

fastx-utils之将RNA序列文件转换成DNA序列： rna2dna

一、fastx-utils rna2dna介绍

功能描述：

`fastx-utils rna2dna` 将RNA序列文件转换成DNA序列。

命令行接口：

```
$ fastx-utils rna2dna

Usage: fastx-utils rna2dna <in.fa>
```

二、使用场景实例及其用法

使用场景经典案例：

1. `Silva` 数据库提供的数据文件为 `A,U,C,G` 需要转成 `A,T,C,G`

示例演示：

示例文件： `SILVA_138_SSURef.fasta`

```
$ cat SILVA_138_SSURef.fasta | head -n4
```

>HM112333.1.1374

Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Enterobacter;uncultured gamma proteobacterium

GAUUGAACGCUGGCGGCAGGCCUAACACAUGCAAGUCGAACGGUACAGGAAGAAGCUUGCUCUUCUUGCUGACGAGUGGC
GGACGGGUGAGUAAUGUCUGGGAAACUGCCUGAUGGAGGGGGUAACUACUGGAAACGGUAGCUAAUACCGCAUAACGUC
GCAAGACCAAAGAGGGGGACCUUAGGGCCUCUUGCCAUCGGAUGUGCCCAGAUUGGGAUUAGCUAGUAGGUGGGGUAAAGG
CUCACCUAGGCGACGAUCCCUAGCUGGUCUGAGAGGAUGACCAGCCACACUGGAACUGAGACACGGUCCAGACUCCUACG
GGAGGCAGCAGUGGGGAUAUUGCACA AUGGGCGCAAGCCUGAUGCAGCCAUGCCGCGUGUAUGAAGAAGCCUUCGGGU
UGUAAAGUACUUCAGCGGGGAGGAAGGUGUUGUGGUUAAUAACCACAGCAAUUGACGUUACCCGAGAGAAGACCCGG
CUAACUCCGUGCCAGCAGCCGCGGUAAUACGGAGGGUGCAAGCGUUAUACGGAAUACUGGGCGUAAAGCGCACGCAGGC
GGUCUGUCAAGUCGGAUGUGAAAUCCCGGGCUCAACCUGGGAACUGCAUUCGAAACUGGCAGGCUUGAGUCUCGUAGAG
GGGGGUAGAAUCCACGUGUAGCGGUGAAAUGCGUAGAGAUUCUGGAGGAUACUCGGUGGCGAAGGCGGCCCCUGGACGA
AGACUGACGCUCAGGUGCGGAAGCGUGGGGAGCAAAAGGAUUGAUUAGAUACCGGUAGUCCACGCUGUAAACGAUGUCUAAU
UGGAGGUUGUGCCCCUGAGGUGUGGCUUCCGGAGCUAACGCGUUAUAAUAGACCGCCUGGGGAGUACGGCCGCAAGGUUAA
AACUCAAAUGAAUUGACGGGGGCCCGCACAAAGCGGUGGAGCAUGUGGUUUAUUCGAUGCAACGCGAAGAACCUUACCG
GUCUUGACAUCACAGAACUUUCCAGAGAUGGAGGGGUGCCUUCGGGAACUGUGAGACAGGUGCUGCAUGGCUGUCGUCU
GCUCGUGUUGUGAAUUGUUGGUUAAAGUCCCGCAACGAGCGCAACCCUUAUCCUUGUUGCCAGCGGUUAGGCCGGGAAC
UCAAGGAGACUGCCAGUGAUAAACUGGAGGAAGGUGGGGAUGACGUCAAGUCAUAGGCCCUUACGACCAGGGCUACA
CACGUGCUACA AUGGCGCAUACAAAGAGAAGCGAACUCGCGAGAGCAAGCGGACCUCAUAAAGUGCGUCGUAGUCCGGAC
UGGAGUCUGCAACUCGACUCCAUGAAGUCGGAUUCGCUAGUAAUCGUGGAUCAGAAUGCCACGGUGAAUACGUUCCGGG
CCUUGUACACACCG

>KF089970.1.1357

Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas;uncultured bacterium

AUUGAACGCUGGCGGCAGGCCUAACACAUGCAAGUCGAGCGGAUGAAGGGAGCUUGCUCUGGAUUCAGCGGCGGACGGG
UGAGUAAUGCCUAGGAUUCUGCCUGGUAGUGGGGUAACGUCCGGAAACGGGCGCUAAUACCGCAUACGUCCUGAGGGA
GAAAGUGGGGAUCUUCGGACCUCACGCUAUCAGAUAGCCUAGGUCGGAUUGCUAGUUGGUGGGGUAAAGGCCUGCCA
AGGCGACGAUCCGUAACUGGUCUGAGAGGAUGAUACAGUCACACUGGAACUGAGACACGGUCCAGACUCCUACGGGAGGCA
GCAGUGGGGAUAUUGGACAAUGGGCGAAAGCCUGAUCCAGCCAUGCCGCGUGUGUGAAGAAGGUCUUCGGAUUGUAAAG
CACUUUAAGUUGGGAGGAAGGGCAGUAAGUUAUACCUUGCUGUUUUGACGUUACCAACAGAAUAAAGCACCGGCUAACUU
CGUGCCAGCAGCCGCGGUAAUACGAAGGGUGCAAGCGUUAUACGGAAUACUGGGCGUAAAGCGCGCUAGGUGGUUCAG
CAAGUUGGAUGUGAAAUCCCGGGCUCAACCUGGGAACUGCAUCCAAAACUACUGAGCUAGAGUACGGUAGAGGGUGGUG
GAAUUUCCUGUGUAGCGGUGAAUUGCGUAGAUUAGGAAGGAACACCAGUGGCGAAGGCGACCACCUGGACUGAUACUGA
CACUGAGGUGCGAAAGCGUGGGGAGCAAAAGGAUUGAUUACCCUGGUAGUCCACGCGUAAACGAUGUCGACUAGCCGU
UGGGAUCCUUGAGAUUUAGUGGCGCAGCUAACGCGAUAAUGCGACCGCCUGGGGAGUACGGCCGCAAGGUUAAAACUCA
AAUGAAUUGACGGGGGCCCGCACAAAGCGGUGGAGCAUGUGGUUUAUUCGAAGCAACGCGAAGAACCUUACCUUGGCCUUG
ACAUGCUGAGAACUUUCCAGAGAUGGAUUGGUGCCUUCGGGAACUCAGACACAGGUGCUGCAUGGCUGUCGUCAGCUCGU
GUCGUGAGAUUUGGUUAAAGUCCCGUAACGAGCGCAACCCUUGUCCUAGUUACCAGCACCUCGGGUGGGCACUCUAAAG
GAGACUGCCGGUGACAAACCGGAGGAAGGUGGGGAUGACGUCAAGUCAUAGGCCCUUACGGCCAGGGCUACACACGUG
CUACA AUGGUCGGUACAAAGGUUGCCAAGCCGCGAGGUGGAGCUAAUCCCAUAAAACCGAUUCGUAGUCCGGAUCGCAGU
CUGCAACUCGACUGCGUGAAGUCGGAUUCGCUAGUAAUCGUGAAUCAGAAUGUCACGGUGAAUACGUUCCCGGGCCU

运行命令:

```
$ fastx-utils rna2dna SILVA_138_SSURef.fasta |head -n4
```

```
>HM112333.1.1374
Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;
Enterobacter;uncultured gamma proteobacterium
GATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAACGGTAACAGGAAGAAGCTTGCTTCTTTGCTGACGAGTGGC
GGACGGGTGAGTAATGTCTGGGAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTC
GCAAGACCAAAGAGGGGGACCTTAGGGCCTCTTGCCATCGGATGTGCCAGATGGGATTAGCTAGTAGGTGGGGTAAAGG
CTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAAGTGAAGACACGGTCCAGACTCCTACG
GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGT
TGTAAGTACTTTTCAGCGGGGAGGAAGGTGTTGTGGTTAATAACCACAGCAATTGACGTTACCGCAGAAGAAGCACCGG
CTAACTCCGTGCCAGCAGCCGCGGTAAACGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGC
GGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAAGTGCATTGCAAACTGGCAGGCTTGAGTCTCGTAGAG
GGGGGTAGAATTCACGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACTCGGTGGCGAAGGCGGCCCTGGACGA
AGACTGACGCTCAGGTGCGGAAGCGTGGGGAGCAAACAGGATTAGATACCTGGTAGTCCACGCTGTAAACGATGTCTATT
TGGAGGTTGTGCCCTGAGGTGTGGCTTCCGGAGCTAACGCGTTAAATAGACCGCTGGGGAGTACGGCCGAAGGTTAA
AACTCAAATGAATTGACGGGGGGCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTG
GTCTTGACATCCACAGAACTTTCCAGAGATGGAGGGGTGCCTTCGGGAAGTGTGAGACAGGTGCTGCATGGCTGTCGTCA
GCTCGTGTGTGAAATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCCTTATCCTTTGTTGCCAGCGGTTAGGCCGGGAAC
TCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGCCCTTACGACCAGGGCTACA
CACGTGCTACAATGGCGCATACAAAGAGAAGCGAACTCGCGAGAGCAAGCGGACCTATAAAGTGCCTCGTAGTCCGGAC
TGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGTTCCCGGG
CCTTGACACACCG
>KF089970.1.1357
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pse
udomonas;uncultured bacterium
ATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAGCGGATGAAGGGAGCTTGCTCCTGGATTAGCGGCGGACGGG
TGAGTAATGCCTAGGAATCTGCCTGGTAGTGGGGGATAACGTCCGGAAACGGGCGCTAATACCGCATACGTCCTGAGGGA
GAAAGTGGGGGATCTTCGGACCTCACGCTATCAGATGAGCCTAGGTGCGATTAGCTAGTTGGTGGGGTAAAGGCCTGCCA
AGGCGACGATCCGTAAGTGGTCTGAGAGGATGATCAGTCACACTGGAAGTGAAGACACGGTCCAGACTCCTACGGGAGGCA
GCAGTGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTCGGATTGTAAAG
CACTTTAAGTTGGGAGGAAGGGCAGTAAGTTAATACCTTGCTGTTTTGACGTTACCAACAGAATAAGCACCGGCTAACTT
CGTGCCAGCAGCCGCGTAATACGAAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCTAGGTGGTTTCAG
CAAGTTGGATGTGAAATCCCCGGGCTCAACCTGGGAAGTGCATCCAAAAGTACTGAGCTAGAGTACGGTAGAGGGTGGTG
GAATTTCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAAGTGGCGAAGGCGACCACCTGGACTGATACTGA
CACTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCGTAAACGATGTCGACTAGCCGT
TGGGATCCTTGAGATCTTAGTGCGCAGCTAACGCGATAAGTCGACCGCCTGGGGAGTACGGCCGAAGGTTAAACTCA
AATGAATTGACGGGGGCCGCACAAGCGGTGGAGCATGTGGTTTAATTGCAAGCAACGCGAAGAACCTTACCTGGCCTTG
ACATGCTGAGAACTTTCCAGAGATGGATTGGTGCCTTCGGGAAGTGCAGACACAGGTGCTGCATGGCTGTCGTGAGCTCGT
GTCGTGAGATGTTGGGTAAAGTCCCGTAACGAGCGCAACCCCTTGCTTAGTTACCAGCACCTCGGGTGGGCACTCTAAG
GAGACTGCCGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGCCCTTACGGCCAGGGTACACACGTG
CTACAATGGTCCGTACAAAGGGTTGCCAAGCCGCGAGGTGGAGCTAATCCCATAAAACCGATCGTAGTCCGGATCGCAGT
CTGCAACTCGACTGCGTGAAGTCGGAATCGCTAGTAATCGTGAATCAGAATGTCACGGTGAATACGTTCCCGGGCCT
```

本文材料为 **BASE (Biostack Applied bioinformatic SEies)** 课程 **Linux Command Line Tools for Life Scientists** 材料，版权归 **上海逻捷信息科技有限公司** 所有。

Last Update: 2020-08-10 11:56 AM