The genome of the basidiomycetous yeast and human pathogen Cryptococcus neoformans

Loftus B.J.(1), Fung E.(2), Roncaglia P.(3), Rowley D.(2), Amedeo P.(1), Bruno D.(2), Vamathevan J.(4), Miranda M.(2), Anderson I.J.(1), Fraser J.A.(5), Allen J.E.(6), Bosdet I.E.(7), Brent M.R.(8), Chiu R.(7), Doering T.L.(9), Donlin M.J.(10), D'Souza C.A.(11), Fox D.S.(5,12), Grinberg V.(4), Fu J.(13),Fukushima M.(2), Haas B.J.(6), Huang J.C.(5), Janbon G.(14), Jones S.J.M.(7), Krzywinski M.I.(7), Kwon-Chung J.K.(15), Lengeler K.B.(5,16), Maiti R.(6), Marra M.A.(7), Marra R.E.(5,17), Mathewson C.A.(7), Mitchell T.G.(5), Pertea M.(6), Riggs F.R.(4), Salzberg S.L.(6), Schein J.E.(7), Shvartsbeyn A.(4), Shin H.(7), Specht C.A.(18), Suh B.B.(19), Tenney A.(8), Utterback T.R.(20), Wickes B.L.(13), Wortman J.R.(1), Wye N.H.(7), Kronstad J.W.(11), Lodge J.K.(10), Heitman J.(5), Davis R.W.(2), Fraser C.M.(1), and Hyman R.W.(2)

(1) The Institute for Genomic Research (TIGR), Rockville, MD, U.S.A. (2) Stanford Genome Technology Center, Stanford University, Paolo Alto, CA, U.S.A. (3) Neurobiology Sector, International School for Advanced Studies (SISSA-ISAS), Trieste, Italy (4) DNA Sequencing Facility, The Institute for Genomic Research (TIGR), Rockville, MD, U.S.A. (5) Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC, U.S.A. (6) Bioinformatics, The Institute for Genomic Research (TIGR), Rockville, MD, U.S.A. (7) Genome Sciences Centre, Vancouver, BC, Canada (8) Laboratory for Computational Genomics, Washington University, St. Louis, MO, U.S.A. (9) Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO, U.S.A. (10) Department of Biochemistry and Molecular Biology, Saint Louis University School of Medicine, St. Louis, MO, U.S.A. (11) The Michael Smith Laboratories, The University of British Columbia, Vancouver, BC, Canada (12) Research Institute for Children and the Department of Pediatrics, Lousiana State Health Science Center, Children's Hospital, New Orleans, LA, U.S.A. (13) University of Texas Health Science Center, San Antonio, TX, U.S.A. (14) Unité de Mycologie Moléculaire, Institut Pasteur, Paris, France (15) Molecular Microbiology Section, Laboratory of Clinical Investigation, National Institutes of Health (NIAID/NIH), Bethesda, MD, U.S.A. (16) Institut für Mikrobiologie, Heinrich-Heine-Universität, Düsseldorf, Germany (17) Plant Pathology & Ecology, The Connecticut Agricultural Experiment Station, New Haven, CT, U.S.A. (18) Department of Medicine, Boston University, Boston, MA, U.S.A. (19) Department of Biomolecular Engineering, University of California, Santa Cruz, CA, U.S.A. (20) Joint Technology Center, J. Craig Venter Foundation, Rockville, MD, U.S.A.

Motivation

Cryptococcus neoformans is an opportunistic human pathogen of global importance. The two strains of C. neoformans serotype D JEC21 and B-3501A are highly related but markedly different in their pathogenicity. We report their genome sequences as an important step in the elucidation of the molecular basis for Cryptococcus virulence and for its difference with other fungal pathogens.

Methods

Whole genome shotgun sequencing was performed followed by assembly via CAP4 or the Celera genome assembler. To ensure accurate gene structure annotation, 23,000 JEC21 cDNA clones were also sequenced. Gene models in JEC21 were created combining together the predictions of three different gene-finders (GlimmerM, Phat and Twinscan), the similarity matches with publicly available protein sequences and with Cryptococcus neoformans cDNAs/ESTs using the program Combiner. UTRs were defined using PASA. All gene model structures obtained by this automated

pipeline were then manually curated. Predicted proteins from the JEC21 genome were searched against a variety of completed and ongoing fungal genome projects using the AAT package. Sequence homologs were identified by filtering the results with a stringency requiring a minimum of 60% of overlap within the original coding sequence and >20% identity at the amino acid level. Basidiomycete-specific genes and Cryptococcus-specific genes were identified on the basis of their presence or absence in public databases. A large fraction (60%) of the Cryptococcus proteome was annotated using Gene Ontology (GO) terms. The B-3501A genome assembly was aligned to the reference JEC21 genome using MUMmer and the output was then parsed to determine the presence of insertions, deletions and SNPs. SNPs and Indels were assigned on the basis of their presence in genic and intergenic regions. To map the distribution of SNPs and Indels, the chromosomes of JEC21 were divided into 10kb regions and the frequency of SNPs and Indels within each region were mapped using a color-coding system. As expected, the frequency of SNP/indels is greater in intronic and intergenic regions than in predicted exons and a bias exists towards fewer SNP/indels in coding versus non-coding exons. However, a relatively high percentage of the SNPs (45.45%) and indels (30.35%) are predicted within the coding regions, and likely contribute to the phenotypic differences between these strains.

Results

The 20-Mb genome of C. neoformans contains ~6500 intron-rich genes and encodes a transcriptome abundant in alternatively spliced and antisense messages. The gene organization is considerably more complex than in ascomycetes and is comparable to that observed in Arabadopsis thaliana or Caenorhabditis elegans. ~80 genes were identified that are likely involved in biosynthesis of the capsule (a major virulence factor) and of the cell wall, an essential and unique component of fungi. Most of these genes are unique to C. neoformans and possible drug targets. Comparison of two phenotypically distinct strains reveals variation in gene content in addition to genome sequence polymorphisms. Sequences of JEC21 and B-3501A are 99.5% identical. SNPs and indels are distributed in blocks of high and low sequence polymorphism reflecting the recombination events that occurred during production of these sibling strains. To investigate the genetic basis for their different virulence, genomic regions encompassing JEC21 genes were compared directly with the B-3501A assembly. 99.7% of genes display >98% nucleotide identity. Strain-specific genes were experimentally verified and included a Ras GTPase-activating protein and two proteins of unknown function specific to B-3501A, and four proteins of unknown function specific to JEC21. ~65% of C. neoformans genes have conserved sequence homologs in a sampling of completed fungal genomes, and of these 12% are restricted to the basidiomycete genome Phanerochaete chrysosporium. Another 10% appear to be unique to C. neoformans based on the absence of identifiable homologs in the current public databases. Lineage-specific gene family expansions do not represent the most abundant protein domains within the C. neoformans genome, which are similar to those of ascomycetous fungi. Two of the 11 gene families that appear unique to C. neoformans are involved in capsule formation, and another encodes nucleotide sugar epimerases associated with cell wall formation. Comparison with S. cerevisiae reveals a similar distribution of genes across nearly all functional categories, except for an expansion of the drug efflux transporters of the major facilitator superfamily in C. neoformans, suggesting enhanced transport capability in this environmental yeast. Genome comparison between the divergent fungal pathogens C. albicans and C. neoformans reveals that cell surface proteins implicated in the former's adhesion to epithelial cells are absent in the latter, suggesting C. neoformans binds host cells via distinct mechanisms.

Contact email: roncagli@sissa.it