ENTOMOPATHOGENIC *RICKETTSIELLA* BACTERIA: PHYLOGENOMICS OF A POSSIBLE SOURCE FOR BIOINSECTICIDE DEVELOPMENT

Andreas Leclerque

Institute for Biological Control
Julius Kühn Institute (JKI) - Federal Research Centre for Cultivated Plants
Heinrichstraße 243, 64287 Darmstadt, Germany
tel.: +49-6151-407229
fax: +49-6151-407290
andreas.leclerque@jki.bund.de; leclerque@hotmail.com

Abstract

The application of molecular techniques to the phylogenetics of "rickettsias of insects" has triggered a recent controversy on the most appropriate taxonomic organization of the bacterial genus *Rickettsiella*. Making use of the first whole genome sequence data available from this taxon - namely from an American isolate stemming from the pill-bug, *Armadillidium vulgare* - a combination of phylogenetic reconstruction and likelihood-based significance testing has been employed to establish its phylogeny beyond the 16S rRNA gene level. On the basis of an evaluated set of 181 single-copy orthologous gene (SCOG) families, the present study demonstrates that the pathotype `*Rickettsiella armadillidii*´ is correctly assigned to the gammaproteobacterial order *Legionellales*, but that its further assignment to the taxonomic family *Coxiellaceae* is unsubstantiated. `*R. armadillidii*´ and related *Rickettsiella* bacteria are therefore equally closely related to vertebrate pathogenic bacteria of the genera *Coxiella* and *Legionella*. The consequences of this phylogenomic result for the possible development of a new bioinsecticide will be discussed.

Key words: `*Rickettsiella armadillidii*´, *Legionellales*, *Coxiellaceae*, bioinsecticide

Introduction

Rickettsiella bacteria are intracellular pathogens of a wide range of arthropods that typically multiply in vacuolar structures within fat body cells and are frequently associated with protein crystals (Figure 1). As such, they are of principal interest as a possible biocontrol agent. Due to early descriptions, "rickettsias of insects" were comprised under the genus *Rickettsiella* (Philip) of the alpha-proteobacterial order *Rickettsiales* (Weiss et al. 1984). The genus is primarily subdivided into bacterial pathotypes designated after a strain's original host. The resulting pathotype designation is superposed by the distinction of three recognized species, namely the type species *R. popilliae* (Dutky & Gooden) as well as *R. grylli* (Vago & Martoja), and *R. chironomi* (Weiser). Several pathotypes have been placed in synonymy with each of these species, while others await conclusive species assignment (Garrity et al. 2005).

In conflict with this taxonomic classification, determination of a 16S rRNA-encoding sequence from *R. grylli* (Roux et al. 1997) has revealed highest homology to orthologous genes from the gamma-proteobacterial genera *Coxiella* and *Legionella*, and the entire genus *Rickettsiella* has subseqeuently been reorganized in the taxonomic family *Coxiellaceae* within the order *Legionellales* (Garrity et al. 2005). At the order level, this reorganization receives support from the determination of 16S rRNA-encoding sequences from further *Rickettsiella* pathotypes, e.g. from ticks (Kurtti et al. 2002), collembola (Czarnetzki & Tebbe 2004), aquatic isopods (Cordaux et al. 2007), scarabaeids (Leclerque & Kleespies 2008b), and dipteran insects (Leclerque & Kleespies 2008d). However, further arthropod-associated bacteria originally described as *Rickettsiella* pathotypes (Drobne et al. 1999; Radek 2000) were removed from this taxon and reorganized instead in the candidate genus `*Rhabdochlamydia*´ of the order *Chlamydiales* after the respective 16S rRNA-encoding sequences had

been determined (Kostanjsek et al. 2004; Corsaro et al. 2007). Nevertheless, monophyly of the genus *Rickettsiella* has been claimed (Cordaux et al. 2007) and critically discussed (Leclerque 2008b)
Moreover, the genome sequence of the pathotype `*Rickettsiella armadillidii*´, a pathogen of the pill-bug, *Armadillidium vulgare*, has meanwhile been established and found to contain two identical copies of a 16S rRNA gene with highest homology to orthologs from *Rickettsiella*-like bacteria assigned to the order *Legionellales* (Leclerque & Kleespies 2008a). In particular, the existence of multiple spatially dispersed 16S rRNA operons in the *Rickettsiella* genome makes interoperonic heterogeneity a possible explanation of the observed diverging 16S rRNA-based phylogenies of different *Rickettsiella* pathotypes. Due to these possible ambiguities at the16S rRNA level and as unambiguous systematic characterization is a prerequisite for the evaluation of an organism's biocontrol potential, the present study aims to deliver a sound taxonomic assignment of the genus *Rickettsiella* beyond the 16S rRNA level by extensively mining the `*R. armadillidii*´ genome data.

Material and Methods

The `*R. armadillidii*´ genome sequence (GenBank accession number NZ_AAQJ00000000) was compared to 20 further completely annotated bacterial genomes. In order to identify gene families consisting of single-copy orthologs (SCOGs) across diverse subsets produced from this genome space, the criterion of "best reciprocal hits" was implemented at the level of deduced amino acid sequences by a series of BlastP searches (Altschul et al. 1997) with the presumed ortholog from each of the genomes in question being in turn used as query sequence. A gene family was selected for downstream analyses if each BlastP run identified the same set of best hits containing exactly one sequence from each genome. As annotation of the *Rickettsiella grylli* genome sequence is not yet complete, identified orthologs from *R. grylli* were used as query sequence in an additional tBlastN search on the genome draft to make sure that no unannotated paralogs had escaped our notice.

Nucleotide and amino acid sequence alignments were produced with the CLUSTALW function (Thompson et al. 1994) of the MEGA 4 program (Tamura et al. 2007) using an IUB DNA and a Gonnet protein weight matrix, respectively, with alignments of protein-encoding DNA sequences being filtered for potentially hypervariable sites allowing for synonymous substitution. The TREE-PUZZLE 5.2 program (Schmidt et al. 2002) was used to estimate data set-specific parameters as amino acid and nucleotide frequencies, percentages of invariable sites, transition/transversion ratios, and alpha parameters for the gamma-distribution based correction of rate heterogeneity among sites. Based on these parameters, the Hasegawa-Kishino-Yano (HKY) model (Hasegawa et al. 1985) was chosen along the rationale outlined by Posada & Crandall (1998) as the most appropriate common model of sequence evolution for phylogenetic analyses of our nucleotide data (including 16S rRNA gene sequence alignments). Amino acid sequence comparisons were performed using the Jones-Taylor-Thornton (JTT) model of substitution (Jones et al. 1992). Under these model assumptions, organismal phylogenies were reconstructed from both DNA and protein sequence alignments by three alternative approaches: (i) with the Maximum Likelihood (ML) method as implemented in the PhyML software tool (Guindon & Gascuel 2003) and in MEGA 4 with both (ii) the Minimum Evolution (ME) method using the Close-Neighbor-Interchange (CNI) algorithm and (iii) the Neighbor Joining (NJ) method. In all cases, pair-wise deletion for missing data and a□-distribution-based model of rate heterogeneity (Yang 1993) allowing for eight different rate categories were applied. Tree topology confidence limits were explored in non-parametric bootstrap analyses over 1,000 pseudo-replicates using the respective function of the software tools used for phylogenetic reconstruction. Individual phylogenetic trees were condensed into extended majority rule consensus tree topologies by means of the CONSENSE module of the PHYLIP 3.6 software package (Felsenstein 2004). Candidate topologies for significance testing were generated manually as text files employing Newick format.

Likelihood-based non-parametric significance testing of these candidate tree topologies was performed using the Shimodaira-Hasegawa or SH-test (Shimodaira & Hasegawa 1999) at the 5% significance level as implemented in the TREE-PUZZLE 5.2 software package. From a given set of candidate phylogenetic tree topologies, the test firstly determines which topology fits best the phylogenetic information present in an underlying DNA or protein sequence alignment, i.e. it selects a

ML tree. Secondly, the test positively rejects those topologies that are significantly worse interpretations of the sequence data than this best tree with respect to an exogenously established significance threshold. For a more detailed description of the relevant bioinformatic methodology as well as the identity of the genomes and SCOG families compared and the complete set of permutative tree topologies evaluated in the SH-test please refer to Leclerque (2008a).

Results

Order level classification of *`Rickettsiella armadillidii´*

Bioinformatic analysis of the *`R. armadillidii´* genome sequence revealed the presence of single copies for 194 out of 211 gene families that had been identified earlier as a core set of single-copy orthologs (SCOGs) across the *Gammaproteobacteria* (Ciccarelli et al. 2006; Lerat et al. 2003). In order to reassess the taxonomic position of *`R. armadillidii´*, its genome sequence was compared to 16 gamma-proteobacterial and one rickettsial, i.e. alpha-proteobacterial, genomes, and 132 genes from the SCOG core set were identified across this genome space. These were further grouped into three subsets of gene families made up of 88 informational, 44 operational, and 35 universally panorthologous families, respectively. For each SCOG subset, metaprotein data sets comprising 35,772, 19,685, and 11,831 sites, respectively, were created by sequence concatenation and alignment. Subsequent phylogenetic reconstruction by the ML, ME, and NJ method gave rise to nine different hypothetical organismal phylogenies represented in Figure 2.

These nine phylogenies firstly coincide in placing the metaproteins generated from genes of *`R. armadillidii´* in a single branch together with the sequences representing the genera *Legionella* and *Coxiella.* Optimal bootstrap support at the root of this branch makes it a consistent representation of the presumed *Legionellales* clade within the *Gammaproteobacteria.*

Family level classification of *`Rickettsiella armadillidii´*

Secondly, the nine metaprotein trees can be subdivided into two groups of topologies (Figure 2) according to the *internal* organization of the *Legionellales* clade, with the *`R. armadillidii´* sequence being located in a sister clade position either relative to the *Coxiella* (Figure 2A) or with respect to the *Legionella* clades (Figure 2B). To solve this apparent contradiction, we have used likelihood-based significance testing in a model system comprising 6 bacterial genomes: two sequences each from the genera *Coxiella* and *Legionella*, the *`R. armadillidii´* sequence, and the *E. coli* genome as an exogenously fixed outgroup. Exhaustive permutation of these genomes over two bifurcative topological backbone structures generated a total of 105 possible candidate tree topologies that were evaluated for the 181 SCOG families from the basic core set identified across this genome space using the SH-test. In line with expectations from the above analysis, all but candidate topologies A through C (Figure 3) were positively rejected as significantly worse descriptions of *each* of the 181 SCOG family-specific alignment data. The 181 SCOG family-specific best trees were unevenly distributed over these topologies: 87, 78, and 23 best-trees had topology A, B, and C, respectively. However, second best trees of topology A, B or C were *in no case* rejected.

Discussion

The phylogenomic approach presented here on the one hand unambiguously corroborates the current assignment of the pathotype *`Rickettsiella armadillidii´* to the taxonomic order *Legionellale*s; irrespective of the bioinformatic method employed, metaprotein sequences generated from the *`R. armadillidii´* genome are placed with 100% bootstrap support into a single clade together with the orthologous sequences from *Legionella* and *Coxiella* bacteria.

On the other hand, the present analysis does *not* lend support to the current assignment of the genus *Rickettsiella* to the taxonomic family *Coxiellaceae*. Likelihood-based significance testing identified three possible internal structures (Figure 3) of the presumed *Legionellales* clade that are not rejected as descriptions of any of the 181 sets of SCOG family-specific sequence data, i.e. the SH-test does not discriminate among these topologies. As the evaluated set of 105 candidate topologies has

been permutatively complete and the remaining 102 clade structures were clearly and always rejected, this failure to discriminate is not due to a lack of phylogeny useful information in the underlying sequence data, i.e. topologies 3A, 3B, and 3C are the best and equally good descriptions of a *bifurcative* internal structure of the *Legionellales* clade. In other words, and with the important limitation that the `*R. armadillidii*´ is currently the only available representative of its genus, the present phylogenomic analysis shows that no pair of the genera *Legionella*, *Coxiella*, and *Rickettsiella* is more closely related to each other than to the third genus. This result contradicts the current family-level assignment of the genus *Rickettsiella* and is consistent with the introduction of an own taxonomic family *Rickettsiellaceae* within the order *Legionellales*.

The above corroborated taxonomic classification of the pathotype `*Rickettsiella armadillidii*´ systematically places *Rickettsiella* bacteria in close vicinity to vertebrate and even human pathogenic *Legionella* and *Coxiella*. Under biosafety and legalization aspects, this finding puts some doubt on the suitability of living *Rickettsiella* bacteria as biocontrol agents. This might be the more relevant as preliminary evidence suggests that pathogenesis mechanisms in these bacteria are operated by the same key virulence factors, e.g. nuclear genome encoded type-IVB secretion systems (Leclerque & Kleespies 2008c). At the same time, these similarities might offer a starting point for the development of *Rickettsiella*-based biocontrol agents beyond the whole organism approach.

List of Abbreviations

CNI	Close Neighbor Interchange algorithm
ME	Minimum Evolution method of phylogenetic reconstruction
ML	Maximum Likelihood method of phylogenetic reconstruction
MP	Maximum Parsimony method of phylogenetic reconstruction
NJ	Neighbor Joining method of phylogenetic reconstruction
SCOG	single copy orthologous genes
SH	Shimodaira Hasegawa likelihood-based significance test

References

ALTSCHUL SF, MADDEN TL, SCHAFFER AA, ZHANG J, ZHANG Z, MILLER W, & LIPMAN DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.

CICCARELLI, F.D., DOERKS, T., VON MERING, C., CREEVEY, C.J., SNEL, B. & BORK, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. Science 311: 1283-1287.

CORDAUX, R., PACES-FESSY, M., RAIMOND, M., MICHEL-SALZAT, A., ZIMMER, M. & BOUCHON, D. (2007). Molecular characterization and evolution of arthropod pathogenic *Rickettsiella* bacteria. Appl. Environm. Microbiol. 73: 5045-5047.

CORSARO, D., THOMAS, V., GOY, G., VENDITTI, D., RADEK, R. & GREUB, G. (2007). 'Candidatus *Rhabdochlamydia crassificans'*, an intracellular bacterial pathogen of the cockroach *Blatta orientalis* (Insecta: Blattodea). Syst. Appl. Microbiol. 30: 221-228.

CZARNETZKI, A.B. & TEBBE, C.C. (2004). Diversity of bacteria associated with Collembola – a cultivation-independent survey based on PCR-amplified 16S rRNA genes. FEMS Microbiol. Ecol. 49: 217-227.

DROBNE, D., STRUS, J., ZNIDARSIC, N. & ZIDAR, P. (1999). Morphological description of bacterial infection of digestive glands in the terrestrial isopod *Porcellio scaber* (Isopoda, Crustacea). J. Invertebr. Pathol. 73: 113-119.

FELSENSTEIN J (2004). PHYLIP (Phylogeny Inference Package), version 3.6. Department of Genome Sciences, University of Washington, Seattle, WA, 2004. Available from http://evolution.gs.washington.edu/phylip.html.

GARRITY, G.M., BELL, J.A. & LILBURN, T. (2005). Family II. *Coxiellaceae* fam. nov.. *Bergey's Manual of Systematic Bacteriology*, Vol. II, Part B (Garrity GM, Brenner DJ, Krieg NR & Staley JT, eds), pp. 237-247. Springer, New York.

GUINDON S & GASCUEL O (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696-704.

HASEGAWA M, KISHINO H & YANO T-A (1985). Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22: 160–174.

JONES DT, TAYLOR WR & THORNTON JM (1992). The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8: 275–282.

KOSTANJSEK, R., STRUS, J., DROBNE, D. & AVGUSTIN, G. (2004). 'Candidatus *Rhabdochlamydia porcellionis*', an intracellular bacterium from the hepatopancreas of the terrestrial isopod *Porcellio scaber* (Crustacea: Isopoda). Int. J. Syst. Evol. Microbiol. 54: 543-549.

KURTTI, T.J., PALMER, A.T. & OLIVER, J.H. (2002). *Rickettsiella*-like bacteria in *Ixodes woodi* (Acari: Ixodidae). J. Med. Entomol. 39: 534-540.

LECLERQUE, A. (2008a). Whole genome-based assessment of the taxonomic position of the arthropod pathogenic bacterium *Rickettsiella grylli*. FEMS Microbiol. Lett. 283: 117-127.

LECLERQUE, A. (2008b). Reorganization and Monophyly of the Genus *Rickettsiella*: All in Good Time. Appl. Environm. Microbiol. 74: 5263-5264.

LECLERQUE, A. & KLEESPIES, R.G. (2008a). The taxonomic position of the entomopathogenic bacterium *Rickettsiella grylli*: 16S rRNA genes and beyond. IOBC/wprs Bull. 31: 99-104.

LECLERQUE, A. & KLEESPIES, R.G. (2008b). 16S ribosomal RNA, GroEL, and MucZ based assessment of the taxonomic position of *Rickettsiella melolonthae* and its implications for the organization of the genus *Rickettsiella*. Int. J. Syst. Evol. Microbiol. 58: 749-755.

LECLERQUE, A. & KLEESPIES, R.G. (2008c). Type IV secretion system components as phylogenetic markers of entomopathogenic bacteria of the genus *Rickettsiella*. FEMS Microbiol. Lett. 279: 167-173.

LECLERQUE, A. & KLEESPIES, R.G. (2008d). Genetic and electron-microscopic characterization of *Rickettsiella tipulae*, an intracellular bacterial pathogen of the crane fly, *Tipula paludosa*. J. Invertebr. Pathol. 98: 329-334.

LERAT, E., DAUBIN, V. & MORAN, N.A. (2003). From gene trees to organismal phylogeny in prokaryotes: The case of the □-proteobacteria. PLoS Biol. 1: 101-109.

POSADA D & CRANDALL KA (1998). MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817-818.

RADEK, R. (2000). Light and electron microscopic study of a *Rickettsiella* species from the cockroach *Blatta orientalis*. J. Invertebr. Pathol. 76: 249-256.

ROUX, V., BERGOIN, M., LAMAZE, N. & RAOULT, D. (1997). Reassessment of the taxonomic position of *Rickettsiella grylli*. Int. J. Syst. Bacteriol. 47: 1255-1257.

SCHMIDT HA, STRIMMER K, VINGRON M & VON HAESELER A (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502-504.

SHIMODAIRA, H. & HASEGAWA, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16: 1114–1116.

TAMURA K, DUDLEY J, NEI M & KUMAR S (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596-1599.

THOMPSON JD, HIGGINS DG & GIBSON TJ (1994). CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.

WEISS, E., DASCH, G.A. & CHANG, K.-P. (1984). Genus VIII. *Rickettsiella* Philip 1956. *Bergey's Manual of Systematic Bacteriology*, Vol. I (Krieg NR & Holt JG, eds.), pp. 713-717. Williams & Wilkins, Baltimore, MD.

YANG Z (1993). Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol Biol Evol 10: 1396-1401.

Figures

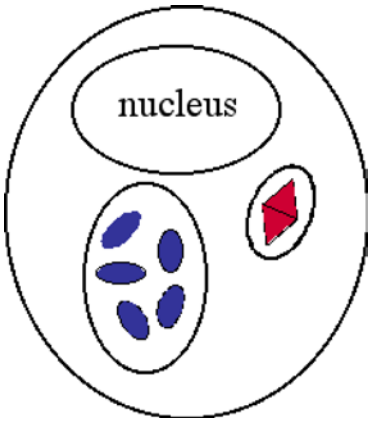

Figure 1 Schematic representation of the cytology of *Rickettsiella* infected eukaryotic cells. *Rickettsiella* bacteria (blue) replicate inside cytoplasmic vacuoles. Bacterial replication is often associated with the appearance of membrane-bounded protein crystals (red) in the cytoplasm.

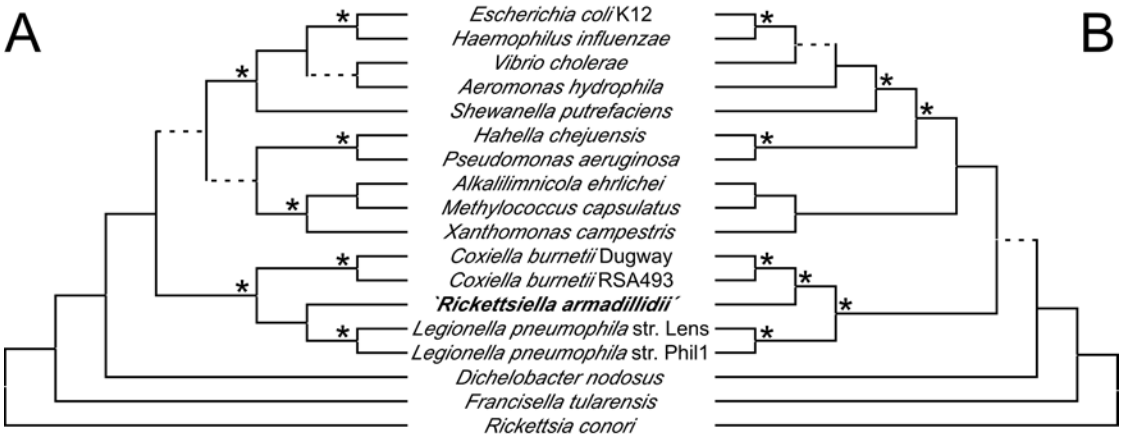

Figure 2. Extended majority rule consensus trees generated from nine hypothetical organismal phylogenies. Original phylograms were grouped according to the alternative *Legionellales* clade topologies. Both groups are represented by consensus trees A (6 phylogenies) and B (3 phylogenies). Dashed lines denote branches that are not present in all underlying topologies; the asterisk indicates optimal bootstrap support in all phylograms.

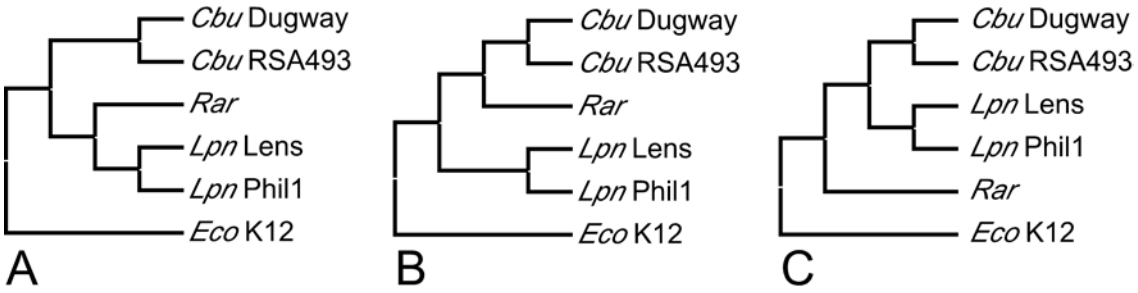

Figure 3. *Legionellales* clade topologies that are not rejected by the SH-test. *Cbu, Coxiella burnetii; Lpn, Legionella pneumophila; Rar, Rickettsiella armadillidii; Eco, Escherichia coli.*