

Overexpression of Molecular Chaperone Genes in Nucleomorph Genomes

Yoshihisa Hirakawa,^{*,†,1} Shigekatsu Suzuki,^{†,2} John M. Archibald,^{3,4} Patrick J. Keeling,^{4,5} and Ken-ichiro Ishida¹

¹Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

²Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

³Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada

⁴Canadian Institute for Advanced Research, Program in Integrated Microbial Biodiversity

⁵Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: hirakawa.yoshi.fp@u.tsukuba.ac.jp.

Associate editor: Howard Ochman

Abstract

Chlorarachniophytes and cryptophytes possess complex plastids that were acquired by the ingestion of a green and red algal endosymbiont, respectively. The plastids are surrounded by four membranes, and a relict nucleus, called the nucleomorph, remains in the periplastidal compartment, which corresponds to the remnant cytoplasm of the endosymbiont. Nucleomorphs contain a greatly reduced genome that possesses only several hundred genes with high evolutionary rates. We examined the relative transcription levels of the genes of all proteins encoded by the nucleomorph genomes of two chlorarachniophytes and three cryptophytes using an RNA-seq transcriptomic approach. The genes of two heat shock proteins, Hsp70 and Hsp90, were highly expressed under normal conditions. It has been shown that molecular chaperone overexpression allows an accumulation of genetic mutations in bacteria. Our results suggest that overexpression of heat shock proteins in nucleomorph genomes may play a role in buffering the mutational destabilization of proteins, which might allow the high evolutionary rates of nucleomorph-encoded proteins.

Key words: endosymbiosis, plastid, algae, evolutionary rates, transcriptome.

Plastids were acquired by multiple endosymbiotic events of photosynthetic organisms. Plants and several algae (e.g., glaucophytes, green, and red algae) acquired plastids through a primary endosymbiosis of a cyanobacterium (Rodríguez-Ezpeleta et al. 2005; Price et al. 2012). Many other algal groups have more complex plastids, called secondary plastids, which originated from green or red algal endosymbionts via secondary endosymbioses (Gould et al. 2008; Archibald 2009; Keeling 2010). In these events, many of the genes residing in the endosymbiont genomes have been lost or transferred to the host nuclear genome, resulting in the highly reduced organellar genome in modern-day plastids (Martin et al. 1998; Dorrell and Howe 2012). Chlorarachniophytes and cryptophytes possess secondary plastids derived from green and red algal endosymbionts, respectively. These two algal groups are of special interest because their secondary plastids still possess a relict nucleus, called the “nucleomorph,” of the engulfed algae, whereas many other algae have lost this organelle (Gilson 2001; Archibald 2007). Nucleomorphs are located in the plastid intermembrane space, the so-called periplastidal compartment (PPC), which corresponds to the remnant cytoplasm of the algal endosymbiont. Nucleomorphs contain a greatly reduced genome under 1 Mb in size, which is the smallest among the known eukaryotic nuclear genomes (Moore and Archibald 2009; Ishida et al. 2011). Complete

nucleomorph genome sequences have been reported in a chlorarachniophyte (Gilson et al. 2006) and four cryptophyte species (Douglas et al. 2001; Lane et al. 2007; Tanifuji et al. 2011; Moore et al. 2012). Nucleomorph genomes sequenced so far consist of three linear chromosomes possessing only several hundred genes. All nucleomorph genomes encode a small number of plastid-associated proteins and hundreds of housekeeping proteins, suggesting that nucleomorphs are essential for the maintenance of plastids. However, numerous important genes for nucleomorph biogenesis are absent from nucleomorph genomes (e.g., DNA polymerase genes). Recently, complete nuclear genomes were sequenced for the chlorarachniophyte *Bigelowiella natans* and the cryptophyte *Guillardia theta* (Curtis et al. 2012), revealing the existence of more than 1,000 nucleus-encoded proteins predicted to be imported into the PPC to carry out essential functions not encoded by nucleomorph DNA.

Nucleomorph genomes are characterized by a highly compact structure with very short intergenic regions (Keeling and Slamovits 2005; Williams et al. 2005). The impact of this gene-dense structure on the regulation of nucleomorph gene expression is poorly understood. Recently, Tanifuji et al. (2014) reported the transcription patterns of nucleomorph genes using genome mapping analyses with RNA-seq data sets in the chlorarachniophyte *B. natans* and three cryptophytes. In

all four species, RNA transcripts covered over 99% of the entire nucleomorph genomes including intergenic regions, and global transcript levels were equal or higher for nucleomorph genes than for nuclear homologs. We have studied the nucleomorph mRNA expression in these four species as well as another chlorarachniophyte *Amorphochlora amoebiformis* and discovered that nucleomorph genes for two heat shock proteins (NmHsp70 and NmHsp90) are transcribed to a remarkable degree. Our results imply a relationship between higher levels of Hsp transcripts and higher evolutionary rates of nucleomorph genes.

To calculate the relative mRNA expression levels of nucleomorph-encoded proteins, we analyzed transcriptome data from two chlorarachniophytes, *A. amoebiformis* CCMP2058 (formerly *Lotharella amoebiformis*) and *B. natans* CCMP2755, and three cryptophytes, *G. theta* CCMP2712, *Cryptomonas paramecium* CCAP977/2A, and *Chroomonas mesostigmatica* CCMP1168, that were generated by the National Center for Genome Resources Marine Microbial Eukaryotic Transcriptome Sequencing Project (sample IDs are MMETSP0038, 0042, 0045, 0046, and 0047; <https://www.marinemicroeukaryotes.org/>, last accessed March 12, 2014). Several million RNA-seq reads were mapped onto the nucleomorph genome sequences, and the depth of coverage in each protein-coding region was measured. Relative transcript levels were estimated by calculating RPKM (Reads Per Kilobase of exon per Million mapped reads) values (Mortazavi et al. 2008). In the two chlorarachniophytes, the average transcript levels among all nucleomorph genes were found to be 3,032 and 3,691 in *A. amoebiformis* and *B. natans*, respectively. Interestingly, the transcript levels of two heat shock protein genes, *NmHsp70* and *NmHsp90*, were significantly higher, at 54,375 and 45,761 in *A. amoebiformis*, and at 35,738 and 38,000 in *B. natans*, respectively (fig. 1A and B and supplementary tables S1 and S2, Supplementary Material online); none of the genes for known cochaperones of NmHsp70 and NmHsp90 or small heat shock proteins have been detected in the nucleomorph genomes. The mRNA derived from these two genes is predicted to comprise 11% and 7% of the total nucleomorph-derived mRNA in *A. amoebiformis* and *B. natans*, respectively.

To compare the mRNA expression levels of nucleus- and nucleomorph-encoded heat shock proteins, we carried out real-time quantitative polymerase chain reaction with gene-specific primers for *A. amoebiformis*. The *NmHsp70* and *NmHsp90* genes were highly transcribed in comparison with the nucleomorph-encoded plastid translocon *NmToc75*, which showed an average transcript level (fig. 1C), in agreement with the transcriptome data. The transcript levels of the NmHsp genes were equal to or greater than those of nucleus-encoded cytoplasmic homologs (fig. 1C). Because translation rates of nuclear and nucleomorph genes are unknown, we were not able to compare the amount of protein based on mRNA transcript levels. However, it would be interesting to see whether levels of nucleomorph-encoded heat shock proteins are as abundant as their transcripts, because the volume of the PPC in which the nucleomorph resides is much smaller than the cytoplasm. Although the transcript levels of cytoplasmic *Hsp70* and *Hsp90* changed dramatically under light

and dark conditions, the *NmHsp70* and *NmHsp90* genes were constantly highly expressed in the PPC (fig. 1C). Light induction of Hsp transcripts has been reported in various photosynthetic organisms including the cyanobacterium *Synechocystis* (Hihara et al. 2001), the green alga *Chlamydomonas* (von Gromoff et al. 1989), and land plants (Li et al. 2000; Rossel et al. 2002). Our results suggest that the transcript levels of nucleus-encoded Hsps in chlorarachniophyte also increase in response to light.

Exceptionally high levels of Hsp gene transcription were also observed in cryptophyte nucleomorph genomes. The transcript levels of *NmHsp70* and *NmHsp90* were found to be 57,544 and 43,706 in *Ch. mesostigmatica*, 58,270 and 31,472 in *G. theta*, and 25,571 and 15,252 in *C. paramecium*, whereas the average transcript levels among all nucleomorph genes of these species were 2,021, 2,333, and 2,768, respectively (fig. 2A–C, supplementary tables S3–S5, Supplementary Material online). The mRNA transcripts from NmHsp genes were predicted to comprise 10%, 8%, and 3% of the total nucleomorph-derived mRNA in *Ch. mesostigmatica*, *G. theta*, and *C. paramecium*, respectively. *NmHsp70* and *NmHsp90* were the most highly expressed genes in the nucleomorph genomes of *Ch. mesostigmatica* and two chlorarachniophytes. However, in *G. theta* and *C. paramecium*, the mRNA levels of a few genes for plastid stroma proteins (CbbX and ClpP1) and nucleomorph-localized proteins (Gsp2 and H2B) were greater than those of NmHsp genes (fig. 2B and C, supplementary tables S4 and S5, Supplementary Material online).

Molecular chaperones, including heat shock proteins, are essential proteins that play an important role in the folding, disaggregation, and intracellular transport of proteins in cells (Saibil 2013). The PPC is predicted to contain fewer proteins than the cytoplasm, because the volume of PPC and the number of nucleomorph genes are much smaller than that of cytoplasm/nuclear genes. This raises the question: Why are the Hsp70 and Hsp90 molecular chaperones so highly expressed in the nucleomorph genomes? One intriguing possibility is that it is related to the evolutionary rate of nucleomorph genomes. In *Escherichia coli*, it has been reported that the overexpression of GroEL/GroES chaperonins mask the deleterious effects of mutated proteins, which allows for an increase in the number of accumulating mutations (Maisnier-Patin et al. 2005; Tokuriki and Tawfik 2009). In eukaryotic cells, Hsp90 is thought to act as a buffer for genetic variations (Taipale et al. 2010). Furthermore, members of the bacterial genus *Buchnera*, which are endosymbionts of aphids, have greatly reduced genomes, which evolve faster than their homologs in closely related free-living bacteria (Moran 1996). Most of the heat shock proteins in *Buchnera aphidicola* are overexpressed even under nonstress conditions (Wilcox et al. 2003), implying that these chaperones mask the destabilizing effects of mutations and allow the genetic variation in the endosymbiotic genome (McCutcheon and Moran 2012). A similar situation is seen in the nucleomorph genomes of chlorarachniophytes and cryptophytes.

To estimate the sequence divergence of nucleomorph genomes, we used 26 and 28 homologous sequences of

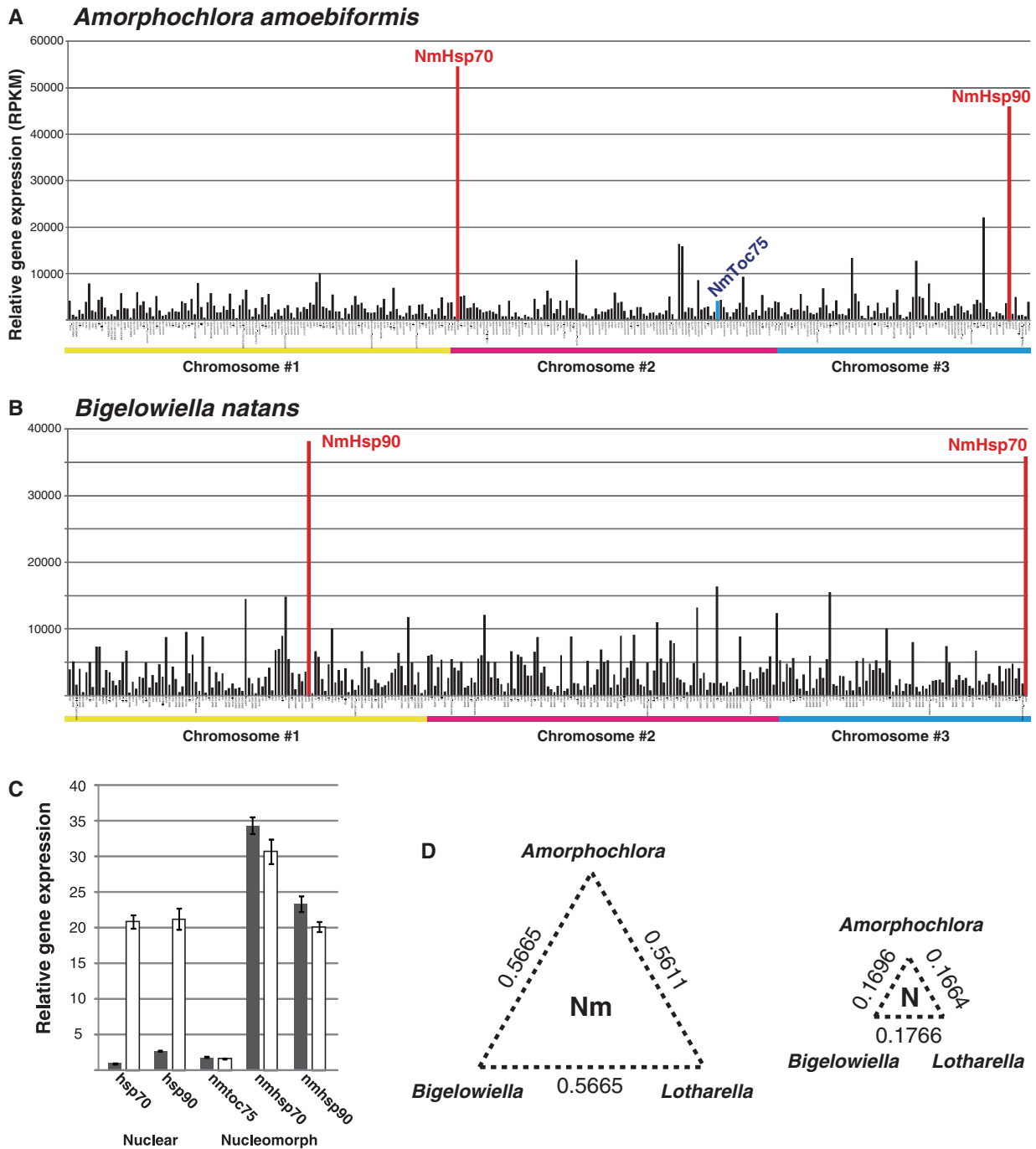


Fig. 1. Relative gene expression and evolutionary rates of nucleomorph-encoded proteins in chlorarachniophytes. (A, B) Bars show relative mRNA transcript levels (RPKM) for all nucleomorph genes in the chlorarachniophyte *Amorphochlora amoebiformis* and *Bigelowiella natans*. (C) Relative transcription levels of nucleus- and nucleomorph-encoded heat shock proteins and NmToc75 estimated by real-time quantitative polymerase chain reaction in a dark- (black bars) and a light-phase (white bars) culture. The value of dark-phase nuclear *Hsp70* is set to 1. Error bars represent the standard deviation of triplicate experiments. (D) Evolutionary rates of nuclear (N) and nucleomorph (Nm) ribosomal protein sequences among three chlorarachniophytes. The length of dotted lines indicates the evolutionary distances estimated by amino acid substitution rates (the numbers on lines).

nucleus- and nucleomorph-encoded ribosomal proteins in three chlorarachniophytes (*A. amoebiformis*, *B. natans*, and *L. globosa/L. vacuolata*) and three cryptophytes (*Ch. mesostigmatica*, *G. theta*, and *C. paramecium*), respectively. Amino acid substitution rates of the nucleomorph-encoded proteins were calculated to be between 0.5611 and 0.5665, and 0.4292 and 0.4882 per site among the three species of

chlorarachniophytes and cryptophytes, respectively, whereas the substitution rates in nucleus-encoded proteins were between 0.1664 and 0.1766, and 0.1997 and 0.2055, respectively (figs. 1D and 2D). The nucleomorph-encoded proteins were clearly evolving faster than their nuclear counterparts in both chlorarachniophytes and cryptophytes, and a high evolutionary rate of nucleomorph genes has been reported in *B. natans*

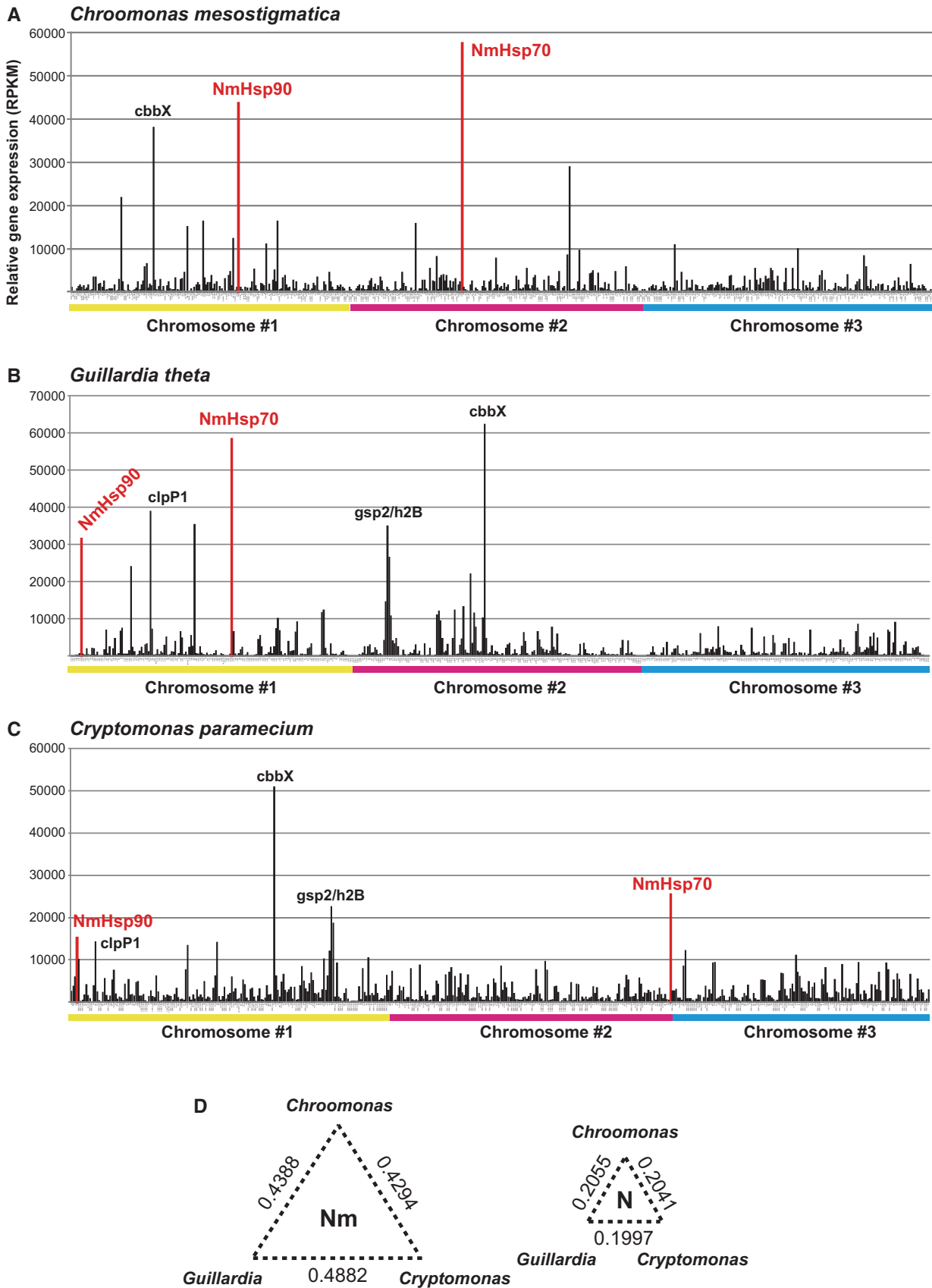


Fig. 2. Relative gene expression and evolutionary rates of nucleomorph-encoded proteins in cryptophytes. (A–C) Bars show relative mRNA transcript levels (RPKM) for all nucleomorph genes in three cryptophyte species. (D) Evolutionary rates of nuclear (N) and nucleomorph (Nm) ribosomal protein sequences among three cryptophytes. The length of dotted lines indicates the evolutionary distances estimated by amino acid substitution rates (the numbers on lines).

previously (Patron et al. 2006). These data raise the possibility that the higher expression of molecular chaperones might compensate for the fast evolution of nucleomorph genes by buffering the deleterious effects of mutated proteins in the PPC. In addition, such accelerated evolution could also be attributed to the loss of genes involved in nucleomorph DNA replication and repair (Curtis et al. 2012), like bacterial endosymbionts (McCutcheon and Moran 2012).

In our analyses, the *NmHsp70* and *NmHsp90* transcript levels of *Ch. mesostigmatica* and two chlorarachniophytes were significantly higher than the other transcripts, whereas those of *G. theta* and *C. paramecium* were somewhat less conspicuous (figs. 1 and 2). Interestingly, although the nucleomorph genomes of these two cryptophytes possess genes that encode proteasome subunits (Douglas et al. 2001; Tanifuji et al. 2011, Stork et al. 2012), *Ch. mesostigmatica* and two chlorarachniophytes completely lack these genes (Gilson et al. 2006; Moore et al. 2012). This observation is consistent with the possibility that *Ch. mesostigmatica* and

the two chlorarachniophytes are more susceptible to protein misfolding because they lack a proteasomal degradation system in the PPC. It is possible that these three species express their *NmHsp* genes at higher levels to mitigate these risks.

Our analysis of expression levels revealed a second noteworthy bias in base composition of nucleomorph genes. It has been shown that highly expressed genes are less divergent and have an amino acid compositional bias leading to a higher GC content in AT-rich genomes of bacterial endosymbionts (Schaber et al. 2005) and the eukaryotic parasite, *Plasmodium* (Chanda et al. 2005). Nucleomorph genomes are also extremely AT-rich genomes. We calculated the GC content of each gene in five nucleomorph genomes of chlorarachniophytes and cryptophytes and found a positive correlation between the GC content and the relative transcription level in all five genomes; the GC content tends to be increased depending on the transcription level (fig. 3A–E). Furthermore, we identified the 10 most highly and

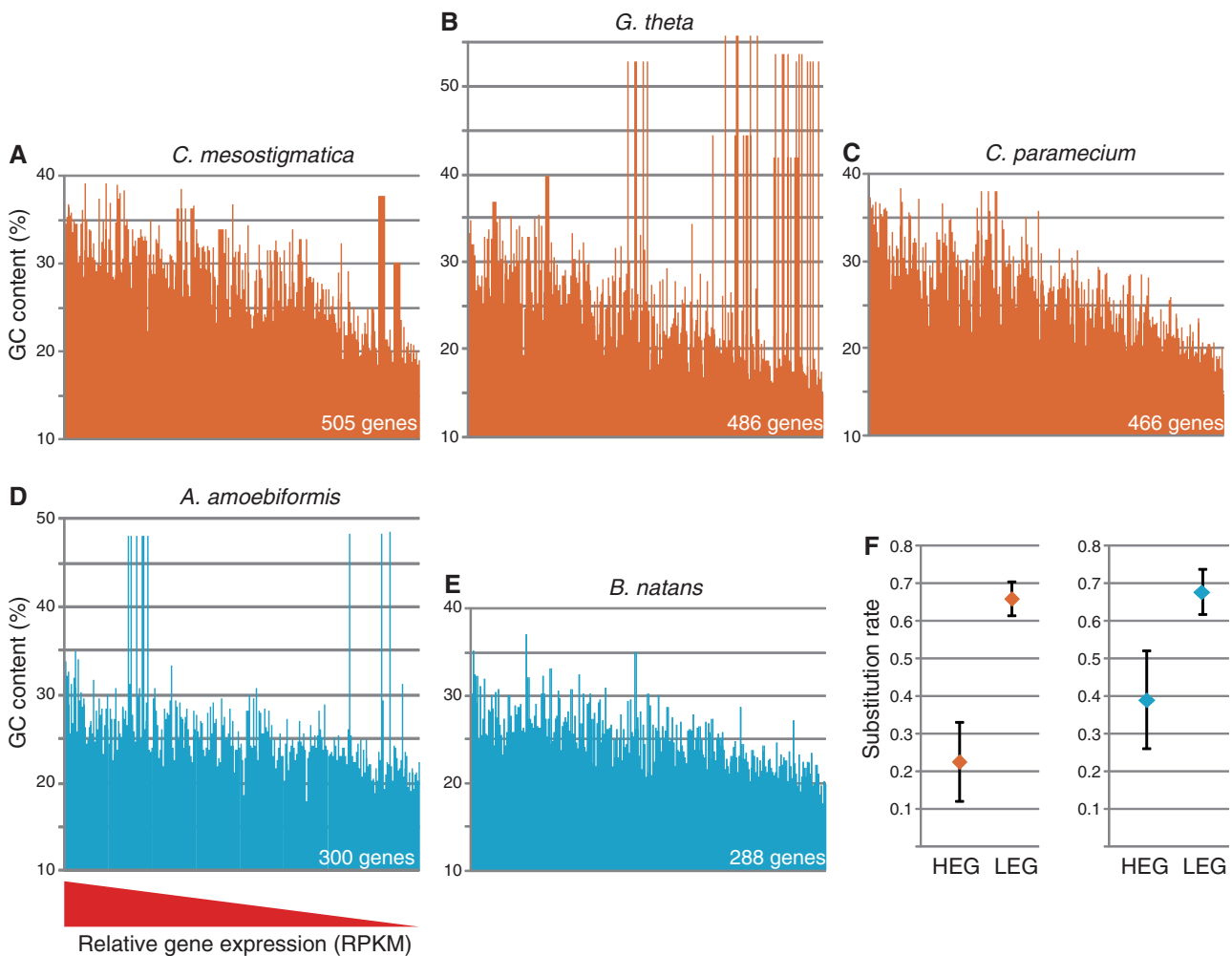


Fig. 3. Correlation between relative transcription level and GC content of nucleomorph genes. (A–E) Each bar indicates the GC content of nucleomorph genes ordered by their relative transcription levels, RPKM values, in three cryptophytes (orange) and two chlorarachniophytes (blue). Highly expressed genes reside in the left side of graphs. (F) Amino acid substitution rates of HEG/LEG between homologs from two species of cryptophytes, *Chroomonas mesostigmatica* and *Guillardia theta* (orange), and chlorarachniophytes, *Amorphochlora amoebiformis* and *Bigelowiella natans* (blue). Error bars indicate the standard deviation of substitution rates from 10 HEG/LEG.

lowly expressed genes (HEG and LEG) from two species of chlorarachniophytes/cryptophytes and calculated their amino acid substitution rates between homologous sequences. Average substitution rates of the HEG/LEG between *A. amoebiformis* and *B. natans*, and *Ch. mesostigmatica* and *G. theta*, were found to be 0.39/0.68 and 0.23/0.66, respectively (fig. 3F), suggesting that the sequence conservation of HEG was clearly higher than that of LEG. It has been reported that the highly expressed NmHsp70 and NmHsp90 of *G. theta* are well conserved (Archibald et al. 2001). These results suggest that highly transcribed genes with higher GC content are more conserved in nucleomorph genomes. Therefore, the selective pressures acting on highly expressed genes in nucleomorph genomes may be similar to those in other endosymbiotic/parasitic genomes.

Supplementary Material

Supplementary materials and methods and tables S1–S5 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

The authors thank Dr Goro Tanifuji (University of Tsukuba) for kindly providing the cryptophyte transcriptome data sets. All RNA-seq data sets were generated by the National Center for Genome Resources (NCGR) as a part of Marine Microbial Eukaryote Transcriptome Sequencing Project of Gordon and Betty Moore Foundation. This work was supported by JSPS KAKENHI Grant Number 23117004. Y.H. was a JSPS Postdoctoral Fellowship for Research Abroad. J.M.A. and P.J.K. are Senior Fellows of the Canadian Institute for Advanced Research, Program in Integrated Microbial Biodiversity.

References

Archibald JM. 2007. Nucleomorph genomes: structure, function, origin and evolution. *Bioessays* 29:392–402.

Archibald JM. 2009. The puzzle of plastid evolution. *Curr Biol* 19: R81–R88.

Archibald JM, Cavalier-Smith T, Maier U, Douglas S. 2001. Molecular chaperones encoded by a reduced nucleus: the cryptomonad nucleomorph. *J Mol Evol* 52:490–501.

Chanda I, Pan A, Dutta C. 2005. Proteome composition in *Plasmodium falciparum*: higher usage of GC-rich nonsynonymous codons in highly expressed genes. *J Mol Evol* 61:513–523.

Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y, et al. 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* 492: 59–65.

Dorrell RG, Howe CJ. 2012. What makes a chloroplast? Reconstructing the establishment of photosynthetic symbioses. *J Cell Sci* 125: 1865–1875.

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

Gilson PR. 2001. Nucleomorph genomes: much ado about practically nothing. *Genome Biol* 2:REVIEWS1022.

Gilson PR, Su V, Slamovits CH, Reith ME, Keeling PJ, McFadden GI. 2006. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. *Proc Natl Acad Sci U S A* 103: 9566–9571.

Gould SB, Waller RF, McFadden GI. 2008. Plastid evolution. *Annu Rev Plant Biol* 59:491–517.

Hihara Y, Kamei A, Kanehisa M, Kaplan A, Ikeuchi M. 2001. DNA microarray analysis of cyanobacterial gene expression during acclimation to high light. *Plant Cell* 13:793–806.

Ishida K, Endo H, Koike S. 2011. *Partenskyella glossopodia* (Chlorarachniophyceae) possesses a nucleomorph genome of approximately 1 Mbp. *Phycol Res* 59:120–122.

Keeling PJ. 2010. The endosymbiotic origin, diversification and fate of plastids. *Philos Trans R Soc Lond B Biol Sci* 365:729–748.

Keeling PJ, Slamovits CH. 2005. Causes and effects of nuclear genome reduction. *Curr Opin Genet Dev* 15:601–608.

Lane CE, van den Heuvel K, Kozera C, Curtis BA, Parsons BJ, Bowman S, Archibald JM. 2007. Nucleomorph genome of *Hemiselmis andersenii* reveals complete intron loss and compaction as a driver of protein structure and function. *Proc Natl Acad Sci U S A* 104: 19908–19913.

Li QB, Haskell D, Zhang C, Sung DY, Guy C. 2000. Diurnal regulation of Hsp70s in leaf tissue. *Plant J* 21:373–378.

Maisnier-Patin S, Roth JR, Fredriksson Å, Nyström T, Berg OG, Andersson DI. 2005. Genomic buffering mitigates the effects of deleterious mutations in bacteria. *Nat Genet* 37:1376–1379.

Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M, Kowallik KV. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393:162–165.

McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10:13–26.

Moore CE, Archibald JM. 2009. Nucleomorph genomes. *Annu Rev Genet* 43:251–264.

Moore CE, Curtis B, Mills T, Tanifuji G, Archibald JM. 2012. Nucleomorph genome sequence of the cryptophyte alga *Chroomonas mesostigmatica* CCMP1168 reveals lineage-specific gene loss and genome complexity. *Genome Biol Evol* 4: 1162–1175.

Moran NA. 1996. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A* 93:2873–2878.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5:621–628.

Patron NJ, Rogers MB, Keeling PJ. 2006. Comparative rates of evolution in endosymbiotic nuclear genomes. *BMC Evol Biol* 6:46.

Price DC, Chan CX, Yoon HS, Yang EC, Qiu H, Weber AP, Schwacke R, Gross J, Blouin NA, Lane C, et al. 2012. *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science* 335:843–847.

Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, Bohnert HJ, Philippe H, Lang BF. 2005. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr Biol* 15:1325–1330.

Rossel JB, Wilson IW, Pogson BJ. 2002. Global changes in gene expression in response to high light in *Arabidopsis*. *Plant Physiol* 130: 1109–1120.

Saibil H. 2013. Chaperone machines for protein folding, unfolding and disaggregation. *Nat Rev Mol Cell Biol* 14:630–642.

Schaber J, Rispe C, Wernegreen J, Bunes A, Delmotte F, Silva FJ, Moya A. 2005. Gene expression levels influence amino acid usage and evolutionary rates in endosymbiotic bacteria. *Gene* 352:109–117.

Stork S, Moog D, Przyborski JM, Wilhelm I, Zauner S, Maier UG. 2012. Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of ubiquitin-dependent translocation from degradation. *Eukaryot Cell* 11:1472–1481.

Taipale M, Jarosz DF, Lindquist S. 2010. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat Rev Mol Cell Biol* 11:515–528.

Tanifuji G, Onodera NT, Wheeler TJ, Dlutek M, Donaher N, Archibald JM. 2011. Complete nucleomorph genome sequence of the non-photosynthetic alga *Cryptomonas paramecium* reveals a core nucleomorph gene set. *Genome Biol Evol* 3:44–54.

- Tanifuji G, Onodera NT, Moore CE, Archibald JM. 2014. Reduced nuclear genomes maintain high gene transcription levels. *Mol Biol Evol.* 31: 625–635.
- Tokuriki N, Tawfik DS. 2009. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* 459:668–673.
- von Gromoff ED, Treier U, Beck CF. 1989. Three light-inducible heat shock genes of *Chlamydomonas reinhardtii*. *Mol Cell.* 9:3911–3918.
- Wilcox JL, Dunbar HE, Wolfinger RD, Moran NA. 2003. Consequences of reductive evolution for gene expression in an obligate endosymbiont. *Mol Microbiol.* 48:1491–1500.
- Williams BAP, Slamovits CH, Patron NJ, Fast NM, Keeling PJ. 2005. A high frequency of overlapping gene expression in compacted eukaryotic genomes. *Proc Natl Acad Sci U S A.* 102: 10936–10941.