**BorreliaBase**: a phylogeny-centered browser of *Borrelia* genomes

Lia Di¹, Pedro E Pagan¹, Daniel Packer², Che L Martin³, Saymon Akther¹, Girish Ramrattan¹, Emmanuel F Mongodin⁵, Claire M Fraser⁵, Steven E Schutzer⁶, Benjamin J Luft⁷, Sherwood R Casjens⁸ and Wei-Gang Qiu¹,³,⁴,⁹*

**Abstract**

**Background:** The bacterial genus *Borrelia* (phylum Spirochaetes) consists of two groups of pathogens represented respectively by *B. burgdorferi*, the agent of Lyme borreliosis, and *B. hermsii*, the agent of tick-borne relapsing fever. The number of publicly available *Borrelia* genomic sequences is growing rapidly with the discovery and sequencing of *Borrelia* strains worldwide. There is however a lack of dedicated online databases to facilitate comparative analyses of *Borrelia* genomes.

**Description:** We have developed *BorreliaBase*, an online database for comparative browsing of *Borrelia* genomes. The database is currently populated with sequences from 35 genomes of eight Lyme-borreliosis (LB) group *Borrelia* species and 7 Relapsing-fever (RF) group *Borrelia* species. Distinct from genome repositories and aggregator databases, *BorreliaBase* serves manually curated comparative-genomic data including genome-based phylogeny, genome synteny, and sequence alignments of orthologous genes and intergenic spacers.

**Conclusions:** With a genome phylogeny at its center, *BorreliaBase* allows online identification of hypervariable lipoprotein genes, potential regulatory elements, and recombination footprints by providing evolution-based expectations of sequence variability at each genomic locus. The phylo-centric design of *BorreliaBase* (http://borreliabase.org) is a novel model for interactive browsing and comparative analysis of bacterial genomes online.

**Keywords:** Lyme disease, Vector-borne relapsing fever, Genome browser, Recombination, Population genomics

**Background**

Spirochetes belonging to the genus *Borrelia* are obligate, tick- or louse-borne parasites of vertebrates and cause zoonotic infections in humans worldwide [1]. Phylogeny based on multiple genomic loci has confirmed the presence of two major *Borrelia* lineages, one represented by *B. burgdorferi*, the agent of Lyme borreliosis (LB), and the other by *B. hermsii*, the agent of tick-borne relapsing fever (RF) [2]. Members of the *Borrelia* appear to have the most complex genomic architecture among known prokaryotes. Sequencing of the first *Borrelia* genome revealed a 910-kilobase linear chromosome and twenty-one linear and circular plasmids with a total length of 610 kilobases in *B. burgdorferi* B31, a tick isolate from New York, USA [3,4]. Subsequent sequencing efforts showed that the linear chromosome is conserved and co-linear between the LB and RF groups [5-13]. A group of 30-kilobase circular plasmids appear to be conserved and syntenic across the entire genus [14]. A large ~160-kilobase linear plasmid specific to the RF group shares considerable syntenic regions with the lp54 plasmids of the LB group due to lateral transfer [15]. The linear plasmid lp23 in two RF group genomes is syntenic to the cp26 plasmids in LB group genomes [6]. Within the *B. burgdorferi* sensu lato species group (the LB group), a linear plasmid lp54 and a circular plasmid cp26 are universally present and largely syntenic among all species while the presence and the gene order of other plasmids are much less conserved [16,17]. Nevertheless, the total gene complement is quite stable within the LB group *Borrelia* [16,18].

* Correspondence: weigang@genectr.hunter.cuny.edu

¹Department of Biological Sciences, Hunter College, The City University of New York, 10065 New York, NY, USA

³Department of Biology, The Graduate Center, City University of New York, 10016 New York, USA

Full list of author information is available at the end of the article

© 2014 Di et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
The number of sequenced *Borrelia* genomes is rising rapidly. At the time of writing, the Bacterial Bioinformatics Resource Center (the PATRIC database, http://www.patricbrc.org) lists thirty-two completed and draft *Borrelia* genomes [19]. These genomic sequences represent twenty-eight LB *Borrelia* strains and six RF *Borrelia* strains, covering eight recognized or proposed species of the LB Group including *B. burgdorferi* sensu stricto, *B. afzelii*, *B. garinii*, *B. bavariensis*, *B. spielmanii*, *B. valaisiana*, *B. bissettii*, and *B. finlandensis*, and five species of the RF group including *B. recurrentis*, *B. duttonii*, *B. crocidurae*, *B. hermsii*, and *B. turicatae*. The chromosome of *B. miyamotoi*, an RF species, has also recently been sequenced [7]. A search on the GOLD genome-project registry (http://genomesonline.org/) yielded 47 completed, in-progress, and targeted *Borrelia* genome projects [20].

The rapid increase of sequenced *Borrelia* genomes allows for understanding *Borrelia* genome functions through comparative analysis, such as identification of host-interacting genes, regulatory elements, and pathways that are conserved among all species as well as species- and strain-specific genomic variations [21]. The comparative analysis of *Borrelia* genomes is presently hindered by a lack of dedicated online portal for their archival, dissemination, and comparative analysis. The Spirochete Genome Browser (SGD, http://sgb.fli-leibniz.de) currently provides web-based tools for analyzing the genomes of four strains of LB-group *Borrelia* (PKo, ZS7, B31, and PBi), five strains of RF-group *Borrelia* (DAH, A1, 91E135, and Ly), three species of *Leptospira* (*L. biflexa*, *L. borgpetersenii*, and *L. interrogans*), and two species of *Treponema* (*T. denticola* and *T. pallidum*). Web portals devoted to the comparative analysis of a single bacterial genus or species have been developed. A notable example is the *Pseudomonas* Genome Database, which facilitates population and evolutionary analysis of *Pseudomonas* genomes by providing up-to-date, computer-predicted orthologous gene sequences at both within- and between-species levels [22]. GenoList, a comprehensive database of over 700 completed bacterial genomes, provides tools and information for comparative proteomics, such as identification of proteins specific to a bacterial group [23]. At present, GenoList hosts genomic and proteomic information of three major *Borrelia* species causing Lyme disease (*B. burgdorferi* sensu stricto, *B. garinii*, and *B. afzelii*) and three *Borrelia* species causing relapsing fever (*B. recurrentis*, *B. duttonii*, and *B. hermsii*). The PATRIC database hosts a more comprehensive list of genomes of pathogenic bacterial genera including 34 genomes of *Borrelia* [19]. Comparative genomic information is also available from OrthologueDB, another comprehensive database of orthologous genes, which uses a rigorous, phylogeny-based algorithm to improve predictions of orthology in completed bacterial and archaeal genomes [24]. To the best of our knowledge, at present there is no genome database dedicated specifically to *Borrelia* or to providing comparative information such as sequences of orthologous ORFs in this genus.

We have recently released complete and draft genome sequences of twenty-one additional LB Group strains [10-13]. Population genomic analysis revealed genomewide recombination among co-existing strains and adaptive genome diversification driven by antigenic variations within local *Borrelia* populations [17,25]. Despite large variations in the plasmid composition and the gene order on the majority of plasmids, the total gene content is remarkably stable among LB-group genomes, suggesting that species divergence is driven mainly by copy-number and adaptive sequence variations of surface lipoprotein genes [16]. Here we describe the design, content, and usage of *BorreliaBase*, a database developed during the course of the aforementioned studies. *BorreliaBase* is not only the first publicly available database dedicated to *Borrelia* genomes, it also implements a novel graphic user-interface design that encourages comparisons of bacterial genomes under a framework of their shared phylogenetic history.

**Construction and content**

**Borrelia genomes**

*BorreliaBase* is currently populated with sequences of 35 genomes representing eight LB-group *Borrelia* species, including *B. burgdorferi* sensu stricto, *B. garinii*, *B. afzelii*, *B. bavariensis*, *B. spielmanii*, *B. valaisiana*, *B. bissettii*, and *B. finlandensis*, and seven RF-group *Borrelia* species, including *B. miyamotoi*, *B. hermsii*, *B. parkeri*, *B. turicatae*, *B. crocidurae*, *B. recurrentis*, and *B. duttonii*. *B. burgdorferi* sensu stricto, a major pathogenic species causing Lyme disease in North America and Europe, is represented by genomes of eleven North American and two European strains. Table 1 lists sequenced *Borrelia* genomes that are presently in the *BorreliaBase*.

While *BorreliaBase* includes basic genome annotations such as plasmid identity, ORF descriptions, and graphic maps of ORF locations, single-genome information is not the focus of *BorreliaBase*, considering that such information is available in primary or aggregate databases such as the NCBI Microbial Genomes website (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html), the UCSC Microbial Genome Browser (http://microbes.ucsc.edu/), and the PATRIC database (http://patricbrc.org). Instead, *BorreliaBase* focuses on providing high-quality comparative information for sequenced *Borrelia* genomes including genome phylogeny, genome synteny, and orthologous ORFs and intergenic spacers (IGSs) (Table 2).
Orthologous ORFs and IGSs

At its current state, the Borreliabase contains sequences of 750, 26, and 61 orthologous ORF families on the main chromosome, the cp26 plasmid, and the lp54 plasmid, respectively [17], as well as sequences of 541, 17, and 29 sets of unpublished orthologous IGSs on the main chromosome, the cp26 plasmid, and the lp54 plasmid, respectively. Methods on the identification of orthologous ORFs on the main chromosome, lp54, and cp26 have been described previously [17]. Briefly, the ORF orthology was determined with a combination of sequence-similarity search using BLAST [30], clustering of homologous
proteins using MCL [31], and manual validation based on genome synteny. Subsequently, we identified orthologous IGS sequences as those sharing a pair of flanking orthologous ORFs. To ensure the quality of IGS sequences in individual genomes, a consensus start-codon position was identified for each orthologous ORF family with the assumption that start-codon position and N-terminus sequences are conserved within as well as between species. As a result, many ORF sequences have been manually adjusted and their N-terminus sequences may differ from those in NCBI and other genome repositories. Plans have been made to reconcile ORF coordinates between BorreliaBase and the NCBI genome repository.

**Genome-based phylogeny**

Currently, comparative data on BorreliaBase include a statistically robust genome phylogeny. First, 266 sets of single-copy ortholog ORF families present in all 35 Borrelia genomes were identified from the ortholog database (Figure 1). Subsequently, each ORF family was translated into protein sequences, which were then aligned with MUSCLE (v3.7) [32]. Codon alignments corresponding to aligned protein sequences were obtained and then concatenated using customized PERL scripts based on BioPerl [33]. A maximum likelihood tree was obtained by using RAxML with three partitions of nucleotide substitution rates corresponding to the three codon positions [34]. Most of these orthologs are chromosome-borne. The phylogenetic position of the strain 297, the main chromosome of which remains to be sequenced, was determined and manually added to the RAxML tree based on phylogenies previously derived using plasmid-borne SNPs [18].

**Sequence alignments**

In addition to serving the raw orthologous sequences, BorreliaBase provides alignments of these sequences to allow direct interactive exploration of sequence variability (see Utility and Discussion). For orthologous IGS sequences, we provide online visualization of nucleotide alignments produced by MUSCLE (v3.7) with individual sequences ordered according to the genome phylogeny (see Use Case 3 below). For each set of orthologous ORF sequences, we provide MUSCLE-aligned amino-acid sequences as well as the corresponding codon alignment based on the amino-acid alignment (see Use Case 2 below).

**Relational database and web interface**

BorreliaBase is supported by a database in the back end with five main relational entities (Figure 1). The database is implemented with the PostgreSQL database management system (server version 9.1.9, http://www.postgresql.org). A dynamic web front was developed using the Asynchronous HTTP Request (AJAX) technology. On the server side, customized Perl-based Common Gateway Interface (CGI) scripts receive user inputs, fetch sequences from the back-end database, perform alignment if necessary, and output requested sequences and alignments as JavaScript Object Notation (JSON) objects. On the client side, customized JavaScript functions built with the jQuery (http://jquery.com) and D3 (http://d3js.org) libraries render these JSON objects into dynamic HTML contents.

**Utility and discussion**

With BorreliaBase, we aim to facilitate comparative analysis of Borrelia genomes in a phylogenetic framework. For
example, users can download manually curated orthologous ORF and IGS sequences or directly visualize sequence variations online. We do not duplicate existing single-genome features already provided by general-purpose microbial genome browsers such as the NCBI Microbial Genomes and the Integrated Microbial Genomes (IMG) databases. These databases provide comprehensive genome annotations including genome maps, plasmid identities, and ORF descriptions. In the following, we describe four examples to illustrate how BorreliaBase may be productively used for comparative exploration of Borrelia genomes.

**Use case 1: synteny analysis- identify gene gains and losses**

Synteny analysis compares co-linearity of genes on a replicon to identify evolutionary gains and losses of genes in genomes. A reliable synteny analysis requires the identification of a set of orthologous genes across multiple genomes as well as a phylogeny of these genomes. In this example, it has been identified that orthologs of *ospB* is absent in the genomes of two *B. garinii* strains PBr and Far04 (Figure 2). By displaying gene synteny in the context of the genome phylogeny, BorreliaBase allows us to conclude based on the principle of maximal parsimony that there is a single, lineage-specific deletion of *ospB* on the branch leading to *B. garinii*. This evolutionary history is consistent with the function of *ospB* being a gene expressed primarily in a mammalian environment [35] and with the fact that *B. garinii* uses birds as primary reservoir hosts [36]. Since the synteny view on BorreliaBase serves dynamic rather than static images, one can obtain customized synteny views centered on a specific gene (*ospA* in this case, Figure 2). A further customization is that one may select or hide genomes by clicking on the branches or tips of the tree.

**Use case 2: phylogenomics- visualize synonymous and nonsynonymous variations**

Lipoproteins, especially those localized to the outer cell membrane, are primary candidates for the development
of diagnostics for and vaccines against *Borrelia* infection, since they play key roles in establishing infections in arthropod vectors and vertebrate hosts [37,38]. For example, *B. burgdorferi* undergoes phase changes in its outer surface protein composition through the expression of *ospA* in the tick phase, the expression of *ospC* during host invasion, and the expression of *vls* genes in the mammalian phase [39]. Genes encoding outer surface lipoproteins tend to show elevated sequence variability within species, between species, or both within and between species [16,40]. As such, sequence variability is a reliable guide for identifying *Borrelia* proteins and protein regions that interact directly with the host.

A frequent request *Borrelia* researchers make to us is the identification of sequences orthologous to a gene of interest in all sequenced genomes. *BorreliaBase* fulfills this need by providing online access to orthologous ORF sequences.

A user begins by searching for a gene of interest using either a locus name (e.g., a15, the locus name for *ospA* in strain B31, the reference genome for LB group *Borrelia*) or a gene name (e.g., *ospA*). If the gene exists, a synteny view centered on the requested ORF appears showing its genomic context (Figure 2). At the same time, both the amino-acid alignment and the corresponding codon alignment of its orthologs are displayed (Figure 3). The integration of codon and amino-acid alignments allows easy determination of whether a nucleotide substitution is synonymous or non-synonymous (Figure 3). With the codon-alignment view, a user is able to visually scan for genes or regions that are potentially host-interacting based on a high density of non-synonymous substitutions before conducting a formal nonsynonymous-to-synonymous (*Ks/Ka*) ratio test [41]. Conversely, a user may identify gene regions with both low rates of synonymous and nonsynonymous substitutions between the *Borrelia* species. For example, using the codon-alignment view of *BorreliaBase*, one can visualize that the 5′ leader sequences of *ospC* is highly conserved across the eight LB-group *Borrelia* species in both synonymous and nonsynonymous nucleotide sites (Figure 3). Conservation of synonymous sites at the 5′ regions of a gene is common in bacteria and is typically associated with a preference for rare codons and a high level of gene expression [42].

A user may download aligned and unaligned DNA and protein sequences of a set of orthologous ORFs. Further, users may select or de-select strains using the genome phylogeny.

### Use case 3: phylogenetic footprinting- identify regulatory sequences

For each set of orthologous IGS sequences, a user can browse a nucleotide alignment to identify potential *cis*-regulatory sequences by looking for nucleotide sites conserved among the *Borrelia* species. Functional elements embedded in IGSs, e.g., *cis*-regulatory sequences and sequences encoding small RNAs, tend to be conserved among closely related bacterial species, allowing identification of such elements through comparative genomics with an approach called phylogenetic footprinting [43-45]. Using *BorreliaBase*, for example, we
Figure 3 Use case 2. Identify synonymous and nonsynonymous variations. Integrated visualization of codon (top panel) and amino-acid (bottom panel) alignments, with individual sequences arranged according to the genome phylogeny, allows identification of synonymous and nonsynonymous variations. This view also facilitates exploration of selective mechanisms underlying sequence variations. Here, three notable patterns of sequence variability are highlighted for the first 44 codons of *ospC*. Some sites vary synonymously within and between species (highlighted in blue), indicating strong purifying selection. Two sites vary non-synonymously within a single species (in orange), suggesting balancing selection. The first 21 sites (in red) are nearly constant not only in amino acids but also in codons, hinting strong selection for codon usage in this gene region.

Figure 4 Use case 3. Identify conserved regulatory sequence. In this alignment view of the promoter region of the *ospA*/*ospB* operon, a dot (•) represents a nucleotide or a gap that is the same as in strain SV1, the closest outgroup of *B. burgdorferi* sensu stricto and used as the reference. Known cis-regulatory elements within this promoter region are highly conserved across LB-group *Borrelia* species, including, from left to right, a direct repeat (DR, in purple) and the T-rich region (in yellow) associated with RpoS-mediated repression in mammalian environment, the putative RpoD-binding sites (−35 region in green and −10 region in blue), the transcription start site (+1, in pink), and the ribosome-binding site (RBS, in orange).
were able to show a high level of cross-species conservation of cis-regulatory sequences upstream of the ospAB operon, including the ribosomal-binding site, the transcription start position, the −35 and −10 RpoD (σ70)-binding sites, and a T-rich region required for ospA/B repression in mammalian hosts [46] (Figure 4).

**Use case 4: population genomics - identify recombination**

Homologous recombination in bacteria, which typically takes the form of gene conversion, facilitates adaptive evolution in bacteria [47,48]. Indeed, gene conversion underlines much of the hypervariability observed at loci encoding dominant antigens such as ospC in LB-group *Borrelia* and the rfb operon (coding for the O-antigen) and the fim locus (coding for an adhesion) in *Escherichia coli* [17,49]. We have previously established that geographically coexisting *Borrelia* strains recombine at a rate approximately three times the mutation rate [17,25]. Identification of homoplasy, a genetic variation that is inconsistent with the overall genome phylogeny, forms the basis of many tests of recombination such as the four-gamete test and the homoplasy test [50-52]. With sequence alignments displayed in the context of a genome phylogeny, a user of *BorreliaBase* can visually identify genomic regions affected by recombination and even estimate local recombination rates based on identification of homoplasies. In the first example, among strains of the same *Borrelia* species (*B. burgdorferi* sensu stricto), the presence of all four haplotypes at a pair of SNP sites is indicative of at least one past recombination event involving one of the SNP sites (Figure 5A). The second example shows that a portion of the b22 gene of the

![Figure 5](https://example.com/figure5.png)

**Figure 5 Use case 4.** Identify recombination. Recombination among bacterial genomes leaves genomic footprints that are identifiable by tests of homoplasy (i.e., phylogenetic inconsistency). (A) The “four-gametes” test. Two SNP sites within the b18-b19 IGS region are shown. All four possible haplotypes (labeled “1” through “4” and highlighted in yellow) are present in this region among the 14 *B. burgdorferi* sensu stricto genomes, suggesting at least one recombination event between these two SNP sites. (B) Cross-species gene conversion. A codon alignment of a region of b22 is shown. The *B. burgdorferi* sensu stricto strain BOL26 has a haplotype in this region that closely resembles its homologs in two *B. afzelii* strains (PKo and ACA-1), caused apparently by a replacement of the allele in BOL26 by an allele originated from a *B. afzelii* strain. (C) The “sister-group” test. Two SNP sites in a18 segregating between 156a and 297, two phylogenetic sister-group strains, are shown. The “C” at the first SNP site is a singleton while both alleles at the second SNP site have multiple copies. Since sister-group genomes are most closely related to each other, the first SNP suggests a recent point mutation (non-synonymous in this case) while the second SNP is likely to be introduced by recombination. The latter two analyses require a phylogenetic framework.
**Conclusions**

The phylogeny-centered design of **BorreliaBase** encourages comparing **Borrelia** genomes in the framework of their shared phylogenetic history. More than simply serving as a graphic guide for selecting and de-selecting strains, the genome phylogeny provides visual, evolution-based expectations on comparative genomic information such as an average level of sequence variability and the expected phylogenetic consistency of SNPs. When, and only when, an observed pattern deviates significantly from these evolutionary expectations should functional importance be conjectured. For example, the observed high levels of within-species sequence variability of lipo-protein genes such as *ospC*, *dpsA*, and *vls* deviate greatly from the expected sequence variability, suggesting strong diversifying selection at these loci driven by, e.g., immune escape from the host. Conversely, the observed high levels of sequence conservation of cis-regulatory elements deviate from the expected level of sequence variations between **Borrelia** species, consistent with gene-regulatory roles these IGS elements play (Figure 4). A phylogenetic expectation is especially valuable for identifying genomic footprints of recombination (Figure 5). Nexplorer, an online tool for analyzing evolution of characters (e.g., presence and absence of introns) associated with a protein family, similarly employs a phylogeny-centered user-interface design [54].

While allowing interactive comparative exploration of **Borrelia** genomes, the current version of **BorreliaBase** (public release 1.0) is admittedly limited in genomic scope. For example, comparative features, including gene synteny and orthologs, are only available for the three universally present replicons. We plan to include results of other replicons in future releases of **BorreliaBase**.

**Availability and requirements**

The **BorreliaBase** is publicly accessible online at http://BorreliaBase.org. The source code of the website is publicly available at http://sourceforge.net/projects/phylobrowse/.

**Abbreviations**

IGS: Intergenic Spacer; ORF: Open Reading Frame; **B. burgdorferi** s.l.: **Borrelia burgdorferi** sensu lato; **B. burgdorferi** s.s.: **Borrelia burgdorferi** sensu stricto; LB: Lyme borreliosis; RF: Relapsing fever.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

LD designed and implemented the web interface. PP and CM developed prototype websites for ORFs and IGSs respectively. PP matched GenBank Locus Tags to individual ORFs in the back-end genome database. PP and SA identified consensus start-codon positions for each orthologous ORF family. GR prepared figures and online documentation. DP configures and administrates the website. EM, CF, SE BL, SC, and WQ sequenced and annotated the majority of the genomes in the current **BorreliaBase**. SC identified plasmids in these genomes. WQ conceived of the project, developed the back-end database, and drafted the manuscript. All authors read and approved the final manuscript.

**Acknowledgements**

We thank two anonymous reviewers for constructive critiques. We thank Sabeel Kazi for preparing the Entity-Relation diagram. We thank Sviatoslav Kendall for assistance in literature search. We thank Patricia Rosa for bringing our attention to the codon-usage conservation in *ospC* leader sequences. This work was supported by Public Health Service grants AI07955 (to WQ); AI37256 (JL), AI49003 (SRC), and AI30071 (CMF and SES) from the National Institute of Allergy and Infectious Diseases (NIAID), and the grant MD007599 (Hunter College) from the National Institute on Minority Health and Health Disparities (NIMHD) of the National Institutes of Health (NIH). The content of this manuscript is solely the responsibility of the authors and do not necessarily represent the official views of NIAID, NIMHD, or NIH.

**Author details**

1. Department of Biological Sciences, Hunter College, The City University of New York, 10065 New York, NY, USA.
2. Department of Computer Science, Hunter College, The City University of New York, 10065 New York, NY, USA.
3. Department of Biology, The Graduate Center, City University of New York, 10016 New York, USA.
4. Center for Translational and Basic Research, Hunter College, The City University of New York, 10065 New York, NY, USA.
5. Institute for Genome Sciences, University of Maryland School of Medicine, 21201 Baltimore, MD, USA.
6. Department of Medicine, New Jersey Medical School, Rutgers, The State University of New Jersey, 07103 Newark, NJ, USA.
7. Department of Medicine, Health Science Center, Stony Brook University, 11794 Stony Brook, NY, USA.
8. Department of Pathology, Division of Molecular Cell Biology and Immunology, University of Utah School of Medicine, 84112 Salt Lake City, UT, USA.
9. Department of Biological Sciences, Hunter College of the City University of New York, 695 Park Avenue, 10065 New York, NY, USA.

Received: 13 February 2014 Accepted: 26 June 2014

**References**

1. Barboux AG, Haynes SF. *Biology of Borrelia* species. *Microbiol Rev* 1986, 50:381–400.
2. Wu D, Jospin G, Eisen JA: Systematic identification of gene families for use as “Markers” for phylogenetic and phylogeny-driven ecological studies of bacteria and archaea and their major subgroups. *PLoS One* 2013, 8:e77033.
3. Fraser CM, Caijens S, Huang WM, Sutton GG, Clayton R, Lathigra R, White O, Ketchum KA, Dodson R, Hickey EK, Gwinn M, Dougherty B, Tomb J-F,
Fleischmann RD, Richardson D, Peterson JK, Kerlavage AR, Quackenbush J, Salzberg S, Hanson M, van Vugt R, Palmer N, Adams MD, Gocayne J, Weidman J, Utterback T, Watthey L, McDonald L, Atriach P, Bowman C, et al. Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi. Nature 1997, 390:580–586.

4. Casjens S, Palmer N, Van Vugt R, Mun Huang W, Stevenson B, Rosa P, Lathigra R, Sutton G, Peterson J, Dodson RJ, Halldorson D, Hickey E, Gwinn M, White O, Fraser MC. A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNA in an infectious isolate of the Lyme disease spirochete Borrelia burgdorferi. Mol Microbiol 2000, 35:490–516.

5. Elbir H, Gimenez G, Robert C, Bergström S, Cutler S, Raoult D, Drancourt M: 2013, a tick-borne agent of deadly louse-borne relapsing fever, is a degraded subset of Borrelia sensu lato isolate.

6. Lescoat M, Audic S, Robert C, Nguyen TT, Blanc G, Cutler SJ, Wincker P, Pouloux A, Claverie J-M, Raoult D, Ducrocq M. The genome of Borrelia recurrentis, the agent of deadly louse-borne relapsing fever, is a degraded subset of tick-borne Borrelia duttonii. PLoS Genet 2008, 4:e1000185.

7. Huf N, Langergäu AG, Barbour AG: Chromosome sequence of Borrelia miyamotoi, an uncultivable tick-borne agent of human infection. Genome Announc 2013, 1:00713–13.

8. Glöckner G, Lehmann R, Romualdi A, Pratelle S, Schulte-Schuchter U, Schuhlab M, Wilke B, Suhnel J, Plater M: Comparative analysis of the Borrelia garinii genome. Nucleic Acids Res 2004, 32:e6.038–6046.

9. Glöckner G, Schuhlab-Schuchter U, Schuhlab M, Felder M, Suhnel J, Wilke B, Plater M: Comparative genome analysis: selection pressure on the Borrelia vls cassettes is essential for infectivity. BMC Genomics 2006, 7:211.

10. Casjens SR, Fraser-Liggett CM, Qiu W-G, Dunn JJ, Luft BJ, Schutzer SE: Whole-genome sequences of two Borrelia afzelii and two Borrelia garinii Lyme disease agent isolates. J Bacteriol 2011, 193:6959–6969.

11. Casjens SR, Fraser-Liggett CM, Qiu W-G, Dunn JJ, Luft BJ, Schutzer SE: Whole genome sequence of an unusual Borrelia burgdorferi sensu lato isolate. J Bacteriol 2011, 193:1489–1490.

12. Schutzer SE, Fraser-Liggett CM, Qiu W-G, Kraicz P, Munodgın EF, Dunn JJ, Luft BJ, Casjens SR: Whole-genome sequences of Borrelia bissetti, Borrelia holandica, and Borrelia spielmani. J Bacteriol 2012, 194:545–546.

13. Schutzer SE, Fraser-Liggett CM, Casjens SR, Qiu W-G, Dunn JJ, Munodgın EF, Luft BJ: Whole-genome sequences of thirteen isolates of Borrelia burgdorferi. J Bacteriol 2011, 193:1018–1020.

14. Stevenson B, Porcella SF, Oie KL, Fitzpatrick CA, Raffel SJ, Lubke L, Schrumpf E, Bocci OS, Pochett MC, Schrumpf E, Senger M, Stein LD, Stupka E, Wilson MJC, Yoo HS, Zhang C, Zhang Y, Sobral В, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Gough R, Hix D, Kenyon R, . Machi D, Mao C, Nordberg EK, Olson R, Machi D, Schober BW: PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res 2012, 40:D1581–D1591.

15. Miller SC, Porcella SF, Raffel SJ, Schwan TG, Barbour AG: Large linear plasmids of Borrelia species that cause relapsing fever. J Bacteriol 2013, 195:3629–3639.

16. Casjens SR, Munodgın EF, Qiu W-G, Luft BJ, Schutzer SE, Gilcrease EB, Huang WM, Vujadinovic M, Aron JR, Vargas LC, Freeman S, Radune D, Weidman ЈF, Dimitrov G, Khoury HM, Sosa JE, Halpin RA, Dunn JJ, Fraser CM: Genome stability of Lyme disease spirochetes: comparative genomics of Borrelia burgdorferi plasmids. PLoS One 2012, 7:e33280.

17. Haven J, Vargas LC, Munodgın EF, Xue V, Hernandez Y, Pagan P, Fraser-Liggett CM, Schutzer SE, Luft BJ, Casjens SR, Qiu W-G: Pervasive recombination and sympatric genome differentiation driven by frequency-dependent selection in Borrelia burgdorferi, the Lyme disease bacterium. Genetics 2011, 189:951–966.

18. Munodgın EF, Casjens SR, Bruno JF, Xu Y, Drabek EF, Riley DR, Canarell CR, Pagan PE, Hernandez YA, Vargas LC, Dunn JJ, Schutzer SE, Fraser CM, Qiu W-G, Luft BJ: Inter- and intra-specific pan-genomes of Borrelia burgdorferi sensu lato: genome stability and adaptive radiation. BMC Genomics 2013, 14:953.

19. Wattram AR, Abrahams D, Daiqy O, Drax TL, Driscoll T, Gabbard JL, Gillespie JI, Gough R, Hx D, Kenyon R, Mach D, Mao C, Nordberg EK, Olson R, Overbeck R, Puschn G, Shukla M, Schumann J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MIC, You HS, Zhang C, Zhang Y, Sobral BW: PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res 2014, 42(D1):DS81–DS91.

20. pagani I, Lolioc K, Janssen J, Chen I-HA, Smimova T, Nosrat B, Markowitz YM, Kyrides NC: The Genomes Online Database (GOLD) v. 4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res 2012, 40(Database issue):D571–D579.

21. Norris SR, Lin T: Out of the woods: the remarkable genomes of the genus Borrelia. J Bacteriol 2011, 193:6812–6814.

22. Winsor GL, Lam DKW, Fleming L, Lo R, Whiteside MD, Yu Y, Hancock REW, Brinkman FSL: Pseudomonas genome database: improved comparative analysis and population genomics capabilities for Pseudomonas genomes. Nucleic Acids Res 2011, 39(Database issue):D560–D566.

23. Lechat P, Hummel L, Rousseau S, Mosser I: GenoList: an integrated environment for comparative analysis of microbial genomes. Nucleic Acids Res 2008, 36(Database issue):D469–D474.

24. Whiteside MD, Winsor GL, Laird MR, Brinkman FSL: OrthologueDB: a bacterial and archaeal orthology resource for improved comparative genomics analysis. Nucleic Acids Res 2012, 41:2366–2376.

25. Qiu W-G, Schutzer SE, Bruno JF, Attine O, Xu Y, Dunn JJ, Fraser CM, Casjens SR, Luft BJ: Genetic exchange and plasmid transfers in Borrelia burgdorferi sensu stricto revealed by three-way genome comparisons and multilocus sequence typing. Proc Natl Acad Sci U S A 2004, 101:H1450–1455.

26. Brenner EV, Kurthshok AM, Strømøn CN, Fomemo NK: Whole-genome sequencing of Borrelia burgdorferi Bygfr, isolated from taiga ticks (Ixodes persulcatus). J Bacteriol 2012, 194:5713.

27. Jiang B, Yao H, Tong Y, Yang X, Huang Y, Jiang J, Cao W: Genome sequence of Borrelia garinii strain NMJ11, isolated from China. J Bacteriol 2012, 194:6660–6661.

28. Jiang B-G, Zheng Y-C, Tong Y-G, Jia N, Huo Q-B, Fan H, Ni X-B, Ma L, Yang X, Jiang J-F, Cao WC: Genome sequence of Borrelia afzelii strain HLJ01, isolated from a patient in China. J Bacteriol 2012, 194:7014–7015.

29. Barbour AG, Campeau Miller S: Genome Sequence of Borrelia parkeri, an agent of enzootic relapsing fever in Western North America. Genome Announc 2014, 2(11): doi:10.1128/genomeA.00118-14.

30. Camacho C, Louriul G, Avagyan V, Ma L, Papadopoulos J, Bealer K, Madden TL: BLAST+: architecture and applications. BMC Bioinformatics 2009, 10:42.

31. Enright AJ, Van Dongen S, Ouzounis CA: An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 2002, 30:1575–1584.

32. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004, 32:1792–1797.

33. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdelen C, Fuellen G, Gilbert JGR, Korf I, Lapp H, Lethavia S, Matsalla C, Mungall CJ, Osborne BR, Pochett MR, Schattler P, Senger M, Stein LD, Stupka E, Wilson MJC, Yoo HS, Zhang C, Zhang Y, Sobral В, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Gough R, Hix D, Kenyon R, . Machi D, Schober BW: PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res 2012, 40:D1581–D1591.

34. Siddiqui AR, Alam N, Gokal MC, Sahu JP: Comparative genomics analysis of Borrelia burgdorferi. Prog Biophys Mol Biol 2010, 100:9–47.
45. Degnan PH, Ochman H, Moran NA: Sequence conservation and functional constraint on intergenic spacers in reduced genomes of the obligate symbiont Buchnera. PLoS Genet 2011, 7:e1002252.
46. Xu Q, McShan K, Liang FT: Two regulatory elements required for enhancing ospA expression in Borrelia burgdorferi grown in vitro but repressing its expression during mammalian infection. Microbiology 2010, 156:2194–2204.
47. Didelot X, Maiden MCJ: Impact of recombination on bacterial evolution. Trends Microbiol 2010, 18:315–322.
48. Fraser C, Hanage WP, Spratt BG: Recombination and the nature of bacterial speciation. Science 2007, 315:476–480.
49. Touchon M, Hoede C, Tenillon O, Barbe V, Baenishyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, Calteau A, Chiapello H, Cleemont Q, Cruveiller S, Danchin A, Diard M, Dossat C, Karoul ME, Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Le Bouguénec C, Lesca P, Mangenot S, Martinez-Jéhanne V, Matic I, Nassif X, Oztas S, et al. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet 2009, 5:e1000344.
50. Hudson RR, Kaplan NL: Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 1985, 111:147–164.
51. McVean G, Awadalla P, Fearnhead P: A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics 2002, 160:1231–1241.
52. Smith JM, Smith NH: Detecting recombination from gene trees. Mol Biol Evol 1998, 15:590–599.
53. Guttmann DS, Dykhuizen DE: Clonal divergence in Escherichia coli as a result of recombination, not mutation. Science 1994, 266:1380–1383.
54. Gopalan V, Qiu W-G, Chen MZ, Stoltzfus A: Nexplorer: phylogeny-based exploration of sequence family data. Bioinforma Oxf Engl 2006, 22:120–121.

doi:10.1186/1471-2105-15-233
Cite this article as: Di et al. BorreliaBase: a phylogeny-centered browser of Borrelia genomes. BMC Bioinformatics 2014 15:233.