Genetic diversity and geographic pattern in early South American cotton domestication

Ola Tveitereid Westengen

Master Thesis in Tropical Ecology and Management of Natural Resources

Department of Ecology and Natural Resource Management

Agricultural University of Norway

2004

Preface

‘Caral. Cotton. Peace.’ This riddle contains the background for this thesis. Professor Manfred Heun launched it, and I have spent the last one and a half year trying to solve it. The answer lays in the issue of cotton domestication, and its first traces in Peru. I am deeply thankful for the fun I have had pursuing this issue. It has led me through fabulous experiences in Peru, through the fascinating field of molecular biology, and through thought provoking literature on human prehistory.

I would like to express my sincere appreciation to my supervisor, Professor Manfred Heun for his guidance and encouragement through all steps of this research work. His initiatives and advices have been invaluable, and his support and patience admirable.

My gratitude also goes to Peru, where I met hospitality from so many people. Thanks to J. Lazo, G. Arevalo, F. Balavarca, C. Basurto, M. White, C. Deza, V. Medina, J. Escurra Puicon, P. M. Reyes More, E. Manco, and P. Azang Huaman for information and collaboration in collecting efforts in different parts of the country. Special thanks to Eric Rodriguez in Trujillo and Maria Olivos in Chiclayo, who even let me stay in their homes while they enthusiastically helped me in my mission. Finally, the success of my field work in Peru is largely due to Dr. Zosimo Huaman who helped with obtaining the legal permissions for collecting and exporting the plant material, and also accompanied me in parts of the expedition.

I would also like to thank Kari Vollan and Jørn Henrik Sønstebø for advices and technical assistance during laboratory work.

At last, a special thank goes to my dearest. Eilen, my family, and my friends are the foundation for everything I do.

Abstract

AFLP-fingerprinting was applied in a survey of genetic diversity in the cotton genus; *Gossypium*. Three diploid species (*G. raimondii, G. arboreum,* and *G. herbaceum*) and four allotetraploid species (*G. barbadense, G. hirsutum, G. mustelinum,* and *G. tomentosum*) were included in the study. Main focus was on the South American *G. barbadense* and its domestication history. A field collection of germplasm was conducted along northern coastal Peru and over an Andean transect to get a representative sample from the region containing most archaeological evidence on the first domestication of the crop. Gene bank material was added for comparison. Eight primer combinations yielded 340 polymorphic bands among the 131 accessions included. The obtained NJ and UPGMA analyses are fully congruent with cytogenetic evidence on the relationships between the diploid and the tetraploid species. Also at the allotetraploid interspecific level the accessions cluster according to taxonomic classification, well supported by bootstrap values. At the intraspecific level the genetic diversity of *G. barbadense* reveals geographic patterns. The accessions from coastal Peru display a distinct genetic diversity paralleled by their primitive agro-morphological traits. Accessions from the northernmost coast of Peru and provinces west of Andes in Ecuador cluster basal to accessions with origin east of Andes. The remaining accessions from Bolivia, Brazil, Columbia, Venezuela and the Caribbean- and Pacific- islands cluster with east of Andes accessions. NW Peru/ SW Ecuador (the area flanking the Guayaquil gulf) appear to be the center of primitive domesticated *G. barbadense* cottons from where it spread over the Andes and expanded into its pre-Columbian range. The locally maintained types from Peru serve traditional uses and represent a genetic resource pool of great importance.

Sammendrag

Denne studien tar for seg genetisk diversitet i bomullsslekten *Gossypium* ved hjelp av DNA-markøren AFLP. Tre diploide arter (*G. raimondii*, *G. arboreum*, og *G. herbaceum*) og fire allotetraploide arter (*G. barbadense*, *G. hirsutum*, *G. mustelinum*, og *G. tomentosum*) er inkludert i studien. Hovedfokus ligger på den Sør Amerikanske bomullsarten *G. barbadense* og dens domestiseringshistorie. Genetisk materiale ble innsamlet i Peru. Innsamlingen ble gjort langs de nordlige kystområdene og øst over Andesfjellene for å få et representativt utvalg fra den regionen hvor arkeologisk bevis for den første bruken og domestiseringen av denne landbruksplanta er funnet. Dette materiale ble komplementert med genbankmateriale av større geografisk spredning, samt andre arter som ikke forekommer i Peru. Åtte primerkombinasjoner ga 340 polymorfe AFLP-bånd blant de 131 inkluderte planteprøvene. NJ og UPGMA analyser av det totale datasettet er i full overensstemmelse med de foreliggende cytogenetiske konklusjoner angående slektskapsforholdet mellom de diploide og de allotetraploide artene. Også mellom de tetraploide artene skilte analysene i henhold til taksonomisk klassifikasjon, med høye bootstrap verdier på nodene. Innen G. *barbadense* framkommer det geografiske mønster i utbredelsen av den genetiske variasjonen. For det første viser plantene fra kystområdene i Peru en distinkt og enestående genetisk variasjon som samsvarer med deres primitive agromorfologiske trekk. Videre framkommer det at plantene fra provinsene vest for Andes i Ecuador, samt planter fra det nordvestligste Peru, forgreiner seg basalt i forhold til plantene med opprinnelse øst for Andes. Endelig viser det seg at planter fra Bolivia, Colombia, Venezuela, og øyene i stillehavet og Karibia viser størst affinitet med denne undergruppen med opprinnelse øst for Andes. På grunnlag av disse dataene, sammenstilt med paleobotaniske bevis fra den arkeologiske litteraturen, foreslås nordvest Peru/ sørvest Ecuador (området rundt Guayaquil gulfen) som området hvor *G. barbadense* bomullskultivering spredde seg over Andes og derfra ekspanderte utover sitt panamerikanske utbredelsesområde. De lokalt bevarte bomullstypene fra Peru utviser en variasjon i farge og egenskaper som gjør de viktige i tradisjonell bruk fra medisin til veving av fiskegarn. I tillegg representerer det genetiske mangfoldet i disse bomullstypene en potensielt viktig genetisk ressurs for planteforedling, og videre bevaring av denne diversiteten må sikres.

1. INTRODUCTION

1.1 Crop domestication

The study of crop domestication is a main focus within the study of human prehistory. The transition to agriculture, the so called Neolithic revolution, constitutes a profound change in humankind's organizational way of life and relation to the environment. The emergence of agriculture and the emergence of civilization are tightly linked (Price & Gebauer 1995, Diamond 1998) and consequently many different scientific disciplines have taken interest and part in the quest for where and when agriculture was first initiated. The investigation of the very material those first agents of change worked with, namely the plants and the animals that underwent domestication, is contributed to by archaeology, history, linguistics, anthropology, classical botanical taxonomy, biogeography, and genetics.

The where and when question concerning the emergence of agriculture, is as most profound questions not an easy one to answer, but as evidence is gathered in the field a fascinating and complex history unravel. The transition from different forms of hunter and gatherer societies to the first sedentary horticulturalists seems to have happened quite simultaneously in several distinct areas in the world in the change between Pleistocene and Holocene some 10 000 years ago (Smith 1995, Diamond 1998, Zohary & Hopf 2000, Gepts 2003). In the search for the geographical origins there are particularly two approaches that have been mutually fruitful; that of archaeology revealing remains of plants and equipment associated with the new subsistence form, and that of botany in its search for the wild relatives of crop plants. The 19th century Swiss botanist Alphonse de Candolle is called the father of the study of crop evolution. He proposed that the distribution range of the wild forms of domesticated plants also must represent the areas where man initiated his 'artificial' selection. This approach was refined by the distinct Russian geneticist Nicolai Vavilov in the 1920s and 1930s who drew up eight centers of agricultural origin around the world based on a simple but elegant hypothesis stating that the area containing the highest diversity of a given crop represents its cradle of domestication (Vavilov 1992). By comparing the variation in all the cultivars he encountered on his frequent collecting expeditions he recognized an overlap of diversity or 'genecenter' of different crops as if they where domesticated as crop complexes. Thus he formulated the notion of domestication centers, and even though changes and additions have been made to Vavilovs map, this notion of distinct centers have survived in most accounts of the matter (e.g. Smith 1995; Diamond 1999; Gepts 2002). Harlan (1971, 1975, 1992, 1995)

among others opposes the centric view and rather proposes a mosaic of change in subsistence systems and initiation of cultivation and domestication of indigenous plants in a wide belt between 35°N and 35 °S around the world. Although Harlan always emphasized the importance of Vavilov's work he came to question not only the eight centers' reflection of the crop evolution, but also the connection between the diversity within a crop and its origin as a crop plant *per se*; the human selection pressure might in fact work phenotypic diversifying when differential selection is imposed in distinct geographical regions, thus secondary diversification centers can be mistaken as primary gene centers. 'The single crop approach' proposed by Harlan emphasis the necessity of close crop by crop investigations to avoid oversimplifying. This is an ideal I pursue in this work on South American cotton domestication.

One of the important insights of Vavilov was, that studies on the present patterns of genetic diversity in crop plants not only are the key to their past history, but also to genetic resources for plant breeding. Vavilov conducted expeditions to collect germplasm from crops all over the world to ensure conservation of genetic diversity for future plant breeding. Also this study is conducted with this twofold aim; both to contribute to the historic-philosophical investigation of the Neolithic revolution, and to the research on genetic diversity as a priority for conservation.

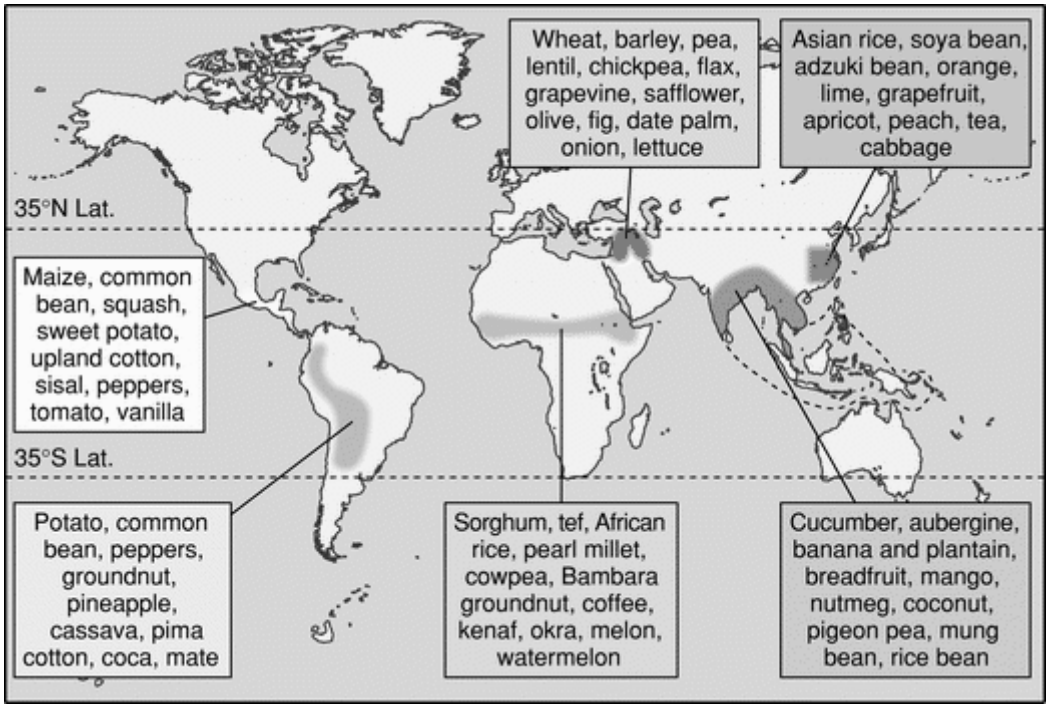


Figure 1.1 Illustration of 'domestication centers' with the crops that defines them. (From Gepts & Papa. Encyclopedia of Life Sciences. 2002)

1.2 From morphological to molecular markers in domestication studies

The study of crop evolution is a study of microevolutionary changes. It is a population genetic endeavor relying on analysis of genetic variation within and between wild and domesticated populations (intraspecific) to infer phylogeograpic patterns (Avis 1994).
A classical taxonomic approach to the study of crop evolution relies on morphological variation and analysis of derived and modified phenotypic traits (e.g. Vavilov 1992). Such 'morphological markers' have provided the whole framework for studies on crop domestication, and clear lines of evidence on origin area for many plants are established (Harlan 1992, 1995). Nevertheless the relatively small number of morphological traits is a limitation when one seeks to pinpoint domestication events for a crop whose progenitors are geographically widely spread, or when many different progenitor candidate species/subspecies are recognized.

Isozymes, defined as variants of the same enzyme differing between individuals and populaions, have been the subject for many studies in population genetics, including crop evolution (Lewontin and Hubby 1966, Stebbins 1989, Doebley 1989). Isozymes are the primary products of genes and are highly variable, thus they represent the next stage of refinement in the study of crop plants and their wild relatives. Doebley (1989) reviews the contribution to the crop evolution field by isozyme analysis and emphasizes a quality they possess in contrast to morphological traits; isozymes are neutral traits in relation to the artificial selection pressure imposed on morphological traits, thus they are not apt to have been altered by the process of domestication. This feature is a quality both when the crop is drastically altered from its putative wild progenitors such as for *Zea mays* and *Brassica oleracea,* and when the question of relationship is not resolvable due to close phenotypic similarity and many candidate origin areas (e.g. *Gossypium barbadense*). Isozymes were in fact the first molecular markers (Avis 1994).

The extension of this line of revealing evolutionary pathways through analysis of genetic variation came with methods for sequencing specific DNA regions and DNA marker techniques such as the Restriction Fragment Length Polymorphism (RFLP) method (Botstein et al 1980). Now the variation could be detected right at the DNA molecular level. By the introduction of the Polymerase Chain Reaction (PCR) and the possibility to rapidly amplify specific genetic regions the road was paved for a wide range of molecular markers such as microsatelites (Tautz 1989) based on amplification of highly mutable simple sequence repeat

loci (SSRs), Random Amplified Polymorphic DNA (RAPDs) (Welsh & McClelland 1990; Williams et al. 1990) based on amplification of arbitrary regions flanked by inverted primer sites, and Amplified Fragment Length Polymorphism (AFLP) (Vos et al. 1995) based on selective amplification of restricted DNA. All of the above mentioned molecular markers have been used in the study of crop evolution: e.g. RFLPs on the cotton species *Gossypium hirsutum* (Brubaker & Wendel 1994); sequencing of a single copy gene in manioc (Olsen & Schaal 2001); microsatelites on maize (Matsuoka 2002); RAPDs on buckwheat (Tsuji & Ohnishi 2000); and AFLPs on Fertile Crescent cereals (Heun et al. 1997, Badr et al. 2000, Özkan et al. 2002). This study on *G. barbadense* domestication is thus embedded in a methodological tradition by utilizing one of its newest and most effective tools, namely AFLP fingerprinting.

AFLP markers are nuclear genomic markers, generally assumed to be randomly distributed throughout the genome (Vos et al. 1995, Mueller & Wolfenbarger 1999). This assumption is verified numerous times in mapping studies where AFLP markers are found dispersed on all chromosomes (e.g. Lacape et al 2003). AFLP is a powerful tool for revealing genetic variation between and within populations as many polymorphisms can be detected per primer combination. AFLP is one of those high resolution markers that can generate 'DNA fingerprints', as a fragment composition specific for each individual organism can be generated (Parker et al.1998, Ouborg et al. 1999, Mueller & Wolfenbarger 1999). The technique has proven to be robust and highly reliable regarding reproducibility among laboratories (Jones et al.1997). Since total genomic DNA is the basis, the mode of AFLP inheritance is Mendelian. Thus the marker can be used in population genetics, but some limitations must be recognized: AFLPs are dominant markers, meaning that the banding pattern of homozygotes and heterozygotes can not be directly distinguished. This again means that genotype can be estimated, but allele frequencies can not be inferred without having a pedigree (Mueller & Wolfenbarger 1999, Ouborg et al. 1999). Prudence is needed when choosing taxonomic level of investigation as similarity of bands do not guarantee homeology of alleles, thus at higher taxonomic levels the similarity between distant taxa can be reduced to the random level (Mueller & Wolfenbarger 1999). Generally AFLP markers are best suited for intraspecific, or genealogical, investigations, but have also been found to resolve interspecific relations; depending on the level of polymorphism.

1.3 Some fundamental assumptions on genetic diversity in crop plants

A general, and probably the most fundamental assumption on the relationship between a wild progenitor and its descendant is that the former displays a broader genetic diversity than the latter (Doebley 1989, Tanksley & McCouch 1997, Gepts & Papa 2002). This assumption relies on features of the domestication process: The first farmers probably experimented with only a few plants to begin with and for every generation this small founder population faced hard selection pressure progressively narrowing the genetic base of the crop. Furthermore it is reasonable to assume that crop failures, random genetic drift in the small populations, and dispersal out of the origin area etc. have induced severe genetic bottleneck effects on the crop plants, leaving them with only a subset of the original gene pool (Tanksley & McCouch 1997, Gepts 2002) (see figure 1.2). Opposing these unifying evolutionary forces are forces such as introgression of variation from wild relatives, exchange of germplasm between farmers, selection against inbreeding, and maintenance of diverse landraces (Doebley 1989). But generally the assumption of lower diversity following domestication and dispersal underlies analysis such as ours.

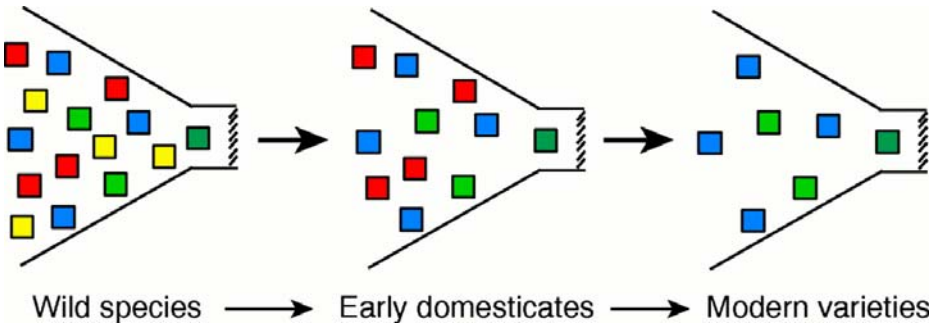


Figure 1.2 The narrowing of the gene pool during the bottlenecks of domestication. (From Tansksley & McCough. Science, Vol 277, 1997).

This study of *Gossypium barbadense* cotton domestication relies on genotyping accessions from all over its pre-Columbian geographical range by AFLP fingerprinting. As stated above this embeds the study in the tradition from Vavilov by applying the modern approach of Heun et al. (1997). Since this first AFLP fingerprinting based study on domestication centers, other studies have revisited the approach on other cereal crops (Badr et al. 2000, Özkan et al. 2002). The assumptions underlying the approach are stated by Heun et al. (1997): 1) Genetic distance within a species can be evaluated by multiple dominant DNA-markers. 2) The progenitor of crop plants have not undergone significant genetic change during the past 10 000 years.

1.4 Cotton domestication and the aim of this study

In the annals of crop domestication cotton holds a unique position as four of the genus' species were domesticated in four assumable independent events for the same trait (Wendel 1995, Brubaker et al. 1999). The elongated epidermal seed trichomes of two Old World hemisphere cottons (*G. herbaceum* and *G. arboreum*), and two New World hemisphere cottons (*G. hirsutum* and *G. barbadense*), were recognized as useful spinnable fibers by different groups of aboriginal people distantly located. Archeological evidence of cotton remains in association with human settlements goes back to the 6th millennium BC for the Greater Indus area in the Old World (Moulherat et al. 2002), and to layers dated to 6400 – 5000 years before present (BP) for the Zaña river valley in present northwest Peru (Rossen et al. 1996, Dillehay et al. 1997). Concerning the Old World species, domestication studies based on morphology and allozyme variation together with phytogeographic and historical considerations, propose that *G. herbaceum* has its domestication centre in Arabia, and *G. arboreum* in the Indus Valley of Pakistan (Hutchinson et al. 1947, Wendel et al. 1989, Brubaker et al 1999). The New World species have received more attention on this matter due to their present day economic importance. *G. hirsutum* has undergone the whole array of crop evolutionary studies from morphological- (e.g. Hutchinson 1951, Stephens 1976a; 1976b) via isozyme- (Wendel et al 1992) to genetic- (Brubaker & Wendel 1994) marker based studies of the intraspecific variation coupled with geographical associations. A well founded hypothesis finally best supported by a restriction fragment length polymorphism (RFLP) based study propose the Yucatan Peninsula in Mexico as the domestication center of this species (Brubaker & Wendel 1994).

G. barbadense domestication has also been extensively studied, though dominantly from the archeological and paleoethno-botanical fields of science. Along the coast of present day Peru and Ecuador a tremendous archeological record favorable for investigation due to the dry and conserving climate has revealed cotton remains in such abundances that even degree of domestication along the wild-to-domesticate continuum could be inferred (Stephens 1975, Stephens & Moseley 1973; 1974). The archeological evidence of a maritime subsistence for the Andean civilizations depending on cotton for fishing-nets etc. has led to the general perception of *G. barbadense* domestication to have taken place somewhere along this coastline. (This Archaeological context is treated in detail in chapter four.) Molecular evidence from an isozyme diversity study (Percy & Wendel 1990) is generally in accordance with the domestication scenario derived from archaeology: 153 *G. barbadense* germplasm

bank accessions representing the species' broad native range plus several improved cultivars of worldwide origin were analyzed based on allozymes from 24 polymorphic loci. The unspecified geographic region 'West of Andes' slightly exceed other regions in measures of genetic variability and a broad definition of the crops center of variability and origin in 'northwestern South America west of the Andes'(between Bolivia and Colombia), is stated. Representing the 'state of the art' in the field the study by Percy and Wendel (1990) contain certain weaknesses necessary to be pointed out. Their approach to the analysis of the allozyme diversity is a principal component analysis (PCA) where grouping into regional groups is done by subjective considerations. The PCA groups are treated as distinct categories in downstream statistical analyses, and the interpretation of these contains circle argumentation; i.e. the topology of the downstream cluster analysis is taken as supportive evidence for the grouping pattern of the PCA it is derived from. Further, the west of Andes group's slightly higher measures of genetic diversity can not be seen independent from the higher number of accessions included in this group, and in fact the vaguely defined origin area is stated so because trans-Andean and Bolivian accessions are included in the group (Bolivia's rough topology only allows cotton to grow in the Amazonian regions *east* of the Andes). In summary this allozyme study only contributes with a broad framework on the issue of locating the domestication center for *G. barbadense*. Unpublished isozymic evidence, available in Lazo (1991) propose a coast of Peru/ Ecuador domestication center for *G. barbadense* based on data for 14 polymorphic enzyme loci and phenotypic observations of 63 accessions. Piperno and Pearsall (1998) sums up the archaeological evidence on cotton domestication from Peru, integrates their own paleoethno-botanical evidence from coastal Ecuador (Damp & Pearsall 1994) and consolidates the isozymic evidence of Percy and Wendel (1990) when they place *G. barbadense* in a NW Peru/ SW Ecuador proposed domestication center together with *Phaseolus lunatus* (Andean), *Cucurbita ecuadoriensis*, and *Canavalia plagiosperma*.

Until now no study applying DNA- markers to assess the intraspecific genetic diversity and its phylogeographic implications for the domestication of *G. barbadense* has been conducted. The aim of this thesis is thus to apply the AFLP molecular marker technique to survey genetic diversity in *G. barbadense* and possibly reveal geographic patterns that can clarify its domestication history. In addition, the study includes accessions from closely related species: *G. herbaceum* and *G. arboreum* -the two A-genome diploid species that were domesticated in the Old World; *G. raimondii* –the endemic Peruvian wild D-genome diploid species; and *G.*

mustelinum, *G. tomentosum* and *G. hirsutum* –other members of the AD-genome tetraploid group where *G. barbadense* belongs. The AFLP method is earlier found suitable to reveal relationships between species (interpecific) in *Gossypium* (Abdalla et al. 2001), and these other species are included to contribute to the long lasting discussion on whether the American AD-tetraploids containing an Old World A-genome and a New World D-genome is the result of a pre-human natural transoceanic hybridization (Philips 1963, Senchina 2003) or originated as a crossing brought about by humans (Hutchinson et al. 1947, Heyerdahl 1968).

The core outcome of this study is presented in the paper in chapter five. Chapter two holds an extended methodological part, followed by a brief account in chapter three of the results that are more thoroughly presented in the paper. Chapter four is a review of the archaeological literature on the origins of agriculture in South West America in general and the excavated traces of *G. barbadense* cotton domestication in particular. The sixth chapter, 'Economic Botany', is a brief survey on some aspects of human use and dependence of crop diversity that are untypical inclusions in a genetic study; the policy issue on genetic resources conserning Peruvian cotton, and the etymologic and ethnographic role of the crop. Finally, chapter seven speaks for itself: 'Concluding Remarks and Future Directions.'

Figure 1.3 The native Peruvian *Gossypium barbadense* cotton. The naturally pigmented cotton still serves important purposes in weaving as well as in traditional medicine.

2. MATERIALS AND METHODS

2.1 Germplasm

The tissue samples included in this study are of two origins: 1) Self-collected material from Peru. 2) Seeds from the Cotton Collection of USDA-ARS. The accessions were selected to get a representative sample from the crop's pre-Columbian range with special emphasis on the native habitats in Peru.

2.1.1 Germplasm from Peru

Because of strong regulations on export of genetic recourses from Peru, the material in germplasm repositories are limited from this country. As a non-commercial research project we were able to collect material in field from this region considered a likely origin area of cotton domestication on the South American continent.

2.1.1.1 Permissions

Permission to import germplasm material from Peru and USDA-ARS seed-bank were granted from the Norwegian authority on these matters, 'Statens Landbrukstilsyn`.

The Peruvian embassy in Sweden and the Convention on Biological Diversity (CBD) focal point in Peru provided information on germplasm policy in Peru. Instituto Nacional de Recursos Naturals (INRENA) is the institution in charge of granting the permit for collecting activities on biodiversity in Peru. Upon arrival in Peru, together with project collaborator Dr. Z. Huaman (ProBioAndes), an application for germplasm collection was developed in accordance to the protocol of INRENA. The permission classified as 'Autorizacion N^{o} 025 - 2003 –INRENA-IFFS-DCB' were granted. This authorization was later expanded to other areas as information on previous collection sites etc. was obtained. The final collecting permit allowed us to collect in the departments of Ancash, La Libertad, Lambayeque, Piura, Tumbes, Cajamarca, Amazonas, San Martin, Loreto, Ucayali and Huànuco, with the exception of protected natural areas. A Material Transfer Agreement (MTA) was granted by the same institution upon presentation of a new application containing documentation of all the material and deposition of 50% of it in Peruvian institutions. Voucher specimens were deposited both at the Herbario Weberbauer at Universidad Nacional Agraria La Molina and at Herbarium Truxillense at Universidad Nacional de Trujillo. Germplasm in the form of seeds were deposited in the seedbank of 'Instituto Nacional de Investigacion Agraria' (INIA).

2.1.1.2 Locality, geographic and herbarium data

Prior to the expedition we compiled locality data from International Plant Genetic Recourse Institute (IPGRI) reports. Previous sampling of *Gossypium* germplasm in this area is published in Simpson et al. (1985) Schwendiman et al. (1985), and Percival & Kohel (1990). *G. raimondii* was sampled by Simpson et al. in 1983 in Chicama, Santa Ana, and Casca river valleys. In Chicama river valley also sampling of wild *G. barbadense* was reported. Schwendiman et al. in 1985 reported collection of wild *G. barbadense* in the provinces of Guayas and Los Rios in Ecuador. The collecting effort also aimed to cover areas where archaeological evidence of cotton cultivation had been excavated. Clues to interesting areas were obtained from archaeological literature; e.g. Stephens & Moseley (1973; 1974), Stephens (1975), Quilter et al. (1991), Damp & Pearsall (1994), and Solis (2001). Of more specific and accurate information value on geographic localities of plants and population were the consultancies with scientists and laymen upon arrival and during the travel in Peru. Finally the broad collection of voucher specimens at the herbarium of Universidad Nacional Mayor de San Marcos in Lima contained a well of more or less accurate information on localities of previous collections.

2.1.1.3 Field collection

The first phase of collecting was made in the coastal regions and means of transportation such as public busses, mc-taxies, and university vehicles was used. The second phase where I was accompanied by Dr. Huaman went to more remote and difficult accessible areas in the mountains and high jungle areas, and a four wheel drive pickup was used. Collecting was done by foot. Local residents led us to most of the plants and populations. Longitude, latitude and altitude data were obtained using a Global Positioning System (GPS) receiver (Garmin, eTrex Summit) and phenotypic observations and herbarium passport information were added. A sample of one small and fresh leaf was taken from each plant registered, and instantly put to dry on silica gel. Seeds were collected from those plants having mature bolls. Some representative plants had three branches cut off as herbarium vouchers. In total 100 accessions were made, five of these are *G. raimondii* and the rest are *G. barbadense*. The geographic localities of samples are plotted on a map in figure 1 in the paper. All information on collected accessions is presented in table form in appendix 1.

Figure 2.1 Cotton plant with boll and flower having its geographical position determined by GPS

2.1.2 Germplasm from USDA-ARS

A total of 77 accessions of *Gossypium* were kindly provided by The Cotton Collection of the United States Department of Agriculture –Agricultural Research Service (USDA-ARS), College Station, Texas. The germplasm repository provided 5 accessions of *G. arboreum*, 60 *G. barbadense*, 5 *G. herbaceum*, 3 *G. hirsutum*, 1 *G. mustelinum*, 3 *G. raimondii*, and 1 *G. tomentosum*. The accessions were chosen to represent a wide geographic range within *G. barbadense* and to include the phylogenetic closest species. The accessions included in the analyses are presented in table 1 in the paper.

2.1.3 Greenhouse growing

Those accessions having seeds were grown under controlled conditions in the greenhouse facilities of 'Senter for Klimaregulert Planteforskning' (SKP) at the Agricultural University of Norway. Tree seeds were planted for every accession in separate pots. Temperature was fixed at 25^{o}C and pot-soil was kept constantly humid. The day-length was set to 14 hours of artificial light. Germination happened between tree days and tree weeks after sowing. In total 126 plants grew cotyledons suitable for extraction.

2.2 DNA extraction

2.2.1 Tissue sampling

One plant from every accession was sampled. Preferably the plants were harvested when cotyledons were in the size range between ca 2 and 6 cm^2, the cotyledons were cut just above the meristem and air-brushed for eventual impurities before one was put to dry in a silica-gel filled plastic bag for storage, and the other immediately brought to grinding. Grinding was done with a pestle in an eppendorf-tube in liquid nitrogen. As soon as the leaf was ground to fine powder and the nitrogen had evaporated the sample was transferred to a -80 C^o freezer.

2.2.2 Extraction

DNA was extracted by using a DNeasy Plant Mini Kit from Qiagen. The manufacturer protocol was followed with minor modifications. Briefly, 400 µl of buffer AP1 and 4 µl RNase was added to about 100 mg of the wet weight tissue powder in an eppendorf-tube and left to incubate for 10 min at 65^{o}C to lyse the cells and digest the RNA. 130µl buffer AP2 was added to the lysate mix and incubated for 5 min to precipitate detergents, proteins, and polysaccharides. This resulted in a turbid and viscous lysate wich had to be centrifuged (Biofuge from Heraeus) for at least 5 min at full speed (13 000 rpm) before supernatant could be transferred to the QIAshredder spin column sitting in a 2 ml collection tube. The flow through was carefully removed to be added a 1.5 x volume of Buffer AP3 with ethanol. Between 500 and 700 µl flow through was obtained and two full volume rounds of centrifuge through the DNeasy spin column was needed to filtrate the sample. The DNeasy column was then placed in a new collection tube, and 500µl Buffer AW was centrifuged through in two steps: for 1 min at 8000 rpm; then for 2 min at maximum speed. For some samples the membrane was darkly colored at this stage and an additional spin with 96% ethanol was done according to the troubleshooting manual. To dry the membrane completely it was necessary to spin another min with no buffer added. Finally the column was transferred to a 1.5 ml eppendorf-tube, 100 µl preheated (65^{o}C) Buffer AE was pipeted directly ont the membrane and left to incubate for 5 min before it was spun through in centrifuge for 1 min at 8000 rpm to elute. This elution was repeated to yield two DNA eluates for every plant before storage at 4^{o}C.

2.2.3 Testing for quality and quantity

This crude DNA eluate was tested for quality and quantity by running an agarose gel electrophoresis. 3 µl eluate of every sample were mixed with 10 µl MQ-water and 2 µl loading dye and run on a 0.8% 350 ml ethidiumbromide agarosegel at 100 V for

approximately 80 min. 0.5% TBE was used as buffer in the running chambers. The gel was exposed to UV light and a Polaroid picture was taken. The quality of the sample is reflected by the distinctiveness of the light band on the photo: A smear indicates protein contamination and/or degraded DNA, while a clear cut band indicates high purity and quality. The DNA concentration was estimated by comparison against λ-DNA (Fermentas).

2.3 The AFLP procedure

AFLP fingerprinting was performed according to Vos et al. (1995). The total genomic DNA underwent the three main steps of AFLP fingerprinting: 1) Digestion of the DNA by restriction endonucleases and ligation of double stranded adapters to the loose ends to generate template DNA for amplification. 2) Selective amplification of a subset of restriction fragments by the use of PCR primers with selective nucleotides. 3) Amplified fragments are analyzed by denaturing polyacrylamide gel electrophoresis.

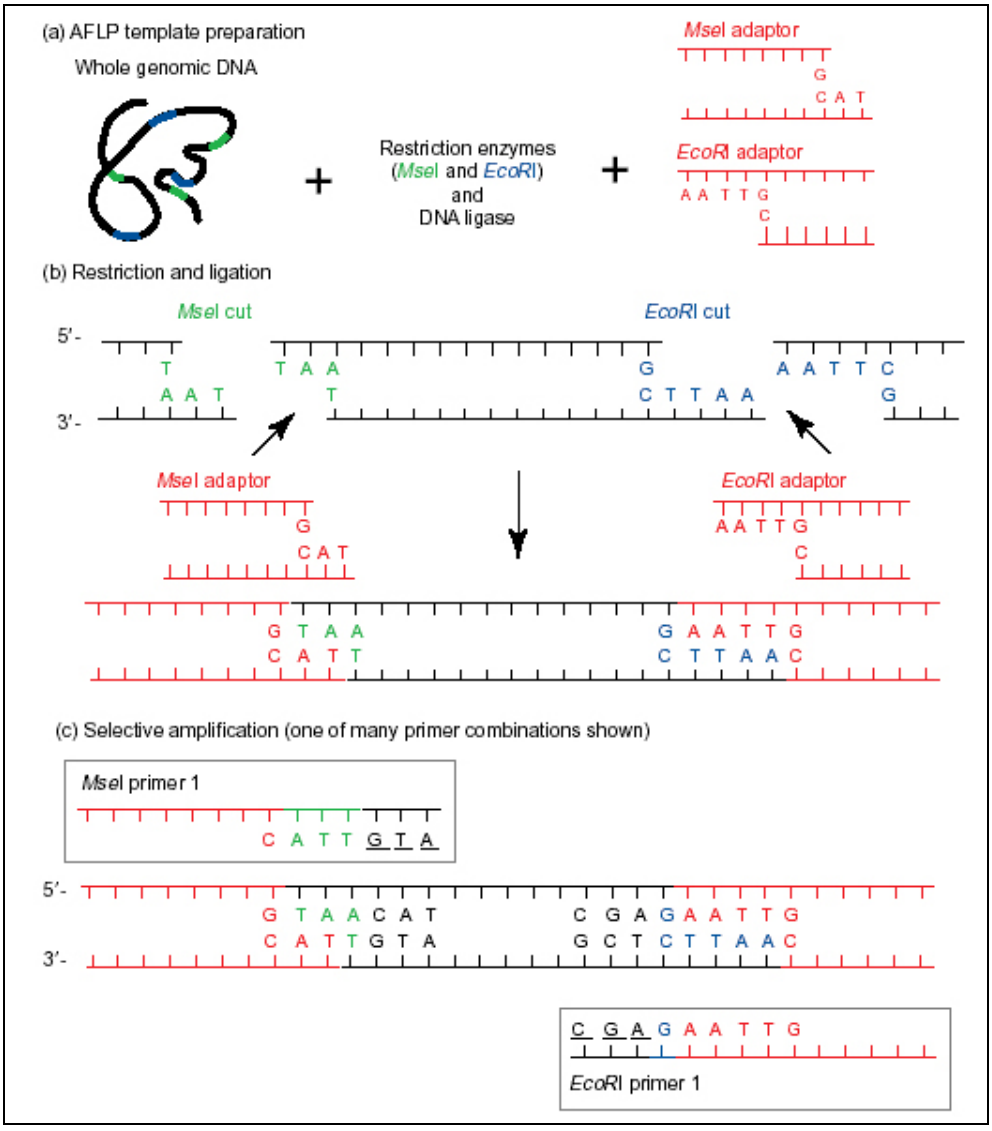


Figure 2.2 The AFLP procedure. (From Mueller & Wolfenbarger, TREE, Vol. 14. 1999)

2.3.1 Digestion and ligation

A protocol derived from Becker et al. (1995) was followed. The DNA samples were digested by two restriction endonucleases simultaneously: *Eco*RI has a 6-bp recognition site and the frequent cutter *Mse*I has a 4-bp recognition site, together they generate small DNA fragments in the optimal size range of less than 1 kb for separation on polyacrylamide gels. The samples were diluted and pipeted into 200 µl tubes on PCR-plates to equalize the DNA concentration in every sample. A final amount of 500ng DNA was prepared for every sample and added a mixture of 5U E*co*RI (Fermentas), 5U *Mse*I (NEB), 5µl 10 x RL-buffer in a total volume of 50 µl. The DNA was digested for 2 hours at 37^{o}C.

The restricted DNA was then added 10µl of a ligation mix containing 1.0 µl *Eco*RI adapter (5'-CTC GTA GAC TGC GTA CC-3'/ 3'-CTG ACG CAT GGT TAA-5'), 1.0 µl *Mse*I adapter (5'-GAC GAT GAG TCC TGA G-3'/ 3'-TA CTC AGG ACT CAT-5') (Invitrogen), 1.2 µl 10 mM ATP, 1.0 µl 10xRL-buffer, 1.0 µl T4 DNA ligase (1U/µl) (Fermentas), and 4.8 µl ddH_2O. The restricted DNA was incubated with the ligation mix for a minimum of tree hours at 37^{o}C.

2.3.2 Selective preamplification

A selective preamplification step with primers complementary to each adapter sequence with a one-base extension was performed to amplify a subset of the total restriction fragments. It was decided to perform the preamplification with E01 and M02 (Invitrogen) based on prior reports of AFLP fingerprinting on cotton (Abdalla et al. 2001, Lacape et al. 2003). A preamplification mix containing 1.5 µl of each +1 primer, 1.0 µl 10 mM dNTP mix, 0.25 µl Taq polymerase (5U/µl) (Fermentas), 5.0 µl 10 x PCR-buffer, 5.0 µl 25nM $MgCL_2$, and 30.75 ddH_2O was added to 5.0 µl template DNA of every sample. PCR reaction was carried through on a PTC-200 thermocycler (MJ Research) running 30 cycles of: 94^{o}C in 30 sec, 56^{o}C in 60 sec, 72^{o}C in 60 sec, and 4^{o}C as soak temperature. 10 µl of the preamplification product was diluted 1:19 with ddH_2O prior to selective +3/+3 amplification.

2.3.3 Selective Amplification

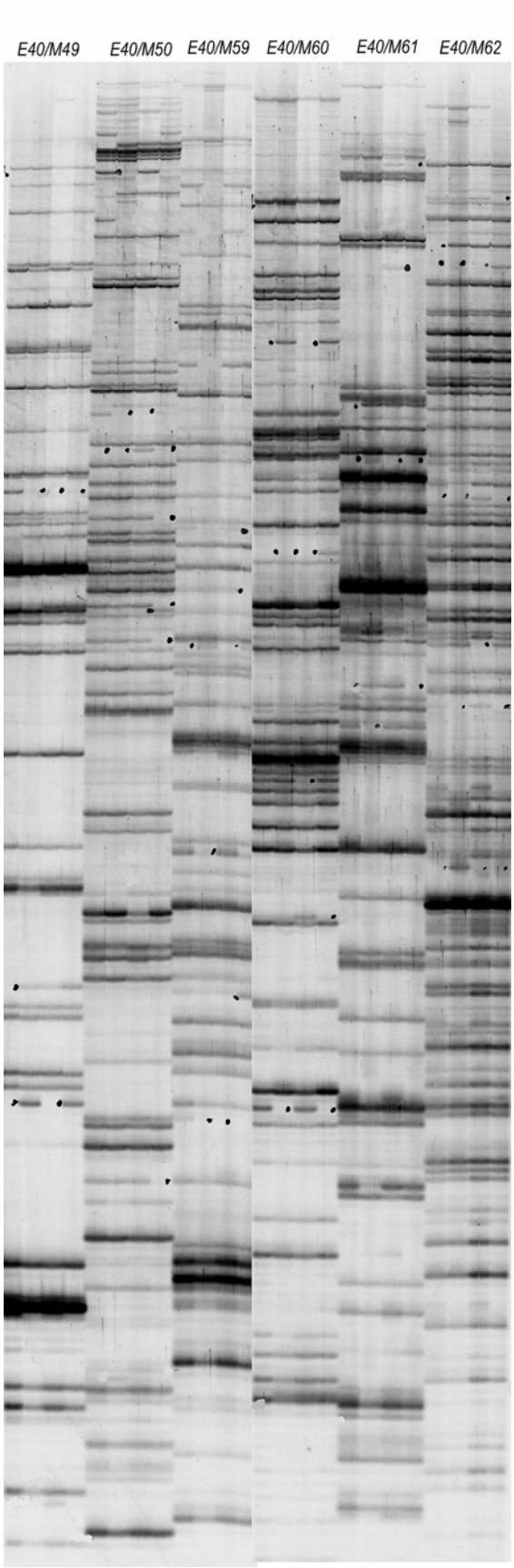
Primer labeling was done by phosphorylating the 5'-end of the *Eco*RI primers with radioactive γ^{33}P and T4 kinase (Fermentas). The primer was diluted to 50ng/µl and the mix for one reaction consist of 0.05µl 10 x buffer A, 0.29 µl ddH_2O, 0.01 µl T4 PNK (Fermentas), 0.05µl γ^{33}P ATP, 0.1 µl *Eco*RI +3 primer. The labeling of the *Eco*RI primer is a forward

reaction performed in a 500µl eppendorf tube at 37°C for 2 hours, followed by inactivation at 67°C for 15 min.

The selective amplification with +3/+3 primers was performed in a hot PCR mix with the following composition (1x rxn): 0.5µl *E primer, 0.6µl M primer, 0.4µl dNTP (10 mM), 2.0 µl PCR buffer (10x), 1.2µl $MgCL_2$ (25 mM), 0.1µl Taq polymerase (5U/µl), 10.2µl ddH_2O. 5 µl of the diluted preamplificated DNA template had this mix added and directly placed on a GeneAmp thermocycler (Applied Biosystems) for the final +3/+3 PCR run: 1 cycle at; 94°C for 30 sec, 65°C for 30 sec, 72°C for 60 sec; then the annealing temperature was lowered by 0.7°C for each cycle during 12 cycles; and finally for 23 cycles at 94°C for 30 sec, 56°C for 30 sec, and 72°C for 60 sec. The PCR product was then added an equal amount of loading buffer for PAGE (80% formamide, 10 mM EDTA (pH 8), 1.0 mg/ml Xylene Cyanol FF, 1.0 mg/ml Bromphenol Blue.) and stored at -20°C prior to gel run.

2.3.4. Gel analysis

The samples were runned on 5 % polyacrylamide gels (19:1 acrylamide:bis; 8M Urea; 1xTBE buffer). Polymerization was initiated by adding 300µl 10% APS (amoniumpersulfat) and 30 µl TEMED (N,N,N,N-Tetramethylenediamine) to 60 ml polyacrylamide stock and poured with 0.4 mm spacers and sharkstooth combs. The gels were mounted in a vertical gel elctrophoresis apparatus (Model S2/ S2001Gibco BRL Life Technologies) and preheated for about 30 min. Running buffers were 2xTBE in bottom chamber and 1xTBE in top chamber. The PCR products were denatured at 94°C for 10 min and directly chilled on ice upon loading the gel. 3µl of each sample was loaded and 1.5 µl of a labeled 30-330 bp AFLP ladder was loaded for comparison. The gels were run at 80 watt power for 90 min. After the run the gel was fixed in 10% acetic acid for 30 min and finally washed 2 x 1 min and 1x 10 min in dH_2O before drying over night. The gels were then exposed to X-ray film (Kodac BioMax MR) from 8 to 48 hours, depending on radiation intensity, and developed on a Curix 60 from AGFA. The resulting autoradiograph images could then be scored for presence or absence of bands using a light box.



2.4 Fingerprinting

2.4.1 Screening for primer combinations

All 56 combinations of the primers E32, E33, E35, E38, E40, E41, E42 and M47, M48, M49, M50, M59, M60, M61, and M62 (Invitrogen) were tested on four selected genotypes of *G. barbadense*: P45; P80; Ch 36; Gb 480.

2.4.2 Final fingerprinting

Eight primer combinations were used in the final fingerprinting of all 131 accessions: E35/M48; E38/M48; E38/M49; E40/M50; E40/M59; E40/M60; E41/M47; E41/M60. The sequences for these primers are given in table 2.1. To assure that only reproducible bands were scored and that the different PAGEs were well aligned overlapping of samples was applied both in the selective PCR and when loading the gels. Example of fingerprint image is given in figure 2.2.

Table 2.1 Sequence of primers used in the final fingerprinting.

Primer name	Sequence (5' – 3')
E-01	-GACTGCGTACCAATTC**A**NN-
E-35	- **ACA**-
E-38	- **ACT**-
E-40	- **AGC**-
E-41	- **AGG**-
M-02	-GATGAGTCCTGAGTAA**C**NN-
M-47	- **CAA**-
M-48	- **CAC**-
M-49	- **CAG**-
M-50	- **CAT**-
M-59	- **CTA**-
M-60	- **CTC**-

Figure 2.3 Example of AFLP fingerprints obtained from screening

2.4.3 Data analysis

The X-ray films were scored in a binary manner by giving the presence of a polymorph AFLP band a score of '1' and a '0' for its absence. Only unambiguous bands within an 80 to 800 bp range were scored. The data was assembled in a matrix and analyzed for pairwise genetic similarity using the Dice similarity coefficient (Dice 1945, Nei & Li 1979) and the Jaccard's similarity coefficient (Jacard 1908). Cluster analyses of the total data set and subsets of this were carried out both in the program NTSYS-pc v.2.11f (Rohlf 2000) and in PAUP v.4.0b10 Macintosh (Swofford 1998). Dendrograms were constructed using the unweighted pair-group method with arithmetic averages (UPGMA) (Sokal & Sneath 1963) and the neigbour joining (NJ) method of Saitou & Nei (1987). The robustness of the obtained trees was evaluated by comparing different data analyses (Dice and Jaccard matrixes analyzed in NJ vs. UPGMA, vs. principal coordinate analysis). In addition, bootstrap analyses (Felsenstein 1985) with 1000 re-samplings were done in PAUP.

3. RESULTS

3.1 Germplasm

In total high quality DNA was obtained for 144 accessions, 131 of these were included in the final analysis; 65 obtained from USDA-ARS seedbank and 67 from material collected in Peru.

3.2 Fingerprinting

3.2.1 Screening for primer combinations

The screening of primer combinations to search for suitable primer combinations yielded interpretable results for 39 combinations of E01 and M02 primers. Both total number of band, number of polymorphic band, and resolution of bands were highly variable between combinations. Table 3.1 displays a matrix of the primer combinations, where number of polymorphic bands relative to total number of bands is given as fraction and percent. Table 3.2 is an evaluation matrix of the primer combinations where also resolution of bands is taken into account.

Table 3.1 Matrix of primer combinations. Number of polymorphic bands/ total number of bands. (% Polymorphic bands)

E / M	M47	M48	M49	M50	M59	M60	M61	M62
E35	2/38 (3.4%)	7/68 (12%)	6/62 (9.7%)	1/79 (1.3%)	4/73 (5.3%)	4/71 (5.6%)	6/62 (9.7%)	1/64 (1.6%)
E38	3/80 (3.8%)	7/59 (11.9%)	7/63 (11.1%)	5/85 (5.9%)	0/76 (0%)	3/62 (4.8%)	8/56 (14%)	5/70 (7.1%)
E40	3/59 (5%)	5/52 (9.6%)	4/40 (10%)	14/76 (18.4%)	8/71 (11.2%)	6/54 (11.1%)	5/45 (11.1%)	7/79 (8.9%)
E41	9/76 (11.8%)	4/61 (6.6%)	4/50 (8%)	1/57 (1.8%)	4/53 (7.5%)	12/56 (21.4%)	5/53 (9.4%)	4/57 (7%)
E42	3/64 (4.7%)	5/75 (6.7%)	6/66 (9%)	1/74 (1.3%)	m.d.	1/52 (1.9%)	9/70 (12.9)	5/70 (7.1%)

Table 3.2 Quality of primer combinations considering polymorphism and resolution of bands.

E / M	M47	M48	M49	M50	M59	M60	M61	M62
E35	bad	Good	ok	bad	ok	ok	ok	bad
E38	bad	Good	Good	ok	bad	ok	ok	bad
E40	ok	ok	ok	Good	Good	Good	ok	Good
E41	Good	ok	bad	bad	bad	Good	Good	ok
E42	bad	ok	bad	bad	m.d.	bad	ok	ok

3.2.2 Final fingerprinting

The results of the fingerprinting is presented in the paper and displayed in its figure 2 and 3 as both NJ and UPGMA trees. Inter-specific relationships in agreement with cytogenetic evidence and taxonomic classifications were obtained. At the intra-specific level some geographic associations were apparent and aspects of the domestication and dispersal history of *G. barbadense* could be inferred. This same pattern is also apparent in the principal coordinate analysis presented in appendix 2.

4. THE ARCHAEOLOGICAL BACKDROP

4.1 The origins

The main theory on population of the American continent by humans states a migratory route across the Bering land bridge in late Pleistocene. The theory assumes an ice-free corridor from Siberia to Alaska where the first Americans pursued their subsistence on big-game hunting. The first culture of the Americas universally accepted in the archaeological society is the so-called 'Clovis' culture. The culture is named after a site near Clovis in New Mexico where a spear point was found stuck in the bone remains of a now extinct bison species. This type of points has been recovered from an enormous range of both the northern and southern part of the continent, and a reliable dating of about 12000 years BP has been determined (e.g. Goebel et al. 2003). Undermining the status of the Clovis culture as the earliest expression of human presence on the American continent is a wide array of excavation sites and findings predating the characteristic spear hunters by thousands of years. Most of these earlier claims of human touch have been reported from South America (Harlan 1995, Marshall 2003). Most notably is the consensus of many early man experts in 1997 of a ca.14 850 calender years B.P. occupation of the Monte Verde site in southern South America (Dillehay 1989, Meltzer 1997). This opened up for the acceptance of also other pre-Clovis sites. Since this study deals with the origin of crop domestication this discussion on the first paleoindian footsteps is not of direct relevance, but it is important to bear in mind that no theory on this first discovery of America yet can be justified above the level of hypothesis. This means that the ancestors of the first agriculturalists, and eventually the first cotton growers, have an unknown paleohistory, and also hypotheses on transoceanic voyages (e.g. Heyerdahl 1968, Dillehay 2003) must be taken seriously, with the potential diffusion of knowledge and spreading of plant material this can have caused.

One of the earliest sites containing agricultural remains in South America is the Guitarrero Cave at 2800 m elevation in an inter-montane valley between the Cordillera Blanca and Cordillera Negra mountain ranges in central Peru (Lynch 1980). The plant remains found are classified as common bean (*Phaseolus vulgaris*), lima bean (*Phaseolus lunatus*), oca (*Oxalis* sp.), chili pepper (*Capsicum* sp), *Solanum hispidium,* squash (*Cucurbita* sp.), and pacay (*Inga* sp.). The layers they are found in are dated to 8600 to 5600 BC, this is in fact, as Harlan (1995) points out, evidence indicating plant domestication in the New World as early, or

possibly earlier, than in the Old World. Others, most notably Piperno and Pearsall, authors of the book 'On the origin of agriculture in the lowland neotropics'(1998), consider the low-lying regions the most likely candidate-areas of origin of plant cultivation on the American continent. In accordance to Harlan's non-center view, they propose a minimum of three independent 'origin of domestication' areas: southwestern Ecuador/northern Peru, northern South America (Colombia/Venezuela/the Guianas/northern Brazil), and southwestern Mexico (Piperno & Pearsall 1998).

In excavation of early Holocene sites the first forest tubers, seed plants and tree fruits are recovered in substantial amounts. And for the first time plant grinding implements and tools for working soil shows up in the archaeological record: In the Las Vegas site on the Santa Elena Peninsula of Ecuador, *Cucurbita* phytolits are recovered from layers dated between 11000 and 10000 y. B.P., indicating a early domestication of squash. And at the San Isidro site in present day Colombia a carbon date of 10 050-9539 B.P. is obtained for the finding of stone grinding tools and remains of a wide range of tubers, grasses, legumes and tree fruits.

4.2 Early subsistence in Peru

The earliest traces of paleoindians along the Peruvian coast are from the site Quebrada Jaguay yielding a 13 000 – 11 000 calibrated years before present ^{14}C date (Sandweiss et al. 1998). This site is suggested by the authors to have been a seasonal settlement for nomadic foragers who spent part of the year in the highland areas and the other part on the coast exploiting maritime recourses.

The first coast culture with apparently sedentary coast settlements was the Paijan culture dated between 10 800 and 8 300 B.P. (Dillehay et al. 2003). This culture is characterized by using 'fishtail' points and they seem to have relied on a seafood subsistence base (Moseley 2001). The first Paijan sites will never lend themselves to archaeological investigations since the sea level has risen to about 100 m. above ice age level. In the later Paijan sites the tool technology shifted to more grinding stones and Dillehay et al. (2003) suggest a gradually larger reliance on plant food. Grinding stones are also highly present in the sites excavated by Dillehay et al. (1989, 1997) and Rossen et al. (1996) in the Zaña Valley in the present day department of Lambayeque. These excavations are indeed interesting concerning the subject of this thesis: The Zaña Valley not only contains the earliest evidence of horticultural

societies along the Andean vest flanks, but is also the area where cotton appears for the first time in the archaeological record. The first plants associated with these Early Holocene dwellings are, among others, manioc (*Manihot esculenta*), peanuts (*Arachis hypogea*), squash (*Cucurbita* sp.) and quinoa (*Chenopodium quinua*) dated to 7950 ± 180 B.P. It is an interesting feature with these plants that they all most probably are exogenous crops, meaning that their assumed cradle of origin are as distant as the Andean highlands of today's Bolivia for quinoa (Smith 1995) and the Ecuadorian Amazon for manioc (Olsen & Schaal 1999). This indicates cultural interactions over long distances already at this time.

4.3 Cotton enters the scene

The oldest remains of cotton excavated by archaeologists are excavated in the so-called 'Middle Preceramic', the period between 7700 – 5000 B.P. At the Chilca site of central coastal Peru cotton occurred in tomb context from this period (Piperno & Pearsall 1998), and about 80 km inland in the Zaña river Valley seeds identified as *G. barbadense* are obtained from layers dated to 6 400 – 5 000 B.P., the so called Tierra Blanca phase (Rossen et al. 1996, Dillehay et al.1997). Cotton appears in the same layer as cocoa (*Erythroxylon* sp.) and in connection with small but significant organizational changes in the sites: The first special-purpose, nonresidential sites occur, probably used for ceremonies of some sort (Dillehay 1989, 1997). Also from coastal Ecuador cotton seeds identified as *G. barbadense* are obtained from Middle Preceramic layers. The seeds are recovered from the Real Alto site in the Chanduy valley and Damp and Pearsall (1994) reports a ^{14}C date of 3500 – 3000 BC. on the context layers. Stephens et al. (1973, 1974, 1975) did some extensive studies on cotton remains from a range of early sites along the Peruvian coast. From five sites in the Ancon – Chillon area of central coastal Peru cotton materials of many types were obtained and analyzed: Scraps of twined fabrics and cordage; unprocessed fibers; carbonized seeds; uncarbonized seeds, kernels and boll segments (Stephens & Moseley 1973, 1974). This material, and the examination of cotton remains from the North Peruvian site of Huaca Prieta soon after (Stephens 1975), represent the pioneering research on *G. barbadense* domestication. This work revealed the first morphological changes on the Peruvian cotton induced by a domestication process. Older remains excavated can only suggest some kind of management of cotton (e.g. Damp & Pearsall 1994) while Stephens data provide direct information on a certain domestication status of the plants initiated at the time of 2500 B.C. At this time in the archaeological record the appearance of cotton remains suddenly flourish,

and the name of the period in the cultural chronology speaks for itself: 'Cotton Preceramic', ranging between 4600 – 2800 B.P. In this period cotton is present in 100% of all sites excavated and thus have been assigned a greater role in Peruvian archaeology than any other plant. While as many as 19 crops have been recorded for the period from different sites along the coast (Piperno & Pearsall 1998), there are no others than can match the abundance of cotton. Associated with cotton in many excavations are bottle gourds (*Lagenaria* and *Cucurbita*) and evidence of a marine subsistence base. Apparently cotton fibers were spun and plied for yarn and cord that was looped and knotted to produce fishing nets. The bottle gourds probably worked as flotation. These findings led Moseley to propose the hypothesis about 'the Maritime Foundations of the origins of Andean Civilizations' (MFAC) (Moseley 1975, 2001; Sanweiss & Moseley 2001). According to this theory, the marine subsistence base played a formative role in the development of social complexity, and so sparked the beginning of civilization on the South American continent. The credo in archaeology is that sedentary and social complexity first arose with an agricultural subsistence base such as along the great rivers of the Middle East. So when Moseley and supporters propose that fisher folk were the first to participate in complex organization and to construct the first monumental architecture on this continent they in fact challenge a fundamental notion in archaeology. The growing evidence of a broader subsistence base than first thought by Moseley (1975) (e.g Ugent et al. 1982;1984;1986, Raymond 1980) led to a modification of the MFAC to incorporate plants as important carbohydrate and vitamin sources (Quilter et al.1991, Moseley 2001).

Figure 4.1 Above: The archaeological site Huaca Prieta (occupied between 3100-1300 BC) in todays department La Libertad. Left: A modern descendant of the Moche Indians showing her dooryard cotton plant. Right: Excavated plied cotton fibers of natural colors.

Lately MFAC has been challenged on principal means based on the early dates obtained from the site of Caral in the Central Peru (Solis et al. 2001). Calibrated radiocarbon dates yielded an occupational date range from 2627 BC to 2020 BC for this monumental center placed 23 km inland in the Supe river valley. Haas and Creamer (2001), coauthors of the Shady et al. (2001) report, argue that the exploitation of the great anchovy populations was agriculture dependent from the beginning. In their view inland farming came first and later started the exchange of cotton nets and other commodities for marine goods. This evidence has led some crop evolutionists (e.g. Gepts 2004b) to reconsider the pattern of cotton domestication developed by Stephens and Moseley. Could it not be that the slight inland placement of Caral reflects a more classical pattern where agriculture emerged inland and then later spread both its agents and their technology to the coast? Sandweiss and Moseley (2001) argue that the term ‘monumental architecture’ is a matter of definition, and that the uniqueness in size and

location of Caral is amplified. In their view Caral brings in important data, but nicely embeds in the MFAC both in time, location, and subsistence.

One of the most thorough excavations of a Cotton Preceramic site is that of El Paraiso in close proximity, but slightly closer to the coast than Caral (Quilter et al. 1991). The food refuse reveals that although fish was the primary animal food at the site, the inhabitants also heavily depended on both domesticated and wild plants: Achira (*Canna edulis*) and jicama (*Pachyrrhizus tuberosus*) are tuber producing plants still in use in traditional Peruvian cusine. More generally known are common and lima beans (*Phaseolus vulgaris*, and *P. lunatus*), tomato (*Physialis* sp.) and chili pepper (*Capsicum* sp.). Many fruits were utilized, such as guava (*Psidium guajava*) and pacay (*Inga feuillei*), and gathered roots from wild sedges seems to have played an important role still at this stage. Nevertheless there is one plant that appeared in such substantial amounts that it must be assumed a paramount role in the economy of El Paraiso; cotton. The ubiquity of cotton is probably caused by its twofold role as the major industrial product supporting fishing and also providing the raw material for textiles for clothing. The authors of the El Paraiso report go far in their consideration of the determining role played by cotton:

"A chief cause for the concentration of populations in larger sites may thus have been due to the desire for increased cotton production (…)" (Quilter et al. 1991, p.282)

The intensification of cotton farming with the introduction of irrigation agriculture together with the first ceramics and weavings marks the beginning of the next archaeological period called 'The Initial Period' (Pozorski and Pozorski 1990). It is not before in the next period, the 'Early Horizon', that cotton is absent from some excavated sites, although still present in about 80% (Piperno & Pearsall 1998). At this time (2800-2200 BP.) social complexity and diversity is flourishing along the Peruvian coast, more niches become exploited, and the foundations are laid for a long row of native civilizations culminating in the great Inca empire in the 15th century A.C.

Table 4.1 Cultural Chronology. 'Horizons' refer to times where a wave of stylistic influence swept across regional boundaries and 'periods' refer to times characterized by regional styles. (sources: Dillehay et al. 1997, Davies 1997, Moseley 2001)

Time	Cultural feature/name
Lithic- / Early Preceramic- Period ~7700 BP	Paijan Culture and the Guitarrero Cave Dwellers.
Middle Preceramic Period 7700-5000 BP	Earliest cotton excavated.
Cotton / Late Preceramic- Period 4600-2800 BP	Earliest monumental architecture. Cotton present in all sites
Initial Period 4000-3500 BP	Earliest ceramics.
Early Horizon 2800-2200 BP	The Chavín civilization
Early Intermediate Period 200 B.C. – 600 A.D.	Moche Culture / Nazca Culture / Tiwanaku Culture
Middle Horizon 600 – 950 A.D.	Sican / Moche / Huari / Ica / Tiwanaku
Late Intermediate Period 950 – 1400 A.D.	Chimu / Chancay / Ica / Kilke / Aymara
Late Horizon 1400-1532 A.D	Inca Kingdom
The Spanish Conquest 1532 A.D.	Pizarro and his men, aided by guns, germs and horses conquer the realm of the Incas.

Figure 4.2 The Inca site Machu Pichu. Agricultural terraces surround the inner city. It was built by the Inca ruler Pachacutec only 60-70 years before the Spanish conquest.

5. PAPER:

Genetic diversity and geographic pattern in early South American cotton domestication

Ola T. Westengen[1], Zósimo Huamán[2] and Manfred Heun[1,3]

[1] Department of Ecology and Natural Resource Management, Agricultural University of Norway, Po.Box 5003, 1432 Aas -Norway

[2] PROBIOANDES, Calle El Herraje 120, Lima 12 - Peru

[3] Communicating author

Keywords AFLP fingerprinting, *Gossypium*, genetic diversity, domestication

Submitted to Theoretical and Applied Genetics 08.05.04.

Abstract

AFLP-fingerprinting was applied to survey genetic diversity of primitive South American *Gossypium barbadense* cotton for establishing a possible link to its pre-Columbian spread. New germplasm was collected along coastal Peru and over an Andean transect in areas where most archaeological evidence relating to cotton domestication has been recorded. Gene bank material of three diploid (*G. raimondii, G. arboreum,* and *G. herbaceum*) and four allotetraploid cotton species (*G. hirsutum, G. mustelinum, G. tomentosum* and additional *G. barbadense*) was added for inter- and intraspecific comparison. Eight primer combinations yielded 340 polymorphic bands among the 131 accessions. The obtained NJ and UPGMA analyses are in full agreement with the known cytogenetics of the tetraploid cottons and their diploid genome donors. The four tetraploid species are well distinguished according to taxonomic classification. The genetic diversity within *G. barbadense* reveals geographic patterns. The locally maintained cottons from coastal Peru display a distinct genetic diversity paralleled by their primitive agro-morphological traits. Accessions from the northernmost coast of Peru and from SW Ecuador cluster basal to east of Andes accessions. The remaining accessions from Bolivia, Brazil, Columbia, Venezuela and the Caribbean- and Pacific-islands cluster with east of Andes accessions. NW Peru/ SW Ecuador (the area flanking the Guayaquil gulf) appear to be the center of primitive domesticated *G. barbadense* cottons from where it spread over the Andes and expanded into its pre-Columbian range.

Introduction

The cotton genus, *Gossypium* L., consists of about 45 diploid and five tetraploid species forming a monophyletic group (Fryxell 1992, Wendel & Cronn 2003). Eight diploid genomes (A-G and K) each comprising 13 chromosome pairs have been identified (Endrizzi et al. 1985). The native distribution of all five tetraploid cottons (*2n* = 4x = 52, AADD) is restricted to the New World although its emergence involves a combination of an Old World A-genome (derived from an ancestor of *G. arboreum* L. and *G. herbaceum* L) and a New World D-genome (derived from an ancestor of *G. raimondii* Ulbrich) (Stephens 1944, Phillips 1963; 1964, Seelanan et al. 1997, Cronn et al. 1999, Liu et al. 2001). The underlying transoceanic dispersal of the A genome donor and the timing of the polyploidization event has been controversial, but the current view is that a single Mid-Pleistocene (1-2 mya) polyploidization event has occurred (Phillips 1963, Wendel & Albert 1992, Senchina et al. 2003). It is further assumed, that the emergence of tetraploid cotton was followed by long distance separation into five species: Three of which are truly wild species: *G. mustelinum* Miers ex Watt (limited distribution in northeast Brazil, Wendel et al. 1994), *G. darwinii* Watt (endemic to the Galapagos Islands, Wendel & Percy 1990) and *G. tomentosum* Nutall ex Seemann (endemic to the Hawaiian Islands, DeJoode & Wendel 1992). The other two species have undergone domestication: *G. hirsutum* L. (predominantly distributed in Meso America and the Caribbean) and *G. barbadense* L.(main distribution in South America and the Caribbean). Molecular diversity within and among the tetraploid cottons is limited (Small et al. 1999, Wendel & Cronn 2003), yet data indicate that *G. mustelinum* forms a basal branch in the phylogram of the tetraploids while the remaining four species forms two sister groups; one constituted by *G. hirsutum* and *G. tomentosum* and the other by *G. barbadense* and *G. darwinii* (Small et al. 1998, Liu et al. 2001).

The elongated epidermal seed trichomes of *G. herbaceum*, *G. arboreum*, *G. hirsutum* and *G. barbadense* were recognized by humans as useful spinnable fibers and led to four independent domestication events (Wendel 1995, Brubaker et al. 1999). Archaeological evidence of cotton remains in association with human settlements dates back to the 6th millennium BC for the Greater Indus area in the Old World (Moulherat et al. 2002), and to layers dated to 6400 – 5000 years BP for the Zaña river valley in Peru in the New World (Rossen et al. 1996, Piperno & Pearsall 1998). The cotton *G. hirsutum* has received much attention in this matter due to its prime economic importance in modern cotton production (e.g. Brubaker & Wendel 1994). The South American *G. barbadense* domestication has also been studied, but mainly by archaeological approaches. Cotton remains along the coast of

Peru and Ecuador are abundant and show a gradient in traits from wild-to-domesticated (Stephens & Moseley 1973; 1974, Stephens 1973). Purely wild *G. barbadense* are mentioned in the literature (e.g. Percy & Wendel 1990), but are not clearly defined in terms of fitness related traits for wild survival. The wild-to-domesticated continuum is according to Percy & Wendel (1990) delimited in four general categories in gene bank collections: (i) wild; (ii) 'dooryard cottons' or 'commensals', meaning single plants found near habitations thought to be derived directly from local wild populations; (iii) landraces, and finally; (iv) improved modern cultivars. The maritime subsistence for the Andean civilizations depending in part on cotton fishing-nets has led to the perception that the *G. barbadense* domestication has taken place along the coastline, but cotton is also found in more inland located sites like Caral (Solis et al. 2001). An allozyme study (Percy & Wendel 1990) yielded geographic clusters largely congruent with this maritime based scenario, but only an unspecified geographic domestication region 'north-western South America west of the Andes' including Colombia, Ecuador, Peru and Bolivia, was stated.

Until now no DNA marker based study has assessed the intraspecific genetic diversity of *G. barbadense* with the objective to describe a geographic pattern that can help in understanding the pre-Columbian domestication events of this species. To this end we conducted a new field collection of dooryard/ feral cottons of *G. barbadense* in Peru and added material from the USDA Cotton Collection, to get a representative sample from its pre-Columbian range. We included the two wild tetraploid species *G. mustelinum* and *G. tomentosum* and the domesticated *G. hirsutum* to get an interspecific comparison at the polyploid species level. We also added the three diploid species involved in the polyploidization event to have a direct comparisons between the diploid and the derived polyploid cottons. We applied the widely used AFLP (amplified fragment length polymorphism, Vos et al. 1995) fingerprinting method which has been used both to pinpoint domestication events (Heun et al. 1997), and to reveal relationship among diploid and tetraploid cottons (Abdalla et al. 2001). Besides, our study provides knowledge concerning the presently grown genetic diversity of *G. barbadense* in the 'Andean' countries in general and Peru in particular, with potential value for conservation strategies.

Material and methods

The analyzed cotton accessions were obtained in two different ways: I) Seed of 77 accessions was obtained from the Cotton Collection of the USDA-ARS, College Station, USA; via Dr. E. Percival, who selected the accessions based upon our wishes on which area to cover (Table 1). II) New material from Peru was collected by the authors (OW, ZH); its distribution in Peru is shown in figure 1. A collecting permit was obtained from the Instituto Nacional de Recoursos Naturales (INRENA) to collect *Gossypium* spp. in the departments of Ancash, La Libertad, Lambayeque, Piura, Tumbes, Cajamarca, Amazonas, San Martin, Loreto, Ucayali and Huanuco. A Material Transfer Agreement (MTA) was signed to send collected material to Norway. Voucher specimens were deposited both at the Herbario Weberbauer at the Universidad Nacional Agraria La Molina and at the Herbarium Truxillense at the Universidad Nacional de Trujillo. Seed samples were deposited in the Instituto Nacional de Investigacion Agraria (INIA) Lima, Peru. Collecting data included passport information on each locality; longitude, latitude and altitude data were obtained using a Global Positioning System (GPS) receiver (Garmin, eTrex Summit) and Phenotypic observations was added (see electronic supplementary material). For the DNA extraction one small, fresh leaf was taken from each cotton plant and dried in sealed plastic bags containing dry silica gel. Seed was collected from those plants having mature bolls. Branches were cut off from a few representative plants as herbarium vouchers. In total 100 cotton accessions were collected in this way, five of these are *G. raimondii* and the rest are *G. barbadense*.

Table 1. *Gossypium* accessions from the USDA-ARS Cotton Collection, College Station, USA, included in this study.

Species	Inventory no.	Accession no.	Origin country	Locality
G. hirsutum	TEX 624	PI 154104	Mexico	Chiapas
	TEX 2335	PI 607661	Costa Rica	Playa Ostinal, Nicoya Peninsula
G. mustelinum	AD_4 7	P1 530741	Brazil	Unspecified
G. tomentosum	AD_3 17	PI 530723	Hawaii	Unspecified
G. raimondii	D_5 4	PI530901	Peru	Unspecified
	P 6	O.C.	Peru	Cajamarca S07°26 / W78°56
G. herbaceum	A_1 9	PI 175456	Turkey	Unspecified
	A_1 17	PI 408775	Afghanistan	Unspecified
	A_1 24	PI 408782	Uzbekistan	Unspecified
	A_1 137	PI 630012	China	Unspecified
G. arboreum	A_2 190	PI 615699	Myanmar	Unspecified
	A_2 192	PI 615701	Iran	Unspecified
	A_2 244	PI 615745	Pakistan	Unspecified
	A_2 248	PI 614759	Russia	Unspecified
	A_2 256	PI 615757	Thailand	Unspecified
G. barbadense	Gb 61	PI 528320	Ecuador	Manabi
	Gb 619	PI 608062	Ecuador	Manabi
	Gb 602	PI 528118	Ecuador	Los Rios
	Gb 618	PI 608061	Ecuador	Guayas
	Gb 672	PI 608108	Ecuador	Guayas
	Gb 763	PI 608155	Ecuador	Guayas
	Gb 307	PI 528155	Ecuador	Guayas
	Gb 308	PI 528156	Ecuador	Guayas
	Gb 315	PI 528163	Ecuador	Guayas
	Gb 353	PI 528201	Ecuador	El Oro
	Gb 354	PI 528202	Ecuador	Loja
	Gb 673	PI 608109	Ecuador	Loja
	Gb 674	PI 608110	Ecuador	Loja
	Gb 675	PI 608111	Ecuador	Loja
	Gb 395	PI 528240	Ecuador	Unspecified
	Gb 594	PI528110	Ecuador	Unspecified
	Gb 596	PI528112	Ecuador	Unspecified
	Gb 63	PI528322	Ecuador	Unspecified
	Gb 314	PI 528162	Ecuador	Unspecified
	Gb 352	GB 352	Ecuador	Unspecified
	Gb 624	PI 608067	Ecuador	Galapagos Islands
	Gb 627	PI 608070	Ecuador	Galapagos Islands
	Gb 625	PI 608062	Ecuador	Galapagos Islands
	Gb 667	PI 358963	Ecuador	Galapagos Islands
	Gb 414	PI 528259	Peru	Tumbes
	Gb 819	PI 435250	Peru	Piura, S05°30 / W80°30
	Gb 989	GB 989	Peru	Lambayeque
	Gb 540	PI 528056	Peru	La Libertad
	Gb 541	PI 528057	Peru	La Libertad
	Gb 758	PI 608151	Peru	Unspecified
	Gb 417	PI 528262	Peru	Unspecified
	Gb 199	PI 276476	Brazil	Saenz Pena, Chaco
	Gb 1048	PI 608347	Brazil	Federal district
	Gb 195	PI 528376	Brazil	Unspecified
	Gb 196	PI 528376	Brazil	Unspecified
	Gb 427	PI 528276	Brazil	Unspecified
	Gb 1029	PI 608345	Brazil	Unspecified
	Gb 1021	PI 543760	Bolivia	Pando, S11°00 / W68°45
	Gb 347	PI 528195	Bolivia	Unspecified
	Gb 706	PI 608132	Bolivia	Unspecified
	Gb 1012	PI 608340	Venezuela	Zulia
	Gb 371	PI 528218	Colombia	Unspecified
	Gb 388	PI 528232	Colombia	Unspecified
	Gb 523	PI 528039	Colombia	Unspecified
	Gb 380	PI 528227	Dominica	Unspecified
	Gb 275	PI 528133	Haiti	Unspecified
	Gb 480	PI 527996	Haiti	Unspecified
	Gb 216	PI 528378	U.S.	Hawaii
	Gb 444	PI 528287	U.S.	Hawaii
	Gb 451	PI 528288	U.S.	Hawaii

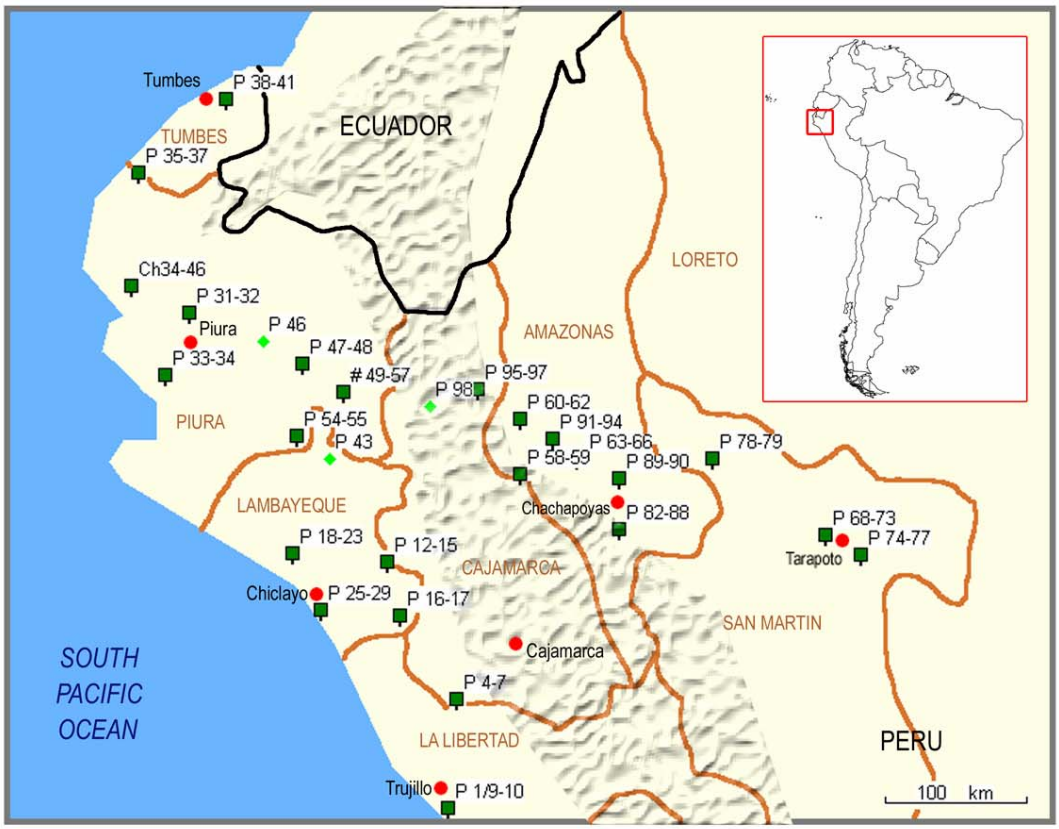


Figure 1. Map of Northern Peru displaying department boundaries and the Andean mountain chain in relation to collecting sites of new *Gossypium barbadense* and *G. raimondii* accessions. ◊ Indicates single accessions, □ indicates groups of closely located accessions.

DNA extraction

From those accessions with seed, three kernels were sown in separate pots under controlled conditions at the Agricultural University of Norway (NLH), Ås, Norway, as requested by the Norwegian authorities (import permit obtained from Landbrukstilsynet, Ås, Norway). Temperature was set at 25^{o}C, the day length at 14 hours, and pot-soil was kept constantly humid. Seeds germinated between three days and three weeks after sowing. One plant from every accession was used for DNA extraction by cutting cotyledons at the size of 2 to 6 cm^2. The cotyledons were cut just above the meristem and they were air-brushed to avoid possible impurities. The tissue was ground in liquid nitrogen and transferred to a -80^{o}C freezer. A total of 131 accessions representing seven *Gossypium* species were included in the final analysis; 65 samples obtained from the USDA-ARS gene bank and 67 from the material

collected in Peru. Five samples originate from silica gel dried material whereas the rest was obtained from freshly harvested tissue obtained from seed. DNA was extracted by using a DNeasy Plant Mini Kit from QIAGEN. The quality and quantity of the obtained undigested DNA was checked and quantified against undigested λ-DNA (Fermentas).

AFLP fingerprinting

AFLP fingerprinting was performed according to the original protocol of Vos et al. (1995); omitting the use of streptavidin beads for selecting *EcoRI* biotinylated DNA fragments. Briefly, 500ng DNA per sample was digested for 2 hours at 37^{o}C by 5U EcoRI (Fermentas) and 5U MseI (NEB) in 1x RL-buffer (10mM Tris acetate pH 7.5, 10mM Mg acetate, 50mM potassium acetate, 5mM DTT). Ligation of adapters fitting the cutting sites was done by adding T4 DNA ligase (Fermentas), 10 mM ATP, 10xRL-buffer, and by incubating the mixture for three hours at 37^{o}C. Thereafter the selective pre-amplification with primers complementary to adapter sequence with a one base extension was performed with E01 and M02. This primer combination was chosen based on the results of Abdalla et al. (2001) and Lacape et al. (2003). The PCR reactions (5.0 µl of the above mentioned DNA digestion/ ligation mix, 1.5µl of each +1 primer (75 ng), 1.0µl 10mM dNTP mix, 0.25µl *Taq* polymerase (5U/µl) (Fermentas), 5.0µl 10x PCR-buffer, 5.0 µl 25nM $MgCL_2$, in a total volume of 45 µl) were carried out on a PTC-200 thermocycler (MJ Research) with the temperature time profile given by Vos et al. (1995). The resulting pre-amplification products were diluted 1:19 prior to the selective +3/+3 amplification. Primer labeling was done by phosphorylating the 5'-end of the E- primers with radioactive γ^{33}P and T4 kinase (Fermentas). The selective amplification mix with +3/+3 primers (5µl of the diluted pre-amplified DNA, 0.5µl *E primer, 0.6µl M primer (see below for primer sequences), 0.4µl dNTP (10 mM), 2.0 µl PCR buffer (10x), 1.2µl $MgCL_2$ (25 mM), 0.1µl Taq polymerase (5U/µl), in a total volume of 20µl) was performed on a GeneAmp thermocycler (Applied Biosystems) according to Vos et al. (1995). Loading buffer (99% formamide, 10 mM EDTA (pH 8), 1.0 mg/ml xylene cyanol FF, 1.0 mg/ml bromophenol blue.) was added to the final PCR products (50%) and DNA was denatured by heat prior to running the samples (3µl) on 5% polyacrylamid (Acrylamid/ bis 19:1) gels under denaturing conditions. The electrophoresis apparatus (S2/ S2001, Gibco BRL Life Technologies) was filled with 2x TBE in the lower chamber and with 1x TBE in the upper chamber. For comparison 1.5 µl of a labeled 30-330 bp AFLP ladder from Invitrogen was loaded on the PAGE. The gels ran at 80 Watt for 90 min, thereafter fixed, dried and exposed to X-ray film (Kodac BioMax MR) for 8 to 48 hours, depending on

radiation intensity. X-ray films were developed with a AGFA Curix 60. The resulting images were scored by naked eye for the presence/ absence of AFLP bands using a light box. Overlapping samples as well as overlapping PCR reactions guarantied that only reproducible AFLPs were scored and the different PAGE's were well aligned. Based on a screening of 40 E01/M02 primer combinations with four samples (results not presented here) the following primer combinations were used to fingerprint all 131 cottons: E35/M48; E38/M48; E38/M49; E40/M50; E40/M59; E40/M60; E41/M47; E41/M60 (Invitrogen). Core sequence for *Eco*RI primers is 5'-GACTGCGTACCAATTCNNN-3', and 5'-GATGAGTCCTGAGTAANNN-3' for *Mse*I. The -3' end sequences for the primers used in this study are as follows: E01-A; E35-ACA; E38-ACT; E40-AGC; E41-AGG; M02-C; M47-CAA; M48-CAC; M49-CAG; M50-CAT; M59-CTA; M60-CTC.

Data analysis

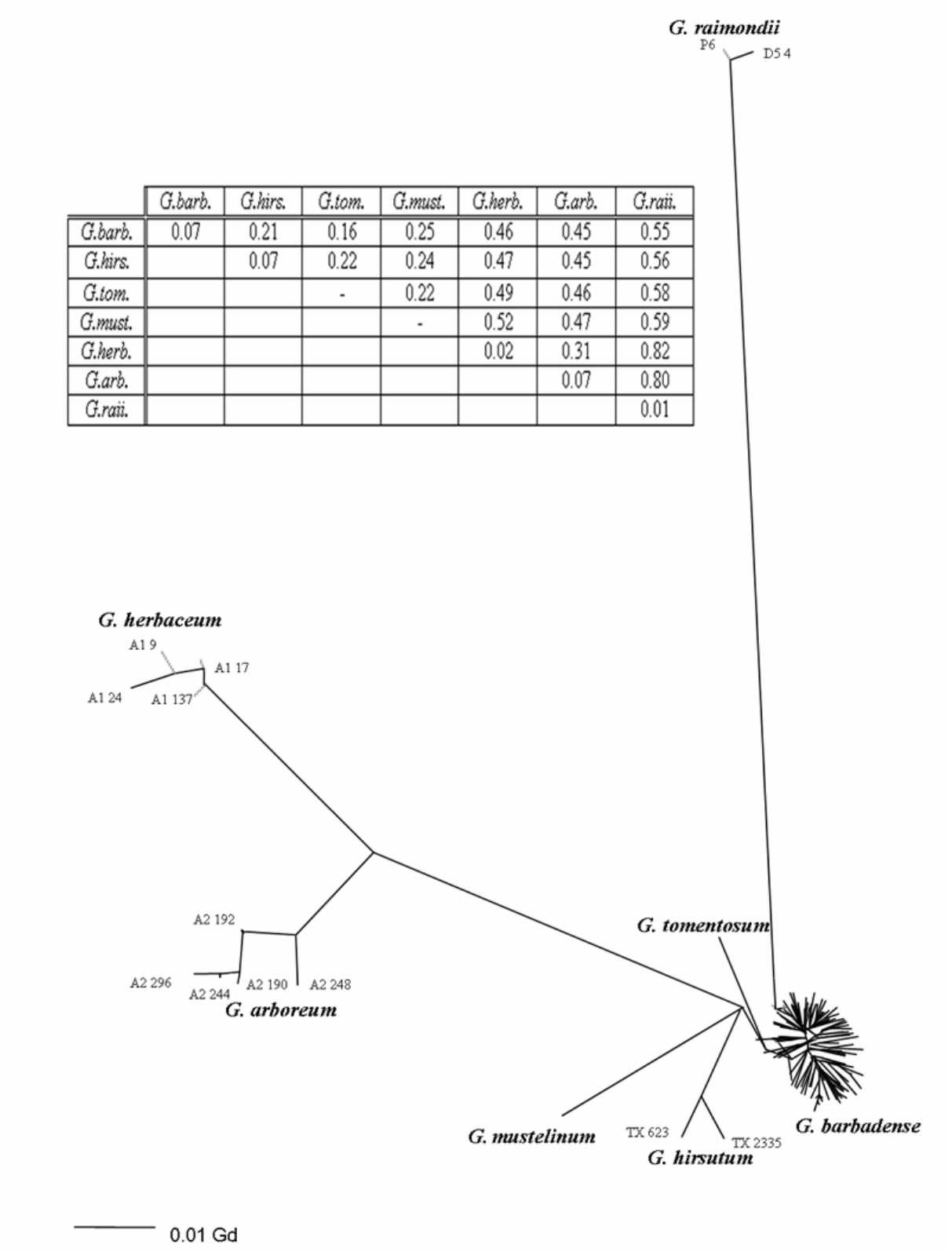
The X-ray films were scored in a binary manner by giving the presence of a polymorphic AFLP band a score of '1' and a '0' for its absence. Only unambiguous bands within an 80 to 800 bp range were scored. The data was assembled in a matrix and analyzed for pairwise genetic similarity (Gs) using the Dice similarity coefficient (Dice 1945, Nei & Li 1979) in the SIMQUAL option of NTSYS-pc v.2.11f (Rohlf 2000). The Dice similarity is given by the equation: $Gs_{ij} = 2a / (2a + b + c)$ where Gs_{ij} express the genetic similarity between line i and j, a is the number of bands present in both accessions, b is the number of bands present in i and absent in j, and c is the number of bands absent in i and present in j. Genetic distance (Gd) measures were obtained for all Gs coefficients by the equation: $Gd_{ij} = 1 - Gs_{ij}$. The Dice based Gd matrix was used for the neighbor joining (NJ) method of Saitou & Nei (1987). The Gs data were used for calculating unweighted pair-group method with arithmetic averages (UPGMA) (Sokal & Sneath 1963). The trees were calculated with the program NTSYS and PAUP (version4.0b10 for Macintosh, Swofford 1998). The robustness of the obtained trees was evaluated by comparing the different data analyses (NJ, UPGMA and principal coordinate analysis) and by bootstrapping (Felsenstein 1985) with 1000 re-samplings (PAUP).

Results

A total of 340 unambiguous, polymorphic bands were scored with eight AFLP primer combinations. 185 of these AFLP bands were polymorphic among the tetraploid species, and 93 were polymorphic within the final (see below) *G. barbadense* sample. The level of intraspecific variation in *G. barbadense* ranges from seven polymorphic bands out of a total of 44 (16%) in primer combination E38/M48 to 16 out of 60 (27%) in E35/M48. Figure 2 displays the NJ tree of the full data set. The intermediate position of the AD-tetraploids relative to the A- and D- diploids is well visible. *G. herbaceum, G. arboreum* and *G. raimondii* are well separated and so are the four tetraploids. The average genetic distances between species (see inlet of figure 2, which also contains the within species distances) range from 0.16 (*G. barbadense* vs. *G. tomentosum)* to 0.82 (*G. herbaceum* vs. *G. raimondii*).

To address the tree topology of the tetraploid species all diploid species (and eight tetraploid accessions that had missing data in more than one primer combination) were discarded from the data matrix. The resulting Dice similarity based UPGMA is shown in figure 3a. The robustness of the clustering was evaluated by bootstrapping. All species defining branches are well supported and equal those obtained in figure 2. The *G. barbadense* sub-cluster in figure 3a displays a first split distinguishing a first subset comprising 33 coastal Peruvian accessions. The remaining accessions from northernmost coastal Peru (the departments of Piura and Tumbes) and accessions from coastal Ecuador are clustered as basal to east of Andes accessions and accessions from Bolivia, Brazil, Columbia, Venezuela and the Caribbean- and Pacific-islands.

To further investigate these geographic patterns a *G. barbadense* NJ tree was computed on the basis of the 96 accessions displaying no missing data in any of the primer combinations (Fig. 3b). This NJ intraspecific analysis ‘zooms in’ on the topology of the *G. barbadense* cluster of figure 3a and displays the same general pattern in an unrooted way. The highest distance between two accessions of *G. barbadense* within this subset is found between the coastal Peruvian P1 and the Hawaiian Gb480 (0.13), while the average distance is limited to 0.06. A principal coordinate analysis (not shown, available as electronic supplementary material) revealed the same overall pattern for these 96 accessions and the two first principal coordinates accounted for 23.9% of the total variation.



	G.barb.	G.hirs.	G.tom.	G.must.	G.herb.	G.arb.	G.raii.
G.barb.	0.07	0.21	0.16	0.25	0.46	0.45	0.55
G.hirs.		0.07	0.22	0.24	0.47	0.45	0.56
G.tom.			-	0.22	0.49	0.46	0.58
G.must.				-	0.52	0.47	0.59
G.herb.					0.02	0.31	0.82
G.arb.						0.07	0.80
G.raii.							0.01

Figure 2. NJ of seven *Gossypium* species represented by 131 accessions calculated from pairwise distances (see bar for resolution) obtained with 340 AFLPs. Inlet table shows the average genetic distance between and within species.

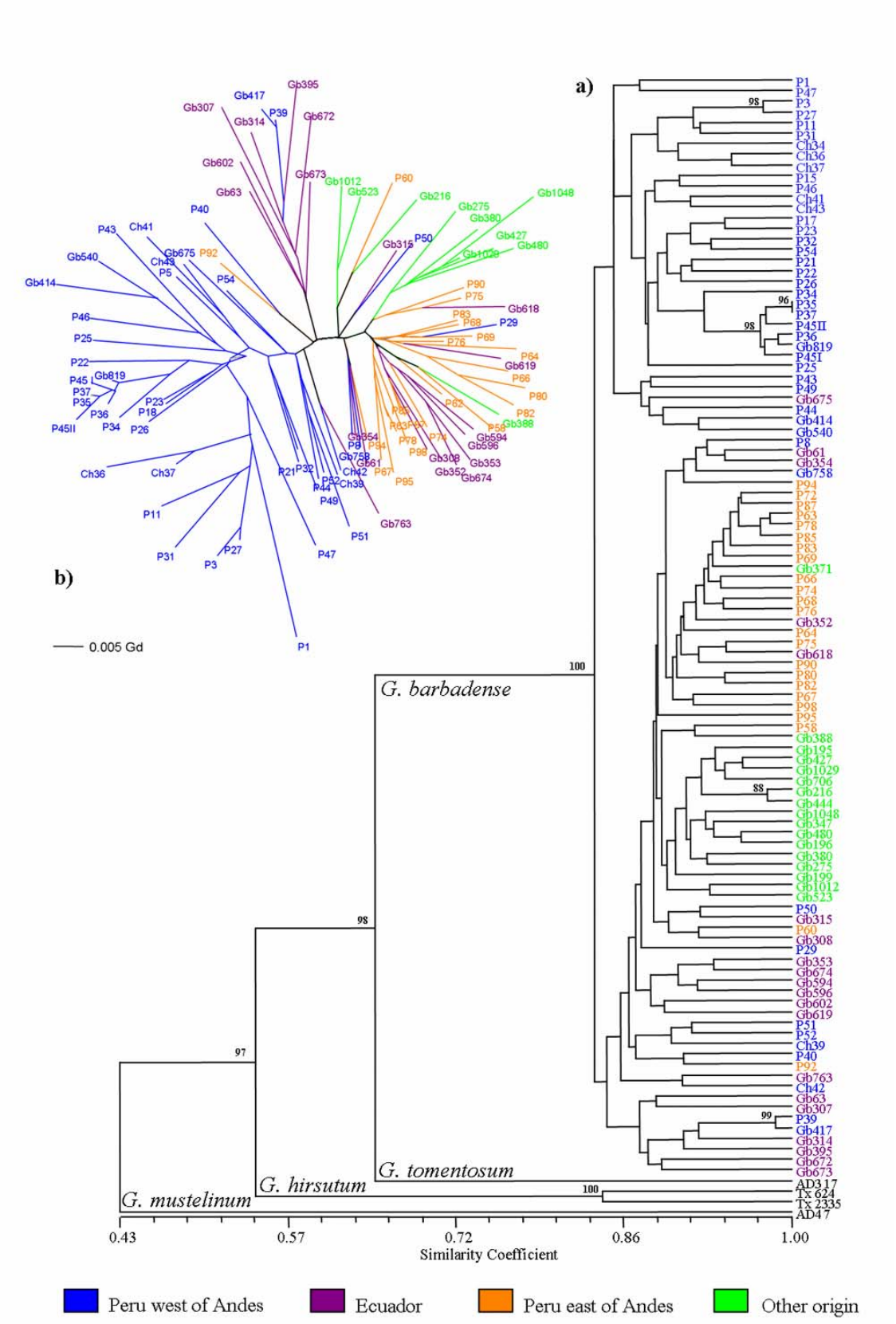


Figure 3. (a) UPGMA of 108 tetraploid cotton accessions obtained with 185 AFLPs. Species names and bootstrap values (>80) are indicated within the cluster. (b) NJ of 96 *G. barbadense* accessions with 93 AFLPs. Inventory no. (See table 1 and appendix 1) are colored according to their origin.

Discussion

Diploid vs. tetraploid relationships

The intermediate position of the four allotetraploid AD-cotton species relative to the A and the D diploids in the NJ analysis (fig. 2) is in full agreement with the established cytogenetic theory (Endrizzi et al. 1985). The genetic distances between the two different diploid species (A vs. D) and the derived tetraploids are large (0.45-0.59) compared to the distance among the tetraploids (0.16-0.25), as is expected by the bottleneck effect of a single polyploidization event (Cronn et al. 1999, Wendel & Cronn 2003). It is also apparent that *G. raimondii* is less related to the tetraploids than the two A- genome species (*G. arboreum* and *G. herbaceum*), again in agreement with cytogenetic and molecular studies (Endrizzi et al. 1985, Wendel 1989, Cronn et al. 1999, Small et al. 1999, Abdalla et al. 2001, Senchina et al. 2003). This consistency of AFLP analysis with cytogenetic studies is important as precaution is recommended for the use of AFLP markers in interspecific comparisons (e.g. El-Rabey et al. 2002), since homology of bands will be less well preserved the more evolutionary distant the analyzed material is. Abdalla et al. (2001) identify AFLP bands that are shared between a subset of the A diploids and tetraploids and call those 'A-related' markers, and do the same for identifying 'D related' markers. Applying this classification to our data, we count 136 A-related and 75 D-related markers which is a proportion similar to that obtained by Abdalla et al. (2001). These numbers have later been used by Wendel & Cronn (2003) in supporting the relative contribution of the two diploids for the emergence of the tetraploid cottons. Similar interspecific comparisons have shown that, e.g., *Aegilops* genomes are arranged by AFLP/NJ in accordance with known cytogenetics (Sasanuma et al 2004) and that also holds true for *Avena* (Drossou et al. 2004), *Solanum* (Kardolus et al. 1998), *Musa* (Ude et al. 2002) and *Cicer* (Sudupak et al. 2004). Even in El-Rabey et al. (2002), where the problem surrounding the homology assumption of co-migrating AFLP bands was studied, the AFLP clustering of *Hordeum* genomes was 'in full concordance with geographic origin, cytology, and taxonomic status'.

Interspecific relations among AD tetraploid cottons

Extensive work has studied the molecular phylogeny of *Gossypium* (Wendel 1989, Wendel & Albert 1992, Cronn et al. 1996, Seelanan et al. 1997, Small et al. 1998; 1999, Seelanan et al.1999, Small & Wendel 2000, Liu et al 2001, Cronn et al. 2002). Still the phylogenetic relationship among the tetraploids has remained elusive due to their short evolutionary separation and low levels of nucleotide diversity (Small et al. 1998, 1999). The

radiation of the tetraploids have only had an estimated 1.5 million years (Senchina et al. 2003), and the separation into 'good species' is confounded, as can be concluded from detection of gene flow across the taxonomic species borders (Percy & Wendel 1990, Wendel & Percy 1990, Brubaker & Wendel 1994) and from examples of testcrosses yielding fertile F_2 progenies (e.g. Hutchinson 1947, Endrizzi 1985). Nevertheless, Small et al. (1998) described two main branches within the tetraploid group; one branch leading to *G. mustelinum* and one branch leading to the remaining four species with *G. barbadense* and *G. darwinii* forming a sister group to *G. hirsutum* and *G. tomentosum*. Later studies have either not addressed the relationship between the tetraploids (Small et al. 2000, Cronn et al 2002), or failed to resolve the relative position of *G. tomentosum* within the group (Liu et al. 2001). Our data support the basal branching of *G. mustelinum*, but indicate a closer clustering of *G. tomentosum* towards *G. barbadense* than towards *G. hirsutum*. The difference in genetic distance is low (0.16 between *G. barbadense* and *G. tomentosum* vs. 0.22 between *G. hirsutum* and *G. tomentosum*) and can not be taken as a final evidence for a phylogenetic revision. Yet, Hutchinson et al. (1947) reported about chromosome pairing and testcross progeny studies among those tetraploids and concluded that *G. tomentosum* is closer to *G. barbadense* than to *G. hirsutum* as suggested by our data. Another interesting feature was observed when five *G. barbadense* accessions from the Galapagos Islands were included; two of the accessions (Gb624 / Gb625) clustered separate from all other *G. barbadense.* This could have been caused by the fact that some of the Galapagos accessions in the USDA-ARS Cotton Collection contain introgressions from *G. darwinii* growing in close proximity (Percival pers. com). Due to that uncertainty, we eliminated all Galapagos accessions from our analyses. The observed bias possibly caused by the indirect involvements of the fifth tetraploid cotton (i.e., *G. darwinii*) would make future analyses of defined Galapagos material (*G. barbadense* and *G. darwinii*) important.

Domestication and genetic diversity

G. barbadense is widely distributed, covering the whole range of tropical South America including an overlapping distribution with *G. hirsutum* in the northern part of the continent and in the Caribbean (Brubaker et al 1999, Wendel & Cronn 2003). Present-day indigenous *G. barbadense* is grown in gardens and peasants fields in northwestern South America and is referred to as 'dooryard' cotton or 'commensals'. The plants occur as perennial shrubs with thick basal stems, up to tree-four meters tall and they produce lint in an array of brown colors. The record concerning the native distribution of wild *G. barbadense* is

rather anecdotal, yet extant wild populations were reported from Guayas and Los Rios in Ecuador and Tumbes, Peru (Stephens & Moseley 1973; 1974, Schwendiman et al. 1985, Percival & Kohel 1990). The search for truly wild accessions is even more complicated since the wild-to-domesticate continuum in *G. barbadense* hardly allows categorical distinctions. Cotton remains from archeological excavation sites from central- and northern- coastal Peru show this wild-to-domesticated continuum for seed size, boll size, and fiber width. It is also proposed that a selection for greater differentiation between fuzz and lint led to morphs with strongly reduced fuzz layer called 'tufted' seeds, and the so-called 'kidney seeded' type of *G. barbadense*, as they were more easily ginned by hand (Turcotte & Percy 1990, Brubaker et al. 1999). Stephens (1975) found that the remains from the archaeological site Huaca Prieta (from about 2500 BC) display chocolate-colored fibers and that the oldest excavation layers only contain fuzzy seeds while tufted seeds appear in more recent layers. Selection also eliminated the hard seed coat and the delayed germination (Hutchinson 1947). In addition a low selection pressure was probably induced on traits like percent lint, lint length and strength, and eventually also on color differences for artistic purposes in weaving and improved catching abilities of fishing nets(Hutchinson et al 1947, Vreeland 1999). More specialized accessions have undergone selection for photoperiod neutrality, early maturing and fine white lint etc., but these later domestication traits were selected in colonial times and thereafter (Hutchinson et al. 1947, Brubaker et al.1999) and should not be considered here. In summary, for primitive cottons there exists no clear separation between wild and present-day 'dooryard' cottons. In fact, these dooryards are envisioned to be derived directly from local wild progenitors (Hutchinson 1947, Percy & Wendel 1990, Brubaker et al. 1999). Despite changing climate, intermixing etc., the geographic pattern of nowadays dooryards might provide an insight into the pre-Columbian distribution of this plant. Thus, we focused our re-sampling on dooryard cottons to get (in combination with gene bank material) a representative sample of *G. barbadense* across its assumed pre-Colombian range. The most striking observation we obtained is the unique diversity in the large majority of accessions from coastal Peru. A few accessions from NW Peru cluster with SW Ecuadorian accessions and together these accessions are basal to the remaining *G. barbadense*. This indicates that primitive domesticated *G. barbadense* has its domestication center in the NW Peru/ SW Ecuador as proposed (mainly derived from archaeological evidence) by Piperno and Pearsall (1998, page 164). Further, we also suggest, that cotton from this core area was transported across the Andes, and spread thereafter south into Bolivia, east into Brazil, and north into Colombia, Venezuela, and the Caribbean- and Pacific- islands. This scenario is proposed with

caution as the data do not yield high bootstrap-values on the most basal nodes of the *G. barbadense* cluster, but the pattern is evident in all the different analyses.

Our field observations (see appendix 1 of the supplement material and Honores et al. 2003) support the DNA data as they indicate a high diversity of 'primitive' types in the coastal departments of Peru. We encountered dooryard- and feral- plants with strongly arborescent growth habit, small bolls, and lint colors varying from reddish-brown to light yellow-brown, and purplish-brown to deep chocolate-brown. It was also apparent that the cottons east of the Andes displayed only a subset of this variation with generally larger bolls (long and slender) and lighter shades of brown to white lint. In the 'Selva Alto' of San Martin, a region where the Amazonian rainforest meets the Andes, plants in the next stage of domestication, i.e., landraces, are grown in small-scale commercial production of organic cotton. The domestication trait 'kidney-seeded' is known only from east of Andes (Turcotte & Percy 1990) and was present in some of our accessions from Bolivia and Brazil, indicating that our geographic interpretation of the data follows a spreading towards increased domestication level.

Conclusion

Our DNA analyses confirm the relationship among diploid and tetraploid cottons. They also confirm the archeology/ allozyme derived assumption of the South American cotton domestication having its origin in the coastal zone of NW Peru and SW Ecuador. We add that cotton spread out from this center (probably in the frame of domestication) over the Andes and from there into other parts of the continent. We also observed a tremendous diversity of *G. barbadense* found along the remaining northern Peruvian coast. This diversity is facing serious threats due to habitat destruction and the replacement of local types, and both *ex-situ* and *in-situ* conservation strategies should be implemented.

Acknowledgments

We thank Dr. E. Percival (USDA-ARS, Texas, USA) for providing seed material. In Peru, we thank J. Lazo, G. Arevalo, F. Balavarca and C. Basurto for assistance and information, K. Ramirez and M. R. Pastor at INRENA for obtaining the permission to collect and export the germplasm, M. White, C. Deza, E. Rodriguez and V. Medina for accompanying the collecting efforts in La Libertad., M. I. Olivos Farro and J. Escurra Puicon for collaboration in Lambayeque, P. M. Reyes More for information; E. Manco for access to PRONIRGEB germplasm and P. Azang Huaman for collaboration in San Martin. Thanks are also due to K. Vollan and J. H. Sønstebø for advices and technical assistance during laboratory work at AUN, Norway.

References

Abdalla AM, Reddy OUK, El-Zik KM, Pepper AE (2001) Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. Theor Appl Genet 102:222-229

Brubaker CL, Bourland FM, Wendel JF (1999) The origin and domestication of cotton. In: Smith CW, Cothren JT (eds) Cotton: origin history, technology, and production. 1st ed. New York, John Wiley & Sons, pp 3-31

Brubaker CL, Wendel JF (1994) Reevaluating the Origin of Domesticated Cotton (*Gossypium hirsutum* Malvaceae) using Nuclear Restriction Fragment Length Polymorphicisms (RFLP). Am J Bot 81:1309-1326

Cronn RC, Small RL, Haselkorn T, Wendel JF (2002) Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. Am J Bot 89:707-725

Cronn RC, Small RL, Wendel JF (1999) Duplicated genes evolve independently after polyploid formation in cotton. Proc Natl Acad Sci USA 96:14406-14411

Cronn RC, Zhao XP, Paterson AH, Wendel JF (1996) Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. J Mol Evol 42:685-705

Dejoode DR, Wendel JF (1992) Genetic Diversity and Origin of the Hawaiian-Islands Cotton, *Gossypium tomentosum*. Am J Bot 79:1311-1319

Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297-308

Drossou A, Katsiotis A, Leggett JM, Loukas M, Tsakas S (2004) Genome and species relationships in genus *Avena* based on RAPD and AFLP molecular markers. Theor Appl Genet (in press)

El-Rabey HA, Badr A, Schafer-Pregl R, Martin W, Salamini F (2002) Speciation and species separation in *Hordeum* L. (Poaceae) resolved by discontinuous molecular markers. Plant Biology 4:567-575

Endrizzi JE, Turcotte EL, Kohel RJ (1985) Genetics, cytology, and evolution of Gossypium. Adv Genet 23:271-375

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-791

Fryxell PA (1992) A revised taxonomic interpretation of *Gossypium* L., (Malvacea). Rheedea 2:108-165

Heun M, SchaferPregl R, Klawan D, Castagna R, Accerbi M, Borghi B, Salamini F (1997) Site of einkorn wheat domestication identified by DNA fingerprinting. Science 278:1312-1314

Honores A.M.F, Rodriguez E.F.R., Westengen O.T. (2003) Biologia y Etnobotanica del Algodon nativo Peruano (*Gossypium barbadense* L., Malvacea) Arnaldoa 10:1-12

Hutchinson JB, Silow RA, Stephens SG (1947) The Evolution of *Gossypium* and the Differentiation of the Cultivated Cottons 1st edn. London, Oxford University Press.

Kardolus JP, van Eck HJ, van den Berg RG (1998) The potential of AFLPs in biosystematics: a first application in *Solanum* taxonomy (Solanaceae). Plant Syst Evol 210:87-103

Lacape JM, Nguyen TB, Thibivilliers S, Bojinov B, Courtois B, Cantrell RG, Burr B, Hau B (2003) A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* x *Gossypium barbadense* backcross population. Genome 46:612-626

Liu Q, Brubaker CL, Green AG, Marshall DR, Sharp PJ, Singh SP (2001) Evolution of the FAD2-1 fatty acid desaturase 5 ' UTR intron and the molecular systematics of *Gossypium* (Malvaceae). Am J Bot 88:92-102

Moulherat C, Tengberg M, Haquet J, Mille B (2002) First Evidence of Cotton at Neolithic Mehrgarh, Pakistan: Analysis of Mineralized Fibres from a Copper Bead. Journal of Archaeological Science 29:1393-1401

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

Percival AE, Kohel RJ (1990) Distribution, collection, and evaluation of *Gossypium*. Adv Agron 44:225-256

Percy RG, Wendel JF (1990) Allozyme Evidence for the Origin and Diversification of *Gossypium barbadense* L. Theor Appl Genet 79:529-542

Phillips LL (1964) Segregation in new allopolyploids of *Gossypium*. V. Multivalent formation in New World x Asiatic and New World x wild American Hexaploid Am J Bot 51:324-329

Phillips LL (1963) The cytogenetics of *Gossypium* and the origin of New World cottons. Evolution 17:460-469

Piperno DR, Pearsall DM (1998) The origins of agriculture in the lowland Neotropics. San Diego, Academic Press

Rohlf FJ (2000) Numerical taxonomy and multivariate analysis system ver. 2.11. Applied Biostatistics, New York

Rong JK, Abbey C, Bowers JE, Brubaker CL, Chang C, Chee PW, Delmonte TA, Ding XL, Garza JJ, Marler BS, Park CH, Pierce GJ, Rainey KM, Rastogi VK, Schulze SR, Trolinder NL, Wendel JF, Wilkins TA, Williams-Coplin TD, Wing RA, Wright RJ, Zhao XP, Zhu LH, Paterson AH (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166:389-417

Rossen J, Dillehay TD, Ugent D (1996) Ancient Cultigens or Modern Intrusions? Evaluating Plant Remains in an Andean Case Study. Journal of Archaeological Science 23:391-407

Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425

Sasanuma T, Chabane K, Endo TR, Valkoun J (2004) Characterization of genetic variation in and phylogenetic relationships among diploid *Aegilops* species by AFLP: incongruity of chloroplast and nuclear data. Theor Appl Genet 108:612-618

Schwendiman J, Ano G, Percival AE (1986) Cotton collecting in continental Ecuador and Galapagos Islands. FAO/IPBGR Plant Genetic Resources Newsletter 64:33-37

Seelanan T, Brubaker CL, Stewart JM, Craven LA, Wendel JF (1999) Molecular systematics of Australian Gossypium section Grandicalyx (Malvaceae). Syst Bot 24:183-208

Seelanan T, Schnabel A, Wendel JF (1997) Congruence and consensus in the cotton tribe (Malvaceae). Syst Bot 22:259-290

Senchina DS, Alvarez I, Cronn RC, Liu B, Rong JK, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF (2003) Rate variation among nuclear genes and the age of polyploidy in Gossypium. Mol Biol Evol 20:633-643

Small RL, Wendel JF (2000) Phylogeny, duplication, and intraspecific variation of Adh sequences in new world diploid cottons (*Gossypium* L., Malvaceae). Mol Phylogenet Evol 16:73-84

Small RL, Ryburn JA, Wendel JF (1999) Low levels of nucleotide diversity at homoeologous Adh loci in allotetraploid cotton (*Gossypium* L.). Mol Biol Evol 16:491-501

Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF (1998) The tortoise and the hare: Choosing between noncoding plastome and nuclear ADH sequences for phylogeny reconstruction in a recently diverged plant group. Am J Bot 85:1301-1315

Sokal RR, Sneath PHA (1963) Principles of numerical taxonomy. Freeman. San Fransisco

Solis RS, Haas J, Creamer W (2001) Dating caral, a preceramic site in the Supe Valley on the central coast of Peru. Science 292:723-726

Stephens SG (1944) Phenogenetic evidence for the amphidiploid origin of New World cottons. Nature 153:53-54

Stephens SG (1975) A Reexamination of the Cotton Remains from Huaca Prieta, North Coastal Peru. American Antiquity 40:406-419

Stephens SG, Moseley ME (1974) Early domesticated cottons from archaeological sites in central coastal Peru. American Antiquity 39:109-122

Stephens SG, Moseley EM (1973) Cotton Remains from Archeological Sites in Central Coastal Peru. Science 180:186-188

Stephens SG (1944) Phenogenetic Evidence for the Amphidiploid Origin of New World Cottons. Nature 153:3871-3872

Sudupak MA, Akkaya MS, Kence A (2004) Genetic relationships among perennial and annual *Cicer* species growing in Turkey assessed by AFLP fingerprinting. Theor Appl Genet 108:937-944

Swofford DL (1998) PAUP*: Phylogenetic analysis using parsimony (*and other methods) vers. 4.0b10 for Macintosh. Sinauer, Sunderland, Massachusetts

Turcotte EL, Percy RG (1990) Genetics of kidney seed in *Gossypium barbadense* L. Crop Sci 30:384-386

Ude G, Pillay M, Nwakanma D, Tenkouano A (2002) Analysis of genetic diversity and sectional relationships in *Musa* using AFLP markers. Theor Appl Genet 104:1239-1245

Vreeland JM (1999) The revival of colored cotton. Sci Am 280:112-118

Vos P, Hogers R, Bleeker M, Reijans M, Vandelee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP - a New Technique for DNA-Fingerprinting. Nucleic Acids Res 23:4407-4414

Wendel JF (1995) Cotton. In: Simmonds S, Smartt J (eds) Evolution of Crop Plants. 1st ed. London, Longman, pp 358-366

Wendel JF, Cronn RC (2003) Polyploidy and the evolutionary history of cotton. Adv Agron 78:139-186

Wendel JF, Rowley R, Stewart JM (1994) Genetic Diversity in and Phylogenetic-Relationships of the Brazilian Endemic Cotton, *Gossypium mustelinum* (Malvaceae). Plant Syst Evol 192:49-59

Wendel JF, Albert VA (1992) Phylogenetics of the Cotton Genus (Gossypium) - Character-State Weighted Parsimony Analysis of Chloroplast-DNA Restriction Site Data and its Systematic and Biogeographic Implications. Syst Bot 17:115-143

Wendel JF, Percy RG (1990) Allozyme Diversity and Introgression in the Galapagos-Islands Endemic *Gossypium darwinii* and its Relationship to Continental *Gossypium barbadense*. Biochem Syst Ecol 18:517-528

Wendel JF (1989) New World Tetraploid Cottons Contain Old-World Cytoplasm. Proc Natl Acad Sci USA 86:4132-41

6. ECONOMIC BOTANY

6.1 Genetic resources

Most major agricultural crops were domesticated in the present-day poorest regions of the world, while today's prosperous countries have offered few genetic resources in return. As Falcon and Fowler (2002) points out, the US agriculture is composed almost entirely of imported crop species while the regions contribution to the agricultural system of the world is constricted to cranberries and sunflowers. Needless to say the geneflow out of Norway has contributed even less to the world's bread basket. Peru is one of those countries on historic ground when it comes to crop domestication; potato, common bean, chili pepper, peanut, and cassava are just a few of the Peruvian plants that have shaped world's agriculture and nutrition substantially (Harlan 1992). In addition there is a wealth of other crops that was cultivated by the Incas, their ancestors, and their descendants (National Research Council 1989). Thus *Gossypium barbadense* cotton adds to a long list of crops with their original gene-pool in the poor remnants of the Inca Empire. It is a political question of great concern in leading forums of the world (FAO, WTO, CBD) how the intellectual property related to crops should be shared and rewarded, but the current legislations doubtlessly protects corporate patent rights better -and in many instances to the expense of- the contribution to crop domestication/improvement by third world farmers and their forefathers (Gepts 2004a). The plant genetic industry is heavily concentrated in a half-dozen major firms holding substantial numbers of key patents on germplasm (Falcon & Fowler 2002, Gepts 2004a), and it should come as no surprise that also cotton has its share of patents attached to it. The annual 20 billion $ cotton industry (Rong et al. 2004) offers great incitements for biotechnological research. For example; it is estimated that cotton production accounts for 23% of the insecticide- and 10% of the total pesticide- consumption in the world (Vreeland 1999), and obviously a reduced need for chemical input would have both economic and ecologic benefits. The susceptibility of a crop to pathogens and insect pests depends both on ecological and genetic factors; at the ecological level monocultures are more prone to devastating pests than more diverse production systems (Vasey 1992), but the ultimate resistance factor is the genetic makeup of the crop (Harlan 1992). A general rule of thumb is that the narrower the genetic base of a crop becomes the more vulnerable it gets; diversity is the prerequisite for evolution in a fluctuating environment (Tanksley & McCouch 1997). In this perspective new genes for pest resistance are needed to improve cotton production. With the advance of biotechnology plants can be genetically modified to incorporate genes from other organisms.

Cotton plants with toxin-genes from the bacteria *Bacillus thuringiensis*, so-called *Bt*-cotton, is developed, and in fact now dominates the Chinese cotton agriculture (Linnestad 2002). This is a controversial and technically demanding strategy for crop improvement, and the approach of classical breeding is still of more significance and does not contain the possible ecological and health risks of GMO. In classical breeding, genes of interest, for example resistance genes to pathogens, are sought among different lines of cultivars, or eventually among the wild relatives of the crop (eventually aided by molecular linkage maps), and then crossed into the line of interest via regular sexual propagation. The genetic diversity in *G. barbadense* in Peru observed during the collecting efforts, and later quantified by DNA-fingerprinting, represent such a genetic resource for breeding. These naturally pigmented cottons can grow in arid soils with high levels of salinity and boron toxicity, and they are considerably hardier towards different pests than commercial hybrids (Vreeland 1981; 1999, Reyes More 1999, Nordt et al. 2004). The colored cotton has attracted interest with regard to fiber strength (Price et al. 2001) and colorfastness (Oktem et al. 2003). And, besides being a potential treasury for development of more viable and resistant commercial white cotton, the naturally colored cotton is a resource in itself. The growing demand in western markets for environmentally friendly products has fueled the success of the 'Native Cotton Project' – a network of Peruvian small scale cotton farmers growing indigenous landraces with no input of agrochemicals, distributing their produce to the US and Europe (Azang pers com., Vreeland 1999). This kind of viable on farm or *in-situ* conservation of the *G. barbadense* genetic resources of Peru is positive contrast to the overall tendency of loss of agro-biodiversity following commercialization, the phenomenon known as 'genetic erosion'. A typical story of genetic erosion followed the growing dominance of all-white commercial cotton in Peru; the government issued laws and decrees that obliged farmers to cut down and destroy perennial, colored cottons as they where envisioned to 'contaminate' white varieties through cross-pollination and serve as vectors for harmful insects and diseases (Simpson et al. 1985). These laws where enforced from 1931 and adhered to 1990 and even aimed at eradication of the lintless endemic *G .raimondii* and the kapok tree, *Bombax discolor* (Vreeland 1999).

As the policy on genetic resources has moved from a 'common heritage' conception to the prevailing view of genes as something than can be sold and bought (Fowler 2002), the legislations on collecting and exporting germplasm became significantly stricter in Peru in the 1990s. The final material transfer agreement (MTA) that was signed for the Peruvian material included in this study is in agreement with the principles of the International Treaty on Plant

Genetic Recourses for Food and Agriculture about access and benefit sharing on genetic material, though cotton is not included in this treaty (FAO 2004). One of the statements made clear in the MTA is that the holder can not claim any form of intellectual property right, such as patents, on the genetic material exported. In this statement lays the reasons for the tightening of the Peruvian germplasm policy; profit driven bio-prospecting efforts of pharmaceutical and agricultural companies have extended the colonial gene and money flow from the south to the north. In the case of cotton, naturally pigmented cotton bred for a few generations from material obtained from genebank accessions of South American origin has been patented as 'Coyote' and 'Green', and gives the patent owner the legal right to exclude others from selling these varieties in the US (ETC-group 2004). The report from the ETC-group evoke the term 'biopiracy' when considering that the peasants of coastal Peru who has cultivated and maintained these cultivars some 5000 years now in theory can be denied growing them for sale in the northern part of their continent.

6.2 Ethno-botanical considerations

The subdicipline of archaeology that deals with human-plant interactions throughout history is called 'paleoethnobotany' (Pearsall 2000). The 'paleo' prefix has obvious relevance to domestication history, but also the more current meaning of the discipline 'ethnobotany' as the multidisciplinary study of the dynamic relationship between people and plants (Salick et al. 2003) deserves some considerations here. The cultural meaning and significance of plants adds a value beyond the mere economic when it comes to rationales for conserving genetic diversity.

When referring to 'cotton' in everyday speech the meaning is usually the cotton fibers themselves. Throughout this thesis 'cotton' has been used in the botanical way as the common English name of the *Gossypium* genus, but when looking at the etymological roots of the word the everyday use is of more relevance. The earliest word defined as meaning cotton is the Sanskrit 'karpasa-i', while it is rather the Arabic 'al-' or 'el-kutum' which served as the origin of as well the English 'cotton' as the Spanish 'algodon' (Percival & Kohel 1990). The establishment of these words in the European languages probably happened as a consequence of the Moorish introduction of cotton production and textiles to Spain (Federacion Nacional de Algodoneros 1990). The lingua franca of the Inca Empire was Quechua, an ancient language with origin in the high Andean plains around the capitol Cuzco. Quechua is still the

mother tongue among the indigenous people of the Andes, and especially among the elders there are still many who do not speak the language of the conquistadores. The Quechua word for cotton is ‘utcu’ or ‘uchto’, and one of the river valleys in the department of Amazonas where collections were done for this study had the name ‘Utcubamba’, meaning ‘the valley where cotton grows’. There also exist a nomenclature of common names distinguishing cottons according to color of the fibers, appearance of the plant, and the use of the plant. ‘Pacucho’ is the Quechua name of brown fibered cotton. ‘Sani utcu’ is Quetchua for *G. barbadense* with red stem, a cultivar type only encountered east of Andes during our collecting efforts. ‘Ratansho’ was the name used for those *G. barbadense* having seeds grown together in clusters, earlier in the thesis referred to as ‘kidney-seeded’. ‘Pardo’ is the Spanish collective term for colored cotton followed by numerous sub-classifications according to color and other properties: blanco nativo; pardo catildo; pardito; ozcura; maron; morado; algodon del pais; pardo paisano; pardo mestizo; pardo mezclado; suave; blanco del pais; yema de huevo; fifo; and crema. Distinctions’ telling that cotton is not merely cotton.

The most important use of *G. barbadense* is spinning and weaving of the fibers for textiles. As stated earlier it might have been for fishing nets and lines the first selection on different colors was induced; the darker shades were less visible to fish and thereby improved catches (Stephens and Moseley 1974, Vreeland 1999). Another early use of cotton was for mummification; in the Chancay Valley the bodies were filled with cotton that would absorb the fluids and aid the process of mummification (Vreeland 1981, Vreeland 1999). The use of naturally pigmented cotton for textiles in garments and artistic tapestry etc. is possibly equally ancient (Stephens and Moseley 1974, Moseley 1992), and is still widely practiced in Peru.

Figure 6.1 Tapestry (1000 AD) made of naturally pigmented *Gossypium barbadense* yarn depicting a cotton plant. (Picture from Vreeland, Scientific American, Vol 280, 1999).

Ethnobotanical studies by Vreeland (1981) found that not only is the naturally pigmented cotton used in a vigorous artisan textile craft, but also in the practice of traditional medicinal cures and in the exercise of prehispanic beliefs and religious rituals. In the herbarium at the University San Marcos in Lima I encountered several voucher specimens of *G. barbadense* that was collected in different ethnobotanical surveys. A voucher from the department of Loreto had the following information: 'Juice of bracts and green immature fruits used to cure ear-aches.' Two others, also from Amazonian regions: 'Infusion of leaves of 'algodon rojo' used as cure for colic, white cotton not used this way.'; 'Leaves boiled and decoction drunk against stomach ache.', and; 'Young leaf decoction drunk to treat colic and fright (susto).' Clearly these statements are at the brink of the acceptable following a scientific rationality, and when talking to people maintaining and harvesting the pigmented cottons it became apparent that indigenous Peruvian perceptions on health matters follows quite other categories than those of school medicine; at the medicine market in Chiclayo brown cotton was sold in small packets as remedy for the above mentioned 'susto', explained as some sort of a mild curse. The fibers are also used as a remedy against 'mal de ojo' (the 'evil eye'). According to Vreeland (1981) this is a psychosomatic condition affecting many children, and when I inquired on the actual symptoms of this condition I was explained that it was a general condition of apathy and depression. The 'Brujos' are people envisioned to possess

supernatural powers, and probably it is one of those it is referred to in this most spiritual ethnobotanical description on a San Marcos voucher specimen: 'Cotton burnt and person affected by the devil is surrounded in smoke and healers hands drawn over affected individual's body.' At the other end of the spectrum from superstition to school medicine I was given detailed information on the antibacterial effect of naturally pigmented cotton on 'mal de aire' the common name for the medical diagnose 'conjunctivitis'- a bacterial infection of the eye causing secretion in the tear ducts. Further, several other uses of *G. barbadense* is given in 'Vocabulario de los nombres vulgares de la flora Peruana' (Soukup 1970), e.g.: the ash of cotton fibers is used to dry clean wounds in the head (Ambo, Arequipa); the ash is used as remedy for spider bites (north); infusion of red cotton fibers drunk to cure heart pain (Cajamarca); decoction of roots used as diuretic treat (South); leaves used against hemorrhoids (Loreto). In summary, it is no exaggeration to state that the naturally pigmented cottons of Peru have a cultural significance far beyond that of its all-white cash crop relative.

Figure 6.2 Scene from the herb market in Cusco

7. CONCLUDING REMARKS AND FUTURE DIRECTIONS

The present study has revealed evidence on both natural- and human- induced evolution of *Gossypium*. The main goal of the study was to highlight the domestication history of the South American textile crop *G. barbadense*. Other species were included in the task as it has been a long lasting discussion whether the American AD-tetraploids containing an Old World lint bearing A-genome and a New World D-genome is the result of a natural transoceanic hybridization or originated as a human caused cross.

The obtained results are in agreement with the cytogenetic evidence on the relationship between the diploids and the tetraploids; the tetraploids cluster intermediate between the A and D diploids. The distance between the diploids and the tetraploids is large. I ascribe this distance to the fact that extant diploids only are the closest related descendants, and not the actual diploids that were involved in the hybridization, and to the bottleneck created by monophyletic polyploidization. The matter deserves further investigation by including also other diploid genome groups (eight such groups exists worldwide). A logical extension of these conclusions on genetic distance is to support the view that polyplyploidization is considerably more distant in time than suggested by a human transoceanic dispersal scenario (Prevailing view is a 1.5 mya polyploidization event, Senchina et al. 2003).

Among the four tetraploid species included in this study the AFLP data led to well separated groups in concordance with taxonomic classification, though our phenetic data cluster the Hawaiian wild species *G. tomentosum* closer to *G. barbadense* in contrast to prevailing phylogenetic theory which suggest a closer relationship of this species towards the Central American *G. hirsutum*. This discrepancy deserves further investigations (especially since the sequence data underpinning the phylogenetic theories are based on few nucleotide substitutions, Small et al.1999) and in addition the last tetraploid *G. darwinii* from Galapagos should be included.

Concerning the intraspecific analysis of the focus species *G. barbadense*, clear patterns of genetic diversity in relation to geographic origin of accessions are apparent: The diversity along the Peruvian coast is relatively distinct and unique compared to that found among the remaining accessions, and within the latter accessions from northernmost coastal Peru and SW Ecuador appears to be the source from where all accessions originating east of Andes and

throughout the pre-Columbian range are derived (see figure 7.1). By proposing NW Peru/ SW Ecuador as the area of expansion for cotton cultivation we are careful not to equalize this with the crops domestication center; the locally maintained cottons along the pacific coast of Peru and Ecuador are probably derived from many different wild stocks once sprawling in the river valleys and domestication was probably a byproduct gradually induced along the whole coastal zone. Future directions for the domestication studies on *G. barbadense* should include more sophisticated approaches like mapping and screening for polymorphisms in genes related to early domestication traits such as 'tufted' and 'kidney-seeded'. Further elaboration on domestication traits could also be done by comparison towards parallel traits and their genetic control in *G. hirsutum*. It would also be desirable to include more locality specific material from a broader range of the coastline than covered in this field collection for an extended analysis. Especially the shores flanking the Guayaquil gulf on the Ecuadorian side should be searched for its asserted populations of truly wild *G. barbadense*. Finally, in the cross disciplinary study of crop evolution, also other scientific sources not included here can potentially contribute, e.g. paleo-climatology: The driving force behind agricultural transitions certainly includes climatic factors, and eventual correlations with global and regional changes in vegetation and weather regimes should be investigated.

A last remark should be made on the conservation biological aspects. Further work on *G. barbadense* domestication relies on close cooperation with local experts, and it should be a prime goal to use the knowledge on distribution of present genetic diversity to contribute in conservation efforts. Genetic diversity in agriculture is a prerequisite for human prosperity, and hopefully a fairer share of this wealth will pass to the managers of this diversity; the peasants of Peru.

Figure 7.1 Map of South America depicting the dispersal scenario for *Gossypium barbadense* cotton derived in this study. Grey area contains unique genetic diversity while black area is the probable expansion region for spread of *G. barbadense* across the Andes and further throughout it pre-Columbian range.

LITERATURE CITED

Abdalla A. M., O. U. K. Reddy, K. M. El-Zik, and A. E. Pepper. 2001. Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. Theoretical and Applied Genetics **102**:222-229.

Avise J. C. 1994. Molecular Markers, Natural History and Evolution. Chapman & Hall, New York.

Badr A., K. Muller, R. Schafer-Pregl, H. El Rabey, S. Effgen, H. H. Ibrahim, C. Pozzi, W. Rohde, and F. Salamini. 2000. On the origin and domestication history of barley (*Hordeum vulgare*). Molecular biology and evolution **17**:499-510.

Becker J., P. Vos, M. Kuiper, F. Salamini, and M. Heun. 1995. Combined Mapping of AFLP and RFLP Markers in Barley. Molecular & General Genetics **249**:65-73.

Botstein D., R. L. White, M. Skolnick, and R. W. Davis. 1980. Construction of a Genetic-Linkage Map in Man using Restriction Fragment Length Polymorphisms. American Journal of Human Genetics **32**:314-331.

Brubaker C. L., F. M. Bourland, and J. F. Wendel. 1999. The origin and domestication of cotton. Pages 3-31 In C.W. Smith and J.T. Cothren, editors. Cotton: origin history, technology, and production, John Wiley & Sons, New York.

Brubaker C. L., J. F. Wendel. 1994. Reevaluating the Origin of Domesticated Cotton (*Gossypium hirsutum* Malvaceae) using Nuclear Restriction Fragment Length Polymorphisms (RFLPs). American Journal of Botany **81**:1309-1326.

Brubaker C. L., A. H. Paterson, and J. F. Wendel. 1999. Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome **42**:184-203.

Cronn R. C., R. L. Small, and J. F. Wendel. 1999. Duplicated genes evolve independently after polyploid formation in cotton. Proceedings of the National Academy of Sciences of the United States of America **96**:14406-14411.

Cronn R. C., R. L. Small, T. Haselkorn, and J. F. Wendel. 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. American Journal of Botany **89**:707-725.

Cronn R. C., X. P. Zhao, A. H. Paterson, and J. F. Wendel. 1996. Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. Journal of Molecular Evolution **42**:685-705.

Damp J. E., D. M. Pearsall. 1994. Early Cotton from Coastal Ecuador. Economic Botany **48**:163-165.

Davies N. 1997. The Ancient kingdoms of Peru, The Penguin Group, London.

Dejoode D. R., J. F. Wendel. 1992. Genetic Diversity and Origin of the Hawaiian-Islands Cotton, *Gossypium tomentosum*. American Journal of Botany **79**:1311-1319.

Diamond J. 1999. Guns, Germs, and Steel. The fates of human societies. W.W. Norton & Company, New York.

Dice L. R. 1945. Measures of the amount of ecologic association between species. Ecology **26**:297-308.

Dillehay T. D., J. Rossen, and P. J. Netherly. 1997. The Nanchoc Tradition: the beginnings of Andean Civilization. American Scientist **85**:46-55.

Dillehay T. D., J. Rossen, G. Maggard, K. Stackelbeck, and P. Netherly. 2003. Localization and possible social aggregation in the late Pleistocene and early Holocene on the north coast of Peru. Quarternary International **109-110**:3-11.

Dillehay T. D. 2003. Palaeoanthropology: Tracking the first Americans. Nature **425**:23-24.

Dillehay T. D. 1989. Monte-Verde. Science **245**:1436-1436.

Doebley J. 1989. Isozymic evidence and the evolution of crop plants. Pages 165-191 In D. E. Soltis and P. S. Soltis, editors. Isozymes in Plant Biology, Dioscorides press, Oregon

Drossou A., A. Katsiotis, J. M. Leggett, M. Loukas, and S. Tsakas. 2004. Genome and species relationships in genus *Avena* based on RAPD and AFLP molecular markers. Theoretical and Applied Genetics **108**: *on line first*

El-Rabey H. A., A. Badr, R. Schafer-Pregl, W. Martin, and F. Salamini. 2002. Speciation and species separation in *Hordeum* L. (Poaceae) resolved by discontinuous molecular markers. Plant Biology **4**:567-575.

Endrizzi J. E., E. L. Turcotte, and R. J. Kohel. 1985. Genetics, cytology, and evolution of *Gossypium*. Advances in Genetics **23**:271-375.

ETC-group; Action Group on Erosion, Technology and Concentration. 2004. BIO-PIRACY: The story of natural coloured cottons of the Americas. www.etcgroup.org (25.05.04)

FAO; Food and Agriculture Organization of the United Nations. 2004. www.fao.org (25.05.04)

Falcon W. P., C. Fowler. 2002. Carving up the commons - emergence of a new international regime for germplasm development and transfer. Food Policy **27**:197-222.

Federacion Nacional de Algodoneros. 1990. Bases Technicas para el Cultivo del Algodon en Colombia. Bogota.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783-791.

Fowler C. 2002. Sharing agriculture's genetic bounty. Science **297**:157-157.

Fryxell P. A. 1992. A revised taxonomic interpretation of *Gossypium* L., (Malvacea). Rheedea **2**:108-165.

Gepts P., R. Papa. 2002. Evolution during Domestication. Encyclopedia of life sciences: www.els.net (10.10.03.)

Gepts P. 2004a. Who owns biodiversity, and how should the owners be compensated? Plant physiology **134**:1295-1307.

Gepts P. 2004b. Presentation: The Sacred City of Caral. http://agronomy.ucdavis.edu/gepts (25.05.04)

Goebel T., M. R. Waters, and M. Dikova. 2003. The archaeology of Ushki Lake, Kamchatka, and the Pleistocene peopling of the Americas. Science **301**:501-505.

Haas J., W. Creamer. 2001. Amplifying importance of new research in Peru - Response. Science **294**:1652-1653.

Harlan J. R. 1975. Geographic Patterns of Variation in some Cultivated Plants. Journal of Heredity **66**:182-191.

Harlan J. R. 1971. Agricultural Origins - Centers and Noncenters. Science **174**:468-474.

Harlan J. R. 1995. The living fields: Our agricultural heritage. Cambridge University Press, Cambridge.

Harlan J. R. 1992. Crops & man. American Society of Agronomy: Crop Science Society of America, Madison, Wisconsin.

Heun M., R. SchaferPregl, D. Klawan, R. Castagna, M. Accerbi, B. Borghi, and F. Salamini. 1997. Site of Einkorn Wheat Domestication Identified by DNA Fingerprinting. Science **278**:1312-1314.

Heyerdahl T. 1968. Sjøveier til Polynesia. Gyldendal, Norway. Oslo.

Honores A. M. F., E. F. R. Rodriguez, and O. T. Westengen. 2003. Biologia y Etnobotanica del Algodon nativo Peruano (*Gossypium barbadense* L., Malvacea). Arnaldoa **10**(2):1-12.

Hutchinson J. B. 1951. Intra-specific differentiation in *Gossypium hirsutum*. Heredity **5**:161-193.

Hutchinson J. B., R. A. Silow, and S. G. Stephens. 1947. The Evolution of *Gossypium* and the Differentiation of the Cultivated Cottons Oxford University Press, London.

Jaccard P. 1908. Nouvelles recherches sur la distribution florale. Bull. Soc. Vaud. Sci. Nat. **44**:223-270.

Jones C. J., K. J. Edwards, S. Castaglione, M. O. Winfield, F. Sala, C. vandeWiel, G. Bredemeijer, B. Vosman, M. Matthes, A. Daly, R. Brettschneider, P. Bettini, M. Buiatti, E. Maestri, A. Malcevschi, N. Marmiroli, R. Aert, G. Volckaert, J. Rueda, R. Linacero, A. Vazquez, and A. Karp. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Molecular Breeding **3**:381-390.

Kardolus J. P., H. J. van Eck, and R. G. van den Berg. 1998. The potential of AFLPs in biosystematics: A first application in *Solanum* taxonomy (Solanaceae). Plant Systematics and Evolution **210**:87-103.

Lacape J. M., T. B. Nguyen, S. Thibivilliers, B. Bojinov, B. Courtois, R. G. Cantrell, B. Burr, and B. Hau. 2003. A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* x *Gossypium barbadense* backcross population. Genome **46**:612-626.

Lazo Alvarez J. 1991. Contribucion a l'etude de origin, la domesticacation et la dispersion de l'especie Gossypium barbadense L: Polymorphisme enzymatique, morphologique et agronomic des populations sauvages, subspontanees et cultives. Phd thesis Universite Paris Sud. Paris.

Linnestad C. 2002. Kina velger genmodifisert bomull.Genialt **2**:22-23.

Liu B., C. L. Brubaker, G. Mergeai, R. C. Cronn, and J. F. Wendel. 2001a. Polyploid formation in cotton is not accompanied by rapid genomic changes. Genome **44**:321-330.

Liu Q., C. L. Brubaker, A. G. Green, D. R. Marshall, P. J. Sharp, and S. P. Singh. 2001b. Evolution of the FAD2-1 fatty acid desaturase 5 ' UTR intron and the molecular systematics of Gossypium (Malvaceae). American Journal of Botany **88**:92-102.

Lynch T. F. 1980. Guitarrero Cave: early man in the Andes. Academic Press, New York.

Marshall E. 2001. Pre-Clovis sites fight for acceptance. Science **291**:2317-2317.

Matsuoka Y., Y. Vigouroux, M. M. Goodman, G. J. Sanchez, E. Buckler, and J. Doebley. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. Proceedings of the National Academy of Sciences of the United States of America **99**:6080-6084.

Meltzer D. J. 1997. Anthropology - Monte Verde and the Pleistocene peopling of the Americas. Science **276**:754-755.

Moseley M. E. 1975. Chan Chan - Andean Alternative of Preindustrial City. Science **187**:219-225.

Moseley M. E. 2001. The Incas and their ancestors: The archaeology of Peru. Thames and Hudson, London.

Moulherat C., M. Tengberg, J. Haquet, and B. Mille. 2002. First Evidence of Cotton at Neolithic Mehrgarh, Pakistan: Analysis of Mineralized Fibres from a Copper Bead. Journal of Archaeological Science **29**:1393-1401.

Mueller U. G., L. L. Wolfenbarger. 1999. AFLP genotyping and fingerprinting. Trends in Ecology & Evolution **14**:389-394.

National Research Council. 1989. Lost crops of the Incas: Little known plants of the Andes with promise for worldwide cultivation. National Academy Press, Washington, D.C.

Nei M., W. H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA **76**:5269-5273.

Nordt L., F. Hayashida, T. Hallmark, and C. Crawford. 2004. Late prehistoric soil fertility, irrigation management, and agricultural production in northwest coastal Peru. Geoarchaeology-an International Journal **19**:21-46.

Oktem T., A. Gurel, and H. Akdemir. 2003. The characteristic attributes and performance of naturally colored cotton. AATCC Review **3**:24-27.

Olsen K. M., B. A. Schaal. 1999. Evidence on the origin of cassava: Phylogeography of *Manihot esculenta.* Proceedings of the National Academy of Sciences of the United States of America **96**:5586-5591.

Ouborg N. J., Y. Piquot, and J. M. Van Groenendael. 1999. Population genetics, molecular markers and the study of dispersal in plants. Journal of Ecology **87**:551-568.

Ozkan H., A. Brandolini, R. Schafer-Pregl, and F. Salamini. 2002. AFLP analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast Turkey. Molecular biology and evolution **19**:1797-1801.

Parker P. G., A. A. Snow, M. D. Schug, G. C. Booton, and P. A. Fuerst. 1998. What molecules can tell us about populations: Choosing and using a molecular marker. Ecology **79**:361-382.

Pearsall D. M. 2000. Paleoethnobotany: A handbook of procedures. Academic Press, San Diego, Calif.

Percival A. E., R. J. Kohel. 1990. Distribution, collection, and evaluation of *Gossypium*. Advances in Agronomy **44**: 225-256

Percy R. G., J. F. Wendel. 1990. Allozyme Evidence for the Origin and Diversification of Gossypium-Barbadense L. Theoretical and Applied Genetics **79**:529-542.

Phillips L. L. 1964. Segregation in new allopolyploids of *Gossypium*. V. Multivalent formation in New World x Asiatic and New World x wild American Hexaploid American Journal of Botany. **51**:324-329.

Phillips L. L. 1963. The cytogenetics of *Gossypium* and the origin of New World cottons. Evolution **17**:460-469.

Piperno D. R., D. M. Pearsall. 1998. The origins of agriculture in the lowland Neotropics. Academic Press, San Diego.

Pozorski S., T. Pozorski. 1990. Reexamining the Critical Preceramic Ceramic Period Transition - New Data from Coastal Peru. American Anthropologist **92**:481-491.

Price T. D., A. B. Gebauer. 1995. Last hunters, First farmers. School of American Research Press Santa Fe, New Mexico.

Price J. B., X. L. Cui, T. A. Calamari, and R. G. McDaniel. 2001. Assessing the quality of four naturally colored cottons. Textile Research Journal **71**:993-999.

Quilter J., B. Ojeda, D. M. Pearsall, D. H. Sandweiss, J. G. Jones, and E. S. Wing. 1991. Subsistence Economy of El-Paraiso, an Early Peruvian Site. Science **251**:277-283.

Raymond J. S. 1980. The Maritime Foundations of Andean Civilization: A Reconsideration of the Evidence. American Antiquity **46**:806-821.

Reyes More P. M. 1999. El algodon nativo: Caracterizasion, Evaluacion y potencia lidales de utilizacion. Ponencia presentada en el forum taller "Evaluacion de los recursos fitigeneticos disponibles en la region Piura 18-19 Nov.1999:1-9.

Rohlf F.J. 2000. Numerical taxonomy and multivariate analysis system version 2.11. Applied Biostatistics, New York.

Rong J. K., C. Abbey, J. E. Bowers, C. L. Brubaker, C. Chang, P. W. Chee, T. A. Delmonte, X. L. Ding, J. J. Garza, B. S. Marler, C. H. Park, G. J. Pierce, K. M. Rainey, V. K. Rastogi, S. R. Schulze, N. L. Trolinder, J. F. Wendel, T. A. Wilkins, T. D. Williams-Coplin, R. A. Wing, R. J. Wright, X. P. Zhao, L. H. Zhu, and A. H. Paterson. 2004. A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (Gossypium). Genetics **166**:389-417.

Rossen J., T. D. Dillehay, and D. Ugent. 1996. Ancient Cultigens or Modern Intrusions?: Evaluating Plant Remains in an Andean Case Study. Journal of Archaeological Science **23**:391-407.

Salick J. 2003. Intelectual Imperatives in Ethnobiology. NSF Biocomplexity Workshop Report; www.econbot.org/home.html (25.05.04)

Saitou N., M. Nei. 1987. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. Molecular biology and evolution **4**:406-425.

Sandweiss D. H., M. E. Moseley. 2001. Amplifying importance of new research in Peru. Science **294**:1651-1652.

Sandweiss D. H., H. McInnis, R. L. Burger, A. Cano, B. Ojeda, R. Paredes, M. D. Sandweiss, and M. D. Glascock. 1998. Quebrada Jaguay: Early South American maritime adaptations. Science **281**:1830-1832.

Sasanuma T., K. Chabane, T. R. Endo, and J. Valkoun. 2004. Characterization of genetic variation in and phylogenetic relationships among diploid Aegilops species by AFLP: incongruity of chloroplast and nuclear data. Theoretical and Applied Genetics **108**:612-618.

Schwendiman J., G. Ano, and A. E. Percival. 1985. Cotton collecting in continental Ecuador and Galapagos Islands. FAO/IPBGR Plant Genetic Resources Newsletter No. **64**: 33-37

Seelanan T., A. Schnabel, and J. F. Wendel. 1997a. Congruence and consensus in the cotton tribe (Malvaceae). Systematic Botany **22**:259-290.

Seelanan T., A. Schnabel, and J. F. Wendel. 1997b. Congruence and consensus in the cotton tribe (Malvaceae). Systematic Botany **22**:259-290.

Seelanan T., C. L. Brubaker, J. M. Stewart, L. A. Craven, and J. F. Wendel. 1999. Molecular systematics of Australian Gossypium section Grandicalyx (Malvaceae). Systematic Botany **24**:183-208.

Senchina D. S., I. Alvarez, R. C. Cronn, B. Liu, J. K. Rong, R. D. Noyes, A. H. Paterson, R. A. Wing, T. A. Wilkins, and J. F. Wendel. 2003. Rate variation among nuclear genes and the age of polyploidy in Gossypium. Molecular biology and evolution **20**:633-643.

Simpson B. B., J. M. Vreeland, and R. Ferreyra. 1985. *Gossypium raimondii*: collections in northern Peru. FAO/IPBGR Plant Genetic Resources Newsletter No.**62**:17-20

Small R. L., J. F. Wendel. 2000. Phylogeny, duplication, and intraspecific variation of Adh sequences in new world diploid cottons (Gossypium L., Malvaceae). Molecular phylogenetics and evolution **16**:73-84.

Small R. L., J. A. Ryburn, and J. F. Wendel. 1999. Low levels of nucleotide diversity at homoeologous Adh loci in allotetraploid cotton (Gossypium L.). Molecular biology and evolution **16**:491-501.

Small R. L., J. A. Ryburn, R. C. Cronn, T. Seelanan, and J. F. Wendel. 1998. The tortoise and the hare: Choosing between noncoding plastome and nuclear ADH sequences for phylogeny reconstruction in a recently diverged plant group. American Journal of Botany **85**:1301-1315.

Smith B. D. 1995. The emergence of agriculture. Scientific American Library, New York.

Sokal R.R., and Sneath P.H.A.1963. Principles of numerical taxonomy. W. H. Freeman, San Francisco.

Solis R. S., J. Haas, and W. Creamer. 2001. Dating Caral, a Preceramic site in the Supe Valley on the Central Coast of Peru. Science **292**:723-726.

Soukup J. 1970. Vocabulario de los nombres vulgares de la flora Peruana. Collegio Salesiano, Lima.

Stebbins G. L. 1989. Introduction. Pages 1-5 In D. E. Soltis and P. S. Soltis, editors. Isozymes in plant biology, Dioscorides Press, Portland, Oregon.

Stephens S. G. 1976. The origin of Sea Island Cotton. Agricultural history **50**:391-399.

Stephens S. G. 1975. A Reexamination of the Cotton Remains from Huaca Prieta, North Coastal Peru. American Antiquity **40**:406-419.

Stephens S. G. 1944. Phenogenetic Evidence for the Amphidiploid Origin of New World Cottons. Nature **153**:3871-3872.

Stephens S. G., M. E. Moseley. 1974. Early domesticated cottons from archaeological sites in central coastal Peru. American Antiquity **39**:109-122.

Stephens S. G., E. M. Moseley. 1973. Cotton Remains from Archeological Sites in Central Coastal Peru. Science **180**:186-188.

Stephens S. G. 1976. Some observations on photoperiodism and the development of annual forms of domesticated cottons. Economic Botany **30**:409-418.

Sudupak M. A., M. S. Akkaya, and A. Kence. 2004. Genetic relationships among perennial and annual Cicer species growing in Turkey assessed by AFLP fingerprinting. Theoretical and Applied Genetics **108**:937-944.

Swofford DL (1998) PAUP*: Phylogenetic analysis using parsimony(*and other methods) version 4.0b10 for Macintosh. Sinauer, Sunderland, Massachusetts

Tanksley S. D., S. R. McCouch. 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. Science **277**:1063-1066.

Tautz D. 1989. Hypervariability of Simple Sequences as a General Source for Polymorphic DNA Markers. Nucleic acids research **17**:6463-6471.

Tsuji K., O. Ohnishi. 2000. Origin of cultivated Tatary buckwheat (*Fagopyrum tataricum* Gaertn.) revealed by RAPD analyses. Genetic Resources and Crop Evolution **47**:431-438.

Turcotte E. L., R. G. Percy. 1990. Genetics of kidney seed in *Gossypium barbadense* L. Crop Science **30**:384-386.

Ude G., M. Pillay, D. Nwakanma, and A. Tenkouano. 2002. Analysis of genetic diversity and sectional relationships in *Musa* using AFLP markers. Theoretical and Applied Genetics **104**:1239-1245.

Ugent D., S. Pozorski, and T. Pozorski. 1986. Archaeological Manioc (*Manihot*) from Coastal Peru. Economic botany **40**:78-102.

Ugent D., S. Pozorski, and T. Pozorski. 1984. New Evidence for Ancient Cultivation of *Canna edulis* in Peru. Economic Botany **38**:417-432.

Ugent D., S. Pozorski, and T. Pozorski. 1982. Archaeological Potato Tuber Remains from the Casma Valley of Peru. Economic botany **36**:182-192.

Vasey D. E. 1992. An ecological history of agriculture : 10,000 B.C.-A.D. 10,000. Iowa State University Press, Ames.

Vavilov N. I. 1992. Origin and geography of cultivated plants. Cambridge University Press, Cambridge.

Vos P., R. Hogers, M. Bleeker, M. Reijans, T. Vandelee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP - a New Technique for DNA-Fingerprinting. Nucleic acids research **23**:4407-4414.

Vreeland JM. Colored Cotton: Return of the Native; 1981.International Development Research Center Report nr **10** (2):4-5

Vreeland J. M. 1999. The Revival of Colored Cotton. Scientific American **280**:112-118.

Welsh J., M. Mcclelland. 1990. Fingerprinting Genomes using PCR with Arbitrary Primers. Nucleic acids research **18**:7213-7218.

Wendel J. F. 1995. Cotton. Pages 358-366 In S. Simmonds and J. Smartt, editors. Evolution of Crop Plants, Longman, London.

Wendel J. F. 1989. New World Tetraploid Cottons Contain Old World Cytoplasm. Proceedings of the National Academy of Sciences of the United States of America **86**:4132-4136.

Wendel J. F., R. C. Cronn. 2003. Polyploidy and the evolutionary history of cotton. Advances in Agronomy **78**:139-186.

Wendel J. F., V. A. Albert. 1992. Phylogenetics of the Cotton Genus (*Gossypium*) Character State Weighted Parsimony Analysis of Chloroplast DNA Restriction Site Data and its Systematic and Biogeographic Implications. Systematic Botany **17**:115-143.

Wendel J. F., R. G. Percy. 1990. Allozyme Diversity and Introgression in the Galapagos-Islands Endemic *Gossypium darwinii* and its Relationship to Continental *Gossypium barbadense.* Biochemical systematics and ecology **18**:517-528.

Wendel J. F., R. Rowley, and J. M. Stewart. 1994. Genetic Diversity in and Phylogenetic Relationships of the Brazilian Endemic Cotton, *Gossypium mustelinum* (Malvaceae). Plant Systematics and Evolution **192**:49-59.

Wendel J. F., C. L. Brubaker, and A. E. Percival. 1992. Genetic Diversity in *Gossypium hirsutum* and the Origin of Upland Cotton. American Journal of Botany **79:**1291-1310.

Wendel J. F., P. D. Olson, and J. M. Stewart. 1989. Genetic Diversity Introgression and Independent Domestication of Old World Cultivated Cottons. American Journal of Botany **76**:1795-1806.

Williams J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey. 1990. DNA Polymorphisms Amplified by Arbitrary Primers are Useful as Genetic-Markers. Nucleic acids research **18**:6531-6535.

Zohary D., and M. Hopf 2000. Domestication of plants in the Old World: The origin and spread of cultivated plants in West Asia, Europe and the Nile Valley. Oxford University Press, Oxford.

Inventory #	Accession #	Species	Locality	District	Province	Department	Latitude	Longitude	Elevation	Plant description (native color names)	Soil	Leaf	Seed	Herbarium
P1	OWMW 001	Gbb	Garden in Moche river valley, in vicinity of Huaca de Sol	Trujillo	Trujillo	La Libertad	S.08°08'29.5"	W.79°00'07.4"	38	3-4m. Blanco nativo. In ditch between garden and dirt road.	Sandy soil	Yes		
P2	OWVM 002	Gbb	Botanical garden at Universidad nacional de Trujillo	Trujillo	Trujillo	La Libertad	S.08°06'42.1"	W.79°02'17.7"	10	2-3m. Pardo. Shrub growth.	Dry sand soil	Yes	W-P	N-U
P3	OWVM 003	Gbb	Botanical garden at Universidad nacional de Trujillo	Trujillo	Trujillo	La Libertad	S.08°06'42.1"	W.79°02'17.7"	10	2-3m. Blanvo nativo.		yes	W-P	
P4	OWCD 004	Grii	La Playa	San Benito		Cajamarca	S.07°27'03.4"	W.78°56'50.1"	993	4-5 m. Arborescent growth	Sandy soil	yes		N-T-U
P5	OWCD 005	Grii	Primera Agua	San Benito		Cajamarca	S.07°26'21.1"	W.78°56'17.6"	1180	4-5 m. Arborescent growth	Sandy soil			
P6	OWCD 006	Grii	La Cueva	San Benito		Cajamarca	S.07°26'54.9"	W.78°56'40.6"	1079	4-5 m. Arborescent growth	Sandy soil			
P7	OWCD 007	Gbb	Pueblo Jaguey	Jaguey	Ascope	La Libertad	S.07°36'54.0"	W.78°56'28.5"	417	1m. Colour unknown. In center of this small remote village.	Sandy soil	Yes		
P8	OWVM 008	Gbb	Behind farm building in Moche river valle	Trujillo	Trujillo	La Libertad	S.08°08'28.7"	W.79°00'07.7"	32	3-4m. White lint.	Sandy soil	Yes	W-P	
P9	OWVM 009	Gbb	Garden in Moche river valley	Trujillo	Trujillo	La Libertad	S.08°08'32.2"	W.79°00'04.1"	39	2m. Pardo.	Sandy soil	Yes		
P10	OWVM 010	Gbb	Neighbouring garden to GPS 2(	Trujillo	Trujillo	La Libertad	S.08°08'32.2"	W.79°00'04.1"	39	50cm. Pardo	Sandy soil	Yes		
P11	OW 011	Gbb	Within area of Tucume pyramid site museum	Tucume	Lambayeque	Lambayeque	S.06°30'48.6"	W.79°50'51.6"	65	1,5-2m. Three different colours. Planted for display.	Sandy soil	Yes	W-P	N-U
P12	OWMO 012	Grii	Cerca Caserio Juana Rios (Camino a Jomoche)	Chongoyape	Lambayeque	Lambayeque	S.06°35´39.8"	W.79°24´13.3'	264	5-6m. Arborecent plants. Flowers white. Small populations	Sandy soil	yes		N-T-U
P13	OWMO 013	Grii	Cerca Caserio Juana Rios (Camino a Jomoche	Chongoyape	Lambayeque	Lambayeque	S.06°35´34.6"	W.79°24´15.1"	264					
P14	OWMO 014	Grii	Cerca Caserio Juana Rios (Camino a Jomoche	Chongoyape	Lambayeque	Lambayeque	S.06°35´33.2"	W.79°24´13.4"	264					
P15	OWMO 015	Gbb	By road from Chiclayo to Chongoyape		Chiclayo	Lambayeque	S.06°41'20.2"	W.79°27'38.2"	183	2m. Blanco. Shrub.	sandy silt	Yes		
P16	OWMO 016	Gbb	Caserio Las Canteros	Patapo	Chiclayo	Lambayeque	S.06°44'22.9"	W.79°32'04.9"	148	1.5m. Maron (brown). Shrub	Sandy soil	Yes		N-U
P17	OWMO 017	Gbb	Road from Chiclayo to Chongoyape		Chiclayo	Lambayeque	S.06°32'04.9"	W.79°35'38.1"	138	2m. Semispontaneous along roac	Sandy soil	yes		
P18	OWMO 018	Gbb	Caserio Pampas	Morrope	Lambayeque	Lambayeque	S.06°31'36.5"	W.79°59'48.7"	34	2m. Pardo (dark) Shrub. In cluster with the two following samples. Outskirt of remote peasants Chacra	Sand	Yes		
P19	OWMO 019	Gbb	Caserio Pampas	Morrope	Lambayeque	Lambayeque	S.06°31'36.5"	W.79°59'48.7"	34	2m. Pardo (light) Shrub		yes		N-U
P20	OWMO 020	Gbb	Caserio Pampas	Morrope	Lambayeque	Lambayeque	S.06°31'36.5"	W.79°59'48.7"	34	2m. Blanco nativo		yes		
P21	OWMO 021	Gbb	Along dirtroad northeast of Morrope	Morrope	Lambayeque	Lambayeque	S.06°31'31.0"	W.79°59'34.1"	38	2m. Semispontaneous along road.Morrope IV	Sand	yes		
P22	OWMO 022	Gbb	Along dirtroad northeast of Morrope	Morrope	Lambayeque	Lambayeque	S.06°31'31.0"	W.79°59'34.1"		2m. Semispontaneous. Morrope V		yes		
P23	OWMO 023	Gbb	Along dirtroad northeast of Morrope	Morrope	Lambayeque	Lambayeque	S.06°30'26.9"	W.79°59'13.0"	44	3m. Large plant covering some 15m2. Intense pardo/maron	Sand	Yes	W-P	N-U
P24	OWVMER 024	Gbb	Along road between Roma and Ascope in Chicama r.v.		Ascope	La Libertad	S.07°44'31.2"	W.79°07'21.1"	200	2m. By sugarcane fields and small creek.	Sandy soil	yes		N-U
P25	OWMO 025	Gbb	In dooryard of Senora Batasca	Chiclayo	Chiclayo	Lambayeque	S.06°44'17.1"	W.79°50'53.5"	48	2-3m. Treelike growth	Saltsaturated soil	yes	W-P	
P26	OW 026	Gbb	El Cortigo, field of victorino Tullume Chanchafe	Monsefu	Monsefu	Lambayeque	S.06°52'26.7"	W.79°52'01.7"	22	2m. Pardo (light), big shrub plant	Fertile Soil	yes	W-P	N-T-U
P27	OW 027	Gbb	El Cortigo, field of victorino Tullume Chanchafe	Monsefu	Monsefu	Lambayeque	S.06°52'26.7"	W.79°52'01.7"	22	2m. Pardo (dark) Shrub.		yes		
P28	OW 028	Gbb	El Cortigo, field of victorino Tullume Chanchafe	Monsefu	Monsefu	Lambayeque	S.06°52'26.7"	W.79°52'01.7"	22	1m. Redish plant. Small capsuls, ca 1cm.		yes		
P29	OW 029	Gbb	El Cortigo, field of victorino Tullume Chanchafe	Monsefu	Monsefu	Lambayeque	S.06°52'26.7"	W.79°52'01.7"	22	2-3m Monopodial growth. Very small capsuls, <1cm		yes		
P30	OW 030	Gbb	El Cortigo, field of victorino Tullume Chanchafe	Monsefu	Monsefu	Lambayeque	S.06°52'26.7"	W.79°52'01.7"	22	2m. Pardo (light)			W-P	
P31	OW 031	Gbb	Caretera PanAm norte, 200m past bridge south of Piura		Piura	Piura	S.05°13'27.9"	W.80°39'20.1"	45	2-3m.Pardo.Treelike growth. Capsuls intermediate <3.5cm>	Sand	yes		
P32	OW 032	Gbb	Zona Industrial		Piura	Piura	S.05°11'48.2"	W.80°37'44.5"	8	3-4m Pardo. Trelike growth. Big capsuls>4cm.	Sand	yes		
P33	OW 033	Gbb	Coronado	Coronado	Piura	Piura	S.05°28'07.7"	W.80°43'55.7"	16	1-2m. Pardo.	Sandy silt			
P34	OW 034	Gbb	Santa Clara		Piura	Piura	S.05°29'35.3"	W.80°45'00.1"	22	3m. Pardo. In dooryard.	Sand/clay	yes		
P35	OW 035	Gbb	Mancora. Start of dirtroad to Cerro de Amotape	Mancora	Talara	Piura	S.04°06'03.5"	W.81°01'40.2"	14	3m. Pardo.Thick stem. Small capsuls<2.5cm By road.	Sandy clay	yes		N-T-U
P36	OW 036	Gbb	By dirtroad to Cerro Amotape	Mancora	Talara	Piura	S.04°06'51.1"	W.80°58'50.9"	45	3m. Pardo Capsuls and flowers small	Sand	yes		N-T-U
P37	OW 037	Gbb	By dirtroad to Cerro Amotape	Mancora	Talara	Piura	S.04°06'51.1"	W.80°58'50.9"	45	2m. Pardo.		yes		
P38	OW 038	Gbb	Casserio Cerro Blanco, Margen derecha rio Tumbes			Tumbes	S.03°38'28.3"	W.80°25'21.9"	30	1-2m. Blanco nativo	dry & sandy	yes		
P39	OW 039	Gbb	Casserio Cerro Blanco. Margen derecha rio Tumbes. Fields of Artemio Carasco Romero			Tumbes	S 03°38'10.2"	W.80°25'14.6''	33	2-3m. Pardo	Redish clay soil	yes		N-T-U
P40	OW 040	Gbb	Casserio Cerro Blanco. Margen derecha rio Tumbes. Fields of Artemio Carasco Romero			Tumbes	S 03°38'10.2"	W.80°25'14.6''	33	2-3m Blanco Nativo	Redish clay soil	yes		
P41	OW 041	Gbb	Casserio Cerro Blanco. Margen derecha rio Tumbes. Fields of Artemio Carasco Romero			Tumbes	S 03°38'10.2"	W.80°25'14.6''	33	2-3m. Mestizo	Redish clay soil	yes		
P42	OW 042	Gbb	Pueblo Arena		Paita	Piura	S.04°52'53.0"	W.81°01'28.4''	31	1m Pima	Clay soil close to rio Chira	yes		
P43	OWZH 043	Gbb	Parcela El Zapatero. Km 98 Chiclayo - Chulucanas. Field of Gerardo Dalmacio Sanchez	Olmos	Olmos	Lambayeque	S.05°55'08.2"	W.79°45'16.8''	167	2m Pardo Capsuls intermediate (2cm) Shrub aperance	sandy soil	yes		N-T-U
P44	OWZH 044	Gbb	Naupe La capilla. Km 188 Chiclayo - Chulucanas	Olmos	Olmos	Lambayeque	S.05°36'47.8"	W.79°53'26.4''	166	4-5m blanco nativo, "Algodón del pais", treelike growth	Sandy soil	yes		
P45	OWZH 045	Gbb	Naupe La capilla. Km 189 Chiclayo - Chulucanas	Olmos	Olmos	Lambayeque	S.05°36'39.2"	W.79°53'35.3''	174	4-5m Tick stems, small capsuls<1cm, leaves small with tick cuticule. Six plants in garden of abandoned house.	dry & sandy	yes	W-P	N-T-U
P46	OWZH 046	Gbb	El Cincuenta, about 50 m from crossroad to Chulucanas	Chulucanas	Morropon	Piura	S.05°10'15.4"	W.80°11'04.0"	87	2m Algodon del Pais	Sand	yes	W-P	
P47	OWZH 047	Gbb	Caserio Juan Velazco, 29 km from crossroad Panamericana to Canchaque. Field of Humberto Correa	Buenos Aires	Morropon	Piura	S.05°19'22.7"	W.79°55'52.0"	187	2m Pardo Paisano	clay/sand	Yes		

P48	OWZH	048	Gbb	Caserio Juan Velazco, 29 km from crossroad Panamericana to Canchaque. Field of Humberto Correa	Chulucanas	Morropon	Piura	S.05°19'22.7"	W.79°55'52.0"	187						
P49	OWZH	049	Gbb	Loma Larga Bajo. 66 km from crossroad Panamericana to Canchaque. Field of Luis Felipe Jibaja	San Miguel del Faique	Huancabamba	Piura	S.05°25'09.8"	W.79°40'29.0"	457	2m Pardo	sandy soil	yes	W-P	N-T-U	
P50	OWZH	050	Gbb	Loma Larga Bajo. 66 km from crossroad Panamericana to Canchaque. Field of Francisco Cania Travieso	San Miguel del Faique	Huancabamba	Piura	S.05°25'14.1"	W.79°40'21.5"	483	1-2m Pardo	sandy soil	yes	W-P		
P51	OWZH	051	Gbb	Loma Larga Alto. 68 km from crossroad Panamericana to Canchaque. Field of Luis Laban Choquehuanca	San Miguel del Faique	Huancabamba	Piura	S.05°24'59.2"	W.79°39'27.9"	612	2m Pardo, capsuls big >3cm	sandy soil	yes	W-P	N-T-U	
P52	OWZH	052	Gbb	Loma Larga Alto. 68 km from crossroad Panamericana to Canchaque. Field of Luis Laban Choquehuanca	San Miguel del Faique	Huancabamba	Piura	S.05°24'59.2"	W.79°39'27.9"	612	2m Blanco del Pais	sandy soil	yes			
P53	OWZH	053	Gbb	Loma Larga Alto. 68 km from crossroad Panamericana to Canchaque. Field of Luis Laban Choquehuanca	San Miguel del Faique	Huancabamba	Piura	S.05°24'59.2"	W.79°39'27.9"	612	1-2m Pardo catildo			W-P		
P54	OWZH	054	Gbb	Loma Larga Alto. 68 km from crossroad Panamericana to Canchaque. Field of Luis Laban Choquehuanca	San Miguel del Faique	Huancabamba	Piura	S.05°24'59.2"	W.79°39'27.9"	612	2m pardo color yema de huevo		yes	W-P		
P55	OWZH	055	Gbb	Villa Canchaque, 75 km from crossroad Panamericana to Canchaque. Quinta San Juan	Chanchaque	Huancabamba	Piura	S.05°24'59.2"	W.79°39'27.8"	928	1m nativo	fertile soil	yes			
P56	OWZH	056	Gbb	San Cristobal, 6 km road Canchaque to Huarmaca. Field of Geronimo Huaman	San Miguel del Faique	Huancabamba	Piura	S.05°24'31.3"	W.79°35'41.7"	1231	3m nativo, blomstring i sept.	sandy soil	yes			
P57	OWZH	057	Gbb	Chambelico, 20 km road Canchaque to Huarmaca. Field of Arturo Lizana Huancas	San Miguel del Faique	Huancabamba	Piura	S.05°28'14.1"	W.79°36'19.6"	1358	2-3m Pardo	sandy soil	yes			
P58	OWZH	058	Gbb	La Curva de Chiple. Km 146.5 from Olmos Corral Quemado to Moyobamba. Field of Etebrina Perez	Cayayu	Gutervo	Cajamarca	S.06°02'17.0"	W.78°54'49.4"	720	1-2m nativo, monopodial	gravel	yes			
P59	OWZH	059	Gbb	San Juan de Chiple. Km 147 from Olmos Corral Quemado to Moyobamba.	Cayayu	Gutervo	Cajamarca	S.06°02'15.6"	W.78°54'47.6"	707	3m nativo	fertile soil	yes			
P60	OWZH	060	Gbb	Chamaya Molino, Km 184 from Olmos Corral Quemado to Moyobamba.	Chamaya	Jaen	Amazonas	S.05°49'13.5"	W.78°44'35.0"	477	3-4m, monopodial, erect growth.Semispontaneous.	Gravel, sand	yes	W-P	N-T-U	
P61	OWZH	061	Gbb	Siempre Viva, 200 km from Olmos Corral Quemado to Moyobamba. In backyard	Milagro	Utcubamba	Amazonas	S.05°44'48.7"	W.78°39'51.32	444	3m (Only seeds)	Sandy		W-P		
P62	OWZH	062	Gbb	Cruz de Cayalti. Km 210 from Olmos Corral Quemado to Moyobamba. Field of Ines Pacheco.	Milagro	Utcubamba	Amazonas	S.05°42'40.5"	W.78°34'59.0"	536	4-5m (only seeds sampled)	Sandy	yes	W-P		
P63	OWZH	063	Gbb	La Caldera. Km 257 from Olmos Corral Quemado to Moyobamba. In crossroad to Jamalca. Field of Ruben Comeca	Jamalca	Utcubamba	Amazonas	S.05°51'26.22"	W.78°13'39.3"	581	2m Pardo, Big leaves, red stem	fertile soil	yes	W-P	N-T-U	
P64	OWZH	064	Gbb	La Caldera. Km 257 from Olmos Corral Quemado to Moyobamba. In crossroad to Jamalca. Field of	Jamalca	Utcubamba	Amazonas	S.05°51'26.22"	W.78°13'39.3"	581	2m Blanco del Pais, Big leaves, green stem		yes			
P65	OWZH	065	Gbb	Magunchal, 264 km from Olmos Corral Quemado to Moyobamba.	Jamalca	Utcubamba	Amazonas	S.05°53'23.3"	W.70°11'35.5"	645	4m. Aparently wild, blanco, big narrow capsuls.	Sandy gravel		W-P		
P66	OWZH	066	Gbb	Pache. 305 km from Olmos Corral Quemado to Tarapoto.	Pomachocos	Tunsilla	Amazonas	S.05°53'38.2"	W.77°58'25,6"	1607	4-5m, tall treelike,blanco, seeds collected	Sand		W-P		
P67	OWZH	067	Gbb	Tres de Mayo. 420 km from Olmos Corral Quemado to Tarapoto	Naranjos	Rioja	San Martin	S.05°44'47.9"	W.77°28'40,8"	981	1-2m, healthy plants in peasants Chacra	Humus		W-P		
P68	OWZH	068	Gbb	San Juan del Rio Mayo, 34 km from Tarapoto to Moyobamba. Field of Raul Sangama	Shanao	Lamas	San Martin	S.06°25'32.9"	W.76°35'37.3"	254	2m. Healthy plant in organic cotton field. Local name: Sani Utcu (Light brown). Stem red.	Clayish soil	yes	W-P	N-T	
P69	OWZH	069	Gbb	San Juan del Rio Mayo, 34 km from Tarapoto to Moyobamba. Field of Raul Sangama	Shanao	Lamas	San Martin	S.06°25'32.9"	W.76°35'37.3"	254	2m.Aspero Blanco. Outskirt of field.	Clayish soil				
P70	OWZH	070	Gbb	Shanao. House of Tedin Huaman	Shanao	Lamas	San Martin	S.06°24'39.6"	W.76°35'35.4"	269	1m. Pardo Pacucho or Pakutco. Small plants in backyard	Clayish soil	yes			
P71	OWZH	071	Gbb	Santa Rosa, San Miguel de Rio Mayc	Tabalasos	Lamas	San Martin	S.06°26'10.9"	W.76°35'14.1"	292	2m. Pardo. In organic cotton field	Clayish soil	yes			
P72	OWZH	072	Gbb	San Antonio de Rio Mayc	Lamas	Lamas	San Martin	S.06°25'19.4"	W.76°31'16.4"	770	2m. Nativo blanco. Capsuls big >3cm	Redish clay soil	yes			
P73	OWZH	073	Gbb	Lamas, Barrio Wayku by handcraft store	Lamas	Lamas	San Martin	S.06°25'19.4"	W.76°31'16.3"	772			yes			
P74	OWZH	074	Gbb	Morales. 610 km from Olmos Corral Quemado to Tarapoto.	Morales	San Martin	San Martin	S.06°27'57.3"	W.76°22'14.1"	312	3m. Aparently wild plant in roadside.Capsuls intermediate.	Clayish soil	yes	W-P		
P75	OWZH	075	Gbb	Tarapoto, Barrio Partido Alto, Sector Coperolta	Tarapoto	San Martin	San Martin	S.06°28'19.12"	W.76°22'14.1"	412	3-4m. Pardo. In dooryard	Clayish soil	yes	W-P		
P76	OWZH	076	Gbb	Tarapoto, Barrio Partido Alto, Sector Coperolta, field of Pompilio Azang Huaman	Tarapoto	San Martin	San Martin	S.06°28'00.0"	W.76°21'35.5"	503	Small plants, 3 months old. Pardo -organic cotton field	humus				
P77	OWZH	077	Gbb	Tarapoto, dirt road	Tarapoto	San Martin	San Martin	S.06°28'37.1"	W.76°21'56.6"	411	2m. Pardo.Capsuls small <2cm. Shrubby growth. Dooryard	sand	yes			
P78	OWZH	078	Gbb	Nueva Cajamarca. 459 km from Olmos Corral Quemado to Tarapoto.	Nueva Cajamarca	Rioja	San Martin	S.05°55'25.1"	W.77°18'57.3"	869	3m. Pardo. Red stem.	Clayish soil	yes	W-P	N-T-U	
P79	OWZH	079	Gbb	Caserio Amangang. 415 km from Olmos Corral Quemado to Tarapoto	Naranjos	Rioja	San Martin	S.05°41'56.9"	W.77°36'15.8"	941	Small plants, 3 months old. Pardo -organic cotton field	Clayish soil	yes			
P80	OWZH	080	Gbb	Carrera, 315 km from Olmos Corral Quemado to Tarapoto	Pomachoca	Bongara	Amazonas	S.05°52'16.0"	W.77°57'35.8"	1844	1-2m. Blanco nativo	Humus				
P81	OWZH	081	Gbb	Nuevo horizonte	Valera San Paulo	Bongara	Amazonas	S.06°02'07.0"	W.77°56'21.0"	1421	3m. Capsuls large <3cm.	Clayish soil	yes			
P82	OWZH	082	Gbb	San Fransisco. 23 km from Chachapoyas to Tingo.	Chachapoyas	Chachapoyas	Amazonas	S.06°17'05.9"	W.77°56'14.4"	1771	2m. Pardo	Sandy silt	yes	W-P		
P83	OWZH	083	Gbb	San Fransisco. 23 km from Chachapoyas to Tingo.	Chachapoyas	Chachapoyas	Amazonas	S.06°17'05.9"	W.77°56'14.4"	1771	2m.Blanco nativo	Sandy silt		W-P		

P84	OWZH	084	Gbb	Nuevo Tingo, on road to Kuelap	Luya	Chachapoyas	Amazonas	S.06°22'18.1"	W.77°54'36.6"	1973	2m.Blanco nativo. In garden	Sandy soil	yes		
P85	OWZH	085	Gbb	Tingo bajo	Luya	Chachapoyas	Amazonas	S.06°22'40.8"	W.77°54'22.4"	1836	2m.Blanco nativo. In garden	Sandy soil	yes		
P86	OWZH	086	Gbb	Tingo bajo	Luya	Chachapoyas	Amazonas	S.06°22'40.8"	W.77°54'22.4"	1836				W-P	
P87	OWZH	087	Gbb	Tingo bajo	Luya	Chachapoyas	Amazonas	S.06°22'40.8"	W.77°54'22.4"	1836				W-P	
P88	OWZH	088	Gbb	Tingo bajo	Luya	Chachapoyas	Amazonas	S.06°22'40.8"	W.77°54'22.4"	1836				W-P	
P89	OWZH	089	Gbb	Nueva Esperanza	Valera	Bongera	Amazonas	S.06°01'49.5"	W.77°56'37.1"	1369	3m. Blanco Nativo. Capsuls intermediate ><2cm. Planted close to house	sandy soil	yes	W-P	
P90	OWZH	090	Gbb	Corontachaca, 289 km from Olmos Corral Quemado to Tarapoto.	Jazan	Bongera	Amazonas	S.06°01'55.6"	W.77°56'35.5"	1215	Only seeds. Collected from chacra above village.			W-P	
P91	OWZH	091	Gbb	Nunaytemple, 245 km from Olmos Corral Quemado to Tarapoto	Bagua Grande	Utcubamba	Amazonas	S.05°48'59.4"	W.78°18'45.8"	528	1m. Blanco nativo	Sandy soil	yes		
P92	OWZH	092	Gbb	Puerto Motupe, 241 km from Olmos Corral Quemado to Tarapoto	Bagua Grande	Utcubamba	Amazonas	S.05°47'43.1"	W.78°20'07.3"	499	4-5m.Pardo. In dooryard.	sandy gravel		W-P	N-T-U
P93	OWZH	093	Gbb	La Lenia, road Bagua Grande - Cajaruro - Bagua Chica	Copia	Bagua	Amazonas	S.05°41'49.6"	W.78°28'30.5"	452	3-4m. Pardo. Capsuls small <2cm. Close to rice field. Semispontaneous.	Red clay		W-P	
P94	OWZH	094	Gbb	Cruce Alenya, on road Bagua Chica to Cajarur	Bagua Chica	Bagua	Amazonas	S.05°40'57.4"	W.78°29'00.1"	449	3m. Pardito. In garden	clay		W-P	
P95	OWZH	095	Gbb	Mochenta, 7 km from Chamaya to Jaen	Jaen	Jaen	Cajamarca	S.05°47'02.7"	W.78°46'59.4"	749	5m. Pardo. Semispontaneous. Close to road to Jaen.	sandy clay	yes	W-P	
P96	OWZH	096	Gbb	Chanango, 30 km from Jaen to San Ignasio	Bellavista	Jaen	Cajamarca	S.05°38'47.4"	W.78°44'43.9"	590	3m. Capsuls intermediate <3cm>. In garden.	clay	yes		
P97	OWZH	097	Gbb	San Augustin, 38 km from Jaen to San Ignasi	Bellavista	Jaen	Cajamarca	S.05°35'30.2"	W.78°47'09.5"	758	Only seeds				
P98	OWZH	098	Gbb	San Antonio, 105 km from Olmos Corral Quemado to Tarapoto	Pumahuaca	Jaen	Cajamarca	S.05°58'34.7"	W.79°12'15.8"	1009	2m.Blanco nativo. Capsuls intermediate, <3cm>	clay			
P99	OWZH	099	Gbb	Tupac Amaru, El Overal, 17.5 km from Olmos Corral Quemado to Tarapoto	Canaris	Olmos	Lambayeque	S.05°56'09.5"	W.79°34'45.8"	594	2m. Gnarled plant	sandy soil			
P100	OWZH	100	Gbb	El Palmo Los Boliches, 3 km km from Olmos Corral Quemado to Tarapoto	Olmos	Olmos	Lambayeque	S.05°58'40.7"	W.79°41'17.5"	287	Pardo. Only seeds.			W-P	
Ch 34-43	Vreeland J.		Gbb	Sampled from material maintained at research station of PRONIRGEB germplasm at the EEA El Porvenir, Tarapoto		Originally from Chira river valley, Piura									

Collectors

OW	Ola Westengen
OWMW	Ola Westengen - Michael White
OWVM	Ola Westengen - Victor Medina
OWCD	Ola Westengen - Carlos Deza
OWMO	Ola Westengen - Maria I. Olivos
OWVMER	Ola Westengen -Victor Medina - Eric Rodrigue
OWZH	Ola Westengen - Zosimo Huama

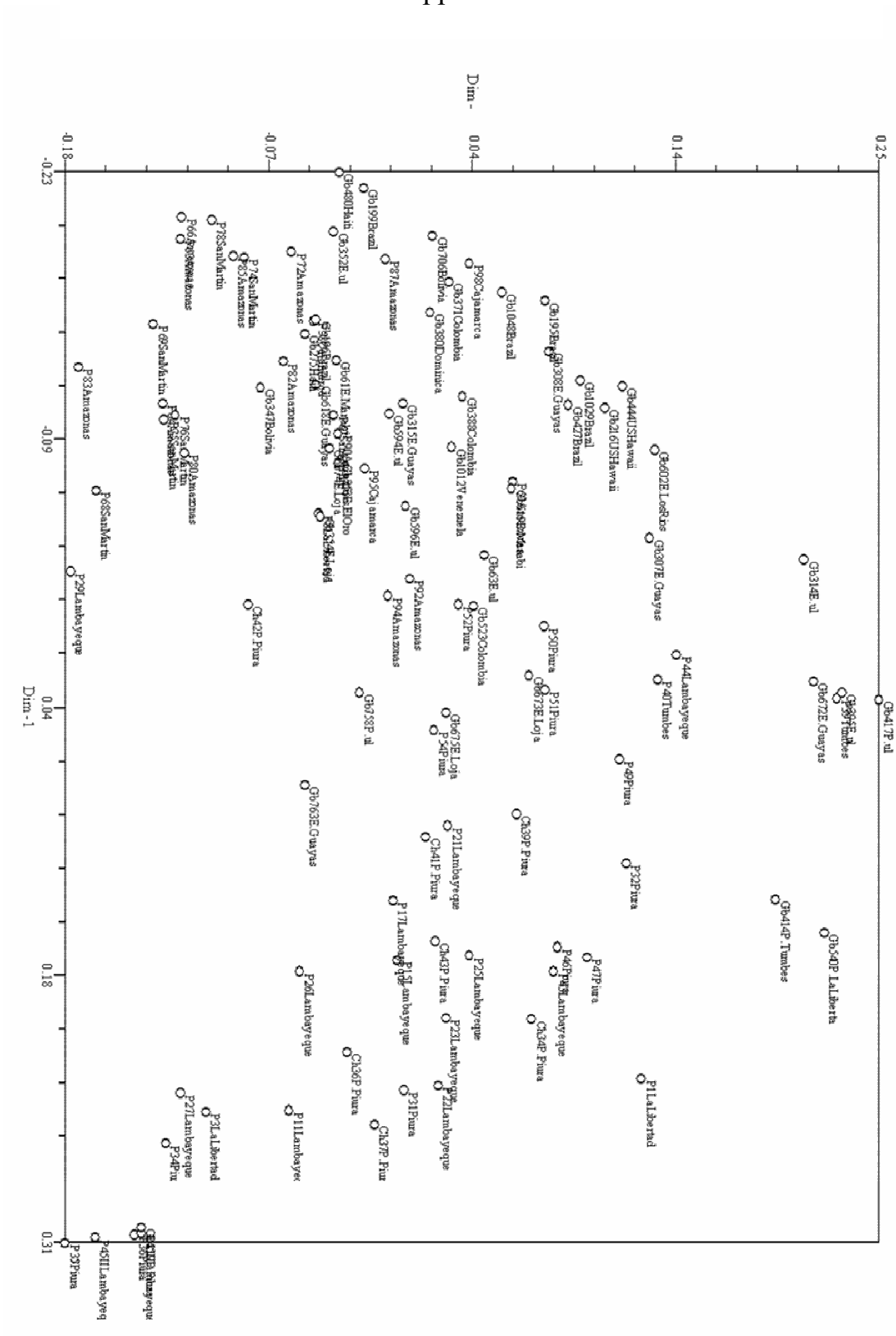
Species

Gbb	*Gossypium barbadense*
Gri	*Gossypium raimondii*

Herbarium

N	Agricultural University of Norway
T	Trujillo University
U	Agrarian University La Molina

Appendix 2



Principal Coordinate Analysis of 96 *Gossypium barbadense* accessions.
First two dimensions account for 23.9 % of the variation.