



www.gena.ucl.ac.be



"The sturdiness and statistical power of SAS tools, combined with their open systems design, has enabled us to build a comprehensive data warehousing system for our fundamental and applied research in a very short period of time"

Dr. Xavier Draye

Faculty of bio-engineering,  
agronomy and environment, UCL

*The Power to Know™*

## UCL

### Sequencing the genomes of the world's leading crops with SAS

*The laboratory of crop physiology and plant improvement at UCL's Faculty of bio-engineering, agronomy and environment (Louvain-la-Neuve) is successfully using the SAS system in its genetic research work – both fundamental (e.g. to gain a better understanding of the root system formation in certain plants) and applied (improving the characteristics and performance of specific food plants).*

#### Sequencing the genomes of the world's leading crops

UCL's Faculty of bio-engineering, agronomy and environment at Louvain-la-Neuve carries out genetic research on populations and uses quantitative genetics to gain a better understanding of the consequences of selective breeding and the evolutionary effects of natural selection, in order to examine the basic prerequisites for improvements to animal and vegetable species.

Many of the world's food and feed crops are members of the Poaceae family, which includes rice, wheat, maize, sorghum, sugarcane, barley, oat, rye, millet, and others.

Sorghum (*Sorghum bicolor* L. Moench.) is a suitable species for positional cloning because of its relatively small genome size. The small genome of sorghum provides an important template for the study of closely-related large-genome crops such as maize and sugar cane. Moreover, sorghum is a leading cereal in arid and semi-arid agriculture, ranking fifth in importance among the world's grain crops.

Dr. Xavier Draye, a researcher at the UCL's laboratory of crop physiology and plant improvement, confirmed that genetic research on sorghum and other plants should allow improvements to crop qualities and yields, while reduc-

ing the environmental impact of growing these crops (reduced fertilizer residues in the soil).

"For example, we try to gather detailed knowledge of the root system size and development of the plant," he stated. "We look in the genome to try and find the genes responsible for a specific growth or form of the roots that makes them more adapted to certain soils and environments. This involves two types of analysis, a comparison of the root system in the soil and an analysis of the plant's genome, plus comparisons between the two. Our goal is to exploit the genetic variations among individuals."

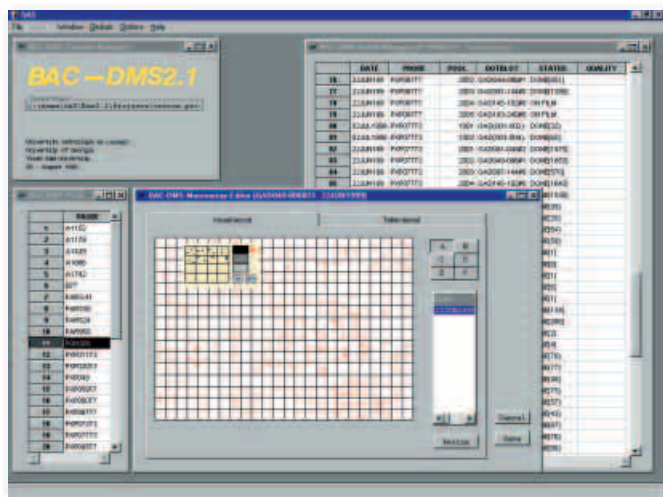
#### A genetic detective's work

During the last years, direct sequencing of the genomes (about 300 millions pairs of bases for rice and billions for maize) has become routine. The struggling work now is to annotate the DNA sequence, viz. to find the genes (each a certain sequence of pairs of bases) and assign their functions. One promising strategy is to look at the association between the presence of certain sequences (also known as DNA markers) in the genome and desired characteristics of the roots.

More than two decades of genomics research have yielded high-density maps of DNA markers for many major

#### The challenge

Develop appropriate database systems with sophisticated query facilities to capture large amounts of data quickly and produce intelligible results efficiently from genetic research on plants. Develop a software system to facilitate the archiving of data from screening macroarrays of ordered DNA clones, and allow the navigation of these data to detect errors and uncertainties.



**SAS software was chosen for the complex statistical work involved in mapping DNA markers for genome sequencing.**

crops. These maps provide a valuable foundation for basic research in genome organization and evolution. One of the essentials for the success of map-based cloning and physical analysis of large chromosomal regions is the availability of libraries containing large inserts of genomic DNA. These libraries have made valuable contributions to the production of physical maps of large regions and to the isolation of many important genes. A large insert Sorghum propinquum BAC (Bacterial Artificial Chromosomes) library has been constructed at Texas A&M University to analyze the physical organization of the sorghum genome and facilitate positional cloning of sorghum genes.

#### The need for a data warehouse

"We are speaking of really high numbers here," said Dr. Xavier Draye. "Just to give you an idea, with 1,000 to 10,000 markers, we try to map 30,000 to 150,000 clones. We use grids to physically spot the occurrence of certain clones, with more than 18,000 clones on a single grid. As grids can be

re-used for examining (but get old and more imprecise to read), as batches tend to overlap and research people also overlap over time, we needed a monitoring system and a robust data warehouse, built upon some 10 different databases."

The system had to enable easy, accurate identification of the grids, while guiding the user through the various steps intuitively.

SAS software was chosen for this complex statistical work. "SAS AF enabled us to develop the BAC-DMS application very rapidly – it took less than 3 months working part-time. SAS has a very open systems design and allows easy links. What's more, SAS is a highly developed language and a very robust and comprehensive statistical analysis tool, which means that an experienced SAS user can create a new application in less than half an hour. The fact that everything is done in one single environment also accelerates the work," added Dr. Xavier Draye.

The program now provides a versatile "click-a-clone" data entry graphical interface, a data warehouse including a sophisticated exploration tool, a routine to export groups of overlapping clones for interactive clone ordering, and a monitoring system for the practical management of a large number of reusable macroarrays.

"The ability to use BAC-DMS does not depend on any SAS skills," Dr. Xavier Draye concluded. "However, all native SAS windows can be accessed during a BAC-DMS session, allowing almost any database transaction to be performed by a knowledgeable SAS user."

#### The solution

Implement a data entry graphical interface and a data warehouse system with a sophisticated exploration tool in the application development environment of the SAS software, to achieve a significant reduction in the time required for experimental setup and data entry, and in the risk of data entry mistakes.



SAS Institute  
Kasteel de Robiano  
Hertenbergstraat 6  
B-3080 Tervuren  
Tel.: +32 (0)2 766 07 00  
Fax: +32 (0)2 766 07 77

SAS Institute s.à.r.l.  
Office City  
6 Circuit de la Foire Internationale  
BP 2507  
L-1025 LUXEMBOURG  
Tel.: 00-352-264.20.410  
Fax: 00-352-264.20.608

[www.sas.com](http://www.sas.com)