

## Research projects for MSc theses

### Metadata management on a bioinformatics platform for plant transcriptomics

#### Supervisor

Dr Andreas Bohn  
mail: abohn@itqb.unl.pt

#### Laboratory

Systems Biodynamics Group, ITQB-UNL, Oeiras, Portugal  
<http://sbd.itqb.unl.pt>

#### Area of research

Computational Biology / Bioinformatics  
Plant molecular biology

#### Project summary

Cork oak forests (the “montado”) are a unique and emblematic resource for Portugal, due to their ecological and socio-economic significance. They are threatened by damaging policies of land use, by climate change, biotic stresses and wild fires. In order to gain sufficient knowledge to preserve these ecosystems and improve their cultivation and selection, the cork oak has become the object of several national research endeavors in plant biology. Among these initiatives is the Cork Oak Expressed Sequence Tags Consortium (COEC, <http://coec.fc.ul.pt>), an initiative in functional genomics consisting of 12 participating laboratories across Portugal.

The Systems Biodynamics Group participated in this consortium in collaboration with the Bioinformatics Unit of the Instituto Gulbenkian de Ciências, developing a bioinformatics platform for the management and the annotation of cork oak expressed sequence tags (ESTs), entitled CorkOakDB (<http://corkoakdb.org>). One objective in the development of this platform was to grant its extensibility and reusability for future research in cork oak and woody plants.

An important element of preserving data accessibility and reusability is the annotation of raw and processed data with metadata (information about data). In the case of the COEC, a challenge in metadata organization is the large heterogeneity of the trees, conditions, and tissues from which RNA samples were taken for EST sequencing. In order to ease the process for the project partners, the provision of the metadata was realized by weakly structured free-text descriptions.

The objective of the present project is to develop an user-friendly interface for metadata caption, to be implemented in CorkOakDB and eventually other bioinformatics platforms. This should be realized using tools with the potential to become common standard, e.g. the open source ISA metadata tracking tool (<http://www.isa-tools.org>).

The practical aims are thus 1) to annotate the cDNA libraries contained in CorkOakDB in a formatted fashion and with ontological terms extracted from the corresponding free-form texts, and 2) to generalize the created workflow for use in other, similar, projects.