

## Bioinformatics Project Plan Proposal

Project title: Identification of regulatory elements in oomycetes genomes  
Project supervisor: Michael F. Seidl  
IBIVU supervisor: Jaap Heringa/Anton Feenstra/Sanne Abeln

### Background

*Phytophthora infestans* is the most destructive pathogen of potato and a model organism for the oomycetes<sup>1</sup>. Oomycetes form a distinct phylogenetic lineage of fungus-like eukaryotic microorganisms. They are filamentous, microscopic, absorptive organisms that reproduce both sexually and asexually. Oomycete lifestyle, switch in lifestyle, infection and maintaining infections need tight and cooperative regulation of cohorts of different genes. Currently, nearly nothing is known about up- or downstream regulators of the expression of oomycetes genes. As more genomic sequences of oomycetes in comprehensive quality become available and gene expression data for *Phytophthora infestans* is also accessible, we want to combine the identification of co-regulated genes and prediction of regulatory motifs identified by phylogenetic footprinting. The combination of the bioinformatic approaches might enable us to define regulatory motifs in the proximity of oomycete genes.

### Approach

The underlying approach of this project has been applied for *Plasmodium falciparum*<sup>2</sup>. The project work can be divided into two parts. First, we determine based on the gene expression data the correlation coefficients between all gene pairs, use clustering algorithm to identify co-expressed gene clusters and select relevant gene clusters for further analysis.

In the second part, we will derive the up-/downstream regions of the co-expressed genes and predict orthologous regions in the other genomes. We will identify regulatory motifs in the orthologous regions using already established algorithms. Possible programming shall be written in Perl or Python.

### Aims/objectives

We want to predict a repertoire of possible regulatory motifs in oomycete genomes defined by integration of phylogenetic footprinting and gene co-expression analysis. Furthermore, we also want to identify common patterns within the presence of regulatory motifs in the proximity of genes.

### Timeline

My internship will take place at the Evolutionary Genomics and Integrative Bioinformatics group, which is part of Theoretical Biology and Bioinformatics of University Utrecht. My supervisor is Michael Seidl and I will be spending 4 months on the project, starting on the first day of June 2010. In the month September I will be spending half a day on the project.

|            |   |
|------------|---|
| June:      | Background reading/information gathering/data analyzing |
| July:      | Data analyzing/ motif Identifying/script writing/       |
| Augustus:  | Motif Identifying/clustering/script writing/            |
| September: | Analyzing results/etc.                                  |
| October:   | Writing my project report                               |

1. *Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans*. Haas BJ. *Nature*. 2009 Sep 17;461(7262):393-8. Epub 2009 Sep 9.

2. *Combinatorial gene regulation in Plasmodium falciparum*. van Noort V, Huynen MA. *Trends Genet*. 2006 Feb;22(2):73-8. Epub 2005 Dec 27. Review.