Poster Communication Abstract – 1.21

THE WHOLE GENOME SSR MARKERS DATABASE OF GLOBE ARTICHOKE

PORTIS E.*, PORTIS F.**, VALENTE L.**, LANTERI S.*, ACQUADRO A.*

*) DISAFA - Plant Genetics and Breeding, University of Torino, 10095 Grugliasco (TO)
**) Yebokey – User Experience Design and Web Development, 10134 Torino (TO)

Cynara cardunculus, database, microsatellite, MISA, SSR

Cynara cardunculus L. ($2n=2x=$34, haploid genome size ~1.08Gbp) belongs to the *Asteraceae* family (formerly *Compositae*) and includes three botanical *taxa*: the globe artichoke (var. *scolymus*), the cultivated cardoon (var. *altilis*), and the wild cardoon [var. *sylvestris* (Lamk) Fiori]. Globe artichoke contributes significantly to the agricultural economy of the Mediterranean basin, and in particular Italy, the world's leading producer with an annual yield of 365,000 tons (FAOSTAT 2012). To gain benefit of the first high-quality draft genome assembly of the globe artichoke, unravelled by DISAFA- Plant Genetics and Breeding of the University of Torino within the *Compositae Genome Project* (Scaglione et al. 2014, 58th SIGA Annual Congress), the bulk mining of its markers in totality is imperative and critically required. Here we report on the first microsatellite marker database *CyMSatDB* (*Cynara cardunculus* MicroSatellite DataBase) based on whole genome based SSR mining.

SSR-containing sequences were mined *in-silico* using the *SciRoKo SSR-search* module (http://kofler.or.at/bioinformatics/SciRoKo/), which allows both perfect and compound SSRs to be detected. Any sequence was considered as an SSR where a motif was repeated at least 10 times (1nt motif), six times (2nt) or five times (3-6nt), allowing for only one mismatch. For compound repeats, the maximum default interruption (spacer) length was set at 100bp. From the 13K scaffolds analysed, representing 725 Mb of genomic sequence, about 295.000 SSR motif were identified, including 66.700 cases of compound SSRs. The SSR loci identified were classified on the basis of repeat motif and the number of repeat units: dinucleotides are the most frequent (59.7%), followed by mono- (30.9%) and trinucleotides (6.8%); tetra- penta-and hexanucleotides are rare (1.5, 0.4 and 0.7% respectively).

CyMSatDB, a user-friendly and freely accessible tool, offers chromosome wise as well as location wise search of primers, using *Primer3* implementation. This web application, based on a LAMP solution stack, organizes microsatellites data in a MySQL database and provides an effective and responsive interface developed in PHP. To cater the customized needs of wet lab, features with a novelty of an automated primer designing tool is added. For plant variety identification along with degree of admixture, SSR is always preferred if they are in multiplex mode; such multiplex designing needs much more number of markers to design multiplex with thermodynamic compatibility, which can be accomplished only from our large marker dataset. The feature of user defined primer designing has great advantage in terms of precise selection from each chromosome, defined location, size of amplicons for ease of rapid genotyping.