

# SCIENTIFIC REPORTS



OPEN

## Trans-Atlantic exchanges have shaped the population structure of the Lyme disease agent *Borrelia burgdorferi sensu stricto*

Received: 17 September 2015

Accepted: 18 February 2016

Published: 09 March 2016

S. Castillo-Ramírez<sup>1</sup>, V. Fingerle<sup>2</sup>, S. Jungnick<sup>2</sup>, R. K. Straubinger<sup>3</sup>, S. Krebs<sup>4</sup>, H. Blum<sup>4</sup>, D. M. Meinel<sup>2</sup>, H. Hofmann<sup>5</sup>, P. Guertler<sup>2</sup>, A. Sing<sup>2</sup> & G. Margos<sup>2</sup>

The origin and population structure of *Borrelia burgdorferi sensu stricto* (s.s.), the agent of Lyme disease, remain obscure. This tick-transmitted bacterial species occurs in both North America and Europe. We sequenced 17 European isolates (representing the most frequently found sequence types in Europe) and compared these with 17 North American strains. We show that trans-Atlantic exchanges have occurred in the evolutionary history of this species and that a European origin of *B. burgdorferi* s.s. is marginally more likely than a USA origin. The data further suggest that some European human patients may have acquired their infection in North America. We found three distinct genetically differentiated groups: i) the outgroup species *Borrelia bissettii*, ii) two divergent strains from Europe, and iii) a group composed of strains from both the USA and Europe. Phylogenetic analysis indicated that different genotypes were likely to have been introduced several times into the same area. Our results demonstrate that irrespective of whether *B. burgdorferi* s.s. originated in Europe or the USA, later trans-Atlantic exchange(s) have occurred and have shaped the population structure of this genospecies. This study clearly shows the utility of next generation sequencing to obtain a better understanding of the phylogeography of this bacterial species.

Genome sequencing has drastically changed our understanding of microbial microevolution over the past decades. This is particularly the case for bacteria causing infectious diseases. Furthermore, the advent of next-generation sequencing (NGS) has allowed scientists to handle large population samples and to address a variety of biological questions concerning the evolution and ecology of bacterial pathogens. For example, NGS data has been used to study pathogen evolution within a single patient<sup>1</sup>, transmission chains within a single hospital<sup>2,3</sup> and clonal diversification over time and space<sup>4,5</sup>. Without any doubt the unprecedented resolution provided by NGS is changing the way in which evolutionary biology and molecular epidemiology studies are conducted<sup>4-6</sup>.

*Borrelia burgdorferi sensu stricto*, a tick-borne spirochete bacterium, belongs to the species complex *B. burgdorferi sensu lato* (s.l.) that now comprises more than 20 accepted and proposed species of which five are regularly isolated from human patients<sup>7,8</sup>. The species that cause Lyme Borreliosis (LB) (also called Lyme disease) include *B. burgdorferi* s.s., *B. afzelii*, *B. spielmanii*, *B. garinii* and *B. bavariensis*. LB can be regarded as a multi-systemic disease, because it presents different symptoms in patients, such as erythema migrans (EM), arthritis, neurological manifestations (neuroborreliosis) or lymphocytoma<sup>9-11</sup>. So far, *B. burgdorferi* s.s. is the only human pathogenic *Borrelia* species of the Lyme borreliosis group of spirochetes in North America<sup>9</sup> and symptoms described in the USA differ somewhat from those in Europe<sup>11-13</sup>.

LB is a zoonotic vector-borne disease, hence, the natural transmission cycles of these bacteria depend on vertebrate reservoir hosts and vector ticks of the *Ixodes persulcatus* species complex<sup>14,15</sup>. This ecological dependence

<sup>1</sup>Programa de Genómica Evolutiva, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Apartado Postal 565-A, CP 62210, Cuernavaca, Morelos, México. <sup>2</sup>National Reference Center for Borreliosis at the Bavarian Health and Food Safety Authority, Veterinärstr. 2, 85764 Oberschleissheim, Germany. <sup>3</sup>LMU Munich, Department of Infection and Zoonoses, Veterinärstr. 13, 80539 Munich, Germany. <sup>4</sup>LMU Munich, Gene Centre, Lafuga, Feodor-Lynen-Strasse 25, 81377 Munich, Germany. <sup>5</sup>TU Munich, Klinik für Dermatologie and Allergologie, 80802 Munich, Germany. Correspondence and requests for materials should be addressed to G.M. (email: gmargos1@gmail.com)

Strain Name	Abbreviation	Diagnosis	Year of isolation	Origin	MLST ST
PMeh	Meh	Arthritis	1996	Garmisch	ST20
PBre	Bre	EM*	1988	Munich	ST20
PDri	Dri	EM	1988	Munich	ST24
PFil	Fil	EM	1985	Munich	ST284
PSst	Sst	EM	1993	Tübingen	ST21
PAli	Ali	EM	1994	Regensburg	ST1
PLue	Lue	EM/multiple	1999	Munich	ST3
PFheI	FheI	Lymphom	2010	Munich	ST21
PFheII	FheII	Lymphom	2010	Munich	ST21
PGI	GI	NB <sup>§</sup>	1993	Bayreuth	ST21
PHas	Has	NB	1992	Ulm	ST1
PKu	Ku	NB	1996	Bad Mergentheim	ST20
PMi	Mi	NB	1994	Munich	ST20
PAbe	Abe	NB	1997	Munich	ST1
PKifI	KifI	Skin/chest	1988	Munich	ST24
PKifII	KifII	Skin/knee	1988	Munich	ST24
Z41293	Z412	N/A - Tick isolate		Germany	ST27

**Table 1. Newly sequenced strains.** *B. burgdorferi* s.s. strains newly sequenced, symptoms caused, year of isolation, and geographic origin within Germany. Reads have been submitted to the NCBI Sequence Read Archive (SRA) under accession number SRP051650 \*Erythema migrans. <sup>§</sup>Neuroborreliosis. N/A = not applicable.

determines the geographic distribution of the bacteria (and the disease) to the temperate zones of the northern hemisphere (roughly from latitude 40 degree - 60 degree N). Within this geographic range of vector and host occurrence, the distribution of *Borrelia* species is not uniform. Some *Borrelia* species (e.g. *B. garinii*, *B. afzelii* and *B. bavariensis*) are only found in Eurasia, whereas *B. burgdorferi* s.s. occurs on both sides of the Atlantic Ocean<sup>16</sup>.

Understanding the movement or spread of bacterial pathogens is important for epidemiological surveillance and monitoring and to predict the spread of such pathogens into new geographic regions. Researchers have used *B. burgdorferi* s.s. to understand the movement of these bacteria within endemic areas but also into new geographic regions<sup>17–24</sup>. Depending on the molecular markers used, the origin of the genospecies *B. burgdorferi* s.s. has been suggested to be in the USA (*ospC*<sup>23,25</sup>) or Europe (multilocus sequence typing [MLST]<sup>20</sup>). Furthermore, population level studies using *ospC* suggested that *ospC* type A strains crossed the Atlantic Ocean approximately 200 years ago<sup>23</sup>. Although *B. burgdorferi* s.s. was one of the first bacterial species being sequenced<sup>26</sup>, completed genomes are available only for 17 isolates, mainly originate from the USA (14 out of 17)<sup>27–30</sup>.

In this study, we exploit the high resolution of NGS to obtain a better understanding of the relatedness of populations from Europe and the USA, with a particular view on human patient isolates from Europe, and to re-address the question of the origin of *B. burgdorferi* s.s.. We included in our analysis some previously sequenced strains to create a more comprehensive data set for this species in terms of genome sequences and geographic origin.

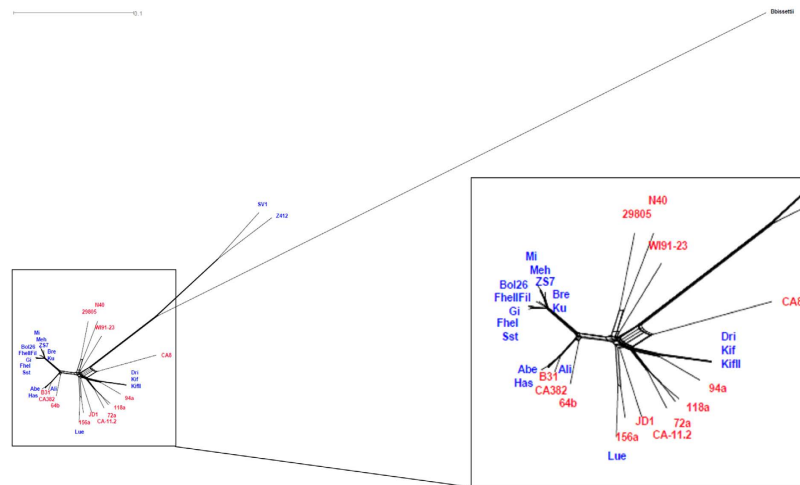
## Results

**The data set and genome sequencing.** We sequenced 17 European isolates (Table 1) using the Illumina MiSeq technology to compare them to fully sequenced genomes available in GenBank, the majority of these strains originated from the USA (see <http://borreliabase.org/>). As we aimed to infer the evolutionary history of strains, we chose to analyse the main chromosome, which contains the genetic information of the core genome. In general, the reads mapped uniformly to the B31 main linear chromosome with more than 90 percent of the chromosome covered. Mean median coverage was 221× (minimum coverage 34× for sample PLue, maximum coverage 419× for sample PAli) (Supplementary Table 1). Regardless of coverage, we observed an alignment gap in the region corresponding to 435 kbp to 442 kbp of B31 (with some variation between strains) for all strains. In B31 this region contains the 5S–23S ribosomal DNA which is duplicated in *LB* spirochetes<sup>31</sup> and this duplication evidently presents a problem for proper mapping of NGS sequence reads. We included in our analyses 17 genome sequences that are publicly available (Table 2); most of them are from the USA (n = 14) while three are from Europe (Bol26, SV1, and ZS7) and the closely related species *B. bissettii* strain DN127 was used as an outgroup. We used progressive Mauve<sup>32</sup> to align the chromosomal genomes of our newly sequenced strains as well as those from the previously published strains. We found 92 locally colinear blocks (LCBs), eight of them contain DNA conserved in all the strains and were used for downstream analysis; these eight LCBs are described in Supplementary Table 2. Of note, we observed similar levels of genetic variation across the eight LCBs (see Supplementary Table 2). In summary, we compiled a data set of 35 strains that contained isolates from both Europe and the USA (including an outgroup).

**Three genetically differentiated groups.** To visualize the genetic relationships among the strains, we first employed a neighbor net (Fig. 1), as this approach accounts for the possibility of recombination. It is

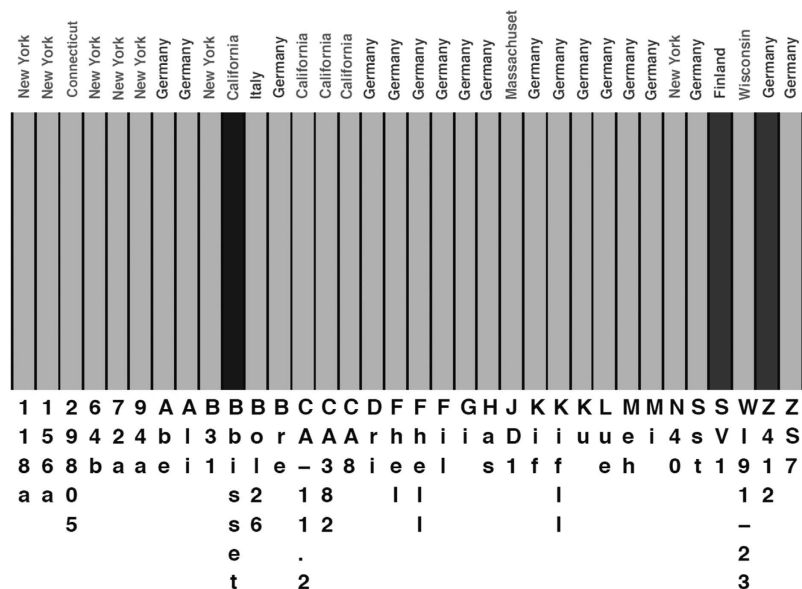
Isolate (GenBank accession)	Geographic origin	Biological source	MLST ST
B31 (AE000783.1)	New York, USA	<i>I. scapularis</i> (tick)	ST1
CA382 (CP005925.1)	California, USA	<i>I. pacificus</i> (tick)	Nd
64b (ABKA02000001.1 - ABKA02000006.1)	New York, USA	Human	ST59
ZS7 (CP001205.1)	Germany, EU	<i>I. ricinus</i> (tick)	ST20
Bol26 (ABCW02000001.1 - ABCW02000005.1)	Italy, EU	<i>I. ricinus</i> (tick)	ST332
W191-23 (ABJW02000001.1 - ABJW02000031.1)	Wisconsin, USA	Bird	ST228
29805 (ABJX02000001.1 - ABJX02000038.1)	Connecticut, USA	<i>I. scapularis</i> (tick)	ST12
N40 (CP002228.1)	New York, USA	<i>I. scapularis</i> (tick)	ST19
JD1 (CP002312.1)	Massachusetts	<i>I. scapularis</i> (tick)	ST11
156a (ABCV02000001.1)	New York, USA	Human	ST4
CA-11.2A (ABJY02000001.1 - ABJY02000014.1)	California, USA	<i>I. pacificus</i> (tick)	ST333
118a (ABGI02000001.1 - ABGI02000008.1)	New York, USA	Human	ST34
72a (ABGJ02000001.1 - ABGJ02000006.1)	New York, USA	Human	ST14
94a (ABGK02000001.1 - ABGK02000009.1)	New York, USA	Human	ST18
CA8 (ADMY01000001.1 - ADMY01000007.1)	California, USA	<i>I. pacificus</i> (tick)	Nd
SV1 (ABJZ02000001.1 - ABJZ02000005.1)	Finland, EU	<i>I. ricinus</i> (tick)	ST414
<i>B. bissettii</i> DN127* (ABJZ02000001.1 - ABJZ02000005.1)	California, USA	<i>I. pacificus</i> (tick)	NA

**Table 2. Publicly available isolates.** We used the BorreliaBase database (<http://borreliabase.org>) to access these chromosomal genomes. This database gives direct access to the GenBank webpages that contain the files for the chromosomal DNA for each isolate. The GenBank accession numbers are in parenthesis. \*This is a closely related species that was used as an outgroup.



**Figure 1. Neighbor Net of the 34 isolates.** Neighbor Net based on all the SNPs, the uncorrected p distance was used and the network was drawn employing the equal angle method. In red are the strains from the USA, whereas blue labels depict the European strains. Considerable reticulation can be observed in the network (see inlet), which is compatible with the notion that recombination has affected this data set. The Phi test for detecting recombination was conducted and significant signals for recombination were found ( $p < 0.001$ ).

known that significant amounts of variation on plasmids of *B. burgdorferi* s.s. have arisen by recombination<sup>33</sup>, and our analysis showed that some recombination seems to occur on the main chromosome (Fig. 1). As would be expected if recombination occurs, the network showed some reticulation reflecting conflicting phylogenetic signals; corroborating this result, the PhiTest gave significant signals for recombination ( $p < 0.001$ ). Neither the European strains (blue) nor the American strains (red) formed a monophyletic group (Fig. 1). Actually, there were several clades that contained both European and American strains. Strains SV1 and Z41293 (which were both isolated from *I. ricinus* ticks) clustered together and were well separated from the rest of the strains, suggesting that these two strains comprised divergent strains of *B. burgdorferi* s.s.. All *B. burgdorferi* s.s. strains were clearly differentiated from the outgroup and although SV1 and Z41293 were considerably separated from the other *B. burgdorferi* s.s. strains, they were much closer to the rest of the *B. burgdorferi* s.s. strains than to the *B. bissettii* outgroup. To further study the genetic relationships of these strains we used the Bayesian Analysis of Population Structure software (BAPS)<sup>34</sup>. Three genetically differentiated groups were found (Fig. 2). The smallest



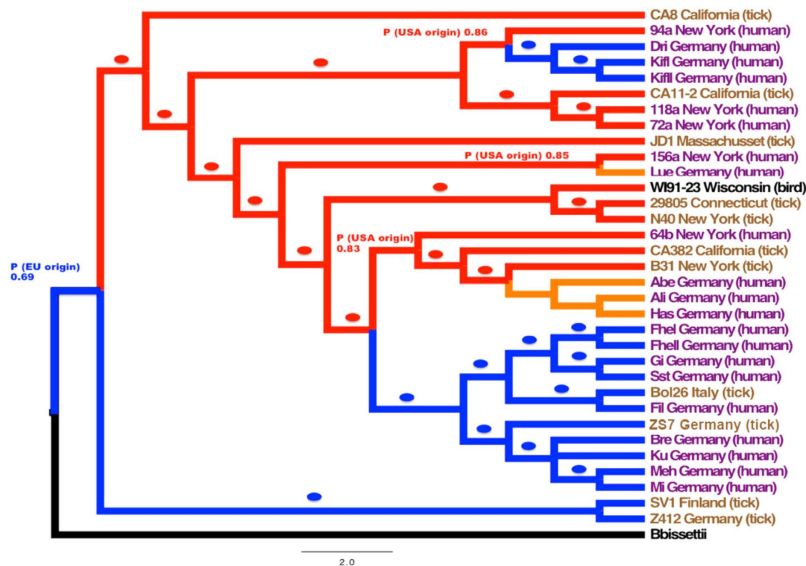
**Figure 2. Mixture and Admixture analyses of the data set.** Mixture and Admixture analyses were conducted through BAPS. Each color represents one of three genetically differentiated groups and each vertical colored bar corresponds to one isolate. The color coding is as follows: black, BAPS group 3; dark grey, BAPS group 2; light grey, BAPS group 1. No evidence for admixture between groups was found.

group (BAPS group 3) contained exclusively the outgroup strain (black bar) and another small group (BAPS group 2) had the two divergent strains (dark grey bars). BAPS group 1 is the largest group and contained all the remaining strains (light grey bars). Notably, neither the American strains nor the European strains formed particular clusters; this result is consistent with genetic exchange/migration between European and North American populations of *B. burgdorferi* s.s.. Remarkably, we did not detect any signals of admixture between the different BAPS groups. Taken together, these results indicate population structuring and trans-Atlantic exchange between Europe and the USA.

**Agreement between chromosomal phylogeny and MLST.** We used phylogenetic analysis to provide additional insights into the population structure of *B. burgdorferi* s.s.. Given the distorting effects of recombination on phylogenetic analysis, we first identified the recombination events between genomes via BRATNextGen<sup>35</sup>; this software implements a Bayesian analysis for detecting recombination events in whole-genome sequences. This analysis suggested that 8,321 out of 68,158 (~12%) of the SNPs reside in regions of the genome that have undergone recombination. Using the remaining SNPs not affected by recombination (the SNP alignment is available upon request), we constructed both Bayesian and Maximum Likelihood (ML) phylogenies, which showed very similar topologies; for simplicity only the Bayesian phylogeny is presented (Fig. 3, Supplementary Figures 1 and 2). All three groups identified by BAPS show very strong node support (posterior probabilities ranged from 1.00 to 0.79).

Given the tree topologies of our Bayesian and ML phylogenies, we wanted to ensure that this result was not due to long-branch attraction (LBA); the wrong grouping of two or more long non-sister branches in a tree due to homoplasies<sup>36</sup>. Specifically, we wanted to corroborate that the position of the divergent group was not a result of LBA and, hence, we conducted two exploratory analyses. The rationale of the long-branch extraction strategy (our first approach) is to take out the long branches - the ones allegedly causing the LBA - one at a time, and see whether this disturbs the groupings present in the tree<sup>37</sup>. Therefore, we constructed one ML phylogeny without *B. bissettii* and another without SV1 and Z41293 (see methods). Notably, the groupings in the trees were hardly affected when either *B. bissettii* or SV1 and Z41293 were removed (Supplementary Figure 4) and the trees were very similar to those in our initial Bayesian and ML phylogenies; thus LBA does not seem to affect our phylogenetic reconstructions. However, to provide additional independent support, we also carried out a partition analysis where the idea is to eliminate the fast-evolving sites, the ones most contributing to LBA. Thus, we created a phylogeny based on the SNP alignment without the fast-evolving sites (see methods); this phylogeny shows identical groups to the Bayesian and ML phylogenies and was very similar to the phylogenies from the previous analysis (see Supplementary Figure 5). These two exploratory analyses convincingly show that LBA is not affecting our data and that the position of the divergent group with respect to the outgroup is not an artifact.

Our tree suggested that different genotypes are likely to have been introduced several times into the same area. For instance, PLue, is genetically distant from PDri, PKifI and PKifII, even though all four strains were collected in Munich. The same applies for the USA strains; for example, although strains CA8 and CA382 were both collected in California, these strains are located on distant branches of the tree. Furthermore, this phylogeny strongly suggests that intercontinental exchanges might have occurred from the USA to EU. By way of illustration, PABe, PAlI, and PHas (all from Germany) cluster within a group that also includes B31 and CA382 (from New York



**Figure 3. Bayesian Phylogeny of the 34 isolates.** Bayesian phylogeny based on SNPs was not affected by recombination. The color-coding of the tips reflects the biological source of the isolates and it is as follows: purple, human isolates and brown tick isolates. The color of the branches describes the results of the ancestral state reconstruction analysis; blue reflects a European origin, whereas red implies USA origin. The little ellipses show those cases where the marginal posterior probability (MPP) of a European origin (blue) or North American origin (red) was higher than 0.95; cases where the MPP was lower than 0.95 are shown in the tree. The orange branches describe those strains that, although they were collected in Germany, the true geographic origin is uncertain. We used FigTree v1.4.0 to transform the branches (cladogram) for clarity; the tree without transformation is available as Supplementary Figure 1. The scale bar shows the estimated number of substitutions per SNP.

and California, respectively). Another example is 156a, which was collected from New York and clusters with PLue, a German isolate. Given the topology of our tree, at least four trans-Atlantic migrations might have happened, all of them within BAPS group 1. Our phylogenetic analysis is in good agreement with the pattern of genetic relatedness among the strains that had been previously shown using MLST based on eight chromosomally located housekeeping genes<sup>20</sup> (Supplementary Figure 3) suggesting that MLST provides a good approximation of strain relatedness. Furthermore, the phylogenetic analysis gives additional support to the suggestion that *Borrelia* populations are genetically structured, have experienced trans-Atlantic exchange and that SV1 and Z412 are the most divergent strains within the *B. burgdorferi* s.s. species. As most of our strains originated from human patients, we used the MLST data to compare the strains analysed here with tick-derived strains in the MLST database (Supplementary Figure 6). All but two STs (i.e. ST1 and ST3, which have so far been only found in human patients) are also found in the European tick vector. The majority of STs found in ticks and humans in the USA differ from those found in Europe, the exception being ST1 and ST3.

**Geographic origin.** To infer the geographic origin of *B. burgdorferi* s.s., we conducted an ancestral state reconstruction analysis using SIMMAP<sup>38</sup>; the results of this analysis are summarized in Fig. 3. This figure shows that a European origin is moderately more probable than a North American origin for the ancestor of all the *B. burgdorferi* s.s. strains (marginal posterior probability (MPP) of EU origin 0.69) but is highly likely for the group composed of the two divergent strains included in BAPS group 2 (MPP of EU origin 0.95). The directionality of subsequent movements is, however, not straightforward. According to our analysis, we were able to detect several trans-Atlantic exchanges and most of them seem to go from the USA to Europe. For instance, if only the three strains PAli, PAbe and PHas are considered in the analysis – a tight cluster of strains from Munich – a European origin is very plausible (MPP of EU origin 0.999). However, if the closest related strain (namely B31 from New York) to those three strains is included, the situation is reversed and now a North American origin is much more likely (MPP of USA origin 0.978, Table 3); this implies an introduction from the USA into Europe. Similarly, given that the ancestor of the group composed of 156s and PLue seems to have been in the USA (MPP of USA origin 0.85, see Fig. 3), PLue may reflect a recent introduction to Europe from the USA. The MLST Sequence Types (ST) were ST1 for B31, PAli, PAbe, PHas and ST3 for PLue. In this context it is noteworthy that both STs have frequently been found in *I. scapularis* ticks in the USA (but not in *I. ricinus* ticks in Europe) further supporting an American origin of the respective clades. Interestingly, STs that are found in European ticks (e.g. ST20, ST24) form their own cluster within the phylogeny and these STs have not been described in either patients or ticks in the USA (Supplementary Figure 6). Thus, the ancestral state reconstruction analysis suggests a moderately more likely European origin for the original population and also clarifies the probable directionality of some of the trans-Atlantic exchanges.

Group	Prob(USA origin)	Prob(EU origin)
All <i>B. burgdorferi</i> s.s. strains	0.311892	0.688108
BAPS group 2 (SV1 and Z412)	0.049205	0.950795
PLue and 156a	0.849954	0.150046
PAlI, PAbe, PHas and B31	0.977758	0.022242
PAlI, PAbe, PHas	0.000002	0.999998

**Table 3. Ancestral State Reconstruction Analysis.** Posterior probabilities of a European or an American origin for all the different groups.

## Discussion

In this study we sequenced 17 European *B. burgdorferi* s.s. strains isolated from human patients and from a tick via NGS. We also included previously sequenced strains into our analyses (to produce a data set of 34 strains) to obtain a better understanding of the relationship of strains between Europe and the USA and the origin of *B. burgdorferi* s.s.. Importantly this dataset, although limited in extent, includes both human and tick isolates from both Europe and the USA (see Tables 1 and 2); without considering the outgroup, 11 isolates (~33%) came from ticks and 13 isolates (~39%) were collected in the USA. Even though the newly sequenced isolates included in the study originated mainly from German patients, they do represent the most frequently found MLST STs of *B. burgdorferi* s.s. in Europe, which have been found in ticks in various regions such as Latvia, France and UK<sup>(39,40)</sup>, [www.pubmlst.org/borrelia/](http://www.pubmlst.org/borrelia/)).

We focused on the main linear chromosome because it is more likely to reflect the core genome<sup>28</sup> and may be less prone to horizontal transfer than plasmid-encoded genes<sup>27,33,41</sup>. However, we found that a significant amount of the SNPs (12%) were located on regions inferred to be affected by recombination. This result was not unexpected as previous studies suggested recombination to mutation rates of 3:1 in *B. burgdorferi* s.s., although this ratio was found in plasmids rather than the main linear chromosome<sup>33</sup>. However, more recent population genomics studies using the main linear chromosome and two plasmids of European strains of *B. burgdorferi* s.s. suggested recombination to mutation ratios of 1.7<sup>42</sup>. These data reinforce the view that recombination is one of the forces creating novel combinations of alleles within this species.

Strain clustering in phylogenetic analysis correlated only poorly with geography. Notably, of three well-differentiated genetic groups, one comprised both European and American strains, supporting the idea of trans-Atlantic exchange<sup>33</sup>. In addition, phylogenetic analysis showed that strains from Europe are nested within the USA clades suggesting that at least four intercontinental migration events have occurred in that direction (i.e. from the USA to Europe). The routes and vehicles of these exchanges are unknown and deserve further investigation (see below) though our data suggest that some European patients may have acquired their infection in North America. One ancestral migration might have occurred at a more basal position of the tree resulting in the split between BAPS1 and BAPS2 clades, an event that gave rise to the USA population. Apart from this, our phylogenetic analysis demonstrates that multiple introductions, with different genotypes, might have occurred into the same geographic area. This is also supported by the BAPS analysis, as we found that strains from sites that are in geographic close proximity belong to different BAPS groups. In accordance with this, no admixture was detected between the BAPS groups, not even between BAPS groups 2 and 1, both of which have *B. burgdorferi* s.s. strains. Therefore, besides geography with intercontinental exchange, additional unknown factors probably operate to generate population structuring.

Whether the divergence of the two strains from BAPS group 2 reflects also differences in ecology is currently unknown. Strains related to Z41293 (here abbreviated Z412) have been found in *I. ricinus* ticks in Switzerland and France<sup>20</sup>, but further ecological information on these strains is not available. Z41293 and related strains have been used to determine the species threshold for *B. burgdorferi* s.l.<sup>39,43</sup>. Strain Z41293 and related strains were considered “borderline” *B. burgdorferi* s.s.<sup>43</sup> because DNA-DNA hybridization studies found that this strain showed 74 percent similarity and a  $T_m$  of 2 °C to strain B31. Using MLST, strain SV1 also falls within the borders of *B. burgdorferi* s.s. species. As the former strains served to set the threshold for species delineation, in our view it is problematic to designate them as a new species<sup>44</sup>. Despite the contentious status of Z41293 and related strains, these divergent strains have so far only been detected in Europe, but not in North America supporting the idea of a European origin of *B. burgdorferi* s.s..

We conducted an ancestral state reconstruction analysis to infer the geographic origin of the species and it suggests that the ancestor of all the strains is moderately more likely to have had a European origin than a North American origin. This relates to the fact that the most divergent group (BAPS group 2) has a probable European origin. In this context, we would like to point out that, in order to correct for the sampling bias (more European isolates), we employed an equal prior so that *a priori* both a European and a USA origin had an equal probability of occurrence. This finding combined with the fact that these strains have been found only in Europe further supports the notion that the seminal population originated in Europe as suggested previously by<sup>20</sup>. Our analysis also implies that after the origin of the primordial population, several trans-Atlantic exchanges have evidently occurred.

However, considering the position of *B. burgdorferi* s.s. in a whole species tree<sup>7</sup>, a number of potential alternative scenarios can be considered: A) the common ancestor of *B. burgdorferi* s.s. originated in Europe and following the split from the divergent clade (BAPS group 2), the resulting *B. burgdorferi* s.s. migrated to North America. Since then several trans-Atlantic migrations have occurred forming the European subpopulations and a North American subpopulation. B) Another possibility is that all *B. burgdorferi* s.s. evolved in North America.

The divergent group migrated to Europe and died out in North America. C) *B. burgdorferi* s.s. evolved in North America and the divergent group still exists in North America but at very low frequency and has simply not been found in natural transmission cycles. It needs to be emphasized that the species tree shown by<sup>7</sup> was based on eight chromosomal housekeeping loci and may change if more genetic information is incorporated. We acknowledge that denser sampling from both Europe and the USA – especially of isolates from this divergent group – and time dating of population migrations is needed to be able to resolve these issues unambiguously. We also suggest that in order to resolve the phylogeography of *B. burgdorferi* s.s., other genospecies of the LB group of spirochetes need to be included into future phylogenetic investigations.

It is difficult to reconcile the population structure of *B. burgdorferi* s.s. with natural transmission cycles. Two different tick vector species maintain the transmission cycles in Europe and North America, namely *I. ricinus* and *I. scapularis*, respectively. Trans-Atlantic migration and mixing of tick vector species has not been reported suggesting that the Atlantic Ocean represents a barrier to migration/mixing of vector species<sup>14</sup>. Although *B. burgdorferi* s.s. can utilize many hosts as reservoirs including rodents, insectivores and birds<sup>16</sup>, none of these hosts is known to regularly migrate East-to-West or *vice versa* across the Atlantic. Whether sea birds may be involved in transcontinental migration of *B. burgdorferi* s.s.<sup>45</sup> needs to be further investigated. We would like to highlight that strains of *B. burgdorferi* s.s. found in Europe – and showing the same ST as North American strains (ST1 and ST3) – have only been isolated from human patients (see below)<sup>(23,46)</sup>, and this study), so humans are potentially the vehicle to transport strains of *B. burgdorferi* s.s. between continents. If this is the case, these strains are highly unlikely to enter natural transmission cycles as humans are not considered to contribute to the natural transmission of *Borrelia*<sup>14</sup>. Intriguingly, ST1 strains of *B. burgdorferi* s.s. are commonly found in ticks and patients in the USA, but they have not been reported from ticks in Europe (see [pubmlst.org/borrelia/](http://pubmlst.org/borrelia/) database<sup>46</sup>). There are several possible scenarios that might explain the finding of such closely related strains to B31 in patients in Europe. First, these patients may have acquired their *Borrelia* infections while travelling in the USA. Second, such strains may be present at a very low prevalence in ticks in Southern Germany as ticks investigated for *B. burgdorferi* s.s. had been collected in Latvia, Scotland, France or Italy and this ST has not been found<sup>47,48</sup>. As pointed out before, *B. burgdorferi* s.s. is not very prevalent in Europe; LB species such as *B. garinii* or *B. afzelii* are much more common than *B. burgdorferi* s.s.<sup>49</sup>. In this scenario, the fact that B31-related strains are not found in ticks but only in patients might suggest a high potential for establishing an infection in humans. However, a discrepancy between tick and human isolates has been observed in the USA, not all STs isolated from human patients have been detected in ticks<sup>50</sup> suggesting that some strains although rare in natural transmission cycles can be found in human patients.

The NGS data provided very robust phylogenies and it was interesting that the distribution of STs matched well with clades generated by MLST analysis. This suggests that MLST can be used as a good approximation to explore/investigate the relationship of *B. burgdorferi* strains, although some authors reported otherwise<sup>28</sup>. Interestingly and, in contrast to other bacterial pathogens, there was a limited amount of variation between strains belonging to the same ST. For example, strains that belonged to ST1 (PAlI, PABe, PHas) differed in only 63, 64 and 261 SNPs, respectively, to B31 on the main linear chromosome. In comparison, an ST (i.e. ST239) of *Staphylococcus aureus* strains comprising identical MLST types differed in more than 4,000 SNPs<sup>51</sup>. On the other hand, due to the high resolution provided by NGS, no two strains show the same genotype – not even FheI and FheII that were collected from the same patient at the same time. That was still true when only the non-recombinant SNPs were considered.

In conclusion, our study shows that population structuring and trans-Atlantic exchanges have occurred during the evolutionary history of *B. burgdorferi* s.s. and may still occur today. However, we do acknowledge that more data are required to infer migration routes and transport vehicles. Importantly, it seems that some European patients may have acquired their infection in North America. Thus, our study clearly gives a demonstration of the utility of NGS platforms for studying the evolutionary history of vector-borne bacterial pathogens, especially for those where genetic variation is low.

## Materials and Methods

**The data set and genome sequencing.** We analyzed 17 strains that were previously typed to be *B. burgdorferi* s.s. and were available at the German National Reference Centre for *Borrelia* (Table 1). *B. burgdorferi* s.s. is less common in Europe than in North America (13% of all European species in the MLST database). The strains included in the study were chosen to account for the genetic variation found by MLST in European *B. burgdorferi* s.s. to attempt to limit phylogenetic discovery bias<sup>52</sup>. The MLST STs included were ST20, ST21, ST24, ST284 and ST27, which represent the most common STs in *Ixodes ricinus* ticks from Germany, Latvia, UK, France and human patients (see [PubMLST.org/database/borrelia/](http://PubMLST.org/database/borrelia/)). ST27 represents a divergent strain, which was used to determine the species threshold in *Borrelia*<sup>39</sup>. ST1 was frequently found in human patients and was included to determine its relationship to strains found in North America<sup>23</sup>. Strains were isolated between 1985 and 2010 from patients and one *I. ricinus* tick. Details are given in Table 1. *Borrelia* cultures grown under standard conditions<sup>53</sup> were used for DNA extraction via a Maxwell<sup>®</sup> 16 genomic DNA purification kit (Promega, Germany). Following DNA quantification, libraries were prepared according to the Nextera DNA sample preparation guide (Qiagen, Germany). The samples were diluted to a concentration of 50 ng per microliter and “tagmented” by simultaneously fragmenting DNA using transposomes as provided by the manufacturer and adding adapters. After tagmentation, samples having adapters on both ends underwent 5 PCR cycles to amplify the product and to add index primers. The resulting libraries were then validated using an Agilent 2100 Bioanalyzer (Agilent, Germany). Paired end sequence reads (250 bp) were generated using an Illumina Miseq platform (Illumina, San Diego CA, USA). We used the B31 and SV1 main linear chromosomes (B31, GenBank accession number NC 001318; SV1, GenBank accession number NZ ABJZ00000000.2) as reference. We concentrated on the main chromosome because it most likely represents the core genome for *Borrelia*<sup>27</sup> which is best suited to infer the evolutionary

history of strains. For *de novo* assembly, reads were quality checked and filtered using FastQ Trimmer (window size 10, step size 2, Q score 20)<sup>54</sup>, kmernator (k-mer size 31) and velvetoptimizer<sup>55</sup> implemented in the Galaxy platform<sup>56,57</sup> provided by the Gene Centre, LMU, Munich, Germany. The following parameters were used in velvetoptimizer: start Hash length 35, end hash length 85, no. of threads 4, k-mer optimization metric N50. Resulting contigs were mapped onto the reference chromosomes with Burrows-Wheeler Aligner (BWA)<sup>58</sup> and re-arranged and sorted using the Mauve Contig Mover algorithm<sup>59</sup> which is within the Mauve Genome Alignment Software<sup>60</sup>. The reference chromosomes were B31 for 16 strains and SV1 for strain Z41293. Contigs that did not align to the main linear chromosome were discarded. Some of the publicly available genomes are in a draft status and we therefore used the Mauve Contig Mover to order the contigs; in such cases B31 was used as the reference genome. We used progressive Mauve<sup>32</sup> to align the chromosome genomes of our newly sequenced strains as well as those from the previously published strains. We employed the default parameters and a SNP file was created; 92 colinear sub-alignments (LCBs) were found, but only eight of them contained DNA conserved in all the strains. An SNP alignment was created which is available upon request. For comparison with the MLST scheme used by<sup>20</sup> and<sup>50</sup>, sequence data of the eight housekeeping loci (*clpA*, *clpX*, *nifS*, *pepX*, *pyrG*, *recG*, *rplB*, *uvrA*) were extracted from whole genome sequences if not already available from the MLST database hosted at the University of Oxford (PubMLST.org).

**Neighbor net and detection of recombination.** We employed SplitTree4 software on the SNP alignment to construct a Neighbor Net<sup>61</sup>, the uncorrected P distance was used. We use the BRATNextGen software<sup>35</sup> to identify the recombination events exclusively in the eight LCBs that contained DNA conserved in all the strains. We used similar settings as those used by Castillo-Ramirez *et al.*<sup>4</sup>. A p-value of less than 0.01 was set to determine the significance of the recombinant segments, this was done via a bootstrap test with 100 replicates.

**Maximum Likelihood and Bayesian phylogenies.** We constructed phylogenies by means of PhyML (Maximum Likelihood; ML)<sup>62</sup> and MrBayes<sup>63</sup>. The multiple alignment on which the phylogenies were run, contained only SNPs that were located in the eight conserved LCBs and that were not present in regions affected by recombination. We used the GTR model and, because we only considered SNPs, we did not include a proportion of invariable sites. Base frequencies as well as the relative substitution rates between them were calculated by maximizing the likelihood of the phylogeny. We set the GTR model, allowing gamma-distributed rate variation across sites. We ran the analysis for 8,000,000 generations, sampling every 1000 generations – 25% were discarded as burn-in. To ensure that two runs had converged we checked that the average standard deviation of the split frequencies were below 0.01.

**Bayesian Analysis of Population Structure (BAPS).** The mixture and admixture analyses were implemented through BAPS 5<sup>34</sup>. To conduct the population mixture analysis we selected the “Clustering of individuals” module. To properly evaluate the number of genetically diverged groups, we ran the analysis several times with different values for the maximum number of genetically diverged groups; these values ranged from 2–20. We chose the “Admixture based on mixture clustering” module to implement the population admixture analysis. The minimum population size was set to 2; we used 100 iterations to calculate the admixture coefficient for the individuals and the number of reference individuals from each population was set to 200. Finally, 20 iterations were used to calculate the admixture coefficient for the reference individuals.

**Ancestral Reconstruction Analysis.** The ancestral state reconstruction analysis was carried out by means of SIMMAP<sup>38</sup>. The ancestral state reconstruction module was employed on the trees that were derived from a Bayesian analysis conducted with MrBayes. The geographic source of the strains was coded as 1 (Europe) and 0 (USA) and we configured the models as follows: for the bias parameter we selected an equal prior (1/k) – this was in order to not bias our analysis given the overrepresentation of European strains – and for the rate parameter we specified a gamma distribution with an alpha of 1.25, beta of 0.25 and k equal to 90.

**Detecting long-branch attraction (LBA).** We used two strategies to detect LBA. First, we conducted a long-branch extraction approach<sup>37</sup> in which the branches that are potential candidates for causing LBA are extracted one at a time. If LBA is the problem this should affect the clustering pattern in the remaining taxa. Therefore, we constructed two new ML phylogenies: one without the outgroup and another without strains SV1 and Z41293 that are located on the second longest branch (see Supplementary Figure 4). As a second strategy, we conducted a partition analysis, in which the fast-evolving sites are discarded as these mostly contribute to LAB. For this, MEGA6<sup>64</sup> was used to estimate position-by-position rates under the GTR model employing a discrete Gamma distribution (with 5 categories) to model the rate differences across the SNPs. Then, a new alignment was created that included only the SNPs with slower relative evolutionary rates than the two fast evolving categories from the discrete Gamma distribution. A ML phylogeny was made on this alignment (see Supplementary Figure 5).

## References

1. Mwangi, M. M. *et al.* Tracking the *in vivo* evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. *Proc Natl Acad Sci USA* **104**, 9451–9456 (2007).
2. Snitkin, E. S. *et al.* Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. *Science translational medicine* **4**, 148ra116 (2012).
3. Torok, M. E. *et al.* Rapid whole-genome sequencing for investigation of a suspected tuberculosis outbreak. *J Clin Microbiol* **51**, 611–614 (2013).
4. Castillo-Ramirez, S. *et al.* Phylogeographic variation in recombination rates within a global clone of methicillin-resistant *Staphylococcus aureus*. *Genome Biol* **13**, R126 (2012).
5. Holden, M. T. *et al.* A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res* **23**, 653–664 (2013).



6. Croucher, N. J. *et al.* Rapid pneumococcal evolution in response to clinical interventions. *Science* **331**, 430–434 (2011).
7. Margos, G., Vollmer, S. A., Ogden, N. H. & Fish, D. Population genetics, taxonomy, phylogeny and evolution of *Borrelia burgdorferi* sensu lato. *Infect Genet Evol* **11**, 1545–1563 (2011).
8. Rudenko, N., Golovchenko, M., Grubhoffer, L. & Oliver, J. H., Jr. Updates on *Borrelia burgdorferi* sensu lato complex with respect to public health. *Ticks Tick Borne Dis* **2**, 123–128 (2011).
9. Bacon, R. M., Kugeler, K. J. & Mead, P. S. Surveillance for Lyme disease—United States, 1992–2006. *MMWR Surveill Summ* **57**, 1–9 (2008).
10. Stanek, G. *et al.* Lyme borreliosis: Clinical case definitions for diagnosis and management in Europe. *Clin Microbiol Infect* **17**, 69–79 (2010).
11. Steere, A. C. Lyme borreliosis in 2005, 30 years after initial observations in Lyme Connecticut. *Wien Klin Wochenschr* **118**, 625–633 (2006).
12. Strle, F. & Stanek, G. Clinical manifestations and diagnosis of Lyme borreliosis. *Current Problems in Dermatology* **37**, 51–110 (2009).
13. Steere, A. C. Lyme disease. *N Engl J Med* **345**, 115–125 (2001).
14. Kurtenbach, K. *et al.* In *Bacterial Population Genetics in Infectious Disease*, Edn. 1. (eds D. A. Robinson, D. Falush & E. J. Feil) (John Wiley & Sons, Inc., 2010).
15. Radolf, J. D., Caimano, M. J., Stevenson, B. & Hu, L. T. Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes. *Nat Rev Microbiol* **10**, 87–99 (2012).
16. Piesman, J. & Gern, L. Lyme borreliosis in Europe and North America. *Parasitology* **129** Suppl, S191–220 (2004).
17. Brisson, D. *et al.* Evolution of northeastern and midwestern *Borrelia burgdorferi*, United States. *Emerg Infect Dis* **16**, 911–917 (2010).
18. Hoen, A. G. *et al.* Phylogeography of *Borrelia burgdorferi* in the eastern United States reflects multiple independent Lyme disease emergence events. *Proc Natl Acad Sci USA* **106**, 15013–15018 (2009).
19. Margos, G., Castillo-Ramirez, S. & Hoen, A. G. Phylogeography of Lyme borreliosis-group spirochetes and methicillin-resistant *Staphylococcus aureus*. *Parasitology* **139**, 1952–1965 (2012).
20. Margos, G. *et al.* MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. *Proc Natl Acad Sci USA* **105**, 8730–8735 (2008).
21. Mechai, S. *et al.* Phylogeographic analysis reveals a complex population structure of *Borrelia burgdorferi* in southeastern and south central Canada. *Appl Environ Microbiol* (2014).
22. Ogden, N. H. *et al.* Investigation of genotypes of *Borrelia burgdorferi* in *Ixodes scapularis* ticks collected in surveillance in Canada. *Appl Environ Microbiol* **77**, 3244–3254 (2011).
23. Qiu, W. G. *et al.* Wide distribution of a high-virulence *Borrelia burgdorferi* clone in Europe and North America. *Emerg Infect Dis* **14**, 1097–1104 (2008).
24. Qiu, W. G., Dykhuizen, D. E., Acosta, M. S. & Luft, B. J. 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 (2002).
25. Marti Ras, N., Postic, D., Foretz, M. & Baranton, G. *Borrelia burgdorferi* sensu stricto, a bacterial species “made in the U.S.A.”? *Int J Syst Bacteriol* **47**, 1112–1117 (1997).
26. Fraser, C. M. *et al.* Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390**, 580–586 (1997).
27. Casjens, S. R. *et al.* Genome stability of Lyme disease spirochetes: comparative genomics of *Borrelia burgdorferi* plasmids. *PLoS ONE* **7**, e33280 (2012).
28. Mongodin, E. F. *et al.* Inter- and intra-specific pan-genomes of *Borrelia burgdorferi* sensu lato: genome stability and adaptive radiation. *BMC Genomics* **14**, 693 (2013).
29. Schutzer, S. E. *et al.* Whole-genome sequences of *Borrelia bissettii*, *Borrelia valaisiana*, and *Borrelia spielmanii*. *J Bacteriol* **194**, 545–546 (2012).
30. Schutzer, S. E. *et al.* Whole-genome sequences of thirteen isolates of *Borrelia burgdorferi*. *J Bacteriol* **193**, 1018–1020 (2011).
31. Schwartz, J. J., Gazumyan, A. & Schwartz, I. rRNA gene organization in the Lyme disease spirochete, *Borrelia burgdorferi*. *J Bacteriol* **174**, 3757–3765 (1992).
32. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5**, e11147 (2010).
33. 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 USA* **101**, 14150–14155 (2004).
34. Tang, J., Hanage, W. P., Fraser, C. & Corander, J. Identifying currents in the gene pool for bacterial populations using an integrative approach. *PLoS computational biology* **5**, e1000455 (2009).
35. Marttinen, P. *et al.* Detection of recombination events in bacterial genomes from large population samples. *Nucleic acids research* **40**, e6 (2012).
36. Bergsten, J. A review of long-branch attraction. *Cladistics* **21**, 163–193 (2005).
37. Siddall, M. E. & Whiting, M. F. Lang-Branch Abstractions. *Cladistics* **15**, 9–24 (1999).
38. Bollback, J. P. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* **7**, 88 (2006).
39. Margos, G. *et al.* A new *Borrelia* species defined by Multilocus Sequence Analysis of Housekeeping Genes. *Appl Environ Microbiol* **75**, 5410–5416 (2009).
40. Vollmer, S. A. *et al.* Host migration impacts on the phylogeography of Lyme Borreliosis spirochaete species in Europe. *Environ Microbiol* **13**, 184–192 (2011).
41. Seifert, S. N., Khatchikian, C. E., Zhou, W. & Brisson, D. Evolution and population genomics of the Lyme borreliosis pathogen, *Borrelia burgdorferi*. *Trends in genetics : TIG* **31**, 201–207 (2015).
42. Jacquot, M. *et al.* Comparative population genomics of the *Borrelia burgdorferi* species complex reveals high degree of genetic isolation among species and underscores benefits and constraints to studying intra-specific epidemiological processes. *PLoS ONE* **9**, e94384 (2014).
43. Postic, D., Garnier, M. & Baranton, G. Multilocus sequence analysis of atypical *Borrelia burgdorferi* sensu lato isolates - description of *Borrelia californiensis* sp. nov. and genomospecies 1 and 2. *Int J Med Microbiol* **297**, 263–271 (2007).
44. Casjens, S. R. *et al.* Whole genome sequence of an unusual *Borrelia burgdorferi* sensu lato isolate. *J Bacteriol* **193**, 1489–1490 (2011).
45. Duneau, D. *et al.* Prevalence and diversity of Lyme borreliosis bacteria in marine birds. *Infect Genet Evol* **8**, 352–359 (2008).
46. Crowder, C. D. *et al.* Genotypic variation and mixtures of Lyme *Borrelia* in *Ixodes* ticks from North America and Europe. *PLoS One* **5**, e10650 (2010).
47. James, M. C., Gilbert, L., Bowman, A. S. & Forbes, K. J. The Heterogeneity, Distribution, and Environmental Associations of *Borrelia burgdorferi* Sensu Lato, the Agent of Lyme Borreliosis, in Scotland. *Frontiers in public health* **2**, 129 (2014).
48. Vollmer, S. A. *et al.* Spatial spread and demographic expansion of Lyme borreliosis spirochaetes in Eurasia. *Infect Genet Evol* **14C**, 147–155 (2013).
49. Rauter, C. & Hartung, T. Prevalence of *Borrelia burgdorferi* sensu lato genospecies in *Ixodes ricinus* ticks in Europe: A metaanalysis. *Appl Environ Microbiol* **71**, 7203–7216 (2005).
50. Hanincova, K. *et al.* Multilocus sequence typing of *Borrelia burgdorferi* suggests existence of lineages with differential pathogenic properties in humans. *PLoS One* **8**, e73066 (2013).
51. Harris, S. R. *et al.* Evolution of MRSA During Hospital Transmission and Intercontinental Spread. *Science* **327**, 469–474 (2010).

52. Pearson, T. *et al.* Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proc Natl Acad Sci USA* **101**, 13536–13541 (2004).
53. Ruzic-Sabljić, E. & Strle, F. Comparison of growth of *Borrelia afzelii*, *B. garinii*, and *B. burgdorferi sensu stricto* in MKP and BSK-II medium. *Int J Med Microbiol* **294**, 407–412 (2004).
54. Blankenberg, D. *et al.* Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**, 1783–1785 (2010).
55. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821–829 (2008).
56. Giardine, B. *et al.* Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* **15**, 1451–1455 (2005).
57. Goecks, J., Nekrutenko, A. & Taylor, J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11**, R86 (2010).
58. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
59. Rissman, A. I. *et al.* Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics* **25**, 2071–2073 (2009).
60. Darling, A. E., Tritt, A., Eisen, J. A. & Facciotti, M. T. Mauve assembly metrics. *Bioinformatics* **27**, 2756–2757 (2011).
61. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**, 254–267 (2006).
62. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704 (2003).
63. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**, 539–542 (2012).
64. Tamura, K. *et al.* MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**, 2725–2729 (2013).

## Acknowledgements

The study was supported by the Bavarian Ministry for Health and Care and by the German Federal Ministry of Health via the Robert Koch-Institute and its National Reference Laboratories Network. This work was partially supported by “Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica PAPIIT”, grant number IA200515 to S.C.R. The authors would like to thank Cecilia Hizo-Teufel and Silvia Stockmeier for excellent technical assistance. V.F., A.S. and G.M. are members of ESGBOR. S.C.R. would like to dedicate this work to the 43 trainee teachers from Ayotzinapa, who were abducted in September 2014 in the town of Iguala, Mexico.

## Author Contributions

G.M., V.F., R.S. and A.S. designed the study, S.K., H.B., S.J., D.M.M. and P.G. performed experiments and sequencing, S.C.R., G.M. and D.M.M. analysed data and HH contributed new reagents. G.M., S.C.R. and V.F. prepared figures and wrote the manuscript and all co-authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Castillo-Ramírez, S. *et al.* Trans-Atlantic exchanges have shaped the population structure of the Lyme disease agent *Borrelia burgdorferi sensu stricto*. *Sci. Rep.* **6**, 22794; doi: 10.1038/srep22794 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>