Sequencing the Major *Mycosphaerella* Pathogens of Wheat and Banana

G.H.J. Kema[a]
Plant Research International B.V.
Wageningen University and Research Center
P.O. Box 16, 6700 AA Wageningen
The Netherlands

Keywords: Black Sigatoka, black leaf streak disease, *Musa*

Abstract

***Mycosphaerella* is one of the largest genera of plant-pathogenic fungi with more than 1,000 named species, many of which are important pathogens causing leaf spotting diseases in a wide variety of crops including cereals, citrus, banana, eucalypts, soft fruits and horticultural crops. A few species of *Mycosphaerella* cause disease in humans and other vertebrates. An international project was initiated to sequence the genomes of *M. graminicola* and *M. fijiensis*, two of the most economically important pathogens of wheat and banana, respectively, along with 40,000 ESTs from *M. fijiensis* and the related maize pathogen *Cercospora zeae-maydis*. The 9x *M. graminicola* genome size is 39.8 Mb with chromosome sizes from 548 kb to 6 Mb and a complete circular mitochondrial genome of 43,947 bp. Our data indicate that *M. graminicola* has both the largest chromosome number and the smallest chromosome sizes recorded among filamentous ascomycetes. The *Mycosphaerella* Genomics Consortium, which was established in 2003, decided to use *M. graminicola* as the model to develop more genetic and genomic research on *M. fijiensis*. Since 2003, *M. fijiensis* EST sequencing has resulted in more than 30,000 ESTs, and the genome sequencing was recently finished at 7.8x. The genome size of *M. fijiensis* is 80% larger than that of *M. graminicola*. The completed mitochondrial sequence is more then twice as large, and the estimated nuclear genome size is approximately 72 Mb. The extension of the genome size of *M. fijiensis* seems to be mostly due to additional repeated sequences. The status of *Mycosphaerella* sequencing will have a significant effect on future studies aimed at the control of black leaf streak disease. The current status of both sequencing projects and other initiatives to exploit this information and to put it into a multidisciplinary approach focusing on sustainable management of the disease will be discussed.**

INTRODUCTION

Mycosphaerella is one of the largest genera of plant-pathogenic fungi with more than 1,000 named species, many of which are important pathogens causing leaf spotting diseases in a wide variety of crops including cereals, citrus, banana, eucalypts, soft fruits and horticultural crops. A few species of *Mycosphaerella* cause disease in humans and other vertebrates. Through the Community Sequencing Program sponsored by the U.S. DOE-Joint Genome Institute (JGI), the USDA-ARS at Purdue University and Plant Research International initiated an international project to sequence the genomes of *M. graminicola* and *M. fijiensis*, two of the most economically important pathogens of wheat and banana, respectively, along with 40,000 ESTs from *M. fijiensis* and the related maize pathogen *Cercospora zeae-maydis*. The 9x *M. graminicola* sequencing is complete and was made public on 1 November 2006 following automated and manual annotation (http://genome.jgi-psf.org/Mycgr1/Mycgr1.home.html). Due to very good assembly statistics, as well as a >2000-marker Diversity Array Technology (DArT) linkage map that was aligned to the genome, JGI decided to complete the sequencing of the *M. graminicola* genome at the Stanford Human Genome Center. The genome size is 39.8 Mb, and the majority of chromosomes, with sizes from 548 kb to 6 Mb, have been

[a] On behalf of the International Mycosphaerella Genomics Consortium (IMGC; http://imgc.inibap.org/).

sequenced completely including both telomeres, with currently only eight gaps remaining, and a complete circular mitochondrial (mt) genome of 43,947 bp. Our data indicate that *M. graminicola* has both the largest chromosome number and the smallest chromosome sizes recorded among filamentous ascomycetes. In conclusion, these data make *M. graminicola* the first filamentous fungus with a completely finished genome sequence.

The International Mycosphaerella Genomics Consortium (IMGC; http://imgc.inibap.org/), which was established in 2003, decided to take *M. graminicola* as the model to develop more genetic and genomic research in *M. fijiensis*, the causal agent of black leaf streak disease in banana. Since then, *M. fijiensis* EST sequencing has resulted in 30,011 ESTs, and the genome sequencing was recently finished at 7.1x and made public on August 3, 2007. The genome size of *M. fijiensis* is approximately 72 Mb, which is 80% larger than the *M. graminicola* genome and the completed mt sequence is 74 kb, which is almost twice the *M. graminicola* mt genome (44 kb). The extension of the genome size of *M. fijiensis* seems to be mostly due to additional repeated sequences, as the estimated number of genes is 10,327 compared to the predicted 11,414 genes in *M. graminicola*. The EST data sets from both pathogens support this conclusion.

The status of *Mycosphaerella* sequencing will have a drastic effect on future studies aiming at the control of black leaf streak of banana. Many initiatives have started, including the International Pesticide Reduction Plan for Banana (PRPB), to exploit this information in a multidisciplinary approach focusing on a sustainable management of the disease.

HOW TO EXPLOIT A DNA SEQUENCE IN A MANAGEMENT CONTEXT?

Our current programme will exploit these genome sequences to retrieve information that can be used in the development of management tools and strategies. Some can be applied at a relatively short notice, while others will be more distantly available to contribute to controlling black leaf streak. The overview below provides a summary of opportunities that will contribute to understanding the diversity and dynamics of natural populations. In addition, genomic data will contribute to the elucidation of the host-pathogen interaction and hence to new breeding strategies or novel leads for innovative disease control. Moreover, they support detection and quarantine policies.

Short-Term Utilisation of Genomic Data

1. Microsatellite and Mating Type Loci. The sexual cycle of *M. fijiensis* plays a crucial role in the epidemiology, and genetic variation among field isolates is large (Gauhl et al., 2000; Rivas et al., 2004). Molecular markers, particularly restriction fragment length polymorphisms (RFLP) and random amplified polymorphic DNA (RAPD), have contributed to the genetic analyses of this pathogen. Extensive population genetics analyses of *M. fijiensis* have been conducted using RFLP markers (Carlier et al., 1996). However, handling these markers requires substantial technical and operational expertise that is not required for other types of markers, such as microsatellites (Selkoe and Toonen, 2006). These microsatellite markers, or simple-sequence repeats (SSR), short tandem repeats (STR) or variable number tandem repeats (VNTR) comprise a specific sequence of nucleotides that is tandemly repeated in the genome. They are ideal for genetics and population biology. Like all genetic loci, SSRs are subject to point mutations that generate large numbers of alleles at a single locus, each differing by one or more copies of the repeat unit that can be employed for diversity studies. The flanking regions of microsatellite loci usually do not contain mutations, hence these can anchor to specific primers for amplification, yielding a high level of easily scored polymorphisms (Selkoe and Toonen, 2006). However, the identification and validation of useful SSRs is a time-consuming effort, which resulted only in 11 SSR loci in *M. fijiensis* (Neu et al., 1999). Using a software pipeline that was developed at Plant Research International for the analysis of the 30,137 ESTs from *M. graminicola* IPO323, we found 77 SSR loci that could be easily used in diversity PCR screens and genetic mapping experiments (Goodwin et al., 2006). A similar approach using the sequence information that JGI

generated from *M. fijiensis* CIRAD86 resulted in 203 SSR loci from the 31,011 ESTs and approximately 4,600 SSR loci from the genomic sequence of this isolate. Both sets of SSR loci are currently being validated. These markers will be essential for future population studies and genetic mapping experiments, as well as for specific isolate genotyping. As these markers will provide a deep insight in the population structure of *M. fijiensis*, they can be readily used in areas where the disease is still establishing itself, such as in parts of Brazil (São Paulo State, for instance, where it arrived in 2004; compared to the Amazon area, where it arrived in 1997). In addition, they can be used by the community to understand adaptation processes to colder temperatures as *M. fijiensis* is successfully competing with *M. musicola* at higher altitudes in Costa Rica (Arzanlou et al., 2007).

Similarly, mating type loci can be used for a quick population analysis. We recently isolated and characterised the mating type loci from *M. fijiensis* (Conde et al., 2007) that can now be used for rapid mating type genotyping and population analyses in order to test whether natural populations are randomly mating (Zhan et al., 2002). In addition, thorough analyses of these loci will also reveal whether species in the Sigatoka leaf spot complex of banana do sexually interact with each other potentially resulting in new hybrid *Mycosphaerella* species that may even adapt to other host systems (Arzanlou, unpublished data). Genomic data are essential for these analyses.

2. Targets for Quantitative Real Time PCR. Genomic information is very helpful for the development of specific primers and/or probes that can be used in quantitative assays. Although a DNA sequence is not a prerequisite for the development of these tools, it definitely enables a quick generation of multiple target genes. Quantitative real time PCR assays for the Sigatoka complex in banana have been developed (Arzanlou et al., 2007). *Mycosphaerella fijiensis*, *M. musicola* and *M. eumusae* can now be easily identified and quantified in leaves with Sigatoka leaf spot symptoms. However, using this technology, such data can also be retrieved from symptomless leaves. Evidently, quantitative data on non-symptomatic or early disease-diagnosed leaves will enable timely control measures. Compared to the traditional visual estimation of disease pressure in banana plantations, this technology enables the collection and processing of many more samples, which aids accuracy that will contribute to an improved management of black leaf streak disease. Genomic information will now also enable rapid development of additional molecular diagnostic tools for other *Mycosphaerella* species that are part of the complex interactions that may take place in Sigatoka leaf spot-infested plants. The development of such molecular diagnostics is now relatively simple, and similar approaches for *Radopholus similis* and potentially *Fusarium oxysporum* f. sp. *cubense* are now under development. However, the application of these tools in practical disease management requires their thorough validation in the field. Such projects will provide data on damage and action thresholds that are required for their implementation. There is no doubt about the use of molecular diagnostic in future enhanced disease management strategies; during the growing cycle, but certainly for quarantine measures, as has been demonstrated in Australia (Henderson et al., 2006). Furthermore, such diagnostics are excellent research tools to quantify fungal biomass in precise phenotyping experiments. Particularly in the *M. fijiensis*-banana interaction where symptom development takes long, molecular diagnostics will contribute to understanding the complexity of pathogenesis and resistance.

3. Genomics and Fungicide Resistance. Fungicides play a significant role in the management of *Mycosphaerella* diseases of cereal and banana with a total global expenditure of ~€ 1,100 million/year. Septoria tritici blotch is the major wheat disease in Western Europe. Wheat in this region has a total acreage of 27 million ha. Therefore, *M. graminicola* is the primary target for the agrochemical industry. Genomic information from *M. graminicola* contributes significantly to the identification of new targets and the development of new fungicides. Similarly, *M. fijiensis* genomics will contribute to new products. However, this is typically a longer-term application of a genomic sequence. Resistance monitoring in fungal populations will benefit more quickly from a genome

sequence. The loss of efficacy of strobilurins to *Mycosphaerella* species represents the most recent and eye-catching resistance development phenomenon (Sierotzki et al., 2000; Gisi et al., 2005). In most fungal species, including *M. graminicola* and *M. fijiensis*, this resistance is based on a guanine/cytosine replacement converting the amino acid glycine to alanine at position 143 of the cytochrome b protein (G143A), which is encoded by the mt genome. Evidently, having the sequence of this mt DNA contributes strongly to the identification of other (rare) point mutations that lead to resistance. Furthermore, the mt DNA provides the development of diagnostic tools to track the male/female contributions of parental strains in crossing experiments that provide a deep insight in the distribution of mutant alleles in natural populations (Ware, 2006). Azole fungicides represent another class of chemistry that is commonly used in *Mycosphaerella* disease management. Resistance development to these compounds is characterised by a slow reduction of efficacy, probably due to the quantitative inheritance of this character (Stergiopoulos et al., 2003). Recently, Selmecki et al. (2006) discovered a strong correlation between aneuploidy and azole resistance in *Candida albicans* by using a whole genome micro-array. Similar events might also contribute to azole resistance development in *Mycosphaerella*, where aneuploidy is very common. Therefore, micro-arrays enable whole genome comparisons between individual *M. fijiensis* strains to study the effects of aneuploidy on gene copy number and expression studies to understand how these relate to the gradual accumulation of azole resistance in *Mycosphaerella*. A genomic sequence, therefore, can strongly support the discovery and monitoring of fungicide resistance alleles and hence, will contribute to optimise their application in disease management.

Long-Term Utilisation of Genomic Data

1. Whole Genome Comparisons. As pointed out above, a genomic sequence of *M. fijiensis* will enable the construction of micro-arrays for a range of different experiments that will provide a deep insight in the genomic and genetic complexity of this species. However, as we have shown recently (Arzanlou et al., 2007), the Sigatoka leaf spot complex of banana could, apart from *M. fijiensis*, *M. musicola* and *M. eumusae*, also harbour unknown *Mycosphaerella* species or hybrids of *Mycosphaerella* species. Micro-array-aided whole genome comparisons will provide unparalleled insight in the relationship of and the genomic differences between these *Mycosphaerella* species. Hence, a genomic sequence of *M. fijiensis* will be very useful to accelerate our understanding of other *Mycosphaerella* species on banana and how they interact. Such information will potentially provide clues for the (genetic) origin of *M. fijiensis* and its competitiveness over other *Mycosphaerella* species on banana. A better understanding of these aspects of fungal complexes will significantly contribute to long-term management and quarantine strategies.

2. Genetic Mapping. A DNA sequence is not required for genetic mapping studies, but will enable the generation of many molecular markers that will provide an overview of the genetic structure of an organism. The more markers one can generate, the higher the resolution. High-density linkage maps are also required for the assembly of fungal genomes, particularly for *M. fijiensis*, where deep sequencing is not envisaged. We have used this technology to generate a high-density DArT linkage map in *M. graminicola* (Wittenberg, 2006) and are currently in the process of developing a DArT linkage map for *M. fijiensis* using a segregating population between *M. fijiensis* isolates CIRAD86 (the sequenced isolate) and CIRAD139a. As these isolates differ significantly in pathogenicity/aggressiveness, these data will provide a first and deep insight in the genetic basis and inheritance of effectors that control pathogenicity/aggressiveness. Linking phenotyping and marker data will eventually help us to physically locate these effectors in *M. fijiensis*. Understanding the nature and dissemination of these loci within natural populations and in experimental intrageneric segregating populations is crucial to understand and model the adaptation of *M. fijiensis* to host resistance and its potential interaction with other *Mycosphaerella* species over time. In addition, high-density genetic mapping provides unparalleled insight in recombination events in *M. fijiensis,* and will

enable us to identify genes in quiet and active genomic regions that will be subject for further studies and exploitation. Finally, injection of candidate effector proteins can be used to screen for resistant responses in banana germplasm in order to identify and eventually exploit matching resistance genes.

3. Expression Studies. Unveiling the nature of the aforementioned effector genes that control pathogenicity in *Mycosphaerella* is a major research area. Functional analyses of these genes by 'knock-out' strategies (Balint-Kurti et al., 2001) will provide more information on their requirement for pathogenicity. Expression arrays and proteomic experiments will help us to understand their expression during pathogenesis and their interaction with other genes. Eventually, the data will help us to understand the role of these effectors, but also of, for example, secondary metabolism (toxins and their secretion) in pathogenicity. Once these genes are identified, their presence and distribution can be monitored in natural *M. fijiensis* populations, which is crucial for effective disease management.

CONCLUSION

The decision of the IMGC to adopt the wheat pathogen *M. graminicola* as a model to develop research in *M. fijiensis* turned out to be an excellent choice. The rapid generation of *Mycosphaerella* genomic data sets and materials for genetic studies will direct major changes in the *M. fijiensis* research area. Deployment of well characterised isolates is essential for high-quality phenotyping of banana germplasm. The genes that will be discovered from these genomic datasets that play essential roles in the host pathogen interactions, such as avirulence effectors, can be used to develop strategies for genetic modification with the aim of providing durable resistance to *M. fijiensis*. Sustainability of banana production systems, however, can only be achieved in a multidisciplinary context. To achieve this goal, high-quality data on the pathogens to be combated need to be generated. Those data should be thoroughly matched with expertise in agronomy, breeding, biological control, disease forecasting and modeling, soil science and nutrition as well as post-harvest, social and extension aspects for appropriate exploitation. The goal is to bring these different disciplines together in a network of scientists that will focus on pesticide reduction.

Literature Cited

Arzanlou, M., Abeln, E.C.A., Kema, G.H.J., Waalwijk, C., Carlier, J., De Vries, I., Guzmán, M. and Crous, P.W. 2007. Molecular diagnostics for the Sigatoka complex in of banana. Phytopathology 97:1112-1118.

Balint-Kurti, P.J., May, G.D. and Churchill, A.C.L. 2001. Development of a transformation system for *Mycosphaerella* pathogens of banana: a tool for the study of host/pathogen interactions. FEMS Microbiology Letters 195(1):9-15.

Carlier, J., Lebrun, M.H., Zapater, M.F., Dubois, C. and Mourichon, X. 1996. Genetic structure of the global population of banana black leaf streak fungus, *Mycosphaerella fijiensis*. Mol. Ecol. 5:499-510.

Conde, L., Waalwijk, C., Canto-Canché, B.B., Kema, G.H.J., Crous, P.W., James, A.C. and Abeln, E.C.A. 2007. Isolation and characterization of the mating type locus of *Mycosphaerella fijiensis*, the causal agent of black leaf streak disease of banana. Molec. Plant Pathol. 8:111-120.

Gauhl, F., Pasberg-Gauhl, C. and Jones, D.R. 2000. Black leaf streak. Disease cycle and epidemiology. p.56-62. In: D.R. Jones (ed.), Disease of Bananas, Abacá and Enset. CABI, Wallingford, UK.

Gisi, U., Pavic, L., Stanger, C., Hugelshofer, U. and Sierotzki, H. 2005. Dynamics of *Mycosphaerella graminicola* populations in response to selection by different fungicides. p.89-101. In: H. Lyr, P.E. Russell, H.W. Dehne, U. Gisi and K.H. Kkuch (eds.), Modern Fungicides and Antifungal Compounds II, 14th International Reinhardsbrunn Symposium, AgroConcept, Bonn, Verlag Th. Mann Gelsenkirchen.

Goodwin, S.B., Van der Lee, T.A.J., Cavaletto, J.R., Te Lintel Hekkert, B., Crane, C.F. and Kema, G.H.J. 2006. Identification and genetic mapping of highly polymorphic microsatellite loci from an EST database of the septoria tritici blotch pathogen *Mycosphaerella graminicola*. Fungal Genet. Biol. 44:398-414.

Henderson, J., Pattemore, J.A., Porchun, S.C., Hayden, H.L., Van Brunschot, S., Grice, K.R.E., Peterson, R.A., Thomas-Hall, S.R. and Aitken, E.A.B. 2006. Black Sigatoka disease: new technologies to strengthen eradication strategies in Australia. Australasian Plant Pathology 35:181-193.

Neu, C., Kämmer, D., Kahl, G., Fischer, D. and Weising, K. 1999. Polymorphic microsatellite markers for the banana pathogen *Mycosphaerella fijiensis*. Mol. Ecol. 8:523-525.

Rivas, G.G., Zapater, M.F., Abadie, C. and Carlier, J. 2004. Founder effects and stochastic dispersal at the continental scale of the fungal pathogen of bananas *Mycosphaerella fijiensis*. Molecular Ecology 13:471-482.

Selkoe, K.A. and Toonen, R.J. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol. Lett. 9:615-629.

Selmecki, A., Forche, A. and Berman, J. 2006. Aneuploidy and Isochromosome Formation in Drug-Resistant Candida albicans. Science 313:367-370.

Sierotzki, H., Parisi, S., Steinfeld, U., Tenzer, I., Poirey, S. and Gisi, U. 2000. Mode of resistance to respiration inhibitors at the cytochrome bc1 enzyme complex of *Mycosphaerella fijiensis* field isolates. Pest Management Science 56:833-841.

Stergiopoulos, I., Van Nistelrooy, J.G.M., Kema, G.H.J. and De Waard, M.A. 2003. Multiple mechanisms account for variation in base-line sensitivity to azole fungicides in field isolates of *Mycosphaerella graminicola*. Pest Management Sci. 59:1333-1343.

Ware, S. 2006. Aspects of Sexual Reproduction in *Mycosphaerella* Species on Wheat and Barley. Wageningen University and Research Center. PhD thesis, 190p.

Wittenberg, A.H.J., 2006. Genetic Mapping using Diversity Array Technology (DArT). PhD thesis Wageningen University. 223p.

Zhan, J., Kema, G.H.J., Waalwijk, C. and McDonald, B.A. 2002. Distribution of mating type alleles in the wheat pathogen *Mycosphaerella graminicola* over spatial scales from lesions to continents. Fungal Genet. and Biol. 26:128-136.