Sequencing chickpea genomes: genotyping and high-quality reference genomes for association studies and breeding.

Peter Chang1, Vasantika Suryawanshi2, Matilde Cordeiro2, Noelia Carrasquilla-Garcia1, Wendy Vu2, Sripada Udupa3, Eric vonWettberg4, R. Varma Penmetsa1, Sergey Nuzhdin2, Douglas R. Cook1

1 University of California, Davis, USA
2 University of Southern California, Los Angeles, USA
3 International Center for Agricultural Research in the Dry Areas (ICARDA), MOROCCO. 4 Florida International University, Miami, USA

Abstract:

Natural populations of species closely related to cultivated chickpea (Cicer arietinum) were identified and collected systematically using ecological principles from their native range in southeastern Turkey. Collection was focused primarily on C. reticulatum (wild progenitor of the cultigen) and C. echinospermum, a sister species, and with limited co-incident collection of the more distantly related species C. bijugum and C. pinnatifidum.

Over 1,000 wild individuals sampled in 2013 were genotyped using Restriction- enzyme Associated DNA Genotyping By Sequencing (RAD-GBS). Based on this sequencing, allele-frequency based population assignment was conducted for all genotypes leading to the choice of the focal genotypes as donor parents for introgression population development.

Both the currently available reference genomes for cultivated chickpea are draft assemblies containing several ambiguous regions, and whole genome assemblies of wild relatives are currently unavailable. To address these limitations, focal genotypes for high- resolution reference genomes for each of the three species were selected based on: i) their use in introgression population development; ii) genetic relationships to genotypes from other sites of the same species; and iii) likelyhood of long-term stability of the collection site for potential future in-situ studies. Genotypes CDCFrontier (C.ari), Besev_079 (C. ret) and S2Drd_065 (C. ech) represent the three species. For each genotype, sequence data from ~60x short-read Illumina and ~30x long-read PacBio are being integrated with BioNano optical mapping data. Assemblies will be assessed via high-density linkage mapping (RAD-GBS) of early generation progenies derived from wild x wild and wild x cultivated crosses.

In addition, 26 wild accessions that represent ecological and molecular variation within the species and serving as potential introgression donors and recipient cultivars were sequenced to ~30x coverage via Illumina short read sequencing, data that allow for analysis of genome-wide signatures of selection and for trait-gene associations. To identify rare alleles among populations and to calculate linkage disequilibrium, and association studies with native site ecological parameters, ~200 genotypes from the same populations were sequenced to medium depth (~10x) via Illumina short-read sequencing.

Together these genome data represent a novel resource for chickpea biology, to improve the accuracy and precision of association mapping, trait-marker discovery and introgression breeding.