

```
In [1]: ! all_entities_Genome -show-fields
```

```
Available fields: pegs rnas scientific_name complete prokaryotic dna_size contigs  
domain genetic_code gc_content phenotype md5 source_id
```

```
In [2]: ! all_entities_Genome -f scientific_name | grep "Streptococcus pneumoniae"
```

```
kb|g.3478      Streptococcus pneumoniae SPN034183
kb|g.3264      Streptococcus pneumoniae GA47901
kb|g.21524     Streptococcus pneumoniae SP9-BS68
kb|g.10034     Streptococcus pneumoniae 70585
kb|g.162Streptococcus pneumoniae R6
kb|g.21849     Streptococcus pneumoniae ATCC 700669
kb|g.3483      Streptococcus pneumoniae SPN032672
kb|g.1116      Streptococcus pneumoniae D39
kb|g.3576      Streptococcus pneumoniae GA17545
kb|g.1878      Streptococcus pneumoniae G54
kb|g.1779      Streptococcus pneumoniae Taiwan19F-14
kb|g.9774      Streptococcus pneumoniae SP23-BS72
kb|g.161Streptococcus pneumoniae TIGR4
kb|g.1785      Streptococcus pneumoniae P1031
kb|g.9944      Streptococcus pneumoniae CDC1873-00
kb|g.3484      Streptococcus pneumoniae SPN033038
kb|g.2032      Streptococcus pneumoniae TCH8431/19A
kb|g.1341      Streptococcus pneumoniae SP23-BS72
kb|g.1339      Streptococcus pneumoniae SP18-BS74
kb|g.1804      Streptococcus pneumoniae SP14-BS292
kb|g.160Streptococcus pneumoniae TIGR4
kb|g.21723     Streptococcus pneumoniae Taiwan19F-14
kb|g.9570      Streptococcus pneumoniae D39
kb|g.21523     Streptococcus pneumoniae SP6-BS73
kb|g.1340      Streptococcus pneumoniae SP19-BS75
kb|g.3485      Streptococcus pneumoniae SPN7465
kb|g.1881      Streptococcus pneumoniae BS458
kb|g.9945      Streptococcus pneumoniae SP195
kb|g.1780      Streptococcus pneumoniae Hungary19A-6
kb|g.1574      Streptococcus pneumoniae CDC1873-00
kb|g.1900      Streptococcus pneumoniae CGSP14
kb|g.109Streptococcus pneumoniae INV200
kb|g.21724     Streptococcus pneumoniae Hungary19A-6
kb|g.21649     Streptococcus pneumoniae CDC1087-00
kb|g.1573      Streptococcus pneumoniae CDC1087-00
kb|g.1880      Streptococcus pneumoniae BS457
kb|g.1337      Streptococcus pneumoniae SP11-BS70
kb|g.1879      Streptococcus pneumoniae BS397
kb|g.3262      Streptococcus pneumoniae GA41317
kb|g.3263      Streptococcus pneumoniae GA47368
kb|g.3515      Streptococcus pneumoniae SPN061370
kb|g.3480      Streptococcus pneumoniae SPN021198
kb|g.20651     Streptococcus pneumoniae
kb|g.1336      Streptococcus pneumoniae SP9-BS68
kb|g.2593      Streptococcus pneumoniae BS455
kb|g.21725     Streptococcus pneumoniae JJA
kb|g.3477      Streptococcus pneumoniae SPN034156
kb|g.3475      Streptococcus pneumoniae INV200
kb|g.1334      Streptococcus pneumoniae SP3-BS71
kb|g.1338      Streptococcus pneumoniae SP14-BS69
kb|g.3476      Streptococcus pneumoniae INV104
kb|g.1335      Streptococcus pneumoniae SP6-BS73
kb|g.3260      Streptococcus pneumoniae GA17570
kb|g.187Streptococcus pneumoniae 670-6B
kb|g.3474      Streptococcus pneumoniae OXC141
kb|g.1576      Streptococcus pneumoniae CDC0288-04
kb|g.9770      Streptococcus pneumoniae SP3-BS71
kb|g.3261      Streptococcus pneumoniae GA41301
kb|a.21650     Streptococcus pneumoniae CDC0288-04
```

```
In [3]: ! all_entities_Genome -f scientific_name | grep "Streptococcus pneumoniae" | genomes_t
```

kb g.3478	Streptococcus pneumoniae	SPN034183	kb g.3478.peg.287	Triosephosphate isomerase (EC
kb g.21524	Streptococcus pneumoniae	SP9-BS68	kb g.21524.CDS.1282	triosephosphate isomerase
kb g.21524	Streptococcus pneumoniae	SP9-BS68	kb g.21524.CDS.1614	triosephosphate isomerase
kb g.162	Streptococcus pneumoniae R6	kb g.162.peg.336	Triosephosphate isomerase (EC	
kb g.21849	Streptococcus pneumoniae	ATCC 700669	kb g.21849.CDS.1471	triosephosphate isomerase
kb g.3483	Streptococcus pneumoniae	SPN032672	kb g.3483.peg.1960	Triosephosphate isomerase
kb g.1116	Streptococcus pneumoniae	D39	kb g.1116.peg.63	Triosephosphate isomerase
kb g.1878	Streptococcus pneumoniae	G54	kb g.1878.peg.102	Triosephosphate isomerase
kb g.1779	Streptococcus pneumoniae	Taiwan19F-14	kb g.1779.peg.617	Triosephosphate isomerase
kb g.9774	Streptococcus pneumoniae	SP23-BS72	kb g.9774.peg.1174	triosephosphate isomerase
kb g.9774	Streptococcus pneumoniae	SP23-BS72	kb g.9774.peg.1418	triosephosphate isomerase
kb g.161	Streptococcus pneumoniae TIGR4	kb g.161.peg.903	Triosephosphate isomerase (EC	
kb g.1785	Streptococcus pneumoniae	P1031	kb g.1785.peg.461	Triosephosphate isomerase
kb g.3484	Streptococcus pneumoniae	SPN033038	kb g.3484.peg.1907	Triosephosphate isomerase
kb g.2032	Streptococcus pneumoniae	TCH8431/19A	kb g.2032.peg.428	Triosephosphate isomerase
kb g.1341	Streptococcus pneumoniae	SP23-BS72	kb g.1341.peg.225	Triosephosphate isomerase
kb g.1339	Streptococcus pneumoniae	SP18-BS74	kb g.1339.peg.1032	Triosephosphate isomerase
kb g.1339	Streptococcus pneumoniae	SP18-BS74	kb g.1339.peg.331	Triosephosphate isomerase
kb g.1804	Streptococcus pneumoniae	SP14-BS292	kb g.1804.peg.833	Triosephosphate isomerase
kb g.160	Streptococcus pneumoniae TIGR4	kb g.160.peg.549	Triosephosphate isomerase (EC	
kb g.9570	Streptococcus pneumoniae	D39	kb g.9570.peg.1439	triosephosphate isomerase
kb g.21523	Streptococcus pneumoniae	SP6-BS73	kb g.21523.CDS.334	triosephosphate isomerase
kb g.21523	Streptococcus pneumoniae	SP6-BS73	kb g.21523.CDS.683	triosephosphate isomerase
kb g.1340	Streptococcus pneumoniae	SP19-BS75	kb g.1340.peg.783	Triosephosphate isomerase
kb g.3485	Streptococcus pneumoniae	SPN7465	kb g.3485.peg.936	Triosephosphate isomerase
kb g.1881	Streptococcus pneumoniae	BS458	kb g.1881.peg.318	Triosephosphate isomerase
kb g.9945	Streptococcus pneumoniae	SP195	kb g.9945.peg.554	triosephosphate isomerase
kb g.1780	Streptococcus pneumoniae	Hungary19A-6	kb g.1780.peg.722	Triosephosphate isomerase
kb g.1574	Streptococcus pneumoniae	CDC1873-00	kb g.1574.peg.2069	Triosephosphate isomerase
kb g.1574	Streptococcus pneumoniae	CDC1873-00	kb g.1574.peg.2405	Triosephosphate isomerase
kb g.1900	Streptococcus pneumoniae	CGSP14	kb g.1900.peg.651	Triosephosphate isomerase
kb g.21649	Streptococcus pneumoniae	CDC1087-00	kb g.21649.CDS.1827	triosephosphate isomerase
kb g.1573	Streptococcus pneumoniae	CDC1087-00	kb g.1573.peg.829	Triosephosphate isomerase
kb g.1880	Streptococcus pneumoniae	BS457	kb g.1880.peg.339	Triosephosphate isomerase
kb g.1337	Streptococcus pneumoniae	SP11-BS70	kb g.1337.peg.2003	Triosephosphate isomerase
kb g.1879	Streptococcus pneumoniae	BS397	kb g.1879.peg.153	Triosephosphate isomerase
kb g.3515	Streptococcus pneumoniae	SPN061370	kb g.3515.peg.1042	Triosephosphate isomerase
kb g.3480	Streptococcus pneumoniae	SPN021198	kb g.3480.peg.1923	Triosephosphate isomerase
kb g.1336	Streptococcus pneumoniae	SP9-BS68	kb g.1336.peg.490	Triosephosphate isomerase
kb g.2593	Streptococcus pneumoniae	BS455	kb g.2593.peg.305	Triosephosphate isomerase
kb g.3477	Streptococcus pneumoniae	SPN034156	kb g.3477.peg.1423	Triosephosphate isomerase
kb g.3475	Streptococcus pneumoniae	INV200	kb g.3475.peg.191	Triosephosphate isomerase
kb g.1334	Streptococcus pneumoniae	SP3-BS71	kb g.1334.peg.281	Triosephosphate isomerase
kb g.1338	Streptococcus pneumoniae	SP14-BS69	kb g.1338.peg.515	Triosephosphate isomerase
kb g.3476	Streptococcus pneumoniae	INV104	kb g.3476.peg.501	Triosephosphate isomerase
kb g.1335	Streptococcus pneumoniae	SP6-BS73	kb g.1335.peg.2064	Triosephosphate isomerase
kb g.187	Streptococcus pneumoniae 670-6B	kb g.187.peg.508	Triosephosphate isomerase (EC	
kb g.3474	Streptococcus pneumoniae	OXC141	kb g.3474.peg.105	Triosephosphate isomerase
kb g.1576	Streptococcus pneumoniae	CDC0288-04	kb g.1576.peg.986	Triosephosphate isomerase
kb g.9770	Streptococcus pneumoniae	SP3-BS71	kb g.9770.peg.1344	triosephosphate isomerase
kb g.9770	Streptococcus pneumoniae	SP3-BS71	kb g.9770.peg.1935	triosephosphate isomerase
kb g.21650	Streptococcus pneumoniae	CDC0288-04	kb g.21650.CDS.378	triosephosphate isomerase
kb g.3479	Streptococcus pneumoniae	SPN072838	kb g.3479.peg.620	Triosephosphate isomerase
kb g.1575	Streptococcus pneumoniae	SP195	kb g.1575.peg.1499	Triosephosphate isomerase
kb g.9771	Streptococcus pneumoniae	SP14-BS69	kb g.9771.peg.1228	triosephosphate isomerase
kb g.9771	Streptococcus pneumoniae	SP14-BS69	kb g.9771.peg.1728	triosephosphate isomerase
kb g.110	Streptococcus pneumoniae OXC141	kb g.110.peg.1350	Triosephosphate isomerase	
kb g.108	Streptococcus pneumoniae INV104B	kb g.108.peg.1305	Triosephosphate isomerase	
kb g.20736	Streptococcus pneumoniae TIGR4	kb g.20736.CDS.1519	triosephosphate isomerase	

```
In [4]: %%bash
all_entities_Genome -f scientific_name | grep "Streptococcus pneumonia" > all_strep
ls -a
echo ">>>> head"
head all_strep
echo ">>>> tail"
tail all_strep

.
..
all_strep
CDM_tutorial.ipynb
communities_analysis_tutorial.ipynb
genome_analysis_tutorial.ipynb
images
js
qc_tutorial.ipynb
REST_tutorial.ipynb
Rplots.pdf
>>>> head
kb|g.3478      Streptococcus pneumoniae SPN034183
kb|g.3264      Streptococcus pneumoniae GA47901
kb|g.21524     Streptococcus pneumoniae SP9-BS68
kb|g.10034     Streptococcus pneumoniae 70585
kb|g.162Streptococcus pneumoniae R6
kb|g.21849     Streptococcus pneumoniae ATCC 700669
kb|g.3483      Streptococcus pneumoniae SPN032672
kb|g.1116      Streptococcus pneumoniae D39
kb|g.3576      Streptococcus pneumoniae GA17545
kb|g.1878      Streptococcus pneumoniae G54
>>>> tail
kb|g.21525     Streptococcus pneumoniae SP11-BS70
kb|g.398Streptococcus pneumoniae 23F
kb|g.3482      Streptococcus pneumoniae SPN1041
kb|g.1115      Streptococcus pneumonia pneumoniae D39
kb|g.10035     Streptococcus pneumoniae P1031
kb|g.9773      Streptococcus pneumoniae SP19-BS75
kb|g.10086     Streptococcus pneumoniae G54
kb|g.1577      Streptococcus pneumoniae CDC3059-06
kb|g.3259      Streptococcus pneumoniae GA04375
kb|g.21652     Streptococcus pneumoniae MLV-016
```

```
In [5]: ! all_entities_Publication -a | head
```

```
2673545 SPT15, the gene encoding the yeast TATA binding factor TFIID, is required
for normal transcription initiation in vivo.      http://www.ncbi.nlm.nih.gov/pubmed
/26735451331480968
10613857Role of HrcA and CIRCE in the heat shock regulatory network of
Bradyrhizobium japonicum.      http://www.ncbi.nlm.nih.gov/pubmed/10613857
1331479535
7961401 Sulfide dehydrogenase from the hyperthermophilic archaeon Pyrococcus
furiosus: a new multifunctional enzyme involved in the reduction of elemental
sulfur. http://www.ncbi.nlm.nih.gov/pubmed/7961401      1331160047
8165147 A putative regulatory gene downstream of recA is conserved in
gram-negative and gram-positive bacteria.      http://www.ncbi.nlm.nih.gov/pubmed
/81651471331526910
8487307 Primary structure and crystallization of orotate phosphoribosyltransferase
from Salmonella typhimurium.      http://www.ncbi.nlm.nih.gov/pubmed/8487307
1331526958
10644761Thioredoxin-dependent hydroperoxide peroxidase activity of
bacterioferritin comigratory protein (BCP) as a new member of the thiol-specific
antioxidant protein (TSA)/Alkyl hydroperoxide peroxidase C (AhpC) family.
http://www.ncbi.nlm.nih.gov/pubmed/10644761      1331479544
9030723 Identification and purification of translation initiation factor 2 (IF2)
from Thermus thermophilus.      http://www.ncbi.nlm.nih.gov/pubmed/9030723
1331527025
16216571YjjX: from structure "Tu" function.      http://www.ncbi.nlm.nih.gov/pubmed
/16216571      1331159922
15209499Nickel superoxide dismutase structure and mechanism.
http://www.ncbi.nlm.nih.gov/pubmed/15209499      1331480208
9103608 Citrate synthase gene comparison, a new tool for phylogenetic analysis,
and its application for the rickettsiae.http://www.ncbi.nlm.nih.gov/pubmed
/91036081331527032
```

```
In [6]: ! echo 'kb|g.0' | get_entity_Genome -f 'scientific_name,contigs,dna_size'
```

```
kb|g.0 Escherichia coli K12      1      4639221
```

```
In [7]: ! echo 'kb|g.3857' | get_relationship_IsComposedOf -to id | contigs_to_sequences
```

```
>kb|g.3857.c.0
agagattacgtctggttgcaagagatcatgacagggggaattggtgaaaataaatatat
cgccagcagcacatgaacaagtttcggaatgtgatcaatttaaaaaatttattgacttagg
cgggcagatactttaaccaatataggaatacaagacagacaaataaaaaatgacagagtag
acaacatccatgaaccgcatcagcaccaccaccattaccaccatcaccattaccacaggt
aacggtgcgggctgacgcgtacaggaaacacagaaaaaagcccgccacctgaacagtgcgg
gcttttttttcgaccagagatcacgaggttaacaacctgcgagtggtgaagttcggcggt
acatcagtggtcaaatgcagaacgtttttctgcgtgttgccgatattctggaaagcaatgcc
aggcaagggcaggtagcgaccgtactttccgcccccgcaaaattaccaacctctggtg
gcaatgattgaaaaaactatcgcgggccaggtgctttgccgaatatcagcgatgcagaa
cgtattttttctgacctgctcgcaggacttgccagcgccagccggattccccgcttgca
cgggttgaatatggttgctgaacaagaattcgctcagatcaaacatgttctgcatggtatc
agcctgctgggtcagtgcccgatagcatcaacgcccgcgctgatttgccgtggcgaaaaa
atgctcgatcgcatatggtgggacttctggagggcgctgggcatcgcgtcacgggtgatc
gatccggtagaaaaattgctggcggtggccattacctgaatctaccggttgatcgcg
gaatcgactcgccgtatcgccgccagccagatcccgccgatcacatgatcctgatggcg
ggctttaccgcccgtaatgaaaagggtgaactggtggtgctgggcccgtaatggtccgac
tattccgcccgtgctggccgctgtttacgcgctgactgctgtgaaatctggactgac
gtcgatggcggtgtatacctgtgacccgcgccaggtgccggacgccaggtgctgaaatcg
atgtcctaccaggaagcgatggaactctcttacttcggcgccaaagtccctcaccctcgc
accattacgcccacgcgccagttccagatccccctgtctgattaaaaataccggtaatccg
caggcgccaggaacgctgatcggcgctccagcgacgatgataacctgccggttaaaggg
atctctaaccttaacaacatggcgatgtttagcgtctccggcccgggaatgaaagggatg
attgggatggcgcgcgctgttttcgcccacatgtctcgcgcgggatctcgggtggtgctc
attaccagtcctcctctgagtacagcatcagtttctgtgtgccgagagtgactgcgcg
cgtgcccgccgtgcgatgcaggatgagttctatctggagctgaaagaggggctgctggag
ccgctggcggttacggagcggttggtgattatctctgttgctggcgacggtatgcgcacg
ctacgcggcatttcagcgaaattcttcgcccgcgtggcgcggttaatatcaatatcgtg
gcatcgctcaggatcttctgagcgcttctatttctgtggtggtgaataacgacgatgcc
accaccggcggtgcgggtaacgcaccagatgctgttcaataccgatcaggtgattgaagtg
tttgatattggcgctcggcggtcggcgcgctactggaacagcttaaactcagcaa
acctggtgaagaacaagcacatcgatctacgcgtgtgcggcggtggcgaactcaaaggcg
ttgctaaccaatgtgatggcctgaatctggacaactggcagggcgaactggcgcaagcg
aacgcgcggttcaatctgggacgcttaattcgccgtggtgaaagaatatcatctactcaat
ccggtgattggtgattgtacctccagtcaggcggtggccgaccagatgcggacttccgtg
cgtgaaaggggtccatgtggtgacgccaacaagaaagcgaacacctcgtcaatggactac
taccatcagctacgttttcgcccgcgcgaatcacggcgcaaatcttctgtatgacaccaac
gtcggcgccggtttgcccgttaactgaaaacctgcaaaacctgctgaatgccggtgatgaa
ctgcaaaaaattttccggcattctttccgggtcgtctctttttattttcggtaaaactggaa
gaggggatgagtcctcacaggcgaccgcccgtggcgcgagatgggctataccgaaccc
gatccgcgcgacgatctttccggtatggatgtggcgcgtaaaactgttgatcctcgccgc
gagacggggcgcgagctggagctttccgatatgtgattgaaccggtgttgccgaacgag
tttgacgectccggcgatgtgaccgcctttatggcacatctgccgcagcttgacgacgcg
tttgccgcccgtgtggcgaaagctcgtgatgaaggtaaggtattgcgctatgtgggcaat
atcgaagagggtggcggtgctgcccgtgaaaattgccgaagttgatggtaacgatccgctc
ttcaaaagtgaaaaaacgggtgaaaacgcgctggcggttctacagccactattatcagccctg
ccggttggtgctgcgcgggtacggcgaggaatgatgtgacggcgcgggcggtgtttgcc
gatctgttacggaccctctcatggaagttaggagtttaacatggtgaaagtgtatgcccc
ggcttcagcgcgcaacatgagcgtcggtttcgacgtgttggcgcgacgggtcacaccggt
tgacggcacgttgctggcgatgtggtatccggtgaagcagcgaccatttccgtctgca
taacctggggcgatttgccgataaaactgccgccggagccgcgtgaaaatattgtttatca
gtgctgggaacgtttttgccagggcattggggaaaacctcccgggtggcgatgacgctgga
aaaaaatatgccgattggttccgggttaggggtccagcgccgtgttccgctcgtcgcgcgct
ggtcgcgatgaatgagcactgcggcaaacggttaaacgacacgcgctctgttggcgctgat
ggcgagctggaaggccgtatctccggcagcatccattacgataacgctcgcgcgctgctt
tcttggcggtgaaggcgttgatgattgaagaaaacggcattattagtcagcaggtgccggg
ctttgatgagtggtactgggtactggcttatccgggcattaaagtgttctaccgcagaagc
acaaacacattttacctacacaatatcaccatcaaaattacattacacataaacacacatct
```

In [7]: