A global map of genetic diversity in *Babesia microti* reveals strong population structure and identifies variants associated with clinical relapse

Supporting information

Supplemental Text:

Sequencing and Enrichment of Global and US *Babesia microti*:

Zoonotic B. microti in the US was first described on Nantucket in 1969[1] and later recognized at multiple sites throughout the Northeast and Midwest during the 1980s[2,3]. Phylogenetic analysis of 18S RNA and beta-tubulin genes has established a species complex with at least three major clades, with the original Nantucket strains along with strains from Switzerland and Russia labeled as Clade 1 or *B. microti* sensu stricto[4], and *B. microti*-like strains such as those from Japan[5,6], Alaska, and Europe classified as Clades 2 and 3, or *B. microti* sensu lato[4], though others have proposed that *B. microti* be reclassified as a genus[7]. We studied samples from Japan[5,6], Alaska[8], Russia[9], and multiple sites in the United States. The origin, date of collection, and sample identifiers for each sample are listed in Supplemental Table 1. The identifiers used to refer to geographic groupings of samples are given below in the section "Nomenclature and Geographic Groupings".

For clinical, rodent, and tick samples, we developed three methods to enrich for parasite material for sequencing: Leukocyte depletion based on cellulose filtration[10], which produced 4.9 (+/- 5.4) -fold enrichment (Supplementary Fig. 1), and two methods of hybrid selection[11] (Supplementary Fig. 1), which yielded an 85.9 (+/- 78.6)-fold enrichment, sufficient to sequence the low quantities of parasite DNA from ticks.

Reads from *B. microti* sensu lato samples aligned poorly to the *B. microti* R1 reference genome[12] and were assembled de novo (see methods), producing draft assemblies of 5.87 Mb for AW-1[5] (Japan), 5.87 Mb for Hobetsu[5] (Japan), and 6.14 Mb for CR400[8] (Alaska), which are included as supplemental files. *B. microti*-like genomes displayed substantial nucleotide divergence, with a mean nucleotide diversity of 84.4% for the Alaskan strain and 85.5% for the Japanese strain as compared to the R1 reference (Supplemental Fig 12a-c). Despite this, the assembled contigs from Alaskan and Japanese strains possessed a conserved genomic architecture (Supplemental Fig. 5a-c) and aligned in contiguous blocks with *B. microti* R1 reference. The distribution of contig sizes for the draft assemblies is given in Supplemental Figure 12c-f.

Nomenclature and Geographic Groupings:

In order to refer to groups of strains consistently, we used the following naming conventions:

Nantucket (NAN): Samples from Nantucket (includes Bab14, which groups with Nantucket samples).

Mainland New England (MNE): Samples from Connecticut, Rhode Island, Massachusetts, New Hampshire, and Maine. Excludes samples from NAN lineage above. This includes ND11, which was isolated from a North Dakota resident, but types with other MNE samples.

Coastal New England (CNE): Samples from Mainland New England and Nantucket, i.e. the union of NAN and MNE.

Reference group (REF): This includes the R1 reference[12] and samples that group with it.

Northeast (NE): All samples from the Northeastern United States, i.e. the union of NAN, MNE, and REF.

Midwest (MW): Samples from Minnesota and Wisconsin. This group excludes ND11, as above, which groups with MNE samples.

Continental US (CUS): All samples from the Northeast and Midwest, i.e. the union of MW and NE.

Global: All study samples, including those from Alaska, Japan, Russia, and the Continental United States (CUS).

B. microti sensu stricto (BMSS): All Clade 1 samples, i.e. the union of CUS and the sample from Russia.

Within-Host Evolution and Differences between Zoonotic and Enzootic Strains:

Among the 35 distinct BMSS strains studied (excluding the 3 serial samples of GI and 1 serial sample of RMNS), 32 were obtained from clinical cases of human babesiosis (plus the reference R1). We sequenced two samples from tick (PI-2000, SN-1988), which fell into the MNE and REF group lineages, respectively. One BMSS strain (MYS/Russia) was from a vole (*Clethrionomys glareolus*). Both PI-2000 and SN-1998 were phylogenetically situated within major lineages and were indistinguishable by nucleotide or structural differences from zoonotic samples, although the study was not powered to detect differences between zoonotic and enzootic samples. The three lineages that have been detected in the Northeast by VNTR typing[13] were identified in ticks, and we see the same lineages in clinical cases. Thus, it appears that all major enzootic lineages are capable of producing human disease, though whether certain lineages are more likely to do so remains is not yet known.

We sequenced serial samples of Bab16 over a course of 5 days. We did not observe any mutations. Together with the observation that only three mutations accumulated in GI and RMNS strains sampled over a period of 32 years of laboratory propagation, it appears rates of within-host evolution are slow, and this is consistent with inferred rates of mutation (Supplemental Figure 6D).

Time to Most Recent Common Ancestry:

Root-to-tip analysis of the Mainland New England and Nantucket lineages, for which samples were available over the time period 1969 – 2015, supported the existence of a molecular clock (Supplemental Figure 6a-b).

We used dates of collection at tip-dates to infer a rate of the molecular clock and to estimate the time of most recent common ancestry for the identified lineages. The sequence of Gray[1] (1969) and Peabody[14] (1973) strains was identical, which we attribute to either an extreme lack of genetic diversity during the 1960s and 1970s on Nantucket or to contamination of ATCC stocks at some point in the preservation or procurement process. Thus, for the purposes of BEAST analysis, only one of these genomes was considered and this was assigned a collection date of 1971 with an uncertainty of +/- 2 years. Chromosome 2 sequences were excluded from the analysis based on the uncertain ancestry of select regions of chromosome 2 (see below). The precise collection time of the R1 reference isolate[12] was not known and this was assigned a collection date of 2005 +/- 5 years. Following Drummond et al.[15], we compared strict molecular clock (CLOC) models, uncorrelated exponential (UCED) and log-normal (UCLN) relaxed clock models. We used an HKY84 model for nucleotide substitution with gamma-distributed rates with 4 site categories. We used the complete genomes including both coding and non-coding sequences since most *B. microti* DNA is coding, genes are densely spaced, and the assumption that the small intergenic regions are neutral is doubtful. We also fit codon-partitioned models on the concatenated coding sequence for all coding genes. We used an infinite uniform (improper) prior over the interval $[-\infty,\infty]$ for CLOC_IU model and a Gamma(4, $5x10^{-9}$) prior based on empirical substitution data for CLOC_G, UCED_G, and UCLN_G models. The choice of prior was based on observed mutation data for laboratory-propagated strains. The GI strain, which was continuously maintained for 28 years with tick and rodent passage, accumulated 2 SNPs over this interval. The RMNS strain, maintained similarly for 4 years, accumulated 1 SNP over this interval. If we model mutations as a Poisson process, $P(n) = (\lambda T)^n e^{-\lambda T}/n!$, for an observed n = 3 events observed over an interval of T = 32 years x 6395000 sites, the maximum likelihood estimate for λ is $n/T = 1.47x10^{-8}$ variants/site/year. The shape of this likelihood is closely approximated by a Gamma(4, $5x10^{-9}$) (Supplemental Figure 6C). We therefore chose Gamma(4, $5x10^{-9}$) as a prior, justified by observed rates of evolution in laboratory lines and encompassing a range of plausible rates from $5.5x10^{-9}$ to $4.5x10^{-8}$ (95%HPD). This prior closely matched the inferred substitution rate for MNE and NAN lineages (Supplemental Figure 6D) obtained using an uninformative (improper uniform) prior in both lineages independently.

The four models produced similar estimates for TMRCA (Figure 3B, Supplemental Figure 6, Supplemental Table 6). Importantly, estimates for CLOC_IU and CLOC_G45 were very similar, suggesting that the gamma(4,$5x10^{-9}$) is appropriate for these data and not resulting in aberrant or unreliable estimates. Generally, relaxed clock models were favored by model comparison using the harmonic mean estimator[16] and Akaike Information Criteria through MCMC [17] (Supplemental Table 6A-B). However, strict clock models may be more biologically appropriate as CUS samples are closely related members of a single species. We present models of both types. All models estimated TMRCA for NAN, REF, and MNE at between 40-600 years. TMRCA between NE and MW was earlier, between 300 – 5,000 years. Estimates for TMRCA placed divergence of CUS samples within the last 15,000 years. We estimated a median mutation rate of

1.92×10^{-8} / site / year [median; 95% HPD 3.84×10^{-9} – 3.53×10^{-8}] under the CLOC_IU model (Supplemental Figure 4) and of 1.85×10^{-8} / site / year [median; 95% HPD 8.01×10^{-9} – 3.00×10^{-8}] under the CLOC_G model. This was consistent among sublineages (MNE – 2.2 $\times 10^{-8}$ / site / generation [95% HPD 4.40×10^{-10} – 4.7×10^{-8}] and NAN - 2.2 $\times 10^{-8}$ / site / generation [95%HPD 2.89×10^{-9} – 4.35×10^{-8}]) analyzed independently under CLOC_IU models. This held for trees computed on NAN samples alone, MNE samples alone (lineages for which adequate numbers of reliably dated longitudinal samples were collected), and for the set of CUS samples (Supplemental Figure 6). Median and 95% HPD estimates for TMRCA are shown in Figures 3c-d. We also used a codon-partitioned model for aligned coding sequences, which produced similar results (Supplemental Figure 6G), with the exception of CLOC_IU models, which had slightly wider HPD intervals. We estimated divergence between CUS and Russian *B. microti,* but these estimates were less precise. Most models placed this divergence between 200,000 and 1.5 million years ago. The CLOC yielded a slightly higher upper bound (up to 2.6 million years ago). In cases where the CLOC_IU models produced wider HPD intervals, this is attributable to the model's allowance of mutation rates approaching zero, which are biologically implausible, and it is likely that inference under models with informative priors is more accurate. Overall, the estimates obtained from all models indicate 1) deep divergence between NE and MW lineages with subsequent radiation into sub-lineages (NAN, REF, MNE) along geographic lines 2) timing of the divergence among CUS samples consistent with proposed models of population isolation at the conclusion of the most recent ice age approximately 15,000 years ago[18] and 3) separation from Eurasian populations much earlier, hundreds of thousands to millions of years ago.

One potential weakness of these estimates is that they rest on a relatively simple demographic and mutational model and do not explicitly model changes in population size, geographic range, and recombination. As additional samples become available, these estimates may be refined. The absence of recombination in these models is not likely to have affected estimates for EC lineages, which do not show strong evidence of recombination (Supplemental Figure 5), particularly as chromosome 2 was excluded from BEAST analysis because of the unusually large proportion of alleles shared with MW samples (see below), but MW samples and estimates for CUS may be affected. Recombination would have a substantial effect on the TMRCA estimates if populations of different ages were interbreeding. However, had this happened, blocks of the genome within a given lineage would have a different mutational density than others, and we did not see evidence of such differences (Supplemental Figure 4E). Despite these limitations, the clear convergence to similar rates of evolution in independent lineages (Supplemental figure 6D), the concordance with the empirical molecular clock (Supplemental figures 6C-D) and consistency with proposed biogeographic models[18], all argue in favor of the utility of our approach and support the estimates of most recent common ancestry among *B. microti* lineages.

Discordant loci on chromosome 2

Two segments of chromosome 2, covering roughly positions 0 – 122 KB and 522 – 1000 KB, respectively, display unusual levels of genetic divergence between the MNE and

Nantucket samples. This is reproducible between samples collected at different times, sequenced in different batches, and at different sites, and cannot be attributed to an artifact of sequencing or variant calling. These regions, which amount to ~10% of the nuclear genome, contain the majority of fixed differences between the samples (61 out of 82); π measured between the two populations within the segments is 1.1×10^{-4}, but 8.8×10^{-6} in the remainder of the genome.

The same regions show much more allele sharing between the Midwest and either the MNE or the Nantucket samples than elsewhere in the genome. Table S2 breaks down the distribution of loci for which two of these populations share an allele and the third has a different allele. In the genome as a whole, more than 98% of the time the different allele occurs in the Midwest. In the chromosome 2 regions, however, the different allele is far more likely to occur in one of the eastern populations, with similar levels appearing in Nantucket and MNE. The reason for this signature is not entirely clear. It is possible that these regions represent introgression, i.e. acquisition of ancestral sequence to promote survival or fitness, but the presence of two distinct regions argues against a single event. Further analysis of this region as additional samples become available may shed light on the unusual pattern of variation seen on this chromosome.

Analysis of Recombination:

Electron micrograph data have suggested that *Babesia microti* undergoes sexual development [19]; however, nuclear fusion has not observed, nor have recombinant organisms been demonstrated. We applied the pairwise homoplasy index[20] (PHI) to search for recombination in our samples. This revealed strong evidence of recombination in CUS samples and also BMSS. There was no evidence of recombination in the MNE or NAN lineages, and only on chromosome 4 in NE samples (Figure S5A). We attempted to localize signals of recombination by applying the PHI test within 100Kb windows of each chromosome. This analysis revealed a single localizable signal within the BMSS group, on chromosome 4, and equivocal signals on chromosomes 1 and 3. No other statistically significant regions were identified. Thus, we find evidence of recombination in *B. microti*, as evidence by an excess of incompatible sites[20], but this does not appear to have strongly shaped variation within recently diverged lineages.

Multi-copy gene families and genomic distribution of variants:

B. microti possesses a unique multi-copy gene family, BMN[12,21]. These genes occur at chromosome ends in subtelomeric regions as well as internal clusters and resemble multi-copy gene families of other bacteria and parasites. Multi-copy gene families are common among pathogens[22], particularly eukaryotic parasites[23-26]. Within the Apicomplexa, such multi-copy gene families are found in *Plasmodia, Babesia, Crytosporidia, and Eimeria,* and have recently been reviewed in depth by Reid[27]. Members of these gene families are often rapidly evolving through accelerated mutation, recombination, and gene conversion. For example, *P. falciparum var* genes*,* encoding

the PfEMP-1 proteins, mediate antigenic variation through a mutually exclusive expression mechanism [28]. Within *Babesia,* Variant Erythrocyte Surface Antigen (VESA) genes in *B. bovis* [29-31]*,* function in a manner similar to *var* genes in *P. falciparum* [31]*.*

The function of BMN genes in *B. microti* is not known, but they are postulated to contribute to chronic infection or immune evasion through differential expression or recombination[21], though their degenerate repeat structure has raised speculation that their primary function is structural[21]. Among CUS samples, we did not observe an increased substitution rate among BMN gene families (P = 0.7, Wilcoxon Rank-Sum test), but diversity was extremely limited among these samples, which limits the power of this comparison. Comparing Russian *B. microti* to the R1 reference, we found evidence of an accelerated substitution rate in BMN genes (nucleotide diversity was a mean of 1.9 fold greater than other coding sequences; P = 9.8 $x10^{-3}$, Wilcoxon Rank-Sum test, two-tailed alternative, Supplemental Figure 8F). BMN genes also had elevated dN/dS ratios when compared to the genome as a whole (Supplemental Figure 8E; P = 1.26 $x10^{-5}$, Wilcoxon Rank-Sum test, two-tailed alternative). BMN genes in which zero variants were called were excluded from this analysis due to the likelihood that the reference sequence was too divergent to align reads. Not all members of the BMN family were under the same selective pressure, with BBM_I0004 and BBM_I03513 showing the strongest evidence of increased substitution and positive selection. BBM_II01570 appeared to show weak evidence of negative selection, although manual inspection of this locus revealed a drop in coverage and a small number of called variants, suggesting that sequence may have diverged so substantially in this region that variants cannot be called with short reads. Supporting this, in the analysis of unfiltered variants (Supplemental Figure 7 and Supplemental Table 3), highly substituted regions contained multiple members of the BMN family. In general, resequencing approaches using short reads likely underestimate the true diversity in the BMN family, much of which may be generated by recombination or gene conversion [32-34].

This issue likely affected other multi-copy gene families as well. *B. microti* contains three *vesa*-like genes and four *Theileria parva tpr*-like genes in mosaic subtelomeric structures[12]. The three VESA-like genes and one TPR-like protein (BBM_III04845) were so highly polymorphic that we were unable to identify any variants. Three additional TPR-like proteins (BBM_II04270, BBM_III00015, and BBM_I00005) had limited diversity; inspection of the aligned sequence found only small islands of short reads aligning with confidence allowing for variant detection, although those that aligned for BBM_I0005 did show an excess of non-synonymous variants, providing weak evidence of positive selection (adjusted P = 0.22). Furthermore, both BBM_II04270 and BBM_III00015 were found in highly substituted regions as identified by the analysis of unfiltered variants (Supplemental Table 3). Thus, given their multi-copy nature, presence at chromosome ends, and our inability to align short reads reliably to these sequences, we suspect that members of both families are also hyper-variable, perhaps more so than BMN.

Clinical details of Relapsing Cases:

Bab05: The patient from whom Bab05 was isolated was a 48 year-old woman with a severe case of babesiosis (16% parasitemia on first admission). Her past medical history was notable for cystic fibrosis status post double lung transplantation, idiopathic thrombocytopenic purpura status post splenectomy, and diabetes. Her home medications included trimethoprim/sulfamethoxazole (TMP/SMX), mycophenolate 750 mg po three times daily, prednisone 15 mg po daily, and tacrolimus 4mg po twice daily. She presented with abdominal pain in July 2014 and was found to have babesiosis due to *B. microti* infection (initial parasitemia at with 16%). She was initiated on atovaquone 750 mg po twice daily and azithromycin 500 mg po daily; her TMP/SMX was stopped at that time. Testing for Lyme antibodies (IgG and IgM) was negative, and PCR assays to *Anaplasma phagocytophilum*, and *Ehrlichia spp.* (*E. Chaffeensis, E. Ewingii, E. muris*-like) were all negative.

She underwent RBC exchange transfusion on hospital day (HD) 2. Her parasitemia sharply declined, clinical status improved, and she was discharged on HD9 on atovaquone and azithromycin with a plan to continue on these medications for two weeks until her parasitemia cleared or at least 6 weeks. Clearance of microscopic parasitemia was documented 91 days after initial presentation. She continued on this regimen for 110 days after her initial presentation, at which time atovaquone was interrupted and the dose of azithromycin reduced to 250 mg po daily, for the purposes of parasite suppression and *Mycobacterium avium* complex (MAC) prophylaxis. 21 days later, she was readmitted with fevers and malaise. Parasitemia was at 21.7%, at which point a sample study was collected. She was initiated on an antibiotic regimen of clindamycin 600 mg po every 6hrs, atovaquone 750 mg po twice daily, and azithromycin 500 mg po daily. She underwent exchange transfusion on day 129 and day 135, and her parasitemia decreased, with clinical improvement. Her parasitemia became undetectable by PCR on day 255, at which point clindamycin was discontinued (after a total of 124 days), and the dose of azithromycin was reduced to 250 mg po daily, and atovaquone was changed to 1500mg po daily.

Bab 14: Bab14 was isolated from a 77 year-old man with a past medical history notable diffuse large B cell lymphoma and IgG4-related disease for which he had been treated with rituximab in April 2015. He was found to have babesiosis in August 2015 and treated with a 10-day course of atovaquone and azithromycin. He returned to care in November 2015 complaining of severe fatigue, at which time his Babesia PCR was noted to be positive (record of quantitative parasitemia not available) and he was restarted on atovaquone and azithromycin after transfusion of four units of packed red blood cells. Six days after re-initiation of treatment with atovaquone/azithromycin, his parasitemia was noted to 4.5% and his hematocrit 21. He received two additional units of packed red blood cells and was started on clindamycin/quinine. His parasitemia was 3.8% two days after starting clindamycin/quinine and then 1% five days after starting, at which time a study sample was obtained. He experienced ototoxicity while on quinine was transitioned to atovaquone/azithromycin after five days of clindamycin/quinine. After restarting atovaquone/azithromycin, his parasitemia decreased to 0.5% after two days and then was less than 0.1% after 25 days and was microscopically undetectable 38

days later; his treatment was discontinued 40 days after transitioning back to atovaquone/azithromycin. Two months after discontinuing treatment he returned to care and was again found to have a positive Babesia PCR (no report of parasitemia available), at which time was restarted on atovaquone/azithromycin with a plan for ongoing monitoring by PCR and blood smear.

MGH2001 has been reported previously[35]. **BWH2003 and MORNS2015** were noted by clinical providers to be relapsing cases and propagated in Hamsters to establish laboratory isolates. As a result of the protocol under which they were obtained, additional details of human infection beyond their relapsing status were unavailable.

Possible evidence for locally imported babesiosis in Bab14:

Bab14 was isolated from a resident of South Dennis, MA. The Bab14 parasite is separated from the Gray[1]/Peabody[36] strains, isolated on Nantucket in 1969 and 1973 respectively. The patient had not travelled to Nantucket in over 10 years. While he had received blood transfusions as a part of his treatment for relapsed babesiosis, his diagnosis of babesiosis preceded transfusion. The presence of a Nantucket group parasite in this patient suggests the possibility that Nantucket group parasites have recently established a focus on mainland MA, or conversely that Nantucket group parasites are dispersed over a region that includes portions of mainland MA, and that the Gray/Peabody parasites were themselves imported to Nantucket some time after their divergence from other Nantucket-group parasites (i.e. RMNS, GI, BWH-2014), which differ from the Gray/Peabody/Bab14 group by 50 – 70 SNPs.

Copy Number Amplification containing *B. microti* MRP:

In the Bab05 case, we also noted a three-fold amplification of a 15KB region on chromosome 2 (658,075-672,981); this region includes the gene BBM_II01855 (Supplemental Figure 13), which encodes an ABC transporter with homology to multidrug resistance-associated proteins (MRPs). Copy number variation is an established mechanism of drug resistance in *P. falciparum,* most notably the ABC transporter MDR1[37,38]. The *P. falciparum* homolog of MRP is known to influence quinine and chloroquine susceptibility in *P. falciparum*[39]*.* Given the uniqueness of this event in a case of severe relapse, we hypothesize that copy number variants in bmMRP may contribute to *B. microti* survival during atovaquone or azithromycin treatment.

Supplementary Table 1:

Sample Name	Source	Site	Origin	Year	Enrichment Method	Coverage	VNTR Type
Bab01	Patient	MGH	Natick, MA	2014	Cellulose	16.0 (114.0)	CT/RI
Bab02	Patient	MGH	Millis, MA	2014	Cellulose	60.6 (92.6)	CT/RI
Bab03	Patient	MGH	Topsfield, MA	2014	Cellulose	27.2 (38.6)	CT/RI
Bab04	Patient	MGH	Bedford, MA	2014	Cellulose	8.9 (14.5)	CT/RI
Bab05	Patient	MGH	Lyndeborough, NH	2014	Cellulose	597.2 (882.7)	CT/RI
Bab06	Patient	BWH	N/A	2015	None	44.0 (65.3)	N/A
Bab07	Patient	Faulkner	N/A	2015	Cellulose	23.0 (35.8)	N/A
Bab08	Patient	BWH	N/A	2015	None	87.8 (447.1)	N/A
Bab10	Patient	BWH	Stonington, CT	2015	Cellulose	209.1 (277.7)	N/A
Bab11	Patient	MGH	Norwell, MA	2015	None	107.4 (153.8)	N/A
Bab12	Patient	BWH	N/A	2015	Cellulose	45.8 (71.8)	N/A
Bab13	Patient	MGH	Kennebunk, ME	2015	None	28.7 (46.6)	N/A
Bab14	Patient	MGH	South Dennis, MA	2015	None	12.9 (23.1)	N/A
Bab15	Patient	MGH	Winchester, MA	2015	None	96.6 (140.6)	N/A
Bab16	Patient	MGH	Gloucester, MA	2015	None	79.9 (105.5)	N/A
UMMS1	Patient	UMMS	N/A	2014	None	187.2 (212.5)	N/A
UMMS2	Patient	UMMS	N/A	2014	WGB HS	421.2 (778.6)	N/A
UMMS3	Patient	UMMS	N/A	2014	WGB HS	521.9 (749.4)	N/A
UMMS4	Patient	UMMS	N/A	2014	WGB HS	691.1 (1120.4)	N/A
UMMS5	Patient	UMMS	N/A	2014	WGB HS	974.5 (1389.4)	N/A
ND11	Patient	Mayo	North Dakota	2003	None	32.3 (41.6)	N/A
WI07	Patient	Mayo	Wisconsin	2002	None	10.5 (17.5)	N/A
MNBO10	Patient	Mayo	Minnesota	2005	None	11.1 (17.6)	N/A
Gray	HSPS	N/A	Nantucket	1969	None	45.0 (67.8)	Nan
Peabody	HSPS	N/A	Nantucket	1973	None	49.9 (74.0)	Nan
GI1986	HSPS	N/A	Nantucket	1986	None	9.4 (17.6)	Nan
GI1990	HSPS	N/A	Nantucket, MA	+4yrs	None	60.67 (82.5)	Nan
GI2004	HSPS	N/A	Nantucket, MA	+18 yrs	None	66.1 (102.6)	Nan
GI2014	HSPS	N/A	Nantucket, MA	+28 yrs	None	17.9 (23.1)	Nan
RMNS1997	HSPS	N/A	Nantucket, MA	1997	None	111.2 (183.3)	Nan
RMNS2001	HSPS	N/A	Nantucket, MA	+4 yrs	None	217.3 (274.5)	Nan
MGH2001	HSPS	MGH	N/A	2001	None	119.2 (166.2)	N/A
MN-1	HSPS	Minnesota	N/A	1995	None	89.2 (172.8)	N/A
BWH2014	HSPS	NCH/BWH	N/A	2014	None	149.2 (216.6)	N/A
BWH2003	HSPS	BWH	N/A	2003	None	34.4 (49.0)	N/A
MORNS	HSPS	MWH	N/A	2015	SureSelect	185.6 (205.7)	N/A
PI2000	Tick	N/A	Prudence Island, RI	2000	SureSelect	220.5 (3312.1)	N/A
SandyNeck	Tick	N/A	Sandy Neck, MA	1998	SureSelect	298.8 (449.2)	SN
MYS-Russia	Vole	N/A	Ural Mountains, Russia	1998	None	20.4 (51.4)	N/A
CR400	Vole	N/A	Alaska	2001	None	De novo assembly	N/A
AW1	HSRS	N/A	Awaji, Japan	2001	None	De novo assembly	N/A
Hobetsu	HSRS	N/A	Hobetsu, Japan	2001	None	De novo assembly	N/A

List of strains with date and place of origin, enrichment method (if applicable) and coverage (mean and standard deviation). The Gray[1], Peabody[36], and GI[40] strains were isolated from the initial cases of babesiosis reported on Nantucket. MN-1 was isolated from a Minnesota resident [21]. PI2000 and SN1988 were collected from Prudence Island and Sandy Neck (on Cape Cod, MA)[13]. Mys-Russia was isolated from the Ural Mountains[9]. CR400 was isolated from Alaska[8], and AW-1 and Hobetsu were isolated from Japan.[5]

Abbreviations: HSPS = Hamster Strain from Patient Sample; HSRS = Hamster strain from rodent sample. BWH = Brigham and Women's Hospital; MGH = Massachusetts General Hospital; NCH = Nantucket Cottage Hospital; MWH = Melrose Wakefield

Hospital; UMMS = University of Massachusetts Medical School; WGB HS = Whole Genome Bait Hybrid Select. CT/RI = Connecticut/Rhode Island. SN = Sandy Neck. PI = Prudence Island. VNTR = variable nucleotide tandem repeat.

Supplementary Table 2: Discordant Loci on Chromosome 2

Discordant Loci	Genome-Wide	Chromosome 2 Regions
MNE allele = Nantucket allele	1202	64
Midwest allele = MNE allele	5	35
Midwest allele = Nantucket allele	16	20

Table S2. Number of loci with discordant fixed alleles among the three CUS populations (MNE, Nantucket and Midwest), broken down by the populations that share an allele. The genome-wide values exclude the two regions on chromosome 2.

Supplemental Table 3: Chromosomal regions of elevated variation

Chromosome	Start	End	Genes Contained	Description
Chromosome 1	160007	161007	BBM_I00435	
Chromosome 1	547024	548024	BBM_I01510	
Chromosome 1	548024	549024		
Chromosome 1	939041	940041	BBM_I02590	solute carrier family 25 (mitochondrial carrier; adenine nucleotide
Chromosome 1	1303057	1304057	BBM_I03535	BMN family (1)
Chromosome 2	0	776	BBM_II00005	BMN family (1)
Chromosome 2	1776	2777	BBM_II00010	BMN family (2)
Chromosome 2	2777	3777	BBM_II00010	BMN family (2)
Chromosome 2	3777	4777	BBM_II00015	Vesa-like
Chromosome 2	292789	293789	BBM_II00780	
Chromosome 2	527800	528800	BBM_II01465	
Chromosome 2	1467841	1468841	BBM_II01485, BBM_II01490	
Chromosome 2	1470841	1471841	BBM_II01495	tRNA dimethylallyltransferase [EC:2.5.1.75]
Chromosome 2	1472841	1473841	BBM_II04200	exosome complex exonuclease DIS3/RRP44 [EC:3.1.13.-]
Chromosome 2	1480842	1481842	BBM_II04220	
Chromosome 2	1482842	1483842		
Chromosome 2	1483842	1484842		
Chromosome 2	1484842	1485842	BBM_II04225, BBM_II04230	BBM_II4230 - DDX54, DBP10, ATP-dependent RNA helicase DDX54/DBP10
Chromosome 2	1485842	1486842	BBM_II04230	DDX54, DBP10, ATP-dependent RNA helicase DDX54/DBP10
Chromosome 2	1486842	1487842		
Chromosome 2	1488842	1489842		
Chromosome 2	1490842	1491842	BBM_II04235	translation initiation factor IF-2
Chromosome 2	1492842	1493842	BBM_II04240	
Chromosome 2	1493842	1494842	BBM_II04245	
Chromosome 2	1495842	1496842	BBM_II04250, BBM_II04255	
Chromosome 2	1496842	1497842	BBM_II04255, BBM_II04260	BBM_II04260 - BMN family (1)
Chromosome 2	1497842	1498842	BBM_II04260	BMN family (1)
Chromosome 2	1498842	1499842	BBM_II04265	BMN family (2)
Chromosome 2	1501842	1502843	BBM_II04270	Tpr related protein, putative
Chromosome 2	1504843	1505843	BBM_II04280	BMN family (2)
Chromosome 2	1505843	1506843	BBM_II04280	BMN family (2)
Chromosome 2	1506843	1507843		
Chromosome 2	1507843	1508843	BBM_II04285	
Chromosome 3	3458	4458	BBMIII_00015	Tpr related protein, putative
Chromosome 3	4458	5458	BBMIII_00020	BMN family (2)
Chromosome 3	6458	7458	BBMIII_00020	BMN family (2)
Chromosome 3	90462	91462	BBMIII_00255	BMN family (2)
Chromosome 3	91462	92462	BBMIII_00255	BMN family (2)
Chromosome 3	93462	94462		
Chromosome 3	272470	273470	BBMIII_00785	BMN family (1)
Chromosome 3	280470	281470		
Chromosome 3	307471	308471	BBMIII_00885	
Chromosome 3	458478	459478	BBMIII_01295	CCR4-NOT transcription complex subunit 1
Chromosome 3	1232512	1233512		
Chromosome 3	1459522	1460522	BBMIII_04060	
Chromosome 3	1460522	1461522	BBMIII_04065, BBMIII_04070	BBMII_04070 - splicing factor 45
Chromosome 3	1743534	1744535	BBMIII_04820, BBMIII_04825	BBMIII_04825 - DNA-directed RNA polymerase II subunit A [EC:2.7.7.6]
Chromosome 3	1751535	1752535		
Chromosome 4	126132	127132	BBMIII_05195	
Chromosome 4	535150	536150		
Chromosome 4	1011171	1012171	BBM_III07535	N1-15 protein, maltese-cross seroactive antigen
Chromosome 4	1012171	1013171	BBM_III07535	N1-15 protein, maltese-cross seroactive antigen
Chromosome 4	1013171	1014171	BBM_III07535	N1-15 protein, maltese-cross seroactive antigen
Chromosome 4	1014171	1015171	BBM_III07535	N1-15 protein, maltese-cross seroactive antigen
Chromosome 4	1015171	1016171	BBM_III07535	N1-15 protein, maltese-cross seroactive antigen
Chromosome 4	1016171	1017171	BBM_III07535	N1-15 protein, maltese-cross seroactive antigen
Chromosome 4	1017171	1018171	BBM_III07535	N1-15 protein, maltese-cross seroactive antigen
Chromosome 4	1018171	1019171		
Chromosome 4	1221180	1222180	BBM_III08155	
Chromosome 4	1222180	1223180	BBM_III08155	
Chromosome 4	1223180	1224180	BBM_III08155, BBM_III08157	
Chromosome 4	1549195	1550195	BBM_III09150	
Chromosome 4	1813206	1814206	BBM_III09975	BMN family (1)
Chromosome 4	1815206	1816206	BBM_III09980	BMN family (2)

Supplemental Table 4: dN/dS ratios

Strongest Positive Selection:

Gene	dN	dS	dNdS	Z	P	Adjusted.P	Description
BBM_I00435	0.066647441	0.028337435	2.351922163	2.743518253	0.006078467	0.035782867	
BBM_II04205	0.048260353	0	0	2.320506431	0.020313497	0.070133592	
BBM_III03095	0.006793583	0	0	2.228442553	0.025851018	0.080798922	
BBM_I03535	0.055619019	0.010544974	5.274457541	2.148390578	0.031682741	0.090134164	BMN family (1)
BBM_I00004	0.072710427	0.008243002	8.820867387	2.114578873	0.034465856	0.093692284	BMN family (1)
BBM_III00955	0.042120083	0.00580892	7.250931531	2.029297821	0.042427965	0.105539942	
BBM_II02785	0.019279497	0	0	1.980506389	0.047646656	0.11259956	
BBM_III01305	0.035225402	0.009937892	3.544554835	1.796820985	0.072364044	0.143492551	
BBM_III00255	0.053908407	0.034401947	1.567016182	1.712735688	0.086761176	0.159933645	BMN family (2)
BBM_III08865	0.017707341	0	0	1.628289273	0.103463563	0.177533489	
BBM_III03830	0.021225398	0	0	1.620646409	0.105093493	0.179446753	
BBM_I01000	0.018447415	0.004759799	3.875670752	1.606824426	0.108092889	0.18313258	
BBM_III00960	0.053935564	0.025169206	2.142918822	1.597259478	0.110207879	0.185543217	
BBM_I01510	0.080061241	0.048956297	1.635361446	1.515057934	0.129757733	0.20781512	
BBM_II04270	0.004628382	0	0	1.449321589	0.147247795	0.228372477	Tpr related protein, putative
BBM_III05650	0.006568239	0	0	1.409551396	0.158672187	0.24055249	
BBM_III07835	0.007233399	0	0	1.409077296	0.158812314	0.240635179	
BBM_II01360	0.009367956	0	0	1.407553438	0.159263343	0.240674559	
BBM_II01000	0.016282946	0	0	1.406659591	0.159528354	0.240702724	
BBM_III04145	0.011364126	0	0	1.406125034	0.159687	0.240704669	
BBM_I02865	0.028380027	0.010308514	2.75306688	1.397410523	0.162290157	0.242847996	
BBM_II01310	0.030778946	0	0	1.392066638	0.16390222	0.243989347	
BBM_I02960	0.074106745	0.024514344	3.022995213	1.346992364	0.177982692	0.25673414	
BBM_II02595	0.021664146	0	0	1.321421585	0.186360833	0.264656684	
BBM_III08115	0.035340247	0	0	1.317533019	0.187660021	0.265851696	
BBM_III03105	0.016411996	0.004650285	3.529245535	1.287934215	0.197768864	0.275579565	
BBM_III09660	0.054595448	0.02363225	2.310209479	1.253055443	0.21018553	0.288157582	
BBM_II01405	0.025553536	0	0	1.17657713	0.23936433	0.319244046	
BBM_I00955	0.044164741	0.018293871	2.414182315	1.172436736	0.241021758	0.321331609	
BBM_II02130	0.032136944	0	0	1.147037298	0.251366206	0.332074007	
BBM_III08265	0.023740606	0	0	1.144409261	0.252453949	0.332880065	
BBM_III01535	0.029255535	0.009933494	2.945140518	1.133030949	0.257201257	0.337479812	
BBM_III02900	0.022191581	0	0	1.078183498	0.280951875	0.361015475	
BBM_III06360	0.018339842	0.007118774	2.57626422	1.075075221	0.282341052	0.361749473	N-acetylglucosaminylphosphatidylinositol deacetylase [EC:3.5.1.89]
BBM_III03305	0.01358998	0.0058794	2.311457049	1.071932757	0.283750236	0.3630926	
BBM_II04015	0.027347286	0.008629285	3.169125334	1.071808409	0.283806095	0.3630926	
BBM_III00885	0.012051583	0.005816191	2.072074776	1.052232596	0.292692848	0.371056593	
BBM_II03720	0.008867936	0	0	1.034481769	0.300910995	0.378993429	
BBM_II03840	0.01392351	0	0	1.025183662	0.305276536	0.383798242	
BBM_III04300	0.006050914	0	0	1.011830694	0.311619011	0.389191896	
BBM_II04280	0.001124228	0	0	0.999437518	0.317582793	0.394431943	BMN family (2)
BBM_I02025	0.004081655	0	0	0.997954326	0.318301507	0.395024765	
BBM_I02355	0.006451702	0	0	0.996762059	0.318880019	0.395024765	
BBM_II00585	0.008196905	0	0	0.995882054	0.319307456	0.395024765	
BBM_III08560	0.015626272	0	0	0.992116365	0.321140764	0.395024765	
BBM_III06505	0.012233029	0.004462896	2.741051876	0.986438045	0.323918204	0.39622848	
BBM_I02345	0.027119277	0.012776338	2.122617456	0.968861846	0.332614124	0.403888579	
BBM_III00950	0.024534584	0.016089181	1.524911964	0.963947559	0.335072218	0.406024576	deoxyhypusine monooxygenase [EC:1.14.99.29]
BBM_I02820	0.014253586	0	0	0.96052078	0.336793178	0.407119051	
BBM_III07225	0.063603915	0.028792247	2.20906396	0.949253768	0.342491561	0.412518188	

Strongest Purifying Selection:

Gene	dN	dS	dNdS	Z	P	Adjusted.P	Description
BBM_III05345	0.005977652	0.057142586	0.104609403	-8.374889449	0	0	
BBM_II00600	0.006174884	0.053529571	0.115354633	-8.314808196	0	0	E3 ubiquitin-protein ligase HUWE1 [EC:6.3.2.19]
BBM_III07380	0.006950063	0.043083465	0.161316247	-7.587982039	3.24E-14	3.77E-11	
BBM_III02770	0.000962636	0.03854262	0.024975893	-7.193801512	6.30E-13	5.49E-10	pre-mRNA-processing factor 8
BBM_III00050	0.004138043	0.050623641	0.081741313	-6.922213694	4.45E-12	3.10E-09	
BBM_III01570	0.003455963	0.048836085	0.070766581	-6.482868183	9.00E-11	5.23E-08	BMN1-5B
BBM_III04825	0.001657985	0.049174278	0.033716513	-6.336763948	2.35E-10	1.17E-07	DNA-directed RNA polymerase II subunit A [EC:2.7.7.6]
BBM_III00735	0.008407229	0.051829167	0.162210378	-6.232298657	4.60E-10	2.00E-07	
BBM_III00165	0.008616785	0.047383162	0.181853319	-6.15125625	7.69E-10	2.98E-07	
BBM_III06375	0.010341525	0.041245637	0.250730158	-5.771876467	7.84E-09	2.44E-06	
BBM_I00830	0.007992112	0.054522631	0.146583392	-5.764871776	8.17E-09	2.44E-06	
BBM_III08600	0.017156599	0.05503755	0.311725339	-5.759873623	8.42E-09	2.44E-06	
BBM_III01825	0.010170009	0.046231958	0.219977898	-5.553132248	2.81E-08	7.52E-06	
BBM_III04060	0.008776307	0.048519883	0.180880637	-5.424939017	5.80E-08	1.35E-05	
BBM_I01550	0.006469062	0.043987181	0.147066978	-5.367119957	8.00E-08	1.74E-05	ubiquitin carboxyl-terminal hydrolase 7 [EC:3.1.2.15]
BBM_III02575	0.005859245	0.045479917	0.128831471	-5.343788976	9.10E-08	1.87E-05	
BBM_III03720	0.003134315	0.03959305	0.079163269	-5.329088276	9.87E-08	1.91E-05	myosin ATPase [EC:3.6.4.1]
BBM_III04225	0.02645499	0.118083799	0.224035728	-5.284076192	1.26E-07	2.32E-05	
BBM_III04180	0.006804056	0.081812818	0.083166143	-5.256016023	1.47E-07	2.38E-05	
BBM_III00795	0.005385592	0.048517274	0.111003599	-5.252411581	1.50E-07	2.38E-05	
BBM_III09210	0.008938516	0.035515694	0.25167793	-5.249633042	1.52E-07	2.38E-05	
BBM_III09715	0.005074705	0.033713098	0.150526213	-5.244354684	1.57E-07	2.38E-05	
BBM_III06700	0.010370491	0.047838157	0.216782821	-5.188252781	2.12E-07	3.08E-05	
BBM_III05535	0.013049775	0.035788342	0.364637587	-5.179088536	2.23E-07	3.11E-05	MDN1, REA1, midasin
BBM_I02860	0.001604837	0.030468912	0.0526713	-5.165930428	2.39E-07	3.21E-05	DNA-directed RNA polymerase II subunit B [EC:2.7.7.6]
BBM_III04235	0.02766337	0.119073772	0.232321275	-5.146077939	2.66E-07	3.43E-05	translation initiation factor IF-2
BBM_III01450	0.013996549	0.044950313	0.311378225	-5.133974868	2.84E-07	3.53E-05	
BBM_III05350	0.008156596	0.041787482	0.195192327	-5.123034697	3.01E-07	3.60E-05	
BBM_III06670	0.008217173	0.040671184	0.202039186	-5.114025117	3.15E-07	3.60E-05	DNA polymerase epsilon subunit 1 [EC:2.7.7.7]
BBM_I01085	0.004309301	0.045387811	0.094944011	-5.111170687	3.20E-07	3.60E-05	
BBM_III02890	0.010833752	0.057577734	0.188158711	-5.073584728	3.90E-07	4.25E-05	
BBM_III08125	0.005281424	0.044128847	0.119681902	-5.004413002	5.60E-07	5.80E-05	
BBM_III01095	0.002764964	0.067534598	0.040941443	-5.002383845	5.66E-07	5.80E-05	
BBM_III04195	0.003796458	0.114824551	0.033063124	-4.981805904	6.30E-07	6.27E-05	cleavage stimulation factor subunit 2
BBM_III00990	0.010931533	0.057680288	0.189519386	-4.895542589	9.80E-07	9.49E-05	
BBM_III01235	0.007022644	0.04354815	0.161261602	-4.888433614	1.02E-06	9.57E-05	DNA polymerase I [EC:2.7.7.7]
BBM_I02285	0.004628565	0.044513116	0.103982038	-4.865419513	1.14E-06	0.000104747	carbamoyl-phosphate synthase [EC:6.3.5.5]/ aspartate
BBM_III00980	0.006132812	0.07085804	0.086550681	-4.854561364	1.21E-06	0.000107815	26S proteasome regulatory subunit N1
BBM_I01860	0.011294601	0.048944126	0.230765204	-4.812092085	1.49E-06	0.000130129	
BBM_III01295	0.008153357	0.045659099	0.178570249	-4.802317185	1.57E-06	0.000133314	CCR4-NOT transcription complex subunit 1
BBM_III02935	0.011340995	0.052825739	0.214686909	-4.793939655	1.64E-06	0.000135697	
BBM_I03030	0.005063307	0.04326818	0.117021502	-4.71328378	2.44E-06	0.000197556	
BBM_I02740	0.003571542	0.045330955	0.078788138	-4.679248478	2.88E-06	0.000224402	structural maintenance of chromosome 4
BBM_III01960	0.00221525	0.033150103	0.066824816	-4.677948877	2.90E-06	0.000224402	intron-binding protein aquarius
BBM_III04140	0.004930402	0.041286296	0.119419831	-4.663479434	3.11E-06	0.000235546	chromodomain-helicase-DNA-binding protein 1 [EC:3.6.4.12]
BBM_I01620	0.009195155	0.046370718	0.198296589	-4.64422353	3.41E-06	0.000253113	
BBM_III04250	0.002134682	0.044342385	0.048140897	-4.620120799	3.84E-06	0.000278449	splicing factor 3B subunit 1
BBM_III04460	0.00093518	0.049703808	0.018815056	-4.524488554	6.05E-06	0.000430588	elongation factor EF-2 [EC:3.6.5.3]
BBM_III03480	0.008051411	0.057408998	0.140246497	-4.512957825	6.39E-06	0.00044559	
BBM_I01920	0.004588896	0.038844712	0.11813438	-4.480796194	7.44E-06	0.000508162	clathrin, heavy polypeptide

Supplementary Table 5: Highly substituted genes

A)

Gene	SNPs	Length	Diversity	Z	P	Adjusted.P	Description
BBM_II04245	60	687	0.087336245	9.991352954	1.66E-23	5.87E-20	
BBM_II04285	40	534	0.074906367	8.26855884	1.36E-16	2.39E-13	
BBM_I01510	92	1356	0.067846608	7.290068755	3.10E-13	3.64E-10	
BBM_III08185	8	123	0.06504065	6.901160123	5.16E-12	4.55E-09	
BBM_II04230	168	2712	0.061946903	6.472363426	9.65E-11	6.81E-08	DDX54, DBP10, ATP-dependent RNA helicase DDX54/DBP10
BBM_II04255	39	645	0.060465116	6.266986273	3.68E-10	1.93E-07	
BBM_II04250	60	993	0.060422961	6.261143469	3.82E-10	1.93E-07	
BBM_I00425	90	1539	0.058479532	5.991782224	2.08E-09	9.16E-07	
BBM_I00004	20	351	0.056980057	5.783953384	7.30E-09	2.86E-06	BMN family (1)
BBM_II04220	63	1128	0.055851064	5.627473744	1.83E-08	5.87E-06	
BBM_II00780	101	1842	0.054831705	5.486189491	4.11E-08	1.15E-05	
BBM_I03513	12	219	0.054794521	5.481035732	4.23E-08	1.15E-05	BMN family (2)
BBM_I02960	26	480	0.054166667	5.394014523	6.89E-08	1.74E-05	
BBM_III07225	10	192	0.052083333	5.105262329	3.30E-07	7.77E-05	
BBM_II04225	124	2388	0.051926298	5.083497088	3.71E-07	8.17E-05	
BBM_I00435	100	1947	0.051361068	5.005155636	5.58E-07	0.000115869	
BBM_II04235	135	2679	0.050391937	4.870833056	1.11E-06	0.000217874	translation initiation factor IF-2
BBM_III03565	27	549	0.049180328	4.702902714	2.56E-06	0.000476394	
BBM_II04115	5	102	0.049019608	4.680626749	2.86E-06	0.000504646	
BBM_III00255	145	2991	0.04847877	4.605666004	4.11E-06	0.000690924	BMN family (2)
BBM_III02125	12	255	0.047058824	4.408859978	1.04E-05	0.00166691	
BBM_III07964	18	387	0.046511628	4.333018088	1.47E-05	0.002256704	CCAAT-box DNA binding protein subunit B - CBFD_NFYB_HMF
BBM_I02465	36	789	0.045627376	4.210459901	2.55E-05	0.003747377	
BBM_II04240	56	1242	0.045088567	4.135780323	3.54E-05	0.004901018	
BBM_III08155	128	2841	0.045054558	4.131066699	3.61E-05	0.004901018	
BBM_I01980	51	1134	0.044973545	4.119838173	3.79E-05	0.004955482	
BBM_III02340	8	180	0.044444444	4.046504283	5.20E-05	0.006326423	
BBM_III09660	16	360	0.044444444	4.046504283	5.20E-05	0.006326423	
BBM_I03535	24	543	0.044198895	4.012470875	6.01E-05	0.007068175	BMN family (1)
BBM_III03785	44	1002	0.043912176	3.972731267	7.11E-05	0.008088604	
BBM_II01535	24	549	0.043715847	3.945519909	7.96E-05	0.008781367	
BBM_III00960	49	1125	0.043555556	3.923303346	8.73E-05	0.009340414	
BBM_I02095	13	303	0.04290429	3.833037314	0.000126571	0.013137299	
BBM_III05105	27	636	0.04245283	3.770464449	0.000162944	0.01597304	
BBM_III07935	10	237	0.042194093	3.734603178	0.000188011	0.017932223	
BBM_III09760	45	1068	0.042134831	3.72638949	0.000194242	0.018038964	BMN family (2)
BBM_III02305	39	933	0.041800643	3.68007063	0.000233169	0.021098838	
BBM_III04830	7	168	0.041666667	3.661501357	0.000250742	0.022121672	
BBM_I01335	10	243	0.041152263	3.590204519	0.000330419	0.02776303	
BBM_III05850	15	366	0.040983607	3.566828506	0.000361328	0.029654086	
BBM_III09960	15	375	0.04	3.430499601	0.000602471	0.047247104	
BBM_III03525	55	1380	0.039855072	3.410412492	0.000648647	0.049762502	
BBM_III07810	92	2322	0.039621016	3.37797207	0.000730225	0.054829019	
BBM_III00785	39	987	0.039513678	3.363094834	0.000770739	0.05666536	BMN family (1)
BBM_III09725	18	459	0.039215686	3.321792893	0.000894411	0.064415813	
BBM_I02825	63	1620	0.038888889	3.276498431	0.001051029	0.074181624	
BBM_III07375	5	129	0.03875969	3.258591318	0.001119668	0.077476659	
BBM_III02585	12	315	0.038095238	3.166497595	0.001542866	0.104707209	
BBM_III06235	85	2250	0.037777778	3.122497261	0.001793238	0.117191407	

B)

Gene	Diversity	Z	P	Adjusted.P	Description
BBM_mt00020	0.00170925	9.77344876	1.46E-22	8.61E-20	cytb, ubiquinol-cytochrome c reductase cytochrome b subunit
BBM_III09915	0.00153496	8.7351951	2.43E-18	9.54E-16	
BBM_III06145	0.00131976	7.45329945	9.10E-14	2.68E-11	histone H4
BBM_mt00035	0.0012674	7.14138573	9.24E-13	2.33E-10	cox3, cytochrome c oxidase subunit 3
BBM_III00690	0.00119514	6.71099922	1.93E-11	4.55E-09	
BBM_II02210	0.00109242	6.09908446	1.07E-09	1.79E-07	
BBM_III08185	0.00095648	5.28932235	1.23E-07	1.88E-05	
BBM_III04950	0.00092428	5.09748532	3.44E-07	4.67E-05	
BBM_III04080	0.00090747	4.99738126	5.81E-07	7.60E-05	
BBM_II00935	0.0008993	4.948682	7.47E-07	9.42E-05	glutaredoxin 3
BBM_II02225	0.00089498	4.92297356	8.52E-07	0.00010373	
BBM_III09795	0.00088951	4.89038078	1.01E-06	0.00011839	
BBM_III02625	0.00086507	4.74481459	2.09E-06	0.00023758	
BBM_I02275	0.00085318	4.67396817	2.95E-06	0.00032581	large subunit ribosomal protein L24e
BBM_II00295	0.00082457	4.50357352	6.68E-06	0.00071457	
BBM_I02065	0.00081507	4.44695322	8.71E-06	0.00090401	deoxyhypusine synthase [EC:2.5.1.46]
BBM_I02530	0.00078941	4.29409922	1.75E-05	0.00176857	
BBM_I01885	0.00076786	4.16574759	3.10E-05	0.00288202	
BBM_III00955	0.00068399	3.66618027	0.0002462	0.02227798	
BBM_I01335	0.0006822	3.65551027	0.00025667	0.02264479	
BBM_II02290	0.00067806	3.63083476	0.00028251	0.02431619	
BBM_II03255	0.00066576	3.55758863	0.00037427	0.031448	
BBM_III06620	0.00065243	3.47815505	0.00050488	0.0414352	
BBM_I00005	0.00064561	3.43752688	0.00058705	0.04708427	Tpr related protein, putative
BBM_III00785	0.00063572	3.3785905	0.00072858	0.05713721	BMN family (1)
BBM_III01850	0.00062854	3.33584351	0.00085041	0.06524125	
BBM_II03600	0.00062439	3.31112486	0.00092922	0.06977039	cyclin-dependent kinase regulatory subunit CKS1
BBM_II02815	0.00060908	3.21995769	0.0012821	0.09426071	
BBM_II01100	0.00057041	2.98957997	0.00279361	0.19717321	
BBM_II01460	0.00056325	2.94691331	0.00320963	0.22209392	
BBM_III00590	0.00055258	2.88339817	0.0039341	0.26195155	
BBM_II01055	0.00054711	2.85080772	0.00436083	0.27980691	
BBM_I01270	0.00053958	2.80591416	0.00501741	0.31618615	
BBM_III00080	0.00053476	2.77721638	0.00548267	0.3394444	prefoldin alpha subunit
BBM_III05115	0.00052627	2.72665361	0.00639802	0.38268813	
BBM_II01490	0.00052216	2.70213834	0.00688951	0.40468164	
BBM_I00955	0.00052131	2.69707917	0.00699506	0.40468164	
BBM_III01965	0.00051886	2.68250022	0.00730741	0.41593309	
BBM_II00310	0.00051388	2.65285931	0.00798131	0.44708021	
BBM_I02135	0.00051191	2.64108586	0.00826408	0.45068299	
BBM_III06860	0.00051165	2.6395733	0.00830105	0.45068299	
BBM_I02395	0.00050515	2.60082316	0.00930004	0.49727022	
BBM_II01845	0.00050383	2.59300234	0.00951421	0.50112919	
BBM_III05330	0.00050235	2.58415856	0.00976169	0.50660303	
BBM_II02610	0.00049894	2.56382126	0.01035268	0.52948729	
BBM_III03830	0.00049782	2.55720004	0.01055185	0.53196408	
BBM_III04280	0.00049546	2.54314148	0.01098607	0.54605427	
BBM_III06190	0.00049338	2.53072291	0.01138277	0.55791397	
BBM_II03800	0.00048901	2.50471441	0.01225503	0.58443245	DNA-directed RNA polymerase III subunit C11 [EC:2.7.7.6]
BBM_III01725	0.00048901	2.50471441	0.01225503	0.58443245	

Supplemental Table 5: A) List of the 50 most substituted genes comparing Russian *B. microti* to the R1 reference. Full table available in supplemental files. B) 50 most substituted genes within the CUS samples. Full table available in supplemental files.

Supplemental Table 6: BEAST estimates and Model Comparisons

A)

HME:

	ln P(model | data)	S.E.	CUS_CLOC_IU	CUS_CLOC_G	CUS_UCED_G	CUS_UCLN_G
CUS_CLOC_IU	-6618014.212	+/- 0.146	-	-0.542	-22.09	-20.713
CUS_CLOC_G	-6618012.963	+/- 0.148	0.542	-	-21.548	-20.171
CUS_UCED_G	-6617963.348	+/- 0.155	22.09	21.548	-	1.377
CUS_UCLN_G	-6617966.518	+/- 0.163	20.713	20.171	-1.377	-

AICM:

	AICM	S.E.	CUS_CLOC_IU	CUS_CLOC_G	CUS_UCED_G	CUS_UCLN_G
CLOC_IU	13236055.2	+/- 0.086	-	-4.707	-89.429	-78.319
CLOC_G45	13236050.4	+/- 0.077	4.707	-	-84.723	-73.613
UCED_G45	13235965.7	+/- 0.114	89.429	84.723	-	11.11
UCLN_G45	13235976.8	+/- 0.089	78.319	73.613	-11.11	-

B)

Summary Statistic	EC	MNE	MW	NAN	REF	CUS	CNE
mean	1649	219	2237	268	215	6316	410
stderr of mean	116	5	114	6	7	444	10
median	1293	177	1837	220	160	5043	336
geometric mean	1357	186	1868	231	171	5291	348
95% HPD lower	327	51	379	73	41	1341	99
95% HPD upper	3933	499	5100	583	543	14708	909
auto-correlation time (ACT)	145220	19046	88392	19516	24064	166960	23842
effective sample size (ESS)	124	945	204	922	748	108	755

A) Log10 Bayes factors (BF) based on the marginal likelihood method[16] and Akaike Information Criteria by Markov Chain Monte Carlo (AICM)[17] comparing strict clock (CLOC), uncorrelated exponential (UCED) and uncorrelated log-normal (UCLN) distributions for CUS samples. Positive values between compared models favor the model in the row over the model in the column. B) Summary statistics (in years) for time to most recent common ancestry by population for UCED (the favored model in A). Positive values between compared models favor the model in the row over the model in the column. HPD = highest posterior density.

Supplemental Table 7: Non-Synonymous Variants in Relapsing Cases. Each relapsing case was compared to its nearest neighbor (as measured by p-distance) and substitutions resulting in pairwise amino acid differences were identified. Variants in which the relapsing case contained the wild-type allele are marked as WT.

Sample	Gene	Chromosome	Base	Description	Amino Acid Substitution
Bab05	BBM_I00580	chromosome 1	210944	n/a	F115L
Bab05	BBM_I01510	chromosome 1	547252	n/a	Y323C
Bab05	BBM_I03030	chromosome 1	1106570	n/a	I1294T (WT)
Bab05	BBM_II02835	chromosome 2	1000673	n/a	S221P
Bab05	BBM_mt00020	mitochondrion	3648	cytochrome b	L277P (shared by MORNS2015)
Bab05	BBM_mt00035	mitochondrion	7430	cytochrome oxidase 3	H179Y
Bab05	rpl4	apicoplast	7864	ribosomal subunit L4	R86H
Bab14	BBM_I00115	chromosome 1	37648	n/a	
Bab14	BBM_I02075	chromosome 1	762763	inorganic pyrophosphatase	P264S
Bab14	BBM_I03445	chromosome 1	1264406	coatomer protein complex, subunit alpha (xenin)	A924V
Bab14	BBM_I03460	chromosome 1	1269400	n/a	D355G
Bab14	BBM_mt00020	mitochondrion	3177	cytochrome b	Y120C
Bab14	rpl4	apicoplast	7863	ribosomal subunit L4	R86C
MGH2001	BBM_I00213	chromosome 1	60729	Cystein Rich Modular Protein, CRMP	W335L
MGH2001	BBM_III02665	chromosome 3	943961	n/a	L21I
MGH2001	BBM_III03370	chromosome 3	1190774	n/a	V170A (WT)
MGH2001	BBM_III03470	chromosome 3	1225089	n/a	T300A
MGH2001	BBM_III06975	chromosome 4	813896	n/a	R655H (WT)
MGH2001	BBM_III07040	chromosome 4	849012	aspartate aminotransferase	P240L
MGH2001	BBM_mt00020	mitochondrion	3102, 3220	cytochrome b	I95S, M134I
BWH2003	BBM_I01325	chromosome 1	483564	n/a	D159N (shared by Bab15)
BWH2003	BBM_II01475	chromosome 2	536568	n/a	F1998L (WT)
BWH2003	BBM_II03485	chromosome 2	1222818	n/a	S141G
BWH2003	BBM_III06135	chromosome 4	493919	n/a	A326T
BWH2003	BBM_III06175	chromosome 4	505781	n/a	D319E (WT)
BWH2003	BBM_III07180	chromosome 4	898529	n/a	S227I (shared by Bab15)
BWH2003	BBM_III07380	chromosome 4	956965	n/a	T3445K
BWH2003	BBM_III07915	chromosome 4	37108, 11371	n/a	K475Q, I480N
BWH2003	BBM_III09545	chromosome 4	1669128	n/a	V82I
BWH2003	BBM_mt00020	mitochondrion	3648	cytochrome b	L277P
MORNS2015	BBM_I00213	chromosome 1	60729	Cystein Rich Modular Protein, CRMP	W335L
MORNS2015	BBM_I02415	chromosome 1	875249	n/a	E115K (shared by Bab03, Bab08, Bab10, Bab12, Bab16, MGH2001)
MORNS2015	BBM_I02455	chromosome 1	884392	TSR2, pre-rRNA-processing protein TSR2	W41C
MORNS2015	BBM_I03030	chromosome 1	1106570	n/a	I1294T (WT)
MORNS2015	BBM_II00950	chromosome 2	348600	n/a	L89I
MORNS2015	BBM_II02600	chromosome 2	921987	n/a	A283V (WT)
MORNS2015	BBM_III03100	chromosome 3	1110447	mRNA (guanine-N7-)-methyltransferase	L469I
MORNS2015	BBM_III04730	chromosome 3	1699962	n/a	T259M
MORNS2015	BBM_mt00020	mitochondrion	3237	cytochrome b	T140K
MORNS2015	rpl4	apicoplast	7825	ribosomal subunit L4	S73L

Supplemental Table 8: Mutations associated with atovaquone and azithromycin resistance in other Apicomplexa and bacterial species.

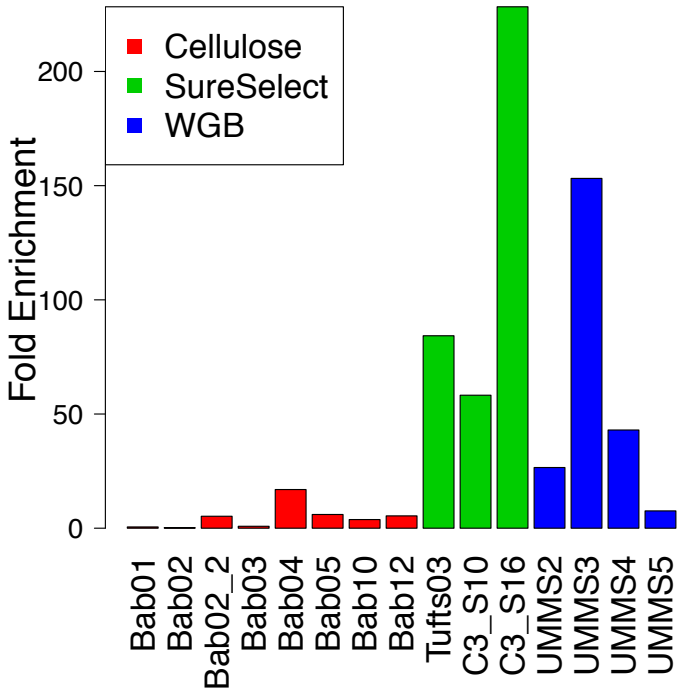
A. Cytochrome B mutations associated with atovaquone resistance

Species	Position	Corresponding Position in *B. microti*	Mutation	Reference
B. gibsoni	108	121	A>T	Sakuma et al. 2009[41]
B. gibsoni	121	134	M>I	Matsuu et al. 2006[42]
T. gondii	129	134	M>L	McFadden et al. 2000[43]
P. falciparum, P. berghei	133	134	M>I	Korsinczky et al. 2000[44]
P. berghei	144	145	L>(F,S)	Korsinczky et al. 2000
T. gondii	254	262	I>L	McFadden et al. 2000
P. yoelii	258	262	I>M	Korsinczky et al. 2000
P. yoelii	267	271	F>I	Korsinczky et al. 2000
P. falciparum, P. yoelii	268	272	Y>(S,C)	Korsinczky et al. 2000
P. yoelii	271	275	L>V	Korsinczky et al. 2000
P. falciparum, P. yoelii	272	276	K>R	Korsinczky et al. 2000
P. falciparum	275	279	P>T	Korsinczky et al. 2000
P. falciparum	280	284	G>D	Korsinczky et al. 2000

B. Ribosomal protein L4 mutations associated with azithromycin resistance.

Species	Position	Corresponding Position in *B. microti*	Mutation	Reference
E. coli	63	80	K>N, E	Sidhu et al. 2007[45], Chittum and Champney 1994[46]
S. pneumoniae	69	81	G>(C,T, V)	Sidhu et al. 2007
S. pneumoniae	70	82	T>P	Sidhu et al. 2007
S. pneumoniae	71	83	G>(S, R)	Sidhu et al. 2007
P. falciparum	76	83	G>V	Sidhu et al. 2007

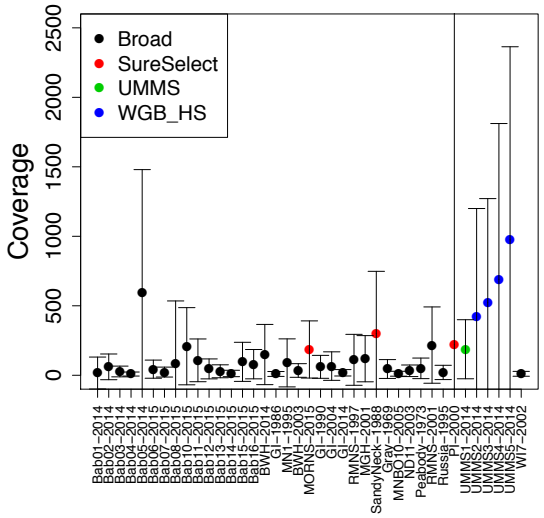
Supplementary Figure 1: Enrichment of *Babesia microti* DNA by three methods



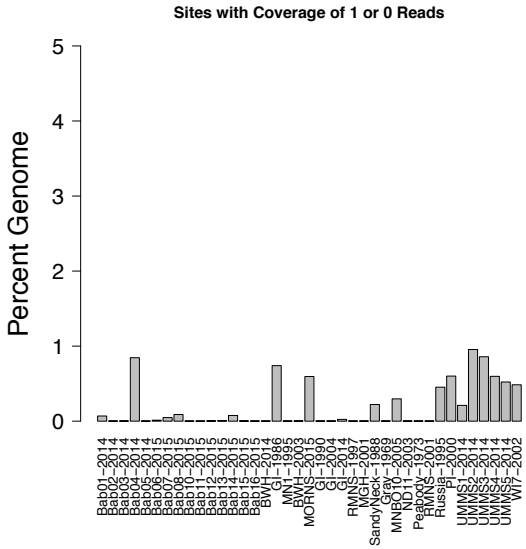
Fold enrichment by method for the strains which underwent an enrichment procedure. Bab02 and Bab02_2 denote samples separated by consecutive days of collection from the same patient. Short reads from this sample were pooled in the remainder of the analyses.

Supplementary Figure 2: Coverage Statistics

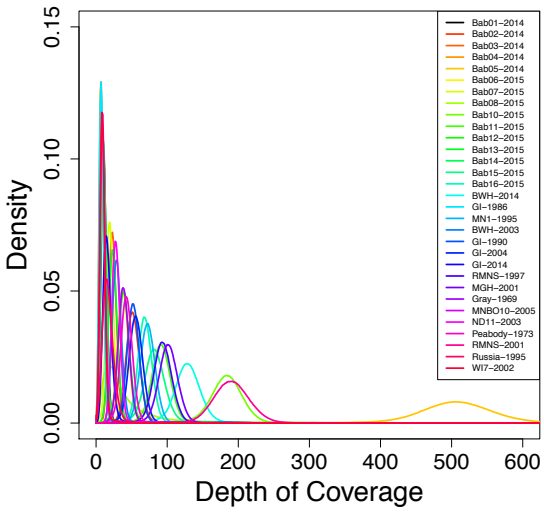
A)



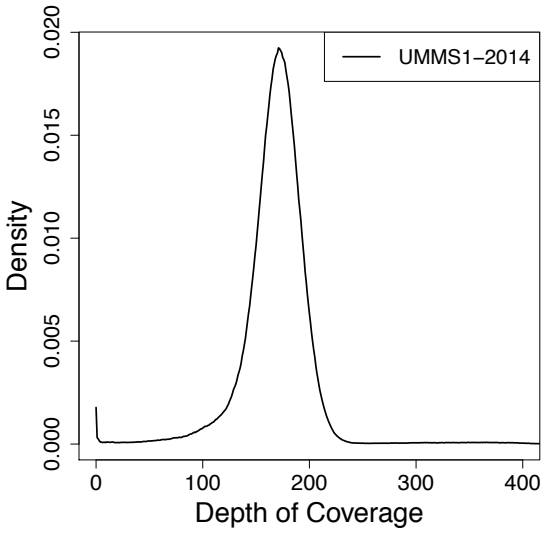
B)



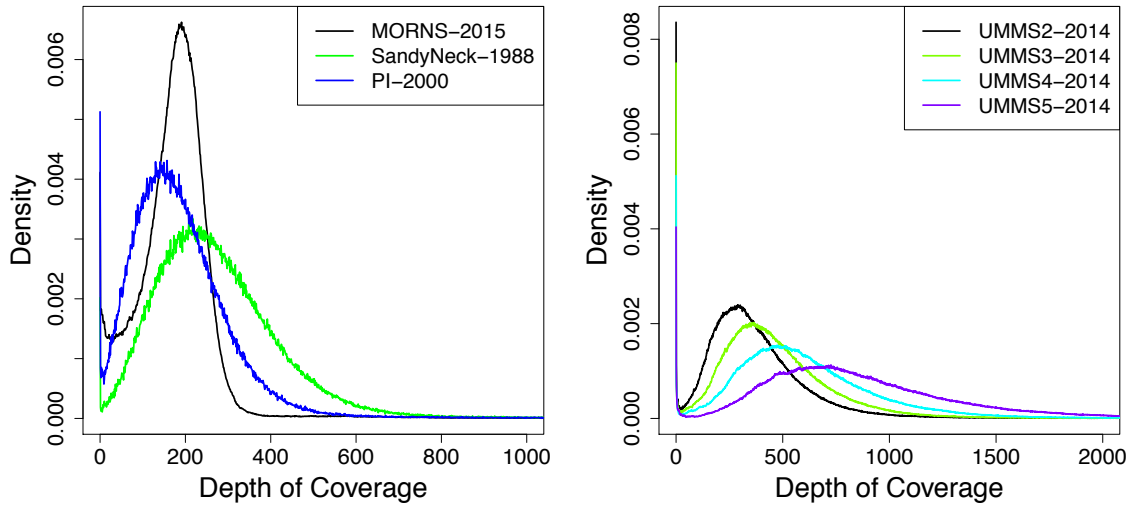
C)



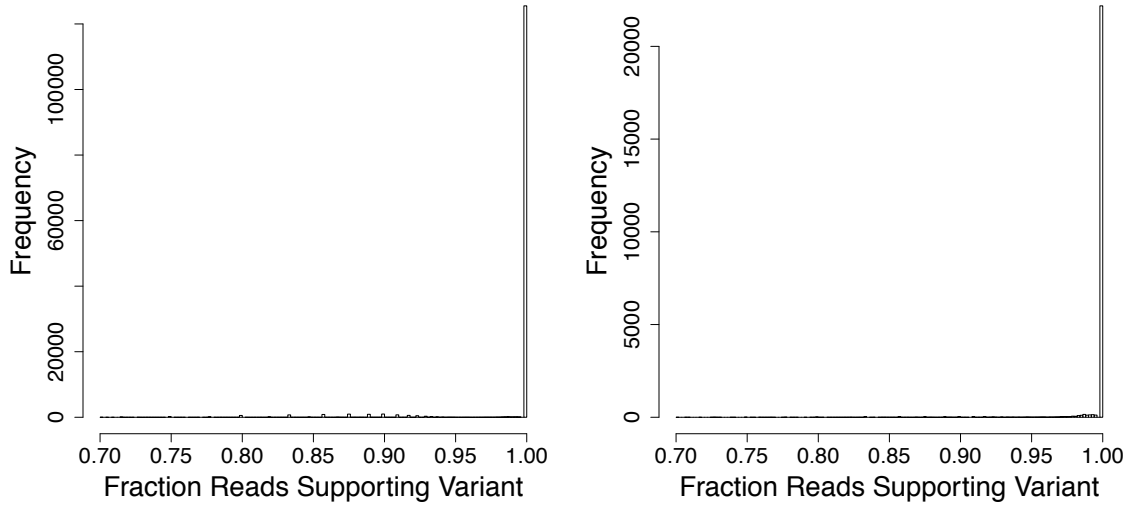
D)



E) **F)**

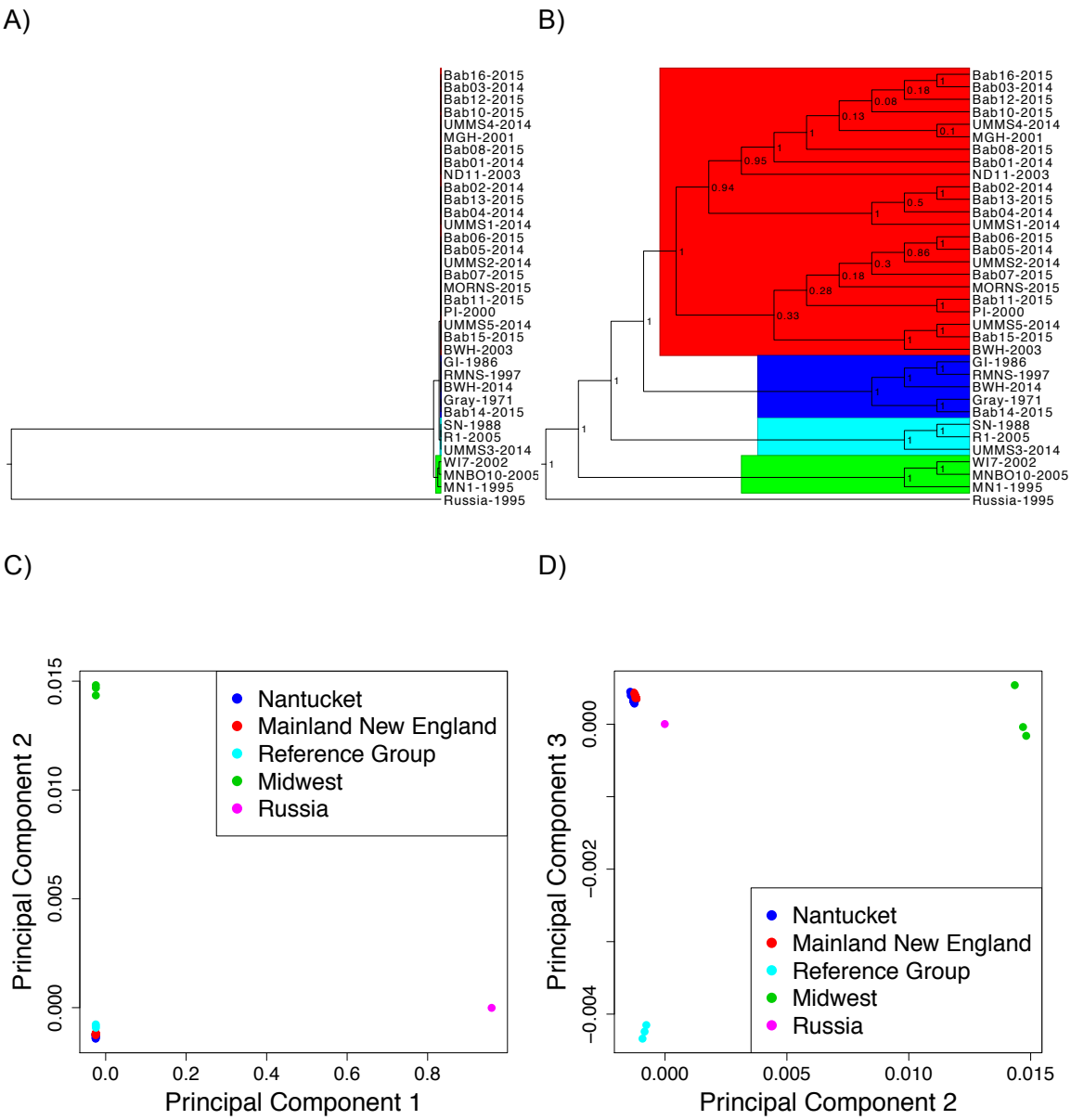


G) **H)**



A) Mean coverage per site +/- 1SD for all of the libraries in the study B) Percentage of bases in the genome with fewer than 2 reads for each library sequenced. C-F) Coverage histograms for libraries prepared with C) Apollo protocol D) Agilent SureSelect E) TruSeq protocol and F) TruSeq protocol with hybrid selection. G) Fraction of reads supporting alternate variants in BMSS and H) CUS samples.

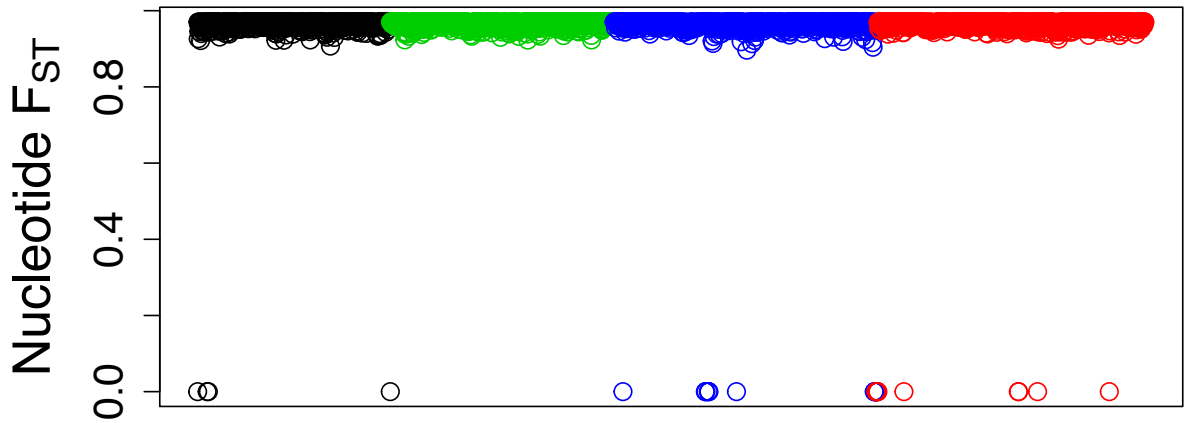
Supplementary Figure 3: Additional Phylogenetic Analysis of *B. microti* samples



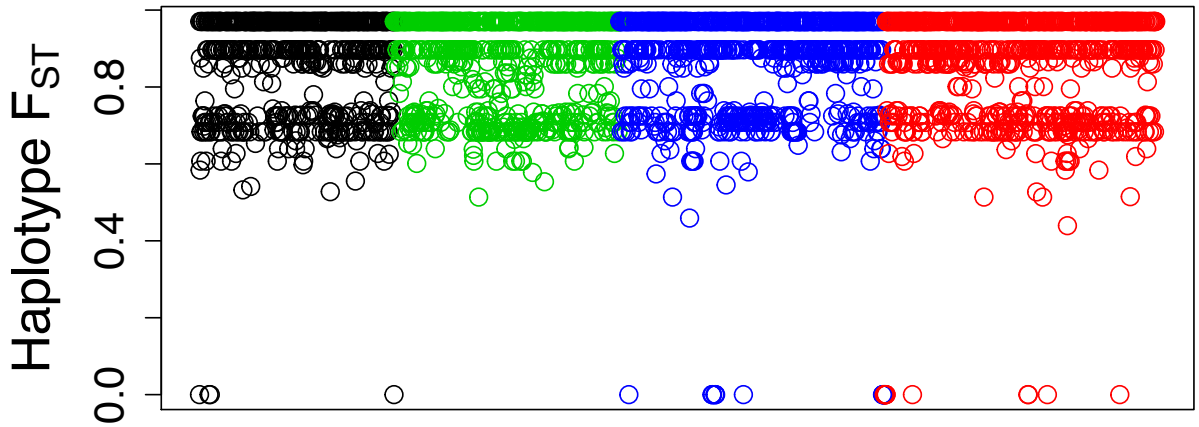
A) Maximum clade credibility tree from core chromosomal sequences (chromosome 2 removed – see supplement) with groups of samples colored by lineage; Green – Midwest, Cyan – Reference Group, Blue – Nantucket, Red – Mainland New England. B) Cladeogram of the tree in A) showing posterior support for each node. C-D) Principal component plots of genetic relationships among strains based p-distance (i.e. the proportion of nucleotides that differ between two sequences).

Supplemental Figure 4: Population Genetic Summary Statistics

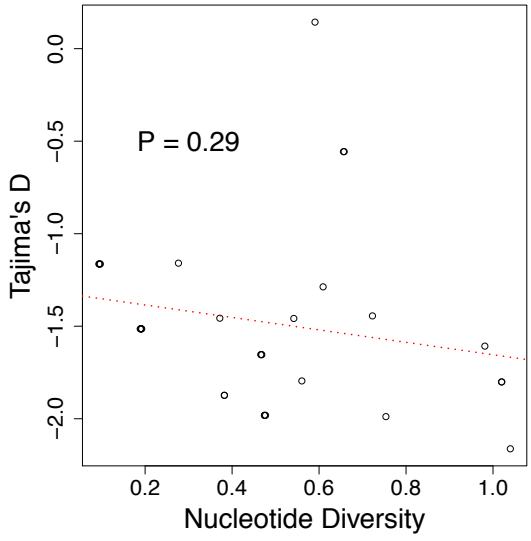
A)



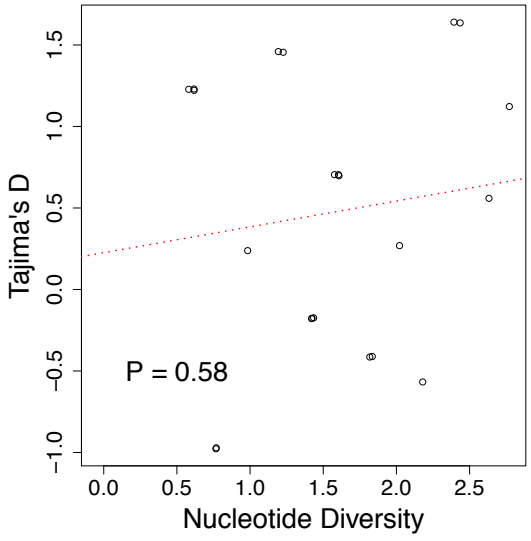
B)



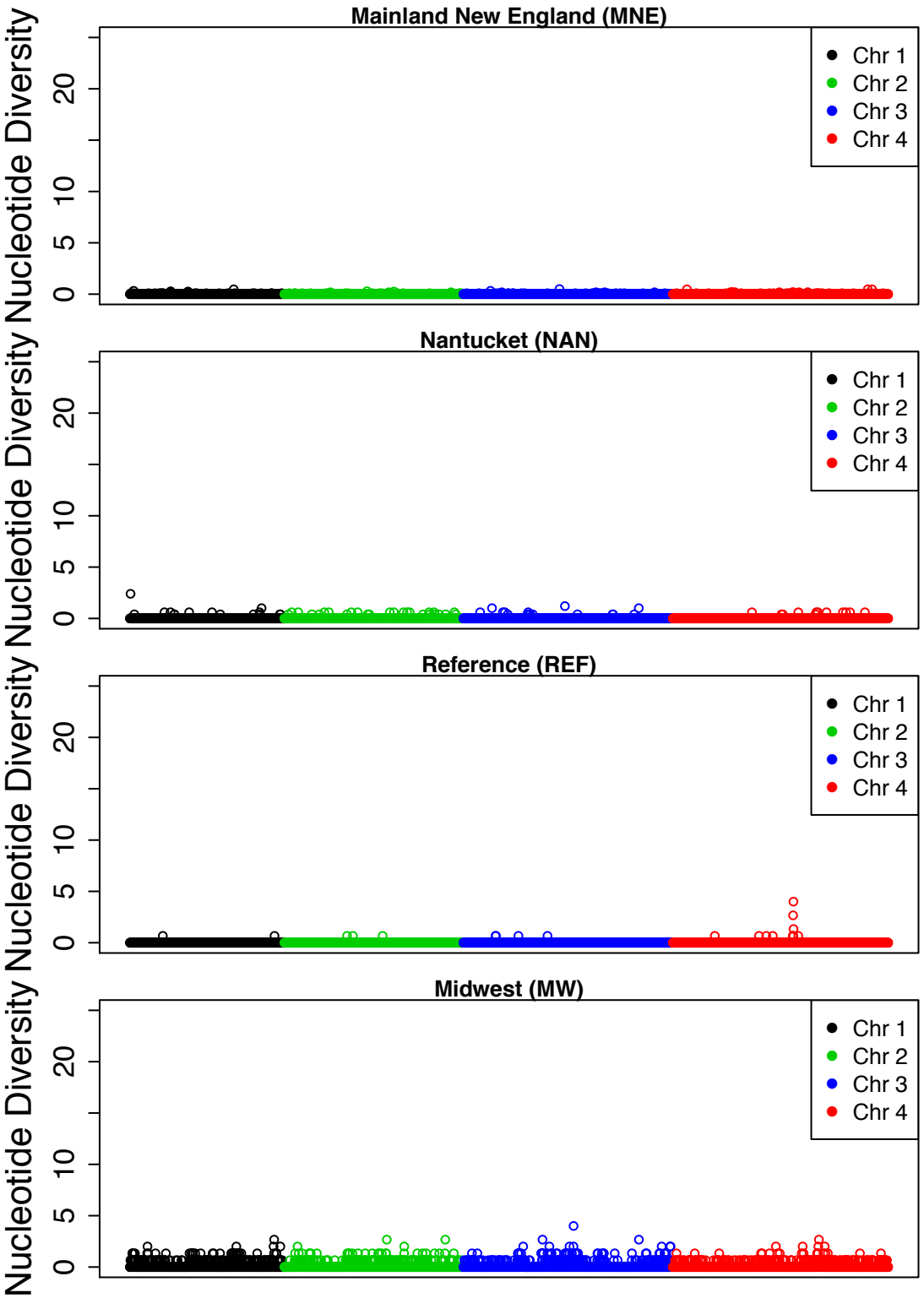
C)



D)



476 **E)**



477
478
479

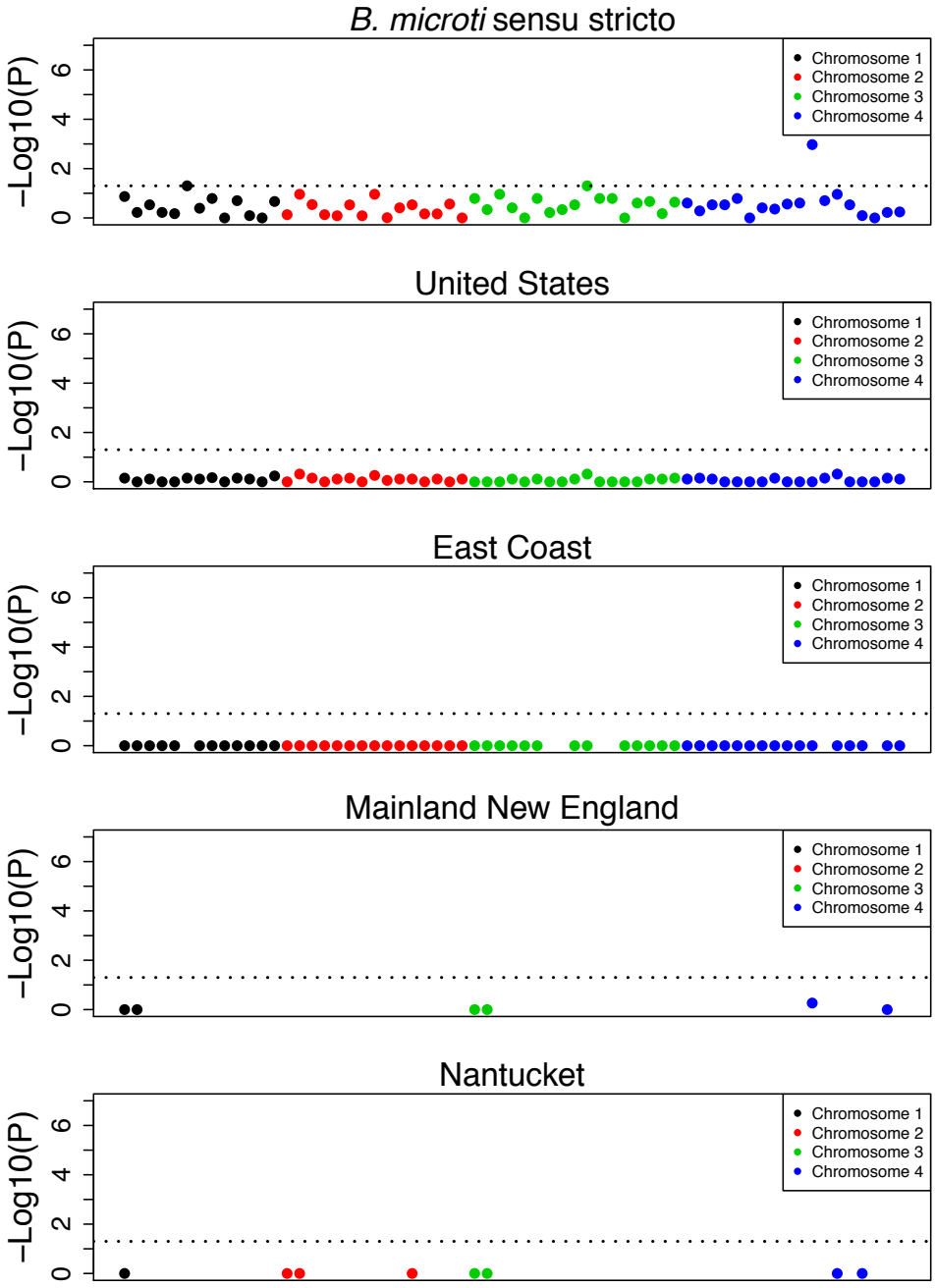
A) Genome-wide values of Fst calculated by nucleotide (upper panel) and haplotype (lower panel) methods. The x-axis shows concatenated chromosomes (chromosome 1 – black; chromosome 2 – red; chromosome 3 – green; chromosome 4 – blue). **B)** Relationship between Tajima's D and π for Mainland New England linage samples; **C)** Relationship between Tajima's D and π for Nantucket linage samples. **E)** Nucleotide diversity within lineages. The peak on chromosome 4 in REF corresponds to BBM_III07535, which had extreme polymorphism such that reads aligned to the reference only for other samples within the REF group.

Supplementary Figure 5: Analysis of Recombination in *B. microti*

A)

	B. microti sensu stricto	United States	EastCoast	Mainland New England	Nantucket
Chromosome 1	2.24E-06	4.52E-04	3.71E-01	6.27E-01	1.00E+00
Chromosome 2	2.85E-05	4.52E-06	2.50E-01	2.07E-01	1.00E+00
Chromosome 3	2.23E-12	3.81E-05	9.88E-01	1.00E+00	1.00E+00
Chromosome 4	6.22E-16	3.10E-04	2.28E-04	4.61E-01	1.00E+00

B)

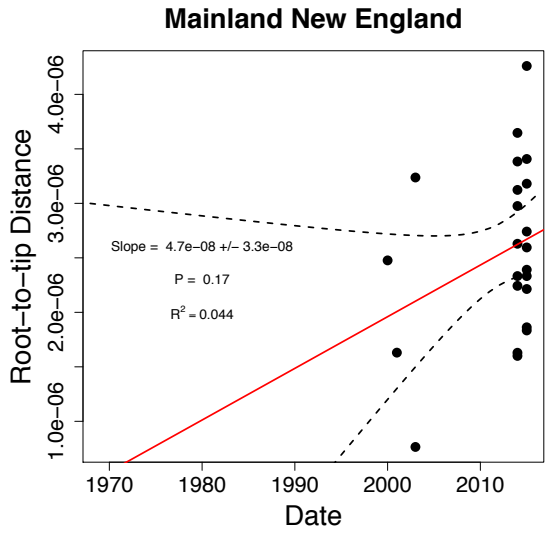


A) Results of the PHI[20] test for all chromosomes in each lineage. B) PHI test in 200Kb windows throughout the genome. No point is plotted if the interval contained an insufficient number of polymorphic sites to evaluate the test statistic. The test could not

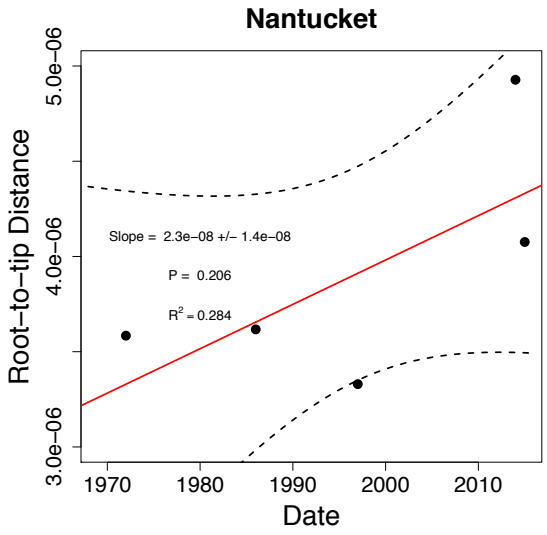
be conducted on a lineage with three samples, so a separate panel for MW is not included. These samples are incorporated into CUS.

Supplementary Figure 6: Divergence Times of *B. microti*

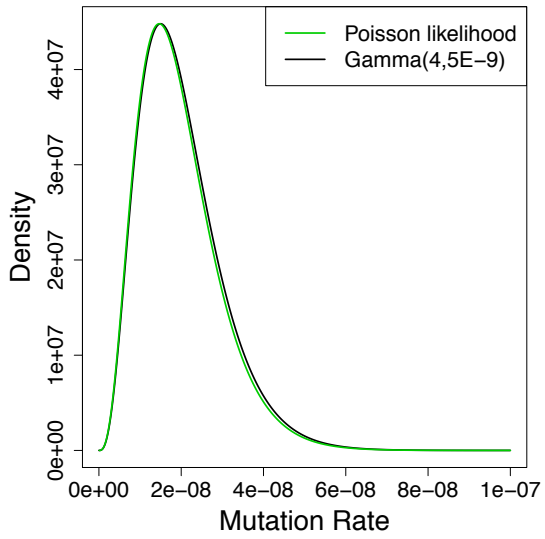
A)



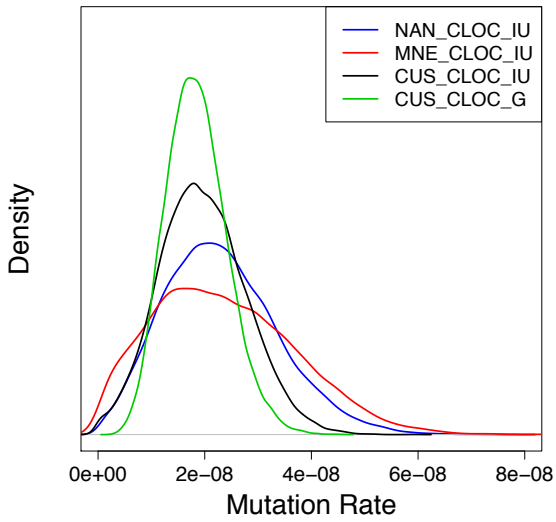
B)



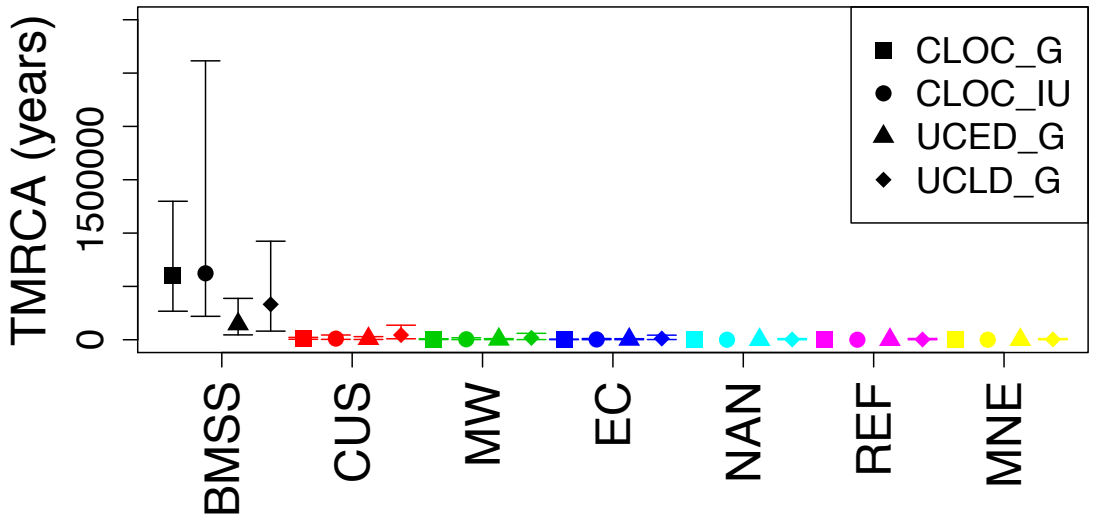
C)



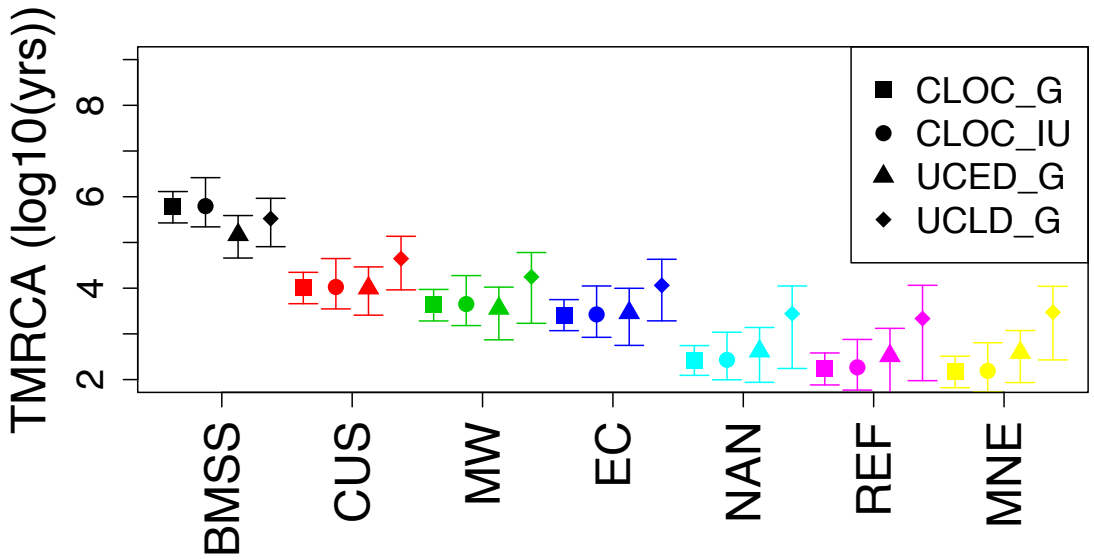
D)



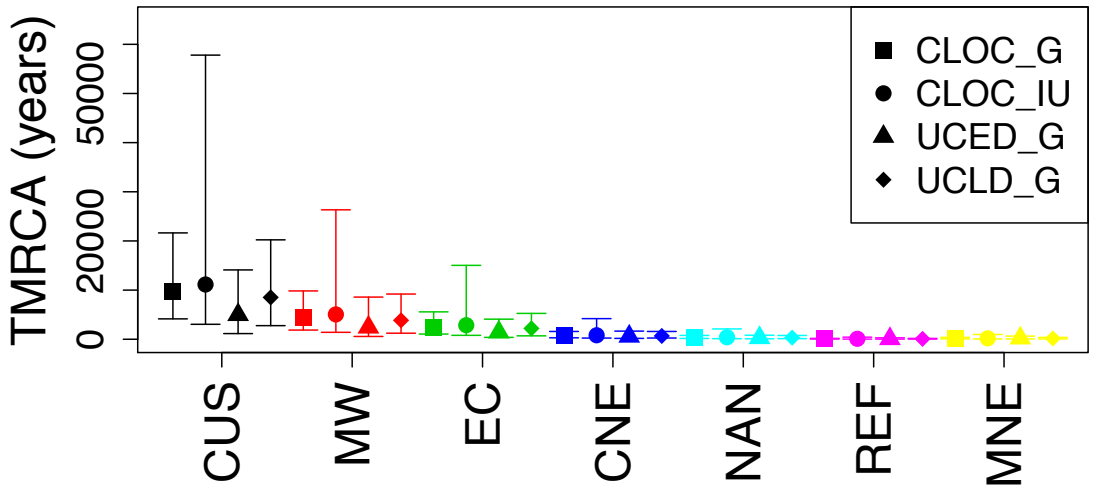
E)


TMRCA (years)
0
1500000
CLOC_G
CLOC_IU
UCED_G
UCLD_G
BMSS
CUS
MW
EC
NAN
REF
MNE

F)

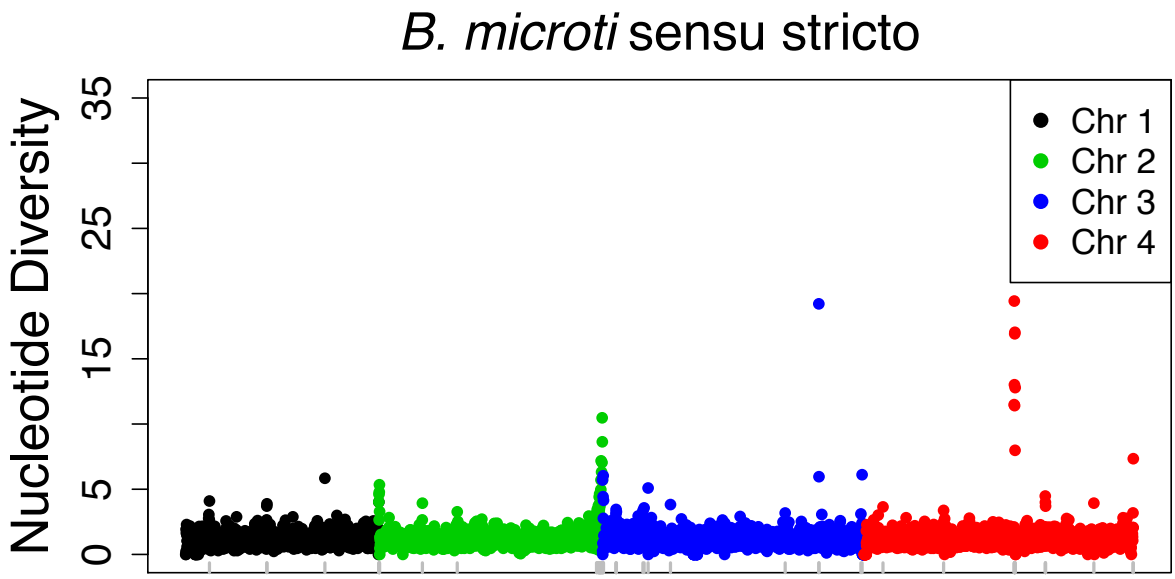

TMRCA (log10(yrs))
2
4
6
8
CLOC_G
CLOC_IU
UCED_G
UCLD_G
BMSS
CUS
MW
EC
NAN
REF
MNE

G)



A) Root-to-tip distance for MNE and **B)** NAN samples. **C)** Poisson likelihood for empirical rate from laboratory propagated isolates and Gamma(4,5x10^{-9}) which was used to construct a prior (see supplemental note). **D)** Posterior distributions for mutation rate for NAN and MNE lineages run independently with an uninformative (improper uniform) prior, for CUS samples with an uninformative (improper uniform) prior, and for CUS samples with a Gamma(4,5x10^{-9}) prior. TMRCA estimates for continental US lineages are given in Figure 3c and Supplemental Table 6. **E-F)** TMRCA (median plotted as a point, with shape denoting the model, and error bars corresponding to 95% HPD) for all BMSS lineages are shown in E) (linear scale) and F) (log scale). **G)** TMRCA estimates obtained under a codon-partitioned model based on an alignment of all protein coding genes on nuclear chromosomes (median plotted as a point, with shape denoting the model, and error bars corresponding to 95% HPD).

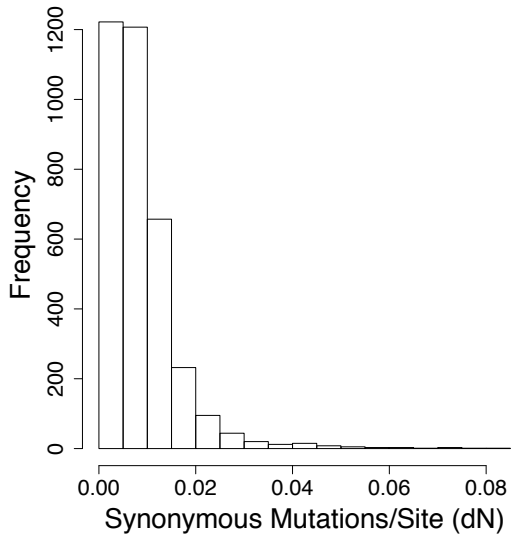
Supplementary Figure 7: Genomic Distribution of Variants (Unfiltered Set)



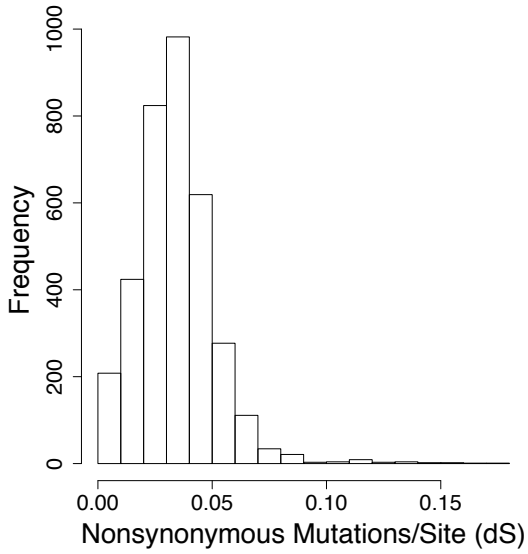
Pairwise nucleotide diversity (π, in 1Kb bins) using the unfiltered set of variant calls (methods). Discrepancies with Figure 2a represent places where there is likely sequence variation (e.g. subtelomeric regions), but we cannot confidently call the variant. Regions marked with gray ticks were in the top 1% of diversity by bin. A list of these regions is provided in Supplemental Table 3.

Supplementary Figure 8: Analysis of Substitution Rate, Evolution and BMN genes.

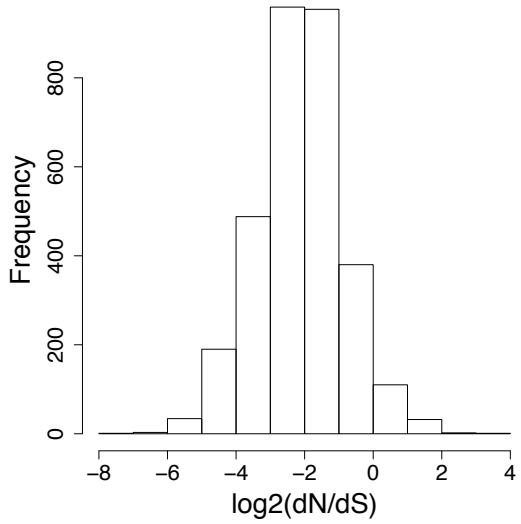
A)



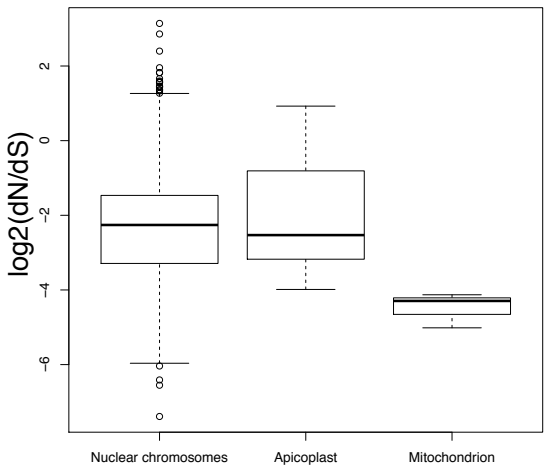
B)



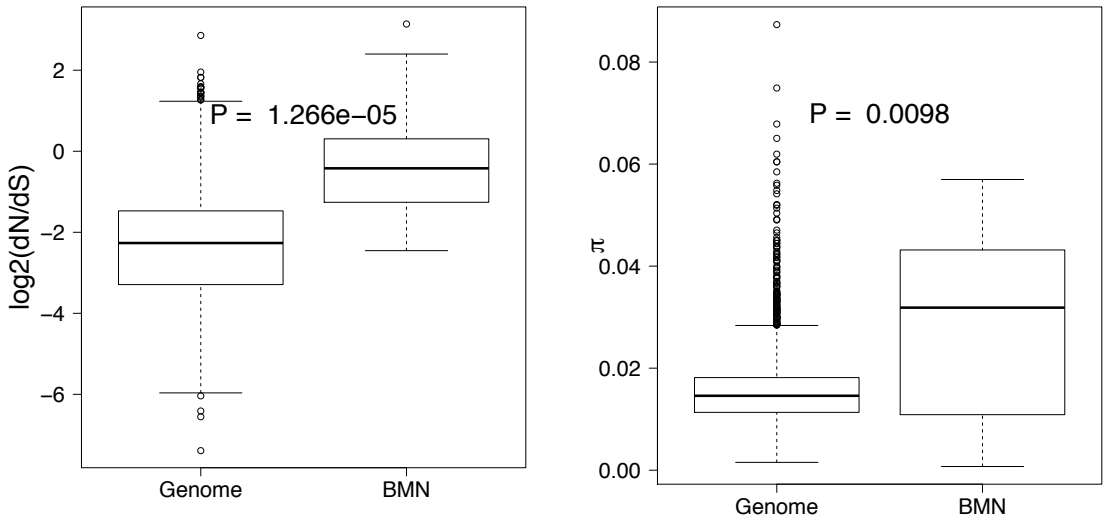
C)



D)



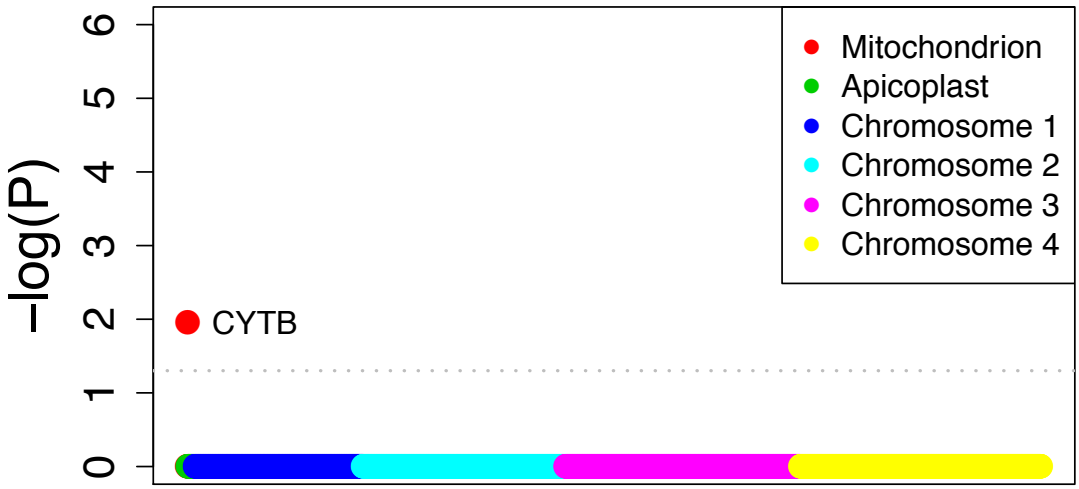
E) F)



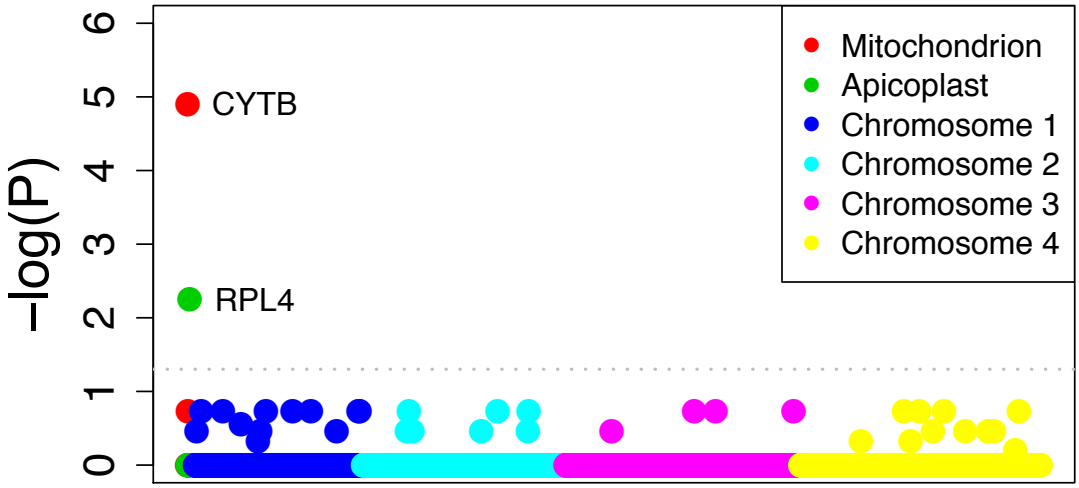
A) Distribution of dS, the rate of synonymous mutations per synonymous site, B) dN, the rate of non-synonymous mutations per non-synonymous site and C) the ratio of dN/dS. D) Box-and-whisker plot showing median and 1st and 3rd quartiles of dN/dS ratio by sequence type. Whiskers extend to 1.5 times the interquartile range, or the most extreme data point, whichever is larger. E) Box-and-whisker plot showing median and 1st and 3rd quartiles for dN/dS ratios for BMN genes as compared to the genome as a whole ($P = 1.26 \times 10^{-5}$, Wilcoxon Rank-Sum test, two-tailed alternative). Whiskers are marked as in D. F) Box-and-whisker plot showing median and 1st and 3rd quartiles for SNP density among BMN genes compared to the genome as a whole ($P = 9.8 \times 10^{-3}$, Wilcoxon Rank-Sum test, two-tailed alternative). Whiskers are marked as in D).

Supplemental Figure 9: Variants Associating with Relapsing Babesiosis and Timeline of Parasitemia in Bab05

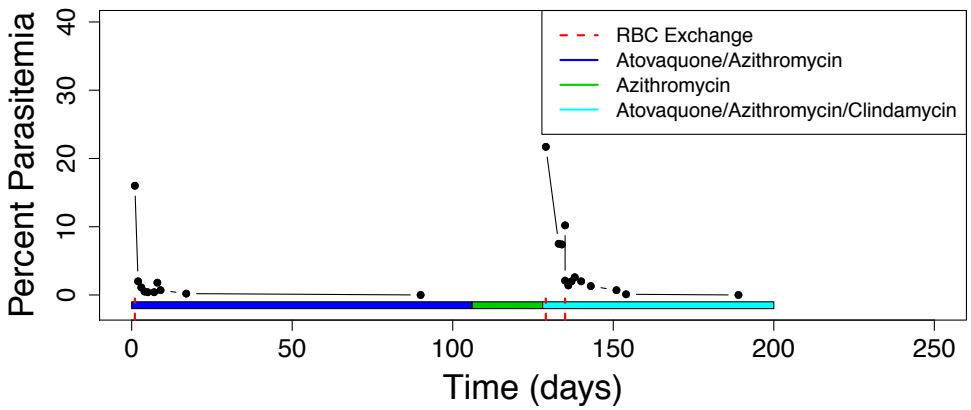
A)



B)

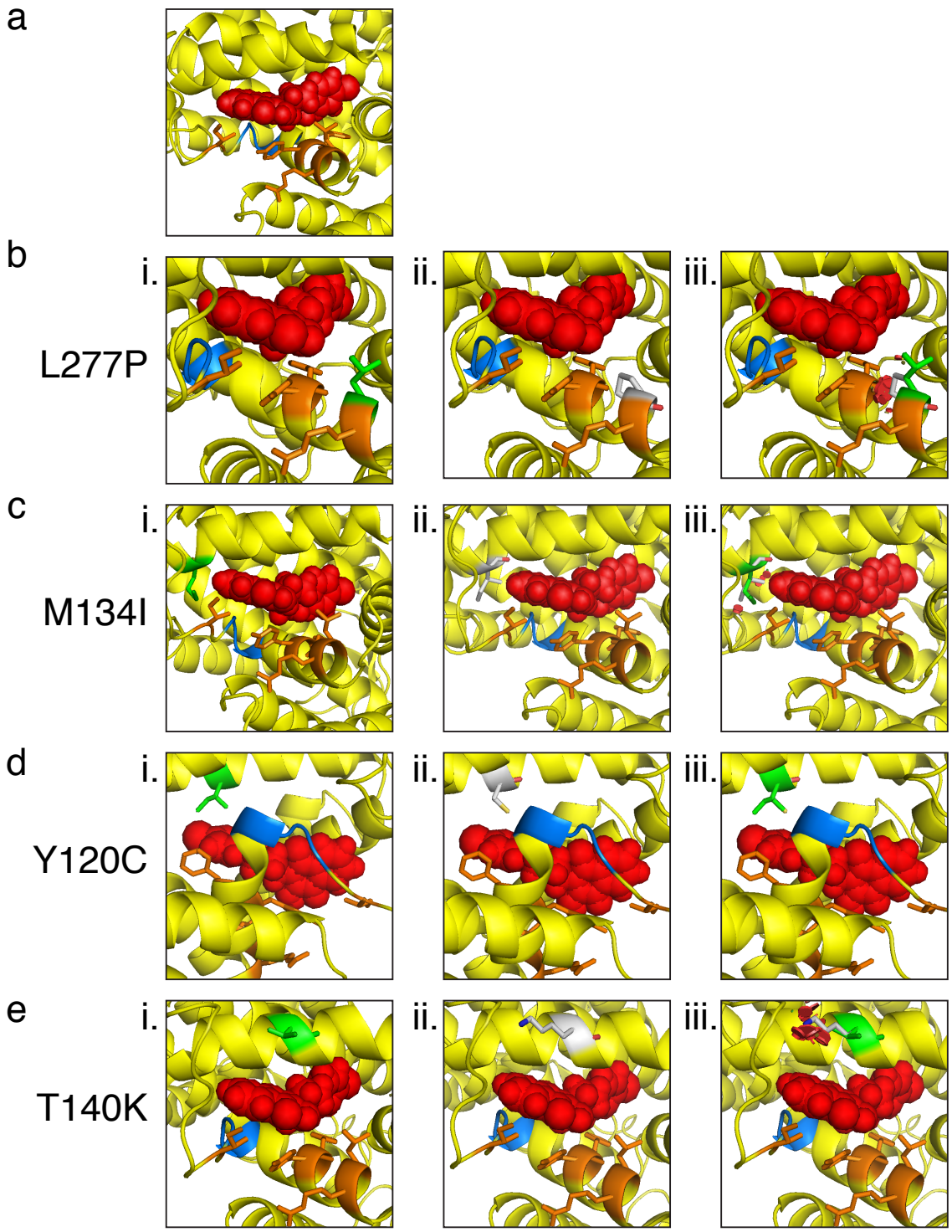


C)



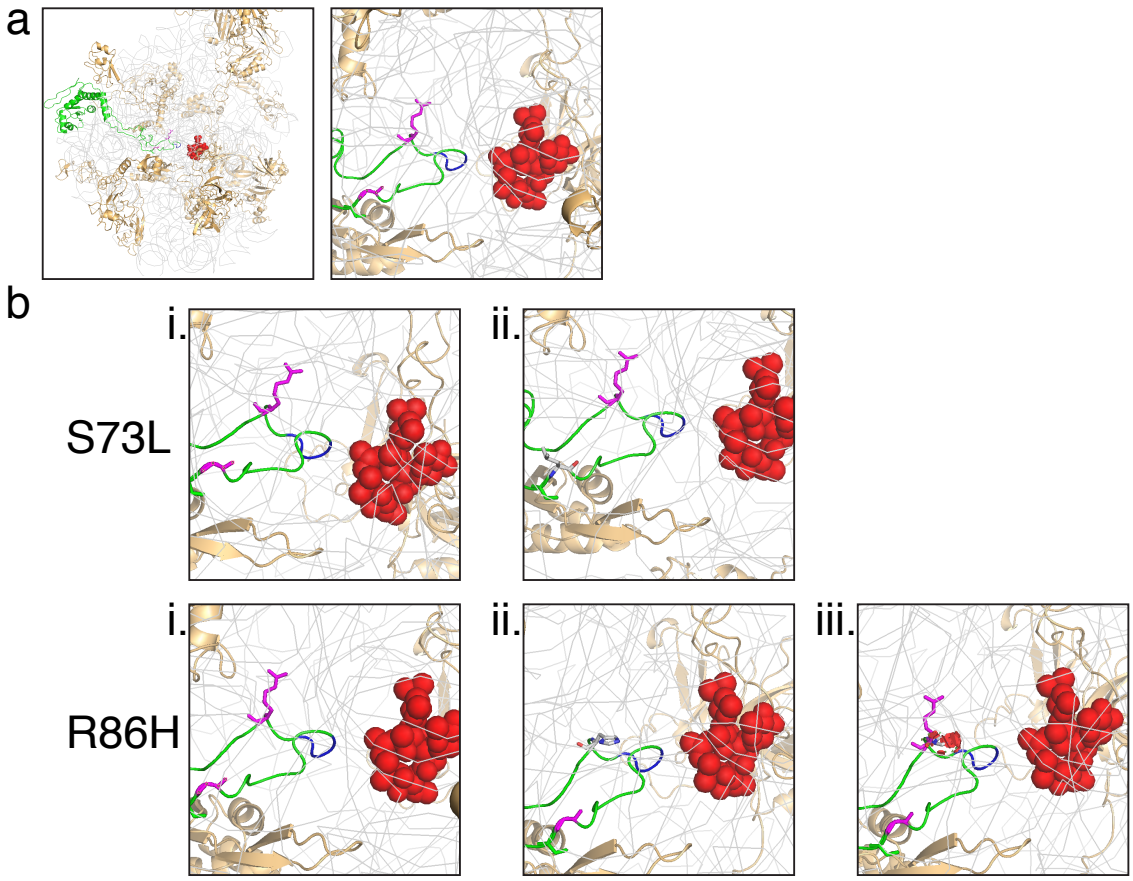
A) Fisher's exact test for association between the proportion of relapsing cases that contain a non-synonymous variant in a given protein vs. the proportion of non-relapsing cases that contain such variants. Top plot gives correct P values by the method of Benjamini and Hochberg [47]; uncorrected P values are in the bottom panel. C) Timeline of parasitemia and treatment for the Bab05 case.

Supplementary Figure 10: Modeling of cytochrome b mutations identified in relapsing *B. microti* cases.



The solved structure of *S. cerevisiae* cytochrome bc1 complex (yellow ribbon) complexed with atovaquone[48] (red) was used to model mutations found in atovaquone-resistant *B. microti.* (PDB ID: 4PD4)
A. Mutations conferring atovaquone resistance in *P. falciparum* have been described[44] in residues (orange) that are in close proximity with the atovaquone binding pocket, as well as the highly conserved PEWY motif (blue).
B. Visualization of L277 in *B. microti* (corresponding to L282 in the structural model), which has been colored green (i). An L>P substitution has been observed at this site in atovaquone-resistant strains Bab05 and BWH2003; the mutant residue is shown in gray (ii). The Mutagenesis wizard in Pymol was used to model the L277P mutation and the highest probability rotamer was selected for representation (iii). Red disks denote significant overlap of atomic van der Waals radii and thus indicate potential steric hindrance.
C. Same as in B, but modeling M134I in *B. microti* (corresponding to M139 in the structural model). The M>I substitution has been observed at this site in the atovaquone-resistant *B. microti* strain MGH2001, as well as in other Apicomplexan species.
D. Same as in B, but modeling Y120C in B. microti (corresponding to Y125 in the structural model). The Y>C substitution has been observed at this site in the atovaquone-resistant *B. microti* strain Bab14.
E. Same as in B, but modeling T140K in B. microti (corresponding to T145 in the structural model). The T>K substitution has been observed at this site in the atovaquone-resistant *B. microti* strain MORNS2015.

Supplementary Figure 11: Modeling of ribosomal protein L4 mutations identified in relapsing cases

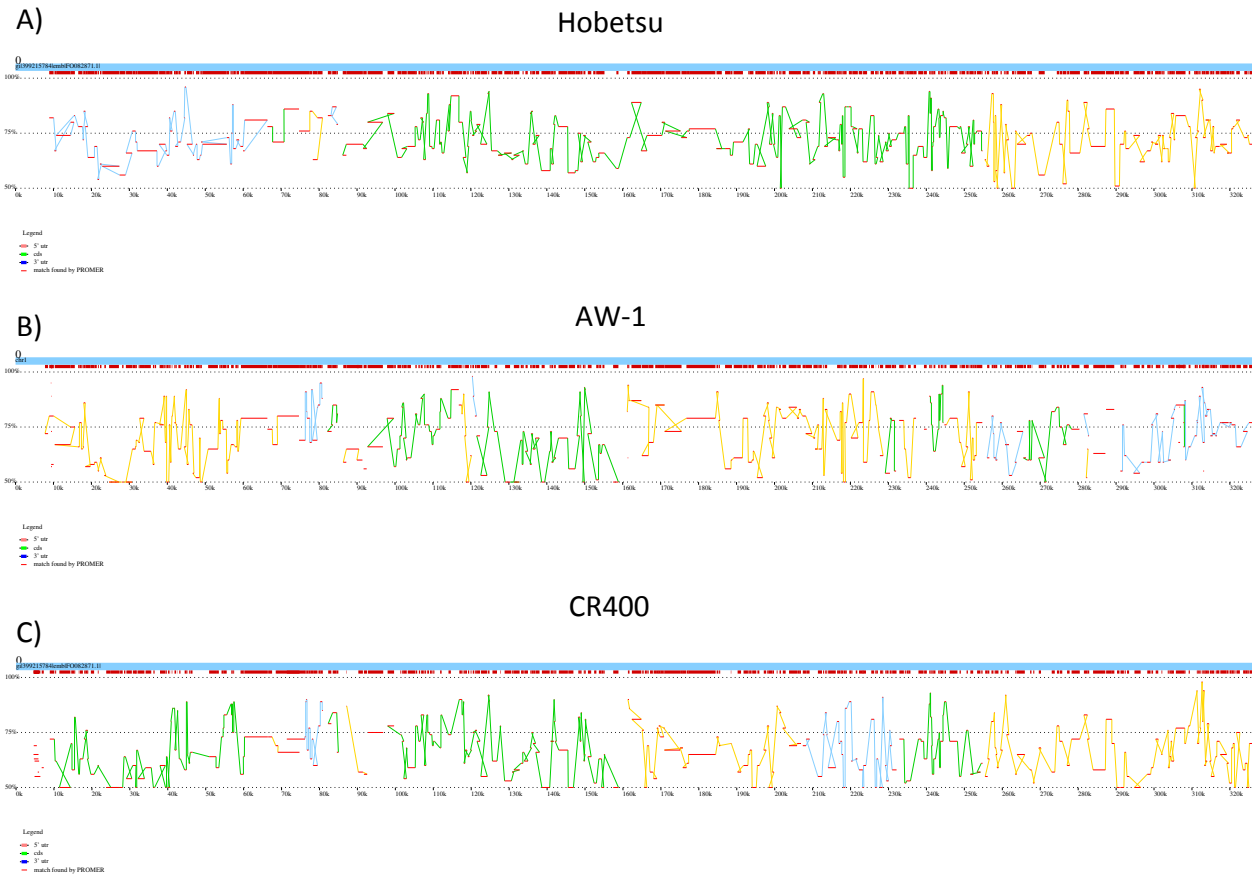


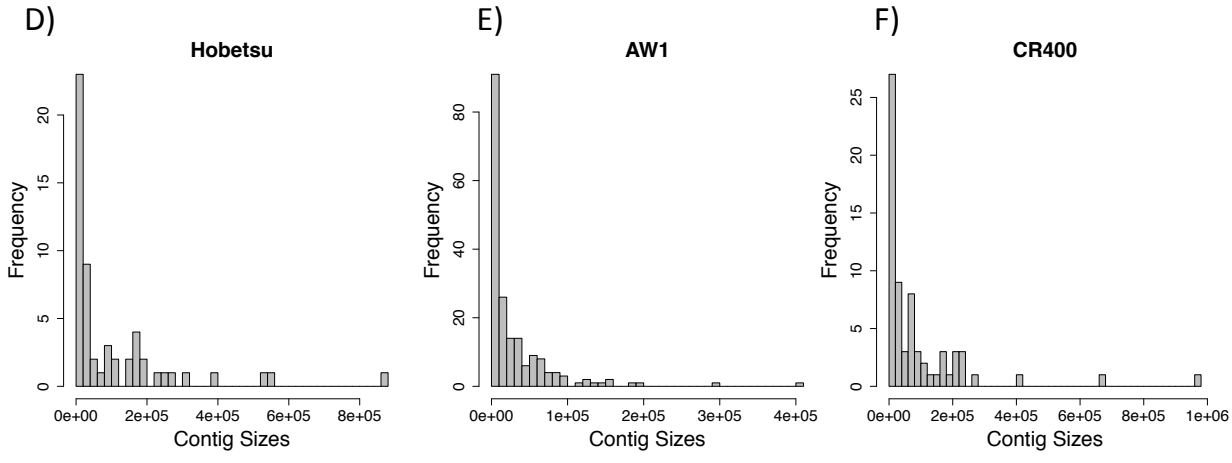
A. The published structure of *T. thermophilus* ribosomal protein L4[49] (green ribbon) was used to model *RPL4* mutations in azithromycin-resistant B. microti described in this study (PDB ID: 4V7Y). 23S, 16S, and 5S rRNA (gray) have been simplified to enable visualization of the interaction between L4 and azithromycin (red spheres). The positions of previously described mutations conferring azithromycin resistance in other Bacterial[50] and Apicomplexan[45] species (see Figure 4) have been highlighted in blue.

B. Visualization of S73 (corresponding to G61 in the structural model), which has been colored in green (i). The *B. microti* MORNS2015 sample has an S>L substitution at this site; the mutant residue is shown in gray (ii). Due to the discordance in the wild-type residue occurring at this site between species, the steric hindrance resulting from the conversion to leucine cannot be appropriately modeled.

C. Same as in B, but modeling R86H in *B. microti* (corresponding to R74 in the structural model, colored in green). The *B. microti* Bab05 sample has an R>H mutation at this site; the mutant residue is shown in gray (ii). Red disks denote significant overlap of atomic van der Waals radii and thus indicate possible steric hindrance.

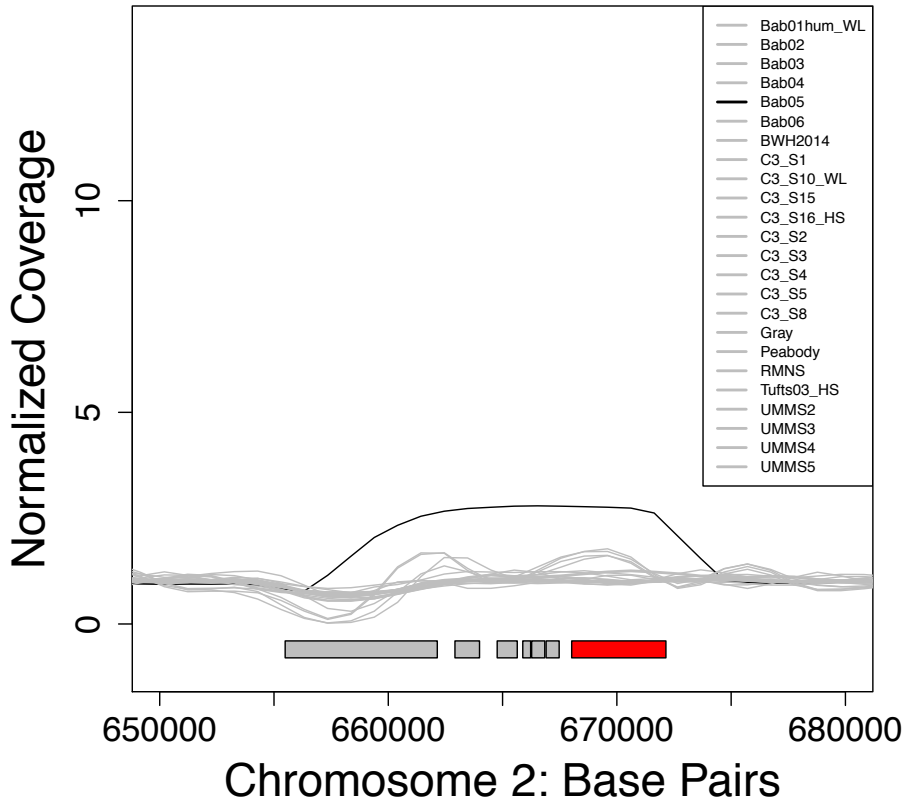
Supplementary Figure 12: Contig Alignment of *B. microti*-like strains and Draft Assemblies





A-C) Promer alignments for the first 320kb of chromosome 1 for (A) Hobetsu B) AW-1 samples, both from Japan, and C) CR-400 from Alaska. Distinct contigs are represented by different colors, with blue representing the reference sequence and red regions representing areas of alignment to the reference. D-F) Distribution of contig sizes for the each of the assemblies.

Supplemental Figure 13: Amplification of bmMRP



Amplification of a 15KB region on chromosome 2 (658,075-672,981), containing bmMRP (red), which coverage data suggested was present in three copies in Bab05.

References:

1. Western, K. A., Benson, G. D., Gleason, N. N., Healy, G. R. & Schultz, M. G. Babesiosis in a Massachusetts Resident. *N Engl J Med* **283,** 854–856 (1970).
2. Steketee, R. W. *et al.* Babesiosis in Wisconsin: A New Focus of Disease Transmission. *JAMA* **253,** 2675–2678 (1985).
3. Joseph, J. T. *et al.* Babesiosis in Lower Hudson Valley, New York, USA. *Emerg. Infect. Dis.* **17,** 843–847 (2011).
4. Goethert, H. K. & Telford, S. R. What is Babesia microti? *Parasitology* **127,** 301–309 (2003).
5. Tsuji, M. *et al.* Human babesiosis in Japan: epizootiologic survey of rodent reservoir and isolation of new type of Babesia microti-like parasite. *J. Clin. Microbiol.* **39,** 4316–4322 (2001).
6. Wei, Q. *et al.* Human babesiosis in Japan: isolation of Babesia microti-like parasites from an asymptomatic transfusion donor and from a rodent from an area where babesiosis is endemic. *J. Clin. Microbiol.* **39,** 2178–2183 (2001).
7. Nakajima, R. *et al.* Babesia microti-group parasites compared phylogenetically by complete sequencing of the CCTeta gene in 36 isolates. *J Vet Med Sci* **71,** 55–68 (2009).
8. Goethert, H. K., Cook, J. A., Lance, E. W. & Telford, S. R., III. Fay and Rausch 1969 Revisited: Babesia microti in Alaska Small Mammals. *journal of parasitology* **92,** 826–831 (2009).
9. Telford, S. R. *et al.* [Detection of natural foci of babesiosis and granulocytic ehrlichiosis in Russia]. *Zh Mikrobiol Epidemiol Immunobiol* 21–25 (2002).
10. Sriprawat, K. *et al.* Effective and cheap removal of leukocytes and platelets from Plasmodium vivax infected blood. *Malaria Journal* **8,** 115 (2009).
11. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* **27,** 182–189 (2009).
12. Cornillot, E. *et al.* Sequencing of the smallest Apicomplexan genome from the human pathogen Babesia microti. *Nucleic Acids Research* **40,** 9102–9114 (2012).
13. Not ‘out of Nantucket’: Babesia microti in southern New England comprises at least two major populations. **7,** 546 (2014).
14. Ruebush, T. K. & Spielman, A. Human Babesiosis in the United States. *Ann Intern Med* **88,** 263–263 (1978).
15. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed Phylogenetics and Dating with Confidence. *Plos Biol* **4,** e88 (2006).
16. Suchard, M. A., Weiss, R. E. & Sinsheimer, J. S. Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular Biology and Evolution* **18,** 1001–1013 (2001).
17. Baele, G. *et al.* Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* **29,** 2157–2167 (2012).
18. Telford, S. R., III. Babesial infections in humans and wildlife. *Parasitic protozoa* **5,** 1–47 (1993).
19. Rudzinska, M. A., Spielman, A., Lewengrub, S., Trager, W. & Piesman, J. Sexuality in piroplasms as revealed by electron microscopy in Babesia microti. *Proceedings of the National Academy of Sciences* **80,** 2966–2970 (1983).
20. Bruen, T. C., Philippe, H. & Bryant, D. A Simple and Robust Statistical Test for Detecting the Presence of Recombination. *Genetics* **172,** 2665–2681 (2006).

21. Homer, M. J. *et al.* A Polymorphic Multigene Family Encoding an Immunodominant Protein from Babesia microti. *J. Clin. Microbiol.* **38,** 362–368 (2000).
22. Deitsch, K. W., Moxon, E. R. & Wellems, T. E. Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. *Microbiol. Mol. Biol. Rev.* **61,** 281–293 (1997).
23. Barry, J. D., Ginger, M. L., Burton, P. & McCulloch, R. Why are parasite contingency genes often associated with telomeres? *International Journal for Parasitology* **33,** 29–45 (2003).
24. Jackson, A. P. Genome evolution in trypanosomatid parasites. *Parasitology* **142 Suppl 1,** S40–56 (2015).
25. Zarowiecki, M. & Berriman, M. What helminth genomes have taught us about parasite evolution. *Parasitology* **142 Suppl 1,** S85–97 (2015).
26. Reid, A. J. Large, rapidly evolving gene families are at the forefront of host–parasite interactions in Apicomplexa. *Parasitology* **142,** S57–S70 (2014).
27. Reid, A. J. Large, rapidly evolving gene families are at the forefront of host–parasite interactions in Apicomplexa. *Parasitology* **142,** S57–S70 (2014).
28. Kyes, S., Horrocks, P. & Newbold, C. Antigenic Variation at the Infected Red Cell Surface in Malaria. *Annu. Rev. Microbiol.* **55,** 673–707 (2001).
29. Allred, D. R. & Al-Khedery, B. Antigenic variation and cytoadhesion in Babesia bovis and Plasmodium falciparum: different logics achieve the same goal. *Mol. Biochem. Parasitol.* **134,** 27–35 (2004).
30. Allred, D. R., Cinque, R. M., Lane, T. J. & Ahrens, K. P. Antigenic variation of parasite-derived antigens on the surface of Babesia bovis-infected erythrocytes. *Infect. Immun.* **62,** 91–98 (1994).
31. Al-Khedery, B. & Allred, D. R. Antigenic variation in Babesia bovis occurs through segmental gene conversion of the ves multigene family, within a bidirectional locus of active transcription. *Mol. Microbiol.* **59,** 402–414 (2006).
32. Frank, M. *et al.* Frequent recombination events generate diversity within the multi-copy variant antigen gene families of Plasmodium falciparum. *International Journal for Parasitology* **38,** 1099–1109 (2008).
33. Claessens, A. *et al.* Generation of Antigenic Diversity in Plasmodium falciparum by Structured Rearrangement of Var Genes During Mitosis. *PLoS Genet* **10,** e1004812 (2014).
34. Borst, P. & Cross, G. A. M. Molecular basis for trypanosome antigenic variation. *Cell* **29,** 291–303 (1982).
35. Vyas, J. M., Telford, S. R. & Robbins, G. K. Treatment of refractory Babesia microti infection with atovaquone-proguanil in an HIV-infected patient: case report. *Clin Infect Dis.* **45,** 1588–1590 (2007).
36. Ruebush, M. J. & Hanson, W. L. Susceptibility of Five Strains of Mice to Babesia microti of Human Origin. *The Journal of Parasitology* **65,** 430 (1979).
37. Price, R. N. *et al.* Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number. *The Lancet* **364,** 438–447 (2004).
38. Venkatesan, M. *et al.* Polymorphisms in Plasmodium falciparum chloroquine resistance transporter and multidrug resistance 1 genes: parasite risk factors that affect treatment outcomes for P. falciparum malaria after artemether-lumefantrine and artesunate-amodiaquine. *Am. J. Trop. Med. Hyg.* **91,** 833–843 (2014).
39. Raj, D. K. *et al.* Disruption of a Plasmodium falciparum multidrug resistance-associated protein (PfMRP) alters its fitness and transport of antimalarial drugs and glutathione. *Journal of Biological Chemistry* **284,** 7687–7696 (2009).
40. Piesman, J., Karakashian, S. J., Lewengrub, S., Rudzinska, M. A. & Spielmank,

A. Development of Babesia microti sporozoites in adult Ixodes dammini. *International Journal for Parasitology* **16,** 381–385 (1986).
41. Sakuma, M., Setoguchi, A. & Endo, Y. Possible Emergence of Drug-Resistant Variants of Babesia gibsoni in Clinical Cases Treated with Atovaquone and Azithromycin. *Journal of Veterinary Internal Medicine* **23,** 493–498 (2009).
42. Matsuu, A., Miyamoto, K., Ikadai, H., Okano, S. & Higuchi, S. Short report: cloning of the Babesia gibsoni cytochrome B gene and isolation of three single nucleotide polymorphisms from parasites present after atovaquone treatment. *American Journal of Tropical Medicine and Hygiene* **74,** 593–597 (2006).
43. McFadden, D. C., Tomavo, S., Berry, E. A. & Boothroyd, J. C. Characterization of cytochrome b from Toxoplasma gondii and Qo domain mutations as a mechanism of atovaquone-resistance. *Mol. Biochem. Parasitol.* **108,** 1–12 (2000).
44. Korsinczky, M. *et al.* Mutations in Plasmodium falciparum cytochrome b that are associated with atovaquone resistance are located at a putative drug-binding site. *Antimicrob. Agents Chemother.* **44,** 2100–2108 (2000).
45. Sidhu, A. B. S. *et al.* In vitro efficacy, resistance selection, and structural modeling studies implicate the malarial parasite apicoplast as the target of azithromycin. *Journal of Biological Chemistry* **282,** 2494–2504 (2007).
46. Chittum, H. S. & Champney, W. S. Erythromycin inhibits the assembly of the large ribosomal subunit in growing Escherichia coli cells. *Current Microbiology* **30,** 273–279 (1995).
47. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57,** 289–300 (1995).
48. Birth, D., Kao, W.-C. & Hunte, C. Structural analysis of atovaquone-inhibited cytochrome bc1 complex reveals the molecular basis of antimalarial drug action. *Nature Communications* **5,** (2014).
49. Bulkley, D., Innis, C. A., Blaha, G. & Steitz, T. A. Revisiting the structures of several antibiotics bound to the bacterial ribosome. *Proceedings of the National Academy of Sciences* **107,** 17158–17163 (2010).
50. Pihlajamäki, M. *et al.* Ribosomal mutations in Streptococcus pneumoniae clinical isolates. *Antimicrob. Agents Chemother.* **46,** 654–658 (2002).