FOLIA PARASITOLOGICA 52: 8–14, 2005

REVIEW ARTICLE

Comparative genomics of microsporidia

Patrick J. Keeling, **Naomi M. Fast**, **Joyce S. Law**, **Bryony A.P. Williams** and **Claudio H. Slamovits**

Canadian Institute for Advanced Research, Botany Department, University of British Columbia, 3529–6270 University Boulevard, Vancouver, BC, V6T 1Z4, Canada

Key words: Microsporidia, genome compaction, comparative genomics, parasites

Abstract. Microsporidia have been known for some time to possess among the smallest genomes of any eukaryote. There is now a completely sequenced microsporidian genome, as well as several other large-scale sequencing efforts, so the nature of these genomes is becoming apparent. This paper reviews some of the characteristics of microsporidian genomes in general, and some of the recent discoveries made through comparative genomic analyses. In general, microsporidian genomes are both reduced and compacted. Reduction takes place through gene loss, which is understandable in obligate intracellular parasites that rely on their host for many metabolites. Compaction is a more complex process, and is as yet not fully understood. It is clear from genomes surveyed thus far that the remaining genes are tightly packed and that there is little non-coding sequence, resulting in some extraordinary arrangements, including overlapping genes. Compaction also seems to affect certain aspects of genome evolution, like the frequency of rearrangements. The force behind this compaction is not known, and is especially interesting in light of the fact that surveys of genomes that are significantly different in size yield similar complements of protein-coding genes. There are some interesting exceptions, including catalase, photolyase and some mitochondrial proteins, but the rarity of these raises an interesting question as to what accounts for the significant differences seen in the genome sizes among microsporidia.

INTRODUCTION – THE UNUSUALLY SMALL GENOMES OF MICROSPORIDIA

In many ways, eukaryotic and prokaryotic genomes are fundamentally different: at the gross level, eukaryotic genomes are generally organized into multiple linear chromosomes, whereas prokaryotes most often have a single circular chromosome that is divided and segregated by different means. There are several exceptions within prokaryotes (both multiple and linear chromosomes exist in some species), but these are derived conditions. There are also important differences at the level of genome structure and function. In particular, virtually all eukaryotic genes are each expressed on a unique mRNA while prokaryotic genes are often organized as operons which are co-expressed as polycistronic mRNAs. Eukaryotic genomes are also widely perceived to be larger and less gene-dense than their prokaryotic counterparts: whereas prokaryotic genomes are generally tightly packed with genes, genes in the human genome can be separated by intergenic regions larger than some entire prokaryotic genomes! Moreover, eukaryotic genes are themselves often exceedingly large, as they may contain many very large introns which are post-transcriptionally spliced to yield mature mRNA, so even a single human gene can be larger than the smallest prokaryotic genome. These generalisations are true for most eukaryotes, but eukaryotic genome diversity is so poorly understood it can be misleading to extend them beyond the cases that are well studied.

The range of sizes for known eukaryotic genomes is immense, as shown by the examples in Table 1, but what accounts for these differences is not always so obvious. Some organisms clearly contain more genes than others, but the lack of connection between an organism's perceived "complexity" and its genome size (called the C-value paradox) has been puzzled over for some time. For example, as interesting as the dinoflagellates are, *Gonyaulax* is not 30 times more complex than humans despite the fact that its nucleus apparently contains that much more DNA. These differences in genome complexity are likely to arise for many different reasons under different circumstances.

On the other end of the spectrum from dinoflagellates are the microsporidia. Since their first investigation, microsporidian genomes were recognized to be special because they are so small (Biderre et al. 1994, 1995, 1999). There is no saying how big a "typical" eukaryotic genome should be, but it is nonetheless clear that those of many microsporidia are unusually diminutive. Indeed, the microsporidia contain some of the smallest eukaryotic genomes known, some smaller than those of many prokaryotes. Similarly-sized genomes are only found in some picoplankton (Courties et al. 1998) and smaller ones in nucleomorphs, the nuclei of ultra-reduced endosymbiotic red and green algae found in

This paper was presented at the NATO Advanced Research Workshop "Emergent Pathogens in the 21st Century: First United Workshop on Microsporidia from Invertebrate and Vertebrate Hosts", held in České Budějovice, Czech Republic, July 12–15, 2004.

Address for correspondence: P.J. Keeling, Canadian Institute for Advanced Research, Botany Department, University of British Columbia, 3529–6270 University Boulevard, Vancouver, BC, V6T 1Z4, Canada. Phone: ++1 604 822 4906; Fax: ++1 604 822 6089;
E-mail: pkeeling@interchange.ubc.ca

Table 1. Genome sizes of representative eukaryotes.

ORGANISM	GROUP	GENOME SIZE (Mbp)	REFERENCE
Gonyaulax polyedra	Dinoflagellate	98,000	Shuter et al. 1983
Heterocapsa pygmaea	Dinoflagellate	4,450	Triplett et al. 1993
Toxoplasma gondii	Apicomplexan	87	Blaxter and Ivens 1999
Plasmodium falciparum	Apicomplexan	23	Gardner et al. 2002)
Cryptosporidium parvum	Apicomplexan	9	Spano and Crisanti 2000
Paramecium caudatum	Ciliate	8,600	Shuter et al. 1983
Thalassiosira pseudonana	Diatom	32	Armbrust et al. 2004
Coscinodiscus asteromphalus	Diatom	25,000	Shuter et al. 1983
Amoeba proteus	Amoeba	290,000	Friz 1968
Amoeba dubia	Amoeba	670,000	Friz 1968
Dictyostelium discoideum	Slime Mold	34	Glockner et al. 2002
Entamoeba histolytica	Archamoeba	<20	http://www.sanger.ac.uk/ Projects/E-histolytica/
Trichomonas vaginalis	Parabasalian	60–80	http://www.tigr.org/ tdb/e2k1/tvg/
Trypanosoma	Kinetoplastid	39	El-Sayed et al. 2000
Leishmania major	Kinetoplastid	33	Myler et al. 2000
Cyanidioschyzon merolae	Red Alga	16	Matsuzaki et al. 2004
Guillardia theta (nucleomorph)	Red Alga	0.55	Douglas et al. 2001
Chlamydomonas reinhardtii	Green Alga	100	Harris 1993
Ostreococcus tauri	Green Alga (picoeukaryote)	10	Courties et al. 1998
Bigelowiella natans (nucleomorph)	Green Alga	0.38	Gilson and McFadden 2002
Oryza sativa	Plant	466	Yu et al. 2002
Zea mays	Plant	3,000	Arumuganathan and Earle 1991
Arabidopsis thaliana	Plant	125	Arabidopsis Genome Iniciative 2000
Mouse	Animal	2,500	Waterston et al. 2002
Human	Animal	2,900	Waterston et al. 2002
Fugu rubripes	Animal	365	Aparicio et al. 2002
Drosophila melanogaster	Animal	137	Adams et al. 2000
Ciona intestinalis	Animal	156	Dehal et al. 2002
Saccharomyces cerevisiae	Fungus	12	Blandin et al. 2000
Cryptococcus neoformans	Fungus	20	Wickes et al. 1994
Neurospora crassa	Fungus	43	Schulte et al. 2002
Encephalitozoon intestinalis	Microsporidian	2.3	Peyretaillade et al. 1998
Encephalitozoon cuniculi	Microsporidian	2.9	Katinka et al. 2001
Antonospora locustae	Microsporidian	5.4	Streett 1994
Spraguea lophii	Microsporidian	6.2	Biderre et al. 1994
Glugea antherinae	Microsporidian	19.5	Biderre et al. 1994

cryptomonads (Douglas et al. 2001) and chlorarachniophytes (Gilson and McFadden 2002), respectively. A range of microsporidian genome sizes is also given in Table 1, most of which are estimates based on pulsed-field gel electrophoresis karyotyping. There is some variability in genome sizes within certain species (e.g., *Encephalitozoon cuniculi* and *E. hellem*) (Biderre et al. 1999, Delarbre et al. 2001), but these values likely represent reasonable estimates. The largest microsporidian genomes are just under 20 million base pairs (Mbp) (Biderre et al. 1994), which is not particularly unusual. The smaller, however, are between 2 and 3 Mbp, the smallest being about half the size of the *Escherichia coli* genome. These smaller genomes have been the focus of most of the attention in microsporidian genomics, and the nature of these genomes and the implications of the range of sizes will be discussed in turn.

GENOMIC RESOURCES FROM MICROSPORIDIA

Microsporidian genomes have attracted interest due to a combination of factors, including their importance as vertebrate and invertebrate pathogens, their controversial evolutionary history, as well as their small size. Consequently, the genomic resources for microsporidia have steadily increased in the last few years, and will continue to do so. Presently there is a complete genome sequence for *Encephalitozoon cuniculi* (genoscope.cns.fr/externe/English/Projets/Projet_AD/AD.html) (Katin-

ka et al. 2001), a partial genome sequence for *Antonospora locustae* (formerly *Nosema locustae*) (jbpc.mbl.edu/Nosema/index.html), genome sequence surveys (GSS) from *Spraguea lophii* (jbpc.mbl.edu/SpragueaHTML/) (Hinkle et al. 1997), *Vittaforma corneae* (Mittleider et al. 2002), and *Antonospora locustae* (botany.ubc.ca/keeling/AntonosporaGSS.html) (Slamovits et al. 2004a), and an expressed sequence tag (EST) survey for *A. locustae* (amobidia.bcm.umontreal.ca/pepdb/pep.php) (unpublished data).

WHAT MAKES A GENOME SMALL?

Microsporidian genomes are small, but what does that mean? There are two basic ways to make a genome small: (1) by reduction, or loss of genes, and (2) by compacting what genes remain into a smaller and smaller space (Keeling 2004). Microsporidia have done both.

In terms of gene loss, the *E. cuniculi* genome tells a fairly straightforward story (Katinka et al. 2001). It contains relatively few protein-coding genes (1,997 identified), suggestive of massive gene loss due to host dependence. Indeed, this is supported by the non-random nature of the genes that have been lost. Missing are genes for proteins involved in pathways for the biosynthesis of small compounds like nucleotides, fatty acids, and most amino acids. Conversely, complete or near-complete sets of genes for a variety of other processes, structures or pathways like DNA replication, ribosomes, spliceosomes, glycolysis and others are found. These features are not really surprising; as they are intracellular parasites, we expect microsporidia to be highly reliant on their hosts for energy and many metabolites, so the absence or reduction of genes for proteins involved in these processes is expected (Katinka et al. 2001). Genes for proteins involved in basic housekeeping tasks like expression and replication naturally cannot be lost and we therefore expect to find similar complements of these in various microsporidian genomes. Some of the potentially interesting areas for differential gene loss and retention will be in processes like metabolism, regulation of expression, stress responses, infection and host interactions. These last two areas are particularly interesting, and will likely account for some proportion of the "unidentifiable" genes in a given microsporidian genome. These genes are difficult to study, but determining their function will likely be very important.

In terms of compaction, the *E. cuniculi* genome and the sequence survey of *A. locustae* tell similar stories, but this is somewhat more difficult to explain. *E. cuniculi* chromosomes are made up of gene-rich cores flanked on both ends by rRNA operons, non-coding subtelomeric regions and telomers (Katinka et al. 2001). The gene-rich cores are composed of protein-coding genes in very high density (about 0.97 genes/Kbp). Genes are separated by short intergenic regions (average of 129 bp), have few introns, and the genes themselves are on average 15% shorter than homologues in yeast. The *A. locustae* regions that have been sequenced are quite similar, the gene density is about 0.94 genes/Kbp and the average intergenic region is about 200 bp, with gene lengths similar to *E. cuniculi* homologues (Slamovits et al. 2004a). Why are microsporidian genomes compacted? This question is not so easy to answer compared to considering why the gene complement has been reduced. It is possible that genome compaction gives a competitive advantage to parasites infecting the same host (although this seems like it would provide only a minor advantage), or it is possible that there are biases in favour of deletions in replication and/or recombination systems that simply ratchet the genome size down for no particular reason. Whether one of these or some other reason provides an explanation, it is clear that the force that led to this condition is relatively strong, since the gene density of these two microsporidian genomes is considerably higher than any other well-sampled eukaryote. For instance, the yeast genome, which is considered to be relatively compact, has a gene density of about half that of *E. cuniculi* or *A. locustae*.

GENOME STRUCTURE CONSERVATION

Another characteristic found to be similar between the *E. cuniculi* and *A. locustae* genomes is the relative order of genes in the genome, or synteny. The order of genes in a eukaryotic genome can be maintained by a variety of forces, but these tend to be weak forces so that genes drift apart relatively quickly, mostly by small inversions, but also by large inversions, transpositions, and the breakup of chromosomes. Through time, the gene order is randomized.

The first data on comparative gene order came from closely related members of the genus *Encephalitozoon*, where the relative orientations of polar tube protein genes was found to be conserved (Delbac et al. 2001). Some degree of conservation is expected among such close relatives, but as one looks at more distantly related species the processes described above should lead to complete randomisation of the genomes. Indeed, when the genomes of *E. cuniculi* and *A. locustae* were compared much of the genome was organized differently, but a significant fraction of the known genes were retained in the same orientation, or close by. Comparing the relative positions of 94 gene pairs, 13% were found to be shared by both genomes, while 30% of gene pairs were found within 10 genes of one another (Slamovits et al. 2004a). Without knowing the length of time between the present and the common ancestor of two organisms, it is impossible to know how much synteny to expect to have been retained, so there is no concrete answer to the question of whether this is more synteny than we should expect in other genomes. However, there are indirect reasons, based on comparisons with other organisms, to suspect these genomes are evolving

relatively slowly. We know that *A. locustae* and *E. cuniculi* represent a relatively deep divergence within the microsporidia (Slamovits et al. 2004b), so any conservation between them goes back to near the diversification of extant microsporidian species. If we examine the degree of synteny shared between different species of ascomycete fungi, we find that there is no conservation whatsoever between *Saccharomyces* and *Schizosaccharomyces* (which represent the entire range of diversity of ascomycetes) and there is about 9% pairwise conservation between *Saccharomyces cerevisiae* and *Candida albicans* (two relatively closely related members of the Saccharomycetales) (Keogh et al. 1998, Seoighe et al. 2000). The conservation of genome order between *A. locustae* and *E. cuniculi* is about the same as that between *S. cerevisiae* and *C. albicans* (it is about 1.5 times higher, but to be conservative, we will treat them as about equal). Either the genome diverged at about the same time, or they are evolving at different rates.

Saccharomyces and *Candida* are estimated to have diverged about 200 million years ago (Berbee and Taylor 2001), which is about the time that marsupials diverged from placental mammals. For the evolutionary rates of microsporidian genomes to be about the same as these fungal genomes, microsporidia would have to have evolved at about the same time. For this to be true, the emerging parasites would have to have evolved a mechanism to infect some animal species, and then quickly spread through the entire Kingdom. This would be a remarkable colonisation of a hostile habitat, and does not likely account for the conservation of microsporidian genomes (Slamovits et al. 2004a).

A more likely explanation is that the genomes of *A. locustae* and *E. cuniculi* did not diverge at the same time as those of *S. cerevisiae* and *C. albicans*, but that they are evolving at different rates. When the nature of microsporidian genomes is considered, it is not difficult to see why this could happen. It has been shown that genome order is affected by gene density in fungal systems, but this was shown to be a relatively minor force in preserving the overall structure in these genomes (Hurst et al. 2002). However, the much higher degree of compaction in microsporidian genomes could elevate the importance of this force by making it increasingly difficult to make the breakpoints needed to shuffle genes without serious deleterious effects. Indeed, the correlation between intergenic distances and conservation of gene order that is expected if compaction is reducing genome flexibility (Hurst et al. 2002) is found in both *E. cuniculi* and *A. locustae* genomes (Slamovits et al. 2004a). It is likely that other forces are also operating to preserve the gene order of these genomes, but for now we can only suggest that their extreme nature may be showing us a normally insignificant force of genome evolution in its most exaggerated state.

LIMITED VARIATION IN GENE CONTENT

The *E. cuniculi* genome is among the smallest of microsporidia, and the *Encephalitozoon* group in general has particularly small genomes. In contrast, the *A. locustae* genome is estimated to be 5.4 Mbp, or about 1.86 times that of *E. cuniculi*. Since the gene densities appear to be about the same, we would expect to find near one out of every two genes in *A. locustae* to be absent from *E. cuniculi*, but this does not seem to be the case. Of the 138 genes presently reported in *A. locustae*, 130 are also found in *E. cuniculi*. Five are not detectably related to any other gene, and three are found in other organisms but not *E. cuniculi*. These differences raise some questions of general interest, and some of the genes also raise specific questions.

First, how can the *A. locustae* genome be 1.85 times larger than that of *E. cuniculi* if the gene density is about the same and the gene complement is similar? Assuming the current sampling is about random, and also assuming there are not large numbers of genes in *E. cuniculi* that *A. locustae* lacks (these are impossible to identify without the entire *A. locustae* genome, but there must be several such genes and they would make this discrepancy more pronounced), then there are several possible explanations. One trivial explanation is that the estimated size of the *A. locustae* genome is wrong and that it is closer to 3 Mbp. Alternatively, there may be large gene-poor regions of the genome that have not been sampled. The region around catalase (Fast et al. 2003) could represent one such area since this 13,000 bp region only contains three genes (for a density of only 0.23 genes/Kbp – over four times lower than the average found in gene-rich areas). Lastly, it is also possible that most of the genome shares a high gene-density, but that more than one copy of some genes exist. These each have interesting implications. If the genome has islands of vastly different density then why are some regions under such strong compacting selection while others are not? If, alternatively, most of the genome has the same gene density, then why are many extra copies of some genes retained when the genome is apparently under selection to reduce its size. Overall it raises the question of whether compaction is not entirely connected to reduction in size.

A few of the genes that *A. locustae* does contain are worth noting, as they are an indication of some of the things we might hope to find with more extensive comparative genomics in microsporidia. First, the microsporidian mitochondrion has been a source of interest for both evolutionary and cell biology reasons because (1) it was hypothesized not to exist (Cavalier-Smith 1983), and (2) it is now recognized a highly reduced, cryptic organelle that has lost many of the primary functions of typical mitochondria (Williams and Keeling 2003). The complement of mitochondrial proteins in *E. cuniculi* has been carefully examined and the functions of the organelle hypothesized based on this list of genes

(Katinka et al. 2001). *A. locustae* has been shown to have most of these genes, but also has at least two others that are lacking in *E. cuniculi* (Williams and Keeling 2005). One of these is a transporter homologous to a yeast mitochondrial transporter, and most similar to one that specifically transports pyruvate. This is an interesting possibility since pyruvate dehydrogenase is present in both *E. cuniculi* and *A. locustae* (Fast and Keeling 2001, Katinka et al. 2001), but its function is not known. The second gene is homologous to mitochondrial inner membrane protease. This protein is part of a small complex which cleaves the signal peptide from proteins that use the endosymbiont secretion system to target mitochondrial proteins to the inner membrane or intermembrane space. Neither this protein, other components of the complex, or any known target of the complex are known from *E. cuniculi*. Its presence in *A. locustae* suggests that other mitochondrial proteins are likely encoded in the *A. locustae* genome, and other unrecognized mitochondrial metabolic functions will be found. This underscores the value of comparative genomics in expanding our model of what a "typical" microsporidian looks like: in many ways *E. cuniculi* may be unusual (it does have one of the smallest genomes known), so more data will be helpful in generating a well-informed model for predicting what to expect in other microsporidia.

Two other genes of interest that are not found in *E. cuniculi* are related to coping with environmental stress. These are catalase and photolyase. Catalase converts hydrogen peroxide to water and hydrogen gas, and is a marker enzyme for peroxisomes, although it is also found in many prokaryotes in eukaryotes outside of the organelle. *A. locustae* has been shown to encode a catalase of the non-peroxisomal, group II variety that is functional and expressed in spores (Fast et al. 2003). The gene is likely not ancestral to microsporidia, since it is not of fungal ancestry, but is instead derived from a proteobacterium by lateral gene transfer. Photolyase is also involved in protection from the environment, in this case DNA repair. This enzyme absorbs a photon of visible light and uses the energy to repair UV-generated lesions, in the case of the *A. locustae* photolyase, cyclobutane pyrimidine dimers (Slamovits and Keeling 2004). Once again, the gene is expressed in spores and functional (it was shown to complement an *E. coli* mutant) and this class of enzyme (class II) has not been found in fungi to date. The phylogeny of photolyase is not as robust as that of catalase, but the *A. locustae* enzyme appears to be most related to animal homologues, so its history in microsporidia will also be interesting to determine. Together these genes reinforce the importance of comparative genomics in building a model for the "typical" microsporidia.

FUTURE PROSPECTS

Given the prevalence and importance of microsporidia as parasites of humans and commercially, environmentally, and medically important animals, together with their very small genome sizes, it is inevitable that additional genome projects will soon be underway. Each genome project of a parasite has the potential to generate spectacular new insights into the parasite's function at the molecular level and to transform the way we investigate its molecular biology. At the same time, however, each new microsporidian investigated at the genomic level will also add to a body of data on the forces that shape these unusual genomes and this will provide information on genome dynamics with importance beyond the microsporidia. It remains to be seen how far the conservation of genome order observed between *E. cuniculi* and *A. locustae* will extend to other groups, although the limited data from *Spraguea* (Hinkle et al. 1997) suggests it will be extended. This is not only of interest to genome evolution, but also provides a predictive power that may be experimentally useful since it gives one more way to search for a gene of interest in the genome. It also remains to be seen whether there are certain genes or regions of the genome that are especially prone to conservation and why this may be, or if it is simply a general process that speeds up and slows down around the genome at random. It will also be of considerable interest to examine the differential retention and loss of genes between species: the *A. locustae* genome survey represents only about 10% of its genome but has revealed a number of interesting differences with *E. cuniculi*. The rest of this genome and data from many others will provide a better-informed model of "typical" microsporidian metabolism and molecular biology.

Acknowledgements. This manuscript represents work presented at a NATO Advanced Research Workshop in České Budějovice in July 2004. We thank the organisation committee of that workshop for their organisation and for travel support. We also thank the Burroughs-Wellcome Fund, the Canadian Institutes for Health Research, and Genome Canada for support of microsporidian research in the lab of PJK. PJK is a Scholar of the Canadian Institutes for Advanced Research and a new investigator of the CIHR and the Michael Smith Foundation for Health Research.

REFERENCES

ADAMS M.D., CELNIKER S.E., HOLT R.A., EVANS C.A., GOCAYNE J.D., AMANATIDES P.G., SCHERER S.E., LI P.W., HOSKINS R.A., GALLE R.F., GEORGE R.A., LEWIS S.E., RICHARDS S., ASHBURNER M., HENDERSON S.N., SUTTON G.G., WORTMAN J.R., YANDELL M.D., ZHANG Q., CHEN L.X., BRANDON R.C., ROGERS Y.H., BLAZEJ R.G., CHAMPE M., PFEIFFER B.D., et al. 2000: The genome sequence of *Drosophila melanogaster*. Science 287: 2185–2195.

ARMBRUST E.V., BERGES J.A., BOWLER C., GREEN B.R., MARTINEZ D., PUTNAM N.H., ZHOU S., ALLEN A.E., APT K.E., BECHNER M., BRZEZINSKI M.A., CHAAL B.K., CHIOVITTI A., DAVIS A.K., DEMAREST M.S., DETTER J.C., GLAVINA T., GOODSTEIN D., HADI M.Z., HELLSTEN U., et al. 2004: The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306: 79–86.

APARICIO S., CHAPMAN J., STUPKA E., PUTNAM N., CHIA J.M., DEHAL P., CHRISTOFFELS A., RASH S., HOON S., SMIT A., GELPKE M.D., ROACH J., OH T., HO I.Y., WONG M., DETTER C., VERHOEF F., PREDKI P., TAY A., LUCAS S., RICHARDSON P., SMITH S.F., CLARK M.S., EDWARDS Y.J., DOGGETT N., et al. 2002: Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297: 1301–1310.

ARABIDOPSIS GENOME INITIATIVE 2000: Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796–815.

ARUMUGANATHAN K., EARLE E.D. 1991: Nuclear DNA content of some important plant species. Plant. Mol. Biol. Rep. 9: 208–218.

BERBEE M.L., TAYLOR J.W. 2001: Fungal molecular evolution: gene trees and geologic time. In: D.J. Mclaughlin, E.G. Mclaughlin and P.A. Lemke (Eds.), The Mycota: a Comprehensive Treatise on Fungi as Experimental Systems for Basic and Applied Research. Volume VII: Systematics and Evolution, Part B. Springer-Verlag, Berlin, Heidelberg, pp. 229–245.

BIDERRE C., MATHIS A., DEPLAZES P., WEBER R., MÉTÉNIER G., VIVARÈS C.P. 1999: Molecular karyotype diversity in the microsporidian *Encephalitozoon cuniculi*. Parasitology 118: 439–445.

BIDERRE C., PAGÈS M., MÉTÉNIER G., CANNING E.U., VIVARÈS C.P. 1995: Evidence for the smallest nuclear genome (2.9 Mb) in the microsporidium *Encephalitozoon cuniculi*. Mol. Biochem. Parasitol. 74: 229–231.

BIDERRE C., PAGES M., MÉTÉNIER G., DAVID D., BATA J., PRENSIER G., VIVARÈS C.P. 1994: On small genomes in eukaryotic organisms: molecular karyotypes of two microsporidian species (Protozoa) parasites of vertebrates. C. R. Acad. Sci. III 317: 399–404.

BLANDIN G., DURRENS P., TEKAIA F., AIGLE M., BOLOTIN-FUKUHARA M., BON E., CASAREGOLA S., DE MONTIGNY J., GAILLARDIN C., LEPINGLE A., LLORENTE B., MALPERTUY A., NEUVEGLISE C., OZIER-KALOGEROPOULOS O., PERRIN A., POTIER S., SOUCIET J., TALLA E., TOFFANO-NIOCHE C., WESOLOWSKI-LOUVEL M., MARCK C., DUJON B. 2000: Genomic exploration of the hemiascomycetous yeasts: 4. The genome of *Saccharomyces cerevisiae* revisited. FEBS Lett. 487: 31–36.

BLAXTER M., IVENS A. 1999: Reports from the cutting edge of parasitic genome analysis. Parasitol. Today 15: 430–431.

CAVALIER-SMITH T. 1983: A 6-kingdom classification and a unified phylogeny. In: H.E.A. Schenk and W. Schwemler (Eds.), Endocytobiology Vol. II. Walter de Gruyter & co., Berlin, New York, pp. 1027–1034.

COURTIES C., PERASSO R., CHRETIENNOT-DINET M.J., GOUY M., GUILLOU L., TROUSSELLIER M. 1998: Phylogenetic analysis and genome size of *Ostreococcus tauri* (Chlorophyta, Prasinophyceae). J. Phycol. 34: 844–849.

DEHAL P., SATOU Y., CAMPBELL R.K., CHAPMAN J., DEGNAN B., DE TOMASO A., DAVIDSON B., DI GREGORIO A., GELPKE M., GOODSTEIN D.M., HARAFUJI N., HASTINGS K.E., HO I., HOTTA K., HUANG W., KAWASHIMA T., LEMAIRE P., MARTINEZ D., MEINERTZHAGEN I.A., NECULA S., NONAKA M., PUTNAM N., RASH S., SAIGA H., SATAKE M., et al. 2002: The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science 298: 2157–2167.

DELARBRE S., GATTI S., SCAGLIA M., DRANCOUR M. 2001: Genetic diversity in the microsporidian *Encephalitozoon hellem* demonstrated by pulsed-field gel electrophoresis. J. Eukaryot. Microbiol. 48: 471–474.

DELBAC F., PEUVEL I., METENIER G., PEYRETAILLADE E., VIVARES C.P. 2001: Microsporidian invasion apparatus: identification of a novel polar tube protein and evidence for clustering of *ptp1* and *ptp2* genes in three *Encephalitozoon* species. Infect. Immun. 69: 1016–1024.

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.

El-SAYED N.M., HEGDE P., QUACKENBUSH J., MELVILLE S.E., DONELSON J.E. 2000: The African trypanosome genome. Int. J. Parasitol. 30: 329–345.

FAST N.M., KEELING P.J. 2001: Alpha and beta subunits of pyruvate dehydrogenase E1 from the microsporidian *Nosema locustae*: mitochondrion-derived carbon metabolism in microsporidia. Mol. Biochem. Parasitol. 117: 201–209.

FAST N.M., LAW J.S., WILLIAMS B.A., KEELING P.J. 2003: Bacterial catalase in the microsporidian *Nosema locustae*: implications for microsporidian metabolism and genome evolution. Eukaryot. Cell 2: 1069–1075.

FRIZ C.T. 1968: The biochemical composition of the free-living amoebae *Chaos chaos*, *Amoeba dubia* and *Amoeba proteus*. Comp. Biochem. Physiol. B 26: 81–90.

GARDNER M.J., SHALLOM S.J., CARLTON J.M., SALZBERG S.L., NENE V., SHOAIBI A., CIECKO A., LYNN J., RIZZO M., WEAVER B., JARRAHI B., BRENNER M., PARVIZI B., TALLON L., MOAZZEZ A., GRANGER D., FUJII C., HANSEN C., PEDERSON J., FELDBLYUM T., PETERSON J., SUH B., ANGIUOLI S., PERTEA M., ALLEN J., et al. 2002: Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11 and 14. Nature 419: 531–534.

GILSON P.R., McFADDEN G.I. 2002: Jam packed genomes – a preliminary, comparative analysis of nucleomorphs. Genetica 115: 13–28.

GLOCKNER G., EICHINGER L., SZAFRANSKI K., PACHEBAT J.A., BANKIER A.T., DEAR P.H., LEHMANN R., BAUMGART C., PARRA G., ABRIL J.F., GUIGO R., KUMPF K., TUNGGAL B., COX E., QUAIL M.A., PLATZER M., ROSENTHAL A., NOEGEL A.A.

2002: Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. Nature 418: 79–85.

HARRIS E.H. 1993: *Chlamydomonas reinhardtii*. In: S.J. O'Brien (Ed.), Genetic Maps: a Compilation of Linkage and Restriction Maps of Genetically Studied Organisms. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp. 2.156–2.169.

HINKLE G., MORRISON H.G., SOGIN M.L. 1997: Genes coding for reverse transcriptase, DNA-directed RNA polymerase, and chitin synthase from the microsporidian *Spraguea lophii*. Biol. Bull. 193: 250–251.

HURST L.D., WILLIAMS E.J., PAL C. 2002: Natural selection promotes the conservation of linkage of co-expressed genes. Trends Genet. 18: 604–606.

KATINKA M.D., DUPRAT S., CORNILLOT E., MÉTÉNIER G., THOMARAT F., PRENIER G., BARBE V., PEYRETAILLADE E., BROTTIER P., WINCKER P., DELBAC F., El ALAOUI H., PEYRET P., SAURIN W., GOUY M., WEISSENBACH J., VIVARÈS C.P. 2001: Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 414: 450–453.

KEELING P.J. 2004: Reduction and compaction in the genome of the apicomplexan parasite, *Cryptosporidium parvum*. Develop. Cell 6: 614–616.

KEOGH R.S., SEOIGHE C., WOLFE K.H. 1998: Evolution of gene order and chromosome number in *Saccharomyces*, *Kluyveromyces* and related fungi. Yeast 14: 443–457.

MATSUZAKI M., MISUMI O., SHIN I.T., MARUYAMA S., TAKAHARA M., MIYAGISHIMA S.Y., MORI T., NISHIDA K., YAGISAWA F., YOSHIDA Y., NISHIMURA Y., NAKAO S., KOBAYASHI T., MOMOYAMA Y., HIGASHIYAMA T., MINODA A., SANO M., NOMOTO H., OISHI K., HAYASHI H., OHTA F., NISHIZAKA S., HAGA S., MIURA S., MORISHITA T., et al. 2004: Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature 428: 653–657.

MITTLEIDER D., GREEN L.C., MANN V.H., MICHAEL S.F., DIDIER E.S., BRINDLEY P.J. 2002: Sequence survey of the genome of the opportunistic microsporidian pathogen, *Vittaforma corneae*. J. Eukaryot. Microbiol. 49: 393–401.

MYLER P.J., SISK E., McDONAGH P.D., MARTINEZ-CALVILLO S., SCHNAUFER A., SUNKIN S.M., YAN S., MADHUBALA R., IVENS A., STUART K. 2000: Genomic organization and gene function in *Leishmania*. Biochem. Soc. Trans. 28: 527–531.

PEYRETAILLADE E., BIDERRE C., PEYRET P., DUFFIEUX F., MÉTÉNIER G., GOUY M., MICHOT B., VIVARÈS C.P. 1998: Microsporidian *Encephalitozoon cuniculi*, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSU rRNA reduced to the universal core. Nucleic Acids Res. 26: 3513–3520.

SCHULTE U., BECKER I., MEWES H.W., MANNHAUPT G. 2002: Large scale analysis of sequences from *Neurospora crassa*. J. Biotechnol. 94: 3–13.

SEOIGHE C., FEDERSPIEL N., JONES T., HANSEN N., BIVOLAROVIC V., SURZYCKI R., TAMSE R., KOMP C., HUIZAR L., DAVIS R.W., SCHERER S., TAIT E., SHAW D.J., HARRIS D., MURPHY L., OLIVER K., TAYLOR K., RAJANDREAM M.A., BARRELL B.G., WOLFE K.H. 2000: Prevalence of small inversions in yeast gene order evolution. Proc. Natl. Acad. Sci. USA 97: 14433–14437.

SHUTER B.J., THOMAS J.E., TAYLOR W.D., ZIMMERMAN A.M. 1983: Phenotypic correlates of genomic DNA content in unicellular eukaryotes and other cells. Am. Nat. 122: 26–44.

SLAMOVITS C.H., FAST N.M., LAW J.S., KEELING P.J. 2004a: Genome compaction and stability in microsporidian intracellular parasites. Curr. Biol. 14: 891–896.

SLAMOVITS C.H., KEELING P.J. 2004: Class II photolyase in a microsporidian intracellular parasite. J. Mol. Biol. 341: 713–721.

SLAMOVITS C.H., WILLIAMS B.A., KEELING P.J. 2004b: Transfer of *Nosema locustae* (Microsporidia) to *Antonospora locustae* n. comb. based on molecular and ultrastructural data. J. Eukaryot. Microbiol. 51: 207–213.

SPANO F., CRISANTI A. 2000: *Cryptosporidium parvum*: the many secrets of a small genome. Int. J. Parasitol. 30: 553–565.

STREETT D.A. 1994: Analysis of *Nosema locustae* (Microsporidia: Nosematidae) chromosomal DNA with pulsed-field gel electrophoresis. J. Invertebr. Pathol. 63: 301–303.

TRIPLETT E.L., GOVIND N.S., ROMAN S.J., JOVINE R.V.M., PREZELIN B.B. 1993: Characterization of the sequence organization of DNA from the dinoflagellate *Heterocapsa pygmaea* (*Glenodinium* sp.). Mol. Mar. Biol. Biotechnol. 2: 239–245.

WATERSTON R.H., LINDBLAD-TOH K., BIRNEY E., ROGERS J., ABRIL J.F., AGARWAL P., AGARWALA R., AINSCOUGH R., ALEXANDERSSON M., AN P., ANTONARAKIS S.E., ATTWOOD J., BAERTSCH R., BAILEY J., BARLOW K., BECK S., BERRY E., BIRREN B., BLOOM T., BORK P., BOTCHERBY M., BRAY N., BRENT M.R., BROWN D.G., BROWN S.D., et al. 2002: Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520–562.

WICKES B.L., MOORE T.D., KWON-CHUNG K.J. 1994: Comparison of the electrophoretic karyotypes and chromosomal location of ten genes in the two varieties of *Cryptococcus neoformans*. Microbiology 140: 543–550.

WILLIAMS B.A.P., KEELING P.J. 2003: Cryptic organelles in parasitic protists and fungi. Adv. Parasitol. 54: 9–67.

WILLIAMS B.A.P., KEELING P.J. 2005: Microsporidian mitochondrial proteins: expression in *Antonospora locustae* spores and identification of two novel genes. J. Eukaryot. Microbiol. (In press.)

YU J., HU S., WANG J., WONG G.K., LI S., LIU B., DENG Y., DAI L., ZHOU Y., ZHANG X., CAO M., LIU J., SUN J., TANG J., CHEN Y., HUANG X., LIN W., YE C., TONG W., CONG L., GENG J., HAN Y., LI L., LI W., HU G., et al. 2002: A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296: 79–92.

Received 23 August 2004

Accepted 22 December 2004