FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*

Frédéric Bouché[†], Guillaume Lobet[†], Pierre Tocquin and Claire Périlleux[*]

PhytoSYSTEMS, Laboratory of Plant Physiology, University of Liège, Quartier Vallée 1 Sart Tilman Campus, 4 Chemin de la Vallée, 4000 Liège, Belgium

ABSTRACT

Flowering is a hot topic in Plant Biology and important progress has been made in *Arabidopsis thaliana* toward unraveling the genetic networks involved. The increasing complexity and the explosion of literature however require development of new tools for information management and update. We therefore created an evolutive and interactive database of flowering time genes, named FLOR-ID (Flowering-Interactive Database), which is freely accessible at http://www.flor-id.org. The hand-curated database contains information on 306 genes and links to 1595 publications gathering the work of >4500 authors. Gene/protein functions and interactions within the flowering pathways were inferred from the analysis of related publications, included in the database and translated into interactive manually drawn snapshots.

INTRODUCTION

The timing of flowering significantly affects plant fitness and crop yield so that the understanding of the underlying mechanisms is a primary source for further improvement of agricultural productivity. Reproductive success requires synchronization with seasonal cycling, thus plants have evolved a complex network of regulatory pathways to sense and integrate external and endogenous signals. Among the environmental cues, daylength and temperature are the main determinants of flowering time (1). As a result, pioneering forward genetic studies focused on the model plant *Arabidopsis thaliana* led to the categorization of flowering-time mutants according to their altered sensitivity to photoperiod ('photoperiodic pathway') and to winter cold ('vernalization pathway') (2,3). Mutants that were late flowering but remained sensitive to both environmental factors were classified in an 'autonomous pathway' whereas the limiting effect of gibberellins, a group of phytohormones, gave its name to the the 'gibberellin pathway' (4–6). Until

recently, the convergence of these four flowering pathways towards a few transcription factors, called 'integrators', was still the most visible crosstalk reflecting the fine-tuning of flowering time by known stimuli (7).

The view of flowering time control has however exploded over the last years and a picture of great complexity is emerging (8). First, dissection of the genetic networks underlying the pathways revealed multiple links between their components, such as light signaling and circadian timing in the photoperiodic pathway (9), which explained the pleiotropic phenotypes of some flowering-time mutants (10,11). Second, the identification of RNA processing and epigenetic regulation as major mechanisms of the autonomous and vernalization pathways revealed actors with rather generic roles and even became instrumental in unraveling new layers of gene regulation in plants (12,13). Third, additional pathways were uncovered after the investigation of the mechanisms promoting flowering in the absence of photoperiodic cues, as the plant ages or as the surrounding temperature rises (14–16). MicroRNAs were shown to have an important role (17), as well as sugar status and signaling (18–20) or other phytohormones (21,22).

This fast progress clearly demonstrates that the genetic dissection of flowering time control was greatly facilitated by the focus on Arabidopsis. As the view of the process evolved from a set of discrete pathways toward a complex network of interconnected hubs, the number of genes involved increased from 80 (23) to 180 (24). The highly intricate nature of these genetic networks, together with the diversity of the experimental evidence supporting their building, create a need to have a consolidated basis on which new players will be added. The question is even more challenging as new -omics high-throughput technologies are being used. We therefore undertook the construction of a core database named FLOR-ID (Flowering-Interactive Database) that could be progressively and regularly updated on the basis of new knowledge. We performed a careful literature survey and created a curated database of 306 flowering-time genes in Arabidopsis. We developed this tool as a website, meeting its dynamic and evolutive aims. A col-

[*]To whom correspondence should be addressed. Tel: +32 4 3663833; Fax: +32 4 3663831; Email: cperilleux@ulg.ac.be
[†]These authors contributed equally to this work.

lection of manually drawn snapshots provides an interactive user interface.

DATABASE CREATION

Identification of flowering-time associated genes

The FLOR-ID database aims at gathering information about genes involved in the regulation of flowering time. We defined 'flowering-time genes' as genes whose mutation and/or overexpression alters flowering time in any Arabidopsis accession. We allocated those genes among seven pathways whereby flowering is regulated by: photoperiod, vernalization, aging, ambient temperature, hormones, sugar or autonomously. Genes under the control of several converging pathways were defined as 'flowering-time integrators'.

To draw up an exhaustive list of flowering-time genes in Arabidopsis, we merged the results of four complementary approaches: (i) we started with a gene list published by Fornara *et al.* (24); (ii) we searched the UniProt database (http://www.uniprot.org) (25) for all the Arabidopsis proteins associated with the keyword 'flowering', excluding those isolated from pleiotropic mutants; (iii) we performed a literature search in NCBI PubMed (http://www.ncbi.nlm.nih.gov/pubmed), limited to the publications of the last 10 years; (iv) we analyzed recent reviews on the topic. More details are provided as Supplementary Material S1, including the full list of the reviews (8–11, 13–17, 22, 24, and supplementary references 28–114).

Gene information retrieval

For each identified flowering-time gene, we retrieved the relevant publications by querying TAIR10 (http://www.arabidopsis.org) (26), PubMed and Google Scholar (http://scholar.google.com) with the AGI locus identifier or the gene name(s). We narrowed these publication collections to the most representative and informative articles by primarily retaining those describing the cloning of the gene(s) or specifically dedicated to flowering time. When necessary, a further round of selection followed to keep most recent articles and/or those having a high number of citations.

The resulting compilation of more than 3000 full-text PDF files was then analyzed to find information about the phenotypes of the mutants, overexpressors and other transgenic lines. This was performed by contextual searches using FoxTrot Professional Search (CTM Development). Each phenotype was associated with the corresponding publication(s) and, when available, complementary information was added such as the growth conditions used for mutant phenotyping and functional data on the encoded protein. Genes were further sorted according to the flowering pathway in which they are primarily involved.

Database structure

All collected information about the flowering-time genes was incorporated into a normalized MySQL relational database: FLOR-ID (Figure 1). In addition to the information gathered from our literature survey, we used the AGI locus identifier and the reference PubMed ID to link
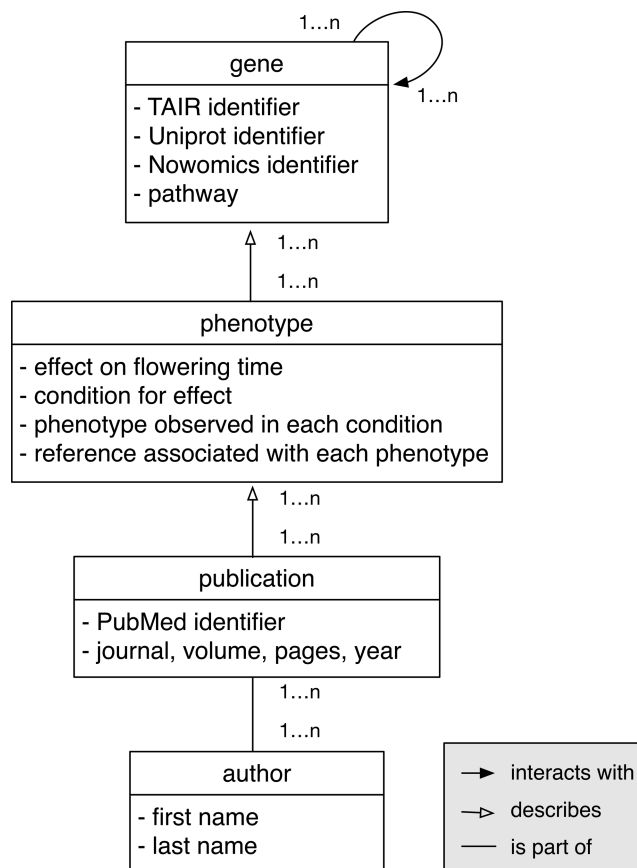

Figure 1. Simplified representation of the relational structure of the FLOR-ID database.

FLOR-ID to external databases: TAIR, PubMed, Uniprot and Nowomics (https://nowomics.com/).

USER INTERFACE

A freely accessible website was designed as a user frontend of FLOR-ID (http://www.flor-id.org). The database can be consulted by two ways: pre-compiled tabulated views and interactive schemes featuring flowering pathways and gene interactions.

Tabulated data

The raw content of the database is available through three thematic tables focused respectively on genes, publications and authors. Data can be further filtered by any textual variable and thereafter exported as text (CSV), Excel or PDF files. On a gene-by-gene basis, the 'Gene Details' pages of FLOR-ID give access to phenotypes of mutants and overexpressors, publications, protein function and interactors of the selected gene. 'Interactors' here refers to genes that have been shown to act upstream or downstream of the selected gene, whether their interaction is direct or not, or to proteins if protein-protein interactions were demonstrated (Figure 2). FLOR-ID also supports programmatic extraction of information for meta-analysis purposes. Custom URL addresses are used to access the gene information as
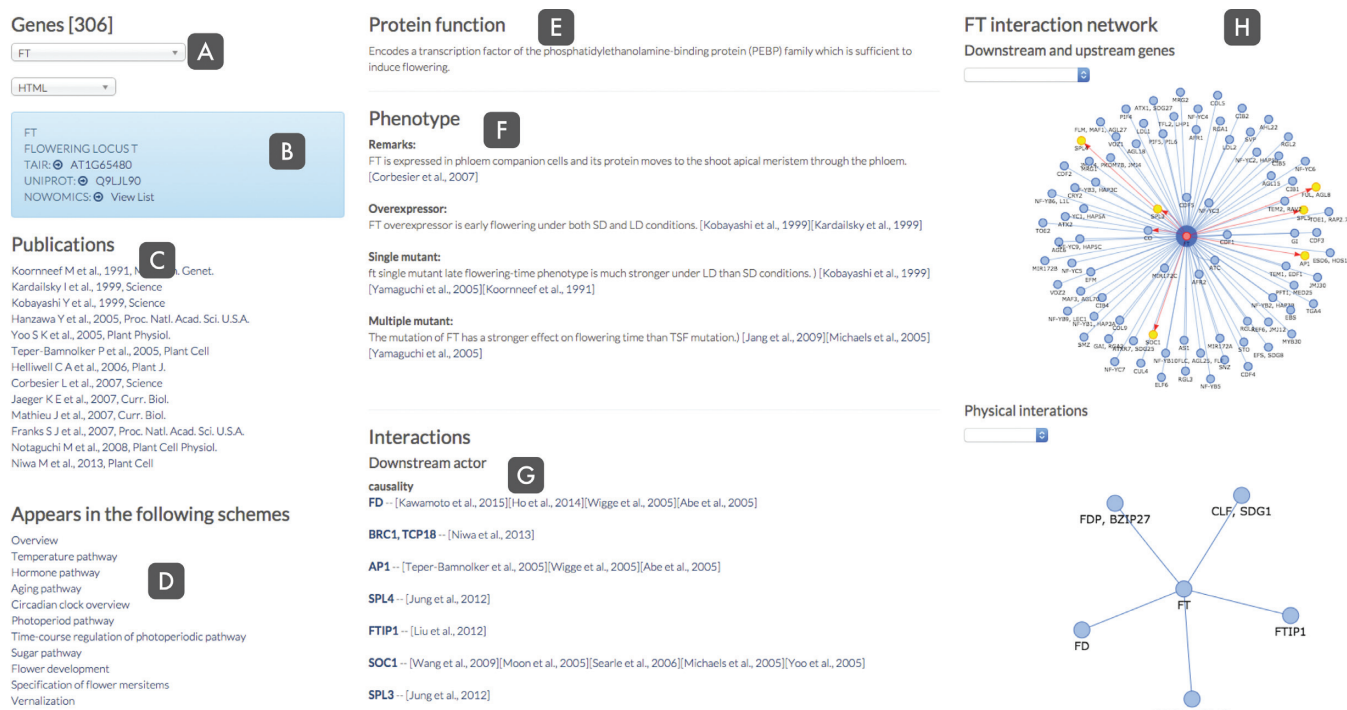
Genes [306] (A)

FT

HTML

FT
FLOWERING LOCUS T
TAIR: ⊕ AT1G65480
UNIPROT: ⊕ Q9LJL90
NOWOMICS: ⊕ View List (B)

Publications (C)

Koornneef M et al., 1991, M... n. Genet.
Kardailsky I et al., 1999, Science
Kobayashi Y et al., 1999, Science
Hanzawa Y et al., 2005, Proc. Natl. Acad. Sci. U.S.A.
Yoo S K et al., 2005, Plant Physiol.
Teper-Bamnolker P et al., 2005, Plant Cell
Helliwell C A et al., 2006, Plant J.
Corbesier L et al., 2007, Science
Jaeger K E et al., 2007, Curr. Biol.
Mathieu J et al., 2007, Curr. Biol.
Franks S J et al., 2007, Proc. Natl. Acad. Sci. U.S.A.
Notaguchi M et al., 2008, Plant Cell Physiol.
Niwa M et al., 2013, Plant Cell

Appears in the following schemes

Overview
Temperature pathway
Hormone pathway
Aging pathway (D)
Circadian clock overview
Photoperiod pathway
Time-course regulation of photoperiodic pathway
Sugar pathway
Flower development
Specification of flower mersitems
Vernalization

Protein function (E)

Encodes a transcription factor of the phosphatidylethanolamine-binding protein (PEBP) family which is sufficient to induce flowering.

Phenotype (F)

Remarks:
FT is expressed in phloem companion cells and its protein moves to the shoot apical meristem through the phloem. [Corbesier et al., 2007]

Overexpressor:
FT overexpressor is early flowering under both SD and LD conditions. [Kobayashi et al., 1999][Kardailsky et al., 1999]

Single mutant:
ft single mutant late flowering-time phenotype is much stronger under LD than SD conditions.) [Kobayashi et al., 1999] [Yamaguchi et al., 2005][Koornneef et al., 1991]

Multiple mutant:
The mutation of FT has a stronger effect on flowering time than TSF mutation.) [Jang et al., 2009][Michaels et al., 2005] [Yamaguchi et al., 2005]

Interactions

Downstream actor (G)
causality
FD -- [Kawamoto et al., 2015][Ho et al., 2014][Wigge et al., 2005][Abe et al., 2005]

BRC1, TCP18 -- [Niwa et al., 2013]

AP1 -- [Teper-Bamnolker et al., 2005][Wigge et al., 2005][Abe et al., 2005]

SPL4 -- [Jung et al., 2012]

FTIP1 -- [Liu et al., 2012]

SOC1 -- [Wang et al., 2009][Moon et al., 2005][Searle et al., 2006][Michaels et al., 2005][Yoo et al., 2005]

SPL3 -- [Jung et al., 2012]

FT interaction network (H)
Downstream and upstream genes

Physical interactions

FDP, BZIP27   CLF, SDG1

FT

FD   FTIP1

Figure 2. Gene-details window displayed in FLOR-ID. (**A**) Users can select any gene from the database (here *FLOWERING LOCUS T*) and find information displayed either in an HTML or XML format (for automated data retrieval). (**B**) Gene information, with links to TAIR, UniProt and Nowomics. (**C**) Relevant publications, with direct links to full-length articles. (**D**) FLOR-ID schemes providing network information for the gene. (**E**) Protein function. (**F**) Known flowering-time phenotype(s) of the corresponding mutant(s), associated with the original publication(s). (**G**) Known interactors, based on stated publications. (**H**) Graphical representation of the gene interaction network.

eXtended Markup Language (XML) or JavaScript Object Notation (JSON) files. For instance, retrieving information about *FLOWERING LOCUS T,* (AT1G65480) in an XML format is done with the URL http://www.phytosystems.ulg. ac.be/florid/details/?gene=AT1G65480&type=xml.

Interactive schemes

To give a visual structure to the FLOR-ID database, we produced snapshots illustrating different levels of complexity and enabling interactive access to the data. The snapshots are based on a careful analysis of the literature and supported by experimental evidence. One overview level, seven flowering pathways and two complementary pictures on circadian clock and flower development are connected to detailed nested schemes (37 in total). In each scheme, interactions between genes or proteins are shown by different line types and ending styles. From a technical point of view, FLOR-ID houses the schemes as Scalable Vector Graphics (SVG). SVG's are natively supported in every modern browser (Chrome 43, Firefox 38, Safari 8, Internet Explorer 11). Their vectorial nature ensures a minimal loading time. It also enables the tagging of each individual element (gene or line) which, combined with modern web technologies (JavaScript), makes the schemes fully interactive (Figure 3). As a result, users can access the information stored in the underlying database by clicking on any element: gene names direct to mutant phenotypes, locus information and key publications, while lines and arrows lead to the papers that provided evidence for the interactions shown. This chain of information is displayed in a dynamic lateral panel. This panel also contains links to external resources (TAIR10, UniProt and direct links to the publications). All snapshots can be downloaded as PDF files.

DATA CURATION

The curation of the FLOR-ID database is manual by design and was performed by the authors. Manual curation filters the information contained in the database with strong, peer-reviewed experimental data and ensures its quality. Most importantly, all gene and protein interactions shown in the interactive snapshots are supported by experimental data whereas interactions inferred from -omics approaches are mainly predictive.

The flowering community is large and active. In order to leverage its assets, we created a user form allowing to suggest modifications and submit new genes, connections or publications. In a first step, this information will be displayed in a provisional table. In a second step, the database curators will validate the new item. Finally, the update will be incorporated to the public version of the database.

CONCLUSIONS AND PROSPECTS

Information management is a challenge as new biology approaches disclose increasing levels of complexity and regulatory networks. Evolutive databases provide a mean to cope with the increasing amount of data and to gather information from the scientific community. Tools exist to
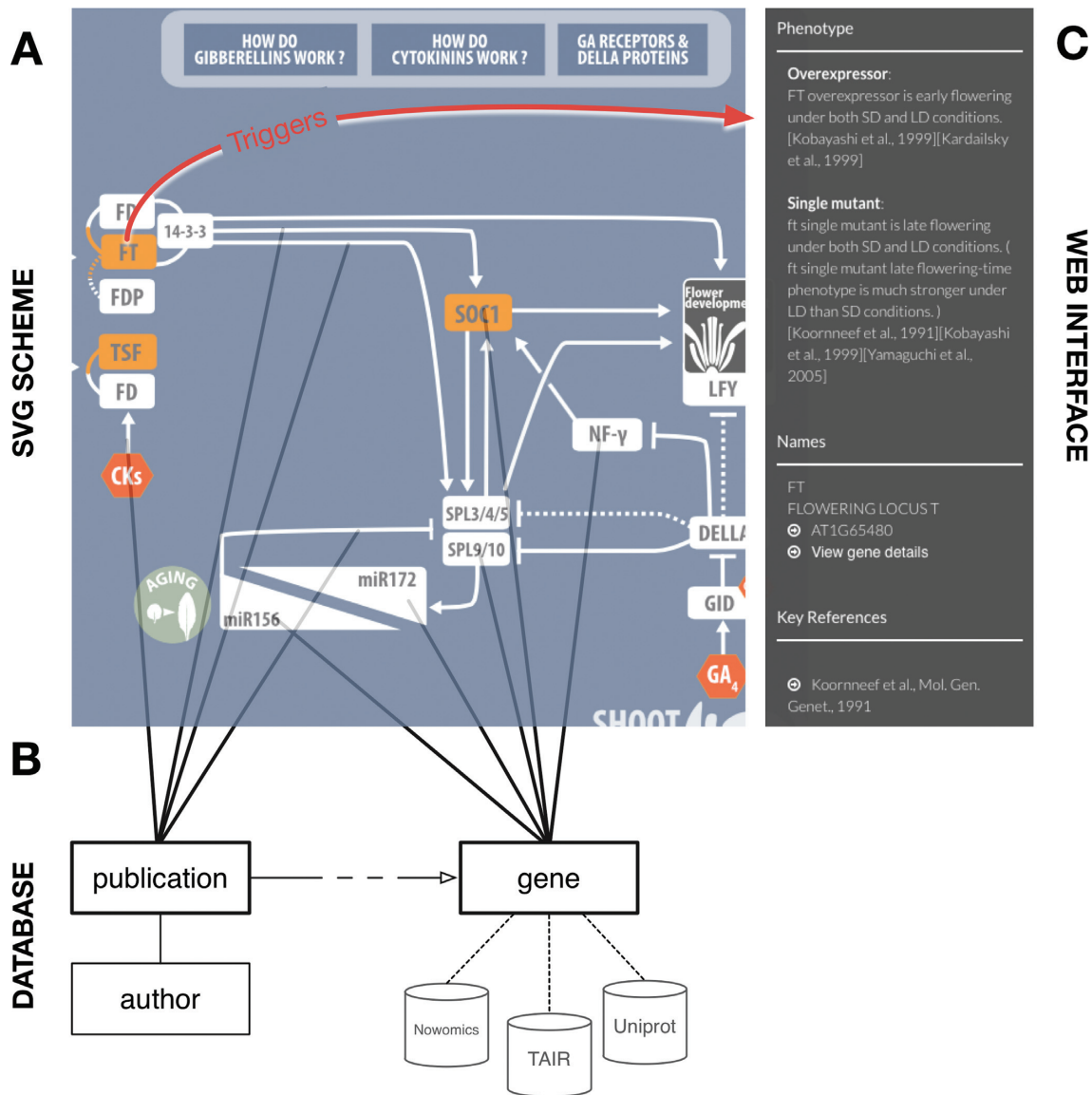
Figure 3. Interactive snapshots of FLOR-ID. (**A**) SVG scheme in which each element (gene, line) is clickable. (**B**) Database connected to the scheme. (**C**) Web panel opened after clicking on an element of the scheme. Information is retrieved from the database and displayed in a human-reading form.

manage raw experimental data sets, but literature is similarly exploding. The core of FLOR-ID is build around the flowering-time gene networks and contains hand-curated data. It therefore offers a reliable tool to define flowering time genes in comparative genomic studies or transcriptomic analyses. Complementary plugins can be created to extend its functionality even further, e.g. for mapping raw gene expression data onto the flowering pathways.

The content of the FLOR-ID database could be easily incremented with modules expanding beyond flowering time, for example to gametophyte development. The curation of the database could then be shared between different laboratories in order to provide expert updates and ensure content accuracy. We thus hope that FLOR-ID will give the impetus to a collective, shared and structuring approach of literature and knowledge recording. Such initiative will logically contribute to community resources like Araport (27).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Bernier,G. and Périlleux,C. (2005) A physiological overview of the genetics of flowering time control. *Plant Biotechnol. J.*, **3**, 3–16.
2. Koornneef,M., Hanhart,C.J. and van der Veen,J.H. (1991) A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Mol. Gen. Genet.*, **229**, 57–66.
3. Martinez-Zapater,J.M., Coupland,G., Dean,C. and Koornneef,M. (1994) The transition to flowering in Arabidopsis. In: Meyerowitz,EM and Somerville,CR (eds). *Arabidopsis*, Cold Spring Harbor Laboratory Press, NY, pp. 403–433.
4. Koornneef,M., Alonso-Blanco,C., Peeters,A.J. and Soppe,W. (1998) Genetic control of flowering time in Arabidopsis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **49**, 345–370.
5. Levy,Y. and Dean,C. (1998) The transition to flowering. *Plant Cell*, **10**, 1973–1990.
6. Reeves,P.H. and Coupland,G. (2001) Analysis of flowering time control in Arabidopsis by comparison of double and triple mutants. *Plant Physiol.*, **126**, 1085–1091.
7. Putterill,J., Laurie,R. and Macknight,R. (2004) It's time to flower: the genetic control of flowering time. *BioEssays News Rev. Mol. Cell. Dev. Biol.*, **26**, 363–373.
8. Blümel,M., Dally,N. and Jung,C. (2015) Flowering time regulation in crops—what did we learn from Arabidopsis? *Curr. Opin. Biotechnol.*, **32**, 121–129.
9. Song,Y.H., Shim,J.S., Kinmonth-Schultz,H.A. and Imaizumi,T. (2015) Photoperiodic flowering: time measurement mechanisms in leaves. *Annu. Rev. Plant Biol.*, **66**, 441–464.
10. Mishra,P. and Panigrahi,K.C. (2015) GIGANTEA – an emerging story. *Plant Genet. Genomics*, **6**, 8.
11. Shim,J.S. and Imaizumi,T. (2015) Circadian clock and photoperiodic response in Arabidopsis: from seasonal flowering to redox homeostasis. *Biochemistry*, **54**, 157–170.
12. Marquardt,S., Boss,P.K., Hadfield,J. and Dean,C. (2006) Additional targets of the Arabidopsis autonomous pathway members, FCA and FY. *J. Exp. Bot.*, **57**, 3379–3386.
13. Berry,S. and Dean,C. (2015) Environmental perception and epigenetic memory: mechanistic insight through *FLC*. *Plant J.*, **83**, 133–148.
14. Wang,J.-W. (2014) Regulation of flowering time by the miR156-mediated age pathway. *J. Exp. Bot.*, **65**, 4723–4730.
15. Verhage,L., Angenent,G.C. and Immink,R.G.H. (2014) Research on floral timing by ambient temperature comes into blossom. *Trends Plant Sci.*, **19**, 583–591.
16. Capovilla,G., Schmid,M. and Posé,D. (2015) Control of flowering by ambient temperature. *J. Exp. Bot.*, **66**, 59–69.
17. Teotia,S. and Tang,G. (2015) To bloom or not to bloom: role of microRNAs in plant flowering. *Mol. Plant*, **8**, 359–377.
18. Wahl,V., Ponnu,J., Schlereth,A., Arrivault,S., Langenecker,T., Franke,A., Feil,R., Lunn,J.E., Stitt,M. and Schmid,M. (2013) Regulation of flowering by trehalose-6-phosphate signaling in *Arabidopsis thaliana*. *Science*, **339**, 704–707.
19. Yu,S., Cao,L., Zhou,C.-M., Zhang,T.-Q., Lian,H., Sun,Y., Wu,J., Huang,J., Wang,G. and Wang,J.-W. (2013) Sugar is an endogenous cue for juvenile-to-adult phase transition in plants. *eLife*, **2**, e00269.
20. Yang,L., Xu,M., Koo,Y., He,J. and Poethig,R.S. (2013) Sugar promotes vegetative phase change in *Arabidopsis thaliana* by repressing the expression of *MIR156A* and *MIR156C*. *eLife*, **2**, e00260.
21. D'Aloia,M., Bonhomme,D., Bouché,F., Tamseddak,K., Ormenese,S., Torti,S., Coupland,G. and Périlleux,C. (2011) Cytokinin promotes flowering of Arabidopsis via transcriptional activation of the *FT* paralogue *TSF*. *Plant J.*, **65**, 972–979.
22. Davis,S.J. (2009) Integrating hormones into the floral-transition pathway of *Arabidopsis thaliana*. *Plant Cell Environ.*, **32**, 1201–1210.
23. Putterill,J. (2001) Flowering in time: genes controlling photoperiodic flowering in Arabidopsis. *Philos. Trans. R. Soc. Lond. Ser. B*, **356**, 1761–1767.
24. Fornara,F., de Montaigu,A. and Coupland,G. (2010) SnapShot: Control of flowering in Arabidopsis. *Cell*, **141**, 550.e2.
25. The UniProt Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
26. Lamesch,P., Berardini,T.Z., Li,D., Swarbreck,D., Wilks,C., Sasidharan,R., Muller,R., Dreher,K., Alexander,D.L., Garcia-Hernandez,M. *et al.* (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.*, **40**, D1202–D1210.
27. Krishnakumar,V., Hanlon,M.R., Contrino,S., Ferlanti,E.S., Karamycheva,S., Kim,M., Rosen,B.D., Cheng,C.-Y., Moreira,W., Mock,S.A. *et al.* (2015) Araport: the Arabidopsis information portal. *Nucleic Acids Res.*, **43**, D1003–D1009.