Recombination Within and Between Species of the Alpha Proteobacterium *Bartonella* Infecting Rodents

Anna Paziewska · Philip D. Harris · Lucyna Zwolińska · Anna Bajer · Edward Siński

Abstract *Bartonella* infections from wild mice and voles (*Apodemus flavicollis*, *Mi. oeconomus*, *Microtus arvalis* and *Myodes glareolus*) were sampled from a forest and old-field habitats of eastern Poland; a complex network of *Bartonella* isolates, referrable to *B. taylorii*, *B. grahamii*, *B. birtlesii* and *B. doshiae*, was identified by the sequencing of a *gltA* fragment, comparable to previous studies of *Bartonella* diversity in rodents. Nested clade analysis showed that isolates could be assigned to zero- and one-step clades which correlated with host identity and were probably the result of clonal expansion; however, sequencing of other housekeeping genes (*rpoB*, *ribC*, *ftsZ*, *groEl*) and the *16S RNA* gene revealed a more complex situation with clear evidence of numerous recombinant events in which one or both *Bartonella* parents could be identified. Recombination within *gltA* was found to have generated two distinct variant clades, one a hybrid between *B. taylorii* and *B. doshiae*, the other between *B. taylorii* and *B. grahamii*. These recombinant events characterised the differences between the two-step and higher clades within the total nested cladogram, involved all four species of *Bartonella* identified in this work and appear to have played a dominant role in the evolution of *Bartonella* diversity. It is clear, therefore, that housekeeping gene phylogenies are not robust indicators of *Bartonella* diversity, especially when only a single gene (*gltA* or *16S RNA*) is used. *Bartonella* clades infecting *Microtus* were most frequently involved in recombination and were most frequently tip clades within the cladogram. The role of *Microtus* in influencing the frequency of *Bartonella* recombination remains unknown.

A. Paziewska (✉) · L. Zwolińska · A. Bajer · E. Siński
Department of Parasitology, Institute of Zoology,
Faculty of Biology, University of Warsaw,
Miecznikowa 1,
02-096 Warsaw, Poland
e-mail: apaziewska@biol.uw.edu.pl

P. D. Harris
National Centre for Biosystematics,
Natural History Museum, University of Oslo,
P.O. Box 1172, Blindern,
Oslo, Norway

Introduction

Bartonella is a parasite of mammalian erythrocytes and endothelial cells transmitted by blood feeding arthropod ectoparasites [5]. This genus has a well-established alpha taxonomy consisting of 27 species and three sub-species, and several taxa cause significant human disease: *B. bacilliformis*, responsible for Oroya fever; *B. quintana*, causing trench fever and *B. henselae*, causing cat-scratch disease are the most important, but others are recorded as adventitious zoonotic infections, giving the genus the reputation as an emerging human pathogen [6, 7]. Since La Scola et al. [28] suggested that species could be described on the basis of DNA sequences from housekeeping genes, of which RNA polymerase (*rpoB*) and citrate synthase (*gltA*) were considered most useful, there has been an explosion of research into the taxonomy of the genus, and GenBank (http://www.ncbi.nlm.nih.gov) now contains over 1,000 sequences of *gltA* from the genus, mostly from rodents and insectivores, although isolates from hosts as diverse as marsupials and marine mammals have been sequenced.

Surveys of *gltA* diversity have revealed an impressive range of *Bartonella* isolates from rodents and their arthropod ectoparasites, which are generally treated as

clonal and recombinantly isolated (e.g. [21, 23]). Indeed, the paradigm of isolation from recombination and inferred clonality underpins almost all recent work on *Bartonella* in a clinical or veterinary setting [8, 19, 30]. Great diversity can occur even across short geographical distances; thus, Welc-Falęciak et al. [44] recorded 16 genotypes of *B. taylorii* and *B. grahamii* from a small area of NE Poland, treating all as isolated non-recombinant strains. This level of diversity suggests either implausibly rapid mutational evolution of *gltA* or a very long independent evolutionary history for clades. The alternative hypothesis is that diversity is a result of recombination. To test this, we have systematically searched for evidence of recombination within a population of *Bartonella* isolates from rodents in NE Poland, part of the same sample area as Welc-Falęciak et al. [44], utilising a nested clade [41] approach to describe the most parsimonious structure for *Bartonella* clades within this population, which could then be used to test specific hypotheses about the prevalence of recombination within and between clades.

Methods

Mice (*Apodemus flavicollis*) and voles (*Microtus arvalis*, *Mi. oeconomus* and *Myodes glareolus*) were live-trapped at monthly intervals between June 2007 and May 2009 in regenerating old-field habitats derived from abandoned agricultural land (post 1990) and in a managed forest at Urwitałt in the Mazury Lake District, NE Poland. These habitats were within one of the locations utilised by Welc-Falęciak et al. [44]. Within this location, the most distant trapping sites (between forest and old fields) were no more than 5 km apart. Detailed trapping protocols have been described previously [35] and conformed with permission granted by the Polish Ethical Committee (permission number 737/2007). The four sampled rodent species represent the greater part of the small mammal fauna in this area. *Micromys minutus*, *Apodemus agrarius*, *Sicista betulina*, *Sorex araneus* and *S. minutus* also occur infrequently but were not sampled. At each capture, 50 μl of blood was taken from a tail vein directly into 200 μl of 0.001 M EDTA and used for PCR amplification. The *gltA* gene locus was amplified as previously described by Norman et al. [33], using primers specific to a fragment corresponding to amino acids 260 to 370 of the *Escherichia coli* sequence [32]. For selected isolates, additionally fragments of a further five genes were amplified: RNA polymerase beta-subunit gene (*rpoB*), cell division protein gene (*ftsZ*), riboflavin synthase gene (*ribC*), 60 kDa heat-shock protein gene (*groEl*) and 16S RNA coding gene. The 333 bp *rpoB* gene fragment was amplified using primers described previously by Renesto et al. [37], modified on the

basis of alignment of sequences from GenBank (accession numbers, AF165991, AF165993, AF165995, AB426700, AB426701), with the forward primer, 5′-GCACGATTYG CATCATCATTTTCC-3′ and reverse primer, 5′-CGCATT ATGGTCGTATTTGTCC-3′. The PCR of the *groEl* gene fragment was conducted using the forward primer of Zeaiter et al. [45], 5′-GGAAAAAGTGGGCAATGAAG-3′ and reverse primer designed in this study (on the basis of sequences deposited in GenBank, AF304017, AB426677, AF014833), 5′-TCCTTTAACGGTCAACGCATT-3′, and the obtained product was 752 bp. Primers to 420 bp fragment of *ribC* gene (forward, 5′-TYGGTTGTGTKG AAGATGT-3′; reverse, 5′-AATAATMAGAACATC AAAAA-3′), 515 bp fragment of *ftsZ* gene (forward, 5′-CATATGGTTTTCATTACTGCYGGTATGG-3′; reverse, 5′-TTCTTCGCGAATACGATTAGCAGCTTC-3′) and 369 bp fragment of *16S RNA* gene (forward, 5′-TCA GAACGAACGCTGG CGGC-3′; reverse, 5′-CGTCATTATCTTCACCGG-3′) were designed in this study on the basis of sequences from GenBank (accession numbers: *ribC*-AY116635, AY116627, AB426690, AB426689, *ftsZ*-AF467754, AF467756, AB426641, AB426642, AB426645, AB426647, AB426649, AB426651, NC012846, *16S RNA*-Z31349, Z31350, Z31351). All of the genes, after first denaturation in 94 °C for 2 min, were amplified with 40 cycles of 94 °C for 45 s, 47 °C (*groEl*), 54 °C (*16R RNA*), 52 °C (*rpoB* and *ribC*) or 61 °C (*ftsZ*) for 45 s, 72 °C for 45 s and followed by a single 7-min extension step in 72 °C. All PCR products were sequenced in both directions. Chromatogram quality was inspected visually, and only sequences derived from a single PCR product (i.e. no ambiguous peaks on the chromatogram) were analysed further. Potential mixed infections were therefore excluded from analysis.

Analysis

Phylogenetic analysis and alignments were carried out using the Mega 4.1 software [39], but examination of resulting clado-grams revealed poor resolution of clades at a subspecific level. A network of zero-step and one-step clades was therefore established using the methodology of Nested Clade Analysis and statistical parsimony [41–43] to describe the most likely mutational relationships between *Bartonella* isolates collected during this work. The cladogram and the limit for statistical parsimony were calculated using TCS [10]. Four isolates, EU014267, EU014269, EU014274 and EU014275, collected from the same location by Welc-Falęciak et al. [44] in 2005, were also included because they represented otherwise missing internal steps within the clade network. Log-linear models were implemented using SPSS v 14.00 to establish significant departures from randomness in the host range of isolates within each nested *Bartonella* clade [41]. To identify

recombination within the *gltA* gene, the sequenced fragment was first divided into three 100 bp segments and phylogenies generated using the minimum evolution algorithm in Mega 4.1. Discrepancies between these phylogenies were then used to identify potential recombinant *gltA* sequences, which were analysed further to confirm or reject recombination using the RDP-2 software package [18]. To identify potential recombination events between disparate parts of the genome, isolates from the range of *Bartonella gltA* clades were sequenced at the other genes described. All distinct genotypes of each gene were treated as distinct alleles and coded as such. Using an MLST approach [14, 29], distinct alleles were then plotted on to the cladogram and evidence was sought of disjunctions between the overall distribution of housekeeping genes and of connections between disparate clades, which could be taken as evidence of a recombinant event. The congruence of the gene phylogenies to each other and to the *gltA* phylogeny was tested by generating consensus (100 bootstraps) maximum likelihood phylogenies using PhyML ([20], performed on the Montpellier bioinformatics platform and the University of Oslo Bioportal), after first establishing optimal DNA evolution models for each gene using jModelTest [36]. The congruence between trees generated in this way for each gene and trees constrained by the assumption that the *gltA* phylogeny reflected the evolutionary history of the *Bartonella* isolates was tested using maximum likelihood ratio tests.

The unique sequences were deposited in GenBank under accession numbers GU338880-GU338885 (*16S*), GU338887-GU338901 (*ftsZ*), GU338903-GU338915 and GU338917-GU338924 (*ribC*), GU338925-GU338936 and GU338938-GU338941 (*rpoB*), GU338942-GU338976 (*gltA*), and GU559862-GU559871 and GU559873 (*groEl*).

Results

From 1,457 rodents sampled over 2 years (2007–2009), 2,582 blood samples were obtained, of which 579 (22.4%) were positive for *Bartonella*. Of these, 147 isolates were characterized more fully, sequencing of *gltA* revealing a rich, complex diversity of strains. Phylogenetic analysis (minimum evolution algorithm; 1,000 bootstrap replicates, Mega 4.1) resulted in 26 (102 isolates) poorly resolved variants (Ur01-Ur26), which clustered in a clade that included the *B. taylorii* type strain and other *B. taylorii* isolates from rodents and a further five variants (Ur28-Ur32, 39 isolates), which grouped with type isolates of *B. grahamii*. A single variant (Ur33, one isolate) could be referred to *B. birtlesii* while three variants (Ur35-Ur37, three isolates) showed closest similarity to *B. doshiae*. One isolate (Ur27) clustered persistently with *B. grahamii* in the cladistic analyses although it showed

greatest percent similarity at *gltA* to *B. taylorii* while another (Ur34) could not be linked with any previously described species (Fig. 1). Because of the poor resolution within the *B. taylorii* isolates, nested clade analysis was used to establish hypotheses of mutational relationships between clades and was therefore used to infer the clonal structure of *Bartonella* isolates. For the 292 bp fragment of *gltA*, the probability that an individual step in the cladogram is parsimonious is 0.9985 (calculated using TCS). We have therefore treated isolates which cannot be linked to their nearest neighbour by less than three missing mutational steps ($P=0.9899$) or to the central clade of the cladogram (as identified by the TCS) by five or less mutational steps ($P=0.9748$) as belonging to separate clades. In fact, there is only one example of a clade at this threshold, the case of Ur01 within the *B. taylorii* clade A. In all other cases, the difference between clades is of the order of a minimum of seven to ten changes at the *gltA* locus, and assignation to these higher level clades was unambiguous.

The *gltA* Network

Using NCA, the 39 *B. grahamii* isolates grouped into a network of five zero-step clades differing by single mutations within *gltA* (Fig. 2) and a sixth clade (Ur28) separated from its nearest neighbour by two mutations. The probability that this overall 6-step network is parsimonious is 0.9650 (TCS), and the links between the *B. grahamii gltA* clades were therefore accepted at the 95% level.

The 102 isolates of *B. taylorii*, on the other hand, exhibited a much more complex pattern breaking up into three clusters of clades (*B. taylorii* clades A, B and C) which were only distantly related to each other and two other clades (Ur01 and Ur27) which could not be parsimoniously linked to any higher grouping. The distances between the major clades are considerable; clade B is 13 nucleotides different to the type A clade while the type C clade differs by seven nucleotides from type A. Ur01 differs by five nucleotides from the central clade of the type A clade and is therefore close to the parsimony limit for this clade. We treat it as independent of the type A clade, the most conservative interpretation of its position.

The *B. taylorii* clade A consists of fifteen zero-step clades (Fig. 2). None of these clades are more than the three steps from the centre of the network, and most are directly connected by a single step to another clade. Two clades, Ur08 and Ur16, are connected to their nearest neighbour by one and two missing transitions, respectively, but these still remain within the parsimonious limit ($P=0.9530$), and the links between these clades were accepted at the 95% level.

Figure 1 Phylogeny of a 292 bp fragment of citrate synthase (*gltA*) generated using minimum evolution algorithm (MEGA 4.1) with 1,000 bootstrap replicates. Species identities based on nearest match to type sequences from GenBank, clade identities based on second-step clades from the nested clade analysis. Two clades (Ur27 and Ur34) could not be assigned to a species with confidence
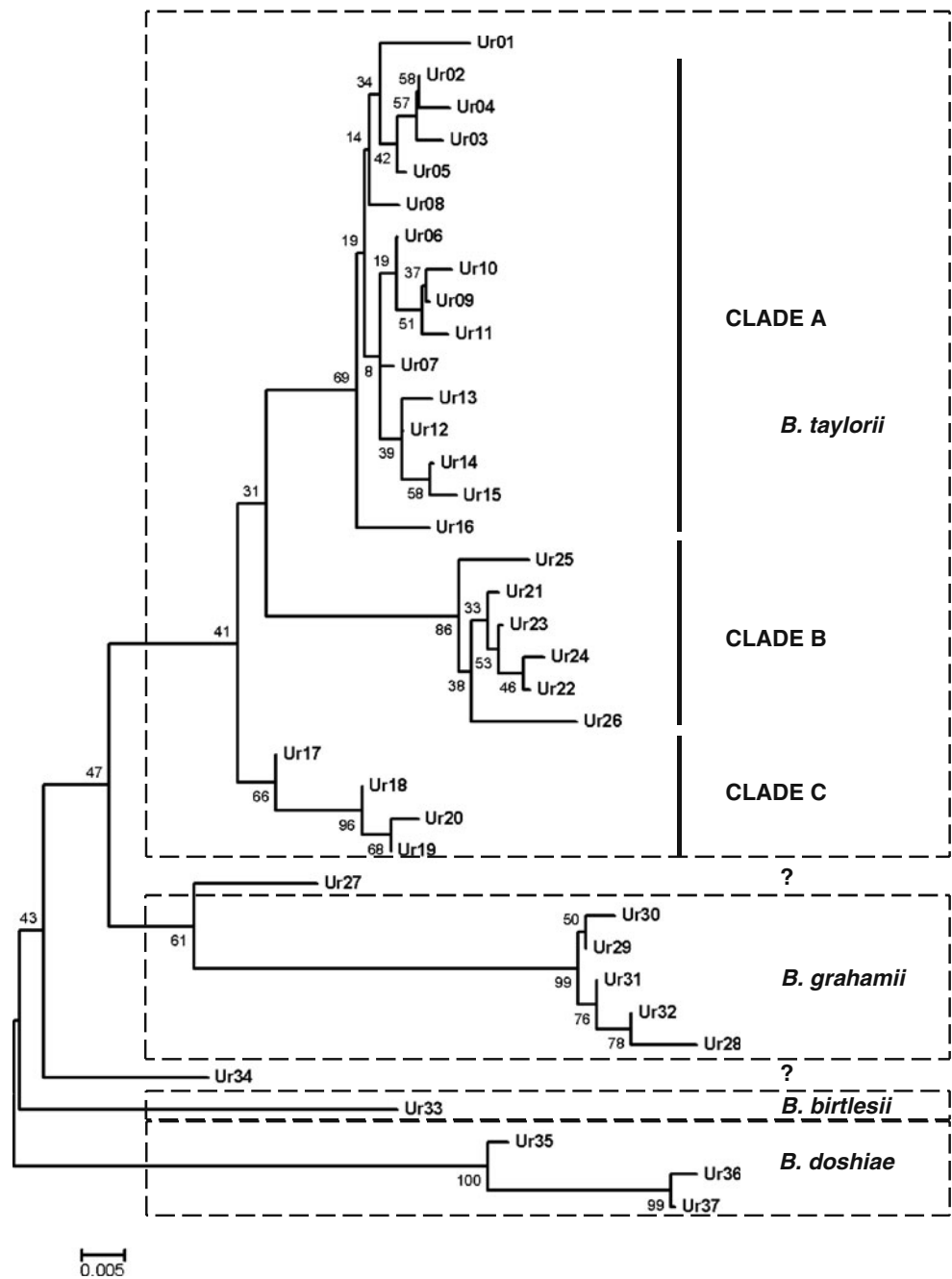
The *B. taylorii* clade B represents the most complex of the *B. taylorii* clades (Fig. 2). It consists of a core of six zero-step clades (Ur21, Ur22, Ur23, Ur24, EU014274 and EU014275), which are all joined by no more than two mutational steps into a network with Ur21 at its centre. This part of the *B. taylorii* clade B is clearly parsimonious and is accepted at the 95% level. There is less certainty over the placement of the remaining parts of the network as both Ur25 and a clade containing Ur26 and EU014269 connect into Ur21 by two missing mutational steps. Strictly, according to the criteria established using the TCS, this is a parsimonious network with Ur26 connected to the central clade (Ur22) by five steps, including only two missing steps (*P*=0.9530); however, in view of the central position of Ur26 in a number of recombinational events within the network, this may not be the correct position for this clade. The *B. taylorii* clade C represents a simple and parsimonious network of four zero-step clades.

The *B. doshiae* clade consists of two zero-step clades which are parsimoniously joined (Ur37 and Ur36) and a third clade (Ur35) which appears to be *B. doshiae* based on overall sequence similarity, but which differs from Ur36 and Ur37 by, respectively, seven and eight base substitutions.
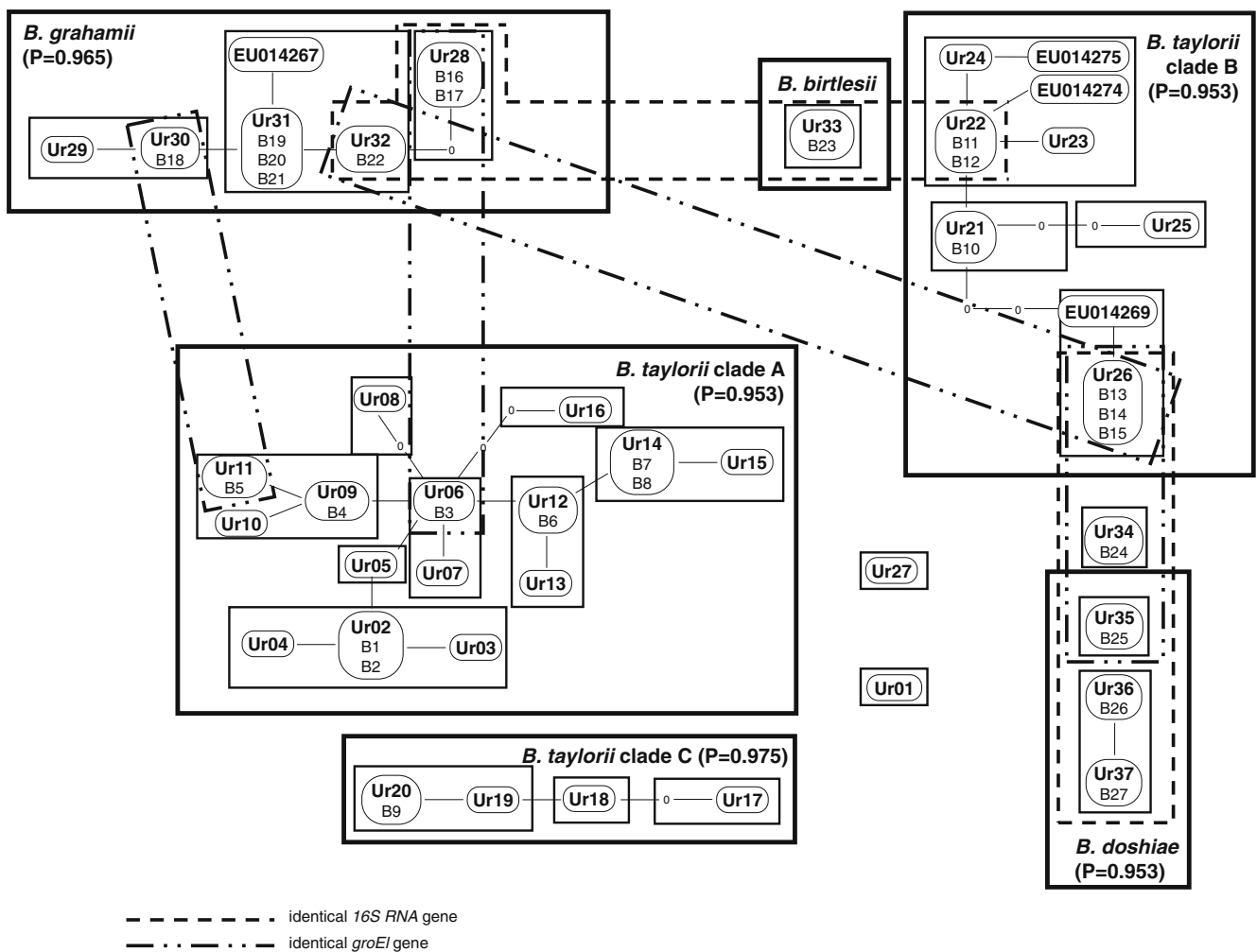
Figure 2 Cladogram of the *Bartonella* isolates collected in this work, and relevant haplotypes from Welc-Falęciak et al. [44] based on a 292 bp fragment of citrate synthase showing known (perfect match) recombinant events within the cladogram. "Ur01-Ur37" represents unique variants recovered from *Bartonella* isolates corresponding to zero-step clades in the terminology of Templeton et al. [41]. *Boxes* represent one-step clades. 'Missing' haplotypes (i.e. predicted but not collected) are indicated by *0*. Larger clades corresponding to 'species' and other higher clades surrounded by heavy boxes. Distances between clades of more than five base changes not marked because of likelihood that these will be recombinant events. Clades sharing *groEl* and *16S RNA* haplotypes between 'species' are marked

Recombination Within *gltA* and the Origin of Novel *Bartonella* Clades

Within this network, four clades remained problematical; Ur01, which appears to be a *B. taylorii* isolate, Ur35 which appears to represent a form of *B. doshiae*, and Ur27 and Ur34 which could not be identified with confidence based on their *gltA* sequences. Recombination within *gltA* appeared to be involved in the generation of two of these four clades. Ur27 appears, by sequence analysis, to lie between *B. taylorii* and *B. grahamii*. On closer examination, it became clear that the first part of the sequenced 292 bp *gltA* fragment was identical with the *gltA* sequence for Ur20, part of the *B. taylorii* clade C. The final part of the *gltA* fragment, however, was identical at every variable base with the equivalent part of the sequence for Ur31, a clade which forms part of *B. grahamii*. The recombinant

break point was between bases 1024 and 1025 of the gene (numbering according to [32]). Statistical support for this event was high and provided by four algorithms [$P=$ 0.0033, GENECONV; $P=0.00027$, MaxChi and Chimaera; $P=0.00000399$, 3Seq] within the RDP-2 package and was obvious on visual inspection (Fig. 3).

Recombination also appears to have been involved in the generation of the *B. doshiae* clade Ur35, which seems to be a recombinant clade between *B. doshiae* (the type sequence, which has not been recovered at Urwitałt) and Ur26 (Fig. 3). In this case, of the 41 variable sites between the three sequences, all are consistent with the parental sequences with a break point between bases 1051 and 1058 of the gene. This event has high statistical support ($P=0.00032$, GENECONV; $P=0.021$, MaxChi and Chimaera; 0.000007, 3Seq) and is clear on visual inspection (Fig. 3).
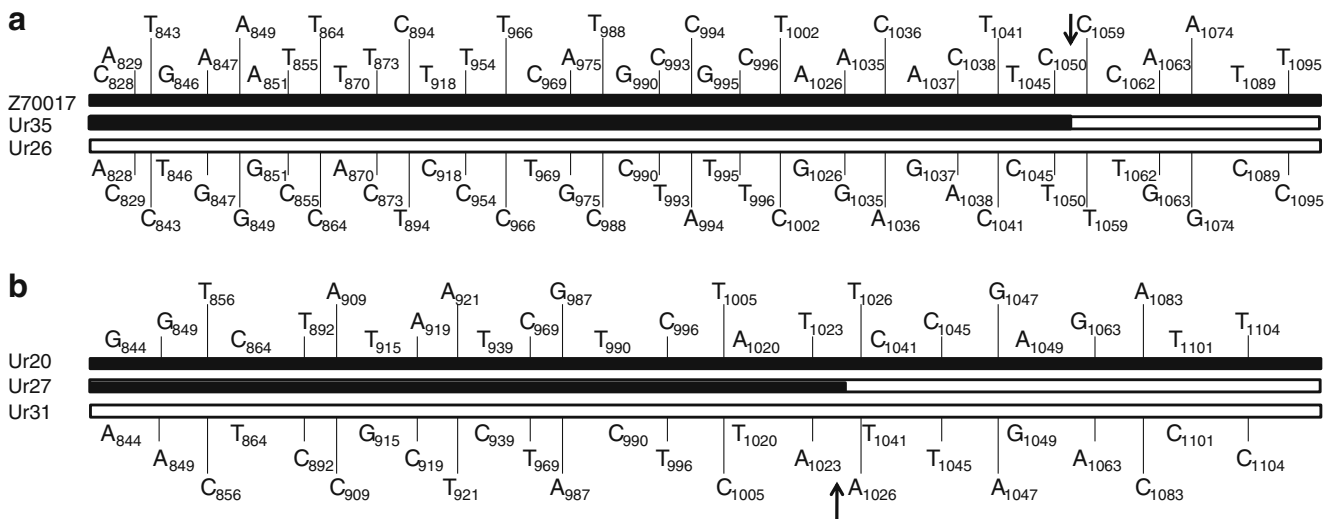
a

| A_{829} | T_{843} | A_{847} | A_{849} | T_{855} T_{864} | T_{873} | C_{894} | T_{966} | A_{975} T_{988} | C_{993} | C_{994} C_{996} | T_{1002} | C_{1036} | C_{1038} | T_{1041} | $\downarrow C_{1059}$ | A_{1074} |
| C_{828} G_{846} | A_{851} | T_{870} | T_{918} T_{954} | C_{969} | G_{990} | G_{995} | A_{1026} A_{1035} | A_{1037} | T_{1045} | C_{1050} C_{1062} | A_{1063} | T_{1089} T_{1095} |

Z70017
Ur35
Ur26

| A_{828} | T_{846} | G_{851} | A_{870} | C_{918} | T_{969} | C_{990} | T_{995} | G_{1026} | G_{1037} | C_{1045} | T_{1062} | C_{1089} |
| C_{829} C_{843} | G_{847} G_{849} | C_{855} C_{864} | C_{873} T_{894} | C_{954} C_{966} | G_{975} C_{988} | T_{993} A_{994} | T_{996} C_{1002} | G_{1035} A_{1036} | A_{1038} C_{1041} | T_{1050} T_{1059} | G_{1063} G_{1074} | C_{1095} |

b

| G_{844} | G_{849} | T_{856} | C_{864} | T_{892} | A_{909} | A_{919} | A_{921} | C_{969} | G_{987} | C_{996} | T_{1005} | T_{1023} | T_{1026} | C_{1045} | G_{1047} | G_{1063} | A_{1083} | T_{1104} |
| | | | | T_{915} | T_{939} | T_{990} | | A_{1020} | C_{1041} | A_{1049} | T_{1101} |

Ur20
Ur27
Ur31

| A_{844} | T_{864} | G_{915} | C_{939} | C_{990} | T_{1020} | T_{1041} | G_{1049} | C_{1101} |
| A_{849} C_{856} | C_{892} C_{909} | C_{919} T_{921} | T_{969} A_{987} | T_{996} | C_{1005} A_{1023} $\uparrow A_{1026}$ | T_{1045} A_{1047} | A_{1063} C_{1083} | C_{1104} |

Figure 3 Examples of intra-gene recombination events into the *gltA* gene with polymorphic sites and break points marked on. **a** Ur35 (*B. doshiae*) as a recombinant of Ur26 (*B. taylorii* clade B) and *B. doshiae* type sequence (Z70017). **b** Ur27 (species not defined) as a recombinant of Ur20 (*B. taylorii* clade B) and Ur31 (*B. grahamii*)

Recombination Across the Genome

Twenty-seven isolates (B1-B27), chosen to give the greatest possible phylogenetic coverage of obtained *gltA* phylogeny, were sequenced at the other gene loci. The genes that were used allowed analysis of recombinant events spanning the genome. In general, the pattern of alleles of *rpoB*, *ribC* and *ftsZ* mirrored the pattern of *gltA* clades and supported the assignation of clades based on *gltA* (Fig. 4), but although the consensus phylogenies of *ribC* and *rpoB* were similar, the difference between them and that of *gltA* was nevertheless significant ($0.01 < P < 0.05$, maximum likelihood ratio test). The *ftsZ* gene phylogeny was less congruent with the *gltA* gene phylogeny ($P > 0.05$, maximum likelihood ratio test).

The pattern of alleles of these other genes established conclusively that Ur35 and Ur34 were indeed related to Ur36 and Ur37 within *B. doshiae*. The most notable differences in the distribution of these housekeeping gene alleles are listed in Table 1. The most important concerned Ur26. Three isolates of this clade were sequenced, all with identical *gltA*; however, while two isolates resembled *B. taylorii* based on other housekeeping genes and supported placement within the *B. taylorii* clade B, the third isolate was identical at all other genes with Ur34 and Ur35 and was *B. doshiae*-like. This appears therefore to be a case of a recombinant exchange of *gltA* from *B. taylorii* into a *B. doshiae* clade. The next most distinctive example of recombination is the presence of an alternative *ribC* allele (differing by 32 or 33 base substitutions) in clades Ur28 and Ur32, suggesting a recombinant event dividing *B. grahamii* into two groups. The other important difference is the presence of an alternative *ftsZ* allele (differing by eight substitutions) within Ur02, suggesting a recombinant event into this part of the *B. taylorii* clade A.

The other two genes sequenced, *groEl* and the *16S RNA* gene, show much greater evidence of recombination across the network of *Bartonella* clades, and for both loci, the consensus phylogenies were very different from that of *gltA* ($P > 0.001$, maximum likelihood ratio test). Although the *16S RNA* gene of all isolates was sequenced successfully, there were several cases of *groEl* not amplifying successfully, complicating interpretation for this gene. Nevertheless, there were clear examples of *groEl* exchanged between clades (Table 2). Three examples were identified where *B. taylorii* isolates shared different *groEl* alleles with *B. grahamii* isolates. In addition, there were two examples of *gltA* clades (Ur28, Ur31) in which different isolates had radically different *groEl* alleles. An intriguing observation was that two *B. taylorii* clade A clades had *groEl* alleles most similar to that of the *B. birtlesii* type sequence (AM690315). It is not clear whether these represent two distinct recombinant events involving as yet uncollected *B. birtlesii* isolates from the Urwitałt environment or evidence of an older recombinant event into *B. taylorii* which has subsequently undergone mutational drift and disruption by other recombinations from *B. grahamii*. A similarly complex pattern of recombination was noted for the *16S RNA* gene. Eleven *16S RNA* alleles were collected from the sequenced isolates, some of which were widely distributed. One allele, found in the *B. birtlesii* isolate, was also identified in the *B. grahamii* clades Ur32 and Ur28 and in Ur22 (*B. taylorii* clade B). These links between clades are illustrated in Fig. 2.

The Ecology of *Bartonella* Clades

The first-step *gltA* clades showed a degree of both host and habitat specificity (Table 3). Log-linear contingency analysis revealed a significant association between clade and host identity (first-step clade x host species, $\chi^2 = 266.63$, df=69,
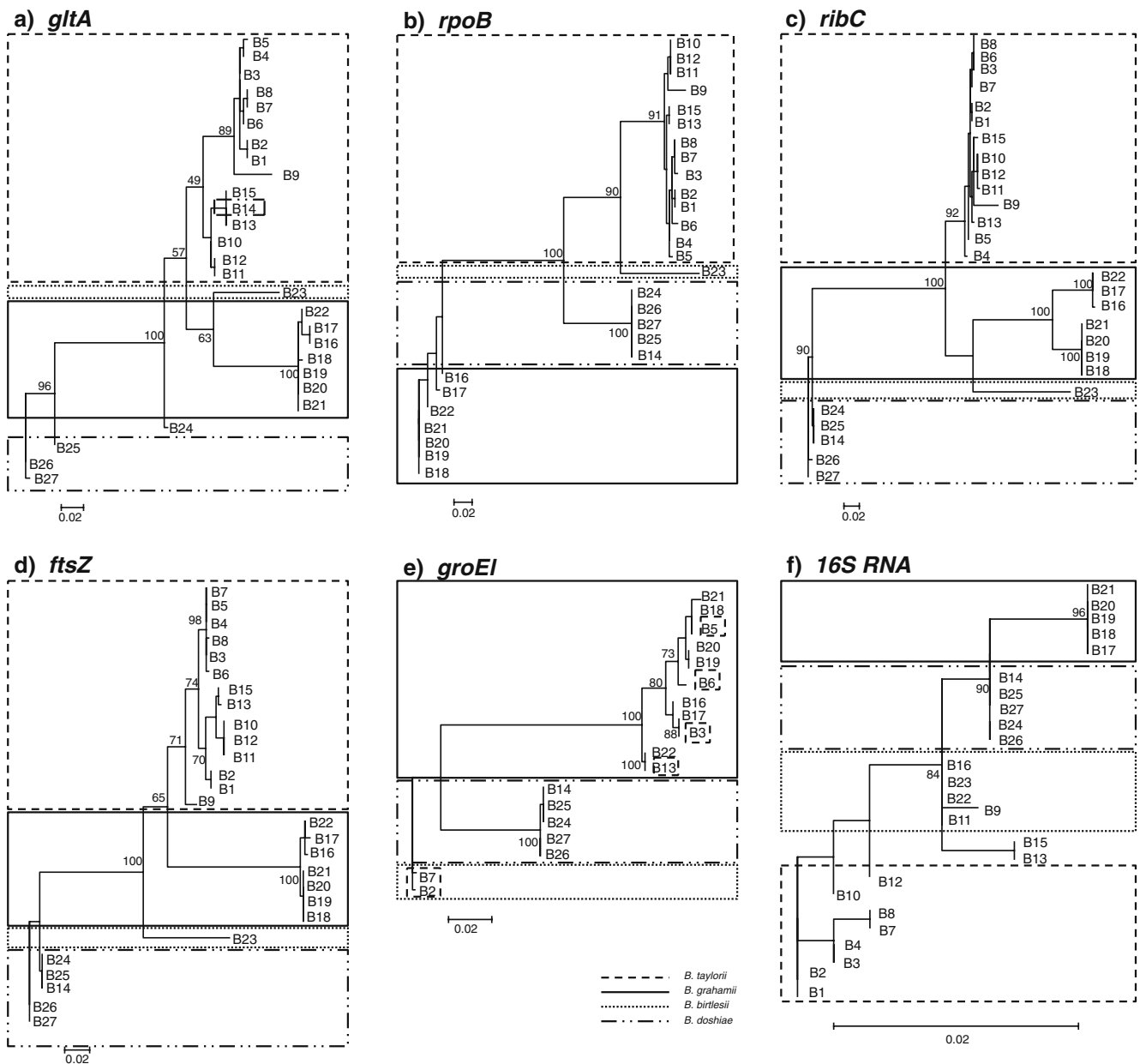
Figure 4 Comparison of phylogenies obtained with the metabolic genes analysed in this work: *gltA*, *ftsZ*, *ribC*, *rpoB*, *groEl* and gene coding 16S RNA. Phylogenies generated using maximum likelihood (PhyML) with 100 bootstrap replicates. Species identities *boxed*, based on nearest match to type sequences from GenBank. Note the lack of support for *B. taylorii* with *groEl*

$P < 0.001$). A significant association was also noted with habitat, but it is not clear whether this is due rather to the habitat specificity of the rodent species. At second-step level, significance was noted between clade identity and host family (second-step clade x host family, $\chi^2 = 139.51$, df=13, $P < 0.001$).

The frequency of occurrence of the *gltA* clades was consistent with predictions of the nested clade analysis, although clade prevalence was not included in the original algorithm. More isolates were collected from internal clades (Table 3) than from the tip zero-step clades, and this distribution also showed a relationship with host identity, with tip clades more commonly encountered in *Microtus* than the stem clades, which occurred in *Myodes* and *Apodemus* ($\chi^2 = 42.50$, df=6, $P < 0.001$). This was reflected in the molecular diversity of isolates collected from each host. In the case of *Microtus* spp., an average of only 2.5 isolates per clade were identified (13 distinct *gltA* variants from 33 sequenced isolates). In *My. glareolus* on the other hand, seven isolates per clade (seven distinct *gltA* variants from 53 isolates) were collected. *Bartonella* from *A. flavicollis* showed intermediate diversity, with 3.1 isolates per clade (20 distinct *gltA* variants from 62 isolates).

Table 1 Evidence of recombination within clades

	B. taylorii clade A								B. taylorii clade B					B. grahamii							B. doshiae			
	B3	B4	B5	B6	B7	B8	B1	B2	B11	B12	B13	B15	B10	B18	B19	B20	B21	B22	B16	B17	B26	B27	B25	B24
gltA	–	1	2	1	2	2	2	2	–	0	5	5	1	1	–	0	0	1	3	3	–	1	**7**	**30**
rpoB	–	3	4	4	1	1	3	3	–	0	4	4	0	0	–	0	0	3	4	6	–	0	0	0
ribC	–	6	3	0	1	0	3	3	–	1	5	5	1	0	–	0	0	**32**	**33**	**32**	–	2	5	5
ftsZ	–	0	0	1	0	1	**8**	**8**	–	0	5	4	0	0	–	0	0	3	4	5	–	0	5	5
groEl	–	ND	**10**	**11**	**68**	ND	ND	**70**	ND	ND	ND	ND	ND	3	–	0	6	**16**	**9**	**9**	–	0	1	1
16S RNA	–	0	ND	ND	1	1	1	1	–	2	2	2	5	0	–	0	0	4	4	0	–	0	0	0

Number of pairwise differences between an isolate considered typical of each clade (B3, B11, B19 and B26) and other isolates from that clade. Numbers in italics denote changes of up to five bases which are assumed due to mutational drift. Numbers in bold denote larger changes which suggest recombination

ND not done

Discussion

It is customary to treat *Bartonella* isolates as clonal [21] and to assume that meaningful phylogenies can be generated from single gene alignments [23, 25] despite convincing genomic evidence that the genus undergoes genetic re-organisation and recombination [2, 38]. The present work makes it clear that these bacteria frequently recombine, to the extent that it proved impossible to use an algorithm such as eBurst [13] to demonstrate clonal multiplication. Instead, nested clade analysis [41] allowed us to visualise the diversity of *gltA* alleles in *Bartonella* and then to test specific hypotheses about their distribution. Nested clade analysis in phylogeography [40] has been criticised [31], but we have used the method as originally envisaged [41] to partition character states (host specificity or sequence variation in other genes) between the network of clades established using one gene (*gltA*). The network of first-step clades (joined by single mutational steps) are probably clonal; three such clones were identified as 'B. grahamii', and 16 as 'B. taylorii', with two clones of 'B. doshiae' and a single clone of 'B. birtlesii' within the overall *Bartonella* population from rodents within this small area of forest and old fields. There was compelling evidence for recombination between all four *Bartonella* taxa identified, hitherto regarded as valid species, and recombination within the *gltA* locus had generated two clades which were otherwise difficult to assign to species.

Bartonella has an established binomial taxonomy, which is assumed to reflect underlying biological reality. There has been frequent speculation concerning the extreme diversity of *Bartonella* (over 1,000 sequences of *gltA* from this genus are already deposited in GenBank) as an evolutionary and phylogenetic problem [12, 21, 23] while at the same time the contribution of recombination to genome evolution in the genus is clear [2, 38]. Despite frequent cautions against the use of single-gene phylogenies in microbial taxonomy, the reliance on binomial taxonomy and single gene sequences (most frequently *gltA* or *16S RNA*) for diagnosis has led to important judgements in zoonotic medicine and wildlife disease. Thus the reputation of rodent-infecting *B. grahamii* and *B. vinsonii arupensis* as emerging human pathogens (e.g. [15, 24]) rests upon the isolation of single clades from humans diagnosed by gene sequences which are prone to recombination. Similarly, *B. henselae* (cat and human parasite transmitted by fleas) has been recorded from marine mammals, transmitted by *Cyamus* [16]; the use of a wider range of loci in this work might have clarified the exact strains capable of such an ecologically diverse zoonosis. The present work makes it clear that this approach is both unreliable and potentially highly misleading, and the use of these sequences, especially the *16S RNA* marker, for

Table 2 Examples of whole-gene exchange between distant clades of *Bartonella* within the recombinant network

Isolates (zero-step clade, bacteria species)	Genes with identical sequence
B_{17} (Ur28, *B. grahamii*) and B_{18}–B_{21} (Ur30, Ur32; *B. grahamii*)	*16S RNA*
B_{16} (Ur28, *B. grahamii*); B_{22} (Ur32, *B. grahamii*); B_{23} (Ur23, *B. birtlesii*) and B_{11} (Ur22, *B. taylorii* clade B)	*16S RNA*
B_{13}, B_{15} (Ur26, *B. taylorii* clade B) and B_{14} (Ur26, *B. doshiae*)	*gltA*
B_{17} (Ur28, *B. grahamii*) and B_3 (Ur06, *B. taylorii* clade A)	*groEl*
B_{18} (Ur30, *B. grahamii*) and B_5 (Ur11, *B. taylorii* clade A)	*groEl*
B_{22} (Ur32, *B. grahamii*) and B_{13} (Ur26, *B. taylorii* clade B)	*groEl*

Only examples where genes were identical in sequence in both partners are included

Table 3 Identity of zero-step, one-step and two-step clades of *Bartonella* collected during this work and their relationship with infected host

Second-step clade	First-step clade	Zero-step clade	Internal clade	Number of isolates			
				My. glareolus	*A. flavicollis*	*Mi. arvalis*	*Mi. oeconomus*
1	1	Ur01	NA		1		
2	2	Ur02	Y		3 (B1, B2)		
		Ur03	N		1		
		Ur04	N		1		
	3	Ur05	Y		3		
3	4	Ur06	Y		14	2 (B3)	
		Ur07	N		2		
	5	Ur08	N		1		
	6	Ur09	Y		13 (B4)		
		Ur10	N		1		
		Ur11	N		1 (B5)		
4	7	Ur12	Y			2	1 (B6)
		Ur13	N			1	
	8	Ur14	Y	23 (B7, B8)			
		Ur15	N	2			
5	9	Ur16	N				1
6	10	Ur17	N		1		
	11	Ur18	Y		1		
	12	Ur29	Y		4		
		Ur20	N		1 (B9)		
7	13	Ur21	Y	5 (B10)			
	14	Ur22	Y				3 (B11, B12)
		Ur23	N				1
		Ur24	Y				1
	15	Ur25	N	1			
8	16	Ur26	N			10 (B13, B15)	1 (B14)
9	17	Ur27	NA		1		
10	18	Ur28	N			4 (B16, B17)	
	19	Ur29	N		1		
		Ur30	Y	1	2 (B18)		
	20	Ur31	Y	16 (B19, B20)	10 (B21)	1	
		Ur32	Y			4 (B22)	
11	21	Ur33	NA		1 (B23)		
12	22	Ur34	NA				1 (B24)
13	23	Ur35	NA				1 (B25)
14	24	Ur36	NA			1 (B26)	
		Ur37	NA			1 (B27)	

Information about the placement of the clade. Codes in parentheses indicate isolates characterised in more detail

Y internal clade, *N* tip clade, *NA* does not apply

diagnostic purposes should be undertaken only with extreme caution.

Although genetic re-organisation is clear at a genomic level [2], the stability of housekeeping genes is assumed, between which pathogenicity islands are free to evolve [38]. Saenz et al. [38] suggest that recombinant duplication of pathogenicity-related loci led to the evolution of the largest, most complex *Bartonella* genomes, with up to 69 virulence-related genes (including plasmid encoded copies, see [2]). However, this phylogeny [38] was based only on type isolates; while *B. grahamii* is a relatively homogenous taxon [3], we have no such confidence in respect of *B. taylorii* or *B. doshiae*. We have shown that the latter intergrades with *B. taylorii*, and there is more evidence of interchange between *B. grahamii*, *B. doshiae* and *B. birtlesii* and the different clades of *B. taylorii* than there is of interchange between *B. taylorii* clades themselves. This network of links within *B. taylorii* presents a problem for wider *Bartonella* taxonomy. As currently constituted, phylogenies based on *gltA* (see Fig. 1) which include the three *B. taylorii* clades identified in this work span a huge range of isolates from rodents, Eurasian insectivores and bats [11]. If the isolates from insectivores and bats are considered not to be *B. taylorii*, then the current *B. taylorii* will need to be divided. The scope of *B. taylorii* as a natural taxon therefore needs experimental review.

Sequencing of *B. tribocorum* and *B. grahamii* genomes has shown that *groEl* and *16S RNA* genes lie within a high plasticity region [2, 38], 1624722 and 2084545 bases from the origin of replication. The position of these genes, adjacent to *virB/virD* and *vbh* pathogenicity islands, suggests that they could be the most useful genes in predicting pathogenic potential in *Bartonella* isolates. The situation with *gltA* is most remarkable; previously the stability of this gene has been assumed, but it is now clear that novel variants of this gene are actively generated by recombination. In this respect, *Bartonella* appears similar to *Wolbachia* [1], and recombination within *gltA* may be quite widespread in this group of bacteria.

The recombinant network of *Bartonella* isolates extends across four species of rodents in two distinct but related biotypes, old managed forest and recently (20 years) abandoned agricultural land. Nested clade analysis based on *gltA* had sufficient resolving power to identify clades of *B. taylorii* and *B. grahamii* associated with particular host species and with habitat. *B. taylorii* and *B. grahamii* as species are normally considered to lack host specificity [3, 4], but in the present work both zero-step clades and, to a large extent, first-step clades nested within the cladogram were host specific. Thus, *B. taylorii* clade B was recovered only from voles; three of the first-step clades within this clade came from *Microtus* spp. in the abandoned field system, and two first-step clades came from *My. glareolus* in the forest. Similarly, within the *B. taylorii* clade C, all isolates (eight belonging to four one-step clades) came from *A. flavicollis*. The larger *B. taylorii* clade A and *B. grahamii* were more complex, and it was only within these clades that first-step clades infecting more than one host species were noted. The most characteristic was the large *B. grahamii* clade Ur31; 27 isolates were recorded from three rodent species, but the majority (26 isolates) came from the woodland rodents, *My. glareolus* and *A. flavicollis*. The stem zero-step clades (Ur06, Ur31, Ur22) within each clade contained the greatest number of isolates, as would be predicted for the oldest clades from which most others were descended [9]. More interestingly, these were clades involving forest rodents; whereas, isolates infecting *Microtus* in the abandoned field system tended to belong to tip clades collected from only one or two hosts. As a result, clade diversity was highest in *Microtus* spp. and lowest in *My. glareolus*. Almost all of the recombinant connections noted in this work, with the exception of that between *B. grahamii* Ur30 and *B. taylorii* Ur11 (exchange of a *groEl*), involved *Microtus* as one of the partners. It is worth pointing out that Ur26 in this context is a large clade of 11 isolates from *Microtus*, all with an identical *gltA* sequence but composed of both *B. doshiae* and *B. taylorii*. This clade took part in many of the long-range recombinant exchanges across the network. This suggests a role for *Microtus* in the diversification of *Bartonella*, although this rodent has the strictest habitat specificity [26, 34] and is unlikely to act as a bridge between rodent communities. The low diversity of isolates from *My. glareolus* supports the role of this host supporting clonal expansion within the forest environment but not favouring genetic diversification or exchange of genetic material with other isolates. The competence of natural *Bartonella* clades is unknown, but *B. grahamii* at least carries a diverse complement of plasmids and phages and contains numerous phage sequences intercalated into the genome [2]. It is not clear therefore whether this diversity of transforming factors extends across all the species and isolates examined in this work or is restricted to the 'B. grahamii sensu stricto' clades such as Ur29, 30 and 31, which are most similar to the Uppsala isolate sequenced by Berglund et al. [2]. Assuming that all isolates are equally competent, then the importance of *Microtus* relative to *Myodes* in fostering recombinant isolates is perhaps due to differences in ectoparasite vector ecology between the two species [17, 35]. This is a topic worthy of further research in understanding the role of recombination in generating novel genotypes of differing pathogenicity.

This study has revealed a huge diversity of *Bartonella* over very short geographical distances; sequencing just the 300 bp *gltA* fragment revealed 37 distinct terminal clades among 147 isolates from a small area of forest and

abandoned agricultural land. Sequencing of the additional housekeeping loci revealed further diversity, at least within some clades, and possibly all isolates would prove unique if sequenced at sufficient loci. Other studies (e.g. [21, 22, 44]) report similar diversity, which may be a feature of the biology of *Bartonella*. The agricultural land was abandoned only in 1990, prior to which *Microtus* must have been relatively rare [27]. We may therefore have witnessed an explosion in *Bartonella* diversity linked to the expansion of *Microtus* following ecological succession in the abandoned fields. Whatever the reasons for the considerable recombinant diversity of *Bartonella*, this is clearly an ideal system in which to study recombination in natural populations of parasitic bacteria, as recombinant events are sufficiently common to be observed in real time but not so frequent as to obscure all signal of clonal lineages. This system provides a model for the study of *Bartonella* evolution at a fine scale, to identify in natural populations the mechanisms which allow specific evolutionary lineages to become human pathogens.

References

1. Baldo L, Bordenstein S, Wernegreen JJ, Werren JH (2005) Widespread recombination throughout *Wolbachia* genomes. Mol Biol Evol 23:437–449

2. Berglund EC, Frank AC, Calteau A, Petterson OV, Granberg F, Eriksson AS, Naslund K, Holmberg M, Lindroos H, Andersson SGE (2009) Run off replication of host adaptability genes is associated with gene transfer agents in the genome of mouse infecting *Bartonella grahamii*. PLoS Genet 5:e1000546

3. Berglund EC, Ehrenborg C, Pettersson OV, Granberg F, Näslund K, Holmberg M, Andersson SGE (2010) Genome dynamics of *Bartonella grahamii* in micro-populations of woodland rodents. BMC Genomics 11:152

4. Birtles RJ, Harrison TG, Molyneux DH (1994) *Grahamella* in small woodland mammals in the UK—isolation, prevalence and host specificity. Ann Trop Med Parasitol 88:317–327

5. Birtles RJ (2005) Bartonellae as elegant haemotropic parasites. Ann NY Acad Sci 1063:270–279

6. Boulouis HJ, Chang CC, Henn JB, Hasten RW, Chomel BB (2005) Factors associated with the rapid emergence of zoonotic *Bartonella* infections. Vet Res 36:383–410

7. Breitschwerdt EB, Maggi RG, Chomel BB, Lappin MR (2009) Bartonellosis: an emerging infectious disease of zoonotic importance to animals and human beings. Vet Emerg Crit Care (San Antonio) 20:8–30

8. Breitschwerdt EB, Maggi RG, Cadenas MB, de Paiva V, Diniz PP (2009) A groundhog, a novel *Bartonella* sequence and my father's death. Emerg Infect Dis 15:2080–2086

9. Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. Genetics 134:959–969

10. Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene geneologies. Mol Ecol 9:1657–1659

11. Concannon R, Wynn-Owen K, Simpson VR, Birtles RJ (2005) Molecular characterisation of haemoparasites infecting bats (Microchiroptera) from bats. Parasitology 131:489–496

12. Ehrenborg C, Handley S, Ellis B, Mills J, Holmberg M (2003) *Bartonella grahamii* infecting rodents display high genetic diversity over short geographic distances. Ann NY Acad Sci 990:233–235

13. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 186:1518–1530

14. Feil E, Zhou J, Maynard Smith J, Spratt BG (1996) A comparison of the nucleotide sequences of the *adk* and *recA* genes of pathogenic and commensal *Neisseria* species: evidence for extensive inter-species recombination within *adk*. J Mol Evol 43:631–640

15. Fenollar F, Sire S, Raoult D (2005) *Bartonella vinsonii* subsp. *arupensis* as an agent of blood culture-negative endocarditis in a human. J Clin Microbiol 43:945–947

16. Harms CA, Maggi RG, Breitschwerdt EB, Clemons-Chevis CL, Solangi M, Rotstein DS, Fair PA, Hansen LJ, Hohn AA, Lovewell GN, McLellan WA, Pabst DA, Rowles TK, Schwacke LH, Townsend FI, Wells RS (2008) *Bartonella* species detection in captive, stranded and free ranging cetaceans. Vet Res 39:59

17. Harris PD, Paziewska A, Zwolińska L, Siński E (2009) Seasonality of the ectoparasite community of woodland rodents in a Mazurian forest, Poland. Wiad Parazytol 55:377–388

18. Heath L, van der Walt E, Varsani A, Martin DP (2006) Recombinant patterns in aphthoviruses mirror those found in other picornaviruses. J Virol 80:11827–11832

19. Henn JB, Chomel BB, Boulois HJ, Kasten RW, Murray WJ, Bar-Gal GK, King R, Courreau JF, Baneth G (2009) *Bartonella rochalimae* in raccoons, coyotes and red foxes. Emerg Infect Dis 15:1984–1986

20. Guindon S, Gascuel O (2003) A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

21. Inoue K, Kabeya H, Kosoy MH, Bai Y, Smirnov G, McColl D, Artsob H, Maruyama S (2009) Evolutionary and geographical relationships of *Bartonella grahamii* isolates from wild rodents by multi-locus sequencing analysis. Microb Ecol 57:534–541

22. Inoue K, Kabeya H, Shiratori H, Ueda K, Kosoy MY, Chomel BB, Boulouis HJ, Maruyama S (2009) *Bartonella japonica* sp. nov. and *Bartonella silvatica* sp. nov., isolated from *Apodemus* mice in Japan. Int J Syst Evol Microbiol. doi:10.1099/ijs.0.011528-0

23. Inoue K, Maruyanra S, Kabeya H, Hagiya K, Izumi Y, Une Y, Yoshikawa Y (2009) Exotic small mammals as potential reservoirs of zoonotic *Bartonella* spp. Emerg Infect Dis 15:526–532

24. Kerkhoff FT, Bergmans AM, van Der Zee A, Rothova A (1999) Demonstration of *Bartonella grahamii* DNA in ocular fluids of a patient with neuroretinitis. J Clin Microbiol 37:4034–4038

25. Kosoy MY, Regnery RL, Tzianabos T, Marston EL, Jones DC, Green D, Maupin GO, Olson JG, Childs JE (1997) Distribution, diversity and host specificity of *Bartonella* in rodents from the Southeastern United States. Am J Trop Med Hyg 57:578–588

26. Kozakiewicz M, Gortat T, Kozakiewicz A, Barkowska M (1999) Effects of habitat fragmentation on four rodent species in a Polish farm landscape. Landscape Ecol 14:391–400

27. Kozakiewicz M, Kozakiewicz A (2008) Long-term dynamics and biodiversity changes in the mosaic of agricultural and forest habitats. Ann Zool Fenn 45:263–269

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

29. Maiden MCJ, Bygraves JA, Feil E (1998) Multilocus sequence typing: a portable approach to the identification of clades within populations of pathogenic microorganisms. Proc Natl Acad Sci USA 95:3140–3145

30. Morick D, Osinga N, Gruys E, Harrus S (2009) Identification of a *Bartonella* species in the harbour seal (*Phoca vitulina*) and in seal lice (*Echinophthirius horridus*). Vector Borne Zoonotic Dis 9:751–753

31. Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. Mol Ecol 18:1034–1037

32. Nguyen NT, Maurus R, Stokell DJ, Ayed A, Duckworth HW, Brayer GD (2001) Comparative analysis of folding and substrate binding sites between regulated hexameric type II citrate synthases and unregulated dimeric type I enzymes. Biochemistry 40:13177–13187

33. Norman AF, Regnery R, Jameson P, Greene C, Krause DC (1995) Differentiation of *Bartonella*-like isolates at the species level by PCR-restriction fragment length polymorphism in the citrate synthase gene. J Clin Microbiol 33:1797–1803

34. Panzacchi M, Linnell JDC, Melis C, Odden M, Odden J, Gorini L, Andersen R (2010) Effect of land use on small mammal abundance and diversity in a forest-farmland mosaic in south-eastern Norway. For Ecol Manage 259:1536–1545

35. Paziewska A, Zwolińska L, Harris PD, Bajer A, Siński E (2010) Utilisation of rodent species by larvae and nymphs of hard ticks (Ixodidae) in two habitats in North East Poland. Exp Appl Acarol 50:79–91

36. Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25:1253–1256

37. Renesto P, Gouvernet J, Drancourt M, Roux V, Raoult D (2001) Use of *rpoB* gene analysis for detection and identification of *Bartonella* species. J Clin Microbiol 39:430–437

38. Saenz HL, Engel P, Stoeckli MC, Lanz C, Raddatz G, Vayssier-Taussat M, Birtles R, Schuster SC, Dehio C (2007) Genomic analysis of *Bartonella* identifies type IV secretion systems as host adaptability factors. Nat Genet 39:1469–1476

39. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

40. Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Mol Ecol 7:381–397

41. Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in Drosophila. Genetics 117:343–351

42. Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data: III Cladogram estimation. Genetics 132:619–633

43. Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. Genetics 134:659–669

44. Welc-Falęciak R, Paziewska A, Bajer A, Behnke JM, Siński E (2008) *Bartonella* spp. infections in rodents from different habitats in the Mazury Lake District, Northeast Poland. Vector Borne Zoonotic Dis 8:467–474

45. Zeaiter Z, Fournier PE, Ogata H, Raoult D (2002) Phylogenetic classification of *Bartonella* species by comparing *groEL* sequences. Int J Syst Evol Microbiol 52:165–171