



Transcriptome atlas of *Phalaenopsis equestris*

Anna V. Klepikova¹, Artem S. Kasianov¹, Margarita A. Ezhova^{1,2},
Aleksey A. Penin^{1,3} and Maria D. Logacheva^{1,2,3}

¹Laboratory of Plant Genomics, Institute for Information Transmission Problems of the Russian Academy of Sciences, Moscow, Russia

²Skolkovo Institute of Science and Technology, Moscow, Russia

³Lomonosov Moscow State University, Moscow, Russia

ABSTRACT

The vast diversity of Orchidaceae together with sophisticated adaptations to pollinators and other unique features make this family an attractive model for evolutionary and functional studies. The sequenced genome of *Phalaenopsis equestris* facilitates Orchidaceae research. Here, we present an RNA-seq-based transcriptome map of *P. equestris* that covers 19 organs of the plant, including leaves, roots, floral organs and the shoot apical meristem. We demonstrated the high quality of the data and showed the similarity of the *P. equestris* transcriptome map with the gene expression atlases of other plants. The transcriptome map can be easily accessed through our database Transcriptome Variation Analysis (TraVA) for visualizing gene expression profiles. As an example of the application, we analyzed the expression of *Phalaenopsis* “orphan” genes—those that do not have recognizable similarity with the genes of other plants. We found that approximately half of these genes were not expressed; the ones that were expressed were predominantly expressed in reproductive structures.

Subjects Bioinformatics, Molecular Biology, Plant Science

Keywords Orchidaceae, *Phalaenopsis equestris*, Transcriptome map, RNA-seq, Database, Orphan genes

Submitted 26 November 2020

Accepted 15 November 2021

Published 10 December 2021

Corresponding author

Anna V. Klepikova,
annklepikova@gmail.com

Academic editor

Francisco Balao

Additional Information and
Declarations can be found on
page 9

DOI 10.7717/peerj.12600

© Copyright
2021 Klepikova et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

INTRODUCTION

The enormous diversity of orchids traditionally attracts the attention of plant biologists. Orchidaceae comprises approximately 25 thousand species, which makes it the largest plant taxon (Cai *et al.*, 2015). The diversification of orchids has evolved along with complex pollinator-adapted flower structure (Cozzolino & Widmer, 2005), CAM photosynthesis and epiphytism (Silvera *et al.*, 2009). Orchids are highly valuable ornamental plants; among them, *Phalaenopsis* species are one of the most widely grown and sold (Griesbach, 2002). There is a need to develop tools for the identification of species and cultivars and for marker-assisted breeding.

A genome assembly of *Phalaenopsis equestris* (horse phalaenopsis) (Cai *et al.*, 2015) and its improvement with long reads (Zhang *et al.*, 2017) provided novel opportunities for evolutionary and functional studies of Orchidaceae. Genome assembly was used for functional studies of transcription factor families (Lin *et al.*, 2016; Valoroso *et al.*, 2019), somatic embryogenesis (Chen *et al.*, 2019), and retrotransposon insertions (Hsu *et al.*,

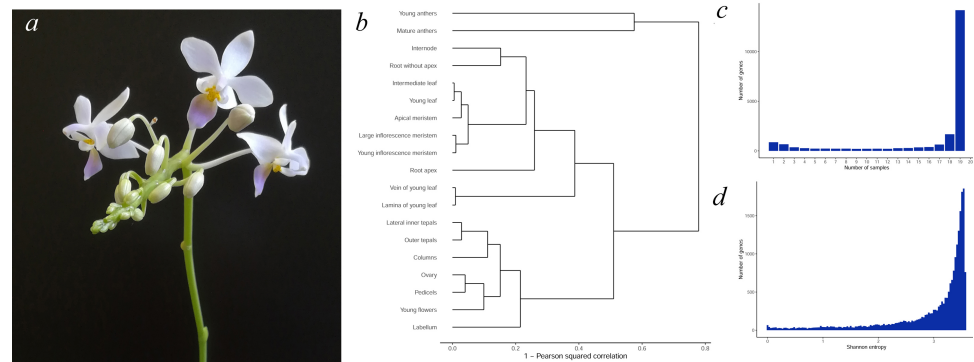


Figure 1 (A) The general view of *P. equestris* var. *cyanochilus*; (B) hierarchical clustering tree of transcriptome map samples; (C) the distribution of genes by the number of samples where the gene is expressed. Only expressed genes with 5 or more normalized read counts in each biological replicate were considered; (D) the distribution of the Shannon entropy of *P. equestris* genes.

Full-size [DOI: 10.7717/peerj.12600/fig-1](https://doi.org/10.7717/peerj.12600/fig-1)

2019), as well as for evolutionary studies of ancient polyploidy (Barrett *et al.*, 2019). However, transcriptome resources of *P. equestris* remain limited even though *de novo* transcriptome assembly was performed based on RNA sequencing of 11 organs (Niu *et al.*, 2016). Transcriptome data are essential for many research directions. In particular, they provide insight into gene function in multigene families (see, *e.g.*, Su *et al.*, 2013; Kuo *et al.*, 2021). Additionally, if available for different species or varieties, these data help to characterize genetic diversity in coding regions (see, *e.g.*, Tsai *et al.*, 2015).

In our study, we present a transcriptome map of *P. equestris* consisting of 19 samples in two biological replicates. The ornamental orchid *P. equestris* comprises three varieties and numerous hybrids of various flower colors and sizes (Hsu & Chen, 2016). To create a transcriptome atlas, we chose *P. equestris* var. *cyanochilus* because clonal plants are available for this cultivar, which helps to reduce interindividual variability (Fig. 1A). High-quality RNA of orchid organs and tissues was sequenced using Illumina technology, resulting in 1,687 M reads. We compared the expression characteristics of the *P. equestris* transcriptome map with gene expression atlases of other plants to provide evidence of the reliability of our data. Transcriptome map of *P. equestris* can be applied in a great variety of functional studies.

MATERIALS & METHODS

Growth conditions

Plants of *P. equestris* var. *cyanochilus* (the name of the variety means “blue lip”, var. blue, according to the provider, orchidee.su) were grown in a climate chamber under a 16 h light/8 h dark cycle at 22 °C and 50–60% relative humidity. Samples were collected in two biological replicates; each replicate consisted of at least seven plants. Sample collection was performed within two hours (Zeitgeber time ZT8-10) to reduce the influence of the circadian cycle.

RNA extraction, library preparation and sequencing

RNA was extracted using the RNeasy mini kit (Qiagen, The Netherlands) following the manufacturer's protocol. To ensure a high quality of *Phalaenopsis* samples, RNA was analyzed using capillary electrophoresis on an Agilent Bioanalyzer 2100. cDNA libraries for Illumina sequencing were constructed using the NEBNext Ultra II RNA Library Prep Kit for Illumina (New England BioLabs, MA, USA) following the manufacturer's protocol in 0.5 of the recommended volume (due to low RNA quantity in samples such as the shoot apical meristem). cDNA libraries were sequenced with HiSeq4000 and NextSeq500 (Illumina, CA, USA) instruments (50 bp and 75 bp single read runs).

Read mapping

Read trimming was performed using Trimmomatic version 0.36 (Bolger, Lohse & Usadel, 2014) in single read mode with the parameters "ILLUMINACLIP:common.adapters.file:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:15 MINLEN:30". For read mapping, genome assembly and annotation of *P. equestris* from the PLAZA database (version 4.5, https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v4_5_monocots/organism/view/Phalaenopsis+equestris) was used. Trimmed reads mapped on the genome assembly using Spliced Transcripts Alignment to a Reference (STAR) version 2.4.2 (Dobin et al., 2013) in the "GeneCounts" mode and parameters "-sjdbOverhang 59 -sjdbGTFfeatureExon exon -sjdbGTFtagExonParentTranscript gene_id" to obtain counts of uniquely mapped reads on each gene.

Expression characteristics of the transcriptome map

Gene read counts obtained with STAR were normalized to library size using size factors, as described in (Anders & Huber, 2010). A threshold of five or higher normalized read counts in each biological replicate was used to define expressed genes.

To describe the gene expression pattern, Shannon entropy values were calculated for expressed in at least one sample genes (Schug et al., 2005). To avoid overrepresentation of certain plant organs, the samples were grouped using distances on a clustering tree: gene expression levels were averaged if samples had distance $(1 - \text{Pearson } r^2)$ less than 0.1. Sample groups are listed in Table S1.

Differential expression analysis

Differential expression between each pair of samples was analyzed using the R packages "DESeq2" (Love, Huber & Anders, 2014), "edgeR" (Robinson, McCarthy & Smyth, 2010), and "baySeq" (Hardcastle & Kelly, 2010). The thresholds "FDR-corrected p -value < 0.05 " and "fold change ≥ 2 " were used to consider a gene as differentially expressed.

Data availability

The raw RNA-seq data of the transcriptome map were deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession PRJNA667255. The TraVA database can be accessed at http://travadb.org/browse/Species=Phalaenopsis_equestris/.

RESULTS AND DISCUSSION

Transcriptome map construction

We collected 31 samples covering main plant organs and developmental stages, such as roots, young and mature leaves, floral organs, flower buds, and meristems. Each sample was collected in two biological replicates, and each replicate was pooled from at least seven plants. Sample RNA was sequenced on an Illumina platform, resulting in 29 M–65 M raw single reads (38 M median) for each sample (for sequencing statistics, see [Table S2](#)). After removing low-quality reads and adapter sequenced 98.7–99.8% of reads remained ([Table S2](#)).

Reads were mapped to the reference genome of *P. equestris* ([Cai et al., 2015](#)) with only one match allowed (unique mapping); 9.2–89.6% of high-quality reads were successfully mapped ([Table S2](#)). Twelve samples showed an extremely low mapping percentage; unmapped reads were identified as sequences belonging to Cymbidium mosaic virus (GenBank accession [MK816927](#)), which is known to persist in the majority of the *P. equestris* population and affect mainly mature and senescent tissues ([Koh, Lu & Chan, 2014](#)). As the library size of infected samples was insufficient and could distort the conclusions, we excluded samples with a percentage of mapped reads lower than 35% in at least one biological replicate. The remaining samples had 37.3–89.6% uniquely mapped reads with a median of 81.6%.

Thus, we constructed a transcriptome map of *P. equestris* covering 19 organs and parts of the plant. Floral organs (anthers, labellum, inner and outer tepals), leaves at different developmental stages, axes (inflorescence and pedicel), shoot apical and inflorescence meristems, and root parts were taken into the analysis (for a detailed description of the samples, see [Table S3](#)). The biological replicates showed high consistency (median Pearson $r^2 = 0.99$, [Table S4](#)).

Clustering of samples generally reflects the plant body plan and groups organs with similar morphology and physiology ([Klepikova & Penin, 2019](#)). Hierarchical clustering of *P. equestris* samples showed the same pattern ([Fig. 1B](#)). Sample clusters were formed by floral organs, leaf parts, meristems and young leaves, inflorescence internode and root; young and mature anthers were the most distant from the other samples, similar to *A. thaliana*, rice, and maize ([Nobuta et al., 2007](#); [Wang et al., 2010](#); [Stelpflug et al., 2016](#); [Klepikova et al., 2016](#)). The distances between samples on the clustering tree were closer than those in the other species we observed ([Klepikova et al., 2016](#); [Penin et al., 2019](#)), which can be explained by the lack of older tissues in the *P. equestris* transcriptome atlas.

We compared our samples with publicly available *P. equestris* transcriptomes ([Table S5](#)). In general, the clustering of samples was consistent ([Fig. S1](#)), although leaves and columns from BioProject [PRJNA288388](#) ([Niu et al., 2016](#)) clustered outside the other samples.

Expression characteristics of *P. equestris*

The *Phalaenopsis* genome annotation (PLAZA database, version 4.5) includes 29 431 protein-coding genes. Among them, 14 174 (48%) genes were expressed in all samples (using five reads in each biological replicate as a threshold), while transcripts of 21 671

(74%) genes were found in at least one sample. These values are in the range of typically expressed gene numbers across plant transcriptome maps (Klepikova & Penin, 2019). As in other species, samples demonstrated similarity in the number of expressed genes, which varied from 15 612 (53%) in the shoot apical meristem to 18 947 (64%) in ovules before pollination (Table S6).

Expression patterns of *P. equestris* genes

The study of gene expression patterns can shed light on the biological function of the gene and place it among essential for a plant existence ubiquitously expressed genes or precise regulators of tissue features –sample-specific expressed genes. We used two approaches to define gene expression patterns. Counting the number of samples where a gene is expressed is the simplest method to characterize expression pattern width, as was shown for *Nicotiana tabacum* (Edwards et al., 2010) and *Vigna unguiculata* (Yao et al., 2016). Here, the majority of genes (16,486) were expressed in 17 or more samples; the second peak (1,896 genes) of the distribution was formed by genes expressed in three or fewer samples (Fig. 1C). Most tissue-specific genes were found in anthers (56% of tissue-specific genes), roots (11%), and meristems (both shoot apical and inflorescence meristems, 8%). A high number of anther-associated genes have been found in *A. thaliana* (Klepikova et al., 2016) and are expected for *P. equestris*, as young and mature anthers are the most distant samples on the clustering tree (Fig. 1B).

While useful, this first approach depends on an arbitrary threshold that separates expressed and nonexpressed genes and does not take into account the variation in the expression level between samples. To overcome this issue, we used Shannon entropy as a measure of expression pattern width: low entropy values correspond to tissue-specific genes, while high values mark ubiquitously expressed genes (Schug et al., 2005). The distribution of Shannon entropy in *P. equestris* was significantly skewed to the right, revealing a large proportion of widely expressed genes (Fig. 1D), similar to *A. thaliana*, *Solanum lycopersicum* and *Zea mays* (Sekhon et al., 2013; Klepikova et al., 2016; Penin et al., 2019).

Using a Shannon entropy value lower than 0.25, we identified 521 tissue-specific genes. As was found under direct counting, the majority of genes were associated with anthers or roots (Fig. 2). According to GO enrichment, genes uniquely expressed in the mature anthers were involved in cell wall organization, biogenesis and modification and had pectin esterase and enzyme inhibitor activity (Table S7). Young anthers were characterized by genes encoding products with amine and amino acid binding activity (Table S8). Root-specific genes (expressed in the sample “Root without apex”) were described by the terms “response to chemical stimulus”, “response to oxidative stress”, “oxidation–reduction”, and “heme binding” (Table S9).

To identify genes that were uniformly expressed across tissues, we selected genes with a Shannon entropy 3.55 or higher and calculated the coefficient of variance (CV) as a measure of expression stability. For 899 out of 1,340 genes, the CV was less than 0.25, indicating uniform expression in all samples and biological replicates. Stable genes had GO enrichment in terms associated with vesicles, membranes, RNA processing and localization. The list of

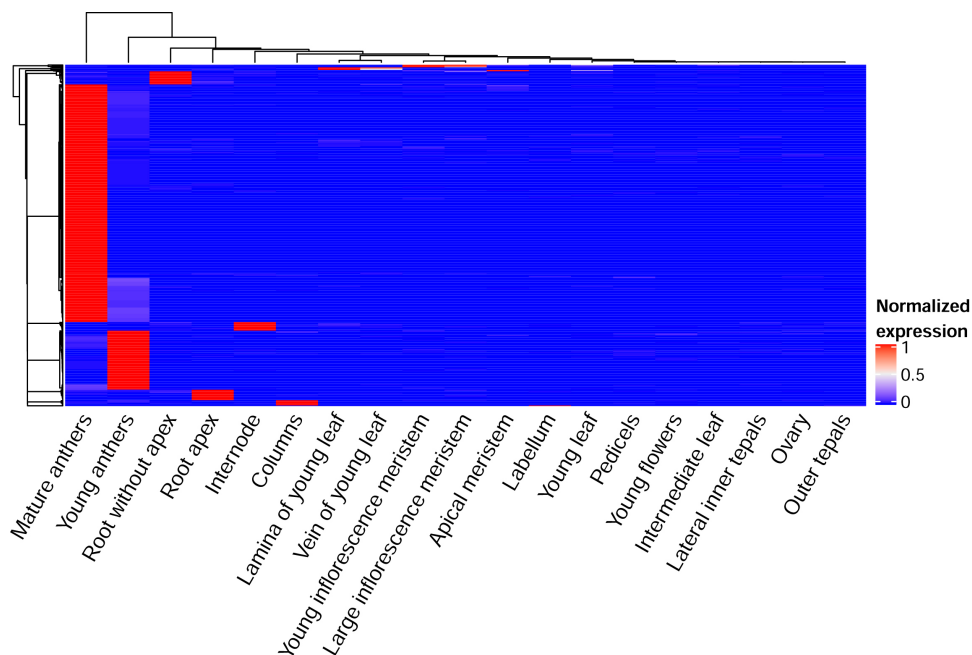


Figure 2 The heatmap of tissue-specific genes. Expression levels of each gene in each sample were normalized on its maximal expression level.

Full-size DOI: [10.7717/peerj.12600/fig-2](https://doi.org/10.7717/peerj.12600/fig-2)

GO categories strongly overlapped with the enrichment of uniformly expressed genes in *A. thaliana*, indicating interspecies universality of basic biological processes (Table S10).

P. equestris transcriptome variation database

We aimed to make our transcriptome data easily accessible and ready to use, so we uploaded *P. equestris* transcriptomes into our database Transcriptome Variation Analysis (TraVA, http://travadb.org/browse/Species=Phalaenopsis_equestris/). Gene IDs which are used in TraVA database (e.g., PEQU_41727) match the IDs in PLAZA genome assembly (see Materials and Methods). The TraVA interface demonstrates a color chart of gene expression profiles in a single- or multiple-gene view. A user can choose to show or hide expression values in a chart and choose between several types of read count normalization (Fig. 3).

We included the results of differential expression analysis in the TraVA database. In single-gene mode, the user can select one of three tools (DESeq2, edgeR or baySeq) and a sample against which other samples will be tested for the presence of differential expression.

Application of the TraVA database to the characterization of orchid genes

The graphical interface of TraVA facilitates gene expression pattern analysis and comparison and can be widely used in *P. equestris* functional studies. Orchids are a large and highly diverse plant family whose species are adapted to a number of ecological niches (typical terrestrial plants, epiphytes, nonphotosynthetic plants). These adaptations are reflected in their genome –for example, in *P. equestris*, which has a sophisticatedly differentiated



Figure 3 Database view. Expression profiles of *P. equestris* genes *PEQU_39433*, *PEQU_02900*, *PEQU_33696*.

Full-size DOI: 10.7717/peerj.12600/fig-3

perianth, the number of *AP3* orthologs is higher than that of *Apostasia shenzhenica*, the basal orchid species with an undifferentiated perianth (Zhang et al., 2017). Conversely, *P. equestris*, which is an epiphyte and does not develop typical terrestrial roots, lacks *AGL12* and several genes of the *ANR1* clade, in contrast to *A. shenzhenica* (Zhang et al., 2017). This stresses the importance of the study of lineage-specific genes and gene families. For the genes that do not have orthologs in model species, the analysis of their expression profiles is the first step towards functional characterization.

We identified 181 (160 after filtering out the proteins that had Xs across more than 50% of its length) *P. equestris* proteins that did not have significant similarity to any *Arabidopsis* protein (*e*-value cutoff = 10). Of them, 118 share similarity with the proteins of *A. shenzhenica* and are thus presumably orchid-specific, while 42 have no hits and thus emerged after the divergence of Apostasioideae and Epidendroideae. The survey of the expression profiles showed that 93 of them were not expressed in any of the samples in the map (Fig. S2). Among those that are expressed, most are expressed at very low levels. Higher expression levels were associated with reproductive structures, particularly anthers (Fig. 4).

Among vegetative structures, the most distinct is the root (root apex), where three genes—*PEQU_39433*, *PEQU_02900*, and *PEQU_33696*—have the highest expression levels. *Phalaenopsis* roots are unique (compared to most other plants, including *A. shenzhenica*, but not other epiphytic orchids) in many respects—in particular, they are photosynthetic and develop a special structure called velamen. Velamen is a tissue

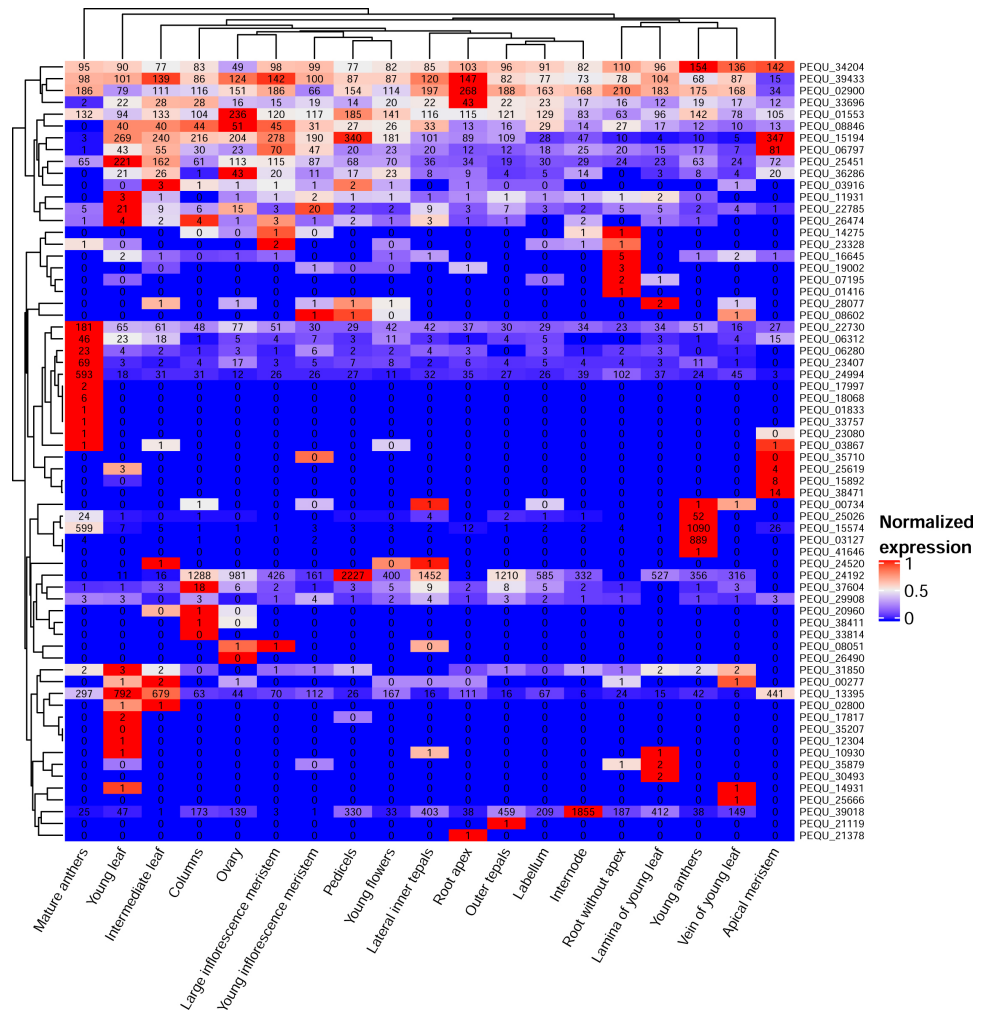


Figure 4 The heatmap of *P. equestris*-specific genes. Expression levels of each gene in each sample were normalized on its maximal expression level for the color key. The numbers on the figure represent normalized gene read count averaged over biological replicates.

Full-size [DOI: 10.7717/peerj.12600/fig-4](https://doi.org/10.7717/peerj.12600/fig-4)

of epidermal origin that consists of several layers of dead cells that help to absorb water and protect photosynthetic tissues of the root from UV damage. Notably, *PEQU_39433* and *PEQU_33696* do not have homologs in *A. shenzhenica*. *PEQU_02900* has marginal similarity (34%) with the *A. shenzhenica* protein encoded by the Ash001570 gene.

The topic of orphan genes—those that lack detectable homologs in other lineages—is widely discussed, particularly regarding plants (Arendsee, Li & Wurtele, 2014). While some orphan genes might represent artifacts of the annotation, others have a function (for example, *A. thaliana* orphan gene QQS, which acts in starch metabolism) (Li et al., 2009). The functional analysis of orphan genes, however, lags behind that of typical genes, as orphan genes are overlooked in annotations based on homology; they are also usually expressed at lower levels and in a narrower range of tissues (reviewed in Schlötterer, 2015). The study of expression levels and patterns of a potential orphan gene is a first step towards

its characterization—the detectable level of expression is evidence that the ORF is indeed a gene, not an annotation artifact.

Notably, orphan genes in well-characterized animal genomes (e.g., those of *Drosophila* and primates) have expression patterns biased towards male reproductive structures (Begun et al., 2007; Xie et al., 2012). According to the “out-of-testis” hypothesis (Kaessmann, 2010), this phenomenon is mediated by the unique epigenetic state of chromatin during male gametogenesis. We observed the same bias in *Phalaenopsis*; the growing availability of plant transcriptome maps will enable us to determine if this phenomenon is universal for plants.

CONCLUSIONS

In this study, we present a transcriptome map of the orchid *Phalaenopsis equestris* covering 19 organs at various stages of development. We identified 521 tissue-specific genes, the majority of which were expressed in anthers, roots (11%), and meristems. The uniformly expressed genes were associated with similar processes as those in *Arabidopsis thaliana*, i.e., vesicles, membranes, RNA processing and localization. To improve the use of these data, we integrated a transcriptome map into our database TraVA and demonstrated its usability in the study of *P. equestris* orphan genes. We expect that this resource will help to further investigate the genetic basis of unusual traits typical for this species and/or other orchids (CAM photosynthesis, unique floral structure, the development of aerial roots). Moreover, since *Phalaenopsis equestris* is a monocot and there is a bias towards grasses in large-scale transcriptome projects (see Klepikova & Penin, 2019), the characterization of the nongrass monocot transcriptome will be helpful not only for partial genetics of this species but also for large-scale comparative transcriptomics of flowering plants.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The reported study was funded by RFBR according to the research project No. 18-29-13017. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
RFBR: No. 18-29-13017.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Anna V. Klepikova analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Artem Sergeevich Kasianov analyzed the data, prepared figures and/or tables, and approved the final draft.

- Margarita A. Ezhova performed the experiments, prepared figures and/or tables, and approved the final draft.
- Aleksey A. Penin conceived and designed the experiments, performed the experiments, prepared figures and/or tables, and approved the final draft.
- Maria D. Logacheva analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The RNA-seq raw data of transcriptome map are available at NCBI Sequence Read Archive (SRA): [PRJNA667255](https://www.ncbi.nlm.nih.gov/sra/PRJNA667255).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.12600#supplemental-information>.

REFERENCES

- Anders S, Huber W. 2010.** Differential expression analysis for sequence count data. *Genome Biology* 11:R106 DOI [10.1186/gb-2010-11-10-r106](https://doi.org/10.1186/gb-2010-11-10-r106).
- Arendsee ZW, Li L, Wurtele ES. 2014.** Coming of age: orphan genes in plants. *Trends in Plant Science* 19:698–708 DOI [10.1016/j.tplants.2014.07.003](https://doi.org/10.1016/j.tplants.2014.07.003).
- Barrett CF, McKain MR, Sinn BT, Ge X-J, Zhang Y, Antonelli A, Bacon CD. 2019.** Ancient polyploidy and genome evolution in palms. *Genome Biology and Evolution* 11:1501–1511 DOI [10.1093/gbe/evz092](https://doi.org/10.1093/gbe/evz092).
- Begun DJ, Lindfors HA, Kern AD, Jones CD. 2007.** Evidence for *de Novo* evolution of testis-expressed genes in the *Drosophila yakuba*/*Drosophila erecta* Clade. *Genetics* 176:1131–1137 DOI [10.1534/genetics.106.069245](https://doi.org/10.1534/genetics.106.069245).
- Bolger AM, Lohse M, Usadel B. 2014.** Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120 DOI [10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170).
- Cai J, Liu X, Vanneste K, Proost S, Tsai W-C, Liu K-W, Chen L-J, He Y, Xu Q, Bian C, Zheng Z, Sun F, Liu W, Hsiao Y-Y, Pan Z-J, Hsu C-C, Yang Y-P, Hsu Y-C, Chuang Y-C, Dievart A, Dufayard J-F, Xu X, Wang J-Y, Wang J, Xiao X-J, Zhao X-M, Du R, Zhang G-Q, Wang M, Su Y-Y, Xie G-C, Liu G-H, Li L-Q, Huang L-Q, Luo Y-B, Chen H-H, Van de Peer Y, Liu Z-J. 2015.** The genome sequence of the orchid *Phalaenopsis equestris*. *Nature Genetics* 47:65–72 DOI [10.1038/ng.3149](https://doi.org/10.1038/ng.3149).
- Chen J-C, Tong C-G, Lin H-Y, Fang S-C. 2019.** *Phalaenopsis* LEAFY COTYLEDON1-induced somatic embryonic structures are morphologically distinct from protocorm-like bodies. *Frontiers in Plant Science* 10:1594 DOI [10.3389/fpls.2019.01594](https://doi.org/10.3389/fpls.2019.01594).
- Cozzolino S, Widmer A. 2005.** Orchid diversity: an evolutionary consequence of deception? *Trends in Ecology & Evolution* 20:487–494 DOI [10.1016/j.tree.2005.06.004](https://doi.org/10.1016/j.tree.2005.06.004).
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013.** STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21 DOI [10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635).

- Edwards KD, Bombarely A, Story GW, Allen F, Mueller LA, Coates SA, Jones L. 2010.** TobEA: an atlas of tobacco gene expression from seed to senescence. *BMC Genomics* **11**:142 DOI [10.1186/1471-2164-11-142](https://doi.org/10.1186/1471-2164-11-142).
- Griesbach RJ. 2002.** Development of *Phalaenopsis* orchids for the mass-market. In: Janick J, Whipkey A, eds. *Trends in new crops and new uses*. Alexandria: ASHS Press, 458–465.
- Hardcastle TJ, Kelly KA. 2010.** baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* **11**:22 DOI [10.1186/1471-2105-11-422](https://doi.org/10.1186/1471-2105-11-422).
- Hsu C-C, Chen W-H. 2016.** The breeding achievements from *Phalaenopsis equestris*. In: Khew GS, ed. *Malayan orchid review*. Vol. 49. Singapore: The Orchid Society of South-East Asia, 41–47 DOI [10.1142/9789814749954_0007](https://doi.org/10.1142/9789814749954_0007).
- Hsu C-C, Su C-J, Jeng M-F, Chen W-H, Chen H-H. 2019.** A *HORT1* retrotransposon insertion in the *PeMYB11* promoter causes harlequin/black flowers in *Phalaenopsis* orchids. *Plant Physiology* **180**:1535–1548 DOI [10.1104/pp.19.00205](https://doi.org/10.1104/pp.19.00205).
- Kaessmann H. 2010.** Origins, evolution, and phenotypic impact of new genes. *Genome Research* **20**:1313–1326 DOI [10.1101/gr.101386.109](https://doi.org/10.1101/gr.101386.109).
- Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA. 2016.** A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *The Plant Journal* **88**:1058–1070 DOI [10.1111/tpj.13312](https://doi.org/10.1111/tpj.13312).
- Klepikova AV, Penin AA. 2019.** Gene expression maps in plants: current state and prospects. *Plants* **8**:309 DOI [10.3390/plants8090309](https://doi.org/10.3390/plants8090309).
- Koh KW, Lu H-C, Chan M-T. 2014.** Virus resistance in orchids. *Plant Science* **228**:26–38 DOI [10.1016/j.plantsci.2014.04.015](https://doi.org/10.1016/j.plantsci.2014.04.015).
- Kuo S-Y, Hu C-C, Huang Y-W, Lee C-WChin-Wei, Luo M-J, Tu C-W, Lee S-C, Lin N-S, Hsu Y-H. 2021.** Argonaute 5 family proteins play crucial roles in the defence against Cymbidium mosaic virus and Odontoglossum ringspot virus in *Phalaenopsis aphrodite* subsp. *formosana*. *Molecular Plant Pathology* **22**:627–643 DOI [10.1111/mpp.13049](https://doi.org/10.1111/mpp.13049).
- Li L, Foster CM, Gan Q, Nettleton D, James MG, Myers AM, Wurtele ES. 2009.** Identification of the novel protein QQS as a component of the starch metabolic network in *Arabidopsis* leaves. *The Plant Journal* **58**:485–498 DOI [10.1111/j.1365-3113X.2009.03793.x](https://doi.org/10.1111/j.1365-3113X.2009.03793.x).
- Lin Y-F, Chen Y-Y, Hsiao Y-Y, Shen C-Y, Hsu J-L, Yeh C-M, Mitsuda N, Ohme-Takagi M, Liu Z-J, Tsai W-C. 2016.** Genome-wide identification and characterization of *TCP* genes involved in ovule development of *Phalaenopsis equestris*. *Journal of Experimental Botany* **67**:5051–5066 DOI [10.1093/jxb/erw273](https://doi.org/10.1093/jxb/erw273).
- Love MI, Huber W, Anders S. 2014.** Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:550 DOI [10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8).

- Niu S-C, Xu Q, Zhang G-Q, Zhang Y-Q, Tsai W-C, Hsu J-L, Liang C-K, Luo Y-B, Liu Z-J. 2016. De novo transcriptome assembly databases for the butterfly orchid *Phalaenopsis equestris*. *Scientific Data* 3:160083 DOI 10.1038/sdata.2016.83.
- Nobuta K, Venu RC, Lu C, Beló A, Vemaraju K, Kulkarni K, Wang W, Pillay M, Green PJ, Wang G, Meyers BC. 2007. An expression atlas of rice mRNAs and small RNAs. *Nature Biotechnology* 25:473–477 DOI 10.1038/nbt1291.
- Penin AA, Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD. 2019. Comparative analysis of developmental transcriptome maps of *Arabidopsis thaliana* and *Solanum lycopersicum*. *Genes* 10:50 DOI 10.3390/genes10010050.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140 DOI 10.1093/bioinformatics/btp616.
- Schlötterer C. 2015. Genes from scratch –the evolutionary fate of de novo genes. *Trends in Genetics* 31:215–219 DOI 10.1016/j.tig.2015.02.007.
- Schug J, Schuller W-P, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ. 2005. Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biology* 6:R33 DOI 10.1186/gb-2005-6-4-r33.
- Sekhon RS, Briskine R, Hirsch CN, Myers CL, Springer NM, Buell CR, de Leon N, Kaepler SM. 2013. Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. *PLOS ONE* 8:e61005 DOI 10.1371/journal.pone.0061005.
- Silvera K, Santiago LS, Cushman JC, Winter K. 2009. Crassulacean acid metabolism and epiphytism linked to adaptive radiations in the orchidaceae. *Plant Physiology* 149:1838–1847 DOI 10.1104/pp.108.132555.
- Stelpflug SC, Sekhon RS, Vaillancourt B, Hirsch CN, Buell CR, de Leon N, Kaepler SM. 2016. An expanded maize gene expression atlas based on rna sequencing and its use to explore root development. *The Plant Genome* 9(1):plantgenome2015.04.0025 DOI 10.3835/plantgenome2015.04.0025.
- Su C-I, Chen W-C, Lee A-Y, Chen C-Y, Chang Y-CA, Chao Y-T, Shih M-C. 2013. A modified ABCDE model of flowering in orchids based on gene expression profiling studies of the moth orchid *Phalaenopsis aphrodite*. *PLOS ONE* 8:e80462 DOI 10.1371/journal.pone.0080462.
- Tsai C-C, Shih H-C, Wang H-V, Lin Y-S, Chang C-H, Chiang Y-C, Chou C-H. 2015. RNA-Seq SSRs of moth orchid and screening for molecular markers across genus *Phalaenopsis* (Orchidaceae). *PLoS ONE* 10:e0141761 DOI 10.1371/journal.pone.0141761.
- Valoroso MC, Sobral R, Saccone G, Salvemini M, Costa MMR, Aceto S. 2019. Evolutionary conservation of the orchid MYB transcription factors DIV, RAD, and DRIF. *Frontiers in Plant Science* 10:1359 DOI 10.3389/fpls.2019.01359.
- Wang L, Xie W, Chen Y, Tang W, Yang J, Ye R, Liu L, Lin Y, Xu C, Xiao J, Zhang Q. 2010. A dynamic gene expression atlas covering the entire life cycle of rice. *The Plant Journal* 61:752–766 DOI 10.1111/j.1365-313X.2009.04100.x.

- Xie C, Zhang YE, Chen J-Y, Liu C-J, Zhou W-Z, Li Y, Zhang M, Zhang R, Wei L, Li C-Y. 2012.** Hominoid-specific de novo protein-coding genes originating from long non-coding RNAs. *PLOS Genetics* **8**:e1002942 DOI [10.1371/journal.pgen.1002942](https://doi.org/10.1371/journal.pgen.1002942).
- Yao S, Jiang C, Huang Z, Torres-Jerez I, Chang J, Zhang H, Udvardi M, Liu R, Verdier J. 2016.** The *Vigna unguiculata* Gene Expression Atlas (VuGEA) from *de novo* assembly and quantification of RNA-seq data provides insights into seed maturation mechanisms. *The Plant Journal* **88**:318–327 DOI [10.1111/tpj.13279](https://doi.org/10.1111/tpj.13279).
- Zhang GQ, Liu KW, Li Z, Lohaus R, Hsiao YY, Niu SC, Wang JY, Lin YC, Xu Q, Chen LJ, Yoshida K, Fujiwara S, Wang ZW, Zhang YQ, Mitsuda N, Wang M, Liu GH, Pecoraro L, Huang HX, Xiao XJ, Lin M, Wu XY, Wu WL, Chen YY, Chang SB, Sakamoto S, Ohme-Takagi M, Yagi M, Zeng SJ, Shen CY, Yeh CM, Luo YB, Tsai WC, Van de Peer Y, Liu ZJ. 2017.** The *Apostasia* genome and the evolution of orchids. *Nature* **549**:379–383 DOI [10.1038/nature23897](https://doi.org/10.1038/nature23897).