bioRxiv preprint doi: https://doi.org/10.1101/2020.11.03.366260; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

1 # Transcriptome atlas of *Phalaenopsis equestris*
2
3 Anna V. Klepikova[1], Artem S. Kasianov[1], Margarita A. Ezhova[1], Aleksey A. Penin[1,2], Maria D.
4 Logacheva[1,2,3]
5
6 [1] Laboratory of Plant Genomics, Institute for Information Transmission Problems of the Russian
7 Academy of Sciences, Moscow, Russia
8 [2] Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia
9 [3] Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russia
10
11 Corresponding Author:
12 Anna V. Klepikova [1]
13 Bolshoy Karetny per. 19, build. 1, Moscow 127051, Russia
14 Email address: annklepikova@gmail.com
15
16 ## Abstract
17 The vast diversity of Orchidaceae together with sophisticated adaptations to pollinators and other
18 unique features make this family an attractive model for evolutionary and functional studies. The
19 sequenced genome of *Phalaenopsis equestris* facilitates Orchidaceae research. Here we present
20 an RNA-seq based transcriptome map of *P. equestris* which covers 19 organs of the plant
21 including leaves, roots, floral organs and shoot apical meristem. We demonstrated the high
22 quality of the data and showed the similarity of *P. equestris* transcriptome map with gene
23 expression atlases of other plants. The transcriptome map can be easily accessed through our
24 database Transcriptome Variation Analysis (TraVA) visualizing gene expression profiles. As an
25 example of the application we analyzed the expression of *Phalaenopsis* “orphan” genes – the
26 ones that do not have recognizable similarity with genes of other plants. We found that about a
27 half of them are not expressed; the ones that are expressed have a predominant expression
28 pattern in reproductive structures.
29
30 ## Introduction
31 The enormous diversity of orchids traditionally attracts attention of plant biologists. Orchidaceae
32 comprises about 25 thousand of species, which makes it the largest plant taxon (Cai et al., 2015).
33 The diversification of orchids has evolved along with complex pollinator-adapted flower
34 structure (Cozzolino & Widmer, 2005), CAM-photosynthesis and epiphytism (Silvera et al.,
35 2009).
36 Genome assembly of *Phalaenopsis equestris* (the horse phalaenopsis) (Cai et al., 2015) provided
37 novel opportunities for evolutionary and functional studies of Orchidaceae. Genome assembly
38 was used for the functional studies of transcription factor families (Lin et al., 2016; Valoroso et
39 al., 2019), somatic embryogenesis (Chen et al., 2019), retrotransposon insertions (Hsu et al.,
40 2019), as well as for evolutionary studies of ancient polyploidy (Barrett et al., 2019). However,

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.03.366260; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

41 transcriptome resources of *P. equestris* remain limited even though the de novo transcriptome
42 assembly was performed based on RNA sequencing of 11 organs (Niu et al., 2016).
43 In our study we present a transcriptome map of *P. equestris* consisting of 19 samples in two
44 biological replicates. High-quality RNA of orchid organs and tissues was sequenced using
45 Illumina technology resulting in 1 687 M reads. We compared expression characteristics of
46 *P. equestris* transcriptome map with gene expression atlases of other plants to provide evidence
47 of reliability of our data. Transcriptome map of *P. equestris* can be applied in a great variety of
48 functional studies.
49

50 Materials & Methods

51 Growing conditions

52 Plants were grown in a climate chamber under a 16 h light/8 h dark cycle at 22°C and 50–60%
53 relative humidity. Samples were collected in two biological replicates; each replicate consists of
54 at least seven plants. Sample collection was performed within two hours (Zeitgeber time ZT8-10)
55 to reduce the influence of the circadian cycle.

56 RNA extraction, library preparation and sequencing

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

65 Read mapping

66 Read trimming was performed using Trimmomatic version 0.36 (Bolger, Lohse & Usadel, 2014)
67 in a single read mode and parameters "ILLUMINACLIP:common.adapters.file:2:30:10
68 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:15 MINLEN:30". For read mapping
69 genome assembly and annotation of *P. equestris* from PLAZA database (version 4.5) was used.
70 Trimmed reads mapped on the genome assembly using Spliced Transcripts Alignment to a
71 Reference (STAR) version 2.4.2 (Dobin et al., 2013) in the "GeneCounts" mode and parameters
72 "--sjdbOverhang 59 --sjdbGTFfeatureExon exon --sjdbGTFtagExonParentTranscript gene_id" to
73 obtain counts of uniquely mapped reads on each gene.

74 Expression characteristics of transcriptome map

75 Gene read counts obtained with STAR were normalized on library size using size factors, as
76 described in (Anders & Huber, 2010). A threshold of five or higher normalized read counts in
77 each biological replicate was used to define expressed genes.
78 To describe gene expression pattern Shannon entropy values were calculated for expressed in at
79 least one sample genes (Schug et al., 2005). In order to avoid overrepresentation of certain plant
80 organs, the samples were grouped using distances on clustering tree: gene expression levels were

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.03.366260; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

81 averaged if samples had distance (1 - Pearson r^2) less than 0.1. Sample groups are listed in Table
82 S1.

83 **Data availability**

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

87

88 Results

89 **Transcriptome map construction**

90 Ornamental orchid *P. equestris* comprises three varieties and numerous hybrids of various flower
91 colors and sizes (Hsu & Chen, 2016). To create transcriptome atlas we chose *P. equestris* var.
92 blue (orchidee.su) as clonal plants are available for the cultivar which helps to reduce
93 interindividual variability. We have collected 31 samples covering main plant organs and
94 developmental stages such as roots, young and mature leaves, floral organs, flower buds, and
95 meristems. Each sample was collected in two biological replicates, and each replicate was pooled
96 from at least seven plants. Sample RNA was sequenced on Illumina platform resulting in 29 M -
97 65 M raw single reads (38 M median) for each sample (for sequencing statistics see Table S2).
98 After removing low-quality reads and technical sequenced 98.7-99.8% of reads remained (Table
99 S2).

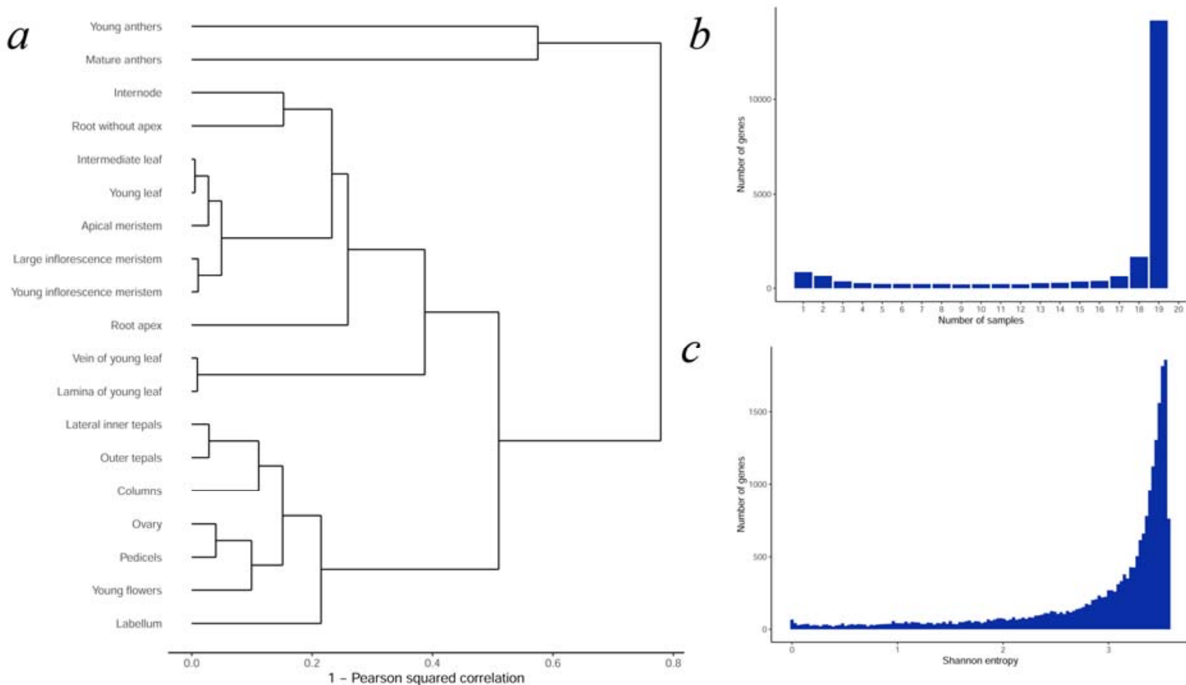
100 Reads were mapped on the reference genome of *P. equestris* (Cai et al., 2015) with only one
101 match allowed (unique mapping); 9.2-89.6% of high-quality reads were successfully mapped
102 (Table S2). 12 samples showed extremely low percentage of read mapping; unmapped reads
103 were identified as sequences belonging to Cymbidium mosaic virus (GenBank accession
104 MK816927) which are known to persist in the majority of *P. equestris* population and affect
105 mainly mature and senescent tissues (Koh, Lu & Chan, 2014). As the library size of infected
106 samples was insufficient and can distort the conclusions we excluded samples with a percentage
107 of mapped read lower than 35% in at least one biological replicate. The remained samples had
108 37.3-89.6% of uniquely mapped reads with median of 81.6%.

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

114 Clustering of samples generally reflects plant body plan and groups organs with similar
115 morphology and physiology (Klepikova & Penin, 2019). Hierarchical clustering of *P. equestris*
116 samples showed the same pattern (Fig. 1A). Sample clusters were formed by floral organs, leaf
117 parts, meristems and young leaves, inflorescence internode and root; young and mature anthers
118 were an outgroup for the other samples, similar to *A. thaliana*, rice, and maize (Nobuta et al.,
119 2007; Wang et al., 2010; Stelpflug et al., 2016; Klepikova et al., 2016). The distances between
120 samples on clustering tree were closer than in other species we observed (Klepikova et al., 2016;

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.03.366260; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

121 Penin et al., 2019), which can be explained by the lack of older tissues in *P. equestris*
122 transcriptome atlas.



123
124 **Figure 1.** Expression characteristics of the P. equestris transcriptome map: (a) Hierarchical
125 clustering tree of transcriptome map samples; (b) The distribution of genes by the number of
126 samples where gene is expressed. Only expressed genes with 5 or more normalized read counts
127 in each biological replicate were considered; (c) The distribution of Shannon entropy of P.
128 equestris genes.
129 We compare our samples with publicly available *P. equestris* transcriptomes (Table S5). In
130 general, the clustering of samples was consistent (Fig. S1), though leaf and column from the
131 BioProject PRJNA288388 (Niu et al., 2016) form outgroup to all other samples.

132 **Expression characteristics of *P. equestris***

133 *Phalaenopsis* genome annotation (PLAZA database, version 4.5) includes 29 431 protein-coding
134 genes. Among them 14 174 (48%) genes were expressed in all samples (using five reads in each
135 biological replicate as a threshold), when transcripts of 21 671 (74%) genes were found in at
136 least one sample. These values are in the range of typical expressed gene numbers across plant
137 transcriptome maps (Klepikova & Penin, 2019). As in other species, samples demonstrated
138 similarity in the number of expressed genes, which varied form 15 612 (53%) in shoot apical
139 meristem to 18 947 (64%) in ovules before pollination (Table S6).

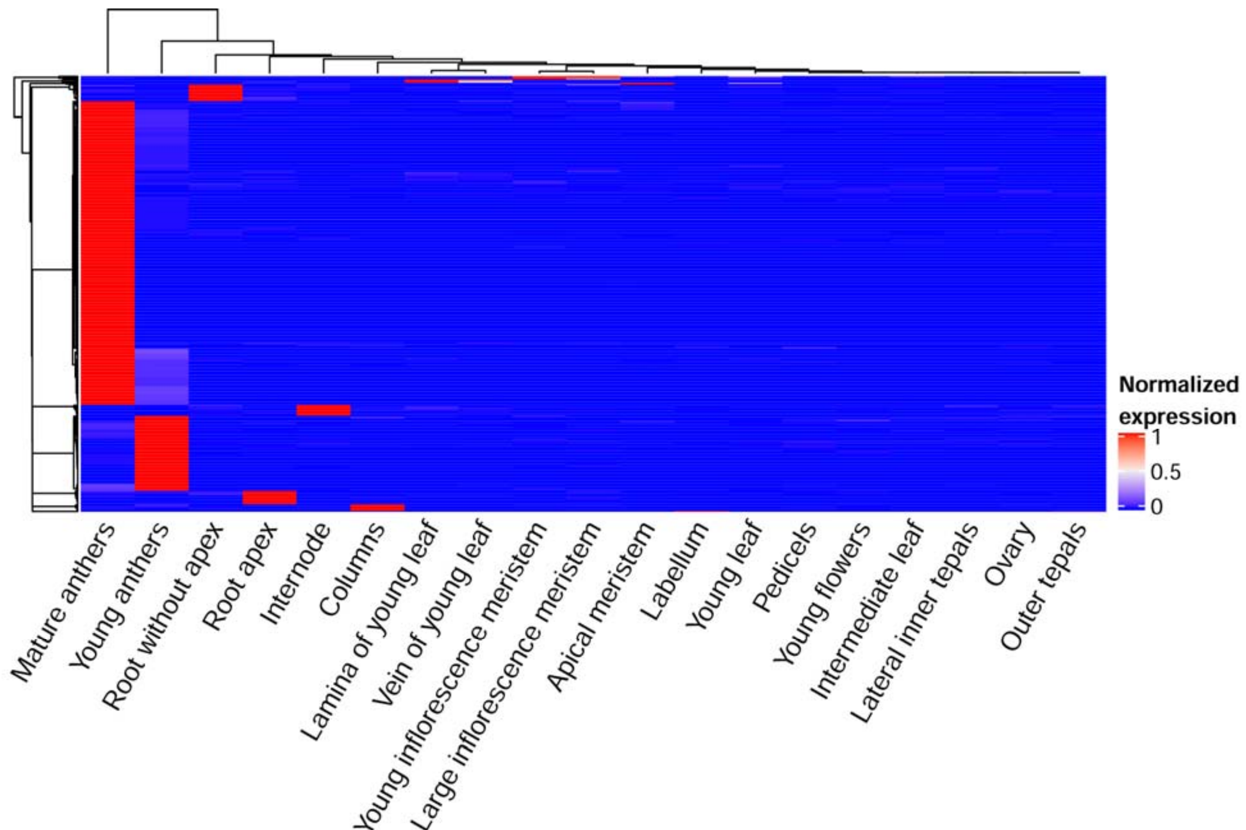
140 **Expression patterns of *P. equestris* genes**

141 The study of gene expression pattern can shed light on the biological function of the gene and
142 place it among essential for a plant existence ubiquitously expressed genes or precise regulators
143 of tissue features – sample-specific expressed genes. We used two approaches to define gene
144 expression patterns. A number of samples where gene is expressed is the simplest method to

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.03.366260; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

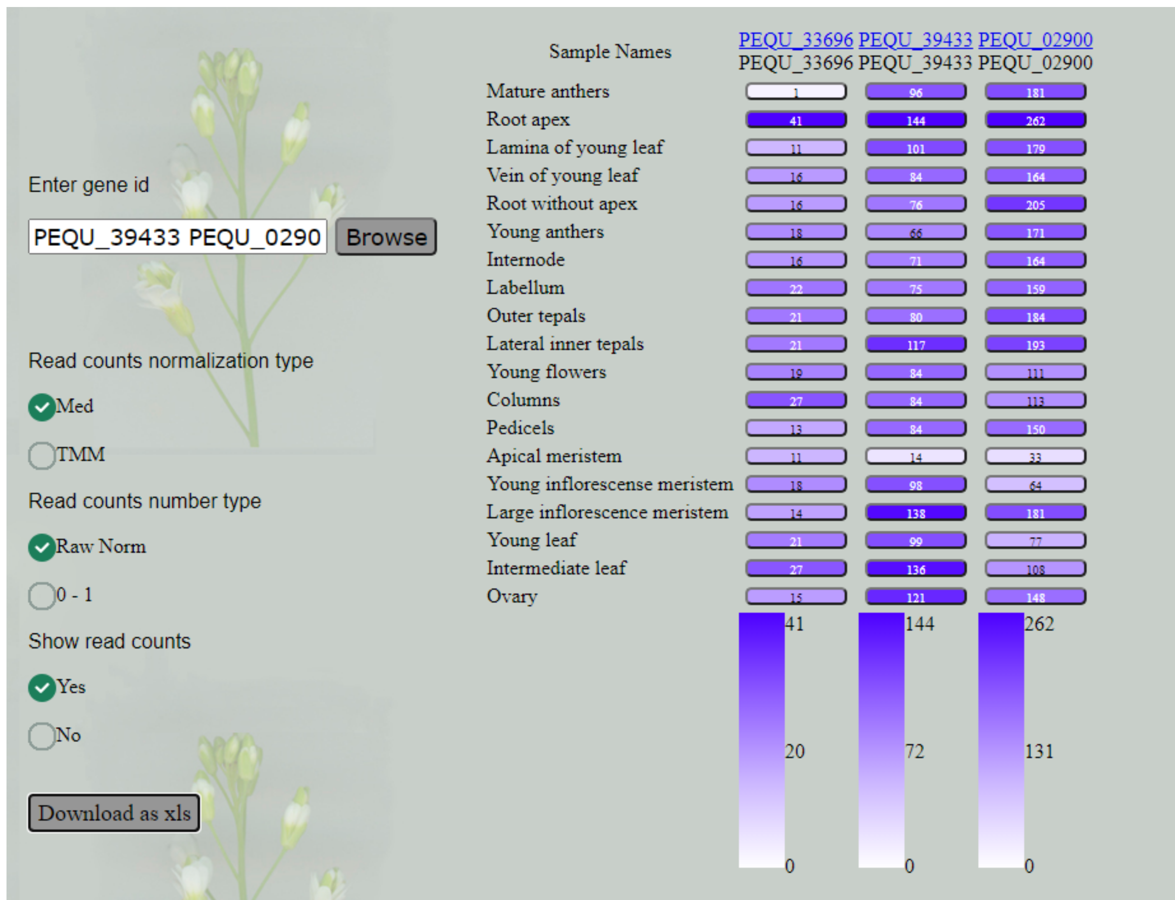
145 characterize expression pattern width, as was shown for *Nicotiana tabacum* (Edwards et al.,
146 2010) or *Vigna unguiculata* (Yao et al., 2016). The majority of genes (16 486) were expressed in
147 17 or more samples; the second peak (1 896 genes) of the distribution is formed by genes
148 expressed in 3 or less samples (Fig. 1B). The main patterns of tissue-specific genes were anthers
149 (56% of tissue-specific genes), roots (11%), and meristems (both shoot apical and inflorescence
150 meristem, 8%). The high number of anther-associated genes are known for *A. thaliana*
151 (Klepikova et al., 2016) and is expected for *P. equestris* as young and mature anthers are the
152 most distant samples on clustering tree (Fig. 1A).
153 While useful, such approach depends on an arbitrary threshold which separates expressed and
154 non-expressed genes and does not take into account the variation of expression level between
155 samples. To overcome the issue, we used Shannon entropy as a measure of expression pattern
156 width: low entropy values correspond to tissue-specific genes, while high values mark
157 ubiquitously expressed genes (Schug et al., 2005). The distribution of Shannon entropy in
158 *P. equestris* was significantly skewed to the right revealing major part of wide-expressed genes
159 (Fig. 1C) similarly to *A. thaliana*, *Solanum licopersicum* and *Zea mays* (Sekhon et al., 2013;
160 Klepikova et al., 2016; Penin et al., 2019).
161 Using Shannon entropy value lower than 0.25 we identified 521 tissue-specific genes. As in case
162 of direct count the majority of genes was associated with anthers or roots (Fig. 2). According to
163 GO enrichment, genes uniquely expressed in the mature anthers were involved in cell wall
164 organization, biogenesis and modification and had pectinesterase and enzyme inhibitor activity
165 (Table S7). Young anthers were characterized by genes encoding products with amine and amino
166 acid binding activity (Table S8). Root-specific genes (expressed in the sample “Root without
167 apex”) were described by terms “response to chemical stimulus”, “response to oxidative stress”,
168 “oxidation reduction”, and “heme binding” (Table S9).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.03.366260; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



169
170 **Figure 2.** The heatmap of tissue-specific genes. Expression levels of each gene in each sample
171 were normalized on its maximal expression level.
172 To find genes with opposite behaviour which uniformly expressed across tissues we selected
173 genes with Shannon entropy 3.55 or higher and calculated coefficient of variance (CV) as a
174 measure of expression stability. For 899 out of 1 340 genes CV was less than 0.25, indicating
175 uniform expression in all samples and biological replicates. Stable genes had GO enrichment in
176 terms associated with vesicles, membranes, RNA processing and localization. The list of GO
177 categories strongly overlapped with the enrichment of *A. thaliana* uniformly expressed genes
178 indicating inter-species universality of basic biological processes (Table S10).
179 ***P. equestris* Transcriptome Variation Database**
180 We aimed to make our transcriptome data easily accessible and ready to use, so we uploaded
181 *P. equestris* transcriptomes into our database Transcriptome Variation Analysis (TraVA,
182 http://travadb.org/browse/Species=Phalaenopsis_equestris/). TraVA interface demonstrates a
183 color chart of gene expression profiles in a single- or multiple-gene view. A user can prefer to
184 show or hide expression values in a chart and choose between several types of read count
185 normalization (Fig. 3).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.03.366260; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



186

187 **Figure 3.** Database view.

188 **The application of the TraVA database to the characterization of orchid genes**

189 Graphical interface of TraVA facilitate gene expression patterns analysis and comparison and
190 can be widely used in *P. equestris* functional studies. Orchids are a large and highly diverse plant
191 family whose species adapted to a number of ecological niches (typical terrestrial plants,
192 epiphytes, non-photosynthetic plants). These adaptations reflect in their genome – for example,
193 *P. equestris* which has sophisticatedly differentiated perianth the number of *AP3* orthologs is
194 higher compared to *Apostasia schenzhenica*, the basal orchid species with undifferentiated
195 perianth. Vice versa, *P. equestris* which is an epiphyte and does not develop typical terrestrial
196 roots, lacks *AGL12* and several genes of the *ANR1* clade, in contrast to *A. schenzhenica*. This
197 stresses the importance of the study of lineage-specific genes and gene families. For the genes
198 that do not have orthologs in model species the analysis of the expression profiles is the first step
199 towards functional characterization.
200 We identified 181 (160 after filtering of the proteins that had X on more than 50% of length)
201 *P. equestris* proteins that do not have significant similarity to any *Arabidopsis* protein (e-value
202 cut-off = 10). Out of them 118 share similarity with the proteins of *A. schenzhenica* and are thus
203 presumably orchid-specific while 42 have no hits and thus emerged after the divergence of

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.03.366260; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Apostasioideae and Epidendroideae. The survey of the expression profiles showed that 93 of them are not expressed in any of the samples of the map (Fig. S2). Among the ones which are expressed most are expressed at very low levels. Higher expression levels are associated with reproductive structures, in particular, anthers (Fig. 4).

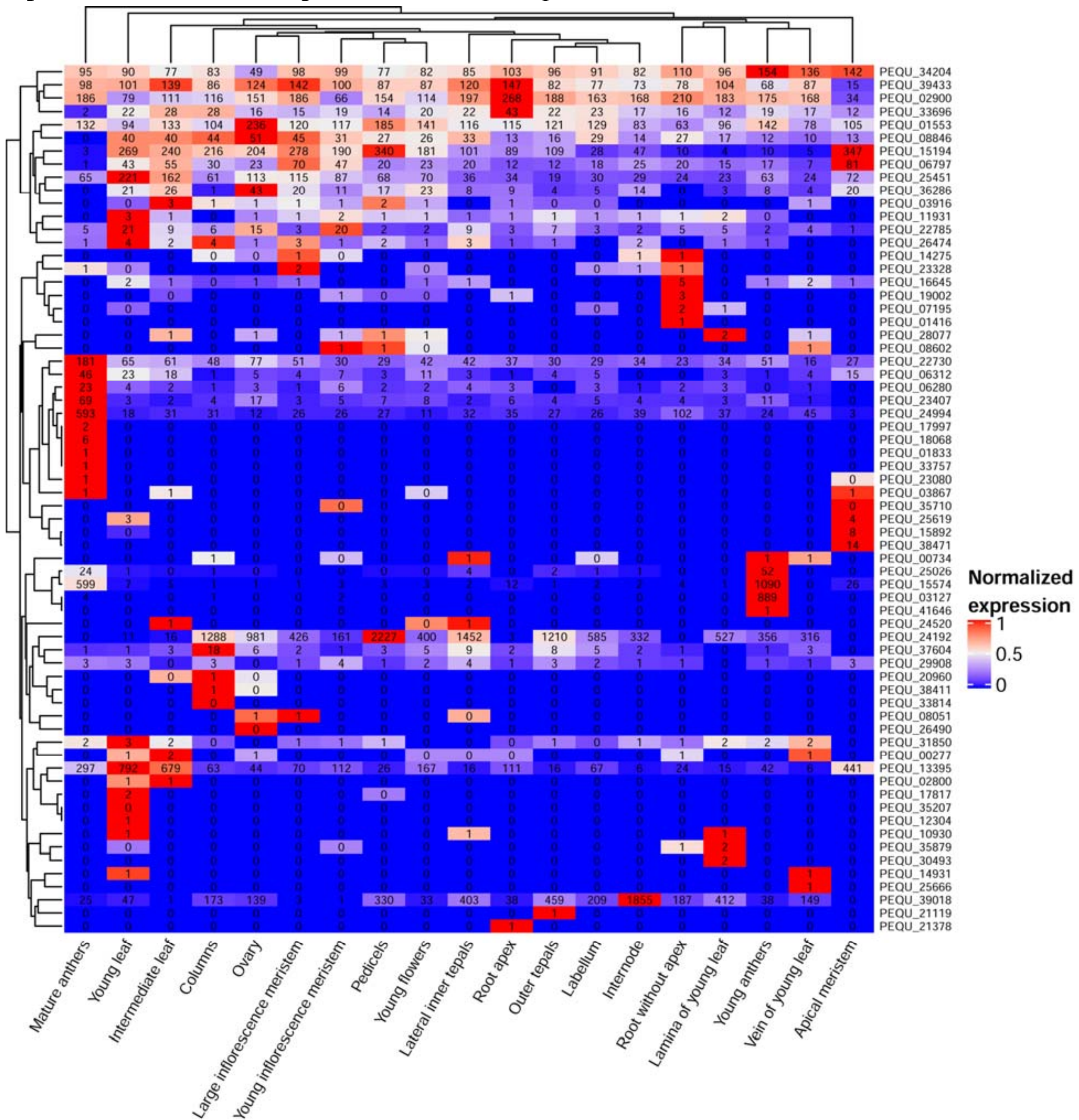


Figure 4. The heatmap of P. equestris-specific genes. Expression levels of each gene in each sample were normalized on its maximal expression level for the color key. The numbers on the figure represent normalized gene read count averaged over biological replicates.

Among vegetative structures the most distinct is root (root apex) where three genes – *PEQU_39433*, *PEQU_02900*, *PEQU_33696* – have the highest expression levels. *Phalaenopsis* roots are unique (compared to most other plants, including *A. schenzhenica*, but not to other

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.03.366260; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

215 epiphytic orchids) in many respects – in particular, they are photosynthetic and develop a special
216 structure called velamen. Velamen is a tissue of epidermal origin that consists of several layers
217 of dead cells which help to absorb water and protect photosynthetic tissues of the root from the
218 UV damage. Notably, *PEQU_39433* and *PEQU_33696* do not have homologs in
219 *A. schenzhenica*. *PEQU_02900* has a marginal similarity (34%) with *A. schenzhenica* protein
220 encoded by Ash001570 gene.
221 The topic of orphan genes – the ones that lack detectable homologues in other lineages - is
222 widely discussed, in particular in application to plants (Arendsee, Li & Wurtele, 2014). While a
223 part of the orphan genes might represent the artifacts of the annotation, others have a function
224 (for example, *A. thaliana* orphan gene *QQS* which acts in starch metabolism (Li et al., 2009).
225 The functional analysis of orphan genes however lags behind the typical genes; they are
226 overlooked in the annotations based on the homology; they are also usually expressed at lower
227 levels and in a narrower range of tissues (reviewed in Schlötterer, 2015). The study of expression
228 levels and patterns of a potential orphan gene is a first step towards its characterization – the
229 detectable level of expression is an evidence of that the ORF is indeed a gene, not an annotation
230 artifact.
231 Notably that orphan genes in the well-characterized animal objects (Drosophila, primates) have
232 expression patterns biased towards male reproductive structures (Begun et al., 2007; Xie et al.,
233 2012). According to the “out-of-testis” hypothesis (Kaessmann, 2010) this is mediated by the
234 unique epigenetic state of the chromatin during male gametogenesis. We observed the same bias
235 in *Phalaenopsis*; the growing availability of plant transcriptome maps will enable to find out if
236 this is universal for plants.

237 ## Conclusions

238 In this study we present a transcriptome map of orchid *Phalaenopsis equestris* covering 19
239 organs at various stages of the development. We identified 521 tissue-specific genes the majority
240 of which expressed in anthers, roots (11%), and meristems. The uniformly expressed genes were
241 associated with the similar processes as in *Arabidopsis thaliana*, i.e. vesicles, membranes, RNA
242 processing and localization. In order to improve reuse of the data we integrated transcriptome
243 map in our database TraVA and demonstrated its usability in the study of *P. equestris* orphan
244 genes.

245 ## Author Contributions

246 Conceptualization, A.A.P.; Data curation, A.V.K.; Formal analysis, A.V.K. and A.S.K.; Funding
247 acquisition, A.A.P.; Investigation, A.A.P., M.D.L. and A.V.K.; Methodology, A.A.P and
248 M.A.E.; Project administration, A.A.P.; Software, A.V.K. and A.S.K.; Supervision, A.A.P.;
249 Writing—original draft, A.V.K. and M.D.L..; Writing—review & editing, A.A.P., A.V.K. and
250 M.D.L.

251 ## Funding

252 The reported study was funded by RFBR according to the research project № 18-29-13017.

253 ## Conflicts of Interest

254 The authors declare no conflict of interest.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.03.366260; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

255

256 ## References

257 Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome*
258 *Biology* 11:R106. DOI: 10.1186/gb-2010-11-10-r106.
259 Arendsee ZW, Li L, Wurtele ES. 2014. Coming of age: orphan genes in plants. *Trends in Plant*
260 *Science* 19:698–708. DOI: 10.1016/j.tplants.2014.07.003.
261 Barrett CF, McKain MR, Sinn BT, Ge X-J, Zhang Y, Antonelli A, Bacon CD. 2019. Ancient
262 Polyploidy and Genome Evolution in Palms. *Genome Biology and Evolution* 11:1501–
263 1511. DOI: 10.1093/gbe/evz092.
264 Begun DJ, Lindfors HA, Kern AD, Jones CD. 2007. Evidence for *de Novo* Evolution of Testis-
265 Expressed Genes in the *Drosophila yakuba* / *Drosophila erecta* Clade. *Genetics* 176:1131–
266 1137. DOI: 10.1534/genetics.106.069245.
267 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
268 data. *Bioinformatics (Oxford, England)* 30:2114–2120. DOI:
269 10.1093/bioinformatics/btu170.
270 Cai J, Liu X, Vanneste K, Proost S, Tsai W-C, Liu K-W, Chen L-J, He Y, Xu Q, Bian C, Zheng
271 Z, Sun F, Liu W, Hsiao Y-Y, Pan Z-J, Hsu C-C, Yang Y-P, Hsu Y-C, Chuang Y-C,
272 Dievart A, Dufayard J-F, Xu X, Wang J-Y, Wang J, Xiao X-J, Zhao X-M, Du R, Zhang G-
273 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
274 Peer Y, Liu Z-J. 2015. The genome sequence of the orchid *Phalaenopsis equestris*. *Nature*
275 *Genetics* 47:65–72. DOI: 10.1038/ng.3149.
276 Chen J-C, Tong C-G, Lin H-Y, Fang S-C. 2019. *Phalaenopsis LEAFY COTYLEDON1*-Induced
277 Somatic Embryonic Structures Are Morphologically Distinct From Protocorm-Like
278 Bodies. *Frontiers in Plant Science* 10:1594. DOI: 10.3389/fpls.2019.01594.
279 Cozzolino S, Widmer A. 2005. Orchid diversity: an evolutionary consequence of deception?
280 *Trends in Ecology & Evolution* 20:487–494. DOI: 10.1016/j.tree.2005.06.004.
281 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
282 TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*
283 29:15–21. DOI: 10.1093/bioinformatics/bts635.
284 Edwards KD, Bombarely A, Story GW, Allen F, Mueller LA, Coates SA, Jones L. 2010.
285 TobEA: an atlas of tobacco gene expression from seed to senescence. *BMC Genomics*
286 11:142. DOI: 10.1186/1471-2164-11-142.
287 Hsu C-C, Chen W-H. 2016. The breeding achievements from *Phalaenopsis equestris*. In:
288 *Malayan Orchid Review*. WORLD SCIENTIFIC, 41–47. DOI:
289 10.1142/9789814749954_0007.
290 Hsu C-C, Su C-J, Jeng M-F, Chen W-H, Chen H-H. 2019. A *HORT1* Retrotransposon Insertion
291 in the *PeMYB11* Promoter Causes Harlequin/Black Flowers in *Phalaenopsis* Orchids.
292 *Plant Physiology* 180:1535–1548. DOI: 10.1104/pp.19.00205.
293 Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Research*
294 20:1313–1326. DOI: 10.1101/gr.101386.109.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.03.366260; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

295 Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA. 2016. A High
296 Resolution Map of the *Arabidopsis thaliana* Developmental Transcriptome Based on
297 RNA-seq Profiling. *The Plant Journal*. DOI: 10.1111/tpj.13312.
298 Klepikova AV, Penin AA. 2019. Gene Expression Maps in Plants: Current State and Prospects.
299 *Plants* 8:309. DOI: 10.3390/plants8090309.
300 Koh KW, Lu H-C, Chan M-T. 2014. Virus resistance in orchids. *Plant Science* 228:26–38. DOI:
301 10.1016/j.plantsci.2014.04.015.
302 Li L, Foster CM, Gan Q, Nettleton D, James MG, Myers AM, Wurtele ES. 2009. Identification
303 of the novel protein QQS as a component of the starch metabolic network in *Arabidopsis*
304 leaves. *The Plant Journal* 58:485–498. DOI: 10.1111/j.1365-313X.2009.03793.x.
305 Lin Y-F, Chen Y-Y, Hsiao Y-Y, Shen C-Y, Hsu J-L, Yeh C-M, Mitsuda N, Ohme-Takagi M,
306 Liu Z-J, Tsai W-C. 2016. Genome-wide identification and characterization of *TCP* genes
307 involved in ovule development of *Phalaenopsis equestris*. *Journal of Experimental Botany*
308 67:5051–5066. DOI: 10.1093/jxb/erw273.
309 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.
310 2016. De novo transcriptome assembly databases for the butterfly orchid *Phalaenopsis*
311 *equestris*. *Scientific Data* 3:160083. DOI: 10.1038/sdata.2016.83.
312 Nobuta K, Venu RC, Lu C, Beló A, Vemaraju K, Kulkarni K, Wang W, Pillay M, Green PJ,
313 Wang G, Meyers BC. 2007. An expression atlas of rice mRNAs and small RNAs. *Nature*
314 *Biotechnology* 25:473–477. DOI: 10.1038/nbt1291.
315 Penin AA, Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD. 2019. Comparative
316 Analysis of Developmental Transcriptome Maps of *Arabidopsis thaliana* and *Solanum*
317 *lycopersicum*. *Genes* 10:50. DOI: 10.3390/genes10010050.
318 Schlötterer C. 2015. Genes from scratch – the evolutionary fate of de novo genes. *Trends in*
319 *Genetics* 31:215–219. DOI: 10.1016/j.tig.2015.02.007.
320 Schug J, Schuller W-P, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ. 2005. Promoter
321 features related to tissue specificity as measured by Shannon entropy. *Genome Biology*
322 6:R33. DOI: 10.1186/gb-2005-6-4-r33.
323 Sekhon RS, Briskine R, Hirsch CN, Myers CL, Springer NM, Buell CR, de Leon N, Kaeppler
324 SM. 2013. Maize Gene Atlas Developed by RNA Sequencing and Comparative Evaluation
325 of Transcriptomes Based on RNA Sequencing and Microarrays. *PLoS ONE* 8:e61005.
326 DOI: 10.1371/journal.pone.0061005.
327 Silvera K, Santiago LS, Cushman JC, Winter K. 2009. Crassulacean Acid Metabolism and
328 Epiphytism Linked to Adaptive Radiations in the Orchidaceae. *Plant Physiology*
329 149:1838–1847. DOI: 10.1104/pp.108.132555.
330 Stelpflug SC, Sekhon RS, Vaillancourt B, Hirsch CN, Buell CR, de Leon N, Kaeppler SM. 2016.
331 An Expanded Maize Gene Expression Atlas based on RNA Sequencing and its Use to
332 Explore Root Development. *The Plant Genome* 9:0. DOI:
333 10.3835/plantgenome2015.04.0025.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.03.366260; this version posted November 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

334 Valoroso MC, Sobral R, Saccone G, Salvemini M, Costa MMR, Aceto S. 2019. Evolutionary
335 Conservation of the Orchid MYB Transcription Factors DIV, RAD, and DRIF. *Frontiers*
336 *in Plant Science* 10:1359. DOI: 10.3389/fpls.2019.01359.
337 Wang L, Xie W, Chen Y, Tang W, Yang J, Ye R, Liu L, Lin Y, Xu C, Xiao J, Zhang Q. 2010. A
338 dynamic gene expression atlas covering the entire life cycle of rice. *The Plant Journal*
339 61:752–766. DOI: 10.1111/j.1365-313X.2009.04100.x.
340 Xie C, Zhang YE, Chen J-Y, Liu C-J, Zhou W-Z, Li Y, Zhang M, Zhang R, Wei L, Li C-Y.
341 2012. Hominoid-Specific De Novo Protein-Coding Genes Originating from Long Non-
342 Coding RNAs. *PLoS Genetics* 8:e1002942. DOI: 10.1371/journal.pgen.1002942.
343 Yao S, Jiang C, Huang Z, Torres-Jerez I, Chang J, Zhang H, Udvardi M, Liu R, Verdier J. 2016.
344 The *Vigna unguiculata* Gene Expression Atlas (VuGEA) from *de novo* assembly and
345 quantification of RNA-seq data provides insights into seed maturation mechanisms. *The*
346 *Plant Journal* 88:318–327. DOI: 10.1111/tpj.13279.