

NGS at geneXplain GmbH

Introducing GenomeTraveler

geneXplain GmbH



Products

- ▶ Bioinformatics and systems biology: geneXplain platform
- ▶ Bioinformatics and NGS: IMC and GenomeTraveler
- ▶ Cheminformatics: PASS, PharmaExpert, GUSAR

Systems Biology and Cheminformatics

- ▶ geneXplain platform
 - Collection, storage and analysis of experimental data
 - Network clustering and search for master regulators
 - Graphical programming of workflows
 - Possibility to write new scripts and add-ons
- ▶ PASS and PharmaExpert
 - Qualitative biological activity prediction for libraries of chemical compounds
 - Drug-drug interactions and identification of most promising candidates
- ▶ GUSAR
 - Quantitative biological activity prediction for libraries of chemical compounds
 - (Q)SAR model building

geneXplain GmbH



Products

- ▶ Bioinformatics and systems biology: geneXplain platform
- ▶ **Bioinformatics and NGS:** IMC and **GenomeTraveler**
- ▶ Cheminformatics: PASS, PharmaExpert, GUSAR

Contents

- ▶ Primary analysis
 - *de novo* assembly
 - Scaffolding
 - Mapping
 - Import of mapping results
- ▶ Secondary analysis
 - Profile analysis
 - Mutation analysis
 - Expression analysis
- ▶ Comparative genomics
 - *in silico* cloning
 - Genome analysis
 - Array analysis
- ▶ Why you should try GenomeTraveler

Primary Analysis

- » Getting started with NGS data

General Information



- ▶ GT supports NGS data from the following sources:
 - ABI SOLiD (CSFastA)
 - Illumina GA (FastQ)
 - Roche/454 GS FLX/Junior

- ▶ It is possible to map NGS sequences against sequence data in any of these formats:
 - ABI/SCF
 - GenBank
 - FastA

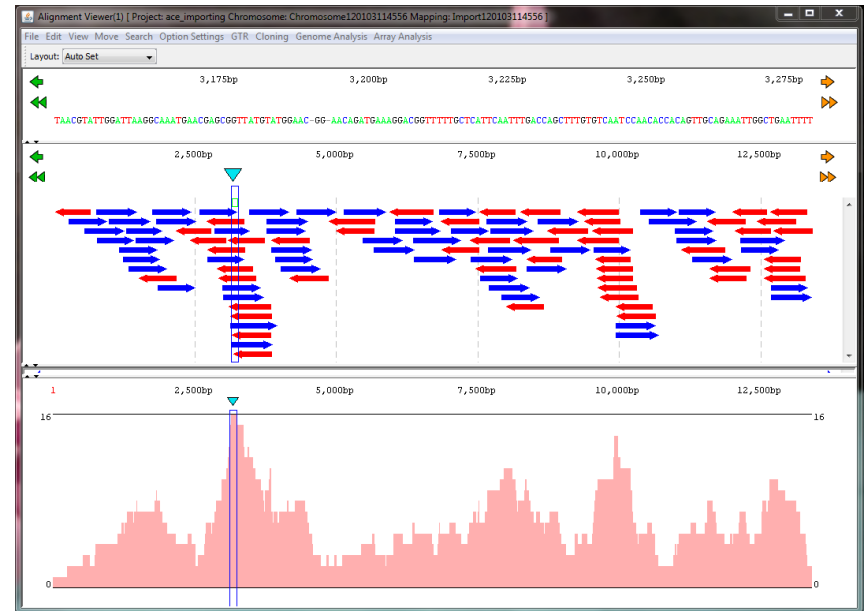
de novo Assembly

▶ *de novo* assembly can be done with 200 bp or larger reads, saved in the following formats:

- ABI/SCF/SFF/PHD
- FastA/FastQ/CSFastA
- Solexa Raw
- GenBank/EMBL
- TXT

▶ GT also supports the import of assembly files from:

- Velvet (AFG)
- Phrap (ACE)

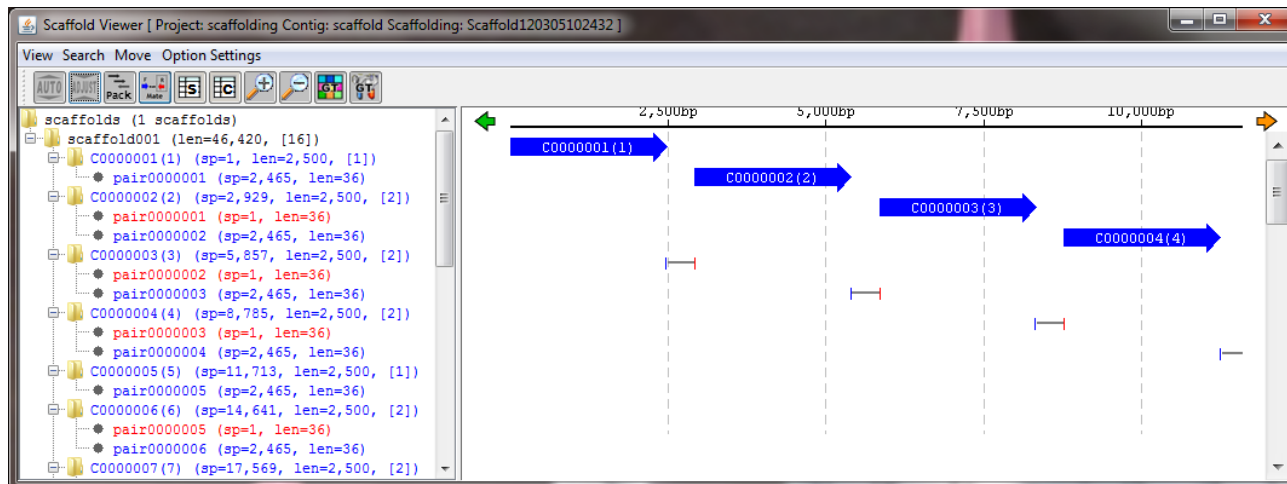


Example: assembly results imported from Phrap.

Scaffolding

The *de novo* assembly of large genomes results in a large number of contigs. These can be combined using paired-end short reads.

Contig positions in a scaffold can be viewed in GT's "Scaffold Viewer".



Mapping Functions

The mapping algorithm is based on NCBI Blast results for reference sequences compared with NGS short reads.

fragment
short reads

paired-end
short reads

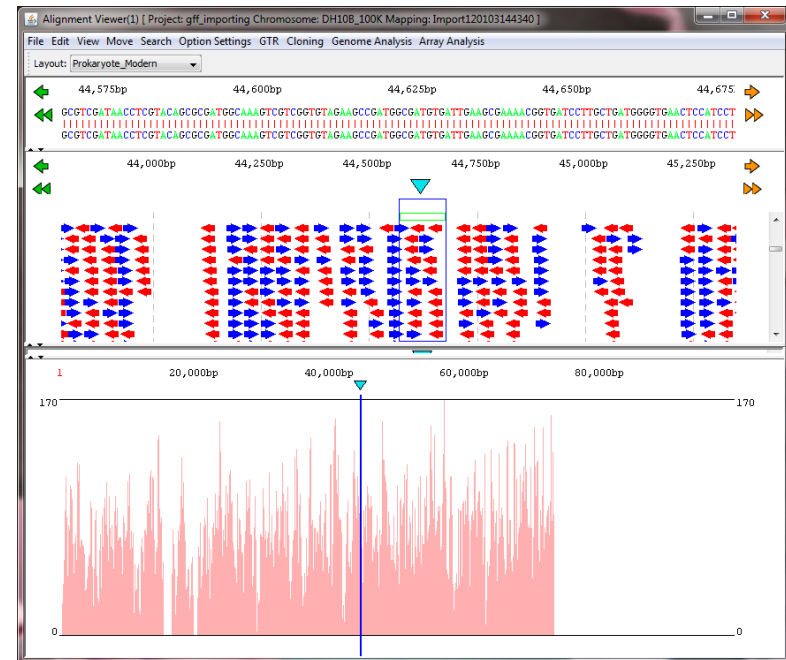
multiple sites

multiple
chromosomes
or contigs

Import of Mapping Results

- ▶ GenomeTraveler can also import mapping results from other software, namely:
 - SAMtool (SAM/BAM)
 - SOLiD (GFF)

- ▶ Mapping and import results can be viewed in the "Profile and Alignment Viewer" of GT.



Example: imported GFF file.

Primary Analysis Summary

Long read *de novo* assembly and import of assembly results from Phrap and Velvet.

Scaffolding using paired-end short reads.

GenomeTraveler has a number of options for the primary analysis of NGS data.

Mapping: single-end, paired-end, multiple chromosome/contig/site mapping.

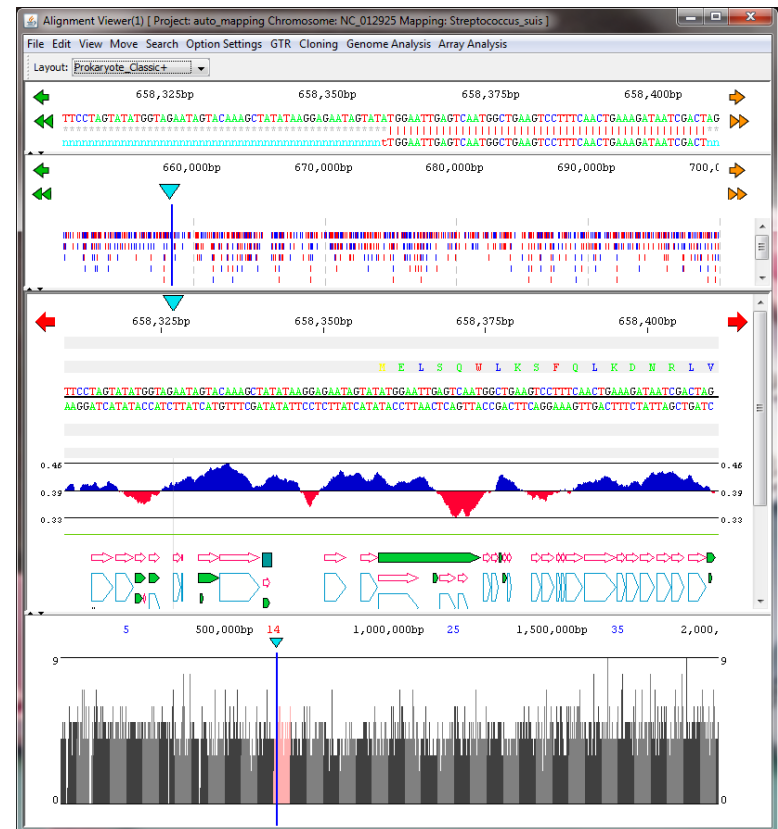
Import of mapping result files from SAMtool and SOLiD.

Secondary Analysis

»» Analysis of mapped data

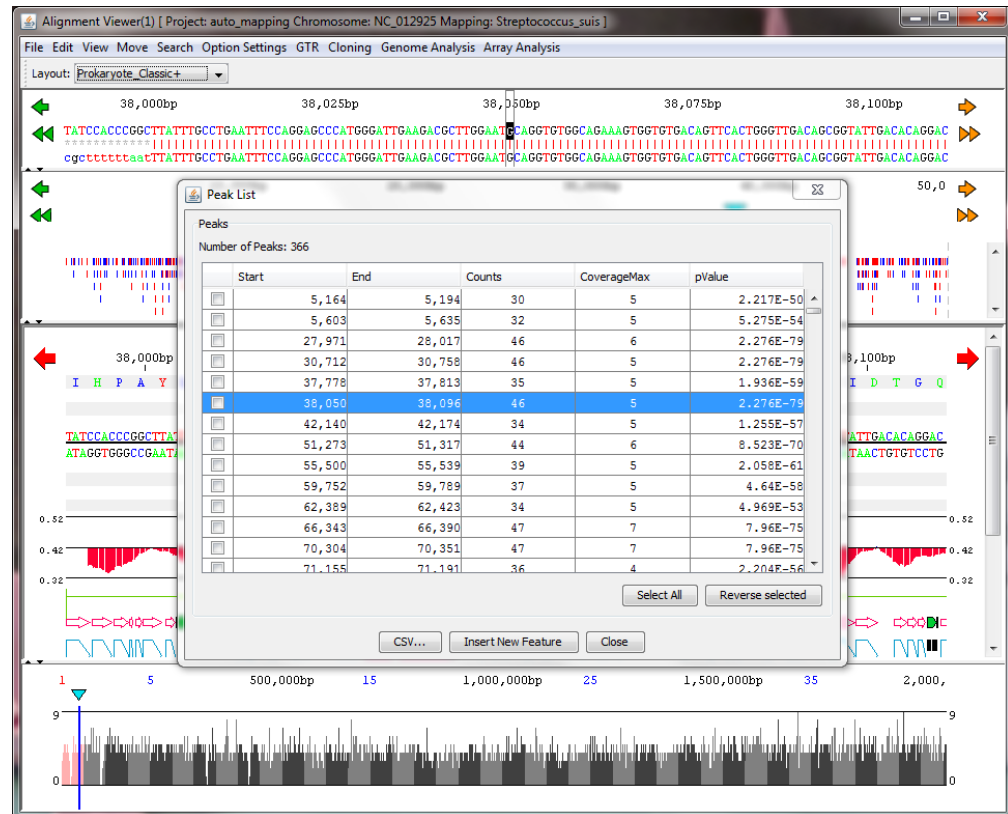
General Information

- ▶ Once NGS data is mapped, it can be analyzed in the "Profile and Alignment Viewer".
- ▶ The viewer consists of four panels:
 - Whole genome navigation panel
 - Profile panel
 - Genome (feature) map
 - Sequence alignment



Profile Analysis

- ▶ Peak detection can be done for one, a few or all contigs in one alignment.
- ▶ Peaks can be saved as CSV file or inserted as new features into the genome map.

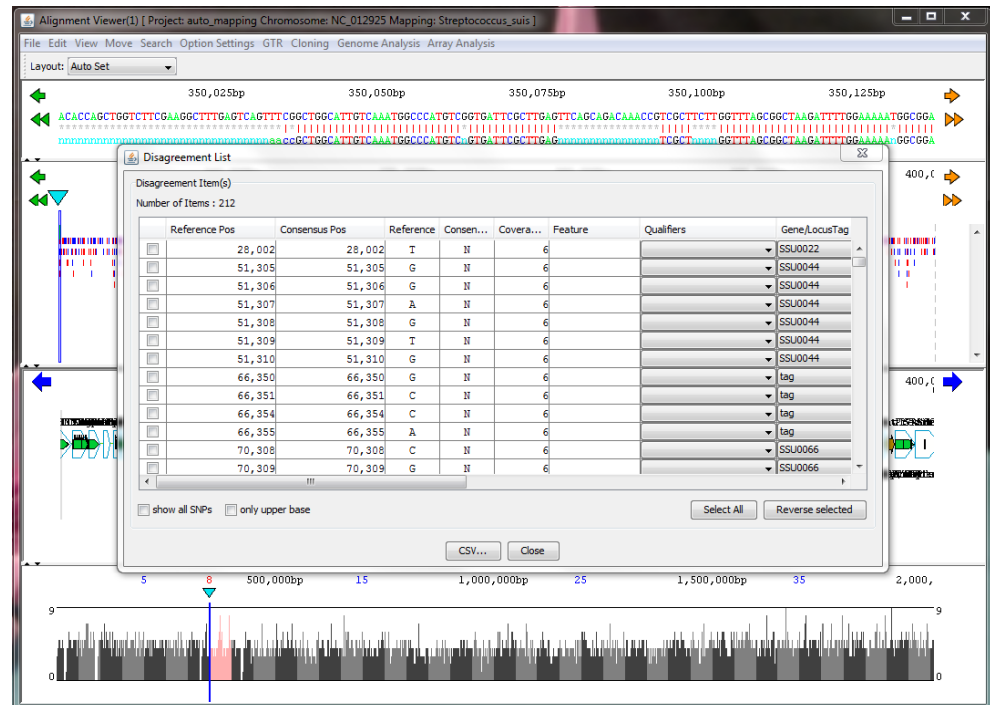


Mutation Analysis

- ▶ The "Disagreement List" is a list of bases which differ between the reference and the consensus of mapped reads (example).

- ▶ SNP detection
 - Based on disagreement list
 - Feature mapping
 - Link generation with JSNP

- ▶ Indel detection



Expression Analysis

- ▶ GT provides profiles and a comparison with expression profiles of tiling arrays for:
 - RNA-Seq
 - ChIP-Seq

Secondary Analysis Summary

Profile analysis: detect and register peaks.

Mutation analysis: find divergent bases, SNPs and insertion/deletion sites.

GenomeTraveler provides three fields for secondary analysis of NGS data.

Expression analysis: GT provides profiles for RNA- and ChIP-Seq data.

These are GT specific functions that can be accessed in the "Profile and Alignment Viewer".

Comparative Genomics

» Functions inherited from IMC

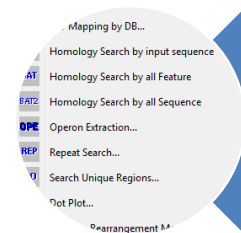
General Information

GenomeTraveler is based on IMC (*in silico* Molecular Cloning), which offers three major *in silico* experimentation areas.

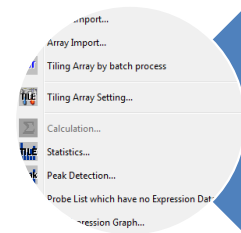
These are also featured in GT and can be accessed in the "Profile and Alignment Viewer".



Cloning



Genome analysis



Array analysis

in silico Cloning

The "Cloning" menu lists a number of possible actions for *in silico* molecular cloning experiments. Included are all necessary functions for:

- PCR (including gel electrophoresis of results)
- Ligation and sequence editing
- Restriction enzymes and digestion (example below)

The screenshot shows the 'Cloning' menu with various options. Below it, a window displays a DNA sequence with restriction sites identified. The sequence is: L K E F V S S E D Q L M I L G G D G T L S K A L R F W P A S L P F A Y Y P T. The highlighted recognition sites are: T T A A G G A G T T T G T C A G T A G T A A G A T C A A T T A A T G A T T T T A G G C G G A G A C G G C A C A C T A T C T A A A G C T T T G C G T T T T T G G C C A G C T A G T C T A C C G T T T G C T T A T T A T C C A A C A A T T T C C T C A A A C A G T C A T C A C T T C T A G T T A A T T A C T A A A A T C C G C C T C T G C C G T G T A T A G A T T T C G A A A C G C A A A A A C C G G T C G A T C A G A T G G C A A A C G A A T A A T A G G T T G.

Enzyme	Count	Enzyme	Recognition site	Position	Dam/...	Upstream...	Downstre...
<input checked="" type="checkbox"/> HindIII	606	<input checked="" type="checkbox"/> HindIII	5' A^AGCTT 3' 3' TTCGA^A 5'	2970		"SSU0003"	"SSU0003"
<input checked="" type="checkbox"/> NdeI	168	<input checked="" type="checkbox"/> HindIII	5' A^AGCTT 3' 3' TTCGA^A 5'	4316		"SSU0005"	"SSU0006"
<input checked="" type="checkbox"/> XbaI	332	<input checked="" type="checkbox"/> HindIII	5' A^AGCTT 3' 3' TTCGA^A 5'	4332		"SSU0005"	"SSU0006"
<input checked="" type="checkbox"/> XhoI	80	<input checked="" type="checkbox"/> HindIII	5' A^AGCTT 3' 3' TTCGA^A 5'	5635		"SSU0006"	"pth"
		<input checked="" type="checkbox"/> XbaI	5' T^CTAGA 3' 3' AGATC^T 5'	6777		"trcF"	"trcF"
		<input checked="" type="checkbox"/> XbaI	5' T^CTAGA 3' 3' AGATC^T 5'	6888		"trcF"	"trcF"

Buttons at the bottom: Select All, Reverse selected, More Search..., Digestion..., Insert New Feature, Gel Electrophoresis..., CSV..., FastA..., Close.

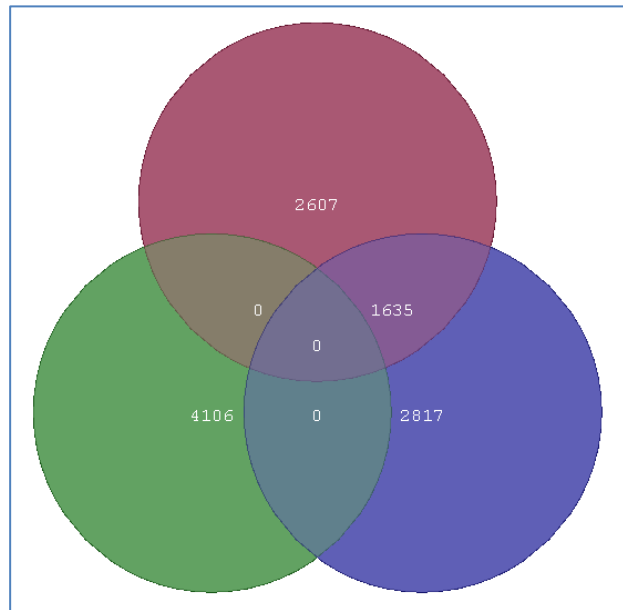
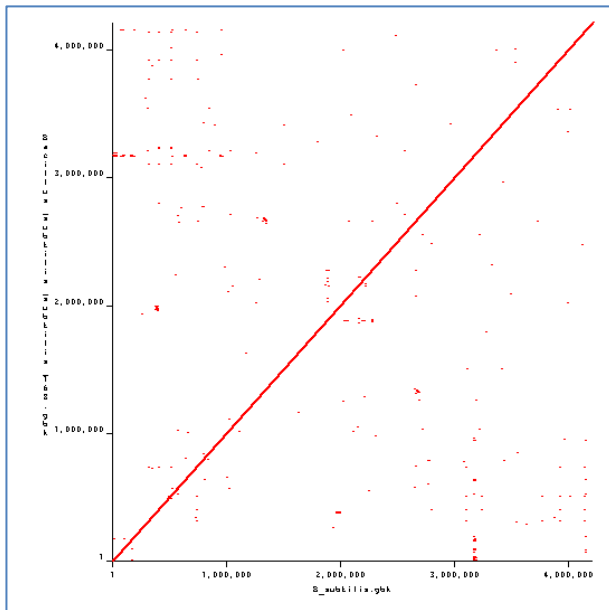
The 'Cloning' menu includes the following options:

- RE Recognition...
- Primer Registration...
- PCR Primer Design...
- Batch PCR Primer Design...
- Whole Genome PCR Primer Design...
- PCR [current]...
- PCR [files]...
- In Fusion...
- Ligation...
- Plasmid Map Viewer...
- Edit Sequences...
- Cutting off Sequences...
- Add Enzyme Site...
- Add T-Base to both 3' ends
- Blunting
- Phosphorylation

Genome Analysis

The "Genome Analysis" menu is filled with functions from statistics to visualization to working with sequences:

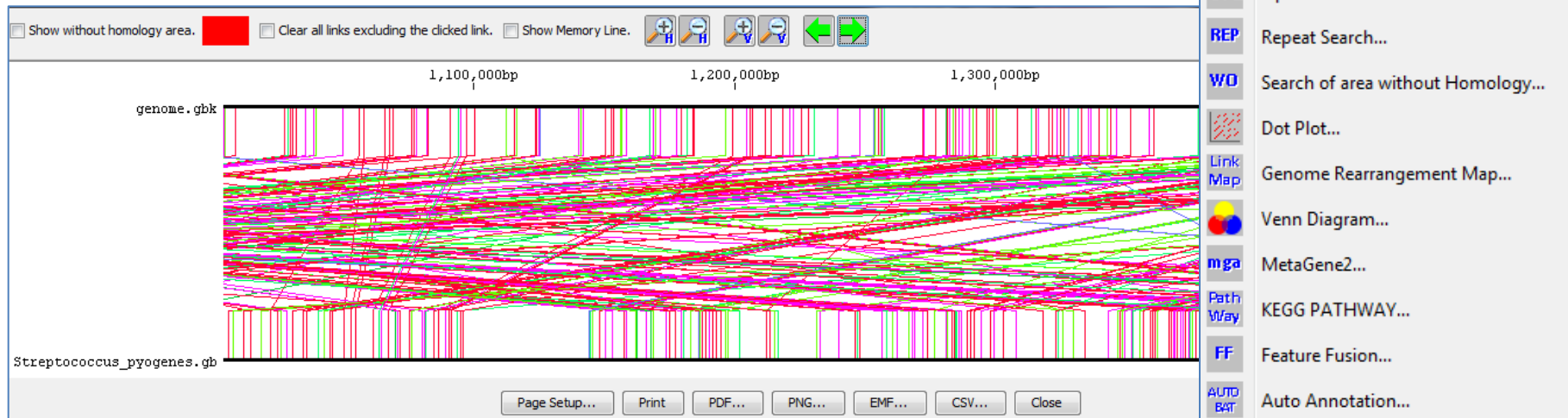
- Translation
- Annotation
- EST and trace mapping
- Homology searches
- Dot plot, Venn diagram (examples below), KEGG pathways



Genome Analysis	Array Analysis
	Show Codon Usage...
	Statistics...
ORF	ORF Extract...
TR	Translation
	EST Mapping...
Trace	Trace Mapping...
A.A.	A.A. Sequence Mapping...
EST DB	EST Mapping by DB...
HM	Homology Search by Input Sequence
BAT	Homology Search by all Feature
BAT2	Homology Search by all Sequence
OPE	Operon Extract...
REP	Repeat Search...
WD	Search of area without Homology...
	Dot Plot...
Link Map	Genome Rearrangement Map...
	Venn Diagram...
mga	MetaGene2...
Path Way	KEGG PATHWAY...
FF	Feature Fusion...
AUTO BAT	Auto Annotation...

Genome Analysis

The genome rearrangement map, as part of the genome analysis functionality, compares the distribution of features on your mapped sequence compared to any reference sequence you set (in GenBank/EMBL/GPFF formats), as seen in the example below.















Array Analysis

The "Array Analysis" menu contains all functions necessary for the creation and analysis of array data:

- Probe design (example below)
- Import of probe and array data
- Statistics and peak detection
- Two expression graphs and clustering

Array Analysis

-  Probe Design...
-  Probe Import...
-  Array Import...
-  Tiling Array by batch process
-  Tiling Array Setting...
-  Calculation...
-  Statistics...
-  Peak Detection...
-  Probe List which have no Expression Data(s)...
-  Two Expression Graph...
-  Clustering...
-  Read Clustering Result File...

Probe Design

Probe length in bp:

Step size in bp:

Strand: Forward Reverse Both

The area that overlaps with Feature is excluded from the Probe candidate.

- 10_signal
- 35_signal
- 3'dip
- 3'UTR
- 5'clip

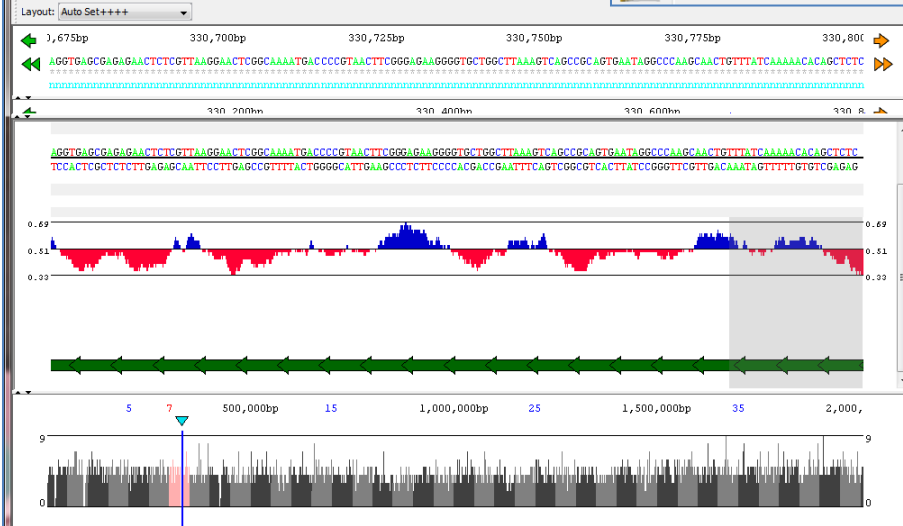
Alignment Viewer(1) [Project: auto_mapping Chromosome: NC_012925 Mapping: Streptococcus_suis]

File Edit View Move Search Option Settings GTR Cloning Genome Analysis Array Analysis

Layout: Auto Set++++

330,675bp 330,700bp 330,725bp 330,750bp 330,775bp 330,800

AGGTGAGCGGAGAACTCTGTTAAGGAATCTGGCAAAATGACCCCGTAACCTCGGAGAAAGGGGTCTGGCTTAAAGTCAGCCGCAGTGAATAGGCCAAGCAACTGTTATCAAAAACACAGCTCTC
 TCACCTGCCCTCTTGAAGCAATCTCTTGAAGCCGTTTACTGGGCACTTGAAGCCCTCTCCCAAGACGAAATTCAGTCGGCTCACTTATCGGGTCTGTGACAAATAGTTTTTGTGTGAGAG



0.69 0.51 0.33

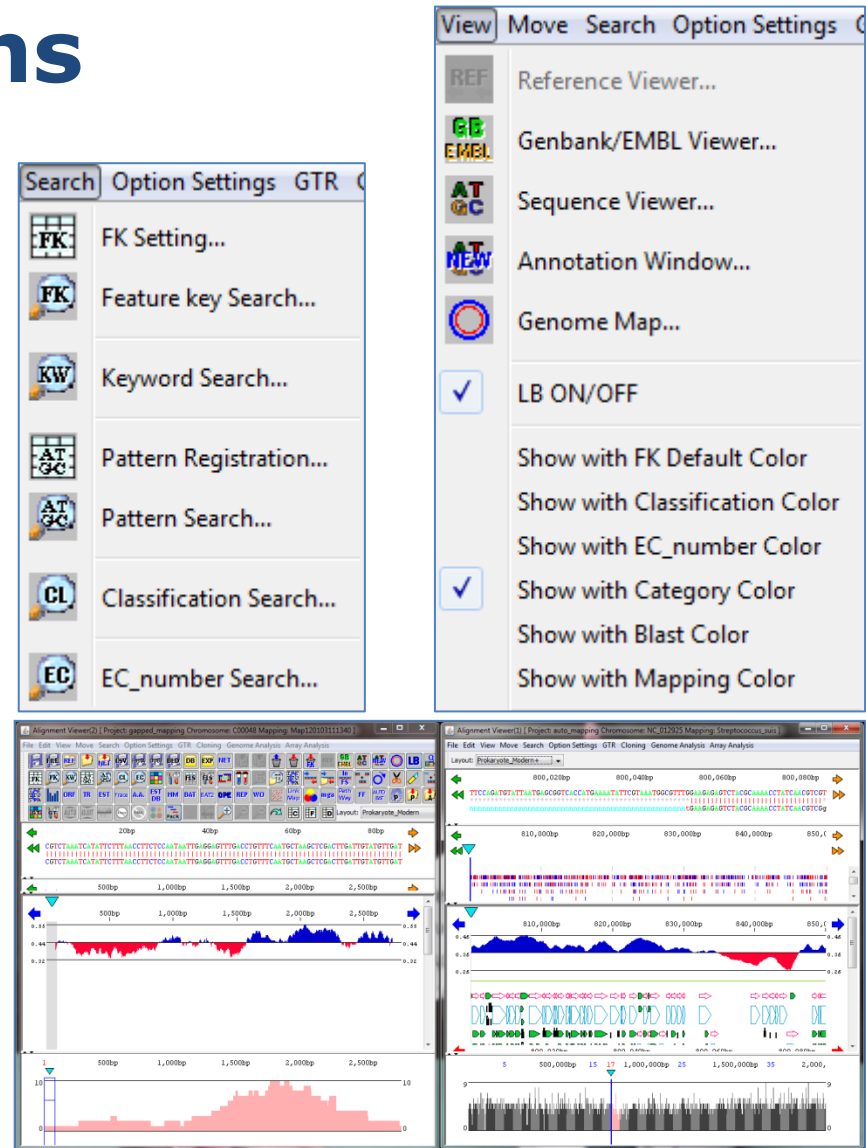
5 7 15 25 35 2,000

9 0

Additional Functions

The "Profile and Alignment Viewer" has more to offer, though:

- ▶ Search for pattern matches, keywords, and more.
- ▶ View your sequence in different ways.
- ▶ Customize the genome feature map and hide the program's quick access buttons.



Comparative Genomics Summary

in silico cloning options include PCR, restriction and ligation of sequences.

Genome analysis features translation, annotation, homology searches and more.

With GenomeTraveler you can work *in silico* with your NGS data.

Array analysis offers probe design, data import, peak detection and more.

The additional functions of the "Profile and Alignment Viewer" render GT more flexible.

Why You Should Try GenomeTraveler

» Advantages of the Software

Key Notes

- ▶ GenomeTraveler has a number of options for the primary analysis of NGS data.
- ▶ GenomeTraveler provides three fields for secondary analysis of NGS data.
- ▶ With GenomeTraveler you can work *in silico* with your NGS data.

- ▶ You can download a free two-week trial version of GenomeTraveler at <http://www.genexplain.com/genome-traveler>