MINUTES FROM THE 2nd MEETING OF THE EDEN PROJECT

Campus de Gambelas, Universidade do Algarve, Faro, Portugal
20-22 June 2007

This was the second meeting of participants in the EDEN project. It is part of one of the deliverables (D.6.2.1) and was also used as a meeting of the Project Steering Committee.

The objectives were to share information of the advances during the first months of the project, and to plan for the next steps in project development.

List of attendants to the meeting:

IMEDEA-UIB:
- Emilio Hernández-García (emilio(_at_)imedea.uib.es)
- E. Alejandro Herrada (alejandro(_at_)imedea.uib.es)
- Alejandro Rozenfeld (alex(_at_)imedea.uib.es)

CCMAR:
- Ester Serrão (eserrao(_at_)ualg.pt)
- Gareth Pearson (gpearson(_at_)ualg.pt)
- Sophie Arnaud-Haond (sarnaud(_at_)ualg.pt)

BioInf Leipzig:
- Konstantin Klemm (klemm(_at_)bioinf.uni-leipzig.de)
- Markus Riester (markus(_at_)bioinf.uni-leipzig.de)
- Camille Stephan-Otto Attolini (camille(_at_)bioinf.uni-leipzig.de)

TKK/LCE:
- Kimmo Kaski (kimmo.kaski(_at_)hut.fi)
- Jari Saramäki (jsaramak(_at_)cc.hut.fi)
- Tapio Heimo (taheimo(_at_)lce.hut.fi)

June 20, 2007

During the morning, Faro and Helsinki teams discussed on the questions relevant to network analysis which the proposed software toolbox should address. It was agreed that the software will analyze fragment/proteins data (allele frequency/genotyping) instead of sequences. Faro team will send to Helsinki (and also deposit on the web site) datasets of the most used molecular markers as examples of the types of data and formats which should be analyzed, as well as a selection of distances measures used in the literature.

During the afternoon session, all the teams discussed on the topics described below. References and data on these topics will be uploaded to the EDEN web.

GENOTYPE DATASETS: microsatellite data for 15 locus of *Fucus* (a nonclonal alga) and 10 locus of *Zimodocea* (a seagrass, both from Mediterranean as from Atlantic meadows) are nearly ready and will be uploaded to the web by the Faro team as soon as possible. Different mating systems are represented. The main interest of analyzing these new data sets with network methods is to elucidate the impact of habitat differences, and of mating systems on the genetic network structure. Can network methodologies help to distinguish a situation of recent hybridization between two species from the alternative scenario of an incipient speciation event? Correlations with geography will be different in the two scenarios.

SOFTWARE TOOLBOX: It is becoming increasingly clear that the project should not only address network *analysis*, but more importantly, network *building*. Developing methodologies for scanning different threshold values instead of a single one, and extracting information from the sequences of outcome, would be important.

COMMUNITY DETECTION: Participants reviewed different methodologies to detect community structure.
From classical genetics, F_{st} and methods based on it are the most important: one performs a statistical test to check the hypothesis that there is panmixia among two different groups. The groups are given a priori. With AMOVA one compares the variance of some genetic property inside groups and between groups, and checking statistically the hypothesis that they are really different groups. Here again the groups should be given a priory and there is a single level of hierarchy. With Bayesian methods, one sets up a model for evolution, and fits the parameters to data, which allow to estimate if there have been population differentiation on the data set analyzed or not.
From recent network theory, there is a bunch of methods based on the modularity approach of Girman and Newman, including Potts-model type. Nevertheless, the most consistent methods seem to be the ones based on Clique Percolation. New approaches by Arenas et al. seem to improve the modularity approach but not strongly enough (only communities of particular size are detected). One should try also to understand recent information theoretic methods.

RESULTS FROM ANALYSIS OF POSIDONIA NETWORKS: Sophie Arnaud-Haond and Alex Rozenfeld described recent results on the analysis of interpopulation and intrapopulation networks built from microsatellite data from *Posidonia*.

June 21, 2007

On Thursday morning. all groups discussed on the following topics:

NETWORK PHYLOGENIES: Markus Riester exposed that the convenient method to build Network Phylogenies is Neighbor-Net. It searches for conflicts in tree-reconstruction methods (violation of triangular identities, e.g.) and solves them by introducing cycles.

SEQUENCE DATA: After discussing several candidates (picoplankton, pathogens from TreeBase, transposable elements …) it was agreed that the sequence data set that would be used as benchmark for Network Phylogenies construction (deliverable D1.1, due in September) would be data from *Fucus*, and from *Cyanobacteria*. The Faro team will collect that and upload to the web as soon as possible.

MODELLING: The need to have some models of population dynamics running to really understand the *networks of genetic similarity* constructed so far (deliverable D3.1) was stressed. It seems that, at the interpopulation level, migration is an important process to consider, being mutations more important only at the largest distances. Development of other models of biological interactions and evolution is also a subject of EDEN. In this context, Camille Stephan-Otto Attolini described her recent work on a model of interacting cells including genes which interact with the environment and express factors leading to cell mobility and metabolism.

On Thursday afternoon:

PHYLOGENETIC TREE TOPOLOGY: Alejandro Herrada exposed results from Palma analyzing a large set of interspecific and intraspecific phylogenies. The topological analysis reveals correlations either in the extinction or in the branching frequency inside related lineages. In fact one would expect that neighbours of misfit taxa (i.e. the closely related taxa) will have a high probability of being misfit and getting extinct as well that the other taxa, and the same may apply to positively selected taxa having suddenly an evolutionary advantage they share. But this feature is missing in the macroevolutionary models usually used to model phylogenetic topology. Competition among lineages may be another factor to be taken into account.

GENETIC FLOW: Konstantin Klemm came out with an idea on how to quantify the amount of gene flow between populations. It was based in a formula describing how well is the set of alleles in one population contained in the set of alleles from another one. The *most contained* set should have exported genes to the *container* one. The proposal seems very interesting and the teams will develop it.

June 22, 2007

During the morning the different teams discussed informally in small groups on different topics including:

- Bayesian coalescence methods (Pritchard software Structure) applied to *Posidonia* populations evolution.
- Circadian rhythms in gene expression in algae, and the impacts of the tide.
- Genetic and spatial data for Cimodocea in the Ebro delta.
- Folding of RNA of Fucus.
- Networks of genetic similarity for *Trypanosoma* cruzi.
- Sampling problems for genetic data arising from clones which may have a power law size distribution.

It was agreed that the next EDEN meeting will take place in Finland next winter.