NamesforLife Release 20190727

During July 2019, there were a total of 154 changes as compared to **NamesforLife Release 20190614**, including 119 novel taxa, 1 replacement name, 5 rank elevations, 2 rank reductions, 8 new combinations, no corrections, no neotype/proxy types, 3 transfers of taxa and 7 changes in the preferred names appearing in the NamesforLife condensed taxonomy.

NamesforLife maintains and distributes two views of the monthly state of prokaryotic taxonomy, based on recent updates of the published record and supporting data. The taxonomies are purely hierarchical, and based on the validly published names (those which conform to the principles and rules set forth in the *International Code of Nomenclature of Prokaryotes*) appearing in the *International Journal of Systematic and Evolutionary Microbiology*. These taxonomies represent a consensus view of experts who apply a contemporary method of classification including phylogenetic analysis of the small ribosomal subunit (16S rRNA gene), pairwise comparison of genome sequences and phenotypic properties.

Table 1. A summary of the current state of taxonomy of prokaryotes with validly published names.

	Complete Taxonomy[a]	Condensed Taxonomy[b]	HQ16S[c]	Genomes (type)[d]	Genomes (non-type)[e]	Genomes (combined)
Phyla	50	39	39	38	35	36
Classes	162	97	98	91	74	89
Orders	392	244	244	116	125	124
Families	804	570	567	498	345	505
Genera	3,538	3,077	3,033	2,223	981	2,299
Species/Subsp.	20,596	16,421	16,218	8,405	3,852	9,521

[a] The *Complete Taxonomy* includes all published synonyms, homonyms and names that may be considered illegitimate, rejected, orthographically or grammatically incorrect or not validly published for a documented reason. It is used to establish nomenclatural accuracy and determining the correct current state of a name and to correctly interpreting the names appearing in older literature. It also includes a growing subset of published *Candidatus* taxa. Only those that are considered preferred names are presented here.

[b] The *Condensed Taxonomy* is a view of the current state of prokaryotic taxonomy and nomenclature that leverages features of the NamesforLife Information Architecture. Each species/subspecies is uniquely represented in a single point in the hierarchy, based on its most recent validly published name or revision in its circumscription or properties. Mapping to earlier states and all associated data and literature is addressed using NamesforLife DOIs.

[c] The *HQ16S* data sets consists of curated, high-quality 16S rRNA gene sequences used in the published descriptions of type strains of species/subspecies of bacteria and archaea with validly published names. Linking to verified deposits of viable type material in over 125 culture collections as well as earlier synonyms, the *HQ16S* dataset allows for accurate identification and naming of >98.75% of bacteria and archaea with validly published names.

[d-e] NamesforLife genome sequence data is a continuously updated version of publicly available prokaryotic genome assemblies. **Release 20190727** contains 205,822 records including 10,614 assemblies that were verified as sourced from 8,417 type strains with validly published names. This number was reduced to 8,025 type strains when excluded assemblies were removed from consideration.

Five hundred and twenty-nine type strain genome records were re-annotated to reflect recent changes in nomenclature. An additional 154,003 assemblies could be associated with a total of 4,213 taxa with validly published names at varying levels of taxonomic resolution. The taxonomic coverage of the combined data set remained at 9,521 validly named species/subspecies. There were 35,061 records in which the nomenclature was re-annotated. Of the remaining sequence records, 6,181 were identified as *Candidatus* taxa and could be placed into 566 discrete "groups" at varying levels of taxonomic resolution. This decrease represents an increase from June 2019 and partially reflects our effort to refine the groupings of candidate taxa based on depositor assigned names. As noted previously,

we expect that this number will continue to fluctuate over the next few months as we work towards achieving a more uniform naming of candidate taxa. The remaining 35,024 sequences were associated with names that have no standing in the nomenclature of prokaryotes. As in previous months we have also observed that assemblies present in a prior month have been suppressed or excluded. In **NamesforLife Release 20190727**, 16,493 assemblies present in **NamesforLife Release 20190614** do not appear this release and 10,454 sequences are new, which is a sharp departure from what we have previously observed.

Table 2. Summary of reported NCBI exclusions for N4L re-annotated genomes.

Reported exclusion category[a]	N4L type	N4L non-type	N4L *Candidatus*	N4L invalid
assembly from type material	10,301	628	10	643
assembly from synonym type material	69	158	0	0
assembly from proxytype material	6	38	0	11
assembly designated as neotype	9	0	0	0
assembly from pathotype material	0	28	0	0
assembly designated as reftype	0	19	1	1
untrustworthy as type	77	21	0	2
derived from environmental source	2	1,465	54	2,164
derived from metagenome	0	1,2034	5,279	13,858
derived from single cell	0	71	228	1,238
derived from surveillance project	0	1	0	0
chimeric	0	3	0	2
contaminated	25	224	6	53
mixed culture	3	4	0	2
unverified source organism	1	0	0	1
hybrid	0	1	0	0
misassembled	3	28	0	7
validation errors	1	253	0	3
genome length too large	37	687	19	13
genome length too small	10	317	41	8
partial	1	74	35	57
high contig L50	64	456	7	222
low contig N50	99	2,144	438	1,848
abnormal gene to sequence ratio	6	180	13	15
low gene count	0	1	3	0
low quality sequence	29	414	16	79
many frameshifted proteins	124	1,627	5	92
missing ribosomal protein genes	2	76	5	1
missing rRNA genes	17	81	11	8
missing tRNA genes	12	204	27	10

[a] Exclusions and relation to type material reported by NCBI and mapped to N4L categories of re-annotated genome assemblies. Genomes assemblies may have zero, one or more than class of exclusion reported.

For more information about NamesforLife services and data products visit https://namesforlife.com.