International Scholars Journals

African Journal of Virology Research ISSN 3421-7347 Vol. 2 (2), pp. 135-141, February, 2008. Available online at www.internationalscholarsjournals.org © International Scholars Journals

Full Length Research Paper

A study of polymerase chain reaction-computer database for species determination of human pathogenic microsporidia

Scot E. Dowd[1]* and Jeanette A. Thurston-Enriquez[2]

[1]USDA-ARS, Livestock Issues Research Unit, Lubbock, TX, Texas.

[2]USDA-ARS, Soil and Water Conservation Research Unit, Lincoln, NE 68583, Texas.

Accepted 11 January, 2008

The potential for waterborne disease and zoonotic transmission of at least two species of human pathogenic microsporidia has heightened interest in clinical and environmental detection methods for these organisms. Detection using the polymerase chain reaction (PCR) followed by computer database homology comparison (CDHC) (PCR-CDHC) was reported previously by this research group. As a result, PCR-CDHC has been employed by many research groups around the world for species determination of human pathogenic microsporidia. To validate the CDHC speciation approach, a phylogenetic tree was generated using the small subunit ribosomal DNA sequences (SSU-rDNA) of a large number of microsporidia. An index of similarity was created and used as part of an assessment of CDHCs ability to differentiate between closely related species. Polymerase chain reaction followed by dye termination PCR sequencing and subsequent CDHC of the sequences was performed on 8 species of microsporidia including four human pathogenic strains. The four non-human pathogenic microsporidia tested by this approach were those shown by the phylogenetic analyses to be very closely related to the other human pathogenic species as determined by branch length. In all cases the CDHC approach was able to correctly identify the eight species of microsporidia evaluated. To provide an example of PCR-CDHC, a "universal" and two previously published pathogen-specific microsporidia PCR protocols followed by PCR-CDHC was conducted to assess their ability to detect naturally occurring microsporidia species in swine wastewater. Only one primer set resulted in a PCR-CDHC analysis where presumptive human pathogenic microsporidia was detected. Subsequent CDHC showed these presumptive positive PCR results were actually false positives. With the appropriate primer set, PCR-CDHC proves to be a reliable method that can be used for specific species determination of human pathogenic microsporidia in samples where non-pathogenic species may be present.

Key words: Microsporidia, Sequencing, BLAST, PCR, detection.

INTRODUCTION

The term microsporidia is used to describe a large group of obligate intracellular parasites with diverse hosts that range from humans to insects. The microsporidia have been associated with disease in humans since the mid-1980s (Desportes et al., 1985) and are becoming a major cause of concern, especially, with confirmation of the presence of human pathogenic species in water (Dowd et al., 2003c; Dowd et al., 1998) and in animal feces (Lores et al., 2002; Fayer et al., 2003). These findings suggest that at least two human-pathogenic species (*Enterocytozoon bieneusi* and *Encephalitozoon intestinalis*) may be of concern in environments impacted by human and animal wastes, such as water and soil.

Several methods have been used for the detection and characterization of these emerging pathogenic fungi in environmental samples (Sparfel et al., 1997;et al., 2003; Dowd et al., 1999; Dowd et al., 1998). Of these methods, one of the most widely employed, specific, sensitive and

*Corresponding author. E-mail: sdowd@lbk.ars.usda.gov. Tel: 806-746-5356 ext 122 office. Fax: 806-744-5028.

sensitive and rapid methods for the detection of microsporidia is the polymerase chain reaction (PCR) . In addition, the use of PCR followed by sequencing of the PCR products have provided the ability to quickly determine the species of microsporidia, a task, which traditionally required the use of transmission electron microscopy (Canning et al., 1987). There are two molecular methods that show considerable utility for detection of the range of human pathogenic microsporidia species. The first is the use of the PCR followed by restriction fragment-length polymorphic (RFLP) analysis, to help differentiate species based upon varied restriction patterns (Fedorko et al., 1995). The second and the most promising is PCR-CDHC, the combined use of PCR using specific PCR primers followed by PCR product sequencing and Computer Database Homology Comparison (CDHC). This method relies on species- specific identification based on genetic differences of an amplified portion of the SSU rDNA (Thurston-Enriquez et al., 2002; et al., 1998). The use of restriction analysis is no more cost effective and leaves considerable doubt as to the actual identity of the organisms from which the amplicon was derived. For instance, in clinical samples, where only a few species of microsporidia are expected, the *Encephalitozoonidae* (*Encephalitozoon intestinalis*, *Encephalitozoon hellem*, and *Encephalitozoon cuniculi*) can often be distinguished by restriction analysis of their PCR products (Fedorko et al., 1995). However, for environmental samples, there is also the potential for other microsporidia genera, not important to human or animal health, to be present. It is not unreasonable to consider that these other microsporidia species, some of which have never been characterized, could have the same restriction patterns as the human pathogens, thus lending uncertainty to such analyses. Thus, the use of PCR-CDHC, which allows for a nucleotide-by-nucleotide analysis of the PCR product, appears to be a better method for molecular identification of the human pathogenic microsporidia. Until now, this method has not been evaluated for discrimination between closely related species.

One of the issues critical to accessing PCR-CDHC is the close genetic similarity between the small subunit ribosomal DNA sequences (SSU-rDNA) that are targeted by PCR protocols that employ CDHC for species-specific microsporidia detection. While these protocols have been used by this and other research groups, the specificity and utility of this identification approach has not been experimentally validated. The current study challenged PCR-CDHC by assessing its ability to specifically identify human and non-human pathogenic strains that are genetically closely related based upon 16s phylogeny. This study also evaluated phylogenetic relationships of a large number of microsporidia based on SSU-rDNA sequences acquired through available bioinformatics databases, and then used this data to further evaluate the CDHC methodology for species determination of microsporidia and to re-examine phylogenetic relationships among microsporidia species. Finally, two swine lagoon samples were assayed using three different PCR primer sets to determine the ability of PCR-CDHC to identify human pathogenic microsporidia species.

MATERIALS AND METHODS

Protozoa

Microsporidia spores were obtained directly from the American Type Culture Collection (ATCC) to ensure their identity. These spores included *Encephalitozoon intestinalis* ATCC 50506, *Encephalitozoon hellem* ATCC 50504, *Encephalitozoon cuniculi* ATCC 50503, *Vittaforma cornea* ATCC 50505, *Endoreticulatus sherburg* ATCC 50040, *Nosema necatrix* ATCC 30460, and *Nosema trachiplusieae* ATCC 30702.

Phylogenetic analysis

Extensive phylogenetic analysis of the microsporidia sequences found in the GenBank, was performed using the Genetics Computer Groups Wisconsin Package Version 9.1 (Accelrys, Madison, WI). A multiple alignment was created using CLUSTAL W (11) adjusted with high penalties for gap creation and extension. Manual editing of this alignment was performed using the Seqlab interface of the Wisconsin Package. Because of the variety of lengths associated with the database sequences, the ends and beginnings of sequences were truncated to ensure that all sequences were homologous and of the same length. Three strains of *Encephalitozoon cuniculi* were included in the analysis as grouping controls, providing a similarity index for phylogram branch lengths. Finally, as a check for the accuracy of the resulting alignment, all gaps were removed and Jotun Hein and Clustal W methods used to realign the sequences. These alignments were then compared to the first alignment in order to ensure reproducibility and accuracy.

A variety of methods were used to measure the evolutionary relationship of the aligned sequences during editing, but the final extensive confirmative phylogenetic analysis was performed using Phylogenetic Analysis Using Parsimony (PAUP, Smithsonian Institute). Bootstraping (10,000 replications) that included a heuristic search. The starting tree was obtained by stepwise addition. Tree-bisection-reconnection branch swapping was performed and maximum parsimony used as the criterion for optimization. The total characters considered in the alignment were 1156. The resulting data was displayed as an unrooted phylogram generated using PAUPdisplay and is seen in Figure 1.

Swine lagoon sample analysis

In order to provide a set of environmental samples to exemplify incorrect use of PCR-CDHC we collected and analyzed wastewater samples from two different swine farms. These samples were expected to have a high diversity of fungi, including microsporidia, and thus the probability of false positive results. Swine wastewater samples from Nebraska (USA) were collected from two different farms in sterile 1L polypropylene bottles and stored at 4^{o} C using standard USEPA methods for the collection of wastewater effluent samples. 15 ml of wastewater was resuspended in 35 ml of 50 mM Tris/0.5% polyoxyethlene-sorbitan monooleate (Tween 80) and concentrated by centrifugation (1800 x g for 10 min). The supernatant was aspirated off and the sample concentrate was resuspended in 20 ml 50 mM Tris/0.5% polyoxyethlene- sorbitan monooleate solution. This solution was overlaid onto 30 ml of room temperature 2.5 M sucrose (specific gravity of 1.34 g/ml) and

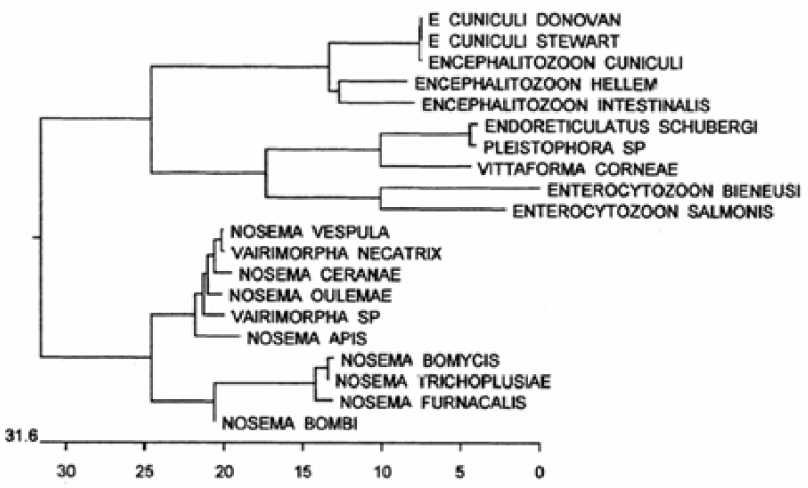

Figure 1. phylogram showing the relationships between the microsporidia based upon SSU-RNA. The close relationship of *V. necatrix* and *N. vespula* provide the best challenge for PCR- CDHC the close relationship between *N. bombycis* and *N. trichoplusiae* are also used as a challenge. Three varieties of *E. cuniculi* were utilized as a grouping control. From this we can see that *N. vespula* and *V. necatrix* are highly similar and *E. schubergi* and *Pleitophora* sp. have high similarity. Differentiation of these very similar species provides the best challenge for the PCR-CDHC methodology.

centrifuged at 512 x g for 10 min. After centrifugation, the top portion and 5 - 8 ml past the interface was collected and deionized water was used to bring the volume up to 50 ml. This solution was centrifuged at 512 x g for 10 min. The supernatant was aspirated off and the sample pellet was resuspended in an equal volume of 0.01 M phosphate buffered saline (pH 7.4). DNA was extracted from 0.2 ml of the pellet solution the Qiagen QIAmp DNA Stool Mini Kit (Qiagen, Valencia, CA) as described by -Enriquez et al. (Thurston-Enriquez et al., 2002).

PCR

Two sets of microsporidia primers were designed for this study. The first set is designed to be more specific for the human pathogenic microsporidia, while the second set was designed to be more universal in order to amplify a wide variety of small ribosomal subunits of the microsporidia group. The pathogen- specific primers were designated micpF1 for the forward primer (5'AGG TTG ATT CTG CCT GAC-3') and micpR1 for the reverse primer (5'GCG CCT GCT GCC RTC CT 3') . These primers generate products of approximately 305 bp for the Encephalitozoonidae. The universal primers were designated micuF1 for the forward primer (5'TGT GCG CTT AAT TTG ACT CA 3') and micuR1 for the reverse primer (5' ATA RCG ACG GGC GGT GTG 3'). The universal primers generate products between 370 and 400 bp for most microsporidia. The third PCR primer set was previously reported by Fedorko et al (Fedorko et al., 1995). The Fedorko primer sequences are: 5'-CAC CAG GTT GAT TCT GCC TGA C-3' for the forward primer and the reverse primer is 5'-CCT CTC CGG AAC CAA ACC CTG- 3'. All PCR reactions were performed using standard conditions, PCR buffer (Perkin-Elmer Corp., Norwalk, Conn), and *Taq* Gold (Perkin-Elmer Corp., Norwalk, Conn.). The cycling conditions started with 10 min initial enzyme activation step at 95^{O}C, followed by 30 cycles of the PCR consisting of 1 min denaturation at 94^{O}C, 1 min annealing at 57^{O} C, and 1 min extension at 72^{O}C. A final 5 min extension period at 72^{O}C was also included after the cycling.

The SSU-rDNA of each of the three separate species of *Encephalitozoon* were amplified five separate times using the PCR. This resulted in 15 separate reactions, 5 for each species of *Encephalitozoon*.

PCR sequencing

For PCR reactions that resulted in only one product, the PCR product was purified using the QIAquick PCR purification kit (QIAGEN, Santa Clarita, CA) and subsequently resuspended in sterile nanopure water (dnase/rnase free). For PCR reactions resulting in more than one product, PCR products were extracted from agarose gels and purified using the Wizard SV Gel and PCR Clean-up System (Promega, Madison, WI) and resuspended in sterile nanopure water (dnase/rnase free) . The forward PCR primer was then used for dye termination PCR sequencing. The products from each of these reactions were then sequenced in duplicate or triplicate using standard BIGdye3 chemistry and methods on an ABI3100 (Applied Biosystems, CA).

Computer Database Homology Analysis (CDHC)

CDHC used to identify the species of microsporidia. This procedure was briefly described by Dowd et al. (2003). This was procedure was utilized in order to determine if this approach would consistently identify the species of *Encephalitozoon* from which the amplicon was derived. In addition, *Vittaforma corneae*, a human pathogen, and three non- human pathogens which were *Endoreticulatus schubergi*, *Nosema trichoplusiae*, and *Vairimorpha necatrix* were used to challenge CDHC. These species were chosen because of the similarity shown in the phylogenetic analyses (Figure 1). The initial database (NCBI) search was performed using BLASTn (Altschul et al., 1990) provided by the National Center for Biotechnology Information (NCBI) at their web site [http://www.ncbi.nlm.nih.gov]. Results were then carefully analyzed

Table 1. Results of CDHC for the *Encephalitozoon* sp.

Test Organism[a]	BLASTn ID[b] Accession number[c]	BLAST bit Score[d]	BLAST e-value[e]	BLAST Identity[f]	e-value[g] negative ID[h]
E. cuniculi	*E.cuniculi* AJ005581	416	e-115	99	e-35/ *E. hellem*
E. cuniculi	*E. cuniculi* AJ005581	412	e-114	100	2e-34/*E hellem*
E. cuniculi	*E. cuniculi* AJ005581	422	e-116	100	2e-37/*E. hellem*
E. cuniculi	*E. cuniculi* AJ005581	412	e-114	100	e-34/*E. hellem*
E. cuniculi	*E. cuniculi* AJ005581	422	e-117	100	2e-37/*E. hellem*
E. intestinalis	*E. intestinalis* /L39113	389	e-106	97	1e-44/*E. hellem*
E. intestinalis	*E. intestinalis* L39113	335	2e-90	97	5e-44/*E. hellem*
E. intestinalis	*E. intestinalis* L39113	389	e-106	97	2e-44/*E. hellem*
E. intestinalis	*E. intestinalis* L39113	337	4e-91	97	1e-44/*E. hellem*
E. intestinalis	*E. intestinalis* L39113	389	e-106	97	2e-37/*E. hellem*
E. hellem	*E. hellem* AF039230	444	e-123	100	1e-39/ *E. intestinalis*
E. hellem	*E. hellem* AF039230	446	e-124	100	1e-39/ *E. intestinalis*
E. hellem	*E. hellem* AF039230	440	e-122	100	1e-38/ *E. intestinalis*
E. hellem	*E. hellem* AF039230	446	e-124	100	1e-39/*E. intestinalis*
E. hellem	*E. hellem* AF039230	444	e-123	100	6e-41/*E. intestinalis*

This table provides the results of PCR-CDHC for the different varieties of human pathogenic microsporidia. The ATCC designation of genus and species is indicated in the first column. In the second column is the primary CDHC identification along with the appropriate accession number of this hit. In columns 3-5 we provide the CDHC statistics that are most relevant for evaluation of the results. In the final column we provide the first false identification of the CDHC and its associated E-value. We can see by comparing the E-values of the top hit to the first incorrect identification that the method is robust and able to differentiate among the closely related species. [a]Organisms from pure culture used as subjects for the PCR sequencing CDHC validation, [b]CDHC genus and species determined for subjects by CDHC, [c]Accession number for database sequence identified by CDHC, [d]Statistical bit score generated by BLASTn, [e]Statistical e-value generated for highest scoring pair by BLASTn, [f]Percent similarity between subject and CDHC identified organism, [g]Statistical e-value of the first incorrectly identified organism. [h]genus and species of first incorrectly identified organism.

in graphical and statistical form. The graphical form was used to ensure that the majority of the query sequence entered into the search was subsequently aligned with the high scoring database sequence returned. Following this initial search the high scoring database sequence and the PCR product (query or unknown sequence) can be aligned again using pair-wise (Martinez or NW Method) alignment. The Martinez (Martinez, 1988) algorithm identifies regions of perfect match between two sequences, while the Needleman-Wunsch (Needleman et al., 1970) method then optimizes the alignments between regions containing perfect matches. This second alignment is performed separately from the alignment generated by BLASTn, which is often incomplete. This step may not be necessary if the entire query sequence length corresponds to the alignment length detailed in the BLAST statistics.

RESULTS AND DISCUSSION

For the phylogenetic analysis maximum parsimony was utilized for evolutionary analysis for the reason best stated by “Occam’s razor”: the best explanation of the data is the simplest i.e. maximum parsimony requires the fewest *ad hoc* assumptions. It was also felt that minimum evolution or maximum likelihood methods would be unlikely to produce statistically better trees as the initial trees that were generated and based upon the initial alignments performed, were very similar to the final tree generated using maximum parsimony and bootstrapping. The length of the terminal branches along with the short internal internodes of the final tree might have indicated that maximum likelihood would be a more appropriate optimization method. However, given the low levels of divergence expected within the data set, the use of maximum parsimony prevailed. The resulting phylogram is shown in Figure 1.

The goal of the phylogenetic analysis was to create an index showing relationships of the Encephalitozoonidae. We can see from the phylogram that *Encephalitozoon intestinalis*, *Encephalitozoon hellem* and *Encephalitozoon cuniculi* (strains) are all very closely related. The similarity index provided by this phylogram can now be used to evaluate the ability of Computer Database Homology Comparison (CDHC) to determine the species of human pathogenic microsporidia. The great importance of microsporidia identification is due to the great need not only to determine genus, but also to differentiate between species of microsporidia both in clinical settings and in environmental settings. For instance, when screening agricultural water samples for human pathogenic microsporidia, this is especially critical due to the potential presence of literally thousands of other genus/species. Thus, when using molecular methods such as the PCR for environmental detection of microsporidia, such an approach for species determination must be able to differen-

Table 2. Results of CDHC for the non-*Encephalitozoon* microsporidia including the human pathogenic *Vittaformacorneae*.

Test Organism[a]	BLASTn ID[b] Accession number[c]	BLAST bit Score[d]	BLAST e-value[e]	BLAST Identity[f]	e-value[g] negative ID[h]
Vittaforma corneae	*V. corneae*	694	0.0	99	5e-86/*E. schubergi*
Endoreticulatus	*Pleistophora sp.*	694	0.0	99	e-168/*Pleistophora sp.*
schubergi	*E .schubergi*	678	0.0	98	
Nosema trichoplusiae	*N. trichoplusiae*	647	e-118	99	e-109/*N. furnacalis*
Vairimorpha necatrix	*N. necatrix*	678	0.0	99	e-176/Vairimorpha sp.

This table provides the results of PCR-CDHC for the different varieties of non-human pathogenic microsporidia including the very closely related species identified using phylogeny. The ATCC designation of genus and species is indicated in the first column. In the second column is the primary CDHC identification along with the appropriate accession number of this hit. In columns 3-5 we provide the CDHC statistics that are most relevant for evaluation of the results. In the final column we provide the first false identification of the CDHC and its associated E-value. We can see by comparing the E-values of the top hit to the first incorrect identification that the method is robust and able to differentiate among the closely related species. [a]Organisms from pure culture used as subjects for the PCR sequencing CDHC validation, [b]CDHC genus and species determined for subjects by CDHC, [c]Accession number for database sequence identified by CDHC, [d]Statistical bit score generated by BLASTn, [e]Statistical e-value generated for highest scoring pair by BLASTn, [f]Percent similarity between subject and CDHC identified organism, [g]Statistical e-value of the first incorrectly identified organism, [h]genus and species of first incorrectly identified organism.

tiate the human pathogenic microsporidia from the non-pathogenic species. Application of CDHC was described by et al. (Dowd et al., 1998) and later applied by Thurston-Enriquez et al. (Thurston-Enriquez et al., 2002).

The results of the current study show that in each case the correct species of *Encephalitozoon* was identified by the BLASTn database search (Table 1). The results of the PCR-CDHC for the non-*Encephalitozoon* are indica-ted in Table 2. A positive identification using CDHC was generated for *V. corneae*, e value 0.0 and a score of 694. *E. schubergi* was correctly identified though in this same BLAST search *Pleistophora sp.* ATCC 50040 also pro-duced the same e value of 0.0 (scores of 678 and 694, respectively) and included complete query- hit alignments. Subsequent alignments scored at 99% or better and all included the full range of the sequenced PCR product. Although *Pleistophora sp*., listed as ATCC 50040, was originally submitted as belonging to this genus it was later found to be of the genus *Endoreticulatus* and the species *schubergi,* thus explaining these two results. Also, CDHC analyses of *N. trichoplusiae* generated identities of both *N. trichoplusiae* and *N. bombycis* with e values of -118 and scores of 426 bits with full input sequence align-ments. According to the phylogram, these two organisms are very closely related. In fact, Pieniazek *et al.* (Pienia-zek et al., 1996), called *N. trichoplusiae* a “synonym of *N. bombycis*” due to their analysis of ribosomal RNA se-quences and suggested that *N. trichoplusiae* be consi-dered a subspecies of *N. bombycis*. The phylogenetic analysis done for this research supports their findings. Not only morphological and host-parasite associations, but also molecular data should be considered in the taxonomic evaluation of microsporidia. Finally, *Varimo-rpha necatrix* was identified as *N. necatrix* and *Vairimor-pha sp*. with e values of 0.0 and full input sequence align-ments. Many of the *Vairmorpha sp.* have been classified as two different species until it was found that the organisms were dimorphic, i.e., capable of producing two spore types. Because of their dimorphic lifestyle these organisms were placed in a new genus, *Varimorpha.* As a result, caution must be taken in the interpretation of BLASTn and CDHC results since changes in classify-cation are continuously taking place (especially with microsporidia) and not always corrected for in the public databases. Nonetheless, CDHC was still able to deter-mine the correct species from which the various ampli-cons were amplified. The CDHC results of the non-Ence-phalitozoonidae are shown in Table 2. With the Encepha-litozoonidae there are over 22 base pair differences between the two closest relatives (*E. intestinalis* and *E. hellem*). Thus, given the ability of the CDHC to distin-guish not only between the species of microsporidia it can be seen that this is a very efficient and sensitive means to determine the species of fungi.

Two environmental samples were analyzed for micro-sporidia. The results of these analyses can be used as an example of the potential for false identification of human pathogenic microsporidia if the guidelines for the inter-pretation of CDHC are not careful followed. Two separate swine lagoon wastewater samples were collected and assayed for microsporidia using three different PCR pro-tocols. The two primer sets designed for this study’s phylogenetic analysis, the universal and pathogen-speci-fic primer sets, and a primer set widely used for micro-sporidia detection in clinical and environmental samples (Fedorko et al., 1995). CDHC followed the PCR reactions and results of this analysis are shown in Table 3. Both environmental samples showed positive PCR reactions using each of the primer sets. Following BLAST of the resultant sequenced PCR product, the Fedorko et al (Fedorko et al., 1995) primer apparently identified *E. bienusi* (based on highest % similarity and lack of gaps). However, the percent identity, bit score, and number of gaps were 91%, 101, and 0/79, respectively. Therefore

Table 3. Results from PCR-CDHC analysis of a swine lagoon wastewater sample using three primer sets developed for microsporidia species detection.

Sample	Primer set	CDHC Results for Swine Lagoon Wastewater Sample NCBI nt (June 2005)	Interpretation
Swine Lagoon Sample 1	"Pathogen-Specific" 300 bp PCR product	100% Identity (48 bits; 0/24 gaps): *Uncultured forest soil bacteria*	False identification Unknown or putative fungi
		96% Identity (54 bits; 0/31 gaps): *Orthosomella operophterae*	
	"Universal" 500 bp PCR product	100% Identity (329 bits; 0/166 gaps): *Polytoma sp.*	False identification Unknown or putative fungi
	"Fedorko et al. (1995)" 300 bp PCR product	**91% Identity (101 bits; 0/79 gaps): *Enterocytozoon bieneusi*	False identification Unknown or putative fungi
		87% Identity (107 bits; 1/118 gaps): *Orthosomella operophterae*	
Swine Lagoon Sample 2	"Pathogen-Specific" 200 bp PCR product	96% identity (44.1 bits; 0/26 bits): *Deinococcus radiodurans*	False identification Unknown or putative fungi
		93% Identity (44.1 bits; 0/30 gaps): *Methanosarcina acetivorans*	
	"Universal" 500 bp PCR product	89% Identity (89.7 bits; 0/77 gaps): *Geminella sp*.	False identification
		89% Identity (85.7 bits; 0/75 gaps): *Fusaruim sp.*	Unknown or putative fungi
	"Fedorko et al. (1995)" 300 bp PCR product	100% Identity (42.1 bits; 0/21 gaps) *Homo sapiens* cDNA	False identification Unknown or putative fungi
		95% Identity (40.1 bits; 0/24 gaps): *Mus musculus* chromosome	

This table provides the results of PCR-CDHC for 2 separate swine wastewater samples. These samples are expected to have both fungi and non-pathogenic microsporidia present. In each case an appropriate sized PCR product was generated. However upon sequencing and CDHC it is shown that these samples do not contain human pathogenic microsporidia. It is important to note that cloning of these products prior to sequencing may be suggested especially in wastewater samples. In this case evaluating 10-20 clones from each sample will provide better ability to identify among different populations. **Particularly important is the use of appropriate guidelines for species determination. In this case a human pathogenic species of microsporidia was identified. Yet when considering the bit score and e-value we must conclude that this is not *Enterocytozoon bieneusi* but a distant relative. This table exemplifies the ability of PCR-CDHC to differentiate species and the importance of following appropriate guidelines in evaluating results of this method.

Therefore based upon the low identity (requires 97% of greater), lack of alignment length, and the low bit score it is unlikely that the identification of this product as *E. bienusi* is accurate. The target size for these primers and *E. bienusi* is 268 bp yet only 79 bp aligned with the *E. bienuesi* database sequence. At the very least we would expect alignments in the range of 210-240 bp if this were an actual identification. This would produce bit scores of roughly 230 with a Blossum62 matrix. These results exemplified by the supposed detection of *E. beinusi*, elucidate the potential for false identification and the need for careful examination of BLAST results when performing CHDC. This is because, by design, the BLAST search engine will return results for almost any sequence entered (depending on the initial E-value selected).

Thus, the following guidelines for the use of CDHC as a method for the determination or identification of microsporidia after PCR and sequencing should be followed in order to prevent misinterpretation of the results. The BLAST search should return complete alignments for the entire query (unknown) sequence entered into the BLAST engine. This means that if a 300 bp query sequence is entered, then the BLAST search should return approximately a 300 bp alignment. Most fungi have regions of homology within the small ribosomal subunit so partial matches cannot be considered relevant for species determination. If in doubt, the top hit sequence should be downloaded from the database and a local alignment performed using an appropriate algorithm as discussed in the materials and methods. Secondly, even if the complete query sequence is included in the alignment there should be a stringent cut-off for % similarity. With the negligible error rate of both PCR enzymes and enzymes used in cycle sequencing, combined with the potential for errors in base calling software we can, at the extreme of conservative estimation, expect no more than 3 out of 100 bases to have errors. Thus, if the sequenced 300 bp PCR product (query sequence) is fully aligned with the high scoring sequence derived from the BLAST search (HSP) in the database, we can be reasonably confident in our identification if the subsequent percentage match for pair-wise alignment between the query and database sequence is 97% or greater. It is important to note that our use of TaqGOLD polymerase which has no proof reading function was based upon previous work. Modern enzymes with proof reading ability are now suggested to

provide improved results and reduce potential errors during enzymatic amplification. Publications that utilize BLAST as a method for species determination should include in their results not only the e-value of the BLAST, but also the % identity, the bit score, PCR product size, the query lengths, and the hit length (or alignment length). The PCR sequences derived should also be submitted to GenBank (www.ncbi.nlm.nih.gov) either as unknown or putative microsporidia or with putative identification at the genus level (e.g. putative *Enterocytozoon* spp.).

When reporting or submitting sequence data to GenBank, extreme care should be utilized in how the sequences are annotated. Authors should remain conservative in their naming of environmental or unknown sequence isolates. For example, if PCR -CDHC determines that *E. intestinalis* was detected in swine feces, the use of "presumptive *Enterocytozoon* spp. originating from swine fecal sample E-value -118 bit score 300" as an annotation might be appropriate rather than using annotations that indicate absolute identifications such as "*E. intestinalis* derived from swine fecal sample".

This is especially important when using only portions of SSU-RNA as methods of identification.

Conclusions

This study has used the phylogenetic tree generated as a similarity index and way to challenge the CDHC methodology. Because the length of the branches connecting any two organisms on the tree is related to their sequence similarity we can see that the three species of *Encephalitozoon* are very closely related. These results, in conjunction with the results of the CDHC performed in this study using known controls, allow us to justify the use of PCR-CDHC as a method for specifies determination of microsporidia. The CDHC method is highly significant and illustrates the problems with other published molecular methods used for species determination of microsporidia such as restriction pattern analysis, probes, or nested PCRs. These methods would be unable to differentiate between closely related species with statistical validity as they rely on much shorter sequences of the PCR product. Probes for instance are generally in the 30 - 50 nucleotide size ranges, whole nested PCR only rely on the presence of 20 - 24 oligonucleotides internal to the original two primers as a confirmation assay. CDHC looks at the entire PCR product generated nucleotide-by-nucleotide providing a reliable method for species determination of microsporidia detected by PCR. Guidelines for the use of CDHC are provided along with recommendations for reporting results in databases and publications.

ACKNOWLEDGEMENT

This research was supported by SED's USEPA STAR fellowship.

USDA Disclaimer: The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the United States Department of Agriculture.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EWan, Lipman DJ (1990). Basic local alignment search tool. J. Mol. Biol. 215: 403-410.

Canning EU and Hollister WS (1987). Microsporidia of mammals--widespread pathogens or opportunistic curiosities? Parasitol. Today. 3: 267-273.

Desportes I, Le CY, Galian A, Bernard F, Cochand-Priollet B, Lavergne A, Ravisse P, Modigliani R (1985). Occurrence of a new microsporidan: Enterocytozoon bieneusi n.g., n. sp., in the enterocytes of a human patient with AIDS. J. Protozool. 32: 250-254.

Dowd SE, Gerba CP, Enriquez FJ, Pepper IL (1998). PCR amplification and species determination of microsporidia in formalin-fixed feces after immunomagnetic separation. Appl. Environ. Microbiol. 64: 333-336.

Dowd SE, Gerba CP, Kamper M, Pepper IL (1999). Evaluation of methodologies including immunofluorescent assay (IFA) and the polymerase chain reaction (PCR) for detection of human pathogenic microsporidia in water. J. Microbiol. Methods 35: 43-52.

Dowd SE, Gerba CP, Pepper IL (1998). Confirmation of the human-pathogenic microsporidia Enterocytozoon bieneusi, Encephalitozoon intestinalis, and Vittaforma corneae in water. Appl. Environ. Microbiol. 64: 3332-3335.

Dowd SE, John D, Eliopolus J, Gerba CP, Naranjo J, Klein R, Lopez B, de MM, Mendoza CE, Pepper IL (2003). Confirmed detection of Cyclospora cayetanesis, Encephalitozoon intestinalis and Cryptosporidium parvum in water used for drinking. J. Water Health. 1: 117-123.

Fayer R, Santin M, Trout JM (2003). First detection of microsporidia in dairy calves in North America. Parasitol. Res. 90:383-386.

Fedorko DP, Nelson NA, Cartwright CP (1995). Identification of microsporidia in stool specimens by using PCR and restriction endonucleases. J. Clin. Microbiol. 33:1739-1741.

Lores B, del A, Arias C (2002). Enterocytozoon bieneusi (microsporidia) in faecal samples from domestic animals from Galicia, Spain. Mem. Inst. Oswaldo Cruz. 97: 941-945.

Martinez HM (1988) A flexible multiple sequence alignment program. Nucleic Acids Res. 16: 1683-1691.

Needleman SB, Wunsch CD (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48: 443-453.

Pieniazek NJ, da Silva AJ, Slemenda SB, Visvesvara GS, Kurtti TJ, Yasunaga C (1996) Nosema trichoplusiae is a synonym of Nosema bombycis based on the sequence of the small subunit ribosomal RNA coding region. J. Invertebr. Pathol. 67: 316-317.

Sparfel JM, Sarfati C, Liguory O, Caroff B, Dumoutier N, Gueglio B, Billaud E, Raffi F, Molina JM, Miegeville M and Derouin F (1997). Detection of microsporidia and identification of Enterocytozoon bieneusi in surface water by filtration followed by specific PCR. J. Eukaryot. Microbiol. 44: 78S.

Thurston-Enriquez JA, Watt P, Dowd SE, Enriquez R, Pepper IL, Gerba CP (2002). Detection of protozoan parasites and microsporidia in irrigation waters used for crop production. J. Food Prot. 65: 378-382.