Genome Evolution of fungal pathogens from *Magnaporthe oryzae/grisea* clade

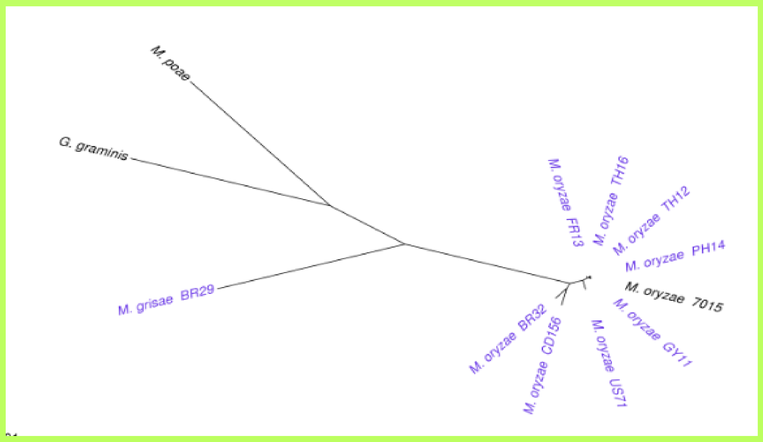
M-H Lebrun (1), L. Mallet (2), C. Guerin (2), H. Chiapello (2), E. Ortega-Abboud (3), A. Gendrault (2), J. Kreplak (4),T. Kroj (3), A. Couloux (5), C. Cruaud (5), J. Amselem (4), D. Tharreau (3), E. Fournier (3)

1-INRA, UR BIOGER, AgroParisTech, Thiverval-Grignon, France,
2-INRA, UR MIG, 78352 Jouy-en-Josas, France,
3-INRA-CIRAD, UMR BGPI, 34398 Montpellier, France
4-INRA, UR URGI, 78026 Versailles, France,
5-Genoscope, 91 507 Evry, France
marc-henri.lebrun@versailles.inra.fr

Introduction

Magnaporthe oryzae/grisea **species complex is composed of host specific forms pathogenic on different Poaceaes. This complex includes a subgroup composed of closely related isolates pathogenic on rice as well as a subgroup of isolates responsible of epidemics on wheat in South America. The GEMO project aims at characterizing genes and evolutionary events involved in the adaptation of these different related pathogens to their specific host plant using comparative genomics. Such host adaptation may rely on variations in gene or transposon content or location, as well as modifications in coding and regulatory sequences. Eight strains from *M. oryzae* pathogenic on either rice, wheat, *Setaria* or *Eleusine* and one strain of *M. grisea* pathogenic on *Digitaria,* were sequenced using NGS technologies (see Figure 1).**

Figure 1: Phylogenetic relationships among *M. grisea/oryzae* isolates estimated by whole genome comparison



Results, Glogal genome comparisons

De novo **annotation was carried out with Eugene for genes and with REPET for transposons (TEs). Most frequent TE families are LTR retro-transposons, but DNA transposon families were found. TEs cover about 10-12% of these genomes.**

Differences in genome size and gene content (12 300-20 500 genes) were observed between isolates (see Table 1). Gene number was overestimated in 4 fragmented genomes (FR13, GUY11, TH12, PH14) with poor scaffolding (short length and truncated CDS). However, significant differences were found also between well-assembled genomes (39-42 Mb, 12.616 – 14.013 genes), suggesting a variability in genome organisation.

Table 1 *Magnaporthe* genome features

Genomes	*70-15	†BR29	BR32	CD156	FR13	GY11	TH12	PH14	TH16	US71
Main host	rice	Digitaria	wheat	Eleusine	rice	rice	rice	rice	rice	Setaria
Size (Mb)	40.9	40.9	41.9	42.7	43.1	46.3	48.5	49.8	39.1	41.2
GC%	51.51	47.88	48.29	47.6	39.81	49.1	50.33	47.87	48.55	48.31
Nb contigs	216	9,644	6,044	26,535	79,619	13,188	9,908	11,772	4,114	7,398
Nb scaffolds	53	169	111	237	2,051	1,964	940	711	171	220
Depth coverage (x)	‡ -	60	55	50	40	42	48	56	53	80
Scaff. N50 (kb)	6,607	955	1,760	1,066	101	187	590	597	938	813
%N	0.18	4.09	4,96	6,59	22.55	7.98	5.83	10.23	5.80	5.45
% Unannotated DNA	37.0	52.2	42.2	43.5	54.1	39.7	38.1	41.0	41.3	42.6
Predicted genes	12,827	12,616	14,781	14,415	15,035	20,621	19,811	20,067	13,725	14,013
Mono-exonic genes	2,507	3,755	4,184	4,268	3,094	6,122	7,297	6,695	3,8	2,376
CDS ≤ 300bp	1,274	1,537	2,541	2,473	3,095	3,58	2,827	3,029	2,296	4,172
Accessory genes	3,635	2,632	5,53	5,269	5,344	10,773	10,548	10,502	5,112	5,092
Specific genes	968	1,506	665	644	1,402	1,55	772	1,21	305	557
SSPs, Signal P detect	1,456	1,391	1,646	1,627	1,451	1,751	1,847	1,832	1,531	1,579
No SP, 1 N-Ter TMD	237	250	269	277	317	406	363	374	252	276
Total SSPs	1,693	1,641	1,915	1,904	1,768	2,157	2,21	2,206	1,783	1,855
% of genes	13.03	13.35	13.35	13.54	11.87	10.55	11.35	11.09	13.14	13.44

SSP Small secreted proteins (300 aa <)

‡,* The reference strain *M.o.* 70-15 is sequenced with Sanger technology. † BR29 is a strain from the *M. grisea* species.

Results, Secreted proteins

Putative secreted proteins (SPs) were detected with Signal-P. SPs have a median length of 260 aa (see Figure 2).

Small secreted proteins (300 aa<) correspond to 10,5 to 13,3 % of the CDS. Most of them do not encode for secreted enzymes (CAZY, protease, lipase) and may behave as effectors.

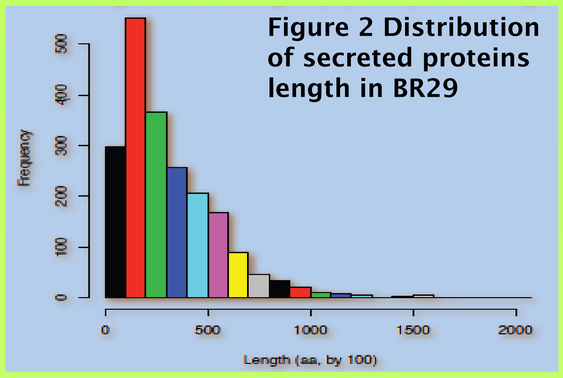
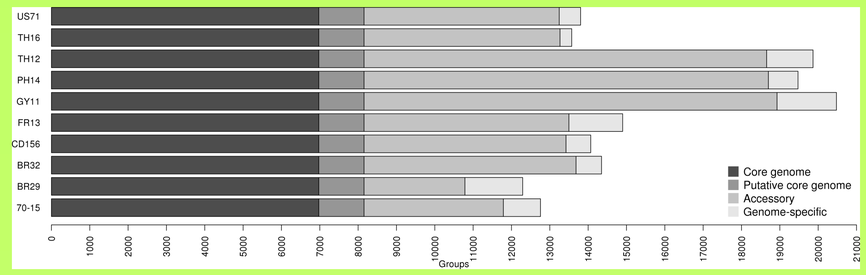


Figure 2 Distribution of secreted proteins length in BR29

Results, Gene families

OrthoMCL analysis including 70-15 genome identified 20.443 clusters of related genes (families), including 8.154 single copy gene families shared by all isolates (core genome), families shared only by some isolates (accessory) and genome-specific gene families (305-1550, see Figure 3).

Figure 3 Distribution of gene families detected by OMCL in different *Magnaporthe* genomes



Conclusion

A significant variability in genome size and gene content was observed between *Magnaporthe oryzae/grisea* isolates pathogenic on different host plants. Characterization of accessory and genome-specific genes and comparison of genome organization (synteny, mobile elements) are expected to facilitate identifying genes and evolutionary events associated with adaptation of *Magnaporthe* to different host plants.