

FUNCTIONAL GENOMICS GROUP
AT FRAUNHOFER IGB STUTTGART



AN ANNOTATE-BY-CLICK BROWSER FOR INSPECTION OF
TRANSCRIPTIONAL LANDSCAPES

Provisional Manual

Author:
Stefan Lorenz

May 6, 2013

Contents

1 Summary	3
2 Overview Of Functionality	3
3 Requirements	3
4 Installation	3
5 Launching GeneScapes	4
5.1 Start Session	4
5.2 Navigation	5
6 Visualisation Setup	6
6.1 The Size Menu	7
6.2 The Look Menu	7
6.3 The Color Menu	8
6.4 Save/Load Visualisation Setup	8
7 Loading Additional Files	9
7.1 Open Track Table	9
7.2 Save/Load GeneScapes Session	9
7.3 An Overview Of Import Formats	10
8 Information Window	11
8.1 Feature Information Field	12
8.2 Feature Editing	12
8.3 Create/Remove Features	12
8.4 Feature Editor	13
9 Select Features	14
10 Export Data	15
10.1 Export Features	15
10.2 Export ABC track	15
10.3 Overview Of Export Formats	15
11 Annotate By Click (ABC)	16
11.1 Create A New Bucket	16
11.2 Via Information Window	16
11.3 Via Wiggle Plot	16
11.4 Via Selection Frame	16
11.5 Add Gene And CDS Feature	17
11.6 Other Feature Types	17
12 Visualise Orthologs	18
12.1 Prepare Files	18
12.2 Anchor & Shift Positions	18
12.3 Reset Positions	19
13 Create Sequence Logos	20
14 Blast Features	21
15 Conclusion	22
16 Missing Manual Parts	22

List of Figures

1	Main window of GeneScapes	4
2	Main menu of GeneScapes	4
3	Entry field for downloading files from NCBI	4
4	Embedded browser window for visualisation of a data set	5
5	Visualisation menu is split into three submenues, Look, Size and Color	6
6	Size menu of genomic features	7
7	Buckets in GeneScapes	7
8	The look submenu	7
9	The color chooser	8
10	Somehow configured the visualisation	8
11	The track table assists handling of files	9
12	Imported annotations and HTS reads	10
13	The indicator shows the status of different modes	11
14	Ttoolbox for annotating and receiving information about features	11
15	The feature editor enables changing information of features	13
16	Mark an region by dragging the mouse cursor	14
17	Selected features marked by highlighted background	14
18	GeneScapes requests for genomic positions	16
19	Tool box for annotating	17
20	Find orthologs with GeneScapes	18
21	Table to shift positions for each track individually	19
22	Use the shift function to match positions	20
23	The first few nucleotides of the sequence logo starting with the start codon	20

1 Summary

The annotation process of automatically assembled transcripts is often insufficient [Martin and Wang, 2011]. Therefore we developed a tool for manual curation of transcriptomic landscapes. With the aid of visualisation of sequenced reads, assemblies as well as gene orthologues GeneScapes? helps to reduce errors in annotations. Beside the easy use of (re)annotation functionalities the software is highly adaptable in its visualisation for different data sets and linked to web services and databases. Written in Java (1.7) it runs on nearly all platforms.

2 Overview Of Functionality

GeneScapes is a tool for:

- inspection of high throughput sequencing data (HTS) [Wang et al., 2009]
- manual (re)annotation of transcriptional active regions
- visualisation of orthologs via track shifting
- big variety of visualisation

3 Requirements

GeneScapes is written in Java which makes it platform independent. It requires the Java Runtime Environment (JRE) 1.7 or higher, which is freely available. 2 GB of RAM are recommended. To use full functionality of the software, install 'Python' as well as the additional library 'Biopython' [Cock et al., 2009]. Python should be already installed on Linux systems by default. Help for downloading and installing one can find on the web sites listed below.

- Java JRE 1.7 or higher

For full functionality:

- Python
- Biopython
- Internet

4 Installation

Download and extract the tar archive. Start it from the terminal to set memory.

```
tar xfv GeneScapes<version>.tar # extract
cd GeneScapes<version>          # change directory
java -jar -Xmx2G GeneScapes.jar # start genescapes
```

5 Launching GeneScapes

Launching the program the main window of GeneScapes opens, looking similar to figure 1. On the top the main menu is located, which contains functions to start a new session, including loading files or changing the look&feel settings. Sessions are shown in separated embedded windows. On the bottom left the memory bar indicates the used memory of the system.

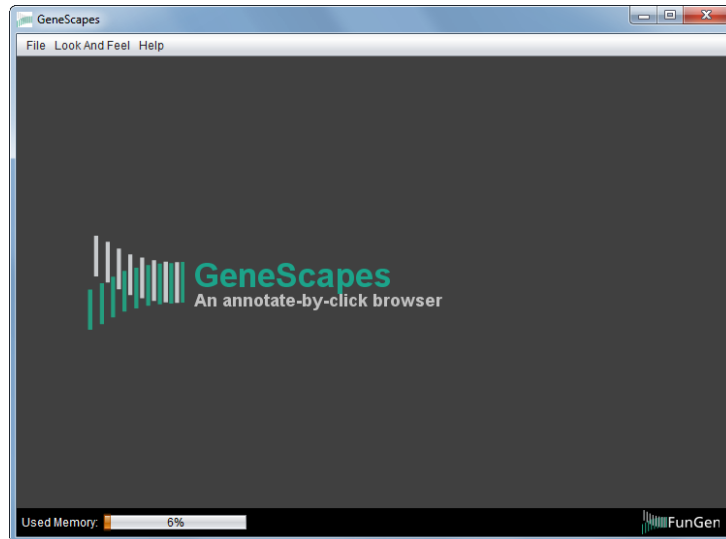


Figure 1: Main window of GeneScapes

5.1 Start Session

As first step we want to download and visualise the *Yersinia pestis* genome sequence and its annotation, which has the accession id 'NC_009381' on NCBI [Pruitt et al., 2007]. Click on the menu 'File' → 'Download from NCBI' (Internet is required). If one selects 'Open Local File...' a filechooser pops up for selecting a file stored on your pc. Select the 'example.gbk' file in the '../GeneScapes/genomes/' directory. The *Yersinia pestis* genome gets also loaded. You might have to extract the zipped example file. If the download option is chosen enter the accession id 'NC_009381' as shown in figure 2 & 3.

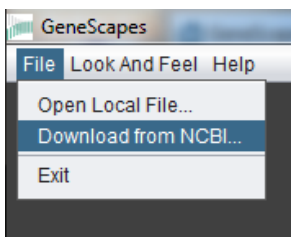


Figure 2: Main menu of GeneScapes

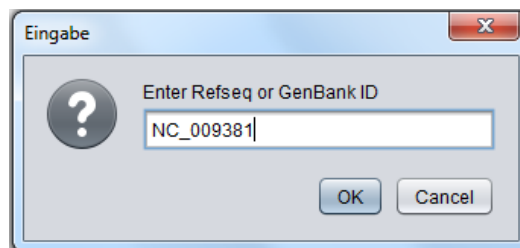


Figure 3: Entry field for downloading files from NCBI

GeneScapes downloads the GenBank file with all its genomic features and sequences into the 'genomes' folder named like the accession id and the 'gbk' file extension.

The embedded session window looks similar to this (Fig. 4):

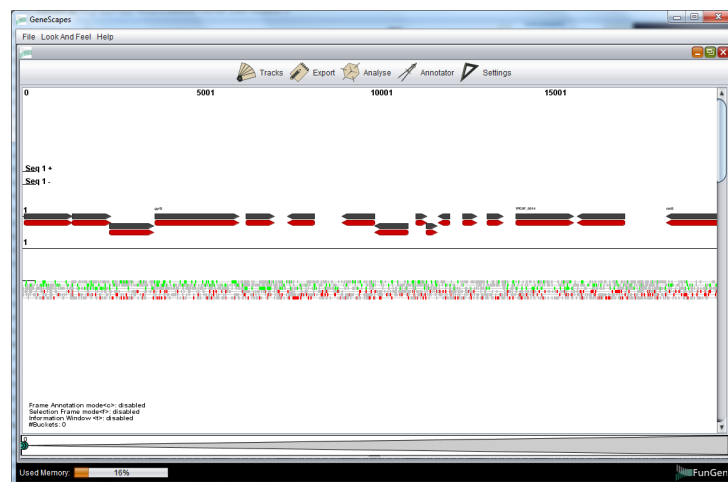


Figure 4: Embedded browser window for visualisation of a data set

5.2 Navigation

There are different ways to navigate the browser through the loaded pestis genome. One can either use the keyboard, mouse, the navigation slider or a combination of all. There are four directions: zoom in, zoom out, go left and go right.

- By keyboard

Go left	a
Go right	d
Zoom in	w
Zoom out	s

- By mouse

By dragging the mouse cursor navigation is supported in the same intuitive principle commonly known e.g. for google maps.

- By navigation slider

The navigation slider consists of a start and stop knob and a range bar which represents the displayed region.

- Bar - The bar is draggable and changes start and stop position and keeps the range
- Start knob - Drag left or right to change start position of the browser
- Stop Knob - Drag left or right to change stop position of the browser

6 Visualisation Setup

Let's be honest, the standard settings aren't the most beautiful ones! Now we want to adopt the visualisation.

To open the track settings do a right click on the short horizontal line on the left side numbered by the track index, which opens the track settings menu. This line also represents the position of the track and is relocatable in its height by dragging the mouse cursor. The menu is separated into three submenues named 'Size', 'Look' and 'Color'.

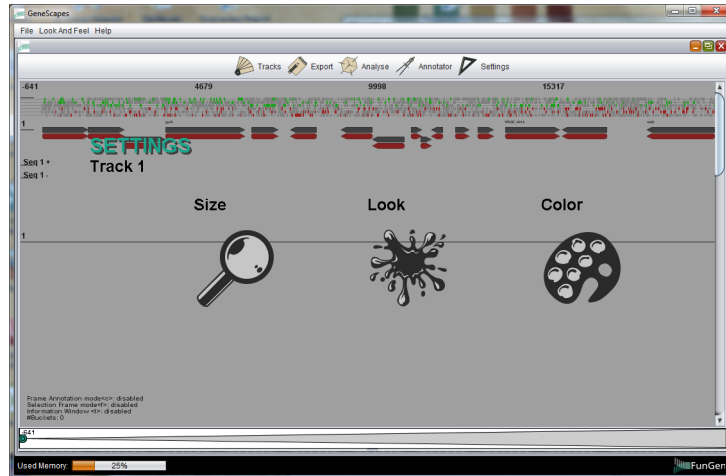


Figure 5: Visualisation menu is split into three submenues, Look, Size and Color

GeneScapes handles the different genomic feature types like 'gene', 'CDS', 'ncRNA' etc. separately. One has to select the desired type in the combobox to change the settings for it. All submenues are structured the same way, but they can differ a bit e.g. the six-frame-translation does not have all setting parameters genomic features have for example. Possible feature types are listed here. To move the track vertically drag this line to the desired position.

6.1 The Size Menu

In this submenu sizes of plots, genomic features or sequences are changeable. For a better understanding how features are stored in GeneScapes and what is meant by 'Buckets', 'Spaces' etc. see figure 7.

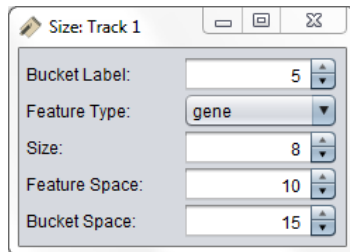


Figure 6: Size menu of genomic features

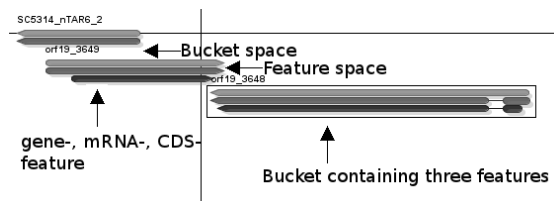


Figure 7: Buckets in GeneScapes

Bucket Label	Changes the size of the bucket label
Feature Type	Type of feature you want to change
Size	Changes size of the feature
Feature Space	Changes the size of the spaces between overlapping features in a bucket
Bucket Space	Changes the size of overlapping buckets

6.2 The Look Menu

In this submenu shapes of features are changeable as well as the visibility. In figure 8 the submenu of a genomic feature and the wiggle plot is shown. Wiggle plots are used to visualize scores of features like RPKM [Mortazavi et al., 2008] or FPKM [Trapnell et al., 2010] values or as most common application, the distribution of NGS reads.

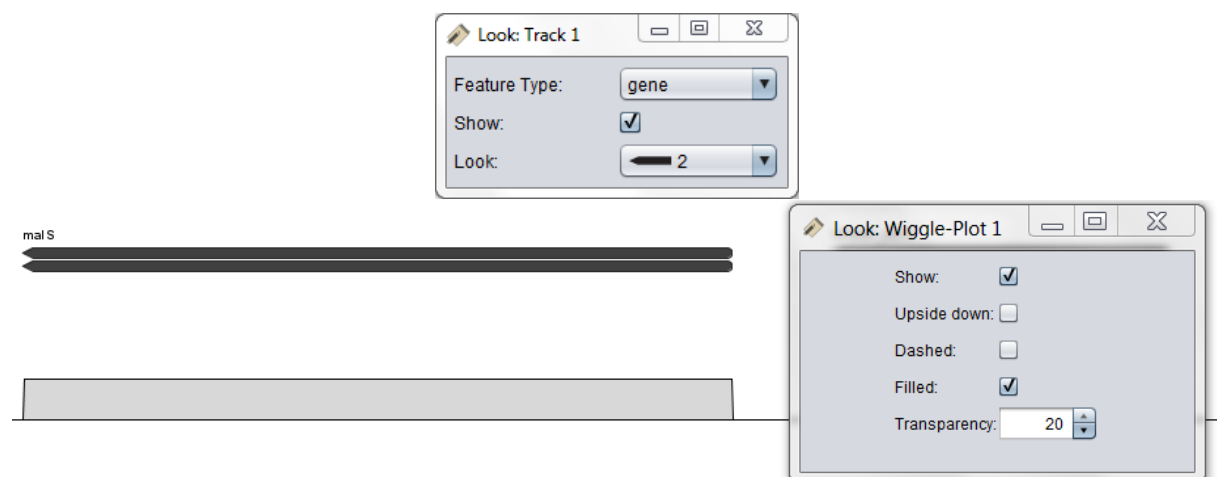


Figure 8: The look submenu

6.3 The Color Menu

The color menu enables changing colors of features, wiggle plots, sequences etc.. On the bottom left the combobox shows the types of features if one selected a feature track. The combobox is also filled with nucleotide foreground/background color, start- and stop codon etc. depended on the selected track

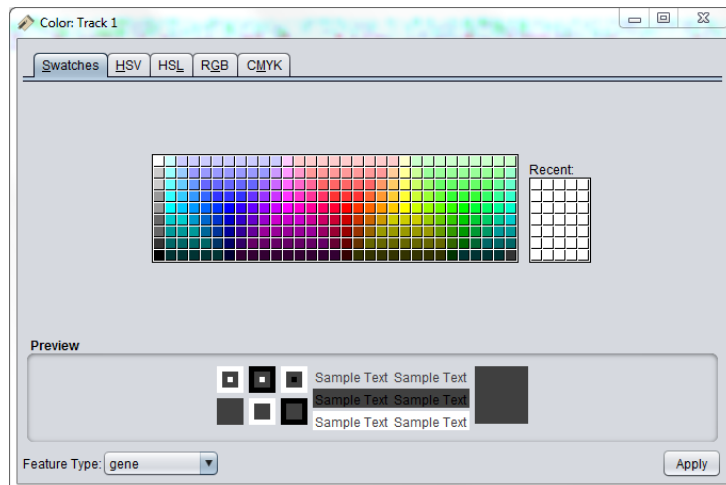


Figure 9: The color chooser

Finally it might look like this...



Figure 10: Somehow configured the visualisation


6.4 Save/Load Visualisation Setup

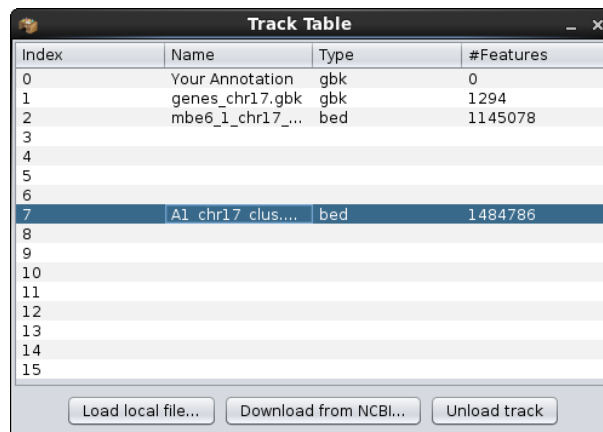
It is possible to store your settings persistently as XML file [Slominski, 2006]. Therefore use the 'Viewer Menu → Settings → Save Track Settings'. Either one can save it as default, so GeneScapes will load this settings the next time oen starts it or one can save it as an extra file.

7 Loading Additional Files

Having started a session one often needs more information than just from a single file to visualize and analyse biological data. One needs to load these files into so called tracks. A track represents different kinds of visualisation, depending on the file format. One can either load a local file from your computer or download a GBK file from the NCBI database by entering the accession ID.

7.1 Open Track Table

The track table, 'Viewer Menu → Tracks  → Open Track Table', shows several so called 'tracks' in a table, numbered in the first column (track index). 16 tracks are provided which can be filled with different data sets. The track with the track index 0 is your annotation track named 'ABC-track'. It is possible to unload a certain track by clicking 'Unload Track'. Once a track is unloaded you are able to refill it with another file again. If you load data in a already filled track the loaded one gets removed and the new one takes place. The handling is the same for all supported file formats, the corresponding visualisation and the possibilities of analysing it are different because every file format has differences in its information content.




Index	Name	Type	#Features
0	Your Annotation	gbk	0
1	genes_chr17.gbk	gbk	1294
2	mbe6_1_chr17_...	bed	1145078
3			
4			
5			
6			
7	A1_chr17 clus....	bed	1484786
8			
9			
10			
11			
12			
13			
14			
15			

Figure 11: The track table assists handling of files

- Load local file... - To import local data select a track by clicking on the desired table row and press the 'Load Local File' button and select a file via the filechooser.
- Download from NCBI... - Another possibility is to download data as a GBK file from the NCBI database via the accession ID. The file is stored in the '../GeneScapes/genomes/' directory and gets automatically loaded.
- Unload track - Removes the data in the selected track.

7.2 Save/Load GeneScapes Session

GeneScapes provides saving whole Sessions in *.gss format. One can save a session via the 'Viewer Menu → Settings  → Save Current Session'. A filechooser pops up requesting directory and name of the gss file. Reconstructing a session works the same way explained in 5.1. Load gss files via the 'Main Frame Menu → Open Local File...'. GeneScapes creates a temporary session file if loaded data changes. This 'LastOpened.gss' you find in the GeneScapes folder.

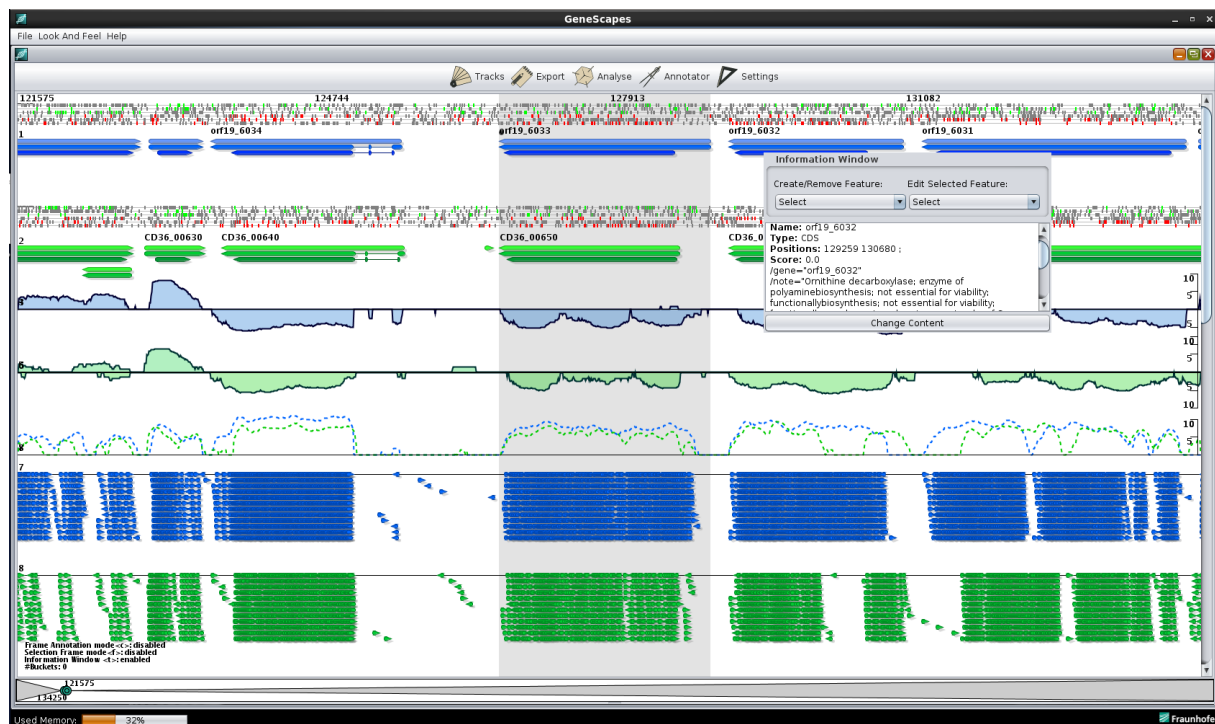


Figure 12: Imported annotations and HTS reads

7.3 An Overview Of Import Formats

Up to now many different file formats exist. Files containing genomic features and optional sequences like 'SAM', 'EMBL' or 'GenBank' belonging to one group. Files containing only features like 'GFF', 'GTF' or 'BED' belonging to the second one. Files containing continuous data for quantitative representation like 'WIG' and 'BEDGRAPH' represents the third group. The fourth group contains a reference sequences, the format is called 'FASTA'. The specifications of the provided file formats are listed below.

- BED12
- BEDGRAPH
- EMBL
- GBK
- GFF3
- GTF
- SAM
- WIGGLE
- FastA
- GSS - The GSS GeneScapeSession format allows to reconstruct previous GeneScapes sessions
- Track Settings - The adaptation of visualisation can get persistently saved in a XML file, which enables a quick change of the display setup

8 Information Window

The information windows enables receiving meta data from a genomic features. Therefore the user needs to enable the information window mode. Pressing 't' enables/disables this tool box, which pops up when the mouse cursor hovers a feature. Another possibility is to set visibility via the 'Viewer Menu' → 'Analyse' → 'Show Information Window'. To close the window do a right click on the textarea in the middle.

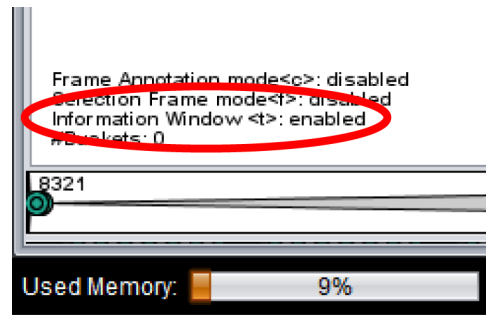


Figure 13: The indicator shows the status of different modes

The information window is divided into four parts, the information panel in the middle, the feature editing and feature creation comboboxes and the 'Change Content' button at the bottom.

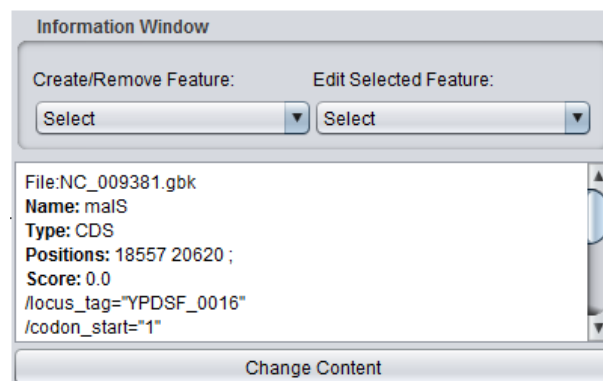


Figure 14: Ttoolbox for annotating and receiving information about features

The single parts are described in the next subsections.

8.1 Feature Information Field

In the middle of the information window a textarea shows all information of a hovered feature. Typically, a feature contains a name, type, positions and a score as well as some additional information. In this case (figure 14) the name is the gene name 'malS', the type in this case is 'CDS' indicating coding regions and the positions of the gene on the reference genome/contig. The score either represents an expression value like FPKM coming from a Cufflinks assembled transcriptome annotation [Trapnell et al., 2010] or mapping scores e.g. coming from Tophat mapping [Trapnell et al., 2009]. Visualising read distribution is the most common application. Enabling the wiggle plot (see 6.2) visualizes the scores. Other additional information are shown in the form /tag="information". Underlined blue colored tags are links to different databases like NCBI or ENSEMBL. Clicking the link opens the default internet browser of your system and loads the corresponding database entry.

8.2 Feature Editing

The 'Edit Selected Feature' combobox lists different functions to edit a feature. Changing the strand affinity, setting the score or add new tags is done with one click. Note the tag form: /tag-name="note" for insertion of new own tags.

+	set strand affinity to '+'
-	set strand affinity to '-'
gene	set type to 'gene'
mRNA	set type to 'mRNA'
CDS	set type to 'CDS'
ncRNA	set type to 'ncRNA'
Other Type	set type to another self entered type 'String'
3'antisense	add tag to point to 3'antisense transcript
5'antisense	add tag to point to 5'antisense transcript
fused	add tag to point to not complete covered transcript
predicted orientation	add tag to point to unknown orientation
dubious	add tag to point to some dubious looking transcript
Set Tag...	set own tag 'String' → /tag-name="note"
Set Score...	set score for feature 'float'

8.3 Create/Remove Features

The 'Create/Remove Feature' combobox contains functions to create features or remove existing ones. Every time one creates a 'Novel' or 'Template' feature it is within a also new created bucket. Its type is 'mRNA' by default. One can add more features to this bucket e.g. by selecting 'Add Gene & CDS' or choose the option 'Add Other Feature' which puts a new feature into this bucket. One always has to choose start and stop positions, Genescapes suggests. To create spliced features select a new pair of positions. This can be done by selecting the start/stop positions out of the combobox items 'Start' and 'Stop' or by enabling the 'Frame Annotation Mode' and the 'Selection Frame Mode' or do a double click on the wiggle plot.

More about annotate manual created features and about different modes you will find in section 11.

8.4 Feature Editor

The button 'Change Content' on the bottom of the information window opens a window which gives an editable view about all information a feature has.

- **Position Field**
In this text field every exon is represented by a new line containing two integers separated by ','. Every position pair has to be in a new line. One can simply change, add, remove or copy this position information and paste it to another feature.
- **Tag Field**
This textarea contains further information as tags. All changes have only affects to your temporary data and will not change the original file! Every tag has to be written in a new line. Please make sure to enter tags in the right format. Form: /tag-name="note"

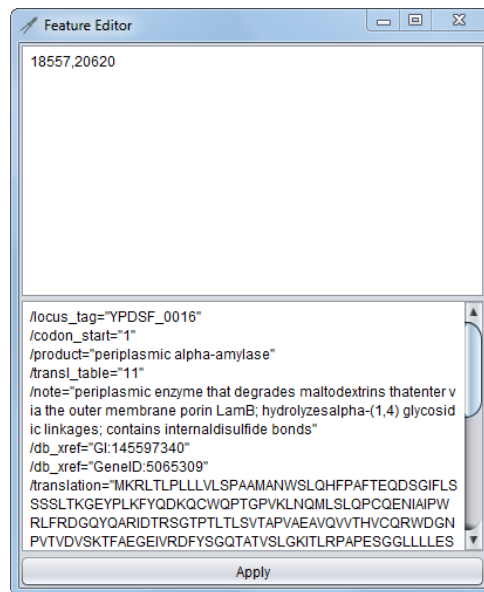


Figure 15: The feature editor enables changing information of features

9 Select Features

Selecting features enables the export and further analysis on the selected ones including blast, creating sequence logos or exporting the nucleotide or amino acid sequence. Selected features are marked by highlighted background.

- **Single Selection**
A single left click on a feature selects it. Pressing 'Ctrl' and click on an unselected feature, one adds to the list of already selected features. A click on selected ones will remove them. To clear the selection click on a free area.
- **Multiple Selection**
Check the status of the 'Frame Annotation Mode' and the 'Selection Frame Mode'. 'Frame Annotation Mode' must be disabled (change with key 'c') and 'Selection Frame Mode' (change with key 'f') must be enabled. This mode is the multiple selection mode. Dragging the mouse cursor paints an selection frame, all features in between this frame get selected. This function is also compatible with pressing 'Ctrl', which leads to the removal of already selected and the new selection of unselected features.

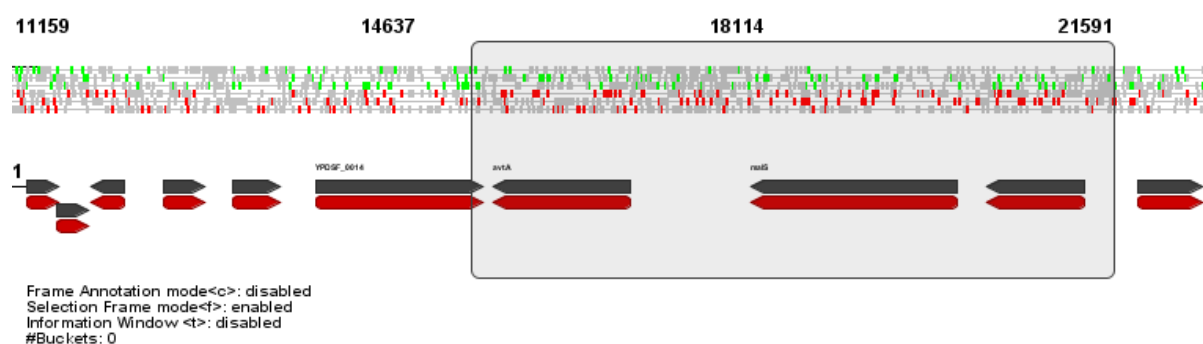


Figure 16: Mark an region by dragging the mouse cursor

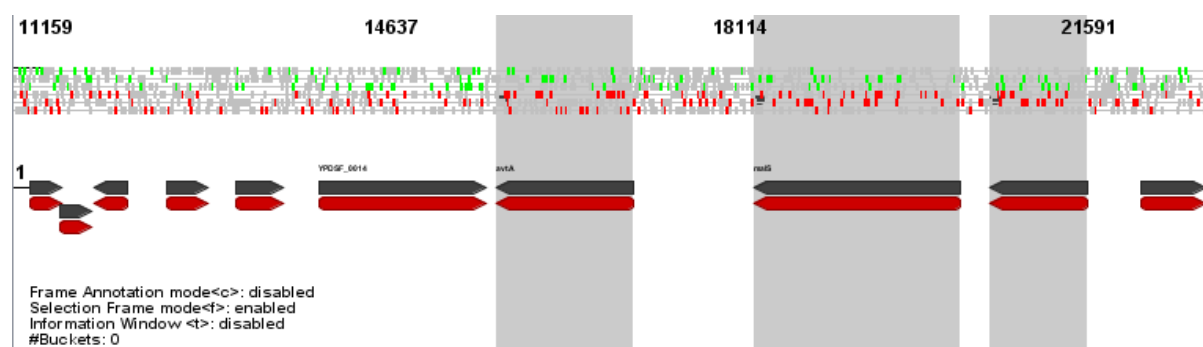



Figure 17: Selected features marked by highlighted background


10 Export Data

This section shows how to export features or your own created annotation.

10.1 Export Features

The selection allows to export the features in different file formats (FastA, GBK, GFF3, GTF) via the 'Viewer Menu' → 'Export'  → 'Export Selected Features...'.

10.2 Export ABC track

One can export the annotation track via the 'Viewer Menu' → 'Export'  → 'Export ABC Track' in GBK format. More about annotate your own features you find in section 11.

10.3 Overview Of Export Formats

GeneScapes enables the export of different data formats. Extracting information from selected Features, Blast results or sequence logos and also Track Settings and Session files. The following list gives an overview about all formats:

- Selected Features - See Import Formats for further details (FastA, GBK, GFF3, GTF)
- ABC-track GBK, also see section 11
- BLAST TXT + XML, also see section 14
- Screenshot/Weblogo PDF, also see section 13 [Crooks et al., 2004][Lowagie, 2007]
- GSS - The GSS GeneScapeSession format allows to reconstruct previous GeneScapes sessions, see also section 7
- Track Settings - The adaptation of visualisation can persistently be saved in a XML file which enables a quick change of the display setup, see also section 6

11 Annotate By Click (ABC)

The ABC function enables a fast manual (re)annotation of (novel) features. GeneScapes stores genomic features in so called feature buckets. Features of different types including 'mRNA' or 'CDS' coming from one common gene are put together in one bucket.

11.1 Create A New Bucket

As first step annotating a new feature one needs to create a new bucket with a new feature in it, 'Viewer Menu' → 'Annotator' → 'Create New Bucket' or via the shortcut 'Ctrl+y'. GeneScapes will request for the first position pair, the start and the stop position of the new feature. The different ways to set the positions are described in the next sections. For further position pairs, which means new exons, repeat the procedure of position selection. The type of a newly created ones is 'mRNA'.



Figure 18: GeneScapes requests for genomic positions

11.2 Via Information Window

The information window provides a lot of functionalities. It is possible to set positions of a new annotated feature via the information window. Move the mouse cursor over the feature which positions you want to set to the new annotated feature and click 'Start' or 'Stop' in the left combobox named 'Create/Remove Feature'. Do this step again to generate a further exon.

11.3 Via Wiggle Plot

It is possible to set positions of a new annotated feature via the wiggle plot. With a double left click on the wiggle plot (the score tooltip must be shown) one can set a position to a new annotated feature. Do this step again to generate a further exon.

11.4 Via Selection Frame

It is possible to set positions of a new annotated feature via the selection frame. Dragging the mouse leads to different result, notice that the 'Frame Annotation Mode' and the 'Selection Frame Mode' is enabled. Do this step again to generate a further exon.

11.5 Add Gene And CDS Feature

To add gene and CDS features to the newly created bucket with your mRNA feature in it, select 'Add Gene & CDS Feature' in the 'Create/Remove Feature' combobox of the information window by hovering your annotated transcript.

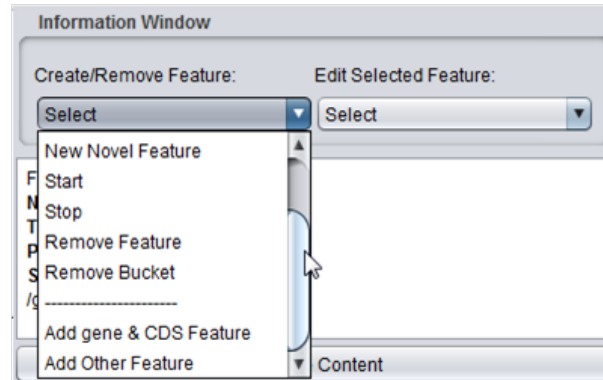


Figure 19: Tool box for annotating

The gene feature has the absolute positions of the mRNA feature, the most left and the most right one. The ORF finder of GeneScapes calculates the longest ORF and adds it to the bucket as CDS feature with the translated amino acid sequence. Notice to select the right translation table for your organism in the 'Viewer Menu' → 'Annotator' → 'Translation Table'.

11.6 Other Feature Types

There are many more feature types than 'gene', 'mRNA' and 'CDS'. To set another type hover the feature you want to change with the mouse cursor to open the information window and select 'Other Type' in the 'Edit Selected Feature' combobox. A pop-up opens, enter the type here.

12 Visualise Orthologs

One can visualise orthologue gene pairs at a glance by selecting a feature and execute the 'Find Orthologue Of Selected' in the menu under 'Analyse' (figure 20). This is up to now only possible if there are two or more GBK files loaded and the gene feature contains a specific tag, described in the next subsection.

12.1 Prepare Files

Your GBK files need a specific tag to identify the ortholog. The tag must be equal to:

```
/note="ref\_name orthologue: GeneXY"
```

where 'ref_name' is the name of the organism and 'GeneXY' is the name of the orthologues gene. In the GBK file it looks similar to:

```
[...]
gene          15906..16668
              /gene="SC5314_nTAR1.1"
              /note="C.dub orthologue: CD36_nTAR1.1"
mRNA          15906..16668
              /gene="SC5314_nTAR1.1"
              /transcript_id="tid_SC5314_nTAR1.1.1"
              /note="experimentally verified"
CDS           16104..16631
              /gene="SC5314_nTAR1.1"
              /note="homologues tblastn hit in Chr1_C.dubliniensis_CD36"
[...]
```

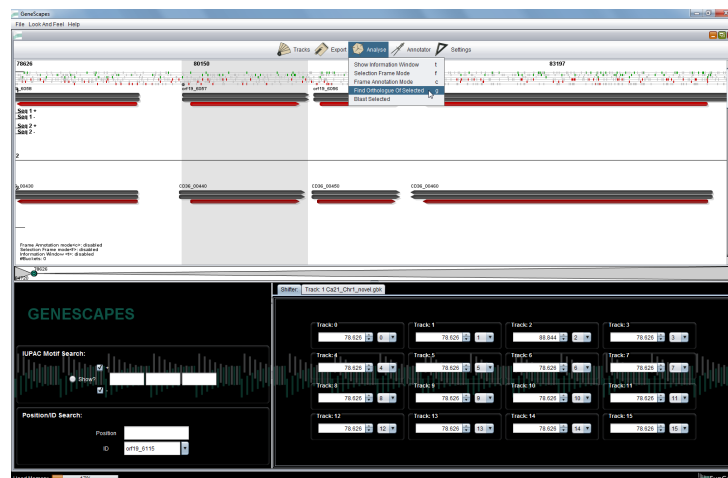


Figure 20: Find orthologs with GeneScapes

12.2 Anchor & Shift Positions

Orthologue gene pairs differ in their chromosomal/contig positions therefore a shift function is required to visualise the pairs at a glance. In GeneScapes every track has its own position which is shiftable. By default all tracks are anchored to itself. By dragging up the divider below the position slider the anchor table on the right side becomes visible. The numbered counter boxes enable a shift of single position by the up and down buttons or entering a position. The combobox on the right side of the position counter indicates the track index to which this track is anchored to. So if one wants to visualise orthologues genes and their associated reads, set the anchor of the track where the reads are stored in to the index

of the track number where the annotation is loaded in. If GeneScapes shifts the track of the genes the tracks anchored to this track gets also shifted.

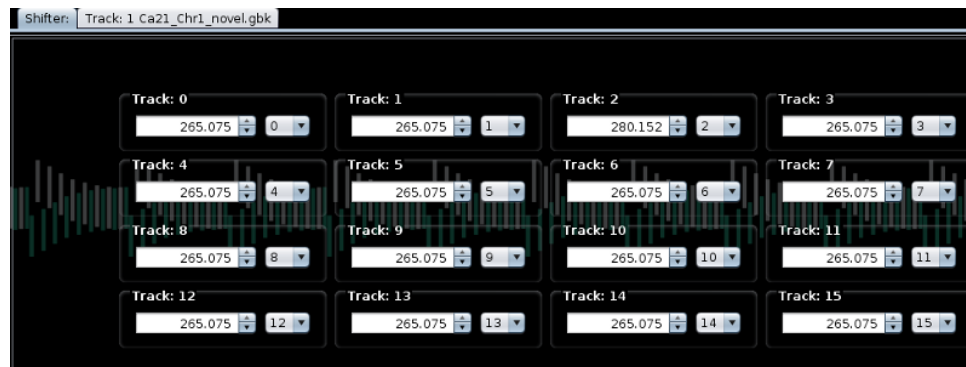


Figure 21: Table to shift positions for each track individually

12.3 Reset Positions

To reset the positions of the tracks because orthologs have different genomic positions on chromosomes or contigs click 'Viewer Menu' → 'Tracks' → 'Reset Track Positions' or the shortcut 'Ctrl+r'.

Anchoring tracks is also helpful exporting a nucleotide sequence from a track which does not contain sequence information. One can link this track to another track e.g. FastA file with the needed sequence and export it. So if a reference genome and the annotation is loaded one can easily export all sequences of all genes.

13 Create Sequence Logos

GeneScapes is linked to a web based application called WebLogo3 [Crooks et al., 2004].

Definition:

"A sequence logo is a graphical representation of an amino acid or nucleic acid multiple sequence alignment[...] Each logo consists of stacks of symbols, one stack for each position in the sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino or nucleic acid at that position. In general, a sequence logo provides a richer and more precise description of, for example, a binding site, than would a consensus sequence."

Via the 'Viewer Menu' → 'Export' → 'Export Sequence Logo...' it can be created. One can select either the 'DNA', 'RNA' or 'AA' sequences of the features. The sequences of the selected features get shortened to the shortest one that all sequences have same length, which is necessary for the sequence logo. Use the shift function to visualise your orthologues genes so that the nucleotide positions of selected features are above each other e.g. 'ATG' as start codon for orthologues ORFs for shifting tracks as shown in figure 22.

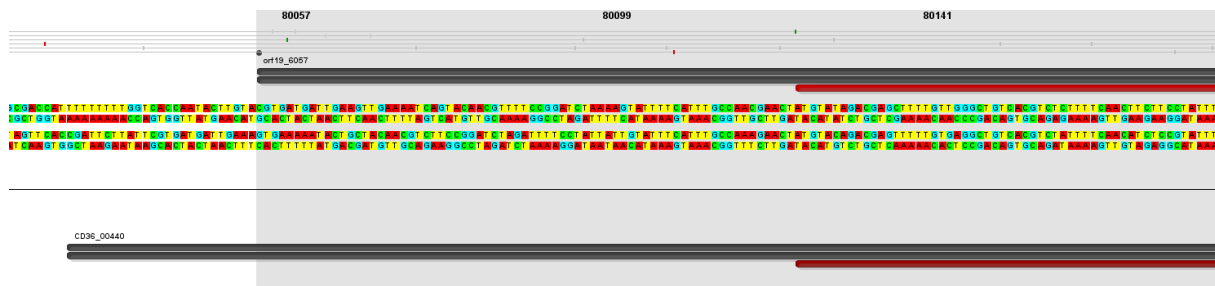


Figure 22: Use the shift function to match positions

The resulting PDF file should look similar to this:

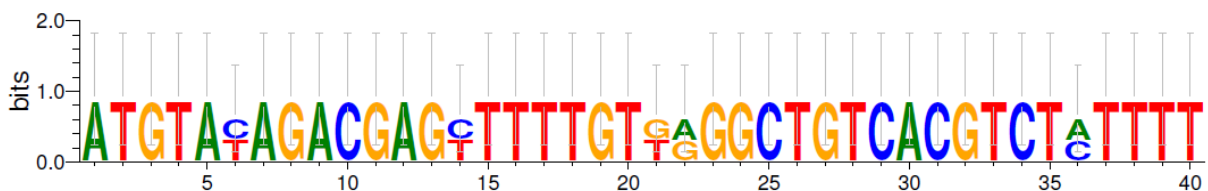



Figure 23: The first few nucleotides of the sequence logo starting with the start codon

14 Blast Features

GeneScapes is linked to the NCBI blastn web server.

Definition:

"The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families."

First step is to select all features one wants to blastn against the nucleotide collection (nt). Then execute the blast via the 'Viewer Menu' → 'Analyse'  → Blast Selected. All hits are reported in the XML file. A second resulting file is the table format which contains all hits with an evalue ≤ 1 .

15 Conclusion

GeneScapes enables a practical manual curation platform for transcripts in genetic landscapes and gives the possibility to add biological knowledge to the data directly by the biologist himself. It is already used for manual annotation of pathogenic fungi [Grumaz et al., 2013]. Visualisation works on all organisms, independent of complexity. The variety of visual settings can bring a clear setup to all data sets, which is necessary for inspection. For lower organisms like fungi or bacteria it annotating works fast and stable, for complex transcriptomes it might take too a long time annotating big numbers of exons and isoforms. For future work it is designated to add external tools for gene- and orf finding, splice site prediction and implement new plots for the detection of SNPs and allele specific nucleotides.

16 Missing Manual Parts

- findings motifs
- jump to position or gene via feature table or position and gene id entry
- change look & feel layout
- flash screen capturing tutorial
- tips & tricks

References

- [Cock et al., 2009] Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and de Hoon, M. J. L. (2009). Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423.
- [Crooks et al., 2004] Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004). Weblogo: a sequence logo generator. *Genome research*, 14(6):1188–1190.
- [Grumaz et al., 2013] Grumaz, C., Lorenz, S., Stevens, P., Lindemann, E., Schöck, U., Retey, J., Rupp, S., Sohn, K., et al. (2013). Species and condition specific adaptation of the transcriptional landscapes in candida albicans and candida dubliniensis. *BMC genomics*, 14(1):1–17.
- [Lowagie, 2007] Lowagie, B. (2007). itext in action. *Manning Publications Co.*
- [Martin and Wang, 2011] Martin, J. A. and Wang, Z. (2011). Next-generation transcriptome assembly. *Nature Reviews Genetics*, 12(10):671–682.
- [Mortazavi et al., 2008] Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by rna-seq. *Nature methods*, 5(7):621–628.
- [Pruitt et al., 2007] Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2007). Ncbi reference sequences (refseq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research*, 35(suppl 1):D61–D65.
- [Slominski, 2006] Slominski, A. (2006). Mxp1: Xml pull parser 3rd edition (xpp3).
- [Trapnell et al., 2009] Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). Tophat: discovering splice junctions with rna-seq. *Bioinformatics*, 25(9):1105–1111.
- [Trapnell et al., 2010] Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010). Transcript assembly and quantification by rna-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology*, 28(5):511–515.
- [Wang et al., 2009] Wang, Z., Gerstein, M., and Snyder, M. (2009). Rna-seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1):57–63.

Iconset: Vintage Icons by DesignContest (49 icons)
<http://www.designcontest.com/free-icons/>