1. **cp ../../bio425/data/mystery\_seq1.fas mystery\_seq1.fas**

mystery\_seq1.fas :

***>mystery\_seq\_1***

***TTTAAAACTTTTCTATTGGATAGATTTTATACAAAGAAGGTAATAATGTATAAACAACAA***

***TATTTTATTTCTGGCAAGGTGCAAGGTGTTGGTTTTAGATTTTTCACAGAGCAAATAGCA***

***AATAATATGAAACTAAAAGGATTTGTAAAAAATCTCAACGATGGAAGGGTAGAAATTGTA***

***GCTTTCTTTAATACTAAAGAACAAATGAAAAAATTTGAAAAATTATTAAATGGGAATAAG***

***TATTCAAACATTAAAAACATTGAAAAAATAGTTTTAGATGAAAATTATCCTTTTCAATTT***

***AATGATTTTAAAATTTATTATTAGGGCTTGCCTCTCGTTTAACAAGTACCTTAACCTTAT***

***TTTTTGGTTTAATATTGTGCACATAAGAATTGTTATTCTTATAGCAAGACACACTAGTTC***

***TAAAAAAATGTTCGACTTTAAATTTCAAAAACTCTAAAGACTTTCTGTTTCTACAAAAAA***

***TATTCAAATTGCCATCAGAGAATTTAATATCAAGCCCACAATTACAACTTGTTTTTAAAA***

***AACTGCACGAAATAAAACTACCATTTGAACTATTCCTATTAAAAAGAATAAAATATTTTT***

***TACCTCTTTTGGAGTCTTTAAAATAAATTCTAAGCCATTCTCCTGCTGGCTTAACTGAAG***

***CCAATCTATCATAAAGATAAGAAATGTAAATCGTTCTATTGTCATTAGATTTTAAAATTT***

***TTTTAAAAAAATAATGAGGACCTATTTTCATACAAATTATATAATATCTTATTAATAAAA***

***TATTTCCTAATATTTCCCAAATATATTGATAATGCCTGAATTTAAAAAAACAAAAACTAT***

1. **vi .bashrc**
2. **export PATH=$PATH:/data/biocs/b/bio425/bp-utils**
3. **sourc .bashrc**
4. **../../bio425/bin/long-orfs ../../bio425/data/mystery\_seq1.fas mystery\_seq1.coord**

***Sequence file = ../../bio425/data/mystery\_seq1.fas***

***Excluded regions file = none***

***Circular genome = true***

***Initial minimum gene length = 90 bp***

***Determine optimal min gene length to maximize number of genes***

***Maximum overlap bases = 30***

***Start codons = atg,gtg,ttg***

***Stop codons = taa,tag,tga***

***Sequence length = 840***

***Final minimum gene length = 232***

***Putative Genes:***

***00001 16 324 +1 0.929***

***00002 751 308 -2 0.911***

1. **grep "^[0-9]" mystery\_seq1.coord > mystery\_seq1.coord2**

***00001 16 324 +1 0.929***

***00002 751 308 -2 0.911***

1. **../../bio425/bin/extract ../../bio425/data/mystery\_seq1.fas mystery\_seq1.coord2 > mys.nuc**

***>00001 16 324 len=309***

***TTGGATAGATTTTATACAAAGAAGGTAATAATGTATAAACAACAATATTTTATTTCTGGC***

***AAGGTGCAAGGTGTTGGTTTTAGATTTTTCACAGAGCAAATAGCAAATAATATGAAACTA***

***AAAGGATTTGTAAAAAATCTCAACGATGGAAGGGTAGAAATTGTAGCTTTCTTTAATACT***

***AAAGAACAAATGAAAAAATTTGAAAAATTATTAAATGGGAATAAGTATTCAAACATTAAA***

***AACATTGAAAAAATAGTTTTAGATGAAAATTATCCTTTTCAATTTAATGATTTTAAAATT***

***TATTATTAG***

***>00002 751 308 len=398***

***TACAAATTATATAATATCTTATTAATAAAATATTTCCTAATATTTCCCAAATATATTGAT***

***AATGCCTGAATTTAAAAAAACAAAAACTATTTTAAAACTTTTCTATTGGATAGATTTTAT***

***ACAAAGAAGGTAATAATGTATAAACAACAATATTTTATTTCTGGCAAGGTGCAAGGTGTT***

***GGTTTTAGATTTTTCACAGAGCAAATAGCAAATAATATGAAACTAAAAGGATTTGTAAAA***

***AATCTCAACGATGGAAGGGTAGAAATTGTAGCTTTCTTTAATACTAAAGAACAAATGAAA***

***AAATTTGAAAAATTATTAAATGGGAATAAGTATTCAAACATTAAAAACATTGAAAAAATA***

***GTTTTAGATGAAAATTATCCTTTTCAATTTAATGATTT***

1. **bioseq -t1 mys.nuc > mys.pep**

***>00001 16 324 len=309***

***LDRFYTKKVIMYKQQYFISGKVQGVGFRFFTEQIANNMKLKGFVKNLNDGRVEIVAFFNT***

***KEQMKKFEKLLNGNKYSNIKNIEKIVLDENYPFQFNDFKIYY\****

***>00002 751 308 len=398***

***YKLYNILLIKYFLIFPKYIDNA\*I\*KNKNYFKTFLLDRFYTKKVIMYKQQYFISGKVQGV***

***GFRFFTEQIANNMKLKGFVKNLNDGRVEIVAFFNTKEQMKKFEKLLNGNKYSNIKNIEKI***

***VLDENYPFQFND***

1. **bioseq ../../bio425/data/mystery\_seq1.fas mystery\_seq1.coord2 > mys.nuc2**

***>mystery\_seq\_1***

***TTTAAAACTTTTCTATTGGATAGATTTTATACAAAGAAGGTAATAATGTATAAACAACAA***

***TATTTTATTTCTGGCAAGGTGCAAGGTGTTGGTTTTAGATTTTTCACAGAGCAAATAGCA***

***AATAATATGAAACTAAAAGGATTTGTAAAAAATCTCAACGATGGAAGGGTAGAAATTGTA***

***GCTTTCTTTAATACTAAAGAACAAATGAAAAAATTTGAAAAATTATTAAATGGGAATAAG***

***TATTCAAACATTAAAAACATTGAAAAAATAGTTTTAGATGAAAATTATCCTTTTCAATTT***

***AATGATTTTAAAATTTATTATTAGGGCTTGCCTCTCGTTTAACAAGTACCTTAACCTTAT***

***TTTTTGGTTTAATATTGTGCACATAAGAATTGTTATTCTTATAGCAAGACACACTAGTTC***

***TAAAAAAATGTTCGACTTTAAATTTCAAAAACTCTAAAGACTTTCTGTTTCTACAAAAAA***

***TATTCAAATTGCCATCAGAGAATTTAATATCAAGCCCACAATTACAACTTGTTTTTAAAA***

***AACTGCACGAAATAAAACTACCATTTGAACTATTCCTATTAAAAAGAATAAAATATTTTT***

***TACCTCTTTTGGAGTCTTTAAAATAAATTCTAAGCCATTCTCCTGCTGGCTTAACTGAAG***

***CCAATCTATCATAAAGATAAGAAATGTAAATCGTTCTATTGTCATTAGATTTTAAAATTT***

***TTTTAAAAAAATAATGAGGACCTATTTTCATACAAATTATATAATATCTTATTAATAAAA***

***TATTTCCTAATATTTCCCAAATATATTGATAATGCCTGAATTTAAAAAAACAAAAACTAT***

1. **bioseq -t1 mys.nuc2 > mys.pep2**

***>mystery\_seq\_1***

***FKTFLLDRFYTKKVIMYKQQYFISGKVQGVGFRFFTEQIANNMKLKGFVKNLNDGRVEIV***

***AFFNTKEQMKKFEKLLNGNKYSNIKNIEKIVLDENYPFQFNDFKIYY\*GLPLV\*QVP\*PY***

***FLV\*YCAHKNCYSYSKTH\*F\*KNVRL\*ISKTLKTFCFYKKYSNCHQRI\*YQAHNYNLFLK***

***NCTK\*NYHLNYSY\*KE\*NIFYLFWSL\*NKF\*AILLLA\*LKPIYHKDKKCKSFYCH\*ILKF***

***F\*KNNEDLFSYKLYNILLIKYFLIFPKYIDNA\*I\*KNKNY***