

Expression browsers

The reduction in cost of next generation sequencing has generated a large increase of gene expression data from RNA-Seq studies in wheat in the last 5 years. However, until recently most of this data has been inaccessible to wheat researchers and breeders due to the complicated bioinformatics analysis required to unlock this information. The development of gene expression atlases for wheat allows researchers to access information about the expression of genes in different circumstances. For example, a list of candidate genes within a mapping interval can be narrowed down using the expression patterns of these genes.

Currently three expression browsers exist which will be described below: expVIP (www.wheat-expression.com), Wheat eFP and WheatExp.

a) expVIP

expVIP is a flexible RNA-Seq analysis platform that can produce an interactive display of RNA-Seq data for any species with a transcriptome. expVIP was used to generate a wheat expression browser (www.wheat-expression.com) which incorporates data from 32 studies (1,145 RNA-Seq samples) including:

- different tissues across developmental stages
- different cultivars
- grain development
- grain layers
- senescence
- abiotic stresses (phosphate starvation, drought and heat stress)
- disease infection (yellow rust, powdery mildew, septoria, rice blast and fusarium)

expVIP is able to compare these different studies by organizing the data into specific categories; this is referred to as metadata (**Figure 1**). Metadata are divided into high-level categories: variety, tissue, developmental stage and stress. Each high-level category is further subdivided into low-level categories. High-level categories are broad filters, whereas low-level categories allow for finer differentiation. Using these metadata categories, website users can dynamically filter the displayed data to focus on their traits of interest (**Figure 2**).

Variety		Tissue		Dev. stage		Stress	
High level	Variety	High level	Tissue	High level	Dev stage	High level	Stress
Other	CS Nullitetra	Grain	Endosperm	Seedling	7 days	None	None
	Bobwhite		Transfer cells		14 days	Disease	Septoria 4 dpi
	Avalon			Septoria 10 dpi
	Cadenza	Shoots	Flag leaf	Vegetative	3 leaf stage		Mildew 24 h
	...		2 nd leaf		5 leaf stage		Yellow rust 3 dpi
		
Chinese Spring	Chinese Spring	Spike	Spikelets	Reproductive	Anthesis	Abiotic	Heat (1 h)
			Stamen		5 dpa		Drought (6h)
			...		20 dpa		P starvation (10 d)
		Roots	roots	

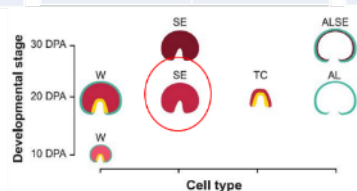


Figure 1: Example of metadata categories

This figure illustrates some of the high- and low-level categories that are used to partition the data. As an example, a Chinese Spring starchy endosperm grain sample harvested 20 days post anthesis would be categorised as shown by the green highlighted text.

expVIP displays expression values of your gene(s) of interest by default in all studies. These are visualized either as bar graphs for single genes (**Figure 2**, top) or as heatmaps for multiple genes (**Figure 3**). By default, the data is not ordered, but this can be changed by simply clicking the coloured tiles underneath the metadata category of interest, e.g. High-level Tissue or Variety (**Figure 2**, bottom).

The heatmap display mode is particularly useful for determining the expression of candidate genes underlying a QTL. In the example shown in **Figure 3**, the expression pattern of a set of genes shows that one gene is only expressed in the spike, while the other genes show a broad expression pattern. One can thus query genes underlying a QTL and investigate their expression across the studies to narrow down the list of candidate genes.

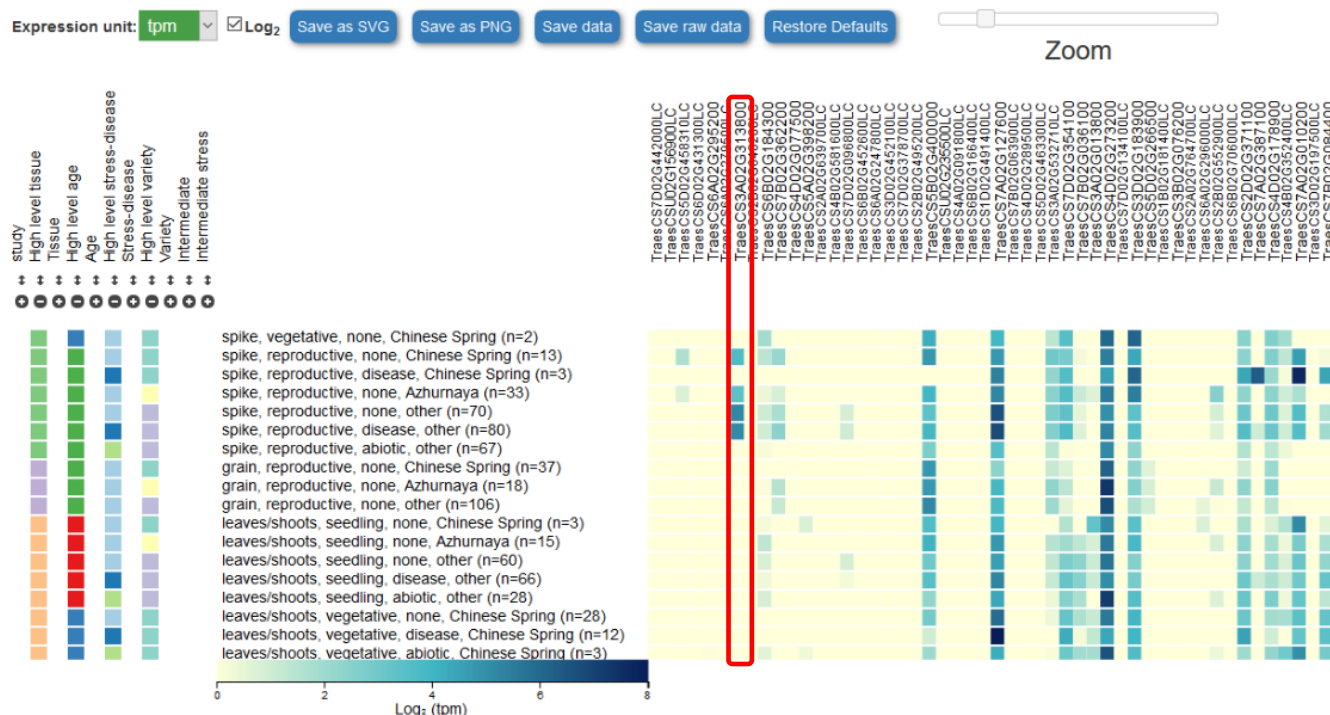


Figure 3: Expression of an example gene set displayed as a heatmap on expVIP.

This figure exemplifies the heatmap display on expVIP. The data is sorted by High-level Tissue. The red box highlights a gene that is only expressed in the spike, as opposed to the other genes which are either broadly expressed or not expressed at all. Visualising expression data in such a way can be useful when comparing lists of genes within a QTL, for example, to identify putative candidate genes. Note that most non-expressed genes are low-confidence gene models.

Another feature allows users to visualize all homoeologs of a gene of interest. This is simply done by ticking the box labelled “Homoeologs” at the top of the webpage (**Figure 4**). Homoeologs are defined as a triad of genes located at the same syntenic position across the three genomes in all selected samples based on the Ensembl Plants Plant Compara pipeline (Vilella et al., 2009; Ensembl Plants, 2018).

A ternary plot showing the expression of a given triad is available via ticking the corresponding box and allows users to determine two things: (i) is my gene of interest part of a triad that has homoeolog expression bias, and (ii) are the homoeologs expressed in a stable or dynamic manner (see **Figure 5**)?

Homoeolog expression bias refers to a marked difference in expression between homoeologs, i.e. if one homoeolog is expressed less (suppressed) or more (dominant) than the other two homoeologs. We also measure the variation of each triad across tissues and catalogue the 10% of triads with the least variation across tissues (stable triads) and the 10% of triads with the largest change across tissues (dynamic triads). This terminology is based on [Ramirez-Gonzalez et al., \(2018\)](#), which users are encouraged to read.

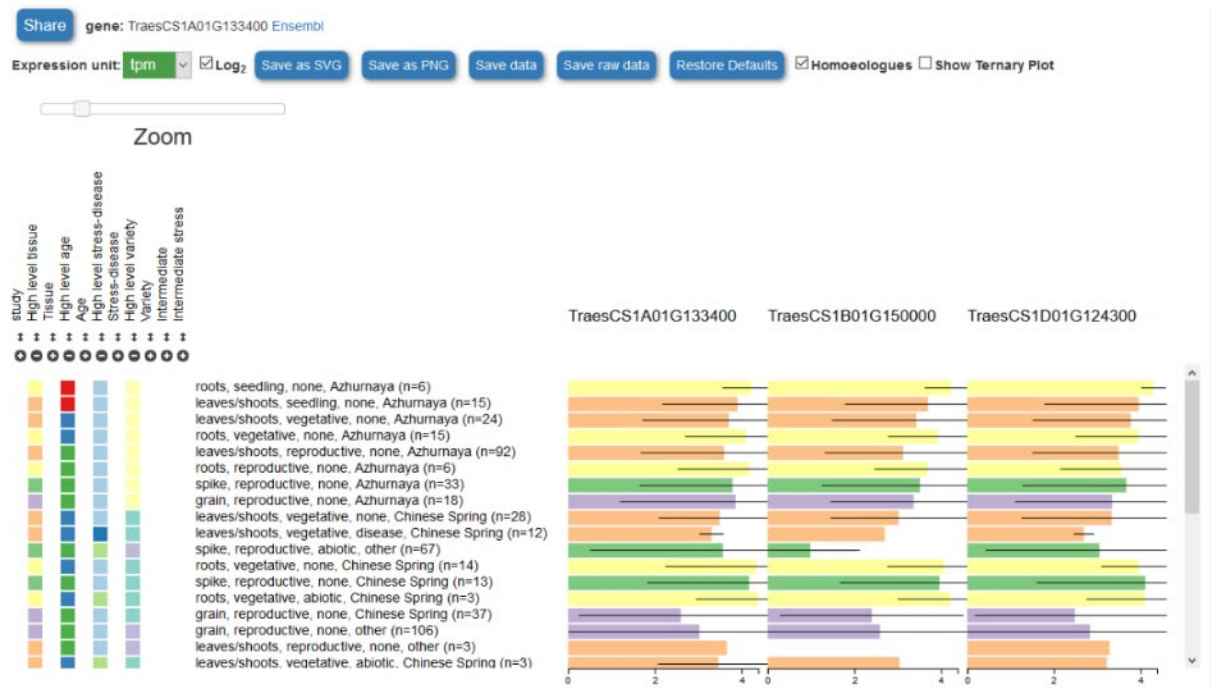


Figure 4: Homoeolog view

By ticking the “Homoeologs” box users can simultaneously compare the expression profile of all homoeologous copies of a given triad.

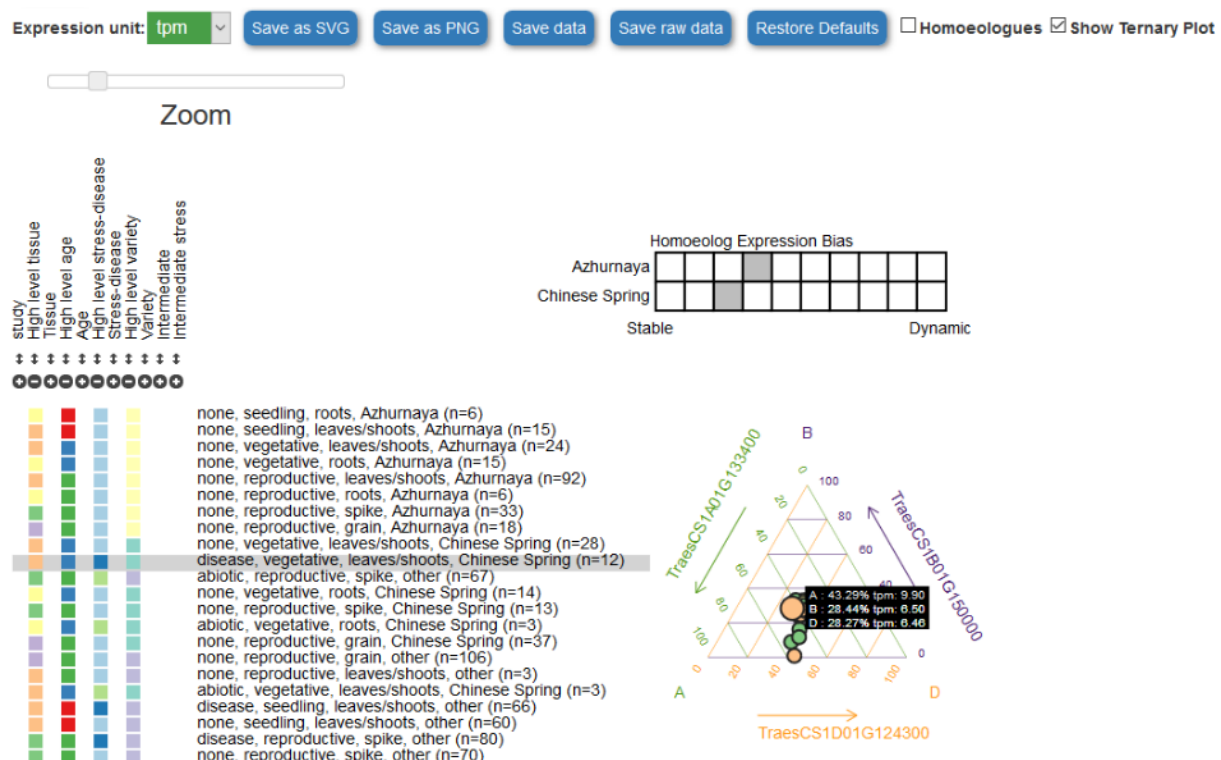


Figure 5: Triad view

By ticking the “Show Ternary Plot” box users can quickly investigate whether a triad displays an homoeolog expression bias and whether the triad is expressed in a stable or dynamic fashion. Each sample is displayed based on the selected metadata. In this way, the homoeolog expression bias of the triad can be compared across all of the selected conditions. In this case, the highlighted sample is biased towards the A genome (accounting for 43% of the expression), while the other orange sample is completely suppressed for the B genome.

Virtual machine

expVIP is a highly flexible platform which requires 3 inputs: metadata (as described above), reads from RNA-Seq experiments and a reference transcriptome (**Figure 6**). This allows expVIP to be used not just on wheat but any species with a reference transcriptome.

A user-friendly virtual machine which runs the entire expVIP platform on a desktop PC is available (<https://github.com/homonecloco/expvip-web/wiki/LoadingVM>) and has an extensive tutorial available at the same site, putting RNA-Seq analysis and visualisation within the reach of non-bioinformaticians.

The virtual machine comes in two versions:

- with the existing wheat data pre-loaded so a user can add in their own wheat samples to compare
- empty so that RNA-Seq data for any species can be used

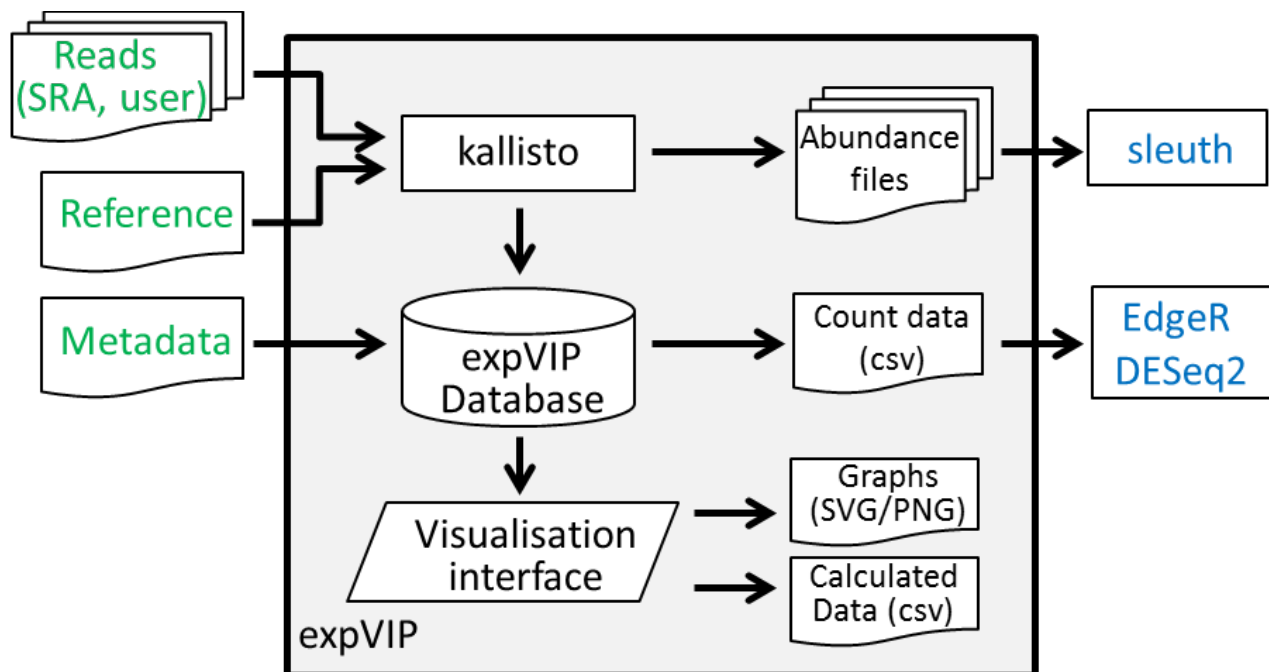


Figure 6: The expVIP platform

The virtual machine generates a visual display exactly the same as found at www.wheat-expression.com but including the users own samples, which are stored locally (user data is not made public). The expVIP platform also generates data which is ready to use in downstream differential gene expression analysis, for example using DESeq2, EdgeR or sleuth.

Graphs can be saved from the website display in both png and svg format, and all data can be extracted at the push of a button.

Tutorials

A tutorial explaining how to use expVIP's features is available both as a video:

<https://www.youtube.com/watch?v=axDfuVn6GIk>

And as a walk-through pdf:

[https://github.com/homonecloclo/expvip-web/wiki/Tutorial-expVIP-Graphical-Interface-\(Wheat-Expression-Browser-example\)](https://github.com/homonecloclo/expvip-web/wiki/Tutorial-expVIP-Graphical-Interface-(Wheat-Expression-Browser-example))

expVIP website and references

expVIP website: www.wheat-expression.com

expVIP reference:

- Borrill et al., 2016, DOI: 10.1104/pp.15.01667
- Ramirez-Gonzalez et al., 2018, DOI: 10.1126/science.aar6089

b)Wheat eFP

Wheat eFP (http://bar.utoronto.ca/efp_wheat/cgi-bin/efpWeb.cgi) contains gene expression data from a developmental time course of the cultivar Azhurnaya (Bayer Crop Science, Ghent). It includes 15 developmental stages and 71 tissues, all visible as pictograms with gene expression shown as a colour scale (**Figure 2A**).

Gene expression can be displayed in three different manners: (i) absolute, (ii) relative to its control value, which is the median level of expression of that gene across all tissues displayed in the browser, or (iii) compared between two genes. It is important to note that the scale in the absolute view varies according to the highest expression value of a gene at any given stage. For example, if a gene is lowly expressed across all samples and its highest expression value is 0.7, the associated pictogram will be coloured in red. But the same is true if the expression value is 800. It is thus critical to check the scale when assessing gene expression and not simply rely on the colours within the pictograms.

Expression values can be downloaded as a table or displayed as a chart (**Figure 2B** and **2C**)

Wheat eFP website: http://bar.utoronto.ca/efp_wheat/cgi-bin/efpWeb.cgi

Wheat eFP citations: Ramirez-Gonzalez et al., (2018) [doi:10.1126/science.aar6089](https://doi.org/10.1126/science.aar6089)

Winter et al., (2007) <https://doi.org/10.1371/journal.pone.0000718>

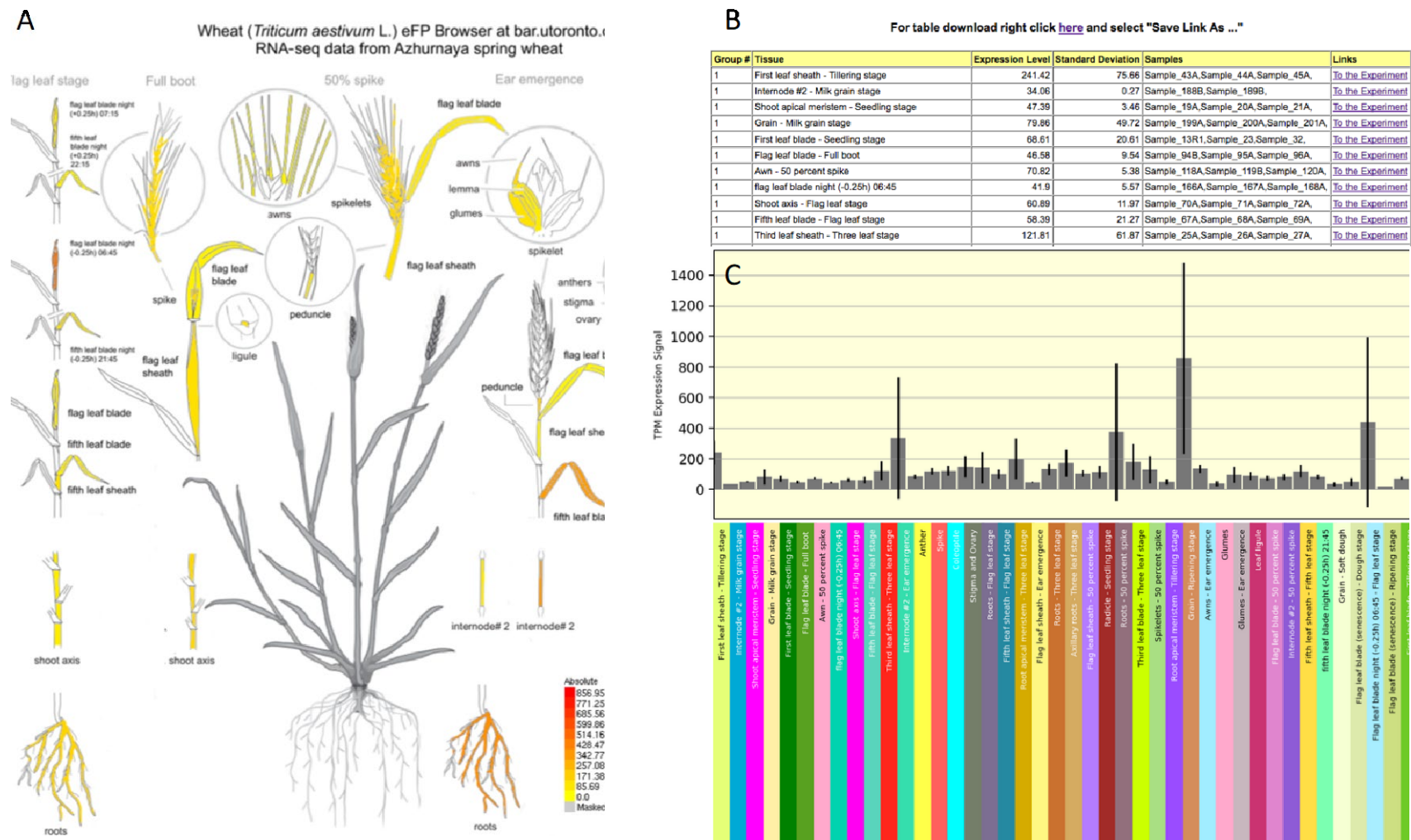


Figure 7. Screenshots from the Wheat eFP expression browser
Pictograms showing different wheat developmental stages and tissues (A). Expression is displayed using a colour scale, the values of which are shown at the bottom right. Note that the values change relative to each gene displayed! Gene expression values shown as a downloadable table (B) and as a bar chart across all samples (C).

c) WheatExp

WheatExp (available at <http://wheat.pw.usda.gov/WheatExp/>) includes gene expression data from 6 studies:

- different tissues over development
- grain development
- grain layers
- senescence time-course
- drought and heat stress
- photomorphogenesis

Each study is displayed separately with a graphic interface describing the particular study that is being viewed. For example, expression of the gene *Traes_4DS_1CF5D9107* in the “grain layer” study reveals higher expression in the endosperm than in the pericarp (**Figure 8**). The display of multiple genes side by side is possible and the database can be searched either by gene ID or a BLAST interface. Note that WheatEXP only uses CSS gene models (see the [Gene Models](#) section).

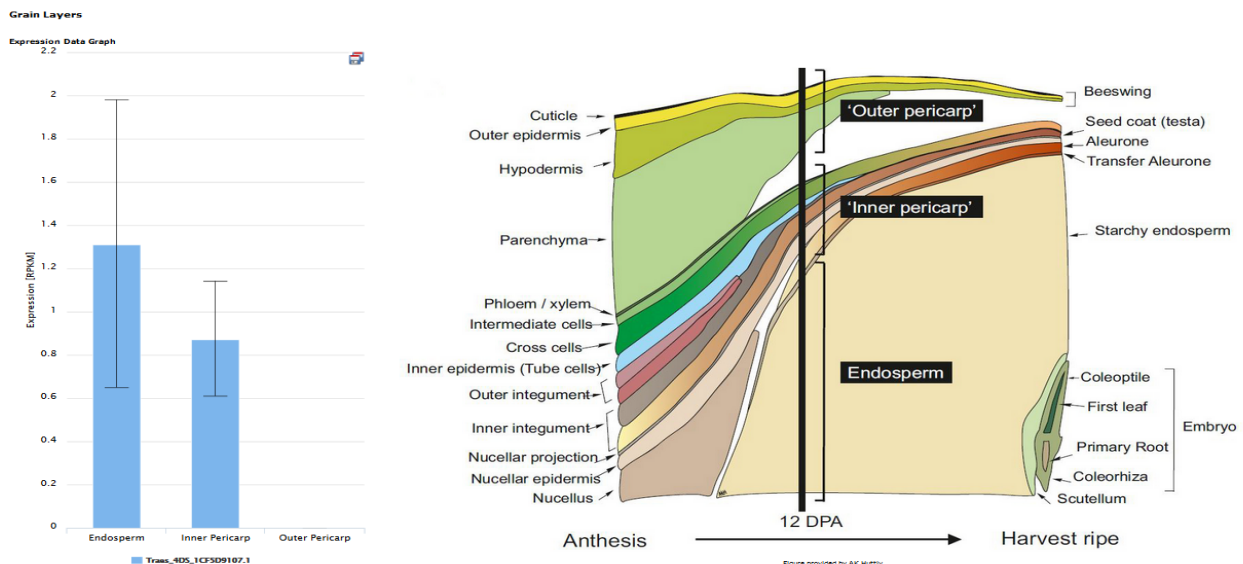


Figure 8: Expression of *Traes_4DS_1CF5D9107* in the “grain layer” study on WheatExp.

WheatExp website and citation

WheatExp website: <http://wheat.pw.usda.gov/WheatExp/>

WheatExp reference: Pearce *et al.*, 2015. DOI: 10.1186/s12870-015-0692-1