Michael J. Skvarla · Ronald Ochoa ·
Jose Carlos Verle Rodrigues ·
H. Joel Hutcheson *Editors*

Contemporary Acarology

2017

Chapter 1
Use of the Ion Torrent PGM for Determining the Genomic Sequences of *Francisella* and *Coxiella*-Like Endosymbionts and *Rickettsia* Directly from Hard Ticks

Gregory A. Dasch, Arunachalam Ramaiah, Zachary C. Holmes, Maria L. Zambrano and T. Brian Shirey

Abstract Microbiome assessments based on amplification and sequencing of their 16S rRNA genes or targeted surveys of specific agents have greatly augmented our understanding of the distribution and identity of bacteria found in ticks. The most prevalent tick bacteria include approximately 10 species, such as *Coxiella*-like (CLE) and *Francisella*-like (FLE) endosymbionts, and spotted fever group *Rickettsia* (SFG). However, genomes of only a few of the many tick agents that are not cultivable have been characterized by very deep next generation sequencing (Illumina HiSeq). As an alternative approach, we have assessed use of the Ion Torrent Personal Genome Machine (IT-PGM) for direct sequencing and assembly of these agent genomes. Agents were identified and quantitated in three genera of ticks (*Amblyomma*, *Dermacentor*, *Rhipicephalus*) by qPCR assays. Geneious, CLC Genomics Workbench, BWA, and SPAdes tools were used for read mapping and assembly of the sequences obtained from size-fractionated libraries made from DNA that was extracted from single alcohol-preserved ticks on single 318 or 314 IT-PGM chips with 200 bp chemistry. Plasmid and chromosome sequences were obtained for six SFGs, four FLEs, and three CLEs. The depth and percentage of genome coverage for symbionts and *Rickettsia* were enhanced by use of Qiagen Repli-G amplification of the DNA used for the libraries. The IT-PGM is a relatively inexpensive sequencing platform for initial genomic characterization of some of the abundant bacterial agents found in hard ticks.

G. A. Dasch (✉) · A. Ramaiah · Z. C. Holmes · M. L. Zambrano
Rickettsial Zoonoses Branch, US Centers for Disease Control and Prevention, 1600 Clifton Rd., Atlanta, GA 30333, USA
e-mail: ged4@cdc.gov

T. B. Shirey
Enteric Diseases Laboratory Branch, US Centers for Disease Control and Prevention, 1600 Clifton Rd., Atlanta, GA 30333, USA

© Springer Nature Switzerland AG 2019

M. J. Skvarla et al. (eds.), *Contemporary Acarology*,
https://doi.org/10.1007/978-3-030-17265-7_1

Keywords Ion torrent PGM · Hard ticks · Endosymbiont · *Rickettsia* · *Coxiella* · *Francisella*

1.1 Introduction

The intimate association of bacterial symbionts with arthropods has been recognized for decades. Buchner (1965), the influential father of this field, developed and advanced the hypothesis that nutritional deficiencies in the diet of many arthropods made such associations advantageous for them. Indeed, many such symbionts are maintained as obligate heritable associations in specialized organs called mycetomes and they are passed transovarially and transtadially. Arthropods dependent on diets consisting only of plant sap or vertebrate blood generally have symbionts, a finding consistent with the nutritional hypothesis. The advent of molecular methods for the characterization of these bacteria have led to an expanding number of symbionts whose genomes have been sequenced, thus permitting the development of molecular hypotheses about the precise beneficial physiological and immunological interactions that occur between each host and bacterium (Wernegreen 2015). The range of bacterial symbiotic associations in arthropods ranges from obligatory to facultative; besides nutritional benefits, they can confer secondary effects, including reproductive manipulation of the host, protection of the host against parasites and other enemies, and a repertoire of additional metabolic capabilities that facilitate adaptation to changing environments and variable marginal dietary resources (Moran et al. 2008; Wernegreen 2012; Corbin et al. 2017). The associations can be extremely complex, as two or more bacterial symbionts may be present and may work in concert or in competition in their interactions with their hosts and one may be obligate and the other(s) facultative. Different populations of the same host species may have vastly different complements of symbionts; indeed, the symbionts in different host populations may be genetically variable and confer different phenotypes on those hosts (Henry et al. 2015; O'Fallon 2008; Douglas and Werren 2016). However, many relatively stable associations of dominant heritable primary symbionts with their hosts ("holobionts") remain poorly understood.

One example of this paucity of detailed information is the various associations of ticks with bacterial endosymbionts. Duron et al. (2017) recently summarized information on ten maternally inherited bacteria found in ticks. Four were Gamma-Proteobacteria belonging to the genera *Coxiella*, *Francisella*, *Rickettsiella*, and *Arsenophonus*. The Alpha-Proteobacteria also had members of four genera, including *Wolbachia*, *Rickettsia*, *Midichloria*, and *Lariskella*. The remaining genera included *Spiroplasma* in the Mollicutes and *Cardinium* in the Bacteroidetes.

Direct experimental evidence for the role of tick symbionts in providing essential vitamins, cofactors, and/or amino acids is very limited. Zhong et al. (2007) demonstrated that survival and reproduction of *Amblyomma americanum* was adversely affected by reduction of its *Coxiella*-like agent with antibiotics, a classical method for inferring the essential role of a symbiont in the nutrition of its host. Since *A. amer-*

icanum lineages do not always harbor *R. amblyommatis* while its *Coxiella* is present in nearly 100% of the ticks, that agent appears to be more essential to tick survival. Zhong (2012) has reviewed the literature on this symbiosis. More recently, the value of genomic sequences in supporting a role for vitamin synthesis in 50 symbionts of insects and arachnids has been reviewed by Serbus et al. (2017). The variety of vitamin, cofactor, and amino acid pathways employed in different symbionts suggests that the symbiont-host interactions are probably not uniform in all tick-symbiont associations, so that complete genome sequencing of these agents will be an essential first step in understanding that interaction and how it may be exploited for new approaches to tick control. Furthermore, symbiont-pathogen interactions may be equally diverse (Oliva Chávez et al. 2017).

Although the widespread use of PCR assays for detecting both the symbionts and bacterial pathogens transmitted by ticks (Mixson et al. 2006) has greatly increased our understanding of the distribution of such agents across tick species and habitats, this approach is limited by sampling and sensitivity. For example, the survey of ticks by Epis et al. (2008) for the fascinating intramitochondrial symbiont, *Midichloria mitochondrii*, missed its presence in *Amblyomma americanum*. Deeper sequencing, using 16S microbiome of a larger sample of ticks from additional locations, demonstrated that three genotypes of *M. mitochondrii* were present (Williams-Newkirk et al. 2012). *Wolbachia*, which is estimated to be present in 10% of arthropods, was only recently found to be an important agent in some aphid populations. Consequently, microbiome surveys are more suited than are PCR screens to sensitive evaluation of the prevalence and relative numbers of symbionts in significant size samples of ticks (Williams-Newkirk et al. 2014). We were impressed by the large amounts of *Rickettsia* and *Coxiella* present relative to the other bacteria in the *A. americanum* microbiome. While it was possible for us to routinely isolate *R. amblyommatis*, the commonplace rickettsial agent of this tick (Mixson et al. 2006), in cell culture and to sequence its complete genome (chromosome and 3 plasmids) (strain GAT-3OV, NCBI GCA_000284055.1, Karpathy et al. 2016), the *Coxiella*-like agent (CLE) has not been cultivated. Both Smith et al. (2015) and our laboratory were able to assemble the complete sequence from shotgun metagenomics libraries sequenced on an Illumina HiSeq (Williams-Newkirk et al. 2015; Ramaiah et al. 2017). However, no complete tick genome sequences have been obtained to date, so it is not possible to filter out the reads obtained from the large tick genomes. Both de novo assembly and read mapping sequences compared with near relatives can be used as approaches for recovering the genome sequences from metagenomic data. However, the FLE from *Amblyomma maculatum* (Gerhart et al. 2016), the *Rhipicephalus turanicus* CLE (Gottlieb et al. 2015), and the *Rickettsia* endosymbiont (REIS) from *Ixodes scapularis* (Gillespie et al. 2011) were all assembled as scaffolds from this type of data. A better sequence was obtained for REIS (now *Rickettsia buchneri* sp. nov.) from tick cell culture stocks of the agent (Kurtti et al. 2015). Similarly, *Francisella persica*, the symbiont of the soft tick *Argas persicus* was cultivated and then sequenced (Larson et al. 2016). The goal of the present investigation was to determine whether a less expensive approach could be used to characterize the genomes of tick symbionts than Illumina deep sequencing. We hoped that this approach would provide new

insight into a number of these organisms when neither cell culture nor pure isolates could be obtained. We also wanted to determine if sequences could be obtained from individual ticks since pooled DNA samples had been used previously with the *A. americanum* and *R. turanicus* CLE agents.

We chose the Ion Torrent Personal Genome Machine because it can provide 200–400 bp reads and a variety of inexpensive chips were available. It has been used to assemble relatively large and complete genome sequences of bacteria and for sequencing complete 16S rRNA molecules for microbiome analysis. The Ion library technology we employed could also be applied on the new more automated Ion S5 machine which has a bigger chip and greater automation, should greater read depth and more sampling be required.

1.2 Materials and Methods

1.2.1 Tick Collection and Storage, DNA Extraction and Storage

Host-seeking adult ticks (Table 1.1) were collected by dragging a cloth over vegetation (i.e. flagging) (California, Georgia, Ohio) (Bermúdez et al. 2009; Wikswo et al. 2008). Other ticks (Table 1.1) were purchased from colonies maintained at Oklahoma State University (OSU) under an NIAID (BEI Resources) contract. Origins of the ticks used for initiating the colonies were provided by OSU. Ticks were transferred into ≥70% ethanol for storage at 4 C until their DNA was extracted. Ticks were surface disinfected with sequential washes of 10% bleach and 70% ethanol, then rinsed 3 times in sterile distilled water, frozen in a close-fitting eppdendorf 1.5 ml centrifuge tube in a LN2 bath and pulverized with a Kontes pestle prior to DNA extraction with the Promega Wizard SV Genomic DNA Kit or Qiagen DNA easy tissue kit (Williams-Newkirk et al. 2014). DNA extracts were stored at 4 °C until screening by PCR for different agents and subsequent library preparation.

1.2.2 PCR Testing of Samples for Their Content of CLE, FLE, and Spotted Fever Group **Rickettsia**

Rickettsia were initially detected and quantitated in all tick DNAs by SsoFastEVA-Green qPCR (BioRad) amplification of an *omp*B fragment (Eremeeva et al. 2003), *Amblyomma americanum* and *A. maculatum* were also tested for *Rickettsia*, *Ehrlichia chaffeensis*, and *E. ewingii* with the multiplex TaqMan assay using iTaq Universal Probes Supermix (BioRad) of Killmaster et al. (2014). The identity of agents in positive ticks was determined and quantitated by subsequent qPCR with an *R. parkeri*-specific TaqMan assay (Jiang et al. 2012) or by DNA sequencing of a 5′ fragment of

Table 1.1 Summary of tick samples and metagenomic sequencing results on Ion Torrent-PGM

Tick samples	USA origin	Stage	Total number of useable reads	Percentage of useable total reads (%)	Median length of reads	IT-PGM chemistry	Repli-G method
Amblyomma maculatum Amac21	Georgia	Female	4,007,976	62	113 bp	200	No
Amblyomma maculatum Amac50	Oklahoma	Female	3,810,034	43	137 bp	200	No
Dermacentor variabilis Dvar22	Georgia	Female	6,211,538	66	244 bp	200	No
Dermacentor variabilis Dvar100	Ohio	Female	3,717,748	74	218 bp	200	No
Dermacentor variabilis DvarSlov	Virginia	Adult	5,810,262	67	227 bp	200	No
Dermacentor occidentalis DoccEMCF9	California	Female	6,518,864	69	243 bp	200	No
Dermacentor andersoni DandM9	Montana	Male	4,690,550	50	220 bp	400	No
Amblyomma americanum Aam46	Georgia	Male	5,242,310	52	272 bp	200	No
Amblyomma americanum Aam10SC	Georgia	Female	3,490,856	35	138 bp	200	Single cell
Rhipicephalus sanguineus RhsangF23	Oklahoma	Female	4,490,171	54	123 bp	200	Ultrafast
Dermacentor variabilis DvarF4M4	Oklahoma	Female	2,891,366	51	80 bp	200	Mini
Dermacentor variabilis DvarF4SC4	Oklahoma	Female	72,986[a]	14.6	206 bp	200	Single cell
Dermacentor variabilis DvarF1SC1	Oklahoma	Female	19,602[a]	6.2	107 bp	200	Single cell

[a]Ion Torrent 314 chips, rest 318 chips

*omp*A (Bermúdez et al. 2009; Wikswo et al. 2008). *Francisella*-like endosymbionts (FLE) were detected and quantitated by EVA-Green qPCR using the 16S rRNA primers NC-Fran16S-F and NC-Fran16S-R of Dergousoff et al. (2012). *Coxiella*-like endosymbionts were detected by standard PCR with 16S primers (Scoles 2004) or detected and quantitated by the *fus*A TaqMan assay with primers AAMFUSA-F and AAMFUSA-R and probe AAMFUSA-Pr of Jasinkas et al. (2007). All EVA-Green and TaqMan assays were performed on a CFX96 Touch Real-Time PCR Detection System (BioRad).

1.2.3 DNA Sequencing of OmpA Amplicons

Amplicons were purified from amplicon bands cut from 2% ethidium bromide-stained agarose gels with the Wizard SVGel and PCR clean-up system (Promega) and sequenced with the ABI BigDye Terminator v3.1 cycle sequencing kit (ABI) according to the manufacturer's recommendations on an ABI 3130*xl* genetic analyzer (Bermúdez et al. 2009). Reads from both forward and reverse directions were analyzed and assembled with DNAStar Lasergene and the sequence identify evaluated with BLASTn on the NCBI server.

1.2.4 IT-PGM Library Preparation and Sequencing

Tick DNA extracts were quantitated with the Qubit HS-DNA assay (Life Technologies). Tick DNA (250 ng) was treated with the Ion Xpress Fragment Library kit (Life Technologies #4471269). Following treatment (37 C for 8.5 min) with the Ion Shear Plus Reagents Kit (#4471248) and purification with the DNA Clean and Concentrator-10 Kit (Zymo Research), the Ion Plus Fragment Library Kit (Life Technologies #4471252) was then used to nick-repair the DNA and ligate the DNA (25 C for 15 min, 72 C for 5 min) to IT-PGM primers. Following purification and concentration with the DNA and Concentrator-10 kit, 330 bp fragments were selected and eluted with the E-Gel Agarose Gel Electrophoresis System on Size Select 2% gels (Life Technologies). The size-selected ligated fragments were then amplified with Platinum PCR SuperMix High buffer and Library Amplification Primer Mix from the Ion Plus Fragment library kit (Denature 5 min at 95 C, 8 cycles of 15 s at 95 C, 15 s at 58 C, and 1 min at 70 C). The library was purified and concentrated again with the DNA and Concentrator-10 kit and quantified with the Qubit HS-DNA assay. Templates were generated with 100 pM of library and beads enriched on the Ion Torrent One Touch 2 system with 200 or 400 bp V2 Template kits; sequencing was performed with the Ion Torrent Personal Genome Machine, according to the manufacturer's instructions.

1.2.5 Repli-G Procedures

Qiagen Repli-G mini (M) (>10 ng, 50 ul, 16 h at 30 C), ultrafast (UF) (>10 ng, 30 ul, 1.5 h at 30 C), or single cell (SC) (10–100 pg, 50 ul, 16 h 30 C) kits were used to amplify tick DNA in 25 µl reactions prior to preparation and sequencing of IT-PGM libraries using the full fragmentation and templating procedure described above.

1.2.6 Bioinformatic Analysis of IT-PGM Sequences

Default Ion Torrent PGM parameters for exclusion of sequences (polyclonal, low quality, and adapter dimer ISPs) were used. The reads were first analyzed with Geneious 8.05 (Kearse et al. 2012) by read mapping to available genome assemblies for FLE (*A. maculatum*, *D. variabilis*, *Francisella persica*), CLE (*A. maculatum*, *Rhipicephalus turanicus*, *R. sanguineus*) and known US spotted fever group *Rickettsia* agents and their plasmids (*R. montanensis*, *R. rhipicephali* CWPP 3-7-F6, *R. bellii* OSU85-389, *R. parkeri* Portsmouth, *R. amblyommatis* GAT-3OV). The Geneious contigs were then analyzed by BLASTn and annotated by Prokka (Seemann 2014). The reads were also analyzed by use of BWA v0.7.12 (Li and Durbin 2010) for read mapping to libraries of *Francisella*, *Coxiella*, *Rickettsia* genome sequences available at NCBI or to individual genome sequences. The mapped reads were extracted using SAMtools v1.3.1 (Li et al. 2009) and subsequently these reads were assembled into contigs using SPAdes v3.10.1 (Bankevich et al. 2012) or CLC Genomics Workbench v9.5.2 (https://www.qiagenbioinformatics.com/). Assembled contigs were then analyzed by BLASTn, Prokka, and by tBLASTn using NCBI protein lists for individual genomes as well as the limited sequences available from "*Candidatus* Rickettsia andeanae".

1.3 Results

1.3.1 Library Preparation and IT-PGM Sequence Reads Obtained

Thirteen metagenome libraries were prepared using Ion Torrent shearing and ligation reagents and templating on the One Touch 2 for sequencing on the Ion Torrent Personal Genome Machine (IT-PGM) (Table 1.1). Most of the difficulties that were encountered with achieving consistency in the recovery of useable sequencing reads were related to the manual loading of the 314 and 316 Ion chips and sizing of the fragmented libraries. The small 7 µl volume used on the 314 chip and limited number of sequence reads on that chip also made accurate library preparation and quantitation more crucial; however, several later loadings of other arthropod single

cell Repli-G libraries resulted in a greater recovery of useable reads than the initial two experiments described here (178,750 useable reads, 189 bp median read length; 275,890 useable reads, 211 bp median read length). For unknown reasons, more unusable polyclonal sequences were also obtained with the smaller chip; however, this chip is convenient and less expensive for gaining experience with alternative library preparation methods, including size fractionation and amplification methods. We obtained more consistent results with library yields by cutting out specific band sizes on SYBR-Green stained standard 2% agarose gels than with the E-gel system and by using the Promega SVS gel purification method to recover the DNA from the agarose, as used for DNA sequencing of amplicons. The variable results obtained with median read lengths on the E-gels can be directly attributed to uncertainty about when to stop the gel migration and recover the correct size products. We improved this procedure by using several different DNA ladders (the BioRad EZLoad 500 bp ladder was particularly useful) to achieve greater certainty about recovery of the correct sizes of fractionated and ligated SYBR-Green stained DNA either directly from the E-gel or cut from standard 2% agarose gels on a blue light box.

1.3.2 Repli-G Enhancement of Read Depth

Although alcohol-preserved single adult males or females were used to make all of the libraries sequenced here, tick nymphs and larvae (even as pools derived from the same engorged female) contain substantially less DNA than the adults and we wished to avoid introducing more polymorphisms by pooling samples. In other studies, we initially found that even adult head lice had too little DNA for accurate quantitation and library preparation, so we evaluated the use of mini, ultrafast, and single cell Repli-G kits from Qiagen to amplify the amounts of starting DNA to have enough for library preparations. Five of the 13 samples sequenced were amplified with one or another of these kits (Table 1.1). We found the single cell kit to be the most reliable and flexible because it can be used with the least input DNA and larger volumes of the dilute DNA obtained from our conventional DNA extractions (typically 100–200 μl volumes). Mini kits provided the least amplified material and less latitude in initial DNA volume and concentration. The ultrafast kit offered little advantage other than cost over the single cell kit and did not work as well with low concentration DNA extracts (not shown). In some cases we concentrated all of the low concentration starting material (with the DNA purification and concentration kits from Zymo) to 20 μl, but adequate yields of amplified material could be obtained even with samples below the sensitivity of the Qubit HS-DNA kit by using the single-cell Repli-G kit. We did not observe any difficulties during bioinformatic analysis of the Repli-G samples when compared to the standard libraries (Tables 1.2 and 1.3).

Table 1.2 Effect of Repli-G amplification on the relative recovery of sequence reads mapping with Geneious against FLE agents

		#Reads mapping to FLE references			FLE % total reads			
Ticks (State) Sex	**# IT-PGM Reads**	**AmacFLE**	**DvarHS28 FLE**	**Fpersica**	**AmacFLE**	**DvarHS28 FLE**	**Fpersica**	**Avg**
Aam10SC (GA) F	34,90,856	55	113	151	0.0016	0.0032	0.0043	0.0030
Aam46 (GA) M	52,42,310	179	51	54	0.0034	0.0010	0.0010	0.0018
Amac21 (GA) F	40,07,976	36393	31786	33512	0.9080	0.7931	0.8361	0.8457
Amac50 (OK) F	38,10,034	5036	1448	4557	0.1322	0.0380	0.1196	0.0966
DandM9 (MT) M	46,90,550	10758	11277	86596	0.2294	0.2404	1.8462	0.7720
DoccEMCF9A (CA) F	65,18,864	26002	26242	25124	0.3989	0.4026	0.3854	0.3956
Dvar22 (GA) F	62,11,538	60304	33810	15948	0.9708	0.5443	0.2567	0.5906
Dvar100 (OH) F	37,17,748	8034	12537	12104	0.2161	0.3372	0.3256	0.2930
DvarSlov (VA) adult	58,10,262	13859	9195	14054	0.2385	0.1583	0.2419	0.2129
DvarF4M4 (OK) F	28,91,366	51567	63247	59814	1.7835	2.1874	2.0687	2.0132
DvarF1SC1 (OK) F	19,602	1316	1063	1103	6.7136	5.4229	5.6270	5.9212
DvarF4SC4 (OK) F	72,986	3171	2709	2586	4.3447	3.7117	3.5431	3.8665
RhsangF23UF6 (OK) F	44,90,171	16974	14944	53130	0.3780	0.3328	1.1833	0.6314

Green: Positive identification of FLE with wide genomic coverage (large contigs or nearly complete)
Orange: Positive presence of FLE sequences but with low and interspersed coverage (many contigs)
Pink: Positive maps of reads to FLE but with very low coverage and big gaps (Few contigs)
White: Negative for FLE, reads map to sites of low complexity

Table 1.3 Effect of Repli-G amplification on the relative recovery of sequence reads mapping with Geneious against CLE agents

		#Reads mapping to CLE references			CLE % total reads			
Ticks (State) Sex	# IT-PGM Reads	AamGA-CLE	AamOK-CLEAA	CLE-Rt	AamGA-CLE	AamOK-CLEAA	CLE-Rt	Avg
Aam10SC (GA) F	34,90,856	14668	14646	1449	0.4202	0.4196	0.0415	0.2937
Aam46 (GA) M	52,42,310	539	3508	1721	0.0103	0.0669	0.0328	0.0367
Amac21 (GA) F	40,07,976	198	198	631	0.0049	0.0049	0.0157	0.0085
Amac50 (OK) F	38,10,034	3754	3755	5765	0.0985	0.0986	0.1513	0.1161
DandM9 (MT) M	46,90,550	190	195	2773	0.0041	0.0042	0.0591	0.0224
DoccEMCF9A (CA) F	65,18,864	111	111	829	0.0017	0.0017	0.0127	0.0054
Dvar22 (GA) F	62,11,538	13746	13749	16116	0.2213	0.2213	0.2595	0.2340
Dvar100 (OH) F	37,17,748	443	443	854	0.0119	0.0119	0.0230	0.0156
DvarSlov (VA) adult	58,10,262	19	19	1299	0.0003	0.0003	0.0224	0.0077
DvarF4M4 (OK) F	28,91,366	39141	39140	64474	1.3537	1.3537	2.2299	1.6458
DvarF1SC1 (OK) F	19,602	787	788	1012	4.0149	4.0200	5.1627	4.3992
DvarF4SC4 (OK) F	72,986	2290	2298	3001	3.1376	3.1485	4.1117	3.4660
RhsangF23UF6 (OK) F	44,90,171	694425	695342	2248213	15.4654	15.4859	50.0697	27.0070

Green: Positive identification of CLE with wide genomic coverage (large contigs or nearly complete)
Orange: Positive presence of CLE sequences but with low and interspersed coverage (many contigs)
Pink: Positive maps of reads to CLE but with very low coverage and big gaps (Few contigs)
White: Negative for CLE, reads map to sites of low complexity

1.3.3 *Tick Samples with* Francisella*-Like Endosymbionts (FLE)*

An EVA-Green PCR assay detecting a fragment of 16S rRNA was used to screen samples of 6 species of ticks for the presence of *Francisella*-like agents. Some samples (4 species, 5 samples, Table 1.4) with lower Ct values (20–25 using 1/200th of DNA extracted from a tick) were selected for further analysis from a larger set of samples obtained at the same time; these 5 samples, as well as two additional samples (DvarSlov and Amac21), were found to contain FLE by IT-PGM sequence analysis. None of these samples had been amplified by Repli-G prior to sequencing. Geneious 10.0.9 was used initially to map and assemble reads using the *A. maculatum* FLE (OK), which was in 7 scaffolds (Gerhart et al. 2016); *Francisella persica* VR331, which was isolated from the soft tick *Argas persicus* and subsequently cultivated in vitro before sequencing (Larson et al. 2016); and a partial sequence we have obtained and assembled from Illumina HiSeq sequencing of a Georgia female of *Dermacentor variabilis*, obtained from the same site as Dvar22GA (Dasch et al. 2016). Read mapping of the other tick sequence libraries without FLE provided an indication of the sensitivity and background of this approach to identifying FLE-specific sequences in these metagenomic libraries (Table 1.5).

The two *A. americanum* ticks (Aam10SC, Aam46) without detectable FLE by qPCR had low numbers of sequences, coverage, and percentage of genomes covered relative to those with FLE by Geneious, but higher numbers of reads, as well as background, in negative ticks were obtained with CLC (Table 1.4). Geneious 10.0.9 mapped more reads than Geneious 8.0.5 but while the read coverage increased in most cases, the percentage of the genome mapping remained largely unchanged (not shown). Neither the FLE qPCR Ct value nor the number of useable reads had any obvious relationship to the number of reads mapping to any of the three assembled FLE that are of similar size and have significant sequence homology (Table 1.4). In general, the total number of reads and percentage coverage were also prone to be misleading because some read maps had high read coverage at specific chromosome locations (appearing as spikes in the maps) but these were not consistent between the three reference FLE used for mapping. This inconsistency can be seen clearly in Fig. 1.1, especially in Fig. 1.1f, where abundant reads were clustered but only for *F. persica* VR331. These aberrant regions did not interfere with interpreting the data. For example, Amac21GA had similar high quality read coverage for all three FLE maps (Fig. 1.1a–c), while *Dermacentor andersoni* (Fig. 1.1d–f) had many gaps and lower and less uniform coverage along the chromosome, particularly with respect to *F. persica* (Fig. 1.1f). Therefore, the percentage of the FLE sequence with read coverage was more consistent with all three FLE reference sequences than was the number of reads mapping or median read coverage; that parameter gave the best indication of how much related FLE sequence had been obtained. While sequences unique to the new FLE agents (i.e., not present in any of the reference genomes), cannot be evaluated accurately without full genome coverage and assembly, it is clear that the Amac21 FLE agent from Georgia mapped similarly to that of the *A.*

Table 1.4 Bioinformatic analysis of IT-PGM sequence libraries for the presence of CLE and FLE agent sequences

Ticks (State) Sex	# IT-PGM Reads	FLE-16S qPCR	AmacFLE (OK) (G)	AmacFLE (OK) (CLC)	DvarHS28 FLE (GA) (G)	DvarHS28 FLE (GA) (CLC)	F persica VR331 (G)	F persica VR331 (CLC)
Size of ref seq (bp)			15,56,261	15,56,261	15,75,539	15,75,539	15,16,676	15,16,676
Aam10SC (GA) F	34,90,856	neg	55 (0.0x, 0.1%)	13,392 (2.8x, 11.4%)	113 (0.0x, 0.1%)	16,385 (3.2x, 11.4%)	151 (0.0x, 0.1%)	14,381 (3.0x, 11.7%)
Aam46 (GA) M	52,42,310	neg	179 (0.0x, 0.2%)	6,681 (2.6x, 6.1%)	51 (0.0x, 0.2%)	7,871 (2.9x, 6.1%)	54 (0.0x, 0.2%)	6,849 (2.6x, 6.3%)
Amac21 (GA) F	40,07,976	nd	36,393 (2.3x, 79.4%)	28,425 (2.3x, 80.9%)	31,786 (1.8x, 74.7%)	31,042 (2.3x, 76.5%)	33,512 (2.1x, 70.6%)	27,070 (2.3x, 73.3%)
Amac50 (OK) F	38,10,034	Ct 25.1	5,036 (0.4X, 12.4%)	7,110 (1.6x, 17.3%)	1,448 (0.1x, 11.6%)	8,203 (1.7x, 16.5%)	4,557 (0.4x, 10.5%)	7,412 (1.7x, 16.1%)
DandM9 (MT) M	46,90,550	Ct 25.9	10,758 (0.8x, 25.8%)	16,092 (2.2x, 34.4%)	11,277 (0.8x, 26.0%)	18,638 (2.3x, 34.6%)	86,596 (23.5x, 23.6%)	16,315 (2.3x, 33.0%)
DoccEMCF9A (CA) F	65,18,864	Ct 25.4	26,002 (1.8x, 67.5%)	12,559 (1.9x, 69.1%)	26,242 (1.9x, 68.4%)	16,741 (2.1x, 69.9%)	25,124 (1.9x, 62.6%)	13,469 (2.0x, 65.9%)
Dvar22 (GA) F	62,11,538	Ct 22.0	60,304 (11.6x, 26.9%)	6,271 (1.4x, 29.8%)	33,810 (4.1x, 28.3%)	9,809 (1.6x, 31.0%)	15,948 (0.9x, 24.5%)	7,141 (1.6x, 28.2%)
Dvar100 (OH) F	37,17,748	Ct 21.5	8,034 (0.7x, 37.9%)	12,230 (0.8x, 38.2%)	12,537 (1.0x, 40.5%)	8,920 (1.6x, 42.0%)	12,104 (0.8x, 35.8%)	7,180 (1.6x, 38.4%)
DvarSlov (VA) adult	58,10,262	nd	13,859 (0.7x, 24.2%)	5,482 (1.4x, 26.9%)	9,195 (0.5x, 25.6%)	14,054 (0.7x, 25.7%)	14,054 (0.7x, 25.7%)	6,455 (1.5x, 25.9%)
DvarF4M4 (OK) F	28,91,366	neg	51,567 (4.4x, 2.8%)	88,370 (8.4x, 28.7%)	63,247 (5.6x, 3.0%)	97,637 (9.0x, 28.5%)	59,814 (4.7x, 3.2%)	93,438 (8.6x, 30.0%)
DvarF1SC1 (OK) F	19,602	neg	1,316 (0.2x, 14.3%)	1,066 (2.1x, 1.3%)	1,063 (0.7x, 50.0%)	1,309 (0.2x, 13.0%)	1,103 (0.7x, 51.5%)	1,357 (0.4x, 16.2%)
DvarF4SC4 (OK) F	72,986	neg	3,171 (1.0x, 34.3%)	1,903 (2.7x, 2.3%)	2,709 (1.4x, 64.2%)	3,306 (1.1x, 32.9%)	2,586 (1.3x, 61.3%)	3,263 (0.9x, 30.7%)
RhsangF23UF6 (OK) F	44,90,171	neg	16,974 (1.9x, 1.8%)	26,820 (7.1x, 14.4%)	14,944 (1.6x, 1.7%)	34,617 (8.1x, 14.2%)	53,130 (4.3x, 2,1%)	30,442 (7.8x, 15.2%)

Green: Positive identification of FLE with wide genomic coverage (large contigs or nearly complete)
Orange: Positive maps of reads to FLE but with very low coverage and big gaps (Few contigs)
Pink: Positive maps of reads to FLE but with very low coverage and big gaps (Few contigs)
White: Negative for FLE, reads map to sites of low complexity
(G) reads mapping with Geneious 10.0.9
(CLC) reads mapping with CLC Workbench 8
Boxes: number of reads mapping (median read coverage per base, percentage of reference genome mapped)
nd: not done

Table 1.5 Bioinformatic analysis of IT-PGM sequence libraries for the presence of CLE agent sequences

Ticks (State) Sex	# IT-PGM Reads	AamGA-CLE (G)	CLEAA (OK) (G)	CLEAA (OK) (CLC)	CLE-RsangF23 Final (G)	CLE-Rturanicus (G)	CLE-Rturanicus (CLC)
Size of ref seq (bp)		656,933	656,901	656,901	1,115,981	1,733,840	1,733,840
Aam10SC (GA) F	3,490,856	14,668 (2.0x, 35.9%)	14,646 (2.0x, 35.8%)	8,485 (1.9x, 43.6%)	13,015 (1.3x, 1.3%)	1,449 (0.2x, 1.0%)	15,074 (3.3x, 12.4%)
Aam46 (GA) M	5,242,310	539 (0.1x, 3.6%)	3,508 (0.5x, 3.6%)	3,338 (2.3x, 9.7%)	1,081 (0.1x, 0.5%)	1,721 (0.6X, 0.3%)	7,449 (3.7x, 5.4%)
Amac21 (GA) F	4,007,976	198 (0.0x, 0.8%)	198 (0.0x, 0.8%)	3,742 (3.5x, 7.8%)	750 (0.1x, 0.5%)	631 (0.1x, 0.4%)	7,568 (4.0x, 5.4%)
Amac50 (OK) F	3,810,034	3,754 (0.9x, 0.2%)	3,755 (0.9x, 0.2%)	2,990 (2.9x, 6.5%)	7,067 (2.5x, 0.2%)	5,765 (0.7x, 0.2%)	8,296 (6.1x, 4.8%)
DandM9 (MT) M	4,690,550	190 (0.1x, 0.2%)	195 (0.1x, 0.2%)	6,923 (3.3x, 12.1%)	5,577 (3.8x, 0.2%)	2,773 (0.5x, 0.2%)	17,180 (5.6x, 9.5%)
DoccEMCF9A (CA) F	6,518,864	111 (0.0x, 0.6%)	111 (0.0x, 0.6%)	2,704 (4.5x, 4.7%)	18,822 (1.5x, 0.4%)	829 (0.1x, 0.3%)	7,166 (8.2x, 3.3%)
Dvar22 (GA) F	6,211,538	13,746 (3.3x, 0.3%)	13,749 (3.3x, 0.3%)	2,310 (2.9x, 4.4%)	15,254 (1.5X, 0.2%)	16,116 (2.2x, 0.2%)	6,706 (7.6x, 3.2%)
Dvar100 (OH) F	3,717,748	443 (0.1x, 0.8%)	443 (0.1x, 0.8%)	1,838 (6.2x, 3.7%)	613 (0.1x, 0.5%)	854 (0.1X, 0.3%)	4,578 (9.3x, 2.2%)
DvarSlov (VA) adult	5,810,262	19 (0.0x, 0.3%)	19 (0.0x, 0.3%)	1,962 (2.9x, 3.8%)	831 (0.2x, 0.2%)	1,299 (0.1x, 0.2%)	5,969 (7.7x, 2.8%)
DvarF4M4 (OK) F	2,891,366	39,141 (6.0x, 10.0%)	39,140 (6.0x, 10.0%)	43,190 (8.4x, 37.1%)	54,919 (4.7x, 70.0%)	64,474 (3.6x, 62.5%)	96,820 (4.5x, 69.3%)
DvarF1SC1 (OK) F	19,602	787 (0.6x, 30.6%)	788 (0.6x, 29.7%)	482 (2.3x, 1.5%)	860 (0.4x, 14.0%)	1012 (0.3x, 14.1%)	662 (1.7x, 1.0%)
DvarF4SC4 (OK) F	72,986	2,290 (1.4x, 38.6%)	2,298 (1.5x, 41.4%)	996 (2.9x, 3.7%)	2,569 (0.7x, 24.0%)	3,001 (0.7x, 23.3%)	1,921 (1.6x, 5.1%)
RhsangF23UF6 (OK) F	4,490,171	694,425 (170.8, 66.1%)	695,342 (170.6x, 66.1%)	254,141 (88.4x, 60.5%)	1,675,572 (219.5X, 100%)	2,248,213 (191.0, 97.2%)	2,366,498 (192.8x, 95.6%)

Green: Positive identification of CLE with wide genomic coverage (large contigs or nearly complete)
Orange: Positive presence of CLE sequences but with low and interspersed coverage (many contigs)
Pink: Positive maps of reads to CLE but with very low coverage and big gaps (few contigs)
White: Negative for CLE, reads map to sites of low complexity
(G) reads mapping with Geneious 10.0.9
(CLC) reads mapping with CLC Workbench 8
boxes: number of reads mapping (median read coverage per base, percentage of reference genome mapped)
nd: not done

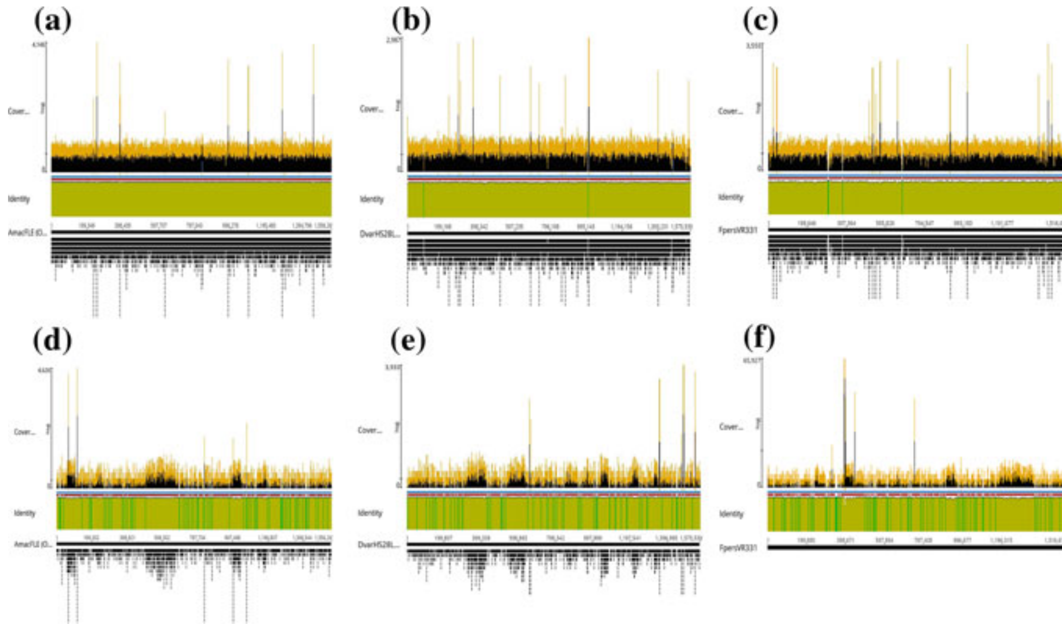

Fig. 1.1 Geneious read mapping of Amac21 and DandM9 IT-PGM sequences against *Francisella*—like endosymbiont (FLE) reference genomes. **a–c** Amac21, **d–f** DandM9; **a**, **d** *Amblyomma maculatum* OK FLE; **b**, **e** *Dermacentor variabilis* HS28 FLE; **c**, **f** *Francisella persica* VR331. Identity bar over coverage at that coordinate: greeny-brown (100%), green (low or no coverage). Coverage: log scale, black mean coverage, orange maximum coverage, blue bar minimum coverage greater than 5, red bar, contiguous coverage, black bars-reads mapped (**f** not displayed because of very high coverage spikes)

maculatum FLE agent from the Oklahoma tick colony and slightly less well to the *D. variabilis* HS28 FLE or *F. persica*. This difference was confirmed by Prokka analysis of the predicted ORFs from their consensus sequences. However, while all of the positive libraries had reads distributed along the reference target FLE sequences, only 10–80% of the reference sequence was obtained in the assemblies. CLC generally gave comparable positive coverage of read maps to Geneious with the same specific individual reference FLE sequences but the reads and contigs recovered had no predictable trend to each. This inconsistency is probably due to the relatively low genome coverage obtained because the majority of the sequence reads obtained were actually from the tick's large genome. However, the *D. andersoni* and *D. occidentalis* FLE appear to have genomes similar in size to the three reference FLE agents.

A higher percentage of the target FLE genomes was recovered by mapping IT-PGM sequences with a database comprised of all complete *Francisella* genomes and the FLE sequences by use of BWA, followed by SPAdes or CLC workbench assembly (not shown). This approach increased the expected genome coverage of novel FLE by Prokka annotation. While identification of much of the expected complement of core genes had been obtained, many protein sequences were fragmented, so it was difficult to discern if those proteins were pseudogenes undergoing degradation or if this breakage was solely an artefact of insufficient read depth.

Four libraries (RhsangF23, DvarF4M4, DvarF4SC4, DvarF1SC1) from three ticks from the Oklahoma tick colony provided surprising data from the FLE read mapping. All were negative by FLE 16S qPCR. Their DNAs had been Repli-G amplified, in the hope that their bacterial sequences would be selectively amplified relative to the host tick DNA, as well as to obtain more DNA for library preparation. This selective amplification appeared to have occurred, since the average of the percentage of FLE-mapped reads compared to the total number of reads obtained was higher with those samples treated with the Repli-G kits (Table 1.2). The RhsangF23 sequences mapped to the FLE reference sequences but in greatest abundance at specific locations across the chromosome (Fig. 1.2a, b), in contrast to the broader coverage seen with homologous mapping of Amac21 (GA) to its highly related *A. maculatum* OK FLE agent. This abundance and site restriction suggested that the RhsangF23 bacterium was not a FLE agent and that another organism with sequence similarity at the core genome level was being detected. This conclusion is consistent with the FLE qPCR results. A similar result was also obtained for the DvarF4M4 sequences, with even higher FLE coverage (Fig. 1.3a, b). These results are also consistent with the idea that these sequences are not spurious artefacts due to the Repli-G amplification.

Despite the incomplete genome coverage, several features of these previously uncharacterized FLE genomes in *D. andersoni*, *D. variabilis*, and *D. occidentalis* could be discerned. (1) The genome sizes of the novel FLE, as well as for the one from Georgia *A. maculatum*, clearly approached that of the few FLE sequences now available. (2) The sequences obtained were accurate and permitted identification of SNPs and INDELs in comparison to the three assembled FLE available. (3) The base composition and contiguity of a number of adjacent genes could be confirmed as well as intergenic sequences between contiguous genes. (4) tRNAs and much of the rRNA sequences could be extracted. (5) Housekeeping genes being analyzed for a FLE-specific endosymbiont Multiple Locus Sequence Typing (MLST) protocol could be compared to identify conserved primer sites for amplifying these genes. On the other hand, the IT-PGM data was insufficient for completing the assembly of the 7 contigs for *A. maculatum* OK FLE or for improved scaffolding of the partial *D. variabilis* (Dvar28) FLE agent assembly we had obtained previously from Illumina HiSeq data.

1.3.4 Tick Samples with **Coxiella**-*Like Endosymbionts (CLE)*

Three CLE sequences were available to us for read mapping the IT-PGM sequences (Table 1.5). Two sequences were the highly genome-reduced CLE agent from *Amblyomma americanum*, one from a pool of Oklahoma State University colony ticks (CLEAA-OK) (Smith et al. 2015) and the other a highly related sequence from our own work: AamGA-CLE was assembled from Illumina Hi-Seq 2500 sequences obtained from a single female tick from GA (Williams-Newkirk et al. 2015). The two sequences are nearly the same length but differ in 360 sites (39 INDELS, 321 SNPs) scattered over the chromosome (Ramaiah et al. 2017). The other, much larger CLE

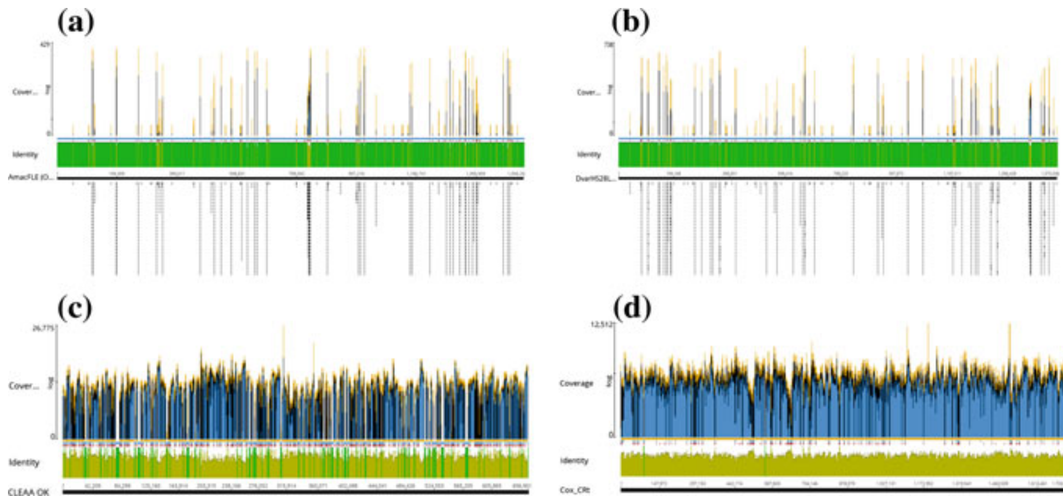

Fig. 1.2 Geneious read mapping of RhsangF23 IT-PGM sequences against *Francisella*—like endosymbiont (FLE) and *Coxiella*-like endosymbiont (CLE) reference genomes. **a** *Amblyomma maculatum* OK FLE; **b** *Dermacentor variabilis* HS28 FLE, *Francisella persica* VR331; **c** *Amblyomma americanum* OK CLE (CLEAA); **d** *Rhipicephalus turanicus* FLE. See Fig. 1.1 for other details

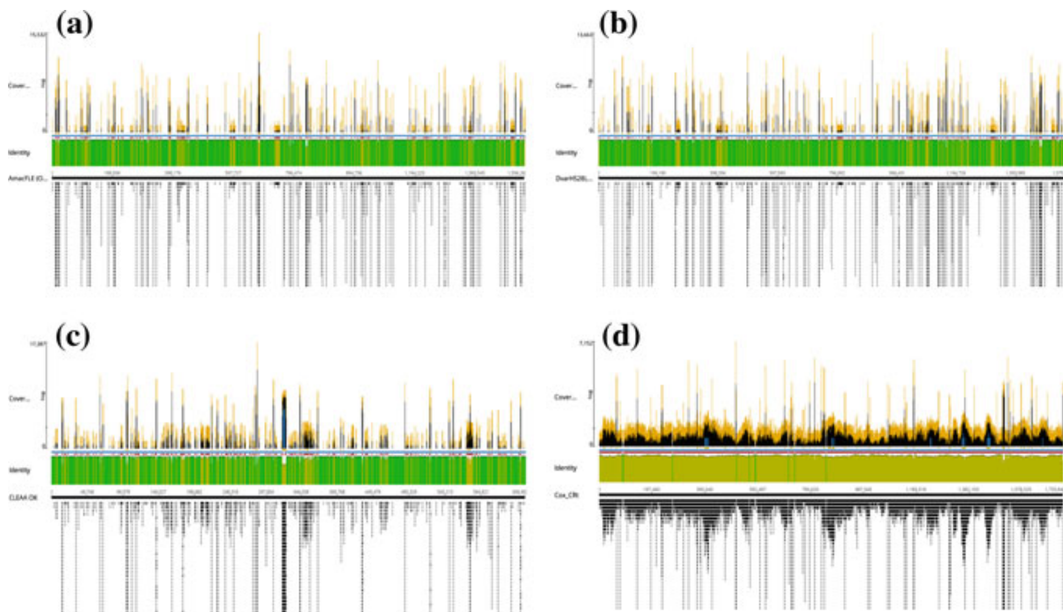

Fig. 1.3 Geneious read mapping of DvarF4M4 IT-PGM sequences against *Francisella*—like endosymbiont (FLE) and *Coxiella*-like endosymbiont (CLE) reference genomes. **a** *Amblyomma maculatum* OK FLE; **b** *Dermacentor variabilis* HS28 FLE, *Francisella persica* VR331; **c** *Amblyomma americanum* OK CLE (CLEAA); **d** *Rhipicephalus turanicus* FLE. See Fig. 1.1 for other details

sequence is from a pool of tissues from nine *Rhipicephalus turanicus* from Israel and sequence polymorphisms were detected between CLE in individual field collected ticks (Gottlieb et al. 2015). It is also smaller than the genome of the pathogen *Coxiella burnetii*. We could detect a large portion of the CLE agent in both of our samples of *A. americanum* [one from GA obtained at the same time as the sample used for Illumina sequencing, another from the same Oklahoma State University colony used for CLEAA by Smith et al. (2015)] as well as a larger but partial CLE sequence assembled from RhsangF23 (Ramaiah et al. 2017). We used the CLC workbench assembly we got for CLE-Rs from the latter tick to do read mapping of all the other samples with Geneious. As was expected, the CLE-Rs assembly remapped in Geneious to a mean coverage of 216.2 and 100% of this reference sequence (Table 1.5). CLE has not been reported previously from *Dermacentor variabilis* and, while we detected it in both of the Oklahoma tick colony samples, we also detected it in the 3 *D. variabilis* from Virginia, Ohio, and Georgia (Table 1.5). The CLE in both *Rhipicephalus sanguineus* and the two Oklahoma samples of *D. variabilis* (F4M4/F4SC4 and F1SC1) also mapped to the 3 FLE sequences as previously noted (Table 1.5), which is unsurprising, given that both agents are in the gamma proteobacteria. Similarly, the FLE agents found in eastern *D. variabilis* from Virginia, Ohio, and Georgia and in *D. andersoni*, *D. occidentalis*, and *Amblyomma maculatum* were also detected by read mapping with CLE but at low levels. In some cases, the relative numbers of reads mapped to CLE and FLE for a tick permitted direct determination of the correct agent but it was obvious from the Geneious read mapping that the FLE agent sequences were more closely related to the sequences obtained for both RhsangF23 (Fig. 1.2c, d) and for DvarF4M4 (Fig. 1.3c, d). Indeed, both of these CLE agents were much closer to the *Rhipicephalus turanicus* lineage than the smaller 657 Kb *A. maculatum* CLE and they both had reads mapping to much of the 2.5 times larger CRt genome sequence. However, while highly related, these two new CLE agents do not appear to be identical because the RhsangF23 agent mapped more closely to the A. *americanum* CLE agent (Fig. 1.2c) as did DvarF4M4 (Fig. 1.3c). It is possible that the greater read coverage for RhsangF23 may account for this difference, since DvarF4M4 mapped to 70% of CLE-Rsanguineus (Table 1.5). PROKKA, BLASTn, and tBLASTn analyses of the two assembled contigs clearly confirmed their correct assignments to *Coxiella* and not to *Francisella* (not shown).

The CLE Geneious read mapping data again demonstrated an apparent selective enhancement of the number of bacterial reads following Repli-G amplification (Table 1.3).

1.3.5 **Amblyomma americanum** ***Tick Samples Containing*** **Rickettsia amblyommatis** ***and AamGA-CLE***

The genome sequence of the reference strain *Rickettsia amblyommii* GAT-30V [recently renamed as *R. amblyommatis*—(Karpathy et al. 2016)] was obtained from DNA purified from purified rickettsiae obtained from a cell cultivated isolate from

the ovaries of a female tick (by G. Dasch) from Georgia (Panola Mountain State Park) by Roche 454 and Sanger sequencing (NCBI Assembly GCA_000284055.1). It contains a 1,407,796 bp chromosome and, atypically for species of *Rickettsia*, 3 plasmids (pRam1-18,263 bp, pRam2-22,851 bp, and pRam3-31,974 bp) (Table 1.6). Two other GA tick samples from Barnesville (Midland Piedmont), GA were obtained in June 2014; one adult female (Aam10) underwent Repli-G single cell amplification and the other adult male (Aam46) was not amplified before IT-PGM sequencing (Tables 1.1 and 1.6). Neither sample had high levels of *Rickettsia amblyommatis* (Ct 25.3 and 24.3, respectively) when compared to many GA ticks (Ct of 20 is common) but good assemblies of both the chromosome and all three plasmids were obtained with both the Geneious 10.0.9 and the BWA-SPAdes/CLC assembly pipelines (Table 1.6). The Geneious maps for the chromosome and plasmids are shown in Fig. 1.4 for Aam46, which had greater read depth, more useable reads, and a larger median sequence read length (Table 1.1). While amplification with Repli-G did not appear to improve the final results in this case (because the library preparation and chip loading was suboptimal for Aam10), much less of the original tick DNA extracted was needed for the analysis so it was still available for other PCR analyses or additional deep sequencing. The coverage of all 3 plasmids was 100% for both ticks and nearly identical with that of the reference isolate GAT-30V (Table 1.7). The plasmids read mapped much the same with either Geneious or CLC analysis but the lower *R. amblyommatis* coverage was better analyzed by CLC. The same bias was observed for the CLE agent found in each tick, but the relative amount detected was substantially higher in Aam10 than in Aam46 (Table 1.7). The relative amounts of *R. amblyommatis* and CLE have been found to vary in different adult ticks by both quantitative PCR and by microbiome analysis (Zhong et al. 2007; Williams-Newkirk et al. 2014). In Oklahoma *Amblyomma americanum*, the males often lack detectable *R. amblyommatis* and they are lower in abundance in general than the CLE (Zhong et al. 2007). Georgia ticks generally have consistently high amounts (Williams-Newkirk et al. 2014); the Aam46 Georgia male used here clearly had abundant *Rickettsia*. Whether the relative abundance of *R. amblyommatis* and CLE can also affect the relative sequence read coverage of each agent obtained on the IT-PGM is difficult to know without more testing of samples with known relative amounts of each agent. Since both agents are so abundant in most ticks, that could be evaluated on the IT-PGM by barcoding of samples so they can be pooled and assayed on the same sequencing chip.

1.3.6 **Amblyomma maculatum** ***Tick Samples Containing*** **Rickettsia parkeri** ***or*** **R. andeanae** ***and Amac-FLE***

The genome sequence of the human reference isolate *Rickettsia parkeri* Portsmouth was obtained from a human case of spotted fever rickettsiosis that occurred in the Tidewater area of Virginia (Whitman et al. 2007). It has a genome size similar to those of other core (classic) spotted fever rickettsiae, including *R. rickettsii*, *R. peacockii*,

Table 1.6 Comparison of the chromosome and plasmid assemblies obtained with Geneious (G) and CLC genomics workbench (CLC) from IT-PGM sequence libraries obtained from two Georgia *Amblyomma americanum* female ticks with those of the reference Georgia isolate *Rickettsia amblyommatis* GAT-3OV and the CLEAA agent from the Oklahoma *A. americanum* colony tick

R. amblyommatis domain	GAT-3OV size	Aam10 (G)	Aam10 (CLC)	Aam46 (G)	Aam46 (CLC)
Chromosome	1,407,796 bp	263,648 bp	399,706 bp	988,360 bp	1,017,362 bp
pRam1	18,263 bp	18,263 bp	18,263 bp	18,267 bp	18,263 bp
pRam2	22,851 bp	22,851 bp	22,844 bp	22,851 bp	22,855 bp
pRam3	31,974 bp	31,930 bp	31,965 bp	31,974 bp	32,005 bp
	AamOK CLEAA	Aam10 (G)	Aam10 (CLC)	Aam46 (G)	Aam46 (CLC)
CLEAA-AamOK domain	656,901 bp	235,285 bp	286,832 bp	23,779 bp	64,192 bp
Total IT-PGM reads		**3,490,856**		**5,242,310**	

Table 1.7 Geneious 10.0.9 bioinformatic analysis of IT-PGM sequence libraries for the presence of Rickettsia agent chromosome and plasmid sequences

Ticks	# IT-PGM Reads	*R. ambylommatis* GAT-3OV	pRam1	pRam2	pRam3	*R. peacockii* Rustic	pRpr1
Size of ref		14,07,796	18,263	22,851	31,974	12,88,492	26,406
Amac21 (GA)	40,07,976	24,697 (1.3x, 54.2%)	8 (0.0x, 4.5%)	0 (0.0x, 0%)	1 (0.0x, 0.3%)	17,880 (1.2x, 57.9%)	1 (0.0x, 0.4%)
Amac50 (OK)	38,10,034	14,270 (1.3x, 60.4%)	1,389 (10.6x, 87.0%)	597 (3.6x, 38.4%)	786 (3.4x, 46.7%)	13,167 (1.2x, 60.4%)	165 (0.9x, 16.3%)
DandM9	46,90,550	7,706 (0.5x, 0.1%)	0 (0.0x, 0%)	1 (0.0x, 0.1%)	1 (0.0x, 0.1%)	43 (0.0x, 0.1%)	0 (0.0x, 0.0%)
DoccEMCF9A(CA)	65,18,864	28,491 (3.1x 74.3%)	20,134 (94.7x, 82.2%)	593 (6.4x, 21.3%)	18,008 (38.5x, 15.7%)	27,993 (2.8x, 76.5%)	94 (0.9x, 3.3%)
Aam10 (GA)	34,90,856	20,938 (2.2x, 18.7%)	172,806 (1519.7x, 100.0%)	97,707 (740.6x, 100.0%)	60,630 (491.0x, 99.9%)	28,507 (4.1x, 18.1%)	22,997 (232.3x, 22.9%)
Aam46 (GA)	52,42,310	25,943 (7.0x, 70.2%)	322,013 (4352.8x, 100.0%)	278,577 (3110.8x, 100.0%)	138,142 (1553.1x, 100.0%)	40,377 (12.3x, 65.6%)	59,287 (979.6x, 25.7%)
RhsangF23 UF6(OSU)	44,90,171	107,536 (9.2x, 87.0%)	1 (0.0x, 0.8%)	19 (0.1x, 1.4%)	21 (0.1x, 1.3%)	103,933 (10.0x, 90.9%)	32,840 (155.8, 0.3%)
Dvar22 (GA)	62,11,538	13,249 (1.1x, 0.1%)	1 (0.0x, 0.1%)	0 (0.0x, 0%)	0 (0.0x, 0%)	9 (0.0x, 0.1%)	0 (0.0x, 0%)
Dvar100 (OH)	37,17,748	65,754 (9.1x, 68.4%)	1,090 (14.5x, 14.4%)	1,845 (18.4x, 33.3%)	3,038 (22.0x, 24.0%)	63,901 (9.7x, 73.2%)	1,556 (13.7x, 24.7%)
DvarSlov(VA)	58,10,262	12 (0.0x, 0.1%)	0 (0.0x, 0%)	0 (0.0x, 0%)	0 (0.0x, 0%)	9 (0.0x, 0.1%)	0 (0.0x, 0%)
DvarF4 M4(OK)	28,91,366	1,884,592 (137.3x, 88.9%)	152 (0.8x, 47.5%)	977 (5.4x, 8.1%)	2,464 (8.8x, 9.1%)	1,807,884 (142.4x, 93.2%)	30 (0.1x, 0.8%)
DvarF1 SC1 (OK)	19,602	17,223 (1.7x, 71.9%)	277 (2.9x, 72.1%)	206 (1.0x, 32.7%)	226 (2.1x, 42.1%)	16,583 (1.8x, 73.3%)	294 (3.0x, 54.2%)
DvarF4 SC4 (OK)	72,986	61,536 (8.5x, 90.0%)	11 (2.2x, 62.5%)	277 (5.4x, 87.3%)	460 (4.3x, 52.2%)	59,263 (9.0x, 95.0%)	572 (7.7x, 69.8%)

Green: Positive identification of *Rickettsia* with wide genomic coverage (large contigs or nearly complete)
Orange: Positive presence of *Rickettsia* sequences but with low and interspersed coverage (many contigs)
Pink: Positive maps of reads to *Rickettsia* but with very low coverage and big gaps (few contigs)
White: Negative for *Rickettsia* , reads map to sites of low complexity
boxes: number of reads mapping (median read coverage per base, percentage of reference genome mapped)

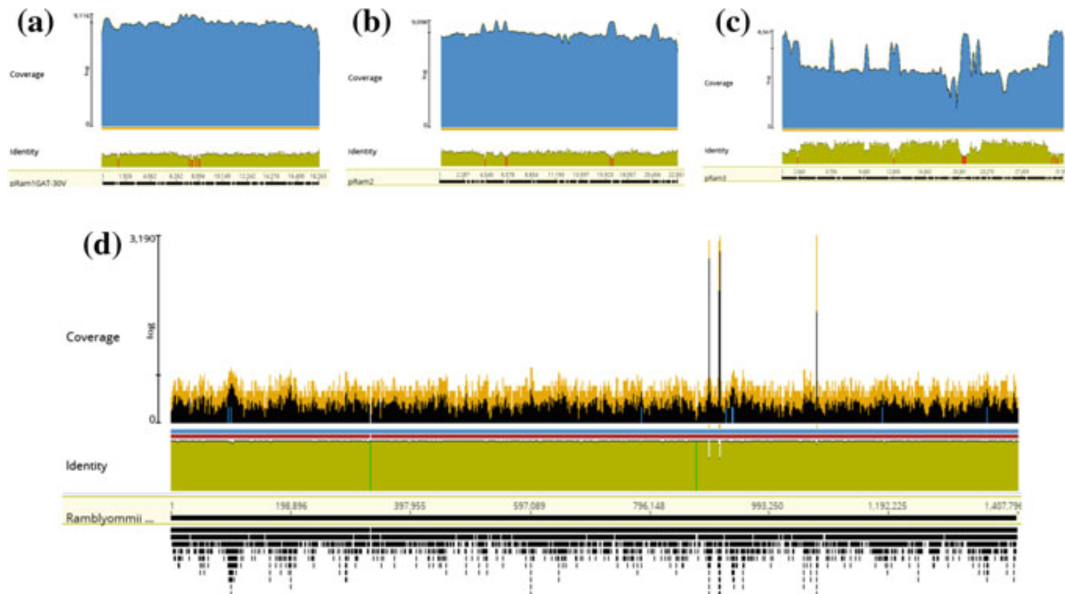

Fig. 1.4 Geneious read mapping of Aam46 IT-PGM sequences against *Rickettsia amblyommatis* GAT-30V reference genome. **a** Plasmid *pRam1;* **b** plasmid *pRam2*; **c** plasmid *pRam3*; **d** chromosome. See Fig. 1.1 for other details

and *R. philipii* in the USA, *R. sibirica* in Europe and Asia, and *R. conorii* in Europe, Africa, and the Indian subcontinent, and it lacks a plasmid (Table 1.8). *Amblyomma maculatum* are widely infected with *R. parkeri* at high rates of infection (Pagac et al. 2014). We analyzed a female tick (Amac21) collected in June 2014 from Barnesville (Midland Piedmont), GA (Table 1.1), which was shown to be infected with *R. parkeri* and the *A. maculatum* FLE agent by qPCR. A slightly better assembly was obtained with CLC and the total number of reads mapping to Portsmouth was greater than observed with other spotted fever group rickettsiae (Tables 1.8, 1.9 and 1.10). As found for *R. amblyommati*s in *A. americanum* ticks, the IT-PGM read coverage was not sufficient for full genome assembly but BLAST analysis of protein contigs identified in Prokka confirmed the tick contained *R. parkeri*. The second *A. maculatum* female tick (Amac50), obtained from the Oklahoma State University Tick Colony, contained a spotted fever *Rickettsia* that was not detected by the *R. parkeri* specific assay and was presumed to be "*Candidatus* R. andeanae" (hereafter *R. andeanae*), which is common in ticks from Oklahoma and present with varying abundance in other populations of *A. maculatum* (Fornadel et al. 2011; Jiang et al. 2012; Paddock et al. 2015). Because no genome sequence has been obtained yet for *R. andeanae*, in part because it has proven difficult to obtain an isolate in sustained cell culture (Luce-Fedrow et al. 2012), we employed Geneious and CLC read mapping of the IT-PGM Amac50 sequences against a number of other spotted fever rickettsiae with available genome sequences in order to extract its reads and to assemble its genome. We confirmed our hypothesis about the identity of its *Rickettsia* by comparing our partial sequence to several gene fragments available for *R. andeanae* (*sca*4-GU395298, *omp*B-GU395297, AY652981, *omp*A-GU395296, AY590796) and affirmed their ori-

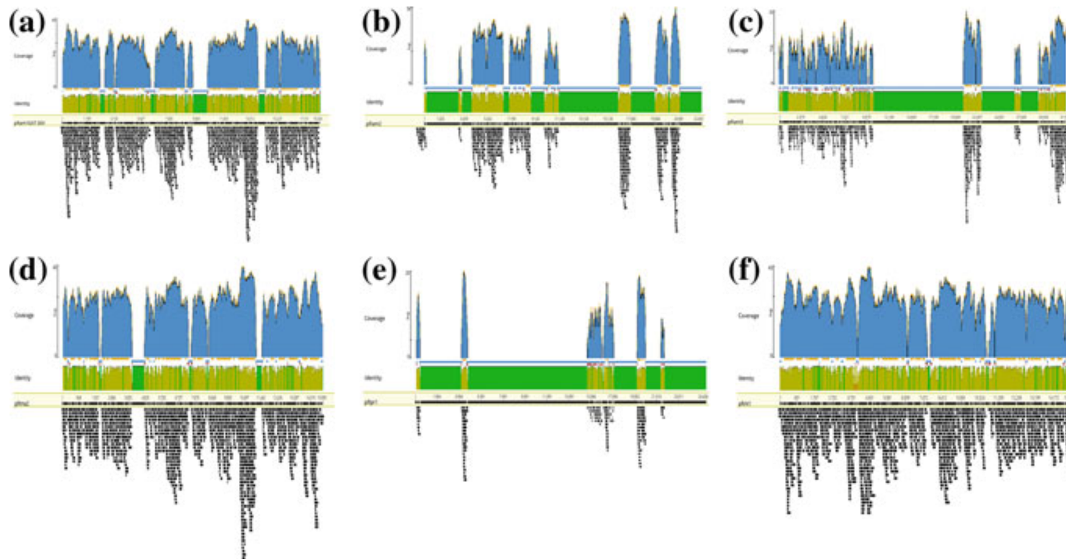

Fig. 1.5 Geneious read mapping of Amac50 IT-PGM sequences against selected *Rickettsia* plasmids. **a–c** *R. amblyommatis* GAT-30V pRam1, pRam2, pRam3; **d** R. massiliae AZT80 pMa2; **e** *R. peacockii* Rustic pRpr1; **f** *R. rhipicephali* 3-7-F6 CWPP pRrh1. See Fig. 1.1 for other details

gin. The prior phylogenetic placement of *R. andeanae* in a clade with *R. montanensis*, *R. rhipicephali*, *R. massiliae*, and *R. amblyommatis* was confirmed as well. Since all of these species except *R. montanensis* contain a plasmid, we also mapped the Amac50 sequence reads against six plasmid sequences from this clade and to the plasmid of *R. peacockii* Rustic with Geneious 10.0.9 (Fig. 1.5). This analysis confirmed that *R. andeanae* has a plasmid that is most related to those of the similar plasmids from *R. rhipicephalus* 3-7-F6 (pRrh1) and *R. massiliae* AZT80 (pRma2) (Fig. 1.5d, f; Tables 1.7, 1.9 and 1.10); pRam1, the smallest *R. amblyommatis* plasmid and one most similar to pRh1 and pRma2, also mapped to many Amac50 reads (Fig. 1.5a), while both pRam2 and pRam3 had regions of high homology (Fig. 1.5b, c) to the plasmid (name proposed here as pRan1) of *R. andeanae* but it had very incomplete coverage. Similarly, *R. peacockii* plasmid pRpR1 had only a few domains with high similarity (Fig. 1.5e). Because pRrh1, pRma2, and pRam1 all had domains with no homology to *R. andeanae* reads, it is likely that pRan1 has a unique sequence. It is probably at least 15–16 Kb in size (98.2% coverage of the 15 Kb plasmid pRrh1, whereas the 18 Kb plasmid pRam1 had 87.0% coverage) and belongs to this widespread small rickettsial plasmid family.

The sequence reads mapping to the FLE agent present in both the Amac21 GA and Amac50 OK ticks confirmed that both of these agents were similar to the AmacFLE agent assembled from Illumina HiSeq (7 contigs) by Gerhart et al. (2016). Although the number of *Rickettsia* reads mapping to chromosomes of *R. parkeri* and *R. andeanae*, respectively, were similar in the two ticks, 4.7–6.4 times as many FLE reads mapped from Amac21 data than from Amac50 (Table 1.4). Whether the presence of the plasmid pRan1 reduced FLE detection in Amac50 is not known but that

Table 1.8 Comparison of the chromosome assemblies obtained with Geneious (G) and CLC genomics workbench (CLC) from IT-PGM sequence libraries obtained from two *Amblyomma maculatum* female ticks with those of reference isolates of spotted fever group *Rickettsiae* and the CLEAA agent from the Oklahoma *A. americanum* colony tick. *R. parkeri* identified in tick (Green). Candidatus R. andeanae, closest relative, identified in tick (orange)

Rickettsia domain	Size	Amac21GA (G)	Amac21GA (CLC)	Amac50-OK (G)	Amac50-OK (CLC)
R. parkeri Portsmouth	1,300,386 bp	802,742 bp	837,142 bp	807,327 bp	
R. rhipicephali 3-7-F6 CWPP	1,290,368 bp	743,661 bp		812,157 bp	846,610 bp
R. montanensis OSU 85-930	1,279,798 bp	738,210 bp		811,465 bp	845,102 bp
R. massiliae AZT80	1,263,719 bp	732,041 bp		792,829 bp	827,566 bp
R. amblyommatis GAT-3OV	1,407,796 bp	763,164 bp		850,057 bp	888,178 bp
		Amac21 (G)	Amac21 (CLC)	Amac50 (G)	Amac50 (CLC)
FLE-AmacOK domain	1,556,261 bp	1,236,324 bp	1,259,409 bp	192,960 bp	269,992 bp
Total IT-PGM reads		**4,007,976**		**3,810,034**	

Table 1.9 Geneious 10.0.9 bioinformatic analysis of IT-PGM sequence libraries for the presence of *Rickettsia* agent chromosome and plasmid sequences

Ticks	# IT-PGM Reads	*R. parkeri* Portsmouth	*R. rhipicephali* 3-7-F6 CWPP	pRrh1	*R. massiliae* AZT80	pRma2
Size of ref		13,00,386	12,90,368	15,099	12,73,719	15,000
Amac21 (GA)	40,07,976	26,442 (2.2x, 61.7%)	25,639 (1.8x, 57.6%)	5 (0.0x, 3.6%)	24,871 (1.7X, 57.9%)	6 (0.0X, 4.2%)
Amac50 (OK)	38,10,034	13,771 (1.4x, 62.1%)	13,843 (1.4x, 62.9%)	15,979 (11.9x, 98.2%)	13,676 (1.5x, 62.7%)	1,200 (11.2x, 91.4%)
DandM9	46,90,550	38 (0.0x, 0.1%)	42 (0.0x, 0.1%)	0 (0.0x, 0%)	46 (0.0x, 0.1%)	0 (0.0x, 0%)
DoccEMCF9A(CA)	65,18,864	28,433 (2.9x, 79.1%)	28,826 (2.8X, 83.2%)	18,203 (110.8x, 98.8%)	28408 (2.7x, 81.3%)	2,600 (44.4x, 96.5%)
Aam10 (GA)	34,90,856	20,689 (2.5x, 17.4%)	21,957 (2.8x, 17.2%)	136,383 (1371.2 (99.3%)	15,845 (1.3x, 17.1%)	123,975(1348.8, 98.3%)
Aam46 (GA)	52,42,310	23,948 (7.1x, 66.6%)	32,257 (10.5x, 66.0%)	243,649 (3783.6, 99.3%)	13,710 (2.9x, 66.0%)	245,441 (3774.0x, 98.4%)
RhsangF23 UF6(OSU)	44,90,171	82,212 (8.6x, 92.7%)	100,535 (9.3x, 93.2%)	3 (0.0x, 3.0%)	106,231 (10.4x, 94.9%)	4 (0.0x, 2.5%)
Dvar22 (GA)	62,11,538	8 (0.0x, 0.1%)	7 (0.0x, 0.1%)	0 (0.0x, 0%)	7 (0.0x, 0.1%)	0 (0.0x, 0%)
Dvar100 (OH)	37,17,748	63,770 (9.4x, 71.7%)	64,035 (9.6x, 72.8%)	654 (10.7x, 10.5%)	63,058 (10.0x, 73.1%)	647 (10.6x, 10.5%)
DvarSlov(VA)	58,10,262	9 (0.0x, 0.1%)	0 (0.0x, 0%)	0 (0.0x, 0%)	8 (0.0x, 0.1%)	0 (0.0x, 0%)
DvarF4 M4(OK)	28,91,366	1,865,192 (144.5, 94.3%)	1,795,425 (144.9x, 95.2%)	222 (1.4x, 62.0%)	1,851,895 (146.8x, 96.7%)	139 (1.0x, 58.7%)
DvarF1 SC1 (OK)	19,602	16,997 (1.8x, 74.3%)	16,891 (1.8x, 75.0%)	220 (1.7x, 17.9%)	16,898 (1.8x, 76.2%)	77 (1.3x, 23.0%)
DvarF4 SC4 (OK)	72,986	60,813 (9.0x, 94.6%)	60,229 (9.0x, 95.8%)	34 (1.5x, 75.2%)	60,530 (9.2x, 97.3%)	95 (4.7x, 70.2%)

Green: Positive identification of *Rickettsia* with wide genomic coverage (large contigs or nearly complete)
Orange: Positive presence of *Rickettsia* sequences but with low and interspersed coverage (many contigs)
Pink: Positive maps of reads to *Rickettsia* but with very low coverage and big gaps (few contigs)
White: Negative for *Rickettsia* , reads map to sites of low complexity
Boxes: number of reads mapping (median read coverage per base, percentage of reference genome mapped)

Table 1.10 Geneious 10.0.9 bioinformatic analysis of IT-PGM sequence libraries for the presence of *Rickettsia* agent chromosome and plasmid sequences

Ticks	# IT-PGM Reads	*R. montanensis OSU85-930*	*R. bellii* OSU85-389	pRbe1	*R. rickettsii* SSmith
Size of ref		12,79,798	15,28,980	48,775	12,57,710
Amac21 (GA)	40,07,976	25,275 (2.0x, 57.7%)	18,921 (1.1X, 19.9%)	0 (0.0x, 0%)	24,523 (1.6x, 60.1%)
Amac50 (OK)	38,10,034	13,765 (1.3x, 63.4%)	11,148 (1.4x, 23.9%)	425 (1.2x, 17.2%)	13,586 (1.4x, 62.9%)
DandM9	46,90,550	45 (0.0x, 0.1%)	327 (0.1x, 0.2%)	1 (0.0x, 0.1%)	41 (0.0x, 0.1%)
DoccEMCF9A(CA)	65,18,864	28,270 (2.9x, 78.7%)	23,708 (1.6x, 36.2%)	288 (1.3x, 7.4%)	28,279 (2.8x, 80.2%)
Aam10 (GA)	34,90,856	20,632 (2.5x, 17.4%)	30,515 (4.4x, 5.1%)	53,180 (311.4, 31.2%)	20,326 (2.5x, 17.2%)
Aam46 (GA)	52,42,310	23,498 (6.8x, 66.3%)	55,922 (18.0x, 28.4%)	126,218 (1128.9x, 33.4%)	23,257 (7.1x, 66.2%)
RhsangF23 UF6(OSU)	44,90,171	109,892 (11.5x, 97.1%)	82,974 (5.8x, 55.2%)	1 (0.0x, 0.1%)	101,051 (9.3x, 94.7%)
Dvar22 (GA)	62,11,538	8 (0.0x, 0.1%)	13,234 (1.3x, 0.1%)	0 (0.0x, 0%)	9 (0.0x, 0.1%)
Dvar100 (OH)	37,17,748	65,136 (9.7x, 74.7%)	98,731 (12.9x, 99.1%)	18,926 (82.4x, 100.0%)	61,925 (9.5x, 92.4%)
DvarSlov(VA)	58,10,262	11 (0.0x, 0.1%)	13 (0.0x, 0.1%)	1 (0.0x, 0.0%)	9 (0.0x, 0.1%)
DvarF4 M4(OK)	28,91,366	1,925,309 (151.1x, 97.8%)	1,491,092 (106.2x, 69.3%)	2,345 (6.4x, 5.5%)	1,831,678 (145.6x, 96.5%)
DvarF1 SC1 (OK)	19,602	17,561 (1.9x, 78.4%)	12,571 (1.2x, 51.5%)	234 (1.7x, 34.7%)	16,742 (1.8x, 76.3%)
DvarF4 SC4 (OK)	72,986	62,416 (9.4x, 97.7%)	51,493 (7.2x, 88.6%)	658 (4.5x, 34.2%)	59,920 (9.2x, 96.5%)

Green: Positive identification of *Rickettsia* with wide genomic coverage (large contigs or nearly complete)

Orange: Positive presence of *Rickettsia* sequences but with low and interspersed coverage (many contigs)

Pink: Positive maps of reads to *Rickettsia* but with very low coverage and big gaps (few contigs)

White: Negative for *Rickettsia* , reads map to sites of low complexity

Boxes: number of reads mapping (median read coverage per base, percentage of reference genome mapped)

might have occurred since the *R. rhipicephali* chromosome and plasmid had similar read coverage in the Amac50 IT-PGM library despite their 85-fold difference in size.

1.3.7 Analysis of Bacterial Sequences Present in the Ticks **Dermacentor andersoni** *and* **D. occidentalis**

Dermacentor andersoni is a vector for the agent of Rocky Mountain spotted fever, *R. rickettsii*, and commonly harbors the so-called East Side agent, *R. peacockii*, which is thought to block transovarial maintenance of *R. rickettsii* (Niebylski et al. 1997). Our adult male sample from Montana from the Oklahoma State University tick colony (DandM9) had no detectable *Rickettsia* DNA by qPCR and thus served as an excellent negative control for the read mapping procedure as only minimal background reads mapped to a few sites in both *Rickettsia* chromosomal and plasmid targets (Tables 1.7, 1.9 and 1.10) and to CLE agents (Table 1.5). The library and sequencing worked because its FLE agent (Scoles 2004) was readily detected (Table 1.4). This library was the only one sequenced with 400 bp chemistry, but that difference did not result in a significantly longer median read length (220 bp). CLC mapping was better than Geneious for assembling its FLE agent (Table 1.4).

Dermacentor occidentalis is the vector of *Rickettsia philipii*, the etiologic agent of Pacific Coast Tick fever on the West Coast of the USA (Padgett et al. 2016). It also more commonly harbors *Rickettsia rhipicephali* (Wikswo et al. 2008; Stephenson et al. 2017). This latter agent is not identical to the *R. rhipicephali* from *Rhipicephalus sanguineus* (Wikswo et al. 2008), whose genome sequence was obtained from the type strain 3-7-F6 CWPP, which was isolated from a *Rhipicephalus sanguineus* tick from Mississippi in 1973 (Burgdorfer et al. 1975). Our female adult *D. occidentalis* sample (EMCF9) was collected on February 14, 2014 at Crystal Cove State Park, El Moro Canyon, California. *Rickettsia rhipicephali* was present, as shown by qPCR and DNA sequencing. 83% of the chromosome and 98.8% of the pRh1 plasmid sequence could be mapped and assembled with Geneious (Table 1.9); CLC analysis provided similar results (not shown).

1.3.8 Analysis of **Dermacentor variabilis** *from Virginia, Georgia, and Ohio for* **Rickettsia** *and FLE*

Dermacentor variabilis is commonly infected with *R. montanensis* and *R. bellii* (Pagac et al. 2014; Stephenson et al. 2017; Wood et al. 2016; Yunik et al. 2015) and is the major vector of sporadic cases of Rocky Mountain spotted fever in two thirds of the USA (Anderson et al. 1986). It can also be infected with *R. rhipicephali* in populations on the West Coast of the USA (Wikswo et al. 2008). Much like the *D. andersoni* tick DandM9 from Montana, two samples of *D. variabilis* adults from

Georgia (Dvar22, female) and Virginia (DvarSlov) gave insignificant numbers of sequence reads mapping to *Rickettsia* chromosomes or plasmids (Tables 1.7, 1.9 and 1.10), further confirming the accuracy of the mapping methodology. However, another qPCR negative *D. variabilis* female tick, obtained from The Wildlands Park in Ohio in August 2014 (Dvar100), contained many sequence reads, mapping to both the chromosome and plasmid of *R. bellii* OSU85-389 (Fig. 1.6); this reference isolate had originally been obtained from an adult female collected in Franklin County, Ohio 5/13/1989 (Fuerst et al. 1990), a location about 50 miles west of Wildlands Park. *R. bellii* is not detected with the rickettsia qPCR assays employed (Eremeeva et al. 2003). Remarkably, this quality sequence assembly had 99.1% coverage of the *R. bellii* chromosome and 100.0% of the 48,775 bp plasmid (Fig. 1.6). None of the other plasmids in US ticks had more than patchy similarity to this sequence (only the large *Rickettsia felis* plasmid pRF1 is shown-Fig. 1.6b).

All three eastern *D. variabilis* ticks had sequences reads that each mapped to the three FLE agents at comparable levels, but at much lower levels than to *R. bellii* OSU85-389 in Dvar100 (Tables 1.2 and 1.10).

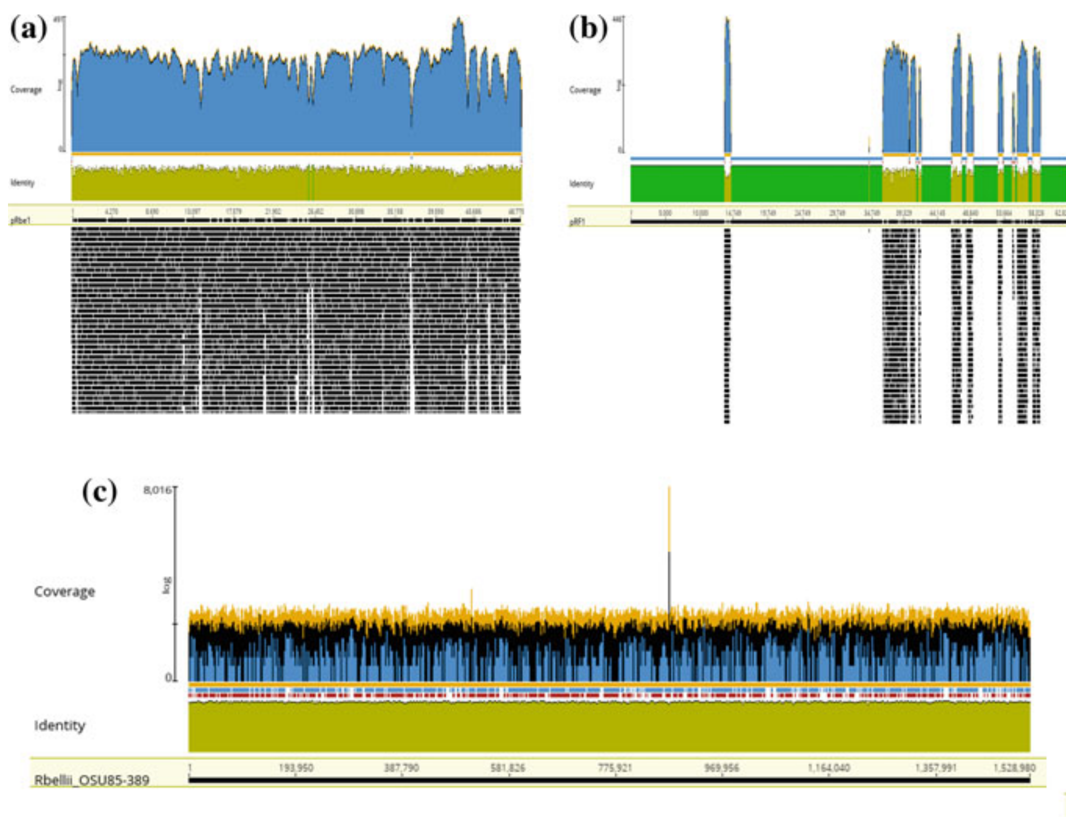

Fig. 1.6 Geneious read mapping of Dvar100 IT-PGM sequences against selected *Rickettsia* sequences. **a** *R. bellii* OSU85-389 plasmid pRbe1; **b** *R. felis* URRWXCal2 plasmid pRF1; **c** *R. bellii* OSU 85-389 chromosome. See Fig. 1.1 for other details

1.3.9 *Detection of* Rickettsia montanensis *in* Dermacentor variabilis *and* Rhipicephalus sanguineus *from Oklahoma*

Two *D. variabilis* females (DvarF4 and DvarF1) from the Oklahoma State University Tick colony were analyzed following Repli-G amplification of their DNA. The mini-Repli-G amplified library (DvarF4M4) was sequenced on the standard Ion 318 chip while the other two libraries from single cell (SC) amplification (DvarF4SC4, DvarF1SC1) were analyzed on the small Ion 314 chips to mimic the results to be expected from individual barcoded library samples. Because of problems occurring during loading the small chip and fractionation of the two SC libraries, this experiment was imperfect but it provides a sense of the lower boundaries of results that might be expected by the library input reductions required by mixing barcoded samples at equivalent loadings. A *Rickettsia* agent, mapping most strongly to the reference strain *R. montanensis* OSU-85-930, was obtained with all three libraries (Fig. 1.7c; Table 1.10) at 97.8, 97.7 and 78.4% coverage for DvarF4M4 (1,914,722 reads mapping), DvarF4SC4, and DvarF1 SC1 libraries, respectively. However, unexpectedly, since the reference isolate of *R. montanensis* has no plasmid, these samples also contained sequences that mapped against rickettsial plasmids, particularly pRrh1 (Fig. 1.7a) and pRma2 (not shown) and their abundance decreased with the total number of sequences obtained (Table 1.9). In comparison, while a *Rhipicephalus sanguineus* female (RhsangF23) from the same Oklahoma State University tick colony had about twice the coverage (109,892 reads) as DvarF4SC4 of the *R. montanensis* chromosome (62,416 reads, 97.7% coverage) (Fig. 1.7d; Table 1.10); it had only 3 reads to pRrh1 (Fig. 1.7b) while DvarF4SC4 had 34 reads mapping. The low plasmid coverage in RhsangF23 suggests that if this plasmid is indeed from *R. montanensis*, it has a very low copy number. Possibly of more significance is the fact the *R.*

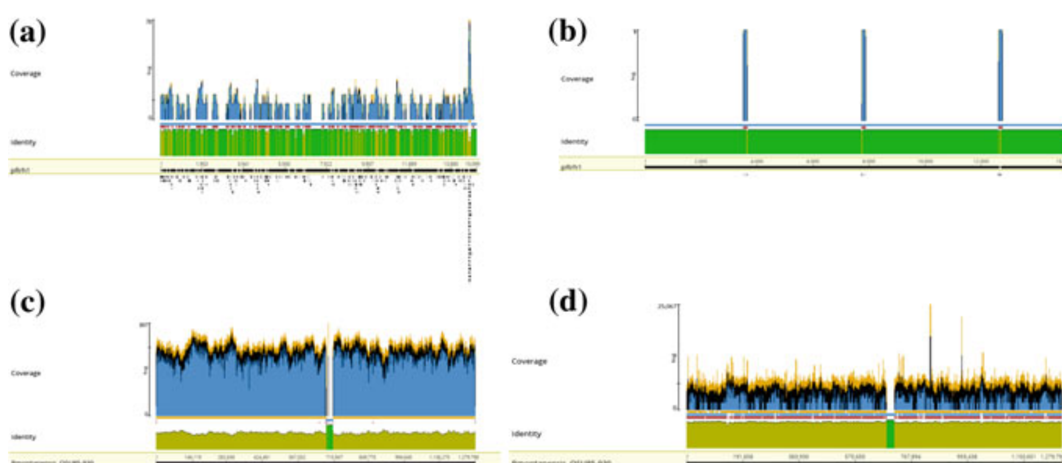

Fig. 1.7 Geneious read mapping of DvarF4M4 (**a**, **c**) and RhsangF23 (**b**, **d**) IT-PGM sequences against selected *Rickettsia* sequences. **a**, **b** *R. rhipicephali* 3-7-F6 CWPP plasmid pRrh1; **c**, **d** *R. montanensis* OSU 85-930 chromosome. See Fig. 1.1 for other details

montanensis-like agents in both DvarF4M4 and RhsangF23 contained no sequences mapping to the same site of 28.7 Kb located close to coordinate 700 Kb in *R. montanensis* OSU85-930 (Fig. 1.7c, d). This variant type of *R. montanensis* may in fact have been exchanged by horizontal transmission at some point, possibly originating from *D. variabilis*, since wild *Rhipicephalus sanguineus* have been reported only rarely to contain *R. montanensis* (Trout-Fryxell et al. 2015), while it is commonplace in *D. variabilis* (Fuerst et al. 1990; Stromdahl et al. 2011). Because *omp*A detection was inefficient relative to *omp*B assays in the Trout-Fryxell paper, surveillance with the former target may account for its low detection rate; however, the *omp*A region used in the Eremeeva et al. (2003) EvaGreen assay was present in the *R. montanensis* assembled from both the DvarF4M4 and RhsangF23 sequences, so other factors may account for its low detection rate rather than its absence in some agents.

1.4 Discussion and Conclusions

1.4.1 Library Preparation and Sequencing

The most variable aspects of sequencing on the IT-PGM were obtaining consistent mean read lengths and recovery of sufficient amounts of the library for loading. In part, this variability was due to the fact that tick samples with different histories and yields of DNA were used, as well as different approaches to fractionation of the DNA. Manual loading of the chips improved greatly with experience and this step can now be automated with the Ion Chef system. The use of the Repli-G single cell procedure did not appear to affect adversely read mapping or genome assembly. Indeed, it had significant advantages in stretching limited single sample DNAs and what appears to be preferential amplification of the microbial and mitochondrial (not shown) sequences over the tick chromosomal reads. This result may be due to preferential fragmentation of the tick DNA relative to the microbial and mitochondrial DNA, so that amplification of the preferred latter longer more intact DNA occurs. This effect was most spectacular with the 20 mitochondrial minicircles from human lice (not shown) and varied more among the tick samples. Because the *Rickettsia* plasmids detected here are relatively large, no consistent or marked effect was seen relative to the chromosomal reads. No plasmid was detected in the CLE or FLE agents.

1.4.2 Genome Assembly and Annotation

Geneious proved to be a very efficient tool for mapping the amount of IT-PGM reads on a desk top computer (usually less than 10 min). The de novo assembly did not provide significantly better results than the contigs obtained by read mapping.

This similarity is probably due to the low read depth, but it could be improved by removing short reads and aberrant read pileup spikes. The key to successful read mapping assembly is having reference sequences that are very closest to that of the agent being sequenced. The biggest advantage in using CLC genomics workbench on a high performance cluster (and with the other command line tools available) was being able to map against a database containing all of the related genome sequences available, a task that Geneious cannot perform on a laptop. However, the quality of the final CLC assemblies was again only marginally better than that obtained with Geneious. This program is also very suitable for rapid screening of the quality and usefulness of IT-PGM data by non-bioinformaticians. It can also be used for annotations using GenBank reference sequences. In general, we were pleasantly surprised at the amount of useful data that was obtained on the platform.

1.4.3 Limitations of Approach and Possible Remedies

The limitations of read depth for assembling symbiont genomes from metagenomics samples could be overcome only partially with Repli-G amplification. The larger Ion S chip will help but it also increases cost. It is likely that successful implementation of the 400 bp chemistry and improvements in accurate library size fractionation would also help but these aspects were not tested enough to confirm any advantages. One of our goals was to obtain enough symbiont genome coverage that long range PCR could be employed to make baits for enrichment of the target symbiont DNAs (Dunning Hotopp et al. 2017; Kent et al. 2011; Jones and Good 2016) and to design primers suitable for multiple locus sequence typing (MLST) of the symbionts, for population studies of mitochondria evolution in the ticks (Ketchum et al. 2009). We confirmed that primers for MLST target genes could be recovered for FLE, CLE and *Rickettsia* (not shown). The bait enrichment strategy should also work since reads were mapped over much of each of these chromosomes for multiple agents and for the *Rickettsia* plasmids. The tick mitochondrial assemblies (14.5–15.0 Kb) also provided enough assembled data to permit PCR amplification of target regions using specific primers rather than relying solely on conserved or degenerate tick primers and long-range PCR (not shown).

1.4.4 Value of Partial Symbiont Genome Sequences from the IT-PGM

Besides facilitating direct approaches to obtaining partial or complete genome sequences for symbionts and non-cultivable *Rickettsia*, we have demonstrated that substantial unknown information could be extracted from the DNA of single alcohol-preserved ticks. In many cases involving outbreaks of disease or tick samples that

must be shipped from other countries, alcohol-preserved ticks may be the only available material.

It was possible to determine if each agent had been previously sequenced and how similar it was to available sequences. This comparison was demonstrated unequivocally with the *Rickettsia* sequences, wherein *R. parkeri*, *R. amblyommatis*, and *R. bellii* could be compared with genome sequences from isolates; a significant variant of *R. montanensis* was identified in both *D. variabilis* and *R. sanguineus*; a plasmid was confirmed to be present in *R. andeanae* and *R. bellii*; and a large amount of new sequence was obtained for *R. andeanae.* Sufficient DNA sequence was obtained to determine the %GC content and minimal genome sizes for the *R. sanguineus* CLE and *D. variabilis* FLE agents and the genome sizes of the *A. maculatum* FLE and *A. americanum* CLE could be confirmed. For some tick bacterial agents, the read depth was sufficient to identify SNPs and INDELs in the assembled genomes relative to the available reference sequences. Finally, phylogenetic trees could also be constructed with concatenated sequences to further confirm agent identity. These features will permit routine evaluation and selection of specific tick DNA samples, which can warrant the cost of deeper metagenomic sequencing both from the standpoint of DNA quality and relative agent abundance and the uniqueness of the target agent. However, one confounding factor in metagenome-derived assemblies may be the presence of symbiont sequences in the host genome that have been inserted following endosymbiont-host lateral gene transfers (Dunning Hotopp et al. 2017).

Acknowledgements The authors thank the following individuals for providing some of the individual ticks used in this study from other previous collaborative investigations: Sarah Billeter, Renjie Hu, Alece Powery, Jasmine Hensley, Lindsey Killmaster, and Amanda Jo Williams-Newkirk. We also thank Lisa Coburn, manager of the Oklahoma State University Tick Laboratory, for providing details on the origin of ticks from their tick colonies which are supported in part by NIAID BEI Resources. This research was supported in part by appointment of Arunachalam Ramaiah in the Postdoctoral Bioinformatics Fellowship Program administered by the Association of Public Health Laboratories (APHL) and funded by the Centers for Disease Control and Prevention (CDC) as part of Cooperative Agreement 1U60OE000103 (CFDA No. 93.322). The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the U.S. Centers for Disease Control and Prevention.

Data Submission The Ion Torrent PGM sequence data for each sample (Table 1.1) is available at the NCBI SRA resources site under Bioproject numbers PRJNA408163 (*Rhipicephalus sanguineus* F23UF), PRJNA413819 (*Dermacentor andersoni* M9), PRJNA413914 (*Dermacentor occidentalis* EMCF9), PRJNA413920 (*Amblyomma americanum* Aam10 and Aam46), PRJNA413923 (*Amblyomma maculatum* Amac21, Amac50), and PRJNA413929 (*Dermacentor variabilis* Dvar22, Dvar200, DvarSlov, DvarF4M4, DvarFvSC4, and DvarF1SC1).

Conflict of Interest Statement The authors declare they have no conflicts of interest affecting this work.

References

Anderson JF, Magnarelli LA, Philip RN, Burgdorfer W (1986) *Rickettsia rickettsii* and *Rickettsia montana* from ixodid ticks in Connecticut. Am J Trop Med Hyg 35:187–191

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477

Bermúdez SE, Eremeeva ME, Karpathy SE, Samudio F, Zambrano ML, Zaldivar Y, Motta JA, Dasch GA (2009) Detection and identification of rickettsial agents in ticks from domestic mammals in eastern Panama. J Med Entomol 46:856–861

Buchner P (1965) Endosymbiosis of animals with plant microorganisms. Interscience Publishers, New York

Burgdorfer W, Sexton DJ, Gerloff RK, Anacker RL, Philip RN, Thomas LA (1975) *Rhipicephalus sanguineus*: vector of a new spotted fever group rickettsia in the United States. Infect Immun 12:205–210

Corbin C, Heyworth ER, Ferrari J, Hurst GD (2017) Heritable symbionts in a world of varying temperature. Heredity (Edinb) 118:10–20

Dasch GA, Shirey TB, Zambrano ML, Sheth M, Holmes ZC, Williams-Newkirk AJ (2016) Ion torrent PGM and Illumina HiSeq analysis of the metagenome of *Dermacentor variabilis* (Ixodidae). In: XXV international congress of entomology, Orlando, FL, September 25–30, 2016

Dergousoff SJ, Chilton NB (2012) Association of different genetic types of *Francisella*-like organisms with the Rocky Mountain wood tick (*Dermacentor andersoni*) and the American dog tick (*Dermacentor variabilis*) in localities near their northern distributional limits. Appl Environ Microbiol 78:965–971

Douglas AE, Werren JH (2016) Holes in the hologenome: why host-microbe symbioses are not holobionts. MBio 7(2):e02099

Dunning Hotopp JC, Slatko BE, Foster JM (2017) Targeted enrichment and sequencing of recent endosymbiont-host lateral gene transfers. Sci Rep 7:857

Duron O, Binetruy F, Noël V, Cremaschi J, McCoy KD, Arnathau C, Plantard O, Goolsby J, Pérez de León AA, Heylen DJA, Van Oosten AR, Gottlieb Y, Baneth G, Guglielmone AA, Estrada-Peña A, Opara MN, Zenner L, Vavre F, Chevillon C (2017) Evolutionary changes in symbiont community structure in ticks. Mol Ecol 26:2905–2921

Epis S, Sassera D, Beninati T, Lo N, Beati L, Piesman J, Rinaldi L, McCoy KD, Torina A, Sacchi L, Clementi E, Genchi M, Magnino S, Bandi C (2008) *Midichloria mitochondrii* is widespread in hard ticks (Ixodidae) and resides in the mitochondria of phylogenetically diverse species. Parasitology 135:485–494

Eremeeva ME, Dasch GA, Silverman DJ (2003) Evaluation of a PCR assay for quantitation of *Rickettsia rickettsii* and closely related spotted fever group rickettsiae. J Clin Microbiol 41:5466–5472

Fornadel CM, Zhang X, Smith JD, Paddock CD, Arias JR, Norris DE (2011) High rates of *Rickettsia parkeri* infection in Gulf Coast ticks (*Amblyomma maculatum*) and identification of "*Candidatus* Rickettsia andeanae" from Fairfax County, Virginia. Vector Borne Zoonotic Dis 11:1535–1539

Fuerst PA, Poetter KP, Pretzman C, Perlman PS (1990) Molecular genetics of populations of intracellular bacteria: the spotted fever group rickettsiae. Ann N Y Acad Sci 590:430–438

Gerhart JG, Moses AS, Raghavan R (2016) A *Francisella*-like endosymbiont in the Gulf Coast tick evolved from a mammalian pathogen. Sci Rep 6:33670

Gillespie JJ, Joardar V, Williams KP, Driscoll T, Hostetler JB, Nordberg E, Shukla M, Walenz B, Hill CA, Nene VM, Azad AF, Sobral BW, Caler E (2011) A *Rickettsia* genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of an obligate intracellular lifestyle. J Bacteriol 194:376–394

Gottlieb Y, Lalzar I, Klasson L (2015) Distinctive genome reduction rates revealed by genomic analyses of two *Coxiella*-like endosymbionts in ticks. Genome Biol Evol 7:1779–1796

Henry LM, Maiden MC, Ferrari J, Godfray HC (2015) Insect life history and the evolution of bacterial mutualism. Ecol Lett 18:516–525

Jasinskas A, Zhong J, Barbour AG (2007) Highly prevalent *Coxiella* sp bacterium in the tick vector *Amblyomma americanum*. Appl Environ Microbiol 73:334–336

Jiang J, Stromdahl EY, Richards AL (2012) Detection of *Rickettsia parkeri* and *Candidatus* Rickettsia andeanae in *Amblyomma maculatum* Gulf Coast ticks collected from humans in the United States. Vector Borne Zoonotic Dis 12:175–182

Jones MR, Good JM (2016) Targeted capture in evolutionary and ecological genomics. Mol Ecol 25:185–202

Karpathy SE, Slater KS, Goldsmith CS, Nicholson WL, Paddock CD (2016) *Rickettsia amblyommatis* sp nov, a spotted fever group *Rickettsia* associated with multiple species of *Amblyomma* ticks in North, Central and South America. Int J Syst Evol Microbiol 66:5236–5243

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649

Kent BN, Salichos L, Gibbons JG, Rokas A, Newton IL, Clark ME, Bordenstein SR (2011) Complete bacteriophage transfer in a bacterial endosymbiont (*Wolbachia*) determined by targeted genome capture. Genome Biol Evol 3:209–218

Ketchum HR, Teel PD, Coates CJ, Strey OF, Longnecker MT (2009) Genetic variation in 12S and 16S mitochondrial rDNA genes of four geographically isolated populations of Gulf Coast ticks (Acari: Ixodidae) J Med Entomol 46:482–489

Killmaster LF, Loftis AD, Zemtsova GE, Levin ML (2014) Detection of bacterial agents in *Amblyomma americanum* (Acari: Ixodidae) from Georgia, USA, and the use of a multiplex assay to differentiate *Ehrlichia chaffeensis* and *Ehrlichia ewingii*. J Med Entomol 51:868–872

Kurtti TJ, Felsheim RF, Burkhardt NY, Oliver JD, Heu CC, Munderloh UG (2015) *Rickettsia buchneri* sp nov, a rickettsial endosymbiont of the blacklegged tick *Ixodes scapularis*. Int J Syst Evol Microbiol 65:965–970

Larson MA, Nalbantoglu U, Sayood K, Zentz EB, Cer RZ, Iwen PC, Francesconi SC, Bishop-Lilly KA, Mokashi VP, Sjöstedt A, Hinrichs SH (2016) Reclassification of *Wolbachia persica* as *Francisella persica* comb nov and emended description of the family Francisellaceae. Int J Syst Evol Microbiol 66:1200–1205

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAM tools. Bioinformatics 25:2078–2079

Luce-Fedrow A, Wright C, Gaff HD, Sonenshine DE, Hynes WL, Richards AL (2012) In vitro propagation of "*Candidatus* Rickettsia andeanae" isolated from *Amblyomma maculatum*. FEMS Immunol Med Microbiol 64:74–81

Mixson TR, Campbell SR, Gill JS, Ginsberg HS, Reichard MV, Schulze TL, Dasch GA (2006) Prevalence of *Ehrlichia*, *Borrelia*, and rickettsial agents in *Amblyomma americanum* (Acari: Ixodidae) collected from nine states. J Med Entomol 43:1261–1268

Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet 42:165–190

Niebylski ML, Schrumpf ME, Burgdorfer W, Fischer ER, Gage KL, Schwan TG (1997) *Rickettsia peacockii* sp nov, a new species infecting wood ticks, *Dermacentor andersoni*, in western Montana. Int J Syst Bacteriol 47:446–452

O'Fallon B (2008) Population structure, levels of selection, and the evolution of intracellular symbionts. Evolution 62:361–373

Oliva Chávez AS, Shaw DK, Munderloh UG, Pedra JH (2017) Tick humoral responses: marching to the beat of a different drummer. Front Microbiol 14:223

Paddock CD, Denison AM, Dryden MW, Noden BH, Lash RR, Abdelghani SS, Evans AE, Kelly AR, Hecht JA, Karpathy SE, Ganta RR, Little SE (2015) High prevalence of "*Candidatus* Rickettsia andeanae" and apparent exclusion of *Rickettsia parkeri* in adult *Amblyomma maculatum* (Acari: Ixodidae) from Kansas and Oklahoma. Ticks Tick Borne Dis 6:297–302

Padgett KA, Bonilla D, Eremeeva ME, Glaser C, Lane RS, Porse CC, Castro MB, Messenger S, Espinosa A, Hacker J, Kjemtrup A, Ryan B, Scott JJ, Hu R, Yoshimizu MH, Dasch GA, Kramer V (2016) The Eco-epidemiology of Pacific Coast tick fever in California. PLoS Negl Trop Dis 5:e0005020

Pagac BB, Miller MK, Mazzei MC, Nielsen DH, Jiang J, Richards AL (2014) *Rickettsia parkeri* and *Rickettsia montanensis*, Kentucky and Tennessee, USA. Emerg Infect Dis 20:1750–1752

Ramaiah A, Williams-Newkirk AJ, Frace MA, Zambrano ML, Dasch GA (2017) Evolution of the genomes of the *Coxiella*-like endosymbionts of ticks. J Mississippi Acad Sci 62:21

Scoles GA (2004) Phylogenetic analysis of the *Francisella*-like endosymbionts of *Dermacentor* ticks. J Med Entomol 41:277–286

Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069

Serbus LR, Rodriguez BG, Sharmin Z, Momtaz AJMZ, Christensen S (2017) Predictive genomic analyses inform the basis for vitamin metabolism and provisioning in bacteria-arthropod endosymbiosis. G3 (Bethesda) 7:1887–1898

Smith TA, Driscoll T, Gillespie JJ, Raghavan R (2015) A *Coxiella*-like endosymbiont is a potential vitamin source for the Lone Star tick. Genome Biol Evol 7:831–838

Stephenson N, Blaney A, Clifford D, Gabriel M, Wengert G, Foley P, Brown RN, Higley M, Buckenberger-Mantovani S, Foley J (2017) Diversity of rickettsiae in a rural community in northern California. Ticks Tick Borne Dis 8:526–531

Stromdahl EY, Jiang J, Vince M, Richards AL (2011) Infrequency of *Rickettsia rickettsii* in *Dermacentor variabilis* removed from humans, with comments on the role of other human-biting ticks associated with spotted fever group rickettsiae in the United States. Vector Borne Zoonotic Dis 11:969–977

Trout Fryxell RT, Steelman CD, Szalanski AL, Billingsley PM, Williamson PC (2015) Molecular detection of *Rickettsia* species within ticks (Acari: Ixodidae) collected from Arkansas United States. J Med Entomol 52:500–508

Wernegreen JJ (2012) Strategies of genomic integration within insect-bacterial mutualisms. Biol Bull 223:112–122

Wernegreen JJ (2015) Endosymbiont evolution: predictions from theory and surprises from genomes. Ann NY Acad Sci 1360:16–35

Whitman TJ, Richards AL, Paddock CD, Tamminga CL, Sniezek PJ, Jiang J, Byers DK, Sanders JW (2007) *Rickettsia parkeri* infection after tick bite, Virginia. Emerg Infect Dis 13:334–336

Wikswo ME, Hu R, Dasch GA, Krueger L, Arugay A, Jones K, Hess B, Bennett S, Kramer V, Eremeeva ME (2008) Detection and identification of spotted fever group rickettsiae in *Dermacentor* species from southern California. J Med Entomol 45:509–516

Williams-Newkirk AJ, Burroughs M, Changayil SS, Dasch GA (2015) The mitochondrial genome of the lone star tick (*Amblyomma americanum*). Ticks Tick Borne Dis 6:793–801

Williams-Newkirk AJ, Rowe LA, Mixson-Hayden TR, Dasch GA (2012) Presence, genetic variability, and potential significance of "Candidatus Midichloria mitochondrii" in the lone star tick *Amblyomma americanum*. Exp Appl Acarol 58:291–300

Williams-Newkirk AJ, Rowe LA, Mixson-Hayden TR, Dasch GA (2014) Characterization of the bacterial communities of life stages of free living lone star ticks (*Amblyomma americanum*). PLoS ONE 9:e102130

Wood H, Dillon L, Patel SN, Ralevski F (2016) Prevalence of *Rickettsia* species in *Dermacentor variabilis* ticks from Ontario, Canada. Ticks Tick Borne Dis 7:1044–1046

Yunik ME, Galloway TD, Lindsay LR (2015) Assessment of prevalence and distribution of spotted fever group rickettsiae in Manitoba, Canada, in the American dog tick, *Dermacentor variabilis* (Acari: Ixodidae). Vector Borne Zoonotic Dis 15:103–108

Zhong J (2012) *Coxiella*-like endosymbionts. Adv Exp Med Biol 984:365–379

Zhong J, Jasinskas A, Barbour AG (2007) Antibiotic treatment of the tick vector *Amblyomma americanum* reduced reproductive fitness. PLoS ONE 2:e405