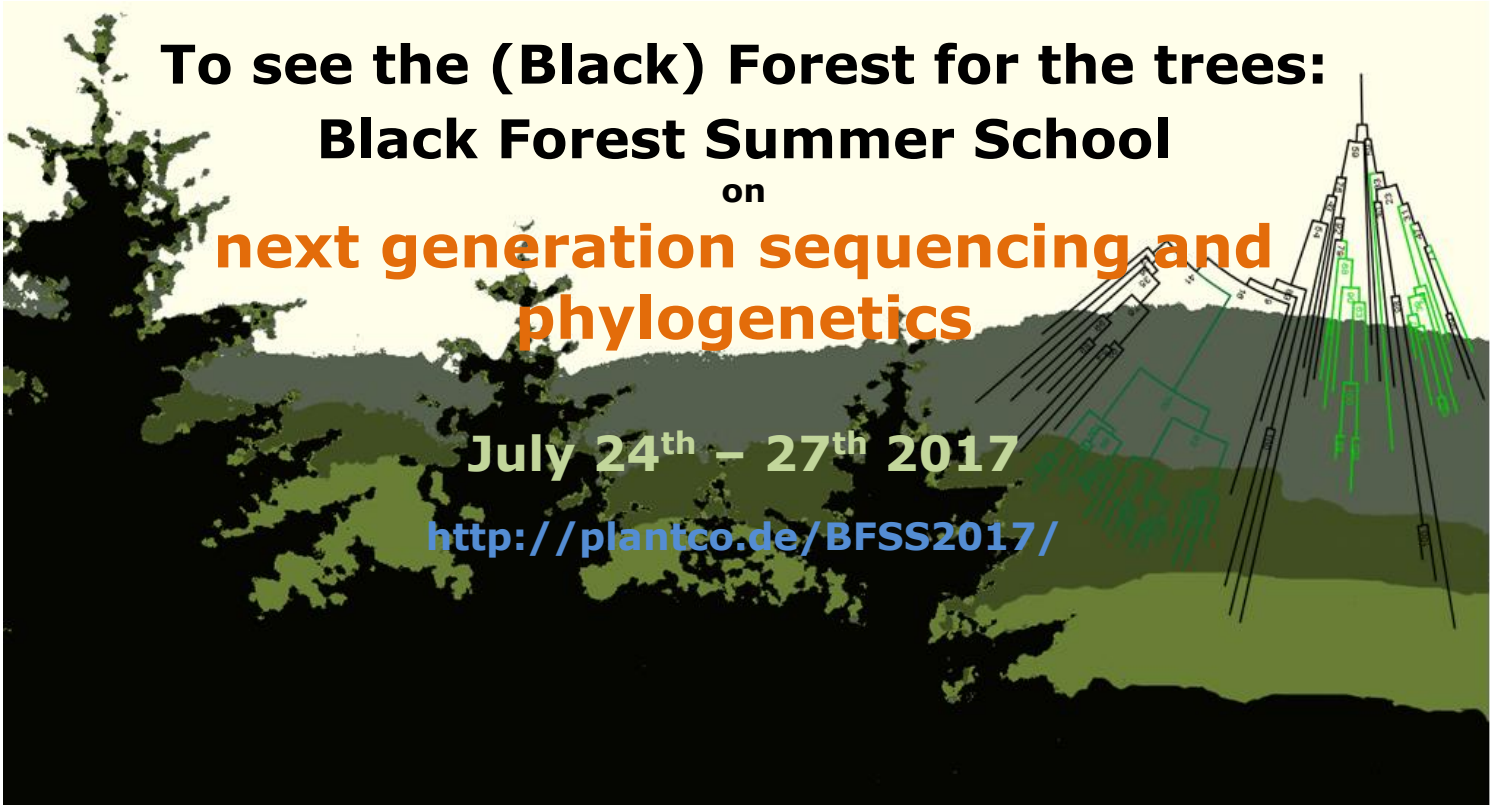


Abstract Book

Edited by Stefan A. Rensing
Marburg, Germany, July 2017

Venue: Leistungszentrum Herzogenhorn (Black Forest Highlands, Germany)
~1,300 mtrs above sea level

Acknowledgements

The Black Forest Summer School 2017 is organized by Plantco.de e.V. (Freiburg/Marburg, Germany) as a joint action with the University of Marburg and EVOLTREE; support by the companies shown below is gratefully acknowledged.

Thank you very much!

Contents

Program

Venue: Leistungszentrum Herzogenhorn

All workshops and talks as well as poster sessions and industry exhibits take place in the hall (gymnasium). Breakfast, lunch, supper, coffee breaks and evening entertainment are located in the dining rooms.

Monday July 24th

15:15, 16:15, 17:15	Bus shuttles from Feldberg-Bärental train station to venue
19:00	Welcome reception with food and beverages
20:15	Opening remarks
	Welcome lecture: "**Plastid genome evolution in parasitic plants**" (Julia Naumann, PennState/University of Technology Dresden)
later	*"Moss cocktail workshop"*

Tuesday July 25th

9:00 – 9:30	Lecture I: **NGS data generation: tools, terms and pitfalls** (Stefan Rensing, University of Marburg)
9:30 – 9:45	Discussion
9:45 – 10:30	Lecture II: **NGS data processing and differential expression analysis** (Fabian Haas & Noe Fernandez Pozo, University of Marburg)
10:30	Coffee break

10:45 – 11:00 …continuation of lecture II
11:00 – 11:15 Discussion

11:15 – 11:45 Lecture III: **Introduction to Galaxy**
(Anika Erxleben, University of Freiburg)
11:30 – 11:45 Discussion

12:00 Lunch

13:00 – 14:15 *Poster session I* with coffee

14:15 – 15:15 Lecture IV: **Mapping: short read alignment to a reference**
(Noe Fernandez Pozo, University of Marburg)
15:15 – 15:30 Discussion

15:45 – 16:15 Lecture V: **Principles of phylogenetics**
(Stefan Rensing, University of Marburg)
16:15 – 16:45 Discussion

16:55 Group foto

17:00 – 21:00 Excursion

17:00	*Departure to Seebuck*
17:15	*Excursion (hike) to Feldsee*
18:30	*Black Forest Food at Raimartihof*
20:30	*Return to Herzogenhorn*

18:00 Supper (for those not on excursion)

Wednesday July 26th

9:00 – 9:45 *Oral session I*
T1 Jake Chandler, London
"SeedAdapt – Unravelling the Molecular Mechanisms Controlling Germination and Dormancy in *Aethionema arabicum* – a Model Species for Diaspore Heteromorphism"
T2 Katharina Winkel, Hannover
"Identification and Characterization of Enzymes and Transporters Involved in the Ureide Biosynthesis in the Soybean Nodule"
T3 Norico Yamada, Konstanz
"Comparative transcriptional analyses in three intermediate-stage plastids in dinoflagellates"

9:45 – 10:15 Lecture VI: **How to infer my own phylogeny?**
(Stefan Rensing, University of Marburg)
10:15 – 10:30 Discussion

10:30 Coffee break & industry exhibits

11:00 – 11:45 Lecture VII: **Phylogenomics: From gene trees to the species tree**
(Julia Naumann, PennState/University of Technology Dresden)
11:45 – 12:00 Discussion

12:00 Lunch

13:00 – 15:00 *Poster session II* with coffee & industry exhibits

15:00 – 15:45	*Oral session II* T4 Veli Vural Uslu, Heidelberg "Dynamics of Nutritional Stress Response" T5 Christian Wever, Düsseldorf "Creating a new Crop – genetic evaluation and collection of *Silphium perfoliatum* L." T6 Anne Christina Genau, Marburg "Single copy ortholog identification in alternation of generation/sexual reproduction in bryophytes"
15:45 – 16:15	Lecture VIII: **Genedata Selector and metagenomics of biogas production process** (Steffen Fehrmann, Genedata)
16:15 – 16:45	Lecture IX: **Using CLC Genomics Workbench for analysis of NGS data** (André Koper, Qiagen)
17:00 – 17:45	Round table sessions: topics t.b.d. on-site
18:00	Supper
19:00 – 20:00	Round table sessions cntd.: topics t.b.d. on-site
later	*Farewell party*

Thursday July 27th

9:00 – 9:45	Lecture X: **Metagenome and network analyses** (Sina Beier, University of Tübingen)
9:45 – 10:00	Discussion
10:00	Coffee break
10:30 – 11:15	Lecture XI: **ChIP-seq analyses** (Ioana Lemnian, University of Halle)
11:15 – 11:30	Discussion
11:30	Wrapup, prizes and concluding remarks
12:00	Lunch, end of summer school
13:30	Bus shuttle to Feldberg-Bärental train station

Lecture Abstracts

L0 Plastid genome evolution in parasitic plants

Julia Naumann
PennState/University of Technology Dresden
jxn25@psu.edu

Plastid genomes of photosynthetic flowering plants are usually highly conserved in both structure and gene content. However, the plastomes of parasitic and mycoheterotrophic plants are released from selective constraints due to the reduction or loss of photosynthetic ability. Their plastomes typically show decay in gene content and sometimes structure, too. Phylogenetic markers of those reduced and divergent plastomes are particularly difficult to sequence traditionally. Whole plastome sequences obtained via "high throughput sequencing" became extremely valuable for obtaining insights into the evolution of parasitic plants. However, reconstructing and annotating plastomes of extreme parasites is still challenging due to their divergence. To date, there are about 40 plastome sequences of plants that range from facultative hemiparasites to holoparasites. We now know that independent lineages show highly similar patterns of gene loss. The holoparasitic lineages have split from their photosynthetic sisters up to 100 MYA, but yet they show a surprisingly consistent gene set that allows us to determine the genes that are essential for the retention of the plastid chromosome. What is this minimal gene set of a functional plastome and could it be lost entirely in a plant?

L1 NGS data generation: tools, terms and pitfalls

Stefan Rensing
Plant Cell Biology, University of Marburg, Germany
stefan.rensing@biologie.uni-marburg.de

In this lecture, I shall first compare highly parallel ("NGS") sequencing technologies with Sanger and each other and talk about pros and cons of the different technologies.

I will then define technical terms that will be needed for comprehension of the summer schools' topics, and will talk about some methodological details like fastq format, phred scores, fragment sizes, regional bias, multiplexing, paired ends, mate pairs etc.

Some potential pitfalls will be highlighted and finally I will introduce some tools that might be useful but are not covered by individiual workshops.

L2 NGS data processing and transcriptome assembly

Fabian Haas & Noe Fernandez Pozo
Plant Cell Biology, University of Marburg, Germany
fabian.haas@biologie.uni-marburg.de & noe.fernandezpozo@biologie.uni-marburg.de

Next Generation Sequencing (NGS) has become the method of choice to address a multitude of biological questions in all areas of natural sciences from ecology to biotechnology. NGS datasets are used to reconstruct genomes, unravel population and evolutionary relationships, build expression profiles and many more applications. This workshop provides a detailed introduction how to process NGS datasets by using state of the art programs and methods. Short tutorials will cover all necessary steps from raw sequencing data to cleaned good quality reads which are mandatory for follow-up applications. Further, the focus of this workshop will be on one of these applications, transcriptome assembly. All relevant key points and key terms will be highlighted to discuss how to assemble contigs and to do further analysis steps such as transcript prediction.

L3 The Galaxy framework as a unifying bioinformatics solution for HTS data analysis

Anika Erxleben
University of Freiburg, Germany
erxleben@informatik.uni-freiburg.de

The Freiburg Galaxy Team is member of the German Network for Bioinformatics Infrastructure (de.NBI) and aims to provide comprehensive bioinformatics services to users in life sciences research, industry and medicine. Within this network, we are part of the RNA Bioinformatics Center (RBC) and we are responsible for supporting RNA related research in Germany. In this talk we will present our analysis platform Galaxy which makes advanced bioinformatics software accessible to biologists directly by providing an intuitive web interface to these applications while fostering reproducibility through the automatic creation of re-runnable protocols of each analysis. We describe the use of Galaxy for HTS data analysis in genomics, proteomics, imaging and metabolomics. Focusing on the perspective of a biological user, we will demonstrate the benefits of Galaxy for these analyses, as well as its value for software developers seeking to publish new software.

L4 Mapping: short read alignment to a reference

Noe Fernandez Pozo
Plant Cell Biology, University of Marburg, Germany
noe.fernandezpozo@biologie.uni-marburg.de

Read alignments are the starting point for most re-sequencing projects using next generation sequencing technologies. No matter if you want to identify single nucleotide polymorphisms between your reference and another individual/population of the same species, find structural rearrangements or do transcriptome/expression analyses between individuals, you always have to start with mapping the reads against an available reference. This workshop aims at giving a brief introduction to the underlying principle of establishing short read alignments and discuss in this context problems that usually occur during the alignment process. We will also discuss what kind of information we can extract from different read alignment patterns in both whole genome and transcriptome short read data and introduce different tools commonly used in the different analyses. In the end we will look at the limits of short read alignment analysis and give an outlook on future perspectives in next generation sequencing.

L5 Principles of phylogenetics

Stefan Rensing
Plant Cell Biology, University of Marburg, Germany
stefan.rensing@biologie.uni-marburg.de

How does sequence evolution occur and why does it allow to infer gene and species phylogenies? What are substitution matrices and why do we need them all over the place?

Everybody knows BLAST, everybody uses it. But how do you define homology from a BLAST result? By E-value? By bit score? By alignment length? Or by % identity?

Once you have determined homology, how do you go on? How do you generate an alignment and how do you visualize and curate it?

Based on an alignment, which methods are there to infer phylogenetic trees and what are their pros and cons? Finally, how do you interpret a phylogenetic tree?

L6 How to infer my own phylogeny?

Stefan Rensing
Plant Cell Biology, University of Marburg, Germany
stefan.rensing@biologie.uni-marburg.de

This workshop will essentially cover the same topics as lecture 4, namely the principles of phylogenetics. We shall talk about why and how we can infer molecular phylogenies, look at the term homology in detail, and aim to understand duplication events and what the difference between gene and species trees is.
The workshop will present you with conceptual questions with regard to phylogeny and will mainly be a discussion forum. I will also introduce useful tools that you can use for retrieving homologs, for aligning and visualizing them, for tree inference, model selection and tree visualization.
Finally, we will leave space for your specific questions – phylogenetic methods are an expected aim of many or most of your projects (given the topic of the summer school), so we will try and come up with helpful suggestions for your work.

L7 Phylogenomics: From gene trees to the species tree

Julia Naumann
PennState/University of Technology Dresden
jxn25@psu.edu

High throughput sequencing (HTS) has taken phylogenetics to a new level: phylogenomics. Hand crafted alignments of Sanger sequences and the obtained individual gene trees reached their limitations when it came to very closely related taxa (i.e. low level phylogenetics). The expectations on HTS data (i.e. multiple genes from the nucleus and the other organelles) were very high to resolve relationships that were not solved in previous efforts. However, with big data came more conflicts and challenges. How to identify and extract good phylogenetic markers from genomic data? How to computationally and bioinformatically process many genes from many taxa? How to get trustworthy, highly resolved and supported phylogenies that reflect the species tree? There are two major approaches to deal many markers: concatenation or tree reconciliation. The latter is using a coalescence-based approach where a species tree is estimated based on individual gene trees. In this interactive lecture we will be discussing the highs and lows of both methods and strategies to get the best species tree.

L8 Genedata Selector and metagenomics of biogas production process

Stefan Fehrmann
Genedata
steffen.fehrmann@genedata.com

During anaerobic digestion (AD) of biogas plants, the microbial population is key to methane production in biogas plants, yet its composition is poorly characterized. In collaboration with the Fraunhofer IGB in Stuttgart, we uncovered growth and transcriptional adaptations of the microbial community due to temperature changes and total reactor running time during AD of slurry and maize silage. Over two hydraulic retention times (HRTs), three biogas reactors were assessed for microbial composition and gene expression reflecting pathway activities using whole genome shotgun (WGS) and metatranscriptome sequencing, respectively. The reactors were fed with maize silage and yielded a stable methane production early on but showed large shifts in microbial diversity over time. Remarkably, an early temperature drop had only a minimal effect on the population's diversity but rather a significant effect on gene expression, especially on the archaeal methane metabolism. Here I will present our findings and highlight how I used a combination of tools to mine the raw sequencing data, and subsequently manage the processed data in Genedata Selector to detect and confirm trends during the analysis.

L9 Using CLC Genomics Workbench for analysis of NGS data

André Koper
Qiagen
Andre.koper@qiagen.com

L10 Metagenome and network analyses

Sina Beier
University of Tübingen
sina.beier@uni-tuebingen.de

This lecture will introduce how NGS and bioinformatics can lead us from an environmental sample to detailed information about the microbial community in that sample. We will discuss the workflow of data analysis as well as the interpretation and evaluation of the results.

This way we can answer the common taxonomic metagenomics questions (Who is there, how abundant are the different taxa?) and provide a functional analysis (What are they doing, how do microbes interact with each other and their environment?).

We will mainly use MEGAN 6 Community Edition, but also discuss additional tools and methods.

L11 ChIP-seq analyses

Ioana Lemnian
University of Halle
ioana.lemnian@informatik.uni-halle.de

Chromatin immunoprecipitation followed by high-throughput DNA sequencing (ChIP-seq) is a powerful method for locating DNA-protein interactions, like transcription factors binding sites, or histone modification sites across genomes. After a brief introduction to the general ChIP-seq workflow, we will focus on the peak calling, the computational method of identifying genomic regions that have been enriched with mapped reads from the ChIP-seq experiment. We are going to discuss the statistical models, the assumptions used in state-of-the-art peak callers, and continue with possible downstream analyses like the annotation of the peaks, motif enrichment, and de-novo motif discovery. At the end of this workshop you will have gotten an overview of the biological questions that can be answered by ChIP-seq experiments and of the existing tools and best practices for the data analysis.

Contributed Talk Abstracts

T1 SeedAdapt – Unravelling the Molecular Mechanisms Controlling Germination and Dormancy in *Aethionema arabicum* – a Model Species for Diaspore Heteromorphism

Chandler J.O.[1], Graeber K.[1], Merai Z.[2], Mittelsten Scheid O.[2], Rensing S.A.[3], Grosche C.[3], Wilhelmsson P.K.I.[3], Lenser T.[4], Theissen G.[4], Sperber K.[5], Mummenhoff K.[5], Nguyen T.-P.[6], Schranz M.E.[6], Strnad M.[7], Leubner-Metzger G.[1,7]

1. School of Biological Sciences, Royal Holloway University of London, Egham, Surrey, TW20 0EX, United Kingdom, 2. Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, Vienna Biocenter, 1030 Vienna, Austria, 3. Plant Cell Biology, Faculty of Biology, University of Marburg, 35043 Marburg, Germany, 4. Department of Genetics, Friedrich Schiller University, 07743 Jena, Germany, 5. Department of Biology, University of Osnabruck, 49076 Osnabruck, Germany, 6. Biosytematics Group, Wageningen University, 6708 PB Wageningen, The Netherlands, 7. Laboratory of Growth Regulators, Centre of the Region Hana for Biotechnological and Agricultural Research, Palacky University and Institute of Experimental Botany, Academy of Sciences of the Czech Republic, 78371 Olomouc, Czech Republic
Jake.Chandler@rhul.ac.uk

An adaptation to harsh unpredictable environments is the trait of heterodiaspory, the ability of a single plant to produce diaspores with different morphologies and physiological behaviors. *Aethionema arabicum* is an ideal model for studying the mechanisms of diaspore heteromorphism that produces two distinct fruit and seed morphs on the same infructescence. A large fruit morph containing multiple mucilaginous M+ seeds dehisces at maturity leading to the dispersal of the bare M+ seeds. In contrast, a smaller indehiscent fruit morph contains a single non-mucilaginous M– seed which is dispersed encased in its fruit coat via abscission. The diaspores differ in germination physiology with M– seed exhibiting more dormancy partly due to the fruit coat. *Ae. arabicum* combines bet-hedging with phenotypic plasticity. The numbers and ratios of each fruit morph depend on maternal growth conditions with M+ seed development favored under stress. Germination of both morphs also has a temperature response dependent on parental plant growth conditions. We hypothesized that the hormonomes, epigenomes, and transcriptomes of the dimorphic diaspores provide syndrome×stress memories that are dispersed to establish the next generation. Integrating results obtained through RNAseq and hormone profiling we will provide a holistic view on fruit-seed stress physiology during dormancy release and germination relevant to the ecological significance of diaspore dimorphism in *Ae. Arabicum*.

T2 Identification and Characterization of Enzymes and Transporters Involved in the Ureide Biosynthesis in the Soybean Nodule

Katharina Winkel, Marco Herde, Claus-Peter Witte

Institute of Plant Nutrition, Leibniz Universität Hannover
winkel@pflern.uni-hannover.de

Two major transport systems exist by which legumes, in symbioses with rhizobia, can export the fixed atmospheric nitrogen from their nodules. While temperate legumes mainly transport amides, tropical plants show higher export levels of ureides. The ureide biosynthesis pathway is not completely elucidated and some enzymes and especially transporters are still not identified. For the identification of candidates potentially involved in the ureide biosynthesis, publicly accessible RNAseq data were compiled comparing amide-exporters, *Medicago truncatula* and *Lotus japonicus*, and ureide-exporters, *Phaseolus vulgaris* and *Glycine max*, on transcriptional basis. Assuming that the ureide biosynthesis is transcriptionally up-regulated in nodules compared to roots of ureide exporting plants and not in amide exporting plants, this data set should yield a list of candidate genes potentially involved in ureide biosynthesis. A phosphatase was found in this list, which was already shown by us to dephosphorylate XMP to xanthine *in vitro*. This enzyme might be directly involved in ureide biosynthesis. This and further candidate proteins will be biochemically characterized in vitro and CRISPR/Cas constructs will be generated and used in a soybean hairy root system for in vivo analysis. Metabolic profiling of transgenic roots and nodules and the biochemical characterization will be combined to gain information on the biological function of those candidate genes.

T3 Comparative transcriptional analyses in three intermediate-stage plastids in dinoflagellates

Norico Yamada, Peter G. Kroth

Department of Biology, University of Konstanz
norico.yamada@uni-konstanz.de

It has been accepted that plastids of all phototrophs except Paulinella chromatophora derived from a cyanobacterium by an endosymbiosis event. As this event occurred about a billion years ago, the process of endosymbiotic conversion to an organelle cannot be observed experimentally. However, some eukaryotic organisms show intermediate stages of plastid acquisition, including a part of dinoflagellates. Such dinoflagellates, called "dinotoms", possess plastids that originate from diatoms; however, the diatoms still possess their own nuclei, mitochondria, and ribosomes. Recently, we revealed that dinotoms possess different species of endosymbiont diatoms, depending on the dinoflagellate species. From this, we hypothesised that the stage of endosymbiotic conversion might be different within the dinotoms.Here, we plan to study the degree of endosymbiotic conversion of the DNA sequence level. We study three species of dinotoms, which show different endosymbiosis stage, a temporarily retained stage (kleptoplastids), a permanently retained stage, and a permanently retained stage with a reduced endosymbiotic organelle. We have extracted mRNA from two out of them. Transcript sequencing was conducted by poly-A selected strand-specific cDNA synthesis to check the gene expression level of the both nuclei from hosts and endosymbionts. By analyses of the sequences, we hope to identify key genes for retention of an endosymbiont permanently, and for losing endosymbiont organelles.

T4 Dynamics of Nutritional Stress Response

Veli Vural Uslu, Rüdiger Hell

Center for Organismal Studies, Heidelberg
veli.uslu@cos.uni-heidelberg.de

The dramatic drop in the soil nutrient content worldwide in arable lands has recently become a major problem that directly impinges on food production and economy. Plants cope with suboptimal environmental conditions by activating their stress response mechanisms. Therefore, improving stress response in plants has a tremendous potential in improving crop yield in stress conditions. All organisms from bacteria to humans and to plants enhance their stress response when they are pretreated with the stress factor. The mechanisms that underlie this phenomenon called "priming" have remained almost fully unidentified. Our preliminary data with the model plant *Arabidopsis* implicate the presence of nutrient priming, which at least in part operates via DNA methylation. To understand the principles of this priming, we employ a parallel liquid chromatography and non-biased sequencing approaches to monitor and relate the dynamics of metabolite, transcript and chromatin landscapes, respectively. These chromatin profiles will also yield the first histone landscape of col-0 mature root to lay the ground for further investigation of root chromatin in different genetic backgrounds and environmental conditions. In addition, using reverse genetics with DNA modification mutants sheds light on the mechanisms behind priming for the first time. Yet, beyond DNA or histones modifications, we implicate RNA degradation machinery in the regulation of transcriptional dynamics during priming.

T5 Creating a new Crop – genetic evaluation and collection of *Silphium perfoliatum* L.

Christian Wever[1], Lukas Becker[1], Elena Pestsova[1], Martin Hoeller[2], Ralf Pude[2], Peter Westhoff[1]

1. Developmental and Molecular Biology of Plants, Heinrich-Heine University, Duesseldorf,
2. INRES - Nachwachsende Rohstoffe, Friedrich-Wilhelms-University, Bonn
christian.wever@hhu.de

The cup plant (*Silphium perfoliatum* L.) is like sunflower a member of the Asteraceae and native to the US. Due to its high biomass yield, Silphium is a promising alternative for energy maize. Silphium perfoliatum is a perennial plant with a broad range of ecological benefits: a long flowering period, an efficient growing under low-input agriculture, less weed killers and soil erosion. Until today almost no breeding attempts have been made for domestication of the cup plant and all field trials were done with the few available European genotypes of unknown origin. A selected part of the European genepool was analysed via Tunable Genotyping By Sequencing (tGBS) technology (Data2Bio®). The data showed a broad genetic diversity for this non-model species but also one clear pattern of population stratification. The genotypes from the Ukraine indicated a decreased genetic diversity, but simultaneously the highest frequency of unique SNPs. For making sure that future breeding will be based on sufficent germplasm diversity, a plant hunting trip to the US has been performed. Based on herbarium data 40 accessions of *Silphium* were collected, which are covering its native distribution. The subsequently aims are now to analyse these new genotypes and compare them with the European genotypes on two levels, genetically via tGBS and phenotypically via field trials. Based on this data and an association mapping study, it will be possible to evaluate best genotypes and traits for agriculture.

T6 Single copy ortholog identification in alternation of generation/sexual reproduction in bryophytes

Anne C. Genau, Per K.I. Wilhelmsson, Kristian K. Ullrich, Christopher Grosche, Fabian Haas, Stefan A. Rensing

Plant Cell Biology, University of Marburg, Germany
genau@staff.uni-marburg.de

In course of evolution land plants evolved a peculiar haplodiplontic life cycle in which both the haploid gametophyte and the diploid sporophyte are multicellular. However, within land plants the dominant generation (gametophyte or sporophyte) changes gradually from bryophytes to angiosperms. With this, following the alternation of generation on molecular level is difficult to analyze in flowering plants due to their highly derived life cycle having reduced their gametophytic tissue to basically a few (hidden) cells. Bryophytes, in contrast, as haploid-dominant (gametophyte) plants, have more easily tractable generations and have already the fundamental regulatory networks for switch of vegetative and reproductive growth. This switch is most often regulated by transcription associated proteins (TAPs) which regulate the alternation of the transcriptional program. In many cases the genes encoding TAPs are single copy genes. To avoid a dosage imbalance of their gene products it is selected against the duplicates of those fundamental regulators after duplication events. We were interested in identifying those single copy genes with special emphasis on embryo development. Combining orphan identification (proteinortho), protein-family categorization (TAPscan), literature search and phylogeny we were able to identify single copy genes potentially involved in embryo development in bryophytes. Here we present the workflow, also adaptable for other genes of interest.

Poster Abstracts

P1 Towards establishment of a genetic transformation system in charophyte green algae – Zygenematophyceae

Hong Zhou, Dieter Hanelt, Klaus von Schwarzenberg

University of Hamburg
fbnv870@uni-hamburg.de

The Zygenematophyceae occupy an important phylogenetic position as the sister group of land plants. However, little is known about this class of green algae, especially concerning functional genomics. Until now, no genomic sequence data for Zygenematophyceae are available and no robust genetic experimental platform is established. We use our Microalgae and Zygnematophyceae Collection Hamburg (MZCH) as a source to screen for streptophytic algae strains allowing high-throughput transformation as a prerequisite for potential model organisms. Finaly, *Cosmarium regenesii* and *Spirogyra pratensis* were chosen as fast growing candidate species representing the orders Desmidiales and Zygnematales. For both, axenic cultures were established. Different zeocin resistance cassettes driven by heterologous promoters have been successfully introduced to the nuclear genome of *Cosmarium regenesii* using microparticle bombardment. However, none of the heterologous promoters seem to work in *Spirogyra pratensis*. Currently, isolation of two homologous promoters and trials of genetic transformation of *Spirogyra pratensis* are undergoing.

P2 The role of specific post-translational modifications of trehalase on drought stress tolerance in *Arabidopsis thaliana*

Le Cong Huyen Bao Tran Phan, Patrick Van Dijck

KU Leuven, Belgium
tran.phan@kuleuven.vib.be

Trehalose is a non-reducing disaccharide and a well-established osmoprotectant in microorganisms. Nevertheless, only trace amounts of trehalose are present in the majority of higher plants, suggesting that trehalose does not act as compatible solute in plants. Trehalase is an enzyme participating in a three-step enzymatic trehalose biosynthesis pathway. Most plant genomes contain large trehalose biosynthesis gene families, while the trehalase enzyme is encoded by a single gene. Recently, lowering the level of trehalose by overexpressing the endogenous trehalase AtTRE1 gene has been shown to result in improved survival during drought stress in Arabidopsis plants, while Attre1 mutants with increased trehalose levels were more drought stress sensitive. The precise mechanism of trehalase regulation in responding to stress conditions is not yet known. Here, we present that AtTRE1 is posttranslationally modified. Putative phosphorylation sites at position Ser71 and Thr128 or potential glycosylation sites at position Asn193 and Asn292 were changed to Ala. These point mutations significantly lower the specific activity. This finding suggests that trehalase is regulated by both phosphorylation and glycosylation.

P3 Search for a putative ribose transporter linking nucleotide catabolism with sugar recycling in Arabidopsis

Rebekka Schroeder, Claus-Peter Witte

University of Hannover
rschroeder@pflern.uni-hannover.de

Nucleoside catabolism is a vital degradation mechanism of RNA for the recycling of nutrients like nitrogen and ribose. Recently, we characterized a ribokinase (RBSK), a member of the PfkB kinase family, which phosphorylates the ribose moiety released by Nucleoside Hydrolase 1 (NSH1). Ribose-1-phosphate is subsequently entering the pentose phosphate pathway or the salvage pathway of nucleosides. Metabolomic analyses of several mutants of the nucleoside degradation pathway showed that the accumulation of ribose in the RBSK mutant is exclusively caused by the continuous degradation of nucleosides, which takes place in the cytosol. However, we observed a co-localisation of the plant RBSK with several downstream enzymes in the chloroplast, raising the question how ribose released by nucleoside catabolism is transported into the chloroplast stroma for further processing. So far the permease, transporting ribose into the plastidic lumen is unknown in plants.This study aims to identify the plastidic ribose transporter increasing our understanding of ribose processing in the cell.

P4 Enhancer evolution of a homeobox gene shaped leaf diversity

Francesco Vuolo[1], Remco Mentink[1], Mohsen Hajheidari[1], C Donovan Bailey[2], Dmitry Filatov[3], Miltos Tsiantis[1]

[1] Max Planck Institute for Plant Breeding Cologne, [2] New Mexico State University, [3] University of Oxford
mentink@mpipz.mpg.de

We identified Reduced Complexity (RCO), a homeobox TF both necessary and sufficient for leaf complexity in Brassicaceae. RCO duplicated from Late Meristem Identity 1 (LMI1), a leaf margin growth repressor that contributes to serration development in Arabidopsis. RCO however, displays a more restricted expression pattern at the base of leaflets in Cardamine and was deleted in Arabidopsis, contributing to leaf simplification. Comparing LMI1 and RCO 5\' upstream regions of multiple Brassicaceae, we identified a 500 bp enhancer that regulates specific expression. The RCO enhancer underwent increased base substitution, indicating that positive selection likely contributed to its novel expression domain. However, relevant base changes that led to altered RCO expression are still unknown. We plan to whole genome sequence additional Brassicaceae species and create multiple alignments of LMI1 and RCO genes using target capture. Allowing us to study LMI1/RCO cluster structure and predict TF binding sites. Reporter genes / genetic assays can be used for functional validation. Furthermore, through phylogenetic trees we can place this in an evolutionary context and see how the expression changes progress. Finally, we will perform ChIP-seq experiments to explore differential binding of TFs predicted to bind. Our findings will increase our understanding of how specific base changes alter organ morphology and are then selected by evolution.

P5 Usage of mating-type loci to identify strains from the mycophenolic acid producer *Penicillium brevicompactum*

Yasaman Mahmoudjanlou, Tim A. Dahlmann, Ulrich Kück

Ruhr-Universität Bochum, Lehrstuhl für Allgemeine und Molekulare Botanik, Universitätsstraße 150, Building ND7/122, 44801 Bochum, Germany
yasaman.mahmoudjanlou@rub.de

In heterothallic ascomycetes, two non-allelic idiomorphs, termed mating-type loci (MAT1-1 or MAT1-2) specify the sex of individual strains and control mating as well as sexual reproduction. They carry genes encoding transcription factors, which have either an alpha- or a high mobility group- DNA-binding domain. After cloning of MAT loci by using PCR primers for conserved sequences flanking the MAT loci, we discovered the genomic organization of the MAT1-2-1 and MAT1-1-1 open reading frames from at least 13 strains of filamentous fungus *Penicillium brevicompactum*. This is of substantial biotechnological and medical importance, because of its ability to produce the immunosuppressant mycophenolic acid. The open reading frames were verified by cDNA cloning and sequencing. Comparing MAT amino acid sequences with those from other Penicillium species revealed a high homology in the DNA binding domains. However other regions of the proteins were less similar. Beside 2 molecular markers, Internal transcribed spacer (ITS), ß-tubulin, MAT loci were also used for taxonomic characterization of 36 wild type strains provided from different culture collections. Remarkably, from 36 strains, previously described as P. brevicompactum 16 were identified as another related species. Our data suggests that MAT loci can be used as a novel molecular marker to identify strains from *P. brevicompactum* and point to the potential of this gene for the taxonomic identification of other Penicillium species.

P6 Transcriptomic and Hormonal Regulation of Agricultural Weed Seed Dormancy and Germination as Targets for Weed Management Tools

Thomas Holloway[1], Kazumi Nakabayashi[1], Safina Khan[1], David Stock[2], Tim Hawkes[2] & Gerhard Leubner-Metzger[1]

[1] The Seed Biology Place, Royal Holloway University of London, Egham Hill, Egham, Surrey, UK, TW20 0EX,
[2] Syngenta Ltd. Jealott's Hill International Research Centre, Bracknell, RG42 6EY
Thomas.Holloway.2012@live.rhul.ac.uk

Understanding the germination and dormancy of agricultural weed seeds is of key importance to predicting and controlling weed emergence in the field. Many weed species have become highly adapted to different agronomic systems, synchronising their germination with the cropping cycle to avoid herbicides during fallow periods. The aim of this project is to use a cross-species approach to identify the key regulators of seed dormancy and germination in a variety of economically important weed species. Using this framework we investigate how novel dormancy release and germination stimulant compounds could be used to manipulate weed seed germination to reduce weed fitness. Having developed a physiological understanding of the dormancy and germination of the weeds, we are aiming to use transcriptomic and targeted metabolomic approaches to identify novel regulators of dormancy and germination that are affected by the application of germination manipulating compounds. The key challenges in this project will include working with non-model species with little or no genome information available (particularly with polyploid species) and making meaningful comparisons between transcriptomes of related species. A further challenge will be to integrate findings from metabolomic and transcriptomic datasets. The project has the potential to improve the prediction of weed emergence in the field and to aid the development of new agrochemicals that target the weed soil seed bank.

P7 The mitochondrial genome of the fern *Vittaria lineata* documents abundant repeats and recombination, frequent invasion by chloroplast DNA and horizontal gene transfer from bacteria

Simon Zumkeller, Monika Polsakiewicz and Volker Knoop

University of Bonn, IZMB
zumkeller@uni-bonn.de

Chloroplast genomes are easily assembled from next generation sequence (NGS) data given their conserved structure and stoichiometric dominance in plant tissues. In contrast, vascular plant mitochondrial genomes are more complex for assembly given their variable gene complements, lack of gene syntenies and their recombinant structures affected by invasion of chloroplast or nuclear DNA or even via horizontal gene transfer (HGT). Leptosporangiate ferns remain the last major clade of land plants for which no mitochondrial genome has been assembled, likely owing to a combination of the above facts. We here present the chloroplast and mitochondrial genomes of a first leptosporangiate fern, *Vittaria lineata* (Pteridaceae, Polypodiales). The *V. lineata* cpDNA is expectedly highly conserved in structure, whereas we find a very complex, non-orthodox mtDNA with numerous repeats separating single copy regions. Nevertheless, the Vittaria mtDNA has a rich gene complement lacking only the ccm gene suite. Importantly, it features several sequences indicative of bacterial HGT, mainly from Rickettsiales, but also including tRNA genes of chlamydial origin. A disrupted gene structure is found for rrn26, which is split in two parts to be reassembled on RNA level via a trans-splicing group I intron. Some 400 events of C-to-U and U-to-C RNA editing are expected for the maturation of the chloroplast transcripts and ca. 1000 such events for the mitochondrial transcriptome of *V. lineata*.

P8 A better understanding of the evolution of the behavior of carnivorous plants

Alberto Davila Lara, Axel Mithoefer, Ralf Omuller

Max Planck Institute for Chemical Ecology, Jena
adavila-lara@ice.mpg.de

Carnivory in higher plants for additional nutrient supply is a fascinating topic in plant science and has been known for centuries. Recent studies orientated in the understanding of the evolutionary behavior of carnivorous plants have been increased and revealed new findings. The sequenced genome of *Cephalotus follicularis*, Oxalidales, revealed that this species has an independent evolution history\; furthermore, by analyzing phylogeny-based orthologue-paralog genes it was suggested that *Nepenthes alata* and *C. follicularis* are closely related. This result is based on a comparison with the sequences of proteins: Purple acid phosphatase, PR-1 like protein, RNase T2 and Beta-1,3-glusanase, Thaumatin-like protein and GH19 chitinase. However, this does not prove the same when these two same species were analyzed taking only as reference sequences of the following proteins Aspartic protease, Clase III peroxidase, and GH18 chitinase where N. alata and C. follicularis are distant in phylogenetic analysis. Using pitcher plants of the genus Nepenthes, we will address several important questions in order to gain a better understanding of molecular events during carnivory in plants and its evolution. As well, Ortholog genes involved in the digestion process in *Nepenthes alata* that are shared with other carnivorous and non-carnivorous plants will be identified and studies for their evolution

P9 Physiological and molecular mechanisms of *Lepidium sativum* seed longevity and vigour

Marta Pérez, Bliss Buttery and Gerhard Leubner

School of Biological Sciences, Plant Molecular Sciences, Royal Holloway, University of London, Egham, Surrey. TW20 0EX, U.K.
Marta.Perez@rhul.ac.uk

Seeds have a central role in almost all the food supply chains important to human and animal survival. Seed longevity and vigour are key quality-defining traits that determine plant species adaptability to changing environmental conditions and seed performance after storage. Seed vigour is a complex trait with genetic and environmental components determining the rate and uniformity of seed germination and seedling growth under harsh weather conditions and abiotic stresses. The mechanisms that regulate seed germination have been widely studied; however, very little is known about the mechanisms related to seed longevity and vigour. Moreover, the relationship between seed pigmentation, dormancy and vigour, and their roles in environmental adaptation are poorly understood. Therefore, the objective of the present study is to investigate the physiological and molecular mechanisms underpinning the regulation of seed longevity and vigour in different accessions of *Lepidium sativum* by using accelerated ageing techniques in combination with naturally aged lines and a natural transparent testa (tt) mutant line. The results show different sensitivity of the lines to the ageing treatment related to the seed coat colour. Moreover, differential expression patterns of genes involved in genome integrity and stability processes were observed among the different treatments and lines indicating that this is a promising system for further comparative transcriptomic analyses.

P10 Fertility variation between *Physcomitrella patens* ecotypes

Rabea Meyberg, Manuel Hiss, Jens Westermann, Lucas Schneider, Rebecca Hinrichs, Stefan A. Rensing

University of Marburg
rabea.meyberg@biologie.uni-marburg.de

This project focuses on the factors influencing sexual reproduction of *Physcomitrella patens* ecotypes Reute, Villersexel and different Gransden variants. We find that gametangiogenesis and ripening of the gametangia occur in a similar time frame and without any gross morphological differences. Yet, the sporophyte analysis revealed a severe and significant reduction of sporophytes per gametophore in the Gransden strains, one of which is close to sterile. Subsequent crossing experiments showed that Gransden strains were capable of developing sporophytes on up to 95% of the gametophores if fertilized by Villersexel or Reute spermatozoids. This observation points out apparently recurring male fitness problems in the Gransden strains and is reinforced by a spermatozoid analysis revealing aberrant morphology and strongly reduced motility of Gransden spermatozoids. A comparative transcriptomic analysis between antheridia of the studied ecotypes is planned in order to gain detailed knowledge of expression differences and putative SNPs of genes expressed in the antheridia. In parallel, candidate genes were selected for their putative involvement in the process of sexual reproduction. They were validated by qPCR and analyzed regarding their genetic and epigenetic differences between ecotypes. Deletion mutants are being generated and will be analyzed. We aim to gain an understanding of factors influencing the efficiency of sexual reproduction in the moss *Physcomitrella patens*.

P11 Targeting the grana membranes with chimeric Light Harvesting Complex Stress Related Proteins

Christo Schiphorst, Alberta Pinnola, Manuel Benedetti, Roberto Bassi

University of Verona
christo.schiphorst@univr.it

Plants and algae require light for growth, but excess light can be harmful. They have therefore developed a wide variety of protection mechanisms, commonly referred to as Non-Photochemical Quenching (NPQ). The most important component of NPQ is energy quenching (qE), it is activated within seconds by a decreased pH in the thylakoid lumen. The precise mechanism of how the energy is quenched remains unknown, but there are several generally accepted theories that all depend on an interaction between chlorophylls and carotenoids. Two proteins are essential for the activation of NPQ, PSBS and LHCSR, which can be found in plants and algae respectively. The exception to this rule is the moss P. patens, an evolutionary intermediate between algae and plants, which expresses both PSBS and LHCSR. The main difference between PSBS and LHCSR is that PSBS does not bind any pigments. Therefore PSBS itself cannot be the site of quenching. LHCSR on the other hand contains approximately 8 chl a and 4 xanthophylls, therefore making it likely that LHCSR is also the site of quenching. Initial expression of LHCSR1 in A. thaliana located the protein in stroma membranes only, hampering the interaction with PSII and yielding a reduced level of functional complementation. In this project we complement the quenching activity in A. thaliana npq4 strains by using a chimeric PpLHCSR1-AtLhcb1.1 construct. The purpose of the chimeric construct is to direct LHCSR1 to the grana membranes where PSII is located.

P12 DOE JGI Plant Flagship Gene Atlas: *Physcomitrella patens*

Fabian B. Haas[1], Pierre-François Perroud[1], Jeremy Schmutz[2,3], Jane Grimwood[2], Gary Stacey 4, Avinash Sreedasyam[2], Stefan A. Rensing[1]

[1] Plant Cell Biology, Faculty of Biology, University of Marburg, Germany [2] HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA [3] DOE Joint Genome Institute, Walnut Creek, CA, USA 4 Divisions of Plant Science and Biochemistry, National Center for Soybean Biotechnology, University of Missouri, Columbia, MO, USA
fabian.haas@biologie.uni-marburg.de

In the context of the Department of Energy's Joint Genome Institute Plant Flagship Gene Atlas project, we present the dataset generated for the model moss P. patens. Our in-house RNA-seq data management pipeline was used to map and associate sequence reads with differently expressed genes (DEGs). Comparison across all replicates of treated and untreated stages allowed us to define stage specific transcript sets. Diverse and specific expression patterns emerge from the hormonal and metabolic treatments.The homogeneity of protonemata makes this tissue, in principle, the simplest to experimentally manage and should be easy to compare between laboratories. Yet, replication of the same conditions between laboratories underscored the difficulty to perform such comparison. Even with identical medium and growth conditions, different laboratories generated significantly different transcriptome profiles.Normalization is difficult for any expression profiling methodology and not well developed for RNA-seq data yet. A methodology to make use of External RNA Controls Consortium (ERCC) spike-ins for cross-experiment normalization is under development. So far, DEG calling using spike-ins as a normalization is significantly different from the DEG calling with in silico normalization methods.All presented experiment results are based on the first set of 99 samples. Another 72 samples have been processed and will significantly broaden the range of conditions and perturbations.

P13 Gaining insights into the molecular mechanisms behind allelochemical effects of myrigalone A on *Lepidium sativum* seed germination

Kazumi Nakabayashi[1], David Stock[2] and Gerhard Leubner[1]

[1]Royal Holloway University of London, Egham, TW20 0EX, United Kingdom [2]Syngenta Ltd. Jealott's Hill International Research Centre, Bracknell, RG42 6EY, United Kingdom
Kazumi.Nakabayashi@rhul.ac.uk

Myrigalones are naturally occurring flavonoids found in exudates in leaves and fruits of *Myrica gale*. They are rare flavonoids from the class of dihydrochalcones, and some of which show several biological activities including antioxidant or phytotoxic activity. Our previous studies showed that leachate from *M. gale* fruits and myrigalone A (MyA), a major myrigalone compound in the leachate, inhibited *Lepidium sativum* germination in a dose-dependent manner. Application of MyA also results in atypical endosperm rupture in the germination process and eventually pale green seedlings. Preceding analysis showed MyA treatment did not change ABA levels, whereas the levels of bioactive GAs were significantly lower in imbibed seeds with an increased accumulation of direct precursors. These data suggested that MyA potentially inhibited GA3 oxidases, which lead to the reduced level of GA. However, interestingly, exogenous application of GA did not overcome retarded germination or pale green phenotype, suggesting deficiency of bioactive GAs is not a direct cause of these MyA effects. To identify the molecular mechanisms underlying these allelopathic properties of MyA, transcriptome and comprehensive hormonome analyses are underway. These omic analyses are expected to provide a basis to target key genes or pathways explaining MyA biological activities.

P14 Genome improvement of *Aethionema arabicum*

Thu-Phuong Nguyen[1], Erik van der Berg[1], Cornelia Muehlich[2], Stefan A. Rensing[2], M. Eric Schranz[1]

[1] Biosystematics Group, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands;
[2] AG Rensing, Faculty of Biology, University of Marburg, Karl-von-Frisch Straße 8, 35043 Marburg, Germany
cornelia.muehlich@biologie.uni-marburg.de

Aethionema arabicum is part of an early branching sister group of thecore Brassicaceae. It is displaying fruit and seed dimorphism and isliving in variable environments. It is therefore suited to study genomicevolution and adaption to abiotic stress. With these two aims in mind, we set to improve its existing genome draft. We started our project withthe present genome draft (v2.5) that was established usinga progressive assembly approach. First, preexisting scaffolds wereassembled by the use of a newly created genetic map. The genetic map wasgenerated using recombinant inbred lines of crosses between twowild ecotypes, Cyprus and Turkey. This approach led to the detection ofeleven linkage groups, reducing the number of scaffolds from 3,166 to 2,990. Second, we successfully generated long read sequences for bothecotypes using two approaches: PacBio and MinION sequencing. Thesubsequent superscaffolding performed with PBjelly further reduced thenumber of scaffolds by four. In parallel, the number of Ns was loweredfrom 25,785,896 to 13,790,434 (by 46.5%) while raising the number ofhigh quality bases of the genome from 170,236,799 to 189,660,650. Mostof the improvements generated by PBjelly approach were related to gapfilling. With these improvements, we plan to release the new v3.0 A.arabicum genome.

P15 RNA editing from C to U and in reverse: investigating the hornwort *Anthoceros agrestis*

Philipp Gerke[1], Péter Szövényi[2], Mareike Schallenberg-Rüdinger[1], Volker Knoop[1]

[1]Institut für Zelluläre und Molekulare Botanik (IZMB), Bonn, Germany, [2]Department of Systematic and Evolutionary Botany, University of Zurich, Zurich, Switzerland
PhilippGerke@gmx.de

RNA editing converting cytidines into uridines is a common feature of land plant organelles. The key factors targeting specific editing sites are nuclear-encoded, RNA-binding Pentatricopeptide Repeat (PPR) proteins. All 13 organellar RNA editing sites have been assigned to their respective PPR-type editing factors in the model moss *Physcomitrella patens*. Intriguingly, all of them feature a carboxyterminal "DYW"-type cytidine deaminase domain likely performing the C-to-U conversion. The occurrence of "reverse" RNA editing among plants has remained unclear and no U-to-C editing factor is known. We now find reverse U-to-C editing to occur exclusively in ferns, lycophytes and hornworts calling for a novel model system for functional studies. Among hornworts, the most ancient plant clade showing reverse editing, *Anthoceros agrestis* is emerging as such a model. We have assembled the chloroplast and mitochondrial genomes for A. agrestis and predict a total of 737 C-to-U and 948 U-to-C edits in the two organelles. Intriguingly, the *A. agrestis* nuclear genome encodes numerous DYW-type PPR proteins with a characteristic variant of their essential "PG motif". This "WW-type" PG variant is absent in mosses, liverworts and seed plants, which show C-to-U editing exclusively. We suggest establishing A. agrestis as a first model organism featuring both types of RNA editing towards a future functional characterization of first U-to-C editing factors.

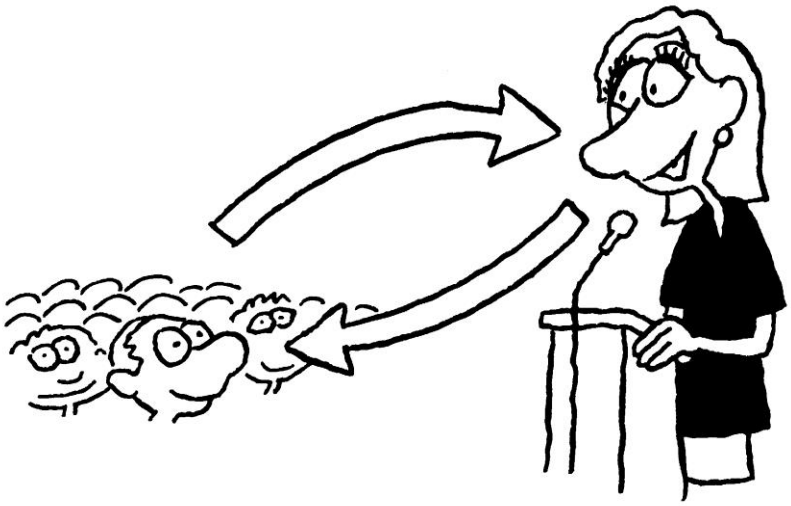
List of participants

last name	first name	affiliation	country	email -address
Beier	Sina	University of Tübingen	Germany	sina.beier@uni-tuebingen.de
Becker	Lukas	Heinrich-Heine-University	Germany	lukas.becker@hhu.de
Chandler	Jake	Royal Holloway University of London	United Kingdom	Jake.Chandler@rhul.ac.uk
Davila Lara	Alberto	Max Planck Institute for Chemical Ecology	Germany	adavila-lara@ice.mpg.de
Erickson	Jessica	Martin Luther University	Germany	jessica.erickson2525@gmail.com
Erxleben	Anika	University of Freiburg	Germany	erxleben@informatik.uni-freiburg.de
Fehrmann	Steffen	Genedata	Switzerland	Steffen.Fehrmann@genedata.com
Fernandez Pozo	Noe	University of Marburg	Germany	noe.fernandezpozo@biologie.uni-marburg.de
Genau	Anne Christina	University of Marburg	Germany	genau@staff.uni-marburg.de
Gerke	Philipp	University of Bonn	Germany	PhilippGerke@gmx.de
Gong	Xin	University of Goettingen	Germany	xgong@gwdg.de
Grosche	Christopher	University of Marburg	Germany	cgrosche@biologie.uni-marburg.de
Haas	Fabian	University of Marburg	Germany	fabian.haas@biologie.uni-marburg.de
Holloway	Thomas	Royal Holloway University	United Kingdom	Thomas.Holloway.2012@live.rhul.ac.uk
Koper	André	Qiagen	Germany	Andre.koper@qiagen.com
Lemnian	Ioana	University of Halle	Germany	ioana.lemnian@informatik.uni-halle.de
Mahmoudjanlou	Yasaman	Ruhr University Bochum	Germany	yasaman.mahmoudjanlou@rub.de
Mentink	Remco	Max Planck Institute for Plant Breeding Research	Germany	mentink@mpipz.mpg.de
Meyberg	Rabea	University of Marburg	Germany	rabea.meyberg@biologie.uni-marburg.de
Muehlich	Cornelia	University of Marburg	Germany	cornelia.muehlich@biologie.uni-marburg.de
Nakabayashi	Kazumi	Royal Holloway University of London	United Kingdom	Kazumi.Nakabayashi@rhul.ac.uk
Naumann	Julia	Technical University of Dresden	Germany	jxn25@psu.edu
Omrani	Maryam	San Raffaele Telethon Institute for Gene Therapy	Italy	omrani.maryam@hsr.it
Pereira Torres	Denise	Technical University of Munich	Germany	d.torres@wzw.tum.de
Perez	Marta	Royal Holloway University of London	United Kingdom	Marta.Perez@rhul.ac.uk
Phan	Le Cong Huyen Bao Tran	University of Leuven	Belgium	tran.phan@kuleuven.vib.be
Rensing	Stefan	University of Marburg	Germany	stefan.rensing@biologie.uni-marburg.de
Renteria	Zaida	Institute of Parasitology, University of Leipzig	Germany	zaida_melina.renteria_solis@uni-leipzig.de
Schiphorst	Christo	University of Verona	Italy	christo.schiphorst@univr.it
Schroeder	Rebekka	Leibniz University Hannover	Germany	rschroeder@pflern.uni-hannover.de

last name	first name	affiliation	country	email -address
Seidl	Anna	Universitaet fuer Bodenkultur Wien	Austria	anna.seidl@boku.ac.at
Tintelnot	Stefanie	BMG Labtech	Germany	stefanie.tintelnot@bmglabtech.com
Uslu	Veli Vural	Center for Organismal Studies at Uni Heidelberg	Germany	veli.uslu@cos.uni-heidelberg.de
Wang	Yi	Max Planck Institute for Plant Breeding Research	Germany	wang@mpipz.mpg.de
Wever	Christian	Heinrich-Heine-University	Germany	christian.wever@hhu.de
Winkel	Katharina	Leibniz University Hannover	Germany	winkel@pflern.uni-hannover.de
Xiao	Zhenggao	University of Neuchatel	Switzerland	zhenggaoxiao@gmail.com
Yamada	Norico	University of Konstanz	Germany	norico.yamada@uni-konstanz.de
Zhang	Bing	University of Goettingen	Germany	zhangbing061989@126.com
Zhou	Hong	University of Hamburg	Germany	fbnv870@uni-hamburg.de
Zumkeller	Simon	University of Bonn	Germany	zumkeller@uni-bonn.de

Black Forest Summer School 2017

Interactive Lectures



The Black Forest Summer School concept

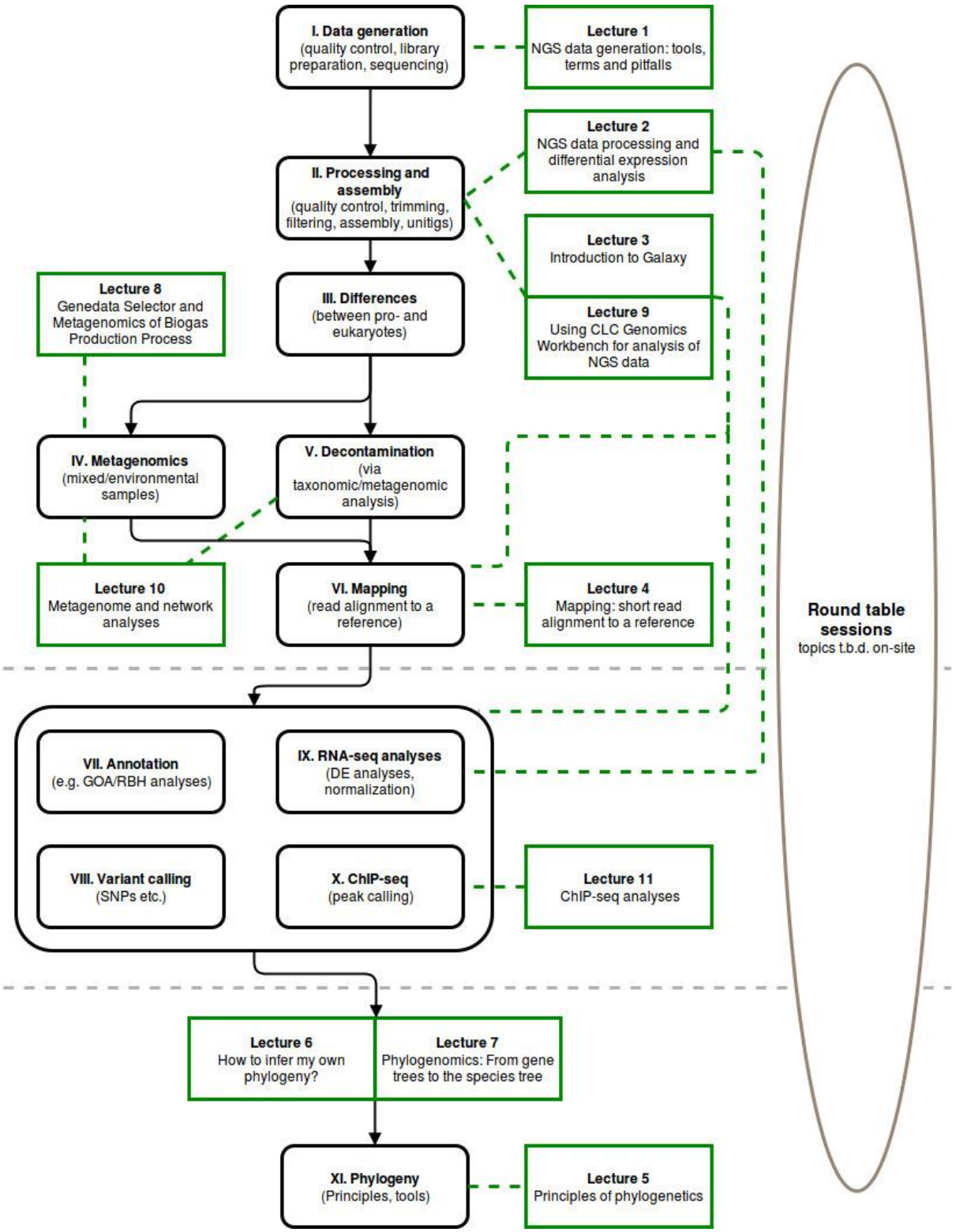
We will not teach you how to program – we will teach you which existing tools are out there to process NGS data and to infer phylogenies.

BFSS workshops are interactive lectures, in which the participants are encouraged to ask questions and will be charged with questions / exercises by the lecturer.

Lectures may include **life demo sessions** (but typically no hands-on part).

You will receive digital materials (on your memory stick) that will enable you to repeat what was demonstrated at the school, as well as helpful lists of software applications.

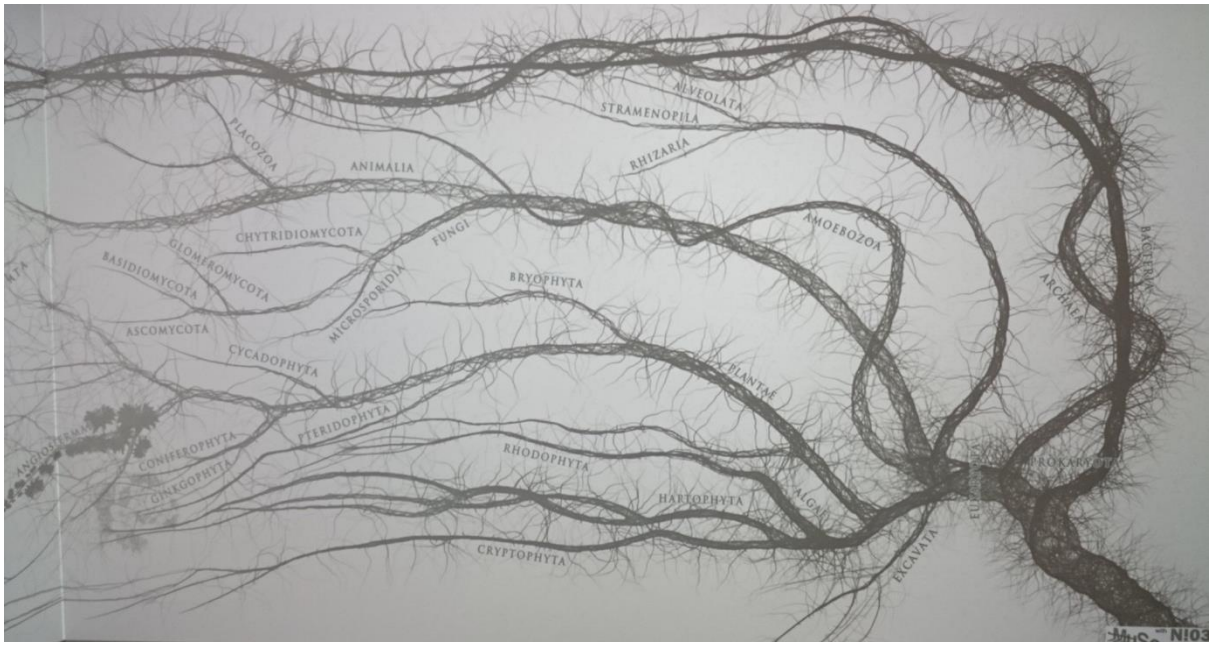
While we encourage participants to bring their own laptop/tablet, **computer use is not necessary** to follow the schools' program.


I. Data generation
(quality control, library preparation, sequencing)
Lecture 1
NGS data generation: tools, terms and pitfalls
Lecture 2
NGS data processing and differential expression analysis
II. Processing and assembly
(quality control, trimming, filtering, assembly, unitigs)
Lecture 3
Introduction to Galaxy
Lecture 9
Using CLC Genomics Workbench for analysis of NGS data
Lecture 8
Genedata Selector and Metagenomics of Biogas Production Process
III. Differences
(between pro- and eukaryotes)
IV. Metagenomics
(mixed/environmental samples)
V. Decontamination
(via taxonomic/metagenomic analysis)
Lecture 10
Metagenome and network analyses
VI. Mapping
(read alignment to a reference)
Lecture 4
Mapping: short read alignment to a reference
VII. Annotation
(e.g. GOA/RBH analyses)
IX. RNA-seq analyses
(DE analyses, normalization)
VIII. Variant calling
(SNPs etc.)
X. ChIP-seq
(peak calling)
Lecture 11
ChIP-seq analyses
Lecture 6
How to infer my own phylogeny?
Lecture 7
Phylogenomics: From gene trees to the species tree
XI. Phylogeny
(Principles, tools)
Lecture 5
Principles of phylogenetics
Round table sessions
topics t.b.d. on-site

L6 How to infer my own phylogeny?

Stefan Rensing
Plant Cell Biology, University of Marburg, Germany
stefan.rensing@biologie.uni-marburg.de

This lecture will essentially cover the same topics as lecture 5, namely the principles of phylogenetics, but from a more practical perspective. We shall talk about why and how we can infer molecular phylogenies, look at the term homology in detail, and aim to understand duplication events and what the difference between gene and species trees is.
The lecture will present you with conceptual questions with regard to phylogeny and will mainly be a discussion forum. I will also introduce useful tools that you can use for retrieving homologs, for aligning and visualizing them, for tree inference, model selection and tree visualization.
Finally, we will leave space for your specific questions – phylogenetic methods are an expected aim of many or most of your projects (given the topic of the summer school), so we will try and come up with helpful suggestions for your work.



['Universum', Bremen]

i) Background knowledge: [see whether you remember all the terms correctly]

What is (sequence) homology?

Qualitative, exclusive *vs.* Quantitative, relative

Substitution matrix, conservative exchange

What are homologs, paralogs, orthologs and xenologs? Mind the frame of reference…

Bit score, E-value – or something else?

The Twilight Zone of protein alignments

ii) Some tasks

Query: Your favourite *Arabidopsis* protein
Desired subjects: Other plant proteins
Question: which kind of BLAST?

Query: Your favourite maize protein
Desired subjects: homologs from animals
Question: which kind of BLAST/parameters?

Query: Somatic embryogenesis receptor kinase (SERK)
Desired subjects: other SERKs
Question: how to filter?

Is this homologous?

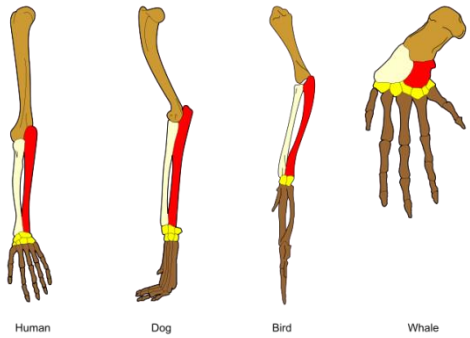
```
sp|P38851|YHV5_YEAST HYPOTHETICAL 143.6 KD PROTEIN IN SPO16-REC104
INTERGENIC REGION
            Length = 1228

 Score = 32.5 bits (72), Expect = 0.097
 Identities = 17/50 (34%), Positives = 26/50 (52%), Gaps = 3/50 (6%)
Query: 10   LNLMARKMKDTDSEEELKEAFKVFDKDQNGFISAAELRHVMTNLGEKLTD 59
            LNL+ RKM D D E  +   FK FD+D+N +      + +     + LT+
Sbjct: 1119 LNLLTRKMMDNDQENPV---FKRFDEDKNAYQYKGTRQEIAIKRNQVLTE 1165
```

Explain the highest values in the PAM250 matrix by biological knowledge.

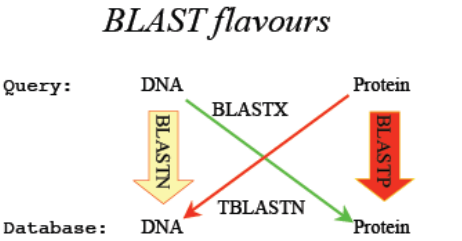
How many conservative exchanges in this hit?

```
Identities = 186/288 (65%), Positives = 240/288 (83%), Gaps = 0/288 (0%)
```

iii) Material



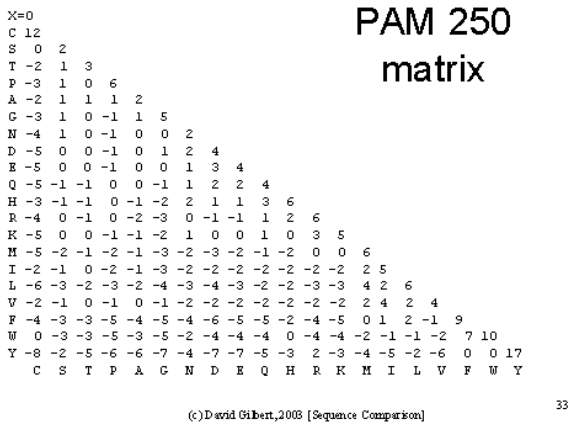
[en.wikipedia.org]

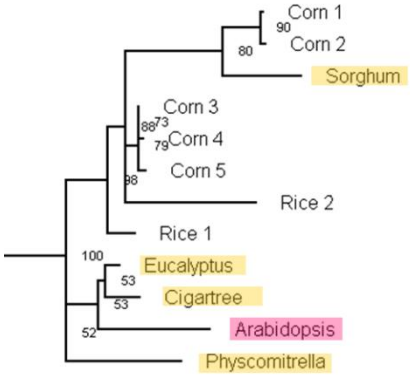


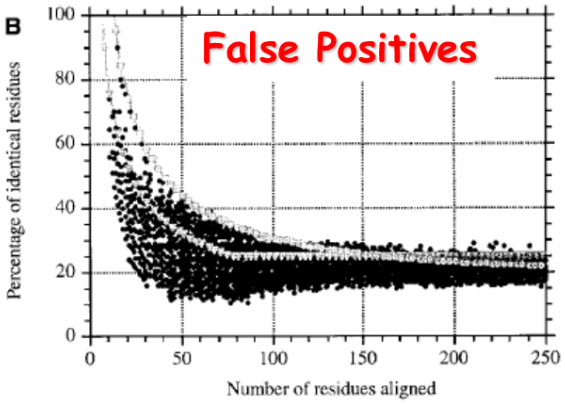
TBLASTX: DNA Query to DNA Database *via* translation

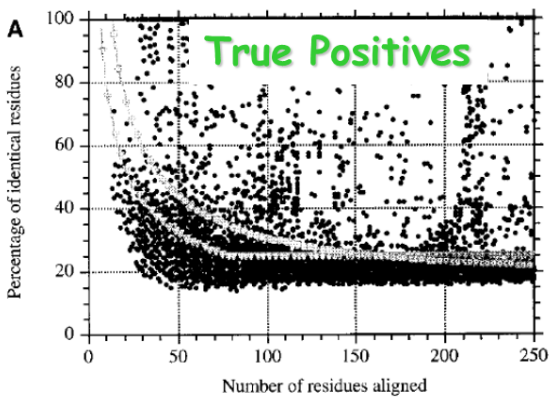
[NCBI]

```
AATGGCATCC          VLLVIRPWHD
|xx|x|x|||          |:|:::x||:
ATAGCCGTCC          VVLILKYWHE
```









[Rost et al. (1999) Protein Engineering 12:85–94]

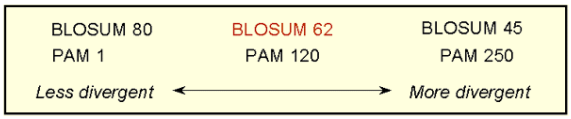
[PDF] **Basic local alignment search tool**
SF **Altschul**, W Gish, W Miller, EW Myers... - Journal of molecular ..., 1990 - cmu.edu
A new approach to rapid sequence comparison, **basic local alignment search tool** (BLAST), directly approximates alignments that optimize a measure of local similarity, the maximal segment pair (MSP) score. Recent mathematical results on the stochastic properties of **...**
Cited by 41804 Related articles All 168 versions

Gapped **BLAST** and PSI-**BLAST**: a new generation of protein database search programs
SF **Altschul**, TL Madden, AA Schäffer... - Nucleic acids ..., 1997 - Oxford Univ Press
Abstract The **BLAST** programs are widely used tools for searching protein and DNA databases for sequence similarities. For protein comparisons, a variety of definitional, algorithmic and statistical refinements described here permits the execution time of the **...**
Cited by 42612 Related articles BL Direct All 271 versions

[scholar.google.de]

The BLOSUM family

- BLOSUM matrices are based on local alignments.
- BLOSUM 62 is a matrix calculated from comparisons of sequences with no less than 62% divergence.
- All BLOSUM matrices are based on observed alignments; they are not extrapolated from comparisons of closely related proteins.
- BLOSUM 62 is the default matrix in BLAST 2.0. Though it is tailored for comparisons of moderately distant proteins, it performs well in detecting closer relationships. A search for distant relatives may be more sensitive with a different matrix.



The relationship between BLOSUM and PAM substitution matrices. BLOSUM matrices with higher numbers and PAM matrices with low numbers are both designed for comparisons of closely related sequences. BLOSUM matrices with low numbers and PAM matrices with high numbers are designed for comparisons of distantly related proteins. If distant relatives of the query sequence are specifically being sought, the matrix can be tailored to that type of search.

[NCBI]

Self hit of a Calmodulin fragment

```
sp|P27161|CALM_LYCES CALMODULIN
           Length = 149

 Score =  123 bits (307), Expect = 3e-29
 Identities = 60/60 (100%), Positives = 60/60 (100%)
Query: 1   NGTIDFPEFLNLMARKMKDTDSEEELKEAFKVFDKDQNGFISAAELRHVMTNLGEKLTDE 60
           NGTIDFPEFLNLMARKMKDTDSEEELKEAFKVFDKDQNGFISAAELRHVMTNLGEKLTDE
Sbjct: 61  NGTIDFPEFLNLMARKMKDTDSEEELKEAFKVFDKDQNGFISAAELRHVMTNLGEKLTDE 120
```

FtsZ query – older BLASTP vs. nr

```
>SP_PL:O65875 O65875 pisum sativum (garden pea). cell division
           protein ftsz. 11/1999
           Length = 423
 Score =  354 bits (900), Expect = 3e-97
 Identities = 181/272 (66%), Positives = 210/272 (76%), Gaps = 1/272 (0%)

Query: 18  VQIGEALTRGLGTGGKPFLGXXXXXXXXXXXXXXVVDADLVFIXXXXXXXXXXXXXPVVA 197
           ++IGE LTRGLGTGG P LG              +  +DLVFI             PVVA
Sbjct: 112 IKIGELLTRGLGTGGNPLLGEQAAEESKEAIANALKGSDLVFITAGMGGGTGSGAAPVVA 171

Query: 198 RVAKEAGQLTVGVVTYPFTFEGRRRSQQAVEAIENLRKSVDSLIVIPNDRLLDVSGDKTP 377
           +++KEAG LTVGVVTYPF+FEGR+RS QA+EAIE L+K+VD+LIVIPNDRLLD++ ++ P
Sbjct: 172 QISKEAGYLTVGVVTYPFSFEGRKRSLQALEAIEKLQKNVDTLIVIPNDRLLDIADEQMP 231

Query: 378 LQEAFSLADDVLRQGVQGISDIITTPGLVNVDFADVRAVMSNSGTAMLGVGSSSGKNRXX 557
           LQ+AF LADDVLRQGVQGISDIIT PGLVNVDFADV+AVM +SGTAMLGVG SSGKNR
Sbjct: 232 LQDAFRLADDVLRQGVQGISDIITIPGLVNVDFADVKAVMKDSGTAMLGVGVSSGKNRAE 291

Query: 558 XXXXXXXXXPLIERSIEQATGIVYNITGGPDLTLQEVNTVSEIVTGLADPSANIIFGAVV 737
                   PLI  SI+ ATG+VYNITGG D+TLQEVN VS++VT LADPSANIIFGAVV
Sbjct: 292 EAAEQATLAPLIGSSIQSATGVVYNITGGKDITLQEVNRVSQVVTSLADPSANIIFGAVV 351

Query: 738 DDKYTGEIHVTIIATGFSHSFQKSLV-DPNVSR 833
```

```
           DD+YTGEIHVTIIATGFS SFQK L+ DP  ++
Sbjct: 352 DDRYTGEIHVTIIATGFSQSFQKKLLTDPRAAK 384
```

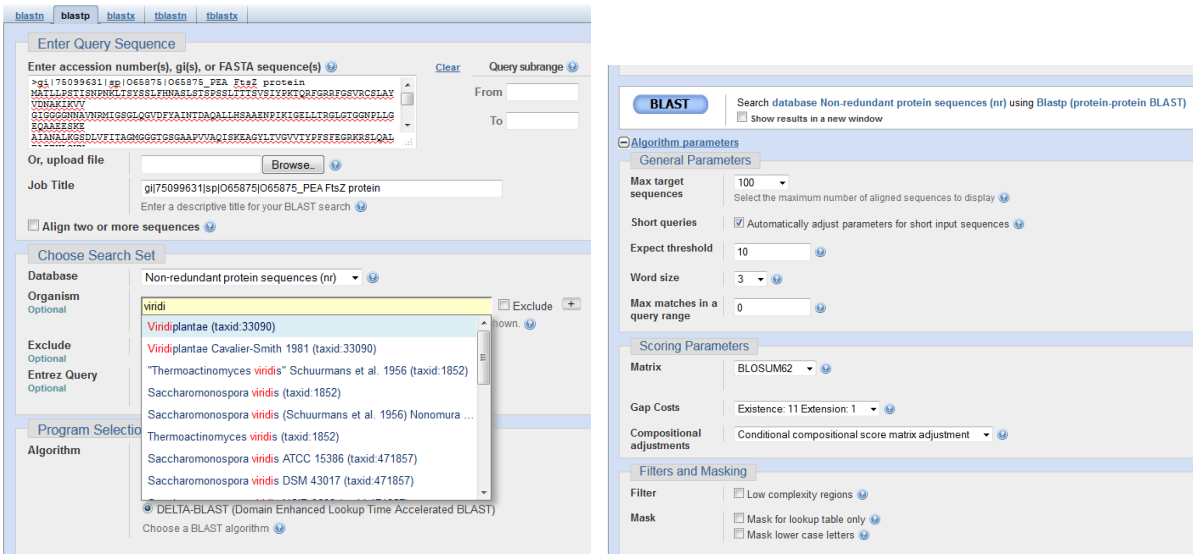
FtsZ query – recent BLASTP vs. nr, option „filter low complexity regions“ off

```
>gb|AAA85526.1|  FtsZ [Nostoc sp. PCC 7120]
Length=379

 Score =  313 bits (801),  Expect = 1e-99, Method: Compositional matrix adjust.
 Identities = 186/288 (65%), Positives = 240/288 (83%), Gaps = 0/288 (0%)

Query  82   RMIGSGLQGVDFYAINTDAQALLHSAAENPIKIgelltrglgtggnpllgeQAAEESKEA  141
            RMI S + GV+F++INTDAQAL  + A + ++IG+ LTRGLG GGNP +G++AAEES++
Sbjct  32   RMIESDVSGVEFWSINTDAQALTLAGAPSRLQIGQKLTRGLGAGGNPAIGQKAAEESRDE  91

Query  142  IANALKGSDLVFItagmgggtgsgaaPVVAQISKEAGYLTVGVVTYPFSFEGRKRSLQAL  201
            IA AL+G+DLVFITAGMGGGTG+GAAP+VA+++KE G LTVGVVT PF FEGR+R+ QA
Sbjct  92   IATALEGADLVFITAGMGGGTGTGAAPIVAEVAKEMGALTVGVVTRPFVFEGRRRTSQAE  151

Query  202  EAIEKLQKNVDTLIVIPNDRLLDIADEQMPLQDAFRLADDVLRQGVQGISDIITIPGLVN  261
            + IE L+  VDTLI+IPN++LL++  EQ P+Q+AFR ADDVLRQGVQGISDIITIPGLVN
Sbjct  152  QGIEGLKSRVDTLIIIPNNKLLEVIPEQTPVQEAFRYADDVLRQGVQGISDIITIPGLVN  211

Query  262  VDFADVKAVMKDSGTAMLGVGVSSGKNRAEEAAEQATLAPLIGSSIQSATGVVYNITGGK  321
            VDFADV+AVM D+G+A++G+GVSSGK+RA EAA  A  +PL+  SI+ A GVV+NITGG
Sbjct  212  VDFADVRAVMADAGSALMGIGVSSGKSRAREAAIAAISSPLLECSIEGARGVVFNITGGS  271

Query  322  DITLQEVNRVSQVVTSLADPSANIIFGAVVDDRYTGEIHVTIIATGFS  369
            D+TL EVN  ++ +  + DP+ANIIFGAV+DDR  GE+ +T+IATGF+
Sbjct  272  DLTLHEVNAAAETIYEVVDPNANIIFGAVIDDRLQGEVRITVIATGFT  319
```

FtsZ query – Delta BLAST vs. Viridiplantae

```
>ref|XP_001767462.1| UniGene info linked to XP_001767462.1Gene info linked to
XP_001767462.1Genome view with mapviewer linked to XP_001767462.1 predicted protein
[Physcomitrella patens subsp. patens]
 gb|AAQ88118.1|  beta-tubulin 5 [Physcomitrella patens]
 gb|EDQ67786.1| Gene info linked to EDQ67786.1 predicted protein [Physcomitrella
patens subsp. patens]
Length=443

 GENE ID: 5930639 PHYPADRAFT_165677 | hypothetical protein
[Physcomitrella patens subsp. patens] (10 or fewer PubMed links)

 Score =  115 bits (289),  Expect = 6e-30, Method: Composition-based stats.
 Identities = 44/363 (12%), Positives = 99/363 (27%), Gaps = 90/363 (25%)

Query  67   IKVVGIGGGGNNAVNRMIGS---------------------------------GLQGVDF  93
            I  +  G  GN    +                                      G
Sbjct  4    ILHIQGGQCGNQIGAKFWEVVCEEHGIDPTGTYKGLSDLQLERINVYYNEASGGRYVPRA  63

Query  94   YAINTDAQALLHSAAE--NPIKIGELLTRGL-GTGGNPLLG-----EQAAEESKEAIANA  145
              ++ +   +    +     I   +    G  G G N   G      +  +   + +
Sbjct  64   VLMDLEPGTMDSVRSGPYGQIFRPDNFVFGQTGAGNNWAKGHYTEGAELIDSVLDVVRKE  123

Query  146  LKGSD---LVFITAGMGGGTGSGAAP-------------VVAQISKEAGY---LTVGVVT  186
             +  D      +                            ++++I +E      LT  V
Sbjct  124  AESCDCLQGFQVCH------------SLGGGTGSGMGTLLISKIREEYPDRMMLTFSVFP  171

Query  187  YPFSFEGRKRSLQALEAIEKLQKNVDTLIVIPNDRLLDIADEQMP-LQDAFRLADDVLRQ  245
             P   +       A  ++ +L +N D  +V+ N+ L DI    +  +  +F   + ++
Sbjct  172  SPKVSDTVVEPYNATLSVHQLVENADECMVLDNEALYDICFRTLKLITPSFGDLNHLISA  231

Query  246  GVQGISDIITIPGLVNVDFADVKAVMKDSG---TAMLGVGVSSG----KNRAEEAAEQAT  298
             + GI+  +  PG +N D   +   +         M+G    +     + R+    E
Sbjct  232  TMSGITCCLRFPGQLNSDLRKLAVNLIPFPRLHFFMVGFAPLTSRGSQQYRSLTVPE--L  289
```

```
Query  299  LAPLIG-SSIQSATGVVY--NITGGK----DITLQEVNRVSQVVTSLADPSANIIFGAVV  351
               +    ++  A    +   +T         ++ +EV+     V +    S  + +
Sbjct  290  TQQMWDAKNMMCAADPRHGRYLTASAMFRGKMSTKEVDEQMINVQNKNS-SYFVEWIPNN  348
```



[NCBI]

Part 2

Based on the definition of sequence homology by BLAST, we will go through the motions on how to create phylogenetic trees. The lecture will confer information on multiple sequence alignment generation, visualization and curation (clipping), as well as on the subsequent generation, visualization and interpretation of phylogenetic trees.

2i) Background knowledge:
[check whether you understand the terms and their relevance]

Mutation rate *vs.* observed mutations / positive and negative selection / conservation

Columns in an alignment / clipping / 50% rule / overlaps

Neighbour-Joining, Maximum Likelihood, Maximum Parsimony, Bayesian Inference…

Terminology (including bootstrap / confidence / support)

Frame of reference: in- and out-paralogs / species vs. gene trees

Model organisms: orthology and evo-devo; informative phylogenetic position

2ii) Some tasks

Right or wrong?! If wrong -> correct it…

The small ribosomal subunit RNA of maize and tomato is homologous.

The small ribosomal subunit RNA of maize is 95% homologous to the one from rice.

The DNA binding domain of the *A. thaliana* transcription factor ARF4 is homologous to the DNA binding domain of the *A. thaliana* transcription factor ABI3.

The *A. thaliana ARF4* gene is homologous to the *P. patens ARF4* gene.

A. thaliana ARF4 is homologous to *A. thaliana* ABI3.

The DNA binding domain of the *A. thaliana* transcription factor ARF4 is quite homologous to the DNA binding domain of the *A. thaliana* transcription factor ABI3.

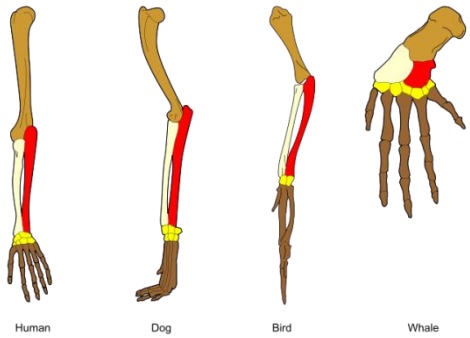
How to detect a xenolog/horizontal gene transfer?

Were the FIE paralogs acquired early and lost in the lineages with only one copy, or gained several times independently?

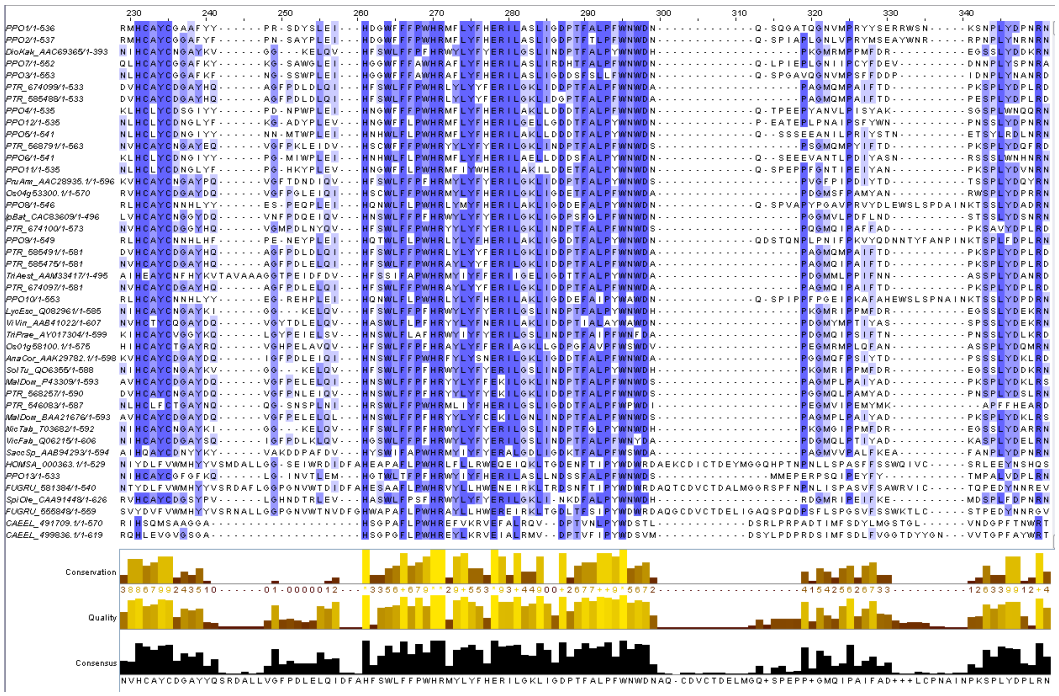
Where does the plant nuclear encoded *ftsZ* gene come from?

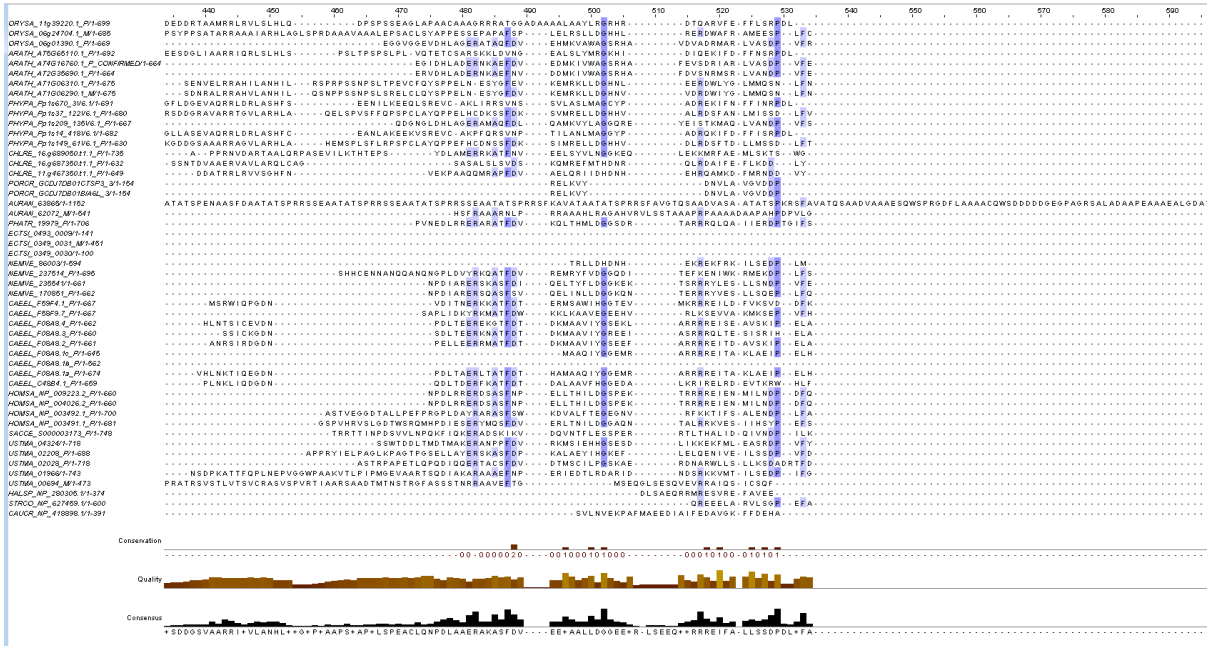
If there were a *WUS* gene in *P. patens*, what could you potentially learn from its loss of function mutant?

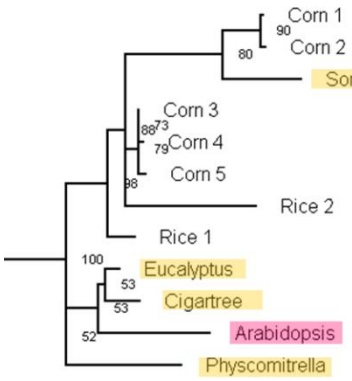
2iii) Material



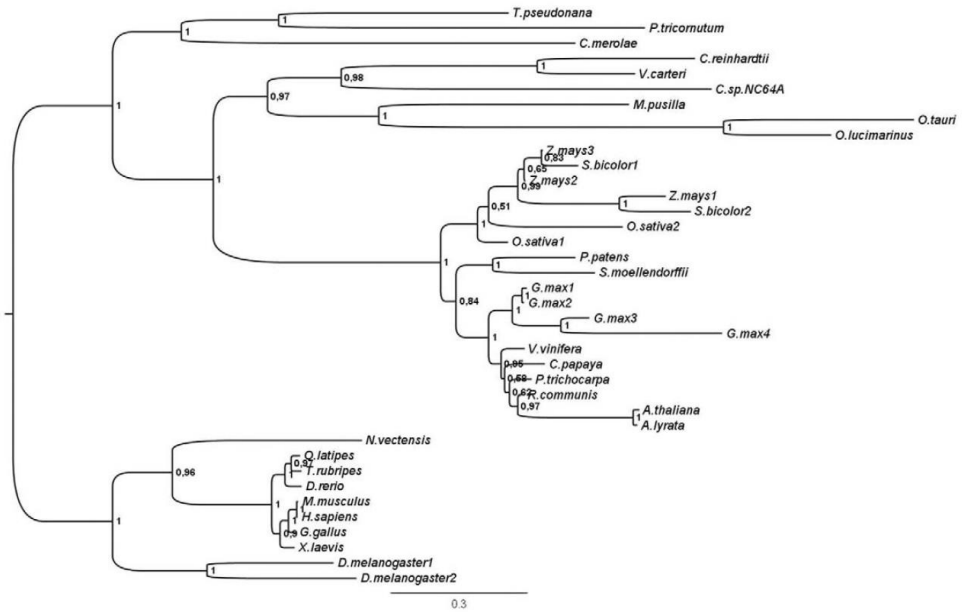
[en.wikipedia.org]







Two FIE trees



[Mosquna et al. (2009) Development 136:2433]

FtsZ tree

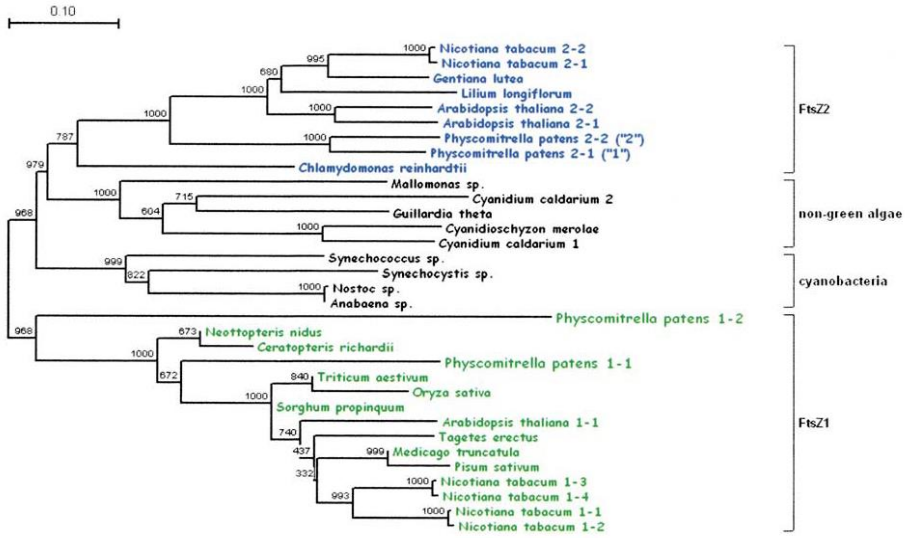


Fig. 3. Neighbor-joining phylogenetic tree based on 530 amino acid positions. Numbers at the branches result from 1000 bootstrap resamplings. See text for details.

[Rensing et al. (2004) J Mol Evol 58:154]

The closest **BLAST** hit is often not the nearest neighbor
LB Koski, GB Golding - Journal of Molecular Evolution, 2001 - Springer
... Nevertheless, it is often common practice to report simply the most similar **BLAST** (**Altschul** et al. **... Altschul** SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped **BLAST** and PSI-**BLAST**: a new gen- eration of protein database search programs. **...**
Cited by 330 Related articles BL Direct All 13 versions

[scholar.google.de]

P. patens evo-devo papers

Diversification of gene function: homologs of the floral regulator FLO/LFY control the first zygotic cell division in the moss Physcomitrella patens.
Tanahashi T, Sumikawa N, Kato M, Hasebe M.
Development. 2005 Apr;132(7):1727-36.

An ancient mechanism controls the development of cells with a rooting function in land plants. bHLH
Menand B, Yi K, Jouannic S, Hoffmann L, Ryan E, Linstead P, Schaefer DG, Dolan L.
Science. 2007 Jun 8;316(5830):1477-80.

Regulation of stem cell maintenance by the Polycomb protein FIE has been conserved during land plant evolution.
Mosquna A, Katz A, Decker EL, Rensing SA, Reski R, Ohad N.
Development. 2009 Jul;136(14):2433-44.

A polycomb repressive complex 2 gene regulates apogamy and gives evolutionary insights into early land plant evolution.
Okano Y, Aono N, Hiwatashi Y, Murata T, Nishiyama T, Ishikawa T, Kubo M, Hasebe M.
Proc Natl Acad Sci U S A. 2009 Sep 22;106(38):16321-6.

Role of ABA and ABI3 in desiccation tolerance.
Khandelwal A, Cho SH, Marella H, Sakata Y, Perroud PF, Pan A, Quatrano RS. ABI3
Science. 2010 Jan 29;327(5965):546.

450 MY old orthologs

Science AAAS

The Physcomitrella Genome Reveals Evolutionary Insights into the Conquest of Land by Plants
Stefan A. Rensing, *et al.*
Science **319**, 64 (2008);
DOI: 10.1126/science.1150646

2iv) Tools

Jalview – generating, loading, looking at, clipping, saving alignments
http://www.jalview.org/
(+ is able to call some state of the art alignment algorithms: probcons, muscle, mafft, tcoffee)

Phylogeny.fr – from alignment to tree visualization
(offers tree inference by NJ, ML, MP, BI)
http://www.phylogeny.fr/
or **MEGA**, http://www.megasoftware.net/

FigTree – load, look at, alter, save/export trees
http://tree.bio.ed.ac.uk/software/figtree/

If you dare:
NJ: Quicktree, Quicktree_SD (http://www.mybiosoftware.com/quicktree-sd-1-2-implemented-scroredist-distances-quicktree.html)
ML: TREE-PUZZLE (http://www.tree-puzzle.de/), RAxML (http://sco.h-its.org/exelixis/software.html)
BI: MrBayes (http://mrbayes.sourceforge.net/)
Model testing: ProtTest: http://darwin.uvigo.es/software/prottest.html , JModeltest http://jmodeltest.org/login or http://www.molecularevolution.org/software/phylogenetics/jmodeltest
Alignment trimming: http://trimal.cgenomics.org/ or http://www.phylobabble.org/t/pruning-an-alignment-rows-or-columns/280
Mesquite: http://mesquiteproject.org/

Notes

Notes