Genet Resour Crop Evol (2011) 58:727–739
DOI 10.1007/s10722-010-9615-z

RESEARCH ARTICLE

Development and use of microsatellite markers for genetic diversity analysis of cañahua (*Chenopodium pallidicaule* Aellen)

A. Vargas · D. B. Elzinga · J. A. Rojas-Beltran · A. Bonifacio · B. Geary · M. R. Stevens · E. N. Jellen · P. J. Maughan

Received: 26 April 2010 / Accepted: 8 September 2010 / Published online: 23 October 2010

Abstract Cañahua (*Chenopodium pallidicaule* Aellen) is a poorly studied, annual subsistence crop of the high Andes of South America. Its nutritional value (high in protein and mineral content) and ability to thrive in harsh climates make it an important regional food crop throughout the Andean region. The objectives of this study were to develop genetic markers and to quantify genetic diversity within cañahua. A set of 43 wild and cultivated cañahua genotypes and two related species (*Chenopodium quinoa* Willd. and *Chenopodium petiolare* Kunth) were evaluated for polymorphism using 192 microsatellite markers derived from random genomic cañahua sequences produced by 454 pyrosequencing of cañahua genomic DNA. Another 424 microsatellite markers from *C. quinoa* were also evaluated for cross-species amplification and polymorphism in cañahua. A total of 34 polymorphic microsatellite marker loci were identified which detected a total of 154 alleles with an average of 4.5 alleles per marker locus and an average heterozygosity value of 0.49. A cluster analysis, based on Nei genetic distance, clearly separated from wild cañahua genotypes from the cultivated genotypes. Within the cultivated genotypes, subclades were partitioned by AMOVA analysis into six model-based clusters, including a subclade consisting sole of erect morphotypes. The isolation by distance test displayed no significant correlation between geographic collection origin and genotypic data, suggesting that cañahua populations have moved extensively, presumably via ancient food exchange strategies among native peoples of the Andean region. The molecular markers reported here are a significant resource for ongoing efforts to characterize the extensive Bolivian and Peruvian cañahua germplasm banks, including the development of core germplasm collections needed to support emerging breeding programs.

Keywords Cañahuaf · Genetic diversity · Microsatellite markers · Population structure

A. Vargas · D. B. Elzinga · B. Geary · M. R. Stevens · E. N. Jellen · P. J. Maughan
Department of Plant & Wildlife Sciences, Brigham Young University, Provo, UT 84602, USA

A. Vargas · J. A. Rojas-Beltran · A. Bonifacio
Fundacion PROINPA, Casilla Postal 4285, Cochabamba, Bolivia

P. J. Maughan (✉)
Department of Plant & Wildlife Sciences, Brigham Young University, 285 WIDB, Provo, UT 84602, USA
e-mail: Jeff_Maughan@byu.edu

Introduction

Cañahua (*Chenopodium pallidicaule* Aellen) is a nutritionally important annual crop with regional importance in Bolivia and Peru. It is an annual diploid ($2n = 2x = 18$) species of the large, but

poorly studied, Amaranthaceae family, subfamily Chenopodioideae (Muller and Borsch 2005). Cañahua's growth habits can be classified as erect, semiprostrate or prostrate. Height (25–60 cm) and maturity (95–173 days) range dramatically among ecotypes (Flores 2006; Gade 1970; Galwey 1989; Rodríguez 2007). Current cultivation of cañahua occurs in two major centers in the general proximity of Lake Titicaca on the northern Altiplano, namely in the departments of La Paz, Bolivia and Puno, Peru, although minor cultivation extends into the Bolivian departments of Oruro, Cochabamba, and Potosi, as well as scattered valleys in the Peruvian Andes. Cañahua yields in the northern Altiplano range from 375 to 2,968 kg/ha (Flores 2006; Marin 2002; Rodríguez 2007). Cañahua is uniquely adapted to the harsh climatic conditions that characterize much of the Andean Altiplano, including extreme elevations (~4,000 m.a.s.l), frequent frosts and hail, and arid-saline soils (rain fall between 500 and 600 mm; pH 4.8; Risi and Galwey 1984). The Altiplano region has approximately three million inhabitants, most of whom are subsistence farmers, reinforcing cañahua's importance as a regional food security crop.

The percent protein content of the cañahua seed ranges from 14 to 19% (IPGRI et al. 2005), with a superior balance of essential amino acids—rivaling that of soybean (Repo-Carrasco et al. 2003). Moreover, cañahua flour is high in dietary fiber, lignin, and high antioxidant activity in the phenolic components analyzed, as well as iron and calcium (Repo-Carrasco-Valencia et al. 2009; Woods Páez and Eyzaguirre 2004). Traditional uses of cañahua include a milled flour or *pito* made from toasted cañahua grain used to treat altitude sickness and fatigue (Bonifacio 2003). Anciently cañahua was an integral part of the subsistence farmers' reciprocal food exchange/acquisition strategy—a seasonal food-trading network practiced between farming communities located at different altitudes (Gumerman 1997; Marti and Pimbert 2007; Rist 2000). Such strategies were important for ensuring food security and nutrition.

Genetic diversity is an important aspect of any crop improvement program (Friedt et al. 2007). A primary concern regarding the future improvement of cañahua is the loss of *in situ* genetic diversity. Poverty, political unrest and climate destabilization have accelerated a massive human exodus from the Andean Altiplano. Efforts to conserve the genetic diversity of cañahua have led to the collection and development of two main germplasm banks (Flores 2006; IPGRI et al. 2005). The largest collection, consisting of 801 accessions, is maintained at the National Germplasm Bank of High Andean Grains at the Foundation for the Promotion and Investigation of Andean Products (PROINPA) in Quipaquipani, Bolivia, while another 460 accessions are maintained by the Universidad del Altiplano in Illpa-Puno, Peru (UNA). Initial attempts are being made to develop core-breeding collections based on eco-geographical descriptors, morphology and molecular data.

Here we report the development of the first molecular markers for cañahua. Several of these markers are based on microsatellite sequences first identified in *Chenopodium quinoa* Willd. (Mason et al. 2005), a related Andean crop species, while others are cañahua-specific. The specific aims of this project were to (1) assess the potential use of *C. quinoa* microsatellite markers in cañahua; (2) develop and characterize new microsatellite markers specific to cañahua from random 454-pyrosequencing of the cañahua genome; (3) evaluate the correlation between morphological characteristics and genetic distance using microsatellite genotypic data; and (4) determine the patterns of population structure among cañahua genotypes.

Materials and methods

Plant material

A total of 43 diverse cañahua genotypes were used in this study, including thirty from Bolivia and 13 from the USDA-ARS germplasm collection (Table 1). Additionally, we included a single accession of *C. quinoa* (cv. 'Ollague') provided by A. Mujica (UNA, Puno, Peru) and a single genotype of *Chenopodium petiolare* Kunth, a sympatric diploid weed species collected by F. Fuentes (UNAP, Iquique, Chile), for comparative purposes. We use the terminology "genotypes" to describe the cañahua samples as they represent subsamples from potentially heterogeneous germplasm collections. Passport (origin) data, domestication status and morphological data are reported in Table 1. Morphological characteristics, collected from greenhouse-grown plants (Provo, Utah, 2008) and field plots (Letanias, Bolivia, 2005)

included: (1) growth habit (prostrate, bushy, erect); (2) plant color (yellow, pink, red); (3) seed coat color (white, brown, black); (4) pericarp color (beige, grey, khaki, light orange, dark orange, light brown, dark brown, firebrick, saddle brown); and (5) downy mildew symptoms (present, absent).

Cañahua plants grown in the greenhouses at Brigham Young University (Provo, Utah, USA) were sown in 12 cm pots using Sunshine Mix II (Sun Grow, Bellevue, WA), supplemented with 2 g per pot of Osmocote Plus® fertilizer (Scotts, Marysville, OH). Plants were maintained at 18°C under broad-spectrum halogen lamps with a 16-h light regime.

Cañahua specific microsatellite discovery, primer design and PCR

Genomic DNA used for sequencing and PCR was extracted from 30 mg of freeze-dried apical leaf tissue from greenhouse-grown plants following procedures described by (Todd and Vodkin 1996). To identify cañahua-specific microsatellites, randomly sheared genomic DNA of the genotype Ames 13223 was sequenced using 454 FLX pyrosequencing technology at Brigham Young University-DNA Sequencing Center (Provo, Utah). Raw DNA sequences were cleaned of remnant vector and scanned for perfect microsatellite motifs using the computer program MISA (Thiel et al. 2003) with the following parameters: di-, tri- and tetra-nucleotide motif repeat units of 8, 6 and 4, respectively. Forward and reverse primers were designed using the computer program Primer3 v2.0 (Rozen and Skaletsky 2000) with the following parameters: product size = 150–225, max Tm difference = 1°C and max polyX = 3. All primers were screened on a preliminary panel of eight cañahua genotypes (Table 1). The screening panel was used to eliminate primer pairs that failed to amplify or amplified complex banding patterns. Primer pairs that produced simple banding patterns (presumable single loci) were then screened on the full panel of 43 diverse cañahua genotypes. Cross-species *C. quinoa* microsatellite primer pairs were also screened for polymorphism using a smaller panel of three cañahua genotypes and one *C. quinoa* genotype (Table 1). Polymorphic markers derived specifically from cañahua genomic sequence were sequentially named with the suffix "CP" (*C. pallidicaule*) followed by microsatellite motif type (AT, AAT, and AAAT). *C. quinoa* derived microsatellites were previously labeled with the suffix "Q" or "K" as described by Mason et al. (2005) and Jarvis et al. (2008) (Table 2).

Amplification of the microsatellite markers was performed in 10-µl PCR reactions containing 30 ng genomic DNA, 0.5 units of HotStarTaq DNA polymerase (Qiagen-Aldrich, Inc., Saint Louis, MO), 0.5 µM forward primer, 0.5 µM reverse primer, and 0.1 mM cresol red. Thermal cycling profiles were as follows: 94°C for 30 s, followed by 19 cycles of 94°C for 60 s, 55°C for 30 s (decreasing 0.5°C every cycle), 72°C for 60 s; 34 cycles of 94°C for 30 s, 55°C for 60 s, 72°C for 60 s, followed by a final extension at 72°C for 10 min. PCR products and molecular ladder were separated on 3% Metaphor agarose gels (Cambrex Bio Science, East Rutherford, NJ) at 120 V for 5 h. All gels were run in 0.5× TBE buffer and visualized using ethidium bromide staining with UV transillumination.

Statistical analysis

Alleles for each microsatellite marker were scored as present, absent, or missing and converted into a binary matrix and analyzed using various statistical analysis programs. Evaluation of microsatellite marker polymorphism was measured by the heterozygosity (H) value for each marker as described by Nei (1978) using PowerMarker v3.0 (Liu and Muse 2005). The degree of inbreeding in the cañahua was estimated by fixation index (F), while the apparent outcrossing rates (t_a) were obtained from the inbreeding coefficient F [$t_a = (1 - F)/(1 + F)$] on the basis of the average expected and observed heterozygosities (Vaz et al. 2009).

The genetic distance among genotypes was calculated based on the matrices of allele frequencies using PowerMarker v3.0 (Liu and Muse 2005) and Nei distance to infer intrapopulation variation (Nei and Li 1979). The clustering criterion used was neighbor joining and the resulting dendrogram was rooted using *C. quinoa* and *C. petiolare* genotypes as the out-groups.

The population structure analysis was based on a Bayesian approach using the computer programs structurama (http://fisher.berkeley.edu/structurama/) and STRUCTURE (Evanno et al. 2005; Pritchard et al. 2000). Structurama was used to assign individuals to

Table 1 Passport data and morphological characteristics for the 43 cañahua genotypes used in the microsatellite analyses

ID	Name	Source	Origin	Latitude	Longitude	Altitude	DS[a]	GH[b]	PT[c]	SC[d]	PC[e]	DI[f]
cp1	0012-1	PROINPA	N/A	N/A	N/A	N/A	2	2	1	2	1	2
cp2*	0102-1	PROINPA	N/A	N/A	N/A	N/A	2	3	2	2	1	2
cp3*	0112-1	PROINPA	N/A	N/A	N/A	N/A	2	3	2	2	1	2
cp4	0113-1	PROINPA	N/A	N/A	N/A	N/A	2	3	1	2	1	2
cp5	0120-1	PROINPA	N/A	N/A	N/A	N/A	2	2	2	2	1	2
cp6	0181-1	PROINPA	N/A	N/A	N/A	N/A	2	2	2	2	1	2
cp7*	0271-1	PROINPA	Bolivia, Murillo	−16.5	−68.3	4100.8	2	2	2	3	6	2
cp8	0351-1	PROINPA	Bolivia, Murillo	−16.6	−68.2	3602.4	2	2	1	2	1	2
cp9	LRN-2	PROINPA	Bolivia, Desaguadero	−16.6	−69.0	3886.8	2	2	2	3	2	2
cp10	LRB-1	PROINPA	Bolivia, Desaguadero	−16.6	−69.0	3886.8	2	2	2	2	1	2
cp11**	Line-1(05)	PROINPA	Bolivia, Caquiaviri	−16.9	−68.5	4022.4	2	2	2	2	8	2
cp12*	Line-2(05)	PROINPA	Bolivia, Comanche	−16.8	−68.4	3960.6	2	2	2	2	1	2
cp13	Line-3(05)	PROINPA	Bolivia, Curahuara	−17.5	−68.5	3832.9	2	3	3	2	4	2
cp14	Line-4(05)	PROINPA	Bolivia, Curahuara	−17.5	−68.5	3832.9	2	2	3	2	8	2
cp15	0019-1	PROINPA	N/A	N/A	N/A	N/A	2	3	2	2	1	2
cp16	0102-2	PROINPA	N/A	N/A	N/A	N/A	2	2	2	3	2	2
cp17*	0441-1	PROINPA	Bolivia, Aroma	−17.3	−68.0	3813.7	1	2	2	3	6	2
cp18	0472-1	PROINPA	Bolivia, Aroma	−17.3	−68.0	3813.7	2	2	1	2	1	2
cp19	0475-1	PROINPA	Peru, Puno	−15.7	−70.1	3828.6	2	2	3	3	9	2
cp20	0517-1	PROINPA	Bolivia, Ingavi	−16.8	−69.0	3828.0	2	2	2	2	1	2
cp21	0017-1	PROINPA	N/A	N/A	N/A	N/A	2	2	2	2	3	2
cp22	0225-2	PROINPA	N/A	N/A	N/A	N/A	2	2	2	2	1	2
cp23	0468-1	PROINPA	Bolivia, Aroma	−17.3	−68.0	3813.7	2	2	2	2	1	2
cp24	0468-2	PROINPA	Bolivia, Aroma	−17.3	−68.0	3813.7	2	2	2	2	2	1
cp25	0568-1	PROINPA	Bolivia, Omasuyus	−16.0	−68.7	3842.6	2	3	3	2	5	2
cp26	Line-9(05)	PROINPA	Bolivia, Omasuyus	−17.6	−68.4	3864.9	2	3	2	2	1	2
cp27	0568-2	PROINPA	Bolivia, Omasuyus	−16.0	−68.7	3842.6	2	2	3	2	7	2
cp28	0636-1	PROINPA	Bolivia, Coromata	−16.2	−68.5	3983.7	2	2	2	3	2	2
cp29**	PI 478406	USDA	Bolivia, La Paz	−16.3	−68.6	3891.7	2	2	2	3	2	2
cp30	PI 478407	USDA	Bolivia, La Paz	−16.3	−68.5	3912.1	2	3	3	2	4	2
cp31	PI 510525	USDA	Peru, Puno	−15.7	−70.1	3839.0	2	2	2	2	1	2
cp32	PI 510526	USDA	Peru, Puno	−16.4	−69.2	3883.5	2	2	2	2	1	2
cp33*	Ames 13221	USDA	Bolivia, La Paz	−16.2	−68.8	4259.0	2	2	2	2	1	2
cp34	Ames 13222	USDA	Bolivia, La Paz	−16.2	−68.8	4258.1	2	2	2	2	1	2
cp35	PI 510531	USDA	Peru, Puno	−16.0	−69.8	3834.1	2	2	2	2	6	2
cp36	PI 510529	USDA	Peru, Puno	−15.9	−70.0	3833.5	2	2	2	2	1	2
cp37	PI 510528	USDA	Peru, Puno	−16.1	−69.7	3853.9	2	2	2	2	1	2
cp38	Ames 13223	USDA	Bolivia, La Paz	−16.3	−68.6	3880.4	2	2	3	2	4	2
cp39**	Ames 13224	USDA	Bolivia, La Paz	−16.3	−68.5	3912.7	2	2	2	3	2	2
cp40	PI 510527	USDA	Peru, Puno	−16.2	−69.4	4007.5	2	2	2	2	1	2
cp42	0351Mut-45	PROINPA	Bolivia, Ingavi	−16.6	−68.2	3603.0	2	2	1	2	1	2
cp43	Wild-2	PROINPA	Bolivia, Ingavi	−16.6	−68.2	3603.7	1	2	1	3	7	1
cp44*	Wild-4	PROINPA	Bolivia, Ingavi	−16.6	−68.2	3604.3	1	1	2	3	7	2
C. petiolare	564	UAP	Chile, Chijo	−19.4	−68.6	3918.0	1	N/A	1	3	6	N/A

Table 1 continued

ID	Name	Source	Origin	Latitude	Longitude	Altitude	DS[a]	GH[b]	PT[c]	SC[d]	PC[e]	DI[f]
*C. quinoa***	cv. Ollague	CIP-FAO	Chile, Ollague	−21.2	−68.3	4122.4	2	N/A	3	1	N/A	1

The preliminary screening panels for *C. quinoa* based and *C. pallidicaule* specific microsatellites are marked with "*" and "**", respectively. Morphological traits include domestication status (DS), growth habit (GH), seed coat color (SC), pericarp color (PC), and disease presence (DI)

N/A Data not available

[a] 1 = Wild, 2 = Cultivated

[b] 1 = Prostrate, 2 = Bushy, 3 = Erect

[c] 1 = Yellow, 2 = Pink, 3 = Red

[d] 1 = White, 2 = Brown, 3 = Black

[e] 1 = Beige, 2 = Grey, 3 = Khaki, 4 = Orange, 5 = Dark orange, 6 = Brown, 7 = Dark brown, 8 = Firebrick, 9 = Saddle brown

[f] 1 = Present, 2 = Absent

each subpopulation. The subpopulation numbers (k) were summarized using a Bayesian Markov Chain Monte Carlo (MCMC) analysis of population structure using the mean partition (Huelsenbeck and Andolfatto 2007). The sitting parameters for structurama were: MCMC ngen = 2,000,000 samplefre = 1,000, printfreq = 1,000 and nchains = 2. Later the predefined population number was placed into STRUCTURE to infer the population structure using 1,000,000 MCMC reps after a burn-in period of 100,000 and a model allowing for admixture and correlated alleles frequencies (Achigan-Dako 2008). An admixture proportion was given by the average value of Q, which gave the probability that each genotype corresponded to each cluster.

An analysis of molecular variance (AMOVA) (Excoffier et al. 1992) was carried out using Arlequin 3.5.1.2 (Excoffier et al. 2005) to estimate genetic variance components and to partition the total variance within and among clusters generated from the STRUCTURE program. The significance of variance components was tested using non-parametric permutations of the data set with 1,000 permutations (Excoffier et al. 1992; Weir 1996; Weir and Cockerham 1984). The partitioned genetic effects were attributable to differences among hierarchical clusters (F_{ST}), among clusters within hierarchical cluster (F_{SC}) and among genotypes across the entire collection (F_{CT}). A Pearson correlation analysis was used to identify pattern of association between genetic distances, morphological characteristic, and clusters found by the STRUCTURE program.

The isolation by distance (IBD) (Wright 1946) analysis was performed using a Mantel test in the ade4 package in the R program (Bohonak 2002; Dray and Dufour 2007). First, the matrix pairwise Nei genetic distances (Nei and Tajima 1983) were calculated using PowerMarker v3.0 (Liu and Muse 2005). Second, geographic locations of the genotypes were converted into kilometer distances and then arranged in a matrix. Third, the mantel test was used to assess the correlation between these two matrices in order to understand the correlation between spatial location and the genetic distance of the genotypes. Then, the correlation matrix was calculated and plotted. This correlation tested the null hypothesis that the Bolivian and Peruvian genotypes' correlation is exactly zero.

Results

Marker characterization in cañahua

454-pyrosequencing produced 116,912 sequences (~28.986 Mb). With a genome size of 1,044 Mb (estimated by flow cytometry, data not reported), approximately 0.0277× of the total genome was sequenced. Automated searches for microsatellite motifs within the preassembled sequence using the computer program MISA identified a total of 1,000 potential microsatellite markers, including 389 di-, 538 tri- and 73 tetra-nucleotide repeat motifs or an approximate microsatellite density of 1 per 29 kb.

Table 2 Cañahua and *C. quinoa* microsatellite marker characteristics, including GenBank accession ID, primary motif, complexity, type, PCR primers, product size (PRO), observed number of alleles (ONA), and heterozygosity (*H*)

	Marker name[a]	GenBank accession ID[b]	Primary motif	Complexity	Type	Forward primer (5′-3′)	Reverse primer (5′-3′)	PRO	ONA	*H*
1	CPAAT003	HQ156786	$(AAT)_{14}$	Simple	Perfect	TTTGACTTTCTTGCCCTATTTACA	CGCTATGTATTCCAATGCGA	228	2	0.3084
2	CPAAT015	HQ156787	$(AAT)_{13}$	Simple	Perfect	CGTACAAAGTATAATTAGTGGCGA	CAGAGCACGACAGAGCAGAA	152	4	0.5680
3	CPAC001	HQ156788	$(AC)_{8}$	Simple	Perfect	GTTAAGCTCCACAATCCTTTGA	ATTCCACCGGTTGTCACTTT	225	2	0.2832
4	CPAT002	HQ156789	$(AT)_{8}$	Simple	Perfect	AAGTGTATTAACCGCCCGTG	AGCTCCTGGCTTAGTTTCAAGA	214	2	0.3261
5	CPAT005	HQ156790	$(AT)_{9}$	Simple	Perfect	GGAATCCAACACTTCAGTGGA	AGTTAGTCCGGGTAAAGCCC	202	2	0.4955
6	CPAT034	HQ156791	$(AT)_{9}$	Simple	Perfect	CAAACAGAAAGTAAAATAGCAAAAGAA	ATATGCAGGGTCTTGGTTGG	141	4	0.6675
7	CPAT045	HQ156792	$(AT)_{9}$	Simple	Perfect	ATGGTATCGGAGCCAACAAG	GCTGCGAGACCAGCTTTATT	138	8	0.6936
8	CPAT055	HQ156793	$(AT)_{10}$	Simple	Perfect	AGTCAATGTCGTGTCTCAATATACA	GGTTAATGAAAGGTCCAGGC	172	4	0.5944
9	CPATA010	HQ156794	$(ATA)_{9}$	Simple	Perfect	AAGAGGAGGAGGAGTAGTAGTTGTAG	AGAAAACCAGAAGACTGGTCC	182	7	0.7670
10	CPATT005	HQ156795	$(ATT)_{13}$	Simple	Perfect	GTTTGGGTGAAGAGCGAATC	CCGCGTTAGTATCATCAATCG	126	5	0.5635
11	CPATT006	HQ156796	$(ATT)_{30}$	Simple	Perfect	TTGTTAATTTGGGCAAGCAA	TCTCGTGACCTGACTGTGAGA	210	4	0.4551
12	CPATT013	HQ156797	$(ATT)_{12}$	Simple	Perfect	TCCGATTTAATTCAAATTCTCTTACA	AAGTGTGACACATATAATCGGAGC	162	5	0.1363
13	CPATTT009	HQ156798	$(ATTT)_{6}$	Simple	Perfect	TTTCTGTTGAAAACTCATTGCC	CAACTTGGCCAAAATGATGA	149	3	0.2506
14	CPCA003	HQ156799	$(CA)_{12}$	Simple	Perfect	CCACATCACTTCTGCCGATA	AAGTTAAATCAGTTTGGCTCCTTT	136	2	0.4124
15	CPCAG001	HQ156800	$(CAG)_{14}$	Simple	Perfect	GGGTTATTGTCAATGCTAGTGAA	TTGAACACTCTTGACCTTCTTCT	107	4	0.4082
16	CPGTA002	HQ156801	$(GTA)_{9}$	Simple	Perfect	AGTTAGGGCGGTAGGGAGAG	GGCTATGGAGGTCGACAAAT	103	2	0.4444
17	CPTA001	HQ156802	$(TA)_{8}$	Simple	Perfect	TCAAACTCAAGTAATATTGCAGATCA	TTCCATGGCATTTCTTCACA	179	4	0.4536
18	CPTAA001	HQ156803	$(TAA)_{12}$	Simple	Perfect	GGTGGAGGAGAACAACCTCA	GACATTATAATTGGAAGTCCACCC	187	5	0.3602
19	CPTAT012	HQ156804	$(TAT)_{13}$	Simple	Perfect	TTGCTGACCCCTTCTTCTTC	TCACAACAATTAAAATAAAACCCC	198	5	0.4033
20	CPTTA016	HQ156805	$(TTA)_{9}$	Simple	Perfect	AGTTTGACTCCTTCACCCCC	CCTACCAGCTTCACGTCACA	172	4	0.5152
21	CPTTA017	HQ156806	$(TTA)_{10}$	Simple	Perfect	CGTTGACTTGCAGAGCTGAG	CGAATTGAGCTGAATCGAACT	201	2	0.4800
22	CPTTA022	HQ156807	$(TTA)_{14}$	Simple	Perfect	AGTTGGTGTGAAACCCGTACT	TGCGATTCAATATGCTTTATCA	136	7	0.6851
23	KCAA010	Jarvis et al. (2008)	$(GGT)_{4}$	Simple	ImPerfect	TGGGTCGTAGTTCTGGGTTC	CTTATCACCAGCAGCAGCAC	191	4	0.5809
24	KCAA041	Jarvis et al. (2008)	$(GTT)_{8}$	Simple	Perfect	TGGTCGTAGACCACCATTT	CGGATCACTCCACCCTTGTA	197	6	0.7290
25	KCAA132	Jarvis et al. (2008)	$(CAA)_{5}$	Simple	ImPerfect	CAAACTGCAGGCACCACA	CAACTTCACCATAGGCCAAAA TGATGA	220	2	0.4444
26	KCAA139	Jarvis et al. (2008)	$(CAA)_{3}$	Simple	ImPerfect	GAACACCCAACCTGCAAACT	CAACTTCACCATACGCATTCA	180	4	0.2101
27	QAAT076	Mason et al. (2005)	$(ATT)_{30}$	Simple	Perfect	GCTTCATGTGTTATAAAATGCCAAT	TCTCGGCTTCCCACTAATTTT	197	6	0.7215

Table 2 continued

	Marker name[a]	GenBank accession ID[b]	Primary motif	Complexity	Type	Forward primer (5′-3′)	Reverse primer (5′-3′)	PRO	ONA	*H*
28	QAAT100	Mason et al. (2005)	$(ACTACC)_8$	Compound	Perfect	GGCATCCAGAGGTCAGTCTT	GCAATTCTTCCTAATAACAACAACAA	349	6	0.3611
29	QATG001	Mason et al. (2005)	$(CAT)_8$	Simple	Perfect	CATAACATCCTATAGAACCGTTGC	TGCTCCTACTTTACTCATTGACTCAT	200	4	0.3682
30	QCA005	Mason et al. (2005)	$(CA)_{16}$	Simple	Perfect	GTGGTTCATGGCTGATCCTT	CTTGCCATCAGGGCATATCT	185	3	0.3534
31	QCA024	Mason et al. (2005)	$(CA)_{17}$	Compound	Perfect	AGATGAGCTTGAATCATTACATC	TACATACTGTAAATCAGCCAAA	250	3	0.5714
32	QCA038	Mason et al. (2005)	$(CT)_{10}(CA)_{13}$	Compound	Perfect	CATTTCCCAAACTGCATGAAT	ATGTGTGTTGCGTGTGAGTG	197	3	0.5559
33	QCA 055	Mason et al. (2005)	$(TG)_{14}$	Simple	Perfect	GGGCATATCTGAAGGAAAGA	ACGCAGGTAGCACTTCCAGT	198	3	0.3534
34	QCA076	Mason et al. (2005)	$(CA)_{12}$	Simple	Perfect	AGGGAGGGTTGAAGGAAAGA	TCTTGTGGACTTGCATGCTT	150	3	0.5214
	Mean							183	3.9	0.4718

[a] Prefix "CP" primers derived from *C. pallidicaule*; prefix "Q" and "K" primers derived from *C. quinoa*

[b] GenBank accession numbers are given for all cañahua-specific micsatellites, while publication references are provided for the *C. quinoa* based microsatellites

The most abundant di-, tri-, and tetra-nucleotide motifs identified were AT (71%), AAT (68%), and AAAT (70%), respectively (Fig. 1). When 192 of these putative microsatellite loci were tested for polymorphism, 161 (85%) produced simple amplification products; however, only 22 (11.5%) were polymorphic in the full screening panel and non-redundant (Table 2; Fig. 2). Of these polymorphic microsatellite loci, the AT-rich motifs were clearly the most polymorphic, representing greater than 82% of all polymorphic motifs. To accumulate additional microsatellite loci, we also tested the transferability of microsatellite markers, originally identified in *C. quinoa*, to cañahua. In total, 424 *C. quinoa* microsatellite markers (Jarvis et al. 2008; Mason et al. 2005) were similarly evaluated for polymorphism using the cañahua preliminary and full screening panels. Of these primer pairs, 170 (40%) amplified simple banding patterns including 12 (6%) that were clearly polymorphic. Of these 12 microsatellite loci, the most common polymorphism motifs were the di-nucleotide CA (42%) and the tri-nucleotide CAA (33%). We note that the change in the most common polymorphic motif is likely due to the use of enriched motif libraries during the development of the *C. quinoa* markers (Mason et al. 2005). Table 2 describes each polymorphic marker according to marker name, primary motif, complexity (simple/compound), type (perfect/imperfect), amplification primer sequences, and expected PCR product size.

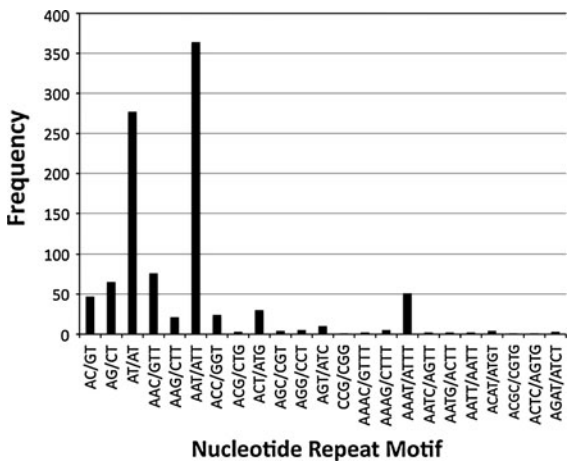

Fig. 1 Frequency of microsatellite repeat motifs identified in cañahua based on 1,000 microsatellite motifs identified from the 454-pyrosequencing data using the computer program MISA

Genetic diversity

The information content of a microsatellite marker is a measure of the observed number of alleles (ONA) and their corresponding relative frequencies within the test population and is often expressed in terms of

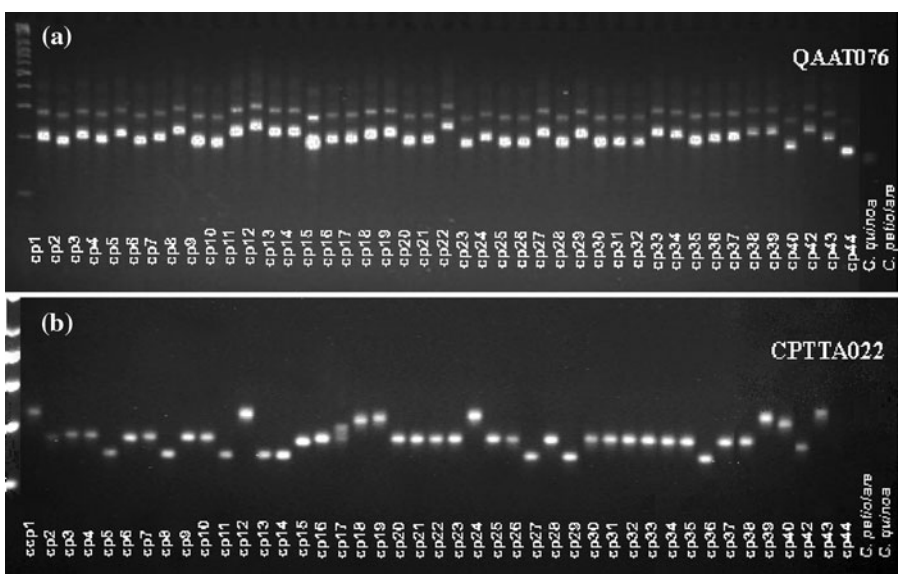

Fig. 2 Example of polymorphic microsatellite markers in cañahua (*lanes 2–44*), *C. petiolare* and *C. quinoa* (*lanes 45–46*). *Panel a* shows marker QAAT076 and *panel b* shows marker CPTTA022. Standards are loaded in *lane 1*

heterozygosity (H) (Höglund 2009). Screening the full panel of 43 cañahua genotypes identified 154 alleles spread across the 34 microsatellite loci, ranging from a low of 2 to a high of 8 alleles per locus, or an average of 4.5 alleles per marker locus. Heterozygosity values for the microsatellite loci ranged from 0.176 to 0.767, with a mean value of 0.487. According to Ott's (1992) criteria, a marker locus is considered highly polymorphic if $H \geq 0.70$, and polymorphic if $H \geq 0.10$. Thus, all of the markers were considered polymorphic and three markers were considered highly polymorphic (Table 2). We note that these H values may be underestimated since 3% Metaphor agarose was used to resolve the microsatellite alleles. In our experiments, 4-bp resolution was achieved; however, others have reported 2-bp resolution with 3.5% Metaphor agarose (Groben and Wricke 1998). The use of Metaphor agarose parallels electrophoresis technology available in many developing countries and represents a cost-effective method to analyze microsatellite markers in terms of technical expertise and reagent cost.

Relatedness of the cañahua genotypes

The unrooted tree shows a sub-branch consisting of the wild cañahua genotypes (cp43, cp17 and cp44) and the outgroup species (Fig. 3a). Similarly, the rooted Neighbor Joining (NJ) analysis revealed a sub-clade characterized by a single wild cañahua genotype (cp44) and the out-group species (Fig. 3b). A second sub-clade consisted of the remaining two wild cañahua genotypes (cp43 and cp17). We note that the wild genotypes (cp44, cp43, and cp17) were collected at the periphery of cultivated fields and were categorized as wild cañahua based solely on morphological criteria. The positioning of the wild genotypes, cp43 and cp17, is of interest as they are grouped much closer to the cultivated cañahuas, but are clearly genetically dissimilar, suggesting their potential to serve as novel sources of genetic variation within emerging cañahua breeding programs.

The cultivated genotypes were clearly distributed into two main clades, each with numerous subclusters. Analysis of the data with the computer program STRUCTURE produced the highest log-likelihood scores when the number of subclusters (k) was set at six, including five subclusters of cultivated genotypes and one clustered that contained all of the wild genotypes (Fig. 3b). Of the five cultivated clusters, little differentiation based on morphological characteristics was observed, except for cluster 1, which consisted of seven of the eight erect growth-habit genotypes. The remaining erect growth-habit genotype was found in cluster 4. In addition to the 26

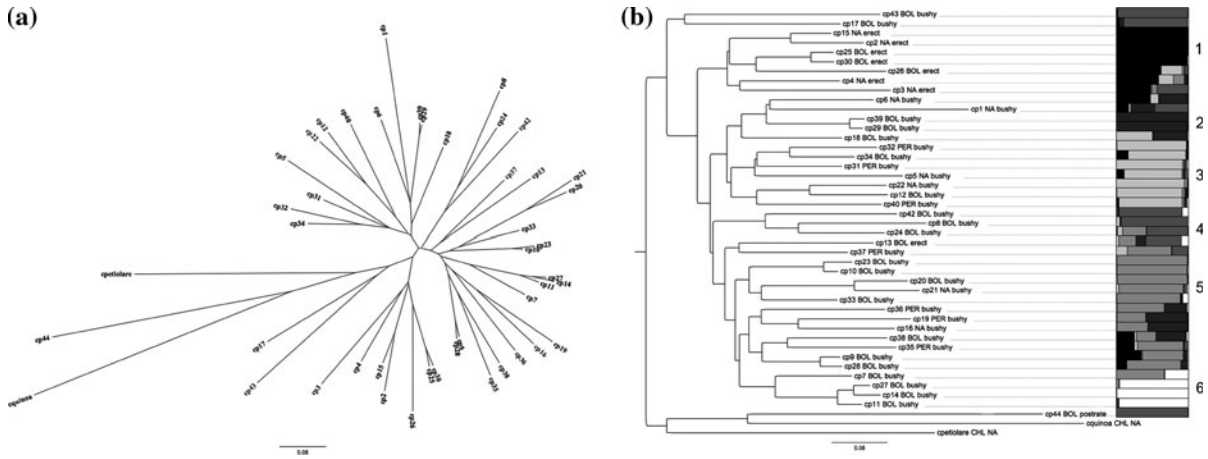

Fig. 3 **a** Unrooted neighbor joining tree showing the genetic relationship among cañahua genotypes based on microsatellite marker data. **b** Rooted neighbor joining dendrogram and population structure based on 43 genotypes and 48 microsatellite markers using the computer program STRUCTURE. Identification codes of genotypes correspond to those given in Table 1. Branch lengths are proportional to genetic distance. Each genotype is represented by a line partitioned in six *gray* colored segments that represent the individual's estimated membership fractions (Q) to each one of the six clusters. Outgroup species are *C. petiolare* and *C. quinoa* (cv. Ollague)

genotypes that were clearly identified as belonging to a single model-based cluster, 17 genotypes were identified as having admixed ancestry (Fig. 3b). The genetic differentiation among the clusters was analyzed by AMOVA and significant ($P < 0.001$) genetic variance was detected within and among clusters. The variance within clusters accounted for the largest portion (73.8%) of the total variance, whereas 26.2% of the variance was conserved among clusters.

An isolation by distance test, used to determine the relative influences of geographic distance (origin of collection) on genetic similarity among genotypes, was not significant ($R^2 = 4.07E - 5$; $P > 0.05$). Correlation analysis of the genetic and phenotypic data for all morphological traits revealed a significant positive, but low correlation ($R^2 = 0.105$; $P = 0.0217$). Similarly, the determination coefficient of individual traits with genetic distance was also low (growth habit, $R^2 = 0.105$; plant color, $R^2 = 0.00814$; seed coat color, $R^2 = 0.0272$; and pericarp color, $R^2 = 0.0271$) and suggests that individually these morphological traits are not predictive of genetic diversity, perhaps due to the limited variation within the traits. Lastly, the average fixation index, which refers to the degree of inbreeding in the population, was high ($F = 0.97$), and was reflected as a low level of predicted outcrossing ($t_a = 0.018$).

Discussion

The transferability of microsatellite loci across plant genera and species is based on the successful amplification of simple and polymorphic products in the target species. In this study we showed the potential transferability of microsatellite loci originally identified in *C. quinoa* to its sister taxon cañahua. Forty percent of all *C. quinoa* microsatellite markers successfully amplified a single band of similar length in cañahua—a result that confirms the close ancestry of the two species. While the number of polymorphic cross-species microsatellites was low (12), these markers will undoubtedly be useful for establishing syntenic relationships between the *C. quinoa* and cañahua genomes and suggest the possibility to exploit genetic information across the two species. Similar levels of cross-species transferability have been reported among *Vitis* species (Di Gaspero et al. 2000), *Phaseolus* species (Gaitan-Solis et al. 2002) and *Amaranthus* species (Mallory et al. 2008).

The most common microsatellite motifs in the cañahua-specific microsatellites were AAT and AT. High frequencies of AT-rich microsatellites have also been observed in many other species of the Chenopodioideae, including *C. quinoa* and *Beta vulgaris* L. (Jarvis et al. 2008; Morchen et al. 1996). Mason et al. (2005) correlated the degree of polymorphism with the motif length in *C. quinoa* and suggested that the rate of polymorphism increases dramatically when the tandem repeat length is greater than 20 bp. Similar results were found by Mallory et al. (2008) in the related Amaranthoideae subfamily species amaranth (*A. hypochondriacus* L.), suggesting that the future development of microsatellites in cañahua should focus on the identification of markers with repeat lengths of greater than 20 bp in AT-rich motifs in order to ensure high discovery rates of polymorphic markers. Such parameters are invaluable, especially in light of the ease of microsatellite discovery from genomic DNA using next-generation sequencing technology. For example, a single pico-titer plate of 454-pyrosequencing Titanium® technology is relatively inexpensive (~$5,000 USD) and can produce more than 500 Mb of DNA sequence. Given the high density of microsatellite loci in plant genomes and automated microsatellite discovery algorithms (e.g., MISA), a single next-generation sequencing run can identify thousands of putative microsatellite loci. The use of motif type and repeat length can be important parameters for minimizing screening cost while maximizing the polymorphic loci discovery rate.

Genetic relatedness, as measured by NJ analysis, revealed a fairly close relationship between two of the wild cañahua accessions and the cultigens. Unfortunately, essentially no phylogenetic research has been conducted in this genus, thus the question of relationships between cultivated and wild, potential progenitor-genotypes are unknown. Our analysis suggests that at least one of the wild genotypes (cp44) is genetically more closely associated with the outgroup species *C. petiolare* and *C. quinoa* than it is with the cultivated or other wild genotypes, indicating that it may have been misclassified as *C. pallidicaule* and may represent a different (although related) species or an interspecific hybrid.

The cultivated cañahua clusters found in STRUCTURE did not show any geographical differentiation. Indeed, Bolivian and Peruvian genotypes are mixed throughout the clusters (Fig. 3b). The lack of geographic differentiation among the cañahua genotypes is suggestive of extensive mixing of the cañahua genotypes across its endemic cultivation area, presumably as a result of trade or bartering among native peoples. The inconsistency of geographic distance data was corroborated by the isolation by distance test. Interestingly, the marker data does identify several admixed genotypes, suggesting that genotypes, once traded, experienced gene flow—presumably through a low level of outcrossing, estimated here at about 1.8%. This low-level of outcrossing is also reflected in the high degree of homozygosity ($F = 0.97$) identified at the microsatellite loci and is likely a function of cañahua's cleistogamous breeding strategy (Risi and Galwey 1984). From a practical germplasm perspective, inbreeding simplifies the maintenance of long-term germplasm collections, as curators do not have to be concerned about gene flow during the propagation of the germplasm bank. Similarly, inbreeding facilitates the development of pure (fixed) lines that are important to core-breeding collections and the development of commercial varieties.

Anciently, cañahua was an integral part of the Andean farmers reciprocal food exchange/acquisition strategy—a seasonal food trading network practiced between farming communities located at different altitudes (Altiplano, yunga, valley and low lands regions; (Rist 2000). Besides cañahua, crops traditionally incorporated in the food strategy (reciprocity and barter market) included potato (*Solanum andigenum* Juz. et Bukasov., *S. tuberosum* L., *S. juzepczukii* Bukasov.), oca (*Oxalis tuberosa* Molina.), isaño or mashua (*Tropaeolum tuberosum* Ruiz. & Pav.), quinoa (*C. quinoa*), coca (*Erythroxylum coca* Lam.), yuca (*Manihot esculenta* Crantz), dehydrated maize (*Zea mays* L.), legumes, and several medicinal plants (Marti and Pimbert 2007). Trading foods items from one Andean region to another continues to be an important strategy for ensuring food security and nutrition—even the development of motor transportation seems to have had little impact on this practice. Consequently, the effects of these trading systems on cañahua germplasm diversity and distribution should be an important consideration in any future germplasm conservation efforts.

Conclusions

We report the first set of polymorphic molecular markers for cañahua (*C. pallidicaule*), a food crop of regional importance for subsistence farmers across the Andean region of South America. The markers reported are of particular value in ongoing efforts to extend and characterize the cañahua germplasm collections in Bolivia and Peru, including the development of core breeding collections needed for emerging cañahua improvement programs across the region. Initial phylogeny analyses using these markers indicate that geographic origin and morphology are not fully predictive of genetic distance, suggesting that future germplasm collection efforts should account for ancient food exchange strategies and incorporate molecular marker data to assure comprehensive germplasm collections. Current efforts are aimed at developing new molecular markers, including single nucleotide polymorphisms, and diverse segregating populations to facilitate the development of linkage maps for cañahua—an important first step towards the development of marker-assisted breeding programs.

Acknowledgments This research was supported by grants from the McKnight Foundation, as well as the Ezra Taft Benson Agriculture and Food institute and Holmes Family Foundation. We gratefully acknowledge the advice and support provided by Dr. Alejandro Bonifacio, PROINPA Foundation, Bolivia and are grateful to Dinesh Adhikary, Nathan Mahler, Joshua Raney, and Shawna and James Daley for technical support in developing the microsatellite protocols. We are also grateful to David Brenner of the USDA NC-7 (Ames, IA) and Dr. Angel Mujica of the National University of the Altiplano (Puno, Peru) for providing seed for this study.

References

Achigan-Dako EG (2008) Phylogenetic and genetic variation analyses in cucurbit species (*Cucurbitaceae*) from West Africa: definition of conservation strategies. Cuvillier Verlag, Göttingen

Bohonak AJ (2002) IBD (isolation by distance): a program for analyses of isolation by distance. J Hered 93:153–154

Bonifacio A (2003) *Chenopodium* sp.: genetic resources, ethnobotany, and geographic distribution. Food Rev Int 19:1–7

Di Gaspero G, Peterlunger E, Testolin R, Edwards KJ, Cipriani G (2000) Conservation of microsatellite loci within the genus Vitis. Theor Appl Genet 101:301–308

Dray S, Dufour AB (2007) The ade4 package: implementing the duality diagram for ecologists. J Stat Softw 22:1–20

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

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

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform 1:47–50

Flores R (2006) Evaluación preliminar agronómica y morfológica del germoplasma de cañahua (*Chenopodium pallidicaule* Aellen) en la estación experimental Belen. Universidad Mayor de San Andrés, La Paz

Friedt W, Snowdon RJ, Ordon F, Ahlemeyer J (2007) Plant breeding: assessment of genetic diversity in crop plants and its exploitation in breeding. In: K. Esser UL, W. Beyschlag, J. Murata (eds). Springer, Berlin, pp 151–178

Gade D (1970) Ethnobotany of cañahua (*Chenopodium pallidicaule*), rustic seed crop of the altiplano. Econ Bot 24:55–61

Gaitan-Solis E, Duque MC, Edwards KJ, Tohme J (2002) Microsatellite repeats in common bean (*Phaseolus vulgaris*): isolation, characterization, and cross-species amplification in *Phaseolus* ssp. Crop Sci 42:2128–2136

Galwey NW (1989) Exploited plants—Quinoa. Biologist 36:267–274

Groben R, Wricke G (1998) Occurrence of microsatellites in spinach sequences from computer databases and development of polymorphic SSR markers. Plant Breed 117: 271–274

Gumerman G (1997) Food and complex societies. J Archaeol Meth Theor 12:105–139

Höglund J (2009) Evolutionary conservation genetics, 1st edn. Oxford University Press, New York

Huelsenbeck JP, Andolfatto P (2007) Inference of population structure under a Dirichlet process model. Genetics 175:1787–1802

IPGRI, PROINPA, IFAD (2005) Descriptores para cañahua (*Chenopodium pallidicaule* Aellen). Instituto Internacional de Recursos Fitogenéticos, Roma, Italia; Fundación PROINPA, La Paz, Bolivia; International Fund for Agricultural Development, Roma, Italia

Jarvis DE, Kopp OR, Jellen EN, Mallory MA, Pattee J, Bonifacio A, Coleman CE, Stevens MR, Fairbanks DJ, Maughan PJ (2008) Simple sequence repeat marker development and genetic mapping in quinoa (*Chenopodium quinoa* Willd.). J Genet 87:39–51

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129

Mallory MA, Hall RV, McNabb AR, Pratt DB, Jellen EN, Maughan PJ (2008) Development and characterization of microsatellite markers for the grain Amaranths. Crop Sci 48:1098–1106

Marin W (2002) Distanciamiento entre surcos y plantas en dos ecotipos de kañawa (*Chenopodium pallidicaule* Aellen) en el Altiplano Norte. Universidad Mayor de San Andrés, La Paz

Marti N, Pimbert M (2007) Barter markets for the conservation of agro-ecosystem multi-functionality: the case of the chalayplasa in the Peruvian Andes. Int J Agr Sustain 5:51–69

Mason SL, Stevens MR, Jellen EN, Bonifacio A, Fairbanks DJ, Coleman CE, McCarty RR, Rasmussen AG, Maughan PJ (2005) Development and use of microsatellite markers for germplasm characterization in quinoa (*Chenopodium quinoa* Willd.). Crop Sci 45:1618–1630

Morchen M, Cuguen J, Michaelis G, Hanni C, SaumitouLaprade P (1996) Abundance and length polymorphism of microsatellite repeats in *Beta vulgaris* L. Theor Appl Genet 92:326–333

Muller K, Borsch T (2005) Phylogenetics of Amaranthaceae based on matK/trnK sequences data: evidence from parsimony, likelihood, and bayesian analysis. Ann Missouri Bot Gard 92:66–102

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Nei M, Li WH (1979) Mathematical-model for studying genetic-variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A 76:5269–5273

Nei M, Tajima F (1983) Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics 105:207–217

Ott J (1992) Strategies for characterizing highly polymorphic markers in human gene-mapping. Am J Hum Genet 51: 283–290

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Repo-Carrasco R, Espinoza C, Jacobsen SE (2003) Nutritional value and use of the Andean crops quinoa (*Chenopodium quinoa*) and kañiwa (*Chenopodium pallidicaule*). Food Rev Int 19:179–189

Repo-Carrasco-Valencia R, de La Cruz AA, Alvarez JCI, Kallio H (2009) Chemical and functional characterization of kañiwa (*Chenopodium pallidicaule*) grain, extrudate and bran. Plant Foods Hum Nutr 64:94–101

Risi CJ, Galwey NW (1984) The Chenopodium grains of the Andes: Inca crops for modern agriculture. Adv Appl Biol 10:145–216

Rist S (2000) Linking ethics and the market—Campesino economic strategies in the Bolivian Andes. Mt Res Dev 20:310–315

Rodríguez M (2007) Evaluación de las pérdidas de grano y grado de impurezas en cuatro métodos de cosecha de cañahua (*Chenopodium pallidicaule* Aellen) en la comunidad de Quipaquipani, Viacha. Universidad Mayor de San Andrés, La Paz

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386

Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

Todd JJ, Vodkin LO (1996) Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. Plant Cell 8:687–699

Vaz ARD, Borba TCD, Brondani C, Rangel PHN, Camargo GSD, Telles MPD, Filho JAF, Brondani RPV (2009) Genetic analysis of a local population of *Oryza glumaepatula* using SSR markers: implications for management and conservation programs. Genetica 137:221–231

Weir B (1996) Genetic Data Analysis II: Methods for discrete population genetic data. Sinauer Assoc., Inc, Sunderland

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. Evolution 38:1358–1370

Woods Páez A, Eyzaguirre P (2004) Cañahua deserves to come back. Leisa 20:11–13

Wright S (1946) Isolation by distance under diverse systems of mating. Genetics 31:39–59