

Workshop So-Long (FOR 2281): Transcriptome analyses

10-15 October, University of Freiburg



Venue: Univ. Freiburg, Institute for Biologie II (Zoology), Evol. Biology & Ecology, Hauptstr. 1, 79104 Freiburg

Rooms: Seminar Room Biology I; CIP Pool Biology II. We work with a virtual machine (VMware) based on Ubuntu 16.06 with necessary software, tools, packages installed.

Participants: *So-Long members*: Abel Bernadou, Matteo Negroni, Denise Aumer, Alice Séguret, Boris Kramer, Marisa Rodrigues, Daniel Elsner, Manuel Monroy Kuhn; *additional collaborators Muenster / Mainz / Regensburg*: Maximilian Körner, Pinar Güler, Philipp Sprenger, Claudia Gstöttl

Lecturers: Daniel Elsner, Manuel Monroy Kuhn, Mark Harrison, Barbara Feldmeyer, Karen Meusemann

Overall organization: Florentine Schaub, Karen Meusemann with help of Daniel Elsner (computer facilities, VM)

Workshop Schedule (daily, 9 a.m. - ca. 5 p.m. Lunch daily around 12 :30 pm)

Sunday 9th arrival in the evening at the accommodation

Monday 10th morning:

Introduction in the Seminar Room

Basics: Introduction CIP Pool: [Linux / working on command line](#)

afternoon:

Quality check, Trimming: [FastQC](#), [Trimmomatic](#) incl. explanation of fastq format & and few basic terms (scaffolds, contigs)

short introduction of tools [BUSCO \[DOGMA\]](#) (possibility to for completeness against a reference set of genes / proteins)

evening: optional: Dinner at the traditional restaurant "Schlappen" (<http://www.schlappen.com/>)

Tuesday 11th morning:

Introduction in differential gene expression analyses in the Seminar Room

Introduction into R: basic commands in [R](#)

Mapping of raw reads on available genomes with [TopHat2](#) [TopHat will be not maintained in the future but is still commonly used. alternative tools: HiSat2, Kallisto].

afternoon:

adjust output of TopHat (with [HTSeq](#)); preparation for DESeq2

Statistical Analyses I: PCA as quality and outlier check; statistical analyses of mapping (read counts) with [PCA \(R\)](#)

What is differentially expressed, what is not, meaning fold change values, thresholds, incl. one factor and multifactor analysis [DESeq2](#)

generating lists of (shared) genes, listing subsets, comparisons, Venn Diagrams and meaning (e.g. [venny](#), [R](#))

evening: optional: guided city tour & dinner at the restaurant "Weber's Weinstube" (<http://www.webers-weinstube.de/>)

Wednesday 12th morning:

Statistical Analyses I (continued)

afternoon:

Statistical Analyses I (continued)

discussion of problems, results, analyses design, challenges etc. (Seminar Room)

evening: optional: Dinner in one of the numerous restaurants of Freiburg

Thursday 13th morning:

Statistical Analyses II: Functional Analyses: Introduction & general overview

Analyses examples with [InterProScan](#) (incl [Pfam](#)), [GOTerms](#) ([topGO](#)), discussion of difficulties, challenges, pitfalls (e.g. quality of annotated genomes, databases etc.)

afternoon:

Statistical Analyses II: Enrichment analyses (incl. Fishers exact test) using [R](#) ([topGO](#)) including boxplots, term-clouds molecular functions, biological processes, what is meaningful to compare

evening: Original Badisches Dinner with "Kässpätzle & Salat" hand-made by Flo in the Biology I building ☺, drinks will be organized as well

Friday 14th morning:

Statistical Analyses II: Pathway analyses based on GO-Terms (lists, [KEGG](#); [KAAS](#))

afternoon:

plenary discussion, problems, conclusions, closing remarks (Seminar Room)

evening: optional: Dinner at the restaurant "Waldsee" (<http://www.waldsee-freiburg.de/>)

Saturday 15th morning: Checkout

day-tour: optional: trip to the "Schauinsland" with guided tour "Bergwerksmuseum" and Lunch at the restaurant "Die Bergstation"

List of installed tools (some extra packages / dependencies are installed, most R packages are installed via Bioconductor, we will use the majority, not all)

Python3, Python2.7, Java (open jdk 1.8.0_91), NCBI Blast+ (2.31), Hmmer 3.1b2, Exonerate, Sqlite3, Pfam Scan (current release, v. 30), BUSCO, R 3.3.1, RStudio FastQC, Trimmomatic v.0.36, IGV 2.3.81, InterproScan, SAMtools, bedtools, pandas vim, bowtie 2 v. 2.2.6, TopHat v2.1 [HiSat2, BBmap], HTSeq, [eXpress], GOAtools [Bioconductor, within R]: DESeq2, topGO, SeqGSEA, Gosummaries, [edgeR] [Jupyter Notebook, Seaview, Mafft; not installed: HiSat2, Trinity, Cufflinks, DOGMA, Orthograph]