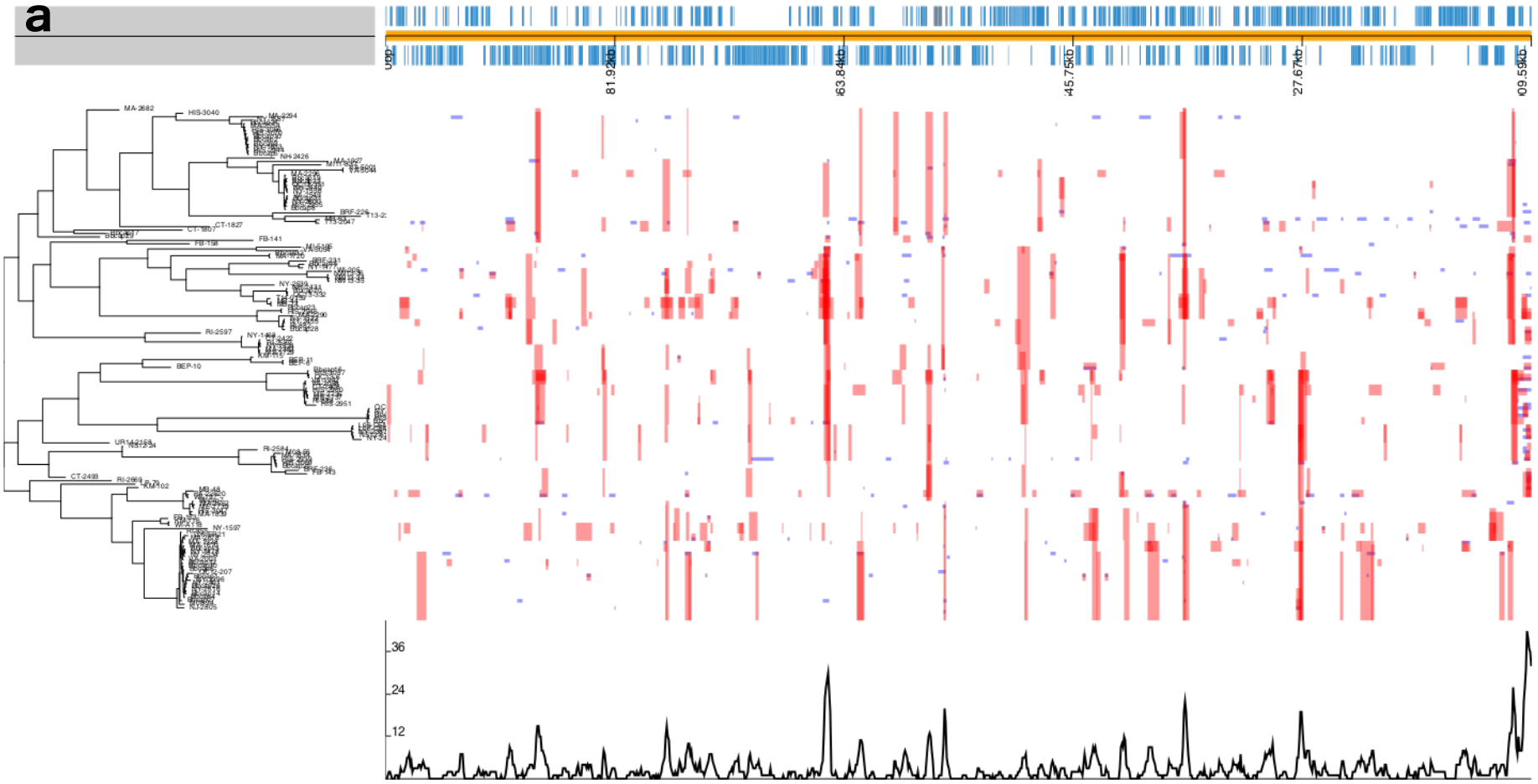
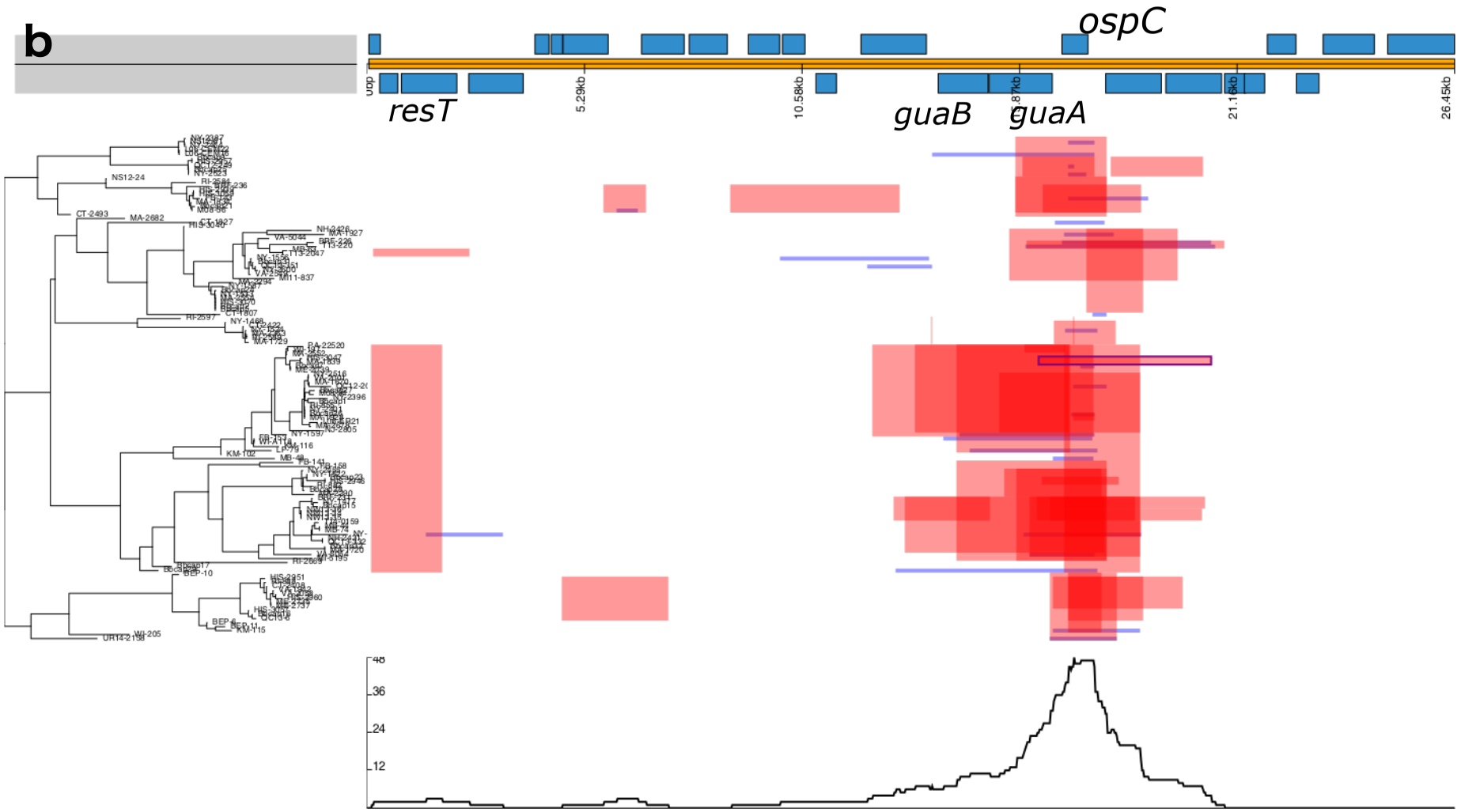
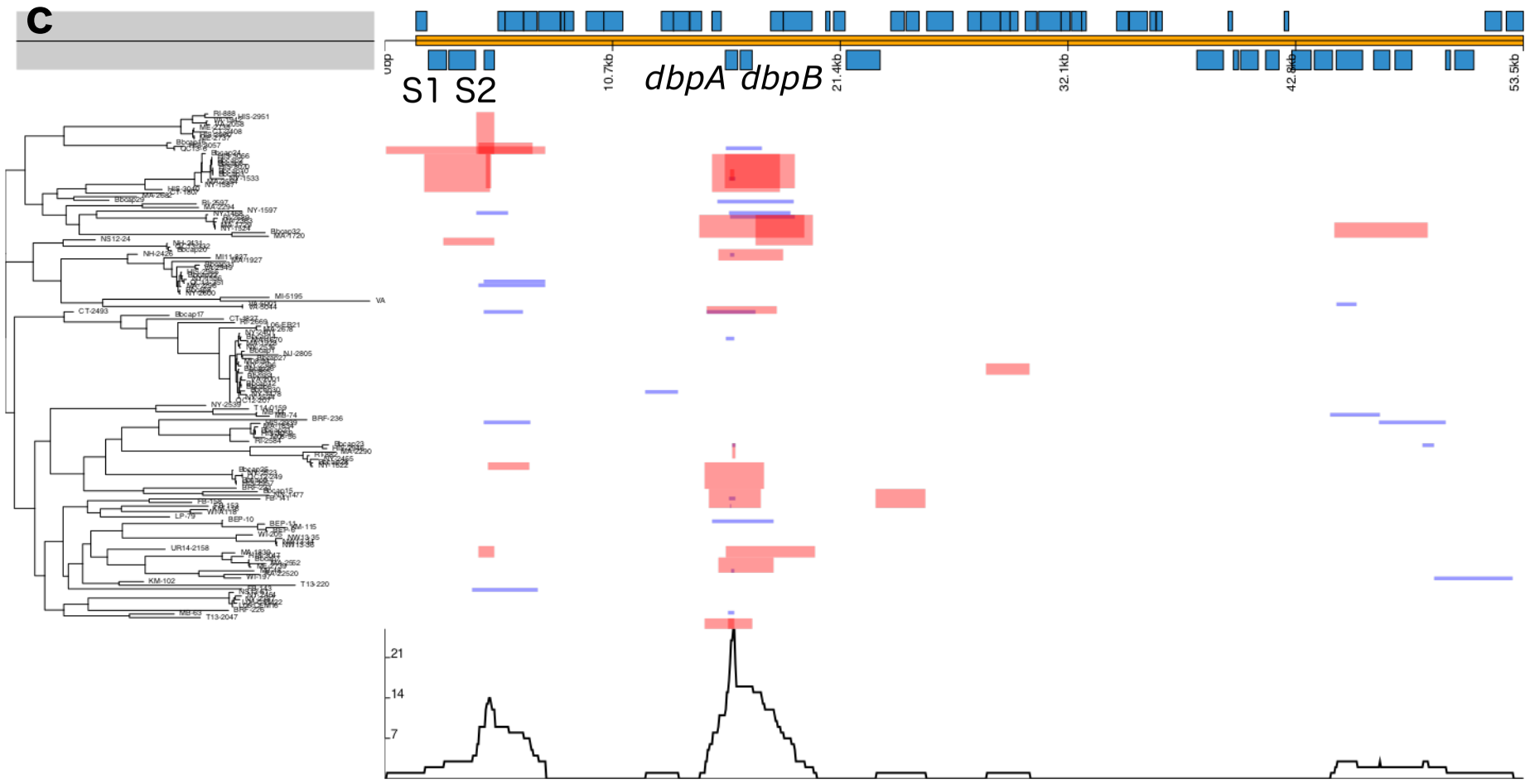
Genomic insights into the ancient spread of Lyme disease across North America

Katharine S. Walter, Giovanna Carpi, Adalgisa Caccone*, Maria A. Diuk-Wasser*

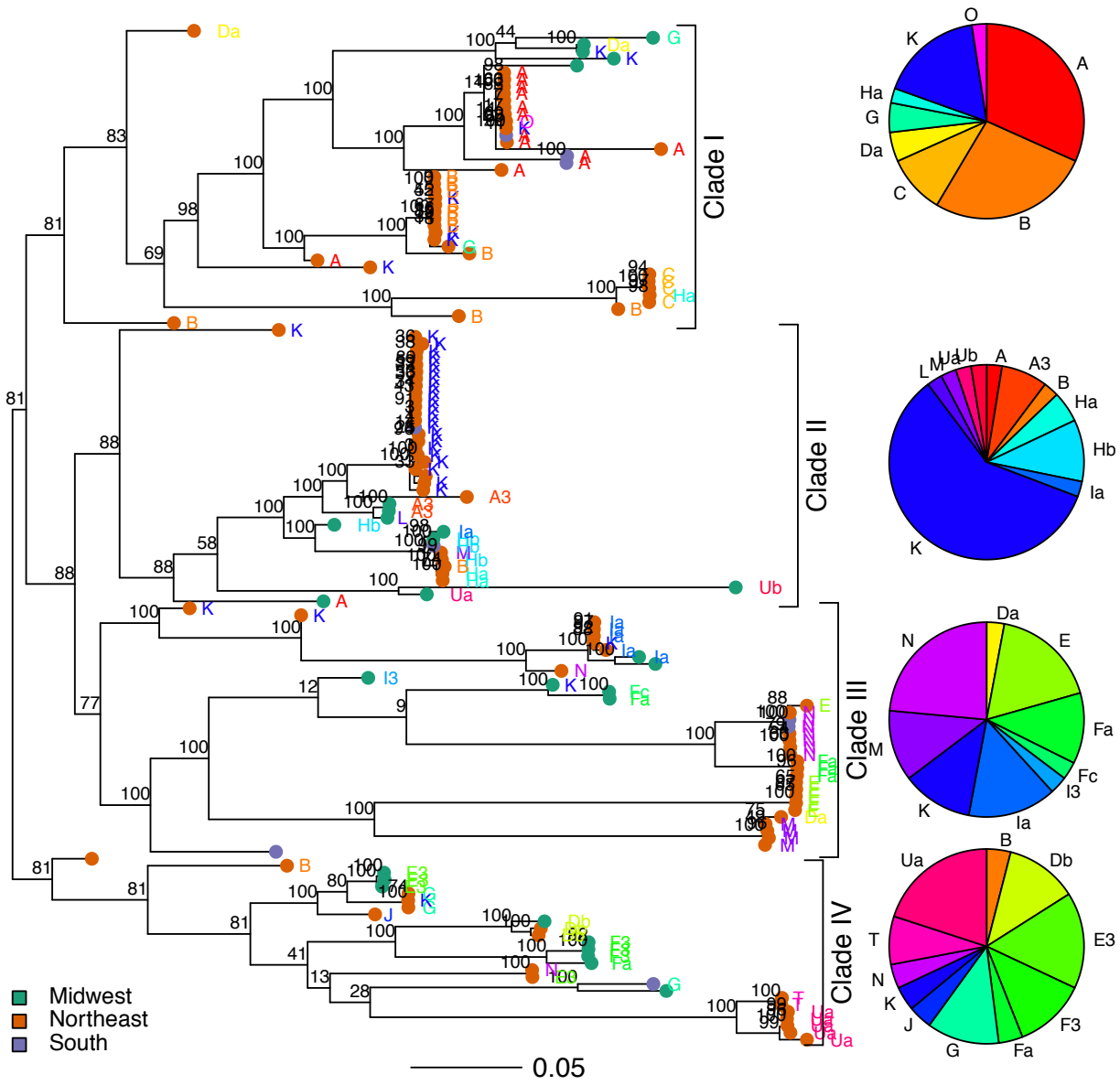
*These authors contributed equally.


a
MA-2682
HIS-3040
NH-2426
CT-1827
FB-141
FB-198
RI-2597
BEP-10
CT-2493
NY-1597
81.92kb
63.84kb
45.75kb
27.67kb
09.59kb
36
24
12

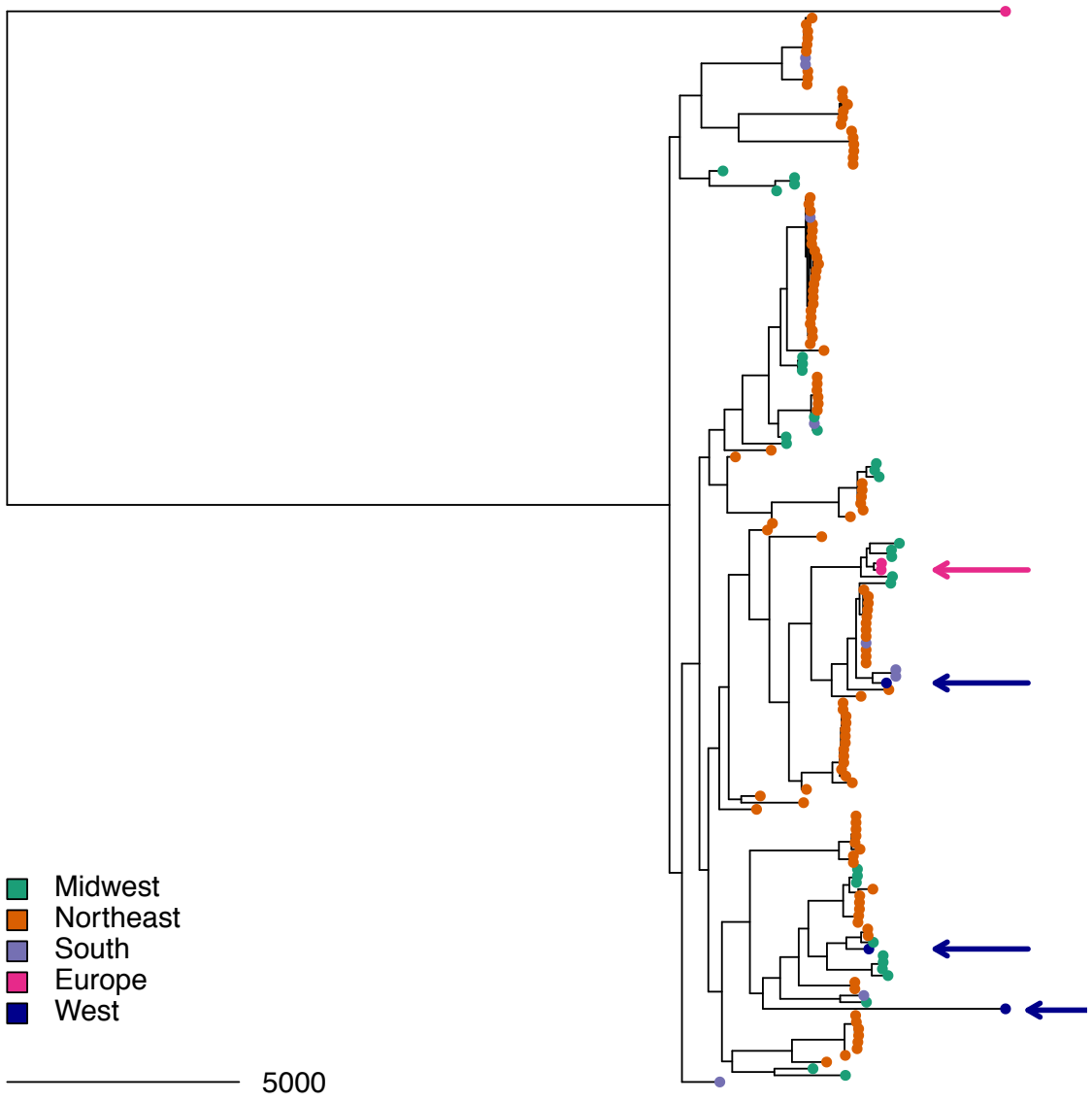

b
ospC
resT
guaB
guaA
5.29kb
10.58kb
21.1kb
26.45kb
48
36
24
12


C
S1 S2
dbpA dbpB
10.7kb
21.4kb
32.1kb
42.8kb
53.5kb
21
14
7

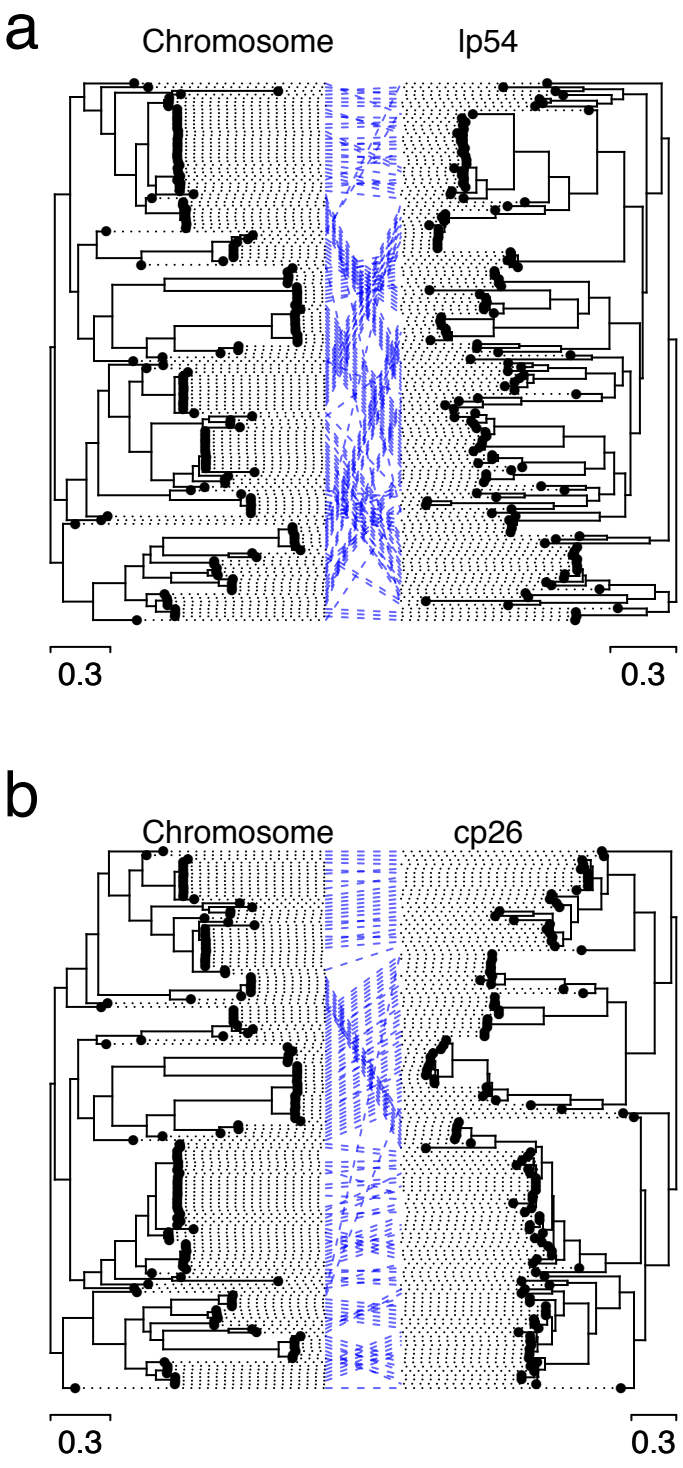
Supplementary Figure 1. Frequent recombination along the *B. burgdorferi* genome. Maximum likelihood phylogenies are pictured to the left of predicted recombinant tracts along the (a) chromosome (910724-bp), (b) plasmid cp26 (26498-bp), and (c) plasmid lp54 (53657-bp). Red blocks indicate recombination events occurring on an internal branch, shared by multiple samples through common descent. Blue blocks represent recombinations that occur on terminal branches and are unique to individual tip samples. The track at the top of each panel is a gene map; genes of interest are labeled.

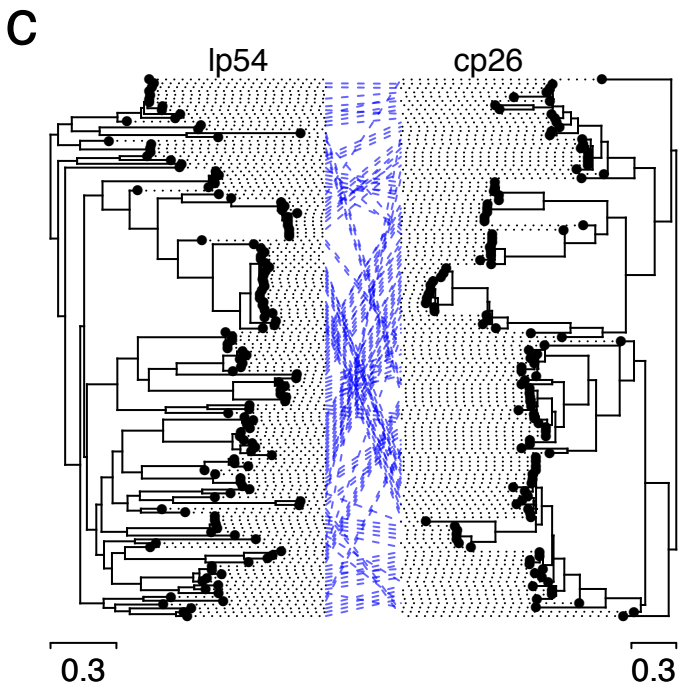


Supplementary Figure 2. Maximum likelihood *B. burgdorferi* tree inferred from a recombination-free alignment of 16,370 SNPs. As depicted in Figure 2, with each tip labeled by *ospC* serotype (*ospC* label is also colored by serotype). Tips are colored by sampling region. Branches with greater than 50% bootstrap support are labeled with support value. Branch lengths indicate the estimated number of substitutions per variable site. Pie charts depict the distribution of *ospC* serotypes within each clade.

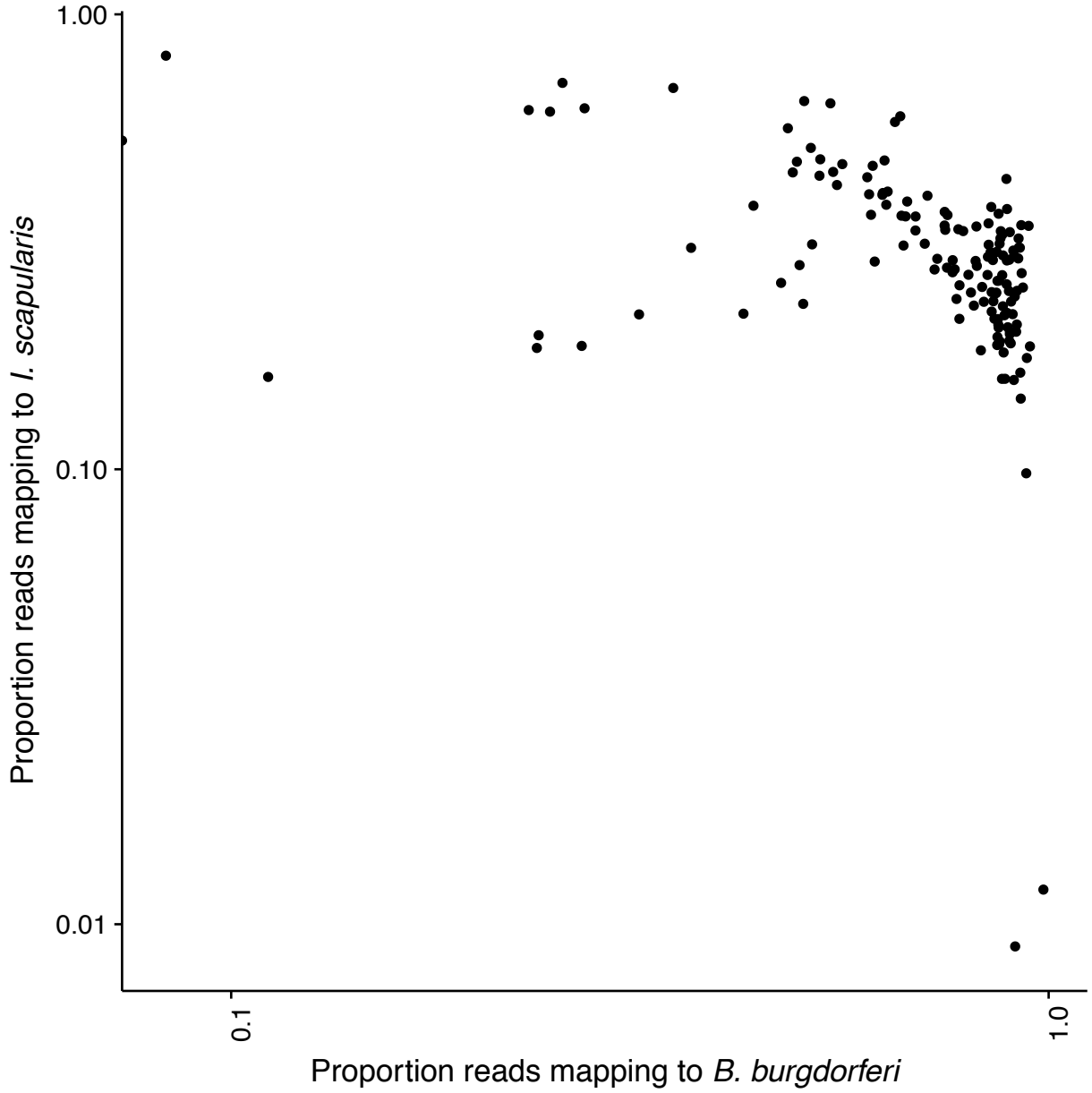


Supplementary Figure 3. *B. burgdorferi* maximum likelihood phylogeny including all 15 previously published genomes. Tips are colored by sampling region. *Borrelia finlandensis*, the most closely related sister species to *B. burgdorferi* sensu stricto, is the outgroup (Table S1). A pink arrow identifies the two monophyletic European samples. Blue arrows identify the three California samples. Branches with greater than 50% bootstrap support are labeled with support value. Branch lengths indicate the estimated number of substitutions per variable site.

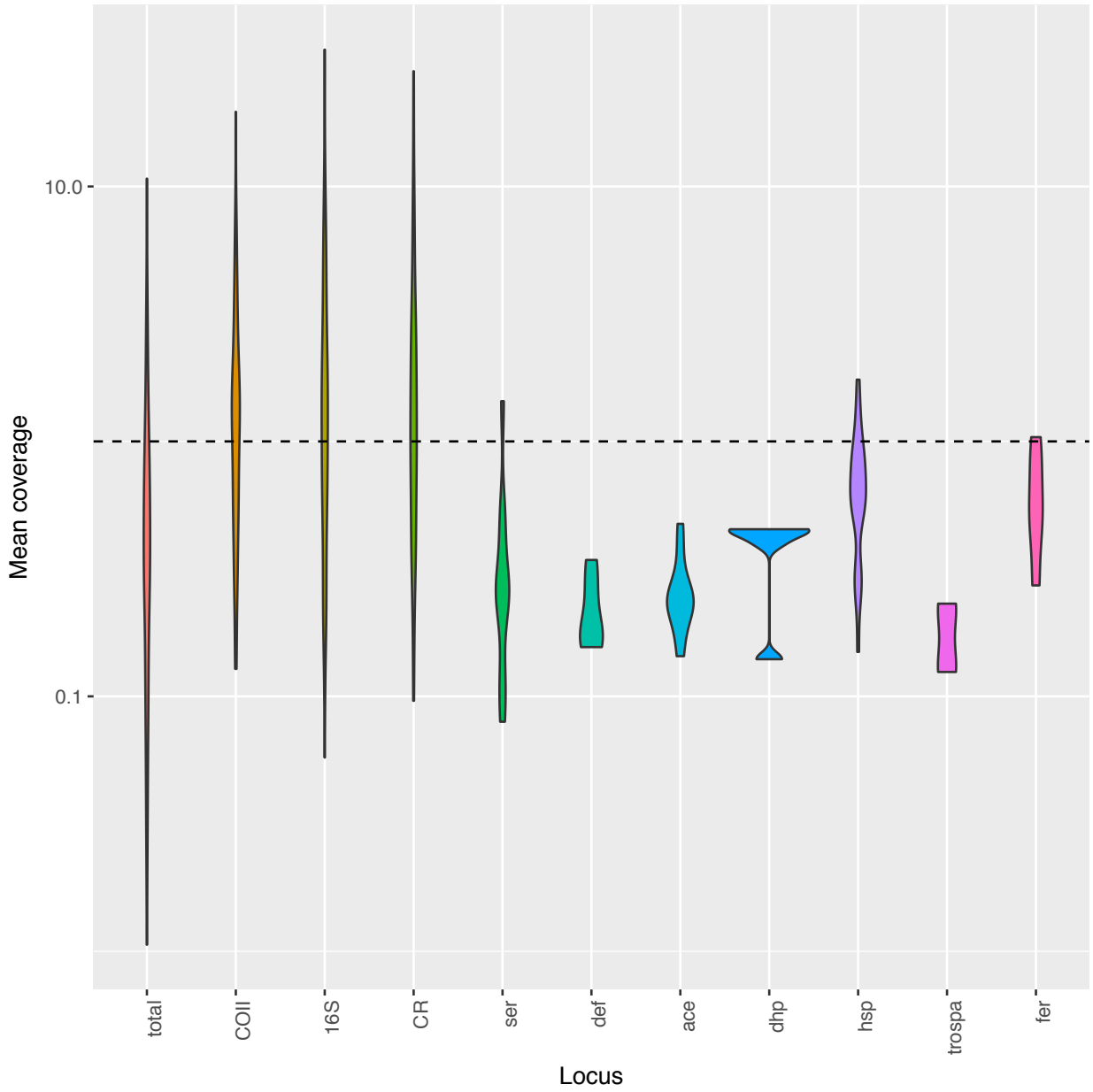

a
Chromosome
lp54
0.3
0.3
b
Chromosome
cp26
0.3
0.3



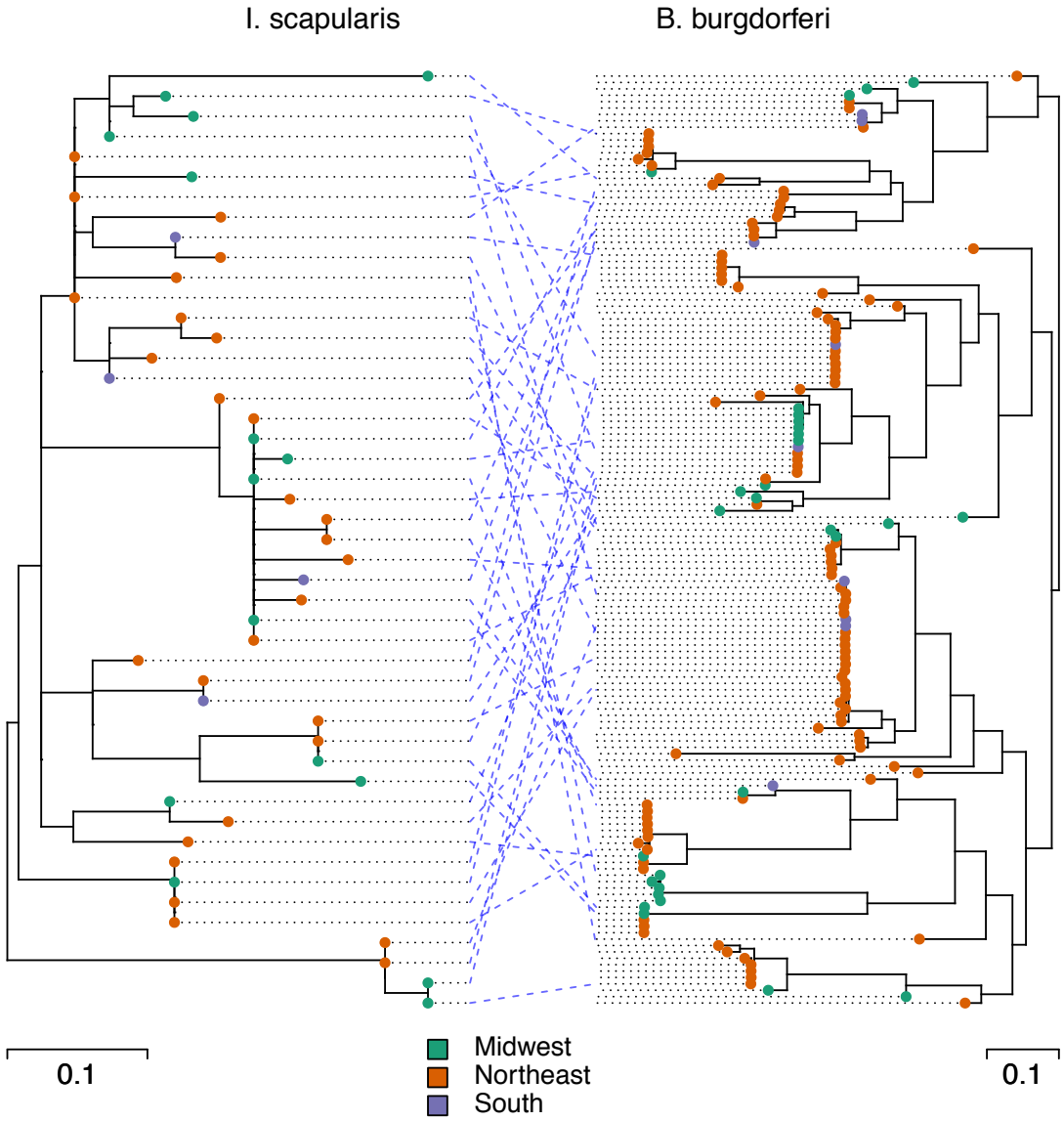
Supplementary Figure 4. Co-evolution of *B. burgdorferi* chromosome and plasmids lp54 and cp26. Co-phylogenies of (a) the chromosome and plasmid lp54, (b) the chromosome and plasmid cp26, and (c) plasmids lp54 and cp26. Phylogenies are unrooted maximum likelihood trees. Branches are rotated to optimize matching of tips[1]. Blue dotted lines connect tips corresponding to the same tick sample. Branches with greater than 50% bootstrap support are labeled with support value. Branch lengths indicate the estimated number of substitutions per variable site. Co-phylogenies show significant support for co-evolution of chromosome and plasmids with several historic plasmid exchanges between lineages. A global test (Parafit)[2] supports co-evolution of the chromosome and plasmid lp54 ($p < 0.01$), chromosome and plasmid cp26 ($p < 0.01$), and both plasmids ($p < 0.01$).



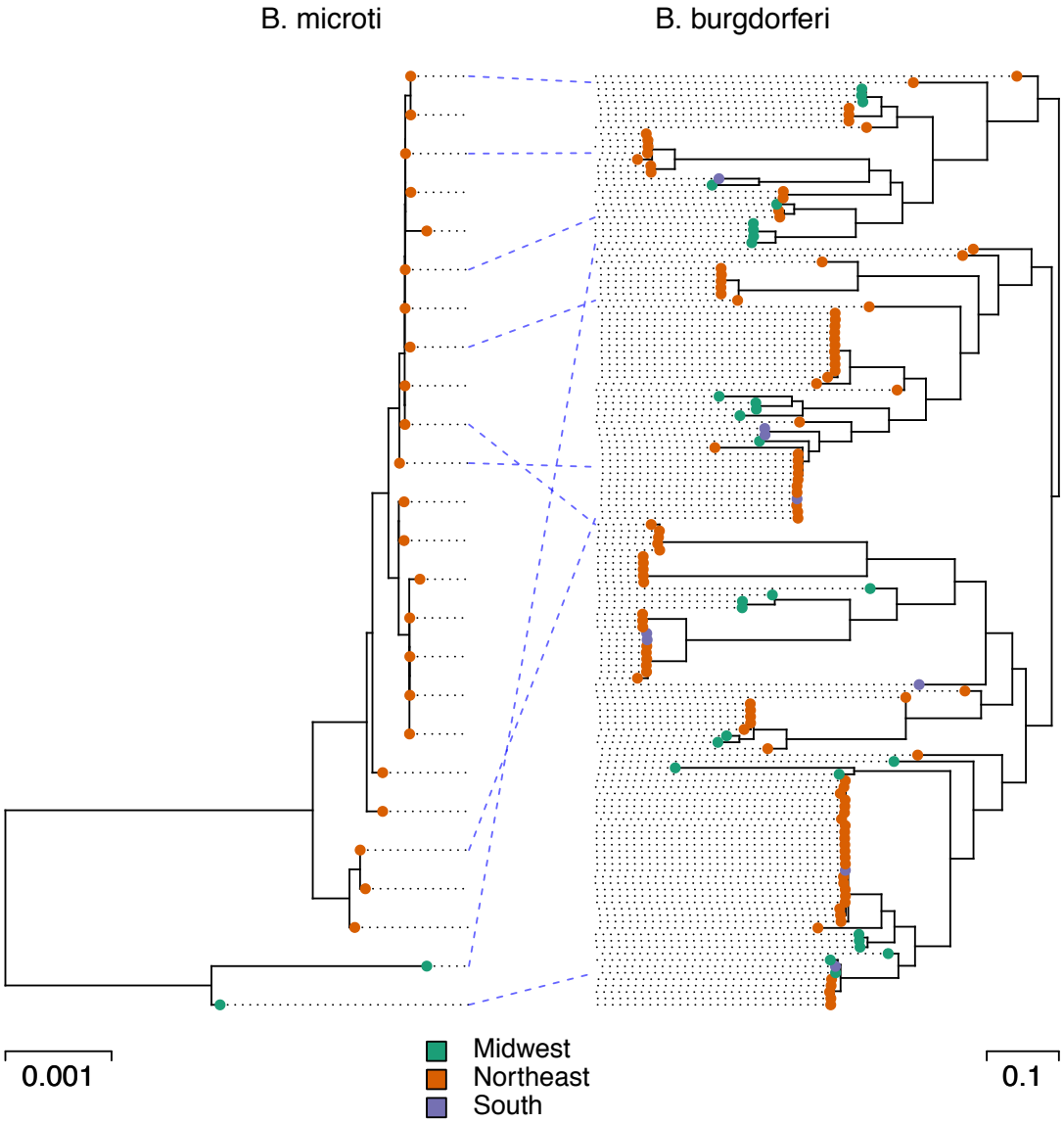
Supplementary Figure 5. Sequence by-catch maps to the tick genome. Each point represents a single sample; the proportion of reads (log-scale) mapping to *B. burgdorferi* (x-axis) and the tick, *I. scapularis* (y-axis). This sequence by-catch allows us to use hybrid capture data to investigate variation in the tick vector.



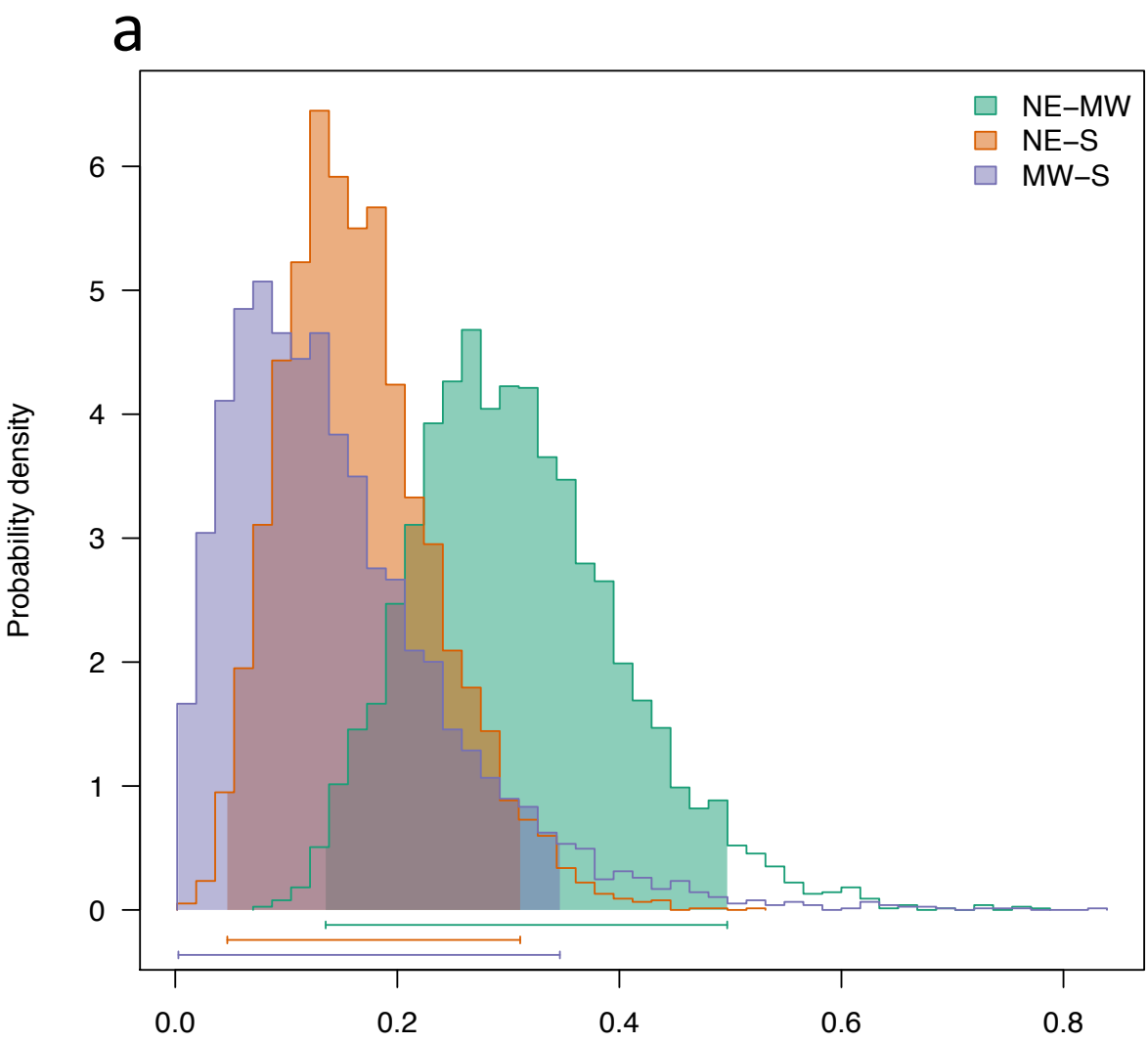
Supplementary Figure 6. Violin plot of coverage of 10 tick genes. The violin plot shows the probability density of mean coverage or the proportion of samples with a mean coverage of *y* for each gene. The dotted line represents 1 X mean coverage. *COII*, *16S*, and *CR* (cytochrome oxidase II, 16S, and the control region) are mitochondrial genes with > 1.5 X mean coverage and were used for population genetic study of *I. scapularis*. The remaining nuclear genes have < 1X average coverage, precluding analysis of these loci.



Supplementary Figure 7. Sequence by-catch enables co-evolutionary study of *B. burgdorferi* and its tick vector. Co-phylogeny of *I. scapularis* (left) and *B. burgdorferi* (right). Branches have been rotated to maximize links between tree tips and tips are colored by sampling region. The *I. scapularis* maximum likelihood phylogeny is based on three concatenated mitochondrial genes (*COII*, *16S*, and *CR*). Branch lengths indicate the estimated number of substitutions per variable site. While the midwestern *B. microti* samples are monophyletic and strongly differentiated from northeastern *B. microti*, the *B. burgdorferi* sampled from the same ticks fall into two divergent clades (Supplementary Figure 7). A global test of co-evolution for the tick and bacteria confirms their independent evolutionary histories (p-value = 0.403) [2,3].

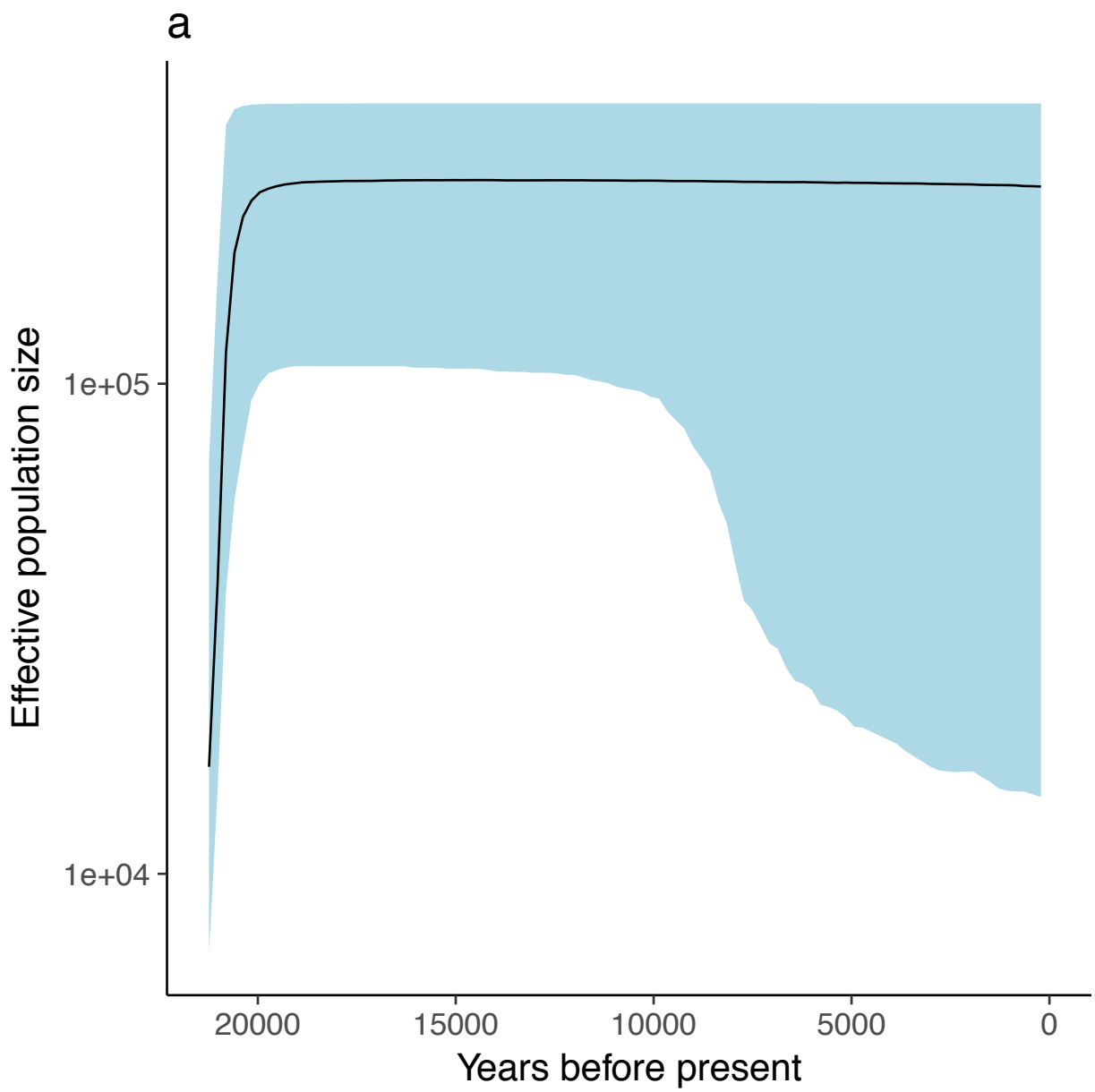


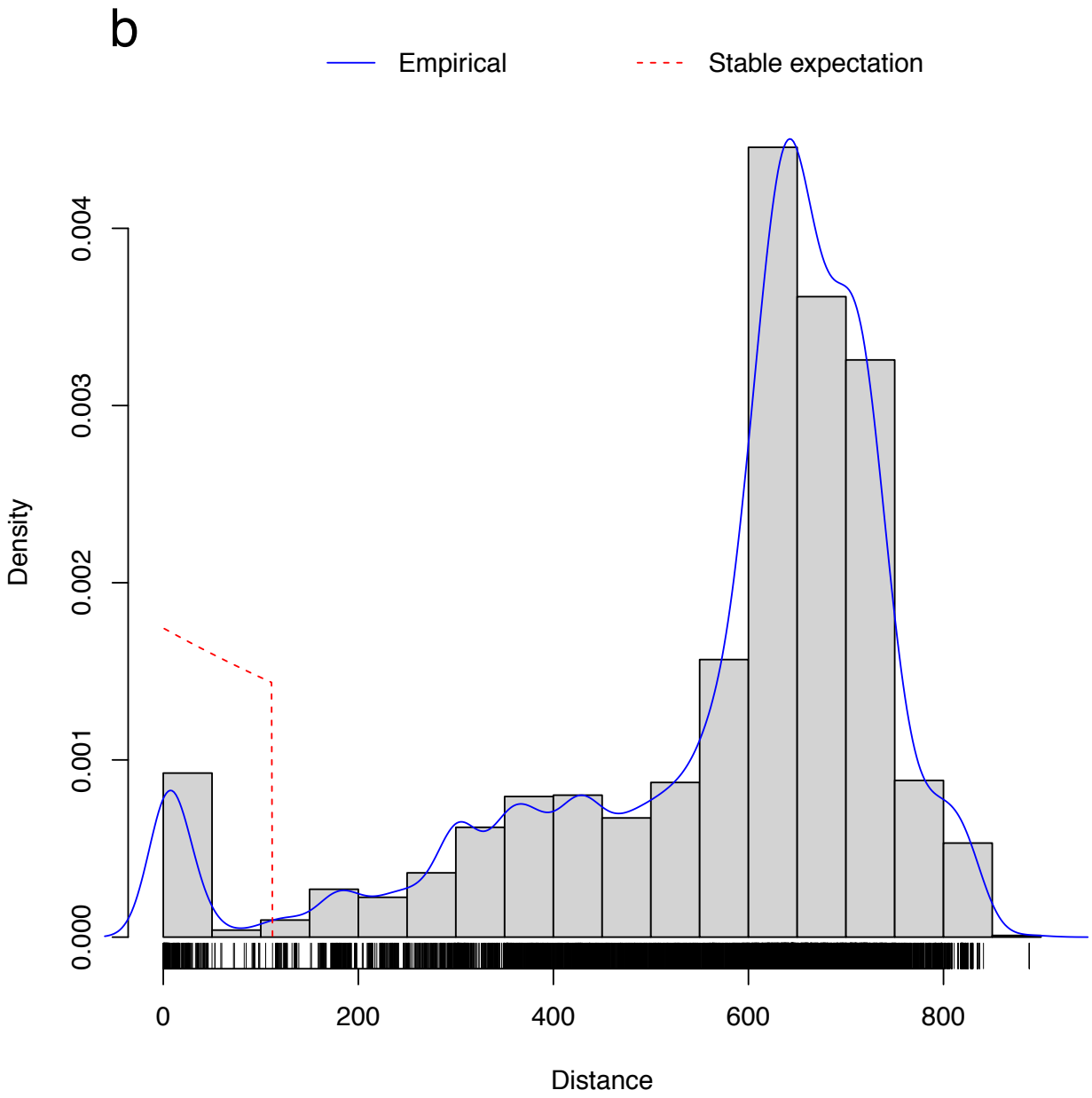
Supplementary Figure 8. Co-captured *Babesia microti* sequence enables co-evolutionary study of the two co-vectored parasites. Co-phylogeny of the *B. microti* apicoplast (left) and *B. burgdorferi* (right). Branches have been rotated to maximize links between tree tips and tips are colored by sampling region. Branch lengths indicate the estimated number of substitutions per variable site. A global test of coevolution of the two parasites reveals no support for coevolution (p-value = 0.251)[2].


a
NE–MW
NE–S
MW–S
Probability density
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8


b

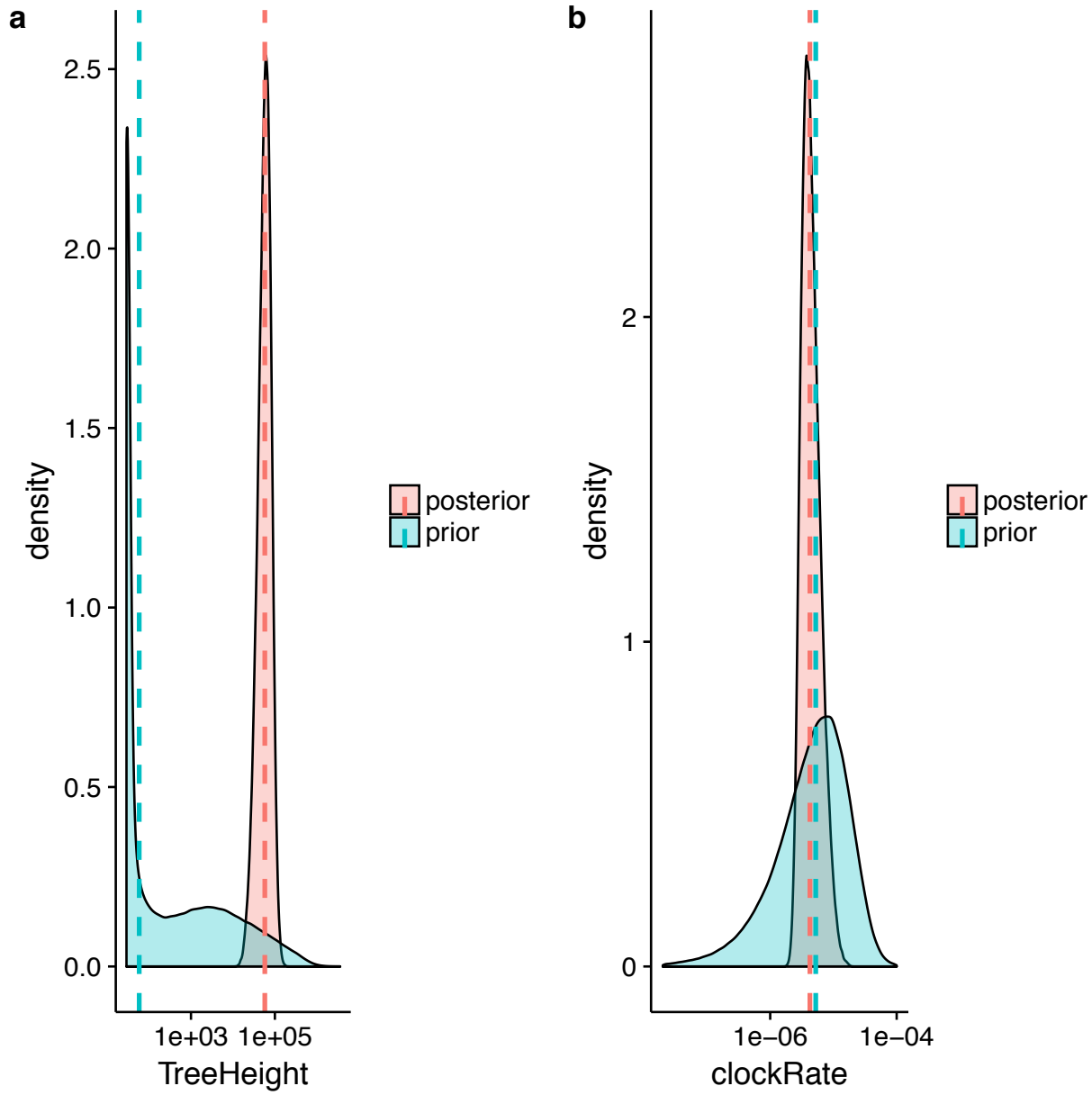
Supplementary Figure 9. Long distance migration shapes diversity. To perform ancestral state reconstruction, we estimated migration rates between each sampled region (Northeast, Midwest, and South) with a Markov model implemented in diversitree[4]. (a) The posterior distribution of bidirectional migration rates. Horizontal bars beneath the distributions correspond to the 95% credibility intervals and overlap, revealing a lack of support for differential migration rates between regions. However, there is a hierarchy (though not statistically significant) in migration rates. Migration between the Northeast and Midwest is most frequent, followed by migration between the Northeast and South. Migration between the South and Midwest is least frequent. (b) Relative rates of migration between sampled regions visualized on a map of North America.


a
Effective population size
1e+05
1e+04
20000
15000
10000
5000
0
Years before present



Supplementary Figure 10. Historic population expansion. (a) Bayesian skyline plot showing changes in estimated mean *B. burgdorferi* effective population size over time (black line) with 95% credible intervals (blue shading)[5]. (b) A histogram of the observed distribution of pairwise distances between samples. The lines show an empirical density estimate (blue) and the

expected mismatch distribution for a stable population (red)[6,7].



Supplementary Figure 11. Posterior parameter estimates inferred from sequence data and from the priors. Posterior density plot for (a) tree height (years) and (b) clock rate (substitutions/site/year) inferred from fitting the best-fitting evolutionary model to the *B. burgdorferi* sequence data (i.e. posterior, pink) and to the prior, without sequence data (blue), reveal that priors do not overwhelm sequence data. Dashed lines indicate the median parameter estimate.

Supplementary Table 1. Previously published *B. burgdorferi* sensu stricto genomes used in this study. For a subset of analyses, we included all 15 previously published *B. burgdorferi* genomes, representing cultured isolates from ticks, humans, and a song sparrow[8–18].

Sample Name	Sample Type	Location	Lat*	Long*	Year*	Ref	ENA Assembly
Borrelia_burgdorferi_b31	*I.scapularis*	NY	41.0572	-72.3163	1982	[8,9]	ASM868v2
Borrelia_burgdorferi_29805	*I. scapularis*	CT	41.5834	-71.7622	1993	[10]	ASM17229v2
Borrelia_burgdorferi_64b	skin	NY	41.122	-73.7949	2003	[11]	ASM17233v2
Borrelia_burgdorferi_72a	skin	NY	41.122	-73.7949	2003	[11]	ASM17175v2
Borrelia_burgdorferi_94a	skin	NY	41.122	-73.7949	2003	[11]	ASM18171v2
Borrelia_burgdorferi_118a	skin	NY	41.122	-73.7949	2003	[11]	ASM17173v2
Borrelia_burgdorferi_156a	skin	NY	41.122	-73.7949	2003	[11]	ASM18155v2
Borrelia_burgdorferi_bol26	skin	Italy	41.8719	12.5674	2003	[11]	ASM18157v2
Borrelia_burgdorferi_ca_11_2a	*I. pacificus*	CA	38.5816	-121.494	1985	[12]	ASM17231v2
Borrelia_burgdorferi_jd1	*I.scapularis*	MA	42.6792	-70.8412	1986	[13]	ASM16665v1
Borrelia_burgdorferi_n40	*I.scapularis*	NY	41.122	-73.7949	1988	[14]	ASM16663v1
Borrelia_burgdorferi_wi91_23	*Melospiza melodia* (song sparrow)	WI	45.4162	-92.6468	1991	[15]	ASM18185v2
Borrelia_burgdorferi_zs7	*I. ricinius*	Germany	51.1657	10.4515	1989	[16]	ASM2140v1
Borrelia_finlandensis	*I. ricinius*	Finland	61.9241	25.7482	2008	[17]	ASM18187v2
Borrelia_burgdorferi_ca8	*I. pacificus*	California	38.5816	-121.494	1998	[18]	ASM38256v1
Borrelia_burgdorferi_ca382	*I. pacificus*	California	38.5816	-121.494	1998	[3]	ASM44446v1

*Some previously published genomes do not have geographic coordinates or sampling year; we approximated coordinates based on geographic information available in source publication and year based on year of first publication discussing sample.

Supplementary Table 2. Coverage of ten tick genes present in sequence by-catch. The accession number and reference indicates the reference tick haplotype used for mapping. The three mitochondrial genes (COII, 16S, and CR) with mean coverage > 1.5 X were used for population genetic study.

Locus	Gene Name	Mean Cov.	Std. dev.	Location	Accession number	Ref.
16S	16S	2.68	4.97	mitochondria	KR092226	[19]
COII	Cytochrome oxidase II	1.89	2.69	mitochondria	KR092237	[19]
CR	Control region	2.54	4.26	mitochondria	KR092282	[19]
ace	Acetylcholinesterase	0.26	0.09	nucleus	KT847953	[20]
def	Defensin	0.22	0.08	nucleus	KT849839	[20]
dhp	Dihydropyrimidine	0.38	0.13	nucleus	KT849259	[20]
fer	Ferredoxin-glutamate synthase	0.62	0.30	nucleus	KT848579	[20]
hsp	Heat shock protein	0.69	0.38	nucleus	KT847275	[20]
ser	Serotonin	0.36	0.38	nucleus	KT846601	[20]
trospa	TROSPA	0.18	0.06	nucleus	KT850389	[20]
Total		**0.77**	**1.42**			

Supplementary Table 3. Posterior probability distributions of parameters estimated in BEAST. Mean, standard error, standard deviation, median, lower (5%) and upper (95%) limits of the highest posterior density (HPD) interval, auto-correlation time (ACT), and effective sample size (ESS). ESS > 200 is frequently used to assess parameter convergence. UCLD mean rate is the mean rate under the uncorrelated log-normally distributed (UCLD) relaxed molecular clock model, allowing variation in substitution rate across lineages, in units of substitutions per site per year. The strict clock rate parameter is substitution rate under a strict molecular clock model (no substitution rate variation across lineages), again in units of substitutions per site per year. Tree height is the age of the most recent common ancestor of sampled *B. burgdorferi*, in years.

Parameter	Mean	Std. Err.	Std. Dev.	Median	$HPD_{5\%}$	$HPD_{95\%}$	ACT	ESS
Posterior	-8.44E+04	1.17E+00	5.33E+01	-8.44E+04	-8.45E+04	-8.43E+04	1.15E+05	2.08E+03
Likelihood	-8.26E+04	2.24E-01	1.18E+01	-8.26E+04	-8.27E+04	-8.26E+04	8.68E+04	2.77E+03
Prior	-1.77E+03	1.11E+00	5.19E+01	-1.77E+03	-1.86E+03	-1.66E+03	1.10E+05	2.19E+03
Tree height	5.89E+04	8.54E+02	2.08E+04	5.70E+04	2.05E+04	9.83E+04	4.03E+05	5.95E+02
UCLD mean rate	7.49E-06	6.93E-08	2.92E-06	6.79E-06	3.32E-06	1.34E-05	1.35E+05	1.78E+03
UCLD std. dev.	1.04E+00	4.80E-03	7.34E-02	1.04E+00	9.00E-01	1.19E+00	1.03E+06	2.34E+02
Strict clock mean rate	4.75E-06	3.90E-08	1.85E-06	4.27E-06	2.28E-06	8.56E-06	1.07E+05	2.24E+03
Strict clock rate variance	1.05E-10	3.05E-12	1.08E-10	7.18E-11	9.97E-12	2.96E-10	1.92E+05	1.25E+03
Strict clock rate coef. of variation	1.28E+00	8.63E-03	1.66E-01	1.27E+00	9.84E-01	1.61E+00	6.50E+05	3.69E+02

Supplementary Table 4. Comparison of four clock and demographic models in BEAST. The marginal likelihood of four models were calculated with path sampling. The relaxed clock, skyline population size model has the highest marginal likelihood, referred to as M_1. Bayes Factors (BF), or the ratio of marginal likelihoods, are calculated by subtracting each marginal likelihood from that of M_1. Positive Bayes factors indicate support for M_1 with respect to the other models. Bayes factors greater than 20 are generally considered strong support for a given model[5].

Model	Marginal likelihood	Bayes factor
Strict clock, constant pop	-85546.24987	1862.962164
Strict clock, skyline pop	-85477.90002	1794.61232
Relaxed clock, constant pop	-83765.47245	82.18474709
Relaxed clock, skyline pop (M_1)	**-83683.2877**	-

Supplementary File 1. Sample descriptions. For each bacterial sample, the biological source, sampling location, and *B. burgdorferi* mapping statistics.

Supplementary References

1. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3,** 217–223 (2012).
2. Legendre, P., Desdevises, Y. & Bazin, E. A Statistical Test for Host–Parasite Coevolution. *Syst. Biol* **51,** 217–234 (2002).
3. Humphrey, P. T., Caporale, D. A. & Brisson, D. Uncoordinated phylogeography of *Borrelia burgdorferi* and its tick vector, *Ixodes scapularis*. *Evolution (N. Y).* **64,** 2653–63 (2010).
4. FitzJohn, R. G. Diversitree : comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* **3,** 1084–1092 (2012).
5. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7,** 214 (2007).
6. Rogers, A. R. & Harpending, H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9,** 552–69 (1992).
7. Paradis, E. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* **26,** 419–420 (2010).
8. Fraser, C. M. *et al.* Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390,** 580–6 (1997).
9. Burgdorfer, W. *et al.* Lyme disease-a tick-borne spirochetosis? *Science* **216,** 1317–9 (1982).
10. Zingg, B. C., Anderson, J. F., Johnson, R. C. & LeFebvre, R. B. Comparative analysis of genetic variability among *Borrelia burgdorferi* isolates from Europe and the United States by restriction enzyme analysis, gene restriction fragment length polymorphism, and pulsed-field gel electrophoresis. *J Clin Microbiol* **31,** 3115–3122 (1993).
11. Qiu, W.-G. *et al.* Genetic exchange and plasmid transfers in *Borrelia burgdorferi* sensu stricto revealed by three-way genome comparisons and multilocus sequence typing. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 14150–5 (2004).
12. Schwan, T. G. *et al.* Distribution and molecular analysis of Lyme disease spirochetes, *Borrelia burgdorferi*, isolated from ticks throughout California. *J. Clin. Microbiol.* **31,** 3096–3108 (1993).
13. Piesman, J., Mather, T. N., Sinsky, R. J. & Spielman, A. Duration of tick attachment and *Borrelia burgdorferi* transmission. *J. Clin. Microbiol.* **25,** 557–558 (1987).
14. Barthold, S. W. *et al.* Experimental Lyme arthritis in rats infected with *Borrelia burgdorferi*. *J Infect Dis* **157,** 842–846 (1988).
15. McLean, R. G., Ubico, S. R., Hughes, C. A. N., Engstrom, S. M. & Johnson, R. C. Isolation and characterization of *Borrelia burgdorferi* from blood of a bird captured in the Saint Croix River Valley. *J. Clin. Microbiol.* **31,** 2038–2043 (1993).
16. Simon, M. M. *et al.* Recombinant outer surface protein a from *Borrelia burgdorferi* induces antibodies protective against spirochetal infection in mice. *J. Infect. Dis.* **164,** 123–132 (1991).
17. Casjens, S. R. *et al.* Whole genome sequence of an unusual *Borrelia burgdorferi* sensu lato isolate. *J. Bacteriol.* **193,** 1489–90 (2011).
18. Brown, R. N., Peot, M. A. & Lane, R. S. Sylvatic maintenance of *Borrelia burgdorferi* (Spirochaetales) in Northern California: untangling the web of transmission. *J. Med. Entomol.* **43,** 743–751 (2006).
19. Khatchikian, C. E. *et al.* Recent and rapid population growth and range expansion of the Lyme disease tick vector, *Ixodes scapularis* , in North America. *Evolution (N. Y).* **69,** 1678–1689 (2015).
20. Van Zee, J., Piesman, J. F., Hojgaard, A. & Black, W. C. Nuclear markers reveal

predominantly north to south gene flow in *Ixodes scapularis* , the tick vector of the Lyme disease spirochete. *PLoS One* **10,** 1–22 (2015).