Barchi L, Acquadro A, Portis E, Alonso D, Gramazio P, Vilanova S, Diez MJ, Salinier J, Unlu A, Toppino L, Bassolino L, Tripodi P, D'Agostino N, Börner A, Brandt R, Scaglione D, Dicenta E, Pinosio S, Aprea G, Ferrante P, Finkers R, Boyaci HF, Bovy A, Stein N, Schafleitner R, Daunay MC, Lefebvre V, Lanteri S, Prohens J, Rotino GL, Giuliano G. **2019**. *Capsicum* spp. and eggplant genome sequencing and resequencing provide new tools for the characterization of genetic resources | BG-K/01. ***17th Eucarpia Meeting on Genetics and Breeding of Capsicum and Eggplant***, 11-13 septembre 2019, Avignon (France). p.173-175.

Barchi L (Orateur),

[Communication orale invitée + Résumé]

URL du congrès ou éditeur	https://symposium.inra.fr/eucarpiacapsicumeggplant2019/

Title

Capsicum spp. and eggplant genome sequencing and resequencing provide new tools for the characterization of genetic resources

Authors

Barchi L1 , Acquadro A1 , Portis E1 , Alonso D2 , Gramazio P2 , Vilanova S2 , Diez MJ2 , Salinier J3 , Unlu A4 , Toppino L5 , Bassolino L5 , Tripodi P6 , D'Agostino N6 , Börner A7 , Brandt R7 , Scaglione D8 , Dicenta E8 , Pinosio S8 , Aprea G9 , Ferrante P9 , Finkers R10, Boyaci HF4 , Bovy A10, Stein N7 , Schafleitner R11, Daunay MC3 , Lefebvre V3 , Lanteri S1 , Prohens J2 , Rotino GL5 , Giuliano G9

DISAFA, University of Torino, Grugliasco (TO), Italy. 2 Instituto de Conservación y Mejora de la Agrodiversidad Valenciana (COMAV), Universitat Politècnica de València, Valencia, Spain. 3 INRA, UR1052 GAFL, Avignon, France. 4 Bati Akdeniz Agricultural Research Institute (BATEM), Antalya, Turkey. 5 CREA-GB, Montanaso Lombardo (LO), Italy. 6 CREA , Pontecagnano Faiano (SA), Italy. 7 IPK, Gatersleben, Germany. 8 IGA Technology Services Srl, Udine, Italy. 9 ENEA, Casaccia Res Ctr, Roma, Italy. 10 WUR, Wageningen, The Netherlands. 11 The World Vegetable Center, Shanhua, Tainan, Taiwan

Abstract

BACKGROUND We report the current state of the art on genome sequencing and resequencing studies in Capsicum annuum L. and Solanum melongena as well as some of their cultivated and wild related species. The availability of high quality genome sequences in both species has opened the way to the characterization of genetic resources through genotyping-by-sequencing (GBS) approaches. Among them, we focus on the recently developed Single Primer Enrichment Technology (SPET) technique, which we have applied for targeted sequencing of a wide collection of S. melongena accessions and representatives of its primary, secondary and tertiary genepools. RESULTS Pepper Since the beginning of 2014, various consortia have released the genome sequences of domesticated and wild Capsicum species. The whole-genome sequences of C. annuum CM334 and C. chinense PI159236, which carry important disease resistance traits and have been widely used as parental lines of mapping populations, were the first to be released [1]. Next, the genome sequences of C. annuum Zunla-1 and of the wild species Chiltepin (C. annuum var. glabriusculum) were published [2]. Both studies highlighted that the pepper genome size is ~3-3.5 Gb, has a high percentage (over 80%) of repetitive elements, and encodes about 35K genes. Later, the improved version of the reference genome of CM334 and C. chinense

PI159236, together with the sequencing of the domesticated C. baccatum were also made available, and contributed to deciphering the evolutionary relationships among the three species as well as to estimating the lineage-divergence times occurring in Capsicum [3]. By adopting the linked-read sequencing technology, the genome sequence of an F1 hybrid obtained by crossing CM334 with a non-pungent C. annuum breeding line was also published [4]. The availability of pepper genome sequences has allowed the resequencing of Capsicum accessions at both targeted genomic regions and the whole genome level. By applying a bulk segregant analysis, it has been possible to identify markers tightly linked to the Pvr4 locus, conferring dominant resistance to three pathotypes of Potyvirus (PVY, [5]). Furthermore, after the resequencing of resistant and susceptible cultivars, SNPs and putative alleles related to resistance against powdery mildew and bacterial wilt have also been detected [6,7]. To provide insights into the process of pepper domestication, the Zunla-1 genome sequence was analysed together with resequencing (20-30X) information of 18 cultivated accessions representative of the major varieties of C. annuum and two semi-wild/wild peppers [2]. This resulted in the identification of 115 genomic regions affected by artificial selection in cultivated peppers. Lastly, genome resequencing (30X) of four genotypes, representative of the main varietal types grown in the Mediterranean region, has been performed. Distinctive variations in miRNAs and resistance gene analogues (RGAs) have been highlighted as well as mutations in the coding sequences and regulatory regions of genes affecting fruit size and shape (Barchi et al., SOLCUC meeting 2017). Within the G2P-SOL EU project (www.g2p-sol.eu), the analysis of genome-wide genotyping-by-sequencing (GBS) data on 9,659 pepper accessions retrieved from the major European (CGN, INRA, IPK, UPV) and Asian (AVRDC) gene banks, Universities and Research Centres has been recently performed and the results obtained reported by Tripodi et al. (see Proceedings of this Meeting).

Eggplant: In 2014, Hirakawa and co-authors [8] produced the first unanchored draft of the S. melongena genome sequence using the Nakate-Shinkuro accession. The obtained sequence covered about 70% of its projected 1.2 Gb genome size and more than 42K genes were identified. More recently, the Italian Eggplant Genome Consortium (IEGC) [9] developed a high quality and anchored genome assembly of the eggplant line 67/3, the male parent of an F6 RIL (Recombinant Inbred Line) mapping population. A hybrid assembly, covering 1.22 Gb, was obtained by merging Illumina sequencing data and optical mapping. The female parent of the RIL mapping population (line '305E40') was also sequenced (coverage of 34X), and thanks to low coverage resequencing (1X) of the F6 RILs, the genome assembly was anchored to the 12 chromosomes. Recently, based on Illumina sequencing data, a draft genome assembly of 1.02 Gb in size was developed for the cultivated species S. aethiopicum, which contained as for eggplant, about 76% of repetitive sequences. Furthermore, compared to S. melongena, an expansion of gene families involved in drought or salinity tolerance as well as disease resistance including defence responses was identified. Recently, the resequencing of seven eggplant accessions and one accession of the wild relative S. incanum, which are the parents of a MAGIC population, has been performed (Gramazio et al., see Proceedings of this Meeting). The set of identified SNP polymorphisms has been annotated and currently is being used for further analyses in order to efficiently genotype the MAGIC population with the goal of dissecting key agronomic and morphological traits. More recently, 60 genotypes of the cultivated scarlet eggplant (S. aethiopicum) belonging to the varietal groups, "Gilo" and "Shum", as well as 5 accessions of its ancestor S. anguivi, were sequenced at a coverage of 30-60X, with the goal of investigating the evolution, population demography and domestication history of the species [10].

Population structure of eggplant germplasm based on Single Primer Enrichment Technology (SPET) genotyping : Within the G2P-SOL EU project (www.g2p-sol.eu), which brings together the major European (CGN, INRA, IPK, UPV) and Asian (AVRDC) gene banks, Universities and Research Centres, 2,912 S. melongena, 305 S. aethiopicum and 122 S. macrocarpon accessions as well as a set of 266 accessions belonging to 29 wild species have been inventoried. The Single Primer Enrichment Technology (SPET)

genotyping, recently developed by Nugen® was applied for their targeted genotyping. Starting from more than 12K polymorphic sites found in both coding regions and in the introns/UTRs, a set of 5K best performing SPET probes was used for diversity analyses, and a panel of about 25K high confident (min mean read depth of 30 and max missing data of 0.5%) SNPs evenly distributed throughout the genome were detected. The FastSTRUCTURE analysis (Fig. 1) identified 9 main clusters (K), with S. melongena accessions grouping into 6 sub-clusters. Three further clusters were found, including accessions belonging to S. macrocarpon and S. aethiopicum as well as species belonging to the subgenus Solanum, in good agreement with clustering obtained with a maximum likelihood phylogenetic tree. A total of 1,114 accessions were finally classified as admixed. The first two components of the PCA (Fig. 2) explained about 26% of the genetic variation. The first axis, explaining 15% of the genetic variation separated S. macrocarpon from the other species. The second component grouped separately S. melongena and S. aethiopicum as well as species from the Leptostemonum and Solanum subgenera. Several accessions had an unclear assignment, presumably due to misclassifications in the genebanks or gene flow among species. The gathered information was used for the development of a core collection of genotypes for future Genome Wide Association (GWA) studies. For this purpose, 15 S. melongena accessions and 5 wild species will be re-sequenced at 30X with the goal of identifying new SNPs useful for high resolution SPET genotyping.

REFERENCES

[1] Kim S et al., 2014, Nat Genet. 46, 270–8 [2] Qin C et al., 2014, Proc Natl Acad Sci. 111, 5135–5140 [3] Kim S et al., 2017, Genome Biol. 18, 210 [4] Hulse-Kemp AM et al., 2018, Hortic Res. 5, 4 [5] Devran Z et al., 2015, Mol Breed. 35, 101 [6] Kang YJ et al., 2016, BMC Plant Biol. 16, 235 [7] Ahn YK et al., 2018, Sci Reports 8, 5188 [8] Hirakawa H et al., 2014, DNA Res. 21, 649–60 [9] Barchi L et al., Sci Rep Submitted. [10] Song B et al., 2019, bioRxiv 532077
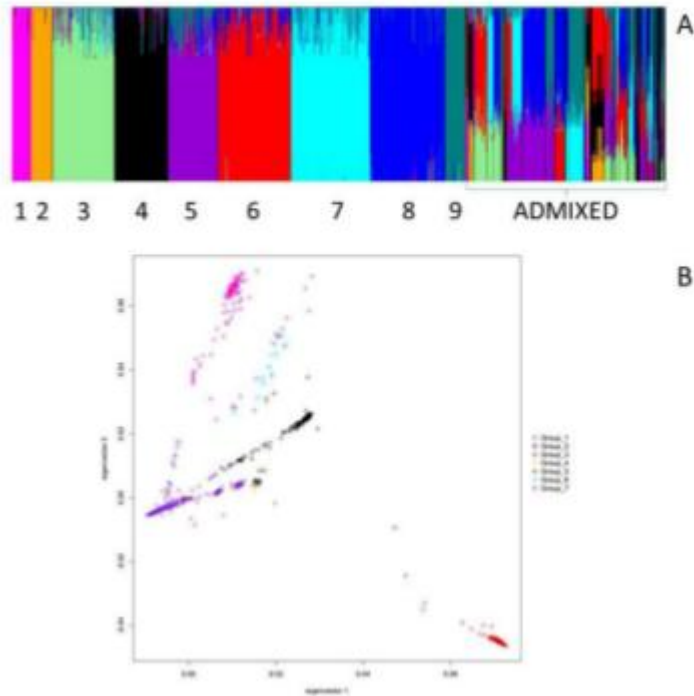
Figure 1.

A The FastSTRUCTURE genetic architecture of the full germplasm panel at K = 9. Group 1 includes subgenus Solanum species, Group 2 mainly includes S. macrocarpon and closest relatives, Group 3, 5, 6, 7, 8 and 9 includes S. melongena and closest relatives, Group 4 includes S. aethiopicum and closest relatives. Admixed groups include among others other African, Asian and Australian wild relatives of eggplants as well as the Torva group-American species.

B: PCA results: Group 1 includes S. melongena and closest relatives, Group 2 includes S. aethiopicum and closest relatives, Group 3 includes S. macrocarpon and closest relatives, Group 4 includes other African, Asian and Australian wild relatives of eggplants, Group 5 includes the Solanum section Torva group-American species, Group 6 includes other American subgenus Leptostemonum species and Group 7 includes subgenus Solanum species.