**Practical session**

**Using BioMart to query genome features**

Pilar Corredor Moreno (pilar.moreno@earlham.ac.uk)

## **Task 1 – finding a gene**

You are looking for genes involved in disease response in wheat. You have found an interesting genomic region in chromosome 7BL (TGACv1\_scaffold\_577914\_7BL). Using BioMart’s Triticum aestivum (TGACv1) gene dataset, filter on chromosomal region.

1. **How many protein-coding transcripts are there in this region?**

Now, you would like to know what protein-coding transcripts involved in the 'detection of biotic stimulus' are present in this region.

1. **Is there any gene with that function? What's the gene ID and its name?**

Using the option Attributes, find the following information:

1. **Which are the gene coordinates?**
2. **Is it on the forward or reverse strand?**

## **Task 2 – Finding similarities**

Using the Attributes tab, select orthologous and try to answer the following questions:

1. **Can you find any orthologous of this gene in barley (Hordeum vulgare L.)?**
2. **If so, what's the percentage of the target that is identical to the query?**
3. **Which are the Gene IDs of the wheat homeologues?**

## **Task 3 – Export results**

Obtain a file containing the unspliced transcript sequences in FASTA format with headers containing the following Attributes:

- Ensembl Gene ID  
- The associated gene name  
- Ensembl Transcript ID  
- Whether the transcript is on the forward (1) or reverse strand (-1)  
- The trancript coordinates

## **Task 4 – Find Variation**

You have two wheat genes you’re interested in and you want to find natural variation in them.

**TRIAE\_CS42\_4AL\_TGACv1\_290734\_AA0988760**

**TRIAE\_CS42\_2BL\_TGACv1\_129359\_AA0380140**

Click on New and choose Plant Variation Mart

1. **Can you find any variations in those genes using BioMart’s Triticum aestivum (TGACv1) variation dataset?**
2. **What's the consequence of such variations?**
3. **Find all the SNP’s located in the 3’-untranslated regions for these 2 genes.**
4. **Obtain a FASTA file for the variations with upstream and downstream flacking regions of 100 bp. The header should include:**

* Variation ID
* Variation source
* Chromosome name
* Position on the chromosome
* Variant sequences

1. **Is any of those variations observed in the Cadenza background?**