Sweetpotato omics at CIP

Jan Kreuze*, Merideth Bonierbale, Marc Ghislain, Wolfgang Gruneberg, Leticia Portal, Dora Quispe, Genoveva Rossel, Roland Schafleitner, Reinhard Simon

International Potato Center (CIP), Apartado 1558, Lima 12, Peru (*j.kreuze@cgiar.org)

Origin and diversity of sweetpotato

Sweetpotato (*Ipomoea batatas*) belongs to the botanical family Convolvulaceae, Genus *Ipomoea*, section Batatas. The crop originates from wild species probably somewhere between the Yucatan Peninsula of Mexico and the Orinoco River in Venezuela. Understanding its origin, domestication, and diversity is vital to design best strategy to sequence its genome(s). We analyzed a representative collection of New World sweet potato landraces (329 accessions from Mexico to Peru) with both chloroplast and nuclear microsatellite markers. Both kinds of markers supported the existence of two geographically restricted genepools, corresponding to accessions from the north-western part of South America and accessions from the Caribbean and Central America region. Different hypotheses subsist concerning its derivation from wild ancestors: (i) autopolyploid derivative of *I. trifida* (2X); (ii) allopolyploid involving *I. trifida* and an unidentified 4X parent. Our analysis suggests at least two independent domestication events and analysis of marker data and gene sequences confirmed diploid *I.trifida* populations from Central America as the most likely progenitors to sweetpotato, but also a contribution by *I.triloba.* Hence, DNA markers and cytological evidences favor the allopolyploid origin and thus the crop may be constituted by two or three related genomes. However, the contribution of *I. trifida* is generally agreed by all authors and is therefore an appropriate start for sequencing the genome(s) of sweetpotato.

Genetic and genomic resources

CIP holds in trust 7,783 sweetpotato accessions, including breeding lines, improved varieties, landraces, and wild accessions from 58 countries. CIP has developed genomic resources to better characterize, understand and utilize the available genetic resources, such as the composite genotype set, hexaploid and diploid mapping populations, molecular markers and a gene index. Complementary efforts in the US, Japan, Korea and China have contributed significantly to pool of genomic resources available for sweetpotato. Nevertheless, sweetpotato breeding is constrained by the complexity of the genetics of this out-crossing hexaploid crop and by still insufficient amount of available genomic resources. Therefore concerted action is required to improve the genomic resources available for sweetpotato to support more rapid and controlled improvement of this important crop. CIP has also recently developed a segregating population from *I. trifida* in the view of supporting the sequencing of sweetpotato genomes. These resources are briefly described hereafter.

Construction of a composite genotype set

A Composite Genotype Set (CGS) was developed consisting of 472 accessions from sweetpotato germplasm collection of CIP based on the geographical origin of the accessions in order to have a workable sample of the germplasm with high level of biological biodiversity. Moreover clones with high beta-carotene, starch, iron and zinc contents, nematode, virus and drought resistance were included. A subset of the CGS was genotyped with 1,088 DArT (Diversity arrays technology) markers and interestingly the *I. trifida* accessions, though separated

from sweetpotato accessions, clustered near to Latin-American landraces. This observation indicate again the ancestry of *I trifida* in the formation of Latin American sweetpotato germplasm. Most of these accessions are now available for international distribution, a complete list of the Composite Genotype Set is available at http://gcpcr.grinfo.net/files/cr_files/gcpcr_file832.xls.

Development of sweetpotato mapping populations

Three potential mapping populations were developed: (I) Zapallo x Wagaboolige, (II) Xuxhu18 x SPK004, and (III) Beaurregard x Tanzania (BxT) and key factors such as yield, pro-vitamin A (β-carotene), starch and dry matter content, sugars (sucrose, glucose, fructose and maltose) and minerals (calcium, magnesium, iron and zinc) were determined in field trials. Genetic maps were generated for the parents Xuxhu18, SPK004, Beaurregard and Tanzania on the basis of Amplified fragment length polymorphism (AFLP) and Simple sequence repeat (SSR) markers. However, finally focus was given to the mapping population BxT for which it was possible to identify 92 and 91 linkage groups for the parent "Beauregard" and for the parent "Tanzania", respectively. For starch, β-carotene and dry matter, seven, seven and four QTLs were identified in the parent Beauregard, respectively, whereas four, five and 6 QTLs were identified in the parent Tanzania. In total, 80% of the markers for starch, β-carotene and dry matter were very closely linked, 10 were linked and 10% unlinked. The hexaploid mapping population BxT, with 201 genotypes is included in CIP's genebank holdings and can be accessed by scientists worldwide.

Development of the sweetpotato gene index and other sequence resources

Two quarter 454 pyrosequencing runs used two normalized cDNA collections from stems and leaves from drought-stressed sweetpotato clone Tanzania and yielded 524,209 reads, which were assembled together with 22,094 publically available expressed sequence tags into 31,685 sets of overlapping DNA segments and 34,733 unassembled sequences. Blastx comparisons with the UniRef100 database allowed annotation of 23,957 contigs and 15,342 singletons resulting in 24,657 putatively unique genes. Further, 27,119 sequences had no match to protein sequences of UniRef100database. On the basis of this gene index, we have identified 1,661 gene-based microsatellite sequences, of which 223 were selected for testing and 195 were successfully amplified in a test panel of 6 hexaploid (*I. batatas*) and 2 diploid (*I. trifida*) accessions. A searchable version of the gene index, including a blastn function, is available at http://www.cipotato.org/sweetpotato_gene_index.

On the other hand a complete run on SOLiD system, and two lanes on a Illumina HiSeq 2000 run were used to generate 50 Gigabase and 70 Gigabase of raw sequence of the sweetpotato genome and bioinformatics assembly will be initiated soon.

Development of a diploid *Ipomoea trifida* mapping population

A linkage map was constructed for *I. trifida* based on the segregations of DArT, AFLP and SSR markers in 76 genotypes from the progeny from a cross between *I. trifida* M9 and *I. trifida* M19. The linkage map is composed of 740 loci (96 AFLP, 602 DArT and 42 SSR) distributed in 15 linkage groups (LOD>4). The map covers 1,335 cM. In addition a set of 121 COS (conserved orthologous set) markers are being mapped in this population and will be exploited for comparative mapping with the hexaploid mapping populations. The population has been introduced in vitro, but is not yet virus tested. Current research aims at linking the *I. trifida* and BxT maps, and virus testing the *I.trifida* population to make it globally available.