Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 326 (2005) 233–241
BBRC
www.elsevier.com/locate/ybbrc

Genomic organization and expression of 23 new genes from *MATα* locus of *Cryptococcus neoformans* var. *gattii*

Ping Ren[a], Paola Roncaglia[a,1], Deborah J. Springer[b], Jinjiang Fan[a,2], Vishnu Chaturvedi[a,b,*]

[a] *Mycology Laboratory, Wadsworth Center, New York State Department of Health, Albany, NY, USA*
[b] *Department of Biomedical Sciences, SUNY School of Public Health, University at Albany, Albany, NY, USA*

Received 2 November 2004

Abstract

The pathogenic yeast *Cryptococcus neoformans* (*Cn*) causes cryptococcosis, a life-threatening disease of the brain. Molecular studies of *Cn* variety *gattii* have lagged behind other two varieties (var. *grubii* and var. *neoformans*) although they have distinct biology and disease patterns. We focused on gene discovery in *MATα* locus because it predominates in clinical strains. A var. *gattii* cosmid library was screened with DNA probes from other two varieties. Two positive clones were sequenced to identify ORFs based on similarities to known proteins, and to ESTs using bioinformatics, and manually by a curator. Approximately 76 kb sequenced DNA revealed 23 genes and ORFs. The existence of predicted genes was verified by RT-PCR analyses designed to amplify spliced sequences. The results confirmed that the transcripts were expressed both at 30 and 37 °C. The var. *gattii MATα* locus genes showed rearrangements in order and orientation vis-à-vis other two varieties. Mating-specific genes showed higher nonsynonymous mutation rates, and gene trees showed var. *gattii* strains in a distinct clade. The identification of the largest number, thus far, of var. *gattii* structural genes should set the stage for future molecular pathogenesis studies.

Keywords: *Cryptococcus neoformans*; Variety *gattii*; *MATα*; Locus; Genes; Sequence; RT-PCR; Transcription

Cryotococcus neoformans (*Cn*) is an encapsulated, pathogenic yeast. It causes life-threatening cryptococcosis in both normal and severely debilitated humans. Of the three *Cn* varieties, var. *gattii* has unique environmental and clinical features. It is mainly found in tropical climates, most commonly on *Eucalyptus* trees. It predominantly infects immunocompetent individuals, while var. *grubii* and var. *neoformans* cause disease in immunocompromised individuals, and are more widely distributed around the globe, found in soil and pigeon droppings [1–3]. The mechanisms underlying differences in host predilection and geographical distribution of the three *Cn* varieties remain unknown. Until recently, *Cn* var. *gattii* was considered rare on the North American continent, but an ongoing outbreak of *Cn* var. *gattii* infection on Vancouver Island in Canada has dispelled this view [4–6]. Prior to 1999, an average of two or three cases of cryptococcosis due to var. *grubii* and var. *neoformans* occurred annually in British Columbia. Since 1999, more than 66 human cases in otherwise healthy individuals, with at least four fatalities, have been attributed to infection with var. *gattii* [7]. There has also been an equally large increase in the rate of infection of animals, including dogs, cats, and porpoises [4]. Thus,

* Corresponding author. Fax: +1 518 486 7811.
E-mail address: vishnu@wadsworth.org (V. Chaturvedi).
[1] Present address: International School for Advanced Studies (SISSA-ISAS) Via Beirut 2-4, 34014 Trieste, Italy.
[2] Present address: Biotechnology Research Institute, National Research Council of Canada, 6100 Royalmount venue, Montreal, Que., Canada H4P 2R2.

0006-291X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2004.11.017

cryptococcosis due to *Cn* var. *gattii* presents an emerging threat to humans and animals in North America.

An intriguing and still incompletely explained observation about cryptococcosis is that more than 95% of clinical and environmental isolates of *Cn* are haploid, *MATα* [5,8,9]. The molecular mechanisms behind this dominance of α-mating type and its significance for human disease are being intensively studied in var. *neoformans* and var. *grubii* [10–13]. Two whole-genome projects on *Cn* var. *neoformans* are near completion, while one on var. *grubii* is well underway. The availability of this genomic information in the near future is likely to provide further impetus for investigations into mechanisms underlying differences in molecular, physiological, clinical, and ecological features, between these two varieties. By contrast, there is little information available on genes and pathogenesis of *Cn* var. *gattii*, even though a genomics project is underway at the University of British Columbia [5,14,15]. The present study was aimed at the characterization of var. *gattii MATα* locus since genes in this locus are known to be important in the biology and virulence of the other two *Cn* varieties. It is hoped that identification of var. *gattii* genes will promote future molecular pathogenesis studies.

Materials and methods

Strain. Cn var. *gattii* NIH 444 (ATCC 32609, *MATα*, serotype B) was chosen for our study, because it has been previously used in a number of mating studies, it has a well-defined pathogenic course in experimental cryptococcosis, and a genetic transformation system is available for molecular pathogenesis studies [15–17].

Cosmid clones. Cn var. *gattii* NIH 444 genomic DNA was extracted by grinding of early log phase cells with a sterilized mortar and pestle under liquid nitrogen, followed by phenol–chloroform extraction, and ligation into the pWEB vector (Epicenter, Madison, WI), which allows efficient cloning of approximate 40 kb genomic DNA fragments [18]. The quality of cosmid library was checked by screening for two housekeeping genes, *ACT1* and *URA5*, (GenBank Accession Nos. U10867 and AF536328). PCR primers for mating related genes *MFα1*, *STE11α*, *STE12α*, *STE20α*, and *MYO2* (Table 1) were designed based on the sequences from either *Cn* var. *grubii* or var. *neoformans* [10,11,19]. These gene fragments were PCR-amplified using NIH 444 genomic DNA and were used as probes in colony hybridization to isolate positive cosmid clones. To re-confirm positive cosmid clones, we used them as templates in PCR with the primers for the gene fragments described earlier.

Nucleotide sequencing. Cosmid clones were sequenced using EZ::TN Transposon Tools ⟨TET-1⟩ Insertion kits, (Epicenter, Madison, WI) to randomly insert primer binding sites into the target DNA in vitro. About 180 insertion clones from each selected cosmid were sequenced using the unlabeled forward and reverse transposon-specific primers provided in the kit. All sequencing was done on both strands of the DNA by ABI PRISM 377 sequencer with the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). The resulting sequence data were analyzed and assembled with the Sequencer 4.1.2 software (Gene Codes, Ann Arbor, MI). Additional sequencing with custom-designed primers was also carried out to fill in sequence gaps.

Bioinformatics. Assembled DNA sequences (about 76 kb) were analyzed by P. Roncaglia, then at The Institute for Genomic Research (TIGR, Rockville, MD). The nucleotide sequence of the *MATα* locus was searched against a database of cDNAs and ESTs from *Cn* var. *neoformans* (strains JEC21 and B3501) and var. *grubii* (strain H99), available on the Web sites of the institutions involved in the genome-sequencing projects (http://www-sequence.stanford.edu/group/C.neoformans/index.html; http://www.tigr.org/tdb/e2k1/cna1/; http://cneo.genetics.duke.edu/; and http://www.genome.ou.edu/cneo.html). The DNA-to-cDNA alignments were performed using the DDS/GAP2 set of programs from the AAT tool package [20]. The stringency of the DNA-to-cDNA alignments was set at 90% identity and 95% similarity. The nucleotide sequences were also searched against the public protein databases. The DNA-to-protein alignments were performed using the DPS/NAP set of programs from the AAT tool package [20]. Additionally, gene sequences were predicted de novo using four gene-finding programs: GlimmerM ([21], http://www.tigr.org/software/glimmerm/), Phat ([22]), TWINSCAN ([23]; http://genes.cs.wustl.edu/), and Unveil ([21]; http://www.tigr.org/software/Unveil/index.shtml). All of the evidence collected above (alignments to cDNAs, ESTs, and proteins, and automated gene predictions) was then combined computationally using the Combiner program ([24]; http://www.tigr.org/software/combiner/). Output from the alignment searches and the gene-prediction processes as above was displayed using the Annotation Station software (developed under contract to The Institute for Genomic Research by Neomorphic, an Affymetrix affiliate). Such output was then manually analyzed and validated. Once the exon–intron structure of the genes was defined, the annotation of putative functions for the protein-coding genes was carried out using the MANATEE tool (http://manatee.sourceforge.net/). The predicted protein sequences from the mating type locus of *Cn* var. *gattii* were searched against the Pfam ([25]; http://www.sanger.ac.uk/Software/Pfam/) and TIGRFAM (http://www.tigr.org/TIGRFAMs/) set of hidden Markov models using the HMMER program (http://hmmer.wustl.edu/), and against the public protein databases using BLASTP [26]. The putative functions of genes in *Cn* var. *gattii*, plus the suggested gene symbols and protein names, were established according to information available for their homologs in other *Cn* varieties. In the case of gene models that show conserved sequence homology to uncharacterized genes in other organisms (e.g., other species of *Cn*, or yeasts), the name 'conserved hypothetical protein' (CHP) has been adopted.

Phylogenetic analysis. Percentage of nucleotide identity among *MATα* locus genes was compared among three *Cn* varieties and between two strains of *Cn* var. *gattii* NIH444 and WM276 by ClustalW

Table 1
Comparison of nucleotides and deduced amino acids of selected *MATα* genes of *Cryptococcus neoformans* (*Cn*) var. *gattii* NIH 444 with various strains being used in whole genome sequencing

Gene	var. *gattii* WM276	var. *neoformans* JEC21	var. *grubii* H99
CAP1α	96/97[a]	83/85	84/88
CID1α	95/97	84/78	86/80
LPD1α	91/99	88/96	87/96
RPO41α	96/98	89/95	89/95
STE11α	94/92	85/79	85/80
MYO2α	95/94	88/84	88/85
STE20α	96/97	87/90	88/91
PRT1α	97/98	91/89	91/89
ZNF1α	83/91	41/80	42/78
STE12α	93/90	82/74	83/77
RPL39α	96/100	89/96	90/96
MFα1	97/100	94/94	93/86
NOG2	NA	NA	83/85
PAN6	91/90	NA	83/62

[a] Percent nucleotide identity/percent amino acid identity.

(v1.4) multiple sequence alignment, using MacVector 7.1.1 software (Accelrys, San Diego, CA). The PAUP 4.0 program was used to create phylogenetic trees [27]. Distance matrices were computed using the neighbor-joining (NJ) method. Trees were displayed as unrooted phylograms with branch lengths corresponding to the number of substitutions between two nodes. One thousand bootstrap replicates were generated to verify phylogenetic groupings. The rates of synonymous (*Ks*) and nonsynonymous (*Ka*) substitutions were calculated according to the method of Li [28].

RT-PCR analyses. Total RNA was isolated from cells grown in YPD broth (30 and 37 °C, 180 rpm) using the RNAqueous Midi kit (Ambion, Austin, TX) following the mechanical disruption of cells with glass beads. Superscript III One-Step RT-PCR System with Platinum *Taq* DNA polymerase (Invitrogen, Carlsbad, CA) was selected for RT-PCR analyses, because of its high sensitivity and specificity. The primers were selected in adjacent exons flanking a single predicted intron of no less than 50-bp so as to yield an expected spliced product [29–31]. Primer3 program was used for design (http://www.broad.mit.edu/cgi-bin/primer/primer3_www.cgi; [32]). The details of primers, melting temperatures (T_m), and expected product sizes are given in Supplementary Table 1. RT-PCR was also carried out for actin (*ACT1*) as a positive control with primers that were designed from two exonic regions to yield 594-bp DNA and 543-bp RNA products due to splicing of one intron of 51 bp [15]. RT-PCR products were run on 1–2% agarose gels, stained with ethidium bromide, and their sizes were estimated by visual comparisons with 1 kb and 100 bp DNA ladders. The spliced amplicons were excised from agarose gels, purified with a commercial kit, and sequenced with respective primer sets to confirm their identity with the predicted gene sequences.

Accession numbers. The GenBank accession numbers for nucleotide sequences in four fragments are as follows: AY421964 for the 7.2-kb fragment, AY421965 for the 17.27-kb fragment, AY421966 for the 17.32-kb fragment, and AY421967 for the 34.0-kb fragment. The corresponding ORF information in these fragments is available at the GenBank website.

Results

Cn var. gattii cosmid clones

The estimated *Cn* genome size is approximately 23 Mb (Wickes et al., 1994). Therefore, the number of clones required to ensure a 99% probability of a given DNA sequence being contained within a cosmid library composed of an average 40-kb inserts is 2646 clones [33]. Our *Cn* var. *gattii* cosmid library contained 62,950 clones with average insert size of 38 kb (details not shown). One-third of the library was screened for housekeeping genes, yielding a total of 19 clones positive for *ACT1* and 15 clones positive for *URA5*. Insert identity in the representative clones was confirmed by Southern hybridization; one clone each was also sequenced to further confirm its identity either as *ACT1* or *URA5*. This high-efficiency cosmid DNA library of *Cn* var. *gattii* was then screened by colony hybridization with PCR-amplified fragments of *Cn STE20α*, *STE12α*, or *MFα1*. Six, three, and five cosmid clones, respectively, were selected with *STE20α*, *STE12α*, or *MFα1* probes. Southern hybridization was performed on these positive clones. *Bam*HI or *Eco*RI digested cosmid DNA from three of six clones gave positive hybridization signal with *STE20α* probes (data not shown). Similarly, two out of three *STE12α* clones, and one out of five *MFα* clones, tested positive (details not shown). Multiple PCRs were performed on these six positive clones, using the primers for mating-related genes such as *MFα1*, *STE12α*, *STE20α*, *STE11α*, and *MYO2α*. One cosmid clone was PCR-positive for *STE20α*, *STE11α*, and *MYO2α*, while the second cosmid clone gave positive PCR amplicons for *MFα1* and *STE12α* (data not shown). These cosmid clones (~35 and ~40-kb) were selected for final analyses with transposon insertions for directional nucleotide sequencing, and by using custom-made primers for filling in sequence gaps.

MATα locus genes

Sequencing was carried out on about 380 sub-clones and the sequences were aligned to identify overlapping fragments. Four discrete segments (total ~75.8 kb) of the *MATα* locus were fully assembled. Fragment I (GenBank Accession No.: AY421964) 7221 bp contains *PRT1α*, *ZNF1α*, and a conserved hypothetical protein gene (*CHP01*). Fragment II (GenBank Accession No.: AY421965) 17,271 bp contains a conserved hypothetical protein gene (*CHP02*), *STE12α*, *RPL39α*, and *MFα* genes. Fragment III (GenBank Accession No.: AY421966) 17,317 bp contains *NOG2*, *PAN6*, *HSP12*, *APG9*, *YAH1*, *NFS1*, and two conserved hypothetical protein genes (*CHP03* and *CHP04*). Fragment IV (GenBank Accession No.: AY421967) 34,039 bp contains *CAP1α*, *CID1α*, *LPD1α*, conserved hypothetical protein gene (*CHP05*), and the *RPO41α*, *STE11α*, *MYO2α*, and *STE20α* genes. The detailed homologies of these genes are provided in Supplementary Table 2. Fragments I, II, and III came from the cosmid clone 2, and fragment IV originated from cosmid clone 1. The assembly of fragment IV revealed uninterrupted sequences encompassing the whole of the cosmid 1, which allowed this fragment's easy orientation on the *MATα* locus, since it contained *CAP1*, a gene that has a fixed position adjacent to the left or upper boundary of the *MATα* locus in the other two *Cn* varieties. Similarly, fragment III, which originated from cosmid 2 and which revealed a conserved order of non-*MATα* related genes (*NOG2*, *PAN6*, *HSP12*, and *APG9*), allowed identification of the right or lower boundary of the *MATα* locus. Fragment III was sequenced to the very end and expanded non-*MATα* locus region with identification of two genes *YAH1* and *NFS1* not shown on the previous maps from other *Cn* varieties. Fragments I and III are at the two ends of cosmid 1 in pWEB vector, as confirmed by further sequencing. Therefore, it was easy to orient fragments I and II relative to the other two fragments, and propose a preliminary map of the locus vis-à-vis other two *Cn* varieties (Fig. 1).

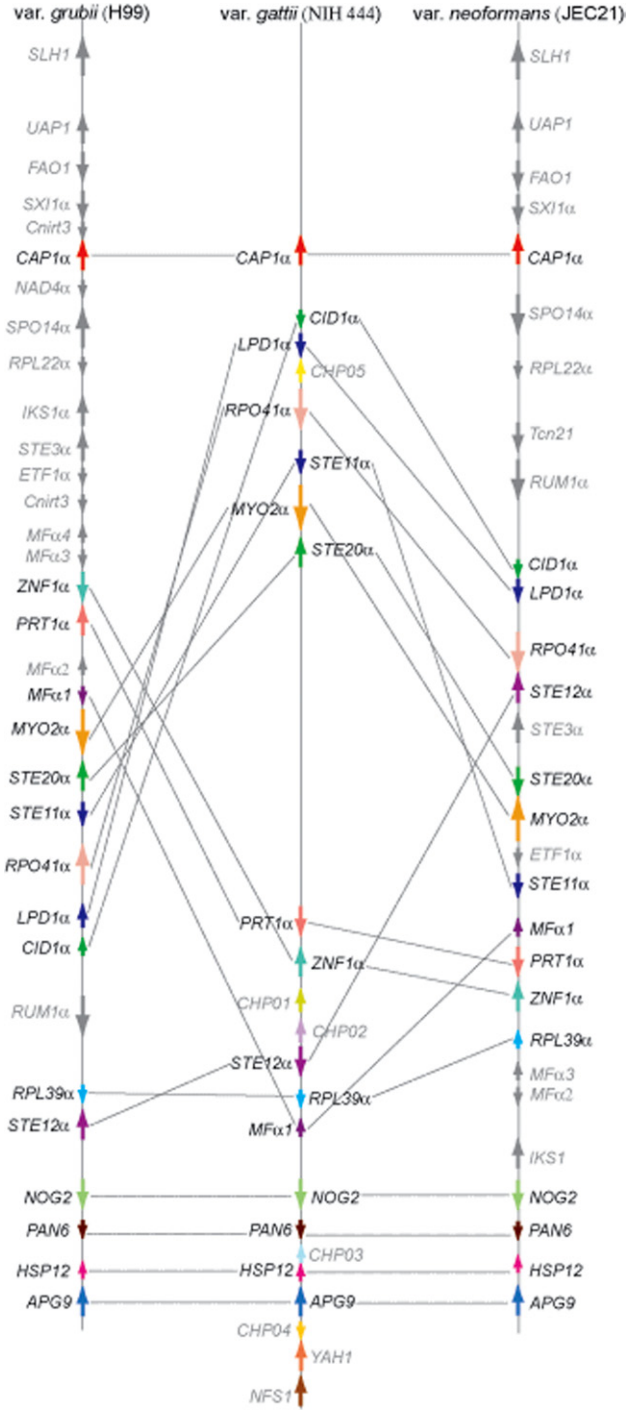

Fig. 1. The genomic organization in *Cn* var. *gattii MATα* locus assembled in this study compared to completed loci in var. *grubii* and var. *neoformans*. The gene order and orientation for the latter two species are according to [35]. Microsynteny at the periphery of the locus and reorganization in genes within are salient features of this comparison.

MATalpha genes in Cn varieties

Twelve genes (*CAP1α*, *CID1α*, *LPD1α*, *RPO41α*, *STE11α*, *MYO2α*, *STE20α*, *PRT1α*, *ZNF1α*, *STE12α*, *RPL39α*, and *MFα1*) identified in our sequencing project in *Cn* var. *gattii* are *MATα* locus-specific in *Cn* var. *grubii MATα* strain H99 as well as in var. *neoformans MATα* strains JEC21. Two genes (*YAH1* and *NFS1*) and the gene for a hypothetical protein (*CHP04*) are uniquely situated at the lower/left periphery of the var. *gattii* locus. Three conserved hypothetical protein genes found in fragments I, II, and IV have very close homologs in both JEC21 and H99. Genes found in fragment III *NOG2*, *PAN6*, *HSP12*, *APG9*, *YAH1*, and *NFS1* flank the *MATα* locus at the lower or right boundary. Two conserved hypothetical protein genes in this fragment have homologs in JEC21 and H99. In this upper or left flanking region (Fig. 1), the *NOG2*, *PAN6*, *HSP12*, and *APG9* show synteny with both *Cn* var. *neoformans* and var. *grubii*. However, the provisional gene order within the var. *gattii MATα* locus is slightly different from that either in var. *neoformans* or var. *grubii* (Fig. 1).

Comparisons of percentage of nucleotide and amino acid identities among *MATα* genes in three *Cn* varieties are summarized in Table 1. We also included corresponding sequences of another *Cn* var. *gattii* strain (WM276, serotype B) recently released by the Genome Center of the University of British Columbia, Vancouver, BC, Canada (http://www.bcgsc.ca/gc). Surprisingly, a number of *MATα*-locus genes from these two var. *gattii* strains do not share a very close sequence identity. This sequence divergence is more noticeable when comparisons are made to var. *grubii* and var. *neoformans* strains (Table 1). Notably, well-known fungal transcription factor *STE12α* is predicted to encode a protein 832 amino acids in var. *gattii* NIH 444 compared to 841 amino acids in var. *grubii* and 855 amino acids in var. *neoformans*. The protein kinase kinase kinase gene *STE11α* is predicted to encode a protein of 1232 amino acids in var. *gattii* NIH 444, compared to 1230 amino acids in var. *grubii* (80% identity) and var. *neoformans* (79% identity). In contrast, *STE20α*, another well-studied protein kinase gene in *MATα* locus, has 645 amino acids in all three *Cn* varieties with a high percentage (around 90%) of identity of nucleotides and amino acids in the three *Cn* varieties.

RT-PCR analyses

The results of PCR and RT-PCR for verification of predicted genes are summarized in Fig. 2. A particular primer pair was first used in PCR with genomic DNA and then in RT-PCR with total RNA, and the products were run on gel in pairs for side-by-side comparison of sizes. Each gel run also included *ACT1* as control for the reactions. The results showed that 23 predicted genes/ORFs were correctly spliced; single bands were obtained for 9 genes, and multiple bands were obtained for 14 predictions. The latter could be due to spliced variants of the target gene, transcript fragments or nonspecific amplifications. We did not try further optimization of RT-PCR parameters as long as the band

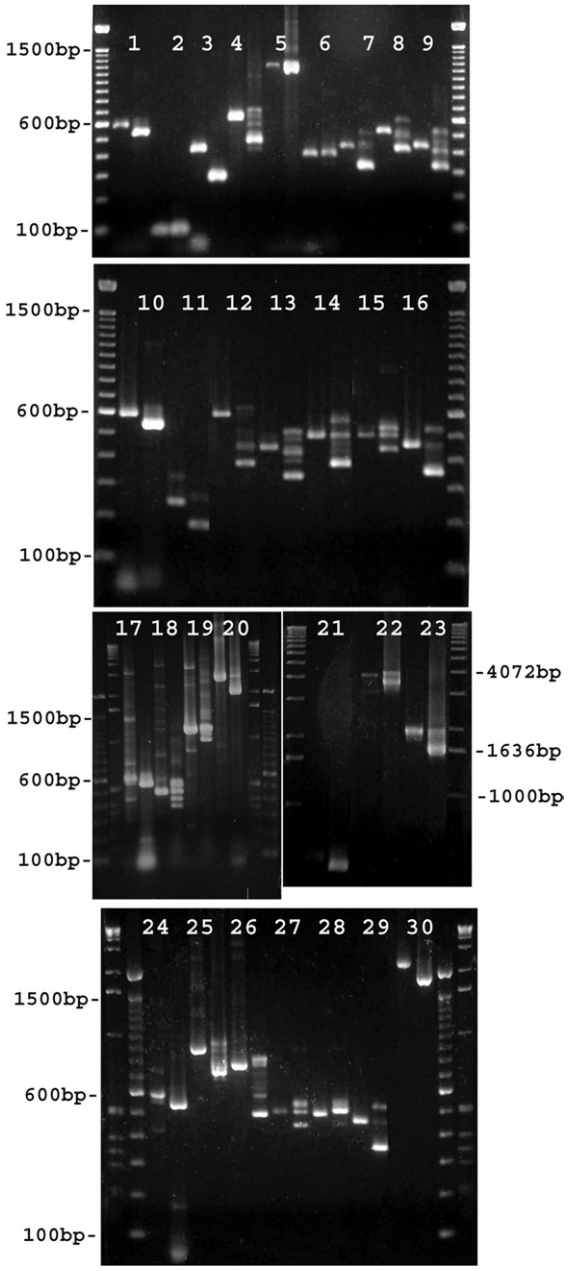

Fig. 2. Detection of predicted spliced sequences in *Cn* var. *gattii* genes by RT-PCR. The primers were designed from adjacent exons flanking a single predicted intron of no less than 50-bp so as to yield an expected spliced product (Supplementary Table 2). Positive control was *Cn ACT1* (actin). One kilobase and 100 bp DNA ladders were used as size markers on ethidium bromide stained agarose gels. PCR and RT-PCR products from a particular gene were loaded in consecutive lanes as labeled: (1, 10, 17, 21, 24) *ACT1*, (2) *MFα1*, (3) *STE20α*, (4) *MYO2α*, (5) *STE12α*, (6) *NSF1*, (7) *LPD1α*, (8) *PRT1α*, (9) *CHP01*, (11) *HSP12*, (12) *PAN6*, (13) *APG9*, (14) *NOG2*, (15, 27) *CHP02*, (16, 29) *CHP05*, (18) *RPL39α*, (19) *YAH1*, (20) *CID1α*, (22) *RPO41α*, (23) *ZNF1α*, (25) *STE11α*, (26) *CHP04*, (28) *CHP03*, and (30) *CAP1α*. The expected spliced products were excised from gels, and sequenced to confirm their identity (details not shown).

of expected size was present, and it could be excised for nucleotide sequencing for confirmation of its identity. No differences in transcription patterns were observed when RNA from cells grown at 37 °C was used in RT-PCR analyses (details not shown).

Phylogenetic analysis

Cn var. *gattii* and var. *grubii*/var. *neoformans* always segregated into two distinct clades in phylogenetic trees constructed by neighbor-joining method for genes with homologs in other *Cn* varieties. The branch lengths separating two clades in different genes were quite variable, from a minimal length corresponding to 0.05 changes (*LPD1α*) to a maximum of 680 changes (*MYO2α*). These branches were supported by high bootstrap values and revealed highly congruent phylogeny for the three varieties (Fig. 3). Notably, *MFα1* did not follow this general trend as it separated from var. *grubii* with good bootstrap support; this is similar to our previous characterization of large number of *MFα1* in various *Cn* strains [13]. Interestingly, the inclusion of a second strain of var. *gattii* yielded two sister groups albeit with low bootstrap values that precluded any meaningful conclusions; this was also consistent with our previous observations on multiple sister groups in *MFα1* of var. *gattii* [13]. The ratio of the number of nonsynonymous nucleotide substitutions per nonsynonymous site (*Ka*) to the number of synonymous nucleotide substitutions per synonymous site (*Ks*) was also determined. The values of *Ka*/*Ks* for mating-associated genes (*STE12α*, *STE11α*, and *MYO2α*) were nearly twice as high as compared to those of non-mating genes in *MATα* locus indicative of higher selection pressure on these genes (Fig. 4). Most notably, *MFα1* was an exception to this pattern in comparisons between var. *gattii* and var. *neoformans* genes, an observation consistent with the phylogenetic tree for this gene, and indicative of a closer evolutionary link between pheromone of these two varieties.

Discussion

New genes in var. gattii

The highlight of our study is the identification of 23 new structural genes in *Cn* var. *gattii*. Additionally, the verification of transcription of these genes both at 30 and 37 °C validated their potential as valid targets for future studies on biology and virulence of this pathogen. Our results also complement key findings previously reported for completed mappings of *MAT* loci in var. *neoformans* and var. *grubii* [34,35]. These included: a relatively large size of the locus, spanning 50–100 kb, with more than 20 genes; conserved gene order adjacent to the locus boundaries; noticeable rearrangement of gene order and orientation within the locus; and presence of many essential genes with no known role in mating. Even though expression analyses

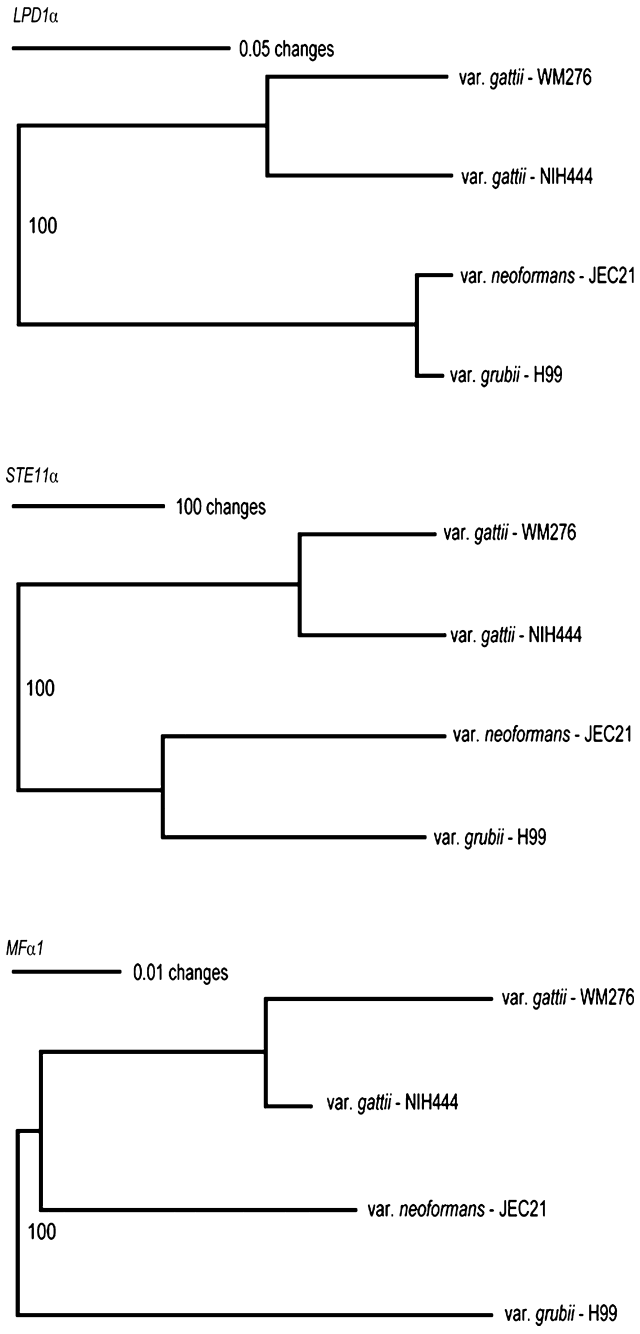

Fig. 3. Neighbor-joining (NJ) phylogenetic trees for three representative genes from var. *gattii* (strains NIH 444 and WM276) compared to var. *grubii* (H99) and var. *neoformans* (JEC21). Bootstrap values are percentages based on 1000 replicates. Var. *gattii* and var. *grubii*/var. *neoformans* segregated into two distinct clades for all genes examined except for *MFα1*. Interestingly, two strains of var. *gattii* also showed two sister groups albeit with weak bootstrap support.

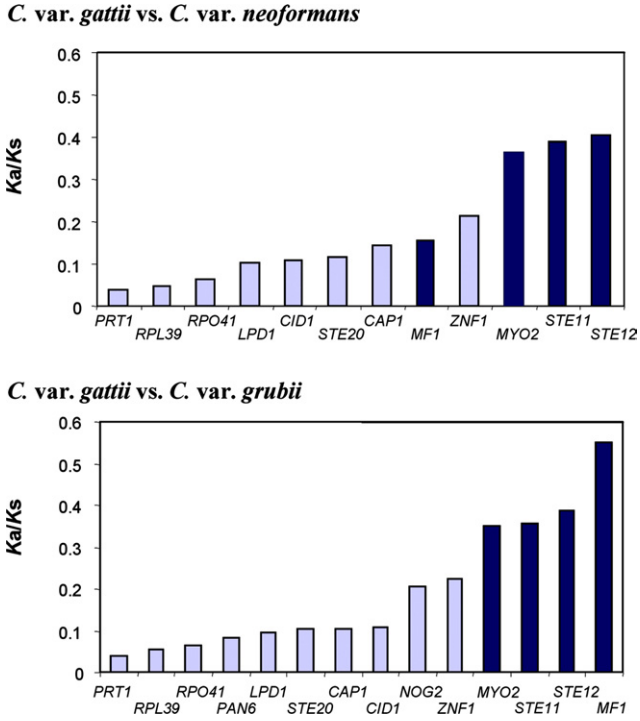

Fig. 4. Comparison of nonsynonymous (*Ka*) to synonymous (*Ks*) substitutions for *MATα* locus genes in three *Cn* varieties. The ratios (*Ka*/*Ks*) were calculated according to the method of [28]. Genes with experimental evidence of a role in *Cn* mating are plotted as darker bars. These genes exhibited highest ratios compared to other genes in *MATα* locus, which is indicative of a positive selection pressure.

have not yet been reported, it is safe to surmise based on sequence similarities that most of the genes in other *Cn* varieties are likely to have similar profiles as observed for var. *gattii*. The band sizes of expected RT-PCR products in this study varied from 100 to 3560 bp, which required a few optimizations steps. However, this product range was easily achieved because the stated efficiency of Superscript III One-Step RT-PCR System is up to 4000 bp. Overall, our results confirm the usefulness of RT-PCR and sequencing to confirm computational predictions of novel genes, which has been a very successful tool in large-scale projects on mouse, human, and rat genomes [25,29,30]. Further application of this approach is presently underway in our laboratory to obtain full-ORF cDNA clones for these genes.

MATα locus complexity

Cn MATα locus in three varieties appeared to be intermediate in organization and complexity when compared to similar loci from other ascomycete and basidiomycete fungi. The *Cn* mating system comprises a single locus, and two-allele organization, which is reminiscent of mating loci in the ascomycetes *Neurospora crassa*, *Podospora anserina*, and *Cochliobolus heterostrophus*, with *Ustilago hordei* being the sole basidiomycete that is known to have this system. However, the physical size of the locus, with a span greatly exceeding 50 kb, is more in line with the complex, tetrapolar mating-type loci known from *Ustilago maydis*, *Schizophyllum commune*, and *Coprinus cinereus* [34–38]. Also, the presence of essential, housekeeping genes within the *Cn* locus is similar to observations reported from such diverse sources as mating-type (*MT*) loci of the haploid green alga

Chlamydomonas reinhardtii and the human Y chromosome [34,39,40]. It has been postulated that *Cn MAT*-locus complexity represents early evolutionary steps in the formation of a 'sex chromosome' [35]. Interestingly, most of the *MATα* locus-specific genes found in var. *neoformans*/var. *grubii* have been localized in this study to a much smaller portion of the var. *gattii* genome. However, we are aware that cosmid clones used for the mapping of the *MATα* locus are limited in the size of their contiguous DNA (~40 kb), and are therefore unlikely to provide an accurate estimation of the actual size of the locus; this limitation may even lead to misassembly. We are confident that the latter was avoided since all peripheral genes in var. *gattii* appeared well conserved, similar to previously published studies in other two varieties [34,35]. However, we are expanding this study with BAC libraries to measure accurately the size of *MATα* and corresponding *MATa* loci in *Cn* var. *gattii*.

MATα locus evolution

The microsynteny in the DNA region flanking the *MATα* loci from all *Cn* varieties clearly suggests a positive selection pressure to conserve gene order adjacent to the locus. It has been observed that the gene order is widely conserved around fungal mating loci; perhaps, this ensures the inheritance of the entire locus following meiotic exchanges [34,41,42]. The evidence for a more rapid evolution of sex-related genes in the *MATα* locus was suggested by rearrangements in gene order and orientation seen in all three varieties, and by higher *Ka*/*Ks* values for genes that have been experimentally shown to be important for mating in *Cn* [10–12] or in *U. maydis* [43]. These findings expand our previous report on evolutionary changes in the *MFα1* pheromone gene [13]. The higher rate of evolutionary changes in mating and sex-associated genes has been reported from other fungi, alga, fruit flies, and mammalian sperm proteins [40,44–46]. Excellent experimental studies in *Drosophila melanogaster* and related species, and mammalian sperm proteins have provided the conceptual framework for the existence of rapidly evolving Sex and Reproduction Related (SRR) genes [44,45]. Further analyses of SRR genes have led to the discovery of 'speciation gene' [47]. Even though experimental studies with fungal systems have not progressed this far, it is pertinent to recall that specialized genes for nonself recognition ('vegetative incompatibility') are well studied in filamentous ascomycetes [48]. It is a moot point whether future characterization of *MAT*-related genes will provide the first clues about the existence of any speciation genes in fungi.

Cn evolution and speciation

No notable deletions or duplications of ORFs were detected in the *MATα* locus of *Cn* var. *gattii*. We were surprised not to find *STE2/3* pheromone receptor, and a second copy of *MFα1*; however, this absence is most likely due to existent gaps in our assembled sequences. Previously, we reported the existence of two copies of *MFα1* in var. *gattii*, three copies in var. *neoformans*, and four copies in var. *grubii*, based on DNA hybridization studies [13]. The presence of multiple *MFα1* in the three *Cn* varieties was subsequently confirmed by genome sequencing [35]. Remarkably, pheromone genes are the only ORFs duplicated in this locus; all other genes are present in single copies in all three *Cn* varieties, suggesting that the evolutionary changes in the locus do not involve gene duplication and loss as a general mechanism, which is extensively recorded in ascomycetous yeasts [49]. The observed re-sorting of *MATα* genes in three loci is unlikely to act as strict barrier for intervariety mating in *Cn*, because changes in gene orders and orientations within closely related taxa are unlikely to have a profound impact, and intervariety mating is known to be feasible in related species, at least in the well-studied model organism *Saccharomyces* [50,51]. Nevertheless, the genomic changes in the *MATα* locus are clearly indicative of either an ongoing or recently completed reproductive isolation in various *Cn* populations. The latter inference is also supported by the failure of recombination in progeny resulting from crosses between var. *gattii* and var. *neoformans* strains [52]. Thus, accumulating evidence suggests the existence of post-zygotic isolation as part of speciation in *C. neoformans*. Along these lines, it appears that *Cn* var. *gattii* best fits the model for allopatric speciation, based on its restricted distribution in tropical and sub-tropical areas, and its recently recognized occurrence in North American coastal temperate rain forests. In contrast, the most abundant of three varieties, var. *grubii*, occurs worldwide, including areas in which its distribution overlaps with that of var. *gattii* [1,5]. Even overlaps in the two populations have not yielded overwhelming evidence of hybrid strains due to intermixing (AB or AC strains). This is unlike the AD hybrids recognized for *Cn* var. *neoformans* and var. *grubii*. The latter strains are true hybrids, based on multiple experimental criteria. They have most prominently been found in Southern Europe, where var. *neoformans* and var. *grubii* are both abundant, with the resultant intermixing and formation of hybrids. However, these hybrids are generally sterile, as would be expected of crosses between independently evolving populations [13,53,54].

Cn species complex

It is pertinent to recall that the current nomenclature of *Cn* var. *gattii* is in a state of flux, even though the evidence for the existence of distinct biological and phylogenetic species is overwhelming. All sequences generated in this study were deposited in the GenBank as *Cn*. var.

gattii, but they were curated as *C. bacillisporus*. Historically, the sexual form or teleomorph of this fungus was given the name *Filobasidiella bacillispora*, with the corresponding anamorphic or asexual form termed *Cryptococcus bacillisporus* [16,55,56]. However, a number of subsequent taxonomic studies led to *C. bacillisporus* being re-classifed as *Cn* var. *gattii* [56]. Recently, it has been proposed that two pathogenic *Cryptococcus* species be recognized, namely, *C. neoformans* and *C. gattii* [17]. All phylogenetic analyses in this study, and similar analyses of other genes, have revealed that *Cn* var. *gattii* and var. *neoformans*/var. *grubii* constitute two distinct groups (reviewed by [17]). The recognition of these two separate species by the medical mycology community is unlikely to pose a practical problem, as many laboratory reagents are already available for routine differentiation. Moreover, laboratory identification of two pathogenic *Cryptococcus* species is likely to be clinically useful, as it will lead to a better understanding of epidemiology, and disease progression and outcome.

Acknowledgments

We thank Dr. Brendan Loftus, The Institute for Genomic Research, Rockville, MD, for his generous help at various stages of this project. Drs. Sudha Chaturvedi and Adriana Verschoor offered valuable editorial comments. Thanks are also due to Jan Galligan, Wadsworth Center Photography and Illustration Unit for artwork, and the Center's Molecular Genetics Core Facility for nucleotide sequencing. This work was partially supported with funds from the National Institutes of Health (AI 053732). We also acknowledge the use of the following genome sequencing projects supported at other institutions: Stanford University (NIH AI47087), TIGR (NIH AI48594), Duke University (NIH AI50113), University of Oklahoma (NIH AI47079), and University of British Columbia (Canadian Institute of Health Research).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbrc.2004.11.017.

References

[1] K.J. Kwon-Chung, J.E. Bennett, Medical Mycology, Lea & Febiger, Philadelphia, PA, 1992.
[2] B. Speed, D. Dunt, Clinical and host differences between infections with the two varieties of *Cryptococcus neoformans*, Clin. Infect. Dis. 21 (1995) 28–34.
[3] S.P. Franzot, I.F. Salkin, A. Casadevall, *Cryptococcus neoformans* var. *grubii*: separate varietal status for *Cryptococcus neoformans* serotype A isolates, J. Clin. Microbiol. 37 (1999) 838–840.
[4] C. Stephen, S. Lester, W. Black, M. Fyfe, S. Raverty, Multispecies outbreak of cryptococcosis on southern Vancouver Island, British Columbia, Can. Vet. J. 43 (2002) 792–794.
[5] J.A. Fraser, R.L. Subaran, C.B. Nichols, J. Heitman, Recapitulation of the sexual cycle of the primary fungal pathogen *Cryptococcus neoformans* var. *gattii*: implications for an outbreak on Vancouver Island, Canada, Euk. Cell 2 (2003) 1036–1045.
[6] P. Kibsey, Update on *Cryptococcus neoformans* var. *gattii*., CPMT Connect. 7 (2003) 1–2.
[7] M. Fyfe, W. Black, M. Romney, P. Kisbey, L. MacDougall, M. Starr, M. Pearce, C. Stephen, L. Stein, S. Mak, B. Emerson, J. Issac-Renton, D. Patrick, in: 5th International Conference of *Cryptococcus* and Cryptococcosis, Melbourne, Australia, 2002.
[8] K.J. Kwon-Chung, J.E. Bennett, Distribution of alpha and a mating types of *Cryptococcus neoformans* among natural and clinical isolates, Am. J. Epidemiol. 108 (1978) 337–340.
[9] C.L. Halliday, T. Bui, M. Krockenberger, R. Malik, D.H. Ellis, D.A. Carter, Presence of alpha and a mating types in environmental and clinical collections of *Cryptococcus neoformans* var. *gattii* strains from Australia, J. Clin. Microbiol. 37 (1999) 2920–2926.
[10] B.L. Wickes, U. Edman, J.C. Edman, The *Cryptococcus neoformans STE12alpha* gene: a putative *Saccharomyces cerevisiae STE12* homologue that is mating type specific, Mol. Microbiol. 26 (1997) 951–960.
[11] C.L. Yue, L.M. Cavallo, J.A. Alspaugh, P. Wang, G.M. Cox, J.R. Perfect, J. Heitman, The *STE12alpha* homolog is required for haploid filamentation but largely dispensable for mating and virulence in *Cryptococcus neoformans*, Genetics 153 (1999) 1601–1615.
[12] D.L. Clarke, G.L. Woodlee, C.M. McClelland, T.S. Seymour, B.L. Wickes, The *Cryptococcus neoformans STE11alpha* gene is similar to other fungal mitogen-activated protein kinase kinase kinase (MAPKKK) genes but is mating type specific, Mol. Microbiol. 40 (2001) 200–213.
[13] V. Chaturvedi, J. Fan, B. Stein, M.J. Behr, W.A. Samsonoff, B.L. Wickes, S. Chaturvedi, Molecular genetic analyses of mating pheromones reveal intervariety mating or hybridization in *Cryptococcus neoformans*, Infect. Immun. 70 (2002) 5225–5235.
[14] G.N. Latouche, T.C. Sorrell, W. Meyer, Isolation and characterisation of the phospholipase B gene of *Cryptococcus neoformans* var. *gattii*, FEMS Yeast Res. 2 (2002) 551–561.
[15] S.D. Narasipura, J.G. Ault, M.J. Behr, V. Chaturvedi, S. Chaturvedi, Characterization of Cu,Zn superoxide dismutase (*SOD1*) gene knock-out mutant of *Cryptococcus neoformans* var. *gattii*: role in biology and virulence, Mol. Microbiol. 47 (2003) 1681–1694.
[16] K.J. Kwon-Chung, J.E. Bennett, T.S. Theodore, *Cryptococcus bacillisporus* sp. nov. serotype B-C of *Cryptococcus neoformans*, Int. J. Syst. Bacteriol. 28 (1978) 616–620.
[17] K.J. Kwon-Chung, T. Boekhout, J.W. Fell, M. Diaz, (1557) Proposal to conserve the name *Cryptococcus gattii* against *C. hondurianus* and *C. bacillisporus* (*Basidiomycota*, *Hymenomycetes*, *Tremellomycetidae*), TAXON 51 (2002) 804–806.
[18] A.J. Cozijnsen, B.J. Howlett, Characterisation of the mating-type locus of the plant pathogenic ascomycete *Leptosphaeria maculans*, Curr. Genet. 43 (2003) 351–357.
[19] S. Chaturvedi, B. Rodeghier, J. Fan, C.M. McClelland, B.L. Wickes, V. Chaturvedi, Direct PCR of *Cryptococcus neoformans MATalpha* and *MATa* pheromones to determine mating type, ploidy, and variety: a tool for epidemiological and molecular pathogenesis studies, J. Clin. Microbiol. 38 (2000) 2007–2009.

[20] X. Huang, M.D. Adams, H. Zhou, A.R. Kerlavage, A tool for analyzing and annotating genomic sequences, Genomics 46 (1997) 37–45.
[21] W.H. Majoros, M. Pertea, C. Antonescu, S.L. Salzberg, GlimmerM Exonomy and Unveil: three ab initio eukaryotic genefinders, Nucleic Acids Res. 31 (2003) 3601–3604.
[22] S.E. Cawley, A.I. Wirth, T.P. Speed, Phat—a gene finding program for *Plasmodium falciparum*, Mol. Biochem. Parasitol. 118 (2001) 167–174.
[23] P. Flicek, E. Keibler, P. Hu, I. Korf, M.R. Brent, Leveraging the mouse genome for gene prediction in human: from whole-genome shotgun reads to a global synteny map, Genome Res. 13 (2003) 46–54.
[24] J.E. Allen, M. Pertea, S.L. Salzberg, Computational gene prediction using multiple sources of evidence, Genome Res. 14 (2004) 142–148.
[25] A. Bateman, L. Coin, R. Durbin, R.D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E.L. Sonnhammer, D.J. Studholme, C. Yeats, S.R. Eddy, The Pfam protein families database, Nucleic Acids Res. 32 (Database issue) (2004) D138–D141.
[26] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[27] D.L. Swafford, PAUP*: Phylogenetic Analysis using Parsimony and Other Methods, Sinauer Associates, Sutherland, MA, 2000.
[28] W.H. Li, Unbiased estimation of the rates of synonymous and nonsynonymous substitution, J. Mol. Evol. 36 (1993) 96–99.
[29] N. Miyajima, C.B. Burge, T. Saito, Computational and experimental analysis identifies many novel human genes, Biochem. Biophys. Res. Commun. 272 (2000) 801–807.
[30] R. Guigo, E.T. Dermitzakis, P. Agarwal, C.P. Ponting, G. Parra, A. Reymond, J.F. Abril, E. Keibler, R. Lyle, C. Ucla, S.E. Antonarakis, M.R. Brent, Comparison of mouse and human genomes followed by experimental verification yields an estimated 1019 additional genes, Proc. Natl. Acad. Sci. USA 100 (2003) 1140–1145.
[31] J.Q. Wu, D. Shteynberg, M. Arumugam, R.A. Gibbs, M.R. Brent, Identification of rat genes by TWINSCAN gene prediction, RT-PCR, and direct sequencing, Genome Res. 14 (2004) 665–671.
[32] S. Rozen, H. Skaletsky, Primer3 on WWW for general users and for biologist programmers, Methods Mol. Biol. 132 (2000) 365–386.
[33] J. Sambrook, D.W. Russell, Molecular Cloning: A Laboratory Manual, third ed., CSH Laboratory Press, Cold Spring Harbor, NY, 2001.
[34] M. Karos, Y.C. Chang, C.M. McClelland, D.L. Clarke, J. Fu, B.L. Wickes, K.J. Kwon-Chung, Mapping of the *Cryptococcus neoformans MATalpha* locus: presence of mating type-specific mitogen-activated protein kinase cascade homologs, J. Bacteriol. 182 (2000) 6222–6227.
[35] K.B. Lengeler, D.S. Fox, J.A. Fraser, A. Allen, K. Forrester, F.S. Dietrich, J. Heitman, Mating-type locus of *Cryptococcus neoformans*: a step in the evolution of sex chromosomes, Eukaryot. Cell 1 (2002) 704–718.
[36] E. Coppin, R. Debuchy, S. Arnaise, M. Picard, Mating types and sexual development in filamentous ascomycetes, Microbiol. Mol. Biol. Rev. 61 (1997) 411–428.
[37] J.W. Kronstad, C. Staben, Mating type in filamentous fungi, Annu. Rev. Genet. 31 (1997) 245–276.
[38] L.A. Casselton, N.S. Olesnicky, Molecular genetics of mating recognition in basidiomycete fungi, Microbiol. Mol. Biol. Rev. 62 (1998) 55–70.
[39] B.T. Lahn, D.C. Page, Functional coherence of the human Y chromosome, Science 278 (1997) 675–680.
[40] P.J. Ferris, E.V. Armbrust, U.W. Goodenough, Genetic structure of the mating-type locus of *Chlamydomonas reinhardtii*, Genetics 160 (2002) 181–200.
[41] G. Butler, C. Kenny, A. Fagan, C. Kurischko, C. Gaillardin, K.H. Wolfe, Evolution of the *MAT* locus and its Ho endonuclease in yeast species, Proc. Natl. Acad. Sci. USA 101 (2004) 1632–1637.
[42] T.Y. James, S.R. Liou, R. Vilgalys, The genetic structure and diversity of the A and B mating-type genes from the tropical oyster mushroom, *Pleurotus djamor*, Fungal Genet. Biol. 41 (2004) 813–825.
[43] I. Weber, C. Gruber, G. Steinberg, A class-V myosin required for mating, hyphal growth, and pathogenicity in the dimorphic plant pathogen *Ustilago maydis*, Plant Cell 15 (2003) 2826–2842.
[44] A. Civetta, R.S. Singh, Sex-related genes, directional sexual selection, and speciation, Mol. Biol. Evol. 15 (1998) 901–909.
[45] D.G. Torgerson, R.S. Singh, Sex-linked mammalian sperm proteins evolve faster than autosomal ones, Mol. Biol. Evol. 20 (2003) 1705–1709.
[46] K. O'Donnell, T.J. Ward, D.M. Geiser, H. Corby Kistler, T. Aoki, Genealogical concordance between the mating type locus and seven other nuclear genes supports formal recognition of nine phylogenetically distinct species within the *Fusarium graminearum* clade, Fungal Genet. Biol. 41 (2004) 600–623.
[47] S. Sun, C.T. Ting, C.I. Wu, The normal function of a speciation gene, *Odysseus*, and its hybrid sterility effect, Science 305 (2004) 81–83.
[48] N.L. Glass, D.J. Jacobson, P.K. Shiu, The genetics of hyphal fusion and vegetative incompatibility in filamentous ascomycete fungi, Annu. Rev. Genet. 34 (2000) 165–186.
[49] C. Seoighe, N. Federspiel, T. Jones, N. Hansen, V. Bivolarovic, R. Surzycki, R. Tamse, C. Komp, L. Huizar, R.W. Davis, S. Scherer, E. Tait, D.J. Shaw, D. Harris, L. Murphy, K. Oliver, K. Taylor, M.A. Rajandream, B.G. Barrell, K.H. Wolfe, Prevalence of small inversions in yeast gene order evolution, Proc. Natl. Acad. Sci. USA 97 (2000) 14433–14437.
[50] H. Fujimura, Cell–cell recognition and pheromone response of the yeast *Saccharomyces globosus*, FEMS Microbiol. Lett. 173 (1999) 63–68.
[51] G. Fischer, C. Neuveglise, P. Durrens, C. Gaillardin, B. Dujon, Evolution of gene order in the genomes of two related yeast species, Genome Res. 11 (2001) 2009–2019.
[52] A. Varma, T. Boekhout, M. Diaz, J. Fell, K.J. Kwon-Chung, in: General Meeting. Amer. Soc. Microbiol., 2002.
[53] M. Cogliati, M.C. Esposito, D.L. Clarke, B.L. Wickes, M.A. Viviani, Origin of *Cryptococcus neoformans* var. *neoformans* diploid strains, J. Clin. Microbiol. 39 (2001) 3889–3894.
[54] Z. Yan, X. Li, J. Xu, Geographic distribution of mating type alleles of *Cryptococcus neoformans* in four areas of the United States, J. Clin. Microbiol. 40 (2002) 965–972.
[55] K.J. Kwon-Chung, A new genus, *Filobasidiella*, the perfect state of *Cryptococcus neoformans*, Mycologia 67 (1975) 1197–1200.
[56] K.J. Kwon-Chung, J.E. Bennett, J.C. Rhodes, Taxonomic studies on *Filobasidiella* species and their anamorphs, Antonie Van Leeuwenhoek 48 (1982) 25–38.