ResearchOnline@JCU

JAMES COOK UNIVERSITY AUSTRALIA

This is the author-created version of the following work:

Elliman, J. Ruth, and Owens, L. (2020) ***Confirmation that candidatus Coxiella cheraxi from redclaw crayfish (Cherax quadricarinatus) is a close relative of Coxiella burnetii, the agent of Q-fever.*** **Letters in Applied Microbiology, 71 (3) pp. 320-326.**

Access to this file is available from:

https://researchonline.jcu.edu.au/65019/

© 2020 The Society for Applied Microbiology

Please refer to the original source for the final version of this work:

https://doi.org/10.1111/lam.13336

1 **Title**

2 Confirmation that candidatus *Coxiella cheraxi* from redclaw crayfish (*Cherax quadricarinatus*) is a

3 close relative of *Coxiella burnetii*, the agent of Q-fever.

4 **Running title:** *C. cheraxi* ex *C. quadricarinatus*

5 Coxiella cheraxi sequencing

6 **Authors**

7 Jennifer Elliman and Leigh Owens

8 **Corresponding author details**

9 Leigh Owens

10 James Cook University

11 leigh.owens@jcu.edu.au

12 ph 0747814632

13

14 **Work done**

15 James Cook University, Australia

16

17 Confirmation that candidatus *Coxiella cheraxi* from redclaw crayfish (*Cherax quadricarinatus*) is a

18 close relative of *Coxiella burnetii*, the agent of Q-fever.

19 Jennifer Elliman and Leigh Owens*

20 College of Public Health, Biomedicine and Veterinary Science

21 1 Solander Drive, James Cook University,

22 Townsville, Australia 4811.

23 *author for correspondence leigh.owens@jcu.edu.au

24 **Significance and Impact of Study**

25 This work reports on the use of newer technologies on archival samples and provides significantly

26 more data on the currently limited genome data of candidatus *Coxiella cheraxi,* one of the few

27 species isolated in the genus *Coxiella. Candidatus Coxiella cheraxi* causes death in redclaw crayfish

28 and has been reported as being closely related to *Coxiella burnettii*, the agent of Q-fever*,* based on

29 16S rRNA sequencing. This work provides confirmation for this claim.

30 **Abstract**

31 A *Coxiella* sp. closely related to the agent of Q-fever, *Coxiella burnetii*, has been associated with

32 mortalities in redclaw crayfish, (*Cherax quadricarinatus*), in farms and experimental facilities for

33 three decades. Limited sequence data including 16S rRNA have placed the rickettsial species as a

34 new species, candidatus *Coxiella cheraxi* closely related to *C. burnetii*. MinION sequencing was

35 conducted on the last remaining isolate from an outbreak of disease, TO-98. The accuracy of base

36 pair reads was mostly 99.9% (error rate 1 in 1000) or better. After filtering for reads of co-isolated

37 *Citrobacter freundii*, 2,629 sequences remained with the longest being 12,585 base pairs (bp). The

38 longest 21 sequences are presented with their single best hit statistics when examined by NCBI

39 Blastn (nucleotides) and the nucleotides translated into proteins NCBI Blastx. All sequences hit with

40 either *Coxiella burnetii* (29/42, 69%) or *Coxiella* (10/42, 24%) or rickettsia (3/42, 7%) with an error
41 rate of less than 1 in 1 million for either bp or amino acids. Sequencing in this report confirms
42 candidatus *C. cheraxi* is a new species very closely related to *C. burnetii.*

43 Keywords: *Coxiella cheraxi*, *Coxiella burnetii,* Q-fever, crayfish, *Cherax quadricarinatus, Citrobacter*
44 *freundii*

45 **Introduction**

46 Rickettsial problems in freshwater crayfish (*Cherax quadricarinatus*) were first recorded in northern
47 Queensland, Australia in 1989 in an experimental facility (Owens *et al.* 1992). By 1990, rickettsia had
48 caused 24,000 deaths in crayfish, a 22% loss at a farm in south eastern Queensland (Ketterer *et al.*
49 1992). Rickettsia continued to be found in farms (Edgerton *et al.* 1995; Powell 2013) and in
50 experimental facilities using farmed crayfish for trials (La Fauce and Owens 2007). Mortalities
51 associated with the experimental trials were often so severe that trials had to be abandoned, (200 of
52 500 crayfish died, 90% of a 2nd batch died) (Powell 2013) or were compromised (La Fauce and
53 Owens 2007). Furthermore, a similar rickettsia identified by electron microscopy (Romero *et al.*
54 2000), were exported to Ecuador with *C. quadricarinatus* where it caused 45-80% mortalities in
55 growout ponds (Jimenez and Romero 1997). The continued irritating recrudescence of rickettsial
56 disease has meant some effort into identification of the rickettsia has occurred, albeit on a
57 shoestring budget.

58 It is difficult to culture rickettsia in cell culture as being bacteria, rickettsia are killed by antibiotics
59 used to kill contaminating bacteria and keep cell cultures viable. The rickettsia sample has to be
60 completely free of contaminating bacteria, a state that is often very difficult to achieve without
61 extremely good axenic techniques. As alternatives, embryo culture in chicken embryos was used
62 previously (Tan and Owens 2000) or suckling mice brains can be trialled but they are ethically
63 unpalatable techniques and difficult to get through ethics committees. Therefore, molecular
64 techniques and sequencing are often a way of providing new information.

65 The first attempt to use molecular methods to identify the Australian crayfish rickettsia was in 1998.
66 PCRs for 16SrRNA were undertaken on an isolate designated TO-98 (Tan and Owens 2000) recovered
67 using embryo culture from crayfish dying on sorting trays when establishing monosex crayfish
68 populations. The rickettsial sequence of 1325 base pairs was 95.6% identical to *Coxiella burnetii*, the
69 agent of Q-fever. The isolate was proposed to be a new species, candidatus *Coxiella cheraxi* (UniProt
70 taxon identifier 426622). The near derailing of the crayfish study of La Fauce and Owens (2007) lead
71 to the next molecular examination of TO-98 (Cooper *et al.* 2007) which produced a real-time qPCR
72 using the *Coxiella* outer membrane 1 (*com*1) PCR which was identical to *C. burnetii*. Unfortunately,
73 this powerful diagnostic test was erroneously missed by the risk assessment team for the
74 importation of Australian redclaw crayfish into Norway (VKM 2016). Further sequence of
75 superoxidase dismutase (*sod*B) with flanking genes (99% identical) and a repeat sequencing of the
76 16SrRNA gene clearly indicated close relationship between the two *Coxiella* species (Cooper *et al.*
77 2007).

78 The epizootic of 2012 in crayfish trials (mentioned above) was traced back using histopathology,
79 16SrRNA and *com*1 PCR (Powell 2013) to one of three farms having production problems.
80 Unfortunately, the sequencing was poor, resulting in 85% and 83% matches for 16SrRNA and *com*1
81 respectively. A PCR for the virulence sequence of *Coxiella burnetii*, IS111a (Duron 2015) was
82 negative. Therefore, the aetiological agent of the 2012 outbreak was concluded to be a rickettsial
83 infection, probably *Coxiella* but not *C. burnetii,* perhaps *C. cheraxi* but conclusions could not be more
84 definitive (Powell 2013). An experiment involving trying to infecting crayfish, *C. quadricarinatus* with
85 10^4 to 10^6 *C. burnetii* lead to 67% infection at the level of 1-200 copies of *com*1 detected via qPCR at
86 three weeks suggesting the crayfish were not acutely susceptible to *C. burnetii* but were slow to
87 clear it from their bodies (Powell 2013).

88 Recently, next generation sequencing has become cheap enough to attempt sequencing with the
89 last, frozen sample of TO-98. This paper reports the use of MinION sequencing to produce extended

90 genome sequencing of candidatus *Coxiella cheraxi*. The relationship of the 21 longest sequences to
91 one other nearest neighbour are documented. In addition, *Citrobacter freundii* was found cohabiting
92 in the TO-98 sample and a partial sequence is included.

93 **Results and Discussion**

94 Gel electrophoresis indicated the DNA sample for sequencing was not sheared and was acceptable
95 for long read sequencing. However, during library preparation, the thermal cycler used in the
96 fragmentation step faulted, resulting in longer digestion of the DNA (about five minutes instead of
97 one). This should result in more, shorter DNA strands for sequencing. Given the lack of another
98 sample to reattempt this step, the library preparation and sequencing was continued to determine
99 how much sequence information could be gained. However, the digestion error in processing
100 confounds our ability to judge the effectiveness of using MinION technology on archival samples
101 under standard conditions. The read length analysis of the passed reads provides evidence of the
102 effect of over-digestion, with most reads being less than 2500 bp (Figure 1).

103 The quality of passed reads was assessed and the quality control Figure (2) has a shoulder between
104 Q12.5 and Q25. At a Q score of 20, this translates to 99% accuracy for the base pair reads (1 in 100
105 bp errors) with a substantial proportion at 99.9% (1 in one thousand bp error rate) or better.
106 Unfortunately, in addition to the large number of short reads, there is a large amount of sequencing
107 with low Q scores (below 12.5) where the error rate is at 10% (1 in 10 bp errors) so alignment was
108 not productive and therefore, the larger reads were directly analysed. As MinION sequencing is
109 more error prone than Illumina sequencing and depends on post-sequencing correction methods
110 (Rang *et al.* 2018), the sequences generated herein should be considered a draft rather than final.
111 Therefore, limited conclusions have been stated to minimise potential error in claims and
112 confirmatory sequencing should be ultimately sought when more samples become available.

113

114 Some of the longest sequences (~9,500 bp) were found to be *Citrobacter freundii* (99% identity,
115 e=0.0) (Table 1), so the sequence library was filtered against *C. freundii* sequences from NCBI
116 Genbank to remove contaminating *Citrobacter* sequences. This left 2,629 sequences with the longest
117 being 12,585 base pairs (bp) (Table 2). The longest 21 sequences are presented with their statistics
118 when examined by NCBI Blastn (nucleotides) and the nucleotides translated into proteins NCBI
119 Blastx presented in bold in Table 2. Only the best hit has similarity recorded while all genes
120 identified by Blastx are listed in order. All sequences hit with either *C. burnetii* (29/42, 69%) or
121 *Coxiella* (10/42, 24%) or rickettsia (3/42, 7%) with an error rate of less than 1 in 1 million for either
122 bp or amino acids.

123 Sequence SRR11188394.2 (Table 2) has the best nucleotide and amino acid similarity (e=0.0, e-124)
124 respectively to *C. burnetii* for carbamoyl-phosphate synthase (large subunit) which is the first step
125 for pyrimidine and arginine synthesis in prokaryotes. Due to its vital function, this gene is likely to be
126 highly conserved. Sequence SRR11188394.16 (Table 2) has the worst nucleotide similarity (e=1.3 to
127 *C. burnetii*). This sequence was retained because of the high amino acid similarity (*Coxiella* sp., e-62)
128 and the interesting nature of the gene. The IS6 family of transposases is associated with jumping
129 genes, is likely to vary considerably between *Coxiella* spp. and isolates and may not be present in all
130 isolates. These sequences are available as a submitted Supplemental File S1. They have been
131 submitted to NCBI Genbank and have been allocated the Biosample number SAMN13874604. Reads
132 are stored in the NCBI Sequence Read Archive (SRA) under run #SRR11188394 with reads allocated
133 numbers SRR11188394.1 to SR11188394.22.

134 *Citrobacter*

135 *Citrobacter* has been a long-time antagonist and/or a commensal of farmed crayfish in Australia
136 (Owens and Evans 1989); recently in China (Shen *et al.* 2005) and indeed in aquaculture in general. It
137 was one of the first bacteria isolated from exoskeleton lesions of marron, *Cherax tenuimanus*
138 (Owens and Evans 1989). It has arisen from time to time in the culture of *C. quadricarinatus* (Owens,

139 unpublished) and silver perch (*Bidyanus bidyanus*) from central Queensland. *Citrobacter* as a
140 member of the Enterobacteriaceae is often considered a common gut commensal of animals, often
141 confused with Salmonella in microbiological cultures and is seen as a secondary pathogen waiting to
142 exploit an immuno-compromised host. However, in a reported Chinese case, a massive mortality in
143 crayfish farms in Zhejiang Province was attributed to *C. freundii.* Unfortunately, the contributing role
144 of Australian RNA iflavirus, bunyavirus, and particularly reovirus, which was known to be
145 transhipped into China in 1992, i.e. Cherax reovirus, was not co-investigated. So, the role of
146 *C. freundii* as a primary pathogen in *Cherax* needs to be further elucidated.

147 There were insufficient reads that passed stringent quality control to fully sequence the genome of
148 *Coxiella cheraxi*, but enough that we believe that this methodology has potential for use on archival
149 samples in the future. The current sequencing should be considered a draft sequence until it can be
150 verified by other sequencing in the future. However, it is clear from the 21 largest sequences and
151 their high statistical significance that this organism is most closely related to *Coxiella burnetti.*

152 **Material and Methods**

153 DNA extraction

154 One rickettsia-positive *C. cheraxi* cephalothorax from the 1998 work had been dorso-ventrally split
155 and stored at -80 ℃. These two halves were homogenized as follows; 2ml of PBS were added to the
156 crayfish in a pestle and this was gently hand ground. The PBS was collected and the process
157 repeated twice. The collected PBS solution was centrifuged at 1000*g* for 5 min to remove
158 particulate matter and the supernatant collected. This process was repeated twice using the
159 supernatant. The supernatant was then centrifuged at 7500*g* for 5 min to produce a bacterial pellet.
160 The presence of Gram-negative rods in the pellet was confirmed with a Gram stain. The pellet was
161 then made up to 200ul with PBS and total DNA extracted using a Qia-amp DNA blood mini kit
162 (Qiagen) utilizing the protocol for isolation of genomic DNA from bacterial suspension cultures. The
163 quantity of DNA was assessed spectroscopically (Implen nanophotometer). The sample was

164 lyophilized in a Telstar Cryodos and resuspended in sufficient molecular biology grade water to
165 produce 400ng of DNA in the volume required for the rapid sequencing kit. The quality of the
166 sample was confirmed via gel electrophoresis.

167 MinION protocol

168 400ng of DNA was processed using the Rapid Sequencing kit (SQK-RAD004, Oxford Nanopore) as per
169 the manufacturer's instructions with the following change; a fault with the thermal cycler during
170 library preparation meant that the sample was digested for a longer time that prescribed. The
171 sample was sequenced on a Flo-Min 106 D flowcell in a MinION (Oxford Nanopore) as per the
172 manufacturer's instructions.

173 Processing of sequencing reads

174 Data was processed in Rstudio version 1.2.1335 (Rstudio team 2015) (http://www.rstudio.com/ 10
175 July 2019) to combine all passed fastq files and adaptors were trimmed using Porechop (Wick 2018)(
176 https://github.com/rrwick/Porechop 10 July 2019). Long read assembly was attempted but there
177 were insufficient long reads for meaniful contigs, so sequence data was directly analysed using
178 Geneious 9.1.8.

179 **Acknowledgements**

180 The MinION sequencing was funded by the Long Read Access Grant (2019) from the James Cook
181 University Centre for Tropical Bioinformatics and Molecular Biology.

182 **References**

183 Cooper AE, Layton R, Owens L, Ketheesan N, and Govan B (2007) Evidence for the classification of a
184 crayfish pathogen as a member of the genus *Coxiella*. *Lett Appl Microbiol* **45**:558-563

185 Duron O (2015) The IS1111 insertion sequence used for detection of *Coxiella burnetii* is widespread
186 in *Coxiella*-like endosymbionts of ticks. *FEMS Microbiol. Lett.s* **362**(17):fnv132. .

187 Edgerton BF and Owens L (1999) Histopathological surveys of the redclaw freshwater crayfish,
188 *Cherax quadricarinatus,* in Australia. *Aquaculture* **180**:23-40

189 Jimenez R and Romero X (1997) Infection by intracellular bacterium in red claw crayfish, *Cherax*
190 *quadricarinatus* (Von Martens), in Ecuador. *Aquac. Res.* **28:** 923-929

191 Ketterer PJ, Taylor DJ, and Prior HC (1992) Systemic rickettsia-like infection in farmed freshwater
192 crayfish, *Cherax quadricarinatus*. In: *Diseases in Asian Aquaculture I.* M. Shariff, R.P. Subasinghe &
193 J.R. Arthur (eds.), p.173-179. Fish Health Section, Asian Fisheries Society, Manila, Philippines.

194 La Fauce K and Owens L (2007) Investigation into the pathogenicity of *Penaeus merguiensis*
195 densovirus (*Pmerg*DNV) to juvenile *Cherax quadricarinatus*. *Aquaculture* **27**:31-38

196 Owens L and Evans LH (1989) Common diseases of freshwater prawns (*Macrobrachium*) and crayfish
197 (marron and yabbies) relevant to Australia. Invertebrate Aquaculture. Proceedings 117 of the
198 Postgraduate Committee Veterinary Science, University of Sydney: 227-240.

199 Owens L, Muir P, Sutton D and Wingfield M (1992) The pathology of microbial diseases in tropical
200 Australian crustacea. In: *Diseases in Asian Aquaculture I.* M Shariff, RP Subasinghe and JR Arthur
201 (eds.), p.165-172. Fish Health Section, Asian Fisheries Society, Manila, Philippines.

202 Powell S (2013) Comparative Study of two Australian *Coxiella* species: *Coxiella burnetii* and
203 *candidatus* Coxiella cheraxi *nov. sp.* (TO-98). MSc Microbiology and Immunology, James Cook
204 University

205 Rang FJ, Kloosterman WP and de Ridder J (2018) From squiggle to basepair: computational
206 approaches for improving nanopore sequencing read accuracy. *Genome Biol.* **19**: 90

207

208 Romero X, Turnball JF and Jimenez R (2000) Ultrastructure and cytopathology of a rickettsia-like
209 organism causing systemic infection of redclaw crayfish, *Cherax quadricarinatus* (Crustacea:
210 Decapoda), in Ecuador. *J. Invertebr.Pathol.* **76**: 95–104

211

212 RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL

213 http://www.rstudio.com/.

214 Shen J, Gu Z, Pan X, Zhou B, Yin W and Cao Z (2005) Isolation and identification of *Citrobacter*

215 *freundii* from *Cherax quadricarinatus*. *Journal of Fishery Sciences of China* **12**(2):197-200 (Abstract

216 only, in Chinese)

217

218 Tan CK and Owens L (2000) The infectivity, transmission & 16S rRNA sequencing of a rickettsia,

219 *Coxiella cheraxi sp. nov.* from the freshwater crayfish, *Cherax quadricarinatus. Dis. of Aquat. Organ.*

220 **41**: 115-122

221 VKM. (2016) Risk assessment on import of Australian redclaw crayfish to Norway, Opinion of the

222 Panel on Animal Health and Welfare, ISBN: 978-82-8259-254-3, Oslo, Norway.

223 Wick, R. (2018) Porechop: adaptor trimmer for Oxford nanopore reads, available at

224 https://github.com/rrwick/Porechop

225 Table 1 Longest sequence similar to *Citrobacter freundii* cosequenced from isolate TO-98.

Sequence No.	Length bp	Best Nucleotide Hit (blastn)	Error rate (e)	Identity (%)	Coverage (%)	Functional Gene (blastx)
SRR11188394.22	9481	*Citrobacter freundii* FDFAARGOS 549	e-0.0	82.6	99	
		Citrobacter freundii	**e-135**	**58**	**15**	**ATP-dependent endonuclease**

226

227

228 Table 2. The longest 21 sequences from TO-98 that are statistically similar to *Coxiella burnettii* or
229 other rickettsia using both blastn (nucleotides) and blastx (translated amino acids, in bold) against
230 the NCBI Genbank database Nucleotide sequences are provided in the supplementary data (S1) and
231 can also be found in the NCBI Sequence read archive (SRA) under run #SRR11188394. Only the best
232 hit of each sequence has been analysed. Where more than one gene hit was found using Blastx, all
233 genes in order are listed under the best hit.

Sequence No.	Length bp	Best Nucleotide Hit (blastn)	Error rate (e)	Identity (%)	Coverage (%)	Functional Gene (blastx)
SRR11188394.1	12585	*C. burnetii* RSA439	e-48	65.3	35	
		candidatus *C. mudrowiae*	**e-49**	**27.5**	**18**	**LPS-assembly protein LptD**
						all genes in order
						aminoglycoside phosphotransferase,LPS-assembly protein LptD,16S RNA S49 family peptidase
SRR11188394.2	11813	*C. burnetii* Dugway	e-0.0	71.9	32	
		C. burnetii	**e-124**	**37.8**	**27**	**carbamoyl-phosphate synthase large subunit**
SRR11188394.3	8247	*C. burnetii* 2014-PE-15890	e-82	69.9	18	
		C. burnetii	**e-46**	**43.4**	**11**	**malate dehydrogenase**
SRR11188394.4	7717	*C. burnetii* RSA439	e-22	73	18	
		***Coxiella* sp.**	**e-33**	**37.1**	**14**	**aminotransferase DegT**
						all genes in order
						ABC transporter ATP-binding protein, aminotransferase DegT
SRR11188394.5	6605	*C. burnetii* Schperling	e-46	66.1	35	
		C. burnetii	**e-24**	**33.5**	**17**	**HlyD family efflux transporter periplasmic adaptor subunit**
						all genes in order
						HlyD family efflux transporter periplasmic adaptor subunit, TolC family protein, primosomal protein N
SRR11188394.6	5365	*C. burnetii* Schperling	e-178	69.6	65	

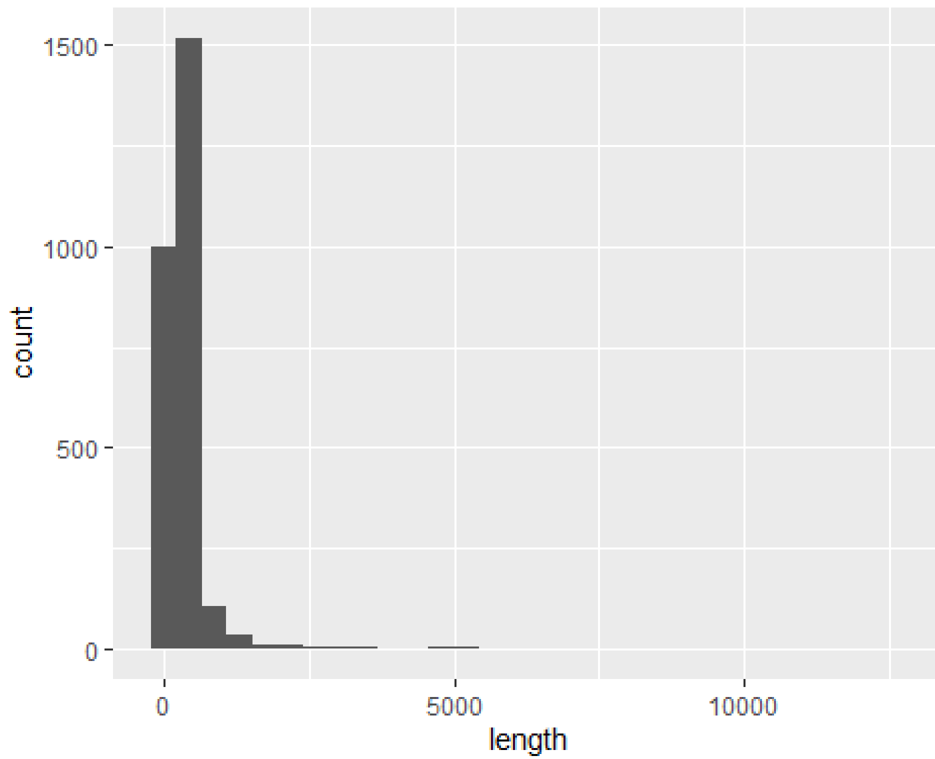
		Coxiella **sp.**	**e-109**	**51**	**35**	**translation initiation factor IF-2**
SRR11188394.7	5011	*C. burnetii* Schperling	e-84	66.5	63	
		C. burnetii	**e-53**	**35.5**	**28**	**NADH-quinone oxidoreductase subunit Nuon** **all genes in order** ***NADH-quinone oxidoreductase subunit Nuon, NADH-quinone oxidoreductase subunit M, NADH-quinone oxidoreductase subunit L***
SRR11188394.8	5006	*C. burnetii* Schperling	e-13	68.5	9	
		C. burnetii	**e-32**	**49**	**14**	**murien L.D-transpeptidase catalytic domain family protein**
SRR11188394.9	4814	*Coxiella*-like from *Amblyomma americanum*	e-12	74.2	2	
		C. burnetii	**e-09**	**57.7**	**3**	**L-aspartate oxidase** **all genes in order** ***L-aspartate oxidase, response regulator transcription factor***
SRR11188394.10	4656	*Wolbachia pipientis* wALB-HN2016	e-14	77.3	5	
		Aquacella lusitana	**e-24**	**37.6**	**17**	**IS481 family transposase/murien L.D-transpeptidase catalytic domain**
SRR11188394.11	4559	*C. burnetii* RSA439	e-15	84.8	17	
		C. burnetii	**e-47**	**32.6**	**47**	**phosphoesterase; TraM recognition domain-containing protein** **all genes in order** ***phosphoesterase; TraM recognition domain-containing protein***
SRR11188394.12	3941	*Aquacella lusitana* SGT-39	e-6	80	1	
		Coxiella **sp.**	**e-8**	**47.9**	**4**	**IS982 family transposase** **all genes in order** ***IS982 family transposase, hyptothetical protein***
SRR11188394.13	3424	*C. burnetii* Scurry	e-29	76.5	17	
		C. burnetii	**e-62**	**38.5**	**38**	**bifunctional 2-methylcitrate dehydratase/aconitate hydratase**

						all genes in order ***bifunctional 2-methylcitrate dehydratase/aconitate hydratase, 2-methylcitrate synthase***
SRR11188394.14	3399	*C. burnetii* MSU goat Q117	e-14	74.6	10	
		C. burnetii	**e-33**	**39.2**	**22**	**MMPL family transporter (multidrug efflux protein)**
SRR11188394.15	3223	*C. burnetii* Heizberg	e-6	88.2	7	
		Coxiella*-like from *Rhipicephalus microplus	**e-21**	**39.7**	**15**	**aspartate-semialdehyde dehydrogenase**
SRR11188394.16	3003	*C. burnetii* RSA439	e = 1.3	100	1	
		***Coxiella* sp.**	**e-62**	**50.6**	**29**	**IS6 family transposase**
SRR11188394.17	2873	*C. burnetii* MSU goat Q117	e-8	86.1	4	
		***Coxiella* sp.**	**e-13**	**38.8**	**10**	**transposase** **all genes in order** ***TSUP family transporter, transposase***
SRR11188394.18	2584	*C. burnetii* 2014-PE-15890	e-10	86.2	4	
		candidatus *C. mudrowiae*	**e-18**	**39.4**	**44**	**NADH-quinone oxidoreductase subunit M** **all genes in order** ***NADH-quinone oxidoreductase subunit M, NADH-quinone oxidoreductase subunit N***
SRR11188394.19	2561	*C. burnetii* RSA439	e-9	81.1	4	
		C. burnetii	**e-27**	**65.8**	**15**	**methylglyoxal synthase**
SRR11188394.20	2388	*C. burnetii* RSA439	e-31	75.1	32	
		***Coxiella* sp.**	**e-65**	**41.7**	**47**	**GTPase HflX**
SRR11188394.21	2371	*C. burnetii* Scurry	e-6	79.6	7	
		C. burnetii	**e-23**	**45**	**18**	**EVE domain-containing protein** **all genes in order** ***transposase, EVE domain-containing protein***

234

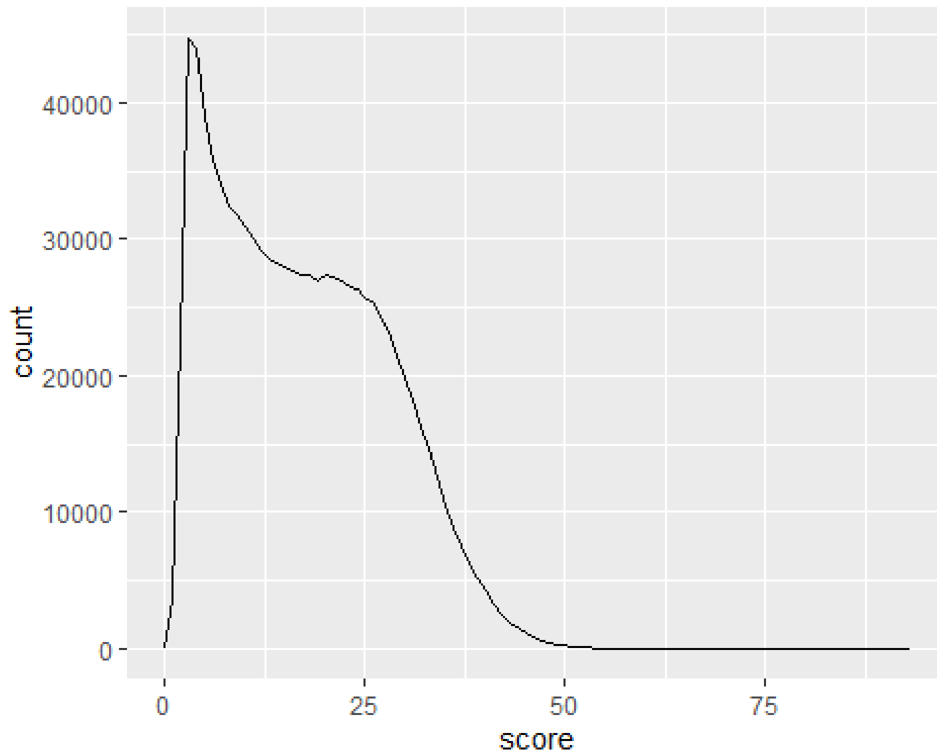
235

236



237

238 Figure 1. Number of reads produced of various lengths. The X axis is length in base pairs. The Y axis
239 is number of reads. Most sequences were less than 2500 bp. Graph produced in RStudio.



240

241 Figure 2. Quality scores of passed reads used for analysis of sequence. There is a peak below Q12.5,

242 a shoulder between Q12.5 and Q25 (up to 99.9% base call accuracy) and a large area up to Q50

243 where base call accuracy was >99.9%. Graph produced in RStudio.

244

245 S1 Sequence data from Table 1

246

247

248

249