| *Species* | Current name in GenBank | GenBank id |
|---|---|---|
| *A. nidulans* | Aspergillus nidulans FGSC A4 | CH236920 |
| *A. niger* | Aspergillus niger CBS 513.88 | AM270980 |
| *A. terreus* | Aspergillus terreus NIH2624 | CH981535 |
| *B. cinerea* | Botryotinia fuckeliana B05.10 | BS264920 |
| *C. immitis* | Coccidioides immitis RS | CH671914 |
| *C. cinereus* | Coprinopsis cinerea okayama7#130 | AACS00000000 |
| *C. neoformans* | Cryptococcus neoformans R265 | AAFP00000000 |
| *F. graminearum* | Gibberella zeae PH-1 | AACM00000000 |
| *F. oxysporum* | Fusarium oxysporum f. sp. lycopersici FGSC 4286 | AAXH00000000 |
| *F. verticillioides* | Gibberella moniliformis 7600 | AAIM00000000 |
| *M. grisea* | Magnaporthe grisea 70-15 | AACU00000000 |
| *N. crassa* | Neurospora crassa OR74A | AABX00000000 |
| *R. oryzae* | Rhizopus oryzae RA 99-880 | AACW00000000 |
| *S. pombe* | Schizosaccharomyces pombe 972h- | CU329670 |
| *S. sclerotiorum* | Sclerotinia sclerotiorum 1980 | AAGT00000000 |
| *S. nodorum* | Phaeosphaeria nodorum SN15 | AAGI00000000 |

**Table S1**. Nomenclature of the 16 fungal genomes.

| *Species* | number of ESTs | Kb |
|---|---|---|
| *A. nidulans* | 16,848 | 10,480 |
| *A. niger* | 15,473 | 10,443 |
| *A. terreus* | 61,051 | 48,256 |
| *B. cinerea* | 11,076 | 11,107 |
| *C. immitis* | 65,754 | 55,777 |
| *C. cinereus* | 15,715 | 14,020 |
| *C. neoformans* | 14,211 | 11,060 |
| *F. graminearum* | 21,325 | 14,829 |
| *F. oxysporum* | 9,248 | 5,454 |
| *F. verticillioides* | 87,086 | 75,347 |
| *M. grisea* | 53,102 | 30,299 |
| *N. crassa* | 28,089 | 14,658 |
| *S. pombe* | 8,131 | 4,720 |
| *S. sclerotiorum* | 1,494 | 1,260 |
| *S. nodorum* | 16,014 | 10,784 |

**Table S2.** Volumes of the EST sequence data used to generate the test sets for assessment of the algorithms performance (Table S5). The test set generation was done by EST to genomic DNA mapping (A. Kislyuk, A. Lomsadze, M. Boroodvsky, unpublished)

| | Iteration 6 | | | Iteration 7 | | |
|---|---|---|---|---|---|---|
| | Instances of upper path transition | Instances of lower path transition | % of upper path transition | Instances of upper path transition | Instances of lower path transition | % of upper path transition |
| *A. nidulans* | 17,336 | 699 | 96.1 | 17,553 | 759 | 95.9 |
| *A. niger* | 22,368 | 1,069 | 95.4 | 22,592 | 1,172 | 95.1 |
| *A. terreus* | 21,263 | 1,150 | 94.9 | 21,605 | 1,132 | 95.0 |
| *B. cinerea* | 13,454 | 1,028 | 92.9 | 13,569 | 1,077 | 92.6 |
| *C. immitis* | 13,073 | 762 | 94.5 | 13,382 | 783 | 94.5 |
| *C. cinereus* | 50,321 | 3,079 | 94.2 | 53,454 | 3,313 | 94.2 |
| *C. neoformans* | 29,124 | 1,092 | 96.4 | 30,085 | 1,200 | 96.2 |
| *F. graminearum* | 5,982 | 261 | 95.8 | 6,025 | 264 | 95.8 |
| *F. oxysporum* | 31,355 | 2,127 | 93.6 | 31,408 | 2,246 | 93.3 |
| *F. verticillioides* | 24,296 | 1,034 | 95.9 | 24,359 | 1,149 | 95.5 |
| *M. grisea* | 12,568 | 5,576 | 69.3 | 14,109 | 4,120 | 77.4 |
| *N. crassa* | 12,945 | 1,264 | 91.1 | 12,880 | 1,358 | 90.5 |
| *R. oryzae* | 15,152 | 28,842 | 34.4 | 10,238 | 34,546 | 22.9 |
| *S. pombe* | 4,712 | 226 | 95.4 | 4,718 | 227 | 95.4 |
| *S. sclerotiorum* | 19,057 | 1,293 | 93.6 | 19,194 | 1,310 | 93.6 |
| *S. nodorum* | 19,895 | 1,385 | 93.5 | 20,083 | 1,403 | 93.5 |

**Table S3.** Counts of traversing of the Viterbi paths computed by the algorithm through upper and lower branches of the enhanced intron sub-model (Fig. 2).

| **set I** | genes | introns per gene |
|---|---|---|
| *C. cinereus* | 167 | 3.4 |
| *C. immitis* | 432 | 2.3 |
| *F. verticillioides* | 327 | 2.0 |
| *M. grisea* | 169 | 2.0 |
| *S. pombe* | 1,277 | 3.1 |

| **set II** | transcripts | introns per transcript |
|---|---|---|
| *A. nidulans* | 1,075 | 2.6 |
| *A. niger* | 955 | 2.8 |
| *A. terreus* | 729 | 2.8 |
| *B. cinerea* | 787 | 2.7 |
| *C. neoformans* | 2,425 | 3.8 |
| *F. graminearum* | 919 | 2.6 |
| *F. oxysporum* | 461 | 2.5 |
| *N. crassa* | 276 | 2.5 |
| *R. oryzae* | 2,169 | 3.3 |
| *S. nodorum* | 413 | 2.7 |
| *S. sclerotiorum* | 587 | 2.9 |

**Table S4.** Size of the test sets and average intron density; sets of type I consist of complete genes; sets of type II include both complete and incomplete genes (neither set contains single exon genes).

| *species* | repeats (nt) | | | | % of all repetitive sequences | | | repeats found in coding regions as % of total size of predicted coding regions | % of total genome size | genome size (MB) |
|---|---|---|---|---|---|---|---|---|---|---|
| | in intergenic regions | in coding regions | in introns | total (nt) | in intergenic regions | in coding regions | in introns | | | |
| *A. nidulans* | 638,497 | 186,062 | 34,227 | 858,786 | 74.3 | 21.7 | 4.0 | 1.3 | 2.8 | 31 |
| *A. niger* | 141,935 | 67,487 | 9,306 | 218,728 | 64.9 | 30.9 | 4.3 | 0.4 | 0.6 | 34 |
| *A. terreus* | 135,794 | 18,713 | 4,159 | 158,666 | 85.6 | 11.8 | 2.6 | 0.1 | 0.5 | 29 |
| *B. cinerea* | 272,539 | 100,341 | 8,753 | 381,633 | 71.4 | 26.3 | 2.3 | 0.8 | 1.5 | 26 |
| *C. immitis* | 379,493 | 51,969 | 54,726 | 486,188 | 78.1 | 10.7 | 11.3 | 0.5 | 1.7 | 29 |
| *C. cinereus* | 51,643 | 387,201 | 24,847 | 463,691 | 11.1 | 83.5 | 5.4 | 2.0 | 1.2 | 38 |
| *C. neoformans* | 194,860 | 139,299 | 27,505 | 361,664 | 53.9 | 38.5 | 7.6 | 1.3 | 1.8 | 20 |
| *F. graminearum* | 99,662 | 40,156 | 2,228 | 142,046 | 70.2 | 28.3 | 1.6 | 0.6 | 0.4 | 40 |
| *F. oxysporum* | 764,914 | 1,276,899 | 88,266 | 2,130,079 | 35.9 | 59.9 | 4.1 | 4.6 | 3.6 | 60 |
| *F. verticillioides* | 72,123 | 20,471 | 4,298 | 96,892 | 74.4 | 21.1 | 4.4 | 0.1 | 0.2 | 42 |
| *M. grisea* | 783,116 | 1,518,185 | 163,183 | 2,464,484 | 31.8 | 61.6 | 6.6 | 8.9 | 6.2 | 40 |
| *N. crassa* | 682,978 | 95,603 | 277,416 | 1,055,997 | 64.7 | 9.1 | 26.3 | 0.6 | 2.7 | 39 |
| *R. oryzae* | 104,687 | 660,911 | 40,280 | 805,878 | 13.0 | 82.0 | 5.0 | 3.7 | 2.0 | 40 |
| *S. pombe* | 133,175 | 66,816 | 14,231 | 214,222 | 62.2 | 31.2 | 6.6 | 0.8 | 1.8 | 12 |
| *S. sclerotiorum* | 436,924 | 301,321 | 25,777 | 764,022 | 57.2 | 39.4 | 3.4 | 1.8 | 2.0 | 39 |
| *S. nodorum* | 309,373 | 33,960 | 31,208 | 374,541 | 82.6 | 9.1 | 8.3 | 0.2 | 1.0 | 37 |

**Table S5.** Statistics of the content of repetitive sequences determined by RepeatMasker in protein-coding and non-coding regions (as predicted by GeneMark.hmm-ES) determined in the 16 fungi genomes.

| | | *A. nidulans* | | | *A. niger* | | | *A. terreus* | | | *B. cinerea* | | | *C. neoformans* | | | *F. graminearum* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | intron submodel | | | intron submodel | | | intron submodel | | | intron submodel | | | intron submodel | | | intron submodel | | |
| | | original | new | δ | original | new | δ | original | new | δ | original | new | δ | original | new | δ | original | new | δ |
| Internal exon | Sn | 77.3 | **87.4** | 10.1 | 85.0 | **91.5** | 6.5 | 85.5 | **91.6** | 6.1 | 79.5 | **87.9** | 8.4 | 85.7 | **92.3** | 6.6 | 88.6 | **92.6** | 4.0 |
| | Sp | 90.5 | **93.1** | 2.6 | 91.4 | **96.3** | 4.9 | 90.9 | **94.8** | 3.9 | 91.4 | **96.5** | 5.1 | 91.1 | **95.1** | 4.0 | 93.6 | **95.9** | 2.3 |
| Intron | Sn | 81.1 | **89.0** | 7.9 | 86.2 | **91.7** | 5.5 | 88.2 | **92.7** | 4.5 | 84.7 | **89.8** | 5.1 | 86.8 | **92.4** | 5.6 | 90.5 | **93.5** | 3.0 |
| | Sp | 93.1 | **96.4** | 3.3 | 93.4 | **96.8** | 3.4 | 94.5 | **97.4** | 2.9 | 94.1 | **96.7** | 2.6 | 93.0 | **96.0** | 3.0 | 96.0 | **97.5** | 1.5 |
| Donor | Sn | 84.9 | **90.5** | 5.6 | 90.1 | **92.9** | 2.8 | 90.7 | **93.7** | 3.0 | 88.4 | **91.1** | 2.7 | 91.3 | **94.6** | 3.3 | 93.2 | **94.7** | 1.5 |
| | Sp | 95.6 | **96.8** | 1.2 | 96.2 | **97.3** | 1.1 | 96.1 | **97.7** | 1.6 | 97.1 | **97.3** | 0.2 | 96.4 | **97.4** | 1.0 | 97.8 | **97.9** | 0.1 |
| Acceptor | Sn | 83.8 | **91.4** | 7.6 | 89.3 | **94.2** | 4.9 | 90.2 | **94.4** | 4.2 | 87.0 | **92.4** | 5.4 | 88.7 | **94.0** | 5.3 | 92.0 | **95.6** | 3.6 |
| | Sp | 94.5 | **97.5** | 3.0 | 95.1 | **98.5** | 3.4 | 95.2 | **97.9** | 2.7 | 95.4 | **98.6** | 3.2 | 94.4 | **97.2** | 2.8 | 96.5 | **98.5** | 2.0 |

| | | *R. oryzae* | | | *F. oxysporum* | | | *N. crassa* | | | *S. sclerotiorum* | | | *S. nodorum* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | intron submodel | | | intron submodel | | | intron submodel | | | intron submodel | | | intron submodel | | |
| | | original | new | δ | original | new | δ | original | new | δ | original | new | δ | original | new | δ |
| Internal exon | Sn | 88.7 | **88.8** | 0.1 | 84.1 | **92.5** | 8.4 | 81.2 | **85.2** | 4.0 | 82.6 | **90.2** | 7.6 | 82.8 | **88.5** | 5.7 |
| | Sp | 94.3 | **94.7** | 0.4 | 87.8 | **90.6** | 2.8 | 92.0 | **95.6** | 3.6 | 91.3 | **94.1** | 2.8 | 90.7 | **94.8** | 4.1 |
| Intron | Sn | 88.8 | **88.9** | 0.1 | 86.7 | **91.3** | 4.6 | 85.9 | **88.6** | 2.7 | 86.3 | **91.3** | 5.0 | 87.3 | **90.8** | 3.5 |
| | Sp | 95.9 | **95.9** | 0.0 | 94.0 | **94.8** | 0.8 | 94.8 | **97.0** | 2.2 | 94.7 | **96.4** | 1.7 | 94.9 | **97.2** | 2.3 |
| Donor | Sn | 91.3 | **91.4** | 0.1 | 89.3 | **93.4** | 4.1 | 88.4 | **89.6** | 1.2 | 90.5 | **93.5** | 3.0 | 90.4 | **92.4** | 2.0 |
| | Sp | 97.0 | **97.2** | 0.2 | 95.0 | **95.5** | 0.5 | 96.7 | **97.6** | 0.9 | **97.5** | 97.4 | -0.1 | 96.6 | **97.6** | 1.0 |
| Acceptor | Sn | 90.3 | **90.4** | 0.1 | 89.3 | **94.3** | 5.0 | 88.9 | **91.3** | 2.4 | 88.2 | **93.8** | 5.6 | 89.8 | **93.3** | 3.5 |
| | Sp | 96.7 | **96.8** | 0.1 | 95.4 | **96.6** | 1.2 | 96.8 | **98.7** | 1.9 | 95.5 | **97.8** | 2.3 | 96.1 | **98.2** | 2.1 |

**Table S6.** Accuracy of prediction of gene structure elements. The Sn and Sp values were determined for the test sets of incomplete genes.

| *Species* | donor | | | branch point | | | acceptor | | | spacer | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | self-training | alignment | δ | self-training | alignment | δ | self-training | alignment | δ | self-training | alignment | δ |
| *A. niger* | 8.0 | 7.8 | 0.2 | 7.3 | 7.6 | -0.3 | 5.1 | 5.0 | 0.1 | 2.1 | 1.7 | 0.4 |
| *A.nidulans* | 7.7 | 7.6 | 0.1 | 7.3 | 7.4 | -0.1 | 5.0 | 5.0 | 0.0 | 2.0 | 1.8 | 0.2 |
| *A. terreus* | 7.9 | 7.7 | 0.2 | 7.5 | 8.0 | -0.5 | 5.1 | 5.1 | 0.0 | 2.1 | 2.1 | 0.0 |
| *B. cinerea* | 7.9 | 8.2 | -0.3 | 7.4 | 8.2 | -0.8 | 5.0 | 5.1 | -0.1 | 2.2 | 2.4 | -0.2 |
| *C. immitis* | 7.8 | 7.4 | 0.4 | 7.2 | 7.0 | 0.2 | 5.3 | 5.0 | 0.3 | 1.9 | 1.4 | 0.5 |
| *C. cinereus* | 7.9 | 7.8 | 0.1 | 5.7 | 6.3 | -0.6 | 5.3 | 5.3 | 0.0 | 1.1 | 0.9 | 0.2 |
| *C. neoformans* | 8.5 | 7.1 | 1.4 | 6.7 | 5.9 | 0.8 | 5.1 | 5.1 | 0.0 | 1.8 | 1.8 | 0.0 |
| *F. graminearum* | 8.4 | 8.6 | -0.2 | 7.6 | 8.3 | -0.7 | 5.0 | 5.0 | 0.0 | 2.3 | 2.5 | -0.2 |
| *F. oxysporum* | 7.5 | 8.7 | -1.2 | 7.2 | 8.1 | -0.9 | 4.8 | 5.6 | -0.8 | 1.7 | 2.3 | -0.6 |
| *F. verticillioides* | 8.2 | 8.3 | -0.1 | 7.5 | 7.8 | -0.3 | 4.9 | 5.2 | -0.3 | 2.0 | 2.3 | -0.3 |
| *M. grisea* | 7.9 | 8.5 | -0.6 | 7.5 | 8.2 | -0.7 | 4.9 | 5.3 | -0.4 | 1.1 | 1.6 | -0.5 |
| *N. crassa* | 8.7 | 8.5 | 0.2 | 8.3 | 8.2 | 0.1 | 5.1 | 5.3 | -0.2 | 2.4 | 1.7 | 0.7 |
| *R. oryzae* | 7.1 | 5.4 | 1.7 | 4.0 | 4.1 | -0.1 | 5.1 | 6.4 | -1.3 | 0.3 | 0.8 | -0.5 |
| *S. sclerotiorum* | 7.8 | 8.2 | -0.4 | 7.3 | 7.8 | -0.5 | 5.0 | 5.2 | -0.2 | 2.0 | 2.7 | -0.7 |
| *S. pombe* | 8.6 | 9.2 | -0.6 | 7.6 | 7.8 | -0.2 | 5.4 | 7.2 | -1.8 | 1.8 | 1.8 | 0.0 |
| *S. nodorum* | 7.6 | 8.5 | -0.9 | 7.2 | 7.8 | -0.6 | 4.8 | 5.3 | -0.5 | 2.1 | 2.2 | -0.1 |

**Table S7.** Relative entropies of the first order models of donor, branch point and acceptor sites as well as the length distributions of the downstream spacers derived from the sets of intron determined by i/ the self-training algorithm and ii/ EST to genome alignment. Differences between the values derived by different methods are shown in columns with label δ.

| *Species* | **Programs used for gene prediction** |
|---|---|
| *A. nidulans* | Fgenesh, Fgenesh+ and Geneid |
| *A. niger* | Unknown |
| *A. terreus* | Fgenesh, Fgenesh+ and Geneid |
| *B. cinerea* | Fgenesh & Geneid |
| *C. immitis* | Fgenesh, Geneid and GENEWISE |
| *C. cinereus* | AUGUSTUS, GeneZilla; SNAP |
| *C. neoformans* | GENEWISE, TWINSCAN, GLEAN |
| *F. graminearum* | Fgenesh and Geneid |
| *F. oxysporum* | Fgenesh and Geneid |
| *F. verticillioides* | Fgenesh and Geneid |
| *M. grisea* | Fgenesh, Geneid and GENEWISE. |
| *N. crassa* | Fgenesh, Geneid, and GENEWISE |
| *R. oryzae* | Fgenesh and Geneid. Fgenesh |
| *S. pombe* | Unknown |
| *S. sclerotiorum* | Fgenesh and Geneid |
| *S. nodorum* | Fgenesh, Fgenesh+, Geneid, GENEWISE |

**Table S8.** Intrinsic (*ab initio*) and extrinsic gene finding methods used to produce annotation of the 16 fungal genomes