Parentage assignment in tetraploid plant species using NGS-derived markers

BERNET, Guillermo P.[1]; OZÁEZ, Jaime[1]; MIRABEL, Sophie [1]; BARBARIN, Andrea [1]; IZQUIERDO, Beatriz [1]; GARCÍA, Jaime [2]; TRIVIÑO, Juan Carlos [2]; GIRÓS, Amparo [1]
[1] Agrigenomics and Molecular Microbiology Department at Sistemas Genómicos
[2] Bioinformatics Department at Sistemas Genómicos
Corresponding authors: amparo.giros@sistemasgenomicos.com; guillermo.bernet@sistemasgenomicos.com

Introduction

Next generation sequencing (**NGS**) technologies have revolutionized many areas of biological research taking the discovery and application of **molecular markers** to high-throughput levels, thus broadening and enhancing plant breeding prospects.
In **plant breeding** programmes it is important to establish and track the exact **parentage** of families and individuals, and early assignment of genetic relationships between parents and offspring can be of undoubtable utility. DNA-based molecular markers have been widely and successfully used in this regard, but they are particularly challenging in **polyploid species** analyses because of the presence of a large number of homologues which difficult reliable polymorphisms detection.
In this work we present the consistent results obtained for the molecular characterization of a tetraploid plant species using **NGS-derived markers**.

Methods

Wet Lab: The experiment comprised the DNA extraction by duplicate analysis of 13 samples corresponding to 3 different families (two fathers and three descendants each). Construction of a reduced-representation library and its sequencing on an Illumina HiSeq 2500 (Fig. 1).

Dry Lab: Data obtained by sequencing were analyzed using a reference-based approach, where variants were identified and processed through house-made filters. Parental relationship between samples was characterized using a combination of novel scripts based on parental scores and phylogenetic distances (Fig. 2 and Fig. 3).

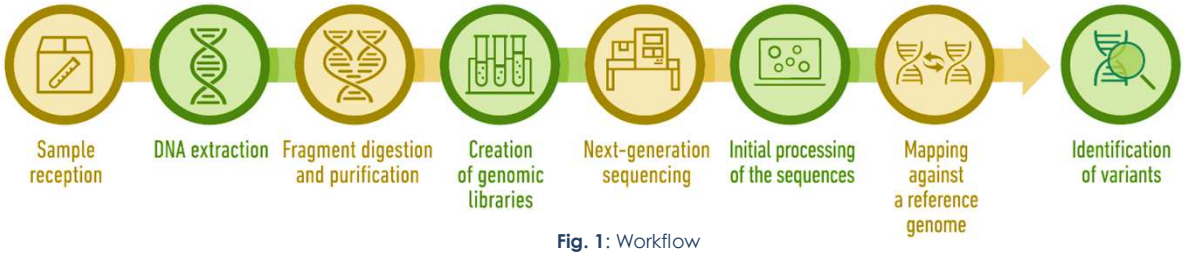

Fig. 1: Workflow

Results

The experiment has shown that the characterization of the familiar relationship through a closed dataset has not been affected by ploidy variability within individuals and the precision of the output was reported with a high level of reliability. In Fig. 2 are plotted phylogenetic distances of two different families (each individual has two replicates). Upper family (in purple and pink) has three descendants with their replicates (*AE006155* until *AE006160*) and are closer to their mother and a brother of their real father (*AE006145-6*). In the family of the bottom (in green) are represented two descendants against their mother (*AE006135-6*). With our combined method we can identify siblings and parents from one individual. We have also seen that there is always a pattern in heritage into descendants that have dominance over the other parent. Thus, although phylogenetic distances could be similar between parents and siblings, parental score allows to infer and find real parents as shown in Fig. 3.

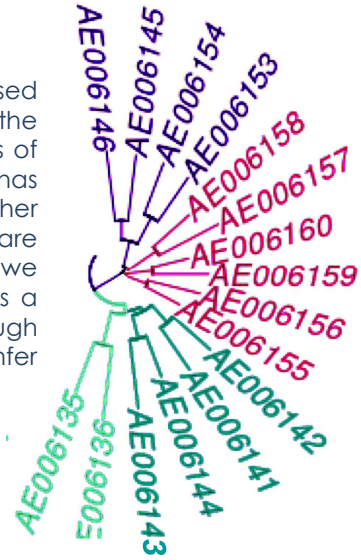

Fig. 2: Dendrogram of two different families that are genetically closer.

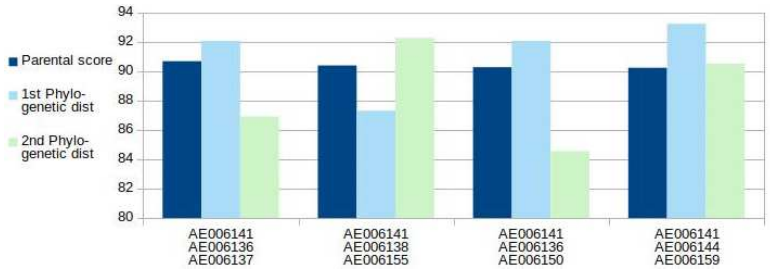

Fig. 3: Parental scores and phylogenetic distances of a real familiar trio (first one) and non familiar trios.

Conclusions

The analytical strategy here presented, using an in silico predicted suitable combination of restriction enzymes to reduce genome complexity and sequenced using the Illumina 2500 ultrasequencing platform, yields reliable results and an amount and quality of newly discovered molecular markers that allow final parentage assignment mainly based on the genetic distances detected among individuals.

1. Toonen RJ, Puritz JB, Forsman ZH, Whitney JL, Fernandez-Silva I, Andrews KR, Bird CE. (2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. PeerJ 1:e203
2. Kamvar ZN, Brooks JC and Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front. Genet. 6:208.
3. Nei, M. (1987). Molecular Evolutionary Genetics. (Chapter 9). New York: Columbia University Press.
4. McKenna, A. et al. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20, 1297-1303.