Genomics Data 3 (2015) 106–111

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Transcriptome profiling of male gametophyte development in *Nicotiana tabacum*

Pavel Bokvaj, Said Hafidh, David Honys *

Laboratory of Pollen Biology, Institute of Experimental Botany, Academy of Sciences of the Czech Republic, Rozvojová 263, 165 02 Praha 6, Czech Republic

ARTICLE INFO

Article history:
Received 23 October 2014
Received in revised form 10 December 2014
Accepted 16 December 2014
Available online 23 December 2014

Keywords:
Pollen development transcriptome
Tobacco
Reproduction
Male gametophyte

ABSTRACT

Pollen, an extremely reduced bicellular or tricellular male reproductive structure of flowering plants, serves as a model for numerous studies covering wide range of developmental and physiological processes. The pollen development represents a fragile and vital phase of plant ontogenesis and pollen was among the first singular plant tissues thoroughly characterized at the transcriptomic level (Honys and Twell [5]). Arabidopsis pollen developmental transcriptome has been published over a decade ago (Honys and Twell, 2004) and transcriptomes of developing pollen of other species have followed (Rice, Deveshwar et al. [2]; Triticeae, Tran et al. [11]; upland cotton, Ma et al. [8]). However, the transcriptomic data describing the development of tobacco pollen, a bicellular model for cell biology studies, have been missing. Here we provide the transcriptomic data covering three stages (Tupý et al., 1983) of wild type tobacco (*Nicotiana tabacum*, cv. Samsun) pollen development: uninucleate microspores (UNM, stage 1), early bicellular pollen (eBCP, stage 3) and late bicellular pollen (lBCP, stage 5) as a supplement to the mature pollen (MP), 4 h-pollen tube (PT4), 24 h-pollen tubes (PT24), leaf (LF) and root (RT) transcriptomic data presented in our previous studies (Hafidh et al., 2012a; Hafidh et al., 2012b). We characterized these transcriptomes to refine the knowledge base of male gametophyte-enriched genes as well as genes expressed preferentially at the individual stages of pollen development. Alongside updating the list of tissue-specific genes, we have investigated differentially expressed genes with respect to early expressed genes. Pollen tube growth and competition of pollen tubes in female pistil can be viewed as a race of the fittest. Accordingly, there is an apparent evolutionary trend among higher plants to store significant material reserves and nutrients during pollen maturation. This supply ensures that after pollen germination, the pollen tube utilizes its resource predominantly for its rapid elongation in the female pistil. Previous transcriptomic data from *Arabidopsis* showed massive expression of genes encoding proteins forming both ribosomal subunits that were accumulated in developing pollen, whereas their expression was not detectable in growing pollen tubes (Honys and Twell, 2004). We observed a similar phenomenon in less advanced bicellular tobacco pollen. Here, we describe in detail how we obtained and analyzed validated microarray dataset deposited in Gene Expression Omnibus (GSE62349).

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Specifications	
Organism/tissue	*Nicotiana tabacum/developing pollen stages 1, 3, 5*
Sex	*Male*
Sequencer or array type	*Agilent Tobacco Gene Expression Microarray 4x44K*
Data format	*Raw and processed*
Experimental features	*Transcriptome profiling of male gametophyte development, analysis in context of previous mature pollen and pollen tubes data*
Consent	*n/a*

Direct link to deposited data

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62349

Experimental design, materials and methods

Tobacco immature pollen isolation

Wild-type *Nicotiana tabacum* (cv. Samsun) seeds were sown in a greenhouse under 22–25 °C and short-day conditions. Adult plants with fully developed roots were transplanted to an outdoor greenhouse to compost soil. There the plants were grown under the natural day–night photoperiod in spring and summer.

* Corresponding author. Tel.: +420 225 106 450; fax: +420 225 106 456.
E-mail address: david@ueb.cas.cz (D. Honys).

http://dx.doi.org/10.1016/j.gdata.2014.12.002
2213-5960/© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

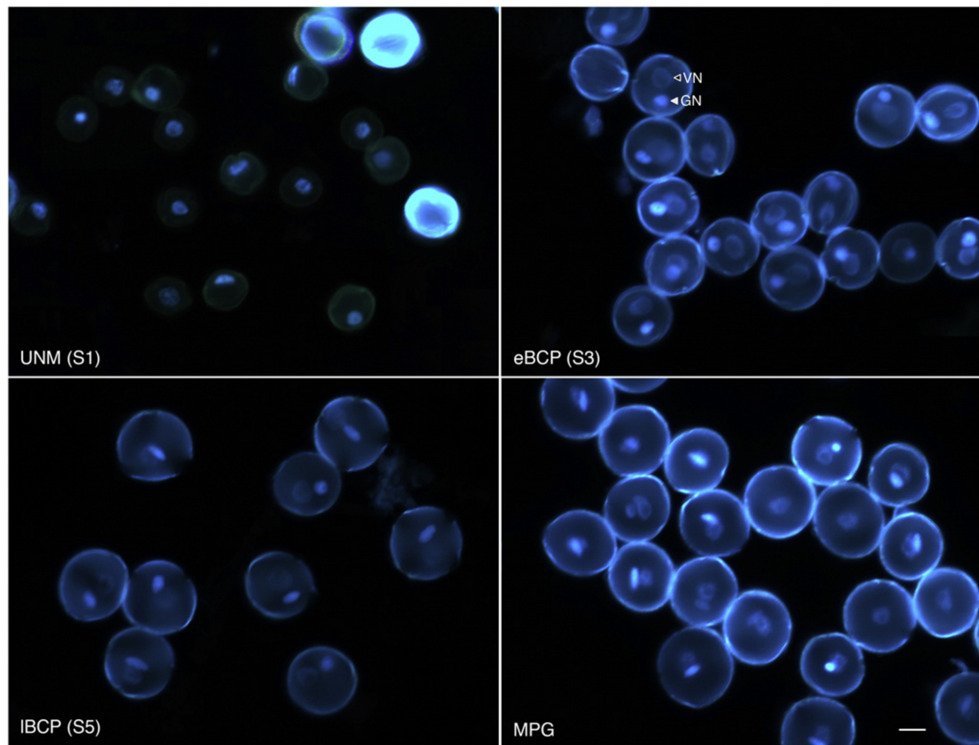


Fig. 1. DAPI-stained populations of developing spores used for the transcriptomic analysis: microspores (UNM); early bicellular pollen (eBCP); late bicellular immature pollen (lBCP); and mature pollen MPG. Scale bar = 10 μm. GN, VN stand for generative nucleus and vegetative nucleus, respectively.

Immature pollen grains at three stages of development were collected from August to September. The appropriate developmental stage was determined by the length of the flower buds including calyx as published previously by [12]. Flower bud lengths were as follows: uninuclear microspores (stage 1), 15–16 mm; early bicellular pollen (stage 3), 25–27 mm, late bicellular pollen (stage 5), 46–49 mm. After flower collection, freshly isolated anthers were immediately processed by gentle crushing in a chilled mortar with 5% sucrose in diethylpyrocarbonate (DEPC)-treated sterile water. The mixture containing released pollen was filtered through a nylon mesh of approximately 100 μm to remove anther debris. Suspended pollen grains were then sedimented by centrifugation (2000 *g*, 5 min, 4 °C), decanted excess supernatant and stored in −80 °C. For the assessment of the correct pollen developmental stage, standard 4′,6-diamidino-2 phenylindole (DAPI) staining in DAPI staining solution (0.1 M sodium phosphate, pH 7; 1 mM EDTA, 0.1% (v/v) Triton X-100, 0.8 mg/ml DAPI) was used ([9]; Fig. 1). Stained pollen grains were visualized with NIKON TE2000 fluorescence microscope and images were captured with NIS element software (NIKON).

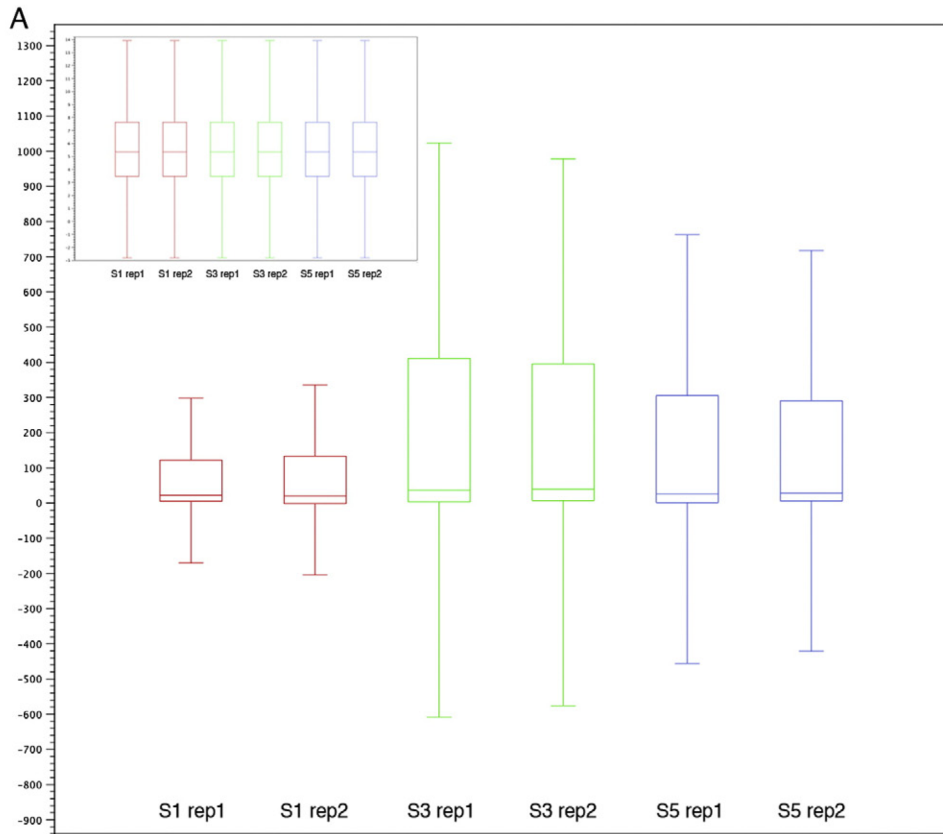
RNA extraction, purification, and quality verification

Total RNA was isolated from 100 mg of immature pollen using the Qiagen RNeasy Plant Kit according to the manufacturer's instructions (Qiagen, Valencia, CA) and treated with DNaseI (Promega, Madison, WI). RNA was quantified using NanoDrop (Thermo Scientific, Wilmington, USA). Prior to shipment, five replicates from each sample were tested using semi-quantitative RT-PCR with two marker genes, Nt-eIF5A and a constitutive 18S rRNA, for reproducibility. RNA concentration, purity and integrity (RIN) were assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Boblingen, Germany) at Imaxio (Clermond-Ferrand, France). In all samples, RNA concentration was higher than 94.54 ng/μl and RIN ranged between 8.20 and 10. Two independent RNA samples showing two intact bands corresponding to 18S and 28S were used for microarray analysis.

Experimental procedures for microarray analysis

Cyanine-3 (Cy3)-labeled cRNA was prepared from 66 ng of total RNA by Low Input Quick Amp Labeling Kit, One-Color (Agilent Technologies, Boblingen, Germany) according to the manufacturer's instruction. Cy3-labeled cRNA was purified by RNeasy Mini Kit (Qiagen, Valencia, CA). NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, Wilmington, USA) was used to evaluate the cRNA yield and labeling efficiency. In all samples, cRNA yield was higher than 8.34 μg and labeling efficiency was over 39.43 pmol Cy/μg cRNA. Cy3-labeled cRNA was then fragmented in a 30-minute incubation at 60 °C. After fragmentation, 1650 ng of Cy3-labeled cRNA was hybridized to Tobacco Gene Expression Microarray 4x44K (Platform GPL10098, Agilent Technologies, Boblingen, Germany) for 17 h at 65 °C in a rotating Agilent hybridization oven (Agilent Technologies, Boblingen, Germany). After washing the slides were scanned on the Agilent High Resolution Microarray Scanner using default parameters for 4x44K microarray formats (Agilent Technologies, Boblingen, Germany). Background-subtracted and

Fig. 2. (A) Box plot of the raw hybridization signal from the four samples, highlighting systematic differences between samples. The inset shows the same data after the normalization. S1 — uninucleate microspore (Stage 1), replicates 1 and 2 (rep1, rep2); S3 — early bicellular pollen (Stage 3); S5 — late bicellular pollen (Stage 5). (B) Outputs of the first and second principal component (PCA) analysis of the log2-transformed data sets. The largest and second-largest principal components (variability projection 1 and 2, respectively) are displayed in orthogonal directions, assessing the overall homogeneity between replicates and variability between samples of different tissue types as reflected in their grouping. S1 — uninucleate microspore (Stage 1); S3 — early bicellular pollen (Stage 3); S5 — late bicellular pollen (Stage 5); MPG — mature pollen grain; PT4 — 4-hr cultivated pollen tube; PT24 — 4-hr cultivated pollen tube; LEAF — leaves; ROOT — roots.


A
S1 rep1
S1 rep2
S3 rep1
S3 rep2
S5 rep1
S5 rep2

B
PT4
PT24
MPG
S5
S3
S1
ROOT
LEAF
Projection on 2
Projection on 1

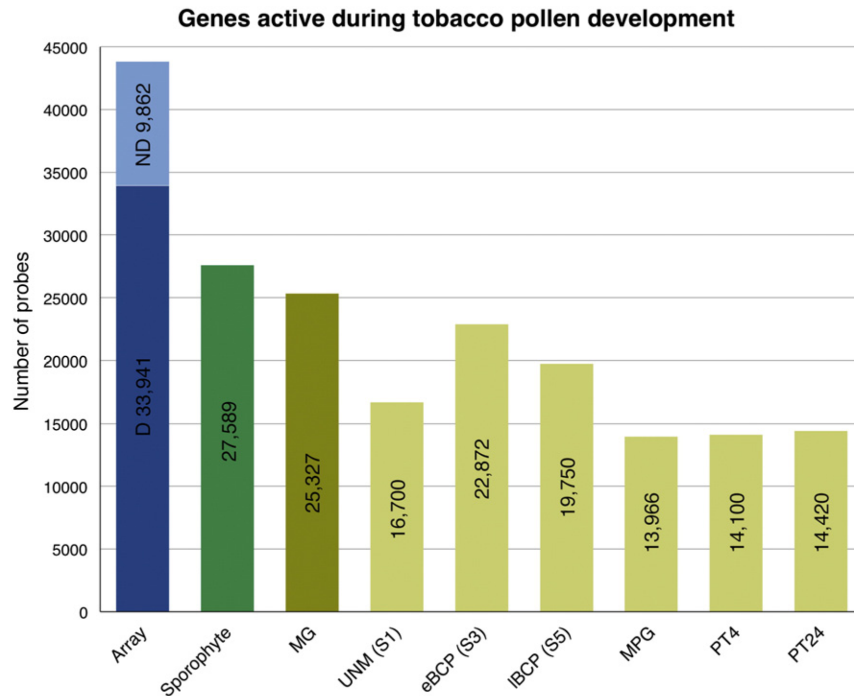


Fig. 3. Quantification of genes expressed in sporophyte and during male gametophyte development. Only genes showing well-above-background value "1" in both biological replicates were considered as reliably expressed. The first column shows the total number of probes represented on the Agilent 4x44K Tobacco Genome Array indicating probes that gave positive signal in at least one sample ("D") and those that showed no expression in any sample ("ND"). MG — male gametophyte; UNM (S1) — uninucleate microspore (Stage 1); eBCP (S3) — early bicellular pollen (Stage 3); lBCP (S5) — late bicellular pollen (Stage 5); MPG — mature pollen grain; PT4 — 4-hr cultivated pollen tube; PT24 — 24-hr cultivated pollen tube.

spatially-detrended Processed Signal intensities were obtained by Feature Extraction Software 11.0.1.1 (Agilent Technologies, Boblingen, Germany) used with default parameters (Protocol GE1_1100_Jul11 and Grid 21113_D_F_20130122).

Data processing and analysis

All transcriptomic data sets were normalized using publicly available dChip 1.3 software (http://www.hsph.harvard.edu/cli/complab/dchip/). The reliability and reproducibility of the analyses were ensured by the use of duplicates in each experiment, the normalization of all arrays to the median probe intensity level, and the use of normalized intensities of all arrays for the calculation of model-based gene expression values based on the Perfect Match-only model [7]. For each sample, only probes with the detection call of 'present' and an expression value detection level of 'well above background' (Boolean flag, two-sided t-test) in both replicates were considered to be expressed.

To determine the quality of the quantile-normalized data set and the correlation between arrays, CLC Main Workbench version 7.0.3 (CLC bio, Aarhus, Denmark) was used to compute mean expression values and corresponding p-values of the log_2-transformed data. The output of this analysis was used to assess the correlation between samples using PCA (Fig. 2). Newly submitted datasets as well as previously published ones [3,4] were used for this analysis and for the quantification of genes active in individual tobacco pollen developmental stages (Fig. 3).

Genes differentially expressed between arrays were statistically determined using the following criteria: for all filtered quantile-normalized datasets, (1) probes must have p-values of <0.05; (2) the p-values must pass FDR correction to minimize false discovery rate; (3) they should have an absolute expression value (fold change) of $\pm >3$; (4) and they must have expression levels of >100, well above background. All genes that passed the above test from pairwise comparisons are reported in Supplementary Tables 1–3.

Gene ontology term enrichment analysis of stage-enriched genes was performed using the AgriGO (http://bioinfo.cau.edu.cn/agriGO/) gene ontology database [13] supplemented by REVIGO (http://revigo.irb.hr/) visualization toolbox [10], together with application of the hypergeometric distribution test, and Benjamini and Honchberg corrected p-value cut off <0.05 [1]. GO terms from the categories 'biological processes', 'cellular components' and 'molecular function' with an adjusted p-value of <0.05 were considered over-represented in a subset of genes analyzed (Fig. 4, Supplementary Fig. 1). Specifically we looked at genes encoding ribosomal proteins that showed early expression profile in Arabidopsis [6] and observed similar expression pattern in less advanced bicellular tobacco pollen visualized by Mapman (http://mapman.gabipd.org/web/guest/mapman, Fig. 5, Supplementary Table 4).

Discussion

We extended our original transcriptomic analysis of tobacco male gametophyte development [3,4] and included three developmental stages of immature pollen. This also supplemented previously published pollen transcriptomic studies [2,5,8,11]. The analysis of normalized datasets revealed gene ontology categories specifically enriched in each developmental stage that corresponded to important cellular and molecular functions characteristic for these stages. We have also confirmed that the characteristic and correlated early expression pattern of genes encoding proteins forming ribosomal subunits.

Conflict of interest

The authors declare no conflicts of interest.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2014.12.002.

Acknowledgments

The authors gratefully acknowledge the financial support from the Czech Scientific Foundation (projects P501/11/1462, P305/12/2611, LD14109 and 13-06943S) and Ministry of Education Youth and Sports of the Czech Republic (LD13049).

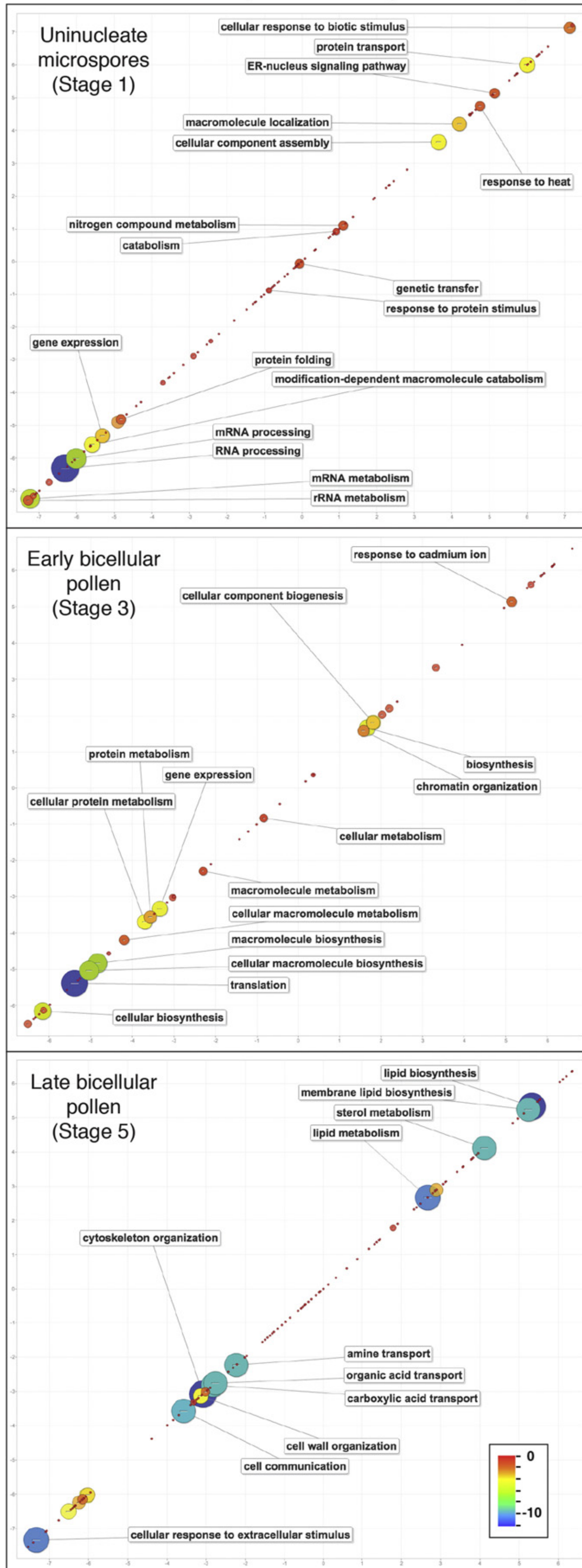


Fig. 4. Gene ontology terms from the 'biological processes' category with an adjusted p-value of <0.05 over-represented in subsets of genes enriched in three stages of tobacco pollen development. The size and color of circles reflects the $\log_{10}$ p-value of the individual GO terms (see inset).

References

[1] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Soc. Ser. B (Methodological) 57 (1) (1995) 289–300.

[2] P. Deveshwar, W.D. Bovill, R. Sharma, J.A. Able, S. Kapoor, Analysis of anther transcriptomes to identify genes contributing to meiosis and male gametophyte development in rice. BMC Plant Biol. 11 (2011) 78.

[3] S. Hafidh, K. Breznenová, D. Honys, De novo post-pollen mitosis II tobacco pollen tube transcriptome. Plant Signal. Behav. 7 (2012) 918–921.

[4] S. Hafidh, K. Breznenová, P. Růžička, J. Feciková, V. Čapková, D. Honys, Comprehensive analysis of tobacco pollen transcriptome unveils common pathways in polar cell expansion and underlying heterochronic shift during spermatogenesis. BMC Plant Biol. 12 (2012) 24.

[5] D. Honys, D. Twell, Comparative analysis of the *Arabidopsis* pollen transcriptome. Plant Physiol. 132 (2003) 640–652.

[6] D. Honys, D. Twell, Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. Genome Biol. 5 (2004) R85.

[7] C. Li, W.H. Wong, Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 31–36.

[8] J. Ma, H. Wei, M. Song, C. Pang, J. Liu, L. Wang, J. Zhang, S. Fan, S. Yu, Transcriptome profiling analysis reveals that flavonoid and ascorbate-glutathione cycle are important during anther development in Upland cotton. PLoS ONE 7 (2012) e49244.

[9] S.K. Park, R. Howden, D. Twell, The *Arabidopsis thaliana* gametophytic mutation gemini pollen1 disrupts microspore polarity, division asymmetry and pollen cell fate. Development 125 (1998) 3789–3799.

[10] F. Supek, M. Bošnjak, N. Škunca, T. Šmuc, REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS ONE (2011). http://dx.doi.org/10.1371/journal.pone.0021800.

[11] F. Tran, C. Penniket, R.V. Patel, N.J. Provart, A. Laroche, O. Rowland, L.S. Robert, Developmental transcriptional profiling reveals key insights into Triticeae reproductive development. Plant J. 74 (2013) 971–988.

[12] J. Tupý, J. Süss, E. Hrabětová, L. Říhová, Developmental changes in gene expression during pollen differentiation and maturation in *Nicotiana tabacum*. Biol. Plant. 25 (1983) 231–237.

[13] Zhou Du, Xin Zhou, Yi Ling, Zhenhai Zhang, Zhen Su, agriGO: a GO analysis toolkit for the agricultural community Nucleic Acids Research Advance Access published on July 1, 2010. Nucl. Acids Res. 38 (2010) W64–W70. http://dx.doi.org/10.1093/nar/gkq310.

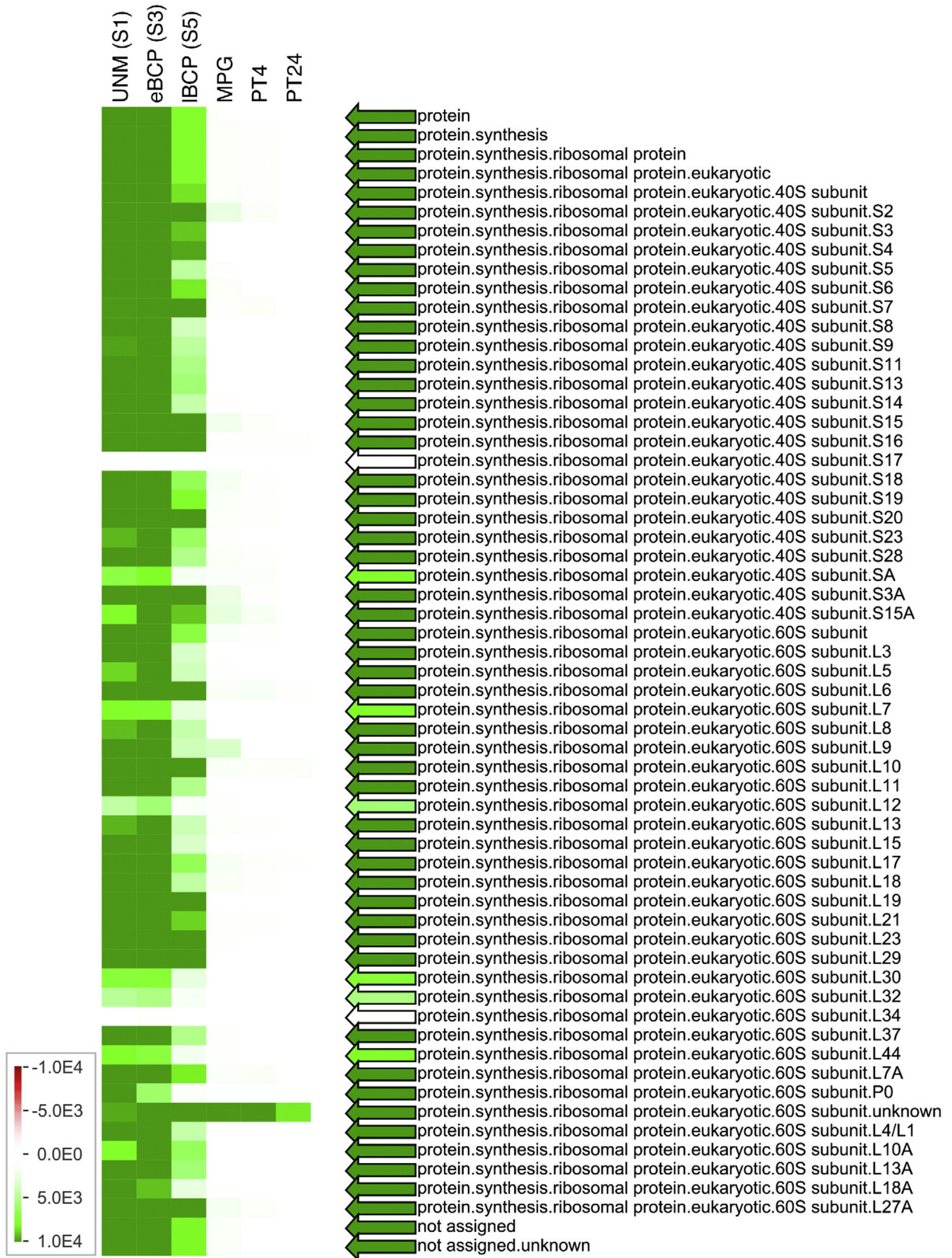


Fig. 5. Expression profiles of genes encoding ribosomal proteins in tobacco male gametophyte. Expression signal of each gene represent the mean value of two biological replicates (Supplementary Table 4). UNM (S1) — uninucleate microspore (Stage 1); eBCP (S3) — early bicellular pollen (Stage 3); lBCP (S5) — late bicellular pollen (Stage 5); MPG — mature pollen grain; PT4 — 4-hr cultivated pollen tube; PT24 — 4-hr cultivated pollen tube.