

SeqCode, a nomenclatural code for prokaryotes described from sequence data

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

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

Supplementary Text

Page

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

Glossary

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

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

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

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

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

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

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

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

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

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

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

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

Later synonym: A subsequently validly published synonym [in some taxonomic literature, this is 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. | <i>Breakout group recommendation:</i> Mostly no. <i>Pros:</i> If so, this may encourage deposition of strains into culture collections, because the ICNP requires this. <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). | <i>Committee action:</i> SeqCode is not restricted to uncultivated organisms. See General Consideration 5 and Rule 23. <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. |
| SeqCode should allow replacement of type sequences with types strains. | <i>Breakout group recommendation:</i> No consensus. <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. <i>Cons:</i> If no, this would enhance the stability of types but lead to conflict between types and names with ICNP. | <i>Committee action:</i> SeqCode does not allow replacement of sequences with strains as types. <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. |
| SeqCode should specify that priority of higher taxa depends on the priority of the type genus. | <i>Breakout group recommendation:</i> Mostly no. <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. <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. | <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. <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. |
| SeqCode should include a proposal for an “Approved List” to incorporate the current <i>Candidatus</i> taxa. | <i>Breakout group recommendation:</i> Yes. <i>Pros:</i> If so, this list would leverage a large body of literature with minimal work and disruption to the community. <i>Cons:</i> None noted. | <i>Committee action:</i> We agree and will work with the research community on an “Approved List” following initial implementation of the SeqCode. <i>Rationale:</i> The list will serve the community and maximize SeqCode utility. |
| SeqCode should include a “digital protologue” as part of the naming process. | <i>Breakout group recommendation:</i> Yes. <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). | <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”. <i>Rationale:</i> A strong consensus emerged from the community that a “digital |

| | | |
|--|--|---|
| | <i>Cons:</i> A “digital protologue” would take substantial work and funding to set up and maintain. | protologue” and automated registration system would greatly support systematics. |
| SeqCode should require ‘perfect Latin’? | <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. | <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. |
| SeqCode should be managed and revised using two international committees mirroring the ICSP and Judicial Commission. | <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. | <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. |
| SeqCode should include the rank of subspecies. | <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. | <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. |

87

88

Description of the SeqCode Registry

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

Vision for adoption of *Candidatus* names

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

References:

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

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

Example of path 1: *Wolframiraptoraceae* and child taxa

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

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

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

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

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

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

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

Description of the family *Wolframiiiraptoraceae*

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

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

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

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

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

Description of the genus *Wolframiiiraptor*

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

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

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

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

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

Description of *Wolframiraptor gerlachensis*^{Ts}

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

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

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

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

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

| Proposed taxon | Etymology | Nomenclatural type |
|---|---|--|
| Family <i>Wolframiiiraptoraceae</i> | Wolf.ra.mi.i.rap.to.ra'ce.ae N.L. masc. n. <i>Wolframiiiraptor</i> , type genus of the family; L. fem. pl. suff. -aceae, ending to denote a family; N.L. fem. pl. n. <i>Wolframiiiraptoraceae</i> , family of the genus <i>Wolframiiiraptor</i> | Genus <i>Wolframiiiraptor</i> |
| Genus <i>Wolframiiiraptor</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>Wolframiiiraptor</i> , snatcher of tungsten | Species <i>Wolframiiiraptor gerlachensis</i> ^{Ts} |
| Species <i>Wolframiiiraptor 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} |

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

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

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

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

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

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

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

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

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

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

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

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

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

Description of the genus *Methanoproducendum*

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

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

Description of the species *Methanoproducendum senex*

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

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

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

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

Example of combination of path 1 and path 2 for validation: class *Kryptonina* and child taxa

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

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

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

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

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

Description of the genus *Kryptonium*

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

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

Description of the species *Kryptonium thompsonii*

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

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

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

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

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. -aceae, 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. -ales, 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. -ia, ending to denote a class; N.L. neut. pl. n. <i>Kryptonia</i> , class of the genus <i>Kryptonium</i> | Genus <i>Kryptonium</i> |

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

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

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

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

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

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

The nomenclatural type is NCBI Assembly GCA_900070205.1.

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

References

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