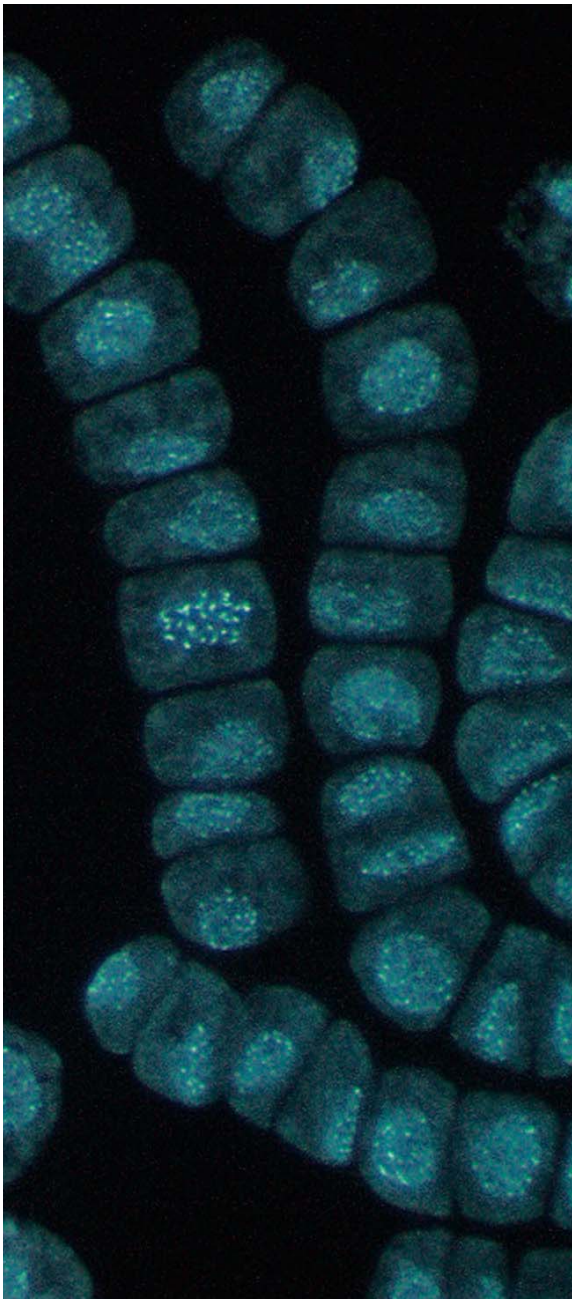


Unlocking the secrets of the banana genome

Will genomics, the science emerging around the sequencing and deciphering of the functions of genes, change the way we grow food?



Genomics should help understand the processes that underlie the problems affecting bananas and, as a result, improve the lives of smallholders.
G. Orjeda, INIBAP

If all that mattered were the number of genes, we wouldn't be able to explain why humans, and not bananas, are the species who solved the genetic code, sequenced their own genome and are busy doing the same for other species. Humans and bananas both have about 30 000 genes. The challenge is to understand the similarities—and what makes us so different.

As genes are identified and functions are assigned to them—the goal of genomics—we shall come to understand how banana genes interact with each other and the environment to make a plant thrive or die. The hope is that this knowledge will translate into sustainably higher yields, especially for the smallholders who grow over 85% of the bananas harvested in the world.

*Breeders have largely confined their search for genes to *Musa acuminata* and *Musa balbisiana*, but other wild species might harbour agronomically-interesting characteristics, such as mechanisms for tolerance to cold, water-logging or drought. From left, *Musa campestris*, *Musa borneensis* and *Musa velutina*.*
M. Häkkinen



This paper is based on a talk given by Pat Heslop-Harrison from the University of Leicester at a symposium on the Conservation and Use of Musa Biodiversity for Improving Livelihoods held in Leuven, Belgium, on 18 October 2005.

One compelling motivation is provided by the fact that most of the existing varieties haven't changed substantially since the time they were selected from the wild and nurtured by farmers over thousands years, whereas the conditions—economic and agronomic—under which they are cultivated have changed spectacularly. Bananas are beset by a host of pests and diseases while the loss of agricultural land to urbanization and unsustainable agricultural practices is forcing many farmers to grow bananas in less than optimal settings. What smallholders need are bananas that can overcome these challenges and still possess the qualities that growers, traders and consumers have come to expect. This is a tall order, given the difficulty of crossing varieties which, for all practical purposes, are sterile (see *Hungry for improvement* in this report).

The same constraints that make breeding bananas a long and laborious process also limit the application of classical genetics to understanding the processes that underlie the problems afflicting bananas. Banana genomics is still in its infancy and its agricultural benefits are still largely theoretical. Like other branches of banana research, it remains relatively under-resourced, but a networking approach is helping researchers make the most of what they do have and this approach has certainly accelerated gene discovery. Since 2001, most of the genomics work on bananas is being done by scientists who are members of the Global *Musa* Genomics Consortium for which INIBAP provides the secretariat. What they are finding has implications not only for agriculture, but also for conservation and fundamental research.

A FIRST GLIMPSE INTO THE GENOME

Since the vast majority of domesticated varieties of banana are related to one or two wild species, scientists have not one but two genomes to look at; *Musa acuminata* donated the so-called A genome and *Musa balbisiana* the B genome. As a prelude to

sequencing the banana genome, the DNA of *acuminata* and *balbisiana* plants was duplicated and broken into smaller pieces that are easier to handle. These DNA fragments were inserted into bacteria for safe keeping as 'bacterial artificial chromosomes' (BACs). Consortium members have developed five BAC libraries: one of *balbisiana* and four of *acuminata*. The BAC libraries are freely available to the research community through the *Musa* Genome Resources Centre, hosted by the Institute of Experimental Botany in the Czech Republic (see *A resourceful centre* in this report).

A first glimpse into the organization of the A genome was provided when the Laboratory of Gene Technology at the *Katholieke Universiteit Leuven* (KULeuven), in Belgium, sequenced two BAC clones from one of the *acuminata* BAC libraries (the wild diploid Calcutta 4). The gene density was, on average, one for every 8700 bases (8.7 kb). Less than 50% of the DNA in these clones was coding for genes. Like most plant genomes, the banana genome seems to comprise gene-rich areas that are separated by long stretches of repetitive sequences. And as scientists have known for some time, the B genome—and maybe the A genome as well—contains viral DNA from the *Banana streak virus*, which, in the course of the evolution of *balbisiana*, has found a niche in the banana's genome (see *The accidental pathogen* in the 2004 INIBAP annual report).

BAC clones have also been used to compare banana sequences with those of rice, a monocotyledon like banana, and *Arabidopsis*, a dicotyledon. By virtue of being sequenced, the genomes of these plants have become the yardstick against which genomes-in-waiting are compared, but their current prominence may diminish as other genomes yield up their secrets. A first round of comparison done at The Institute for Genomic Research (TIGR), in the United States, found less similarity with rice and more correspondence with *Arabidopsis* than anticipated. Banana's evolutionary position, relatively far from the major cereal crops and with similarities to various dicotyledons, may help to make it a useful point of comparison. (see *The comparative advantage of bananas*)

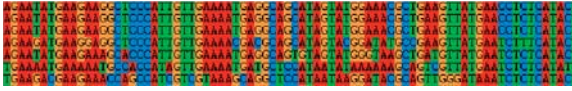
MINING THE GENOME

To obtain the complete sequence of the banana genome, a representative genotype will be chosen and each clone in its BAC library will be sequenced. A computer programme will then match the overlapping ends to reassemble the entire sequence



Calcutta 4 is the common name given to the *Musa acuminata* sub-species *burmannicoides*.
R. Markham





from the tens of thousands of clones in the BAC library. Identifying the genes buried in the litany of As, Cs, Gs and Ts that represent the four DNA bases of the genetic code is only a beginning. The next task is to find out their function.

Indeed, long before the entire genome sequence is available, there are other ways to search for useful genes. Since nature didn't start from scratch with each species, the genome can be mined for genes known to exist in other species, whereas ESTs (expressed sequence tags) and complementary DNA (cDNA) provide a short cut to discover commonly expressed genes and genes that are expressed in specific conditions, such as disease, infection or drought. The procedure starts with the harvest of messenger RNA (mRNA)—the molecule that carries the information encoded on the DNA to be trans-



What farmers need are cultivars that can grow in extreme conditions.
INIBAP,
J. Daniells



lated into proteins. The presence of mRNA in a cell indicates that the genes for which they act as intermediary are active. But because of the transient nature of the RNA molecule, scientists transcribe it into the more stable cDNA. The sequence of a certain number of bases at either or both ends of the cDNA is an EST. It need only be long enough to act as a tag for the gene encoded in the cDNA.

The method is popular with scientists trying to isolate agronomically important genes, such as those involved in a plant's response to biotic or abiotic stresses. In these instances, the mRNA is collected from a plant submitted to the stress in question—be it a pathogen, extremes of temperature or drought—or from a normal or developing plant, when the goal is to zero in on genes involved in its day-to-day functioning or development.

A resourceful centre

Whenever possible, the Global *Musa* Genomics Consortium is committed to placing the products its members have developed in the public domain. To do this it has set up the *Musa* Genome Resources Centre as its distribution arm. The primary aim of the Centre is to support the research activities of the Consortium by making available *Musa* genome resources to its members and by developing new resources. Under some circumstances, the Centre also distributes its resources to individuals and research organizations outside the Consortium. The main limitation is that users have to sign an agreement ensuring that the resources remain in the public domain.

Five BAC libraries are currently available. They include one *balbisiana* BAC library of the wild diploid Pisang klutuk wulung donated by the *Centre de Coopération internationale en recherche agronomique pour le développement* (CIRAD) and four *acuminata* BAC libraries: two are of the wild diploid Calcutta 4 (one from CIRAD and the other from Texas A&M University), one is of the wild diploid Tuu gai from the *Centro de Investigación Científica de Yucatán* (CICY), and the fourth is of the triploid cultivar Grande naine from CIRAD.

The Centre also provides libraries of expressed sequence tags (ESTs), insert DNA, repetitive DNA clones and molecular markers. Information about the Centre is available at <http://www.musagenomics.org/index.php?page=resources>

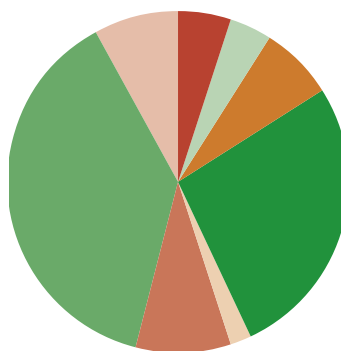


Jaroslav Dolezel of the Institute of Experimental Biology, which hosts the *Musa* Genome Resources Centre.
A. Vezina, INIBAP



Genome sequencing at The Institute for Genomic Research in the United States. TIGR

Breakdown of the 22 candidate genes linked to drought stress identified by scientists from the University of Leicester.



- Transcription factors 5%
- Cell division and growth 4%
- Protein synthesis 7%
- Novel genes 27%
- Other 2%
- Metabolism 9%
- Abiotic stress related 38%
- Cellular communication and signal transduction 8%

In Brazil, scientists from EMBRAPA and the Catholic University of Brasilia, took samples of Calcutta 4 leaves exposed to temperatures as cold as 5°C and as hot as 45°C, to build two cDNA libraries, one for each type of temperature stress. About 10% of the sequenced cDNAs were present in both types of stressed

leaves, while 42% were present only in the leaves exposed to cold and 48% only in the ones exposed to hot temperatures.

Meanwhile in the UK, a team from the Department of Biology at the University of Leicester has isolated candidate genes involved in the tolerance to drought in three dessert cultivars (AAA). The scientists gave one plant its daily share of water (200 ml) but withheld most of it from another. The expression profile of the cDNA led to the identification of 22 candidate genes. The scientists were able to show that 3 of the genes had been activated by the drought stress, 5 had been down-regulated and 14 had been turned off. The genes were then classified according to their putative function. Nearly 40% of them were genes that had never been identified before (pie chart, above).

Many Consortium members are similarly busy building EST and cDNA libraries of genes expressed in male flowers, young leaves, roots of *in vitro* plantlets, the peel of green fruits, leaves exposed to the fungus *Mycosphaerella fijiensis* (the causal agent of black Sigatoka) and plants wounded by weevils, to name only a few. The sequences, as they are elucidated, are posted on the website of the Global *Musa* Genomics Consortium so that others can join the effort of comparing and analysing them.

GETTING A HANDLE ON DIVERSITY

Even when the genes have been identified and their functions revealed, another challenge will remain: to document allelic diversity—which is the difference between knowing that a gene codes for eye colour, for example, and knowing which variant, or allele, codes for blue eyes and which one for brown. This is a crucial piece of the puzzle for breeders, who want to be able to recognise individual alleles in

crosses to ensure that the new variety expresses the characteristic they want to impart. The same holds for scientists who want to genetically engineer bananas.

This level of knowledge would also help to ensure that collections capture the existing allelic diversity, which in turn is expected to increase the use of the diversity conserved in genebanks. At the moment, the search for useful traits is a laborious, or sometimes a rather hit-or-miss affair, involving biological screening in the field or glasshouse, or simply crossing the relatively few materials known to be compatible. Systematic screening for allelic diversity, using molecular methods, can help to pinpoint useful diversity and ensure that it is not lost before we even know what is there.

The development of powerful molecular tools by initiatives such as the Global *Musa* Genomics Consortium provides an unprecedented opportunity to use the mostly untapped diversity available in wild and cultivated *Musa*. Used wisely, a better understanding of the plants on which we rely to produce our food, and of the pests and diseases that constrain our ability to do so, would also allow us to increase the nutritional quality of our food crops and lessen the impact of agriculture on the environment. But improving agriculture and our food are not the only reasons to pursue genomics. Uncovering the hidden genetic diversity of bananas should remind us of our moral duty to conserve it. ☞

The comparative advantage of bananas

Even though it ranks next to rice, wheat and maize in terms of its importance as a food crop, the banana does not command the same kind of attention from researchers and the donors who fund their efforts. However, the banana's unusual genetic make-up could attract a whole new level of interest. The banana has many things to offer scientists interested in the power of comparative genomics to study fundamental questions of biology and evolution. For one thing, the banana provides an opportunity to analyse traits not found in the current model plants.

The reproductive system of bananas includes both sexual and vegetative modes of propagation, the latter being the result of various forms of sterility combined with parthenocarpy (fruit development without fertilization), which is relatively rare in monocots. What is more, Southeast Asia is unique in hosting sexually and asexually reproducing bananas, both of which have co-evolved with pathogens. Enlarging the scope of genomic studies to cultivars that, like plantains, diversified outside Asia in the absence of some key pathogens would provide a valuable tool to study how bananas evolved, with and without their original attackers.

The range of ploidy levels found in bananas also offers a special opportunity to gain insight into the greater-than-additive gains in crop productivity that often accompany polyploidy. More complex than cotton and sugarcane polyploids, which contain only one level of ploidy within the taxon, banana varieties include various levels of ploidy and mixes of the A and B genomes (AA, BB, AB, AAA, AAB, ABB, AAAA, AABB, AAAB). This creates opportunities to study not only the relationship between ploidy and phenotype, but also the causes and consequences of polyploidy for genome organization.