NamesforLife Release 20190930

During September 2019, there were a total of 126 changes as compared to **Release 2019828**, including 94 novel taxa, no replacement names, 1 rank elevation, no rank reductions, 13 new combinations, no corrections, no neotypes/proxy types, 4 transfer of taxa and 14 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	51	39	39	38	35	38
Classes	175	97	98	91	74	91
Orders	407	244	244	223	160	224
Families	822	572	569	499	347	515
Genera	3,547	3,094	3,049	2,249	1,023	2,358
Species/Subsp.	20,768	16,549	16,338	8,601	4,035	9,822

[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 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 set 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 98.73% 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 20190930** contains 212,284 records including 10,921 assemblies that were verified as sourced from 8,607 type strains with validly published names; 1,745 type strains were represented by two or more genome assemblies. This number was reduced to 8,257 type strains when excluded assemblies were removed from consideration.

Five hundred and fifty-four type strain genome records were re-annotated to reflect recent changes in nomenclature. An additional 157,654 assemblies could be associated with a total of 4,420 taxa with validly published names at varying levels of taxonomic resolution. The taxonomic coverage of the combined data has increased to 10,213 validly named species/subspecies. There were 35,760 records in which the nomenclature was re-annotated. Of the remaining sequence records, 6,311 were identified as *Candidatus* taxa and could be placed into 566 discrete "groups" at varying levels of taxonomic resolution. Of those, 238 have names that will appear on the soon-to-be-

published *Lists of names of prokaryotic Candidatus taxa* in the *International Journal of Systematic and Evolutionary Microbiology.* The remaining 37,397 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 20190930**, 294 assemblies present in **NamesforLife Release 20190828** do not appear this release and 6,502 sequences are new. While both numbers are significant lower than in July, the number of remain higher that what was observed earlier in 2019.

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	10,546	745	10	673
assembly from synonym type material	66	161	0	0
assembly from proxy type material	6	37	0	11
assembly designated as neotype	9	0	0	0
assembly from pathotype material	0	28	0	0
assembly designated as ref type	0	19	1	1
untrustworthy as type	79	22	0	2
derived from environmental source	2	1,465	53	2,176
derived from metagenome	0	12,232	5,391	15,289
derived from single cell	0	71	227	1,240
derived from surveillance project	0	0	0	0
chimeric	0	3	0	2
contaminated	26	269	6	66
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	279	0	5
genome length too large	37	707	19	14
genome length too small	11	324	41	8
partial	1	75	35	57
high contig L50	0	0	0	0
low contig N50	0	0	0	0
abnormal gene to sequence ratio	9	279	13	17
low gene count	0	1	3	0
low quality sequence	29	422	16	82
many frameshifted proteins	127	1,596	5	83
missing ribosomal protein genes	5	196	7	4
missing rRNA genes	14	81	11	8
missing tRNA genes	9	326	26	14

[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.