## Lawrence Berkeley National Laboratory

**Recent Work** 

## Title

Comparative Genomics Tools for analysis of Six Dothideomycete Genomes

## Permalink

https://escholarship.org/uc/item/6tp259s3

## Authors

Aerts, Andrea Minovitsky, Simon Cantor, Michael <u>et al.</u>

# Publication Date

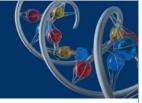
2009-03-17



## Comparative Genomics Tools For Analysis of Six **Dothideomycete Genomes**

Andrea Aerts\*, Simon Minovitsky, Mike Cantor, Igor Ratnere, Inna Dubchak, Igor Shabalov, Igor Grigoriev US DOE Joint Genome Institute, Walnut Creek, CA 94598

\*Correspondence: ALAerts@lbl.gov



Mier fo M

Mar 5 .m

#### \*Abstract

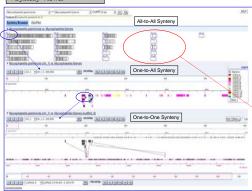
Comparative genomics is a powerful tool for genome annotation and analysis. We recently initiated analysis of six of the Dothideomycete genomes that are publicly available: Mycosphaerella graminicola, Mycosphaerella fijiensis and Cochliobolus heterostropus C5 all sequenced and annotated at the DOE Joint Genome Institute as well as Alternaria brassicicola sequenced at the Washington University School of Medicine, Genome Sequencing Center and annotated by the Christopher Lawrence Lab at Virginia Bioinformatics Institute, Virginia Tech, Pyrenophora triticirepentis and Stagonospora nodorum, both sequenced and annotated by the Broad Institute, MIT. We placed these genomes into a comparative framework equipped with tools to facilitate genome analysis: VISTA genome conservation curves linking genome browsers, gene cluster browser and protein links to homologs in all six genomes, side-by-side comparisons of KEGG pathways and KOG functions, and synteny viewer.

JGI community annotation brought biologists behind each of these genomes together to explore common mechanisms of plant pathogenicity and host interactions.

#### VISTA Conservation Analysis

JGI			lla graminicol Finished Ge	nome			
The real provider of the							
L	e   0   00   00			-1.5+   -3+   -10+		AA 1	NA ]
623	arread area	and Decent	and one		Terrori		
Bare Position	There are		Second Second and Second	Encod Money	(accord)		3000
S62 outdWillmain	An	the end at	star	M	k		alla A
and pointed internation.		11	The Calchibidiate Instantion	franka (\$ progratia)	100	A.A.	109 4
adaine Jereine	48	NAK MA A	м	4	h	Δ	1
and workshowned	111132		VIII - Harrowski - Ha	interne conservation	192		
	AM	MA ME M	A M	A	ân.	1	
statu ordak bearing the	40	NA MA B	Ma			0	
	- 60		PLPS	cepetia severalian	AL	_^	-
	7.m	NO. M. S.	. 6.4	h	to .	A	~
1004.0 (W.)			< Eastign 16 1	Icelfiction. •			
ADALY INF BURLS			Size N	and to			
		2002 Investments	in calolog per firs lief.	a inchesi 200, 137 anter	In cardoot		
					THE	111m	
G		Mycospha	erella fijiensis	v1.0		11140	
	00 KEDG   KDG   AM	enterflearth   De	writeast   Sofe   Harne	HELPY	10.1	14/14/96/16-	
The real president of the C	Interne Portal (inersion 21)	10 to evaluative room			1000	a college	
The real president of the C	A   A   XA   XXA   XXX		a   +3x   +1.5x   -	1.34 -34 -314	1000	AAT	44
The read permitten of the C Abye: ccc ccc ccc pearlold, 41,197	4 . 9 . 99 . 999						ia.] iaffeld>)
The real prevalue of the C Aver: ccc cc cc builder: praffald, 41,197	4   3   35   353 487-212367	Size 24681	(+) Ref. 320	International		Next 5	affeld > 1
The real generalize of the C Nove: <u>ccc</u> <u>cc</u> Position: <u>peafold</u> , 41:197 Deer Position	4   3   35   353 487-212367	Size 24681	(+) Ref. 320	passed person passed	zent an	Next 5	affeld > 1
The real prevalue of the C Aver: ccc cc cc builder: praffald, 41,197	4   3   35   353 487-212367	Size 24681	i ini kef, 3m mi pami provi Navatariti pravi	passed person passed	and an	Next 5	affeld > 1
Reach   BLAST   Remedi The one personalise of the C Nye: <u>ccc</u> <u>cc</u> Position: perfold, 41,187 Reac Position Real VIDS R.g. economic	467-212367 I generi provol :	Size 24681	i ini kef, 3m mi pami provi Navatariti pravi	and   press press press press press press press press press		Next 5	utteld >>
The real generalize of the C Nove: <u>ccc</u> <u>cc</u> Position: <u>peafold</u> , 41:197 Deer Position	a   p.   3.0   553 467-213567 1 gmont proved p	Zoom +10 Size (Heer name zone) po mos	i i i farf. 300 mil premi provel Alla and provel Alla and provel Alla and provel	and   peaksi pressi peaksi wright • wright •	NUM	Next 5	utteld >>
Rearch   63.51   Rearch The next prevention of the C Nove: <<< < <. Souther: Seathold, 41.197 The Province Rear Vittle Rup accounts ASSA working, 4	a   p.   3.0   553 467-213567 1 gmont proved p	Zoom +10 Size (Heer name zone) po mos	i i i farf. 300 mil premi provel Alla and provel Alla and provel Alla and provel	and   press press press press press press press press press	NUM	Next 5	utteld >>

#### Synteny Viewer



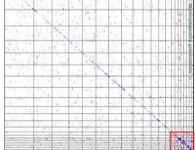
	Genome size (Mbp)	# of genes predicted
Mycosphaerella graminicola	39.7	10,952
Mycosphaerella fijiensis	73.4	10,327
Cochliobolus heterostrophus C5	34.9	9,633
Alternaria brassicicola	30.3	10,688
Pyrenophora tritici-repentis	37.8	12,141
Stagonospora nodorum SN15	37.1	15,983

Genome Statistics

www.jgi.doe.gov/Mgraminicola www.jgi.doe.gov/Mfijiensis www.jgi.doe.gov/Cochliob www.igi.doe.gov/Abrassicicola www.jgi.doe.gov/Pyrenophora www.jgi.doe.gov/Snodorum

VISTA tracks show regions of conservation at the genome level between two genomes. The VISTA peak highlighted on the Maraminicola browser represents conservation between a coding region in M.graminicola and an orthologous region in M.fijensis. By clicking on the VISTA track the user is able to browse orthologous regions in a related genome via an intermediate page.





Smaller chromosomes of M graminicola have higher repeat content, lower gene density, and mostly novel genes indicative of potentially dispensable chromosomes.

M.graminicola v2.0	chr1-chr13	chr14-chr21
fraction of repeats (%)	17 +/- 4	31 +/- 6
gene density (1 per kb)	3.5 +/- 0.2	7.1 +/- 0.8
fraction of gene models with pfam domain (%)	55 +/ 4	3 +/- 3

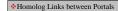


This clustering example demonstrates how the clustering tool can be effectively used to annotate individual genes. The domain view, Fig. 1, shows that the A.brassicicola genes are most likely one gene that has been split. The browser viewer, Fig. 2, shows a gap in the assembly (the red section of bar) which created a split in this gene.



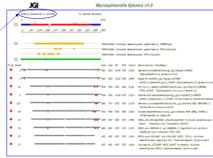
sectio	n of the longer bl	ack		803				1			
000000	in or the longer of	aon		5282	. 1	1		1		2	
			1		1			1		2	
iteins ind	cluded in cluster 1284	-						Fig. 1			
Passing 55 1	Localizes 7	Anna I	Annan Magain, 1	Austine of Austineers 7	Course B	scanat Chill	THE OTHER WICH.	1	DOMANS PINTER	W SENE MODEL	
383.17	CHEMICAL HISTORY, LANSING	2408	1194	1	A beautists			0 -	_		- 110
2808 /**	Contract Indication (Indication	1101	204	A -	A.bessenna			•			-
2010.1	scatters_21.783133-208672	1945	1942	8.	C. Indentity Tue C5			0	_	-	- 15
40012	xuffore.22838818-2545815		1858	1 8	M. Spanes	00000		0	_	-	- 18
338152 [**	mr.2234205-234888	5742	1856	8	N paratesia	00000	0000	0		-	- 18
\$251.7"	NATUR_20.12050-142002	9428	1823		P. Nettorian	00000		o		_	- 18
100681	SQUEETER LATERNEO (429/20)	8021	1943	2	P. Interneta			0		-	- 18

Atlant 6 M



When clicking on the description of the M.fijiensis protein aligned to the M.graminicola protein the user can then jump to the protein analysis page of the M.fijiensis protein thus enabling comparative analysis at the gene level.

JGI	Mycosphaeralla graminicola v2.0 Finistvid Genome
estative, langesticative, 2, step, 2000	K Ta lanae Braser
1	176
6, 045, 107, 364, 28, 8,	0 ,30 ,30 ,69 ,63 ,03
0	Interced to take meety is a spectacity intervel
	Interchic successive exectation spectacity introduced
	iMentone: motore meantylase (Weinfloar)
FLip Shart	But Les ID 128 Some Description (Ladise)
•	top his load and 2775 Represented and Representation and the second seco
• *****	and the set of N2 200 Wightworkshop, pp. 75, 75, 1930 article, ferminel, 6, 6092 Diportantical Figures
• *	epi bit bei ein alle in alle in alle discher einer einer einer dischigten Und lakobat serbenal.
• • •	* 34 54 55 55 15 15 Print Rent Statements and appropriation PRT_2015 (Pyreaglers tribut-repetits)
• •	* 34 54 52 12 12 14 Deterative/setating_g10xhef5,21548 evb/c,bewuel_L_25025 Cochusters reterations
• 112	de loi (11 ell 112 measuristichen/relations_grother102 devel2.)
• • • • • • • • • • • • • • • • • • • •	412 Sel 102 415 1127 or 3,3,1217059 spectratical protein 360,0009 Discussion in reduced 363
• 10	enc est 195 est 1979 re.h.h.f.755est Repotential protocol etters Description stations FIEC MD
• • • • • • • • • • • • • • • • • • • •	This has been and and and and anti-selection probability of the selection
• a	30 40 70 50 100 er.h.b.852421 construct hypothetical proton



Side-by-side comparison of KEGG gene function assignments.

#### \*KEGG Functional Annotation Browser

**IGI** 

# MAP00252 Alanine and aspartate metabolism

Side-by-side comparative analysis of number of genes in different genomes belonging to the same KOG function or KEGG pathway helps to find gene family expansions or possible annotation errors

\*KOG Functional Annotation Browser

function (0)	Function Description	Gene Models in Mycosylumida graninicela 123 Finished Geneme	Gene Models in Wycrogheerstia Tjierwis e1.3	Conte Waters in Contilidatus heterostrophus CS	Gene Models in Abernatis brandicionis	Gene Models in Pyremythers stitui-repetite	Gene Models in Disgenerapers restores EN15
	Ed administration in the prose	( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( )					
	#2022.9.1.1.1.1.0.00 given synthesication synthese categories advantation						
	WEITING Autophysics Statistics and Street and St.						
	ADDITED Antiplement of the dataset of the second of						
	and second and second						
	WEIGHTAN An of Concerning Strappingth						
	Appendial Grands generalized inter-						
	ADD22111 (below aprilliant Patientian Aprilliant		-			1.00	
	#064791 (Drittant				14		
	minuted lineare share						
	POGNETI Constraint I as Remarking and search						

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC52-06NA25396.

Side-by-side comparison of KOG gene function assignments.