

Comparative Reannotation of 21 *Aspergillus* Genomes

Asaf Salamov¹, Robert Riley¹, Alan Kuo¹, Igor Grigoriev¹

¹DOE 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.

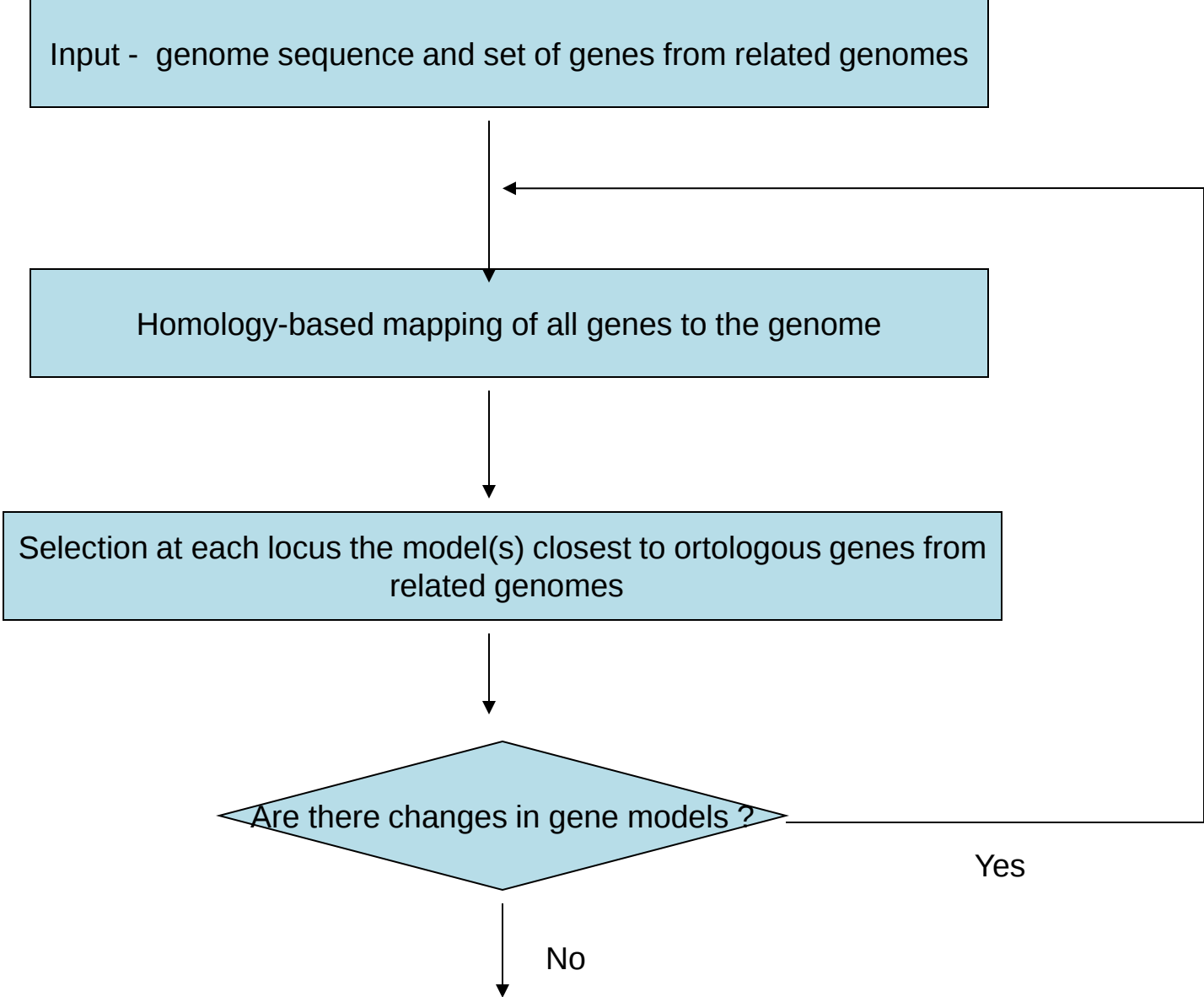
Comparative reannotation of 21 *Aspergillus* genomes

Asaf Salamov, Robert Riley, Alan Kuo, Igor V. Grigoriev
DOE Joint Genome Institute, Walnut Creek, CA, USA; aasalamov@lbl.gov

Abstract

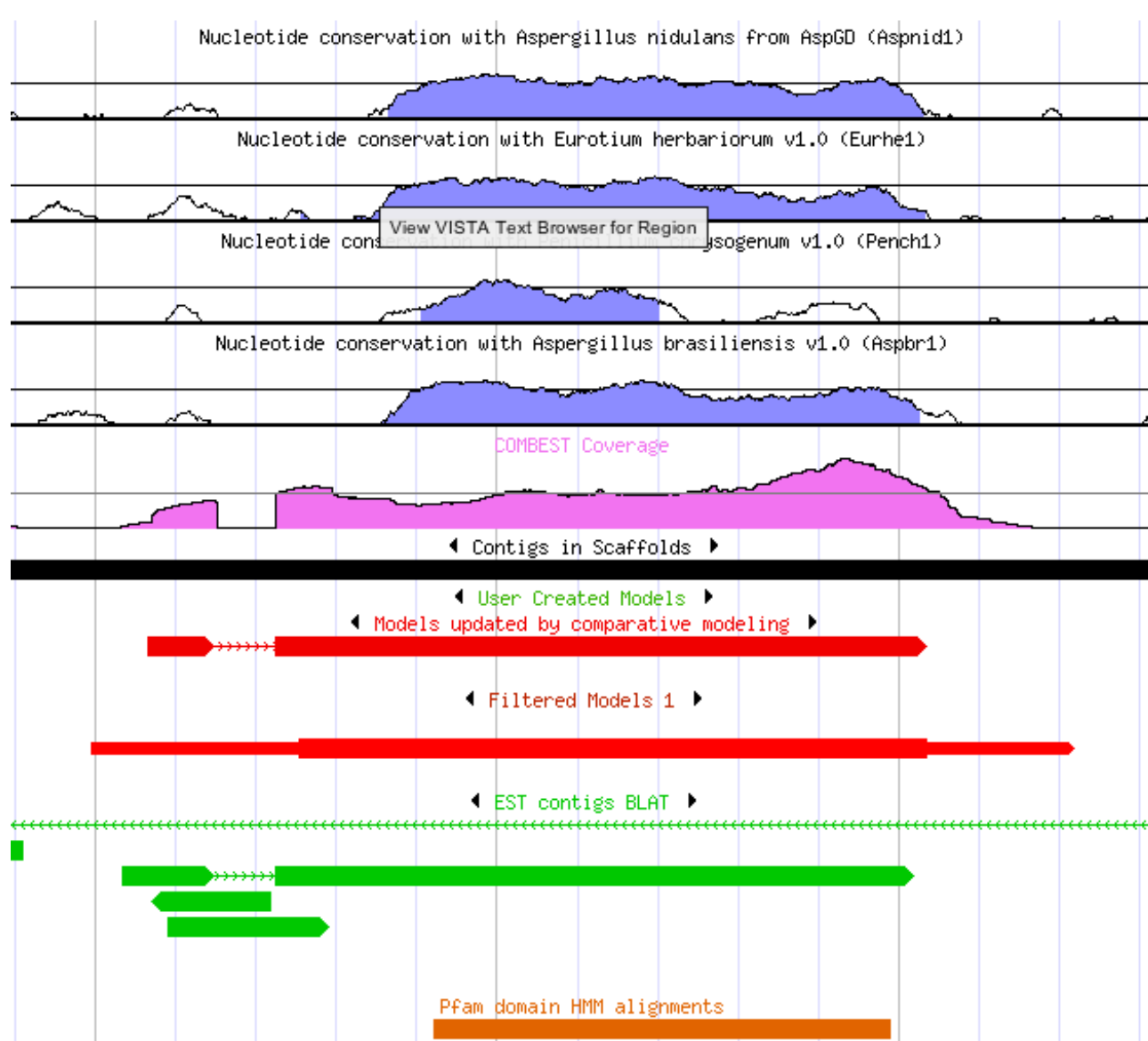
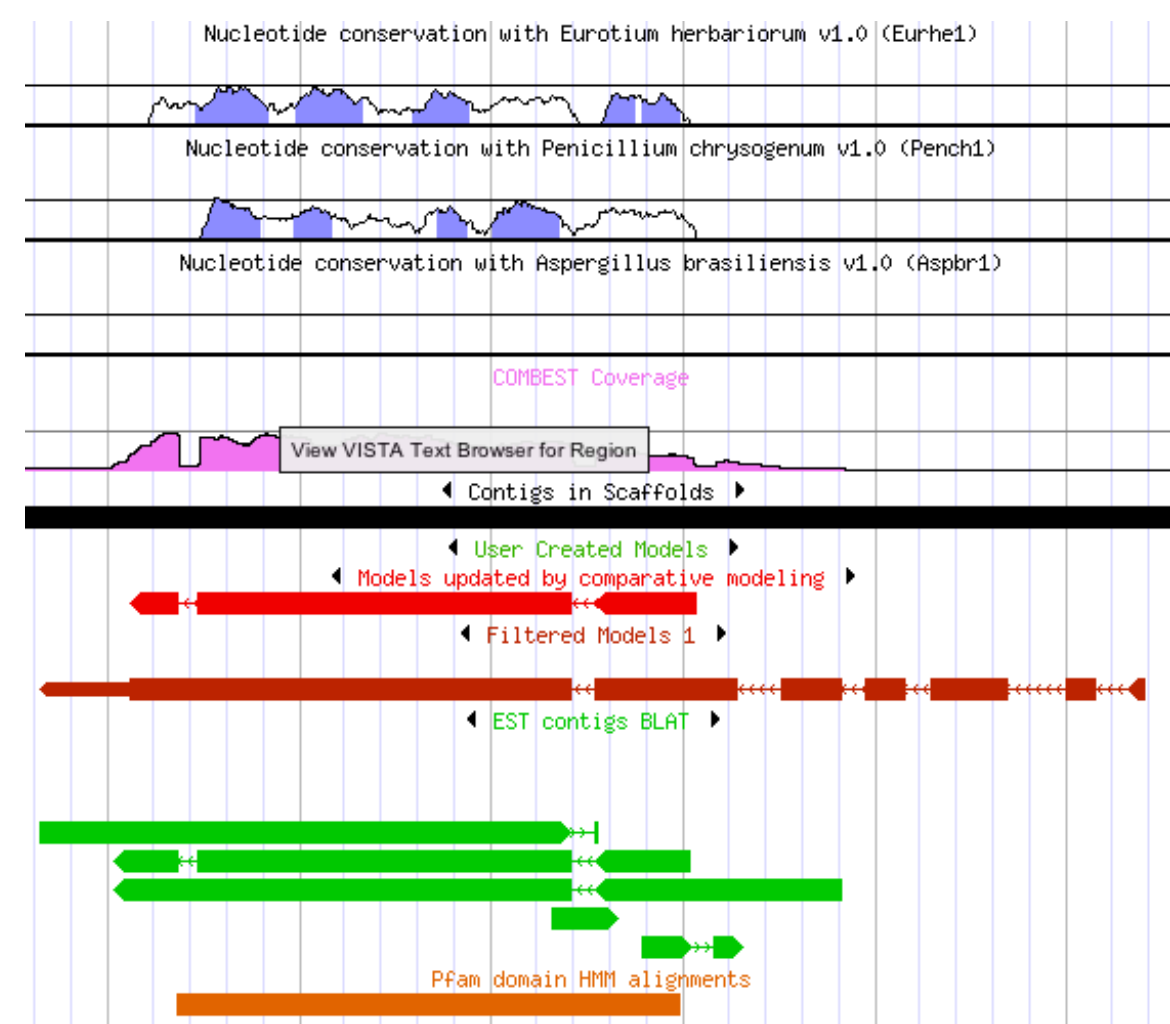
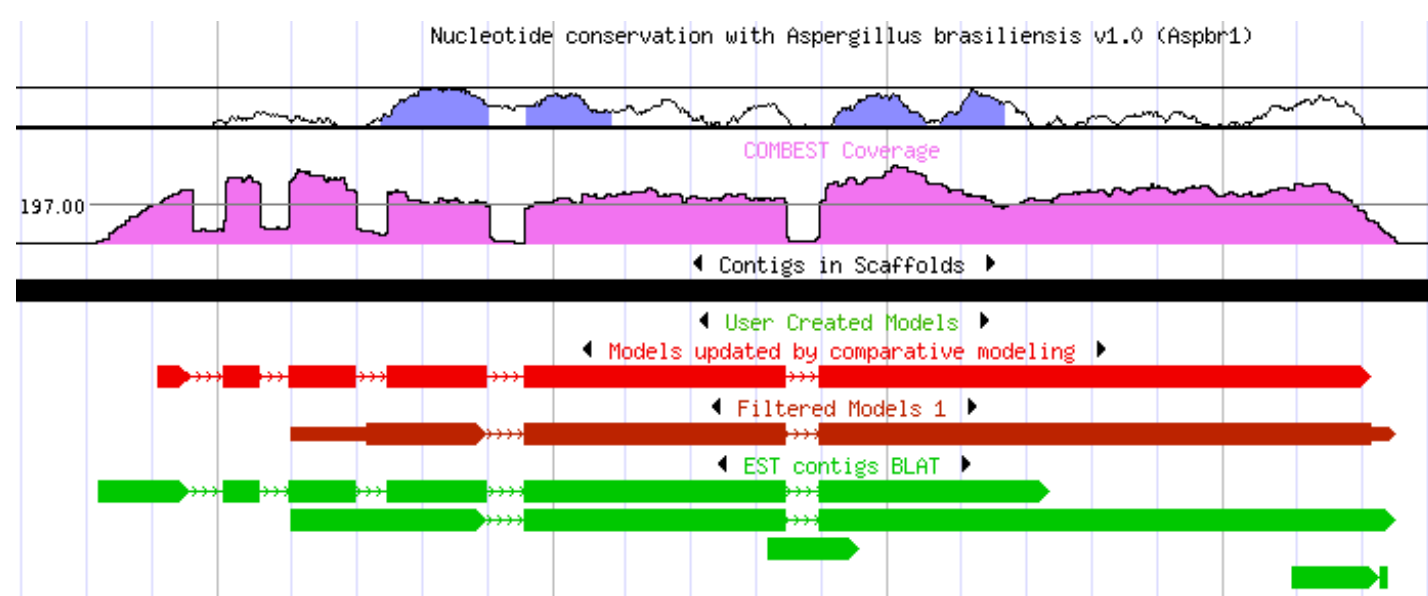
We used comparative gene modeling to reannotate 21 *Aspergillus* genomes. Initial automatic annotation of individual genomes may contain some errors of different nature, e.g. missing genes, incorrect exon-intron structures, 'chimeras', which fuse 2 or more real genes or alternatively splitting some real 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.

Flowchart of algorithm



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
All organisms	324	1617	56

Examples of *Aspergillus zonatus* models improved by reannotation



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

Eurotiomycetes group page

#	Organism	Name	Assembly length	# genes
1	Aspergillus aculeatus	ATCC16872 v1.1	35,424,414	10,845
2	Aspergillus brasiliensis	CBS 101740	35,608,783	13,000
3	Aspergillus carbonarius	ITEM 5010	36,290,756	11,624
4	Aspergillus clavatus	NRRL 1	27,859,441	9,121
5	Aspergillus flavus	NRRL 3557	36,790,245	12,604
6	Aspergillus foetidus		37,468,345	13,530
7	Aspergillus fumigatus	Af293	29,388,377	9,781
8	Aspergillus glaucus	CBS 516.65	27,993,362	11,283
9	Aspergillus nidulans	from AspGD	30,483,994	10,680
10	Aspergillus niger		34,853,277	11,085
11	Aspergillus oryzae	RIB40	37,882,812	12,030
12	Aspergillus sydowii		34,381,026	13,620
13	Aspergillus terreus	NIH 2624	29,331,195	10,406
14	Aspergillus tubingensis	CBS 134.48	35,146,149	12,322
15	Aspergillus versicolor		33,126,810	13,228
16	Aspergillus wentii	DTO 134E9	31,350,635	12,442
17	Aspergillus zonatus		28,924,210	11,362
18	Eurotium herbariorum		26,209,327	10,076
19	Monascus ruber	NRRL 1597 v1.0	24,798,814	9,650
20	Neosartorya fischerii	NRRL 181	32,551,711	10,406
21	Penicillium chrysogenum	v1.0	31,340,922	11,396

Functionalities available include:

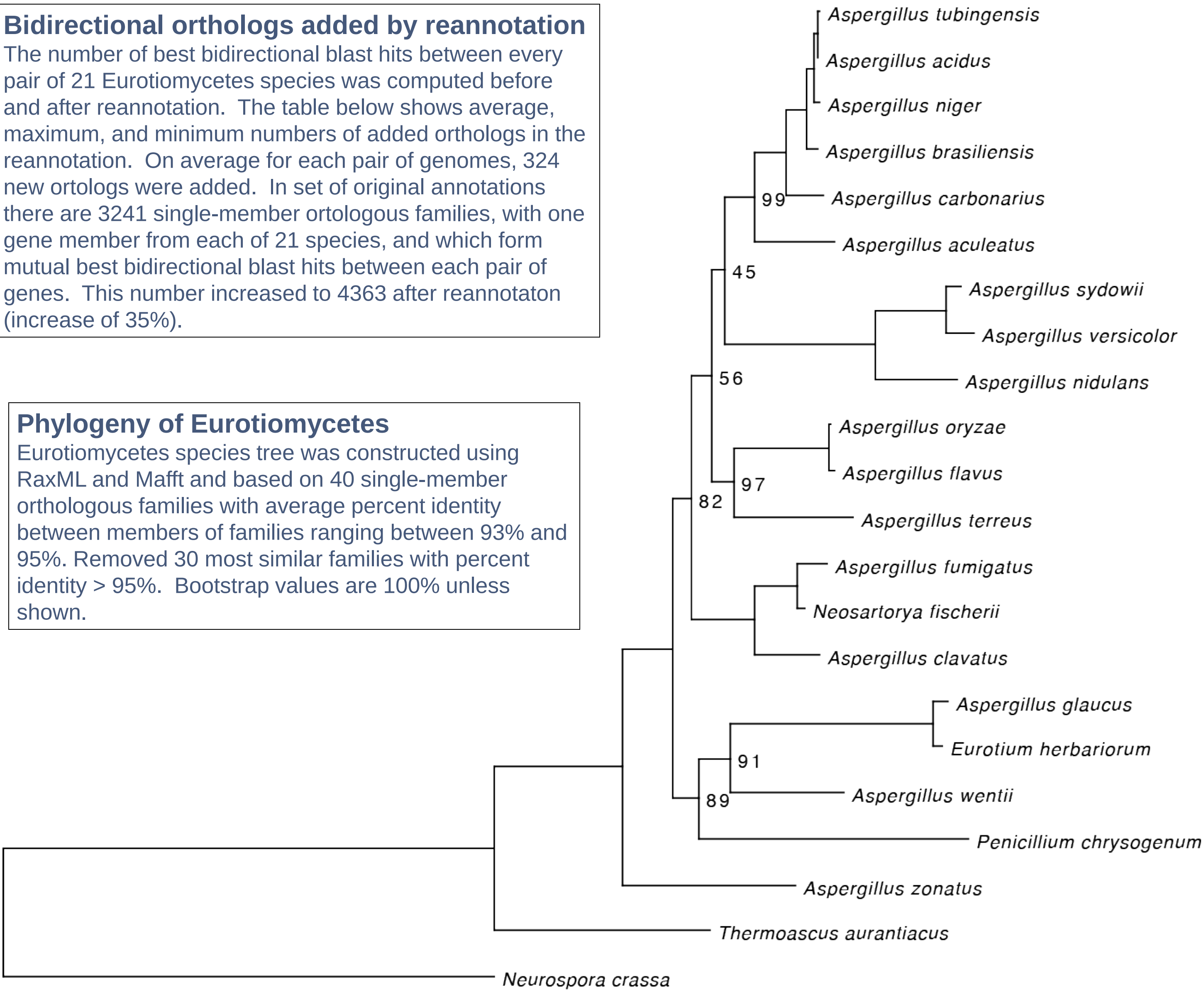
- Keyword search
- BLAST search against all *Aspergilli*
- Data download
- Comparative clustering

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.



Summary of improvements

Organism	Old	New	Changed		Novel		Updated		Split		Joined		Added Pfams			
<i>Aspergillus fumigatus</i>	9763	9994	9994	2.3	2814	28.2	303	3	2406	24.1	26	0.3	79	0.8	183	2.9
<i>Aspergillus oryzae</i>	12023	12725	12725	5.5	4673	36.7	1097	8.6	2864	22.5	276	2.2	436	3.4	843	11.9
<i>Aspergillus clavatus</i>	9121	9209	9209	1	1629	17.7	156	1.7	1376	14.9	22	0.2	75	0.8	92	1.6
<i>Aspergillus flavus</i>	12603	13234	13234	4.8	4566	34.5	423	3.2	3366	25.4	685	5.2	92	0.7	622	8.4
<i>Aspergillus terreus</i>	10403	10930	10930	4.8	4341	39.7	465	4.3	3515	32.2	292	2.7	69	0.6	618	9.6
<i>Neosartorya fischerii</i>	10405	10759	10759	3.3	2221	20.6	403	3.7	1717	16	38	0.4	63	0.6	285	4.4
<i>Aspergillus brasiliensis</i>	12998	13036	13036	0.3	1814	13.9	134	1	1448	11.1	111	0.9	121	0.9	115	1.5
<i>Aspergillus acidus</i>	13529	13538	13538	0.1	1986	14.7	128	0.9	1545	11.4	161	1.2	152	1.1	118	1.6
<i>Aspergillus glaucus</i>	11276	11295	11295	0.2	1755	15.5	127	1.1	1357	12	146	1.3	125	1.1	104	1.7
<i>Aspergillus sydowii</i>	13619	13674	13674	0.4	2525	18.5	130	1	2028	14.8	241	1.8	126	0.9	171	2.1
<i>Aspergillus tubingensis</i>	12322	12379	12379	0.5	2024	16.4	151	1.2	1568	12.7	175	1.4	130	1.1	117	1.6
<i>Aspergillus versicolor</i>	13228	13298	13298	0.5	2122	16	91	0.7	1668	12.5	269	2	94	0.7	146	1.8
<i>Aspergillus wentii</i>	12442	12428	12428	-0.1	1796	14.5	75	0.6	1430	11.5	179	1.4	112	0.9	114	1.5
<i>Aspergillus zonatus</i>	11362	11304	11304	-0.5	1740	15.4	39	0.3	1416	12.5	163	1.4	122	1.1	69	0.9
<i>Eurotium herbariorum</i>	10074	10158	10158	0.8	1707	16.8	163	1.6	1325	13	121	1.2	98	1	124	2.1
<i>Penicillium chrysogenum</i>	11395	11469	11469	0.6	2029	17.7	117	1	1823	15.9	38	0.3	51	0.4	112	1.7
<i>Thermoascus aurantiacus</i>	8798	8944	8944	1.6	1565	17.5	157	1.8	1337	14.9	42	0.5	29	0.3	164	3.1
<i>Aspergillus aculeatus</i>	10828	11026	11026	1.8	2289	20.8	201	1.8	1993	18.1	66	0.6	29	0.3	217	3.2
<i>Aspergillus carbonarius</i>	11624	12046	12046	3.5	2478	20.6	474	3.9	1865	15.5	62	0.5	77	0.6	343	4.8
<i>Aspergillus nidulans</i>	10653	10683	10683	0.3	2585	24.2	114	1.1	2352	22	32	0.3	87	0.8	140	2.1
<i>Aspergillus niger</i>	11189	11958	11958	6.4	5038	42.1	800	6.7	4001	33.5	150	1.3	87	0.7	465	6.6
Total	239655	244087	244087	1.8	53697	22	5748	2.4	42400	17.4	3295	1.3	2254	0.9	5162	3.6