WHOLE GENOME COMPARISON OF THE *A. FUMIGATUS* GROUP

J. Wortman
The Institute for Genomic Research, USA
Saturday, February 25, 2006, 11:50 - 12:10 pm

The publication of the genomes of *A. fumigatus*, a human pathogen, *A. oryzae*, used in food production, and *A. nidulans*, a genetic model organism, represents an important milestone for the Aspergillus research community, expanding knowledge of their physiology and mechanisms of gene regulation. Although members of the same genus, phylogenetic analysis of these species revealed that each pair is approximately as evolutionarily distant as mammals and fish, with *A. fumigatus* and *A. oryzae* more related to one another than *A. nidulans*.

Three additional genome projects are being funded by the NIAID with the goals of better elucidating the *A. fumigatus* genome and improving the genome annotation: *N. fischeri* (*A. fischerianus*), *A. clavatus* and *A. terreus*. *A. terreus* and *N. fischeri* are capable of causing disease in human patients, while *A. clavatus* cannot survive at body temperature, and is therefore not a viable pathogen. Preliminary analysis of genome structure supports the previously determined phylogeny of these species, with *N. fischeri* the most related to *A. fumigatus*, with the longest uninterrupted syntenic blocks; *A. clavatus* intermediate, with numerous local rearrangements and inversions; and *A. terreus* the most distant, with inter-scaffold breaks common.

This presentation will cover our preliminary examination of the genome sequences and predicted gene complements of *N. fischeri* and *A. clavatus* in comparison to two strains of *A. fumigatus,* Af293 and CEA10. Even within the single species, *A. fumigatus* isolates vary significantly in their morphology and pathogenicity. Af293 and CEA10 have a number of unique, strain-specific genes (2%), located predominantly in non-syntenic subtelomeric blocks. A few strain-specific regions contain putative gene clusters involved in secondary metabolism, osmotolerance, or arsenic resistance, highlighting the potential role of the subtelomeric regions in maintaining species variability.

Comparative genomic analysis at the level of these more closely related *Aspergilli* should provide important additional insight into the evolutionary forces at play and their effect on gene content, regulation and expression. Characterization of strain and species specific genes, polymorphisms and differential regulation will foster our understanding of mechanisms of pathogenicity, environmental adaptation and resistance to anti-fungal treatments.

Comparative Genomics of the *Aspergillus* Genus

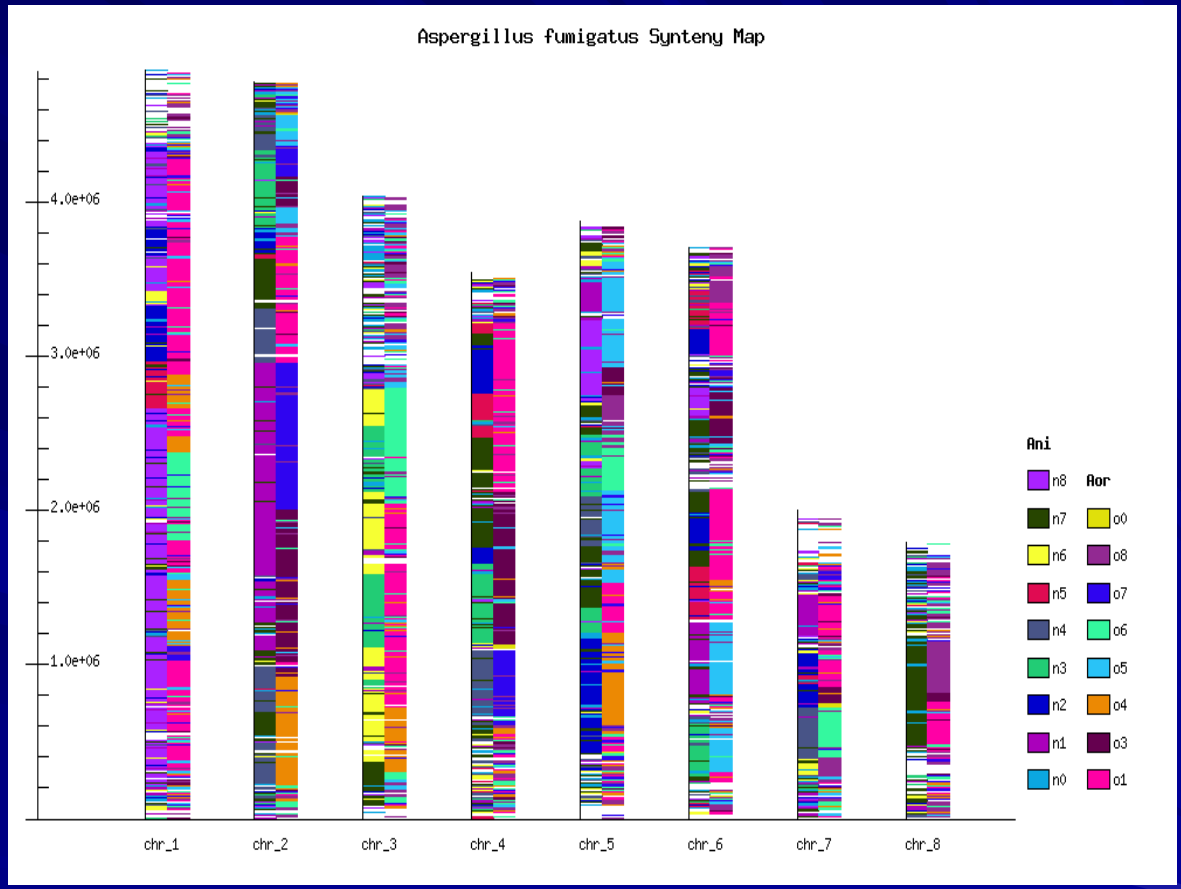
Jennifer Russo Wortman
The Institute for Genomic Research
AAA 2006

TIGR
THE INSTITUTE FOR GENOMIC RESEARCH

Aspergillus Genomes

Nature 438 (22 Dec 2005)

- *A. fumigatus*
  - Human pathogen
- *A. nidulans*
  - Model organism
- *A. oryzae*
  - Food production



Although in the same genus, the three Aspergilli differ considerably in their genome sequences. Predicted orthologs shared by all three species display an average of only 68% amino acid identity. This protein identity is comparable to that between mammals and fish. The three species also differ considerably in genome size and show extensive structural reorganization.

Goals of Project

Three additional genome projects are being funded by the NIAID with the goals of better elucidating the *A. fumigatus Af293* genome:

- *N. fischeri* (*A. fischerianus*)
- *A. clavatus*
- *A. terreus*

The objectives in sequencing these three *Aspergilli* genomes are to use comparative genomics to:

- Improve annotation in *Aspergillus* genomes and provide new targets for experimental studies
- Identify differences in gene content and/or regulatory elements that might contribute to pathogenicity
- Facilitate vaccine component selection, with the aim of preventing invasive aspergillosis

N. fischeri and *A. clavatus* are being sequenced by TIGR

A. terreus is being sequenced by the Broad Institute

A second *A. fumigatus* strain (CEA10) was made available through Merck and is included in this analysis

N. fischeri [A. fischerianus]

- Apart from sister taxa, *A. fumigatus var ellipticus*, *N. fischeri* is the most closely related species to *A. fumigatus*
  - *N. fischeri* is the teleomorph of *A. fischerianus*
- Rarely identified as a pathogen with only two medical cases reported in literature
  - Scarcity in environment
  - Misidentification in the laboratory
  - Relative lack of virulence
- Role in food spoilage
- Homothallic with thermoresistant ascospores
  - Reduced growth at 42ºC relative to 37ºC and no radial growth at 48ºC in contrast to *A. fumigatus* which shows increased growth at 42ºC relative to 37ºC and measurable growth at 48º

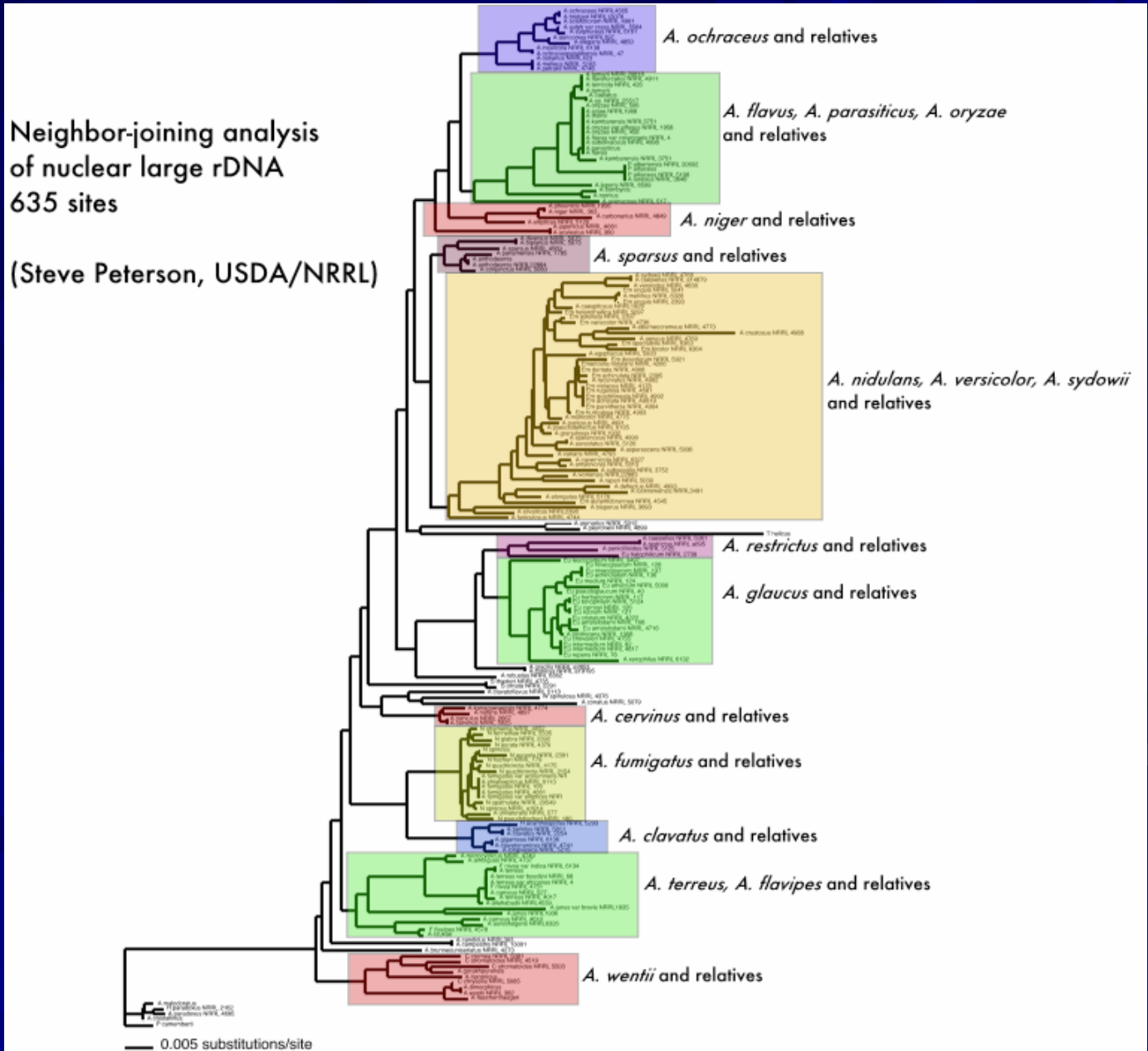
A. clavatus

- *A. clavatus* is a very rare human pathogen with only one medical case reported in literature (post-surgery endocarditis)
  - Grows more slowly at 37^{o}C than *A. fumigatus*
  - Bigger spore size may prevent lung penetration
- Although not a common pathogen, it is probably an important allergen and has been shown to be the cause of an extrinsic allergic alveolitis known as malt worker's lung
- Produces a number of mycotoxins including patulin, kojic acid, cytochalasins and tremorgenic mycotoxins
  - Causes neurotoxicosis in sheep and cattle fed infected grain

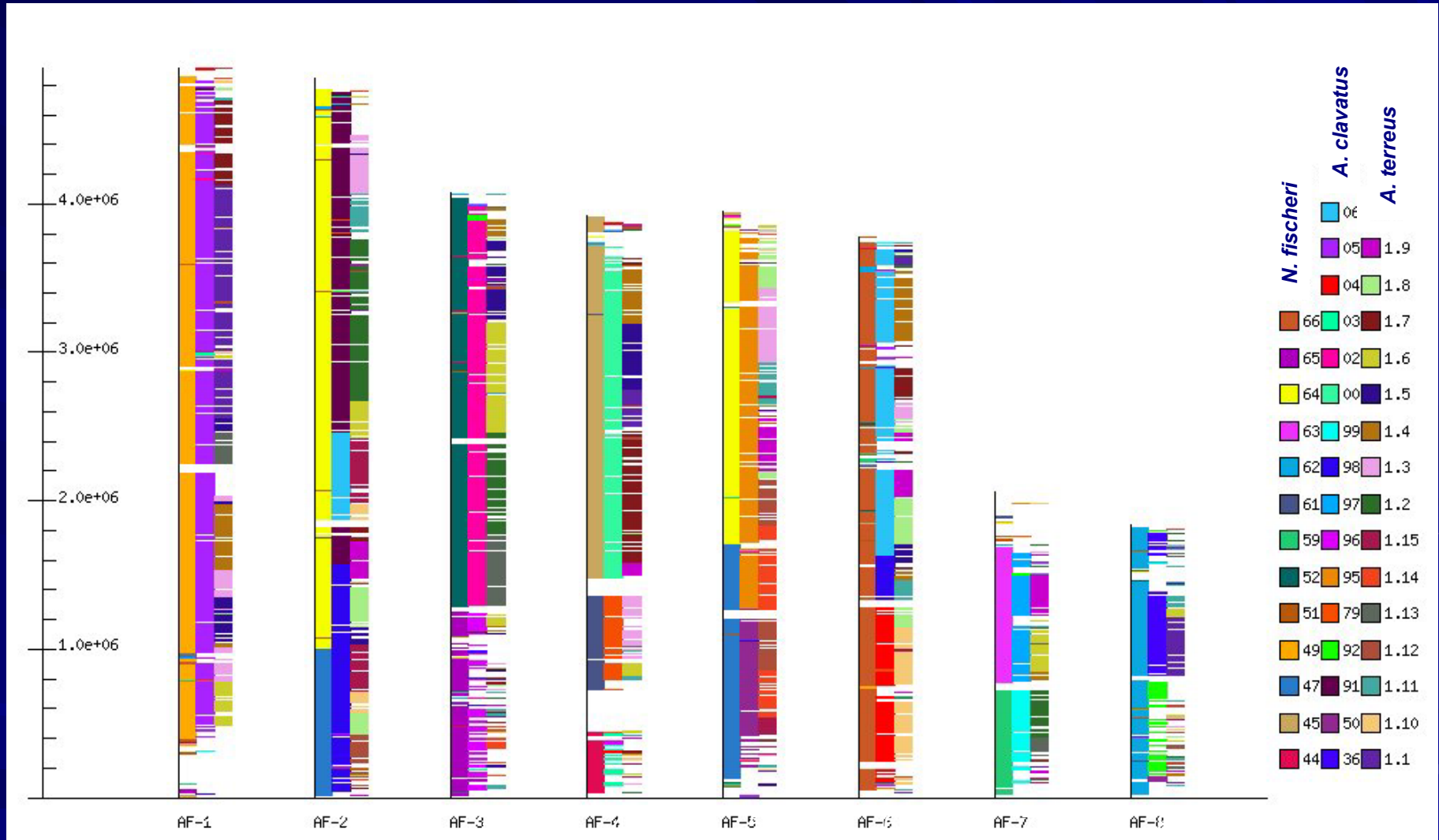
A. terreus

- *A. terreus* is a less frequent cause of invasive aspergillosis
  - An emerging pathogen, number of cases increasing
  - Higher mortality
  - Resistant to amphotericin B
- Allergenic, causing allergic bronchopulmonary aspergillosis
- Used in industry to produce lovastatin and itaconic acid
- Produces a number of secondary metabolites and mycotoxins, including territrem A, citreoviridin, citrinin, gliotoxin, patulin, terrein, terreic acid and terretonin
- *A. terreus* is unique among *Aspergilli* in producing lateral cells termed aleurospores in the absence of typical conidiophore structures in submerged culture

Aspergillus Phylogeny

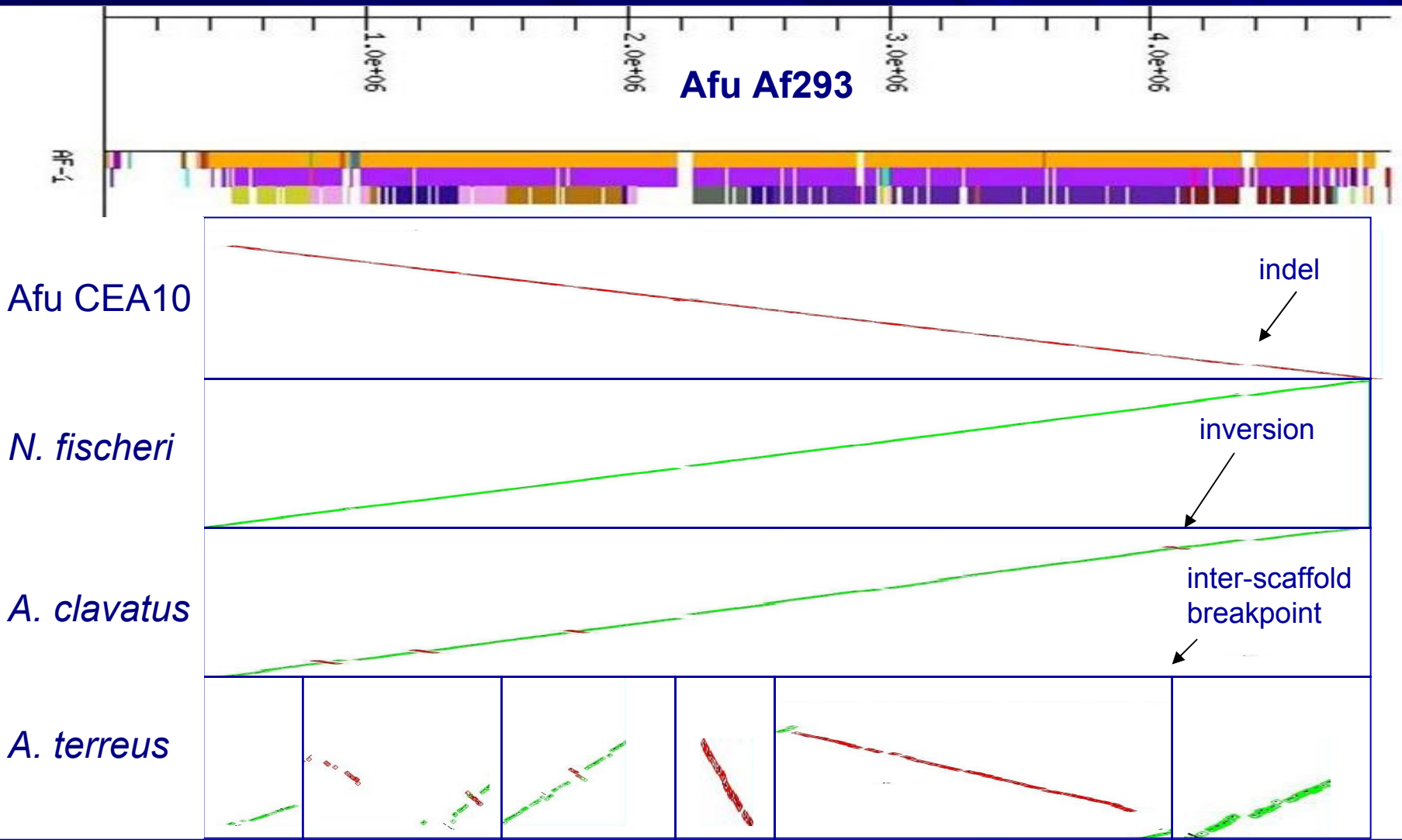


A. fumigatus Synteny Map



Assembled supercontigs > 100 kb from *N. fischeri*, *A. clavatus*, and *A. terreus* were aligned to *A. fumigatus* chromosomes using the MUMmer package (*http://mummer.sourceforge.net/*).

Genome Structure



Dot plots showing pair-wise alignments of *A. fumigatus* strain CEA10, *N. fischeri*, *A. clavatus* and *A. terreus* against *A. fumigatus* strain Af293, Chromosome 1.

Pair-wise Conserved Synteny

			Coverage				Max/Mean	
		A. fumigatus	*N. fischeri*	*A. clavatus*		*A. fumigatus*	*N. fischeri*	*A. clavatus*
A. fumigatus		-	27.7 (96)	25.3 (88)		-	1025	278
N. fischeri		29.9 (93)	-	26.8 (84)		3895	-	294
A. clavatus		25.9 (93)	25.9 (93)	-		2232	1587	-

Table 2 | Characteristics of pairwise conserved synteny

Reference	Coverage (Mb) (percentage of reference)*				Maximum/mean block length (kb)‡		
	A. nidulans	*A. fumigatus*	*A. oryzae*	Either†	*A. nidulans*	*A. fumigatus*	*A. oryzae*
A. nidulans	-	20.5 (68)	20.4 (68)	21.6 (72)	-	175	114
A. fumigatus	20.4 (73)	-	20.7 (74)	21.5 (77)	2,429	-	168
A. oryzae	23.3 (63)	24.3 (66)	-	25.4 (69)	943	1,159	-

*Coverage of reference organism assembly by pairwise conserved syntenic blocks (>10 kb in length) to each target genome.
†Coverage of reference organism assembly by pairwise conserved syntenic blocks (>10 kb in length) in either other genome.
‡Upper right half shows mean blocks sizes and lower left half shows maximum sizes across all blocks using either organism as reference.

Galagan et al. Nature 438, 1105-1115 (22 December 2005)

As expected, *A. fumigatus* shares larger syntenic regions with *N. fischeri* and *A. clavatus*, covering a significantly larger percentage of each genome, than with the more evolutionarily distant *A. oryzae* and *A. nidulans*.

Genome Statistics

Chromosome	AFU Af293	AFU CEA10	NFA	ACLA
Size (Mb)	28.8	29.2	32.0	27.7
Supercontigs > 2kb	19	32	256	28
Supercontigs > 10kb	19	17	19	21
GC Content	49.8	49.4	49.6	49.2
Genes				
# of Genes	**9854**	**10099**	**10929**	**8653**
Mean Gene Length (bp)	1461.8	1425.2	1434.3	1436.8
Gene Density	2924	2868.3	3010.7	3091.7
Percent with Introns	78.4	78.4	78.2	83
Exons				
Mean # per Gene	2.9	2.8	2.9	3
GC Content	54	53.9	54.4	55.4
Mean Length (bp)	511.2	500.8	498.7	471.9
Introns				
GC Content	46.6	46.6	46.4	46.8
Mean Length (bp)	95.4	88.4	86.1	101.5
Intergenic Regions				
GC Content	45.9	45.1	45	44
Mean Length (bp)	1275.3	1275.6	1296.4	1390.6

Proteome Comparison

Ortholog Analysis:

- All vs. all BlastP
- Compute mutual best hits
- Cluster results

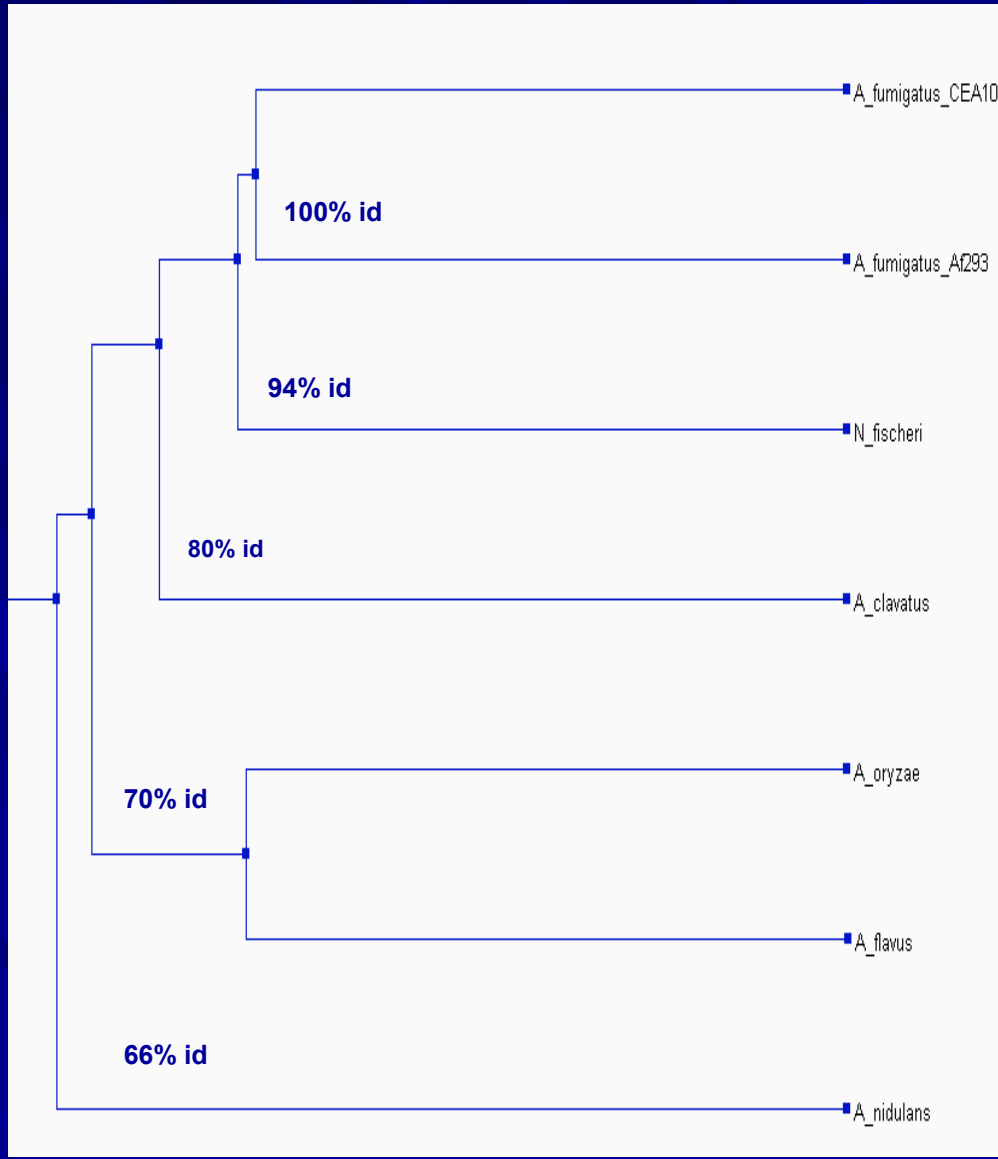
Identified 9845 ortholog clusters

7076 clusters represent all analyzed proteomes

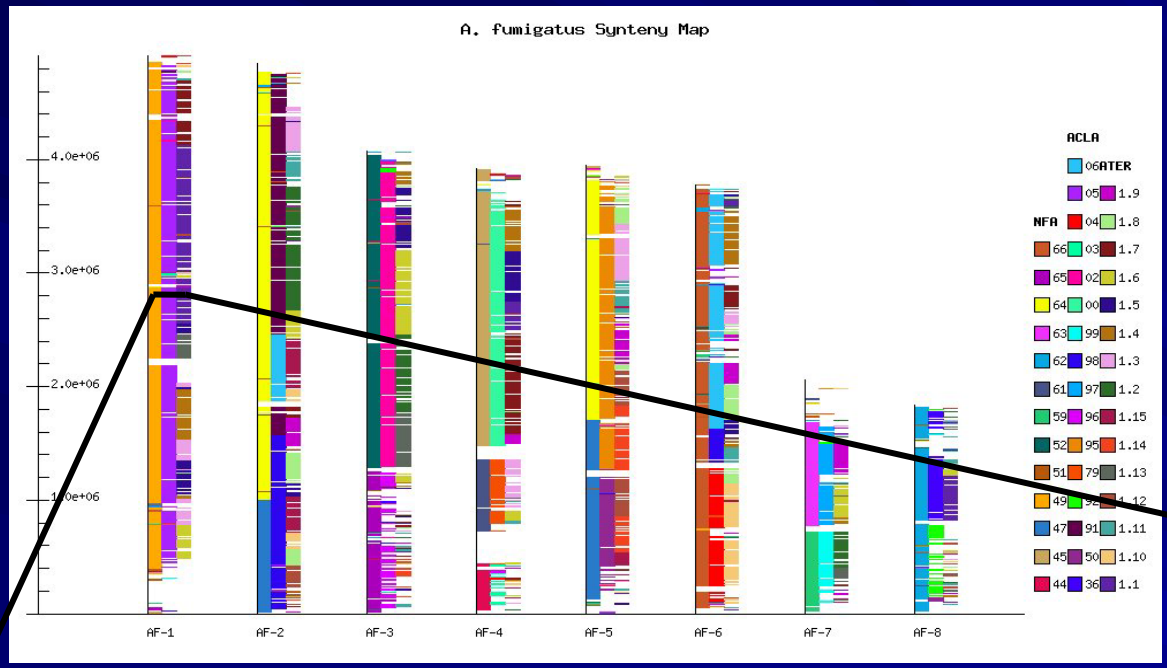
(*A. fumigatus Af293, A. fumigatus CEA10, N. fischeri, A. clavatus*)

6862 'closed' clusters contain one protein from each proteome

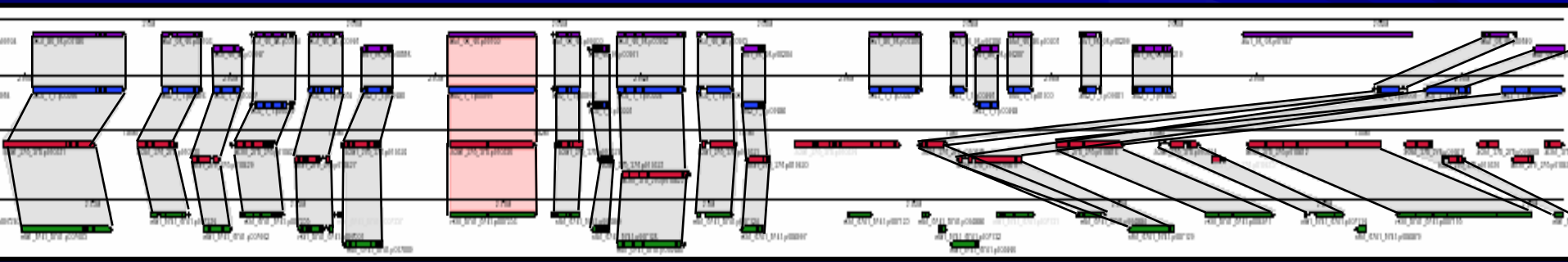
Species	#Genes in COGs	Percent of Proteome
***A. fumigatus* Af293**	8769 (9638)	89% (98%)
***A. fumigatus* CEA10**	8874 (9758)	88% (94%)
N. fischeri	9021	83%
A. clavatus	7672	89%



Orthologs and Synteny

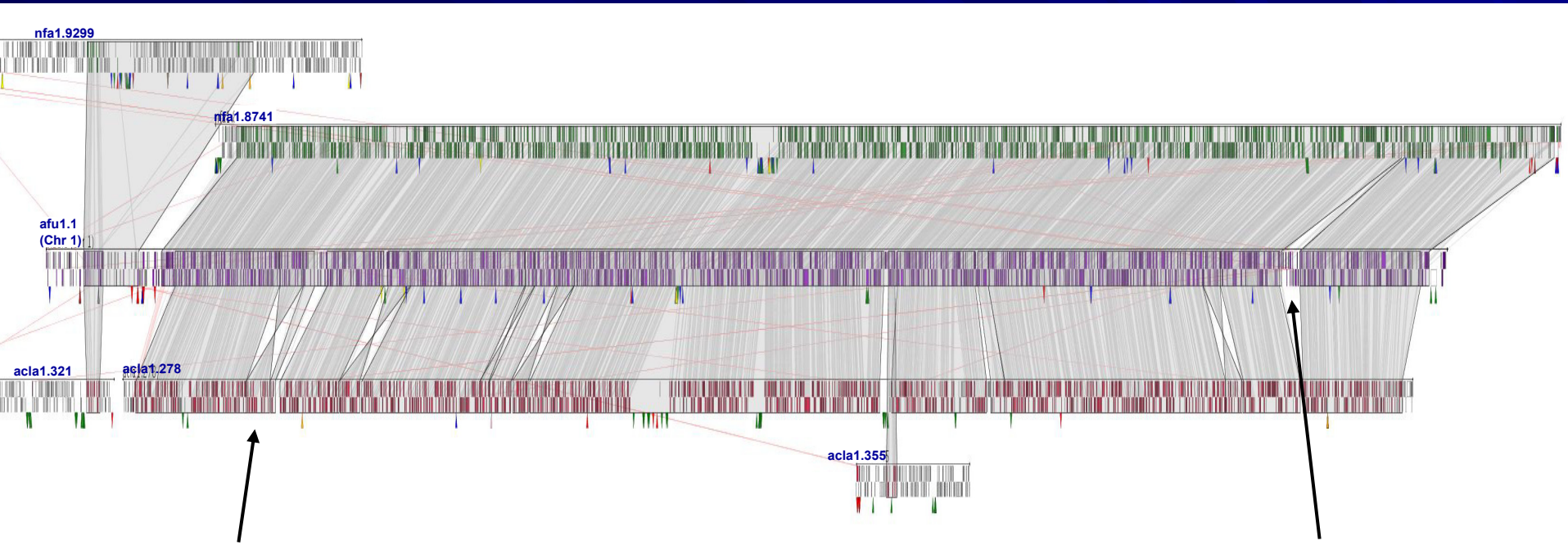


Correlating ortholog clusters with genomic context allows us to define and investigate syntenic blocks and breakpoints



SYBIL: Web-based software for comparative genomics: *http://sybil.sourceforge.net/*

Integrated Chromosome View

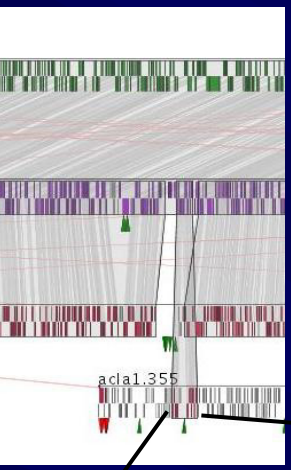


inversion breakpoint indel

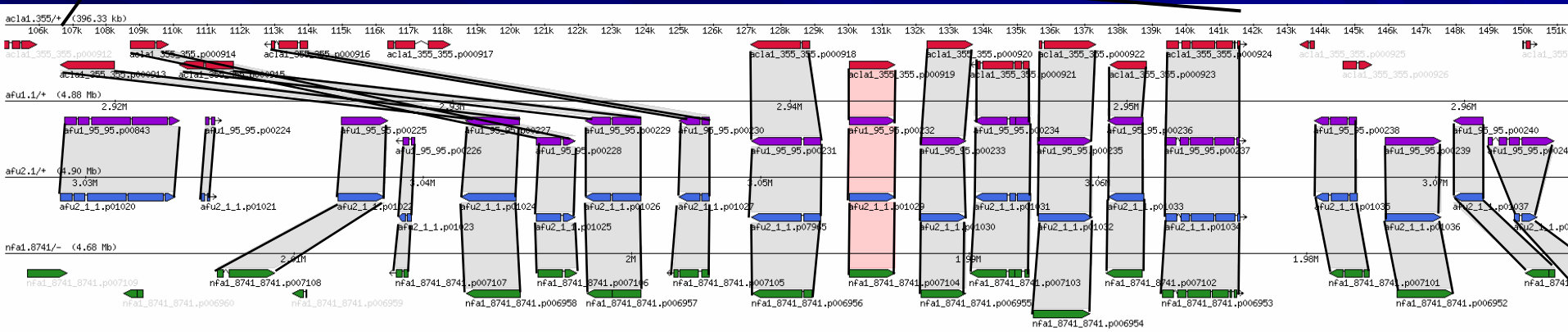
Synteny view of *A. fumigatus* strain Af293, Chromosome 1

- Enables correlation of genome structure differences with gene content, as well as other feature types
- Colored triangles represent different classes of transposable elements

Translocated Cluster – *A. clavatus*



Amino Acid Metabolism



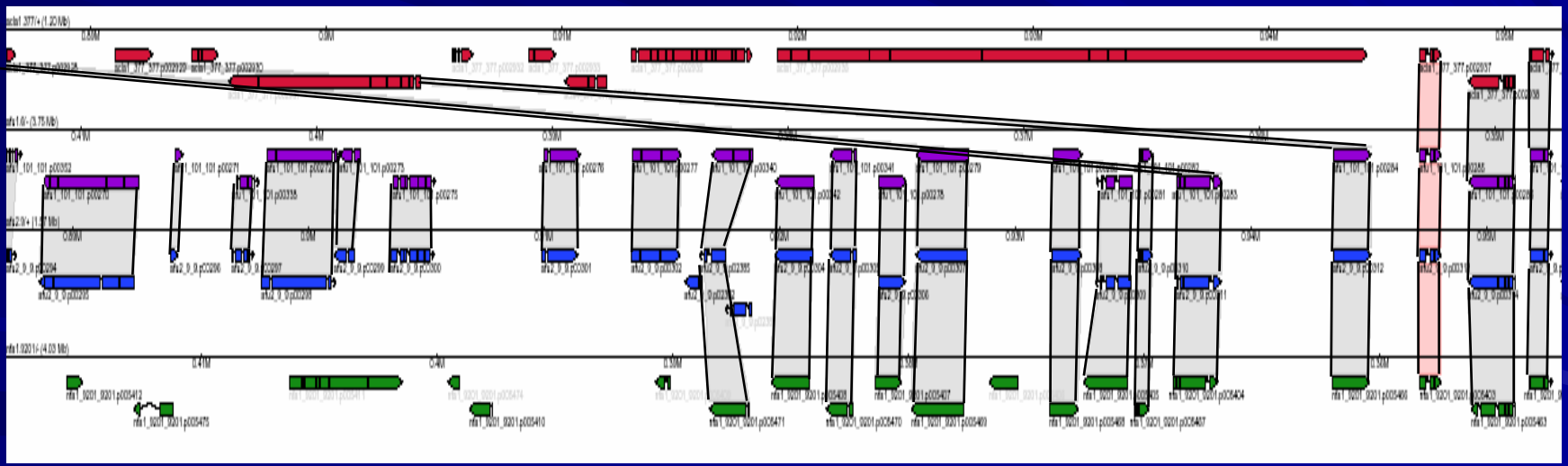
AMP-binding enzyme
DAHP synthetase
lactonohydrolase
glycosyl hydrolase
MFS transporter
aldehyde reductase
GMC oxidoreductase

acla1.245
fatty acid desaturase
cytochrome P450

Unique Cluster – *A. clavatus*

Fatty Acid Synthesis

Beta-ketoacyl synthase
serine hydrolase
MFS transporter
ABC transporter
AMP-binding enzyme



putative pectinesterase

Unique Genes

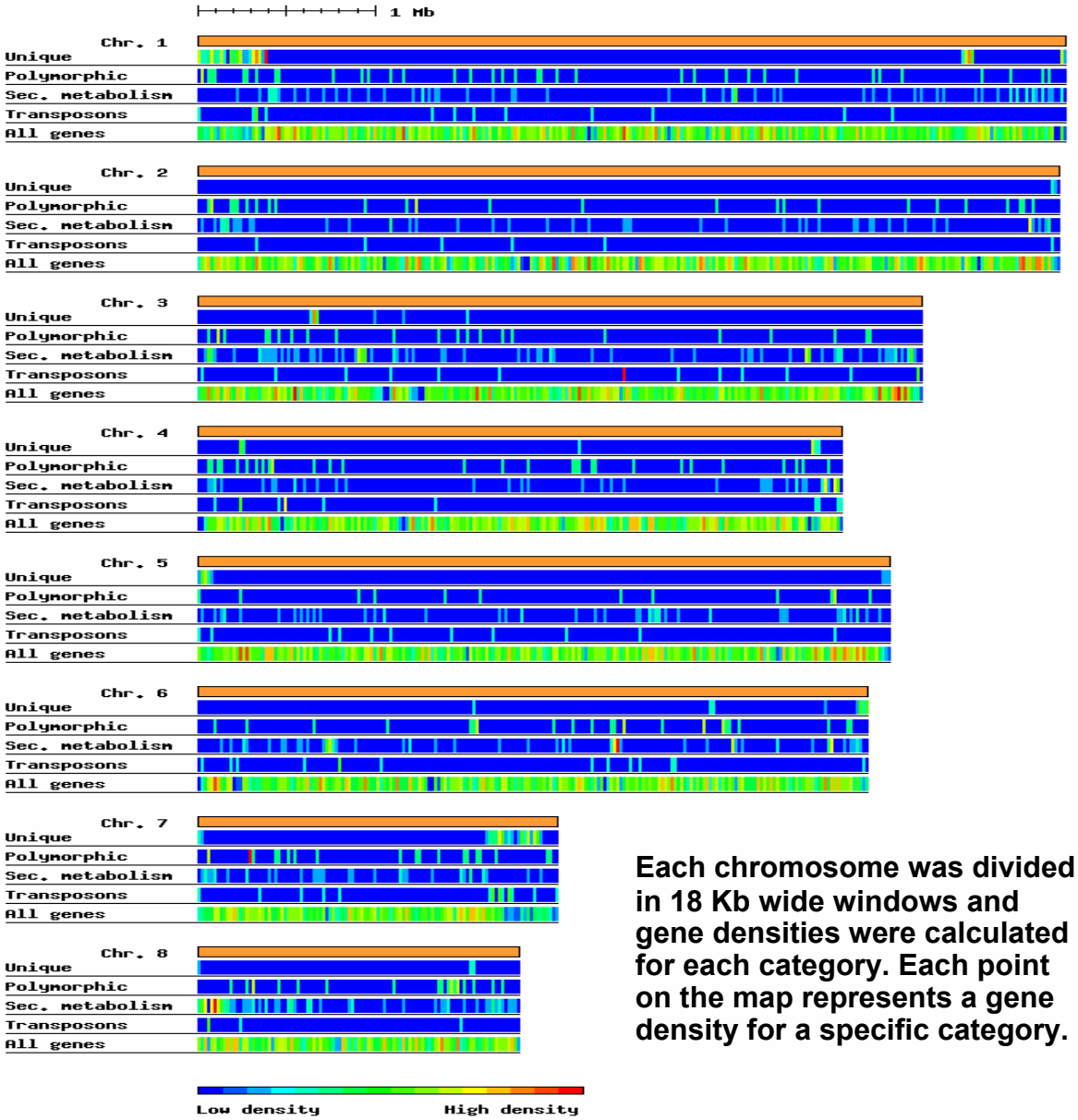
A. fumigatus Af293 vs. *N. fischeri*	1267
A. fumigatus Af293 vs. *A. clavatus*	2447
N. fischeri vs. *A. clavatus*	3509
N. fischeri vs. *A. fumigatus* Af293	2561
A. clavatus vs. *A. fumigatus* Af293	1682
A. clavatus vs. *N. fischeri*	1522

- MFS transporters
- C6 transcription factors
- Secondary metabolism
- Lipid metabolism
- Carbohydrate metabolism
- Protein modification
- Cell wall biosynthesis
- Amino acid metabolism
- Stress response
- Signal transduction

Approximate numbers of species-specific genes in pair-wise genome comparisons, along with the functional categories most represented in these gene sets.

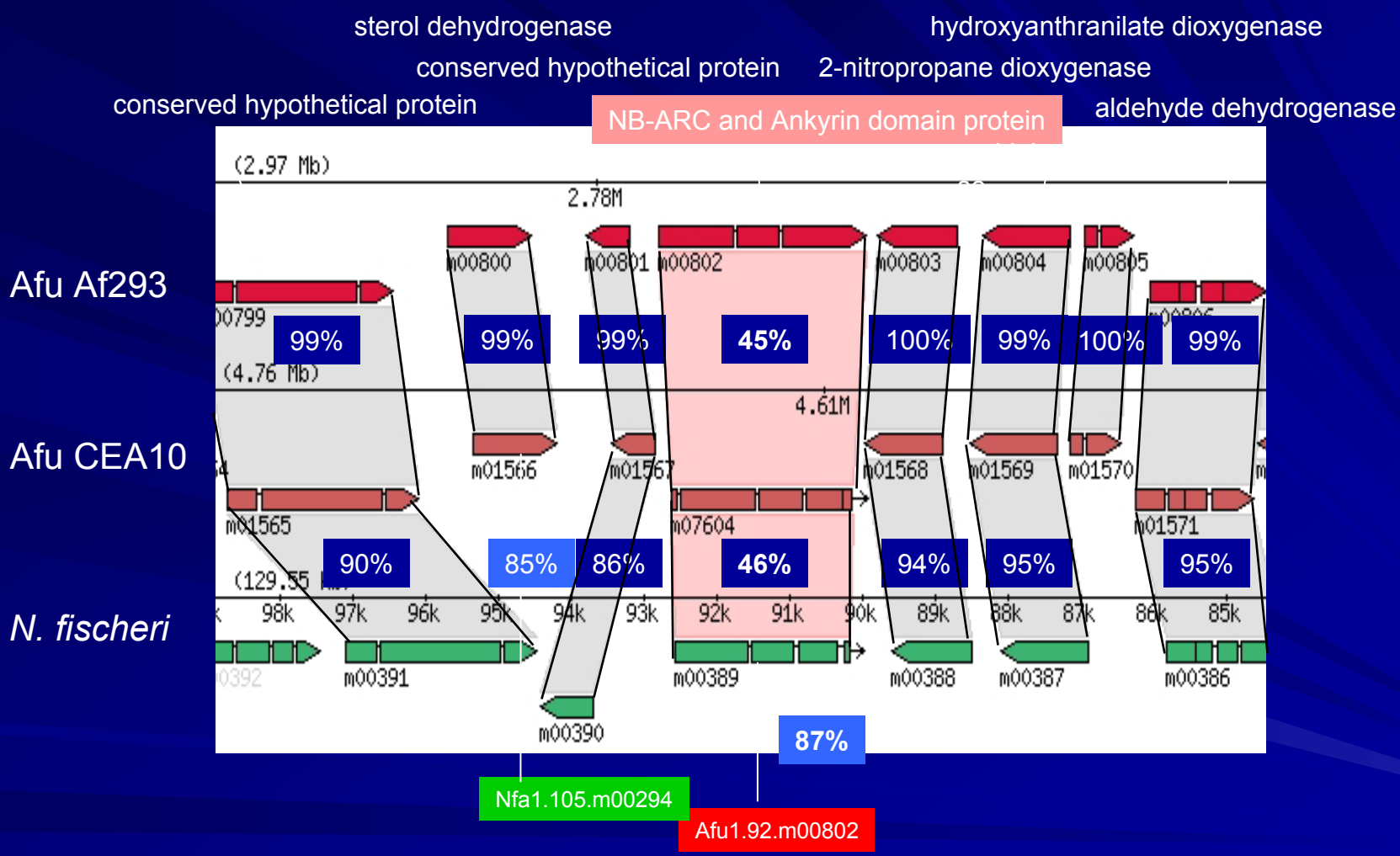
Annotation inconsistencies between genomes can affect the accurate identification of strain and species specific genes.

Strain Specific Genes in Af293



Each chromosome was divided in 18 Kb wide windows and gene densities were calculated for each category. Each point on the map represents a gene density for a specific category.

Polymorphic Genes



Polymorphism in a gene encoding a NB-ARC and Ankyrin domain protein. The protein is distantly related to other fungal proteins involved in self-nonself recognition during hyphal fusion.

Secondary Metabolism

		A. fumigatus Af293	*N. fischeri*	*A. clavatus*
A. fumigatus Af293	NRPS	**16**	9	6
	NRPS-like	**6**	2	1
	hybrid	**4**	1	2
	PKS	**25**	8	4
N. fischeri	NRPS	9	**26**	4
	NRPS-like	2	**4**	2
	hybrid	1	**3**	1
	PKS	8	**17**	5
A. clavatus	NRPS	6	4	**12**
	NRPS-like	1	2	**5**
	hybrid	2	1	**5**
	PKS	4	5	**15**

Numbers of select secondary metabolism enzymes found and conserved in each of the *Aspergillus* species analyzed, including non-ribosomal peptide synthetases (NRPS), polyketide synthases (PKS), and hybrid NRPS-PKS proteins.

Medically Relevant Genes

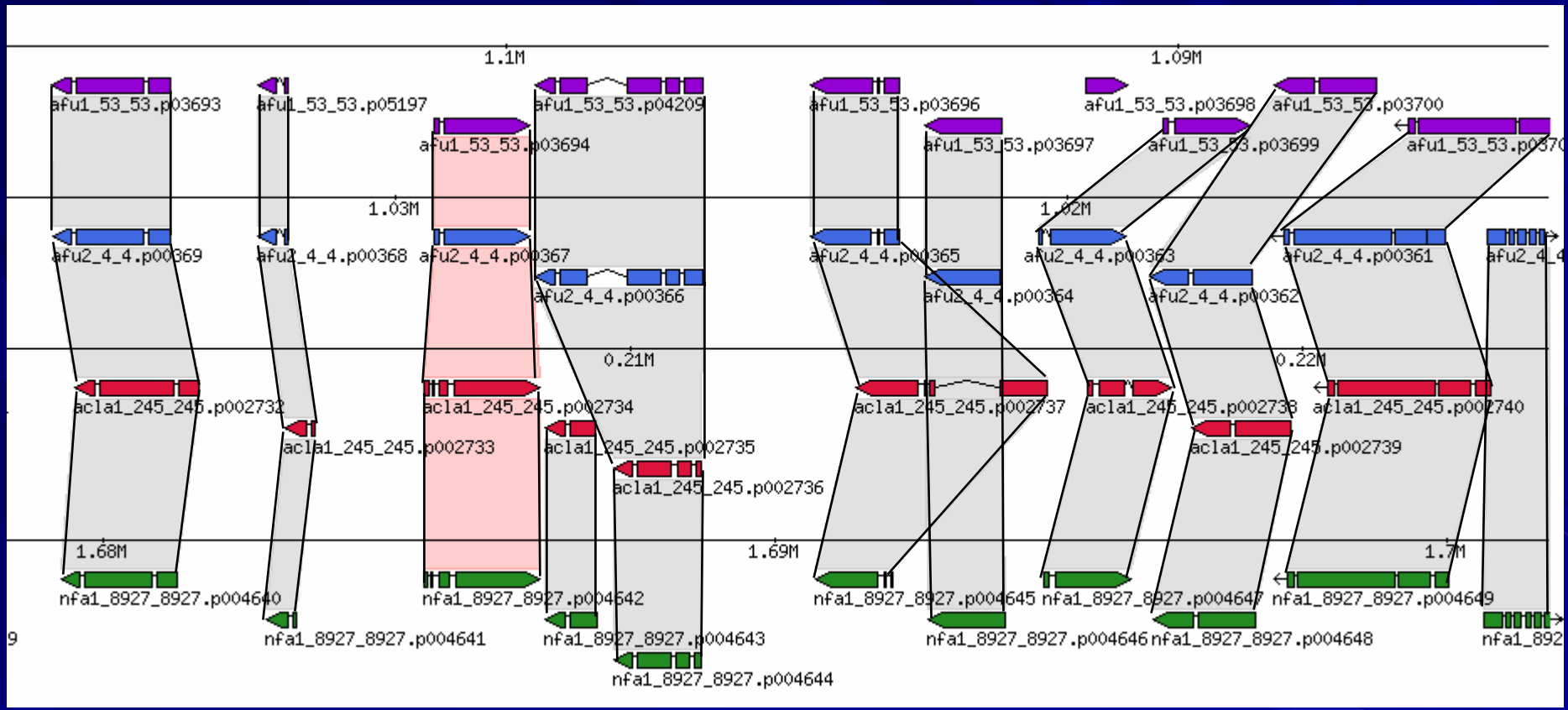
Protein	Cellular Function	Genome Locus Identifier	Reference(s)
PyrG	Pyrimidine biosynthesis	Afu2g08360	d'Enfert et al., 1996
PabaA	Folate biosynthesis	Afu6g04820	Brown et al., 2000
LysF	Lysine biosynthesis	Afu5g08890	Liebmann et al., 2004a
PksP	Pigment biosynthesis	Afu2g17600	Langfelder et al., 1998 Tsai et al., 1998
SidA	Siderophore biosynthesis	Afu2g07680	Schrettl et al., 2004
ChsG	Chitin biosynthesis	Afu3g14420	Mellado et al., 1996
CgrA	Ribosome biosynthesis	Afu8g02750	Bhabhra et al., 2004
AreA	Regulation of nitrogen metabolism	Afu6g01970	Hensel et al., 1998
CpcA	Regulation of amino acid biosynthesis	Afu4g12470	Krappmann et al., 2004
LaeA	Regulation of secondary metabolism	Afu1g14660	Bok and Keller, 2003
Cat1	Oxidative stress	Afu3g02270	Paris et al., 2003
Cat2	Oxidative stress	Afu8g01670	Paris et al., 2003
Fos-1	Stress response	Afu6g10240	Clemons et al., 2002
GpaB	cAMP signaling	Afu1g12930	Liebmann et al., 2004b
PkaC1	cAMP signaling	Afu2g12200	Liebmann et al., 2004b
RhbA	Nutrient sensing	Afu5g05480	Panepinto et al., 2003
AcyA	cAMP signaling	Afu6g08520	Liebmann et al., 2004b

Af293	Af293	CEA10	N. fischeri	*A. clavatus*	Protein name
Afu1g00810	98.m00059	X	X	X	allergen Asp F7-like, putative
Afu3g00710	100.m00052	3.m01280	8673.m002433	X	allergen Asp F4-like, putative
Afu6g03230	101.m00189	9.m00206	9201.m005332	X	cell wall glucanase/allergen F16-like
Afu7g04930	99.m00257	8.m02878	8817.m001520	X	alkaline serine protease (PR1)/allergen F18-like
Afu6g03260	101.m00186	9.m00202	9201.m005333	377.m002813	aspartic endopeptidase (AP1), putative
Afu2g05150	57.m05810	2.m00479	9292.m008856	369.m005229	cell wall galactomannoprotein Mp2
Afu5g08030	94.m00607	4.m00640	9292.m009530	262.m006016	extracellular cellulase CelA, putative
Afu1g16190	95.m00699	1.m01486	8741.m006531	278.m010115	cell wall glucanase Crf1
Afu6g08510	108.m00201	6.m03533	9201.m005820	368.m005722	cell wall glucanase, putative
Afu3g09250	93.m00562	3.m00590	9288.m004839	286.m006347	cell wall glucanase, putative
Afu5g01440	53.m03957	4.m00112	8927.m004719	245.m003006	allergen, putative
Afu2g03720	57.m05668	2.m00344	9292.m008761	369.m005363	peptidyl-prolyl cis-trans isomerase
Afu6g10480	108.m00366	6.m00380	9201.m006108	368.m005919	peptidyl-prolyl cis-trans isomerase, putative
Afu4g09580	96.m00440	5.m00370	9374.m003881	293.m005638	major allergen Asp F2
Afu4g06670	96.m00692	5.m00082	9374.m004144	293.m005362	allergen Asp F7
Afu4g11800	96.m00257	5.m00585	9374.m003749	293.m005834	alkaline serine protease Alp1
Afu4g07650	96.m00605	5.m00181	9374.m004376	293.m005463	peptidyl-prolyl cis-trans isomerase (CypB), putat
Afu4g00870	109.m00070	11.m00069	9105.m000907	355.m000958	antigenic cell wall galactomannoprotein, putative
Afu2g03830	57.m05679	2.m00355	9292.m008766	369.m005352	allergen Asp F4
Afu2g03120	57.m05614	2.m00289	8927.m004057	369.m005459	cell wall glucanase (Utr2), putative
Afu3g02940	100.m00249	3.m01080	8673.m002790	322.m002485	allergen, putative
Afu4g03240	104.m00063	10.m00155	8881.m001236	297.m001803	cell wall serine-threonine-rich galactomannoprot
Afu2g02050	57.m05507	2.m00184	8927.m004349	369.m005561	peptidyl-prolyl cis-trans isomerase, putative
Afu5g01890	53.m03913	4.m00156	8927.m003856	245.m002969	peptidyl-prolyl cis-trans isomerase, putative
Afu2g12630	92.m00394	2.m08163	9292.m010160	344.m006145	allergen Asp F13
Afu1g01750	98.m00138	1.m00199	8741.m007863	278.m011413	peptidyl-prolyl cis-trans isomerase, putative
Afu4g11580	96.m00277	5.m00564	9374.m004174	293.m005816	Mn superoxide dismutase (SodB), putative
Afu5g13350	94.m00852	4.m01147	9292.m008491	262.m006395	peptidyl-prolyl cis-trans isomerase, putative
Afu5g13300	94.m00160	4.m01142	9292.m008496	262.m006390	aspartic endopeptidase Pep1
Afu5g02330	53.m03871	4.m00199	8927.m003833	245.m002917	ribonuclease mitogillin
Afu5g04170	53.m03692	4.m00370	8927.m004639	245.m002731	suppressor of vegetative incompatibility (Mod-E)
Afu2g15950	92.m00681	2.m01426	9292.m010223	344.m006471	extracellular aspartic endopeptidase, putative
Afu6g02140	101.m00282	9.m00310	9201.m005467	377.m002927	peptidyl prolyl cis-trans isomerase (CypC), putat
Afu6g02280	101.m00338	9.m00297	9201.m005481	377.m002913	allergen Asp F3
Afu6g06770	108.m00039	6.m00049	9201.m005942	369.m005136	enolase/allergen Asp F 22
Afu2g11850	92.m00329	2.m01029	9292.m010127	344.m006072	60S ribosomal protein L3
Afu2g10100	92.m00193	2.m00857	9292.m010052	344.m005910	60S acidic ribosomal protein P2/allergen Asp F 8
Afu8g07080	55.m03210	7.m00038	9299.m003227	321.m002310	elastinolytic metalloproteinase Mep
Afu8g03890	55.m03245	7.m00317	9299.m002921	321.m002019	peptidyl-prolyl cis-trans isomerase, putative
Afu3g07430	93.m00945	3.m00762	9288.m004677	286.m006187	peptidyl-prolyl cis-trans isomerase/cyclophilin, p
Afu1g14550	95.m00542	1.m01345	8741.m006683	278.m010266	Mn superoxide dismutase MnSOD
Afu5g09210	94.m00504	4.m00754	9292.m009224	262.m005903	autophagic serine protease Alp2

Published *A. fumigatus* genes involved in pathogenicity are conserved with *N. fischeri*, and *A. clavatus*, suggesting that they might be important for growth and stress response.

Most putative *A. fumigatus* allergen genes are also conserved, with one missing in *N. fischeri* and four missing in *A. clavatus*.

Annotation Errors Affect Analysis



5’ exons missed in both Afu strains

Merged gene calls are a common problem in fungal genome annotation

Most exon differences between these genomes are annotation errors

Comparative data can be leveraged to improve annotation

Preliminary Conclusions

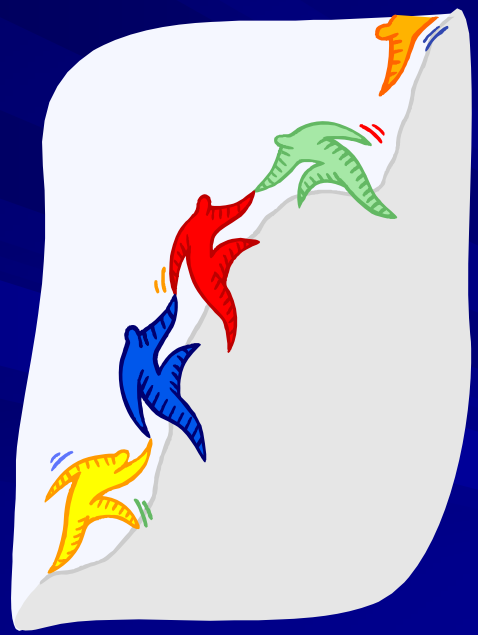
- The *A. fumigatus* core chromosome regions encode predominantly highly conserved housekeeping genes, whereas subtelomeric regions tend to encode strain- and lineage-specific genes.
- The two *A. fumigatus* strains, Af293 and CEA10, have a significant number (2%) of strain-specific genes, which have no detectable orthologs and are located mostly in non-syntenic subtelomeric blocks.
- Both the genome structure and proteome analyses support the previously determined phylogeny of *A. fumigatus*, *N. fischeri* and *A. clavatus.* Breaks in synteny between these genomes are often flanked by repeats and transposable elements.
- In pair-wise species comparison, 10-30% of individual proteomes were not conserved, with certain functional categories more likely to be divergent
- Previously identified genes associated with pathogenicity are well conserved between *A. fumigatus*, *N. fischeri* and *A. clavatus*, suggesting that most have a critical role in cell growth and maintenance.
- In addition to presence or absence of genes, protein mutation rates and regulatory mechanisms must be explored to explain phenotypic and functional differences between species.

Acknowledgements

Natalie Fedorova
Vinita Joardar
Rama Maiti
Brian Haas
Paolo Amedeo
Joshua Orvis

Sam Angiuoli
Jonathan Crabtree
Jay Sundaram

Owen White



Bill Nierman
H. Stanley Kim
Catherine Ronning

Michael Anderson
David Denning
Gustavo Goldman

James Galagan
Bruce Birren

Bo Jiang
Terry Roemer
Phil Youngman
Sebastien Lemieux
Howard Bussey

NIAID