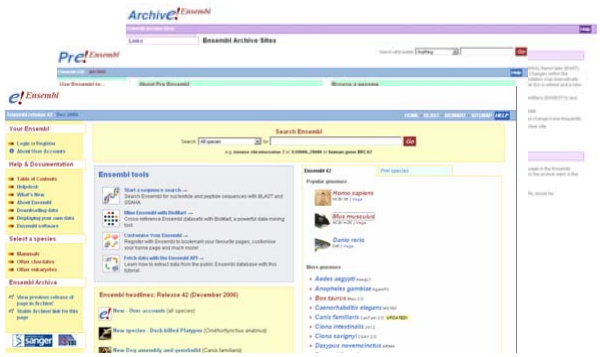
# DAS
## Integration of external annotation data



**Xosé Mª Fernández**
European Bioinformatics Institute

Feb 2007

## Outline of talk

- **DAS**
  - What is DAS?
  - DAS client DAS server
- DAS in Ensembl
  - Ensembl as DAS server
  - DAS essentials
  - Examples

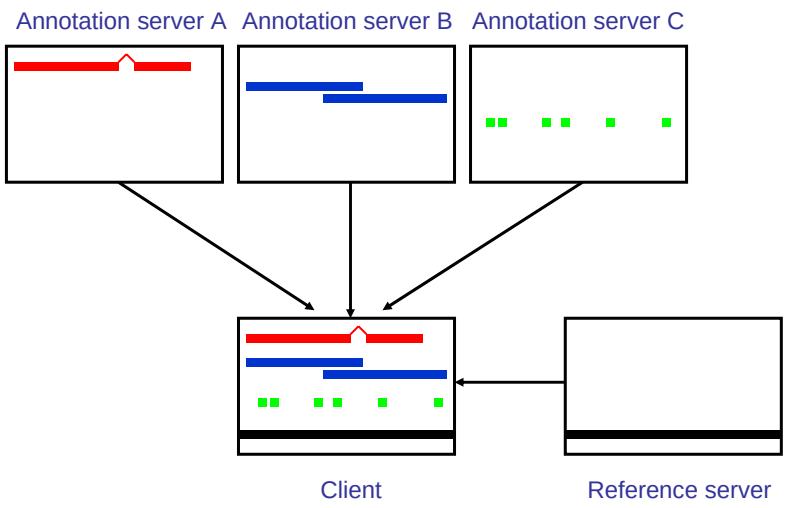
# What is DAS?

**DAS** **D**istributed **A**nnotation **S**ystem

- Developed by Robin Dowell, Sean Eddy and Lincoln Stein (Dowell *et al*., 2001)
- Specification of a protocol for requesting and returning sequence and annotation data
- Allows decentralised storage of sequence annotation, and integration on demand basis by client-side software
- Simple protocol that runs over HTTP

… see **http://biodas.org** for details

# DAS concept



## DAS concept

The annotations may be shared more easily.

The amount of data which needs to be stored locally is decreased.

Responsibility for updating and maintaining the data is left with the original data provider.

Conflicting annotations are permitted, encouraging information dissension and dialogue.

Annotation can be stored in a variety of formats.

Other programs, even other types of programs (not annotation viewers), may use the data.

## DAS - Annotation

- Annotation - An entity which:
  - Is anchored to the reference sequence via a stop and start value.
  - Possesses an ID unique to the server and a structured description of its types, methods, and categories.
- Global Annotation:
  - Annotation that applies to the entire reference sequence; start and stop are void

# DAS client DAS server

Definitions

- A **DAS client** is either an application (***Apollo***) or a web site (***Ensembl***) requesting and displaying annotations for a region of a sequence
- A **DAS server** is designated either as
  - –a **reference server**, that provides essential structural information (e.g. entry points or sequence), or
  - –an annotation server, that supplies specific annotation

  and has the knowledge about the sequence and/or annotation data stored in a flat file or database

# Ensembl, a DAS client

DAS clients in **Ensembl**

- **ContigView**

  Principal data visualisation tool for genome sequence annotation information
- **GeneView**

  Provides detailed information about gene model predictions
- **ProtView**

  Provides detailed information about protein model predictions

## Reference Sequences

- Applicable to **ContigView**
  - Chromosome
  - Contig
  - Clone
- Applicable to **ProtView** or **GeneView**
  - Protein
  - Gene (global annotation only)
  - cDNA (not yet implemented)

## DAS format

URL following a specific, standardised format:

<prefix>/das/<dsn>/<cmd>?<args>

- Site-specific prefix (server URL)
- Keyword “das”
- Data source name (dsn)
- Command (cmd)
- Arguments (args)

Example

http://www.ebi.ac.uk/das-srv/uniprot/das/aristotle/features?segment=P51587

# DAS format

URL following a specific, standardised format:

<prefix>/das/<dsn>/<cmd>?<args>

- Site-specific prefix (server URL)
- Keyword “das”
- Data source name (dsn)
- Command (cmd)
- Arguments (args)

Example

http://www.ebi.ac.uk/das-srv/uniprot/das/aristotle/features?segment=P51587

# DAS format

URL following a specific, standardised format:

<prefix>/das/<dsn>/<cmd>?<args>

- Site-specific prefix (server URL)
- Keyword “das”
- Data source name (dsn)
- Command (cmd)
- Arguments (args)

Example

http://www.ebi.ac.uk/das-srv/uniprot/das/aristotle/features?segment=P51587

# DAS format

URL following a specific, standardised format:

<prefix>/das/<dsn>/<cmd>?<args>

- Site-specific prefix (server URL)
- Keyword “das”
- Data source name (dsn)
- Command (cmd)
- Arguments (args)

Example

http://www.ebi.ac.uk/das-srv/uniprot/das/aristotle/features?segment=P51587

# DAS format

URL following a specific, standardised format:

<prefix>/das/<dsn>/<cmd>?<args>

- Site-specific prefix (server URL)
- Keyword “das”
- Data source name (dsn)
- Command (cmd)
- Arguments (args)

Example

http://www.ebi.ac.uk/das-srv/uniprot/das/aristotle/features?segment=P51587

# DAS essentials

<prefix>/das …

- /dsn

  Retrieves a list of data sources available from this server

- /<dsn>/types[?segment=<ref>]

  Retrieves a list of types available from a data source or particular data source segments

- /<dsn>/features?segment=<ref>

  Retrieves the annotations across a segment

Example

http://www.ensembl.org/das/dsn

## ‘dsn’ - Get the list of available sources

http://www.ensembl.org/das/dsn
http://das.ensembl.org/das/dsn

```
<?xml version="1.0" standalone="yes"?>
<!DOCTYPE DASDSN SYSTEM "http://www.biodas.org/dtd/dasdsn.dtd">
<DASDSN>
  <DSN>
    <SOURCE id="ens_34_affy10ksnp">ens_34_affy10ksnp</SOURCE>
    <MAPMASTER>http://das.ensembl.org/ensembl1834</MAPMASTER>
    <DESCRIPTION>Affymetyrix 10K SNP Array Mapping</DESCRIPTION>
  </DSN>
  <DSN>
    <SOURCE id="ens_35_affy10ksnp">ens_35_affy10ksnp</SOURCE>
    <MAPMASTER>http://das.ensembl.org/ensembl2635</MAPMASTER>
    <DESCRIPTION>Affymetyrix 10K SNP Array Mapping</DESCRIPTION>
  </DSN>
  <DSN>
    <SOURCE id="ens_anoph_dros_5_1">ens_anoph_dros_5_1</SOURCE>
    <MAPMASTER>http://servlet.sanger.ac.uk:8080/das/mosquito51</MAPMASTER>
    <DESCRIPTION>Ensembl-mapped Drosophila Genes</DESCRIPTION>
  </DSN>
  <DSN>
    <SOURCE id="ens_cbr25_fgenesh">ens_cbr25_fgenesh</SOURCE>
    <MAPMASTER>http://servlet.sanger.ac.uk:8080/das/cbriggsae1025</MAPMASTER>
    <DESCRIPTION>C. briggsae fgenesh predictions</DESCRIPTION>
  </DSN>
```

# DAS essentials

<prefix>/das …

- /dsn

  Retrieves a list of data sources available from this server

- /<dsn>/types[?segment=<ref>]

  •Retrieves a list of types available from a data source or particular data source segments

- /<dsn>/features?segment=<ref>

  Retrieves the annotations across a segment

Example

http://genome.dbs.dtu.dk:9000/das/cbs_ptm/tyes?segment=P51587

# DAS essentials

<prefix>/das …

- /dsn

  Retrieves a list of data sources available from this server

- /<dsn>/types[?segment=<ref>]

  Retrieves a list of types available from a data source or particular data source segments

- /<dsn>/features?segment=<ref>

  Retrieves the annotations across a segment

Example

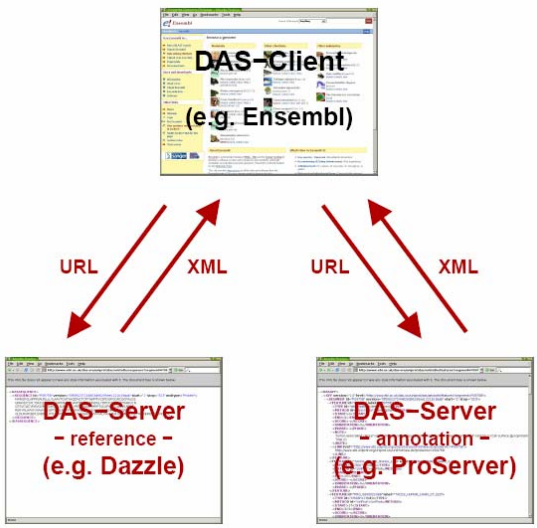
http://das.ensembl.org/das/ens_36_refseq/features?segment=1:40000,60000

## ‘features’ - Get the annotations

http://www.ebi.ac.uk/das-srv/asd/das/atd_human/features?segment=13:3770000,38100000

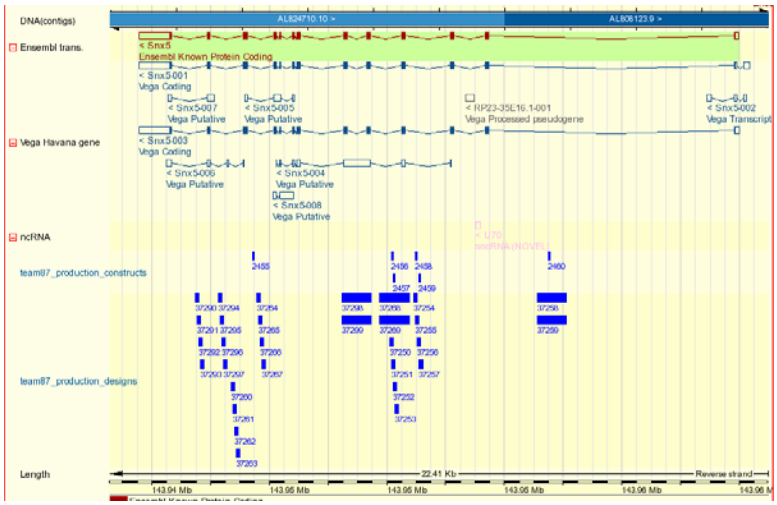
```
<?xml version="1.0" standalone="yes"?>
<!DOCTYPE DASGFF SYSTEM "http://www.biodas.org/dtd/dasgff.dtd">
<DASGFF>
  <GFF version="1.01" href="http://web59-node1.ebi.ac.uk:9100/das/atd_human/features">
    <SEGMENT id="13" version="1.0" start="3770000" stop="38100000">
    <FEATURE id="6438:1.215740" label="6438:1.215740">
      <TYPE id="exon"  reference="no" subparts="no" superparts="no">exon</TYPE>
      <START>32658014</START>
      <END>32658215</END>
      <ORIENTATION>-</ORIENTATION>
      <LINK href="http://www.ebi.ac.uk/asd-srv/Atd.cgi?method=TRANSCRIPT;origin=AT;product·
      <GROUP id="6438" label="ENSG00000133121 AT-1"  />
    </FEATURE>
    <FEATURE id="6438:1.215741" label="6438:1.215741">
      <TYPE id="exon"  reference="no" subparts="no" superparts="no">exon</TYPE>
      <START>32639687</START>
      <END>32639758</END>
      <ORIENTATION>-</ORIENTATION>
      <LINK href="http://www.ebi.ac.uk/asd-srv/Atd.cgi?method=TRANSCRIPT;origin=AT;product·
      <GROUP id="6438" label="ENSG00000133121 AT-1"  />
    </FEATURE>
    <FEATURE id="6438:1.215742" label="6438:1.215742">
      <TYPE id="exon"  reference="no" subparts="no" superparts="no">exon</TYPE>
      <START>32637473</START>
      <END>32637554</END>
      <ORIENTATION>-</ORIENTATION>
      <LINK href="http://www.ebi.ac.uk/asd-srv/Atd.cgi?method=TRANSCRIPT;origin=AT;product·
      <GROUP id="6438" label="ENSG00000133121 AT-1"  />
    </FEATURE>
```

## *Ensembl - Project*



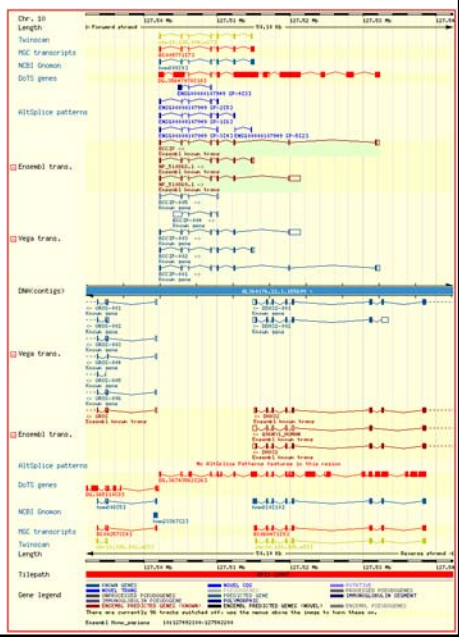
EMBL-EBI wellcome trust sanger institute

# EUCOMM



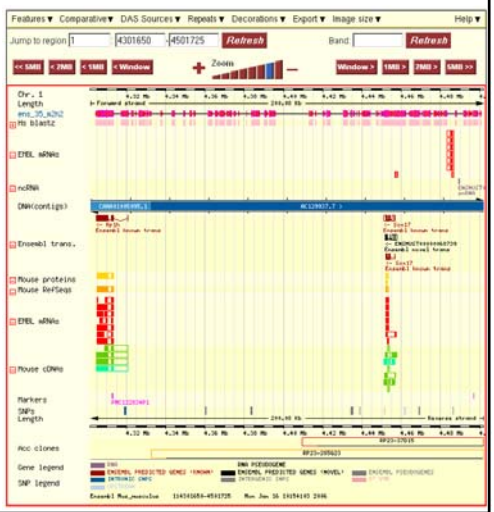

EMBL-EBI wellcome trust sanger institute

# Display 3rd party data



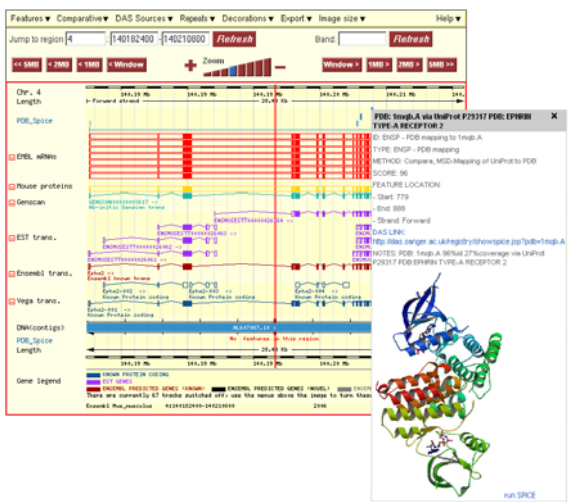
EMBL-EBI wellcome trust sanger institute

# Display
# Custom tracks



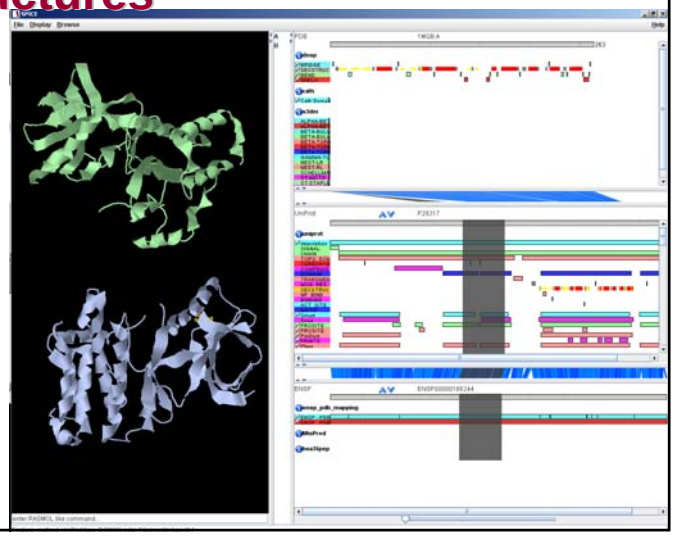

EMBL-EBI wellcome trust sanger institute

# Display
# 3D structures



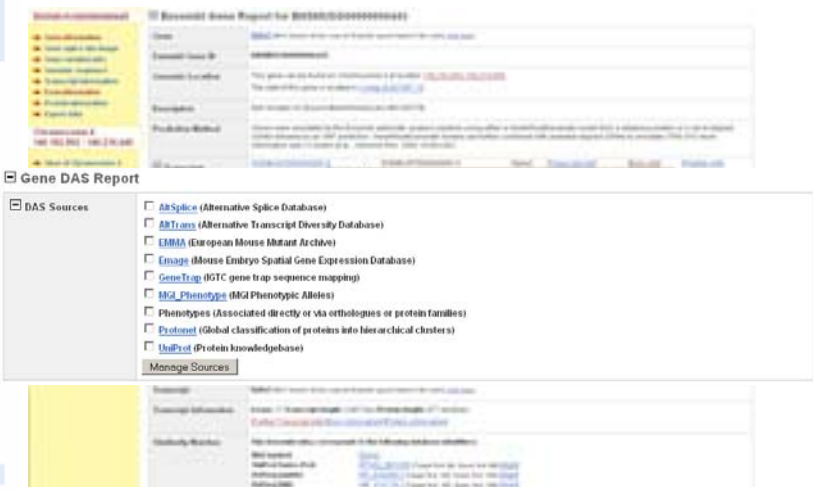
EMBL-EBI wellcome trust sanger institute

# Display 3D structures



EMBL-EBI wellcome trust sanger institute

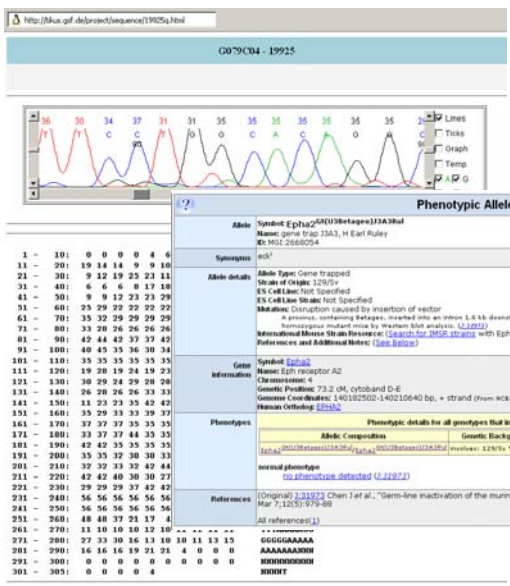
# GeneDAS



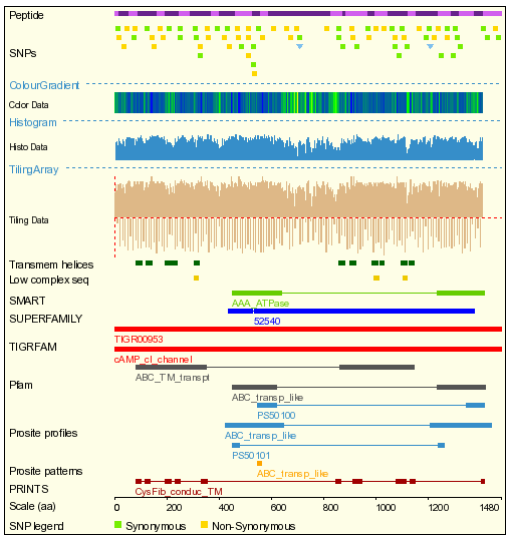
EMBL-EBI wellcome trust sanger institute

# GeneDAS

EMBL-EBI wellcome trust sanger institute



# Style sheets




# Adding custom data to Ensembl Displays

- Add a new DAS server;
  - Switch on pre-configured server,
  - Configure a new server,
  - Set up your own server (see Ensembl docs),
  - ContigView, ProtView and GeneView.
- Use Ensembl DAS server to display your data, made accessible via;
  - file upload,
  - Web page (URL),
  - ContigView only at this time.

# URL-based Upload

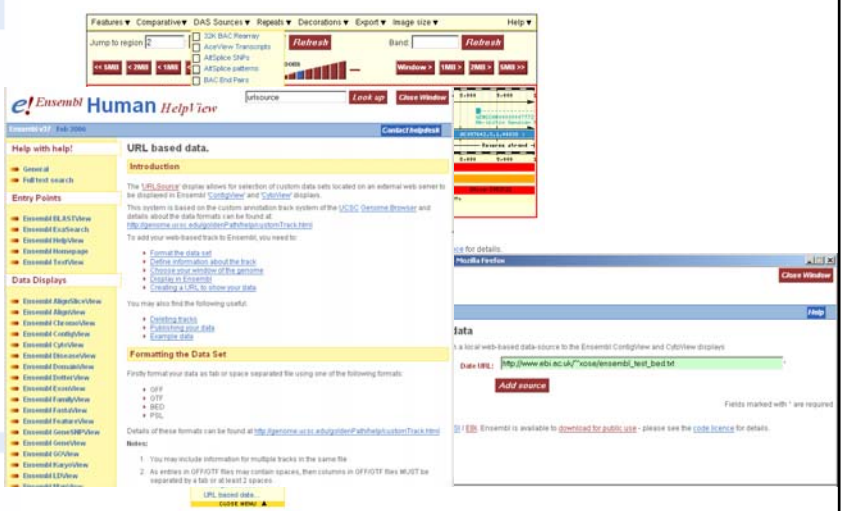
Based on the custom annotation track system of the UCSC browser.

Allowed formats: GFF, GTF, BED, PSL (*see also* http://genome.ucsc.edu/goldenPath/help/customTrack.html).

Display data by either entering the URL of the data file on the Urlsource page (ContigView: Data sources > URL based data …) or by creating the following URL:

http://www.ensembl.org/species/contigview?data_URL=URL

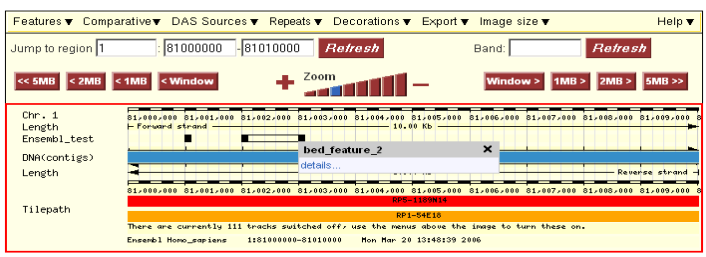
# Display of uploaded data



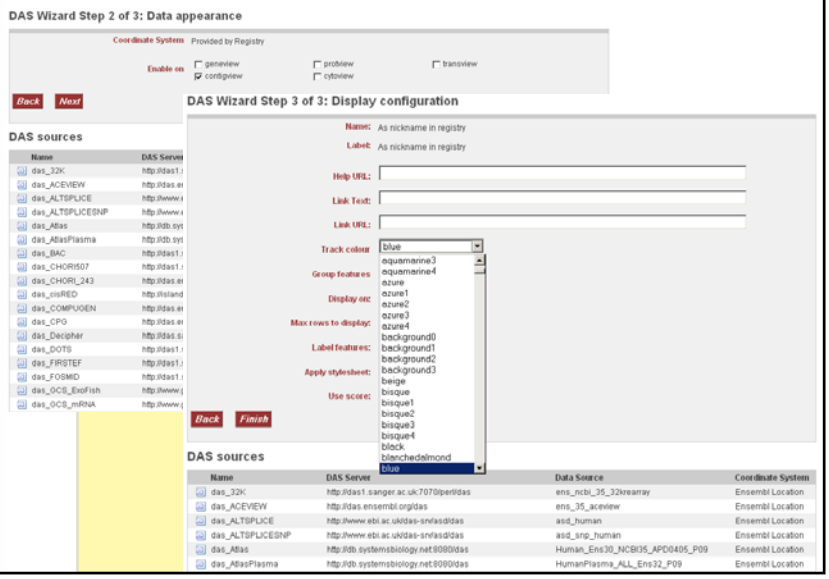
# URL-based

```
browser position chr2:1-10000
track name=Ensembl_test description="Ensembl
workshop (BED)" color=000000
url=http://www.ebi.ac.uk/~xose/ensembl_test.html
2      1000   1100   bed_feature_1 1000   +
2      2000   2100   bed_feature_2 500    +
2      3000   3100   bed_feature_2 100    +
```

# Display of data via URL



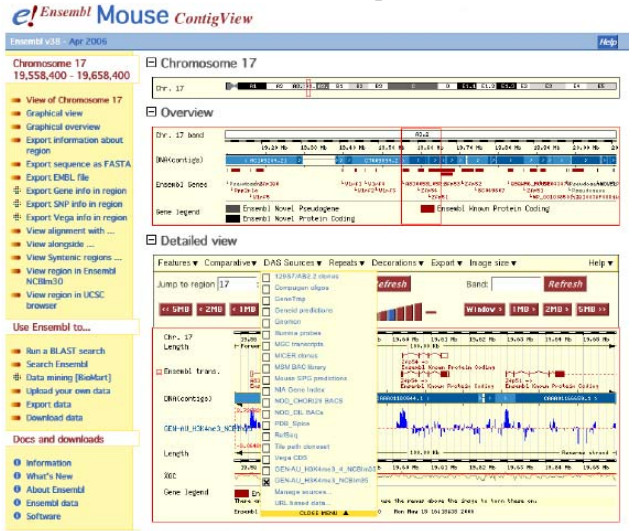
# Display of uploaded data



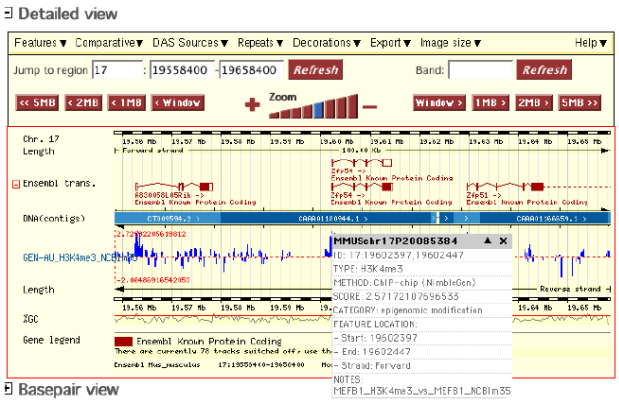

# File upload-based

```
#<col1>       <col2>       <col3>       <col4>
#<group>      <name>       <type>       <subtype>

Similarity    Fake_match_1 homology     wublastn
Transcription Fake_tscr_1  transcript   exon
Transcription Fake_tscr_1  transcript   exon

#<col5> <col6>  <col7> <col8>  <col9>  <col10>
#<chr>  <start> <end>  <strand><phase> <score>

2       4000    4050   +       .       100
2       4200    4300   +       .       100
2       4400    4500   +       .       100
```

# Example: Epigenomic ChIP-chip data




# Example: Epigenomic ChIP-chip data



# DAS Registry

http://www.dasregistry.org

# DAS Server Registration

Central DAS data source registration service at:

http://www.dasregistry.org/

A registered data source

- Becomes publicly available
- can automatically be activated

  in various DAS clients (Ensembl, SPICE, Dasty, etc.)
- gets “health”-checked

# DAS Server Implementations

Most commonly used DAS-Server implementations:

- **LDAS**
  The Lightweight DAS server using Perl, Apache and MySQL
  … see http://www.biodas.org/servers/
- **DAZZLE**
  DAS server written in Java
  … see http://www.derkholm.net/thomas/dazzle
- **ProServer**
  DAS server written in Perl
  … see http://www.sanger.ac.uk/proserver

# Ensembl as DAS server

- list of reference servers

**http://www.ensembl.org/das/dsn**

```
- <DSN>
    <SOURCE id="Homo_sapiens" version="37_35j">Homo_sapiens</SOURCE>
    <MAPMASTER>http://www.ensembl.org/das/Homo_sapiens</MAPMASTER>
  - <DESCRIPTION>
      Homo_sapiens Reference server based on NCBI35. Contains 111 entry points of the highest assembly tier.
    </DESCRIPTION>
  </DSN>
- <DSN>
    <SOURCE id="Macaca_mulatta" version="37_1a">Macaca_mulatta</SOURCE>
    <MAPMASTER>http://www.ensembl.org/das/Macaca_mulatta</MAPMASTER>
  - <DESCRIPTION>
      Macaca_mulatta Reference server based on MMUL_0_1. Contains 192158 entry points of the highest assembly tier.
    </DESCRIPTION>
  </DSN>
- <DSN>
    <SOURCE id="Monodelphis_domestica" version="37_2a">Monodelphis_domestica</SOURCE>
    <MAPMASTER>http://www.ensembl.org/das/Monodelphis_domestica</MAPMASTER>
  - <DESCRIPTION>
      Monodelphis_domestica Reference server based on BROADO2. Contains 5078 entry points of the highest assembly tier.
    </DESCRIPTION>
  </DSN>
```

# Ensembl as DAS server

- list of entry points

`http://www.ensembl.org/das/Homo_sapiens.NCBI36.reference/entry_points`


```
- <DASEP>
  - <ENTRY_POINTS href="Homo_sapiens.NCBI36.reference" version="1.0">
      <SEGMENT id="NT_113917" start="1" stop="19840" orientation="+">NT_113917</SEGMENT>
      <SEGMENT id="NT_113886" start="1" stop="96249" orientation="+">NT_113886</SEGMENT>
      <SEGMENT id="NT_113912" start="1" stop="185143" orientation="+">NT_113912</SEGMENT>
      <SEGMENT id="NT_113878" start="1" stop="106433" orientation="+">NT_113878</SEGMENT>
      <SEGMENT id="NT_113957" start="1" stop="166452" orientation="+">NT_113957</SEGMENT>
      <SEGMENT id="NT_113889" start="1" stop="161147" orientation="+">NT_113889</SEGMENT>
      <SEGMENT id="NT_113905" start="1" stop="183161" orientation="+">NT_113905</SEGMENT>
      <SEGMENT id="NT_113952" start="1" stop="184355" orientation="+">NT_113952</SEGMENT>
      <SEGMENT id="NT_113927" start="1" stop="111864" orientation="+">NT_113927</SEGMENT>
      <SEGMENT id="NT_113908" start="1" stop="13036" orientation="+">NT_113908</SEGMENT>
      <SEGMENT id="NT_113962" start="1" stop="217385" orientation="+">NT_113962</SEGMENT>
      <SEGMENT id="NT_113938" start="1" stop="44580" orientation="+">NT_113938</SEGMENT>
      <SEGMENT id="NT_113941" start="1" stop="37498" orientation="+">NT_113941</SEGMENT>
      <SEGMENT id="NT_113870" start="1" stop="145186" orientation="+">NT_113870</SEGMENT>
      <SEGMENT id="NT_113939" start="1" stop="147354" orientation="+">NT_113939</SEGMENT>
      <SEGMENT id="NT_113901" start="1" stop="182896" orientation="+">NT_113901</SEGMENT>
      <SEGMENT id="NT_113916" start="1" stop="173443" orientation="+">NT_113916</SEGMENT>
      <SEGMENT id="NT_113949" start="1" stop="159169" orientation="+">NT_113949</SEGMENT>
      <SEGMENT id="NT_113880" start="1" stop="185571" orientation="+">NT_113880</SEGMENT>
      <SEGMENT id="NT_113953" start="1" stop="131056" orientation="+">NT_113953</SEGMENT>
      <SEGMENT id="NT_113926" start="1" stop="119514" orientation="+">NT_113926</SEGMENT>
      <SEGMENT id="NT_113951" start="1" stop="152296" orientation="+">NT_113951</SEGMENT>
      <SEGMENT id="NT_113944" start="1" stop="182567" orientation="+">NT_113944</SEGMENT>
      <SEGMENT id="NT_113921" start="1" stop="39615" orientation="+">NT_113921</SEGMENT>
      <SEGMENT id="NT_113873" start="1" stop="51825" orientation="+">NT_113873</SEGMENT>
      <SEGMENT id="NT_113882" start="1" stop="172475" orientation="+">NT_113882</SEGMENT>
      <SEGMENT id="NT_113948" start="1" stop="92689" orientation="+">NT_113948</SEGMENT>
      <SEGMENT id="NT_113924" start="1" stop="139260" orientation="+">NT_113924</SEGMENT>
      <SEGMENT id="NT_113900" start="1" stop="112804" orientation="+">NT_113900</SEGMENT>
      <SEGMENT id="NT_113943" start="1" stop="81310" orientation="+">NT_113943</SEGMENT>
```


# Ensembl as DAS server

- Exploring assembly:

`http://www.ensembl.org/das/ .. /features?segment=1:1,1000000`


```
- <DASGFF>
  - <GFF>
    - <SEGMENT id="1" start="1" stop="1000000">
      - <FEATURE id="1">
          <START>1</START>
          <STOP>1000000</STOP>
          <TYPE id="chromosome" category="component" reference="yes" superparts="no" subparts="yes">chromosome</TYPE>
          <TARGET id="1" start="1" stop="1000000">1</TARGET>
        </FEATURE>
      - <FEATURE id="supercontig:NT_077912">
          <START>357583</START>
          <STOP>511231</STOP>
          <TYPE id="supercontig" category="component" reference="yes" superparts="yes" subparts="yes">supercontig</TYPE>
          <TARGET id="NT_077912" start="1" stop="153649">supercontig NT_077912</TARGET>
        </FEATURE>
      - <FEATURE id="supercontig:NT_077913">
          <START>561232</START>
          <STOP>1000000</STOP>
          <TYPE id="supercontig" category="component" reference="yes" superparts="yes" subparts="yes">supercontig</TYPE>
          <TARGET id="NT_077913" start="1" stop="438769">supercontig NT_077913</TARGET>
        </FEATURE>
      - <FEATURE id="supercontig:NT_077402">
          <START>1</START>
          <STOP>167280</STOP>
```

# Ensembl as DAS server

- Exploring assembly:

`http://www.ensembl.org/das/ .. /features?segment=NT_077912`

```
- <DASGFF>
  - <GFF>
    - <SEGMENT id="NT_077912" start="" stop="">
      - <FEATURE id="NT_077912">
          <START/>
          <STOP/>
          <TYPE id="supercontig" category="component" reference="yes" superparts="yes" subparts="yes">supercontig</TYPE>
          <TARGET id="NT_077912" start="" stop="">NT_077912</TARGET>
        </FEATURE>
      - <FEATURE id="clone:AL732372.15">
          <START>1</START>
          <STOP>153649</STOP>
          <TYPE id="clone" category="component" reference="yes" superparts="yes" subparts="yes">clone</TYPE>
          <TARGET id="AL732372.15" start="1" stop="153649">clone AL732372.15</TARGET>
        </FEATURE>
      - <FEATURE id="chromosome:1">
          <START>1</START>
          <STOP>153649</STOP>
          <TYPE id="chromosome" category="supercomponent" reference="yes" superparts="no" subparts="yes"/>
          <TARGET id="1" start="357583" stop="511231"/>
        </FEATURE>
      </SEGMENT>
    </GFF>
  </DASGFF>
```

# Ensembl as DAS server

- Exploring assembly:

`http://www.ensembl.org/das/ .. /features?segment=AL732372.15.1.153649`

```
- <DASGFF>
  - <GFF>
    - <SEGMENT id="AL732372.15.1.153649" start="" stop="">
      - <FEATURE id="AL732372.15.1.153649">
          <START/>
          <STOP/>
          <TYPE id="contig" category="component" reference="yes" superparts="yes" subparts="no">contig</TYPE>
          <TARGET id="AL732372.15.1.153649" start="" stop="">AL732372.15.1.153649</TARGET>
        </FEATURE>
      - <FEATURE id="clone:AL732372.15">
          <START>1</START>
          <STOP>153649</STOP>
          <TYPE id="clone" category="supercomponent" reference="yes" superparts="yes" subparts="yes"/>
          <TARGET id="AL732372.15" start="1" stop="153649"/>
        </FEATURE>
      </SEGMENT>
    </GFF>
  </DASGFF>
```