

RESEARCH

Open Access



BarleyExpDB: an integrative gene expression database for barley

Tingting Li^{1,2†}, Yihan Li^{1†}, Hongbin Shangguan¹, Jianxin Bian³, Ruihan Luo¹, Yuan Tian⁴, Zhimin Li¹, Xiaojun Nie^{2*} and Licao Cui^{1*}

Abstract

Background RNA-sequencing (RNA-seq) has been widely used to study the dynamic expression patterns of transcribed genes, which can lead to new biological insights. However, processing and analyzing these huge amounts of histological data remains a great challenge for wet labs and field researchers who lack bioinformatics experience and computational resources.

Results We present BarleyExpDB, an easy-to-operate, free, and web-accessible database that integrates transcriptional profiles of barley at different growth and developmental stages, tissues, and stress conditions, as well as differential expression of mutants and populations to build a platform for barley expression and visualization. The expression of a gene of interest can be easily queried by searching by known gene ID or sequence similarity. Expression data can be displayed as a heat map, along with functional descriptions as well as Gene Ontology, Kyoto Encyclopedia of Genes and Genomes, Proteins Families Database, and Simple Modular Architecture Research Tool annotations.

Conclusions BarleyExpDB will serve as a valuable resource for the barley research community to leverage the vast publicly available RNA-seq datasets for functional genomics research and crop molecular breeding.

Keywords Barley, Expression, RNA-sequencing, Database, BarleyExpDB

Background

Over the last decade, RNA-sequencing (RNA-seq) has surpassed microarray to become the common technique in biological studies [1]. It is a powerful analytical tool for

transcriptional profiling to study gene structures, splicing patterns, and gene/isoform expression levels [2]. The proliferation of the use of next-generation sequencing technologies in the plant research community has led to the accumulation of copious amounts of RNA-seq data in different plant species [3]. By sequencing RNA samples collected from different parts of the plant or plants cultivated under different conditions, such as tissues, developmental stages, and biotic and abiotic stress treatments, the hypotheses of the functions of specific genes/isoforms can be determined [4]. These approaches have been widely used in functional studies aimed at uncovering regulatory mechanisms in major crops such as maize, wheat, and rice [5].

Numerous studies have generated large volumes of raw sequencing data that have been deposited in online repositories such as the Sequence Read Archive (SRA) [6], European Nucleotide Archive (ENA) [7], and

[†]Tingting Li and Yihan Li contributed equally to this work.

*Correspondence:

Xiaojun Nie
small@nwsuaf.edu.cn

Licao Cui
cuilicao@jxau.edu.cn

¹ College of Bioscience and Engineering, Jiangxi Agricultural University, Nanchang 330045, Jiangxi, China

² State Key Laboratory of Crop Stress Biology in Arid Areas and College of Agronomy, Northwest A&F University, Yangling 712100, Shaanxi, China

³ Peking University Institute of Advanced Agricultural Sciences, Weifang 261325, Shandong, China

⁴ Xintai Urban and Rural Development Group Co., Ltd, Taian 271200, Shandong, China



Genome Sequence Archive (GSA) [8]. By 2021, these archives have collectively released a total of ~45,000 libraries for major crops, including wheat, rice, maize, and cotton [9]. Retrospective analyses of the massive amount of RNA-seq data can accelerate functional genomics research and illuminate biological insights [10, 11]. However, these data are available as data archival repositories that store raw sequencing reads, whose access is costly for many academic groups that lack specialized computational resources or dedicated bioinformatics personnel [1].

Efforts have been made to standardize and simplify the access to gene expression data generated by RNA-seq to create a unified resource from fragmented repositories [12]. Recently, comprehensive online databases were established to enhance the utilization of these RNA-seq datasets. For example, the *Arabidopsis* RNA-seq (ARS) database provides a comprehensive platform with integrated, user-friendly, and multifaceted functions for exploring *Arabidopsis* RNA-seq libraries [1]. WheatExp provides free access to a comprehensive array of expression data which allows users to decipher homologue-specific gene expression profiles across a broad range of tissues from different developmental stages in polyploid wheat [13]. BnTIR was established using comprehensive RNA-seq datasets from 91 different tissues spanning *Brassica napus* development [14]. Robinson et al. developed a quick and easy-to-use platform (AgriSeqDB) for RNA-seq data visualization and analysis in *Arabidopsis* and five major crops [5]. Yu et al. constructed the Plant Public RNA-seq Database (PPRD) for viewing, analyzing, and interpreting different mutants, tissues, developmental stages, and stress conditions from several species [9]. These resources have become increasingly valuable exploratory tools for deciphering the complex architecture of the regulatory mechanisms that govern biological processes in diverse species.

As one of the prominent crops since the dawn of early agricultural civilization, cultivated barley (*Hordeum vulgare* L. ssp. *vulgare*) is mainly used for feeding animals, malting, and brewing [15]. Barley is one of the most adaptable plants that can withstand harsh conditions better than its close relative wheat, and it maintains an important role in human nutrition in areas with harsh climates [16]. Wild barley (*H. vulgare* L. ssp. *spontaneum*), domesticated ca. 10,000 years ago in the Fertile Crescent, is the progenitor of cultivated barley and serves as a rich genetic resource for barley improvement [17]. As another important variety of barley, Tibetan hulless barley (*H. vulgare* L. var. *nudum*) is the principal cereal cultivated by Tibetans and a key livestock feed in the Tibetan Plateau [18,

19]. Similar to that in other crops, genomics has been a major driver of genetic and breeding advances in barley over the past decade [15]. The barley genomic assemblies, including the first draft genome and its subsequent revisions (Morex V1, V2, and V3), have undergone multiple rounds of refinement with the emergence of computing algorithms (such as TRITEX workflow) and new sequencing technologies (such as PacBio HiFi, 10 × genomics, chromosome conformation capture sequencing (Hi-C), and biological nano optical mapping) [20–22]. The recently released barley genome and pan-genome expanded the range of natural or induced sequence variation to facilitate genetic analysis and breeding [15, 19, 23]. Benefiting from the release of the genome, copious amounts of RNA-seq-based transcriptome data have been produced and are available for comparative analysis [24–26]. To support sharing and utilizing, researchers have constructed various repositories for RNA-seq datasets of barley [27–29]. However, these existing databases are not well integrated with the currently published RNA-seq datasets, especially in the mining of expression profiles of wild barley and Tibetan hulless barley. In addition, the unsupported reference genome, with the most updated version, is also inconvenient to use. Therefore, a large-scale database with comprehensive RNA-seq datasets that can provide visualized transcriptome expression patterns for barley is greatly needed.

Here, we construct the barley expression database (BarleyExpDB: <http://barleyexp.com/>) (Fig. 1), a web-accessible resource that integrates 56 studies consisting of 3,492 publicly available barley RNA-seq library data deposited at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA). BarleyExpDB supports gene ID queries for four different barley genomes, including cultivated barley (Morex V2 and V3), wild barley (B1K-04-12), and Tibetan hulless barley (Zangqing320), to improve their viability and functionality. In addition, the database provides a user-friendly interface to efficiently visualize, organize, and download the expression profiles of different subspecies/varieties, mutants, stages/tissues, and stress treatments, as well as recombinant inbred lines (RIL) and near isogenic lines (NIL) population. With the rapid growth of barley RNA-seq libraries and the continuous improvement of the reference genome, we plan to regularly update BarleyExpDB in the future. Our approach is designed to provide free, user-friendly, and comprehensive expression data to support researchers in gaining new biological insights and generating new hypotheses in molecular evolution and breeding.

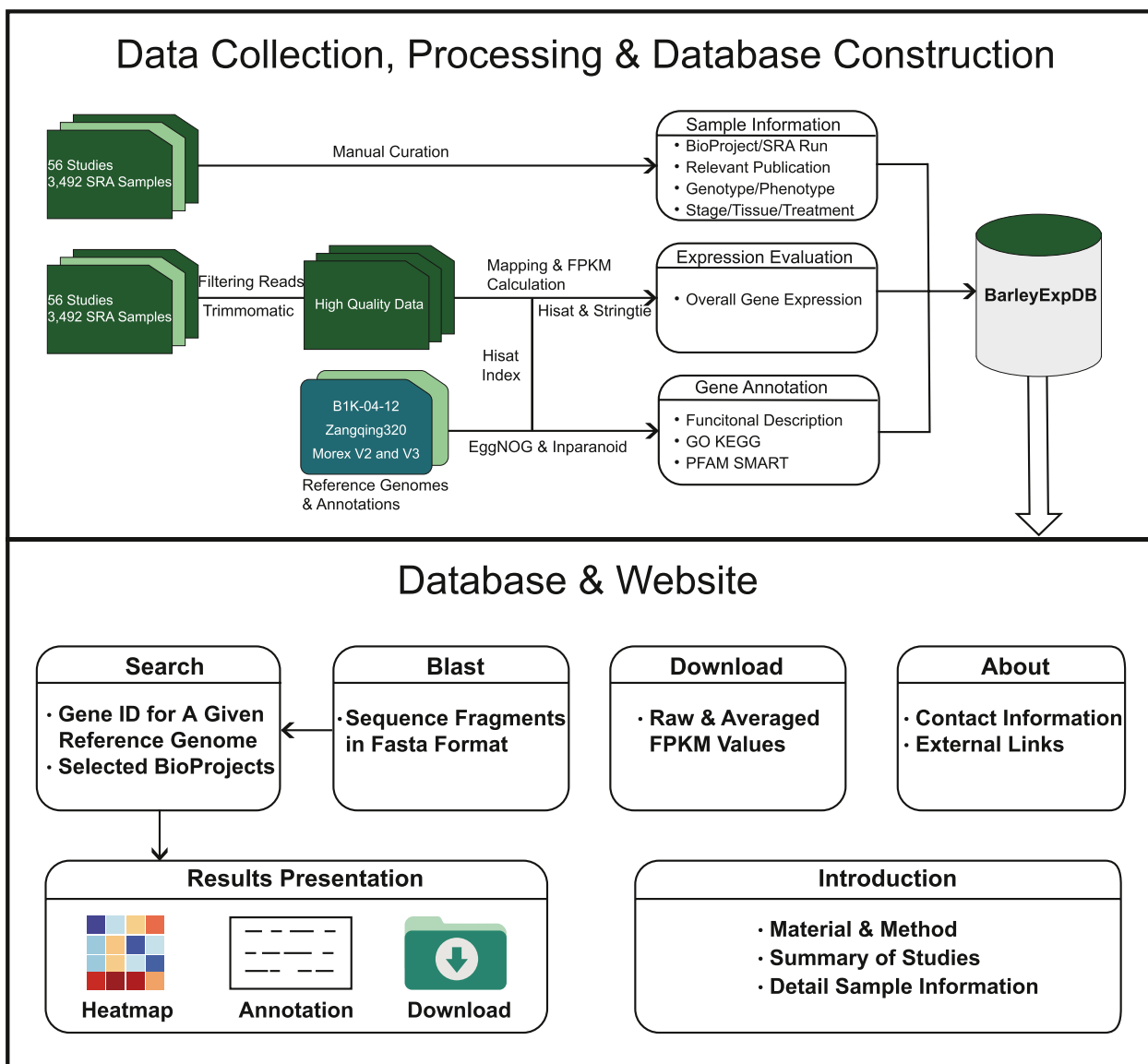


Fig. 1 Workflow of BarleyExpDB

Construction and content

RNA-seq data collection and data processing

A total of 56 studies comprised of 3,492 RNA-seq libraries were included in BarleyExpDB (Supplementary Table S1). These datasets maximally represent the barley expression datasets across multiple subspecies/varieties, developmental stages/tissues, mutants, and stress treatments, as well as RIL and NIL populations. The RNA-seq datasets were downloaded from the NCBI SRA database using the prefetch option in SRAToolkit v2.10.8. The downloaded SRA files were transformed into FASTQ files using the parallel-fastq-dump tool (github.com/rvalieris/parallel-fastq-dump). The quality control of raw

reads was performed using Trimmomatic v0.36 with the following parameter options: Minlen of 90, Trailing of 3, Leading of 3, and SlidingWindow of 4:5.

The barley Morex V2, V3, and B1K-04-12 reference assemblies were downloaded from the IPK database (<https://doi.org/10.5447/ipk/2019/8>, <http://doi.org/10.5447/ipk/2021/3>, and <https://doi.ipk-gatersleben.de/DOI/c4d433dc-bf7c-4ad9-9368-69bb77837ca5/3490162b-3d76-4ba1-b6ee-3eae5f6b644/2>). The reference genome of Zangqing320 was retrieved from WheatOmics (<http://wheatomics.sdau.edu.cn/>). Hisat v2.1.0 was used to build the index for the genomic assembly and to align the RNA-seq reads onto the reference genome.

SAM files were converted into BAM format and then the BAM files were sorted using the 'bs' and 'sort' options of SAMtools v1.3.1. Stringtie v1.3.5 was used for calculating the fragments per kilobase of transcript per million mapped reads (FPKM) values for each gene.

Database implementation and web interface

The web server is hosted on the Tencent Cloud's light-weight application server with four Intel(R) Xeon(R) Platinum 8255C CPUs at 2.50 GHz with 8 GB of RAM and can be freely accessed through its website for non-commercial use. The server uses the Linux CentOS v7.9 operating system (<http://www.centos.org>). The front-end web interface was developed using HTML (<https://www.w3.org/html/>), JavaScript (<https://www.javascript.com/>), and CSS (<http://www.w3.org>). The server-side back-end was implemented by PHP (<https://www.php.net/>). Gene expression matrix storage, maintenance, and operation are supported by MySQL v5.6.50. The custom PHP code was written to enable data searches from MySQL, which were transferred to the front-end.

Development of data mining tools

Functional descriptions, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, Gene Ontology (GO) terms, enzyme commission (EC) numbers, Protein Families Database (PFAM) designations, and Simple Modular Architecture Research Tool (SMART) protein domains were annotated using eggNOG-mapper v2 (<http://egg-nog-mapper.embl.de/>) [30]. The Basic Local Alignment Search Tool (BLAST) was implemented using ViroBlast, a standalone BLAST web server, for sequence homology searches [31]. Orthologs to rice (https://oct2017-plants.ensembl.org/Oryza_sativa/Info/Index) and *Arabidopsis thaliana* (<https://www.arabidopsis.org/>) were identified using Inparanoid v8.0. The interactive heatmap was rendered using Plotly.js (Plotly Technologies Inc., Collaborative Data Science, Montréal, QC, 2015. <https://plotly.com/>) with specified versatile parameters and exported in PNG format.

Utility and discussion

Data quality control and initial mapping statistics

To evaluate the data quality, the sequence quality, GC content, and mapping rate were estimated for each sample. A total of over 12 TB of high-quality clean data were generated using the commonly endorsed criteria of Trimmomatic v0.36. The alignment summary revealed that most reads were aligned concordantly exactly one time to the reference genome, which supported the reliability of the RNA-seq datasets (Fig. 2a-d).

Reproducibility of biological samples

The reproducibility of the gene expression profiles across biological and technical replicates was further used to evaluate the quality of the RNA-seq datasets. The pairwise Pearson correlation coefficient between any two of the samples within the same sample group was calculated based on all genes in the barley genome. Taking the BioProject PRJEB14349 as an example, the mean correlation coefficient value and the standard error per sample group were 0.8069 and 0.1368, respectively, suggesting a high level of sample reproducibility and measurement consistency among replicates (Fig. 2e). Principal component analysis (PCA) confirmed these findings and revealed a strong correlation between replicates of the same stage/tissue, but a significantly lower correlation between samples from different stages/tissues (Fig. 2f).

Comparison with published research

Validation experiments, such as qRT-PCR, were not performed because germplasm resources and sample materials were not available in the corresponding studies. However, existing studies have confirmed the high degree of uniformity among biological replicates of the same stage/treatment as well as the distinctness between stages/treatments by PCA [32–34], hierarchical clustering [34], and multidimensional scaling (MDS) analysis [35]. In addition, qRT-PCR validation of the candidate genes was performed, and the results showed high correlation coefficients with transcriptomic expression levels, confirming the reliability of the transcriptome analysis [25, 32, 34, 36]. Notably, several researchers have performed in-depth experiments, such as overexpression and virus-induced knockdown, to verify the biological functions of candidate genes [25, 37]. Taken together, these findings indicated that the RNA-seq data were valid. We conclude that BarleyExpDB is a valuable resource for the research community to efficiently utilize the vast publicly available RNA-seq datasets, with biological functions and molecular mechanisms to be further investigated by researchers.

Database implementation and practical tools

The BarleyExpDB is publicly accessible through the easy-to-use and intuitive web link <http://barleyexp.com/>. The web interface contains five main sections, namely, Home, BLAST, Introduction, Download, and About (Fig. 3).

Home

The home page of BarleyExpDB provides three main modules for cultivated barley (reference genome: Morex V2 and V3), wild barley (B1K-04-12), and Tibetan hullless barley (Zangqing320). Each module

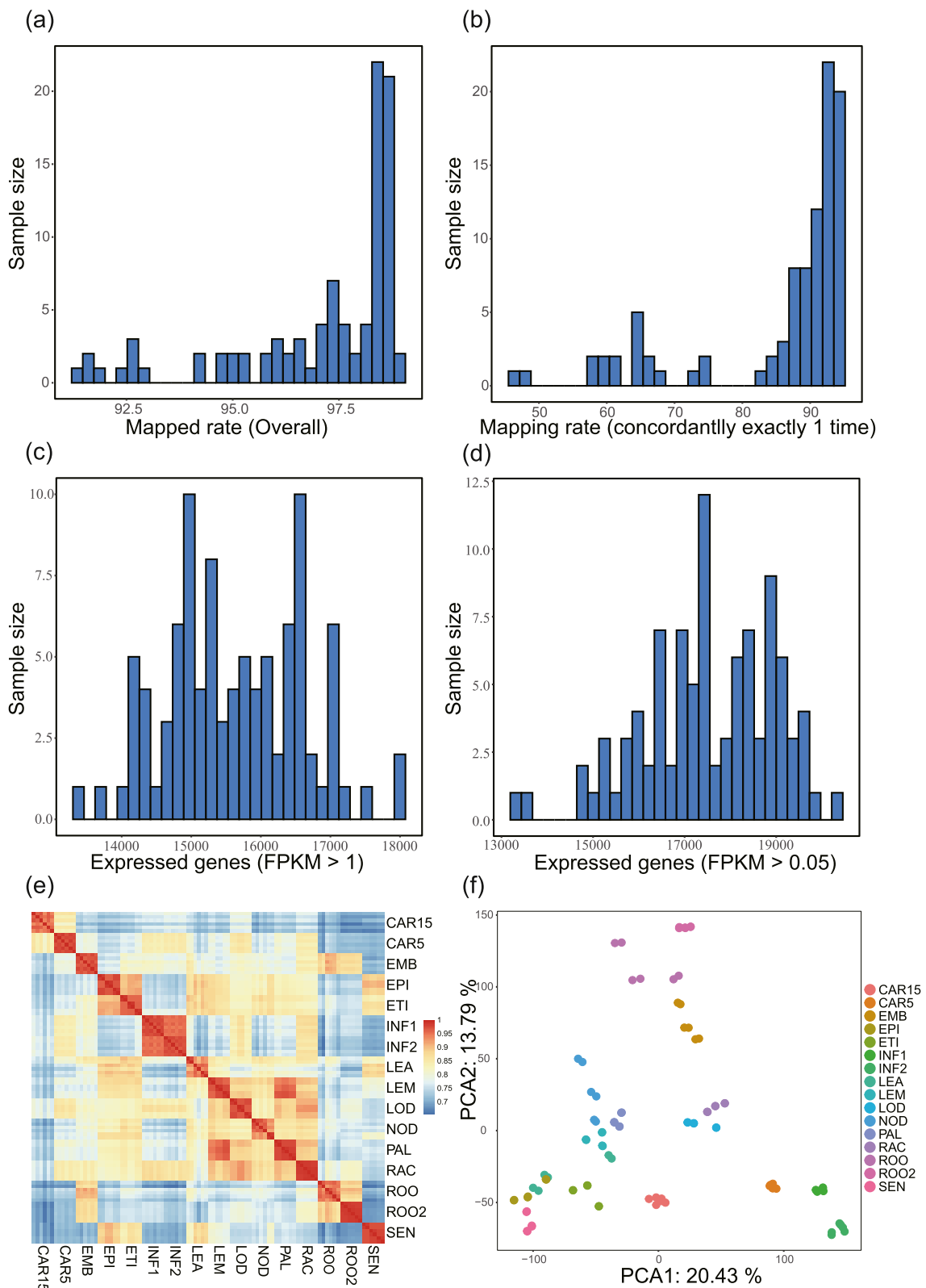
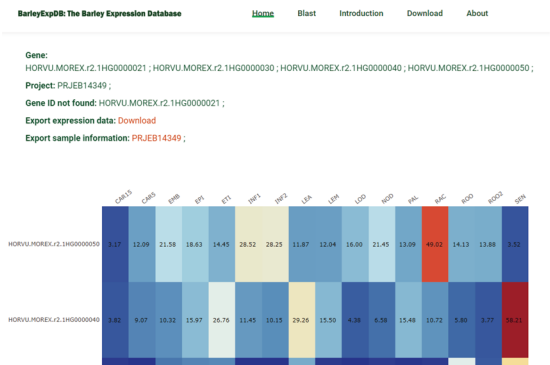
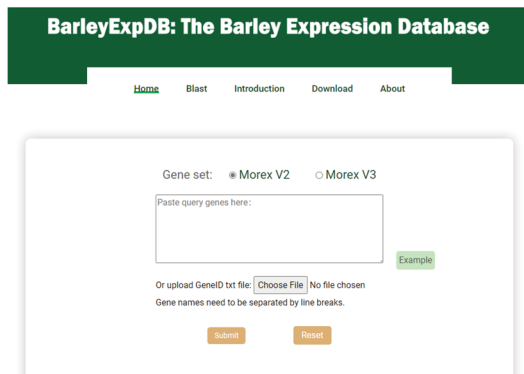
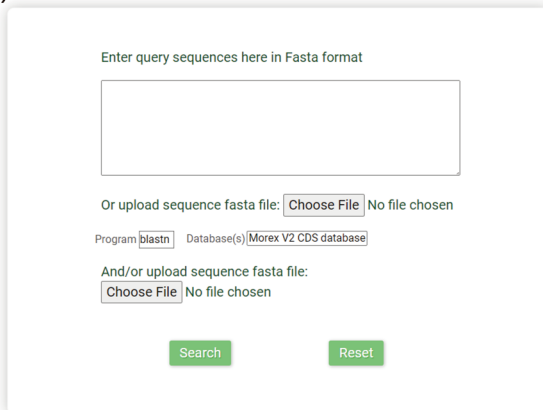


Fig. 2 Landscape of the mapping statistics and expression profiles (e.g., PRJEB14349: RNA-seq of 16 developmental stages of barley). **a** Overall mapping rate. **b** Concordantly exactly one time mapping rate. **c** Gene number (FPKM > 1) distribution. **d** Gene number (FPKM > 0.05) distribution. **e** Heatmap of pair-wise Pearson coefficients between samples. **f** Data representation by principal component analysis for all samples

(a)



(b)



(c)

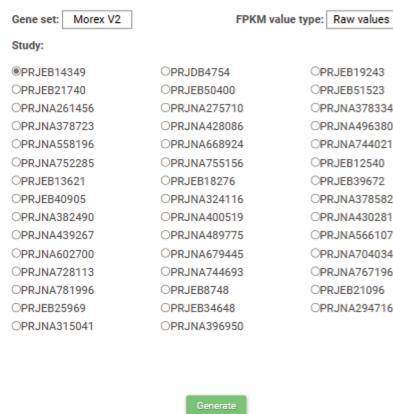
BarleyExpDB is a user-friendly, integrated, multifaceted RNA-seq database that allows users to conveniently retrieve gene expression from barley RNA-seq libraries and returns an intuitive visual display, accompanied by functional descriptions and annotations from GO, KEGG, PFAM, and SMART. Our database contains a total of 56 transcriptome studies, including 3,492 samples covering different developmental stages and tissues, biotic and abiotic stresses, and various mutants and populations (e.g., RIL and NIL). Details are listed in the table below. The platform also provides several useful toolkits implementing search, BLAST and download functions. We believe that BarleyExpDB will help make RNA-seq big data more accessible and available to the barley research community.

Note: The RNA-seq datasets were downloaded from the NCBI-SRA database to study the expression patterns of barley genes. Quality control was performed with Trimmomatic v0.36, Hisat2 v2.1.0 and StringTie v1.3.5 software were used to calculate fragments per kilobase of transcript per million mapped reads (FPKM) values for each gene with default parameters.

Summary of studies

SRA BioProject	Title	Publication	NGS technology	Description
PRJEB14349	RNA-Seq of 16 developmental stages of barley (Morex cultivar).	A chromosome conformation capture ordered sequence of the barley genome	Illumina HiSeq 2000	Cereal grasses of the Triticeae tribe have been the major food source in temperate regions since the dawn of agriculture. Their large genomes are characterized by a high content of repetitive ...more
PRJDB4754	Hordeum vulgare Transcriptome or Gene expression.	Extreme Suppression of Lateral Floret Development by a Single Amino Acid Change in the VRS1 Transcription Factor	Illumina HiSeq 2500	Total RNA was extracted from immature spikes at the awn primordium stage of cultivated barley (<i>Hordeum vulgare</i> L.). Total RNA was measured by using Agilent 2100 Bioanalyzer ...more

(d)



(e)



Fig. 3 Overview of BarleyExpDB. **a** Homepage of BarleyExpDB and example results. **b**. Sequence similarity search tools. **c**. Introduction of BarleyExpDB. **d**. Downloaded pages. **e**. Contact information and external links

provides a brief descriptive list of studies currently available in our database, as well as a user-friendly search box for gene expression profile queries. From this main hub, users should select the corresponding BioProject studies according to their individual research needs. Notably, we provide a tag for each

study that includes summary information to make it easy for users to quickly select the study they are interested in. The search function of the BarleyExpDB can be queried in one of the four gene sets, such as Morex V2 (e.g., HORVU.MOREX.r2.1HG0000021) [22], Morex V3 (e.g., HORVU.MOREX.r3.1HG0000030)

[38], B1K-04-12 (Horvu_FT11_1H01G003300) [15] or Zangqing320 (MLOC_42) [19]. Users can query the genes of interest directly in the search box. For larger-scale analyses, it is recommended that the user provide a processed text file with one gene ID per line and optionally upload it to BarleyExpDB. The maximum single query allowed in the database is 500 genes at a time. If more query genes are needed, users can query by submitting in batches, or directly download the raw data and extract the files by themselves.

The graphical representation profiles of all datasets are presented on a heat map showing the average RPKM values. Graphics can be downloaded in 'png' image format, and FPKM data is presented in an accompanying table that can be exported in '.csv' format, facilitating the analysis required by end-users. In addition to the expression search tools, we provide comprehensive information on PFAM, SMART, GO, KEGG, and functional descriptions for user-submitted genes, as well as valuable information for the selected RNA-seq study, such as genotype/phenotype, stage/tissue, and relevant publications. It should be of note that users have access to homologous genes in rice and *Arabidopsis thaliana*.

Blast

For sequence fragments without definite gene IDs, BarleyExpDB provides an online BLAST service to query across different database collections. Users can submit the amino acid or nucleotide sequences to the search box in Fasta format, or upload them to the database in text file format. Five kinds of BLAST algorithms (e.g., BLASTN, BLASTP, and TBLASTX) can be selected to identify putative homologous sequences. When browsing using the BLAST alignment function, candidate hits are listed in descending order of expectation and are viewed side-by-side in the results window.

Introduction

The "Introduction" page provides a brief description of BarleyExpDB and a drop-down menu where users can browse the "Materials and Methods" applied in BarleyExpDB. The analysis software used to build the database can be accessed directly via a link. The commands and parameters are also displayed.

The interface also provides a comprehensive description of each RNA-seq study in BarleyExpDB, such as sample accessions, stages/tissues, treatments, and related publications, which is valuable information to help users select appropriate samples and conduct subsequent studies.

Download

All the FPKM values of the expression matrix that are available for downloading and reanalyzing are listed on the 'Download' page.

About

The "About" page displays a few generic external links that users can access quickly.

Prospects

BarleyExpDB is in a continual state of incremental improvement. These resources will contribute to our understanding of the complex structures that control the regulatory mechanisms of biological processes in the barley genome. Furthermore, BarleyExpDB will be implemented with additional features and utilities to better serve the barley research community: (i) The assembly of telomere-to-telomere (T2T) genomic sequences has been recently reported in various plants [39, 40]. However, none of the seven chromosomes are completed from end to end, and a large number of unresolved gaps and missing sequences have been observed in the rDNA, centromere, and sub-telomere regions of Morex V3 assembly [41]. The development of a gap-closed, newly annotated T2T assembly for barley (which will likely be called Morex V4) is planned to be released soon, and this is our subsequent update of the genomic gene set; (ii) Given the widespread species-wide catalog of gene presence/absence variants (PAV), a single reference genomic context is not competent for the assessment of dispensable gene expression. We expect to integrate the Barley Pangenome V1, or the forthcoming Barley Pangenome v2, to provide genomic-transcriptomic companion expression profile datasets, which will provide important information on the functional studies of specific genes; (iii) With the growing repository of barley transcriptome datasets in the public domain, our database will be updated continuously with more newly release RNA-seq samples. The BarleyExpDB framework is highly scalable and can efficiently integrate newly released RNA-seq expression datasets, ensuring that we can achieve at least two updates per year; (iv) RNA editing is a process by which genetic information is modified in the RNA sequence corresponding to its DNA template [42]. The next version of BarleyExpDB is under development to study post-transcriptionally regulated RNA editing sites by incorporating paired DNA-RNA data; (v) The rapid and enthusiastic adoption of full-length, single-cell, and spatial RNA-seq revolutionizing the details of whole-transcriptome studies [43, 44]. The data generated by these novel sequencing technologies

will be integrated into BarleyExpDB with a view to reflecting the full spectrum of differentially alternatively spliced transcripts and snapshots from tissue to cell; (vi) New functions and analytical tools will be implemented in BarleyExpDB for the convenience of users, such as identification of tissue-specific genes and online eGWAS prediction of phenotype-related genes. Our approach will maximize the utility of the database in studying different aspects of barley development, enabling researchers and small labs without computing resources to mine complex and valuable expression datasets. We also welcome complementary RNA-seq datasets from third-party research groups to enrich our database.

Conclusions

The rapid development of next-generation sequencing technologies, coupled with the decreasing cost of sequencing, has led to the accumulation of copious amounts of expression data. We present BarleyExpDB, a convenient, web-accessible, and management-flexible RNA-seq database in barley that allows users to quickly scan the abundant information using the known gene ID of Morex V2, V3, B1K-04–12 and Zangqing320. BarleyExpDB is currently the most comprehensive RNA-seq database in barley and provides the expression levels in various tissues, developmental stages, environmental stresses, as well as in different genotypes, phenotypes, mutants and populations. Our database also implements several practical utilities for sequence homology searching, visualization, function annotation, and result downloading. We believe that BarleyExpDB will contribute to the acquisition and utilization of transcriptome big data and advance functional genomics and breeding biology research in barley.

Abbreviations

ARS	Arabidopsis RNA-seq
BLAST	Basic Local Alignment Search Tool
EC	Enzyme Commission
ENA	European Nucleotide Archive
FPKM	Fragments Per Kilobase of Transcript Per Million Mapped Reads
NIL	Near Isogenic Lines
GO	Gene Ontology
GSA	Genome Sequence Archive
KEGG	Kyoto Encyclopedia of Genes and Genomes
MDS	Multidimensional Scaling
NCBI	National Center for Biotechnology Information
PAV	Presence/Absence Variant
PAFM	Proteins Families Database
PCA	Principal Component Analysis
PPRD	Plant Public RNA-seq Database
RIL	Recombinant Inbred Lines
RNA-seq	RNA-sequencing
SMART	Simple Modular Architecture Research Tool
SRA	Sequence Read Archive
T2T	Telomere-to-Telomere

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-023-04193-z>.

Additional file 1: SupplementaryTable 1. Detail sample information of the RNA-seq datasets downloaded from the NCBI SRA database.

Acknowledgements

We are grateful to all the research groups that provided RNA-seq data to the public, and we apologize for not being able to cite all relevant papers in the manuscript due to space limitations. We appreciated Dr. Shengwei Ma for providing the reference genome and annotation information of Zangqing320. We also thank Dr. RuiMin Li for his constructive comments and the High-Performance Computing platform of Northwest A&F University.

Authors' contributions

Z.L., X.N., and L.C. designed and supervised the project. T.L. and R.L. downloaded the RNA-seq datasets. T.L., H.S., and J.B. performed the data processing. Y.L. and L.C. wrote the manuscript. Y.T. revised the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by the National Natural Science Foundation of China (Grant No. 32060458 and 32160219), and the Postdoctoral Foundation of China (Grant No. 2022M713430). The funders had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

Data pertaining to the study have been included in the article, and further inquiries can be directed to the corresponding authors. The barley genomes were downloaded from the given links: <https://doi.org/10.5447/IPK/2019/19> (Morex V2), <http://doi.org/10.5447/ipk/2021/3> (Morex V3), <https://doi.ipk-gatersleben.de/DOI/c4d433dc-bf7c-4ad9-9368-69bb77837c-a5/3490162b-3d76-4ba1-b6ee-3eae5f6b644/2> (B1K-04–12), and <http://wheatomics.sdau.edu.cn/> (Zangqing320). To comprehensively evaluate the expression pattern of barley, publicly available RNA-seq datasets were obtained from the NCBI SRA database with BioProject numbers PRJNA629999, PRJNA507455, PRJNA748178, PRJNA828098, PRJNA227211, PRJNA432492, PRJNA261456, PRJEB4947, PRJNA679445, PRJNA491382, PRJNA665933, PRJNA489775, PRJNA744021, PRJNA543388, PRJNA728483, PRJEB12540, PRJNA324116, PRJNA400519, PRJNA439267, PRJNA546269, PRJNA566107, PRJNA602700, PRJEB40905, PRJNA704034, PRJNA728113, PRJNA744693, PRJEB39672, PRJNA496380, PRJNA428086, PRJEB14349, PRJNA378582, PRJNA558196, PRJEB19243, PRJNA294716, PRJNA315041, PRJNA378334, PRJEB25969, PRJEB21740, PRJDB4754, PRJNA396950, PRJNA378723, PRJNA430281, PRJNA275710, PRJEB8748, PRJEB18276, PRJNA382490, PRJEB13621, PRJEB21096, PRJEB34648, PRJEB50400, PRJEB51523, PRJNA668924, PRJNA752285, PRJNA767196, PRJNA781996 and PRJNA755156. The gene expression matrix containing the raw and averaged FPKM values across 56 studies can be directly downloaded from <http://barleyexp.com/download.html>.

Declarations

Ethics approval and consent to participate

Plant samples were not collected and processed in this study. No specific permits were required for sample collection in this study. We complied with relevant institutional, national, and international guidelines and legislations for plant study.

Consent for publication

Not applicable.

Competing interests

Author Y.T. is employed by Xintai Urban and Rural Development Group Co., Ltd., Taian, Shandong, China. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 2 December 2022 Accepted: 27 March 2023
Published online: 31 March 2023

References

- Zhang H, Zhang F, Yu Y, Feng L, Jia J, Liu B, Li B, Guo H, Zhai J. A comprehensive online database for exploring ~20,000 public arabidopsis RNA-Seq libraries. *Mol Plant*. 2020;13(9):1231–3.
- Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57–63.
- Langmead B, Nellore A. Cloud computing for genomic data analysis and collaboration. *Nat Rev Genet*. 2018;19(4):208–19.
- Liu J, Yin F, Lang K, Jie W, Tan S, Duan R, Huang S, Huang W. MetazExp: a database for gene expression and alternative splicing profiles and their analyses based on 53 615 public RNA-seq samples in 72 metazoan species. *Nucleic Acids Res*. 2022;50(D1):D1046–d1054.
- Robinson AJ, Tamiru M, Salby R, Bolitho C, Williams A, Huggard S, Fisch E, Unsworth K, Whelan J, Lewsey MG. AgriSeqDB: an online RNA-Seq database for functional studies of agriculturally relevant plant species. *BMC Plant Biol*. 2018;18(1):200.
- Katz K, Shutov O, Lapoint R, Kimelman M, Brister JR, O'Sullivan C. The sequence read archive: a decade more of explosive growth. *Nucleic Acids Res*. 2022;50(D1):D387–d390.
- Harrison PW, Ahamed A, Aslam R, Alako BT, Burgin J, Buso N, Courtot M, Fan J, Gupta D, Haseeb M, et al. The European Nucleotide Archive in 2020. *Nucleic Acids Res*. 2021;49(D1):D82–d85.
- Wang Y, Song F, Zhu J, Zhang S, Yang Y, Chen T, Tang B, Dong L, Ding N, Zhang Q, et al. GSA: genome sequence archive. *Genom Proteom Bioinf*. 2017;15(1):14–8.
- Yu Y, Zhang H, Long Y, Shu Y, Zhai J. Plant Public RNA-seq Database: a comprehensive online database for expression analysis of ~45 000 plant public RNA-Seq libraries. *Plant Biotechnol J*. 2022;20(5):806–8.
- Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, Silverstein MC, Ma'ayan A. Massive mining of publicly available RNA-seq data from human and mouse. *Nat Commun*. 2018;9(1):1366.
- Collado-Torres L, Nellore A, Kammers K, Ellis SE, Taub MA, Hansen KD, Jaffe AE, Langmead B, Leek JT. Reproducible RNA-seq analysis using recount2. *Nat Biotechnol*. 2017;35(4):319–21.
- Vivian J, Rao AA, Nothhaft FA, Ketchum C, Armstrong J, Novak A, Pfeil J, Narkizian J, Deran AD, Musselman-Brown A, et al. Toil enables reproducible, open source, big biomedical data analyses. *Nat Biotechnol*. 2017;35(4):314–6.
- Pearce S, Vazquez-Gross H, Herin SY, Hane D, Wang Y, Gu YQ, Dubcovsky J. WheatExp: an RNA-seq expression database for polyploid wheat. *BMC Plant Biol*. 2015;15:299.
- Liu D, Yu L, Wei L, Yu P, Wang J, Zhao H, Zhang Y, Zhang S, Yang Z, Chen G, et al. BnTIR: an online transcriptome platform for exploring RNA-seq libraries for oil crop *Brassica napus*. *Plant Biotechnol J*. 2021;19(10):1895–7.
- Jayakodi M, Padmarasu S, Haberer G, Bonthala VS, Gundlach H, Monat C, Lux T, Kamal N, Lang D, Himmelbach A, et al. The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature*. 2020;588(7837):284–9.
- Dawson IK, Russell J, Powell W, Steffenson B, Thomas WTB, Waugh R. Barley: a translational model for adaptation to climate change. *New Phytol*. 2015;206(3):913–31.
- Liu M, Li Y, Ma Y, Zhao Q, Stiller J, Feng Q, Tian Q, Liu D, Han B, Liu C. The draft genome of a wild barley genotype reveals its enrichment in genes related to biotic and abiotic stresses compared to cultivated barley. *Plant Biotechnol J*. 2020;18(2):443–56.
- Zeng X, Guo Y, Xu Q, Mascher M, Guo G, Li S, Mao L, Liu Q, Xia Z, Zhou J, et al. Origin and evolution of qingke barley in Tibet. *Nat Commun*. 2018;9(1):5433.
- Dai F, Wang X, Zhang XQ, Chen Z, Nevo E, Jin G, Wu D, Li C, Zhang G. Assembly and analysis of a qingke reference genome demonstrate its close genetic relation to modern cultivated barley. *Plant Biotechnol J*. 2018;16(3):760–70.
- Mayer KF, Waugh R, Brown JW, Schulman A, Langridge P, Platzer M, Fincher GB, Muehlbauer GJ, Sato K, Close TJ, et al. A physical, genetic and functional sequence assembly of the barley genome. *Nature*. 2012;491(7426):711–6.
- Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J, et al. A chromosome conformation capture ordered sequence of the barley genome. *Nature*. 2017;544(7651):427–33.
- Monat C, Padmarasu S, Lux T, Wicker T, Gundlach H, Himmelbach A, Ens J, Li C, Muehlbauer GJ, Schulman AH, et al. TRITEX: chromosome-scale assembly of Triticeae genomes with open-source tools. *Genome Biol*. 2019;20(1):284.
- Sato K, Mascher M, Himmelbach A, Haberer G, Spannagl M, Stein N. Chromosome-scale assembly of wild barley accession "OUH602." G3 (Bethesda) 2021;11(10):244.
- Cai S, Shen Q, Huang Y, Han Z, Wu D, Chen ZH, Nevo E, Zhang G. Multi-omics analysis reveals the mechanism underlying the edaphic adaptation in wild barley at evolution slope (Tabigha). *Adv Sci (Weinh)*. 2021;8(20): e2101374.
- Xu C, Zhan C, Huang S, Xu Q, Tang T, Wang Y, Luo J, Zeng X. Resistance to powdery mildew in qingke involves the accumulation of aromatic phenolamides through jasmonate-mediated activation of defense-related genes. *Front Plant Sci*. 2022;13: 900345.
- Steffenson BJ, Olivera P, Roy JK, Jin Y, Smith KP, Muehlbauer GJ. A walk on the wild side: mining wild wheat and barley collections for rust resistance genes. *Aust J Agric Res*. 2007;58(6):532–44.
- Milne L, Bayer M, Rapazote-Flores P, Mayer CD, Waugh R, Simpson CG. EORN, a barley gene and transcript abundance database. *Sci Data*. 2021;8(1):90.
- Rapazote-Flores P, Bayer M, Milne L, Mayer CD, Fuller J, Guo W, Hedley PE, Morris J, Halpin C, Kam J, et al. BaRTv1.0: an improved barley reference transcript dataset to determine accurate changes in the barley transcriptome using RNA-seq. *BMC Genomics*. 2019;20(1):968.
- Coulter M, Entizne JC, Guo W, Bayer M, Wonneberger R, Milne L, Schreiber M, Haaning A, Muehlbauer GJ, McCallum N, et al. BaRTv2: a highly resolved barley reference transcriptome for accurate transcript-specific RNA-seq quantification. *Plant J*. 2022;111(4):1183–202.
- Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol*. 2021;38(12):5825–9.
- Deng W, Nickle DC, Learn GH, Maust B, Mullins JL. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. *Bioinformatics*. 2007;23(17):2334–6.
- Harb A, Simpson C, Guo W, Govindan G, Kakani VG, Sunkar R. The effect of drought on transcriptome and hormonal profiles in barley genotypes with contrasting drought tolerance. *Front Plant Sci*. 2020;11: 618491.
- Thiel J, Koppolu R, Trautewig C, Hertig C, Kale SM, Erbe S, Mascher M, Himmelbach A, Rutten T, Esteban E, et al. Transcriptional landscapes of floral meristems in barley. *Sci Adv*. 2021;7(18):0832.
- Borrego-Benjumea A, Carter A, Tucker JR, Yao Z, Xu W, Badea A. Genome-wide analysis of gene expression provides new insights into waterlogging responses in barley (*Hordeum vulgare* L.). *Plants (Basel)*. 2020;9(2):240.
- Bélangier S, Marchand S, Jacques P, Meyers B, Belzile F. Differential expression profiling of microspores during the early stages of isolated microspore culture using the responsive barley cultivar gobernadora. G3 (Bethesda) 2018;8(5):1603–14.
- Szurman-Zubrzycka M, Chwiałkowska K, Niemira M, Kwaśniewski M, Nawrot M, Gajecka M, Larsen PB, Szarejko I. Aluminum or Low pH - which is the bigger enemy of barley? transcriptome analysis of barley root meristem under al and Low pH stress. *Front Genet*. 2021;12: 675260.
- Pan R, Ding M, Feng Z, Zeng F, Medison MB, Hu H, Han Y, Xu L, Li C, Zhang W. HvGST4 enhances tolerance to multiple abiotic stresses in barley: evidence from integrated meta-analysis to functional verification. *Plant Physiol Biochem*. 2022;188:47–59.
- Mascher M, Wicker T, Jenkins J, Plott C, Lux T, Koh CS, Ens J, Gundlach H, Boston LB, Tulpová Z, et al. Long-read sequence assembly: a technical evaluation in barley. *Plant Cell*. 2021;33(6):1888–906.
- Zhang Y, Fu J, Wang K, Han X, Yan T, Su Y, Li Y, Lin Z, Qin P, Fu C, et al. The telomere-to-telomere gap-free genome of four rice parents reveals SV and PAV patterns in hybrid rice breeding. *Plant Biotechnol J*. 2022;20(9):1642–4.
- Liu J, Seetharam AS, Chougule K, Ou S, Swentowsky KW, Gent JJ, Llacua V, Woodhouse MR, Manchanda N, Presting GG, et al. Gapless assembly of maize chromosomes using long-read technologies. *Genome Biol*. 2020;21(1):121.

41. Navrátilová P, Toegelová H, Tulpová Z, Kuo YT, Stein N, Doležel J, Houben A, Šimková H, Mascher M. Prospects of telomere-to-telomere assembly in barley: analysis of sequence gaps in the MorexV3 reference genome. *Plant Biotechnol J*. 2022;20(7):1373–86.
42. Zhang H, Fu Q, Shi X, Pan Z, Yang W, Huang Z, Tang T, He X, Zhang R. Human A-to-I RNA editing SNP loci are enriched in GWAS signals for autoimmune diseases and under balancing selection. *Genome Biol*. 2020;21(1):288.
43. Shaw R, Tian X, Xu J. Single-cell transcriptome analysis in plants: advances and challenges. *Mol Plant*. 2021;14(1):115–26.
44. Byrne A, Cole C, Volden R, Vollmers C. Realizing the potential of full-length transcriptome sequencing. *Philos Trans R Soc Lond B Biol Sci*. 2019;374(1786):20190097.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

