

CropPedia - The intelligent linking pin between big data and new crops

Roeland van Ham, Rudi van Bavel & Anker Sørensen

How do companies and research organizations with the ambition to develop new or improved crops find their way in the ever-increasing amount of available crop data? The answer is CropPedia: KeyGene's intelligent big data tool. CropPedia combines, searches, compares and visualizes data in an extremely clever and quick way.

Enormous amounts of data

It is not always easy for crop improvers to see the forest for the trees. There is an ever-increasing amount of crop data available, but finding a way through it all is another matter. Take, for instance, the genome of rice, which comprises 430 million bases. Research organizations have already determined the full genome sequence of 3000 rice lines; in total 1.290 billion bases. Another example: there are potato-improving companies that monitor some 100,000 potato plants for 50 traits each year: size, shape, number of fruit, resistance to pathogens and drought, etc. Their next line of interest is the differences in DNA sequence between the best performing plants and successful crossbreeds. To combine such data, you do not only need databases that can intercommunicate, but also tools to select, process and visualize the data, - preferably without long processing times. If companies want to increase the chance to stay ahead of their competitors with respect to developing market leading varieties, they need to be able to use big data for crop improvement. CropPedia is the intelligent big data tool that enables you to do so.

The power of CropPedia

There are more platforms that unlock huge amounts of data to researchers and crop improvers, but CropPedia has three unparalleled qualities.

Firstly, CropPedia has been developed on the basis of more than 25 years of experience and collaboration within KeyGene in the field of molecular crop improvement. KeyGene's know-how in the areas of DNA, molecular markers and plant traits – to name just a few – combined with its full awareness of the challenges faced by crop improvement companies, has been an important cornerstone for the successful development of CropPedia's functionalities.

A second distinct aspect is the fact that CropPedia allows connection between and integration of different data. All the relevant data, both a company's proprietary and public datasets, is available and can be searched in an extremely clever way in a secured environment. CropPedia can help ascertain, for instance, as to which genes are involved in desired traits. Molecular biologists can exploratively search for relationships between DNA markers and genes, on the one hand, and plant traits like drought and fungal resistance, on the other hand. The need to switch between Excel sheets, breeding records, publicly accessible genome or marker charts will be a thing of the past.

A third quality that makes CropPedia stand out from other existing big data tools is its speed. Its use of very modern storage and search technologies brings an end to the minutes long waiting for the requested information to appear on the computer screen. The program comes with an unlimited amount of searchable data – without any need of a super computer. The program can even run on a notebook, thanks to new calculation modules and network solutions for data storage and transport developed by KeyGene.

Smart research tool

The possibilities for smartly searching all the data in CropPedia, blasting, comparing of data of different genomes and examining gene functionalities are important features of CropPedia for crop researchers. CropPedia includes so-called "gene passports"^(1,2,3,4), which present for a selected gene, the summary of all the functional and structural annotations. This provides information about the gene's function, appearance and location within the cell. The CropPedia Variomics viewer is another special attribute. It displays in one view, all structural variants detected between re-sequenced germplasm accessions within a short amount of time. Based on annotations of the structural variants sub-selections can be made in seconds.

Joint workspace

CropPedia has been developed with a very intuitive and visually attractive user interface. CropPedia promotes cooperation

About the authors



Prof. Dr. Roeland van Ham is Vice president Bioinformatics and Modeling at KeyGene and professor Plant Computational Biotechnology at TU Delft.



Rudi van Bavel, BSc. is bioinformatician at KeyGene and head of the development of CropPedia



Anker Sørensen MSc. is responsible for New Business at KeyGene and has a strong background in plant breeding

About KeyGene



KeyGene – The crop innovation company

KeyGene is the go-to AgBiotech company for higher crop yield & quality. With our intellectual capital, solution driven approach and collaborative spirit, we work for the future of global agriculture with partners in the AgriFood sector. www.keygene.com

About big data!



CropPedia Variomics Viewer

between colleagues: it features joint workspaces where company staff in different departments or even at remote locations can work together on the same topic and research question. Data manipulations and outcomes can be stored in the workspace for others to comment. Besides this CropPedia features an authentication module, enabling companies to determine which staff members have access to which information. Moreover, CropPedia can easily be integrated in existing pipelines and systems. If the hardware is in place, the system can be used as soon as it has been installed. All the data are already there.

Unexpected insights

The CropPedia program allows for hypothesis-free examinations that may very well lead to unexpected insights that researchers would not acquire if they only stick to a single gene or trait. This opportunity to conduct explorative research will be further expanded in the near future. CropPedia's new trump card will be the program's ability to also make suggestions after the initial search is performed, such as: 'cucumber also has this resistance gene', or 'for this potato, a blue-color producing gene can be found in the same DNA region'.

Essential to new generation crop improvers

The enormous amount of – genetic, genomic, proteomic and phenotypical – crop data that is, and will be, available offers huge opportunities for both scientists and crop improvement companies. To discover the basis of new traits, make connections with other crops and search for that one specific causal gene. When it comes to developing improved varieties, CropPedia is a very effective tool, the linking pin in genome research, which helps make a difference and stand out.

Identifying yield genes in rice

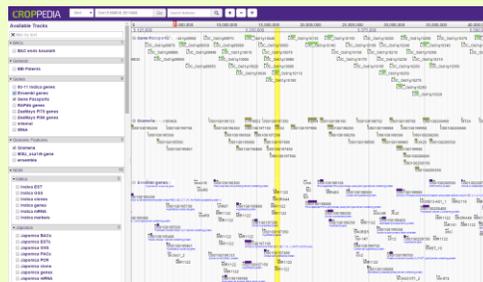
A CropPedia showcase

What we knew:

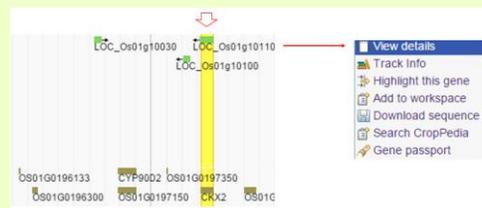
1. the number of grains per panicle is one of the yield component traits in rice (Xing & Zhang⁵)
2. a mapped QTL called Grain number 1a (Gn1a) flanked by two (RFLP) markers: R3192 and C12072S (Ashikari et al.⁶).

The **Key word search** was deployed to find the position of the QTL flanking markers. The sequence of both RFLP markers was **downloaded and blasted** on the IRGSP-1.0 whole genome sequence (Oryza sativa Japonica Group cv. Nipponbare). The blast result showed a 404.5 kb region on chr. 1: 5109949-5514454

The **CropPedia genome browser** was used to zoom in on the 404.5 kb region followed by the **activation of the gene model/annotation track** "Gene passports" and additional tracks such as "Ensemble genes". This resulted in an overview, covering all underlying genes for that region:



The genes were **inspected for their functional annotation** by examining the Gene Passports. This lead to a list of a few dozen functional candidate genes, one of them being the CKX2 gene (cytokinin-oxidase 2):



Ashikari et al.⁽⁶⁾ identified CKX2 as the causal gene and described that the CKX2 allele in the high yielding rice variety 5150 carried a 11-bp deletion that leads to a premature stop codon.

We used the **CropPedia Variomics viewer** to obtain in-depth insight in the CKX2 allelic variation in multiple rice lines. In this way it became directly clear which of the rice lines had similar or different alleles compared to rice variety 5150. Lines with a similar allele may also be used to improve the number of grains per panicle. Additional resequencing of germplasm enables to inspect the variation of this and other loci to an even greater extent in CropPedia. At the same time breeding germplasm can be easily examined for additional variation in the causal gene.

References

1. Blum, T., Briesemeister, S., and Kohlbacher, O. (2009). MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction. *BMC Bioinformatics* 10, 274.
2. Pérez-Rodríguez, P., Riaño-Pachón, D.M., Corréa, L.G.G., Rensing, S.A., Kersten, B., and Mueller-Roeber, B. (2010). *PlnTFDB: updated content and new features of the plant transcription factor database*. *Nucleic Acids Res.* 38, D822–D827.
3. Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., et al. (2009). *InterPro: the integrative protein signature database*. *Nucleic Acids Res.* 37, D211–D215
4. Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). *Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments*. *Genome Biol.* 9, R7
5. Xing, Y. and Zhang, Q. (2010) *Genetic and molecular bases of rice yield*. *Annu Rev Plant Biol.* 61:421-42. doi: 10.1146/annurev-arplant-042809-112209
6. Motoyuki Ashikari, Hitoshi Sakakibara, Shaoyang Lin, Toshio Yamamoto, Tomonori Takashi, Asuka Nishimura, Enrique R. Angeles, Qian Qian, Hidemi Kitano, Makoto Matsuoka (2005). *Cytokinin Oxidase Regulates Rice Grain Production*. *Science*, vol 309, Issue 5735, pp. 741-745 Jaggard, K.W., et al. (2010). *Possible changes to arable crop yields by 2050*. *Philos. Trans. R. Soc. Lond. B Biol.* 365, 2835-2851

Contact details

KeyGene
Alexander Wittenberg

+31 (0)317 466866
Alexander.wittenberg@keygene.com

Request a free trial on:
www.croppedia.com