

## **Comparative Reannotation of 21 *Aspergillus* Genomes**

Asaf Salamov<sup>1</sup>, Robert Riley<sup>1</sup>, Alan Kuo<sup>1</sup>, Igor Grigoriev<sup>1</sup>

<sup>1</sup>DOE Joint Genome Institute, Walnut Creek, CA, USA

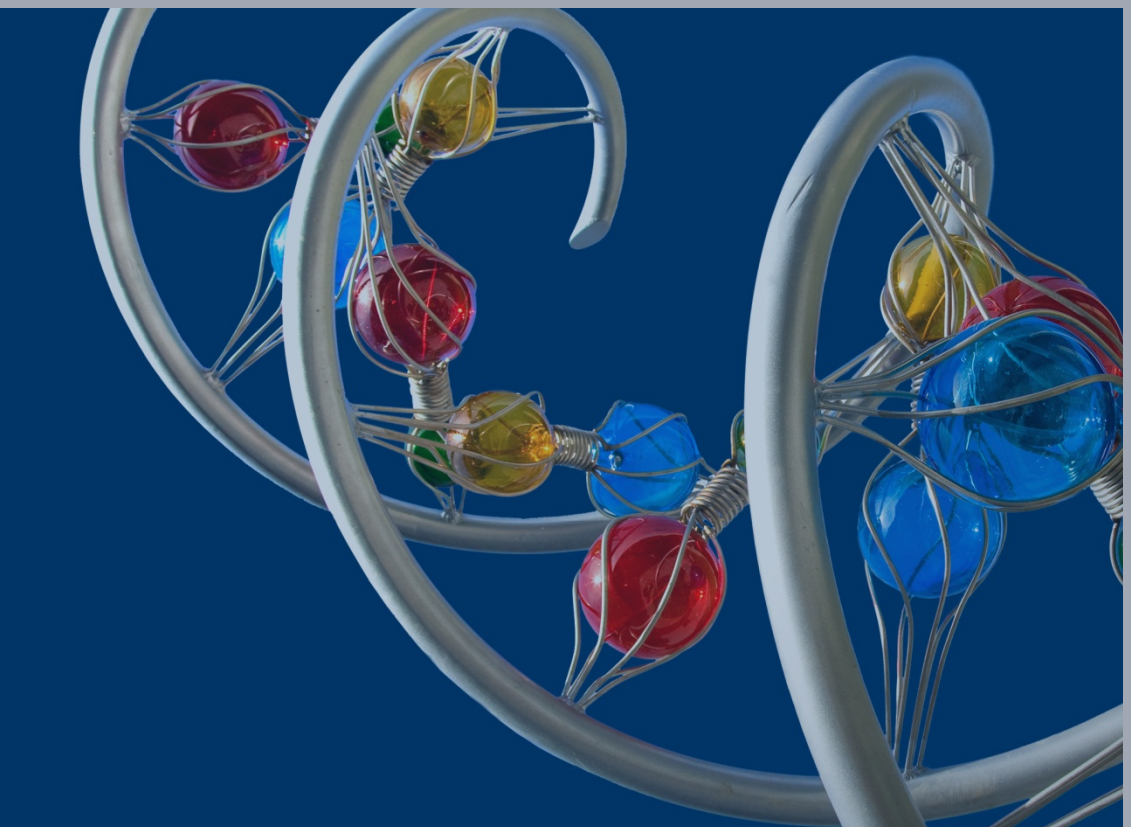
March 2013

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

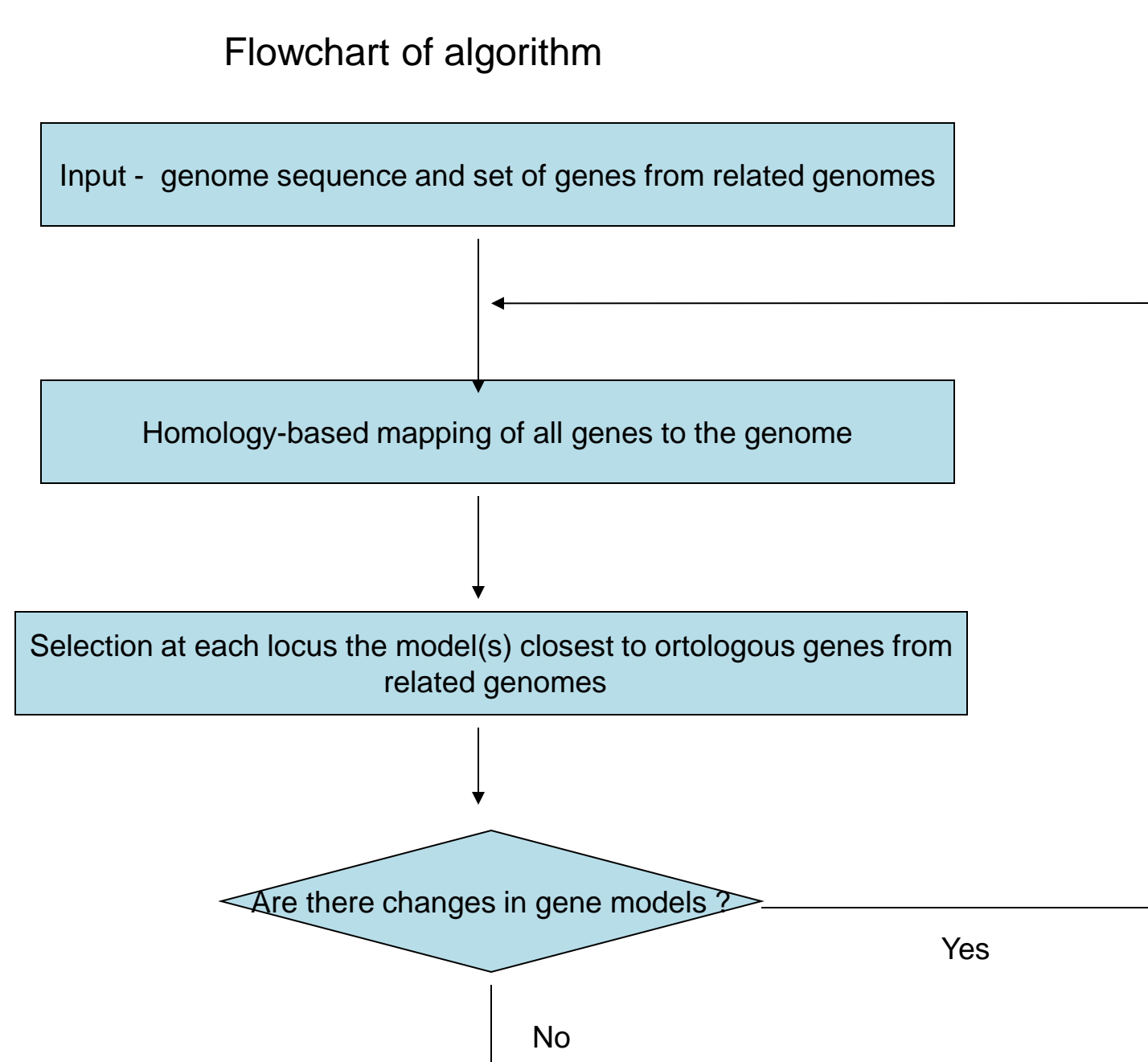
## **DISCLAIMER**

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.



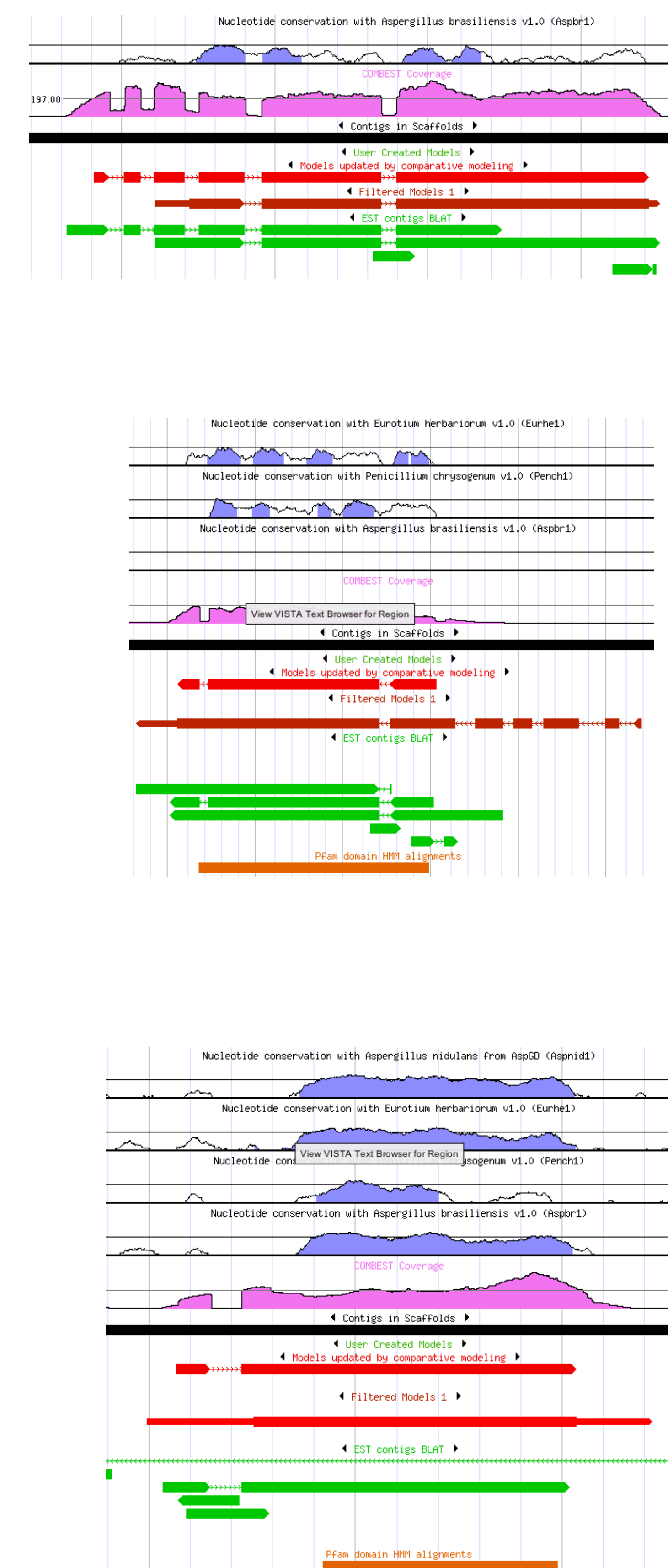


**Abstract**  
We used comparative gene modeling to reannotate 21 *Aspergillus* genomes. Initial automatic annotation of individual genomes may contain some errors of different nature, e.g. missing genes, incorrect exon-intron structures, 'chimeras', which fuse 2 or more real genes or alternatively splitting some real genes into 2 or more models. The main premise behind the comparative modeling approach is that for closely related genomes most orthologous families have the same conserved gene structure. The algorithm maps all gene models predicted in each individual *Aspergillus* genome to the other genomes and, for each locus, selects from potentially many competing models, the one which most closely resembles the orthologous genes from other genomes. This procedure is iterated until no further change in gene models is observed. For *Aspergillus* genomes we predicted in total 4503 new gene models (~2% per genome), supported by comparative analysis, additionally correcting ~18% of old gene models. This resulted in a total of 4065 more genes with annotated PFAM domains (~3% increase per genome). Analysis of a few genomes with EST/transcriptomics data shows that the new annotation sets also have a higher number of EST-supported splice sites at exon-intron boundaries.



Organism	Average	Max	Min
<i>Aspergillus aculeatus</i>	287	802	163
<i>Aspergillus brasiliensis</i>	275	844	107
<i>Aspergillus carbonarius</i>	521	1047	377
<i>Aspergillus clavatus</i>	192	603	56
<i>Aspergillus flavus</i>	512	1617	304
<i>Aspergillus acidus</i>	265	800	84
<i>Aspergillus fumigatus</i>	306	718	168
<i>Aspergillus glaucus</i>	204	668	63
<i>Aspergillus niger</i>	531	981	335
<i>Aspergillus nidulans</i>	217	687	56
<i>Aspergillus oryzae</i>	819	1617	603
<i>Aspergillus sydowii</i>	267	829	113
<i>Aspergillus terreus</i>	539	1028	398
<i>Aspergillus tubingensis</i>	275	811	108
<i>Aspergillus versicolor</i>	264	846	111
<i>Aspergillus wentii</i>	227	774	83
<i>Aspergillus zonatus</i>	195	628	63
<i>Eurotium herbariorum</i>	205	654	80
<i>Neosartorya fischerii</i>	261	726	116
<i>Penicillium chrysogenum</i>	212	728	80
<i>Thermoascus aurantiacus</i>	220	610	126
<b>All organisms</b>	<b>324</b>	<b>1617</b>	<b>56</b>

**Examples of *Aspergillus zonatus* models improved by reannotation**



**Mycocosm: <http://jgi.doe.gov/fungi>**

**Eurotiomycetes group page**

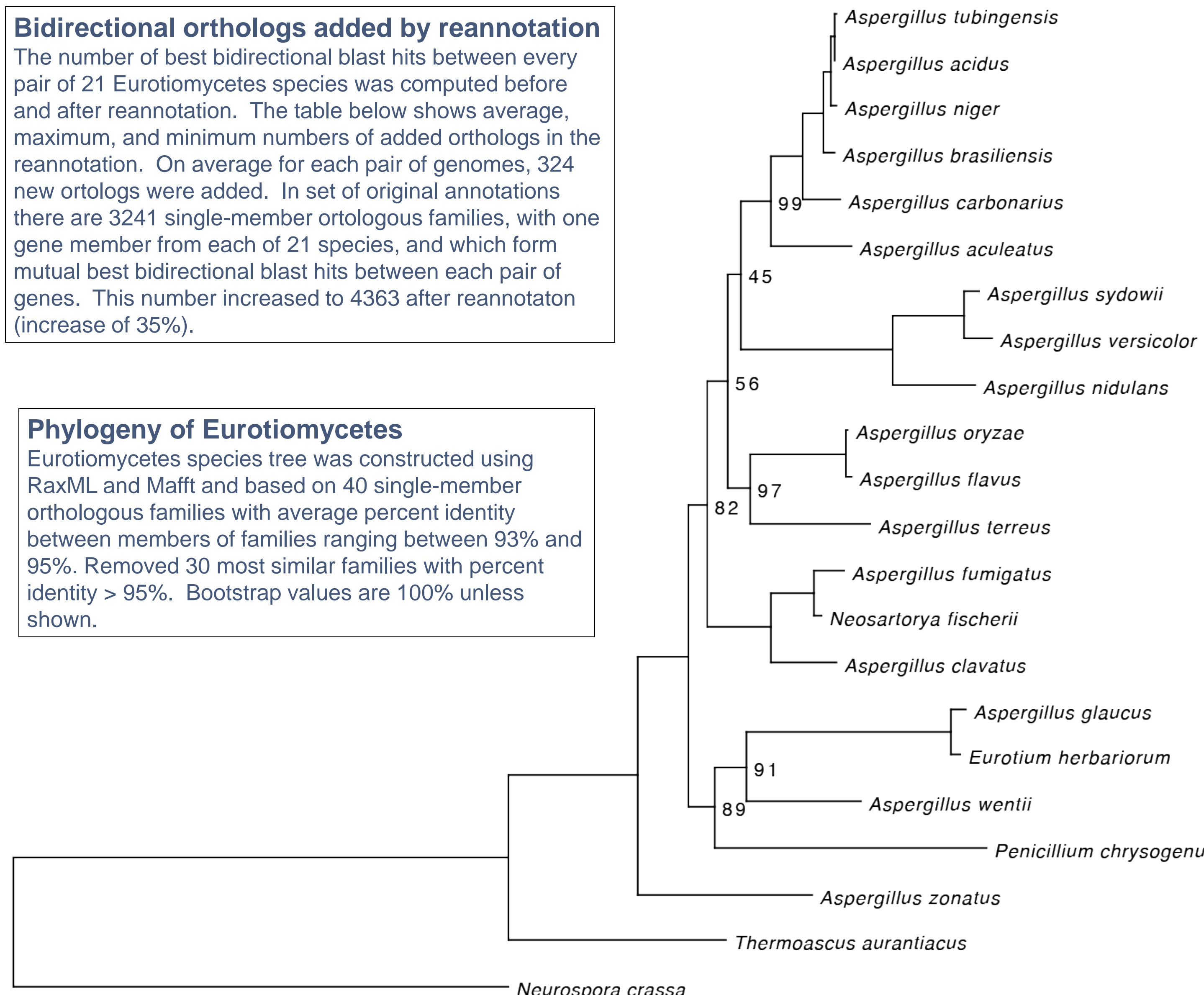
#	Organism	Name	Assembly length	# genes
1	Aspcc1	<i>Aspergillus aculeatus</i> ATCC16872 v1.1	35,424,414	10,845
2	Aspcc1	<i>Aspergillus brasiliensis</i> CBS 101740	35,808,783	13,000
3	Aspcc3	<i>Aspergillus carbonarius</i> ITEM 5010	36,290,756	11,434
4	Aspcc1	<i>Aspergillus clavatus</i> NRRL 1	27,859,441	9,121
5	Aspcc1	<i>Aspergillus flavus</i> NRRL 3557	36,790,245	12,604
6	Aspcc1	<i>Aspergillus foetidus</i>	37,468,345	13,530
7	Aspcc1	<i>Aspergillus fumigatus</i> A293	29,388,377	9,781
8	Aspcc1	<i>Aspergillus glaucus</i> CBS 516.65	27,993,362	11,283
9	Aspcc1	<i>Aspergillus nidulans</i> from AspGD	30,483,994	10,680
10	Aspcc5	<i>Aspergillus niger</i>	34,853,277	11,085
11	Aspcc1	<i>Aspergillus oryzae</i> RIB40	37,882,812	12,030
12	Aspcc1	<i>Aspergillus terreus</i> NIH 2624	34,381,026	13,620
13	Aspcc1	<i>Aspergillus tubingensis</i> CBS 134.48	29,331,195	10,406
14	Aspcc1	<i>Aspergillus versicolor</i> CBS 134.48	35,146,149	12,322
15	Aspcc1	<i>Aspergillus wentii</i> DTO 134E9	33,126,810	13,228
16	Aspcc1	<i>Aspergillus zonatus</i> DTO 134E9	31,350,635	12,442
17	Aspcc1	<i>Aspergillus aculeatus</i> ATCC16872 v1.1	28,924,210	11,362
18	Eurhe1	<i>Eurotium herbariorum</i>	26,209,327	10,076
19	Monru1	<i>Monascus ruber</i> NRRL 1597 v1.0	24,798,814	9,650
20	Neof1	<i>Neosartorya fischerii</i> NRRL 181	32,551,711	10,406
21	Pench1	<i>Penicillium chrysogenum</i> v1.0	31,340,922	11,396

**Bidirectional orthologs added by reannotation**

The number of best bidirectional blast hits between every pair of 21 Eurotiomycetes species was computed before and after reannotation. The table below shows average, maximum, and minimum numbers of added orthologs in the reannotation. On average for each pair of genomes, 324 new orthologs were added. In set of original annotations there are 3241 single-member orthologous families, with one gene member from each of 21 species, and which form mutual best bidirectional blast hits between each pair of genes. This number increased to 4363 after reannotation (increase of 35%).

**Phylogeny of Eurotiomycetes**

Eurotiomycetes species tree was constructed using RaxML and Mafft and based on 40 single-member orthologous families with average percent identity between members of families ranging between 93% and 95%. Removed 30 most similar families with percent identity > 95%. Bootstrap values are 100% unless shown.



**Functionalities available include:**

- Keyword search
- BLAST search against all *Aspergillus*
- Data download
- Comparative clustering

**Summary of improvements**

Organism	Old	New	Changed	Novel	Updated	Split	Joined	Added Pfams
<i>Aspergillus fumigatus</i>	9763	9994	2.3	2814	28.2	303	3	2406
<i>Aspergillus oryzae</i>	12023	12725	5.5	4673	36.7	1097	8.6	2864
<i>Aspergillus clavatus</i>	9121	9209	1	1629	17.7	156	1.7	1376
<i>Aspergillus flavus</i>	12603	13234	4.8	4566	34.5	423	3.2	3366
<i>Aspergillus terreus</i>	10403	10930	4.8	4341	39.7	465	4.3	3515
<i>Neosartorya fischerii</i>	10405	10759	3.3	2221	20.6	403	3.7	1717
<i>Aspergillus brasiliensis</i>	12998	13036	0.3	1814	13.9	134	1	1448
<i>Aspergillus acidus</i>	13529	13538	0.1	1986	14.7	128	0.9	1545
<i>Aspergillus glaucus</i>	11276	11295	0.2	1755	15.5	127	1.1	1357
<i>Aspergillus sydowii</i>	13619	13674	0.4	2525	18.5	130	1	2028
<i>Aspergillus tubingensis</i>	12322	12379	0.5	2024	16.4	151	1.2	1568
<i>Aspergillus versicolor</i>	13228	13298	0.5	2122	16	91	0.7	1668
<i>Aspergillus wentii</i>	12442	12428	-0.1	1796	14.5	75	0.6	1430
<i>Aspergillus zonatus</i>	11362	11304	-0.5	1740	15.4	39	0.3	1416
<i>Eurotium herbariorum</i>	10074	10158	0.8	1707	16.8	163	1.6	1325
<i>Penicillium chrysogenum</i>	11395	11469	0.6	2029	17.7	117	1	1823
<i>Thermoascus aurantiacus</i>	8798	8944	1.6	1565	17.5	157	1.8	1337
<i>Aspergillus aculeatus</i>	10828	11026	1.8	2289	20.8	201	1.8	1993
<i>Aspergillus carbonarius</i>	11624	12046	3.5	2478	20.6	474	3.9	1865
<i>Aspergillus nidulans</i>	10653	10683	0.3	2585	24.2	114	1.1	2352
<i>Aspergillus niger</i>	11189	11958	6.4	5038	42.1	800	6.7	4001
<b>Total</b>	<b>239655</b>	<b>244087</b>	<b>1.8</b>	<b>53697</b>	<b>22</b>	<b>5748</b>	<b>2.4</b>	<b>42400</b>

***Aspergilli* and comparative Eurotiomycetes in MycoCosm**

**Sequenced at JGI**      **Imported from AspGD**

<i>Aspergillus acidus</i> (formerly foetidus)	<i>Aspergillus clavatus</i> NRRL 1
<i>Aspergillus aculeatus</i> ATCC16872	<i>Aspergillus flavus</i> NRRL 3557
<i>Aspergillus brasiliensis</i> CBS 101740	<i>Aspergillus fumigatus</i> A293
<i>Aspergillus carbonarius</i> ITEM 5010	<i>Aspergillus nidulans</i>
<i>Aspergillus glaucus</i> CBS 516.65	<i>Aspergillus niger</i>
<i>Aspergillus sydowii</i>	<i>Aspergillus oryzae</i> RIB40
<i>Aspergillus tubingensis</i> CBS 134.48	<i>Aspergillus terreus</i> NIH 2624
<i>Aspergillus versicolor</i>	<i>Neosartorya fischerii</i> NRRL 181
<i>Aspergillus wentii</i> DTO 134E9	
<i>Aspergillus zonatus</i>	
<i>Eurotium herbariorum</i>	
<i>Monascus ruber</i> NRRL 1597	
<i>Penicillium chrysogenum</i>	<a href="http://www.aspgd.org/">http://www.aspgd.org/</a>