In-silico microRNAs analysis from Cartamus and Cynara spp. EST datasets

Catalano D, Finetti-Sialer MM

Motivation

The family Asteraceae represents one of the largest evolutive radiation of flowering plants, including more than 1500 genera and 23000 species, and comprising economically important as well as ornamental crops (Jansen, 1987; Bremer, 2008). Among members of this dicotyledon family two edible species, Carthamus tinctorius L. and Cynara cardunculus var. scolymus L., also have a phytopharmaceutical interest. The former, known as safflower, is the only member of this genus widely cultivated for industrial oil, as a livestock feed or for use in traditional medicine (Han et al., 2009). The second species, the globe artichoke, apart of its importance as a food, is popular for its dietary and therapeutic potentials, especially for hepato-biliary dysfunctions and digestive complaints (Gebhardt, 1998; López-Molina, 2005). The discovery of microRNAs (miRNAs) in different genomes provided a new paradigm in evolution studies, due to their ancient origin and the important role played in development regulation, since the egress of multicellular organisms. This system is highly conserved in plant and animal cells. Plant miRNAs derive from long primary transcripts giving rise to mature 21-24nt RNAs products, fundamental in RNA-based gene regulation (Bartell, 2004). In plants, miRNAs control messengers degradation or restrain translation, affecting somatic development and the response to biotic and abiotic stresses (Jones-Rhoades, 2006). A feature of miRNAs is their imperfect but extensive complementarity to corresponding mRNA targets, thus making their computational prediction possible. This approach is useful when data mining is performed on the basis of miRNA:mRNA targets conservation, among different species, and was herein applied to the study of Carthamus and Cynara spp.

Methods

A comparative approach was used to identify both evolutionary and functionally conserved miRNAs as well as their targets, by detection of common expressed sequence tags (ESTs) in Carthamus and Cynara spp. A bioinformatic pipeline was developed at this regard to analyze publicly available ESTs dataset and identify miRNA candidates in these two related species. Complete EST datasets from Carthamus and Cynara (36323 and 42011 sequences, respectively) were screened looking for putative miRNA targets by means of RNAhybrid (Rehmsmeier, et al.,

Istituto di Genetica Vegetale, CNR, Via G. Amendola 165/A, 70126, Bari, Italy

2004) software and using the mirBase mature miRNA sequences from Arabidopsis as a reference set. Afterward the NCBI Blast algorithm (Altschul et al., 1990) embedded in a Perl script was used to identify the homologous region in Carthamus and Cynara EST datasets. The RNAhybrid results and the homologous region calculated, have been downloaded into a relational database (DBMS MySql) ad hoc developed. We generate the “mapping” of the miRNA target on the “EST homologous region” by means SQL queries. Finally, the signal-to-noise ratio and specificity were assessed using two approaches: in the first one, for each Arabidopsis mature miRNA, the EST dataset was analyzed with a query using 10 di-nucleotiderandomized sequences. In the second approach, 5000 stochastic EST datasets were used as query and real mature miRNAs sequences as referent set. In both cases the predicted targets obtained for the random miRNA sequences were compared with the real miRNA targets.

Results

We found that 75% of the Carthamus and Cynara ESTs shared at least one homologous region (e-value > 10-4), and about 40% of the dataset entries displayed aligned sequences 400 bp long or more. About 8000 Cynara and 9900 Carthamus ESTs did not show common homologous regions, probably due to incomplete trascriptome sequencing data. The RNAhybrid targets analysis showed that 3475 and 4775 Carthamus and Cynara ESTs have at least one microRNA putative target, with 207 and 433 ESTs characterized by more than one target predicted, respectively. About 746 and 742 Cynara and Carthamus ESTs shared an aligned region with a conserved target, while in 525 and 566 ESTs, respectively, the same target was predicted on both sequences, considered as evolutionarily and functionally conserved. Finally, a pool of 95 different target predicted was identified in both species, ascribed to 51 miRNA families. To date, some of these miRNAs appeared specific for Arabidopsis. Statistical assessment showed that 37 targets belonged to 21 miRNA families and have a signal to noise ratio higher than 2, with 50% specificity or higher.

Contact e-mail

domenico.catalano@igv.cnr.it, mariella.finetti@igv.cnr.it