

Custom preformatted  
databases Lotus

# LotuS custom sequence databases (2)

- LotuS preformatted databases consist of two file:
  - Fasta file (**DNA**, can be translated RNA) – unique, short header to each ref sequence
  - Taxonomy file – contains for each Fasta header taxonomy information
- Fasta file is straightforward to create – both strands will be searched, so sequence orientation is not important
- Taxonomy file contains fasta identifier (without “>”), tab, and “;” delimited list of taxonomic levels:
  - HP451749      k\_\_Eukaryota; p\_\_Basidiomycota; c\_\_Pucciniomycetes; o\_\_Pucciniales; f\_\_Pucciniaceae; g\_\_Puccinia; s\_\_Puccinia triticina

# LotuS custom sequence databases

- Additionally, the level is indicated with a single letter in the tax string:
  - k = kingdom
  - p = phylum
  - c = class
  - o = order
  - f = family
  - g = genus
  - s = species
- Each unknown level is marked with level and “?”
- E.g. HP451XXX k\_\_Eukaryota; p\_\_Basidiomycota; c\_\_Pucciniomycetes; o\_\_Pucciniales; f\_\_Pucciniaceae; g\_\_?; s\_\_?

# Example: Silva

## otusdir/DB/SLV\_123\_SSU.fasta

```
>HP451749
CCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTATAAACAACTATACAGT
AAACTGCGAATGGCTCATTAAATCAGTTATAGTTTATTTGATGATACCTTACTACATGGATAACTGTGGTAATCTAGAG
CTAATACATGCTGAAAAGCCCAACCTTTGGAAGGGGTGATTTATTAGATAAAAAACCAATGGCTTCGGGTCTCTTTG
GTGATTCATAATAACTCTCGAATCGCATGGCCCTGTGCCGGTATGCTTCATTCAAATATCTGCCCTATCAACTTCGA
TGGTAGGATAGAGGCTACCATGGTATGACGGGTAACGGGGAATAAGGGTTCGATTCGGAGAGAGGGGCTGAGAAACG
GCCCTCAAATCTAAGGATTGCAGCAGGCGCGCAAATTACCAATCCTGACACAGGGAGGTAGTGACAATAAATAACAATG
TATGGCTCTTTGGGCTTACAATTGGAATGAGTACAATTTAAATCTTTAACGAGGATCAATGGAGGGCAAGTCTGGT
GCCAGCAGCCGCGTAATCCAGCTCCAATAGCGTATATTTAAATTTGTTGACGTTAAAAAGCTCGTAGTCGAACTTCGGC
CTCTGGCAGTTGGTCCGCTTTGGTGTGACTGATTTGTTGGAGGCTTACCTCTGGTGAATTCATGCACTTACTG
GGTGTGGAAGGGAACAGGACTTTTACTTTGAAAAAATTAGAGTGTCAAAGCAGGCTTATGCCTGAATACATTAGCATG
GAATAATAAATAGGACGTGTGATTCTATTTTGTGGTTTCTAGGATTACCGTAATGATGAATAGGGTCAGTTGGGGGCA
TTGTATTACATCGTCAGAGGTGAAATCTTGGATTGATGTAAGACAACTACTGCGAAAGCATCTGCAAGGATGACTT
CATTGATCAAGAACGAAGGTTAAGGGTTCAAAAACGATCAGATACCGTTGTAGTCTTAACAGTAAACTATGCCGACTGGG
GATCAGACAAGGATTTAATGACTTGTCTGGCACCCAAAGGAAACCTTAAGTTAAGGGGGGGGAGTGAATGAGGATAA
GAGTATGGGATTTGAAAATAGCGAGGCGTAGGGCACCACAGGTTGTGAGCCTGCGGCTTAATTTGACTCAACAGGGG
AAACTCACCAGGTCAGACACAGTAAGGATTGACAGATTGATAGCTTTTTCTTGATTTGTGGTGGTGGTGCATGGCCG
TTCTTAGTTGGGTGGAGTGATTTGTCTGGGTAATCCGATAACAACAACAACTTCTCTGCTAAATAGTCCAGCTGGCT
ACGGCTGGCTGCAGACTTCTAAAGGGACTATCAGACGTTTAGTTGATGGAAGTTGGAGGCAATAACAGGCTGTGATGC
CCTTAGATGTTCTGGGCCACACGCTCTACACTGACCAAGCAACGAGTATATCACTTATCTGAAAAGCGGGGGTAA
TCTTGTGAAACTGGTCTGATGGGGATAGAGCATTGCAATTTGCTCTTCAACGAGGAATACCTAGTAAGCGTATGTC
ATCAGCATGCGTTGATTACGTCCTTGCCTTTGTACACACCGCCGTCGCTACTACCGATTGGCAGGCTTTTTGAGACGT
TCATTCGGATAATGCGTTGGCGGACACCATCAGCGTTGGTTAGCGAATGCGGTCGACCACTGTCAGACGAAGACGCC
CAATCTTGCTATTGGGTTGGGTTCTA
>AB002583
CAAAAAGCGAGTTTATGCTGGCTCAGGATGAACGCTGGCGGTATGCTTAACACATGCAAGTCGAACGCGCGTAAGTG
GCGTGCGAACGCGGTGAGTAACACGTGAGAATCTGCCCTAGGAGTTGGATAAGGCTTGGAAACGAGCGCTAAACCAACA
TATAAGGAAAGGAGATCGCCTAGGGAAGAGCTCGCGGCTGATTAGCTAGTTGGTAGGGTAAAGGCTACCAAGGCGAT
GATCAGTAGCTGGTCTGAGAGGATGACCAGCCACACTGGGACTGAGACACGCGCCACAGCTCTACGGGAGGCAAGGCG
GGAATTTCTGCAATGGCGAAAGCCTGACAGAGCAATCCGCGTAGGATGAAGGCTTAGGTTGTCAACCTCTTTT
.....
```

## lotusdir/DB/SLV\_123\_SSU.tax

```
head SLV_123_SSU.tax
HP451749 k_Eukaryota; p_Basidiomycota; c_Pucciniomycetes;
o_Pucciniales; f_Pucciniaceae; g_Puccinia; s_Puccinia triticina
AB002583 k_Bacteria; p_Cyanobacteria; c_Chloroplast;
o_Cyanidioschyzon merolae; f_; g_; s_
FJ904637 k_Eukaryota; p_Apicomplexa; c_Conoidasida;
o_Coccidia; f_Eimeriorina; g_Calyptosporidae; s_Calyptospora
spīnosa
FJ906773 k_Eukaryota; p_Arthropoda; c_Maxillopoda;
o_Pedunculata; f_?; g_?; s_Lepas anatifera
FJ911809 k_Eukaryota; p_Arthropoda; c_Arachnida; o_?; f_?;
g_?; s_Gamasellodes adrianae
FJ911827 k_Eukaryota; p_Arthropoda; c_Arachnida; o_?; f_?;
g_?; s_Dendrolaelaps sp. 2 APGD-2010
AB012846 k_Eukaryota; p_; c_Chlorophyceae;
o_Sphaeropleales; f_; g_Coelastrella; s_Coelastrella striolata var.
multistriata
AB013182 k_Eukaryota; p_Bangiales; c_?; o_?; f_?; g_?;
s_Porphyra sp. Shimonoseki
FJ911854 k_Eukaryota; p_Arthropoda; c_Arachnida; o_?; f_?;
g_?; s_Ornithonyssus bursa
FJ911859 k_Eukaryota; p_Arthropoda; c_Arachnida; o_?; f_?;
g_?; s_Spelaerhynchus praecursor
```