Reclassification of a *Polynucleobacter cosmopolitanus* strain isolated from tropical Lake Victoria as *Polynucleobacter victoriensis* sp. nov.

Martin W. Hahn,¹* Johanna Schmidt,¹ Grace Ssanyu Asiyo,² Nikos C. Kyripides,³ Tanja Woyke³ and William B. Whitman⁴

**Abstract**

The genus *Polynucleobacter* (family *Burkholderiaceae*) is phylogenetically subdivided into at least four subclusters. One of those, subcluster PnecC, was recognized as a cryptic species complex. Here we test by comparative genome analyses whether subcluster PnecD, currently solely represented by the species *Polynucleobacter cosmopolitanus*, also represents such a cryptic species complex. The genome sequences of the two *P. cosmopolitanus* strains, MWH-Molsø² and MWH-VicM1, were determined. The latter strain was also characterized in the previous description of *P. cosmopolitanus*. These two strains originate from a temperate lake located in Austria and from the large tropical Lake Victoria located in East Africa, respectively. Strains MWH-Molsø² and MWH-VicM1 possess quite small genomes of 1.78 and 1.63 Mbp, respectively, and share similar G+C values of 44.1 and 43.1 mol%, respectively. Both strains encode only a single copy of the ribosomal operon, and their 16S rRNA genes differ only in four positions, equalling a sequence similarity of 99.74%. Both genomes possess characteristics indicating evolutionary genome streamlining, such as high coding densities of 93.9 and 94.6% of bases, respectively. Average nucleotide identity (ANI) comparisons of the genomes of the two strains resulted in a value of 78.4%, suggesting that each of the strains represents a separate species. Our investigation suggests that PnecD represents an additional cryptic species complex within the genus *Polynucleobacter* that was not resolved by 16S rRNA gene sequence analyses. We propose reclassification of strain MWH-VicM1 as *Polynucleobacter victoriensis* sp. nov., with type strain MWH-VicM1¹ (=DSM 21486⁵ =JCM 32005⁷).

The genus *Polynucleobacter* and the species *Polynucleobacter necessarius* were described by Klaus Heckmann and Helmut J. Schmidt as obligate endosymbionts of ciliates [1]. Diversity studies by cultivation-independent methods revealed that bacteria closely related to *P. necessarius* are typically present in the water column of lakes [2], ponds [3] and running waters [4], and that these bacteria frequently occur in such systems with high cell numbers [5]. Isolation of strains [6], investigations of environmental samples by fluorescence *in situ* hybridization (FISH) [3, 7], and cultivation experiments with endosymbiotic *P. necessarius* [8] indicated that *Polynucleobacter* bacteria dwelling in the water column of freshwater systems represent free-living bacteria and not obligate symbionts of ciliates [9]. Recent genome comparisons between an endosymbiotic and a free-living *Polynucleobacter* strain revealed lifestyle-specific genomic signatures, including, for instance, the presence of a large number of pseudogenes and about 30% reduction in genome size of the obligate endosymbiont [10].

A large number of *Polynucleobacter* strains could be isolated from freshwater systems and were characterized by sequencing of phylogenetic markers [6, 11, 12]. Based on phylogenetic analyses of 16S rRNA genes, the strains were grouped into the subclusters PnecA, PnecB (including PnecB1 and PnecB2), PnecC and PnecD [3, 6]. These subclusters are characterized by inter- and intra-subcluster 16S rRNA sequence similarity values of <98% and ≥99%, respectively [6, 9]. Recent investigations demonstrated that subcluster PnecC represents a cryptic species complex of currently seven described species and an unknown but presumably large number of undescribed species [13–16], which cannot be resolved by analysis of 16S rRNA gene sequences. Presently, it is not known if the other three
Polynucleobacter subclusters also represent such cryptic species complexes.

Subcluster PnecD is currently represented exclusively by the species Polynucleobacter cosmopolitanus [17]. The description of this species included the characterization of five strains forming a monophyletic lineage and sharing 16S rRNA sequence similarities ≥99.3%. Owing to generally weak growth of all Polynucleobacter strains in artificial medium (see Suppl. Mat. Fig. S2 in [9]) and the difficulty in extracting sufficient genomic DNA for pairwise DNA–DNA reassociation experiments, investigations on genomic coherence of the five strains were omitted. Instead the five strains were preliminarily described as members of the species P. cosmopolitanus [17].

In the study presented here, we investigated the genomic similarity of two strains previously included in the species P. cosmopolitanus [17]. These strains are the type strain MWH-Moliso2T isolated from Lake Mondsee, Austria, and the tropical strain MWH-VicM1 isolated from Lake Victoria near Kampala, Uganda [6]. Both strains were cultivated by using the filtration-acclimatization method and NSY medium [18] as described previously [6]. A comprehensive phenotypic and chemotaxonomic characterization of both strains was included in the previous species description [17]. Based on large genomic differences revealed by comparative genome analyses, we conclude that the two strains represent different species and propose to establish MWH-VicM1T as the new type strain of P. victoriensis sp. nov.

### GENOMIC CHARACTERIZATION OF STRAINS MWH-MOISO2T AND MWH-VICM1T

Genomic DNA of both strains MWH-Moliso2T and MWH-VicM1T was extracted from biomass grown in liquid NSY medium as described previously [19] and used for genome sequencing. DNA of strain MWH-Moliso2T was sequenced by Illumina MiSeq. Paired-end sequencing (2 x 150 bp) of a fragment library resulted in about 1.2 x 10^8 quality filtered reads with a mean length of 148 bp. Sequence assembly and subsequent closure of some gaps resulted in six contigs with a total length of 1.78 Mbp (Table 1). Sequencing coverage was about 100-fold.

Strain MWH-VicM1T was sequenced at the DOE-Joint Genome Institute as part of the Genomic Encyclopedia of Type Strains, Phase III (KMG-III) study [20] using the Illumina HiSeq 2000 1TB platform. Paired-end sequencing (2 x 150 bp) of a fragment library resulted in about 7.4 x 10^6 quality filtered reads. Assembly of reads resulted in three contigs with a total sequence length of 1.63 Mbp and a sequencing coverage of about 680-fold.

The obtained genome sequences of the two strains, MWH-Moliso2T and MWH-VicM1T, were annotated using the IMG/ER annotation pipeline [21]. The IMG Genome IDs of the two genomes are 2642422582 and 2710264786, respectively. Both genome sequences were also deposited in GenBank/EMBL/DDBJ (accession numbers NJGG00000000 and FYEX00000000, respectively). Gene-finding using the JGI annotation pipeline [22, 23] resulted in 1822 and 1677 ORFs in strains MWH-Moliso2T and MWH-VicM1T, respectively (Table 1).

Despite neither genome being closed, the obtained genome sizes are expected to be quite close to the real genome sizes of the two respective strains. Owing to the high sequencing coverage, the whole or almost the whole genome of the strains should be represented by reads. Previous independent sequencing of two very closely related Polynucleobacter asymbioticus strains isolated from the same habitat at a time interval of 4 years and differing only in 22 single nucleotide polymorphisms clearly resulted in the same genome size despite different sequencing technologies and strategies being used [24]. On the other hand, many Polynucleobacter strains contain a few repetitive sequences, which include insertion elements, giant genes with repetitive sequence elements and the translation elongation factor Tu gene present as two identical copies in all Polynucleobacter genomes investigated so far. Such repetitive sequences frequently do not assemble properly and form contigs in the size range of 0.1–2.0 kbp. Such contigs originating from repetitive sequences are usually characterized by coverage values of two-, three- or fourfold the mean coverage of the whole genome. Contigs smaller than 1 kbp are discarded by the IMG annotation pipeline, and thus do not contribute to the final genome size values of draft genomes. In general, such repetitive sequences not assembling properly results in underestimations of the real genome size; however, analyses of several Polynucleobacter genome assemblies suggest that the underestimation is usually in the range of about 1 to 5 kbp, which equals in typical Polynucleobacter genomes in underestimations of genome size by less than 0.4%. Thus, the genome size data given in Table 1 are assumed to represent the genome sizes of the investigated strains quite well.

Among the free-living Polynucleobacter strains with sequenced genomes, strain MWH-VicM1T has the smallest genome size. Both genomes show signatures of evolutionary genome streamlining previously described for another free-living Polynucleobacter strain [25] but lack large numbers of pseudogenes, a known signature for reductive genome evolution in obligate endosymbionts [10, 26]. Genome streamlining in these two investigated PnecD strains is indicated, for instance, by the high coding densities of the genomes of strains MWH-Moliso2T and MWH-VicM1T of 93.9 and 94.6%, respectively. Such high coding densities have not been reported, so far, for other members of the family Burkholderiaceae not affiliated with the genus Polynucleobacter. The IMG system [21] contains currently (March 2017) 811 genomes of Burkholderiaceae (single cell genomes were excluded) bacteria not affiliated with the genus Polynucleobacter. Among these genomes, only seven (0.9%) had coding densities >90%, and the mean coding density was 85.1%. By contrast, genomes of free-living Polynucleobacter strains usually possess coding densities >92% [13, 14, 16, 19, 24, 25, 27].
The two genome sequences presented here are the first for *Polynucleobacter* genomes of strains affiliated with subcluster PnecD [6]. All previously published *Polynucleobacter* genome sequences represent strains belonging to subcluster PnecC [13, 14, 16, 24, 25]. The sizes of the two PnecD genomes presented in this study and the 18 previously published genomes of free-living *Polynucleobacter* strains affiliated with subcluster PnecC (Fig. 1 and [24]) differ. All published genomes of free-living PnecC strains are characterized by genome sizes >2 Mbp, while the two PnecD genomes possess sizes of <1.78 Mbp (Table 1). Unpublished genome size data of PnecC strains prove that there is no systematic difference in genome size between PnecD and PnecC strains (M. W. Hahn et al., unpublished data). Interestingly, the genome of the free-living (planktonic) strain MWH-VicM1 is only 66 kbp larger than the genome of the obligate endosymbiont *P. necessarius* STIR1 [10] and is even smaller than genomes of a few other endosymbiotic *Polynucleobacter* strains recently presented [26]. This suggests that genome sizes in obligate endosymbiotic and free-living *Polynucleobacter* strains are not necessarily very different. Thus genome size neither is a systematic trait discriminating all PnecC and PnecD strains, nor does it discriminate between free-living and endosymbiotic *Polynucleobacter* strains.

Both of the PnecD strains lack a gene cluster encoding an anoxygenic photosynthesis system previously found in the type strains of *Polynucleobacter duraquae* [14] and *P. wuianus* [16]. Interestingly, *P. cosmopolitanus* strain MWH-Molso2T encodes a special variant of rhodopsins (xanthorhodopsins) that is also found in the obligate methylotroph *Candidatus Methylomarinus turicensis* [28] with an amino acid identity of 82%. These xanthorhodopsins are light-driven proton pumps tuned to green light, which could represent an adaptation to the light conditions in more productive systems [29]. Rhodopsin genes were only found in 1 [27] of the 18 genomes of planktonic *Polynucleobacter* strains and in none of the endosymbiotic strains [10, 26] investigated previously. Strain MWH-Molso2T encodes flagella, which strain MWH-VicM1T lacks. Both strains encode an ABC-type Fe3+ transport system but no FeoAB transporters for uptake of Fe3+ ions. This reflects very well the alkaline pH conditions of their habitats [6, 13]. Neither strain encodes a cytochrome bd-1 terminal oxidase (CydAB) [14–16] or a fumarate reductase (FrdABCD) found in some PnecC genomes [14–16]. Thus, both PnecD strains lack these features typically suggesting adaptation to low oxygen concentrations and facultative anoxygeniosis, respectively.

Analysis of genome similarity between *P. cosmopolitanus* MWH-Molso2T and strain MWH-VicM1T by ANI [30, 31] using the IMG system [21] resulted in a value of 78% ANI (Fig. 1). By contrast, pairwise ANI comparisons with genomes of PnecC strains resulted in values of about 72% ANI, and ANI values shared by strain MWH-VicM1T with the type strains of two *Cupriavidus* species were about 70%. The alignment fractions from which these ANI values resulted were 78, 41–46 and 12–15% of the genome size of strain MWH-VicM1T for the comparisons with strain MWH-Molso2T (intra-PnecD comparison), for comparisons with PnecC strains, and for comparisons with *Cupriavidus* spp. strains, respectively. Both the ANI values [24] and the alignment fraction data indicate that strain MWH-VicM1T shares more homologous genes with the other PnecD strain than with PnecC strains or the closest relatives outside of the genus *Polynucleobacter*.

### Table 1. Genome characteristics of the investigated *Polynucleobacter* strains

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain</th>
<th>Genome size (Mbp)</th>
<th>Scaffolds</th>
<th>G+C content (mol%)</th>
<th>GenBank/EMBL/DDBJ accession no.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. victorienis</em> sp. nov.</td>
<td>MWH-VicM1T (=DSM 21486T)</td>
<td>1.63</td>
<td>3</td>
<td>43.1</td>
<td>FYEX00000000</td>
<td>This study</td>
</tr>
<tr>
<td><em>P. cosmopolitanus</em></td>
<td>MWH-Molso2T (=DSM 21490T)</td>
<td>1.78</td>
<td>6</td>
<td>44.1</td>
<td>NJGG00000000</td>
<td>This study</td>
</tr>
<tr>
<td><em>P. necessarius</em></td>
<td>STIR1 (host Euploetes aediculatus)</td>
<td>1.56</td>
<td>1</td>
<td>45.6</td>
<td>CP001010</td>
<td>[10]</td>
</tr>
<tr>
<td><em>P. aenigmaticus</em></td>
<td>MWH-K35W1T (=DSM 24006T)</td>
<td>2.14</td>
<td>37</td>
<td>46.0</td>
<td>NGU00000000</td>
<td>[27]</td>
</tr>
<tr>
<td><em>P. asymbioticus</em></td>
<td>QLW-P1DMWA-1T (=DSM 18221T)</td>
<td>2.16</td>
<td>1</td>
<td>44.8</td>
<td>CP000655</td>
<td>[19]</td>
</tr>
<tr>
<td><em>P. duraquae</em></td>
<td>MWH-MoK4T (=DSM 21495T)</td>
<td>2.03</td>
<td>1</td>
<td>45.2</td>
<td>CP007501</td>
<td>[14]</td>
</tr>
<tr>
<td><em>P. sinensis</em></td>
<td>MWH-HuW1T (=DSM 21492T)</td>
<td>2.32</td>
<td>19</td>
<td>45.5</td>
<td>LOJ10100000</td>
<td>[14]</td>
</tr>
<tr>
<td><em>P. yangtzensis</em></td>
<td>MWH-JaK3T (=DSM 21493T)</td>
<td>2.05</td>
<td>42</td>
<td>45.4</td>
<td>LOJ10100000</td>
<td>[14]</td>
</tr>
<tr>
<td><em>P. wuianus</em></td>
<td>QLW-P1FAT50C-4T (=DSM 24008T)</td>
<td>2.23</td>
<td>1</td>
<td>44.9</td>
<td>CP015922</td>
<td>[16]</td>
</tr>
<tr>
<td><em>P. spagnophilus</em></td>
<td>MWH-Weng1-1T (=DSM 24018T)</td>
<td>2.04</td>
<td>17</td>
<td>45.6</td>
<td>MPIY01000000</td>
<td>[15]</td>
</tr>
</tbody>
</table>

FL, Free-living; E, endosymbiotic.
Fig. 1. Phylogenetic position of strain MWH-VicM1<sup>T</sup>. (a) Maximum-likelihood (ML) tree calculated with eight concatenated housekeeping gene sequences. Sequences of the type material of <i>P. necessarius</i> [1] could not be included owing to the unavailability of the culture containing this endosymbiont [14]. Instead, sequences of two other <i>P. necessarius</i> strains were included, one of which is most likely identical with the type material of the species [14]. Bootstrap values are shown from left to right for neighbour-joining (NJ), ML and maximum-parsimony (MP) trees calculated with the same sequence set. Pairwise ANI values for whole genome comparisons of strain MWH-VicM1<sup>T</sup> with the other shown taxa are given. Bar, 0.1 substitution per nucleotide position. (b) ML tree calculated with 16S rRNA gene sequences and pairwise 16S rRNA sequence similarity values of strain MWH-VicM1<sup>T</sup> with the other taxa shown. Bootstrap values are shown from left to right for NJ, ML and MP trees calculated with the same sequence set. Bar, 0.01 substitution per nucleotide position. NA, Not available.
PHYLOGENY

As shown previously, phylogenetic reconstructions based on 16S rRNA gene sequences placed strains MWH-Molso2T and MWH-VicM1T in subcluster PnecD of the genus Poly-
nucleobacter [6]; however, the sequences of this ribosomal
gene are too similar (99.74 %) to provide suitable phyloge-
etic resolution (Fig. 1b). In order to establish a phyloge-
netic reconstruction with a higher resolution, multilocus
sequence analysis based on eight housekeeping genes
defined previously [14, 27] was performed. In all but one
case (endosymbiotic P. necessarius strain Ammermann,
[14]), these genes were extracted from the genomes of the
type strains of all previously described PnecC species, from
two PnecC strains representing two separate undescribed
PnecC species [32, 33], and from the two PnecD strains
investigated here. The genomes of two Cupriavidus strains
[34, 35] served as outgroup. Concatenation of partial gene
sequences resulted in an alignment length of 6249 align-
ment positions. Neighbour-joining, maximum-likelihood
and maximum-parsimony trees were calculated by using
the software MEGA7 [36] (Fig. 1b). The obtained multilocus
trees confirm the separation of the investigated strains in
the two subclusters, PnecC and PnecD, and demonstrate
that the phylogenetic distance between the two PnecD
strains is similar to distances between distinct PnecC
species.

ECOLOGY AND BIOGEOGRAPHY

Strain MWH-VicM1T was isolated from a tropical lake. Like
PnecC strains also isolated from tropical freshwater systems
[37], it lacked the ability to grow at temperatures of 4–6 °C.
On the other hand, the strain showed a higher maximum
growth temperature (38 °C) than all other PnecD or PnecC
strains isolated from habitats located in temperate climatic
zones [9, 15–17, 27, 38–40]. Both traits hint at a thermal
adaptation of strain MWH-VicM1T to tropical or at least to
warmer climate conditions [41]. For tropical PnecC strains
a restricted geographical range was indicated by cultivation-
independent investigations [37]. The clade formed by these
tropical strains could not be detected in Central European
habitats but was detected in Ugandan habitats. Similarities
between thermal adaptation and biogeography of PnecC
strains and strain MWH-VicM1T are obvious. However,
characterization of additional strains of the new species are
needed to determine if the geographical range of the new
species is really restricted in a way similar to that in the
PnecC strains.

CRYPTIC SPECIES COMPLEX

The ANI results of the comparisons of the genomes of
strains MWH-Molso2T and MWH-VicM1T stand in
strong contrast to the sequence similarity of their 16S
rRNA genes (Fig. 1). A very similar discrepancy between
ANI values and 16S rRNA gene similarities is well docu-
mented for several strains affiliated with subcluster PnecC
[13, 27]. Only two of the five strains included in the
description of P. cosmopolitanus [17] have had their
genomes sequenced so far; thus, one can only speculate
whether the remaining three strains also represent sepa-
rate species. A recent investigation on four free-living Pol-
y nucleobacter strains included in the emended description of
P. necessarius [9] revealed that these strains represent
four separate species [14]. Sequence comparisons based
on partial glutamine synthetase (glnA) gene sequences
may provide hints as to whether or not the three other P.
cosmopolitanus strains [17] actually represent novel spe-
cies. A previous investigation on PnecC strains suggested
that strains sharing ANI values <85 % also share glnA
sequence similarities <94 % [13]. Analyses of the five glnA
sequences (accession numbers FR732027, FR732028,
FR732031, FR732033 and FR821089) of PnecD strains
included in the previous description of the species P. cos-
mopolitanus resulted in an average sequence similarity of
92.3 %, a minimum value of 88.9 % and a maximum
value of 97.7 %.

POPROAL OF THE NEW SPECIES

**POLYNUCLEOBACTER VICTORIENSIS SP. NOV.**

The performed ANI analyses and the phylogenetic recon-
structions show that the previously described species P.
cosmopolitanus contains at least two distinct species. The
determined value of 78.4 % ANI (Fig. 1a) lies far below
the threshold range of 93–96 % ANI suggested to demar-
cate strains representing separate species [31, 42, 43].
Furthermore, comparison of the revealed phylogenetic
distances between the two investigated PnecD strains with
distances between distinct species within subcluster PnecC
confirms the separation of the two strains into two spe-
cies (Fig. 1a). We propose to establish the new species P.
victoriensis sp. nov. to harbour strain MWH-VicM1T.
Future genome sequence-based investigations are needed
to reveal whether the other three strains included in the
description of P. cosmopolitanus really belong to this spe-
cies, whether they also belong to the new species pro-
posed here, or whether they represent separate, so far
undescribed species.

Strain MWH-VicM1T can be discriminated from Poly-
nucleobacter strains not affiliated with subcluster PnecD by
the presence of the fatty acid C_{12:0} 3-OH, which in Poly-
nucleobacter strains has so far been detected exclusively
in all investigated strains affiliated with subcluster PnecD
[17]. A genotypic trait of all so far investigated strains
affiliated with subcluster PnecD is the presence of the

---

**Hahn et al., Int J Syst Evol Microbiol 2017;67:5087-5093**

Downloaded from www.microbiologyresearch.org by
IP: 5091.114.010.11
On: Sat, 10 Nov 2018 00:11:40
signature sequence 5′-AA(T/G)CCCT(A/T)AGGG-GAAA-3′ within the 16S rRNA gene (Escherichia coli positions 181–197) [44]. Strain MWH-VicM1\(^T\) can be discriminated from the type strain of \textit{P. cosmopolitanus}, which also belongs to subcluster PnecD, by its ability to assimilate malonic acid, as well as by growth at 38°C and lack of growth at 5°C [17].

**DESCRIPTION OF POLYNUCLEOBACTER VICTORIENSIS SP. NOV.**

\textit{Polynucleobacter victoriensis} (vic.to.r.en’sis. N.L. masc. adj. victoriensis of or belonging to Lake Victoria, the lake from which the type strain was isolated).

The description is based on phenotypical data of Hahn [6] and Hahn \textit{et al.} [17], on chemotaxonomical data of Hahn \textit{et al.} [17] and on genomic data presented in the present study (Table 1). Contains free-living \textit{Polynucleobacter} strains dwelling in the water column of freshwater systems. Cells are short curved rods, 0.4–1.1 μm in length and 0.3–0.5 μm in width, depending on cultivation conditions. Chemo-organotrophic, aerobic, and weak anaerobic growth was observed. Colonies grown on NSY agar are non-pigmented, circular and convex with smooth surface. Growth occurs up to 38°C but not at 5°C. Growth occurs in 0–0.5% (w/v) NaCl. Assimilates acetate, propionate, pyruvate, malate, malonate, fumarate, succinate, oxaloacetate, L-alanine and L-cysteine. Weak assimilation of D-galacturonic and D-galacturonic acid. Does not assimilate glycocolate, glyoxylic acid, D-galacturonic and D-galacturonic acid. The type strain is MWH-VicM1\(^T\) (=DSM 21486\(^T\)=JCM 32005\(^T\)), which was isolated from Lake Victoria near Kampala, Uganda. The species epithet indicates the origin of the type strain but does not indicate that the distribution of the taxon is restricted to a certain geographical area or a certain freshwater system. The draft genome of the type strain is characterized by a size of 1.63 Mbp and a G+C content of 43.1 mol%.

**Funding information**

This study was supported by the Austrian Science Fund (FWF) project I482-B09 and the European Science Foundation (ESF) project FREDI. Genome sequencing of strain MWH-VicM1\(^T\) was conducted by the US Department of Energy (DOE) Joint Genome Institute, a DOE Office of Science User Facility, which is supported by the Office of Science of the US DOE under Contract no. DE-AC02-05CH11231.

**Conflicts of interest**

The authors declare that there are no conflicts of interest.

**Ethical statement**

The presented study does not include any experimental work with humans or vertebrates.

**References**


**Five reasons to publish your next article with a Microbiology Society journal**

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologysociety.org.