

1 **Analyses of the complete genome sequence of 2,6-dichlorobenzamide (BAM) degrader**
2 ***Aminobacter* sp. MSH1 suggests a polyploid chromosome, phylogenetic reassignment, and**
3 **functions of (un)stable plasmids.**

4 **Running title: Complete genome of *Aminobacter* sp. MSH1**

5 **Authors**

6 Tue Kjærgaard Nielsen^{a†}, Benjamin Horemans^{b†}, Cedric Lood^{c,d}, Jeroen T'Syen^e, Vera van Noort^c,
7 Rob Lavigne^d, Lea Ellegaard-Jensen^f, Ole Hylling^f, Jens Aamand^g, Dirk Springael^{e‡*}, Lars
8 Hestbjerg Hansen^{a‡*}

9 [†]These authors contributed equally to the work presented

10 [‡] DS and LHH initiated the study and supervised the project equally in separate labs

11 **Institutional Affiliation**

12 ^a University of Copenhagen, Faculty of Science, Department of Plant and Environmental Sciences,
13 Section for Microbiology and Biotechnology, Copenhagen, Denmark

14 ^b BAT Knowledge Centre, Sustainable Materials Unit, Vlaams Instituut voor Technologisch
15 Onderzoek, Mol, Belgium

16 ^c KU Leuven, Faculty of Bioscience Engineering, Department of Microbial and Molecular Systems
17 (M²S), Centre of Microbial and Plant Genetics, Leuven, Belgium

18 ^d KU Leuven, Faculty of Bioscience Engineering, Department of Biosystems, Laboratory of Gene
19 Technology, Leuven, Belgium

20 ^e KU Leuven, Faculty of Bioscience Engineering, Department of Earth and Environmental
21 Sciences, Division of Soil and Water Management, Leuven, Belgium

22 ^f Aarhus University, Department of Environmental Science, Section of Environmental
23 Microbiology and Circular Resource Flow, Roskilde, Denmark

24 ^g Geological Survey of Denmark & Greenland (GEUS), Department of Geochemistry,
25 Copenhagen, Denmark

26 **Corresponding authors**

27 * Professor Dirk Springael

28 KU Leuven

29 Faculty of Bioscience Engineering

30 Department of Earth and Environmental Sciences

31 Division of Soil and Water Management

32 Kasteelpark Arenberg 20 bus 2459

33 3001 Leuven

34 Phone: +32 16321604

35 E-mail: dirk.springael@kuleuven.be

36 and

37 * Professor Lars H. Hansen

38 Microbial Ecology and Biotechnology

39 Department of Plant and Environmental Sciences

40 University of Copenhagen

41 Thorvaldsensvej 40

42 1871 Frederiksberg, Denmark

43 Mobile: +45 28752053

44 E-mail: lhha@plen.ku.dk

45 **Abstract**

46 *Aminobacter* sp. MSH1 (CIP 110285) can use the pesticide dichlobenil and its transformation
47 product, the recalcitrant groundwater micropollutant, 2,6-dichlorobenzamide (BAM) as sole
48 source of carbon, nitrogen, and energy. The concentration of BAM in groundwater often exceeds
49 the threshold limit for drinking water, resulting in the use of additional treatment in drinking water
50 treatment plants (DWTPs) or closure of the affected abstraction wells. Biological treatment with
51 MSH1 is considered a potential sustainable alternative to remediate BAM-contamination in
52 drinking water production. Combining Illumina and Nanopore sequencing, we here present the
53 complete genome of MSH1, which was determined independently in two different laboratories.
54 Unexpectedly, divergences were observed between the two genomes, i.e. one of them lacked four
55 plasmids compared to the other. Besides the circular chromosome and the two previously
56 described plasmids involved in BAM catabolism pBAM1 (41 kb) and pBAM2 (54 kb), we observe
57 that the genome of MSH1 contains two megaplasmids pUSP1 (367 kb) and pUSP2 (366 kb) and
58 three smaller plasmids pUSP3 (97 kb), pUSP4 (64 kb), and pUSP5 (32 kb). The MSH1 substrain
59 from KU Leuven showed a reduced genome lacking plasmids pUSP2 and the three smaller

60 plasmids and was designated substrain MK1, whereas the variant with all plasmids was designated
61 as substrain DK1. Results of a plasmid stability experiment, indicate that strain MSH1 may have
62 a polyploid chromosome when growing in R2B medium with more chromosomes than plasmids
63 per cell. Based on phylogenetic analyses, strain MSH1 is reassigned as *Aminobacter niigataensis*
64 MSH1.

65 **Importance**

66 The complete genomes of the two MSH1 substrains, DK1 and MK1, provide further insight into
67 this already well-studied organism with bioremediation potential. The varying plasmid contents in
68 the two substrains suggest that some of the plasmids are unstable, although this is not supported
69 by the herein described plasmid stability experiment. Instead, results suggest that MSH1 is
70 polyploid with respect to its chromosome, at least under some growth conditions. As the essential
71 BAM-degradation genes are found on some of these plasmids, stable inheritance is essential for
72 continuous removal of BAM. Finally, *Aminobacter* sp. MSH1 is reassigned as *Aminobacter*
73 *niigataensis* MSH1, based on phylogenetic evidence.

74 **Keywords**

75 *Aminobacter*, 2,6-dichlorobenzamide, *Phyllobacteriaceae*, Nanopore sequencing, catabolic
76 plasmids

77 **Introduction**

78 The occurrence of organic micropollutants in different water compartments threatens both
79 ecosystem functioning as well as future drinking water supplies (1). Organic micropollutants are

80 organic chemicals with complex and highly variable structures, and they have in common that they
81 occur in the environment at trace concentrations (in the $\mu\text{g} - \text{ng/L}$ range). Organic micropollutants
82 often have unknown ecotoxicological and/or human health effects. They include a multitude of
83 compounds such as pharmaceuticals, pesticides, ingredients of household products and additives
84 of personal care products. In the European Union, the threshold limit for pesticides and relevant
85 transformation products in drinking water is set at 0.1 $\mu\text{g/L}$ (2). This threshold is frequently
86 exceeded and forces drinking water treatment plants to invest in expensive physicochemical
87 treatment technologies or to close groundwater extraction wells (3). The use of pollutant degrading
88 bacteria in bioaugmentation strategies to remove micropollutants, such as pesticides, from drinking
89 water, is presented as a solution (3,4). The groundwater micropollutant 2,6-dichlorobenzamide
90 (BAM), a transformation product of the herbicide dichlobenil, frequently occurs in groundwater
91 in Europe, often exceeding the threshold concentration (5). *Aminobacter* sp. MSH1 (CIP 110285)
92 was enriched and isolated from dichlobenil treated soil sampled from the courtyard of a plant
93 nursery in Denmark. The strain converts dichlobenil to BAM, which is further fully mineralized
94 (6). Efforts to elucidate the catabolic pathway for BAM degradation in MSH1 revealed the
95 involvement of two plasmids. The first step of BAM-mineralization involves the hydrolysis of
96 BAM to 2,6-dichlorobenzoic acid (2,6-DCBA) by the amidase BbdA encoded on the 41 kb IncP1-
97 β plasmid pBAM1 (7). Further catabolism of 2,6-DCBA to central metabolism intermediates
98 involves enzymes encoded on the 54 kb *repABC* family plasmid pBAM2. The strain mineralizes
99 BAM at trace concentrations (6) and invades biofilms of microbial communities of rapid sand
100 filters used in DWTPs (8). Moreover, it was successfully used in bioaugmentation of rapid sand
101 filters, both in lab scale and pilot scale biofiltration systems, to remove BAM from (ground)water
102 (8–11). On the other hand, long-term population persistence and catabolic activity in the sand

103 filters were impeded, likely due to a combination of predation and wash out (11, 12), as well as
104 to physiological and genetic changes. Reducing flow rate and improving inoculation strategy have
105 demonstrated prolonged persistence and activity of MSH1 in bioaugmented sand filters (13).
106 However, other studies indicate that MSH1 shows a starvation survival response, in the nutrient
107 (especially carbon) limiting environment of DWTPs, leading to reduced specific BAM degrading
108 activity (14). Moreover, a substantial loss of plasmid pBAM2 was observed upon prolonged
109 transfer of MSH1 both in R2A medium and in C-limited minimal medium (15), indicating that
110 the plasmid is not entirely stable. Moreover, mutants lacking the ability to convert BAM into 2,6-
111 DCBA have been reported (7). Clearly, to come to full management of bioaugmentation using
112 MSH1 in DWTP biofiltration units aiming at BAM removal, more knowledge is needed on the
113 physiological as well as genetic adaptations of MSH1 when introduced into the corresponding
114 oligotrophic environment. The elucidation of the full genome sequence is crucial in this.

115 The complete genome sequence presented in this study shows that MSH1 substrain DK1 has a
116 single chromosome and seven plasmids, including the two previously described catabolic plasmids
117 pBAM1 and pBAM2, while substrain MK1 lacks four of these plasmids. The relative sequence
118 coverage of the plasmids compared to the chromosome suggested that there are either multiple
119 copies of the chromosome per cell or that there are, on average, fewer than one copy of six out of
120 the seven plasmids per cell. This was tested in a plasmid stability experiment with substrain DK1
121 where plasmids were found to be overall stable, with the exception of a single loss event of pUSP1.
122 This supports the hypothesis that MSH1 might have a polyploid chromosome, at least under some
123 growth conditions.

124 **Material and Methods**

125 *Growth conditions, genomic DNA preparation and sequencing*

126 The genome sequence of strain MSH1 was independently obtained in two different laboratories,
127 i.e., the KU Leuven in Belgium (MK1) and the Aarhus University lab in Roskilde, Denmark (DK1).
128 In both cases, *Aminobacter* sp. MSH1 was obtained from the strain collection of the laboratory
129 that originally isolated the bacterium (6). Sequencing of substrain DK1 at the Roskilde lab was
130 performed as follows. Directly derived from a cryostock obtained from the original lab of MSH1,
131 two ml of a culture grown in R2B were used for extraction of high molecular weight (HMW) DNA
132 using the MasterPureTM DNA Purification Kit (Epicentre, Madison, WI, USA), using the kit's
133 protocol for cell samples. DNA was eluted in 35 µL 10 mM Tris-HCl (pH 7.5) with 50 mM NaCl.
134 The purity and concentration of extracted DNA were measured with a NanoDrop 2000c and a
135 Qubit[®] 2.0 fluorometer (Thermo Fisher Scientific, Walther, MA, USA), respectively. An Illumina
136 Nextera XT library was prepared for paired-end sequencing on an Illumina NextSeq 500 with a
137 Mid Output v2 kit (300 cycles) (Illumina Inc., San Diego, CA, USA). Paired-end reads (2x151 bp)
138 were trimmed for contaminating adapter sequences and low quality bases (<Q20) at the ends of
139 the reads were removed using Cutadapt (v1.8.3) (16). Paired-end reads that overlapped were
140 merged with AdapterRemoval (v2.1.0) (17). For Oxford Nanopore sequencing, a library was
141 prepared from the same DNA extract using the Rapid Sequencing kit (SQK-RAD004). This was
142 loaded on an R9.4 flow cell and sequenced using MinKnow (v1.11.5) (Oxford Nanopore
143 Technologies, Oxford, UK). Nanopore reads were basecalled with albacore (v2.1.10) without
144 quality filtering of reads. Only reads longer than 5,000 bp were retained and sequencing adapters
145 were trimmed using Porechop (v0.2.3). A hybrid genome assembly with Nanopore and Illumina
146 reads was performed using Unicycler (v0.4.3) (18).

147 The Illumina sequencing of substrain MK1 in the KU Leuven lab was reported previously (7, 19).
148 Briefly, genomic DNA was isolated from a culture grown on R2B using the Puregene Core kit A
149 (Qiagen, Hilden, Germany), according to the manufacturer's instructions, except that DNA
150 precipitation was performed with ethanol. A library was constructed for paired-end sequencing
151 using 500 bp inserts and sequencing was performed on the Illumina GAIIx platform. Generated
152 read lengths were 90 bp. The Illumina reads were quality controlled using FastQC (20) (v0.11.6)
153 and BBduk (21) (v36.47). This included trimming the reads with low scoring regions (Phred < 30),
154 clipping adapters, and removing very short reads (length < 50). For Nanopore sequencing, total
155 genomic DNA was extracted from a culture grown on R2B with 200 mg/L BAM using the DNeasy
156 UltraClean Microbial Kit (Qiagen, Hilden, Germany). Afterwards, the genomic DNA was
157 mechanically sheared using a Covaris g-Tube (Covaris Inc., MA, USA) to an average fragment
158 length of 8 kb. The library for sequencing was prepared using the 1D ligation approach with native
159 1D barcoding (SQK-LSK109) and sequenced on a MinION R9.4 flow cell using the MinION
160 sequencer (Oxford Nanopore Technologies, Oxford, UK). The Nanopore reads were basecalled
161 with Albacore (v2.0.2), and the barcode sequences were trimmed using Porechop (v0.2.3). Hybrid
162 assembly of genome was performed as reported above.

163 *Genome analyses*

164 For both genomes, automatic gene annotation was done using Prokka (22) (v1.14.0). Separately
165 from Prokka, proteins with transmembrane helices were identified using TMHMM v2.0 (23).
166 Genes were assigned to COG functional categories using EggNOG-mapper v4.5.1 (24). Genome
167 comparison was done using EDGAR (25). Metabolic pathways were explored using Pathway
168 Tools (26) and RAST (27). Circularized views of chromosome and plasmids were made using
169 Circos (28). MegaX (29) was used for protein alignment and tree building. Phylogenetic analysis

170 for strain MSH1 was performed using a clustal-omega (30) multiple sequence alignment using 16S
171 ribosomal RNA genes from the set of type strains available in the *Phyllobacteriaceae* family. The
172 tree was inferred using PhyML (31) with a GTR substitution model and a calculation of branch
173 support values (bootstrap value of 1,000). Whole-genome-based taxonomic classification was
174 performed with *in silico* DNA:DNA hybridization using the Type Strain Genome Server (TYGS)
175 (32). Furthermore, average nucleotide identity (ANI) values were calculated for MSH1 against all
176 available *Aminobacter* genomes in NCBI (downloaded January 31, 2021), using FastANI (33) and
177 plotted in R with the *pheatmap* package (34). Genomes of the two MSH1 substrains were
178 compared using the Mauve genome alignment software (35). Plasmids were characterized with
179 regards to relaxase genes and replicon families using MOB-suite (36).

180

181 *Plasmid (in)stability experiment*

182 To test for plasmid stability, MSH1 cells from -80°C cryostock were streaked on R2A plates and
183 DNA from 1 ml of the cryostock was extracted using MasterPure™ DNA Purification Kit. After
184 incubation at 22°C for 11 days, a single colony from the R2A plate was picked and resuspended
185 in 105 µl phosphate-buffered saline (PBS). From this, 5 µl suspension was inoculated in 25 ml
186 R2B for 72 hours. Whole genome sequencing was performed on the remaining 100 µl PBS
187 suspension. After 72 hours of growth in R2B, 1 mL broth culture was sampled for DNA extraction,
188 similarly to the DNA extracted for initial DNA sequencing (above), and 100 µL of dilution series
189 10⁻⁵-10⁻⁸ of the R2B culture were plated onto R2A and the plates incubated at 22°C. After 7 days
190 of growth, DNA was extracted and sequenced, as described above, from 14 individual colonies
191 (originating from a single cell) resuspended in 100 µL PBS. All sequencing was performed on an

192 Illumina NextSeq 550 with a Mid Output v2 kit (300 cycles) using Nextera XT library preparations
193 as described above.

194 Sequencing adapters and poor quality sequences were trimmed from paired end reads using
195 Trimmomatic (v0.39) (37) with the options
196 “ILLUMINACLIP:/usr/share/trimmomatic/NexteraPE-PE.fa:2:30:10 LEADING:3 TRAILING:3
197 SLIDINGWINDOW:4:15 MINLEN:36”. Trimmed and filtered reads from each replicate MSH1
198 sample were mapped with bwa (v0.7.17-r1198-dirty) (38) to the completely assembled MSH1
199 genome including plasmids pBAM1-2 and pUSP1-5. Sequencing coverage in 1,000 bp windows
200 for all replicons per replicate sample was calculated with samtools (v1.9-166-g74718c2) (39) and
201 bedtools (v2.28.0) (40). Coverage data for all replicons were divided by the mean coverage of the
202 chromosome, in order to normalize replicon copy numbers relative to the chromosome.
203 Normalized coverage of all replicons for all replicates were visualized with Circos (v0.69-6) (28).

204

205 *Data availability*

206 The genome sequences of strain MSH1 substrains MK1/DK1 are available under the following
207 GenBank accession numbers CP026265/CP028968 (chromosome), CP026268/CP028967
208 (pBAM1), CP026267/CP028966 (pBAM2), CP026266/CP028969 (pUSP1) and CP028970
209 (pUSP2), CP028971 (pUSP3), CP028972 (pUSP4) and CP028973 (pUSP5).

210 **Results and discussion**

211 *Genome statistics*

212 The MSH1 genome (based on substrain DK1) consists of a chromosome of 5,301,518 bp and seven
213 plasmids. The genome contains two large plasmids pUSP1 of 367,423 bp and pUSP2 of 365,485

214 bp, three smaller plasmids pUSP3, pUSP4, and pUSP5 (respectively 97,029 bp, 64,122 bp, and
215 31,577 bp) and the two previously reported smaller catabolic plasmids pBAM1 and pBAM2 of
216 40,559 bp and 53,893 bp, respectively (Table 1). A total of 6,257 genes could be predicted of
217 which six rRNAs, 53 tRNAs, and four ncRNAs. A total of 6,194 CDS were predicted including
218 190 pseudogenes (Table 2). Circular views of the chromosome and seven plasmids are shown in
219 Figure 1 and 2. The KU Leuven variant, designated as substrain MK1, lacked one of the two larger
220 plasmids, i.e. pUSP2, and the three smaller plasmids pUSP3, pUSP4, and pUSP5. Except for the
221 discrepancy in plasmids, the shared genomes (chromosome, pUSP1, pBAM1, and pBAM2) of the
222 two strains have an average nucleotide identity of 99.9925%. The BAM-catabolic genes were
223 manually checked for mutations that could indicate differences in degradation potential. A single
224 nucleotide change was noted in the *bbdb3* gene on pBAM2, encoding one of three subunits of a
225 TRAP-type transport system potentially involved in the uptake of 2,6-DCBA (19). In this gene, a
226 non-synonomous substitution has changed a glycine to an arginine in the resulting protein in MK1.
227 Currently, it is not known if this change has an effect on the putative function of this tripartite
228 transport system. Furthermore, differences were found in the region of plasmid pUSP1 containing
229 an IS30 family insertion sequence with 38 bp flanking, imperfect, inverted repeats (IRs). The
230 repeats appear complete in DK1, but MK1 shows a deletion of 56 bp and 34 bp up- and
231 downstream of the IS30 transposase gene, including partial deletion of the IR at both ends,
232 suggesting that the MK1 substrain has undergone further genetic changes. The deletions flanking
233 the IS30 element on pUSP1 in MK1 may have been caused by a possible intramolecular
234 transposition event (41). However, this IS30 element with deletion in the IRs in MK1 may still be
235 functional, as the functional core region of IS30 IRs are only part of the complete IR (42).

236

237 *Phylogenetic assignment of MSH1 to Aminobacter niigataensis*

238 A phylogenetic tree based on the 16S rRNA gene sequence indicating the position of MSH1 is
239 shown in Figure 3. The 1,463 bp 16S rRNA gene sequence of MSH1 is 100% identical to that of
240 *Aminobacter niigataensis* DSM 7050^T and 99.6-99.8% to those of other *Aminobacter* species. This
241 is supported by whole-genome *in silico* digital DNA:DNA hybridization using TYGS, which
242 reports that MSH1 (substrain DK1) is 82.5% (recommended d_4 formula) similar to *A. niigataensis*
243 DSM 7050. (Supplementary Table S1). Finally, ANI values against all available *Aminobacter*
244 genomes from NCBI (complete and incomplete assemblies; downloaded January 31, 2021),
245 showed an ANI of 98% against *A. niigataensis* DSM 7050 (Figure 4). Based on these analyses, we
246 reassign *Aminobacter* sp. MSH1 as *Aminobacter niigataensis* MSH1.

247

248 *Chromosomally encoded metabolic features of MSH1*

249 The chromosome of MSH1 possesses all genes required for glycolysis using the Embden-
250 Meyerhof pathway and additionally possesses all genes for glucose metabolism through the
251 Entner-Doudoroff pathway and the pentose phosphate pathway. It also contains all genes of the
252 tricarboxylic acid cycle. MSH1 was previously shown to grow slower on succinate and acetic acid
253 as carbon sources compared to glucose, fructose, and glycerol (43). MSH1 does not possess genes
254 involved in carbon fixation which rules out autotrophic growth. MSH1 further displays the
255 catechol *ortho*-cleavage pathway (44) and possesses genes for conversion of benzoate to catechol
256 allowing the organism to grow on benzoate which was confirmed by culturing the strain on
257 benzoate (data not shown). With regards to nitrogen metabolism, MSH1 contains a gene cluster

258 that encodes the transmembrane ammonium channel AmtB as well as its cognate protein GlnK
259 (45) for controlling ammonium influx in response to the intracellular nitrogen status, indicating
260 that MSH1 can use mineral ammonia as a nitrogen source directly from its environment. In
261 addition, MSH1 encodes for proteins involved in nitrate transport (NrtA and NrtT). The
262 corresponding genes are located upstream of genes for assimilatory nitrate reduction (*nasDEA*) to
263 ammonium suggesting that MSH1 can also use nitrate as a nitrogen source. Finally, ammonia is
264 also released from amino acid metabolism and is further incorporated in L-glutamate for
265 biosynthesis. Furthermore, MSH1 contains a gene cluster which combines a periplasmic
266 dissimilatory nitrate reductase (*napAB*), the membrane-bound cytochrome *c* (*napC*) that is
267 involved in electron transfer from the quinol pool in the cytoplasmic membrane to NapAB, *nirK*
268 (nitrate reductase) and *norBC* (nitrix oxide reductase). However, *narG*, encoding the cytoplasmatic
269 oriented dissimilatory nitrate reductase, is lacking. Dissimilatory nitrate reductases are associated
270 with the cell membrane, and are typically involved in energy acquisition, detoxification, and redox
271 regulation (46) NarG, located at the cytoplasmatic side of the cell membrane, is the typical
272 respiratory nitrate reductase though its function can be replaced by NapAB in cases coupled to
273 formate oxidation (46). However, this is unlikely since MSH1 is unable to grow under nitrate
274 reducing conditions (6). In addition, *nosZ* for reduction of nitrous oxide to dinitrogen (47) is
275 missing. The exact function of the gene cluster containing *napABC*, *nirK* and *norBC* is yet
276 unknown. Besides direct uptake, for sulfur metabolism, MSH1 possesses two nearby located gene
277 clusters encoding the ABC transporter complex CysUWA involved in sulfate/thiosulfate import.
278 One of the two clusters appears directed to the uptake of sulfate while the other to thiosulfate
279 uptake, since they respectively are linked with *sbp* and *cysP* (48). Both genes encode for the
280 periplasmic protein that delivers sulfate or thiosulfate to the ABC transporter for high affinity

281 uptake but Sbp binds sulfate and CysP thiosulfate. Furthermore, the chromosome contains all
282 genes necessary for assimilatory sulfate reduction. The pathway reduces sulfate to sulfide
283 involving ATP sulfurylase (CysND), adenosine 5'-phosphosulfate reductase (CysC), 3'-
284 phosphoadenosine-5'-phosphosulfate reductase (CysH) and sulfite reductase (CisIJ). In addition,
285 MSH1 contains *cysK* encoding O-acetylserine sulfhydrylase that incorporates sulfide into O-
286 acetylserine to form cysteine (48). The assimilation of thiosulfate is less clear but MSH1 encodes
287 for another homologue of CysK as well as several glutaredoxin proteins that are needed to
288 incorporate thiosulfate in O-acetylserine and reductive cleavage reaction of its disulfide bond to
289 form cysteine (48).

290

291 *Plasmids of MSH1*

292 Besides the previously described IncP1-β and *repABC* plasmids, pBAM1 and pBAM2 (7, 19),
293 substrain DK1 harbors the five pUSP1-5 plasmids (Figure 2), while substrain MK1 lacks pUSP2,
294 pUSP3, pUSP4, and pUSP5. Catabolic genes on pBAM1 and pBAM2 enable MSH1 to mineralize
295 the groundwater micropollutant BAM and use it as a source of carbon, nitrogen, and energy for
296 growth. The amidase BbdA on pBAM1 transforms BAM to 2,6-dichlorobenzoic acid (DCBA) (7)
297 which is further metabolized by a series of catabolic enzymes encoded by pBAM2 (19, 49). As
298 previously discussed (19), the gene *bbdI* encoding the glutathione dependent thiolytic
299 dehalogenase responsible for removal of one of the chlorines from BAM together with *bbdJ*
300 encoding glutathione reductase, occur on pBAM2 in three consecutive, perfect repeats followed
301 by a fourth, imperfect repeat. This, together with the placement of the BAM degradation genes on
302 two separate plasmids (pBAM1 and pBAM2) and the bordering of the catabolic gene clusters by
303 remnants of insertion sequences and integrase genes, suggests that the BAM catabolic genes in

304 MSH1 have been acquired by horizontal gene transfer and then evolved to occur in their observed
305 genomic organisation. In addition, pBAM2 has a considerably lower GC content of 56% compared
306 to the chromosome and other plasmids which are between 60.0 and 64.4% (Table 2), which could
307 indicate that pBAM2 was acquired from another, unknown, unrelated bacterium. It was previously
308 shown that mineralization of DCBA is a common trait in bacteria in sand filters and soils, while
309 BAM to DCBA conversion is the rate limiting step in BAM mineralization and is rare in microbial
310 communities (50).

311 Like pBAM2, plasmids pUSP1, pUSP2, and pUSP3 belong to the *repABC* family. *repABC*
312 replicons are known as typical genome components of *Alphaproteobacteria* species (51). The
313 occurrence of more than one *repABC* replicon in one and the same genome has been described
314 before and the plasmid family has been shown to exist of different incompatibility groups. For
315 instance, *Rhizobium etli* CFN42 has 6 *repABC* plasmids (52, 53).

316 Plasmids pBAM2, pUSP2, pUSP3, and pUSP4 contains Type IV secretion system (T4SS) genes
317 (54), while pUSP1 does not. This indicates that pUSP1 is likely not self-transferable, unlike
318 pBAM2, pUSP2, pUSP3, and pUSP4. Besides T4SS genes, plasmid pUSP4 contains a *mobABC*
319 operon. The 31.6 kbp plasmid pUSP5 lacks conjugative transfer genes and appears to be a
320 mobilizable plasmid with genes encoding a VirD4-like coupling protein and a TraA conjugative
321 transfer relaxase likely involved in nicking at an *oriT* site and unwinding DNA before transfer.
322 Furthermore, MOB-suite predicted that pBAM1, pUSP2, and pUSP4 have MOBP-type relaxase
323 genes, while pUSP3 and pUSP5 have MOBQ-type relaxase genes. pBAM2 and pUSP1 were not
324 predicted to have MOB-related genes.

325

326

327 *Specialized functions of plasmids pUSP1-5*

328 In Table 3, all CDS of the different plasmids are categorized according to COGs. Half of the CDS
329 annotated on plasmid pUSP1 (322 CDS) and pUSP2 (346 CDS) are genes primarily associated
330 with the transport and metabolism of amino acids (20% and 12%, resp.), carbohydrates (6% and
331 6%, resp.) and inorganic compounds (10% and 3%, resp.), and genes for energy production and
332 conversion (9% and 8%, resp.). For the plasmids pUSP3, pUSP4, and pUSP5, CDS categorized
333 under the same COGs are lower than 18%. Together, pUSP1 and pUSP2 accounts for about 17%
334 of all genes in MSH1 related to amino acid, carbohydrate transport and metabolism, and energy
335 production and conversion in MSH1. The transport systems encoded by pUSP1 and pUSP2 include
336 multiple ABC-transporters for N and/or S-containing organic compounds. For amino acids,
337 carbohydrates and inorganic compound metabolism and transport, ABC-type transport systems are
338 predicted for polar amino acids (arginine, glutamine), branched chain amino acids, and multiple
339 sugars. In addition, transport systems for spermidine/putrescine, taurine, aliphatic sulphonates,
340 dipeptides, beta-methyl galactoside, polysialic acid, and phosphate were predicted. Putative
341 functions could be assigned by Prokka to 64.2%, 56.8%, 27.4%, 42.2%, and 34.3% of CDS for
342 pUSP1, pUSP2, pUSP3, pUSP5 and pUSP5, respectively. On pUSP1, found in both MSH1
343 substrains, multiple genes could be assigned to metabolic subsystems by RAST. These include
344 folate biosynthesis, cytochrome oxidases and reductases, degradation of aromatic compounds
345 (homogentisate pathway), ammonia assimilation, and several genes related to amino acid
346 metabolism. Some of these functions on pUSP1 do not have functional analogs on the chromosome,
347 which may help to explain why pUSP1 was not lost in substrain MK1, but the other pUSP plasmids
348 were. On pUSP2, which is absent in MK1 substrain, some genes are predicted to be involved in
349 acetyl-CoA fermentation to butyrate, creatine degradation, metabolism of butanol, fatty acids, and

350 nitrile, and a few miscellaneous functions. A large number of CDS on pUSP1 (19%), and pUSP2
351 (23%) are homologues to CDS on the chromosome and could be considered dispensable genes.
352 However, although these CDS might be considered homologues, their functionality might differ
353 considerably in terms of substrate specificity and kinetics.
354 Besides genes encoding conjugative transfer, plasmid replication, and plasmid stability
355 functions, most genes on plasmids pUSP3, pUSP4, and pUSP5 could not be annotated with a
356 function. However, several genes on pUSP3 may have functions related to metabolism of sugars,
357 including inositol and mannose which were not tested in an earlier growth optimization
358 experiment (43). On pUSP4, genes encoding a transmembrane amino acid transporter are
359 situated next to an aspartate ammonia-lyase-encoding gene that enables conversion between
360 aspartate and fumarate that may enter the tricarboxylic acid cycle, as described above. A
361 cytochrome bd-type quinol oxidase, encoded by two subunit genes on pUSP5, also occurs in
362 some nitrogen-fixing bacteria where it is responsible for removing oxygen in microaerobic
363 conditions. Furthermore, a pseudoazurin type I blue copper electron-transfer protein is encoded
364 by a gene on pUSP5, that may act as an electron donor in a denitrification pathway. A chromate
365 transporter, ChrA, encoded by a gene on pUSP5 may confer resistance to chromate. Future
366 studies should look into whether the lack of plasmids pUSP2-5 in substrain MK1 has phenotypic
367 consequences, with regards to the predicted functions, including metabolism of sugars and
368 aspartate, nitrogen metabolism, and resistance to chromate.

369

370

371

372 *Plasmid stability and chromosome ploidy*

373 The Illumina sequencing coverage of several plasmids relative to the chromosome (except for
374 pBAM1) was lower than one, i.e., approximately 0.3 to 0.6 per chromosome. This suggests that
375 either not all cells (only three to six out of ten) contain a copy of the same plasmid due to plasmid
376 loss or that there are multiple copies of the chromosome. Previously, in the MK1 substrain, we
377 observed that pBAM2 is not always perfectly inherited by the daughter cells in cultures grown in
378 R2B and R2B containing BAM (15). To observe whether plasmid instability explained the copy
379 number relative to the chromosome in the sequenced cultures, sequencing was performed directly
380 on the cryo stock as well as on colonies directly derived from this, mimicking the sequenced cell
381 preparation for whole genome sequencing. We hypothesized that if certain plasmids are not stably
382 inherited (i.e. those with copy numbers 0.3 to 0.6), only part of the cell population will harbour
383 those plasmids and picking of multiple colonies from a plate will result in picking of some colonies
384 that have lost one or more plasmids.

385 MSH1 was sequenced directly from the cryostock, from a single colony picked from R2A plates
386 after spreading the cryostock, and from the broth R2B culture that had been inoculated with the
387 same single colony from cryostock. Moreover, after spreading the latter R2B culture on an R2A
388 plate, an additional 14 MSH1 colonies were picked for sequencing. Taking into account a plasmid
389 coverage of 0.3 – 0.6 per chromosome, we expect that around half of the colonies would have lost
390 one or more of the plasmids in the case of poor inheritance. However, only one of the colonies
391 showed loss of a plasmid, i.e., plasmid pUSP1 (Figure 5) indicating polypliody of the chromosome
392 rather than unstable inheritance of plasmids. The loss of *repABC* megaplasmid pUSP1 shows that
393 the possible metabolic features encoded by genes on pUSP1, as described above, are not essential
394 for growth under these conditions, although, remarkably this is the only pUSP plasmid still present

395 in substrain MK1. Interestingly, the plasmid/chromosome-ratio varied according to the growth
396 medium from which DNA was isolated. When growing in R2B (broth), e.g. as done for DNA
397 extraction for genome sequencing and from cryostock and R2B culture (first and third green rings,
398 Figure 6), all plasmids, except pBAM1, have a copy number lower than one per chromosome.
399 When DNA was extracted from colonies grown on R2A plates (though resuspended in PBS prior
400 to DNA extraction), plasmid copy numbers were approx. one per chromosome, except for pBAM1
401 which has a copy number of approx. 2.5 per chromosome.

402 Except for the single loss of pUSP1, nothing here indicates unstable maintenance of plasmids and
403 subsequent loss. Instead, our results indicate that MSH1 regulates the chromosome copy number
404 according to whether it grows as planktonic bacteria or fixed on an agar plate. The results shown
405 here can be explained by MSH1 being polyploid with regards to its chromosome when growing in
406 broth media. Single-copy plasmids (e.g. pBAM2, pUSP1-5) will thereby have copy numbers lower
407 than one relative to the chromosome, when growing in broth R2B. Polyploidy in prokaryotes have
408 been described before, including in *Deinococcus*, *Borrelia*, *Azotobacter*, *Neisseria*, *Buchnera*, and
409 *Desulfovibrio* (55) and may be quite overlooked in many other bacteria. *E. coli* in stationary phase
410 was shown to have two chromosome copies after growing in rich, complex medium, but only 60%
411 of the cells had two copies in stationary phase after slower growth in a synthetic medium (55). It
412 was suggested that monoploidy is not typical for proteobacteria, and that many bacteria are
413 polyploid when growing in exponential phase (55). Possible advantages offered by polyploidy
414 include resistance to DNA damage and mutations, global regulation of gene expression by
415 changing chromosome copy number, and finally polyploidy may enable heterozygosity in bacteria
416 where genes mutate to cope with challenging condition while preserving a copy of the original
417 genes. Despite the stability of the plasmids in MSH1, the MSH1 substrain MK1 lacks plasmids

418 pUSP2-5 and a loss of pBAM2 was previously observed (15). Although pBAM2 encodes its own
419 T4SS, the multiple loss of pBAM2 and pUSP2-5 in MK1 could be hypothetically explained by
420 some uncharacterized plasmid codependence, where one loss leads to another. The dynamics of
421 plasmid loss that has led to formation of substrain MK1 are still unknown.

422 **Conclusions**

423 The full genome of *Aminobacter* sp. MSH1, re-identified here as *Aminobacter niigataensis* MSH1,
424 consisting of a chromosome and seven plasmids, was determined combining both Nanopore and
425 Illumina sequencing. Two smaller plasmids pBAM1 and pBAM2 were previously identified
426 carrying the catabolic genes required for mineralization of the groundwater micropollutant BAM.
427 Both the chromosome and the other five plasmids are described here for the first time. A plasmid
428 stability experiment showed that most plasmids were stably maintained, with exception of a single
429 loss event of plasmid pUSP1. Instead, the results indicate that MSH1 has a polyploid chromosome
430 when growing in broth, thereby reducing plasmid copy numbers per chromosome to below one.
431 When comparing the original strain MSH1 (DK1) and substrain MK1, we observed that plasmids
432 pUSP2, pUSP3, pUSP4, and pUSP5 were below detection limits in MK1. Substrain MK1 may
433 previously have lost these plasmids but maintained pUSP1, pBAM1, and pBAM2, thereby
434 retaining its capacity to degrade BAM. Future studies on growth and degradation kinetics of the
435 MSH1 and its substrain MK1 lacking several plasmids, can reveal if plasmids pUSP2-5 harbour
436 unknown (favorable) functions or if they impose a metabolic burden on MSH1. This will help to
437 elucidate which substrain is preferable for bioaugmentation.

438 **Funding**

439 This project (LHH, LEJ, OH, and JAA) was funded by MEM2BIO (Innovation Fund Denmark,
440 contract number 5157-00004B) at Aarhus University and DS by the KU Leuven C1 project no.
441 C14/15/043 and the BELSPO IAP-project μ -manager no. P7/25 at KU Leuven. BH was supported
442 by postdoctoral fellowship grant 12Q0218N and CL by an SB-FWO PhD fellowship 1S64718N.
443 TKN was supported by AUUFF NOVA project ORIGENE.

444 **Authors' contributions**

445 LHH, TKN, BH and DS initiated, planned, and funded the project. BH and TKN wrote the initial
446 draft manuscript. BH, TKN, and CL performed genome sequencing, assembly annotation, and data
447 analyses. OH performed the plasmid stability experiment. VvN, RL, DS, LEJ, JA, and LHH
448 critically reviewed the paper, assisted in data interpretation, and coordinated experiments.

449 **References**

- 450 1. Schwarzenbach RP, Escher BI, Fenner K, Hofstetter TB, Johnson CA, von Gunten U,
451 Wehrli B. 2006. The challenge of micropollutants in aquatic systems. *Science* (80-)
452 313:1072–1077.
- 453 2. EU. 1998. Council Directive 98/83/EC of 3 November 1998 on the quality of water
454 intended for human consumption. Counc Eur Union.
- 455 3. Benner J, Helbling DE, Kohler HP, Wittebol J, Kaiser E, Prasse C, Ternes TA, Albers
456 CN, Aamand J, Horemans B, Springael D, Walravens E, Boon N. 2013. Is biological
457 treatment a viable alternative for micropollutant removal in drinking water treatment
458 processes? *Water Research* 47(16):5955-5976

459 4. Vandermaesen J, Horemans B, Bers K, Vandermeeren P, Herrmann S, Sekhar A,
460 Seuntjens P, Springael D. 2016. Application of biodegradation in mitigating and
461 remediating pesticide contamination of freshwater resources: state of the art and
462 challenges for optimization. *Appl Microbiol Biotechnol* 100:7361-7376.

463 5. Björklund E, Anskjær GG, Hansen M, Styrihave B, Halling-Sørensen B. 2011. Analysis
464 and environmental concentrations of the herbicide dichlobenil and its main metabolite 2,6-
465 dichlorobenzamide (BAM): A review. *Sci Total Environ* 409:2343-2356.

466 6. Sørensen SR, Holtze MS, Simonsen A, Aamand J. 2007. Degradation and mineralization
467 of nanomolar concentrations of the herbicide dichlobenil and Its persistent metabolite 2,6-
468 dichlorobenzamide by *Aminobacter* spp. isolated from dichlobenil-treated soils. *Appl*
469 *Environ Microbiol* 73:399–406.

470 7. T'Syen J, Tassoni R, Hansen L, Sorensen SJ, Leroy B, Sekhar A, Wattiez R, De Mot R,
471 Springael D. 2015. Identification of the amidase BbdA that initiates biodegradation of the
472 groundwater micropollutant 2,6-dichlorobenzamide (BAM) in *Aminobacter* sp. MSH1.
473 *Environ Sci Technol* 49:11703–11713.

474 8. Horemans B, Raes B, Vandermaesen J, Simanjuntak Y, Brocatus H, T'Syen J, Degryse J,
475 Boonen J, Wittebol J, Lapanje A, Sorensen SR, Springael D. 2016. Biocarriers improve
476 bioaugmentation efficiency of a rapid sand filter for the treatment of 2,6-
477 dichlorobenzamide (BAM)-contaminated drinking water. *Environ Sci Technol* 51:1616–
478 1625.

479 9. Albers CN, Feld L, Ellegaard-Jensen L, Aamand J. 2015. Degradation of trace
480 concentrations of the persistent groundwater pollutant 2,6-dichlorobenzamide (BAM) in

481 bioaugmented rapid sand filters. *Water Res* 83:61–70.

482 10. Albers CN, Jacobsen OS, Aamand J. 2013. Using 2,6-dichlorobenzamide (BAM)
483 degrading *Aminobacter* sp. MSH1 in flow through biofilters—initial adhesion and BAM
484 degradation potentials. *Appl Microbiol Biotechnol* 98:957–967.

485 11. Ellegaard-Jensen L, Albers CN, Aamand J. 2016. Protozoa graze on the 2,6-
486 dichlorobenzamide (BAM)-degrading bacterium *Aminobacter* sp. MSH1 introduced into
487 waterworks sand filters. *Appl Microbiol Biotechnol* 1–9.

488 12. Ellegaard-Jensen L, Horemans B, Raes B, Aamand J, Hansen LH. 2017. Groundwater
489 contamination with 2,6-dichlorobenzamide (BAM) and perspectives for its microbial
490 removal. *Appl Microbiol Biotechnol* 101(13):5235-5245.

491 13. Ellegaard-Jensen L, Schostag MD, Nikbakht Fini M, Badawi N, Gobbi A, Aamand J,
492 Hansen LH. 2020. Bioaugmented Sand Filter Columns Provide Stable Removal of
493 Pesticide Residue From Membrane Retentate. *Front. Water*.

494 14. Sekhar A, Horemans B, Aamand E, Vanhaeke L, Hofkens J, Springael D. 2013.
495 Evaluation of the biofilm forming capacity of the 2, 6-dichlorobenzamide (BAM)
496 degrading *Aminobacter* sp. strain MSH1 at macropollutant and micropollutant BAM
497 concentrations. *Commun Agric Appl Biol Sci*. 2013;78(1):31-6

498 15. Horemans B, Raes B, Brocatus H, T'Syen J, Rombouts C, Vanhaecke L, Hofkens J,
499 Springael D. 2017. Genetic (in)stability of 2,6-dichlorobenzamide catabolism in
500 *Aminobacter* sp. strain MSH1 biofilms under carbon starvation conditions. *Appl Environ
501 Microbiol* 83(11):e00137-17.

502 16. Martin M. 2014. Cutadapt removes adapter sequences from high-throughput sequencing

503 reads. *EMBnet.journal*.

504 17. Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: Rapid adapter trimming,
505 identification, and read merging. *BMC Res Notes* 9:88.

506 18. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome
507 assemblies from short and long sequencing reads. *PLoS Comput Biol* 13(6):e1005595.

508 19. T'Syen J, Raes B, Horemans B, Tassoni R, Leroy B, Lood C, van Noort V, Lavigne R,
509 Wattiez R, Kohler H-PE, Springael D. 2018. Catabolism of the groundwater
510 micropollutant 2,6-dichlorobenzamide beyond 2,6-dichlorobenzoate is plasmid encoded in
511 *Aminobacter* sp. MSH1. *Appl Microbiol Biotechnol* 102:7963–7979.

512 20. Andrews S. 2010. FastQC - A quality control tool for high throughput sequence data.
513 <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Babraham Bioinformatics.

514 21. Bushnell B, Rood J, Singer E. 2017. BBMerge – Accurate paired shotgun read merging
515 via overlap. *PLoS One* 12(10):e0185056.

516 22. Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*.

517 23. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane
518 protein topology with a hidden markov model: application to complete genomes. *J Mol
519 Biol* 305:567–580.

520 24. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P.
521 2017. Fast Genome-Wide Functional Annotation through Orthology Assignment by
522 eggNOG-Mapper. *Mol Biol Evol* 34:2115–2122.

523 25. Blom J, Albaum SP, Doppmeier D, Pühler A, Vorhölter F-J, Zakrzewski M, Goesmann A.
524 2009. EDGAR: A software framework for the comparative analysis of prokaryotic

525 genomes. *BMC Bioinformatics* 10:154.

526 26. Karp PD, Latendresse M, Paley SM, Krummenacker M, Ong QD, Billington R, Kothari
527 A, Weaver D, Lee T, Subhraveti P, Spaulding A, Fulcher C, Keseler IM, Caspi R. 2016.
528 Pathway Tools version 19.0 update: software for pathway/genome informatics and
529 systems biology. *Brief Bioinform* 17:877–890.

530 27. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S,
531 Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil
532 LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O,
533 Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: Rapid annotations using
534 subsystems technology. *BMC Genomics* 9:75.

535 28. Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra
536 MA. 2009. Circos: An information aesthetic for comparative genomics. *Genome Res*
537 19:1639-1645.

538 29. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary
539 Genetics Analysis across Computing Platforms. *Mol Biol Evol* 35:1547–1549.

540 30. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H,
541 Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of
542 high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*
543 7:539.

544 31. Lethiec F, Gascuel O, Duroux P, Guindon S. 2005. PHYML Online—a web server for fast
545 maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* 33:W557–W559.

546 32. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for

547 state-of-the-art genome-based taxonomy. *Nat Commun* 10:2182.

548 33. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High
549 throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries.
550 *Nat Commun* 9:5114.

551 34. Kolde R. 2015. pheatmap : Pretty Heatmaps. R package version 1.0.12. R Package version
552 108.

553 35. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: Multiple alignment of
554 conserved genomic sequence with rearrangements. *Genome Res* 7:1394-403.

555 36. Robertson J, Nash JHE. 2018. MOB-suite: software tools for clustering, reconstruction
556 and typing of plasmids from draft assemblies. *Microb genomics* 4:8.

557 37. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina
558 Sequence Data. *Bioinformatics* 30(15):2114-2120.

559 38. Li H.; Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler
560 transform. *Bioinformatics* 25:1754–1760.

561 39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G,
562 Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
563 25(16):2078-9.

564 40. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
565 features. *Bioinformatics* 26:841–842.

566 41. Olasz F, Stalder R, Arber W. 1993. Formation of the tandem repeat (IS 30)2 and its role in
567 IS 30-mediated transpositional DNA rearrangements. *MGG Mol Gen Genet* 239:177-187.

568 42. Olasz F, Farkas T, Kiss J, Arini A, Arber W. 1997. Terminal inverted repeats of insertion
569 sequence IS30 serve as targets for transposition. *J Bacteriol* 179(23):7551-8.

570 43. Schultz-Jensen N, Knudsen BE, Frkova Z, Aamand J, Johansen T, Thykaer J, Sørensen
571 SR. 2014. Large-scale bioreactor production of the herbicide-degrading *Aminobacter* sp.
572 strain MSH1. *Appl Microbiol Biotechnol* 98:2335-2344.

573 44. Harayama S, Mermod N, Rekik M, Lehrbach PR, Timmis KN. 1987. Roles of the
574 divergent branches of the meta-cleavage pathway in the degradation of benzoate and
575 substituted benzoates. *J Bacteriol* 169(2):558-64.

576 45. Zheng L, Kostrewa D, Bernèche S, Winkler FK, Li XD. 2004. The mechanism of
577 ammonia transport based on the crystal structure of AmtB of *Escherichia coli*. *Proc Natl
578 Acad Sci U S A* 101(49):17090-17095.

579 46. Sparacino-Watkins C, Stolz JF, Basu P. 2014. Nitrate and periplasmic nitrate reductases.
580 *Chem Soc Rev* 43(2):676-706.

581 47. Bedzyk L, Wang T, Ye RW. 1999. The periplasmic nitrate reductase in *Pseudomonas* sp.
582 strain G-179 catalyzes the first step of denitrification. *J Bacteriol* 181(9):2802-6.

583 48. Kawano Y, Suzuki K, Ohtsu I. 2018. Current understanding of sulfur assimilation
584 metabolism to biosynthesize l-cysteine and recent progress of its fermentative
585 overproduction in microorganisms. *Appl Microbiol Biotechnol* 102(19):8203-8211.

586 49. Raes B, Horemans B, Rentsch D, T'Syen J, Ghequire M, De Mot R, Wattiez R, Kohler H-
587 P, Springael D. 2019. *Aminobacter* sp. MSH1 mineralises the groundwater micropollutant
588 2,6-dichlorobenzamide through a unique chlorobenzoate catabolic pathway. *Environ Sci
589 Technol* 53:10146–10156.

590 50. Vandermaesen J, Horemans B, Degryse J, Boonen J, Walravens E, Springael D. 2016.
591 Mineralization of the Common Groundwater Pollutant 2,6-Dichlorobenzamide (BAM)
592 and its Metabolite 2,6-Dichlorobenzoic Acid (2,6-DCBA) in Sand Filter Units of Drinking
593 Water Treatment Plants. *Environ Sci Technol* 50(18):10114-10122.

594 51. Cevallos MA, Cervantes-Rivera R, Gutiérrez-Ríos RM. 2008. The *repABC* plasmid
595 family. *Plasmid* 60(1):19-37.

596 52. González V, Santamaría RI, Bustos P, Hernández-González I, Medrano-Soto A, Moreno-
597 Hagelsieb G, Janga SC, Ramírez MA, Jiménez-Jacinto V, Collado-Vides J, Dávila G.
598 2006. The partitioned *Rhizobium etli* genome: Genetic and metabolic redundancy in seven
599 interacting replicons. *Proc Natl Acad Sci U S A* 103(10):3834-3839.

600 53. Castillo-Ramírez S, Vázquez-Castellanos JF, González V, Cevallos MA. 2009. Horizontal
601 gene transfer and diverse functional constraints within a common replication-partitioning
602 system in Alphaproteobacteria: The *repABC* operon. *BMC Genomics* 10:536.

603 54. Cascales E, Christie PJ. 2003. The versatile bacterial type IV secretion systems. *Nat Rev
604 Microbiol* 1:137–149.

605 55. Pecoraro V, Zerulla K, Lange C, Soppa J. 2011. Quantification of Ploidy in Proteobacteria
606 Revealed the Existence of Monoploid, (Mero-)Oligoploid and Polyploid Species. *PLoS
607 One* 6:e16392.

608 56. Albers P, Lood C, Özturk B, Horemans B, Lavigne R, van Noort V, De Mot R, Marchal
609 K, Sanchez-Rodriguez A, Springael D. 2018. Catabolic task division between two near-
610 isogenic subpopulations co-existing in a herbicide-degrading bacterial consortium:
611 consequences for the interspecies consortium metabolic model. *Environ Microbiol* 20:85–

612 96.

613

614 **Tables**

615 **Table 1.** Genome accession codes

Label	Size (Mb)	GC (%)	Topology	INSDC identifier	RefSeq ID
Chromosome	5.30	63.2	Circular	CP028968.1(CP026265.1)*	NZ_CP028968.1(NZ_CP026265.1)*
Plasmid 1 pBAM1	0.04	64.4	Circular	CP028967.1(CP026268.1)*	NZ_CP028967.1(NZ_CP026268.1)*
Plasmid 2 pBAM2	0.05	56.0	Circular	CP028966.1(CP026267.1)*	NZ_CP028966.1(NZ_CP026267.1)*
Plasmid 3 pUSP1	0.37	63.1	Circular	CP028969.1(CP026266.1)*	NZ_CP028969.1(NZ_CP026266.1)*
Plasmid 4 pUSP2	0.37	60.1	Circular	CP028970.1	NZ_CP028970.1
Plasmid 5 pUSP3	0.10	60.5	Circular	CP028971.1	NZ_CP028971.1
Plasmid 6 pUSP4	0.06	61.9	Circular	CP028972.1	NZ_CP028972.1
Plasmid 7 pUSP5	0.03	62.9	Circular	CP028973.1	NZ_CP028973.1

616 * INSDC identifier and RefSeqID of KU Leuven substrain MK1 submission in brackets

617

618 **Table 2.** Genome statistics based on substrain MK1.

Attribute	Value	% of Total
Genome size (bp)	6321606	100.0
DNA coding (bp)	5587258	88.4
DNA G+C (bp)	3976162	62.9
DNA scaffolds	8	100.0
Total genes	6257	100.0
Protein coding genes	6004	96.0
RNA genes	63	1.0
Pseudo genes	190	3.0
Genes with function prediction	5182	75.9
Genes assigned to COGs	3890	62.2

Genes with Pfam domains	5006	8190
Genes with signal peptides	565	9.0
Genes with transmembrane helices	1423	620 22.7
CRISPR repeats	0	0 621

622

623

624

625

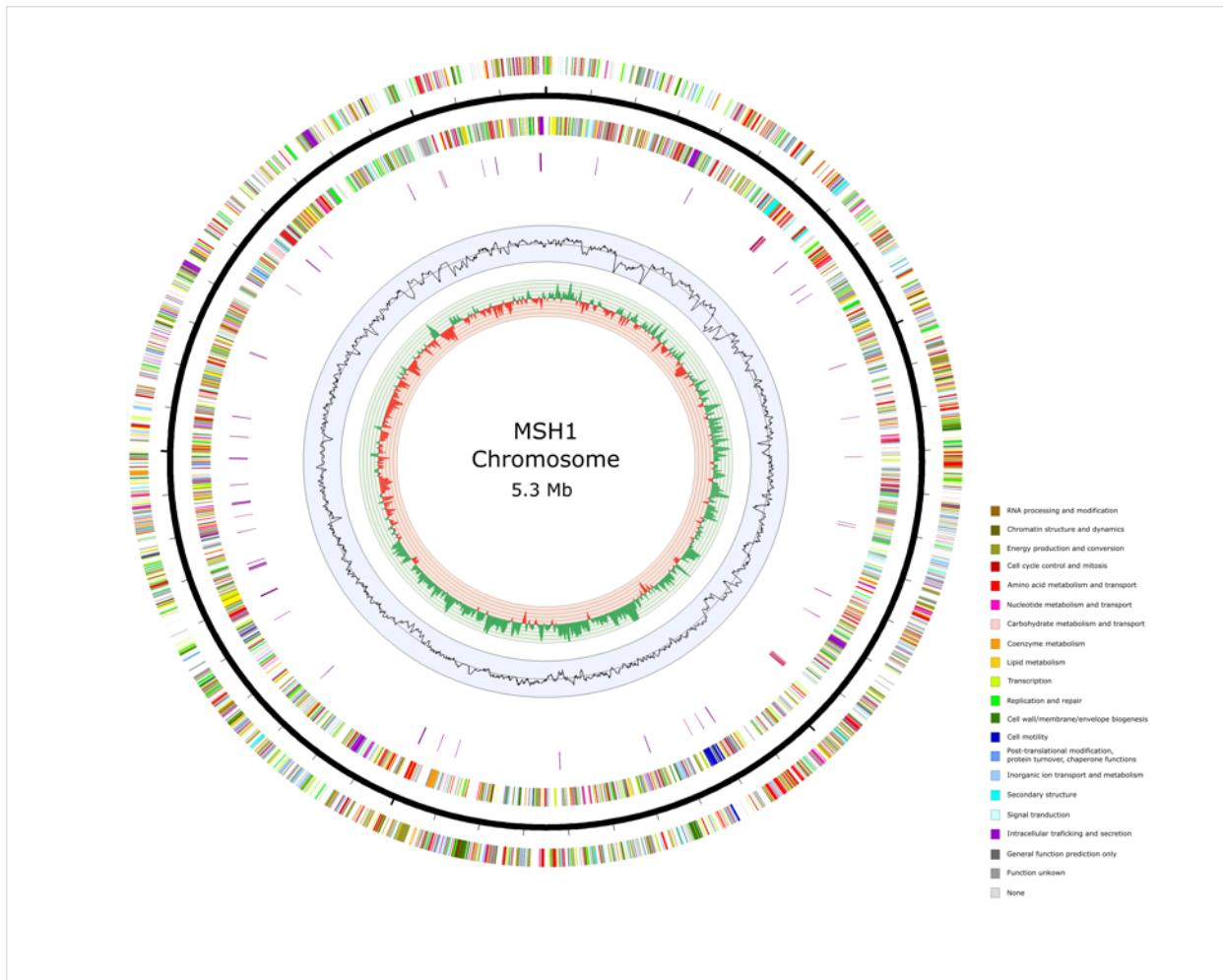
626

627

Table 3. Percentage of genes associated with general COG functional categories in genome and replicons.

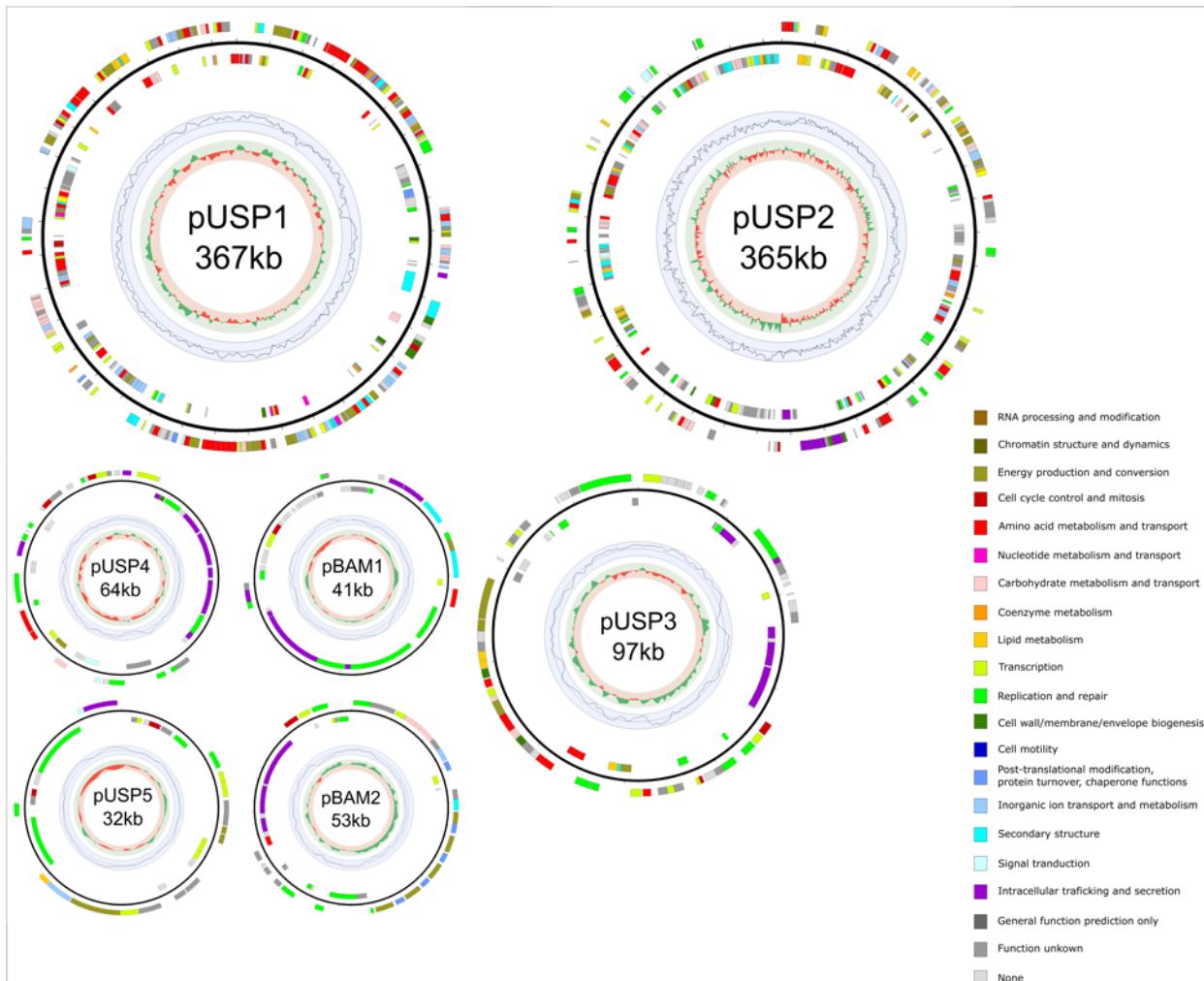
Code	Description	Total	Chr	pBAM1	pBAM2	pUSP1	pUSP2	pUSP3	pUSP4	pUSP5
J	Translation, ribosomal structure and biogenesis	2.9%	3%	0%	0%	2%	1%	0%	0%	0%
A	RNA processing and modification	0.0%	0%	0%	0%	0%	0%	0%	0%	0%
K	Transcription	7.3%	7%	5%	9%	10%	7%	9%	5%	18%
L	Replication, recombination and repair	4.9%	4%	20%	21%	2%	12%	14%	16%	18%
B	Chromatin structure and dynamics	0.1%	0%	0%	0%	0%	0%	0%	0%	0%
D	Cell cycle control, Cell division, chromosome partitioning	0.7%	1%	2%	2%	2%	1%	2%	3%	6%
V	Defense mechanisms	1.0%	1%	0%	0%	0%	1%	0%	2%	0%
T	Signal transduction mechanisms	2.5%	3%	0%	0%	1%	1%	0%	3%	3%
M	Cell wall/membrane biogenesis	3.8%	4%	0%	0%	2%	1%	2%	2%	0%
N	Cell motility	0.6%	1%	0%	0%	0%	0%	0%	0%	0%
U	Intracellular trafficking and secretion	2.5%	2%	24%	17%	0%	3%	11%	22%	3%
O	Posttranslational modification, protein turnover, chaperones	3.0%	3%	0%	17%	1%	0%	0%	0%	0%
C	Energy production and conversion	5.1%	5%	2%	2%	9%	8%	5%	2%	12%
G	Carbohydrate transport and metabolism	4.1%	4%	0%	6%	6%	6%	2%	2%	0%
E	Amino acid transport and metabolism	9.8%	9%	2%	2%	20%	12%	7%	6%	0%
F	Nucleotide transport and metabolism	1.6%	2%	0%	0%	1%	0%	0%	0%	0%
H	Coenzyme transport and metabolism	2.3%	3%	0%	0%	2%	1%	0%	0%	0%
I	Lipid transport and metabolism	2.3%	2%	0%	0%	2%	5%	3%	0%	3%
P	Inorganic ion transport and metabolism	5.8%	6%	0%	0%	10%	3%	0%	0%	3%
Q	Secondary metabolites biosynthesis, transport and catabolism	1.9%	2%	5%	4%	5%	6%	1%	0%	0%
R	General function prediction only	0.0%	0%	0%	0%	0%	0%	0%	0%	0%
S	Function unknown	20.7%	21%	10%	13%	17%	21%	17%	5%	21%
-	Not in COGs	16.9%	17%	29%	8%	9%	12%	27%	33%	15%
CDS		6237	5277	41	53	322	346	100	63	34

630 Figures



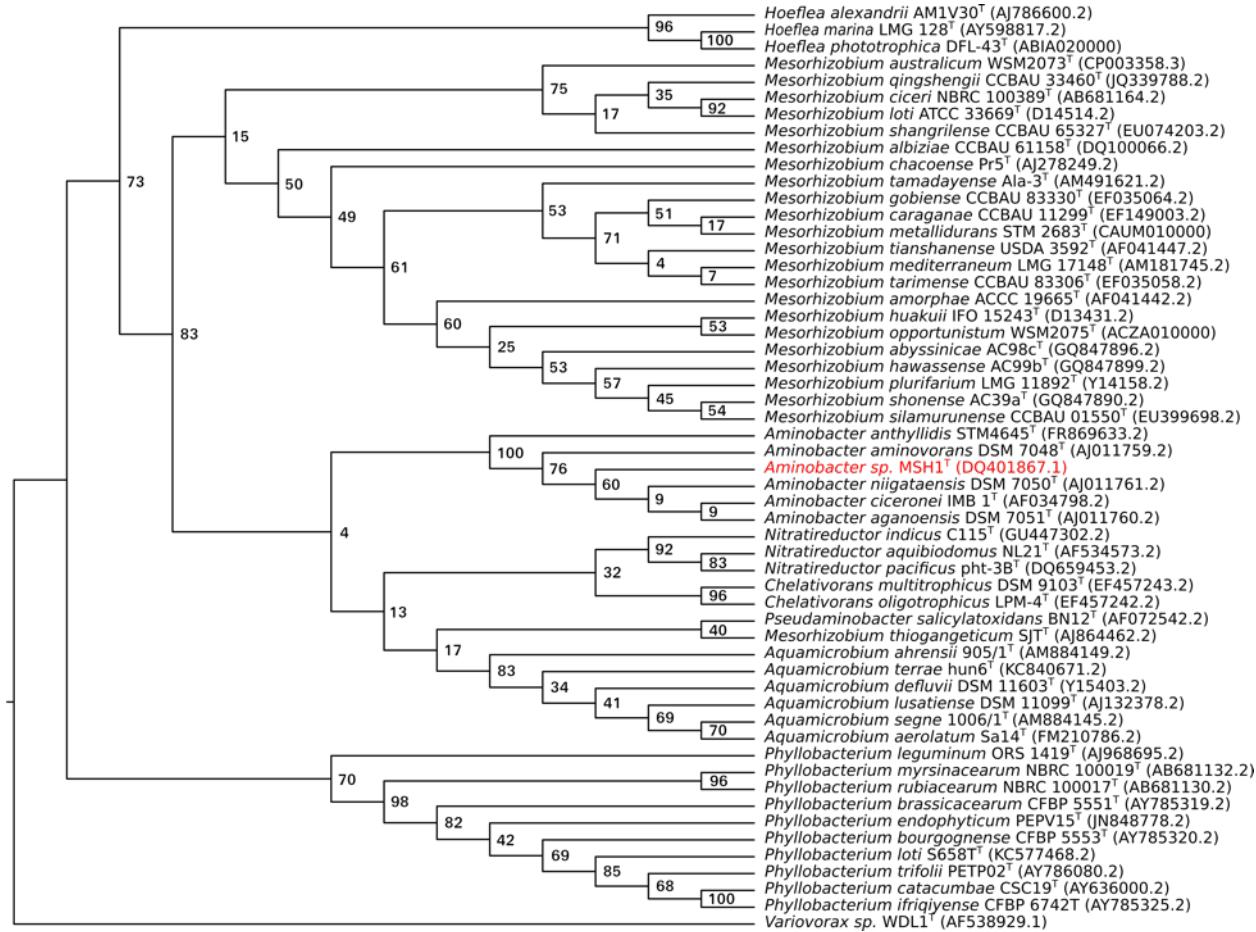
631

632 **Figure 1.** Circular view of the chromosome of *Aminobacter* sp. MSH1. From outer to inner circle:
633 CDS on leading strand, scale (ticks: 100 kb), CDS on lagging strand, tRNA (purple) and rRNA
634 (red) (only chromosome), GC plot and GC skew (>0 : green, <0 : red). CDS are colored according
635 to COG functional categories determined with EggNOG mapper 4.5.1.

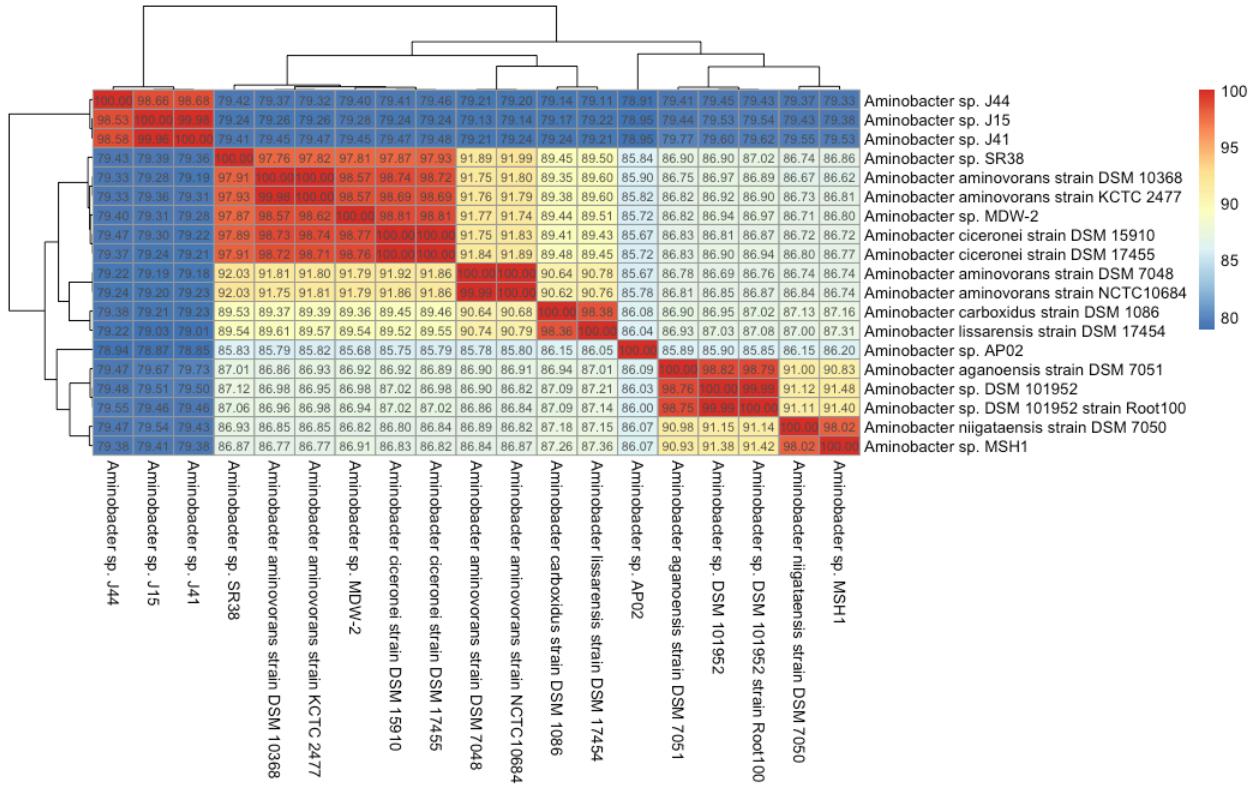


636

637 **Figure 2.** Circular view of the plasmids of the newly assigned *Aminobacter niigataensis* MSH1.
638 From outer to inner circle: CDS on leading strand, scale (ticks: 100 kb), CDS on lagging strand,
639 GC plot and GC skew (>0: green, <0: red). CDS are colored according to COG functional
640 categories determined with EggNOG mapper 4.5.1. The KU Leuven substrain MK1 lacks plasmids
641 pUSP2-5.



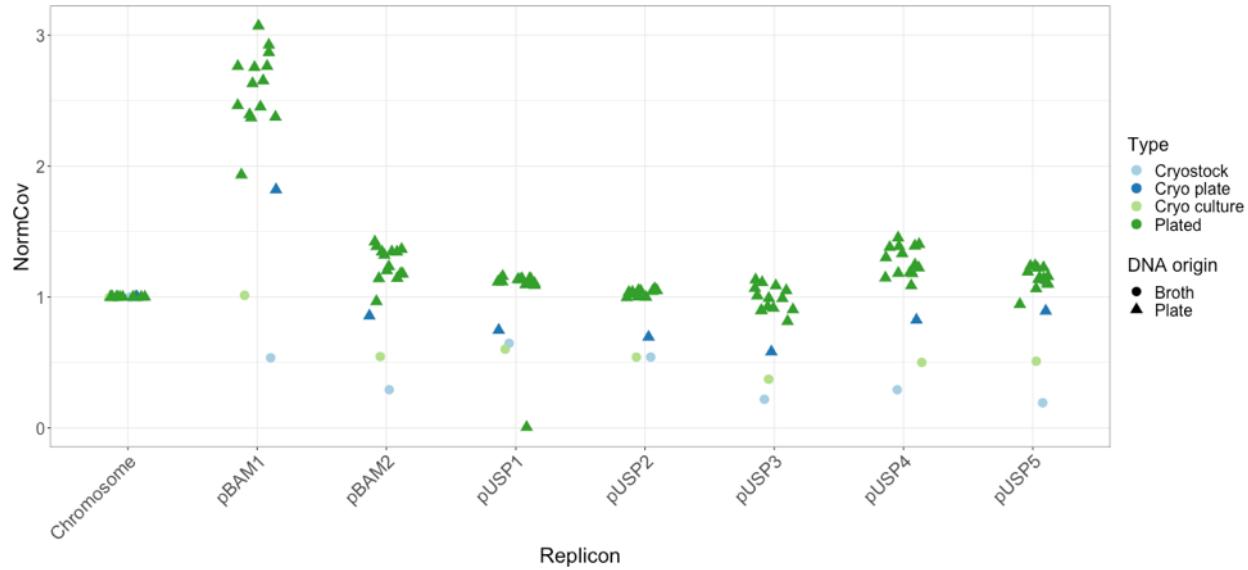
643 **Figure 3.** Phylogenetic relationships of *Aminobacter niigataensis* MSH1 based on the 16S rRNA
644 gene sequence. Maximum likelihood tree visualized as a cladogram with bootstrap values. This tree
645 was created from a clustal-omega (30) multiple sequence alignment using 16S rRNA genes from
646 the set of type strains available in the *Phyllobacteriaceae* family (NCBI accession numbers
647 between parenthesis). The tree was inferred using PhyML (31) with a GTR substitution model and
648 a calculation of branch support values (bootstrap value of 1000). The *Variovorax* sp. strain WDL1
649 was used as an outgroup (56).



650

651 **Figure 4.** Heatmap of ANI values for all available *Aminobacter* genomes from NCBI (downloaded
 652 January 31, 2021). Genomes are clustered using hierarchical clustering of ANI values, as implemented in
 653 the R package “pheatmap” (v1.0.12).

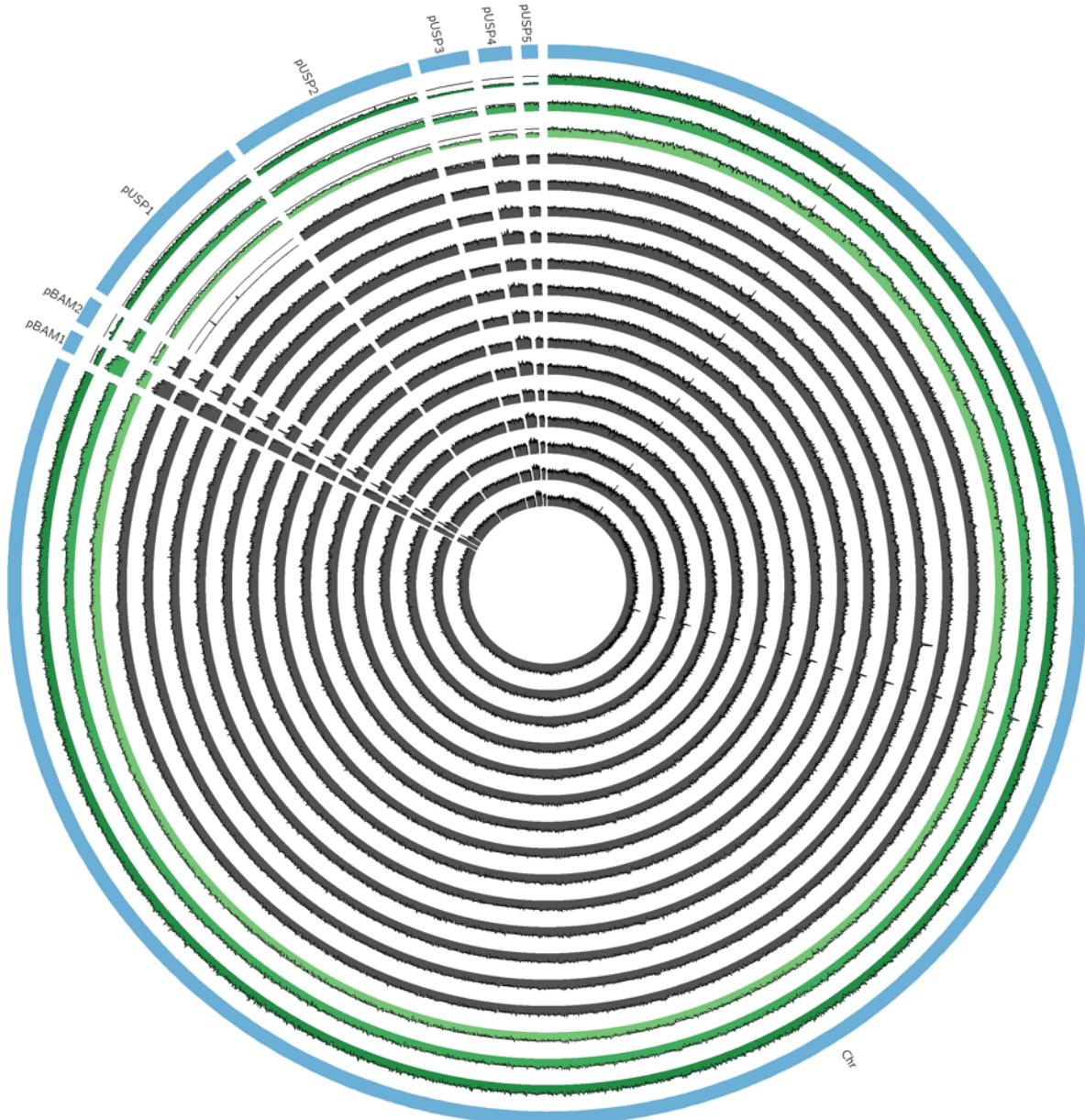
654



655

656 **Figure 5.** Coverage of replicons normalized to chromosome coverage (NormCov). A NormCov of 1
657 indicates a single copy per chromosome of a replicon. A NormCov above 1 indicates that there are more
658 copies of a given plasmid than the chromosome per cell. Points have been slightly jittered horizontally to
659 improve visualization of overlaps.

660



661

662 **Figure 6.** Illumina reads mapped to the chromosome and plasmids of MSH1 (DK1 substrain). The
663 outer blue ring indicates the replicons. The inner rings show read mapping coverage of the
664 replicons, normalized to the coverage of the chromosome per replicate, for all of the 17 sequenced
665 replicates. The three first green rings show replicates cryostock, first colony from R2A plate, first
666 colony from R2B broth, respectively. The subsequent 14 grey rings show the replicates that all
667 originate from the same first colony. A solid line in the background of all tracks indicate the

668 chromosome coverage line. Coverage above this line indicates a replicon copy number higher than
669 1 per chromosome and vice versa.

670 **Supplementary data**

671

672 **Supplementary Table S1.** Digital DNA:DNA hybridization (dDDH) results from the online Type Strain Genome Server (TYGS)
 673 analysis. d0, d4, d6 refers to different algorithms used TYGS. Formula d0 (a.k.a. GGDC formula 1): length of all HSPs divided by total
 674 genome length. Formula d4 (a.k.a. GGDC formula 2): sum of all identities found in HSPs divided by overall HSP length. Formula d6
 675 (a.k.a. GGDC formula 3): sum of all identities found in HSPs divided by total genome length. C.I.: Confidence interval.

676

Type Strain Genome Server								
Query strain	Subject strain	dDDH	C.I.	dDDH	C.I.	dDDH	C.I.	G+C content
		(d0, in %)	(d0, in %)	(d4, in %)	(d4, in %)	(d6, in %)	(d6, in %)	difference (in %)
MSH1	<i>Aminobacter niigataensis</i> DSM 7050	69.7	[65.8 - 73.3]	82.5	[79.7 - 85.0]	74.3	[70.8 - 77.5]	0.51
MSH1	<i>Aminobacter aganoensis</i> DSM 7051	53.4	[49.9 - 56.9]	40.2	[37.7 - 42.7]	50.5	[47.5 - 53.6]	1.01
MSH1	<i>Aminobacter lissarensis</i> DSM 17454	40.6	[37.2 - 44.0]	30.8	[28.4 - 33.3]	37.4	[34.5 - 40.5]	0.19
MSH1	<i>Aminobacter aminovorans</i> DSM 7048	41.3	[37.9 - 44.8]	30	[27.6 - 32.5]	37.7	[34.8 - 40.8]	0.31
MSH1	<i>Aminobacter ciceronei</i> DSM 15910	39.2	[35.8 - 42.6]	29.8	[27.4 - 32.3]	36	[33.1 - 39.1]	0.18

MSH1	<i>Aminobacter ciceronei</i> DSM 17455	39.2	[35.8 - 42.6]	29.8	[27.4 - 32.3]	36	[33.1 - 39.1]	0.18
MSH1	<i>Chelatobacter heintzii</i> DSM 10368	38.7	[35.3 - 42.2]	29.7	[27.3 - 32.2]	35.7	[32.7 - 38.7]	0.26
MSH1	<i>Mesorhizobium plurifarium</i> ORS 1032	18.9	[15.8 - 22.5]	21.9	[19.6 - 24.3]	18.5	[15.9 - 21.5]	1.19
MSH1	<i>Mesorhizobium waimense</i> ICMP 19557	19.1	[16.0 - 22.7]	21.7	[19.5 - 24.2]	18.7	[16.0 - 21.7]	0.49
MSH1	<i>Mesorhizobium australicum</i> WSM2073	19.2	[16.1 - 22.8]	21.7	[19.4 - 24.1]	18.8	[16.1 - 21.8]	0.05
MSH1	<i>Mesorhizobium qingshengii</i> CGMCC 1.12097	19.5	[16.3 - 23.1]	21.6	[19.3 - 24.0]	19	[16.3 - 22.0]	0.24
MSH1	<i>Mesorhizobium sangaii</i> DSM 100039	19.5	[16.3 - 23.1]	21.6	[19.4 - 24.1]	18.9	[16.3 - 22.0]	0.51

677