

Tutorial: Extracting 16S rRNA sequences for particular organisms from NCBI's locally installed nt database using blastdbcmd

Umer Zeeshan Ijaz

Summary of one-liners:

For each command given below, the first blastdbcmd invocation produces 2 entries per sequence (GI and title of the sequence), the awk command selects from the output of that command those sequences which have '16S' in their title and prints their GIs, and finally the second blastdbcmd invocation uses those GIs to print the sequence data for the 16S rRNA sequences in the nt database. Since nt database is huge, the methods given below use nohup to run the command in the background and they will continue even if you log off from a remote SSH session.

Method 1: Specify the organism names in the command itself

Command:

```
[uzi@quince-srv2 ~]$ nohup /home/opt/ncbi-blast-2.2.28+/bin/blastdbcmd -db /home/opt/ncbi-blast-2.2.28+/db/nt -entry all -outfmt "%g;%t" | awk -F";;" '/16S/ && /Archaeoglobus fulgidus|Desulfovibrio piger|Herpetosiphon aurantiacus|Ignicoccus hospitalis|Methanocaldococcus jannaschii|Methanococcus maripaludis|Nanoarchaeum equitans|Pyrobaculum aerophilum|Pyrobaculum calidifontis|Pyrococcus horikoshii|Rhodopirellula baltica|Sulfolobus tokodaii/{print $1}' | /home/opt/ncbi-blast-2.2.28+/bin/blastdbcmd -db /home/opt/ncbi-blast-2.2.28+/db/nt -entry_batch - -out sequences.fasta &
```

Method 2: Specify the organism names in a file organisms.txt

```
[uzi@quince-srv2 ~]$ cat organisms.txt
Archaeoglobus fulgidus
Desulfovibrio piger
Herpetosiphon aurantiacus
```

Ignicoccus hospitalis
Methanocaldococcus jannaschii
Methanococcus maripaludis
Nanoarchaeum equitans
Pyrobaculum aerophilum
Pyrobaculum calidifontis
Pyrococcus horikoshii
Rhodopirellula baltica
Sulfolobus tokodaii

Command:

```
[uzi@quince-srv2 ~]$ nohup /home/opt/ncbi-blast-2.2.28+/bin/blastdbcmd -db /home/opt/ncbi-blast-2.2.28+/db/nt -entry  
all -outfmt "%g; ;%t" | grep -F "$(cat organisms.txt)" | awk -F";;" '{print $1}' | /home/opt/ncbi-blast-  
2.2.28+/bin/blastdbcmd -db /home/opt/ncbi-blast-2.2.28+/db/nt -entry_batch - -out sequences2.fasta &
```

Display the header information of the sequences we have pulled out:

```
[uzi@quince-srv2 ~]$ grep -i ">" sequences.fasta
```

```
>gi|444303911|ref|NR_074334.1| Archaeoglobus fulgidus DSM 4304 strain DSM 4304 16S ribosomal RNA, complete sequence >gi|38792|emb|X05567.1|  
Archaeoglobus fulgidus strain VC-16 16S ribosomal RNA gene  
>gi|444303911|ref|NR_074334.1| Archaeoglobus fulgidus DSM 4304 strain DSM 4304 16S ribosomal RNA, complete sequence >gi|38792|emb|X05567.1|  
Archaeoglobus fulgidus strain VC-16 16S ribosomal RNA gene  
>gi|38805|emb|Y00275.1| Archaeoglobus fulgidus (VC-16) DNA for 16S ribosomal RNA  
>gi|294440|gb|L07510.1|PYBRGD Pyrobaculum aerophilum 16S ribosomal RNA gene  
>gi|1145365|gb|U38484.1|MMU38484 Methanococcus maripaludis 16S ribosomal RNA, complete sequence  
>gi|1145366|gb|U38486.1|MMU38486 Methanococcus maripaludis 16S ribosomal RNA, complete sequence  
>gi|1145367|gb|U38487.1|MMU38487 Methanococcus maripaludis 16S ribosomal RNA, complete sequence  
>gi|1145368|gb|U38941.1|MMU38941 Methanococcus maripaludis 16S ribosomal RNA, complete sequence  
>gi|2232234|gb|AF005049.1|AF005049 Methanococcus maripaludis 16S ribosomal RNA gene, complete sequence  
>gi|2564752|dbj|D45214.1|PYWOT3 Pyrococcus horikoshii 16S rRNA gene, complete sequence  
>gi|343201091|ref|NR_041778.1| Desulfovibrio piger ATCC 29098 strain ATCC29098 16S ribosomal RNA, complete sequence  
>gi|6979515|gb|AF192152.1|AF192152 Desulfomonas pigra 16S ribosomal RNA gene, complete sequence  
>gi|18151366|dbj|AB022438.1| Sulfolobus tokodaii genes for 16S rRNA, 23S rRNA, complete and partial sequence  
>gi|20429035|emb|AJ318041.1| Nanoarchaeum equitans Kin4-M 16S rRNA gene, strain Kin4-M  
>gi|265678650|ref|NR_028955.1| Ignicoccus hospitalis KIN4/I 16S ribosomal RNA, partial sequence >gi|20429036|emb|AJ318042.1| Ignicoccus sp. Kin4-  
I 16S rRNA gene, strain Kin4-I
```

>gi|343200235|ref|NR_040922.1| Pyrobaculum calidifontis JCM 11548 strain VA1 16S ribosomal RNA, complete sequence >gi|27807847|dbj|AB078332.1| Pyrobaculum calidifontis gene for 16S rRNA, complete sequence
>gi|343200235|ref|NR_040922.1| Pyrobaculum calidifontis JCM 11548 strain VA1 16S ribosomal RNA, complete sequence >gi|27807847|dbj|AB078332.1| Pyrobaculum calidifontis gene for 16S rRNA, complete sequence
>gi|2913866|dbj|D87344.1| Pyrococcus horikoshii gene for 16S ribosomal RNA, partial sequence
>gi|89077582|gb|DQ374392.1| Archaeoglobus fulgidus strain L3 16S ribosomal RNA gene, partial sequence
>gi|265678749|ref|NR_029054.1| Pyrococcus horikoshii strain JA-1 16S ribosomal RNA, partial sequence >gi|37543825|gb|AY099169.1| Pyrococcus horikoshii strain DSM 12428 16S ribosomal RNA gene, partial sequence
>gi|265678749|ref|NR_029054.1| Pyrococcus horikoshii strain JA-1 16S ribosomal RNA, partial sequence >gi|37543825|gb|AY099169.1| Pyrococcus horikoshii strain DSM 12428 16S ribosomal RNA gene, partial sequence
>gi|37543862|gb|AY099205.1| Pyrococcus horikoshii strain DSM 12428 16S-23S ribosomal RNA intergenic spacer region, complete sequence
>gi|116496446|gb|EF012748.1| Rhodopirellula baltica strain OJF4 16S ribosomal RNA gene, partial sequence
>gi|116496447|gb|EF012749.1| Rhodopirellula baltica strain OJF5 16S ribosomal RNA gene, partial sequence
>gi|116496448|gb|EF012750.1| Rhodopirellula baltica strain OJF6 16S ribosomal RNA gene, partial sequence
>gi|126211523|gb|EF421451.1| Rhodopirellula baltica strain OJF13 16S ribosomal RNA gene, partial sequence
>gi|126211524|gb|EF421452.1| Rhodopirellula baltica strain OJF14 16S ribosomal RNA gene, partial sequence
>gi|126211525|gb|EF421453.1| Rhodopirellula baltica strain OJF15 16S ribosomal RNA gene, partial sequence
>gi|148357800|gb|EF589343.1| Rhodopirellula baltica strain F1 16S ribosomal RNA gene, partial sequence
>gi|148357801|gb|EF589344.1| Rhodopirellula baltica strain Gt1 16S ribosomal RNA gene, partial sequence
>gi|148357802|gb|EF589345.1| Rhodopirellula baltica strain G1 16S ribosomal RNA gene, partial sequence
>gi|148357804|gb|EF589347.1| Rhodopirellula baltica strain M5 16S ribosomal RNA gene, partial sequence
>gi|148357806|gb|EF589349.1| Rhodopirellula baltica strain Gr16 16S ribosomal RNA gene, partial sequence
>gi|148357811|gb|EF589354.1| Rhodopirellula baltica strain Cc1 16S ribosomal RNA gene, partial sequence
>gi|171879796|gb|EU545801.1| Sulfolobus tokodaii strain TW 16S ribosomal RNA gene, partial sequence
>gi|223016623|gb|FJ624338.1| Rhodopirellula sp. 2SC 16S ribosomal RNA gene, complete sequence; and 16S-23S ribosomal RNA intergenic spacer, partial sequence >gi|223016633|gb|FJ624348.1| Rhodopirellula baltica strain SH159 IFAM 1455 16S ribosomal RNA gene, complete sequence; and 16S-23S ribosomal RNA intergenic spacer, partial sequence
>gi|223016626|gb|FJ624341.1| Rhodopirellula baltica strain SH28 IFAM 1430 16S ribosomal RNA gene, complete sequence; and 16S-23S ribosomal RNA intergenic spacer, partial sequence
>gi|223016627|gb|FJ624342.1| Rhodopirellula baltica strain SH386 IFAM 3184 16S ribosomal RNA gene, complete sequence; and 16S-23S ribosomal RNA intergenic spacer, partial sequence
>gi|223016628|gb|FJ624343.1| Rhodopirellula baltica strain SH398 IFAM 3246 16S ribosomal RNA gene, complete sequence; and 16S-23S ribosomal RNA intergenic spacer, partial sequence
>gi|223016629|gb|FJ624344.1| Rhodopirellula baltica strain SH121 IFAM 1433 16S ribosomal RNA gene, partial sequence
>gi|223016630|gb|FJ624345.1| Rhodopirellula baltica strain SH123 IFAM 1434 16S ribosomal RNA gene, complete sequence; and 16S-23S ribosomal RNA intergenic spacer, partial sequence
>gi|223016636|gb|FJ624351.1| Rhodopirellula baltica strain SH198 IFAM 1735 16S ribosomal RNA gene, complete sequence; and 16S-23S ribosomal RNA intergenic spacer, partial sequence
>gi|223016632|gb|FJ624347.1| Rhodopirellula baltica strain SH155 IFAM 1436 16S ribosomal RNA gene, complete sequence; and 16S-23S ribosomal RNA intergenic spacer, partial sequence
>gi|223016634|gb|FJ624349.1| Rhodopirellula baltica strain SH188 IFAM 1580 16S ribosomal RNA gene, complete sequence; and 16S-23S ribosomal RNA intergenic spacer, partial sequence
>gi|223016635|gb|FJ624350.1| Rhodopirellula baltica strain SH190 IFAM 1582 16S ribosomal RNA gene, complete sequence; and 16S-23S ribosomal RNA intergenic spacer, partial sequence

>gi|223016631|gb|FJ624346.1| Rhodopirellula baltica strain SH126 IFAM 1435 16S ribosomal RNA gene, complete sequence; and 16S-23S ribosomal RNA intergenic spacer, partial sequence
>gi|223016637|gb|FJ624352.1| Rhodopirellula baltica strain SH26 IFAM 1429 16S ribosomal RNA gene, complete sequence; and 16S-23S ribosomal RNA intergenic spacer, partial sequence
>gi|265678307|ref|NR_028609.1| Sulfolobus tokodaii strain 7 16S ribosomal RNA, partial sequence
>gi|288872013|dbj|AB546257.1| Methanococcus maripaludis gene for 16S rRNA, partial sequence, strain: Mic1c10
>gi|288872014|dbj|AB546258.1| Methanococcus maripaludis gene for 16S rRNA, partial sequence, strain: Mic4c08
>gi|291167473|dbj|AB264796.1| Methanococcus maripaludis gene for 16S ribosomal RNA, partial sequence
>gi|313760414|dbj|AB603516.1| Methanocaldococcus jannaschii gene for 16S rRNA, partial sequence, strain: JCM 10045
>gi|321442501|gb|HQ845416.1| Rhodopirellula baltica strain Cc2 16S ribosomal RNA gene, partial sequence
>gi|321442502|gb|HQ845417.1| Rhodopirellula baltica strain CcC11 16S ribosomal RNA gene, partial sequence
>gi|321442503|gb|HQ845418.1| Rhodopirellula baltica strain CcC12 16S ribosomal RNA gene, partial sequence
>gi|321442505|gb|HQ845420.1| Rhodopirellula baltica strain CcC3 16S ribosomal RNA gene, partial sequence
>gi|321442507|gb|HQ845422.1| Rhodopirellula baltica strain CcC5 16S ribosomal RNA gene, partial sequence
>gi|321442508|gb|HQ845423.1| Rhodopirellula baltica strain CcC7 16S ribosomal RNA gene, partial sequence
>gi|321442510|gb|HQ845425.1| Rhodopirellula baltica strain Cor1 16S ribosomal RNA gene, partial sequence
>gi|321442511|gb|HQ845426.1| Rhodopirellula baltica strain Cor2 16S ribosomal RNA gene, partial sequence
>gi|321442512|gb|HQ845427.1| Rhodopirellula baltica strain Cor4 16S ribosomal RNA gene, partial sequence
>gi|321442513|gb|HQ845428.1| Rhodopirellula baltica strain F1 16S ribosomal RNA gene, partial sequence
>gi|321442514|gb|HQ845429.1| Rhodopirellula baltica strain F2 16S ribosomal RNA gene, partial sequence
>gi|321442515|gb|HQ845430.1| Rhodopirellula baltica strain F4 16S ribosomal RNA gene, partial sequence
>gi|321442516|gb|HQ845431.1| Rhodopirellula baltica strain F5 16S ribosomal RNA gene, partial sequence
>gi|321442517|gb|HQ845432.1| Rhodopirellula baltica strain F6 16S ribosomal RNA gene, partial sequence
>gi|321442518|gb|HQ845433.1| Rhodopirellula baltica strain F7 16S ribosomal RNA gene, partial sequence
>gi|321442519|gb|HQ845434.1| Rhodopirellula baltica strain FC1 16S ribosomal RNA gene, partial sequence
>gi|321442520|gb|HQ845435.1| Rhodopirellula baltica strain FC2 16S ribosomal RNA gene, partial sequence
>gi|321442522|gb|HQ845437.1| Rhodopirellula baltica strain FC41 16S ribosomal RNA gene, partial sequence
>gi|321442523|gb|HQ845438.1| Rhodopirellula baltica strain FC42 16S ribosomal RNA gene, partial sequence
>gi|321442524|gb|HQ845439.1| Rhodopirellula baltica strain FC5 16S ribosomal RNA gene, partial sequence
>gi|321442525|gb|HQ845440.1| Rhodopirellula baltica strain FC6 16S ribosomal RNA gene, partial sequence
>gi|321442526|gb|HQ845441.1| Rhodopirellula baltica strain FC7 16S ribosomal RNA gene, partial sequence >gi|321442529|gb|HQ845444.1| Rhodopirellula baltica strain FC101 16S ribosomal RNA gene, partial sequence >gi|321442530|gb|HQ845445.1| Rhodopirellula baltica strain FC102 16S ribosomal RNA gene, partial sequence >gi|321442577|gb|HQ845492.1| Rhodopirellula baltica strain MsF1 16S ribosomal RNA gene, partial sequence
>gi|321442595|gb|HQ845510.1| Rhodopirellula baltica strain UC10 16S ribosomal RNA gene, partial sequence
>gi|321442526|gb|HQ845441.1| Rhodopirellula baltica strain FC7 16S ribosomal RNA gene, partial sequence >gi|321442529|gb|HQ845444.1| Rhodopirellula baltica strain FC101 16S ribosomal RNA gene, partial sequence >gi|321442530|gb|HQ845445.1| Rhodopirellula baltica strain FC102 16S ribosomal RNA gene, partial sequence >gi|321442577|gb|HQ845492.1| Rhodopirellula baltica strain MsF1 16S ribosomal RNA gene, partial sequence
>gi|321442595|gb|HQ845510.1| Rhodopirellula baltica strain UC10 16S ribosomal RNA gene, partial sequence
>gi|321442526|gb|HQ845441.1| Rhodopirellula baltica strain FC7 16S ribosomal RNA gene, partial sequence >gi|321442529|gb|HQ845444.1| Rhodopirellula baltica strain FC101 16S ribosomal RNA gene, partial sequence >gi|321442530|gb|HQ845445.1| Rhodopirellula baltica strain FC102 16S ribosomal RNA gene, partial sequence >gi|321442577|gb|HQ845492.1| Rhodopirellula baltica strain MsF1 16S ribosomal RNA gene, partial sequence
>gi|321442595|gb|HQ845510.1| Rhodopirellula baltica strain UC10 16S ribosomal RNA gene, partial sequence
>gi|321442526|gb|HQ845441.1| Rhodopirellula baltica strain FC7 16S ribosomal RNA gene, partial sequence >gi|321442529|gb|HQ845444.1| Rhodopirellula baltica strain FC101 16S ribosomal RNA gene, partial sequence >gi|321442530|gb|HQ845445.1| Rhodopirellula baltica strain FC102 16S ribosomal RNA gene, partial sequence

>gi|321442570|gb|HQ845485.1| Rhodopirellula baltica strain Gr17 16S ribosomal RNA gene, partial sequence
>gi|321442572|gb|HQ845487.1| Rhodopirellula baltica strain Gtu3 16S ribosomal RNA gene, partial sequence
>gi|321442573|gb|HQ845488.1| Rhodopirellula baltica strain M1 16S ribosomal RNA gene, partial sequence
>gi|321442574|gb|HQ845489.1| Rhodopirellula baltica strain M2 16S ribosomal RNA gene, partial sequence
>gi|321442575|gb|HQ845490.1| Rhodopirellula baltica strain M3 16S ribosomal RNA gene, partial sequence
>gi|321442576|gb|HQ845491.1| Rhodopirellula baltica strain M4 16S ribosomal RNA gene, partial sequence
>gi|321442578|gb|HQ845493.1| Rhodopirellula baltica strain MsF2 16S ribosomal RNA gene, partial sequence
>gi|321442579|gb|HQ845494.1| Rhodopirellula baltica strain MsF4 16S ribosomal RNA gene, partial sequence
>gi|321442581|gb|HQ845496.1| Rhodopirellula baltica strain MsF7 16S ribosomal RNA gene, partial sequence
>gi|321442583|gb|HQ845498.1| Rhodopirellula baltica strain Pd5 16S ribosomal RNA gene, partial sequence
>gi|321442584|gb|HQ845499.1| Rhodopirellula baltica strain Pd6 16S ribosomal RNA gene, partial sequence
>gi|321442587|gb|HQ845502.1| Rhodopirellula baltica strain Sm3 16S ribosomal RNA gene, partial sequence
>gi|321442589|gb|HQ845504.1| Rhodopirellula baltica strain U1 16S ribosomal RNA gene, partial sequence
>gi|321442590|gb|HQ845505.1| Rhodopirellula baltica strain U2 16S ribosomal RNA gene, partial sequence
>gi|321442591|gb|HQ845506.1| Rhodopirellula baltica strain UC1 16S ribosomal RNA gene, partial sequence
>gi|321442592|gb|HQ845507.1| Rhodopirellula baltica strain UC2 16S ribosomal RNA gene, partial sequence >gi|321442602|gb|HQ845517.1|
Rhodopirellula baltica strain UC21 16S ribosomal RNA gene, partial sequence
>gi|321442592|gb|HQ845507.1| Rhodopirellula baltica strain UC2 16S ribosomal RNA gene, partial sequence >gi|321442602|gb|HQ845517.1|
Rhodopirellula baltica strain UC21 16S ribosomal RNA gene, partial sequence
>gi|321442596|gb|HQ845511.1| Rhodopirellula baltica strain UC11 16S ribosomal RNA gene, partial sequence
>gi|321442597|gb|HQ845512.1| Rhodopirellula baltica strain UC12 16S ribosomal RNA gene, partial sequence
>gi|321442600|gb|HQ845515.1| Rhodopirellula baltica strain UC15 16S ribosomal RNA gene, partial sequence
>gi|321442604|gb|HQ845519.1| Rhodopirellula baltica strain UC23 16S ribosomal RNA gene, partial sequence
>gi|321442606|gb|HQ845521.1| Rhodopirellula baltica strain UC25 16S ribosomal RNA gene, partial sequence
>gi|321442612|gb|HQ845527.1| Rhodopirellula baltica strain UC45 16S ribosomal RNA gene, partial sequence
>gi|321442613|gb|HQ845528.1| Rhodopirellula baltica strain UC46 16S ribosomal RNA gene, partial sequence
>gi|321442614|gb|HQ845529.1| Rhodopirellula baltica strain UC47 16S ribosomal RNA gene, partial sequence
>gi|321442618|gb|HQ845533.1| Rhodopirellula baltica strain UF41 16S ribosomal RNA gene, partial sequence
>gi|321442622|gb|HQ845537.1| Rhodopirellula baltica strain UF82 16S ribosomal RNA gene, partial sequence
>gi|321442623|gb|HQ845538.1| Rhodopirellula baltica strain UF14 16S ribosomal RNA gene, partial sequence
>gi|326580967|gb|JF298894.1| Desulfovibrio piger strain EBA23-28 16S ribosomal RNA gene, partial sequence
>gi|343202891|ref|NR_043384.1| Rhodopirellula baltica SH 1 strain 1 16S ribosomal RNA, partial sequence
>gi|358365241|ref|NR_044615.2| Pyrobaculum aerophilum str. IM2 strain im2 16S ribosomal RNA, partial sequence
>gi|374174436|gb|JQ045861.1| Archaeoglobus fulgidus strain delta 16S ribosomal RNA gene, partial sequence
>gi|375311520|gb|JN694982.1| Rhodopirellula baltica strain 2ILRT 16S ribosomal RNA gene, partial sequence
>gi|375311521|gb|JN694983.1| Rhodopirellula baltica strain 4ILRT 16S ribosomal RNA gene, partial sequence
>gi|375311523|gb|JN694985.1| Rhodopirellula baltica strain 3ILRT 16S ribosomal RNA gene, partial sequence
>gi|375311524|gb|JN694986.1| Rhodopirellula baltica strain 5ILRT 16S ribosomal RNA gene, partial sequence
>gi|378405450|gb|JQ346753.1| Methanococcus maripaludis strain JJ 16S ribosomal RNA gene, partial sequence
>gi|378927119|gb|JQ405263.1| Methanococcus maripaludis strain IB1 16S ribosomal RNA gene, partial sequence
>gi|388556231|dbj|AB685272.1| Archaeoglobus fulgidus gene for 16S ribosomal RNA, partial sequence, strain: SA2
>gi|441425469|gb|KC139247.1| Methanococcus maripaludis strain NBRC 101831 16S ribosomal RNA gene, partial sequence
>gi|444303811|ref|NR_074233.1| Methanocaldococcus jannaschii DSM 2661 strain DSM 2661 16S ribosomal RNA, complete sequence
>gi|444303814|ref|NR_074236.1| Herpetosiphon aurantiacus DSM 785 16S ribosomal RNA, complete sequence
>gi|444303925|ref|NR_074348.1| Sulfolobus tokodaii str. 7 strain 7 16S ribosomal RNA, complete sequence

>gi|444303935|ref|NR_074358.1| *Pyrococcus horikoshii* OT3 strain OT3 16S ribosomal RNA, complete sequence
>gi|444303937|ref|NR_074360.1| *Pyrobaculum calidifontis* JCM 11548 strain JCM 11548 16S ribosomal RNA, complete sequence
>gi|444439379|ref|NR_074094.1| *Methanococcus maripaludis* S2 strain S2 16S ribosomal RNA, complete sequence
>gi|470466091|ref|NR_074104.1| *Ignicoccus hospitalis* KIN4/I strain KIN4/I 16S ribosomal RNA, complete sequence