



This site provides online material for the www.cosmoss.org Physcomitrella Genome Workshop that is held in Freiburg.

date

16.-18.3.2008

Contents

- [1 Speakers](#)
- [2 Cosmoss accounts](#)
- [3 First day](#)
 - ◆ [3.1 Morning session](#)
 - ◇ [3.1.1 The P. patens genome](#)
 - ◇ [3.1.2 Overview of the cosmoss.org resources](#)
 - ◇ [3.1.3 Cosmoss sequence retrieval](#)
 - ◇ [3.1.4 BLAST, homology and hit filtering](#)
 - ◇ [3.1.5 Hands-on experience I](#)
 - [3.1.5.1 Menu, documentation, mailing list and wiki](#)
 - [3.1.5.2 Sequence Retrieval](#)
 - [3.1.5.3 Keyword search](#)
 - [3.1.5.4 Sequence Viewer](#)
 - ◇ [3.1.6 The cosmoss.org databases](#)
 - ◇ [3.1.7 mossDB](#)
 - ◆ [3.2 Afternoon session](#)
 - ◇ [3.2.1 Hands-on experience II](#)
 - ◇ [3.2.2 Reciprocal BLAST](#)
 - ◇ [3.2.3 CSV BLAST](#)
 - ◇ [3.2.4 Hands-on experience III](#)
- [4 Second Day](#)
 - ◆ [4.1 Morning session](#)
 - ◇ [4.1.1 The cosmoss.org genome browser](#)
 - ◇ [4.1.2 Hands-on experience IV](#)
 - ◇ [4.1.3 Genome browser: hidden treasures](#)
 - ◇ [4.1.4 Hands-on experience V](#)
 - [4.1.4.1 BLAST gbrowse integration](#)
 - [4.1.4.2 Advanced navigation and zooming](#)
 - [4.1.4.3 Exporting sequence annotations and publication quality images](#)
 - [4.1.4.4 Highlighting](#)
 - [4.1.4.5 Displaying custom annotation](#)
 - ◆ [4.2 Afternoon session](#)
 - ◇ [4.2.1 PlanTAPDB as a knowledge base](#)

- ◇ [4.2.2 Hands-on experience VI](#)
- ◇ [4.2.3 Hands-on experience VII](#)
- ◇ [4.2.4 Hands-on experience VIII](#)

Speakers

- [Stefan Rensing](#)
- [Andreas Zimmer](#)
- [Daniel Lang](#)

Cosmoss accounts

You have to be a registered cosmoss user in order to access some of linked material.

For the workshop participants, accounts have been created (Unless you already had one) with an initial password. The password is on the sheet with the wifi credentials at the bottom. Please change it as soon you [log in](#).

But everyone is invited to register! Please contact helpdesk-cosmoss@uhura.biologie.uni-freiburg.de for an account!

First day

Morning session

The *P. patens* genome

speaker

Stefan

presentation

[Genome.pdf](#)

Overview of the cosmoss.org resources

speaker

Daniel

presentation

[cosmoss_overview.pdf](#)

Cosmoss sequence retrieval

speaker

Andreas

presentation

[SequenceRetrieval.pdf](#)

BLAST, homology and hit filtering

speaker

Stefan

presentation

[BLAST.pdf](#)

Hands-on experience I

Using cosmoss.org (help, documentation), sequence retrieval, databases

prepared by

Andreas and Daniel

Menu, documentation, mailing list and wiki

Tasks

1. Explore the the menu system
2. Find the FAQ
3. Find the BLAST documentation - What are the e-value threshold defaults?
4. Explore the wiki - Find out [Daniel's ICQ](#) number
5. Register to both of the [mailing lists](#)

Sequence Retrieval

documentation

[Sequence Retrieval](#)

Familiarize yourself with the Cosmoss Retrieval system:

Tasks

Try to retrieve the following sequences:

BJ172647

BJ179866

PP015013150R

Cosmoss_workshop_2008

You can access the sequence retrieval via the transcriptome and genome menu.

Hint: These accession numbers above are from the [pp0304](#) annotated virtual transcript database. You can retrieve multiple sequences by providing their accession numbers as a comma- or space-separated

For a vast number of sequences you could upload your request in a text file.

File format: text
Provide accession number per line.

Select the following accession numbers and copy them into a new text file (e.g. notepad) and save it.

Phypa_173430
Phypa_201973
Phypa_221004
Phypa_59935
Phypa_152025
Phypa_225236
Phypa_106210
Phypa_109367
Phypa_110675
Phypa_125839
Phypa_125903
Phypa_168764
Phypa_201189
Phypa_233894
Phypa_61317
Phypa_77574
Phypa_115069
Phypa_123666
Phypa_146969
Phypa_87740
Phypa_87752
Phypa_151552
Phypa_8310

Goto the cosmoss.org sequence retrieval:

Select database: *P.patens.V1.2_proteins*

This database contains all *P.patens* released proteins. In comparison to the release V1.1 all gene models overlapping with transposons, non-protein-coding genes (e.g. tRNA genes) were removed. The *Physcomitrella patens* genome accession numbers work for both transcripts and proteins databases. Just change the database to change to your favored sequence type.

Browse for the previously created file and submit your request

Save the sequences in a new file in FASTA format

Select only a subset of the sequences and save it to a new file

Keyword search

documentation

[Keyword search](#)

How many *geranylgeranyl pyrophosphate synth(et)ases* (GGPS) are in the virtual transcriptome?

Try to find the corresponding [pp0304](#) virtual transcripts by [keyword search](#)!

Tasks

1. Read the documentation
2. Use the simple search menu to find the GGPS's in [pp0304](#)
3. Play around with the advanced search option.

This is the advanced query that works:

```
"geranylgeranyl"[DESC] AND "phosphate"[DESC] AND "synth"[DESC]
```

Finally, here are the two loci in the genome:

- [Phypa_201614](#)
- [Phypa_65868](#)

The two initial [pp0304](#) transcripts are highlighted with a yellow box.

Sequence Viewer

documentation

[Sequence Viewer](#)

The accession numbers (sequence identifiers) in the search results from the above keywords search link out to the (transcriptome) sequence viewer!

Tasks

1. Check it out and play around with the (sometimes hidden, I know) links to inspect all features that we annotated for these sequences.
2. [PP_9795_C1](#) is an assembled (contig) transcript. How many ESTs are included in this contig?
3. Retrieve the contig members in [Genbank format](#)!

In the [databases talk](#) you'll hear more on the process, to help you interpret the things you see in the Sequence Viewer...

The cosmoss.org databases

speaker

Daniel

presentation

[cosmoss_databases.pdf](#)

further information

- [The DDBJ/EMBL/GenBank Feature Table Definition](#)
 - [List of features and qualifiers, with examples](#)
 - [EMBL \(format\) manual](#)
 - [EMBL annotation examples](#)
 - [Sequence Ontology Project \(SO\)](#)
 - [GENERIC FEATURE FORMAT VERSION 3 \(GFF\)](#)
 - [Gene Ontology](#)
-

mossDB

speaker

Stefan

presentation

[mossDB.pdf](#)

Afternoon session

Hands-on experience II

(batch) BLAST & hit filtering

prepared by

Stefan

material

[BLAST hands_on.pdf](#)

Reciprocal BLAST

speaker

Andreas

presentation

[Reciprocal BLAST searches.pdf](#)

CSV BLAST

speaker

Stefan

presentation

[CSV-BLAST.pdf](#)

Hands-on experience III

CSV-BLAST and reciprocal BLAST

prepared by

Stefan and Andreas

material

[CSV-BLAST hands_on.pdf](#)

[Hands_on_reciprocal_blast.pdf](#)

Second Day

Morning session

The cosmoss.org genome browser

speaker

Andreas

presentation

[GenomeBrowser.pdf](#)

Hands-on experience IV

Genome browser basics

prepared by

Andreas and Daniel

presentation

[Hands_on_gbrowse_basics.pdf](#)

Genome browser: hidden treasures

speaker

Daniel

presentation

[cosmoss_gbrowse_treasures.pdf](#)

links

- [Scalable Vector Graphics \(SVG\)](#)
 - [Inkscape](#) An Open Source vector graphics editor, with capabilities similar to Illustrator, CorelDraw, or Xara X, using the W3C standard Scalable Vector Graphics (SVG) file format.
-

Hands-on experience V

Genome browser: customization and special features

prepared by

Daniel

documentation [gbrowse](#)



[general gbrowse help](#)

BLAST gbrowse integration

documentation

[BLAST gbrowse integration](#)

Tasks

1. BLAST with an arbitrary transcript/EST vs the scaffolds and follow the  link.
2. Compare results when BLASTing w/o low complexity filtering!
3. Compare the BLAST to the spliced-alignment results. Are there lonely exons?
4. BLAST with the *Arabidopsis* protein AT5G13930 .1 vs the v1.2 gene models and follow the  link.
5. Compare one of the hit *Physcomitrella* loci vs the *Arabidopsis* locus [AT5G13930](#)

Advanced navigation and zooming

Tasks

1. Find out the definition of a *gene* in [SO](#) by using the *Ontology_term* cross-link in a gene feature's mouse-over window
2. Zoom into a CDS exon and back again to the region of its *mommy* mRNA or gene
3. Play around with the zoom function to inspect EST and cDNA spliced alignments!

Exporting sequence annotations and publication quality images

Tasks

1. Export the upstream 10kbp your most favorite region to [FASTA](#) format and reverse it when necessary to reflect the gene's orientation.
2. *OPTIONAL*: Get another upstream region and try to find shared putative promoter elements using e.g. [AlignACE](#)
3. Save your most favorite region as a [png](#) image.

4. Check out the exemplary PDF created with Inkscape from the SVG of a region:
www.cosmoss.org/inkscape_example.pdf

Highlighting

example locus

[scaffold_29:573541..579040](#)

documentation

[available colors](#) (scroll down to *Colors*)

Tasks

1. Let gbrowse highlight a feature and zoom out again, in order to see whether it overlaps with another feature in another track (*highlight and zoom to the region of this feature* link in the mouse-over)
2. Highlight your favorite gene model for the locus using the *Highlight feature* box in the *Display panel*.
3. Visualize an PCR experiment on the genomic locus! You've just amplified and sequenced an genomic PCR product for the locus using the primer coordinates below. Highlight the genomic region using the *Highlight regions* box in the *Display panel*.

forward primer mapping

[scaffold_29:575397..575417](#)

reverse primer mapping

[scaffold_29:577964..577950](#)

The screenshot shows the gbrowse web interface. At the top, there are several tracks with checkboxes: 'MIPS ANGELA long_terminal_repeat', 'TE related PFAM domains', and 'Helitrons Bennetzen lab'. Below these tracks are buttons for 'Configure tracks...' and 'Update'. A section titled 'Display Settings' is expanded, showing three input fields: 'Highlight feature(s) (feature1 feature2...)' with the value 'all_Phypa_205678@yellow', 'Highlight regions (region1:start..end region2:start..end)' with the value 'scaffold_29:577964-577950', and 'Region Size (bp)' with a dropdown menu set to '10 kbp'. There is an 'Update' button at the bottom right of this section. Below the 'Display Settings' section is another section titled 'Add your own tracks'.

Displaying custom annotation

You can draw your own features and have your own custom track in the browser!

documentation

[gbrowse custom annotation help](#)

Exporting sequence annotations and publication quality images

example locus

[scaffold_29:573541..579040](#)

Tasks

1. Go through the documentation
2. Visualize an RT-PCR experiment on the genomic locus! You've just amplified and sequenced an RT-PCR product for the locus below. Display it as custom annotation using the coordinates:

```
RT-exon1 scaffold_29:575397..575570
RT-exon1 scaffold_29:576849..576875
RT-exon1 scaffold_29:577042..577335
RT-exon1 scaffold_29:577629..577692
```

See the [example file](#) which combines everything. Adjust it, if you like!

Afternoon session

PlanTAPDB as a knowledge base

speaker

Daniel

presentation

[cosmoss_plantapdb.pdf](#)

related resource

[phylogenies from the genome paper](#)

Hands-on experience VI

PlanTAPDB: keyword search, filtering, phylogenies, taxonomic profiles

showed by

Daniel

Hands-on experience VII

?The Summary HowTo I.?

Stefan

material

[HowToSummaryI.pdf](#) and [see this folder](#)

Hands-on experience VIII

?The Summary HowTo II.?

Depending on what you'd like to work upon...