Interpretation for COMP90016 Assignment3

In this assignment, we developed some Python programs to analyze GenBank type file. The programs can extract the tRNA info form the input GenBank files, and output the results to stdout with FASTA file format.

In the stage1, the program read from local GenBank files. In the stage 2/3, the program read directly from GenBank. In the stage3, besides the FASTA format stdout output, a data file assign3-stage3.out is also produced to store some further information about the retrieved tRNA info and input file.

To run the program, we can input the command as:

|  |
| --- |
| *[weih@mundula] assignment3 [1:58] python assign3 stage2.py NC\_003416.gbk* |

The output is:

|  |
| --- |
| *>NC\_003416, tRNA 1, tRNA-Pro*  *CAACATGTAGTTTAATAAAAATTTAATATTTGGGTTATTAAGATAAGTTGTTGA*  *>NC\_003416, tRNA 2, tRNA-Val*  *AAACTTTTAGTTTCATTAAGAATGTTTCATTTACAATGAGGGGGATAAAAGTTTT*  *>NC\_003416, tRNA 3, tRNA-Trp*  *ATAGATTTAAGTTAAGTTAAACTATTAATTTTCAAAATTAAAAATGGATGTCTATA*  *>NC\_003416, tRNA 4, tRNA-Glu*  *GAGACTTTAGTATATAATAGTACGGTTTATTTTCAATAAATTGGATGTAAGTCTTG*  *>NC\_003416, tRNA 5, tRNA-Ser*  *AAACTATTTTTTTTGTTTTGAAAACAAATATTAAGATTTTTTCTTGTAGTTT*  *>NC\_003416, tRNA 6, tRNA-Asn*  *TAAGAGTTAGTTTAATAAGAATTGTTGACTGTTAATCAAAAGGTTTACTCTTAA*  *>NC\_003416, tRNA 7, tRNA-Tyr*  *AAGAATTTAGTTTATTTAAAATGTTAAATTGTAAATTTAAAGTAGTTAATTCTTG*  *>NC\_003416, tRNA 8, tRNA-Lys*  *GAGATGTTAACTTAAGTTTAAAGTGTCAAATTTTTAATTTGAAAATGTATTTACACATCTTA*  *>NC\_003416, tRNA 9, tRNA-Leu*  *GTTAATATAGCATAAGAAGTGCATTTGTTTTAAGCGCAAAAGATAAAAATTAGCT*  *>NC\_003416, tRNA 10, tRNA-Ser*  *AATGAGTTTTATACAAGTCTTCTAAATTTGTTTTAGGTTTTACCTGCTCATTT*  *>NC\_003416, tRNA 11, tRNA-Ile*  *ATTTCATTAGTAAAATTTATTATGTTATCTTGATAAGGTAAAGTTCTTTTGATGAAGTA*  *>NC\_003416, tRNA 12, tRNA-Arg*  *AAATGTTAAATAGATTTACTATGTTTGATTACGGTTCAAAATGATTAACATTTT*  *>NC\_003416, tRNA 13, tRNA-Gln*  *TGTAATTTAGTTTAGATAGAATTTTTATTTTTGGTGTAAAAGGGTTTAATTACAA*  *>NC\_003416, tRNA 14, tRNA-Phe*  *ATCTTGTTAGTTTATTAGTAAAATATGACTCTGAAGAAGTTAAGAATTATAAGATA*  *>NC\_003416, tRNA 15, tRNA-Leu*  *ACATAATTAGTATATAAAATATATTAGATTTAGGTTCTAAAGATTTATTTATGTT*  *>NC\_003416, tRNA 16, tRNA-Thr*  *GCTAATTTAGTTTATGTAGAATTTGTAACTTGTAATTATAAGGTTAAATTAGTA*  *>NC\_003416, tRNA 17, tRNA-Cys*  *AAATCTTTAGTATAATTTAGTATTTTTAATTGCAAATTAAAGGGTTAAGGATTTT*  *>NC\_003416, tRNA 18, tRNA-Met*  *AATAAGATAGGATAATAAAGTCTGTTAGGTTCATACCCTAAGGGTGTTTTCTCTTATTG*  *>NC\_003416, tRNA 19, tRNA-Asp*  *AAAGTTTTAGTATATATAAGTATGGTTTATTGTCAATAAGCAGGTTTAAAACTTTA*  *>NC\_003416, tRNA 20, tRNA-Gly*  *GATTCTTTAGTATAATTAGTATGTTTGATTTCCAATCAAGTGGTTTAGGAATTA*  *>NC\_003416, tRNA 21, tRNA-His*  *AGCTTTGTAGTTTATTTTAAAATTTTTGTTTGTGGTACAAAAGAAATATTTAGCTA*  *>NC\_003416, tRNA 22, tRNA-Ala*  *GGGACTTTAATTTAAGTTAAAATATATGATTTGCATTCATAAAATTTAAGTTCTA* |

To further analyze the output results, we can pipe this assignment programs’ above output to the assignment1’s stage2.py program to get some statistical information such as Entropy of relative tRNA. The input command could be:

|  |
| --- |
| *[weih@mundula] assignment3 [1:62] python assign3-stage2.py NC\_003416.gbk |python stage2.py* |

The output is:

|  |
| --- |
| >NC\_003416, tRNA 1, tRNA-Pro  length: 54  G+C: 20%  entropy: 1.7  >NC\_003416, tRNA 2, tRNA-Val  length: 55  G+C: 25%  entropy: 1.8  >NC\_003416, tRNA 3, tRNA-Trp  length: 56  G+C: 16%  entropy: 1.6  >NC\_003416, tRNA 4, tRNA-Glu  length: 56  G+C: 26%  entropy: 1.8  >NC\_003416, tRNA 5, tRNA-Ser  length: 52  G+C: 15%  entropy: 1.6  >NC\_003416, tRNA 6, tRNA-Asn  length: 54  G+C: 24%  entropy: 1.8  >NC\_003416, tRNA 7, tRNA-Tyr  length: 55  G+C: 14%  entropy: 1.5  >NC\_003416, tRNA 8, tRNA-Lys  length: 62  G+C: 20%  entropy: 1.7  >NC\_003416, tRNA 9, tRNA-Leu  length: 55  G+C: 27%  entropy: 1.8  >NC\_003416, tRNA 10, tRNA-Ser  length: 53  G+C: 26%  entropy: 1.8  >NC\_003416, tRNA 11, tRNA-Ile  length: 59  G+C: 20%  entropy: 1.7  >NC\_003416, tRNA 12, tRNA-Arg  length: 54  G+C: 20%  entropy: 1.7  >NC\_003416, tRNA 13, tRNA-Gln  length: 55  G+C: 20%  entropy: 1.6  >NC\_003416, tRNA 14, tRNA-Phe  length: 56  G+C: 21%  entropy: 1.7  >NC\_003416, tRNA 15, tRNA-Leu  length: 55  G+C: 14%  entropy: 1.6  >NC\_003416, tRNA 16, tRNA-Thr  length: 54  G+C: 20%  entropy: 1.7  >NC\_003416, tRNA 17, tRNA-Cys  length: 55  G+C: 18%  entropy: 1.6  >NC\_003416, tRNA 18, tRNA-Met  length: 59  G+C: 32%  entropy: 1.9  >NC\_003416, tRNA 19, tRNA-Asp  length: 56  G+C: 21%  entropy: 1.7  >NC\_003416, tRNA 20, tRNA-Gly  length: 54  G+C: 25%  entropy: 1.8  >NC\_003416, tRNA 21, tRNA-His  length: 56  G+C: 21%  entropy: 1.7  >NC\_003416, tRNA 22, tRNA-Ala  length: 55  G+C: 20%  entropy: 1.7 |

From above output, we can see that there are 22 tRNA piece in the *NC\_003416.gbk* file. We also can see every tRNA’s amino acid specify, the length of the tRNA piece, the percentage of the G+C amino acid in the relative DNA piece, and the Entropy of relative tRNA.

In the satge3, besides above information, we also output some info into the *assignment3-stage3.out* data file. For the *NC\_003416.gbk* file, the *assignment3-stage3.out* file content is:

|  |
| --- |
| tRNA 1, tRNA-Pro, 0 -> 54, forward  tRNA 2, tRNA-Val, 57 -> 112, forward  tRNA 3, tRNA-Trp, 799 -> 855, forward  tRNA 4, tRNA-Glu, 866 -> 922, forward  tRNA 5, tRNA-Ser, 1621 -> 1673, forward  tRNA 6, tRNA-Asn, 1673 -> 1727, forward  tRNA 7, tRNA-Tyr, 1749 -> 1804, forward  tRNA 8, tRNA-Lys, 3302 -> 3364, forward  tRNA 9, tRNA-Leu, 3370 -> 3425, forward  tRNA 10, tRNA-Ser, 3425 -> 3478, forward  tRNA 11, tRNA-Ile, 4323 -> 4382, forward  tRNA 12, tRNA-Arg, 4385 -> 4439, forward  tRNA 13, tRNA-Gln, 4442 -> 4497, forward  tRNA 14, tRNA-Phe, 4509 -> 4565, forward  tRNA 15, tRNA-Leu, 5680 -> 5735, forward  tRNA 16, tRNA-Thr, 6501 -> 6555, forward  tRNA 17, tRNA-Cys, 9442 -> 9497, forward  tRNA 18, tRNA-Met, 9502 -> 9561, forward  tRNA 19, tRNA-Asp, 9564 -> 9620, forward  tRNA 20, tRNA-Gly, 9628 -> 9682, forward  tRNA 21, tRNA-His, 10378 -> 10434, forward  tRNA 22, tRNA-Ala, 13377 -> 13432, forward  Organism: Necator americanus  Authors: Hu,M., Chilton,N.B. and Gasser,R.B.  Title: The mitochondrial genomes of the human hookworms, Ancylostoma duodenale and Necator americanus (Nematoda: Secernentea)  Journal: Int. J. Parasitol. 32 (2), 145-158 (2002)  Authors: Hu,M.  Title: Direct Submission  Journal: Submitted (12-SEP-2003) Department of Veterinary Science, The University of Melbourne, 250 Princes Highway, Werribee, Victoria 3030, Australia  Authors: Hu,M.  Title: Direct Submission  Journal: Submitted (26-OCT-2001) Department of Veterinary Science, The University of Melbourne, 250 Princes Highway, Werribee, Victoria 3030, Australia  Authors:  Title: Direct Submission  Journal: Submitted (08-SEP-1999) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA |

Form above result, we can see every tRNA piece’s location in the original input file. We also could see tRNA’s strand location info (forward or reverse), and the info such as data source’s authors, original journal, etc.