doi:10.1016/j.jmb.2004.06.032 *J. Mol. Biol.* (2004) **341**, 713–721

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Class II Photolyase in a Microsporidian Intracellular Parasite

Claudio H. Slamovits and Patrick J. Keeling*

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

Photoreactivation is the repair of DNA damage induced by ultraviolet light radiation using the energy contained in visible-light photons. The process is carried out by a single enzyme, photolyase, which is part of a large and ancient photolyase/cryptochrome gene family. We have characterised a photolyase gene from the microsporidian parasite, *Antonospora locustae* (formerly *Nosema locustae*) and show that it encodes a functional photoreactivating enzyme and is expressed in the infectious spore stage of the parasite's life cycle. Sequence and phylogenetic analyses show that it belongs to the class II subfamily of cyclobutane pyrimidine dimer repair enzymes. No photolyase is present in the complete genome sequence of the distantly related microsporidian, *Encephalitozoon cuniculi*, and this class of photolyase has never yet been described in fungi, the closest relatives of Microsporidia, raising questions about the evolutionary origin of this enzyme. This is the second environmental stress enzyme to be found in *A. locustae* but absent in *E. cuniculi*, and in the other case (catalase), the gene is derived by lateral transfer from a bacterium. It appears that *A. locustae* spores deal with environmental stress differently from *E. cuniculi*, these results lead to the prediction that they are more robust to environmental damage.

Keywords: photolyase; cryptochrome; Microsporidia; DNA repair; lateral gene transfer

*Corresponding author

Introduction

Ultraviolet light (UV) induces DNA damage that, if unrepaired, may result in mutations and cell death. Two kinds of UV-induced damage, cyclobutane pyrimidine dimer (CPD) and pyrimidine (6–4) pyrimidone photoproducts, or (6–4) photoproducts are the substrates for the light-dependent repairing activity of photoreactivating (PHR) enzymes (or photolyases). Photolyases are flavoproteins whose mechanisms of action have been studied in detail,[1,2] and the crystal structure of the *Escherichia coli* PHR enzyme has been elucidated.[3] Briefly, flavin–adenine dinucleotide (FAD) in its reduced state donates an electron that splits the UV-induced nucleotide–nucleotide dimer. A second chromophore site (or antenna site) that contains MHTF or 8-HTF captures a blue-light photon and transfers an electron back to the catalytic site, restoring the original excitation state of FAD. All photolyases are evolutionarily related, belonging to a large and evolutionarily complex gene family referred to as the Photolyase-cryptochrome family.[4] This family has achieved a surprisingly high degree of functional diversity, encompassing functions as different as DNA repair, physiological and developmental responses to light in plants,[5] and control of circadian rhythms in animals.[6] It is believed that the ancestor of the family was a photolyase, since photoreactivation is the most common function performed by the members of the family,[4] and because photoreactivation must have played a crucial role during early evolution when harmful solar radiation was significantly higher than it is today.[7]

There are two functional classes of photolyase that repair (6–4) and CPD photoproducts, respectively. (6–4) Photolyases are found only in animals

Supplementary data associated with this article can be found at doi: 10.1016/j.jmb.2004.06.032

Abbreviations used: CPD, cyclobutane pyrimidine dimer; CIP, calf intestinal phosphatase; TAP, tobacco acid pyrophosphatase; UV, ultraviolet light; PHR, photoreactivating; ORF, open reading frame; CFU, colony-forming units.

E-mail address of the corresponding author: pkeeling@interchange.ubc.ca

and are closely related to animal cryptochromes, which are involved in the entrainment of circadian rhythms in insects and mammals.[8] CPD photolyases, on the other hand, are the most abundant and occur in all three domains of life.[1] CPD photolyases can also be subdivided into two classes differentiated by sequence similarity.[1,4] class I CPD photolyases (CPD-I) are mostly found in bacteria and archaea. In eukaryotes, they are found only in fungi. class II CPD photolyases (CPD-II) are the most divergent group of the family and are present in some bacteria,[9] archaea,[10] plants,[11] green algae,[12] dipterans,[13] and vertebrates.[14] Interestingly, this class of photolyase is also present in the genomes of several poxviruses from vertebrates and insects.[15] In fungi, photoreactivation is performed by CPD-I,[16] which is consistent with the absence of CPD-II in any fungus studied to date.

Microsporidia comprises a diverse group of highly derived, obligate intracellular parasitic eukaryotes that parasitize all animal phyla, and 13 species are known to infect humans.[17] Many characteristics of this enigmatic group appear primitive, which led investigators to think that they were among the first eukaryotes.[18] However, it is now known that they are not primitive, but rather extremely specialized and highly reduced relatives of fungi.[19,20] The severe reduction that typifies Microsporidia is exemplified by their genomes, which are among the smallest known in eukaryotes. The genome of the human parasite *Encephalitozoon cuniculi* has been completely sequenced and showed a compact and highly reduced complement of genes.[21] The proteome of *E. cuniculi* includes proteins devoted to repair of DNA, but consistently with its intracellular way of life, no photolyase gene was found in the genome.

Genomic surveys of the comparatively distantly related microsporidian, *Antonospora locustae* (formerly known as *Nosema locustae*; this species has recently been renamed either *Paranosema locustae*[22] or *A. locustae*[23]), have revealed a considerable degree of conservation in both gene content and genome structure compared to *E. cuniculi*, despite their high sequence divergence.[24] Nevertheless, intriguing differences in gene content have been found. Fast *et al.* reported the acquisition of a bacterial catalase gene by *A. locustae*, suggesting that these parasites can integrate foreign genes that appear to provide the spore with additional protection against environmental stress. Here we describe the isolation of an *A. locustae* class II CPD photolyase gene that is absent from the *E. cuniculi* genome and not orthologous to the class I photolyases known from fungi. The gene is expressed in *A. locustae* spores, and confers photoreactivation to a photolyase-deficient strain of *E. coli*. This is the first report of a class II CPD photolyase in an organism of fungal affinity and raises interesting questions about the possible roles of lateral gene transfer in evolution of these enzymes.

Results and Discussion

A. locustae encodes a class II photolyase

An open reading frame (ORF) of 1587 bp with high sequence similarity to members of the photolyase/cryptochrome family was identified from an *A. locustae* genomic sequence survey using blastx. This putative photolyase was located on a fragment of 6214 bp, which was completely sequenced and shown to encode three additional recognisable coding regions. Two are protein-coding regions showing significant similarity to putative bacterial hydrolases of unknown function and to the hypothetical *E. cuniculi* protein NP_597669, respectively. The fragment also encodes a gene for a glutaminyl-tRNA. The putative photolyase coding region was also PCR-amplified from genomic DNA extracted from an independently isolated sample of *A. locustae* spores. The arrangement of genes in the genomic fragment is consistent with microsporidian but not with metazoan genomic architecture, and no other insect DNA appeared in the large genomic sample (about 600 kb[24]). Likewise, no PCR products were obtained when genomic DNA from the uninfected *A. locustae* host, *Locusta migratoria*, was used with primer pairs derived from the cloned photolyase sequence.

Microsporidia are now known to be relatives of fungi,[25,26] which possess class I photolyase. However, the *A. locustae* photolyase shared a higher sequence similarity to class II photolyases from animals, viruses and plants (not shown). Aligning the *A. locustae* sequence with several class II photolyases and the class I photolyases from yeast and *E. coli* confirmed that the *A. locustae* sequence shares a number of substitutions specifically with class II photolyase (Figure 1, grey rectangles). Secondary structures were predicted for the yeast class I photolyase as well as *A. locustae* and four other class II photolyases, and these were compared with the experimentally deduced structure of the *E. coli* class I photolyase.[3] The structural prediction for yeast is similar to the pattern of α and β-motifs in *E. coli*, the only difference being that yeast lacks α helix a19, which is located in a region near the variable carboxy terminus (not shown). The predicted secondary structure for *A. locustae* was less similar than yeast and *E. coli* proteins, but was very similar to predictions for class II photolyases from mammal, fish, *Drosophila* and *Arabidopsis* (Figure 1). Notable differences between class I and II enzymes include the absence of α helix a2, the presence of an additional α helix between a11 and a 12, and a large deletion at a17 and a18. These differences are all predicted to be shared by the *A. locustae* protein and all the class II proteins analysed (Figure 1). The protein sequence of *A. locustae* also contains the two PROSITE signature sequences for class II photolyases: positions 363 to 377 are a perfect match to signature PS01083, while positions 490 to 509 are a

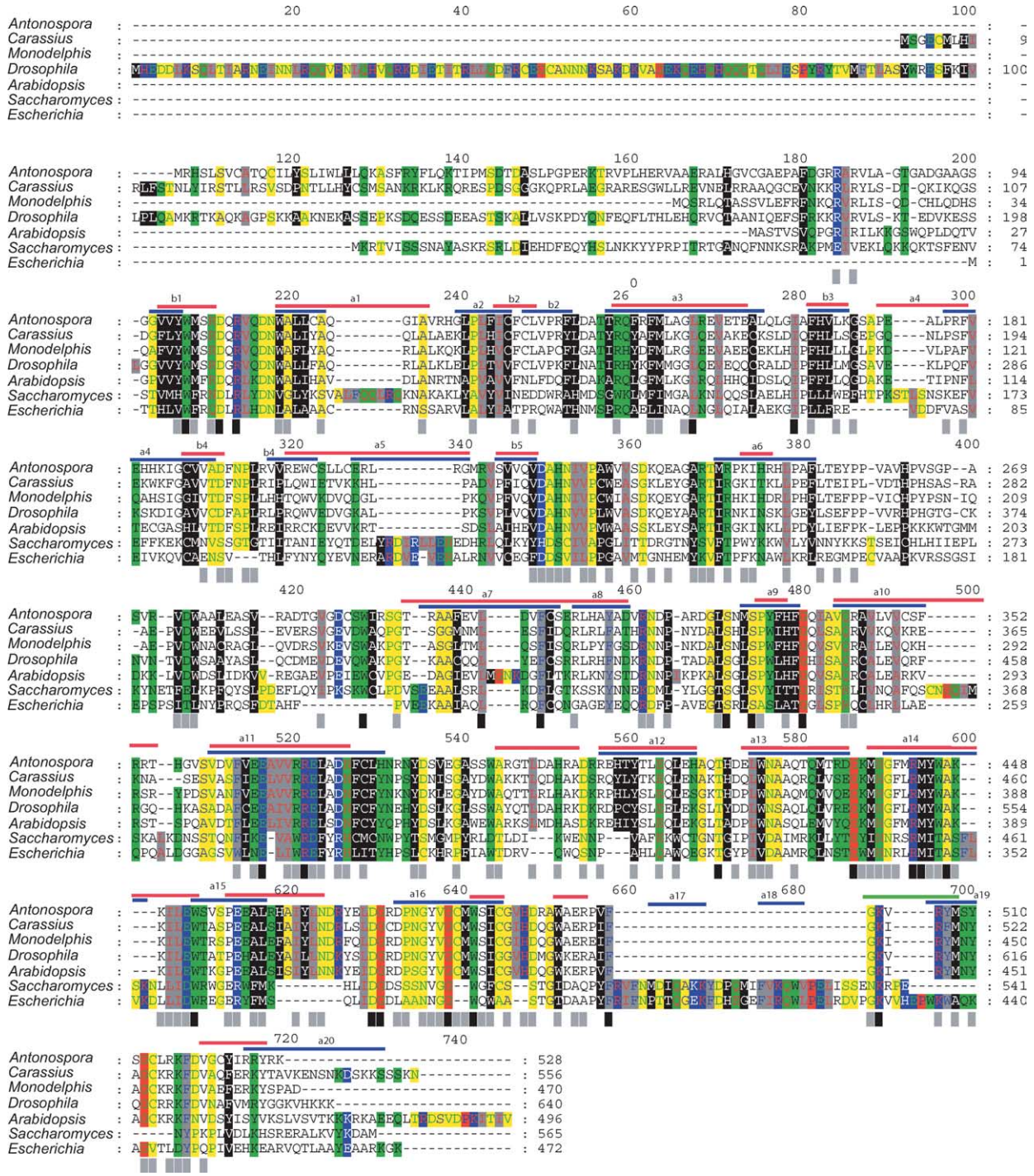

Figure 1. Protein sequence alignment showing the *A. locustae* photolyase, four class II sequences, yeast class I photolyase and the *E. coli* photolyase. Coloured columns show conservation of physicochemical properties. Black rectangles indicate perfectly conserved positions; grey rectangles indicate positions where *A. locustae* and the four class II sequences share the same amino acid. First three lines in the alignment comprise the α-beta domain, and the remainder correspond to the helical domain; blue lines highlight α and β structural domains as experimentally determined in *E. coli*;[3] red lines highlight α and β structural domains predicted for the class II sequences and *A. locustae*; green line marks a β bridge predicted only in *A. locustae* and all class II proteins; numbers on the alignment indicate absolute positions and numbers on the right indicate relative position from the start for each sequence.

close match to PS01084, differing only at the last position. Variation at this position occurs also in photolyases from some poxviruses, *Desulfovibrio* and *Methanosarcina* (not shown).

Complementation of photolyase-deficient *E. coli*

The evolutionary distance between the putative

A. locustae photolyase and other homologues is relatively high, and light-dependent reactions such as photolyase activity are unexpected in an intracellular parasite, so we have assessed the activity of the *A. locustae* enzyme by complementation. Both wild-type *A. locustae* photolyase (AlPHR) and a non-sense mutant bearing a single base-pair deletion at position 1165 (AlPHRΔ1165) were cloned into an *E. coli* expression vector, and the photolyase-deficient strain CSR-603 was transformed with both constructs. Ability to undergo photoreactivation was examined by measuring survival after successive irradiation with UV and white light. If the wild-type protein effectively restores photoreactivation, then recovery of UV-irradiated bacteria would only increase if white light is subsequently applied. By contrast, no increase in survival is expected when white light is suppressed, when the photolyase gene is absent or mutated, or when its expression is prevented. As shown in Figure 2, only IPTG-induced, wild-type photolyase-pGEX containing *E. coli* CSR-603 irradiated with white light showed significant recovery after UV-irradiation. Absence of insert, induction, or white light treatment all resulted in increased UV sensitivity. Bacteria transformed with the mutant pAlPHRΔ1165 showed similar sensitivity. Altogether, expression of the intact gene and white light illumination are both essential for increased recovery after UV illumination.

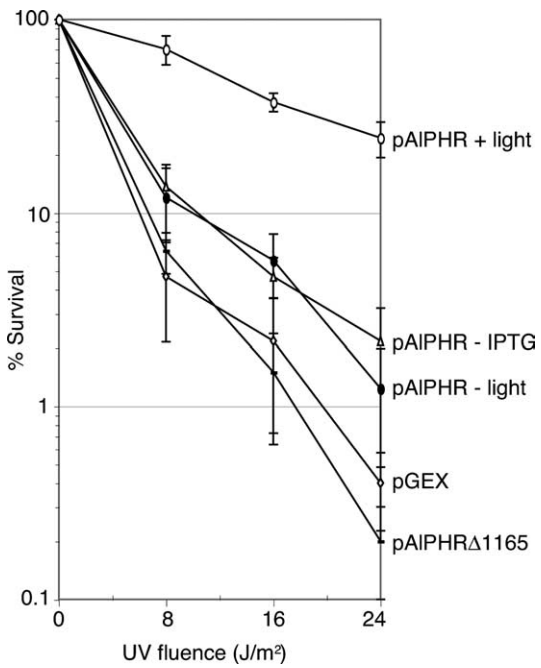

Figure 2. Phenotypic complementation of photolyase-deficient *E. coli* strain CSR-603. Bacteria were transformed with plasmids pAlPHR (intact photolyase) or pAlPHRΔ1165 (mutant photolyase, –) and exponentially grown cultures were induced with IPTG. Plated bacteria were exposed to the indicated intensities of UV and subsequently illuminated (○) or maintained in the dark (●) for one hour at room temperature. After overnight incubation at 37°C, survival was measured as the fraction of survivors referred to the number of CFU in the unexposed area of the plates. Percentage of survival is represented on a logarithmic scale. Non-induced bacteria carrying pAlPHR (△) and induced bacteria transformed with the empty, pGEX plasmid (◇) were included for comparison. Points represent the average among three independent experiments and error bars represent standard error.

Expression of photolyase spores and functional implications

Microsporidia are obligate intracellular parasites: the only stage of their life cycle that is exposed to extracellular conditions, and the stage responsible for host-to-host transmission, is the spore. The microsporidian spore is mechanically protected by a thick wall, and the resistance to different environmental stresses varies from species to species.[17] *A. locustae* propagates vertically (transovarially), and also horizontally through spores that may be deposited in the open environment (mostly on the surfaces of leaves) as host insects die.[17] DNA repair systems (aside from photolyase) are prominent in the proteome of *E. cuniculi*, but since spores are resting entities the importance of these energetically demanding systems in the spore itself is uncertain. Photolyase activity, on the other hand, requires only visible light as an energy source, and would therefore be expected to function normally in even completely dormant cells. Given the pivotal position of the spore in the parasite's life cycle, its exposure to the environment, and the restricted availability of energy within the spore, it would be interesting to know whether photolyase was active in spores of *A. locustae*. It has been recently shown that *A. locustae* also encodes a bacterially derived catalase gene not found in *E. cuniculi*, and that this enzyme is active in spores, altogether suggesting that *A. locustae* may have significantly enhanced protective resources to cope with environmental stress.[27]

To determine if photolyase is expressed in *A. locustae* spores, poly(A) RNA was isolated from purified spores and photolyase transcripts amplified using 5′-RACE. Three individual products were cloned and sequenced and all were found to initiate at −667 (position 1207 of the fragment). Interestingly, this position is well within the coding region of the upstream ORF, suggesting that transcription of the *A. locustae* photolyase is driven by a promoter embedded in another gene. Such overlap between a protein-coding and regulatory sequence is probably a reflection of the extreme compaction of the *A. locustae* genome. In addition, transcripts were also detected by RT-PCR using photolyase-specific primers and also appeared in an ongoing *A. locustae* spore EST sequencing project (B. A. P. Williams & P.J.K., unpublished results). This EST project contains about six

photolyase hits but no known *L. migratoria* gene or other sequences of insect origin. Altogether, these results show that photolyase is expressed at least at the mRNA level in spores, which leads to the prediction that *A. locustae* spores are more resistant to UV radiation than those of *E. cuniculi* (which lacks photolyase). Indeed, the UV resistance of *E. cuniculi* spores has been shown to be comparable to that of a DNA repair-deficient *B. subtilis* strain.[28]

Molecular evolution of the *A. locustae* photolyase

The photolyase/cryptochrome family is believed to be very ancient, since its members appear in all three domains of life.[4] The functional diversity achieved by the family through its long history can be roughly divided into two main categories: the photolyases, which repair DNA photoproducts in a light-dependent manner, and the cryptochromes, which have more diverse functions, but are generally involved in regulatory pathways *via* protein–protein or protein–DNA interactions.[1,8] Phylogenetic studies have suggested that the ancestor of the family was a photolyase,[1,4] leading to the assumption that photoreactivation was key to the success of early life when the intensity of ultraviolet radiation on the Earth's surface was much higher than it is today.[4,7] However, the early events that led to the diversification of this gene family are obscured by the lack of phylogenetic resolution at several nodes of the photolyase/cryptochrome gene tree.

All known CPD photolyases belong to one of two well-defined classes (Figure 3A). class II CPD photolyases are the most divergent group of the family, while class I enzymes are more similar to the cryptochromes and (6–4) photolyases. Curiously, class I photolyases have been found in several ascomycete fungi but in no other eukaryotic group, which has been interpreted as a result of lateral transfer from a bacterium to a fungus.[4] Consistent with characteristics observed in the primary sequence, the *A. locustae* gene falls within the class II CPD photolyase clade with strong support (Figure 3A), ruling out any possibility that known microsporidian and fungal photolyases are orthologous.

Focusing specifically on the class II photolyases (Figure 3B), the *A. locustae* sequence consistently branched at the base of a clade comprising animals, poxviruses, and the apicomplexan *Plasmodium*. Without knowing the nature of an ancestral fungal photolyase, or indeed even if the ancestor of fungi contained photolyase, it is difficult to distinguish between several alternative explanations for the origin of the microsporidian gene. It is possible that the ancestor of fungi and Microsporidia possessed either a class I or class II photolyase, which would mean gene loss in ascomycetes and *E. cuniculi* (if the ancestor possessed class II) or in Microsporidia (if it possessed class I) followed by the regaining of a gene of the other class in either ascomycetes or *A. locustae*, presumably by lateral transfer. If the former (i.e. the *A. locustae* gene represents an ancestral fungal gene), then some of the deeper-branching fungi may still possess a related photolyase, but we failed to amplify class II photolyase from several genera each of zygomycetes and chytridiomycetes using several sets of degenerate primers (not shown). The possibility of the ancestor of fungi and Microsporidia to have both classes of photolyase cannot be ruled out, but is less likely, since both types of enzyme perform the same reaction, and no organism is presently known to have both photolyases. A full understanding of the origin of the microsporidian photolyase is presently impossible due to the limited sampling of class II genes in diverse eukaryotes and a lack of phylogenetic resolution. Nevertheless, it is apparent that photolyases from ascomycetes, *A. locustae*, or both are likely derived from lateral gene transfer.

While fungi (other than Microsporidia) do not yet appear to have a class II CPD photolyase, we have found evidence for a potential fungal (6–4) photolyase. An ORF in ongoing genome sequence of the fungus *Magnaporthe grisea* which annotated as a hypothetical protein shows significant similarity to proteins that fall in the group comprising animal cryptochromes and (6–4) photolyases (Figure 3A). Cryptochromes are generally defined as photolyase-like molecules lacking photolyase activity.[8] This functional definition does not necessarily reflect evolutionary events, since there is evidence that animal and plant cryptochromes have independent origins (*Drosophila* and *Arabidopsis* CRY interact with partner proteins at the amino and carboxy-terminal domains, respectively[8]). In any case, it is apparent that animal cryptochromes share a common origin and that they are related to (6–4) photolyases, but the exact nature of that relationship cannot be deduced from the trees. The existence of a fungal member of this group (Figure 3A) pushes back the date of their estimated diversification and raises questions about the activity of the fungal protein.

Conclusions

The microsporidian parasite *A. locustae* possesses a gene encoding a functional class II CPD photolyase, and expression of this gene in the spore suggests that may be involved in environmental stress response. The phylogeny of the photolyase/cryptochrome family and the known distribution of photolyase both suggest that the microsporidian gene may have originated by lateral gene transfer or, alternatively, that was inherited from fungal ancestors after a complex series of gene losses. It has recently been shown that *A. locustae* spores also express a functional group II catalase that is absent from the *E. cuniculi* genome and was almost certainly acquired by lateral gene transfer from a

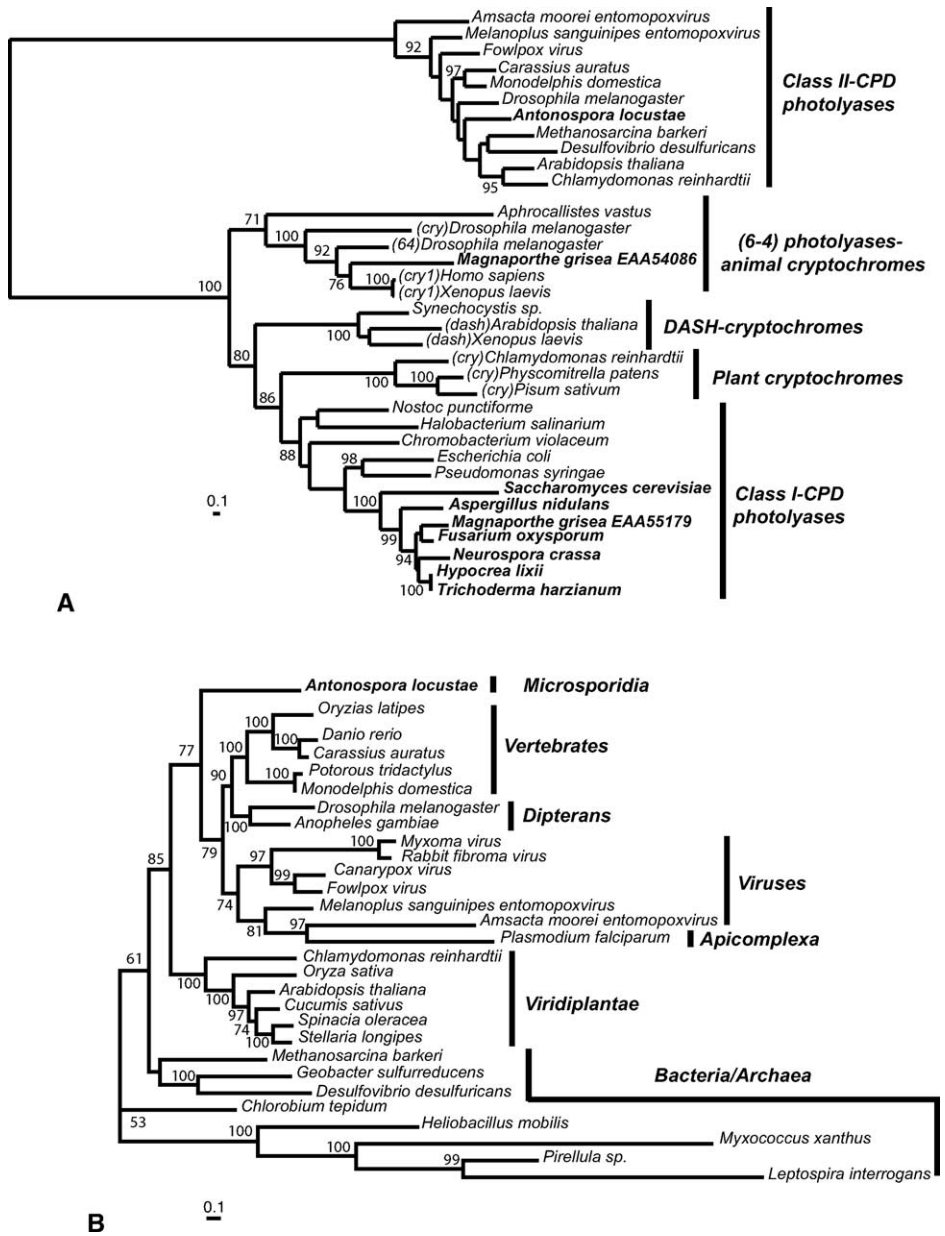

Figure 3. Phylogeny of photolyases and cryptochromes. Numbers indicate bootstrap support (100 replicates). Trees were built with PHYML[34] using the Dayhoff substitution model, gamma and invariant sites. A, ML tree showing the interrelationships among members of all recognized groups of the photolyase/cryptochrome gene family. Sequences from fungi and Microsporidia are highlighted. B, ML tree with all available class II photolyases.

bacterium.[27] If photolyase was also acquired by lateral transfer, then *A. locustae* would have significantly augmented the capacity of its spores to withstand environmental stress (i.e. oxidative stress and UV damage) by capturing foreign coding sequences. This process is only just beginning to be characterized in eukaryotic genomes,[29,30] and the contrast between *A. locustae* and *E. cuniculi* advance the notion that it may be more prevalent in some genomes than others. Regardless of how these genes originated, however, it is clear that *A. locustae* has an enhanced repertoire of stress response proteins compared with *E. cuniculi*, leading to the prediction that *E. cuniculi* spores would be more prone to various kinds of environmental damage.

Materials and Methods

DNA sequencing and analysis of the *A. locustae* photolyase gene

A. locustae spores (ATCC 30860) were kindly provided by M & R Durango, Inc. (Bayfield, CO). Genomic DNA was isolated and multiple genomic libraries were constructed as part of an *A. locustae* random sequence survey as described.[31] Sequences were compared against public databases using the ESTid program (generously

provided by M. Reith, Institute for Marine Biosciences, Halifax, Nova Scotia, Canada) and NetBLAST† using BLASTX and BLASTN. A fragment containing part of the photolyase ORF was completely sequenced and resulted in 5519 bp. The fragment was found to overlap with several additional clones, resulting in the 6210 bp sequence. The photolyase coding sequence from an independent sample of *A. locustae* genomic DNA was also amplified and sequenced, and found to be identical. To rule out an insect host origin for the photolyase gene, we made PCR using a pair of primers that amplify the complete coding sequence of photolyase. Templates were genomic *A. locustae* and *L. migratoria* DNA. PCR conditions were as follows. Initial denaturing step, 95 °C, three minutes; 30 cycles with 95 °C, one minute, 56 °C, 45 minutes and 72 °C, one minute 30 seconds. A three-minute final extension at 72 °C was also performed. The expected 1.6 kb band was present in *A. locustae* but not in *L. migratoria*. Deduced protein sequence photolyases were analysed for secondary structure with SSPRO‡ and compared to the *E. coli* PHR structure.[3] Protein sequences were aligned with T-Coffee[32] and graphic display was prepared with GeneDoc§. Distance matrix was built with MEGA 2.1‖ using *p*-distance, and pair-wise gap deletion from an alignment made with T-Coffee.

Phenotypic complementation of phr-deficient *E. coli*

The *A. locustae* coding sequence for photolyase was amplified and cloned in pGEX expression vector (Amersham Biosciences). Four clones were sequenced, one of which contained a single-base deletion at position 1165. Photoreactivation-deficient *E. coli* strain CSR-603 (provided by the *E. coli* Genetic Stock Center, Yale University) was transformed with the following plasmids: pGEX with intact *A. locustae* photolyase (pAlPHR), pGEX with the deletion-mutant *A. locustae* photolyase (pAlPHRΔ1165), and pGEX alone. Photoreactivation was assayed as follows: transformants were grown in Luria–Bertani (LB) medium[33] with 50 mg/ml ampicillin until they reached an $A_{600\,nm}$ of 0.5. Cultures were divided and isopropyl β-D-thiogalactopyranoside (IPTG) to 1 mM was added to one tube of each culture and incubated at 37 °C in a shaker for two hours to allow expression of the recombinant gene. Cells from 1 ml of medium were pelleted and resuspended in 10 ml of saline. The bacterial suspension was diluted 1/100 and 80 μl were plated on LB-agar. One half of each plate was irradiated with UV-B (peak at 312 nm, fluences 8 J/m^2, 16 J/m^2 and 24 J/m^2) with a FotoPrepI transilluminator (Fotodyne, New Berlin, WI), while the other half was kept covered. Identical plates were divided and one half exposed to white light for one hour while the remainder were kept wrapped in aluminium foil. Recovery was measured by growing plates at 37 °C overnight and dividing the number of CFU in the irradiated side by the number of CFU in the non-irradiated side. UV fluorescence was measured with a Spectroline DRC-100X digital radiometer (Spectronics Corp.) with an UV-B detector attached.

RNA extraction and detection of photolyase expression in *A. locustae* spores

A. locustae spores were ground under liquid nitrogen with a mortar and pestle, and polyadenylated RNA was extracted using MicroPoly(A)Pure (Ambion). Reverse transcription with oligo(dT)$_{20}$ primers was performed using ThermoScript RT-PCR System (Invitrogen). Second-strand synthesis was carried out with Taq DNA polymerase (Invitrogen) using different pairs of primers designed from the *A. locustae* photolyase spanning the whole coding sequence. Five-prime RACE was performed with Ambion's First Choice 5′ RLM-RACE. Briefly, purified polyadenylated RNA was treated with calf intestinal phosphatase (CIP) to remove 5′ phosphates, according to the manufacturer's directions. The resulting dephosphorylated RNA was treated with tobacco acid pyrophosphatase (TAP) to remove the 5′ cap from full-length mRNAs. RNA was then ligated with 5′-RACE RNA adapter oligonucleotide using T4 RNA ligase. First-strand cDNA synthesis was made by reverse-transcribing the ligated mRNA with random decamers. Photolyase transcripts were then PCR amplified using the 5′-RACE outer primer and a 3′ reverse primer homologous to a region 300 bp downstream of the *A. locustae* photolyase start position. The single, discrete band at about 1 kb was extracted from the gel and cloned into a pCRII vector (Invitrogen). Plasmids from positive colonies were purified and three clones were sequenced. Poly(A) RNA was also subject to RT-PCR using pairs of primers that span different regions of the photolyase gene. Products representing four different regions of the gene (between 300 bp and 900 bp) were cloned and their sequences were identical with the photolyase cloned from *A. locustae*.

Phylogenetic analysis

The conceptual amino acid sequence of the *A. locustae* photolyase was added to a global alignment of the photolyase/cryptochrome family and an alignment of class II photolyases. Alignments were made with T-Coffee[32] and ambiguously aligned regions were excluded from further analyses. Phylogenetic trees were constructed by using Maximum Likelihood with PhyML[34] using a Dayhoff substitution model. Site-to-site rate variation was modelled on a gamma distribution with four categories of substitution rates and invariable sites with the α parameter and proportion of invariable sites estimated from the data. Bootstrap values were obtained from 100 replicates for each dataset.

Nucleotide sequence accession numbers

Nucleotide sequence data have been deposited in Genbank under accession number AY608637.

† ftp://ftp.ncbi.nlm.nih.gov/blast/executables/release/

‡ http://www.igb.uci.edu/tools/scratch/

§ Nicholas, K. B. & Nicholas, H. B. J. (2000). GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the author.

‖ http://www.megasoftware.net

Acknowledgements

We thank M. Reith (Institute for Marine Biosciences) for the use of ESTid and M. Upadhyaya and N. Furness (University of British Columbia)

for kindly providing a radiometer. Unpublished *A. locustae* EST data were provided by B. A. P. Williams. *Escherichia coli* strain CSR-603 was kindly provided by The *Escherichia coli* Genetic Stock Center, Yale University. This work was supported by a New Investigator Award to P.J.K. from the Burroughs-Wellcome Fund and a grant from the Canadian Institutes for Health Research (MOP-42517). P.J.K. is a New Investigator of the Michael Smith Foundation of Health Research and Canadian Institute for Health Research and Scholar of the Canadian Institute for Advanced Research.

References

1. Todo, T. (1999). Functional diversity of the DNA photolyase/blue light receptor family. *Mutat. Res.* **434**, 89–97.
2. Sancar, A. (2003). Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. *Chem. Rev.* **103**, 2203–2237.
3. Park, H. W., Kim, S. T., Sancar, A. & Deisenhofer, J. (1995). Crystal structure of DNA photolyase from *Escherichia coli*. *Science*, **268**, 1866–1872.
4. Kanai, S., Kikuno, R., Toh, H., Ryo, H. & Todo, T. (1997). Molecular evolution of the photolyase-blue-light photoreceptor family. *J. Mol. Evol.* **45**, 535–548.
5. Ahmad, M., Jarillo, J. A., Smirnova, O. & Cashmore, A. R. (1998). Cryptochrome blue-light photoreceptors of *Arabidopsis* implicated in phototropism. *Nature*, **392**, 720–723.
6. Emery, P., So, W. V., Kaneko, M., Hall, J. C. & Rosbash, M. (1998). CRY, a *Drosophila* clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell*, **95**, 669–679.
7. Rothschild, L. J. (1999). The influence of UV radiation on protistan evolution. *J. Eukaryot. Microbiol.* **46**, 548–555.
8. Cashmore, A. R. (2003). Cryptochromes: enabling plants and animals to determine circadian time. *Cell*, **114**, 537–543.
9. O'Connor, K. A., McBride, M. J., West, M., Yu, H., Trinh, L., Yuan, K. *et al.* (1996). Photolyase of *Myxococcus xanthus*, a Gram-negative eubacterium, is more similar to photolyases found in Archaea and higher eukaryotes than to photolyases of other eubacteria. *J. Biol. Chem.* **271**, 6252–6259.
10. Yasui, A., Eker, A. P., Yasuhira, S., Yajima, H., Kobayashi, T., Takao, M. & Oikawa, A. (1994). A new class of DNA photolyases present in various organisms including aplacental mammals. *EMBO J.* **13**, 6143–6151.
11. Ahmad, M., Jarillo, J. A., Klimczak, L. J., Landry, L. G., Peng, T., Last, R. L. & Cashmore, A. R. (1997). An enzyme similar to animal type II photolyases mediates photoreactivation in *Arabidopsis*. *Plant Cell*, **9**, 199–207.
12. Petersen, J. L., Lang, D. W. & Small, G. D. (1999). Cloning and characterization of a class II DNA photolyase from *Chlamydomonas*. *Plant Mol. Biol.* **40**, 1063–1071.
13. Kim, S. T., Malhotra, K., Ryo, H., Sancar, A. & Todo, T. (1996). Purification and characterization of *Drosophila melanogaster* photolyase. *Mutat. Res.* **363**, 97–104.
14. Kato, T., Jr, Todo, T., Ayaki, H., Ishizaki, K., Morita, T., Mitra, S. & Ikenaga, M. (1994). Cloning of a marsupial DNA photolyase gene and the lack of related nucleotide sequences in placental mammals. *Nucl. Acids Res.* **22**, 4119–4124.
15. Srinivasan, V., Schnitzlein, W. M. & Tripathy, D. N. (2001). Fowlpox virus encodes a novel DNA repair enzyme, CPD-photolyase, that restores infectivity of UV light-damaged virus. *J. Virol.* **75**, 1681–1688.
16. Sancar, G. B. (1985). Sequence of the *Saccharomyces cerevisiae* PHR1 gene and homology of the PHR1 photolyase to *E. coli* photolyase. *Nucl. Acids Res.* **13**, 8231–8246.
17. Wittner, M. & Weiss, L. M. (1999). *The Microsporidia and Microsporidiosis*, ASM Press, Washington, DC.
18. Vossbrinck, C. R., Maddox, J. V., Friedman, S., Debrunner-Vossbrinck, B. A. & Woese, C. R. (1987). Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature*, **326**, 411–414.
19. Keeling, P. J., Luker, M. A. & Palmer, J. D. (2000). Evidence from beta-tubulin phylogeny that microsporidia evolved from within the fungi. *Mol. Biol. Evol.* **17**, 23–31.
20. Keeling, P. J. & Fast, N. M. (2002). Microsporidia: biology and evolution of highly reduced intracellular parasites. *Annu. Rev. Microbiol*, **56**, 93–116.
21. Katinka, M. D., Duprat, S., Cornillot, E., Metenier, G., Thomarat, F., Prensier, G. *et al.* (2001). Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature*, **414**, 450–453.
22. Sokolova, Y. Y., Dolgikh, V. V., Morzhina, E. V., Nassonova, E. S., Issi, I. V., Terry, R. S. *et al.* (2003). Establishment of the new genus *Paranosema* based on the ultrastructure and molecular phylogeny of the type species *Paranosema grylli* Gen. Nov., Comb. Nov. (Sokolova, Selezniov, Dolgikh, Issi 1994), from the cricket *Gryllus bimaculatus* Deg. *J. Invertebr. Pathol.* **84**, 159–172.
23. Slamovits, C. H., Williams, B. A. & Keeling, P. J. (2004). Transfer of *Nosema locustae* (Microsporidia) to *Antonospora locustae* n. comb. Based on molecular and ultrastructural data. *J. Eukaryot. Microbiol.* **51**, 207–213.
24. Slamovits, C. H., Fast, N. M., Law, J. S. & Keeling, P. J. (2004). Genome compaction and stability in microsporidian intracellular parasites. *Curr. Biol.* **14**, 891–896.
25. Keeling, P. J. & Doolittle, W. F. (1996). Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. *Mol. Biol. Evol.* **13**, 1297–1305.
26. Van de Peer, Y., Ben Ali, A. & Meyer, A. (2000). Microsporidia: accumulating molecular evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi. *Gene*, **246**, 1–8.
27. 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.
28. Marshall, M. M., Hayes, S., Moffett, J., Sterling, C. R. & Nicholson, W. L. (2003). Comparison of UV inactivation of spores of three encephalitozoon species with that of spores of two DNA repair-deficient *Bacillus subtilis* biodosimetry strains. *Appl. Environ. Microbiol.* **69**, 683–685.
29. Bergthorsson, U., Adams, K. L., Thomason, B. &

Palmer, J. D. (2003). Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature*, **424**, 197–201.
30. Archibald, J. M., Rogers, M. B., Toop, M., Ishida, K. & Keeling, P. J. (2003). Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. *Proc. Natl Acad. Sci. USA*, **100**, 7678–7683.
31. 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.
32. Notredame, C., Higgins, D. G. & Heringa, J. (2000). T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217.
33. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edit., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
34. Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704.

Edited by J. Karn

(Received 4 May 2004; received in revised form 11 June 2004; accepted 14 June 2004)

Supplementary Material for this paper is available on Science Direct