
transPLANT 2nd user training workshop

Poznan, Poland, June, 27th, 2013

**A tutorial introduction into the
MIPS PlantsDB barley&wheat
database instances**

TUTORIAL ANSWERS

**Please direct any questions related to this tutorial or MIPS PlantsDB in
general to:**

manuel.spannagl@helmholtz-muenchen.de

**Manuel Spannagl
MIPS/IBIS Munich
Helmholtz Center Munich**

Exercise 1

☞ Go to the MIPS PlantsDB homepage: <http://mips.helmholtz-muenchen.de/plant/genomes.jsp>.

☞ Select *TriticeaeProject* in the left menu or *Hordeum vulgare* in the taxonomy tree as your database and select "Gene annotation" and then "Search" in the left menu. Enter **MLOC_67600.1** in the "Search for Genetic Elements" field and press "Search". Clicking the gene ID in the result list will bring you to the gene report page that contains all information about that gene/locus.

The gene is located on **morex_contig_53323**.

☞ To search for all genes annotated as "WUSCHEL-related homeobox" go back to the search form and type these keywords into the "Search for genetic elements" field.

This free-text search for a functional description returns a total of 8 different barley genes.

Exercise 2

☞ Open the PlantsDB barley instance (<http://mips.helmholtz-muenchen.de/plant/barley/index.jsp>) and navigate to "Search".

☞ Type "Alpha-glucosidase" into the "Search for Genetic Elements" field and press "Search".

☞ Click the gene Id of the fourth result (**MLOC_68876.3**) and proceed to the gene report.

☞ Scroll down to the section "Gene family information" and click the "Gene family link". This report gives you information about closely related genes within the same and/or other organisms (for the methods and the species compared see the top section).

There are six putative orthologs for **MLOC_68876.3**, two in rice, two in maize, one in brachypodium and one in *Arabidopsis thaliana*. Moreover, three putative barley paralogous sequences were identified: AK354012, MLOC_15619.3, MLOC_60415.1.

☞ Go back to the gene report and scroll down a little bit to "Sequence report for MLOC_68876.3"...then click the "CDS" and "Protein" link which will direct you to the download link for both sequences.

Exercise 3

🔗 Navigate to the barley GenomeZipper entry page at: <http://mips.helmholtz-muenchen.de/plant/barley/gz/index.jsp> and proceed to the "Search" section. At the zipper entry page more detailed information about the concept is available incl. further reading. Insert the Brachypodium gene identifier (**Bradi2g39900.1**) for which you want to search associated barley loci into the "Search GenomeZipper for marker or syntenic genes by name" field. The result provides links to the identified reference gene reports (Brachypodium, rice, Sorghum) as well as links to the anchoring marker report and to the GenomeZipper report where all "zipped" elements (incl. barley flcDNAs, 454 genomic reads; ESTs) can be queried and downloaded.

In the barley genome zipper, **Bradi2g39900.1** is anchored on barley chromosome 1H at an approximate genetic location of 0.95 cM. That means that, given local syntenic conservation with the reference genomes Brachypodium, Rice and Sorghum at that region, the barley ortholog of **Bradi2g39900.1** can likely be found at that position (and local gene order) on the barley genome. This anchoring was established directly by a matching marker, "2_0842" (this is a SNP marker, detailed information about this marker can be obtained from the marker report by clicking the identifier).

At that zipper position (loci 31 on barley chromosome 1H), gene models from the reference genomes Sorghum and rice have been identified as well: **Sb09g000380.1** and **Os05g0103500**. Clicking any of these genes directs you to the corresponding gene reports in the species PlantsDB instances. If you want to explore the neighboring (and ordered) anchored loci in that region you can click the "Loci-Position on Chr" number in the GenomeZipper report (third line). This view reports 15 loci upstream and downstream of the current zipped gene.

Exercise 4

🔗 Open the gene report of the Brachypodium gene **Bradi2g39900.1**, either by clicking one of the corr. gene links you still have open from exercise 3 or by searching this gene at <http://mips.helmholtz-muenchen.de/plant/brachypodium/searchjsp/index>. Scroll down to "References" and hit the "CrowsNest_SytenyToBarley" link.

This link brings you directly to the CrowsNest Synteny visualization tool scaled at a micro-synteny display level (the “most detailed view”-maximum zoomed-in). The view visualizes the syntenic relationship of the selected Brachypodium gene with its barley ortholog in the genomic/chromosomal context of both genome sequences. Please note that no continuous pseudo-chromosome sequences are available for the barley genome yet, nevertheless we can visualize the syntenic relationships to reference genomes on the basis of genetically/physically anchored or “genome-zipped” barley genes. CrowsNest is capable of visualizing the syntenic relationships on different levels of hierarchy, from very general views (macro-synteny, CrowsNest levels 1-3) down to gene-by-gene views (micro-synteny, CrowsNest level4, your current level). Navigation between the levels is possible using the navigation bar in the upper left corner. In our example, the Brachypodium genome region is depicted on the lower bar, whereas the barley genome region is shown in the upper bar. The upper overview picture indicates the regions you are currently looking at in context of the full chromosomes. Coordinates are given to help locate the regions.

Moving your mouse cursor over the connecting line gives you details about the established orthologous relationship. The Brachypodium gene in our example is connected with the barley gene **MLOC_57152.1**. Clicking the small purple boxes representing the genes models directs you to the corresponding gene reports in PlantsDB where you can obtain detailed gene information. The chromosomes involved in this syntenic relationship are Brachypodium chromosome 2 and barley chromosome 1. Looking at the small upper overview or/and the higher CrowsNest level views indicates that there is at least one larger-scale inversion observable between these two chromosomes.

Exercise 5

🔗 Download the DNA sequence from ftp://ftpmips.helmholtz-muenchen.de/plants/user_training/UnknownGrassSeq.fa and insert it into the BLAST field of the IPK barley ViroBLAST server at <http://webblast.ipk-gatersleben.de/barley/viroblast.php>.

🔗 Select “blastn” as the Program and “HC_genes_CDS_seq” as your Database. Leave all other parameters at their default value and hit “Basic search”.

The BLAST result indicates a very good match of the query sequence to the Barley HC gene **MLOC_58525.1**. The BLAST database you just searched against consists of the CDS sequences of all barley high-confidence gene models. This dataset is also available as bulk download from the FTP server in case you want to use it in a local BLAST search, e.g. on your machine.

☞ Copy/remember the Barley gene identifier and go to the MIPS PlantsDB barley database at <http://mips.helmholtz-muenchen.de/plant/barley/ga/searchjsp/index.jsp> to search this gene model.

☞ Paste/type the Barley gene identifier into the "Search for Genetic Elements" field and hit "Search".

The result page summarizes all information about this barley gene. According to its predicted functional description this gene encodes for a **CCAAT-box-binding transcription factor-like protein**.

☞ To retrieve all domain annotations (such as PFAM, INTERPRO) for this gene scroll down the gene report and hit the "List protein features and domains in this gene" link in the "Protein domain informations from SIMAP" section.

A number of different domains was annotated for this gene model using INTERPROScan, including **IPR005612 (CCAAT-binding factor)** and **IPR016024 (Armadillo-type fold)**.

Exercise 6

☞ Download the DNA sequence from ftp://ftpmips.helmholtz-muenchen.de/plants/user_training/UnknownGrassSeq.fa and insert it into the BLAST field at <http://mips.helmholtz-muenchen.de/plant/wheat/uk454survey/searchjsp/blast.jsp?organism=wh eatukortho>.

☞ Select "Nucleotide" as Sequence type and "Proteins" or "ORFs" as your Database. Leave all other parameters at their default value and hit "BLAST".

The BLAST result indicates a very good match of the query sequence to the Brachypodium gene **Bradi3g47770.1**. The BLAST database you just searched against consists of the set of representative grass orthologous genes used in the UK wheat genome analysis. More details about this dataset are available from the PlantsDB wheat entry webpage or the publication [Brenchley et al., Nature 2012]. The dataset is also available as bulk download from the FTP server in case you want to use it in a local BLAST search, e.g. on your machine.

☞ Copy/remember the Brachypodium gene identifier and go to <http://mips.helmholtz-muenchen.de/plant/wheat/uk454survey/searchjsp/index.jsp> to search for all wheat genic sequences associated to this reference gene model.

☞ Paste/type the Brachypodium gene identifier into the "Search for Genetic Elements" field and hit "Search".

The result page summarizes all information generated in the UK wheat project associated with this grass reference gene model (in our example it is a representative from Brachypodium). This Brachypodium gene model was selected as a representative for the orthologous gene group given in the table on the lower section.

☞ To download the associated genic sub-assembly sequences from bread wheat move your cursor to the "Wheat ref. sequence" section and select "[FASTA]" to download the sequences in fasta format. Save the file on your local harddisk.

☞ Open the downloaded sequence file using a text editor (of your choice).

This file contains genic sub-assembly sequences of bread wheat which were associated to a corresponding grass reference gene model (NOTE: NOT full-length wheat gene models in most cases!).

The fasta entries look like this:

```
>Traes_Bradi3g47770.1_000017_A wheat UK 454 genic subassembly sequence associated to OG rep gene Traes_Bradi3g47770.1_000017_A  
CTGAGGACCCTGCTAT...
```

The identifier **Traes_Bradi3g47770.1_000017_A** contains information about the associated repr. grass gene model (**Bradi3g47770.1**) as well as its predicted wheat sub-genome origin (classified as coming either from the A, B or D sub-genome). If no classification was possible, a "X" was appended instead of "A", "B" or "D".

Exercise 7

☞ Go to the MIPS PlantsDB barley instance: <http://mips.helmholtz-muenchen.de/plant/genomes.jsp>.

☞ Click on "Triticeae Project".

☞ Click on "Genomes".

☞ Click on "Barley".

☞ Click on "Download".

The download webpage summarizes the data contents available for this organism as well as information on data download policies (if applicable) and/or publications to cite. For barley, the download is organized in anchoring, expression, genes and sequences sections.

☞ Click the link at "Download" (end of page) to proceed to the barley FTP page.

☞ Click "genes".

☞ Download (or click)

`barley_HighConf_genes_MIPS_23Mar12_ProteinSeq.fa`.

This file contains all protein sequences annotated for the barley genome, classified as "high-confidence".

There is a classification scheme for the barley gene sequences into "high-confidence" (defined by homology to reference proteins) and "low-confidence" (showing lower or no homology to reference proteins).

Additional downloads are available from the sections of "sequences" (incl. WGS assemblies), "expression" (expression data) and "anchoring" (physical/genetic map integration).