Castelli Michele (Orcid ID: 0000-0003-2177-5772)
Nardi Tiago (Orcid ID: 0000-0002-1248-9873)
Sassera Davide (Orcid ID: 0000-0003-0442-5252)

Title: "*Candidatus* Sarmatiella mevalonica" endosymbiont of the ciliate *Paramecium* provides insights on evolutionary plasticity among *Rickettsiales*

Running title: *Sarmatiella*'s genome and *Rickettsiales* plasticity

Michele Castelli[1#], Olivia Lanzoni[2,3#], Tiago Nardi[1], Stefano Lometto[1], Letizia Modeo[2,4], Alexey Potekhin[5], Davide Sassera[1]*, Giulio Petroni[2]*

1. Dipartimento di Biologia e Biotecnologie, Università degli studi di Pavia, Pavia, Italy

2. Dipartimento di Biologia, Università di Pisa, Pisa, Italy

3. Department of Food Hygiene and Environmental Health, University of Helsinki, Helsinki, Finland

4. CISUP, Centro per l'Integrazione della Strumentazione dell'Università di Pisa, Pisa, Italy,

5. Department of Microbiology, Faculty of Biology, Saint Petersburg State University, Saint Petersburg, Russia

#These authors contributed equally to the work

*Corresponding authors:

Giulio Petroni: Dipartimento di Biologia, Università di Pisa, Via Volta 4/6, 56126, Pisa, Italy, +39 050 2211384; giulio.petroni@unipi.it

Davide Sassera: Dipartimento di Biologia e Biotecnologie, Università degli studi di Pavia, Via Ferrata 9, 27100, Pavia, Italia, +39 0382 986028, davide.sassera@unipv.it

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1462-2920.15396

This article is protected by copyright. All rights reserved.

Accepted Article

Accepted Article

Originality-Significance Statement

Rickettsiales engage obligate relationships with eukaryotes and include human pathogens, reproductive manipulators of invertebrates, as well as many still poorly investigated intracellular bacteria in aquatic hosts. Phylogenetic relationships of *Rickettsiales* associated with diverse hosts are highly interwoven, i.e., with few exceptions, with no evidence of host-symbiont co-evolution.

The herein characterised "*Candidatus* Sarmatiella mevalonica" is a novel *Rickettsiales* associated with *Paramecium*. "*Candidatus* Sarmatiella" presents the distinctive mevalonate pathway for the synthesis of isoprenoids, typical in eukaryotes but rare in bacteria. Additionally, we found genes of this pathway only in three metagenome-derived *Rickettsiales* assemblies. Accordingly, a scenario of multiple independent horizontal gene transfer events was delineated. We predict this pathway enables the bacteria to scavenge metabolic intermediates from the host, allowing to synthesise multiple key metabolites, and possibly also to obtain energy.

Therefore, an enhanced scenario of the *Rickettsiales* plasticity is presented, in particular concerning the role of horizontal gene transfer. Moreover, this is a sharp example of evolutionary convergence which may offer a basis for re-evaluating the implications of convergence also for other more nuanced genetic and phenotypic traits along *Rickettsiales* evolution.

Summary

Members of the bacterial order *Rickettsiales* are obligatorily associated with a wide range of eukaryotic hosts. Their evolutionary trajectories, in particular concerning the origin of shared or differential traits among distant sub-lineages, are still poorly understood. Here we characterised a novel *Rickettsiales* bacterium associated with the ciliate *Paramecium tredecaurelia,* and phylogenetically related to the *Rickettsia* genus. Its genome encodes significant lineage-specific features, chiefly the mevalonate pathway gene repertoire, involved in isoprenoid precursor

biosynthesis. Not only this pathway has never been described in *Rickettsiales,* it also is very rare among bacteria, though typical in eukaryotes, thus likely representing a horizontally-acquired trait. The presence of these genes could enable an efficient exploitation of host-derived intermediates for isoprenoid synthesis. Moreover, we hypothesise the reversed reactions could have replaced canonical pathways for producing acetyl-CoA, essential for phospholipid biosynthesis. Additionally, we detected phylogenetically unrelated mevalonate pathway genes in metagenome-derived *Rickettsiales* sequences, likely indicating evolutionary convergent effects of independent horizontal gene transfer events. Accordingly, convergence, involving both gene acquisitions and losses, is highlighted as a relevant evolutionary phenomenon in *Rickettsiales*, possibly favoured by plasticity and comparable lifestyles, representing a potentially hidden origin of other more nuanced similarities among sub-lineages.

Keywords

horizontal-gene transfer (HGT), mevalonate, convergent evolution, *Rickettsia*-like organism (RLO), protist, Protozoa, bacterial symbiont, parasite

Accepted Article

Accepted Article

Introduction

Rickettsiales sensu stricto (Szokoli *et al.* 2016a; Muñoz-Gómez *et al.* 2019) are a large monophyletic lineage of bacteria living in obligate association with eukaryotic cells, highly variegated in host range and kinds of interaction (Perlman *et al.* 2006; Beckmann *et al.* 2019; Walker and Ismail 2008; Rihikisa 2010; Castelli *et al.* 2016). The vast majority of characterised cases replicate strictly intracellularly, with the significant exception of the recently described "*Candidatus* Deianiraea vastatrix" (Castelli *et al.* 2019a). Overall, *Rickettsiales* present rather flexible lifestyles, considering that they are not restricted to single host lineages, being able of horizontal switch even between rather different host species, as shown by direct evidence and by inference along their evolutionary history (e.g. Epis *et al.* 2009; Matsuura *et al.* 2012; Braig *et al.* 1994; Modeo *et al.* 2020). This resembles other unrelated eukaryote-associated bacterial lineages, such as *Chlamydiae* and *Legionellales* (Horn *et al.* 2004; Duron *et al.* 2018). The plasticity of *Rickettsiales* is reflected also in their genome features, displaying only part of the paradigmatic traits of vertically transmitted obligate symbionts (McCutcheon and Moran 2012), in particular reduced size and streamlined metabolism. Alongside, they present relevant variability, for example in the genes involved in the interaction with the hosts (Walker and Ismail 2008; Rikihisa 2010; Martijn *et al.* 2015; Beckmann *et al.* 2019), as well as in some metabolic features (Dunning-Hotopp *et al.* 2006; Sassera *et al.* 2011; Driscoll *et al.* 2017; Castelli *et al.* 2019a). Such differences could be partly promoted by their lifestyle, allowing exposure to mobile genetic elements (phages, plasmids, transposons), in contrast to strictly vertically transmitted lineages. Such elements may be potential drivers of intragenomic recombination and of horizontal gene transfer (HGT) from external sources. Multiple reports describe specific HGT events in *Rickettsiales*, but the overall frequency and impact of such mechanisms in the evolution of the members of this order is not fully

understood (Ogata *et al.* 2006; Nikoh *et al.* 2014; El Karkouri *et al.* 2016; Gillespie *et al.* 2012; Kent and Bordenstein 2010; Duplouy *et al.* 2013; Klasson *et al.* 2009; Wang and Wu 2017).

Currently, there is wide consensus that the *Rickettsiales* ancestor (Proto-Rickettsiales) was most likely associated with an aquatic unicellular host (Vannini *et al.* 2004; Weinert *et al.* 2009; Ogata *et al.* 2006; Kang *et al.* 2014). However, many different open questions still remain to be elucidated on its physiological, metabolic and functional features, such as whether it was already intracellular, whether it possessed flagella and used them for motility and/or secretion, whether it was capable of anaerobic respiration, and which were its biosynthetic abilities (Sassera *et al.* 2011; Castelli *et al.* 2019a).

Many different lineages of *Rickettsiales* are currently known (over 30 described genera, new descriptions appearing every year), associated with phylogenetically and environmentally diverse hosts, and host and symbionts phylogenies are highly interwoven (Castelli *et al.* 2016, Matsuura *et al.* 2012; Epis *et al.* 2008; Gruber-Vodicka *et al.* 2019; George *et al.* 2020). However, to date only a minor part of such diversity was investigated by genomics. Indeed, most of our knowledge is derived from vector-borne pathogenic members, in particular *Rickettsia* spp. (Ogata *et al.* 2006; Walker and Ismail 2008; Driscoll *et al.* 2017; El Karkouri *et al.* 2016; Gillespie *et al.* 2012), and from *Wolbachia*. The latter displays peculiar interactions ranging from manipulation of host reproduction to necessary mutualism in its arthropod and nematode hosts (Taylor *et al.* 2005; Beckmann *et al.* 2019; Nikoh *et al.* 2014; Duplouy *et al.* 2013; Klasson *et al.* 2009). Some additional studies (Schulz *et al.* 2016; Floriano *et al.* 2018; Yurchenko *et al.* 2018; Gruber-Vodicka *et al.* 2019; Wang and Wu 2017; Klinges *et al.* 2019; Castelli *et al.* 2019a; George *et al.* 2020; Olivieri *et al.* 2019) provided relevant information on recent evolution of other *Rickettsiales* lineages and on the specific interaction mechanisms with their hosts, but the available insights on the evolution of the whole order are still limited. Indeed, general comparative analyses are strongly hampered by the relatively high evolutionary distance among available *Rickettsiales* genomes,

Accepted Article

making it difficult to discern the effects of different evolutionary events. In practice, it may be hardly possible to distinguish ancestral features derived from the Proto-Rickettsiales from the results of HGT events, especially if these are ancient enough, considering the high sequence evolutionary rates of *Rickettsiales* and multiple plausible scenarios of differential gene retention/loss patterns (Castelli *et al.* 2019a). Thus, extended taxonomic sampling and subsequent genomics will probably be a prerequisite for elucidating the evolutionary origin of many relevant traits of the evolution and diversification of *Rickettsiales*.

Here, we present the morphological, phylogenetic and genomic characterisation of "*Candidatus* Sarmatiella mevalonica", a novel *Rickettsiales* species found as endosymbiont of the ciliate *Paramecium tredecaurelia*, and first representative of a new genus belonging to the family *Rickettsiaceae*. This organism presents distinctive genomic and predicted metabolic traits, in particular the biosynthetic pathways of isoprenoids. This offers a useful model to approach the origin of some peculiar pathways in *Rickettsiales*, for which we propose and discuss a possible evolutionary scenario.

Accepted Article

Accepted Article

Results

Characterisation of *Paramecium tredecaurelia* WO2 and its bacterial endosymbiont

Strain WO2 was morphologically identified as a member of the *Paramecium aurelia* complex. Subsequently, molecular characterisation using three marker genes allowed to assign it by homology searches on NCBI Nucleotide to the species *Paramecium tredecaurelia* (Supplementary table S1).

Preliminary live observations by DIC evidenced numerous and seemingly motile intracellular cytoplasmic bacteria in *P. tredecaurelia* WO2 cells (A. Potekhin, personal observation).

An almost complete 16S rRNA gene (MT984300, 1,398 bp) was obtained by PCR with broad range primers and sequenced. Its best blastn hit on NCBI Nucleotide (release 237) was another endosymbiont of *Paramecium*, the *Rickettsiaceae* bacterium "*Ca*. Spectririckettsia obscura" (MH194582; 92.6% identity; 95% query coverage) (Castelli *et al.* 2019b). FISH with two newly designed specific probes in conjunction with broad range ones confirmed the presence of numerous symbionts inside the host cells (Fig 1; Supplementary figure S1), and excluded the presence of other different intracellular bacteria. The almost universal bacterial probe EUB338 had many (ten) mismatches with the 16S rRNA gene of the symbiont (data not shown), similarly to the closely related "*Ca*. Spectririckettsia obscura", likely due to high sequence divergence (Castelli *et al.* 2019b) and, consistently, produced no signal for the bacterium.

According to electron microscopy observations, the symbionts of *P. tredecaurelia* measured about 1.6 µm x 0.4 µm and showed the typical Gram-negative rod-shape, with two membranes spaced apart about 1.6 nm encircling a homogeneous cytoplasm (Fig 2a,b). Bacterial cytoplasm was

slightly electron dense; no other structures were visible except abundant ribosomes (Fig 2a) and, occasionally, electron-lucid ‘‘holes’’ (diameter: ~0.13 µm) (Fig 2c). According to their size and appearance, these “holes” possibly represent polyhydroxyalkanoate (PHA) granules (Sabaneyeva *et al.*. 2018; Chee *et al.* 2010). No flagella or pili were detected, but, as typical of *Rickettsia* and its relatives (Silverman *et al.* 1978; Lanzoni *et al.* 2019), an electron-lucid halo was found surrounding bacterial cells, which appeared in direct contact with host cytoplasm (i.e., no host symbiosomal membrane was observed) (Fig 2a), frequently located in proximity to host lipid droplets (Fig 2b). Some bacterial cells were observed inside areas delimited by single-layered, host-derived membranes in the vicinity of horseshoe-like cisterns resembling phagophores (Fig 2c). Although the majority of the internalised bacteria did not diverge in morphology from those free in the cytoplasm, some appeared highly degraded, but still recognisable by the presence of the typical encircling electron-lucid halo; this observation suggests that the membranous structures might indeed represent digestive vacuoles (i.e., autophagosomes). The irregular autophagosomes likely correspond to the symbiont aggregations observed in several *Paramecium* cells in FISH experiments (Fig 1b).

Endosymbiont 16S rRNA gene phylogeny

Phylogeny on the 16S rRNA gene produced overall consistent results for maximum-likelihood (ML) and Bayesian inference (BI) approaches (Fig 3; Supplementary figure S2) and with the literature (e.g. Yurchenko *et al.* 2018; Castelli *et al.* 2019a; Gruber-Vodicka *et al.* 2019). The four *Rickettsiales* families were all confirmed with high support, as well as their respective relationships, in particular *Rickettsiaceae* as sister-group of all other families. Additionally, a possible sister-group relationship of *Anaplasmataceae* and “*Ca*. Deianiraeaceae” is indicated, though with low support. As expected, the novel WO2 symbiont falls within *Rickettsiaceae* with “*Ca*. Spectririckettsia” as

sister group (100 ML|1.00 BI). In turn, these symbionts form a monophyletic lineage together with members of other *Rickettsiales* genera - namely *Rickettsia*, "*Ca*. Trichorickettsia", "*Ca*. Gigarickettsia", and "*Ca*. Megaira" - associated with metazoans (Perlman *et al.* 2006) and unicellular organisms, including several ciliates (Castelli *et al.* 2016; Lanzoni *et al.* 2019). However, such grouping found only limited support, suggesting that further analyses will be necessary to fully resolve these phylogenetic relationships. Within the *Rickettsia* genus, the two expected main clades were found, namely clade 1, including bacteria associated with arthropods as symbionts (Perlman *et al.* 2006) or vector-borne pathogens (Walker and Ismail 2008), and clade 2, whose representatives are associated with arthropods as wells as aquatic organisms, e.g. leeches (Kikuchi and Fukatsu 2005; Pilgrim *et al.* 2017; previously termed "torix" group, Weinert *et al.* 2009).

Within *Rickettsiaceae* a large and highly supported (89 ML|1.00 BI) clade was identified, including all the previously mentioned lineages, as well as other quite well characterised organisms, such as the mite-borne pathogen *Orientia* (Min *et al.* 2008), and symbionts of ticks (Mediannikov *et al.* 2014) and unicellular eukaryotes (Yurchenko *et al.* 2018; George *et al.* 2020). From here on, these will be referred to as "classical *Rickettsiaceae*", while the remaining representatives of *Rickettsiaceae*, showing less resolved reciprocal phylogenetic relationships, and associated with ticks (Felsheim *et al.*, unpublished [CP009217]), amoebae (Muñoz-Gómez *et al.* 2019) or environmentally-derived (Martijn *et al.* 2015), as "basal *Rickettsiaceae*".

The highest identity found for the 16S rRNA gene of WO2 symbiont (≤90%, with "*Ca*. Spectririckettsia", Supplementary table S2), is significantly lower than the commonly accepted genus threshold 94.5% (Yarza *et al.* 2014), thus this bacterium can be considered as a representative of a novel genus; we named the new organism "*Ca*. Sarmatiella mevalonica" (from here on, *Sarmatiella mevalonica*), in reference to its geographical origin and predicted metabolism (Taxonomic description in Supplementary text S1).

Accepted Article

General genome features and phylogenomic analyses of the endosymbiont

The final draft genome assembly of *Sarmatiella* consisted of 122 contigs (1,270,819 bp; N50=16,243 bp; L50=26; GC=38.0%), estimated to contain the whole genome sequence (Supplementary text S2). A total of 1202 genes were annotated (1163 CDSs, 34 tRNAs, one tmRNA, one RNAse P, and three rRNAs), accounting for 75.6% coding in total. Several insertion elements (296 transposase sequences, plus 41 among possible passenger and accessory genes) belonging to 12 families (in particular IS5 and IS630) were found (Supplementary table S3). Such a high number of repeated elements, the majority of which (182) within 500 bp from contig ends, can explain the fragmentation level of the genome assembly, consistently with the assembly graph structure (Supplementary text S2). Five phage-related CDSs were found, but no prophages were identified.

Phylogenomic analyses confirmed the overall topology of the 16S rRNA gene phylogeny, besides providing more robust statistical supports (Fig 4; Supplementary figure S3). In detail, the four families of *Rickettsiales* and their relationship were confirmed with high support, as well as the monophyletic classical *Rickettsiaceae* (100 ML|1.00 BI), distinct from basal *Rickettsiaceae*. *Sarmatiella* forms a monophyletic clade with the *Rickettsia* genus (100 ML|1.00 BI), as other close relatives in the 16S rRNA gene phylogeny currently lack published genome sequences. The two main *Rickettsia* clades were confirmed as well. In addition, some unclear relationships in the 16S rRNA gene phylogeny appeared resolved, e.g. "*Ca*. Sneabacter" and "*Ca*. Phycorickettsia" being sister groups with high support (100 ML|1.00 BI), consistently with published phylogenomic analyses (George *et al.* 2020).

Comparative genomic analyses

Accepted Article

A total of 663 unique COGs were identified in *Sarmatiella* (Supplementary table S4). Only nine of these are unique with respect to other *Rickettsiales*, among which, notably, four COGs (five genes) involved in the mevalonate pathway, which will be treated separately in detail (section "Isoprenoid synthesis and related pathways"). The function of the other five *Sarmatiella*-specific COGs (Supplementary table S4) is unclear, as only generically predicted and/or not confirmed by direct sequence comparison, thus they were not further analysed. Otherwise, most of the predicted functional and metabolic capabilities result similar to other members of *Rickettsiales* (592 COGs present in all the families), and specifically of family *Rickettsiaceae* (641 COGs) (Fig. 5; Supplementary figure S4; Supplementary table S5; Detailed metabolic comparison in Supplementary text S3).

For the carbohydrate and energetic metabolism, *Sarmatiella* harbours a partial gluconeogenesis pathway, as other classical *Rickettsiaceae* do, as well as an almost complete non-oxidative pentose-phosphate pathway. This is similar to most *Rickettsiales*, although the closely related *Rickettsia* clade 1 and *Orientia* lack this capability (Driscoll *et al.* 2017; Min *et al.* 2008). No pyruvate dehydrogenase is present, and only a partial Krebs cycle is retained (Fig. 5). Specifically, only the oxoglutarate dehydrogenase complex and the succinyl-CoA synthetase are present, which could be fuelled by deamination products of glutamine and glutamate, as in *Rickettsia* (Driscoll *et al.* 2017). *Orientia* genomes similarly display only partial gene sets for these pathways (Min *et al.* 2008), though less pronouncedly reduced. This similarity, given the respective phylogenetic relationships (Fig. 4), is likely an evolutionary convergence. Consistently, a minimal oxidative phosphorylation apparatus is present in *Sarmatiella*, namely NADH-quinone oxidoreductase, bd-like cytochrome-ubiquinol terminal oxidase, and ATP synthase, while cytochrome c reductase and oxidase are absent, as in other *Rickettsiales* symbionts of ciliates (Floriano *et al.* 2018; Castelli *et al.* 2019a) and other protists (Yurchenko *et al.* 2018). This preferential conservation of the terminal oxidase,

contrarily to other *Rickettsiales* (Dunning-Hotopp *et al.* 2006), could represent an adaptation to hypoxic conditions (Borisov *et al.* 2011).

Predicted biosynthetic abilities of major cell membrane and cell wall components, i.e. phospholipids, peptidoglycan and LPS (lipopolysaccharide) are consistent with *Rickettsia* (Driscoll *et al.* 2017). On the other side, similarly in particular to other *Rickettsiaceae*, amino acid and nucleotide synthesis are scarce. As probable complementation for such deficiencies, four *tlc* nucleotide translocases are present, putative orthologs of the set of five in *Rickettsia* (except *tlc3*) (Andersson *et al.* 1998).

Overall, the genome presents cofactor biosynthetic pathways similar to *Rickettsia* and other *Rickettsiales*, including those for lipoate, ubiquinone, folate, NADP (from NAD) and iron-sulfur clusters. In addition, *Sarmatiella* can synthesise NAD, a trait shared only with *Occidentia* (Mediannikov *et al.* 2014), while some members of other *Rickettsiales* families have only partial pathways. As most classical *Rickettsiaceae* and contrarily to basal *Rickettsiaceae* and other families (Dunning-Hotopp *et al.* 2006; Sassera *et al.* 2011), *Sarmatiella* is incapable of biotin synthesis, but a putative truncated biotin synthase pseudogene is present, possibly indicating a recent loss of the pathway.

Finally, DNA repair capabilities are overall consistent with *Rickettsiales*, though, distinctively, for mismatch repair only *mutS* is present, possibly indicating an impaired mechanism. Interestingly, mismatch repair can work as an inhibitor of interspecies recombination (Matic *et al.* 1995), which may indicate that the *Sarmatiella* genome could be particularly permissive for horizontal transfer and recombination events, such as those hypothesised below.

Isoprenoid synthesis and related pathways

A novel distinctive feature of *Sarmatiella* among *Rickettsiales* are five mevalonate pathway (MEV)

Accepted Article

genes (Supplementary table S6). This pathway catalyses the conversion of acetoacetyl-CoA into isopentenyl-diphosphate (IPP), which can be converted into its isomer dimethylallyl-diphosphate (DMAPP) by the isopentenyl-diphosphate isomerase, a step shared also with other pathways (Fig. 5). Both isomers are precursors of isoprenoids, e.g. the lipid anchor of quinones and peptidoglycan. MEV is typical in eukaryotes and *Archaea*, common in Gram-positive bacteria, but very rare in Gram-negatives, in particular *Proteobacteria*, which display the alternative MEP/DOXP pathway, capable of producing both IPP and DMAPP (Lombard and Moreira 2011; Hoshino and Gaucher 2018). Thus, more thorough analyses were conducted on isoprenoid biosynthetic pathways among *Rickettsiales*. A substantially complete MEP/DOXP pathway was found in all *Anaplasmataceae,* "*Ca*. Midichloriaceae", basal *Rickettsiaceae*, and in "*Ca*. Deianiraea vastatrix". All these organisms lack *idi* gene (encoding isopentenyl-diphosphate isomerase), dispensable since the MEP/DOXP pathway can produce both IPP and DMAPP. Phylogenetic results are consistent with vertical inheritance of this pathway through the Proto-Rickettsiales from an alphaproteobacterial ancestry (Supplementary figure S5).

The only gene related to isoprenoid precursors found among other classical *Rickettsiaceae* is the *idi*, specifically *idi1* (for enzyme isoform 1), as in *Sarmatiella*, in *Occidentia massiliensis* and *Rickettsia* clade 2 (Mediannikov *et al.* 2014; Pilgrim *et al.* 2017; Wang *et al.* 2020), and *idi2* (for enzyme isoform 2) in *Rickettsia* clade 1 (Driscoll *et al.* 2017). Thus, all the five MEV-specific genes were confirmed as unique of *Sarmatiella* among all *Rickettsiales* organisms. However, extending the search to *Rickettsiales* (metagenome-assembled genomes) (MAGs), we found that two of them, phylogenetically affiliated to "*Candidatus* Deianiraeaceae" (Fig 4), present the complete gene set for the MEV pathway including *idi1*, and a third one, belonging to basal *Rickettsiaceae* (Fig. 4), possesses, in addition to the complete MEP/DOXP pathway, a single MEV gene, i.e. hydromethylglutaryl-CoA reductase (*hmgr*). While *Sarmatiella* has isoform 2 (*hmgr2*), the three MAGs have isoform 1 (*hmgr1*). MEV genes in *Sarmatiella* are not grouped in operons, while a

Accepted Article

portion of them (three) form a putative operon in the two "*Candidatus* Deianiraeaceae" MAGs (Supplementary table S6). In single MEV gene phylogenies (Supplementary figure S6), among *Rickettsiales* only genes of very closely related organisms formed monophyletic groups, i.e. the six genes of two "*Candidatus* Deianiraeaceae", and, respectively, the *idi1* or *idi2* of each of the two *Rickettsia* clades. All general tree topologies were poorly supported, preventing to accurately infer the evolutionary origin of each gene found in *Rickettsiales.* Compositional analyses of MEV genes showed consistence with the respective genome assemblies, showing no evidence of recent HGT (Supplementary table S6).

Interestingly, the metabolism of PHA granules, present also in *Rickettsia*, "*Candidatus* Phycorickettsia", basal *Rickettsiaceae*, and possibly *"Candidatus* Jidaibacter" (Driscoll *et al.* 2017; Yurchenko *et al.* 2018; Schulz *et al.* 2016; Muñoz-Gómez *et al.* 2019), shares an intermediate with the MEV pathway. PHA granules represent a carbon storage form, synthesised from acetyl-CoA through an acetoacetyl-CoA intermediate, which is indeed also the precursor of the MEV pathway, possibly representing a link between the two pathways. PHA granules are disassembled into acetoacetate, which can be then reconverted into acetyl-CoA.

Protein secretion and host interaction apparatuses

The available gene set putatively involved in protein secretion and interaction with the host of *Sarmatiella* is similar to *Rickettsiaceae* and *Rickettsiales* in general (Gillespie *et al.* 2015). In particular, the novel genome is equipped with Sec and Tat translocons, as well as with type I secretion system, and six OmpB/Sca-like type V secretion autotransporter beta-barrel proteins (Fig. 5). The typical *Rickettsiales* type IV system is present with almost the same components as in *Rickettsia*, namely *virB2*, *virB3*, virB4 (two copies), *virB6* (five copies), *virB8* (two copies), *virB9* (two copies), *virB10*, *virB11*, and *virD4* (Gillespie et al. 2016). However, no *virB7* gene was

identified, possibly due to high sequence variations (Gillespie *et al.* 2009), as observed in other fast-evolving *Rickettsiales* (Castelli *et al.* 2019a). In addition, a second *virD4* copy was found, shorter on its C-terminal. Interestingly, its best blastp hits on NCBI repositories are similarly short VirD4 isoforms in the basal *Rickettsiaceae* bacterium "*Candidatus* Arcanobacter lacustris" (Martijn *et al.* 2015) (having an additional full-length gene) and in some *Rickettsiales* MAGs, possibly indicating that these genes result from the same ancestral duplication. As hypothesized for other shorter isoforms of rickettsial type IV secretion apparatus (Gillespie *et al.* 2015), it may display regulatory functions.

Although no flagella were observed in microscopy, several genes encoding for components of the flagellar apparatus were found. In detail, components of the basal body, rotor, stator, and the flagellar secretion system are present, but no key components of the motile organelle, i.e. flagellar filament (fliC), cap (fliD), or associated proteins were found (e.g. flgK, flgL, flgM). This condition reminds that of "*Ca.* Sneabacter namystus", a classical *Rickettsiaceae* symbiont with highly reduced genome, for which, given the underlying homologies (Blocker *et al.* 2003), a role of the apparatus as solely type III secretion system was hypothesised (George *et al.* 2020). The main difference between the two organisms is that *Sarmatiella* also presents the flagellar hook (flgE).

In general, the *Sarmatiella* genome encodes for several putative secreted effectors. These include a set of 40 signal peptide predicted proteins (Supplementary table S7), as well as 16 tetratricopeptide-repeat containing proteins, representing potential interactors, and other proteins possibly involved in vacuolar escape, i.e. two hemolysin-like proteins, and a patatin-like protein (Walker and Ismail 2008).

Accepted Article

Discussion

In this work, we characterised and comparatively analysed *Sarmatiella mevalonica,* a novel *Rickettsiales* intracellular symbiont of the ciliate *P. tredecaurelia*, showing distinctive genomic and predicted functional features. Consistently with other *Rickettsiales*, its predicted metabolism indicates an obligate association with the host. Thus, in the absence of isolation in pure culture, it meets the requirements for the provisional *Candidatus* taxonomic status (Murray and Stackebrandt 1995; Oren *et al.* 2020) (see Supplementary text S1 for taxonomic description).

Many features of *Sarmatiella*, both in its morphology and its genome content, are consistent with *Rickettsiales* (e.g. Sassera *et al.* 2011; Dunning-Hotopp *et al.* 2006; Castelli *et al.* 2019a; Min *et al.* 2007), and, specifically, highly comparable to *Rickettsia* (Silverman *et al.* 1978; Driscoll *et al.* 2017; Gillespie *et al.* 2015). These include a prevalent cytoplasmic location, the absence of encircling host vacuoles, the presence of a surrounding electron-lucid halo, highly reduced biosynthetic pathways (especially for nucleotides and amino acids) coupled with many transporters to scavenge precursors from the host (including four *tlc* nucleotide transporters), and multiple secretion systems (including rickettsial type IV and OmpB/Sca-like autotransporter type V), involved in putative release of effectors to the host.

Available data is too limited to reasonably infer the nature of the association, also considering that, while this is the first report of *P. tredecaurelia* harbouring symbionts, very few reports of this ciliate species exist at all (Przyboś *et al.* 2013). We could hypothesise that, considering the absence of clearly host-supportive traits in the genome, *Sarmatiella* could possibly be an intracellular parasite, as most *Rickettsiales*. The observation of some *Sarmatiella* cells inside putative host digestive

Accepted Article

vacuoles could support this hypothesis, suggesting some kind of autophagic defence mechanism by the host (Szokoli *et al.* 2016b). However, ciliates may employ autophagy to get energy supply even regardless of the presence of intracellular bacteria (Görtz and Fokin 2009). In any case, the relationship between *Sarmatiella* and its host might be more complex, similarly to what has been shown for some other bacterial symbionts of ciliates, which, besides clearly parasitic behaviours, also have beneficial or even protective effects under certain circumstances (Bella *et al.* 2016; Schu and Schrallhammer 2018; Schrallhammer and Potekhin 2020), including environmental stress (Duncan *et al.* 2010).

Interestingly, the *Sarmatiella* genome presents significant peculiarities with respect to other *Rickettsiales*, in particular a distinctive gene set encoding for the flagellar apparatus and the unique presence of the MEV genes.

The presence of flagellar genes in *Sarmatiella* is by itself worthy of attention, as recently they were found in many phylogenetically unrelated *Rickettsiales* (e.g. Sassera *et al.* 2011; Schulz *et al.* 2016; Martijn *et al.* 2015; Klinges *et al.* 2019; George *et al.* 2020). Although flagellar-driven motility was observed at least once (Vannini *et al.* 2014), the actual function of these organelles in the typically intracellular *Rickettsiales* is unclear, also considering that, in many representatives, including *Sarmatiella*, no flagella were actually observed. For this reason, given the underlying homologies with type III secretion systems (Blocker *et al.* 2003). an alternative/additional role in delivering effectors to the host was proposed (Sassera *et al.* 2011; George *et al.* 2020). In line with the latter hypothesis, while maintaining most core flagellar components, including the secretion apparatus, *Sarmatiella* lacks the extracellular ones that build up the functional motile organelle. Thus, the observed motility of *Sarmatiella* is unlikely to be driven by flagella, and further analyses will be necessary to clarify the underlying mechanism. Interestingly, the *Sarmatiella* set of flagellar genes is reminiscent of another *Rickettsiaceae* bacterium, "*Ca.* Sneabacter namystus" (George *et al.*

2020). However, considering the closer phylogenetic relationship of *Sarmatiella* with the flagellated and motile "*Ca*. Trichorickettsia" and "*Ca*. Gigarickettsia" (Vannini *et al.* 2014; Sabaneyeva *et al.* 2018), this could be the result of convergent evolution.

The complete MEV gene repertoire is the most distinctive feature found in *Sarmatiella*. Indeed, this pathway, involved in the synthesis of the precursors of isoprenoids and common in eukaryotes and *Archaea*, is very rare in most bacterial lineages, including *Proteobacteria* (Hoshino and Gaucher 2018; Lombard and Moreira 2011), and, to our knowledge, has never been described in *Rickettsiales* before. Nevertheless, the multi-step and carefully revised selective assembly procedure (Supplementary text S2), the gene phylogenies (Supplementary figure S6), the position in the assembled contigs, and the sequence composition (Supplementary table S6) allowed to confidently exclude that these genes originated from misassembly/assembly contaminations by the host (or other bacteria), allowing to reliably assign them to *Sarmatiella*. Basal *Rickettsiaceae* and members of the other three *Rickettsiales* families present the typical bacterial alternative MEP/DOXP pathway, while classical *Rickettsiaceae* lack both MEV and MEP/DOXP pathway. This common presence of MEP/DOXP genes, together with phylogenetic analyses, clearly indicates their vertical inheritance, and suggests that their absence in classical *Rickettsiaceae* is likely due to a lineage-specific loss. Accordingly, classical *Rickettsiaceae* are host-dependent for IPP and DMAPP (Ahyong *et al.* 2019; Driscoll *et al.* 2017), though some of them are able to interconvert these two compounds with isopentenyl-diphosphate isomerase, absent in other *Rickettsiales* lineages, thus being a HGT candidate (Driscoll *et al.* 2017). Besides *Sarmatiella*, our analyses identified a complete MEV gene set also in two *Rickettsiales* MAGs, affiliated to "*Ca*. Deianiraeaceae", and a single gene (*hmgr1*) in another basal *Rickettsiaceae* MAG (in addition to the MEP/DOXP). Phylogenetic reconstructions of MEV and *idi* genes of *Sarmatiella,* MAGs and classical *Rickettsiaceae* are poorly supported, consistently with fast sequence evolution and consequent high divergences, typical in obligatorily host-associated bacteria. Thus, similarly to other rare or unique

Accepted Article

genes in *Rickettsiales* (Sassera *et al.* 2011; Castelli *et al.* 2019a), a clear reconstruction of the evolutionary history of MEV genes is prevented. In particular it is not possible to unambiguously identify potential HGT events and putative donors. Nevertheless, especially taking into account the extreme rarity of this pathway in all *Proteobacteria*, the most parsimonious and reasonable reconstruction suggests that they were acquired through HGT event(s) from unknown sources, most likely rather anciently, considering low sequence identities with other organisms and compositional consistency with the respective genomes (Supplementary table S6). Given the not so robust phylogenetic signal and possibility of successive secondary losses, the precise timing and number of distinct acquisitions is unclear. Considering the lack of evidence, with few exceptions, of close relationship for the *Rickettsiales* MEV genes sequences (Supplementary figure S6), as well as the presence of different gene isoforms (e.g. *hmgr2* in *Sarmatiella* and *hmgr1* in the three MAGs, *idi1* and *idi2* in each distinct *Rickettsia* clade), it seems likely that there could have been multiple distinct HGT events, tentatively up to six (Fig. 4).

The functional significance of MEV in *Sarmatiella* and other *Rickettsiales* should be pondered. As the pathway is typical in eukaryotes, it could offer the chance of a host-symbiont exchange of pathway intermediates, engaging a deeper metabolic interplay. An obvious consequence could be a more efficient production of isoprenoids by the bacterium. Considering the overall predicted metabolic capabilities of *Sarmatiella*, another intriguing and non-mutually exclusive hypothesis could be delineated. MEV enzymes can also catalyse the reverse reactions, thereby enabling some bacteria to grow on mevalonate as sole carbon source (Siddiqi *et al.* 1967; Takatsuji *et al.* 1982), allowing, in conjunction with acetoacetyl-CoA transferase, the production of acetyl-CoA. *Sarmatiella* could be able to perform such reverse reactions, which, moreover, would be fundamental, as the sole putative route to acetyl-CoA production (Fig. 5), considering that the canonical path by pyruvate dehydrogenase is absent. Acetyl-CoA could be then used for biosynthesis of (phospho)lipids, while the *Sarmatiella* Krebs cycle is partial and does not require

Accepted Article

this compound. Produced acetyl-CoA could also be reversibly stored by the bacterium into PHA granules. In addition, reversed MEV reactions could provide reduced NAD(P)H, either representing an additional energy source or being directly involved in biosynthetic pathways (detailed in Supplementary text S3). Thus, the probable acquisition of MEV genes could have equipped *Sarmatiella* with a novel alternative way to efficiently exploit host metabolites, and, potentially, could even have been the "premise" for specific losses of otherwise "basic" pathways (pyruvate dehydrogenase and part of Krebs cycle) that would have become no more strictly required. Alternatively, *Sarmatiella* might be able to import acetyl-CoA directly from its host, just as *Orientia* (Min *et al.* 2008). In any case, since currently no transporter was identified for any of the hypothesised host-symbiont metabolite interchanges, it is not yet possible to clearly discern among the proposed hypotheses (or additional ones).

As for the MEV in the *Rickettsiales* MAGs, its role is more elusive, considering the inherently lower quality information available on the respective organisms. Nevertheless, assuming a host-association and the considerations above on the evolutionary history of those genes, it is reasonable to hypothesise a HGT-driven evolutionary convergent scenario, in which this pathway may enable novel advantageous metabolic interactions, possibly resulting in the production of acetyl-CoA from host-derived MEV intermediates. From a more general perspective, we must consider that among *Proteobacteria* MEV genes are rare and likely originated from HGT events as well (Lombard and Moreira 2011; Hoshino and Gaucher 2018), and that, at the same time, they are found in several eukaryote-associated bacteria (e.g. *Legionella*, *Coxiella*, "*Ca.* Liberibacter") (Gottlieb *et al.* 2015; Gomez-Valero *et al.* 2011; Lin *et al.* 2011). This allows to speculate that similar advantages could hold at least for some of these organisms.

A critical interpretation of the case of MEV genes from the perspective of *Rickettsiales* evolution and diversification offers additional intriguing insights. The numerous lineage-specific differences among *Rickettsiales* are themselves indicative of evolutionary plasticity. However, it is often hard to

Accepted Article

discern how these actually originated, i.e. identify single HGT, gene duplication, loss or rearrangement events, considering the limited phylogenetic signal (due to fast evolution) and the phylogenetically unbalanced set of genomes available (Castelli *et al.* 2019a). In this regard, MEV genes represent a precious example, because, contrarily to other cases, their rarity and patchy phylogenetic distribution all along *Proteobacteria* allow to confidently indicate them as HGT-acquired. This constitutes a reliable basis to convincingly confirm the contribution of HGT events in evolutionary plasticity and diversification of *Rickettsiales*. Thus, even in such a lineage of strictly host-associated bacteria, with relatively limited functional and metabolic background, an efficient and stable metabolic integration of horizontally-acquired features with pre-existing recipient features can occur. Moreover, along the same line of reasoning, multiple probable HGT events of MEV genes were evidenced among *Rickettsiales*, indicating an evolutionary convergence. Accordingly, the plasticity of *Rickettsiales*, together with their common ancestral background derived from the Proto-Rickettsiales, may be seen as permissive for convergent adaptations, and, possibly, even favouring them in the context of comparable lifestyles and hosts. It is worth to underline that evolutionary convergence in *Rickettsiales* was observed also as the result of gene losses, such as for part or the whole set of flagellar genes (e.g. present work, George *et al.* 2020; Floriano *et al.* 2018; Yurchenko *et al.* 2018). Under this perspective, the evolution of many other traits in *Rickettsiales* could have been more complex than expected, potentially following more evolutionary convergence paths than currently recognised, still nuanced or hidden by the high evolutionary and sequence distances.

Overall, many different features of *Rickettsiales* and their evolution still need to be elucidated, with the scenario getting more complex the more data are collected on previously “neglected” lineages. We predict that deep genomic analyses from such novel and/or neglected lineages, as in case of *Sarmatiella* and several recent others (Sassera *et al.* 2011; Mediannikov *et al.* 2014; Pilgrim *et al.* 2017), in particular those associated with aquatic hosts (George *et al.* 2020; Yurchenko *et al.* 2018;

Klinges *et al.* 2019; Gruber-Vodicka *et al.* 2019; Castelli *et al.* 2019a; Floriano *et al.* 2018; Schulz *et al.* 2016), will offer a wider comparative perspective, for a better reconstruction of the evolutionary paths of *Rickettsiales* and the underlying genomic processes. Additionally, a better understanding of their interaction with the hosts, including through transcriptome analyses (see as examples Yang *et al.* 2016; Gruber-Vodicka *et al.* 2019; Pirritano *et al.* 2020; Midha *et al.* 2020), may significantly contribute as well, providing validated functional grounds for comparative and evolutionary inferences.

Accepted Article

Accepted Article

Experimental procedures

Host isolation, cultivation, and *in vivo* observations

Paramecium tredecaurelia strain WO2 was isolated from a sample collected from an effluent of post-treated wastewater in Orenburg, Russia (51.766236, 55.036390) in September 2015. It was maintained at 18°C inside a Sanyo climatic chamber (Osaka, Japan), and fed once in a fortnight with lettuce medium inoculated with *Enterobacter aerogenes*.

All live observations were performed using Differential Interference Contrast (DIC) with a Leica 6000 microscope (Leica Microsystems, Wetzlar, Germany) equipped with a digital camera DFC 500. Prior to observations, live cells were immobilised with the help of a mechanical microcompressor (Yan *et al.* 2014). The host was preliminarily identified as a species representing the *P. aurelia* complex according to morphological features.

Molecular characterisation of host and symbiont

Total genomic DNA was extracted using NucleoSpin Plant II (Macherey-Nagel, Germany) kit from approximately 100 *Paramecium* cells, as described previously (Lanzoni *et al.* 2016). All following PCR reactions were carried out with ExTaq polymerase and reagents (Takara, Otsu, Japan). Host was characterised by amplification and direct sequencing of 18S rRNA gene, cytochrome oxidase I (COI) gene, and complete ITS1-5.8S-ITS2, as previously described (Lanzoni *et al.* 2016).

The 16S rRNA gene of the endosymbiont was amplified as previously described (Castelli et al 2019a), and the product was directly sequenced using internal primers 16S R515ND, 16S F785ND (Vannini *et al.* 2004), and Sarmat_F343 (5'-GTTAGGAAGCAGCAGTG-3'), manuallydesigned on purpose exploiting the dedicated ARB package 6.0.6 function.

Accepted Article

16S rRNA gene phylogeny of the endosymbiont

For the phylogeny on the 16S rRNA gene of the symbiont, its sequence was aligned with ARB 6.0.6 (Westram *et al.* 2011) together with other 35 *Rickettsiales*, and six other *Alphaproteobacteria* as outgroup. The manually revised alignment was trimmed at both ends to the length of the shortest sequence, and, to remove hypervariable positions, only those where the most conserved base was present in at least 10% sequences were kept (1,385 total positions). Best substitution model was predicted with jModelTest 2.1.4 (Darriba *et al.* 2012) according to the Akaike information criterion, and ML and BI phylogenies were inferred as previously described (Lanzoni *et al.* 2019).

Fluorescence *in situ* hybridisation

FISH experiments were conducted following the protocol by Manz *et al.* (1992). A preliminary screening was carried out with the almost universal bacterial probe EUB338 (Amann *et al.* 1990) and the *Alphaproteobacteria*-targeted probe ALF1b (Manz *et al.* 1992). As EUB338 did not show any signal, a species-specific variant was designed using the newly obtained 16S rRNA gene sequence from the WO2 sample. The specificity of this and of an additional probe designed on a different region of the gene was tested *in silico* both on Ribosomal Database Project (RDP, Cole *et al.* 2013: rdp.cme.msu.edu: accessed in May 2020) and on SILVA rRNA database release 138 (Quast *et al.* 2013) (), allowing 0 and 1 mismatches on each database (Supplementary table S8). The respective probe sequences were deposited on Probebase (Greuter *et al.* 2016). The formamide concentration was experimentally tested at (0%, 15%, and 30% w/v), and selected as optimal at 0%. All samples were observed under a Leica TCS SP5 confocal microscope.

Ultrastructural characterisation

Ciliates were prepared for observation under Transmission Electron Microscope (TEM) as described in Szokoli *et al.* (2016a). Briefly, cells were fixed in 2.5% glutaraldehyde and 1.6%

paraformaldehyde in phosphate buffer (0.1 M, pH 7.4), with a post-fixation in 1.5% OsO_4, then dehydrated at increasing percentages of ethanol solutions, and finally embedded in epoxy embedding medium (Fluka, BioChemika). Ultrathin sections were stained with uranyl acetate followed by lead citrate. Samples were observed using a JEM-1400 (JEOL Ltd., Tokyo, Japan) electron microscope.

Symbiont genome sequencing and assembly

In order to get DNA with minimal proportion of contaminants in a sufficient amount for genome sequencing, a whole genome amplification (WGA) was performed using QIAGEN Repli-G single-cell kit from around five *Paramecium* cells as previously described (Serra *et al.* 2020), and sequenced with Illumina HiSeq X by Admera Health (South Plainfield, NJ, USA), obtaining 44,434,254 2x150 bp read pairs. After a preliminary assembly of total reads with SPAdes 3.6.0 (Bankevich *et al.* 2012), the contigs belonging to the endosymbiont were selected using the blobology pipeline (https://github.com/blaxterlab/blobology; accessed January 2021) (Kumar *et al.* 2013), as described previously (e.g. Castelli *et al.* 2019a; Floriano *et al.* 2018; Detailed procedure in Supplementary text S2; Supplementary table S9). Briefly, contigs with log_{10} coverage higher than 1.2 were selected, then, within this selection, those having a best megablast hit on Peniculida (which include *Paramecium*) were removed. After an accurate manual revision of this set, reads mapping to these contigs with Botwie2 2.2.6 (Langmead and Salzberg 2012) were separately reassembled. The assembly output was extensively manually examined and revised, in order to get the final genome assembly of the symbiont. Assembly completeness was further evaluated in comparison with other *Rickettsiales* with BUSCO 3 on a set of 221 conserved orthologs in *Proteobacteria* (Simão *et al.* 2015).

Genomic analyses

Accepted Article

The protein-coding and non-coding RNA genes in the WO2 symbiont genome were annotated using Prokka 1.10 (Seeman 2014), followed by manual curation by detailed inspection of blastp hits on NCBI nr and on *Rickettsiales* proteins. Putative secreted proteins were predicted with SignalP 5.0 (Almagro Armenteros *et al.* 2019). Insertion sequences were predicted with IS finder (Siguier *et al.* 2006), and prophages with and PHASTER (Arndt *et al.* 2016) with the respective web interfaces (https://phaster.ca/; https://isfinder.biotoul.fr/; accessed in May 2020). The obtained results were compared with the curated annotation.

Phylogenomic analyses

Phylogenomic analyses were performed on a previously determined curated set of 120 highly conserved orthologous proteins (Parks *et al.* 2018), adding the novel genome to a dataset of other 28 representative *Rickettsiales,* including four MAGs, plus six other *Alphaproteobacteria* as outgroup. Ortholog sequences in the symbiont and in other organisms absent from the original dataset (Parks *et al.* 2018) were identified with GTDB-tk 1.1.1 (Chaumeil *et al.* 2020), and, as previously described (Castelli *et al.* 2019a), added to the original alignment with MAFFT 7.271 (Nakamura *et al.* 2018), keeping the originally aligned positions 34,747 positions selected by Parks and coauthors (2018) (--add --keeplength options) . Thus, the best substitution model was selected with ProtTest 3.4.2 (Darriba *et al.* 2011), and ML and BI phylogenies were inferred, respectively, with RaxML 8.2.4 (Stamatakis 2015) with 1000 pseudo-replicates, and with MrBayes 3.2.6 (Ronquist *et al.* 2012) with three runs each with one cold and three heated chains, iterating until convergence (i.e. average standard deviation of split frequencies below 0.01) after 500,000 generations, applying 25% burn-in.

Metabolic prediction and comparison

Reconstruction of the predicted metabolic pathways of the symbiont was performed employing

BioCyc (Karp *et al.* 2019) and KEGG (Kanehisa *et al.* 2019) references (https://biocyc.org/: https://www.genome.jp/kegg/; accessed in July 2020), followed by manual inspection. Cluster of Orthology Groups (COGs) were identified using the NCBI pipeline on the 2014 release (ftp://ftp.ncbi.nih.gov/pub/COG/COG2014/data: Galperin *et al.* 2015). The COG repertoire was directly compared with 24 other *Rickettsiales sensu stricto*, in order to identify its peculiarities.

Comparative analyses on isoprenoid synthesis genes

All available *Rickettsiales* genome assemblies and MAGs (May 2020) were downloaded from NCBI. The GenBank annotation was used, or, when not available, a novel annotated with Prokka. The results of blastp queries of genes of the mevalonate pathway (MEV) from the WO2 symbiont and reference organisms (Hoshino and Gaucher 2018) on all *Rickettsiales* protein sequences were manually inspected. In the corresponding selected assemblies, rRNA genes were identified with barrnap 0.7 (Seeman 2013), and results were queried on NCBI nucleotide (release 237), to filter out chimeric assemblies that included sequences of non-*Rickettsiales* organisms. The remaining assemblies were analysed with BUSCO as described above, and only those with >50% conserved orthologs were retained (Supplementary table S10).

All identified MEV genes - only a representative selection for isopentenyl-diphosphate isomerase (idi), common in *Rickettsia* - were added to the single-gene alignments by Hoshino and Gaucher (2018) with MAFFT (--add --keeplength options). Statistical comparisons were performed on the codon adaptation index (CAI) of the MEV genes of the symbiont and the selected *Rickettsiales* MAGs with respect to other CDSs of the respective assemblies with CAIcal 1.3 (Puigbo *et al.* 2008).

For the phylogeny of the MEP/DOXP (2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate) pathway, the seven reference *Escherichia coli* gene sequences were downloaded from BioCyc, and orthologs were identified by best blastp hits on selected *Rickettsiales* and

Accepted Article

representative members of main proteobacterial lineages, as well as on *Verrucomicrobia* and *Flavobacteria* as outgroups (72 total organisms). Each ortholog was aligned with Muscle 3.8.31 (Edgar 2004), polished with Gblocks 0.91b (Talavera and Castresana 2007), and all genes were concatenated together with an in-house perl script (1091 amino acid positions).

For each alignment, after selection of the best substitution model with ProtTest, ML phylogeny was inferred with RAxML with 100 bootstrap pseudo-replicates.

Data Availability

Sequence data have been deposited in GenBank with the listed accession numbers. *P. tredecaurelia* WO2 partia1 8S-ITS1-5.8S-ITS2-28S: MT980797, *P. tredecaurelia* WO2 partial cytochrome oxidase subunit I gene: MT989369, "*Ca*. Sarmatiella mevalonica" WO2 partial 16S rRNA gene: MT984300; "*Ca*. Sarmatiella mevalonica" WO2 genome: JACVVM000000000; WO2 Illumina shotgun sequencing reads: PRJNA662343.

Acknowledgements

This work was funded by the University of Pisa PRA_2018_63 project to GP; the European Community's H2020 Programme H2020-MSCA-RISE 2019 grant 872767 to GP; the Italian Ministry of Education, University and Research (MIUR): Dipartimenti di Eccellenza Programme (2018–2022)—Department of Biology and Biotechnology "L. Spallanzani", University of Pavia to DS; fluorescent and confocal microscopy observations were made possible with support of RSF grant 20-14-00220 to AP.. We would like to thank A. Oren for advice in bacterial nomenclature, E. Gaucher and Y. Hoshino for providing alignments for mevalonate gene phylogenies, and S.S. Liu and L. Teng for providing genome sequence of *Rickettsia* endosymbiont of *Bemisia tabaci*. Confocal microscopy was performed at the "Chromas" Core Facility Center, Saint Petersburg State University, Russia. The WO2 culture is available upon a request from the RC CCM "Culture Collection of Microorganisms", Saint Petersburg State University, Russia.

Accepted Article

References

- Ahyong, V., Berdan, C.A., Burke, T.P., Nomura, D.K., Welch, M.D. (2019) A metabolic dependency for host isoprenoids in the obligate intracellular pathogen *Rickettsia parkeri* underlies a sensitivity to the statin class of host-targeted therapeutics. MSphere 4: e00536-19.
- Andersson, S.G.E., Zomorodipour, A., Andersson, J.O., Sicheritz-Pontén, T., Alsmark, U.C.M., Podowski, R.M., *et al.* (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396: 133–43.
- Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Nordahl Petersen, T., Winther, O., *et al.* (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37: 420–423.
- Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R., Stahl, D.A. (1990) Combination of 16S ribosomal-RNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl Environ Microbiol 56: 1919–25.
- Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S. (2016) PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44: W16-W21.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., *et al.* (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comp Biol 19: 455–77.
- Beckmann, J.F., Bonneau, M., Chen, H., Hochstrasser, M., Poinsot, D., Merçot, H., *et al.* (2019) The toxin-antidote model of cytoplasmic incompatibility: genetics and evolutionary implications. Trends Genet 35: 175-185.
- Bella, C., Koehler, L., Grosser, K., Berendonk, T.U., Petroni, G., Schrallhammer, M. (2016) Fitness impact of obligate intranuclear bacterial symbionts depends on host growth phase.

Accepted Article

Accepted Article

Front Microbiol 7: 2084-

- Blocker, A., Komoriya, K., Aizawa, S.I. (2003) Type III secretion systems and bacterial flagella: insights into their function from structural similarities. Proc Nat Ac Sci 100: 3027-3030.
- Borisov, V.B., Gennis, R.B., Hemp, J., Verkhovsky, M.I. (2011) The cytochrome bd respiratory oxygen reductases. Biochimica et Biophysica Acta (BBA) – Bioenergetics 1807: 1398-1413.
- Braig, H.R., Guzman, H., Tesh, R.B., O'Neill, S.L. (1994) Replacement of the natural *Wolbachia* symbiont of *Drosophila simulans* with a mosquito counterpart. Nature 367: 453–5.
- Castelli, M., Sassera, D., Petroni, G. (2016) Biodiversity of "non-model" *Rickettsiales* and their association with aquatic organisms. In *Rickettsiales*—biology, molecular biology, epidemiology, and vaccine development. Thomas, S. (ed) Cham, Switzerland: Springer International Publishing, pp. 59–91.
- Castelli, M., Sabaneyeva, E., Lanzoni, O., Lebedeva, N., Floriano, A.M., Gaiarsa, S., *et al.* (2019a) *Deianiraea*, an extracellular bacterium associated with the ciliate *Paramecium*, suggests an alternative scenario for the evolution of *Rickettsiales*. ISME J 13: 2280-2294.
- Castelli, M., Serra, V., Senra, M.V.X., Basuri, C.K., Soares, C.A.G., Fokin, S.I., *et al.* (2019b) The hidden world of *Rickettsiales* symbionts: "*Candidatus* Spectririckettsia obscura," a novel bacterium found in Brazilian and Indian *Paramecium caudatum*. Microb Ecol 77: 748-758.
- Chaumeil, P.A., Mussig, A.J., Hugenholtz, P., Parks, D.H. (2020) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36: 1925–1927.
- Chee, J.Y., Yoga, S.S., Lau, N.S., Ling, S.C., Abed, R.M.M., Sudesh, K. (2010) Bacterially produced polyhydroxyalkanoate (PHA): converting renewable resources into bioplastic. In

Accepted Article

Current research, technology and education topics in applied microbiology and microbial biotechnology. Méndez-Vilas, A. (ed). Badajoz, Spain: Formatex Research Center, pp. 1395–1404.

- Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., *et al.* (2014) Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res 42: D633–D642.
- Darriba, D., Taboada, G.L., Doallo, R., Posada, D. (2011) ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27: 1164–5.
- Darriba, D., Taboada, G.L., Doallo, R, Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9: 772.
- Driscoll, T.P., Verhoeve, V.I., Guillotte, M.L., Lehman, S.S., Rennoll, S.A., Beier-Sexton, M., *et al.* (2017) Wholly *Rickettsia*! Reconstructed metabolic profile of the quintessential bacterial parasite of eukaryotic cells. Mbio 8: e00859–17.
- Duncan, A.B., Fellous, S., Accot, R., Alart, M., Sobandi, K.C., Cosiaux, A., Kaltz, O., (2010) Parasite-mediated protection against osmotic stress for *Paramecium caudatum* infected by *Holospora undulata* is host genotype specific. FEMS Microbiol. Ecol 74: 353–360.
- Dunning Hotopp, J.C., Lin, M., Madupu, R., Crabtree, J., Angiuoli, S.V., Eisen, J., *et al.* (2006) Comparative genomics of emerging human ehrlichiosis agents. PLoS Genet 2: e21.
- Duplouy, A., Iturbe-Ormaetxe, I., Beatson, S.A., Szubert, J.M., Brownlie, J.C., McMeniman, C.J., *et al.* (2013) Draft genome sequence of the male-killing *Wolbachia* strain wBol1 reveals recent horizontal gene transfers from diverse sources. BMC Genomics 14: 20.
- Duron, O., Doublet, P., Vavre, F., Bouchon, D. (2018) The importance of revisiting *Legionellales* diversity. Trends Parasitol 34: 1027-1037.
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high

Accepted Article

throughput. Nucleic Acids Res 32: 1792–1797.

- El Karkouri, K., Pontarotti, P., Raoult, D., Fournier, P.E. (2016) Origin and evolution of rickettsial plasmids. PLoS One 11: e0147492.
- Epis, S., Sassera, D., Beninati, T., Lo, N., Beati, L., Piesman, J., *et al.* (2018) *Midichloria mitochondrii* is widespread in hard ticks (Ixodidae) and resides in the mitochondria of phylogenetically diverse species. Parasitology 135: 485–94.
- Floriano, A.M., Castelli, M., Krenek, S., Berendonk, T.U., Bazzocchi, C., Petroni, G., Sassera D. (2018) The genome sequence of "*Candidatus* Fokinia solitaria": insights on reductive evolution in *Rickettsiales*. Genome Biol Evol 10: 1120–6.
- Görtz, H.D., Fokin, S.I. (2009) Diversity of endosymbiotic bacteria in *Paramecium*. In Endosymbionts in *Paramecium*. Fujishima, M. (ed). Berlin-Heidelberg, Germany: Springer, pp. 131-160.
- Galperin, M.Y., Makarova, K.S., Wolf, Y.I., Koonin, E.V. (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res 43: D261–D269.
- George, E.E., Husnik, F., Tashyreva, D., Prokopchuk, G., Horák, A., Kwong, W.K., *et al.* (2020) Highly reduced genomes of protist endosymbionts show evolutionary convergence. Curr Biol 30: 925-933.e3.
- Gillespie, J.J., Ammerman, N.C., Dreher-Lesnick, S.M., Rahman, M.S., Worley, M.J., Setubal, J.C., *et al.* (2009) An anomalous type IV secretion system in *Rickettsia* is evolutionarily conserved. PLoS ONE 4: e4833.
- Gillespie, J.J., Joardar, V., Williams, K.P., Driscoll, T., Hostetler, J.B., Nordberg, E., *et al.* (2012) 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.

Accepted Article

-

- Gillespie, J.J., Kaur, S.J., Rahman, M.S., Rennoll-Bankert, K., Sears, K.T., Beier-Sexton, M., *et al.* (2015) Secretome of obligate intracellular *Rickettsia*. FEMS Microbiol Rev 39: 47–80.
- Gillespie, J.J., Phan, I.Q., Driscoll, T.P., Guillotte, M.L., Lehman, S.S., Rennoll-Bankert, K.E., *et al.* (2016). The *Rickettsia* type IV secretion system: unrealized complexity mired by gene family expansion. Pathogens and disease, 74: ftw058.
- Gomez-Valero, L., Rusniok, C., Jarraud, S., Vacherie, B., Rouy, Z., Barbe, V., *et al.* (2011) Extensive recombination events and horizontal gene transfer shaped the *Legionella pneumophila* genomes. BMC Genomics 12: 536.
- 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.
- Greuter, D., Loy, A., Horn, M., Rattei, T. (2016) probeBase--an online resource for rRNA-targeted oligonucleotide probes and primers: new features. Nucleic Acids Res 44: D586-D589.
- Gruber-Vodicka, H.R., Leisch, N., Kleiner, M., Hinzke, T., Liebeke, M., McFall-Ngai, M., *et al.* (2019) Two intracellular and cell type-specific bacterial symbionts in the placozoan *Trichoplax* H2. Nat Microbiol 4: 1465–1474.
- Horn, M., Collingro, A., Schmitz-Esser, S., Beier, C.L., Purkhold, U., Fartmann, B., *et al.* (2004) Illuminating the evolutionary history of *Chlamydiae*. Science 304: 728-730.
- Hoshino, Y., Gaucher, E.A. (2018) On the origin of isoprenoid biosynthesis. Mol Biol Evol 35: 2185–2197.
- Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., Tanabe, M. (2019) New approach for understanding genome variations in KEGG. Nucl Acids Res 47: D590–D595.

Accepted Article

- Kang, Y.J., Diao, X.N., Zhao, G.Y., Chen, M.H., Xiong, Y., Shi, M., *et al.* (2014) Extensive diversity of *Rickettsiales* bacteria in two species of ticks from China and the evolution of the *Rickettsiales*. BMC Evol Biol 14: 167.
- Karp, P.D., Billington, R., Caspi, R., Fulcher, C.A., Latendresse, M., Kothari, A., *et al.* (2019) The BioCyc collection of microbial genomes and metabolic pathways. Brief Bioinformatics 20: 1085-1093.
- Kent, B.N., Bordenstein, S.R. (2010) Phage WO of *Wolbachia*: lambda of the endosymbiont world. Trends Microbiol 18: 173-181.
- Kikuchi, Y., Fukatsu, T. (2015) *Rickettsia* infection in natural leech populations. Microb Ecol 49: 265-271.
- Klasson, L., Westberg, J., Sapountzis, P., Näslund, K., Lutnaes, Y., Darby, A.C., *et al.* (2009) The mosaic genome structure of the *Wolbachia* wRi strain infecting *Drosophila simulans*. Proc Natl Acad Sci 106: 5725-5730.
- Klinges, J.C., Rosales, S.M., McMinds, R., Shaver, E.C., Shantz, A.A., Peters, E.C., *et al.* (2019) Phylogenetic, genomic, and biogeographic characterization of a novel and ubiquitous marine invertebrate-associated *Rickettsiales* parasite, *Candidatus* Aquarickettsia rohweri, gen. nov., sp. nov. ISME J 13: 2938–2953.
- Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M., Blaxter, M. (2013) Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. Front Genet 4: 237.
- Langmead, B., Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357–359.
- Lanzoni, O., Fokin, S.I., Lebedeva, N., Migunova, A., Petroni, G., Potekhin, A. (2016) Rare freshwater ciliate *Paramecium chlolarelligerum* Kahl, 1935 and its macronuclear symbiotic bacterium "*Candidatus* Holospora parva". PLoS One 11: e0167928.

- Lanzoni, O., Sabaneyeva, E., Modeo, L., Castelli, M., Lebedeva, N., Verni, F., *et al.* (2019) Diversity and environmental distribution of the cosmopolitan endosymbiont "*Candidatus* Megaira". Sci Rep 9: 1179.
- Lin, H., Lou, B., Glynn, J.M., Doddapaneni, H., Civerolo, E.L., Chen, C., *et al.* (2011) The complete genome sequence of '*Candidatus* Liberibacter solanacearum', the bacterium associated with potato zebra chip disease. PLoS One 6: e19135.
- Lombard, J., Moreira, D. (2011) Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. Mol Biol Evol. 28: 87-99.
- Manz, W., Amann, R.I., Ludwig, W., Wagner, M., Schleifer, K.H. (1992) Phylogenetic oligodeoxynucleotide probes for the major subclasses of *Proteobacteria*: problems and solutions. Syst Appl Microbiol 15: 593–600.
- Martijn, J., Schulz, F., Zaremba-Niedzwiedzka, K., Viklund, J., Stepanauskas, R., Andersson, S.G.E., *et al.* (2015) Single-cell genomics of a rare environmental alphaproteobacterium provides unique insights into *Rickettsiaceae* evolution. ISME J 9: 2373-2385.
- Matic, I., Rayssiguier, C., Radman, M. (1995) Interspecies gene exchange in bacteria: the role of SOS and mismatch repair systems in evolution of species. Cell 80: 507-515.
- Matsuura, Y., Kikuchi, Y., Meng, X.Y., Koga, R., Fukatsu, T. (2012) Novel clade of alphaproteobacterial endosymbionts associated with stinkbugs and other arthropods. Appl Environ Microbiol. 78: 4149–56.
- McCutcheon, J.P., Moran, N. (2012) Extreme genome reduction in symbiotic bacteria. Nat Rev Biotechnol 10: 13-26.
- Mediannikov, O., Nguyen, T.T., Bell-Sakyi, L., Padmanabhan, R., Fournier, P.E., Raoult, D. (2014) High quality draft genome sequence and description of *Occidentia massiliensis* gen. nov., sp. nov., a new member of the family *Rickettsiaceae*. Standards in Genomic Sciences

Accepted Article

9: 9.

- Midha, S., Rigden, D.J., Siozios, S., Hurst, G.D.D., Jackson, A.P. (2020) The *Paracaedibacter*-like endosymbiont of *Bodo saltans* (Kinetoplastida) uses multiple putative toxin-antitoxin systems to maintain its host association. BioRxiv: 2020.07.24.217133.
- Min, C.K., Yang, J.S., Kim, S., Choi, M.S., Kim, I.S., Cho, N.H. (2008) Genome-based construction of the metabolic pathways of *Orientia tsutsugamushi* and comparative analysis within the *Rickettsiales* order. Comp Funct Genomics 2008: 623145.
- Modeo, L., Salvetti, A., Rossi, L., Castelli, M., Szokoli, F., Krenek, S., *et al.* (2020) "*Candidatus* Trichorickettsia mobilis", a *Rickettsiales* bacterium, can be transiently transferred from the unicellular eukaryote *Paramecium* to the planarian *Dugesia japonica*. PeerJ 8: e8977.
- Muñoz-Gómez, S., Hess, S., Burger, G., Lang, B.F., Susko, E., Slamovits, C.H., *et al.* (2019) An updated phylogeny of the *Alphaproteobacteria* reveals that the parasitic *Rickettsiales* and *Holosporales* have independent origins. eLife 8: e42535.
- Murray, R.G., Stackebrandt, E. (1995) Taxonomic note: implementation of the provisional status *Candidatus* for incompletely described procaryotes. Int J Syst Bacteriol 45: 186–187.
- Nakamura, T., Yamada, K.D., Tomii, K., Katoh, K. (2018) Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics 34: 2490–2492.
- Nikoh, N., Hosokawa, T., Moriyama, M., Oshima, K., Hattori, M., Fukatsu, T. (2014) Evolutionary origin of insect-*Wolbachia* nutritional mutualism. Proc Natl Acad Sci 111: 10257-10262.
- Ogata, H., La Scola, B., Audic, S., Renesto, P., Blanc, G., Robert, C., *et al.* (2006) Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. PloS Genet 2: e76.
- Olivieri, E., Epis, S., Castelli, M., Varotto Boccazzi, I., Romeo, C., Desirò, A., *et al.* (2019)

Accepted Article

Tissue tropism and metabolic pathways of *Midichloria mitochondrii* suggest tissue-specific functions in the symbiosis with *Ixodes ricinus*. Ticks Tick Borne Dis 10: 1070-1077.

- Oren, A., Garrity, G.M., Parker, C.T., Chuvochina, M., Trujillo, M.E. (2020) Lists of names of prokaryotic *Candidatus* taxa. Int J Syst Evol Microbiol 70: 3956-4042.
- Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A., *et al.* (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36: 996–1004.
- Perlman, S.J., Hunter, M.S., Zchori-Fein, E. (2006) The emerging diversity of *Rickettsia*. Proc R Soc B 273: 2097–2106.
- Pilgrim, J., Ander, M., Garros, C., Baylis, M., Hurst, G.D.D., Siozios, S. (2017) Torix group *Rickettsia* are widespread in *Culicoides* biting midges (Diptera: Ceratopogonidae), reach high frequency and carry unique genomic features. Environ Microbiol 19: 4238-4255.
- Pirritano, M., Zaburannyi, N., Grosser, K., Gasparoni, G., Müller, R., *et al.* (2020) Dual-Seq reveals genome and transcriptome of *Caedibacter taeniospiralis*, obligate endosymbiont of *Paramecium*. Sci Rep 10: 9727.
- Przyboś, E., Tarcz, S., Surmacz, M., Sawka, N., Fokin, S.I. (2013) *Paramecium tredecaurelia*: a unique non-polymorphic species of the *P. aurelia* spp. complex (Oligohymenophorea, Ciliophora). Acta Protozool 52: 257-266.
- Puigbo, P., Bravo, I.G., Garcia-Vallve, S. (2008) CAIcal: a combined set of tools to assess codon usage adaptation. Biology Direct 3: 38.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., *et al.* (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucl Acids Res 41: D590–D596.
- Rikihisa, Y. (2010) *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*: subversive manipulators of host cells. Nat Rev Microbiol 8: 328–39.

Accepted Article

Accepted Article

- Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61: 539–42.
- Sabaneyeva, E., Castelli, M., Szokoli, F., Benken, K., Lebedeva, N., Salvetti, A., *et al.* (2018) Host and symbiont intraspecific variability: The case of *Paramecium calkinsi* and "*Candidatus* Trichorickettsia mobilis". Eur J Protistol 62: 79–94.
- Sassera, D., Lo, N., Epis, S., D'Auria, G., Montagna, M., Comandatore, F., *et al.* (2011) Phylogenomic evidence for the presence of a flagellum and cbb3 oxidase in the free-living mitochondrial ancestor. Mol Biol Evol 28: 3285–96.
- Schrallhammer, M., Potekhin, A. (2020) Epidemiology of nucleus-dwelling *Holospora*: infection, transmission, adaptation, and interaction with *Paramecium*. In: Symbiosis: Cellular, Molecular, Medical and Evolutionary Aspects. Kloc, M. (ed). Results and Problems in Cell Differentiation, 69. Springer, Cham. pp. 105-135.
- Schu, M.G., Schrallhammer, M. (2018) Cultivation conditions can cause a shift from mutualistic to parasitic behavior in the symbiosis between *Paramecium* and its bacterial symbiont *Caedibacter taeniospiralis*. Curr Microbiol 75 : 1099-1102.
- Schulz, F., Martijn, J., Wascher, F., Lagkouvardos, I., Kostanjšek, R., Ettema, T.J.G., *et al.* (2016) A *Rickettsiales* symbiont of amoebae with ancient features. Environ Microbiol 18: 2326–42.
- Seemann, T. (2013) barrnap 0.5 : rapid ribosomal RNA prediction. Available online at: http://www.vicbioinformatics.com/
- Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30: 2068–9.
- Serra, V., Gammuto, L., Nitla, N., Castelli, M., Lanzoni, O., Sassera, D. *et al.* (2020) Morphology, ultrastructure, genomics, and phylogeny of *Euplotes vanleeuwenhoeki* sp. nov.

and its ultra-reduced endosymbiont "*Candidatus* Pinguicoccus supinus" sp. nov. Sci Rep 10: 20311

- Siddiqi, M.A., Rodwell, V.W. (1967) Bacterial metabolism of mevalonic acid. J Bacteriol 93: 207-214.
- Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., Chandler, M. (2006) ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34: D32–D36.
- Silverman, D.J., Wisseman, C.L. Jr, Waddell, A.D., Jones, M. (1978) External layers of *Rickettsia prowazekii* and *Rickettsia rickettsii*: occurrence of a slime layer. Infect Immun 22: 233-246.
- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210–3212.
- Stamatakis, A. (2015) Using RAxML to infer phylogenies. Curr Protoc Bioinform 51: 6.14.1–6.14.14.
- Szokoli, F., Castelli, M., Sabaneyeva, E., Schrallhammer, M., Krenek, S., Doak, T.G., *et al.* (2016a) Disentangling the taxonomy of *Rickettsiales* and description of two novel symbionts ("*Candidatus* Bealeia paramacronuclearis" and "*Candidatus* Fokinia cryptica") sharing the cytoplasm of the ciliate protist *Paramecium biaurelia*. Appl Environ Microbiol 82: 7236–47.
- Szokoli, F., Sabaneyeva, E., Castelli, M., Krenek, S., Schrallhammer, M., Soares, C.A.G., *et al.* (2016b) "*Candidatus* Fokinia solitaria", a novel "stand-alone" symbiotic lineage of *Midichloriaceae* (*Rickettsiales*). PLoS One 11: e0145743.
- Takatsuji, H., Nishino, T., Miki, I., Katsuki, H. (1983) Studies on isoprenoid biosynthesis with bacterial intact cells. Biochem Biophys Res Commun 110: 187-193.
- Taylor, M.J., Bandi, C., Hoerauf, A. (2005)*Wolbachia* endosymbionts of filarial nematodes. Adv Parasitol 60: 245–84.

Accepted Article

Accepted Article

- Talavera, G., Castresana, J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56: 564–577.
- Vannini, C., Petroni, G., Verni, F., Rosati, G. (2004) A bacterium belonging to the *Rickettsiaceae* family inhabits the cytoplasm of the marine ciliate *Diophrys appendiculata* (Ciliophora, Hypotrichia). Microb Ecol 49: 434–42.
- Vannini, C., Boscaro, V., Ferrantini, F., Benken, K.A., Mironov, T.I., Schweikert, M., *et al.* (2014) Flagellar movement in two bacteria of the family *Rickettsiaceae*: a re-evaluation of motility in an evolutionary perspective. PLoS One 9: e8771.
- Walker, D.H., Ismail, N (2008) Emerging and re-emerging rickettsioses: endothelial cell infection and early disease events. Nat Rev Microbiol 6: 375–86.
- Wang, Z., Wu, M. (2017) Comparative genomic analysis of *Acanthamoeba* endosymbionts highlights the role of amoebae as a "melting pot" shaping the *Rickettsiales* evolution. Genome Biol Evol 9: 3214-3224.
- Wang, H.L., Lei, T., Wang, X.W., Maruthi, M.N., Zhu, D.T., Cameron, S.L., *et al.* A newly recorded *Rickettsia* of the Torix group is a recent intruder and an endosymbiont in the whitefly *Bemisia tabaci*. Environ Microbiol 22: 1207-1221.
- Weinert, L.A., Werren, J.H., Aebi, A., Stone, G.N., Jiggins, F.M. (2009) Evolution and diversity of *Rickettsia* bacteria. BMC Biol 7: 6.
- Westram, R., Bader, K., Prüsse, E., Kumar, Y., Meier, H., Glöckner, F.O., *et al.* (2011) ARB: a software environment for sequence data. In Handbook of molecular microbial ecology I: metagenomics and complementary approaches. de Bruijn, F.J. (ed). Hoboken, New Jersey: John Wiley & Sons, pp. 399–406.
- Yan, Y., Jiang, L., Aufderheide, K.J., Wright, G.A., Terekhov, A., Costa, L., *et al.* (2014) A microfluidic-enabled mechanical microcompressor for the immobilization of live single- and multi-cellular specimens. Microsc. Microanal. 20: 141-151.

- Yang, A., Narechania, A., Kim, E. (2016) Rickettsial endosymbiont in the "early-diverging" streptophyte green alga *Mesostigma viride*. J Phycol 52: 219–229.
- Yurchenko, T., Ševčíková, T., Přibyl, P., El Karkouri, K., Klimeš, V., Amaral, R., *et al.* (2018) A gene transfer event suggests a long-term partnership between eustigmatophyte algae and a novel lineage of endosymbiotic bacteria. ISME J 12: 2163–75.

Accepted Article

Accepted Article

Figure legends

Figure 1. Fluorescence microscopy images of *P. tredecaurelia* WO2 cells. In (a) the cell was hybridised with the *Sarmatiella*-specific probe Sarmat_433 labelled with fluorescein-isothiocyanate (FITC; green signal) and with the almost-universal bacterial probe EUB338 labelled with Cy3 (red signal), and subsequently stained with DAPI (4′,6-diamidino-2-phenylindole; blue signal). In (b) the *Paramecium* cell was hybridised with the *Alphaproteobacteria*-targeted probe ALF1b labelled with FITC (green signal) and with the *Sarmatiella*-specific probe Sarmat_338 labelled with Cy3 (red signal), and stained with DAPI (blue). The cytoplasm of the ciliate cells (a, b) is populated by numerous *Sarmatiella* bacteria, coloured in green in (a), as they are not recognised by the EUB338 probe due to many mismatches in the sequence (see Results section for more details), and appearing yellowish in (b), due to the combined signals of the alphaproteobacterial and the specific probe. In some cells, in particular (a), a high number of bacteria appear densely concentrated in clusters (white arrowheads), reminiscent in their shape of digestive vacuoles, and possibly corresponding to the membrane-delimited areas observed in electron microscopy (Fig. 2c). Scale bars: 25 μm.

Figure 2. Transmission electron microscope images of *Sarmatiella* bacteria showing their intracellular localisation and ultrastructure. (a) Cross section through a *Sarmatiella* cell lying in the host cytoplasm, showing the two typical Gram-negative membranes and the surrounding *Rickettsia*-like electron-lucid halo (black arrow). (b) A view through the host cytoplasm with three *Sarmatiella* cells sectioned at variable orientations. Bacteria are frequently localised in proximity to host lipid droplets (L). (c) Some bacteria are located in areas delimited by single-layered host membranes (black arrows). These areas possibly represent autophagic digestive vacuoles, as they are in the vicinity of horseshoe-like cisterns (C) resembling phagophores, and sometimes enclosed bacteria appear degraded (black arrowhead), although still maintaining their typical surrounding electron-lucid halo. Occasionally, electron-lucid ‘‘holes’’ are visible inside the bacteria (white arrowhead).;

according to their size and appearance, these might possibly represent polyhydroxyalkanoate (PHA) granules. Scale bars: 0,2 μm (a), 0,8 μm (b), 1 μm (c).

Figure 3. Bayesian inference phylogenetic tree of selected members of the order *Rickettsiales*, based on 16S rRNA gene sequences. Numbers associated to each node represent bootstrap values inferred after 1000 maximum likelihood pseudo-replicates (obtained with the software phyML) and Bayesian posterior probabilities (values below 50|0.70 are not shown; full values in Supplementary figure S2). The novel *Sarmatiella* sequence characterised in this study is shown in bold. Members of classical *Rickettsiaceae*, basal *Rickettsiaceae* and the other three *Rickettsiales* families are evidenced by black lines on the right, while those of the two *Rickettsia* clades by dashed lines. The outgroup, composed by six other *Alphaproteobacteria*, is shown collapsed as a trapezoidal shape. "*Ca.*" is an abbreviation for "*Candidatus*". The scale bar stands for estimated proportional sequence divergence.

Figure 4. Bayesian inference phylogenomic tree of *Rickettsiales* based on concatenated alignment of 120 orthologous amino acidic sequences (Parks *et al.* 2018). Numbers associated to each node represent bootstrap values inferred after 1000 maximum likelihood pseudo-replicates (obtained with the software RaxML) and Bayesian posterior probabilities (Full trees in Supplementary figure S3). The novel *Sarmatiella* sequence characterised in this study is shown in bold. Members of classical *Rickettsiaceae*, basal *Rickettsiaceae* and the other three *Rickettsiales* families are evidenced by black lines on the right, while those of the two *Rickettsia* clades by dashed lines. The outgroup, composed by six other *Alphaproteobacteria*, is shown collapsed as a trapezoidal shape. The scale bar stands for estimated proportional sequence divergence. As indicated in the legend on the right, the presence of MEP/DOXP pathway (triangular shape), MEV (square shape) and *idi* genes (circular shape) is highlighted next to each organism. The numbers within squares and circles indicate the isoform of hmgr and idi present, respectively. The grey square indicates the presence of only a single MEV gene (*hmgr1*). Inferences on the ancestral gene repertoire in the Proto-

Accepted Article

Rickettsiales, as well as on possible gains (+) and losses (-) in different *Rickettsiales* lineages, are shown along the tree.

Figure 5. Schematic representation of the main metabolic and functional features of *Sarmatiella mevalonica* (Detailed description in Supplementary text S3). Features typically found in other *Rickettsiales*, but absent in *Sarmatiella*, are shown in red, with dashed lines. Among features present in *Sarmatiella*, solid back square boxes with capitalised text summarise multi-step pathways (evidenced in green if present in *Sarmatiella*), regular typeset indicate single compounds, italics indicate enzymes or transporters. Arrows stand for single enzymatic reaction (only selected enzyme names are reported), or links (as precursor or product) between a compound and a pathway, or between pathways sharing products and precursors. Bidirectional arrows stand for reversible reactions, arrows with bifurcations indicate two compounds as substrates/products of the same reaction, crossing arrows (e.g. for oxidative phosphorylation oxidoreductases) indicate the conjunct processing of two substrates into the respective products in the same reaction. Abbreviations (except universally used ones, e.g. ATP): MEV - mevalonate; MEP/DOXP - 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate; T1SS: type I secretion system; T3SS: type III secretion system; T4SS: type IV secretion system; T5SS: type V secretion system; PHA: polyhydroxyalkanoate; PEP: phosphoenolpyruvate; LPS: lipopolysaccharide. The “?” next to the T3SS indicates the still hypothetical status for this additional/alternative function of the partial flagellar apparatus.

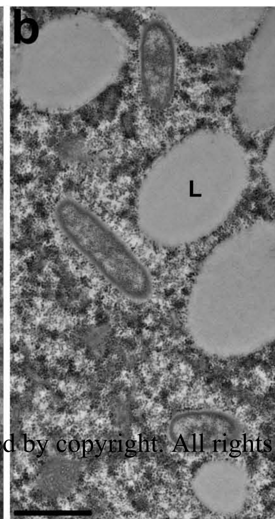
Accepted Article

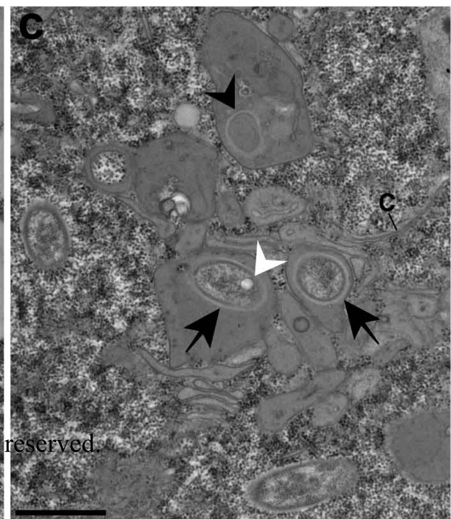
Accepted Article

a

b

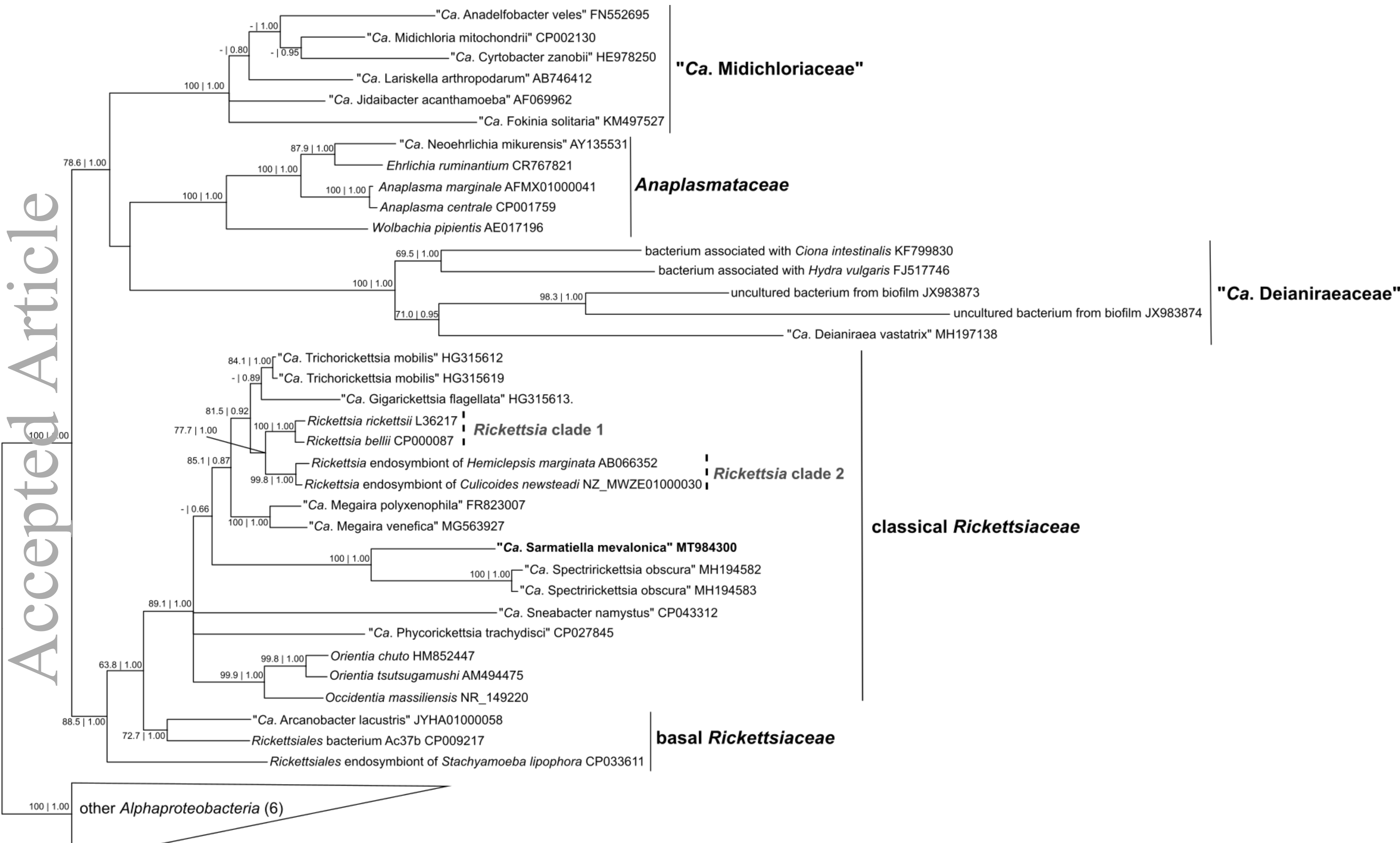
Accepted Article





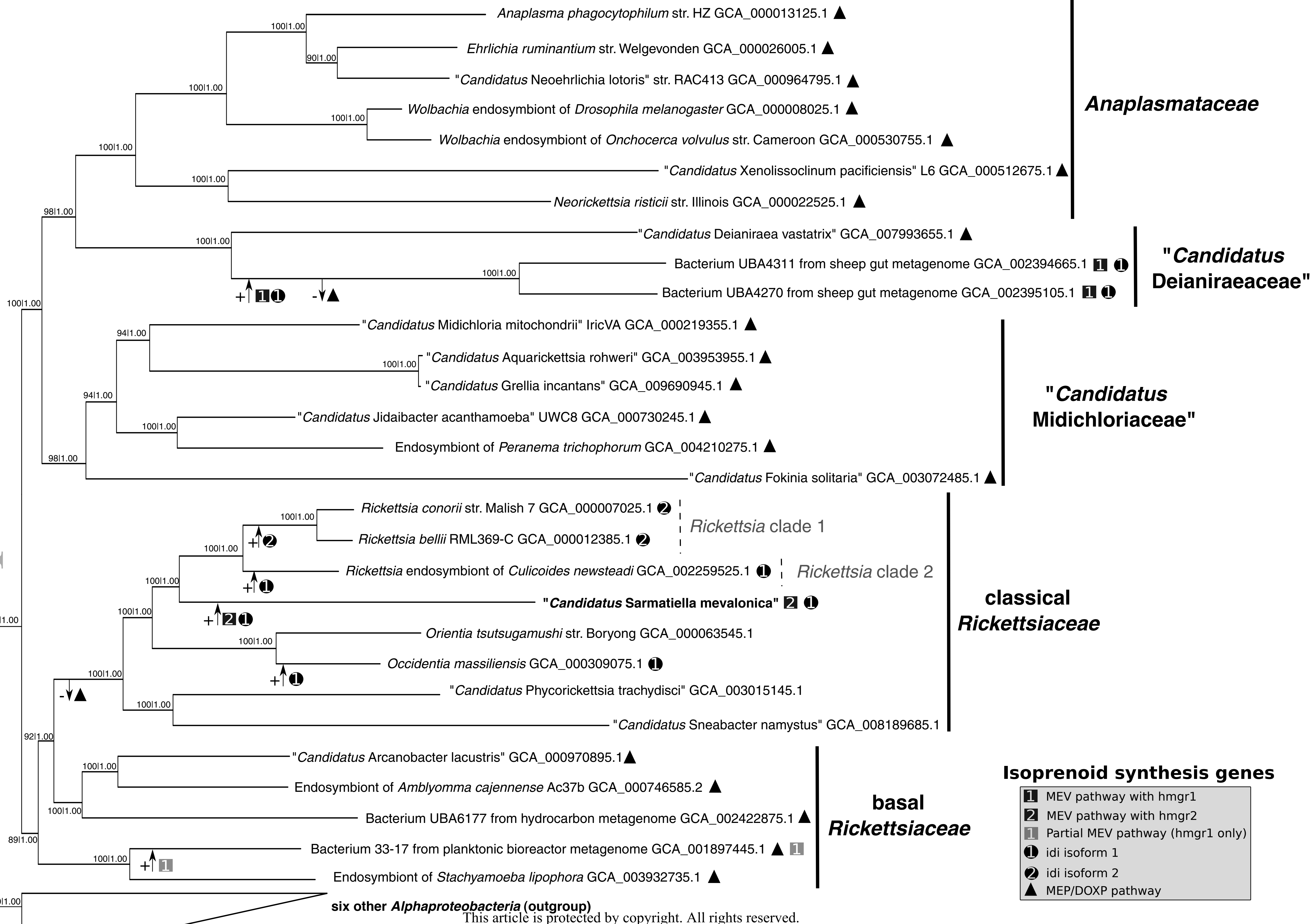
This article is protected by copyright. All rights reserved.

Accepted Article



This article is protected by copyright. All rights reserved.

Accepted Article



This article is protected by copyright. All rights reserved.

Accepted Article

This article is protected by copyright. All rights reserved.