ARTICLE

Received 20 Jul 2012 | Accepted 11 Mar 2013 | Published 23 Apr 2013

DOI: 10.1038/ncomms2730

Identifying sources of tick blood meals using unidentified tandem mass spectral libraries

Özlem Önder[1], Wenguang Shao[2], Brian D. Kemps[1], Henry Lam[2,3] & Dustin Brisson[1]

Rapid and reliable identification of the vertebrate species on which a disease vector previously parasitized is imperative to study ecological factors that affect pathogen distribution and can aid the development of public health programs. Here we describe a proteome profiling technique designed to identify the source of blood meals of haematophagous arthropods. This method employs direct spectral matching and thus does not require *a priori* knowledge of any genetic or protein sequence information. Using this technology, we detect remnants of blood in blacklegged ticks (*Ixodes scapularis*) and correctly determine the vertebrate species from which the blood was derived, even 6 months after the tick had fed. This biological fingerprinting methodology is sensitive, fast, cost-effective and can potentially be adapted for other biological and medical applications when existing genome-based methods are impractical or ineffective.

[1] Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19014-6019, USA. [2] Bioengineering Graduate Program, Division of Biomedical Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China. [3] Department of Chemical and Biomolecular Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China. Correspondence and requests for materials should be addressed to H.L. (email: kehlam@ust.hk) or to D.B. (email: dbrisson@sas.upenn.edu).

Trophic relationships are primary determinants of both population dynamics and the distribution and abundance of organisms[1–3]. For example, the distribution and abundance of potential host species of parasites, as well as the quality of those host species, have a dominant influence on the population densities and geographic ranges of parasites[1,4,5]. Although the distribution and abundance of parasites are important to human health and economics[6,7], little is known about the extent to which different vertebrate species affect the dynamics of most parasite species. Many haematophagous parasites also vector microbial pathogens that are the primary cause of emerging infectious diseases[8]. Further, human disease risk is often correlated with the distribution and abundance of these disease vectors[9–12]. *Ixodes scapularis* ticks are of particular public health importance as they are a common vector of several emerging infectious diseases in the eastern United States including anaplasmosis, babesiosis and Lyme disease[13].

The natural ecologies of many vectors remain obscure because of the inherent difficulties in collecting the quantitative data necessary to assess the relevance of each host species to parasite population dynamics[5,14–18]. Traditionally, labour-intensive quantification of the parasite burden on all potential wildlife species was needed[19,20]. These data are often heavily biased by the variation in trapping success among animal species[5,16]. Analysis of the vertebrate blood remaining in haematophagous insects is an alternative and highly sought-after method to determine the importance of each species to the parasite population[21–26]. Identification of the source of the blood found in vectors that are seeking a subsequent host is also considerably less labour intensive than observing vectors on live vertebrates. Moreover, collections of host-seeking vectors, as opposed to vectors collected from wildlife hosts, provide a random sample resulting in an unbiased quantification of the frequency in which each host species is parasitized.

Current methodologies employed to identify previous blood meals, including serological techniques[22,23,25], DNA-based methods[21,24] and protein analysis[26,27], confront at least one of the hurdles that prevent their practical use. First, remnants of blood from previous hosts are often severely degraded when vectors begin seeking another host, drastically reducing the efficacy of most procedures[21,28–34]. Second, all methods to date select one or several target molecules *a priori* to distinguish among a set of *a priori*-selected host species. Random degradation or digestion of the target molecules limits the success of these methodologies even though non-target molecules that can distinguish among host species may still be present[26,32]. Further, target molecules may be identical among host species that are not included in a study, thus reducing confidence in these procedures. Most importantly, suitable target molecules have not been characterized in the majority of wildlife species, thus limiting these methods to well-characterized model organisms. In summary, a robust, universal method to analyse vector blood meals is needed to identify the impact of host species on the ecology and epidemiology of disease vectors.

In the present study, we describe a non-targeted, genome-free, proteomics-based, spectral-matching[35–40] methodology designed to identify the source of blood meal remnants in haematophagous arthropods. This methodology can be used to determine the ecological importance of each species in a vertebrate community to the maintenance of populations of haematophagous arthropods. We empirically validated this methodology by identifying the source of blood remnants in blacklegged ticks (*I. scapularis*) that had fed on different vertebrate species. Remnants of blood in the ticks can be detected and correctly classified to the vertebrate species in the majority of cases, even for ticks that had completed a moult and were active for up to 6 months since their last blood meal.

Results

Construction of spectral libraries. Whole blood from 24 vertebrate species were individually trypsin-digested and analysed by liquid chromatography–tandem mass spectrometry (LC-MS^2). Reference libraries were built for each species using SpectraST (version 4.0), which was adapted to build spectral libraries from unidentified spectra (Fig. 1). Spectra that are highly similar and therefore likely derived from identical peptides were detected by similarity clustering and merged into higher-quality consensus spectra (Supplementary Methods). Unlike conventional spectral libraries, similarity clustering is necessary for consensus spectrum building when peptide identifications are not known. Collectively, the spectral libraries contained 9,045 consensus spectra, 1,039 (11.5%) of which were detected in multiple species. The distribution and relative abundance of spectra among vertebrate species are visually represented in Fig. 2. Spectra from the blood of *M. musculus*, the only species for which a genome sequence

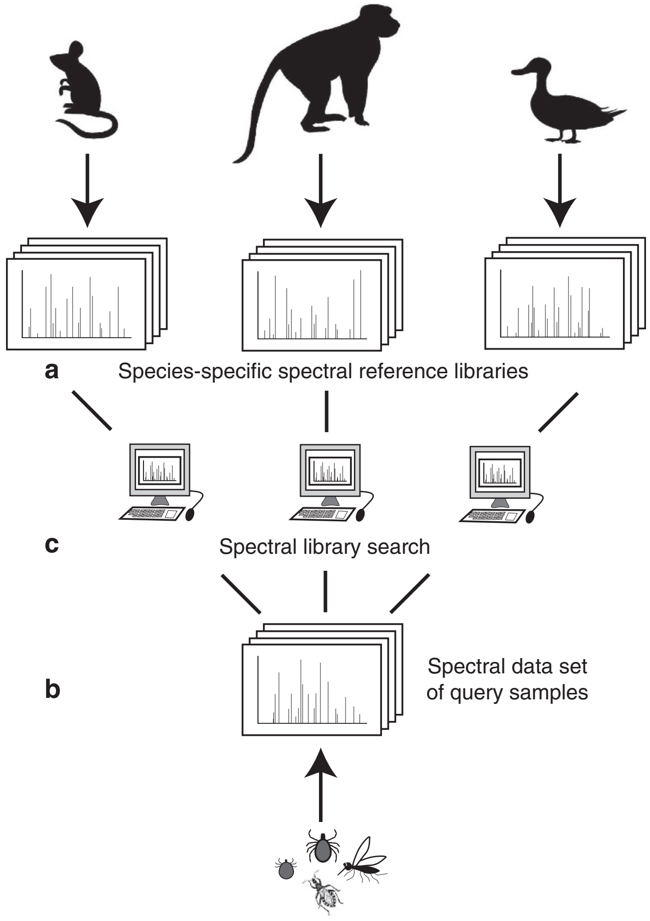


Figure 1 | A schematic overview of blood meal identification using spectral proteomes. (**a**) Species-specific spectral reference libraries are built from LC-MS^2 analyses of trypsin-digested whole blood from potential host vertebrates. Spectra do not require peptide identifications as long as they are reproducible. (**b**) Collection of query LC-MS^2 spectral data sets obtained from vectors days to months after blood feeding. Query spectral data sets include the vector proteome and the proteome of the remnants of blood from the previous blood meal. (**c**) Similarities between the spectral data set and each spectral reference library are determined using a Spectral Data set Similarity Scoring (SDSS) algorithm. The SDSS algorithm provides a quantitative similarity measure between sets of spectra, and hence the proteomes underlying the data sets, using the number of query spectra that matched each reference library spectrum.

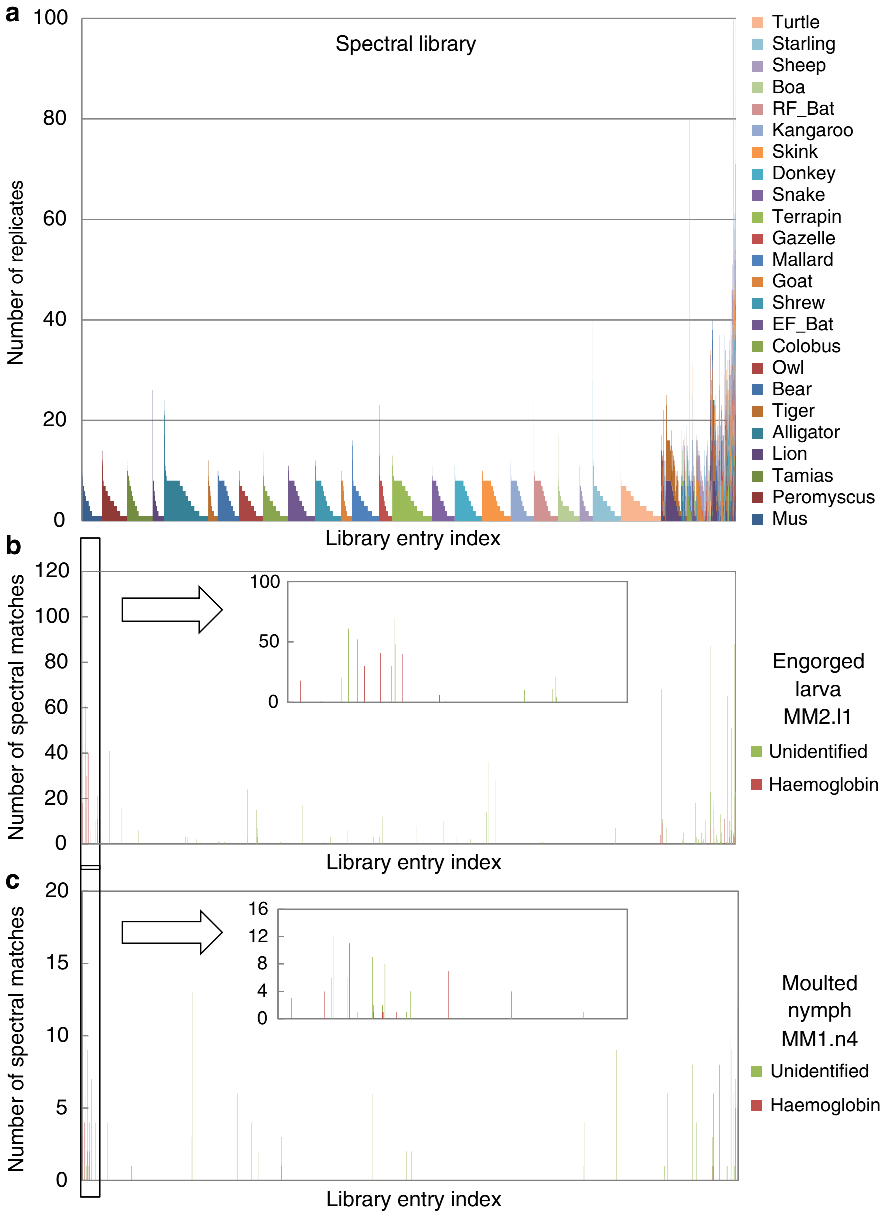


Figure 2 | Reference spectral libraries of the blood proteomes of 24 vertebrate species. (**a**) The diversity and abundance of unique and shared spectra in the spectral reference libraries visualized as a bar chart. Each bar along the abscissa represents a distinct consensus spectrum, and the height of the bar represents the number of replicates merged into that consensus. Bars are color-coded to represent the species from which the replicates originate. The multi-colored bars to the right represent spectra that are observed in multiple species. Example of spectral data set matches for (**b**) one larva engorged with *M. musculus* (MM2.l1) or (**c**) one moulted nymph held at 25 °C for 6 months after feeding on *M. musculus* (MM1.n4). The majority of spectral matches for the tick data sets are to library spectra from the correct species (*M. musculus*), or to spectra from multiple species including *M. musculus*. The matched spectra are subjected to sequence searching against haemoglobin sequences. Although some spectral matches are due to haemoglobin (red), many other matches cannot be identified by sequence searching (green).

database is available and peptide identifications could be made reliably, were identified by conventional sequence searching to reveal the underlying protein distribution of the spectral libraries (Fig. 3).

Accuracy of clustering algorithm. The performance of the clustering algorithm was validated by comparing a spectral library built by similarity clustering without knowledge of peptide identifications, as described above, and a spectral library built from sequence-search results of the same data. The latter is possible for species with comprehensive and high-quality sequence databases such as *M. musculus*. We implemented this strategy on a larger data set containing 40 runs of a SDS–PAGE-fractionated protein digest sample of *M. musculus* blood. In this data set, 20,156 of the 38,010 acquired tandem mass spectra survived the spectrum filter and were similarity-clustered, as described above, into 3,584 clusters. At the same time, traditional sequence searching (Supplementary Methods) identified 10,324 spectra from the same data set to a total of 1,566 distinct identifications (44% of 3,584). Therefore, less than half of all clusters were identified by sequence searching, likely because of incomplete

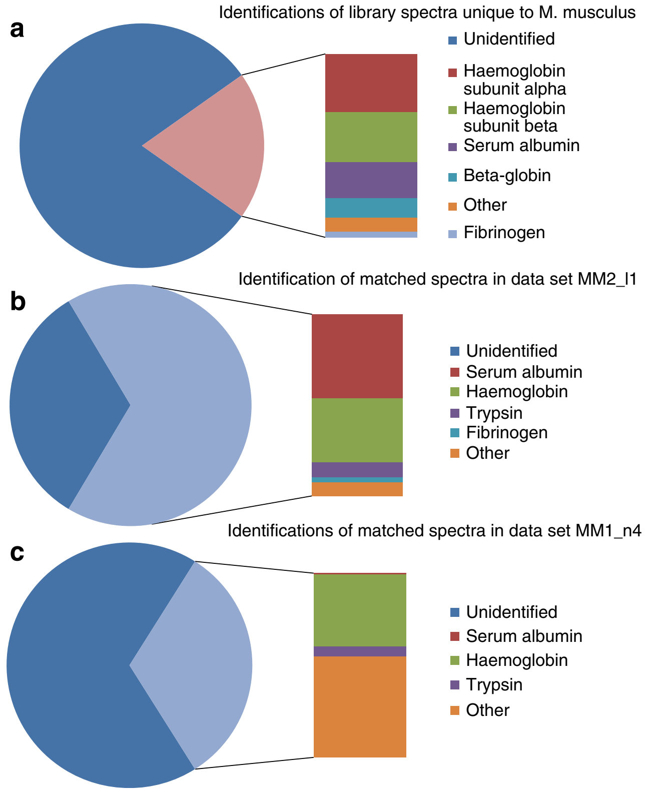


Figure 3 | Underlying protein distribution of the blood proteome of *M. musculus*. Peptide identifications of spectra used to build or search the *M. musculus* library by sequence searching. (**a**) The protein distribution of the identifications of the library spectra unique to *M. musculus*. (**b**,**c**) The protein distributions of identifications of spectra matched to the *M. musculus* library for two data sets: MM1.n4 (a moulted nymphal tick) and MM2.l1 (an engorged larva), respectively. A large fraction of spectra remain unidentified by sequence searching, despite the comprehensiveness of the *M. musculus* protein sequence database. Among identified spectra, subunits α and β of haemoglobin, serum albumin, β-globin and fibrinogen are the most frequent protein identifications found.

search space (for example, due to modifications not considered) and insufficient discriminating power of sequence searching. Of the clusters with at least one originating replicate identified, 1,492 entries (95%) are 'pure' clusters with only one sequence identification, 29 (2%) entries are 'mixed' ones with two or more different identifications, 35 (2%) entries are 'split' clusters, for which the same identification was also found in another cluster, and 10 entries (1%) are both 'mixed' and 'split' (Fig. 4). This result suggests that the clustering algorithm is largely capable of merging spectra that originate from the same peptide, but not those that originate from different peptides.

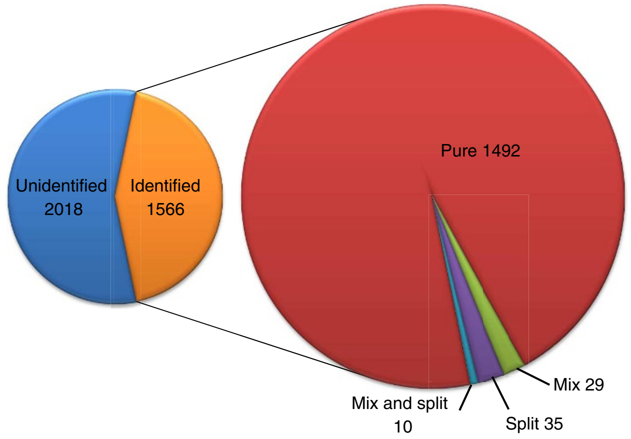


Figure 4 | Correspondence between spectrum clusters and sequence-search results. The clustering algorithm is validated using a much larger *M. musculus* fractionated blood data set. Among the 1,566 clusters that have at least one spectrum identified by SEQUEST, an overwhelming number (1492/1566 = 95%) of them are 'pure' clusters that have a single and unique peptide identification in *M. musculus* not found in any other cluster.

Table 1 | The species on which an engorged larva (*I. scapularis*) previously parasitized can be confidently determined through proteome profiling.

		Proteome profile (spectral library-based)		
Species	**Animal_ID. Tick_ID**	**Best spectral library match**	**Bootstrap support**	**Success rate**
Lab mouse	MM1.l1	Lab mouse	> 99%	5/5
	MM1.l2	Lab mouse	> 99%	
	MM1.l3	Lab mouse	> 99%	
	MM1.l4	Lab mouse	> 99%	
	MM2.l1	Lab mouse	> 99%	
White-footed mouse	PL1.l1	White-footed mouse	> 99%	4/4
	PL2.l1	White-footed mouse	> 99%	
	PL3.l1	White-footed mouse	> 99%	
	PL4.l1	White-footed mouse	> 99%	
Chipmunk	TS1.l1	Chipmunk	> 99%	3/3
	TS2.l1	Chipmunk	> 99%	
	TS3.l1	Chipmunk	> 99%	
Squirrel	SC1.l1	White-footed mouse	69.8%	2/3*
	SC1.l2	Chipmunk	> 99%	
	SC1.l3	Chipmunk	> 93%	
Total				14/15

**S. carolinensis* squirrel was not included in the reference library. The identification was regarded as successful if the best match for the spectral data sets from ticks that fed on squirrels was *T. striatus* chipmunks, evolutionarily the most closely related species (see Fig. 6).

Identification of blood meal source in engorged larvae. The correct vertebrate species was identified as the source of the blood meal in all but one engorged larvae tested, demonstrating proof-of-principle evidence of the efficacy of this methodology (Table 1). The identification is made by finding the reference library with the highest similarity to the spectral data set of the blood meal to be tested. A spectral data set similarity score (SDSS) developed for this application provides a quantitative similarity measure between two spectral data sets, and hence the entire proteomes underlying the data sets. Using this algorithm, the blood in all five of the larval ticks that had fed to repletion on either of a two *M. musculus* mice was determined to be most similar to the *M. musculus* spectral reference library. These results had high bootstrap support (> 99%), indicating that a large number and wide variety of spectra are available to correctly identify the blood in these ticks as *M. musculus* blood (Table 1). An example spectral data set match is shown in Fig. 2. These data also demonstrate that the presence of vector proteins in the sample do not detract from the ability of this methodology to identify the source of the blood meal. Correct identifications were also achieved for all larval ticks that had fed on any of the four *Peromyscus leucopus* mice or any of the three *Tamias striatus* chipmunks, all with high bootstrap support (> 99%) (Table 1).

Interestingly, the blood in the larval ticks that had fed on two of the three *Sciurus carolinensis* squirrels was not identified as squirrel blood, as there was no squirrel blood spectral reference library, but instead matched most closely with the chipmunk spectral reference library (Table 1), the species that is evolutionarily most closely related to *S. carolinensis* among the reference libraries. Thus, this method can also identify an unknown sample to the closest-related species represented in the spectral libraries. This finding is corroborated by the evolutionary accuracy of the phylogenies built from the spectral reference libraries discussed below.

A recent publication reported the protein sequences of haemoglobin subunits of several vertebrate species by *de novo* sequencing and suggested that these sequences can be used to identify the source of blood meals in vectors[27]. To benchmark the methodology presented here, we also searched the proteomic data from our ticks against the haemoglobin sequences of *M. musculus*, *P. leucopus* and *T. striatus* by sequence database searching. Identification of the source of the blood meal from engorged larval ticks was also effective using methods based on detecting haemoglobins by sequence searching algorithms (Supplementary Table S1). Similar to the results by our method, larval ticks that had fed upon *S. carolinensis* could also be identified as ticks that had fed upon *T. striatus* if *S. carolinensis* haemoglobin sequences are not included in the database.

Identification of the blood meal in moulted nymphal ticks. In 23 of the 26 cases, the spectral data sets derived from the remnants of blood in moulted nymphal ticks were identified to the spectral reference libraries of the species on which the tick had fed as a larva (Table 2, Fig. 2). Of the ticks that fed on *M. musculus,* the *M. musculus* spectral library was the most similar for all of the nymphs that were held at 25 °C for 1 month (6/6) and for 3 months (6/6) after blood meal, while the *M. musculus* spectral library was the most similar for the data sets from 4 of 6 nymphs held at 25 °C for 6 months before analyses. The statistical confidence in *M. musculus* as the source of the previous blood meal was >76% (bootstrap support) for each of the correctly classified nymphs. The two data sets that were not identified to the *M. musculus* were not misidentified to another organism; they simply had very low similarity scores with all spectral libraries, suggesting that little MS^2 data from the blood proteome was collected. More importantly, the spectral data sets from seven of eight nymphal ticks that had fed as larvae on natural reservoir hosts (*P. leucopus* (5/6) and *T. striatus* (2/2)) and subsequently held at 25 °C for 5 months were successfully identified to the animal on which the ticks had previously fed, each with high bootstrap support (Table 2). These data strongly support the utility of the spectral matching methodology to identify the source of previous blood meals in nymphal ticks that are collected in nature.

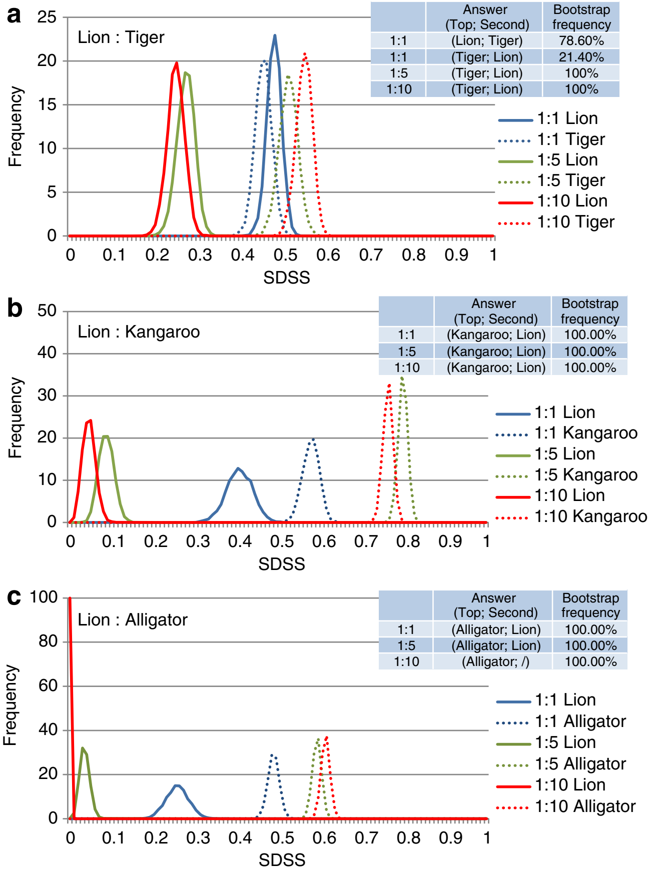
Identification of the source of the blood meal from moulted nymphal ticks was less effective when peptide identification relied upon *de novo* sequenced haemoglobin subunits (Supplementary Table S2). There were no instances in which the traditional sequence-based peptide identification methodology correctly identified the source of a blood meal that could not be identified using the spectral searching algorithm. Furthermore, the sequence-based method failed to identify the source of the blood meal for two ticks that fed on lab mouse, 4 out of 6 ticks that had fed on a white-footed mouse, and both ticks that had fed on a chipmunk. In these tick samples, no confident identification to any haemoglobin sequences were found. These results indicate that relying on *a priori*-selected target proteins (haemoglobin-α and -β) for identification was more prone to failure than the spectral matching algorithm described when the blood meal has undergone digestion over time. This is likely because the target proteins may have been degraded, whereas our proteome profiling strategy is still able to identify the source of the blood meal based on spectra from other, non-target, proteins. Although haemoglobin is the most abundant protein in vertebrate blood, only a small fraction of the matching spectra utilized by our method can be identified to haemoglobin sequences (Figs 2 and 3).

Table 2 | The species on which moulted nymphs (*I. scapularis*) parasitized during their larval stage can be confidently determined.

			Proteome profile (spectral library-based)		
Species	Tick age	Animal_ID. Tick_ID	Best spectral library match	Bootstrap support	Success rate
Lab mouse	1 Month	MM1.n1	Lab mouse	>99%	6/6
		MM1.n2	Lab mouse	>99%	
		MM1.n3	Lab mouse	>99%	
		MM2.n1	Lab mouse	>99%	
		MM2.n2	Lab mouse	>99%	
		MM2.n3	Lab mouse	>99%	
	3 Months	MM1.n4	Lab mouse	>99%	6/6
		MM1.n5	Lab mouse	>99%	
		MM1.n6	Lab mouse	>99%	
		MM2.n4	Lab mouse	>99%	
		MM2.n5	Lab mouse	90.7%	
		MM2.n6	Lab mouse	>99%	
	6 Months	MM1.n7	Lab mouse	77.0%	4/6
		MM1.n8	Lab mouse	94.3%	
		MM1.n9	Lab mouse	>99%	
		MM2.n7	Lab mouse	>99%	
		MM2.n8	No match*	X	
		MM2.n9	No match*	X	
White-footed mouse	5 Months	PL5.n1	WF mouse	>99%	5/6
		PL6.n1	WF mouse	96.1%	
		PL6.n2	WF mouse	76.3%	
		PL7.n1	WF mouse	>99%	
		PL7.n2	No match*	X	
		PL7.n3	WF mouse	98.5%	
Chipmunk	5 Months	CM4.n1	Chipmunk	97.7%	2/2
		CM5.n1	Chipmunk	>99%	
Total					23/26

Abbreviation: WF, white-footed.
*No matches were found to the reference library spectra.

Multiple blood sources. Although *I. scapularis* ticks feed only once per life stage, other arthropod vectors feed multiple times in rapid succession, which could hinder the identification of one or more sources of the blood in the vector. The ability to correctly identify the source of blood from vectors that had fed multiple times using the spectral matching methodology presented here was determined by experimentally mixing whole blood from different species at specified ratios. Whole blood from an African lion (*Panthera leo*) was mixed with blood from either an Amur tiger (*Panthera tigris*), an American alligator (*Alligator mississippiensis*) or a red kangaroo (*Macropus rufus*) at lion: other ratios of 1:1, 1:5 or 1:10. The choice of species was determined by the evolutionary relatedness ranging from very related to distantly related. Proteins from mixed blood samples were analysed by liquid chromatography - tandem mass spectrometry (LC-MS^2) and the SDSS algorithm as described above. The spectral matching methodology can successfully identify one of the two true species in all cases (with high bootstrap support), even if blood from both species is roughly at equal concentrations (Fig. 5). In addition, when one source of blood is at much greater concentrations than the other (ratios of 1:5 and 1:10), the more



	Answer (Top; Second)	Bootstrap frequency
1:1	(Lion; Tiger)	78.60%
1:1	(Tiger; Lion)	21.40%
1:5	(Tiger; Lion)	100%
1:10	(Tiger; Lion)	100%

	Answer (Top; Second)	Bootstrap frequency
1:1	(Kangaroo; Lion)	100.00%
1:5	(Kangaroo; Lion)	100.00%
1:10	(Kangaroo; Lion)	100.00%

	Answer (Top; Second)	Bootstrap frequency
1:1	(Alligator; Lion)	100.00%
1:5	(Alligator; Lion)	100.00%
1:10	(Alligator; /)	100.00%

Figure 5 | Identifying mixed blood samples. The score histograms of matches to spectral libraries of lion (solid lines) and the other animal (dashed lines) are shown for three different mixing ratios (lion: other)—blue 1:1 green 1:5 and red 1:10 for (**a**) mix of lion and tiger blood, (**b**) mix of lion and kangaroo blood and (**c**) mix of lion and alligator blood, respectively. These mixed blood data sets matched the spectral libraries of all other spectral libraries with very low scores (< 0.01). The species with the highest score and the second highest scores, together with the bootstrap frequencies of the highest-scoring identification, are shown in the inset tables.

concentrated source is identified as the top-scoring hit in all cases, while the SDSS for the less-abundant source varies. It is likely that blood derived from one of the multiple feedings will be predominant in a vector because of digestion between feedings or differences in the amount of blood acquired from each host. Under this scenario, this source of blood is likely to be identified by this method, despite contamination with other blood sources. Although the source of the blood making up the minor fraction can potentially be identified when it is at relatively high concentrations, a more sophisticated statistical method, as well as experimental samples that have fed on multiple host species, is needed to properly assess the validity of this method in identifying blood mixtures. Nevertheless, one of the true sources of the blood in a vector that has taken two blood meals in rapid succession can be identified with high confidence.

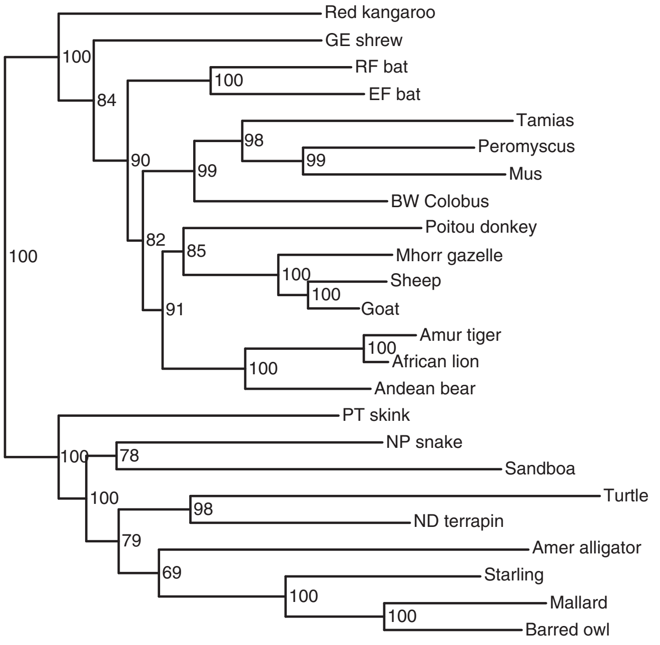


Figure 6 | Evolutionary relationships are accurately resolved from blood proteomes. Bayesian phylogenies constructed in MrBayes using the presence or absence of each spectrum in each reference library in the *variable* evolutionary model designed for restriction digest data resulted in a phylogeny that recapitulates the established relationships among these species.

Blood proteome divergence is evolutionary conserved. Bayesian phylogenies derived from the presence or absence of each spectrum in each spectral reference library suggest that the degree of similarity in spectral libraries derived from blood proteomes is correlated with evolutionary relatedness. Bayesian phylogenies constructed using evolutionary models designed for restriction digest data resulted in topologies that recapitulate the known relationships among the vertebrate species (Fig. 6). All nodes in the phylogeny constructed using the *variable* evolutionary model in MrBayes were supported by the data. The choice of evolutionary model for phylogenetic reconstruction did not affect the relationships among species, although branch lengths differed and posterior support for nodes was generally lower (data not shown).

Evolutionary resolution of the spectral matching methodology. The level of biological organization (species, population or individual) that can be distinguished using the proposed methodology was investigated using spectral data derived from whole blood. To allow for separate reference and testing data sets, we used a single LC-MS^2 run to build the library and used a subsequent run from the same blood sample to search the library. In addition, for two of the species (lion and tiger), we obtained whole blood from two individuals, only one of which was used to build the database. The SDSS values were calculated for each pair of samples in the test data set and spectral libraries. The correct species most closely matched the test sample in all cases, generally with SDSS scores many fold greater than the next closest match (Table 3). Even the evolutionarily most closely related pair of species in our data set (lion and tiger) were accurately identified in all cases, including samples from individuals that were not included in the database. Like all methods that depend on evolutionary relatedness, we expect to reach a limit in resolution. However, this resolution appears to be at a finer scale than between species as closely related as lions and tigers.

Discussion

Public health efforts to control the distribution and abundance of pathogens and their vectors in nature require an accurate

Table 3 | Evolutionary resolution.

	Lion2	Alligator	Tiger2	Bear	Owl	Colobus	EF_Bat	Shrew	Goat	Mallard	Gazelle	Mus	Terrapin	Snake	Pero	Donkey	Skink	Kangaroo	RF Bat	Boa	Sheep	Starling	Tamias	Turtle
Lion2	0.74	0.00	0.14	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lion1	0.42	0.00	0.18	0.00	0.03	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00
Alligator	0.00	0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tiger2	0.12	0.00	0.76	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tiger1	0.05	0.00	0.33	0.01	0.02	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02	0.01	0.01	0.00	0.00
Bear	0.01	0.00	0.01	0.92	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Owl	0.00	0.00	0.00	0.01	0.69	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Colobus	0.00	0.00	0.00	0.00	0.00	0.65	0.00	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
EF_Bat	0.00	0.00	0.00	0.00	0.00	0.00	0.89	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Shrew	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.89	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Goat	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.76	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.10	0.00	0.00	0.00
Mallard	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.65	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Gazelle	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
Mus	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.70	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Terrapin	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Snake	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pero	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Donkey	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.74	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Skink	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Kangaroo	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.89	0.00	0.00	0.00	0.00	0.00	0.00
RF_Bat	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.91	0.00	0.00	0.00	0.00	0.00
Boa	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.73	0.00	0.01	0.00	0.00
Sheep	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.10	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.55	0.00	0.00	0.00
Starling	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.72	0.00	0.00
Tamias	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.56	0.00
Turtle	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.64

Data sets from pure blood samples of all animals, including two individuals of the lion and the tiger species (rows), are searched against the libraries of all animals (columns). The SDSS values between each pair of data set and library are shown, with the top score for each data set shown in red.

assessment of the factors that affect the population dynamics of these species. For example, control strategies that have focused on interrupting the transmission of *B. burgdorferi* in *P. leucopus* mice have resulted in only modest success in reducing infection prevalence in *I. scapularis*, and thus minimal reductions in human Lyme disease risk[41–44]. A likely explanation for the weak effects of mouse-targeted intervention strategies is the existence of important alternative host species[5,16,45]. Methods that rapidly and reliably identify the vertebrate species on which a disease vector previously parasitized can be used to assess the importance of each vertebrate species to the maintenance and proliferation of vector populations. Public health programs can use this knowledge of the natural ecology of the parasite to design intervention strategies that target only the appropriate host species. The methodology presented here is unlikely to provide information concerning which intervention strategies are likely to be effective. However, in multi-host parasite systems where host-targeted interventions are least likely to be practical, knowing that interventions need to target only a limited set of all potential vertebrate species increases the practicality of these public health programs.

In this study, we present a non-targeted, genome-free, mass spectral searching algorithm to determine the vertebrate species from which a disease vector took their last blood meal. Using this method, the remnants of blood in blacklegged ticks were detected and correctly classified to the vertebrate species from which the tick acquired their previous blood meal in 37 out of 41 cases (Tables 1 and 2). Most importantly, heavily degraded remnants of the larval blood meal in nymphal ticks that were active for 5 months or more remained detectable and were correctly classified to the vertebrate species from which the tick acquired the blood in 11 out of 14 cases (Table 2). These data provide proof-of-principle validation of this method in determining the relative importance of vertebrate species as hosts for vector populations in nature.

The presented method compares the entire tandem mass spectral data set derived from individual vectors to spectral reference libraries derived from the whole blood of individual species to determine the source of the blood remaining in the vector. The SDSS provides a quantitative similarity measure between two sets of spectra. Although *I. scapularis* ticks feed only once per life stage, empirical data suggest that the spectral matching methodology can potentially identify the sources of blood in vectors that have fed multiple times in rapid succession (Fig. 5). The SDSS is also evolutionarily accurate such that the evolutionary clade of the species from which a blood meal was acquired can be identified, even if the correct species is not in the reference database. For example, the absence of an *S. carolinensis* spectral reference library did not prevent the identification of the source of blood in engorged larval ticks that had fed on *S. carolinensis* squirrels to the correct evolutionary clade, sciurid rodents, which was represented in the spectral libraries by *T. striatus*. (Table 1). Thus, an unknown blood source can be classified to the most evolutionarily related species represented in the reference library if the true species is absent from the reference library.

The evolutionary accuracy of our spectra matching methodology is perhaps not surprising, given that the spectral reference libraries can be used to accurately reconstruct the evolutionary history of vertebrate hosts (Fig. 6). This finding is practically useful as it indicates that the spectral reference libraries need not necessarily contain a complete set of potential host species to identify the evolutionary clade from which a blood meal was taken. For example, there is only one marsupial in Pennsylvania, the Virginia Opossum; blood meals acquired in Pennsylvania that match to the library of any marsupial can still be identified to the Virginia Opossum by inference. It is important to note that the specificity of this methodology—the level of biological organization (order, genus, species and population) to which one can confidently identify the source of a blood meal—is dependent on the species richness of the library. Thus, adding more species to the database should permit the determination of the vertebrate source of a blood meal to a finer evolutionary resolution. We expect that our method, like any method that depends on the evolutionary changes in protein sequences, will eventually reach a limit in resolution. However, this resolution appears to be at a finer scale than the most closely related species in this study, the lion and the tiger (Table 3).

The spectral data set similarity algorithm correctly identified the source of the blood in 23 of the 26 moulted nymphs that had fed on either *M. musculus, P. leucopus* or *T. striatus* (Table 2), even when analysed up to 6 months after moulting. The efficacy of this method may be further improved by better proteome coverage, which can be readily obtained using state-of-the-art mass spectrometers. In this study, each spectral reference library was built with two replicate LC-MS^2 runs of unfractionated blood using a quadrupole ion trap mass spectrometer, leaving considerable room to improve proteome coverage. One should also note that digestion of the blood meal is likely much faster in these experimental ticks than in natural conditions as nymphs are inactive at very low winter temperatures after their larval blood meal. As environmental conditions affect digestion in ectothermic animals[46,47], the detection limit of the blood proteome from the nymphal vector may be extended further under natural field conditions. Future applications of this work to field-collected ticks with a broader range of potential host species will help to further evaluate the efficacy of this methodology under natural environmental conditions.

This spectral library-based approach has several distinct theoretical advantages over traditional genome-obligated proteomic data analysis methods. On the one hand, genome sequence data are not available from the vast majority of animals that are potential sources of blood meals for haematophagous parasites, which limits the potential animal species that can be included in any study using traditional methods. On the other hand, a spectral library is essentially a record of the MS-detectable proteome, including all post-translational modifications, information that is not readily deducible from the genome sequences. Experimentally observed spectra also contain information on the true intensity of fragment ions, presence of neutral-loss fragments and other unexpected but reproducible minor ions, none of which can be predicted from sequences, leading to more accurate spectral matching. In fact, the spectral matching methodology is shown to perform better than a recently proposed sequence-based methodology that utilizes sequences of dominant blood proteins (haemoglobin subunits α and β), especially in identification of the source of blood from moulted nymphal ticks (Table 2 and Supplementary Table S2). This is likely because the spectral matching methodology performs an unbiased profiling that accounts for any molecules that can be ionized and fragmented in the mass spectrometer, as opposed to restricting detection to an *a priori*-selected set of target peptides. Proteomic profiling methodologies should also have more discriminating power and better resolution among species, as closely related species may not be distinguishable based on the sequences of a few peptides. Moreover, experimental extraction and *de novo* sequencing of specific proteins is also a painstaking and error-prone process.

The spectral library-based approach also has several practical advantages for blood meal analysis. First, building blood spectral libraries is much cheaper and faster than genome sequencing and is accessible to researchers without direct access to such technologies (Supplementary Table S3). Second, the proposed methodology requires neither the identity nor the function of detected proteins for taxonomic identification as long as the recorded mass signal is reproducible and consistent, similar to fingerprinting methodologies using DNA (ref. 48). Third, it is not necessary to remove tick protein contaminants during sample preparation, as this method is robust against impurities. This is because the SDSS as defined is unaffected by unmatched spectra. Lastly, spectra are stored in their entirety and can be used in future classical proteomic studies. Hence, building spectral libraries derived from blood proteomes is a promising strategy that may be applicable to identify the vertebrate source of blood meals in the multitude of vector-borne disease systems.

Although our methodology was tested in a tick vector, it may be similarly successful in many other haematophagous parasites, thereby expanding our understanding of the natural life cycle of important human pathogens and aiding in the control of vectors and vector-borne diseases. More generally, this approach can potentially be used to 'fingerprint' any biologically complex samples for use in medical, public health and research applications, either independently or as a complementary method to DNA-based approaches. Potential future applications include identifying tissue or cell types in clinical settings, detecting the presence and abundance of microorganisms in environmental samples, and elucidating the constituents in the human gut microbiome.

Methods

Software development. SpectraST (version 4.0), a tandem mass spectral library building and searching program originally developed for peptide identification [39,49], was extended to incorporate unidentified spectra. Specifically, spectra imported into a library can be referenced by serial identification numbers without explicit peptide sequence identification. A consensus algorithm based solely on spectral similarity was applied to combine replicate MS^2 spectra acquired multiple times due to fragmentation of the same molecule into a consensus spectrum. This procedure is common in traditional spectral library construction and produces higher-quality spectra[50–53]. The numbers of spectra combined into each consensus spectrum, along with the source of the sample, were recorded. Spectra with limited information, which may cause rampant indiscriminate matching, were removed by spectral filtering. For algorithm details, please refer to Supplementary Methods. SpectraST is part of the Trans-Proteomic Pipeline suite of software, an open-source and continuously maintained tool that supports all major computer platforms and mass spectrometers[54].

Spectral libraries of whole blood. Spectral reference libraries were built from LC-MS^2 analyses of whole blood from each of 24 animal species: *M. musculus* (lab mouse), *P. leucopus* (white-footed mouse), *T. striatus* (eastern chipmunk), *Ovis aries* (domestic sheep), *P. leo* (African lion), *P. tigris* (Amur tiger), *Strix varia* (barred owl), *Anas platyrhynchos* (mallard), *Sturnus vulgaris* (starling), *Colobus guereza* (black and white Colobus), *Chelodina longicollis* (snake-necked turtle), *Colubrinus loveridgei* (Kenyan sand boa), *Rousettus aegyptiacus* (Egyptian fruit bat), *Pteropus rodricensis* (Rodrigues Fruit bat), *Tremarctos ornatus* (Andean bear), *Equus africanus asinus* (Poitou donkey), *Nanger dama mhorr* (Mhorr's gazelle), *Capra aegagrus hircus* (Goat), *M. rufus* (red kangaroo), *Corucia zebrata* (prehensile-tailed skink), *A. mississippiensis* (American alligator), *Pituophis melanoleucus* (northern pine snake), *Malaclemys terrapin* (Northern diamondback terrapin) and *Rhynchocyon petersi* (giant elephant shrew). These species are representative of the major clades of land animals from which blood meals may be taken by haematophagous arthropods. Animal blood was generously donated by veterinarians at the Philadelphia Zoo or acquired from laboratory animals at the University of Pennsylvania following all animal care and use guidelines. Whole-blood samples were trypsin-digested in solution, and the resulting peptide mixtures were run on an analytical C_{18} nanocapilary HPLC column (Acclaim PepMap100; DIONEX) into a LCQ-Deca XP Plus quadrupole ion trap mass spectrometer [55–57]. Duplicate runs were performed, and spectra acquired from each species were used to build spectral reference libraries as described above (Fig. 1) and in Supplementary Methods.

***I. scapularis* tick samples**. Blood-fed larval ticks (*I. scapularis*) were acquired by allowing ticks to feed to repletion on a host animal and detach naturally[16]. The larval blood meal represents the first blood meal as *I. scapularis* ticks feed only once at each of the three life stages (larva, nymph and adult). We experimentally infested two *M. musculus* mice with larval ticks purchased from Oklahoma State University. Larvae that fed to repletion and naturally detached from mice were collected, either frozen immediately ($n = 5$) or held at 25 °C and allowed to moult to the nymphal stage. Nymphal ticks were frozen at 1 month ($n = 6$), 3 months ($n = 6$) or 6 months ($n = 6$) post moult for later processing and blood meal analysis. Engorged larvae that had fed upon wild *P. leucopus, T. striatus* or *S. carolinesis* were acquired as previously described[15]. Six engorged larvae from three different *P. leucopus* mice and two engorged larvae from two different *T. straitus* chipmunks were allowed to moult and held at 25 °C for 5 months post moult before being frozen for processing and blood meal analysis.

Blood meal proteome analysis. Whole ticks were manually crushed in 20 μl of PBS, insoluble materials removed by centrifugation at 25,000 *g* at room temperature, and total protein concentrations of supernatants determined by the Bradford method (Bio-Rad). Protein extracts from engorged larvae, which are composed primarily of host blood, were run on a one-dimensional NuPAGE 4–20% Bis-Tris gel (Invitrogen, Switzerland) and proteins visualized by Coomassie Blue Staining (Bio-Rad). Each lane was cut into 20 equal-width bands, which were subjected to in-gel trypsin digestion and LC-MS^2 (refs 55–57). Total protein extracts from nymphal ticks were trypsin-digested in solution before LC-MS^2 analysis without prior gel separation.

Spectral data set similarity. Data obtained from LC-MS^2 analyses of tick samples were searched against the spectral reference libraries by SpectraST (version 4.0)[49], using a precursor m/z tolerance of 1.5 Da e^{-1} and search parameters identical to those used for traditional spectral searching (Fig. 1). The number of query spectra that matched each reference library spectrum, where a match is defined as dot product >0.7 and dot bias <0.45, were counted. For each tick, the similarity between the query spectral data set and the reference library for each species was quantified by the SDSS, given by:

$$SDSS(q,s)=\frac{\sum_i\left[m_q(i)\times r_s(i)\right]}{\sqrt{\sum_i\left[m_q(i)\right]^2\sum_i\left[r_s(i)\right]^2}}$$

where *SDSS(q, s)* is the measure of similarity between the query data set *q* and the reference library *s*, *i* is an index that enumerates the spectra in the query data set, $m_q(i)$ is the number of spectra in the query data set *q* that matches the *i*th spectrum in the reference library *s* and $r_s(i)$ is the number of replicates that were combined into the *i*th spectrum of the reference library *s*. With the normalization factor of the denominator, this dot product-like function ranges from 0 (no spectral matches between the query data set and the reference library) to 1 (identical distribution of spectral counts between the query data set and the reference library). The abundance of each spectrum is taken into account to improve the discriminatory power of this function. Several alternative data set similarity functions were investigated and are presented in Supplementary Fig. S1.

The data were bootstrapped[58] to obtain a statistical measure of confidence in the identification of the source of the blood meal in each tick. For each query data set, 5,000 bootstrapped samples were generated by randomly selecting, *with replacement*, *N* MS^2 spectra from the data set, where *N* is the total number of MS^2 spectra in the query data set. Each bootstrapped sample was searched against each reference library, and the most similar reference library (greatest SDSS score) was recorded. The confidence in the species identified as the source of the previous blood meal is the fraction of bootstrapped samples for which the most similar reference library is derived from the correct species. This strategy quantifies the uncertainty in species identification that can occur because of the stochastic selection of peptides for fragmentation in the mass spectrometer.

Estimating evolutionary divergence from spectral libraries. The spectral libraries built from the whole-blood samples were used to infer phylogenetic relationships among species. The blood proteome data from each species was transformed into linear vectors representing the presence (1) or absence (0) of each library spectrum, resulting in data structures similar to restriction digest data. Bayesian phylogenies were constructed from these linear vectors in MrBayes[59]. Phylogenies were reconstructed in MrBayes using all five evolutionary models (all, noabsencesites, nopresencesites, variable and informative) originally designed for restriction digest data.

Data and software availability. All data used in this study, including those from LC-MS^2 analyses of the whole blood of 24 animals and of blood-fed *I. scapularis* samples, have been deposited at the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[60] under the accession code PXD000170. The software program SpectraST is part of the Trans-Proteomic Pipeline suite. The source code is available at http://www.sourceforge.net/sashimi/, and a one-click Windows installer is available at http://tools.proteomecenter.org/. Instructions and documentations are available at http://tools.proteomecenter.org/wiki/index.php?title=Software:SpectraST. The software is continually maintained and upgraded.

References

1. Keesing, F. *et al.* Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* **468,** 647–652 (2010).
2. Leibold, M. A. A graphical model of keystone predators in food webs: trophic regulation of abundance, incidence, and diversity patterns in communities. *Am. Nat.* **147,** 784–812 (1996).
3. Rose, M. D. & Polis, G. A. The distribution and abundance of coyotes: The effects of allochthonous food subsidies from the sea. *Ecology* **79,** 998–1007 (1998).
4. Johnson, P. T. J. & Thieltges, D. W. Diversity, decoys and the dilution effect: how ecological communities affect disease risk. *J. Exp. Bio.* **213,** 961–970 (2010).
5. LoGiudice, K., Ostfeld, R. S., Schmidt, K. A. & Keesing, F. The ecology of infectious disease: effects of host diversity and community composition on Lyme disease risk. *Proc. Natl Acad. Sci. USA* **100,** 567–571 (2003).
6. McLeod, R. S. Costs of major parasites to the Australian livestock industries. *Int. J. Parasitol.* **25,** 1363–1367 (1995).
7. Morgan, E. R. & Wall, R. Climate change and parasitic disease: farmer mitigation? *Trends Parasitol.* **25,** 308–313 (2009).
8. Taylor, L. H., Latham, S. M. & Woolhouse, M. E. Risk factors for human disease emergence. *Phil. Trans Roy. Soc. B* **356,** 983–989 (2001).
9. Ezenwa, V. O., Godsey, M. S., King, R. J. & Guptill, S. C. Avian diversity and West Nile virus: testing associations between biodiversity and infectious disease risk. *Proc. Biol. Sci.* **273,** 109–117 (2006).
10. Glass, G. E. *et al.* Environmental risk factors for Lyme disease identified with geographic information systems. *Am. J. Public Health* **85,** 944–948 (1995).
11. Khatchikian, C. E. *et al.* Geographical and environmental factors driving the increase in the Lyme disease vector, *Ixodes scapularis*. *Ecosphere* **3,** 85 (2012).
12. Martens, P. *et al.* Climate change and future populations at risk of malaria. *Global Environ. Change-Human and Policy Dimensions* **9,** S89–S107 (1999).
13. Goodman, J. L., Dennis, D. T. & Sonenshine, D. E. *Tick-borne diseases of humans* (ASM Press, 2005).
14. Brisson, D., Brinkley, C., Humphrey, P. T., Kemps, B. D. & Ostfeld, R. S. It takes a community to raise the prevalence of a zoonotic pathogen. *Interdiscip. Perspect. Infect. Dis.* **2011,** 741406 (2011).
15. Brisson, D. & Dykhuizen, D. E. ospC diversity in Borrelia burgdorferi: different hosts are different niches. *Genetics* **168,** 713–722 (2004).
16. Brisson, D., Dykhuizen, D. E. & Ostfeld, R. S. Conspicuous impacts of inconspicuous hosts on the Lyme disease epidemic. *Proc. Biol. Sci.* **275,** 227–235 (2008).
17. Kilpatrick, A. M., Daszak, P., Jones, M. J., Marra, P. P. & Kramer, L. D. Host heterogeneity dominates West Nile virus transmission. *Proc. Biol. Sci.* **273,** 2327–2333 (2006).
18. Li, W. *et al.* Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310,** 676–679 (2005).
19. Lord, R. D., Lord, V. R., Humphreys, J. G. & McLean, R. G. Distribution of Borrelia burgdorferi in host mice in Pennsylvania. *J. Clin. Microbiol.* **32,** 2501–2504 (1994).
20. Matuschka, F. R., Fischer, P., Musgrave, K., Richter, D. & Spielman, A. Hosts on which nymphal Ixodes ricinus most abundantly feed. *Am. J. Trop. Med. Hyg.* **44,** 100–107 (1991).
21. Allan, B. F., Goessling, L. S., Storch, G. A. & Thach, R. E. Blood meal analysis to identify reservoir hosts for Amblyomma americanum ticks. *Emerg. Infect. Dis.* **16,** 433–440 (2010).
22. Arnold, E. H., Simmons, S. W. & Fawcett, D. G. Precipitin technique for determining mosquito blood meals. *Public Health Rep.* **61,** 1244–1249 (1946).
23. Burkot, T. R., Goodman, W. G. & DeFoliart, G. R. Identification of mosquito blood meals by enzyme-linked immunosorbent assay. *Am. J. Trop. Med. Hyg.* **30,** 1336–1341 (1981).
24. Kent, R. J. Molecular methods for arthropod bloodmeal identification and applications to ecological and vector-borne disease studies. *Mol. Ecol. Resour.* **9,** 4–18 (2009).
25. Tempelis, C. H. & Rodrick, M. L. Passive hemagglutination inhibition technique for the identification of arthropod blood meals. *Am. J. Trop. Med. Hyg.* **21,** 238–245 (1972).
26. Wickramasekara, S., Bunikis, J., Wysocki, V. & Barbour, A. G. Identification of residual blood proteins in ticks by mass spectrometry proteomics. *Emerg. Infect. Dis.* **14,** 1273–1275 (2008).
27. Laskay, U. A. *et al.* Development of a host blood meal database: de novo sequencing of hemoglobin from nine small mammals using mass spectrometry. *Biol. Chem.* **393,** 195–201 (2012).
28. Gariepy, T. D., Lindsay, R., Ogden, N. & Gregory, T. R. Identifying the last supper: utility of the DNA barcode library for bloodmeal identification in ticks. *Mol. Ecol. Resour.* **12,** 646–652 (2012).
29. Humair, P.-F. *et al.* Molecular identification of bloodmeal source in Ixodes ricinus ticks using 12S rDNA as a genetic marker. *J. Med. Entomol.* **44,** 869–880 (2007).
30. Morán Cadenas, F. *et al.* Identification of host bloodmeal source and Borrelia burgdorferi sensu lato in field-collected Ixodes ricinus ticks in Chaumont (Switzerland). *J. Med. Entomol.* **44,** 1109–1117 (2007).
31. Mota, J. *et al.* Identification of blood meal source and infection with Trypanosoma cruzi of Chagas disease vectors using a multiplex cytochrome b polymerase chain reaction assay. *Vector Borne Zoonotic Dis.* **7,** 617–627 (2007).
32. Mukabana, W. R., Takken, W. & Knols, B. G. Analysis of arthropod bloodmeals using molecular genetic markers. *Trends Parasitol.* **18,** 505–509 (2002).
33. Pizarro, J. C. & Stevens, L. A new method for forensic DNA analysis of the blood meal in Chagas disease vectors demonstrated using Triatoma infestans from Chuquisaca, Bolivia. *PLoS ONE* **3,** e3585 (2008).
34. Thiemann, T. C. & Reisen, W. K. Evaluating sampling method bias in *Culex tarsalis* and *Culex quinquefasciatus* (diptera: culicidae) bloodmeal identification studies. *J. Med. Entomol.* **49,** 143–149 (2012).
35. Craig, R., Cortens, J. C., Fenyo, D. & Beavis, R. C. Using annotated peptide mass spectrum libraries for protein identification. *J. Proteome Res.* **5,** 1843–1849 (2006).
36. Dasari, S. *et al.* Pepitome: evaluating improved spectral library search for identification complementarity and quality assessment. *J. Proteome Res.* **11,** 1686–1695 (2012).

37. Frewen, B. E., Merrihew, G. E., Wu, C. C., Noble, W. S. & MacCoss, M. J. Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. *Anal. Chem.* **78,** 5678–5684 (2006).
38. Lam, H. & Aebersold, R. Building and searching tandem mass (MS/MS) spectral libraries for peptide identification in proteomics. *Methods* **54,** 424–431 (2011).
39. Lam, H. *et al.* Building consensus spectral libraries for peptide identification in proteomics. *Nat. Methods* **5,** 873–875 (2008).
40. Stein, S. E. & Scott, D. R. Optimization and testing of mass-spectral library search algorithms for compound identification. *J. Am. Soc. Mass Spec.* **5,** 859–866 (1994).
41. Deblinger, R. D. & Rimmer, D. W. Efficacy of a permethrin-based acaricide to reduce the abundance of *Ixodes dammini* (Acari: Ixodidae). *J. Med. Entomol.* **28,** 708–711 (1991).
42. Dolan, M. C. *et al.* Control of immature *Ixodes scapularis* (Acari: Ixodidae) on rodent reservoirs of Borrelia burgdorferi in a residential community of southeastern Connecticut. *J. Med. Entomol.* **41,** 1043–1054 (2004).
43. Tsao, J. I. *et al.* An ecological approach to preventing human infection: vaccinating wild mouse reservoirs intervenes in the Lyme disease cycle. *Proc. Natl Acad. Sci. USA* **101,** 18159–18164 (2004).
44. Hornbostel, V. L., Ostfeld, R. S. & Benjamin, M. A. Effectiveness of Metarhizium anisopliae (Deuteromycetes) against *Ixodes scapularis* (Acari: Ixodidae) engorging on *Peromnyscus leucopus*. *J. Vector Ecol.* **30,** 91–101 (2005).
45. Brisson, D. & Dykhuizen, D. E. A modest model explains the distribution and abundance of Borrelia burgdorferi strains. *Am. J. Trop. Med. Hyg.* **74,** 615–622 (2006).
46. Phillips, J. S., Adeyeye, O. & Bruni, D. Respiratory metabolism of the soft tick, *Ornithodoros turicata* (Duges). *Exp. Appl. Acarol.* **19,** 103–115 (1995).
47. Ribeiro, J. M. The midgut hemolysin of *Ixodes dammini* (Acari: Ixodidae). *J. Parasitol.* **74,** 532–537 (1988).
48. Cawood, A. H. DNA fingerprinting. *Clin. Chem.* **35,** 1832–1837 (1989).
49. Lam, H. *et al.* Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* **7,** 655–667 (2007).
50. Beer, I., Barnea, E., Ziv, T. & Admon, A. Improving large-scale proteomics by clustering of mass spectrometry data. *Proteomics* **4,** 950–960 (2004).
51. Frank, A. M. *et al.* Clustering millions of tandem mass spectra. *J. Proteome Res.* **7,** 113–122 (2008).
52. Frank, A. M. *et al.* Spectral archives: extending spectral libraries to analyze both identified and unidentified spectra. *Nat. Methods* **8,** 587–591 (2011).
53. Tabb, D. L., MacCoss, M. J., Wu, C. C., Anderson, S. D. & Yates, 3rd J. R. Similarity among tandem mass spectra from proteomic experiments: detection, significance, and utility. *Anal. Chem.* **75,** 2470–2477 (2003).
54. Deutsch, E. W. *et al.* A guided tour of the trans-proteomic pipeline. *Proteomics* **10,** 1150–1159 (2010).
55. Onder, O., Aygun-Sunar, S., Selamoglu, N. & Daldal, F. A glimpse into the proteome of phototrophic bacterium *Rhodobacter capsulatus*. *Adv. Exp. Med. Biol.* **675,** 179–209 (2010).
56. Onder, O., Turkarslan, S., Sun, D. & Daldal, F. Overproduction or absence of the periplasmic protease DegP severely compromises bacterial growth in the absence of the dithiol: disulfide oxidoreductase DsbA. *Mol. Cell Proteomics* **7,** 875–890 (2008).
57. Onder, O. *et al.* Modifications of the lipoamide-containing mitochondrial subproteome in a yeast mutant defective in cysteine desulfurase. *Mol. Cell Proteomics* **5,** 1426–1436 (2006).
58. Efron, B. & Tibshirani, R. *An Introduction to the bootstrap* (Chapman & Hall, 1993).
59. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17,** 754–755 (2001).
60. Vizcaino, J. A. *et al.* The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41,** D1063–D1069 (2013).

Acknowledgements

This work was supported by a grant from the National Institutes of Health (AI076342), the Center for Diseases Control and Prevention (U01CK000170) and the University Grant Council, Hong Kong Special Administrative Region Government (HKUST RPC10EG08, DAG12EG01S). We would like to thank Dr Maarten Voordouw for providing ticks fed on *M. musculus* and *P. leucopus*, and Donna M. Ialeggio and the Philadelphia Zoo Veterinary staff for providing vertebrate blood samples, and the PRIDE team for help with data submission to ProteomeXchange.

Author contributions

H.L., D.B. and Ö.Ö. conceived the study and formulated the experimental and computational strategy. Ö.Ö. and B.K. performed all the experiments, including sample preparation and mass spectrometric analysis of vertebrate blood and *I. scapularis* samples. H.L. and W.S. adapted SpectraST to build spectral libraries from unidentified spectra. W.S. designed and validated the spectral matching algorithm, and performed all the data analysis, including the comparative study against a sequence-based alternative method. Ö.Ö., H.L. and D.B. co-wrote the manuscript, with input from all authors.

Additional information

Accession codes: Data from LC-MS/MS analyses of the whole blood of 24 animals and of blood-fed *I. scapularis* samples have been deposited at ProteomeXchange under the accession code PXD000170.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Önder, Ö. *et al.* Identifying sources of tick blood meals using unidentified tandem mass spectral libraries. *Nat. Commun.* 4:1746 doi: 10.1038/ncomms2730 (2013).