POEAS: Automated Plant Phenomic Analysis Using Plant Ontology

Khader Shameer1,*, Mahantesha BN Naika1,2,*, Oommen K. Mathew1,* and Ramanathan Sowdhamini1

1National Centre for Biological Sciences (TIFR), GKVK Campus, Bangalore, India. 2Department of Plant Biotechnology, University of Agricultural Sciences, GKVK Campus, Bangalore, India. *These authors contributed equally to this work.

ABSTRACT: Biological enrichment analysis using gene ontology (GO) provides a global overview of the functional role of genes or proteins identified from large-scale genomic or proteomic experiments. Phenomic enrichment analysis of gene lists can provide an important layer of information as well as cellular components, molecular functions, and biological processes associated with gene lists. Plant phenomic enrichment analysis will be useful for performing new experiments to better understand plant systems and for the interpretation of gene or proteins identified from high-throughput experiments. Plant ontology (PO) is a compendium of terms to define the diverse phenotypic characteristics of plant species, including plant anatomy, morphology, and development stages. Adoption of this highly useful ontology is limited, when compared to GO, because of the lack of user-friendly tools that enable the use of PO for statistical enrichment analysis. To address this challenge, we introduce Plant Ontology Enrichment Analysis Server (POEAS) in the public domain. POEAS uses a simple list of genes as input data and performs enrichment analysis using Ontologizer 2.0 to provide results in two levels, enrichment results and visualization utilities, to generate ontological graphs that are of publication quality. POEAS also offers interactive options to identify user-defined background population sets, various multiple-testing correction methods, different enrichment calculation methods, and resampling tests to improve statistical significance. The availability of such a tool to perform phenomic enrichment analyses using plant genes as a complementary resource will permit the adoption of PO-based phenomic analysis as part of analytical workflows. POEAS can be accessed using the URL http://caps.ncbs.res.in/poeas.

KEYWORDS: phenomics, plant ontology, phenotype enrichment, plant genomics, Arabidopsis thaliana

Introduction

Phenomics is a recently evolved term to collectively define the measurement of the phenotypic characteristics of biological entities, including the physical and biochemical traits of an organism.1,2 A phenome is a catalog of all phenotypes that is compiled from an experiment or from the collective phenomic knowledge of an organism. Plant phenomics1–7 refers to the systematic study of plant phenotypes. Ontologies, such as plant ontology (PO), play an important role functioning as translational resources between experimental and in silico phenotyping. Ontologies can be used to capture and map out an existing library of phenotypes to a list of new entities (for example, genes, proteins, and metabolites). Biomedical ontologies have improved the unified interpretation of a group of genes (gene lists), proteins, RNA, or metabolites identified from high-throughput genomics, proteomics, transcriptomics, or metabolomics studies. Gene ontology (GO) and the association of GO terms with gene products and statistical enrichment analyses have contributed to the interpretation of gene or protein lists for more than one decade. Ontologies are currently developed to address highly specific domains or subdomains in the biomedical knowledge universe. To illustrate the growth, currently a total of 329 ontologies are available from BioPortal — an ontology repository of the National Centre for Biomedical Ontology (NCBO).8 Along with the unanimous growth of broad spectrum ontology and widely used ontologies,
such as GO, various other biomedical ontologies are under active development. While these resources are available as reference tools, a large subset of biomedical ontologies does not have direct association data to connect different biological entities. Apart from the primary goal of the unification of concepts, definitions, and knowledge in biomedical science, a prominent application of biomedical ontologies is enrichment analysis. Biological enrichment analysis is a collective term used to define a broad area of knowledge-based statistical approaches. It is designed to identify statistically significant terms associated with the list of biological molecules identified from an experiment when compared to the background distribution (annotations of genes in the genome or genes in experimental platforms). Enrichment analysis can be implemented with an ontology or an annotation repository, such as Pfam domains and Swiss-Prot annotations, to understand the functional trend of biological phenomena. Ontology-based phenomic mappings were used in human phenotypes, and cellular phenotypes, fission yeast, disease annotations, and plants. Plant phenomics have been employed to study several aspects of plants, including the phenomic impact of stress-responsive genes.

**Plant Phenomic Enrichment Analyses Using PO**

Plant phenomics is the collective measurement of phenomes that includes the physical and biochemical traits of an organism, and the phenome of an organism can be effectively described using ontologies. PO is a compendium of terms to define the diverse phenotypic characteristics of plant species into two categories (plant anatomy, and morphology and development stages). PO definitions and related annotations are available for several model plant genomes and are integrated into several key plant genome databases, such as The Arabidopsis Information Resource (TAIR), NASC/NASCArrays, Gramene/GrameneMart, Sol Genomics Network (SGN), and MaizeGDB. Additional terms, annotations and genomes, are being added to PO because of the collective effort from experimental biologists, computational biologists, and biocurators. However, tools that are designed specifically to utilize the growing plant phenomic knowledgebase are required to leverage their application in large-scale plant phenomic studies. Currently, generic meta-analyses tools, such as DAVID or PANTHER, do not provide enrichment analyses using PO. A tool reported by Xin et al. provides enrichment analysis using PO terms, but the tool does not offer an option to select enrichment methods, multiple-testing correction methods, or visualization in a cyclic graph format. Recently, while performing a large-scale comparative analysis of stress-responsive genes (n = 3091) in *Arabidopsis thaliana*, we realized this challenge and adapted a widely used GO term enrichment analysis tool (Ontologizer 2.0) to perform phenomic enrichment analyses using genes from STIFB2. In this manuscript, we describe a web-based version of the utility called Plant Ontology Enrichment Analysis Server (POEAS), which has been developed and provided in the public domain for phenomic analyses.

**Materials and Methods**

POEAS is currently available for *A. thaliana*, additional genomes will be added as part of future updates. The latest version of PO files (.obo and .assoc) and TAIR annotations are fetched periodically from PO and TAIR FTP servers, respectively. Currently, POEAS accepts lists of gene names, locus names, or TAIR identifiers (IDs) as input data. The POEAS web interface (Fig. 1) is developed using Javascript, HTML, and CSS. Enrichment analysis was implemented using Ontologizer 2.0, a biomedical ontology enrichment analysis tool that has multiple options available to select the enrichment method and statistical approach. The following types of multiple-testing correction methods are available in the current version of POEAS: Bonferroni, Holm, Benjamini–Hochberg, Benjamini–Yekutieli, Westfall and Young step-down, and Westfall and Young single-step. Options are also provided to run enrichment analyses without multiple-testing corrections to test potential enrichment in small gene lists. Six enrichment calculation methods are available in the current version of POEAS: Model-Based Gene-Set Enrichment Analyses (MGSA), Parent–Child-Intersection, Parent–Child-Union, Term-For-Term, Topology-Elim, and Topology-Weighted. In the backend, the server uses a scheduler script to retrieve updated PO annotations and associations. POEAS also offers interactive options to identify user-defined background population sets, various multiple-testing correction methods, different enrichment calculation methods, and resampling tests to improve statistical significance (Fig. 1).

**Web server construction, the application features, and performance of POEAS.** POEAS provides a web-platform for performing enrichment analyses of PO terms using genes from *A. thaliana*. The user can submit a list of differentially expressed gene IDs from expression profiling (RNASeq or microarray experiments). Depending on the availability, a list of background genes tested in the experiment can also be provided. Further, the user can select multiple-testing correction methods, enrichment calculation methods, and resampling steps to perform the enrichment analyses (Fig. 1). The successful POEAS run provides tables with enriched PO terms associated with the gene list; visualization of the enriched terms in a PO tree diagram can also be accessed. Files are also provided to download enrichment results, annotation tables, and PO diagrams in SVG format. The downloadable files can be used to filter associated PO terms and genes associated with each PO term based on user requirements (Fig. 2).

**A use-case for POEAS: phenomic features of stress-responsive genes upregulated by abscisic acid (ABA).** POEAS can be used for the phenomic inference of genes from different types of experiments. To illustrate the application
Phenome-wide analysis in plants using POEAS server

Figure 1. Web interface of POEAS. (A) Options to input list of genes identified from an experiment (microarray, next-generation sequencing, proteomics, etc.) and background list of genes from the study (for example, list of genes in a microarray, genes in a given genome, etc.). (B) Option to select multiple-testing correction method. (C) Option to select enrichment calculation method. (D) Option to select resampling steps for multiple-testing corrections.

Figure 2. Features of POEAS. (A) Results table providing various information from a phenomic enrichment analysis using POEAS. The table provides information, including PO ID, PO term (name), and P-values (unadjusted and adjusted P-values using multiple-testing correction method (p.adjusted)). (B) Dicyclic graph of terms associated with input gene list. (C) Options to download data in various formats after a phenomic enrichment analysis using POEAS.
of POEAS, we discuss a use-case here. We identified 700 *A. thaliana* genes that were responsive to ABA stress, which were obtained from the Stress Responsive Transcription Factor Database, version 2 (STIFDB2). These were targeted by one or more stress-responsive transcription factors.\(^6,17,35\) This list of 700 TAIR locus IDs was used as input, and the multiple-testing correction method was set to “Bonferroni,” the enrichment calculation method was set to “Term-For-Term,” and the resampling steps were set to “1000.” The output from this analysis provided extensive information on plant phenotypic characteristics represented by these genes. Phenomic analytics revealed that a subset of genes influences plant phenotypes in multiple levels of plant structure development stages (temporal) and plant anatomy. A total of 65 enriched plant anatomy terms (Table 1) and 20 temporal terms (Table 2) were enriched \((P = 0.05; \text{Bonferroni corrected})\). The most significant terms associated with genes that respond to ABA stress treatments are ones like “cotyledon”, “pollen”, “microgametophyte” and “pollen sac”. ABA is a key regulatory plant hormone that acts as a mediator between various physiological processes, including seed dormancy, plant growth, and secondary stress response for various abiotic stressors, such as drought, cold, light, and temperature. Increased levels of ABA were used to replicate environmental stress in the laboratory setting.\(^36-38\) Biological and functional term enrichment analyses of the 700 genes responsive to ABA treatment provided insights into

### Table 1. PO (anatomy/plant anatomical entity) terms associated with ABA responsive genes in *A. thaliana* identified using POEAS.

| PO ID     | PO TERM          | P-VALUE |
|-----------|------------------|---------|
| PO:0020030| Cotyledon        | 8.9E-31 |
| PO:0025281| Pollen           | 1.3E-27 |
| PO:0025280| Microgametophyte | 1.3E-27 |
| PO:0025277| pollen sac       | 1.4E-27 |
| PO:0025202| Microsporangium  | 2.9E-27 |
| PO:0025094| Sporangium       | 3.6E-27 |
| PO:0009007| portion of plant tissue | 6.9E-26 |
| PO:0009002| plant cell       | 1.5E-25 |
| PO:0020038| Petiole          | 2.0E-22 |
| PO:0025066| Stalk            | 3.9E-22 |
| PO:0025139| phyllome apex    | 9.3E-22 |
| PO:0020137| leaf apex        | 9.3E-22 |
| PO:0008019| leaf lamina base | 2.1E-21 |
| PO:0000230| inflorescence meristem | 2.3E-21 |
| PO:0020039| leaf lamina      | 2.5E-21 |
| PO:0025060| Lamina           | 2.5E-21 |
| PO:0000293| guard cell       | 8.3E-21 |
| PO:0000013| cauline leaf     | 9.3E-21 |
| PO:0009013| Meristem         | 1.0E-20 |
| PO:0002000| stomatal complex | 3.4E-20 |
| PO:0025165| shoot epidermal cell | 6.1E-20 |
| PO:0000005| cultured plant cell | 1.1E-18 |
| PO:0000004| in vitro plant structure | 1.2E-18 |
| PO:0006035| shoot epidermis  | 1.7E-18 |
| PO:0009061| Androecium       | 9.2E-18 |
| PO:0009029| Stamen           | 9.2E-18 |
| PO:0009025| vascular leaf    | 2.2E-17 |

(Continued)

### Table 1. (Continued).

| PO ID     | PO TERM          | P-VALUE |
|-----------|------------------|---------|
| PO:0004013| epidermal cell   | 3.2E-17 |
| PO:0025034| Leaf             | 4.7E-17 |
| PO:0009027| Megasporophyll   | 6.6E-16 |
| PO:0009030| Carpell          | 6.6E-16 |
| PO:0025195| pollen tube cell | 7.1E-16 |
| PO:0005679| Epidermis        | 1.3E-15 |
| PO:0005005| shoot internode  | 1.8E-15 |
| PO:0020142| stem internode   | 1.8E-15 |
| PO:0020100| hypocotyls      | 1.8E-15 |
| PO:0009062| Gynoecium        | 8.1E-14 |
| PO:0009001| Fruit            | 1.5E-13 |
| PO:0009028| Microsporophyll  | 2.3E-13 |
| PO:0009081| inflorescence branch | 1.4E-11 |
| PO:0009052| Pedicel          | 1.4E-11 |
| PO:0020122| inflorescence axis | 1.8E-11 |
| PO:0009010| Seed             | 2.4E-11 |
| PO:0000037| shoot apex       | 2.9E-11 |
| PO:0025001| cardinal organ part | 3.1E-10 |
| PO:0009026| Sporophyll       | 1.4E-09 |
| PO:0005052| plant callus     | 1.7E-09 |
| PO:0009009| plant embryo     | 2.2E-09 |
| PO:0009005| Root             | 3.4E-09 |
| PO:0025025| root system      | 6.0E-09 |
| PO:0009047| Stem             | 9.4E-09 |
| PO:0009049| Inflorescence    | 1.3E-08 |
| PO:0009059| Corolla          | 1.5E-08 |
| PO:0009032| Petal            | 1.5E-08 |
| PO:0009060| Calyx            | 1.1E-07 |
| PO:0009031| Sepal            | 1.1E-07 |
| PO:0000084| plant sperm cell | 7.3E-07 |
| PO:0025006| Gynoecete        | 3.9E-06 |
| PO:0009058| Perianth         | 1.0E-05 |
| PO:0025022| collective leaf structure | 1.1E-05 |
| PO:0000003| whole plant      | 1.1E-05 |
| PO:0006001| Phyllome         | 2.0E-05 |
| PO:0000025| root tip         | 1.3E-04 |
| PO:0025029| shoot axis       | 9.2E-04 |
| PO:0025023| collective phyllome structure | 2.4E-02 |

Note: *Bonferroni-adjusted P-values.*
Table 2. PO (temporal/plant structure development stage) terms associated with ABA responsive genes in *A. thaliana* identified using POEAS.

| PO ID     | PO TERM                                | P-VALUE  |
|-----------|----------------------------------------|----------|
| PO:0007095| LP.08 eight leaves visible             | 6.9E-17  |
| PO:0007123| LP.06 six leaves visible               | 2.0E-15  |
| PO:0007064| LP.12 twelve leaves visible            | 3.3E-14  |
| PO:0007103| LP.10 ten leaves visible               | 7.8E-13  |
| PO:0007098| LP.02 two leaves visible               | 2.8E-12  |
| PO:0001017| M germinated pollen stage              | 5.2E-12  |
| PO:0001054| 4 leaf senescence stage                | 4.1E-11  |
| PO:0001050| leaf development stages                | 4.8E-11  |
| PO:0007131| seedling development stage             | 5.9E-11  |
| PO:0001007| pollen developmental stages            | 9.3E-11  |
| PO:0007605| androecium developmental stages        | 9.7E-11  |
| PO:0007134| A vegetative growth                   | 7.5E-10  |
| PO:0007033| whole plant growth stage               | 9.2E-10  |
| PO:0001016| L mature pollen stage                 | 1.9E-09  |
| PO:0007115| LP.04 four leaves visible             | 2.9E-08  |
| PO:0007133| leaf production                        | 3.6E-06  |
| PO:0007112| 1 main shoot growth                   | 3.8E-06  |
| PO:0001185| C globular stage                      | 5.2E-04  |
| PO:0001081| F mature embryo stage                 | 1.8E-03  |
| PO:0004507| D bilateral stage                     | 5.5E-03  |

Note: *Bonferroni-adjusted P-values.

Figure 3. Visualization of PO terms associated with genes responsive to ABA using POEAS. (A) Highlighted terms are 20 temporal terms associated with genes responsive to ABA. (B) Highlighted terms are 65 anatomy terms associated with genes responsive to ABA.
for the user to include additional layers of information, such as a background dataset; select statistical tests, such as Bonferroni correction; and resample to improve plant phenomic enrichment analyses.

Conclusion
We have designed a public web server called POEAS for automated phenomic enrichment analyses of the genes of *A. thaliana*. As phenomic analyses are gaining interest in the plant community, the availability of POEAS would enable the use of phenomic enrichment as a routine analytical step in automated and custom annotation workflows.

Acknowledgments
The authors of this paper would like to acknowledge the authors of Ontologizer 2.0 and PO for their useful discussions. The authors thank the National Centre for Biological Sciences, Tata Institute of Fundamental Research, and the University of Agricultural Sciences (Bangalore) for infrastructural support.

Author Contributions
Conceived and designed the experiments: RS, KS. Analyzed the data: MBNN, KS. Contributed to the writing of the manuscript: KS, MBNN, OKM. Agree with manuscript results and conclusions: KS, MBNN, OKM, RS. Jointly developed the structure and arguments for the paper: KS, MBNN, OKM. Made critical revisions and approved final version: RS. All authors reviewed and approved of the final manuscript.

REFERENCES
1. Bilder RM, Sabb FW, Cannon TD, et al. Phenomics: the systematic study of phenotypes on a genome-wide scale. *Neuroscience*. 2009;164:30–42.
2. rode D, Gowd J, Drabkin H, et al. Ontology: the next challenge. *Nat Rev Genet*. 2010;11:85–66.
3. Finkel E. Imaging. With ‘phenomics,’ plant scientists hope to shift breeding into overdrive. *Science*. 2010;325:380–1.
4. Furbank RT, Tester M. Phenomics—technologies to relieve the phenotyping bottleneck. *Trends Plant Sci*. 2011;16:635–44.
5. Berger B, de Rigt B, Tester M. Trait dissection of salinity tolerance with plant phenomics. *Methods Mol Biol*. 2012;913:399–411.
6. Naika M, Shameer K, Sowdhamini R. Comparative analyses of stress-responsive genes in Arabidopsis thaliana: insight from genomic data mining, functional enrichment, pathway analysis and phenomics. *Mel Biosyst*. 2013;2:988–908.
7. Yang W, Duan L, Chen G, Xiong L, Liu Q. Plant phenomics and high-throughput phenotyping: accelerating rice functional genomics using multidisciplinary technologies. *Curr Opin Plant Biol*. 2013;16:180–7.
8. Whetzel PL, Noy NF, Shah NH, et al. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Res*. 2011;39:WS41–5.
9. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25:25–9.
10. Blake JA, Dolan M, Drabkin H, et al. Gene Ontology annotation and resources. *Nucleic Acids Res*. 2013;41:D530–5.
11. Cote R, Reisinger F, Martens L, Barhun S, Vizecina J, Hermandjok H. The Ontology lookup service: bigger and better. *Nucleic Acids Res*. 2010;38:W155–60.
12. Smith B, Ashburner M, Rosse C, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol*. 2007;25:1251–5.