Project: TALE transcription factors and flowering time

Final research activity report (Giovanna Frugis, IBBA-CNR, IT)

Two degenerated sequences, the 8 bp binding site TGACAG(GC)T (BS8) and the 11 bp TGAYNGAYNGA (GA2ox1-like BS), had been previously used to identify the putative TALE (KNOX and BLH) transcription factors binding sites amongst 170 flowering time genes. This analysis identified more than 50 flowering time genes as possible targets of TALE TFs.

In the framework of a CNR short term fellowship, the 1135 sequenced genomes of *A. thaliana* natural accessions, available in the "1001 Genomes Project", have been screened for SNPs/mutations in the core consensus region of the identified TALE binding sites, which may result in loss of transcriptional regulation. The screen was carried out through the 1001 Genomes tool website (http://tools.1001genomes.org/) using the "VCF Subset" tool to download high quality allelic variants in the TALE binding site regions of the 1135 *A. thaliana* genomes, with the help of Joffrey Fits who provided assistance for all the available tools. The VCF text information was then transferred to Excel files and assembled into complete data files containing all the information about the target flowering time genes, the chromosomal regions analyzed, the allelic variants found and the number of accessions harboring the mutated alleles. Additional Excel files were also created to list the specific *A. thaliana* accessions harboring the mutated alleles for each gene.

According to Dr. Weigel suggestions, the analysis was also extended to find newly evolved binding sites, which may result in acquired TALE-dependent transcriptional regulation of novel flowering time genes. This approach implies the screening of genomic regions (including promoters, coding sequences, introns and UTRs) of 170 flowering time genes, which is in progress. The novel screen was carried out using a three-steps approach: i) download of the whole genomic regions and chromosomal coordinates for each flowering time gene of the *A. thaliana* reference genome through TAIR sequence bulk download (https://www.arabidopsis.org/tools/bulk/sequences/index.jsp); ii) download of the corresponding genomic regions from the 1135 *A. thaliana* accessions thought the "Pseudogenomes" download tool (http://tools.1001genomes.org/); iii) search the downloaded sequences for TALE BS8 and GA2ox1-like BS motives through the Regulatory Sequences Analysis Tools platform (RSAT plants, http://floresta.eead.csic.es/rsat/). Information on novel acquired regulatory TALE binding sites were then collected in an Excel data file which summarizes the results.

RESULTS

For the **lost TALE binding sites**, 34 flowering time genes showed at least one allelic variation in the binding site, either in the core consensus region (39 variants) or in the 4 bp flanking the binding site (28 variants). Notably, allelic variations in 14 flowering time genes were harbored by 10 or more accessions, with the highest score of 658 accessions showing the same mutant allele, which are therefore considered good candidates for

possible adaptive mutations. Putative adaptive mutations were found in the following pathways: gibberellin (1), meristem response and development (3), photoperiod (5) and vernalization (5). Most interestingly, two putative adaptive mutations were found in two different binding sites of the *SHORT VEGETATIVE PHASE* (*SVP*) gene and eight *A. thaliana* accessions were mutated in the core region of both binding sites. A preliminary analysis of the geographical distribution and annotated flowering time of these 8 accessions showed that they belong to Spain, Italy_Balkan_Caucasus and Germany groups and all display a late flowering phenotype compared to the references Col-0 and Ler-0.

For **newly evolved binding sites**, the screen is still in progress (60 genes out of 170 have been screened by now) and will be completed in Italy. New putative binding sites were found in *PHYA*, GA2/KS and VIP6 genes, which already had binding sites for TALEs, and in *SPL4*, which did not display any conserved binding site in the reference accession. The newly acquired binding sites were observed in either 1 or 2 *A. thaliana* accessions.

PERSPECTIVES

The screen identified several putative lost and some newly evolved TALE binding sites, some of them being excellent candidates for adaptive mutations in the flowering time pathway. Initially, 50 *A. thaliana* accessions, selected among those harboring allelic variations for lost binding sites in the genes *SVP, CAL, LFY* or for newly evolved binding sites in the genes *PHYA, SLP4* and *GA2/KS,* will be studied and characterized. Functional genomics strategies to assess the effect of the putative adaptive allelic variants on flowering time will be designed in collaboration with the Prof. Weigel lab, including identification of mutants harboring the same polymorphic sites and CRISPR/Cas9 genome editing in *A. thaliana* reference species.

Date: 05 August 2016

Firma