DNA barcoding parasite organisms found in terrestrial mammal scat using COI sequence data

BIOS 35502: Practicum in Environmental Field Biology

Derryl Miller
Advisor: Dr. Andy Mahon

2009

ABSTRACT

While invasive techniques such as trapping or marking animals for study are a concern for the safety and preservation of established ecosystems, molecular techniques have provided accurate information about population dynamics and the interaction between many species within an ecosystem, oftentimes without unnecessary human interaction. Molecular techniques, involving the sequencing of mitochondrial DNA such as the Cytochrome c Oxidase subunit I (COI), are extremely versatile in the study of invasive, endangered, and evasive or dangerous species. This is possible because the COI is a relatively short strand of DNA which can be found in scat, hair, tissue, and virtually anywhere where target organisms shed cells. This study focuses on the barcoding of parasites found in the scat of terrestrial mammals at the University of Notre Dame Environmental Research Center (UNDERC). From the sequencing and analysis of numerous parasites and the generation of a phylogenic tree, this work finds that a fragment of the COI gene is effective for the analysis of both identity of a specimen and the relatedness between species of parasites (Bayesian statistic > 0.5 for numerous groups) at UNDERC with the help of the National Center for Biotechnology Information (NCBI) database.

INTRODUCTION

Two of the largest concerns in ecology are the discovery and preservation of biodiversity around the world (Colwell and Codington 1994). Within the study of biodiversity of organismal groups, many correlations have been found between a species' size, population density, and ability to travel passively and actively to new areas with their level of overall diversity across a region (Martiny et. al. 2006). As this broad and lofty goal, to discover and map worldwide biodiversity, has been identified by the scientific community, several studies have made efforts to map terrestrial and aquatic biodiversity (e.g. Colwell and Coddington 1994; Burton and Davie 2007; Herbert et. al. 2004). Surprisingly, relatively few studies have addressed the biodiversity of small invertebrates, interstitial organisms, or microorganisms. This is primarily because many types and strains of these organisms have a significantly smaller body size, often resulting in higher population densities. Additionally, the overall trends for the genetic diversification and extinction of these small organisms is still largely unknown (Colwell and Coddington 1994).

Interestingly, the goal of maintaining and studying biodiversity is interrelated with the study of small organisms because these tiny creatures often exist within a host as parasites and can easily be detected with various DNA extraction, amplification, and sequencing techniques described in the literature (Monis and Andrews 2002). Furthermore, studying small parasitic organisms,

which rely on host macroorganisms, can provide a good indication of overall biodiversity at a study site (Monis and Andrews 2002).

There is also further interest in DNA barcoding of parasite biodiversity because of its noninvasive nature and its capability for early detection of invasive species, non-native species which may offset the balance of a pre-established ecosystem where they are unfamiliar (Ficetola et. al. 2008). Catching invasive species early, through the discovery of the introduction of new parasites into an ecosystem, may be a useful technique for the preservation of established ecosystems from dangerous, non-native species.

The overall purpose of this project was to utilize molecular techniques such as DNA extraction, polymerase chain reaction (PCR), and DNA sequencing to barcode parasite species at UNDERC, found in host mammal scat on the property. After DNA extraction, the barcoding will be performed through DNA amplification in the presence of universal oligonucleotide primers, specific to the mitochondrial portions of the DNA. Barcoding of invertebrate parasites by specifically utilizing the Cytochrome c Oxidase subunit I gene (COI) of the mitochondrial DNA is quite common in the literature (e.g. Ficetola et. al. 2008; Herbert et. al. 2003; Gianmarco et. al. 2009; Lunt et. al. 1996).

Cytochrome c Oxidase subunit I is an especially useful portion of DNA because it is very diverse, capable of identifying many different species from a smaller portion of the DNA found in the field. This gene has also been shown to

differ about 5% between closely related species of invertebrate taxa (e.g. Mahon et. al. 2008; Thornhill et. al. 2008). Furthermore, long-term goals of this study include the identification and study of biodiversity and its preservation at UNDERC.

METHODS

The purpose of this project was to provide new information as to the diversity of the population of invertebrate parasites present at the UNDERC property, found in host mammal scat. In order to identify parasite species at UNDERC, these small organisms were collected from the scat of various mammal species in an opportunistic fashion (Figure 1), wherever they could be found and identified with the help of Elbroch's field guide (2003). Study sites were varied across the UNDERC property as much as possible, and GPS coordinates were taken at the site of each scat sample (Table 1). As scat samples were recovered, the samples were processed through 500 micron/millimeter and 125 micron/millimeter sieves, and parasites contained within the scat samples were preserved in 95% ethanol prior to DNA extraction. After scat samples were recovered, the DNA of each parasite specimen from the preserved samples was extracted using a Qiagen DNEasy extraction kit according to the manufacturer's recommendations (Qiagen Inc., Valencia, CA).

PCR amplification of an approximately 600 base pair fragment of the mitochondrial COI gene was performed using universal invertebrate primers, HCO and LCO (Folmer et. al. 1994). Individual PCR samples contained 1 μL sample DNA to 24 μL of Master Mix, comprised of 2.5 μL TAQ Buffer, 0.5 μL dNTP, 2.5 μL Mg^{2+}, 0.5 μL HCO, 0.5 μL LCO, 17.35 μL H_2O, and 0.15 μL TAQ. PCR samples were incubated using COI amplification protocol (Table 3). After PCR amplification, successful attempts were photographed with a gel camera, and the resulting DNA was purified using a Qiagen Qiaquick PCR Gel extraction kit according to the manufacturer's recommendations (Qiagen Inc., Valencia, CA). Prior to sequencing these DNA samples, a second PCR amplification was run containing 2 μL DNA sample to 5 μL Master Mix, comprised of 1 μL TAQ Buffer, 0.32 μL HCO or LCO, 2.68 μL H_2O, and 1 μL Big Dye. The samples of the second PCR were amplified using ABI protocol (Table 4). The resulting DNA samples were then sequenced using an Applied Biosystems ABI3700 Gene sequencer at the University of Notre Dame.

DNA sequences were aligned and screened using Bioedit (v7; Hall 1999) and analyzed with a suite of molecular tools, such as Bayesian analysis, the parsimony based TCS haplotype analysis, and sequence divergence calculations. In order to construct a phylogenic tree, Mr. Modeltest was used to generate a model of evolution prior to Bayesian analysis by Mr. Bayes, indicating statistical support for individual clades within the phylogenic tree. The phylogenic tree was

then visualized using Treeview. The sequencing of the COI in the parasites will therefore identify individual species and, though Bayesian analysis, provide a phylogenic tree, indicating the genetic relationship between parasite specimens found during this study.

RESULTS

After DNA extraction, amplification, and purification, forty eight DNA samples were sent back to Notre Dame for sequencing (Genomic sequencing facility, Dept. of Biological Sciences). Of these forty eight DNA samples, nineteen were successfully sequenced, and these sequences were compared to other COI genes of morphologically similar organisms from the NCBI database. Many of the data sequences collected at UNDERC were found to closely resemble the COI sequences from NCBI with 80%-93% similarity in nucleotide base pairing (Table 2).

Furthermore, the identity of many of the parasite species having a close match of the COI gene were closely related to the parasite in question, taxonomically. This was determined using Bioedit (v7; Hall 1999) and Bayesian analysis, from which a phylogenic tree was constructed (Figure 2). Within the phylogenic tree, clades of Isopoda, Hymenoptera, Coleoptera, Oligochaeta, Nemotoda, and an Unknown order were each found to have statistically significant genetic relationships (Bayesian statistic > 0.5) between sequence data

of parasites found at UNDERC with the published sequences of the NCBI database.

Our hypotheses, that sequencing and statistical comparison through Bayesian analysis between the COI subunit of parasite specimens at UNDERC and NCBI sequence data can accurately predict identity of parasite specimens found in the scat of mammals, is accepted. After the construction of a phylogenic tree, it is furthermore confirmed that the COI can accurately provide statistical support of the relatedness between the identified invertebrate parasites found at UNDERC.

DISCUSSION

As the ecological community continues to realize the depth of the problem of mapping species worldwide to preserve biodiversity, molecular techniques continue to aid in the discovery of new species, the taxonomic arrangement or rearrangement of already discovered species, and even finding statistical evidence of the genetic relatedness between specimens through the construction of phylogenic trees. The COI gene fragment, being a genetically diverse and relatively easy strand to analyze, has been a standard barcoding strand, used around the world to discover and organize global biodiversity. In this project, the COI gene successfully predicted the identity of parasite specimens found in the scat of terrestrial mammals at UNDERC. Analysis of the COI also provided

statistical evidence for the relatedness of these specimen to others found all over the world, recorded in the NCBI.

While the Bayesian analysis of sequence data from this project did yield statistically significant relationships (Bayesian statistic > 0.5) within clades of Isopoda, Hymenoptera, Coleoptera, Oligochaeta, Nemotoda, and an Unknown order, this statistical evidence was not as strong as was expected. In order to increase the effectiveness of the Bayesian analysis, a larger sample size of sequences with greater sequence quality, more nucleotides present per COI sequence, is needed, but, given the financial and time constraints on this project, finding sequence data and statistically relating these sequences to the published sequences of the NCBI database was a success. The Unknown order is not the discovery of a new order, but rather is an unresolved relationship containing organisms from Diptera, Coleoptera, and Decapoda. Because of this ambiguity, it is difficult to label the Unknown clade; however, with a larger sample size and higher quality COI strands from DNA sequencing, this clade would likely resolve to provide statistical support for the relationship of the sequences within this clade to Diptera, Coleoptera, or Decapoda.

Future studies involving the sequencing of parasites and their host macroorganisms using the COI include examining the species richness of parasites found in scat across the property with the presence of physical barriers. These barriers include Tenderfoot Creek or the road systems in the area, which

may inhibit the transportation of parasite species in mammal scat, which may cause parasite populations to differ significantly on either side of these physical barriers. While out collecting, the majority of predator scat, coyote, wolf, and bear, were found out in the open; however, the scat of herbivores at UNDERC, such as deer, porcupine, and snowshoe hair were almost exclusively found in the woods under the cover of trees. This fact might also facilitate the study of parasites found in the scat of terrestrial predators and herbivores separately to see if their diet or social behavior increases susceptibility to parasitism at UNDERC (Rohde 1994). Additionally, there are many ways to track animals, including sign and footprints; however, some sign that is discovered has been either mutilated by the weather or is rendered unrecognizable. Another question to be answered concerning mammal sign is whether or not enough DNA can be extracted from mammal scat to provide genetic proof that the scat belongs to a specific mammal species. This study would be especially useful in examining the distribution or abundance of dangerous, evasive, or endangered mammals in an area, through noninvasive molecular techniques (Janecka et. al. 2008; Fernandes et. al. 2008).

There are many practical applications to the field of molecular biology and DNA barcoding. Mapping biodiversity, detecting invasive or endangered species in an environment early to preemptively begin conservation efforts, and observing the interaction of host and parasite species in an environment are only a few capabilities of molecular techniques. Molecular techniques can also be

exclusively noninvasive, which is ideal for preserving established ecosystems while providing information as to the genetic diversity of a population without catching or disturbing organisms in their habitat. There is an increasing potential for conservation efforts of biodiversity worldwide as our knowledge of global biodiversity increases, while our power to discover new genetic information with statistical support increases significantly as online databases like the National Center for Biotechnology Information (NCBI) continue to grow with newly published molecular literature.

ACKNOWLEDGEMENTS

I would like to sincerely thank my research mentor, Dr. Andy Mahon, for his excellent help and advice throughout this project. I would also like to thank Dr. Michael Cramer, Dr. Heidi Mahon, Dr. Dave Choate, Dr. Gary Belovsky, Dr. Jessica Hellmann, Dr. Sunny Boyd, Dr. Todd Crowl, Dr. Walt Carson, and Andrew Perry for their continued support throughout this course in experiencing field research. In addition, I would like to thank the University of Notre Dame Biology Department for the opportunity to participate in Practicum in Environmental Field Biology (UNDERC), and the entire Hank family, who graciously continue to financially support research at UNDERC with the Bernard J. Hank Family Endowment.

LITERATURE CITED

Burton, T. E. and Davie, P. J. F. 2007. A revision of the shovel-nosed lobsters of the genus Thenus (Crustacea : Decapoda : Scyllaridae), with descriptions of three new species. *Zootaxa*. (1429): 1-38.

Colwell, R. K.; Coddington, J. A. Estimating Terrestrial Biodiversity through Extrapolation. 1994. *Phil. Trans. R. Soc. Lond.* 345:101-118.

Elbroch, Mark. 2003. Mammal Tracks & Sign: A Guide to North American Species. Stackpole Books: Pennsylvania.

Fernandes, Carlos A.; Ginja, Catarina; Pereira, Iris. 2008. Species-specific mitochondrial DNA markers for identificationof non-invasive samples from sympatric carnivores in the Iberian Peninsula. *Conserv Genet*. 9:681–690.

Ficetola, G. F.; Bonini, A.; Miaud, C. 2008. Population Genetics Reveals Origin and Number of Founders in a Biological Invasion. *Molecular Ecology*. 17: 773-782.

Ficetola, G. F.; Miaud, C.; Pompanon, F.; Taberlet, P. 2008. Species Detection Using Environmental DNA from Water Samples. *Biol. Lett.* 4: 423-425.

Folmer, O.; Black, M.; Hoeh, W; Lutz, R.; Vrijenhoek, R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol*. 3:294–299.

Gianmarco, Ferri; Milena, Alù; Beatrice, Corradini; Manuela, Licata; Giovanni, Beduschi. 2009. Species Identification Through DNA ‘‘Barcodes.’’ *Genetic Testing and Molecular Biomarkers*. 13(3): 421-426.

Herbert, P. D.; Cywinska, A.; Ball, S. L.; DeWaard, J. R. 2003. Biological Identifications through DNA Barcodes. *Proc. R. Soc. Lond.* 270: 313-321.

Herbert, Paul D. N.; Stoeckle, Mark Y.; Zemlak, Tyler S.; Francis, Charles M. 2004. Identification of Birds through DNA Barcodes. *Plos Biology*. 2(10):1657-1663.

Janecka, J. E.; Jackson, R.; Yuquang, Z.; Diqiang, L.; Munkhtsog, B.; Buckley-Beason, V.; Murphy, W. J. 2008. Population monitoring of snow leopards using noninvasive collection of scat samples: a pilot study. *Animal Conservation*. 11(5): 401-411.

Lunt, D. H.; Zhang, D. X.; Szymura, J. M.; Hewitt, G. M. 1996. The insect cytochrome oxidase I gene: Evolutionary patterns and conserved primers for phylogenetic studies. *Insect Molecular Biology*. 5(3):153-165.

Mahon, A. R.; Arango, C. P.; Halanych, K. M. 2008. Genetic diversity of *Nymphon* (Arthropoda: Pycnogonida: Nymphonidae) along the Antarctic Peninsula with a focus on *Nymphon australe*. *Marine Biology*. 155: 315-323.

Martiny, J. B. H.; Bohannan, B. J. M.; Brown, J. H.; Colwell, R. K.; Fuhrman, J. A.; Green, J. L.; Horner-Devine, M. C.; Kane, M.; Krumins, J. A.; Kuske, C. R.; Morin, P. J.; Naeem, S.; Ovreas, L.; Reysenbah, A.; Smith, V. H.; Staley, J. T. 2006. Microbial Biogeography; Putting Microorganisms on the Map. *Nature Publishing Group.* 4: 102-112.

Monis, P. T.; Andrews, R. H.; Saint, C. P. 2002. Molecular Biology Techniques in Parasite Ecology. *International Journal for Parasitology.* 32: 551-562.

Rohde, K. 1994. Niche Restriction in Parasites-Proximate and Ultimate Causes. *Parasitology*. 109: 69-84.

Thornhill, D. J.; Mahon, A. R.; Norenberg, J. L.; Halanych, K. M. 2008. Open-ocean barriers to dispersal: a test case with the Antarctic Polar Front and the ribbon worm *Parborlasia corrugatus* (Nemertea: Lineidae). *Molecular Ecology*. 17(23): 5104-5117.

TABLES

Table 1: A table relating scat location and mammal identity to the number and kinds of parasites found within each scat sample collected in the field. Generic descriptions of the parasites found within are included as best as they could be determined. GPS coordinates where each scat sample was discovered are included. See Figure 1 for approximate locations on a map of the property.

Site #	Mammal Scat I.D.	Parasites Found	General Description	GPS Coordinates
1	Wolf	1	Gastropoda	16T 0304256 UTM 5124255
2	Coyote	0		16T 0304122 UTM 5124453
3	Coyote	0		16T 0304269 UTM 5125023
4	Coyote	0		16T 0304333 UTM 5125155
5	Coyote	1	Gastropoda	16T 0305019 UTM 5125673
6	Coyote	2 2 2	Gastropoda Coleoptera Unknown	16T 0305837 UTM 5125439
7	Coyote	1 1 8	Gastropoda Oligochaeta Unknown	16T 0305965 UTM 5125622
8	Wolf	9	Gastropoda	16T 0306006 UTM 5125649
9	Wolf	1 1 2 3	Gastropoda Oligochaeta Coleoptera Unknown	16T 0306072 UTM 5125671
10	Coyote	1 1	Gastropoda Coleoptera	16T 0306123 UTM 5125720
11	Coyote	1 1	Gastropoda Chilopoda	16T 0306128 UTM 5125720
12	Unable to Identify	1	Gastropoda	16T 0306249 UTM 5125847
13	Coyote	3	Gastropoda	16T 0306594 UTM 5125885

14	Coyote	1	Gastropoda	16T 0306728 UTM 5125803
15	Coyote	3 6 6	Gastropoda Coleoptera Unknown	16T 0306805 UTM 5125793
16	Wolf	1	Gastropoda	16T 0306959 UTM 5125762
17	Coyote	2 2	Gastropoda Unknown	16T 0307268 UTM 5125578
18	Coyote	0		16T 0307665 UTM 5125621
19	Deer	1 1	Coleoptera Unknown	16T 0307601 UTM 5125727
20	Porcupine	1 1	Coleoptera Unknown	16T 0307595 UTM 5125815
21	Snowshoe Hair	0		16T 0307358 UTM 5125757
22	Coyote	2	Gastropoda	16T 0306025 UTM 5125666
23	Bear	0		16T 0305415 UTM 5122211
24	Deer	2	Coleopteraa	16T 0305361 UTM 5122155
25	Coyote	1	Hymenoptera	16T 0304023 UTM 5123635
26	Coyote	6	Unknown	16T 0304017 UTM 5123609
27	Wolf	1 6 3 4	Coleoptera Nematoda Unknown Unknown	16T 0304003 UTM 5123570
28	Coyote	1 1 1	Unknown Unknown Unknown	16T 0304001 UTM 5123565
29	Coyote	2 1 6	Coleoptera Hymenoptera Unknown	16T 0303948 UTM 5123516
30	Mouse (*Peromiscus Maniculatus*)	0		
31	Coyote	9	Unknown	16T 0303489 UTM 5123508

32	Wolf	1 1	Coleoptera Unknown	16T 0303415 UTM 5123577
33	Coyote	2	Coleoptera	16T 0302753 UTM 5123787
34	Coyote	1	Unknown	16T 0302733 UTM 5123837
35	Coyote	0		16T 0302702 UTM 5123714
36	Deer	1	Unknown	
37	Deer	7 1 3 4	Coleoptera Coleopteran Larvae Unknown Unknown	16T 307600 UTM 5128981
38	Porcupine	5 1	Unknown Unknown	16T 307408 UTM 5122091
39	Porcupine	0		16T 307058 UTM 5122168
40	Coyote	0		16T 307058 UTM 5122168
41	Porcupine	1	Unknown	16T 302702 UTM 5125094
42	Porcupine	3 8	Nematoda Unknown	16T 302702 UTM 5125094
43	Deer	25 12 15 14	Dipteran Larvae Coleopteran Nematodes Unknown	16T 302780 UTM 5125038
44	Bear	25 1 5	Dipteran Larvae Hymenoptera Unknown	16T 0303013 UTM 5125054

Table 2: A table listing the parasites with successfully sequenced COI genes during this project and the percentage of similarity found between the sequenced COI gene and its closest relative, found with the National Center for Biotechnology Information (NCBI) database. Successful parasite sequence names in the phylogenic tree (Figure 2) are also included.

Site #	**Morphological Order of Parasite**	**Closest Relative from NCBI**	**Names in Figure 2**	**% Overlap of COIs**
1	Isopoda	Nothing	H01	0%
7	Coleoptera	*Raymunida erythrina*	G04	84%
7	Oligochaeta	*Dendrobaena octaedra*	E04	92%
7	Coleoptera	Nothing	G01	0%
16	Gastropoda	Nothing	F04	0%
25	Hymenoptera	*Camponotus pennsylvanicus*	H03	89%
27	Isopoda	*Patelloa sp.*	F03	88%
27	Diptera	*Belvosia sp.*	E03	89%
27	Diptera	*Onthophagus clypeatus*	E06	91%
32	Isopoda	*Drosophila moriwakii*	D05	81%
32	Isopoda	Nothing	D02	0%
41	Lepidoptera	*Bembecia himmighoffeni*	C01	83%
41	Lepidoptera	*Quadrus contubernalis*	C04	92%
42	Hymenoptera	*Myrmecocystus melliger*	D04	81%
43	Coleoptera	*Drosophila rubida*	B01	85%
43	Diptera	*Calliphora hilli*	H02	91%
44	Diptera	*Biopyrellia bipuncta*	F02	91%
44	Diptera	*Neomyia coeruleifrons*	F05	93%
44	Hymenoptera	*Schinia pulchripennis*	E05	86%

Table3: A description of COI PCR amplification protocol, including incubating temperatures and time intervals for the Cytochrome c Oxidase subunit I prior to purification. Step 1 is a denaturization phase. Steps 2-4 are repeated thirty times while amplification occurs. Step 5 is an extension phase. Step 6 is a refrigeration step to preserve DNA after PCR is complete. Samples are removed from the Thermocycler and refrigerated during step 6. Total time is about 1 hour and 40 minutes.

Steps	Incubating Temperature (^{O}C)	Time Intervals
1	94	1.00 minute
2	94	30 seconds
3	50	45 seconds
4	72	1.00 minute
5	72	8.00 minute
6	4	Until Refrigerated

Table 4: A description of ABI PCR protocol, including incubating temperatures and time intervals for the Cytochrome c Oxidase sequence I prior to sequencing. Step 1 is a denaturization phase. Steps 2-4 are repeated 44 times while DNA amplification occurs. Step 5 is a refrigeration step to preserve DNA after PCR is complete. Samples may be removed from the Thermocycler anytime during step 5. This protocol is run twice for each DNA sample, using HCO for the first run and LCO for the second run. Total time is about 3 hours and 46 minutes per run.

Steps	Incubating Temperature (^{O}C)	Time Intervals
1	96	5.00 minutes
2	96	10 seconds
3	50	5 seconds
4	60	4.00 minutes
5	4	Until Refrigerated

FIGURES

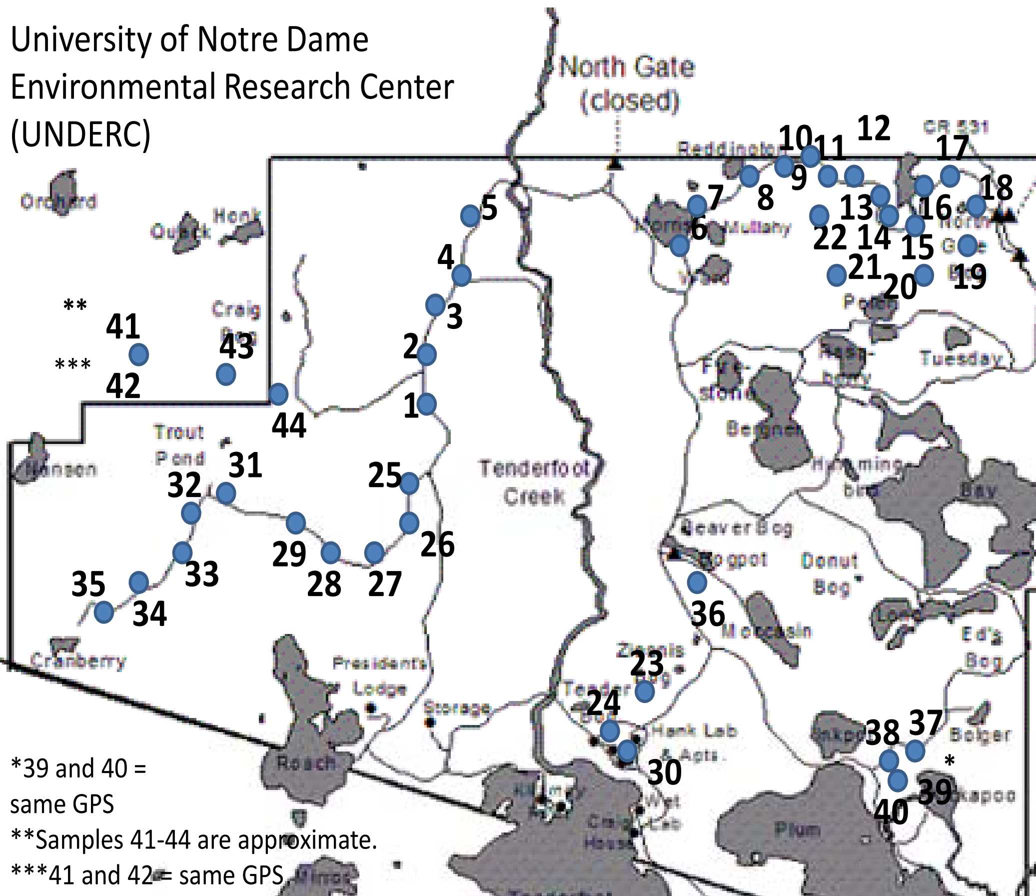

Figure 1: The sites of scat collection across the UNDERC property are labeled on the property map. Refer to Table 1 for information on individual sites as well as GPS coordinates on each site.

Phylogenic Tree

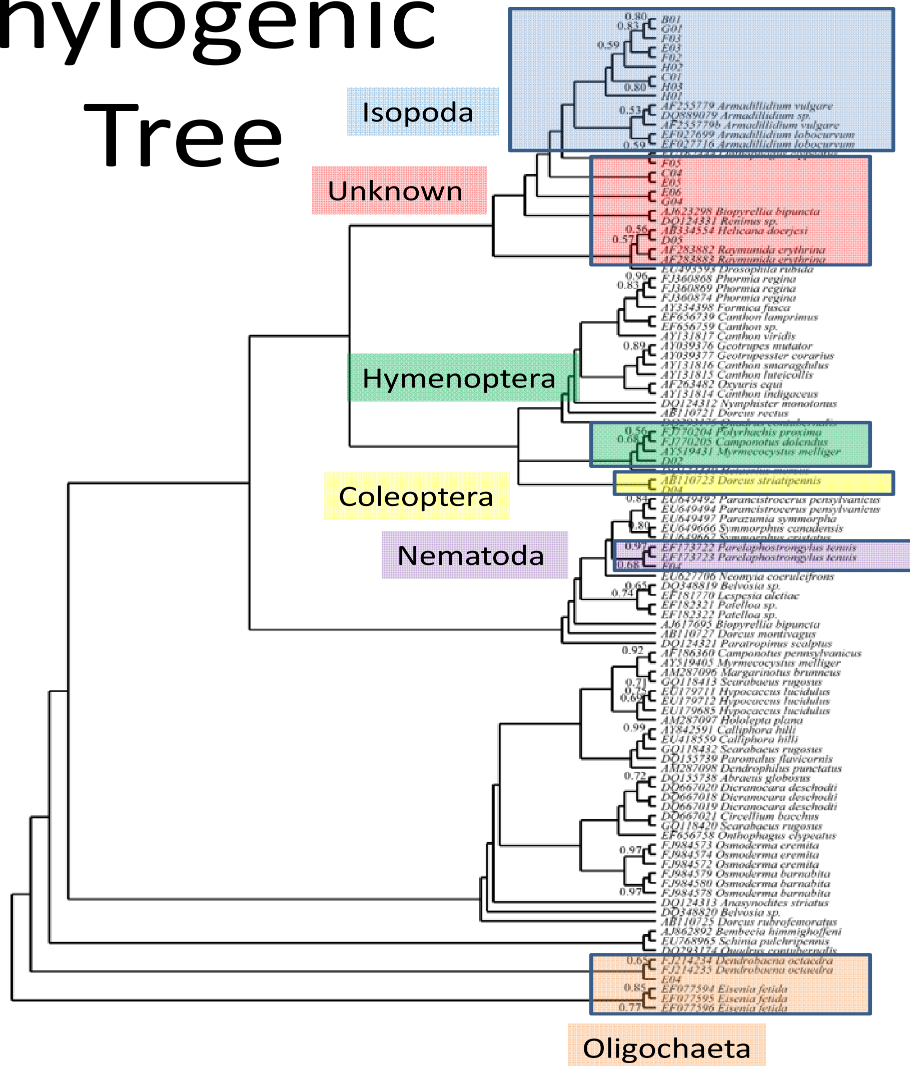

Figure 2: A phylogenic tree associating COI sequences from parasite specimen found at UNDERC with other COI sequences found on the National Center for Biotechnology Information (NCBI) database. Clades of Oligochaeta, Nemotoda, Coleoptera, Hymenoptera, Isopoda, and an Unknown

order were each found within this phylogenic tree to have a statistically acceptable genetic relatedness (Bayesian Statistic > 0.5) based solely on the COI sequence. Only Bayesian statistics > 0.5 remain on this figure.