Standards in
Genomic Sciences

SHORT GENOME REPORT

Open Access

CrossMark

Genome sequence of *Candidatus* Arsenophonus lipopteni, the exclusive symbiont of a blood sucking fly *Lipoptena cervi* (Diptera: Hippoboscidae)

Eva Nováková[1*], Václav Hypša[1], Petr Nguyen[2], Filip Husník[1] and Alistair C. Darby[3]

Abstract

Candidatus Arsenophonus lipopteni (*Enterobacteriaceae*, *Gammaproteobacteria*) is an obligate intracellular symbiont of the blood feeding deer ked, *Lipoptena cervi* (Diptera: Hippoboscidae). The bacteria reside in specialized cells derived from host gut epithelia (bacteriocytes) forming a compact symbiotic organ (bacteriome). Compared to the closely related complex symbiotic system in the sheep ked, involving four bacterial species, *Lipoptena cervi* appears to maintain its symbiosis exclusively with *Ca. Arsenophonus lipopteni*. The genome of 836,724 bp and 24.8 % GC content codes for 667 predicted functional genes and bears the common characteristics of sequence economization coupled with obligate host-dependent lifestyle, e.g. reduced number of RNA genes along with the rRNA operon split, and strongly reduced metabolic capacity. Particularly, biosynthetic capacity for B vitamins possibly supplementing the host diet is highly compromised in *Ca. Arsenophonus lipopteni*. The gene sets are complete only for riboflavin (B2), pyridoxine (B6) and biotin (B7) implying the content of some B vitamins, e.g. thiamin, in the deer blood might be sufficient for the insect metabolic needs. The phylogenetic position within the spectrum of known *Arsenophonus* genomes and fundamental genomic features of *Ca. Arsenophonus lipopteni* indicate the obligate character of this symbiosis and its independent origin within Hippoboscidae.

Keywords: *Arsenophonus*, Symbiosis, Tsetse, Hippoboscidae

Introduction

Symbiosis has for long been recognized as one of the crucial drivers of evolution. In insects, numerous symbiotic relationships, mainly with bacteria, enabled the hosts to exploit various environments and/or life strategies, and supposedly started adaptive radiations in some groups. The mechanisms of such evolutionary processes include for example contribution to the host immunity, modification of the reproductive strategy, or provision of essential compounds to the hosts relying on nutritionally compromised resources. Blood feeding (hematophagous) insects provide an illuminating example of a life strategy shift coupled with symbiosis. Since blood meal lacks

some of the B vitamins, hematophagous insects rely on their supply by symbiotic bacteria. The relationships between bacteria and hematophagous insect displays considerable degree of variability spanning from less intimate associations with entire gut microbial community, e.g. triatomine bugs [1, 2], to highly specialized interactions with few or single obligate symbiont(s), e.g. lice, bed bugs, tsetse flies, louse flies and bat flies [3–7]. With the recent advancement of genomic approaches and genetic manipulations, symbioses in these insect groups, often important disease vectors, have become of a high interest.

Here we describe fundamental biological characteristics and genome properties of the obligate symbiont of a deer ked, *Lipoptena cervi* (Hippoboscidae). In comparison to multipartite symbiotic systems of closely related hosts from families Hippoboscidae (i.e. *Melophagus ovinus* [7]) and Glossinidae (i.e. *Glossina* sp. [6]), *Lipoptena*

* Correspondence: novaeva@paru.cas.cz
[1]Faculty of Science, University of South Bohemia, and Institute of Parasitology, Biology Centre, ASCR, v.v.i. Branisovka 31, 37005 Ceske Budejovice, Czech Republic
Full list of author information is available at the end of the article

Nováková *et al. Standards in Genomic Sciences* (2016) 11:72

Page 2 of 7

cervi harbors a single unaccompanied obligate symbiont from the genus *Arsenophonus*. The genome of *Candidatus* Arsenophonus lipopteni has been sequenced for two reasons. The first was to extend our knowledge on occurrence and genomics of the obligate symbionts across the spectrum of hematophagous hosts involved in strictly bilateral symbiosis, e.g. bed bugs [5], head lice [4], leaches [8, 9]. This is a necessary prerequisite for the future analysis of the origins and evolution of this kind of symbioses. In addition, we intend to use the sequence in a broader comparative framework focused on evolution of bacterial symbiosis, particularly on its role in B vitamin provision to various ecological types of the hosts.

Organism information
Classification and features

Ca. Arsenophonus lipopteni has an obligate association with its host, *L. cervi*, and is therefore uncultivable. In order to localize the bacteria within the host, Fluorescent In Situ Hybridization and Transmission Electron Microscopy was performed on dissected gut tissue as described in detail in [7]. For FISH, the tissue was fixed and hybridized in tubes with eubacterial (EUB338, Flc-GCTGCCTCCCGTAGGA; [10]) and *Ca. Arsenophonus lipopteni* specific probes (ArL, Cy3-CTGACTAACGC TTGCACC; this study). The later was designed in a variable region of 16S rRNA gene taking the target sequence accessibility into account [11].

The distribution of *Ca. Arsenophonus lipopteni* (Fig. 1) in the host body closely resembles that of *Ca. Arsenophonus melophagi* and *Wigglesworthia glossinidia*[T], the obligate symbionts of the blood sucking flies *Melophagus ovinus* and *Glossina* sp., respectively [7]. Highly pleomorphic cells of the Gram negative non-sporulating bacteria from the family *Enterobacteriales* are primarily found in the modified part of the gut wall (bacteriome)

formed by the specialized enlarged epithelial cells (bacteriocytes, Fig. 1c, 1d). Additional key features of *Ca. Arsenophonus lipopteni* are provided as a standardized summary in Table 1.

Apart from the functional characterization, the genome sequence of *Ca. Arsenophonus lipopteni* was also utilized to assess the relationship of this bacterium to other *Arsenophonus* symbionts. Since the sequence compositional shift compromises phylogenetic usage of 16S rDNA, leading to topological artifacts with long branched symbiotic taxa clustering together [12], we carried out a phylogenetic analysis of a multi-gene matrix and used advanced Bayesian approaches. The matrix was generated for all available *Arsenophonus* genomes (incl. *Ca. Riesia pediculicola*), five other symbionts, eight non-symbiotic members of *Enterobacteriaceae*, and two outgroups. A set of 70 orthologous genes was determined as an intersection of COGs shared by these bacteria (generated using the MicrobesOnline database; [13]) with "SICO" gene list [14]. The genes were retrieved from the finished assembly using Blastp searches [15] and processed as described previously [7]. The resulting matrix contained 22618 unequivocally aligned positions. Phylobayes [16], a tool specifically developed to overcome the difficulty with heterogeneous composition of sequences, was used for the tree reconstruction. The analysis was run in 2 chains under the GTR + CAT model with amino acids recoded according to the Dayhoff6 option. When the convergence was not reached after 20,000 cycles, the program was stopped and majority rule consensus was calculated after discarding 4,000 cycles burn-in.

The results confirm *Ca. Arsenophonus lipopteni* membership in the genus *Arsenophonus*. All *Arsenophonus* species (including *Ca.* Rieisa pediculicola) formed a well-supported monophyletic branch clustering as a
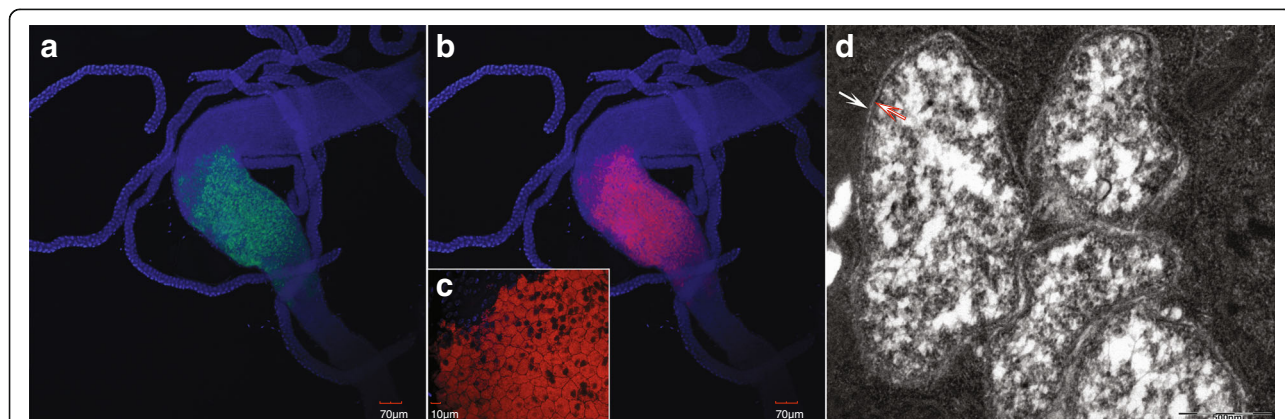

Fig. 1 Visualization of the bacteria in the host tissue using FISH and TEM (D). The symbiotic organ (bacteriome) localized in the midgut section harboring *Ca. Arsenophonus lipopteni* targeted with green (Flc) labeled eubacterial probe (**a**) and red (Cy3) labeled specific probe (**b**). Detail of the host cells (bacteriocytes) filled with the symbionts (**c**). The blue signal is DAPI stained DNA. Four cells of *Ca. Arsenophonus lipopteni* under TEM (**d**). The white arrow points to bacterial outer membrane and the red bordered arrow shows the cytoplasmatic cell membrane

Nováková *et al. Standards in Genomic Sciences* (2016) 11:72

Page 3 of 7

Table 1 Classification and general features of *Ca.* Arsenophpnus lipopteni

MIGS ID	Property	Term	Evidence code[a]	
	Classification	Domain *Bacteria*	TAS	[33]
		Phylum *Proteobacteria*	TAS	[34]
		Class *Gammaproteobacteria*	TAS	[35]
		Order "*Enterobacteriales*"	TAS	[36]
		Family *Enterobacteriaceae*	TAS	[37]
		Genus *Arsenophonus*	TAS	[38]
		Species *Ca. Arsenophonus lipopteni*	IDA	
		Strain: CB	IDA	
	Gram stain	Negative	TAS	[38]
	Cell shape	Pleomorphic	NAS	
	Motility	Non-motile	TAS	[38]
	Sporulation	Non-sporulating	TAS	[38]
	Temperature range	Not determined	IDA	
	Optimum temperature	Not determined	IDA	
	pH range; Optimum	Not determined	IDA	
	Carbon source	Not determined	IDA	
MIGS-6	Habitat	Insect host; bacteriome of *L. cervi*	IDA	
MIGS-6.3	Salinity	Not determined	IDA	
MIGS-22	Oxygen requirement	Not determined	IDA	
MIGS-15	Biotic relationship	Symbiotic	IDA	
MIGS-14	Pathogenicity	Non-pathogen	NAS	
MIGS-4	Geographic location	Ceske Budejovice, Czech Republic	IDA	
MIGS-5	Sample collection date	June 2010	IDA	
MIGS-4.1	Longitude	14.43	IDA	
MIGS-4.2	Latitude	48.97	IDA	
MIGS-4.4	Altitude	399 m	IDA	

[a]Evidence codes, IDA Inferred from Direct Assay, TAS Traceable Author Statement (i.e., a direct report exists in the literature), NAS Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [39]

sister group to *Providencia* (Fig. 2). Despite the length of branches for the obligate symbionts with highly modified genomes, this arrangement was assigned a high posterior probability. Although the six included *Arsenophonus* lineages certainly do not form a monophyletic group within the known *Arsenophonus* spectrum [17], the results indicate that *Ca. Arsenophonus lipopteni* evolved independently from *Ca. Arsenophonus* melophagi housed in related Hippoboscidae host [7].

Genome sequencing information
Genome project history
The host specimens *Lipoptena cervi* were collected from wild populations during summer 2010 in the Czech Republic. Finished genome sequence has been deposited in GenBank under acc. No. CP013920 on January 11, 2016. A summary on the sequencing project is provided in Table 2.

Growth conditions and genomic DNA preparation
Since the bacterium is uncultivable, the host tissue was used for DNA extraction. The gut tissue containing the symbiotic organs were dissected from 6 flies in 1× phosphate buffered saline, homogenized with a sterile mortar and pestle and extracted using QiaAmp DNA Micro Kit (QIAGEN, United Kingdom). The DNA quality was assessed using Agilent 2100 Bioanalyzer (Agilent Technologies).

Genome sequencing and assembly
The paired end 100 bp long reads were generated on one lane of Illumina HiSeq2000 run at Yale Center for Genome Analysis. A5 assembly pipeline with the default settings was used to assemble the reads [18]. Of the 109,640 resulting contigs, the longest contig (836,730 bp) with 40× fold coverage formed a circular molecule with 99 bp overlap at the ends. This contig corresponds to the *Ca.*
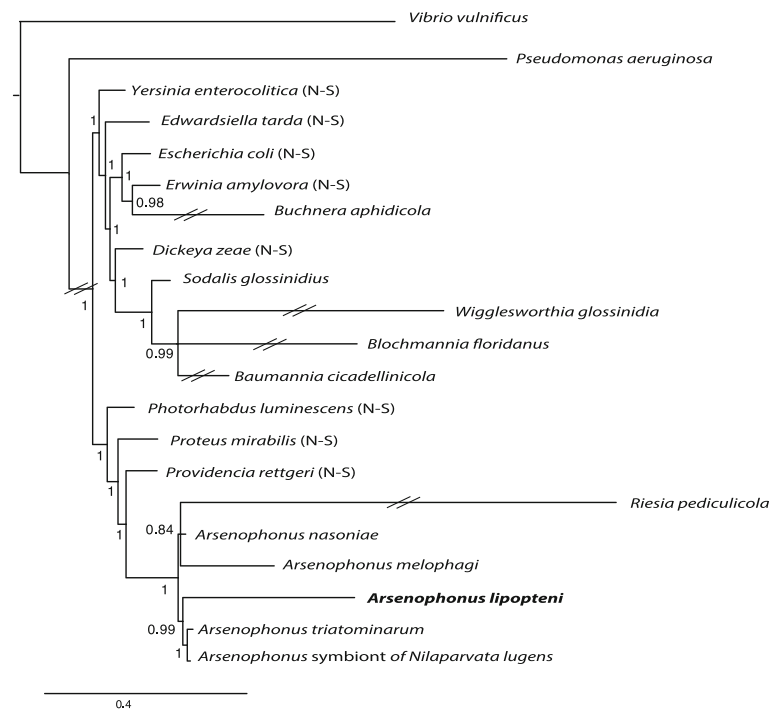
Nováková *et al. Standards in Genomic Sciences* (2016) 11:72

Page 4 of 7

Fig. 2 Phylogenomic reconstruction of *Ca. Arsenophonus lipopteni* position. The length of the double crossed branches was scaled to 1:4. The numbers indicate posterior probability for each node. The "N-S" in the brackets following the taxon name designates the non-symbiotic bacteria included into the dataset

Arsenophonus lipopteni genome. Pilon v1.12 [19] was used to check assembly quality and to improve base calls and small indels.

Genome annotation

The finished genome was annotated using a combination of following tools: RAST [20], PGAP, and Prokka v1.10 [21]. The annotation was then manually curated and

Table 2 Project information

MIGS ID	Property	Term
MIGS 31	Finishing quality	Finished
MIGS-28	Libraries used	2 × 100 bp paired end
MIGS 29	Sequencing platforms	Illumina
MIGS 31.2	Fold coverage	40
MIGS 30	Assemblers	A5
MIGS 32	Gene calling method	RAST, PGAP, PROKKA
	Locus Tag	AUT07
	Genbank ID	CP013920
	Genbank Date of Release	1-25-2016
	GOLD ID	Gp0127464
	BIOPROJECT	PRJNA306001
MIGS 13	Source Material Identifier	Host tissue
	Project relevance	Evolution of bacterial symbiosis

checked for the presence of gene remnants. The final annotation is available in GenBank (CP013920). Metabolic pathways were reconstructed in the RAST server [20] and gene absence was verified using BlastP searches. Proteins were assigned to the clusters of orthologous groups using COGnitor [22], and the presence of signal peptides was detected using SignalP [23]. Pfam domains were predicted using HMMER [24] against the Pfam-A database [25]. Transmembrane predictions were done using TMHMM Server v. 2.0. The search for CRISPR repeats was performed in Geneious [26].

Genome properties

The finished genome consists of 836,724 nucleotides in a single circular chomosome with a low GC content of 24.9 %. The total number of predicted functional genes (667) relative to the genome size implies a lower coding density (75.8 %). The average gene length of 1,001 bp however does not suggest that the genome underwent an extreme economization typical for the obligate symbionts, e.g. *Buchnera aphidicola* str. Cc, *Ca. Sulcia muelleri, Ca. Carsonella ruddii, Ca. Zinderia insecticolla* [27]. Over 99 % of protein coding genes have been assigned to particular COGs and Pfam domains (Tables 3 and 4). Signal peptides and transmembrane helices have been identified for 5 and 124 protein coding genes respectively (Table 3). The noncoding RNA genes consist

Nováková *et al. Standards in Genomic Sciences* (2016) 11:72

Page 5 of 7

Table 3 Statistics for finished genome assebly of *Ca. Arsenophonus lipopteni*

Attribute	Value	% of Total[a]
Genome size (bp)	836,724	100.00
DNA coding (bp)	633,822	75.80
DNA G + C (bp)	208,103	24.90
DNA scaffolds	1	100.00
Total genes	683	100.00
Protein coding genes	625	91.50
RNA genes	42	6.10
Pseudo genes	16	2.30
Genes assigned to COGs	622	99.52
Genes assigned Pfam domains	625	100.00
Genes with signal peptides	5	0.80
Genes with transmembrane helices	124	19.8.00
CRISPR repeats	0	0.00

[a]The total is based on either the size of the genome in bp or the total number of genes

Table 4 Number of protein coding genes assigned to the COG categories

Cat. code	Value	Percentage of total	Description
J	128	20.48	Translation, ribosomal structure and biogenesis
A	1	0.16	Processing and modification
K	17	2.72	Transcription
L	40	6.4	Replication, recombination and repair
B	0	0	Chromatin structure and dynamics
D	15	2.4	Cell cycle control, cell division, chromosome partitioning
Y	0	0	Nuclear structure
V	7	1.12	Defense mechanisms
T	10	1.6	Signal transduction mechanisms
M	70	11.2	Cell wall/membrane biogenesis
N	1	0.16	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	12	1.92	Intracellular trafficking and secretion
O	42	6.72	Posttranslational modification, protein turnover, chaperones
C	22	3.52	Energy production and conversion
G	26	4.16	Carbohydrate transport and metabolism
E	34	5.44	Aminoacid transport and metabolism
F	24	3.84	Nucleotide transport and metabolism
H	49	7.84	Coenzyme transport and metabolism
I	26	4.16	Lipid transport and metabolism
P	24	3.84	Inorganic ion transport and metabolism
Q	1	0.16	Secondary metabolites biosynthesis, transport and catabolism
R	9	1.44	General function prediction only
S	13	2.08	Function unknown
-	51	8.16	Assigned to more than one category
-	3	0.48	Not in COGs

of tmRNA, RNAseP, Alpha RBS, cspA, 35 tRNAs, and 3 rRNA genes (altogether 42 RNA genes). The three ribosomal genes are however not organized into a single operon, a phenomenon previously described for at least 9 unrelated bacterial clades, including gammaproteobacterial symbionts of the genus *Buchnera* and *Candidatus* Blochmannia, and attributed to their host-dependent lifestyle [28].

The genome properties described above coupled with 16 pseudogenes identified in the genome suggest rather recent establishment of the obligate symbiosis resulting in significant but recent gene/function loss without removal of presently non-coding regions. Regarding the coding capacity for B vitamins and related cofactors, the genome of *Ca. Arsenophonus lipopteni* appears to be highly economized. Similar to *Ca. Arsenophonus* melophagi, the bacteria cannot synthesize thiamine (B1), niacin (B3), panthothenic acid (B5) and folic acid (B9). In addition, the genome does not code for heme biosynthesis. Other basic genome characteristics are summarized in Table 3.

Conclusions

Compared to the closely related complex symbiotic system in the sheep ked, *Melophagus ovinus*, *Lipoptena cervi* appears to maintain symbiosis exclusively with *Ca. Arsenophonus lipopteni*. The growing number of genome sequences available for the symbionts and the hematophagous hosts involved in strictly bilateral symbiosis (e.g. [29, 30]) will help elucidating some common requirements on B vitamins, or possibly highlight diverse needs of insects digesting blood of various vertebrates. *Ca. Arsenophonus lipopteni* possesses complete gene sets for biosynthesis of three B vitamins, riboflavin (B2), pyridoxine (B6) and biotin (B7). While the metabolic capacity is directly assessed from genomic data, the presence of any vitamin efflux systems cannot be easily elucidated due to yet poorly understood mechanisms for vitamin export [31]. However, based on recent findings from other hematophagous systems, it has become more clear that the nutritional interaction does not rely on biosynthesis of all B vitamins as originally suggested by Puchta [32]. For instance, similar to all the other *Arsenophonus* genomes, biosynthetic capacity for thiamin is compromised in *Ca. Arsenophonus lipopteni*. The genome however possesses ABC thiamin transporter genes (*thiP, thiQ, tbpA*) implying the content of thiamin or thiamin pyrophosphate, compared to

Nováková *et al. Standards in Genomic Sciences* (2016) 11:72

Page 6 of 7

e.g. biotin or riboflavin, in the host blood might be sufficient for the insect metabolic needs (Novakova, unpublished data). Within the spectrum of known *Arsenophonus* genomes ranging from 0.57 Mb of *Ca. Riesia pediculicola* to 3.5 Mb of *A. nasoniae*, representing various symbiotic types, the genomic sequence of *Ca. Arsenophonus lipopteni* clearly reflects characteristics common for obligate mutualists. Furthermore, the phylogenetic reconstruction suggests an independent origin of this obligate association within Hippoboscidae.

Authors' contributions
EN, AD, VH participated in the design of the study and coordination. EN, FH and PN performed the imaging. VH and FH assembled and annotated the genome sequence. All authors participated in editing of the manuscript and read and approved the final version.

Competing interests
The authors declare that they have no involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

Author details
[1]Faculty of Science, University of South Bohemia, and Institute of Parasitology, Biology Centre, ASCR, v.v.i. Branisovka 31, 37005 Ceske Budejovice, Czech Republic. [2]Faculty of Science, University of South Bohemia, and Institute of Entomology, Biology Centre, ASCR, v.v.i. Branisovka 31, 37005 Ceske Budejovice, Czech Republic. [3]Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK.

References
1. Beard CB, Cordon-Rosales C, Durvasula RV. Bacterial Symbionts of the Triatominae and their potential use in control of Chagas disease transmission 1. Annu Rev Entomol. 2002;47:123–41.
2. Gumiel M, da Mota FF, Rizzo Vde S, Sarquis O, de Castro DP, Lima MM, et al. Characterization of the microbiota in the guts of *Triatoma brasiliensis* and *Triatoma pseudomaculata* infected by *Trypanosoma cruzi* in natural conditions using culture independent methods. Parasit Vectors. 2015;8:245.
3. Hypša V, Křížek J. Molecular evidence for polyphyletic origin of the primary symbionts of sucking lice (Phthiraptera, Anoplura). Microbial Ecol. 2007;54:242–51.
4. Sasaki-Fukatsu K, Koga R, Nikoh N, Yoshizawa K, Kasai S, Mihara M, et al. Symbiotic bacteria associated with stomach discs of human lice. Appl Environ Microbiol. 2006;72:7349–52.
5. Hosokawa T, Koga R, Kikuchi Y, Meng XY, Fukatsu T. *Wolbachia* as a bacteriocyte-associated nutritional mutualist. Proc Natl Acad Sci U S A. 2010;107:769–74.
6. Aksoy S, Chen X, Hypsa V. Phylogeny and potential transmission routes of midgut-associated endosymbionts of tsetse (Diptera: Glossinidae). Insect Mol Biol. 1997;6:183–90.
7. Nováková E, Husník F, Šochová E, Hypša V. *Arsenophonus* and *Sodalis* symbionts in louse flies: an analogy to the *Wigglesworthia* and *Sodalis* system in tsetse flies. Appl Environ Microbiol. 2015;81:6189–99.
8. Nelson M, Graf J. Bacterial symbioses of the medicinal leech *Hirudo verbana*. Gut Microbes. 2012;3:322–31.
9. Manzano-Marín A, Oceguera-Figueroa A, Latorre A, Jiménez-García LF, Moya A. Solving a bloody mess: B-vitamin independent metabolic convergence among gammaproteobacterial obligate endosymbionts from blood-feeding arthropods and the leech *Haementeria officinalis*. Gen Biol Evol. 2015;7:2871–84.
10. Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl Environ Microbiol. 1990;56:1919–25.
11. Behrens S, Rühland C, Inácio J, Huber H, Fonseca A, Spencer-Martins I, et al. In situ accessibility of small-subunit rRNA of members of the domains Bacteria, Archaea, and Eucarya to Cy3-labeled oligonucleotide probes. App Environ Microbiol. 2003;69:1748–58.
12. Charles H, Heddi A, Rahbe Y. A putative insect intracellular endosymbiont stem clade, within the *Enterobacteriaceae*, inferred from phylogenetic analysis based on a heterogeneous model of DNA evolution. C R Acad Sci Ser III Sci Vie. 2001;324:489–94.
13. Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, et al. MicrobesOnline: an integrated portal for comparative and functional genomics. Nucleic Acids Res. 2010;38:D396–400.
14. Lerat E, Daubin V, Moran NA. From gene trees to organismal phylogeny in prokaryotes: the case of the γ-proteobacteria. PLoS Biol. 2003;1:e19.
15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
16. Lartillot N, Lepage T, Blanquart S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics. 2009; 25:2286–8.
17. Nováková E, Hypša V, Moran NA. *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. BMC Microbiol. 2009;9:1–14.
18. Tritt A, Eisen JA, Facciotti MT, Darling AE. An integrated pipeline for de novo assembly of microbial genomes. PLoS One. 2012;7:e42304.
19. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9:e112963.
20. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.
21. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.
22. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28:33–6.
23. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8:785–6.
24. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7: e1002195.
25. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, et al. The Pfam protein families database. Nucleic Acids Res. 2002;30:276–80.
26. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28: 1647–9.
27. McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol. 2012;10:13–26.
28. Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biol Direct. 2009;4:13.
29. Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, et al. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc Natl Acad Sci U S A. 2010;107:12168–73.
30. Attardo GM, Abila PP, Auma JE, Baumann AA, Benoit JB, Brelsfoard CL, et al. Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis. Science. 2014;344:380–6.
31. Jaehme M, Slotboom DJ. Diversity of membrane transport proteins for vitamins in bacteria and archaea. Biochimica et Biophysica Acta (BBA)-General Subjects. 2015;1850:565–76.
32. Puchta O. Experimentelle Untersuchungen über die Bedeutung der Symbiose der Kleiderlaus Pediculus vestimenti Burm. Z Parasitenkd. 1955;17:1–40.
33. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87:4576–9.
34. Garrity GM, Bell JA, Lilburn T. Phylum XIV. *Proteobacteria* phyl. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology, vol. Volume 2, Part B. 2nd ed. New York: Springer; 2005. p. 1.

Nováková *et al. Standards in Genomic Sciences* (2016) 11:72

Page 7 of 7

35. Garrity GM, Bell JA, Lilburn T. Class III. *Gammaproteobacteria* class. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology, vol. Volume 2, Part B. 2nd ed. New York: Springer; 2005. p. 1.

36. Garrity GM, Holt JG. Taxonomic Outline of the Archaea and Bacteria. In: Garrity GM, Boone DR, Castenholz RW, editors. Bergey's Manual of Systematic Bacteriology, vol. Volume 1. 2nd ed. New York: Springer; 2001. p. 155–66.

37. Judicial Commission Conservation of the family name *Enterobacteriaceae*, of the name of the type genus, and designation of the type species OPINION NO. 15. Int Bull Bacteriol Nomencl Taxon. 1958;8:73–4.

38. Werren JH. *Arsenophonus*. Bergey's Manual of Systematic Bacteriology (Vol. 2), G.M. Garrity (ed), New York: Springer-Verlag; 2004:1–4.

39. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.