J. Eukaryot. Microbiol., 53(6), 2006 pp. 515–521

DOI: 10.1111/j.1550-7408.2006.00135.x

Novel Nucleomorph Genome Architecture in the Cryptomonad Genus *Hemiselmis*

CHRISTOPHER E. LANE and JOHN M. ARCHIBALD

Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS B3H 1X5, Canada

ABSTRACT. Cryptomonads are ubiquitous aquatic unicellular eukaryotes that acquired photosynthesis through the uptake and retention of a red algal endosymbiont. The nuclear genome of the red alga persists in a highly reduced form termed a nucleomorph. The nucleomorph genome of the model cryptomonad *Guillardia theta* has been completely sequenced and is a mere 551 kilobases (kb) in size, spread over three chromosomes. The presence of three chromosomes appears to be a universal characteristic of nucleomorph genomes in cryptomonad algae as well as in the chlorarachniophytes, an unrelated algal lineage with a nucleomorph and plastid genome derived from a green algal endosymbiont. Another feature of nucleomorph genomes in all cryptomonads and chlorarachniophytes examined thus far is the presence of subtelomeric ribosomal DNA (rDNA) repeats at the ends of each chromosome. Here we describe the first exception to this canonical nucleomorph genome architecture in the cryptomonad *Hemiselmis rufescens* CCMP644. Using pulsed-field gel electrophoresis (PFGE), we estimate the size of the *H. rufescens* nucleomorph genome to be $\sim$ 580 kb, slightly larger than the *G. theta* genome. Unlike the situation in *G. theta* and all other known cryptomonads, sub-telomeric repeats of the rDNA cistron appear to be absent on both ends of the second largest chromosome in *H. rufescens* and two other members of this genus. Southern hybridizations using a variety of nucleomorph protein gene probes against PFGE-separated *H. rufescens* chromosomes indicate that recombination has been a major factor in shaping the karyotype and genomic structure of cryptomonad nucleomorphs.

Key Words. Cryptophyceae, genome evolution, genome rearrangement, genome reduction, secondary endosymbiosis.

THE process of endosymbiosis has had a profound impact on the evolution of photosynthetic life on Earth. The primary endosymbiotic origin of plastids (chloroplasts), whereby a cyanobacterium became a permanent resident inside a previously heterotrophic eukaryote, is generally believed to have happened only once, at least a billion years ago (Yoon et al. 2004), although a second primary endosymbiosis appears to be in its early stages in the cercozoan protist *Paulinella chromatophora* (Marin, Nowack, and Melkonian 2005). Three modern-day eukaryotic lineages, the glaucophytes, green algae (including land plants), and red algae, harbor plastids whose ancestry can be traced directly back to the primary endosymbiosis (Bhattacharya, Yoon, and Hackett 2004; Keeling 2004). All other photosynthetic eukaryotes contain plastids with a more complicated history, in which red or green algae—and their plastids—were consumed by other eukaryotes, but not digested. This process, referred to as secondary endosymbiosis, has occurred at least three times throughout evolution and has given rise to the bulk of the eukaryotic phototrophs in the world's oceans, including the dinoflagellates, haptophytes, and heterokonts (Archibald and Keeling 2005; Delwiche 1999; Keeling 2004).

The cryptomonads and chlorarachniophytes are two lineages of secondary plastid-containing algae that offer a unique window into the process of secondary endosymbiosis because they still contain the relict nucleus of their eukaryotic endosymbionts in a miniaturized form, termed a nucleomorph. The nucleomorph genomes of the cryptomonad, *Guillardia theta* (Douglas et al. 2001) and of the chlorarachniophyte, *Bigelowiella natans* (Gilson et al. 2006) have been entirely sequenced. The two genomes are similar in size, 551 and 373 kb, respectively, and in their basic architecture. Both genomes are partitioned amongst three linear chromosomes and are extremely gene dense, with little in the way of repetitive sequence. The similarities in the size and structure of the *G. theta* and *B. natans* nucleomorph genomes are particularly striking when one considers that they are the product of independent secondary endosymbioses. While the cryptomonads harbor a plastid derived from a red algal endosymbiont (Van der Auwera et al. 1998), the chlorarachniophyte plastid has a green algal origin (McFadden et al. 1994).

The karyotypic diversity of nucleomorphs within the chlorarachniophytes and cryptomonads has been investigated using pulsed-field gel electrophoresis (PFGE). Maier and colleagues (Rensing et al. 1994; Zauner et al. 2000), McFadden and colleagues (Gilson and McFadden 1999; McFadden et al. 1994), and most recently, Lane et al. (2006) have shown that the presence of three chromosomes appears to be a universal feature of these genomes, although total nucleomorph genome size varies considerably, between $\sim$ 450 and $\sim$ 845 kb in cryptomonads (Lane et al. 2006; Rensing et al. 1994) and $\sim$ 380 to $\sim$ 455 kb in chlorarachniophytes (Gilson and McFadden 1999). Genomic data indicate that another common feature of chlorarachniophyte and cryptomonad nucleomorph genomes is the presence of ribosomal DNA (rDNA) cistrons immediately adjacent to the telomeric repeats on each of the six chromosome ends (Douglas et al. 2001; Gilson et al. 2006). Southern hybridization experiments have shown that rDNA repeats are present in a wide range of cryptomonad and chlorarachniophyte algae (Gilson and McFadden 1999; Lane et al. 2006; Rensing et al. 1994), and the assumption that the rDNA is sub-telomeric has been confirmed where investigated (Douglas et al. 2001; Gilson and McFadden 2002; Gilson et al. 2006; Lane et al. 2006; Zauner et al. 2000). The functional significance of this arrangement is currently unknown.

We have discovered the first known deviation from the canonical nucleomorph genome architecture in the cryptomonad *Hemiselmis rufescens* CCMP644. We used Hoechst dye-cesium chloride (CsCl) density gradient centrifugation to isolate DNA fractions from each of its four genome-containing compartments and show that, in contrast to the situation in the model cryptomonad *G. theta* (Douglas 1988), the nucleomorph genome of *H. rufescens* has a higher A/T content than its mitochondrial genome. Southern hybridizations against PFGE-separated nucleomorph chromosomes using nucleomorph rDNA probes reveal that while *H. rufescens* and its closest relatives possess three nucleomorph chromosomes, only the 5S gene of the rDNA cistron remains on their second largest chromosome. The results of hybridizations using probes for several nucleomorph protein-coding genes also suggest a significant amount of inter-chromosomal rearrangement relative to the nucleomorph genome of *G. theta* (Douglas et al. 2001). These data are the first evidence supporting inter-chromosomal rearrangements as a major force shaping nucleomorph genome size and structure in cryptomonads.

MATERIALS AND METHODS

DNA extraction through nucleomorph 18S rDNA sequencing. Algal cultures were obtained from public collections (Table 1) and maintained under designated conditions. Routine DNA

Corresponding Author: C. Lane, Department of Biochemistry and Molecular Biology, 5850 College St., Dalhousie University, Halifax, N S, Canada B3H 1X5—Telephone number: 902-494-2536; FAX number: 902-494-1355; e-mail: c.lane@dal.ca

Table **1.** Taxa and GenBank accession numbers used for phylogenetic analysis.

Taxon name	Culture strain	GenBank Accession Number
Chroomonas mesostigmatica R. Butcher ex D.R.A. Hill*	CCMP1168	DQ228123
Chroomonas pauciplastida nom. prov. R. Butcher*	CCMP268	**DQ519363**
Chroomonas sp.*	CCMP270	DQ228124
Chroomonas sp.	M1312	AJ420678
Chroomonas sp.	M1318	AJ420679
Chroomonas sp.	M1418	AJ420680
Chroomonas sp.	SAG B980-1	AJ420677
Hemiselmis rufescens Parke*	CCMP439	AJ420690
*Hemiselmis rufescens**	CCMP644	**DQ519365**
Hemiselmis sp.*	RCC659	**DQ519364**
Hemiselmis virescens Droop	CCMP443	AJ420691
Unidentified Cryptomonad	Uncultured	U53191

Numbers in bold were obtained in this study. Culture collection abbreviations are as follows: CCMP, Provasoli-Guillard National Center for Culture of Marine Phytoplankton, USA; M, Dr. Michael Melkonian, Germany; RCC, Roscoff Culture Collection, France; SAG, Sammlung von Algenkulturen Göttingen, Germany.

*Cultures maintained in the Archibald Laboratory.

extractions, used for PCR and sequencing, were performed on 100 ml of dense log-phase culture, whereas DNA was extracted from 8 l of culture for density gradient centrifugation experiments. DNA extraction was performed and PCR was used to amplify the nucleomorph 18S rDNA locus for phylogenetic reconstruction, as previously described by Lane et al. (2006). PCR products were separated on an agarose gel, excised, and purified with the MinElute Gel Extraction Kit (QIAGEN Sciences, MD, USA) according to the manufacturer's instructions. Purified amplicons were cloned using the Topo-TA Cloning Kit (Invitrogen, Carlsbad, CA, USA). Three to five independent clones were sequenced using the PCR primers and internal primers previously described by Lane et al. (2006). Sequencing was performed on a Beckman Coulter CEQ8000 (Beckman Coulter Inc., Fullerton, CA, USA) using the CEQ Dye Terminator Cycle Sequencing Kit. Sequences produced during this study have been submitted to GenBank under the accession numbers DQ519363—DQ519365.

Pulsed-field gel electrophoresis (PFGE). Agarose plugs for PFGE were prepared as previously described by Eschbach et al. (1991) and Lane et al. (2006) using 2 l of log-phase culture as starting material. A 1% agarose gel (1 × TBE) was run in 0.5% (v/v) TBE buffer at 14.0 °C in a CHEF-DR III Pulsed-Field Electrophoresis System (Bio-Rad Laboratories, Hercules, CA, USA). Run time was 60 h at a voltage of 4.1 V/cm with a 30–10 s switch time.

Hoechst dye-CsCl centrifugation and fraction purification. DNA was diluted with TE buffer to a total vol. of 10.5 ml and 11 g of CsCl was added and allowed to thoroughly dissolve. Two mg of Hoechst dye (No 33258, Sigma-Aldrich, St. Louis, MO, USA) were added and the sample was gently inverted for at least an hour to allow the dye to bind DNA. The sample was then transferred to Quick-Seal centrifuge tubes (Beckman Coulter Inc.) and spun at 40,000 *g* for 40 h using a Ti-75 rotor in a L8-M ultracentrifuge (Beckman Coulter, Inc.). DNA bands were visualized under long wave ultraviolet light and individually extracted with a 30 gauge needle with the tip cut off. Bands were removed from top to bottom in the gradient and subjected to at least three rounds of butanol extraction to remove the dye. The resulting samples were ethanol precipitated overnight at 4 °C. The DNA was rehydrated in 10 µl (bands 1–4) or 500 µl (band 5) of TE buffer and 10 µl of each were loaded onto a 0.8% agarose gel and run at 60 V for 2 h. Fraction 1 DNA was purified from a second 8-l batch of culture and digested with *Eco*R1. The DNA was then separated on an 1.0% agarose gel and the 2-kb size fraction was removed, purified, and cloned into puc19. A single 96-well plate of clones derived from this library was sequenced and primers were designed to three putative nucleomorph genes for the synthesis of Southern hybridization probes.

Southern hybridization. DNA transfer to a positively charged membrane was performed as in Lane et al. (2006) and labeled DNA probes were synthesized using the PCR digoxygenin (DIG) Synthesis Kit (Roche Diagnostics Corp., Indianapolis, IL, USA). The oligonucleotide primers used to amplify the probes can be found in Table 2. Hybridizations were done overnight at 57 °C (rDNA probes) or 45 °C (protein-coding genes). Following hybridization, the membranes were processed using the DIG Luminescent Detection Kit and CDP-Star substrate (Roche Diagnostics Corp.).

Sequence analysis. Individual DNA sequence reads were assembled into contigs and edited in Sequencher 4.2 (Gene Codes Corp., Ann Arbor, MI, USA). Multiple sequence alignments were constructed in MacClade v 4.06 (Maddison and Maddison 2003) and aligned by eye. Ambiguously aligned positions were removed before analysis. The nucleomorph encoded 18S rDNA data set was subjected to maximum likelihood (ML) analysis in PAUP v 4.0b10 (Swofford 2002). One-thousand random sequence addition replicates were performed using the heuristic search option, tree bisection reconnection (TBR) branch swapping and the model determined by Modeltest v 3.7 (Posada and Crandall 1998) (TVM+I+Γ). Bootstrap analysis was performed with 100 replicates of 10 random sequence additions. MrBayes v 3.0 (Huelsenbeck and Ronquist 2003) was run for 1,000,000 generations using

Table **2.** Nucleomorph, mitochondrial and plastid genes targeted for Southern hybridization probes and the primers used to amplify them.

Gene	Forward Primer (5′–3′)	Reverse Primer (5′–3′)
5S rDNA	TATACGGCCATCCTCATGG	ATACAGCACCTAGAGTTCC
28S rDNA	ATCTCTTAGGATCGACTCACC	TAAGCAGAACTGGCGATGC
18S rDNA	ATCATTCAAATTTCTGCC	CGACTACGAGCTTTTTAACTGC
*ubc*4	ATGGCTACAAAACGAATCC	CATTCTTGTGCAGTTGCC
*tub*G	TGCTGGAAACAACTGGGC	TTGTAAGGTTGTATCAC
*hsp*70	GTAGAAAATTYGARGARYTNTGYATG	ATCAAATGTTACYTCDATYTGNGG
*iap*100	TCCTAGTGGAAACGAACTAG	GGCCATTTCAGATTTATCC
*rpa*2	AATTATCACAAACTAGTG	TTACCGTACAGATAC
*cox*I	CCAACCAYAAAGATATWGGTA	GGATGACCAAARAACCAAAA
*chl*L	CTAACGTGAAGCTTGCAGTATACG	GCCATACATCTTCATAGTGG

Primers for tubG, iap100, and rpa2 are exact-match to *Hemiselmis rufescens* CCMP644. The primer sets for the 18S and hsp70 are cryptomonad nucleomorph-specific. *cox*I is encoded in the mitochondrion and *chl*L in the plastid of *H. rufescens*.

the GTR+I+Γ model to estimate posterior probabilities. Trees were sampled every 100 generations and the first 4,000 trees were discarded as ''burn-in.'' The remaining 6,000 trees were used to calculate the consensus tree.

RESULTS

Nucleomorph karyotypes and 18S rDNA sequencing. As part of an ongoing investigation of nucleomorph genome size and structural diversity in a wide range of cryptomonad species, we discovered that members of the genus *Hemiselmis* have a nucleomorph genome structure unlike any of those described thus far. PFGE of *H. rufescens* CCMP644 (formerly identified as *Plagioselmis prolonga*) revealed the presence of three small chromosomes, approximately 205, 190, and 185 kb in size, similar to the nucleomorph chromosomes in the model cryptomonad *G. theta* (Fig. 1). Southern hybridizations using a probe designed to a fragment of the nucleomorph 18S rDNA locus failed to produce positive hybridization signals for all three chromosomes, as has been seen in all other examined species (Lane et al. 2006; Rensing et al. 1994). Chromosome II consistently produced no signal, and the hybridization signal derived from chromosome III was approximately half as intense as that seen for chromosome I, suggesting a possible difference in the number of rDNA copies between the two (Fig. 1).

To determine whether this unusual feature is unique to this species or is also present in related taxa, we selected two additional relatively unstudied *Hemiselmis* strains for 18S rDNA amplification, PFGE and Southern hybridization analysis: 1) *H. rufescens* CCMP439 (from which the nucleomorph 18S rDNA sequence was available; (Hoef-Emden, Marin, and Melkonian 2002); and (2) *Hemiselmis* sp. RCC659. A member of a related genus, *Chroomonas pauciplastida*, was also included. In *H. rufescens* CCMP439 and *Hemiselmis* sp. RCC659, chromosome II also failed to hybridize with the 18S rDNA probe, and there was a major difference in the size of the chromosome between the two species (Fig. 1). This size variation is intriguing given that the two strains of *H. rufescens* and *Hemiselmis* sp. RCC659 are extremely closely related to one another, sharing >98% similarity over 1,830 bp of their nucleomorph 18S rDNA sequences (Fig. 2). In contrast, the probe hybridized to all three of the (much larger) nucleomorph chromosomes in *C. pauciplastida* (Fig. 1). These results suggested two possibilities; either the nucleomorph genome of *Hemiselmis* spp. is spread over only two chromosomes (with the middle band on the EtBr-stained gel being derived from a different genome), or chromosome II of these taxa originated from the nucleomorph, but does not encode a copy of the 18S rDNA locus. In order to distinguish between these two possibilities, and to further elucidate the dynamics of nucleomorph

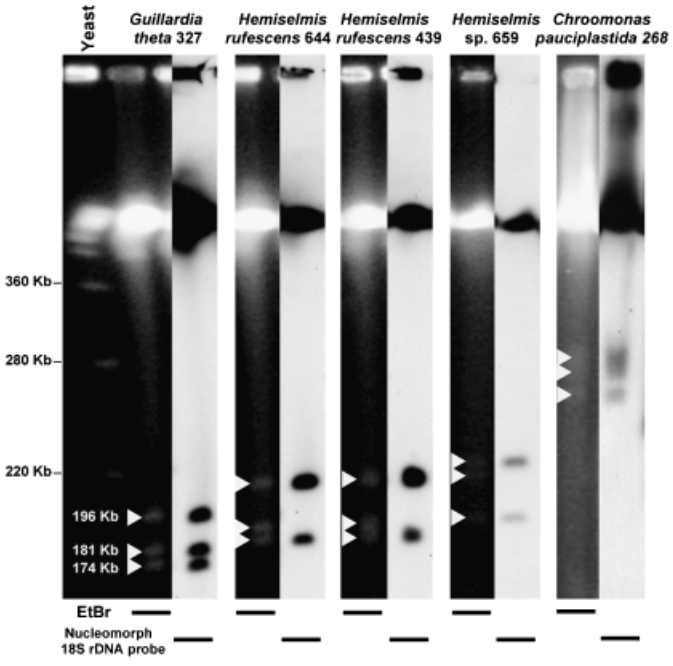

Fig. **1.** Pulsed-field gel electrophoresis separation of cryptomonad nucleomorph chromosomes showing the disparity in 18S ribosomal DNA probe hybridization signal between members of the genus *Hemiselmis* and other cryptomonads. Whereas ethidium bromide staining reveals the presence of three nucleomorph bands in all cryptomonads, members of the genus *Hemiselmis* only show a positive Southern hybridization signal to chromosomes I and III when a portion of the nucleomorph 18S rDNA gene is used as a probe. Additionally, the middle band is significantly (~30 kb) smaller in *H. rufescens* compared with *Hemiselmis* sp. RCC659. Arrowheads indicate putative nucleomorph chromosomes and known chromosome sizes are labeled. Yeast is included as a size standard. Significant cross-hybridization to host nucleus-encoded 18S rDNA gene(s) was also observed.

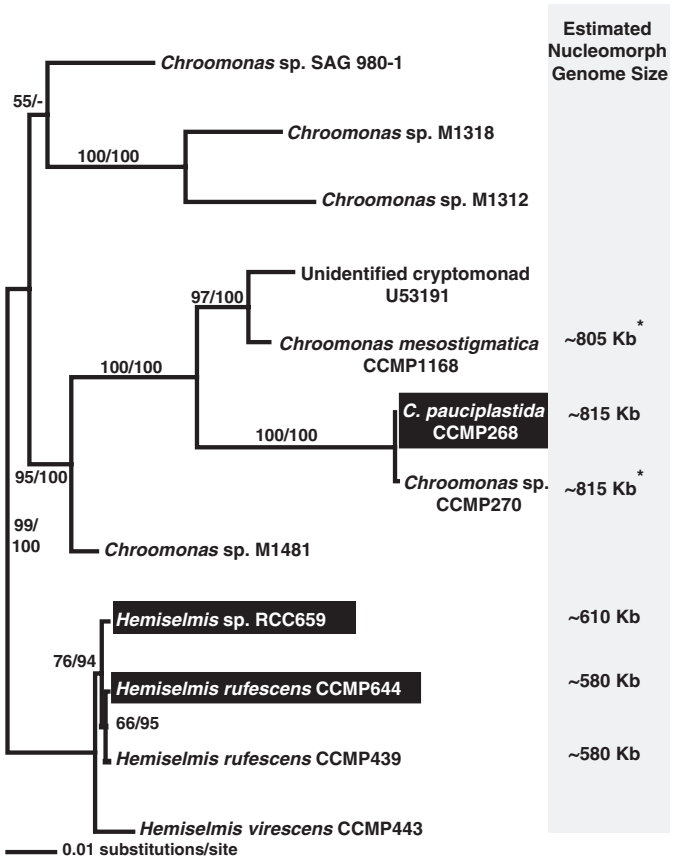

Fig. **2.** Maximum likelihood phylogenetic tree of nucleomorph 18S rDNA gene sequences showing the relationships between members of the genus *Hemiselmis* and the related genus *Chroomonas*. Nucleomorph genome size estimates, based on pulsed-field gel electrophoresis analysis, are listed adjacent to the relevant taxa. New 18S rDNA sequences are highlighted with a black background and ''*'' indicates previously published genome size estimates (Lane et al. 2006). Numbers at nodes indicate likelihood bootstrap (100 replicates) values, followed by Bayesian posterior probabilities. Values less than 50 are indicated by ''-.'' The tree is arbitrarily rooted between *Hemiselmis* and *Chroomonas*.

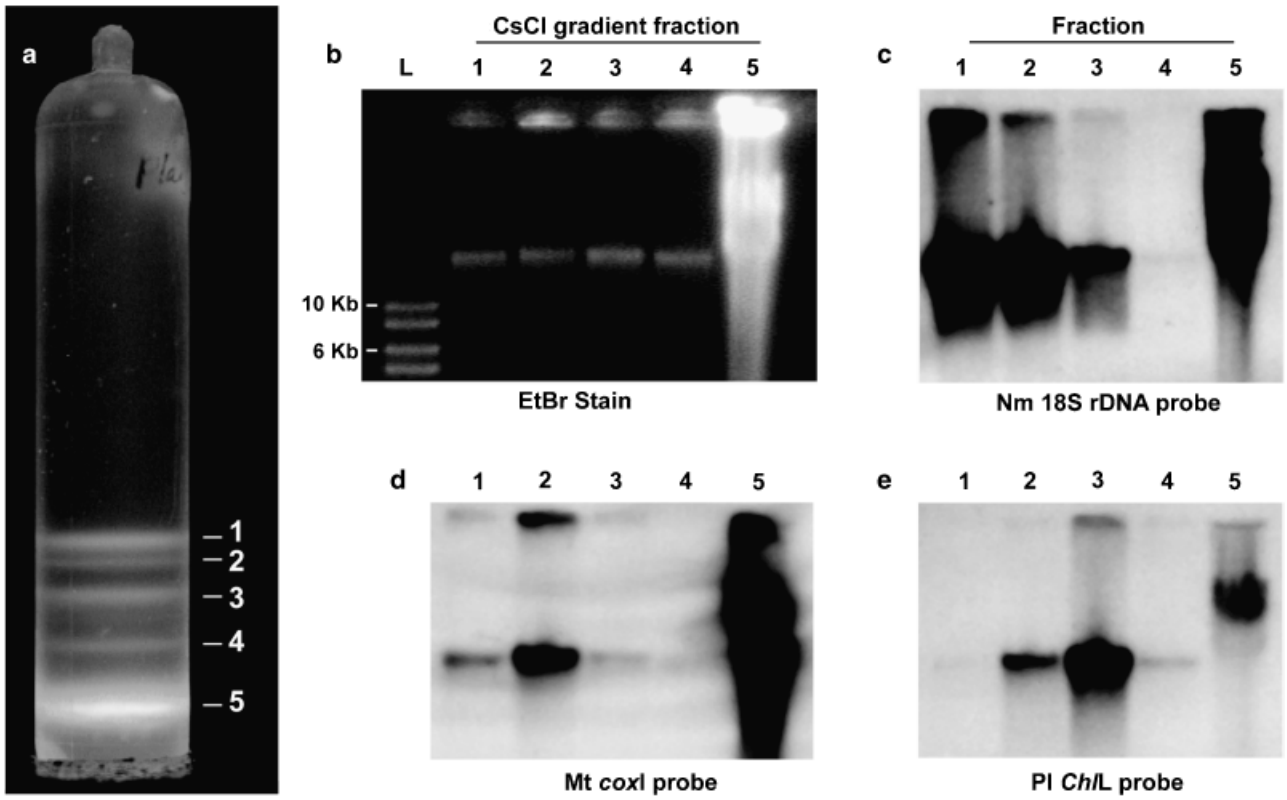

Fig. **3.** Hoechst dye-CsCl gradient centrifugation of total DNA from *Hemiselmis rufescens* CCMP644. (**A**) Five DNA bands were resolved in the gradient and extracted individually. (**B**) The fractions were separated on an agarose gel and visualized using ethidium bromide stain. Southern hybridizations were performed using a portion of (**C**) nucleomorph 18S ribosomal DNA, (**D**) mitochondrion-encoded cytochrome oxidase subunit I (*cox*I), and (**E**) plastid-encoded protochlorophyllide reductase subunit L (*chl*L) as probes to determine which fraction(s) contained DNA from each genome. Fraction 1 showed the strongest nucleomorph signal, whereas fractions 2 and 3 were shown to contain mitochondrial and plastid DNA, respectively. Fractions 4 and 5 contained nuclear DNA (see text). Some cross-hybridization with the nuclear fraction was evident with all probes and may be the result of contamination on the gradient, the presence of copies of organellar genes in the nuclear genome, or both.

genome structure and function, we selected *H. rufescens* CCMP644 for more detailed molecular investigation.

Genomic characterization of *H. rufescens* CCMP644. Five distinct DNA bands were produced after CsCl gradient centrifugation (Fig. 3a). Bands were individually extracted from the gradient and run on an agarose gel (Fig. 3b). To elucidate the identity of each of these bands, gene probes were designed against select mitochondrial, nucleomorph and plastid genes (Table 2). Southern hybridizations against these fractions indicated that the band containing DNA of the lowest density (i.e. most A/T rich) showed a positive signal when probed with a portion of nucleomorph 18S rDNA (Fig. 3c). Bands two and five also produced considerable signal when probed with the nucleomorph 18S rDNA probe, but subsequent hybridizations using mitochondrial *cox*I (Fig. 3d) revealed that the mitochondrial DNA was mainly contained in band 2 and the close proximity of fractions one and two on the gradient resulted in a certain amount of cross-contamination between the two. Band 5 clearly represented the nuclear DNA (based on its abundance and G/C content) and cross hybridization with the nucleomorph 18S rDNA probe was expected based on our initial PFGE results (Fig. 1). A plastid 16S rDNA probe (data not shown) and a probe for the light-independent protochlorophyllide reductase (*chl*L) gene, which is encoded in the *H. rufescens* plastid genome (Archibald Lab, unpubl.), were used to identify band 3 as that containing plastid DNA (Fig. 3e). None of the probes hybridized significantly to band 4, which most likely represents a satellite fraction of nuclear DNA with slightly higher A/T content than the rest of the genome.

The CsCl gradient experiments allowed us to isolate a fraction highly enriched with nucleomorph DNA from an organism with a previously undescribed genomic arrangement. To further characterize the nucleomorph genome of *H. rufescens*, a restriction enzyme library was made from material isolated from fraction 1 (see ''Materials and Methods''). Test sequencing from this library confirmed that this fraction is indeed derived from the nucleomorph genome. To elucidate the identity of Chromosome II, primers were designed to synthesize Southern hybridization probes for three nucleomorph genes: (1) *iap*100—encoding a component of the plastid import machinery; (2) *rpa*2—replication protein A2; and (3) *tub*G—Gamma tubulin. A fourth nucleomorph gene for the 70-kDa heat shock protein (*hsp*70) was amplified using degenerate primers in use in the Archibald Lab (Table 2). The probes for *iap*100 and *rpa*2 produced positive hybridization signals for chromosome II of *H. rufescens* (Fig. 4). These results indicate that chromosome II is of nucleomorph origin and not from another source. The *tub*G and *hsp*70 probes hybridized to chromosome I in *H. rufescens*, providing evidence that these genes are both encoded there, rather than on separate chromosomes, as in *G. theta*.

Having confirmed that chromosome II in *H. rufescens* is indeed a nucleomorph chromosome, we synthesized *H. rufescens* probes corresponding to each major component of the sub-telomeric repeats in the *G. theta* nucleomorph genome (Table 2). Southern hybridizations showed that, as is the case for the 18S rDNA (Fig. 1), the 28S and *ubc*4 loci appear to be absent from chromosome II in *H. rufescens*, but that the 5S rDNA locus persists (Fig. 5). These

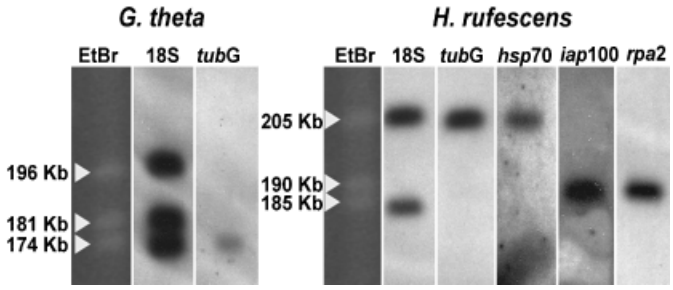

Fig. **4.** Southern hybridization results for nucleomorph-encoded gene probes showing their genomic locations. The 18S ribosomal DNA probe hybridizes to all three chromosomes in *Guillardia theta* (as expected) and only chromosome I and III of *Hemiselmis rufescens* CCMP644. Probes designed against nucleomorph-encoded genes (*iap*100—encoding a component of the plastid import machinery, *rpa*2—replication protein A2, *tub*G—Gamma tubulin, and *hsp*70—the nucleomorph gene for the 70-kDa heat shock protein) hybridized to each of the chromosomes. The probes for *iap*100 and *rpa*2 revealed that chromosomes II of *H. rufescens* is a *bona fide* nucleomorph chromosome even though it lacks copies of the 18S rDNA.

results were also observed for *Hemiselmis* sp. RCC659 (data not shown). Several attempts to amplify a probe to hybridize to the nucleomorph telomere in *Hemiselmis* were unsuccessful.

DISCUSSION

We have shown that members of the little-studied cryptomonad genus *Hemiselmis* possess a nucleomorph genome unlike any previously described. The significance of the absence of complete rDNA cistrons at the ends of chromosome II in *H. rufescens* CCMP644 remains unclear, but in *G. theta*, these repeated portions of the genome have an atypically high G/C content ($\sim$45%) for nucleomorph DNA (Lane et al. 2006). In addition to being a feature of the nucleomorph genome in both *G. theta* and *Bigelowiella natans*, sub-telomeric rDNA is present in the nuclear genomes of *Dictyostelium discoideum* (Eichinger et al. 2005), the diplomonad *Giardia lamblia* (Hou et al. 1995; Le Blancq, Korman, and Van der Ploeg 1991), and the microsporidian *Encephalitozoon cuniculi* (Brugere et al. 2000; Katinka et al. 2001). Convergence on subtelomeric rDNA cistrons in reduced genomes is unlikely to be coincidence, but a functional role has yet to be established. It is believed that rDNA has a role in telomere maintenance in the *D. discoideum* genome (Eichinger et al. 2005). Where studied, the rDNA regions at the ends of nucleomorph chromosomes, and the non-coding region adjacent to them, are almost identical at each chromosome end, indicating a high level of recombination or gene conversion (Douglas et al. 2001; Gilson and McFadden 2002; Gilson et al. 2006). The fact that the entire repeated region in *G. theta* (rDNA cistron and *ubc*4) except the 5S rDNA (which is closest to the telomere) has been lost from chromosome II in *Hemiselmis* suggests that they are not strictly essential for chromosome structure in nucleomorphs, but their maintenance in all other species examined thus far argues for an important role.

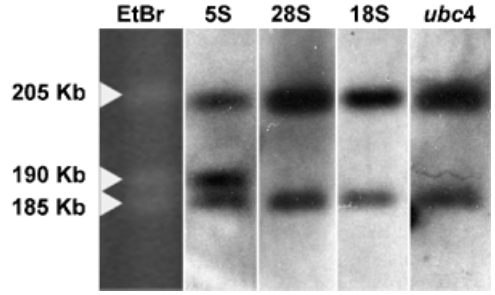

Fig. **5.** Southern hybridizations against the nucleomorph chromosomes of *Hemiselmis rufescens* CCMP644, which had been separated using pulsed field gel electrophoresis (see Fig. 1), using probes designed to each of the main components of ribosomal DNA, as well as ubiquitin conjugating enzyme subunit four (*ubc*4). These genes occur in each of the subtelomeric repeat regions in the sequenced nucleomorph genome of *Guillardia theta*. In *H. rufescens*, whereas all of the probes hybridize to chromosomes I and III, only the 5S rDNA produced a hybridization signal against chromosome II.

In the nucleomorph genome of the chlorarachniophyte *B. natans* (Gilson et al. 2006), the sub-telomeric multi-copy region includes a complete copy of the rDNA cistron on each chromosome end (albeit without a 5S rDNA gene), as well as a truncated form of *dna*K on five of six ends. Gilson et al. (2006) speculated that the *dna*K pseudogene has spread from the single intact *dna*K locus on the left arm of chromosome II to the remaining chromosome ends by recombination acting to homogenize the rDNA repeats. In cryptomonads, if the rDNA arrangement in the *H. rufescens* nucleomorph genome was initially the same as in *G. theta*, it appears that a cross-over event occurred on chromosome II between the 5S gene and the remaining portions of the cistron, which resulted in the loss of the majority of the rDNA locus. This arrangement was likely then propagated to at least the other end of chromosome II.

Recombination and gene conversion at subtelomeric regions are common mechanisms in yeast, whose telomeres have been thoroughly studied (Bhattacharyya and Lustig 2006; McEachern and Iyer 2001), and high rates of recombination have been suggested to be associated with duplication and divergence at the ends of wheat chromosomes (See et al. 2006). If recombination is most likely to occur at the chromosome ends and multiple copies of the rDNA are necessary to carry out sufficient rates of translation, then their subtelomeric location in nucleomorph genomes might be dictated by this structural constraint. Conversely, it may be the rDNA operons themselves that promote recombination at the chromosomal ends, providing an anchor point for DNA exchange between non-homologous chromosomes. It is significant that only one multi-copy gene, designated ORF160, in the *G. theta* nucleomorph genome is found outside of the subtelomeric regions: three copies of ORF160 occur in *G. theta*, one on chromosome I and two copies on III, and all three are 100% identical at the nucleotide level (Douglas et al. 2001). Determining the gene content at the ends of chromosome II in *H. rufescens* may improve our understanding of their evolution and separate ''chicken'' from ''egg.'' Efforts to completely sequence the nucleomorph genome of *H. rufescens* CCMP644 are currently underway.

Nucleomorph genome karyotypes have been shown to vary significantly between species (Lane et al. 2006; Rensing et al. 1994), but the reasons for this variation are unclear. Obvious possibilities include significant variation in amounts of non-coding DNA in different species, differing coding capacities or some combination of both. We have demonstrated previously that there is variation in the size of the non-coding DNA region between the single-copy DNA and rDNA cistrons in the nucleomorph chromosomes of different species (Lane et al. 2006), but whether this is a general trend throughout the chromosomes is unclear. Whereas nucleomorph chromosomes may differ in size simply due to the amount of non-coding DNA, if gene content makes up the majority of the difference, then nucleomorph genomes may lose coding DNA in large portions during recombination events. The results presented here indicate that within *Hemiselmis*, there are large differences ($\sim$30 kb) in the size of chromosome II between *Hemiselmis* sp. RCC659 and the closely related *H. rufescens* clade. The similarity in the size of chromosomes I and III among

these species, in addition to their close phylogenetic relationship, suggests that the size difference of chromosome II represents one or a few recent and large deletion(s) or, less likely in our opinion, an insertion(s). It is very unlikely that enough non-coding DNA could be lost in an already highly reduced genome to account for the difference over such short evolutionary time. Also, given the extremely compact nature of nucleomorph genomes and the almost complete lack of pseudogenes in the *G. theta* sequence, insertion events probably occur very rarely, if at all. It seems more likely that large portions of the nucleomorph genome have been lost in an ancestor of *H. rufescens* in one or a few events.

The idea that inter-chromosomal recombination could play a major role in shaping the nucleomorph genome is supported by our Southern hybridization data. In *G. theta*, *hsp*70 and *tub*G are located on chromosomes I and III, respectively (Douglas et al. 2001). In contrast, these genes are both found on chromosome I in *H. rufescens* CCMP644. Additionally, probes for *rpa*2 and *iap*100 both hybridize to chromosome II in *H. rufescens*, but are on different chromosomes in *G. theta*. Recombination has clearly reorganized the nucleomorph genomes in cryptomonad lineages since their common ancestor, and our karyotype data suggest that this is an ongoing process. The complete sequence of the *H. rufescens* nucleomorph should provide insight into the question of whether large or small deletion events are the major force in nucleomorph genome size variation among cryptomonads, but preliminary evidence presented here points towards large losses. Additionally, a second cryptomonad nucleomorph genome sequence will allow for a comparison of the coding capacity of these miniature genomes and shed light on the core set of proteins required for their maintenance.

ACKNOWLEDGMENTS

We thank Camilla Nesbø, Inaki Ruiz-Trillo, and two anonymous reviewers for helpful comments on the manuscript. We also appreciate the assistance of David Spencer during the CsCl experiments and the work of Cathy Kozera and Sharen Bowman on the *H. rufescens* CCMP644 restriction enzyme library. This work was supported by Genome Atlantic and an NSERC Discovery Grant (283335–2004) awarded to JMA. JMA is a Scholar of the Canadian Institute for Advanced Research, Program in Evolutionary Biology.

LITERATURE CITED

Archibald, J. M. & Keeling, P. J. 2005. On the origin and evolution of plastids. *In*: Saap, J. (ed.), Microbial Phylogeny and Evolution. Oxford University Press, New York. p. 238–260.

Bhattacharya, D., Yoon, H. S. & Hackett, J. D. 2004. Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *Bioessays*, **26**:50–60.

Bhattacharyya, M. K. & Lustig, A. J. 2006. Telomere dynamics in genome stability. *Trends Biochem. Sci.*, **31**:114–122.

Brugere, J.-F., Cornillot, E., Metenier, G., Bensimon, A. & Vivares, C. P. 2000. *Encephalitozoon cuniculi* (Microspora) genome: physical map and evidence for telomere-associated rDNA units on all chromosomes. *Nucleic Acids Res.*, **28**:2026–2033.

Delwiche, C. F. 1999. Tracing the thread of plastid diversity through the tapestry of life. *Am. Nat.*, **154**(Suppl.):S164–S177.

Douglas, S. 1988. Physical mapping of the plastid genome from the chlorophyll c-containing alga, *CryptomonasΦ*. *Curr. Genet.*, **14**: 591–598.

Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L. T., Wu, X., Reith, M., Cavalier-Smith, T. & Maier, U. G. 2001. The highly reduced genome of an enslaved algal nucleus. *Nature*, **410**: 1091–1096.

Eichinger, L., Pachebat, J. A., Glockner, G., Rajandream, M. A., Sucgang, R., Berriman, M., Song, J., Olsen, R., Szafranski, K., Xu, Q., Tunggal, B., Kummerfeld, S., Madera, M., Konfortov, B. A., Rivero, F., Bankier, A. T., Lehmann, R., Hamlin, N., Davies, R., Gaudet, P., Fey, P., Pilcher, K., Chen, G., Saunders, D., Sodergren, E., Davis, P., Kerhornou, A., Nie, X., Hall, N., Anjard, C., Hemphill, L., Bason, N., Farbrother, P., Desany, B., Just, E., Morio, T., Rost, R., Churcher, C., Cooper, J., Haydock, S., van Driessche, N., Cronin, A., Goodhead, I., Muzny, D., Mourier, T., Pain, A., Lu, M., Harper, D., Lindsay, R., Hauser, H., James, K., Quiles, M., Babu, M. M., Saito, T., Buchrieser, C., Wardroper, A., Felder, M., Thangavelu, M., Johnson, D., Knights, A., Loulseged, H., Mungall, K., Oliver, K., Price, C., Quail, M. A., Urushihara, H., Hernandez, J., Rabbinowitsch, E., Steffen, D., Sanders, M., Ma, J., Kohara, Y., Sharp, S., Simmonds, M., Spiegler, S., Tivey, A., Sugano, S., White, B., Walker, D., Woodward, J., Winckler, T., Tanaka, Y., Shaulsky, G., Schleicher, M., Weinstock, G., Rosenthal, A., Cox, E. C., Chisholm, R. L., Gibbs, R., Loomis, W. F., Platzer, M., Kay, R. R., Williams, J., Dear, P. H., Noegel, A. A., Barrell, B. & Kuspa, A. 2005. The genome of the social amoeba *Dictyostelium discoideum*. *Nature*, **435**:43–57.

Eschbach, S., Hofmann, C. J., Maier, U.-G., Sitte, P. & Hansmann, P. 1991. A eukaryotic genome of 660 kb: electrophoretic karyotype of nucleomorph and cell nucleus of the cryptomonad alga, *Pyrenomonas salina*. *Nucleic Acids Res.*, **19**:1779–1781.

Gilson, P. R. & McFadden, G. I. 1999. Molecular, morphological and phylogenetic characterization of six chlorarachniophyte strains. *Phycol. Res.*, **47**:7–19.

Gilson, P. R. & McFadden, G. I. 2002. Jam packed genomes—a preliminary, comparative analysis of nucleomorphs. *Genetica*, **115**:13–28.

Gilson, P. R., Su, V., Slamovits, C. H., Reith, M. E., Keeling, P. J. & McFadden, G. I. 2006. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. *Proc. Natl. Acad. Sci. USA*, **103**:9566–9571.

Hoef-Emden, K., Marin, B. & Melkonian, M. 2002. Nuclear and nucleomorph SSU rDNA phylogeny in the Cryptophyta and the evolution of cryptophyte diversity. *J. Mol. Evol.*, **55**:161–179.

Hou, G. Y., Leblancq, S. M., Yaping, E., Zhu, H. X. & Lee, M. G. S. 1995. Structure of a frequently rearranged ribosomal RNA encoding chromosome in *Giardia lamblia*. *Nucleic Acids Res.*, **23**:3310–3317.

Huelsenbeck, J. & Ronquist, F. 2003. MrBayes, version 3.0. Evolutionary Biology Centre, Uppsala University.

Katinka, M. D., Duprat, S., Cornillot, E., Metenier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretaillade, E., Brottier, P., Wincker, P., Delbac, F., El Alaoui, H., Peyret, P., Saurin, W., Gouy, M., Weissenbach, J. & Vivares, C. P. 2001. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature*, **414**: 450–453.

Keeling, P. J. 2004. Diversity and evolutionary history of plastids and their hosts. *Am. J. Bot.*, **91**:1481–1493.

Lane, C. E., Khan, H., MacKinnon, M., Fong, A., Theophilou, S. & Archibald, J. M. 2006. Insight into the diversity and evolution of the cryptomonad nucleomorph genome. *Mol. Biol. Evol.*, **23**:856–865.

Le Blancq, S. M., Korman, S. H. & Van der Ploeg, L. H. 1991. Frequent rearrangements of rRNA-encoding chromosomes in *Giardia lamblia*. *Nucleic Acids Res.*, **19**:4405–4412.

Maddison, W. & Maddison, D. 2003. MacClade, version 4.06. Sinauer Associates, Sunderland, MA.

Marin, B., Nowack, E. C. & Melkonian, M. 2005. A plastid in the making: evidence for a second primary endosymbiosis. *Protist*, **156**:425–432.

McEachern, M. J. & Iyer, S. 2001. Short telomeres in yeast are highly recombinogenic. *Mol. Cell*, **7**:695–704.

McFadden, G. I., Gilson, P. R., Hofmann, C. J., Adcock, G. J. & Maier, U.-G. 1994. Evidence that an amoeba acquired a chloroplast by retaining part of an engulfed eukaryotic alga. *Proc. Natl. Acad. Sci. USA*, **91**:3690–3694.

Posada, D. & Crandall, K. A. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14**:817–818.

Rensing, S. A., Goddemeier, M., Hofmann, C. J. & Maier, U.-G. 1994. The presence of a nucleomorph *hsp*70 gene is a common feature of Cryptophyta and Chlorarachniophyta. *Curr. Genet.*, **26**:451–455.

See, D. R., Brooks, S., Nelson, J. C., Brown-Guedira, G., Friebe, B. & Gill, B. S. 2006. Gene evolution at the ends of wheat chromosomes. *Proc. Natl. Acad. Sci. USA*, **103**:4162–4167.

Swofford, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (* and other Methods), Version 4.0b10 PPC. Sinauer Associates, Sunderland, MA.

Van der Auwera, G., Hofmann, C. J., De Rijk, P. & De Wachter, R. 1998. The origin of red algae and cryptomonad nucleomorphs: A comparative phylogeny based on small and large subunit rRNA sequences of *Palmaria palmata*, *Gracilaria verrucosa*, and the *Guillardia theta* nucleomorph. *Mol. Phylogenet. Evol.*, **10**:333–342.

Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G. & Bhattacharya, D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.*, **21**:809–818.

Zauner, S., Fraunholz, M., Wastl, J., Penny, S., Beaton, M., Cavalier-Smith, T., Maier, U.-G. & Douglas, S. 2000. Chloroplast protein and centrosomal genes, a novel tRNA intron, and odd telomeres in an unusually compact eukaryotic genome-the cryptomonad nucleomorph. *Proc. Natl. Acad. Sci. USA*, **97**:247–260.

Received: 05/11/06, 07/05/06, 07/21/06; accepted: 08/01/06