

The genetics of flowering: a new database

2/2/16

The database developed by the Laboratory of Plant Physiology at the University of Liège is not only an educational tool, but also a valuable resource for researchers, in a field where knowledge is growing exponentially. It is comprised of fully 'clickable' snapshots allowing users to access the underlying information. On the website, they can also find all the references of the scientific articles (currently 1646!) that have contributed to this data, with direct links to the publications.

The Laboratory of Plant Physiology at ULg has been specialising in the study of flowering for several decades. *"Our diet is mostly based on the availability of fruits. For instance, flour, which is obtained from wheat grain. The grain of wheat is a fruit. So if plants don't flower, there's no fruit"*, Professor **Claire Périlleux**, director of the laboratory, reminds us. While flowering is essential to the human diet, controlling this process is just as important. *"If a lettuce flowers, it can't be sold anymore"*, explains the researcher. Owing to its central role in the preservation and evolution of plant species, flowering, which lies at the basis of reproduction, is also key to the stakes of biodiversity.

"Flowering began to be investigated from a genetic point of view in the 1990s, based on the Arabidopsis thaliana model", continues Professor Claire Périlleux. Nicknamed the 'white mouse of plant biology', this plant had its **genome** entirely **sequenced** in 2000. From this moment on, it served as a gateway for many genetic research in flowering. *"A fundamental discovery concerning this species often opens the door to equally fundamental discoveries in other species. For instance, many Asian teams are working on rice. And up until now, the majority of data concerning Arabidopsis thaliana has improved knowledge about the flowering of rice. But this is also true for tomato, potato, etc."*



1646 articles in just a few clicks

Over the years, knowledge in the genetics of flowering has grown exponentially thanks to this key model. *"For people who are starting out in this field, the subject has become very difficult to deal with. Researchers write reviews several times a year. We also use an increasing number of schemes - known as snapshots - to better visualise this mass of data and keep an overview"*, explains Professor Claire Périlleux. Confronted with this information management problem while doing his PhD at the Laboratory of Plant Physiology, **Frédéric Bouché**, who is also passionate about IT and graphics, set up his own database. *"We realised that what he was doing could be very useful, not only for the researchers of the laboratory who didn't necessarily find it easy to manage and represent the literature, but also for the rest of the scientific community"*, continues Claire Périlleux. Frédéric Bouché then met **Guillaume Lobet**, an FNRS postdoctoral researcher, who was also passionate about new technologies. Together they decided to transform this database into a website called the 'Flowering Interactive Database' [FLOR-ID]. *"This database contains approximately 1646 scientific articles. The interactive format allows the information to be easily accessed."*

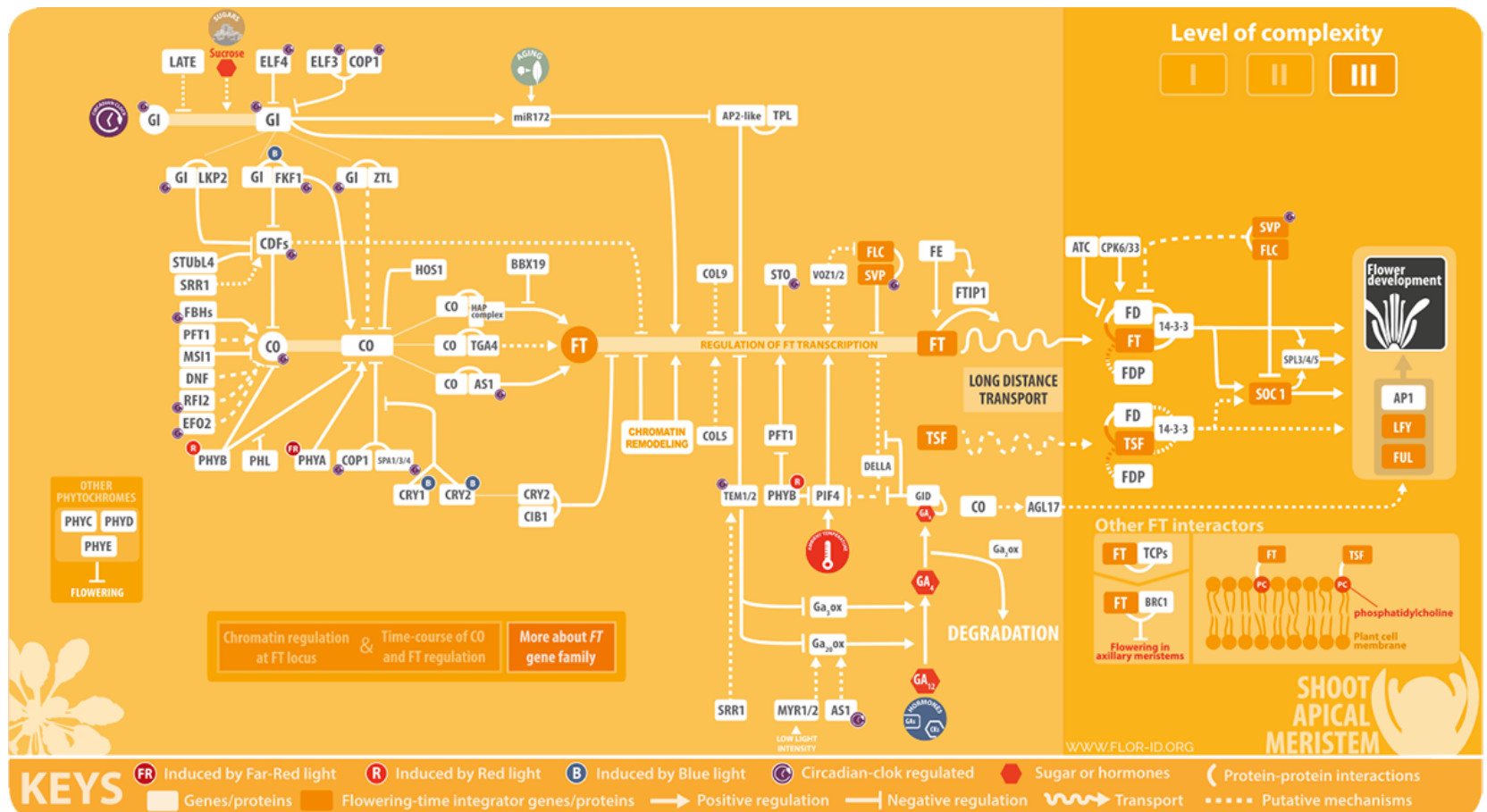
The snapshots are entirely 'clickable' both as regards the listed genes and the lines that show the interactions between these genes. The website also provides all the references of the scientific articles that contribute to this data, with direct links to the publications. *"We make sure that the information is easy to extract, so that researchers can re-use it in their own analyses and presentations. This is a real time-saver"*, Guillaume Lobet points out. *"It's also possible to query the database regarding a specific gene"*, Claire Périlleux adds. These different entries and levels of complexity allow students and senior researchers alike to tailor its use to their needs. *"We offer pre-digested information but this does not mean the reader doesn't need to think"*,

Professor Périlleux underlines. In fact, these shortcuts to knowledge leave more time and availability for the most inventive part of research.

The question of open access

"Making information public is part of our job", Guillaume Lobet considers. "We're not publishers; we're paid by the public authorities and, apart from the fact that if we don't make the information public, we shan't reap the benefits, it is our duty to do it. Furthermore, the raw material is public. We don't divulge anything. It's the intellectual effort of the compilation that we share.

Recently published in the journal *Nucleic Acids Research*, the article (1) relating to FLOR-ID should bring this innovative tool to the scientific community's attention. "It is still very difficult to estimate the value of a free initiative that provides access to information. This review of the literature isn't published in the same way as other reviews of the literature. It's a format that many journals still aren't ready to manage. Therefore, we published this paper in a special issue dedicated to databases. And the article was very well received", Claire Périlleux adds.



In a field where new articles appear every month, it is FLOR-ID's evolving nature that really makes it stand out. Every visitor can submit new information at any time through an online form. The information is checked, then integrated. "This system means that we don't have to produce a new scheme every time. In addition, manual curation - we check every piece of information submitted - ensures far greater reliability", Guillaume Lobet points out.

The fact that an open and evolving database such as this is free, raises questions. *"Within the present context of research, where laboratories run on limited funds, it is essential to make these databases accessible to everyone"*, Frédéric Bouché believes. However, for databases that require considerable curation and the use of powerful servers, a paid-for subscription seems inevitable. Frédéric Bouché gives the example of TAIR (The Arabidopsis Information Resource) database. With an operating budget of USD 1.6 million, an intermediary solution was chosen after the subsidies stopped: paid-for access to the data published in the past year. *"As academics, we don't focus on the economic side. Nevertheless, at a certain stage of development, the issue may arise for FLOR-ID, for instance, if we have to hire someone to keep it up to date. This was also one of Nucleic Acids Research's requirements: that the database be maintained for five years after publication. This is of course what we want, but it will indeed raise other questions"*, the researchers explain.

Finally, the project's continuation implies collaboration with other teams specialising in the genetics of plant development. *"There have been huge progress in the genetics of flowering time but this type of resource could also be created for the genes involved in fertility, pollen formation, etc. Ideally, this database should be similarly expanded because we know that there are other initiatives like this, but laboratories don't necessarily make them public. Making a system user friendly does indeed require an extra effort"*, Professor Claire Périlleux explains. With the appearance of these new tools, the question of sharing knowledge, inside and outside the scientific community, is more than ever crucial.

(1) *FLOR-ID: an interactive database of flowering-time gene networks in Arabidopsis thaliana*, in *Nucleic Acids Research*.