# Borrelia

-The genus Borrelia forms a deeply separated lineage within the Spirochaetes branch of the bacterial tree
-are genomically unique and not closely related to any other bacteria, including the other Spirochaetes.
-Borrelia burgdorferi and Borrelia hermsii are representative of the two major branches within the Borrelia genus
-B. burgdorferi typifies the Lyme disease Borrelia branch (LB)
-B. hermsii typifies the relapsing fever agent branch

-Each LB species carries a linear chromosome about 900 kbp in length
-and multiple circular and linear low copy number plasmids
    -5-84 kbp range.

**”Large” or “Main” chromosome**

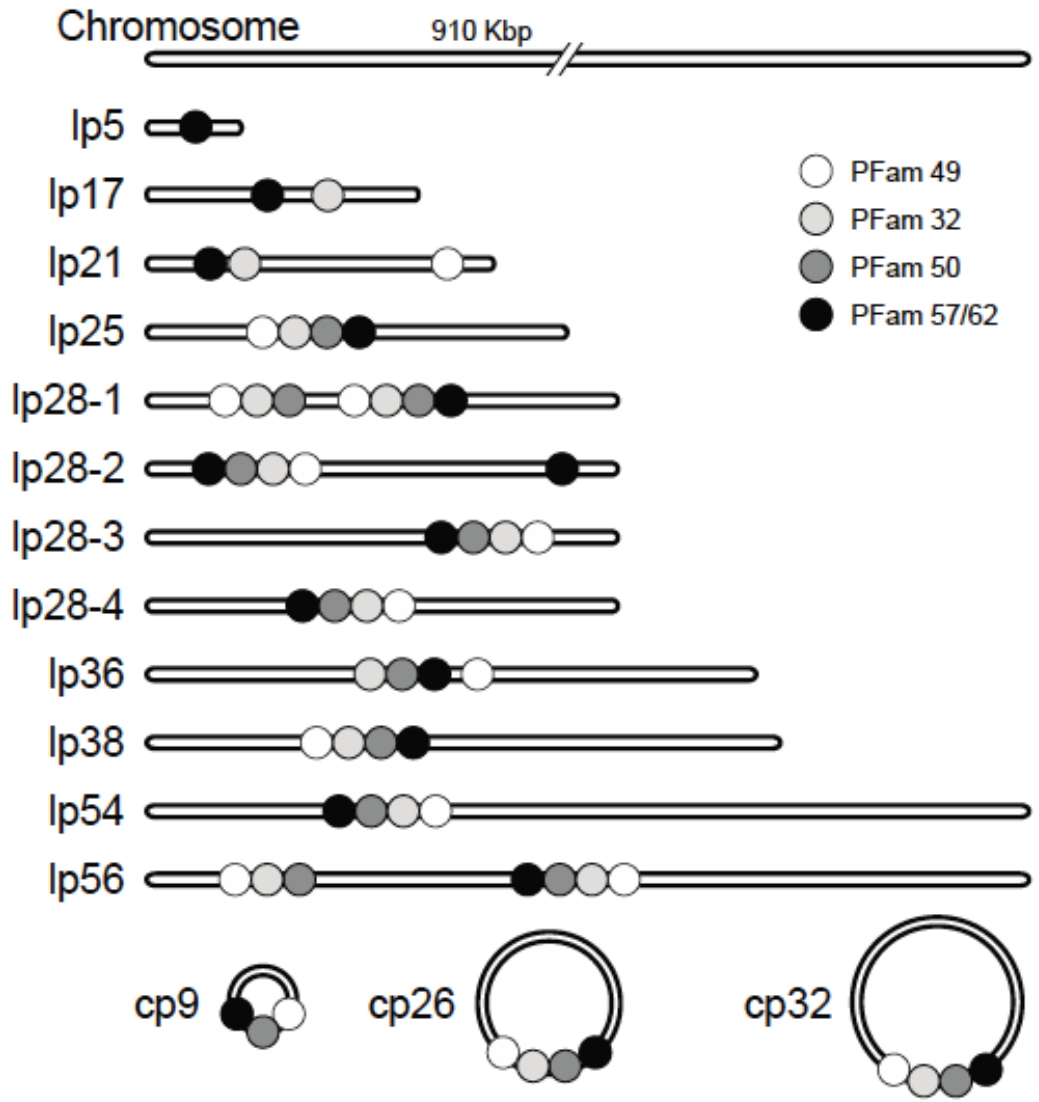
-900 kbp
-carry great majority of the genes that encode metabolic enzymes/ housekeeping genes
-a few surface-exposed proteins encoded
-tightly packed with genes
-evolutionarily stable
-five rRNA genes (two 23S, two 5S and one 16S)
    -clustered very near the center of the chromosome
-32 tRNAs
    -scattered throughout in 7 clusters and 13 single genes
-Variable numbers of tandem repeats were found in three genes: BB_0210, BB_0546 and BB_0801
-lacks or has limited genes that encode for:
    -respiration
    - amino acid synthesis
    - nucleotide synthesis
    -lipid synthesis
    - enzyme cofactor synthesis
-different LB species (27) large chromosome varies in size, ranging from
    -900,694 kbp in B. chilensis
    -922,901 in B. burgdorferi JD1
-right end extensions in B. burgdorferi , B. mayonii and B. valaisiana, B. turdi
    -plasmid fusion

**Borrelia burgdorferi B31 Chromosome**

-type strain
-910,725 bp
-28.6% GC
-803 protein coding genes
-17 pseudogenes

# Plasmids

-members of the Borrelia genus carry more plasmids than any other bacteria.
    -LB isolates have between 7 and 21 different plasmids
-linear and circular
-5 to 84 kbp in size
-except for cp26 are not required for growth in culture
-carry the bulk of surface lipoprotein encoding genes
-have often be found to be necessary for mouse infectivity or tick transmission
-B. burgdorferi plasmids were originally named in the type strain B31
    -“lp” for linear plasmid and “cp” for circular plasmid
    -numerical designations of their approximate kbp sizes
-all of the B31 plasmids larger than 15 kbp carry a set of four genes, usually in a contiguous cluster, that are predicted to be involved in plasmid replication and maintenance.
    -PFam32, PFam49, PFam50 and PFam57/62
    -Smaller plasmids lp5, lp21 and cp9 often lack one or more of these four gene types but always carry a PFam57/62 gene
-PFam32 proteins are quite diverse and can be robustly parsed into different sequence types
    -no LB cell has been found that carries two plasmids that encode PFam32 proteins of the same sequence type




Different B31 cultures have been found to carry one or two related cp9 plasmids and up to eight different cp32 plasmids (prophages).

| Species | Isolate (X[1]) | cp9 | cp26 | cp32[2] | lp5 | lp17 | linears[3] | lp54 | Total[4] | Reference |
|---|---|---|---|---|---|---|---|---|---|---|
| *B. burgdorferi* | | | | | | | | | | |
| | 64b (A) | - | 1 | 9[8] | - | 1 | 7 | 1 | 18 | Casjens, 2017 |
| | B31 (A) | 1[5] | 1 | 8[6] | 1 | 1 | 10 | 1 | 21[2,5] | Casjens, 2000 |
| | Bol26 (A) | - | 1 | 6[7,8] | - | 1 | 4 | 1 | 13 | Casjens, 2017 |
| | PAli (A[1]) | - | 1 | 7[6] | - | 1 | 4 | 1 | 13 | Margos, 2017 |
| | PAbe (A[1]) | - | 1 | 7[6] | - | 1 | 4[9] | 1 | 11 | Margos, 2017 |
| | ZS7(A) | - | 1 | 6[8] | - | 1 | 6 | 1 | 14 | Casjens, 2017 |
| | 29805 (B) | - | 1 | 5[8] | - | 1 | 7 | 1 | 15 | Casjens, 2017 |
| | N40 (B) | 1 | 1 | 6[8] | - | 1 | 6 | 1 | 16 | Casjens, 2017 |
| | WI91-23 (B) | 2 | 1 | 7[8] | 1 | 1 | 9 | 1 | 21 | Casjens, 2017 |
| | 72a (C) | - | 1 | 6[8] | - | 1 | 4 | 1 | 13 | Casjens, 2017 |
| | 94a (C) | 1 | 1 | 5 | - | 1 | 5[7] | 1 | 14 | Casjens, 2017 |
| | 118a (C) | 1 | 1 | 8[8] | - | 1 | 8 | 1 | 19 | Casjens, 2017 |
| | CA11_2a (C) | - | 1 | 5 | - | 1 | 4[10] | 1 | 12 | Casjens, 2017 |
| | MM1 (C[1]) | 1 | 1 | 6 | - | 1 | 5 | 1 | 15 | Jabbari, 2018 |
| | 156a (D) | - | 1 | 8[8] | - | 1 | 9 | 1 | 20 | Casjens, 2017 |
| | 297 (D[11]) | - | 1 | 9[8] | - | 1 | 7 | 1 | 19 | Casjens, 2017 |
| | JD1 (D) | - | 1 | 9[8] | - | 1 | 9 | 1 | 20 | Casjens, 2017 |

[2]Number of cp32 PFam32 compatibility types, not DNA molecules. Strain B31 may have had 9th cp32 (cp32-5) that was lost before genome sequencing (Zuckert and Meyer, 1996; Casjens et al., 1997).

[8]One or more cp32s either truncated or fused with other cp32s.

**Typing Methods**

4 main methods

-ribosomal spacer
-ospC
-MLST
-WGS

-except for OspC, these methods rely largely on chromosomal loci for typing
-discriminatory power varies among different molecular typing methods
- whole genome-based approaches such as PFGE, RAPD, MLST and whole genome sequencing (WGS) have the highest resolution
- ospC typing was frequently used to differentiate strains within the same species, in particular in B. burgdorferi, and for disease association studies
- MLST is better suited than ospC for differentiation of disease associated sequence types
- Most PCR based methods can be used for molecular typing of B. burgdorferi directly in a variety of field-collected and clinical specimens
    - mixed infections may not be disentangled
- Some molecular typing methods (i.e., ribotyping, PFGE, RAPD) require pure cultured spirochetes
- Selection of particular genotypes or strains of B. burgdorferi during in vitro cultivation has been demonstrated

**Typing continued**

-Ribosomal spacer method

-2 different loci

-PCR then RFLP

-3 ribosomal spacer types – RST1,RST2,RST3

-correlate with pathogenic potential

ospC typing (outer surface protein C)

- OspC is essential for B. burgdorferi to establish a productive initial infection in mammals
- PCR a 617 bp region of ospC then Sanger sequencing
- ospC is one of the most highly polymorphic genes in LB species

-approximately 25 ospC major groups have been identified in B. burgdorferi populations

MLST (multi-locus sequence typing)
-several schemes developed for LB
-University of Oxford at https://pubmlst.org/borrelia
-targeted sequencing of 8 housekeeping loci
- -comparison of sequence fragments to MLST database
-single point mutation differences
-allele numbers for each locus
-chain of 8 integers corresponds to allelic profile of isolate
-defines the sequence type
-as of May 2020 >3100 isolates and > 900 STs

| Isolate fields | | | | | | | MLST | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| id | isolate | aliases | country | species | year | source | clpA | clpX | nifS | pepX | pyrG | recG | rplB | uvrA | ST |
| 2776 | JD1_wgs | | USA | Borrelia burgdorferi sensu stricto | | tick | 5 | 7 | 5 | 1 | 6 | 1 | 4 | 9 | 11 |
| 2778 | PAbe_wgs | | Germany | Borrelia burgdorferi sensu stricto | 1997 | human | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2779 | PAli_wgs | | Germany | Borrelia burgdorferi sensu stricto | 1994 | human | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2780 | MM1_wgs | | USA | Borrelia burgdorferi sensu stricto | 1983 | animal host | 7 | 6 | 236 | 265 | 1 | 5 | 5 | 5 | 928 |
| 2781 | ZS7_wgs | | Germany | Borrelia burgdorferi sensu stricto | | tick | 14 | 1 | 11 | 1 | 1 | 1 | 1 | 10 | 20 |
| 2782 | B331_wgs | | USA | Borrelia burgdorferi sensu stricto | | human | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 16 |
| 2783 | N40_wgs | | USA | Borrelia burgdorferi sensu stricto | | tick | 4 | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 19 |
| 2784 | FDAARGOS_196_wgs | | USA | Borrelia burgdorferi sensu stricto | 1981 | tick | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2785 | 156a_wgs | | USA | Borrelia burgdorferi sensu stricto | | human | 8 | 1 | 1 | 1 | 4 | 6 | 1 | 7 | 4 |
| 2787 | Bol26_wgs | | Italy | Borrelia burgdorferi sensu stricto | | human | 14 | 1 | 11 | 119 | 1 | 1 | 1 | 10 | 332 |
| 2788 | 72a_wgs | | USA | Borrelia burgdorferi sensu stricto | | human | 9 | 1 | 1 | 7 | 1 | 6 | 1 | 10 | 14 |
| 2789 | 94a_wgs | | USA | Borrelia burgdorferi sensu stricto | | human | 7 | 6 | 6 | 1 | 1 | 5 | 5 | 5 | 18 |
| 2790 | 64b_wgs | | USA | Borrelia burgdorferi sensu stricto | | human | 6 | 1 | 5 | 1 | 1 | 7 | 1 | 19 | 59 |
| 2791 | 118a_wgs | | USA | Borrelia burgdorferi sensu stricto | | human | 8 | 1 | 1 | 7 | 1 | 6 | 1 | 10 | 34 |
| 2792 | CA11.2A_wgs | | USA | Borrelia burgdorferi sensu stricto | | tick | 8 | 1 | | 120 | 1 | 18 | 1 | 7 | |
| 2793 | WI91-23_wgs | | USA | Borrelia burgdorferi sensu stricto | | animal host | 12 | 2 | 1 | 2 | 1 | 16 | 2 | 2 | 228 |
| 2794 | 29805_wgs | | USA | Borrelia burgdorferi sensu stricto | | tick | 3 | 3 | 2 | 4 | 3 | 4 | 4 | 4 | 12 |
| 2795 | 80a_wgs | | Unknown | Borrelia burgdorferi sensu stricto | | unknown | | 5 | 4 | 6 | 1 | 6 | 1 | 6 | |

**Why are plasmids a problem for sequencing/typing?**

-sequenced BB31 has 21 plasmids (2000)
    -several additional plasmids lost between isolation and complete sequencing
-several plasmids have very similar sizes and cannot be separated by pulse-field gel
    -mapping genes by Southern hybridization difficult
-different plasmids in the same cell often carry very similar paralogous sequences
    -identification confusing
    -assembly difficult
-plasmids can be lost during primary isolation
-plasmids readily lost during passage or in vitro manipulation
-clones of any given strain will have lost one or more plasmids
-contain decaying pseudogenes
-examples of circular plasmids integrated into linear plasmids
    - cp32-10 integrated into lp56 in B. burgdorferi strain B31
-examples of end-to-end fusion of linear plasmids
    - lp36 and lp28-4 in B. burgdorferi strain CA-11.2A
-examples of fusion of two or more circular plasmids
    - cp32-1 and cp32-5 in B. burgdorferi strain JD1
-cp32
    - Most LB isolates carry one or more cp32 plasmids
    -relatively frequent homologous recombination shuffles sequence among the copies
    -nonhomologous organizational rearrangements

## Whole genome sequencing of *Borrelia miyamotoi* isolate Izh-4: reference for a complex bacterial genome

- Konstantin V. Kuleshov, et al *BMC Genomics* **volume 21**, Article number: 16 (2020)

-one linear large chromosome, 2 circular plasmids, 12 linear plasmids

-6 kb to 900 kb

-MiSeq

-nanopore

Canu v1.7

-after correction, 31,584 reads with an average length of 7.3 kb

-coverage pattern revealed two contigs with two peaks

-two separate plasmids merged

-sizes not consistent with PFGE

-coverage revealed 3 contigs with very low long read coverage (2-3x) and high similarity with other contigs

-circular plasmids “lost” during assembly

-pacBIO

-PFGE

-10 fragments

-ranging from 13 kb to 900 kb

-some of the bands contain two different plasmids

-used to reject contigs of anomalous sizes

-fused contigs split

-extracted DNA from bands and sequenced (MiSeq)

-generated lots of contigs

-used to identify plasmids (no reads for 2 circular plasmids)

From: Whole genome sequencing of *Borrelia miyamotoi* isolate Izh-4: reference for a complex bacterial genome

| | GenBank accession numbers | Molecule name | Length, bp | PacBio read coverage before and after correction in brackets | MinION read coverage before and after correction in brackets | Illumina reads coverage |
|---|---|---|---|---|---|---|
| 1 | CP024390.1 | chromosome | 906,129 | 695x (16x) | 668x (200x) | 440x |
| 2 | CP024391.1 | plzh4-lp72 | 73,492 | 1378x (24x) | 1118x (300x) | 402x |
| 3 | CP024392.1 | plzh4-lp70 | 70,072 | 677x (17x) | 573x (122x) | 359x |
| 4 | CP024401.2 | plzh4-lp64 | 64,141 | 321x (5x) | 365x (67x) | 262x |
| 5 | CP024393.1 | plzh4-lp41 | 41,127 | 509x (11x) | 447x (54x) | 523x |
| 6 | CP024395.1 | plzh4-cp30–1 | 30,091 | 1712x (26x) | 591x (162x) | 192x |
| 7 | CP040828.1 | plzh4-cp30–2 | 29,490 | 657x (13x) | 265x (49x) | 177x |
| 8 | CP024396.1 | plzh4-lp29 | 28,667 | 1211x (23x) | 544x (72x) | 614x |
| 9 | CP024397.1 | plzh4-lp23 | 27,717 | 528x (16x) | 329x (37x) | 504x |
| 10 | CP024398.1 | plzh4-lp27 | 26,599 | 862x (17x) | 334x (20x) | 251x |
| 11 | CP024399.2 | plzh4-lp24 | 24,033 | 1263x (18x) | 470x (78x) | 466x |
| 12 | CP024400.2 | plzh4-lp18–2 | 18,334 | 1554x (17x) | 323x (63x) | 722x |
| 13 | CP024405.2 | plzh4-lp18–1 | 18,024 | 771x (14x) | 123x (49x) | 527x |
| 14 | CP024404.1 | plzh4-lp13 | 13,410 | 480x (9x) | 118x (43x) | 327x |
| 15 | CP024407.1 | plzh4-lp6 | 5851 | 578x (3x) | 138x (92x) | 625x |
| | | | Total reads: | 312,224 (2625) | 129,992 (31,584) | 2,642,950 |
| | | | Mapped reads: | 95% (100%) | 100% (100%) | 93% |

Genome Canada/Genome Atlantic
EMERGING ISSUES and OPPORTUNITIES

# Lyme Disease in NS: The influence of strain variation on disease

2. **Objectives**:
This proposal has three main objectives that will directly test the hypotheses listed below:
1. To establish a cohort of patients within the vicinity of Luneburg, Nova Scotia, presenting at different stages of Bb infection in order to collect specimens with which to characterize the strains causing infection and to follow clinical outcomes over a two-year period.
2. To collect specimens for culture and molecular investigations in order to examine the biodiversity of the strains infecting humans and their pathogenic potential.
3. To collect Bb strains from ticks in the environment surrounding Lunenburg to compare the biodiversity in ticks with those that lead to infection in humans.

**Hypotheses**:
a. The diversity of Bb strains infecting humans is less that what is found in ticks collected from the environment
b. Canadian strains of Bb discriminated by WGS are associated with different clinical manifestations and differ from those seen in the US
c. The diversity of Canadian strains infecting humans is sufficiently different that they reduce the sensitivity of current serologic tests to identify infection

In order to test these hypotheses we will obtain representative Bb from host-seeking female adult and nymphal ticks (as these are the life stages that ultimately transmit Bb to humans) and from humans infected with Bb.

Tick and human samples (blood, skin biopsy, synovial fluids and/or CSF)

- -divide specimens in two
  - -one for culture
  - -one for molecular testing
- -cultures screened for mixed strain infections using PCR (on the MLST primer sets)
  - -pure cultures allowed to grow to optimum density
  - -mixed cultures
    - -subsurfacing plating
      - -individual colonies will be inoculated into liquid media and grown to max density

Molecular Testing

-DNA for WGS will be extracted from each culture

-Epicentre Masterpure Complete DNA and RNA Extraction Kit

-Multiplexed libraries will be created with TruSeq kits

-paired-end, 300 bp, indexed reads will be generated on MiSeq platform

-MLST using the standard 8 housekeeping loci

-Key loci on the accessory genome (such as ospC) will be extracted using standard scripting methods?

-WGS using PacBio and/or Nanopore

-selected strains from patients exhibiting different manifestations

-environmental strains not found in human infections

-better characterize genetic determinants of pathogenicity in the accessory genome

Immediate future

-test this procedure immediately on 5 genetically distinct Bb strains selected from 25 NS strains already in hand from previous field seasons?

Analysis

- -comprehensive bioinformatics analysis of strain-specific _
  - -gene content
  - -patterns of recombination
  - -evidence for purifying selection
  - -etc?

- -whether strain types and frequencies are changing over time
  - -Nine MLST sequence types were found in 30 questing nymphal ticks from NS (2018)
    - -30% ST3

**Problems/Issues**

-low GC%
-need long reads but also 5kb ones
-who is doing Illumina sequencing?
-coverage
    -nanopore
        -large chromosome
        -plasmids
    -illumina
-PFGE
    -yes/no and who
-bacterial specific protocols
    -library
    -sequencing
    -assembly
-base calling
-correction/polishing
-assembly
    -lose of circular plasmids

**Problems/Issues continued**

-bioinformatic analyses by whom?

-Dr. Jason Leblanc co-directs the serologic diagnostic laboratory at the QEII laboratory and has expertise in molecular diagnostics and interpretation of whole genome sequencing.