



Genome of *Ca. Pandoraea novymonadis*, an Endosymbiotic Bacterium of the Trypanosomatid *Novymonas esmeraldas*

Alexei Y. Kostygov^{1,2†}, Anzhelika Butenko^{1,3†}, Anna Nenarokova^{3,4}, Daria Tashyreva³, Pavel Flegontov^{1,3,5}, Julius Lukeš^{3,4} and Vyacheslav Yurchenko^{1,3,6*}

¹ Life Science Research Centre, Faculty of Science, University of Ostrava, Ostrava, Czechia, ² Zoological Institute of the Russian Academy of Sciences, St. Petersburg, Russia, ³ Biology Centre, Institute of Parasitology, Czech Academy of Sciences, České Budějovice, Czechia, ⁴ Faculty of Sciences, University of South Bohemia, České Budějovice, Czechia, ⁵ Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia, ⁶ Institute of Environmental Technologies, Faculty of Science, University of Ostrava, Ostrava, Czechia

OPEN ACCESS

Edited by:

João Marcelo Pereira Alves,
University of São Paulo, Brazil

Reviewed by:

Zhao-Rong Lun,
Sun Yat-sen University, China
Vera Tai,
University of Western Ontario, Canada

*Correspondence:

Vyacheslav Yurchenko
vyacheslav.yurchenko@osu.cz

† These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Microbial Symbioses,
a section of the journal
Frontiers in Microbiology

Received: 27 June 2017

Accepted: 21 September 2017

Published: 04 October 2017

Citation:

Kostygov AY, Butenko A,
Nenarokova A, Tashyreva D,
Flegontov P, Lukeš J and
Yurchenko V (2017) Genome of *Ca.*
Pandoraea novymonadis, an
Endosymbiotic Bacterium of the
Trypanosomatid *Novymonas*
esmeraldas. *Front. Microbiol.* 8:1940.
doi: 10.3389/fmicb.2017.01940

We have sequenced, annotated, and analyzed the genome of *Ca. Pandoraea novymonadis*, a recently described bacterial endosymbiont of the trypanosomatid *Novymonas esmeraldas*. When compared with genomes of its free-living relatives, it has all the hallmarks of the endosymbionts' genomes, such as significantly reduced size, extensive gene loss, low GC content, numerous gene rearrangements, and low codon usage bias. In addition, *Ca. P. novymonadis* lacks mobile elements, has a strikingly low number of pseudogenes, and almost all genes are single copied. This suggests that it already passed the intensive period of host adaptation, which still can be observed in the genome of *Polynucleobacter necessarius*, a certainly recent endosymbiont. Phylogenetically, *Ca. P. novymonadis* is more related to *P. necessarius*, an intracytoplasmic bacterium of free-living ciliates, than to *Ca. Kinetoplastibacterium* spp., the only other known endosymbionts of trypanosomatid flagellates. As judged by the extent of the overall genome reduction and the loss of particular metabolic abilities correlating with the increasing dependence of the symbiont on its host, *Ca. P. novymonadis* occupies an intermediate position *P. necessarius* and *Ca. Kinetoplastibacterium* spp. We conclude that the relationships between *Ca. P. novymonadis* and *N. esmeraldas* are well-established, although not as fine-tuned as in the case of Strigomonadinae and their endosymbionts.

Keywords: bacterial endosymbiont, *Pandoraea*, phylogenomics, metabolism, Trypanosomatidae

INTRODUCTION

Pandoraea is a genus of Gram-negative rod-shaped β -proteobacteria belonging to the family Burkholderiaceae of the order Burkholderiales. Members of this genus are phenotypically diverse, reflecting a wide spectrum of life strategies. Several species of these microorganisms were documented as opportunistic pathogens in cystic fibrosis patients or in individuals after lung transplantation (Coenye et al., 2000; Stryjewski et al., 2003). Besides, a number of *Pandoraea* spp. (including some pathogenic ones) were isolated from environmental samples such as soils, hen

ding, and oxic water layer above a sulfide-containing sediment (Coenye et al., 2000; Anandham et al., 2010; Sahin et al., 2011). These free-living species participate in the biodegradation of various organic substances (including important pollutants) or perform chemosynthesis by oxidation of heterotrophic sulfur (Okeke et al., 2002; Graff and Stubner, 2003; Ozaki et al., 2007; Liz et al., 2009; Kumar et al., 2015; Jeong et al., 2016).

Previously, we discovered a new species of *Pandoraea*, which, in contrast to its relatives, is an intracellular symbiont of the flagellate *Novymonas esmeraldas* (Kinetoplastea: Trypanosomatidae) (Flegontov et al., 2016). This endosymbiosis appears to have been established relatively recently as judged by the fact that neither of the two participants has close relatives involved in similar relationships. In addition, the division of *Ca. Pandoraea novymonadis* is not synchronized with that of the host cell. As a result, the number of endosymbionts per *Novymonas* cell is unstable and bacteria-free trypanosomatids appear at a relatively high frequency of ~6%. We hypothesized that the endosymbiosis is favorable for *N. esmeraldas*, since large-scale cloning experiments did not reveal any aposymbiotic clone (Flegontov et al., 2016).

All other studied endosymbioses in trypanosomatids are restricted to flagellates of the subfamily Strigomonadinae (Votýpka et al., 2014) and bacteria *Ca. Kinetoplastibacterium* spp. (Burkholderiales: Alcaligenaceae). These relationships seem to have been established earlier in evolution. As judged from the phylogenies of the prokaryotic and eukaryotic partners, the origin of this endosymbiosis was a single event (Du et al., 1994; Teixeira et al., 2011). The long evolution of *Kinetoplastibacterium* resulted in “one bacterium per host cell” arrangement with fine-tuned mechanisms synchronizing their division (Motta et al., 2010; Catta-Preta et al., 2015). The bacterium provides its host with essential nutrients and is remunerated with a direct access to the ATP-producing glycosomes (Motta et al., 1997; de Souza and Motta, 1999; Alves et al., 2011, 2013a).

The free-living hypotrichous ciliates (*Euplotes aediculatus* and related species) with their intracytoplasmic bacterium *Ca. Polynucleobacter necessarius* (hereafter conventionally called *P. necessarius*) represent yet another endosymbiotic association in a protist, reminiscent of the *Novymonas/Pandoraea* system. Although the ciliates are evolutionary extremely distant from trypanosomatids and represent a different eukaryotic supergroup (SAR versus Excavata), their endosymbiont *P. necessarius* belongs to the same β -proteobacterial family *Burkholderiaceae*. This endosymbiosis seems to be quite recent, since there is a very closely related free-living bacterium formally attributed to a separate species *P. asymbioticus*, but showing 99% identity with *P. necessarius* in their 16S rRNA gene sequences (Vannini et al., 2007). Another sign of the relatively nascent nature of these relationships is that *P. necessarius* is apparently a substitute for a more ancient symbiont (*Ca. Protistobacter heckmanni*), another representative of the family *Burkholderiaceae*, which can be found in some *Euplotes* isolates (Vannini et al., 2012, 2013).

Obligate intracellular bacterial symbionts demonstrate similar patterns of genome evolution: reduction of its size, decrease in GC content, elevated evolutionary rate, loss of genes from certain

functional groups (transcriptional regulation, DNA repair, etc.), shrinkage of the repertoire of metabolic capabilities, gene transfer to host's nucleus, and others (Moya et al., 2008; Nowack and Melkonian, 2010; McCutcheon and Moran, 2011). At early phase of endosymbiosis these changes are accompanied by the expansion of mobile genetic elements, pseudogenization, and multiple genomic rearrangements (Ochman and Davalos, 2006; Toh et al., 2006; Burke and Moran, 2011). In the case of Strigomonadinae/*Ca. Kinetoplastibacterium*, all the above-mentioned traits typical of ancient endosymbiotic associations can be observed (Alves et al., 2013b). The comparison of genomes of *P. necessarius* and *P. asymbioticus* revealed only a limited genome size reduction (~28% on DNA and ~34% on the protein level) with a substantial pseudogenization (~18%), but without any mobile elements (Meincke et al., 2012; Boscaro et al., 2013).

While the host of *Ca. P. novymonadis* is closely related to that of *Ca. Kinetoplastibacterium* spp., the bacterium itself is phylogenetically closer to *Polynucleobacter*. In order to understand the nature of endosymbiotic relationships, their underlying mechanisms and routes of adaptation in the *Novymonas/Pandoraea* system, we analyzed the genome of *Ca. P. novymonadis* and compared it with both endosymbiotic systems discussed above.

MATERIALS AND METHODS

Establishing Aposymbiotic Strain of *Novymonas esmeraldas*

The strain E262AT.01 of *N. esmeraldas* was cultivated at 27°C in RPMI-1640 medium (Sigma—Aldrich, St. Louis, MO, United States) supplemented with heat-inactivated 10% fetal bovine serum (FBS; Thermo Fisher Scientific, Waltham, MA, United States). At the logarithmic phase of growth, cells from the 10 ml culture aliquots were pelleted by centrifugation at 1,500 × g for 10 min and re-suspended in the fresh RPMI-1640 medium containing 10, 50, 125, 250, or 500 µg/ml of azithromycin (Barry et al., 2004). This macrolide antibiotic was chosen because of its ability to cross eukaryotic plasma membrane, accumulate in the cytoplasm at high concentration, and retain its activity under these conditions (Maurin and Raoult, 2001; Carryn et al., 2003). The presence/absence of bacterial endosymbionts was monitored after 7 and 14 days of incubation by fluorescent *in situ* hybridization with universal bacteria-specific probe Eub338 (5'-GCTGCCTCCCGTAGGAGT-3') labeled with 5'-Cy3 fluorescent dye, as described previously (Kostygov et al., 2016). After 14 days incubation with 10 and 50 µg/ml of azithromycin, all observed *N. esmeraldas* cells were free of endosymbionts, while at the higher concentrations of the antibiotic trypanosomatid cells died. The bacteria-free cultures were pelleted and transferred to a fresh azithromycin-free medium. The strain obtained with 10 µg/ml of azithromycin (hereafter named E262-AZI) displayed better growth and was used for all the subsequent experiments. The absence of bacteria in the culture was also confirmed by PCR with universal eubacterial 16S rRNA primers P1seq and 1486R, with the original bacteria-containing strain (hereafter named E262-wt) used as a positive control (Teixeira et al., 2011).

Given the significant deceleration of E262-AZI growth as compared to the E262-wt, for the subsequent work we switched from RPMI to a more nutrient-rich medium, M199 (Sigma—Aldrich, St. Louis, MO, United States) supplemented with 10% FBS, 2 $\mu\text{g/ml}$ hemin (Jena Bioscience, Jena, Germany), 2 $\mu\text{g/ml}$ bioperin, 100 units/ml of penicillin, and 100 $\mu\text{g/ml}$ of streptomycin (all from Thermo Fisher Scientific, Waltham, MA, United States). In these conditions, E262-AZI was able to propagate at higher rate, comparable to that of E262-wt.

Genomic DNA Isolation and Sequencing

Total genomic DNA was isolated from $\sim 10^9$ cells of the strains E262-wt and E262-AZI of *N. esmeraldas* using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The genome of the wild-type *N. esmeraldas* was sequenced using a combination of Illumina Technologies: HiSeq 2000 (Macrogen Inc., Seoul, South Korea) and MiSeq (Palacký University, Olomouc, Czechia), yielding 47,024,780 reads with 145 \times average coverage and 21,715,370 reads with 160 \times average coverage, respectively. The genome of the aposymbiotic *N. esmeraldas* was sequenced solely with the Illumina MiSeq technology, resulting into 17,834,848 of reads with 136 \times average coverage. The lengths of the obtained paired-end reads were 100 nt for the HiSeq and 300 nt for the MiSeq sequences.

Genome Assembly and Annotation

DNA sequencing reads were processed using BBtools package v.36.02¹. The reads were merged and quality-trimmed using BBmerge with the quality threshold of 20. Non-merged reads were quality-trimmed using BBduk with the same parameters. The quality of raw and trimmed reads was assessed using FASTQC program v.0.11.5².

The genome assembly for both strains was performed using Spades Genome assembler v.3.9.0 with recommended options (Bankevich et al., 2012). Genomic reads of E262-wt were mapped onto the contigs of the aposymbiotic E262-AZI and the remaining reads were used for assembling the endosymbiont genome. However, a read mapping rate was low ($\sim 50\%$) and the obtained assembly contained both endosymbiont and host contigs. Hence, we decided to use other methods for identification of the bacterial contigs. Firstly, each of E262-wt contigs was used as a query in BLAST searches against the custom database composed of *Pandoraea* spp. and trypanosomatid genomes. The BLASTN program from the BLAST package v.2.2.31+ (Camacho et al., 2009) was used with an *E*-value cut-off of 10^{-5} and other settings left as default. The total length of a BLAST alignment per contig was calculated using custom Ruby script. For every contig, the query coverage with *Pandoraea* hits was divided by that with trypanosomatid hits. The values above 1 were considered as evidencing the bacterial origin. Secondly, we checked the absence of the putative endosymbiont contigs in the E262-AZI assembly using the BLASTN program as above. The best hits for the presumed bacterial contigs were

those with low coverage ($\sim 1\times$), probably representing technical contamination during sequencing of the E262-AZI sample. Thirdly, the read coverage was considered for distinguishing contigs of *N. esmeraldas* and *Ca. P. novyimonadis*. Typically, a cell of this trypanosomatid bears several endosymbionts in the cytoplasm (Kostygov et al., 2016), and each of them might have multiple copies of the bacterial genome. Therefore, the read coverage of *Ca. P. novyimonadis* contigs is expected to be higher than that of *N. esmeraldas* contigs. Indeed, the mean coverage per position in the putative bacterial contigs was $\sim 874\times$ while the remaining ones had only $\sim 25\times$ read coverage. In addition, the contigs of different origin could be discriminated by their GC content. Trypanosomatid contigs had $\sim 65\%$ GC, while those of the endosymbiont were only 43–49% GC-rich. This is in agreement with the observation that endosymbiotic genomes usually have lower GC content than the genomes of their hosts (Moran et al., 2008). It should be noted that GC-rich sequences are generally harder to sequence than AT-rich, and this effect may impact the coverage difference and result in overestimation of the bacterial load. Lastly, the Bandage software³, analyzing assembly using a BLAST-based approach, was used. The program created contig graphs, showing that all the putative endosymbiont contigs may compose a single circular chromosome (under assumption that the two shortest bacterial contigs having double coverage, as compared to longer ones are duplicated). This also evidenced that our assembly was complete. Despite the results of the Bandage analysis, we were unable to assemble the bacterial contigs into one chromosome due to some ambiguities. Genome completeness analysis was performed using BUSCO software (Simão et al., 2015) with bacteria, proteobacteria, and betaproteobacteria universal gene datasets and the predicted *Ca. P. novyimonadis* proteins.

Parameters of the genome assemblies were estimated using QUAST v.4.3 (Gurevich et al., 2013). DNA reads were mapped on the assemblies using Bowtie2 v.2.2.9 (Langmead and Salzberg, 2012), with the “–very-fast” option. The structural and functional annotation of the *Ca. P. novyimonadis* genome was obtained using Prokka package v.1.12-beta (Seemann, 2014), signal peptides were predicted using SignalP v.4.1 (Petersen et al., 2011).

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession MUHY00000000. The version described in this paper is version MUHY01000000. The raw reads are available at the NCBI Sequence Read Archive under the accession no. SRR5280512.

Gene Family Inference and Analysis

The inference of protein orthologous groups (OGs) was performed with OrthoFinder v.1.1.3 (Emms and Kelly, 2015) using a dataset of 23 bacterial genomes, including *Ca. P. novyimonadis* sequenced in this study, 13 other *Pandoraea* spp., 5 *Ca. Kinetoplastibacterium* spp., 2 *Polynucleobacter* spp., *Cupriavidus basilensis*, and *Burkholderia cepacia* available in GenBank (Supplementary Table S1). Gene family gains and losses were mapped on the reference species tree using the

¹<https://sourceforge.net/projects/bbtools/>

²<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

³<https://academic.oup.com/bioinformatics/article/31/20/3350/196114/Bandage-interactive-visualization-of-de-novo>

COUNT software with the Dollo parsimony algorithm (Csuros, 2010) as described elsewhere (Flegontov et al., 2016). Using UpSetR package for R⁴ and a custom Python script we found OGs exclusively shared between *Ca. P. novymonadis* and the following groups of species: (i) *C. basilensis* and *B. cepacia*, (ii) *Polynucleobacter* spp., (iii) *Pandoraea* spp., and (iv) *Ca. Kinetoplastibacterium* spp. Putative annotations for the *Ca. P. novymonadis*-specific proteins were inferred using HHpred v.2.0.16 against Pfam-A database and *E*-value cut-off set to 1 (Soding et al., 2005).

Phylogenomic Analysis

In the 16 bacterial strains selected for phylogenetic inference (**Supplementary Table S1**) 556 shared OGs contained only one gene. The amino acid sequences of each single gene were aligned using L-INS-i algorithm in MAFFT v. 7.310 (Katoh and Standley, 2013). The resulting alignments were trimmed in Gblocks v.0.91b with relaxed parameters ($-b3 = 8$, $-b4 = 2$, $-b5 = h$) and then used for phylogenetic reconstruction in IQ-TREE v.1.5.3 with LG + I + G4 + F model and 1,000 ultrafast bootstrap replicates (Minh et al., 2013; Nguyen et al., 2015). The amino acid substitution model had been selected in the same program using the supermatrix concatenated from the individual alignments of all 556 genes (Kalyanamoorthy et al., 2017). To estimate the resolution power of single genes, for each of the reconstructed trees the average bootstrap support was calculated. Setting 70% as a threshold, we selected 119 genes, which constituted the final dataset. The alignments of these genes were concatenated, producing a supermatrix with 54,345 characters. Maximum-likelihood tree was reconstructed using IQ-Tree with LG + I + G4 + F model and 1,000 standard bootstrap replicates. Bayesian inference of phylogeny was performed in MrBayes v. 3.2.6 (Ronquist et al., 2012) under mixed model prior, empirical amino acid frequencies, and heterogeneity of rates across sites assessed using Γ -distribution and proportion of invariant sites. The analysis was run for 100,000 generations with sampling every 10th of them. The chains demonstrated efficient mixing and the two runs converged at the early phase of the analysis (after 2,500 generations). As set by default, 25% samples were discarded as burn-in.

Metabolic Pathways Analysis

For the comparative metabolic study an automatic assignment of KEGG Orthology (KO) identifiers to the proteins of 19 bacterial strains including *Ca. P. novymonadis* (**Supplementary Table S1**) was completed using BlastKOALA v.2.1 (Kanehisa et al., 2016). The search was performed against a non-redundant pangenomic databases of prokaryotes on the genus level and of eukaryotes on the family level. KEGG Mapper v.2.8 was used for the reconstruction of metabolic pathways and their comparison (Kanehisa, 2017).

The search for lipolytic enzymes was performed using BLASTP with an *E*-value of 10^{-20} with the lipase and esterase sequences from the study of Arpigny and Jaeger as a query and annotated proteins of *Ca. P. novymonadis* and other bacteria

as a database (Arpigny and Jaeger, 1999). In the case of *Ca. P. novymonadis* the *E*-value threshold was relaxed to 10^{-10} .

Synteny Analysis

The overall level of synteny in *Ca. P. novymonadis* as compared to other species of interest was studied using the reference dataset of 11 bacteria (**Supplementary Table S1**). Syntenic regions were inferred and visualized using SyMAP v.4.2 (Soderlund et al., 2011). The settings were as follows: minimum number of anchors required to define a synteny block, 7; overlapping (or nearby) synteny blocks were automatically merged into larger blocks, and only the larger block was kept if two synteny blocks overlapped on a chromosome.

Search for Pseudogenes, Phages, and Mobile Elements

Pseudogenes in *Ca. P. novymonadis* genome were identified using BLASTX with an *E*-value cut-off of 1 against the dataset of annotated proteins of *C. basilensis*, *B. cepacia*, and *Pandoraea* spp. (**Supplementary Table S1**). Prior to homology searches, *Ca. P. novymonadis* genes were masked with Maskfasta script from BEDTools package v. 2.25.0 (Quinlan and Hall, 2010). Genomic regions with BLAST hits were manually inspected and the coordinates of the BLAST hits were used for annotation of pseudogenes. We also checked the presence of pseudogenes among the features annotated with Prokka package by analyzing the annotations of the adjacent genes and concluded that all of them were functional.

The search for mobile elements and phages in the genome of *Ca. P. novymonadis* was performed algorithmically in Phispy v. 2.3 (Akhter et al., 2012), as well as using database searches on the online web servers Phaster (Arndt et al., 2016) and IS Finder⁵ using *E*-value cut-off of 10^{-2} .

Analyses of Genome Sequence Properties

Files with the genome sequences and corresponding annotations for the species of interest were downloaded from the NCBI Genome database (12.12.2016). Pseudogene sequences were excluded from further analyses. Lengths of genes and intergenic regions were calculated based on the gene coordinates within GFF files containing annotation data.

For the analysis of GC content, nucleotide sequences of all genes were extracted using Artemis genome browser release v. 16.0.0 (Rutherford et al., 2000). GC content was calculated with Infoseq script from EMBOSS package v. 6.6.0.0 (Rice et al., 2000). Statistical significance of the differences in GC content, lengths of genes, and intergenic regions was tested using one-way analysis of variance (ANOVA) combined with Tukey's honest significance test in R with *p*-value < 0.05.

Nucleotide composition by codon position, amino acid composition, and codon usage bias of protein-coding genes were analyzed using MEGA 7.0 software (Kumar et al., 2016) on the concatenated sequences of all these genes within a genome.

⁴<https://cran.r-project.org/web/packages/UpSetR/>

⁵<http://phaster.ca> and <https://www-is.biotoul.fr>

Standard deviation of relative synonymous codon usage (RSCU) values (Sharp et al., 1986) was calculated as an integral measure of codon usage bias in a particular species. Stop codons and the two amino acids coded by only one codon (methionine and tryptophan) were excluded.

RESULTS AND DISCUSSION

General Characterization of *Ca. P. novymonadis* Genome

The genome of *Ca. P. novymonadis* was assembled into six contigs with a total length of approximately 1.16 Mb (**Supplementary Table S1**), which is smaller than in free-living *Pandoraea* spp. (4.46–6.5 Mb) or in both *Polynucleobacter* spp. (1.56–2.16 Mb), but larger than in *Ca. Kinetoplastibacterium* spp. (~0.8 Mb). The average coverage with the paired-end 100 nt Illumina HiSeq and 300 nt MiSeq reads was ~874× and the largest contig had the length of 844,906 nt. The two shortest contigs (5,920 and 1,318 bp), containing genes for ribosomal RNA, translation factor Tu 1 and tRNAs for alanine, isoleucine, and tryptophan had approximately doubled coverage (1,555× and 1,864×, respectively) pointing to the probable duplication of these fragments in the genome. The assessment of genome assembly and annotation completeness with single-copy orthologs using BUSCO demonstrated that 147/148 (99.3%) universal genes from bacteria dataset, 216/221 (97.7%) from proteobacteria, and 529/582 (90.9%) from betaproteobacteria were present. This indicates that our assembly was complete.

Currently, there are 1,015 annotated genes, 968 of which are protein-coding. For comparison, free-living *Pandoraea* spp. have 3,960–5,342, *Polynucleobacter* spp. – 1,401 and 2,115, while *Ca. Kinetoplastibacterium* spp. only 690–732 protein-coding genes (**Supplementary Table S1**). The number of identified pseudogenes in *Ca. P. novymonadis* (13) is significantly smaller than in other species of the genus *Pandoraea* (76–361) but is comparable to that in *Ca. Kinetoplastibacterium* spp. (2–19) (**Supplementary Table S1**). Interestingly, *P. necessarius* possesses a high number of pseudogenes (269), which is apparently indicative of intense process of genome evolution and is in agreement with a recent origin of endosymbiosis in this species (Vannini et al., 2007).

No mobile elements were found in the genome of *Ca. P. novymonadis* with any of the used tools. This appears to be a consequence of genome minimization. The genome of this species has lost ~80% of its length and protein-coding capacity compared to the genomes of its free-living *Pandoraea* spp. (**Supplementary Table S1**). We did not find statistically significant differences between the lengths of genes and intergenic regions of *Ca. P. novymonadis* compared to other *Pandoraea* spp., *Ca. Kinetoplastibacterium* spp., *Polynucleobacter* spp., *C. basilensis*, and *B. cepacia* (**Supplementary Figure S1**).

The comparison of GC content in *Ca. P. novymonadis* with that of *P. apista*, *P. necessarius*, and *Ca. Kinetoplastibacterium* crithidii genomes revealed significant differences both in genes and intergenic regions between *Ca. P. novymonadis* and the other analyzed species (**Supplementary Figure S2**). Interestingly,

these differences were most pronounced in the genomes of trypanosomatid endosymbionts, *Ca. P. novymonadis*, and *Ca. K. crithidii*. The average GC content of the *Ca. P. novymonadis* genome (43.8%) is intermediate between that of the free-living *Pandoraea* spp. (62–65%) and *Ca. Kinetoplastibacterium* spp. (30–33%). However, it is similar to that of both endosymbiotic and free-living *Polynucleobacter* spp. (45.6 and 44.8, respectively). This pattern is also conspicuous when considering nucleotide composition in protein coding genes by codon position, with the most pronounced differences at the third position (**Supplementary Figure S3**). We found 35 genes in the *Ca. P. novymonadis* genome with the GC content higher than 56% and all these genes encode tRNAs. This is in agreement with an earlier observation that in prokaryotes the GC content of such genes does not correlate with that of the whole genome (Kawai and Maeda, 2009).

The amino acid frequencies in *Ca. P. novymonadis* differ from those in its close relatives. The most discordant ones are for alanine, isoleucine, and lysine (**Supplementary Figure S4**). As with the nucleotide composition, the amino acids frequencies in this species are intermediate between those of other *Pandoraea* spp. and *Ca. Kinetoplastibacterium* spp. and appear most similar to those in *Polynucleobacter* spp.

In agreement with the previously described general trend, the codon usage bias in analyzed species correlated with genomic GC content (Sharp et al., 2005). This relationship was represented by a sideways parabola with the vertex (i.e., lowest value of RSCU standard deviation) situated at about 50% GC: further from the equilibrium nucleotide frequencies, the more pronounced was the bias. Most of the Alcaligenaceae and Burkholderiaceae species fitted this parabolic curve. Three notable exceptions were *Ca. P. novymonadis* (possessing the least prominent codon usage bias) and the two *Polynucleobacter* spp. (**Supplementary Figure S5**). It was previously proposed that species under selection for rapid growth have stronger codon usage bias (Sharp et al., 2005, 2010). However, this is not the case here. In terms of growth rate, the outliers *Ca. P. novymonadis* and *P. necessarius* do not differ much from *Ca. Kinetoplastibacterium* spp. fitting to the trend, since all these bacteria are endosymbionts. An alternative explanation appears to be more plausible: the bacteria that have to switch gene expression from time to time (usually owing to the changing environment) have a stronger bias as compared to those living in stable conditions (Botzman and Margalit, 2011). Although *Ca. Kinetoplastibacterium* spp. are endosymbionts, their close interactions with the host, reflected by a tight coordination of their cell divisions, may lead to similar switches. As for *Ca. P. novymonadis*, its relationship with the host cell seems to be more relaxed (Kostygov et al., 2016) and apparently does not require complex gene expression.

Synteny analysis with free-living *Pandoraea* spp. demonstrated that 62–69% of “anchors” (pairwise alignments) in *Ca. P. novymonadis* genome are located within synteny blocks with maximal values observed for *P. faecigallinarum* and *P. vervacti* (**Supplementary Table S2**). The fact that the majority of the synteny blocks are inverted (15/24 and 11/21 for *P. faecigallinarum* and *P. vervacti*, respectively), reflects a relatively long evolutionary distances between these species

and *Ca. P. novymonadis*. The pairwise synteny between *Ca. P. novymonadis* and the genomes of other *Pandoraea* spp. available in GenBank is presented in **Supplementary Figure S6**. This analysis demonstrated the reduction of the *Ca. P. novymonadis* genome compared to those of free-living *Pandoraea* spp. and a high number of genome rearrangements occurring in the evolution of this endosymbiotic bacterium.

Thus, sequencing and annotation of the *Ca. P. novymonadis* genome revealed several features characteristic for other endosymbiotic bacteria: reduced size, massive gene losses, and decrease in GC content as compared to the genomes of its free-living relatives (Boscaro et al., 2017). Taken together, *Ca. P. novymonadis* is closer to *P. necessarius* than to *Ca. Kinetoplastibacterium* spp.

Phylogenomic Analysis

The maximum likelihood and Bayesian trees inferred using the supermatrix containing 119 genes displayed identical topologies with all branches having maximal bootstrap percentage and posterior probabilities. Previous reconstruction, based on the 16S rRNA gene sequences, placed *Ca. P. novymonadis* in the very crown of the *Pandoraea* clade, though with a low support (Kostygov et al., 2016). However, the results presented here, which are based on much more extensive dataset, demonstrate this species to be an early branch diverged next to *P. thiooxidans* (**Figure 1**). The same position of *Ca. P. novymonadis* could be observed in analyses using either 556 genes supermatrix, or concatenated 16S rRNA and 23S rRNA genes, or a popular bacterial marker, *gyrB* (data not shown). As compared to other *Pandoraea* spp., the species under study demonstrated significantly longer branch (**Figure 1**). This is related to multiple amino acid substitutions in conserved sites and may be explained by fast adaptive evolution of this species. However, in comparison with the outgroups *B. cepacia* and *C. basilensis*, the branch of *Ca. P. novymonadis* does not appear to be uniquely long (**Figure 1**).

Analysis of Protein Orthologous Groups

We performed OrthoFinder analysis on a dataset of 23 annotated bacterial genomes (**Supplementary Table S1**). This resulted in 12,248 OGs, of which 5,437 contained only one protein. Similarly to the *Ca. Kinetoplastibacterium* spp. (Alves et al., 2013b), the genome of *Ca. P. novymonadis* is minimized and the vast majority of genes are single-copy: we found only five OGs containing two proteins with the sequence identity varying from 36 to 96%. These proteins were annotated as ATP-dependent RNA helicase, NADP⁺ reductase, BolA family transcriptional regulator, alanine-tRNA ligase, and threonine synthase. According to our analysis, ATP-dependent RNA helicase and NADP⁺ reductase were also duplicated in the genomes of several *Ca. Kinetoplastibacterium* spp. This situation is drastically different from that observed in the free-living *Pandoraea* spp., which have a substantially higher number of OGs containing two or more genes (e.g., 338 OGs in *P. apista* and 324 in *P. pnomensusa*).

We mapped gene family gains and losses on the phylogenomic tree (**Figure 1**). Gene loss is a predominant trend for all the leaves and most of the nodes within the *Pandoraea* clade. It is especially

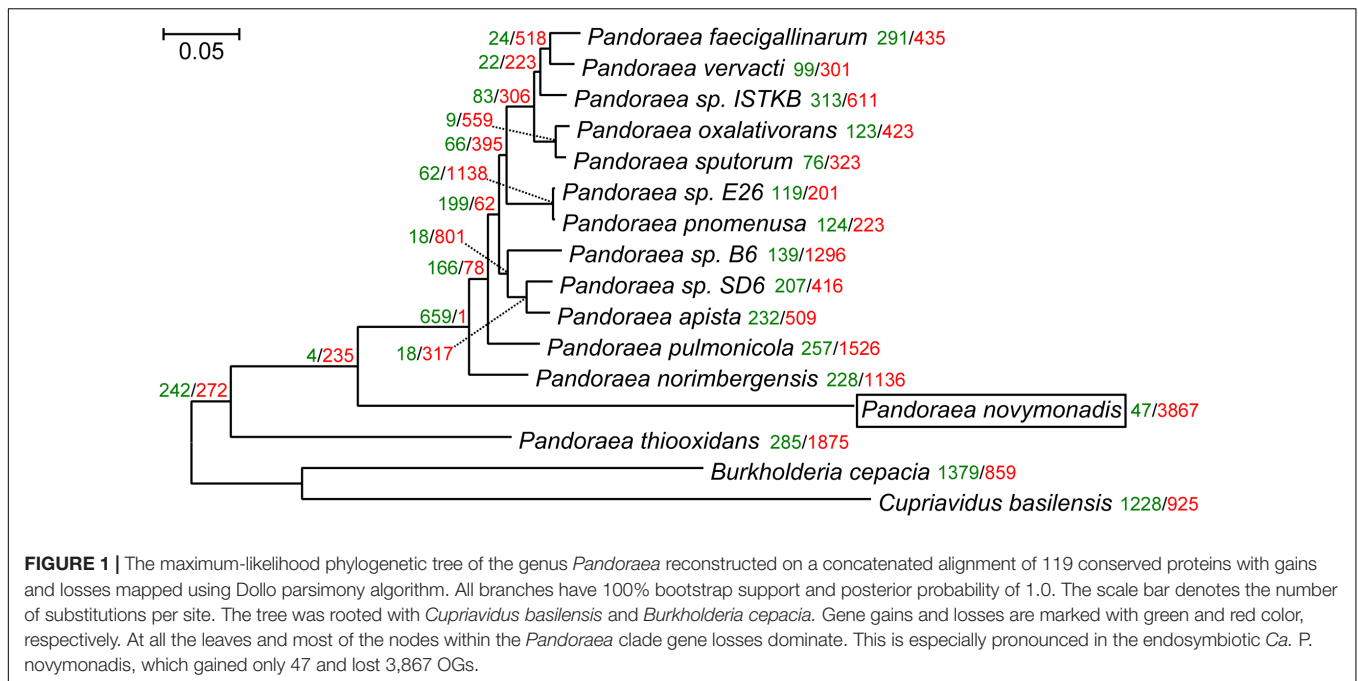
pronounced in the endosymbiotic *Ca. P. novymonadis*, which gained only 47 and lost 3,867 OGs. We used a sensitive HHpred tool attempting to illuminate functions of the proteins within OGs specific for *Ca. P. novymonadis* (Supplementary Table S3). Only 9 out of 47 proteins could be annotated using an *E*-value cut-off of 1. The following putative domains were identified: histidine kinase-like ATPase, cytoplasmic E component of the type III secretion system needle, myristoyl-CoA:protein *N*-myristoyltransferase, and carbohydrate binding domain.

We grouped gene annotations for the 3,867 OGs lost in *Ca. P. novymonadis* according to the KO system. Most of them belong to the following categories: “environmental information processing,” “amino acid metabolism,” “carbohydrate metabolism,” “genetic information processing,” “xenobiotics biodegradation,” and “energy metabolism” (**Supplementary Figure S7**). Out of 3,867 OGs, 1,273 were uniquely lost in *Ca. P. novymonadis*. The composition of functional categories assigned to the proteins within these OGs according to the KO system is similar to that assigned to all 3,867 OGs lost in *Ca. P. novymonadis*. However, the proportion of proteins belonging to the categories “genetic information processing,” “energy metabolism,” and “lipid metabolism” is increased in case of the annotations of OGs uniquely lost in *Ca. P. novymonadis*. The largest portion of OGs lost in *Ca. P. novymonadis* belong to the functional category “environmental information processing,” and more specifically “ATP-binding cassette transporters (ABC transporters).” *Ca. P. novymonadis* has lost many members of this protein family as compared to free-living *Pandoraea* spp.: mineral and organic ion transporters (e.g., for sulfate, nitrate, taurine, molybdate), monosaccharide transporters (e.g., for glycerol-3-phosphate), phosphate and amino acid transporters (e.g., for phosphate, phosphonate, glutamate, aspartate, cystine, urea, D-methionine), and transporters for glutathione and lipooligosaccharides.

Interestingly, there were no OGs uniquely shared between *Ca. P. novymonadis* and either of the endosymbiotic bacterial species investigated here (**Supplementary Figure S8**).

Lipid Metabolism

We identified a full set of enzymes essential for the type-II fatty acid synthesis (FAS) in *Ca. P. novymonadis* and other *Pandoraea* spp., *Ca. Kinetoplastibacterium* spp., *C. basilensis*, *B. cepacia*, and *Polynucleobacter* spp. (Supplementary Table S4). Acetyl-CoA carboxylase, the starting enzyme of the type-II FAS, in bacteria is composed of several polypeptides encoded by four distinct genes: *accA*, *accB*, *accC*, and *accD*. The *accB* and *accC* genes in *Ca. P. novymonadis* are located adjacent to each other and belong to the same operon, similarly to the situation observed in *Escherichia coli* (Janssen and Steinbuchel, 2014). FabF and FabH, 3-ketoacyl-acyl-carrier-protein (ACP) synthases II and III, which catalyze the formation of 3-ketoacyl-ACP by condensation of fatty acyl-ACP with malonyl-ACP, are present, while 3-ketoacyl-ACP synthase II (FabB) is absent in all the analyzed genomes, except for *P. oxalativorans* and *P. vervacti*. FabB participates in the synthesis of unsaturated fatty acids (FAs), catalyzing the condensation of *cis*-3-decenoyl-ACP (formed by the FabA catalyzed reaction), *cis*-5-dodecenoyl-ACP,



and *cis*-7-tetradecenoyl-ACP with malonyl-ACP (Feng and Cronan, 2009). 3-Hydroxydecanoyl-ACP dehydratase/isomerase (FabA), another key player in the synthesis of unsaturated FA is also missing from the analyzed genomes. Interestingly, *C. basilensis* possesses three different enoyl-ACP reductases, catalyzing the last step of the elongation cycle in the synthesis of FA: FabI, FabK, and FabV (Massengo-Tiassé and Cronan, 2009). *Ca. P. novymonadis*, *Ca. Kinetoplastibacterium* spp., and *Polynucleobacter* spp. have only FabI-encoding gene. The majority of the free-living *Pandoraea* spp. retain only FabV, while *B. cepacia*, *P. norimbergensis*, *P. oxalativorans*, *P. pulmonicola*, and *P. thiooxidans* retain FabK along with FabV. The physiological rationale for the presence of multiple enoyl-ACP reductases is poorly understood (Zhu et al., 2013).

All *Pandoraea* spp., *Polynucleobacter* spp., *C. basilensis*, and *B. cepacia* are able to synthesize cardiolipin, phosphatidylethanolamine, and phosphatidyl-L-serine, important components of the bacterial membranes (Supplementary Table S5). In all bacteria analyzed, the end product of the FA biosynthesis, acyl-ACP, can be activated with an inorganic phosphate group by the action of the PlsX component of the PlsX/PlsY/PlsC acyltransferase system, leading to acyl-phosphate, which is subsequently added to glycerol-3-phosphate by the action of the PlsY component (Janssen and Steinbuchel, 2014). The next steps to synthesize diacylglycerol-3-phosphate and cytosine diphosphate diacylglycerol (CDP-diacylglycerol) are performed by 1-acyl-sn-glycerol-3-phosphate acyltransferase (PlsC) and phosphatidate cytidyltransferase (CdsA). CDP-diacylglycerol is the intermediate which is then used for the formation of cardiolipin, phosphatidyl-L-serine, and phosphatidylethanolamine by cardiolipin synthase, CDP-diacylglycerol-serine O-phosphatidyltransferase, and

phosphatidylserine decarboxylase, respectively. All *Ca. Kinetoplastibacterium* spp. lack the capacity to synthesize cardiolipin, while *Ca. K. galatii*, *Ca. K. oncopeltii*, and *Ca. K. blastocrithidii* are not able to produce any of the membrane lipids mentioned above.

Interestingly, no lipases and esterases could be detected in the genome of *Ca. P. novymonadis* even with the *E*-value cut-off of 10^{-10} . We found proteins belonging to the family VI of bacterial lipolytic enzymes in all *Ca. Kinetoplastibacterium* spp. and in *P. necessarius* (Arpigny and Jaeger, 1999). The lipases and esterases belonging to the families I, IV, V, and VI are readily identifiable in the genomes of *C. basilensis* and *B. cepacia*, as well in the free-living *Pandoraea* spp., which in addition possess proteins belonging to the family VII of the lipolytic enzymes.

Importantly, all endosymbionts of trypanosomatids, including *Ca. P. novymonadis*, are unable to oxidize FAs since all the enzymes required for β -oxidation are missing, similarly to the situation observed in bacterial endosymbionts of insects (Zientz et al., 2004).

Carbon Metabolism

All species analyzed in this work preserve enzymes for glycolysis and the central (non-oxidative) part of the pentose phosphate pathway (Supplementary Figure S9). However, only the free-living *Pandoraea* spp. have hexokinase and, thus, are able to utilize glucose. In contrast to the endosymbiotic bacteria, they also can use classic and alternative (i.e., non-phosphorylated) variants of the Entner–Doudoroff pathway. Interestingly, only *P. thiooxidans* possesses phosphofructokinase converting fructose-6-phosphate into fructose 1,6-bisphosphate. Other species must use a bypass through the pentose phosphate pathway for hexose catabolism. Fructose 1,6-bisphosphatase, the

enzyme catalyzing the reverse reaction, is present in all studied species suggesting its importance for anabolic processes, in particular, gluconeogenesis.

We were unable to trace the carbon source that *Ca. Kinetoplastibacterium* spp. utilize instead of glucose. However, for *Ca. P. novyimonadis* and *P. necessarius* this appears to be fructose. Similarly to the situation with glucose, there is no typical phosphorylating enzyme, i.e., fructokinase (it is also absent from all other *Pandoraea* spp.). In all these species we identified the three cytoplasmic components of phosphotransferase system (PTS), namely phosphoenolpyruvate (PEP)-protein phosphotransferase (PTS-EI), histidine phosphocarrier protein (HPr), and PTS system fructose-specific EIIA component (PTS-EIIA^{Fru}). The main function of PTS is a concomitant transfer of sugars inside the cell and their phosphorylation (Saier, 2015). In addition to the three proteins mentioned above, the fully functional PTS must also contain juxtamembrane permease PTS-EIIB and transmembrane PTS-EIIC (sometimes along with PTS-EIID). The phosphate from PEP is successively transferred to PTS-EI, then to HPr, PTS-EIIA, PTS-EIIB, and then to sugar (Saier, 2015). Numerous proteobacteria possess incomplete PTS-lacking EIIB and EIIC components. Such PTSs were proposed to have only regulatory functions (Deutscher et al., 2014). We hypothesize that the incomplete fructose-specific PTS may be used for phosphorylation of fructose. Indeed, in addition to the abovementioned lack of fructokinase, *Ca. P. novyimonadis* also does not have pyruvate kinase, the key enzyme for the production of ATP from PEP at the end of glycolysis. Meanwhile, PTS using PEP as a phosphate donor could substitute this missing link. The lack of hexokinase and fructokinase along with the presence of PTS was also documented in obligate intracellular bacteria of insects (Zientz et al., 2004).

The complete tricarboxylic acid (TCA) cycle is present in all considered bacteria except *Ca. Kinetoplastibacterium* spp., which possess enzymes for two consecutive steps of this cycle: transformation of 2-oxoglutarate to succinyl-CoA and then to succinate. These steps may be preserved because succinyl-CoA is required for lysine biosynthesis. In addition, these bacteria possess malate dehydrogenase interconverting malate and oxaloacetate.

In addition to the TCA cycle, the free-living *Pandoraea* spp. also have the complete glyoxylate pathway, enabling usage of short-chain compounds as a carbon source. Endosymbiotic bacteria in their stable environment do not need such capability. Intriguingly, *P. necessarius* has malate synthase interconverting glyoxylate and malate, whereas other enzymes of this cycle are absent from its genome.

Amino Acid Metabolism

The free-living *Pandoraea* spp. are able to synthesize all 20 amino acids. Meanwhile, the three groups of endosymbionts considered here (*Ca. P. novyimonadis*, *P. necessarius*, and *Kinetoplastibacterium* spp.) demonstrate different phases of gradual loss of those capabilities (Figure 2 and Supplementary Table S6). This process starts with the loss of the pathways for the synthesis of the non-essential amino acids such as alanine, asparagine, and aspartate, a situation observed in

the evolutionary young endosymbiont *P. necessarius*. *Ca. P. novyimonadis* is unable to synthesize three additional amino acids: cysteine, methionine, and proline. *Ca. Kinetoplastibacterium* spp. exhibit the most advanced state, lacking enzymes for the synthesis of 13 amino acids (Figure 2). As judged from previous studies, the metabolic pathways of these endosymbionts and their hosts are interlaced and, for most of the amino acids, the enzymes missing in the bacterium can be substituted by those of the trypanosomatid (Alves et al., 2013a; Alves, 2017). Although the metabolism of *N. esmeraldas* has not been studied yet, it is likely similar to that of its relatives – trypanosomatids of the subfamily Leishmaniinae. This group of flagellates is auxotrophic for arginine, histidine, isoleucine, leucine, phenylalanine, serine, tryptophan, tyrosine, and valine (Opperdoes et al., 2016). Therefore, it is not surprising that *Ca. P. novyimonadis* retained the ability to synthesize them. In return, *N. esmeraldas* may provide the six amino acids, which its symbiont is unable to produce.

In addition to losing the ability to synthesize particular amino acids, the endosymbionts are devoid of some biochemical bypasses. Thus, phenylalanine-4-hydroxylase, converting phenylalanine to tyrosine, is present in free-living *Pandoraea* spp., but absent in all endosymbionts analyzed here. The same concerns arginase, the enzyme transforming arginine to ornithine (Figure 2).

Histidinol-phosphate phosphatase (HPPase), responsible for the penultimate step of histidine biosynthesis, was not found by BlastKOALA in any of the analyzed genomes. Nevertheless, HPPases are present in GenBank genome annotations of all free-living *Pandoraea* spp. Homologous proteins in *Polynucleobacter* spp. and *Ca. Kinetoplastibacterium* spp. are annotated as inositol monophosphatases. The same result was obtained for *Ca. P. novyimonadis* in Prokka annotation. It is known that inositol-monophosphatase-like enzymes may exhibit histidinol-phosphatase activity (Mormann et al., 2006; Petersen et al., 2010; Nourbakhsh et al., 2014). Of note, none of the bacteria analyzed here has other enzymes of inositol metabolism, so it is unlikely that the protein in question is an inositol monophosphatase. Thus, we argue that all analyzed species possess divergent histidinol-phosphatases.

Urea Cycle/Polyamine Synthesis

All free-living *Pandoraea* spp. have complete set of enzymes for the urea cycle and synthesis of important polyamines. *Ca. P. novyimonadis* and *P. necessarius* lack arginase, while preserving ornithine carbamoyltransferase, argininosuccinate synthase, and argininosuccinate lyase (Figure 3). They also possess arginine decarboxylase converting arginine to agmatine, the first intermediate in the synthesis of polyamines, but for the rest of this pathway these bacteria apparently rely on their respective hosts. *Ca. Kinetoplastibacterium* spp. showed the most reduced state with only two enzymes remaining in their arsenal: carbamoyltransferase and arginine decarboxylase (Figure 3).

Vitamins and Cofactors

All bacteria analyzed here are able to synthesize a number of porphyrins, including heme, an essential compound for

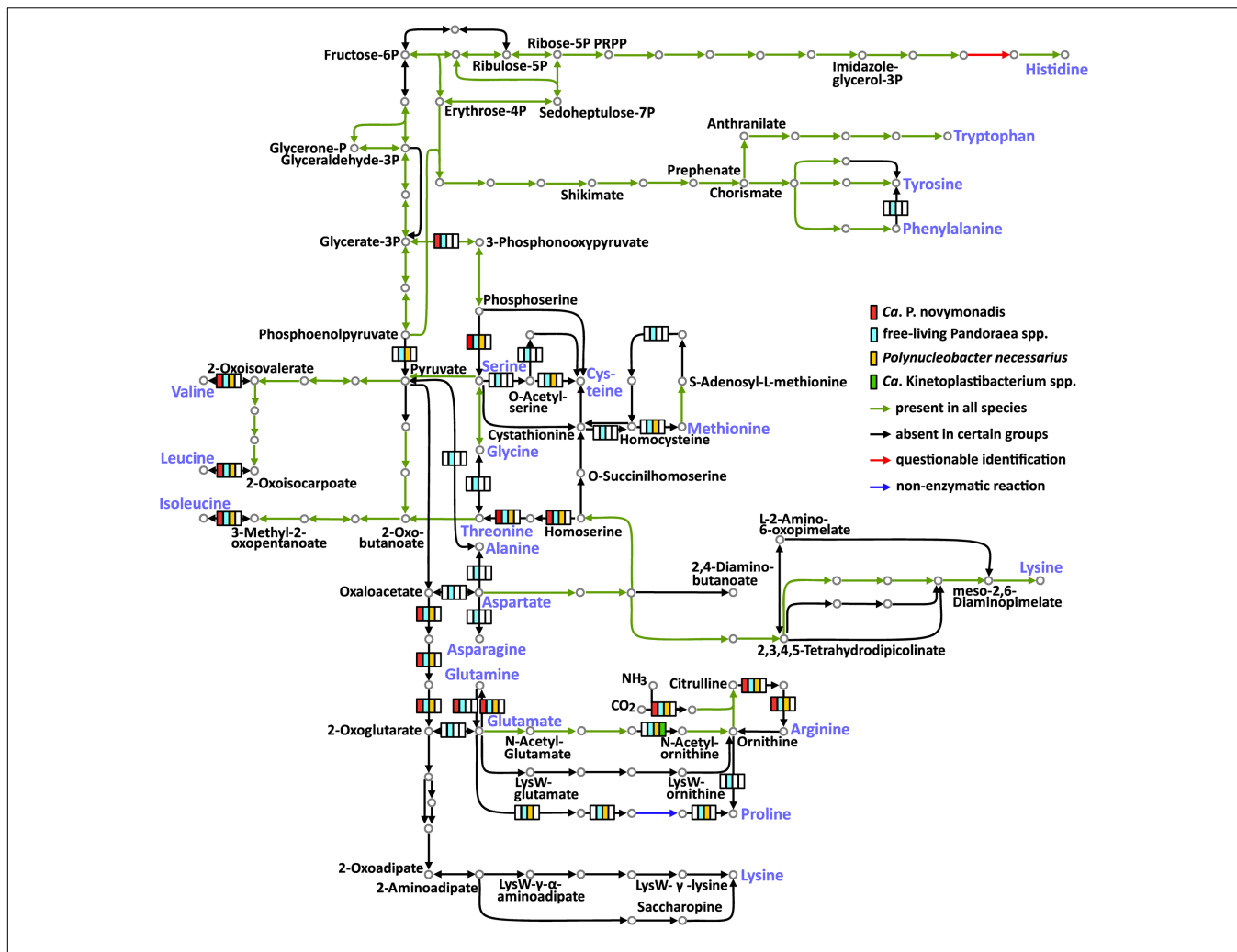


FIGURE 2 | Amino acid metabolism in *Ca. P. novymonadis*, free-living *Pandoraea* spp., *Polynucleobacter necessarius*, and *Ca. Kinetoplastibacterium* spp. The 20 standard amino acids are shown in violet.

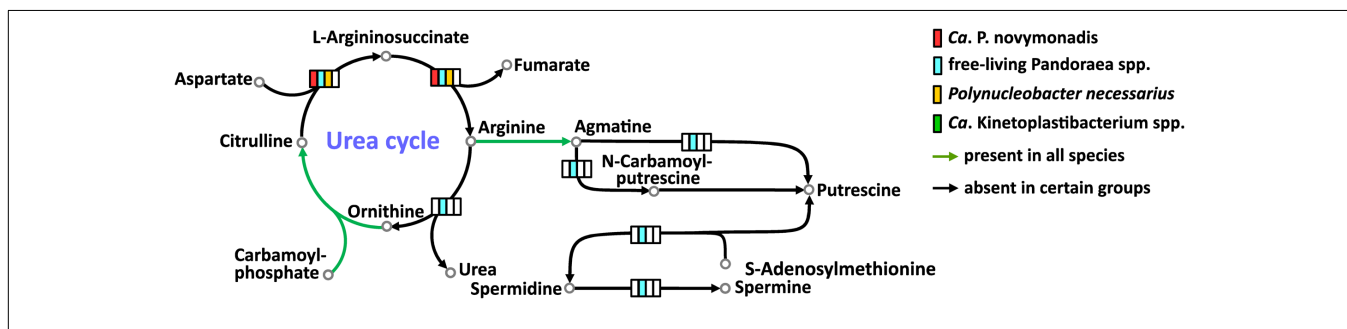


FIGURE 3 | Urea cycle and polyamine synthesis in *Ca. P. novymonadis*, free-living *Pandoraea* spp., *Polynucleobacter necessarius*, and *Ca. Kinetoplastibacterium* spp.

most trypanosomatids (Kořený et al., 2012). The free-living *Pandoraea* spp., *Ca. P. novymonadis* and *P. necessarius*, are prototrophic for all vitamins. As for *Ca. Kinetoplastibacterium* spp., their metabolism of vitamins was previously scrutinized by others (Klein et al., 2013). It has been demonstrated that in

contrast to the rest of bacteria considered here, they are unable to synthesize thiamine, nicotinic acid, and biotin, which are apparently acquired by the trypanosomatid host from the insect's gut content. All enzymes needed to produce folic acid, vitamin B6, and riboflavin essential for the trypanosomatid host are

encoded in the genomes of *Ca. Kinetoplastibacterium* spp., but the pathway of pantothenic acid biosynthesis is interrupted at the very end (Klein et al., 2013). The missing enzyme (ketopantoate reductase) is encoded in the genome of the trypanosomatid host, thus representing an example of deep integration of metabolic pathways in this symbiotic association.

CONCLUSION

Here, we sequenced and analyzed the genome of *Ca. P. novymonadis*, the bacterial endosymbiont of the trypanosomatid *N. esmeraldas*. To better understand the evolution and biology of this bacterium, we compared its genome to those of related prokaryotes, namely the free-living *Pandoraea* spp., two sister *Polynucleobacter* spp., from which one is free-living and the other is endosymbiotic, as well as with *Ca. Kinetoplastibacterium* spp., which are the only other known endosymbionts of trypanosomatids. The genome of *Ca. P. novymonadis* revealed all hallmarks of an endosymbiont genome: size reduction, massive gene losses, decreased GC content, and lowered codon usage bias. At the same time, this genome preserves main metabolic pathways, including biosynthesis of vitamins and heme, essential for the trypanosomatid host. The bacterium does not produce some amino acids, which are likely provided by the host, but retains the ability to synthesize those, for which the trypanosomatid is auxotrophic.

Our data allow first comparative analysis of the endosymbionts of trypanosomatids and strongly indicate that their evolution followed different scenarios, reflected by the fact that they do not have uniquely shared traits. Importantly, from the perspective of both its general genomic features and metabolism, *Ca. P. novymonadis* is closer to the ciliate-dwelling *P. necessarius*, which belongs to the same family Burkholderiaceae, than to *Ca. Kinetoplastibacterium* spp., the only other known endosymbionts of trypanosomatids.

Previously, we proposed that the endosymbiosis between *Ca. P. novymonadis* and *N. esmeraldas* was established relatively recently (Kostygov et al., 2016). This opinion was based on the phylogenetic position of the bacterium and seemingly unsophisticated relationships in this symbiotic association. However, the phylogenomic analysis presented here demonstrates that the endosymbiont diverged earlier than as inferred from its 16S rRNA gene. As judged from its genomic characteristics, *Ca. P. novymonadis* has already passed the intensive period of host adaptation, which can still be observed in *P. necessarius*, the best candidate for a recent endosymbiosis. As judged by the extent of the overall genome reduction and the loss of particular metabolic abilities correlating with the increasing dependence of the symbiont on its host, *Ca. P. novymonadis* occupies an intermediate position *P. necessarius* and *Ca. Kinetoplastibacterium* spp. We conclude that the relationship between *Ca. P. novymonadis* and *N. esmeraldas* is already well-established, although not as fine-tuned as in the case of related flagellates of the family Strigomonadinae and their endosymbionts.

AUTHOR CONTRIBUTIONS

VY and JL jointly conceived the study. AK and AB contributed equally to this work: participated in the design of the study, the analysis and interpretation of data, and the manuscript writing. AN and PF conducted genome assembly and curated annotation, and contributed to the interpretation of data. DT established and analyzed the aposymbiotic strain of *N. esmeraldas*. VY, AK, and JL revised and corrected the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the Grant Agency of Czech Republic awards 17-10656S to VY, 16-18699S to JL and VY, Moravskoslezský kraj research initiative DT01-021358 to VY and AK, the COST action CM1307, and the European Research Council CZ LL1601 to JL. Work in VY lab is financially supported by the Ministry of Education, Youth and Sports of the Czech in the “National Feasibility Program I,” project LO1208 “TEWEP.” AB was funded by grant from the University of Ostrava SGS16/PRF/2017. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.01940/full#supplementary-material>

FIGURE S1 | Graphs showing length distribution for genomic features in *Ca. P. novymonadis* and other bacteria with various lifestyles. (A,C) Distribution of gene lengths; (B,D) distribution of the lengths of intergenic regions. Pnov, *Ca. P. novymonadis*; Bcep, *B. cepacia*; Cbas, *C. basiliensis*; Papi, *Pandoraea apista*; Pfae, *Pandoraea faecigallinarum*; Pnor, *Pandoraea norimbergensis*; Poxa, *Pandoraea oxalativorans*; Ppno, *Pandoraea pnomenusa*; Ppul, *Pandoraea pulmonicola*; Pspu, *Pandoraea sputorum*; Pthi, *Pandoraea thiooxidans*; Pver, *Pandoraea vervacti*; Kbla, *Ca. Kinetoplastibacterium blastocritthidii*; Kcri, *Ca. Kinetoplastibacterium crithidii*; Kdes, *Ca. Kinetoplastibacterium desouzai*; Kgal, *Ca. Kinetoplastibacterium galatii*; Konc, *Ca. Kinetoplastibacterium oncopeltii*; Pasy, *Polynucleobacter asymbioticus*; and Pnec, *Polynucleobacter necessarius*.

FIGURE S2 | Graphs showing GC content distribution for genomic features in *Ca. P. novymonadis* and other bacteria with various lifestyles: solid lines for genes and dotted lines for intergenic regions. Abbreviations of bacterial species names are as in **Supplementary Figure S1**; “ir” stands for intergenic regions.

FIGURE S3 | Nucleotide composition of protein-coding genes by codon positions. Abbreviations of bacterial species names are as in **Supplementary Figure S1**; nucleotides in a particular codon position are denoted below the figure.

FIGURE S4 | Amino acid frequencies. Abbreviations of bacterial species names are as in **Supplementary Figure S1**.

FIGURE S5 | Relationship of codon usage bias and GC content in protein coding genes averaged over the whole genome. The dotted line represents a sideways parabola fitting the distribution of values. Akas, *Advenella kashmirensis*; Tequ, *Taylorella equigenitalis*; Axyl, *Achromobacter xylosoxidans*;

Aspa, *Achromobacter spanius*; Ebro, *Bordetella bronchiseptica*; Bhol, *Bordetella holmesii*; Pnoe, *Pusillimonas noertemanni*; Pind, *Pelistega indica*; Oure, *Oligella urethralis*; Lasp, *Limnobacter* sp. CACIAM 66H1; Ggig, *Ca. Glomeribacter gigasporarum*; Csor, *Caballeronia sordidicola*. All other abbreviations of bacterial species names are as in **Supplementary Figure S1**.

FIGURE S6 | Schematic representation of two-way synteny between *Ca. P. novymonadis* and other *Pandoraea* spp. with sequenced genomes. The four longest *Ca. P. novymonadis* contigs are colored according to the legend. Only scaffolds with synteny blocks are shown. Direct synteny blocks are displayed in red, inverted ones – in green. The contigs are drawn proportionately to their actual length. The genomes of *Pandoraea* spp. shown on the figure are fully assembled to the level of circular chromosomes depicted as the longest colored bars. For some species the shorter colored bars representing plasmids are shown in addition to the chromosomal scaffolds.

REFERENCES

- Akhter, S., Aziz, R. K., and Edwards, R. A. (2012). PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res.* 40, e126. doi: 10.1093/nar/gks406
- Alves, J. M., Klein, C. C., da Silva, F. M., Costa-Martins, A. G., Serrano, M. G., Buck, G. A., et al. (2013a). Endosymbiosis in trypanosomatids: the genomic cooperation between bacterium and host in the synthesis of essential amino acids is heavily influenced by multiple horizontal gene transfers. *BMC Evol. Biol.* 13:190. doi: 10.1186/1471-2148-13-190
- Alves, J. M., Serrano, M. G., Maia, da Silva, F., Voegtly, L. J., Matveyev, A. V., et al. (2013b). Genome evolution and phylogenomic analysis of *Candidatus* Kinetoplastibacterium, the betaproteobacterial endosymbionts of *Strigomonas* and *Angomonas*. *Genome Biol. Evol.* 5, 338–350. doi: 10.1093/gbe/evt012
- Alves, J. M., Voegtly, L., Matveyev, A. V., Lara, A. M., da Silva, F. M., Serrano, M. G., et al. (2011). Identification and phylogenetic analysis of heme synthesis genes in trypanosomatids and their bacterial endosymbionts. *PLOS ONE* 6:e23518. doi: 10.1371/journal.pone.0023518
- Alves, J. M. (2017). “Amino acid biosynthesis in endosymbiont-harboring Trypanosomatidae,” in *The Handbook of Microbial Metabolism of Amino Acids*, ed. J. P. F. D’Mello (Oxfordshire: CAB International), 371–383.
- Anandham, R., Indiragandhi, P., Kwon, S. W., Sa, T. M., Jeon, C. O., Kim, Y. K., et al. (2010). *Pandoraea thiooxydans* sp. nov., a facultatively chemolithotrophic, thiosulfate-oxidizing bacterium isolated from rhizosphere soils of sesame (*Sesamum indicum* L.). *Int. J. Syst. Evol. Microbiol.* 60(Pt 1), 21–26. doi: 10.1099/ijs.0.012823-0
- Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387
- Arpigny, J. L., and Jaeger, K. E. (1999). Bacterial lipolytic enzymes: classification and properties. *Biochem. J.* 343(Pt 1), 177–183. doi: 10.1042/bj3430177
- 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. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Barry, A., Bryskier, A., Traczewski, M., and Brown, S. (2004). Preparation of stock solutions of macrolide and ketolide compounds for antimicrobial susceptibility tests. *Clin. Microbiol. Infect.* 10, 78–83. doi: 10.1111/j.1469-0691.2004.00759.x
- Boscaro, V., Felletti, M., Vannini, C., Ackerman, M. S., Chain, P. S., Malfatti, S., et al. (2013). *Polynucleobacter necessarius*, a model for genome reduction in both free-living and symbiotic bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 110, 18590–18595. doi: 10.1073/pnas.1316687110
- Boscaro, V., Kolisko, 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
- Botzman, M., and Margalit, H. (2011). Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles. *Genome Biol.* 12:R109. doi: 10.1186/gb-2011-12-10-r109
- FIGURE S7** | A pie chart showing gene annotations for 3,867 OGs lost in *Ca. P. novymonadis* (Pnov) categorized according to the KEGG Orthology system. *Pandoraea pnomenus* (Ppno), *Pandoraea norimbergensis* (Pnor), and *Pandoraea vervacti* (Pver) are shown for comparison.
- FIGURE S8** | Analysis of OGs sharing between free-living *Pandoraea* spp., *Ca. Pandoraea novymonadis*, *Ca. Kinetoplastibacterium* spp., *Polynucleobacter* spp., *Burkholderia cepacia*, and *Cupriavidus basilensis*. OGs were categorized according to their presence in the analyzed species. Depicted bars indicate number of OGs that are unique or shared among the genomes of the organisms listed, as indicated by the black dots.
- FIGURE S9** | Carbon metabolism in *Ca. P. novymonadis*, free-living *Pandoraea* spp., *Polynucleobacter necessarius*, and *Ca. Kinetoplastibacterium* spp.
- TABLE S1** | Genomic characteristics of species used in analyses.
- Burke, G. R., and Moran, N. A. (2011). Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol. Evol.* 3, 195–208. doi: 10.1093/gbe/evr002
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Carrin, S., Chanteux, H., Seral, C., Mingeot-Leclercq, M. P., Van Bambeke, F., and Tulkens, P. M. (2003). Intracellular pharmacodynamics of antibiotics. *Infect. Dis. Clin. North Am.* 17, 615–634. doi: 10.1016/S0891-5520(03)00066-7
- Catta-Preta, C. M., Brum, F. L., da Silva, C. C., Zuma, A. A., Elias, M. C., de Souza, W., et al. (2015). Endosymbiosis in trypanosomatid protozoa: the bacterium division is controlled during the host cell cycle. *Front. Microbiol.* 6:520. doi: 10.3389/fmicb.2015.00520
- Coenye, T., Falsen, E., Hoste, B., Ohlen, M., Goris, J., Govan, J. R., et al. (2000). Description of *Pandoraea* gen. nov. with *Pandoraea apista* sp. nov., *Pandoraea pulmonicola* sp. nov., *Pandoraea pnomenus* sp. nov., *Pandoraea sputorum* sp. nov. and *Pandoraea norimbergensis* comb. nov. *Int. J. Syst. Evol. Microbiol.* 50(Pt 2), 887–899. doi: 10.1099/00207713-50-2-887
- Csuros, M. (2010). Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* 26, 1910–1912. doi: 10.1093/bioinformatics/btq315
- de Souza, W., and Motta, M. C. (1999). Endosymbiosis in protozoa of the Trypanosomatidae family. *FEMS Microbiol. Lett.* 173, 1–8. doi: 10.1111/j.1574-6968.1999.tb13477.x
- Deutscher, J., Ake, F. M., Derkaoui, M., Zebre, A. C., Cao, T. N., Bouraoui, H., et al. (2014). The bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system: regulation by protein phosphorylation and phosphorylation-dependent protein-protein interactions. *Microbiol. Mol. Biol. Rev.* 78, 231–256. doi: 10.1128/MMBR.00001-14
- Du, Y., Maslov, D. A., and Chang, K. P. (1994). Monophyletic origin of beta-division proteobacterial endosymbionts and their coevolution with insect trypanosomatid protozoa *Blastocrithidia culicis* and *Crithidia* spp. *Proc. Natl. Acad. Sci. U.S.A.* 91, 8437–8441. doi: 10.1073/pnas.91.18.8437
- Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 157. doi: 10.1186/s13059-015-0721-2
- Feng, Y., and Cronan, J. E. (2009). *Escherichia coli* unsaturated fatty acid synthesis: complex transcription of the *fabA* gene and *in vivo* identification of the essential reaction catalyzed by FabB. *J. Biol. Chem.* 284, 29526–29535. doi: 10.1074/jbc.M109.023440
- Flegontov, P., Butenko, A., Firsov, S., Kraeva, N., Eliáš, M., Field, M. C., et al. (2016). Genome of *Leptomonas pyrrocoris*: a high-quality reference for monoxenous trypanosomatids and new insights into evolution of *Leishmania*. *Sci. Rep.* 6:23704. doi: 10.1038/srep23704
- Graff, A., and Stubner, S. (2003). Isolation and molecular characterization of thiosulfate-oxidizing bacteria from an Italian rice field soil. *Syst. Appl. Microbiol.* 26, 445–452. doi: 10.1078/072320203322497482
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

- Janssen, H. J., and Steinbuchel, A. (2014). Fatty acid synthesis in *Escherichia coli* and its applications towards the production of fatty acid based biofuels. *Biotechnol. Biofuels* 7:7. doi: 10.1186/1754-6834-7-7
- Jeong, S. E., Lee, H. J., Jia, B., and Jeon, C. O. (2016). *Pandoraea terrae* sp. nov., isolated from forest soil, and emended description of the genus *Pandoraea* Coenye et al., 2000. *Int. J. Syst. Evol. Microbiol.* 66, 3524–3530. doi: 10.1099/ijsem.0.001229
- Kalyanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Kanehisa, M. (2017). Enzyme annotation and metabolic reconstruction using KEGG. *Methods Mol. Biol.* 1611, 135–145. doi: 10.1007/978-1-4939-7015-5_11
- Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kawai, Y., and Maeda, Y. (2009). GC-content of tRNA genes classifies archaea into two groups. *J. Gen. Appl. Microbiol.* 55, 403–408. doi: 10.2323/jgam.55.403
- Klein, C. C., Alves, J. M., Serrano, M. G., Buck, G. A., Vasconcelos, A. T., Sagot, M. F., et al. (2013). Biosynthesis of vitamins and cofactors in bacterium-harbouring trypanosomatids depends on the symbiotic association as revealed by genomic analyses. *PLOS ONE* 8:e79786. doi: 10.1371/journal.pone.0079786
- Kořený, L., Sobotka, R., Kovařová, J., Gnypová, A., Flegontov, P., Horváth, A., et al. (2012). Aerobic kinetoplastid flagellate *Phytomonas* does not require heme for viability. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3808–3813. doi: 10.1073/pnas.1201089109
- Kostygov, A. Y., Dobaková, E., Grybchuk-Ieremenko, A., Váhala, D., Maslov, D. A., Votýpka, J., et al. (2016). Novel trypanosomatid-bacterium association: evolution of endosymbiosis in action. *mBio* 7:e01985-15. doi: 10.1128/mBio.01985-15
- Kumar, M., Singh, J., Singh, M. K., Singhal, A., and Thakur, I. S. (2015). Investigating the degradation process of kraft lignin by beta-proteobacterium, *Pandoraea* sp. ISTKB. *Environ. Sci. Pollut. Res. Int.* 22, 15690–15702. doi: 10.1007/s11356-015-4771-5
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Liz, J. A. Z. E., Jan-Roblero, J., de la Serna, J. Z. D., de Leon, A. V. P., and Hernandez-Rodriguez, C. (2009). Degradation of polychlorinated biphenyl (PCB) by a consortium obtained from a contaminated soil composed of *Brevibacterium*, *Pandoraea* and *Ochrobactrum*. *World J. Microbiol. Biotechnol.* 25, 165–170. doi: 10.1007/s11274-008-9875-3
- Massengo-Tiassé, R. P., and Cronan, J. E. (2009). Diversity in enoyl-acyl carrier protein reductases. *Cell. Mol. Life Sci.* 66, 1507–1517. doi: 10.1007/s00018-009-8704-7
- Maurin, M., and Raoult, D. (2001). Use of aminoglycosides in treatment of infections due to intracellular bacteria. *Antimicrob. Agents Chemother.* 45, 2977–2986. doi: 10.1128/AAC.45.11.2977-2986.2001
- McCutcheon, J. P., and Moran, N. A. (2011). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10, 13–26. doi: 10.1038/nrmicro2670
- Meincke, L., Copeland, A., Lapidus, A., Lucas, S., Berry, K. W., Del Rio, T. G., et al. (2012). Complete genome sequence of *Polynucleobacter necessarius* subsp. *asymbioticus* type strain (QLW-PIDMWA-1^T). *Stand. Genomic Sci.* 6, 74–83. doi: 10.4056/signs.2395367
- Minh, B. Q., Nguyen, M. A., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. doi: 10.1093/molbev/mst024
- Moran, N. A., McCutcheon, J. P., and Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* 42, 165–190. doi: 10.1146/annurev.genet.41.110306.130119
- Mormann, S., Lomker, A., Ruckert, C., Gaigalat, L., Tauch, A., Puhler, A., et al. (2006). Random mutagenesis in *Corynebacterium glutamicum* ATCC 13032 using an IS6100-based transposon vector identified the last unknown gene in the histidine biosynthesis pathway. *BMC Genomics* 7:205. doi: 10.1186/1471-2164-7-205
- Motta, M. C., Catta-Preta, C. M., Schenkman, S., de Azevedo Martins, A. C., Miranda, K., de Souza, W., et al. (2010). The bacterium endosymbiont of *Crithidia deanei* undergoes coordinated division with the host cell nucleus. *PLOS ONE* 5:e12415. doi: 10.1371/journal.pone.0012415
- Motta, M. C., Soares, M. J., Attias, M., Morgado, J., Lemos, A. P., Saad-Nehme, J., et al. (1997). Ultrastructural and biochemical analysis of the relationship of *Crithidia deanei* with its endosymbiont. *Eur. J. Cell Biol.* 72, 370–377.
- Moya, A., Pereto, J., Gil, R., and Latorre, A. (2008). Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat. Rev. Genet.* 9, 218–229. doi: 10.1038/nrg2319
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Nourbakhsh, A., Collakova, E., and Gillaspay, G. E. (2014). Characterization of the inositol monophosphatase gene family in *Arabidopsis*. *Front. Plant Sci.* 5:725. doi: 10.3389/fpls.2014.00725
- Nowack, E. C., and Melkonian, M. (2010). Endosymbiotic associations within protists. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 699–712. doi: 10.1098/rstb.2009.0188
- Ochman, H., and Davalos, L. M. (2006). The nature and dynamics of bacterial genomes. *Science* 311, 1730–1733. doi: 10.1126/science.1119966
- Okeke, B. C., Siddique, T., Arbestain, M. C., and Frankenberger, W. T. (2002). Biodegradation of gamma-hexachlorocyclohexane (lindane) and alpha-hexachlorocyclohexane in water and a soil slurry by a *Pandoraea* species. *J. Agric. Food Chem.* 50, 2548–2555. doi: 10.1021/jf011422a
- Oppendoes, F. R., Butenko, A., Flegontov, P., Yurchenko, V., and Lukes, J. (2016). Comparative metabolism of free-living *Bodo saltans* and parasitic trypanosomatids. *J. Eukaryot. Microbiol.* 63, 657–678. doi: 10.1111/jeu.12315
- Ozaki, S., Kishimoto, N., and Fujita, T. (2007). Change in the predominant bacteria in a microbial consortium cultured on media containing aromatic and saturated hydrocarbons as the sole carbon source. *Microbes Environ.* 22, 128–135. doi: 10.1264/jsm.2.22.128
- Petersen, L. N., Marineo, S., Mandala, S., Davids, F., Sewell, B. T., and Ingle, R. A. (2010). The missing link in plant histidine biosynthesis: *Arabidopsis myoinositol monophosphatase-like2* encodes a functional histidinol-phosphate phosphatase. *Plant Physiol.* 152, 1186–1196. doi: 10.1104/pp.109.150805
- Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)02024-2
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Hohna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sy029
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A., et al. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944–945. doi: 10.1093/bioinformatics/16.10.944
- Sahin, N., Tani, A., Kotan, R., Sedlacek, I., Kimbara, K., and Tamer, A. U. (2011). *Pandoraea oxalativorans* sp. nov. *Pandoraea faecigallinarum* sp. nov. and *Pandoraea vervacti* sp. nov., isolated from oxalate-enriched culture. *Int. J. Syst. Evol. Microbiol.* 61(Pt 9), 2247–2253. doi: 10.1099/ijso.026138-0
- Saier, M. H. Jr. (2015). The bacterial phosphotransferase system: new frontiers 50 years after its discovery. *J. Mol. Microbiol. Biotechnol.* 25, 73–78. doi: 10.1159/000381215
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Sharp, P. M., Bales, E., Grocock, R. J., Peden, J. F., and Sockett, R. E. (2005). Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res.* 33, 1141–1153. doi: 10.1093/nar/gki242

- Sharp, P. M., Emery, L. R., and Zeng, K. (2010). Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 1203–1212. doi: 10.1098/rstb.2009.0305
- Sharp, P. M., Tuohy, T. M., and Mosurski, K. R. (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* 14, 5125–5143. doi: 10.1093/nar/14.13.5125
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Soderlund, C., Bomhoff, M., and Nelson, W. M. (2011). SyMAP v3.4: a turnkey synteny system with application to plant genomes. *Nucleic Acids Res.* 39, e68. doi: 10.1093/nar/gkr123
- Soding, J., Biegert, A., and Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33, W244–W248. doi: 10.1093/nar/gki408
- Stryjewski, M. E., LiPuma, J. J., Messier, R. H., Reller, L. B., and Alexander, B. D. (2003). Sepsis, multiple organ failure, and death due to *Pandoraea pnomenusa* infection after lung transplantation. *J. Clin. Microbiol.* 41, 2255–2257. doi: 10.1128/Jcm.41.5.2255-2257.2003
- Teixeira, M. M., Borghesan, T. C., Ferreira, R. C., Santos, M. A., Takata, C. S., Campaner, M., et al. (2011). Phylogenetic validation of the genera *Angomonas* and *Strigomonas* of trypanosomatids harboring bacterial endosymbionts with the description of new species of trypanosomatids and of proteobacterial symbionts. *Protist* 162, 503–524. doi: 10.1016/j.protis.2011.01.001
- Toh, H., Weiss, B. L., Perkin, S. A., Yamashita, A., Oshima, K., Hattori, M., et al. (2006). Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Res.* 16, 149–156. doi: 10.1101/gr.4106106
- 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., Ferrantini, F., Verni, F., and Petroni, G. (2013). A new obligate bacterial symbiont colonizing the ciliate *Euplotes* in brackish and freshwater: ‘*Candidatus* Protistobacter heckmanni’. *Aquat. Microb. Ecol.* 70, 233–243. doi: 10.3354/ame01657
- Vannini, C., Pockl, M., Petroni, G., Wu, Q. L., Lang, E., Stackebrandt, E., et al. (2007). Endosymbiosis in statu nascendi: close phylogenetic relationship between obligately endosymbiotic and obligately free-living *Polynucleobacter* strains (*Betaproteobacteria*). *Environ. Microbiol.* 9, 347–359. doi: 10.1111/j.1462-2920.2006.01144.x
- Votýpka, J., Kostygov, A. Y., Kraeva, N., Grybchuk-Ieremenko, A., Tesařová, M., Grybchuk, D., et al. (2014). *Kentomonas* gen. n., a new genus of endosymbiont-containing trypanosomatids of Strigomonadinae subfam. n. *Protist* 165, 825–838. doi: 10.1016/j.protis.2014.09.002
- Zhu, L., Bi, H., Ma, J., Hu, Z., Zhang, W., Cronan, J. E., et al. (2013). The two functional enoyl-acyl carrier protein reductases of *Enterococcus faecalis* do not mediate triclosan resistance. *mBio* 4:e00613-13. doi: 10.1128/mBio.00613-13
- Zientz, E., Dandekar, T., and Gross, R. (2004). Metabolic interdependence of obligate intracellular bacteria and their insect hosts. *Microbiol. Mol. Biol. Rev.* 68, 745–770. doi: 10.1128/MMBR.68.4.745-770.2004

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 © 2017 Kostygov, Butenko, Nenarokova, Tashyreva, Flegontov, Lukeš and Yurchenko. 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) or licensor 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.