Poster Communication Abstract – 2A.49

A CATALOG OF MOLECULAR DIVERSITY WITHIN *PRUNUS* GERMPLASM INFERRED FROM NEXT-GENERATION SEQUENCING DATA: BIOINFORMATIC APPROACHES AND CHALLENGES

SCALABRIN S.*, POLICRITI A.*,***, NADALIN F.*,***, DEL FABBRO C.*, MICULAN M.*,***, PINOSIO S.*,***, CATTONARO F.*, VENDRAMIN E.****, ARAMINI V.****, VERDE I.****, ROSSINI L.*****, TESTOLIN R.*,**, MORGANTE M.*,**

*) IGA - Istituto di Genomica Applicata, Via J. Linussio 51, 33100 Udine (Italy)
**) DiSA - Dipartimento di Scienze Agrarie e Ambientali, Via delle Scienze 208, 33100 Udine (Italy)
***) DiMi - Dipartimento di Matematica e Informatica, Via delle Scienze 206, 33100 Udine (Italy)
****) CRA - Centro di Ricerca per la Frutticoltura, Via di Fioranello 52, 00134 Roma (Italy)
*****) PTP - Parco Tecnologico Padano, Via Einstein, Loc. Cascina Codazza, 26900 Lodi (Italy)

Next Generation Sequencing, Prunus persica, SNP, indel, bioinformatics

Genome analysis based on next generation sequencing (NGS) technologies provides a powerful approach for surveying molecular diversity among individuals, which in turn can generate tools for linkage mapping, association mapping, gene cloning, molecular breeding, population genetics, germplasm management, crop systematics and evolution.

'De novo' assembly of short reads is challenging as far as the size and complexity of genomes increases. A reference genome correctly assembled and annotated can help solving most of the problems, although several structural variants such as the movement of transposable elements, large insertions/deletions, segmental duplications and other genomic features are still challenging algorithms and automatic procedures.

We sequenced 16 Prunus accessions, that include 14 peach cultivars and species, one almond and one apricot varieties, using the NGS Illumina platform. We produced 64 to 109 bp long paired end reads from approx. 300-500 bp long fragments. The coverage varied from approx. 16 to 75 genome equivalents. Individual genomes were aligned using the doubled haploid peach cultivar 'Lovell' reference sequence recently released by the International Peach Genome Initiative (IPGI) (http://www.rosaceae.org/peach/genome).

In this work we present a catalog of molecular variants that can be mined with different bioinformatic approaches, namely SNPs (Single Nucleotide Polymorphisms), DIPs (Deletion/Insertion Polymorphisms), larger structural variations, which include movement of transposable elements, the so called copy-number variations, segmental duplications and other. Some of these variants, such as SNPs, are easily detected and many commercial and open-access software can perform the search. Others variants, such as the large structural variations, have still analytical approaches to be implemented or improved. For most variants, the possible methodological approaches are discussed and, when available, preliminary results are reported.