The *ste3* Pheromone Receptor Gene of *Pneumocystis carinii* Is Surrounded by a Cluster of Signal Transduction Genes

A. George Smulian,*,† Thomas Sesterhenn,* Reiko Tanaka* and Melanie T. Cushion*,†

**Infectious Disease Division, University of Cincinnati College of Medicine and* †*Cincinnati VA Medical Center, Cincinnati, Ohio 45267-0560*

Manuscript received October 3, 2000
Accepted for publication December 18, 2000

ABSTRACT

Although the clinical aspects of *Pneumocystis carinii* pneumonia are well characterized, the basic biology of the causative organism is poorly understood. Most proposed life cycles of *P. carinii* include both asexual and sexual replicative cycles. The two most prominent morphological forms are a trophic form, thought to undergo asexual replication by binary fission, and a cystic form or ascus containing intracystic bodies or ascospores, the products of sexual replication. To facilitate the Pneumocystis genome project, a *P. carinii* f. sp. *carinii* genomic cosmid library and an additional λ cDNA library were generated. A partial expressed sequence tag database, created as part of the genome project, revealed the transcription of meiosis-specific genes and other genes related to sexual reproduction. The ortholog of Ste3, an **a**-factor pheromone receptor, was cloned and genes surrounding the *ste3* locus were examined. Clustered around the *ste3* gene are genes encoding elements functional in the pheromone response signal transduction cascade of model fungal organisms. These include the Ste20 protein kinase, the Ste12 homoeodomain transcriptional regulator, a potential pheromone mating factor, and other DNA-binding proteins. The genomic organization of the *ste3* locus bears significant similarity to that of the mating locus recently described in *Cryptococcus neoformans*. The *P. carinii* genome contains much of the genetic machinery necessary for pheromone responsiveness, and these data support the existence of a sexual replication cycle.

THE organisms known as "*Pneumocystis carinii*" are in fact a family of related fungal pathogens manifesting genetic and phenotypic differences. Originally classed as protozoans, they have been reclassified as ascomycetous fungi on the basis of 16S ribosomal sequence (Edman *et al.* 1988; Stringer *et al.* 1989). Phylogenetic analyses place them in a deep basal branch within the kingdom fungi in the archaeascomycetes, with their closest extant relative being the fission yeast *Schizosaccharomyces pombe*. Although genetic comparisons clearly support a relationship with *S. pombe*, similarities in biological aspects between these organisms have not been adequately assessed.

Most proposed life cycles of *P. carinii* include both asexual and sexual replicative cycles, although little experimental evidence is available to confirm these proposed mechanisms (Cushion 1998a,b). Life cycle studies have been limited to microscopic and ultrastructural observations of organisms isolated from infected mammalian lungs, as the lack of a robust, long-term culture has precluded detailed dynamic studies. The two most prominent morphological forms noted are a trophic form, thought to undergo asexual replication by binary fission, and a cystic form or ascus containing intracystic bodies or ascospores, the products of sexual replication. Evidence for sexual replication is limited to rare reports of synaptonemal complexes, which are structures formed during meiotic recombination of sister chromatids on electron micrographs of early precysts of *P. carinii* (Matsumoto and Yoshida 1984).

Stringer proposed a system for classification of the members of the Pneumocystis family of organisms based on genetic comparisons (Stringer 1996). The highest level of divergence (class III) was observed in Pneumocystis populations obtained from different mammalian hosts. Class II divergence occurred between organisms within a single mammalian host. The genetic divergence observed within the class II and class III levels is consistent with *bona fide* species differences in other yeast. In recognition of these differences, a system of nomenclature that uses the mammalian host to define the population of *P. carinii* was adopted by the scientific community following the tenets of the Botanical Code. Thus, organisms from human beings are *P. carinii* f. sp. *hominis*, and those from mice, *P. carinii* f. sp. *mus*, for example. *P. carinii* f. sp. *carinii* and *P. carinii* f. sp. *ratti* are two populations found within rat lungs and exhibit class II divergence (Cushion *et al.* 1993). Class I divergence was the lowest level with differences ranging from 0–0.8% in the gene sequences and from 2–4% in the internal transcribed sequence of the rRNA gene regions. Organisms

Corresponding author: A. George Smulian, VA Medical Ctr., 231 Bethesda Ave., Cincinnati, OH 45267-0560.
E-mail: smuliang@fuse.net

with this level of divergence have been found in human beings and rats.

Functional genomics can greatly facilitate our understanding of complex biological systems. This is especially pertinent in organisms, such as *P. carinii*, where direct experimentation is difficult. A *P. carinii* genome project has recently been launched to help understand the basic biology of these organisms (Cushion and Arnold 1997). A physical map and genome sequence of a rat-derived population, *P. carinii* f. sp. *carinii*, are anticipated to be produced upon completion of the project. Creation of a partial expressed sequence tag (EST) database as part of the project revealed the transcription of orthologs of meiosis-specific genes and other genes related to sexual reproduction (http://www.biology.uky.edu/Pc; http://www.gene.genetics.uga.edu). Such genes included *mek1*, a meiosis-specific kinase, and *ste3*, the **a**-mating factor pheromone receptor.

Genetic loci determine sexual compatibility resulting in mating-type phenotypes of fungi. With regard to mating type, fungi may be heterothallic (self-sterile), where gamete nuclei must arise from parents of different mating types, or homothallic (self-fertile), where gamete nuclei are genetically identical and mating type cannot be defined. Pseudohomothallism arises when organisms are self-fertile but the gamete nuclei are not identical. In *Saccharomyces cerevisiae*, a pseudohomothallic ascomycete, haploid strains switch mating types, giving rise to progeny of opposite mating types that can then fuse. Organization of mating-type loci may range from simple, two-mating-type loci containing transcription factors that regulate the expression of pheromone-responsive genes scattered throughout the genome, as seen in *S. cerevisiae* and in the heterothallic ascomycete, *Neurospora crassa* (Kronstad and Staben 1997; Haber 1998). More complex tetrapolar mating loci have been described in many filamentous basidiomycetes. In the mushrooms, *Schizophyllum* spp. and *Coprinus* spp., two unlinked loci are designated A and B, and each have subloci designated α and β. The A loci contain transcriptional factor genes regulating nuclear pairing, hook cell formation fusion, and other functions. The genes at the B loci encode multiple pheromones and pheromone receptors (Vaillancourt *et al.* 1997; Casselton and Olesnicky 1998; O'Shea *et al.* 1998; Fowler *et al.* 1999). The most complex mating locus described has been reported in the heterothallic basidiomycete *Cryptococcus neoformans* where the α-mating locus extends over 50 kb. Here a two-allele (idiomoph) single mating locus contains genes encoding transcription factors, multiple pheromones, a pheromone receptor, and components of the pheromone responsive signal transduction pathway, including Ste20, Ste12, and Ste11 (Moore and Edman 1993; Wickes *et al.* 1997; Karos *et al.* 2000; J. Heitman, personal communication).

In the present study, the Ste3 ortholog was cloned and genes surrounding this putative G-coupled transmembrane receptor were examined. The genomic organization of the *ste3* locus bears significant similarity to the mating locus recently described in the heterothallic basidiomycete *C. neoformans*. These data support the hypothesis that *P. carinii* has a sexual cycle.

MATERIALS AND METHODS

Isolation of organisms and genomic DNA: Karyotype form 1 *P. carinii* f. sp. *carinii* organisms were isolated from the lungs of an individual chronically immunosuppressed rat. Male Lewis rats (140–160 g) received weekly injections of 4 mg methylprednisolone acetate for 8–12 weeks to provoke the infection and sterilized water with 0.2 mg/ml cephadrine (Velosef; Squibb, Princeton, NJ) to prevent secondary bacterial infections. Organisms were extracted from the lungs of rats with severe pneumonia by homogenization of the minced lung tissue in a Stomacher 80 lab blender (Fisher Scientific, Cincinnati, OH) as previously described (Linke *et al.* 1989). After homogenization, organisms were collected by centrifugation; red blood cells were lysed in 0.85% (w/v) NH_4Cl (pH 6.8) for 10 min at 37°, and the remaining material was washed twice in RPMI-1640 (Life Technologies, Grand Island, NY). Organisms used for library construction were filtered exhaustively (10–12 times) through 10-μm filters to remove remaining host cell contamination. After isolation, each organism preparation was enumerated by microscopic methods, divided approximately in half, and prepared for (1) pulsed-field gel analysis by embedding in 0.8% (w/v) low melt agarose (Hong *et al.* 1990) and (2) cryopreservation by suspension in RPMI 1640 supplemented with 7.5% (v/v) DMSO and 10% (v/v) fetal bovine serum, followed by storage in liquid nitrogen. Populations were characterized by karyotype profiles produced by contour-clamped homogeneous electric field (CHEF) electrophoresis. DNA was transferred from the CHEF gels to nylon under neutral conditions, UV crosslinked to the nylon membranes, and hybridized under standard conditions to α-^{32}P-labeled DNA probes, as previously described (Cushion 1998b).

Genomic DNA was isolated for cosmid library construction, for restriction digestion, and for Southern blotting from agarose-embedded organisms. Agarose plugs were melted at 70° and diluted fivefold in TE buffer. An equal volume of saturated phenol was added and the aqueous phase separated by centrifugation through PhaseLock gel (Eppendorf, Westbury, NY). The extraction and separation through PhaseLock gel was repeated with phenol:chloroform:isoamyl alcohol and finally chloroform:isoamyl alcohol, and the DNA was precipitated with ammonium acetate and ethanol, according to vendor instructions. The chromosomal-sized DNA was gently reconstituted in TE buffer and quantified by comparative ethidium bromide staining to known DNA standards. Restriction digestion of genomic DNA was performed under standard conditions, resolved by electrophoresis on a 0.65% (w/v) agarose gel, and transferred to positively charged nylon membrane under alkaline conditions. After UV crosslinking, prehybridization and hybridization were performed in Perfect Hyb solution (Sigma, St. Louis) at 60° with appropriate α-^{32}P-labeled probes. Membranes were stripped with boiling 1% SDS prior to reuse.

Cosmid library construction: Form 1 *P. carinii* f. sp. *carinii* was selected for study in the genome project because it was the most prevalent population in surveys of commercial rat vendors (Cushion 1998b). The form 1 *P. carinii* f. sp. *carinii* preparation that was cloned into the cosmid vector pWEB (Epicentre Technologies, Madison, WI) was shown to lack

visible host DNA, as assessed by CHEF gels stained with SYBR-Gold (Molecular Probes, Eugene, OR) and visualized under UV light. DNA in an agarose block was purified using Phaselock PLG 1 light tubes (Eppendorf). Twenty micrograms of this DNA was randomly sheared by passage through a 23-gauge needle to generate fragments between 35 and 45 kb. Sheared fragments were blunt ended with an enzyme mixture containing T4 DNA polymerase and T4 polynucleotide kinase for 45 min at room temperature. Fragments, similar in size to T7 DNA (35–40 kb), were size selected by gel electrophoresis in SeaPlaque GTG agarose (BioWhittaker Molecular Applications, Rockland, ME). The low-melt agarose plug containing these fragments was treated with β-agarase (Epicentre Technologies) for 30 min at 45° and the DNA was isolated by ethanol precipitation. A total of 250 ng of this DNA was ligated to manufacturer-prepared *Sma*I-digested dephosphorylated pWEB cosmid vector. The pWEB-*P. carinii* genomic DNA ligation mix was packaged in two aliquots using MaxPlax packaging extract (Epicentre Technologies) and used to transfect *Escherichia coli* strain EPI103, and plated on Luria broth (LB)-carbenicillin (50 μg/ml) plates. A total of 2486 colonies were picked into microtiter well plates of LB carbenicillin broth (50 μg/ml), grown overnight, and cryopreserved. Plates were shipped to Dr. Jonathan Arnold (Department of Genetics at the University of Georgia in Athens), where they were replicated and robotically arrayed on nylon membranes in a 4 × 4 high-density array with duplicate colony spotting (Kelkar *et al.* 2001).

cDNA construction: Total RNA was isolated from cryopreserved form 1 *P. carinii* f. sp. *carinii* using Trizol reagent (Life Technologies). To enrich for mRNA, the RNA was fractionated by passage over an oligo(dT) column (Stratagene, La Jolla, CA). mRNA was converted to hemi-methylated cDNA using Moloney murine leukemia virus reverse transcriptase in the presence of methyl-dCTP according to manufacturer's conditions (Stratagene). Double-stranded cDNA was ligated to *Eco*RI linkers and digested with *Xho*I (within the oligo(dT) primer) and *Eco*RI to facilitate cloning. The cDNA was ligated undirectionally between *Eco*RI and *Xho*I sites into λZAPII according to manufacturer's instructions (Stratagene). One-fourth of the ligation reaction was packaged with MaxPlax packing extract and used to infect *E. coli* XL1-blue cells to generate 5×10^5 primary transformants. One hundred thousand unamplified phage were rescued with Exassist helper phage in SOLR cells (Stratagene) to generate a library in pSKII for EST sequencing. The remainder of the primary library was subjected to a single round of amplification to yield a final amplified library at a titer of 1×10^{11} pfu/ml. Individual clones from the replicated cDNA library are available from American Type Culture Collection (Manassas, VA; http://www.atc.org).

Isolation of Ste3 cDNA: A candidate *ste3* EST was identified from the searchable *P. carinii* EST database located at the University of Georgia and created by Michael Weise (http://www.gene.genetics.uga.edu/Pc). To isolate a full-length *ste3* cDNA, 50,000 *P. carinii* cDNA λZapII phage were screened with an [α-^{32}P]dCTP-labeled *ste3* probe generated by PCR amplification from *P. carinii* EST s16D3 template. Twelve phage plaques were isolated following hybridization screening of 5×10^4 pfu. Anchor PCR was performed using vector primers in combination with s16D3 gene-specific primers to characterize which plaques extended to include the 5′ and 3′ ends of the Ste3 cDNA. PCR was performed using *Taq* polymerase (Promega, Madison,WI), HotStart*Taq* polymerase (QIAGEN, Santa Clarita, CA), or *Pfu* polymerase (Promega) with appropriate buffers and conditions defined by the specific primers. PCR products were cloned into pCR2.1 using the TOPO-TA cloning kit (Invitrogen, Carlsbad, CA) and sequenced using vector primers T7 and M13rev. A full-length cDNA was generated by PCR amplification of the longest *ste3* cDNA clone using a primer located at the predicted translational start and within the vector 3′ to the insert cloning site, and the full-length transcript was cloned into pCR2.1. Sequence analysis, sequence translation, and protein secondary structure analysis were performed using DNAMAN software (Lynnon BioSoft, Montreal, Quebec).

Cosmid screening and isolation: High-density array filters of 2486 cosmid clones were hybridized under standard conditions with an [α-^{32}P]dCTP-labeled *ste3* probe generated by PCR amplification from *P. carinii* EST s16D3 template (Kelkar *et al.* 2001). The membranes were stripped in boiling 1% (w/v) SDS and reprobed as necessary with additional PCR-generated probes. Five hybridizing cosmid clones were identified when the cosmid library was probed with the *ste3* probe. One clone (W15A6) was selected for further analysis. Cosmid DNA was isolated from a 500-ml culture of organisms grown in LB-carbenicillin (50 μg/ml) broth using a QIAGEN large construct DNA isolation kit according to manufacturer's instructions. The W15A6 *P. carinii* genomic insert was liberated following *Not*I digestion and isolated from the cosmid vector by gel electrophoresis. Insert DNA was sonicated to yield fragments between 0.5 and 4 kb. Fragmented DNA was blunt ended and phosphorylated using T4 DNA polymerase and T4 polynucleotide kinase under standard conditions. Fragments between 1 and 3 kb were isolated by gel electrophoresis and recovered by adherence to silica glass milk (Geneclean; Bio101, Vista, CA). Cosmid fragments were ligated into dephosphorylated *Eco*RV-digested pSKII+ using T4 DNA ligase and transformed into *E. coli* Top 10 competent cells (Invitrogen). One hundred and fifty recombinant clones were identified by blue/white screening, and plasmid DNA was isolated using a QIAGEN Turbo 96 kit on a Qiarobot 3604. Clones were sequenced at the Molecular Genetics Core Facility at the University of Georgia using T7 and T3 sequencing primers. The sequence of cosmid W15A6 has been deposited in GenBank with the accession no. AF309805.

Informatics and software: DNA analysis and manipulation were performed using DNAMAN version 3.2 software (Lynnon BioSoft). Blast analysis was performed using Blast 2.1 against the National Center for Biotechnology Information (NCBI) nonredundant (nr) database via the NCBI webserver (http://www.ncbi.nlm.nih.gov/blast). Sequence assembly was performed using assembly functions within DNAMAN and by CAP3 contig assembly software via a webserver at http://www.gcg.tigem.it/ASSEMBLY/assemble.html. Analysis for transcription factor binding sites was performed using MatInspector V2.2 at the GSF National Research Center for Environment and Health (http://www.gsf.de/cgi-bin/matsearch.pl). Alignment of DNA and protein sequences was performed using ClustalW 1.8 at http://www.dot.imgen.bcm.tmc.edu.9331/multi-align.

RESULTS

Pc *ste3* encodes a putative pheromone receptor: Analysis of the *P. carinii* ESTs released on the Pneumocystis genome website (http://www.biology.uky.edu/Pc/) revealed an EST, s16D3, with homology to a pheromone receptor Rcb3 of *Coprinus cinereus*. The s16D3 clone was isolated from the arrayed cDNA library, and plasmid DNA was purified and sequenced in its entirety. The clone contained a 700-bp open reading frame (ORF) that demonstrated homology to the **a**-factor pheromone

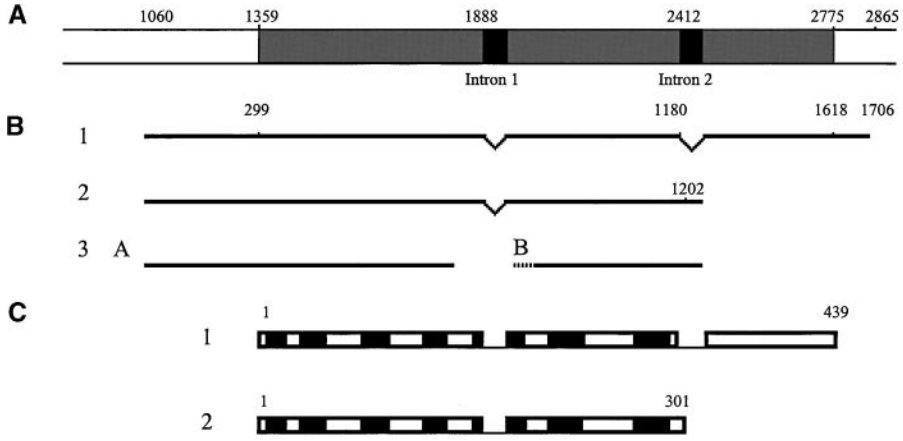

FIGURE 1.—Schematic representation of the *ste3* gene locus, mRNA structure, and protein. (A) The structure of the *ste3* gene within W15A6 is depicted with the coding region shaded gray. The position of the transcriptional start, translational start, and stop is identified. The two introns are shaded in black. (B) Two *ste3* mRNA transcripts (B1 and B2) arise as a result of alternative splicing of the second intron. cDNAs corresponding to each class of transcript were detected in the λZAP cDNA library. The initial EST clone, s16D3, in which homology to pheromone receptors was detected, was generated as an internal fragment (B3A) primed from an internaltract of As. Additional ESTs (s20D1 and s14G9) corresponded to the shorter transcript (B3B). (C) Ste3p encodes a seven-transmembrane G-coupled receptor. The alternatively spliced transcripts give rise to 439- or 301-amino-acid residue polypeptides (C1 and C2), respectively. The seven membrane-spanning helixes (shaded black) are conserved between both putative proteins, while alternative transcripts generated the intracytoplasmic tails of varied length.

receptor Ste3p of *S. cerevisiae*. On the basis of the EST sequence, primers were designed to amplify an internal fragment for use in hybridization screening of the *P. carinii* cDNA λZAP library. Candidate clones were identified by hybridization screening, and anchor PCR was performed to identify λZAP clones extending to include the 5′ and 3′ ends of the *ste3* cDNA. Sequence analysis of the longest λ clone revealed a 5′ untranslated region of 299 bp followed by a 1317-bp ORF and 90-bp 3′ untranslated region (Figure 1B).

The open reading frame, bounded by an ATG start codon in a favorable Kozak orientation, encoded a putative 439-amino-acid protein with a predicted molecular weight of 45 kD. Secondary structure analysis of the putative polypeptide predicted a short extracellular hydrophilic domain, seven membrane-spanning α-helical domains, and a long intracytoplasmic tail (Figure 1C). BlastX analysis (at default settings performed with the NCBI GenPept database) revealed significant homology to several G-coupled pheromone receptors, *e.g.*, Ste3p of *S. cerevisiae*, Pra1p of *Ustilago maydis*, and Bβ1p of *S. commune*, extending through the membrane-spanning regions (Figure 2). The predicted protein includes a histidine asparagine-rich carboxy-terminal tail. Similar simple sequence repeats are seen in the intracytoplasmic tails of a number of pheromone receptors such as Bβ1p of *Schizophyllum commune*. No significant homology was noted to any serpentine receptors other than pheromone receptors.

The sequence of full-length *ste3* cDNA was used to reexamine the EST database to detect additional *ste3*-related ESTs. The original EST, s16D3, appeared to be an internal fragment primed by the oligo(dT) primer from an internal stretch of 10 adenine residues at position +500 (relative to the translational start; Figure 1B). Two additional ESTs, s20D1 and s14G9, were identified by BlastN analysis of the EST database with significant homology to the *ste3* cDNA. These cDNAs were identical to the *ste3* cDNA through position 1180 (Figure 1B). These cDNAs were predicted to encode a 301-amino-acid polypeptide identical to the initial 301 amino acids of Ste3p (Figure 1C). Comparison with the *ste3* gene sequence revealed that these cDNAs arose through alternative splicing of the second intron of *ste3*.

Gene density of cosmid W15A6 containing *ste3*: Genes encoding pheromone receptors may be located within mating loci, as described in *C. cinereus*, *S. commune*, and *C. neoformans* (O'SHEA *et al.* 1998; FOWLER *et al.* 1999; KAROS *et al.* 2000). To better understand the genomic context of the *P. carinii* Ste3p pheromone receptor, the *ste3* gene was cloned from a *P. carinii* genomic library. The organization surrounding the *ste3* gene was explored by sequencing the pWEB cosmid W15A6, one of five cosmids strongly reactive to the *ste3* amplicon used as probe. The W15A6 insert was fragmented by random shearing, subcloned, and sequenced to a 3× overall coverage. Sequences were assembled using CAP3 assembler into 18 contigs. Closure of the gaps was achieved by PCR using primers designed at the end of each contig. Amplicons were directly sequenced, and sequence results were reassembled to generate contiguous sequence spanning the entire 32,083-bp insert.

Putative gene sequences within the cosmid sequence were identified by a combination of three *in silico* methodologies. BlastX analysis identified 14 segments with significant homology [probability values (P) $<10^{-5}$] to sequences within the NCBI database. The coding region of each of these putative genes was determined on the basis of homology to coding sequences of identified orthologs, positioning of potential introns to maintain the open reading frame, and from comparison to EST sequence or sequence amplified from *P. carinii* cDNA. These approaches revealed 14 genes encoded within the W15A6 genomic segment and 1 gene originating

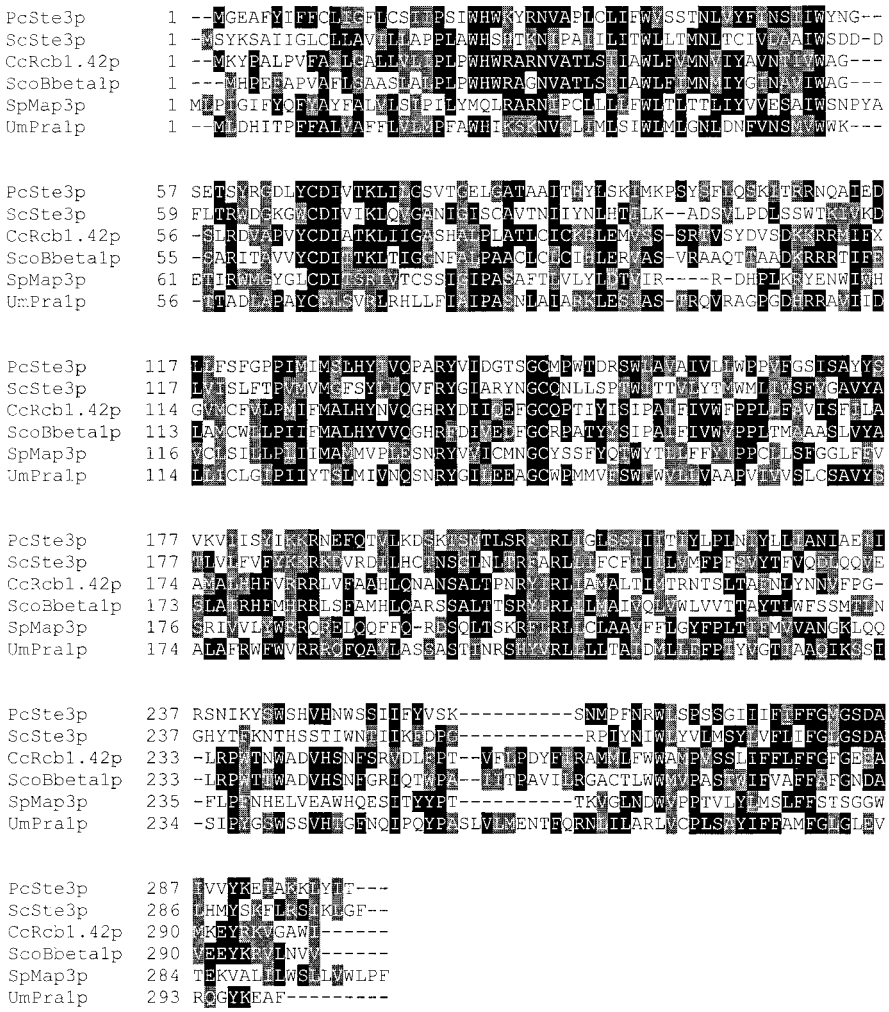

FIGURE 2.—Sequence similarity of pheromone receptors. Alignment of the amino-terminal 300-amino-acid residues of six fungal pheromone receptors demonstrates significant sequence similarity. Alignments that were performed using the ClustalW alignment algorithm show sequence identity (dark shading) and similarity (light shading) between Ste3p of *P. carinii* (PcSte3p), Ste3p of *S. cerevisiae* (ScSte3p), Rcb1.42p of *C. cinereus* (CcRcb1.42p), Bβ1p of *S. commune* (ScoBbeta1p), Map3p of *S. pombe* (SpMam2p), and Pra1p of *U. maydis* (UmPra1p). All aligned pheromone receptors bind **a**-like lipopeptide mating factor.

in a flanking sequence but extending into the W15A6 cosmid (Table 1, Figure 3). The W15A6 sequence is the first extensive segment of *P. carinii* genomic DNA examined. The insert of W15A6 was localized to chromosome 1, the largest-sized chromosome of this 15-band profile (Figure 4). To establish a naming convention, genes were named PCC (*P. carinii* f. sp. *carinii*) A (for genes on chromosome 1) 001–015 (Table 1, column 1). Eight genes bore homology to genes of known function in other organisms. On the basis of the degree of sequence similarity, these eight *P. carinii* genes were named after their orthologs or functions according to recent suggestions by J. R. Stringer and M. T. Cushion (STRINGER and CUSHION 1998) (Table 1, Columns 1 and 2). Six genes were encoded on the Watson strand of DNA (as defined by the coding strand of the *ste3* gene) while nine genes were encoded by the Crick strand of DNA (Figure 3). The borders of the putative open reading frame of each gene are depicted in Table 1, column 3. To identify additional genes not recognized by homology search of GenBank, BlastN analysis was performed against sequences within the *P. carinii* EST database. This identified 17 ESTs encoded from sequences within the W15A6 insert. These included the 8 of the 15 potential genes identified by BlastX search against the NCBI database. Extended intergenic regions were also searched for potential open reading frames using the ORF search function of DNAMAN.

The open reading frames of all gene sequences identified were interrupted by introns (Table 1, column 4). The number of introns identified ranged from 1 to 7 introns per gene with an average of 3.66 per gene sequence or 1 intron per 475 bases of coding sequence. Introns were bounded by consensus donor and acceptor sites and ranged in length from 31 to 235 bases. The average intergenic distance was 498 bases but ranged from 1938 bases between A003 and A004 to −618 where genes A013 and A014 are transcribed off opposite strands and overlap for 618 bases, giving a gene density of one gene/2139 bases for the region examined.

Pc *ste3* locus is flanked by genes functional within a putative pheromone response pathway: W15A6 contained three genes encoding proteins predicted to function within a pheromone response pathway. These included the pheromone receptor *ste3* (A002), the ortholog of the *ste20* protein kinase (A008), and the *ste12* transcrip-

TABLE 1

Sequence homology-based assignment of genes in *P. carinii* cosmid W15A6

Gene	Homologous sequence and accession no.	Gene location	Introns	EST	Comments
A001	Glutathione synthetase	46–844	2	s35F9	5′ end of gene in flanking
gsh1	Gsh2p				sequence
	P35669				
A002	Ste3	1,359–2,775	2	s16D3	**a**-factor pheromone receptor
ste3	B24670			s20D1	
				s14G9	
A003	Rim1p	3,456–2,942	2	s50B9	Zinc-finger single-stranded
	S23548			s37D4	DNA binding protein
A004	Ste12	?–5,426			Homeodomain Transcription
ste12	A33540				factor
A005	YDR124w	8,889–7,254	2		Nuclear transport protein
	S52689				
A006	YAWC	10,390–9,655	2	s40D6	Helix-loop-helix
	T38713			s8F2	DNA-binding protein
A007	T40649	12,475–10,594	4	s18E11	YER124w homologue
A008	Ste20	15,664–13,024	7		Protein kinase
ste20	AF162330				
A009	Sucrose carrier	18,086–15,925	7		Sucrose carrier
sca1	T38541				
A010	Na/H+ exchanger	18,200–20,207	7		Na/H+ exchanger
nhx1	Q04121				
A011	MTS2	20,729–22,230	3	s5E11	26S protease
mts2	P36612				Regulatory subunit 4
A012	AL034583	25,662–22,496		s17H6	Hypothetical protein
A013	Z97052	25,727–28,180	7	s39B6	Hypothetical protein
A014	Glycyl-tRNA	30,445–28,216	9	s20E9	Glycyl-tRNA
grs1	Synthetase			s30H10	Synthetase
	T38704				
A015	YDS4	30,669–31,645	1		Hypothetical coiled-coil
yds4	T38834				protein

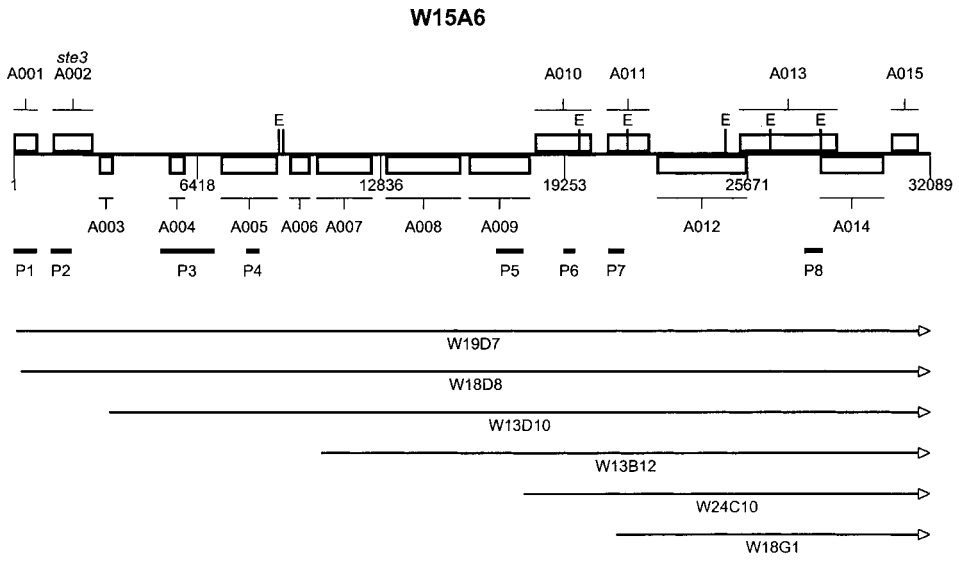

FIGURE 3.—Genetic map of the cosmid W15A6. Schematic representation of the 32,083-bp genomic DNA insert of cosmid W15A6 containing 15 gene sequences (A001–A015) encoding the serpentine receptor, Ste3, and 14 other predicted proteins is shown. Six genes encoded on the sense strand are depicted above the sequence line, while nine genes encoded on the antisense strand are depicted below the sequence line. Predicted *Eco*RI sites within the sequence are marked "E." The positions of hybridization probes (P1–P8) used in the experiments described are shown below the sequence. The relative positions of seven overlapping cosmids used in a chromosomal walk to define the *ste3* locus are shown below the W15A6 sequence line.

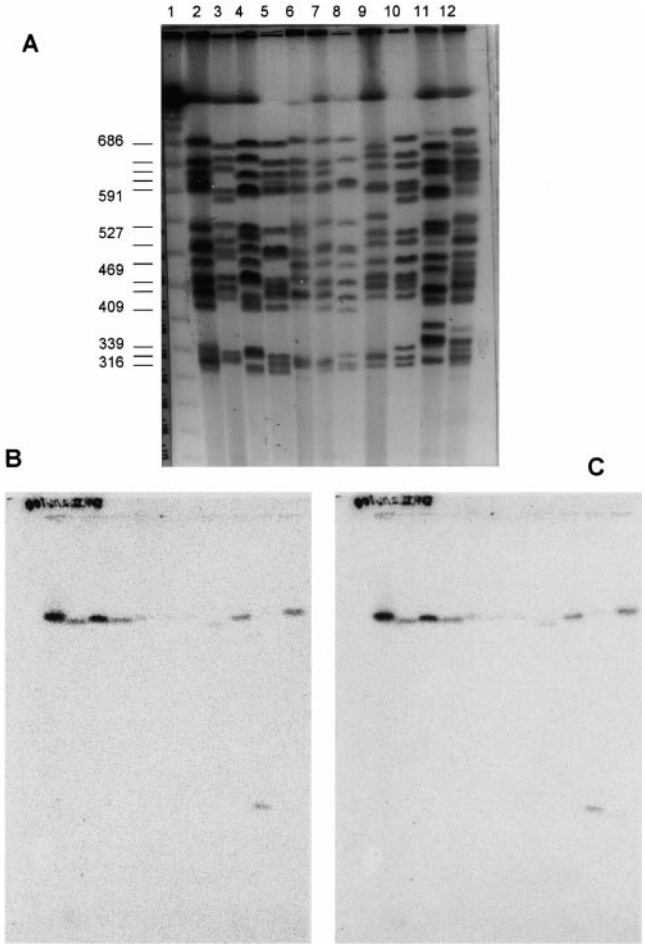

FIGURE 4.—Chromosomal localization of *ste3*. Chromosomes of *P. carinii* were separated by CHEF pulsed-field gel electrophoresis, visualized by Sybr Gold staining, and transferred to nylon membranes for Southern hybridization. Isolates of *P. carinii* f. sp. *carinii* forms 1–9 were separated in lanes 2–10, while an isolate of *P. carinii* f. sp. *ratti* was electrophoresed in lane 11. A mixed isolate *P. carinii* f. sp. *carinii* form 1 and *P. carinii* f. sp. *ratti* was loaded in lane 12. Fourteen chromosomal size bands between 686 and 316 kb are visualized by Sybr Gold staining. (A) For clarity, the inverse image of the fluorescent image is shown. Hybridization with an *ste3* probe (P2 in Figure 3) localized to 686-kb chromosome 1 (B) in all *P. carinii* f. sp. *carinii* isolates and to 345-kb chromosomes in *P. carinii* f. sp. *ratti* isolates. Hybridization with an *mts2* probe (P7) co-localized to chromosome 1. Sizes of selected chromosomes within a *P. carinii* f. sp. *carinii* form 1 pattern are shown on the left in A.

tional activator (A004). Ste20p protein kinase, activated in response to mating factor binding to the appropriate pheromone receptor, phosphorylates the mitogen-activated protein (MAP) kinase kinase kinase, Ste11, within the pheromone-responsive MAP kinase cascade (LEBERER *et al.* 1992; MARCUS *et al.* 1995). Ste12p, a homeodomain transcription factor phosphorylated following activation of the pheromone-responsive MAP kinase cascade, binds to pheromone response elements (PRE) mediating increased transcription of genes in the presence of pheromone (KIRKMAN-CORREIA *et al.* 1993). Two additional genes (A003 and A006) encoded putative DNA-binding protein, which may also function in the regulation of a signal transduction pathway.

PCCA002 corresponded to *ste3*, the putative serpentine pheromone G-coupled receptor. On the basis of the cDNA and EST sequence analysis, the open reading frame appeared to have two alternative 3′ mRNA structures resulting from alternative splicing (Figure 1B). In both transcripts, the ORF initiated with a translational start at position 1359 and spliced out a 46-bp intron between positions 1888 and 1933. A shorter ORF encoding a predicted 301-amino-acid residue peptide resulted from use of a stop codon located at position 2415 within the cosmid sequence. This splice alternative matched the ESTs, s20D1 and s14G9, and a recently released mRNA sequence determined by Thomas and co-workers (GenBank accession no. Aa007236). The alternative splicing product present in the λZAP cDNA clone 2G2, identified by hybridization screening, resulted from the excision of a second intron between positions 2412 and 2462, extending the open reading frame to a stop codon located at position 2775 and resulting in a 439-amino-acid polypeptide (Figure 1, A and C).

The intergenic region between PCCA001 and *ste3* is predicted by MatInspector, a predictor of transcriptional regulation sites, to contain recognition sites for the *S. cerevisiae* transcriptional regulators Mcm1p and a1, involved cooperatively in mating-type-specific gene regulation (DOLAN and FIELDS 1991). Flanking the putative *ste3* gene is a potential short ORF on the Crick strand encoding a peptide of 35 amino acids ending in the prenylation motif -CAAX common to all **a**-factor-like pheromones (SHEN *et al.* 1999). As *ste3* encodes an **a**-factor receptor, the expected pheromone to be expressed in conjunction with this would be an α-factor pheromone; thus the significance of this potential ORF is unknown.

The putative gene, PCCA004, contained an ORF with homology to the homeodomain region of the transcription factor Ste12p of a variety of ascomycetous and basidiomycetous fungi, with greatest homology to the Ste12α protein of *C. neoformans* (WICKES *et al.* 1997). The ORF could not be extended by the presence of introns either upstream or downstream of the homeodomain region. The accuracy of the sequence was verified by comparison within the putative Ste12 region of the sequence in two overlapping, independent cosmids, W10H12 and W19D7. The partial ORF A004 may represent a pseudogene, with a functional *ste12* gene located elsewhere within the genome. Hybridization of the *ste12* probe (Figure 3, P3) to a Southern blot of *P. carinii* chromosomes separated by CHEF pulsed-field gel electrophoresis and a Southern blot of *P. carinii* genomic DNA digested with multiple restriction enzymes provided no evidence of a second *ste12* gene within the genome (data not shown).

To determine if the *ste12* ortholog, A004, was transcribed, Northern analysis was performed and λ phage containing A004 cDNA were sought in the λZAP cDNA library. Three prime anchor PCR using the λcDNA library as template confirmed the presence of cDNAs terminating 90 bp downstream of the stop codon identified immediately following the third helix of the homoedomain motif of the putatitive Ste12 ORF (Kirkman-Correia *et al.* 1993). Northern analysis using a probe generated to the homeodomain encoding region of A004 revealed hybridization with a 4-kb RNA transcript (data not shown). An adjacent gene, PCCA005, encoded on the same strand as A004 encodes an ortholog of the *S. cerevisiae* ORF YDR124w (Table 1). Northern analysis using a PCCA005 probe (Figure 3, P4) identified a 4-kb transcript, identical in size to the RNA species identified by a probe to the adjacent gene PCCA004, suggesting that these constitute a single gene or a polycistronic mRNA transcript. Yeast Ste12p contains multiple functional domains including a DNA-binding domain, transcriptional activation domains resulting in basal and inducible transcription induction, and a domain required for cooperation with the transcriptional regulator Mcm1p (Kirkman-Correia *et al.* 1993). Structural variation has been described among other fungal Ste12p homologs. Although all Ste12p homologs contain a homeodomain, the *C. neoformans* Ste12p also contains a zinc-finger DNA-binding domain (Wickes *et al.* 1997). Full characterization of gene products arising from A004 and A005 will require additional experimentation.

PCCA008 encoded the *P. carinii* ortholog of the protein kinase Ste20, an element of the pheromone-responsive kinase cascade interacting with Cdc42 and Ste11. The Ste20 ORF is interrupted by seven introns ranging in size from 48 to 235 bp in length. The predicted *P. carinii* Ste20 protein demonstrates high homology to both the kinase domain and the Cdc42-binding domains of multiple Ste20-like proteins (Marcus *et al.* 1995; Leberer *et al.* 1996).

Transcriptional regulators are important components of all mating-type loci. Two genes, PCCA003 and PCCA006, encoded two putative DNA-binding proteins that may play a role in the regulation of a pheromone response pathway. These genes encoded orthologs of the zinc-finger single-stranded DNA-binding protein Rim1p of *S. cerevisiae* and the helix-loop-helix DNA-binding protein YAWC of *S. pombe*, respectively.

W15A6 contains genes apparently unrelated to signal transduction: In addition to the genes described above encoding proteins potentially active within a pheromone response pathway, nine additional genes were identified with W15A6. These gene products have no apparent relationship to pheromone response pathways. However, examination of pheromone-responsive genes in *S. cerevisiae* has identified many classes of genes previously not known to be mating related. In addition, unrelated genes, such as a myosin gene described within the *MAT*α locus of *C. neoformans*, may be found within mating loci (Karos *et al.* 2000).

Gene PCCA001 encoded a polypeptide with significant amino acid similarity to glutathione synthetase from a variety of microorganisms, but with greatest homology to the glutathione synthetase of *S. pombe*. ORF PCCA001, encoded on the Watson strand, extended from the left border of W15A6 encoding the carboxy-terminal 200-amino-acid residues of glutathione synthetase. An overlapping cosmid W12G3 extended the ORF an additional 280 bp upstream.

Additional genes in the cosmid W15A6 with significant homology to genes of known function encoded a putative sucrose transporter (PCCA009), a Na/H^+ exchanger (PCCA010), the MTS2p subunit of the 26S proteasome complex (PCCA011), and a glycyl tRNA synthetase (PCCA014). Four genes, PCCA006, PCCA012, PCCA013, and PCCA015, were homologous to hypothetical open reading frames in *S. pombe* with lesser homology to related genes in *S. cerevisiae*, encoding proteins of unknown function. Transcription of each identified gene sequence was confirmed by Northern blot analysis, identification of corresponding ESTs, or cDNAs.

Location of the signal transduction cluster within the genome: Clustering of genes encoding mating-related signal transduction elements including a pheromone receptor, pheromones, Ste11, Ste12, Ste20, DNA-binding proteins, and transcription factors was previously reported within the mating locus of the heterothallic basidiomycete *C. neoformans* (Moore and Edman 1993; Karos *et al.* 2000; J. Heitman, personal communication). The clustering of signal transduction genes including the pheromone receptor identified within *P. carinii* may represent either a mating locus, a silent donor site for genes to a mating locus, or a clustering of signal transduction genes. To help differentiate these possibilities, alternative genes or alleles (idiomorphs) were sought that may occupy the same genomic locale in other *P. carinii* organisms and the copy number of the *ste3* gene and other genes within or flanking a potential mating locus or silent donor site was determined.

Hybridization screening of the arrayed *P. carinii* cosmid library with a *ste3* probe identified five cosmids. When the library was rescreened using a probe to the glutathione synthase gene (Figure 3, P1), the same five cosmids were identified (Figure 3, W19D7, W18D8). A chromosomal walk in the opposite direction using an A010 probe (Figure 3, P6) identified eight cosmids that hybridized with the A010 probe but not with an A001 or *ste3* probe (Figure 3, W13D10, W13B12, W24C10). Screening with an A011 probe (Figure 3, P7) identified one cosmid that hybridized to the A011 probe but did not hybridize with either the A010 or *ste3* probes (Figure 3, W18G1). End sequence analysis of cosmid W12G3 extended the sequence contig border 280 bp upstream of the left border of W15A6 into the glutathione synthe-

tase gene. Cosmids overlapping W15A6 to the right failed to identify any alternative gene sequence linked to the genes A010 or A011. All cosmids extending to the right had the same gene organization as W15A6 by both end sequence analysis (Figure 3) and by PCR amplification using primers specific for each predicted gene within W15A6 (data not shown).

Southern blot analysis was performed on *P. carinii* genomic DNA separated by pulsed-field gel electrophoresis to determine the chromosomal location of *ste3* and other genes within W15A6 across different *P. carinii* populations with class I and class II divergence (Cushion 1998b). Blots were hybridized with probes complementary to genes A001, A002, and A011 (Figure 4). All probes from W15A6 hybridized only with the 686-kb chromosome 1 of *P. carinii* f. sp. *carinii*. The *P. carinii* chromosomal blot contained samples of nine forms of *P. carinii* f. sp. *carinii* as well as *P. carinii* f. sp. *ratti*. The first chromosome in all of the karyotype forms carried *ste3*, *gsh*, and *mts2* with equivalent signal intensity, suggesting that similar linkage between *ste3* and adjacent genes is found in these other *P. carinii* f. sp. *carinii* forms. The genes appeared linked in *P. carinii* f. sp. *ratti* as in *P. carinii* f. sp. *carinii* but were localized to a 345-kb chromosome, the 13th chromosome in this organism.

In an attempt to determine the gene copy number of some of the elements with W15A6, densitometric analysis of the blots probed with *gsh1*, *ste3*, and *mts2* was performed. In a mixed mating-type population of a heterothallic or pseudohomothallic organism, the gene copy number of *ste3* within the population would be lower than that of a gene not within the mating locus. The "difference" in gene copy number would represent those organisms where the α-factor receptor ortholog, *ste2*, lay within the mating locus. Densitometric analysis demonstrated a 1:1 ratio of the signal intensity when probed with *ste3* (Figure 3, P2) compared to the same blots probed with each of the other probes, demonstrating that the copy number of ste3, A001, and A011 is the same. Thus, these populations did not contain organisms where A001 and A011 were present but *ste3* was absent, or such organisms were present as a small minority of the population at a level where such changes could not be detected.

To further analyze gene copy number and gene linkage, genomic *P. carinii* f. sp. *carinii* form 1 DNA that was isolated from six heavily infected animals was restriction-enzyme digested and subjected to Southern analysis using probes to A001, A002, A004, A009, A011, and A013. The hybridization patterns were identical among all six *P. carinii* f. sp. *carinii* isolates (data not shown). The hybridization patterns with probes complementary to A001, A002, A004, and A009 correlated with the predicted restriction sites within the W15A6 sequence and was consistent with a single copy of these genes. Under high stringency conditions, an additional 5-kb restriction fragment hybridized with a probe complementary

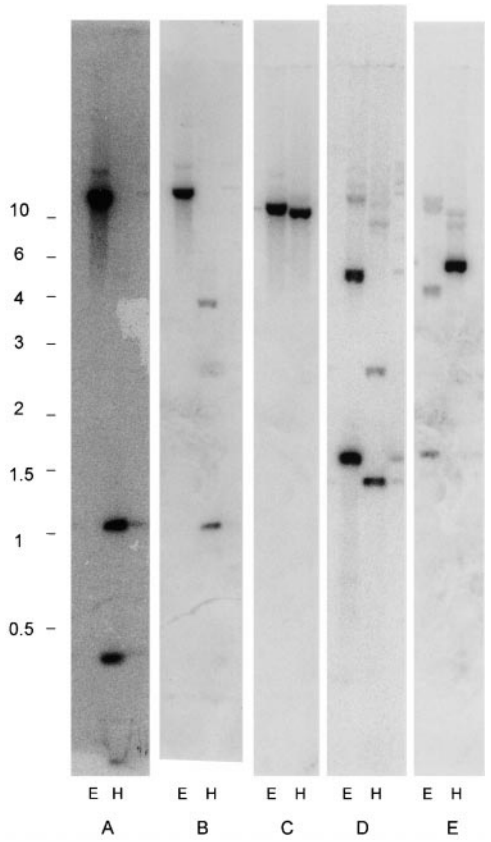

Figure 5.—Southern hybridization of genes within W15A6. Southern hybridization of *P. carinii* f. sp. *carinii* genomic DNA digested with *Eco*RI (lanes E) or *Hin*dIII (lanes H) was performed using hybrization probes specific to genes within cosmid W15A6. Following hybridization, blots were washed under conditions of high stringency (0.2× SSC, 0.2% SDS at 60° for 30 min). Hybridization to *P. carinii* DNA by *gsh1* (P1 in Figure 3) is shown in A, *ste3* probe (P2) in B, *sca* probe (P6) in C, *mts2* probe (P7) in D, and A013 probe (P8) in E. DNA size markers are shown on the left.

to A011, suggesting the presence of restriction polymorphism in the DNA flanking the probe or the presence of an Mts2 paralogue (Figure 5). Restriction fragment polymorphism was also noted with an A013 probe. The presence of a second *mts2* gene within the genome was not supported by the hybridization pattern seen on the *P. carinii* chromosomal blot, where hybridization occurs only to chromosome 1, unless the second gene is also located on the same chromosome. In addition, all 14 cosmids identified in the *P. carinii* genomic library had genomic organization consistent with that detected in W15A6.

DISCUSSION

The Pneumocystis genome sequence project was launched as a community-wide effort to rapidly gain information of biological significance to aid in the understanding of this intractable organism. The first data available from the project included 3896 *P. carinii* ESTs,

reflecting a partial inventory of genes expressed in organisms isolated from a heavily infected mammalian host. These data have already been used to show correlation between the *in vitro* sensitivity to inhibitors of sterol biosynthesis and the presence of the targeted enzymes within the *P. carinii* EST database (Kaneshiro *et al.* 2000).

The sequence data presented here represent the first examination of a large contiguous section of the *P. carinii* f. sp. *carinii* genomic sequence and provide a framework for expectation of the completed *P. carinii* f. sp. *carinii* genome sequence. The full complement of genes in *P. carinii* f. sp. *carinii* would be predicted to be ~3740 genes if one considers the genome size of 8 Mb and the gene density of W15A6. This is slightly lower than previous estimates of 4000 genes made prior to the availability of any gene density analysis (Stringer and Cushion 1998). However, a unique structural aspect of the *P. carinii* f. sp. *carinii* genome (and likely other genomes in this family of organisms) is the extent of repetitive gene families and repetitive DNA sequences at the end of each chromosome. An 11-kb telomeric λ clone has previously been sequenced and contained the telomeric repeats, subtelomeric repeated sequences, and members of the major surface glycoprotein (MSG) gene family, MSG-related (MSR) gene family, and the protease (PRT) gene family but did not extend into unique single copy gene sequences (Wada and Nakamura 1994). Similar arrays occur on each chromosome although the topology of each telomere end remains to be resolved. With a conservative estimate that 11 kb is devoted to MSG/MSR/PRT arrays at each end of the 15 chromosomes, the remaining genome would contain an estimated 3530 genes, only ~65% of the number present in *S. cerevisiae*. Within cosmid W15A6, 23,088 of the 32,083 bp or 71.9% of the sequence is devoted to coding sequence within the 14.5 gene sequences, reflecting the economic use of its small genome. The remaining 28.1% comprises intergenic regions with an average size of 498 bp, and introns with an average of 2.1 introns/kb of coding sequence and an average size of 57 bp. Thus, if W15A6 is reflective of the entire *P. carinii* genome, this would predict 5.71 Mb of coding sequence within the genome and an average gene size of 1620 bp.

The EST database provided partial sequence of a serpentine receptor. Analysis of the complete gene sequence confirmed the sequence of a G-coupled receptor with greatest homology to the **a**-factor receptor of *S. cerevisiae*, Ste3p. Although a multitude of G-coupled receptors are expressed in mammalian cells, only two classes of G-coupled receptors have been described in yeast: the pheromone receptors and those involved in nutritional sensing such as Grp1p (Lorenz *et al.* 2000). Although no functional data demonstrating binding of a pheromone to the putative *P. carinii* receptor nor cellular response to pheromone have been demonstrated, the homology suggests the predicted protein is a pheromone receptor comprising an extracellular domain, seven membrane-spanning α-helical domains, and an intracellular cytoplasmic tail. Sequence analysis of cDNAs revealed two splice alternatives, resulting in proteins with varied length intracytoplasmic tails. Although cDNAs were found that confirmed the presence of both transcripts, no evidence has yet been obtained to determine if both proteins are expressed or the conditions under which alternative splicing occurs. The use of alternative splicing may be an efficient mechanism of the organism to fully exploit its small genome.

Clustering of genes of related function is well described among fungi. Clustering of genes is best described among components of secondary metabolism, such as polyketide synthesis, where large arrays of genes involved in polyketide regulation, synthesis, and processing are frequently clustered together (Keller and Hohn 1997). As compared to operon organization in prokaryotes, eukaryotic clusters usually result in regulated but independent transcription of each gene within the cluster. Clustering has also been reported in nonessential metabolic pathways (Lawrence and Roth 1996). A specialized form of gene clustering is seen within the mating loci. Organization of mating loci may comprise a single-locus two-idiomorph system, containing either solely transcriptional regulator genes or multiple genes within the pheromone response pathway (Kronstad and Staben 1997). Alternatively, complex mating loci have been described in filamentous basidiomycetes, such as *Schizophyllum* spp. and *Coprinus* spp., where the Aα, Aβ, Bα, and Bβ mating loci contain both transcriptional regulators, pheromone receptors, and multiple pheromones organized within more complex A and B complexes (Vaillancourt *et al.* 1997; Casselton and Olesnicky 1998; O'Shea *et al.* 1998; Fowler *et al.* 1999). Maintenance of clusters has advantage for expression or silencing of the entire cluster, regulation of ordered gene expression, and genetic mobility of an entire cluster rather than individual elements within the cluster.

The organization of genes surrounding the putative pheromone receptor of *P. carinii* f. sp. *carinii* fulfills many of the characteristics of a mating locus, namely, the clustering of genes encoding (1) a pheromone receptor (Ste3), (2) a potential pheromone encoding-gene, (3) transcriptional regulators (such as Ste12) that may be involved in regulation of a pheromone response pathway, and (4) members of the pheromone responsive signal transduction cascade. Alternatively, this gene cluster may represent a clustering of signal transduction genes with the true mating locus containing the genes regulating expression of elements within this cluster. Immediately flanking the pheromone receptor is a short open reading frame encoding a possible **a**-mating factor-like polypeptide. The lipopeptide **a**-factor pheromones from different fungi bear little sequence homology, save

for the CAAX prenylation motif at the carboxyl terminus of the predicted peptide (Shen *et al.* 1999). Organisms of the **a**-mating type express **a**-factor pheromone and express α-factor receptor on their surface. The linkage, therefore, of an **a**-factor pheromone receptor and an **a**-factor in the same regulon is unprecedented, suggesting that this short ORF is a random finding, these genes may be independently regulated, or these organisms bear a highly unusual mating structure. On the basis of conservation of signal transduction cascades among other fungi, other adjacent genes that may be involved in a pheromone response pathway include Ste20 and Ste12 orthologs. Two DNA-binding proteins encoded by A003 and A005 may play a role in regulation of the pheromone response. A *P. carinii* Ste11/Byr2 ortholog, found within the *C. neoformans* mating locus, is not located within the Ste3 cluster but rather is located on chromosome 7 of *P. carinii*, while a Ste7/Mkk1 ortholog is located on chromosome 1, but its exact location relative to the Ste3 gene has not been determined.

Additional evidence for this cluster functioning as a mating locus was sought by trying to identify an alternative mating locus. In a pseudohomothallic organism, a second copy of the locus would be found as either the silent copy or at the site of mating-type expression in organisms expressing the Ste3 gene. Restriction fragment polymorphisms were identified with probes at the right end of the gene array (A011 and A013) but no other evidence was found for a second copy of any of the genes within the cluster or for an alternative genomic context of the cluster. In addition, silencing of an entire array of genes, such as this cluster, would be more difficult to achieve than the silencing of a single or a few transcription factors located within the mating loci of homothallic organisms, such as those found in *S. cerevisiae* and *S. pombe*. In a heterothallic organism, an alternative mating locus would be found within the same genomic context in other *P. carinii* organisms. Although it may be postulated that *P. carinii* f. sp. *carinii* and *P. carinii* f. sp. *ratti* may represent populations of opposite mating types, opposite mating types are genetically identical with the exception of the mating idiomorphs, and *P. carinii* f. sp. *carinii* and *P. carinii* f. sp. *ratti* have been shown to have significant differences in gene organization and sequence at multiple loci within the genome (Cushion *et al.* 1993). Populations of *P. carinii* isolated from infected animals are not clonal and may represent one or both mating types, and it is possible that both exist within the same lung. Although the restriction polymorphism noted with an A011 probe hybridized to genomic DNA isolates could reflect hybridization of this probe to an *mts*2a and *mts*2α allele, no other convincing evidence was found to demonstrate an alternative locus within this genomic context. Screening of the cosmid library, screening of multiple populations of *P. carinii* f. sp. *carinii* form 1 organisms, or screening within different forms (2–9) of *P. carinii* f. sp. *carinii* organisms did not produce corroborating evidence of the existence of another mating locus. These findings may be erroneously interpreted if the structure and organization of the mating types vary significantly (Wickes *et al.* 1997), insufficient homology exists between **a**- and α- alleles of genes within the loci, or if the marker used to detect the borders of the locus still lies within the locus. At this junction, the negative findings suggest there is no alternative cluster/locus, or there is marked mating-type disequilibrium seen within the populations examined. This effect has been observed in *C. neoformans*, where mating-type **a** is vastly underrepresented in clinical isolates (Kwon-Chung *et al.* 1992).

P. carinii has much of the genetic machinery necessary for sexual replication. Evidence of its expression as well as the expression of meiosis-specific genes during mammalian infection suggests that the organism does undergo a sexual cycle of replication. However, mating-type loci play roles in processes other than mating. These include sexual dimorphism, vegetative incompatibility, and virulence. Genes within the sexual replication pathway may have been retained from an ancestral organism, and expression of these genes in *P. carinii* may not be related to sexual replication. The exact role for this pathway in sexual replication will need to be experimentally determined. Genomic sequence analysis revealed the presence of mating-type genes and machinery in *Candida albicans*, long thought to be parasexual, and subsequent studies have demonstrated the ability of *C. albicans* to mate under laboratory conditions (Hull and Johnson 1999; Hull *et al.* 2000; Magee and Magee 2000). Similar studies in *P. carinii* will help define the life cycle of the organism.

We are grateful to Drs. Michael Linke and James Stringer for helpful discussion and thoughtful review. This work was supported by the Office of Research and Development, Medical Research Service, Department of Veterans Affairs and the Burroughs Wellcome Fund. George Smulian is a recipient of a New Investigator Award in Pathogenic Mycology from the Burroughs Wellcome Fund.

LITERATURE CITED

Casselton, L. A., and N. S. Olesnicky, 1998 Molecular genetics of mating recognition in basidiomycete fungi. Microbiol. Mol. Biol. Rev. **62:** 55–70.

Cushion, M. T., 1998a Taxonomy, genetic organization, and life cycle of Pneumocystis carinii. Semin. Respir. Infect. **13:** 304–312.

Cushion, M. T., 1998b Genetic heterogeneity of rat-derived Pneumocystis. FEMS Immunol. Med. Microbiol. **22:** 51–58.

Cushion, M. T., and J. Arnold, 1997 Proposal for a Pneumocystis genome project. J. Eukaryot. Microbiol. **44:** 7S.

Cushion, M. T., J. Zhang, M. Kaselis, D. Giuntoli, S. L. Stringer *et al.*, 1993 Evidence for two genetic variants of Pneumocystis carinii coinfecting laboratory rats. J. Clin. Microbiol. **31:** 1217–1223.

Dolan, J. W., and S. Fields, 1991 Cell-type-specific transcription in yeast. Biochim. Biophys. Acta **1088:** 155–169.

Edman, J. C., J. A. Kovacs, H. Masur, D. V. Santi, H. J. Elwood *et al.*, 1988 Ribosomal RNA sequence shows Pneumocystis carinii to be a member of the fungi. Nature **334:** 519–522.

Fowler, T. J., S. M. DeSimone, M. F. Mitton, J. Kurjan and C. A. Raper, 1999 Multiple sex pheromones and receptors of a mush-

room-producing fungus elicit mating in yeast. Mol. Biol. Cell **10:** 2559–2572.

Haber, J. E., 1998 Mating-type gene switching in Saccharomyces cerevisiae. Annu. Rev. Genet. **32:** 561–599.

Hong, S. T., P. E. Steele, M. T. Cushion, P. D. Walzer, S. L. Stringer *et al.*, 1990 Pneumocystis carinii karyotypes. J. Clin. Microbiol. **28:** 1785–1795.

Hull, C. M., and A. D. Johnson, 1999 Identification of a mating type-like locus in the asexual pathogenic yeast Candida albicans. Science **285:** 1271–1275.

Hull, C. M., R. M. Raisner and A. D. Johnson, 2000 Evidence for mating of the "asexual" yeast Candida albicans in a mammalian host. Science **289:** 307–310.

Kaneshiro, E. S., M. S. Collins and M. T. Cushion, 2000 Inhibitors of sterol biosynthesis and amphotericin B reduce the viability of pneumocystis carinii f. sp. carinii. Antimicrob. Agents Chemother. **44:** 1630–1638.

Karos, M., Y. C. Chang, C. M. McClelland, D. L. Clarke, J. Fu *et al.*, 2000 Mapping of the cryptococcus neoformans MATalpha locus: presence of mating type-specific mitogen-activated protein kinase cascade homologs. J. Bacteriol. **182:** 6222–6227.

Kelkar, H. S., J. Griffith, M. E. Case, S. F. Covert, D. A. Hall *et al.*, 2001 The *Neurospora crassa* genome: cosmid libraries sorted by chromosome. Genetics **157:** 979–990.

Keller, N. P., and T. M. Hohn, 1997 Metabolic pathway gene clusters in filamentous fungi. Fungal Genet. Biol. **21:** 17–29.

Kirkman-Correia, C., I. L. Stroke and S. Fields, 1993 Functional domains of the yeast STE12 protein, a pheromone-responsive transcriptional activator. Mol. Cell. Biol. **13:** 3765–3772.

Kronstad, J. W., and C. Staben, 1997 Mating type in filamentous fungi. Annu. Rev. Genet. **31:** 245–276.

Kwon-Chung, K. J., J. C. Edman and B. L. Wickes, 1992 Genetic association of mating types and virulence in Cryptococcus neoformans. Infect. Immun. **60:** 602–605.

Lawrence, J. G., and J. R. Roth, 1996 Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics **143:** 1843–1860.

Leberer, E., D. Dignard, D. Harcus, D. Y. Thomas and M. Whiteway, 1992 The protein kinase homologue Ste20p is required to link the yeast pheromone response G-protein beta gamma subunits to downstream signalling components. EMBO J. **11:** 4815–4824.

Leberer, E., D. Harcus, I. D. Broadbent, K. L. Clark, D. Dignard *et al.*, 1996 Signal transduction through homologs of the Ste20p and Ste7p protein kinases can trigger hyphal formation in the pathogenic fungus Candida albicans. Proc. Natl. Acad. Sci. USA **93:** 13217–13222.

Linke, M. J., M. T. Cushion and P. D. Walzer, 1989 Properties of the major antigens of rat and human Pneumocystis carinii. Infect. Immun. **57:** 1547–1555.

Lorenz, M. C., X. Pan, T. Harashima, M. E. Cardenas, Y. Xue *et al.*, 2000 The G protein-coupled receptor gpr1 is a nutrient sensor that regulates pseudohyphal differentiation in Saccharomyces cerevisiae. Genetics **154:** 609–622.

Magee, B. B., and P. T. Magee, 2000 Induction of mating in Candida albicans by construction of MTLa and MTLalpha strains. Science **289:** 310–313.

Marcus, S., A. Polverino, E. Chang, D. Robbins, M. H. Cobb *et al.*, 1995 Shk1, a homolog of the Saccharomyces cerevisiae Ste20 and mammalian p65PAK protein kinases, is a component of a Ras/Cdc42 signaling module in the fission yeast Schizosaccharomyces pombe. Proc. Natl. Acad. Sci. USA **92:** 6180–6184.

Matsumoto, Y., and Y. Yoshida, 1984 Sporogony in Pneumocystis carinii: synaptonemal complexes and meiotic nuclear divisions observed in precysts. J. Protozool. **31:** 420–428.

Moore, T. D., and J. C. Edman, 1993 The alpha-mating type locus of Cryptococcus neoformans contains a peptide pheromone gene. Mol. Cell. Biol. **13:** 1962–1970.

O'Shea, S. F., P. T. Chaure, J. R. Halsall, N. S. Olesnicky, A. Leibbrandt *et al.*, 1998 A large pheromone and receptor gene complex determines multiple B mating type specificities in Coprinus cinereus. Genetics **148:** 1081–1090.

Shen, W. C., P. Bobrowicz and D. J. Ebbole, 1999 Isolation of pheromone precursor genes of Magnaporthe grisea. Fungal Genet. Biol. **27:** 253–263.

Stringer, J. R., 1996 Pneumocystis carinii: What is it, exactly? Clin. Microbiol. Rev. **9:** 489–498.

Stringer, J. R., and M. T. Cushion, 1998 The genome of Pneumocystis carinii. FEMS Immunol. Med. Microbiol. **22:** 15–26.

Stringer, S. L., J. R. Stringer, M. A. Blase, P. D. Walzer and M. T. Cushion, 1989 Pneumocystis carinii: sequence from ribosomal RNA implies a close relationship with fungi. Exp. Parasitol. **68:** 450–461.

Vaillancourt, L. J., M. Raudaskoski, C. A. Specht and C. A. Raper, 1997 Multiple genes encoding pheromones and a pheromone receptor define the B beta 1 mating-type specificity in *Schizophyllum commune*. Genetics **146:** 541–551.

Wada, M., and Y. Nakamura, 1994 MSG gene cluster encoding major cell surface glycoproteins of rat Pneumocystis carinii. DNA Res. **1:** 163–168.

Wickes, B. L., U. Edman and J. C. Edman, 1997 The Cryptococcus neoformans STE12alpha gene: a putative Saccharomyces cerevisiae STE12 homologue that is mating type specific. Mol. Microbiol. **26:** 951–960.

Communicating editor: J. Arnold