

به نام خداوند بخشنده و مهربان



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

درس مقدماتی بر بیوانفورماتیک

تمرین برنامه‌نویسی چهارم

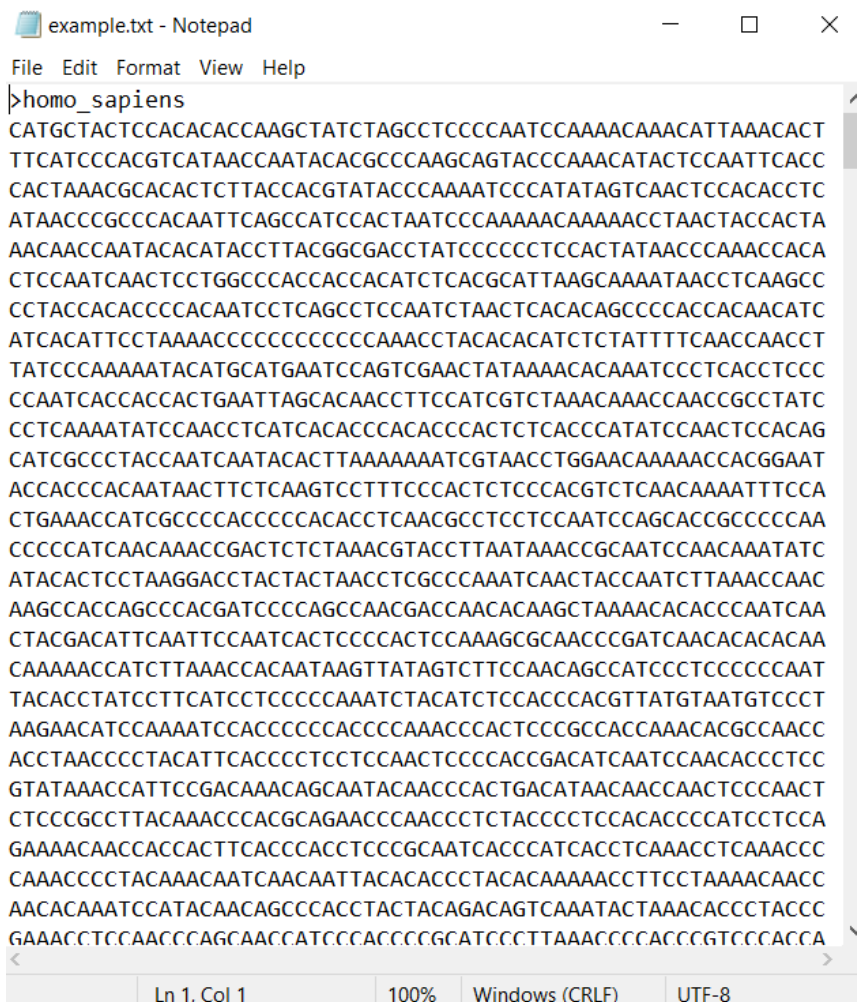
استاد درس: دکتر زینعلی

نام دانشجو:

روزبه قاسمی ۹۵۳۱۴۲۴

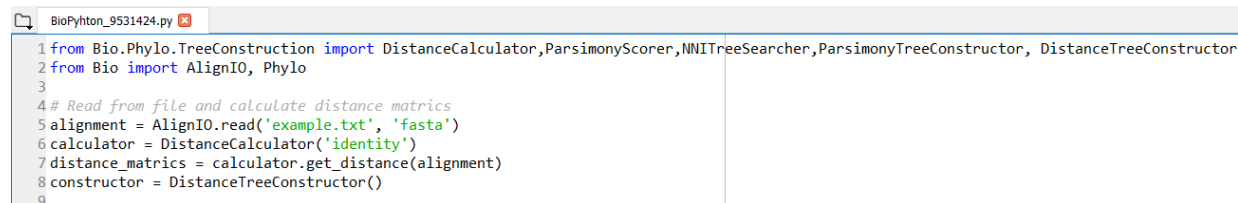
مرداد ۱۳۹۹

ابتدا فایل ورودی را به فرمت **fasta** تبدیل می‌کنیم. برای اینکار کافی بود، در فایل اصلی به جای “#” در ابتدای نام هر گونه، آن را به “>” تبدیل کنیم (به شکل زیر).



```
example.txt - Notepad
File Edit Format View Help
>homo_sapiens
CATGCTACTCCACACCAAGCTATCTAGCCTCCCAATCCAAAACAAACATTAACACT
TTCATCCCACGTCATAACCAATACACGCCAAGCAGTACCCAAACATACTCCAATTCACC
CACTAAACGCACACTCTTACCACGTATACCCAAATCCCATATAGTCAACTCCACACCTC
ATAACCCGCCCACAATTACGGCATCCACTAATCCAAAAACAAAAACCTAACTACCACTA
AACAAACATACACATACCTTACGGCGACCTATCCCCCTCCACTATAACCCAAACACA
CTCCAATCAACTCCTGGCCACCACCACATCTACGCATTAAGCAAAATAACCTCAAGCC
CCTACCACACCCACAATCCTCAGCCTCCAATCTAACTCACACAGCCCACCACAACATC
ATCACATTCTAAAACCCCCCCCCCAACCTACACACATCTCTATTTCAACCAACCT
TATCCAAAAATACATGCATGAATCCAGTCGAATATAAAACACAAATCCCTCACCTCCC
CCAATCACCACCACTGAATTAGCACAACTTCCATCGTCTAAACAAACCAACCGCTATC
CCTCAAAATATCCAACCTCATCACCCACACCCACTCTCACCCATATCCAATCCACAG
CATCGCCCTACCAATCAATACACTTAAAAAATCGTAACCTGGAACAAAAACCACGGAAT
ACCACCCACAATAACTTCTCAAGTCCTTCCCACTCTCCACGTCTCAACAAAATTTCCA
CTGAAACCATCGCCCCACCCCAACCTCAACGCTCCTCCAATCCAGCACCGCCCCCAA
CCCCATCAACAAACCGACTCTCTAAACGTACCTTAATAAACCGCAATCCAACAAATATC
ATACACTCCTAAGGACCTACTACTAACCTCGCCCAATCAACTACCAATCTTAAACCAAC
AAGCCACCAGCCACGATCCCCAGCCAACGACCAACACAAGCTAAACACACCCAATCAA
CTACGACATTCAATTCCAATCACTCCCCACTCCAAAGCGCAACCCGATCAACACACAAA
CAAAACCATCTTAAACCACAATAAGTTATAGTCTTCCAACAGCCATCCCTCCCCCAAT
TACACCTATCCTTCATCTCCCCCAATCTACATCTCCACCCACGTTATGTAATGTCCCT
AAGAACATCCAAATCCACCCCCACCCCAACCCACTCCCGCCACCAAAACAGCCAACC
ACCTAACCCCTACATTACCCCTCCTCCAATCCCCACCGACATCAATCCAACCCCTCC
GTATAAACCATTCGACAAACAGCAATACAACCCACTGACATAACAACCAACTCCCAACT
CTCCCGCCTTACAAACCCACGCAACCCAACCTCTACCCCTCCACACCCCATCTCCA
GAAAACAACCACCACTTCACCACCTCCCGCAATCACCCATCACCTCAAACCTCAAACCC
CAAACCCCTACAAACAATCAACAATTACACACCTACACAAAAACCTTCCTAAAAACAAC
AACACAAATCCATACAACAGCCACCTACTACAGACAGTCAAATACTAAACACCCTACCC
GAAACCTCCAACCCAGCAACCATCCACCCCGCATCCCTTAAACCCACCCGTCCACCA
```

سپس کد مورد نیاز برای این تمرین را می‌نویسیم.



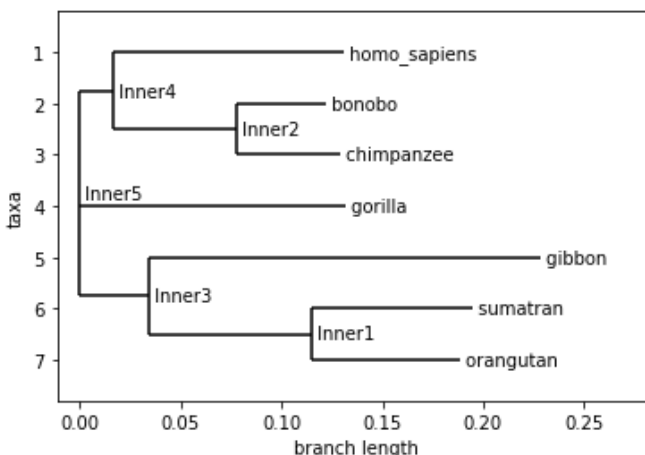
```
BioPhyton_9531424.py
1 from Bio.Phylo.TreeConstruction import DistanceCalculator, ParsimonyScorer, NNITreeSearcher, ParsimonyTreeConstructor, DistanceTreeConstructor
2 from Bio import AlignIO, Phylo
3
4 # Read from file and calculate distance matrices
5 alignment = AlignIO.read('example.txt', 'fasta')
6 calculator = DistanceCalculator('identity')
7 distance_matrices = calculator.get_distance(alignment)
8 constructor = DistanceTreeConstructor()
9
```

حال قسمت درخت UPGMA و درخت NJ را بدست می‌آوریم.

```
10 nj_tree = constructor.nj(distance_matrices)
11 upgma_tree = constructor.upgma(distance_matrices)
12
13 print(nj_tree)
14 Phylo.draw(nj_tree)
15
16 print(upgma_tree)
17 Phylo.draw(upgma_tree)
```

خروجی‌های بدست آمده به شکل زیر است:

```
Tree(rooted=False)
  Clade(branch_length=0, name='Inner5')
    Clade(branch_length=0.016230111077754433, name='Inner4')
      Clade(branch_length=0.11447388171720202, name='homo_sapiens')
      Clade(branch_length=0.06099894926448518, name='Inner2')
        Clade(branch_length=0.04518162713899726, name='bonobo')
        Clade(branch_length=0.05178625037526269, name='chimpanzee')
    Clade(branch_length=0.1321487541278895, name='gorilla')
    Clade(branch_length=0.03431777244070851, name='Inner3')
      Clade(branch_length=0.19416091263884716, name='gibbon')
      Clade(branch_length=0.08083158210747521, name='Inner1')
        Clade(branch_length=0.07934554187931553, name='sumatran')
        Clade(branch_length=0.07376163314320028, name='orangutan')
```

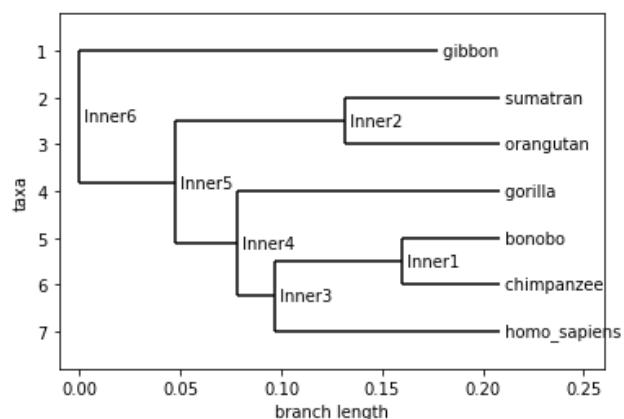


درخت بالا، خروجی درخت NJ می‌باشد.

```

Tree(rooted=True)
  Clade(branch_length=0, name='Inner6')
    Clade(branch_length=0.17745234163914742, name='gibbon')
    Clade(branch_length=0.04727371660162116, name='Inner5')
      Clade(branch_length=0.08458420894626237, name='Inner2')
        Clade(branch_length=0.0765535875112579, name='sumatran')
        Clade(branch_length=0.0765535875112579, name='orangutan')
      Clade(branch_length=0.03095917141999402, name='Inner4')
        Clade(branch_length=0.13017862503752625, name='gorilla')
        Clade(branch_length=0.018200240168117665, name='Inner3')
          Clade(branch_length=0.06349444611227861, name='Inner1')
            Clade(branch_length=0.048483938757129974, name='bonobo')
            Clade(branch_length=0.048483938757129974, name='chimpanzee')
          Clade(branch_length=0.11197838486940859, name='homo_sapiens')

```



درخت بالا یک درخت UPGMA می‌باشد. گونه Gibbon را مانند بقیه در یک ستون نشان نمی‌دهند که حقیقتاً نمی‌دانم مشکل از کجاست)

در قسمت آخر نیز می‌خواهیم درخت Parsimony را بدست بیاوریم.

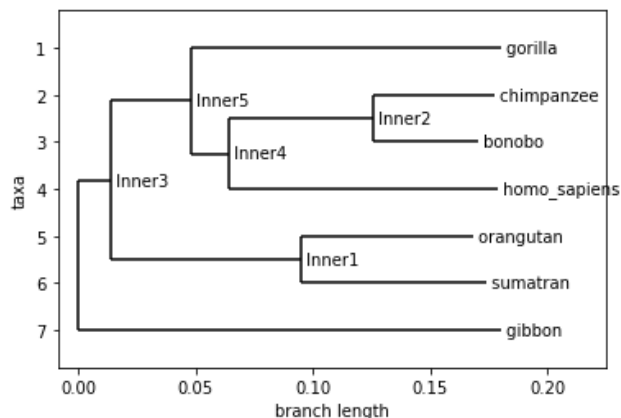
```

20 score_pars_tree = ParsimonyScorer()
21 searcher_pars_tree = NNITreeSearcher(score_pars_tree)
22 constructor_pars_tree = ParsimonyTreeConstructor(searcher_pars_tree, nj_tree)
23 pars_tree = constructor_pars_tree.build_tree(alignment)
24
25 print(pars_tree)
26 Phylo.draw(pars_tree)
27

```

خروجی این درخت برابر است با:

```
Tree(rooted=True)
  Clade(branch_length=0)
    Clade(branch_length=0.01384719303512455, name='Inner3')
      Clade(branch_length=0.03431777244070851, name='Inner5')
        Clade(branch_length=0.1321487541278895, name='gorilla')
        Clade(branch_length=0.016230111077754433, name='Inner4')
          Clade(branch_length=0.06099894926448518, name='Inner2')
            Clade(branch_length=0.05178625037526269, name='chimpanzee')
            Clade(branch_length=0.04518162713899726, name='bonobo')
          Clade(branch_length=0.11447388171720202, name='homo_sapiens')
      Clade(branch_length=0.08083158210747521, name='Inner1')
        Clade(branch_length=0.07376163314320028, name='orangutan')
        Clade(branch_length=0.07934554187931553, name='sumatran')
    Clade(branch_length=0.1803137196037226, name='gibbon')
```



روشی که من استفاده کردم، شاید منحصر به فرد باشد زیرا با کمی تغییر در فایل ورودی به یک فرمت استاندارد تبدیل شد. کد مورد نیاز این تمرین بسیار کم و ساده بود و من این آموزش را از روی وبسایت خود [BioPython](https://biopython.org/) دیدم.

درخت‌های بدست‌آمده مطابق با مطالبی که خواندیم بود. مثلاً باید درخت‌های UPGMA همگی در یک سطح باشند ولی درخت NJ بدین شکل نیست. مثلاً ورودی درخت Parsimony یک Starting Tree است که بر اساس Newick است.

پایان