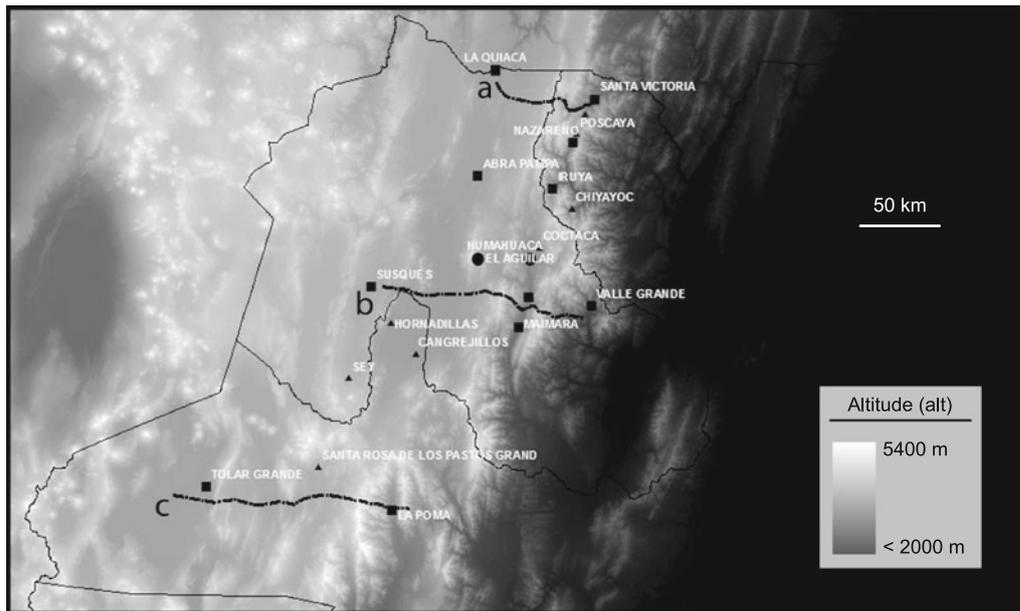
<sup>a</sup> Numbers between brackets correspond to standard error (SE) value

26 % among individuals, and 16 % within individuals. The fixation index (F<sub>st</sub>) showed structure by means of the clustering by region with a value of 0.20. F<sub>st</sub> = 0.58 indicated a great genetic differentiation among populations according to qualitative guidelines suggested by Wright (1978). In addition, F<sub>is</sub> = 0.63 and F<sub>it</sub> = 0.84 indicated a deficiency in heterozygous genotypes in relation to each subpopulation and to the total population respectively. All fixation index values were significant according to a statistical testing performed by using 1,000 random permutations ( $p = 0.001$ ) (Peakall and Smouse 2006).

## Discussion

This is the first molecular characterization of a representative sample of Argentinean quinoa germplasm. In this study, 35 accessions of *C. quinoa* spp. *quinoa* cultivated in the provinces of Salta, Jujuy and Catamarca characterized with SSR markers showed a very high level of genetic diversity. All loci were highly polymorphic presenting a large number of alleles and private alleles. The mean number of alleles per locus (16) was higher than in the

estimates obtained for Chilean quinoa (7.5, Fuentes et al. 2008) and in those reported for the CIP-FAO quinoa collection (11.7, Christensen et al. 2007). This comparison revealed higher levels of allelic diversity in NWA accessions than in Chilean germplasm and CIP-FAO collection. In comparison with these previous works, allelic richness (the total number) showed an intermediate value for Argentinean germplasm (354). The differences found in allelic richness could be the consequence not only of the differences in the number of accessions and their origin, but also of the methodology utilized. Some considerations and corrections were used to make results comparable (Lia et al. 2009). Seven of the 22 SSR loci assessed in our work were shared with these two previous studies. These loci retained 30, 33 and 52 % of total allelic richness observed in the Argentinean, CIP-FAO and Chilean germplasm respectively. Heterozygosity values were quite similar between studies with a high level of polymorphism, and the average number of alleles detected in the 7 loci in NWA were intermediate (15.3, against 20.1 and 11.1 for the CIP-FAO and Chilean germplasm respectively). It is remarkable that this diversity was found in spite of NWA representing a much smaller geographic area than those studied.



**Fig. 4** Digital Elevation Model of NWA with data from the Shuttle Radar Topography Mission (SRTM; Farr et al. 2007). Topographic variation of the region is shown through three transects at different latitudes corresponding to each graphic below (the geographic location is accurate). **a** on latitude 22.25°S from La Quiaca (longitude

65.60°W) to Santa Victoria Oeste (longitude 64.96°W); **b** on latitude 23.50°S from Susques (longitude 66.30°W) to Valle Grande (longitude 65°W); **c** on latitude 24.66°S from Tolar grande (longitude 67.40°W) through La Poma (longitude 66.20°W)

Another similarity among these works is that four out of the five Argentinean accessions characterized in Christensen et al. (2007) were included in our work: CHEN 274, 183, 184 and 272, and were used for comparison purposes. In our study, these four populations showed higher allelic diversity (254 total alleles and 17 private alleles) than in the CIP-FAO study (162 total alleles and 10 private alleles). This difference can be attributed to the larger sample size used in our study (ten individuals per population) which permitted a better detection of alleles and highlighted some of the limitations of previous studies.

The extensive variation in the range of genetic diversity values among quinoa populations is consistent with Schoen and Brown (1991), who concluded that self-pollinating populations exhibit markedly greater variation in Nei's genetic diversity statistic than outbreeding species. The average and range of genetic diversity values reported in our study were even higher than those found by del Castillo et al. (2007) among eight Bolivian quinoa populations characterized with RAPD markers (0.16 and 0.11, respectively). These comparisons should be treated cautiously however, as SSR markers are more informative than RAPD and allozyme markers (used in Schoen and Brown's work) due to SSR codominance and high polymorphism (Tautz 1989; Powell et al. 1996). Comparisons of the range of genetic diversity variation detected in our work are also limited since the other studies of quinoa germplasm characterization using SSR markers were performed using bulks of DNA instead of individual plants as sample.

The highest phenotypic diversity reported for the species is found near Lake Titicaca between Cuzco (Peru) and Lake Poopó (Bolivia), and this region is considered the centre of domestication of this species (Gandarillas 1979b; Pearsall 1992). As an important difference with previous reports (Wilson 1988, Christensen et al. 2007), we demonstrated that high levels of genetic diversity can be found away from the main areas of quinoa cultivation. NWA represents the Southern end of quinoa distribution in the Central Andes and is considered marginal in terms of its cultivation (Bertero 2001). In our work, the Argentinean germplasm presented a high level of allelic diversity (35 accessions and 354 total alleles), comparable to that of accessions from other origins (42 from Bolivia, 28 from Peru and 24 from Chile; Christensen et al. 2007), where the allelic diversity for these origins was 290, 271 and 198 respectively. Argentinean germplasm was underrepresented in previous characterizations, so the level of genetic diversity found in other studies could have been limited by this fact. However, the impact of sample size per population discussed before also needs to be considered.

The genetic structure analysis showed a strong differentiation among populations. While heterozygote deficiency was a result of its breeding system, it may also

reflect the influence of cultural practices since the accessions selected for this study consisted of local landraces donated by farmers who reported their repeated use for long periods. Notwithstanding the marked differentiation among populations, cluster analysis grouped them into four distinct clusters, each of which reflected a different environment of origin (Fig 3). Contrasting environments may have contributed over a long period to the development of a wide genetic diversity of quinoa germplasm adapted to very different climatic conditions. This was also supported by the fixation index which showed a high population subdivision relative to the grouping by region. Although comparisons between studies based on SSR markers and RAPD should be treated cautiously, similar results were revealed by del Castillo et al. (2007) where Bolivian quinoa showed a pattern of genetic differentiation related to the regional eco-geography. The effect of human mediated factors, like different cultures inhabiting the different environments, or the influence of ancient exchange routes affecting germplasm distribution, cannot be discarded as factors contributing to this genetic structure.

Phenotypic characterizations of Bolivian germplasm also found a strong correspondence with landscape structure (Rojas 2003). NWA presents a pattern of environmental variability similar to that of the South of Bolivia with marked longitudinal gradients of aridity associated with increasing altitude and decreasing precipitation westward. From West to East, the four clusters represented the Puna, the Dry Valleys, the high altitude Transition Zone and the Eastern humid valleys described at the beginning of this work (Figs. 1, 4). Even though the Transition zone is not a specific sub-region, it presents higher altitudes and drier conditions than the Eastern humid valleys, and populations from this zone showed a close relationship to each other. In a parallel characterization of phenotypic diversity of the same Argentinean germplasm (Curti et al., 2012), populations showed a similar structure to the results presented here. A decreasing Eastward gradient of genetic diversity and in the number of private alleles of each cluster was observed in our study (Table 4). The presence of private alleles makes it possible the definition of particular genetic configurations (Mondini et al. 2010) and can contribute to the identification of the most divergent accessions, linked to adaptation to extreme conditions, in which favourable genes for tolerance to such conditions can be searched. In our case, the Puna sub-region was genetically the most diverse, and the amount of private alleles detected in this sub-region could possibly be attributed to a better adaptation to the most extreme environment (harsh and dry conditions) included in our study. According to the MLR, the two variables related to precipitation showed statistically significant effects in the model which explains part of the genetic diversity

variation. A similar behaviour connected to variation of genetic diversity was observed in some tree species (Chase et al. 1995; Lowe et al. 2003; Cavers et al. 2003). Chase et al. (1995) speculated that the reason could be greater habitat diversity in dry regions as a consequence of local complexity in water availability within sites. In contrast, moist habitats may have been colonized more recently by moist-adapted ecotypes producing founder effects and resulting in lower population diversity. Detailed studies on quinoa germplasm from the Eastern Bolivian valleys (underrepresented in current collections) and its relations with germplasm from other South American origins may shed more light on quinoa regional patterns of diversity.

This work provides information about germplasm never collected or characterized before. Moreover, this study's results hold several implications for the understanding of the distribution and spatial structure of genetic variability in quinoa as well as for the conservation and future breeding programs.

**Acknowledgments** The authors wish to thank Ing Agr Marcos Angelini for his help in the making of maps and to Natalia Costa Tártara and Cristian Salemne for their help in the treatment. It is also important to thank the generous collaboration of farmers from NWA. This work was supported by FONCYT (Fondo Nacional de Ciencia y Técnica, BID 1728 OC-AR PICT 2005-23382). INTA (Instituto Nacional de Tecnología Agropecuaria) provided the financial support for accession's collection in the NWA region. SMCT gratefully acknowledges the support of a doctoral grant by FONCYT and CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas).

## References

- Anabalón-Rodríguez L, Thomet-Isla M (2009) Comparative analysis of genetic and morphologic diversity among quinoa accessions (*Chenopodium quinoa* Willd.) of the South of Chile and highland accessions. *J Plant Breed Crop Sci* 1:210–216
- Asfaw A, Blair MW, Almekinders C (2009) Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from the East African highlands. *Theor Appl Genet* 120:1–12
- Bertero HD (2001) Quinoa (*Chenopodium quinoa*, Chenopodiaceae) puede llegar a ser un cultivo importante para la agricultura Argentina. *Bol Soc Argent Bot* 36:3–4
- Böhner J, McCloy KR, Strobl J (eds) (2006) SAGA—analysis and modelling applications. Göttinger geographische abhandlungen, vol 115
- Bonifacio A (1990) Caracteres hereditarios y ligamiento factorial en la quinua. Tesis Ing Agr Universidad Mayor de San Simón, Cochamamba
- Brizuela L (2010) Quinoa Argentina. Situación actual de la quinua Argentina, Belén Catamarca. Dissertation, IPAF NOA, Jujuy, Argentina
- Cavers S, Navarro C, Lowe AJ (2003) A combination of molecular markers identifies evolutionarily significant units in *Cedrela odorata* L. (Meliaceae) in Costa Rica. *Conserv Genet* 4:571–580
- Chase MR, Boshier DH, Bawa KS (1995) Population genetics of *Cordia alliodora* (Boraginaceae), a neotropical tree. *Am J Bot* 82:468–475
- Christensen SA, Pratt DB, Pratt C, Nelson PT, Stevens MR, Jellen EN, Coleman CE, Fairbanks DJ, Bonifacio A, Maughan PJ (2007) Assessment of genetic diversity in the USDA and CIP-FAO international nursery collections of quinoa (*Chenopodium quinoa* Willd.) using microsatellite markers. *Plant Genet Resour* 5:82–95
- Curti RN, Andrade AJ, Bramardi S, Velásquez B, Bertero HD (2012) Eco-geographic structure of phenotypic diversity in cultivated populations of quinoa from Northwest Argentina. *Ann Appl Biol*. doi:10.1111/j.1744-7348.2011.00524.x
- Del Castillo C, Winkel T, Mahy G, Bizoux JP (2007) Genetic Structure of quinoa (*Chenopodium quinoa* Willd.) of the Bolivian altiplano as revealed by RAPD markers. *Genet Resour Crop Evol* 54:897–905
- Dellaporta S, Wood J, Hicks J (1983) A plant DNA miniprep: version II. *Plant Mol Biol Report* 1(4):19–21
- DiPEEC (Dirección Provincial de Planeamiento, Estadística y Censos). <http://www.dippeec.jujuy.gov.ar/clima.html>. Accessed Dec 2011
- Elder JK, Southern EM (1987) Computer-aided analysis of one dimensional restriction fragment gels. In: Bishop MJ, Rawlings CJ (eds) Nucleic acid and protein sequence analysis—a practical approach. IRL Press, Oxford, pp 165–172
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491
- Farr TG et al (2007) The shuttle radar topography mission. *Rev Geophys*. doi:10.1029/2005RG000183
- Frankel OH, Brown AHD, Burdon JJ (1995) The conservation of plant biodiversity. Cambridge University Press, Cambridge, pp 10–38
- Fuentes FF, Martínez EA, Hinrichsen PV, Jellen EN, Maughan PJ (2008) Assessment of genetic diversity patterns in Chilean quinoa (*Chenopodium quinoa* Willd.) germplasm using multiplex fluorescent microsatellite markers. *Conserv Genet* 10:369–377
- Gandarillas H (1979a) Botánica. In: Tapia ME et al (eds) La quinua y la kañiwa, cultivos andinos. IICA, Bogotá, pp 20–44
- Gandarillas H (1979b) Genética y origen. In: Tapia ME et al (eds) La quinua y la kañiwa, cultivos andinos. IICA, Bogotá, pp 45–64
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25:1965–1978
- Jarvis DE, Kopp OR, Jellen EN, Mallory MA, Pattee J, Bonifacio A, Coleman CE, Stevens MR, Fairbanks DJ, Maughan PJ (2008) Simple sequence repeats marker development and genetic mapping in quinoa (*Chenopodium quinoa* Willd.). *J Genet* 87:39–51
- Lia VV, Poggio L, Confalonieri VA (2009) Microsatellite variation in maize landraces from Northwestern Argentina: genetic diversity, population structure and racial affiliations. *Theor Appl Genet* 119:1053–1067
- Lorenzini HN, Echeverría MJ, Balmaceda R (1999) Geografía de la Argentina. AZ Editora, Buenos Aires, pp 143–164
- Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure of plant populations. *Ann Rev Ecol Syst* 15:65–95
- Lowe A, Jourde B, Colpaert N, Navarro C, Cavers S (2003) Genetic diversity in Costa Rican populations of mahogany (*Swietenia macrophylla*) II: Impact of logging on level, structure and dynamics. *Heredity* 90:268–275
- Mantel NA (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209–220
- Mason SL, Stevens MR, Jellen EN, Bonifacio A, Fairbanks DJ, Coleman CE, McCarty RR, Rassmussen AG, Maughan PJ (2005) Development and use of microsatellite markers for

- germplasm characterization in quinoa (*Chenopodium quinoa* Willd.). *Crop Sci* 45:1618–1630
- Maughan PJ, Bonifacio A, Jellen EN, Stevens MR, Coleman CE, Ricks M, Mason SL, Jarvis DE, Gardunia BW, Fairbanks DJ (2004) A genetic linkage map of quinoa (*Chenopodium quinoa* Willd.) based on AFLP, RAPD, and SSR markers. *Theor Appl Genet* 109:1188–1195
- McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of rice chromosomes. *Theor Appl Genet* 76:815–829
- Mondini L, Farina A, Porceddu E, Pagnotta MA (2010) Analysis of durum wheat germplasm adapted to different climatic conditions. *Ann Appl Biol* 156:211–219
- Naghavi MR, Hajikram M, Reza Taleei A, Jafar Aghaei M (2010) Microsatellite analysis of genetic diversity and population genetic structure of *Aegilops tauschii* Coss in Northern Iran. *Genet Resour Crop Evol* 57:423–430
- Nassiry MR, Javanmard A, Reza T (2009) Application of statistical procedures for analysis of genetic diversity in domestic animal populations. *Am J Animal Vet Sci* 4:136–141
- National Research Council (1989) Quinoa. In: *Lost crops of the Incas: little-known plants of Andes with promise for worldwide cultivation*. National Academy Press, Washington, pp 148–161
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583–590
- Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities. *Ann Hum Genet* 47:253–259
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295
- Pearsall DM (1992) The origins of plant cultivation in South America. In: Cowan CW, Watson PJ (eds) *The origins of agriculture*. Smithsonian Institution Press, Washington, pp 173–205
- Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. *Trends Plant Sci* 1:215–222
- Risi JC, Galwey NW (1984) The *Chenopodium* grains of the Andes: inca crops for modern agriculture. *Adv Appl Biol* 10:145–216
- Rohlf F (1989) NTSYS-pc numerical taxonomy and multivariate analysis system. Vs 2.0. Exeter Publ. LTD, Setauket
- Rojas W (2003) Multivariate analysis of genetic diversity of Bolivian quinoa germplasm. *Food Rev Int* 19:9–23
- Ruas PM, Bonifacio A, Ruas CF, Fairbanks DJ, Anderson WR (1999) Genetic relationship among 19 accessions of six species of *Chenopodium* L. by random amplified polymorphic DNA fragment (RAPD). *Euphytica* 105:25–32
- Schoen DJ, Brown AHD (1991) Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proc Natl Acad Sci* 88:4494–4497
- Tapia ME, Mujica SA and Cañahua A (1980) Origen, distribución geográfica y sistemas de producción de la quinua. In: *I reunión sobre genética y fitomejoramiento de la quinua*. PISCA-UNTA-IBTA-IICA-CIID, Puno, pp. A1–A8
- Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res* 17:6463–6471
- Ward SM (2000) Allotetraploid segregation for single-gene morphological characters in quinoa (*Chenopodium quinoa* Willd.). *Euphytica* 116:11–16
- Wilson HD (1988) Quinoa biosystematics I: domesticated populations. *Econ Bot* 42:461–477
- Wilson HD (1990) Quinoa and relatives (*Chenopodium* sect. *Chenopodium* subsect. *Cellulata*). *Econ Bot* 44:92–110
- Wright S (1921) Systems of mating. *Genetics* 6:111–178
- Wright S (1978) *Evolution and the genetics of populations*, vol 4. University of Chicago Press, Chicago