Impact of **new technologies** on breeding of field crops

Harold Verstegen, Viktor Korzun, Milena Ouzunova and Leon Broers

Many of the characteristics of field crops are inherited in a quantitative manner; others are more of a qualitative nature. Broad genetic variation combined with highly efficient breeding methods is the prerequisite for continuous breeding progress and developing improved varieties.

Dr. Harold Verstegen is Innovation Manager, Dr. Viktor Korzun is Head of Cereals Biotechnology, Dr. Milena Ouzunova is Head Biotechnology Maize/ Oilseed Rape and Dr. Leon Broers is Member of the Executive Board, responsible for Breeding & Research and Energy Crops at KWS SAAT AG, Einbeck, Germany

In breeding field crops, the main targets are performance – yield and yield stability – as well as a number of agronomic traits like plant height or earliness, resistance to biotic and abiotic stress factors, and a whole range of quality traits – depending on the crop of interest. The use of molecular technologies can improve breeding processes by increasing the heritability and selection efficiency, accelerating breeding cycles or saving costs. Moreover, genetic variation can be enhanced through inducing and identifying new favourable alleles. This has already shown great benefits in various crops.

The latest approaches focus on sequencing technologies combined with bio-informatics

For 25 years KeyGene has been a pioneer in developing new molecular tools and technologies which have repeatedly proven their potential in the professional plant breeding industry. Throughout the years, KeyGene has used collaborations with academia and industrial partners in vegetables and field crops to assess and optimize their technologies already in an early phase of development. KWS – a global seed company in field crops with major breeding programs in corn, sugar beet and cereals as well as potatoes, oilseed rape, sunflower and sorghum – has had the opportunity to collaborate with KeyGene for more than 15 years using its services and participate in technology development programs. Some of the common achievements are briefly presented here.

Male fertility restoration

Hybrid rye breeding and seed production employ an efficient hybridisation system based on cytoplasmic male sterility (Pampa-CMS) and complete restoration of pollen fertility through restorer genes in the commercial seed. However, under unfavourable weather conditions, the amount of pollen is insufficient to allow complete pollination of all female florets. As a consequence, there is an increased risk for ergot infection. Since ergot constitutes a health risk to humans and animals, KWS focused on developing breeding material to reduce the risk of infection. The strategy was to breed for increased pollen shedding ability.

A source for strong restoration was found in IRAN IX, a primitive Iranian rye population, which was then used as a donor for a new restorer gene. Unfortunately, the IRAN IX restorer gene came with many undesirable agronomic characters. To be able to use the restorer gene in the breeding program this linkage drag had to be reduced. To remove unfavourable DNA fragments linked to the restorer gene a fruitful collaboration with KeyGene was initiated. Tightly linked AFLP markers to the restorer gene have been found and specific PCR-based assays were developed. With these DNA assays, it was possible to reduce linkage drag and integrate the IRAN IX restorer efficiently into the breeding process and consequently release the first PollenPlus® hybrid rye varieties. (www.kws-getreide.de/hybridroggen.html).
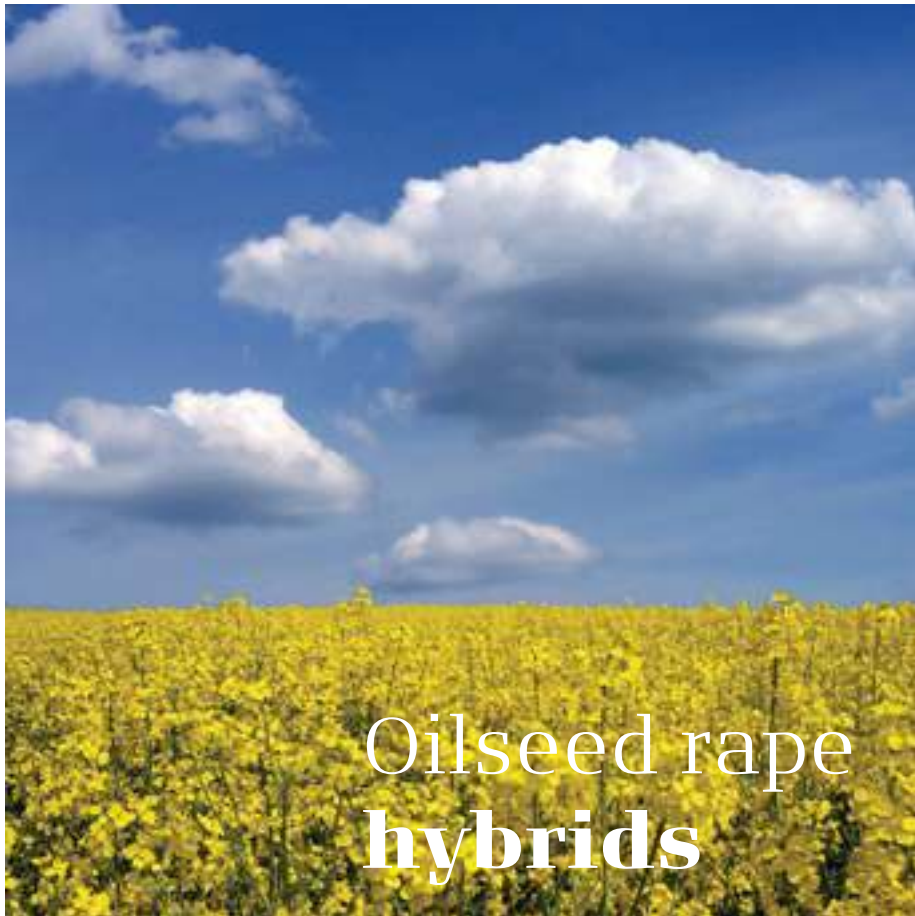
CRoPS® technology

Several years ago no public data on genetic variation was available for Sorghum bicolor. SNPs (Single Nucleotide Polymorphism) – a class of putative markers – are a way of revealing genetic variation in an efficient way. KWS investigated which technology was available to cost-effectively develop markers to be used in the young KWS sorghum breeding program. With the support of KeyGene's CRoPS® SNP discovery technology, several hundreds of validated SNP assays were developed within several months. A result that has been immediately utilized in the sorghum breeding program to study genetic diversity and develop marker assisted selection applications. The CRoPS® technology is a good example of how modern molecular technologies, combined with bio-informatic capabilities can support modern breeding approaches very efficiently even in smaller crops with limited resources.

KeyGene as pathfinder

In plant DNA, polymorphisms – both SNPs and structural variations – can be detected by a variety of flexible and High-Throughput (HTP) genotyping technologies. These technologies differ from each other with respect to costs and sample throughput as well as the number of polymorphisms to validate. The latest approaches increasingly focus on sequenc-

Oilseed rape
hybrids

Bart Lambert

Bayer CropScience has been using various KeyGene technologies for more than 20 years, primarily in support of its oilseed rape R&D activities. It started in 1993 with a license for the AFLP® genotyping technology. AFLP® was initially used for accelerated backcrossing of the SeedLink hybridization technology components in various parental lines of InVigor canola (Brassica napus) hybrids enabling the speeding up of hybrid development.

The technology was also used to assess genetic relationships between various germplasm pools used in canola breeding and for identity control of experimental doubled haploid lines. Later on, with the advent of QTL mapping, AFLP® technology played a key role in the discovery of loci for various native traits, such as flowering time, disease resistance, and fatty acid composition. Building further on this, AFLP® and, later on, whole-genome profiling enabled Bayer CropScience to construct a high-resolution genetic and physical map of B. napus, which formed the basis of the successful sequencing of its genome in 2009, for which KeyGene provided 454 sequencing support. AFLP has now been replaced with SNP-genotyping technology for genome typing. In fact, AFLP® was the first SNP-genotyping technology as it also targets single-nucleotide changes in genomes. More recently, Bayer CropScience licensed-in dwarf alleles for canola discovered by KeyGene with its KeyPoint technology platform. KeyGene continues to be an important partner for Bayer CropScience as it succeeds in developing new, more powerful technologies for trait discovery as well as for speeding up the R&D process, with the focus now shifting to wheat.

Dr. Bart Lambert is Research Manager Oilseeds, Bayer CropScience, Diegem, Belgium

ing technologies combined with bioinformatics. The advantage of sequencing technologies is to combine marker development and validation in one single step enabling the potential to capture more types of variations present in the DNA. This increases the chance of finding associations between genomic variation and phenotypic variation.

A thorough and accurate evaluation of new technologies is necessary to choose the right approaches and combination of technologies. This is not trivial and can save significant capital investments and internal research resources when done properly. KWS has participated in these technology development programs coordinated by KeyGene by being a consortium member.

Enhancement of genetic diversity

Mutation breeding has been known for a very long time. Mutations are often induced applying either chemicals (e.g. EMS, DMS) or radiation (e.g. x- or γ-rays). The main issue is finding the mutant of interest in the mutant population. KeyGene has developed a smart screening technology (KeyPoint® MB) to select very efficiently specific point mutations. KeyPoint®MB has been used to develop various mutant populations in different crops and showed significantly higher efficiency in finding the mutation in genes of interest. This has allowed KWS maize breeding programs to very quickly identify allelic series comprising numerous mutations. Those mutants in the genes of interest are then directly used in a backcrossing approach. KWS is aiming to utilize this technology also for various traits in several other breeding programs.

Conclusion

The impact of new molecular tools and technologies available for plant variety development has proven to be essential to optimize and accelerate breeding programs. Among others, KeyGene has constantly contributed to these developments by offering highly competitive up-to-date technologies for screening genomic variation where available and creating new variation where applicable.

Plant biology enters **Big Data** era

Roeland van Ham and Marcel Reinders

Over the last two decades the role of plant genomics has grown considerably for improving plants in terms of yield, resistance, and nutritional value. Already from the beginning it was clear that information technology and bioinformatics were going to be crucial in this development, from help in storing large genomics datasets to enabling the comparison of DNA sequences to learn more about the function of genes.

In 1989, the year in which KeyGene was founded, a small piece of bioinformatics software called 'BLAST' was being developed at the US National Center for Biotechnology Information. BLAST turned out to be the 'Google' for biological sequences: it can do very fast and highly accurate searches for sequences in a database that match a 'query' DNA sequence. When BLAST was released in 1990, the public sequence database Genbank held some 33,000 sequence records and 40 million nucleotide bases. The ability to compare sequences opened new doors to rationally improve plants and propelled the importance of plant genomics.

Astonishing pace

Although BLAST has changed little over time, our ability to generate DNA data has increased at an astonishing pace, due to revolutionary developments in sequencing technology. Nowadays KeyGene's sequencing facility alone can produce a thousand times the content of the 1989 Genbank database in raw nucleotides, in just a matter of hours, and around the world, laboratories are now producing petabases (1015) of new sequence each year. The Genbank database has grown to over 700 billion carefully curated nucleotides. A significant portion of these data is coming from plants, mostly crops, and it comprises hundreds of whole genome sequences. This gives a wealth of information to refine our understanding of the function of DNA sequences.

Similar revolutions happened in measuring other types of molecular data. It started with measuring RNA expression, first still on microarray technology but now also done by sequencing technology. Throughout the years it became possible to measure a bewildering variety of data types on a similar large scale: not only other molecules such as proteins and metabolites, but also interactions like protein-protein interactions, methylations, protein-DNA bindings, and even phenotypically, like KeyGene's PhenoFab where we image every plant. These different data types revealed that the function of a sequence was much more complex than just that of similar sequences in a database: it is under the control of a wide variety of mechanisms. These influences need to be understood before one can rationally interfere with them to improve a plants behaviour.

Hence, nowadays in a trait development programme for a field-grown crop, data are collected on the plant's genome and physical status, it's gene expression and metabolites, and the environmental conditions: weather, soil composition and pests, and all this at many time points over the course of the experiment. The genetics behind traits is however often highly complex, with contributions of a multitude of genes and environmental factors. Revealing the culprits and the often small impact of each of these requires large datasets to ensure statistical rigor. But how do we manage all those data and derive meaningful insights and applications from them? Taken together, with these developments, plant biology enters the 'Big Data' era.

Big Data

Fortunately, similar developments as in the genomics field happened in many other areas of science and society. No matter if it concerns geographic, physics, climate, or societal data, most modern technologies can produce a wide variety of data in extremely high-throughput and massively parallel ways. Collectively this is denoted as 'Big Data', whose key attributes are often indicated with the three V's of Volume, Velocity and Variety. These tell us that Big Data come to our computers in large quantities, at high speed and in great diversity. Big Data may result from own experiments but can also come to a researcher

Dr. Roeland C.H.J. van Ham is Vice President Bioinformatics and Modeling at KeyGene, Wageningen, the Netherlands

Prof. Dr. Marcel J.T. Reinders is Professor in Bioinformatics at Technical University Delft, the Netherlands

Our ability to generate DNA data has increased at an astonishing pace

from elsewhere; in vast packages produced by large research institutes, or in smaller but diverse packages from many distributed sources around the world. Consequently, datasets are often 'too big to move'. Computational tools are essential for data handling and analysis, and statistics and mathematics creep into all research fields.

What does the Big Data landscape look like in plant biology and how is it evolving? First, there are a large number of conventional databases around the world that collect and structure data on plant genomes, genetics, proteins and metabolites. Not only their number and data contents makes these data big, it is the fact that the data are distributed, heterogeneous and often duplicated, which makes them complex to use and explore in an efficient manner. Public initiatives such as iPlant aim to develop a 'cyber infrastructure' that pulls these distributed databases together and provide computing resources and software to work efficiently on analysis and mining of the data. Where transfer of large datasets becomes the bottleneck, cloud computing may provide an infrastructure in which analysis tools and computation are moved to the data, instead of moving data to those resources. Second, there is an accelerating stream of plant genomics and phenotypic data produced in-house in research laboratories of companies and academia. For its interpretation it is often essential to connect this data to large public datasets, which sets require-

ments to data formats. KeyGene has developed a powerful solution called <Crop>Pedia, that helps in integrating such privately generated datasets, together with data available from the public domain.

Hunt for patterns

While the problems involved in volume and velocity of Big Data, which present the more immediate challenge to conventional IT infrastructures, can increasingly be dealt with through scaling and novel technologies for data storage, transport, security and parallelized computation, data variety is more a matter of biological and mathematical science. It is about understanding and revealing the underlying structures and relationships, because these are what lead to new knowledge and applications. The same way as Facebook exploits Big Data in the hunt for patterns in consumer behaviour that can be translated to lucrative information and personalized advertisement, biologists hunt genomic variation data, gene expression and protein interaction data for mechanisms behind important plant traits. This requires new and clever approaches, often coming from statistics and mathematics, to structure the data as to derive mechanistic knowledge. The interplay with plant biologist is however crucial to successfully implement iterations in this DIKW ('Data Information Knowledge Wisdom') process.

The success of exploiting the digital information enclosed in Big Data in plant sciences in general but certainly also in crop research, will depend on innovations in IT technology, easy access to large-scale computing, and novel statistical and mathematical methods to integrate and model diverse biological data. But perhaps equally important will be to educate and strengthen the computational skills of current and next generation of crop researchers.

KeyGene, from its start, has always been at the forefront of technologies for generating and interpreting large-scale datasets, both in application and development of the required tools. All the ingredients necessary to successfully translate genomic discoveries into breeding applications are deeply rooted in KeyGene's DNA and we can therefore look with confidence to its future at the crossroads of Big Data and crop research.