Leibniz-Institut DSMZ Deutsche Sammlung
von Mikroorganismen und Zellkulturen, GmbH

Mikrobielle Ökologie und Diversitätsforschung

PD Dr. Jörn Petersen

12. April 2019

To the chairman of the
Committee for the PhD defense
Prof. Dr. Miroslav Obornik
University of South Bohemia
Ceske Budejovice

Review

of the PhD thesis of **Karel Kopejtka** with the title:

Genome analysis of the haloalkaliphilic bacterium *Rhodobaca barguzinensis*

Karel Kopejtka has conducted the work for his PhD thesis in the group of Michael Koblizek in the Laboratory of Anoxygenic Phototrophs (Trebon, Czech Republic) with scientific supervision of Jürgen Tomasch (HZI, Braunschweig, Germany). The cumulative thesis is written in English and essentially based on one submitted as well as two published peer-reviewed first author publications. These three papers are framed by (i) an informative introduction about the distribution and evolution of different forms of bacterial photosynthesis, the characteristic photosynthesis gene clusters of aerobic anoxygenic phototrophs and the microbial life in soda lakes, (ii) a short summary of each paper and (iii) a conclusion that sumarizes the aim and outcome of the thesis.

Background and aim of the work

Rhodobacteraceae represent a metabolically versatile group of *Alphaproteobacteria* and many representatives have the capacity of anoxygenic photosynthesis (AAnP). The functional role has been investigated in photoautotrophic and photoheterotrophic model organisms such as *Rhodobacter sphaeroides* 2.4.1 and *Dinoroseobacter shibae* DFL-12, respectively. However, photosynthetic strains show an extremely scattered distribution in the phylogenomic (organismic) framework leading to different hypotheses about the evolution of photosynthesis. The emergence of photoheterotrophs was explained by at least two scenarios, (i) the regressive model predicting a photoautotrophic ancestry followed by the loss of carbon fixation (Koblizek et al. 2013) and (ii) the secondary acquisition of the photosynthesis gene cluster (PGC) by a heterotrophic bacterium via plasmid-mediated

E-Mail: Joern.Petersen@DSMZ.de • Tel: +49-531-2616-209 • Inhoffenstraße 7B • 38124 Braunschweig

horizontal operon transfer (Brinkmann et al. 2018). Karel Kopejtka chose three promising closely related haloalkaliphilic strains from soda lakes, i.e. *Rhodobaca barguzinensis* alga05 (photoheterotroph), *Roseinatronobacter thiooxidans* ALG1 (photoheterotroph) and *Natronohydrobacter thiooxidans* AH01 (heterotroph), to investigate the evolution of phototrophy in *Rhodobacteraceae*.

Spectrum of methods

The methodology included the cultivation of the three bacteria, DNA isolation and genome sequencing based on Roche 454, Illumina HiSeq and PacBio platforms. The focus of the work were bioinformatic analyses starting with three different genome annotation piplines (PGAP [NCBI], RAST, Prokka), phylogenetic analyses (16S rRNA gene, single proteins [e.g. MrpA], concatenated photosynthesis proteins [BchIDHLNB], phylogenomic protein analysis [85 core genes]) and comparative genome analyses (bidirectional best BLAST hits, Ori-finder, CRISPR-finder, detection of orthologs, phage and genomic island search tools) as well as different statistical analyses.

Results of the thesis

The thesis of Karel Kopejtka provided essentially based on the establishment of the complete genome sequence of *Rhodobaca barguzinensis* alga05 a bunch of novel insights about the the biology of this haloalkaliphilic alphaproteobacterium, the evolution of photosynthesis and a conspicuous finding about the genome organization.

(1) The complete genome of strain alga05 that is lacking any extrachromosomal elements was the basis for the reconstruction of the metabolic pathways (Figure 4, page 22) and a comparison of the metabolic potential between facultatively alkaliphilic and non-alkaliphilic *Rhodobacteraceae* (Table 2, page 23). The *mrpB* gene of the *mrpABCDEFG* operon that encodes the B subunit of the MRP Na^+/H^+ antiporter was proposed as a diagnostic marker for haloalkaliphilic strains (Kopejtka et al. 2018, *Extremophiles*).

(2) Genome sequencing of three strains and a phylogenetic 16S rRNA phylogeny proposed a common branching of all haloalkaliphilic strains and a nested positioning of the non-phototrophic representative *N. thiooxidans* AH01 among phototrophic representatives supported by 97% bootstrap support (Figure 1, page 33). The haloalkaliphilic strains ("*Rhodobaca*") and a subtree with strains belonging to the genus *Rhodobacter* were due to their common branching in the 16S rRNA tree operationally denoted as *Rhodobacter-Rhodobacer* group (*RR*-group), but the bootstrap support (BS) was $\leq$50%. The organization of the photosynthesis gene cluster (PGC; Figure 2, page 34) and a concatenated BchIDHLNB protein tree with solid BS provided evidence for the common origin of the bacteriochlorophyll biosynthesis of the *RR*-group (Fig. 4, page 37). A comparative analysis of genes related to photosynthesis showed e.g. that the PGCs of *R. barguzinensis* and *R.*

E-Mail: Joern.Petersen@DSMZ.de • Tel: +49-531-2616-209 • Inhoffenstraße 7B • 38124 Braunschweig

sphaeroides 2.4.1 replaced the genuine *pufC* gene by *pufX* (Table 2, page 35), which is indicative of a derived apparatus emerging comparably late in rhodobacteracean evolution (Brinkmann et al. 2018). Finally, the regressive evolution scenario with a secondary loss of photosynthesis in *N. thiooxidans* AH01 is suggested thus predicting that the last common ancestor of all haloalkaliphilic members of the *RR*-group was phototrophic (Kopejtka et al. 2017; *Genome Biology and Evolution*).

(3) A comparative analysis of complete genomes with circular chromosomes from 105 *Rhodobacteraceae* and four alphaproteobacterial outgroup sequences resulted in a solid phylogenomic tree with 27,688 amino acid positions that based on 85 core genome proteins (Figure 1, page 67). The comparison with the 16S rRNA tree unequivocally documented the paraphyly of the *RR*-group (100% BS) due to the nested positioning of the genus *Paracoccus*, which in turn reflects the limits of the former gold standard of microbial taxonomy, the 16S rRNA gene. However, the most conspicuous finding of this study was the observed difference of the genome organization of 101 *Rhodobacteraceae*, which was investigated by studying the localization of core genes on the chromosome and their average distance from the origin of replication (*OriC*; Figure 2, page 69). *R. barguzinensis* alga5 was the most striking outlier with a significant overrepresentation of core genes at the terminus of replication. The very distantly related roseobacter strain *Loktanella vestfoldensis* showed an analogous pattern, whereas two *Celeribacter* strains exhibit a reciprocal uneven distribution with core genes clustered at the *OriC* (Figure 3, page 70). The underlying genomic forces for the unexpected bias are unclear and provide the potential for promising follow-up studies (Kopejtka et al.; *under review*).

Evaluation of the thesis

I was excited by reading the thesis and gained many new insights into the biology of haloalkaliphilic phototrophs, the evolution of anoxygenic phototrophs and the promising field of genome architecture. The thesis of Karel Kopejtka spans, based on the exciting strain of choice *Rhodobaca barguzinensis* alga5, a broad analytical and intellectual spectrum of state-of-the-art microbial bioinformatics. The number of three first author publications and four additional co-authorships that were not subject of this cumulative thesis is absolutely convincing. Accordingly, and addressed to the Committee for PhD studies in Molecular and Cell Biology, Faculty of Science, University of South Bohemia, Czech Republic, **I explicitly recommend the acceptance of this thesis**. I am looking forward to the defense of Karel Kopejtka on April the 25th 2019. A list of questions that came up during reviewing the very sound thesis is provided below.

Braunschweig, den 12.04.2019

PD Dr. Jörn Petersen

E-Mail: Joern.Petersen@DSMZ.de • Tel: +49-531-2616-209 • Inhoffenstraße 7B • 38124 Braunschweig

List of questions that might be discussed in the defense

(01) Page 1, third paragraph
The 16S rRNA tree of Carl Woese first suggested an evolutionary split between bacteria and a common tree of archaea and eukaryotes. Are there other concepts for the origin of eukaryotes beyond a strict vertical evolution?

(02) Page 4, first paragraph
What are the most important mechanisms of genome evolution?

(03) Page 7, second paragraph & general question
In the light of the current thesis and different ideas regarding regressive evolution and the horizontal transfer of PGC operons, what do you think about the ancestry of *Rhodobacteraceae*? Was the LCA a phototrophic or a heterotrophic bacterium?

(04) Page 7, third paragraph
Eukaryotic photosynthesis can be traced back to the endosymbiosis with a cyanobacterium. Could you estimate the number of endosymbiotic events leading to chloroplasts and how often were Cyanobacteria involved?

(05) Page 9, first paragraph
What is the ecological advantage of aerobic anoxygenic photosynthesis for roseobacters and haloalkaliphilic *Rhodobacteraceae* if the bacteria can not grow photoautotrophically?

(06) Page 15, last paragraph
Based on comparative genomics you proposed that the *mrpB* gene of the MRP Na^+/N^+ antiporter defines the adaptation of haloalkalophilic *Rhodobacterales* to high pH in soda lakes. How would you proceed to experimentally test this prediction (i) in *R. barguzinensis* alga5 and in (ii) the model organism *Phaeobacter inhibens* DSM 17395?

(07) Page 18, „pH homeostasis"
What is the pH optimum of *R. barguzinensis* and non-haloalkaliphilic relatives?

(08) Page 18, „pH homeostasis"
Is there any hint for a localization of *mrpB* genes in genomic islands or on extrachromosomal elements?

(09) Page 17, 31 & 54.
You used three different annotation programms for your publications (PGAP [NCBI], RAST, Prokka). Why? Is the outcome comparable?

(10) Page 33, Figure 1
The 16S rRNA tree shows a paraphyletic grouping of *Roseinatronobacter thiooxidans* ALG1 and *R. monicus*. Would you recommend a taxonomic reclassification?

E-Mail: Joern.Petersen@DSMZ.de • Tel: +49-531-2616-209 • Inhoffenstraße 7B • 38124 Braunschweig

(11) Page 35, 37

R. barguzinensis alga05, *R. thiooxidans* ALG01 and *R. sphaeroides* 2.4.1 contain *pufX* in their PGC instead of *pufC*, which was found in *Rhodovulum sulfidophilum* DSM 2351. What are the implications with respect to the evolutionary timing?

(12) Page 33, 37, 38, 39, 67

The 16S rRNA tree provides no statistical support for the proposed *RR*-group and the phylogenomic analysis clearly shows its non-monophyly due to the placement of the strict heterotrophic genus *Paracoccus*. Which implications has this finding on the proposed regressive evolution scenario? Was the last common ancestor of *Paracoccus* photosynthetic?

(13) Page 57, 67

What are the implications of the conspicuously different placement of *Rhodovulum* spp. within the 16S rRNA and phylogenomic tree for future analyses?

(14) Page 61

The observation of an extremely uneven distribution of core genes near the terminus of replication in *R. barguziensis* alga5 and *L. vestfoldensis* is very intriguing.

- Would you expect that this effect is strain-, species- or genus-specific?
- What are the next steps to understand this phenomenon?
- What are the biological implications of the genomic disproportion for the bacterium?

(15) Page 78

The key bacterium of the current thesis *Rhodobaca barguzinensis* alga5 was designated as „model organism". Is this justified? What are the prerequisites to ennoble a strain as „model organism" and does *R. barguzinensis* fulfill these criteria?

E-Mail: Joern.Petersen@DSMZ.de • Tel: +49-531-2616-209 • Inhoffenstraße 7B • 38124 Braunschweig

Karlsruher Institut für Technologie

Institut für Biologische Grenzflächen (IBG-5)
Dr. John Vollmers
Building 601
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen

KIT-Campus Nord | IBG-5 | Postfach 3640 | 76021 Karlsruhe

Phone: +49 721 608-22515
E-mail: john.vollmers@kit.edu
Web: http://www.ibg.kit.edu/ibg5/

Date: 18.04.2019

To
Prof. Dr. Miroslav Obornik
University of South Bohemia
Ceske Budejovice

Review of the PhD thesis:
"Genome analysis of the haloalkaliphilic bacterium *Rhodobaca barguzinensis*"
of the PhD candidate Karel Kopejtka

The here reviewed thesis, submitted by Karel Kopejtka at the University of South Bohemia, Ceske Budejovice, is written as a cumulative dissertation comprising three publications. Context is provided in the form of an additional detailed introduction, a short summary preceding each publication and a final conclusion.
The thesis presents a thorough comparative genome analyses of members of the order *Rhodobacterales*, with major emphasis on haloalkaliphilic strains of the proposed *Rhodobacter-Rhodobaca* (RR) group, largely focussing on *Rhodobaca barhuzinensis* alga05 as the selected model organism. Consequently, the applied underlying techniques are predominantly of bioinformatic nature, but also cover strain cultivation as well as molecular biology methods required for sequence library preparation for multiple next generation sequencing platforms. The primary bioinformatic methodology comprises sequence processing and Genome assembly, functional annotation and metabolic pathway reconstruction, direct sequence comparisons, phylogenetic analyses as well as the detection of orthologs and genomic islands.
The stated main goals of the thesis were to identify genetic adaptations to haloalkaliphilic environments in *Rhodobaca barguzinensis* alga05 and related alkaliphiles, and to elucidate the evolutionary processes leading to the current distribution of photoautotrophy, photoheterotrophy and (chemo)heterotrophy within haloalkaliphilic *Rhodobacterales*.

Basic overview and structure:
The introduction provides a detailed information on the evolution of phototrophy as well as a concise description of the environmental conditions within soda lakes, providing sufficient details to understand the significance of the, subsequently stated, research objectives. The included publications are not ordered chronologically but rather thematically, which makes most sense in this context.
In the first publication, Karel Kopejtka provides a comprehensive genome analysis of the chosen model organism *Rhodobaca barguzinensis* alga05, giving an overview of its metabolic potential. Here, special focus was placed on potential adaptations to its alkaline and hypersaline environment by direct comparisons to selected alkaliphilic and non-alkaliphilic reference organisms.
The second publication concentrates on a specific feature, phototrophy, and its evolution and distribution among *Rhodobacterales*, using the above described, facultatively anaerobic photohetero-

troph *Rhodobaca (Rca.) barguzinensis*, the obligately aerobic photoheterotroph *Roseinatrobacter thiooxidans* and the obligately aerobic heterotroph *Natronohydrobacter thiooxidans* as major references. These references have the advantage of being closely related and originating from similar haloalkaliphilic environments: soda lakes.
Finally, the third publication provides a framework for multi-marker based phylogenomic analyses of the *Rhodobacteraceae*, thereby complementing 16S based phylogenetic analyses. Most importantly, this last publication also adds additional analyses beyond the scope of the stated original research questions, concerning chromosome structure and the localization of conserved genes within *Rhodobacteraceae* genomes.

Short summary of the results and selected highlights:

The genome of the designated model organism *Rca. barguzinensis* alga05 was analysed in detail. Environmentally relevant features were listed, described and categorized by their corresponding general function: energy, carbon, nitrogen, sulphur and phosphorous metabolism, pH and salinity homeostasis, vitamin-dependence as well as cell motility. A most conspicuous finding was the identification of *mrpB* as a major indicator for haloalkaliphilic lifestyles, due to its presence in all analysed haloalkaliphilic and absence in all analysed non-haloalkaliphilic references. Also noteworthy is the presence of marker genes for the ethylmalonyl-CoA (EMC) pathway, theoretically enabling CO_2-fixation, in all analysed *Rhodobacterales*.
Subsequently, the distribution and synteny of phototrophy gene clusters among haloalkaliphilic *Rhodobacterales* of the proposed RR-Group, was analysed with regard to two hypothetical evolutionary scenarios: a.) a "regressive" scenario, assuming a common photoautotrophic ancestor and subsequent gene loss leading to photoheterotrophic and heterotrophic species; and b.) an alternative scenario assuming a heterotrophic ancestor and subsequent uptake of phototrophy genes via horizontal gene transfer. Karel Kopejtka convincingly argues towards the regressive scenario. This is partly based on the synteny of the analysed phototrophy gene clusters, which were found to be conserved between all phototrophic members of the RR-Group and distinct from other phototrophic *Rhodobacterales*. Another indicator was the sequence similarity of the involved genes, which indicate a closer relationship within the RR group than non-alkaliphilic representatives. Lastly, the presence of singlet oxygen defence system, normally associated with the presence of bacteriochlorophyll *a*, in the heterotrophic reference organism *Natronohydrobacter thiooxidans*, which is interpreted as a remnant of an ancestral photosynthesis system.
A final highlight of the thesis are analyses describing an apparent tendency of conserved genes to be more clustered towards the origin of replication of a chromosome than towards the terminus. Such an observation could well have significant impact on our understanding of genome plasticity and evolution via horizontal gene transfer.

Karlsruher Institut für Technologie

Evaluation of the Thesis:

The thesis is well written. The Introduction is detailed and helpful, the methodology is diverse and appropriate for the questions at hand and the conclusions are generally sound and informative. I was particularly glad to read that the candidate has chosen to reannotate all comparison genomes, using a clearly defined and publicly available pipeline, before performing clustered core- and pan-genome content analyses of *Rhodobacteraceae* chromosomes. This shows that he is aware of methodological artefacts and reproducibility problems caused by different ORF-finding and annotation methods, and chose adequate measures to minimize them, an important consideration which however cannot always be taken for granted.

Apart from that, the fact that Karel Kopejtka has actually first-authored all three of the included publications, in addition to several not-included second-authorships, all within the relatively short timeframe of a PhD, is also quite impressing.

In conclusion, and after careful consideration, I found no concerns that could cause me to reject this thesis. In fact, I found the work to be a valuable and insightful contribution to the scientific field. Therefore, I recommend that this thesis is accepted by the Faculty of Science, University of South Bohemia, and that Karel Kopejtka should be allowed to defend it.

Questions:

In the following I provide a list of questions which I may want to ask the candidate, but which may not all have to be answered exhaustively:

1.) As a speculative question: You identified the presence of the *mrpB* gene as the most prominent genetic determinant for haloalkaliphilic strains. Would you expect that you could now use this for adapting any originally non-alkaliphilic *Rhodobacterales* to soda lake environments, simply by introducing this gene into its genome? Or do you expect there to be additional adaptations, not easily detectable as the mere absence or presence of a particular gene? What could these (roughly) be?

2.) Would you say the terms "photosynthesis" and "phototrophy" are synonymous or is there a difference? If there is a difference, is the term "aerobic anoxygenic photo*synthesis*", as it is commonly used in scientific literature, strictly correct?

3.) As a speculative question: If the EMC-pathway, found in all examined *Rhodobacterales*, enables CO2-fixation, wouldn't that theoretically make all of these strains, in a way, potential autotrophs (including the aerobic anoxygenic phototrophs)?

4.) Can you (roughly) define the term "ortholog" and how it relates to "homolog" and "paralog"?

5.) Could you explain the definitions of core- and pan-genome? Perhaps you may even have heard of, and can (roughly) introduce the extended categories: (hard-)core, soft-core, shell and cloud?

6.) Using strict cutoff-criteria (of 60% sequence identity and 80% coverage) you identified 85 highly conserved core genome proteins for phylogenomic analyses. Were all of these exclusively single copy genes, or were there also paralogs present in some of the references? In case of the latter, how did you deal with such paralogs in the phylogenomic analyses?

Karlsruher Institut für Technologie

7.) Can you also say how large the core-genome of the analysed *Rhodobacteraceae* is when using the relaxed cutoff-criteria (30% identity and 70% coverage) for ortholog-detection?

8.) You correctly state that the determined core genome should not change drastically with the addition of more references, as long as you have originally sampled a broad enough metabolic and phylogenetic spectrum of organisms (which you probably have). But a large part of the core-genome clustering analyses is not based on the actual *core*-genome, but rather on the average number of orthologs (found in the comparison strains) for each gene along the genome. Could *this* value be biased by the overrepresentation of a particular genus or lifestyle within the reference genome dataset?

9.) As a speculative question: What could be the reason for the different trends of conserved gene clustering observed in different *Rhodobacteraceae* genomes (mostly concentration of conserved genes towards the origin but in some cases towards the terminus)?

10.) What could be the structural and/or evolutionary principles behind the clustering of conserved genes towards any structural region of a chromosome?

Karlsruhe, 18.04.2019

Dr. John Vollmers

UNIVERSITY OF OSTRAVA
FACULTY OF SCIENCE

TO WHOM IT MAY CONCERN

Opponent´s report on the dissertation thesis "**Genome analysis of the haloalkaliphilic bacterium *Rhodobaca barguzinensis***" by Karel Kopejtka, School of Doctoral Studies in Biological Sciences, University of South Bohemia

17/04/2019, Ostrava

The thesis concerns an interesting topic of microbiology – the evolution of anoxygenic photosynthetic bacteria. Much less is generally known about these organisms compared to oxygenic phototrophic prokaryotes, i.e. cyanobacteria, so the research carried out by Karel Kopejtka during his PhD study is certainly a valuable contribution to the progress of biology. The thesis consists of a brief introductory chapter followed by three papers, all having Karel as the first author, two of them already published in respected scientific journals and one indicated as under review in *Genome Biology and Evolution*. According to the list of publications Karel's publications provided at the end of the thesis, he is a co-author of three additional papers from the field of photosynthetic microorganisms or microbial genome analysis, so altogether I believe that Karel has met the usual formal criteria of eligibility for PhD thesis defence. Below I comment on the factual content of the key parts of the thesis.

In the Introduction, Karel provides background information on bacteria and their genomics, then focuses on various details of bacterial phototrophs and their evolution, and concludes with discussing the environment of soda lakes as the habitat specific for the organisms that are his main research targets. Overall the Introduction reads well and, but I find it too brief and somewhat superficial in some regards. I do not know what are the standards for the extent of such an introductory chapter of a PhD thesis at the Faculty of Science, University of South Bohemia, but in my opinion the 13 pages of text and figures Karel dedicated to inform the reader about the broader context of his work is perhaps too little.

In addition, the text includes several statements that I find problematic. For example, on page 1 he writes that the "total number of bacterial phyla has been estimated to exceed 1,000". The statement is adopted from a questionable opinion paper by Yarza et al. (2014) that proposes to delimit bacterial phyla based on an arbitrary 16S rRNA gene sequence similarity threshold. Different criteria for delimiting phyla have been proposed by other, such as the approach by Parks et al. (*Nat Biotechnol.* 2018, 36:996-1004) using ranks normalized on the basis of relative evolutionary divergence in multigene trees, which predicts ten times less bacterial phyla than Yarza et al. Without explaining that the very concept of what is a bacterial phylum has not been settled by the community of microbiologists, the statements such as the one quoted above are extremely misleading.

Another statement I dislike is this one: "Evolution can be perceived as a process leading to the adaptation of a species to its environment" (page 3). This sound like a quotation from an old textbook written before the evolutionary theory has integrated ideas such neutral evolution or the selfish gene. It is particularly surprising to read such a sentence in a text written by somebody who has as a daily routine analysing genomes and DNA sequences. All the phylogenetic trees presented in the thesis have been constructed using methods that assume the sequences analysed have evolved

Faculty of Science, University of Ostrava / 30. dubna 22 701 03 Ostrava / Czech Republic
e-mail: lucie.zarnikova@osu.cz / telephone: +420 597 092 102
prf.osu.eu

neutrally! And what about the prophages and other mobile or parasitic genetic elements that are so central in the study presented in the Paper 3 of the thesis? Are they also a manifestation of the evolution as a process by which the bacterial species that carry them adapt to their environments? Similarly questionable is the claim on page 4 that bacterial genomes "can evolve through four mechanisms: spontaneous mutations, gene duplications, horizontal gene transfer [...] and gene loss". So mutations induced by external mutagens (i.e. not spontaneous) do not contribute to bacterial genome evolution? And gene loss resulting from an error during replication is not a form of a spontaneous mutation? I think the classification of "evolutionary mechanisms" presented in the thesis is highly misleading. It is also interesting to note that on page 28, Karel writes that evolution of a prokaryotic genome "is driven by three fundamental processes: random mutations, gain of genes *via* HGT, and gene loss". Why gene duplication has become insignificant between the page 4 and 28?

Truly surprising is that Karel has erred when introducing the topic key to the whole thesis. On page 5 he writes: "Photosynthesis has been found in seven bacterial phyla", and follows with their list culminating with Gemmatimonadetes, whose photosynthetic representative was discovered in 2014. The thesis has thus missed the discovery of an eight bacterial phylum including anoxygenic phototrophs, the candidate phylum WPS-2. Evidence for the presence of anoxygenic photosynthesis in this phylum was published in July 2018 (Holland-Moritz et al., *Environ Microbiol.* 20:2625-38), i.e. half a year before the thesis was completed, so it could have been included. A more recent report in the bioRxiv, admittedly posted only after the submission of the thesis, provides additional fascinating details on this new phototrophic bacterial group (Ward et al., https://doi.org/10.1101/534180). I encourage Karel to discuss at the defence this new phototrophic bacterial group and how it illuminates the evolution of photosynthesis in general.

I also would challenge some additional statements of the thesis concerning evolution of bacterial photosynthesis and phototrophs. On page 5 Karel writes that anoxygenic phototrophs "represent some of the oldest bacterial species on Earth". Is it really possible to speak about species here? I would say that what is in mind in this context are whole phylogenetic lineages or may be even life forms in a general sense, without reference to the taxonomic category of a species. On page 7, Karel writes: "It is assumed that the first photosynthetic organisms emerged in Earth more than 3.5 billion years ago (Awramic 1992)". I was wondering how this compares to the statement on page 1 that the "last common ancestor of Bacteria and Archaea [...] lived 2.5 – 3.2 billion years ago". This would mean that the photosynthetic organisms living 3.5 billion years ago must have been ancestral to the LUCA or must have belonged to lineages that were parallel to the one leading to the LUCA. Is this realistic? I am truly interested in knowing Karel's opinion. On page 7 Karel further comments on the evolutionary succession of different forms of photosynthesis: "phylogenetic analyses of bacterial photosynthesis genes suggest that anoxygenic photosynthesis preceded to oxygenic photosynthesis". Is this a generally accepted scenario? I would point to a recent paper by Cardona (*Open Biol.* 2019, 9:180246) that presents perhaps a dissent perspective on what came first. Again, I would like to know Karel's own opinion on this subject.

My final comment on the Introduction section concerns its formal shape. Despite the brevity of the text, it contains a significant number of typos and mistakes in the language. Just a few examples (problematic words are underlined): "do not possess neither a nucleus" (page 1); "a recently identified phylum Gemmatimonadetes" (page 5); "develope" (page 8); "and its essential

function" (page 10); "Methanocalculus, Methanolobous ..." (page 12) – the genus names should be italicized.

The first two papers included in the thesis have undergone the process of peer review before being published, so I think it is not critical here to provide their detailed analysis. They are both significant pieces of work and deliver important new information on the biology and evolution of the haloalkaliphilic phototrophic bacterium *Rhodobaca barguzinensis* and its relatives in the alphaproteobacterial order Rhodobacterales as gleaned from an analysis of newly sequenced genomes. Nevertheless, I do see a need to raise a few questions concerning these papers.

Paper 1:
The observation that all alkaliphilic Rhodobacterales species investigated possess the *mrpB* gene whereas the non-alkaliphilic ones all lack it is interesting. However, there is no mechanistic explanation offered in the paper as to why the *mrpB* gene might be specifically important for alkaliphiles. Could Karel elaborate on this question by considering what is known about the role of the protein encoded by the gene, or at least to speculate how it possibly functions? Next, according to the paper, *Paracoccus chinensis* possesses a version of the *mrp* operon lacking the *mrpB* gene, but implicitly it follows from the text that the species does have the gene. Does this mean that the gene is placed elsewhere in the genome? If yes, what can be said about the origin of the gene?

Paper 2:
The paper defines the so-called *Rhodobacter-Rhodobaca* group, the key target of the paper and in fact of the whole thesis (section 1.5 Objectives), based on a phylogenetic tree of the 16S rRNA gene constructed using a rather simplistic HKY85 substitution model. Honestly, I do not understand why the GTR model, a standard and generally much better model perfectly appropriate for the type of sequence dataset analysed here, was not used, and it is also unclear from the description in the paper whether and how correction of among-site rate heterogeneity, simply a must in phylogenetic analyses nowadays, was employed. Regardless, the *Rhodobacter-Rhodobaca* group has no bootstrap support in the tree shown in the figure. So it is justifiable to consider it as a biological entity? Indeed, the phylogeny inferred from multiple markers (i.e. a phylogenomic tree) presented as Fig. 1 in the Paper 3 of the thesis shows the *Rhodobacter-Rhodobaca* group as a paraphyletic assemblage (as in fact does a 16S rRNA gene-based tree in the same paper). How does the fact that the *Rhodobacter-Rhodobaca* group apparently is an artificial grouping affects the conclusions of the paper?

Next, I noticed that in Fig. 1. of the Paper 2 (a 16S rRNA gene-based phylogenetic tree) *Gemmobacter aquatilis* is highlighted in bold, which would mean it is a phototrophic species. However, I an equivalent figure in the Paper 1 it is not marked as photosynthetic. So which of these two figures is inaccurate in this regard?

The Paper 3 is an unpublished manuscript that does not address questions directly related to bacterial photosynthesis but instead focuses on certain salient general aspects of bacterial genome evolution, using the partly photosynthetic family Rhodobacteraceae as a model group. Specifically, the study tests on a sample of over 100 genomes of this group the previously proposed trend of

bacterial genome evolution that conserved highly expressed genes are preferentially located in the genome closer to the origin of replication (*oriC*) than to the terminus of replication (*ter*), whereas accessory genes gained by HGT are typically found closer to the *ter* than the *oriC*. My understanding of the conclusions of the paper is such that more Rhodobacteraceae representatives conform to this pattern than those that do not, with some Rhodobacteraceae species, above all *Rhodobaca barguzinensis*, exhibiting an extreme departure from the expected pattern. This type of genome analyse are quite far from my own expertise, so I do not feel truly competent to evaluate the methodology of the study, but me general impression is the analyses are sound and the results robust. Nevertheless, I do have some concerns with the phylogenomic analysis presented in the paper. While inferring the phylogeny of the groups from multiple genetic markers (85 highly conserved proteins in this case) is certainly and advance comparted to the phylogenies based only on the 16S rRNA gene that were used in the previous two papers (see above). However, only the site-homogeneous model LG was used for the tree inference, which is generally inappropriate for phylogenomic matrices of the type employed in the study and may lead to highly supported yet artificial topologies. I recommend to consult the recent paper by Muñoz-Gómez et al. (*Elife*, 2019, 8:e42535), which reports on a rigorous phylogenomic analysis of Alphaproteobacteria (incl. Rhodobacteraceae representatives) using the state-of-the-art methodology. In addition, I see some problems in the presentation of the data in the Paper 3 of the thesis. Disregarding minor formal issues, I found particularly disturbing that for unclear reasons only three out of six supplementary figures have been included in the thesis, which makes understanding of some parts of the main text and evaluation of some of the results difficult, if not impossible.

The thesis closes with a short section summarizing the key findings of Karel's research. The list of particular conclusions included in this section gives a good idea about what has been achieved in the thesis, and my own conclusion is that the extent of Karel's work and the degree of novelty of his results make his thesis a solid piece of scientific work. Hence, despite some criticism mentioned above, **I recommend the thesis for defence.**

I am sorry for not being able to attend the defence in person. Hence, I would like to ask Karel for answering the various question I have raised in this report by e-mail before the defence and for presenting the answers at the defence. Let me add one more question that has puzzled me even before reading Karel's thesis and which is related to his research: is it possible to say at which point of alphaproteobacterial phylogeny photosynthesis appeared in this group? Will we ever know the answer with any confidence?

doc. Mgr. Marek Eliáš, Ph.D.

Department of Biology and Ecology,
Faculty of Science, University of Ostrava
Chittussiho 10, 710 00 Ostrava