2022

qTeller: a tool for comparative multi-genomic gene expression analysis

Margaret R. Woodhouse

Shatabdi Sen

David Schott

John L. Portwood II

Michael Freeling

See next page for additional authors

Authors

Margaret R. Woodhouse, Shatabdi Sen, David Schott, John L. Portwood II, Michael Freeling, Justin W. Walley, Carson M. Andorf, and James Schnable

Bioinformatics, 38(1), 2022, 236–242
doi: 10.1093/bioinformatics/btab604
Advance Access Publication Date: 18 August 2021
Original Paper

OXFORD

Databases and ontologies

qTeller: a tool for comparative multi-genomic gene expression analysis

Margaret R. Woodhouse [1,*,†], Shatabdi Sen[2,†], David Schott[3], John L. Portwood II [1] Michael Freeling[4], Justin W. Walley[2], Carson M. Andorf[1,3] and James C. Schnable[5]

[1]USDA-ARS, Corn Insects and Crop Genetics Research Unit, Ames, IA 50011, USA, [2]Department of Plant Pathology & Microbiology, Iowa State University, Ames, IA 50011, USA, [3]Department of Computer Science, Iowa State University, Ames, IA 50011, USA, [4]Department of Plant & Microbial Biology, University of California, Berkeley, Berkeley, CA 94720, USA and [5]Center for Plant Science Innovation & Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68588, USA

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Associate Editor: Janet Kelso

Received on March 18, 2021; revised on July 23, 2021; editorial decision on August 11, 2021; accepted on August 17, 2021

Abstract

Motivation: Over the last decade, RNA-Seq whole-genome sequencing has become a widely used method for measuring and understanding transcriptome-level changes in gene expression. Since RNA-Seq is relatively inexpensive, it can be used on multiple genomes to evaluate gene expression across many different conditions, tissues and cell types. Although many tools exist to map and compare RNA-Seq at the genomics level, few web-based tools are dedicated to making data generated for individual genomic analysis accessible and reusable at a gene-level scale for comparative analysis between genes, across different genomes and meta-analyses.

Results: To address this challenge, we revamped the comparative gene expression tool qTeller to take advantage of the growing number of public RNA-Seq datasets. qTeller allows users to evaluate gene expression data in a defined genomic interval and also perform two-gene comparisons across multiple user-chosen tissues. Though previously unpublished, qTeller has been cited extensively in the scientific literature, demonstrating its importance to researchers. Our new version of qTeller now supports multiple genomes for intergenomic comparisons, and includes capabilities for both mRNA and protein abundance datasets. Other new features include support for additional data formats, modernized interface and back-end database and an optimized framework for adoption by other organisms' databases.

Availability and implementation: The source code for qTeller is open-source and available through GitHub (https://github.com/Maize-Genetics-and-Genomics-Database/qTeller). A maize instance of qTeller is available at the Maize Genetics and Genomics database (MaizeGDB) (https://qteller.maizegdb.org/), where we have mapped over 200 unique datasets from GenBank across 27 maize genomes.

Contact: margaret.woodhouse@usda.gov

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Since the introduction of RNA-Seq technology over 10 years ago (Wang *et al.*, 2009), the number of available RNA-Seq libraries has increased rapidly (Fig. 1). Many software programs, mostly in R, such as EdgeR, ggplot2, WGCNA and DEvis (Langfelder and Horvath, 2008; Price *et al.*, 2019; Robinson *et al.*, 2010; Wilkinson, 2011), have been created to visualize RNA-Seq abundances across different tissues and time points. However, there are few tools that allow users not trained in programming to visualize RNA-Seq expression patterns across multiple genes or genomic intervals, particularly in an interactive way or to compare any given two genes. In 2012, this need was addressed in the creation of qTeller, a web-hosted RNA-Seq visualization platform that allows users to compare RNA-Seq expression across tissues within a genomic interval, across multiple genes or compare expression between any two genes in a given genome (https://github.com/jschnable/qTeller). The platform displays preanalyzed values from publicly available, published datasets. At the time, qTeller hosted instances for *Zea mays*, *Arabidopsis thaliana* and *Brassica rapa*. Although unpublished, qTeller has been used by many researchers and cited extensively, including in the areas of evolution (Man *et al.*, 2020; Pophaly and

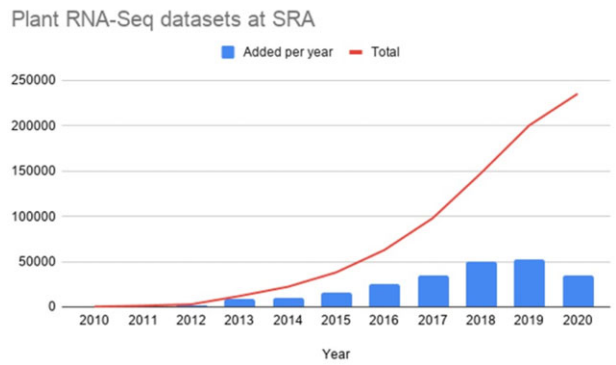

Fig. 1. Plant RNA-Seq datasets at the NCBI Short Read Archive. The chart shows the growth of plant RNA-Seq datasets at the GenBank Short Read Archive from 2010 to 2020. The *x*-axis is labelled by year. The *y*-axis is labeled by the number of RNA-Seq experiments. The blue bars for each year show the number of experiments added during that calendar year. The red line shows the cumulative number of experiments available during the given year

Tellier, 2015; Wang *et al.*, 2019; Woodhouse *et al.*, 2014), meta-analysis (Hawkins *et al.*, 2015; Jia *et al.*, 2018; Zhang *et al.*, 2018), gene and gene family identification (Li *et al.*, 2019; Liu *et al.*, 2019, 2017), quantitative trait and association studies (Liu *et al.*, 2012; Wu *et al.*, 2016b), orthology (Sindhu *et al.*, 2018) and general reviews (Liu *et al.*, 2020; Wang *et al.*, 2016). qTeller's breadth of use demonstrates its value to the research community.

In 2018, the Maize Genetics and Genomics Database (MaizeGDB) (Portwood *et al.*, 2019) released its own version of qTeller for maize (https://qteller.maizegdb.org/). Since then, MaizeGDB has optimized qTeller to host information on protein abundance as well as mRNA abundance (i.e. RNA-Seq data), and to allow comparisons across gene models annotated in different reference genomes. By offering this tool as a web page to the maize community, MaizeGDB helps researchers to quickly compare pre-computed gene expression abundances across maize genes and genomes. Here, we present a description of qTeller and its functionality, and how users can download and run the software themselves. The MaizeGDB qTeller is available at https://github.com/Maize-Genetics-and-Genomics-Database/qTeller.

2 Materials and methods

2.1 qTeller basic functionality

There are three main sections of qTeller: Section 2.1.1, Section 2.1.2 and Section 2.1.3. Each section is accessed through a drop-down menu on the web page and presents gene expression information in a different way to meet the needs of users with distinct use cases. Users may investigate expression within one genome, across multiple genomes or compare RNA expression to protein abundances. The three sections are described below.

2.1.1 Genes in an interval

Section 2.1.1 allows users to select a chromosome and coordinate interval of interest for a given genome (only one genome at a time can be selected) (Fig. 2). The primary use case for this section of qTeller is when a researcher has mapped a gene or QTL to a defined interval in the genome and wishes to use gene expression data within this interval to prioritize among the potential candidate genes within the interval. The interface can be set up for a single reference genome or to have a selectable list of genomes from a set. Next, the user selects the RNA-Seq libraries of interest below the genomic coordinate selection, selects all RNA-Seq libraries under a given set, or selects all RNA-Seq libraries in the database. qTeller then returns the RNA-Seq abundances of all the selected libraries of all the genes within the selected interval for that given genome. A user has the option of selecting 'All Chromosomes' in the dropdown menu and leaving the coordinate boxes blank to return the RNA-Seq abundances for all genes in the genome (excluding unplaced scaffolds). The output is in the format of a table that includes gene model ID, RNA-Seq abundances for each selected library, and a link to visualize the data as a bar chart for every gene model (see Section 2.1.3). A user has the option of either viewing the table as a web page or downloading the table as a .csv file. genome start and end position, and check boxes for each of the RNA-Seq experiments (organized by project or paper). Each set of RNA-Seq experiments has an 'All on' and 'All off' option.

2.1.2 Genes by name

The Section 2.1.2 is similar to Section 2.1.1 except that it allows a user to paste a list of gene models of interest instead of selecting genomic coordinates (Fig. 3). The use cases for this section of qTeller include users with a set of genes of interest identified via other means (e.g. a set of GWAS hits or a cluster of genes linked by protein interaction data). For a multi-genome instance, a mix of gene models across different genomes is permitted, e.g. allowing users to compare expression from gene models across multiple genomes of a pan-genome. Library selection and output tables are the same as for Section 2.1.1.

2.1.3 Visualize expression

The Section 2.1.3 tool draws a bar chart for all libraries for a given input gene model, or draws a dot plot of all libraries between two-gene model inputs (Fig. 4). The latter case is useful for comparing relative expression between two gene models. The dot plot feature for the multi-genome qTeller instance allows a user to input two gene models from any genome. The use cases for this section of qTeller include comparisons of duplicated genes to identify potential evidence of regulatory subfunctionalization, or comparison of patterns of expression of equivalent gene models between different genetic backgrounds/genomics to identify genotype-specific patterns of regulation as the result of *cis*- or *trans*-regulatory divergence. Advanced options for both the bar chart and dot plot allow users to select their libraries of interest instead of visualizing all libraries. Under each visualization output, qTeller generates a shareable link to recreate the bar chart or dot plot images if a user wants to share the data with others or use it in a publication. A user can mouse over the bars in the bar chart, or the dots in the dot plot, to get information about the abundances and how the data were generated experimentally. A user can also change the axes of the dot plot to zoom in on a region of interest for better resolution.

2.2 qTeller expanded functionality

2.2.1 Protein expression visualization

Gene expression is the measurement of how genes produce functional products used to carry out processes in a cell. There are two primary ways to measure gene expression: mRNA abundance using RNA-Seq, and protein abundance using mass spectrometry [reviewed in Zhang *et al.* (2010)]. Gene expression at the mRNA abundance level can be only poorly predicted using data on gene expression at the protein abundance level and vice versa, and both types of data can be used to associate genes with functional characteristics. The functionality of qTeller was expanded to include protein expression as well as RNA-Seq data.

The 'Compare RNA & Protein' tool draws four different types of visualization, a bar chart and three different dot plots. The 'Single-Gene Expression' tool under 'Visualize RNA versus Protein' is similar to the single-genome Section 2.1.3 bar chart, except that the user can select either mRNA abundance (FPKM) or protein abundance (NSAF) for all selected libraries for a single input gene model. The first dot plot (Two-Gene Scatterplot, Fig. 5A) compares two gene model inputs using the same data type, either mRNA versus mRNA or protein versus protein. This dot plot is useful for comparing relative mRNA or protein abundance between two gene models. The dot plot feature for the multi-genome qTeller instance allows a user to input two gene models from any genome as long as the expression data from both genomes were generated by the same project with a consistent methodology. The second dot plot ('Single-

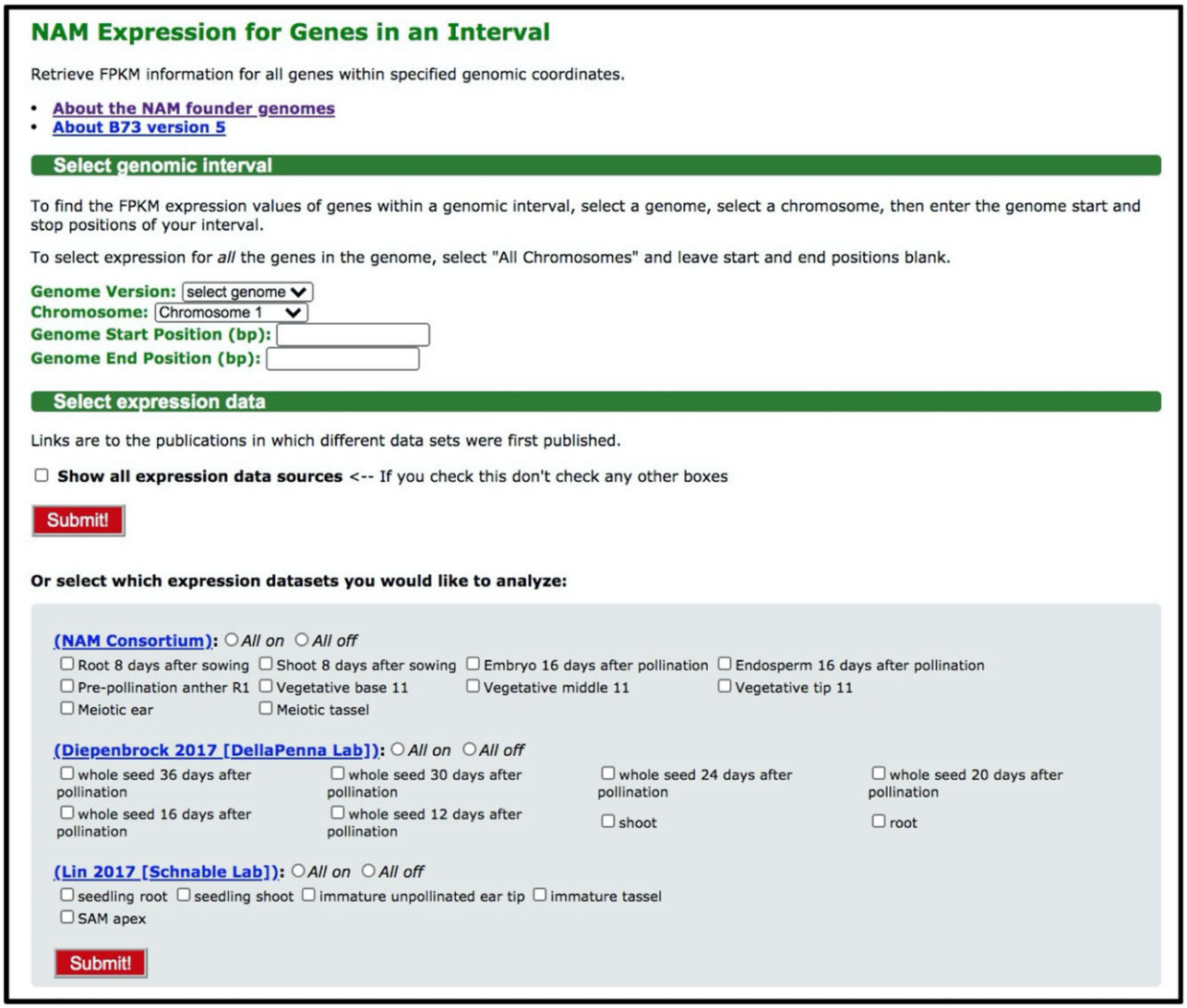

Fig. 2. 'Genes in an Interval' tool. The screenshot is from MaizeGDB's qTeller instance for the 'Genes in an Interval' tool for a set of maize genomes (NAM founders). The input form has a drop-down menu for the genomes and chromosomes (including an 'All Chromosomes' option), text boxes for the genome start and end position, and check boxes for each of the RNA-Seq experiments (organized by project or paper). Each set of RNA-Seq experiments has an 'All on' and 'All off' option

Gene Expression versus Abundance') is used to make a comparison between mRNA abundance and protein abundance for the single input gene model across different tissues. The third dot plot ('Multi-Gene Expression versus Abundance in a Single Tissue', Fig. 5B) is similar to the second dot plot except that it allows a user to select the tissue of interest and enter a list of gene models. This dot plot makes a direct comparison between mRNA and protein abundance for a fixed tissue and set of gene models. The latter two plots also provide a Pearson correlation coefficient that measures linear correlation between two variables and abundance types.

2.2.2 Multi-genome functionality

qTeller now offers a multi-genomic option when building and calling a database; this feature is useful for genomes and/or RNA-Seq datasets that were constructed using the same methods across genome assemblies (e.g. NAM founders in maize, doi:10.1101/2021.01.14.426684). This functionality is specific for multiple genomes within the same species, requiring that all genomes have the same number of chromosomes with the same chromosome ID designation (i.e. chr1, chr2, etc. or 1, 2, etc.). The main technical difference between qTeller and multi-genome qTeller is that an input bed file is required which contains information about each genome ID. The bed file follows the typical structure of a normal bed file, with the gene model ID in Column 4 and the ID of the genome in Column 5 (see Supplementary Table S1). The genome ID can be any alphanumeric string. In order for experiments to be treated as paired data, and thus appropriate for between-genome dot plots and comparisons in multi-genome qTeller, they must be assigned exactly matching values in the 'data_id' column. For instance, if SRR12345 for Genome A is described as 'pollen tube' in the 'data_id' metadata column, then if Genome B's SRR23456 from the same experiment is also from pollen tube tissue, its 'data_id' must also be written as 'pollen tube' exactly, and have the same experiment Source, if the user wishes these experiments to be fetched together.

The biggest difference in the qTeller interface structure for multi-genome is under Section 2.1.1, where users can select a genome of their choice from the drop-down menu at the top. This reflects a change in the database structure wherein each gene model in the multi-genome database is assigned a genome ID (see Software Usage below). Because Section 2.1.1 is based on a genomic coordinate system, more than one genome cannot be fetched at a time. However, under Section 2.1.2 or Section 2.1.3, gene model IDs from more than one genome can be fetched. Ideally, cross-genome RNA-Seq datasets should be matched with identical descriptions only when the RNA samples used for quantification were collected, processed, and sequenced by the same laboratory, to ensure that any differences

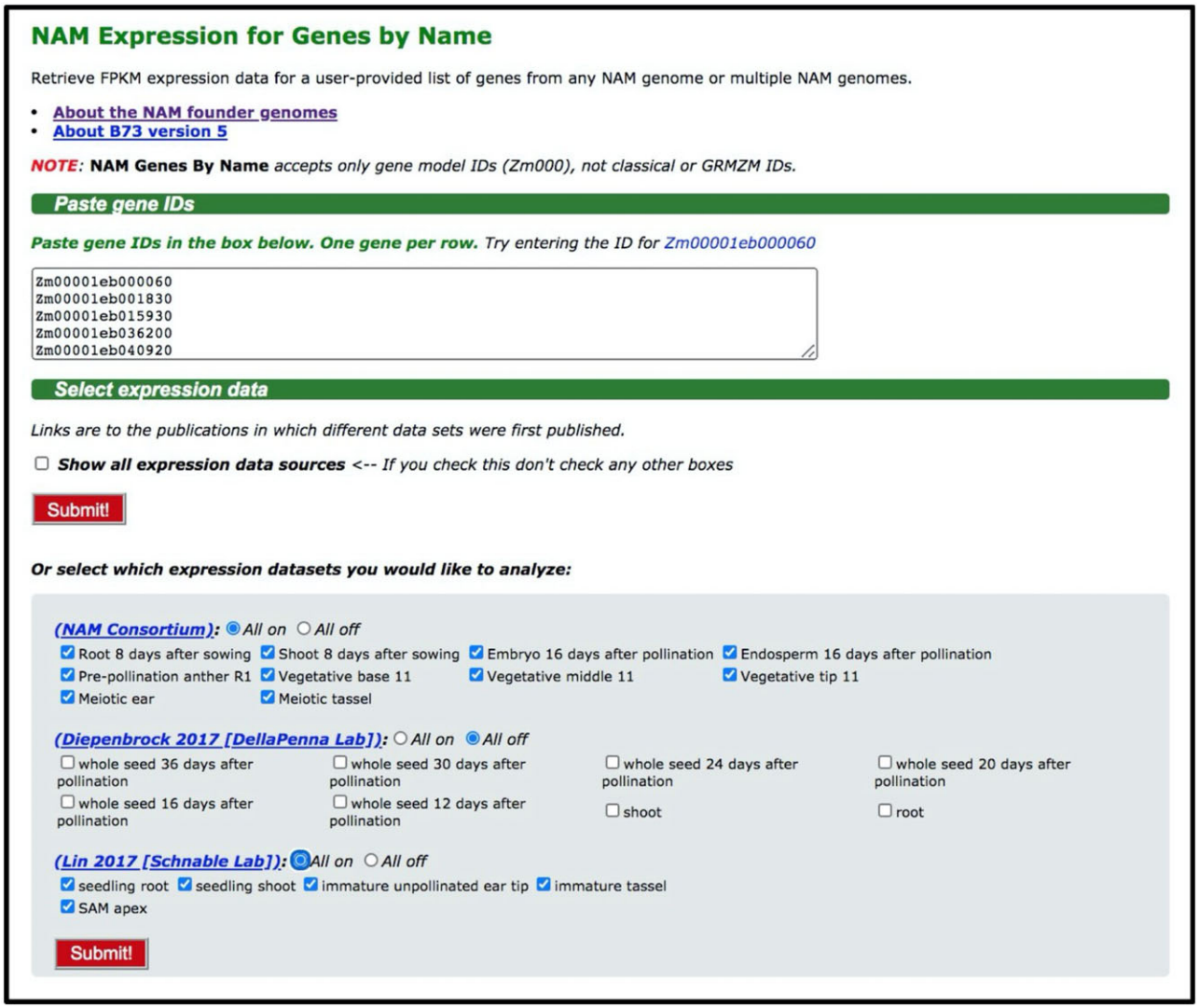

Fig 3. 'Genes by Name' tool. The screenshot is from MaizeGDB's qTeller instance for the 'Genes by Name' tool for a set maize genomes (NAM founders). The input form takes as input a text box for a list of gene model identifiers and check boxes for each of the RNA-Seq experiments (organized by project or paper). Each set of RNA-Seq experiments has an 'All on' and 'All off' option

observed in relative abundances between genomes are not due to differences in laboratory technique or environment.

2.2.3 Expanded qTeller navigation

The expanded qTeller software package includes a reformatted homepage and a menu header on each page for quick access to each of the four tools or links to general information (contact, data sources and FAQs). Each tool menu item has a submenu listing which genomes are available. There is also a new 'News' item feature on the side of the page.

2.3 Basic qTeller software usage

qTeller was originally written in Python 2.7 (http://www.python.org), html and PHP5 for SQlite3 (https://www.sqlite.org/) software. We updated the Python code to Python 3 and ensured that the PHP scripts were PHP7 compatible. Images are drawn using Python Matplotlib (Hunter, 2007). Python3 dependencies are listed in qteller_packagelist_python3.txt in our GitHub and can be installed using Python3 pip.

The most basic form of qTeller requires only three inputs: a gff or bed file of the gene models of interest; a directory containing files with RNA-Seq and/or protein abundances by experiment; and a metadata file structured as described in Supplementary Table S2. There are two main directories: the build_db directory, where the database is built; and the web_interface directory, which contains the dynamic web pages.

qTeller's basic structure allows for most types of RNA-Seq mapping pipelines to be used, since qTeller accepts fpkm abundances calculated by Cufflinks (genes.fpkm_mapping outputs) from genomic mapping pipelines such as GSNAP (Wu *et al.*, 2016a), STAR (Dobin *et al.*, 2013) and TopHat (Kim *et al.*, 2013) or TPM abundances calculated by transcript RNA-Seq mapping programs such as Salmon (Patro *et al.*, 2017) or Kallisto (Bray *et al.*, 2016). However, qTeller's structure is based on gene models, not transcripts; therefore, if a user has Salmon/Kallisto output files that quantify expression at the per-transcript level, it will first be necessary to calculate an aggregate gene-level abundance, whether via averaging or another process, and the resulting gene-level data constructed as .txt file where Column 1 is the Gene ID, and Column 2 is the averaged TPM abundance (see Supplementary Table S3). The qTeller build.py scripts will automatically detect whether the directory containing the RNA-Seq abundances have the .txt or .fpkm_tracking extension, and proceed accordingly. The two-column, gene/abundance .txt file configuration will also work for abundances calculated through EdgeR or some other method. It is important to emphasize that all

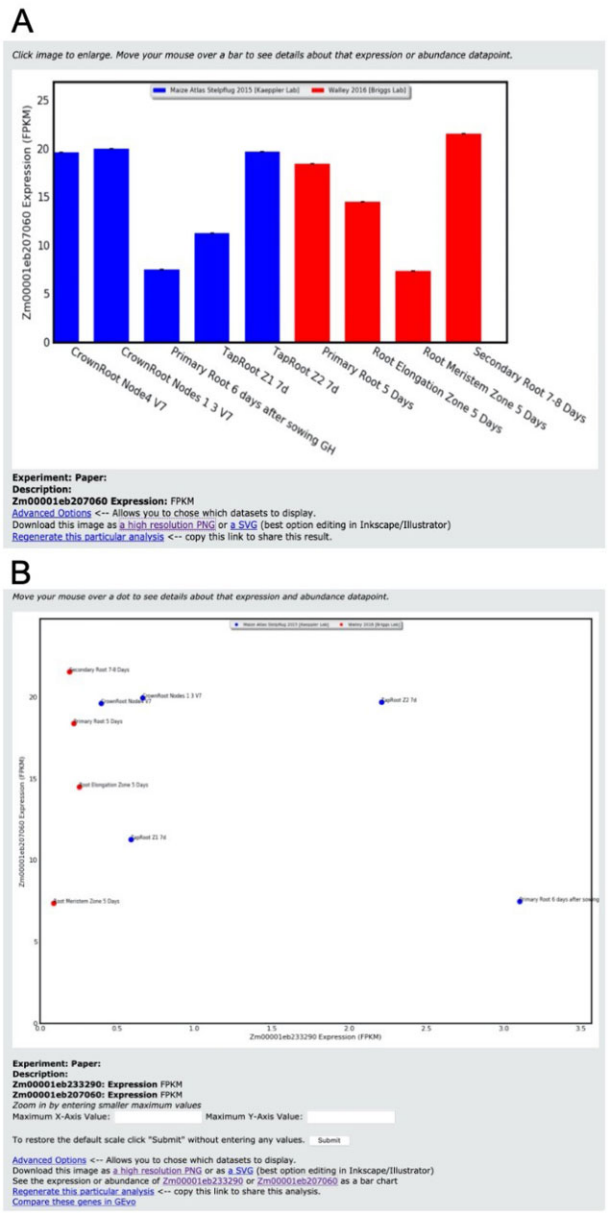

Fig 4. 'Visualize Expression' tool. The screenshot is from MaizeGDB's qTeller instance for the 'Visualize Expression' tool for genes in the B73 v5 maize genome. (**A**) The output of the B73 gene Zm00001eb207060 for a subset of root tissue from the 'Single-Gene Expression' tool that creates a bar chart showing the mRNA abundance from selected RNA-Seq experiments. (**B**) The output from the 'Two-Gene Scatterplot' tool which displays a scatter plot comparing the expression for two genes. Zm00001eb207060 from the bar chart image is compared to its retained homeolog Zm00001eb233290. Notice that Zm00001eb207060 is expressed consistently higher in root tissue than Zm00001eb233290

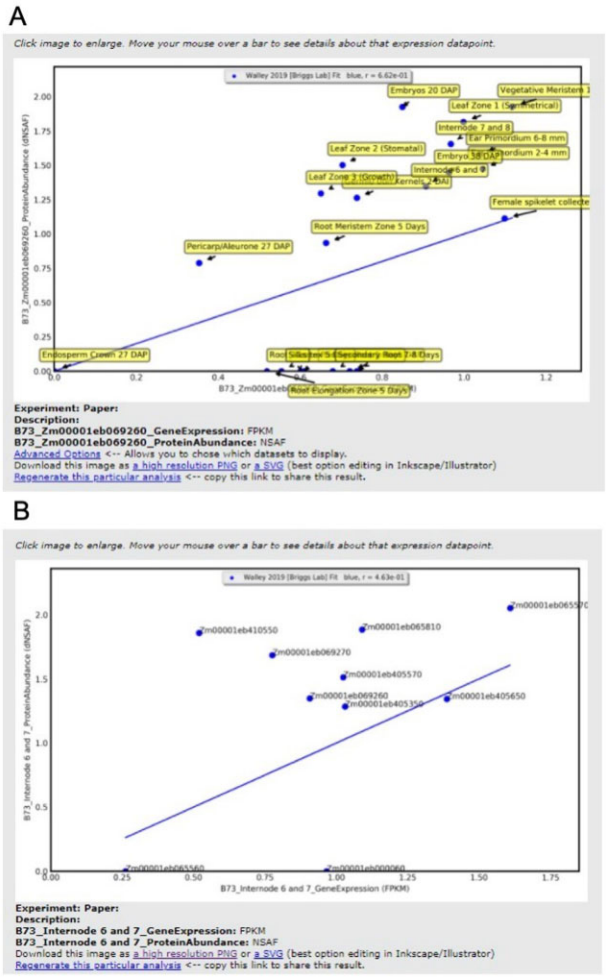

Fig 5. 'Compare RNA and Protein' tool. The screenshot is from MaizeGDB's qTeller instance for the 'Compare RNA and Protein' tool for genes in the B73 v5 maize genome. (**A**) The output from the 'Single-Gene Expression versus Abundance' tool that creates a scatter plot comparing mRNA expression and protein abundance from selected RNA-Seq experiments. (**B**) The output from the 'Multi-Gene Expression versus Abundance in a single tissue' tool which displays a scatter plot comparing mRNA expression and protein abundance for the selected tissue and set of gene models

inputs combined in a single qTeller instance should be mapped using the same pipeline, and use the exact same method of counting abundances (either FPKM or TPM or some other method). The combination of datasets generated using different quantification pipelines will generate many apparent differences in expression, resulting from technical differences in quantification rather than biological differences in expression.

The .txt file or .fpkm_mapping outputs must be preprocessed to replace empty abundances as 'Nan' before running them as input files in the qTeller build.py scripts. This preprocessing of the input files enables qTeller to differentiate between 0 abundance value and null abundance value. This preprocessing of the .txt file or .fpkm_mapping outputs can be either done by the user using Excel, or by the qTeller script metadata_NA_to_Nan.py available in our GitHub. One can also customize the code as per the user input file structure.

qTeller will accept multiple biological replicates. However, if a user has a large number of libraries to work with, it is advised to average the biological replicate abundances by gene, and create a text file as described above, for the RNA input; otherwise, visualization of datasets can become crowded and difficult to resolve visually. This method of averaging biological replicates was used to construct the current MaizeGDB qTeller instances.

Because .gff file structures can vary in Column 9 in terms of whether genes are prefaced by 'ID=' or some other scheme, qTeller allows the user to indicate which identifier is used in the gene model input .gff file by selecting –gff and then –gene_def_tag and typing the identifier afterward. If this option is not selected, qTeller will assume the identifier is 'ID'.

Metadata files need to be organized exactly as the example given in Supplementary Table S2. The qTeller web pages are organized based on the project from which the RNA-Seq data was collected, processed and sequenced (i.e. 'Source' in the metadata files.) This ensures that all the data within a collection has been extracted and sequenced the same way, so as to avoid the issue of artifactual relative abundances due to differences in laboratory technique or

environment. Note that abundances in similar tissues across different laboratories (i.e. leaf) may differ somewhat due to differences in laboratory handling and other factors. Metadata is dynamically organized on each web page based on the contents of the database.

qTeller builds a SQLite3 database from the gene model, RNA (and protein) abundance and metadata information. This database is then the source of all data called by qTeller in the web pages for Section 2.1.1, Section 2.1.2 and Section 2.1.3.

Certain files are hard-coded for drop-down menu information, and must be manually changed by the user. For instance, in the index_*.php files (index_singlegenome.php, index_multigenome.php and Protein_index.php) that correspond to the Section 2.1.1 pages, the chromosome selection drop-down menu must be edited to reflect the number of chromosomes in the target genome(s), and the name of the chromosomes in the target genome. For instance, maize has 10 chromosomes, designated chr1, chr2, etc., whereas Sorghum's chromosomes are designated Chr01, Chr02, etc. in the current Sorghum reference genome release, and *Arabidopsis thaliana* has only five chromosomes; the drop-down menu in index_*.php will need to be edited to reflect the target genome's configuration (see Supplementary Fig. S1). Also, in the index_multigenome.php file for multiple genomes, the drop-down menu for genome selection will need to be manually changed to reflect the genomes used, based on their designation in Column 5 of the gene model bed file input. The default index_multigenome.php in our GitHub download is configured for the multi-genome test data (see below).

Example databases of a subset of incomplete maize data for single-genome, multi-genome and protein data, as well as example metadata, bed and fpkm files to generate these databases, are included in the build_db directory.

3 Results

3.1 Use case: MaizeGDB qTeller

There are several features in MaizeGDB's qTeller instance that are uniquely specified for maize, including the maize genomes, maize datasets and MaizeGDB-specific metadata and other information. The homepage for the MaizeGDB qTeller website (https://qteller.maizegdb.org/) has a general description of qTeller, news items, quick links for each of the four tools, two sections for getting started and frequently asked questions (FAQs) and a Contact page linked to a local JIRA (https://www.atlassian.com/software/jira) instance to track errors or issues. The datasets used in MaizeGDB qTeller are described below.

3.2 Datasets

3.2.1 Maize genomes

MaizeGDB currently hosts three instances of qTeller: RNA-Seq and protein abundance data for the latest two versions (versions 4 and 5) of the reference maize genome B73, and RNA-Seq data for the NAM founder genomes, consisting of the genomes of 26 diverse maize inbred lines (doi:10.1101/2021.01.14.426684, https://nam-genomes.org/).

The well-known and most used public founder maize variety B73 was sequenced in 2009 (Schnable *et al.*, 2009). For nearly a decade, B73 was the reference genome for the maize research community, and most of the genomic tools, resources and datasets at MaizeGDB were oriented around this single reference. MaizeGDB's 2018 release of qTeller centered around version 4 of the B73 genome (RefGen_v4) released in 2017 (Jiao *et al.*, 2017). As sequencing technology became more affordable, additional maize reference-quality genomes were sequenced (Hirsch *et al.*, 2016; Springer *et al.*, 2018; Sun *et al.*, 2018). While these genome assemblies had great potential to advance maize research, the underlying assemblies and supporting datasets (e.g. RNA-Seq) were generated with different methodologies and conditions.

In 2020–2021, the NAM Sequencing Consortium (doi:10.1101/2021.01.14.426684, https://nam-genomes.org/) released the first set of maize genomes sequenced, assembled and annotated in a consistent way. The NAM Sequencing Consortium's data release included a new version of B73 (RefGen_v5) and the 25 founder lines of the Nested Association Mapping (NAM) population, which has been used extensively by maize and other researchers to study maize flowering time (Buckler *et al.*, 2009), leaf architecture (Tian *et al.*, 2011), disease resistance (Poland *et al.*, 2011) and other important agronomic traits (Wallace *et al.*, 2014). These assemblies presented the opportunity for constructing pan-genomes and identifying pan-gene sets (genes conserved across the varieties), as well as making it possible to develop pan-genome tools. The multi-genome version of qTeller at MaizeGDB includes these 25 NAM founder lines and B73 RefGen_v5. The MaizeGDB project currently supports 44 maize genomes that could be included into qTeller as additional multi-genome gene expression datasets become available.

3.2.2 Maize gene expression

The MaizeGDB qTeller has over 200 unique datasets from 12 projects available at MaizeGDB. The B73 version 4 instance of qTeller has RNA-Seq data from six studies (Forestan *et al.*, 2016; Johnston *et al.*, 2014; Kakumanu *et al.*, 2012; Stelpflug *et al.*, 2016; Walley *et al.*, 2016; Waters *et al.*, 2017) covering 158 tissues/conditions. The B73 version 5 instance has data from eight studies (Forestan *et al.*, 2016; Johnston *et al.*, 2014; Kakumanu *et al.*, 2012; Makarevitch *et al.*, 2015; Opitz *et al.*, 2014; Stelpflug *et al.*, 2016; Walley *et al.*, 2016; Warman *et al.*, 2020) covering 172 tissues/conditions. The multi-genome set of NAM founders has three studies with 23 tissues/conditions (10.1101/2021.01.14.426684, 10.1105/tpc.17.00475, 10.1186/s13059-017-1328-6). The 'Compare RNA & Protein' tool has data from one mRNA and protein study (Walley *et al.*, 2016) for 23 tissues/conditions; this dataset is currently the only large-scale gene expression atlas that provides both RNA-Seq and protein abundance data.

3.3 Data processing

All of the RNA-Seq datasets for MaizeGDB qTeller were mapped with a consistent methodology. Fastq files were mapped to either Ensembl AGPv4 B73, Ensembl AGPv5 B73 or the NAM founder genomes using STAR, and abundances calculated using Cufflinks. The protein abundance data was projected to both AGPv4 B73 and AGPv5 B73 based on gene synteny [see methods in Walsh *et al.* (2020)].

4 Conclusion

qTeller was developed to address the need for an accessible tool to organize, integrate, access, compare and visualize gene expression data. Though unpublished, the tool has been used and cited broadly by the plant research community in the study of evolution, meta-analyses, gene and gene family identification, quantitative trait and association studies and ontology. MaizeGDB expanded qTeller's functionality to include multiple genomes and protein abundance data, and enhanced the website layout to make qTeller even easier to use. qTeller was designed for plant species, but is broadly extendable to any species with a sequenced genome and RNA-Seq or protein abundance data.

Funding

This research was supported by the US. Department of Agriculture, Agricultural Research Service, Project Number [5030-21000-068-00-D] through the Corn Insects and Crop Genetics Research Unit in Ames, Iowa. This material is based upon work supported by the Department of Agriculture, Agricultural Research Service under Agreement No. 58-5030-0-036 [Iowa State Award: 022172-00001 to J.W.W.]. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and Employer.

Conflict of Interest: none declared.

References

Bray,N.L. *et al.* (2016) Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.*, **34**, 525–527.

Buckler,E.S. *et al.* (2009) The genetic architecture of maize flowering time. *Science*, **325**, 714–718.

Dobin,A. *et al.* (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

Forestan,C. *et al.* (2016) Stress-induced and epigenetic-mediated maize transcriptome regulation study by means of transcriptome reannotation and differential expression analysis. *Sci. Rep.*, **6**, 30446.

Hawkins,L.K. *et al.* (2015) Characterization of the maize chitinase genes and their effect on *Aspergillus flavus* and *Aflatoxin Accumulation* resistance. *PLoS One*, **10**, e0126185.

Hirsch,C.N. *et al.* (2016) Draft assembly of elite inbred line PH207 provides insights into genomic and transcriptome diversity in maize. *Plant Cell*, **28**, 2700–2714.

Hunter,J.D. (2007) Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.*, **9**, 90–95.

Jia,H. *et al.* (2018) Integrated analysis of protein abundance, transcript level, and tissue diversity to reveal developmental regulation of maize. *J. Proteome Res.*, **17**, 822–833.

Jiao,Y. *et al.* (2017) Improved maize reference genome with single-molecule technologies. *Nature*, **546**, 524–527.

Johnston,R. *et al.* (2014) Transcriptomic analyses indicate that maize ligule development recapitulates gene expression patterns that occur during lateral organ initiation. *Plant Cell*, **26**, 4718–4732.

Kakumanu,A. *et al.* (2012) Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. *Plant Physiol.*, **160**, 846–867.

Kim,D. *et al.* (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.*, **14**, R36.

Langfelder,P. and Horvath,S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.

Liu,J. *et al.* (2020) The past, present, and future of maize improvement: domestication, Genomics, and Functional Genomic Routes toward Crop Enhancement. *Plant Commun.*, **1**, 100010.

Liu,R. *et al.* (2012) Fine mapping and candidate gene prediction of a pleiotropic quantitative trait locus for yield-related trait in Zea mays. *PLoS One*, **7**, e49836.

Liu,Y. *et al.* (2017) Cloning, molecular evolution and functional characterization of ZmbHLH16, the maize ortholog of OsTIP2 (OsbHLH142). *Biol. Open*, **6**, 1654–1663.

Liu,Y. *et al.* (2019) Identification and characterization of the TCA cycle genes in maize. *BMC Plant Biol.*, **19**, 592.

Li,Y. *et al.* (2019) Genome-wide identification, phylogenetic and expression analysis of the maize HECT E3 ubiquitin ligase genes. *Genetica*, **147**, 391–400.

Makarevitch,I. *et al.* (2015) Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet.*, **11**, e1004915.

Man,J. *et al.* (2020) Structural evolution drives diversification of the large LRR-RLK gene family. *N. Phytol.*, **226**, 1492–1505.

Opitz,N. *et al.* (2014) Transcriptomic complexity in young maize primary roots in response to low water potentials. *BMC Genomics*, **15**, 741.

Patro,R. *et al.* (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*, **14**, 417–419.

Poland,J.A. *et al.* (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proc. Natl. Acad. Sci. USA*, **108**, 6893–6898.

Pophaly,S.D. and Tellier,A. (2015) Population level purifying selection and gene expression shape subgenome evolution in maize. *Mol. Biol. Evol.*, **32**, 3226–3235.

Portwood,J.L., 2nd. *et al.* (2019) MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucleic Acids Res.*, **47**, D1146–D1154.

Price,A. *et al.* (2019) DEvis: an R package for aggregation and visualization of differential expression data. *BMC Bioinformatics*, **20**, 110.

Robinson,M.D. *et al.* (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Schnable,P.S. *et al.* (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112–1115.

Sindhu,A. *et al.* (2018) Propagation of cell death in dropdead1, a sorghum ortholog of the maize lls1 mutant. *PLoS One*, **13**, e0201359.

Springer,N.M. *et al.* (2018) The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nat. Genet.*, **50**, 1282–1288.

Stelpflug,S.C. *et al.* (2016) An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. *Plant Genome*, **9**, **1–16**.

Sun,S. *et al.* (2018) Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat. Genet.*, **50**, 1289–1295.

Tian,F. *et al.* (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat. Genet.*, **43**, 159–162.

Wallace,J.G. *et al.* (2014) Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genet.*, **10**, e1004845.

Walley,J.W. *et al.* (2016) Integration of omic networks in a developmental atlas of maize. *Science*, **353**, 814–818.

Walsh,J.R. *et al.* (2020) Tissue-specific gene expression and protein abundance patterns are associated with fractionation bias in maize. *BMC Plant Biol.*, **20**, 4.

Wang,Y. *et al.* (2016) Bioinformatic landscapes for plant transcription factor system research. *Planta*, **243**, 297–304.

Wang,Y. *et al.* (2019) Comparative genomics revealed the gene evolution and functional divergence of magnesium transporter families in Saccharum. *BMC Genomics*, **20**, 83.

Wang,Z. *et al.* (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **10**, 57–63.

Warman,C. *et al.* (2020) High expression in maize pollen correlates with genetic contributions to pollen fitness as well as with coordinated transcription from neighboring transposable elements. *PLoS Genet.*, **16**, e1008462.

Waters,A.J. *et al.* (2017) Natural variation for gene expression responses to abiotic stress in maize. *Plant J.*, **89**, 706–717.

Wilkinson,L. (2011) ggplot2: elegant graphics for data analysis by Wickham, H. *Biometrics*, **67**, 678–679.

Woodhouse,M.R. *et al.* (2014) Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. *Proc. Natl. Acad. Sci. USA*, **111**, 5283–5288.

Wu,T.D. *et al.* (2016a) GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality. *Methods Mol. Biol.*, **1418**, 283–334.

Wu,X. *et al.* (2016b) Exploring identity-by-descent segments and putative functions using different foundation parents in maize. *PLoS One*, **11**, e0168374.

Zhang,G. *et al.* (2010) Protein quantitation using mass spectrometry. *Methods Mol. Biol.*, **673**, 211–222.

Zhang,X. *et al.* (2018) Effects of drought stress and water recovery on physiological responses and gene expression in maize seedlings. *BMC Plant Biol.*, **18**, 68.