An R Package for Data Mining Chili Pepper Fruit Transcriptomes

Christian Escoto-Sandoval

CINVESTAV IPN: Centro de Investigacion y de Estudios Avanzados del Instituto Politecnico Nacional

Alan Flores-Díaz

CINVESTAV IPN: Centro de Investigacion y de Estudios Avanzados del Instituto Politecnico Nacional

M. Humberto Reyes-Valdés

Universidad Autonoma Agraria Antonio Narro

Neftalí Ochoa-Alejo

CINVESTAV: Centro de Investigacion y de Estudios Avanzados del Instituto Politecnico Nacional

Octavio Martinez (✉ octavio.martinez@cinvestav.mx)

Centro de Investigación y de Estudios Avanzados (Cinvestav) Unidad Irapuato <https://orcid.org/0000-0002-7590-0041>

Software article

Keywords: gene expression, RNA-Seq, Capsicum, transcriptome, fruit development

Posted Date: December 21st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-130806/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

SOFTWARE ARTICLE

An R package for data mining chili pepper fruit transcriptomes

Christian Escoto-Sandoval¹, Alan Flores-Díaz¹, M. Humberto Reyes-Valdés², Neftalí Ochoa-Alejo³ and Octavio Martínez^{1*}

*Correspondence:

octavio.martinez@cinvestav.mx

¹Unidad de Genómica Avanzada (Langebio), Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav), Irapuato, Guanajuato, México
Full list of author information is available at the end of the article

Abstract

Background: Open data sharing is instrumental for the advance of biological sciences. Gene expression is the primary molecular phenotype, usually estimated through RNA-Seq experiments. Large scope interpretation of RNA-Seq results is complicated by the extensive gene expression range, as well as by the diversity of biological sources and experimental treatments. Here we present “*Salsa*”, an auto-contained R package for extracting useful knowledge about gene expression during the development of chili pepper fruit.

Methods and Results: Data from 168 RNA-Seq libraries, comprising more than 3.4 billion reads, were analyzed and curated to represent standardized expression profiles (SEPs) for all genes expressed during fruit development in 12 chili pepper accessions. Accessions have representatives of domesticated varieties, wild ancestors and crosses, covering a broad spectrum of genotypes. Data are organized in a relational way, and functions allow data mining from the level of single genes up to whole genomes, grouping profiles by different criteria. Those include any combination of expression model, accession, protein description and gene ontology (GO) term, among others. Also, GO enrichment analysis can be performed over any set of genes.

Conclusions: “*Salsa*” opens endless possibilities for mining the transcriptome of chili pepper during fruit development.

Keywords: gene expression; RNA-Seq; *Capsicum*; transcriptome; fruit development

Background

Measurements of gene expression constitute the primary molecular phenotype. RNA-Seq experiments [1] allow genome-wide estimation of the relative level of gene expression in a particular specie, organ, tissue or even single cells [2]. Gene expression profiles through time estimate the transcriptome landscape of organ developing programs. Phenomena as seed development [3], senescence [4] and aging [5] had been shown to be conserved in plants. In particular, the development of fleshy fruits – an indispensable part of the human diet, is probably conserved throughout the angiosperms [6].

There are various databases to query gene expression profiles, as the NCBI Gene Expression Omnibus [7] or TiGER [8]; however, mining gene expression databases [9] remains a challenge [10], basically because the heterogeneity of organisms, experimental conditions and methods employed to obtain those profiles. Comparisons

of expression profiles between genes within a single experiment is also complicated by the fact that transcript abundance varies in orders of magnitude [11].

Here we present “*Salsa*”, an auto-contained R [12] package that allows genome-wide mining of a large collection of more than 313,000 standardized expression profiles (SEPs), representing expression profile changes during fruit development in chili pepper (*Capsicum annuum* L.). With this application you can look for sets of genes having a particular description or annotation, expressed in one or more accessions and following a specific expression pattern, etc. It also includes functions for the statistical analysis and visualization of SEPs, Gene Ontology (GO) enrichment analyses and web browsing facilities for genes and GO terms.

Even when “*Salsa*” is of primary interest for *Capsicum* research, it will also be useful for researchers interested in fruit development in other Solanaceae, or even to perform comparative analysis of fruit development with taxonomically distant species.

At first sight it might appear excessive to devote an R package to the data mining of a single data set. However, data can be mined with highly different emphases, looking for specific phenomena in the multidimensional space formed by the time profile of almost 30 thousand genes expressed in 12 genotypes of different origin. Considering and analyzing the data from different angles could provide novel insights for a better understanding of transcriptome’s complexity during fruit development.

Implementation

The main factors that hinder comparison between gene expression profiles are the heterogeneity of data sources –different species, organs, tissues, and treatments (environmental conditions, chemicals, mutants, time of development, etc.) as well as differences in data curation and statistical analyses. To alleviate these factors and achieve an equilibrium between generality and specificity, we focus in a set of 12 accessions of a single specie (*Capsicum annuum* L.), growing the plants under uniform conditions and sampling the fruits at fixed times, from flower to maturity [13].

To be able to compare expression profiles from different genes and accessions, we improved the method published in [14] to include a False Discovery Rate (FDR) [15] of approximately 1% when comparing any pair of expression profiles. The method only contrast adjacent time intervals, and determine if gene expression increases (I), decreases (D) or remains steady (S) by applying the method described in [16]. Given that 7 time points are taken into consideration (six time intervals), this categorization results in a total of $3^6 = 729$ discrete expression models. Finally, gene expression is standardized over time to produce “Standardized Expression Profiles” (SEPs). Full details of the method are presented in [13] and commented in Additional file 1.

Data in “*Salsa*” follow the relational database paradigm [17]; gene identifier (“id”) and three other variables link the seven `data.frames`, providing an efficient framework for data querying (for details see section 2.1 in Additional file 1).

At the core of the package is the function “`get.SEP`”, which allows the selection of sets of expression profiles using general and flexible criteria. These criteria include

the selection of genes by description, accessions, expression model, time of maximum expression and expression level (for details see section 2.2 in Additional file 1). Having selected one or more sets of SEPs the user can summarize, plot and analyze the SEPs groups. Summarization and plotting include the calculation of confidence intervals for each one of the expression times, which grant evaluation of differences in relative expression levels among SEPs at each one of the seven time points sampled.

The problem of testing global differences between two SEPs is solved in function “`analyze.2.SEPs`” by estimating Euclidean distances between and within each one of the SEPs and comparing the mean of the distances between the two SEPs with the mean of the distances within each SEP *via* a t-test. This approach reduces the dimension of the problem of simultaneously testing seven time points to the uni-dimensional comparison of two means, offering a powerful approach to decide if two SEPs can be considered equal.

Additional data mining facilities in “*Salsa*” include summarization of expression profile for each one of the genes, web browsing per gene or GO term as well as GO enrichment analysis. GO enrichment is implemented for arbitrary sets of genes in a single GO term (function `analyze.GO`) or for the whole set of GO terms in a single aspect (function `analyze.all.GO`). GO enrichment is performed using Fisher’s exact test, and a filter includes the calculation of FDR when all terms in a GO aspect are analyzed.

“*Salsa*” is implemented in R ($\geq 3.4.4$) and thus it is platform-independent and does not have any restrictions to be used. The binary file to install the package is presented as Additional file 2.

Results and Discussion

“*Salsa*” offers many opportunities to find interesting insights in the chili pepper transcriptome during fruit development. *A priori*, it is difficult to delimitate such universe of possibilities, but it involves as a first step to find an ‘interesting’ set of genes to study. Further steps will involve a more detailed analysis of that gene set, guided by the biological intuition of the researcher.

Two examples of interesting sets of genes that could be analyzed are: 1) genes with a highly concordant and specific expression profile within one or more accessions, but with different and also concordant expression profiles in other accessions; 2) genes associated with a particular GO biological process with contrasting expression patterns between accessions or groups of accessions. In fact, the possibilities of analyses are limitless, and depend on the researcher’s interest.

We decided to illustrate, as an example of data mining, the comparison of expression profiles between two accessions. R code and detailed explanations are presented in section 3 of Additional file 1, while here we present and explain the significance of the results through the figures produced by the package.

Given that the main point of the example is to show the package possibilities, we are not going to discuss in depth the biological implications of findings; in [13] we have described the way to discover interesting facts by employing a set of R data and functions that culminated in the “*Salsa*” package after careful generalization and testing.

Comparing gene expression profiles between accessions with contrasting fruit size.

We begin our analysis by isolating, as separate SEP `data.frames`, the genes expressed in accessions, “AS” (Ancho San Luis), of the domesticated accession set, which produces very large and moderately pungent fruits and “SR” (Sonora Red), a wild accession with very small and highly pungent fruits.

The majority, $> 89\%$, of the genes were consistently expressed in both accessions, while small percentages, $< 3\%$ and $< 7\%$ of the total number of genes, were exclusively expressed in “AS” and “SR”, respectively. Figure 1 presents the plot of the average SEPs for each accession, as well as for the set formed by both of them.

Figure 1 shows that the mean SEPs in “AS” and “SR” significantly differ at some points of the fruit development. For this plot we employed a very stringent threshold for the estimation of confidence intervals (CI) for the means; an Error Type I of $\alpha = 1 \times 10^{-4}$, which implies a 99.99% of confidence. CIs for the mean of each group at each time are shown as thin vertical lines over the circles that stand for the means per time and accession group. Looking at the CIs, we see that the mean SEPs of “AS” and “SR” are highly different at time points 10, 20, 40 and 50 DAA. It is important to consider that the plot of average SEPs, as the one presented in Figure 1 for “AS” (red line) and “SR” (blue line), does not indicate uniformity of expression profiles for individual genes; in fact, the mean of the SEPs hides the large diversity of individual expression profiles among genes (see Figure 3 in Additional file 1).

Finding sets of genes with divergent expression between the two accessions.

The divergence of mean SEP expression between “AS” and “SR”, observed in Figure 1, entails high differences between the transcriptomes of the two accessions; in particular, the peak of mean expression is found at 10 DAA for “AS”, while for “SR” it happens ten days later, at 20 DAA. Peak of mean expression signals maximum transcriptional activity, and is a hallmark in time for each individual gene.

To dissect transcriptome differences between “AS” and “SR”, we selected the sets of genes with simultaneous peak expression at each one of the seven sampled time points. This produces a total of $7 \times 7 = 49$ gene sets (see Box 3 in Additional file 1). Figure 2 shows the matrix of percentages of genes with peak expression at each one of the times.

The total of genes expressed in both accessions was 24720. Of these, 3672, representing a proportion of $3672/24720 \approx 0.1485$ or 14.85%, have their peak expression at 0 DAA in both accessions. That figure is presented in the bottom left hand-side of the matrix in Figure 2. The green dashed line in Figure 2 is over the cases where the peak expression coincides in time in both accessions, and we can see that, except for 0, 10 and 60 DAA, the corresponding percentages are small (less than 2%), which partially explains the differences between mean SEPs observed in Figure 1.

Gene sets that are out the dashed green diagonal of Figure 2 are “interesting”, because they present a pattern where peak expression is out of phase. One of the two sets of genes presenting the highest possible phase difference is the one formed by the 758 genes ($\approx 3.07\%$ of the total; top left hand-side corner in Figure 2) which peak at 0 DAA in “AS” (*X*-axis) while having such maximum at 60 DAA in “SR” (*Y*-axis).

Analysis of the “ASmOSRm60” gene set.

To further show “*Salsa*” capabilities, we performed an in-depth analysis of the set of 758 genes ($\approx 3.07\%$ of the total), which presents its maximum mean expression at the mature flower (0 DAA) in accession “AS”, while having such peak at the mature fruit (60 DAA) in “SR”. We denote that gene set as “ASmOSRm60”. Figure 3 presents the mean expression SEPs in the ASmOSRm60 set.

In Figure 3 we can notice the high difference in phase and contrasting mean standardized expression in the set ASmOSRm60. Mean SEP in “AS” (red line) presents a high peak at 0 DAA, suddenly decreasing from 0 to 10 DAA and then presenting a relatively steady state from 10 up to 60 DAA, forming an ‘L’ shaped expression profile. On the other hand, mean SEP for “SR” (blue line) presents an almost mirrored L shape with a local maximum at 0 DAA and then decreasing from 0 to 10 DAA –where the global minimum of mean expression is reached. From 20 DAA up to 50 DAA mean expression in “SR” stays relatively steady, suffering then a sudden increase to reach the peak of mean expression at 60 DAA. 53 of the 758 genes in ASmOSRm60 ($\approx 6\%$) are transcription factors (TF), and Figure 8 in Additional file 1 shows that these genes display a highly significant difference between “AS” and “SR” only at 0 and 60 DAA, showing a low and not significant steady state between 10 and 50 DAA.

The expression pattern of ASmOSRm60 genes is intriguing because it reverses peak expression from the first stage of fruit development –the mature flower at 0 DAA in domesticated accession “AS”, to the last stage –fully mature fruit at 60 DAA in wild accession “SR”. To understand the biological relevance of this set of genes we performed GO enrichment analyses by running function “`analyze.all.GO`” with a FDR threshold of 10% with categories Biological Process (BP), Cell Component (CC) and Molecular Function (MF).

In the first row of Table 1 we can see that a total of 106 genes from the set of 758 in “ASmOSRm60”, i.e., $\approx 14\%$, are annotated in the BP ‘Transport’ (GO:0006810), while the expected number of such genes under the independence hypothesis is only 70. This implies that small molecule transport is higher in the mature flower (0 DAA) in the domesticated accession “AS”, while conversely it is higher in the mature fruit (60 DAA) in the wild accession “SR” (see Figure 3).

We are not going to extend here the discussion of the biological relevance of the results in Table 1; nonetheless, it must be clear that “*Salsa*” capabilities grant detailed and deep mining of the chili pepper transcriptome during fruit development (see Additional file 1 for more details).

For example, the function “`gene.summary`” gives a plot as well as a numeric summary of mean SEPs for any one of the 29946 genes present in the data. The gene with identifier 3018 was one of the transcription factors found within the “ASmOSRm60” set, and Figure 4 shows the plot resulting from running “`gene.summary(id = 3018)`”.

In Figure 4 we observe that the mean expression per set of SEPs for that gene resembles the mean expression of the genes in the “ASmOSRm60” set, shown in Figure 3. That correspondence is apparent in that the TF with identifier 3018 presents a higher expression in the domesticated set “D” at 0 DAA –set which includes accession “AS”, while presenting a higher expression in the set “W” at 60 DAA

–set which includes accession “SR”; i.e., the expression of gene with identifier 3018 has in general (and not only for accessions “AS” and “SR”) a reversed expression pattern in “D” and “W” accessions.

For brevity we have deepened in the mining of only one of the $49 - 7 = 42$ gene sets that diverge in peak expression between the two accessions (see Figure 2). Certainly, more profound and also more varied searching can be tried with innumerable initial selections of sets of accessions, expression times, expression levels or gene categories. A possibility is to widen the analysis by using the function “`browse.gene()`” (see Figure 10 in Additional file 1). With that function the user can look for orthologous to the *Capsicum* genes, and link to other genomic or transcriptomic databases. Also in “*Salsa*” it is easy to obtain the structure of networks of co-related expression, which can then be studied by graph theory [18] or plot with specialized software [19], as for example the “igraph” R package [20].

Conclusions

We have shown that “*Salsa*” opens endless options for mining the transcriptome of chili pepper during fruit development. *In silico* analyses can suggest interesting hypotheses, which could then be experimentally tested; for example, one possibility is to use SEP similarity to predict a small set of transcription factors candidates to be regulating a given gene.

Acknowledgements

We thank Magda L. Arce-Rodríguez, Fernando Hernández-Godínez, Felipe Cervantes-Hernández, Corina Hayano-Kanashiro, José J. Ordaz-Ortiz and Ana Garcés-Claver for help in the obtention and processing of the data. C.E-S acknowledge Conacyt scholarship for PhD studies (grant number 630487).

Funding

This research was funded by the Consejo Nacional de Ciencia y Tecnología, México (Conacyt project 1570).

Availability of data and materials

All data and functions are included in Additional file 2 as an R package.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Conceptualization, O.M.; Methodology O.M, N.O-A, M.H.R-V, A.F-D and C.E-S; Formal Analysis, O.M, M.H.R-V and C.E-S; Investigation, O.M and C.E-S; Data Curation O.M and C.E-S; Writing –Original Draft O.M; Writing –Review and Editing, O.M, N.O-A, C.E-S, A.F-D and M.H.R-V; Visualization O.M and C.E-S; Supervision O.M and N.O-A; Project Administration N.O-A; Funding Acquisition N.O-A, O.M and M.H.R-V.

Author details

¹Unidad de Genómica Avanzada (Langebio), Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav), Irapuato, Guanajuato, México. ²Department of Plant Breeding, Universidad Autónoma Agraria Antonio Narro, Saltillo, Coahuila, México. ³Departamento de Ingeniería Genética, Unidad Irapuato, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav), Irapuato, Guanajuato, México.

References

- Wang, Z., Gerstein, M., Snyder, M.: RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics* **10**(1), 57–63 (2009)
- Luecken, M.D., Theis, F.J.: Current best practices in single-cell rna-seq analysis: a tutorial. *Molecular systems biology* **15**(6), 8746 (2019)
- Niu, Y., Wu, G.-Z., Ye, R., Lin, W.-H., Shi, Q.-M., Xue, L.-J., Xu, X.-D., Li, Y., Du, Y.-G., Xue, H.-W.: Global analysis of gene expression profiles in brassica napus developing seeds reveals a conserved lipid metabolism regulation with arabidopsis thaliana. *Molecular plant* **2**(5), 1107–1122 (2009)
- Noh, Y.-S., Amasino, R.M.: Regulation of developmental senescence is conserved between arabidopsis and brassica napus. *Plant Molecular Biology* **41**(2), 195–206 (1999)
- McCarroll, S.A., Murphy, C.T., Zou, S., Pletcher, S.D., Chin, C.-S., Jan, Y.N., Kenyon, C., Bargmann, C.I., Li, H.: Comparing genomic expression patterns across species identifies shared transcriptional profile in aging. *Nature genetics* **36**(2), 197–204 (2004)

6. Karlova, R., Chapman, N., David, K., Angenent, G.C., Seymour, G.B., de Maagd, R.A.: Transcriptional control of fleshy fruit development and ripening. *Journal of Experimental Botany* **65**(16), 4527–4541 (2014)
7. Clough, E., Barrett, T.: The gene expression omnibus database. In: *Statistical Genomics*, pp. 93–110. Springer, ??? (2016)
8. Liu, X., Yu, X., Zack, D.J., Zhu, H., Qian, J.: Tiger: a database for tissue-specific gene expression and regulation. *BMC bioinformatics* **9**(1), 1–7 (2008)
9. Creighton, C., Hanash, S.: Mining gene expression databases for association rules. *Bioinformatics* **19**(1), 79–86 (2003)
10. Maleki, F., Ovens, K., Hogan, D.J., Kusalik, A.J.: Gene set analysis: Challenges, opportunities, and future research. *Frontiers in genetics* **11**, 654 (2020)
11. Holland, M.J.: Transcript abundance in yeast varies over six orders of magnitude. *Journal of Biological Chemistry* **277**(17), 14363–14366 (2002)
12. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2013). R Foundation for Statistical Computing. <http://www.r-project.org>
13. Martínez, O., Arce-Rodríguez, M.L., Hernández-Godínez, F., Escoto-Sandoval, C., Cervantes-Hernández, F., Hayano-Kanashiro, C., Ordaz-Ortiz, J.J., Reyes-Valdés, M.H., Razo-Mendivil, F.G., Garcés-Claver, A., Ochoa-Alejo, N.: Transcriptomic analyses throughout chili pepper fruit development reveal novel insights into domestication process. *bioRxiv* (2020). doi:10.1101/2020.10.05.326470. <https://www.biorxiv.org/content/early/2020/10/09/2020.10.05.326470.full.pdf>
14. Martínez-López, L.A., Ochoa-Alejo, N., Martínez, O.: Dynamics of the chili pepper transcriptome during fruit development. *BMC genomics* **15**(1), 143 (2014)
15. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289–300 (1995)
16. Robinson, M.D., McCarthy, D.J., Smyth, G.K.: edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**(1), 139–140 (2010)
17. Codd, E.F.: A relational model of data for large shared data banks. In: *Software Pioneers*, pp. 263–294. Springer, ??? (2002)
18. Pavlopoulos, G.A., Secrier, M., Moschopoulos, C.N., Soldatos, T.G., Kossida, S., Aerts, J., Schneider, R., Bagos, P.G.: Using graph theory to analyze biological networks. *BioData mining* **4**(1), 10 (2011)
19. Pavlopoulos, G.A., Wegener, A.-L., Schneider, R.: A survey of visualization tools for biological network analysis. *Biodata mining* **1**(1), 12 (2008)
20. Csardi, G., Nepusz, T.: The igraph software package for complex network research. *InterJournal, Complex Systems* **1695**(5), 1–9 (2006)

Figures

Figure 1 Plot of mean Standardized Expression Profiles (SEPs) in groups formed by accessions “AS” (in red), “SR” (in blue) and the SEPs including all genes from both accessions (in grey). Thin vertical lines over the circles marking each mean are the 99.99% ($\alpha = 1 \times 10^{-4}$) confidence intervals (CI's) for the means. Plot obtained with function “SEPs.plot()”.

Figure 2 Matrix of percentages of genes peaking at each one of the 49 possible combinations of seven times in accession “AS” (X-axis) and seven times in accession “SR” (Y-axis). The percentage of genes simultaneously presenting the peaks is shown at each intersection, while the size of the circles at each intersection corresponds with the proportion of genes. The green dashed line at the diagonal presents the proportion of genes with identical peak in both accessions.

Figure 3 Mean expression profiles for genes simultaneously having maximum expression at 0 DAA in accession “AS” but maximum expression at 60 DAA in accession “SR” (set “ASm0SRm60”). Obtained with function “SEPs.plot()”; see Box 4 in Additional file 1.

Figure 4 General Expression profiles for gene with id 3018. Obtained by running function “gene.summary(id=3018)”. See section 3.4 in Additional file 1’.

Tables

Table 1 Selected Gene Ontology enrichment analyses results. “O” and “E” are Observed and Expected number of genes in the “ASm0SRm60” set, respectively.

Category	Id	Description	Odds	O	E
Biological Process	GO:0006810	Transport	1.72	106	70
Biological Process	GO:0051649	Establishment of localization in cell	2.02	39	21
Biological Process	GO:0016236	Macroautophagy	22.48	4	0
Cell Component	GO:0009579	Thylakoid	0.00	0	9
Cell Component	GO:0009521	Photosystem	0.00	0	8
Cell Component	GO:0005778	Peroxisomal membrane	15.72	4	0
Molecular Function	GO:0022892	Substrate-specific transporter activity	1.90	42	23

Additional Files

Additional file 1 — R code and supplementary results.

A PDF document presenting and explaining the R code and results for the use of the “*Salsa*” package. This document serves also as a vignette for the package.

Additional file 2 — Binary salsa package.

File “Salsa_0.4.tar.gz” contains the ready to install “*Salsa*” package. For installation use

“R CMD INSTALL Salsa_0.4.tar.gz”

at your operative system command window, or alternatively, use the R function “install.packages()”.

Figures

Figure 1

Plot of mean Standardized Expression Profiles (SEPs) in groups formed by accessions "AS" (in red), "SR" (in blue) and the SEPs including all genes from both accessions (in grey). Thin vertical lines over the circles marking each mean are the 99.99% ($\alpha = 1 \times 10^{-4}$) confidence intervals (CI's) for the means. Plot obtained with function "SEPs.plot()".

Figure 2

Matrix of percentages of genes peaking at each one of the 49 possible combinations of seven times in accession "AS" (X-axis) and seven times in accession "SR" (Y-axis). The percentage of genes simultaneously presenting the peaks is shown at each intersection, while the size of the circles at each intersection corresponds with the proportion of genes. The green dashed line at the diagonal presents the proportion of genes with identical peak in both accessions.

Figure 3

Mean expression profiles for genes simultaneously having maximum expression at 0 DAA in accession "AS" but maximum expression at 60 DAA in accession "SR" (set "ASm0SRm60"). Obtained with function "SEPs.plot()"; see Box 4 in Additional file 1.

Figure 4

General Expression profiles for gene with id 3018. Obtained by running function "gene.summary(id=3018)". See section 3.4 in Additional file 1".

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalFile1.pdf](#)