MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2016) **16**, 1224–1239 doi: 10.1111/1755-0998.12524

SPECIAL ISSUE: SEQUENCE CAPTURE

Simultaneous identification of host, ectoparasite and pathogen DNA via in-solution capture

MICHAEL G. CAMPANA,*[1] MELISSA T. R. HAWKINS,*†[1] LAUREN H. HENSON,* KRISTIN STEWARDSON,* HILLARY S. YOUNG,‡ LEAH R. CARD,§ JUSTIN LOCK,* BERNARD AGWANDA,¶ JORY BRINKERHOFF,** HOLLY D. GAFF,†† KRISTOFER M. HELGEN,† JESÚS E. MALDONADO,*† WILLIAM J. MCSHEA§ and ROBERT C. FLEISCHER*

*Center for Conservation and Evolutionary Genetics, Smithsonian Conservation Biology Institute, 3001 Connecticut Avenue NW, Washington, DC 20008, USA, †Division of Mammals, National Museum of Natural History, Smithsonian Institution, MRC 108, P.O. Box 37012, Washington, DC 20013-7012, USA, ‡Department of Ecology, Evolution and Marine Biology, University of California Santa Barbara, Santa Barbara, CA 93106, USA, §Smithsonian Conservation Biology Institute, National Zoological Park, 1500 Remount Rd., Front Royal, VA 22630, USA, ¶National Museums of Kenya, Box 40658-00100, Nairobi, Kenya, **Department of Biology, B322 Gottwald Center for the Sciences, University of Richmond, 28 Westhampton Way, Richmond, VA 23173, USA, ††Department of Biological Sciences, Old Dominion University, Norfolk, VA 23529, USA

Abstract

Ectoparasites frequently vector pathogens from often unknown pathogen reservoirs to both human and animal populations. Simultaneous identification of the ectoparasite species, the wildlife host that provided their most recent blood meal(s), and their pathogen load would greatly facilitate the understanding of the complex transmission dynamics of vector-borne diseases. Currently, these identifications are principally performed using multiple polymerase chain reaction (PCR) assays. We developed an assay (EctoBaits) based on in-solution capture paired with high-throughput sequencing to simultaneously identify ectoparasites, host blood meals and pathogens. We validated our in-solution capture results using double-blind PCR assays, morphology and collection data. The EctoBaits assay effectively and efficiently identifies ectoparasites, blood meals, and pathogens in a single capture experiment, allowing for high-resolution taxonomic identification while preserving the DNA sample for future analyses.

Keywords: blood meal, DNA, ectoparasite, high-throughput sequencing, in-solution capture, pathogen

Received 7 June 2015; revision received 8 January 2016; accepted 19 January 2016

Introduction

Nearly 30% of emerging infectious diseases of humans are vector-borne, with most of these having a wildlife reservoir host (Jones *et al.* 2008). Ectoparasites are common vectors of pathogen transmission from one host to another, involved in diseases such as Lyme disease (*Borrelia burgdorferi*), Rocky Mountain spotted fever (*Rickettsia rickettsii*), malaria (*Plasmodium* spp.) and anaplasmosis (*Anaplasma phagocytophilum*), among many others (e.g. Pfäffle *et al.* 2013; Brites-Neto *et al.* 2015). Dynamics of disease transmissions are changing at an unprecedented rate due to anthropogenic effects on ecosystems including modification of landscapes, alteration of species distributions and impacts of human-mediated climate change (e.g. Jones *et al.* 2008; Hoberg & Brooks 2015). To understand disease dynamics and the prevalence of specific diseases, intensive field studies are required to collect and screen host and ectoparasite samples for a variety of pathogens (Gilbert *et al.* 2000; Lynen *et al.* 2007; Midilli *et al.* 2009; Sabatini *et al.* 2010). By monitoring the prevalence and distribution of pathogens across hosts, ectoparasites and habitats, we can better predict disease spread, improve our understanding of cycles of disease transmission, and better document movement of pathogens through landscapes, with important applications for predicting emergence of zoonoses globally.

Researchers exploring the dynamics of vector-borne disease in the field often face multiple challenges in correctly identifying hosts, ectoparasites and pathogens (Allan *et al.* 2010). Precise morphological identification of hosts and ectoparasites often requires significant taxonomic expertise (e.g. Nadolny *et al.* 2011; Young

Correspondence: Robert C. Fleischer, Fax: +1 202-633-1237; E-mail: fleischerr@si.edu

[1]Co-first Authors.

et al. 2015a), while molecular techniques are required for accurate and reliable screening of pathogens (Nadolny *et al.* 2011, 2014; Pichon *et al.* 2003). Damage to field-collected specimens can complicate taxonomic identification. Furthermore, ectoparasite samples often yield small volumes and low concentrations of extracted DNA (laboratory observations, Allender *et al.* 2004). Multiple polymerase chain reaction (PCR) assays are thus frequently needed to ensure accurate host and pathogen identification, which can rapidly consume these small quantities of DNA. While multiplex PCR assays can reduce the total number of reactions (e.g. Courtney *et al.* 2004), these strategies require extensive optimization and the use of expensive reagents (e.g. fluorescent probes) or additional experimental steps (e.g. bacterial cloning). Therefore, efficient molecular screening methods are needed to identify DNAs corresponding to a comprehensive number of potential pathogens as well as a variety of hosts and parasites.

DNA capture assays coupled with next generation sequencing are highly suitable for the study of pathogens from a variety of low copy number sources including ectoparasite, noninvasive, museum and ancient samples (Bos *et al.* 2015, and references therein; Hawkins *et al.* 2015; Vuong *et al.* 2013). We developed a method to simultaneously enrich samples from various sources for ectoparasite (primarily ixodid ticks), blood meal, and pathogen DNA. Here we describe an in-solution capture assay (EctoBaits) that, paired with high-throughput sequencing, can simultaneously identify a range of ectoparasites, hosts and pathogen strains from a single library preparation and in-solution hybridization experiment. The EctoBaits assay capitalizes on the wealth of data provided in current DNA barcoding databases to provide accurate species-level taxonomic assignment. While currently limited to northeastern American and East African taxa, the EctoBaits assay can be easily adapted and expanded for other regions of interest.

Materials and methods

Probe design

With the goal of designing an in-solution capture assay for detecting ectoparasites, blood meals and pathogens simultaneously, we comprehensively searched public databases (BOLD and Genbank) for sequences from a variety of taxa relevant to ongoing studies in the northeastern United States and East Africa (Table S1, Appendices S1 and S2), where we are engaged in ongoing fieldwork (Young *et al.* 2014, 2015a; Card *et al.* in press). Both regions are hotspots of vector-borne disease risk (Jones *et al.* 2008). For ectoparasites, hosts, macroparasites and eukaryotic pathogens, we targeted mitochondrial genes due to their higher copy number per cell and corresponding increased likelihood of successful enrichment compared to nuclear genes. For bacterial pathogens, we targeted diagnostic regions (such as 16S rRNA, citrate synthase, flagellin and IS1111 spacers) that had been identified as species-specific in previous research (Eremeeva *et al.* 1994; Levin & Fish 2000; Mediannikov *et al.* 2010; Bai *et al.* 2011; Ghafar & Eltablawy 2011). The final bait targets included 111 American bird species, 47 American mammal species, 108 African mammal species, 84 pathogen strains (59 bacterial strains, 25 eukaryotic strains), 53 hard tick (Ixodidae) species, one soft tick (Argasidae) species, 10 flea species, one sucking louse (Anoplura) species and one filiarial parasite. Pathogen baits were derived from eight *Anaplasma* strains, five *Babesia* strains, seven *Bartonella* strains, three *Borrelia* strains, two *Coxiella* strains, 21 *Ehrlichia* strains, 11 *Hepatozoon* strains, 17 *Rickettsia* strains, nine *Theileria* strains and one strain of *Yersinia pestis* (Table S1, Appendices S1 and S2).

To remove overly similar redundant sequences (<10% divergence) from the probe set, the identified sequences were clustered using CD-HIT-EST (Li & Godzik 2006). Clustering the sequences produced a degenerate set designed to capture a wide range of species. The clustered sequences were split into 100 bp pieces for probe synthesis, resulting in 3,901 different probe sequences after quality control (Appendix S2). Probes did not overlap and were not tiled. Each probe sequence was copied approximately five times to fill a 20 000-probe MYbaits kit (MYcroarray, USA). The effectiveness of the probe set was confirmed by spiking certified pathogen-free chicken (*Gallus gallus*) and rabbit (*Oryctolagus cuniculus*) blood with known concentrations of pathogen-specific PCR products and capturing these samples using EctoBaits (Appendix S1).

Field collection and sample selection

Specimens were obtained from large ongoing investigations of ectoparasites and vector-borne diseases in dry tropical savanna ecosystem of Laikipia County, Kenya (0°17′N, 36°52′ E) and forest and suburban ecosystems in Virginia, USA (38°53′15.6″ N, 78°9′54.6″ W). Ectoparasites were collected between 2010 and 2012 by standardized tick drags (Goddard 1992) or by direct removal from live and dead mammalian and avian hosts (methods in Card *et al.* in press; Young *et al.* 2015a,b). All Kenyan ectoparasites were documented photographically for later reference. Virginian ticks were not photographed as these species have been well-characterized both morphologically and genetically in previous studies. Samples were stored in 95% ethanol until analysis. Twenty-seven ectoparasites

(10 Virginian ticks, 15 Kenyan ticks and two Kenyan fleas) were selected for a 'double blind' system to evaluate the EctoBaits assay (Table 1). Twenty ectoparasites were removed directly from hosts and seven were questing ticks. Most ectoparasite species were identified using morphology and PCR assays (see below). Morphological species identifications were performed with the assistance of taxonomic experts (ticks: Richard Robbins [Armed Forces Pest Management Board], fleas: Katharina Dittmar [State University of New York, Buffalo]). Twelve ectoparasites (KenT01–KenT10, KenF01–KenF02) were directly assessed by taxonomic experts before destruction for DNA analysis, while the remaining 17 (VirT01–VirT10, KenT11–KenT15) were assessed by comparison to identified material. Given difficulties of morphological identification for larval ticks, two of these ectoparasites (KenT07, KenT10) were identified via the strong ecological association between the vectors and their hosts (Table 1). The selected tick samples represented a variety of life stages (nymphal, larval, adult) and feeding conditions (engorged, unengorged but associated with a host, questing on the landscape). The fleas were both adult animals removed from live rodent hosts.

Laboratory methods

DNA extraction. Prior to extraction, all ethanol was removed from the ectoparasite samples. Ectoparasites were then homogenized dry using a BeadBeater (BioSpec Products, USA). DNA was extracted from the homogenates using Qiagen DNA Tissue kits (Qiagen, USA) following the manufacturer's instructions. An extraction negative was included with every extraction set-up.

PCR assays. Ectoparasite species and pathogen infections (*Anaplasma*, *Babesia*, *Bartonella*, *Borrelia*, *Coxiella*, *Ehrlichia*, *Hepatozoon*, *Theileria*) were identified using conventional PCR assays. Positive controls, extraction negatives and no-template controls (containing water instead of DNA) were analysed in each set-up. Each 25 μL reaction contained 1 × AmpliTaq Gold reaction buffer (Life Technologies, USA), 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.4 μM each primer, 20 μg BSA, 1U AmpliTaq Gold (Life Technolo-

Table 1 Ectoparasites selected to evaluate the EctoBaits capture assay

Ectoparasite	Source	Morphology/Ecology	PCR species	Life stage	Source	Accession
VirT01	Virginia	*Ixodes brunneus*	*Ixodes scapularis/dammini*	Adult/Nymph	*Thryothorus ludovicianus*	SRS911751
VirT02	Virginia	*Dermacentor variabilis*		Adult	*Didelphis virginiana*	SRS911765
VirT03	Virginia	*Amblyomma americanum*	*A. americanum*	Nymph	*Vulpes vulpes*	SRS911766
VirT04	Virginia	*I. scapularis*	*Ixodes minor/affinis*	Adult	*Sylvilagus floridanus*	SRS911768
VirT05	Virginia	*A. americanum*	*A. americanum*	Nymph	*Odocoileus virginianus*	SRS911769
VirT06	Virginia	*D. variabilis*	*D. variabilis*	Adult	*V. vulpes*	SRS911770
VirT07	Virginia		*Ixodes cookei/banksi*	Larva	*Marmota monax*	SRS911850
VirT08	Virginia	*D. variabilis*	*D. variabilis*	Adult	*V. vulpes*	SRS911851
VirT09	Virginia	*A. americanum*	*A. americanum*	Adult	*O. virginianus*	SRS911852
VirT10	Virginia	*Ixodes brunneus*	*I. scapularis/dammini*	Nymph	*Buteo jamaicensis*	SRS911933
KenT01	Kenya	*Rhipicephalus* sp.	*Rhipicephalus* sp.	Adult	*Lycaon pictus*	SRS911935
KenT02	Kenya	*Rhipicephalus* sp.	*Rhipicephalus* sp.	Adult	*L. pictus*	SRS911936
KenT03	Kenya	*Rhipicephalus pulchellus*	*R. pulchellus*	Adult	Questing Tick	SRS911939
KenT04	Kenya	*Ixodes* sp.	*Ixodes* sp.	Nymph	*Elephantulus rufescens*	SRS911941
KenT05	Kenya	*Ixodes* sp.	*Ixodes* sp.	Nypmh	*E. rufescens*	SRS911942
KenT06	Kenya	*R. pulchellus*	*R. pulchellus*	Adult	Questing Tick	SRS911957
KenT07	Kenya	*Haemophysalis leachii**	Ixodidae	Nymph	*Mastomys natalensis*	SRS911983
KenT08	Kenya		*Rhipicephalus* sp.	Nymph	*Aethomys hindei*	SRS911984
KenT09	Kenya	*Rhipicephalus pravus*	*Rhipicephalus* sp.	Nymph	*E. rufescens*	SRS911994
KenT10	Kenya	*H. leachii**		Nymph	*Acomys percivali*	SRS911996
KenT11	Kenya	*R. pravus*		Adult	Questing Tick	SRS1133052
KenT12	Kenya	*R. pulchellus*	*Rhipicephalus* sp.	Adult	Questing Tick	SRS1133057
KenT13	Kenya	*Rhipicephalus praetextatus*	*R. pravus*	Adult	Questing Tick	SRS1133060
KenT14	Kenya	*R. praetextatus*		Adult	Questing Tick	SRS1133069
KenT15	Kenya	*R. praetextatus*	*R. pulchellus*	Adult	Questing Tick	SRS1133099
KenF01	Kenya	*Xenopsylla* sp.	Pulicidae	Adult	*Gerbilliscus robustus*	SRS911999
KenF02	Kenya	*Xenopsylla* sp.	Pulicidae	Adult	*G. robustus*	SRS912000

Parasites identified by their strong ecological association with their hosts (rather than morphology) are denoted with an asterisk (*). Sequence Read Archive accession numbers are given for each sample.

Table 2 Polymerase chain reaction assay primers. Primer names are listed with the sequences

Target	Locus	Forward (5′→3′)	Reverse (5′→3′)	References
Ectoparasite	Cytochrome coxidase subunit I	HC02198: TAA CTT CAG GGT GAC CAA AAA TCA	LC01490: GGT CAA CAA ATC ATA AAG ATA TTG G	Folmer *et al.* (1994)
Ectoparasite	Cytochrome c oxidase subunit I	LEPF1: ATT CAA CCA ATC ATA AAG ATA TTG G	LEPR1: TAA ACT TCT GGA TGT CCA AAA ATC A	Hebert *et al.* (2004)
Ectoparasite	16S rRNA	16s+1: CTG CTC AAT GAT TTT TTA AAT TGC TGT	16s-1: GTC TGA ACT CAG ATC AAG T	Nadolny *et al.* (2011)
Anaplasma	16S rRNA	E1: GGC ATG TAG GCG GTT CGG TAA GTT	E2: CCC CCA CAT TCA GCA CTC ATC GTT TA	Ghafar & Eltablawy (2011)
Bartonella	Citrate synthase	BhCS871.p: GGG GAC CAG CTC ATG GTG G	BhC1137.n: AAT GCA AAA AGA ACA GTA AAC A	Bai *et al.* (2011)
Borrelia	Flagellin	FLA297: CGG CAC ATA TTC AGA TGC AGA CAG	FLA652: CCT GTT GAA CAC CCT CTT GAA CC	Levin & Fish (2000)
Coxiella	IS1111 spacers	CbISF: CAA GAA ACG TAT CGC TGT GGC	CbISR: CAC AGA GCC ACC GTA TGA ATC	Mediannikov *et al.* (2010)
Ehrlichia	16S rRNA	HE1F: CAA TTG CTT ATA ACC TTT TGG TTA TAA AT	HE3R: TAT AGG TAC CGT CAT TAT CTT CCC TAT	Ghafar & Eltablawy (2011)
Francisella	Insertion element-like sequence	ISFTu2F: TTG GTA GAT CAG TTG GTG GGA TAA C	ISFTu2R: TGA GTT TTA CCT TCT GAC AAC AAT ATT TC	Versage *et al.* (2003)
Proteus	*ureR* gene	ureRF1: GGT GAG ATT TGT ATT AAT GG	ureRR1: ATA ATC TGG AAG ATG ACG AG	Zhang *et al.* (2013)
Rickettsia	120-kDa protein antigen gene	BG1-21: GGC AAT TAA TAT CGC TGA CGG	BG2-20: GCA TCT GCA CTA GCA CTT TC	Eremeeva *et al.* (1994)
Staphylococcus	*pta* gene	pta_f1: AAA GAC AAA CTT TCA GGT AA	pta_r1: GCA TAA ACA AGC ATT GTA CCG	Bannoehr *et al.* (2009)
Apicomplexa	18S rRNA	HepF300: GTT TCT GAC CTA TCA GCT TTC GAC G	HepR900: CAA ATC AAG AAT TTC ACC TCT GAC	Ujvari *et al.* (2004)
Apicomplexa	18S rRNA	BTH-1F: CCT GMG ARA CGG CTA CCA CAT CT	BTH-1R: TTG CGA CCA TAC TCC CCC CA	Criado-Fornelio *et al.* (2003)
Filariae	18S rRNA, 5.8S rRNA, internal transcribed spacer	FL1: TTC CGT AGG TGA ACC TGC	ITSR: ACC CTC AAC CAG ACG TAC	Namrata *et al.* (2014)
Onchocercidae	12S rRNA	Oncho12SF: TGA CTG ACT TTA GAT TTT TCT TTG G	Oncho12SR: AAT TAC TTT CTT TTC CAA TTT CAC A	Namrata *et al.* (2014)

gies) and 2 μL DNA extract. Primer sequences are listed in Table 2. For the ectoparasite cytochrome c oxidase subunit I (*cox1*) reactions, thermocycling consisted of an initial five minute denaturation step at 95 °C, five cycles of 30 s at 95 °C, 40 s at 45 °C and one minute at 72 °C, 35 cycles of 30 s at 95 °C, 40 s at 51 °C and 1 min at 72 °C, and a final 10 min extension step at 72 °C. *Bartonella*, *Borrelia*, *Coxiella*, *Rickettsia* and ectoparasite 16S rRNA assay programs included an initial 5 min denaturation of 95 °C, 35 (*Bartonella*, *Coxiella*, ectoparasite) or 40 (*Borrelia, Rickettsia*) cycles of one minute at 94–95 °C, one minute at annealing temperature (56 °C for *Bartonella*, 60 °C for *Borrelia* and *Coxiella*, 55 °C for *Rickettsia*, 50 °C for ectoparasites) and one minute at 72 °C, and a final 5 min extension step of 72 °C. Thermocycling for the *Anaplasma* and *Ehrlichia* assays included an initial 5 min denaturation of 95 °C, 35 cycles of 30 s at 94 °C, 30 s at annealing temperature (58 °C for *Anaplasma*, 55 °C for *Ehrlichia*), and 30 s at 72 °C, and a final 5 min extension step of 72 °C. The apicomplexan assays utilized an initial 5 min denaturation of 95 °C, 35 (HepF300/HepR900 primer pair) or 40 (BTH-1F/BTH-1R primer pair) of 30 s at 94 °C, 30 s at 60 °C, and 45 (HepF300/Hep4900) or 60 (BTH-1F/BTH-1R) seconds at 72 °C, and a final 5 min extension step of 72 °C. Representative subsamples of positive PCR products were sequenced on an ABI 3130 (Life Technologies) following standard protocols. False positives (as assessed by sequencing PCR blanks and extraction controls) were discarded.

Library preparation. Single-indexed (22 ectoparasites: VirT01–VirT10, KenT01–KenT10, KenF01–KenF02) and double-indexed (5 ectoparasites: KenT11–KenT15) libraries were prepared using the KAPA Library Preparation Kit – Illumina (KK8201: Kapa Biosystems, USA) following the manufacturer's protocol (Table 1). We included both library and extraction controls for the single-indexed library experiments. We purified products between steps using carboxyl paramagnetic beads (Rohland & Reich 2012). Since 21 of 27 ectoparasite samples had total input DNA of <500 ng, 14–16 cycles of indexing PCR were performed to amplify the adapter–ligated DNA fragments.

Capture and sequencing. Single-indexed libraries were pooled in groups of eight (62.5 ng each library per capture; 500 ng total library per capture pool) for in-solution hybridization following the MYbaits manufacturer's (MYcroarray, USA) protocol. To test the effects of probe concentration, each multiplexed single-indexed library pool was captured twice: once with a 1:5 MYbaits probe dilution (100 ng of probes per capture) and once with a 1:10 dilution (50 ng of probes). Since the results between the captures using the two probe dilutions were similar (see "Probe Dilution" below), we refer to the two captures as 'replicates' for simplicity, despite differing in probe concentration. After capture, single-indexed library pools were amplified by 20 cycles of PCR using Illumina adapter sequences and quantified via quantitative PCR. After library amplification, we size-selected library molecules (200–400 bp) via agarose gel purification using QIAquick Gel Extraction Kits (Qiagen) following the manufacturer's protocol. Library quality was ensured by visual inspection after agarose gel electrophoresis and analysis using a 2100 Bioanalyzer (Agilent Technologies, USA) high sensitivity DNA chip. Pools were sequenced on a HiSeq 2500 (Illumina, USA) with paired-end 100 bp reads by Macrogen (South Korea).

Double-indexed libraries were pooled equally in a group of five (20 ng each library per capture; 100 ng total library per capture pool) and captured using the EctoBaits assay following the manfacturer's instructions. Double-indexed libraries were captured using only a 1:10 dilution of the probes since initial experiments found no significant difference between the 1:5 and 1:10 MYbaits probe dilutions (see "Probe Dilution" below). Captured double-indexed libraries were then reamplified by 18 cycles of PCR using Illumina adapter sequences. Adapter artifacts were removed from reamplified captured libraries (keeping all library molecules longer than 200 bp) using QIAEX® II Gel Extraction Kits (Qiagen) following the manufacturer's protocols. Pools were sequenced on a MiSeq (Illumina, USA) with paired-end 150 bp reads.

Sequence quality control. Using Trimmomatic 0.32 or 0.33 (Bolger *et al.* 2014), we removed adapter sequences and artifacts (maximum seed mismatches 2, palindrome clip threshold 30, simple clip threshold 10) and low quality leading and trailing bases (below PHRED-like quality 3). Additionally, reads were scanned with 4 bp sliding windows and trimmed when average PHRED-like score fell below 15 using Trimmomatic. Trimmed reads below 36 bp were discarded. Paired-end reads were then merged (minimum overlap of 10 bp) using FLASH 1.2.11 (Magoč & Salzberg 2011). The merged, unpaired and unmerged reads were concatenated. Unmerged reads were treated as independent in downstream analyses. The sequences were filtered for quality (minimum base quality score 20) using the QualityFilterFastQ.py script (Kircher 2012). PCR duplicates were removed using CD-HIT-DUP 0.5 (Li & Godzik 2006). Final library qualities were checked using FastQC 0.11.2 (Andrews no date).

Sequence identification and assay validation. The degenerate baits used in EctoBaits precluded identification of species by alignment against the bait sequences. Therefore, sequences were aligned against the Genbank nonredun-

dant nucleotide database using Megablast (BLAST 2.2.30+: Zhang *et al.* 2000) under default settings. Megablast identifications were analysed using MEGAN 5.7 or 5.10 under default parameters (Huson *et al.* 2011). Empirical estimates suggest that ~2 in 1000 reads are incorrectly assigned to libraries in single-indexed, multiplex-captured Illumina experiments due to jumping PCR (Kircher *et al.* 2012). Therefore, to control for jumping PCR and laboratory contaminants, MEGAN sequence identifications that comprised <1/1000 of the total quality-control library were filtered out. Resulting pathogen and blood meal DNAs are expected to be sequenced in a lower frequency than those of the ectoparasite due to the composition of the original DNA extract (eg. Che Lah *et al.* 2015; Bos *et al.* 2015) so the 1/1000 cut-off is a conservative limit. Additionally, since some libraries had very low sequencing depth (fewer than 10 000 sequences after quality control), taxa were not identified unless there were at least 10 corresponding sequences in the filtered library. Although this step could exclude some true blood meal and pathogen sequences, it was necessary to avoid false positives due to laboratory background and jumping PCR. After filtering, the remaining taxa were cross-referenced against the DNA library and extraction controls to identify suspected contaminants.

To validate the assay, each captured library was evaluated against the PCR and ectoparasite morphological results. Putative ectoparasites, blood meal(s) and pathogens were identified using MEGAN, as described below. The metabarcode results were then compared to the expected species based on morphology, data collection and PCR assay data.

Ectoparasites

Preliminary ectoparasite identifications were determined directly from the MEGAN results. The majority ectoparasite clade (Ixodida vs. Insecta in our test samples) was identified for each library (Table S2, Appendix S1). Within the majority clade, we identified the genus to which the plurality of genus-specific sequences within the ectoparasite clade belonged. We accepted the plurality genus as the ectoparasite genus if at least 65% of the genus-specific sequences corresponded to the plurality genus (the 'genus confidence'). Similarly, we identified the species to which the plurality of species-specific sequences within the identified genus belonged. We calculated a 'species confidence' by multiplying the genus confidence by the percentage of the sequences within the identified genus corresponding to the plurality species. The species confidence was defined as zero if no sequences were identifiable to species-level. We accepted the plurality species as the ectoparasite species if the species confidence exceeded 65%. We found empirically that a 65% supermajority reliably identified ectoparasite species, while simultaneously excluding spurious Megablast assignments and sample contaminants (Table S6, Appendix S1).

To refine the preliminary ectoparasite species assignment, all ectoparasite sequences were extracted from the MEGAN results. These were mapped against the *cox1* barcoding region (Genbank accession KC488279.1) and a consensus *cox1* sequence was determined using Geneious 8.1 (medium sensitivity alignment, 10 alignment iterations) (Biomatters, Ltd. 2015). The consensus barcode was then compared against BOLD and Genbank to determine ectoparasite species (with the BOLD-default minimum accepted sequence match percentage identity of 97%) (Table S3, Appendix S1). Final ectoparasite species assignments were based on a combination of the intial MEGAN and barcode results (Table 3).

Blood Meals

The blood meal analysis for the identification of a reservoir host is complicated by the possibility that the ectoparasite may not have consumed blood prior to capture (resulting in no identifiable blood meal) or that the ectoparasite may have fed on multiple hosts immediately before capture (resulting in multiple blood meal signatures). In the latter case, ectoparasites may also have fed on multiple closely related species, further complicating identification through MEGAN analysis of Megablast searches. Nevertheless, multiple blood meal signatures are unlikely in ticks since they typically only feed once per molt (Pfäffle *et al.* 2013). Moreover, public databases are biased towards species of economic importance (e.g. domesticates). Blood meals taken from less-commonly analysed nonmodel species will often be misclassified as the better-characterized species (e.g. wild canid sequences were frequently misidentified as domestic dog [*Canis familiaris*]). Precise blood meal taxonomic identification is further complicated by nuclear mitochondrial DNA transpositions (numts), which are common in many mammalian species and degraded DNA samples (Lopez *et al.* 1994; Den Tex *et al.* 2010).

To combat these taxonomic biases, preliminary blood meal assignments using MEGAN were typically determined to order or family level (Table S2, Appendix S1). A minority of the putative blood meals were identifiable to tribe, genus, or species. Nonspecific sequences (e.g. those limited to higher taxonomic clades such as Boreoeutheria) and putative contaminants (sequences identified in the negative controls) were disregarded as possible blood meals. Additionally, for the blood meal analyses, we discarded libraries with fewer than 10 000 sequences after quality control since these proved unreliable between replicates.

Table 3 Comparison of final EctoBaits ectoparasite and blood meal identifications compared to those expected from morphology and PCR assays. The 'Sequences' column lists the total number of quality-controlled sequences analysed for each library. Samples that either had too few sequences for reliable blood meal identification or had no identifiable blood meal are noted. Sample replicates are clustered together with the varying probe dilutions distinguished by a terminal 'a' (1:5 probe dilution) or 'b' (1:10 probe dilution)

Library	Source	Sequences	EctoBaits ectoparasite	Expected ectoparasite	EctoBaits blood meal	Expected blood meal
VirT01a	Virginia	243 731	*Ixodes scapularis*	*Ixodes* sp.	Passeriformes	*Thryothorus ludovicianus*
VirT01b	Virginia	181 457	*I. scapularis*	*Ixodes* sp.	Passeriformes	*T. ludovicianus*
VirT02a	Virginia	152 120	*Dermacentor variabilis*	*D. variabilis*	*Didelphis virginiana*	*D. virginiana*
VirT02b	Virginia	342 360	*D. variabilis*	*D. variabilis*	*D. virginiana*	*D. virginiana*
VirT03a	Virginia	142 216	*Amblyomma americanum*	*A. americanum*	*Vulpes vulpes*	*V. vulpes*
VirT03b	Virginia	226 975	*A. americanum*	*A. americanum*	*V. vulpes*	*V. vulpes*
VirT04a	Virginia	235 996	*I. scapularis/dentatus*	*Ixodes* sp.	*Sylvilagus floridanus*	*S. floridanus*
VirT04b	Virginia	177 000	*I. scapularis/dentatus*	*Ixodes* sp.	Leporidae	*S. floridanus*
VirT05a	Virginia	417 318	*A. americanum*	*A. americanum*	*Odocoileus virginianus*	*O. virginianus*
VirT05b	Virginia	562 081	*A. americanum*	*A. americanum*	*O. virginianus*	*O. virginianus*
VirT06a	Virginia	186 779	*D. variabilis*	*D. variabilis*	Not identified	*V. vulpes*
VirT06b	Virginia	308 458	*D. variabilis*	*D. variabilis*	Not identified	*V. vulpes*
VirT07a	Virginia	232 105	*Ixodes cookei*	*Ixodes* sp.	*Marmota monax*	*M. monax*
VirT07b	Virginia	240 719	*I. cookei*	*Ixodes* sp.	*M. monax*	*M. monax*
VirT08a	Virginia	242 625	*D. variabilis*	*D. variabilis*	*V. vulpes*	*V. vulpes*
VirT08b	Virginia	337 021	*D. variabilis*	*D. variabilis*	*V. vulpes*	*V. vulpes*
VirT09a	Virginia	246 820	*A. americanum*	*A. americanum*	*O. virginianus*	*O. virginianus*
VirT09b	Virginia	109 877	*A. americanum*	*A. americanum*	*O. virginianus*	*O. virginianus*
VirT10a	Virginia	128 272	*I. scapularis*	*Ixodes* sp.	*O. virginianus*	*Buteo jamaicensis*
VirT10b	Virginia	41 957	*I. scapularis*	*Ixodes* sp.	*O. virginianus*	*B. jamaicensis*
KenT01a	Kenya	186 228	*Rhipicephalus* sp.	*Rhipicephalus* sp.	Not identified	*Lycaon pictus*
KenT01b	Kenya	73 134	*Rhipicephalus* sp.	*Rhipicephalus* sp.	Not identified	*L. pictus*
KenT02a	Kenya	400 268	*Rhipicephalus muhsamae*	*Rhipicephalus* sp.	*L. pictus*	*L. pictus*
KenT02b	Kenya	137 654	*R. muhsamae*	*Rhipicephalus* sp.	*L. pictus*	*L. pictus*
KenT03a	Kenya	51 949	*Rhipicephalus pulchellus*	*R. pulchellus*	Not identified	Questing Tick
KenT03b	Kenya	36 905	*R. pulchellus*	*R. pulchellus*	Bovidae	Questing Tick
KenT04a	Kenya	65 822	*Ixodes* sp.	*Ixodes* sp.	Not identified	*Elephantulus rufescens*
KenT04b	Kenya	20 381	*Ixodes* sp.	*Ixodes* sp.	Not identified	*E. rufescens*
KenT05a	Kenya	182 811	*I. scapularis*	*Ixodes* sp.	*E. rufescens*	*E. rufescens*
KenT05b	Kenya	73 620	*I. scapularis*	*Ixodes* sp.	*Elephantulus* sp.	*E. rufescens*
KenT06a	Kenya	43 962	*R. pulchellus*	*R. pulchellus*	Not identified	Questing Tick
KenT06b	Kenya	13 475	*R. pulchellus*	*R. pulchellus*	Not identified	Questing Tick
KenT07a	Kenya	62 305	*Haemaphysalis leachii*	*H. leachii*	*Mastomys natalensis*	*M. natalensis*
KenT07b	Kenya	67 800	*H. leachii*	*H. leachii*	*M. natalensis*	*M. natalensis*
KenT08a	Kenya	1626	*Rhipicephalus* sp.	*Rhipicephalus* sp.	Too few sequences	*Aethomys hindei*
KenT08b	Kenya	3836	*Rhipicephalus* sp.	*Rhipicephalus* sp.	Too few sequences	*A. hindei*
KenT09a	Kenya	5724	*R. pravus*	*Rhipicephalus* sp.	Too few sequences	*E. rufescens*
KenT09b	Kenya	1467	*R. pravus*	*Rhipicephalus* sp.	Too few sequences	*E. rufescens*
KenT10a	Kenya	37 413	*Haemaphysalis/Ixodes* sp.	*H. leachii*	*Acomys percivali*	*A. percivali*
KenT10b	Kenya	65 247	*Haemaphysalis/Ixodes* sp.	*H. leachii*	*A. percivali*	*A. percivali*
KenT11b	Kenya	14 567	*R. pravus*	*R. pravus*	Not identified	Questing Tick
KenT12b	Kenya	12 551	*R. pulchellus*	*Rhipicephalus* sp.	Not identified	Questing Tick
KenT13b	Kenya	42 110	*R. praetextatus*	*Rhipicephalus* sp.	Not identified	Questing Tick
KenT14b	Kenya	19 188	*R. praetextatus*	*R. praetextatus*	*Canis* sp.	Questing Tick
KenT15b	Kenya	15 752	*R. praetextatus*	*Rhipicephalus* sp.	Not identified	Questing Tick
KenF01a	Kenya	30 236	*Xenopsylla humilis*	*Xenopsylla* sp.	Not identified	*Gerbilliscus robustus*
KenF01b	Kenya	43 596	*X. humilis*	*Xenopsylla* sp.	Not identified	*G. robustus*
KenF02a	Kenya	1772	*X. humilis*	*Xenopsylla* sp.	Too few sequences	*G. robustus*
KenF02b	Kenya	413	Not identified	*Xenopsylla* sp.	Too few sequences	*G. robustus*

We then identified the majority blood meal to species by extracting putative blood meal sequences from the MEGAN results. To maximize the possibility of generating a useable barcode, we included less informative and potential contaminant sequences that had been initially discarded from the MEGAN assignment. The extracted

sequences were mapped against the cytochrome c oxidase I (*cox1*) barcoding region (Genbank accession KC488279.1) and a consensus blood meal *cox1* sequence was determined using Geneious 8.1 (medium sensitivity alignment, 10 alignment iterations) (Biomatters, Ltd. 2015). The consensus barcode was then compared against BOLD and Genbank to determine blood meal species (with a minimum accepted sequence match percentage identity of 97%) (Table S3, Appendix S1). We identified a small number of fox (*Vulpes vulpes*) *cox1* sequences in the controls. We therefore discarded barcodes matching *V. vulpes* unless these results were congruent with the initial MEGAN results. In some cases, we were unable to determine a high-quality barcode sequence (Table S3, Appendix S1), so final blood meal assignments for these individuals were based solely on the preliminary MEGAN results (Table 3).

Pathogens

Pathogens were identified by simple presence/absence after MEGAN sequence filtering. Putative pathogens were identified to family, genus, species or strain depending on sequence specificity (Table 4). Since pathogen sequences were typically expected to be infrequent, we included all pathogen taxa that exceeded the MEGAN sequence filtration described above (see "Sequence Identification and Assay Validation") as putative positive infections. However, we interpreted all pathogen identifications in poorly sequenced single-indexed libraries (<50 000 sequences) with caution since we found that these samples were prone to poor sample reproducibility and therefore potentially spurious results.

Results and discussion

EctoBaits efficiently captured target molecules, enriching the libraries for these DNAs by at least 37.4-fold (see Table S5, Appendix S1). Single-indexed libraries yielded between 413 and 562 081 quality-controlled sequences, with similar taxonomic composition between sample replicates (Figs 1 and 2). The single-indexed extraction and library controls (Sequence Read Archive accessions SRS912001–SRS912002) yielded small quantities of sequences (4983 quality-controlled sequences for the extraction control using a 1:5 probe dilution, 1720 sequences for the extraction control using a 1:10 probe dilution, 395 sequences for the library control) that matched common laboratory contaminants (*Homo sapiens*, *Mus musculus*, Viridiplantae) and cross-contaminants from the other libraries including ticks (*Amblyomma americanum*, *Haemaphysalis* sp., *Ixodes scapularis*, *Rhipicephalus* sp.), Siphonaptera, hosts (*Menetes berdmorei*, *Ictidomys tridecemlineatus*, *V. vulpes*, Bovidae) and pathogens (*Staphylococcus* sp., *Rickettsia* sp., *Proteus mirabilis*). Double-indexed tick libraries yielded between 12 551 and 42 110 quality-controlled sequences. Both replicate libraries from three individuals (KenT08, KenT09 and KenF02) were excluded from the blood meal analyses due to low sequence count (i.e., below 10 000 quality-controlled sequences). Samples for which fewer than 10 000 quality-controlled sequences were generated could be re-run on another sequencing lane to attempt to obtain the recommended minimum sequences. However, some poor-quality samples may not hybridize as expected, and thus not recover the recommended number of sequences. Composition of taxa identified in each library was dependent on the sample's status rather than the probe dilution (Fig. 2). For instance, questing ticks (KenT03, KenT06, KenT11–KenT15) had few identifiable blood meal sequences. Conversely, VirT08 was dominated by blood meal DNA, presumably due to feeding immediately before collection.

Ectoparasites

The EctoBaits assay correctly identified the ectoparasite as a member of the Ixodida or the Pulicidae in all but one very poor quality flea library (total sequences after quality control = 413) (Table 3). The remaining three flea libraries (representing both individuals) were identifiable to species (*Xenopsylla humilis*). The assay identified tick ectoparasites to genus in 24 of 25 individuals (43 of 45 libraries), which matched the expected genus in all cases (Table 3). In addition, 21 ticks (37 libraries) were identifiable to species. This is an improvement over the PCR/morphology approach, which identified only 13 individuals to species reliably (Table 1). Furthermore, the individuals unidentifiable to species by the EctoBaits assay were from Kenya, an area in which ticks remain poorly genetically characterized. The unresolved individuals may be better classified after further taxonomic sampling in the region.

In several cases, the MEGAN-based species assignment proved sensitive to the taxonomic representativeness of the sequences curated in public databases. VirT04 was initially identified by morphology as *I. scapularis* and by PCR assay as *Ixodes* sp. The MEGAN approach identified the individuals as *I. scapularis*, in line with the morphological assignment, but the EctoBaits barcode sequence matched BOLD's *Ixodes dentatus* sequence with very high identity (99.69%). *Ixodes dentatus* is currently only represented in Genbank by five 5.8S ribosomal RNA sequences. Since these species are closely related, the VirT04 sequences are probably misclassified *Ixodes dentatus* sequences. Similarly, KenT05 was identified by

Table 4 Comparison of pathogen and macroparasite identifications between EctoBaits and PCR assays. Sample replicates are clustered together with the varying probe dilutions distinguished by a terminal 'a' (1:5 probe dilution) or 'b' (1:10 probe dilution). The *Francisella tularensis*, *Proteus mirabilis* and *Staphylococcus pseuditermedius* PCR assays were performed after their initial discovery in the EctoBaits assay results. None of the tested primer sets amplified the *Onchocerca ochengi* DNA we observed in the VirT05 and VirT10 EctoBaits results

Library	Sequences	EctoBaits pathogens/Macroparasites	PCR pathogens/Macroparasites
VirT01a	243 731	*Borrelia burgdorferi*	*B. burgdorferi*
VirT01b	181 457	*B. burgdorferi*	*B. burgdorferi*
VirT02a	152 120	*F. tularensis*	*F. tularensis*
VirT02b	342 360	*F. tularensis*	*F. tularensis*
VirT03a	142 216	*Hepatozoon* sp., *Rickettsia* sp.	*Hepatozoon canis*
VirT03b	226 975	Adeleorina	*H. canis*
VirT04a	235 996	None	*Anaplasma phagocytophilum*
VirT04b	177 000	None	*A. phagocytophilum*
VirT05a	417 318	*O. ochengi*, *Rickettsia* sp. (Spotted Fever Group), *Theileria* sp.	*Rickettsia* sp., *Theileria* sp.
VirT05b	562 081	*O. ochengi*, *Rickettsia* sp. (Spotted Fever Group), *Theileria* sp.	*Rickettsia* sp., *Theileria* sp.
VirT06a	186 779	*Francisella* sp., *Rickettsia montanensis*	*F. tularensis*, *R. montanensis*
VirT06b	308 458	*Francisella* sp., *R. montanensis*	*F. tularensis*, *R. montanensis*
VirT07a	232 105	None	None
VirT07b	240 719	None	None
VirT08a	242 625	*Staphylococcus pseudintermedius*	*Ehrlichia chaffeensis*, *H. canis*, *S. pseudintermedius*
VirT08b	337 021	*S. pseudintermedius*	*E. chaffeensis*, *H. canis*, *S. pseudintermedius*
VirT09a	246 820	*Coxiella burnetii* (endosymbiont of *Amblyomma americanum*), *Rickettsia* sp. (Spotted Fever Group)	*C. burnetii*, *Hepatozoon felis*
VirT09b	109 877	*C. burnetii* (endosymbiont of *A. americanum*), *Rickettsia* sp. (Spotted Fever Group)	*C. burnetii*, *Hepatozoon felis*
VirT10a	128 272	*O. ochengi*, *P. mirabilis*, *Rickettsia* sp. (Spotted Fever Group)	*P. mirabilis*, *Rickettsia* sp., Piroplasmida
VirT10b	41 957	*O. ochengi*, *Rickettsia* sp.	*P. mirabilis*, *Rickettsia* sp., Piroplasmida
KenT01a	186 228	*P. mirabilis*, *Rickettsia conorii*	*H. canis*, *Rickettsia conorii*
KenT01b	73 134	Adeleorina, *Rickettsia conorii*	*H. canis*, *Rickettsia conorii*
KenT02a	400 268	*C. burnetii*, *Hepatozoon* sp., *P. mirabilis*, *Rickettsia* sp.	*C. burnetii*, *H. canis*, *P. mirabilis*
KenT02b	137 654	*C. burnetii*, *Hepatozoon* sp., *P. mirabilis*	*C. burnetii*, *H. canis*, *P. mirabilis*
KenT03a	51 949	*P. mirabilis*, *Rickettsia* sp. (Spotted Fever Group)	*A. phagocytophilum*
KenT03b	36 905	*O. ochengi*	*A. phagocytophilum*
KenT04a	65 822	*P. mirabilis*, *Rickettsia* sp. (Spotted Fever Group)	*Rickettsia* sp.
KenT04b	20 381	*Rickettsia* sp. (Spotted Fever Group)	*Rickettsia* sp.
KenT05a	182 811	*P. mirabilis*, *Rickettsia* sp. (Spotted Fever Group)	*Rickettsia* sp.
KenT05b	73 620	*Rickettsia* sp. (Spotted Fever Group)	*Rickettsia* sp.
KenT06a	43 962	*C. burnetii*, *P. mirabilis*, *Rickettsia* sp. (Spotted Fever Group)	None
KenT06b	13 475	*C. burnetii*, *P. mirabilis*, *Rickettsia* sp. (Spotted Fever Group)	None
KenT07a	62 305	*Bartonella* sp., *Rickettsia* sp. (Spotted Fever Group)	*Bartonella* sp., *Rickettsia* sp.
KenT07b	67 800	*Bartonella* sp., *Rickettsia* sp. (Spotted Fever Group)	*Bartonella* sp., *Rickettsia* sp.
KenT08a	1626	Rickettsiaceae	*Hepatozoon* sp.
KenT08b	3836	Coxiellaceae, Rickettsiaceae	*Hepatozoon* sp.
KenT09a	5724	*P. mirabilis*, *Rickettsia* sp. (Spotted Fever Group)	*B. burgdorferi*, *P. mirabilis*
KenT09b	1467	*Rickettsia* sp. (Spotted Fever Group)	*B. burgdorferi*, *P. mirabilis*
KenT10a	37 413	*C. burnetii*, *P. mirabilis*, Rickettsiaceae	*P. mirabilis*
KenT10b	65 247	None	*P. mirabilis*
KenT11b	14 567	*C. burnetii*	*C. burnetii*, *H. canis*
KenT12b	12 551	Coxiellaceae	*C. burnetii*, *Rickettsia* sp.
KenT13b	42 110	*C. burnetii*	*C. burnetii*
KenT14b	19 188	*C. burnetii*	None
KenT15b	15 752	*C. burnetii*	*C. burnetii*, *Rickettsia* sp.
KenF01a	30 236	*Bartonella* sp., *P. mirabilis*, *Rickettsia* sp.	None
KenF01b	43 596	None	None
KenF02a	1772	*P. mirabilis*, *Rickettsia* sp.	None
KenF02b	413	None	None

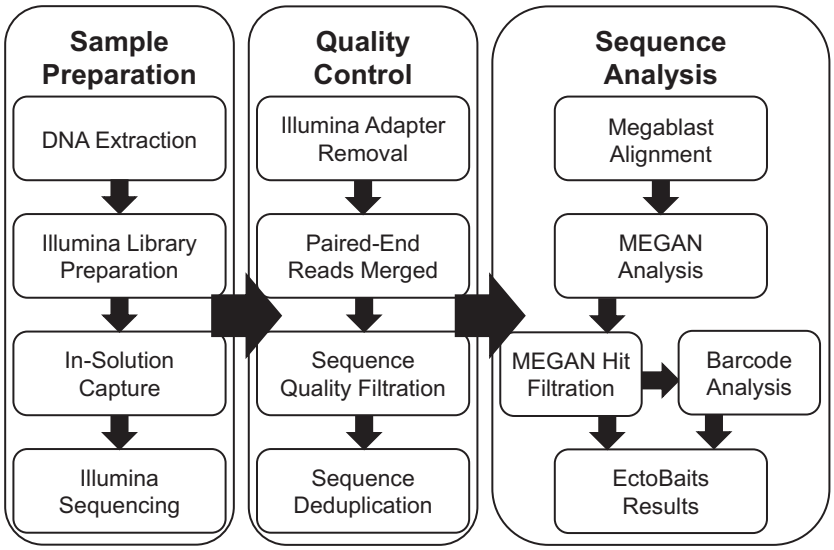

Fig. 1 Flowchart outlining the analysis pipeline of the EctoBaits assay.

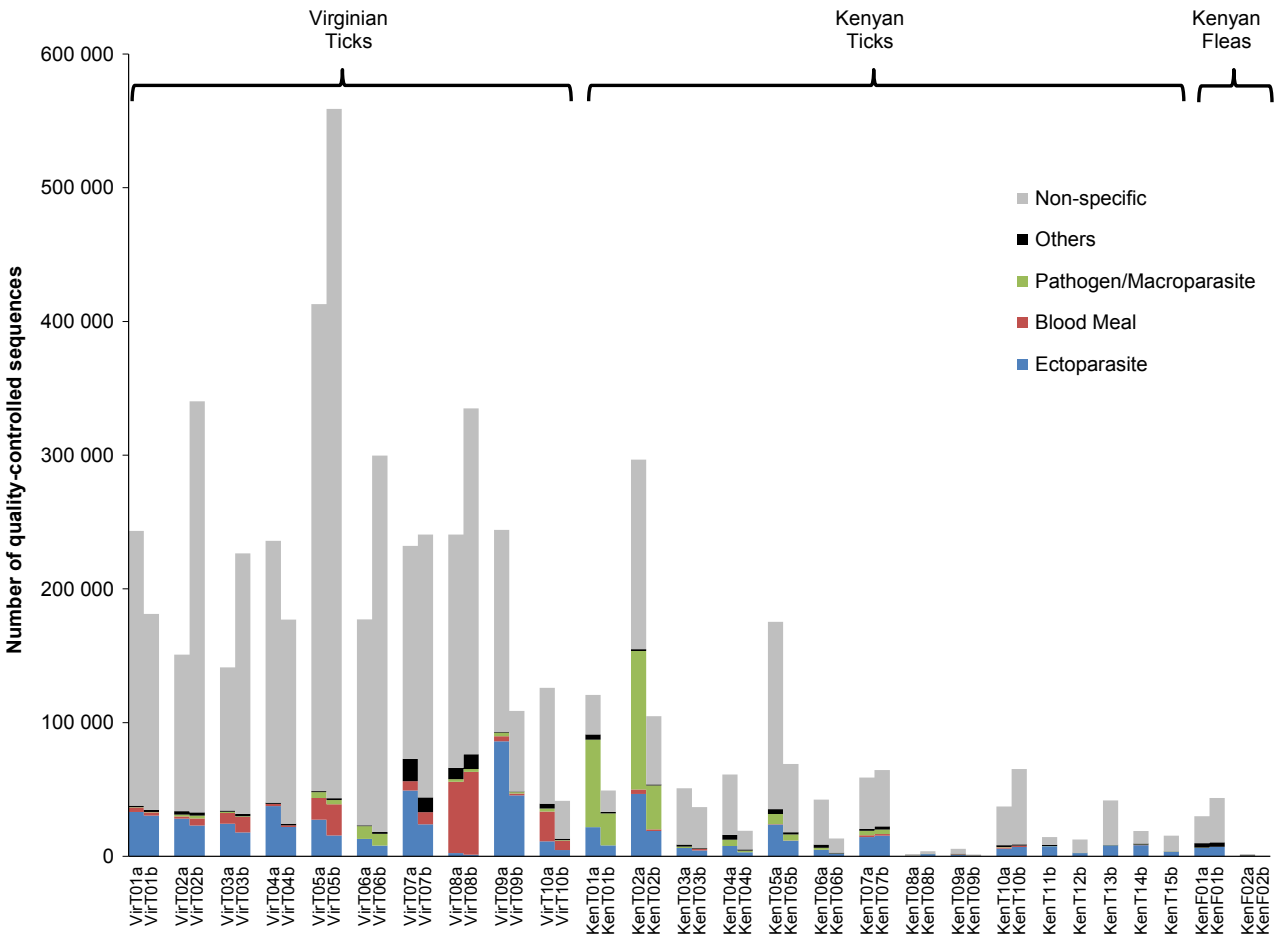

Fig. 2 Total library compositions based on MEGAN analysis. Sample replicates (probe dilutions) are plotted next to each other with labels ending in 'a' or 'b' indicating 1:5 and 1:10 probe dilutions, respectively. Library sequence depths and compositions varied greatly between samples. All pathogen/macroparasite sequences are clustered together for figure clarity. All 'Nonspecific' sequences were unidentifiable by MEGAN analysis since they were either unassigned or assigned to a very high taxonomic rank (such as 'Metazoa'). 'Others' include laboratory contaminants, environmental and microbiomic sequences, and sequences of uncertain origin.

MEGAN assignment as *I. scapularis*. As this individual was from Kenya, which is outside the range of *I. scapularis*, these sequences probably originate from a closely related, unsequenced species.

Blood meals

We identified DNA corresponding to blood meals for 15 of 27 individuals (28 of 49 libraries). EctoBaits identified 12 individuals' blood meals to species level (Table 3). Of these, 10 individuals were identified to species in both replicate libraries, one individual was identified to genus in the replicate library and one to family in the replicate library. One individual's (VirT01) blood meal was identified reliably to the order level. One questing tick (KenT03) had a small quantity of bovid DNA in one replicate, and another (KenT14) had a small number of sequences matching *Canis*. These may represent the remains of previous blood meals or could derive from contamination (e.g. jumping PCR, laboratory reagents).

We noted only one discrepancy between the EctoBaits and the expected results. One tick (VirT10) obtained from a red-tailed hawk (*Buteo jamaicensis*) was found to have fed on white-tailed deer (*Odocoileus virginianus*) blood rather than hawk blood. As VirT10 is a black-legged tick (*I. scapularis*), which commonly feed on deer, this is probably a true result rather than a false positive. Moreover, the swamping out of a hawk blood meal DNA signature by contaminant deer DNA is unlikely since birds have nucleated erythrocytes, while mammals do not. The VirT10 result may be derived from a tick that had recently fed on deer blood and incidentally became associated with a red-tailed hawk before sample collection or from a laboratory error.

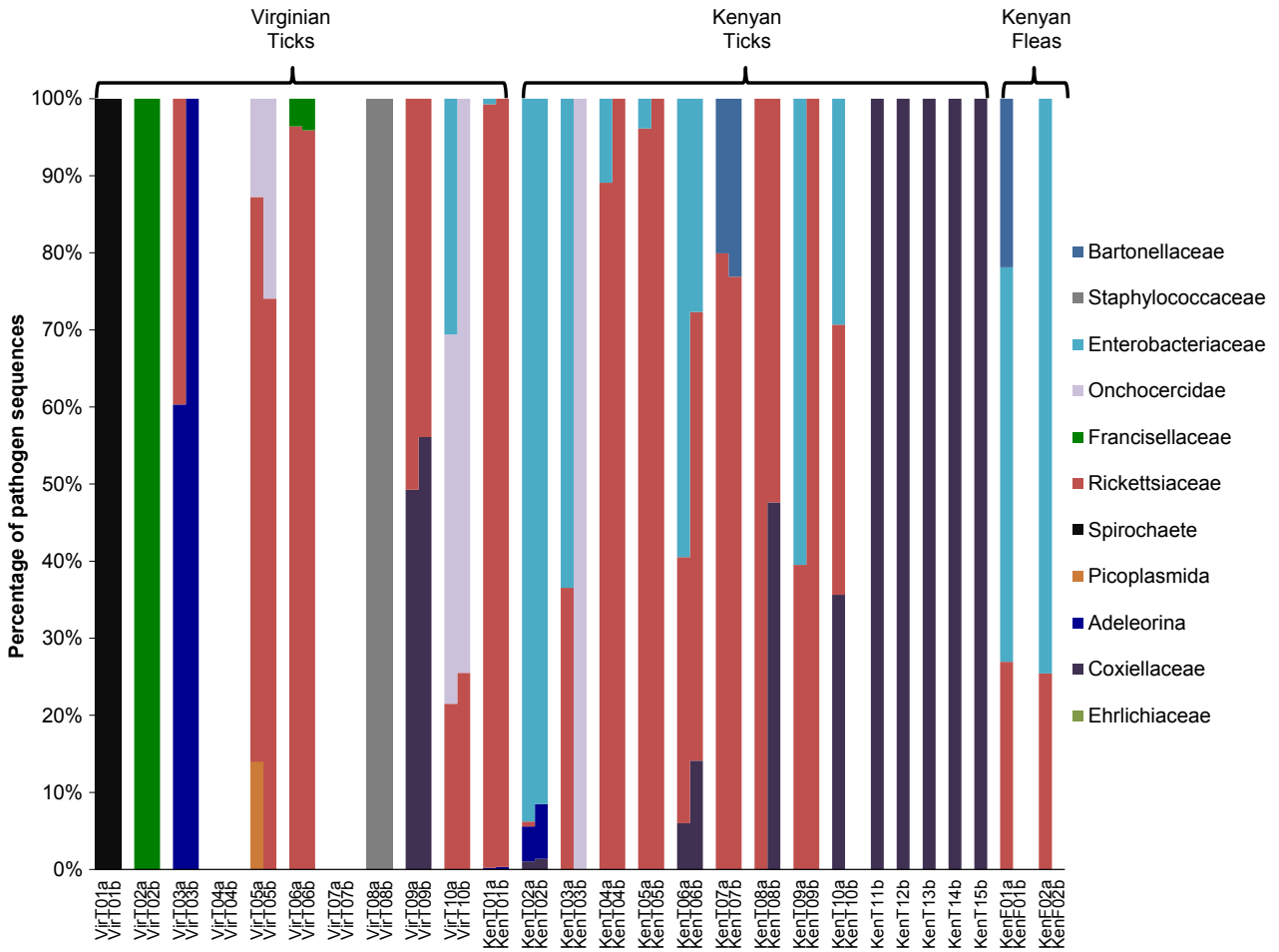

Fig. 3 Composition of library pathogen and macroparasite sequences identified by MEGAN analysis. The pathogen/macroparsite sequences have been broken down by diagnostic pathogen family/suborder. To facilitate comparison and account for differing library sequence depths, the sequence counts are scaled by the total number of pathogen/macroparasite sequences in each library. Sample replicates (probe dilutions) are plotted next to each other with labels ending in 'a' or 'b' indicating 1:5 and 1:10 probe dilutions respectively. Pathogen/macroparasite identifications are highly replicable in high-depth libraries (more than 50 000 sequences after quality control).

Pathogen/Macroparasite	Polymerase chain reaction	EctoBaits	Replicated EctoBaits	Confirmed
Anaplasma	2	0	0	0
Bartonella	1	2	1	1
Borrelia	2	1	1	1
Coxiella	6	8	3	5
Ehrlichia	1	0	0	0
Francisella	2	2	2	2
Hepatozoon	7	2	1	2
Onchocerca	0	3	2	0
Proteus	4	11	2	4
Rickettsia	9	15	11	7
Staphylococcus	1	1	1	1
Theileria	1	1	0	1

Table 5 Summary of EctoBaits and polymerase chain reaction pathogen/macroparasite screens' relative sensitivities. This table only includes genus-level EctoBaits and PCR identifications. Values are frequency counts out of the 27 ectoparasite individuals included in this study. Diagnostic pathogen sequences that were identified in both EctoBaits sample replicates are included in the 'Replicated EctoBaits' column. Positive pathogen results that were positive in both the EctoBaits and polymerase chain reaction assays are included in the 'Confirmed' column

Pathogens

The EctoBaits assay produced highly consistent pathogen results between the EctoBaits sample replicates (Fig. 3; Table 5). Pathogen-detection replicability decreased in poorly sequenced single-indexed libraries (<50 000 quality-controlled sequences) due to reduced power to detect pathogens as well as to exclude contaminant sequences and spurious Genbank assignments (Figs 3 and S2, Appendix S1). Ignoring sequencing depth and excluding the five nonreplicated double-indexed libraries, EctoBaits identified 42 putative infections to genus level, of which 24 (57%) of were identified in both sample replicates (Tables 4 and 5). This replication rate is highly biased by *P. mirabilis* (see below), which was present in small quantities in the majority of the poorly sequenced Kenyan ectoparasites (Table 4; Fig. 3). Moreover, *P. mirabilis* was found to be present in one of the extraction controls, suggesting that some of these may be false positives. Discounting *P. mirabilis*, 22 of 31 (71%) of putative infections were confirmed between the two sample replicates (Table 5).

The EctoBaits assay's detection ability was comparable to PCR assays for most pathogens (Tables 4 and 5). Notably, the EctoBaits assay identified more instances of *Rickettsia* (15 vs. seven individuals for the EctoBaits and PCR-assays respectively) and *Coxiella* (eight vs. six individuals) infections compared to the PCR assays (Table 5). However, some *Rickettsia* results may be false positives given the pathogen's presence in one of the extraction controls and in one of the spiked-blood libraries (Appendix S1). EctoBaits could distinguish between nonpathogenic *Coxiella* endosymbionts and potentially pathogenic *Coxiella* strains, which the short (<160 bp) PCR-derived sequences could not (Table 4). Conversely, the EctoBaits assay had reduced ability to detect Ehrlichiaceae (*Anaplasma* and *Ehrlichia*) and *Hepatozoon* infections. EctoBaits detected no Ehrlichiaceae infections, while the PCR assays detected two *Anaplasma* infections and one *Ehrlichia* infection. This is presumably due to an insufficient number of high-affinity Ehrlichiaceae-specific baits currently included in the EctoBaits assay. EctoBaits identified two individuals infected with *Hepatozoon*, while the PCR-assays identified seven individuals (Tables 4 and 5). In this case, the PCR assays provided species-level taxonomic resolution, while EctoBaits did not (Table 4). Additionally, individual PCR assays may have greater detection sensitivity than EctoBaits for low-level infections (see Appendix S1).

We identified several pathogens and a macroparasite in the EctoBaits assay that we had not previously assayed by PCR (Table 4). These included two instances of *Francisella tularensis* (the causative agent of tularemia), 11 instances of *P. mirabilis* and one *Staphylococcus pseudintermedius* infection. *Francisella tularensis* is a well-known tick-borne pathogen (e.g. Pfäffle *et al.* 2013). *Proteus mirabilis* has been isolated from African ticks (Omoya *et al.* 2013). We also identified three instances of the onchocercid filiarial nematode, *Onchocerca ochengi*, which we include here due to its vector-based transmission and pathogenicity. Onchocercid infections have been previously reported in North American ixodid ticks (Namrata *et al.* 2014). Furthermore, all *Francisella* and *Staphylococcus* instances were replicated between the two dilutions of the libraries, strongly supporting their authenticity. Two each of the *Proteus* and onchocercid instances were replicated between the two dilutions of the libraries.

Since these infections were unexpected, it was necessary to confirm their authenticity in order to rule out the possibility that they were derived from artifacts of the EctoBaits assay. To confirm these newly identified pathogens, we subjected the putatively infected samples (with additional samples included as negative controls) to PCR amplification following previously published protocols (Versage *et al.* 2003; Bannoehr *et al.* 2009; Zhang *et al.*

2013; Namrata *et al.* 2014). We were able to confirm the three pathogens (*F. tularensis*, *P. mirabilis* and *S. pseudintermedius*) using these PCR assays. The PCR assays for onchocercid filiarial disease exhibited nonspecific amplification using published primer sets. Further investigation would be required to confirm the onchocercid assay results.

The novel infections probably hybridized with nonspecific probes from related organisms included in the assay. This is not unexpected, as some studies have noted hybridization with up to 19% divergence from probe sequences (Hawkins *et al.* 2015) and hybridization arrays designed from modern strains can capture divergent ancient pathogen genomes (e.g. Bos *et al.* 2015).

Assay validation and optimization

Library sequence depths. There was a significant, but nonlinear, relationship between single-indexed library sequence-depth and identification confidence (Spearman's $\rho = 0.642$, two-tailed $P = 2.68 \times 10^{-5}$ for genus; $\rho = 0.554$, $P = 9.62 \times 10^{-5}$ for species). However, we observed a pattern of diminishing returns of phylogenetic information with increasing sample sequence depth. Therefore, to maximize cost and time efficiency, it is critical to optimize the sample sequence depth. We plotted species and genus confidence vs. sequence depth (Fig. 4) and fit logarithmic regression curves to these data. We excluded the double-indexed libraries from these comparisons due to their differing library construction and sequencing chemistry. For a minimum 65% species confidence (84% genus confidence), our model predicted that 88 429 sequences were required. Increasing the sequence depth beyond this point produced minimal benefit. For instance, 399 308 sequences were required to obtain an 80% species confidence. Conversely, we found that libraries that had fewer than 10 000 and 50 000 sequences after quality control were difficult to analyse for blood meals and pathogens respectively due to background noise and capture stochasticity. Caution is therefore required to ensure erroneous calls are not made when analysing low-depth libraries. We recommend a target sequence depth of ~90 000 sequences per sample to ensure high-quality identifications and maximize sample throughput. Additionally, replicating each library helps ensure identification reliability and sensitivity. Samples which yielded fewer than 10 000 sequences may require a second enrichment without multiplexing to remove competition for bait sequences. We recommend that EctoBaits bud-

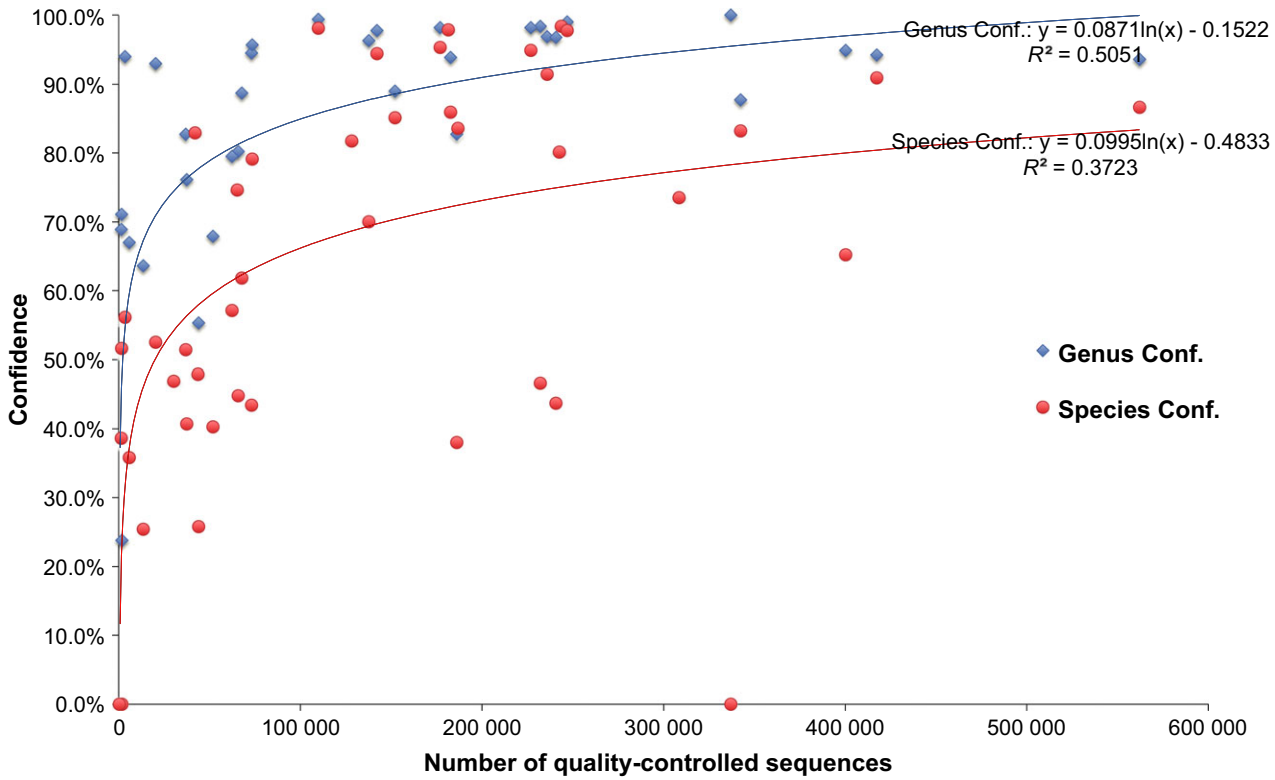

Fig. 4 Total single-indexed library sequence depth vs. EctoBaits ectoparasite identification genus and species confidences. The fit logarithmic models and their corresponding R^2 values are also given.

gets include an additional 20–30% to cover the samples that require individual enrichment.

Probe dilution. Increasing the probe concentration did not improve capture of identifiable sequences. There were no differences in terms of sequence yield, ectoparasite genus confidence or ectoparasite species confidence (two-tailed paired *t*-tests, $P > 0.05$ in all cases). The two dilutions identified putative blood meals equally accurately (Table S2, Appendix S1, Fig. 4). We therefore recommend that 1:10 dilution of probes be used for future capture assays to minimize laboratory costs.

EctoBaits cost comparison. The EctoBaits assay permits a higher throughput of samples, with a lower per-sample cost, than traditional PCR and Sanger sequencing strategies. A single PCR assay (including Sanger sequencing) costs ~$20 per sample. Thus, the 12 PCR assays used here (including the pathogens identified initially via EctoBaits) costs ~$240 per sample. The in-solution capture costs ~$50 per sample (based on 1:10 dilution of the probes and 200 000 paired-end reads on an Illumina MiSeq) and is completed within a single experiment.

Future improvements of the EctoBaits assay

Currently, the EctoBaits assay is primarily restricted to the analysis of ixodid ticks and a limited sampling of flea taxa. For future improvements of the EctoBaits design, we propose to add probes for increased ectoparasite taxonomic coverage (such as argasid ticks, additional fleas, mites, lice and mosquitoes). The assay currently omits viral diseases (such as tick-borne encephalitis) and only includes one filarial disease. Future revisions of the assay will increase the probe density for anaplasmosis and ehrlichiosis, which we found to be poorly captured by the current assay. Additional species-specific *Hepatozoon* probes will also improve *Hepatozoon* detection and taxonomic resolution.

Jumping PCR artifacts complicated identification of blood meals and pathogens in low-sequencing-depth libraries (<10 000 and <50 000 sequences for blood meals and pathogens, respectively). Jumping PCR occurs when molecules of two independent origins switch templates during PCR and create a chimeric molecule of hybrid origin. In the low-coverage libraries, we identified sequences that clearly derived from other libraries even after sequence filtration. Additionally, background laboratory contamination in the low-coverage libraries was difficult to exclude since contamination levels were relatively constant across the libraries (rather than proportional to sequence depth). Other studies have found jumping PCR to be a problem for multiplexed high-throughput sequencing experiments, particularly with single-indexed libraries (Kircher *et al.* 2012; Hawkins *et al.* 2015). To mitigate jumping-PCR artifacts, we recommend double-indexed sequencing in the future usage of the assay and increased sequence sequence depth to ensure that any remaining sequences are filtered out of the final data set.

Increasing numbers of sequences deriving from non-model organisms in public databases such as Genbank and BOLD will also improve the genetic identification of ectoparasites and blood meals. Genetic data are limited for many of the taxa included in this study. Currently, EctoBaits more accurately identifies North American than African taxa, mostly due to bias in genomic sampling (ticks identified to species: 90% of Virginian ticks vs. 67% of Kenyan ticks; blood meals identified to species: 80% of Virginian ticks vs. 44% of Kenyan nonquesting ticks; Table 3). Furthermore, due to stochasticity during DNA extraction, library preparation, and in-solution capture, we recommend that researchers replicate and validate putative positive pathogen results from the current version of the EctoBaits assay.

Finally, little is known about blood meal DNA taphonomy. Research is needed to test how long after feeding EctoBaits can detect a blood meal and whether EctoBaits can detect previous blood in ectoparasites that feed only once per life stage. We also have yet to establish whether there are differences in pathogen and blood meal detectability across developmental stages of ectoparasites.

Conclusions

The EctoBaits assay is an accurate, efficient and effective method to identify northeastern American and east African ectoparasites and their most recent blood meals while simultaneously screening them for various pathogens. It provides data comparable to a multitude of PCR assays while consuming only a fraction of the DNA samples that can be utilized for many more assays and future analyses. EctoBaits can also identify unknown pathogens that would otherwise be overlooked using a targeted PCR-based approach. However, EctoBaits is prone to false positives (especially for *Rickettsia* sp. and *P. mirabilis*) and is currently limited to ixodid ticks and several flea taxa. Furthermore, its ability to identify remnant blood meals in unengorged questing ticks is limited. We demonstrate the utility of EctoBaits for studies dealing with disease transmission in the eastern United States and East Africa using a single enrichment for identification of the host, parasite and pathogen. The assay's power will only improve with the increased comprehensiveness of sequence databases. The time and cost required to identify the same (and in some cases more) pathogens are much

lower using EctoBaits than multiple PCRs. Finally, EctoBaits allows for the documentation of multiple hosts and mixed infections of pathogens.

Acknowledgements

MYcroarray assisted with quality control steps for the bait production (J-M Rouillard). Logistical support and testing of these probes occurred in the Center for Conservation and Evolutionary Genetics with assistance from Nancy McInerney. Brian Allan, Lorenza Beatti, Katharina Dittmar, Ralph Eckerlin, Ashley Hintz, John Montinieri, Richard Robbins and Serena Zhao helped in collection, preparation and identification of the ectoparasites. We are grateful for the assistance of Patrick Jansen, Margaret Kinnaird, Allen Richards, and Chelsea Wright throughout this project. We thank the Smithsonian Grand Challenges Biodiversity Consortium, the Smithsonian DNA Barcode Network, the Morris Animal Foundation (D14ZO-308), the National Geographic Society (8846-10), the Smithsonian Institution Women's Committee (SWC 22 and SWC 44), and the Center for Conservation and Evolutionary Genetics for funding. We thank the Kenya Wildlife Service, National Museums Kenya, and Mpala Research Centre for logistical support and permission to conduct this research.

References

Allan BF, Goessling LS, Storch GA, Thach RE (2010) Blood meal analysis to identify reservoir hosts for *Amblyomma americanum* ticks. *Emerging Infectious Diseases*, **16**, 433–440.

Allender C, Easterday WR, Van Ert MN, Wagner DM, Keim P (2004) High-throughput extraction of arthropod vector and pathogen DNA using bead milling. *BioTechniques*, **37**, 730–734.

Andrews S (no date) FastQC: a quality control tool for high throughput sequence data. Version 0.11.2. (www.bioinformatics.babraham.ac.uk/projects/fastqc/). Downloaded December 2014.

Bai Y, Kosoy M, Recuenco S *et al.* (2011) *Bartonella* spp. in bats, Guatemala. *Emerging Infectious Diseases*, **17**, 1269–1272.

Bannoehr J, Franco A, Iurescia M, Battisti A, Fitzgerald JR (2009) Molecular diagnostic identification of *Staphylococcus pseudintermedius*. *Journal of Clinical Microbiology*, **47**, 469–471.

Biomatters, Ltd. (2015) *Geneious 8.1*. Biomatters, Ltd, Auckland, New Zealand.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bos KI, Jäger G, Schuenemann VJ *et al.* (2015) Parallel detection of ancient pathogens via array-based DNA capture. *Philosophical Transactions of the Royal Society B Biological Sciences*, **370**, 20130375.

Brites-Neto J, Roncato Duarte KM, Fernandes Martins T (2015) Tick-borne infections in human and animal population worldwide. *Veterinary World*, **8**, 301–315.

Card LR, McShea WJ, Calabrese JM *et al*. (2016) Mapping the interaction network for pathogen-tick-host interactions in Virginia. *Journal of Medical Entomology*, (in press).

Che Lah EF, Yaakop S, Ahamad M, Md Nor S (2015) Molecular identification of blood meal sources of ticks (Acari, Ixodidae) using cytochrome b gene as a genetic marker. *ZooKeys*, **478**, 27–43.

Courtney JW, Kostelnik LM, Zeidner NS, Massung RF (2004) Multiplex real-time PCR for detection of *Anaplasma phagocytophilum* and *Borrelia burgdorferi*. *Journal of Clinical Microbiology*, **42**, 3164–3168.

Criado-Fornelio A, Martinez-Marcos A, Buling-Saraña A, Barba-Carretero JC (2003) Molecular studies on *Babesia, Theileria* and *Hepatozoon* in southern Europe. Part I. Epizootiological aspects. *Veterinary Parasitology*, **113**, 189–201.

Den Tex RJ, Maldonado JE, Thorington R, Leonard JA (2010) Nuclear copies of mitochondrial genes: another problem for ancient DNA. *Genetica*, **138**, 979–984.

Eremeeva M, Yu X, Raoult D (1994) Differentiation among spotted fever group rickettsiae species by analysis of restriction fragment length polymorphism of PCR-amplified DNA. *Journal of Clinical Microbiology*, **32**, 803–810.

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. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Ghafar MW, Eltablawy NA (2011) Molecular survey of five tick-borne pathogens (*Ehrlichia chaffeensis*, *Ehrlichia ewingii*, *Anaplasma phagocytophilum*, *Borrelia burgdorferi* sensu lato and *Babesia microti*) in Egyptian farmers. *Global Veterinaria*, **7**, 249–255.

Gilbert L, Jones LD, Hudson PJ, Gould EA, Reid HW (2000) Role of small mammals in the persistence of Louping-ill virus: field survey and tick co-feeding studies. *Medical and Veterinary Entomology*, **14**, 277–282.

Goddard J (1992) Ecological studies of adult *Ixodes scapularis* in central Mississippi: questing activity in relation to time of year, vegetation type, and meteorologic conditions. *Journal of Medical Entomology*, **29**, 501–506.

Hawkins MTR, Hofman CA, Callicrate T *et al.* (2015) In-solution hybridization for mammalian mitogenome enrichment: pros, cons, and challenges associated with multiplexing degraded DNA. *Molecular Ecology Resources*. doi: http://dx.doi.org/10.1111/1755-0998.12448.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812–14817.

Hoberg EP, Brooks DR (2015) Evolution in action: climate change, biodiversity dynamics and emerging infectious disease. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **370**, 20130553.

Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4. *Genome Research*, **21**, 1552–1560.

Jones KE, Patel NG, Levy MA *et al.* (2008) Global trends in emerging infectious diseases. *Nature*, **451**, 990–993.

Kircher M (2012) Analysis of high-throughput ancient DNA sequencing data. *Methods in Molecular Biology*, **80**, 197–228.

Kircher M, Sawyer S, Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, **40**, e3.

Levin ML, Fish D (2000) Acquisition of coinfection and simultaneous transmission of *Borrelia burgdorferi* and *Ehrlichia phagocytophila* by *Ixodes scapularis* ticks. *Infections and Immunity*, **68**, 2183–2186.

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.

Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ (1994) Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution*, **39**, 174–190.

Lynen G, Zeman P, Bakuname C *et al.* (2007) Cattle ticks of the genera *Rhipicephalus* and *Amblyomma* of economic importance in Tanzania: distribution assessed with GIS based on an extensive field survey. *Experimental and Applied Acarology*, **43**, 303–319.

Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963.

Mediannikov O, Fenollar F, Socolovschi C *et al.* (2010) *Coxiella burnetii* in humans and ticks in rural Senegal. *PLoS Neglected Tropical Diseases*, **4**, e654.

Midilli K, Gargılı A, Ergonul O *et al.* (2009) The first clinical case due to AP92 like strain of Crimean-Congo hemorrhagic fever virus and a field survey. *BMC Infectious Diseases*, **9**, 90.

Nadolny RM, Wright CL, Hynes WL, Sonenshine DE, Gaff HD (2011) *Ixodes affinis* (Acari: Ixodidae) in southeastern Virginia and implications for the spread of *Borrelia burgdorferi*, the agent of Lyme disease. *Journal of Vector Ecology*, **36**, 464–467.

Nadolny RM, Wright CL, Sonenshine DE, Hynes WL, Gaff HD (2014) Ticks and spotted fever group rickettsiae of southeastern Virginia. *Ticks and Tick-Borne Diseases*, **5**, 53–57.

Namrata P, Miller J, Shilpa M *et al.* (2014) Filarial nematode infection in *Ixodes scapularis* ticks collected from southern Connecticut. *Veterinary Sciences*, **1**, 5–15.

Omoya OF, Kelly BA, Akinyosoye FA (2013) Impact of electromagnetic field on the pathogenicity of selected entomopathogenic bacteria (*Proteus* sp. and *Bacillus* sp.) on tick (*Rhipicephalus decoloratus*). *African Journal of Biotechnology*, **12**, 4683–4690.

Pfäffle M, Littwin N, Muders SV, Petney TN (2013) The ecology of tick-borne diseases. *International Journal of Parasitology*, **43**, 1059–1077.

Pichon B, Egan D, Rogers M, Gray J (2003) Detection and identification of pathogens and host DNA in unfed host-seeking *Ixodes ricinus* L. (Acari: Ixodidae). *Journal of Medical Entomology*, **40**, 723–731.

Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, **22**, 939–946.

Sabatini GD, Pinter A, Nieri-Bastos FA, Marcili A, Labruna MA (2010) Survey of ticks (Acari: Ixodidae) and their *Rickettsia* in an Atlantic rain forest reserve in the State of São Paulo, Brazil. *Journal of Medical Entomology*, **47**, 913–916.

Ujvari B, Madsen T, Olsson M (2004) High prevalence of *Hepatozoon* spp. (Apicomplexa, Hepatozoidae) infection in water pythons (*Liasis fuscus*) from tropical Australia. *Journal of Parasitology*, **90**, 670–672.

Versage JL, Severin DDM, Chu MC, Petersen JM (2003) Development of a multitarget real-time TaqMan PCR assay for enhanced detection of *Francisella tularensis* in complex specimens. *Journal of Clinical Microbiology*, **41**, 5492–5499.

Vuong N-M, Villemur R, Payment P *et al.* (2013) Fecal source tracking in water using a mitochondrial DNA microarray. *Water Research*, **47**, 16–30.

Young HS, Dirzo R, Helgen KM *et al.* (2014) Declines in large wildlife increase landscape-level prevalence of rodent-borne disease. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 7036–7041.

Young HS, Dirzo R, McCauley DJ *et al.* (2015a) Drivers of intensity and prevalence of flea parasitism on small mammals in East African savanna ecosystems. *Journal of Parasitology*, **101**, 327–335.

Young HS, McCauley DJ, Dirzo R *et al.* (2015b) Context-dependent effects of large wildlife declines on small-mammal communities in central Kenya. *Ecological Applications*, **25**, 348–360.

Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, **7**, 203–214.

Zhang W, Niu Z, Yin K, Liu P, Chen L (2013) Quick identification and quantification of *Proteus mirabilis* by polymerase chain reaction (PCR) assays. *Annals of Microbiology*, **63**, 683–689.

M.G.C., M.T.R.H., J.L., J.E.M., W.J.M., K.M.H., H.D.G., J.B. and R.C.F. designed the research. H.S.Y., L.R.C., B.A., K.M.H. and W.J.M. provided tick samples and identifications. H.D.G. and J.B. facilitated and interpreted the PCR methods. M.T.R.H. and J. L. designed the capture assay. M.T.R.H., M.G.C., K.S., and L.H.H. conducted the laboratory analyses. M.G.C. developed and conducted the bioinformatics analyses. M.G.C. and M.T.R.H. wrote most of the manuscript with input from all authors

Data accessibility

Sanger DNA sequences: Genbank accessions KR262473–KR262493, KR262495– KR262508, KR262510–KR262515, KT956186–KT956194

Illumina DNA sequences: BioProject PRJNA281123

EctoBaits probe sequences: online supporting information (Appendix S2).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 Total library compositions based on MEGAN analysis scaled by the library sequence depths.

Fig. S2 Composition of library pathogen and macroparasite sequences identified by MEGAN analysis.

Table S1 Sequences and species used to develop the EctoBaits assay.

Table S2 Preliminary metagenomics ectoparasite and blood meal identification results.

Table S3 EctoBaits cytochrome c oxidase 1 (*cox1*) results compared to the expected ectoparasite and blood meal species.

Table S4 Results of the blood spiking experiment.

Table S5 Enrichment of libraries by EctoBaits capture using a 1:10 dilution of the baits.

Table S6 The effects of species confidence thresholds on metagenomic ectoparasite identifications.

Appendix S1 Supporting methods and tables.

Appendix S2 Final probe sequences used in the EctoBaits capture assay.