SPECIAL ISSUE REVIEW

Genotyping of *Bartonella* bacteria and their animal hosts: current status and perspectives

M. KOSOY[1]*†, C. MCKEE[1,2]†, L. ALBAYRAK[3] *and* Y. FOFANOV[3]

[1] *Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, Fort Collins, CO 80521, USA*
[2] *Department of Biology, Colorado State University, Fort Collins, CO 80523, USA*
[3] *Department of Pharmacology, University of Texas Medical Branch, Galveston, TX 77555, USA*

(*Received 8 May 2017; revised 7 June 2017; accepted 15 June 2017; first published online 2 August 2017*)

SUMMARY

Growing evidence demonstrates that bacterial species diversity is substantial, and many of these species are pathogenic in some contexts or hosts. At the same time, laboratories and museums have collected valuable animal tissue and ectoparasite samples that may contain substantial novel information on bacterial prevalence and diversity. However, the identification of bacterial species is challenging, partly due to the difficulty in culturing many microbes and the reliance on molecular data. Although the genomics revolution will surely add to our knowledge of bacterial systematics, these approaches are not accessible to all researchers and rely predominantly on cultured isolates. Thus, there is a need for comprehensive molecular analyses capable of accurately genotyping bacteria from animal tissues or ectoparasites using common methods that will facilitate large-scale comparisons of species diversity and prevalence. To illustrate the challenges of genotyping bacteria, we focus on the genus *Bartonella*, vector-borne bacteria common in mammals. We highlight the value and limitations of commonly used techniques for genotyping bartonellae and make recommendations for researchers interested in studying the diversity of these bacteria in various samples. Our recommendations could be applicable to many bacterial taxa (with some modifications) and could lead to a more complete understanding of bacterial species diversity.

Key words: barcoding, bacterial species, *Bartonella*, genomics, genotyping, parasite ecology, parasite–host association.

INTRODUCTION

Since the proposal of DNA barcoding by Hebert *et al*. (2003) as a new methodology for identification of biological species, it has been utilized on a wide variety of taxa for the purposes of identifying museum specimens, evaluation of population and community diversity, discovery of cryptic species, and other forensic applications. Barcoding was adapted for species-level identification by recovery of short DNA sequences from a specific genome fragment and has been applied widely in processing and identifying animal and plant tissues. Attempts have also been made to apply this barcoding paradigm to other eukaryote taxa and some microscopic organisms (Blaxter, 2016). For example, the nuclear ribosomal internal transcribed spacer (ITS) region has been recognized as a potentially universal DNA barcode marker for fungi (Bellemain *et al*. 2010; Schoch *et al*. 2012). Nevertheless, acceptance of DNA barcoding for the identification of organisms is controversial among some taxonomists because of fears that a universal, DNA-based approach in identification of species will replace traditional methods (Lebonah *et al*. 2014).

* Corresponding author: Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, Fort Collins, CO 80521, USA. E-mail: mck3@cdc.gov
† These authors contributed equally to this work.

Although bacterial DNA analysis has been readily accepted for measuring the assembly, diversity and distribution of entire microbial communities in different environments (DeLong and Pace, 2001), application of universal barcoding approaches to bacteria may face challenges unique to prokaryotes. First, many microbial species are challenging to culture by traditional methods (Schloss and Handelsman, 2004), which will curtail a microbiologist's ability to characterize these species beyond molecular approaches. Second, there are still discussions about the precise definition of bacterial species and the numerous, competing methods used to characterize them (Konstantinidis *et al*. 2006; Fraser *et al*. 2007). Finally, the phylogenetic diversity observed in microbes is far more complex than in eukaryotes (Hug *et al*. 2016), with predictions of global bacterial diversity ranging from 10^7 to 10^{12} species (Curtis *et al*. 2002; Schloss and Handelsman, 2004; Dykhuizen, 2005; Locey and Lennon, 2016). We argue that although our ability to measure bacterial diversity at a massive scale has increased in recent decades, the identification of bacterial species is complex and may not be amenable to one universal approach.

Meanwhile, microbiologists are faced with the dilemma of processing the large bank of specimens sitting in laboratory freezers and museum collections

Parasitology (2018), **145**, 543–562.
doi:10.1017/S0031182017001263

across the globe, containing uncharacterized bacterial species and potential pathogens. These specimens may include irreplaceable tissue samples from rare or hard to reach animal species, samples from human subjects or enormous collections of arthropod ectoparasites. How should microbiologists approach assessing the diversity of bacteria parasitizing these diverse animals in a consistent manner that will facilitate comparative epidemiological, evolutionary and ecological analyses? For the remainder of this review, we will focus on the diverse genus *Bartonella* (Alphaproteobacteria: Rhizobiales) to make specific recommendations for the genotyping of these bacteria that could be applicable to a wider array of bacterial taxa. Bartonellae are facultative, fastidious, intracellular bacteria commonly found in many taxonomically diverse mammalian species globally (Kosoy *et al*. 2012). Bartonellae are hypothesized to be transmitted (some possibly harboured) by a variety of arthropod vectors, including ticks, mites, lice, fleas, flies and other insects (Billeter *et al*. 2008; Tsai *et al*. 2011). Given the high prevalence and broad host range of *Bartonella* species, we expect a wide variety of *Bartonella* species and genotypes to be present in animal tissues collected during field investigations or archived in museum and laboratory collections. Furthermore, this genus exemplifies many of the challenges to characterizing bacterial species, including the limitations of 16S ribosomal RNA (rRNA) sequencing (Kosoy *et al*. 2012), challenges with culturing and frequent homologous recombination among genomic loci (Berglund *et al*. 2010; Chaloner *et al*. 2011; Paziewska *et al*. 2011, 2012; Guy *et al*. 2012; Buffet *et al*. 2013*a*; Bai *et al*. 2015*a*). Our analyses in this review will be structured as follows: (1) highlight the value of accurate bacterial genotyping for epidemiological and ecological research, (2) address the challenge of identifying a bacterial species in animal tissues using a short sequence of one or a few selected genomic fragments, (3) analyse the literature on molecular identification of bartonellae in different animal tissues and from diverse animal taxa, (4) compare genetic markers used for genotyping bartonellae with and without culturing and (5) evaluate the phylogenetic resolution of candidate loci based on analysis of genomes available for multiple *Bartonella* species. In the Discussion, we will make specific recommendations for consistent methods for genotyping bartonellae that will facilitate comparative studies.

NEED FOR ACCURATE GENOTYPING OF *BARTONELLA* SPECIES IN ANIMAL TISSUES

Importance of genotyping bartonellae in epidemiology and for biological threat preparation

Genotyping of pathogenic zoonotic bacteria can be a very important part of epidemiological investigations in an effort to define the source of human and animal diseases. The importance of tracing pathogenic bacteria in environment was highlighted by threats presented by select biological agents, particularly anthrax (Keim *et al*. 2008). In a more common situation, extensive databases containing sequences of multiple bacterial strains can provide irreplaceable information for a comparison of gene sequences between a presumptive human pathogen and potential zoonotic sources. As an example related to the study of bartonellae, a recent case of lymphadenopathy in Tbilisi, Georgia was linked to infected rats (Kandelaki *et al*. 2016). This was only possible because of the accurate genetic characterization of related *Bartonella* strains from commensal rats in Israel by Harrus *et al*. (2009). Similar connections have been made between human cases of myocarditis and meningitis in the USA and *Bartonella* genotypes found in ground squirrels (Kosoy *et al*. 2003; Osikowicz *et al*. 2016). Genotyping bartonellae and other infectious bacteria using consistent molecular approaches that generate a common repertoire of gene sequences will surely increase the feasibility and frequency of these comparisons.

Bartonellae as a popular tool for ecological studies

Bartonella species span the symbiont–pathogen continuum (Segers *et al*. 2017) and are an extremely diverse group of bacteria, especially in rodents and bats (Lei and Olival, 2014). Moreover, these vertebrate host–arthropod vector–*Bartonella* systems appear to be globally distributed and phylogenetically complex. Such features make these tripartite systems a popular tool for ecological comparative analyses (Buffet *et al*. 2013*b*; Klangthong *et al*. 2015; Brook *et al*. 2017). There are some ecological projects where identification of *Bartonella* is not essential and where a priority is given to estimation of *Bartonella* prevalence in animal populations without identification of the species (Bai *et al*. 2009; Young *et al*. 2014). However, most ecological and epidemiological studies require accurate identification of specific bacterial species and/or genotypes. The level of discrimination between obtained strains or genotypes depends on the objectives of the studies. In most situations, the investigators prefer to report bacteria at the species level or compare sequence identity with a specific *Bartonella* type strain. There is however a potential pitfall in reporting PCR-positive samples without sequencing of positive products. Some ecologists interested in using simple techniques for estimating prevalence of common animal infection may not be aware that the primers and real-time PCR probes selected for molecular detection of *Bartonella* DNA may not always be specific for the genus *Bartonella* (Maggi and Breitschwerdt, 2005). In the

https://doi.org/10.1017/S0031182017001263 Published online by Cambridge University Press

absence of sequence data, reporting PCR-positive samples alone may overestimate *Bartonella* prevalence in such ecological studies. Therefore, we advocate that studies of *Bartonella* prevalence, and ideally all surveys of infectious bacteria, should adhere to the standard of reporting only sequence-positive samples. Besides the clarity of the results, assessing the diversity and prevalence of *Bartonella* species in ecological studies with sequence-based approaches will facilitate comparisons of bacterial prevalence and diversity across various spatial and temporal scales and among host species and communities. These analyses are necessary for a deeper understanding of the ecology and evolution of *Bartonella* species and other infectious bacteria.

CURRENT APPROACHES TO GENOTYPING BARTONELLAE

Modern microbiologists rely heavily on molecular techniques and associated thresholds to assess the diversity of bacteria in various environments and classify new species. Sequencing of the 16S rRNA gene (Woese and Fox, 1977) and DNA–DNA hybridization experiments (Wayne *et al*. 1987) are two methods capable of delineating bacterial species that were developed fairly early, yet in the intervening years, the limitations of these approach have become more clear (Konstantinidis *et al*. 2006). We will highlight some of these limitations below, and where applicable, make specific connections to the study of bartonellae.

Is bacterial isolation a necessary step for genotyping of bartonellae?

The isolation of *Bartonella* bacteria from infected animals is the preferred method for the diagnosis and characterization of species (Gutiérrez *et al*. 2017). Gutiérrez *et al*. (2017) discuss various methods that are successful in culturing bartonellae and make recommendations for particular sample types (tissue *vs* ectoparasite). However, one persistent challenge is that many microbes will be challenging to isolate with known culturing techniques. We do acknowledge that cultured isolates are crucially important to the identification of bacterial species and obtaining a culture will facilitate all varieties of morphological, biochemical and genetic analyses. There are projects already underway to sequence the whole genome of all known bacterial-type strains (Kyrpides *et al*. 2014), and these data will no doubt expand our knowledge of bacterial diversity and genomic architecture. For bartonellae specifically, culturing can be very time-consuming due to the slow growth of the bacteria (which can be complicated by overgrowth of other contaminating bacteria) and may not be able to detect some species that do not grow quickly on standard blood agar. Thus, a strictly culture-based assessment may be severely biased towards cultivable strains.

The 16S ribosomal RNA sequencing and the problem of identification of Bartonella *species*

Assessing bacterial diversity using 16S rRNA sequences has become a very popular technique, especially with the advent of high-throughput sequencing instruments (e.g. Roche 454, Illumina MiSeq/HiSeq and Ion Torrent). This approach has been able to uncover an enormous diversity of bacterial and archaeal taxa, some of it consisting of heretofore uncultured microbial 'dark matter' (Rinke *et al*. 2013; Saw *et al*. 2015), in environments ranging from mammalian guts and feces (Manichanh *et al*. 2006; Bittar *et al*. 2014) and parasitic arthropods (Qiu *et al*. 2014; Razzauti *et al*. 2015) to marine habitats (Logares *et al*. 2014).

Nevertheless, there is substantial evidence that 16S rRNA sequencing (and its high-throughput applications) may be inadequate for accurately identifying bacterial species and may not be sensitive enough to recover the complex evolutionary histories of many microbial species. Assessment of microbial taxonomic diversity using 16S rRNA sequences commonly follows a threshold of 97% sequence identity to delineate operational taxonomic units, which can obscure the distinction between closely related species and even genera (e.g. *Escherichia* and *Shigella*). Many authors have determined that *Bartonella* species exhibit very high levels of 16S rRNA gene sequence similarity (Birtles and Raoult, 1996). Comparing sequences of 17 *Bartonella* species and subspecies, La Scola *et al*. (2003) reported the lowest discriminatory power (99·7%) and highest interspecies similarity (99·8%) for 16S rRNA, making this genetic locus an ineffective tool for the systematic classification of related bacterial species.

Metagenomics of microbial communities and needs for Bartonella *genotyping*

While reporting the low discriminatory power of 16S rRNA for identification of *Bartonella* species, La Scola *et al*. (2003) acknowledged that this locus is still reliable for differentiation of all *Bartonella* species from *Brucella* species (94% similarity), the genus taxonomically closest to *Bartonella*. This fact can justify the application of ribosomal primers for identification of bartonellae as components of microbial communities using 16S rRNA amplicon sequencing. Few surveys based on metagenomic evaluation of rodent-associated bacteria, including bartonellae, were conducted recently (Razzauti *et al*. 2015; Galan *et al*. 2016; Koskela *et al*. 2017). These studies highlight the utility of metagenomic techniques; however, they are not without their

https://doi.org/10.1017/S0031182017001263 Published online by Cambridge University Press

own limitations regarding the distinction among related bacterial species and potential sequence amplification biases.

Razzauti *et al*. (2015) compared two next-generation sequencing approaches (transcriptome RNA sequencing and 16S metagenomics) according to their ability to survey multiple bacteria in rodent populations in the French Ardennes region. Among vector-borne bacteria, *Bartonella* was the most prevalent (>5 reads in 89% of the rodents by 16S sequencing). The authors acknowledged that an important limitation of these approaches is the accuracy of the taxonomic assignation. RNA sequencing allowed taxonomic classification at the species level, while 16S metagenomics classification was generally restricted to the genus level. Analysing this problem, Razzauti *et al*. (2015) stressed the point that the 16S rRNA gene is difficult to sequence in its totality because of the size (~1550 bp) using current high-throughput sequencing methods. The method proposed by Miller *et al*. (2011) allows to assembly steps, but is not frequently used because of the increased experimental complexity and cost. Instead, a portion of the 16S rRNA gene is usually amplified using specific sets of universal primers. The nine hypervariable (V) regions of the 16S rRNA gene differ between species, and depending on the V region chosen, one can discriminate some species but not others. In their paper, Razzauti *et al*. (2015) have also reported an interesting observation about a large difference in the relative abundance of *Bartonella* reads detected by the 16S MiSeq (95%) *vs* RNA sequencing (<1%).

Galan *et al*. (2016) investigated the potential for recent developments in 16S rRNA-based high-throughput sequencing (Illumina MiSeq) to facilitate the multiplexing of urban rodents in West Africa. This study reported significant difference in *Bartonella* prevalence between rodent species varying from 0·5% in *Mus musculus* to 79% in *Mastomys natalensis*. Praising advances in this screening strategy, the authors admit that 16S rRNA amplicon sequencing based on a short sequence did not yield results sufficiently high in resolution to distinguish between *Bartonella* species (Galan *et al*. 2016). Another metagenomic evaluation of bacteria in voles from Finland (Koskela *et al*. 2017) reported commonality of *Bartonella* species in the voles, although identification of the species was not clear. André *et al*. (2017) used 16S metagenomics to investigate the liver microbiome of *Peromyscus leucopus* mice in Canada, finding no difference between the microbiome assemblages of mouse genotypes separated by the Saint Lawrence River. In contrast to the other studies above, the authors used an additional marker (16S–23S intergenic spacer, ITS) to identify all of the *Bartonella* species as *B. vinsonii arupensis*, a known zoonotic agent in humans (Welch *et al*. 1999; Bai *et al*. 2012).

Banskar *et al*. (2016) used 16S metagenomic sequencing (Ion Torrent) to investigate the fecal microbiome of *Rousettus leschenaultii* bats in India. They found a high abundance of Proteobacteria in some of the samples, which contains a large number of pathogenic genera, including *Bartonella*. However, the authors claim to have detected *Bartonella henselae* in two of the bat samples, which is highly unlikely given the strong association of *B. henselae* with cats. This misidentification is most likely due to the authors' use of a 97% sequence identity threshold, which is insufficient to distinguish among *Bartonella* species. Dietrich *et al*. (2017) also applied 16S metagenomic sequencing (Illumina MiSeq) to characterize the microbiome in saliva, urine and feces from four species of insectivorous bats from South Africa. Similarly to Banskar *et al*. the authors found a high abundance of Proteobacteria in bat feces, but also in saliva and urine. Sequences mapping to *Bartonella* were found predominantly in feces, but also to some extant in saliva and urine. No attempts were made to identify the specific *Bartonella* species found in these samples; however, as we have discussed above, this would likely not be possible using only 16S sequences.

In review, the utility of 16S sequencing will largely depend on the questions investigators wish to pursue, and the scale of phylogenetic resolution needed to answer such questions. If investigators wish to assess bacterial diversity in specimens at the genus level or higher, 16S metagenomics would be an excellent approach. As we have reviewed above, below the genus level however, this gene will not be sufficient to accurately distinguish among related species. Depending on the focus of the study, investigators could then target a few genera of interest for characterization with more discriminating genetic loci (André *et al*. 2017). For example, investigators may target genera with high abundance in the 16S dataset that may contain pathogenic species. This approach has been used with success recently to describe *Bartonella* species in bats (Veikkolainen *et al*. 2014; Wilkinson *et al*. 2016).

DNA–DNA hybridization and its limitations for identification of bacteria

The DNA–DNA hybridization experiments used by Wayne *et al*. (1987) represented a potentially more robust approach for delineating bacterial species using the whole genome. A threshold of 70% hybridization has been used as the 'gold standard' criterion for distinguishing new bacterial species (Tindall *et al*. 2010); however, this technique requires the use of cultured isolates, specialized equipment and

https://doi.org/10.1017/S0031182017001263 Published online by Cambridge University Press

multiple confirmatory tests due to variation across experimental runs. Alternative genome-wide distance measures included average nucleotide identity (Konstantinidis and Tiedje, 2005; Konstantinidis *et al*. 2006; Goris *et al*. 2007; Richter and Rosselló-Móra, 2009) and digital DNA–DNA hybridization (Auch *et al*. 2010; Meier-Kolthoff *et al*. 2014*a*, *b*). One considerable advantage of these techniques is that they do not necessarily require cultured isolates and can be calculated from draft genomes assembled from metagenome and transcriptome sequencing. However, as noted above, these high-throughput techniques are currently not accessible to very many research groups and will not have much utility unless investigators have a bacterial isolate or a draft genome.

Recombination as an important complication for genotyping

Another important issue that can arise when attempting to genotype a bacterial strain is that separate genes may indicate the presence of different species. These conflicts arise due to lateral gene transfer (LGT) among bacteria, either directly through conjugation or indirectly *via* phage-mediated transduction or transformation by uptake of free DNA in the environment. LGT is the predominant mechanism by which bacteria acquire antibiotic resistance genes and can be an important part of bacterial evolution (Vos, 2009). Homologous recombination is a specific form of LGT whereby homologous genes of a donor genome replace the gene variant in the recipient genome. Homologous recombination is a common feature among some bacterial species (Vos and Didelot, 2008) and even among distantly related bacteria (Hanage *et al*. 2006), thus severely complicating phylogenetic inference. This problem has been documented in several studies of *Bartonella* strains from cats, rodents and bats based on sequencing multiple protein-coding loci (Berglund *et al*. 2010; Chaloner *et al*. 2011; Paziewska *et al*. 2011, 2012; Guy *et al*. 2012; Buffet *et al*. 2013*a*; Bai *et al*. 2015*a*) and have provided valuable information about the mechanisms that generate *Bartonella* diversity and the gene flow among co-occurring species. We note here that these studies have been limited to cultured strains. As we will discuss later in the paper, attempts to genotype bartonellae from genomic DNA extracted from whole blood, tissue or from ectoparasites may be further complicated by the presence of multiple *Bartonella* species in the sample. However, sequencing multiple loci will clarify if recombination or multiple species are present and phylogenetic concordance among sequenced loci can be sufficient to describe a potentially novel *Bartonella* species or subspecies.

Gene-sequence-based paradigm for identification of Bartonella *isolates*

There are existing methods, particularly multi-locus sequence typing (MLST; Stackebrandt *et al*. 2002), which can balance the tradeoffs of culturing bias, phylogenetic resolution, homologous recombination and gene conservation across species. MLST of house-keeping genes (i.e. genes under stabilizing selection encoding metabolic functions) remains a powerful technique that can be used on uncultured bacteria to detect evidence of mixed infections and/or homologous recombination, provide sufficient phylogenetic resolution for the delineation of bacterial species, and will provide consistency in the usage of genetic loci that can facilitate global assessments of parasitic bacterial diversity.

Using such an approach, La Scola *et al*. (2003) compared the similarities of seven genetic loci among the 17 species and subspecies of genus *Bartonella*. This comparison led to both the definition of similarity values that discriminated *Bartonella* at the species level and assessment of the relative discriminatory power of each gene examined. The *gltA*, *groEL*, *rpoB* and *ftsZ* genes, and ITS all have good discriminating power ranging from 92·6 to 94·4%. Overall, two genes (*rpoB* and *gltA*) were found to be the most potent markers for demarcation of *Bartonella* species (La Scola *et al*. 2003). This paper was very influential for characterization of *Bartonella* cultures and defining their status as a species. Many studies have now used MLST approaches to characterize *Bartonella* genotypes and species, and based on the available methods, some form of multi-locus sequencing appears to be the most viable method available to most researchers.

Confirmation of species status based on comparison of gene profiles among related species

Characterizing the novel *Bartonella* species (*B. melophagi*) isolated from sheep blood and sheep keds, Kosoy *et al*. (2016) reported presence of 183 genes specific for this species, being absent in genomes of other *Bartonella* species associated with ruminants to support their argument for the separation of this bacterial species from species of other ruminant-associated *Bartonella* species. The authors identified that out of the 1338 genes, the number of homologous but unique genes was estimated to be 1274, out of which 156 genes appeared to be specific to *B. melophagi* and absent in any of the 21 reference genomes *Bartonella* species. Comparison of the gene profile of *B. melophagi* with related *Bartonella* species associated with ruminants (*B. bovis* and *B. schoenbuchensis*) demonstrated that 183 genes were present only in the genome of *B. melophagi*, while 1027 genes present in one or more copies in each genome were conserved between these three

https://doi.org/10.1017/S0031182017001263 Published online by Cambridge University Press

bacterial strains. The remaining 27 genes were present in *B. melophagi* and absent in related species (*B. bovis* and *B. schoenbuchensis*), but found in at least one of the other *Bartonella* species. This analysis indicates that even among related bacterial species, genomes can be very flexible in gene content, and can be useful criterion for describing novel species (Konstantinidis and Tiedje, 2005; Konstantinidis *et al*. 2006).

Applying multi-locus approaches to detection and sequencing without cultured isolates

Due to the challenges of culturing bartonellae, investigators may choose to characterize *Bartonella* infections in animal samples directly from extracted DNA. Despite the convenience of such an approach, it has its own challenges. For instance, some primers used to amplify genetic loci may be insufficiently sensitive to amplify *Bartonella* DNA from some animal tissues. The presence of PCR inhibitors that carry through from blood, tissue or the extraction process may also interfere with detection. The use of nested PCR reactions may be able to overcome some of these deficiencies among sensitivity among loci for detection and may not require much additional primer design. For instance, protocols exist for amplifying *gltA* and *ftsZ* sequences using nested reactions with known primer sets (Norman *et al*. 1995; Birtles and Raoult, 1996; Zeaiter *et al*. 2002; Colborn *et al*. 2010; Gundi *et al*. 2012*a*).

Another challenge with amplification directly from extracted DNA is the potential presence of multiple bacterial species in the sample, and coinfections of multiple *Bartonella* genotypes in one animal are not uncommon based on culturing. When multiple species are present in a sample, the abundance of their DNA may vary in the sample and even across tissue types. Furthermore, different primer sets may have amplification bias towards particular species based on the annealing affinity. These complications may cause the observed *Bartonella* diversity to differ depending on which marker was used for amplification, so a single marker may not be a robust indicator of total *Bartonella* diversity in a set of samples (Buffet *et al*. 2013*a*). Furthermore, investigators have observed recombination events even within a single gene (*gltA*), interfering with phylogenetic inference, but may not be present in other sequenced loci (Paziewska *et al*. 2012; Buffet *et al*. 2013*a*).

Nevertheless, this approach has one very important disadvantage – when sequenced loci are in conflict regarding the bacterial species identified in the sample, one must determine if this is caused by homologous recombination or the presence of multiple infections. In these cases, researchers may choose to report the conflicting results as is and simply note this caveat with the understanding that culturing, cloning sequences into vectors before sequencing, or deep sequencing approaches may differentiate these possible scenarios. In some cases, multiple peaks may be visible in chromatograms of sequences, so cloning would be useful in these cases, but not all cases of multiple infections show this pattern, probably due to varying abundances of DNA that are not detected in the consensus sequence reads.

Even with these known limitations, this multi-locus sequencing approach has been used recently to detect and characterize *Bartonella* genotypes from bats, rodents and carnivores. Lilley *et al*. (2015) used a combination of *rpoB* and *gltA* sequences to characterize a novel *Bartonella* species (*Candidatus B. hemsundetiensis*) in *Myotis daubentonii* bats from Finland. Similar *Bartonella* species have subsequently been cultured and characterized from related insectivorous bats in the Republic of Georgia (Urushadze *et al*. 2017). Martin-Alonso *et al*. (2016) used multiple loci (ITS, *gltA* and *rpoB*) to detect *Bartonella* infections in rodent species from Benin. Based on these markers, the authors describe a distinct *Bartonella* species (*Candidatus B. mastomydis*) from *M. natalensis*. Sequences very similar to this candidate species had previously been acquired from related rodent species in Ethiopia. The authors also reported the presence of multiple peaks in their sequencing results, so they used cloning to distinguish the coinfections. However, there were additional conflicts between *gltA* and *rpoB* sequences for some samples that did not show multiple sequence peaks, with one locus indicating the presence of *Bartonella elizabethae* and the other indicating *Bartonella tribocorum* (Martin-Alonso *et al*. 2016). The authors hypothesize that these conflicts may have arisen by recombination; however, as we noted above, multiple infections (with no evidence of multiple sequence peaks) may be an alternative explanation.

These studies, although fairly recent, demonstrate the potential of this multi-locus approach to characterizing bartonellae without a culturing step. New *Bartonella* species can be described across multiple genes showing phylogenetic concordance, or in other cases, interesting cases of potential recombination or multiple infection can be noted. In this way, multi-locus sequencing can be an important first step, focusing on detection and partial genotyping, with other analyses following after to fully characterize novel or recombinant genotypes by culturing and MLST (or full genome analyses). Multi-locus sequencing therefore strikes a valuable balance by providing potentially more robust assessments of *Bartonella* diversity than single-locus approaches, and is also more accessible to a wider community of researchers than full genomic approaches since it requires only standard molecular techniques (PCR and Sanger sequencing).

https://doi.org/10.1017/S0031182017001263 Published online by Cambridge University Press

Metagenome and transcriptome sequencing – the way of the future?

Shotgun metagenome and transcriptome sequencing techniques that target many coding loci are becoming popular methods for identifying bacteria at the species level with better phylogenetic resolution at low per-base cost (Venter *et al*. 2004; Rinke *et al*. 2013; Logares *et al*. 2014; Hug *et al*. 2016) and may represent a new way forward for identifying bacterial pathogens in a large number of samples and tissue types. However, the cost of deep sequencing and absence of comprehensive reference databases (the majority of environmental microorganisms have yet to be sequenced) making this approach unavailable for all but the most well-funded laboratories. We look forward to seeing these high-throughput metagenome and transcriptome approaches applied more frequently (and we expect they will as the cost of machinery and computing resources become more available and affordable), but for now we seek to make recommendations for genotyping bartonellae that are accessible to a wider community of researchers. As noted above, a multi-locus sequence approach may be the best option for many studies and could facilitate broad-scale comparisons of *Bartonella* diversity across systems if researchers use a consistent set of markers.

REVIEW OF STUDIES FOCUSING ON GENOTYPING BARTONELLAE FROM ANIMALS SAMPLES

In order to make recommendations for sequence-based approaches for genotyping bartonellae from archived animal samples, we performed a literature review to identify commonly used genetic markers. Based on the results of this survey, we will identify some candidate markers that could become consistent features of the multi-locus sequencing approach we described above, and thus facilitate valuable comparative studies of *Bartonella* ecology and evolution.

Analysis of literature on identification of bartonellae in animal hosts

We surveyed a sample of published literature (>400 studies) using paired key words 'bartonella-rodents', 'bartonella-bats', 'bartonella-wildlife', 'bartonella-cats', 'bartonella-dogs' and 'bartonella-ectoparasites'. Of the processed literature, 293 studies were selected with available information on application of diverse genetic markers for genotyping of bartonellae in identified tissues of vertebrate animals and/or their ectoparasites. These studies report investigations conducted in 79 countries of Africa (19), the USA (11), Asia (21), Australia/Oceania (5) and Europe (23), and from a broad diversity of animal taxa, including rodents, bats, carnivores, ruminants and marine mammals (Table 1). In the 101 studies, bartonellae were cultured from blood, followed by genotyping of the isolates. Of those, culturing work was accompanied with molecular detection of *Bartonella* DNA in tissues only in 16 studies, while detection of *Bartonella* DNA in ectoparasites along with culturing bacteria from their hosts was attempted in 21 studies.

Selection of animal tissues for detection and genotyping of Bartonella *species*

The most frequent tissue for targeting and genotyping *Bartonella* DNA by PCR and sequencing was blood: 62 studies where only blood was used and three studies where other tissues along with blood were analysed (Table 1). Other tissues used for *Bartonella* genotyping are: spleen (29), liver (8), heart (7), kidney (5), lung (2), and ear, skin and nail by one study. Besides blood samples, only one tissue type was analysed in 34 studies, two tissues in eight studies and more than two tissues in three studies.

Overall, there are limited reports about significant variation in detection of *Bartonella* DNA between tissues. Razzauti *et al*. (2015) noted that the choice of organ likely has an important impact on the detection or misdetection of *Bartonella*. To explain the huge difference in relative abundance of *Bartonella* reads detected by 16S MiSeq *vs* RNA-Seq cited above, the authors used the currently accepted model of *Bartonella* infection described by Harms and Dehio (2012). This model posits that immediately after infection, bartonellae colonize an unknown primary niche in the mammalian host, most likely vascular endothelial cells. Every 5 days, some of the bacteria in the endothelial cells are released into the blood stream, where they infect erythrocytes. Then bacteria invade a phagosomal membrane inside the erythrocytes, where they multiply until they reach a critical population density. At this point, they simply wait until they are taken up with the erythrocytes by a blood-sucking arthropod. The spleen plays important roles with regard to erythrocytes by removing old erythrocytes, and may thereby hold a reserve of erythrocytes that are highly infected by non-replicating bartonellae, which do not produce RNA molecules. Moreover, due to its central role in recycling erythrocytes, the spleen could also store a large amount of degraded DNA of dead bartonellae (Razzauti *et al*. 2015).

Genetic markers used for identification of bartonellae

Based on our review, the total number of genetic loci that have been used for genotyping *Bartonella* DNA, either from bacterial culturing or tissue extracts, reached 41 (Fig. 1a). The applied markers include both coding genes and intergenic regions. From

https://doi.org/10.1017/S0031182017001263 Published online by Cambridge University Press

Table 1. Studies aimed to identifying bartonellae in different tissues of vertebrate animals and their ectoparasites (only studies with identified animal hosts are selected; studies with testing off-hosts arthropods are excluded)

Host	Number of studies	Culture	DNA extracted directly from tissues						DNA extracted from ectoparasites
			Blood	Heart	Spleen	Liver	Kidney	Lung	
Bats	28	11	9	2	1	0	1	0	11
Rodents	120	61	12	3	23	5	2	2	43
Lagomorphs	4	3	1	0	0	0	0	0	2
Cats	33	4	14	0	0	0	0	2	21
Dogs	15	2	5	0	0	0	0	0	9
Domestic ruminants[a]	13	6	4	0	0	0	1	0	7
Wild ruminants[b]	15	3	8	2	2	2	0	0	9
Wild carnivores[c]	15	6	6	2	2	0	0	0	3
Marine mammals[d]	4	0	3	0	0	0	0	0	1
Other animals[e]	11	5	3	0	1	0	0	0	3
Total	258	101	65	9	29	2	4	2	109

[a] Cattle, sheep, goats, horses, camels.
[b] Deer, moose, feral pigs.
[c] Coyotes, foxes, jackals, raccoons, others.
[d] Seals, dolphins, porpoises, sea otters.
[e] Exotic animals, marsupials, primates, birds, turtles.

1994 when the genotyping of *Bartonella* was initiated through 2002, the procedure of genotyping was limited to the application of two markers (*gltA* and 16S). The number of genetic loci used has steadily increased after 2002 (Fig. 1a), although the majority of studies only use one or two markers (Fig. 1c). Most of the genetic loci were used in only a few studies and were not repeated in laboratories other than ones where they were proposed. Only 10 genetic loci were used >10 times in multiple laboratories (*gltA*, ITS, *rpoB*, 16S rRNA, *ftsZ*, *groEl*, *ribC*, *pap31*, *nuoG* and *ssrA*) with *gltA* being the most frequently used marker across all studies surveyed (Fig. 1a).

Beyond detection, genetic targets that provide sufficient sequence diversity to allow differentiation of *Bartonella* species are required to fully understand the distribution and host specificity of various *Bartonella* species and identification of the strains associated with human illness. The citrate synthase gene (*gltA*), originally proposed by Norman *et al*. (1995), remains the most popular genetic target for *Bartonella* detection and is considered a reliable tool for distinguishing genotypes. In our review, *gltA* was used in 48 of 56 of the studies where one or two markers were applied for identification of *Bartonella* cultures [Table 2(a)]. Other markers, particularly *rpoB* and *ftsZ*, are common when at least four markers are used for direct detection of *Bartonella* DNA in tissues by PCR; ITS is used frequently, comparable with the *gltA* and more often than *rpoB* and *ftsZ* [Table 2(b)].

Few attempts have been made to culture bartonellae from arthropod ectoparasites, so identification of *Bartonella* from ectoparasites is typically performed by PCR on extracted DNA. Studies have identified *Bartonella* DNA in a number of ectoparasite groups: fleas (80), ticks (40), lice (13), bat flies (9), deer and sheep keds (9), mites (6), *Cimex* spp. bugs (2), bees (2) and ants (1). Detection and genotyping primarily target ITS and *gltA*, with *rpoB* being the third most common marker [Table 2(c)].

Comparison of genetic markers for detection and genotyping of Bartonella *DNA in animal tissues*

Of 54 publications where identification and genotyping of bartonellae in mammalian tissues were conducted with at least two different genetic markers, only 13 studies provided data for comparing the effectiveness of using different genetic loci (Table 3). In almost all of these studies, the ITS target was the most sensitive marker for identification of *Bartonella* DNA in blood. Only one study focused on detecting and genotyping *Bartonella* DNA in cat blood found the ITS and *gltA* targets to be equally productive (Bai *et al*. 2015*b*). While detecting and genotyping *Bartonella* in rodent spleens, two studies reported successful identification in more specimens by targeting the *rpoB* gene compared with the *gltA* (Gundi *et al*. 2010, 2012*b*).

Birtles and coworkers described the use of PCR-based amplification of ITS fragments to detect and identify bartonellae in the blood of rodents. Direct detection was of particular use in the longitudinal

https://doi.org/10.1017/S0031182017001263 Published online by Cambridge University Press

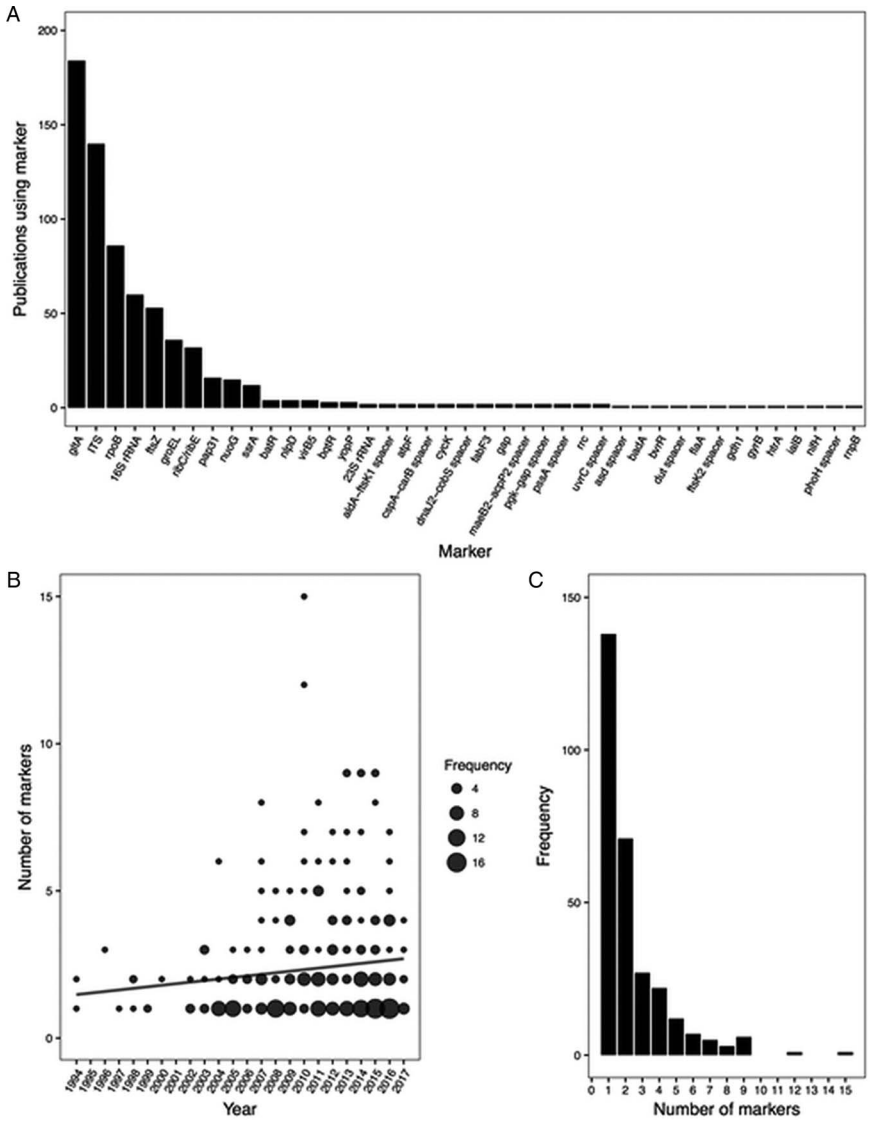

Fig. 1. Frequency of genetic loci used for genotyping bartonellae (a), temporal trend in the number of markers used for genotyping (b) and the frequency distribution of the number of markers used across 293 reviewed studies.

survey of *Bartonella* bacteraemia that involved the field collection of very small amounts of blood from live, wild rodents (Birtles *et al*. 2000). As most of the ITS is non-coding, it is prone to hypervariability, and its sequence variation is markedly higher than that observed at other genetic loci (Roux and Raoult, 1995). Although comparison of ITS sequences is useful for the allocation of detected organisms into one of the recognized *Bartonella* species, detection of a novel ITS sequence can be problematic because of the difficulties with sequencing of amplified fragments and problems with alignment of the obtained sequences (Knap *et al*. 2007). ITS sequences have many insertions and deletions that can complicate phylogenetic analysis. In some of the studies, screening was conducted by conventional or real-time PCR of ITS, followed by sequencing of additional markers, usually *gltA* (Miceli *et al*. 2013; Gutiérrez *et al*. 2015; Bai *et al*. 2016).

Comparison of genetic markers for detection and genotyping of Bartonella *DNA in ectoparasites*

Since only few attempts of culturing *Bartonella* from arthropods have been successful, genotyping of *Bartonella* in insects and acarines relies mostly on detection from extracted DNA by PCR, typically relying on only one marker. In spite of a large number of publications reporting investigation of *Bartonella* in ectoparasites (>140), we were able to select only 14 publications, which provided data on comparison of at least two genetic markers for genotyping *Bartonella* in DNA extracted from arthropods (Table 4). In four of the 14 studies, ITS was shown to be most sensitive marker for detection and genotyping *Bartonella*; however, in some other studies, success with *gltA* gene was similar (De Sousa *et al*. 2006; Pérez-Martínez *et al*. 2009).

https://doi.org/10.1017/S0031182017001263 Published online by Cambridge University Press

Table 2. Summary of markers used for the characterization of bartonellae (a) from cultures, (b) tissues and (c) ectoparasites

Number of markers	Number of studies	*gltA*	ITS	16S	*rpoB*	*ftsZ*	*ribC*	*nuoG*	*ssrA*	Other
(a) Cultures										
1	44	37	2	5	2	0	0	0	0	0
2	12	11	5	3	3	0	0	0	0	2
3	7	6	5	1	4	0	0	1	1	2
4	9	9	4	4	8	8	2	0	1	0
5	8	8	5	6	8	8	2	0	0	2
>5	20	19	9	16	19	18	16	5	5	16
Total	100	90	30	35	44	34	20	6	7	22
(b) Tissues										
1	55	21	23	3	3	1	2	2	0	0
2	24	17	14	3	9	2	0	1	0	1
3	17	14	12	3	10	1	1	3	1	1
4	8	8	6	1	6	3	1	2	3	2
5	3	3	2	1	3	2	2	0	0	0
>5	3	3	3	2	3	2	2	1	1	2
Total	110	66	60	13	34	11	8	9	5	6
(c) Ectoparasites										
1	79	28	37	5	2	1	2	1	2	0
2	35	18	22	9	5	3	2	1	0	4
3	10	8	6	1	3	1	0	2	1	1
4	7	5	7	1	5	2	0	1	2	2
5	8	8	7	6	8	8	1	0	0	0
>5	2	2	1	1	2	2	2	0	0	1
Total	141	69	80	23	25	17	7	4	3	12

When Morick *et al*. (2010) genotyped bartonellae in fleas collected from rodents in the Negev Desert of Israel using three genetic markers (*gltA*, ITS and *rpoB*), they found the 313 bp *gltA* fragment to be the best target for screening fleas for *Bartonella* and for identification to species level. All flea pools that were found positive by *rpoB* or ITS screening were also positive by *gltA*. Pérez-Martínez *et al*. (2009) investigated 82 fleas collected from cats and dogs in Chile. When *rpoB* primers were used, *Bartonella* genotypes were found in four *Ctenocephalides felis* fleas from cats (4·8%) and in four *Pulex irritans* fleas from dogs (4·8%). The same eight samples were positive when primers for *gltA* and ITS were used. None of the 82 specimens were positive when primers targeting the *groEL* gene were used. Conducting surveillance of Egyptian fleas for agents of public health significance, Loftis *et al*. (2006) detected more *Bartonella*-positive fleas using *groEL* than ITS (17 *vs* 11) and were successful in conducting phylogenetic analysis based on comparison of the *groEL* sequences rather than ITS sequences.

Contribution of analyses of complete Bartonella *genomes to primer design*

Cross-referencing the *gltA* primer set against the GenBank dataset showed that despite their common use for *Bartonella* detection, these primers have high cross-reactivity both to potential *Bartonella* host DNA (such as *Rattus*, *Mus* and *Homo sapiens*) and to bacterial species that could inhabit similar ecological niches (such as *Ehrlichia*) (Colborn *et al*. 2010). To identify genus-specific and host-blind primer sets, a whole-genome scan of three *Bartonella* genomes (*B. henselae*, *B. quintana* and *B. bacilliformis*) available at that time was performed (Colborn *et al*. 2010), and the NADH dehydrogenase γ subunit (*nuoG*) primer set was identified and met all the required conditions. A few years later, another genetic locus (*ssrA*), also known as transfer-messenger RNA, was proposed as a target for a genus-specific real-time PCR assay based on analyses on whole genomes (Diaz *et al*. 2012). These markers have been used in a number of studies for the detection of *Bartonella* DNA in animal tissues and ectoparasites, with successful detection at rates similar to other loci but still lower than ITS (Gutiérrez *et al*. 2014; Brook *et al*. 2015; Bai *et al*. 2015*b*).

Recommendations for marker usage in a multi-locus sequencing framework

Overall, the *Bartonella gltA* sequence database in GenBank is the largest and most frequently updated among the different collections of deposited sequences, and therefore allows a more accurate differentiation between *Bartonella* species and strains. A proteomic analysis of *gltA* indicates that most amino substitutions are synonymous, highlighting the important and critical function of the citrate synthase (*gltA*) enzyme. Nevertheless, numerous

https://doi.org/10.1017/S0031182017001263 Published online by Cambridge University Press

Table 3. Comparison of genetic loci used for genotyping of *Bartonella* DNA in animal tissues by their detection frequency

Citation	Host/tissue	Number of samples	ITS	*gltA*	*rpoB*	*ftsZ*	*ribC*	*nuoG*	*ssrA*	*pap31*	*groEL*
Bai *et al.* (2015*b*)	Cat/blood	142	44/142	44/142				29/142			
Gutiérrez *et al.* (2015)	Cat/blood	36	23/36	21/23[a]							
Miceli *et al.* (2013)	Cat/blood	163	3/163	3/3[a]	3/3[a]		0/3[a]			0/3[a]	
Braga *et al.* (2012)	Cat/blood	200	8/200	6/200	3/200		7/200				5/200
Kamani *et al.* (2014)	Bat/blood	148	76/148[b]	14/148	0/148						
Brook *et al.* (2015)[c]	Bat/blood	76	24/76	5/76							
Gutiérrez *et al.* (2014)	Cattle/blood	50	15/50[b]						14/50[b]		
Carrasco *et al.* (2014)	Sea otter/heart	51	23/51		2/51					7/51	
Bai *et al.* (2016)	Wild carnivores/spleen	292	38/292	36/38[a]		33/38[a]					
Ko *et al.* (2013)	Deer/spleen	70	20/70		5/70						
Ko *et al.* (2016)	Rodents/spleen	200	124/200		63/200						
Gundi *et al.* (2012*b*)	Rodents/spleen	98		20/98	24/98						
Gundi *et al.* (2010)[c]	Rodents/kidney, liver, lung	324		42/324	76/324						

[a] The genetic markers used as the second step for genotyping *Bartonella* in positive DNA after initial screening.
[b] Real-time PCR assay was used for detection of *Bartonella* DNA without genotyping.
[c] Data are not in publication, but provided from a private communication.

studies have indicated that ITS is a highly sensitive marker that is invaluable for the detection of *Bartonella* DNA. In order to maximize detection success and differentiation among related *Bartonella* species, we advocate for a multi-locus sequencing approach. Although many markers have been used in different studies, there is a growing consensus of frequently used markers – specifically, *gltA*, ITS, *rpoB*, *ftsZ*, *ribC*, *groEL*, *nuoG* and *ssrA* – that are generally capable of differentiating among *Bartonella* species, particularly when used together in a multi-locus genotyping framework (La Scola *et al.* 2003). Usage of these markers consistently across studies will facilitate ecological analyses of *Bartonella* prevalence and diversity across systems and comprehensive phylogenies of known *Bartonella* species.

EVALUATION OF PHYLOGENETIC RESOLUTION AMONG CANDIDATE LOCI BASED ON ANALYSIS OF *BARTONELLA* GENOMES

La Scola *et al.* (2003) used seven protein-coding loci to genotype *Bartonella* strains, but as we noted above, these genes varied considerably in their power to discriminate among *Bartonella* species. In the intervening years, genomes of many *Bartonella* species have been sequenced and assembled. Using these data, we will evaluate the phylogenetic resolution of a number of candidate loci found in the genomes of *Bartonella* species and compare these results to the genetic markers frequently used to genotype bartonellae.

Detection of gene clusters in Bartonella *genomes*

Genomes from 22 publically available *Bartonella* species were downloaded from GenBank. Every pair of gene sequences from each genome was aligned using the Needleman–Wunsch global alignment algorithm to all other genes. The resulting alignment scores were placed in a square similarity matrix and genes were assigned to clusters using a single-linkage (non-centroid-based, non-greedy), exhaustive clustering algorithm. The clustering threshold was chosen in a way so that each gene cluster is expected to contain the same genes originating from different species. The constituent sequences of each gene cluster were then partitioned into separate FASTA files for subsequent analyses (L. Albayrak and C. McKee, unpublished data).

https://doi.org/10.1017/S0031182017001263 Published online by Cambridge University Press

Table 4. Comparison of genetic loci used for genotyping of *Bartonella* DNA in animal ectoparasites by their detection frequency

Citation	Host/ectoparasite	Number of samples	ITS	*gltA*	*rpoB*	*ftsZ*	*nuoG*	*groEL*
Bai *et al*. (2015*b*)	Cats/fleas	152	31/152	25/152				
Gutiérrez *et al*. (2015)	Cats/fleas	90	68/90	54/90				
Brook *et al*. (2015)[a]	Bats/flies	25		14/25		10/25	12/25	
Kamani *et al*. (2014)	Bats/flies	24	10/24[b]	7/24	7/24			
Loftis *et al*. (2006)	Rats/fleas	400	11/400					17/400
Morick *et al*. (2010)	Rodents/fleas	245	66/245	94/245	74/245			
Pérez-Martínez *et al*. (2009)	Cats and dogs/fleas	82	8/82	8/82	8/82			0/82
Rojas *et al*. (2015)	Cats and dogs/fleas	72		8/72	5/72			
De Sousa *et al*. (2006)	Rodents/fleas	56	4/56	4/56				
Bonilla *et al*. (2009)	Humans/lice	153		49/153	39/153			
Bonilla *et al*. (2014)[a]	Humans/lice	11	2/11	1/11	0/11	0/11		
Morick *et al*. (2009)	Seals/lice	5	1/5		1/5			
Kabeya *et al*. (2010)	Mites/rodents	40	29/40	9/40				
Billeter *et al*. (2011)	Ticks	10	2/10		3/10			

[a] Data are not in publication but provided from a private communication.
[b] Real-time PCR assay was used for detection of *Bartonella* DNA without genotyping.

Ranking of gene clusters by sequence diversity

The overall sequence diversity of each gene cluster was estimated as the ratio between the numbers of unique 32-base long subsequences present in all sequences in the cluster over the total number of 32-base long subsequences present in all sequences in the cluster. For each gene cluster, all 32-base long subsequences (32-mers) from each position in the nucleotide sequences in the FASTA file were collected. The reverse complements of the extracted subsequences were added to the complete set of 32-mers. Unique 32-base long subsequences in the set were identified and their ratio to the total number of subsequences (including duplicates) was calculated. We refer to this measurement as the proportion of unique 32-mers, and it varies between near zero and one. This measure is equal to one if all 32-base long subsequences identified in the clusters are unique.

Gene clusters were sorted in descending order based on the proportion of unique 32-mers and assigned numerical ranks accordingly. The top 10 (most diverse) and bottom 10 (most conserved) clusters containing sequences from each of 22 *Bartonella* genomes were selected. We also identified seven genes commonly used to genotype bartonellae (*ftsZ*, *gltA*, *groEL*, *gyrB*, *nuoG*, *ribC* and *rpoB*) from the ranking, corresponding to the following ranks out of 665 total gene clusters: 171 (*ribC*), 386 (*gyrB*), 543 (*nuoG*), 565 (*gltA*), 611 (*ftsZ*), 635 (*rpoB*) and 658 (*groEL*). This resulted in a list of 26 gene clusters since *groEL* was part of the bottom 10. We then added 16S rRNA (rank 665/665) and ITS (rank 652/665) to this ranking separately for each *Bartonella* species for which these sequences were available (L. Albayrak and C. McKee, unpublished data). The proportion of unique 32-mers is used here to assess the diversity of the nucleotide sequences in each cluster and is a useful measure for ranking many gene clusters by their sequence diversity. However, this measure does not necessarily reflect phylogenetic differentiation among congeneric taxa. Hence, we then quantified the sequence diversity and phylogenetic resolution of each of these 28 gene clusters using additional measures. Sequence diversity was assessed based on the proportion of segregating sites, Watterson's estimator of genetic diversity and nucleotide diversity. Phylogenetic resolution was measured by calculating Tamura-Nei sequence distances and storing the minimum, median and maximum distances. All calculations for these measures were performed in MEGA (Kumar *et al*. 2016).

Across all of the 28 gene clusters, other measures of sequence diversity generally followed a declining trend that corresponded to the ranking by proportion of unique 32-mers (Fig. 2a), and all of the measures were moderately to highly correlated ($0{\cdot}65 < r < 1$). However, there was some variation present in these estimates that was not captured in the proportion of unique 32-mers, particularly in the proportion of segregating sites. Tamura-Nei sequence distances similarly declined across the ranking of gene clusters, with some noticeable variation in the median and maximum distances (Fig. 2b). The nine genetic loci we analysed (16S rRNA, ITS, *ftsZ*, *gltA*, *groEL*, *gyrB*, *nuoG*, *ribC* and *rpoB*) fell between the top 10 and bottom 10 based on the proportion of unique 32-mers, with the exception of *groEL* and 16S, which had the eighth lowest and the lowest rankings, respectively.

https://doi.org/10.1017/S0031182017001263 Published online by Cambridge University Press

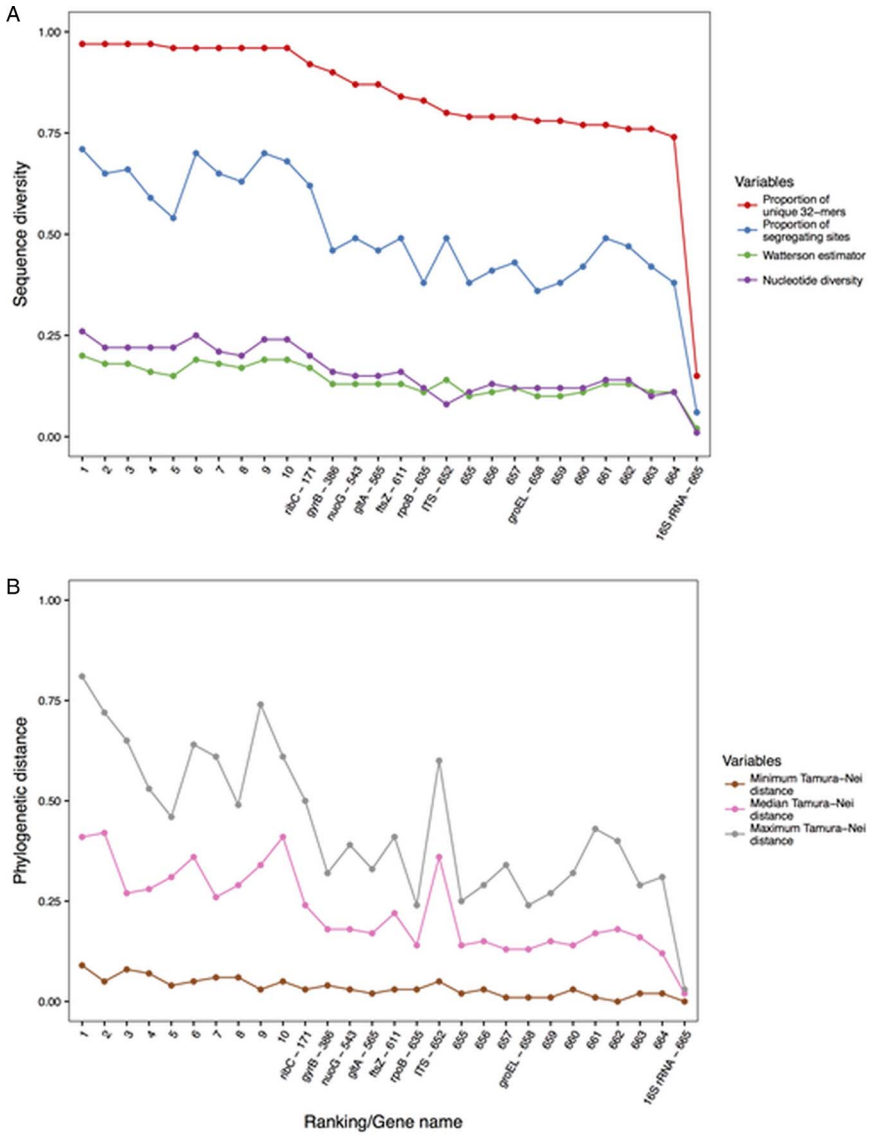

Fig. 2. Measures of sequence diversity (a) and phylogenetic distance (b) across 28 genetic loci occurring in all 22 *Bartonella* genomes currently available. Loci are numerically ranked in descending order along the bottom axis according to the proportion of unique 32-mers. Commonly used markers for genotyping are labelled with their gene names next to their numerical rank.

These two regions also had low sequence diversity by other measures and low Tamura-Nei distances, indicating that they have poor phylogenetic resolution. Overall, the top 10 candidate loci do show significantly higher measures of sequence diversity and phylogenetic distance than the 18 other loci (Fig. 3); however, the distributions of minimum Tamura-Nei distance among these groups do overlap (Fig. 1e). These minimum distances correspond to the inverse of the maximum sequence similarity that La Scola *et al*. (2003) used to assess discriminatory power among loci. Generally, these minimum distances are small among all loci, ranging from 0·002 for 16S rRNA to just 0·085 for the top-ranked gene cluster, an unnamed membrane protein (Fig. 2b). Among the eight other commonly used markers (ITS, *ftsZ*, *gltA*, *groEL*, *gyrB*, *nuoG*, *ribC* and *rpoB*), the minimum distances ranged were 0·012 for *groEL* to 0·038 for *gyrB* and 0·045 for ITS. The majority of these minimum distances were between *B. melophagi* and *B. schoenbuchensis*, two *Bartonella* species found in ruminants (deer and sheep).

Our results largely confirm what La Scola *et al*. (2003) found; however, our rankings of the markers with the ability to distinguish closely related species were somewhat different, and this is partly due to our usage of entire gene sequences for our measurements (La Scola *et al*. used only partial gene sequences). In both of our analyses, 16S rRNA displays the lowest ability to discriminate among *Bartonella* species. The other eight genetic loci commonly used for genotyping perform much better than 16S rRNA, with minimum distances

https://doi.org/10.1017/S0031182017001263 Published online by Cambridge University Press

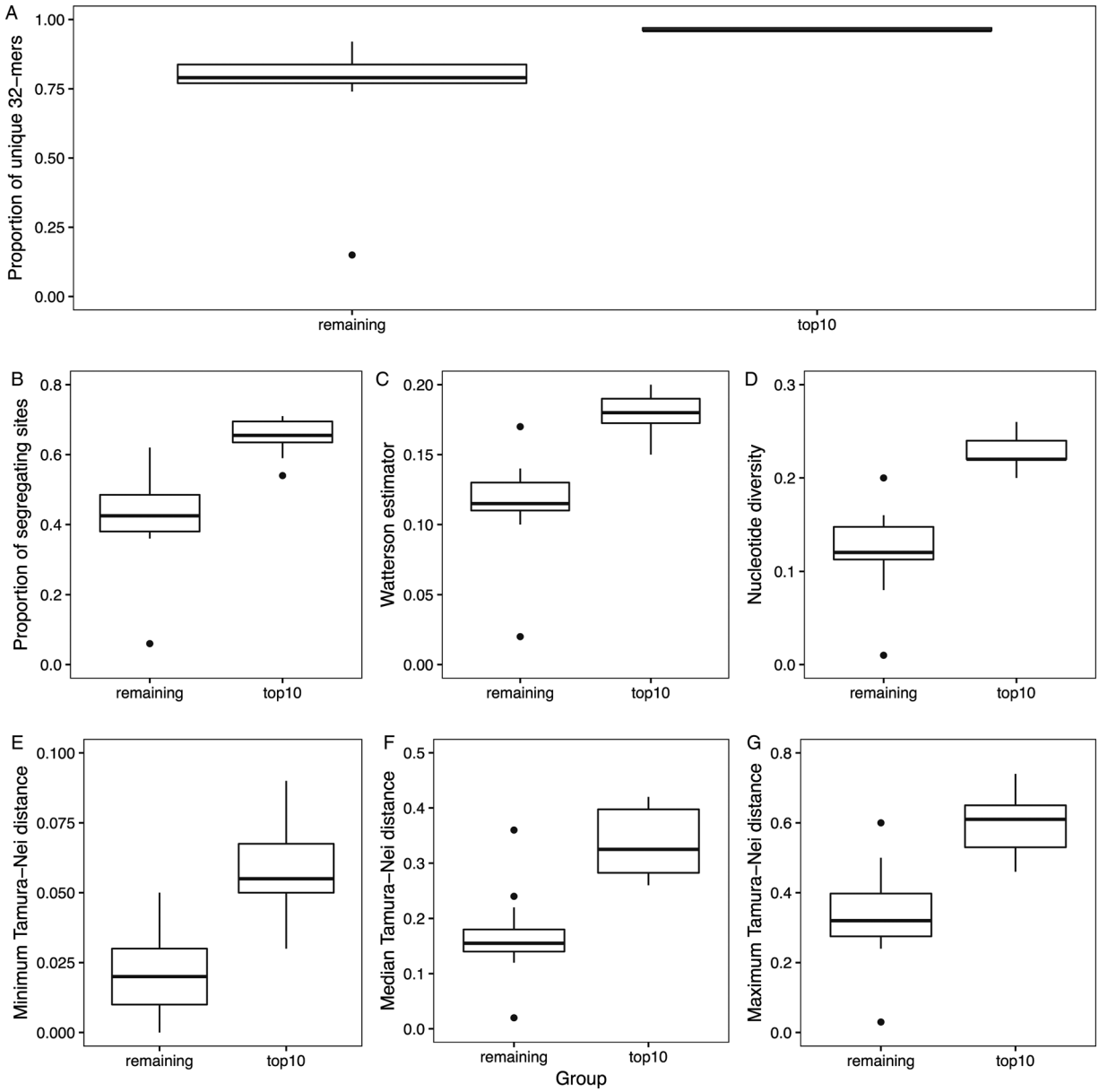

Fig. 3. Comparisons of the proportion of unique 32-mers (a), proportion of segregating sites (b), Watterson's estimator (c), nucleotide diversity (d), Tamura-Nei sequence distance (minimum, median and maximum); (e–f) across 28 genetic loci occurring in all 22 *Bartonella* genomes currently available. Separate boxes in each plot represent the top 10 loci as ranked by the proportion of unique 32-mers ('top 10') and the other 18 loci ('remaining').

exceeding 1%, which should be sufficient for identifying two genotypes as distinct in a phylogenetic analysis, particularly when used together in MLST analyses.

We were able to identify the presence of candidate loci that can discriminate among closely related *Bartonella* species better than the commonly used markers. We suggest that these loci may be useful for the characterization of *Bartonella* species and assessment of phylogenetic relationships; however, there currently exist no known primers for amplifying these loci by conventional PCR. *Bartonella* species have high nucleotide diversity across their genomes (Fig. 2a) with few highly conserved regions, especially in highly diverse genes. Therefore, primer design is a very challenging problem, especially the design of universal primers capable of binding to all possible species (L. Albayrak and C. McKee, unpublished data). There is likely a tradeoff in between phylogenetic resolution of genes and the ability to design universal primers, so the clustering of the eight genetic loci commonly used for genotyping (ITS, *ftsZ*, *gltA*, *groEL*, *gyrB*, *nuoG*, *ribC* and *rpoB*) between the most diverse genes and the least diverse genes (including 16S rRNA) may be a function of this tradeoff. The other potential disadvantage of using any of the top 10 most diverse loci for genotyping bartonellae is that these genes have only been sequenced for 22 *Bartonella* species. There are an

https://doi.org/10.1017/S0031182017001263 Published online by Cambridge University Press

enormous number of potentially new *Bartonella* species and genotypes that have been characterized by only ITS and/or *gltA* sequences that considerably expand our knowledge of *Bartonella* diversity; thus, there is an advantage to continued usage of these markers to facilitate comparative ecological analyses. Switching to different markers would inevitably ignore this diversity and would require considerable time and effort to restore.

DISCUSSION

The practicality of single-locus barcoding of bacteria

The utility of DNA barcoding for animal species is partly due to special features of the genetic markers it targets. The commonly targeted cytochrome *c* oxidase I (COI) gene for barcoding animals is one of the conserved oxidative phosphorylation subunits of the mitochondrial genome. Mitochondria are passed solely from the female parent to offspring in animals, so individuals are typically haploid at all mitochondrial loci and no recombination occurs in the mitochondrial genome. Additionally, there are sufficiently conserved portions of the COI gene that nearly universal PCR primers have been developed (Hajibabaei *et al*. 2006). Since bacterial genomes are haploid, it may be tempting to simply find any genetic marker that has better phylogenetic resolution than 16S rRNA and use it for DNA barcoding of bacteria. There are four primary problems with this approach: (1) increasing sequence diversity in a gene diminishes the ability to design conserved primers that can be utilized across a broad taxonomic diversity of bacteria; (2) homologous recombination is widespread (Vos and Didelot, 2008) and can obscure phylogenetic inference (Fraser *et al*. 2007) and estimates of species diversity if only one marker is used; (3) bacterial genomes are very flexible in gene content, even for closely related species (Konstantinidis *et al*. 2006), so a locus may not exist in all species surveyed; and (4) multiple species may be present in the sample, but may not be detectable due to variation in abundance or primer amplification bias. Thus, we argue that for bacterial species, and in particular *Bartonella*, there is probably no perfect analogue to single-locus DNA barcoding that could be used for all sample types.

There are existing methods, particularly MLST (Stackebrandt *et al*. 2002), which could balance the tradeoffs of culturing bias, phylogenetic resolution, homologous recombination and gene conservation across species. We believe that MLST of housekeeping genes (i.e. genes under stabilizing selection encoding metabolic functions) remains a powerful technique that can be used on uncultured bacteria to detect evidence of mixed infections and/or homologous recombination, provide sufficient phylogenetic resolution for the delineation of bacterial species and will provide consistency in the usage of genetic loci that can facilitate a global assessments of parasitic bacterial diversity. This approach can be appropriately modified for the detection and characterization of parasitic bacteria directly from extracted DNA from a range of sample types, with the caveat that culturing should be attempted if feasible since it is the best way to fully characterize a novel bacterial species. In the following section, we will make specific recommendations for the detection and genotyping *Bartonella* in collected samples (Box 1), but we recognize that with some modifications to collection of appropriate samples and the molecular protocols, this approach is likely generalizable to a variety of bacterial taxa.

Recommendations for the genotyping of Bartonella *species*

The first step to successful detection and potential isolation of bartonellae is to collect appropriate tissues and store them properly [Box 1(a)]. Gutiérrez *et al*. (2017) recommend the collection of whole blood due to the haemotrophic nature of these bacteria using appropriate sterility requirements, especially if culturing is to be attempted. If animals are sacrificed and organs pulled out, spleen is probably the most valuable organ for *Bartonella* detection, although as we have reviewed above, liver, heart, kidney and lung may show evidence of infection, and *Bartonella* species may vary in abundance across these tissue types within individuals. All tissue samples should be transported at low temperature and stored at −20 or −80 °C if not processed immediately. For ectoparasite samples, storage in 70% ethanol at room temperature is convenient and suitable for molecular detection; however, if culturing is planned, then live specimens are preferred and additional surface sterilization protocols will be required (Gutiérrez *et al*. 2017). During the process of collecting and analysing specimens, we recommend that investigators attempt to identify all animals as close as possible to the species level by morphological traits [Box 1(a)]. When morphological identification is not feasible (e.g. when cryptic species of rats and ectoparasites are morphologically undistinguished or when accurate records do not exist), then DNA barcoding of host samples (tissues or whole ectoparasites) at mitochondrial loci (e.g. COI) can be incorporated into molecular analyses (Hajibabaei *et al*. 2006). These data are valuable for understanding the ecology and evolution of *Bartonella* species, particularly for understanding the host range and specificity, vector potential and evolutionary codivergence of parasites with their hosts and vectors.

We recommend that investigators use homogenization techniques appropriate for particular tissue

https://doi.org/10.1017/S0031182017001263 Published online by Cambridge University Press

Box 1. Recommendations for consistent approaches to genotyping bartonellae

(a) Target multiple animal tissues (e.g. blood, spleen, liver and/or heart) and ectoparasites for detection. Identify vertebrate and arthropod hosts to the species level where possible using morphological traits and/or barcoding of mtDNA to facilitate ecological analyses.

(b) Use homogenization and DNA extraction protocols that maximize yield while reducing the presence of PCR inhibitors. Extended lysis or pre-enrichment culture steps may be needed for some samples. After homogenization of samples but before DNA extraction, retain some samples if culturing will be attempted.

(c) When possible, attempt to culture isolates, especially when sequence data indicate the presence of novel species or genotypes. Genotyping should be regarded as just the first step towards the description of *Bartonella* species, with additional trait and genomic data providing valuable information for species descriptions.

(d) Screen samples by ITS or real-time PCR (ITS, rpoB or ssrA) or alternatively, 16S metagenome or transcriptome sequencing (if available) followed by sequencing of multiple house-keeping genes.

(e) At the very least, sequence gltA to facilitate comparison with other studies.

(f) Sequence at least one additional marker to confirm the species identity based on gltA. Conflicting identifications may indicate the presence of multiple infections or recombinant genotypes. More markers provide more robust results, but at least three is recommended.

(g) Additional targets can vary in detection success, but rpoB, ftsZ, groEL, ribC, nuoG and ssrA are popular (in order of frequency used). Nested PCR reactions can increase sensitivity of these markers to be more comparable with ITS results.

(h) Attempt to identify the phylogenetic lineage (Harms and Dehio, 2012) or associated *Bartonella* species complex (Kosoy *et al*. 2012) based on sequence data for any novel genotypes.

or ectoparasite specimens, and follow extraction protocols that maximize DNA yield and quality while minimizing the presence of PCR inhibitors that may be present in the specimens [Box 1(b)]. Some specimens may benefit from pre-enrichment in liquid growth medium before extraction (Maggi *et al*. 2005; Duncan *et al*. 2007; Riess *et al*. 2008; Bai *et al*. 2010) or extended lysis steps during the extraction process. Gutiérrez *et al*. (2017) provide an excellent review of recommended protocols for homogenization and extraction. In cases where culturing might be attempted, we recommend retaining samples of homogenate (either used immediately or frozen at −20 °C or below). As we have advocated above, culturing is vital for the complete description of bacterial species and should be attempted in all studies where appropriate samples are available, especially if sequence data indicate the presence of novel *Bartonella* species or genotypes [Box 1(c)]. Culturing can provide information on valuable traits such as *in vitro* growth rate, bacterial morphology, presence of multiple coinfecting bartonellae, biochemical profiles, etc. Additional genomic (e.g. MLST or whole genome sequencing) analyses that clarify evolutionary histories can be facilitated, if *Bartonella* genotypes are isolated.

For direct detection of *Bartonella* DNA from extracted DNA, there are several options for markers that appear to have good sensitivity. Conventional PCR targeting the 16S–23S intergenic spacer region (ITS) or real-time PCR targeting various loci (e.g. ITS, *ssrA*, *rpoB*) are amenable for screening many samples for potential positives [Box 1(d)]. Primers and protocols for these approaches are published and are reviewed in Gutiérrez *et al*. (2017). We caution against reporting results from real-time PCR assays or conventional PCR in the absence of sequencing, since not all primers are entirely specific for *Bartonella* DNA (Maggi and Breitschwerdt, 2005; Colborn *et al*. 2010; Diaz *et al*. 2012) and may amplify host DNA, leading to false positives. Usage of next-generation sequencing approaches (e.g. 16S metagenomics, transcriptomics) are also valuable at this stage, especially if investigators are interested in describing the microbiome of particular tissues or investigating a broad range of pathogenic bacterial taxa [Box 1(d)]. However, as we reviewed above, the phylogenetic resolution of 16S sequences is limited, so additional genomic loci will need to be sequenced to confirm the species identity of targeted bacterial taxa.

For accurate genotyping, there are a variety of markers that have good phylogenetic resolution (La Scola *et al*. 2003) and validated primer sets (Gutiérrez *et al*. 2017). Based on our review of the literature, *gltA* is the most widely used marker and has the most extensive database of sequences available on GenBank. Therefore, we advocate for all studies to sequence this marker, at least from any novel genotypes or species, to support comparisons of *Bartonella* diversity across studies [Box 1(e)].

https://doi.org/10.1017/S0031182017001263 Published online by Cambridge University Press

Nested PCR reactions (Bai *et al*. 2016) that combine published primers (Norman *et al*. 1995; Birtles and Raoult, 1996) can increase the sensitivity of this marker.

In addition to *gltA*, we feel it is important to acquire additional sequence data from other loci to confirm the species identification by *gltA* [Box 1 (f)]. Phylogenetic concordance among loci may be sufficient to describe novel, candidate *Bartonella* species (Lilley *et al*. 2015; Martin-Alonso *et al*. 2016), which can be further described after culturing isolates. Conflicts among multiple loci may indicate the presence of multiple infections or recombinant infections – molecular cloning or examination of multiple peaks in chromatograms may be able to distinguish these scenarios. In general, three loci should be sufficient to accurately genotype *Bartonella* if a single infection is present or detect mixed infections; however, more loci may produce more robust results or distinguish *Bartonella* genotypes that may be closely related to known species. The number of markers to use will depend on the phylogenetic resolution needed (most MLST studies require only 5–9 markers), and more markers, especially uncommon markers, may have limited usefulness if not repeated in other laboratories. Additional targets can vary in sensitivity and amplification bias for particular species [Box 1(g)], but *rpoB*, *ftsZ*, *groEL*, *ribC*, *nuoG* and *ssrA* are popular (in order of frequency used based on our literature review) and have published primers and amplification protocols (Gutiérrez *et al*. 2017), including some nested protocols (Zeaiter *et al*. 2002; Colborn *et al*. 2010; Bai *et al*. 2016). Our analysis of 22 *Bartonella* genomes indicates that there may be single loci that are capable of distinguishing among *Bartonella* species better than these popular markers alone, but primer design for these regions will be challenging due to their sequence diversity (i.e. with few conserved regions), and the utility of these sequences will lag behind these popular markers unless many laboratories adopt them. Researchers should experiment with multiple markers and modified protocols to find the best ones for their sample type, but some consistency among sequenced markers across studies will facilitate comparative studies of *Bartonella* prevalence and diversity across systems.

Finally, our understanding of *Bartonella* ecology and evolution would benefit from the increased description of *Bartonella* phylogenetic lineages (Harms and Dehio, 2012) and species complexes (Kosoy *et al*. 2012), especially for novel genotypes [Box 1(h)]. Species complexes can include clusters genetically similar species found in a group of related hosts, such as *B. elizabethae*, *Bartonella queenslandensis*, *Bartonella rattimassiliensis* and *B. tribocorum* associated with murine rodents. These species complexes can indicate the presence of evolutionary radiations through codivergence and speciation within related hosts, providing information about the biological niche of these *Bartonella* species. Multiple species complexes linked deeper in evolutionary time may form well-supported clades or lineages that illuminate the longer term diversification processes of this diverse genus. Researchers describing new *Bartonella* species or genotypes could increase the impact of their findings by making these substantial evolutionary and ecological connections, particularly when clinical cases of bartonellosis can be traced to a potential zoonotic origin.

Concluding remarks

Using *Bartonella* bacteria as examples, we have highlighted the substantial challenges that exist in the accurate genotyping of bacteria from environmental samples. Issues related to isolation of cultures, homologous recombination, coinfections, sensitivity and phylogenetic resolution of molecular markers, variation in detection across different tissues, and variation in marker usage across studies are certainly not restricted to studies of *Bartonella*. For environmental samples stored in laboratories and museums around the world, our recommendations for sensitive detection assays (including real-time PCR or high-throughput metagenomics), followed by conventional PCR amplification and sequencing of multiple house-keeping genes are surely applicable to a wide array of other zoonotic bacteria. These may include many proteobacteria (e.g. *Anaplasma*, *Bordatella*, *Brucella*, *Burkholderia*, *Campylobacter*, *Coxiella*, *Ehrlichia*, *Yersinia*, *Francisella*, *Helicobacter*, *Legionella*, *Neorickettsia*, *Orientia*, *Pasteurella*, *Pseudomonas*, *Rickettsia* and *Wolbachia*), spirochetes (e.g. *Borrelia*, *Leptospira*, *Treponema*) and other bacteria (e.g. *Chlamydia*, *Listeria*, *Mycobacterium*, *Mycoplasma*). Although recommendations for the collection of appropriate animal samples and molecular markers used for detection and characterization will vary according to each bacteria (and likely requires standardization as we observed with *Bartonella*), the general process of direct detection from extracted DNA, multi-locus sequencing and subsequent attempts to culture (followed by additional genomic or biochemical characterization) is generalizable. Databases already exist, e.g. http://www.mlst.net/databases/default.asp and https://pubmlst.org/databases/), that contain primers and protocols for multi-locus sequencing approaches for many of the zoonotic bacteria listed above. Standardized genotyping approaches will greatly expand our knowledge of the phylogenetic diversity and ecology of parasitic bacteria infecting animals, and help to measure and mitigate the risks posed by these bacteria to public health.

https://doi.org/10.1017/S0031182017001263 Published online by Cambridge University Press

ACKNOWLEDGMENTS

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of CDC.

FINANCIAL SUPPORT

Numerous investigations of zoonotic bacterial agents around the globe were supported by the CDC' Global Disease Detection program.

REFERENCES

André, A., Mouton, A., Millien, V. and Michaux, J. (2017). Liver microbiome of *Peromyscus leucopus*, a key reservoir host species for emerging infectious diseases in North America. *Infection, Genetics and Evolution* **52**, 10–18.

Auch, A. F., von Jan, M., Klenk, H.-P. and Göker, M. (2010). Digital DNA–DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards in Genomic Sciences* **2**, 117–134.

Bai, Y., Kosoy, M. Y., Calisher, C. H., Cully, J. F. and Collinge, S. K. (2009). Effects of rodent community diversity and composition on prevalence of an endemic bacterial pathogen – *Bartonella*. *Biodiversity* **10**, 3–11.

Bai, Y., Kosoy, M. Y., Boonmar, S., Sawatwong, P., Sangmaneedet, S. and Peruski, L. F. (2010). Enrichment culture and molecular identification of diverse *Bartonella* species in stray dogs. *Veterinary Microbiology* **146**, 314–319.

Bai, Y., Kosoy, M. Y., Diaz, M. H., Winchell, J., Baggett, H., Maloney, S. A., Boonmar, S., Bhengsri, S., Sawatwong, P. and Peruski, L. F. (2012). *Bartonella vinsonii* subsp. *arupensis* in humans, Thailand. *Emerging Infectious Diseases* **18**, 989–991.

Bai, Y., Hayman, D. T. S., McKee, C. D. and Kosoy, M. Y. (2015a). Classification of *Bartonella* strains associated with straw-colored fruit bats (*Eidolon helvum*) across Africa using a multi-locus sequence typing platform. *PLoS Neglected Tropical Diseases* **9**, e0003478.

Bai, Y., Rizzo, M. F., Alvarez, D., Moran, D., Peruski, L. F. and Kosoy, M. (2015b). Coexistence of *Bartonella henselae* and *B. clarridgeiae* in populations of cats and their fleas in Guatemala. *Journal of Vector Ecology* **40**, 327–332.

Bai, Y., Gilbert, A., Fox, K., Osikowicz, L. and Kosoy, M. (2016). *Bartonella rochalimae* and *B. vinsonii* subsp. *berkhoffii* in wild carnivores from Colorado, USA. *Journal of Wildlife Diseases* **52**, 844–849.

Banskar, S., Bhute, S. S., Suryavanshi, M. V., Punekar, S. and Shouche, Y. S. (2016). Microbiome analysis reveals the abundance of bacterial pathogens in *Rousettus leschenaultii* guano. *Scientific Reports* **6**, 36948.

Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P. and Kauserud, H. (2010). ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiology* **10**, 189.

Berglund, E. C., Ellegaard, K., Granberg, F., Xie, Z., Maruyama, S., Kosoy, M. Y., Birtles, R. J. and Andersson, S. G. E. (2010). Rapid diversification by recombination in *Bartonella grahamii* from wild rodents in Asia contrasts with low levels of genomic divergence in Northern Europe and America. *Molecular Ecology* **19**, 2241–2255.

Billeter, S. A., Levy, M. G., Chomel, B. B. and Breitschwerdt, E. B. (2008). Vector transmission of *Bartonella* species with emphasis on the potential for tick transmission. *Medical and Veterinary Entomology* **22**, 1–15.

Billeter, S. A., Cáceres, A. G., Gonzales-Hidalgo, J., Luna-Caypo, D. and Kosoy, M. Y. (2011). Molecular detection of *Bartonella* species in ticks from Peru. *Journal of Medical Entomology* **48**, 1257–1260.

Birtles, R. J. and Raoult, D. (1996). Comparison of partial citrate synthase gene (gltA) sequences for phylogenetic analysis of *Bartonella* species. *International Journal of Systematic Bacteriology* **46**, 891–897.

Birtles, R. J., Hazel, S., Bown, K., Raoult, D., Begon, M. and Bennett, M. (2000). Subtyping of uncultured bartonellae using sequence comparison of 16 S/23 S rRNA intergenic spacer regions amplified directly from infected blood. *Molecular and Cellular Probes* **14**, 79–87.

Bittar, F., Keita, M. B., Lagier, J.-C., Peeters, M., Delaporte, E. and Raoult, D. (2014). *Gorilla gorilla gorilla* gut: a potential reservoir of pathogenic bacteria as revealed using culturomics and molecular tools. *Scientific Reports* **4**, 7174.

Blaxter, M. (2016). Imagining Sisyphus happy: DNA barcoding and the unnamed majority. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20150329.

Bonilla, D. L., Kabeya, H., Henn, J., Kramer, V. L. and Kosoy, M. Y. (2009). *Bartonella quintana* in body lice and head lice from homeless persons, San Francisco, California, USA. *Emerging Infectious Diseases* **15**, 912–915.

Bonilla, D. L., Cole-Porse, C., Kjemtrup, A., Osikowicz, L. and Kosoy, M. (2014). Risk factors for human lice and bartonellosis among the homeless, San Francisco, California, USA. *Emerging Infectious Diseases* **20**, 1645–1651.

Braga, M. do S. C. de O., Diniz, P. P. V. de P., André, M. R., de Bortoli, C. P. and Machado, R. Z. (2012). Molecular characterisation of *Bartonella* species in cats from São Luís, state of Maranhão, north-eastern Brazil. *Memórias do Instituto Oswaldo Cruz* **107**, 772–777.

Brook, C. E., Bai, Y., Dobson, A. P., Osikowicz, L. M., Ranaivoson, H. C., Zhu, Q., Kosoy, M. Y. and Dittmar, K. (2015). *Bartonella* spp. in fruit bats and blood-feeding ectoparasites in Madagascar. *PLoS Neglected Tropical Diseases* **9**, e0003532.

Brook, C. E., Bai, Y., Yu, E. O., Ranaivoson, H. C., Shin, H., Dobson, A. P., Metcalf, C. J. E., Kosoy, M. Y. and Dittmar, K. (2017). Elucidating transmission dynamics and host-parasite-vector relationships for rodent-borne *Bartonella* spp. in Madagascar. *Epidemics* In press.

Buffet, J.-P., Pisanu, B., Brisse, S., Roussel, S., Félix, B., Halos, L., Chapuis, J.-L. and Vayssier-Taussat, M. (2013a). Deciphering *Bartonella* diversity, recombination, and host specificity in a rodent community. *PLoS ONE* **8**, e68956.

Buffet, J.-P., Kosoy, M. Y. and Vayssier-Taussat, M. (2013b). Natural history of *Bartonella*-infecting rodents in light of new knowledge on genomics, diversity and evolution. *Future Microbiology* **8**, 1117–1128.

Carrasco, S. E., Chomel, B. B., Gill, V. A., Kasten, R. W., Maggi, R. G., Breitschwerdt, E. B., Byrne, B. A., Burek-Huntington, K. A., Miller, M. A., Goldstein, T. and Mazet, J. A. K. (2014). Novel *Bartonella* infection in northern and southern sea otters (*Enhydra lutris kenyoni* and *Enhydra lutris nereis*). *Veterinary Microbiology* **170**, 325–334.

Chaloner, G. L., Palmira Ventosilla and Birtles, R. J. (2011). Multi-locus sequence analysis reveals profound genetic diversity among isolates of the human pathogen Bartonella bacilliformis. PLoS Neglected Tropical Diseases 5, e1248.

Colborn, J. M., Kosoy, M. Y., Motin, V. L., Telepnev, M. V., Valbuena, G., Myint, K. S. A., Fofanov, Y., Putonti, C., Feng, C. and Peruski, L. F. (2010). Improved detection of *Bartonella* DNA in mammalian hosts and arthropod vectors by real-time PCR using the NADH dehydrogenase gamma subunit (nuoG). *Journal of Clinical Microbiology* **48**, 4630–4633.

Curtis, T. P., Sloan, W. T. and Scannell, J. W. (2002). Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10494–10499.

DeLong, E. F. and Pace, N. R. (2001). Environmental diversity of Bacteria and Archaea. *Systematic Biology* **50**, 470–478.

De Sousa, R., Edouard-Fournier, P., Santos-Silva, M., Amaro, F., Bacellar, F. and Raoult, D. (2006). Molecular detection of *Rickettsia felis*, *Rickettsia typhi* and two genotypes closely related to *Bartonella elizabethae*. *American Journal of Tropical Medicine and Hygiene* **75**, 727–731.

Diaz, M. H., Bai, Y., Malania, L., Winchell, J. M. and Kosoy, M. Y. (2012). Development of a novel genus-specific real-time PCR assay for detection and differentiation of *Bartonella* species and genotypes. *Journal of Clinical Microbiology* **50**, 1645–1649.

Dietrich, M., Kearney, T., Seamark, E. C. J. and Markotter, W. (2017). The excreted microbiota of bats: evidence of niche specialisation based on multiple body habitats. *FEMS Microbiology Letters* **364**, fnw284.

Duncan, A. W., Maggi, R. G. and Breitschwerdt, E. B. (2007). A combined approach for the enhanced detection and isolation of *Bartonella* species in dog blood samples: pre-enrichment liquid culture followed by PCR and subculture onto agar plates. *Journal of Microbiological Methods* **69**, 273–281.

Dykhuizen, D. (2005). Species numbers in bacteria. *Proceedings of the California Academy of Sciences* **56**, 62–71.

Fraser, C., Hanage, W. P. and Spratt, B. G. (2007). Recombination and the nature of bacterial speciation. *Science* **315**, 476–480.

Galan, M., Razzauti, M., Bard, E., Bernard, M., Brouat, C., Charbonnel, N., Dehne-Garcia, A., Loiseau, A., Tatard, C., Tamisier, L., Vayssier-Taussat, M., Vignes, H. and Cosson, J.-F. (2016). 16S rRNA amplicon sequencing for epidemiological surveys of bacteria in wildlife. *mSystems* **1**, e00032–16.

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P. and Tiedje, J. M. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology* **57**, 81–91.

Gundi, V. A. K. B., Kosoy, M. Y., Myint, K. S., Shrestha, S. K., Shrestha, M. P., Pavlin, J. A. and Gibbons, R. V. (2010). Prevalence and genetic diversity of *Bartonella* species detected in different tissues of small mammals in Nepal. *Applied and Environmental Microbiology* **76**, 8247–8254.

Gundi, V. A. K. B., Billeter, S. A., Rood, M. P. and Kosoy, M. Y. (2012a). *Bartonella* spp. in rats and zoonoses, Los Angeles, California, USA. *Emerging Infectious Diseases* **18**, 631–633.

Gundi, V. A. K. B., Kosoy, M. Y., Makundi, R. H. and Laudisoit, A. (2012b). Identification of diverse *Bartonella* genotypes among small mammals from Democratic Republic of Congo and Tanzania. *American Journal of Tropical Medicine and Hygiene* **87**, 319–326.

Gutiérrez, R., Cohen, L., Morick, D., Mumcuoglu, K. Y., Harrus, S. and Gottlieb, Y. (2014). Identification of different *Bartonella* species in the cattle tail louse (*Haematopinus quadripertusus*) and in cattle blood. *Applied and Environmental Microbiology* **80**, 5477–5483.

Gutiérrez, R., Nachum-Biala, Y. and Harrus, S. (2015). Relationship between the presence of *Bartonella* species and bacterial loads in cats and cat fleas (*Ctenocephalides felis*) under natural conditions. *Applied and Environmental Microbiology* **81**, 5613–5621.

Gutiérrez, R., Vayssier-Taussat, M., Buffet, J.-P. and Harrus, S. (2017). Guidelines for the isolation, molecular detection, and characterization of *Bartonella* species. *Vector-Borne and Zoonotic Diseases* **17**, 42–50.

Guy, L., Nystedt, B., Sun, Y., Näslund, K., Berglund, E. C. and Andersson, S. G. E. (2012). A genome-wide study of recombination rate variation in *Bartonella henselae*. *BMC Evolutionary Biology* **12**, 65.

Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W. and Hebert, P. D. N. (2006). DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 968–971.

Hanage, W. P., Spratt, B. G., Turner, K. M. and Fraser, C. (2006). Modelling bacterial speciation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **361**, 2039–2044.

Harms, A. and Dehio, C. (2012). Intruders below the radar: molecular pathogenesis of *Bartonella* spp. *Clinical Microbiology Reviews* **25**, 42–78.

Harrus, S., Bar-Gal, G. K., Golan, A., Elazari-Volcani, R., Kosoy, M. Y., Morick, D., Avidor, B. and Baneth, G. (2009). Isolation and genetic characterization of a *Bartonella* strain closely related to *Bartonella tribocorum* and *Bartonella elizabethae* in Israeli commensal rats. *American Journal of Tropical Medicine and Hygiene* **81**, 55–58.

Hebert, P. D. N., Cywinska, A., Ball, S. L. and DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**, 313–321.

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C. N., Hernsdorf, A. W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D. A., Finstad, K. M., Amundson, R., Thomas, B. C. and Banfield, J. F. (2016). A new view of the tree of life. *Nature Microbiology* **1**, 1–6.

Kabeya, H., Colborn, J. M., Bai, Y., Lerdthusnee, K., Richardson, J. H., Maruyama, S. and Kosoy, M. Y. (2010). Detection of *Bartonella tamiae* DNA in ectoparasites from rodents in Thailand and their sequence similarity with bacterial cultures from Thai patients. Vector Borne and Zoonotic Diseases (Larchmont, N.Y.) **10**, 429–434.

Kamani, J., Baneth, G., Mitchell, M., Mumcuoglu, K. Y., Gutiérrez, R. and Harrus, S. (2014). *Bartonella* species in bats (Chiroptera) and bat flies (Nycteribiidae) from Nigeria, West Africa. *Vector-Borne and Zoonotic Diseases* **14**, 625–632.

Kandelaki, G., Malania, L., Bai, Y., Chakvetadze, N., Katsitadze, G., Imnadze, P., Nelson, C., Harrus, S. and Kosoy, M. (2016). Human lymphadenopathy caused by ratborne *Bartonella*, Tbilisi, Georgia. *Emerging Infectious Diseases* **22**, 544–546.

Keim, P., Pearson, T. and Okinaka, R. (2008). Microbial forensics: DNA fingerprinting of *Bacillus anthracis* (Anthrax). *Analytical Chemistry* **80**, 4791–4800.

Klangthong, K., Promsthaporn, S., Leepitakrat, S., Schuster, A. L., McCardle, P. W., Kosoy, M. and Takhampunya, R. (2015). The distribution and diversity of *Bartonella* species in rodents and their ectoparasites across Thailand. *PLoS ONE* **10**, e0140856.

Knap, N., Duh, D., Birtles, R., Trilar, T., Petrovec, M. and Avšič-Županc, T. (2007). Molecular detection of *Bartonella* species infecting rodents in Slovenia. *FEMS Immunology & Medical Microbiology* **50**, 45–50.

Ko, S., Kim, S., Kang, J., Won, S., Lee, H., Shin, N., Choi, K., Youn, H. and Chae, J. (2013). Molecular detection of *Bartonella grahamii* and *B. schoenbuchensis*-related species in Korean water deer (*Hydropotes inermis argyropus*). *Vector-Borne and Zoonotic Diseases* **13**, 415–418.

Ko, S., Kang, J.-G., Kim, H.-C., Klein, T. A., Choi, K.-S., Song, J.-W., Youn, H.-Y. and Chae, J. S. (2016). Prevalence, isolation and molecular characterization of *Bartonella* species in Republic of Korea. *Transboundary and Emerging Diseases* **63**, 56–67.

Konstantinidis, K. T. and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 2567–2572.

Konstantinidis, K. T., Ramette, A. and Tiedje, J. M. (2006). The bacterial species definition in the genomic era. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **361**, 1929–1940.

Koskela, K. A., Kalin-Mänttäri, L., Hemmilä, H., Smura, T., Kinnunen, P. M., Niemimaa, J., Henttonen, H. and Nikkari, S. (2017). Metagenomic evaluation of bacteria from voles. *Vector-Borne and Zoonotic Diseases* **17**, 123–133.

Kosoy, M. Y., Murray, M., Gilmore, R. D., Jr., Bai, Y. and Gage, K. L. (2003). *Bartonella* strains from ground squirrels are identical to *Bartonella washoensis* isolated from a human patient. *Journal of Clinical Microbiology* **41**, 645–650.

Kosoy, M. Y., Hayman, D. T. S. and Chan, K.-S. (2012). *Bartonella* bacteria in nature: where does population variability end and a species start? *Infection, Genetics and Evolution* **12**, 894–904.

Kosoy, M., Bai, Y., Enscore, R., Rizzo, M. R., Bender, S., Popov, V., Albayrak, L., Fofanov, Y. and Chomel, B. (2016). *Bartonella melophagi* in blood of domestic sheep (*Ovis aries*) and sheep keds (*Melophagus ovinus*) from the southwestern US: cultures, genetic characterization, and ecological connections. *Veterinary Microbiology* **190**, 43–49.

Kumar, S., Stecher, G. and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**, 1870–1874.

Kyrpides, N. C., Hugenholtz, P., Eisen, J. A., Woyke, T., Göker, M., Parker, C. T., Amann, R., Beck, B. J., Chain, P. S. G., Chun, J., Colwell, R. R., Danchin, A., Dawyndt, P., Dedeurwaerdere, T., DeLong, E. F., Detter, J. C., De Vos, P., Donohue, T. J., Dong, X.-Z., Ehrlich, D. S., Fraser, C., Gibbs, R., Gilbert, J., Gilna, P., Glöckner, F. O., Jansson, J. K., Keasling, J. D., Knight, R., Labeda, D., Lapidus, A., *et al*. (2014). Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. *PLoS Biology* **12**, e1001920.

La Scola, B., Zeaiter, Z., Khamis, A. and Raoult, D. (2003). Gene-sequence-based criteria for species definition in bacteriology: the *Bartonella* paradigm. *Trends in Microbiology* **11**, 318–321.

Lebonah, D. E., Dileep, A., Chandrasekhar, K., Sreevani, S., Sreedevi, B. and Pramoda Kumari, J. (2014). DNA barcoding on bacteria: a review. *Advances in Biology* **2014**, 541787.

Lei, B. R. and Olival, K. J. (2014). Contrasting patterns in mammal–bacteria coevolution: *Bartonella* and *Leptospira* in bats and rodents. *PLoS Neglected Tropical Diseases* **8**, e2738.

Lilley, T. M., Veikkolainen, V. and Pulliainen, A. T. (2015). Molecular detection of *Candidatus Bartonella hemsundetiensis* in bats. *Vector-Borne and Zoonotic Diseases* **15**, 706–708.

Locey, K. J. and Lennon, J. T. (2016). Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 5970–5975.

Loftis, A. D., Reeves, W. K., Szumlas, D. E., Abbassy, M. M., Helmy, I. M., Moriarity, J. R. and Dasch, G. A. (2006). Surveillance of Egyptian fleas for agents of public health significance: *Anaplasma*, *Bartonella*, *Coxiella*, *Ehrlichia*, *Rickettsia*, and *Yersinia pestis*. *American Journal of Tropical Medicine and Hygiene* **75**, 41–48.

Logares, R., Sunagawa, S., Salazar, G., Cornejo-Castillo, F. M., Ferrera, I., Sarmento, H., Hingamp, P., Ogata, H., de Vargas, C., Lima-Mendez, G., Raes, J., Poulain, J., Jaillon, O., Wincker, P., Kandels-Lewis, S., Karsenti, E., Bork, P. and Acinas, S. G. (2014). Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environmental Microbiology* **16**, 2659–2671.

Maggi, R. G. and Breitschwerdt, E. B. (2005). Potential limitations of the 16S-23S rRNA intergenic region for molecular detection of *Bartonella* species. *Journal of Clinical Microbiology* **43**, 1171–1176.

Maggi, R. G., Duncan, A. W. and Breitschwerdt, E. B. (2005). Novel chemically modified liquid medium that will support the growth of seven *Bartonella* species. *Journal of Clinical Microbiology* **43**, 2651–2655.

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R., Jarrin, C., Chardon, P., Marteau, P., Roca, J. and Dore, J. (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55**, 205–211.

Martin-Alonso, A., Houemenou, G., Abreu-Yanes, E., Valladares, B., Feliu, C. and Foronda, P. (2016). *Bartonella* spp. in small mammals, Benin. *Vector-Borne and Zoonotic Diseases* **16**, 229–237.

Meier-Kolthoff, J. P., Hahnke, R. L., Petersen, J., Scheuner, C., Michael, V., Fiebig, A., Rohde, C., Rohde, M., Fartmann, B., Goodwin, L. A., Chertkov, O., Reddy, T. B. K., Pati, A., Ivanova, N. N., Markowitz, V., Kyrpides, N. C., Woyke, T., Göker, M. and Klenk, H.-P. (2014a). Complete genome sequence of DSM 30083T, the type strain (U5/41T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Standards in Genomic Sciences* **9**, 2.

Meier-Kolthoff, J. P., Klenk, H.-P. and Göker, M. (2014b). Taxonomic use of DNA G+C content and DNA–DNA hybridization in the genomic age. *International Journal of Systematic and Evolutionary Microbiology* **64**, 352–356.

Miceli, N. G., Gavioli, F. A., Goncalves, L. R., Andre, M. R., Sousa, V. R. F., de Sousa, K. C. M. and Machado, R. Z. (2013). Molecular detection of feline arthropod-borne pathogens in cats in Cuiaba, state of Mato Grosso, central-western region of Brazil. *Revista Brasileira de Parasitologia Veterinária* **22**, 385–390.

Miller, C. S., Baker, B. J., Thomas, B. C., Singer, S. W. and Banfield, J. F. (2011). EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biology* **12**, R44.

Morick, D., Osinga, N., Gruys, E. and Harrus, S. (2009). Identification of a *Bartonella* species in the harbor seal (*Phoca vitulina*) and in seal lice (*Echinophtirius horridus*). *Vector-Borne and Zoonotic Diseases* **9**, 751–753.

Morick, D., Krasnov, B. R., Khokhlova, I. S., Shenbrot, G. I., Kosoy, M. Y. and Harrus, S. (2010). *Bartonella* genotypes in fleas (Insecta: Siphonaptera) collected from rodents in the Negev Desert, Israel. *Applied and Environmental Microbiology* **76**, 6864–6869.

Norman, A. F., Regnery, R., Jameson, P., Greene, C. and Krause, D. C. (1995). Differentiation of *Bartonella*-like isolates at the species level by PCR-restriction fragment length polymorphism in the citrate synthase gene. *Journal of Clinical Microbiology* **33**, 1797–1803.

Osikowicz, L. M., Billeter, S. A., Rizzo, M. F., Rood, M. P., Freeman, A. N., Burns, J. E., Hu, R., Juieng, P., Loparev, V. and Kosoy, M. (2016). Distribution and diversity of *Bartonella washoensis* strains in ground squirrels from California and their potential link to human cases. *Vector-Borne and Zoonotic Diseases* **16**, 683–690.

Paziewska, A., Harris, P. D., Zwolińska, L., Bajer, A. and Siński, E. (2011). Recombination within and between species of the alpha proteobacterium *Bartonella* infecting rodents. *Microbial Ecology* **61**, 134–145.

Paziewska, A., Siński, E. and Harris, P. D. (2012). Recombination, diversity and allele sharing of infectivity proteins between *Bartonella* species from rodents. *Microbial Ecology* **64**, 525–536.

Pérez-Martínez, L., Venzal, J. M., González-Acuña, D., Portillo, A., Blanco, J. R. and Oteo, J. A. (2009). *Bartonella rochalimae* and other *Bartonella* spp. in Fleas, Chile. *Emerging Infectious Diseases* **15**, 1150–1152.

Qiu, Y., Nakao, R., Ohnuma, A., Kawamori, F. and Sugimoto, C. (2014). Microbial population analysis of the salivary glands of ticks; a possible strategy for the surveillance of bacterial pathogens. *PLoS ONE* **9**, e103961.

Razzauti, M., Galan, M., Bernard, M., Maman, S., Klopp, C., Charbonnel, N., Vayssier-Taussat, M., Eloit, M. and Cosson, J.-F. (2015). A comparison between transcriptome sequencing and 16S metagenomics for detection of bacterial pathogens in wildlife. *PLoS Neglected Tropical Diseases* **9**, e0003929.

Richter, M. and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 19126–19131.

Riess, T., Dietrich, F., Schmidt, K. V., Kaiser, P. O., Schwarz, H., Schafer, A. and Kempf, V. A. J. (2008). Analysis of a novel insect cell culture medium-based growth medium for *Bartonella* species. *Applied and Environmental Microbiology* **74**, 5224–5227.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B. K., Gies, E. A., Dodsworth, J. A., Hedlund, B. P., Tsiamis, G., Sievert, S. M., Liu, W.-T., Eisen, J. A., Hallam, S. J., Kyrpides, N. C., Stepanauskas, R., Rubin, E. M., Hugenholtz, P. and Woyke, T. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437.

Rojas, N., Troyo, A., Castillo, D., Gutierrez, R., Harrus, S., Gutiérrez, R. and Harrus, S. (2015). Molecular detection of *Bartonella* species in fleas collected from dogs and cats from Costa Rica. *Vector-Borne and Zoonotic Diseases* **15**, 630–632.

Roux, V. and Raoult, D. (1995). The 16S-23S rRNA intergenic spacer region of *Bartonella* (Rochalimaea) species is longer than usually described in other bacteria. *Gene* **156**, 107–111.

Saw, J. H., Spang, A., Zaremba-Niedzwiedzka, K., Juzokaite, L., Dodsworth, J. A., Murugapiran, S. K., Colman, D. R., Takacs-Vesbach, C., Hedlund, B. P., Guy, L. and Ettema, T. J. (2015). Exploring microbial dark matter to resolve the deep archaeal ancestry of eukaryotes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **370**, 20140328.

Schloss, P. D. and Handelsman, J. (2004). Status of the microbial census. *Microbiology and Molecular Biology Reviews* **68**, 686–691.

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., Chen, W. and Consortium, F. B. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 6241–6246.

Segers, F. H., Kešnerová, L., Kosoy, M. and Engel, P. (2017). Genomic changes associated with the evolutionary transition of an insect gut symbiont into a blood-borne pathogen. *The ISME Journal* **11**, 1232–1244.

Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A. D., Kämpfer, P., Maidem, M. C. J., Nesme, X., Rosselló-Morá, Swings, J., Trüper, H. G., Vauterin, L., Ward, A. C. and Whitman, W. B. (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology* **52**, 1043–1047.

Tindall, B. J., Rosselló-Móra, R., Busse, H.-J., Ludwig, W. and Kämpfer, P. (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *International Journal of Systematic and Evolutionary Microbiology* **60**, 249–266.

Tsai, Y.-L., Chang, C.-C., Chuang, S.-T. and Chomel, B. B. (2011). *Bartonella* species and their ectoparasites: selective host adaptation or strain selection between the vector and the mammalian host? *Comparative Immunology, Microbiology and Infectious Diseases* **34**, 299–314.

Urushadze, L., Bai, Y., Osikowicz, L. M., McKee, C. D., Kandaurov, A., Kuzmin, I. V., Sidamonidze, K., Putkaradze, D., Imnadze, P. and Kosoy, M. Y. (2017). Prevalence, diversity, and host associations of *Bartonella* strains in bats from Georgia (Caucasus). *PLoS Neglected Tropical Diseases* **11**, e0005428.

Veikkolainen, V., Vesterinen, E. J., Lilley, T. M. and Pulliainen, A. T. (2014). Bats as reservoir hosts of human bacterial pathogen, *Bartonella mayotimonensis*. *Emerging Infectious Diseases* **20**, 960–967.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.-H. and Smith, H. O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74.

Vos, M. (2009). Why do bacteria engage in homologous recombination? *Trends in Microbiology* **17**, 226–232.

Vos, M. and Didelot, X. (2008). A comparison of homologous recombination rates in bacteria and archaea. *The ISME Journal* **3**, 199–208.

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Moore, W. E. C., Murray, R. G. E., Stackebrandt, E., Starr, M. P. and Trüper, H. G. (1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *International Journal of Systematic and Evolutionary Microbiology* **37**, 463–464.

Welch, D. F., Carroll, K. C., Hofmeister, E. K., Persing, D. H., Robison, D. A., Steigerwalt, A. G. and Brenner, D. J. (1999). Isolation of a new subspecies, *Bartonella vinsonii* subsp. *arupensis*, from a cattle rancher: identity with isolates found in conjunction with *Borrelia burgdorferi* and *Babesia microti* among naturally infected mice. *Journal of Clinical Microbiology* **37**, 2598–2601.

Wilkinson, D. A., Duron, O., Cordonin, C., Gomard, Y., Ramasindrazana, B., Mavingui, P., Goodman, S. M. and Tortosa, P. (2016). The bacteriome of bat flies (Nycteribiidae) from the Malagasy region: a community shaped by host ecology, bacterial transmission mode, and host-vector specificity. *Applied and Environmental Microbiology* **82**, 1778–1788.

Woese, C. R. and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5088–5090.

Young, H. S., Dirzo, R., Helgen, K. M., McCauley, D. J., Billeter, S. A., Kosoy, M. Y., Osikowicz, L. M., Salkeld, D. J., Young, T. P. and Dittmar, K. (2014). Declines in large wildlife increase landscape-level prevalence of rodent-borne disease in Africa. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 7036–7041.

Zeaiter, Z., Liang, Z. and Raoult, D. (2002). Genetic classification and differentiation of *Bartonella* species based on comparison of partial ftsZ gene sequences. *Journal of Clinical Microbiology* **40**, 3641–3647.