Bioinformatics within the IASMA grape project: tools for data mining and sequences annotation - (session: Comparative Genomics and Molecular Evolution)

Fontana P.+, Segala C.+, Toppo S.**, Moser C.+, Grando S.+, Valle G.* and Velasco R.+

+ Istituto Agrario di S. Michele all'Adige (TN)
* CRIBI Univ. di Padova
**Dip. Chimica Biologica Univ. Padova

Although grapevine (Vitis Vinifera L.) is one of the most economically important and widely cultivated crop, grape biology is relatively unknown. Our group is involved in a functional genomic project to discover and determine the function of genes expressed in Vitis vinifera with the aim to provide molecular insights into essential physiological processes like photosynthesis, plant defense and biosynthesis of secondary metabolities. We have based our strategy on sequencing Expressed Sequence Tags (ESTs) obtained from cDNA libraries from different Vitis vinifera tissues such as leaf, root, berry, bud, shoot tips and inflorescence.

The Vitis vinifera species is philogenetically distant to other plants available in public databases. New informations about the organization of Vitis vinifera genome can derive from a wide range comparison of grape sequences to well characterized model organisms such as Arabidopsis thaliana and rice.

The functional characterization of ESTs starts with the clustering process to reduce sequences redundance and pass trough similarity search of public annotated sequences. In order to obtain the most correct annotation we compared the secondary structure prediction of the translated sequences with proteins of known 3D structure.

On the basis of our annotated dataset, some genes of interest, were selected to identify single nucleotide polimorphisms (SNPs), then used in marker-assisted selection and functional mapping.

In parallel, spotting of the amplified ESTs on membrane filters or glass supports will permit true genome-wide sampling of gene expression patterns. Experiments to compare gene expression profiles in leaves on different developmental stages are under way.

All collected data are then stored in a relational database based on a SQL engine. A set of bioinformatic visual tools has been developed to follow all the phases of the project, from the production of cDNA clones, to storage of annotated sequences, microarray profiles and SNPs data. Information retrieve is supported by web interfaces that allow an easy access into the data to users unfamiliar with SQL language.