

## **Opportunities for Bridging the Gap between Genomics and Genetic Improvement in *Musa* spp.**

N. Roux<sup>1</sup>, M. Rouard<sup>1</sup>, X.L. Huang<sup>2</sup> and M. Smith<sup>3</sup>

<sup>1</sup> Commodities for Livelihoods programme, Bioversity International, Parc Scientifique Agropolis II, 34397 Montpellier, France

<sup>2</sup> School of Life Sciences, Sun Yat-Sen University, China

<sup>3</sup> Department of Primary Industries & Fisheries, Maroochy Research Station, 47 Mayers Road, Nambour, Qld 4560, Queensland, Australia

**Keywords:** Banana, consortium, genetic improvement, genomics.

### **Abstract**

**Banana (*Musa* spp.) represents one of the most important commodity crops in the world. They are important as an export crop but also play a major role in local food security in developing countries. Banana is susceptible to an ever-increasing range of pests and diseases requiring increased use of pesticides that have adverse environmental and health impacts and threaten the sustainability of the crop. There is an urgent need to develop improved banana cultivars with a wider range of pest and disease resistance. In addition, new cultivars better adapted to the environment (i.e. tolerant to abiotic constraints) and which satisfy consumer needs are also needed. Breeding programmes aiming at broadening the genetic basis and providing new genes of interest face many obstacles (e.g. low fertility, structural heterozygosity, polyploidy). The Global *Musa* Genomics Consortium (<http://www.musagenomics.org>) is an international network of investigators committed to understanding genomic evolution in relation to biotic and abiotic stresses in a polyploid, vegetatively propagated crop and also to provide meaningful insights for the plant community. Indeed, *Musa* lies taxonomically within the monocots, although distant from the grass family (Graminiaceae), in a position that is important for comparative and evolutionary genomics. The Consortium currently brings together expertise from 40 institutions in 24 countries. Members are committed to close collaboration and agree to share materials and resources, including sequence data and enabling technologies. Wherever possible, the products of the Consortium are placed in the public domain. With the soon-to-be-available *Musa acuminata* sequence, it is essential that the community be ready to use this information as efficiently as possible. This paper summarises the discussions and recommendations of participants at a workshop held during this Symposium to address the gap between genomics and breeding and suggest a number of collaborative areas of work to accelerate *Musa* breeding efforts around the world.**

### **INTRODUCTION**

*Musa* (banana and plantain) are grown in 126 countries with an annual production exceeding 100 million tonnes (<http://faostat.fao.org/>). More than 87% of the global production is consumed locally and provides a staple food and major source of income for more than 400 million people in developing countries.

The genus includes some 25 species, divided into four sections: Callimusa, Rhodochlamys, Australimusa and Eumusa. This last section contains two major species,

*Musa acuminata* (A genome) and *Musa balbisiana* (B genome), both diploid ( $2n = 2x = 22$ ) which are at the origin of most of the edible bananas through intra- and inter-specific hybridisation. The export industry is dependent on only three triploid (AAA) clones, all from the same Cavendish subgroup. They are highly susceptible to black leaf streak and vulnerable to emerging diseases such as the Fusarium tropical race 4 (TR4) fungal disease. The 87% locally consumed bananas (including Cavendish clones, plantains, East African Highland bananas, Silk, Pisang Awak and others) that are essential for food security in developing countries are also threatened by major pest and diseases. For some of the pathogens, pesticides can be used, but these are not environmentally friendly, hazardous to human health and too expensive for small-scale farmers. In addition, new cultivars better adapted to the environment (i.e. tolerant to abiotic constraints) and which satisfy consumer needs are also needed.

Hybrids are currently being developed by six major banana breeding programmes and evaluated in many sites around the globe (see these Proceedings for examples of the breeding efforts currently underway): *Fundación Hondureña de Investigación Agrícola* (FHIA) in Honduras; *Empresa Brasileira de Pesquisa Agropecuária* (EMBRAPA) in Brazil, *Centre de Coopération Internationale en Recherche Agronomique pour le Développement* (CIRAD) in French West Indies, *Centre Africain de Recherches sur Bananiers et Plantains* (CARBAP) in Cameroon; International Institute of Tropical Agriculture (IITA) in Nigeria and Uganda, and National Research Centre for Banana (NRCB) in India. These breeding programmes are breeding different types of banana depending on the needs of the region and on the source of funding and their mandate. Most of these breeding programmes have been using the same breeding scheme over the years, as well as the same set of genotypes as parents. A wider diversity could be used, but the complete diversity pool that could be tapped is not fully known. In addition, the origin of the triploid clones most commonly used is not known, which would greatly help in understanding *Musa* genetics and thus genetic improvement. The vegetative propagation, sterility and polyploidy of edible *Musa* clones makes genetics of *Musa* very difficult to study and understand. The development of new genomics tools is expected to facilitate germplasm exploitation and speed up the conventional genetic improvement process by using, for example, molecular markers in marker-assisted breeding. The workshop organised within the framework of ProMusa was a unique opportunity to bridge the gap between genomics and breeding of *Musa* spp.

## **BANANA AND PLANTAIN BREEDING: WANTS AND NEEDS**

There are a number of issues identified by breeders, and those associated with breeding programmes, that will help accelerate their efforts. Better access to germplasm and high-quality associated information is a central requirement. Areas for further prospection need to be identified together with phenotyping tools to better use both new and existing germplasm. Where improved germplasm for breeding has been identified, there is a need to have a workable strategy to deal with intellectual property (IP) management and rights. This also applies to hybrids generated from breeding programmes and is particularly relevant in protecting IP during evaluation.

Germplasm enhancement has also been identified as an area that needs attention. Pre-breeding to evaluate compatibility and knowledge/sources of genetic determinants of important traits such as resistance/tolerance to biotic and abiotic stresses, yield potential, parthenocarpy, fruit quality and nutritional value are all highly desirable. 'Ready-to-use' molecular marker technologies are seen by breeders as a first step in bridging the gap

between breeding and genomics. They are also seen as a tool for identifying and protecting each unique hybrid developed from their programmes. Molecular characterisation can not only lead to discovery of genes for useful traits but also provide information about alleles in diverse germplasm. If this can be combined with molecular information gained on host-pathogen interactions, as well as pathogen diversity, great strides can be made in resistance breeding. An important requirement for disease evaluation is the availability of sites for screening pre-breeding materials and hybrids that do not jeopardise quarantine measures in either the country where breeding takes place or in the countries undertaking regional evaluation. Emphasis was given for an urgent need to screen germplasm for resistance to *Fusarium* tropical race 4.

Aside from molecular genetics, further conservation efforts, particularly of wild types and improved diploids, is seen as another important area. Better ways to increase seed production during crossing, to rescue embryos from seed and to conserve seed were also identified as areas needing further attention. Last, but certainly not least, is the need for enhanced financial support for conventional breeding programmes where genetic gains are already being made in banana and plantain improvement.

### **MUSA GENOMICS: WANTS AND NEEDS**

Genomics provides a powerful tool for identifying genes and markers of agronomic importance. Over the last years, the Global *Musa* Genomics Consortium (GMGC) has been increasing genomic resources. Genome information produced by *Musa* research projects has been made available through the GMGC website (<http://www.musagenomics.org/>). Most of the genomic sequences were obtained by sequencing BAC (Bacterial Artificial Chromosomes) clones bearing genes of interest with a focus on biotic and abiotic stresses. In addition, ESTs (Expressed Sequence Tags), and BAC ends have also been sequenced, and other initiatives are ongoing (454, Solexa etc.).

A set of bioinformatics tools and resources have been developed to manage and analyse *Musa* datasets. For instance, all the publicly available BACs were processed through pipelines of analyses, and predictions of genes and repeat elements were stored and made accessible through a genome browser via the GMGC website (<http://gnpannot.musagenomics.org/>).

Since *Musa* genome sequencing started, *Musa* has entered the post-genomic era. If gene structure prediction tools are a prerequisite for gene discovery studies, other activities to predict and validate the function of genes are necessary. During the workshop, comparative genomics was mentioned as one of the key elements for gene function prediction. Indeed, recent years have seen a large increase in the production of genomic sequences for the major crops. Some species such as *Oryza sativa*, *Arabidopsis thaliana* or *Brachypodium distachyon* have been the most experimentally studied for some years, and reliable information could be transferred to other plants such as *Musa*.

Considering gene validation, transcriptomics, expression array, RNA interference (RNAi), proteomics and genetic transformation were all identified as important areas of research. Re-sequencing has been pointed out as a necessary way to go. Information about a B genome would be valuable. Sequencing of the parent of 'DH Pahang', for example, would be interesting to study, especially considering heterozygosity and translocations. Whole-genome scanning (high-density genotyping) was also mentioned because of the usefulness of markers (e.g. Single Nucleotide Polymorphism-SNP) for re-sequencing.

## **PRIORITISING THE ISSUES NEEDED TO BRIDGE THE GAP BETWEEN GENOMICS AND GENETIC IMPROVEMENT**

A survey of the *Musa* breeding and genomics research community was taken to assess their needs, and issues that were prioritised during the survey were further short-listed by participants during the workshop. As a result, four areas were identified for further discussion and for identifying what is currently available, what is needed, what the priorities are, some potential collaborative areas of work and how tasks can be distributed between different partners. These four areas included: 1) germplasm collection and characterisation; 2) phenotyping; 3) molecular markers for gene discovery and for studying diversity; and 4) identification of parents for recombination and cultivar development.

### **Germplasm Collection and Characterisation**

*Musa* germplasm is currently available to those groups involved in *Musa* crop improvement through Bioversity's International Transit Centre (ITC) hosted by the Katholieke Universiteit Leuven, Belgium. However, availability and access to wild-type germplasm and improved diploids, particularly B types, is limited. Material may also be found in national collections and in the wild but IP issues hinder access. Areas such as Myanmar, islands in the Pacific and islands in East Africa still need to be explored.

Linked with access to the actual germplasm is providing reliable information about the material. Much information is already available through Bioversity's *Musa* Germplasm Information System (MGIS, <http://www.crop-diversity.org/banana/>) and the International *Musa* Testing Programme (IMTP), but new sources of resistant germplasm or accessions with traits of interest for new breeding targets (e.g. human nutrition, drought hardiness) have hardly been described. Standardised protocols for characterisation have to be agreed upon and implemented. These need to be rigorous enough to characterise germplasm across a range of edaphic, climatic and management regimes. Training will need to be put in place so that information that is collected and shared will be of use to those involved in utilising germplasm resources. National programmes are urged to not only characterise their germplasm but to share this information with the international community.

Priority needs to be given to developing an IP management strategy and lobbying governments to share their genetic resources. Bioversity can play a role in these lobbying efforts, as well as provide protocols for collection, characterisation and distribution. Already, centralised facilities exist for virus indexing, cryopreservation, DNA extraction and genotyping. They will continue to play their part in the further utilisation of *Musa* germplasm.

Several international initiatives are being put in place to stimulate the exchange of genetic plant material, such as the International Treaty for Plant Genetic Resources for Food and Agriculture (IT-PGRFA). This treaty is being implemented in several countries and through a project that Bioversity is coordinating where ten priority collections are safeguarding their germplasm through characterisation *in vivo*, duplication *in vitro* and safety duplication of a subset in the ITC as a backup. Nevertheless, ITC is still lacking many wild types which are crucial to breeders as a source of new genes, such as for disease resistance.

## **Phenotyping**

The phenotyping working group identified many areas in common with the group discussing germplasm collection and characterisation. These included: the need to phenotype across environments, the need for standardised protocols for assessing responses to biotic and abiotic stresses, yield and postharvest traits, the need for a network to share knowledge, the need for virus indexing centres and the need to place more germplasm into the ITC while acknowledging that IP issues still need further resolution.

Some areas were discussed in more detail. For instance, it was suggested that many of the quarantine issues associated with screening germplasm for resistance to Fusarium wilt or black leaf streak could be circumvented by phenotyping in a closed facility and particularly one located in a non-banana producing country. It was felt that the criteria for selection of test sites should be re-evaluated.

The group concluded that one of the limitations to utilising genomics is a lack of material, knowledge and precision tools/standards for phenotyping. There is a wide range of diversity to phenotype and a large number of diverse genotypes are needed for association genetics. Databases that are easy to use and understand will be very important to make best use of the information generated through phenotyping. In addition, novel phenotypes for specific applications may be considered, such as the ability to undergo somatic embryogenesis for ease in genetic transformation studies. Finally, in order to build effective collaboration, project proposals need to be written and sources of funding sought.

## **Molecular Markers**

Two types of markers were discussed: markers for diversity studies and markers for gene discovery. Segregating populations are key resources that are already available with the *Musa* research community but are they all accessible? Through known projects, such as the recent project supported by the Generation Challenge Programme: “Targeted *Musa* genome sequencing and frame map construction“, a segregating population was produced by CIRAD for the development of a genetic map “BORLI” (coming from the cross *Musa acuminata* ‘Borneo’ X *Musa acuminata* ‘Pisang Lilin’) (Bakry et al., 2007; Hippolyte et al., 2010). Other segregating populations are reported by members of GMGC for biotic (Fusarium wilt, black leaf streak and Sigatoka leaf spot) and abiotic (drought, low temperature) related traits. Detailed information of those segregating populations is however missing. In order to have a high-quality segregating population, the first step is to have well-characterised and phenotyped parents. Ideally, breeders need to be consulted in the choice of parents and their experience with crosses. There is a need also from breeders to incorporate new wild types as parents. Population size is important and is related to quality of F1 and the type of traits available. The group estimated that an F2 population would vary between 100 and 400 individuals. Accurate phenotyping is then required to further be able to link markers to traits.

Populations need to be available across the community but how can populations be accessed and distributed? Should this be via in-vitro plantlets? In this case, how would expensive virus indexing requirements be managed? The absolute minimum requirement would be through DNA (but in this case highly standardised, accurate phenotyping is required in at least one area). A suggestion was made to perform the cross between virus-free parents and F1 selfing under controlled conditions (e.g. insect-proof greenhouse). F2

individuals would then be multiplied and disseminated as in-vitro plantlets to at least three locations (five copies per individual).

As a first priority, the group suggested that Bioversity takes the lead in making an inventory of segregating populations currently available, with all the necessary information including conditions to make them available. Once this is done, a workshop should be organised to define priority traits and strategies for genetic mapping. The group agreed that it was essential to have collaboration because the tasks are too large for a single institution. Progeny genotyping can be distributed according to the expertise available. It is however important that the data are centralised and made public. The whole *Musa* genome sequence will soon be available so integration of sequence data will enrich the database.

### **Cultivar Development**

The potential for cultivar development in *Musa* still needs to be fully exploited, especially with the emerging diseases that are threatening the crop. But genetic improvement in *Musa* is difficult as most of the edible varieties are parthenocarpic triploids (production of fruits without fertilisation of ovules), thus seedless. On the other hand, this is a characteristic that breeders would want to keep. Polyploidy complicates the crossing schemes. Compared to other crops, the number of breeding programmes in *Musa* is seen as excessive. Nevertheless, more and more national programmes are now becoming important such as in India, Brazil, Uganda, China and the Philippines.

Due to limited resources and breeding-related difficulties breeders should produce suitable tools (see Table 1) and exchange them to boost genetic improvement in *Musa*. It seems that with the *Musa* genome sequence on the horizon, this is the moment to start exchanging material. Genetic stocks also represent a valuable means to improved understanding of *Musa* genetics. Pre-breeding, also called “genetic enhancement”, has three major uses: 1) To prevent genetic uniformity and consequent genetic vulnerability; 2) To raise yields to new heights; and 3) To bring in new quality traits not found in local cultivars. A process needs to be put in place to stimulate exchange of genetic stocks between breeders.

### **THE WAY AHEAD**

The International *Musa* Testing Programme (IMTP) has allowed the exchange within the *Musa* research community of improved germplasm resistant to major pests and diseases such as black leaf streak, Fusarium wilt and nematodes. Feedback will inform the breeder on how his improved variety is behaving under various environments. The recipient receiving the material benefits from the scheme since it brings new resistant cultivars that can be directly cultivated in his/her country. Nevertheless, to better understand the genetics, it would be necessary to take one step back to study the inheritance of specific traits. For this, the community needs more easily available and representative germplasm. Some crosses are being performed by breeding programmes but information is not available on the progenies. There are approximately ten populations in total available at NRCB (drought, Fusarium), EMBRAPA (three crosses between AA), NARO (characterising one), CIRAD, IITA, CARBAP.

Workshop participants identified the following research needs:

1. Information on genetic mechanisms underlying important characters (a population dedicated to association studies);

2. A population segregating for parthenocarpy, bunch weight, number of hands, Fusarium etc... designed for understanding genetic determinants;
3. Adequate mechanisms to phenotype Fusarium interactions with banana plants;
4. Gene constructs for breeding and appropriate genes for genetic engineering;
5. Validation of gene constructs;
6. Several technical guidelines for breeding (e.g. embryo rescue);
7. Accurate and updated information on germplasm in MGIS that is pertinent to breeders (more data on agronomic traits, protocols to evaluate new germplasm and enter the data into MGIS).

The top priorities are germplasm collection (wild species), development of segregating populations, Fusarium screening and gene constructs.

A lot of these areas need collaboration between NARS, international programmes, universities, private sector and other actors:

1. Collaborative effort in collection and study of wild populations, making controlled crosses between accessions, evaluate the crosses for segregation, compatibility etc. Generating information on combining ability, variance components, etc. Nevertheless, sharing of segregating populations is still, in the current policy climate, not feasible.
2. Collaborative effort to develop and refine techniques for Fusarium-*Musa* interactions (e.g. TR4 cannot be evaluated in Brazil).
3. Characterising and validating gene constructs.
4. To perform those tasks, scientists need support from their institutes and their country. The tasks should be distributed according to type of germplasm and the specialty of each group. For example, the plants may be physically located in individual South(east) Asian countries, while technical support can come from the whole *Musa* community according to interests and expertise. It is only through collaboration that our understanding of *Musa* genetics can be improved to a point that new strategies can be put into place for the development of a range of new banana and plantain cultivars.

## ACKNOWLEDGEMENTS

We acknowledge the individual and collective contributions of the workshop participants and thank everyone who contributed to the questionnaire that was sent out before the workshop.

## References

- Bakry, F., Baurens, F.C., Dolezel, J., Heslop-Harrison, J.S., Hippolyte, I., Matsumoto, T., Miller, R.G., Rouard, M., Roux, N., Sidibe-Bocs, S., Souza, M. and Sasaki, T. 2007. Targeted *Musa* genome sequencing and frame map construction. p.115-117. In: 2007 Project Mid-Year and Final Reports: Competitive and Commissioned Projects. Generation Challenge Programme, Mexico DF, Mexico.
- Hippolyte, I., Bakry, F., Seguin, M., Gardes, L., Rivallan, R., Risterucci, A.M., Jenny, C., Perrier, X., Carreel, F., Argout, X., Piffanelli, P., Khan, I.A., Miller, R.N.G, Pappas, G.J., Mbéguié-A-Mbéguié, D., Matsumoto, T., De Bernardinis, V., Huttner, E., Kilian, A., Baurens, F.C., D'Hont, A., Cote, F., Courtois, B. and Glaszmann, J.C. 2010. A saturated SSR/DArT linkage map of *Musa acuminata* addressing genome rearrangements among bananas. *BMC Plant Biology* 10(1):65. doi:10.1186/1471-2229-10-65.

## Tables

Table 1. Bridging the gap between breeders and molecular biologists.

<b>Breeders need from molecular biologists:</b>	<b>Molecular biologists need from breeders:</b>
Markers unique to each variety	Well-characterised (genotyping and phenotyping) mapping populations; Access to breeding population (e.g. improved diploids)
Markers for traits; MAS	Access to germplasm; Wild type germplasm (including A & B, geographical origin)
Knowledge of genetic determinants for traits including polymorphisms	Markers for traits
High density genotyping	Core set of priority traits for farmers
Gene discovery	Non-destructive phenotyping
Their participation in pre-breeding, phenotyping, genetic characterisation	More genetic studies (e.g. cytogenetics, association mapping)
Information on host/pathogen interactions	Further understanding of cell biology - Plant physiology
Information on basis of nutritional value	Breeding strategies (is transgenic to be accepted?) or crossing strategies (diploid, tetraploids etc.); Routes to market