

Evidence of mRNA-Mediated Intron Loss in the Human-Pathogenic Fungus *Cryptococcus neoformans*†

Jason E. Stajich and Fred S. Dietrich*

Department of Molecular Genetics and Microbiology, Center for Applied Genomics and Technology, and Institute for Genome Sciences and Policy, Duke University, Box 3568, Durham, North Carolina 27710

Received 2 February 2006/Accepted 2 March 2006

Introns are a defining feature of eukaryotic genomes, though the mechanism of intron gain or loss is not well understood. Reverse transcription of mRNA followed by homologous recombination with the genome has been posited as a mechanism of intron loss, though little direct evidence of recent loss events has been described to support this model. We find supporting evidence for an mRNA-mediated mechanism of loss through comparative genome analyses that revealed a recent loss of 10 adjacent introns in a 22-exon gene in the human-pathogenic fungus *Cryptococcus neoformans*. We surveyed the gene structures of the entire genomes of *Cryptococcus gattii*, which diverged from the *C. neoformans* lineage 37 million years ago (Mya), and *C. neoformans* var. *grubii* and var. *neoformans*, which diverged 18 Mya. Our comparison revealed greater than 99.9% intron conservation, with evidence from 20 genes showing evidence of intron loss, but no convincing evidence of intron gain. Our findings confirm that *Cryptococcus* introns have been quite stable over recent evolutionary time, with occasional mRNA-mediated intron loss events.

The mechanism and implications of change in the exon-intron structure of genes are a poorly understood aspect of genome evolution. Recent work on intron gain, loss, and conservation in metazoan, plant, and fungal lineages has shown that gene structure is a slowly evolving phylogenetic character, but that work has been unable to identify specific recent events due to the deep evolutionary distances compared and is unable to address the mechanism of intron loss (23, 29–31). Loss and gain of group I rRNA gene introns has been explored (2, 13) but appears to be distinctly different from the process of intron loss and gain in protein-coding genes (21) because of the highly repetitive nature of rRNA genes.

While loss of a single intron can occur by deletion of genomic DNA (19), simultaneous precise loss of multiple adjacent introns is suggestive of a mechanism involving a spliced RNA intermediate. Homologous recombination of a reverse-transcribed transcript has been previously proposed as a model of intron loss (3, 11, 15, 22–24), but whether this occurs through direct recombination of the transcript with the locus or through integration of a transcript into the genome to form a pseudogene (7) followed by gene conversion (10) has not been established.

In addition to intron loss, several molecular mechanisms for intron gain have been proposed. These mechanisms include transposition of introns (4, 21), transposon insertions that mutate to become introns (4), and tandem duplication of sequences where one copy mutates to become an intron (5, 21). These proposed mechanisms share one trait in common, in that they require a newly acquired intron to be a copy of another sequence in the genome. Thus, new introns created by

one of the mechanisms should be identifiable through the presence of homologous sequences in the genome until sufficient sequence divergence has occurred.

We sought to identify examples of recent changes in gene structure to uncover the signatures of the molecular process of intron gain or loss. Comparing entire genome sequences of closely related species provided an opportunity to search for genes whose exon-intron structure had changed. For closely related species a signature of the molecular mechanism of intron loss or gain should be identifiable, since insufficient time will have passed to obscure the signals through accumulation of random mutations.

Cryptococcus neoformans is a generally haploid, single-celled basidiomycete fungus with an intron-rich genome (an average of five introns per protein-coding gene) (20). The fungus is an opportunistic human pathogen with three varieties (17) causing cryptococcosis: *Cryptococcus gattii* has been shown to infect immunocompetent hosts (14), and *C. neoformans* var. *neoformans* and var. *grubii* are more commonly found in immunocompromised hosts (26). The *C. gattii* and *C. neoformans* ancestor diverged approximately 37 million years ago (Mya), while *C. neoformans* var. *grubii* and var. *neoformans* diverged 18.5 Mya (36). The genomes of two *C. gattii* strains, one *C. neoformans* var. *neoformans* strain, and one *C. neoformans* var. *grubii* strain were recently sequenced, providing a data set to identify recent intron loss and gain (20; <http://cneo.genetics.duke.edu/>, <http://www.broad.mit.edu/>, and <http://www.bcgsc.ca>).

MATERIALS AND METHODS

Genome annotation and analysis. Genome annotation of the *C. neoformans* var. *grubii* and *C. gattii* genomes was performed with an automated pipeline using a combination of ab initio and evidence-based gene prediction methods incorporating cDNA, protein, and fungal genome sequence data into composite gene predictions, utilizing the GLEAN tool (A. J. Mackey, Q. Liu, F. C. N. Pereira, and D. S. Roos, unpublished data). All the generated gene predictions can be downloaded from our public web server for fungal genome annotation, including a genome browser for visualizing the annotation (33; <http://fungal.genome.duke.edu>).

* Corresponding author. Mailing address: Department of Molecular Genetics and Microbiology, Duke University, Box 3568, Durham, NC 27710. Phone: (919) 684-2720. Fax: (919) 681-1035. E-mail: dietr003@mc.duke.edu.

† Supplemental material for this article may be found at <http://ec.asm.org/>.

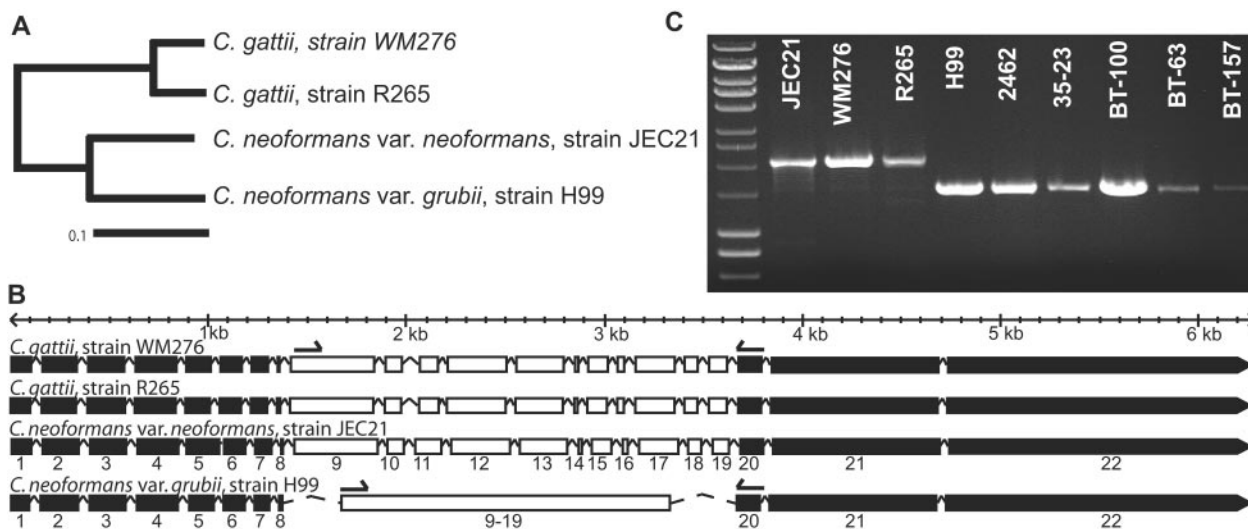


FIG. 1. A. Phylogenetic tree based on average synonymous site differences for 5,298 orthologous loci across the four *Cryptococcus* genomes, *C. neoformans* var. *neoformans* (strain JEC21, serotype D), *C. neoformans* var. *grubii* (strain H99, serotype B), and two strains of *Cryptococcus gattii* (strains WM276 and R265, serotypes B and B/C). B. Schematic of the locus CNI01550 exon-intron structure of the orthologous locus in *C. neoformans* var. *neoformans*, *C. gattii*, and *C. neoformans* var. *grubii*. As indicated by open boxes, exons 9 to 19 of *C. neoformans* var. *grubii* correspond to introns 9 through 19 in the other three strains. Exons and introns are drawn to scale, except where indicated by a dashed line. PCR primers are shown as arrows above the exon boxes. C. Agarose gel showing PCR products amplified across exons 9 to 19. Lanes are in the following order: DNA ladder I (Gene Choice, Frederick, Md.); *C. neoformans* var. *neoformans* (JEC21; size, ~2.0 kb); *C. gattii* strains (WM276 and R265; size, ~2.0 kb); *C. neoformans* var. *grubii* strains (H99, 2462, 35-23, BT-100, BT-63, and BT-157; size ~1.5 kb).

A more thorough description of the genome annotation protocol is in preparation.

The *C. neoformans* var. *grubii* strain H99 genome sequence was obtained from the Broad Institute and Center for Genome Technology at Duke University (Duke assembly dated October 2004; available from the fungal genome website). The *C. gattii* strain R265 sequence (assembly dated January 2005) was obtained from the Broad Institute website. The *C. gattii* strain WM276 was obtained from the BC Genome sequencing center (assembly dated March 2004). The annotated *C. neoformans* var. *neoformans* strain JEC21 genome was obtained from GenBank (accession numbers AE017341.1 to AE017353.1 and AE017356.1).

The algorithm for identifying and evaluating intron positions was implemented in the Perl language utilizing the BioPerl tool kit (version 1.5.1) (32). Orthologous genes for species pairs among the four genomes were identified via best reciprocal hits of protein sequences using FASTP (version 3.4t25). We identified four-way orthologous genes by creating single-linkage clusters from the pairwise orthologs and identifying clusters with exactly one member from each genome. These four-way orthologous proteins were aligned with MUSCLE (version 3.6) (9, 25), and introns were inserted back into the alignment. Alignment columns containing an intron were classified by the number of species that shared the intron, allowing the inspection of patterns of intron conservation or acquisition. Intron columns in the alignment were only considered if there were no gaps in three adjacent columns on either side of the intron in the protein alignment and if the six-column region had an average percent identity of 90% or higher at the protein sequence level. The 90% cutoff was based on a manual investigation of the alignments to minimize false positives. There were 33,494 total intron positions across the 5,298 orthologous alignments, and 18,731 positions from 4,690 loci met the filtered requirement of no gaps and 90% or better average sequence identity. Functional annotation of the genes was obtained through the EBI InterProScan SOAP service (27).

The average identity of the orthologous introns was computed by identifying orthologous introns which fall into the same column of the protein alignment and aligning the intron sequence with ClustalW (version 1.83) (35) and computing the overall identity with the alstat application (SQUID package [http://selab.wustl.edu/cgi-bin/selab.pl?mode = software#squid]). The R package (version 2.2.0) was used for summary statistics of the sequence alignments (28).

Genomic PCR. Genomic PCR was performed with *Ex Taq* polymerase (Takara, Madison, Wis.) and the primers forward (5'-CAA GGT CAG AGA GTT GAC TG-3') and reverse (5'-ATA TGC GCT TGG AGC AGG AT-3') located in exons 9 and 20 of CNI01550 flanking the region of intron loss to produce a 2,116- or 2,124-bp (intron-containing) or 1,574-bp (intron missing)

product. The PCR routine consisted of 95°C for 5 min followed by 35 rounds of 30 s at 95°C, 30 s at 55°C, and 2 min at 72°C and finished with 10 min at 72°C. J. A. Fraser, J. Heitman, A. P. Litvintseva, and T. G. Mitchell generously provided strains.

RESULTS

Evidence for intron loss. Using predicted protein sequences of the four genomes, we identified 5,298 orthologous genes among the four genomes of *C. neoformans* and *C. gattii* strains. Based on four-way coding sequence alignments of the orthologous loci, we estimated an average synonymous substitution rate (K_s) between *C. gattii* and *C. neoformans* of 0.35 substitutions per site and, between *C. neoformans* var. *neoformans* and *C. neoformans* var. *grubii*, an average K_s of 0.22 (Fig. 1A), which is roughly equivalent to the divergence of mouse and rat (12).

We aligned the predicted protein sequences of the orthologous genes and mapped the introns into the alignment to study how introns were conserved among these species. Intron positions were classified by observing the alignment column where an intron occurred and identifying how many species shared an intron in the same position. The comparison revealed that between *C. neoformans* and *C. grubii*, 99% of intron positions are conserved and the sequences of these introns have a mean 82% sequence identity, while the mean coding sequence identity is 90%. Within these conserved genes only 80 intron positions in 46 loci differed by presence in at least one of the sequences out of 18,731 filtered intron positions and 4,690 loci considered.

In particular, two loci had loss of more than three consecutive introns in one lineage. CNI01550, a putative RNA helicase, and CNN02320/FKS1, a 1,3-beta-glucan synthase, show

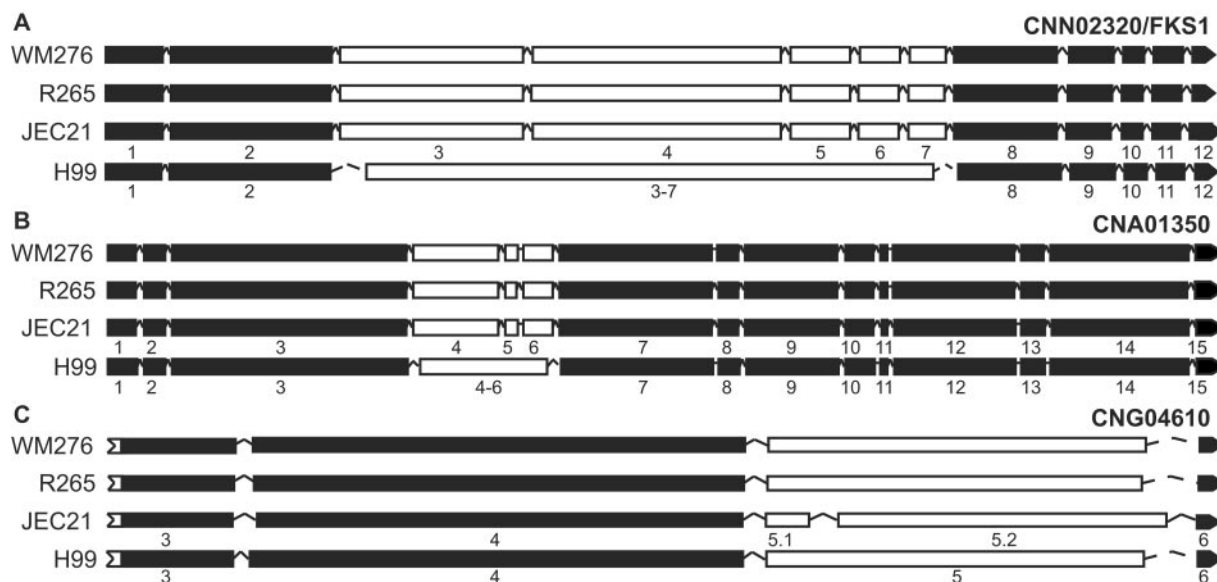


FIG. 2. Annotation for genes in the *Cryptococcus* genomes, showing changes in intron-exon structure. These changes have been confirmed by evaluating genomic sequence alignments. Open boxes indicate portions of the gene where exon-intron structure differs. A. CNN02320/FKS1 locus, a 1,3-beta-glucan synthase, showing changes in four introns between exons 3 and 7 in *C. neoformans* var. *grubii*. B. CNA01350 locus, a *PRP8* homolog and putative splicing factor, showing changes in two introns after exons 4 and 5 in *C. neoformans* var. *grubii* strain H99. C. CNG04610, a putative ubiquitin-protein ligase, showing changes in exon 5 of *C. neoformans* var. *neoformans* strain JEC21. The first two introns are not shown.

evidence of deletion of 10 and 4 introns, respectively, in *C. neoformans* var. *grubii*. We examined the CNI01550 locus and the surrounding genes and found that these 10 consecutive introns were lost precisely at the intron splice sites (Fig. 1B; see also Fig. S1 in the supplemental material). CNI01550 genes are orthologous based on synteny and sequence similarity, and no homologous copy of the locus was detected in a similarity search of any of the *Cryptococcus* genomes.

To confirm the missing introns in the genomic sequences among the population of *C. neoformans* var. *grubii* individuals, we used PCR to amplify this locus (Fig. 1C) in six strains of *C. neoformans* var. *grubii* (H99, 2462, 35-23, BT-100, BT-63, and BT-157) that have been shown to be members of three distinct population clades (18). In addition, we amplified the orthologous region from the *C. neoformans* var. *neoformans* strain JEC21 and two *C. gattii* strains, R265 and WM276. The intronless form was only seen in the *C. neoformans* var. *grubii* strains and was not present in any other *Cryptococcus* strains sampled. Additional examples of intron loss in a single lineage include the loss of four introns as previously reported (18, 34) in CNN02320/FKS1 (Fig. 2A) and loss of two introns from a *PRP8* homolog and putative splicing factor, CNA01350 (Fig. 2B).

Putative intron gains. We did not observe strong evidence of intron gain in any of the loci surveyed. We inferred gain by two criteria, namely that the intron position was unique in one of the species and that this intron position was not seen in any other fungi. However, we identified two loci that are possible cases of intron gain. One case is the fifth intron in CNG04610 (Fig. 2C), a putative ubiquitin-protein ligase, where the intron is found only in *C. neoformans* var. *neoformans*. An alignment of this locus to orthologs in other available fungal sequences reveals this intron position is not shared with any other ho-

molog. Similarly, in the gene CNK02730, a putative galactose transporter, introns 5 through 10 in JEC21 are missing in *C. gattii* while intron 6 is only present in *C. neoformans* var. *grubii* (see the alignment in Fig. S1 in the supplemental material). This circumstance could be the result of an intron insertion in the *C. neoformans* var. *grubii* lineage or parallel loss in the *C. neoformans* var. *neoformans* and *C. gattii* lineages.

As models for intron gain generally assume duplication of an existing sequence, we examined alignments of introns between the four sequenced *Cryptococcus* strains. The mean percent identity was 82% (standard deviation, 6.2%) for orthologous introns shared in all four species, while the mean percent identity was 90% (standard deviation, 4.7%) for the orthologous coding regions. This high similarity indicates that introns can be recognized by sequence similarity alone, and it should be possible to identify the source of newly generated introns between these closely related varieties. A search of all introns for significantly similar sequences within the genome did not identify any clear examples of intron duplication. In addition, the introns identified as putatively gained had no significant sequence similarity to any regions in the host genome or in the genomes of any of the *Cryptococcus* species.

DISCUSSION

The change of exon-intron structure occurred after the divergence of *C. neoformans* var. *grubii* and *C. neoformans* var. *neoformans*, as confirmed by comparison of all three *Cryptococcus* genomes (Fig. 1A). The recent divergence of these strains suggests that the introns were lost from CNI01550 over a short period of time or, more likely, all at once. Considering that intron loss is a rare event—less than 1% of the loci (46 out of 4,690) surveyed had evidence of intron-exon structure

change in the four-way comparison—it is highly unlikely that independent losses of 10 adjacent introns through a DNA deletion mechanism would occur. The deletions are precise excisions of the intron sequences as seen in the alignment of the genomic DNA (see Fig. S1 in the supplemental material). These introns are of the typical length found in *C. neoformans* (51 to 64 bp) and appear to have canonical GT-AG splice sites; furthermore, the genes observed to have intron loss events do not fall into a single type of functional pathway or gene ontology category. The sequence identity of CNI01550 between the strains is highly similar (96% DNA sequence identity between *C. neoformans* var. *neoformans* and *C. neoformans* var. *grubii*), as is typical of orthologs in these strains, indicating that neither copy is a result of a recent horizontal transfer event from a more distant species. There is no evidence of a paralogous copy in any of these genomes, ruling out intron loss through gene conversion with a second copy. The deletion of consecutive introns suggests that the loss event involved a spliced mRNA intermediate, and this is consistent with homologous recombination of a reverse-transcribed transcript into the locus. While reverse transcriptase activity has not been directly shown, cDNA sequencing from *C. neoformans* (16, 20) confirms that transcripts from putative reverse transcriptase-containing loci are expressed (e.g., accession no. CF679825). For CNI01550, the eight 5'-most introns and three 3'-most introns and are still present in the *C. neoformans* var. *grubii* lineage, indicating an event involving the central portion of the spliced mRNA. This is consistent with the observed intron loss among eukaryotic fungi that lacked a 3' loss bias (23). Previous work has demonstrated that reverse-transcribed mRNA can serve as a template for gene conversion of an artificial construct in *Saccharomyces cerevisiae* (6). This example from *C. neoformans* appears to be an unambiguous example of intron loss in a fungus driven by a spliced mRNA intermediate.

While CNI01550 and CNN02320/FKS1 showed unambiguous evidence of intron loss, no definitive examples of intron gain were identified. The apparent intron gain in CNG04610 could also be explained by two independent loss events in the *C. gattii* and *C. neoformans* var. *grubii* lineages. That the additional intron seen in *C. neoformans* var. *neoformans* has no significant sequence similarity elsewhere in the genome of JEC21 or the other *Cryptococcus* genomes suggests that this intron could not have been gained by a mechanism copying the intron, such as intron transposition.

The intron composition of the *Cryptococcus* genome appears to be quite stable. Assuming all differences in introns are due to loss, a constant rate of loss of 0.4% of introns over 18 million years implies an intron half-life of 3 billion years. To arrive at the current number of approximately 33,000 introns, the common ancestor of the fungi (estimated to have emerged 800 Mya [1, 8]) would have had approximately 40,000 introns in 5,300 genes. In contrast, the intron-poor hemiascomycete yeasts, including *Saccharomyces cerevisiae*, must have experienced a significantly higher intron loss rate in their evolutionary history.

All of the loss events observed in these *Cryptococcus* species are precise splice site deletions, suggesting mRNA-mediated intron loss not only occurs but also is the primary means of intron loss. These findings suggest that reverse transcriptase may have an important role in maintenance of transcript fidelity. When an intron is mutated and inefficiently spliced, reverse

transcriptase can act to purge the defective intron from the genome. This may explain the widespread existence of reverse transcriptase in eukaryotic genomes.

ACKNOWLEDGMENTS

The draft genome sequence of *C. neoformans* var. *grubii* strain H99 was produced by the sequencing project at Duke Center for Genome Technology (<http://cneo.genetics.duke.edu>) and Broad Institute of Harvard and MIT (<http://www.broad.mit.edu>). The draft genome sequence of *C. gattii* strain WM276 was produced at Canada's Michael Smith Genome Sciences Centre with funding from Genome Canada and is available at <http://www.bcgsc.ca>. The draft genome sequence of *C. gattii* strain R265 was produced at the *C. neoformans* Serotype B Sequencing Project at the Broad Institute of Harvard and MIT. Thanks to G. R. Fink and J. Heitman for helpful discussions.

J.E.S. was supported by an NSF graduate research fellowship. F.S.D. was supported by NIH grant NS042263-03.

REFERENCES

- Berbee, M. L., and J. W. Taylor. 2000. Systematics and evolution, p. 229–246. In D. J. McLaughlin and E. McLaughlin (ed.), *The mycota*, vol. VIII. Springer-Verlag, New York, N.Y.
- Bhattacharya, D., V. Reeb, D. M. Simon, and F. Lutzoni. 2005. Phylogenetic analyses suggest reverse splicing spread of group I introns in fungal ribosomal DNA. *BMC Evol. Biol.* 5:68.
- Boeke, J. D., D. J. Garfinkel, C. A. Styles, and G. R. Fink. 1985. Ty elements transpose through an RNA intermediate. *Cell* 40:491–500.
- Cavalier-Smith, T. 1985. Selfish DNA and the origin of introns. *Nature* 315:283–284.
- Coghlan, A., and K. H. Wolfe. 2004. Origins of recently gained introns in *Caenorhabditis*. *Proc. Natl. Acad. Sci. USA* 101:11362–11367.
- Derr, L. K., and J. N. Strathern. 1993. A role for reverse transcripts in gene conversion. *Nature* 361:170–173.
- Derr, L. K., J. N. Strathern, and D. J. Garfinkel. 1991. RNA-mediated recombination in *S. cerevisiae*. *Cell* 67:355–364.
- Douzery, E. J., E. A. Snell, E. Baptiste, F. Delsuc, and H. Philippe. 2004. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc. Natl. Acad. Sci. USA* 101:15386–15391.
- Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Ernst, J. F., J. W. Stewart, and F. Sherman. 1981. The *cyc1-11* mutation in yeast reverts by recombination with a nonallelic gene: composite genes determining the iso-cytochromes c. *Proc. Natl. Acad. Sci. USA* 78:6334–6338.
- Fink, G. R. 1987. Pseudogenes in yeast? *Cell* 49:5–6.
- Gibbs, R. A., G. M. Weinstock, M. L. Metzker, D. M. Muzny, E. J. Sodergren, S. Scherer, G. Scott, D. Steffen, K. C. Worley, P. E. Burch, G. Okwuonu, S. Hines, L. Lewis, C. DeRamo, O. Delgado, S. Dugan-Rocha, G. Miner, M. Morgan, A. Hawes, R. Gill, C. Elera, R. A. Holt, M. D. Adams, P. G. Amanatides, H. Baden-Tillson, M. Barnstead, S. Chin, C. A. Evans, S. Ferriera, C. Fosler, A. Glodek, Z. Gu, D. Jennings, C. L. Kraft, T. Nguyen, C. M. Pfannkoch, C. Sitter, G. G. Sutton, J. C. Venter, T. Woodage, D. Smith, H. M. Lee, E. Gustafson, P. Cahill, A. Kana, L. Doucette-Stamm, K. Weinstock, K. Fichtel, R. B. Weiss, D. M. Dunn, E. D. Green, R. W. Blakesley, G. G. Bouffard, P. J. De Jong, K. Osoegawa, B. Zhu, M. Marra, J. Schein, I. Bosdet, C. Fjell, S. Jones, M. Krzywinski, C. Mathewson, A. Siddiqui, N. Wye, J. McPherson, S. Zhao, C. M. Fraser, J. Shetty, S. Shatsman, K. Geer, Y. Chen, S. Abramzon, W. C. Nierman, P. H. Havlak, R. Chen, K. J. Durbin, A. Egan, Y. Ren, X. Z. Song, B. Li, Y. Liu, X. Qin, S. Cawley, K. C. Worley, A. J. Cooney, L. M. D'Souza, K. Martin, J. Q. Wu, M. L. Gonzalez-Garay, A. R. Jackson, K. J. Kalafus, M. P. McLeod, A. Milosavljevic, D. Virk, A. Volkov, D. A. Wheeler, Z. Zhang, J. A. Bailey, E. E. Eichler, et al. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428:493–521.
- Haugen, P., D. M. Simon, and D. Bhattacharya. 2005. The natural history of group I introns. *Trends Genet.* 21:111–119.
- Hoang, L. M., J. A. Maguire, P. Doyle, M. Fyfe, and D. L. Roscoe. 2004. *Cryptococcus neoformans* infections at Vancouver Hospital and Health Sciences Centre (1997–2002): epidemiology, microbiology and histopathology. *J. Med. Microbiol.* 53:935–940.
- Jeffares, D. C., T. Mourier, and D. Penny. 2006. The biology of intron gain and loss. *Trends Genet.* 22:16–22.
- Kupfer, D. M., S. D. Drabenstot, K. L. Buchanan, H. Lai, H. Zhu, D. W. Dyer, B. A. Roe, and J. W. Murphy. 2004. Introns and splicing elements of five diverse fungi. *Eukaryot. Cell* 3:1088–1100.
- Kwon-Chung, K. J., and J. E. Bennett. 1984. Epidemiologic differences between the two varieties of *Cryptococcus neoformans*. *Am. J. Epidemiol.* 120:123–130.

18. Litvintseva, A. P., R. Thakur, R. Vilgalys, and T. G. Mitchell. 1 December 2005, posting date. Multilocus sequence typing reveals three genetic subpopulations of *Cryptococcus neoformans* var. *grubii* (serotype A), including a unique population in Botswana. *Genetics* [Online.] doi:10.1534/genetics.105.046672.
19. Llopert, A., J. M. Comeron, F. G. Brunet, D. Lachaise, and M. Long. 2002. Intron presence-absence polymorphism in *Drosophila* driven by positive Darwinian selection. *Proc. Natl. Acad. Sci. USA* **99**:8121–8126.
20. Loftus, B. J., E. Fung, P. Roncaglia, D. Rowley, P. Amedeo, D. Bruno, J. Vamathevan, M. Miranda, I. J. Anderson, J. A. Fraser, J. E. Allen, I. E. Bosdet, M. R. Brent, R. Chiu, T. L. Doering, M. J. Donlin, C. A. D'Souza, D. S. Fox, V. Grinberg, J. Fu, M. Fukushima, B. J. Haas, J. C. Huang, G. Janbon, S. J. Jones, H. L. Koo, M. I. Krzywinski, J. K. Kwon-Chung, K. B. Lengeler, R. Maiti, M. A. Marra, R. E. Marra, C. A. Mathewson, T. G. Mitchell, M. Pertea, F. R. Riggs, S. L. Salzberg, J. E. Schein, A. Shvartsbeyn, H. Shin, M. Shumway, C. A. Specht, B. B. Suh, A. Tenney, T. R. Utterback, B. L. Wickes, J. R. Wortman, N. H. Wye, J. W. Kronstad, J. K. Lodge, J. Heitman, R. W. Davis, C. M. Fraser, and R. W. Hyman. 2005. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* **307**:1321–1324.
21. Logsdon, J. M., Jr., A. Stoltzfus, and W. F. Doolittle. 1998. Molecular evolution: recent cases of spliceosomal intron gain? *Curr. Biol.* **8**:R560–R563.
22. Mourier, T., and D. C. Jeffares. 2003. Eukaryotic intron loss. *Science* **300**:1393.
23. Nielsen, C. B., B. Friedman, B. Birren, C. B. Burge, and J. E. Galagan. 2004. Patterns of intron gain and loss in fungi. *PLoS Biol.* **2**:e422.
24. Niu, D. K., W. R. Hou, and S. W. Li. 2005. mRNA-mediated intron losses: evidence from extraordinarily large exons. *Mol. Biol. Evol.* **22**:1475–1481.
25. Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**:2444–2448.
26. Perfect, J. R., and A. Casadevall. 2002. Cryptococcosis. *Infect. Dis. Clin. North Am.* **16**:837–874.
27. Pillai, S., V. Silventoinen, K. Kallio, M. Senger, S. Sobhany, J. Tate, S. Velankar, A. Golovin, K. Henrick, P. Rice, P. Stoehr, and R. Lopez. 2005. SOAP-based services provided by the European Bioinformatics Institute. *Nucleic Acids Res.* **33**:W25–W28.
28. R Development Core Team. 2005. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
29. Rogozin, I. B., Y. I. Wolf, A. V. Sorokin, B. G. Mirkin, and E. V. Koonin. 2003. Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Curr. Biol.* **13**:1512–1517.
30. Roy, S. W., and W. Gilbert. 2005. Complex early genes. *Proc. Natl. Acad. Sci. USA* **102**:1986–1991.
31. Roy, S. W., and W. Gilbert. 2005. The pattern of intron loss. *Proc. Natl. Acad. Sci. USA* **102**:713–718.
32. Stajich, J. E., D. Block, K. Boulez, S. E. Brenner, S. A. Chervitz, C. Dagdigan, G. Fuellen, J. G. Gilbert, I. Korf, H. Lapp, H. Lehvaslaiho, C. Matsalla, C. J. Mungall, B. I. Osborne, M. R. Pocock, P. Schattner, M. Senger, L. D. Stein, E. Stupka, M. D. Wilkinson, and E. Birney. 2002. The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* **12**:1611–1618.
33. Stein, L. D., C. Mungall, S. Shu, M. Caudy, M. Mangone, A. Day, E. Nickerson, J. E. Stajich, T. W. Harris, A. Arva, and S. Lewis. 2002. The generic genome browser: a building block for a model organism system database. *Genome Res.* **12**:1599–1610.
34. Tanaka, R., Y. Imanishi, and K. Nishimura. 2003. Difference in FKS1 gene sequences between serotypes A and D of *Cryptococcus neoformans*. *J. Clin. Microbiol.* **41**:4457–4459.
35. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
36. Xu, J., R. Vilgalys, and T. G. Mitchell. 2000. Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*. *Mol. Ecol.* **9**:1471–1481.