MS-DMind : multiscale data “minding” for molecular process related to biotic and abiotic stresses: pilot study with the nsLTP superfamily of proteins

Genomique, Edition 2008

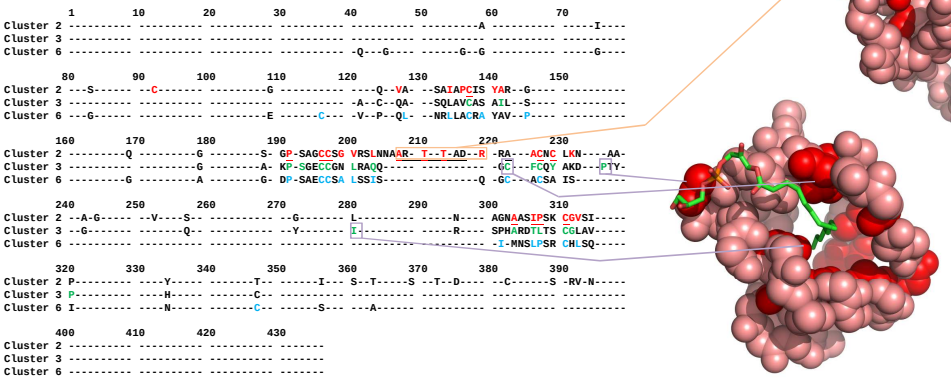
Colloque "Plant Genomics 2012", 2012/04/03-05, Pont-Royal en Provence, France.

Manuel Ruiz[1]
Cécile Fleury[1], Frédéric de Lamotte[1], Marie-Françoise Gautier[1], Jean-François Dufayard[1], Franck Molina[2]

Aims

Understanding biological processes requires managing many complex data sets. MS-DMind project aims at integrating the different ”-omics” spaces corresponding to the different domains of knowledge (genomics, transcriptomics, proteomics, metabolomics, structural biology, molecular dynamics, cellular biology, molecular/biological functions, interactomics and phenomics) in order to allow inferring reliable hypotheses from the data.

The non specific Lipid Transfer Proteins (nsLTPs) show large variations in their sequences, biological roles, quaternary associations and the nature of bound hydrophobic ligands. Besides, they are involved in a large number of biological processes relative to plant development and defense. However they share a conserved eight-cysteine-residue pattern which plays an important role in the structural scaffold. Thus, because its members show a high evolutionary divergence but a conserved common fold, the nsLTP superfamily constitutes a very interesting case of study to validate a method designed for the investigation of protein structure-function relationships.

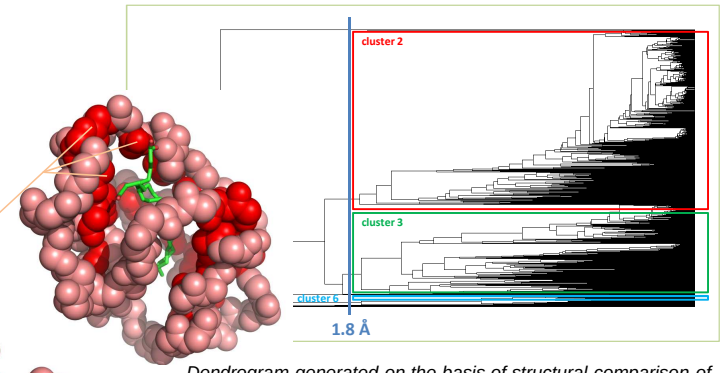


Structure-based sequence alignment of the reference proteins of the 3 main structural clusters.

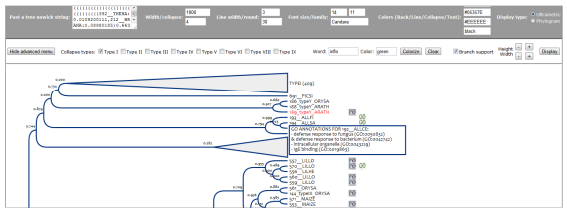
Results

Numerous proteins have been annotated as nsLTPs but we focused the study on the monodomain proteins which present the strict and only nsLTP domain. Eight hundreds mature amino acid sequences belonging to more than 100 plant species have thus been selected. They have been submitted to phylogenic analysis and classified according to sequence identity. For ten of them, three-dimensional structures were available in the Protein DataBank as they had been experimentally determined. Theoretical structures have been calculated for the other proteins, using homology modeling method.

A second classification has been established on the basis of the overall structural alignment of all proteins of the dataset. Using the evolutionary trace method, the observation of structurally equivalent positions allowed identifying either evolutionarily important residues potentially involved in the structural integrity or class-specific conserved residues that may present a functional importance. A functional annotation has been performed manually and the data have been organized and stored in a dedicated multi-scale information system. The comparative structure/function analysis is currently being carried out and is already bringing insights of the ligand binding mechanisms of the nsLTPs.



Dendrogram generated on the basis of structural comparison of all the nsLTPs.



Online phylogenetic tree viewer displaying functional annotations according to Gene Ontology (GO) and Plant Ontology (PO) terms.

Perspectives

In a short term, the developed method will be fully automated and the analysis pipeline will become available to the community. The method has been conceived to be generic enough to handle any protein family, but specific adaptations may need to be made for particular proteins; for example proteins which contain repeats in their sequences and/or 3D structures (alignment tools, structure prediction method). Lipid binding assays are being carried out in SysDiag laboratory and will bring the experimental support to our hypothesis.

CONTACT :
manuel.ruiz@cirad.fr

[1]UMR AGAP CIRAD/INRA, avenue Agropolis, 34398, Montpellier, Cedex 5, France
[2] SysDiag, UMR3145 CNRS/Bio-Rad, 1682 rue de La Valsière, CS 61003, Montpellier, Cedex 4, France