Prof. K. Nieselt, Faculty of Science

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Bioinformatics Master thesis

Deciphering the transcriptome landscape of the antibiotics producing *Streptomyces coelicolor*

Background:
Streptomyces coelicolor is a model organism of the antibiotics producing genus Streptomyces. In the Systems Biology project STREAM the aim was to model gene regulation in this bacterium on various levels – including transcriptomics, proteomics and metabolomics. During that project, our group has conducted almost 350 genome-wide expression experiments using a custom-designed Affymetrix microarray. These experiments encompassed several different time series, during which the wildtype Streptomyces strain M145 as well as different mutants were grown under different growth conditions. STREAM has now been superseded by a follow-up project called SysterACT. The main goal of SysterACT is to develop the model actinomycete Streptomyces coelicolor into a 'Superhost' for the efficient heterologous production of bioactive compounds.
In this project we are currently producing similar transcriptome data, now using RNA-seq, to measure the expression. So far we have produced 60 runs from 2 different fermentations (with 3 replicates each), with about 120 more to come in the next weeks.

Goal of the thesis:
In this master thesis the RNA-seq data produced in SysterACT shall be analysed. Differentially expressed genes along a time series as well as between different time series shall be detected and functionally analysed. Which classes of genes are significantly enriched for differential expression? Are there specific transcribed mutations (allele-specific expression) and if yes, in which genes do they reside? Since the dataset is so rich, many other questions can be pursued during this thesis.

Milestones:

1. Comparison of the expression data derived from mapping against the reference genome with data derived from *de novo* assembly.
   (a) Comparison of several different mapping methods
   (b) Comparison of several different assembly methods
   (c) And how about a hybrid approach?
2. Differentially expressed genes
   (a) along each time series
   (b) between different conditions / time series
3. Analysis of the non-coding RNAs: which are detected, which are differentially expressed?
4. Comparison of the transcript sequences with the reference genome. Are there transcribed mutations (SNPs, indels, ...)?

References:

- Nieselt, Kay and Battke, Florian and Herbig, Alexander and others (2010) The dynamic architecture of the metabolic switch in Streptomyces coelicolor. BMC genomics 11: 10.
- Battke, F and Herbig, Alexander and Wentzel, A and Jakobsen, ØM and Bonin, M and Hodgson, David A and Wohlleben, W and Ellingsen, TE and Nieselt, K and STREAM Consortium (2011) A technical platform for generating reproducible expression data from Streptomyces coelicolor batch cultivations. In: Software Tools and Algorithms for Biological Systems, pages 3–15, Springer.
- Battke, Florian and Symons, Stephan and Nieselt, Kay (2010). Mayday-integrative analytics for expression data. BMC bioinformatics 11:121.
- Battke, Florian and Nieselt, Kay (2011) Mayday SeaSight: combined analysis of deep sequencing and microarray data. PLoS One 6: e16345.

Prerequisites: Good knowledge of scripting languages, good knowledge of (transcriptome) mapping of NGS data, and de novo assembly methods, the visit of the lecture 'Advanced transcriptomics' and / or its respective practical course 'Practical transcriptomics' certainly helps.

Start:
Thesis can be started right away. The data is waiting :).

Contact and inquiries:

Kay Nieselt (kay.nieselt@uni-tuebingen.de)