A high-density genotyping strategy based on gene capture in pepper: perspectives for genome wide association study and genetic mapping

Sonia Elbelt, Jacques Lagnel, Bernard Caromel, Jacques David, Judith Hirsch, Emmanuel Szadkowski, Benoît Moury, Caroline Caporalino, N. Stein, G. Giuliano, et al.

A high-density genotyping strategy based on gene capture in pepper: perspectives for genome wide association study and genetic mapping

Elbelt S[1], Lagnel J[1], Caromel B[1], David J[2], Hirsch J[3], Szadkowski E[1], Moury B[3], Djian-Caporalino C[4], Stein N[5], Giuliano G[6], Lefebvre V[1]

[1]INRA, UR1052 GAFL, Avignon France. [2]INRA, Montpellier SupAgro, UMR Amélioration Génétique et Adaptation des Plantes, Montpellier, France. [3]INRA, UR407 Pathologie Végétale, Avignon, France. [4]INRA, UMR IPMSV 1064, Sophia Antipolis, France. [5]Leibniz Institute of Plant Genetics and Crop Plant Research IPK, OT Gatersleben, Germany. [6]Agenzia Nazionale per le Nuove Tecnologie, l'Energia e lo Sviluppo Economico Sostenibile (ENEA), Casaccia Research Center, Roma, Italy

BACKGROUND The genetic diversity within cultivated peppers (*Capsicum annuum*) has been reduced through its domestication and since its first introduction from the West Indies into Europe during the first travel of Christopher Columbus in XVth. For a long time, it was consequently difficult to get dense genotyping in *C. annuum*. Moreover, pepper is characterized by a large genome size (~3.5 Gb) resulting from its expansion by accumulation of transposable elements (81.5% of the genome), preventing the discovery of SNPs evenly distributed throughout the genome [1]. To overcome this difficulty, we chose a targeted sequence gene capture strategy [2]. We present here the design of baits combining different approaches. First, in a genome wide approach, baits were designed in polymorphic regions identified in RNAseq and genotyping by sequencing (GBS) datasets from pepper genome. Second, in a candidate gene approach, genes of interest from published datasets of other species were in the focus of our investigation in pepper.

MATERIALS & METHODS In order to identify SNPs, reads from two sequencing datasets (21 genotypes from RNAseq, 282 from GBS) were mapped to the 35,884 genes from the reference pepper genome CM334 v1.6. In addition, a set of 10K SNPs from the G2P-SOL project on 871 INRA accessions was included. SNPs calling was performed using the pipeline of Holtz et al. [2]. Since the bait hybridization is efficient when >92-95% similarity occur, we designed baits on polymorphic sites of exons as they are well conserved between genotypes. Plant genes involved in oomycete, virus and nematode resistance as well as in abiotic stresses were selected from literature and their homologs identified in pepper using BLAST.

RESULTS After filtering a total of 463,525 unique SNPs, 26,777 genes (74.6% of the genes annotated on the genome) were found to contain at least one SNP. The majority of SNPs, and consequently the majority of the genes containing SNPs, were found using the RNAseq dataset compared to GBS datasets (Figure 1). The genes containing at least one SNP were evenly distributed throughout the genome. We identified 700 candidate genes from literature leaded to 1,646 homologous in CM334 (Table 1). A total of 1,352 candidate genes contained SNPs (82% of the candidate genes).

DISCUSSION & CONCLUSION Our aim was to obtain at least one polymorphic SNP for each gene. For 25% of pepper genes where no SNP was detected, we designed baits on the first or the last exon in order to catch UTR regions. In the candidate gene approach, we maximized the coverage of each gene to detect novel SNPs by designing baits on each exon taking into account their polymorphic sites. Moreover, to overcome the difficulty of designing specific baits on resistance genes that belong mainly to highly conserved NB-LRR families, we preferentially designed baits on the last exon close to the 3'UTR. Finally, a total of 60,000 baits were designed for bait sequencing capture and will be used further for genome wide association study and genetic mapping.

REFERENCES

[1] Kim S, Park J et al, 2017, Genome Biol, 18, 210. DOI 10.1186/s13059-017-1341-9
[2] Holtz Y, Ardisson M et al, 2016, PLoS ONE, 11, e0154609. DOI: 10.1371/journal.pone.0154609

ACKNOWLEDGEMENTS

The G2P-SOL project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 677379

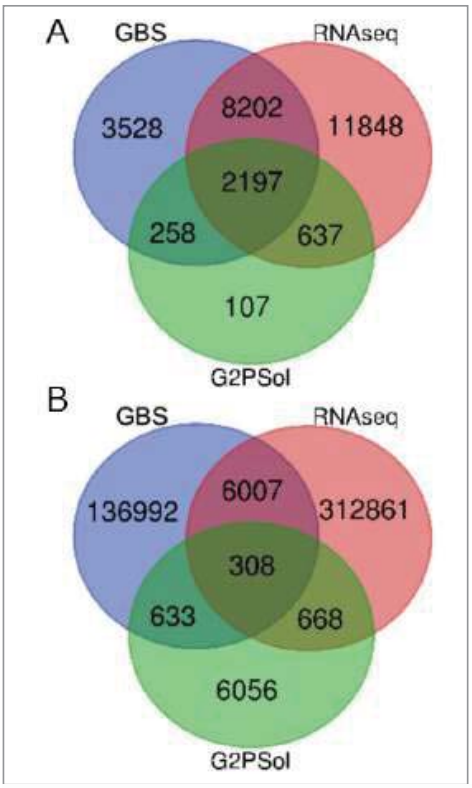

Figure 1. (A) Venn diagram showing genes with at least one SNP for the three datasets. **(B)** Venn diagram showing SNPs for the three datasets.

Type of GC	GC number	Pathogenes family
Resistance genes	57	Oomycetes
		Virus
		Nematodes
		Others
Susceptibility genes	15	Oomycetes
		Others
Interactants	304	Oomycetes
		Virus
		Others
Small RNA factors	14	Virus
ESCRT factors	12	Virus
TIR and non-TIR-NB-LRR analogous	80	NA
Temperature tolerance genes	238	NA
Photosynthesis genes	38	NA
Quantitative trait loci (QTL)	18	Nematodes

Table 1. Genes from Solanaceae, Rosaceae, Fabaceae, Brassicaceae and Poaceae family were identify from literature and sorted in function of their category. Resistance and susceptibility genes are known genes conferring resistance; interactants are plant proteins interacting with the pathogen or target of effectors; small RNA factors are dicer-like (DCL) proteins or targets of miRNA; ESCRT (Endosomal Sorting Complexes Required for Transport) factors are part of the membrane trafficking machinery.

Innovations in Genetics and Breeding of Capsicum and Eggplant

Proceedings of the 17th EUCARPIA Meeting on Genetics and Breeding of Capsicum and Eggplant,

September 11-13, 2019 | Avignon - France

Editors: Véronique Lefebvre & Marie-Christine Daunay

Editors
Véronique Lefebvre & Marie-Christine Daunay

Title
Innovations in Genetics and Breeding of Capsicum and Eggplant

Sub-title
Proceedings of the 17th EUCARPIA Meeting on Genetics and Breeding of Capsicum and Eggplant
September 11-13, 2019 | Avignon - France

Publisher
Institut National de la Recherche Agronomique (INRA)

Centre de recherche Provence-Alpes-Côte d'Azur
228 route de l'aérodrome
CS 40 509 - Domaine Saint Paul, Site Agroparc,
84914 Avignon Cedex 9 - France

Visual identity
© Armelle Favery

Artistic director
© Lyonel Liger assisted by Sabine Laugier

Layout design and editing
Salima Kherchache

Printed by
SUD LABO, 35 avenue Pierre Sémard, 84000 Avignon - France