Borrelia Genomes in the Year 2000

Sherwood Casjens*

Department of Oncological Sciences, University of Utah Medical School, Salt Lake City, UT 84132 USA

Abstract

All analyzed members of the spirochete genus Borrelia contain a linear chromosome about 910 kbp long. The complete sequence of the B. burgdorferi B31 genome predicts that its chromosome carries essentially all of this organism's housekeeping genes. In accordance with these bacterial species' obligate parasitic lifestyle, its genes encode enzymes that are capable of only a minimal metabolism, in which all nucleotides, amino acids, fatty acids and enzyme cofactors must be scavenged from the host. In addition to the chromosome, all Borrelia isolates examined carry multiple linear and circular plasmids with lengths between 5 and 200 kbp. The plasmids, which account for over 600 kbp in isolate B31, carry very few genes with homology to genes outside of the Borrelia genus. But they do carry numerous predicted lipoprotein genes, many of which have been shown to be or are expected to be outer surface proteins. Ten of the linear plasmids have strikingly low protein coding potential for bacterial DNA. These plasmids have enjoyed numerous past duplicative rearrangements, which have resulted in the presence of a substantial fraction of the DNA that appears to be currently undergoing mutational decay, presumably because it is no longer under selection for function.

Borrelia Genome Structure and Complexity

The bacterial spirochete genus Borrelia forms a well-defined phylogenetic clade which has no very close relatives, and their genomes are unique in several respects. Barbour and coworkers (e.g., Barbour and Garon, 1987; Barbour, 1988) originally found that Borrelia burgdorferi and Borrelia hermsii both carry multiple linear plasmids, and soon thereafter multiple circular plasmids were also found (e.g., Hyde and Johnson, 1988; Schwan et al., 1988; Simpson et al., 1990a). These plasmids have all been found to be of low copy number when carefully analyzed (Hinnebusch and Barbour, 1992; Kitten and Barbour, 1992; Casjens and Huang, 1993). The multiple extrachromosomal DNAs in Borrelia cells was unusual, but the presence of linear replicons was not known in other bacteria at that time. Barbour, St. Girons and their colleagues soon obtained evidence that the B. burgdorferi chromosome was also a linear DNA molecule (Baril et al., 1989; Ferdows and Barbour, 1989), and subsequent studies confirmed that apparently all members of this genus do indeed harbor linear chromosomes (Davidson et al., 1992; Casjens and Huang, 1993; Ojaimi et al., 1994; Casjens et al., 1995; van Vugt et al., 2000). A number of workers discovered that Borrelia's many plasmids often carry "repeated" sequences (Simpson et al., 1990b; Marconi et al., 1996b; Porcella et al., 1996; Stevenson et al., 1996; Zuckert and Meyer, 1996; Casjens et al., 1997b; Misonne et al., 1997), however the true complexity of the their genomes was only revealed by the project to determine the entire nucleotide sequence the genome of B. burgdorferi B31, the type strain for the species.

The genome project showed that the culture (called “MI”) of isolate B31 that was analyzed carries one linear chromosome 910 kbp in length, 12 linear plasmids that vary in length from 5 kbp to 54 kbp, and 9 circular plasmids that vary in length from 9 to 32 kbp (Table 1) (Fraser et al., 1997; Casjens et al., 2000). This uncloned isolate, from an Ixodes scapularis tick captured on Shelter Island, New York, in 1981 (Burgdorfer et al., 1982), was propagated a minimal number of passages in culture medium and then passed four successive times through C57/HeJ mice at MedImmune Inc., Gaithersburg, Maryland (Fraser et al., 1997). In addition to these 21 plasmids, other parallel B31 cultures with different laboratory histories are known that carry at least three additional circular plasmids (table 1) (Casjens et al., 1997b; Miller et al., 2000). There is no evidence for non-clonality of the B31 isolate, and the three plasmids not present in B31 MI are apparently not required for infectivity in mice. The only clonal B31 sub-culture to be exhaustively analyzed carries 18 of these 24 plasmids (Casjens et al., 1997b; Casjens et al., 2000), but since these plasmids are easily lost in culture and sometimes lost during mouse passage (Barbour, 1988; Schwan et al., 1988; Persing et al., 1994; Norris et al., 1995; Xu et al., 1996), there is reason to suspect that all 24 may be able to

---

Table 1. The B. burgdorferi Strain B31 Plasmids

<table>
<thead>
<tr>
<th>Circular</th>
<th>Linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>cp9-1 3</td>
<td>lp5 3</td>
</tr>
<tr>
<td>cp9-2 2,3</td>
<td>lp17</td>
</tr>
<tr>
<td>cp26</td>
<td>lp21</td>
</tr>
<tr>
<td>cp32-1</td>
<td>lp25</td>
</tr>
<tr>
<td>cp32-2 2,3</td>
<td>lp28-1</td>
</tr>
<tr>
<td>cp32-3</td>
<td>lp28-2</td>
</tr>
<tr>
<td>cp32-4</td>
<td>lp28-3 3</td>
</tr>
<tr>
<td>cp32-5 3</td>
<td>lp28-4 3</td>
</tr>
<tr>
<td>cp32-6</td>
<td>lp36</td>
</tr>
<tr>
<td>cp32-7</td>
<td>lp38</td>
</tr>
<tr>
<td>cp32-8</td>
<td>lp54</td>
</tr>
<tr>
<td>cp32-9</td>
<td>lp56</td>
</tr>
</tbody>
</table>

~320,000 bp ~363,000 bp

Total Plasmids ~683,000 bp. Chromosome 910,725 bp

1 Numbers in plasmid names indicate the approximate size in kbp.
2 Not present in culture B31 MI.
3 Not present in B31 clone 4a.
Further Reading

Caister Academic Press is a leading academic publisher of advanced texts in microbiology, molecular biology and medical research. Full details of all our publications at caister.com

- MALDI-TOF Mass Spectrometry in Microbiology
  Edited by: M Kostrzewa, S Schubert (2016)
  www.caister.com/malditof

- *Aspergillus* and *Penicillium* in the Post-genomic Era
  Edited by: RP Vries, IB Gelber, MR Andersen (2016)
  www.caister.com/aspergillus2

- The Bacteriocins: Current Knowledge and Future Prospects
  Edited by: RL Dorfl, SM Roy, MA Riley (2016)
  www.caister.com/bacteriocins

- Omics in Plant Disease Resistance
  Edited by: V Bhaudaula (2016)
  www.caister.com/opdr

- Acidophiles: Life in Extremely Acidic Environments
  Edited by: R Quattrini, DB Johnson (2016)
  www.caister.com/acidophiles

- Climate Change and Microbial Ecology: Current Research and Future Trends
  Edited by: J Marxsen (2016)
  www.caister.com/climate

- Biofilms in Bioremediation: Current Research and Emerging Technologies
  Edited by: G Lear (2016)
  www.caister.com/biorem

- Microalgae: Current Research and Applications
  Edited by: MN Tsioleglou (2016)
  www.caister.com/microalgae

- Gas Plasma Sterilization in Microbiology: Theory, Applications, Pitfalls and New Perspectives
  Edited by: H Shintani, A Sakudo (2016)
  www.caister.com/gasplasma

- Virus Evolution: Current Research and Future Directions
  Edited by: SC Weaver, M Denison, M Roossinck, et al. (2016)
  www.caister.com/virusevol

- Arboviruses: Molecular Biology, Evolution and Control
  Edited by: N Vasilakis, DJ Gubler (2016)
  www.caister.com/arbo

- Shigellosis: Molecular and Cellular Biology
  Edited by: WD Picking, WL Picking (2016)
  www.caister.com/shigella

- Aquatic Biofilms: Ecology, Water Quality and Wastewater Treatment
  Edited by: AM Romani, H Guasch, MD Balagué (2016)
  www.caister.com/aquaticbiofilms

- Alphaviruses: Current Biology
  Edited by: S Mahalingam, L Herrero, B Herring (2016)
  www.caister.com/alpha

- Thermophilic Microorganisms
  Edited by: F Li (2015)
  www.caister.com/thermophile

- Flow Cytometry in Microbiology: Technology and Applications
  Edited by: MG Wilkinson (2015)
  www.caister.com/flow

- Probiotics and Prebiotics: Current Research and Future Trends
  Edited by: K Venema, AP Carmo (2015)
  www.caister.com/probiotics

- Epigenetics: Current Research and Emerging Trends
  Edited by: BP Chadwick (2015)
  www.caister.com/epigenetics2015

- *Corynebacterium glutamicum*: From Systems Biology to Biotechnological Applications
  Edited by: A Burkovski (2015)
  www.caister.com/cory2

- Advanced Vaccine Research Methods for the Decade of Vaccines
  Edited by: F Bagnoli, R Rappuoli (2015)
  www.caister.com/vaccines

- Antifungals: From Genomics to Resistance and the Development of Novel Agents
  Edited by: AT Coste, P Vandeputte (2015)
  www.caister.com/antifungals

- Bacteria-Plant Interactions: Advanced Research and Future Trends
  www.caister.com/bacteria-plant

- Aeromonas
  Edited by: J Graf (2015)
  www.caister.com/aeromonas

- Antibiotics: Current Innovations and Future Trends
  Edited by: S Sánchez, AL Demain (2015)
  www.caister.com/antibiotics

- Leishmania: Current Biology and Control
  Edited by: S Adak, R Datta (2015)
  www.caister.com/leish2

- *Acanthamoeba*: Biology and Pathogenesis (2nd edition)
  Author: NA Khan (2015)
  www.caister.com/acanthamoeba2

- Microarrays: Current Technology, Innovations and Applications
  Edited by: Z He (2014)
  www.caister.com/microarrays2

- Metagenomics of the Microbial Nitrogen Cycle: Theory, Methods and Applications
  Edited by: D Marco (2014)
  www.caister.com/n2

Order from caister.com/order
exist in the same cell (but see Stevenson et al. (2000b) for argument that cp32-2 and cp32-7 may not be compatible). These 24 plasmids and one chromosome are by far the largest number of different DNA replicons known in any bacterium. The exact genomic complexity of other B. burgdorferi isolates and other Borrelia species is not known, but where it has been studied (in about 20 Borrelia species at present), display of plasmids in electrophoresis gels suggests all Borrelias carry similarly large numbers of DNA replicons.

The B. burgdorferi B31 genome contains a chromosome that is 910,725 bp in length, 9 circular plasmids that total 249,330 bp (plus ~71,000 bp in the three unsequenced plasmids; table 1), and 12 linear plasmids that total about 363,000 bp (this later value is not known precisely because a few bp, on average <100, were not sequenced at most of the linear plasmids’ extreme termini) (Fraser et al., 1997; Casjens et al., 2000). This gives 1.52 to 1.59 mbp for the total genome size, depending upon whether the plasmids not present in B31 MI are included. Bacterial genomes vary in size from 0.56 to ~10 Mbp, and the Borrelia chromosome is near the small end of this range, which is commensurate with its lifestyle (below).

The linearity of the Borrelia chromosomes and the structure of their termini are extremely unusual. At each of the DNA ends one strand simply turns around and becomes the other strand to form “covalently-closed hairpin telomeres”. The iridopox and poxvirus chromosomes have similar termini, as do a small number of plasmids in other organisms (e.g., Hinnebusch and Tilly, 1993; Ravin et al., 2000), but no other cellular organisms are known to have this type of chromosomal telomere. Another type of linear bacterial replicon is now known in some actinomycete chromosomes and plasmids which have open ends with a covalently-bound protein at each 5'-terminus (summarized in Chen, 1996; Casjens, 1998). Known cellular DNA replication enzymes cannot initiate at ends without primers, so genome circularity and eukaryotic telomerases presumably evolved to handle end replication, and covalently-closed hairpin ends represent a different solution to this problem (discussed in detail in Casjens (1999)).

Sequence analysis of nine Borrelia telomeres has shown that there is an approximately 25 bp nucleotide sequence that is common to all these hairpin ends (Kitten and Barbour, 1990; Hinnebusch and Barbour, 1991; Hinnebusch and Tilly, 1993; Casjens et al., 1997a; Zhang et al., 1997), but the details of their replication remain mysterious. The fact that genes are preferentially oriented out from the center of the chromosome and the GC skew of the two strands is symmetrically arranged with respect to the center strongly suggested that the origin of replication was very near the center of the chromosome (Fraser et al., 1997; McLean et al., 1998; Lafay et al., 1999), and elegant work by Picardeau et al. (1999) has confirmed that at least the bulk of the chromosome is indeed replicated outward from the center.

Why do the Borrelias have linear chromosomes and how did they become linear? It is parsimonious to imagine that the circular chromosome of a progenitor of the Borrelias become linear, since nearly all other bacteria, including all the other spirochetes whose chromosome geometry is known, have circular chromosomes. But how might this linearization have occurred? Integration of a linear DNA molecule (plasmid or virus?) into a circular chromosome would create a larger linear molecule with the ends of the linear parent. Hinnebusch and Barbour (1992) noted that the telomeric sequences of Borrelia and the iridovirus African Swine Fever Virus (ASFV) have some similarity, and that the same ticks can carry both ASFV and Borrelia, and so suggested that the bacteria could have acquired the ends from the virus. This is a very interesting hypothesis, but since no poxvirus-like genes are currently recognizable in the genome, it seems equally likely that a Borrelia progenitor “invented” this type of telomere independently (and perhaps even donated them to the virus?). The advantages, if any, of DNA linearity are unknown, but the recombinogenic nature of DNA ends could perhaps make them useful in some circumstances.

**Predicted Borrelia Genes and Molecular Lifestyle**

**Genes of the B. burgdorferi Chromosome**

Analysis of the B. burgdorferi B31 genome sequence showed that essentially all of the potential “housekeeping” genes lie on the chromosome. Of the 846 intact ORFs (open reading frames) identified on the chromosome, 503 match a gene in the extant database whose biological function is at least partly understood, 70 match genes of unknown function, and the remainder are novel (Fraser et al., 1997). The putative cellular functions that were deduced from these sequence similarities give a picture of the limited metabolic capabilities of this organism that fits with its obligatory parasitic lifestyle. It is apparently able to convert various sugar molecules to lactate, which may be its only way of generating ATP, but there are no Krebs cycle enzymes and no oxidative phosphorylation proteins encoded by the genome. In addition, it can apparently combine fatty acids with glycerol to make phospholipids, can synthesize murien cell wall components from the necessary sugars and amino acids, can make AMP from glucose-6-phosphate and adenine, and can interconvert some nucleotides (but has no recognizable ribonucleotide reductase). It appears not to have other major metabolic capabilities, although a small number of additional pathways are apparently present, for example arginine catabolism and activated isoprene synthesis, which await biological role assignment. Thus, Borrelia metabolic capability is most notable for what it cannot do - de novo synthesis of amino acids, lipids, nucleotides and enzyme co-factors. We predict that these organisms must obtain all these essential molecules from their vertebrate or arthropod hosts, and the experimental evidence that exists supports this notion (Kelly, 1971; Wyss and Johnson, 1996). The presence of possible oligopeptide importers and a number of putative proteases suggests that oligopeptides could be a major source of amino acids for this organism. Another interesting point to emerge from this type of analysis is that B. burgdorferi does not appear to encode any proteins that require iron for their function, nor does it appear to encode any iron import system. Posey and Gherardindi (2000) recently confirmed this experimentally, showing that B. burgdorferi is among the extremely small number of organisms which have evolved not to require iron. It seems likely, given the substantially larger genome
sizes and greater metabolic capabilities of many other spirochetes, that Borrelia's small genome evolved from a progenitor with a larger genome by loss of genetic material as its parasitic lifestyle was adopted.

Curiously, in view of its nearly complete inability to make amino acids, lipids, nucleotides and co-factors, Borrelia appears to encode as few as 16 small molecule import systems. It is quite possible that there are some as yet unrecognized transporters, but it also seems likely that some of the transporters Borrelia has may have reduced specificity compared to those known in other organisms. The chromosome also encodes proteins required to replicate DNA, synthesize RNA and protein, as well as perform cell division, protein secretion (and associated N-terminal lipidation), genetic recombination and DNA repair. In each of these cases, by comparison with other bacterial genomes, there appear to be minimal sets of genes present, with little redundancy and few accessory factors, for each of these tasks. The Borrelia chromosome also carries a full set of motility-related genes (including several paralogous sets of chemotaxis genes), which occupies over 6% of the chromosome, suggesting that cell movement is a very important aspect of its lifestyle.

Genes on the chromosome are tightly packed (93% of the sequence is protein coding) and quite highly clustered according to function, both of which are typical for many bacteria. About 67% of genes are oriented so that they are transcribed away from the center of the chromosome. Very few homologs to known gene expression regulatory proteins are predicted to be encoded by the Borrelia genome; either there is, compared to the more well-studied bacteria, relatively little active regulation of gene expression in these organisms (but they are known to adjust gene expression upon temperature or pH changes (Stevenson et al., 1995; Carroll et al., 1999)), or they use novel proteins for this purpose. There are more than 250 novel genes on the chromosome, so the later is clearly possible. In addition, analysis of the putative genes identified on the chromosome and plasmids found little evidence for homologies to virulence factor genes in other better studied bacterial pathogens.

**Genes of the B. burgdorferi Plasmids**

Unlike the chromosomal genes, most genes on the strain B31 plasmids have no recognizable similarity to known genes outside of the Borrelia genus. Of the 670 predicted plasmid ORFs, only 5.8% have convincing similarity to genes in other organisms whose function is known and only 2% to genes whose function is unknown (Casjens et al., 2000). The former category includes genes involved in plasmid partitioning, small molecule transport, and nucleotide and DNA metabolism. Of particular interest are genes on the circular plasmid cp26 for conversion of inosine monophosphate to guanosine monophosphate, sugar import, and a possible oligopeptide transporter component. Nearly all of the B31 plasmids can be lost in culture and their loss has no apparent effect on growth in culture (their loss often results in loss of virulence in mice, however (e.g., Schwan et al., 1988; Xu et al., 1996)). The facts that all natural isolates of B. burgdorferi appear to carry a cp26 homolog, and that, unlike nearly all of the other plasmids, the cp26’s have never been lost from cultured organisms, suggest that this plasmid, unlike the others, may carry genes essential for growth in the laboratory.

Among the Borrelia-specific genes on the plasmids are an unusually large number that are predicted to encode proteins whose N-termini will be lipidated (Casjens et al., 2000). The genome contains at least 132 such genes, 91 of which lie on plasmids, and a number of the latter have been shown directly to be lipidated and encode proteins that are located on the outer surface of the organism where they are in contact with the host. Thus, even though they harbor no apparent sequence similarity to genes the more well-studied pathogens use to interact with their hosts, Borrelia, like those organisms, appears to concentrate such genes on extrachromosomal elements.

**Relationship to Other Spirochetes**

The Borrelia genus is genetically well defined (no known outliers) and quite well separated from other spirochetes. One of its closer relatives is the genus Treponema (Paster et al., 1991; Wang et al., 1999b). The complete genome of Treponema pallidum has also been sequenced (Fraser et al., 1998), and comparison with the genome of B. burgdorferi shows that they are surprisingly different. Only about 40% of B. burgdorferi genes have a clear ortholog in T. pallidum, and nearly half of these are highly conserved housekeeping functions that are present in most bacteria (Fraser et al., 1998; Subramanian et al., 2000). There are relatively few “spirochete specific” genes that are unique to these two genomes, but there are about 75 genes (all of unknown function) that are shared by the two sequences that have not been found elsewhere (Fraser et al., 1998).

Even when the two bacteria appear to be able to perform the same molecular function, they often do it in different ways. For example, B. burgdorferi only encodes the machinery to perform genetic recombination by a RecBC-Delab type pathway, and T. pallidum has none of these genes but encodes a full set of RecF recombination pathway genes. Similarly, B. burgdorferi imports sugars with a number of phosphotransferase type transporters while T. pallidum uses ABC type sugar transporters. The low fraction of orthologs between the two bacteria, and the fact that many of B. burgdorferi’s genes do not have the corresponding spirochete orthologs as their most similar homologs, are likely due to a combination of lineage-specific gene loss, horizontal gene transfer and perhaps rapid divergence.

**Borrelia Genome Uniformity in Nature**

**Chromosome Uniformity Among B. burgdorferi (sensu stricto) Isolates**

The chromosomes of different individuals within most bacterial species show substantial variation in length, and often in gene order (summarized in Casjens, 1998). These variations include the sporadic presence of such elements as integrated prophages, integrated plasmids, transposons, pathogenicity islands, retrons, integrons, etc., which can account for substantial fractions (often as great as >10%) of these genomes. B. burgdorferi, on the other hand, appears to have a remarkably invariant chromosomal structure. Complete physical maps of eight independent
isolates have been constructed and with one small exception (the extreme right end, see below) no substantial differences in length or gene position have been detected (Davidson et al., 1992; Casjens et al., 1995); in addition, no substantial structural differences have been noted when particular chromosomal regions have been analyzed in more detail in different isolates. The resolution of these chromosomal maps is such that if length differences as small as 5 kbp were present they would probably have been identified as such. These direct measurements, plus the analysis of the B31 genome sequence which found no evidence for genes related to any of the above mobile elements in the chromosome, suggest the bulk of the B. burgdorferi chromosome may not harbor such elements.

The one exception to this uniformity is length variation of about 20 kbp at the extreme right terminus of the chromosome. Compared to the sequenced B31 chromosome, the shortest B. burgdorferi chromosomes are missing ~7 kbp from the right end, and the longest chromosomes have an additional ~13 kbp of DNA at the right end. The shortest chromosomes have tightly packed genes up to ~200 bp from the right telomere. In about two-thirds (22 of the 31) of natural isolates analyzed, the chromosomes have "extra" DNA inserted within or replacing the rightmost terminal 200 bp of the shortest chromosomes, since the position of the rightmost gene (BB0843) is invariant relative to the remainder of the chromosome (Casjens et al., 1997a). This extra right-end DNA, in the two cases examined in detail (Casjens et al., 2000; M. Robertson, S. Casjens and W. M. Huang, unpublished), is essentially all similar to Borrelia linear plasmid DNA sequences. The simplest explanation for these observations is that parts of linear plasmid DNAs occasionally replace the extreme right chromosomal terminus. We have no explanation for how this occurs, why it appears to be limited to only the right terminus, or what advantage if any it might give the organism.

**Plasmid Uniformity Among B. burgdorferi (sensu stricto) Isolates**

Plasmid contents of independent B. burgdorferi isolates are rarely identical, but are nonetheless often surprisingly uniform. A number of previous studies have shown that, for example, a 26-28 kbp circular plasmid (cp26) carrying the outer surface protein OspC gene is always present, and the 49-55 kbp linear plasmid (lp54) carrying the Ospa/B genes is essentially always present in natural isolates (e.g., Samuels et al., 1993; Casjens et al., 1995; Anderson et al., 1996; Guttman et al., 1996; Mathiesen et al., 1997; Tilly et al., 1997). Likewise, multiple homologous 29-32 kbp plasmids (cp32’s) are always present (summarized in Casjens et al., 1997b; Stevenson et al., 2000b). Recently a more systematic study has shown that homologs of multiple probes from seven of the twelve strain B31 linear plasmids are present in at least 14 members of a panel of 15 independent, geographically diverse, North American B. burgdorferi (sensu stricto) isolates (Palmer et al., 2000). Apparent homologs of the remaining five B31 linear plasmids were present in 10, 10, 3, 2 and 2 members of the panel. Thus, although plasmid complements are not identical, and plasmid loss in culture could cause underestimation of the number of plasmids present in the wild, it can be concluded that “most plasmids have homologs in most natural isolates”, something that is not true in many other bacterial species. It has been pointed out that plasmids which are universally present and/or are absolutely required in the natural life cycle of an organism should perhaps be thought of as “mini-chromosomes” to more accurately reflect their importance to the organism (Barbour, 1993; Casjens, 1999; Palmer et al., 2000).

It should be emphasized however, that there is considerable plasmid variation in the face of the above overall uniformity in plasmid content. Apparently cognate plasmids often vary in size by as much as 10-20%, and sometimes much more, suggesting that the plasmids are not nearly as uniform in structure as the chromosome. For example, the lp36 kbp linear plasmid in B31 contains genes which are often found on 25-29 kbp linear plasmids in other isolates, and in several other cases genes which lie on the same plasmid in strain B31 are known to be on different plasmids in another strain (Feng et al., 1998; Palmer et al., 2000).

**Genomes of Other Borrelias**

Measurement of the length of whole Borrelia chromosomes has been performed on about fifteen species and all values so obtained indicate lengths of 900 to 1000 kbp. More accurate values derived from chromosomal macro-restriction maps agree that all members of the Lyme agent species cluster have been examined, B. burgdorferi, B. afzelii, B. garinii, B. bissettii, B. andersonii, B. valaisiana, B. japonica and B. lusitaniae, have chromosomes in the 930±20 kbp range (Davidson et al., 1992; Ojaimi et al., 1994; Casjens et al., 1995; van Vugt et al., 2000). In all cases studied, the particular genes examined lie at indistinguishable positions on the chromosomes of these species, suggesting that the chromosomal uniformity discussed above within the B. burgdorferi species may well extend to all members of this species cluster (the current resolution for “indistinguishable” varies considerably for different genes, but the rRNA genes, which have been mapped most accurately, lie at essentially identical positions in all these species), and in addition, particular chromosomal regions analyzed in more detail haven’t shown major structural differences (e.g., Bunikis et al., 1998; Gorbacheva et al., 2000).

The right end chromosomal length variation observed in B. burgdorferi has not been seen in other most Borrelia species examined, however an apparently similar phenomenon occurs at the left end of the B. japonica chromosome (Casjens et al., 1997a). At present we do not know if this chromosomal uniformity extends to throughout Borrelia genus, but the fact that at least B. hermsii, B. coriaeae, B. turicatae, B. anserina and B. parkeri appear to have similarly sized linear chromosomes makes this idea attractive (Kitten and Barbour, 1992; Casjens and Huang, 1993). In addition, multiple linear plasmids in the 5 to 200 kbp range and circular plasmids in the 9 to 70 kbp range have been observed in every Borrelia species studied (e.g., Kitten and Barbour, 1992; Casjens et al., 1995; Busch et al., 1996; Marconi et al., 1996a), but in no case except in B. burgdorferi B31 MI, (above) is the full complexity of the plasmid content known.

Are the plasmids the same in different Borrelia
species? Do plasmids move between species? Too little is known at present to draw any specific conclusions about overall plasmid contents of the non-Lyme Borrelias, however one cp32-like plasmid and one cp9-1-like plasmid have been completely sequenced from *B. hermsii* and *B. garinii*, respectively. The structure of the cp32s appears to be very constant in *B. burgdorferi* (e.g., Casjens et al., 1997b; Caimano et al., 2000), and the *B. hermsii* homologs are very similar to those, but with several inversions, deletions and insertions in the known sequences (Stevenson et al., 2000a). It will be of interest to see if these structural differences between the cp32-like plasmids of the two species proves to be universal among the *B. hermsii* cp32s. The cp9-1 homologs in *B. burgdorferi* and *B. garinii* are very similar except for 2 regions of non-similarity which total about 2 Kbp (Dunn et al., 1994; Casjens et al., 2000). Although the different species in the Lyme agent cluster certainly have similarities in the genetic content of their plasmids, there also appear to be systematic size differences between plasmids in different species (Palmer et al., 2000). For example, the B31 lp54 homolog is universally present in this species cluster, but appears to be consistently 5-10% larger in *B. garinii* and *B. afzelii* than in *B. burgdorferi* (Samuels et al., 1993; Casjens et al., 1995). These species-specific differences suggest that there may not be free flow of these plasmids among species, however, sequence analysis of several genes from many isolates has led some investigators to postulate that such transfer can occur (e.g., Marconi et al., 1994; Stevenson and Barthold, 1994; Jauris-Heipke et al., 1995; Livey et al., 1995; Wang et al., 1999a).

In spite of the current unknowns, it seems fair to conclude that the genomes of all Borrelias are likely to be made up of linear chromosomes of similar ~910 kbp sizes and of multiple linear and circular plasmids that make up significant fractions of their genomes (e.g., Hayes et al., 1988; Perng and LeFebvre, 1990; Casjens et al., 1995; Cutler et al., 1997; Hinnebusch et al., 1998; Carlyn et al., 2000a; Carlyn et al., 2000b). In the few cases where it has been examined, plasmids of *B. hermsii* and *B. turicatae* (relapsing fever Borrelias) carry genes that are similar to genes on the *B. burgdorferi* plasmids (Marconi et al., 1993; Pennington et al., 1999; Zuckert et al., 1999; Carlyn et al., 2000b; Stevenson et al., 2000a). It thus also seems likely that all Borrelias will be found to carry at least some related host-interaction (e.g., lipoprotein encoding) genes on their plasmids. Although there will certainly be many specific variations among them, when compared to other bacterial phyla the species of the *Borrelia* genus appear to have relatively invariant genome contents.

**Recent Rapid Plasmid Evolution?**

In spite of the overall genome uniformity discussed above, the recent detailed history of the *B. burgdorferi* plasmids appears to have been genetically turbulent. Analysis of the strain B31 linear plasmid sequences has shown that nearly all of these plasmids’ current structures must have been the result of genetic duplications and rearrangements (Casjens et al., 2000). One type of duplication seen is the duplication of whole plasmids, presumably by mutation to a new compatibility type (to be compatible with its parent). The cp32 family of circular plasmids in B31 is an excellent example of such duplications. The B31 MI genome carries seven homologous but different cp32 plasmids, one additional homologous copy recently integrated into a linear plasmid, and two additional circular homologs were most likely lost between strain isolation and sequencing (above). Why it might be advantageous for a bacterium to carry as many as ten homologs of this plasmid is not known, but

---

**Figure 1.** Interplasmid paralogies among *B. burgdorferi* B31 linear plasmids. Predicted genes on four linear plasmids are represented by the arrow shaped boxes whose shading indicates the following: black, putative intact genes; white, mutationally damaged pseudogenes; horizontal crosshatch, less than 300 bp in length with no paralog in the B31 genome. Beneath each plasmid map, thick lines indicate regions of paralogy to other B31 plasmids (indicated below thick lines) and thin lines indicate gaps in similarity with the indicated paralogous regions. “All” indicates regions with paralogous sequence on all of the larger B31 plasmids (all plasmids except lp5, lp17, lp21 and cp9); in all other regions where the plasmids in the figure have similarity to more than one other plasmid, only one of the other paralogies (chosen arbitrarily) was noted in the figure. “chrm” indicates paralogy to the plasmid-like region at the right end of the B31 chromosome. The slanting black line across lp56 indicates the location of the integrated cp32 plasmid (cp32 not shown here) (Casjens et al., 2000).
antigenically active and divergent lipoproteins are encoded by the different cp32 plasmids, and multiple cp32 homologs are present in all strains that have been carefully examined. It has been suggested the multiple, related copies of plasmid genes could be part the mechanism that allows a broad range of vertebrate hosts to be parasitized (Stevenson et al., 2000b). The genome sequence also shows that some *B. burgdorferi* plasmids are the apparent result of large deletions or insertions relative to homologs present in the same cell. For example, the structure of the linear plasmid lp21 of strain B31 appears to be largely the result of the insertion of 12.8 kbp block of DNA into a linear, lp5-like progenitor (Casjens et al., 2000). Similarly, circular plasmid cp18’s of strains N40 and 297 appear to be the results of simple deletions of ~12 kbp from cp32-like plasmids (Stevenson et al., 1997; Akins et al., 1999; Yang et al., 1999; Caimano et al., 2000). B31’s circular cp9-1 and linear lp54 evolutionary histories are somewhat more complex in that they appear to have been generated by more ancient and more complex rearrangements of cp32-like ancestors. Ten of the strain B31 linear plasmids (all except lp28-2 and lp54) appear to have had even more complex histories. Figures 1 and 2 show the very complex sequence relationships among several of these plasmids. They seem to have been built by chaotic random duplication and joining of various linear plasmid segments. Three apparent consequences of these past duplications are the high degree of paralogy among plasmid genes, the presence of a substantial number of pseudogenes on the plasmids, and the low protein coding potential of the affected plasmids (Casjens, 1999; Casjens et al., 2000). Figure 3 shows the distribution of the sizes of the 161 families of paralogous genes in the B31 genome. Of the 535 apparently intact genes ≥300 bp on the plasmids, only 63 have no paralog elsewhere on the plasmids; thus the large majority of plasmid genes (those with paralogs on other plasmids) must have been generated by past duplication events. The latter two features, pseudogenes and low coding density are likely related. We define pseudogenes as DNA regions which are very unlikely to contain functional genes, but which have similarity to putative intact genes elsewhere. Most of these have reading frames that have been damaged by frame-disrupting mutations and/or large deletions and insertions. There are 152 of this type of pseudogene on the B31 plasmids, and 9 pseudogenes (and 2 apparently intact genes) in the plasmid-like extreme right 7.2 kbp of the chromosome (above), most of which are very severely damaged (Casjens et al., 2000). In addition there are 15 known undamaged, unexpressed pseudogenes on lp28-1 which serve as a reservoir of diversity for the *vlsE* gene cassette (Zhang et al., 1997). For comparison, there is only one recognizably
frameshifted gene in the 843 gene “constant portion” of the chromosome. We interpret these pseudogenes (and other long stretches of plasmid non-coding DNA) to be regions that were duplicated but for which two copies are not necessary, and one copy is now undergoing random mutational degradation.

The ten plasmids which contain most (87%) of the pseudogenes have a total predicted protein coding fraction of 0.42, a value that is much lower than any other known comparably sized region of bacterial DNA. These 10 plasmids also carry the bulk (71%) of the predicted genome ORFs that are ≤300 bp long; 88 of 204 predicted genes on these plasmids are ≤300 bp long. Most of these predicted very short ORFs lie in regions that are also rich in pseudogenes and have especially low predicted ORF densities (Figure 4). It is thus quite possible that at least some of these short ORFs are not functional genes, but are the result of fortuitous ORFs predicted in DNA that no longer serves any function, and so the 0.42 protein coding fraction of the DNA of these plasmids may even be an overestimate.

Most sequenced bacterial genomes appear to be essentially free of this type of apparently decaying DNA (i.e., they are nearly completely “ameliorated,” at least in terms of removal of unneeded DNA). Thus, the fact that ten of the B31 linear plasmids are currently in an extremely unameliorated state suggests that, if amelioration rates are comparable in all bacteria, these plasmids are currently in a state of rapid evolutionary change. Work on B. hermsii and B. turicatae has noted the presence of both extensive duplicated sequences and pseudogenes on their plasmids (Restrepo and Barbour, 1994; Restrepo et al., 1994; Pennington et al., 1999; Stevenson et al., 2000b). In B. turicatae surface antigen serotype switching, large duplicative rearrangements of the linear plasmids appear to be generated during reassortment of sequences between

![Figure 4. Ten linear plasmids contain most of the pseudogenes, small genes and non-protein coding DNA of the B. burgdorferi B31 genome. A. The twelve “ameliorated” plasmids are cp9-1, cp26, the seven cp32s, lp28-2, lp54 and the cp32-like portion of lp56. The ten “rapidly evolving” plasmids are lp5, lp17, lp21, lp25, lp28-1, lp28-3, lp28-4, lp36, lp38 and the non-cp32-like portion of lp56 (see Casjens et al., 2000). B. Maps of two “rapidly evolving” linear plasmids, lp25 and lp28-3 with regions of low protein coding potential expanded (coordinates as in Casjens et al., 2000). Lp28-2, a linear plasmid with high coding potential, is shown for comparison. Genes shading is as described in figure 1.](image-url)
The Future?

The determination and analysis of the complete nucleotide sequence of *B. burgdorferi* genome has already led to many new insights into this organism’s lifestyle, which range from prediction of the location of the replication origin to deducing its overall metabolic capabilities and recognizing its iron-free metabolism to identification of possible new outer surface proteins. These certainly could not have been obtained so quickly in other ways. Perhaps unfortunately in the short run, genome analysis has not yet shed much light on the molecular basis for the pathogenic nature of this organism. Perhaps unfortunately, it may be, it must involve most of the B31 linear plasmids.

Acknowledgments

I wish to thank Brian Stevenson and Frank Gherardini for critical reading of this manuscript and for access to unpublished results, respectively.

References


