SIFGD manual for users

Zhen Su Lab at Dept. Bioinformatics Center
Outline

1. Home page 2
2. Gbrowse 3
   Gbrowse 4
   Gbrowse synteny
3. Search 5
4. Motif analysis 6
   Motif Scan analysis
   Motif significance analysis 7
5. Detail page 8
   Gene detail information
   Pathway detail information 9
   miRNA detail information 10
Setaria italica Functional Genomics Database

Setaria italica

Foxtail millet (Setaria italica) is a member of the Poaceae grasses, which distributes in natural and agricultural ecosystems worldwide and grows as a cereal crop or grain for both human food and fodder. It is an ancient crop originated from China and has more than 10,000 years of cultivation history. Millet is C4 grass species with significant characteristics of drought resistance, short growing season, and high light efficiency. Thus millet is favored as an important crop in the semi-arid tropics and could be a great model cereal, as well as for potential biofuel use. Recently two research groups published foxtail millet's whole genome sequencing results, but there is still limited gene function information. We try to establish a database of foxtail millet for gene function and protein network analysis through bioinformatics methods. We applied genome browser as a platform to integrate Setaria italica genome sequence, transcripts sequence, protein sequence, miRNA-seq and RNA-seq data, from public data sources such as phytozyme, NCBI, BGI, etc. In the meanwhile, some gene functional analysis tools were also provided in our database, including gene family identification, gene network prediction and motif analysis.

Update news in SIFGD

2014-05-20: Relationship of two genome version have been displayed in synteny model of GBrowse;
2014-04-20: 277 Pe and related agronomic traits have been added in GBrowse;
2014-03-10: Poster participated The Ist International Setaria Genetics Conference (ISGC);
2013-10-09: Poster participated National Congress of Plant Biology 2013(NCPB);
Annotations like gene structure, expressed sequence tag (EST) location, microRNAs, expression profile of mRNA-seq and miRNA-seq are shown here.
Gbrowse synteny

Gbrowse_syn can show genomic synteny block by submitting a chromosome region. Users can zoom in or out easily.
Search

1. Gene detail search
   Please type interested gene locus: (Example: Si005032m)

2. Metabolism pathway search
   Please type interested pathway name: (Example: Steroid biosynthesis)

3. Transcripts ID to locus ID
   Please input a list of Setaria italica ID, and we'll convert all these ID to locus ID
   Example: Setaria italica

4. Ortholog search
   Please input a list of gene ID, and we'll display their orthologous gene among Arabidopsis or Setaria italica
   Example: Setaria italica, Example: Arabidopsis

5. Batch search
   Please input a list of Setaria italica ID, and we'll extract their annotation
   Example: Setaria italica

Conversion between locus ID and transcripts ID

Ortholog relationship between Arabidopsis and S. italica

Multiple genes detail information search
Motif Scan analysis

Users can submit a promoter sequence by FASTA type. The “Sequence Scan” will find potential cis-elements appear in the promoter and their occurrence number and the whole genome background frequency of each potential cis-elements appear in 1 to 3kb (kilo base pair) promoter region will return as results.
Motif significance analysis

Analysis for known motifs

1. The locations of the motifs
2. Significance analysis

Analysis for novel motifs

custom motif scan:

Motif location
Motif annotation
**Gene detail information**

- **Synteny gene information in Zhang-gu**
- **Transcripts annotation and sequences download**
- **Gene Ontology id and annotation**
- **Enzyme id and pathway information**
- **Expression profile of 4 tissues, including FPKM and raw read count**

### Table 1: Transcript annotation information

<table>
<thead>
<tr>
<th>Transcript ID</th>
<th>Location</th>
<th>Phenotype description</th>
<th>Sequence</th>
<th>Download</th>
</tr>
</thead>
<tbody>
<tr>
<td>5005512m</td>
<td>277 bp, leading strand</td>
<td>shiga toxin-activating enzyme 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5005676m</td>
<td>307 bp, leading strand</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Gene family

<table>
<thead>
<tr>
<th>Category</th>
<th>Family</th>
<th>Subfamily</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang-gu</td>
<td>E1</td>
<td>Thf</td>
</tr>
</tbody>
</table>

### Table 3: Gene Ontology

<table>
<thead>
<tr>
<th>GO Description</th>
<th>GO:0086841</th>
<th>GO:0016544</th>
<th>GO:0051982</th>
</tr>
</thead>
<tbody>
<tr>
<td>small proteolysis activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cellular proteolysis activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein modification process</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 4: Enzyme and pathway information

<table>
<thead>
<tr>
<th>Enzyme ID</th>
<th>Pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1:4.3.1.10</td>
<td>Thf</td>
</tr>
</tbody>
</table>

### Table 5: Expression profile of 4 tissues

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Description</th>
<th>FPKM</th>
<th>Reads count</th>
</tr>
</thead>
<tbody>
<tr>
<td>5005512m</td>
<td>shiga toxin-activating enzyme 1 307 bp, leading strand</td>
<td>30.64</td>
<td>13.7</td>
</tr>
<tr>
<td>5005676m</td>
<td>shiga toxin-activating enzyme 1 307 bp, leading strand</td>
<td>25.58</td>
<td>16.3</td>
</tr>
<tr>
<td>5005512m</td>
<td>shiga toxin-activating enzyme 1 277 bp, leading strand</td>
<td>16.31</td>
<td>8.3</td>
</tr>
<tr>
<td>5005676m</td>
<td>shiga toxin-activating enzyme 1 307 bp, leading strand</td>
<td>13.34</td>
<td>5.7</td>
</tr>
</tbody>
</table>

### Diagrams

- **Overview of gene structure in Gbrowse**
- **Ortholog annotation in Arabidopsis**
- **Gene family annotation**
- **Annotation of function domain**
A list of enzymes involved in the pathway

A wiring diagrams of the pathway and enzymes present in *S. italica* were highlighted in green
miRNA detail information

- Start and end position
- Sequences of mature and precursor miRNA
- Predicted target genes and annotation of alignment
- Expression profile of four tissues, including raw and normalization reads count
- Predicted secondary structure of precursor miRNA

Overview of location mature and precursor miRNA in Gbrowse