

Figure S1. Taxonomic relationships of the fungal species under consideration (www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy).

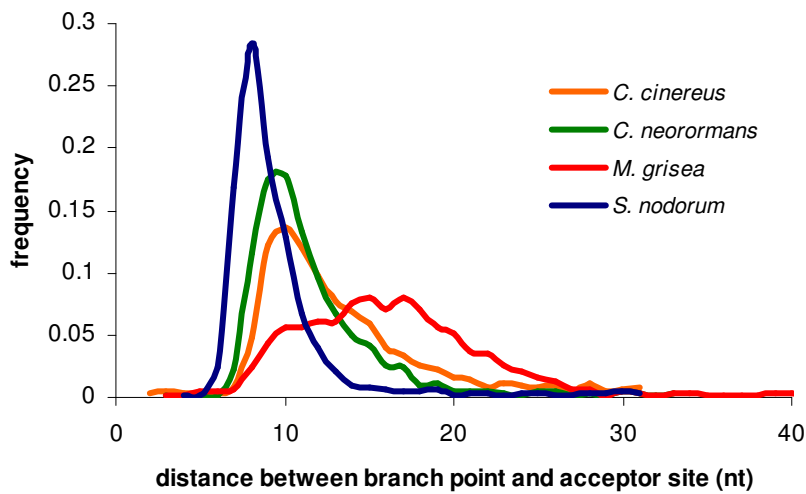


Figure S2. Length distributions of sequences between branch point and acceptor site determined for four fungal species at the final iteration of the algorithm.

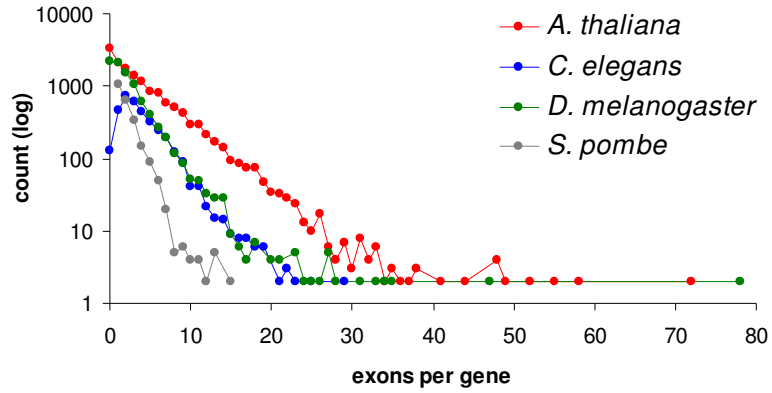


Figure S3. Counts of genes with a given number of exons per gene as calculated i/ from sets of genes confirmed by cDNA (*A. thaliana*, *C. elegans*); ii/ all annotated genes (*D. melanogaster*); iii/ a set of genes with protein product showing full length similarity to a protein in the SwissProt database (*S. pombe*).

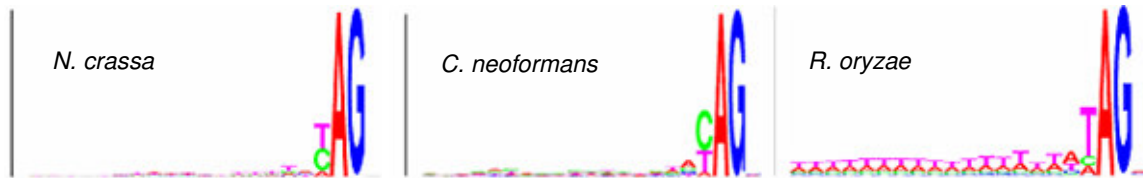


Figure S4. The logos of acceptor site models for *N. crassa* (Ascomycota), *C. neoformans* (Basidiomycota) and *R. oryzae* (Zygomycota). The models were derived from the sets of acceptors identified by EST genome alignments.

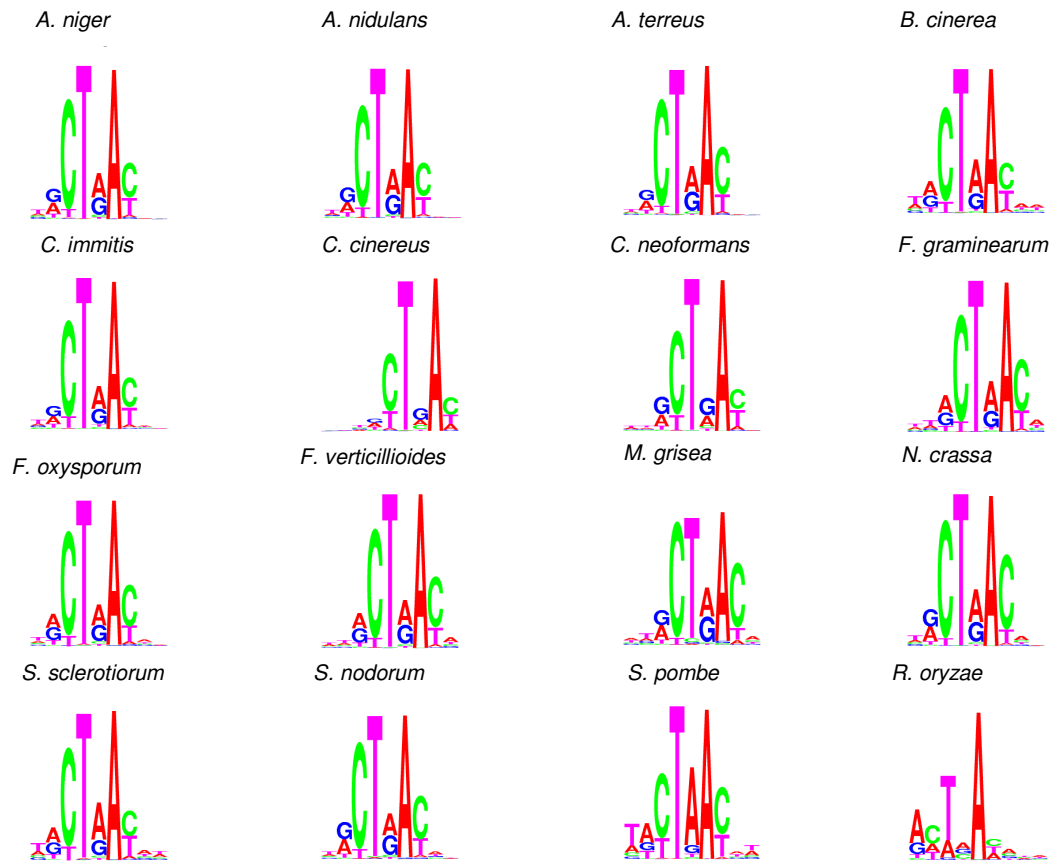


Figure S5. The logos of the zero order models of the branch point site for 16 fungal genomes as determined by the Gibbs sampling alignment of introns predicted at the algorithm convergence.

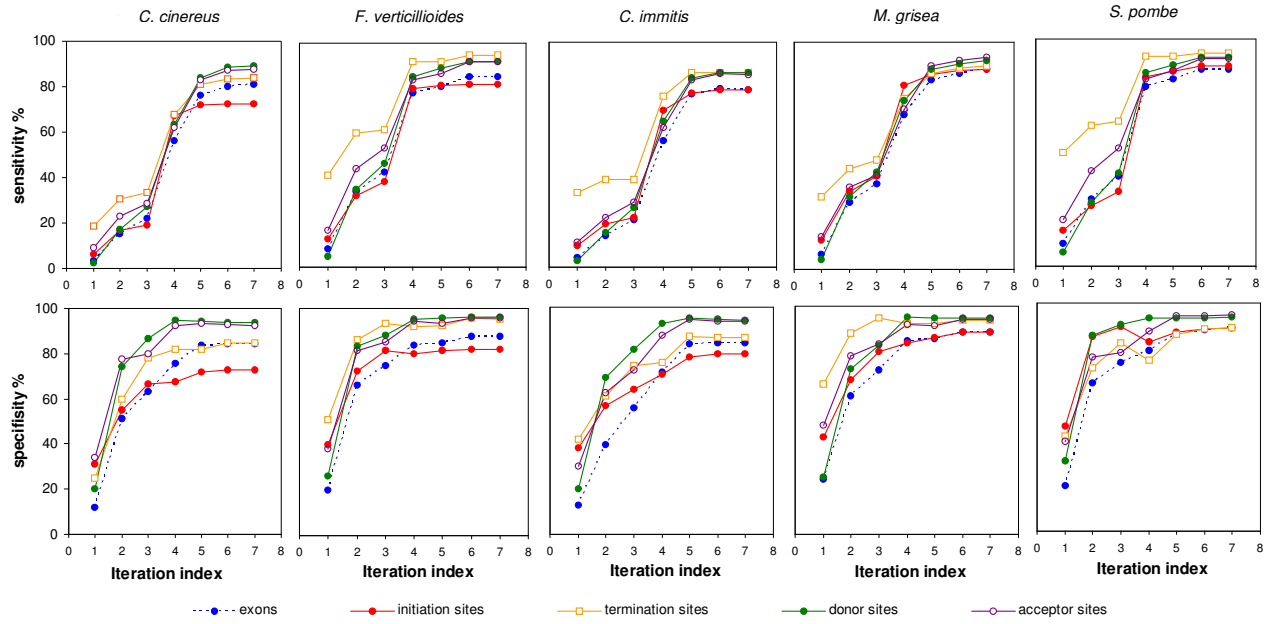


Figure S6. Changes in characteristics of prediction accuracy (Sn, Sp) of the exon-intron structures of genes from five fungal species are caused in subsequent iterations by the changes of the architecture and parameters of the HMM model underlying the gene prediction algorithm.

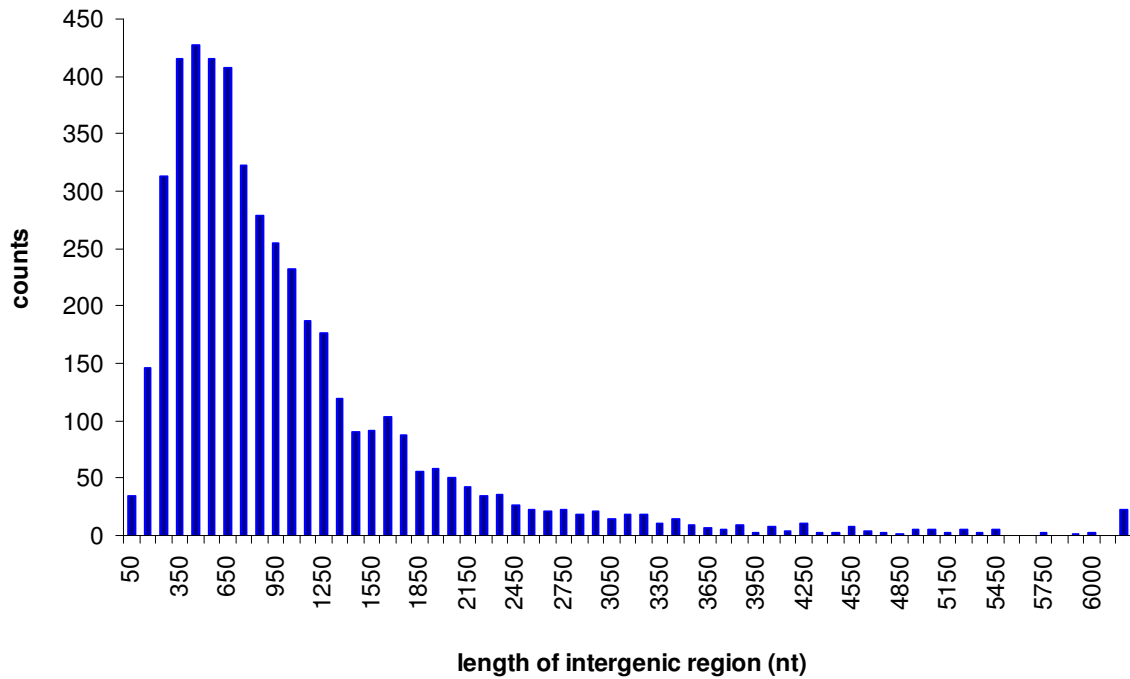


Figure S7. The length distribution of intergenic regions in the *S. pombe* genome (as annotated).

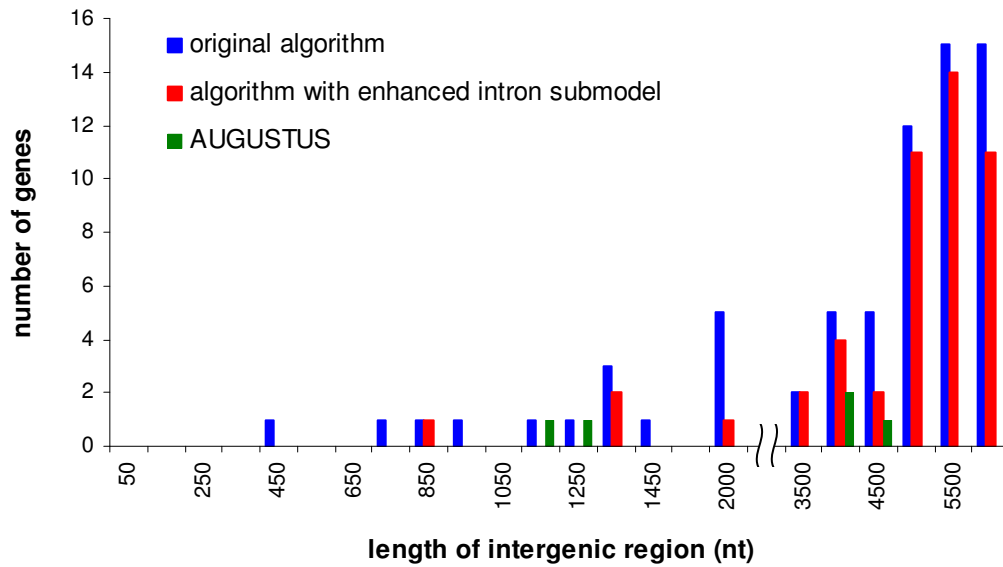


Figure S8. Number of genes erroneously predicted in 1276 intergenic regions of the *S. pombe* artificial chromosomes.

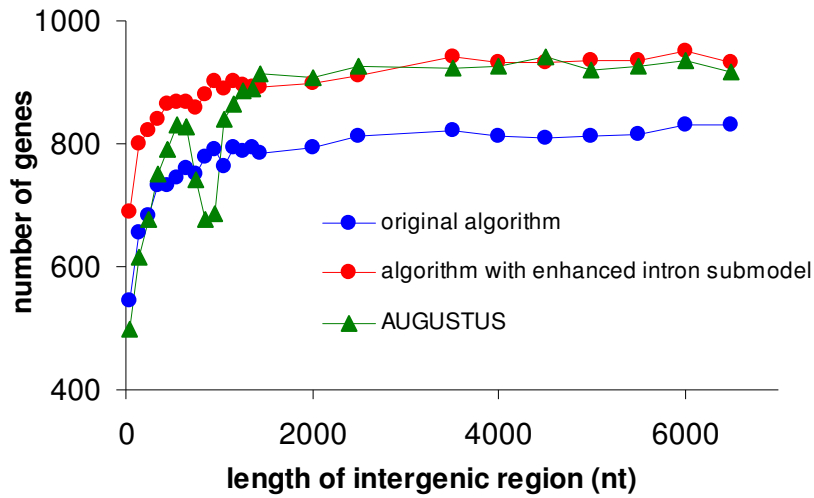


Figure S9. The number of exactly predicted genes in *S. pombe* artificial chromosomes (with 1277 genes) as a function of the length of intergenic region.

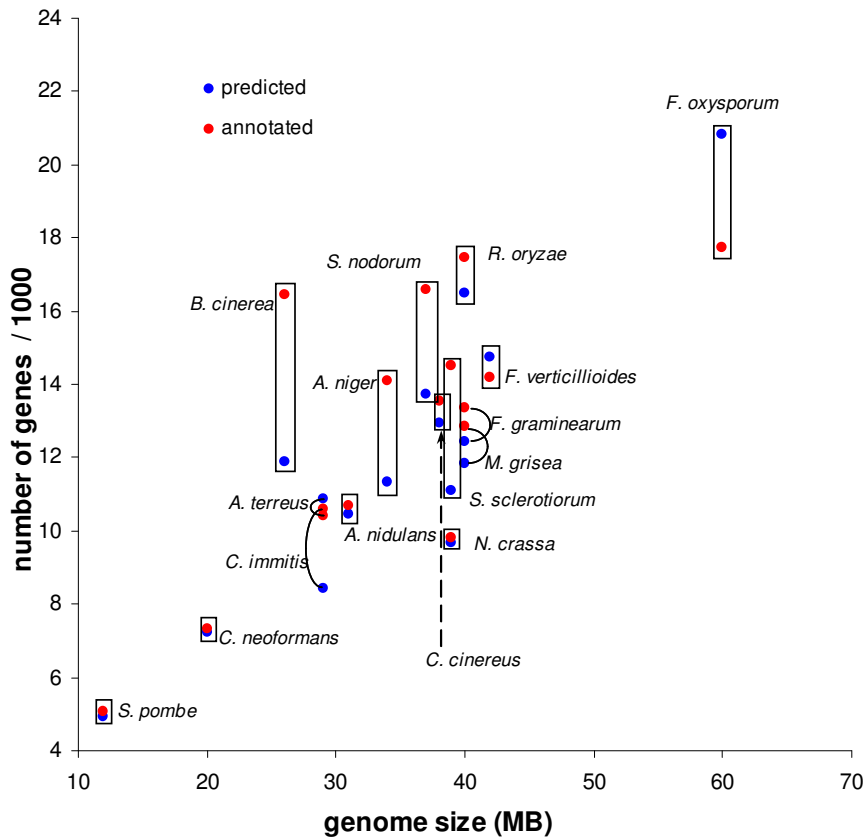


Figure S10. Numbers of predicted and annotated genes in the 16 fungal genomes. For some genomes (*B. cinerea*, *F. oxysporum*) the numbers of predicted and annotated genes are significantly different while for the others (*C. neoformans*, *N. crassa*, *S. pombe*, etc.) they are close to each other. The number of genes hosted per 1 MB of sequence varies from 250 to 430 genes per 1MB (*N. crassa* and *R. oryzae*, respectively).

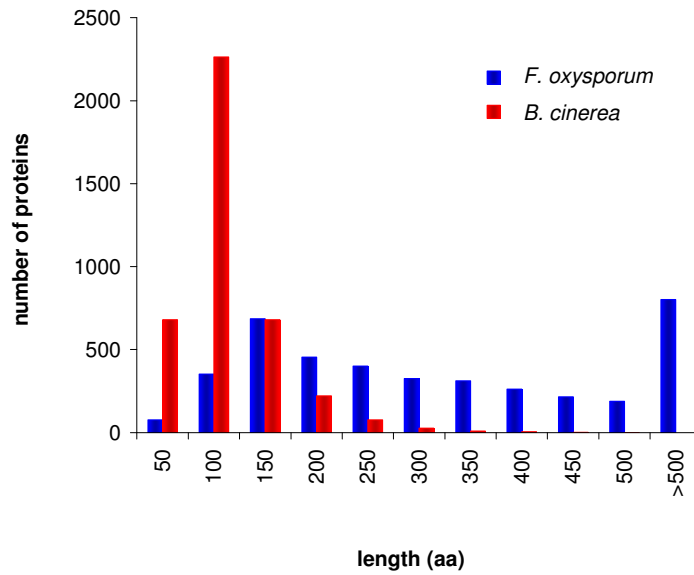


Figure S11. Length distributions of i/ the 4,558 genes annotated in *B. cinerea* genome that did not overlap with the predicted genes (red bars); ii/ the 3,108 genes predicted in *F. oxysporum* genome that did not overlap with the annotated genes (blue bars).

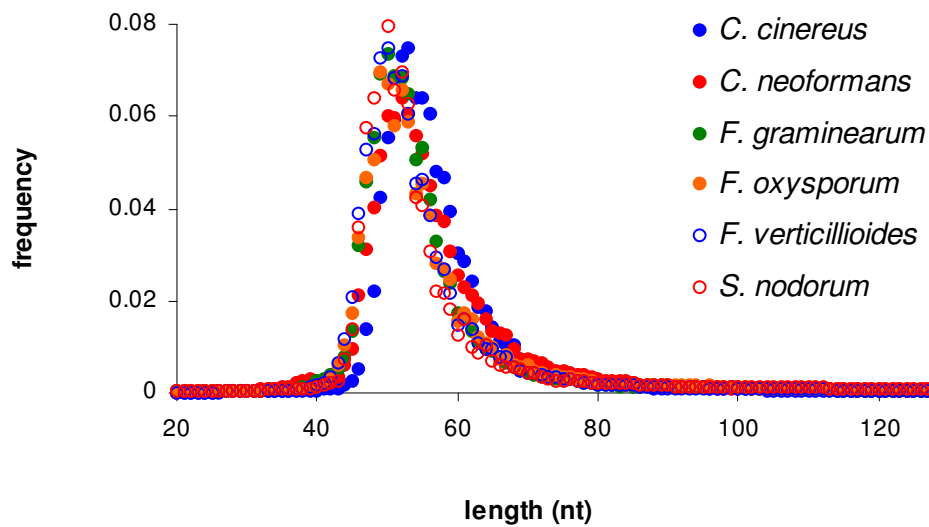
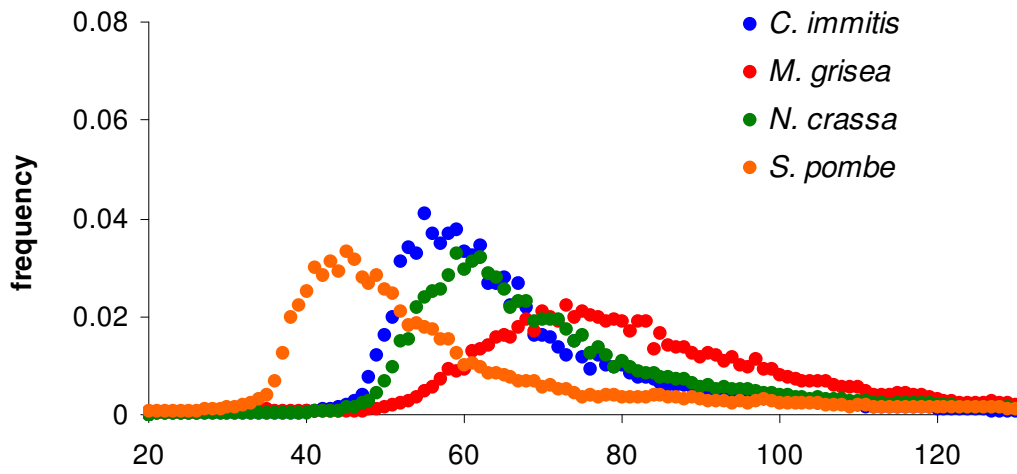
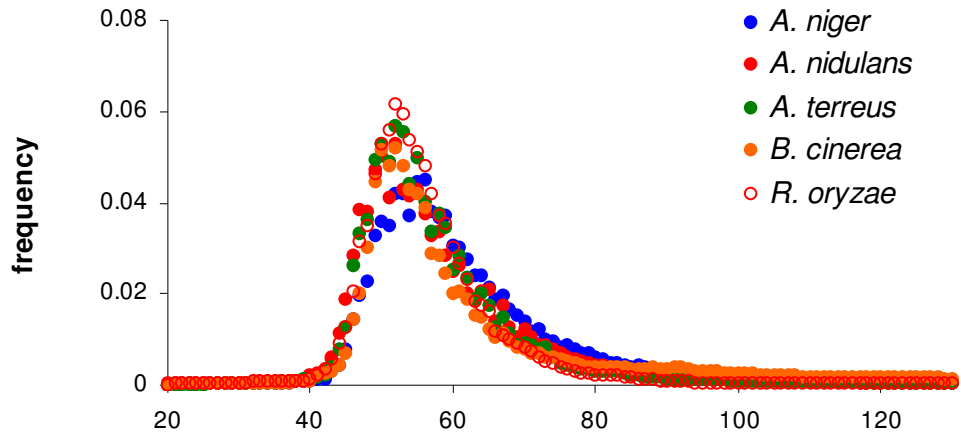


Figure S12. The intron length distributions as determined for the 16 fungal genomes.

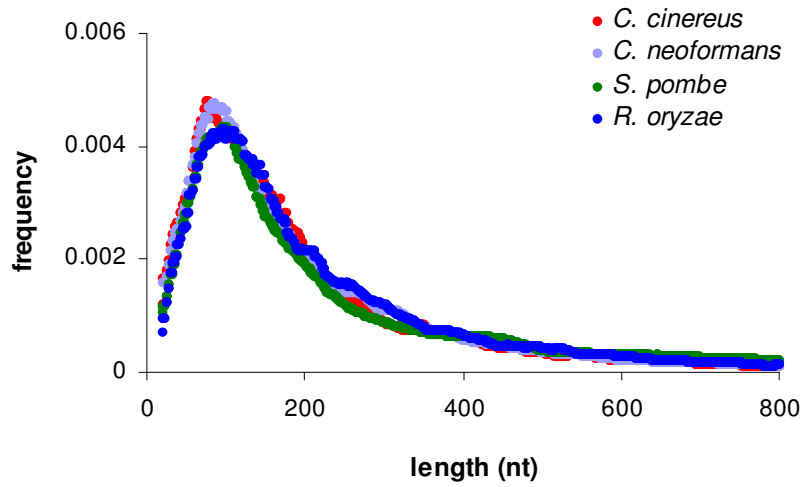
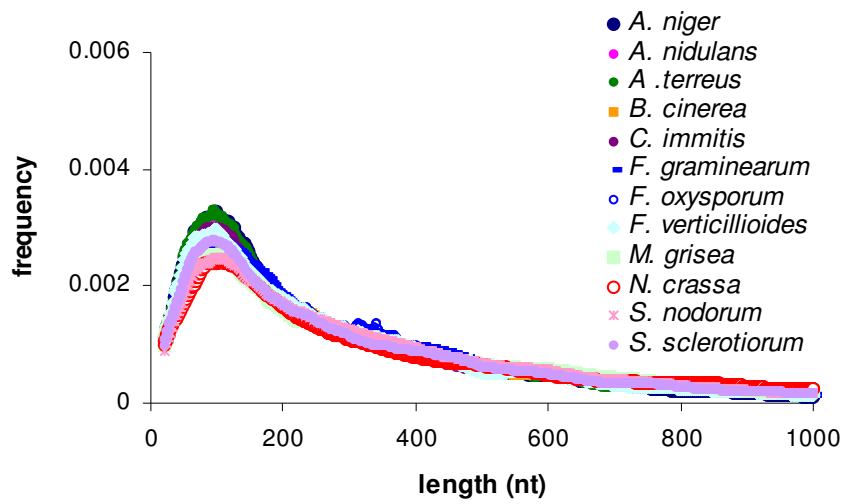


Figure S13. The exon length distributions as determined for the 16 fungal genomes.

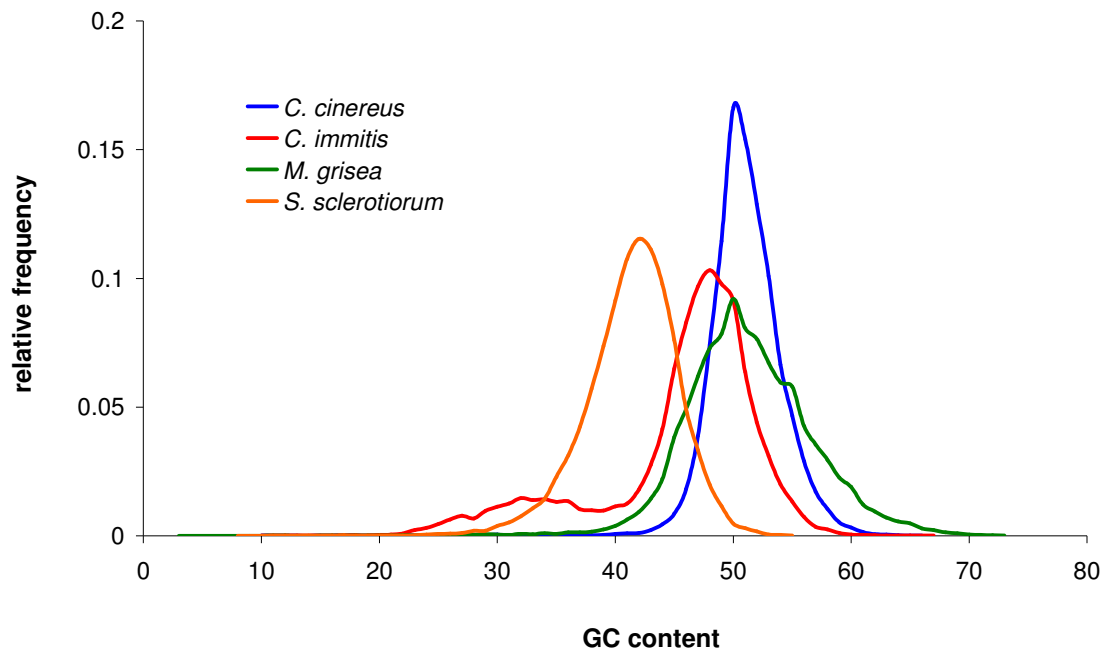


Figure S14. Distribution of genomic DNA GC content for the four fungal species (as determined in 1 kb long non-overlapping fragments).