Comparative Analysis of Twelve Dothideomycete Plant Pathogens

Robin Ohm[1], Andrea Aerts[1], Asaf Salamov[1], Steven Goodwinr[2] and Igor Grigoriev[1]

[1]DOE Joint Genome Institute and [2] USDA ARS Purdue University Campus

[1]*To whom correspondence may be addressed. E-mail:* raohm@lbl.gov.

March 14, 2011

ACKNOWLEDGMENTS:

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Additionally, we wish to thank and acknowledge the following for their efforts: Rosie Bradshaw, Lynda Ciuffetti, Richard Hamelin, Barbara Howlett, Gert Kema, Christopher Lawrence, Richard Oliver, Joey Spatafora & Gillian Turgeon.

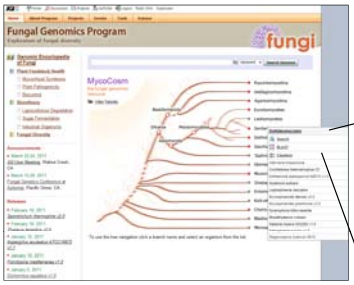
Comparative analysis of twelve Dothideomycete plant pathogens

Robin A. Ohm[1], Andrea Aerts[1], Asaf Salamov[1], Stephen B. Goodwin[2], Igor V. Grigoriev[1]
raohm@lbl.gov
[1] DOE - Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
[2] USDA-ARS, Crop Production and Pest Control Research Unit, 915 West State Street, Purdue University campus, West Lafayette, IN 47907-2054, USA

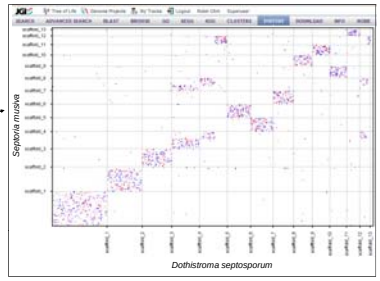
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Introduction

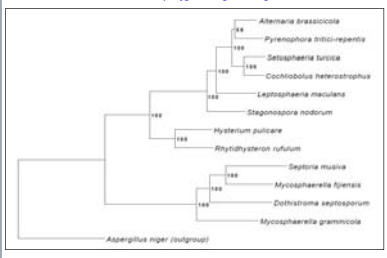
The Dothideomycetes are one of the largest and most diverse groups of fungi. Many are plant pathogens and pose a serious threat to agricultural crops grown for biofuel, food or feed. Most Dothideomycetes have only a single host and related Dothideomycete species can have very diverse host plants. Twelve Dothideomycete genomes have currently been sequenced by the Joint Genome Institute and other sequencing centers. They can be accessed via Mycocosm which has tools for comparative analysis.



Mycocosm. The web portal Mycocosm contains the genomes of all 12 sequenced Dothideomycetes, as well as 51 other fungal genomes sequenced by the JGI and other sequencing centers. Organism-specific and comparative tools are available to the user on http://jgi.doe.gov/fungi

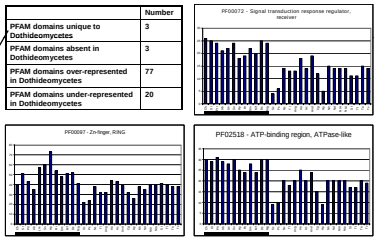


Mesosynteny. Dotplot of pairwise genome comparisons. Many intra-chromosomal, but few inter-chromosomal rearrangements have occurred during evolution. This observation is called 'mesosynteny' and is observed in all 12 Dothideomycetes. The mechanism behind it is unknown.



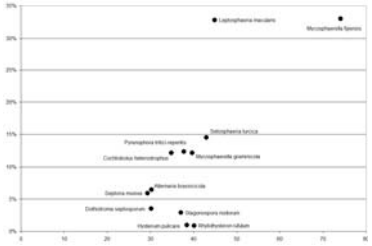
Phylogenetic tree of the 12 species used in this comparative study. The tree was calculated with the program RAxML using 80 proteins that are present with one ortholog in each genome. Bootstrap values are indicated at the nodes.

Unique cluster in the Dothideomycetes. An example of an MCL cluster that is unique to the Dothideomycetes when compared to 18 other Ascomycetes. The proteins contain a Bin3 methyltransferase domain. There is a strong conservation of gene order, as indicated on the right (homologous genes have identical colors)



	Number
PFAM domains unique to Dothideomycetes	3
PFAM domains absent in Dothideomycetes	3
PFAM domains over-represented in Dothideomycetes	77
PFAM domains under-represented in Dothideomycetes	20

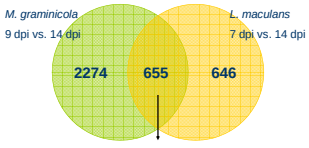
Protein families in the Dothideomycetes compared to 18 other Ascomycetes. Unique and over-represented PFAM domains are involved in a wide variety of functions. Examples from signal transduction, transcriptional regulation and metabolism are shown. The 12 Dothideomycetes are indicated with a black bar.



Repeat content. There is a wide variety in genome size and repeat content among the Dothideomycetes. In *M. fijiensis* and *L. maculans*, massive invasion by transposable elements has taken place.

Species	*Cochliobolus heterostrophus*	*Mycosphaerella graminicola*	*Mycosphaerella fijiensis*	*Pyrenophora tritici-repentis*	*Stagonospora nodorum*	*Alternaria brassisicola*
Host plant	Maize	Wheat	Banana	Grass	Wheat	Brassica species
Sequencing center	JGI	JGI	JGI	BROAD	BROAD	BROAD

Species	*Dothistroma septosporum*	*Septoria musiva*	*Setosphaeria turcica*	*Hysterium pulicare*	*Rhytidhysteron rufulum*	*Leptosphaeria maculans*
Host plant	Pine tree	Poplar tree	Maize	Saprotroph	Saprotroph	Oil seed rape plants
Sequencing center	JGI	JGI	JGI	Oregon State University	Oregon State University	URGI-INRA



	Orthologous pairs	Up-regulated during pathogenesis in both species
All orthologous gene pairs	5822	655
Secondary metabolites biosynthesis, transport and catabolism (KOG)	293	83
P450 domain (PFAM)	37	14
Dienelactone hydrolase (KOG, PFAM)	10	7
Aldo/keto reductase (KOG)	27	12
Various reductases (KOG, PFAM)	366	70
DNA photolysase - FAD-binding (PFAM)	24	10
Alcohol dehydrogenase (PFAM, KOG)	72	24
Glutathion-S-transferase (PFAM)	20	9

Comparative transcriptomics. Microarray data is available for *M. graminicola* (pathogen of wheat, Keon et al. 2005 and 2007) and *L. maculans* (pathogen of oil seed rape plants, Rouxel et al. 2011). In both cases gene expression was analyzed during early and late stage of infection, allowing comparative analysis. There are 17 annotation terms that were over-represented in this group ($p < 10^{-5}$), meaning that they may be involved in the pathogenesis process.

Mycosphaerella graminicola

Chromosome	GC-content (%)	Repeat content (%)	Gene density (genes / Mbp)	Proteins with PFAM domain (%)
Whole genome	52%	18%	276	57%
Chr_14	48%	37%	147	22%

Pyrenophora tritici-repentis

Chromosome	GC-content (%)	Repeat content (%)	Gene density (genes / Mbp)	Proteins with PFAM domain (%)
Whole genome	51	16	321	53%
Supercontig_1.19	50%	40%	157	38%
Supercontig_1.20	49%	38%	154	34%

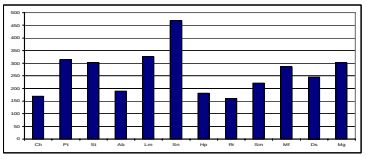
Leptosphaeria maculans

Chromosome	GC-content (%)	Repeat content (%)	Gene density (genes / Mbp)	Proteins with PFAM domain (%)
Whole genome	45	35	278	56
Lm_Super-contig_22_v2	35	90	48	14

Mycosphaerella fijiensis

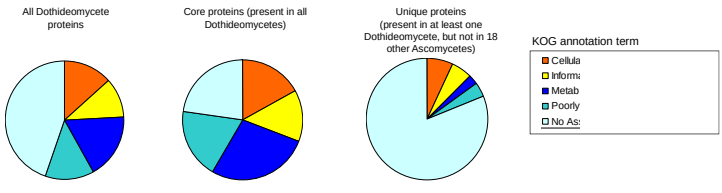
Chromosome	GC-content (%)	Repeat content (%)	Gene density (genes / Mbp)	Proteins with PFAM domain (%)
Whole genome	45	33	147	60
Scaffold_11	40	65	35	13
Scaffold_13	41	63	45	12

Dispensable chromosomes. *M. graminicola* has experimentally been shown to have chromosomes that are dispensable (not necessary for survival). These chromosomes have lower GC content, higher repeat content, lower gene density and a lower percentage of proteins with a PFAM domain. As an example, chromosome 14 is shown. Similar chromosomes (or scaffolds) have been computationally identified in other Dothideomycetes. Examples are shown here. Their dispensability will have to be confirmed in the lab.



Small Secreted Proteins (SSPs). SSPs have been implicated in fungus-plant interactions in several cases, for example during pathogenesis or ectomycorrhizal symbiosis. Within the Dothideomycetes there are large differences in numbers of SSPs. The definition of SSP that was used here is < 200 amino acids, presence of a secretion signal and absence of a transmembrane domain.

	All proteins	SSPs	Conclusion
Cysteine residues (%)	1.2	2.8	SSPs are cysteine-rich
PFAM domain (%)	49.8	11.8	Few SSPs have PFAM domains
In singleton MCL clusters (%)	17%	43%	SSPs are poorly conserved. Does this account for host-specificity?



Protein families in the Dothideomycetes. A large part of the Dothideomycete protein families (as determined by MCL clustering) has no known function. The core protein family set (having at least one member in all Dothideomycetes) has much more KOG annotation terms. Protein families that are unique to the Dothideomycetes are very poorly described, showing that a lot remains to be learned about the Dothideomycetes.

Conclusions

- Genome size and repeat content vary widely in the twelve Dothideomycete genomes which are now available via Mycocosm.
- Many intra-chromosomal, but few inter-chromosomal rearrangements have taken place during Dothideomycete evolution.
- Several potentially dispensable chromosomes have been identified, similar to the ones in *Mycosphaerella graminicola*
- Small Secreted Proteins (SSPs) are found in varying numbers across the Dothideomycetes.
- Many protein families unique to or over-represented in the Dothideomycetes have been identified.
- Comparative transcriptomics gives insight into conserved fungal responses during pathogenesis, leading to new targets to fight infections.

Acknowledgements: Rosie Bradshaw, Lynda Ciuffetti, Steve Goodwin, Richard Hamelin, Barbara Howlett, Gert Kema, Christopher Lawrence, Richard Oliver, Joey Spatafora & Gillian Turgeon

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.