Diverse Lifestyles and Strategies of Plant Pathogenesis Encoded in the Genomes of Eighteen *Dothideomycetes*

Robin A. Ohm[1]*, Nicolas Feau[2], Bernard Henrissat[3], Conrad L. Schoch[4], Benjamin A. Horwitz[5], Kerrie W. Barry[1], Bradford J. Condon[6], Alex C. Copeland[1], Braham Dhillon[2], Fabian Glaser[7], Cedar N. Hesse[8], Idit Kosti[5,7], Kurt LaButti[1], Erika A. Lindquist[1], Susan Lucas[1], Asaf A. Salamov[1], Rosie E. Bradshaw[9], Lynda Ciuffetti[8], Richard C. Hamelin[2], Gert H. J. Kema[10], Christopher Lawrence[11], James A. Scott[12], Joseph W. Spatafora[8], B. Gillian Turgeon[6], Pierre J.G.M. de Wit[13], Shaobin Zhong[14], Stephen B. Goodwin[15], Igor V. Grigoriev[1]

1. Joint Genome Institute; 2. University of British Columbia, Canada; 3. Aix-Marseille Université, France; 4. NIH/NLM/NCBI; 5. Technion - IIT, Israel; 6. Cornell University; 7. BKU - Technion - IIT, Israel; 8. Oregon State University; 9. Massey University, New Zealand; 10. Plant Research International, The Netherlands; 11. Department of Biological Sciences, Blacksburg; 12. Dalla Lana School of Public Health, Canada; 13. Wageningen University, The Netherlands; 14. North Dakota State University; 15. Purdue University

March 2013

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

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

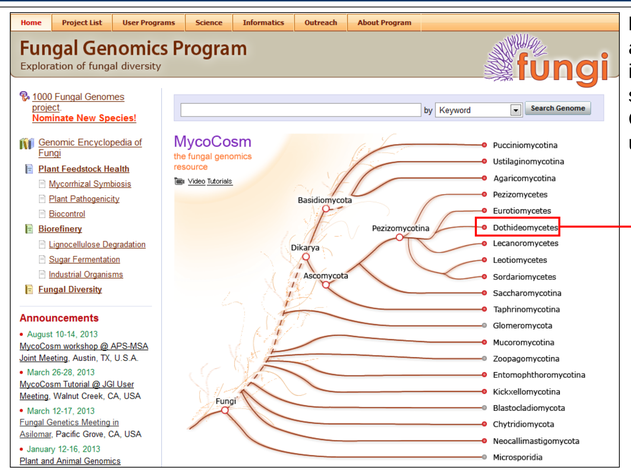
Diverse Lifestyles and Strategies of Plant Pathogenesis Encoded in the Genomes of Eighteen *Dothideomycetes*

Robin A. Ohm[1]*, Nicolas Feau[2], Bernard Henrissat[3], Conrad L. Schoch[4], Benjamin A. Horwitz[5], Kerrie W. Barry[1], Bradford J. Condon[6], Alex C. Copeland[1], Braham Dhillon[2], Fabian Glaser[7], Cedar N. Hesse[8], Idit Kosti[5,7], Kurt LaButti[1], Erika A. Lindquist[1], Susan Lucas[1], Asaf A. Salamov[1], Rosie E. Bradshaw[9], Lynda Ciuffetti[8], Richard C. Hamelin[2], Gert H. J. Kema[10], Christopher Lawrence[11], James A. Scott[12], Joseph W. Spatafora[8], B. Gillian Turgeon[6], Pierre J.G.M. de Wit[13], Shaobin Zhong[14], Stephen B. Goodwin[15], Igor V. Grigoriev[1]

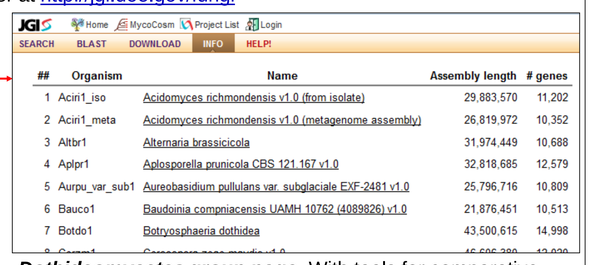
* Email: raohm@lbl.gov

Introduction

The class of *Dothideomycetes* is one of the largest and most diverse groups of fungi. Many are plant pathogens and pose a serious threat to agricultural crops that are grown for biofuel, food or feed. Most *Dothideomycetes* have only a single host plant, and related species can have very diverse hosts. Eighteen genomes of *Dothideomycetes* have currently been sequenced by the Joint Genome Institute and other sequencing centers. Here we describe the results of comparative analyses of the fungi in this group.



Mycocosm. The web portal MycoCosm contains the genomes and annotations of 34 *Dothideomycetes* (including the 18 used in this study), as well as over 150 other fungal genomes, sequenced by the JGI and other sequencing centers. Organism-specific and comparative tools are available to the user at http://jgi.doe.gov/fungi

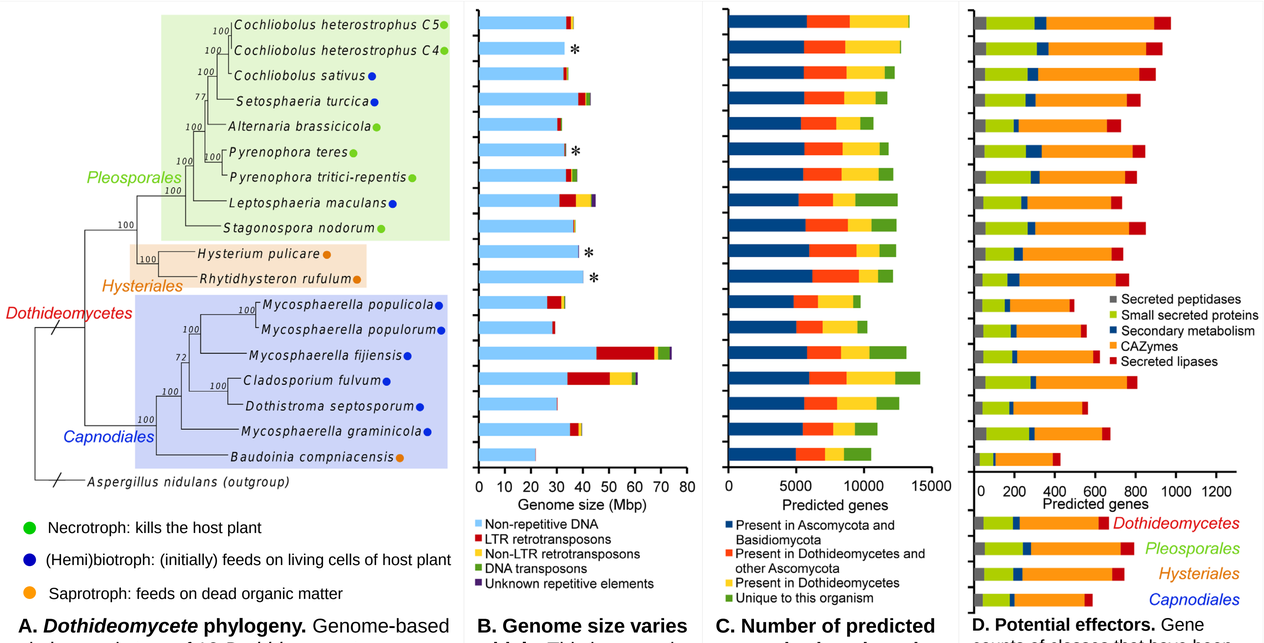


##	Organism	Name	Assembly length	# genes
1	Aciri1_iso	Acidomyces richmondensis v1.0 (from isolate)	29,883,570	11,202
2	Aciri1_meta	Acidomyces richmondensis v1.0 (metagenome assembly)	26,819,972	10,352
3	Altbr1	Alternaria brassicicola	31,974,449	10,688
4	Aplpr1	Aplosporella prunicola CBS 121.167 v1.0	32,818,685	12,579
5	Aurpu_var_sub1	Aureobasidium pullulans var. subglaciale EXF-2481 v1.0	25,796,716	10,809
6	Bauco1	Baudoinia compniacensis UAMH 10762 (4089826) v1.0	21,876,451	10,513
7	Botdo1	Botryosphaeria dothidea	43,500,615	14,998

***Dothideomycetes* group page.** With tools for comparative analysis

Effector genes involved in infection

Lifestyles are largely phylogenetically separated. This is also reflected in the gene classes that are potentially involved in plant pathogenesis (effectors). Necrotrophs and saprotrophs generally have more effector genes than (hemi)biotrophs.



- Necrotroph: kills the host plant
- (Hemi)biotroph: (initially) feeds on living cells of host plant
- Saprotroph: feeds on dead organic matter

A. *Dothideomycete* phylogeny. Genome-based phylogenetic tree of 18 *Dothideomycetes* computed using 51 conserved protein families. Different lifestyles are largely phylogenetically separated. For example, all necrotrophs are Pleosporales, all Hysteriales are saprotrophs, and most Capnodiales are (hemi)biotrophs.

B. Genome size varies widely. This is caused mainly by differences in repetitive content. Asterisks indicate Illumina-sequenced genomes. In these genomes repeat content is likely an under-estimate.

C. Number of predicted genes, broken down by level of conservation.

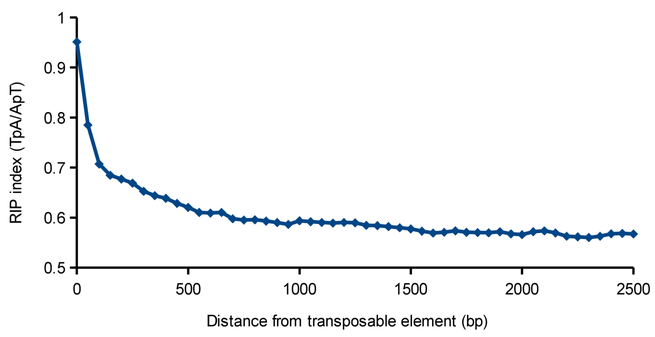
D. Potential effectors. Gene counts of classes that have been implicated in plant infection. Members of *Capnodiales* generally have fewer genes in these classes than *Pleosporales* and *Hysteriales*. This trend is also illustrated by the estimated gene counts for the last common ancestors of the indicated taxa (below the x-axis), which correspond to the taxa in (A)

Effector genes near Transposable Elements are subjected to RIP

Some potential effector gene classes are over-represented near Transposable Elements. There, they are subjected to point mutations caused by the RIP machinery (Repeat-Induced Point mutations). This may speed up their evolution, allowing the fungi to evade the host plant's defenses.

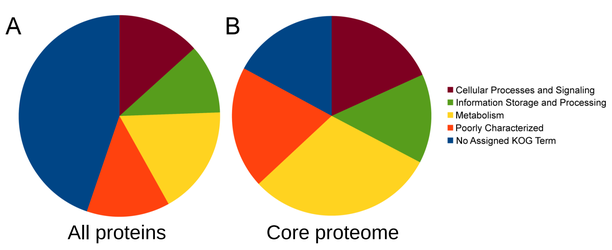
	TE repeat content (%)	Over-representation of gene classes in repeat regions: Small secreted proteins	All secreted proteins	Secondary metabolism	Expanded orphan multi-gene families
Alternaria brassicicola	5.58			Yes	
Baudoinia compniacensis	0.4				
Cladosporium fulvum	44.24				
Cochliobolus heterostrophus C5	7.77	Yes	Yes		Yes
Cochliobolus heterostrophus C4 (*)	0				
Cochliobolus sativus	5.44	Yes		Yes	Yes
Dothistroma septosporum	0.67				
Hysterium pulicare (*)	0.57				
Leptosphaeria maculans	30.93	Yes	Yes	Yes	Yes
Mycosphaerella fijiensis	38.97		Yes		
Mycosphaerella graminicola	11.66	Yes			Yes
Mycosphaerella populicola	20.81			Yes	
Mycosphaerella populorum	3.56	Yes	Yes		Yes
Pyrenophora tritici-repentis	11.44	Yes	Yes	Yes	Yes
Pyrenophora teres f. teres (*)	1.98				Yes
Rhytidhysteron rufulum (*)	0.18				
Setosphaeria turcica	11.16	Yes	Yes	Yes	Yes
Stagonospora nodorum	2.37		Yes		Yes

Potential effectors that are over-represented near Transposable Elements (TEs). Asterisks indicate Illumina-sequenced genomes. In these genomes TE content is likely an under-estimate.



Repeat-Induced Point mutations (RIP) occur in genes near Transposable Elements (TEs). TEs are a target for the RIP machinery, which causes point mutations in the TE to inactivate it. These point mutations also occur near these TEs (Higher RIP index means more point mutations).

Core proteome

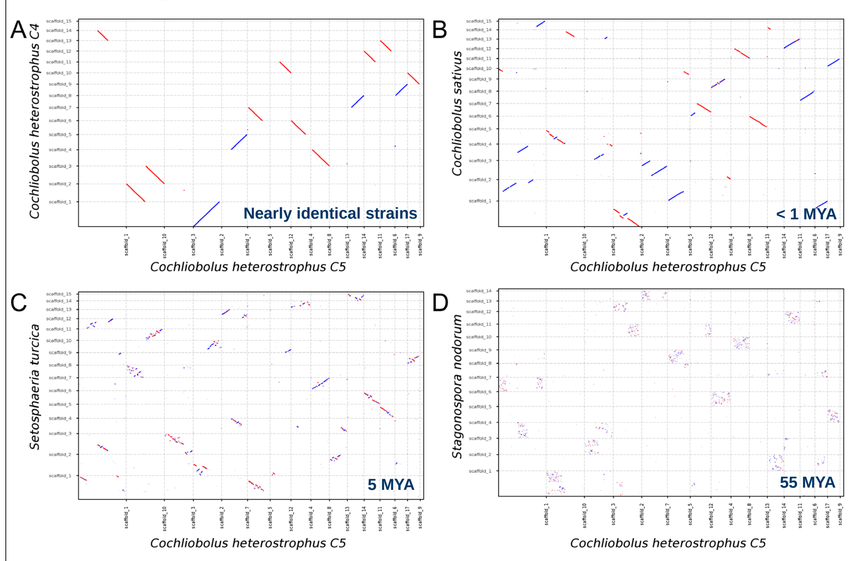


A. The full proteome of the *Dothideomycetes* contains 215225 proteins, and for the majority of these the function according to KOG is unknown or poorly characterized. **B.** The core proteome contains the 66761 proteins from multi-gene families that had at least one member in each *Dothideomycete*. Relative to (A), this set of proteins has more KOG annotations than the full proteome. In particular genes involved in metabolism are over-represented.

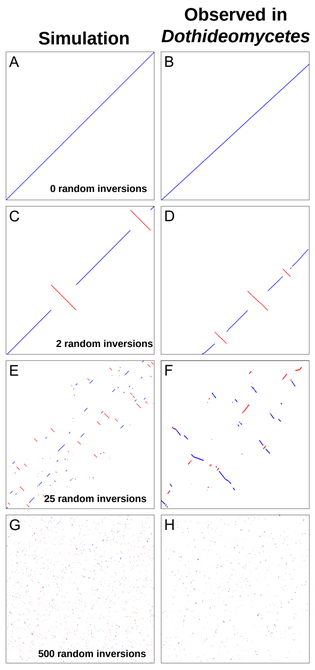
Comparative transcriptomics

- Microarray data are available for *M. graminicola* (Keon et al. 2005 and 2007) and *L. maculans* (Rouxel et al. 2011).
- In both cases gene expression was analyzed during early and late stage of infection, allowing comparative analysis.
- 98 orthologous gene-pairs are up-regulated in both organisms during pathogenesis.
  - 24 are oxidoreductases
  - 3 are transcription factors
- Because these 98 genes have a conserved expression responses during infection, they may very well be involved in the pathogenesis process.
- Currently, a CSP project is in progress to analyze gene expression in *Dothideomycetes* and their hosts during infection.

Whole-genome DNA comparison



From macrosynteny to mesosynteny. Closely related species show a pattern of nearly perfect macrosynteny (A). This pattern degrades in comparisons with progressively distantly related species (B-D). The syntenic regions become short and spread across scaffold pairs. This phenomenon is most predominant among *Dothideomycetes*. It has previously been called mesosynteny (Hane et al. 2011). Here we show that mesosynteny appears to be due to intra-chromosomal inversions. Very few inter-chromosomal rearrangements take place. Interestingly, simple repeats (i.e. low-complexity DNA) are over-represented near inversion breakpoints (data not shown).

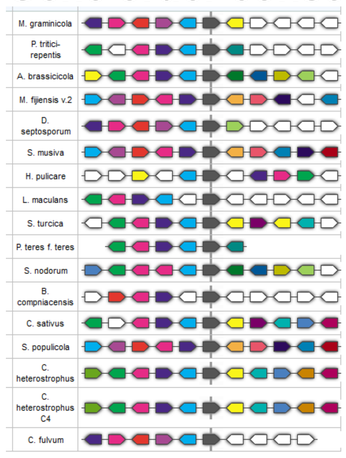


Simulations show that intra-chromosomal inversions lead to a pattern of mesosynteny that is observed among *Dothideomycetes*.

Dispensable chromosomes

Several *Dothideomycetes* have (putatively) dispensable chromosomes. Eight of the 21 chromosomes of *Mycosphaerella graminicola* are dispensable (i.e. not necessary for survival). These dispensable chromosomes are smaller, less gene-dense and more repeat-rich than the core chromosomes. Proteins encoded by genes on these chromosomes less frequently contain a PFAM domain. Their function is unknown (Goodwin *et al*. 2011). Scaffolds with similar characteristics are also present in five other *Dothideomycetes*: *Mycosphaerella fijiensis, Cochliobolus heterostrophus C5, Setosphaeria turcica, Leptosphaeria maculans,* and *Stagonospora nodorum*.

Gene order conservation



Only two syntenic blocks of genes are conserved in *Dothideomycetes*. Identical color means the genes are part of the same multi-gene family. The syntenic block of genes shown here contains mostly oxidoreductases and dehydrogenases. Interestingly, in *L. maculans* these 5 genes are all up-regulated during pathogenesis.

Conclusions

- The genomes of 18 *Dothideomycetes* (of which 15 are plant pathogens) are currently sequenced and available via MycoCosm.
- Necrotrophs and saprotrophs generally have more effector genes than (hemi)biotrophs.
- Some classes of effector genes are over-represented near Transposable Elements, where they are subjected to Repeat Induced Point mutations (RIP). This may speed up their evolution.
- During Dothideomycete evolution many intra-chromosomal inversions, but few inter-chromosomal rearrangements have taken place. This process may keep dispensable chromosomes intact.
- Comparative transcriptomics gives insight into conserved fungal responses during pathogenesis, leading to new targets to fight infections.

Affiliations
1. Joint Genome Institute; 2. University of British Columbia, Canada; 3. Aix-Marseille Université, France; 4. NIH/NLM/NCBI; 5. Technion - IIT, Israel; 6. Cornell University; 7. BKU - Technion - IIT, Israel; 8. Oregon State University; 9. Massey University, New Zealand; 10. Plant Research International, The Netherlands; 11. Department of Biological Sciences, Blacksburg; 12. Dalla Lana School of Public Health, Canada; 13. Wageningen University, The Netherlands; 14. North Dakota State University; 15. Purdue University

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.