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
GAUUGAACGCUGGCGGCAGGCCUAACACAUGCAAGUCGAACGGUAACAGGAAGAAGCUUGCUUCUUUGCUGACGAGU
GGCGGACGGGUGAGUAAUGUCUGGGAAACUGCCUGAUGGAGGGGGAUAACUACUGGAAACGGUAGCUAAUACCGCAU
AACGUCGCAAGACCAAAGAGGGGGACCUUAGGGCCUCUUGCCAUCGGAUGUGCCCAGAUGGGAUUAGCUAGUAGGUG
GGGUAAAGGCUCACCUAGGCGACGAUCCCUAGCUGGUCUGAGAGGAUGACCAGCCACACUGGAACUGAGACACGGUC
CAGACUCCUACGGGAGGCAGCAGUGGGGAAUAUUGCACAAUGGGCGCAAGCCUGAUGCAGCCAUGCCGCGUGUAUGA
AGAAGGCCUUCGGGUUGUAAAGUACUUUCAGCGGGGAGGAAGGUGUUGUGGUUAAUAACCACAGCAAUUGACGUUAC
CCGCAGAAGAAGCACCGGCUAACUCCGUGCCAGCAGCCGCGGUAAUACGGAGGGUGCAAGCGUUAAUCGGAAUUACU
GGGCGUAAAGCGCACGCAGGCGGUCUGUCAAGUCGGAUGUGAAAUCCCCGGGCUCAACCUGGGAACUGCAUUCGAAA
CUGGCAGGCUUGAGUCUCGUAGAGGGGGGUAGAAUUCCACGUGUAGCGGUGAAAUGCGUAGAGAUCUGGAGGAAUAC
UCGGUGGCGAAGGCGGCCCCUGGACGAAGACUGACGCUCAGGUGCGGAAGCGUGGGGAGCAAACAGGAUUAGAUACC
UGGUAGUCCACGCUGUAAACGAUGUCUAUUUGGAGGUUGUGCCCCUGAGGUGUGGCUUCCGGAGCUAACGCGUUAAA
UAGACCGCCUGGGGAGUACGGCCGCAAGGUUAAAACUCAAAUGAAUUGACGGGGGCCCGCACAAGCGGUGGAGCAUG
UGGUUUAAUUCGAUGCAACGCGAAGAACCUUACCUGGUCUUGACAUCCACAGAACUUUCCAGAGAUGGAGGGGUGCC
UUCGGGAACUGUGAGACAGGUGCUGCAUGGCUGUCGUCAGCUCGUGUUGUGAAAUGUUGGGUUAAGUCCCGCAACGA
GCGCAACCCUUAUCCUUUGUUGCCAGCGGUUAGGCCGGGAACUCAAAGGAGACUGCCAGUGAUAAACUGGAGGAAGG
UGGGGAUGACGUCAAGUCAUCAUGGCCCUUACGACCAGGGCUACACACGUGCUACAAUGGCGCAUACAAAGAGAAGC
GAACUCGCGAGAGCAAGCGGACCUCAUAAAGUGCGUCGUAGUCCGGACUGGAGUCUGCAACUCGACUCCAUGAAGUC
GGAAUCGCUAGUAAUCGUGGAUCAGAAUGCCACGGUGAAUACGUUCCCGGGCCUUGUACACACCG
>KF089970.1.1357
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas;uncultured bacterium
AUUGAACGCUGGCGGCAGGCCUAACACAUGCAAGUCGAGCGGAUGAAGGGAGCUUGCUCCUGGAUUCAGCGGCGGAC
GGGUGAGUAAUGCCUAGGAAUCUGCCUGGUAGUGGGGGAUAACGUCCGGAAACGGGCGCUAAUACCGCAUACGUCCU
GAGGGAGAAAGUGGGGGAUCUUCGGACCUCACGCUAUCAGAUGAGCCUAGGUCGGAUUAGCUAGUUGGUGGGGUAAA
GGCCUGCCAAGGCGACGAUCCGUAACUGGUCUGAGAGGAUGAUCAGUCACACUGGAACUGAGACACGGUCCAGACUC
CUACGGGAGGCAGCAGUGGGGAAUAUUGGACAAUGGGCGAAAGCCUGAUCCAGCCAUGCCGCGUGUGUGAAGAAGGU
CUUCGGAUUGUAAAGCACUUUAAGUUGGGAGGAAGGGCAGUAAGUUAAUACCUUGCUGUUUUGACGUUACCAACAGA
AUAAGCACCGGCUAACUUCGUGCCAGCAGCCGCGGUAAUACGAAGGGUGCAAGCGUUAAUCGGAAUUACUGGGCGUA
AAGCGCGCGUAGGUGGUUCAGCAAGUUGGAUGUGAAAUCCCCGGGCUCAACCUGGGAACUGCAUCCAAAACUACUGA
GCUAGAGUACGGUAGAGGGUGGUGGAAUUUCCUGUGUAGCGGUGAAAUGCGUAGAUAUAGGAAGGAACACCAGUGGC
GAAGGCGACCACCUGGACUGAUACUGACACUGAGGUGCGAAAGCGUGGGGAGCAAACAGGAUUAGAUACCCUGGUAG
UCCACGCCGUAAACGAUGUCGACUAGCCGUUGGGAUCCUUGAGAUCUUAGUGGCGCAGCUAACGCGAUAAGUCGACC
GCCUGGGGAGUACGGCCGCAAGGUUAAAACUCAAAUGAAUUGACGGGGGCCCGCACAAGCGGUGGAGCAUGUGGUUU
AAUUCGAAGCAACGCGAAGAACCUUACCUGGCCUUGACAUGCUGAGAACUUUCCAGAGAUGGAUUGGUGCCUUCGGG
AACUCAGACACAGGUGCUGCAUGGCUGUCGUCAGCUCGUGUCGUGAGAUGUUGGGUUAAGUCCCGUAACGAGCGCAA
CCCUUGUCCUUAGUUACCAGCACCUCGGGUGGGCACUCUAAGGAGACUGCCGGUGACAAACCGGAGGAAGGUGGGGA
UGACGUCAAGUCAUCAUGGCCCUUACGGCCAGGGCUACACACGUGCUACAAUGGUCGGUACAAAGGGUUGCCAAGCC
GCGAGGUGGAGCUAAUCCCAUAAAACCGAUCGUAGUCCGGAUCGCAGUCUGCAACUCGACUGCGUGAAGUCGGAAUC
GCUAGUAAUCGUGAAUCAGAAUGUCACGGUGAAUACGUUCCCGGGCCU
```

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
GATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAACGGTAACAGGAAGAAGCTTGCTTCTTTGCTGACGAGT
GGCGGACGGGTGAGTAATGTCTGGGAAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCAT
AACGTCGCAAGACCAAAGAGGGGGACCTTAGGGCCTCTTGCCATCGGATGTGCCCAGATGGGATTAGCTAGTAGGTG
GGGTAAAGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAACTGAGACACGGTC
CAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGA
AGAAGGCCTTCGGGTTGTAAAGTACTTTCAGCGGGGAGGAAGGTGTTGTGGTTAATAACCACAGCAATTGACGTTAC
CCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACT
GGGCGTAAAGCGCACGCAGGCGGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTCGAAA
CTGGCAGGCTTGAGTCTCGTAGAGGGGGGTAGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATAC
TCGGTGGCGAAGGCGGCCCCTGGACGAAGACTGACGCTCAGGTGCGGAAGCGTGGGGAGCAAACAGGATTAGATACC
TGGTAGTCCACGCTGTAAACGATGTCTATTTGGAGGTTGTGCCCCTGAGGTGTGGCTTCCGGAGCTAACGCGTTAAA
TAGACCGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATG
TGGTTTAATTCGATGCAACGCGAAGAACCTTACCTGGTCTTGACATCCACAGAACTTTCCAGAGATGGAGGGGTGCC
TTCGGGAACTGTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTCCCGCAACGA
GCGCAACCCTTATCCTTTGTTGCCAGCGGTTAGGCCGGGAACTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGG
TGGGGATGACGTCAAGTCATCATGGCCCTTACGACCAGGGCTACACACGTGCTACAATGGCGCATACAAAGAGAAGC
GAACTCGCGAGAGCAAGCGGACCTCATAAAGTGCGTCGTAGTCCGGACTGGAGTCTGCAACTCGACTCCATGAAGTC
GGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGTTCCCGGGCCTTGTACACACCG
>KF089970.1.1357
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas;uncultured bacterium
ATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAGCGGATGAAGGGAGCTTGCTCCTGGATTCAGCGGCGGAC
GGGTGAGTAATGCCTAGGAATCTGCCTGGTAGTGGGGGATAACGTCCGGAAACGGGCGCTAATACCGCATACGTCCT
GAGGGAGAAAGTGGGGGATCTTCGGACCTCACGCTATCAGATGAGCCTAGGTCGGATTAGCTAGTTGGTGGGGTAAA
GGCCTGCCAAGGCGACGATCCGTAACTGGTCTGAGAGGATGATCAGTCACACTGGAACTGAGACACGGTCCAGACTC
CTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGT
CTTCGGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCAGTAAGTTAATACCTTGCTGTTTTGACGTTACCAACAGA
ATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTA
AAGCGCGCGTAGGTGGTTCAGCAAGTTGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCCAAAACTACTGA
GCTAGAGTACGGTAGAGGGTGGTGGAATTTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGC
GAAGGCGACCACCTGGACTGATACTGACACTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAG
TCCACGCCGTAAACGATGTCGACTAGCCGTTGGGATCCTTGAGATCTTAGTGGCGCAGCTAACGCGATAAGTCGACC
GCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTT
AATTCGAAGCAACGCGAAGAACCTTACCTGGCCTTGACATGCTGAGAACTTTCCAGAGATGGATTGGTGCCTTCGGG
AACTCAGACACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGTAACGAGCGCAA
CCCTTGTCCTTAGTTACCAGCACCTCGGGTGGGCACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGA
TGACGTCAAGTCATCATGGCCCTTACGGCCAGGGCTACACACGTGCTACAATGGTCGGTACAAAGGGTTGCCAAGCC
GCGAGGTGGAGCTAATCCCATAAAACCGATCGTAGTCCGGATCGCAGTCTGCAACTCGACTGCGTGAAGTCGGAATC
GCTAGTAATCGTGAATCAGAATGTCACGGTGAATACGTTCCCGGGCCT
```

本文材料为 **BASE** (**B**iostack **A**pplied bioinformatic **SE**ies) 课程 **Linux Command Line Tools for Life Scientists** 材料，

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