Physiologia Plantarum
An International Journal for Plant Biology

The MORPH-R web server and software tool for predicting missing genes in biological pathways

David Amar[a], Itziar Frades[b], Tim Diels[c,d], David Zaltzman[e], Netanel Ghatan[e], Pete E. Hedley[f], Erik Alexandersson[b], Oren Tzfadia[e,c,*] and Ron Shamir[a]

[a]Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel
[b]Department of Plant Protection Biology, Swedish University of Agricultural Sciences, Alnarp, Sweden
[c]Department of Plant Systems Biology, VIB, Ghent, Belgium
[d]Department of Mathematics and Computer Science, University of Antwerp, Antwerp, Belgium
[e]Department of Plant Science, The Weizmann Institute of Science, Rehovot, Israel
[f]Cell and Molecular Sciences, The James Hutton Institute, Dundee, United Kingdom

Correspondence
*Corresponding author,
e-mail: oren.tzfadia@psb.ugent.be

Received 23 December 2014

doi:10.1111/ppl.12326

A biological pathway is the set of molecular entities involved in a given biological process and the interrelations among them. Even though biological pathways have been studied extensively, discovering missing genes in pathways remains a fundamental challenge. Here, we present an easy-to-use tool that allows users to run MORPH (MOdule-guided Ranking of candidate PatHway genes), an algorithm for revealing missing genes in biological pathways, and demonstrate its capabilities. MORPH supports the analysis in tomato, Arabidopsis and the two new species: rice and the newly sequenced potato genome. The new tool, called MORPH-R, is available both as a web server (at http://bioinformatics.psb.ugent.be/webtools/morph/) and as standalone software that can be used locally. In the standalone version, the user can apply the tool to new organisms using any proprietary and public data sources.

Introduction

A biological pathway can be summarized as a set of molecular entities involved in a single biological process and the interactions among those entities. Understanding how pathways work and identifying the participating genes in a pathway of interest are crucial for understanding biology, organizing biological knowledge and enhancing biotechnological development. While current knowledge about some biological pathways is substantial and useful for systems-level analyses, not all the genes that participate or affect such pathways are known. Therefore, closing gaps in our current knowledge about biological pathways is a fundamental challenge.

We previously developed the MORPH (MOdule-guided Ranking of candidate PatHway genes) algorithm for revealing missing genes in biological pathways and demonstrated its robustness in tomato and Arabidopsis (Tzfadia et al. 2012). The MORPH algorithm is based on two main learning tasks. First, of a large variety of possible data sources [e.g. gene expression matrices, protein–protein interactions (PPI) and clustering solutions], it learns which datasets are more informative for the pathway of interest. Second, using the selected data, it ranks genes by their association with the pathway of interest.

Here, we present MORPH-R, an easy-to-use R package that allows users to run MORPH conveniently on their own PC or on a web server. The web server and the standalone software are available at http://bioinformatics.psb.ugent.be/webtools/morph/. The MORPH-R package currently supports tomato, potato, rice and Arabidopsis. Users that want to analyze additional datasets, possibly of new organisms, can use the standalone tool. We demonstrate the power of

Abbreviations – GO, gene ontology; GUI, graphical user interface; JA, jasmonic acid; LOOCV, leave-one-out cross-validation; MD, metabolic dependency; MORPH, MOdule-guided Ranking of candidate PatHway genes; PPI, protein–protein interaction.

MORPH-R on pathways of the newly sequenced potato genome. The analysis covers 694 Gene Ontology (GO) categories, retrieved from BioMart (Kasprzyk 2011) and 96 metabolic pathways retrieved from MapMan (Thimm et al. 2004, Urbanczyk-Wochniak et al. 2006). We show that MORPH-R reaches high performance using the potato data and also identifies novel candidate genes.

Materials and methods

Scoring the accuracy of a ranking algorithm

To evaluate an algorithm that ranks genes by their likelihood to be related to a given pathway, we use leave-one-out cross-validation (LOOCV) (Kharchenko et al. 2004). That is, we repeatedly remove a gene from the pathway, run the algorithm on the remaining pathway genes as input and ask what the ranking of the excluded gene is (in perfect ranking, the excluded gene is always at the top of the ranking). The LOOCV process results in a single score called area under the curve of the self-ranked genes (AUSR) between 1 and 0, where 1 is a perfect score and scores close to 0 indicate a random ranking of the candidate genes (see Kharchenko et al. 2006 for details).

How MORPH works

We briefly describe here how MORPH ranks candidate genes for participating in or affecting a pathway of interest. Our goal here is to provide the biological intuition behind the algorithm (for full details, see Tzfadia et al. 2012).

MORPH combines the power of cluster analysis with available large-scale data. First, any large-scale data can be used to partition the genes from that organism into coherent clusters that are expected to share similar biological function. For example, gene expression datasets can be used to detect co-expressed gene clusters. PPI networks can be used to detect protein complexes and pathways, and metabolic dependency (MD) interactions can be used to detect gene groups that participate in the same metabolic processes. For each supported species, MORPH's internal database contains a set of clustering solutions derived from such large-scale datasets. Given a particular clustering solution and a set of genes from a pathway of interest, we are only interested in the clusters that contain the pathway genes. A candidate gene is scored by its co-expression level with the known pathway genes present in its cluster. These scores are used to rank the candidate genes. The process thus depends on the particular clustering solution and on the gene expression matrix used. A combination of clustering solution and gene expression matrix is called a *learning configuration*.

MORPH selects among the different learning configurations in order to optimize the inference for the target pathway. For example, when learning photosynthesis-related pathways, we expect a learning configuration that is based on clustering of metabolic information and gene expression data of experiments using leaves to perform better than learning configurations that use a signaling network and gene expression from experiments using seeds. To cope with this, MORPH uses an internal LOOCV process to select the best learning configuration (e.g. the one with the highest AUSR score). Hence, the output of MORPH is the selected learning configuration, and the candidate gene ranking based on it.

How to run MORPH-R with default data

The new MORPH-R tool enables users to run MORPH via a simple graphical user interface (GUI) and evaluates biological pathways in four plants: tomato, potato, rice and Arabidopsis. Here, we describe how to use the GUI. For documentation and usage of the R functions, see Appendix S1, section 4. The input is the organism name and a list of genes known to participate in the target pathway (see Appendix S1, section 2 for the allowed types of gene identifiers). This list can be uploaded either manually in the text box or as a text file. Then, running MORPH is done by pressing the 'Submit' button (Fig. 1).

MORPH-R's internal database, which is provided with the standalone version, contains a large collection of gene expression profiles. The database supports analysis for four species by default – Arabidopsis, tomato, potato and rice. For Arabidopsis, it contains 216 expression profiles divided into four data sets: (1) *seedling* (64 profiles), (2) *tissues*: a collection of different tissues (99 profiles), (3) *DS1*: a union of the seedlings and tissues datasets and (4) *seed*: seed tissues at different developmental stages (53 profiles). The tomato gene expression repository contains 53 microarray expression profiles reflecting responses to specific stimuli, developmental stages, and selected mutants, divided into two data sets: *root* and *leaf* (21 profiles), and *fruit* (32 profiles). In addition, for tomato and Arabidopsis, a PPI network and an MD network are available (reference to the sources of dataset and network are available in Tzfadia et al. 2012). The potato gene expression repository summarizes over 20 studies and contains 326 profiles. These data are partitioned into four datasets: (1) *All tissues* (326 profiles), (2) *root* (24 profiles), (3) *tuber* (60 profiles) and (4) *leaf* (242 profiles). Potato interaction networks were predicted using sequence homology analysis (See Tzfadia et al. 2012 for details). The rice data now included in MORPH have eight microarray data sets (see

Fig. 1. The main screen. The input specified by the user is the organism to analyze and a group of gene IDs of a specific pathway of interest. To activate the program click the 'Run Morph' button. Once MORPH-R is done running, the results are printed on the screen to right (see Fig. 2). This page will appear on the web-server and on the user's default web browser upon execution of the standalone tool.

Table S9) (92 profiles), collected from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/), and one RNAseq data set (16 profiles) collected from the Rice Genome Annotation portal (http://rice.plantbiology.msu.edu/expression.shtml). In addition to gene expression, partitioning of genes into groups, the rice protein interaction network (Bian et al. 2012), was used to generate clustering solutions.

For each organism, a set of clustering solutions was generated by analyzing the gene expression datasets and the interaction networks (PPIs and MDs). To cluster an expression dataset, we used the CLICK (Sharan and Shamir 2000) and SOM (Tamayo et al. 1999) algorithms, which are available within the EXPANDER suite (Ulitsky et al. 2010). We used MATISSE for combined analysis of gene expression data and the networks (Ulitsky and Shamir 2007). We used MCL (Enright et al. 2002) to cluster interaction networks. Using combinations of the gene expression datasets described above, and clustering solutions, 12–32 learning configurations are defined for each organism (the exact number of configurations depend on the number of datasets and networks available for each organism).

The output of MORPH-R contains three main parts shown in different tabs. The first tab contains the AUSR score of the pathway and a ranking of candidate genes (Fig. 2). We observed empirically that scores greater than 0.7 denote that on the pathway of interest MORPH performs significantly better than expected by chance (Tzfadia et al. 2012). All genes that were not clustered with any pathway gene are placed at the bottom of the ranking. All the other genes are sorted in descending order according to their z-scores (Tzfadia et al. 2012). The z-score of a gene indicates how strong a candidate

New Query

If you use Morph, please cite us at Plant Cell

Download standalone Morph R package with documentation (.zip)

MOdule guided Ranking of candidate PatHway genes

Results

Best AUSR	0.92
Gene expression	Seedlings
Clustering	Metabolic matisse
Genes present	at5g17230 at4g14210 at1g10830 at3g04870 at1g06820 at3g10230 at5g57030 at4g25700 at5g52570 at1g31800 at3g53130 at5g67030 at1g08550

Candidates

Rank	Gene ID	Score	Annotation
1	at4g37760	2.60	squalene epoxidase 3
2	at3g63520	2.57	carotenoid cleavage dioxygenase 1
3	at4g32770	2.46	tocopherol cyclase, chloroplast / vitamin E deficient 1 (VTE1) / sucrose export defective 1 (SXD1)
4	at1g17050	2.37	solanesyl diphosphate synthase 2
5	at2g26800	2.37	Aldolase superfamily protein. Aldolase superfamily protein. Aldolase superfamily protein
6	at2g41680	2.35	NADPH-dependent thioredoxin reductase C
7	at5g16715	2.26	ATP binding valine-tRNA ligases aminoacyl-tRNA ligases nucleotide binding ATP binding aminoacyl-tRNA ligases
8	at3g48730	2.23	glutamate-1-semialdehyde 2,1-aminomutase 2
9	at2g35840	2.22	Sucrose-6F-phosphate phosphohydrolase family protein. Sucrose-6F-phosphate phosphohydrolase family protein. Sucrose-6F-phosphate phosphohydrolase family protein
10	at1g36160	2.20	acetyl-CoA carboxylase 1. acetyl-CoA carboxylase 1
11	at5g17050	2.19	UDP-glucosyl transferase 78D2
12	at4g11570	2.16	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein. Haloacid dehalogenase-like hydrolase (HAD) superfamily protein
13	at4g27600	2.13	pfkB-like carbohydrate kinase family protein
14	at5g38520	2.12	alpha/beta-Hydrolases superfamily protein. alpha/beta-Hydrolases superfamily protein
15	at1g19920	2.10	Pseudouridine synthase/archaeosine transglycosylase-like family protein
16	at5g19850	2.07	alpha/beta-Hydrolases superfamily protein
17	at1g56500	2.04	haloacid dehalogenase-like hydrolase family protein
18	at3g51820	2.03	UbiA prenyltransferase family protein
19	at1g22430	2.02	GroES-like zinc-binding dehydrogenase family protein. GroES-like zinc-binding dehydrogenase family protein
20	at1g31190	2.00	myo-inositol monophosphatase like 1
21	at5g13930	1.99	Chalcone and stilbene synthase family protein

Fig. 2. A table of ranked candidate genes, the main output of MORPH-R. The AUSR score of the pathway is shown at the top of the table. This score ranges between 1 and 0, where 1 is a perfect score and scores close to 0 denote a random ranking of the candidate genes. The table lists the candidate genes for the input pathway, ordered by their z-scores in descending order.

is when compared to all other candidates. The other tabs present the selected learning configuration, and the number of pathway genes that were used in the analysis, and are present in MORPH-R's internal database and additional list reports the genes that are missing from the analysis.

Adding new data sets and new species

When using the standalone version of MORPH-R, the internal database can be easily modified by adding or excluding gene expression datasets, networks or clustering solutions. In addition, it can be customized for running on new data from any organism once gene

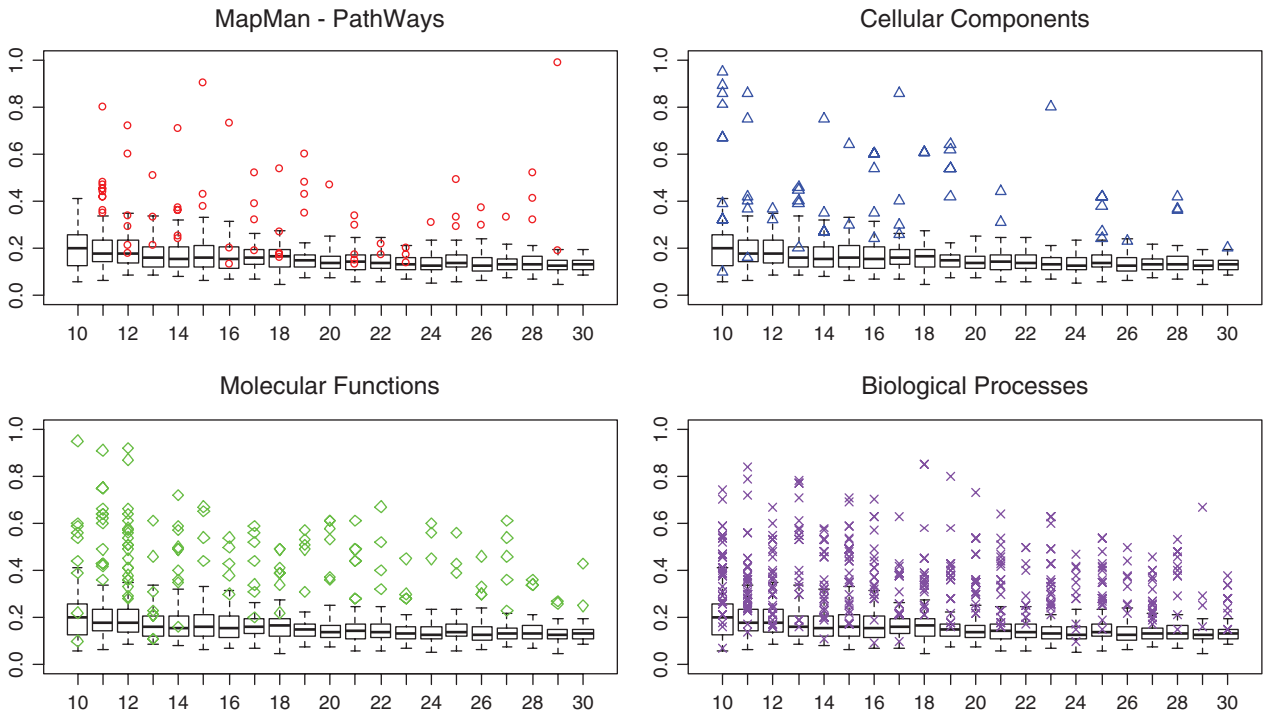

Fig. 3. Performance on potato functional groups. AUSR scores for MapMan pathways and three GO categories: cellular components, molecular functions and biological processes. For each group size, individual marks are results for real groups. The boxplots represent the distribution of AUSR scores of random gene sets (100 repeats). Boxes contain the 25–75 percentile of the distribution.

expression data and clustering solutions are given. MORPH-R contains a suite of R functions that allow computational biologists to pre-process and cluster their datasets (See Appendix S1, sections 3 and 4).

Potato microarray data preprocessing and normalization

We have integrated potato gene expression data from over 20 studies based on the Agilent JHI *Solanum tuberosum* 60 k v1 microarray (ArrayExpress ID: E-MTAB-1655) processed at the James Hutton Institute using standard Agilent recommended methodologies (Hancock et al. 2014). The studies included 326 conditions derived from the following treatments: moderate heat-stress (Hancock et al. 2014), short- and long-day growth regimes (Morris et al. 2014), bruising (unpublished data), phosphorous growth regimes (unpublished data), acidity, *Phytopthora infestans* infection (Ali et al. 2014) and β-aminobutyric acid (BABA) (Bengtsson et al. 2014), phosphite, abscisic acid (ABA), brassinosteroid and salicylic acid (SA) treatments (unpublished data). Different potato accessions and tissues (tuber, stem and leaf) were included.

We applied quantile normalization using the Limma package (Smyth 2005) and subtracted the background intensity from the foreground intensity for each spot using the normexp method (Ritchie et al. 2007). The normalized expression matrix contained 52 998 probes. In order to reduce statistical noise and to focus on genes with high variation we removed both probes with consistently low expression values across the samples and probes with low variance. Thresholds for probe removal were adjusted as proposed in (Tzfadia et al. 2012), which left 14 000 probes. These probes were mapped to 12 956 genes, approximately the same number of genes analyzed in Tzfadia et al. (2012).

Results

We illustrate the potential of MORPH-R by applying the tool to 698 potato GO categories, retrieved from BioMart (Kasprzyk 2011) and 96 potato metabolic pathways retrieved from MapMan (Thimm et al. 2004, Urbanczyk-Wochniak et al. 2006). The output of MORPH-R for each pathway or GO term is the AUSR score and the list of ranked candidate genes. In addition, we ran MORPH on random gene sets of sizes 10–30, with 100 repeats for each size. The maximum AUSR over these random groups of genes was 0.35 (Fig. 3). For a complete list of GO categories and MapMan pathways and their AUSR scores, see Tables S1–S4.

Table 1. MapMan pathways and GO term categories with high AUSR scores from the potato data profiles. The group size is the number of the input genes that were covered in MORPH-R's internal potato data.

	AUSR	Group size
MapMan pathways		
PS light reaction photosystem II. LHC-II	0.99	29
PS light reaction photosystem I. PSI polypeptide subunits	0.90	15
Stress abiotic touch/wounding	0.80	11
PS light reaction cyclic electron flow-chlororespiration	0.73	16
DNA synthesis/chromatin structure histone	0.62	53
Hormone metabolism auxin induced-regulated-responsive-activated	0.54	79
Stress abiotic heat	0.50	105
GO: molecular functions		
Phosphofructokinase activity	0.92	12
Diacylglycerol O-acyltransferase activity	0.91	11
GO: cellular components		
6-phosphofructokinase complex	0.95	10
Photosystem I reaction center	0.89	10
GO: biological process		
Regulation of peptidase activity	0.85	18
Negative regulation of peptidase activity	0.85	18
Folic acid-containing compound biosynthetic process	0.84	11

Some of the MapMan pathways and GO term categories obtained exceptional AUSR scores (Table 1). For example, three large MapMan pathways (>50 genes) obtained very high AUSR scores, which is unusual for pathways of such size (Tzfadia et al. 2012): DNA synthesis/chromatin structure (MapMan 28.1.3, 53 genes) AUSR = 0.62; hormone metabolism auxin induced-regulated-responsive-activated (17.2.3, 79 genes) AUSR = 0.54; and stress abiotic heat (20.2.1; 105 genes) AUSR = 0.50.

Potato GO example: folate biosynthesis

One of the best scoring GO terms was folate biosynthesis (GO: 0009396; AUSR 0.84). In spite of its importance to humans who are dependent on folates from plant and microbial sources, little is known on regulation of folate content and biosynthesis in plants. In tomato, a two-gene strategy overexpressing GTP cyclohydrolase 1 and aminodeoxychorismate synthase gave a 25-fold increase in folate content (Diaz de la Garza et al. 2007). However, a similar approach in potato was not successful (Blancquaert et al. 2013), which calls for a better understanding of the folate pathway in order to find an engineering strategy for potato and other crops.

A rather low number of genes were associated with folic acid-containing compound biosynthetic process by MORPH-R (11 genes were used as input and MORPH-R suggested 21 new candidate genes; see Tables S5 and S6 for the candidate genes and the list of pathway genes). As could be expected, an additional isoform of an aminodeoxychorismate lyase was among them. These enzymes have been shown to be part of folate biosynthesis where one family member catalyzes the last step of the *p*-aminobenzoate branch (Basset et al. 2004). One of the embryogenesis-related genes associated with the pathway according to MORPH-R was SufD. This prediction is in line with previous findings showing that folate biosynthesis is essential for embryogenesis in *Arabidopsis* (Ishikawa et al. 2003).

Although well studied in animal systems, gene products important for the transportation of folates between cellular compartments and organs remain largely unknown in plants. This is true although it is evident that biosynthesis and storage require movement over plastid, mitochondria and vacuolar membranes (Hanson and Gregory 2011). A few transport proteins were linked to folate biosynthesis, e.g. an organic cation transporter. However, neither these proteins nor their homologs have been studied in detail in potato or *Arabidopsis*. Our examples show how MORPH-R can point out novel gene targets that might affect folate biosynthesis.

Potato MapMan term example: wound response

We focused on the MapMan pathway abiotic stress touch/wounding (20.2.4). The AUSR score was 0.8, and MORPH-R suggested 172 genes. Wounding gives rise to a broad set of responses in plants, and it is important to identify genes involved in wound response because they can confer resistance to a broad set of stresses, both abiotic and biotic. Not surprisingly, jasmonic acid (JA)-dependent elements such as ornithine N-delta-acetyltransferase, which forms the defense metabolite N-delta-acetylornithine (Adio et al. 2011), were found among the genes associated by MORPH-R with wound response (see Tables S7 and S8 for the candidate genes and the list of pathway genes). JA is central in wounding response and JA-Ile accumulates in leaves within minutes after damage (Glauser et al. 2008). Another example is the RS5 stachyose synthase gene (ranked 16th) which is nearly identical to the *Arabidopsis* raffinose synthase (At5g40390), recently suggested to be alone responsible for abiotic-induced raffinose biosynthesis in Arabidopsis leaves (Egert et al. 2013). Raffinose

production is increased by numerous abiotic stresses and thought to be involved in protection in oxidative stress (Nishizawa et al. 2008), and over-expression of the stachyose synthase could potentially increase quenching capacity.

Discussion

Even in well studies model systems as Arabidopsis and rice, we still know little about functional annotation of genes. For example, in approximately 40% of Arabidopsis and 1% of rice (*Oryza sativa*), protein-coding genes have been functionally annotated (Rhee and Mutwil 2014). Moreover, the number of functionally annotated genes based on experimental validation in non-model species is scarce. Therefore, gene discovery is still a major challenge in the plant biology research. Several computational approaches to protein function annotation exist, although most are not dedicated to plants, or they are restricted to model plant species only. In this report, we described MORPH-R, an easy-to-use R package that allows users to run MORPH conveniently on their own PC or via a web portal. MORPH-R takes a set of genes that are known to participate in the target pathway as input. Then, it uses a large compendium of data sources (gene expression datasets and interaction networks) to identify and rank new candidate genes.

Because the interactions among pathway genes and products might manifest in a particular data source but not in others, MORPH-R learns which data sources are most informative for each input pathway. It then uses the selected data sources for ranking candidate genes. The package presented here contains both a GUI that can be used without any programming skills and easy-to-use R functions, which allow computationally oriented users more flexibility and additional functionality. The web server provides data and direct analysis capabilities for four major plants: Arabidopsis, potato, tomato and rice, which adds two new organisms. In addition, incorporating new data sources for the supported organisms and adding new species to MORPH can be done easily by utilizing the R scripts provided in the standalone version of the software.

The yardstick for evaluating the quality of MORPH predictions is the AUSR score. We demonstrated the high performance of MORPH-R on potato pathways and showed that many of them achieved significant scores (Fig. 3). We studied the results of two pathways in detail, folate biosynthesis and wound response, and found new candidate genes involved in these processes. The candidates can be categorized into three main groups: additional isoforms of known pathway-genes, such as the aminodeoxychorismate lyase associated with folate biosynthesis; genes whose functionality is supported by literature evidence in another species, such as the RS5 stachyose synthase gene, which was studied in association to wounding in Arabidopsis but not in potato; or genes of unknown function.

In order to make MORPH more broadly useful as a gene discovery framework for plant researchers, we re-implemented the original MORPH algorithm in R and C++. The new version is more modular and thus enables adding new organisms and/or new data sources easily by the user. For the purpose of this report, we used these capabilities to add potato and rice data as additional default species. Those, alongside Arabidopsis and tomato, are available both in the web server and in the standalone version. Moreover, we provide the code of MORPH (R and C++ versions), so users with some experience in command line software execution can run MORPH in 'batch mode'. The new version is also two orders of magnitude faster than the original one. Consequently, a batch mode user can obtain scores and novel predictions for hundreds of biological pathways (originating from MapMan or GO) in just a couple of hours. When the AUSR score of a pathway is high, MORPH-R suggests novel candidate genes that may belong to the pathway. A systematic application of MORPH-R can be used to build and extend biological pathways and regulatory networks.

Author contributions

D. A., O. T. and RS designed the study. D. A., I. F., D. C., D. Z., N. G. and O. T. performed the research. D. A., I. F. and D. C. contributed new analytic computational tools. P. H. provided the gene expression data sets. E. A. analyzed the data. D. A., I. F., E. A., O. T. and R. S. wrote the paper.

Acknowledgements – D. A. was supported in part by fellowships from the Azrieli foundation, and the Edmond J. Safra center for Bioinformatics at Tel Aviv University. O. T. was supported by iCORE. Erik Alexandersson and Itziar Frades were supported by Crafoord grant (20120533) and the Swedish Foundation for Strategic Research (RB608-0006). R. S. was supported by the Israel Science Foundation (grant 317/13). We thank the following researchers from the James Hutton Institute for sharing their microarray datasets: Glenn Bryan, Jayne Davis, Eleanor Gilroy, Wayne Morris, Louise Shepherd, Mark Taylor, Dionne Turnbull and Philip White. We thank Sandeep Kumar Kushwaha for providing helpful comments on the manuscript.

References

Adio AM, Casteel CL, De Vos M, Kim JH, Joshi V, Li B, Juéry C, Daron J, Kliebenstein DJ, Jander G (2011) Biosynthesis and defensive function of N*δ*-acetylornithine, a jasmonate-induced Arabidopsis metabolite. Plant Cell 23: 3303–3318

Ali A, Alexandersson E, Sandin M, Resjö S, Lenman M, Hedley P, Levander F, Andreasson E (2014) Quantitative proteomics and transcriptomics of potato in response to *Phytophthora infestans* in compatible and incompatible interactions. BMC Genomics 15: 497

Basset GJ, Ravanel S, Quinlivan EP, White R, Giovannoni JJ, Rebeille F, Nichols BP, Shinozaki K, Seki M, Gregory JF 3rd, Hanson AD (2004) Folate synthesis in plants: the last step of the *p*-aminobenzoate branch is catalyzed by a plastidial aminodeoxychorismate lyase. Plant J 40: 453–461

Bengtsson T, Weighill D, Proux-Wéra E, Levander F, Resjö S, Burra DD, Moushib LI, Hedley PE, Liljeroth E, Jacobson D (2014) Proteomics and transcriptomics of the BABA-induced resistance response in potato using a novel functional annotation approach. BMC Genomics 15: 315

Bian H, Xie Y, Guo F, Han N, Ma S, Zeng Z, Wang J, Yang Y, Zhu M (2012) Distinctive expression patterns and roles of the miRNA393/TIR1 homolog module in regulating flag leaf inclination and primary and crown root growth in rice (*Oryza sativa*). New Phytol 196: 149–161

Blancquaert D, Storozhenko S, Van Daele J, Stove C, Visser RG, Lambert W, Van Der Straeten D (2013) Enhancing pterin and para-aminobenzoate content is not sufficient to successfully biofortify potato tubers and *Arabidopsis thaliana* plants with folate. J Exp Bot 64: 3899–3909

Diaz de la Garza RI, Gregory JF 3rd, Hanson AD (2007) Folate biofortification of tomato fruit. Proc Natl Acad Sci USA 104: 4218–4222

Egert A, Keller F, Peters S (2013) Abiotic stress-induced accumulation of raffinose in Arabidopsis leaves is mediated by a single raffinose synthase (RS5, At5g40390). BMC Plant Biol 13: 218

Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 30: 1575–1584

Glauser G, Grata E, Dubugnon L, Rudaz S, Farmer EE, Wolfender J-L (2008) Spatial and temporal dynamics of jasmonate synthesis and accumulation in Arabidopsis in response to wounding. J Biol Chem 283: 16400–16407

Hancock RD, Morris WL, Ducreux LJ, Morris JA, Usman M, Verrall SR, Fuller J, Simpson CG, Zhang R, Hedley PE, Taylor MA (2014) Physiological, biochemical and molecular responses of the potato (*Solanum tuberosum* L.) plant to moderately elevated temperature. Plant Cell Environ 37: 439–450

Hanson AD, Gregory JF 3rd (2011) Folate biosynthesis, turnover, and transport in plants. Annu Rev Plant Biol 62: 105–125

Ishikawa T, Machida C, Yoshioka Y, Kitano H, Machida Y (2003) The GLOBULAR ARREST1 gene, which is involved in the biosynthesis of folates, is essential for embryogenesis in *Arabidopsis thaliana*. Plant J 33: 235–244

Kasprzyk A (2011) BioMart: driving a paradigm change in biological data management. Database (Oxford) 2011: bar049

Kharchenko P, Chen L, Freund Y, Vitkup D, Church GM (2006) Identifying metabolic enzymes with multiple types of association evidence. BMC Bioinformatics 7: 177

Kharchenko P, Vitkup D, Church GM (2004) Filling gaps in a metabolic network using expression information. Bioinformatics 20(Suppl 1): i178–i185

Morris WL, Hancock RD, Ducreux LJM, Morris JA, Usman M, Verrall SR, Sharma SK, Bryan G, McNicol JW, Hedley PE (2014) Day length dependent restructuring of the leaf transcriptome and metabolome in potato genotypes with contrasting tuberization phenotypes. Plant Cell Environ 37: 1351–1363

Nishizawa A, Yabuta Y, Shigeoka S (2008) Galactinol and raffinose constitute a novel function to protect plants from oxidative damage. Plant Physiol 147: 1251–1263

Rhee SY, Mutwil M (2014) Towards revealing the functions of all genes in plants. Trends Plant Sci 19: 212–221

Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, Smyth GK (2007) A comparison of background correction methods for two-colour microarrays. Bioinformatics 23: 2700–2707

Sharan R, Shamir R (2000) CLICK: a clustering algorithm with applications to gene expression analysis. Proc Int Conf Intell Syst Mol Biol 8: 307–316

Smyth GK (2005) Limma: linear models for microarray data. In: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S (eds) Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer, New York, pp 397–420

Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci USA 96: 2907–2912

Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37: 914–939

Tzfadia O, Amar D, Bradbury LM, Wurtzel ET, Shamir R (2012) The MORPH algorithm: ranking candidate genes for membership in Arabidopsis and tomato pathways. Plant Cell 24: 4389–4406

Ulitsky I, Shamir R (2007) Identification of functional modules using network topology and high-throughput data. BMC Syst Biol 1: 8

Ulitsky I, Maron-Katz A, Shavit S, Sagir D, Linhart C, Elkon R, Tanay A, Sharan R, Shiloh Y, Shamir R (2010) Expander: from expression microarrays to networks and functions. Nat Protoc 5: 303–322

Urbanczyk-Wochniak E, Usadel B, Thimm O, Nunes-Nesi A, Carrari F, Davy M, Blasing O, Kowalczyk M, Weicht D, Polinceusz A, Meyer S, Stitt M, Fernie AR (2006) Conversion of MapMan to allow the analysis of transcript data from Solanaceous species: effects of genetic and environmental alterations in energy metabolism in the leaf. Plant Mol Biol 60: 773–792

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. MORPH-R documentation.

Table S1. AUSR scores obtained using potato biological process GO terms as input to MORPH-R.

Table S2. AUSR scores obtained using potato cellular component GO terms as input to MORPH-R.

Table S3. AUSR scores obtained using potato molecular function GO terms as input to MORPH-R.

Table S4. AUSR scores obtained using MapMan pathways as input to MORPH-R.

Table S5. Folic acid candidate genes generated by MORPH-R.

Table S6. Folic acid GO term list of known genes.

Table S7. Wound response MapMan pathway candidate genes generated by MORPH-R.

Table S8. Wound response (20.2.4) MapMan pathway list of known genes.

Edited by S. Pelaz