

1 **SeqCode, a nomenclatural code for prokaryotes described from sequence data**

2

3 Brian P. Hedlund, Maria Chuvochina, Philip Hugenholtz, Konstantinos T. Konstantinidis, Alison
4 E. Murray, Marike Palmer, Donovan H. Parks, Alexander J. Probst, Anna-Louise Reysenbach,
5 Luis M. Rodriguez-R, Ramon Rossello-Mora, Iain C. Sutcliffe, Stephanus N. Venter, and William
6 B. Whitman*

7 * Correspondence to: William B. Whitman, whitman@uga.edu

8

9 **Supplementary Text**

	Page
10 Glossary	1
11 Table S1. Summary of breakout group discussions from SeqCode workshops	4
12 Description of SeqCode Registry	6
13 Vision for adoption of <i>Candidatus</i> names	6
14 Example cases of validation of names under the SeqCode	7
15 Example of path 1: family <i>Wolframiraptoraceae</i> and child taxa	7
16 Table S2: Example spreadsheet format for nomenclature proposals	12
17 Example of path 2: <i>Methanoproducendum senex</i>	14
18 Example of combination of paths 1 and 2: class <i>Kryptonionia</i> and child taxa	16
19 Table S3. Etymology and typification of higher taxa for <i>Kryptonium thompsonii</i>	19

20 Glossary

21 **Basonym:** The original or earliest validly published name of a taxon on which a new combination
22 is based when this taxon has been changed, typically as a result of reclassification.

23 **comb. nov. (*combinatio nova*):** Abbreviation to denote transfer of a species to a new genus.
24 Because the species name includes both the genus name and species epithet, the new name is a
25 combination of the new genus name and the original species epithet. See Rule 29.

26 **Correct:** Validly published names or epithets that must be adopted for a taxon under the rules in
27 the proposed taxonomy (i.e., circumscription, position and rank). Note that validly published
28 names may be legitimate but not correct, depending on the taxonomy. For instance, a given
29 taxonomy may merge two species. While both names are validly published and legitimate, only
30 the name of the earlier synonym is correct in this taxonomy.

31 **Diagnosis:** A statement of the characters that allows one to distinguish the taxon from other taxa.

32 **Earlier synonym:** The first validly published synonym [in some taxonomic literature, this is
33 referred to as senior synonym].

34 **Epithet:** An adjective or descriptive phrase expressing a quality characteristic of the person or
35 thing mentioned. The species name is an epithet of the genus name.

36 **Effectively published:** Names and epithets that have appeared in print and/or electronic matter
37 that is peer reviewed and made generally available to the scientific community. Under the
38 SeqCode, preprints are not eligible as an effective publication. See Rule 24 for more explanation.

39 **Effective publication:** The peer-reviewed publication in which a name first appears. Under the
40 SeqCode, preprints are not eligible as an effective publication. See Rule 24 for more explanation.

41 **Homotypic synonyms:** Two or more names associated with the same type [in some taxonomic
42 literature, these are referred to as objective synonyms].

43 **Heterotypic synonyms:** Two or more names associated with different types that in the opinion of
44 the microbiologist concerned belong to the same taxon [in some taxonomic literature, these are
45 referred to as subjective synonyms].

46 **Homonymy:** A situation in which two names are the same but have different types, often due to
47 uniting two genera each of which contains species with the same species epithet or when a name
48 is chosen that is already in use by another taxon, including taxa named under other international
49 codes of nomenclature.

50 **Illegitimate:** Names or epithets that are contrary to the rules of the SeqCode.

51 **Later synonym:** A subsequently validly published synonym [in some taxonomic literature, this is
52 referred to as junior synonym].

53 **Legitimate (names and epithets):** In accordance with the rules of the SeqCode.

54 **Name:** The label applied to an object. A word or set of words by which a taxon is known,
55 addressed, or referred to. A term used for identification.

56 **nom. nov. (*nomen novum*):** A name that is created to replace another scientific name in cases
57 where an earlier name cannot be used for technical, nomenclatural reasons. See Rules 29 and 43.

58 **Nomenclature:** The body or system of names in a particular field.

59 **Nomenclatural type (or just type):** For a species, it is the evidence for the taxon, such as genome
60 sequence, with which the name is permanently associated. In the ICNP, the type is a viable strain
61 deposited in and available from two culture collections. For higher taxa, the type is a lower taxon.
62 The nomenclatural type is not necessarily the most typical or representative element of the taxon.

63 **Orthography:** Formation of words or names.

64 **Orthographic variant:** Names or epithets that are variant spellings of the same name. Under the
65 SeqCode, Orthographic errors may be corrected without affecting the priority of the name. See
66 Rules 46 and 47.

67 **Position:** The higher taxon in which a taxon is placed when there may be alternatives.

68 **Principle of Priority:** A principle that states that the correct name of a taxon is the oldest
69 legitimate, validly published name applied to it. It is the fundamental guiding precept that
70 preserves the stability of biological nomenclature. It was first formulated in 1842 by a committee
71 appointed by the British Association to consider the rules of zoological nomenclature. The
72 committee's report was written by Hugh Edwin Strickland.

73 **Rank:** A level within a taxonomic hierarchy. In the SeqCode, ranks are restricted to phylum, class,
74 order, family, genus, species, and subspecies. See Rule 7.

75 **Species name:** A binomial combination of the genus name and species epithet. See Rules 11 and
76 12.

77 **Taxon:** Any group of organisms treated as a named group in a formal taxonomy.

78 **Validation (of a name or epithet):** The process by which a name becomes “official” under a
79 particular code of nomenclature. Under the SeqCode, validation is completed when the registration
80 process is complete on the SeqCode Registry. See Rule 26.

81 **Validly published:** Names and epithets that are effectively published and registered according to
82 the rules of the code. See Rule 26.

83

84 **Supplementary Table S1. Summary of breakout group discussions from SeqCode workshops.** Workshops
 85 occurred during February 2021, the pros and cons as identified by the breakout groups and resulting action under the
 86 SeqCode.

Discussion topic	Breakout group recommendation and rationale	Committee action and rationale
SeqCode should be restricted to uncultured organisms.	<p><i>Breakout group recommendation:</i> Mostly no.</p> <p><i>Pros:</i> If so, this may encourage deposition of strains into culture collections, because the ICNP requires this.</p> <p><i>Cons:</i> If no, SeqCode would provide a mechanism to name fastidious organisms and those isolated from countries from which export of organisms is difficult or prohibited by law. This would unify naming of all microbes based on common currency (i.e., genomes).</p>	<p><i>Committee action:</i> SeqCode is not restricted to uncultivated organisms. See General Consideration 5 and Rule 23.</p> <p><i>Rationale:</i> SeqCode encourages use of the ICNP when possible; however, genomes are a common currency for all microorganisms. Restricting SeqCode to uncultivated organisms would limit the impact and efficacy of SeqCode.</p>
SeqCode should allow replacement of type sequences with types strains.	<p><i>Breakout group recommendation:</i> No consensus.</p> <p><i>Pros:</i> If so, this would promote the importance of type strains for functional studies. Incentivizes and gives credit to authors who isolate and describe strains rather than genomes. Allows strains to be types under both SeqCode and ICNP.</p> <p><i>Cons:</i> If no, this would enhance the stability of types but lead to conflict between types and names with ICNP.</p>	<p><i>Committee action:</i> SeqCode does not allow replacement of sequences with strains as types.</p> <p><i>Rationale:</i> Doing so would undermine the stability of names formed under the SeqCode and the resulting taxonomies. Genomes are common currency for both cultivated and uncultivated microorganisms under the SeqCode.</p>
SeqCode should specify that priority of higher taxa depends on the priority of the type genus.	<p><i>Breakout group recommendation:</i> Mostly no.</p> <p><i>Pro:</i> If so, this would be less confusing to the research community and simplify automation of name changes based on the simplicity of rules governing priority throughout the ranks.</p> <p><i>Con:</i> If so, this would depart from the current system under the ICNP where the priority of a name depends on the date of validation and the type of the class is the order. This would lead to some differences between SeqCode and ICNP names.</p>	<p><i>Committee action:</i> SeqCode requires that priority of higher taxa formed after Jan. 1, 2022, depends on the priority of genus. See Rule 23d.</p> <p><i>Rationale:</i> We feel that this decision is forward-thinking and corrects a logical lapse in the ICNP. This action will minimize confusion and simplify automated nomenclature changes through the ranks. We plan to propose a change to the ICNP so rules of priority are identical for higher taxa.</p>
SeqCode should include a proposal for an “Approved List” to incorporate the current <i>Candidatus</i> taxa.	<p><i>Breakout group recommendation:</i> Yes.</p> <p><i>Pros:</i> If so, this list would leverage a large body of literature with minimal work and disruption to the community.</p> <p><i>Cons:</i> None noted.</p>	<p><i>Committee action:</i> We agree and will work with the research community on an “Approved List” following initial implementation of the SeqCode.</p> <p><i>Rationale:</i> The list will serve the community and maximize SeqCode utility.</p>
SeqCode should include a “digital protologue” as part of the naming process.	<p><i>Breakout group recommendation:</i> Yes.</p> <p><i>Pros:</i> A machine-readable “digital protologue” to accompany all names would maximize utility of the SeqCode. The submission system should also include automated checks to support the community and minimize problems (e.g., synonymy).</p>	<p><i>Committee action:</i> Registration of names within the SeqCode Registry includes automated checks and curation. A number of recommended data fields with machine-readable outputs act as a “digital protologue”.</p> <p><i>Rationale:</i> A strong consensus emerged from the community that a “digital</p>

	<p><i>Cons:</i> A “digital protologue” would take substantial work and funding to set up and maintain.</p>	<p>protologue” and automated registration system would greatly support systematics.</p>
<p>SeqCode should require ‘perfect Latin’?</p>	<p><i>Breakout group recommendation:</i> Mostly yes. <i>Pros:</i> If so, SeqCode names could be merged with ICNP names. Naming is simplified by recent resources. <i>Cons:</i> Latin rules are somewhat difficult and promote inequity based on language and cultural groups.</p>	<p><i>Committee action:</i> SeqCode requires Latin, although minor orthographic variants (i.e., mistakes) are tolerated. See Principle 3, Rules 8, 45, and 46. Furthermore, the SeqCode Registry has been set up so that most errors will be corrected during registration. In cases where they are not, minor corrections can be made any time at the SeqCode Registry. <i>Rationale:</i> Allows future merging with ICNP names.</p>
<p>SeqCode should be managed and revised using two international committees mirroring the ICSP and Judicial Commission.</p>	<p><i>Breakout group recommendation:</i> Mostly yes. <i>Pros:</i> These committees are necessary to emend SeqCode and resolve disputes. <i>Cons:</i> These committees can become bureaucratic and autocratic.</p>	<p><i>Committee action:</i> SeqCode will be managed by two committees mirroring the ICSP committees. These committees will be codified in statutes with cooperation of the research community outside of the SeqCode. See General Consideration 6 and Rule 2. <i>Rationale:</i> These committees will be necessary to implement and improve SeqCode.</p>
<p>SeqCode should include the rank of subspecies.</p>	<p><i>Breakout group recommendation:</i> No consensus. <i>Pros:</i> Genomic matrices show multimodal relationships between species that could be considered subspecies. Naming these groups could simplify communication about these groups. <i>Cons:</i> There are no clear or universal genomic thresholds within species; therefore, the meaning is unclear, and implementation complicated.</p>	<p><i>Committee action:</i> SeqCode includes subspecies. See Rule 7a, 7c and 13. <i>Rationale:</i> The subspecies rank is optional and may be useful for some researchers.</p>

87

88

89 **Description of the SeqCode Registry**

90 The SeqCode Registry is a web application supporting three main objectives: (1) the registry and
91 evaluation of names to be proposed anew in accordance with the SeqCode, (2) the identification
92 of *Candidatus* names currently used in the literature for the normalization and standardization of
93 their use through validation under the SeqCode, allowing one to drop the *Candidatus* qualifier,
94 and (3) the maintenance of a standardized, publicly available list of names validated under the
95 SeqCode. This web application is available at <https://seqco.de/>. All its public data is accessible and
96 reusable through the Creative Commons Attribution 4.0 License except where otherwise noted,
97 and the underlying code is released as open source under the terms of the Artistic License 2.0. The
98 SeqCode Registry provides user-friendly graphical-interface access to its resources, as well as
99 computer-readable entries in JSON format for easy integration by third-party services. Examples
100 of the system's use are provided below for the registration of names under different publication
101 circumstances.

102

103 **Vision for adoption of *Candidatus* names**

104 A catalog of over 1,000 *Candidatus* names has been compiled (Oren et al., 2020), and recently 917
105 *Candidatus* names were published as part of a study of the chicken fecal microbiome (Gilroy et
106 al., 2021). The SeqCode was deliberately developed with very few requirements in the effective
107 publication to allow these and other names to be validated. In fact, any *Candidatus* name in the
108 literature can be validated under Path 2 (Figure 1) as long as the taxa are named in the effective
109 publication and a genome meets data quality standards required to serve as a nomenclatural type
110 (Table 1). This is possible because critical data, including designation of the nomenclatural type,
111 can be captured under Path 2 in the SeqCode Registry during validation. We plan to initiate this
112 effort, which will be done in collaboration with the community. However, the authors of
113 *Candidatus* taxa themselves will be welcome to validate names that are already effectively
114 published and meet the sequence standards. Because the SeqCode Registry is already operational,
115 this could begin immediately. The basic procedure to validate *Candidatus* names *en masse* would
116 be: (1) to assess genomes assigned to each *Candidatus* taxon for data quality; (2) where a genome
117 is of sufficient quality to serve as a type, contact authors to check auto filled templates generated
118 by the SeqCode Registry and fill in missing data fields; (3) complete validation in the SeqCode
119 Registry; and (4) publish a paper with collaborators from the community announcing validation of
120 the names. This project would result in validation of *Candidatus* names, centralize names and
121 metadata for these taxa, serve an important outreach function to educate the community about the
122 principles and implementation of the SeqCode, and provide a conduit for community feedback.

123 References:

124 1. Oren A, Garrity GM, Parker CT, Chuvochina M, Trujillo ME. 2020. Lists of names of
125 prokaryotic *Candidatus* taxa. *Int J System Evol Microbiol.* 70:3956-4042.
126 <https://doi.org/10.1099/ijsem.0.003789>.

127 2. Gilroy R, Ravi A, Geino M, Pursley I, Horton DL, Alikhan N-F, Baker D, Gharbi K, Hall H,
128 Watson M, Adriaenssens EM, Foster-Nyarko E, Jarju S, Secka A, Antonio M, Oren A, Chaudhuri
129 RR, La Ragione R, Hildebrand F, Pallen MJ. 2021. Extensive microbial diversity within the
130 chicken gut microbiome revealed by metagenomics and culture. PeerJ 9:e10941. doi:
131 10.7717/peerj.10941. eCollection 2021.

132

133 **Example of path 1: *Wolframiraptoraceae* and child taxa**

134 **Rationale for Path 1.** Path 1 (main text Figure 1) is the preferred route to name new taxa under
135 the SeqCode because the SeqCode Registry includes automated checks and curator input. In the
136 case of Path 1, the names are pre-registered with the SeqCode Registry prior to peer-review of the
137 effective publication describing the new taxa. By doing so, Path 1 serves two important roles for
138 the research community: (1) Automated checks and curator input during pre-registration can
139 prevent mistakes such as synonymy or problems with Latinization before names are published and
140 thus prevent confusion resulting from name changes after publication. This process is thus
141 somewhat similar to manual nomenclatural checks during peer-review at IJSEM; however, by
142 automating the process as much as possible, we aim to maximize speed and scalability and
143 minimize human error. (2) SeqCode identifier URLs generated during pre-registration can be
144 included in the effective publication and allow peer-reviewers and editors to access the pre-
145 registered names to ensure they have passed SeqCode checks. This process should improve and
146 simplify peer-review of new names because approval by the SeqCode Registry at the pre-
147 registration phase can provide confidence that the names are free of problems such as synonymy
148 and poor Latinization. It should also be noted that minor orthographic variants of names that are
149 validated under the SeqCode can be proposed by anyone at any time on the SeqCode Registry
150 without publishing errata, which is also aimed at minimizing workload and confusion for the
151 research community. Decisions on these orthographic variants will be refereed by curators.

152 **Overview.** As an example case for Path 1, several authors of the SeqCode (Palmer, Reysenbach,
153 Hedlund) recently completed a combined cultivation/metagenomics study of a novel group of
154 archaea in the GTDB family designated as NZ13-MGT within the GTDB order “Caldarchaeales”,
155 class *Nitrososphaeria*, and phylum *Thermoproteota*. The study initially focused on anaerobic
156 enrichment cultures from sediments of Great Boiling Spring, Nevada, USA, with a single member
157 of the taxon that requires tungsten for growth on corn stover or a sugar mix. The taxon was shown
158 to prefer xylose by combining fluorescence *in situ* hybridization (FISH) with nanometer-scale
159 secondary ion mass spectrometry (nanoSIMS). It is currently represented by a single high-quality
160 MAG, although that MAG formed a >99.5% average nucleotide identity (ANI) cluster with MAGs
161 of lower quality from separate samples of the same enrichment culture and sediments from which
162 the enrichment culture was derived. Analysis of the MAG revealed a putative TupA tungstate
163 transporter and six annotated tungsten-dependent ferredoxin oxidoreductases. To expand the
164 study, 77 additional high-quality MAGs assigned to the GTDB family NZ13-MGT by GTDB-tk
165 (Chaumeil et al., 2020) were assembled from metagenomes from terrestrial and marine
166 hydrothermal systems globally.

167 **Importance of multiple genomes per species.** It is recommended (main text Table 1) that species
168 or subspecies named under the SeqCode include more than one genome. This parallels the general
169 recommendation under the ICNP to characterize multiple strains for proposals of new taxonomic
170 names and is especially important for MAGs and SAGs because of challenges associated with
171 accurately binning metagenomic data and the low completeness of most SAGs. Here, FastANI was
172 used to dereplicate the 78 high-quality MAGs into 11 >95% ANI clusters (i.e., species clusters),
173 and phylogenetic analyses of concatenated marker gene sets confirmed that each ANI cluster was
174 monophyletic. In total, nine of the species clusters were represented by multiple high-quality
175 MAGs derived from metagenomes from different sampling dates and/or geothermal springs,
176 ranging from two to 16 MAGs. Comparison of the multiple MAGs per species cluster allowed us
177 to assess: (1) monophyly of the species by using a multiple marker gene set; (2) the true presence
178 of multiple copies of normally single-copy, conserved marker genes and true absence of conserved
179 marker genes used to assess genome completeness and contamination; (3) the existence of
180 homologs of genes encoding important functions (in this case, tungstate transporters,
181 tungstoenzymes, and genes related to energy conservation) to support the association of those
182 genes/pathways with the species; (4) shared gene content in general; and (5) similar genome sizes
183 for the genomes within a species. These comparisons strengthened our conclusions about the
184 proposed mode of energy conservation and evolution of the organisms and allowed us to identify
185 and reject problematic MAGs. Briefly, various bioinformatic pipelines were used for generation
186 of MAGs, but metagenomic assembly was typically performed with metaSPAdes (Nurk et al.,
187 2017), followed by binning of contigs > 2.5 kbp in length with the MetaBAT (Kang et al., 2015)
188 or MetaBAT2 (Kang et al., 2019), and/or MaxBin2 (Wu et al., 2016), and/or CONCOCT v.1.1.0
189 (Alneberg et al., 2014) algorithms. In some cases, the “bin_refinement” module of MetaWRAP
190 was also used to identify the highest quality MAGs among the three binning strategies, and quality
191 of all genomes were assessed based on CheckM v.1.1.39 estimates of completeness and
192 contamination. Only those MAGs exhibiting > 90 % estimated completeness and < 5 % estimated
193 contamination are reported here. Here, we rejected all MAGs generated by MaxBin2 (Wu et al.,
194 2016) because all were massively contaminated, as evidenced by reproducibly larger and more
195 variable genome sizes than MAGs generated with other pipelines, but without any additional
196 conserved marker genes. These MAGs were therefore not flagged as contaminated by automated
197 contamination detection software such as CheckM (Parks et al., 2018), so comparison of the MAGs
198 within each species was critical.

199 In this study, we named two species based on single high-quality MAGs. In one case,
200 *Terraquivivens tikiterensis*, we justified this decision based on the very high quality of the MAG
201 (99.0 % completeness with 0.49 % estimated contamination; 75x coverage including sequences
202 for 5S, 16S and 23S rRNAs and tRNAs coding for 19 amino acids) and the availability of 24
203 MAGs from other species of the genus for comparison. In the other case, *Benthortus lauensis*, we
204 justified this decision again on the high quality of the MAG (94.0 % completeness with 2.91 %
205 estimated contamination; 34x coverage including sequences or 5S, 16S and 23S rRNAs and tRNAs
206 coding for 20 amino acids) and on the difficulty of obtaining additional samples from the only
207 known habitat of this genus, the Lau Hydrothermal Field.

208 **Data quality under the SeqCode.** Only MAGs exhibiting estimated completeness > 90 % and
209 estimated contamination < 5 % were considered possible candidates for naming under the
210 SeqCode. We furthermore accounted for rRNAs and tRNAs and selected the best MAG
211 representing each species cluster as type genome, ensuring that we adhered to data quality
212 standards (Table 1). We note that several GTDB species representatives did not contain any high-
213 quality MAGs; those MAGs were analyzed phylogenetically in our study, but they should not be
214 named under the SeqCode (Table 1). Nevertheless, they can be identified using GTDB placeholder
215 names.

216 **Nomenclature under the SeqCode.** We decided to propose names for all 11 species clusters under
217 the SeqCode as well as the parent taxa as prescribed under the SeqCode. Names were formed under
218 the rules of Latin following general recommendations of Appendix 9 in the ICNP and were
219 checked by the nomenclature expert Aharon Oren. To follow Rules 18a and 26 in the SeqCode,
220 the now-submitted effective publication includes the following for each taxonomic name that is
221 proposed: (i) clear designation of their nomenclatural type; (ii) designation of the taxonomic rank;
222 and (iii) the etymology of the new name. We also followed Recommendation 26 in the SeqCode
223 by clearly separating the nomenclature proposals in the manuscript within a protologue. An
224 annotated example protologue (with annotations [*in brackets in italics*]) for each taxonomic rank
225 is shown below, preceded by a brief discussion of the elements of the protologue. To be clear,
226 standardized protologues are not required under the SeqCode as long as the rules of the SeqCode
227 are applied, particularly Rule 18a and Rule 26 for new taxon names that don't require unions or
228 divisions of existing taxa. Authors of publications proposing many new taxa may choose to
229 summarize the required elements as a table, for example. A summary of the key elements required
230 by the SeqCode are listed in Supplementary Table S2 for these same three taxonomic names as an
231 example.

232 **Example protologue for a new family.** The protologue below describes the new family
233 *Wolframiiraptoraceae*. Note that for a family or higher taxon, the nomenclatural type is a genus.
234 Here, that genus is *Wolframiiraptor* from which the name of the family must be derived under
235 Rule 15 of the SeqCode.

236 **Description of the family *Wolframiiraptoraceae***

237 *Wolframiiraptoraceae* (Wolf.ra.mi.i.rap.to.ra'ce.ae. N.L. masc. n. *Wolframiiraptor*, type genus of
238 the family; L. fem. pl. suff. *-aceae*, ending to denote a family; N.L. fem. pl. n.
239 *Wolframiiraptoraceae*, family of the genus *Wolframiiraptor*). [*This text designates the taxonomic*
240 *rank (family) and the etymology under Rule 26.4 and 26.5. Both will also be captured in defined*
241 *fields in the SeqCode Registry.*]

242 Members of this family are associated with thermal aquatic environments and have been identified
243 from geothermal springs in China, New Zealand and the USA and a marine hydrothermal vent in
244 the Western Pacific. Phylogenomic inference robustly recovers the genomes of these organisms as
245 a well-supported monophyletic lineage within the order *Caldarchaeales*, and delineation of these
246 taxa as a family is supported by Relative Evolutionary Divergence (RED) and Average Amino
247 Acid Identity (AAI). AAI values among designated type genomes for species in this family range

248 between 65 and 85 % within proposed genera, and between 49 and 57 % among members of
249 different genera. The distribution of genes required for oxidative phosphorylation indicate that
250 members of the family may either be strict or facultative anaerobes. Sulfide-dependent respiration
251 may also occur in some members of the family, but this trait is not conserved for all genera. Several
252 putative tungsten-dependent ferredoxin oxidoreductases, specifically aldehyde ferredoxin
253 oxidoreductases (AORs), formaldehyde ferredoxin oxidoreductases (FORs), and glyceraldehyde-
254 3-phosphate ferredoxin oxidoreductases (GAPORs) are encoded by genomes belonging to this
255 family. [This text includes a description of the taxon, following Recommendation 26. Such text is
256 recommended but not required under the SeqCode.]

257 The nomenclatural type for the family is the genus *Wolframiiiraptor*. [This text designates the
258 nomenclatural type under Rule 26.3. Note that the nomenclatural type for ranks of class, order,
259 and family is the first legitimately named genus in the taxon. These dates are clearly shown in the
260 SeqCode Registry. Rule 26.3 embodies Principle 5 and serves to unambiguously identify the taxon.
261 See Rule 16, especially Table 2, Rule 18, and Rule 22. Under the SeqCode, the name of newly
262 proposed higher taxon must be derived from the root of the type genus *Wolframiiiraptor*
263 (*Wolframi(-um)*) plus a connecting vowel (*i*), *raptor* (genitive case *-oris*, dropping the ending of
264 genitive case gives us the root *raptor-*) followed by the appropriate suffix to denote the rank (*-*
265 *aceae*). See Rule 15 and Table 1.]

266 **Example protologue for a new genus.** The protologue below describes the new genus
267 *Wolframiiiraptor*. Note that for a genus, the nomenclatural type is a species.

268 **Description of the genus *Wolframiiiraptor***

269 *Wolframiiiraptor* (Wolf.ra.mi.i.rap'tor N.L. neut. n. *wolframium*, tungsten; L. masc. n. *raptor*,
270 snatcher or thief; N.L. masc. n. *Wolframiiiraptor*, snatcher of tungsten). [This text designates the
271 taxonomic rank (genus) and the etymology under Rule 26.4 and 26.5.]

272 Members of this genus have been identified from geothermal springs from the Great Basin,
273 Yellowstone National Park, USA, and the Rehai and Ruidian geothermal fields, Tengchong,
274 China. AAI values among genomes representing separate species within the genus range between
275 81 and 90 %. Based on ancestral state reconstruction analysis, likely losses of the genes encoding
276 cytochrome C oxidase subunits, the aerobic carbon-monoxide dehydrogenase large subunit, and
277 sulfide:quinone oxidoreductase (Sqr), indicate that members of this genus are likely strict
278 anaerobes and are also incapable of sulfide oxidation. Genomes of this genus encode a *tupA* subunit
279 of the tungstate (Tup) ABC transporter and contain multiple genes encoding tungsten-dependent
280 oxidoreductases, including three putative AOR-like, one FOR-like and one GAPOR-like proteins.
281 This taxon is supported as a genus-level group by phylogenomics, AAI and RED. [This text
282 includes a description of the taxon, following Recommendation 26. Such text is recommended but
283 not required under the SeqCode.]

284 The nomenclatural type of the genus is *Wolframiiiraptor gerlachensis*^{Ts}. [This text designates the
285 nomenclatural type under Rule 26.3. Note that the nomenclatural type for rank of genus is a

286 *species, typically the first legitimate species in the genus. These dates are clearly shown in the*
287 *SeqCode Registry. Rule 26.3 embodies Principle 5 and serves to unambiguously identify the taxon.*
288 *See Rule 16, especially Table 2, and Rule 22. Note that genus names do not need to have a standard*
289 *suffix like family, order and above, but they should avoid suffixes used for other taxonomic ranks*
290 *to prevent confusion. See Rule 15 and Table 1. Under Chapter 4, the superscript Ts can be added*
291 *when this species is a nomenclatural type under the SeqCode.]*

292

293 **Example protologue for a new species.** The protologue below describes the new species
294 *Wolframiiiraptor gerlachensis*. Note that for a species, the nomenclatural type is a DNA sequence,
295 typically a genome assembly (Table 1).

296 **Description of *Wolframiiiraptor gerlachensis*^{Ts}**

297 *Wolframiiiraptor gerlachensis* (ger.lach.en'sis N.L. masc. adj. *gerlachensis*, of Gerlach, the region
298 where Great Boiling Spring is located in Nevada, and this organism was obtained from). [*This text*
299 *designates the taxonomic rank (species) and the etymology under Rule 26.4 and 26.5. Under*
300 *Chapter 4, the superscript Ts can be added to denote that this species is the type for the genus.]*

301 A MAG representing this species was recovered from metagenomic sequencing of a stable
302 enrichment culture, established from an *in-situ* corn stover enrichment from Great Boiling Spring,
303 Nevada, USA. Enrichment and maintenance of this species within the mixed-culture community
304 was optimal at an incubation temperature of 80 °C with lignocellulose and sugars, at circumneutral
305 pH. This species is dependent on tungsten for growth, with significant decline in its abundance
306 within the community without tungsten added to the growth medium. Additionally, several
307 tungstoenzymes conserved within the genus were expressed at a higher rate during growth on corn
308 stover, suggesting direct involvement of tungstoenzymes in complex carbohydrate metabolism.
309 Cells of this organism showed significant isotope enrichment when grown on isotopically labeled
310 xylose-amended medium, with limited isotope enrichment during growth on medium amended
311 with amino acids, glucose, ribose, and starch, indicating preferential assimilation of xylose. The
312 genome sequence for this organism is 1,277,965 bp in size, consists of 27 contigs, and has a G+C
313 content of 52 %. Completeness is estimated at 98.06 % with 0.49 % contamination, as estimated
314 with CheckM. ANI comparisons among this genome and those of closely related species were
315 below 86 %, supporting the delineation of this taxon as unique and distinct to other species in the
316 genus. [*This text includes a description of the taxon, following Recommendation 26. Such text is*
317 *recommended but not required under the SeqCode.]*

318 The genome Wger_A8^{Ts}, available under the GenBank assembly accession number
319 (GCA_021323375.2^{Ts}), is the designated nomenclatural type for the species, and was recovered
320 from an enrichment culture, established from an *in-situ* enrichment from Great Boiling Spring,
321 Nevada, USA. [*This text designates the nomenclatural type under Rule 26.3. Note that the*
322 *nomenclatural type for rank of species or subspecies is a DNA sequence, typically a genomic*
323 *assembly (Table 1). Rule 26.3 embodies Principle 5 and serves to unambiguously identify the*

324 *taxon. Metadata for this sequence is included in the GenBank entry. Under Chapter 4, the*
 325 *superscript Ts can be added to denote that this genomic assembly is the nomenclatural type of the*
 326 *species.]*

327
 328 **Supplementary Table S2. Example spreadsheet format for nomenclature proposals.** Example
 329 proposals for family, genus, and species ranks. Such tables are simple alternatives to protologues
 330 and may be especially valuable when many names are proposed.

Proposed taxon	Etymology	Nomenclatural type
Family <i>Wolframiiraptoraceae</i>	Wolf.ra.mi.i.rap.to.ra'ce.ae N.L. masc. n. <i>Wolframiiraptor</i> , type genus of the family; L. fem. pl. suff. <i>-aceae</i> , ending to denote a family; N.L. fem. pl. n. <i>Wolframiiraptoraceae</i> , family of the genus <i>Wolframiiraptor</i>	Genus <i>Wolframiiraptor</i>
Genus <i>Wolframiiraptor</i>	Wolf.ra.mi.i.rap'tor N.L. neut. n. <i>wolframium</i> , tungsten; L. masc. n. <i>raptor</i> , snatcher or thief; N.L. masc. n. <i>Wolframiiraptor</i> , snatcher of tungsten	Species <i>Wolframiiraptor gerlachensis</i> ^{Ts}
Species <i>Wolframiiraptor gerlachensis</i> ^{Ts}	ger.lach.en'sis N.L. masc. adj. <i>gerlachensis</i> , of Gerlach, the region where Great Boiling Spring is located in Nevada, and this organism was obtained from	Genomic assembly: GCA_021323375.2 ^{Ts}

331
 332 **Pre-registration of names in the SeqCode Registry.** For pre-registration of names prior to
 333 publication (Path 2), create a profile in the SeqCode Registry (<https://seqco.de/>) and request
 334 contributor privileges. Once approved, a “Contributor” dashboard will become available. For
 335 registering new names, select the “Create a name” option. Provide the proposed name to be
 336 registered in the textbox and select “Register” (e.g., *Wolframiiraptor gerlachensis*). This will only
 337 reserve the name and does not entail validation. During this registration step, automated quality
 338 checks will be performed, and red and orange flags will be generated with descriptions of what is
 339 required/recommended and how each relates to Rules and Recommendations of the SeqCode. Red
 340 flags need to be addressed before progressing further. For “Missing rank” red flags, select “Define
 341 rank” and select the appropriate rank for the taxon for which the name is registered (e.g., species),
 342 and select “Submit”. For addressing the “Missing type” red flags, select “Edit type”, and under
 343 material, select the appropriate description of the genome and provide the associated accession
 344 number (e.g., NCBI assembly; GCA_021323375.2). To add the etymology, select “Edit
 345 etymology” under the “Missing etymology” red flag. Provide the syllabication and language

346 information of components of the name to be registered, or alternatively, provide a formatted
347 etymology in the text box (e.g., ger.lach.en'sis N.L. masc. adj. *gerlachensis*, of Gerlach, the region
348 where Great Boiling Spring is located in Nevada, and this organism was obtained from). To
349 address orange flags, the same process can be followed, although these flags are not obligatory for
350 registration. Orange flags can include, but are not limited to, “Missing description”, “Missing
351 parent”, and “Missing effective publication” flags. If a description for the taxon is available, add
352 this to the “Missing description” flag (e.g., protologue included for publication). If possible, link
353 a parent taxon for the name, even if the parent taxon is also a newly proposed name to be registered
354 (e.g., *Wolframiraptor*). As the pre-registration of the names is occurring before effective
355 publication, the “Missing effective publication” flag remains on the entry until effective
356 publication. If additional names or parent taxon names are proposed, complete this process for all
357 proposed taxa. This can be accessed and/or evaluated under the “My names” list in the
358 “Contributor” dashboard. Once all required information for review is provided for each entry, the
359 “Propose name” option under “Contributor Actions” can be selected for each entry. This allows
360 the addition of the name to a new register list, or, as part of larger projects with a single intended
361 effective publication, all proposed taxa can be added to the same register list for curator review.
362 Select “Register name” to generate a draft register list. This draft register list is assigned an
363 accession URL that should be included in the effective publication (e.g., seqco.de/r:slp2ijs4).
364 Submit the draft register list for curator review. Once a curator provisionally accepts the names,
365 this completes pre-registration. At this point, the proposed names are hidden from the public but
366 can be accessed via the URL, allowing reviewers and editors to review the proposed names.

367 Finally, authors are required to complete the registration/validation process by entering the DOI
368 and submitting the published article, at which point the names become valid and are visible to the
369 public. If the register list accession is included in the effective publication, then the SeqCode
370 Registry will mine that accession from the literature and automatically complete the registration
371 process, which validates the names under the SeqCode. The names are then visible to the public.
372 If the effective publication does not contain the register list accession/s, authors are required to
373 manually complete the registration/validation process by entering the DOI, at which point the
374 names become valid and are visible to the public, this process is completed by expert review of
375 the reported name/s. Minor edits/corrections can be made at any time following validation by
376 registered contributors as long as they are validated by curators.

377

378 **Example of path 2: *Methanoproducendum senex*^{Ts}**

379 Path 2 (main text Figure 1) should be used in cases where scientists want to register and validate
380 names under the SeqCode after the effective publication is already published. This could be done
381 by the authors or by other scientists, although we recommend working with authors to register
382 their names whenever feasible. The names of the taxa must appear in the effective publication
383 including online supplementary material, although minor orthographic variants are allowable.

384 **Overview.** As an example case for Path 2, one of the authors of the SeqCode (Hedlund) was a co-
385 author on a publication of MAGs from geothermal springs in Tengchong, China, containing
386 methyl-coenzyme M reductase (*mcr*) genes, which were interpreted in the context of potential
387 energy conservation from methanogenesis, methanotrophy, or alkanotrophy. In the paper, six
388 novel genera are proposed. A type species is designated for each, and a MAG is designated as the
389 type for each species. In some cases, novel families and orders were also proposed. For the sake
390 of simplicity, we discuss the process of naming the monospecific genus *Methanoproducendum*
391 with the single species *Methanoproducendum senex*.

392 **Best practices for cooperativity for Path 2.** We consider it best practice to communicate with
393 key authors before registering and validating effectively published names under the SeqCode. In
394 this case, one of the authors of the original publication (Hedlund) and Palmer collaborated with
395 the first author (Hua) and corresponding author (Li) to register these names within the SeqCode
396 Registry.

397 **Taxonomic decisions and example of taxonomic freedom.** Per Principle 1, the SeqCode does
398 not endorse any particular taxonomy. This is true of all major codes of nomenclature except the
399 *International Code of Virus Classification and Nomenclature*. The taxonomic decisions described
400 in this section were made by Hedlund, Palmer, Hua, and Li. The original manuscript proposes
401 names for a novel genus and species, *Methanoproducendum senex*, and a novel family,
402 *Methanoproducendaceae*. However, elsewhere in the paper *Methanoproducendum senex* is
403 described as a member of the *Archaeoglobaceae*, a taxon with a validly published name under the
404 ICNP. Upon reassessment, we choose to assign *Methanoproducendum senex* to the
405 *Archaeoglobaceae* based on: (i) monophyly with the *Archaeoglobaceae* with strong support in
406 phylogenetic analyses using conserved marker genes and (ii) RED values consistent with a single
407 family as implemented within GTDB. We also note that only a single high-quality MAG (92.8 %
408 completeness with 0.98 % estimated contamination and sequences of 16S and 23S rRNAs and 42
409 total tRNAs) is described for this taxon. Thus, we feel that a new family is not warranted based on
410 this single MAG as a matter of conservatism.

411 **Deposition of type DNA sequences for species and subspecies into an INSDC database.** The
412 MAG designated as the type for the proposed species (GMQP bin_32 (Ga0263258)) is currently
413 only available in the IMG database. According to Rule 18a, type DNA sequences must be available
414 from INSDC databases, a decision made to promote stable access to the data. Thus, this MAG has
415 been submitted to NCBI (**awaiting accession**).

416 **Minor changes to the etymology for *Methanoproducendum senex*.** According to Rule 45 and
417 Rule 46, names under the SeqCode should comply with rules of Latin; however, minor
418 orthographic errors are tolerated. The etymology should describe the orthography. Here, we will
419 correct a minor error in the etymology. Such errors will be automatically flagged during
420 registration in the SeqCode Registry, which is designed to minimize such minor errors.

421 **Original protologue for *Methanoproducendum senex* (Hua et al., 2019)**

422 “*Methanoproducendum*” (Me.tha.no.pro.du.cen’dum) M. L. n. methanum methane; M. L. adj.
423 producer; L. part. *Methanoproducendum*, the methane producer. The type species is “*Ca.*
424 *Methanoproducendum senex*”. “*Ca. Methanoproducendum senex*” (se’nex). L. n. old man,
425 referring to the deep-branching position of the organism. The type material is the metagenomic
426 bin GMQP bin_32 (Ga0263258).

427
428 **Modified etymology and typification of *Methanoproducendum senex*^{Ts}.** Hua et al., 2019
429 remains the effective publication.

430 Description of the genus *Methanoproducendum*

431 *Methanoproducendum* (Me.tha.no.pro.du.cen’dum N.L. neut. n. *methanum*, methane; N.L. pres.
432 part. *producendum*, produce; N.L. neut. n. *Methanoproducendum*, the methane producer).

433 The nomenclature type of the genus is *Methanoproducendum senex*^{Ts}.

434 Description of the species *Methanoproducendum senex*

435 *Methanoproducendum senex* (se’nex L. masc. n. *senex*, old man, referring to the deep-branching
436 position of the organism).

437 The nomenclatural type is GMQP bin_32^{Ts} (awaiting accession^{Ts}).

438
439 **Registration of effectively published names in the SeqCode Registry.** Registration of names in
440 the SeqCode Registry can occur after effective publication. Register as “Contributor”. Select the
441 “Create a name” option from the “Contributor” dashboard and provide the effectively published
442 name in the textbox and select register (e.g., *Methanoproducendum senex*). Similar as described
443 before, automatic checks are performed, and red and orange flags related to Rules and
444 Recommendations in the SeqCode are generated. In addition to addressing all flags described
445 before, the orange “Missing effective publication” flag needs to be addressed. Select the “Register
446 publication” option under this flag, provide the DOI to the effective publication (10.1038/s41467-
447 019-12574-y) in the provided space, and select “Register”. From here, the linked publication
448 shows which names are linked to the specific publication. Select the taxon name being edited (e.g.,
449 *Methanoproducendum senex*). This publication will appear in the “Publications” section of the

450 entry (end of page). Several publications proposing, emending, correcting, or citing the taxon in
451 question can be linked to the name in this fashion. Under the “Actions” associated with the
452 publication, select “Proposed this name” for publications proposing the name, “Corrigendum” for
453 publication correcting names or “Emended this name” for published emendations. Repeat this
454 process for any higher-level taxonomic names and link the associated effective publication with
455 these names. Notify the SeqCode curators by submitting these names (typically with all associated
456 names in a single Register List) as a draft register list for curator review (e.g.,
457 seqco.de/r:kb0uupm9), and click “Validate Published Names” to submit the manuscript PDF and
458 (if relevant) supplementary material. If proposed names satisfy all criteria for validation under the
459 SeqCode, proposals will be reviewed, and contributors will be notified of validation under the
460 SeqCode or further recommendations on corrections to the submission.

461

UNDER REVIEW

462 **Example of combination of path 1 and path 2 for validation: class *Kryptonion*** 463 **and child taxa**

464 In some cases, some names are published in an effective publication, but additional names are also
465 warranted, such as a case where one or more new higher ranks for a proposed taxon is not
466 published. This would require a combination of path 1 (registration of new names) and path 2
467 (registration of effectively published names).

468 **Overview.** As an example case for a combination of Path 1 and Path 2, Eloë-Fadrosh et al., (2016)
469 proposed the new genus and species *Candidatus Kryptonium thompsonii* based on 13 SAGs and a
470 single MAG from geothermal springs in British Columbia Canada, Yunnan Province China, and
471 Nevada USA. In addition, *Candidatus Kryptonium thompsonii* was visualized by FISH in cells
472 collected from Dewar Creek, British Columbia. *Candidatus Kryptonium thompsonii* was proposed
473 as one of four *Candidatus* genera, each with a single species, belonging to a new candidate phylum
474 named Kryptonion. None of the ranks between genus and phylum were named. In Supplementary
475 Table 4 of that paper, the proposed candidate taxa were named, and the etymology described.
476 Ranks were described in the manuscript. However, no nomenclatural type was designated because
477 there was no path to validate the names under the ICNP or any other authority at the time the paper
478 was published. Many problems like this exist in the past literature because the community hadn't
479 come together to develop standards or best practices, and this is part of the impetus for the
480 SeqCode. The SeqCode is deliberately designed to allow critical information to be captured in the
481 SeqCode Registry, allowing most previous *Candidatus* names to be named under the SeqCode,
482 provided the names appear in an effective publication. However, substantial changes to published
483 names will need to be described separately in a peer-reviewed paper, which then becomes the
484 effective publication. This test case is a good example, as described below.

485 **Taxonomic decisions and example of taxonomic freedom.** Per Principle 1, the SeqCode does
486 not endorse any particular taxonomy. This is true of all major codes of nomenclature except the
487 *International Code of Virus Classification and Nomenclature*. The taxonomic decisions described
488 in this section were made by one of the authors of the original publication describing these
489 genomes (Hedlund) and Palmer in collaboration with the first author (Eloë-Fadrosh) and
490 corresponding author (Ivanova). Again, we consider it best practice to communicate with key
491 authors before registering and validating names under the SeqCode. It is our position that the four
492 genera proposed in the effective publication should be combined into one genus based on (i)
493 monophyly with strong support in phylogenetic analyses using conserved marker genes; (ii) RED
494 values consistent with a single genus as implemented within GTDB; and (iii) average amino acid
495 identity (AAI) values consistent with a single genus (>65% (Konstaninidis et al., 2017)). In this
496 case, we will retain the genus and species name *Kryptonium thompsonii* for the largest group of
497 genomes described from Dewar Creek. This is the largest species cluster, and it was also visualized
498 by FISH. This MAG is of high quality per Genome Standards Consortium recommendations
499 (Bowers et al., 2018) (18 contigs, 95.61 % completeness with 1.91 % estimated contamination;
500 68x coverage and sequences of the 5S, 16S and 23S rRNAs and tRNAs coding for 20 amino acids).

501 Below is the original etymology from Supplemental Table 4 from the effective publication, but
502 with the *Candidatus* prefix removed. This etymology with or without the *Candidatus* prefix and
503 information contained in the paper is sufficient to satisfy Rules 26.1, 26.4, and 26.5 to name both
504 the genus and species. Rules 26.2 and 26.3 will be satisfied when the name is registered with the
505 SeqCode Registry. In this case, the nomenclatural type for the species will be the highest quality
506 genome in the species cluster, which is the MAG JGI-4. The superscript ^{Ts} is added after the species
507 epithet to denote that this is the type species of the genus, following Chapter 4. We also note that
508 the species name has been corrected to *thompsonii*, as the second “i” is the correct suffix to denote
509 a genitive masculine noun, in this case for a species named after a male (David Thompson).
510 “Perfect” Latin is recommended under the SeqCode but not required. We consider this to be a
511 minor orthographic variant on the original name. Also, the etymology is modified from the original
512 to remove the definition of ‘bios’ because the root ‘bios’ does not appear in the name. We note
513 here that the GTDB species representative (JGI-8) is actually a medium-quality MAG. We
514 speculate that the GTDB used this species representative, a SAG, as it was erroneously labeled an
515 “isolate” in NCBI, which was apparently carried over to GTDB. Thus, we recommend caution and
516 care at all levels when designating nomenclatural types to ensure that the highest quality genomic
517 data are used.

518

519 **Modified etymology and typification of *Kryptonium thompsonii*^{Ts}.** Eloë-Fadrosh et al., 2016
520 remains the effective publication.

521 Description of the genus *Kryptonium*

522 *Kryptonium* (Kryp.to’ni.um. Gr. adj. *krypton*, hidden; N.L. neut. n. *Kryptonium* a hidden life form).

523 The nomenclature type of the genus is *Kryptonium thompsonii*^{Ts}.

524 Description of the species *Kryptonium thompsonii*

525 *Kryptonium thompsonii* (thomp.son’i.i. N.L. gen. masc. n. after David Thompson, explorer of the
526 region around Dewar Creek).

527 The nomenclatural type is GCA_001442925.1^{Ts}.

528 Following Recommendation 7, the ranks above the genus *Kryptonium* should be named until an
529 existing name of a parent taxon exists. All available phylogenetic analyses of this genus indicate
530 that it is very distant from the closest relatives named under the ICNP. However, there is
531 disagreement between the original publication describing the SAGs and MAGs and the GTDB
532 taxonomy as to whether the monophyletic lineage containing *Kryptonium* should be considered a
533 phylum or a class. As a matter of conservatism, we choose to only propose names up to the rank
534 of class as suggested by the GTDB taxonomy and ascribe that class to the phylum *Bacteroidota*.
535 The GTDB has suggested the following names at ranks of family, order, and class: *Kryptoniaceae*,
536 *Kryptoniales*, and *Kryptonina*. Those names follow the rules of the SeqCode, particularly Rule 15,

537 which states that the “The name of a family, order, class, or phylum is formed by the addition of
 538 the appropriate suffix to the stem of the type genus name”. Thus, we will use the GTDB names
 539 but require an etymology, which is provided in the table below. The current manuscript then
 540 becomes the effective publication for those names (Path 1).

541

542 **Table S3. Etymology and typification of higher taxa for *Kryptonium thompsonii*^{Ts}**

Proposed taxon	Etymology	Nomenclatural type
Family <i>Kryptoniaceae</i>	Kryp.to.ni.a'ce.ae N.L. neut. n. <i>Kryptonium</i> , type genus of the family; L. fem. pl. suff. <i>-aceae</i> , ending to denote a family; N.L. fem. pl. n. <i>Kryptoniaceae</i> , family of the genus <i>Kryptonium</i>	Genus <i>Kryptonium</i>
Order <i>Kryptoniales</i>	Kryp.to.ni.a'les N.L. neut. n. <i>Kryptonium</i> , type genus of the order; N.L. fem. pl. suff. <i>-ales</i> , ending denoting an order; N.L. fem. pl. n. <i>Kryptoniales</i> , order of the genus <i>Kryptonium</i>	Genus <i>Kryptonium</i>
Class <i>Kryptonia</i>	Kryp.to'ni.a N.L. neut. n. <i>Kryptonium</i> , type genus of the class; N.L. neut. n. suff. <i>-ia</i> , ending to denote a class; N.L. neut. pl. n. <i>Kryptonia</i> , class of the genus <i>Kryptonium</i>	Genus <i>Kryptonium</i>

543

544 **Additional species of the genus *Kryptonium*.** As described above, we have chosen to merge four
 545 genera originally proposed into the single genus *Kryptonium*. In the GTDB, these genomes are
 546 instead divided into six species clusters based on circumscription radii determined from ANI
 547 clustering, splitting the original taxon *Candidatus* *Kryptobacter tengchongensis* into three species.
 548 However, these two additional species clusters likely represent artificial clusters with very high
 549 sequence similarity because these clusters are based on three replicate MAG assemblies each.
 550 Regardless, only one other species cluster has a high-quality genome representative that could
 551 serve as a type under minimal standards described in Figure 1 (main text). Thus, we will only name
 552 a single additional species for the cluster originally proposed as *Candidatus* *Kryptothermus*
 553 *mobilis*, from Great Boiling Spring, Nevada, USA. This taxon was enriched *in situ* in Great Boiling
 554 Spring sediments following the addition of corn stover, consistent with many annotated glycoside
 555 hydrolases in the genome. The original publication has only one MAG; the GTDB cluster again
 556 has three MAGs, although all are publicly available replicate assemblies from the same

557 metagenomic sequencing run. We nevertheless choose to name this species based on the high
558 quality of the MAG (97.25 % completeness with 2.73 % estimated contamination; 44x coverage
559 and sequences of 5S, 16S and 23S rRNAs and tRNAs coding for 20 amino acids) and based on
560 recruitment of the genome at high identity in two different cellulolytic enrichments in Great
561 Boiling Spring that were incubated at 77 °C and 85 °C, providing clear evidence of growth at in
562 situ temperature and consistent with a role in consortial degradation of plant biomass. In addition,
563 21 MAGs and SAGs from different samples representing the genus are available for comparison,
564 providing confidence of genome size, assembly fidelity, and the presence/conservation of key
565 genes.

566 In the case of the union of the genus *Kryptonium* and *Candidatus* *Kryptothermus*, we retain the
567 genus name *Kryptonium*, as justified above. However, we consider this a nom. nov. because not
568 only is a different genus name used, but also the species name should be changed.

569 **Etymology and typification of *Kryptonium mobile* nom. nov.**

570 *Kryptonium mobile* (mo'bi.le. L. neut. adj. *mobile*, motile, moving).

571 Basonym: *Candidatus* *Kryptothermus mobilis* (Eloe-Fadrosh et al., 2016)

572 The nomenclatural type is NCBI Assembly GCA_900070205.1.

573

574 **Pre-registration of names in the SeqCode Registry.** To register names for taxa that encompass
575 names for a lineage where some were proposed in the effective publication and others that are new
576 (and have been effectively published), register as a “Contributor” to the SeqCode Registry. Select
577 “Create a name” option on the “Contributor” dashboard and register all names associated with the
578 lineage as described above in the *Wolframiraptoraceae* example (e.g., *Kryptonina*, *Kryptoniales*,
579 *Kryptoniaceae*, *Kryptonium*, *Kryptonium thompsonii*, and *Kryptonium mobile*). Link effective
580 publications for the taxon names proposed previously (e.g., *Kryptonium thompsonii* proposed by
581 Eloe-Fadrosh et al., (2016) [10.1038/ncomms10476], corrected by Oren et al., (2020)
582 [10.1099/ijsem.0.003789]) and create new names for those that have not been proposed before
583 (e.g., *Kryptonina*, *Kryptoniales*, *Kryptoniaceae*, and *Kryptonium mobile*). As these taxa would have
584 different effective publications associated with the taxon names, separate register lists need to be
585 created and associated with the appropriate effective publications. Add all of the taxa associated
586 with the previously published effective publication to a single register list for approval by a curator
587 (e.g., seqco.de/r:jvapsuy2) and all newly proposed names to a separate register list (e.g.,
588 seqco.de/r:rgrmqbgb). The provided register list accession URL for the newly proposed names
589 should be included in the effective publication to allow automated checks and validation upon
590 publication if approval was given by the curator. In this case, this manuscript serves as the effective
591 publication for *Kryptonina*, *Kryptoniales*, *Kryptoniaceae*, and *Kryptonium mobile*, and should be
592 provided for validation under “Validate Published Names” in PDF along with supplementary
593 material (if relevant).

594

References

- 596 1. Alneberg, J., et al. Binning metagenomic contigs by coverage and composition. *Nature*
597 *Methods* **11**, 1144–1146 (2014).
- 598 2. Beussecker S, Palmer M, Lai D, Dimapilis J, Mayali X, Mosier D, Jiao J-Y, Colman DR,
599 Keller LM, St. John E, Miranda M, Gonzalez C, Gonzalez L, Sam C, Villa C, Zou M,
600 Bodman N, Robles F, Boyd ES, Cox AD, St. Clair B, Hua Z-S, Li W-J, Reysenbach A-L,
601 Stott MB, Weber PK, Pett-Ridge J, Dekas AE, Hedlund BP, Dodsworth JA. Tungsten is
602 essential for carbohydrate metabolism in members of a novel family of anaerobic,
603 thermophilic *Caldarchaeales* (Aigarchaeota). In review *Nature Communications*.
- 604 3. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to
605 classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36** (6) 1925-1927
606 (2020).
- 607 4. Eloë-Fadrosh EA, Paez-Espino D, Jarett J, Dunfield PF, Hedlund BP, Dekas AE, Grasby
608 SE, Brady AL, Dong H, Briggs BR, Li WJ, Goudeau D, Malmstrom R, Pati A, Pett-Ridge
609 J, Rubin EM, Woyke T, Kyrpides NC, Ivanova NN. Global metagenomic survey reveals a
610 new bacterial candidate phylum in geothermal springs. *Nature Communications*. 7:10476.
611 (2016).
- 612 5. Hua ZS, Wang YL, Evans PN, Qu YN, Goh KM, Rao YZ, Qi YL, Li YX, Huang MJ, Jiao
613 JY, Chen YT, Mao YP, Shu WS, Hozzein W, Hedlund BP, Tyson GW, Zhang T, Li WJ.
614 Insights into the ecological roles and evolution of methyl-coenzyme M reductase-
615 containing hot spring archaea. *Nature Communications* 10(1):4574. (2019).
- 616 6. Kang, D. D, Froula, J., Egan, R., & Wang, Z. MetaBAT, an efficient tool for accurately
617 reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165
618 (2015).
- 619 7. Kang, D. D., et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient
620 genome reconstruction from metagenome assemblies. *PeerJ* **7**,e7359 (2019).
- 621 8. Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. MEGAHIT: an ultra-fast single-
622 node solution for large and complex metagenomics assembly via succinct de Bruijn graph.
623 *Bioinformatics* **31** (10), 1674-1676 (2015).
- 624 9. Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. metaSPAdes: a new versatile
625 metagenomic assembler. *Genome Research* **27** (5), 824-834 (2017).
- 626 10. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM:
627 assessing the quality of microbial genomes recovered from isolates, single cells, and
628 metagenomes. *Genome Research* **25**, 1043–1055 (2015).
- 629 11. Wu, Y.-W., Simmons, B. A., & Singer, S. W. MaxBin 2.0: an automated binning algorithm
630 to recover genomes from multiple metagenomic datasets, *Bioinformatics* **32** (4), 605–607
631 (2016).
- 632 12. Uritskiy, G. V., DiRuggiero, J., & Taylor, J. MetaWRAP-a flexible pipeline for genome-
633 resolved metagenomic data analysis. *Microbiome* **6** (1), 158 (2018).
- 634