LEAFDATA: a literature-curated database for Arabidopsis leaf development

Dóra Szakonyi*

Abstract

Background: In the post-genomic era, biological databases provide an easy access to a wide variety of scientific data. The vast quantity of literature calls for curated databases where existing knowledge is carefully organized in order to aid novel discoveries. Leaves, the main photosynthetic organs are not only vital for plant growth but also essential for maintaining the global ecosystem by producing oxygen and food. Therefore, studying and understanding leaf formation and growth are key objectives in biology. Arabidopsis thaliana to this date remains the prime experimental model organism in plant science.

Description: LEAFDATA was created as an easily accessible and searchable web tool to assemble a relevant collection of Arabidopsis leaf literature. LEAFDATA currently contains 13,553 categorized statements from 380 processed publications. LEAFDATA can be searched for genes of interest using Arabidopsis Genome Initiative identifiers, for selected papers by means of PubMed IDs, authors and specific keywords. The results page contains details of the original publications, text fragments from the curated literature grouped according to information types and direct links to PubMed pages of the original papers.

Conclusions: The LEAFDATA database offers access to searchable entries curated from a large number of scientific publications. Due to the unprecedented details of annotations and the fact that LEAFDATA already provides records about approximately 1600 individual loci, this database is useful for the entire plant research community.

Keywords: Database, Arabidopsis thaliana, Leaf development, Manual literature curation

Background

Leaf development from primordium initiation to organ senescence is an intricate process controlled by interconnected regulatory pathways [1, 2]. Many of the key genes have been thoroughly characterized, while the role of numerous other factors with clear leaf phenotypes has not been studied in the context of leaf organogenesis. The shoot apical meristem (SAM) gives rise to the aboveground differentiated organs. The position of leaf initiation is determined by polarized auxin accumulation generated by the YUCCA auxin biosynthesis genes [3] and the PIN-FORMED1 (PIN1) hormone transporter [4]. Leaf identity is established by suppression of meristem identity genes at this marked region by the MYB-family transcription factor ASYMMETRIC LEAVES1 (AS1) and AS2, a LOB domain protein coding gene [5, 6]. A defined boundary region separates the meristem from the organ primordium and provides a border between neighboring organs. Organization of this domain depends on factors including CUP-SHAPED COTYLEDON (CUC) genes, LATERAL ORGAN BOUNDARIES (LOB), LATERAL ORGAN FUSION (LOF1), and JAGGED LATERAL ORGAN (JLO) genes [7]. The early leaf primordium emerges as radially symmetrical, cylindrical structure that soon differentiates along the proximodistal, mediolateral and dorsoventral axes. For the formation of a flattened leaf structure, mutually antagonistic developmental programs define the dorsal and ventral organ identity [8]. AS1, AS2 and the HD-ZIPIII genes act as ventral determinants, while the KANADI (KAN) genes, the YABBY genes and several AUXIN RESPONSE FACTORS (ARFs) promote ventral fate. Leaf growth is a coordinated process of cell division and cell expansion. Cell divisions are...
stress was studied in several stages of leaf development using transcript profiling and quantitative proteomics experiments [27]. These datasets along with metabolite measurements, photosynthesis and respiration rates, enzyme activities, ribosome numbers and lipid content are accessible at the project’s data integration and data sharing portal [28]. In the framework of this project a novel literature curation method was developed using the Leaf Knowtator tool and 283 key publications were processed as a community effort [29]. It was demonstrated that the collected information could be integrated with other public resources and a relational database, KnownLeaf was created. Furthermore, a graphical network was built to facilitate knowledge mining. However, access to the curated is data is hindered by the lack of a web interface. Therefore, our main aim was to establish a convenient resource with reliable query functions for easy access to this curated library.

Here, we present LEAFDATA, a high-quality and freely available literature database for Arabidopsis leaf development. By searching and manually curating 380 primary research publications, we collected 13,553 statements about genes that were experimentally linked to leaf organogenesis. We have created LEAFDATA to support fundamental research and provide a solid information resource for our users.

**Construction and content**

**Data collection**

LEAFDATA records were collected by employing the customized Leaf Knowtator annotation tool [29]. This interface runs in Protégé software version 3.3.1 using and the Knowtator plug-in version 1.9 beta [30]. Results sections of full-text primary research papers are processed. Entries are collected into ten major categories: phenotype, gene expression, feature, DNA–protein interaction, protein–protein interaction, genetic interaction, process, regulation of gene expression, regulation of process, and regulation of phenotype (Table 1). All categories have predefined structures and information slots attached to them that can be filled with ontology terms already uploaded into Leaf Knowtator (Table 2). The main controlled ontology collections that are included in this project are Plant Ontology (PO) [31], BRENDA Tissue Ontology (BTO) [32], Phenotype, Attribute and Trait Ontology (PATO) [33], Plant Trait Ontology (TO) [34], Molecular Interaction (MI) [35], Plant Environment Ontology (EO) [36] and Gene Ontology (GO) [37]. Genes were associated with the specific AGI identifiers derived from the TAIR10 genome annotation [38]. In addition, the Knowtator plug-in automatically saves further details such as the annotated file, annotator and annotated text. The curation system is flexible and can be easily modified...
to other annotation projects. Required slots are filled with terms closely following the original text. In addition to the community curations from 283 publications from the AGRON-OMICS project, 97 new papers were processed.

**Database construction**

Annotations were exported as XML files from Leaf Knowtator. These files are small and easy to share. The XML files were transformed into a single table with a custom made Perl script [24] and loaded in bulk using Structured Query Language (SQL) queries. The LEAFDATA Database resides on the MS SQL Server 2008 platform. The website design is fully responsive in line with current industry standards and is based on the Bootstrap Framework. Bootstrap utilizes HTML, CSS, and JS frameworks for developing responsive projects on the web. For database integration the server side engine Adobe Coldfusion 9 running over MS IIS was chosen for its relatively inexpensive hosting costs, its rapid development credentials and powerful data collation functions. The employment of these cutting-edge technologies offers a modern, literature-curated website that can be used on any device and provide fast access to our data in any research environment.

**Utility**

**LEAFDATA home**

The main site (Fig. 1) provides direct access to the search functions. There is a visual representation of the database content including number of curated publications and individual statements, and details of the different categories. Upon selecting any categories, all annotations can be retrieved. On the bottom of the page, a news section can be found directly connected to an active Twitter account with announcements of relevant publications and database updates. Moreover, a direct contact form is available for any enquiries.

**LEAFDATA search tools**

LEAFDATA provides four convenient search functions. Genes of interest can be queried by using unique AGI identifiers based on the last TAIR10 genome release. All annotations can be retrieved from a selected publication using the PubMed ID. In addition to an author query, we also offer a keyword search. Results are arranged according to distinct categories and individual publications. For illustration, records from an AGI search for the **HD-ZIPIII** transcription factor **REVOLUTA (REV)** is shown (Fig. 2; Table 3). This query resulted in 78 statements from 17 different papers. The keyword tool is particularly helpful to attain required information. It allows combining multiple keywords and limits the search results to only those documents that contain all the terms. This function can be used effectively to find plant lines that share a certain phenotype, genes with the same biological function or similar expression domains. Recent publications revealed that genetic combinations of plant lines with increased leaf size can further enhance growth [39]. In order to find all the large-leaf Arabidopsis lines curated in LEAFDATA, we performed a search for the terms **size_PATO:0000586** and **increased size_PATO:0000117** and retrieved a preliminary list of 173 statements (Additional file 1). Ontology terms were used to minimize the recovery of false positive records and ‘plant part’ was not specified to maximize the number of genuine hits. Terms with similar meanings can be used for this query. For example, large leaves, big leaves, increased leaf size gave 162, 12, and 373 results, respectively (Additional file 2: Table S1). Ten statements were randomly selected for additional data mining (Table 4). First AGI codes were collected from the LEAFDATA gene list available under the SEARCH LEAFDATA tab (see also Additional File

**Table 1 Information types annotated in LEAFDATA**

| Category                  | Example                                                                 | Number |
|---------------------------|-------------------------------------------------------------------------|--------|
| Phenotype                 | The venation in each as2 leaf lamina was bilaterally asymmetrical       | 6559   |
| Gene expression           | YAB3 is detected in the abaxial regions of the developing leaves        | 4617   |
| Feature                   | the AP2/EREBP domain of LEP is located close to the N-terminus of the protein | 731    |
| DNA–protein interactions  | BES1 binding to the promoter of SAUR-15                                  | 151    |
| Protein–protein interaction | AtCul1 … co-immunoprecipitated with … myc-tagged ASK1                 | 345    |
| Genetic interaction       | se quantitatively and qualitatively enhanced the lobing of as1 … leaves | 382    |
| Process                   | CYCD3;1 … important for the initiation of cell division at the G1 phase in leaves | 348    |
| Regulation of gene expression | we conclude that STM negatively regulates AS1                          | 171    |
| Regulation of process     | WRKYS3, is an important positive regulator of senescence                | 206    |
| Regulation of phenotype   | NEK6 … promotes biomass levels                                            | 42     |

Ten major classes of information are curated in our database. Examples of these categories and number of statements are shown here.
Table 2 Phenotype annotation exported from Leaf Knowtator

| Database columns | Entries |
|------------------|---------|
| File             | 21401745.txt.knowtator.xml |
| Class            | Phenotype |
| Annotator        | Dora Szakonyi, LEAFDATA |
| Spanned text     | The rid2-1 mutant was temperature sensitive for seedling growth as well as for callus formation. In rid2-1 seedlings grown at 22 °C, the true leaves were pointed |
| Annotated text   | Genotype ID = rid2-1|Property Slot = NULL|Value ID = pointed|Plant part ID = leaves |
| Growth condition |                     |
| Developmental stage |                 |
| Plant part       | leaf_PO:0025034 |
| Localisation     |                     |
| Property         | shape_PATO:000052 |
| Process          |                     |
| Value            | pointed_PATO:0002258 |
| Regulation       |                     |
| Gene expression  |                     |
| Gene studied     |                     |
| Interaction type |                     |
| Protein studied  |                     |
| Interactor protein |                 |
| Protein          |                     |
| Gene target      |                     |
| Genetic interactor |                |
| DNA target       |                     |
| Genotype ID      | mutated gene_MI:0804 |
| Genotype details | Gene ID = AT5G57280 | Genotype_Zygo- sity = homozygous diploid_APO:0000229 | Mutant LOF_GOF ID = loss of function_ APO:0000011 |

Factuality

In LEAFDATA, a complex range of information is attached to each displayed statements deposited as spanned text. Different classes have distinct predefined slots however records are converted into a single table. Database columns corresponds to all available Leaf Knowtator slots. File, Annotator, Spanned text and Annotated text information is automatically added to curated statements.

Table 2: Phenotype annotation exported from Leaf Knowtator

Discussion

Leaves are essential organs for plant life and the location of multiple biological processes. Organogenesis from emergence of leaf primordium through pattern formation, maturation, maintenance until senescence is regulated by diverse regulatory pathways. Genetic and molecular roles of numerous genes were described in great detail. These genes are classified as key players in leaf morphogenesis. However, numerous additional genes causing altered leaf morphology have been isolated. In many cases, characterization of the observed leaf phenotypes are not main scope of these studies. Furthermore, these information are scattered throughout the existing scientific literature. Our aim was to create a convenient public collection of relevant leaf literature that provides simple query functions and easy access to a large library at the same time. Here, we demonstrate that our published annotation method and the Leaf Knowtator interface [29] can be used effectively for establishing high-quality literature resources. Employing this system guaranteed several unique database features. With a quick workflow, we are able to retain a large amount of information. In LEAFDATA, not only are the curated text fragments from the original publications kept and displayed but ontology terms from established structured vocabularies are simultaneously attached to these statements. Using these standardized terms helps building complex queries and can facilitate data sharing and integration [40]. We adhere to further community standards by employing the entity–attribute–value (EAV) model for phenotype annotations [41]. On average, more than 35 annotations per publication are generated adding up to a total of 13,553 independent statements about nearly 1300 genes. A major advantage of our database is that our curations are not restricted to single genotypes or information types. For instance, phenotype annotations can cover descriptions of single and multiple mutants (Table 3) as well as constitutive or inducible overexpressors, transgenic plants expressing chimeric constructs or modified versions of the gene of interest. Also, gene expression records provide an exceptional range of information including quantification of expression levels and spatial distribution in wild type or various mutant backgrounds (Table 3). Most of our annotations belong to the phenotype and gene expression class however numerous protein–protein interaction, genetic interaction and DNA–protein interaction records can be accessed (Table 1). The original publication details (author, title, PubMed ID) are clearly displayed with each statements and a direct link is provided to the dedicated PubMed
The LEAFDATA resource collects information from primary research articles focusing on Arabidopsis leaf growth and development. The result section of selected papers is manually annotated and organized into distinct categories. Text fragments of the manuscripts are displayed with links to the original papers on PubMed. The database currently hosts 13,553 annotations from 380 publications.

Fig. 1 LEAFDATA home. Four key search tools, a summary of the database content, a news section connected to an active Twitter account and a contact form can be reached from the LEAFDATA main page.
page. The search functions were designed to give a quick access to records from a chosen gene, paper or author. The keyword query allows more detailed data mining e.g. for a specific genotype using multiple terms. In summary, the combination of the LEAFDATA tools can be used effectively to collect wide-range of information (Table 4).

LEAFDATA is a useful platform not only for researchers interested in leaf development but for scientists working with other traits, plant species or model organisms. There are possible applications for our dataset in large-scale projects, mutagenesis screens and developing text-mining tools. University students, interested professionals and the general public can benefit from free and
Table 3 Results using AGI search for AT5G60690

| Phenotype | PubMed ID: 10353913 | Berná G, Robles P, Micol JL—A mutational analysis of leaf morphogenesis in Arabidopsis thaliana rev ... displaying large ... rosette ... leaves
| PubMed ID: 16284709 | Horiguchi G, Ferjani A, Fujikura U, Tsukaya H—Coordination of cell proliferation and cell expansion in the control of leaf size in Arabidopsis thaliana known large-leaf mutants ... rev
| PubMed ID: 20807212 | Kim HS, Kim SJ, Abbasi N, Bressan RA, Yoon DJ, Yoo SD, Kwon SY, Choi SB—The DOF transcription factor Dof5.1 influences leaf axil patterning by promoting Revoluta transcription in Arabidopsis the 35S:REV Arabidopsis plant displayed an upwardly curled leaf phenotype (Mallory et al. 2004)
| PubMed ID: 22003085 | Wang W, Xu B, Wang H, Li J, Huang H, Xu L—YUCCA genes are expressed in response to leaf adaxial-abaxial juxtaposition and are required for leaf margin development
| | as2-1 rev-1 ... there were many more margin cell patches than in the wild type
| | as2-1 rev-1 yuc1 yuc2 yuc4 ... The margin formation ... was severely defective
| | as2-1 rev-1 ... leaves ... had fewer hydathodes than the wild type
| | finger-shaped protrusions ... were observed on the adaxial side of late-appearing leaves of as2-1 rev-1
| | as2-1 rev-1 ... cell patches ... could occur anywhere among the margin cells
| | The rosette leaves of one as2-1 rev-1 yuc1 yuc2 yuc4 plant ... leaf margin curled upward
| | as2-1 rev-1 yuc1 yuc2 yuc4 ... leaves, which were filamentous
| | treatment with increased NPA concentrations resulted in decreased plant size of as2-1 rev-1
| | as2-1 rev-1 ... had lobes on their rosette leaves
| | Compared with the as2-1 rev-1 double mutant ... as2-1 rev-1 yuc1 yuc2 yuc4/+ plants were reduced in size
| | The rosette leaves of one as2-1 rev-1 yuc1 yuc2 yuc4 plant ... were narrow
| | as2-1 rev-1 yuc1 yuc2 yuc4 ... lacking typical margin cells
| PubMed ID: 7555701 | Talbert PB, Adler HT, Parks DW, Comai L—The REVOLUTA gene is necessary for apical meristem development and for limiting cell divisions in the leaves and stems of Arabidopsis thaliana
| | leaves of rev-1 mutants grew abnormally large
| | The rosette leaves of rev-1 plants were not readily distinguishable from wild-type No-0 leaves prior to bolting. As bolting began, however, the youngest rosette leaves became abnormally large ... as they matured
| | The rev-1 mutation caused overgrowth of both rosette and cauline leaves
| | The rosette leaves of rev-1 plants were not readily distinguishable from wild-type No-0 leaves prior to bolting. As bolting began, however, the youngest rosette leaves became ... distorted or uneven in shape as they matured
| | mutants with a syndrome con-sisting of revolute (downwardly curled) leaves ... rev-1, rev-2 and rev-4
| | Both the leaves and the primary shoots of rev-1 mutants were often darker green than those of wild type
| | The difference in leaf size between wild-type and rev-1 plants was not obvious in the earlier rosette leaves, but we measured significant size differences in the cotyledons and first and third leaves from cohort 3 (Table 2). Later leaves differed more dramatically: the mean length of the longest rosette leaf (ordinarily the youngest leaf) of rev-1 plants was about 39% longer than wild-type controls, and rev-1 cauline leaves became up to twice as long as their wild-type counterparts
| Gene expression | PubMed ID: 11525739 | Eshed Y, Baum SF, Perea JV, Bowman JL—Establishment of polarity in lateral organs of plants
| | mRNA of REV ... Later expression is confined to the provascular and vascular tissues of leaves
| | mRNA of REV ... restricted to the adaxial domain as developing primordia
| | REV ... initiates normally in kan1 kan2 leaf primordia
| | mRNA of REV ... is localized to the SAM
| | REV ... in kan1 kan2 leaf primordia ... confinement to the adaxial domain is delayed
| | mRNA of REV ... is localized to the SAM, throughout leaf primordia anlagen
| PubMed ID: 12615938 | Nelissen H, Clarke JH, De Block M, De Block S, Vanderhaeghen R, Zielinski RE, Dyer T, Lust S, Inzé D, Van Lijsebettens M—DRL1, a homolog of the yeast TOT4/KTI12 protein, has a function in meristem activity and organ growth in plants
| | pREVOLUTA:GUS ... were introgressed into drl1-2 ... marker lines displayed promoter activity in the dorsal part of the leaf primordium, including the vascular bundles
| PubMed ID: 16682355 | Garcia O, Collier SA, Byrne ME, Martienssen RA—Specification of leaf polarity in Arabidopsis via the trans-acting siRNA pathway
| | REVOLUTA ... Expression is re-stricted to the adaxial domain
Xu L, Yang L, P.L, Liu Q, Ling Q, Wang H, Poethig RS, Huang H—Genetic interaction between the AS1-AS2 and RDR6-SGS3-AGO7 pathways for leaf morphogenesis leaves of the sgs3-11 as2-101 leaves contained reduced levels of REV transcripts REV transcripts were detected in the earlier stage leaf primordia in as2-101 leaves of the ... zip as1-101 leaves contained reduced levels of REV transcripts REV transcripts were detected in the earlier stage leaf primordia in sgs3-11 as2-101 rdr6-3 as2-101 leaves contained reduced levels of REV transcripts REV transcripts were detected in the earlier stage leaf primordia in as2-101 leaves of the zip as2-101 leaves contained reduced levels of REV transcripts REV transcripts were detected in the earlier stage leaf primordia in sgs3-11 as1-101 leaves of the zip as2-101 leaves contained reduced levels of REV transcripts REV transcripts were detected in the earlier stage leaf primordia in sgs3-11 as2-101

Iwakawa H, Iwasaki M, Kojima S, Ueno Y, Soma T, Tanaka H, Semiarti E, Machida Y, Machida C—Expression of the ASYMMETRIC LEAVES2 gene in the adaxial domain of Arabidopsis leaves represses cell proliferation in this domain and is critical for the development of properly expanded leaves in as2-1 ... No significant differences in levels of transcripts were detected for REV in as1-1 ... No significant differences in levels of transcripts were detected for REV

Vandenbussche M, Horstman A, Zethof H, Koes R, Rijpkema AS, Gerats T—Differential recruitment of WOX transcription factors for lateral development and organ fusion in Petunia and Arabidopsis In Arabidopsis, organ polarity regulation has been characterized extensively at the molecular level. We therefore have monitored the expression levels of a selection of Arabidopsis genes known to be involved in abaxial/adaxial patterning. The selection of genes comprises REVOLUTA ... None of the monitored Arabidopsis genes exhibited significant changes in transcript levels between wox1 prs mutant samples and the wild type

Kim HS, Kim SJ, Abbasi N, Bressan RA, Yun DJ, Yoo SD, Kwon SY, Choi SB—The DOF transcription factor Dof5.1 influences leaf axilary patterning by promoting Revoluta transcription in Arabidopsis REV and ATHB-15 transcripts were strongly enhanced in Dof5.1-1D (Figure 6a); however, ATHB-15 was not increased in DEX:Dof5.1 plants upon DEX treatment whereas REV transcript was enhanced (Figure 6b). Increased expression of ATHB-15 in Dof5.1-D is probably due to a secondary effect on REV expression was decreased in 35S::Dof5.1ΔAct plants although the levels were different depending on transgenic lines

Kim HS, Kim SJ, Abbasi N, Bressan RA, Yun DJ, Yoo SD, Kwon SY, Choi SB—The DOF transcription factor Dof5.1 influences leaf axilary patterning by promoting Revoluta transcription in Arabidopsis REV and ATHB-15 trans- ... REV:REV-VENUS expressed in the apical and central regions of wild-type embryos

Horiguchi G, Mollá-Morales A, Pérez-Pérez JM, Kojima K, Robles P, Ponce MR, Micol JL, Tsukaya H—Differential contributions of ribosomal protein genes to Arabidopsis thaliana leaf development agreement with the synergistic polarity defects observed in rpl4d-3 as2-1 ... expression ... REV were expressed at similar levels in the ... parents

Wang W, Xu B, Wang H, Li J, Huang H, Xu L—YUCCA genes are expressed in response to leaf adaxial-abaxial juxtaposition and are required for leaf margin development pDR5::GUS staining in the as2-1 rev-1 leaves was also concentrated in the top portion of the protrusions YUC4 was expressed in the leaf protrusions of the mock-treated as2-1 rev-1 as2-1 rev-1 ... when ... treated with 1 μM NPA ... we did not observe the small GUS-staining spots on leaf surfaces

Wang W, Xu B, Wang H, Huang H, Cui X—Elongator complex is critical for cell cycle progression and leaf patterning in Arabidopsis We next examined expression of leaf polarity marker genes REVOLUTA (REV) and FILAMENTOUS FLOWER (FIL) in the elo2 as2 background. rev-9 is a T-DNA enhancer trap line in which β-glucuronidase (GUS) staining represents the expression of the leaf adaxial marker REV (Emery et al. 2003; Hawker and Bowman 2004). Compared with that in the rev-9/+ plant (Fig. 1q), GUS staining was not detected from needle-like leaves of the elo2 as2 rev-9/+ plant

Ben Chaabane S, Liu R, Chinnusamy V, Kwon Y, Park JH, Kim SY, Zhu JK, Yang SW, Lee BH.—STA1, an Arabidopsis pre-miRNA processing factor 6 homolog, is a new player involved in miRNA biogenesis Compared with WT, the accumulation of REV transcripts was higher in sta1-1, which is linked to decreased miR164/166 levels and explains the serrated leaf phenotype of sta1-1

Table 3 continued

Table 3

PubMed ID: 16699177
Xu L, Yang L, P.L, Liu Q, Ling Q, Wang H, Poethig RS, Huang H—Genetic interaction between the AS1-AS2 and RDR6-SGS3-AGO7 pathways for leaf morphogenesis leaves of the sgs3-11 as2-101 leaves contained reduced levels of REV transcripts REV transcripts were detected in the earlier stage leaf primordia in as2-101 leaves of the zip as1-101 leaves contained reduced levels of REV transcripts REV transcripts were detected in the earlier stage leaf primordia in sgs3-11 as2-101 rdr6-3 as2-101 leaves contained reduced levels of REV transcripts REV transcripts were detected in the earlier stage leaf primordia in as2-101 leaves of the zip as2-101 leaves contained reduced levels of REV transcripts REV transcripts were detected in the earlier stage leaf primordia in sgs3-11 as2-101

PubMed ID: 17559509
Iwakawa H, Iwasaki M, Kojima S, Ueno Y, Soma T, Tanaka H, Semiarti E, Machida Y, Machida C—Expression of the ASYMMETRIC LEAVES2 gene in the adaxial domain of Arabidopsis leaves represses cell proliferation in this domain and is critical for the development of properly expanded leaves in as2-1 ... No significant differences in levels of transcripts were detected for REV in as1-1 ... No significant differences in levels of transcripts were detected for REV

PubMed ID: 19717616
Vandenbussche M, Horstman A, Zethof H, Koes R, Rijpkema AS, Gerats T—Differential recruitment of WOX transcription factors for lateral development and organ fusion in Petunia and Arabidopsis In Arabidopsis, organ polarity regulation has been characterized extensively at the molecular level. We therefore have monitored the expression levels of a selection of Arabidopsis genes known to be involved in abaxial/adaxial patterning. The selection of genes comprises REVOLUTA ... None of the monitored Arabidopsis genes exhibited significant changes in transcript levels between wox1 prs mutant samples and the wild type

PubMed ID: 20628155
Sarojam R, Sappl PG, Goldshmidt A, Efroni I, Floyd SK, Eshed Y, Bowman JL—Differentiating Arabidopsis shoots from leaves by combined YABBY activities YABBY triple mutants display reduced expression of REV

PubMed ID: 20707212
Kim HS, Kim SJ, Abbasi N, Bressan RA, Yun DJ, Yoo SD, Kwon SY, Choi SB—The DOF transcription factor Dof5.1 influences leaf axilary patterning by promoting Revoluta transcription in Arabidopsis REV and ATHB-15 transcripts were strongly enhanced in Dof5.1-1D (Figure 6a); however, ATHB-15 was not increased in DEX:Dof5.1 plants upon DEX treatment whereas REV transcript was enhanced (Figure 6b). Increased expression of ATHB-15 in Dof5.1-D is probably due to a secondary effect on REV expression was decreased in 35S::Dof5.1ΔAct plants although the levels were different depending on transgenic lines

PubMed ID: 21223391
Szakonyi D, Byrne ME—Ribosomal protein L27a is required for growth and patterning in Arabidopsis thaliana REV:REV-VENUS ... expressed in the apical and central regions of wild-type embryos

PubMed ID: 21251100
Horiguchi G, Mollá-Morales A, Pérez-Pérez JM, Kojima K, Robles P, Ponce MR, Micol JL, Tsukaya H—Differential contributions of ribosomal protein genes to Arabidopsis thaliana leaf development agreement with the synergistic polarity defects observed in rpl4d-3 as2-1 ... expression ... REV were expressed at similar levels in the ... parents

PubMed ID: 22003085
Wang W, Xu B, Wang H, Li J, Huang H, Xu L—YUCCA genes are expressed in response to leaf adaxial-abaxial juxtaposition and are required for leaf margin development pDR5::GUS staining in the as2-1 rev-1 leaves was also concentrated in the top portion of the protrusions YUC4 was expressed in the leaf protrusions of the mock-treated as2-1 rev-1 as2-1 rev-1 ... when ... treated with 1 μM NPA ... we did not observe the small GUS-staining spots on leaf surfaces

PubMed ID: 22026817
Xu D, Huang W, Li Y, Wang H, Huang H, Cui X—Elongator complex is critical for cell cycle progression and leaf patterning in Arabidopsis We next examined expression of leaf polarity marker genes REVOLUTA (REV) and FILAMENTOUS FLOWER (FIL) in the elo2 as2 background. rev-9 is a T-DNA enhancer trap line in which β-glucuronidase (GUS) staining represents the expression of the leaf adaxial marker REV (Emery et al. 2003; Hawker and Bowman 2004). Compared with that in the rev-9/+ plant (Fig. 1q), GUS staining was not detected from needle-like leaves of the elo2 as2 rev-9/+ plant

PubMed ID: 23268445
Ben Chaabane S, Liu R, Chinnusamy V, Kwon Y, Park JH, Kim SY, Zhu JK, Yang SW, Lee BH.—STA1, an Arabidopsis pre-miRNA processing factor 6 homolog, is a new player involved in miRNA biogenesis Compared with WT, the accumulation of REV transcripts was higher in sta1-1, which is linked to decreased miR164/166 levels and explains the serrated leaf phenotype of sta1-1
easy access to the LEAFDATA library offering processed scientific records.

We envision future improvements for LEAFDATA. The current database contains approximately 15–20% of the published Arabidopsis leaf literature, is constantly being updated. However, it will take significant effort to annotate every existing leaf development paper and at the same time keep up with the steady flow of new research. We plan to develop advanced search functions for instance queries for specific phenotypic characteristics,

Table 3 continued

| PubMed ID | Authors | Title | Relevant Information |
|-----------|---------|-------|----------------------|
| 24464295  | Huang et al. | Arabidopsis KANADI1 acts as a transcriptional repressor by interacting with a specific cis-element and regulates auxin biosynthesis, transport, and signaling in opposition to HD-ZIPIII factors | REV increases transcription, most likely by direct activation, of NPY1 |
| 20807212  | Kim et al. | The DOF transcription factor Dof5.1 influences leaf axial patterning by promoting Revoluta transcription in Arabidopsis | REVOLUTA, specifying adaxial identity |
| 20807212  | Garcia et al. | Specification of leaf polarity in Arabidopsis via the trans-acting siRNA pathway | REVOLUTA, specifying adaxial identity |

Seventy-eight records are available in LEAFDATA for AT5G60690, REV gene. These statements are organized according to information type and original publication.
Table 4 Mining LEAFDATA for increased leaf size phenotype

| AGI       | Phenotype                                                                 | Gene expression                                                                 | Process                                                                                           |
|-----------|---------------------------------------------------------------------------|----------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|
| AT4G36380 | The rot3-2 allele causes enlarged leaf blades (10430960)                  | Leaves, epidermis, palisade tissue, and the spongy layer (10430960)              | Cell elongation (11889033, 9694802)                                                              |
| AT1G56010 | 3SS:NAC1 overexpressing lines ... were bigger (11114891)                  | Leaf primordia, nucleus, low levels in leaves (11114891)                         | Elongation of leaves (10430960)                                                                   |
| AT3G59900 | 3SS-ARGOS ... lines showed an enlarged ... leaf size (12566580)            | Young rosette leaves, juvenile leaf ... petioles, juvenile leaf blades, cytosol, nucleus, leaf primordia (12953103) Juvenile leaves (16824178) | Longitudinal cell expansion (17038516)                                                           |
| AT3G62000 | Homozygous plants of the arf2-6 ... have ... large ... rosette leaves (15960614) | ARF2 is expressed in all major plant organs including roots, rosette and cauline leaves, flowers and siliques (15960614, 16339187) | Controls later organ growth by affecting the duration of cell proliferation (16824178)             |
| AT3G13960 | ATGRF5 overexpressers ... developed leaves that were 20–30 % larger than those of the wild type (15960617) | Low in mature stems and leaves, shoot tips containing the shoot apical meristem (SAM) (12974814) Primordium, restricted to the lower half of the leaf primordium, undetectable in mature leaves (15960617) | De-ubiquitinating enzyme (18485060)                                                              |
| AT1G17110 | UBP15 overexpression lines revealed larger overall stature of the plants as well as larger rosette leaves (18485060) | Higher in rosette leaves, increased from the early to late leaf stages, with higher expression in the leaf margin in the late stage, present in both the cytosol and nucleus (18485060) | De-ubiquitinating enzyme (18485060)                                                              |
| AT4G18390 | TCP2 ... loss-of-function ... had ... slightly enlarged leaves (18816164)   | Shoot tips and shoot apical meristems (SAM), young leaves (19200151)             | De-ubiquitinating enzyme (18485060)                                                               |
| AT4G22270 | AtMRB1 overexpressor plants ... exhibited enlarged organ sizes (19200151)   | Shoot tips and shoot apical meristems (SAM), young leaves (19200151)             | De-ubiquitinating enzyme (18485060)                                                               |
| AT4G29040 | rpt2a-2 mutant ... displayed a phenotype of enlarged rosette leaves (19500299) | SAM, all the organs that we tested (flower bud, stem, leaf and root) (15073153) All organs tested, trichomes, expanded cotyledons, vascular cells, shoot meristem (19500299) | Proteasome activity (15073153)                                                                    |
| AT3G44200 | At the flowering stage, the two NEK6-overexpressing lines exhibited ... larger rosette than Col (21801253) | Leaves, young leaves, mature rosette leaves, vascular tissues, petioles (21801253) | Stress response, rosette growth, suppresses expression of several ethylene biosynthesis (21801253) |

Ten representative phenotype records were chosen from the keyword query for the terms size, _PATO:0000586_ and increased size, _PATO:0000117_. Subsequently, gene expression, process and regulation of biological process statements were collected from specific AGI searches. PubMed IDs of the parent publications are shown in brackets.
combinations of features or exclusion certain traits. Similarly, gene expression statements can be further explored by genotypes, changes in certain target genes or expression in special subcellular compartments, cell types and organs. Lastly, we are interested in data visualization and integration with other datasets.

Conclusions
The sheer amount of scientific literature is calling for carefully curated database summarizing experimental results. We employed the Leaf Knowtator curation system and constructed a unique, comprehensive database focusing on Arabidopsis leaf development. In addition to previously described regulators, genes with clear leaf phenotypes are included. The LEAFDATA collection gives access to 380 publications organized according to papers and information types. Four query functions provide easy access to high-quality annotations and direct links to the original papers. LEAFDATA serves as a valuable resource and reference point for the research community. Finally, our annotation approach, data organization and database structure can serve as a prototype for other literature curation projects.

Availability and requirements
LEAFDATA is an open access database at www.leafdata.org. The collection is updated on a regular basis. Questions, comments and requests regarding this database should be sent to Dóra Szakonyi at info@leafdata.org.

Details of LEAFDATA content and screenshots were recorded on 08/11/2015.

Additional files

- **Additional file 1.** Results using keyword search for size_PATO:0000586 increased size_PATO:0000117. A preliminary list of 173 results were recovered. Confirmed records are highlighted in yellow.
- **Additional file 2: Table S1.** Number of results retrieved by keywords search using different terms for large-leaf phenotype.
- **Additional file 3: Table S2.** List of genes curated in LEAFDATA.
- **Additional file 4: Table S3.** List of papers curated in LEAFDATA.

Acknowledgements
We would like to acknowledge Ross Austen at Databweaver International for all the help with the webpage creation. We thank Sofie Van Landeghem and Dean Ravenscroft for feedback on the LEAFDATA resource. We are grateful for Amanda Rasmussen, Ken W. Berendzen and Gerda Szakonyi for useful comments on the database and the manuscript. This work was conducted using the Protégé resource, which is supported by Grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.

Competing interests
The author declares that he has no competing interests.

References
1. Kalve S, De Vos D, Beemster GT. Leaf development: a cellular perspective. Front Plant Sci. 2014;5:362.
2. Rodriguez RE, Debernardi JM, Palatnik JF. Morphogenesis of simple leaves: regulation of leaf size and shape. Wiley Interdiscip Rev Dev Biol. 2014;3(1):41–57.
3. Cheng Y, Dai X, Zhao Y. Auxin synthesized by the YUCCA flavin monoxygenases is essential for embryogenesis and leaf formation in Arabidopsis. Plant Cell. 2007;19(8):2430–9.
4. Reinhart DJ, Pesce ER, Stieger P, Mandel T, Baltensperger K, Bennett M, et al. Regulation of phyllotaxis by polar auxin transport. Nature. 2003;426(6964):255–60.
5. Byrne ME, Barley R, Curtis M, Arroyo JM, Dunham M, Hudson A, Martienssen RA. Asymmetric leaves:1 mediates leaf patterning and stem cell function in Arabidopsis. Nature. 2000;408(6815):967–71.
6. Semeraro E, Ueno Y, Tsukaya H, Ikawaoka H, Machida C, Machida Y. The ASYMERIC LEAVES2 gene of Arabidopsis thaliana regulates formation of a symmetric lamina, establishment of venation and repression of meristem-related homeobox genes in leaves. Development. 2001;128(10):1771–83.
7. Zadnikova P, Simon R. How boundaries control plant development. Cur Opin Plant Biol. 2014;17:116–25.
8. Szakonyi D, Moschopoulos A, Byrne ME. Perspectives on leaf dorsoventral polarity. J Plant Res. 2010;123(3):281–90.
9. Blomme J, Inze D, Gonzalez N. The cell-cycle interactome: a source of growth regulators? J Exp Bot. 2014;65(21):7215–30.
10. Lau OS, Bergmann DC. Stomatal development: a plant’s perspective on cell polarity, cell fate transitions and intercellular communication. Development. 2012;139(20):3683–92.
11. Pattanakij S, Patra B, Singh SK, Yuan L. An overview of the gene regulatory network controlling trichome development in the model plant, Arabidopsis. Front Plant Sci. 2014;5:259.
12. Gentry M, Hennig L. Remodelling chromatin to shape development of plants. Exp Cell Res. 2014;321(1):40–6.
13. Tsukaya H, Byrne ME, Horiguchi G, Sugiyama M, Van Lijsebettens M, Lenhardt M. How do ‘housekeeping’ genes control organogenesis?—Unexpected new findings on the role of housekeeping genes in cell and organ differentiation. J Plant Res. 2013;126(1):3–15.
14. Kidner CA. The many roles of small RNAs in leaf development. J Genet Genomics. 2010;37(1):13–21.
15. Kurepin LV, Pharis RP. Light signaling and the phytohormonal regulation of shoot growth. Plant Sci. 2014;229:280–9.
16. Perez-Perez JM, Candela H, Robles P, Quesada V, Ponce MR, Micol JL. Lessons from a search for leaf mutants in Arabidopsis thaliana. Int J Dev Biol. 2009;53(8–10):1623–34.
17. PHENOPSIS DB. http://bioweb.supagro.inra.fr/phenopsis/. Accessed 08 Nov 2015.
18. Perez-Perez JM, Dauzat M, Negre V, Wuys N, Tireau A, Gennari E, et al. PHENOPSIS DB: an information system for Arabidopsis thaliana phenotypic data in an environmental context. BMC Plant Biol. 2011;11:77.
19. Graniere C, Aguirrezabal L, Chenu K, Cookson SJ, Dauzat M, Hamard P, et al. PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in Arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. New Phytol. 2006;169(3):623–35.
20. Wilson-Sanchez D, Rubio-Diaz S, Munoz-Viana R, Perez-Perez JM, Jover-Gil S, Ponce MR, et al. Leaf phenomics: a systematic reverse genetic screen for Arabidopsis leaf mutants. Plant J. 2014;79(5):878–91.
21. PhenoLeaf. http://genetics.unh.edu/phenoLeaf/index.php. Accessed 08 Nov 2015.
22. Das A, Bucksch A, Price CA, Weitz, JS. ClearedLeavesDB: an online database of cleared plant leaf images. Plant Methods. 2014;10(1):18.
23. ClearedLeavesDB. http://clearedleavesdb.org/. Accessed 08 Nov 2015.
24. Liu X, Li Z, Jiang Z, Zhao Y, Peng J, Jin J, et al. LSD: a leaf senescence database. Nuc Acids Res. 2011;39(Database issue):D103–7.
25. Li Z, Zhao Y, Liu X, Peng J, Guo H, Luo J. LSD 2.0: an update of the leaf senescence database. Nucl Acids Res. 2014;42:D1200–5.
26. Leaf Senescence Database. http://www.eplantsenesence.org/. Accessed 08 Nov 2015.
27. Baerenfaller K, Massonnet C, Walsh S, Baginsky S, Buhlmann P, Hennig L, et al. Systems-based analysis of Arabidopsis leaf growth reveals adaptation to water deficit. Mol Syst Biol. 2012;8:606.
28. AGRON-OMICS data integration and data sharing portal. https://agron-omics.ethz.ch/. Accessed 08 Nov 2015.
29. Szakonyi D, Van Landerghem S, Baerenfaller K, Baeyens L, Blomme J, Casanova-Sáez P, et al. The KnownLeaf literature curation system captures knowledge about Arabidopsis leaf growth and development and facilitates integrated data mining. Curr Plant Biol. 2015;2015(2):1–11.
30. Ogren PV. Knowtator: a protégé plug-in for annotated corpus construction. In: Proceedings of the 2006 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume: demonstrations; New York, New York. 1225791: Association for Computational Linguistics; 2006. p. 273–275.
31. Jaiswal P, Avraham S, Ilic K, Kellogg EA, McCouch S, Pujar A, et al. Plant ontology (PO): a controlled vocabulary of plant structures and growth stages. Comp Funt Genomics. 2005;6(7–8):388–97.
32. Gremsle M, Chang A, Schomburg I, Grote A, Scheer M, Ebeling C, et al. The BREnda tissue ontology (BTO): the first all-integrating ontology of all organisms for enzyme sources. Nucl Acids Res. 2011;39(Database issue):D507–13.
33. Gkoutos GV, Green EC, Mallon AM, Hancock JM, Davidson D. Using ontologies to describe mouse phenotypes. Genome Biol. 2005;6(1):R8.
34. Jaiswal P, Ware D, Ni J, Chang K, Zhao W, Schmidt S, et al. Gramene: development and integration of trait and gene ontologies for rice. Comp Funct Genomics. 2003;3(2):132–6.
35. Hemmajob H, Montecchi-Palazzi L, Bader G, Wojcik J, Salwinski L, Ceol A, et al. The HUPO PSIs molecular interaction format—a community standard for the representation of protein interaction data. Nat Biotechnol. 2004;22(2):177–83.
36. Youens-Clark K, Buckler E, Casstevens T, Chen C, Dedeker G, Derwent P, et al. Gramene database in 2010: updates and extensions. Nucl Acids Res. 2011;39(Database issue):D1085–94.
37. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25–9.
38. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucl Acids Res. 2012;40(Database issue):D1202–10.
39. Vanhaeren H, Inze D, Gonzalez N. Plant growth beyond limits. Trends Plant Sci. 2016;21(2):102–9.
40. Krajewski P, Chen D, Cwiek H, van Dijk AD, Fiorani F, Kersey P, et al. Towards recommendations for metadata and data handling in plant phenotyping. J Exp Bot. 2015;66(18):5417–27.
41. Miyoshi NS, Pinheiro DG, Silva WA Jr, Felipe JC. Computational framework to support integration of biomolecular and clinical data within a translational approach. BMC Bioinformatics. 2013;14:180.