Development of Genetic Barcodes for Hosts of the Blacklegged Tick (*Ixodes scapularis*) in Southern New York

Christine Zolnik
Department of Biology, Fordham University

Abstract

In the northeastern United States, blacklegged ticks (*Ixodes scapularis*) are responsible for transmitting several disease-causing pathogens to humans, including the causative agents of Lyme disease, Babesiosis, and Anaplasmosis. Studies that examine the role of host species in pathogen transmission cycles are important for understanding tick-borne disease dynamics, and subsequently disease risk to humans. However, most studies that examine the role of host species use indirect measures. Previous work has been conducted on developing molecular protocols to test ticks for bloodmeal sources, thereby identifying the host species for each tick without having to concurrently capture the host animal. Use of these techniques would provide direct evidence as to the host species of each blacklegged tick and elucidate the role that different hosts have on tick-borne disease transmission and risk.

The purpose of this study was to develop genetic barcodes for hosts of the blacklegged tick in southern New York State (an area with high prevalence of blacklegged ticks). Barcodes were created by amplifying a segment of the Cytochrome *c* Oxidase Subunit I (COI) gene from various vertebrate tissue sources and comparing them to published sequences in GenBank. Successful extraction and sequencing of DNA from these tissues generated barcodes for 22 species, five of which were new species barcodes not currently present in GenBank. Furthermore, identification of single nucleotide polymorphisms in the sequences of five species demonstrates genetic variation that may indicate different populations of these species.

Key words: blacklegged ticks, Cytochrome *c* Oxidase Subunit I, COI, genetic barcode

Introduction

In the northeastern United States, blacklegged ticks (*Ixodes scapularis*) transmit a number of tick-borne pathogens to humans including the causative agents of Lyme disease, Babesiosis, and Anaplasmosis (Dumler et al. 2005, Kjemtrup and Conrad 2000, Lane et al. 1991, Speilman et al. 1985, Schwartz et al. 1997). Immature blacklegged ticks are generalist feeders and feed on multiple host species throughout their life. Each potential host species differs in its ability to transmit a pathogen to a tick. Consequently, some host species may be more important for maintaining disease cycles than others. Various studies using indirect measures, such as numbers of ticks found on hosts, have suggested that host species community composition influences infection prevalence in ticks. However, good empirical data supporting this idea are lacking.

Previous studies have developed molecular techniques to test ticks, and other bloodfeeding arthropods, for their host species by identifying sources of host bloodmeals. A reverse line-blot hybridization technique has been used in Europe to determine which hosts the sheep tick (*Ixodes ricinus*; closely related to *I. scapularis*) has fed on, using a segment of the 12S rDNA gene in remnants of its earlier bloodmeal (Humair et al. 2007, Pichon et al. 2003). However this technique requires development of oligonucleotide probes for each species of interest, which can be time consuming and may require a different set of probes for the same species based on geographic area. More recently a study by Alcaide et al. (2009) demonstrated that a segment of the vertebrate mitochondrial Cytochrome *c* Oxidase Subunit I (COI) gene could be used to detect bloodmeal sources from a variety of hematophagous arthropods. This barcoding technique was tested on a number of biting fly species and one tick species, but was not tested on *I. scapularis.*

Although a number of genes have been used for barcoding projects including: mitochondrial gene regions of 12S rRNA 16S rRNA, and cytochrome *b* (Bradley & Baker 2001, Lambert et al. 2005, Vences et al. 2005), the cytochrome *c* oxidase subunit I (COI) gene has been proposed by

many as a universal marker and standard for species-level barcodes for animal species (Herbert et al. 2003a, Kress & Erickson 2008). The third-position nucleotides of the COI gene show high rates of base substitutions resulting in higher rates of molecular evolution than other genes used for barcoding, thus allowing for differentiation between closely allied species (Hebert et al. 2003). A study by Hebert et al. (2004) analyzed the DNA of 260 bird species from North America and found that all species had different COI sequences and differences in COI among most of the 260 species was far greater than difference within species.

In addition to the COI gene being variable enough among species and conserved enough within species for appropriate species-level identification, there are large numbers of published sequences for this gene. The International Barcode of Life project (IBOL) is a large biodiversity genomics initiative that aims to construct a DNA barcode reference library for the COI gene in animals (International Barcode of Life 2011). Furthermore, GenBank currently has over 667,000 nucleotide sequences for the COI gene. The existence of large databases and the continued development of a comprehensive barcode library for the COI gene greatly facilitate the use of this gene in barcoding projects.

The goal of this study was to develop genetic barcodes for potential host species of blacklegged ticks that occur in southern New York State (an area with high prevalence of blacklegged ticks and tick-borne diseases). A segment of the COI gene was used to develop barcodes and compare them to sequences in GenBank. Results from this study are needed for the development of DNA barcoding projects aimed at using tick bloodmeals to examine the role of different host species in tick-borne disease dynamics in southern New York State.

Materials and Methods

Vertebrate Tissue Samples

Vertebrate tissue samples were provided by a tissue bank in the Vector Ecology Laboratory, at Fordham University. Tissues were collected between the years 2000-2011 in Westchester County, New York, an area in southern New York where blacklegged ticks are highly prevalent. For this project, samples from 60 individual mammals and birds (31 species) were used. Samples included; 29 mammal and bird internal organ and skeletal muscle samples, 10 bird red blood cell samples, and nine mammal ear punches (section of ear collected from live mammal). Additionally, two human whole blood samples were kindly provided by Alex Bulanov. All samples were stored frozen or in ethanol until DNA extraction. When possible, at least two individuals of each species were used in order to compare sequences. Care was taken throughout the project to limit DNA contamination.

DNA Extraction

DNA was extracted from each tissue sample using DNeasy® Tissue Kits (Qiagen, Venlo, Netherlands). Individual organ, muscle and ear tissue samples were cut into 25 mg pieces using a sterile razor blade for each sample. Prior to the tissue lysis step, each tissue was individually ground with liquid nitrogen using a sterile, plastic pestle. Bird red blood cells (5 µl) and human whole blood (100 µl) were processed according the blood protocol for the DNeasy® Tissue Kit. Ground tissue samples and blood samples were incubated at 56^{o}C for 18 hours in lysis buffer and proteinase K. After incubation, RNase A was added to each sample to remove any RNA. The remainder of the extraction process followed the DNeasy® Tissue Kit protocol.

Primers and PCR Amplification

Alcaide et al. (2009) developed primer pools designed to capture all vertebrate sequences, regardless of nucleotide base differences. The universal-forward primer (HAA YCA YAA RGA YAT YGG) was designed for a conserved nucleotide position at the 5' end of the COI gene and the reverse primer (GCY CAN ACY ATN CCY ATR TA) is specific for vertebrate COI sequences. Additionally, Alcaide et al. (2009) added a M13-tail on the 5' end of the forward primer. I used these published primer pools but modified the hanging tags on the forward primer by adding a T7 tag and added an SP6 tag to the reverse primer for ease when sending for sample for sequencing.

The expected length of the PCR product was 758 bp. I followed the optimized reaction protocol from Alcaide et al. (2009) with slight modification. The reaction protocol consisted of a denaturation step of 4 min at 95^{o}C, 50 cycles (increased from 35 cycles) of 40 s at 45^{o}C, 1 min at 72^{o}C, and 40 s at 94^{o}C, and a final extension step of 7 min at 72^{o}C. Each PCR reaction included 10 µl GoTaq® Hot Start Green Master Mix (Promega, Fitchburg, Wisconsin, U.S.A.), 3.2 pmol of each primer, DNA and nuclease-free water to a final volume of 20 µl. PCR reactions were performed using a 2720 Thermal Cyler or Gene Amp PCR System 9700 (Applied Biosystems, Foster City, CA, U.S.A.). All PCR products were run on a 1% agarose gel and amplified products were purified using QIAquick® PCR Purification Kits (Qiagen, Venlo, Netherlands) and sent for sequencing to GENEWIZ, Inc. (South Plainfield, NJ, U.S.A).

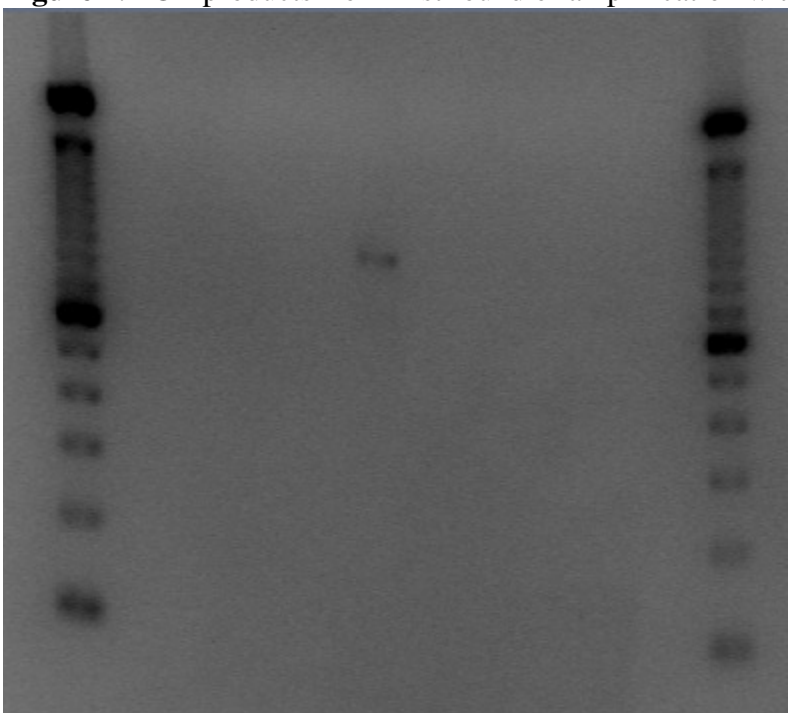
Sequencing

Each sample was sequenced in both directions using primers for the T7 and SP6 tags. Forward and reverse sequences were aligned and compared to published sequences in the NCBI BLAST nucleotide collection.

Results

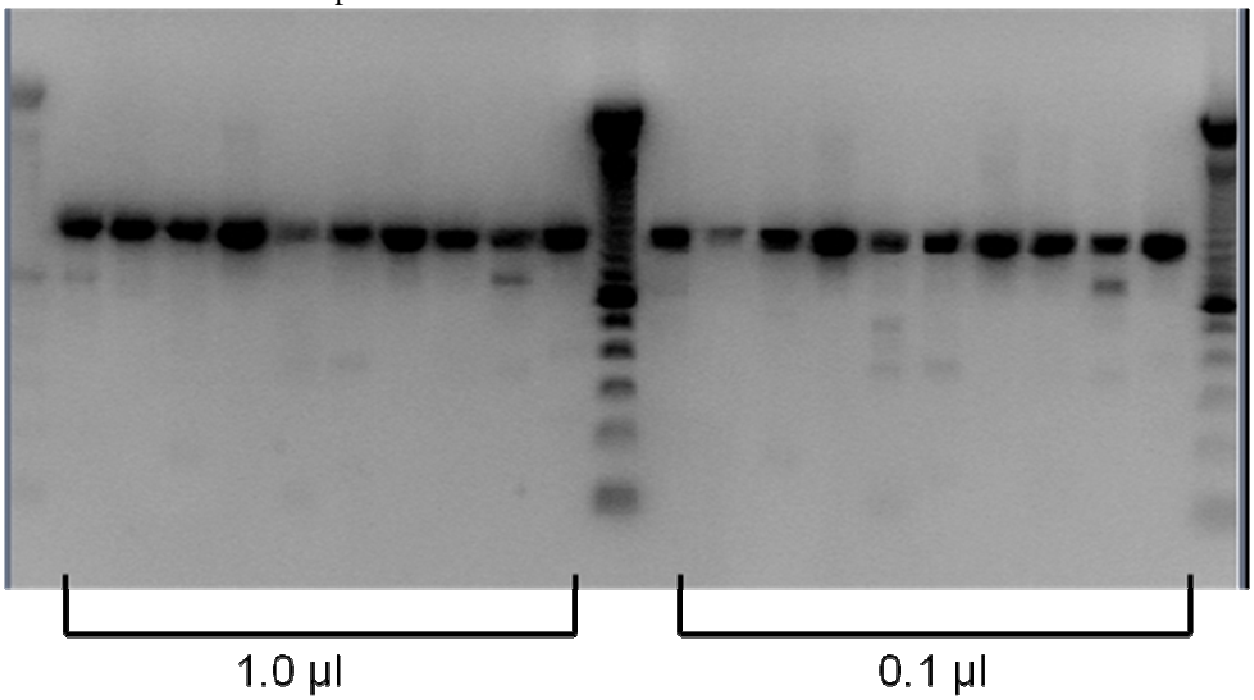
Initial amplifications on 10 samples were performed on varying concentrations of DNA (20, 10, 5, 2, 0.5 ng). A single band was observed in a sample containing 5 ng of DNA (Figure 1). An additional 10 samples were tested using the same varying concentrations of DNA and again a single band was observed. Samples that produced PCR products were sent for sequencing.

Figure 1. PCR products from first round of amplification with 5 ng of DNA.



A second round of amplification was performed using alloquots of 1 µl and 0.1 µl of the PCR product (5 ng samples) from the first round of amplification. Bands were observed for every sample for both dilutions (Figure 2). DNA extracted from the remainder of tissue samples were subjected to two rounds of amplification. All samples for the first round of amplification contained 5 ng of DNA and 1 µl of this PCR product was then used for the second round of amplification. All samples that produced PCR products on the second round of amplification were sent for sequencing.

Figure 2. PCR products from second round of amplification with 1.0 µl and 0.1 µl of PCR product from the first round of amplification.



Of the 60 individuals processed, DNA from 37 individuals (22 species) yielded PCR products during the second round of amplification and were sequenced and analyzed in NCBI BLAST (Table 1). The DNA isolated from 24 individuals (17 species), matched the expected species in GenBank. The DNA of 11 individuals (five species) matched to the appropriate order but there were no matching sequences for lower taxonomic groups (family, genus, species) for these species. These species all were members of the order Rodentia (rodents) or Lagomorpha (rabbits and hares). Two samples matched incorrect species, an ovenbird (*Seiurus aurocapillus*) matched an arrowhead warbler (*Dendroica pharetra*), and an American Robin (*Turdus migratorius*) matched a Gray Catbird (*Dumetella carolinensis*).

Table 1. Number of each species with sequenced DNA and sequence matches in GenBank.

Species	**Common Name**	**# of each Species Sequenced**	**Matched to Species**	**Matched to Order**	**Incorrect Match**
Cardinalis cardinalis	Northern Cardinal	**3**	3		
Carduelis tristis	American goldfinch	**1**	1		
Carpodacus mexicanus	House Finch	**1**	1		
Catharus guttatus	Hermit Thrush	**3**	3		
Didelphis virginiana	Common Opossum	**2**	2		
Dumetella carolinensis	Gray Catbird	**2**	2		
Homo sapiens	Human	**1**	1		
Hylocichla mustelina	Wood Thrush	**1**	1		
Marmota monax	Woodchuck	**1**		1	
Melospiza melodia	Song Sparrow	**1**	1		
Odocoileus virginianus	White-tailed deer	**1**	1		
Peromyscus leucopus	White-footed Mouse	**3**		3	
Poecile atricapillus	Black Capped Chickadee	**1**	1		
Procyon lotor	Raccoon	**2**	2		
Rattus norvegicus	Norway Rat	**1**	1		
Sciurus carolinensis	Gray squirrel	**2**		2	
Seiurus aurocapillus	Ovenbird	**2**	1		1
Spizella passerina	Chipping Sparrow	**1**	1		
Sylvilagus floridanus	Eastern Cottontail	**2**		2	
Tamias striatus	Eastern chipmunk	**3**		3	
Turdus migratorius	American Robin	**2**	1		1
Zonotrichia albicollis	White-throated Sparrow	**1**	1		

Of the 24 isolated DNA samples that matched expected species, six (five species) had unique single nucleotide polymorphisms (SNPs) when compared to species sequences in GenBank (Table 2). Of the two individual Gray Catbirds with SNPs, both share the same SNP and one bird had two additional SNPs.

Table 2. Sequenced species with SNPs

			Number of SNPs			
Species	**Common Name**	**Total # Sequenced**	**0**	**1**	**2**	**3**
Seiurus aurocapillus	Ovenbird	2	1		1	
Carduelis tristis	American Goldfinch	1		1		
Dumetella carolinensis	Gray Catbird	2		1		1
Cardinalis cardinalis	Northern Cardinal	3	2		1	
Procyon lotor	Raccoon	2	1			1

Most of the nucleotide differences were purine to purine, or pyrimidine to pyrimindine differences, with the exception of one SNP for the raccoon DNA that was thymine to guanine (Table 3).

Table 3. Location of each SNP for each sequenced DNA compared to published GenBank sequences (SNPs are in red).

SNP location for Ovenbird (*Seiurus aurocapillus*) compared to GenBank: GU932043.1

```
Query  241   CAACATAAGCTTCTGATTACTACCACCATCATTTCTCCTACTCCTAGCATCCTCCACAGT  300
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Sbjct  1696  CAACATAAGCTTCTGACTACTACCACCATCATTTCTCCTACTCCTAGCATCCTCCACAGT  1755
```

SNP location for American Goldfinch (*Carduelis tristis*) compared to GenBank: DQ434517.51

```
Query  541   GCAACCTCAACACTACATTCTTCGATCCTGCAGGAGGAGGTGACCCAGTCCTATACCAAC  600
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  589   GCAACCTCAACACCACATTCTTCGATCCTGCAGGAGGAGGTGACCCAGTCCTATACCAAC  648
```

SNP location shared by both Gray Catbirds (*Dumetella carolinensis*) compared to GenBank: EF484216.1

```
Query  61    TAGGCCAGCCCGGAGCCCTACTAGGCGACGACCAGGTCTACAATGTGATCGTTACAGCCC  120
             ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  125   TAGGCCAACCCGGAGCCCTACTAGGCGACGACCAGGTCTACAATGTGATCGTTACAGCCC  184
```

SNPs locations for the Gray Catbird (*Dumetella carolinensis*) with three SNPs compared to GenBank: EF484216.1

```
Query  121   ACGCCTTCGTAATAATCTTCTTTATAGTTATACCAATCATAATCGGGGGATTCGGAAACT  180
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Sbjct  185   ACGCCTTCGTAATAATCTTCTTTATAGTTATACCAATCATAATCGGAGGATTCGGAAACT  244


Query  541   TACTAGCTGCTGGAATCACCATGCTTCTCACAGACCGCAATCTCAACACCACCTTCTTTG  600
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct  605   TACTAGCTGCTGGAATCACCATGCTCCTCACAGACCGCAATCTCAACACCACCTTCTTTG  664
```

SNP location for Northern Cardinal (*Cardinalis cardinalis*) compared to GenBank: DQ434508.1

```
Query  241  TTTTGACTCCTACCTCCATCCTTCCTTCTCCTCCTAGCATCTTCTACAGTCGAAGCGGGT  300
            |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Sbjct  256  TTTTGACTCCTACCTCCATCTTTCCTTCTCCTCCTAGCATCTTCTACAGTCGAAGCGGGT  315
```

SNPs locations for Raccoon (*Procyon lotor*) compared to GenBank: AM711899.1

```
Query  481   TTTTTGTATGGTCAGTACTTATTACAGCAGTACTTCTCTTATTATCTCTGCCAGTACTAG  540
             |||||||||||||||||||||||||||||||||||||||||||||| || ||||||||||
Sbjct  5907  TTTTTGTATGGTCAGTACTTATTACAGCAGTACTTCTCTTATTATCGCTACCAGTACTAG  5966

Query  541   CAGCAGGCATCACCATATTACTCACCGACCGAAATCTGAACACAACCTTCTTTGACCCAG  600
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct  5967  CAGCAGGCATCACCATACTACTCACCGACCGAAATCTGAACACAACCTTCTTTGACCCAG  6026
```

For all nucleotide differences between the sequenced vertebrate samples in this study and the species they aligned with in GenBank, none of the SNPs resulted in changes in the protein sequence.

Discussion

This study resulted in the successful extraction and sequencing of vertebrate DNA from various tissues sources. Genetic barcodes were generated for 22 species that are found in southern New York and are potential hosts of blacklegged ticks (primarily ground dwelling mammals and birds). DNA was initially extracted from 60 tissue samples, however only 37 samples yielded sufficient DNA for sequencing. Unsuccessful extraction could have been due to DNA degradation of the tissue samples or insufficient lysis for some samples that may not have been ground completely.

Most of the 37 individuals that were successfully sequenced matched the expected species in GenBank, demonstrating that these 17 species have published sequences for the COI gene. Four species within the order Rodentia and one species within the order Lagomorpha did not match species within GenBank, indicating a deficiency in the nucleotide collection for this gene and these species. This study resulted in the generation of genetic barcodes for these five common species

found in southern New York and will eventually result in addition of these sequences to the GenBank database. Addition of these sequences is of particular importance for tick-borne disease research that aims to use genetic barcodes of the COI gene for host studies, since these five species are known to harbor large numbers of ticks and are also important in the disease amplification in wildlife populations (LoGiudice et al. 2003).

Additionally, two individuals (two species) matched incorrect species with GenBank sequences. An ovenbird matched an arrowhead warbler, and an American Robin matched a Gray Catbird. In total, two ovenbirds and two American robins were sequenced, one of each matched incorrectly and one of each matched to species. This indicates that there may have been an issue with proper labeling or morphological identification of the incorrectly matched specimens. However, arrowhead warblers are endemic to Jamaica (Raffaele et al. 2003) making it highly unlikely that this species with captured in southern New York and misidentified. Furthermore, catbirds and robins have distinct morphological differences, making misidentification unlikely. There may be an error in GenBank for these sequences.

The unique SNPs found in five species, as compared to GenBank sequences, demonstrates sequence variation in these species for this gene do exist. Of the four bird species with SNPs, three are known to be migratory (American Goldfinch, Gray Catbird, Ovenbird). The genetic variation observed for these migratory species may represent unique populations that converge in southern New York, either as stop over points, overwintering or breeding grounds. However, two of the species with SNPs (Northern Cardinal and Raccoon) are known to reside in southern New York where they were collected. The genetic differences observed in these five species highlight the importance of sampling multiple individuals within each species, a goal that was attempted in this study. Further study of these differences would be supported by sequencing DNA from more individuals of each species.

The information generated from this study will be useful in future studies that aim to use barcodes of potential host species of blacklegged ticks, such as studies that wish to determine bloodmeal sources of these ticks within southern New York State.

Acknowledgements

I would like to thank Dr. Berish Rubin for his guidance with this project's conception and development and for training in molecular techniques. Thank you to Alex Bulanov and Xie Xie for their assistance with all laboratory and computer work as well as their advice throughout the project. This work was supported by funds for the Techniques in Cell and Molecular Biology course at Fordham University.

Literature Cited

Alcaide, M., C. Rico, S. Ruiz, R. Soriguer, J. Munoz, and J. Figuerola. 2009. Disentangling vector-borne transmission networks: A universal DNA barcoding method to identify vertebrate hosts from arthropod bloodmeals. PLos ONE:1-6.

Bradley, R. D., and R. J. Baker. 2001. A test of the genetic species concept: cytochrome *b* sequences and mammals. J Mammal 92: 960-973.

Dumler, J. S., K Choi, J. C. Garcia-Garcia, N. S. Barat, D. G. Scorpio, J. W. Garyu, D. J. Grab, and J. S. Bakken. Human Granulocytic Anaplasmosis and Anaplasma phagocytophilum. Emerg Infect Dis 11: 1828-1834.

Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. deWaard. 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B. 270: 313-321.

Hebert, P. D. N., M. Y. Stoeckle, T. S. Zemlak, and C. M. Francis. 2004. Identification of birds through DNA barcodes. PLOS Biol 2: 1657-1663.

Humair, P. F., V. Douet, F. M. Cadenas, L. M. Schouls, I. Van de Pol, and L. Gern. 2007. Molecular identification of bloodmeal source in Ixodes ricinus ticks using 12S rDNA as a genetic marker. Journal of Medical Entomology 44:869-880.

International Barcode of Life. 2011. Retreived May 12, 2011, from http://www.barcodeoflife.org/
Kress, W. J., and D. L. Erickson. 2008. DNA barcodes: Genes, genomics, and bioinformatics. PNAS, 107: 2761-2762.

Kjemtrup, A. M., and P. A. Conrad. 2000. Human babesiosis: an emerging tick-borne disease. International Journal for Parasitology 30: 1323-1337.

Lambert, D. M., A. Baker, L. Huynen, O. Haddrath, P. D. N. Hebert, and C. D. Millar. 2005. Is a large-scale DNA-based inventory of ancient life possible? J Hered 96: 279-284.

Lane, R.S., J. Piesman, W. Burgdorfer. 1991. Lyme borreliosis: Relation of its causative agent to its vectors and hosts in North America and Europe. Annual Review of Entomology 36:587-609.

LoGiudice, K., R. S. Ostfeld, K. A. Schmidt, and F. Keesing. 2003. The ecology of infectious disease: Effects of host diversity and community composition on Lyme disease risk. PNAS 100: 567-571.

Pichon, B., D. Egan, et al. (2003). "Detection and identification of pathogens and host DNA in unfed host-seeking Ixodes ricinus L. (Acari : Ixodidae)." Journal of Medical Entomology 40(5): 723-731.

Raffaele, Herbert; James Wiley, Orlando Garrido, Allan Keith & Janis Raffaele (2003). Birds of the West Indies, Christopher Helm, London.

Schwartz, I., D. Fish, and T. J. Daniels. 1997. Prevalence of the rickettsial agent of Human Granulocytic Ehrlichiosis in ticks from a hyperendemic focus of Lyme disease. New England Journal of Medicine 337:49-50.

Spielman, A., M.L. Wilson, J.F. Levine, et al 1985. Ecology of *Ixodes dammini*-borne human babesiosis and Lyme disease. Annual Review of Entomology 30:439-460.

Vences, M., M. Thomas, A. van der Meijden, Y. Chiari, and D. R. Vieites. 2005. Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. Frontiers in Zoology 2: 1-12.