Gene Pools and the Genetic Architecture of Domesticated Cowpea


Abstract

Cowpea [Vigna unguiculata (L.) Walp.] is a major tropical legume crop grown in warm to hot areas throughout the world and especially important to the people of sub-Saharan Africa where the crop was domesticated. To date, relatively little is understood about its domestication origins and patterns of genetic variation. In this study, a worldwide collection of cowpea landraces and African ancestral wild cowpea was genotyped with more than 1200 single nucleotide polymorphism markers. Bayesian inference revealed the presence of two major gene pools in cultivated cowpea in Africa. Landraces from gene pool 1 are mostly distributed in western Africa while the majority of gene pool 2 are located in eastern Africa. Each gene pool is most closely related to wild cowpea in the same geographic region, indicating divergent domestication processes leading to the formation of two gene pools. The total genetic variation within landraces from countries outside Africa was slightly greater than within African landraces. Accessions from Asia and Europe were more related to those from western Africa while accessions from the Americas appeared more closely related to those from eastern Africa. This delineation of cowpea germplasm into groups of genetic relatedness will be valuable for guiding introgression efforts in breeding programs and for improving the efficiency of germplasm management.

Cowpea is the most important grain legume and fodder crop of the semiarid warm tropics and subtropics. Across wide swaths of sub-Saharan Africa and northeastern Brazil, in particular, cowpea is an important component of cereal and starchy tuber cropping systems because it supplies high protein grain and fodder while also helping to build the typically poor and fragile soils that predominate across much of these agro-ecologies (Ehlers and Hall, 1997). In contrast to many other important world crops, relatively little is understood about the domestication history, worldwide dispersal, and distribution of genetic variation of cowpea. Although domestication of cowpea was presumed to have occurred in Africa given the exclusive presence of wild cowpea in Africa (Steele, 1976), knowledge about the general region or regions of origin and number of domestication events within Africa is fragmented. Faris (1965) presented a review of earlier studies investigating the origin of cultivated cowpea and, along with his own extensive work involving morphological descriptors, suggested that there was evidence for a West or Central African center of domestication for cowpea. However, Coulibaly et al. (2002) provided evidence based on molecular markers that early domestication occurred in northeastern Africa; cowpea in these regions is more closely related to landraces from the Americas than from those from Africa.
Cowpea and sorghum are adapted to the same agro-ecologies and are often intercropped. Therefore, it is tempting to speculate that cowpea may have followed the same route out of Africa as sorghum, moving first from eastern Africa to the Arabian peninsula and then onto the Asian subcontinent (Faris, 1965; Pant et al., 1982) and to East Asia. Subsequently, cowpea may have moved westward to Europe through the Middle East because cowpea was known in southern Europe during Roman times (Tosti and Negri, 2002). It seems plausible also that cowpea first moved from western Africa to the New World with African people during the slave-trading period, but little or no documentation exists to support the extent of this movement. More recently, during the early 20th century, cowpea germplasm moved to the New World through purposeful informal and formal germplasm collecting and introduction activities conducted by the USDA, particularly from Central Asia as collectors sought germplasm of other major temperate-zone crops such as wheat (Triticum aestivum L.) and barley (Hordeum vulgare L.).

Hypotheses concerning the relationship of African germplasm to that found in other parts of the world where it has been introduced can be put on a sound footing with analysis of molecular markers. Fang et al. (2007) used amplified fragment length polymorphism markers to examine 15 landrace accesses of diverse origin and 72 advanced breeding lines and improved cultivars from four West African and two U.S. breeding programs. Their results showed that cowpea in Asia and North America did not share common genetic backgrounds with those from West Africa. However, that study used mostly breeding lines in which introgression of extraregional germplasm would have occurred, potentially obscuring more ancestral domestication relationships. In the present study, we used a much larger set of single nucleotide polymorphism (SNP) markers applied to a larger panel of cowpea landraces collected throughout Africa and in other cowpea-growing regions of Asia, Europe, North America, and South America (Fig. 1). To minimize the potential inclusion of admixed accessions in the sample set, only cowpea landraces that were entered into germplasm collections before 1975 are included because before this date there was very little international transfer of cowpea germplasm between breeding programs. Our study also includes a collection of African wild annual cowpea V. unguiculata subsp. dekindtiana (Harms) Verdc. from both East and West Africa. The subspecies dekindtiana has been documented as the likely progenitor of domesticated cowpea (Coulibaly et al., 2002; Pasquet, 1999). We aimed to examine the gene pool structure of African cowpea landraces and to determine their relatedness to African wild cowpea and non-African domesticated cowpea to clarify the origin and dispersal of this crop and help guide present-day and future breeding efforts.

Materials and Methods

Plant Materials and Genotyping

A total of 422 cowpea landraces collected from 56 countries were used in the present study. Major subdivisions included 323 landraces from North, West, Central, East, southeastern, and southern Africa and the other 99 landraces distributed throughout the rest of the world (Supplemental Table S1). Forty-six accessions of wild cowpea from three countries of West Africa and five countries of East, Southeast, and southern Africa (Supplemental Table S1) were obtained from the USDA germplasm collection in Griffin, GA. Genomic DNA from each line was isolated using Plant DNeasy (Qiagen) starting with 100 mg of young trifoliate leaves. The concentration of DNA was determined using Quant-itTM dsDNA Assay Kit Broad Range (Q33130) (Invitrogen) and fluorescence (excitation at 485 nm and emission at 535 nm for 1 s) measured using a microplate reader (Wallac Victor2 1420 Multilabel counter; PerkinElmer Life Sciences). The DNA concentration was adjusted to approximately 80 ng μL⁻¹ in a Tris-ethylenediaminetetraacetic acid buffer containing 10 mM Tris-hydrochloric acid and 1 mM ethylenediaminetetraacetic acid adjusted to a pH of 8.0 using sterile deionized water. Single nucleotide polymorphism genotyping with a 1536-SNP GoldenGate genotyping assay, as described in Muchero et al. (2009), was then performed at the University of California Los Angeles genotyping facility by Joe DeYoung and Marical Almonte.

Analyses of Gene Pool Structure

Informative markers were filtered based on their observed minor allele frequency, heterozygosity, and missing genotype calls. Population structure was analyzed using both the full set of these SNPs as well as a random set of SNPs at 3-cM intervals based on a cowpea consensus map (Lucas et al., 2011). An unpublished in-house program MAKE STRUCTURE was used to select different SNP sets based on predefined centimorgan distance and to convert the markers from ACGT calls into numerical genotypes. A Bayesian model-based clustering method was implemented in the software STRUCTURE 2.3.3 (Pritchard et al., 2000) under the admixture model with a burn-in period of 100,000 followed by 100,000 replications of Markov-Chain Monte Carlo. The number of clusters (K) was varied from 1 to 10, each including five independent runs. The web-based program STRUCTURE HARVESTER (Earl and von Holdt, 2012) was used to calculate the rate of change in the probability of data between successive K values (ΔK) to determine the optimum K value (i.e., the number of major gene pools) at which ΔK is highest (Evanno et al., 2005). The software CLUMPP (Jakobsson and Rosenberg, 2007) was used to align cluster assignment from independent runs using the in-files generated by STRUCTURE HARVESTER. Memberships of individuals...
assigned to specific gene pools were visualized using the unpublished in-house program “MARK IN MAP” and the software DISTURUCT (Rosenberg, 2004).

Analyses of Genetic Diversity
The geographical location from which each cowpea landrace was collected was noted. Analysis of molecular variance was performed with the software Arlequin 3.5 (Excoffier and Lischer, 2010) applied to all informative markers. Pairwise genetic distances between accessions were measured with the software GGT 2.0 (van Berloo, 2008) based on the allele-sharing method (Bowcock et al., 1994). Phylogenetic relationships were generated based on the genetic-distance matrix using the neighbor-joining method (Saitou and Nei, 1987) and visualized using the software MEGA 5.05 (Tamura et al., 2011).

Results
Single Nucleotide Polymorphism Diversity
Genotyping of 422 cowpea landraces with the 1536 SNPs showed that 1123 were polymorphic between the accessions (73%), 301 were monomorphic (20%), and 113 (7%) could not be called unambiguously and therefore were not used in any of the analyses (Supplemental Table S2). Genotyping of 46 wild cowpea accessions showed that 869 SNP markers (57%) were polymorphic, 554 were monomorphic (36%), and 113 (7%) could not be called unambiguously (Supplemental Table S3). Combining the two sets, 1133 markers were polymorphic (74%), 292 were monomorphic (19%), and 111 (7%) could not be called unambiguously. Of the 1133 polymorphic markers, 1051 have been mapped on 11 linkage groups representing the 11 cowpea chromosomes, based on a cowpea consensus genetic map (Lucas et al., 2011), while the other 82 SNPs remain unmapped (Supplemental Table S4). Linkage group 3 had the highest number of polymorphic markers. All possible SNP types were found in the world landrace and wild cowpea collection; the majority included A/G (or T/C) followed by A/C (or T/G), G/C, and A/T. In the landrace collection, heterozygosity at each polymorphic marker ranged from 0 to 6.6% (1.6% on average) except for four markers with more extreme heterozygosity (10, 15, 18, and 29%). In the wild cowpea, SNP heterozygosity at each polymorphic marker ranged from 0 to 47% (8% on average). In some cases, two or more accessions were found to have the same SNP genotypes at all loci except those with missing genotype calls (Supplemental Tables S2 and S3). For each of these duplicated sets, one accession with the smallest number of missing genotype calls was kept. Consequently, 397 landraces and 34 wild cowpea accessions with unique SNP genotypes were retained for further analyses.

Gene Pool Structure
Of the 1123 markers polymorphic in the landrace collection, 904 with minor allele frequency at least 0.05, less than 10% missing data, and less than 10% heterozygosity were used in population structure analysis. Clustering inference using all 904 SNPs showed that the rate of change in the probability of data between successive $K$ values ($\Delta K$) was highest at $K = 2$ ($\Delta K = 7088$) and a major decline in $\Delta K$ occurred at $K = 3$ (Supplemental Fig. S1). Clustering inference using a customized set of 195 SNPs excluding tightly linked markers (within 3 cM) showed that the rate of change in the probability of data between successive $K$ values ($\Delta K$) was also highest at $K = 2$ ($\Delta K = 1887$) followed by $K = 3$ ($\Delta K = 529$) and a major decline in $\Delta K$ occurred at $K = 4$ (Supplemental Fig. S1). The clustering inference indicated the existence of two major subpopulations (gene pools) in the world landrace population. Cluster assignment for each gene pool was highly consistent ($r = 0.99, P < 0.001$) between the full and
reduced SNP sets (Supplemental Fig. S1). Using a likelihood threshold of 0.7, 165 accessions (42%) were assigned to gene pool 1, 146 accessions (37%) were assigned to gene pool 2, and the other 86 accessions (21%) were intermediate (Fig. 2). The majority of accessions in gene pool 1 were from countries in West, North, and central Africa while the majority of accessions in gene pool 2 were from countries in East, southeast, and southern Africa (Fig. 3). In the “international set” outside Africa, 29 accessions (31%) were grouped in gene pool 1, 25 accessions (27%) were grouped in gene pool 2, and 40 accessions (43%) were intermediate. Applying K values from 3 to 7 introduced more subgroups in gene pool 1 while the majority of landraces in gene pool 2 were still grouped together (Supplemental Fig. S2); use of 195 SNPs at intervals of every 3 cM improved the clustering assignment by reducing fractional memberships of landraces assigned to specific subgroups.

Genetic Diversity among Cowpea Landraces

Analysis of molecular variance applied to all 904 informative SNP markers showed that the majority of genetic variance resided among landraces within countries (69%) while relatively small genetic variance (3%) existed between the African collection and the non-African international collection. Pairwise genetic distances based on allele sharing among 397 landraces varied from 0.01 to 0.72, with an average of 0.38 (Table 1). Landraces within East, southeast, and southern Africa were more variable relative to each other (average distance 0.34, ranging from 0.01 to 0.61) than accessions within West, North, and central Africa (average distance 0.31, ranging from 0.04 to 0.67). Landraces from Asia and Europe were most related
to those in West, North, and central Africa (average distances 0.38 and 0.35, respectively) while landraces from North America and South America were closer to those in East, southeast, and southern Africa (average distances 0.41 and 0.38, respectively). A neighbor-joining phylogenetic tree clearly showed a separation between African landraces from the west and the east while landraces in countries outside Africa clustered with the African accessions in most clades that were intermediate between the western and eastern gene pools (Supplemental Fig. S2).

Relatedness between African Landraces and Wild Cowpea

A total of 322 polymorphic SNPs with less than 10% missing scores and less than 10% heterozygosity in the wild cowpea collection and that were also informative in landraces were used for genetic comparison. Population structure analyses applied to this SNP set also confirmed the presence of two major gene pools in the world cowpea landrace collection (Supplemental Table S5). Pairwise genetic distances showed that landraces from West, North, and central Africa were closer to wild cowpea from the west (average distance 0.39) than to wild cowpea from the east (average distance 0.43). In contrast, African landraces from East, Southeast, and southern Africa were closer to wild cowpea from these regions (average distance 0.42) than to wild cowpea from the west (average distance 0.48). Phylogenetic analyses involving wild cowpea and “pure” representatives of two African gene pools (admixture score less than 0.01 based on structure analyses) showed that wild cowpea were clustered next to each other and there was a clear separation between wild accessions from West Africa and those from East, Southeast, and southern Africa, except for two accessions from West Africa (PI 632895 and PI 632896), which were clustered in the other clade (Fig. 4); wild cowpea from West Africa were clustered next to gene pool 1 while wild cowpea from East, Southeast, and southern Africa were clustered next to gene pool 2.

Discussion

Genetic Relationships among Domesticated and Wild Cowpea

Cowpea landraces used in this study were entered into germplasm collections more than 30 yr ago when there was very little international transfer of cowpea among regional breeding programs and thus little chance of admixture across large geographic regions among landraces. More than 10% of these materials came from the IITA cowpea “mini-core” collection, which was designed to capture the genetic variation present in the wider world collection (Mahalakshmi et al., 2007). Other major subsets of germplasm used in the present study included 48 landrace accessions from Mozambique in southeastern Africa and 35 landrace accessions from Angola in southwestern Africa, both areas of the continent that have lacked representation in previous cowpea diversity studies. Another major subset included 40 accessions from India, recognized as a secondary center of cowpea diversity (Faris, 1965; Pant et al., 1982). The remaining cowpea landraces were collected from 50 other countries in Africa and other continents where cowpea has been introduced. Thus, these genetic materials represent a comprehensive sample of diversity in domesticated cowpea. This was confirmed by results from genotyping using the Illumina GoldenGate Assay in which high levels of SNP diversity were observed among these materials (Supplemental Table S4). This SNP assay was developed based on a diverse discovery panel that included 15 domesticated cowpea accessions from different origins (Muchero et al., 2009) and therefore is expected not to bias gene pool assignment of the landraces in this study although it might provide greater sensitivity to detect variation within domesticated cowpea than the species as a whole.

Population structure analyses delineated the larger African landrace germplasm into two major gene pools. The two gene pools were distributed in two distinct geographical zones separated by the dense and vast rainforests of the Congo River basin (Fig. 3). This region is too wet and not

Table 1. Pairwise genetic distances measured based on allele sharing among the world cowpea landraces.

<table>
<thead>
<tr>
<th>Source (number of accessions)</th>
<th>Africa (West, North, and Central)</th>
<th>Africa (East, Southeast, and South)</th>
<th>Asia</th>
<th>Europe</th>
<th>North America</th>
<th>South America</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa (West, North, Central) (170)</td>
<td>0.31 (0.01–0.61)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa (East, Southeast, South) (133)</td>
<td>0.43 (0.03–0.71)</td>
<td>0.34 (0.05–0.67)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia (66)</td>
<td>0.38 (0.03–0.67)</td>
<td>0.42 (0.05–0.67)</td>
<td>0.36 (0.09–0.66)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe (13)</td>
<td>0.36 (0.11–0.66)</td>
<td>0.45 (0.19–0.72)</td>
<td>0.37 (0.09–0.62)</td>
<td>0.35 (0.07–0.57)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America (9)</td>
<td>0.41 (0.12–0.58)</td>
<td>0.35 (0.05–0.71)</td>
<td>0.39 (0.06–0.66)</td>
<td>0.41 (0.07–0.61)</td>
<td>0.33 (0.15–0.53)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South America (6)</td>
<td>0.38 (0.16–0.6)</td>
<td>0.37 (0.21–0.65)</td>
<td>0.39 (0.15–0.66)</td>
<td>0.40 (0.20–0.6)</td>
<td>0.35 (0.21–0.55)</td>
<td>0.38 (0.28–0.49)</td>
<td></td>
</tr>
<tr>
<td>Total (397)</td>
<td>0.36 (0.01–0.71)</td>
<td>0.40 (0.03–0.72)</td>
<td>0.37 (0.06–0.66)</td>
<td>0.39 (0.07–0.61)</td>
<td>0.34 (0.15–0.55)</td>
<td>0.38 (0.28–0.49)</td>
<td>0.38 (0.01–0.72)</td>
</tr>
</tbody>
</table>

Bold numbers are average genetic distances followed by min. and max. distances in parentheses and italics.
suited to cultivation of cowpea and represents a significant barrier to movement of germplasm. In our study, wild ancestral cowpea of the subspecies *dekindtiana* from West and East Africa also formed two distinct groups and these groups of wild cowpea were clustered relatively closer to the cultivated group from the same geographic region (Fig. 4). Therefore, the broad pattern is suggestive of divergent domestication processes from West and East African *dekindtiana*, respectively, which may be analogous to common bean (*Phaseolus vulgaris* L.) where the genetic architecture of the species supports the existence of two distinct domesticated gene pools, Meso-American and Andean, each derived from different wild beans (Chacón S et al., 2005; Gepts, 1998). However, because gene flow can occur between cultivated and wild cowpea (Rawal, 1975), introgression with local wild cowpea would tend to result in a similar pattern of relatedness, with the local cultivated types becoming relatively closer to local wild forms regardless of their geographical origin.

Given the relatively wider genetic diversity observed among landraces in eastern Africa (Table 1), another plausible hypothesis is that a single early domestication might have occurred in this region followed by movement via human migration to western Africa, bringing cowpea into an area of narrower genetic diversity where gene flow from wild cowpea and directional selection could have led to the formation of a distinct gene pool in the West. There is evidence of only very recent admixture in western African landraces as revealed by structure analyses. Admixed individuals

*Figure 4. Phylogenetic relationships between wild annual cowpea (*Vigna unguiculata* subsp. *dekindtiana*) and “pure” representatives of two gene pools with admixture scores less than 0.01 based on structure analyses. Accessions from West, North, and central Africa are coded blue and accessions from East, southeast, and southern Africa are coded red.*
inherit large genome regions from an external population and thus are difficult to separate by structure using tightly linked markers (Pritchard et al., 2000). By applying a SNP set that excluded tightly linked markers we observed a clearer separation of subgroups in the western Africa collection (Supplemental Fig. S2).

The small genetic differentiation observed between the African and non-African collections indicated that the entire genetic diversity in the African germplasm might already have spread over cowpea-growing regions in the world as a whole although not completely within any single region. Dispersal probably occurred through different routes as revealed by typical patterns of genetic relatedness between world cowpea collections relative to the two primary gene pools in Africa. Although only nine accessions from North America were included in this study, the majority (6 accessions) were assigned to gene pool 2, implying that much of domesticated cowpea in North America did not move directly from West Africa, in contrast to the popular view that cowpea was introduced directly from this region during the slave-trading period (Whit, 2007). Among 66 Asian landraces, about 50% and 20% of them were assigned to gene pool 1 and gene pool 2, respectively, suggesting that representatives from both gene pools were taken to Asia and have existed there for a long time. It is tempting to speculate that cowpea from West Africa was moved to India and other Asian countries along with sorghum and pearl millet from the same region at the time when these crops are presumed to have been introduced to the Asian subcontinent (Faris, 1965; Steele, 1976) following their domestication in Africa. A logical assumption is that when cowpea moved farther east into Asia and encountered more humid conditions poorly suited to dry grain production, human selection for use of the immature pods gave rise to a unique form of vegetable cowpea called “long bean” or “asparagus bean” [Vigna unguiculata subsp. sesquipedalis (L.) Verdc.], which is not found in African domesticated forms. Long bean cowpea has extremely long (50–90 cm) pods that are used as a “snap bean” when young and tender and have a vigorous climbing growth habit quite unlike other domesticated forms of cowpea that are either prostrate vines or bush types that do not climb readily. Indeed, in a recent study involving 95 asparagus bean accessions collected across China, Xu et al. (2012) also reported two distinct subgroups existing in the collection, which may align with the two major African gene pools of cowpea reported in this study.

Implications for Preserving and Using Cowpea Germplasm

The description of patterns of genetic relatedness and clustering of relatively similar individuals into groups of gene pools reported here provides important insights that can improve the efficiency of germplasm preservation and breeding efforts for cowpea. The information will enable rational planning by gene banks to help reduce duplicates (as shown in Supplemental Tables S2 and S3) and to ensure an adequate and balanced representation of the major cowpea gene pools. For breeding programs, members within a gene pool or a race within a gene pool may exhibit common adaptive complexes of physiological traits coupled with a relatively restricted range of morphological and underlying genetic variation. Therefore, crosses within gene pools or races are expected to produce a high frequency of relatively similar-looking progeny while crosses between members of different gene pools or races are expected to produce more variable progeny, perhaps with a relatively lower average performance in early generations. Breeding strategies involving one or more backcross steps may be needed to increase the frequency of useful progeny in such cases (Ehlers and Foster, 1993).

Breeding programs generally work within restricted pools of genetic variation. If specific attempts are not made to introgress new germplasm into the programs, genetic variation is reduced over time thereby limiting short or longer term genetic gain. From our study, the delineation of the broader germplasm of cowpea landraces into gene pools could help guide introgression efforts to expand the genetic diversity within breeding materials and may lead ultimately to development of more efficient strategies and greater genetic gain within future breeding programs. The SNP genotypic database developed from this study (Supplemental Table S2) also can be useful directly for this purpose by allowing users to conduct genomewide association studies and to generate a customized list of polymorphic SNP markers for a biparental breeding population for application in marker-assisted selection.

Supplemental Information Available

Supplemental material is available at http://www.crops.org/publications/tpg.

Supplemental Figure S1. Detecting the number of subpopulations.

Supplemental Figure S2: Population structure of world cowpea landraces.

Supplemental Figure S3: Phylogenetic relationships among cowpea landraces.

Supplemental Table S1: Taxonomy and locations of cowpea landraces and wild relatives.

Supplemental Table S2: Single nucleotide polymorphism (SNP) genotype data of cowpea landraces.

Supplemental Table S3: Single nucleotide polymorphism (SNP) genotype data of wild cowpeas.

Supplemental Table S4: Genomewide distribution of polymorphic markers.

Supplemental Table S5: Membership coefficients for cowpea landraces.

Acknowledgments

This work was supported by the Generation Challenge Program of the Consultative Group on International Agricultural Research and the USAID Dry Grain Pulses CRSP (grant EDH-A-00-07-00005).
References
tary trees with polymorphic microsatellites. Nature 368:453–457. doi:10.1038/368455a
Earl, D., and B. vonHoldt. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and imple-
Jakobsson, M., and N.A. Rosenberg. 2007. CLUMPP: A cluster match-
Steele, W. 1976. Cowpea, Vigna unguiculata (Leguminosae-Papilionatae). In: N. Simmonds, editor, Evolution of crop plants. Longman, Lon-
don, UK.
rion in Chinese asparagus bean (Vigna unguiculata ssp. sesqui-