Genet Resour Crop Evol
DOI 10.1007/s10722-011-9680-y

RESEARCH ARTICLE

Genetic diversity and spatial structure in a new distinct *Theobroma cacao* L. population in Bolivia

Dapeng Zhang · Windson July Martínez · Elizabeth S. Johnson · Eduardo Somarriba · Wilberth Phillips-Mora · Carlos Astorga · Sue Mischke · Lyndel W. Meinhardt

Received: 26 September 2010 / Accepted: 28 February 2011

Abstract Cacao (*Theobroma cacao* L.) is an important economic crop in the Bolivian Amazon. Bolivian farmers both cultivate cacao, and extract fruits from wild stands in the Beni River region and in valleys of the Andes foothills. The germplasm group traditionally used is presently referred to as "Cacao Nacional Boliviano" (CNB). Using DNA fingerprinting technology based on microsatellite markers, we genotyped 164 Bolivian cacao accessions, including both cultivated and wild CNB accessions sampled from the Amazonian regions of La Paz and Beni, and compared their SSR profiles with 78 reference Forastero accessions from Amazonian cacao populations, including germplasm from the Ucayali region of Peru. Results of multivariate ordination and analysis of molecular variance show that CNB cacao has a unique genetic profile that is significantly different from the known cacao germplasm groups in South America. The results also show that cultivated CNB and wild CNB populations in the Beni River share a similar genetic profile, suggesting that the cultivated CNB is of indigenous origin in Bolivia. The level of genetic diversity, measured by allele richness and gene diversity in the Bolivian cacao, is moderately high, but was significantly lower than gene diversity in the other Amazonian cacao populations. Significant spatial genetic structure was detected in the wild CNB population, using analysis of autocorrelation ($rc = 0.232$; $P < 0.001$) and Mantel tests ($Rxy = 0.276$; $P < 0.001$). This finding is also highly valuable to support *in situ* conservation and sustainable use of CNB genetic diversity in Bolivia.

Keywords Alto Beni · Bolivia · *Theobroma cacao* · Spatial genetic structure · Wild population

D. Zhang (✉) · E. S. Johnson · S. Mischke · L. W. Meinhardt
USDA-ARS PSI SPCL, 10300 Baltimore Avenue, Bldg. 001, Rm. 223, Beltsville, MD 20705, USA
e-mail: Dapeng.Zhang@ars.usda.gov

W. J. Martínez · E. Somarriba · W. Phillips-Mora · C. Astorga
Research and Development Division, CATIE, Turrialba, Costa Rica

Present Address:
E. S. Johnson
CABI Caribbean and Latin America, Gordon Street, Curepe, Trinidad, Tobago, West Indies

Introduction

Cacao (*Theobroma cacao* L.) is indigenous to Bolivia and widely distributed in the Bolivian Amazon, including the departments of La Paz, Beni, Pando, Santa Cruz and Cochabamba. Among these regions, Alto (Upper) Beni is the major production site with an annual output of approximately 1,000 tons, representing

Published online: 24 March 2011

approximately 80% of Bolivia's cocoa production and 13% of the household income in this region (July Martínez 2007; Somarriba and Trujillo 2005).

Historical records show that the practice of cacao management in Bolivia was introduced by Franciscan and Jesuit missionaries in the eighteenth century (Cortés 1997, Villegas and Astorga 2005). The germplasm group cultivated by Bolivian natives, called "Criollo" by the farmers, is now known as "Cacao Nacional Boliviano" (CNB). It is characterized by its small pod with the shape of an Amazonian amelonado. The pods usually have a dark green color when young, and are elongate with a short apex, slightly rough with 10 superficial ridges and a thin husk. The seeds are small and the cotyledon color is purple, but there is no pigmentation in the filament and the stamen of the flower (Villegas and Astorga 2005; Bartley 2005; Soria 1966). Although the cultivated CNB in Bolivia is called "Cacao Criollo" by the native people, the morphological characteristics of the crop (small and round shaped pods and small beans with purple color) are obviously different from the genetic group of Central America Criollo (Bartley 2005; Sánchez 1983). The CNB is susceptible to witches' broom disease (*Moniliophthora perniciosa* (Stahel) Aime and Phillips-Mora 2006). Nevertheless, the CNB trees produce fruits earlier than the introduced clones, thus can escape infection by black pod disease (*Phytophthora palmivora* (E. J. Butler) E. J. Butler 1919), which requires cold and wet periods for the inocula to develop (Sánchez 1983; Villegas and Astorga 2005).

Based on the extent of domestication, the Bolivian cacao germplasm was traditionally classified into two main groups (Davies 1986). The first group is wild cacao, which refers to trees that have not been cultivated by man, but reproduce through animal mediated dispersal. The wild cacao is widely distributed along the river banks of Beni and Ichilo. Although they were not cultivated, it is a common practice that the local people harvest pods from these trees (Villegas and Astorga 2005). This material is referred to in this research as "CNB wild" The second group is the so called 'Cacao Criollo', which refers to CNB that has been cultivated for more than 200 years by the Mosetenes Indians (Soria 1965; July Martínez (2002, 2007). However, the distinction between "CNB wild" and "CNB cultivated" is often not clear. In some regions, 'Criollo' and 'wild' have been used interchangeably (Villegas and Astorga 2005; July Martínez 2007). It has also been suggested that "Cacao Criollo" might have originated from "CNB Wild' (Soria 1965), because the cacao trees cultivated in Alto Beni were phenotypically similar to the "CNB wild" that is native to Lower Beni, as well as to some wild cacao in the Brazilian Amazon (Soria 1965, 1966). Villegas and Astorga (2005) characterized the phenotypic variation among 73 CNB cultivated accessions from Alto Beni. They found that this germplasm differed from the five known Forastero and Trinitarios clones used as references in their analysis, and suggested that Alto Beni is perhaps the southwest extreme in terms of the spatial distribution of the Forastero natural populations.

The CNB is well adapted to the local climate conditions and has survived well in unmanaged subspontaneous status despite local diseases (Milz 1990). In the 1960s, cacao cultivation was promoted by the government of Bolivia as the main source of income for farmers in the Bolivian Amazon (Zeballos and Terrazas 1970). The new settlers in the Amazonia regions were provided with hybrid seeds originally brought from Ecuador, and then locally propagated through controlled pollination of international clones introduced from Trinidad and CATIE (Tropical Agricultural Research and Higher Education in Sapecho and Alto Beni). Therefore, the cultivated CNB in farmer fields may have been mixed with the introduced exotic germplasm. In the past 10 years, the 'Central Cooperative Ceibo' has collected more than 60 elite genotypes from some 500 trees, through farmer participatory selections (July Martínez 2007), and preserved these superior genotypes in the cacao germplasm collections. However, little is known about the genetic diversity and population structure of these selections. Knowledge is also lacking regarding relationships between Bolivian cacao and the other known Forastero groups, and whether the germplasm used by the local farmers is indigenous to Bolivia.

In the present study, we report our investigation on molecular characterization of the CNB germplasm groups with different degrees of domestication, including CNB wild, Cacao Criollo (CNB cultivated) that has been maintained in the Bolivian cacao genebank, and elite farmer selections. Specifically, we intend to answer questions including (1) are the cultivated Bolivian cacao introduced or of indigenous origin? What is the genetic relationship between the

Bolivia cacao and the other wild cacao populations in the Amazon? (2) Is there a spatial genetic structure in the wild population of Beni river and what does the diversity distribution imply about their historical dispersal?

The outcome of the proposed research will enable us to describe the genetic architecture of the wild cacao in Bolivia and to understand the processes of cacao dispersal that have given rise to this structure. This knowledge is essential for the sustainable management, production and conservation of cacao germplasm in Bolivia and its neighboring countries.

Materials and methods

Cacao samples

The cacao accessions analyzed in this study were comprised of three types of germplasm, from genebank collections, farmer cultivars, and wild populations. Three collecting expeditions were taken from December 2006 to October 2007 to collect leaf samples and record morphological and agronomic traits. The sampled geographical sites include:

1. Communities in the northern department of La Paz, specifically in the TCO (Tierra Comunitaria de Origen, i.e. Indigenous communal lands), and in the Madidi National Park.
2. The Beni River and some of its tributaries in the departments of La Paz and Beni.
3. Alto Beni and La Paz, where CNB cultivation is thought to be introduced by the Jesuits and Franciscans in the eighteenth century.

The specific sampling sites are presented in Fig. 1. In total, we sampled 164 genotypes of CNB cacao, of which 57 were from wild populations in the Beni River region, and 107 were cultivated CNB genotypes from the genebank of the cooperative in the town of Ceibo, Sapecho, Alto Beni, and farmer selections from the northern department of La Paz. The full list of the accessions is presented in Table 1.

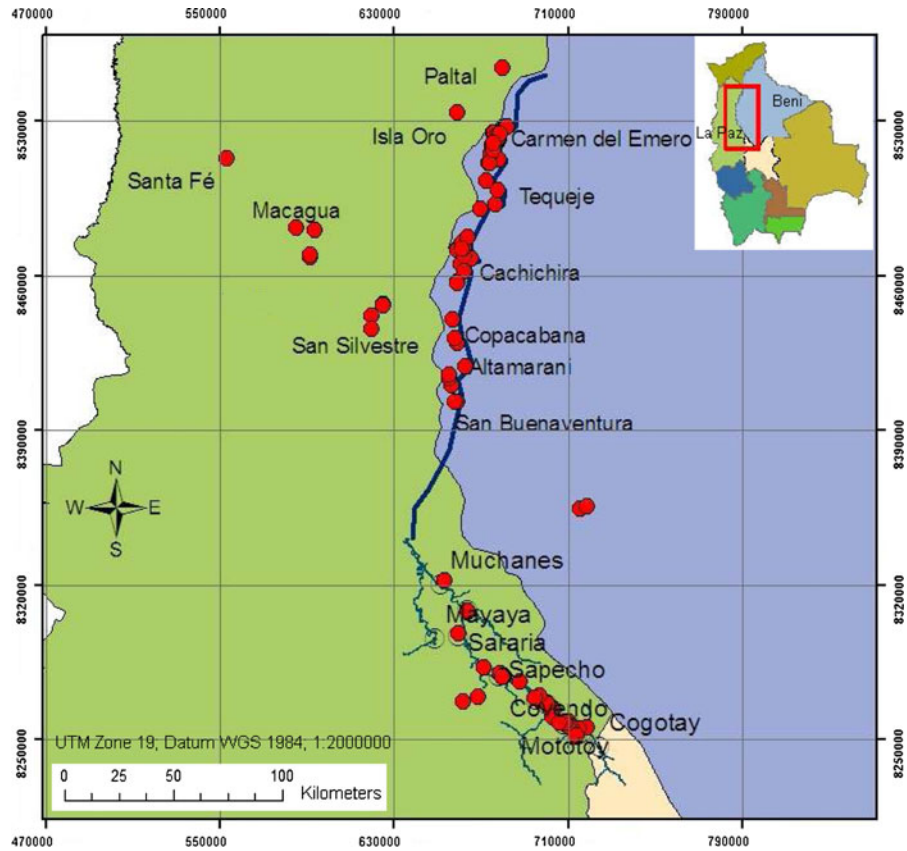

Fig. 1 The geographical sites in the Department of Beni and La Paz, Bolivia, where the cacao samples were taken

Table 1 List of the 164 Bolivian cacao accessions and their collecting locations

Code	Dept	Community	Type	Latitude	Longitude	Altitude (MASL)
TC-1541	Beni	San Silvestre	Wild	8446911	625027	185
TC-1542	Beni	San Silvestre	Wild	8446652	625055	187
TC-1543	Beni	San Silvestre	Wild	8446656	625055	185
TC-1544	La Paz	Tumupasa	Wild	8441852	619781	194
TC-1545	La Paz	Tumupasa	Wild	8435894	619793	190
TC-1546	La Paz	Tumupasa	Wild	8441727	619872	193
TC-1547	La Paz	Tumupasa	Wild	8441693	619752	194
TC-1548	La Paz	Tumupasa	Wild	8441647	619891	194
TC-1549	Beni	Macagua	Wild	8468428	591652	350
TC-1550	Beni	Macagua	Wild	8468482	591612	360
TC-1551	Beni	Macagua	Wild	8468406	591532	300
TC-1552	Beni	Macagua	Wild	8468520	591687	346
TC-1553	Beni	Macagua	Wild	8469170	591614	397
TC-1554	La Paz	Santa Fe	Wild	8481662	584917	215
TC-1555	LaPaz	SantaFé	Wild	8481682	584911	238
TC-1618	Beni	San Marcos	Wild	8471966	661307	178
TC-1619	Beni	San Marcos	Wild	8471879	659300	180
TC-1620	Beni	San Marcos	Wild	8471945	659299	179
TC-1621	Beni	San Marcos	Wild	8471751	659309	180
TC-1623	Beni	San Marcos	Wild	8471890	659208	181
TC-1624	Beni	San Marcos	Wild	8471678	659320	185
TC-1625	Beni	San Marcos	Wild	8471735	659313	175
TC-1626	La Paz	Carmen Emero	Wild	8516708	675539	163
TC-1627	La Paz	Carmen Emero	Wild	8516608	675345	163
TC-1628	La Paz	Carmen Emero	Wild	8516890	675367	160
TC-1629	La Paz	Carmen Emero	Wild	8516670	675548	161
TC-1630	La Paz	Carmen Emero	Wild	8516799	675516	164
TC-1631	La Paz	Carmen Emero	Wild	8516415	675501	159
TC-1632	La Paz	Carmen Emero	Wild	8516567	675524	160
TC-1633	Beni	Copacabana	Wild	8440124	656804	184
TC-1634	Beni	Cacahichira	Wild	8457286	659227	150
TC-1635	Beni	Cacahichira	Wild	8457248	659121	177
TC-1636	Beni	Cacahichira	Wild	8457328	659254	187
TC-1637	Beni	Tequeje	Wild	8511378	673955	181
TC-1638	Beni	Tequeje	Wild	8511332	673929	184
TC-1639	Beni	Tequeje	Wild	8511212	673632	174
TC-1640	Beni	Tequeje	Wild	8511132	673716	180
TC-1641	Beni	Tequeje	Wild	8511132	673719	185
TC-1642	Beni	Tequeje	Wild	8511300	673890	179
TC-1643	Beni	Paltal	Wild	8527290	682131	163
TC-1644	Beni	Paltal	Wild	8527724	681651	148
TC-1645	Beni	Paltal	Wild	8527624	681752	124
TC-1646	Beni	Paltal	Wild	8527626	681732	170
TC-1647	Beni	Paltal	Wild	8527472	681857	148

Table 1 continued

Code	Dept	Community	Type	Latitude	Longitude	Altitude (MASL)
TC-1648	Beni	Paltal	Wild	8527242	681965	156
TC-1649	Beni	Paltal	Wild	8521310	676826	139
TC-1650	Beni	Isla del Oro	Wild	8521452	678038	122
TC-1651	Beni	Isla del Oro	Wild	8520312	677859	161
TC-1652	Beni	Isla del Oro	Wild	8520898	676425	168
TC-1653	Beni	Isla del Oro	Wild	8520580	677032	159
TC-1654	Beni	Isla del Oro	Wild	8522072	678154	146
TC-1655	Beni	Isla del Oro	Wild	8522904	678113	150
TC-1656	Beni	Isla del Oro	Wild	8523678	676243	155
TC-1657	Beni	Isla del Oro	Wild	8516708	675540	163
TC-1658	Beni	Isla del Oro	Wild	8524754	675600	164
TC-1659	Beni	Isla del Oro	Wild	8524546	676345	164
TC-1660	Beni	San Silvestre	Wild	8446648	625112	185
TC-1662	La Paz	Covendo	Wild	8254571	715338	482
TC-1663	La Paz	Covendo	Wild	8354566	715342	482
TC-1664	La Paz	Covendo	Wild	8254543	715341	489
TC-1665	La Paz	Covendo	Wild	8255373	718578	497
TC-1666	La Paz	Covendo	Wild	8255360	718572	528
TC-1667	La Paz	Covendo	Wild	8255374	718598	539
TC-1668	La Paz	Covendo	Wild	8255372	718600	538
TC-1669	La Paz	Covendo	Wild	8355489	718590	544
TC-1670	La Paz	Covendo	Wild	8255484	718596	542
TC-1671	La Paz	Covendo	Wild	8255503	718605	523
TC-1672	La Paz	Covendo	Wild	8255471	718582	529
TC-1673	La Paz	Covendo	Wild	8255478	718587	521
TC-1674	La Paz	Covendo	Wild	8255475	718609	533
TC-1675	La Paz	Covendo	Wild	8255471	718609	543
TC-1676	La Paz	Covendo	Wild	8255466	718611	544
TC-1677	La Paz	Covendo	Wild	8255476	718629	515
TC-1678	La Paz	Covendo	Wild	8254625	715258	376
TC-1661	La Paz	Covendo	Cultivated	8254630	715296	524
TC-1679	La Paz	Remolmo	Cultivated	8263203	703316	491
TC-1680	LaPaz	Cocochi	Cultivated	8258146	710140	483
TC-1681	La Paz	Cocochi	Cultivated	8258127	710090	446
TC-1682	La Paz	Cocochi	Cultivated	8258095	710123	488
TC-1683	LaPaz	Cocochi	Cultivated	8258122	710121	488
TC-1684	La Paz	Mototoy	Cultivated	8256880	708640	510
TC-1685	La Paz	Mototoy	Cultivated	8256951	708612	509
TC-1686	La Paz	Mototoy	Cultivated	8256938	708599	494
TC-1687	LaPaz	SanJosé	Cultivated	8256096	711416	489
TC-1688	LaPaz	SanJosé	Cultivated	8256083	711437	470
TC-1689	LaPaz	SanJosé	Cultivated	8256087	711420	500
TC-1690	La Paz	San José	Cultivated	8256109	711422	477
TC-1691	La Paz	San Antonio	Cultivated	8267901	697093	442

Table 1 continued

Code	Dept	Community	Type	Latitude	Longitude	Altitude (MASL)
TC-1692	La Paz	San Antonio	Cultivated	8267927	697105	422
TC-1693	La Paz	San Antonio	Cultivated	8267466	696663	430
TC-1694	La Paz	San Miguel Huachi	Cultivated	8266354	699723	460
TC-1695	La Paz	San Miguel Huachi	Cultivated	8266690	699846	469
TC-1696	La Paz	San Miguel Huachi	Cultivated	8266578	699967	445
TC-1697	La Paz	San Miguel Huachi	Cultivated	8266396	699974	436
TC-1698	La Paz	San Miguel Huachi	Cultivated	8266424	700025	419
TC-1699	La Paz	San Miguel Huachi	Cultivated	8266485	700035	444
TC-1700	La Paz	San Miguel Huachi	Cultivated	8265880	700045	440
TC-1701	La Paz	San Miguel Huachi	Cultivated	8265879	700052	438
TC-1702	La Paz	San Miguel Huachi	Cultivated	8265916	700063	415
TC-1703	La Paz	San Miguel Huachi	Cultivated	8265958	700058	421
TC-1704	La Paz	Remolmo	Cultivated	8263212	703299	496
TC-1705	La Paz	Remolmo	Cultivated	8263220	703303	489
TC-1556	La Paz	Sapecho	Cultivated	8278592	680563	414
TC-1557	La Paz	Sapecho	Cultivated	8278591	680565	413
TC-1558	LaPaz	Sapecho	Cultivated	8278527	680518	421
TC-1559	LaPaz	Sapecho	Cultivated	8278570	680496	391
TC-1560	La Paz	Sapecho	Cultivated	8278546	680517	431
TC-1561	LaPaz	Sapecho	Cultivated	8278611	680514	411
TC-1562	LaPaz	Sapecho	Cultivated	8278588	680508	384
TC-1563	LaPaz	Sapecho	Cultivated	8278587	680507	398
TC-1564	La Paz	Sapecho	Cultivated	8278585	680516	408
TC-1565	LaPaz	Sapecho	Cultivated	8278573	680513	412
TC-1566	LaPaz	Sapecho	Cultivated	8278565	680515	413
TC-1567	LaPaz	Sapecho	Cultivated	8278570	680511	413
TC-1568	La Paz	Sapecho	Cultivated	8278546	680520	411
TC-1569	LaPaz	Sapecho	Cultivated	8278568	680518	412
TC-1570	LaPaz	Sapecho	Cultivated	8278558	680524	415
TC-1571	LaPaz	Sapecho	Cultivated	8278559	680525	414
TC-1572	La Paz	Sapecho	Cultivated	8278563	680517	422
TC-1573	LaPaz	Sapecho	Cultivated	8278588	680516	417
TC-1574	LaPaz	Sapecho	Cultivated	8278590	680518	414
TC-1575	LaPaz	Sapecho	Cultivated	8278576	680526	419
TC-1576	La Paz	Sapecho	Cultivated	8278572	680510	420
TC-1577	LaPaz	Sapecho	Cultivated	8278561	680532	412
TC-1578	LaPaz	Sapecho	Cultivated	8278561	680534	413
TC-1579	LaPaz	Sapecho	Cultivated	8278559	680517	414
TC-1580	La Paz	Sapecho	Cultivated	8278557	680526	407
TC-1581	LaPaz	Sapecho	Cultivated	8278568	680533	409
TC-1582	LaPaz	Sapecho	Cultivated	8278560	680534	413
TC-1583	La Paz	Sapecho	Cultivated	8278547	680531	429
TC-1584	La Paz	Sapecho	Cultivated	8278557	680529	406
TC-1585	La Paz	Sapecho	Cultivated	8278565	680532	406

Table 1 continued

Code	Dept	Community	Type	Latitude	Longitude	Altitude (MASL)
TC-1586	La Paz	Sapecho	Cultivated	8278541	680520	406
TC-1587	LaPaz	Sapecho	Cultivated	8278578	680514	406
TC-1588	LaPaz	Sapecho	Cultivated	8278581	680535	409
TC-1589	LaPaz	Sapecho	Cultivated	8278570	680516	419
TC-1590	La Paz	Sapecho	Cultivated	8278593	680527	413
TC-1591	LaPaz	Sapecho	Cultivated	8278566	680537	406
TC-1592	LaPaz	Sapecho	Cultivated	8278590	680536	417
TC-1593	LaPaz	Sapecho	Cultivated	8278561	680536	414
TC-1594	LaPaz	Sapecho	Cultivated	8278561	680537	412
TC-1595	La Paz	Sapecho	Cultivated	8278566	680540	401
TC-1596	LaPaz	Sapecho	Cultivated	8278553	680531	425
TC-1597	La Paz	Sapecho	Cultivated	8278567	680534	407
TC-1598	LaPaz	Sapecho	Cultivated	8278570	699863	410
TC-1599	La Paz	Sapecho	Cultivated	8278574	680537	421
TC-1600	La Paz	Sapecho	Cultivated	8278574	680535	419
TC-1601	LaPaz	Sapecho	Cultivated	8278595	680533	416
TC-1602	LaPaz	Sapecho	Cultivated	8278596	680536	419
TC-1603	La Paz	Sapecho	Cultivated	8278593	680544	416
TC-1604	La Paz	Sapecho	Cultivated	8278600	680543	418
TC-1605	La Paz	Sapecho	Cultivated	8278598	680544	423
TC-1606	LaPaz	Sapecho	Cultivated	8278585	680530	418
TC-1607	LaPaz	Sapecho	Cultivated	8278595	680533	417
TC-1608	La Paz	Sapecho	Cultivated	8278590	680544	413
TC-1609	LaPaz	Sapecho	Cultivated	8278571	680551	417
TC-1610	LaPaz	Sapecho	Cultivated	8278573	680549	421
TC-1611	LaPaz	Sapecho	Cultivated	8278572	680539	414
TC-1612	LaPaz	Sapecho	Cultivated	8278567	680549	419
TC-1613	LaPaz	Sapecho	Cultivated	8278570	680546	428
TC-1614	LaPaz	Sapecho	Cultivated	8278564	680548	412
TC-1615	LaPaz	Sapecho	Cultivated	8278588	680280	400
TC-1616	LaPaz	Sapecho	Cultivated	8278559	680638	417
TC-1617	La Paz	Sapecho	Cultivated	8278551	680567	420

DNA extraction, PCR amplification, and capillary electrophoresis

DNA extraction was as previously described (Johnson et al. 2007, 2009). Amplification of microsatellite loci were achieved using 15 primers with sequences previously described (Lanaud et al. 1999; Risterucci et al. 2000; Saunders et al. 2004). These 15 loci have been agreed upon, by multiple international and government-sponsored laboratories in the cacao research community, as standardized SSR primers to characterize all *T. cacao* germplasm collections (Saunders et al. 2004). These standard loci have been used for cacao genotyping in several germplasm collections (Zhang et al. 2006a, 2009). Primers were synthesized by Proligo (Boulder, CO) and forward primers were 5′-labeled using WellRED fluorescent dyes (Beckman Coulter, Inc., Fullerton, CA). PCR was performed as described in Saunders et al. (2004), using commercial hot-start PCR SuperMix that had been fortified with an additional 30 U/ml of hot-start *Taq* DNA polymerase (Invitrogen Platinum *Taq*,

Carlsbad, CA; or Eppendorf HotMaster *Taq*, Brinkman, Westbury, NY).

The amplified microsatellite loci were separated by capillary electrophoresis as previously described (Saunders et al. 2004; Zhang et al. 2006a). Data analysis was performed using the CEQ^{TM} 8000 Fragment Analysis software version 7.0.55 according to manufacturers' recommendations (Beckman Coulter, Inc.). SSR fragment sizes were automatically calculated to two decimal places by the CEQ^{TM} 8000 Genetic Analysis System. Allele calling was performed using the CEQ^{TM} 8000 binning wizard software (CEQ^{TM} 8000 software version 7.0.55, Beckman Coulter, Inc.) and edited, based on the bin list, using a SAS program (SAS 1999).

Data analysis

Summary statistics for each marker locus, including allele richness, observed heterozygosity (H_O), expected heterozygosity (H_E) and inbreeding coefficient (F_{IS}) were separately computed for CNB cultivated and CNB wild CNB, as well as for the combined data of all the studied Bolivian germplasm using FSTAT (Goudet 1995). To assess the level of genetic diversity in the wild CNB population relative to other known Forastero germplasm, we compared the CNB with a previously reported population from the Ucayali River in the Peruvian Amazon (Zhang et al. 2006b). The multilocus SSR data of 43 accessions from the Ucayali population were compared with the wild Beni population. Allele richness, expected and observed heterozygosity, and inbreeding coefficient were computed using the program FSTAT (Goudet 1995). In addition, analysis of molecular variance (AMOVA), implemented in the program GenAlEx 6 (Peakall and Smouse 2006), was used to compute the within and between population variation.

To understand the relationships among the individual CNB genotypes and the reference Forastero clones, we also computed the genetic distances among the CNB individuals together with 35 reference international clones from various known cacao genetic groups. Principal Coordinates Analysis (PCoA), implemented in GenAlEx 6 (Peakall and Smouse 2006), was used for computation. Pair-wise Euclidian distances were computed for every pair of the accessions and the distance matrix is presented in a three-axes plot.

Table 2 Number of alleles (Na), observed and expected heterozygosity (H_O and H_E), inbreeding coefficient, and probability of Chi-Square tests for Hardy–Weinberg Equilibrium in 164 Bolivian cacao accessions

Locus	Na	Ho	He	F_{IS}	χ^2 test HWE
Y16981	6.0	0.53	0.67	0.20**	64.9***
Y16980	6.0	0.08	0.11	0.27**	87.3***
Y16995	5.0	0.39	0.60	0.36**	187.1***
Y16996	7.0	0.58	0.71	0.18*	108.1***
Y16982	7.0	0.23	0.48	0.53***	218.1***
Y16883	4.0	0.57	0.66	0.14^{NS}	33.9***
Y16985	7.0	0.40	0.53	0.25**	94.2***
Y16986	9.0	0.65	0.78	0.17*	63.9**
Y16988	7.0	0.38	0.56	0.32**	191.1***
AJ271942	10.0	0.44	0.66	0.34**	416.6***
AJ271826	6.0	0.29	0.43	0.31**	179.4***
Y16991	4.0	0.29	0.56	0.49***	58.7***
Y16998	5.0	0.18	0.23	0.19*	33.5***
AJ271943	21.0	0.26	0.83	0.69***	1,702.1***
AJ271958	6.0	0.37	0.61	0.40***	144.1***
Mean	7.3	0.38	0.56	0.32**	

NS Not significant

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ respectively

To quantify the extent of population differentiation between the wild CNB and the nearby Ucayali population from Peru, F-statistics were estimated by measuring population variation in allelic distributions between the two populations. Allelic composition was assessed using a contingency table test (Weir and Cockerham 1984), implemented in GENEPOP version 3 (Raymond and Rousset 1995), with a null hypothesis of identical allelic distributions in both CNB wild and Ucayali populations. Permutation of the individual genotypes between populations was carried out with the probability of non-differentiation being estimated over 10,000 randomizations.

Nonrandom geographical distribution of genotypes in a population is known as spatial genetic structure (subsequently abbreviated SGS). To assess SGS in the wild Beni population, we used a pairwise comparison of genetic similarity of individuals with respect to spatial distance separating those individuals within populations, as implemented in GenAlEx 6 (Peakall and Smouse 2006). The significance of the

Table 3 Intrapopulation genetic diversity in cultivated and wild CNB cacao populations compared with a neighboring population from Ucayali River, Peru

	Bolivian (Cultivated)		Bolivian (Wild)		Ucayali (Wild)	
Locus	Allele richness	Gene diversity	Allele richness	Gene diversity	Allele richness	Gene diversity
Y16981	5.75	0.651	5.69	0.684	3.24	0.464
Y16980	2.86	0.097	4.47	0.120	7.87	0.738
Y16995	4.66	0.637	4.00	0.522	8.15	0.815
Y16996	5.60	0.656	5.93	0.742	8.80	0.795
Y16982	6.05	0.474	4.00	0.492	12.40	0.905
Y16883	4.00	0.674	4.00	0.622	10.50	0.843
Y16985	6.15	0.491	5.00	0.595	9.77	0.752
Y16986	6.96	0.765	6.00	0.784	9.64	0.805
Y16988	5.29	0.525	6.93	0.609	7.32	0.587
AJ271942	6.88	0.589	9.29	0.758	9.07	0.793
AJ271826	4.10	0.431	3.72	0.422	10.05	0.780
Y16991	3.41	0.549	3.00	0.552	6.10	0.460
Y16998	4.73	0.204	3.99	0.271	9.18	0.746
AJ271943	13.24	0.813	6.61	0.843	11.16	0.881
AJ271958	5.51	0.634	4.99	0.533	10.65	0.742
Mean	5.68	0.550	5.84	0.570	8.93	0.74

Allelic richness is an adjusted estimated based on the different sample sizes using the program FSTAT (Goudet 1995)

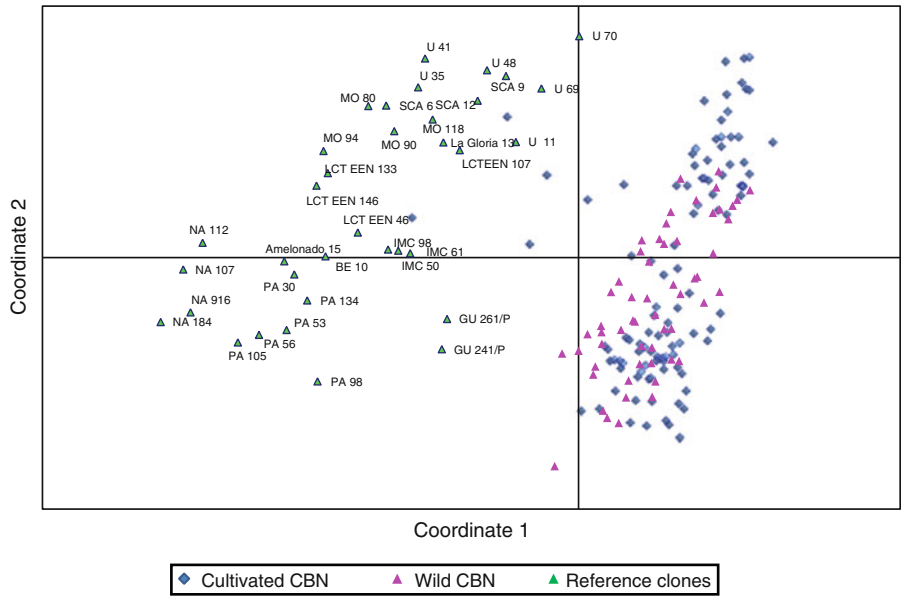

Fig. 2 PCoA plot of 199 cacao accessions, including 57 wild CNB genotypes from the Beni River, 107 cultivated CNB genotypes from La Paz and Beni, and 35 reference clones from the International Genebank, Trinidad. First axis = 44.6% of total information, the second = 18.6% and the third = 11.2

autocorrelation coefficient (r) was tested by constructing a classic 2-tailed 95% confidence interval around the null hypothesis of no SGS (i.e., r = 0) and by performing 999 random permutations of genotypes among geographic locations (Peakall and Smouse 2006). In addition, the Mantel test was

Table 4 Interpopulation variation among wild and cultivated populations from Beni and Ucayali River

	Beni wild population (n = 57)	
Ucayali wild population (n = 43)	F_{ST}^{a}	F_{ST}^{b}
	0.202	0.210
	$P < 0.001$[c]	$P < 0.001$

[a] Definition of F_{ST} follows Weir and Cockerham (1984)

[b] Definition of F_{ST} follows Schneider et al. (2000). Number of permutations = 10,000

[c] Tests of differentiation are performed not assuming H–W equilibrium with each sample. The pairwise significance was presented after standard Bonferroni corrections

Table 5 Analysis of molecular variance (AMOVA) for SSR variation between and within the CNB and Ucayali wild populations

Source	Df	SSD[a]	MSD[b]	Variance component	% Total[c]	*P* value[d]
Among Pops	1	105.6	105.6	1.98	19.0	>0.0001
Within Pops	98	825.9	8.43	8.43	81.0	
Beni	56	423.8	7.57			
Ucayali	42	402.1	9.57			
Total	99	931.5	114.0	10.41		

[a] Sum of squared deviations

[b] Mean squared deviations

[c] Percent of total molecular variance

[d] Probability of obtaining a larger component estimate. Number of permutations = 10,000

performed between the matrix of genetic distances and the linear pairwise geographical distances using the Mantel procedure in the same program.

Results

Genetic diversity in the cultivated and wild Bolivian cacao germplasm

The summary statistics of the 15 microsatellite loci used in the study are listed in Table 2. A total of 110 different alleles were detected across the 164 Bolivian cacao samples, with each cacao tree characterized by a unique multilocus genotype. The mean expected heterozygosity (H_E) over all loci was 0.56 in the studied CNB germplasm and the observed heterozygosity (H_O) is 0.38. The average inbreeding coefficient (F_{IS}) is 0.32 over the 15 loci, which is significantly greater from zero by a permutation test (Goudet 1995). Tests for departure from Hardy–Weinberg Equilibrium (HWE) revealed that all 15 loci have highly significant diversion from HWE (Table 2).

Table 3 presents the comparative diversity measurements among the cultivated and the wild CNB germplasm as well as in the reference population from Ucayali, Peru. The cultivated and wild CNB have similar levels of genetic diversity as measured by allele richness (5.68 vs. 5.84) and gene diversity (0.550 vs. 0.570). These measures are significantly lower than those in the Ucayali population (allele richness = 8.93; Gene diversity = 0.740).

The relationship among the CNB germplasm accessions and the reference international clones is presented by the plot of Principal Coordinates Analysis (Fig. 2). The plane of the first three main PCoA axes, which accounted for 74.4% of total variation, showed that the cultivated CNB genotypes overlapped with the wild CNB population, suggesting their common origin and genetic background. Most of the CNB germplasm is well separated from the reference international clones representing different

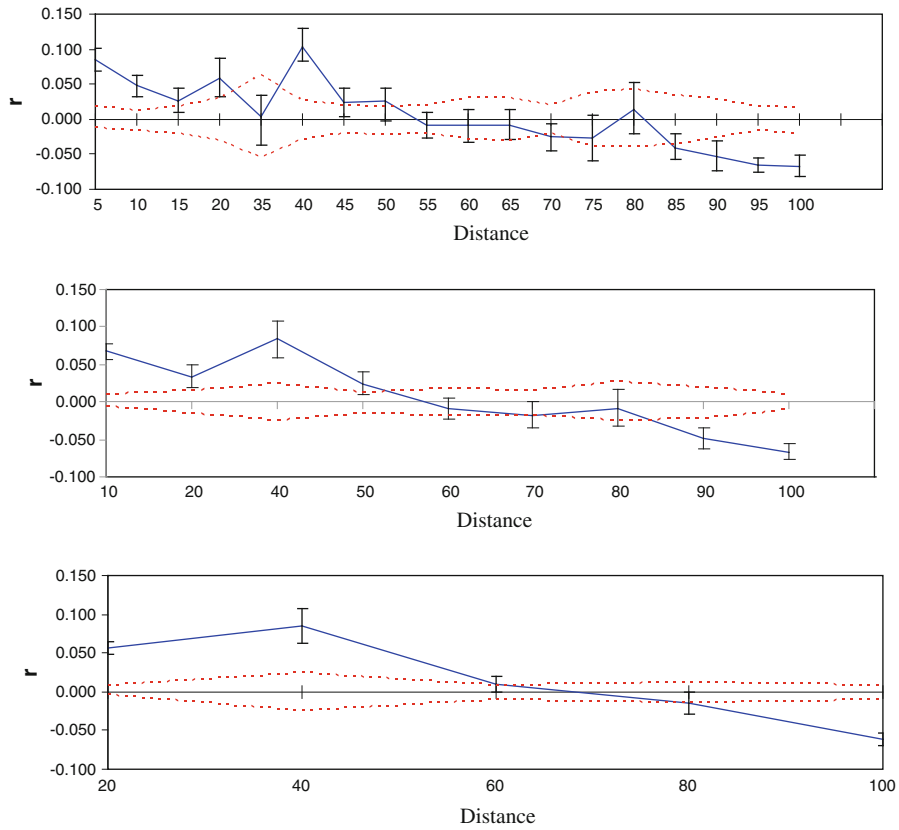

Fig. 3 Spatial autocorrelograms for wild cacao population from the Beni River, Bolivia, showing the global spatial correlation (rc) as a function of geographical distance, 95% confidence interval about the null hypothesis of a random distribution of the wild cacao as determined by bootstrapping. **a** Autocorrelation for distance class sizes of 5 km. **b** Autocorrelation for distance class sizes of 10 km and **c** Autocorrelation for distance class sizes of 20 km

groups of Lower and Upper Amazon Forasteros. Six cultivated CNB genotypes fell within or close to the Upper Amazon Forastero, which is likely due to the introduction of germplasm originated from Peru and Ecuador. The overall pattern of distribution suggests that there was a large genetic differentiation between the CNB germplasm and the rest of the Amazonian Forastero groups.

Table 4 summarizes the F-statistics used to test genetic differentiation between the CNB and the Ucayali populations. The significant population differentiation (Fst = 0.202; $P < 0.001$) between the two Forastero populations was further supported by the result of a hierarchical AMOVA (Table 5), where both the within-group and the between-group variations were found to be highly significant, with 19.0% of the total molecular variance contributed by the differentiation between populations, whereas 81.0% was partitioned into within-populations.

Spatial genetic structure (SGS) in the wild population

The result of combined spatial genetic autocorrelation analysis across the 15 loci is shown in Fig. 3 for distance class sizes of 5, 10 and 20 km. The correlograms show the genetic correlation as a function of distance between genotypes. In all three distance sizes, the values of the correlation coefficients (rc) are positive and significant. There was a little change of rc value as distance class size changes (Fig. 3). The results of Mantel tests further support the result of global spatial autocorrelation, showing a significant positive relationship between geographic and genetic distance (Rxy = 0.276; $P < 0.001$). Therefore, both types of analyses support the conclusion that the distribution of wild cacao genotypes is not random at the sampled geographical scale of the present study. Proximate genotypes tend to be

more genetically similar than those distant ones, which fits the isolation by distance pattern of spatial structure in the natural population of many tropical tree species.

Discussion

Cacao has been long known to exist along the Alto Beni river, especially in the valleys of the transitional region of the foothills of the Andes known as the 'Yungas' (Soria 1965, 1966; Bartley 2005; July Martínez 2002). However, the understanding of the cacao gene pool in the Americas has long been impeded by lack of knowledge regarding the genetic identity and population structure of this germplasm, now referred to as "Cacao Nacional Boliviano" (CNB). It is believed that missionaries introduced the management practices for farming the CNB germplasm sometime between 1,739 and 1,809 (Cortés 1997), then assisted the native Mosetene Indians in the cultivation during the next 150–200 years. Cacao farming remained largely unchanged until the beginning of colonization (1960–1980) when foreign hybrid seeds were introduced from Ecuador, Peru, Trinidad and Tobago (University of West Indies), and CATIE, Costa Rica. However, it is not known if the CNB germplasm is native to the region or was introduced after the start of commercial cacao cultivation in the Americas. Neither do we know the scope of genetic diversity in the Bolivian cacao germplasm. Based on the morphological characteristics, Soria (1965) suggested that the Bolivian cacao cultivated in Alto Beni might be of indigenous origin as the fruits of the local variety were similar to the Amazonian Amelonado type, with very small fruits of a dark green color when young, that yellowed when ripened. Using 30 qualitative characteristics of cacao flowers, pods, seeds and leaves, Villegas and Astorga (2005) characterized 73 genotypes of Cacao Nacional Boliviano (CNB) sampled from nine farms in four localities of Alto Beni, Bolivia. They reported that cultivated CNB from Alto Beni shared principal characteristics with the Forastero group from Rio Beni in the Amazon watershed, and suggested that Alto Beni may be the most extreme Southwest point of the natural distribution of the Forastero group. They also reported that these 73 CNB could be classified into two subgroups, suggesting that that the farmer cultivars were derived from a limited number of wild trees which may have resulted in reduced diversity.

In the present study, we fingerprinted 164 Bolivian cacao accessions, including both cultivated and wild CNB germplasm, using 15 microsatellite markers. The PCoA plot shows that the cultivated and wild CNB germplasm overlapped, demonstrating their common genetic background. The result supports the observations of Soria (1965) and Villegas and Astorga (2005) that the cultivated and wild CNB belong to the same genetic group. Our result demonstrates that the cultivated CNB was domesticated from the wild populations of Beni River.

The allele richness (5.68–5.84 per locus) and gene diversity (0.55–0.57) in the CNB germplasm are roughly 60 and 75% of those found in the Ucayali River germplasm (Table 3). This lower level of genetic diversity, plus the significant inbreeding ($F_{IS} = 0.32$; $P < 0.01$), suggested southward genetic drift from the center of diversity, which substantiates the hypothesis that CNB is a group of Forastero on the southwest extreme of the cacao gene pool distribution in the Americas (Soria 1965; Villegas and Astorga 2005, Bartley 2005). However, the result also shows that the level of allele richness and gene diversity in the cultivated CNB from Alto Beni are comparable to the wild CNB population from Beni River (Table 3), indicating there is no significant reduction of genetic diversity after domestication. The cultivated CNB might have been derived from a relatively large number of wild trees. In fact, it is still a common practice for the local farmers to harvest cacao from the semi-spontaneous cacao trees along the Beni river, which indicates a tradition of multi-site domestication.

The cultivated cacao has been traditionally subdivided into three main groups, including Criollo, Forastero and Trinitario (Cheesman 1944; Wood and Lass, 1985). Among the three main groups, Criollo cacao was domesticated more than 3,000 years ago in Mesoamerica (Henderson et al. 2007). It was believed that Criollo was the only cacao variety cultivated in Mesoamerica before the arrival of the Europeans (Bartley 2005; Motomayor et al. 2003). The Forastero cacao encompasses a diverse range of populations from South America (Bartley 2005). Forastero cacao was not used in production until the mid of eighteenth century. They were brought to the traditional cocoa producing regions (including Central America and the Caribbean) when the cacao plantations were

devastated by the unknown diseases. Trinitario is believed to be the natural hybrid of Criollo and Forastero which started in Trinidad after an introduction of Forastero materials in the mid of eighteenth century (Cheesman 1944; Bartley 2005). Motamayor et al. (2008) analyzed the population structure among the cacao accessions maintained in various *ex situ* collections in South America. They reported that the existing cacao germplasm could be classified into ten distinctive populations based on the results of Bayesian cluster analysis. In a more recent presentation in the International Cocoa Producer's Conference held in Indonesia, Motamayor et al. added Beni as one of the three new clusters (the other two were Huallaga and Ucayali clusters), based the SSR data of twenty Bolivian accessions. The result of the present study further illuminated the unique genetic profile of the CNB germplasm. As shown by the PCoA plot, the CNB germplasm, both cultivated and wild, was completely separated from all the 35 reference clones representing the known Forastero genotypes. Although the pods of CNB appeared similar to the Lower Amazon Amelonado (Bartley 2005; July Martínez 2007), the PCoA plot showed that these two are very different genetic groups. The result of inter-population differentiation (Fst) and AMOVA further demonstrated that the CNB germplasm is genetically different from the other reported germplasm groups in South America. These results support the conclusion that the CNB germplasm is a new group of Forastero cacao that has not been genetically described so far. The diversity level and distribution we describe is likely a combined result of natural forces and human intervention. However, since the wild CNB genotypes used in this study were sampled from one segment along the Beni River, the diversity we reported here probably only represents a fraction of the overall diversity in wild CNB in Bolivia.

Another noteworthy observation is the significant spatial genetic structure (SGS) in the wild CNB population. Although collection of wild cacao for use in breeding started 70 years ago, little information is available regarding the spatial genetic structure of natural cacao populations. Although it is hypothesized that gene flow in cacao is limited and mating is likely confined within patches due to the short-distance seed and pollen dispersal (Chapman and Soria 1983; Dias 2001), gene flow in wild populations has not been investigated. Isolation by distance was reported in the cacao population from the Ucayali river in Peru, but spatial autocorrelation was detected only in a long geographical range (700 km), and not in a local basin or over short distances (Zhang et al. 2006b). The present study provides the first examples showing that SGS actually exists in a cacao population, which stretches over approximately 100 km (Fig. 3) along the Beni River. The autocorrelation analyses we used explore the genetic correlation at multiple distance classes, and significant autocorrelation was detected at each of the classes. Moreover, the highly significant result of Mantel's test also supports our finding of a strong SGS signal across the entire data set. The results are consistent with the short-distance gene flow presumed to exist in natural cacao populations. Our study appears to be the first microsatellite-based study to report consistent positive spatial genetic structure in a wild cacao population. The result is useful for future studies on the relationship between gene flow and local spatial genetic structure, which is essential to understand cacao dispersal in response to landscape change and habitat fragmentation in the Amazon.

The information generated in this study is useful to clarify CNB germplasm, to exploit the genetic resources of local cacao, and to guide the selection of new planting materials to be used in cacao plantations in Bolivia. In the context of international *ex situ* conservation, the Bolivian cacao is invaluable to help fill the diversity gap among the several International cacao collections, and it provides potential new genetic variations that may be useful for cacao improvement programs. It is certainly worthwhile to undertake further exploration of this region to collect and further characterize this new population of cacao.

Acknowledgments We thank Adriana Arciniegas Leal and Stephen Pinney for their assistance in the SSR genotyping. We also thank the Bolivian farmers that participated in the identification of trees and in on-farm interview. Special thanks are due to Dr. Ainong Shi for his critical review of the manuscript.

References

Bartley BGD (2005) The genetic diversity of cacao and its utilization. CABI Publishing, Wallingford

Chapman RK, Soria SJ (1983) Comparative Forcipomyia (Diptera, Ceratopogonidae) pollination of cacao in Central

America and Southern Mexico. Rev Theobroma 13: 129–139

Cheesman EE (1944) Notes on the nomenclature, classification and possible relationships of cocoa populations. Trop Agricult 21:144–159

Cortés PAL (1997) Algunos ejemplos del control gubernamental sobre los jesuitas tras la expulsión. In: Mestre Sanchís A, Fernández Albaladejo P, Gimenes Lopez E,(coord) Disidencias y exilios en la España Moderna. Alicante V 2:691–701

Davies P (1986) Sistemas alternativos de producción para cacao en la zona norte de colonización: un análisis económico exploratorio. CIAT (Centro de Investigación Agrícola Tropical) documento Trabajo No 55. Santa Cruz, Bolivia

Dias LAS (2001) Origin and distribution of *Theobroma cacao* L: A new scenario. In: Dias LAS (ed) Genetic improvement of cacao. Translation available at: http://ecoport.org/ep?SearchType=earticleView&earticleId=197&page=-2 (January 2011)

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. J Hered 86:485–486 Availability: http://www2.unil.ch/popgen/softwares/fstat.htm (January 2011)

Henderson JS, Joyce RA, Hall GR, Hurst WJ (2007) Chemical and archaeological evidence for the earliest cacao beverages. Proc Natl Acad Sci USA 104:18937–18940

Johnson ES, Mora A, Schnell RJ (2007) Field Guide efficacy in the identification of reallocated clonally propagated accessions of cacao. Gen Res Crop Evol 54:1301–1313

Johnson ES, Bekele F, Brown JS, Song JH, Motamayor J, Zhang D, Schnell RJ II, Meinhardt LW (2009) Population structure and genetic diversity of the Trinitario Cacao (*Theobroma Cacao* L.) from Trinidad and Tobago. Crop Sci 49:564–572

July Martínez W (2002) Comportamiento de estacas de tres variedades de cacao (*Theobroma cacao* L.) a tres tratamientos de fitohormonas en la región de Alto Beni. Tesis Ingeniero Agrónomo UMSA, La Paz Bolivia

July Martínez W (2007) Caracterización morfológica y molecular del Cacao Nacional Boliviano y de selecciones élites del Alto Beni, Bolivia. M Sc thesis CATIE, Turrialba, Costa Rica

Lanaud C, Risterucci AM, Pieretti I, Falque M, Bouet A, Lagoda PJL (1999) Isolation and characterization of microsatellites in Theobroma cacao L. Mol Ecol 8:2141–2143

Milz J (1990) Cacao amazónico: posibilidades y perspectivas de producción y comercialización. La Paz, Bolivia, Servicio de Cooperación Técnica y Social Alemana: 61–65

Motamayor JC, Risterucci AM, Heath M, Lanaud C (2003) Cacao domestication II: progenitor germplasm of the Trinitario cacao cultivar. Heredity 91:322–330

Motamayor JC, Lachenaud P, da SilvaeMota JW, Loor R, Kuhn DN, Brown JS, Schnell RJ (2008) Geographic and genetic population differentiation of the Amazonian chocolate tree (Theobroma cacao L). PLoS ONE 3(10): e3311. doi:10.1371/journal/pone.0003311

Motamayor JC, Lachenaud P, da SilvaeMota JW, Loor R, July Martinez W, Graham J, Kuhn DN, Brown JS, Schnell RJ (In press) El nombre 'Forastero' no más: A new protocol for meaningful cacao germplasm classification. In: Proceedings of the International Cocoa Producer's Conference, Nov. 16–21, 2009. Bali, Indonesia

Peakall R, Smouse PE (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Raymond M, Rousset F (1995) GENEPOP (version 1.3): population genetic software for exact tests and ecumenicism. J Heredity 86:248–249, Availability: http://genepop.curtin.edu.au/ (January 2011)

Risterucci AM, Grivet L, N'Goran JAK, Pierevtti I, Flament MH, Lanaud C (2000) A high density linkage map of *Theobroma cacao* L. Theor Appl Genet 101:948–955

Sánchez A (1983) Cultivos de plantación. Manuales para la producción agropecuaria, área producción vegetal. Editorial Trillas, SA de CV (Mexico City), Mexico DF Segunda Reimpresión MX: 11–24

SAS (1999) SAS Version 8.02: SAS/STAT Software: Changes and enhancements through Release 8.02. SAS Institute Inc, Cary, NC

Saunders JA, Mischke S, Leamy EA, Hemeida AA (2004) Selection of International molecular standards for DNA fingerprinting of Theobroma cacao. Theor Appl Genet 110:41–47

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN: A software for population genetics data analysis, version 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland, Availability: http://cmpg.unibe.ch/software/arlequin3/ (January 2011)

Somarriba E, Trujillo L (2005) El Proyecto. Modernización de la cacaocultura orgánica del Alto Beni, Bolivia. Agroforestería en las Américas 43–44:6–14

Soria VJ (1965) Notes on the native cacaos in the vicinity of Iquitos (Peru) and of Alto Beni (Bolivia). Cacao (Costa Rica) 10:14–16

Soria J (1966) Principales variedades de cacao cultivadas en América Tropical. Turrialba 16(3):261–265

Villegas RP, Astorga C (2005) Caracterización morfológica del cacao Nacional Boliviano, Alto Beni, Bolivia. Agroforestería en las Américas 43–44:81–85

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

Zeballos H, Terrazas E (1970) El cultivo de cacao en el Alto Beni. Instituto nacional de colonización. Boletín de divulgación técnica No 1 La Paz, Bolivia

Zhang DP, Mischke S, Goenaga R, Hemeida AA, Saunders JA (2006a) Accuracy and reliability of high-throughput microsatellite genotyping for cacao individual identification. Crop Sci 46:2084–2092

Zhang DP, Arevalo-Gardini E, Mischke S, Zúñiga-Cernades L, Barreto-Chavez A, Adriazola del Aguila J (2006b) Genetic diversity and structure of managed and semi-natural populations of cacao (*Theobroma cacao*) in the Huallaga and Ucayali valleys of Peru. Ann Bot 98:647–655

Zhang DP, Mischke BS, Johnson ES, Mora A, Phillips-Mora W, Meinhardt LW (2009) Molecular characterization of an international cacao collection using microsatellite markers. Tree Genetics & Genomics 5:1–10