

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



- Anil S. Thanki  
Scientific Programmer – Sequencing Informatics  
[Anil.Thanki@tgac.ac.uk](mailto:Anil.Thanki@tgac.ac.uk)

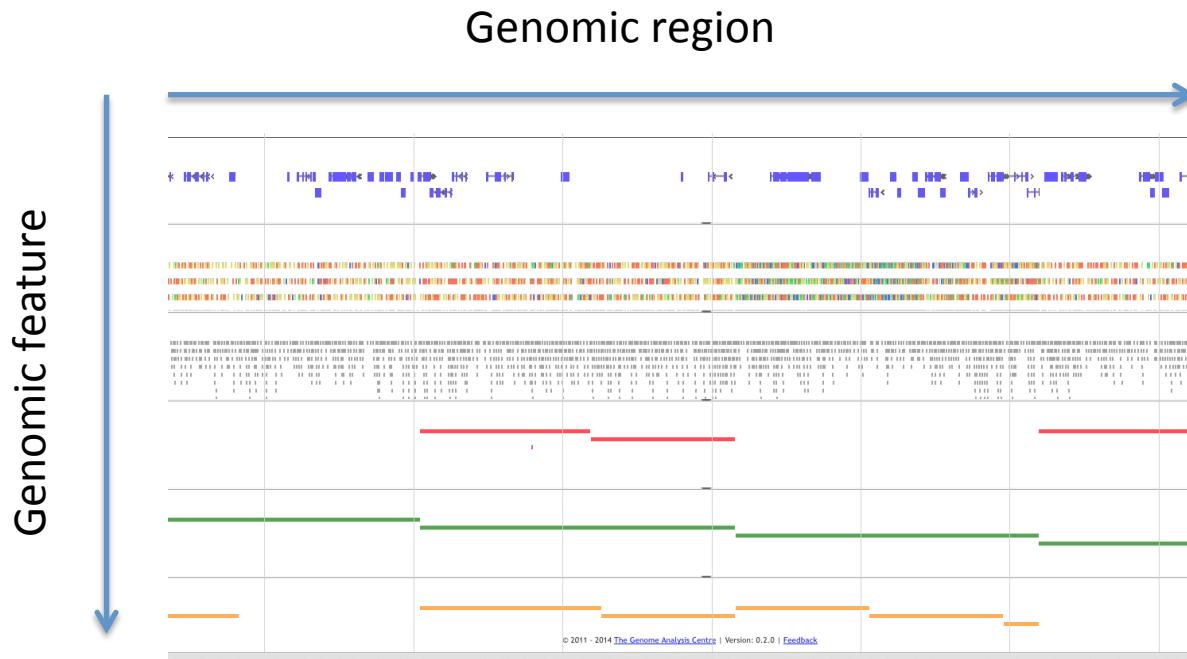
@anilthanki and @tgacbrower

July 11, 2014  
BOSC



#Poster 6

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



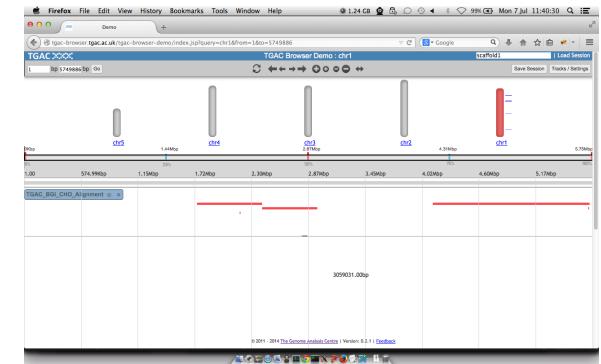
# TGAC Browser



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

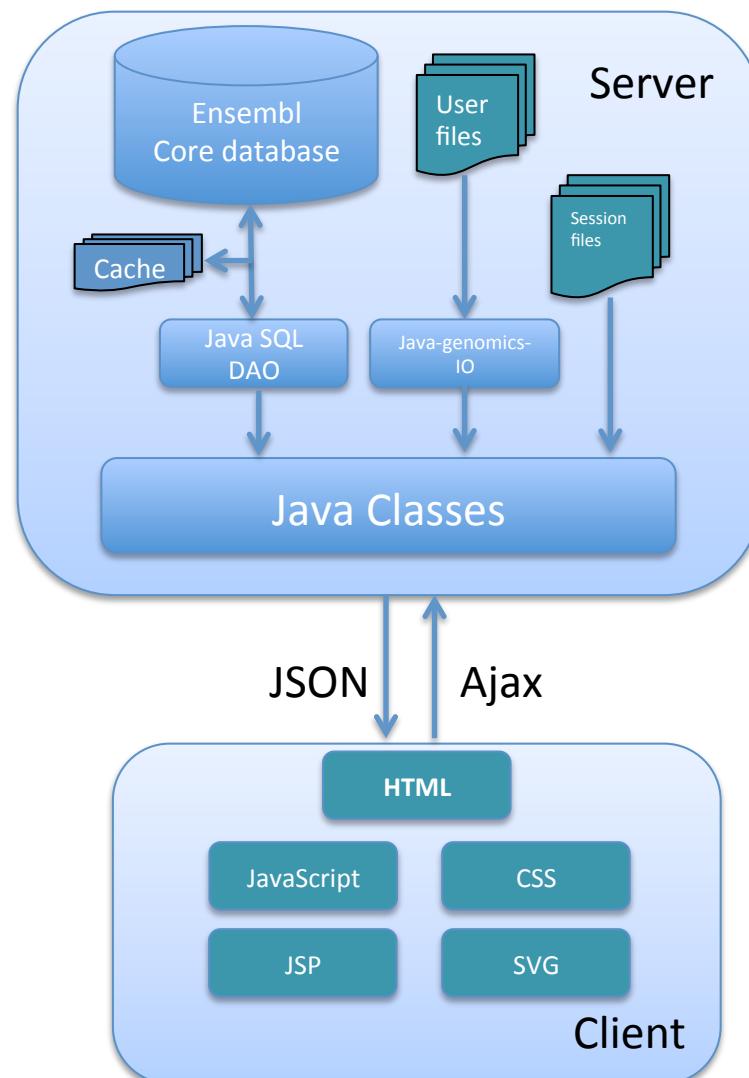
# TGAC Browser server-client

- 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



## Genomics Data



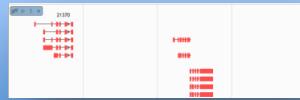
Ensembl core



SAM/BAM



Wig/BigWig



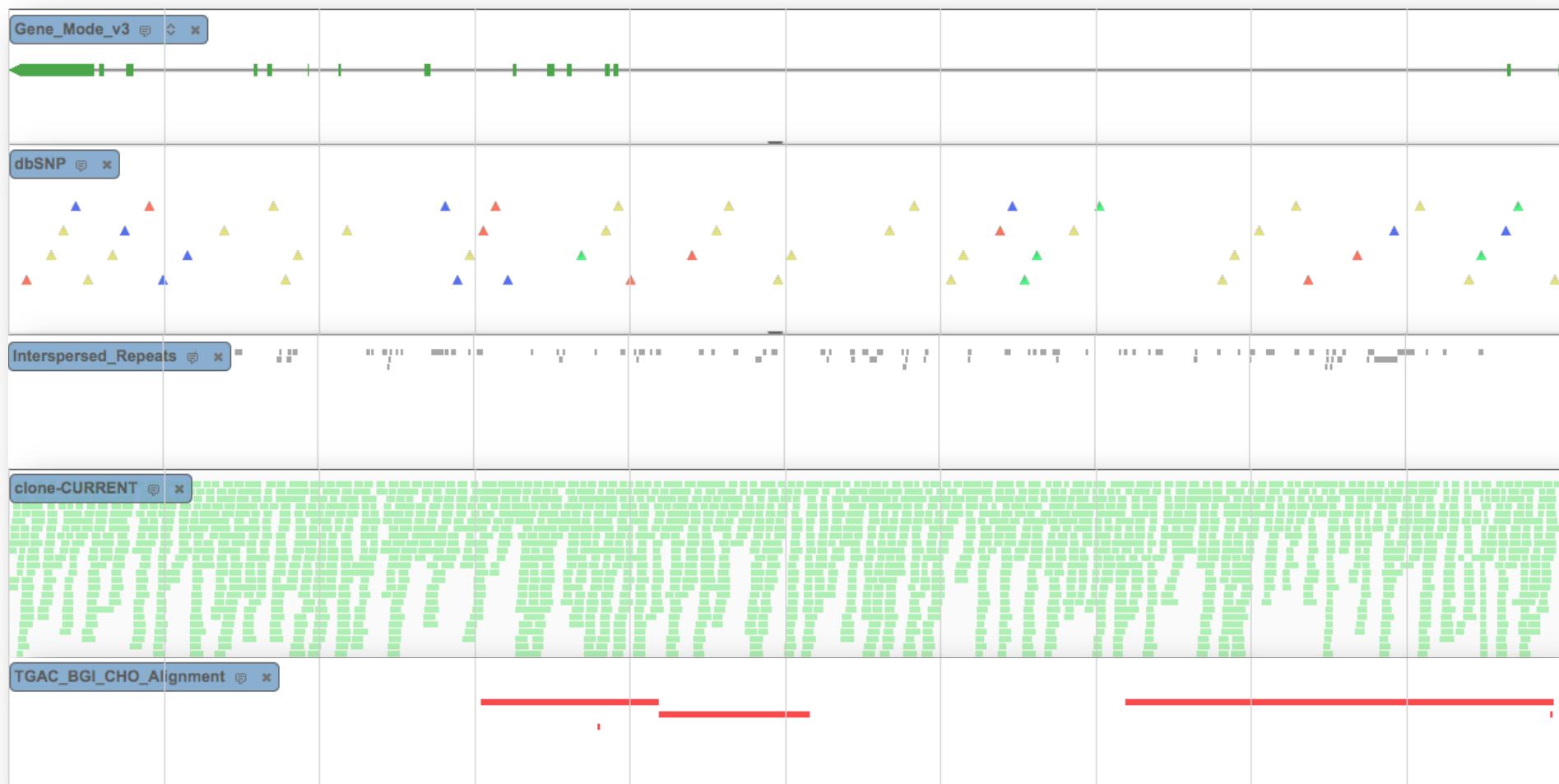
GFF



VCF

# Ensembl Data

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



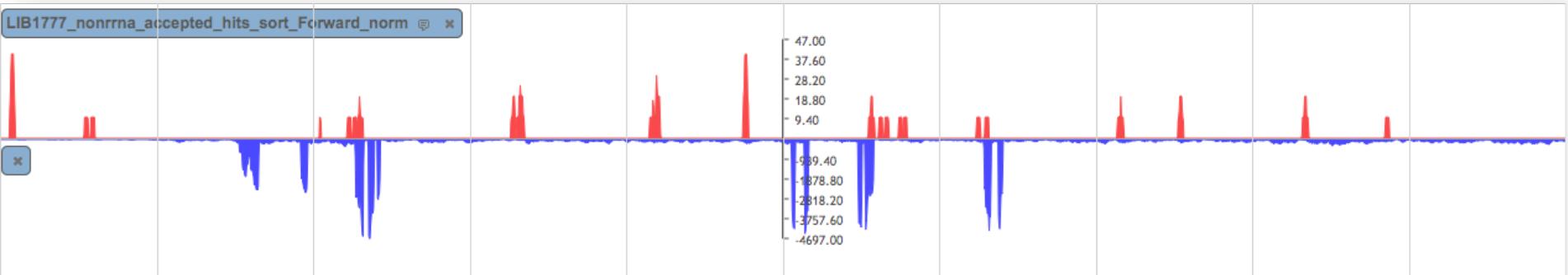
## Visualising reads directly from SAM/BAM file

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



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



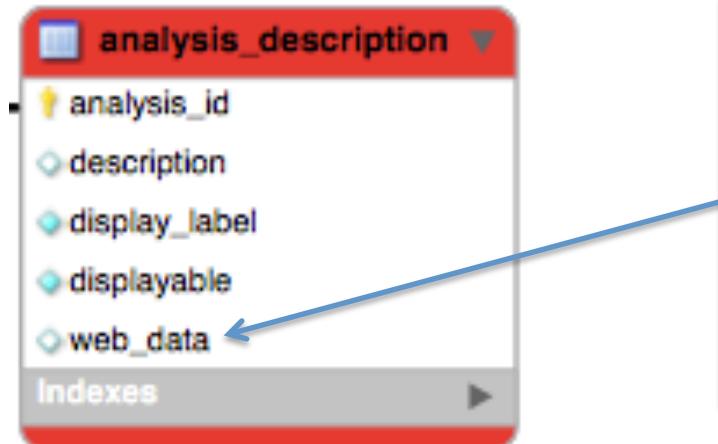
Visualising Genes data directly from gff file

- Exons, Introns, CDS



# Non Ensembl Data

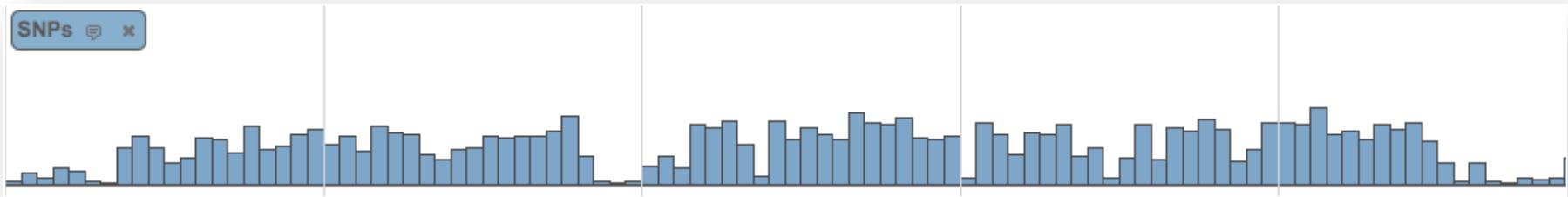
- Adding non Ensembl data in TGAC Browser
- *analysis\_description* table of Ensembl Core schema
- *web\_data* column for file information



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

# Visual Types

- 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

- Search with keyword among data

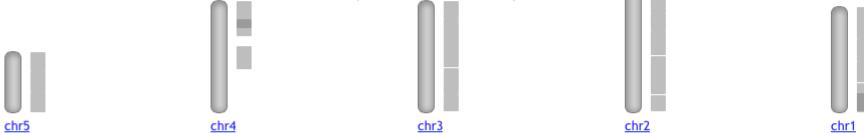
TGAC XXXX

TGAC Browser Demo : scaffol

Result for the search

(Limited to first 100 match)

scaffold



Parent	coord-sys	Name	Position	Link
chr4	1	scaffold15,1-size1490742	1:1490743	<a href="#">♂</a>
chr4	1	scaffold16,1-size978190	1018669:1996859	<a href="#">♂</a>
chr4	1	scaffold17,1-size1280545	2468932:3749477	<a href="#">♂</a>

Result on Chromosomes

TGAC XXXX

CHO : CHO\_feb\_

CHO\_feb\_ | Blast Search | Load Session

Results for searched query

(Limited to first 100 match)

Search Result

Genes Transcripts GO

Track	Gene	Reference Name	Reference Coord Sys	Position	Link
Gene_Gene_Mode_v2	CHO_feb_000001	scaffold1,1-size1749886	1	32456-39240	<a href="#">♂</a>
Gene_Gene_Mode_v2	CHO_feb_000002	scaffold1,1-size1749886	1	41039-64695	<a href="#">♂</a>
Gene_Gene_Mode_v2	CHO_feb_000003	scaffold1,1-size1749886	1	106835-128525	<a href="#">♂</a>
Gene_Gene_Mode_v2	CHO_feb_000005	scaffold1,1-size1749886	1	240094-248775	<a href="#">♂</a>
Gene_Gene_Mode_v2	CHO_feb_000006	scaffold1,1-size1749886	1	252306-270305	<a href="#">♂</a>
Gene_Gene_Mode_v2	CHO_feb_000007	scaffold1,1-size1749886	1	324872-407195	<a href="#">♂</a>
Gene_Gene_Mode_v2	CHO_feb_000008	scaffold1,1-size1749886	1	420063-458286	<a href="#">♂</a>
Gene_Gene_Mode_v2	CHO_feb_000009	scaffold1,1-size1749886	1	583102-641746	<a href="#">♂</a>
Gene_Gene_Mode_v2	CHO_feb_000010	scaffold1,1-size1749886	1	654422-785316	<a href="#">♂</a>
Gene_Gene_Mode_v2	CHO_feb_000013	scaffold1,1-size1749886	1	1239478-1243511	<a href="#">♂</a>
Gene_Gene_Mode_v2	CHO_feb_000015	scaffold1,1-size1749886	1	1335395-1343664	<a href="#">♂</a>
Gene_Gene_Mode_v2	CHO_feb_000018	scaffold1,1-size1749886	1	1548357-1569384	<a href="#">♂</a>

Result as list

# BLAST Integration

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

TGAC  CHO : scaffold1621.1-size467779 | Blast Search | Load Session

Enter sequence below in FASTA or RAW format

```
ttgtgcataatagctgtccccccgggtactgagactttatcatctcagaacaacagtggta  
gca  
ggtaggggttcattgggtggatatt  
aactccgtcaccgtactgtcacctggggccacagtcccaactccgaggatccacaggatggaa  
gag  
agatatatggcttaataccaggaaacttcaaccacaagaatgaatccccagctgacttcatgtt  
gag
```

Blast DB: TGAC\_CHO\_v1 Type: blastn  Include Repeats

BLAST Clear

BLAST History

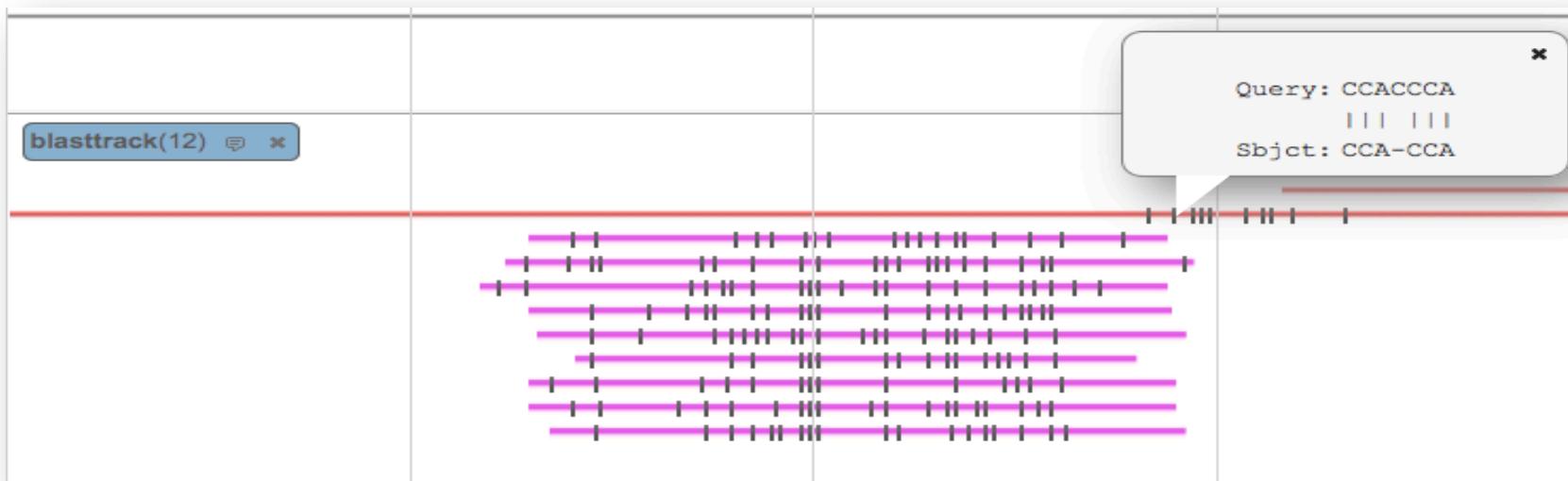
- BLAST job WJIVOTv2 ✓
- BLAST job ctiNDVSL ✓
- BLAST job Ngu4uTwJ ✓

Query id	Subject id	% identity	Link to TGAC Browser										e-value	bit score	Subject db	Download Sequence
scaffold1621.1-size467779:	<a href="#">scaffold_v1.2_11254.1 size7889 unplaced</a>	99.40											0.0	7116	TGAC_CHO_v1	
scaffold1621.1-size467779:	<a href="#">scaffold_v1.2_11254.1 size7889 unplaced</a>	99.89											0.0	1683	TGAC_CHO_v1	
scaffold1621.1-size467779:	<a href="#">scaffold_v1.2_11254.1 size7889 unplaced</a>	94.38	676	7	15	170446	171121	2631	1987				0.0	1011	TGAC_CHO_v1	
scaffold1621.1-size467779:	<a href="#">scaffold_v1.2_11254.1 size7889 unplaced</a>	95.32	555	10	9	172350	172902	818	278				0.0	869	TGAC_CHO_v1	
scaffold1621.1-size467779:	<a href="#">scaffold_v1.2_11254.1 size7889 unplaced</a>	100.00	331	0	0	170086	170416	2980	2650	6e-170	612		0.0	1683	TGAC_CHO_v1	
scaffold1621.1-size467779:	<a href="#">scaffold_v1.2_11254.1 size7889 unplaced</a>	100.00	309	0	0	165386	165694	7490	7182	1e-157	571		0.0	1011	TGAC_CHO_v1	
scaffold1621.1-size467779:	<a href="#">scaffold_v1.2_11254.1 size7889 unplaced</a>	100.00	241	0	0	165741	165981	7180	6940	6e-120	446		0.0	7116	TGAC_CHO_v1	
scaffold1621.1-size467779:	<a href="#">scaffold_v1.2_11254.1 size7889 unplaced</a>	100.00	155	0	0	173147	173301	155	1	4e-72	287		0.0	1683	TGAC_CHO_v1	
scaffold1621.1-size467779:	<a href="#">scaffold_v1.2_11254.1 size7889 unplaced</a>	98.31	59	1	0	49337	49395	7889	7831	4e-17	104		0.0	1011	TGAC_CHO_v1	
scaffold1621.1-size467779:	<a href="#">scaffold_v1.2_11254.1 size7889 unplaced</a>	100.00	55	0	0	109364	109418	7835	7889	2e-16	102		0.0	7116	TGAC_CHO_v1	
scaffold1621.1-size467779:	<a href="#">scaffold_v1.2_11254.1 size7889 unplaced</a>	100.00	55	0	0	408996	409050	7835	7889	2e-16	102		0.0	1683	TGAC_CHO_v1	

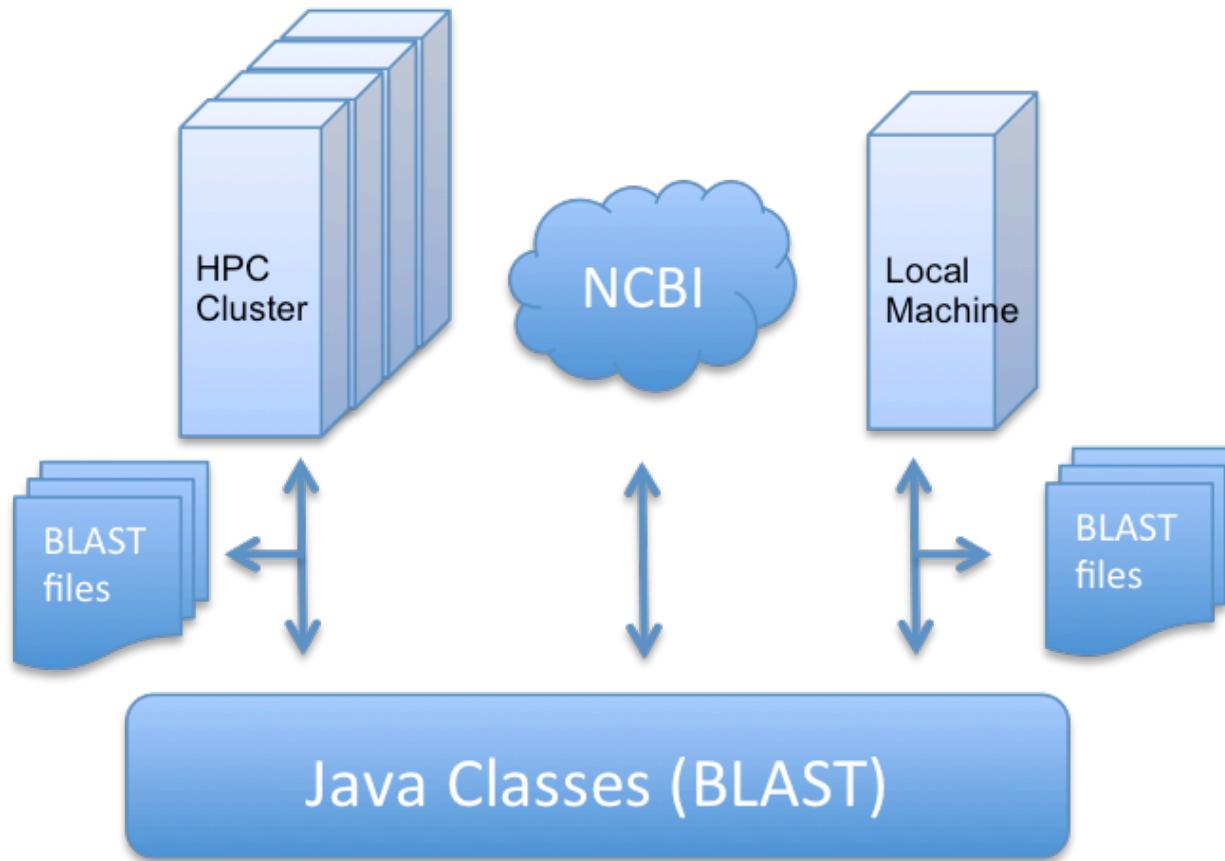
© 2011 - 2014 The Genome Analysis Centre | Version: 0.2.1 | Feedback

# BLAST Integration

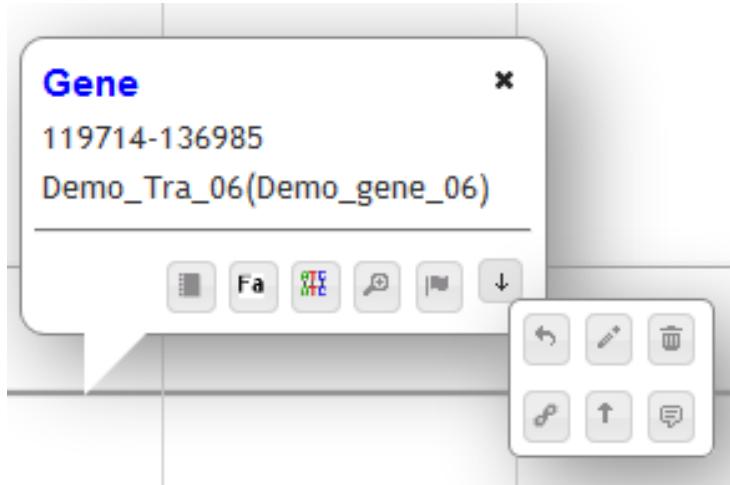
- BLAST run on a track or selected region
- BLAST results showing as a track run
- Coloured based on score and with indel information



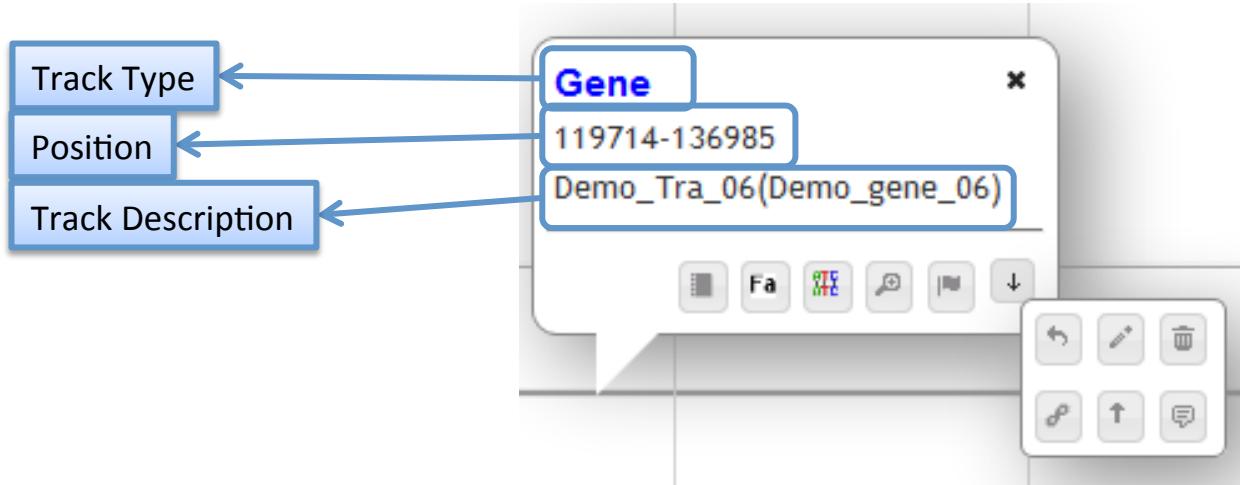
# BLAST Integration



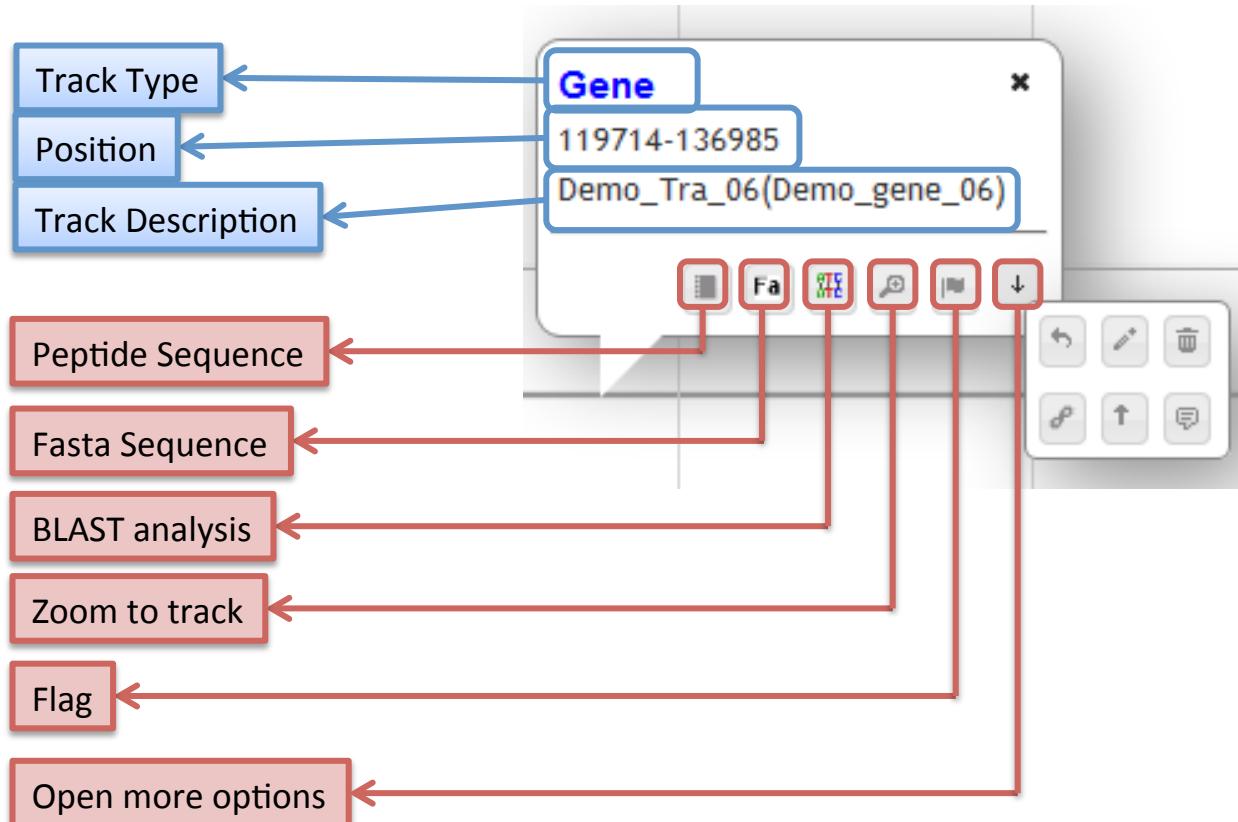
# Popup



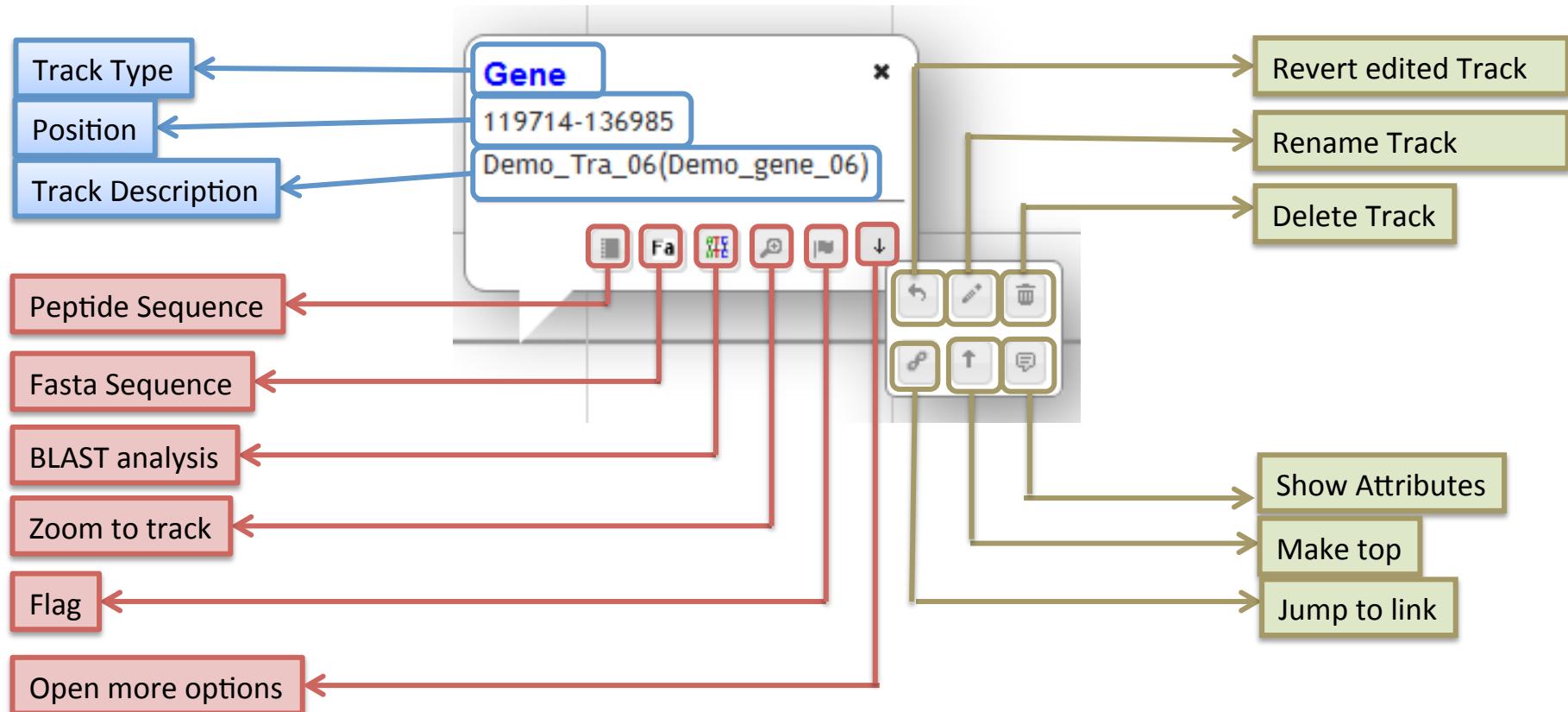
# Popup



# Popup

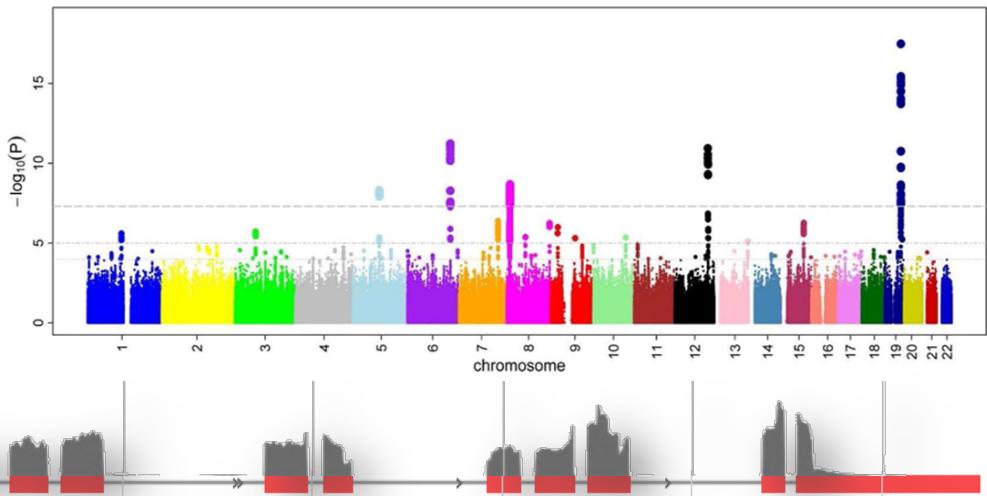


# Popup



# Future Work

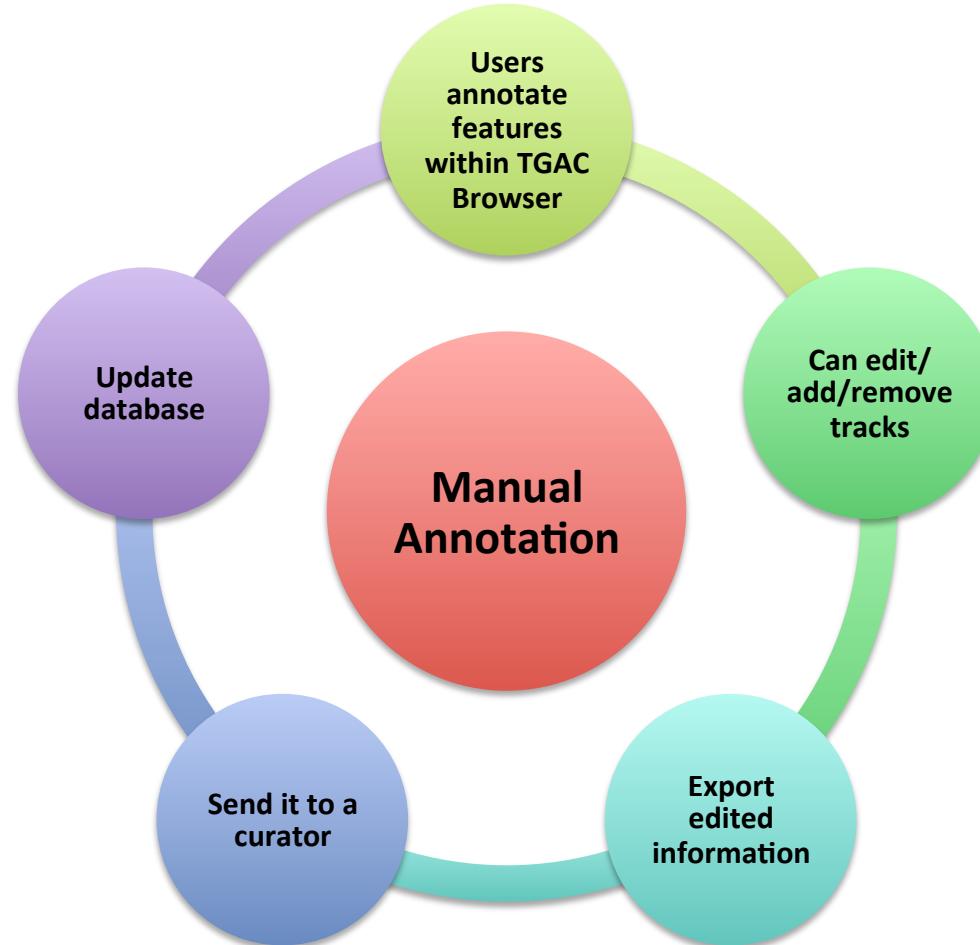
- **Manual Annotation**
- **Aequatus Browser integration**
- Upload user's data
- New Visuals
  - Manhattan Plots,
  - Expression data on Genes
- Integrate HMMER and BLAT analysis
- Data Download from TGAC Browser
- REST API to load data from Ensembl
- Write a Paper



# Manual Annotation

Read Only

Read and Write



# Manual Annotation

- Edit various genomic track credentials
- Add new tracks

Edit

Gene\_Mode\_v2.8.0

Gene Position :  :

Gene Strand:  Forward  Reverse

Gene Name:

Transcript Position (ORF/isoforms):  :

Transcript Name:

Transcription Position (CDS):  :

Exon 1182 Position :  :  ✖ ⚡

Exon 1183 Position :  :  ✖ ⚡

Exon 1184 Position :  :  ✖ ⚡

Exon 1185 Position :  :  ✖ ⚡

Exon 1186 Position :  :  ✖ ⚡

Exon 1187 Position :  :  ✖ ⚡

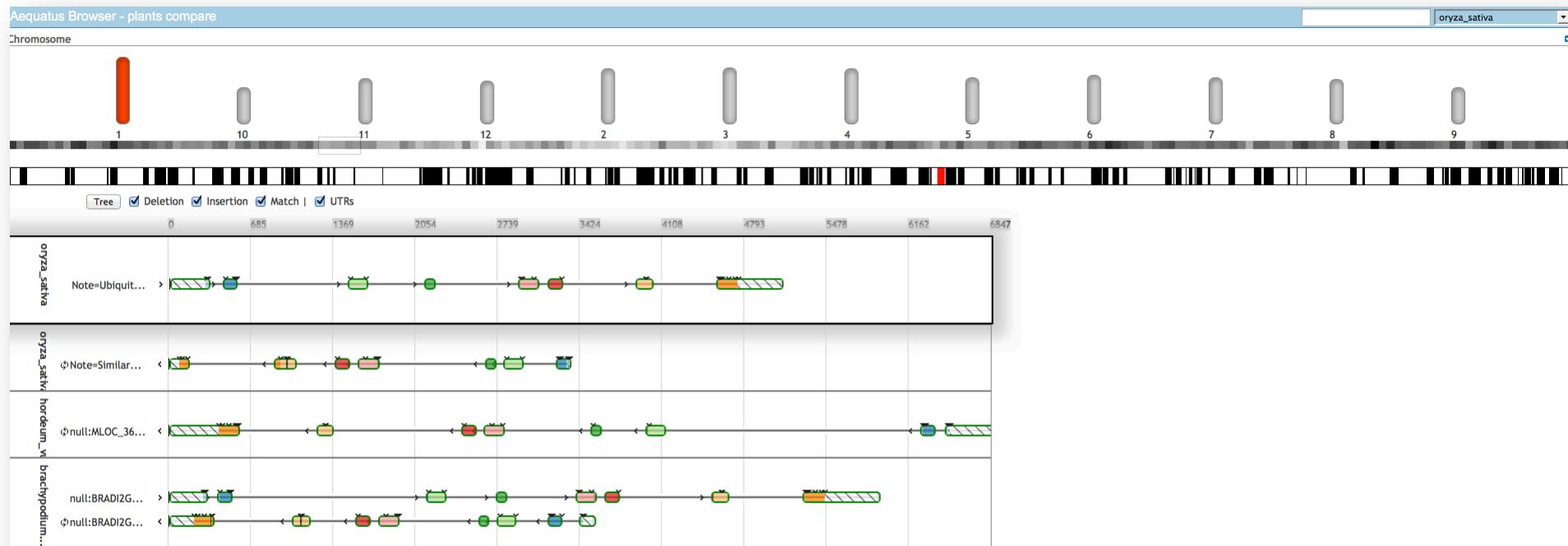
Exon added7 Position :  :  ✓ ✖ ⚡

Exons Add Exon

Save Changes Cancel

# Aequatus Browser

- 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](http://tgac-browser.tgac.ac.uk/plants_compara)
- Slides: <http://tinyurl.com/aequatous-browser>

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



Jinhong Li



Mariella Ferrante  
Remo Sanges

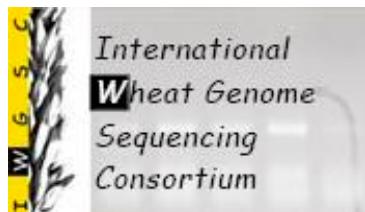
The Sainsbury Laboratory



Burkhard Steuernagel



Paul Linehan



<http://tgac-browser.tgac.ac.uk>



[tgac.browser@tgac.ac.uk](mailto:tgac.browser@tgac.ac.uk)



@anilthanki and @tgacbrower

