



Comparative Genomic Analysis of *Holospora* spp., Intranuclear Symbionts of Paramecia

Sofya K. Garushyants^{1,2*}, Alexandra Y. Belavskaya^{3,4}, Dmitry B. Malko⁵, Maria D. Logacheva^{1,6}, Maria S. Rautian⁴ and Mikhail S. Gelfand^{1,2,5,7*}

¹ Skolkovo Institute of Science and Technology, Moscow, Russia, ² Kharkevitch Institute for Information Transmission Problems, Moscow, Russia, ³ Institute of Integrative Biology, University of Liverpool, Liverpool, United Kingdom, ⁴ Department of Invertebrate Zoology, Faculty of Biology, Saint Petersburg State University, Saint Petersburg, Russia, ⁵ National Research Center for Hematology, Moscow, Russia, ⁶ Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia, ⁷ National Research University Higher School of Economics, Moscow, Russia

OPEN ACCESS

Edited by:

Haiwei Luo,
The Chinese University of Hong Kong,
China

Reviewed by:

Joseph James Gillespie,
University of Maryland, Baltimore,
United States
Claudia Vannini,
Università degli Studi di Pisa, Italy

*Correspondence:

Sofya K. Garushyants
garushyants@iitp.ru
Mikhail S. Gelfand
gelfand@iitp.ru

Specialty section:

This article was submitted to
Evolutionary and Genomic
Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 06 December 2017

Accepted: 29 March 2018

Published: 16 April 2018

Citation:

Garushyants SK, Belavskaya AY,
Malko DB, Logacheva MD,
Rautian MS and Gelfand MS (2018)
Comparative Genomic Analysis
of *Holospora* spp., Intranuclear
Symbionts of Paramecia.
Front. Microbiol. 9:738.
doi: 10.3389/fmicb.2018.00738

While most endosymbiotic bacteria are transmitted only vertically, *Holospora* spp., an alphaproteobacterium from the *Rickettsiales* order, can desert its host and invade a new one. All bacteria from the genus *Holospora* are intranuclear symbionts of ciliates *Paramecium* spp. with strict species and nuclear specificity. Comparative metabolic reconstruction based on the newly sequenced genome of *Holospora curviuscula*, a macronuclear symbiont of *Paramecium bursaria*, and known genomes of other *Holospora* species shows that even though all *Holospora* spp. can persist outside the host, they cannot synthesize most of the essential small molecules, such as amino acids, and lack some central energy metabolic pathways, including glycolysis and the citric acid cycle. As the main energy source, *Holospora* spp. likely rely on nucleotides pirated from the host. *Holospora*-specific genes absent from other *Rickettsiales* are possibly involved in the lifestyle switch from the infectious to the reproductive form and in cell invasion.

Keywords: *Holospora*, *Rickettsiales*, endosymbionts, paramecia, ciliates, nuclear symbionts, genome

INTRODUCTION

Symbiotic associations with bacteria are common to many, if not all, eukaryotes. Endosymbiotic bacteria are well studied in plants, insects, and some other organisms (Moran et al., 2005; Kawaguchi and Minamisawa, 2010; McCutcheon and Moran, 2010; Bennett and Moran, 2013; Becker et al., 2015; Wu et al., 2015; Wexler et al., 2016; Zipfel and Oldroyd, 2017). By the original definition (Oulhen et al., 2016), a symbiotic relationship does not imply benefits for the host. Nevertheless, in symbiotic relationships, symbiont explores its host as a habitat. In turn, host can acquire some necessary products and/or properties which it unable to synthesize, e.g., ammonium produced for plants by nitrogen fixing bacteria, essential amino acids produced by endosymbionts of sap-sucking insects (Moran et al., 2005; Archibald, 2015; Sabater-Muñoz et al., 2017), or vitamins for blood-feeding tsetse fly (Dale and Welburn, 2001). Some symbionts, like some strains of *Wolbachia* spp., protect the host from pathogenic viruses (Hedges et al., 2008); or are necessary for the maturation of the immune system, as e.g., *Wigglesworthia* for tsetse flies (Weiss et al., 2011). *Caedibacter* creates growth advantages for infected paramecia and helps them to out-compete uninfected ones (Kusch et al., 2000). An endosymbiont of ciliates from the genus *Euplotes* is

essential for the host cell proliferation (Vannini et al., 2012, 2017). However, most relations are not well understood, and an advantage for the host is not always demonstrated, especially for facultative endosymbionts (Lu et al., 2016; Yañez et al., 2016).

Endosymbiotic bacteria and endonuclear parasites typically have several characteristic features, such as small genome size, low GC content, and short intergenic spacers (reviewed in McCutcheon and Moran, 2012; Batut et al., 2014; Martínez-Cano et al., 2015). Another important trait of these bacteria is an accelerated rate of genome evolution, caused by the minimal gene flow and substantial genetic drift due to the small effective population size of endosymbionts (Marais et al., 2008; Boscaro et al., 2017; Sabater-Muñoz et al., 2017). This leads to formation of pseudogenes and gene loss, and an overall decrease of the metabolic capacity (McCutcheon and Moran, 2012; Boscaro et al., 2017). The loss of genes involved in the reparation drives an even faster evolution of symbiotic species (McCutcheon and Moran, 2012). It also can lead to the decrease in the GC content (Horst et al., 1999; Moran et al., 2005; Long et al., 2018). A stable environment in the host cell does not require complicated regulatory systems, yielding reduction of intergenic regions (McCutcheon and Moran, 2012).

Holospora spp. are endonuclear symbiotic *Alphaproteobacteria* from the order *Rickettsiales* that inhabit either macro- or micronucleus of *Paramecium* spp. It has been suggested that *Holospora* spp. are parasites, as other *Rickettsiales*. A recent analysis of 16S rRNA lead some researchers to separate the basal rickettsial lineage of *Holospora*-like bacteria in a separate order *Holosporales* (Ferla et al., 2013; Szokoli et al., 2016), but for convenience we discuss both orders here together as *Rickettsiales*, because they have similar lifestyle and metabolic capacities. The genus *Holospora* currently is comprised of nine species, each showing clear nuclear and host specificity (Görtz and Schmidt, 2015). *Holospora* are not able to reproduce outside the host cell but can leave the host in order to invade a new one. Unlike most of the studied symbiotic species, *Holospora* have a complex life cycle involving two morphologically different forms. The short reproductive form exists only in the host nucleus and multiplies by binary fission, while the long infectious form does not multiply and is capable of infecting new host cells (Gromov and Ossipov, 1981; Fokin and Görtz, 2009). This form shows a unique cytological organization with a pronounced polarity (a large perinuclear space and a special tip), which seems to be of functional significance for the infection process. The infectious form enters new host via *Paramecium* food vacuole together with food bacteria, but escapes digestion, exits the vacuole and enters the cytoplasm followed by nuclear invasion. Two main mechanisms of this process have been proposed: *Holospora* either disrupt the digestive vacuole or enter the transport vacuoles (Schweikert et al., 2013). Then *Holospora* reach the nucleus by utilizing the host's actin cytoskeleton (Fokin and Görtz, 2009; Sabaneyeva et al., 2009). In *Holospora obtusa*, the 89 kDa periplasmic protein was shown to be associated with cell invasion (Iwatani et al., 2005). The invasion of macronucleus by *H. obtusa* is associated with release of the 63 kDa periplasmic protein into the macronucleus of the host (Abamo et al., 2008). Additionally, 15 and 39 kDa periplasmic proteins are released from the cell

tip during the macronucleus infection (Fujishima et al., 1997). Mechanisms underlying the transition from the reproductive to the infectious form and back are not well understood, but the 5.4 kDa protein with a signal peptide is only detected in the intermediate and infectious forms (Dohra et al., 1997).

In the presence of *Holospora*, *Paramecium* is able to grow, divide and mate (Schweikert et al., 2013). *Holospora* contribute to the heat-shock resistance in *Paramecium caudatum*, as cells infected with *H. obtusa* express high levels of hsp70 mRNA (Fujishima et al., 2005). They may also assist in acquiring high-salt and osmotic-shock resistance to the host (Fujishima, 2009). The presence of *H. caryophila* in the *Paramecium biaurelia* nucleus is advantageous in several cell lines during the exponential growth (Bella et al., 2016). Despite these observations of benefits for the host, *Holospora* have been shown to negatively affect host cells. For example, the presence of *Holospora elegans* in *Paramecium* leads to the formation of dysfunctional macronucleus during conjugation (Görtz and Fujishima, 1983). *Holospora undulata* increases mortality of the host, especially at low-food treatment (Restif and Kaltz, 2006), and high concentrations of the infectious form in the macronucleus inhibit the host cell growth (Fujishima, 2009). The relationship between *Holospora* spp. and their paramecia hosts seems to be a complex system in which benefits or damages for the host can be highly context-dependent.

The metabolism of *Rickettsiales*, and in particular of *Rickettsia* spp., has been studied in detail (Andersson et al., 1998; Hotopp et al., 2006; Fuxelius et al., 2007; Georgiades et al., 2011). *Rickettsia* export at least 51 metabolites from the host (Driscoll et al., 2017). They are not able to synthesize amino acids, nucleotides, lack glycolysis, and have to import such compounds as coenzyme A, pyruvate, FAD, biotin, etc.

So far, three *Holospora* genomes have been sequenced, all of which are endosymbionts of *Paramecium caudatum*: macronucleus-specific *H. obtusa* and micronucleus-specific *H. elegans* and *H. undulata* (Dohra et al., 2013, 2014). Analysis of common orthologous genes has yielded 572 single-copy core genes shared by the three genomes, and *Holospora* have been shown to rely on the host for energy production (Dohra et al., 2014). However, no detailed metabolic pathway reconstruction has been performed.

Here, we report a comparative analysis of four *Holospora* species, including *Holospora curviuscula*, a newly sequenced macronuclear endosymbiont of *P. bursaria*. We propose that *Holospora* use host-produced nucleotides as its energy source. We also identify the essential compounds that *Holospora* are likely able to synthesize, and compare their metabolism to that of other *Rickettsiales*.

MATERIALS AND METHODS

Growth Conditions and Genomic DNA Preparation

Holospora curviuscula has an obligate association with its host, *Paramecium bursaria*, and is therefore uncultivable, so the bacteria have been grown inside host cells. *Holospora*

curviuscula strain NRB217 from *Paramecium bursaria* isolated in the Novosibirsk Akademgorodok were obtained from the infected clones maintained in the CCCS (Culture Collection of Ciliates and their Symbionts), Research Park, Saint-Petersburg State University. The host cells were cultivated at the room temperature on the lettuce medium inoculated with *Enterobacter aerogenes* as a food resource for paramecia. The culture of *P. bursaria* bearing *H. curviuscula* was concentrated by centrifugation (10 min at 4500 g) and then homogenized using 10% solution of detergent Nonidet P-40 (Sigma-Aldrich Cat No. 21-3277 SAJ). The infectious forms of *H. curviuscula* were isolated from the homogenate by centrifugation in Percoll density gradient (Sigma-Aldrich Cat No. P1644) as described previously (Rautian and Wackerow-Kouzova, 2013). Genomic DNA was isolated with the DNeasy Blood and Tissue kit (QIAGEN Cat No. 69504) using a modified protocol — the time of incubation of cell homogenate with ATL buffer and proteinase K was increased to 16 h. All the subsequent steps were performed according to the standard Quick-Start Protocol.

Genome Assembly and Annotation

Two libraries were generated, paired-end MiSeq Illumina library (2 × 250) (PE), and mate-pair library with insertion size 4 kbp (MP). The initial MP library size was 2.5 million reads, and the PE library size, 4.5 million. MP reads were filtered with NextClip v1.3 (Leggett et al., 2014) and only category A pairs of reads (both reads in a pair contain adapters) were selected for further analysis. Both MP and PE were filtered by quality with trimmomatic version 0.33 (Leading 15, sliding −15, slidingwindow −4:25), after filtering 1.3 million MP reads and 1.7 million PE reads were retained for assembly. Processed reads were assembled with platanus version 1.2.1. The Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession PHHC00000000. The version described in this paper is version PHHC01000000.

To estimate the completeness of the *Holospora* assemblies, we used HMMer to search for essential PFAM domains (Rinke et al., 2013). A domain was considered to be present, if it was found by hmmsearch (with default parameters) (Mistry et al., 2013) and the bias was lower than the hit score. For all missing genes we checked whether they were present in 60 genomes of *Rickettsiales* and 23 other endosymbiotic *Proteobacteria* with tiny genomes less than 0.8 Mb (the complete list of genomes is given in Supplementary Table 1).

The resulting contigs were annotated with the PROKKA pipeline (Seemann, 2014). Metabolic pathways were reconstructed with SEED (Overbeek et al., 2005) and BioCyc (Caspi et al., 2016). To search for known bacteriocins we used APD3 (Wang et al., 2016), BAGEL3 (van Heel et al., 2013), and searched for known bacteriocin-associated PFAM domains. AntiSMASH (Weber et al., 2015) was used to predict biosynthesis of secondary metabolites. Transmembrane helices in proteins were predicted with TMHMM¹. Transporter protein

specificity was predicted by TCDB (Saier et al., 2016). SignalP and SecretomeP were used to predict signal peptides and secreted proteins, respectively (Bendtsen et al., 2005a; Nielsen, 2017). To predict the twin-arginine motif we utilized TatP server (Bendtsen et al., 2005b). Autotransporter proteins were predicted by search for PFAM domains PF03797, PF12951, PF05662, PF05658, PF11924, PF11557, PF16168, PF15403, PF03895 with HMMer package. To find phage-like regions in the genome assemblies, we used PHAST (Zhou et al., 2011). Additionally, we searched for all PFAM-domains with the keywords “transposase” and “phage” (328 PFAM domains).

16S rRNA-Based Phylogeny

To reconstruct the phylogenetic tree of *Holospora* spp., the PhyML software (Guindon et al., 2010) was used. 16S rRNAs from *H. curviuscula* NRB217, *H. elegans* E1, *H. obtusa*, and *H. undulata* HU1 were extracted from the full genome sequences, while 16S rRNAs from *H. acuminata* and *Candidatus Gortzia infectiva* were obtained from the SILVA database (accession numbers: KC164379 and HE797907, respectively). To root the tree, we used 16S rRNA of *Rickettsiales* (SILVA accession number: CP009217). The phylogenetic tree was constructed with PhyML v. 3.1 with 100 bootstrap replicates, GTR substitution model, and -o tlr parameter.

Analysis of Repeats

To estimate the fraction of repetitive DNA in the *Holospora* genome assemblies, reads were realigned with Bowtie 2 (−end-to-end mode) (Langmead and Salzberg, 2012) back to the respective assemblies, and the coverage for each nucleotide was calculated. In the case of *H. curviuscula*, we used PE reads generated in this study, reads for *H. obtusa* and *H. undulata* were downloaded from DDBJ (accession numbers DRP001203 and DRA001008, respectively). A fragment was presumed to be duplicated if it was longer than 20 bp, and its coverage was at least 1.6-fold higher than the median contig coverage.

Orthologous Groups and Genome Comparisons

To investigate the phylogenetic relations between *H. undulata* and *H. elegans*, pairwise genome alignments of all *Holospora* genomes were constructed with MAUVE (snapshot 2015-02-13) (Darling et al., 2010). Alignments of orthologous genes from the essential list (see above) were extracted from the MAUVE output, and all gaps were removed. The number of substitutions was calculated for each gene for each pair of genomes.

To investigate the gene repertoire of all *Holospora* and other *Rickettsiales*, we constructed groups of orthologous genes using OrthoMCL v. 2.0.9 (Li et al., 2003) with the MCL software v. 14-137. The output gene clusters were grouped in five subgroups using an *ad hoc* perl script which considered orthologs, co-orthologs and inparalogs. All *Rickettsiales* and other *Holospora* genomes (for the complete list see Supplementary Table 2) were downloaded from NCBI Genbank (Benson et al., 2013).

¹<http://www.cbs.dtu.dk/services/TMHMM/>

RESULTS

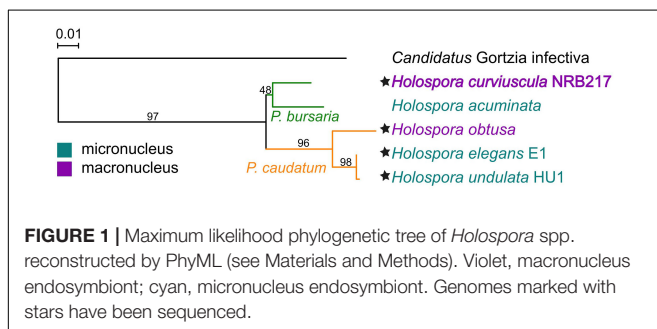
Holospora curviuscula Genome

The *H. curviuscula* draft genome sequenced here is comprised of 152 scaffolds (210 contigs) with N50 of ~39 kbp, total length of 1.7 Mb, and GC content of 37.6% (Table 1). The longest scaffold is 153367 bp and contains 126 CDSs. We predicted 1594 genes, including 1555 protein-coding genes with the average encoded protein length of 218 aa (Supplementary Figure 1). We assigned protein function to 683 genes. The genome assembly contains all rRNA genes and a set of 36 tRNAs necessary to recognize all codons. The 16S rRNA analysis revealed that the sequenced genome indeed belongs to the genus *Holospora*, and specifically to *H. curviuscula*, and is closer to *Holospora acuminata*, another endosymbiont of *P. bursaria*, than to *Holospora* from *P. caudatum* (Figure 1). The *H. curviuscula* genome is the largest *Holospora* genome sequenced so far (Table 1). Its GC content is slightly higher than that of most other non-*Rickettsiales* obligate endosymbionts (McCutcheon and Moran, 2012), but is typical for *Rickettsiales* (Supplementary Figure 2).

We estimated the quality of the assembly by searching for nearly universal bacterial genes (see Materials and Methods). Among the 139 universal PFAM domains, 121 are present in the *H. curviuscula* genome. Nine of the 18 missing domains have been found with a lower similarity threshold, which makes their presence uncertain. The remaining nine of the 18 missing domains are absent in all other *Holospora* as well. Moreover, these domains are also absent either in some endosymbiotic bacteria with small genomes or in some *Rickettsiales* genomes (Figure 2), meaning that their absence can be tolerated. We further searched for these 18 missing domains in unassembled sequencing reads of *H. curviuscula* (14% of all reads), and found no indication of their presence.

TABLE 1 | Results of genome assembly and annotation of *H. curviuscula*, and comparison with other *Holospora* species.

	<i>H. curviuscula</i>	<i>H. obtusa</i>	<i>H. undulata</i>	<i>H. elegans</i>
Genome (bp)	1715500	1334837	1402636	1268333
No. of contigs	210	91	203	152
GC content (%)	37.6%	35.2%	36.1%	36.0%
CDS	1594	1117	1224	1212



We manually checked 615 ORFs for frameshifts, and found only nine ORFs with frameshifts. This number may be an underestimate, because *H. curviuscula* genome contains short ORFs with unknown functions that have no homologs outside *Holospora* and can be remnants of some ancestral genes. Still, it implies that the *H. curviuscula* genome contains few pseudogenes.

Rickettsiales have been reported to have an atypical rRNA operon (Andersson et al., 1998). However, all *Holospora* have a standard rRNA operon, with 23S and 16S genes located close to each other and separated by several tRNA genes.

Reanalysis *H. undulata* and *H. elegans* Genomes

Holospora undulata HU1 and *H. elegans* E1 extracted from the micronucleus of *P. caudatum* have been recognized as separate species based on differences in the cell morphology (Görtz and Schmidt, 2015). However, our comparison of 16S rRNA genes from the available genome assemblies (Genbank IDs NZ_ARPM00000000 and NZ_BAUP00000000, respectively) of these species revealed just a single-nucleotide difference (Figure 1), yielding > 99.9% sequence identity, which is above the common thresholds of 97 or 98.7% used to define species (Janda and Abbott, 2007). That led us to inquire whether the genomes of *H. undulata* and *H. elegans* indeed represented different species or just strains of the same species. We constructed whole-genome pairwise alignments of the *Holospora* genomes and counted mismatches in the essential genes (Figure 3). The *H. elegans*–*H. undulata* pair carried ~100-fold fewer mismatches than all other genome pairs.

Holospora spp. Have Multiple Repeat Sequences and Contain Prophages

All available *Holospora* assemblies are comprised of multiple contigs, and reassembly has not led to better assemblies even with added genome coverage. Moreover, the genome length of *H. curviuscula* has been estimated at ca. 2 Mb (unpublished data), while the total assembly length is around 1.7 Mb. This suggests that *Holospora* genomes may contain a considerable fraction of repetitive DNA. To test this, we searched for fragments with high read coverage (see Materials and Methods). In *H. curviuscula*, we found ~300 repeats with the total length of ~300 kb (Figure 4A) and the estimated copy number varying from 2 to 21 (Figure 4B). In *H. obtusa* and *H. undulata*, the similarly estimated lengths of repetitive DNA were 60 and 161 kb, respectively. We were not able to estimate the fraction of repetitive DNA for *H. elegans*, as no raw sequencing data were available. The calculated repeat lengths could be underestimates, as we have applied strict coverage cutoffs yielding conservative repeat boundaries. For the same reason, we could overestimate the number of repeats by splitting long repeats into individual repeats in regions of low coverage. Still, the match between the difference in the assembly length and the genome length and the total coverage of repetitive DNA suggests that unassembled repeats comprise ~15% of the genome in *H. curviuscula*. The high copy-number repetitive regions in *H. curviuscula* include short ORFs of

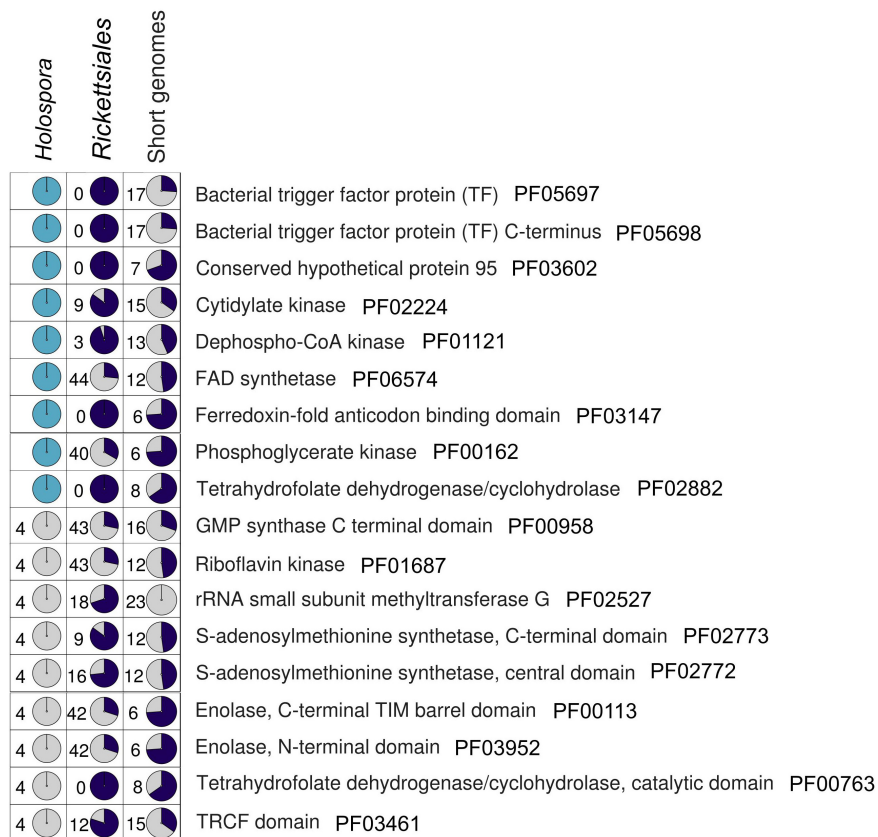


FIGURE 2 | The presence/absence pattern of 18 essential PFAM domains in *Holospira*, endosymbiotic bacteria with tiny genomes, and *Rickettsiales* (Supplementary Table 1). Color code: dark blue, the domain is present; gray, the domain is absent; cyan represents domains found with decreased similarity threshold (see Materials and Methods). The number in each cell is the number of genomes where the domain has not been found.

possible transposases. In addition, the identified repeats include two copies of rRNA operons with the similar structure, and also genes encoding prophage-like proteins.

To characterize prophages further, we searched for complete prophages and, additionally, for phage-related PFAM domains in all *Holospira* genomes. We found two possible prophage regions in *H. elegans*, one in *H. obtusa*, and two in *H. undulata*. One of the *H. undulata* prophages was intact, contained all proteins necessary for the phage replication and was surrounded by integration sequences (Supplementary Figure 3); the remaining copies were severely disrupted and lacked insertion sequences. Although no intact prophages were found in *H. curviuscula*, scaffold565 had a locus with two DDE superfamily endonucleases and two putative transposases, and other scaffolds carried genes encoding phage capsid and phage portal proteins similar to the ones found in other *Holospira* as well as some full-length and fragmentary transposases (data not shown).

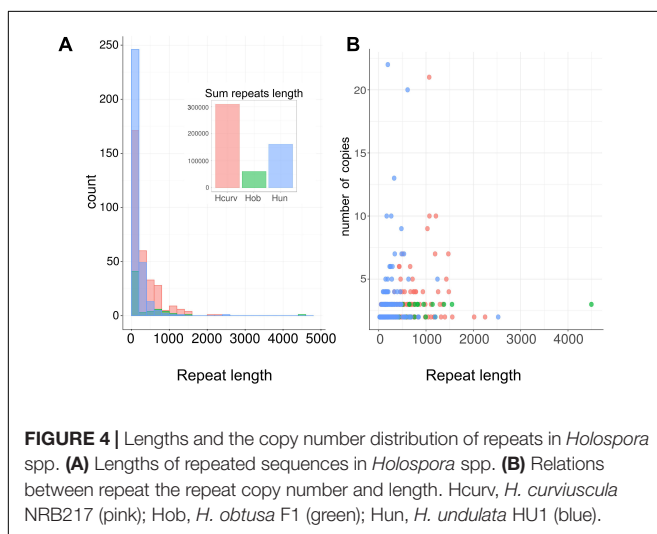
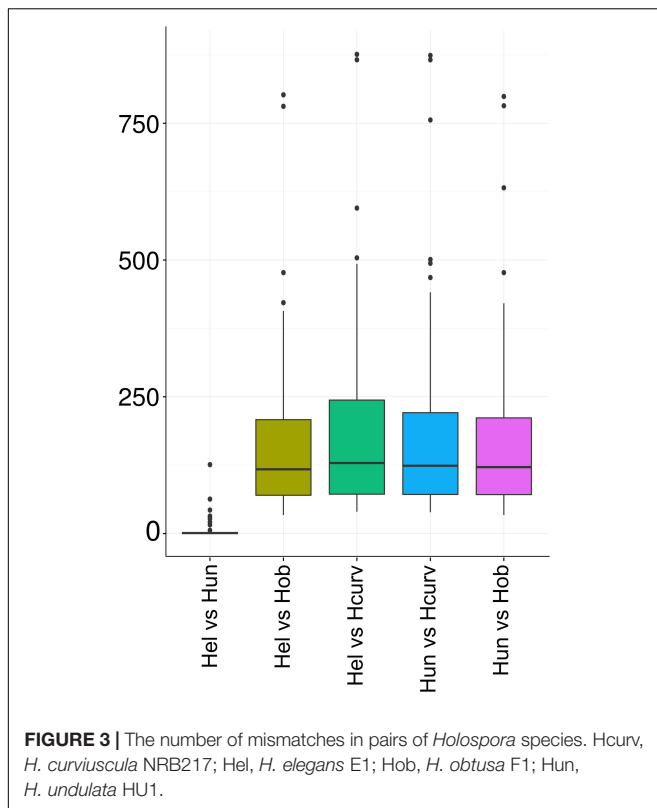
***Holospira* spp. Cannot Synthesize Amino Acids and Some Other Essential Small Molecules**

Holospira have a reduced metabolism even by the standards of the already gene-poor order *Rickettsiales*

(Georgiades et al., 2011). In particular, *Holospira* lack genes encoding enzymes of the citric acid cycle (CAC) (Supplementary Figure 4). By contrast, *Rickettsia* and other *Rickettsiales* have at least some of these enzymes and use them to convert small molecules (Driscoll et al., 2017).

All sequenced *Holospira* lack amino acid synthesis pathways. For most amino acids, all enzymes of the pathway are missing. The exceptions are the partial pathways for the biosynthesis of L-tryptophane, L-lysine, and L-glutamate and for the conversion of L-alanine to D-alanine, an essential compound for formation of the cell wall (Figure 5A). This means that all amino acids have to be imported from the host. While *Rickettsia* also cannot produce amino acids and probably import them from the host (Driscoll et al., 2017), other *Rickettsiales* are capable of synthesizing some amino acids such as L-lysine and L-glutamine (Fuxelius et al., 2007). Similar to *Rickettsia*, *Holospira* cannot produce chorismate or use it as a substrate for downstream reactions, and need to import it or its derivatives from the host.

Like *Rickettsia*, *Holospira* are not capable of purine and pyrimidine synthesis (Figure 5B), although, also like *Rickettsia*, they can convert UTP to NTP. Therefore, they need to obtain all nucleoside triphosphates from the host.



On the other hand, *Holospira* carry genes for ribonucleotide reductases and therefore can convert ribonucleotides into deoxyribonucleotides.

The ubiquinone biosynthesis pathway is also partial, as it is comprised of only dimethylallyl diphosphate (DMAPP) and the enzymes downstream of it (Supplementary Figure 6), similarly to what has been reported for *Rickettsia* (Driscoll et al., 2017). Unlike *Rickettsia*, *Holospira* are unable to convert DMAPP to isopentenyl diphosphate, although they carry geranyl-diphosphate synthase which is missing in *Rickettsia*.

Therefore, DMAPP seems to be obligatorily imported from the host.

The only major metabolic pathway that is almost intact is the fatty acids synthesis (Supplementary Figure 5).

In order to determine how the missing compounds are imported from the host, we analyzed predicted transporters. In *H. curviuscula*, thirty transport-associated genes have been identified. As expected, *Holospira* genomes encode oligopeptide and amino acid transporters, as well as proteases, some of which are periplasmic. In particular, we have found arginine/glutamine, proline, acetyl-serine/cysteine, choline/glycine/betaine, and putative branched amino acids candidate transporters. In addition, we found putative transport systems for magnesium, ferric ions, ribose, purines, sulfoacetate, L-galactonate or other sugars, and putrescine. However, these transport systems are insufficient to deliver all missing compounds to *Holospira*; in particular, it is not clear, how the remaining amino acids are delivered, unless as components of oligopeptides.

Holospira spp. Use Nucleotides as the Main Energy Source

All *Holospira* lack most genes involved in energy production. In particular, all enzymes necessary for the glycolysis except phosphoglycerate mutase (Figure 5C) and all enzymes of the citric acid cycle except malate dehydrogenase (Supplementary Figure 4) are missing. The F_1F_0 -ATPase is also missing. In addition, *Holospira* cannot produce coenzyme A from scratch, although as *Rickettsia*, they seem to have CoaE (PF01121, Figure 2) and thus are able to synthesize coenzyme A from dephospho-CoA (Driscoll et al., 2017). Of the pentose phosphate pathway, only the non-oxidative branch is present, as well as the downstream enzyme ribose-phosphate mutase that converts phosphorybosyl pyrophosphate (PRPP) to sugars (Figure 5C); the energy-producing oxidative pathway leading to the ribulose-5-phosphate is missing. By contrast, all *Holospira* have the pyruvate dehydrogenase complex and are able to convert pyruvate to acetyl-CoA, and further to acetoacetyl-CoA and acetoacetate, with the production of ATP (Figure 5D). All *Holospira* have a set of ribonucleotide reductases, which would allow them to use either nucleotides or ribonucleotides as an energy source. No obvious source of energy other than nucleotides was found.

Secretory Systems and Putative Invasins

Although most *Rickettsiales* are parasites and their genomes are highly reduced, they retain secretory systems such as the Tat and Sec pathways, a type IV system, and the TolC protein from a type I secretion system (Gillespie et al., 2015). Moreover, the VirB protein from a type IV secretory system is thought to be essential for the host invasion in most *Rickettsiales* (Rennoll-Bankert et al., 2015; Gillespie et al., 2016). By contrast, all studied *Holospira* demonstrate a significant decay of secretion pathways. They still possess the complete Sec system, additional systems helping to translocate proteins to the outer membrane (LolA, LolD, and possibly LolE) or transport them outside the cell (BamA, BamB, BamD, and chaperone DegP), and a TolC-like protein. However,

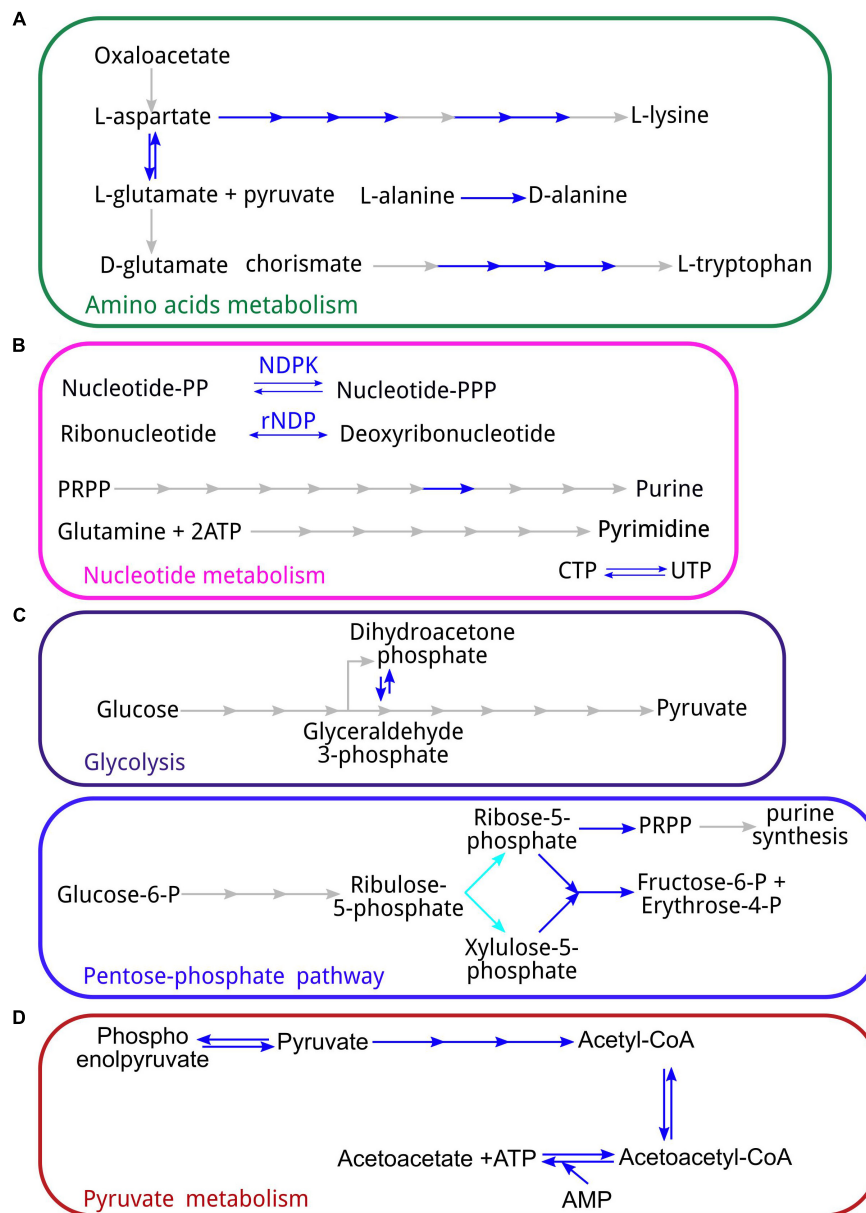


FIGURE 5 | Some aspects of the *Holospira* metabolism: **(A)** amino acid metabolism; **(B)** metabolism of purines and pyrimidines; **(C)** energy metabolism (top – glycolysis, bottom – pentose phosphate pathway); **(D)** pyruvate metabolism. Blue arrows represent reaction for which the corresponding enzyme has been found in all *Holospira* spp.; cyan – enzymes absent in *H. curviuscula*, but present in *H. undulata* and *H. obtusa*. PRPP, phosphoribosyl pyrophosphate; NDPK, nucleoside-diphosphate kinase; rNDP, ribonucleotide reductase.

we found no proteins similar to components of the Tat-system (TatA, TatB, or TatC). Since Tat-system proteins may be hard to identify (Gillespie et al., 2015), we performed a genome-wide search for proteins with the twin-arginine signal required for the recognition by the Tat-system and found no significant hits. Together, these observations show that the Tat transport system is indeed missing, which means that *Holospira* are unable to export folded substrates.

In *Rickettsia*, invasion is associated with proteins RaF, RickA, and Sca (Gillespie et al., 2015) which are missing in *Holospira*,

implying that *Holospira* use other mechanisms to invade the host cell. There are also no proteins with ankyrin domains assumed to play a role in the pathogenesis in *Rickettsia* (Gillespie et al., 2015). Although Sec proteins are present in all *Holospira*, only one protein with a putative autotransporter domain (PF03797) was found in *H. curviuscula* (HCUR_00103). Even though this gene has homologs in other *Holospira*, this domain was not predicted in them, implying that it can be a false positive. However, we found multiple (3–9 copies per genome) genes similar to *ompA* in all *Holospira*. *OmpA* has been previously shown to be involved

in the pathogenesis by *Rickettsiales* (reviewed in Palmer and Noh, 2012).

Holospira-Specific Genes

To search for genes that could be essential for the survival in the nucleus, we compared orthologous groups (OGs) for all *Holospira* with a set of all complete *Rickettsiales* genomes (Figure 6). We identified 102 OGs that occur in all *Holospira* and do not occur in any other *Rickettsiales* species (hereafter, *Holospira*-specific OGs, HOGs), 97 of HOGs contained no paralogs. HOGs encoded transporters, regulators, and hypothetical proteins. We expect proteins essential for survival and adaptation to the host either to be located on the surface of the bacterial cell or to be secreted. Therefore, we focused on those proteins that contained transmembrane helices or signal peptides. Thirty seven HOGs contained proteins with predicted transmembrane helices; 12 other OGs, proteins with predicted classic signal peptides; eighteen OGs were predicted to be secreted by the non-classical pathway (Supplementary Table 3). One of the latter HOGs contained the 89 kDa periplasmic protein which has been reported to be associated with cell invasion in *H. obtusa* (Iwatani et al., 2005). One HOG contained proteins similar to OmpA. These proteins are highly conserved among *Holospira*, and all *Holospira* except *H. curviuscula* carry two paralogs of this gene.

Holospira genomes encode numerous short (<100 aa) peptides with unknown functions. Short proteins have been shown to be involved in spore formation, regulation of transport, regulation of transcription, and signal transduction, or to possess antimicrobial or other toxic activities (Wang et al., 2008; Storz et al., 2014). Of the 102 HOGs, seventeen contained short proteins, including five HOGs that contained proteins with secretory signal peptide sequences. One of these HOGs has been described earlier as the 5.4 kDa protein involved in the switch between the reproductive and infectious forms (Dohra et al., 1997).

To determine whether there are proteins that could determine the nuclear specificity of *Holospira*, we searched for HOGs specific to the two macronuclear species, *H. curviuscula* and *H. obtusa*. We found eight HOGs present in both *H. curviuscula* and *H. obtusa* and absent in *H. undulata*, *H. elegans* and in other studied *Rickettsiales*. Of these, two HOGs encoded alpha/beta hydrolases; one, a possible aspartate/glutamate racemase or a malate isomerase; one, nucleotide sugar epimerase; one HOG had an unknown function; and three remaining HOGs encoded short proteins, one of which was similar to DDE superendonuclease, and another one had a predicted signal peptide.

DISCUSSION

Here we report a comparative analysis of the newly sequenced genome of *H. curviuscula* NRB217 and other available *Holospira* genomes. The phylogenetic analysis of all available *Holospira* has confirmed (Rautian and Wackerow-Kouzova, 2013) that they are clustered by the host species rather than by micro- vs. macronuclear specificity, so that *H. curviuscula* is the outgroup to the previously analyzed genomes. Furthermore, the fact that we found only one substitution in the 16S rRNA gene and few substitutions genome-wide between the sequenced samples of *H. elegans* and *H. undulata* suggests that these are in fact strains of the same species.

The previous genomic analysis of *Holospira* spp. (Dohra et al., 2014) included only a general description of the available COG categories and some reconstruction of the metabolism. It has been observed that many pathways are missing in *Holospira*. The pyruvate dehydrogenase complex that is present in *Holospira* has been proposed to be a possible relic of an ancestral pathway. Additionally, it has been suggested that *Holospira* strongly relies on the host for energy production.

Addition of *H. curviuscula* allows for a detailed comparative genomic analysis of the metabolism of *Holospira* spp. We found that all *Holospira* have reduced metabolic capacities and have to import many metabolites from the host.

Even though the genomes of *Holospira* are relatively large (Table 1) in comparison with other symbiotic species such as *Buchnera*, *Candidatus* Baumannia, or *Candidatus* Carsonella, *Holospira* are unable to produce most of the essential compounds. Furthermore, all available *Holospira* genomes seem to contain a large fraction of repetitive sequences, which complicates the genome assembly. Arguably, as all available

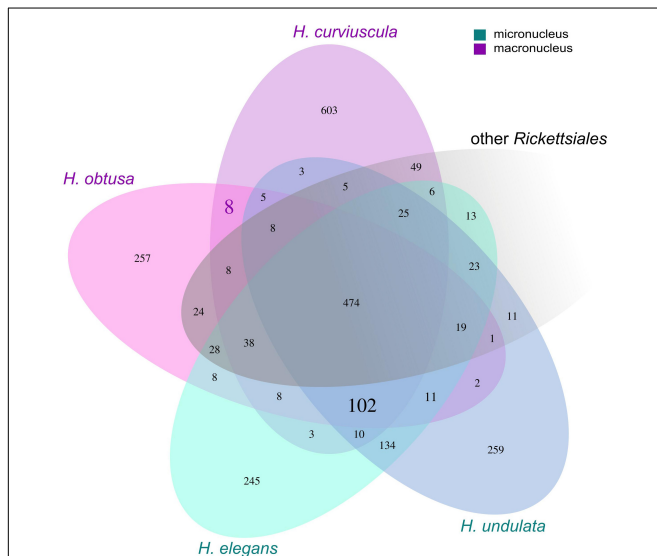


FIGURE 6 | Groups of orthologous proteins in *Holospira* spp. and *Rickettsiales*. Numbers in cells indicate either the number of orthologous groups, or, in the case of unique genes, the number of singletons. Numbers in bold represent groups of orthologous groups common for either macroneuclear (8) or all (102) *Holospira*. OGs common for the macroneuclear endosymbionts are two OGs encoding alpha/beta hydrolases; one, a possible aspartate/glutamate racemase or a malate isomerase; one, nucleotide sugar epimerase; one OG had an unknown function; and three remaining OGs encode short proteins, of which one was similar to the DDE superendonuclease, and another one has a predicted signal peptide. Of 102 *Holospira*-specific OGs: 30 are secreted, 37 contain putative transmembrane helices; 17 HOGs contain short peptides, of which 5 are potentially secreted.

Holospira genomes consist of multiple contigs, some of the enzymes could have been missed. However, quantitative analysis suggests that all universal genes are indeed present and that nearly all non-repetitive regions are contained in contigs; any potential missing part of the genome would have to be small, and cannot account for a large number of missing genes. Furthermore, our analysis is based on independent assemblies of multiple moderately related species, and it is unlikely that the same gene would be missing in several assemblies.

While smaller endosymbionts can produce at least some amino acids or retain some parts of the central metabolism, *Holospira* are unable to synthesize any amino acid, and have to acquire them from the host, despite the fact that some relics of pathways are seen, such as a partial pathway for the tryptophan synthesis. *Holospira* lack glycolysis, the Entner-Doudoroff pathway, and the pentose phosphate pathway. Surprisingly, *Holospira* have no enzymes of the citric acid cycle, which is unusual for *Rickettsiales*, as even the most reduced *Rickettsia* retain it.

Although we have not performed a complete metabolic reconstruction, it is evident that a broad range of compounds, including amino acids, DMAPP, and chorismate derivatives have to be imported from the host for *Holospira* to survive. It is unlikely that these compounds are accumulated by *Holospira* during its brief residence outside the nucleus. Indeed, the process of *Holospira* infection is a rapid one, whereas *Holospira* stay in the nucleus for long periods of time, which means that all these compounds have to be acquired from the host. Amino acids can be obtained by degrading host nuclear proteins, and indeed the *Holospira* infection is associated with increased proteolytic activity in the macronucleus (Freiburg, 1985). The situation with other compounds is more intriguing. All these small molecules can passively diffuse through the nuclear pore complex (Knockenbauer and Schwartz, 2016). Further, *Holospira* alter the host's gene expression and increase RNA synthesis (Freiburg, 1985; Fujishima, 2009). These alterations may help bacteria to acquire the necessary nutrients. To understand how the missing nutrients are delivered to *Holospira*, we searched for transport proteins and attempted to predict their specificity. This has explained some, but not all of the missing nutrients, and this topic has to be investigated further.

We propose that *Holospira* not only relies on the host for the energy production, but specifically use nucleotides or ribonucleotides as the energy source. Indeed, they are able to interconvert them and to convert UTP to CTP, and putative ribose transport proteins, which can also transport nucleotides, are present. We suggest that ribonucleotides are the preferred energy source, as the intracellular abundance of ribonucleotides can be 10- to 100-fold higher than that of dNTPs (Traut, 1994). This seems natural, given the nuclear habitat of *Holospira*.

The interactions between *Holospira* and their hosts have been studied in some depth. *Holospira* infection can either decrease or

increase the *Paramecium* viability under a variety of conditions (Fujishima et al., 2005; Hori et al., 2008; Fujishima, 2009; Bella et al., 2016). Addition of *Holospira curviuscula* to the analysis allowed us not only to investigate proteins specific to *Holospira* in a particular host, but also to find proteins that may play a role in the macronucleus infection. While the details of the mechanism remain unclear, it is likely that the interaction and stable infection rely on *Holospira*-specific proteins, in particular, secreted short conserved peptides. The 89 kDa periplasmic protein shown to be involved in the *H. obtusa* infection (Iwatani et al., 2005) is conserved in all studied *Holospira* spp., which indicates its importance. Finally, we propose that OmpA-like proteins may be involved in the *Holospira* invasion.

Overall, the analysis of the *Holospira* genomes demonstrates that their metabolic capabilities are unusually restricted, especially given the genome size, and provides a list of candidate genes essential for their unique lifestyle.

AUTHOR CONTRIBUTIONS

SG, MR, and MG designed the research. AB and MR isolated and cultivated bacteria. ML prepared the sequencing libraries and sequenced the genome. ML and SG assembled the genome. SG, AB, DM, and MG annotated the genome and performed the comparative analysis. SG, AB, MR, and MG wrote the paper.

FUNDING

This study was supported by the Russian Science Foundation under grant 14-50-00150.

ACKNOWLEDGMENTS

We thank Georgii Bazykin (IITP, Skoltech) for useful discussions. Scientific research was performed in part at the Research Park of the St. Petersburg State University Center for Molecular and Cell Technologies and the Center for Culturing Collection of Microorganisms. We are grateful to students of the Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University (classes of 2016 and 2017) who have annotated fragments of the *H. curviuscula* genome as a part of the regular course work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.00738/full#supplementary-material>

REFERENCES

- Abamo, F., Dohra, H., and Fujishima, M. (2008). Fate of the 63-kDa periplasmic protein of the infectious form of the endonuclear symbiotic bacterium *Holospira obtusa* during the infection process. *FEMS Microbiol. Lett.* 280, 21–27. doi: 10.1111/j.1574-6968.2007.01023.x
- 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, 133–140. doi: 10.1038/24094
- Archibald, J. M. (2015). Endosymbiosis and eukaryotic cell evolution. *Curr. Biol.* 25, R911–R921. doi: 10.1016/j.cub.2015.07.055
- Batut, B., Knibbe, C., Marais, G., and Daubin, V. (2014). Reductive genome evolution at both ends of the bacterial population size spectrum. *Nat. Rev. Microbiol.* 12, 841–850. doi: 10.1038/nrmicro3331
- Becker, M. H., Walke, J. B., Murrill, L., Woodhams, D. C., Reinert, L. K., Rollins-Smith, L. A., et al. (2015). Phylogenetic distribution of symbiotic bacteria from Panamanian amphibians that inhibit growth of the lethal fungal pathogen *Batrachochytrium dendrobatidis*. *Mol. Ecol.* 24, 1628–1641. doi: 10.1111/mec.13135
- Bella, C., Koehler, L., Grosser, K., Berendonk, T. U., Petroni, G., and Schrällhammer, M. (2016). Fitness impact of obligate intranuclear bacterial symbionts depends on host growth phase. *Front. Microbiol.* 7:2084. doi: 10.3389/fmicb.2016.02084
- Bendtsen, J. D., Kiemer, L., Fausbøll, A., and Brunak, S. (2005a). Non-classical protein secretion in bacteria. *BMC Microbiol.* 5:58. doi: 10.1186/1471-2180-5-58
- Bendtsen, J. D., Nielsen, H., Widdick, D., Palmer, T., and Brunak, S. (2005b). Prediction of twin-arginine signal peptides. *BMC Bioinformatics* 6:167. doi: 10.1186/1471-2105-6-167
- Bennett, G. M., and Moran, N. A. (2013). Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a Phloem-feeding insect. *Genome Biol. Evol.* 5, 1675–1688. doi: 10.1093/gbe/evt118
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., et al. (2013). GenBank. *Nucleic Acids Res.* 41, D36–D42. doi: 10.1093/nar/gks1195
- Boscaro, V., Kollisko, M., Felletti, M., Vannini, C., Lynn, D. H., and Keeling, P. J. (2017). Parallel genome reduction in symbionts descended from closely related free-living bacteria. *Nat. Ecol. Evol.* 1, 1160–1167. doi: 10.1038/s41559-017-0237-0
- Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., et al. (2016). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 44, D471–D480. doi: 10.1093/nar/gkv1164
- Dale, C., and Welburn, S. C. (2001). The endosymbionts of tsetse flies: manipulating host–parasite interactions. *Int. J. Parasitol.* 31, 628–631. doi: 10.1016/S0020-7519(01)00151-5
- Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. doi: 10.1371/journal.pone.0011147
- Dohra, H., Suzuki, H., Suzuki, T., Tanaka, K., and Fujishima, M. (2013). Draft genome sequence of *Holospira undulata* strain hu1, a micronucleus-specific symbiont of the ciliate *Paramecium caudatum*. *Genome Announc.* 1:e664-13. doi: 10.1128/genomeA.00664-13
- Dohra, H., Tanaka, K., Suzuki, T., Fujishima, M., and Suzuki, H. (2014). Draft genome sequences of three *Holospira* species (*Holospira obtusa*, *Holospira undulata*, and *Holospira elegans*), endonuclear symbiotic bacteria of the ciliate *Paramecium caudatum*. *FEMS Microbiol. Lett.* 359, 16–18. doi: 10.1111/1574-6968.12577
- Dohra, H., Yamamoto, K., Fujishima, M., and Ishikawa, H. (1997). Cloning and sequencing of gene coding for a periplasmic 5.4 kDa peptide of the macronucleus-specific symbiont *Holospira obtusa* of the ciliate *Paramecium caudatum*. *Zoolog. Sci.* 14, 69–75. doi: 10.2108/zsj.14.69
- Driscoll, T. P., Verhoeve, V. I., Guillothe, 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:e859-17. doi: 10.1128/mBio.00859-17
- Ferla, M. P., Thrash, J. C., Giovannoni, S. J., and Patrick, W. M. (2013). New rRNA gene-based phylogenies of the *Alphaproteobacteria* provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. *PLoS One* 8:e83383. doi: 10.1371/journal.pone.0083383
- Fokin, S. I., and Görtz, H.-D. (2009). “Diversity of *Holospira* bacteria in *Paramecium* and their characterization,” in *Endosymbionts in Paramecium Microbiology Monographs*, ed. M. Fujishima (Berlin: Springer), 161–199. doi: 10.1007/978-3-540-92677-1_7
- Freiburg, M. (1985). Isolation and characterization of macronuclei of *Paramecium caudatum* infected with the macronucleus-specific bacterium *Holospira obtusa*. *J. Cell Sci.* 73, 389–398.
- Fujishima, M. (2009). “Infection and maintenance of *Holospira* species in *Paramecium caudatum*,” in *Endosymbionts in Paramecium*, ed. M. Fujishima (Berlin: Springer), 201–225. doi: 10.1007/978-3-540-92677-1_8
- Fujishima, M., Dohra, H., and Kawai, M. (1997). Quantitative changes in periplasmic proteins of the macronucleus-specific bacterium *Holospira obtusa* in the infection process of the ciliate *Paramecium caudatum*. *J. Eukaryot. Microbiol.* 44, 636–642. doi: 10.1111/j.1550-7408.1997.tb05971.x
- Fujishima, M., Kawai, M., and Yamamoto, R. (2005). *Paramecium caudatum* acquires heat-shock resistance in ciliary movement by infection with the endonuclear symbiotic bacterium *Holospira obtusa*. *FEMS Microbiol. Lett.* 243, 101–105. doi: 10.1016/j.femsle.2004.11.053
- Fuxelius, H.-H., Darby, A., Min, C.-K., Cho, N.-H., and Andersson, S. G. E. (2007). The genomic and metabolic diversity of *Rickettsia*. *Res. Microbiol.* 158, 745–753. doi: 10.1016/j.resmic.2007.09.008
- Georgiades, K., Merhej, V., Karkouri, K. E., Raoult, D., and Pontarotti, P. (2011). Gene gain and loss events in *Rickettsia* and *Orientia* species. *Biol. Direct* 6:6. doi: 10.1186/1745-6150-6-6
- 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. doi: 10.1111/1574-6976.12084
- Gillespie, J. J., Phan, I. Q. H., Driscoll, T. P., Guillothe, 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. *Pathog. Dis.* 74:ftw058. doi: 10.1093/femspd/ftw058
- Görtz, H. D., and Fujishima, M. (1983). Conjugation and meiosis of *Paramecium caudatum* infected with the micronucleus-specific bacterium *Holospira elegans*. *Eur. J. Cell Biol.* 32, 86–91.
- Görtz, H.-D., and Schmidt, H. J. (2015). “*Holospira*,” in *Bergey’s Manual of Systematics of Archaea and Bacteria*. Hoboken, NJ: John Wiley & Sons, Ltd.
- Gromov, B. V., and Ossipov, D. V. (1981). *Holospira* (ex Hafkine 1890) nom. rev., a genus of bacteria inhabiting the nuclei of paramecia. *Int. J. Syst. Evol. Microbiol.* 31, 348–352. doi: 10.1099/00207713-31-3-348
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Hedges, L. M., Brownlie, J. C., O’Neill, S. L., and Johnson, K. N. (2008). Wolbachia and virus protection in insects. *Science* 322:702. doi: 10.1126/science.1162418
- Hori, M., Fujii, K., and Fujishima, M. (2008). Micronucleus-specific bacterium *Holospira elegans* irreversibly enhances stress gene expression of the host *Paramecium caudatum*. *J. Eukaryot. Microbiol.* 55, 515–521. doi: 10.1111/j.1550-7408.2008.00352.x
- Horst, J.-P., Wu, T., Marinus, M. G., Horst, J.-P., Wu, T., Marinus, M. G., et al. (1999). *Escherichia coli* mutator genes. *Trends Microbiol.* 7, 29–36. doi: 10.1016/S0966-842X(98)01424-3
- Hotopp, J. C. D., 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. doi: 10.1371/journal.pgen.0020021
- Iwatani, K., Dohra, H., Lang, B. F., Burger, G., Hori, M., and Fujishima, M. (2005). Translocation of an 89-kDa periplasmic protein is associated with *Holospira* infection. *Biochem. Biophys. Res. Commun.* 337, 1198–1205. doi: 10.1016/j.bbrc.2005.09.175
- Janda, J. M., and Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J. Clin. Microbiol.* 45, 2761–2764. doi: 10.1128/JCM.01228-07
- Kawaguchi, M., and Minamisawa, K. (2010). Plant–microbe communications for symbiosis. *Plant Cell Physiol.* 51, 1377–1380. doi: 10.1093/pcp/pcq125
- Knockenbauer, K. E., and Schwartz, T. U. (2016). The nuclear pore complex as a flexible and dynamic gate. *Cell* 164, 1162–1171. doi: 10.1016/j.cell.2016.01.034
- Kusch, J., Stremmel, M., Breiner, H.-W., Adams, V., Schweikert, M., and Schmidt, H. J. (2000). The toxic symbiont *Caedibacter caryophila* in the cytoplasm of *Paramecium novaeurelia*. *Microb. Ecol.* 40, 330–335.
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

- Leggett, R. M., Clavijo, B. J., Clissold, L., Clark, M. D., and Caccamo, M. (2014). NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. *Bioinformatics* 30, 566–568. doi: 10.1093/bioinformatics/btt702
- Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503
- Long, H., Sung, W., Kucukyildirim, S., Williams, E., Miller, S. F., Guo, W., et al. (2018). Evolutionary determinants of genome-wide nucleotide composition. *Nat. Ecol. Evol.* 2, 237–240. doi: 10.1038/s41559-017-0425-y
- Lu, M., Hulcr, J., and Sun, J. (2016). The role of symbiotic microbes in insect invasions. *Annu. Rev. Ecol. Syst.* 47, 487–505. doi: 10.1146/annurev-ecolsys-121415-032050
- Marais, G. A. B., Calteau, A., and Tenaillon, O. (2008). Mutation rate and genome reduction in endosymbiotic and free-living bacteria. *Genetica* 134, 205–210. doi: 10.1007/s10709-007-9226-6
- Martínez-Cano, D. J., Reyes-Prieto, M., Martínez-Romero, E., Partida-Martínez, L. P., Latorre, A., Moya, A., et al. (2015). Evolution of small prokaryotic genomes. *Front. Microbiol.* 5:742. doi: 10.3389/fmicb.2014.00742
- McCutcheon, J. P., and Moran, N. A. (2010). Functional convergence in reduced genomes of bacterial symbionts spanning 200 million years of evolution. *Genome Biol. Evol.* 2, 708–718. doi: 10.1093/gbe/evq055
- McCutcheon, J. P., and Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10, 13–26. doi: 10.1038/nrmicro2670
- Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 41:e121. doi: 10.1093/nar/gkt263
- Moran, N. A., Degnan, P. H., Santos, S. R., Dunbar, H. E., and Ochman, H. (2005). The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 16919–16926. doi: 10.1073/pnas.0507029102
- Nielsen, H. (2017). Predicting secretory proteins with SignalP. *Methods Mol. Biol.* 1611, 59–73. doi: 10.1007/978-1-4939-7015-5_6
- Oulhen, N., Schulz, B. J., and Carrier, T. J. (2016). English translation of heinrich anton de Bary's 1878 speech, 'Die Erscheinung der Symbiose' ('De la symbiose'). *Symbiosis* 69, 131–139. doi: 10.1007/s13199-016-0409-8
- Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H.-Y., Cohoon, M., et al. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33, 5691–5702. doi: 10.1093/nar/gki866
- Palmer, G. H., and Noh, S. M. (2012). Rickettsial entry into host cells: finding the keys to unlock the doors. *Infect. Immun.* 80, 3746–3747. doi: 10.1128/IAI.00836-12
- Rautian, M. S., and Wackerow-Kouzova, N. D. (2013). Phylogenetic placement of two previously described intranuclear bacteria from the ciliate *Paramecium bursaria* (Protozoa, Ciliophora): "*Holospira acuminata*" and "*Holospira curviuscula*." *Int. J. Syst. Evol. Microbiol.* 63, 1930–1933. doi: 10.1099/ijso.046631-0
- Rennoll-Bankert, K. E., Rahman, M. S., Gillespie, J. J., Guillotte, M. L., Kaur, S. J., Lehman, S. S., et al. (2015). Which way in? The RalF Arf-GEF orchestrates *Rickettsia* host cell invasion. *PLoS Pathog.* 11:e1005115. doi: 10.1371/journal.ppat.1005115
- Restif, O., and Kaltz, O. (2006). Condition-dependent virulence in a horizontally and vertically transmitted bacterial parasite. *Oikos* 114, 148–158. doi: 10.1111/j.2006.0030-1299.14611.x
- Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499, 431–437. doi: 10.1038/nature12352
- Sabaneyeva, E. V., Derkacheva, M. E., Benken, K. A., Fokin, S. I., Vainio, S., and Skovorodkin, I. N. (2009). Actin-based mechanism of *Holospira obtusa* trafficking in *Paramecium caudatum*. *Protist* 160, 205–219. doi: 10.1016/j.protis.2008.11.006
- Sabater-Muñoz, B., Toft, C., Alvarez-Ponce, D., and Fares, M. A. (2017). Chance and necessity in the genome evolution of endosymbiotic bacteria of insects. *ISME J.* 11, 1291–1304. doi: 10.1038/ismej.2017.18
- Saier, M. H., Reddy, V. S., Tsu, B. V., Ahmed, M. S., Li, C., and Moreno-Hagelsieb, G. (2016). The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res.* 44, D372–D379. doi: 10.1093/nar/gkv1103
- Schweikert, M., Fujishima, M., and Görtz, H.-D. (2013). "Symbiotic Associations Between Ciliates and Prokaryotes," in *The Prokaryotes*. Berlin: Springer, 427–463. doi: 10.1007/978-3-642-30194-0_18
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Storz, G., Wolf, Y. I., and Ramamurthi, K. S. (2014). Small proteins can no longer be ignored. *Annu. Rev. Biochem.* 83, 753–777. doi: 10.1146/annurev-biochem-070611-102400
- Szokoli, F., Castelli, M., Sabaneyeva, E., Schrällhammer, M., Krenke, S., Doak, T. G., et al. (2016). 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, 7236–7247. doi: 10.1128/AEM.02284-16
- Traut, T. W. (1994). Physiological concentrations of purines and pyrimidines. *Mol. Cell. Biochem.* 140, 1–22. doi: 10.1007/BF00928361
- van Heel, A. J., de Jong, A., Montalbán-López, M., Kok, J., and Kuipers, O. P. (2013). BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res.* 41, W448–W453. doi: 10.1093/nar/gkt391
- Vannini, C., Ferrantini, F., Ristori, A., Verni, F., and Petroni, G. (2012). Betaproteobacterial symbionts of the ciliate *Euplotes*: origin and tangled evolutionary path of an obligate microbial association. *Environ. Microbiol.* 14, 2553–2563. doi: 10.1111/j.1462-2920.2012.02760.x
- Vannini, C., Sigona, C., Hahn, M., Petroni, G., and Fujishima, M. (2017). High degree of specificity in the association between symbiotic betaproteobacteria and the host *Euplotes* (Ciliophora, Euplotia). *Eur. J. Protistol.* 59, 124–132. doi: 10.1016/j.ejop.2017.04.003
- Wang, F., Xiao, J., Pan, L., Yang, M., Zhang, G., Jin, S., et al. (2008). A systematic survey of mini-proteins in bacteria and archaea. *PLoS One* 3:e4027. doi: 10.1371/journal.pone.0004027
- Wang, G., Li, X., and Wang, Z. (2016). APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* 44, D1087–D1093. doi: 10.1093/nar/gkv1278
- Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Brucocoleri, R., et al. (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 43, W237–W243. doi: 10.1093/nar/gkv437
- Weiss, B. L., Wang, J., and Aksoy, S. (2011). Tsetse immune system maturation requires the presence of obligate symbionts in larvae. *PLoS Biol.* 9:e1000619. doi: 10.1371/journal.pbio.1000619
- Wexler, A. G., Bao, Y., Whitney, J. C., Bobay, L.-M., Xavier, J. B., Schofield, W. B., et al. (2016). Human symbionts inject and neutralize antibacterial toxins to persist in the gut. *Proc. Natl. Acad. Sci. U.S.A.* 113, 3639–3644. doi: 10.1073/pnas.1525637113
- Wu, M., McNulty, N. P., Rodionov, D. A., Khoroshkin, M. S., Griffin, N. W., Cheng, J., et al. (2015). Genetic determinants of in vivo fitness and diet responsiveness in multiple human gut *Bacteroides*. *Science* 350:aac5992. doi: 10.1126/science.aac5992
- Yañez, O., Gauthier, L., Chantawannakul, P., and Neumann, P. (2016). Endosymbiotic bacteria in honey bees: *Arsenophonus* spp. are not transmitted transovarially. *FEMS Microbiol. Lett.* 363:fnw147. doi: 10.1093/femsle/fnw147
- Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39, W347–W352. doi: 10.1093/nar/gkr485
- Zipfel, C., and Oldroyd, G. E. D. (2017). Plant signalling in symbiosis and immunity. *Nature* 543, 328–336. doi: 10.1038/nature22009

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Garushyants, Beliavskaia, Malko, Logacheva, Rautian and Gelfand. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.