

TOWARDS A NEW PHASED GENOME REFERENCE OF FIG (*FICUS CARICA* L.), A CRUCIAL RESOURCE FOR FIG BREEDING

USAI G.*, GIORDANI T.*, VANGELISTI A.*, MASCAGNI F.*, VENTIMIGLIA M.*,
SIMONI S.*, NATALI L.*, CAVALLINI A.*

*) Department of Agriculture, Food and Environment, University of Pisa,
Pisa, Italy

Ficus carica L., genome assembly, genome annotation, genetic variability

The genome sequence is a crucial prerequisite to apply modern breeding procedures to crops, and it is fundamental to decipher the sequence of the two haplotypes which constitute the genome. This is even more true for fruit trees, whose condition of heterozygosity is maintained through clonal propagation. In these species, the heterozygosity can be the basis of favorable traits that must be maintained and/or improved. The fig tree (*Ficus carica* L.) has a great potential for commercial expansion, in fact, this species has esteemed nutritional and nutraceutical characteristics, combined with its ability of adaptation to marginal soils and difficult environmental conditions. High-quality genomic resources have been released only in recent years. Here we report our work-in-progress haplotype-phased genome assembly, which was achieved combining the last published genome reference, produced through single-molecule, real-time (SMRT) sequencing technology (Usai *et al.*, The Plant Journal 102: 600-614, 2020), and the latest methodologies of chromosome conformation capture.

A total of ~55× Hi-C data were integrated with the previously produced fig assembly, resulting in two pseudo-haplotypes of 538 sequences with mean size of 0.65 Mb and N50 of 1.99 Mb and 1.93 Mb, respectively. The pseudo-haplotypes represented ~98% of the estimated 356 Mb fig genome. A total of 400 out of 538 sequences (~96% of both pseudo-haplotypes) were associated to the 13 fig chromosomes generating the fig assembly v2.0. Proper approaches based on *de novo* prediction, RNA-seq data and protein alignment allowed us to predict 34,288 and 33,698 protein-coding genes per pseudo-haplotype, respectively. About 82% of the total protein-coding genes were

functionally annotated. The characterization of the repetitive DNA, fundamental for studying genetic variability, was performed with the same level of accuracy.

The high-quality phased genome reference is being the basis to assess the genetic variability of fig varieties on available Spanish, Tunisian and Turkish fig collections using a genotyping by sequencing (GBS) approach in the frame of a PRIMA (Partnership for Research and Innovation in the Mediterranean Area) project, FIGGEN. This data will be the prerequisite for genome-wide association studies (GWAS) with the final purpose of unveil genes or molecular markers linked to traits related to fruit quality and to environmental adaptation to difficult conditions, consequence of climate change, leading to the genetic improvement of fig.