TGAC Browser: visualisation solutions for big data in the genomic era

- Anil S. Thanki
Scientific Programmer – Sequencing Informatics
Anil.Thanki@tgac.ac.uk

@anilthanki and @tgacbrowser

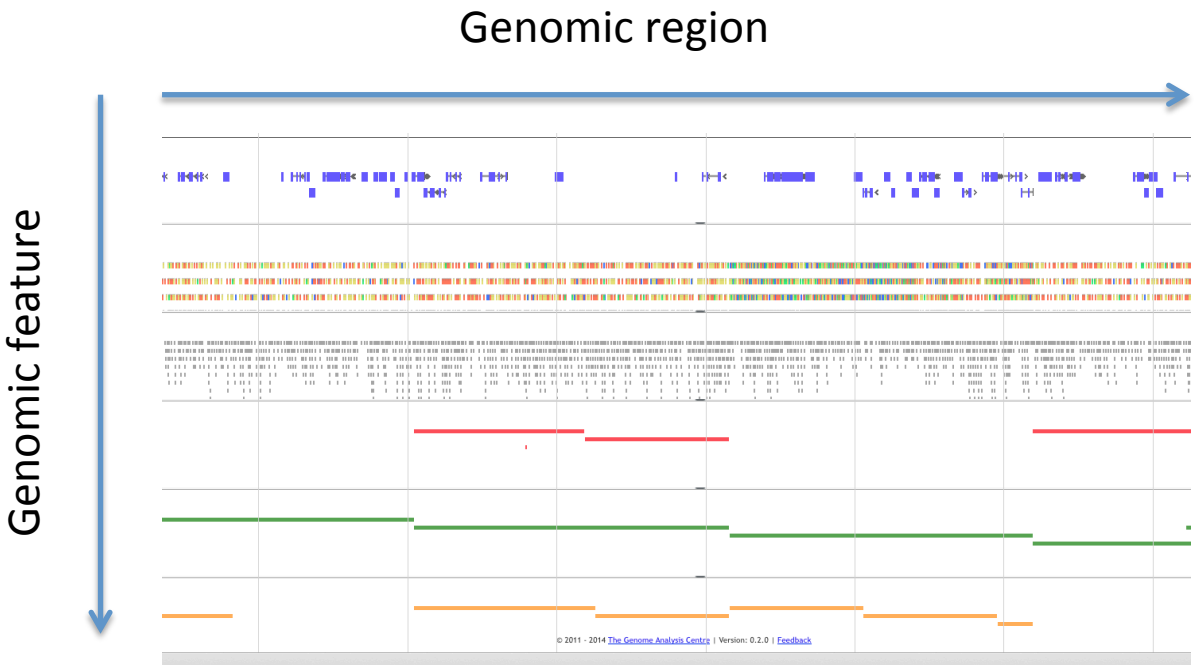
July 11, 2014
BOSC

#Poster 6

Genome Browsers

TGAC
The Genome Analysis Centre™

In bioinformatics, a **genome browser** is a graphical interface for display of information from a biological database for genomic data.



TGAC Browser

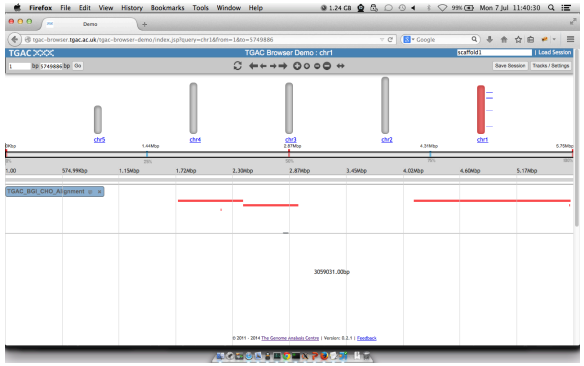
TGAC
The Genome Analysis Centre™

TGAC Browser developed at TGAC from scratch, works on top of Ensembl Core database

TGAC Browser server-client

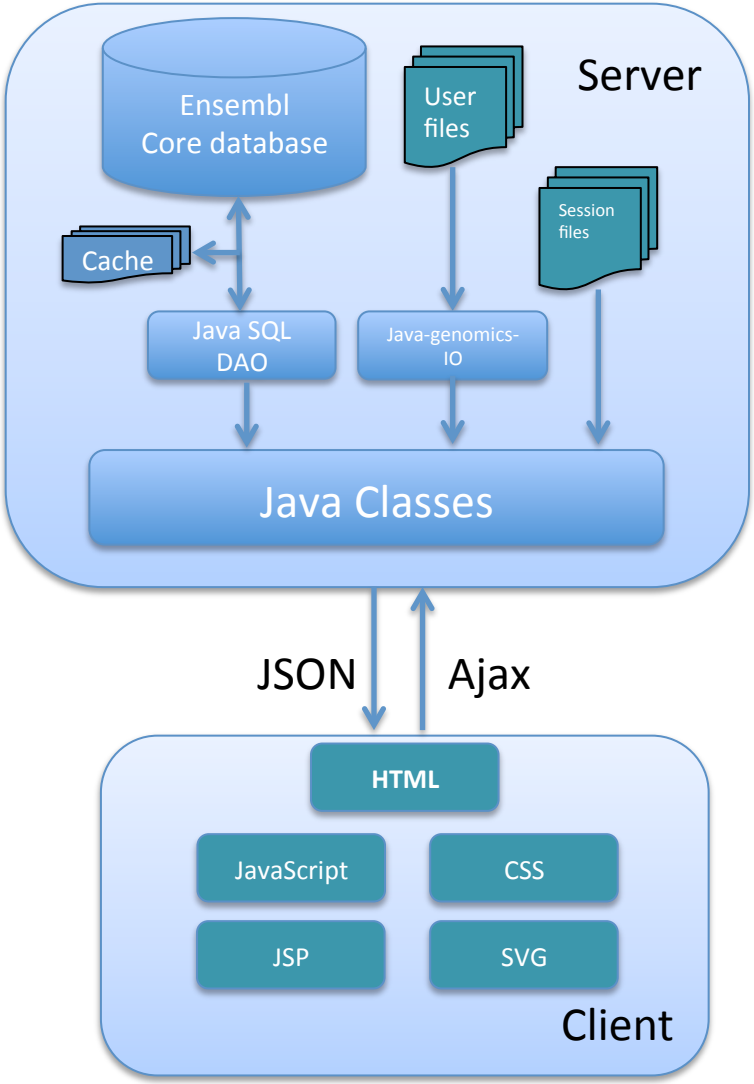
TGAC
The Genome Analysis Centre™

- Heavy database queries and parsing on server
- Transfer information in Text format
- Utilise Client system to generate and render images



- Performance and easy access of data from server implementation
- Flexibility of web Browsers for sharing data

TGAC Browser server-client

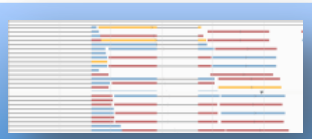


Supported Data

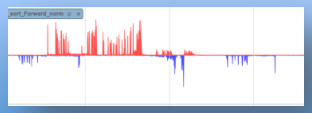
TGAC
The Genome Analysis Centre™

Genomics Data

Ensembl core



SAM/BAM



Wig/BigWig

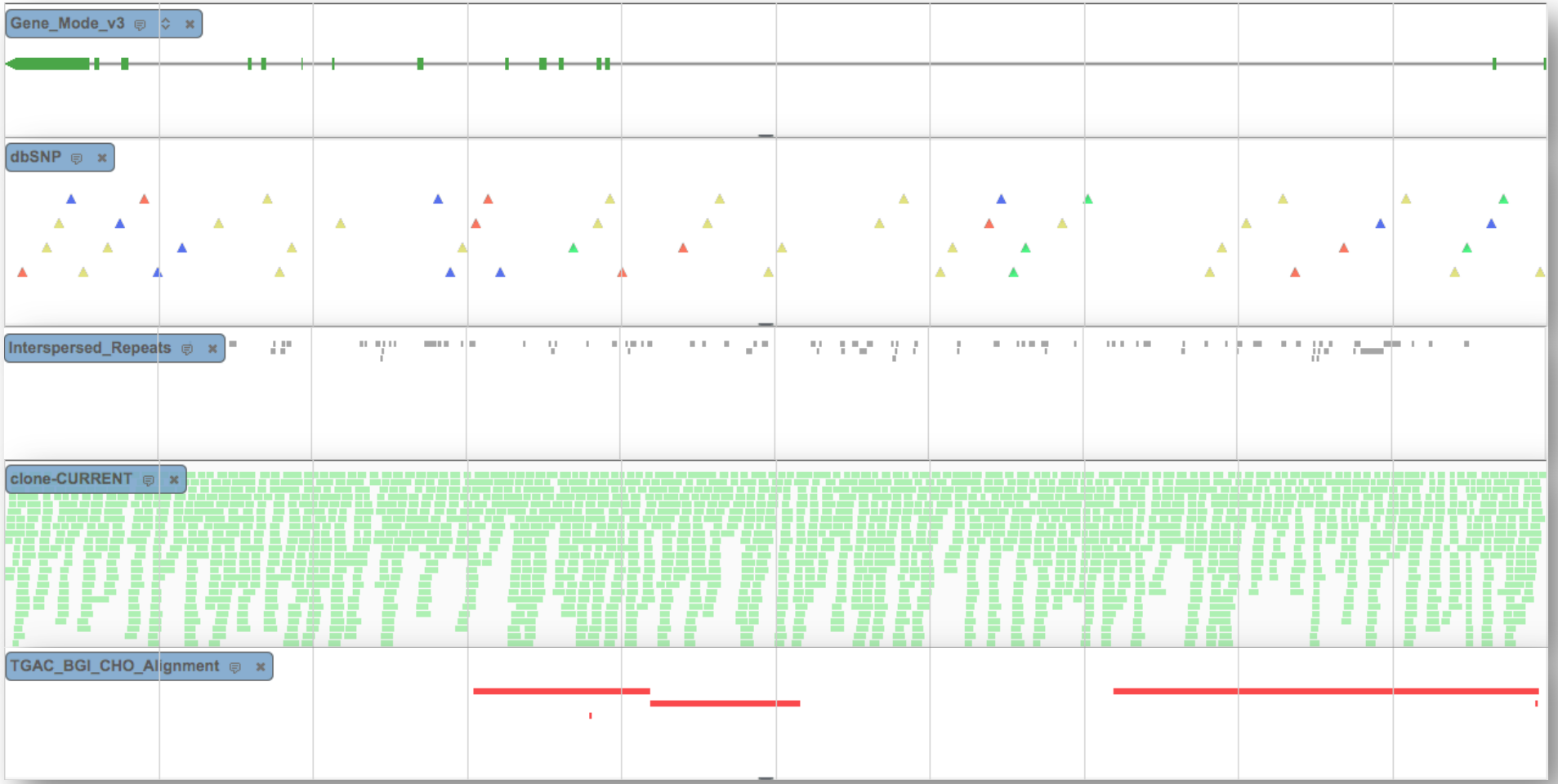
GFF

VCF

Ensembl Data

TGAC
The Genome Analysis Centre™

- Genomic features from Ensembl Database
  - Genes, SNPs, repeats, assembly, alignments, markers, etc



BAM/SAM

Visualising reads directly from SAM/BAM file

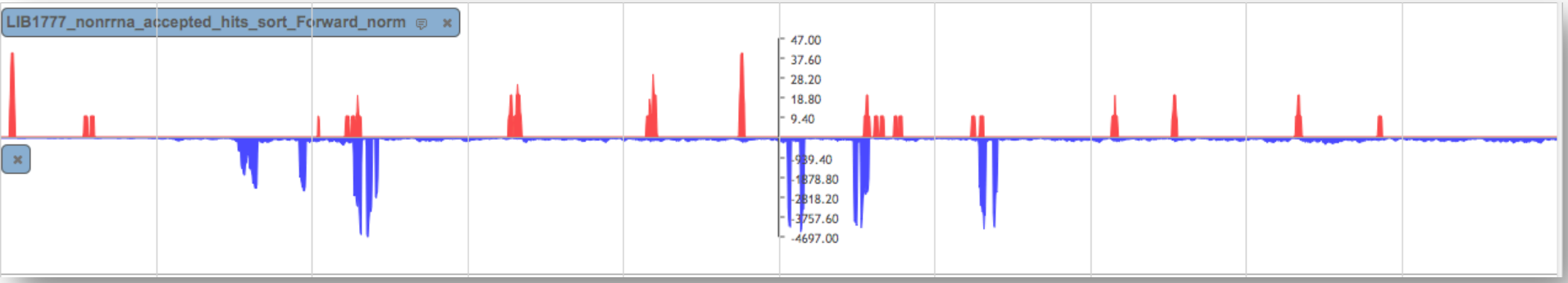
- Coloured Paired end reads
  - Blue First in Pair
  - Brown Second in Pair
  - Orange unpaired
  - Skipping deletions

Wig/bigwig

TGAC
The Genome Analysis Centre™

Visualising expression data directly from wig/bigwig file

- Coloured and oriented peaks
  - Upwards red are positive
  - Downwards blue are negative



VCF and GFF

- Visualising variant data directly from vcf file
- Coloured based on base pairs
- Visualise insertion, deletion, mutation

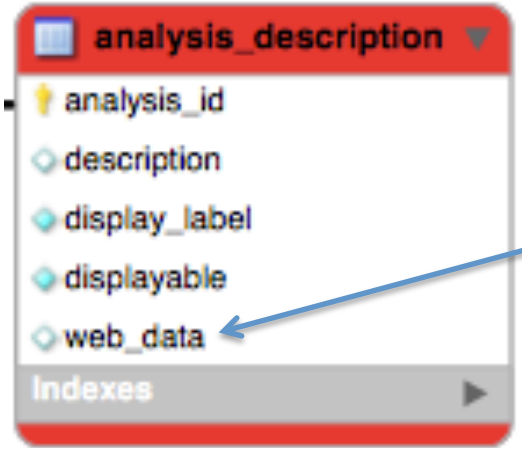
Visualising Genes data directly from gff file

- Exons, Introns, CDS

Non Ensembl Data

TGAC
The Genome Analysis Centre™

- Adding non Ensembl data in TGAC Browser
- *analysis_description* table of Ensembl Core schema
- *web_data* column for file information

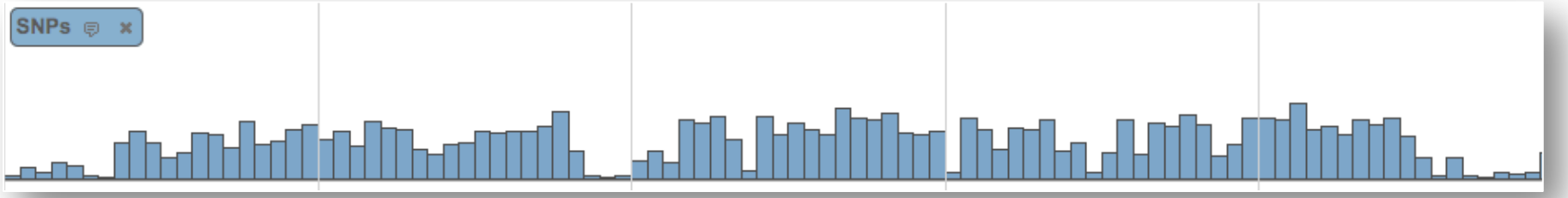



```
{
"colour": "blue",
"source": "file",
"filepath": "/storage/browser/test.bw",
”track-group": LIB1777,
}
```

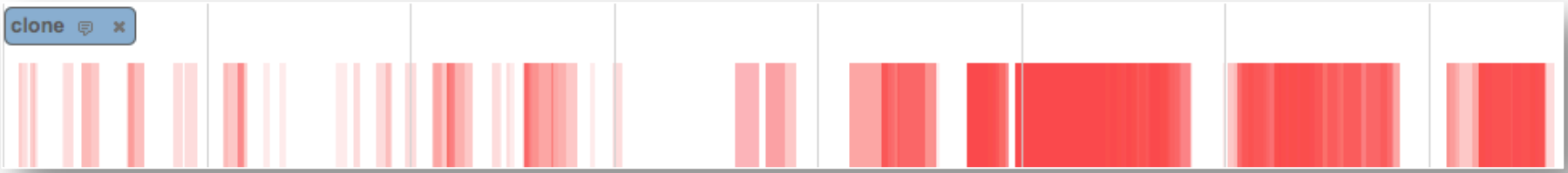
Visual Types

TGAC
The Genome Analysis Centre™

- Select Visual types based on amount of feature
  - Bar charts
  - Heat Map



- Graphical presentation of large data e.g. SNPs, alignments density

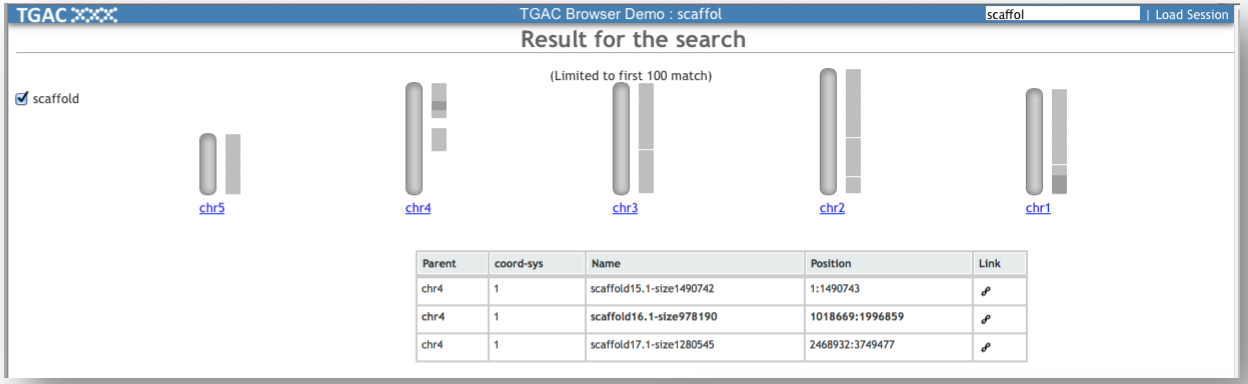


- Heat map presentation of large data

Search

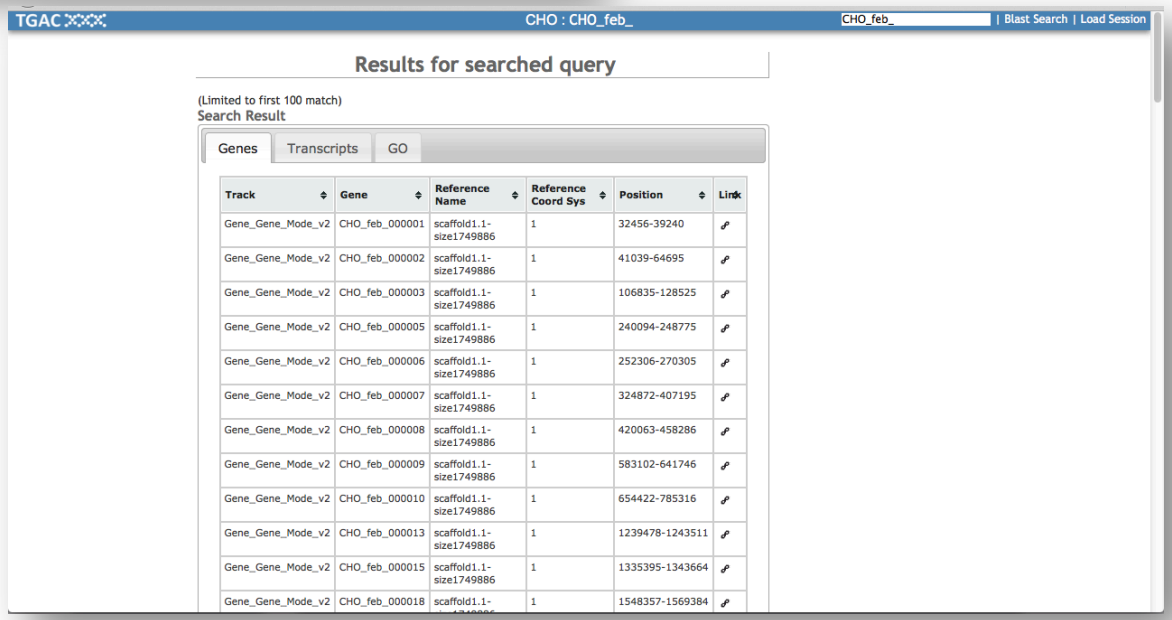
TGAC
The Genome Analysis Centre™

- Search with keyword among data



Result on Chromosomes

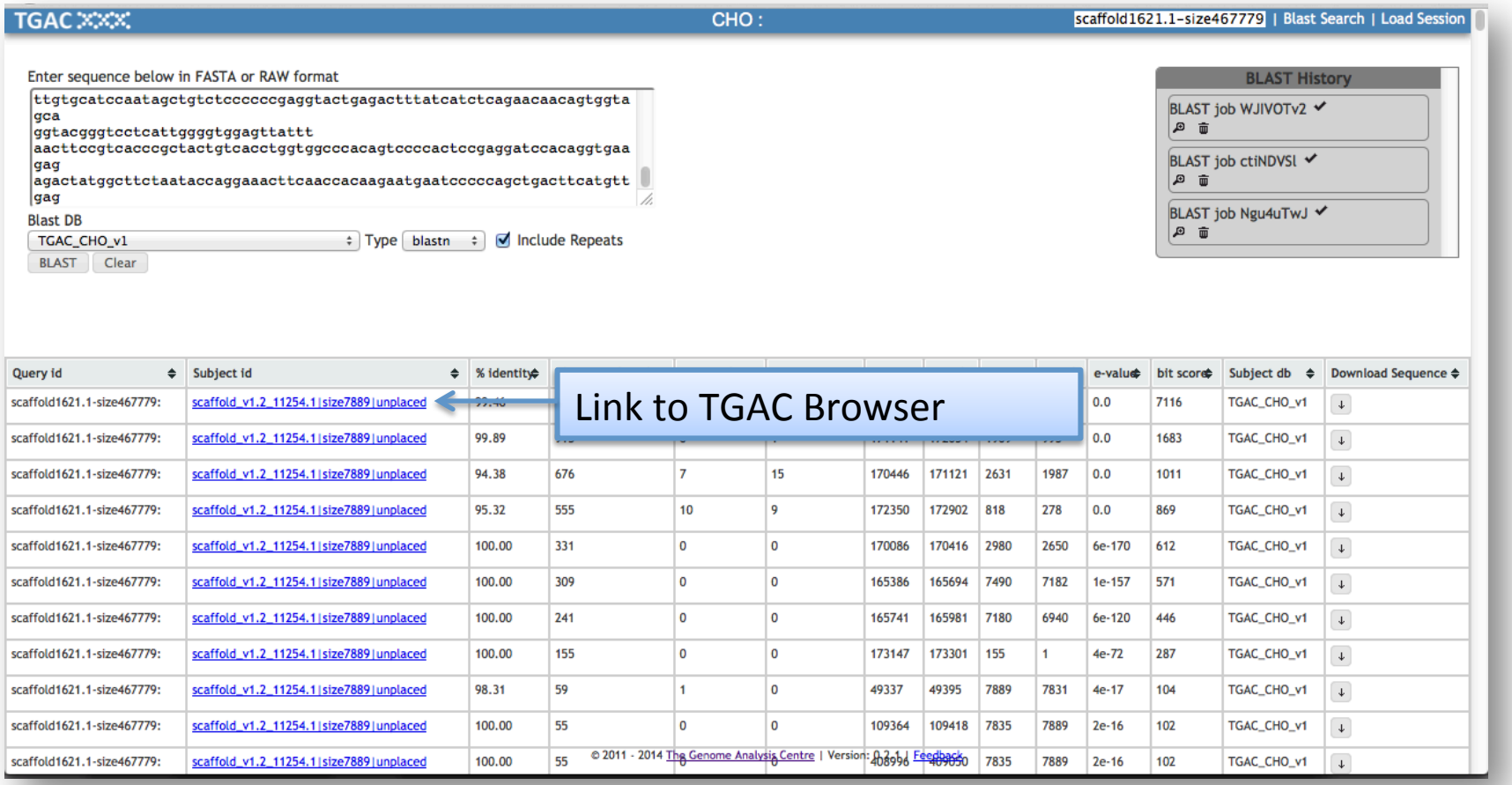
Result as list



BLAST Integration

TGAC
The Genome Analysis Centre™

- BLAST results for sequence search
- BLAST Manager
  - BLAST history logs
  - Run multiple BLAST simultaneously and toggle between result



Query id	Subject id	% identity								e-value	bit score	Subject db	Download Sequence
scaffold1621.1-size467779:	scaffold_v1.2_11254.1|size7889|unplaced									0.0	7116	TGAC_CHO_v1	
scaffold1621.1-size467779:	scaffold_v1.2_11254.1|size7889|unplaced	99.89								0.0	1683	TGAC_CHO_v1	
scaffold1621.1-size467779:	scaffold_v1.2_11254.1|size7889|unplaced	94.38	676	7	15	170446	171121	2631	1987	0.0	1011	TGAC_CHO_v1	
scaffold1621.1-size467779:	scaffold_v1.2_11254.1|size7889|unplaced	95.32	555	10	9	172350	172902	818	278	0.0	869	TGAC_CHO_v1	
scaffold1621.1-size467779:	scaffold_v1.2_11254.1|size7889|unplaced	100.00	331	0	0	170086	170416	2980	2650	6e-170	612	TGAC_CHO_v1	
scaffold1621.1-size467779:	scaffold_v1.2_11254.1|size7889|unplaced	100.00	309	0	0	165386	165694	7490	7182	1e-157	571	TGAC_CHO_v1	
scaffold1621.1-size467779:	scaffold_v1.2_11254.1|size7889|unplaced	100.00	241	0	0	165741	165981	7180	6940	6e-120	446	TGAC_CHO_v1	
scaffold1621.1-size467779:	scaffold_v1.2_11254.1|size7889|unplaced	100.00	155	0	0	173147	173301	155	1	4e-72	287	TGAC_CHO_v1	
scaffold1621.1-size467779:	scaffold_v1.2_11254.1|size7889|unplaced	98.31	59	1	0	49337	49395	7889	7831	4e-17	104	TGAC_CHO_v1	
scaffold1621.1-size467779:	scaffold_v1.2_11254.1|size7889|unplaced	100.00	55	0	0	109364	109418	7835	7889	2e-16	102	TGAC_CHO_v1	
scaffold1621.1-size467779:	scaffold_v1.2_11254.1|size7889|unplaced	100.00	55	0	0	408996	409050	7835	7889	2e-16	102	TGAC_CHO_v1	

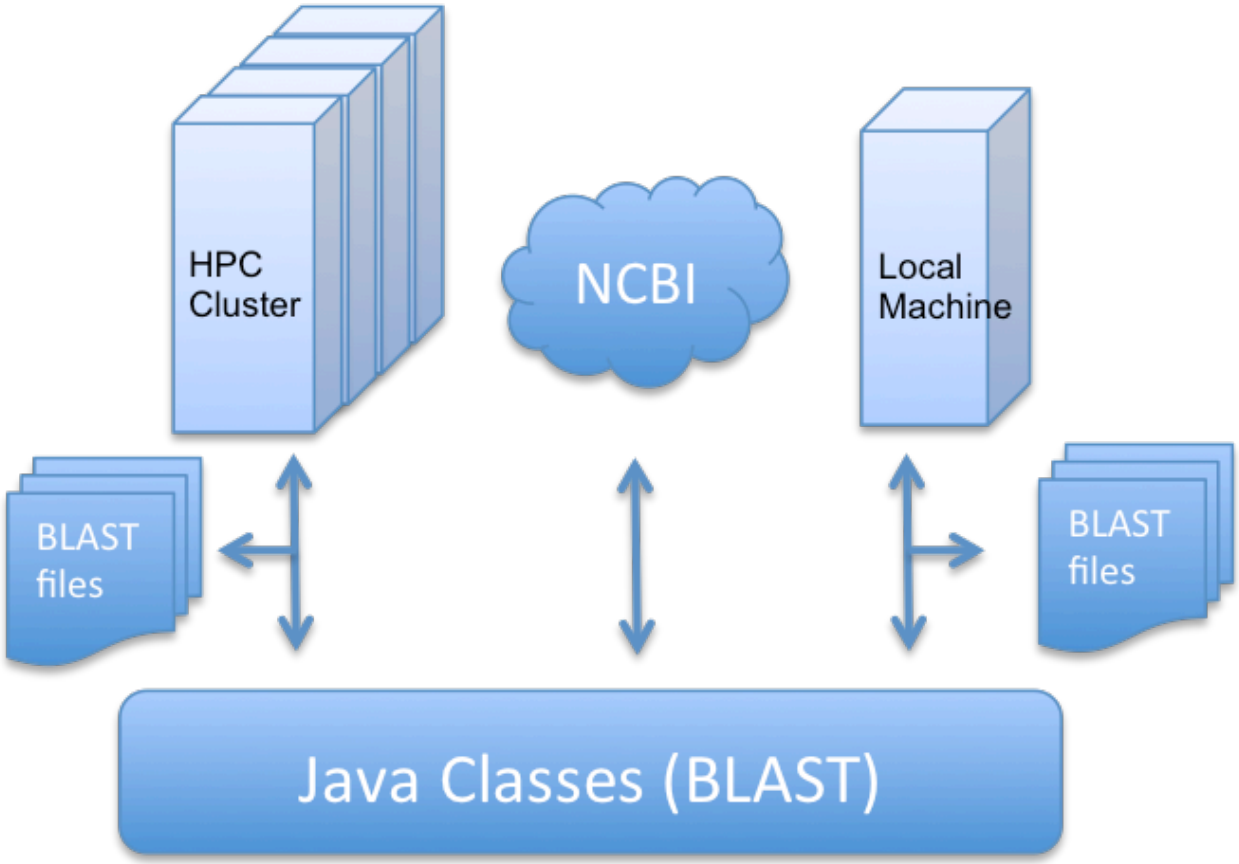
BLAST Integration

TGAC
The Genome Analysis Centre™

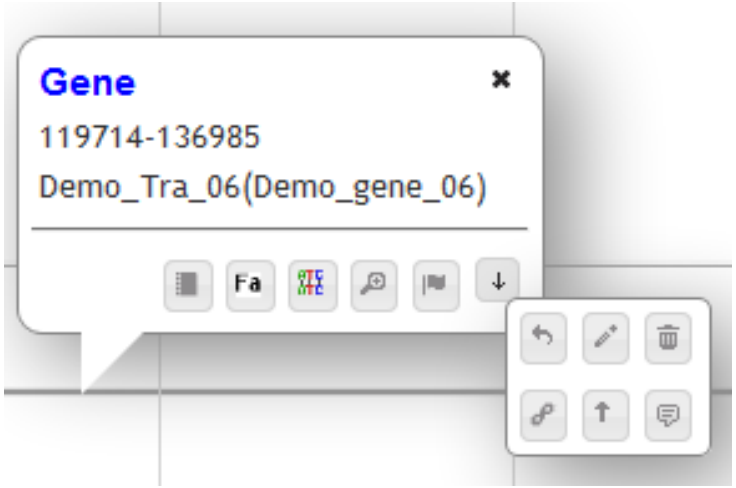
- BLAST run on a track or selected region
- BLAST results showing as a track run
- Coloured based on score and with indel information

BLAST Integration

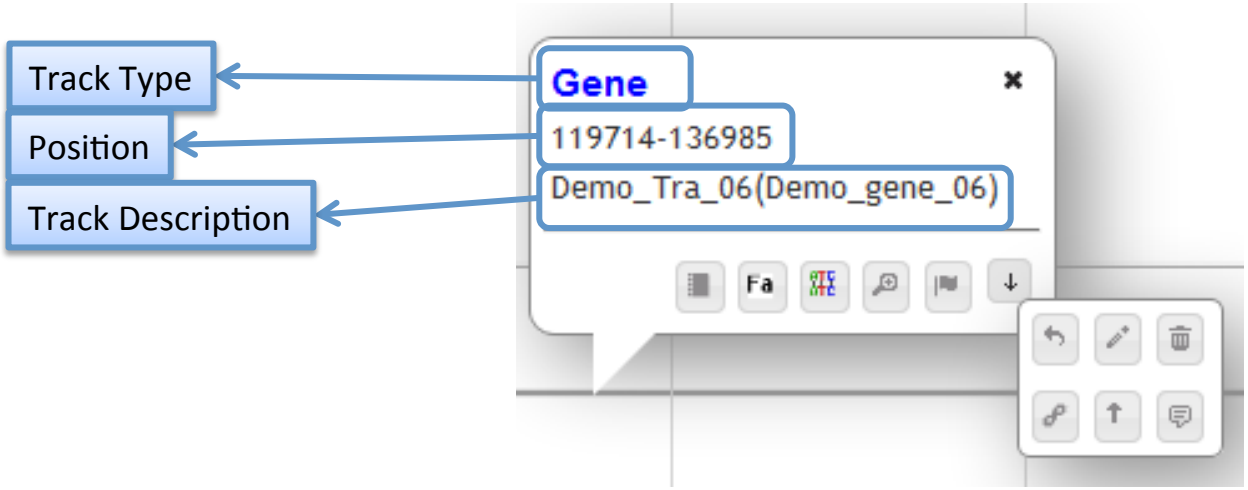
TGAC
The Genome Analysis Centre™



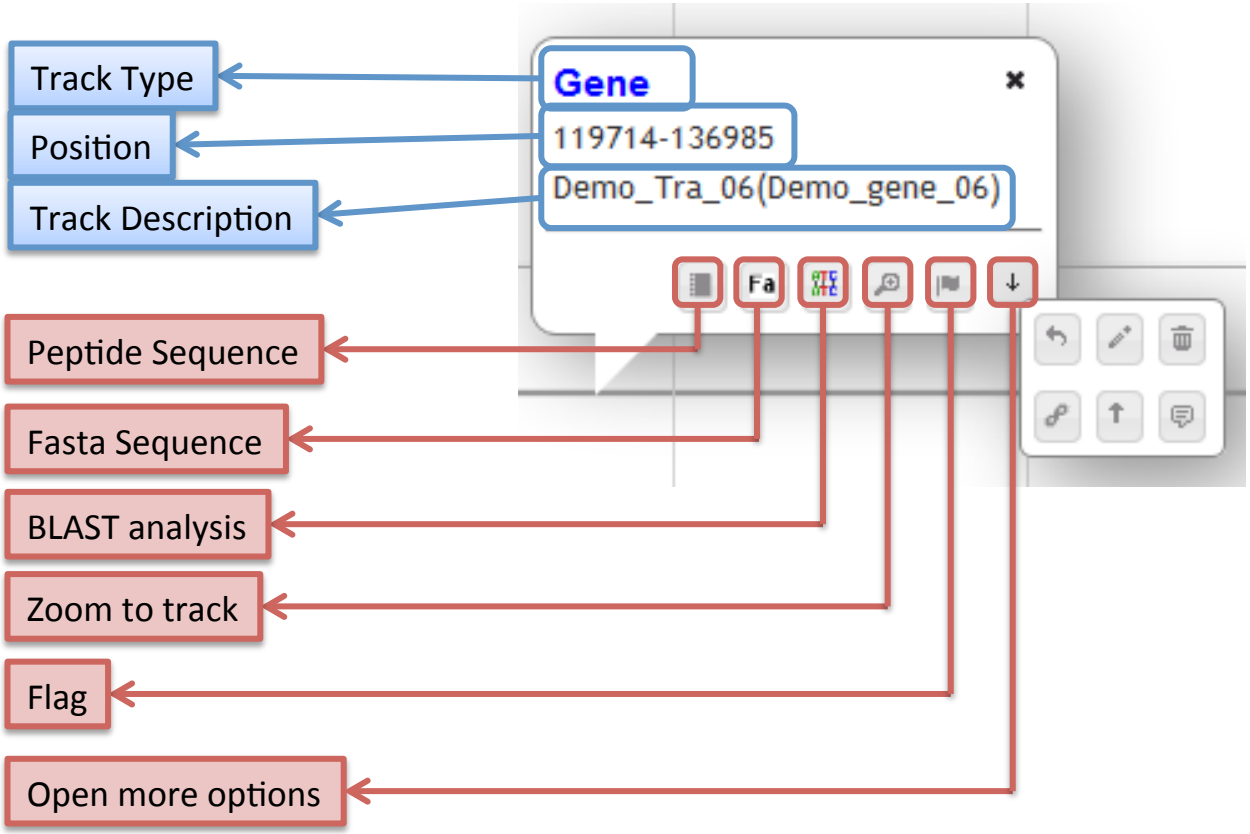
Popup



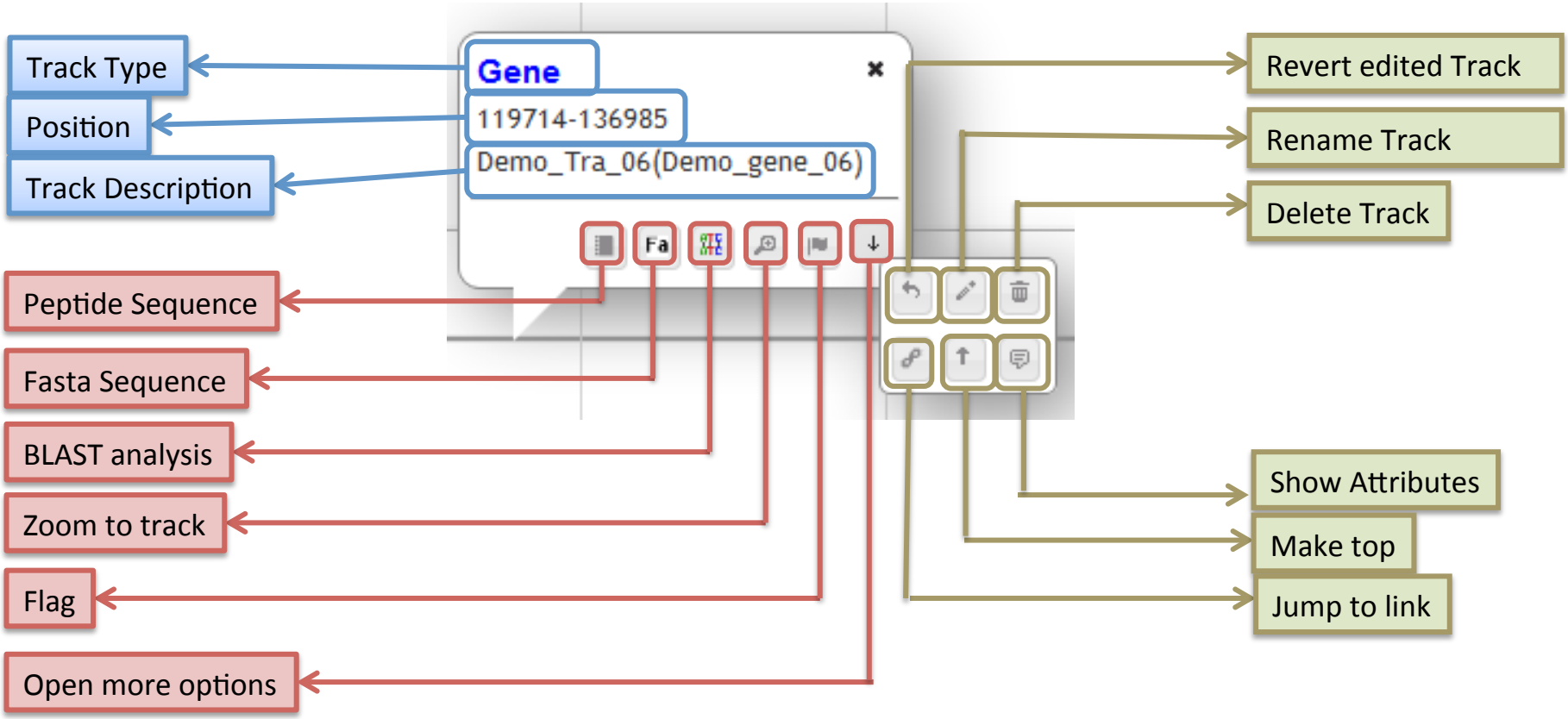
Popup


Track Type
Position
Track Description
Gene
119714-136985
Demo_Tra_06(Demo_gene_06)
Fa

Popup


Track Type
Position
Track Description
Gene
119714-136985
Demo_Tra_06(Demo_gene_06)
Fa
Peptide Sequence
Fasta Sequence
BLAST analysis
Zoom to track
Flag
Open more options

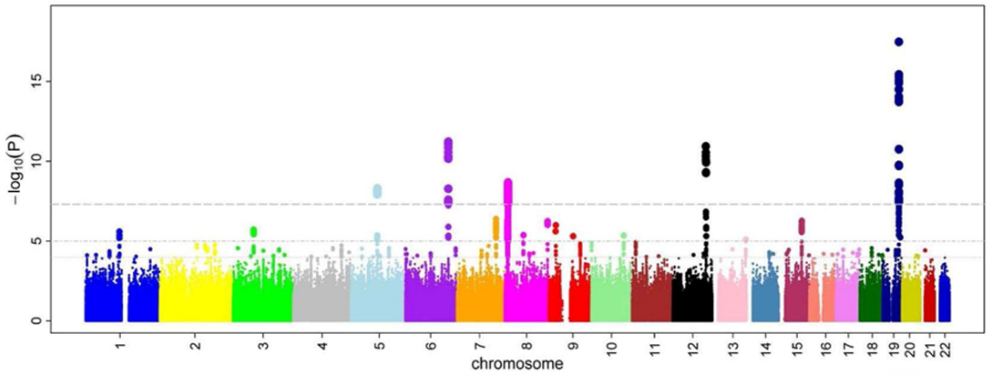
Popup

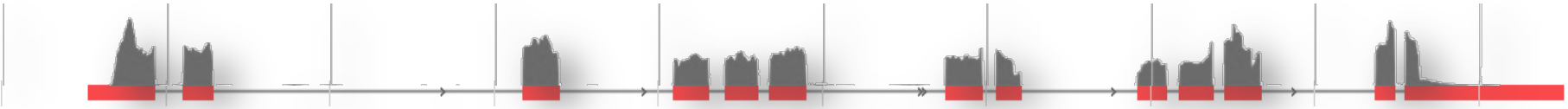

Track Type
Position
Track Description
Gene
119714-136985
Demo_Tra_06(Demo_gene_06)
Fa
Peptide Sequence
Fasta Sequence
BLAST analysis
Zoom to track
Flag
Open more options
Revert edited Track
Rename Track
Delete Track
Show Attributes
Make top
Jump to link

Future Work

TGAC
The Genome Analysis Centre™

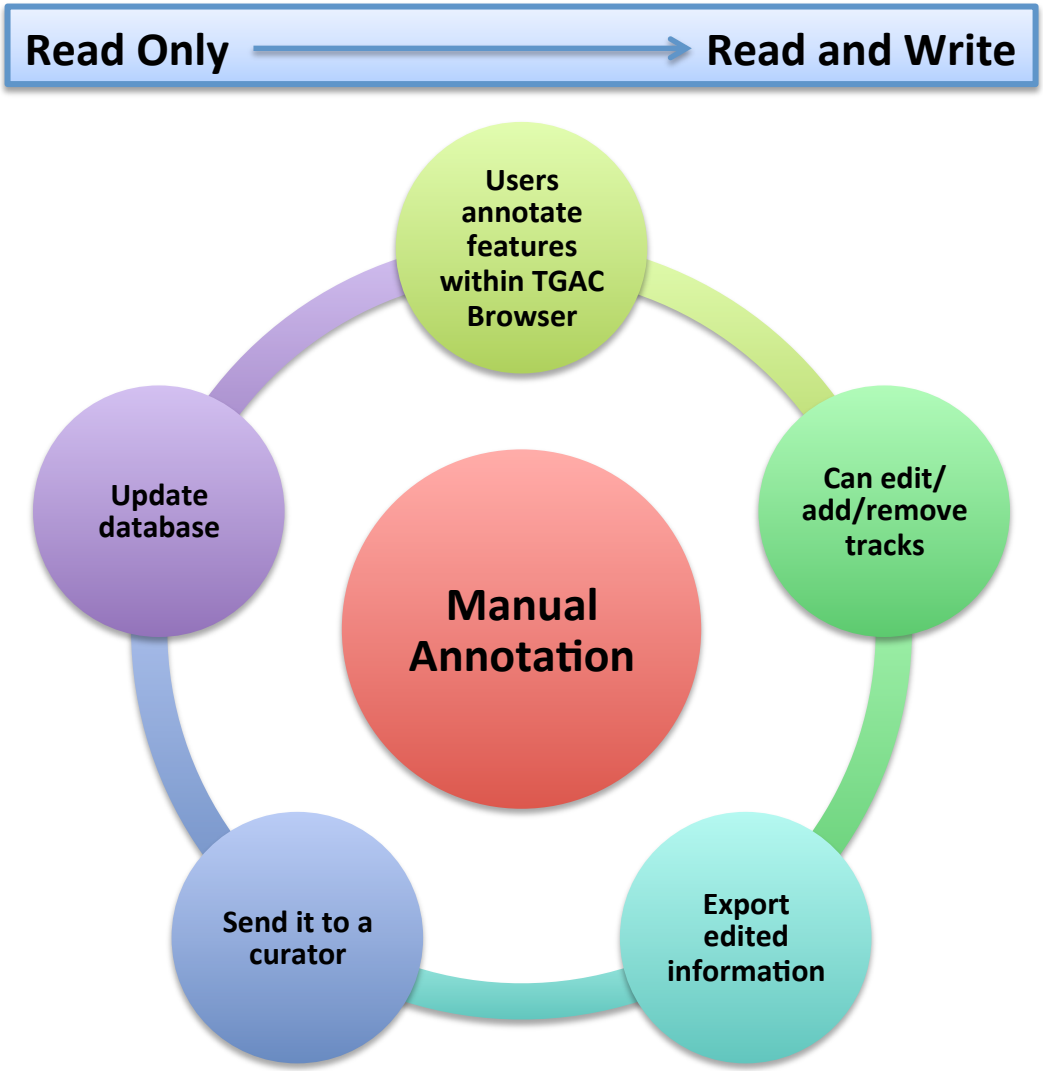
- **Manual Annotation**
- **Aequatus Browser integration**
- Upload user's data
- New Visuals
  - Manhattan Plots,
  - Expression data on Genes
- Integrate HMMER and BLAT analysis
- Data Download from TGAC Browser
- REST API to load data from Ensembl
- Write a Paper





Manual Annotation

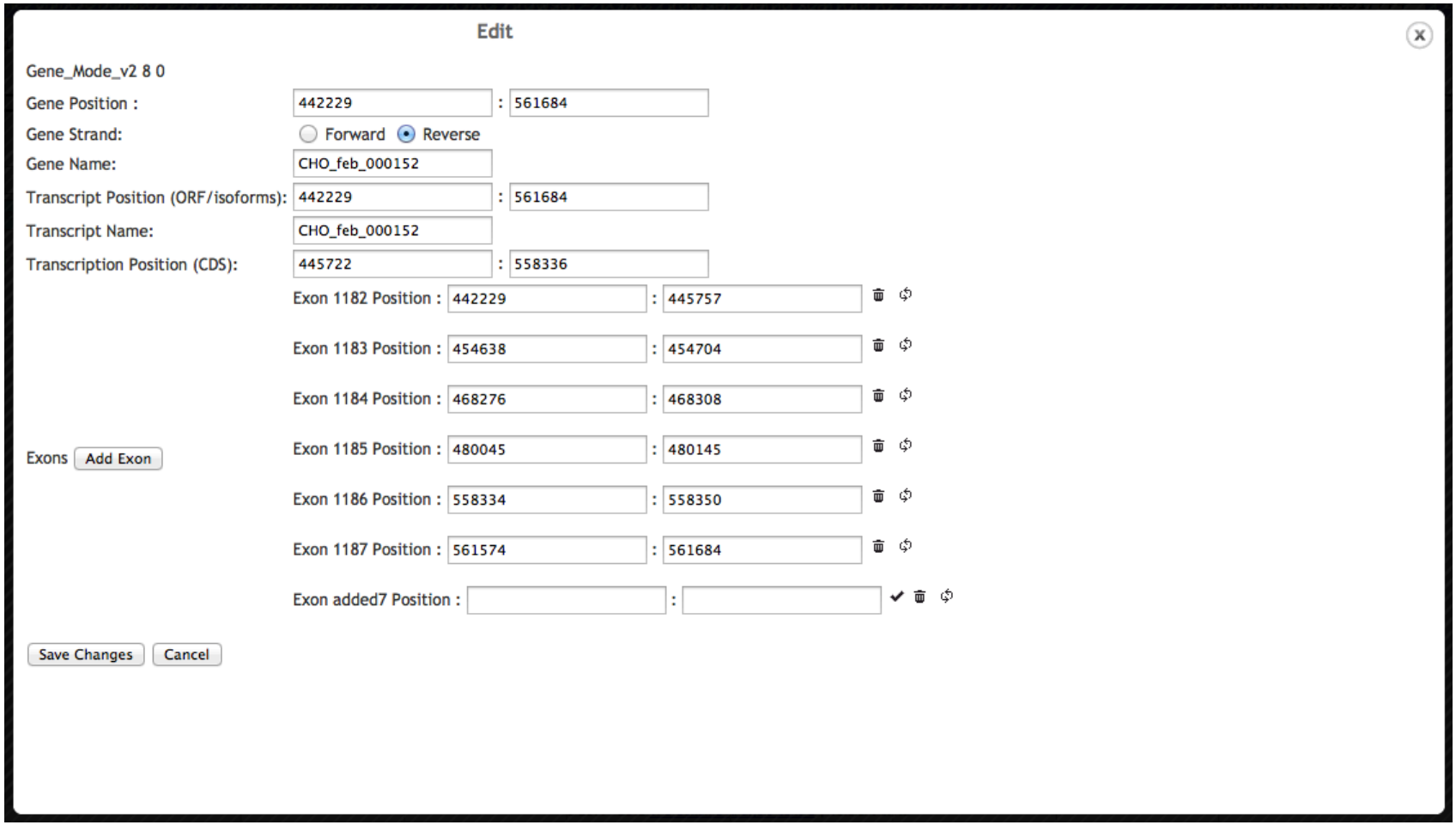
TGAC
The Genome Analysis Centre™



Manual Annotation

TGAC
The Genome Analysis Centre™

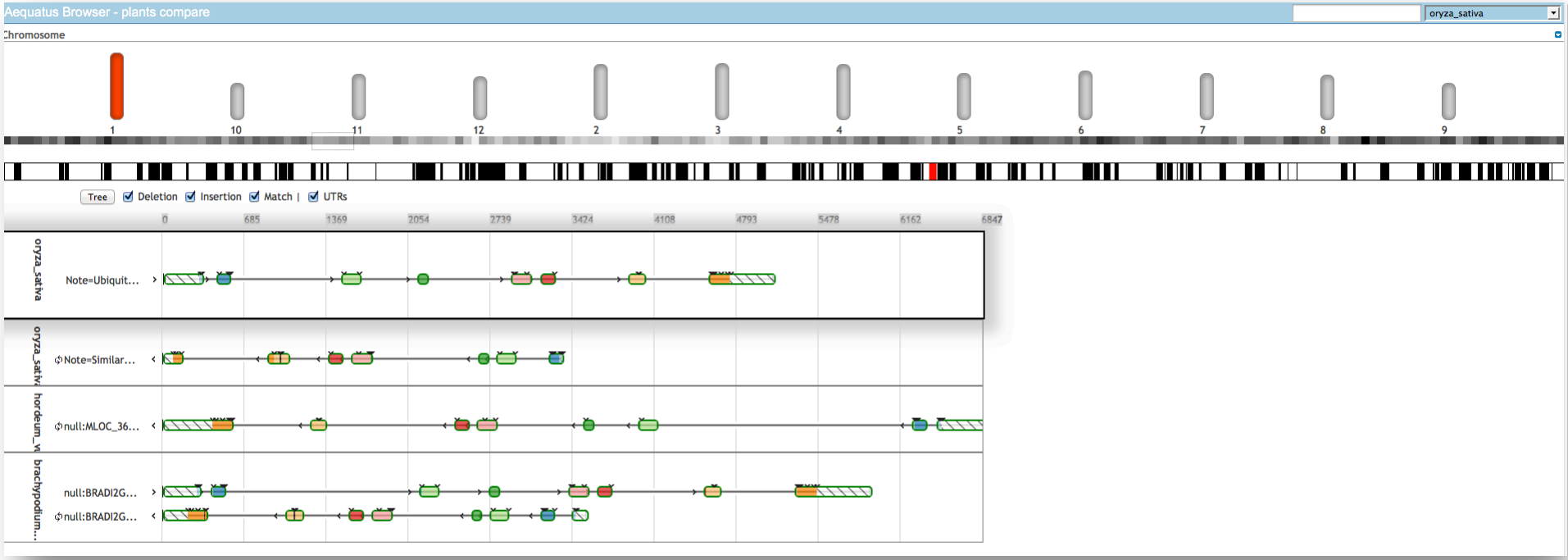
- Edit various genomic track credentials
- Add new tracks



Aequatus Browser

TGAC
The Genome Analysis Centre™

- Integration of Aequatus Browser within TGAC Browser
  - Under development at TGAC



- Visualising complex similarity relationships among species
- Beta: http://tgac-browser.tgac.ac.uk/plants_compara
- Slides: **http://tinyurl.com/aequatous-browser**

TGAC Browser

TGAC
The Genome Analysis Centre™

Started	• Oct 2011
Presented at NGS DeepSeq	• August 2012
V0.1.1	• May 2013
Presented at ISMB	• July 2013
v0.1.2	• Aug 2013
Presented at UK Genome Science	• Sept 2013
V0.2.0	• Dec 2013
Presented at VizBi	• March 2014
v0.2.1	• May 2014
Presenting at BOSC 2014	• July 2014
V0.2.2	• Soon…

Instances:

- TGAC Browser Demo
- Chinese Hamster
- Wheat Yellow Rust
- Chalara Fraxinus
- Brassica
- Homo Sapiens
- Vietnamese Rice
- Lactobacillus salivarius
- IWGSC Wheat
- Hordeum
- Oryza Sativa
- Brachypodium

Documentation: **https://documentation.tgac.ac.uk/display/TB/TGAC+Browser**

Acknowledgements

Robert Davey

Xingdong Bian

Sarah Ayling

Mario Caccamo

Jon Wright

Gemy Kaithakottil

Daniel Mapleson

Martin Ayling

Acknowledgements

John Innes Centre — Jinhong Li

Stazione Zoologica Anton Dohrn Napoli — Mariella Ferrante, Remo Sanges

The Sainsbury Laboratory TSL — Burkhard Steuernagel

Trinity College Dublin — Paul Linehan

International Wheat Genome Sequencing Consortium

http://tgac-browser.tgac.ac.uk
tgac.browser@tgac.ac.uk
@anilthanki and @tgacbrowser

BBSRC bioscience for the future

transPLANT