Nanopore sequencing of genomic DNA from *Magnaporthe oryzae* isolates from different hosts

Joe Win[1], Emilie Chanclud[1,2], C. Sarai Reyes-Avila[1],
Thorsten Langner[1], Tofazzal Islam[3], Sophien Kamoun[1]

[1]The Sainsbury Laboratory, University of East Anglia, Norwich Research Park, Norwich, United Kingdom
[2]Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, Switzerland
[3]Department of Biotechnology, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur, Bangladesh

We report long-range sequencing of eight isolates of *Magnaporthe oryzae* (Syn. *Pyricularia oryzae*) from wheat, rice, foxtail millet and goosegrass using nanopore MinION. Our aim is to obtain chromosome-level genome assemblies that are freely available for public access to be scrutinized for genome rearrangements and structural variation.

Magnaporthe oryzae (Syn. *Pyricularia oryzae*) is a notorious fungus known for causing blast disease on rice and wheat with devastating effect on grain yield. *M. oryzae* host range includes several other cereal crops such as oat, finger millet and foxtail millet as well as wild grasses. *M. oryzae* is found all over the world wherever warm temperature and high humidity are common. Although lineages of *M. oryzae* tend to be adapted to a particular host, this pathogen can shift from one host to another when the conditions permit (Couch *et al.* 2005). We hypothesize that structural variation contributes to adaptation to new hosts and environmental conditions following, for example, the model proposed by Chuma *et al.* (2011). To generate the genomics data that would enable such analyses, we sequenced eight isolates of *M. oryzae* from four different hosts using long-range nanonpore MinION with the aim of providing chromosome-level genome assemblies that are freely available for public access.

Results

We sequenced the genomes of four isolates of *M. oryzae* from different hosts (wheat, rice, foxtail millet and goosegrass) used in the GEMO project (Chiapello *et al.* 2015), and four isolates collected from two consecutive time points (2016 and 2017) from the wheat blast outbreak in Bangladesh (Islam *et al.* 2016). We used long-range sequencing based on nanopore technology and obtained N50 read lengths ranging from ~9,000 to 25,000 bp and total base counts of ~1.3 to 8.1 Gbp (~28x to 180x coverage) (Table 1). The sequence reads were assembled using Canu software (v1.6 and v1.7) (Koren *et al.* 2017) (Table 2). The number of contigs varied from 16 to 84 with the largest contigs in the ~10-11 Mbp range.

To assess the quality of the nanopore assemblies, we compared the rice isolate FR13 assembly to the only publicly available chromosome quality assembly of *M. oryzae* rice isolate 70-15 (Dean *et al.* 2005; Okagaki *et al*. 2015). We noted significant co-linearity between our FR13 assembly and that of 70-15 indicating that the nanopore assembly is of acceptable quality (Figure 1). Future quality control analyses will further determine the overall quality of this and the other genome assemblies we generated.

We deposited the raw reads and the assemblies in the European Nucleotide Archive and OpenWheatBlast website (Table 1-2).

Conclusions

The total lengths of the genome assemblies (41.5 to 46.4 Mbp) indicate near-complete sequencing of the *M. oryzae* isolates shown in Table 1. The lengths of the individual contigs were also long enough to enable structural variation studies. Although further quality controls are needed, we are optimistic that the assemblies are of sufficient quality to initiate structural variation analyses judging from the level of collinearity observed between our assembly of FR13 genome and the reference 70-15 genome. We have ensured open access to these genome data to inspire community involvement in analyzing these data. We hope that these

resources will contribute to tackling the blast disease of rice, wheat and other crops using cutting-edge genomic tools.

Materials and methods

M. oryzae isolates CD156, BR32, FR13 and US71 were acquired from Elisabeth Fournier as reported in the GEMO project (Chiapello *et al.* 2015). Bangladesh isolates BTJP4-1, BTMP13-1, BTGP1-b and BTGP6-f were acquired from Tofazzal Islam via the OpenWheatBlast project.

High molecular weight genomic DNA from *M. oryzae* was extracted from mycelia of 7-day old cultures by following the method described by (Schwessinger and Rathjen 2017). Genomic DNA was quantified on a TapeStation (Agilent) and treated with DNAse-free RNAse. RNAse-treated DNA was sheared using either a gTUBE or a 22 Gauge needle. Sheared DNA was captured using AMPure beads (Beckman Coulter, Indianapolis, US) and eluted in 45 µl water and used for library construction following the 1D protocol from Oxford Nanopore. Sequencing runs were performed using MinION R9.4 (Oxford Nanopore Technologies, Oxford, UK). Sequence reads were assembled into contigs using Canu software (v1.6 and v1.7) (Koren *et al.* 2017).

To validate the nanopore assembly method, we aligned the de novo nanopore assembly of isolate FR13 to the chromosome quality reference genome of isolate 70-15 (MG08) (Dean *et al.* 2005; Okagaki *et al*. 2015) using the NUCmer utility of MUMmer3 alignment program (Kurtz *et al.* 2004). We filtered the resulting coordinate output file for sequence alignments with a similarity >70% across >10 kB regions (delta-filter options -i 70 -l 10000). For visualization we used MUMmerplot (with the arguments -l and --color) followed by gnuplot to generate the output plot. We adjusted the color scale to display a similarity range between 80% and 100%.

Acknowledgements

This research was funded by The Gatsby Charitable Foundation and European Research Council BLASTOFF project. Nanopore sequencing and assembly services were provided by Future Genomics Technologies (Leiden, The Netherlands).

References

Chiapello H, Mallet L, Guerin C, Aguileta G, Amselem J, Kroj T, Ortega-Abboud E, Lebrun MH, Henrissat B, Gendrault A *et al.* 2015. Deciphering Genome Content and Evolutionary Relationships of Isolates from the Fungus Magnaporthe oryzae Attacking Different Host Plants. *Genome Biol Evol* **7**: 2896-2912.

Chuma I, Isobe C, Hotta Y, Ibaragi K, Futamata N, Kusaba M, Yoshida K, Terauchi R, Fujita Y, Nakayashiki H *et al.* 2011. Multiple translocation of the *AVR-Pita* effector gene among chromosomes of the rice blast fungus *Magnaporthe oryzae* and related species. *PLoS Pathog* **7**: e1002147.

Couch BC, Fudal I, Lebrun MH, Tharreau D, Valent B, van Kim P, Notteghem JL, Kohn LM. 2005. Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics* **170**: 613-630.

Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H *et al.* 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**: 980-986.

Islam MT, Croll D, Gladieux P, Soanes DM, Persoons A, Bhattacharjee P, Hossain MS, Gupta DR, Rahman MM, Mahboob MG *et al.* 2016. Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. *BMC Biol* **14**: 84.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**: 722-736.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol **5**: R12.

Okagaki LH, Nunes CC, Sailsbery J, Clay B, Brown D, John T, Oh Y, Young N, Fitzgerald M, Haas BJ, Zeng Q, Young S, Adiconis X, Fan L, Levin JZ, Mitchell TK, Okubara PA, Farman ML, Kohn LM, Birren B, Ma LJ, Dean RA. 2015. Genome sequence of Three Species of the Magnaporthaceae Family of fungi. G3 28;5(12):2539-46

Schwessinger B, Rathjen JP. 2017. Extraction of High Molecular Weight DNA from Fungal Rust Spores for Long Read Sequencing. *Methods Mol Biol* **1659**: 49-57.

Table 1. *M. oryzae* isolates used in this study and summary statistics for their nanopore sequencing runs

M. oryzae isolate	Host	Source	Year collected	Reference	Flowcell ID[1]	Run type	Basecaller	# Reads	Base count (Gbp[2])	Read N50 (bp)	ENA Accession[3]
BTJP4-1	*Triticum aestivum*	Jhenaidah, Bangladesh	2016	Islam et al., 2016	FAH45397	1D	Albacore v2.1.10	464,107	3.1	9,437	ERR2612748
BTMP13-1	*Triticum aestivum*	Meherpur, Bangladesh	2016	Islam et al., 2016	FAH37911	1D	guppy v0.3.0	181,812	1.4	25,365	ERR2612753
BTGP1-b	*Triticum aestivum*	Meherpur, Bangladesh	2017	Islam et al., 2016	FAH43895	1D	Albacore v2.1.10	158,055	1.3	11,213	ERR2612754
BTGP6-f	*Triticum aestivum*	Meherpur, Bangladesh	2017	Islam et al., 2016	FAH43895	1D	Albacore v2.1.10	284,212	1.9	9,448	ERR2612755
BR32	*Triticum aestivum*	Brazil	1990	Urashima et al., 1999	FAH38591	1D	guppy v0.3.0	655,377	5.5	11,021	ERR2612751
FR13	*Oryza sativa*	France	1988	Chiapello et al., 2015; Gladieux et al., 2018	FAH37081	1D	guppy v0.5.1	1,237,702	8.1	10,859	ERR2612749
US71	*Setaria italica*	United States	1998	Chiapello et al., 2015; Gladieux et al., 2018	FAH36864	1D	guppy v0.5.1	661,505	4.9	10,352	ERR2612750
CD156	*Eleusine indica*	Ferkessedougou, Ivory Coast	1989	Chiapello et al., 2015; Gladieux et al., 2018	FAH45397	1D	Albacore v2.1.10	566,528	3.5	9,234	ERR2612752

[1]Sequencing was performed by Future Genomics Technologies
[2]1 Gbp = 1,000,000,000 base pairs
[3]Sequence reads were deposited at European Nucleotide Archive (ENA) with study accession PRJEB27137

Table 2. Summary statistics for *M. oryzae* genomes assembled from nanopore reads

M. oryzae isolate	Host	CANU version[1]	# Contigs	Assemby length (bp)	N25 (bp)	N50 (bp)	N75 (bp)	Max length (bp)	Mean length (bp)	Min length (bp)	GenBank Accession[2]
BTJP4-1[3]	*Triticum aestivum*	1.7	59	44,506,712	6,840,169	4,344,896	3,373,527	7,174,201	754,351	13,054	GCA_900474225.2
BTMP13-1[3]	*Triticum aestivum*	1.6	16	43,978,087	7,837,192	6,037,509	4,385,994	10,783,101	2,748,630	7,390	GCA_900474375.2
BTGP1-b[4]	*Triticum aestivum*	1.7	74	44,406,102	3,690,742	2,814,025	1,269,883	6,505,875	600,082	5,533	GCA_900474635.2
BTGP6-f[4]	*Triticum aestivum*	1.7	57	44,234,333	5,243,043	3,705,381	2,027,069	6,048,575	776,041	8,312	GCA_900474435.2
BR32	*Triticum aestivum*	1.6	21	41,471,325	11,366,628	5,047,693	3,895,412	11,366,628	1,974,825	18,099	GCA_900474545.2
FR13	*Oryza sativa*	1.7	46	46,415,940	6,634,785	5,357,033	2,121,955	7,257,380	1,009,042	19,712	GCA_900474655.2
US71	*Setaria italica*	1.7	84	45,673,611	3,535,243	2,015,667	979,014	4,788,334	543,733	6,882	GCA_900474175.2
CD156	*Eleusine indica*	1.7	44	43,859,562	6,040,961	4,257,479	3,430,372	6,066,300	996,808	8,777	GCA_900474475.2

[1]Genome assembly was performed by Future Genomics Technologies using Canu assembly software
[2]Sequence assemblies were deposited at European Nucleotide Archive (ENA) with study accession PRJEB27137
[3]*M. oryzae* isolates collected from wheat during 2016 epidemic in Bangladesh
[4]*M. oryzae* isolates collected from wheat during 2017 epidemic in Bangladesh

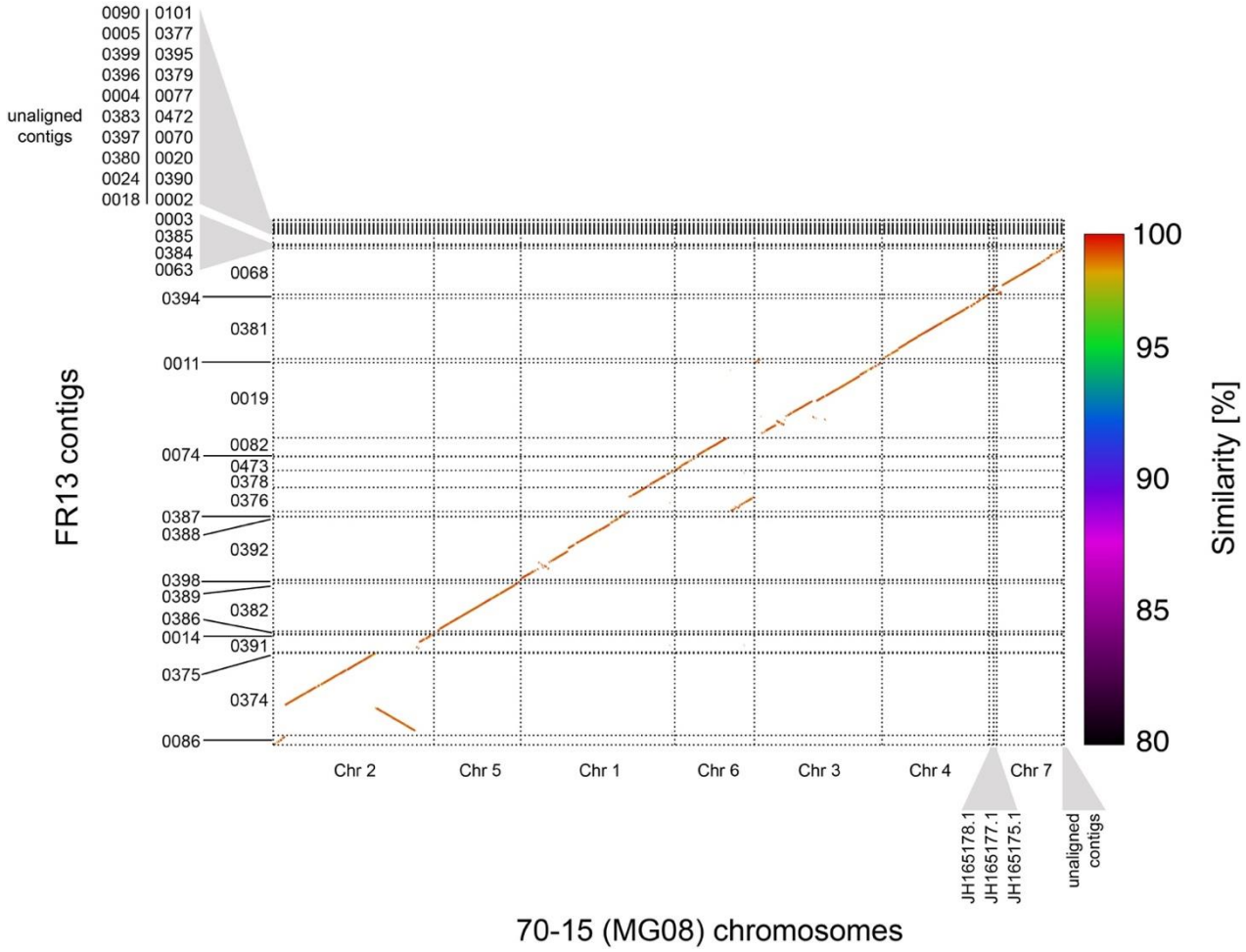

Figure 1. Genome assemblies of *M. oryzae* rice isolate FR13 and the reference genome of 70-15 are overall colinear. Collinearity plot of the genomes of strains FR13 and 70-15. Nanopore contigs of FR13 (y-axis) were aligned and plotted against chromosomes of the reference genome of 70-15 (x-axis). The orange coloured line indicates an overall sequence conservation of ~98% across the genomes.