

## Infopipe

Characterisation of microbial production strains by NGS data analysis integrated in a user friendly web application

<b>Programm / Ausschreibung</b>	FORPA, Forschungspartnerschaften NATS/Ö-Fonds, FORPA NFTE2018	<b>Status</b>	abgeschlossen
<b>Projektstart</b>	17.09.2018	<b>Projektende</b>	16.09.2021
<b>Zeitraum</b>	2018 - 2021	<b>Projektlaufzeit</b>	37 Monate
<b>Keywords</b>	next generation sequencing (NGS), genomics, transcriptomics, enzymology, databases, protein expression, computational biology, bioinformatics		

### Projektbeschreibung

Genetic modification of microbial organisms to produce specific chemical compounds or to express recombinant proteins is routine since several decades. However, any targeted modification (either targeted or random) of the organism's genome can be accompanied by off-target changes of the genomic sequence. Furthermore, these modifications can have considerable desirable and non-desirable effects on gene expression. The characterization of changes across the whole genome and the measurement of the expression of the transcripts has become feasible only in recent years through the development of next generation sequencing technologies. In particular, the technologies introduced by Illumina and Pacific Biosciences facilitate time and cost effective sequencing of genomes (DNA-seq) and transcriptomes (RNA-seq).

The fast and inexpensive generation of sequencing data has now created a new bottleneck: the reliable and timely analysis of rather large amounts of data. To overcome this hurdle, we aim at developing a fully automated, comprehensive and standardized workflow covering the following features:

- Characterization of genome changes compared to a reference genome:  
determination of integration sites and copy number of expression cassettes, identification of SNPs and InDels and their effects on coding sequences, detection of structural variations (larger deletions, insertions and translocations)
- Identification of contamination with bacterial DNA e.g. E. coli genomic DNA integrated into the engineered strains due to transformation with contaminated vector preparations
- Characterization of transcriptome changes compared to a reference transcriptome:  
identification of deregulated genes and pathways, functional grouping of differentially regulated genes to biologic processes
- Generation of standardized, transparent and human readable strain reports
- Management of sequencing data together with the relevant strain meta information
- Integration of this data into a database of available promoters
- General applicability of the new tools for different microbial hosts although the focus of this thesis will be on P. pastoris

The features will be accessible via a user friendly web application, which can be used without requiring any prior bioinformatics knowledge. For validation and demonstration purposes (including publication of new findings for employed model strains) these new bioinformatics tools will be applied for the analysis of exemplary P. pastoris expression strains, mainly generated by random integration and showing varying efficiencies in recombinant protein expression with the goal to

provide rational input for next generation expression strain designs.

## **Projektpartner**

- bisy GmbH