

The Golden Delicious Apple Genome Sequencing Project: progress and perspectives

Velasco R R¹, Zharkikh A², Troggio M¹, Pruss D², Costa, F.¹, Bhatnagar S², Pindo, M. M¹, Lanchbury, J.², Micheletti, D.¹, Coppola, G.¹, Mraz, A.⁵, Stormo, K.⁵, Tao, Q.⁵, Bogden, R.⁵, Magnago P¹, Komjanc M¹, Harkins T⁴, Malnoy M¹, Cestaro A¹, Baldi, P.¹, Salvi, S., S.¹, Fontana P¹, Gutin², Affourtit J³, Salamini F⁶, Egholm M³, Skolnick M² and Viola, R¹

¹IASMA Research Center, Via E. Mach 1, 38010 San Michele all'Adige (TN), Italy

²Myriad Genetics Inc, 320 Wakara Way, Salt Lake City, Utah 84108, USA

³454 Life Sciences Corp, 20 Commercial Street, Branford, Connecticut 06405, USA

⁴Roche Diagnostics Corp, Roche Applied Science, 9115 Hague Road, Indianapolis, Indiana 46250, USA

⁵Amplicon Express Inc., 2345 NE Hopkins Court, Pullman, Washington 99163, USA

⁶Technology Park Lodi, Località Cascina Codazza, Via Einstein, 26900 Lodi, Italy

Apple is one of the most diffuse fruit crops over the temperate climates, and one of the most important representatives of the Rosaceae large family. In our region of Italy, Trentino Alto-Adige, apple represents the most important agricultural resource. We have therefore concentrated our efforts on apple genome sequencing, selecting the elite cultivar Golden Delicious, grown worldwide and representing over 80% of apple in Trentino.

Our research has the multiple goals of genome assembly, gene identification and annotation, and identification of a maximum number of polymorphisms. Golden Delicious is highly polymorphic with two clearly distinguishable haplotypes, expected to reveal several million SNPs and small indels. They represent a substantial resource for molecular breeding programs, as well as trait and QTL marker association. Based on our previous experience on the large heterozygous grapevine genome, we have used novel algorithms to address this challenge, which will be applied also to the apple genome. A total coverage of 4 genome equivalents of libraries of ascending size sequenced by the Sanger method, coupled with 10 genome equivalents of 454 Life Science™ sequences, will allow us to create an effective genome sequence. Assembly will be based on adding sequences of a BAC libraries and a fosmid library, end-sequenced to assemble large meta-contigs. Contigs will be oriented and ordered on appropriate chromosomes by high throughput marker development and genotyping in an F1 cross of Golden delicious x Scarlet.

Currently, over 3 billions of nucleotides and 1,2 billions of pyrosequencing, Sanger and 454 technique, have been produced. Further 6 billions of nucleotides will be developed to the final goal of 14 genome equivalents of apple (4x Sanger and 10x pyrosequencing, respectively). Following ESTs clustering in tentative consensus (TCs) and TC blast against the genome sequence, 2,000 SNPs of Golden Delicious are currently under development. Sequencing and mapping data will be public available at IASMA, NCBI and GDR databases.

Keywords: Apple, Sequencing genome