environmental microbiology

Environmental Microbiology (2011) **13**(9), 2453–2467 doi:10.1111/j.1462-2920.2011.02515.x

Genetic structure of marine *Borrelia garinii* and population admixture with the terrestrial cycle of Lyme borreliosis

Elena Gómez-Díaz,[1,2]* Thierry Boulinier,[3] Natacha Sertour,[4] Muriel Cornet,[4] Elisabeth Ferquel,[4] and Karen D. McCoy[1]

[1]*MIVEGEC, CNRS 5290-IRD 224-UM1-UM2, IRD, 34394, Montpellier, France.*

[2]*Institute of Evolutionary Biology, CSIC-UPF, 08003, Barcelona, Spain.*

[3]*Centre d'Ecologie Fonctionnelle et Evolutive, UMR CNRS 5175, 34294, Montpellier, France.*

[4]*Centre National de Référence des Borrelia, Institut Pasteur, 75724 Paris Cedex 15, France.*

Summary

Despite the importance of population structure for the epidemiology of pathogenic bacteria, the spatial and ecological heterogeneity of these populations is often poorly characterized. Here, we investigated the genetic diversity and population structure of the Lyme borreliosis (LB) spirochaete *Borrelia garinii* in its marine cycle involving colonial seabirds and different host races of the seabird tick *Ixodes uriae*. Multilocus sequence analyses (MLSA) on eight chromosomal and two plasmid loci (*ospA* and *ospC*) indicate that *B. garinii* circulating in the marine system is highly diverse. Microevolution in marine *B. garinii* seems to be mainly clonal, but recombination and selection do occur. Sequence types were not evenly distributed among geographic regions, with substantial population subdivision between Atlantic and Pacific Ocean basins. However, no geographic structuring was evident within regions. Results of selection analyses and phylogenetic discordance between chromosomal and plasmid loci indicate adaptive evolution is likely occurring in this system, but no pattern of host or vector-associated divergence was found. Recombination analyses showed evidence for population admixture between terrestrial and marine strains, suggesting that LB spirochaetes are exchanged between these enzootic cycles. Importantly, our results highlight the need to explicitly consider the marine system for a complete understanding of the evolutionary ecology and global epidemiology of Lyme borreliosis.

Received 1 December, 2010; accepted 20 April, 2011. *For correspondence. E-mail elena.gomez@ibe.upf-csic.es; Tel. (+34) 932309500; Fax (+34) 932211011.

Introduction

Many pathogenic bacteria display clonal population structure, but even within a single species, genetic structure can be dynamic and vary in space and time (Halkett *et al*., 2005). A number of genetic processes can be responsible for this variability and may affect the emergence of new variants (Feil and Enright, 2004). For instance, bacterial exchange and homologous recombination can lead to population admixture and reduced linkage disequilibrium with important consequences for pathogen evolution and population demography (Feil and Spratt, 2001; Feil *et al*., 2001; Spratt *et al*., 2001). Likewise, geographic isolation, or ecological niche specialization, can result in population differentiation and lead to spatial or temporal differences in the distribution and frequency of bacterial clones (Spratt and Maiden, 1999). The evolution of distinct bacterial evolutionary entities, with distinct ecologies and/or epidemiological characteristics, is an essential question for disease control strategies (Woolhouse *et al*., 2005). However, the global ecological diversity and population structure of many bacterial pathogens remain poorly understood (Spratt *et al*., 2001; Woolhouse *et al*., 2005), and particularly so in the case of vector-borne pathogens, where multiple host and vector-associated factors may influence the occurrence and frequency of these diversifying processes.

Lyme borreliosis (LB) is the most common vector-borne zoonosis in temperate regions of the Northern hemisphere and is caused by spirochaetes of the *Borrelia burgdorferi sensu lato* group (Stanek and Strle, 2003; Kurtenbach *et al*., 2006). The obligate enzootic life cycle of these spirochaetes involves primarily *Ixodes* ticks and a variety of vertebrate hosts, including mammals, reptiles and birds (Gray *et al*., 2002). Despite a wide host/vector spectrum, some degree of ecological specialization exists among the different species of the complex, particularly towards the vertebrate host (Kurtenbach *et al*., 2002; 2006). Although LB bacteria are considered to be largely clonal (Dykhuizen *et al*., 1993), recombination of large

genomic segments by plasmid transfers among closely related *B. burgdorferi sensu lato* strains is known to occur and is considered to be the main mechanism driving diversification (Dykhuizen and Baranton, 2001; Stevenson, 2001; Qiu *et al*., 2004). However, the nature of recombination and its implications for the evolution and emergence of new variants depends on the population structure of different clones and the infection dynamics within host and vector populations, which condition opportunities for different strains to exchange genetic material. In particular, recombination among LB spirochaetes is thought to most likely occur in the tick vector owing to high bacteria loads during transmission (Kurtenbach *et al*., 2006). However, to date, we know little about how frequently recombination occurs or how the population structure of the vector may act in shaping that of the bacterium (but see Qiu, 2002).

The marine cycle of LB was discovered almost two decades ago when Olsen and colleagues (1993) demonstrated the circulation of one of the pathogenic species of the LB complex, *Borrelia garinii*, among seabirds via the tick *Ixodes uriae*. However, the potential interaction with the terrestrial cycle and thus, the possible contribution of the marine cycle to LB epidemiology is rarely considered (but see Comstedt *et al*., 2009). Indeed, seabirds typically have large populations, wide geographic distributions and high mobility. Previous evidence suggests they may act as potential reservoirs in the global circulation and large-scale dispersal of LB bacteria (Olsen *et al*., 1995; Olsen, 2003; Duneau *et al*., 2008; Staszewski *et al*., 2008). At the local scale, colonial seabirds tend to breed in large aggregations where several species can nest sympatrically. Although this situation can promote interspecific pathogen exchange, the tick vector in this system consists of multiple, distinct host races (McCoy *et al*., 2001; 2003) which should limit the local mixing of pathogens (Duneau *et al*., 2008; McCoy, 2008). This ecological scenario may therefore favour the evolution of specialization of *Borrelia* bacteria circulating in different seabird host species and tick populations. A recent study by Comstedt and colleagues (2009), based on the intergenic spacer (IGS) region, suggested that the population structure of *B. garinii* is complex due to the different avian and vector species potentially implicated in its circulation. In addition to the problems associated with inferring strain identity from a single locus (Gómez-Díaz, 2009; Travinsky *et al*., 2010), this study was based on a limited number of geographic locations with no way to disentangle geographic and host-related structure. Based on existing information, it therefore remains difficult to make clear predictions about the population structuring and the role of different seabirds and their ticks in the global epidemiology of this pathogen.

Here, we examine the population structure of *B. garinii* in its marine cycle using a comprehensive approach based on the multilocus sequence analysis (MLSA) of eight chromosomal and two plasmid loci. In particular, we examine the genetic diversity and phylogeography of *B. garinii* associated with seabird populations from both the Atlantic and Pacific Ocean basins, explicitly taking into account geographic and host associations within each basin. Using this data set, we also explore the degree of structure and/or potential exchange between the two ecological enzootic cycles of *B. garinii*, the marine and terrestrial LB systems. By integrating different types of loci (plasmid and chromosomal genes) and an extensive geographic sample, we uncover patterns of population structure of these spirochaetes and improve our understanding of the factors that shape their evolution.

Results

Genetic statistics

We characterized diversity at 10 genes for a total of 53 *Borrelia* isolates from the marine and terrestrial LB systems (Fig. 1A; Table 1). Genetic statistics for the concatenated data set and for each individual locus are shown in Table 2. Both plasmid loci showed higher nucleotide diversity estimates compared with chromosomal genes. Among chromosomal loci, *recA* showed the highest variability. Most alleles were unique, highlighting the great allelic diversity among marine strains. However, a few were shared between North Atlantic and North Pacific regions, and between terrestrial and marine *B. garinii* strains (Fig. 1B, Table S1). The comparison with previously published allele sequences from the terrestrial cycle revealed a moderate degree of allele matching; this similarity was more evident for *pyrG* and *recG* genes for which a relatively large sample was available (see Table S1).

Phylogenetic analyses

The phylogenetic tree, built using the eight chromosomal genes, showed that all *Borrelia* isolates from the present study grouped with the *20047 B. garinii* reference strain (Fig. 2). In general, there was no phylogenetic structure according to tick host race/seabird host species, but a certain degree of geographic subdivision was apparent as Pacific, Atlantic and terrestrial *B. garinii* strains tended to group in separate clades (Fig. 2). The tree showed two major well-supported clusters [Bootstrap Support (BS) 100%]: the most basal included two highly divergent Pacific-marine isolates, and the second grouped all other *B. garinii* strains. Within this latter cluster, Pacific isolates also appeared well differentiated and tended to group apart from both Atlantic and terrestrial *B. garinii* isolates (BS 98–100%). However, deeper phylogenetic relation-

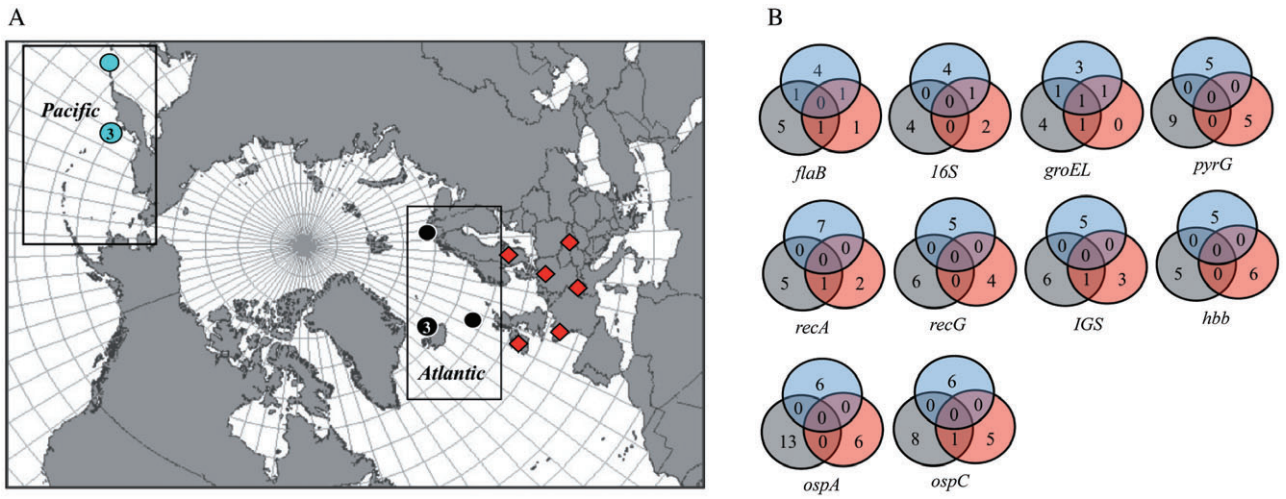

Fig. 1. A. Geographic map of sampling locations for the 53 *Borrelia garinii* isolates analysed in this study. Circles and diamonds correspond to marine and terrestrial isolates, respectively, and colours indicate the geographic region: North Atlantic (black), North Pacific (blue) or terrestrial (red). In the marine system, several distinct colonies were sampled in Iceland and in the Commander Islands (numbers within circles, see text for details).
B. Venn diagram of shared alleles between Atlantic (grey), Pacific (blue) and terrestrial (red) *B. garinii* isolates. Values inside the circles are exclusive alleles of these three groups, whereas the numbers in the overlap areas indicate the number of shared alleles. It should be noted that allele numbers are not based on the same sample sizes for Atlantic ($n = 35$), Pacific ($n = 7$) and terrestrial ($n = 11$) systems and therefore cannot be directly compared in terms of diversity estimates.

ships among Pacific, Atlantic and terrestrial clades were not well resolved and appeared polyphyletic. Atlantic isolates formed five well-supported internal clades, whereas most terrestrial isolates grouped into two well-supported clades (BS 97% and 98%), with the interesting exception of the reference strain *20047* that appeared in a more basal position. This lack of phylogenetic structure was also evident when comparing *recG* and *pyrG* alleles with terrestrial strains from public reference databases; a star-like network was obvious with marine strains intermingled with those of the terrestrial cycle and high levels of either homoplasy or recombination between the two cycles (Figs S1 and S2).

Compared with chromosomal loci, the *ospC* and *ospA* maximum likelihood (ML) phylogenetic trees showed a different picture with no general agreement between clusters from the two sets of loci. Basal clades showed low bootstrap support, but external nodes were well supported and several highly divergent groups were evident (Figs S3 and S4). Despite the fact that genetic diversity was higher in the two plasmid loci compared with the other genes (see Table 2), little phylogeographic structure was apparent, particularly for the *ospC* gene. In the case of the *ospA* gene, four groups were identified in the Atlantic and three in the Pacific. In contrast to marine strains, the majority of terrestrial isolates clustered together, with the exception of *20047*, which grouped with Atlantic marine isolates and *BR14*, which formed a clade with some Pacific marine strains (Fig. S3). Similarly, the *ospC* locus revealed five well-supported groups in the Atlantic and three in the Pacific. In this case, however, terrestrial strains were intermingled with marine strains throughout the phylogenetic tree (Fig. S4). More generally, comparison with *ospA* and *ospC* sequences available in GenBank confirmed the high diversity of marine strains and the similarity to those of the terrestrial cycle (Figs S5 and S6).

In terms of previously defined major serotype groups for *ospA* and *ospC* genes (i.e. different *Borrelia* phenotypes based on reactivity to monoclonal antibodies), phylogenetic and distance analyses clustered marine *B. garinii* strains with sequences representing OspA serotype 3 (18 isolates) and serotype 5 (2 isolates), which appeared to be restricted to the Atlantic and the Pacific regions respectively (Fig. S3). The remaining marine isolates (52%) could not be assigned to any of the eight described OspA serotype reference sequences. Terrestrial OspA sequences grouped with serotypes 6 and 8. Similarly, only a small proportion of marine *B. garinii* sequences matched previously described OspC serotype groups (Wilske *et al.*, 1996). In particular, eight Atlantic isolates matched *B. garinii* group OspC P, two Pacific isolates matched group OspC R and two others group OspC G (Fig. S4). Among terrestrial strains, only the *20047* reference sequence matched a known serotype group (OspC S).

Population structure

Considering only chromosomal loci, but controlling for the possible impact of recombination, the genealogy inferred

Table 1. *Borrelia garinii* isolates typed in this study at eight chromosomal genes.

Isolate	Species	Location	Abbreviation	Tick vector	Host
Atlantic					
T1131	*B. garinii*	Fair Is. – Britain	Fair	*I. uriae*	CG
T1377	*B. garinii*	Fair Is. – Britain	Fair	*I. uriae*	PF
T1638	*B. garinii*	Breidafjordur – Iceland	Brei	*I. uriae*	PF
T1656	*B. garinii*	Grimsey – Iceland	Grim	*I. uriae*	PF
T1666	*B. garinii*	Skrudur – Iceland	Skru	*I. uriae*	PF
T1667	*B. garinii*	Skrudur – Iceland	Skru	*I. uriae*	PF
T1669	*B. garinii*	Skrudur – Iceland	Skru	*I. uriae*	PF
T1673	*B. garinii*	Skrudur – Iceland	Skru	*I. uriae*	PF
T1682	*B. garinii*	Grimsey – Iceland	Grim	*I. uriae*	CG
T1697	*B. garinii*	Grimsey – Iceland	Grim	*I. uriae*	KT
T1699	*B. garinii*	Grimsey – Iceland	Grim	*I. uriae*	KT
T1708	*B. garinii*	Grimsey – Iceland	Grim	*I. uriae*	CG
T1712	*B. garinii*	Grimsey – Iceland	Grim	*I. uriae*	CG
T1727	*B. garinii*	Skrudur – Iceland	Skru	*I. uriae*	PF
T1789	*B. garinii*	Grimsey – Iceland	Grim	*I. uriae*	KT
T1793	*B. garinii*	Grimsey – Iceland	Grim	*I. uriae*	KT
T1794	*B. garinii*	Grimsey – Iceland	Grim	*I. uriae*	PF
T1795	*B. garinii*	Grimsey – Iceland	Grim	*I. uriae*	PF
T1796	*B. garinii*	Grimsey – Iceland	Grim	*I. uriae*	PF
T1798	*B. garinii*	Grimsey – Iceland	Grim	*I. uriae*	PF
T1799	*B. garinii*	Grimsey – Iceland	Grim	*I. uriae*	PF
T1815	*B. garinii*	Breidafjordur – Iceland	Brei	*I. uriae*	RZ
T1866	*B. garinii*	Breidafjordur – Iceland	Brei	*I. uriae*	PF
T1867	*B. garinii*	Breidafjordur – Iceland	Brei	*I. uriae*	PF
T1869	*B. garinii*	Breidafjordur – Iceland	Brei	*I. uriae*	PF
T1874	*B. garinii*	Breidafjordur – Iceland	Brei	*I. uriae*	PF
T1913	*B. garinii*	Breidafjordur – Iceland	Brei	*I. uriae*	FU
T1938	*B. garinii*	Hornøya – Norway	Horn	*I. uriae*	PF
T1940	*B. garinii*	Hornøya – Norway	Horn	*I. uriae*	PF
T2017	*B. garinii*	Grimsey – Iceland	Grim	*I. uriae*	PF
T2021	*B. garinii*	Skrudur – Iceland	Skru	*I. uriae*	PF
T2022	*B. garinii*	Skrudur – Iceland	Skru	*I. uriae*	PF
T2023	*B. garinii*	Skrudur – Iceland	Skru	*I. uriae*	PF
T2024	*B. garinii*	Skrudur – Iceland	Skru	*I. uriae*	PF
T2299	*B. garinii*	Hornøya – Norway	Horn	*I. uriae*	PF
Pacific					
T2645	*B. garinii*	Kuril Is. – Russia	Ku-KA	*I. uriae*	KT
T2712	*B. garinii*	Commander Is. – Russia	Co-Ka	*I. uriae*	KT
T2885	*B. garinii*	Commander Is. – Russia	Co-Ka	*I. uriae*	TPF
T2916	*B. garinii*	Commander Is. – Russia	Co-Ka	*I. uriae*	TPF
T2929	*B. garinii*	Commander Is. – Russia	Co-Ka	*I. uriae*	TPF
T2937	*B. garinii*	Commander Is. – Russia	Co-Ka	*I. uriae*	TPF
T3221	*B. garinii*	Kuril Is. – Russia	Ku-KA	*I. uriae*	PU
Terrestrial					
IPT28	*B. garinii*	Alsace – France		*I. ricinus*	
IPT114	*B. garinii*	Alsace – France		*I. ricinus*	
IPT156	*B. garinii*	Auvergne – France		*I. ricinus*	
IPT167	*B. garinii*	Limousin – France		*I. ricinus*	
IPT189	*B. garinii*	Normandy – France		*I. ricinus*	
Conn A3/3	*B. garinii*	Ireland		*I. ricinus*	
G25	*B. garinii*	Sweden		*I. ricinus*	
152-B17	*B. garinii*	Germany		*I. ricinus*	
BR-14	*B. garinii*	Czech Republic		*I. ricinus*	
CNE 83	*B. garinii*	Switzerland		*I. ricinus*	
Reference					
20047	*B. garinii*	Brittany – France		*I. ricinus*	
B31	*B. burgdorferi*				
VS461	*B. afzelii*				
PotiB1	*B. lusitaniae*				

All isolates in the marine system were obtained from *Ixodes uriae* ticks and the seabird host for each tick is provided in the host column. For terrestrial strains, the original source of laboratory cultures was always *Ixodes ricinus*. Isolate *20047* is the reference *B. garinii* strain. See Fig. 1 for the geographic origins of *Borrelia* isolates.
CG, Common guillemot; KT, Black-legged Kittiwake; PF, Atlantic puffin; T-PF, Tufted puffin; RZ, Razorbill; FU, Northern fulmar; PU, Red-faced Cormorant.

Table 2. Genetic estimates for the 53 *Borrelia garinii* isolates analysed in this study based on eight chromosomal and two plasmid genes (*ospA* and *ospC*).

Locus	Function	L	P_S	N_A	A_D	π	D	ρ	ω_{M2a}	n_{M2a}
flaB	Structural flagellar protein	310	19	13	0.824	0.0089	2.4935	0.007	1.00	0
16S	16S ribosomal RNA	446	14	11	0.799	0.0051	2.2525	0.014	*ne*	*ne*
groEL	Chaperonin	303	13	9	0.831	0.0071	2.1205	0.000	1.00	0
pyrG	CTP synthase	637	54	19	0.901	0.0113	7.2199	0.000	1.00	1
recA	Recombinase A	221	34	25	0.954	0.0198	4.3832	0.036*	1.00	0
recG	ATP-dependent DNA helicase	682	49	17	0.888	0.0099	6.6952	0.004	1.00	2
IGS	5S-23S rRNA intergenic spacer	254	22	14	0.795	0.0116	2.521	0.009	*ne*	*ne*
hbb	Small DNA-binding protein	402	22	13	0.883	0.0090	3.6255	0.000	1.00	1
ospA	Outer surface lipoprotein	619	192	26	0.95	0.0761	48.148	0.020*	2.51	9
ospC	Outer surface lipoprotein	517	220	21	0.872	0.1325	68.2562	0.020*	4.40	17

L = sequence fragment length, P_S = number of polymorphic sites, N_A = number of alleles, A_D = allelic diversity, π = nucleotide diversity, D = average number of nucleotide differences between sequences, ρ = recombination rate 2Ner (r is region sequence length) by LDhat. Values for rho (ρ) were obtained by dividing the per-locus recombination rate estimate from LDhat by the sequence length [the asterisk (*) means significant $P < 0.05$]. $\omega = dN/dS$, and n = number of positively selected sites according to M2a in PAML. Selection estimates for non-coding genes were not performed (*ne*). All comparisons between models M2a and M1a and M7 and M8 for plasmid loci were significant ($P < 0.001$).

by Clonalframe was star-like in shape, likely due to the fact that most polymorphic sites in the data set were unique to each isolate (Fig. 3). This pattern is consistent with rapid clonal expansion, which also agrees with the mismatch distribution analyses on the North Atlantic *B. garinii* marine population (see below). The analysis by Clonalframe assigned individual sequence types to clades that are largely congruent with the phylogenetic analysis, including the separation between Pacific, Atlantic and terrestrial strains. In this case, however, when accounting for recombination, posterior probabilities increased and most genealogical relationships were resolved. Atlantic strains grouped into five independent clonal complexes that appeared well differentiated from both Pacific and terrestrial strains (with the exception of strains T2645 and *20047*), but did not group together and were paraphyletic with respect to Pacific strains. In this analysis, all terrestrial strains clustered together with the exception of *20047*, whereas Pacific bacteria formed several highly divergent and independent clonal complexes consisting of only one or two sequences.

Population differentiation based on chromosomal loci was highly significant between Atlantic and Pacific marine *B. garinii* populations ($\Phi_{ST} = 0.30$, $P < 0.001$), but also between marine and terrestrial isolates (Pacific $\Phi_{ST} = 0.29$, $P < 0.001$, Atlantic $\Phi_{ST} = 0.27$, $P < 0.001$). However, we could not reject the presence of a single geographically unstructured population of *B. garinii* within the Atlantic basin based on the geographic arrangement of populations (Britain, Iceland and Norway pairwise Φ_{ST} comparisons; $P_{BR\text{-}IC} = 0.50$, $P_{BR\text{-}NO} = 0.86$, $P_{IC\text{-}NO} = 0.15$). Likewise, the AMOVA analysis did not detect any evidence of host-associated genetic structure within this region and indices of population differentiation among different host-associated groups were not significant (kittiwake, puffin and common guillemot, pairwise Φ_{ST} comparisons $P_{KT\text{-}PF} = 0.71$, $P_{KT\text{-}CG} = 0.64$, $P_{PF\text{-}CG} = 0.07$). In relation to population demographics, the mismatch distribution of our data set resembled a bell curve with a mean of 27 mismatches and a tail representing a high frequency of pairs with only a few mismatches (Fig. S7). The observed distribution did not differ from the expectations of both the population size expansion and the spatial expansion model at the 5% significance threshold, but did provide much stronger support for the spatial expansion model (Harpending's raggedness index, $P = 0.09$ and $P = 0.89$ respectively). In contrast, no support was provided for a stable, non-expanding population ($P = 0.01$).

Recombination and selection

Borrelia garinii concatenated sequences considering all loci (chromosomal and plasmid genes) showed moderate levels of recombination ($\rho = 40$, Lk = –6112694.7, $P = 0$) (Pérez-Losada *et al*., 2006). This contrasts with the hypothesis of a strictly clonal model of population structure. However, different gene partitions showed variable rates of recombination; in this case recombination appeared to be most frequent in plasmid genes (*ospA* and *ospC*), but was also detected in one chromosomal gene, *recA* (Table 2). When only marine isolates were considered, removing terrestrial isolates, the detected recombination rate dropped considerably ($\rho = 12$, Lk = –4452055.9, $P < 0.001$), suggesting that genetic exchange occurs between the two enzootic cycles. In agreement, results by the RDP3 program further confirmed that significant intragenic recombination occurred at the *ospC* and *ospA* plasmid regions, and that terrestrial isolates were involved in the recombination events detected (Table S2).

Selection estimates on individual genes of *B. garinii* sequences are shown in Table 2. Only plasmid loci

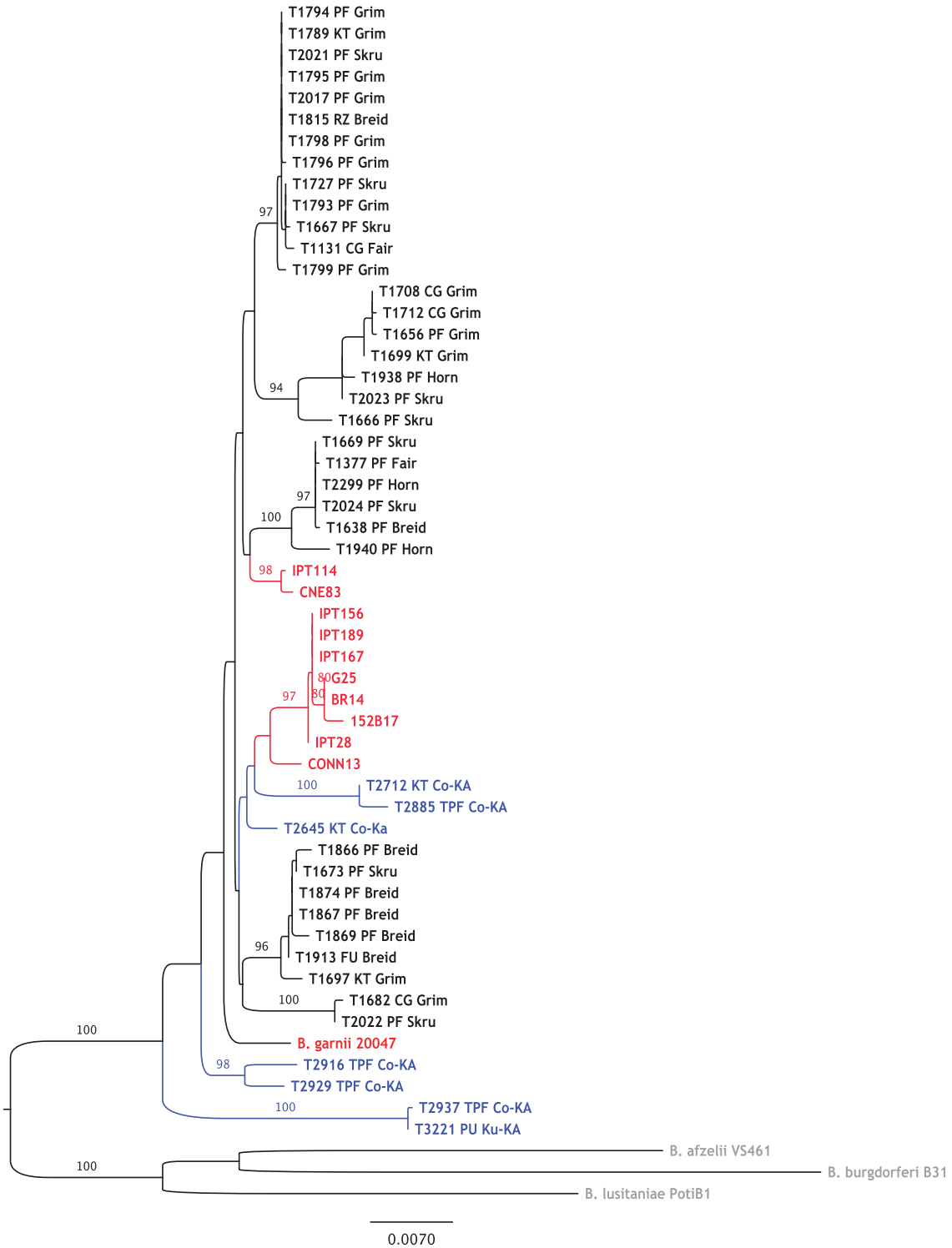

Fig. 2. Maximum likelihood (ML) phylogenetic tree of 53 *B. garinii* isolates and three *B. burgdorferi* s.l. reference strains, *B. afzelii* (VS461), *B. burgdorferi* s.s. (B31) and *B. lusitaniae* (PotiB1), based on concatenated sequences of eight chromosomal genes. The figured topology and branch lengths are for the sampled tree with the highest likelihood by RaxML (1000 searches, log likelihood = –8274.75). Labels at branch tips refer to strain number, seabird host and collection site (see Table 1). Terrestrial isolates are indicated in red and North Pacific and North Atlantic marine isolates are highlighted in blue and black respectively. The three non-*B. garinii* reference strains are in grey. Only bootstrap supports greater than 70% are shown.

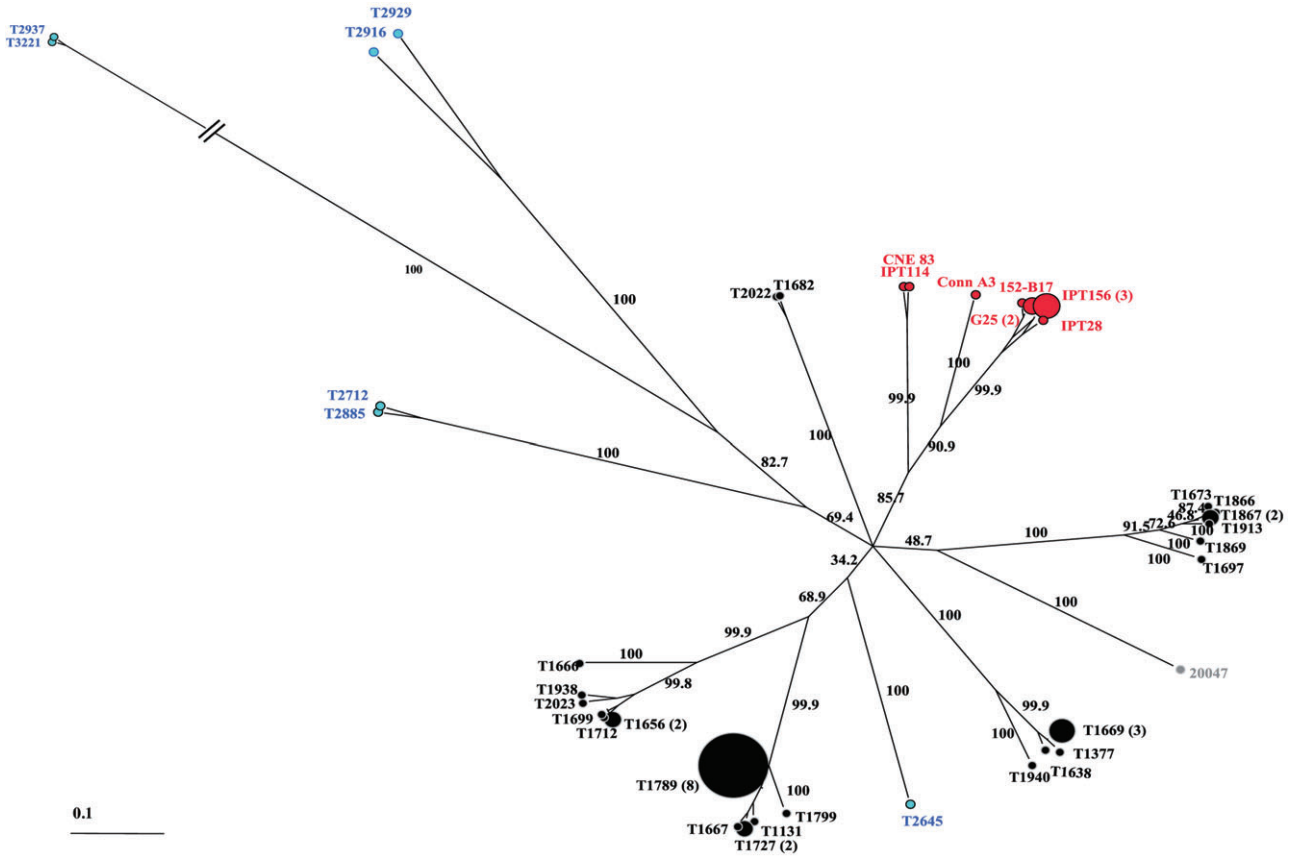

Fig. 3. Consensus network showing clonal relationships of *B. garinii* strains based on eight chromosomal genes. Red labels and branches highlight terrestrial isolates. Circle size is proportional to the number of isolates with identical allelic profiles and colours represent major geographic regions of marine strains (black: North Atlantic and blue: North Pacific).

showed significant departures from neutrality, and for both genes, models that took into account sites under positive selection fit the data significantly better than those that did not [with $P < 0.001$ for both M1a–M2a and M7–M8 likelihood ratio tests (LRTs)]. The Bayes Empirical Bayes (BEB) approach identified nine positive selection sites for *ospA* and 17 for *ospC* that were consistent using both M2a and M8 models (Table 2). Analyses by PAML on the marine data set alone showed similar results for both *ospA* and *ospC* (results not shown), thus indicating that positive selection operates on these genes and that the detected patterns are not simply due to different selection pressures between marine and terrestrial strains or to recombination within each subset (see results of recombination tests on marine strains).

Discussion

Despite the potential importance of population structure in pathogenic bacteria for their epidemiology, the spatial and ecological heterogeneity of these populations is often poorly characterized. Indeed, pathogen variability and its spatial arrangement can affect both transmission rates and the evolution of new variants. As such, this information is key to better understand disease dynamics and evaluate the risk posed to humans. Here, we investigated the genetic diversity and population structure of a LB bacterium *B. garinii* in its marine ecosystem, involving seabirds and the tick *I. uriae* using multilocus sequence analyses on isolates from diverse seabird tick races and geographic locations. These data are directly compared with terrestrial isolates to understand the potential role of this marine ecosystem in the evolution and global epidemiology of this pathogen. *Borrelia garinii* in the marine cycle showed high diversity with no population genetic structure within regions, but significant structure between Atlantic and Pacific regions. There was also evidence of significant structure between marine and terrestrial ecosystems, but the level of divergence measured was of the same order of magnitude as that between the two ocean basins, thus suggesting that the two enzootic cycles do not function independently (i.e. as much mixing occurs between ecosystems as within). Recombination analyses confirmed mixing between the two cycles, but no evi-

dence for divergent selection between the cycles was found at the plasmid loci examined, even though positive selection was evident at these genes. Overall, results suggest that the marine cycle may indeed play an important role in the circulation and evolution of new variants of *B. garinii*.

Diversity of B. garinii *in the marine ecosystem*

Although at least three species of the *B. burgdorferi* s.l. complex may occur in the marine cycle, the dominant species appears to be *B. garinii* (Olsen *et al*., 1993; Gylfe *et al*., 1999; Dietrich *et al*., 2008; Duneau *et al*., 2008). In the present study, we found high levels of genetic diversity in *B. garinii* and most alleles were unique. These results are concordant with previous studies in the marine ecosystem (Duneau *et al*., 2008; Comstedt *et al*., 2009) and with evidence from terrestrial systems showing that *B. garinii* is among the most heterogeneous species of LB bacteria, both genetically and antigenically (Gern and Humair, 2002). The diversity of *B. garinii* represented in the present study is likely an underestimate of the true diversity of this system. Indeed, the sequences analysed for the 10 genes here represent those bacteria that were present in high copy numbers in *I. uriae* ticks. *Borrelia* strains with low copy numbers are difficult to amplify using standard PCR procedures (Gómez-Díaz *et al*., 2010) and were therefore unlikely to be successfully amplified at all 10 genes. This suggests that the marine strains analysed in this study are those best adapted to exploit marine hosts and vectors, and those strains which may be more frequently present in the terrestrial system could be under-represented. As this type of bias is also known for cultured *Borrelia* isolates (Norris *et al*., 1997), the results of this study likely underestimate the degree of interaction between the two ecosystems.

As expected, heterogeneity was particularly marked in the case of the *ospA* and *ospC* loci; we found seven OspA and eight OspC genotype groups in the marine cycle of *B. garinii*. These distinct groups correspond to phenotypically distinguishable *Borrelia* strains (i.e. serotypes, Wang *et al*., 1999a), which have been associated with distinct clinical manifestations of LB and different ecotypes (Kurtenbach *et al*., 2002; Qiu, 2002). Previous studies showed that OspA serotypes 3 and 5–8 of *B. garinii* are maintained in terrestrial birds and seabirds (Kurtenbach *et al*., 2002; 2006; Hanincova *et al*., 2003). Accordingly, here the most abundant OspA genotype in the marine system matched the genetic profile of serotype 3, but the majority of genotype groups we found did not correspond to previously described serotypes and may be exclusive to this system. A similar pattern was found for *ospC* alleles; again in this case, only a few marine sequences matched previously described OspC serotype groups (i.e. OspC G, P and R), none of which appears specifically associated with birds. Of these, only OspC G appears to be pathogenic, causing local infection at the tick bite site, but not systemic disease (Seinost *et al*., 1999). The potential pathogenicity of the newly described OspC genotypes is unknown.

Population structure of marine B. garinii *spirochaetes*

Highly clonal bacterial populations typically display a limited number of clusters of closely related genotypes, and these clonal complexes are often globally distributed and stable over time (Feil and Enright, 2004). In contrast, the population structure of marine *B. garinii* is complex with several independent and highly divergent clonal groups and the distribution of these clones is spatially heterogeneous. Using chromosomal genes only, our results indicate marked differentiation and strong barriers to gene flow between Pacific and Atlantic *B. garinii* marine strains. Although data are still limited, this corresponds with some observed patterns of *Borrelia* infection in these regions, i.e. about 2% prevalence in the Pacific (E. Gómez-Díaz, pers. data) compared with 26% in the Atlantic (Duneau *et al*., 2008). In contrast, no significant population genetic structure appeared within regions, despite high levels of genetic diversity.

Vector-borne microbes necessarily co-occur with their vertebrate hosts and vectors and these hosts are expected to have a strong influence on pathogen population structure and infection dynamics (Woolhouse *et al*., 2002; McCoy, 2008). Population subdivision in pathogens may therefore arise as a consequence of geographic or ecological differences among hosts and/or vector populations that act as barriers to gene flow (e.g. multiple niche polymorphism; see review by Kurtenbach *et al*., 2006). In the marine system of LB, the tick vector *I. uriae* displays marked population substructure in the form of multiple sympatric seabird–host races (McCoy *et al*., 2001; Kempf *et al*., 2009) and previous studies have reported significant differences in infection prevalence and intensity among seabird tick races and geographic locations within the Atlantic (Duneau *et al*., 2008; Gómez-Díaz *et al*., 2010). However, despite these patterns in infection parameters, we did not find an imprint of host-associated divergence in marine *Borrelia* spirochaetes and isolates from different tick races often shared the same bacterial strain. Our results are therefore concordant with recent work in the terrestrial cycle showing uncorrelated patterns of population structure between *B. burgdorferi* and its tick vector *Ixodes scapularis* in North America (Humphrey *et al*., 2010). Variation among tick races in vector competence or interspecific variation in seabird immune responses may affect bacterial fitness and thus explain differences *Borrelia* infection prevalence. However, it

seems that even if certain seabird species-tick race combinations are more competent than others (Gómez-Díaz *et al*., 2010), there is not enough isolation among these different host–vector systems for bacterial population divergence to have occurred.

Selection can significantly alter historical divergence and gene flow patterns among pathogen populations (Gupta and Anderson, 1999). In the marine system, seabird host colonies are often multispecific and the high genetic diversity of marine *B. garinii* at a local scale could be the consequence of strong balancing selection on local *Borrelia* clones (Qiu, 2002) that maintains ancestral variation and masks spatial population structure. Both *ospA* and *ospC* genes code for proteins that are primary targets for the host immune system and, as expected, we found evidence for positive selection on these genes. Recombination is known to affect the accuracy of selection analyses and can be mistaken as evidence of positive selection (Anisimova *et al*., 2003). We feel that this is not the case in our study; certain tests of selection that we used here are less affected by recombination than others (i.e. BEB) and still showed clear evidence of positive selection. In addition, signs of positive selection were similar when considering only marine isolates, for which recombination was low. Despite positive selection at the plasmid loci, there was no evidence of divergence patterns in these loci that could be associated with different seabird hosts, tick races or geographic clade structure (Figs S3 and S4), suggesting that selection at these loci may be linked to within-population variation in susceptibility among individual birds (Dykhuizen and Brisson, 2010). To better understand how host-associated selection may shape the population structure and evolution of these pathogens, further analyses examining the nature of the sites under positive selection are called for, along with more detailed information on temporal changes in the frequency of *Borrelia* variants within seabird colonies and on strict associations between host genotypes/phenotypes and *Borrelia* infection.

Like selection, contemporary gene flow can reduce signs of clonal population structure and historical patterns of population subdivision by homogenizing alleles at different spatial scales; the importance of this force has been shown for a number of bacterial pathogens (i.e. *Campylobacter jejuni*, Sheppard *et al*., 2008). In the present study, clear spatial subdivision was found between Atlantic and Pacific *B. garinii* strains, which appear to function relatively independently. This corresponds to patterns of genetic variation in both seabirds and their ticks and confirms limited movements between Atlantic and Pacific Ocean basins (McCoy *et al*., 2005; Friesen *et al*., 2007; Morris-Pocock *et al*., 2008). However, we could not reject gene flow within these two regions. Bacterial gene flow by the tick is unlikely given that *I. uriae* displays marked geographic subdivision among seabird colonies (McCoy *et al*., 2002; 2005). Even if seabirds are highly vagile and can disperse ticks at large spatial scales, gene flow in the tick is limited in both space and time by different habitat and host-related factors (i.e. host behaviour during the breeding season combined with seasonality in tick infestation; see McCoy *et al*., 2003). Therefore, genetic exchange of *Borrelia* spirochaetes is likely linked to seabird movement. Indeed, previous studies have proposed long-range dispersal by seabirds to explain the presence of identical *B. garinii* gene sequences in ticks from both hemispheres (Olsen *et al*., 1995; Gylfe *et al*., 1999) and recent work in the terrestrial cycle advocates host-mediated dispersal as an explanation for population admixture in *B. burgdorferi* s.l. (Humphrey *et al*., 2010; Vollmer *et al*., 2011). The high admixture and allelic diversity in the Atlantic is consistent with recent population growth and rapid geographic range expansion of the bacterium in the marine system (see mismatch distributions, Fig. S7). This fits with current knowledge about the relatively recent colonization of many seabird colonies of the North Atlantic since the last glacial maxima (max age of < 10 000 years), and patterns of among-colony gene flow in the birds (Riffaut *et al*., 2005; Friesen *et al*., 2007).

Links between marine and terrestrial systems? Implications for the global epidemiology of LB

Several lines of evidence support the existence of admixture between marine and terrestrial strains of *B. garinii*. First, paraphyly is present in marine isolates; certain marine strains are more closely related to terrestrial strains than to other marine strains (Figs 2 and 3). Likewise, estimates of population differentiation were as high between Atlantic and Pacific *B. garinii* isolates as they were between isolates of each marine region and those from the terrestrial ecosystem. Tests of recombination further showed that some exchange between terrestrial and marine variants likely occurs and single-loci comparative analyses with terrestrial allele sequences in public databases highlight the similarity between marine and terrestrial strains (Figs S1, S2, S5 and S6). More surprisingly, the reference strain *20047* clustered with marine isolates at both chromosomal and plasmid loci and showed an admixed ancestry of marine and terrestrial alleles (results not shown). The existence of contact zones in certain seabird colonies, combined with the relatively generalist nature of vectors like *Ixodes ricinus* could provide opportunities for *Borrelia* transmission between these two enzootic systems (Dietrich *et al*., 2011). Indeed, the origin of the *20047* strain was an *I. ricinus* tick sampled in Brittany, France (Postic *et al*., 1990), where seabird colonies occur and where the distributions of *I.*

ricinus and *I. uriae* overlap. Our findings also agree with recent work involving Pacific isolates that demonstrated the existence of shared genetic IGS variants of *B. garinii* in these two tick species (Comstedt *et al*., 2009). The possibility that *I. uriae* may also attack vertebrates other than seabirds, including passerines, small mammals and humans, has been previously suggested, but may not be frequent (Bergström *et al*., 1998). Rather than database reference comparisons of isolates from multiple sources and geographic locations, more work dedicated to typing *Borrelia* within potential contact areas is now called for to help clarify where and how frequently contact between the two cycles occurs.

Although we currently have limited knowledge about the spatial and temporal dynamics of the marine *B. garinii* populations, our results suggest that the inclusion of data from the marine system may alter our current perception of LB epidemiology. Indeed, the wide geographic distribution of seabirds, their great movement capacity and large population sizes may significantly contribute to the epizootic spread of LB spirochaetes. We therefore advocate that future studies of LB epidemiology take the marine system into account more explicitly and that more research effort focuses on understanding the population structure of this system and the nature of potential contact zones between the two enzootic cycles.

Experimental procedures

Study sites, tick sampling and Borrelia *isolates*

Ixodes uriae ticks were collected from seabird species breeding in several geographic locations in the North Atlantic [Iceland (Skrudur, Grimsey and Breidafjordur), Norway (Hornøya) and Britain (Fair Isle)], and the North Pacific [Commander Islands (Toporkov, Ariy Kamen and Bering), Kuril Islands (Ptichi) and Starichkov Island, off the peninsula of Kamchatka] (Fig. 1). DNA extractions were performed on individual *I. uriae* ticks using a DNeasy Tissue Kit (Qiagen, Valencia, CA). *Borrelia* detection was carried out by a standard nested PCR amplification of the flagellin B (*flaB*) gene as described previously (Duneau *et al*., 2008).

Of more than 2000 ticks tested from multiple populations distributed worldwide (i.e. Antarctica, Pacific and Atlantic Oceans), 199 ticks were found positive, corresponding to the populations mentioned above. Among those, the majority corresponded to *B. garinii* (about 80%), although atypical genospecies for the marine ecosystem such as *Borrelia afzelii* or *B. burgdorferi* s.s. were also present (Dietrich *et al*., 2008; Duneau *et al*., 2008). We limited our analysis here to *B. garinii*. In addition to marine isolates, we used 11 cultured *B. garinii* strains from the terrestrial enzootic cycle isolated from *I. ricinus* ticks from several geographic regions (Table 1, Fig. 1). We also typed cultured reference strains of three other genospecies within the *B. burgdorferi* s.l. complex including: *B. afzelii* (VS461), *B. burgdorferi sensu stricto* (B31) and *B. lusitaniae* (PotiB1) (see *Phylogenetic analyses*).

Multilocus sequence analysis

Borrelia isolates were amplified and sequenced at eight chromosomal loci (*16S, flaB, recA, groEL, hbb, 5S–23S IGS, pyrG, recG*) and two plasmid loci (*ospA* and *ospC*). These genes have all been previously used in multilocus studies on the *B. burgdorferi* s.l. species complex (Bunikis *et al*., 2004; Richter *et al*., 2006; Margos *et al*., 2008). Amplification conditions and primer sequences used in this study are listed in Table S3. Chromosomal genes, which consist mostly of housekeeping genes, tend to be under strong selective constraints and often fit neutral expectations. They are thus frequently use to infer phylogenetic and ancestral relationships among bacterial species (Pérez-Losada *et al*., 2006). In contrast, plasmid genes (here *ospA* and *ospC*) are loci that can be subject to strong amounts of recombination and selection. As these genes are often more variable, they can be more appropriate for revealing adaptive population genetic structure in clonal bacterial species, such as *B. burgdorferi* (Didelot and Falush, 2007; Gómez-Díaz, 2009).

All PCR reactions were run with both positive (DNA from cultured *B. garinii*-20047) and negative (distilled water) controls. Amplification products were sent for direct sequencing (Cogenics, France). Chromatographs were checked manually and sequences were aligned using MAFFT v6 (http://align.bmr.kyushu-u.ac.jp/mafft/online/server/) (Katoh *et al*., 2009). We applied Gblocks to eliminate poorly aligned positions of individual gene sequences (http://molevol.cmima.csic.es/castresana/Gblocks_server.html) (Castresana, 2000), but in no case was more than 5% of the sequence length eliminated, mainly gaps and missing data. The final length of the concatenated *B. garinii* sequence was 3255 bp for the eight chromosomal, and 517 bp and 619 bp for the *ospC* and *ospA* plasmid genes.

Not all isolates amplified equally well across the different loci, likely due to the relative quantity of bacterial DNA compared with tick DNA (Gómez-Díaz *et al*., 2010), and multiple infections were sometimes detected (data not shown). After removing incomplete and mixed sequences, a subset of 53 *B. garinii* isolates (including both marine and terrestrial *B. garinii* isolates) was subsequently used for multilocus sequence analyses (Table 1). The sequences reported in this article are available from the GenBank database (Accession No. JF330910–JF331429). Individual allele sequences for some of the genes analysed in this study can also be found online on the MLSA database http://pubmlst.org/bburgdorferi (Table S1).

Sequences of individual genes that differed by one or more nucleotides were assigned distinct allele numbers using DNAsp v5.1 (Rozas *et al*., 2003), with the option of excluding sites with alignment gaps and/or missing data. Individual sequence types (STs) are defined as each unique combination of alleles. *ospC* and *ospA* sequences were assigned to previously defined serotype groups based on neighbour-joining distance analyses using BioNJ (http://www.phylogeny.fr/) on sequences from a selection of strains representing OspA 1–8 and OspC A–U serotype groups, available from GenBank (Wilske *et al*., 1993; Seinost *et al*., 1999; Wang *et al*., 1999b). We calculated basic genetic statistics for all genes using DNAsp v5.1 (Rozas *et al*., 2003).

Phylogenetic analyses

Both *ospA* and *ospC* plasmid loci showed evidence of intragenic recombination and positive selection (see *Results*). We thus examined phylogenetic relationships based on the concatenated sequences of the eight chromosomal genes and in each of the two plasmid loci (*ospA* and *ospC*), separately. Concatenated and individual *ospA* and *ospC* sequences of reference strains of *B. afzelii* (VS461), *B. burgdorferi sensu stricto* (B31) and *B. lusitaniae* (PotiB1) were also included in the analyses as outgroups. Phylogenetic analyses were performed under ML using RAxML GUI (Silvestro and Michalak, 2010), a graphical front–end for RAxML-VI-HPC (Randomized Axelerated Maximum Likelihood; Stamatakis, 2006). In the case of the concatenated data set, we performed a partitioning scheme using eight distinct models/data partitions with individual per partition branch length optimization in order to deal with possible phylogenetic incongruence among genes. ML, with the thorough bootstrap option, was run from random seeds to generate 1000 non-parametric bootstrap replicates. All ML analyses used the general time-reversible (GTR) model, with a gamma model of rate heterogeneity. Individual α-shape parameters, GTR rates and base frequencies were estimated and optimized for each partition.

Comparative single-locus network analyses of *ospA*, *ospC*, *recG* and *pyrG* allele sequences, available from GenBank and the MLST *B. burgdorferi* database (http://borrelia.mlst.net/), were performed using the program SplitsTree v4.10 and the NeighborNet protocol (http://www.splitstree.org/) (Huson, 1998).

Clonality and population structure

To visualize the extent of clonal reproduction and the effect of recombination on our tree-based interpretation of clade structure in the marine cycle of *B. garinii*, we applied Clonalframe, a sequence-based approach that uses concatenated sequences to infer clonal relationships among isolates (Didelot and Falush, 2007). Compared with traditional phylogenetic methods, ClonalFrame analyses do not force relationships to be tree-like, i.e. enables a network analysis, and can incorporate recombination into the evolutionary reconstruction. However, as both *ospA* and *ospC* showed strong evidence of positive selection (see *Results*), these sequences were excluded from the analysis. The program was run using a burn-in of 100 000 iterations with 200 000 subsequent iterations. The resulting consensus network was drawn using SplitsTree v4 (Huson, 1998).

Using chromosomal loci only, genetic structure between marine and terrestrial subsets and between Atlantic and Pacific marine regions, was calculated using an estimator of Wright's F_{ST} statistic (Φ) in the program Arlequin v3 (Excoffier, 2005). Within the North Atlantic, we also tested for genetic differentiation among geographic localities and host-associated tick groups. Within this region, where marine *B. garinii* population parameters were found to fit panmictic expectations (see *Results*), we also used Arlequin to estimate the mismatch frequency distribution and to model the expected distributions under three different demographic scenarios: (i) stable population size, (ii) population demographic expansion and (iii) spatial expansion. Model fit was examined by comparing the Harpending's raggedness statistic of the observed data relative to the models (Schneider and Excoffier, 1999).

Recombination and selection

The composite likelihood coalescent method implemented in LDhat (McVean *et al.*, 2002) was applied to estimate rates of recombination (ρ) for individual gene partitions and for the concatenated loci data set. Using this program, we also performed a non-parametric permutation test (LPT) to test the hypothesis of no recombination ($\rho = 0$). In addition, we used several recombination detection methods implemented in RDP3 (Martin *et al.*, 2005) to visualize hotspots of recombination along the chromosome and in each plasmid loci separately in order to identify the individual sequences involved in major recombination events. The program was run using default parameters and only recombination events detected by at least five methods were accepted.

We applied the ML phylogenetic framework implemented in PAML 4.2b (Yang, 2007) to test for positive selection by comparing the fixation rates of non-synonymous and synonymous substitutions ($\omega = dN/dS$). In this framework, $\omega = 1$ suggests neutral evolution, $\omega < 1$ purifying selection and $\omega > 1$ diversifying positive selection. We estimated ω per site using the codon-based models M1a (nearly neutral) and the alternative model M2a (positive selection) (Nielsen and Yang, 1998). We then compared the null model M7 (where ω follows a beta distribution) and the alternative model M8 (which adds sites with $\omega > 1$). The two model pairs differ in the number of parameters considered (i.e. M1a and M2a are regarded as simpler versions of M7 and M8). We then applied the BEB analysis implemented in PAML (Yang *et al.*, 2005) to identify the potential sites under selection, indicated by a posterior probability (pP) > 0.95. Finally, LRTs (2df) were used to determine which model (i.e. neutral evolution versus selection) best fit our data set using MODELTEST 3.7 (Posada and Crandall, 1998).

Acknowledgements

We thank G. Baranton, R.T. Barrett, M. Dietrich, D. Duneau, M. Garnier, L. Gern, J. González-Solís, A. Petersen, A. Ploshnitsa, D. Postic, D. Shaw and T. Tverra for help and advice at different stages of this work. We also thank members of the Working group 'Tique et Maladies à Tiques' (REID, France) and Dog Cancer Club for stimulating discussions. Dr Nikolay N. Pavlov, Director of Komandorskiy State Nature Biosphere Reserve, Dr Yuri Artukhin and Dr S. Mironov are gratefully acknowledged for their professional advice and logistical support. We also thank NINA, the University of Tromsø – Tromsø Museum (Norway) and Fair Bird Observatory (Scotland). Three anonymous reviewers provided helpful comments on an earlier version of the manuscript. Financial support was provided by the 'Institut Polaire-Paul Emile Victor' (IPEV, programme No. 333), the Centre National de la Recherche Scientifique (CNRS – Emerging Infectious Disease Interdisciplinary Program) and the Agence National de la Recherche (ANR-06-JCJC-0095-01). E.G.-D. was supported by a Marie Curie fellowship No. PIEF-GA-2008-221243.

References

Anisimova, M., Nielsen, R., and Yang, Z. (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* **164:** 1229–1236.

Bergström, S., Haemig, P.D., and Olsen, B. (1998) Distribution and abundance of the tick *Ixodes uriae* in a diverse subantarctic seabird community. *J Parasitol* **85:** 25–27.

Bunikis, J., Garpmo, U., Tsao, J., Berglund, J., Fish, D., and Barbour, A.G. (2004) Sequence typing reveals extensive strain diversity of the Lyme borreliosis agents *Borrelia burgdorferi* in North America and *Borrelia afzelii* in Europe. *Microbiology* **150:** 1741–1755.

Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17:** 540–552.

Comstedt, P., Asokliene, L., Eliasson, I., Olsen, B., Wallensten, A., Bunikis, J., and Bergström, S. (2009) Complex population structure of Lyme borreliosis group spirochete *Borrelia garinii* in subarctic Eurasia. *PLoS ONE* **4:** e5841.

Didelot, X., and Falush, D. (2007) Inference of bacterial microevolution using multilocus sequence data. *Genetics* **175:** 1251–1266.

Dietrich, M., Gómez-Díaz, E., Boulinier, T., and McCoy, K.D. (2008) Local distribution and genetic structure of tick-borne pathogens: an example involving the marine cycle of Lyme disease. In *Integrative Acarology, Proceedings of the 6th European Congress of the European Association of Acarologists*. Bertrand, M., Kreiter, S., McCoy, K.D., Migeon, A., Navajas, M., Tixier, M.-S., and Vial, L. (eds). Montpellier, France: European Association of Acarologists, pp. 33–42.

Dietrich, M., Gómez-Díaz, E., and McCoy, K.D. (2011) Worldwide distribution and diversity of seabird ticks: implications for the ecology and epidemiology of tick-borne pathogens. *Vector Borne Zoonotic Dis* **11:** 453–470.

Duneau, D., Boulinier, T., Gómez-Díaz, E., Petersen, A., Tveraa, T., Barrett, R.T., and McCoy, K.D. (2008) Prevalence and diversity of Lyme borreliosis bacteria in marine birds. *Infect Genet Evol* **8:** 352–359.

Dykhuizen, D.E., and Baranton, G. (2001) The implications of a low rate of horizontal transfer in *Borrelia*. *Trends Microbiol* **9:** 344–350.

Dykhuizen, D.E., and Brisson, D. (2010) Evolutionary genetics of *Borrelia burgdorferi sensu lato*. In *Borrelia: Molecular Biology, Host Interaction and Pathogenesis*. Samuels, S.D., and Radolf, J. (eds). Norfolk, UK: Caister Academic Press, pp. 221–249.

Dykhuizen, D.E., Polin, D.S., Dunn, J.J., Wilske, B., Preac-Mursic, V., Dattwyler, R.J., and Luft, B.J. (1993) *Borrelia burgdorferi* is clonal: implications for taxonomy and vaccine development. *Proc Natl Acad Sci USA* **90:** 10163–10167.

Excoffier, L. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* **1:** 47–50.

Feil, E.J., and Enright, M.C. (2004) Analyses of clonality and the evolution of bacterial pathogens. *Curr Opin Microbiol* **7:** 308–313.

Feil, E.J., and Spratt, B.G. (2001) Recombination and the population structures of bacterial pathogens. *Annu Rev Microbiol* **55:** 561–590.

Feil, E.J., Holmes, E.C., Bessen, D.E., Chan, M.-S., Day, N.P.J., Enright, M.C., *et al*. (2001) Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc Natl Acad Sci USA* **98:** 182–187.

Friesen, V.L., Burg, T.M., and McCoy, K.D. (2007) Mechanisms of population differentiation in seabirds. *Mol Ecol* **16:** 1765–1785.

Gern, L., and Humair, P.F. (2002) Biology of the infectious agents and epidemiology of disease. In *Lyme Borreliosis*. Gray, J.S., Kahl, O., Lane, R.S., and Stanek, G. (eds). Wallingford, Oxon, UK: CABI Publishing, pp. 149–174.

Gómez-Díaz, E. (2009) Linking questions to practices in the study of microbial pathogens: sampling bias and typing methods. *Infect Genet Evol* **9:** 1418–1423.

Gómez-Díaz, E., Doherty, P.F., Duneau, D., and McCoy, K.D. (2010) Cryptic population structure can mask vector-specific patterns of infection: an example from the marine cycle of Lyme borreliosis. *Evol Appl* **3:** 391–401.

Gray, J.S., Kahl, O., Lane, R.S., and Stanek, G. (2002) *Lyme Borreliosis: Biology, Epidemiology and Control*. New York, USA: CABI Publishing.

Gupta, S., and Anderson, R.M. (1999) Population structure of pathogens: the role of immune selection. *Parasitol Today* **15:** 497–501.

Gylfe, A., Olsen, B., Strasevicius, D., Marti Ras, N., Weihe, P., Noppa, L., *et al*. (1999) Isolation of Lyme disease *Borrelia* from puffins (*Fratercula arctica*) and seabird ticks (*Ixodes uriae*) on the Faeroe Islands. *J Clin Microbiol* **37:** 890–896.

Halkett, F., Simon, J.-C., and Balloux, F. (2005) Tackling the population genetics of clonal and partially clonal organisms. *Trends Ecol Evol* **20:** 194–201.

Hanincova, K., Taragelova, V., Koci, J., Schafer, S.M., Hails, R., Ullmann, A.J., *et al*. (2003) Association of *Borrelia garinii* and *B. valaisiana* with songbirds in Slovakia. *Appl Environ Microbiol* **69:** 2825–2830.

Humphrey, P.T., Caporale, D.A., and Brisson, D. (2010) Uncoordinated phylogeography of *Borrelia burgdorferi* and its tick vector, *Ixodes scapularis*. *Evolution* **64:** 2653–2663.

Huson, D.H. (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14:** 68–73.

Katoh, K., Asimenos, G., and Toh, H. (2009) Multiple alignment of DNA sequences with MAFFT. *Methods Mol Biol* **537:** 39–64.

Kempf, F., Boulinier, T., De Meeûs, T., Arnathau, C., and McCoy, K.D. (2009) Recent evolution of host-associated specialization in the seabird tick *Ixodes uriae*. *Mol Ecol* **18:** 4450–4462.

Kurtenbach, K., DeMichelis, S., Etti, S., Schäfer, S.M., Sewell, H.-S., Brade, V., and Kraiczy, P. (2002) Host association of *Borrelia burgdorferi sensu lato* – the key role of host complement. *Trends Microbiol* **10:** 74–79.

Kurtenbach, K., Hanincova, K., Tsao, J.I., Margos, G., Fish, D., and Ogden, N.H. (2006) Fundamental processes in the evolutionary ecology of Lyme borreliosis. *Nat Rev Microbiol* **4:** 660–669.

McCoy, K.D. (2008) The population genetic structure of vectors and our understanding of disease epidemiology. *Parasite* **15:** 444–448.

McCoy, K.D., Boulinier, T., Tirard, C., and Michalakis, Y. (2001) Host specificity of a generalist parasite: genetic evidence of sympatric host races in the seabird tick *Ixodes uriae*. *J Evol Biol* **14:** 395–405.

McCoy, K.D., Boulinier, T., Schjørring, S., and Michalakis, Y. (2002) Local adaptation of the ectoparasite *Ixodes uriae* to its seabird host. *Evol Ecol* **4:** 441–456.

McCoy, K.D., Boulinier, T., Tirard, C., and Michalakis, Y. (2003) Host-dependent genetic structure of parasite populations: differential dispersal of seabird tick host races. *Evolution* **57:** 288–296.

McCoy, K.D., Boulinier, T., and Tirard, C. (2005) Comparative host–parasite population structures: disentangling prospecting and dispersal in the black-legged kittiwake *Rissa tridactyla*. *Mol Ecol* **14:** 2825–2838.

McVean, G., Awadalla, P., and Fearnhead, P. (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160:** 1231–1241.

Margos, G., Gatewood, A.G., Aanensen, D.M., Hanincova, K., Terekhova, D., Vollmer, S.A., *et al*. (2008) MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. *Proc Natl Acad Sci USA* **105:** 8730–8735.

Martin, D.P., Williamson, C., and Posada, D. (2005) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21:** 260–262.

Morris-Pocock, J.A., Taylor, S.A., Birt, T.P., Damus, M., Piatt, J.F., Warheit, K.I., and Friesen, V.L. (2008) Population genetic structure in Atlantic and Pacific Ocean common murres (*Uria aalge*): natural replicate tests of post-Pleistocene evolution. *Mol Ecol* **17:** 4859–4873.

Nielsen, R., and Yang, Z. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148:** 929–936.

Norris, D.E., Johnson, B.J., Piesman, J., Maupin, G.O., Clark, J.L., and Black, W.C., IV (1997) Culturing selects for specific genotypes of *Borrelia burgdorferi* in an enzootic cycle in Colorado. *J Clin Microbiol* **35:** 2359–2364.

Olsen, B. (2003) The role of birds in the ecology and epidemiology of Lyme borreliosis and ehrlichiosis. *Infect Dis* **2:** 4–7.

Olsen, B., Jaenson, T.G.T., Noppa, L., Bunikis, J., and Bergström, S. (1993) A Lyme borreliosis cycle in seabirds and *Ixodes uriae* ticks. *Nature* **362:** 340–342.

Olsen, B., Duffy, D.C., Jaenson, T.G.T., Gylfe, A., Bonnedahl, J., and Bergström, S. (1995) Transhemispheric exchange of Lyme disease spirochetes by seabirds. *J Clin Microbiol* **33:** 3270–3274.

Pérez-Losada, M., Browne, E.B., Madsen, A., Wirth, T., Viscidi, R.P., and Crandall, K.A. (2006) Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. *Infect Genet Evol* **6:** 97–112.

Posada, D., and Crandall, K.A. (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14:** 817–818.

Postic, D., Edlinger, C., Richaud, C., Grimont, F., Dufresne, Y., Perolat, P., *et al*. (1990) Two genomic species in *Borrelia burgdorferi*. *Res Microbiol* **141:** 465–475.

Qiu, W.-G. (2002) Geographic uniformity of the Lyme disease spirochete (*Borrelia burgdorferi*) and its shared history with tick vector (*Ixodes scapularis*) in the northeastern united states. *Genetics* **160:** 833–849.

Qiu, W.-G., Schutzer, S.E., Bruno, J.F., Attie, O., Xu, Y., Dunn, J.J., *et al*. (2004) Genetic exchange and plasmid transfers in *Borrelia burgdorferi sensu stricto* revealed by three-way genome comparisons and multilocus sequence typing. *Proc Natl Acad Sci USA* **101:** 14150–14155.

Richter, D., Postic, D., Sertour, N., Livey, I., Matuschka, F.R., and Baranton, G. (2006) Delineation of *Borrelia burgdorferi* sensu lato species by multilocus sequence analysis and confirmation of the delineation of *Borrelia spielmanii* sp nov. *Int J Syst Evol Microbiol* **56:** 873–881.

Riffaut, L., McCoy, K.D., Tirard, C., Friesen, V.L., and Boulinier, T. (2005) Population genetics of the common guillemot *Uria aalge* in the North Atlantic: geographic impact of oil spills. *Mar Ecol Prog Ser* **291:** 263–273.

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X., and Rozas, R. (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19:** 2496–2497.

Schneider, S., and Excoffier, L. (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* **152:** 1079–1089.

Seinost, G., Dykhuizen, D.E., Dattwyler, R.J., Golde, W.T., Dunn, J.J., Wang, I.-N., *et al*. (1999) Four clones of *Borrelia burgdorferi sensu stricto* cause invasive infection in humans. *Infect Immun* **67:** 3518–3524.

Sheppard, S.K., McCarthy, N.D., Falush, D., and Maiden, M.C.J. (2008) Convergence of *Campylobacter* species: implications for bacterial evolution. *Science* **320:** 237–239.

Silvestro, D., and Michalak, I. (2010) *RAxML GUI: a graphical front–end for RAML* [WWW document]. URL http://sourceforge.net/projects/raxmlgui/.

Spratt, B.G., and Maiden, M.C.J. (1999) Bacterial population genetics, evolution and epidemiology. *Philos Trans R Soc Lond B Biol Sci* **354:** 701–710.

Spratt, B.G., Hanage, W.P., and Feil, E.J. (2001) The relative contributions of recombination and point mutation to the diversification of bacterial clones. *Curr Opin Microbiol* **4:** 602–606.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690.

Stanek, G., and Strle, F. (2003) Lyme borreliosis. *Lancet* **362:** 1639–1647.

Staszewski, V., McCoy, K.D., and Boulinier, T. (2008) Variable exposure and immunological response to Lyme disease *Borrelia* among North Atlantic seabird species. *Proc R Soc Lond B Biol Sci* **275:** 2101–2109.

Stevenson, B. (2001) *Borrelia burgdorferi*: a (somewhat) clonal bacterial species. *Trends Microbiol* **9:** 471–472.

Travinsky, B., Bunikis, J., and Barbour, A.G. (2010) Geographic differences in genetic locus linkages for *Borrelia burgdorferi*. *Emerg Infect Dis* **16:** 1147–1150.

Vollmer, S.A., Bormane, A., Dinnis, R.E., Seelig, F., Dobson, A.D.M., Aanensen, D.M., *et al*. (2011) Host migration impacts on the phylogeography of Lyme borreliosis

spirochaete species in Europe. *Environ Microbiol* **13:** 184–192.
Wang, G., Van Dam, A.P., Schwartz, I., and Dankert, J. (1999a) Molecular typing of *Borrelia burgdorferi* sensu lato: taxonomic, epidemiological, and clinical implications. *Clin Microbiol Rev* **12:** 633–653.
Wang, I.-N., Dykhuizen, D.E., Qui, W., Dunn, J.J., Bosler, E.M., and Luft, B.J. (1999b) Genetic diversity of *ospC* in a local population of *Borrelia burgdorferi* sensu stricto. *Genetics* **151:** 15–30.
Wilske, B., Preac-Mursic, V., Gobel, U.B., Graf, B., Jauris, S., Soutschek, E., *et al*. (1993) An OspA serotyping system for *Borrelia burgdorferi* based on reactivity with monoclonal antibodies and OspA sequence analysis. *J Clin Microbiol* **31:** 340–350.
Wilske, B., Busch, U., Fingerle, V., Jauris-Heipke, S., Preac Mursic, V., Rössler, D., and Will, G. (1996) Immunological and molecular variability of OspA and OspC. Implications for *Borrelia* vaccine development. *Infection* **24:** 208–212.
Woolhouse, M.E.J., Webster, J.P., Domingo, E., Charlesworth, B., and Levin, B.R. (2002) Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat Genet* **32:** 569–577.
Woolhouse, M.E.J., Haydon, D.T., and Antia, R. (2005) Emerging pathogens: the epidemiology and evolution of species jumps. *Trends Ecol Evol* **20:** 238–244.
Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24:** 1586–1591.
Yang, Z., Wong, W.S.W., and Nielsen, R. (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* **22:** 1107–1118.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Unrooted distance network on *recG* allele sequences of *Borrelia garinii* strains analysed in this study together with *recG* alleles available at the MLST *B. burgdorferi* database (http://www.mlst.net/). Identical sequences for marine and terrestrial isolates in this study have been collapsed into unique alleles. Major nodes with bootstrap support > 70% are shown. The scale bar indicates the distance. Pacific marine (PAC), Atlantic marine (ATL) and terrestrial isolates are indicated in blue, black bold and red respectively. Reference strains from the MLST database are in grey. For terrestrial strains, two-digit country codes indicate the origin and are those supplied by the ISO (International Organization for Standardization, http://www.iso.org). For more information on terrestrial strains visit the MLST database. The network was built using SplitsTree (see text).
Fig. S2. Unrooted distance network on *pyrG* allele sequences of *Borrelia garinii* strains analysed in this study together with *pyrG* alleles available at the MLST *B. burgdorferi* database (http://www.mlst.net/). Identical sequences for marine and terrestrial isolates in this study have been collapsed into unique alleles. Major nodes with bootstrap support > 70% are shown. The scale bar indicates the distance. Pacific marine (PAC), Atlantic marine (ATL) and terrestrial isolates are indicated in blue, black bold and red respectively. Reference strains from the MLST database are in grey. For terrestrial strains, two-digit country codes indicate the origin and are those supplied by the ISO (International Organization for Standardization, http://www.iso.org). For more information on terrestrial strains visit the MLST database. The network was built using SplitsTree (see text).
Fig. S3. Maximum likelihood phylogenetic tree of *ospA* sequences of 53 *Borrelia garinii* strains analysed in this study. Labels at branch tips refer to strain number, seabird host and collection site. Terrestrial isolates are indicated in red and North Pacific marine isolates are highlighted in blue. Cultured non-*B. garinii* references strains are in grey. Only bootstrap supports greater than 70% are shown. Major OspA serotype groups for which there was a significant match are indicated (see text). Information on branch tip labels can be found in Table 1.
Fig. S4. Maximum likelihood phylogenetic tree of *ospC* sequences of 53 *Borrelia garinii* strains analysed in this study. Labels at branch tips refer to strain number, seabird host and collection site. Terrestrial isolates are indicated in red and North Pacific marine isolates are highlighted in blue. Only bootstrap supports greater than 70% are shown. Major OspC serotype groups for which there was a significant match are indicated (see text). Information on branch tip labels can be found in Table 1.
Fig. S5. Unrooted distance network of *ospA* allele sequences of *Borrelia garinii* strains analysed in this study together with *ospA* reference sequences available from GenBank (see Table S4). Identical sequences for marine and terrestrial isolates in this study have been collapsed into unique alleles. Major nodes with bootstrap support > 70% are shown. The scale bar indicates the distance. Pacific marine (PAC), Atlantic marine (ATL) and terrestrial isolates are indicated in blue, black bold and red respectively. Reference strains from the MLST database are in grey. For terrestrial strains, two-digit country codes indicate the origin and are those supplied by the ISO (International Organization for Standardization, http://www.iso.org). The network was built using SplitsTree (see text).
Fig. S6. Unrooted distance network of *ospC* allele sequences of *Borrelia garinii* strains analysed in this study together with *ospC* reference sequences available from GenBank (see Table S5). Identical sequences for marine and terrestrial isolates in this study have been collapsed into unique alleles. Major nodes with bootstrap support > 70% are shown. The scale bar indicates the distance. Pacific marine (PAC), Atlantic marine (ATL) and terrestrial isolates are indicated in blue, black bold and red respectively. Reference strains from the MLST database are in grey. For terrestrial strains, two-digit country codes indicate the origin and are those supplied by the ISO (International Organization for Standardization, http://www.iso.org). The network was built using SplitsTree (see text).
Fig. S7. Mismatch frequency distribution of observed pairwise nucleotide differences in concatenated multilocus sequences (eight loci) of the North Atlantic *B. garinii* population (grey bars) and mismatch distributions modelled under three different demographic scenarios (see legend).
Table S1. Allelic profiles of 53 *Borrelia garinii* isolates analysed in this study based on eight chromosomal and two plasmid gene sequences. For *flaB*, *groEL*, *recA, IGS* and

hbb; alleles match those of the MLSA *B. burgdorferi* database (http://pubmlst.org/bburgdorferi/). Alleles that have been previously described are indicated (*). In the case of *pyrG* and *recG* genes, the alleles for which a significant match was found in the MLST *B. burgdorferi* database (http://www.mlst.net/) are listed (our allele number/MLST database allele number). For some genes, allele lengths submitted to the databases do not match sequence lengths submitted to GenBank. See Table 1 for details on the sampling origin of isolates and Table 2 for information on the genes used.

Table S2. Recombination events estimated by RDP3 on the 10 loci *B. garinii* data set. Events listed correspond to those detected by more than five different recombination detection methods with a significant *P*-value ($P < 0.05$). Some breakpoint positions are undetermined. The minor and major parents are the strains contributing the smaller and the larger fraction of the recombinant sequence respectively. The question mark means missing parental sequence in the alignment. Terrestrial strains contributing to recombination events are highlighted in bold.

Table S3. Amplification conditions and primer sequences used in this study.

Table S4. Reference sequences used in the comparative network analysis of the *ospA* gene (see Fig. S5).

Table S5. Reference sequences used in the comparative network analysis of the *ospC* gene (see Fig. S6).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.