BMC Genomics

RESEARCH Open Access

Oxford nanopore technology R10 improved assembly metrics of *Borrelia* plasmids compared to R9

Dunia Velázquez[1,2,3], Sabrina Hepner[1,2*], Frederic D. B. Schedel[1], Magdalena Bichler[1,2], Andreas Sing[1,2], Dirk Metzler[3], Gabriele Margos[1,2*] and Volker Fingerle[1,2]

Abstract

In this project, we compared whole-genome *Borrelia* assemblies using reads produced by Oxford Nanopore Technology (ONT) with R9 and R10 chemistry and flow cells to investigate potential improvements between versions. For this, we sequenced two *B. garinii* isolates and one *B. tillae* isolate using ONT (R9 and R10) and reconstructed the genome using three different assemblers (Canu, Flye, Shasta). The generated consensus sequences were compared to references (Illumina for *B. garinii*, PacBio for *B. tillae*).

A general improvement of genome assemblies between R9 and R10 ONT versions was visible for both *B. garinii* isolates and *B. tillae*. This improvement was observed for all assemblers and can be clearly seen in the completeness and the decrease in the number of gaps. Using ONT, it was possible to generate at least one contig for each expected plasmid for both isolates of *B. garinii*, and even close some gaps that were present in Illumina references (Illumina-REF). Additionally, previously undefined contigs of the Illumina-REF assembly could be identified as parts of plasmid lp28-2, lp28-3, and lp32-10 based on the ONT assemblies. For the more complex genome of *B. tillae*, ONT was only able to generate complete linear plasmids, but all circular plasmids containing sequence stretches of high similarity described in a PacBio reference assembly resulted in a "wrong fusion". For all samples the use of different assemblers (Canu and Flye) was necessary to generate a whole genome sequence, with choice of assembler used for the final consensus sequence of individual genome elements varying between *Borrelia* species.

ONT R10 chemistry and flow cells greatly improved assembly of *B. garinii* genomes lacking complex circular plasmids and even fully resolved incomplete linear plasmids (gap closure) generated using Illumina sequencing. However, circular plasmids of *B. tillae* having sequence stretches of high similarity were not properly assembled. This implies that additional sequencing methods are needed when dealing with complex bacterial genomes.

Keywords Next generation sequencing, Oxford Nanopore Technology, R9, R10, *Borrelia garinii*, *Borrelia tillae*

*Correspondence:
Sabrina Hepner
sabrina.hepner@lgl.bayern.de
Gabriele Margos
gmargos1@gmail.com

[1]Bavarian Health and Food Safety Authority, Oberschleissheim, Germany
[2]German National Reference Center for Borrelia, Oberschleissheim, Germany
[3]Department of Biology, Division of Evolutionary Biology, Ludwig-Maximilians University, LMU Munich, Martinsried, Planegg, Germany

Background

Bacteria of the genus *Borrelia* Swellengrebel 1907 include species that can cause Lyme borreliosis (LB) or relapsing fever (RF) in humans, avian spirochetosis or bovine abortion in animals [1]. These bacteria have a small genome of approximately 1.5 Mbp, consisting of a linear main chromosome, with an approximate length of 900 kbp (approximately 60% of the total genome in isolate B31 of *Borrelia burgdorferi* sensu stricto Johnson et al. 1984 (s.s.)) and numerous linear and circular plasmids with varying lengths between 5 and 170 kbp (approximately 40% of the total genome in isolate B31 of *B. burgdorferi* s.s.) [2, 3]. The mol% G + C of the genome is comparatively low at approximately 30%. Furthermore, the genomes contain many repetitive elements, especially on plasmids [4–6]. The first completely sequenced genome was that of *B. burgdorferi* s.s. isolate B31 with a 910 kbp chromosome and a large proportion (5–30%) of genes encoding *Borrelia*-specific proteins without orthologues in nucleotide or protein databases [3, 7]. There are up to 12 linear and 9 circular plasmids with sizes of 5 – ~60 kbp, containing approximately 600 kbp and 700 coding sequences. Smaller genomes harboring fewer plasmids have been found in some species such as *B. maritima* Margos et al. 2020 [8] or in seabird-associated *B. garinii* Baranton et al. 1992 with only 6–9 plasmids [9]. RF spirochetes have a similar genome structure to *B. burgdorferi* sensu lato (s.l.) containing a linear chromosome and 5–23 plasmids [4]. The total length of the currently known genomes varies between 1.2 and 1.6 Mbp. Chromosomes have a length between 903 and 930 kbp. The length of the plasmids can also vary between 6 kbp and > 165 kbp [2]. These larger plasmids are considered megaplasmids or mini-chromosomes [10].

The chromosomes of different species or strains may have plasmid-like extensions at their right ends [6, 11, 12]. The *Borrelia* chromosome alone does not explain the high diversity found across different *Borrelia* species [4]. The biggest genetic differences are found in plasmids with the number, length, and gene content varying between species, but also between isolates of the same species [2, 5, 13–15]. The gain or loss of genetic elements via plasmid or gene acquisition or loss, as well as phage transduction, can lead to new environmental adaptation or change in bacterial pathogenicity, affecting human and animal health [16, 17]. Thus, plasmid-located genes encoding surface proteins are necessary for borreliae to complete the transmission cycle in nature involving interactions with vertebrate reservoir hosts (invasion, dissemination, persistence) and vectors. Plasmid-encoded proteins – although primarily required for host- or vector-interaction – may also be involved in virulence and pathogenesis of *Borrelia* in humans, and may also influence the clinical manifestations [18–20]. The genomes of LB species can also contain plasmids encoding phage proteins (named cp32) which have similar sequence stretches on several plasmids which can be problematic for genome assembly [21–23].

Thus, although *Borrelia* genomes are small compared to other bacteria [21, 24], when it comes to genome assembly, a big challenge for the plasmid analyses and comparison is not only the variation between strains and isolates but also the lack of reference plasmid sequences [25, 26]. The low GC content (~ 30%), sequence similarity between different plasmids, and a high proportion of repetitive elements, complicate the assembly process, especially from short-read technologies. Illumina sequencing has been a standard sequencing method due to its high accuracy (99-99.99%) and resulting low error rates (between 0.01 and 1%) [27]. And although short-read sequences such as Illumina allow successful assembly of the main chromosome and some plasmids, they are also known to have problems dealing with plasmids containing repetitive elements or stretches of similar sequences such as lp17, lp28 or the cp32 plasmid family. Because plasmid sequences contain information (genes) that are crucial to understanding *Borrelia*'s ecological adaptation, the use of long-read sequencing technologies such as Oxford Nanopore Technologies (ONT) or Pacific Bioscience single molecule real-time technology (PacBio SMRT) is required to improve proper assembly of missing or incomplete *Borrelia* plasmids and has increased over the past years [11, 15, 25, 28].

ONT allows the sequencing of ultra-long reads from 1 kbp to 2.273 Mbp, with relatively low material requirements [29]. The nanopore acts as a biosensor and is embedded into a resistant polymer membrane, which separates a reservoir with conductive electrolytes into "cis" and "trans" compartments. When constant voltage is applied across the membrane, an ionic current across the nanopore is generated and the negatively charged DNA is then driven through the nanopore into the positively charged "trans" side. DNA molecules moving through the pores block the constant ionic current and interrupt the signal, generating measurable changes in the current, that are then decoded using computational algorithms. The DNA translocation process is controlled by a motor protein, connected to the DNA by an adapter ligated to the DNA fragments. The motor protein does not only control the number of bases translocated across the nanopore per second but also has a helicase activity unwinding the double-stranded DNA into single-strands [27]. The parallelization of sequencing and translation process (current changes into nucleotides) allows sequencing in real-time [27, 30].

The concept of Nanopore sequencing was developed in the 1980s, but sequencing was first commercialized in 2014. Since then, there has been a constant improvement

in the nanopore technology, with 9 different flow cell chemistries released: R6 (June 2014), R7 (July 2014), R7.3 (October 2014), R9 (May 2016), R9.4 (October 2016), R9.5 (May 2017), R10 (March 2019) and R10.3 (January 2020), R10.4.1 (2022) [27] (see also https://nanoporetech.com/about/continuous-development-and-improvement and https://nanoporetech.com/about/history#timeline-Introducing%20the%20technology). Yet, translating electric current signals correctly into bases obtained with R9.4.1 was difficult for long homopolymer runs such as polyA or polyT, because the DNA signal was determined by five consecutive nucleotides in the nanopore at the same time [27]. To increase the accuracy in R10 technologies, especially for homopolymer regions, the pores were optimized with a longer pore head containing two different sensing regions, also called a double reader-head [27, 31].

The motor protein was also changed to provide more control over the DNA translocation. An improvement between R9 and R10 have been already reported for example in the sequencing of SARS-CoV with 96.52% accuracy for R9 and 98.34% in R10 or in mitochondrial genomes of higher organisms [27, 31, 32].

Previous studies on *Borrelia*, have already implemented the use of ONT due to the potential advantages [4, 8, 15, 26, 33]. But complete and high-quality genome reconstruction, especially concerning plasmids remains challenging. To this point, in the genome assemblies based on Illumina reads (referred to as Illumina-REF further down) reported by Margos and co-authors, two *B. garinii* isolates (17-54Z3 and NG-Z6) had unidentified contigs that could not be attributed to any of the plasmids (Fig. 1) [9].

Therefore, here we have analyzed, if recent improvements in the ONT sequencing chemistry and technology would translate to improving or solving difficulties associated with *Borrelia* whole genome assembly, considering the complexity of its genome. For this, we generated ONT sequences (R9 and R10) for whole genome assemblies and compared them to previous Illumina genome assemblies (Illumina-REF) of two different *B. garinii* isolates from seabirds (NG-Z6 and 17-54Z3, see Fig. 1) [9].

In addition, we tested if the improvements of the ONT chemistries would also be reflected in generating whole genome sequence data for RF spirochetes with a previously unknown genome such as *B. tillae*. RF *Borrelia* contain a megaplasmid with lengths ranging from 100 to 194 kb and high quantities of repetitive sequences, which may be challenging for genome assembly [4].

As ONT flow cell and kit chemistry are not the only factors that may affect the quality of the genome assembly, we also tested the role of different assemblers, namely Canu [34], Flye [35] and Shasta [36].

Materials and methods

Strains

For this study two different *B. garinii* isolates, 17-54Z3 and NG-Z6, were used which were originally isolated from *Ixodes uriae* White 1852 collected from seabirds (guillemot) on Horneya island, Norway. We choose these isolates because both Illumina-REF genomes had unidentified contigs of 2974 bp, 3243 bp (17-54Z3) and 3049 bp (NG-Z6) (Fig. 1) to examine whether these may belong to any of the identified plasmids [9]. *Borrelia tillae* Approved Lists 1980 [37] isolate Krampitz was obtained from Prof. Krampitz, Institute of Tropical Medicine, LMU München in 1984 in mouse blood.

Cultivation and DNA extraction

Borrelia garinii cultures were grown at 33 °C under a 5% CO_2 atmosphere in complete MKP medium containing 6% rabbit serum as described [38, 39]. *Borrelia tillae* was cultivated in basic MKP medium supplemented with 50% human serum (Biomex, Germany) [40]. DNA was extracted using the Maxwell 16 DNA extractor and a Maxwell LEV Blood DNA kit (Promega, Germany) according to the manufacturer´s instructions. DNA was eluted in EDTA-Free Elution Buffer (Qiagen, Germany, cat no. 19086).

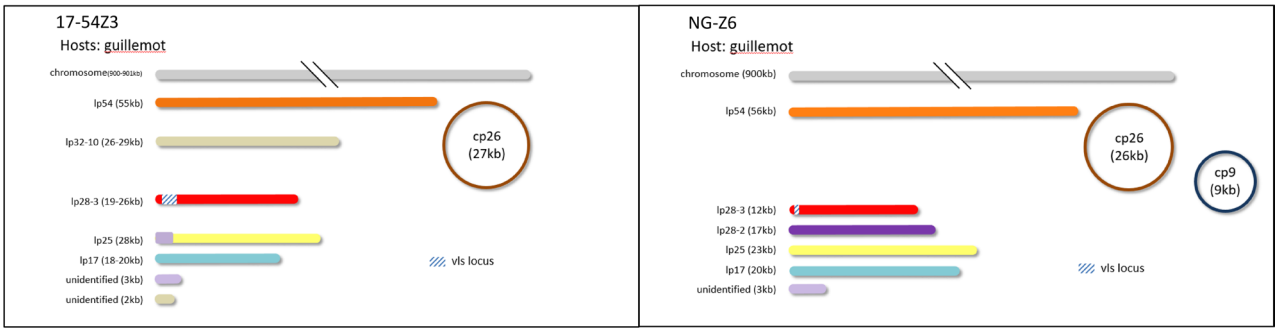

Fig. 1 Schematic drawing for the genome of *B. garinii* isolates 17-54Z3 and NG-Z6. Unidentified plasmids have no PFam32. Modified from [9]

Oxford nanopore sequencing

ONT sequencing was performed on a MinION Mk1B device using R9 (PCR Barcoding Kit SQK-RPB004 and FLO-MIN106 flow cell, R9.4.1) or R10 (Rapid Sequencing Kit V14 SQK-RAD114 together with FLO-MIN114 flow cell, R10.4.1) chemistry and flow cells and sequenced for approximately 5 h. Base calling was conducted using Guppy v6.5.7 high-accuracy (hac). The same base caller was used for R9 and R10 runs to maintain comparability of performance. Afterward, the data was processed following the visual overview (Fig. 2).

Assembly, QC, and contig preparation

Each sequencing run resulted in several fastq files per sample, which were merged into one fastq file per sample using the fastcat-tool v0.13.1 (available under: https://github.com/epi2me-labs/fastcat). ONT reads (R9 and R10) were assembled using Flye v2.8.1-b1676 [35] with --nano-raw -i 1 -g 1.5 m option (1 iteration, 1.5 m genome size), Canu v2.2 [34] with genomeSize = 1.5 m -nanopore option and Shasta v0.11.1 [36] with --config Nanopore-Phased-May2022 --Reads.minReadLength 1000 option. Other parameters were as per default. Given the low-quality assemblies for the *B. garinii* samples, we omitted Shasta assembling for *B. tillae*.

All assemblies were polished using one round of each, Racon v1.5.0 [41] and afterwards Medaka v1.8.0 (available under https://github.com/nanoporetech/medaka) (model: r941_min_high_g303 (R9), r1041_e82_400bps_hac_v4.2.0 (R10)). The assembly quality was evaluated using QUAST (QUality ASsessment Tool) v5.2.0 [42]. Contigs were then compared to the web-based NCBI-BLASTN core nucleotide databae (core nt) optimized for highly similar sequences (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with default parameters [43]. The NCBI-BLASTN hits (information about plasmid type, query coverage, and percent identity) were documented. Plasmids were typed according to PFam32 sequences [25, 26, 44]. Plasmid completeness was analyzed by checking for wraparounds (terminal inverted repeats indicate the linear structure and completeness of genome elements) and terminal direct repeats (indicate the circular structure and completeness of plasmids) [25]. For this, contigs were aligned with itself using the web-based nucleotide NCBI-BLASTN (align two or more sequences option) with default parameters resulting in dotplots and alignments [43]. Wraparounds or terminal direct repeats were then trimmed using MEGA5 (with default settings) [45]. To confirm that the sequence was trimmed at the right position cut sequences were aligned to the contig. The linear plasmids were considered complete when trimmed contigs included the "properly" spaced telomere sequence " TAGTATA" (usually 14 bp before the trimming site) [25]. Circular plasmids were considered complete if terminal direct repeats were present in the untrimmed sequence. For contigs without direct repeats the contig sequence was extended with a polyN tail (5,000 bp) at both ends. Subsequently, ONT reads were mapped on these polyN extended contigs using the CLC Genomic Workbench v.23. Contig sequences that were extendable by 1,000 bp on both sides were rechecked for direct terminal repeats.

Plasmid comparison

To compare the different plasmid sequences obtained from different assemblies and flow cells, they were compared to each other as well as to the previously known sequences Illumina-REF, i.e. Illumina sequences assembled with SPAdes [9, 46]. For this, the plasmids were

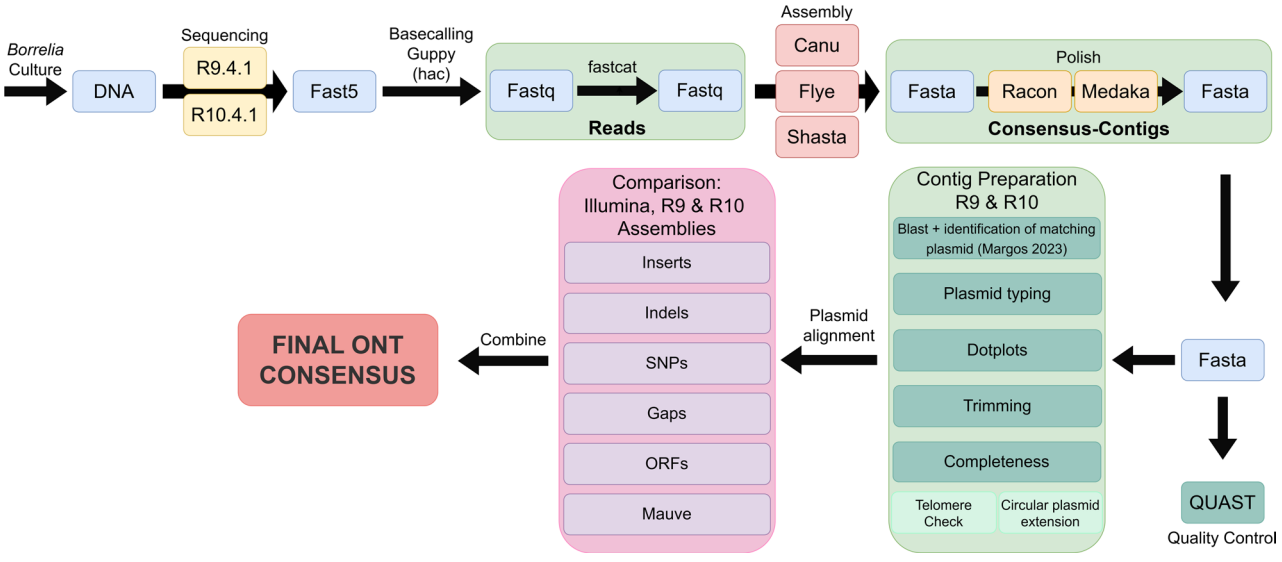

Fig. 2 Schematic overview for the creation of final ONT consensus sequences

aligned to each other using MEGA5 or MEGA6 with default settings [45, 47]. Some alignments had to be performed manually using the web-based nucleotide NCBI-BLASTN function "align two or more sequences" with default parameters and optimized for 'highly similar sequences (megablast)' [43]. Indels, inserts, and gaps were counted manually and noted, while Single Nucleotide Variants (SNVs) were determined using MEGAX [48]. Indels, inserts and gaps are defined as follows: Indels define single to four consecutive base deletions (< 5) compared to the Illumina reference (lack of two consecutive bases = two indels); Inserts define single base insertions compared to the reference and Gaps are defined as absence of ≥ 5 up to > 200 consecutive bases compared to the reference. We made this distinction between indels and gaps to better distinguish single base deletions from larger gaps.

These differences were later visualized via R Studio v4.4.3 [49] and Microsoft Excel software.

Open reading frames (ORF) were then analyzed via the CLC Genomic Workbench v23, with start codons defined for AUG, UUG, and UTG. ORFs were then compared and numbered based on their length, position, start and end codon.

In addition, plasmid sequences were annotated (gbk files) using the RAST annotation server [50] and afterward aligned and visualized using Easyfig v.2.2.5 [51].

Results

Comparison of *Borrelia* genome assemblies using ONT R9 and R10 chemistry

We analyzed if *B. garinii* NG-Z6 and 17-54Z3 ONT R9 and R10 runs resulted in comparable genome sequence assemblies as the Illumina-REF (Tables 1 and 2) assemblies published by [9]. In addition, we compared assemblies obtained with R9 and R10 chemistries and flow cells for one relapsing fever species, *B. tillae* (Tables 1 and 2). Since plasmid reconstruction comprises the biggest challenge in *Borrelia* whole genome assemblies [5, 25], in this comparison we focused on the plasmids. Results of QUAST analysis performed for the different assemblies and flow cell types are shown in Table 1. While QUAST provides useful initial statistics on contiguity and general assembly characteristics, such as the degree of fragmentation, due to the highly fragmented genome architecture of *Borrelia*, these metrics represent only first insights into assembly quality. Results of analyses of the differences such as indels, inserts, gaps, and SNVs compared to the

Table 1 QUAST analysis for *B. garinii* NG-Z6 and 17-54Z3 and *B. tillae* data assembled via Canu, Flye, or Shasta using R9.4.1 (R9) or R10.4.1 (R10) flow cells and chemistries from ONT. For the *B. tillae* assembly we omitted Shasta because of low quality assemblies for *B. garinii* samples

NG-Z6	**Canu R9**	**Canu R10**	**Flye R9**	**Flye R10**	**Shasta R9**	**Shasta R10**
Raw reads Mb[a]	173.7	94.6	173.7	94.6	173.7	94.6
number contigs	11	13	10	8	14	12
Largest contig	905,582	927,509	463,193	932,013	21,517	23,463
Total length	1,112,674	1,341,210	1,085,603	1,226,413	92,141	101,125
N50	905,582	927,509	442,641	932,013	5572	10,211
L50	1	1	2	1	4	3
GC (%)	27.94	27.54	28. 2	27.86	25.58	27.48
17-54Z3	**Canu R9**	**Canu R10**	**Flye R9**	**Flye R10**	**Shasta R9**	**Shasta R10**
Raw reads Mb[a]	167.1	49.7	167.1	49.7	167.1	49.7
number contigs	7	14	8	10	15	11
Largest contig	904,902	342,068	463,175	484,580	24,979	342,666
Total length	1,097,692	1,370,207	1,132,267	1,250,561	119,634	861,138
N50	904,902	191,178	442,037	447,290	8962	230,248
L50	1	3	2	2	5	2
GC (%)	28. 3	35.56	27.87	27.87	26.35	27.93
B. tillae	**Canu R9**	**Canu R10**	**Flye R9**	**Flye R10**	**Shasta R9**	**Shasta R10**
Raw reads Mb[a]	666.2	32.8	666.2	32.8	nd*	nd
number contigs	32	13	18	15	nd	nd
Largest contig	917,787	928,010	920,365	968,580	nd	nd
Total length	1,529,623	1,498,861	1,333,528	1,452,658	nd	nd
N50	917,787	928,010	920,365	968,580	nd	nd
L50	1	1	1	1	nd	nd
GC (%)	29.46	29.78	29.21	29.05	nd	nd

[a]the total number of raw reads in Mb included in assembly

*nd = no data

Table 2 Complete (green) and incomplete (red) plasmids generated by different assemblers and ONT chemistries for all samples. Contigs selected for the consensus sequences are marked with "+", contigs that could have been selected due to identical sequence to the consensus sequence are marked with "(+)" (in this case the possible plasmid number used in the consensus sequence is shown in brackets after the actual number). Incompletely assembled plasmids are represented in dark gray. For *B. tillae* only the circular megaplasmid and the linear plasmids are shown. The remaining circular plasmids are excluded as they were mostly wrongly fused due to high sequence similarity (Table S4)

		Canu R9	Canu R10	Flye R9	Flye R10	Shasta R9	Shasta R10
B. garinii NG-Z6	cp9				+		
	lp17				+		
	lp25				+		
	cp26			+			
	lp28-2				+		
	lp28-3				+		
	lp54		+				
	Complete plasmids	2	7	2	6	0	0
	Plasmids use in the consensus sequence	0	1	1	5	0	0
B. garinii 17-54Z3	lp17		(+)		+		
	lp25		+				
	cp26				+		(+)
	lp28-3				+		
	lp32-10		(+)		+		
	lp54		+				
	Complete plasmids	1	5	3	6	0	3
	Plasmids use in the consensus sequence	0	2(+2)	0	4	0	(1)
B. tillae Krampitz	cp169	+					
	lp33		+				
	lp32		+				
	lp25		+				
	lp20				+		
	lp8		+				
	Complete plasmids	1	4	0	2		
	Plasmids use in the consensus sequence	1	4	0	1		

Illumina-REF for the plasmids are shown in Supplementary Table S1.

For isolate NG-Z6 (chromosome and seven plasmids, Fig. 1), the chromosome was assembled in one contig using both assemblers, Canu (R9 and R10) and Flye (R10) (Table 1). In both Canu assemblies (R9 and R10) a higher number of contigs than expected was found (expected: eight, found: 11 and 13, respectively) (Table 1 and Supplementary Table S2). In both cases several contigs represented duplicates of one plasmid (whole plasmid or only part of it): Canu R9 assembled four different duplicate contigs of plasmid cp9 and Canu R10 four different duplicate contigs of lp17. While Flye R10 generated the expected 8 contigs, a higher contig number of 10 was observed in the Flye R9 assemblies (Table 1). This was due to the division of the chromosome into two different contigs, and the duplication of lp25 (Table S2). Shasta assemblies R9 and R10 also showed a high number of contigs (14 and 12, respectively), but most of them were small fragments of the chromosome: eight different contigs, with lengths between ~2–12 kbp for Shasta R9, and seven contigs with lengths varying between ~3–10 kbp for Shasta R10. The difficulties for Shasta to generate long contigs were obvious by the low N50 (5 kbp and 10 kbp) and higher L50 values (four and three) for both flow cell/

chemistry versions. Additionally, Shasta was not able to assemble any circular plasmids.

For isolate 17-54Z3 (Fig. 1; Table 1 and Supplementary Table S3), interestingly, only Canu R9 presented the chromosome in one contig. Canu R10 presented the chromosome in five different contigs, with an N50 value of 191,178 bp. Additionally, plasmid lp28-3 was present in two duplicated contigs. The GC content of Canu R10 with 35.56% is much higher than the expected ~ 27–28%. Flye R9 presented the chromosome in two different contigs, while R10 presented the chromosome in three different contigs as well as a plasmid duplicate for lp28-3. Finally, both Shasta ONT versions (R9 and R10) presented the chromosome in several contigs (eight contigs and L50 = 5 for R9 and five contigs and L50 = 2 for R10, Table 1 and S3). However, even in Shasta an improvement was visible using R10 (as compared to R9), with a lower L50 value, an increase in the N50 value (N50 = 230,248 bp), and in the total length generated by contigs over 25 kbp (Table 1).

Due to the poor performance of Shasta in the assemblies of *B. garinii*, Shasta was not included in the analyses of *B. tillae* (Table 1). ONT sequencing and genome assembly was done in parallel to an independent sequencing project of *B. tillae* using the long-read PacBio SMRT technology (Margos et al. personal communication). Therefore, in that case we were able to compare our results with the PacBio sequences. The consensus sequence generated by PacBio consisted of 16 different plasmids (10 circular, five linear, and one circular megaplasmid) (unpublished, NCBI Bioproject PRJNA1346408, Biosample SAMN52808808). Due to the lack of a standard plasmid typing scheme for RF *Borrelia* [52], plasmids were named according to plasmid topology (linear or circular, i.e. lp or cp., respectively), plasmid size and gene structure (see Supplementary Table S4) to allow an easier comparison of the plasmids based on different assembler and ONT versions.

In case of *B. tillae*, a reduction in the number of generated contigs between the R9 and R10 chemistries (indicating an improvement of the R10 assemblies) was visible for both Canu and Flye (see Table 1), although this was more noticeable for Canu (R9: 32 contigs, R10: 13 contigs) while for Flye, the number of contigs was reduced from 18 (R9) to 15 (R10). All ONT data and versions presented the chromosome in one contig (Table S2), also indicated by L50 = 1 and N50 > 900 kbp (Table 1). ONT R10 versions for both assemblies had a slightly higher total length (~ 200 kbp longer) than their respective R9 versions. The GC content was also similar across all assemblies and versions. Finally, contigs representing circular plasmids generated by ONT were present, but in most cases, plasmids were wrongly fused due to high sequence similarity (indicated in orange in Supplementary Table S4). For this reason, only the megaplasmid (cp169) and linear plasmids were aligned and compared with each other but also with the assembly generated from PacBio reads after processing each contig as described in the methods section (see Assembly, QC, and contig preparation).

Genome completeness of ONT R10 flow cell and chemistry compared to ONT R9

The completeness of plasmids was defined by wraparounds (terminal inverted repeats, Table S5) in linear plasmids and terminal direct repeats of circular plasmids [25]. Table S5 shows the first and terminal 50 bp of linear plasmids after cutting at the inverted repeat sequences (wraparound); the telomere signature sequence is indicated in red. The completeness of plasmids visibly improved in all assemblies of R10 ONT compared to R9 (Table 2) for all *Borrelia* isolates investigated. The number of complete contigs, independent of the number of differences compared to the reference data, was similar for both Canu R10 and Flye R10 in *B. garinii* (Table 2, S2 and S3). In the case of *B. tillae*, Canu R10 performed slightly better than Canu R9; Canu also performed better than Flye for both flow cell chemistries (Table S4). Completeness was used as the main criterion to define consensus sequences (Table 2, S2-S4). Some new complete plasmids generated by ONT sequences, such as lp28-2 and lp28-3 for *B. garinii* isolate NG-Z6 and lp28-3 and lp32-10 for isolate 17-54Z3, were also used to align previously incomplete or unclassified Illumina data (Fig. 3A and B). For consensus sequence determination, the number of ORF was also used (Fig. 4A and B). These are described in the following section.

For isolate NG-Z6 we established that the unidentified contig of 3,049 bp found previously [9], was part of plasmid lp28-2 (Fig. 3A). The linear plasmid lp28-2 was completely generated by Canu R10 and Flye R10 (Table 2, Supplementary Fig S1). Similarly, lp28-3 of NG-Z6 was not completely assembled based on Illumina short-read data. However, we were able to completely reconstruct the plasmid using data generated with Flye R10 and Canu R10 whilst the R9 versions of both assemblers missed over 1000 bp on the left side of the plasmid (Table 2, Supplementary Fig S2). Both Shasta versions generated only short contigs representing parts of lp28-3. By analyzing the ORFs we found that Flye R10 had 2 extra ORFs, not present in any other ONT assembly (black arrows in Fig. 4B). These ORFs were part of the *vls* (variable major protein (VMP)-like sequence) cassette (visible in the Easyfig alignment Fig S2) missing in the Illumina data. Flye R10 was the only assembly able to completely reconstruct this segment of the plasmid.

For lp28-3 of isolate 17-54Z3 we could determine that the unidentified 3,243 bp contig of the Illumina-REF [9] corresponded to the missing sequence on the left side of

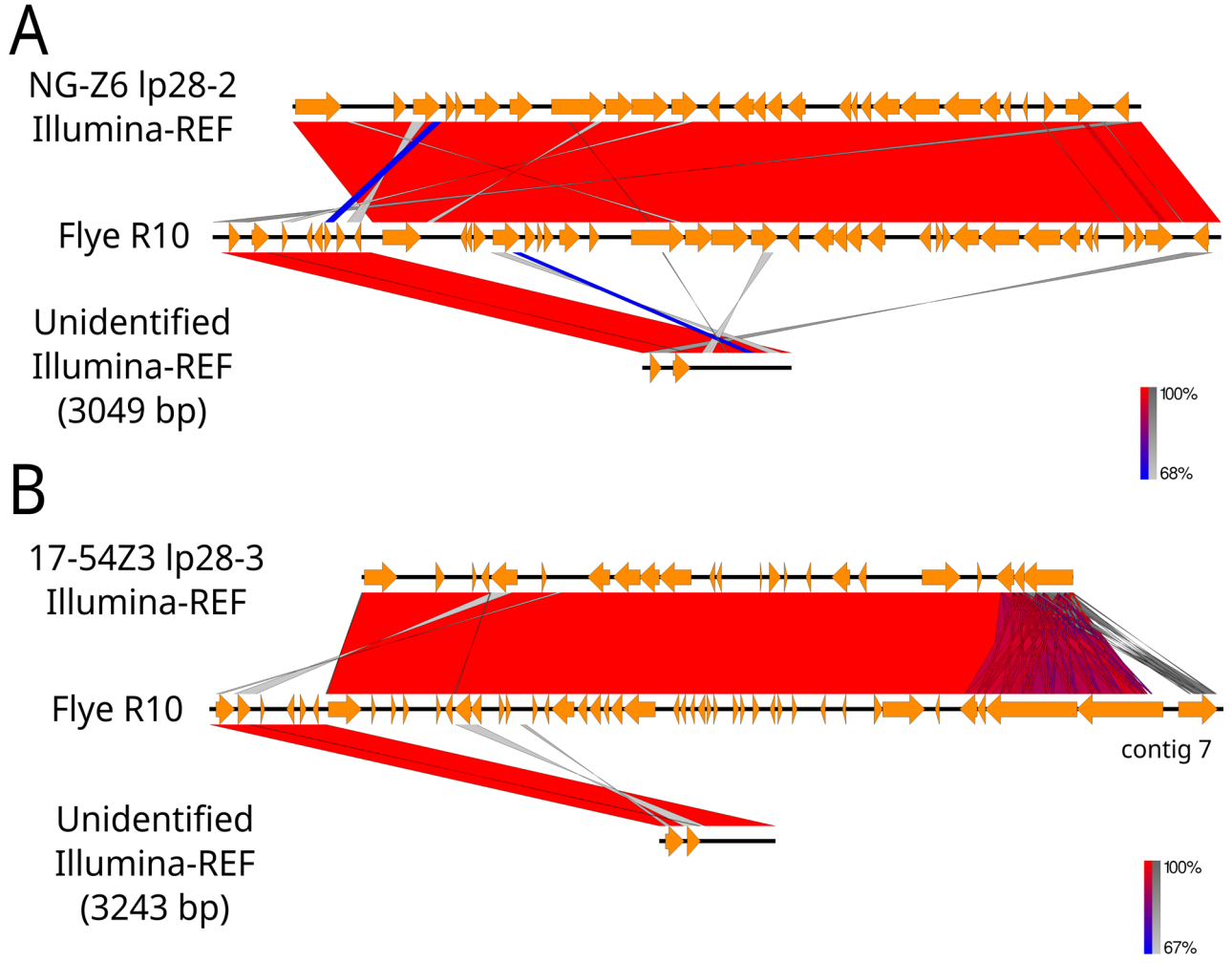

Fig. 3 Easyfig alignment (**A**) lp28-2 of *B. garinii* NG-Z6 and (B) lp28-3 of *B. garinii* 17-54Z3 including unidentified contigs (3049 bp for **A** and 3243 bp for **B**) for Illumina original data from Margos et al. 2023 and Flye R10 assembly data from the present study. CDS are represented in orange and identities between sequences are represented in red for 100% and turning slowly into blue for identities of 68% (**A**) or 67% (**B**) between sequences. In gray the equivalent scale for inverted sequences

the complete Flye R10 contig (defined by the telomere sequences on both ends). The Illumina-REF of lp28-3 missed the last 5,260 bp on the right end, representing the *vls* cassettes (Fig. 3B). This part of the plasmid was problematic in most assemblies and was only present in Flye R10. Interestingly, all R10 assemblies presented two duplicated contigs for plasmid lp28-3. Canu R10 and Shasta R10 contigs covered only part of the plasmid (Supplementary Fig S3). The second contig generated by Flye R10 covered only a small portion of the plasmid (4,497 bp). All R9 assemblers, including Shasta, generated only one contig covering most of the plasmid, but these contigs showed many differences to the ORFs generated by Flye R10 (data not shown, CDS of Illumina-REF and Flye R10 assemblies are shown in orange in Fig. 3B).

The ONT consensus sequences for plasmid lp17 were classified as complete for both *B. garinii* isolates, since they contained the telomere sequences and wraparounds. This plasmid also contains repetitive sequence, which is challenging when assemblies are based solely on Illumina data [9]. For lp17 of isolate NG-Z6 assembled using Illumina data the sequence "GTAATTAATATATGATAT AAA" was found repeated nine times between positions 18,226–18,407 bp. However, in all ONT data, at least one repetition is missing generating a 23 bp-long gap in the ONT sequences. In this case, the gap did not affect any ORF in the area. Due to the long-read nature of ONT, we assumed a higher fidelity in the number of repeats compared to Illumina short-read data. For lp17 of isolate 17-54Z3 the repetitive sequence "GATATAATATAATT AATATAT", starting at 18,371 bp, was found increased by ~600 bp which was present in all ONT assemblies compared to Illumina-REF (Supplementary Fig S4). This resulted in two additional ORFs in ONT data which were not present in the Illumina-REF (Fig. 4A).

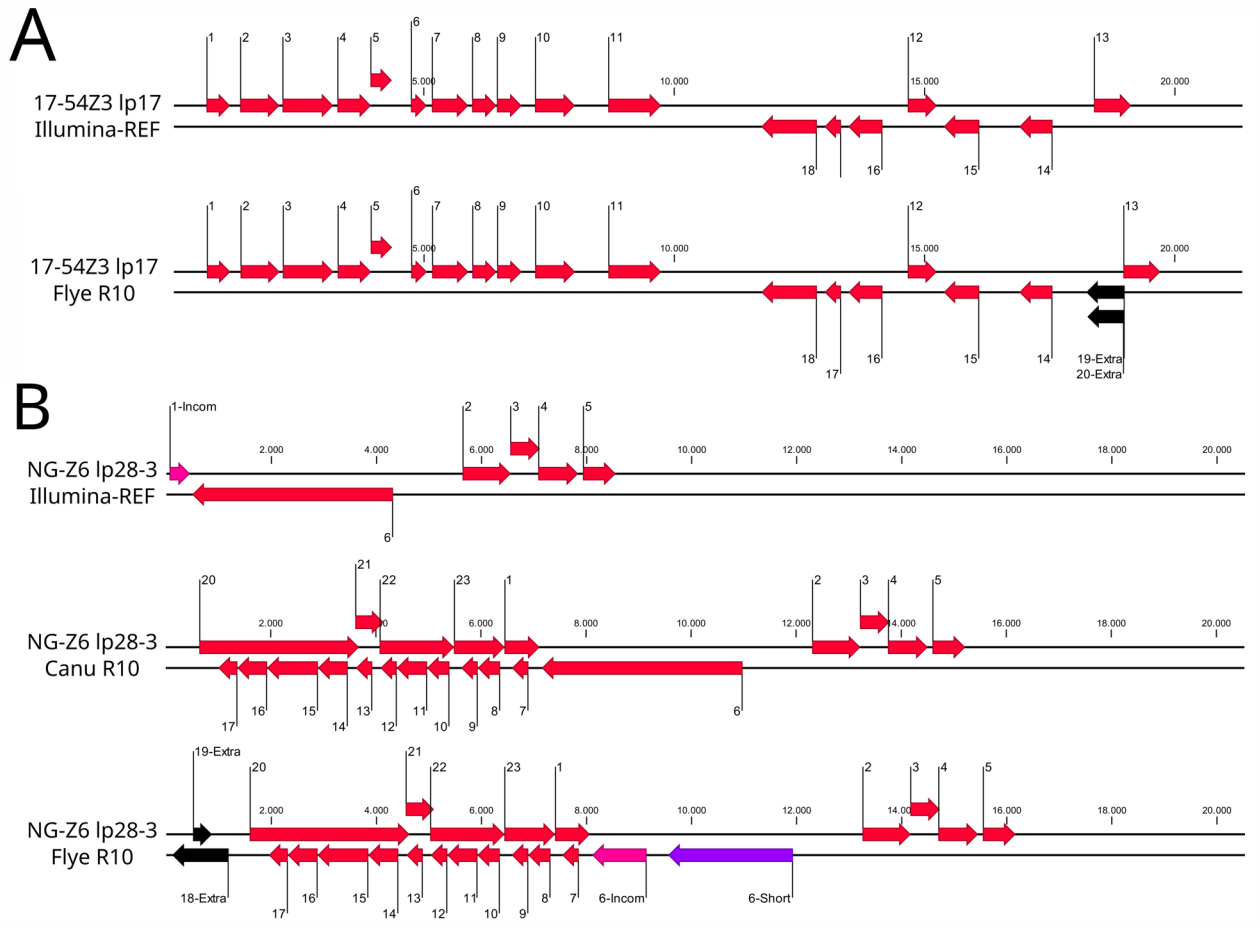

Fig. 4 ORFs for (**A**) *B. garinii* isolate 17-54Z3 plasmid lp17 and (**B**) *B. garinii* isolate NG-Z6 plasmid lp28-3 Illumina-REF compared to the completely reconstructed plasmid based on Flye R10 and/or Canu R10. As shown in Fig S5, assemblies of lp28-3 from ONT R10 data are longer than the Illumina assembly due to the reconstruction of the *vls/vlsE* locus. Please note that the Flye R10 assembly is longer than the Canu R10 assembly, having additional open reading frames (shown in black). Arrows in red represent identical ORFs compared to the Illumina-REF. Purple arrows represent ORFs with an additional stop codon, pink arrows ORFs with a different start codon but same stop codon, and black arrows extra generated ORFs. Numbers are assigned for individual ORFs. For lp28-3 both, ORF-18 and ORF-19 (black), found in Flye R10 are part of the vls cassettes. Figure was generated in CLC v. 23.0

Consensus sequence and comparison of numbers of indels, inserts, gaps and SNVs

After comparing the different contigs generated by the different assemblers, we generated a final combined consensus sequence which includes what we consider the best sequences. The decision which contig from which assembler was used as final consensus was based on: (i) completeness (i.e. presence of telomeres in linear or overlaps in circular plasmids), (ii) the lowest number of sequence differences compared to the Illumina-REF assembly, and (iii) number and position of ORFs.

For the *B. garinii* isolate NG-Z6, plasmids cp9 and lp28-2 were not only completely generated by Flye R10 but also had no differences to the Illumina-REF data. Therefore, these were chosen for the consensus sequence (Table 2 and S1). Completeness was also the main criteria by choosing Flye R10 for lp28-3 of NG-Z6, as it was the only assembly able to complete reconstruct the *vlsE/vls* locus. Two of the ORFs belonging to the vls/*vlsE* locus were missing in the contig generated by Canu R10. For lp17, assembler Canu and Flye R10 completely reconstructed the plasmid, but the lowest number of differences compared to Illumina-REF was found in the Flye R10 contig (Table S1). The differences in the sequences (SNVs, indels, inserts and gaps) (Fig. 5 and Table S1) did not affect the length or number of the ORFs between Flye R10 and Illumina data. Thus, the Flye R10 contig was considered the best and used for the final ONT consensus. For lp25, Flye R10 had the lowest number of differences in the ORFs compared to the Illumina-REF. Finally, Canu R10 was chosen for lp54 due to the lower number of differences compared to Flye R10 (Table S1).

For cp26 all assemblers, except Shasta, were able to generate contigs representing this circular plasmid.

	Indels	Canu R9	Canu R10	Flye R9	Flye R10
B. garinii NG-Z6	cp9	+	+	+	++
	lp17	-	+	-	++
	lp25	-	+	-	++
	cp26	+	+	++	-
	lp28-2	-	+	-	++
	lp28-3	-	+	-	++
	lp54	-	++	-	+
B. garinii 17-54Z3	lp17	-	+(+)	+	++
	lp25	-	++	-	+
	cp26	+	+	+	++
	lp28-3	-	-	-	++
	lp32-10	-	+(+)	+	++
	lp54	-	++	-	+
B. tillae Krampitz	cp169	++	+	-	-
	lp33	-	++	-	+
	lp32	-	++	-	-
	lp25	-	-+	-	-
	lp20	-	-	-	++
	lp8	-	++	-	-

	Inserts	Canu R9	Canu R10	Flye R9	Flye R10
B. garinii NG-Z6	cp9	+	+	+	++
	lp17	-	+	-	++
	lp25	-	+	-	++
	cp26	+	+	++	-
	lp28-2	-	+	-	++
	lp28-3	-	+	-	++
	lp54	-	++	-	+
B. garinii 17-54Z3	lp17	-	+(+)	+	++
	lp25	-	++	-	+
	cp26	+	+	+	++
	lp28-3	-	-	-	++
	lp32-10	-	+(+)	+	++
	lp54	-	++	-	+
B. tillae Krampitz	cp169	++	+	-	-
	lp33	-	++	-	+
	lp32	-	++	-	-
	lp25	-	-+	-	-
	lp20	-	-	-	++
	lp8	-	++	-	-

	SNVs	Canu R9	Canu R10	Flye R9	Flye R10
B. garinii NG-Z6	cp9	+	+	+	++
	lp17	-	+	-	++
	lp25	-	+	-	++
	cp26	+	+	++	-
	lp28-2	-	+	-	++
	lp28-3	-	+	-	++
	lp54	-	++	-	+
B. garinii 17-54Z3	lp17	-	+(+)	+	++
	lp25	-	++	-	+
	cp26	+	+	+	++
	lp28-3	-	-	-	++
	lp32-10	-	+(+)	+	++
	lp54	-	++	-	+
B. tillae Krampitz	cp169	++	+	-	-
	lp33	-	++	-	+
	lp32	-	++	-	-
	lp25	-	-+	-	-
	lp20	-	-	-	++
	lp8	-	++	-	-

	Gaps	Canu R9	Canu R10	Flye R9	Flye R10
B. garinii NG-Z6	cp9	+	+	+	++
	lp17	-	+	-	++
	lp25	-	+	-	++
	cp26	+	+	++	-
	lp28-2	-	+	-	++
	lp28-3	-	+	-	++
	lp54	-	++	-	+
B. garinii 17-54Z3	lp17	-	+(+)	+	++
	lp25	-	++	-	+
	cp26	+	+	+	++
	lp28-3	-	-	-	++
	lp32-10	-	+(+)	+	++
	lp54	-	++	-	+
B. tillae Krampitz	cp169	++	+	-	-
	lp33	-	++	-	+
	lp32	-	++	-	-
	lp25	-	-+	-	-
	lp20	-	-	-	++
	lp8	-	++	-	-

Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Heatmap representing the number of indels, inserts, SNVs, and gaps for all samples. Green fields represented no differences (=0) with REF data [except for gaps, where plasmids with longer contigs compared to Illumina (minus gap values) are represented in green]. Indels define single to four consecutive base deletions (< 5) compared to the Illumina reference; inserts define single base insertions compared to the reference and gaps are defined as absence of ≥ 5 up to > 200 consecutive bases compared to the reference. Red fields represent the highest number of differences for each plasmid. Orange, yellow and light green fields represent intermediates between both. Complete contigs are represented with an +, and incomplete contigs are represented with an -. Contigs chosen for the consensus sequence are represented with a second + symbol, contigs that could have been used for the consensus sequence were represented by (+)

Although all generated contigs were defined as completed via circular plasmid extension, we found that one ORF (location 18,468–19,592 bp in Illumina-REF) was missing in the Flye R10 assembly (see Supplementary Fig S5). Because this gap was only present in Flye R10 we assume it was misassembled. Comparing all ONT assemblies to the Illumina-REF we recognized a high number differences (< 400 SNVs) suggesting that the cp26 of isolate NG-Z6 was misassembled previously due to unknown reasons.

In case of lp17 and lp32-10 of isolate 17-54Z3, two different assembly contigs could be considered as final consensus sequences (Table 2), since both contigs did not show any or showed the exact number of differences between their ONT data and the Illumina-REF (Table S1). There were no differences between Canu R10, Flye R10, and Illumina-REF. Plasmid cp26 showed only one indel between Shasta R10, Flye R10 and Illumina-REF. In cases where ONT assemblies showed no differences, we chose Flye R10 for consensus sequences over others. For lp28-3 of 17-54Z3, only Flye R10 could be considered complete, thus, was taken as final consensus. Finally, Canu R10 was selected for lp25 and lp54 due to the lower amounts of differences to Illumina-REF compared to other assemblers.

For the *B. tillae* genome, candidates for final consensus plasmids were much lower in number than for the previously described *B. garinii* isolates. In many cases only one contig was considered complete (telomeres, overlaps): complete were lp32-Canu R10, lp25-Canu R10 (almost complete), lp20-Flye R10; a low number of differences compared to PacBio contigs (NCBI Genome Bioproject PRJNA1346408, Biosample SAMN52808808) had cp169-Canu R9 and lp33-Canu R10 (Fig. 5).

Finally, to compare how the different assemblies and ONT chemistries performed across the different plasmids, isolates, and species, a heat map comparing the number of indels, inserts, gaps, and SNVs was generated (Fig. 5).

For *B. garinii* the number of indels across the different isolates were lower for Canu R10 compared to Flye R10. Contrary, a lower number of inserts and SNVs was visible for Flye R10 compared to Canu R10. These patterns were not observed for the same assemblers based on R9 chemistry. Although contigs assembled with Shasta had a lower number of inserts, indels and SNVs, all of the contigs generated with Shasta were incomplete. Incomplete contigs tend to be shorter and therefore have a lower number of differences compared to the Illumina or PacBio data (Table S1). In the case of *B. tillae*, no clear trend was discernible for SNVs, indels and inserts between the different assemblies. A reason for this may be the low coverage noted in this assembly [53], the overall coverage using R10 was 27x (range 16x-42x) while in case of R9 the average coverage was > 700x. R10 generated more complete contigs compared to R9, although R9 tended to result in a lower number of SNVs, indels and inserts.

A convincing example of the improvement of R10 compared to R9 chemistry is plasmid lp25 of isolate 17-54Z3. The plasmid was challenging to assemble based only on Illumina data (the plasmid was fragmented with gaps between different contigs) [9]. In the contigs generated with the Flye R9 flow cell and chemistry, a high number of indels (611), SNVs (624), and inserts (62) compared to the Illumina-REF were found, most of them caused by consistent differences in PolyA and PolyT regions. Indeed (and as to be expected by marketing of ONT), an improvement was noticeable in Flye R10 generated contigs, with 41 indels, 19 inserts, and 107 SNVs compared to Illumina-REF.

Discussion

The use of long-read sequencing techniques to achieve whole *Borrelia* genome data, with a special interest in the accurate representation of completed plasmids (including correct reconstruction of repetitive motifs), has been a topic of interest in recent years across *Borrelia* researchers [5, 15, 25, 26, 28, 54, 55]. The low accuracy of ONT long-read sequencing compared to highly accurate Illumina short-read sequencing, especially in the homopolymer regions, was a crucial point regarding ONT technology. Improvements in the pore structure and the inclusion of a double reader-head in the pore, promised a higher sequencing accuracy, especially for homopolymer regions [27, 31].

In this project, we aimed to compare whole-genome *Borrelia* assemblies using ONT to investigate potential improvements between runs using R9 and R10 chemistry and flow cells. For this, we sequenced two *B. garinii* isolates and one *B. tillae* isolate using ONT (R9 and R10) and reconstructed the genome using three different assemblers (Canu, Flye, Shasta). The generated sequences were compared to references (Illumina for *B. garinii*,

PacBio for *B. tillae*). A general improvement of genome assemblies between ONT versions R9 and R10 was visible for both *B. garinii* isolates and *B. tillae*. This improvement was observed for all assemblers and can be clearly seen in the completeness and the decrease in the number of gaps.

R10 technology improved assembly of *B. garinii* plasmids

The final consensus obtained by Illumina sequencing for *B. garinii* isolate NG-Z6 consisted of the chromosome and 7 plasmids (lp54, lp28-3, lp28-2, cp26, lp25, lp17, and cp9) and isolate 17-54Z3 of a chromosome and 6 different plasmids (lp17, lp25, cp26, lp28-2, lp32-10, and lp54) [9]. We therefore expected ideally 8 and 7 contigs for these isolates, respectively (if every genome element was assembled in one contig). Using ONT, it was possible to generate at least one contig for each expected plasmid, and even close some gaps in sequences compared to the Illumina-REF. Specifically, previously undefined contigs in NG-Z6 and 17-54Z3 of the Illumina-REF assembly [9] could be identified as parts of plasmid lp28-2, lp28-3, and lp32-10 based on ONT assemblies.

Not surprisingly, more complex sequence stretches, such as the *vls* locus were incomplete or missing in the Illumina-generated data, due to the short length of reads (150–250 bp). The *vls* system is responsible for antigenic variation in *Borrelia* and involves an expression site (*vlsE*) and a contiguous array of 15 *vls* silent cassettes (repetitive sequences of ~500 bp). These silent cassettes result in repetitive sequence stretches that may complicate genome assembly from short read technologies. The *vlsE* encodes a protein with a lipoprotein leader sequence localized on the surface of the outer membrane. The site can undergo sequence variation by replacement of DNA segments of the *vlsE* section with corresponding regions of the silent cassettes [56]. Some *B. garinii* isolates such as Far04 have shown intercassette frameshifts between the different *vls* regions [57], adding more complexity to the sequence reconstruction. The known limitation of Illumina short-read data to properly assemble highly repetitive regions [3] could be overcome with ONT long-read sequencing. It needs to be mentioned though, that not all assemblers were equally able to reconstruct the *vls* locus including the cassettes, highlighting the importance of use of different assemblers.

For the *B. garinii* isolates analyzed here, Flye generated the highest amount of complete and high quality (low number of differences compared to the Illumina data) contigs, therefore, the majority of the plasmids in the final combined consensus are based on Flye R10. For some particular plasmids such as lp54, the contigs generated by Canu R10 presented a higher quality and were therefore chosen for the final combined consensus.

R10 technology improved assembly of linear but not circular plasmids in *B. tillae*

In the case of the RF species *B. tillae*, an improvement from R9 to R10 became obvious through an improvement in contig completeness between R9 and R10 data. In general, the genome assembly of *B. tillae* was challenging, due to the high number of circular plasmids found through PacBio sequencing and the high similarities shared between them. In *Borrelia* genomics it has become obvious that plasmids with long sequence stretches of high similarity such as the cp32 plasmid family [44] that encodes prophage sequences in *B. burgdorferi* s.l. are challenging to assemble using next generation sequencing [5, 11, 25]. In comparison, our *B. garinii* genomes did not have any cp32 plasmids facilitating the assembly process.

Based on PacBio-generated data, contigs for 10 different circular plasmids were generated in *B. tillae* (unpublished). When comparing the PacBio assemblies to the ONT-generated contigs, we were able to establish that circular plasmids were merged in ONT assemblies (wrong fusion, indicated in orange in Table S4) using both Flye and Canu. The current lack of a standardized plasmid typing system for RF *Borrelia* [2, 26], prevented us from having a control system to identified fused plasmids.

In the case of RF *Borrelia*, previous studies by [4, 26] have shown success in whole genome sequencing by using a hybrid approach between short-read and long-read sequencing methods. This hybrid approach has been previously recommended to generate highly accurate genome reconstructions in bacteria [27, 31].

The need to use several assemblers

Independent of the sequencing method, manual analysis for each contig is necessary to generate an accurate and high-quality consensus sequence for *Borrelia*. Quality checks by QUAST analysis may provide a quick overview of the generated contigs but it needs to be critically assessed. For example, discrepancies to the expected contig (plasmid) length may be caused by the presence of wraparound and overhangs (trimming is required) and duplicate contigs representing the same plasmid need to be removed. Although plasmids are considered complete when wraparound and telomere sequences or overhangs are present (for linear and circular plasmids, respectively) [25], it needs to be considered that internal gaps may lead to incompleteness. This was the case for plasmid cp26 of isolate NG-Z6 generated by Flye R10. Only after comparison with the other assemblies, we determined an 1,125 bp gap in the center of the plasmid. This underlines the importance of using different assemblers and subsequent comparison of contigs/plasmids to prevent further errors. Each assembler we used operates on

a different computational approach. Canu is based on an overlap-layout consensus algorithm, which builds a graph overlapping similar sequences, then a layout, and finally generates a consensus sequence. Flye also constructs the assembly by building a repeat graph that can tolerate some noise in the sequence from concatenated disjoint genomic segments. Finally, Shasta is based on a marker graph representing a marker (predetermined short subsequence) found to be aligned into a set of several reads [27, 31, 36]. Previous studies using ONT to sequence *Borrelia* genomes, have mostly used Canu for assembly [4, 26], although other authors used both Canu and Flye to generate the genomes of reptile-associated Borreliae [15]. In our case, Flye R10 performed best for *B. garinii* in producing final consensus sequences, five for isolate NG-Z6 (and one from R9) compared to one Canu R10 and four for isolate 17-54Z3 compared to one from Canu R10 (additionally two Canu R10 and one Shasta R10 plasmid sequences were identical to Flye R10). Although Shasta generated some completed plasmids for isolate 17-54Z3, it was not able to generate circular plasmids for isolate NG-Z6. The better performance of Flye over Shasta for bacteria de novo sequencing was also shown in a previous study [58].

Limitations of the study

Using ONT sequencing we were able to complete plasmid sequences for some *B. garinii* isolates that were incompletely assembled using Illumina sequencing only. However, even with improved R10 chemistry and flow cell technology, using only ONT reads even in combination with different assembler software, we were unable to completely assemble the complex genome of *B. tillae.* A problem may have been the low coverage (27x, range 16–42) of the genome using R10 and reasons for this need to be further investigated [53]. In the future more *Borrelia* species should be analyzed to delimit the limitations of ONT for sequencing and its reads for assembling complex bacterial genomes. Additionally, to allow comparability across the study, the outdated Guppy basecaller v6.5.7 was used. Further studies are required using the basecaller Dorado, to establish whether this may improve assembly outcome.

Concluding remarks

ONT offers the possibility of whole genome sequencing for less complex genomes of *Borrelia* species. It also provides the possibility to complete incomplete plasmid sequences generated using Illumina sequencing or check repetitive motifs in the sequence. Improvements in the flow cell chemistry over the past years, have improve the quality of the sequenced data.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12749-0.

Supplementary Material 1.

Acknowledgements

We would like to thank Sylvia Stockmeier, NRZ Borrelia, Oberschleissheim, Germany for expert Borrelia cultivation.

Authors' contributions

VF, GM conceptualized the study; DV, SH, FDBS, MB designed the workflow, acquired and analysed the data; VF, GM, DM, SH supervision; DV wrote the draft manuscript, all authors revised the manuscript and agreed to the final version.

Funding

The NRZ Borrelia is funded by the Robert Koch-Institut, Berlin, Germany.

Data availability

The data sets generated and/or analysed during the current study are available in NCBI. GenBank repository under the following links: [https://www.ncbi.nlm.nih.gov/bioproject/](https:/www.ncbi.nlm.nih.gov/bioproject) and https://www.ncbi.nlm.nih.gov/biosample/. NG-Z6 Bioproject PRJNA327303, Biosample SAMN3286046. 17-54Z3 Bioproject PRJNA327303, Biosample SAMN32860465. *B. tillae* Krampitz Bioproject PRJNA1346408, Biosample SAMN53417731; Accession JBTGSB000000000. In addition, data sets generated and analysed are available in BIGSdb under [https://pubmlst.org/organisms/borrelia-spp](https:/pubmlst.org/organisms/borrelia-spp) with the following IDs: NG-Z6, id: 4149, isolate: NG-Z6_wgs. 17-54Z3, id: 4148, isolate: 17-54Z3_wgs. *B. tillae* , id: 4150, isolate: Krampitz. Please note that registration to the BIGSdb and Borrelia MLST database is required to access the data.

Declarations

Competing interests

VF discloses research support: RKI/BMG, ESCMID, ECDC, StMGP/StMUG, INSTAND; lecturing activities (fees, travel expenses): DIAMEDIS, Diasorin, Mikrogen, Seramun, Siemens, HLR, Pfizer; Consultancy, round robin tests (no fees): QCMD, Pfizer, INSTAND, ECDC.

Received: 3 December 2025 / Accepted: 10 March 2026
Published online: 23 March 2026

References

1. Margos G, Henningsson AJ, Hepner S, Markowicz M, Sing A, Fingerle V. *Borrelia* Ecology, Evolution and Human Disease: A Mosaic of Life. In: Sing A, editor. Zoonoses: Infections Affecting Humans and Animals. Volume 1. Switzerland: Springer Cham; 2023.
2. Kingry LC. Genomics of Relapsing Fever Spirochetes. In: Radolf JD, Samuels DS, editors. Lyme Disease and Relapsing Fever Spirochetes: Genomics, Molecular Biology, Host Interactions and Disease Pathogenesis. Poole, UK: Caister Academic; 2021. pp. 63–72.
3. Schwartz I, Margos G, Casjens SR, Qiu WG, Eggers CH. Multipartite Genome of Lyme Disease *Borrelia*: Structure, Variation and Prophages. Curr Issues Mol Biol. 2021;42:409–54.
4. Kneubehl AR, Krishnavajhala A, Leal SM, Replogle AJ, Kingry LC, Bermudez SE, et al. Comparative genomics of the Western Hemisphere soft tick-borne relapsing fever borreliae highlights extensive plasmid diversity. BMC Genomics. 2022;23(1):410.
5. Margos G, Hepner S, Mang C, Marosevic D, Reynolds SE, Krebs S, et al. Lost in plasmids: next generation sequencing and the complex genome of the tick-borne pathogen *Borrelia burgdorferi*. BMC Genomics. 2017;18(1):422.
6. Casjens SR, Mongodin EF, Qiu WG, Luft BJ, Schutzer SE, Gilcrease EB, et al. Genome stability of Lyme disease spirochetes: comparative genomics of *Borrelia burgdorferi* plasmids. PLoS ONE. 2012;7(3):e33280.

7. Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, Lathigra R, et al. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. Nature. 1997;390(6660):580–6.
8. Margos G, Fedorova N, Becker NS, Kleinjan JE, Marosevic D, Krebs S, et al. *Borrelia maritima* sp. nov., a novel species of the *Borrelia burgdorferi* sensu lato complex, occupying a basal position to North American species. Int J Syst Evol Microbiol. 2020;70(2):849–56.
9. Margos G, Hofmann M, Casjens S, Dupraz M, Heinzinger S, Hartberger C, et al. Genome diversity of *Borrelia garinii* in marine transmission cycles does not match host associations but reflects the strains evolutionary history. Infect Genet Evol. 2023;115:105502.
10. Barbour AG. Linear DNA of *Borrelia species* and antigenic variation. Trends Microbiol. 1993;1(6):236–9.
11. Akther S, Mongodin EF, Morgan RD, Di L, Yang X, Golovchenko M, et al. Natural selection and recombination at host-interacting lipoprotein loci drive genome diversification of Lyme disease and related bacteria. mBio. 2024;15(9):e0174924.
12. Margos G, Becker NS, Fingerle V, Sing A, Ramos JA, Carvalho IL, et al. Core genome phylogenetic analysis of the avian associated *Borrelia turdi* indicates a close relationship to *Borrelia garinii*. Mol Phylogenet Evol. 2019;131:93–8.
13. Casjens SR, Di L, Akther S, Mongodin EF, Luft BJ, Schutzer SE, et al. Primordial origin and diversification of plasmids in Lyme disease agent bacteria. BMC Genomics. 2018;19(1):218.
14. Casjens SR, Gilcrease EB, Vujadinovic M, Mongodin EF, Luft BJ, Schutzer SE, et al. Plasmid diversity and phylogenetic consistency in the Lyme disease agent *Borrelia burgdorferi*. BMC Genomics. 2017;18(1):165.
15. Gofton AW, Popa-Baez A, Takano A, Soennichsen K, Michie M, Short M, et al. Characterisation and comparative genomics of three new Varanus-associated *Borrelia* spp. from Indonesia and Australia. Parasit Vectors. 2023;16(1):317.
16. Becker NS, Rollins RE, Nosenko K, Paulus A, Martin S, Krebs S, et al. High conservation combined with high plasticity: genomics and evolution of *Borrelia bavariensis*. BMC Genomics. 2020;21(1):702.
17. Margos G, Fingerle V, Reynolds SE. *Borrelia bavariensis*: Vector Switch, Niche Invasion, and Geographical Spread of a Tick-Borne Bacterial Parasite. Front Ecol Evol. 2019;7.
18. Strle K, Jones KL, Drouin EE, Li X, Steere AC. *Borrelia burgdorferi* RST1 (OspC type A) genotype is associated with greater inflammation and more severe Lyme disease. Am J Pathol. 2011;178(6):2726–39.
19. Strle F, Ruzic-Sabljic E, Cimperman J, Lotric-Furlan S, Maraspin V. Comparison of Findings for Patients with *Borrelia garinii* and *Borrelia afzelii* Isolated from Cerebrospinal Fluid. Clin Infect Dis. 2006;43:704–10.
20. Stanek G, Strle F. Lyme borreliosis: a European perspective on diagnosis and clinical management. Curr Opin Infect Dis. 2009;22(5):450–4.
21. Casjens S. *Borrelia* genomes in the year 2000. J Mol Microbiol Biotechnol. 2000;2(4):401–10.
22. Eggers CH, Caimano MJ, Clawson ML, Miller WG, Samuels DS, Radolf JD. Identification of loci critical for replication and compatibility of a *Borrelia burgdorferi* cp32 plasmid and use of a cp32-based shuttle vector for the expression of fluorescent reporters in the lyme disease spirochaete. Mol Microbiol. 2002;43(2):281–95.
23. Stevenson B, Zuckert WR, Akins DR. Repetition, conservation, and variation: the multiple cp32 plasmids of *Borrelia* species. J Mol Microbiol Biotechnol. 2000;2(4):411–22.
24. Casjens S. The diverse and dynamic structure of bacterial genomes. Annu Rev Genet. 1998;32:339–77.
25. Hepner S, Kuleshov K, Tooming-Kunderud A, Alig N, Gofton A, Casjens S, et al. A high fidelity approach to assembling the complex *Borrelia* genome. BMC Genomics. 2023;24(1):401.
26. Kuleshov KV, Margos G, Fingerle V, Koetsveld J, Goptar IA, Markelov ML, et al. Whole genome sequencing of *Borrelia miyamotoi* isolate Izh-4: reference for a complex bacterial genome. BMC Genomics. 2020;21(1):16.
27. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol. 2021;39(11):1348–65.
28. Hoornstra D, Kuleshov KV, Fingerle V, Hepner S, Wagemakers A, Strube C, et al. Combining short- and long-read sequencing unveils geographically structured diversity in Borrelia miyamotoi. iScience. 2024;27(9):110616.
29. Lu H, Giordano F, Ning Z. Oxford Nanopore MinION Sequencing and Genome Assembly. Genom Proteom Bioinform. 2016;14(5):265–79.
30. Feng Y, Zhang Y, Ying C, Wang D, Du C. Nanopore-based fourth-generation DNA sequencing technology. Genom Proteom Bioinform. 2015;13(1):4–16.
31. Sanderson ND, Kapel N, Rodger G, Webster H, Lipworth S, Street TL et al. Comparison of R9.4.1/Kit10 and R10/Kit12 Oxford Nanopore flowcells and chemistries in bacterial genome reconstruction. Microb genomics. 2023;9(1).
32. Baeza JA, Minish JJ, Michael TP. Assembly of Mitochondrial Genomes Using Nanopore Long-Read Technology in Three Sea Chubs (Teleostei: Kyphosidae). Mol Ecol Resour. 2025;25(1):e14034.
33. Kuleshov KV, Koetsveld J, Goptar IA, Markelov ML, Kolyasnikova NM, Sarksyan DS et al. Whole-Genome Sequencing of Six *Borrelia miyamotoi* Clinical Strains Isolated in Russia. Genome announcements. 2018;6(1).
34. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5):722–36.
35. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019.
36. Shafin K, Pesout T, Lorig-Roach R, Haukness M, Olsen HE, Bosworth C, et al. Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. Nat Biotechnol. 2020;38(9):1044–53.
37. Skerman VBD, McGowan V, Sneath PHA. In: McGowan V, Sneath PHA, editors. Approved Lists of Bacterial Names Skerman VBD. Washington (D.C.): ASM; 1980.
38. Preac-Mursic V, Wilske B, Schierz G. European *Borrelia burgdorferi* isolated from humans and ticks culture conditions and antibiotic susceptibility. Zentralbl Bakteriol Mikrobiol Hyg A. 1986;263(1–2):112–8.
39. Ruzic-Sabljic E, Lotric-Furlan S, Maraspin V, Cimperman J, Logar M, Jurca T, et al. Comparison of isolation rate of *Borrelia burgdorferi* sensu lato in MKP and BSK-II medium. Int J Med Microbiol. 2006;296(Suppl 40):267–73.
40. Margos G, Stockmeier S, Hizo-Teufel C, Hepner S, Fish D, Dautel H, et al. Long-term *in vitro* cultivation of *Borrelia miyamotoi*. Ticks Tick Borne Dis. 2015;6(2):181–4.
41. Vaser R, Sovic I, Nagarajan N, Sikic M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017;27(5):737–46.
42. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072–5.
43. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
44. Casjens S, Palmer N, van Vugt R, Huang WM, Stevenson B, Rosa P, et al. A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. Mol Microbiol. 2000;35(3):490–516.
45. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28(10):2731–9.
46. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput biology: J Comput Mol cell biology. 2012;19(5):455–77.
47. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–9.
48. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35(6):1547–9.
49. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2025.
50. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.
51. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics. 2011;27(7):1009–10.
52. Gofton AW, Popa-Baez A, Takano A, Soennichsen K, Michie M, Short M, et al. Characterisation and comparative genomics of three new Varanus-associated Borrelia spp. from Indonesia and Australia. Parasites Vectors. 2023;16(1):317.
53. Sanderson ND, Hopkins KMV, Colpus M, Parker M, Lipworth S, Crook D et al. Evaluation of the accuracy of bacterial genome reconstruction with Oxford Nanopore R10.4.1 long-read-only sequencing. Microb genomics. 2024;10(5).
54. Kingry LC, Batra D, Replogle A, Rowe LA, Pritt BS, Petersen JM. Whole Genome Sequence and Comparative Genomics of the Novel Lyme Borreliosis Causing Pathogen, *Borrelia mayonii*. PLoS ONE. 2016;11(12):e0168994.
55. Kneubehl AR, Osikowicz LM, Parise CM, Van Gundy TJ, Replogle AJ, Lopez JE, et al. Complete genome sequence of *Borrelia miyamotoi* strain MN18-0001, an Am-East-2 strain type isolate derived from an *Ixodes scapularis* tick (Minnesota). Microbiol Resour Announc. 2024;13(11):e0049024.

56. Norris SJ. vls Antigenic Variation Systems of Lyme Disease Borrelia: Eluding Host Immunity through both Random, Segmental Gene Conversion and Framework Heterogeneity. Microbiol Spectr. 2014;2(6).
57. Norris SJ, Brangulis K. Meta-analysis of the Vmp-like sequences of Lyme disease *Borrelia*: evidence for the evolution of an elaborate antigenic variation system. Front Microbiol. 2024;15:1469411.
58. Khrenova MG, Panova TV, Rodin VA, Kryakvin MA, Lukyanov DA, Osterman IA et al. Nanopore Sequencing for De Novo Bacterial Genome Assembly and Search for Single-Nucleotide Polymorphism. Int J Mol Sci. 2022;23(15).

Publisher's Note