Allelic Richness and Diversity in Global Composite Collection and Reference Sets in Chickpea *(Cicer arietinum L.)*

Hari D Upadhyaya[1], Sangam L Dwivedi[1], Michael Baum[2], Rajeev K Varshney[1], Sripada M Udupa[2], CL Laxmipathi Gowda[1], Valiyaparambil Prasanth[3], Bonnie J Furman[2], David A Hoisington[1], Subhash Chandra[4], and Sube Singh[1]

[1]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru PO, 502324, AP, India Email h.upadhyaya@cgiar.org

[2]Internationtional Center for Agricultural Research in the Dry Areas (ICARDA), PO Box 5466, Aleppo, Syrian Arab Republic

[3]Agricultural Research Station, Jinnaram Mandal, Vavilala, 502319, AP, India

[4] Department of Primary Industries, Tatura 3616, Victoria, Australia

Abstract

Chickpea is the fourth largest grain legume crop globally. A composite collection of 3000 accessions was formed and genotyped using 50 SSR markers. The accessions were also field evaluated for seven qualitative traits. Analysis of 48 SSR markers data on 2915 accessions detected 1683 alleles, of which 935 were rare and 748 common. Gene diversity varied from 0.533 to 0.975. Kabuli chickpea as a group were genetically more diverse than other seed types. Several group-specific unique alleles were detected: 104 in Kabuli, 297 in desi, and 69 in wild *Cicer*; 114 each in West Asia and Mediterranean, 117 in South and South East Asia, and 10 in African region accessions. A genotype-based reference set captured 1315 alleles compared to 1237 alleles in the reference set based on qualitative traits or 1354 alleles based on SSRs and qualitative traits data. The relative usefulness of these reference sets in chickpea breeding and genomics studies are being further investigated.

Media summary

Mining allelic variation in reference set will facilitate identification of diverse germplasm with beneficial traits for enhancing the genetic potential of chickpea.

Keywords

Genetic diversity, Population structure, rare and unique alleles, polymorphism information content

Introduction

Chickpea (*Cicer arietinum* L.) is the 4^{rth} largest grain-legume crop after soybean, bean, and pea. Globally, 8.24 million tons are produced from an area of 9.4 million ha, with an average production of 0.77 t ha^{-1}(FAO 2006). Some 48 countries grow chickpea and report a large variation in grain yield (0.34 to 3.33 t ha^{-1}). Several biotic and abiotic stresses, a narrow genetic base probably as a consequence of its monophyletic origin (Abbo et al. 2003) and lack of adapted varieties, contribute to fluctuations in yield. Chickpea is a self-pollinated crop, with 2n = 2x = 16 chromosomes. Two major forms of cultivated chickpeas are desi (small, angular, and colored seeds with high fibre) and kabuli (large, ram-head shaped, beige colored seeds with low fibre). A

third type, designated as pea-shaped, is characterized by round and medium to small sized seeds. Desi cultivars are primarily grown in South Asia while kabuli cultivars are found in the Mediterranean region. Chickpea is a good source of carbohydrates and proteins, together contributing 80% of the seed dry weight. The chickpea grains are rich in minerals and vitamins, and also used as livestock feed.

Knowledge and management of genetic diversity are critical for any crop improvement programs. Vast collections of chickpea germplasm are maintained by national and international genebanks. Development of core and mini core subsets has been suggested as a gateway to the utilization of genetic diversity in crop improvement. Core and minicore subsets have been reported in chickpea (Upadhyaya et al. 2001; Upadhyaya and Ortiz 2001). More recently, Upadhyaya et al. (2006) developed a composite collection of 3000 accessions, selected from over 29,000 cultivated and wild *Cicer* (*C. reticulatum* and *C. echinospermum*) accessions maintained in ICRISAT and ICARDA genebanks. This study was initiated to dissect the structure, diversity, and allelic richness in the composite collection using SSR markers, and to form a reference set of 300 accessions for chickpea improvement.

Methods

A single plant from each of the 3000 accessions, including controls (Annigeri and ICCV2), was harvested from the field and seeds obtained from such plants were used to raise seedlings for DNA extraction. Young leaf tissues of each accession from the greenhouse grown plants were harvested and immediately stored in 96-well plates. High quality DNA was extracted, and part of it was shipped to ICARDA. Fifty SSR markers, mostly di- and tri-nucleotide repeat motifs, were used to genotype the composite collection at ICRISAT and ICARDA. ICRISAT generated data for 35 SSRs on 3000 accessions using ABI3700, while ICARDA generated data for 15 SSRs on 3000 accessions using ABI3100.

Allele calling was done using GENOTYPER, and alleles binned using ALELOBIN, ICRISAT-developed software implementing the algorithm of Idury and Cardon (1997). Two markers were dropped: TA28 (unacceptable allele size) and TR2 (high heterozygosity most likely due to a duplicate locus). Data for 48 SSRs on 2915 accessions (with less than 3.25% missing data) was used for statistical analysis. PowerMarker V3.0 (Liu and Muse 2005) was used to estimate polymorphism information content (PIC), allelic richness, gene diversity, heterozygosity, and unique, rare and common alleles. Unique alleles are those that present in one accession or one group of accessions but absent in other accessions or group of accessions. Rare alleles are those with <1% frequency while common alleles with ≥1% frequency in the investigated materials.

The composite collection was also field evaluated for seven qualitative traits (IBPGR, ICRISAT, and ICARDA 1993) in augmented design during 2004/05 post-rainy season at Patancheru, India, and data was recorded on five competitive plants per accession.

Simple matching allele frequency-based distance matrices were used in DARwin-5.0 program (Perrier et al. 2003) to dissect the genetic structure of the composite collection and for the identification of a reference set of 300 accessions.

Results

Genetic structure, diversity and allelic richness in composite collection

Forty-eight SSRs on 2915 accessions produced 1683 alleles, ranging from 14 to 67 alleles per SSR, with an average of 35 alleles locus^{-1}. Gene diversity varied from 0.533 to 0.975, averaging 0.869. A very low level of heterozygosity was detected, 0.00% to 3.23%, averaging 0.80%. Of the 1683 alleles, 935 were rare and 748 were common alleles. Rare and common alleles were detected at all the 48 SSR loci. Rare alleles ranged from 7 to 47 averaged 19.5 alleles locus^{-1}, while common alleles from 3 to 39 averaged 15.6 alleles locus^{-1}. Markers with tri-nucleotide repeat motifs showed greater average allele range size (135bp) than those either with di-nucleotide (89bp) or compound (131bp) repeat motifs markers.

This study detected many rare, common and unique alleles within each group. Desi accessions contained a higher proportion of rare alleles (53%) than kabuli's (46%), while wild *Cicer* were devoid of rare alleles. These groups also shared a number of common alleles. For example, desi and kabuli accessions shared 436 alleles while wild *Cicer* with desi and kabuli shared only 17 and 16 alleles, respectively. Pea-shaped accessions shared 7 alleles with desi and 8 alleles with kabuli. South and South East Asia (SSEA) and West Asia (WA) accessions shared 74 alleles while those from Mediterranean with WA and SSEA shared 38 and 33 alleles, respectively. Accessions from Africa shared more alleles with SSEA (11) than those from Mediterranean (3) and WA (5). Frequency of common alleles between desi and kabuli ranged from 47% to 54% while pea-shaped had 99% common alleles. Accessions from Africa had more common alleles (76%) than those from WA (59%), Mediterranean (54%), and SSEA (49%). Desi types contained the largest number of unique alleles (297) followed by kabuli (104) and pea-shaped (4) types. Sixty-nine unique alleles differentiated wild *Cicer* from cultivated types. WA and Mediterranean accessions each had 114 unique alleles while SSEA accessions had 117 unique alleles. Accessions from Africa had 10 unique alleles. A very high proportion of common alleles (99-100%) found in Confederation of Independent States, European, North Central America, and South America accessions probably revealed homogeneity in the genetic materials from these regions. These are the regions that also detected a very low number of unique and rare alleles.

Several differences were detected in allelic richness between groups in the composite collection: desi and kabuli types possess greater average number of alleles (27-31) than those from pea-shaped and wild *Cicer* (7-14), with more alleles in desi than kabuli (31 compared to 27). Africa region accessions had less alleles than those from Mediterranean, SSEA, and WA accessions (26-28 compared to 17 in Africa). This difference could be due to variable sample size. However, we also detected differences in allele size range. For example, the average allele size range in desi and kabuli types differ by 12bp while in pea-shaped types by 47-59bp. Mediterranean and SSEA region accessions had no difference in mean allele-size range (103.5 to 104.8bp). African accessions, in contrast, differ by 36-37bp from Mediterranean and SSEA region accessions. The WA region accessions differ by 3-4bp from the Mediterranean and SSEA.

Neighbour-joining tree of the composite collection broadly separated kabuli from desi, with pea-shaped dispersed in both the groups, and wild *Cicer* falling within kabuli types. However, *C. echinospermum* separated from *C. reticulatum*, though both belong to primary gene pool and cross compatible with *C. arietinum*.

Allelic richness and diversity in reference sets

A SSRs-based reference set of 300 accessions was formed capturing 78% alleles (1315) of the 1683 composite collection (2915 accessions) alleles. The alleles ranged from 8 to 56, averaged 27 alleles locus^{-1}. This reference set contained 463 rare and 852 common alleles. The rare alleles ranged from 2-20, averaging 9.6 alleles locus^{-1} while common alleles ranged from 3 to 41, averaging 17.8 alleles locus^{-1}. The reference set based on qualitative traits captured 73.5% alleles

(1237) while SSRs and qualitative trait-based reference set captured 80.5% alleles (1354). The neighbour-joining tree indicated diversity in all branches of the composite collection tree.

Conclusions

This is the largest and most extensive molecular dataset for chickpea. The SSRs-based reference set captured more allelic variability (1315) than the qualitative traits -based reference set (1237). However, reference set formed based on SSRs and qualitative traits data captured more alleles (1354). These reference sets are useful resources for detecting new sources of genetic variation and allelic variants of candidate gene(s) associated with beneficial traits, identifying diverse lines for use in functional and comparative genomics, in mapping and cloning gene(s), and in applied breeding. Limited seed stock of this reference set is available upon request to researchers under the CGIAR's standard Material Transfer Agreement.

Acknowledgement: Authors acknowledge and thank the Generation Challenge Program for a commissioned grant to develop and genotype the chickpea composite collection.

References

Abbo S, Berger J, Turner NC. 2003. Evolution of cultivated chickpea: Four bottlenecks limit diversity and constrain adaptation. Funct Plant Biol. 30:1081-1087.

IBPGR, ICRISAT, ICARDA 1993. Descriptors for chickpea (*Cicer arietinum* L.). International Board of Plant Genetic Resources, Rome, Italy; International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India; and International Center for Agricultural Research in the Dry Areas, Aleppo, Syria.

FAO 2006. (http://faostat.fao.org/site/408/DesktopDefault.aspx?PageID=408)

Idury RM, Cardon LR. 1997. A simple method for automated allele binning in microsatellite markers. Genome Res. 7:1104-1109.

Liu K, Muse SV. 2005. PowerMarker: Integrated analysis environment for genetic marker data. Bioinformetics 21:2128-2129, http://www.powermarker.net.

Perrier X, Flori A, Bonnot F. 2003. Data analysis methods. Pages 43-76 *in* P Hamon, M Seguin, X Perrier, JC Glaszmann (eds.) Genetic diversity of cultivated tropical plants. Enfield, Science Publishers. Montpellier.

Upadhyaya HD, Bramel PJ, Singh S. 2001. Development of a chickpea core subset using geographic distribution and quantitative traits. Crop Sci. 41:206-210.

Upadhyaya HD, Furman BJ, Dwivedi SL, Udupa SM, Gowda CLL, Baum M, Crouch JH, Buhariwalla HK, Singh S. 2006. Development of a composite collection for mining germplasm possessing allelic variation for beneficial traits in chickpea. Plant Genet Resour. 4:13-19.

Upadhyaya HD, Ortiz R. 2001. A minicore subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. Theor Appl Genet. 102:1292-1298.