

SeqIO

June 2, 2020

1 SeqIO Biopython

In order to use the SeqIO module, first we need to import it:

```
[2]: from Bio import SeqIO
```

1.0.1 Excellent material on SeqIO:

<https://biopython.org/DIST/docs/api/Bio.SeqRecord.SeqRecord-class.html>

The SeqIO object is a generator:

```
[75]: myseq = SeqIO.parse("Cebus_albifrons_NC_002763.1.gb", "genbank")
      print(myseq)
```

```
<generator object parse at 0x7fd7441b24f8>
```

So, in order to use it, we need to iterate over its values. Doing so in a for loop will return lots of info on the genbank file

```
[172]: for i in SeqIO.parse("Cebus_albifrons_NC_002763.1.gb", "genbank"):
      print(i)
```

```
ID: NC_002763.1
Name: NC_002763
Description: Cebus albifrons mitochondrion, complete genome
Database cross-references: Project:11945, BioProject:PRJNA11945
Number of features: 55
/molecule_type=DNA
/topology=circular
/data_file_division=PRI
/date=01-FEB-2010
/accessions=['NC_002763']
/sequence_version=1
/keywords=['RefSeq']
/source=mitochondrion Cebus albifrons (white-fronted capuchin)
/organism=Cebus albifrons
/taxonomy=['Eukaryota', 'Metazoa', 'Chordata', 'Craniata', 'Vertebrata',
'Euteleostomi', 'Mammalia', 'Eutheria', 'Euarchontoglires', 'Primates',
'Haplorrhini', 'Platyrrhini', 'Cebidae', 'Cebinae', 'Cebus']
```

```

/references=[Reference(title='Molecular estimates of primate divergences and new
hypotheses for primate dispersal and the origin of modern humans', ...),
Reference(title='Direct Submission', ...), Reference(title='Direct Submission',
...)]
/comment=REVIEWED REFSEQ: This record has been curated by NCBI staff. The
reference sequence was derived from AJ309866.
COMPLETENESS: full length.
Seq('GTTAATGTAGCTTAATACTCAAAGCAAGGCACTGAAAATGCCTAGACGGGTATT...TAA',
IUPACAmbiguousDNA())

```

If we look at the attributes of the object, we will see that there are several methods (e.g. id, features, annotations, seq, etc...) that could be used to provide direct access to some info from the genbank file:

```

[70]: for i in SeqIO.parse("Cebus_albifrons_NC_002763.1.gb", "genbank"):
      print(dir(i))

```

```

['__add__', '__bool__', '__class__', '__contains__', '__delattr__', '__dict__',
'__dir__', '__doc__', '__eq__', '__format__', '__ge__', '__getattr__',
'__getitem__', '__gt__', '__hash__', '__init__', '__init_subclass__',
'__iter__', '__le__', '__le__', '__len__', '__lt__', '__module__', '__ne__',
'__new__', '__nonzero__', '__radd__', '__reduce__', '__reduce_ex__', '__repr__',
'__setattr__', '__sizeof__', '__str__', '__subclasshook__', '__weakref__',
'_per_letter_annotations', '_seq', '_set_per_letter_annotations', '_set_seq',
'annotations', 'dbxrefs', 'description', 'features', 'format', 'id',
'letter_annotations', 'lower', 'name', 'reverse_complement', 'seq', 'translate',
'upper']

```

For instance, if we print the “annotation” field, we will get a dictionary, that could be used to obtain specific informations:

```

[111]: for i in SeqIO.parse("Cebus_albifrons_NC_002763.1.gb", "genbank"):
      print(i.annotations)
      print(type(i.annotations))

```

```

{'molecule_type': 'DNA', 'topology': 'circular', 'data_file_division': 'PRI',
'date': '01-FEB-2010', 'accessions': ['NC_002763'], 'sequence_version': 1,
'keywords': ['RefSeq'], 'source': 'mitochondrion Cebus albifrons (white-fronted
capuchin)', 'organism': 'Cebus albifrons', 'taxonomy': ['Eukaryota', 'Metazoa',
'Chordata', 'Craniata', 'Vertebrata', 'Euteleostomi', 'Mammalia', 'Eutheria',
'Euarchontoglires', 'Primates', 'Haplorrhini', 'Platyrrhini', 'Cebidae',
'Cebinae', 'Cebus'], 'references': [Reference(title='Molecular estimates of
primate divergences and new hypotheses for primate dispersal and the origin of
modern humans', ...), Reference(title='Direct Submission', ...),
Reference(title='Direct Submission', ...)], 'comment': 'REVIEWED REFSEQ: This
record has been curated by NCBI staff. The\nreference sequence was derived from
AJ309866.\nCOMPLETENESS: full length.'}
<class 'dict'>

```

Suppose we want to get the species name. “organism” is a key in that dictionary, and its value is

the name of the species. So, we could extract this information using a dictionary “get” method:

```
[60]: for i in SeqIO.parse("Cebus_albifrons_NC_002763.1.gb", "genbank"):
      print(i.annotations.get("organism"))
```

Cebus albifrons

We can also use the “format” field to convert a genbank file into a fasta:

```
[87]: for i in SeqIO.parse("Cebus_albifrons_NC_002763.1.gb", "genbank"):
      print(i.format("fasta"))
```

```
>NC_002763.1 Cebus albifrons mitochondrion, complete genome
GTTAATGTAGCTTAATACTCAAAGCAAGGCACTGAAAATGCCTAGACGGGTATTTACTAC
CCCATAAACACACAGGTTTGGTCCTAGCCTTTCTATTAGCCCTCAGTGAGATTACACATG
CAAGCATCTACTATCCTGTGAGAATGCCCTCTAGAACACCAAATTATGAGGAGCGAGTAT
CAAGCACGCATATATGCAGCTCAAAACACTTTGCTTAGCCACACCCCCACGGGAAACAGC
AGTGACAAACTTTTAGCAATAAACGAAAGTTTAACTAAGCTACACTGACAATAGAGTTGG
TCAATTTCTGCCAGCCACCGCGCCATACGATTAACTCAAGTTAATAAAGTCCGGCGTA
AAGAGTGTTTAAGGCCCCACCCTCAATAAAGCTAACCTATACTAAGTTGTGAAAACTC
CAGTTATAGTGAAATACCCTACGAAAGTGGCTTTAATATTCCTGAATACCCATAGCTAA
GACACAACTGGGATTAGATACCCCACTATGCCTAGCCCTAACTCCAATAACTCTACCA
ACAAAATTACTCGCCAGAACACTACAAGCAATAGCTTGAAACTCAAAGGACCTGGCGGTG
CTTTACATCCGTCTAGAGGAGCCTGTTCTGTAATCGATATACCCGATAAACCTTACCAC
CTCTTGCCCCCAGCCTGTATACCGCCATCCTCAGCAAACCTCCCTAAAGATCGTAAAGTAA
GCAAAAGTATTACCATAAAAACGTTAGGTCAAGGTGCAGCCAATGAAGTGGAAGAAAATG
GGCTACATTTTCTAATCTAGAAAATTACACGATAGCCTTTATGAAATTTAAAGGCCCAAG
GTGGATTTAGCAGTAAATCAAGAATAGAGAGCTTGATTGAAGCAAGGCCATTAAGCACGC
ACACACCGCCCGTCACCCCTCCTCAACCATCATGTAGAAAATATATTAACAATTTAATCCG
CTTCATATTGTATGCAGAGGGGATAAGTCGTAACATGGTAAGTGTACTGGAAAGTGTACTT
GGACGAATCAAAGTGTAGCTTAAATTAAGCATCTGGCCTACACCCAGAAGATCTCACAA
CAACCGATCACTTTGAGCTAACTCTAGCCCAAACAATCTCTCTATACTAATACCTAAATT
CAATAACCAAAATCATTTACCTACTATAAAAGTATAGGCGATAGAAATTATATCCGGGCGC
AATAGATGAAGTACCGTAAGGAAAGAAACCCCAACAAAGCATAAAAAAGCAAAGACAAG
CCCTTATACCTTGCGCATAATGAATTAAGTAACTAGAAATAACTTTATACAAAGAATTTAAATA
AAGCTCCCCGAAACAGACGAGCTACCCAAAGATAGCTAAAAGAGCGCACCCGTCTATGT
GGCAAAATAGTGGGAAAATCTATGGGTAGTGGCGACAAACCTACCGAGCCTGGTGATAGC
TGGTTGTCCAAGACAGAGTCTTAGTTCAACTTTATATTTACTTACAGAACCATTTAATCC
CCTCGTAAATTTAATTGTTAGTCTAAAGGGGGCAGCCCTTTAGACCCTAGGAAAAAACCC
TTACTATAGTGAGTAAAACACTTCACCACCATAGTTGGCCTAAAAGCAGCCATCAATTAA
GAAAGCGTTCAAGCTCAATATCCAAACACCCCTCTAATTCCATTAAGTTCACTGAACTCCT
AAAACACATTGGACTAATCTATTATCTAATAGAAGCAATAATGTTAATATAAGTAACATG
AAAACATTTCTCCCCGCATAAGCTTATTACAGATCGAAACAACCTACTGCTAGTTAACAG
GACAATCATTACAACCCACAACCTTAATACCCCATTAAGTAACTGTTAACCCAACACAGG
CATGCACCCAGGAAAGATTAAGGAAAGTAAAGGAACTCGGCAAACTTTACCCCGCCTG
TTTACCAAAAACATCACCTCTAGCATTAATAGTATTAGAGGCACTGCCTGCCCGGTGACA
CATGTTTAAACGGCCGCGGTACCCTGACCGTGCAAAGGTAGCATAATCACTTGTTCTCTAA
ATAGGGACTTGTATGAACGGCAACACGAGGGCTTAAGTGTCTCTTACTTTTAATCAGTGA
AATTGACCTATCCGTGAGGAGGCGGATATACACAAACAAGACGAGAAGACCCTATGGAGC
```

TTCAATTTAATAATACAAACTAGCACACTAAAAACCAACAGGCACTAACACACCATCAAT
GTATTATAAATTTTCGGTTGGGGTGACCTCGGAGTACAGTACAACCTCCGAAAACATAAGC
CATGACCTAACTAGTCTAAGTAAATTACACTCATTGACCCAATAACTTGATCAACGGA
AAGTTACCCTAGGATAACAGCGCAATCCTATTCTAGAGTCCATATCGACAATAGGGTTA
CGACCTCGATGTTGGATCAAGACATCCCAATGGTGTAGCCGCTATTAAGGGTTCGTTTGT
TCAACGATTAAAGTCTTACGTGATCTGAGTTCAGACCGGAGCAATCCAGGTCGGTTTCTA
TCTATTAAAGATTCTCCCAGTACGAAAGGACAAGAGAAATAGAGCCCACTTTACAAAGC
GCCCTCATAATGTCAGATGACTAATATCTTAACCTTACAAATTAATATGCATAGCCCAAA
AACAGGGCTTAGTTAAGATGGCAGAGCCCGGCAATTGCATAAACTTAAACTTTACAAT
CAGAGGTTCAACTCCTCTTCTTAACAATATGTTTATAATCAACCTACTCTTATTAAATTAC
CCCAGCCCTAGTTGCTATAGCATTTTTTAACACTTACAGAACGAAAAATCTTAGGCTATAT
ACAACTCCGCAAGGCCCAACATTGTAGGCCCATATGGGGTACTCCAACCAATCGCCGA
TGCAATAAACTCTTTACAAAAGAACCTCTACTACCTATCACATCTACCACAACCTATA
TATAATCGCCCCAACCTTAGCCCTAACCATTTCCCTTCTTCTATGGAGTCCCTCCCTAT
ACCGTACTCCCTAGTTAACTTCAATCTAGGCCTCCTATTTATACTGGCAACATCAAGTCT
AGCCGTTTACTCAACCTATGATCCGGATGAGCATCAAACCTCTAACTACGCATTAATTGG
CGCACTACGAGCCGTAGCCCAAACATTTCTACGAAGTTACCCTAGCCATTATTCTACT
ATCAACACTACTAATAAGTGGCTCCTTTAATCTACAATCACTAATTACAACACAAGAACA
ATCTTGACTTCTACTCCCATCATGACCCCTAACGATAATATGATTTATCTCCACACTAGC
AGAAACCAATCGAGCCCCCTTCGACCTAACAGAAGGTGAATCAGAACTAGTCTCAGGCTT
CAATATTGAATATGCCGAGGTTCAATTCGCCCTGTTCTTTATGGCAGAATATATAAACAT
TATTATAATAAATGCTCTAACTACTACCATCTTTACAGCAACATCCTACAATATAATCAC
AACAGAACTTTTACTTTAAACTTCCAACTAAAACACTTCTACTAACCACCCTATTTTT
ATGAATCCGTACGGCATACCCCCGTTTTTCGCTACGACCAATTAATGTATCTCCTATGAAA
AAAATTCCTACCCTACACTGGCACTATGTATATGATATATTTCAATACCTATATTATT
ATCCGGCATCCCACCCCAAACATAAGAAATATGTCTGACAAAAGAATTACTTTGATAGAG
TAACTATAGAGGTTTAAATCCTCTTATTTCTAGAATTATAGGACTCGAACCTATTCCTA
AGAACTCAAAATCCTCCGTGCTACCTATTACACCATACTCTAAACAGTAAGGTCAGCTAA
ATAAGCTATCGGGCCCATACCCCGAAAATGTTGGTCCAATCCTTCCCGTACTAATATTAA
TCCCCTAGCCACCTAATTATCTCCCTAACCATTCCAATAGGAACCATAATCACAATCCT
AAGCTCACACTGATTCTAGTATGAACAGGCTTAGAACTGAACATATTAGCTATTATACC
CATACTCGTAAAAATATAAATCCCCGCTCCACAGAAGCATCCACTAAGTATTTCTTAAC
ACAAGCATCCGCGTCCATAATCCTTCTAATAACTATTTTTCTCAACAATTTACTCTACGG
CCAGTGAACAATTAATCCACCTTCTAATCAAGTCTATCCACAATAATATTAATGGCTCT
CGTATTAAAATTAGGTATAGCTCCCCCTCCACTTCTGATTGCCAGAAGTAACCCAAGGCAT
TCCCCTAATTCCCCTACTATCTTACCTTACATGACAAAACTAGCCCCCATGTCAATTAT
TATTCAAATCTTCCCCTATCTCAACTCAAGTGCCCTACTAATAATCTCACTTATATCAAT
CATAATTGGCAGCTGAGGAGGACTTAACCAACACAACTACGCAAAAGCCTAGCCTATAC
CTCAATCACCCTATAGGGTGAATAATGGCAGTACTATTCTATAACCCAAACATTACCGC
CTTAAGTTTACTTATCTACATCTTCTTAACAATTTCCACACTTATACTTTCTACTCAAA
CTCAACACAACAACCTTATCATTATCCCACACTTGAAATAAACTCACATGAGTTATACC
AATAATCCCAATAATAATAATATCCTTAGGAGGCCTCCCCCACTAACAGGCTTCTCCCC
CAAATGAGCTATTATACAAGAAATTACAAAAACAACAGCCTTATTCTCCCCCTCACCAT
AACAATACTTACACTAATAAACTTATATTTCTACATACGCCTAACATACTCAATCTCGAT
AACAATATTTCTACATCCAACAACACAAAAATTAGCTGACAACTAAAACATCTAAAACC
AATACCCTTCTACCTCCCCTCATAATTTCTTCTCCCTCCTACTACCAATCACACCACT
AATGCTAATGACCTAGAAATTTAGGTTAATAAGACCAAGAGCCTTCAAAGCCCCTAGTAA
GTAAATTTTACTTAATTTCTGCACCGCTCTAAGGACTGCAAACTATACCTTGCATCAATT

GAACGCAAATCAACTACTTTAATTAAGCTAAGCCCTCCCTAGACTGATGGGACTTTAACC
CACAAAAATTTAGTTAACAGCTAAATAACCTAATCAACTGGCTTCAATCTAATTCTCCCG
CCTTTAAGGGGAAAAAAGGCGGGAGAAGCCCCGGCAGCATTGAAGCTGCTTCTTTGAAT
TTGCAGTTCAACATGACAGTTCACCTCAGGGCTGGTAAAAAGAGGGGTCACCTCTCTGT
CTTTAGATTTACAGTCTAATGCTTACTCAGCCATTTTACCTTCTACTTATGTTTCATAAAT
CGCTGACTATTTTCAACTAACCATAAAGATATTGGTACACTGTATTTAATATTTGGTGCA
TGAGCCGGAGCAACAGGAACAGCCTTAAGTCTTCTAATTCGAGCTGAGCTGGGCCAACCA
GGAAGCCTAATAGAAGACGACCATGTTTACAATGTTATTGTTACCTCTCACGCATTTATT
ATAATCTTCTTCATGGTCATACCAATTATAATTGGCGGCTTCGGGAATTGATTAGTGCCT
CTAATAATCGGCGCCCCGATATAGCTTTTCTCTGTATAAATAATATAAGCTTCTGACTC
CTCCCCCATCCCTTCTTCTCCTACTTGCCCTTCAACTCTAGAGGCTGGTGTGGAACT
GGCTGAACAGTCTACCCTCCCCTGGCAGGAAATATATCACACCCTGGAGCCTCTGTAGAC
TTAACTATCTTTTCACTGCATCTAGCGGTATTTCTCTATTCTAGGGGCTATTAACTTT
ATTACAACAATTATTAATATAAAAACCACCAGCCACAACCCAATATCAAACACCCCTATTT
GTATGATCCGTACTCATTACAGCAATCCTTCTACTTCTCTCCCTCCCAGTCCTAGCTGCT
GGAATTACTATACTATTAACCGACCGTAACCTTAACACCACCTTTCTTCGACCCTGCTGGT
GGTGGTGACCCCATTTCTATATCAACACCTATTTTGATTTTTTGGTCACCCCGAAGTTTAT
ATCCTTATTTTACCAGGGTTTGGGATAATCTCACATATTGTAACATATTACTCTAACAAA
AAAGAGCCATTTGGGTATATAGGAATAGTTTGAGCCATAATATCCATTGGCTTCTAGGC
TTTATCGTATGAGCTCACCATATATTACAGTAGGAATAGATGTGGATACACGCGGTAT
TTTACATCAGTACCATAATCATTGCCATCCCCACTGGGGTCAAAGTATTTAGCTGGTTG
GCTACACTACATGGAGGCAACATCAAGTGATCTCCTGCAATACTATGAGCCCTGGGTTTT
ATTTTTCTCTTCACTGTGGGTGGATTGACAGGAGTTGTATTAGCTAACTCATCATTAGAT
ATTGTCCTACATGATACATACTACGTGGTAGCCCATTTCCATTACGTCTTATCAATAGGA
GCAGTGTTTGGCATTATAGGAGGCTTCATTCACTGATTTCCATTATTTTCAGGCTATACA
CTCGACCAAACCTACGCTAAAATCCACTTTACTATTATATTTGTAGGTGTAAACATAACC
TTCTTTCCACAACACTTTCTCGGCCTATCTGGAATACCTCGACGGTACTCAGACTACCCC
GATGCATATACTGCATGAAATATTATCTCATCCGTAGGCTCATTTATTTCACTGGCAGCA
GTCATTCTAATAATCTTCATAATTTGAGAAGCCTTCTCTTCAAAAACGAAAAGTTCTAACC
ATCGAACAAATATCTACCAACCTAGAGTGGTTATATGGCTGCCCCCTCCTTATCACACA
TTTGAAGAGGCTACCTACGGTAAACTCCTAAACGAAAAAGGAAGGATTTGAACCCCAA
AAATTGGTTTCAAGCCAATCCCATAGACCCTATGACTTTTTCTATGAGATATTAGTAAAA
TAAATTACATAACTTTGTCAAAGTTAAATTATAAGACCAAAACCTATATATCTTAGTGGC
AACACCAGCTCAACTAGGCCTACAAAACGCTACATCCCCATTATAGAAGAACTCATTGC
CTTCCACGACCACACCTTATAATTATTTTCTAATTAGTTCACTAGTCCTATACATTAT
CTCTTTAATGCTTACCACAAAACCTAATCATACCAGCACTATAAATGCCCAAGAAATCGA
AACAATTTGAACATATCTACCCGCAATTATTCTTATTATAATTGCCCTTCCATCCCTACG
CATTTTATATATAACAGATGAATTTAATAAGCCCTACTTAACCCTTAAAGCAATTGGCCA
CCAATGGTATTGAAGCTATGAGTACTCAGACTATGAGGACCTATTCTTTGATTCTACAT
TATGCCAACCTACTACCTCCAACCAGGCGAATTCCGACTTCTTGAAGTGGACAATCGAAC
AACCTTACCAATAGAAGCAGATATTCGATATCTAATCTCATCACAAGATGTATTACACTC
ATGAACCGTCCCATCATTAGGCGTTAAAGCAGATGCAATTCAGGCCGATTGAACCAAGC
CATACTAGCCTCAATACGACCAGGACTATTTTACGGGCAATGCTCAGAAATTTGCGGGTC
AAATCACAGCTTTATACCTATTGTCCTAGAATTTATCTATTTCCAAGATTTCGAAGTATG
GGCTTCATACTTATATATTGTATCACTGTAAAGCTACTTAGCATTAAACCTTTTAAGTTAA
AGACAGAGAGAATCTCTCTACAGTGAGTGCCACAACCTAGACATCTCACCATGACCAATGG
TGACTTTTATCAATGATTTTAAACCCTGTTTTATGCTATACAACCTAAAAATATTAATAATTA
TTTTCCACACTACCCCACTATCAAAATTAACCTAAAAATTCAAAACCAAAAAACACCTGAG

AACTAAATGAACCAAAATCTATTTGCCTCTTTCAATATACCAATAATCCTGGGAATCCC
CCTAGTAGCACTATTTATTTTTATTTCCCACTATACTAATTACACCCTCTAACAACCTAAA
CAATAACCGATCCTCCTCCCTTCAACAATGACTAATCCAACCTTGTAATAAACAAATAAT
GACAAACCATAACCACAAAGGACAAACCTGATCCCTTATACTTTTAACCCTAATTACCTT
TATTACCCTGAACAACCTTCTCGGAATTACACCCTACGCATTTACACCTACCACACAACCT
ATCATTAATCTAGGCATGGCAATTCCTTATGAATAGCAACCGTACTTATAGGACTTCG
ATTTAAACAAAAGCAACCTTGCTCACCTCTTACCTCAAGGAACACCTATCCCACTCAT
CCCTATACTTATTATTATTGAGACTATCAGCCTCTTCAATTCAACCTATGGCACTAGCCGT
ACGATTAACAGCCAACATCACAGCAGGCCACCTATTAATGCATCTACTAGGAGACACTAT
ATTAATTCTTCTTTCTATTTACCTCTCTTCCCTCGTAATCACCGTTATCGTTATTATTCT
ATTAATTACCCTAGAACTAGGCGTAGCCCTAATCCAAGCCTATGTATTCACACTCCTAGT
AAGTCTCTATTTGCACAATAACTCATAATAATTTCCCAACCAATTATCTATAATGACACA
CCAAACACACGCTTATCATATAGTAAATCCAAGCCCCTGACCACTAACAGGAGCTCTATC
AGCTTTCCTATTAACCTCCGGCCTAGTTATATGGTTCCATTTCTACTACACCCCCCTATT
TAATGCAGGCTTACTAGCCAGTGCAATAACAATAGCTCAATGATGACGCGACGTAGTGCG
AGAAAGTACATATCAAGGCCACCACACTATGCCCGTTCAAAAAGGCCTTCGGTATGGAAT
AATCCTATTTATTATTTTCAAGTTTTCTTCTTCGCGGATTCTTCTGGGCATTCTATCA
CTCAAGCCTAGCTCCAACCCCTCAAACAGGAGGACATTGACCTCCACAGGTATCTTTCC
CCTAAACCAATAGAACTACCACTTCTAAATACAGCTGTCTTACTCGCATCAGGAGTCAC
AATTACCTGAGCACATCATAGCCTAATAGAAGCCAACCAAGAAAATTCAACCCAAGCATT
GTCCTTAACCTATCGCACTAGGAATTTACTTTACCTATCTTCAAGTATCAGAGTACTCTGA
AACACCCCTTCTCCATCTCCGACGGAATTTACGGCTCCACATTCTTTATAGCCACAGGCTT
CCATGGTCTTCATGTTATTATCGGAACCACCTTCCTTGCCACATGCTATATTGCGCAACA
ATTATATCACTTTACACCTAATCACCACCTTTGGATTTGAAGCCGCGCATGATACTGACA
TTTTGTAGATGTGGTATGACTATTCCTCTACATCTCTATCTACTGGTGAGGATCCTATTC
TCTTAGTATAAACAGTACAATTGACTTCCAATTAATAGGCCTTGATAAACCCAAGAGAGA
ATAATTAACCTAATATTAACCCCTAGCAACCAGTACCCTACTAGCCCTACTACTAATTACT
ATTACATTCTGAATACCACAATTAATAAATACACAGAAAAACACAACCCCTATGAATGC
GGATTTGATCCTACAACCTCCGCCCACCTACCCTTTTCTATAAAATTTTTCCTAGTAGCT
ATCACATTCCCTCTATTTGACCTAGAAATTGCCCTACTCCTACCCTTACCGTGAGCTACC
CAAACAGATAACTTAATACTGATAACTAACATAGTCTTTACTCTACTTATTATTCTAGCA
CTAGGACTAGCCTACGAGTGGACCCAAAAGGGATTAGACTGAGTTGATTTGGTATATAGT
TTAAACAAAATAAGTGATTTGCACTCATTAGACTATGGTAAACCGTATATACCAAAATGC
CTTTTATCTATATCAACACCGCACTGGCATACTCTATATCTCTACTAGGACTGTAAATTT
ACCGATCCCACCTAATATCGTCCCTATTATGCTTAGAAGGTATGATACTATCACTATTTA
TTATAATAACAACCATAACCTTAAATATACACCTAATATTAATATACATACTACCCATTA
TTCTACTGGTATTCGCGCATGTGAAGCTGCAGTAGGCCTAGCCTTATTAATTTTAATCT
CCAACCTTATATGGCTTAGACTACGTACAAAACCTAACTTACTCCAATGCTAAAAATTAT
TCTCCCGACAATCATAATAATCCCAACAATATATCTATCAAAAAACCATATAATATGAAT
CAATACAATAATCTGCAGCCTATCAATTAGCGCCCTAACCCCTATAATTCTACATATACC
CAACAACCCATGCAACTTATCGCTAACCTTTCTTCTCAGACCCATTAACATCACCACCTTTT
AGCTCTAACAACTGACTACTGCCACTAATAATCTTAGCAACACAACAACATATTTACAG
CAACTCCCTTCCCCGAAAAAAATTATACACTTCAATACTAATTATCCTACAAATTTCCCT
AATTATAACATTTGCAGCTACAGAACTAATCCTATTCTACATTTTATTTGAACTACTCT
AATTCCAACCCTAATTATTATTACCCGCTGAGGATACCAGCCAGAGCGCCTCAACGCCGG
CTCGTACTTTCTATTCTACACACTAGCTGGATCACTTCCACTACTCATTACACTCCTTTA
CTGCCTCAACAACCTTAGGAACCTAAATATCCTAACAATGACAATTAACACCAAGAATT
ACTAACATCTTGAACCTAATAATATTATATGACTAGGATGCACGATAGCCTTCATAGTAAA

AATACCTCTGTACGGCCTTCACCTATGACTCCCCAAAGCCCATGTAGAAGCCCCAATTGC
CGGCTCAATAGTACTTGCAGCAATCCTGCTAAAACTAGGTAGCTATGGCATAATACGAAT
CATTCCTACCCTCAACCCCTTAACAGAAAAATAAGCTACCCCTTCCTCATCCTATCCCT
ATGAGGCATGGTAATAACAAGCTCCATCTGCTTACGACAGGCCGACCTAAAATCATTAAT
TGCCTACTCTTCCATCAGCCATATGGCGCTTATTACCCTGGCTATTCTTATCCAAACCCC
CTGAAGCCTTACCGGTGCAATACTACTAATAGTCGCACATGCATTTACCTCATCCCTACT
ATTCTGTCTAGCAAACCTCCAATTACGAACGTATTCACAGCCGAACCTATAATATTTACCCG
AGGCCTCCAAGCACTATTTCCACTCCTAACTCTCTGATGACTTCTAGCAAACCTTGCCAA
TCTTGCTCTACCTCCAATACTAATCTACTAGGAGAATTATCTACAATTTTAGCCGCCTT
CTCCTGATCTAATTTTACCATCGCATTTACAGGATTCAATATACTTATCACAGCACTATA
CTCACTACATATATTTACCTCAACACAACGAGGACCATTAACACACAGCACCAGCAATGT
AAAACCCCTATTTACAGGAGAAAAATACTAATACTAATAACATATAGCACCATCCTTCT
TCTCACTTTAAACCCCAAAAATAATGATAGGCCTAATACCATGTAGCTATAGTTTAATTA
AACATTAGATTGTGAATCTAATAATAGAAGACTACAACCTTCTTAACTACCGAGAAAAGTAT
GCAAGAAGTGTAACTCCTGCCTCCAGACTTAATATCCTGGCCTTCTCAGCTTTTAAAGG
ATAGTAGTTATCCATTGGTCTTAGGAGCCAAAAATATTGGTGCAACTCCAAATAAAAGCA
AAATGCACCTTTTCTATTACTCTAATAACACTAATTCCCTTACTAGTACCTATCATAACTA
CCCTAGTTAAATCTAACAAAAATCCCTTATATCCACACTATGTAAAAGTACCCATTATTT
ATGCCCTTATTACTAGCACCTTGTCTATAACAATATTTATTCTCACAGGCCAAGAATCAA
TAATTTCAAAGTACATTGAGCGACAATCCAACTATCAAAGTATCACTAACTTTAAAC
TAGACTTCTTCCATAATATTACGCCTGTAGCACTATTGTTACCTGATCAATTGTAG
AATTCTCAATATGATATATAAACTCAGACCCAAATATTAATCAATTCCTCAAGTACTTAC
TTATATTCCTCATCACAATATTAATCTTAATTACCGCTAATAACTTATTCCAAGTATTTA
TCGGGTGAGAAGGAATAGGTATTATATCCTTTCTACTAATTAGCTGGTGGTATGGCCGAA
CAGAAGCTAACACAGCAGCCTTACAAGCAATCTTATATAACCGTATTGGAGATATTGGCC
TTATCCTTGCAATAGCATGATTTTTTTTTTACACTCCAACCTCATGAGATTTCCAGCAAAATAT
TTATTCTTAACCCCACTACAACTTTTTTCCCTGATAAGCCTTCTCCTAGCAGCAACAG
GAAAATCAGCCCAATTTGGTCTCCACCCATGACTTCCCTCCGCCATAGAAGGACCCACCC
CAGTCTCAGCACTACTTCACTCCAGTACAATAGTCGTGCGAGGTGTCTTCTAATTATCC
GCTTTCATCCTCTAATAGAAAACAACCTTACTCATCCAAACACTTACCTTATCGCTAGGGG
CCATTACCACCCTATTACATCAATCTGTGCTCTAACACAAAATGACATAAAAAAAATTG
TGGCCTTCTCAACCTCAAGCCAACCTAGGCCTTATAACAGTAACAGTCGGTATTAACCAAC
CACATTTAGCCTTTCTACACATCTGTACTCATGCCTTTTTTCAAAGCTATATTATTTCTAT
CCGCAGGATCCATTATCCACAATCTCAACAACGAACAAGACATTCGAAAAATAGGAGGCC
TATTTAAACCATACTTTCACTGCCTCCTCCCTCATTATTGGCAATCTTGCACTTATAG
GAACACCATTTCTTACAGGTTTCTACTCAAAAGATCTAATTATCGAAACCATCAACACGT
CATATACCAACGCCTGAGCCCTTACAACCTACCCTCGTAGCCACCTCCCTTACAGCAATAT
ACAGTGGCCGATTATATTTTTTTACCTTAACAGGATACCCTCGCTTTACAACCTCCACCC
TAATCAACGAAAAATAACAATCTCCTAATAAATCCAATTAGCCGCCTAGCAGTAGGTAGTA
TCCTTGCCGGGTTTCTTATTTCTAATTGCTTACCCCTACTATATACCCCAAAATAACCA
TACCTTTCCACCTTAAGCTTACAGCTCTAAGTGTAAACCACCTAGGACTCCTTATAGCAA
TAGAACTTAACTCTTTAACTAACAACATAAACTATATACCCAGTAAAAATTTACTACT
TCTCTAATATACTAGGCTTCTACTCAATTACTACTCACCGCCTCAACCCCATTTAAACC
TAACCACAAGTCAAAAATTTACCTCCACTTTACTAGACCTATTCTGACTAGAAAAATCTA
TACCAAAATTAACAACACAAACACAAATTTCAATAGCTACAATTACATCAACTCAAAAAG
GCCTAATTAAACTCTACTTCTTTACCTTCTTTATTCCACTCATCCTAACGCTTTTTCTTA
CAATTTAACCCCTCCCCGAGTTAACTCAATTGCAATATGTATACCCATAAATAACGCCC
AACAAGTAACTAAAACAACTCAAACACCATAATCATATAAAGCAGCAGCACCTGCAGGAT

CCTCACGAATCAAACCCGGCCCTCACCTCATAAATCATTCAACCTGCCACAGTATCAT
AATTAACAGCAATCTCCACCGTTTTATTAGGACTACCCCCAGTAATAATACCACCACTA
CCTCCATTGCTAAACCCAATACAAGCATCCCTAAAAATATCAACACTCGATACCCATGTCT
CAGGGTATTTCATCGACTGCTATCGCTGCAGTATAACCAAAAAACAACCATTATACCACCCA
AATAAATTAAAAAAACTATAAGCCCTATATAAGATCCACCAAGATACAGTATAAGTACAC
AACCCACAGCACCCTAAAAACCAACACCAACCCCCATAAATAGGTGAAGGTTTAGAGG
AAAATCCCACAAAACCTGTTACCAATATAATACTTAGCAAAAAATAAAATATACATCACTA
TTCCACATGGATTATAACCATGACTAATGATATGAAAAGCCATCGTTGTATTTCAACTA
TAAGAATACTAATGACCTCTCCCGCAAAACACACCCATTAATAAAAAATTATTAATAGTT
CATTTATTGATCTGCCCACACCATCCAACATCTCCTCCTGATGAAACTTCGGATCACTTC
TAGGCGCCTGCCTAATAATTCAAATTACCACAGGCCTATTCTTAGCGATACACTATACGC
CAGACACCTCAACCGCCTTCTCCTCAGTAGCACATATTACCGAGATATTAATTACGGTT
GAATAATCCGCCTCCTACACGCCAATGGTGCCTCCATATTTTTTGTGTGCTTATTTCTCC
ACACTGGCCGAGGCCTCTACTACGGATCTTTTCTCTTTCTAAACACCTGAAATATTGGTA
CAATCCTATTATTAATAACAATAGCCACAGCCTTTATAGGCTATGTCTTACCGTGAGGCC
AAATATCATTCTGAGGAGCCACAGTTATTACAAATCTTCTATCAGCCATCCCCTATACCG
GACATAACCTTGTACAATGAATCTGAGGTGGCTTTTCAGTAGATAAACCCACCCTCACAC
GATTCTTTACCTTTCACTTTATTTTACCTTTTATTATCACAGCCCTAGCAACTATTCACC
TTTTATTTCTACATGAAACAGGCTCAAATAATCCATCAGGAATAGCATCTAGCCCCGATA
AAATTATATTCCATCCCTACTACACAACCAAAGATATTTTTGGATTAAACCCTTCTTCTCC
TACTCCTTACAAGCCTAACCTATTTACCCCCGACCTTTTAACTGACCCAGATAAECTACA
CACTAGCCAACCCCTTAATACTCCACCCCATATTAAGCCAGAGTGATACTTTCTATTTCG
CATACACAATTTTACGATCTATTCCAAATAAACTAGGAGGTGTTCTAGCTCTTCTATTAT
CTATTATAATCCTAACAATTATCCCTGCCACTCACCTATCCAAACAACAAAGTATAATAT
TCCGACCAATCACCCAAATCCTATTCTGAACCCTAGCAGCCGATCTACTTACACTTACAT
GAATTGGAGGCCAACCCAGTAGAATACCCCTTTGAAGCCATTGGCCAAACCGCATCTATTG
CTTACTTCCTTATTACTCTAATTCCTCTATCAGCCCTAACTGAAAATAAGCTACTTA
AATGATAACGTCTTGTAGTATATCCAATTACCCCGCCTTGTAACCGGAAAAAGGAGGC
ACGCTAACTCCCAGGACAATCAGGAAAAGAATAATTAATTCTACCATCAACACCCAAAG
CTGAAATTCTACATTAACTATTCCCTGAGCATAGTAAATAATAGATATATATACTATTT
GATGGCTTACCCTAAAGTACTCTAAAGGCATACACAATTTTTTTTCCCCTATGTAATTA
GTGCATTATTGCTTATCCCCATGAATAATACATAGTACTACACATGCTTAATTATACATA
GCACATAGAGTCAAAACGTGCATACTACGTACAGAACATGCTTACAAGCAGGAACTATT
ATTTACAGCGGACTATAGCGCATTAACTCAAACCCACGCTCCAATTACAGTCACCACG
GATATCGTACAGCCCATAGAATCATTGATAGTACATCTGCACATCAAATGATTAACCGGA
CATAGCGCATCCTATTTCTTCAGTCCTTCTAACCATGGATATCCCCTTGTATATTTGGTC
TCTTAATCTACCAACCTCCGTGAAACCACCAACCCGCCCCTTCTGCGGCTCTTCTCGCT
CCGGGCCCATATAGACAGGGCTTGGTTATACTGAACTATACCTGGCATTGTTGTTCTAC
CTCAGGGCCATCTCACTAAGACCGTGTCCACGTTCTCTTAAATAAGACATCACGATGGT
GTGGCGCTATCACCTCTTAACCGCGTCACTGGATGCATAGGGTGCCTGGGTGGGGAAAAG
GGGGGGGAGGGAAATCCTCAGCATTGCCGTAGGCTCCGGTAGAAGTCCCGCCCCCGTC
CTGTGGGACCTGTCTGTGTACTGCCATGCCTTATGCTATCATCGCACCTAAATTGAATGT
CTTGGCCCCAACCCGCCCCTAGGTGTTATTTCAGTCAATGGTCTCAGGACATAATAGAT
AATTACAGCACTATACCCAAAAATTTAACACACCAACCCCTTCAACCTTATCCCATATTTT
CTGGGGCGAGCGTCGTAAAGAGATATTTTACCATTAGCAACTAAATTTTATAAATCAACA
AATTAATAGAATTTTACCCCCATAGTACACACCTATTAATAATACTAGATACAATT
ACCTACCGCATACCTCAATCTCTATCCCTCAGAAAAACAGATCTATACCACATAATCGCCC
TGACCTCAAATCTATATAACCTACTGCTATATCCTTATACTGGGCCCTAATTAA

We can also get the reverse complement of a sequence. If we only use `reverse_complement`, we will get some information regarding the other strand:

```
[92]: for i in SeqIO.parse("Cebus_albifrons_NC_002763.1.gb", "genbank"):
        print(i.reverse_complement())
```

```
ID: <unknown id>
Name: <unknown name>
Description: <unknown description>
Number of features: 55
Seq('TTAATTAGGGCCAGTATAAGGATATAGCAGTAGGTTATATAGATTTGAGGTCA...AAC',
IUPACAmbiguousDNA())
```

But if we look at the methods of the `reverse_complement`, we will see that there are many methods to extract more specific data:

```
[91]: for i in SeqIO.parse("Cebus_albifrons_NC_002763.1.gb", "genbank"):
        print(dir(i.reverse_complement()))
```

```
['_add__', '__bool__', '__class__', '__contains__', '__delattr__', '__dict__',
 '__dir__', '__doc__', '__eq__', '__format__', '__ge__', '__getattr__',
 '__getitem__', '__gt__', '__hash__', '__init__', '__init_subclass__',
 '__iter__', '__le__', '__le__', '__len__', '__lt__', '__module__', '__ne__',
 '__new__', '__nonzero__', '__radd__', '__reduce__', '__reduce_ex__', '__repr__',
 '__setattr__', '__sizeof__', '__str__', '__subclasshook__', '__weakref__',
 '_per_letter_annotations', '_seq', '_set_per_letter_annotations', '_set_seq',
 'annotations', 'dbxrefs', 'description', 'features', 'format', 'id',
 'letter_annotations', 'lower', 'name', 'reverse_complement', 'seq', 'translate',
 'upper']
```

For instance, we could get the reverse complement:

```
[113]: for i in SeqIO.parse("Cebus_albifrons_NC_002763.1.gb", "genbank"):
        print(i._seq)
```

```
GTTAATGTAGCTTAATACTCAAAGCAAGGCACTGAAAAATGCCTAGACGGGTATTTACTACCCCATAAACACACAGGTTTG
GTCCTAGCCTTTCTATTAGCCCTCAGTGAGATTACACATGCAAGCATCTACTATCCTGTGAGAATGCCCTCTAGAAGACC
AAATTATGAGGAGCGAGTATCAAGCACGCATATATGCAGCTCAAAACACTTTGCTTAGCCACACCCCCACGGGAAACAGC
AGTGACAAACTTTTAGCAATAAACGAAAGTTTAACTAAGCTACACTGACAATAGAGTTGGTCAATTTTCGTGCCAGCCACC
GCGGCCATACGATTAAGTCAAGTTAATAAAGTCCGGCGTAAAGAGTGTTTAAAGGCCCCACCCTCAATAAAGCTAACCTAT
AACTAAGTTGTGAAAACTCCAGTTATAGTGAATAACCTACGAAAGTGGCTTTAATATTCCTGAATACACCATAGCTAA
GACACAACTGGGATTAGATACCCCACTATGCCTAGCCCTAACTCCAATAACTCTACCAACAAAATTACTCGCCAGAAC
ACTACAAGCAATAGCTTGAAACTCAAAGGACCTGGCGGTGCTTTACATCCGTCTAGAGGAGCCTGTTCTGTAATCGATAT
ACCCCGATAAACCTTACCACCTCTTGCCCCAGCCTGTATACGCCATCCTCAGCAAACTCCCTAAAGATCGTAAAGTAA
GCAAAAGTATTACCATAAAAACGTTAGGTCAAGGTGCAGCCAATGAAGTGGAAGAAATGGGCTACATTTTCTAATCTAG
AAAATTACACGATAGCCTTTATGAAATTTAAAGGCCAAGGTGGATTTAGCAGTAAATCAAGAATAGAGAGCTTGATTGA
AGCAAGGCCATTAAGCACGCACACACCGCCCGTCACCCTCCTCAACCATCATGTAGAAAATATATTAACAATTTAATCCG
CTTCATATTGATGCAGAGGGGATAAGTCGTAACATGGTAAGTGTACTGGAAAGTGTAAGTGGACGAATCAAAGTGTAGCT
```

TAAATTAAAGCATCTGGCCTACACCCAGAAGATCTCACAACAACCGATCACTTTGAGCTAACTCTAGCCCAAACAATCTC
TCTATACTAATAACCTAAATTCAATAACCAAATGATTTACCTACTATAAAAGTATAGGCGATAGAAATTATATCCGGGCGC
AATAGATGAAGTACCGTAAGGAAAGAAACCCCAACAAAGCATAAAAAAGCAAAGACAAGCCCTTATACCTTGCGCATAA
TGAATTAAGTAGAAATAACTTTATACAAAGAATTTAAATAAAGCTCCCCGAAACCAGACGAGCTACCCAAAGATAGCTAA
AAGAGCGCACCCGTCTATGTGGCAAAATAGTGGGAAAATCTATGGGTAGTGGCGACAAACCTACCGAGCCTGGTGATAGC
TGGTTGTCCAAGACAGAGTCTTAGTTCAACTTTATATTTACTTACAGAACCATTTAATCCCCTCGTAAATTTAATTGTTA
GTCTAAAGGGGGCAGCCCTTTAGACCCTAGGAAAAACCTTACTATAGTGAGTAAAACACTTCACCACCATAGTTGGCC
TAAAAGCAGCCATCAATTAAGAAAGCGTTCAAGCTCAATATCCAAACACCCTCTAATTCCATTAACTTCACTGAACTCCT
AAAACACATTGGACTAATCTATTATCTAATAGAAGCAATAATGTTAATATAAGTAACATGAAAACTATTCTCCCCGCAT
AAGCTTATTACAGATCGAAACAACACTACTGCTAGTTAACAGGACAATCATTACAACCCACAACCTTAATACCCATTAACTA
AACTGTTAACCCAACACAGGCATGCACCCAGGAAAGATTAAGGAAAGTAAAGGAACTCGGCAAACTTTACCCCGCCTG
TTTACCAAAAACATCACCTCTAGCATTAAATAGTATTAGAGGCACTGCCTGCCCGTGACACATGTTTAAACGGCCGCGTA
CCCTGACCGTGCAAGGTAGCATAATCACTTGTCTCTAAATAGGGACTTGTATGAACGGCAACACGAGGGCTTAACTGT
CTCTTACTTTTAAATCAGTGAAATTGACCTATCCGTGAGGAGGCGGATATACACAAAACAAGACGAGAAGACCCTATGGAGC
TTCAATTTAATAATACAACTAGCACACTAAAAACCAACAGGCACTAACACACCATCAATGTATTATAAATTTTCGGTTGG
GGTGACCTCGGAGTACAGTACAACCTCCGAAAACATAAGCCATGACCTAACTAGTCTAAGTAAATTACACTCATTGACCC
AATAACTTGATCAACGGACTAAGTTACCCTAGGATAACAGCGCAATCCTATTCTAGAGTCCATATCGACAATAGGGTTTA
CGACCTCGATGTTGGATCAAGACATCCCAATGGTGTAGCCGCTATTAAGGGTTCGTTTGTTCACGATTAAAGTCTTACG
TGATCTGAGTTCAGACCGGAGCAATCCAGTCCGTTTCTATCTATTAAGATTCCCTCCAGTACGAAAGGACAAGAGAAA
TAGAGCCCACTTTACAAAGCGCCCTCATAATGTCAGATGACTAATATCTTAACCTTACAAATTAATATGCATAGCCCAAA
AACAGGGCTTAGTTAAGATGGCAGAGCCCGGCAATTGCATAAAACTTAAACTTTTACAATCAGAGGTTCAACTCCTCTTC
TTAACAATATGTTTATAATCAACCTACTCTTATTAATTACCCAGCCCTAGTTGCTATAGCATTTTTAACTTACAGAA
CGAAAAATCTTAGGCTATATACAACTCCGCAAAGGCCCAACATTGTAGGCCCATATGGGGTACTCCAACCAATCGCCGA
TGCAATAAAACTCTTTACAAAAGAACCTCTACTACCTATCACATCTACCACAACCTATATATAATCGCCCCAACCTAG
CCCTAACCATTTCCCTTCTTCTATGGAGTCCCTCCCTATACCGTACTCCCTAGTTAACTTCAATCTAGGCCTCCTATTT
ATACTGGCAACATCAAGTCTAGCCGTTTACTCAACCTATGATCCGGATGAGCATCAAACTCTAACTACGCATTAATTGG
CGCACTACGAGCCGTAGCCCAACCATTTTCATACGAAGTTACCCTAGCCATTATTCTACTATCAACACTACTAATAAGTG
GCTCCTTTAATCTACAATCACTAATTACAACACAAGAACAATCTTGACTTCTACTCCCATCATGACCCCTAACGATAATA
TGATTTATCTCCACACTAGCAGAAACCAATCGAGCCCCCTTCGACCTAACAGAAGGTGAATCAGAACTAGTCTCAGGCTT
CAATATTGAATATGCCGAGGTTTCATTGCGCCTGTTCTTTATGGCAGAATATATAAACATTATTATAATAAATGCTCTAA
CTACTACCATCTTTACAGCAACATCCTACAATATAATCACAACAGAACCTTTTACTTTAACTTCACAATAAAACACTT
CTACTAACCCACTATTTTTATGAATCCGTACGGCATACCCCGTTTTTCGCTACGACCAATTAATGTATCTCCTATGAAA
AAAATTCCTACCACTCACACTGGCACTATGTATATGATATATTTCAATACCTATATTATTATCCGGCATCCACCCGAAA
CATAAGAAATATGTCTGACAAAAGAATTACTTTGATAGAGTAACTATAGAGGTTTAAATCCTCTTATTTCTAGAATTAT
AGGACTCGAACCTATTCTTAAGAACTCAAAATCCTCCGTGCTACCTATTACACCATACTCTAAACAGTAAGGTCAGCTAA
ATAAGCTATCGGGCCCATACCCCGAAAATGTTGGTCCAATCCTTCCCGTACTAATATTAATCCCCTAGCCACCTAATTA
TCTCCCTAACCATTTCCAATAGGAACCATAATCACAATCCTAAGCTCACACTGATTCCTAGTATGAACAGGCTTAGAACTG
AACATATTAGCTATTATACCCATACTCGCTAAAAATATAAATCCCCGCTCCACAGAAGCATCCACTAAGTATTTCTTAAC
ACAAGCATCCGCGTCCATAATCCTTCTAATAACTATTTTTCTCAACAATTTACTCTACGGCCAGTGAACAATTAATCCAC
CTTCTAATCAAGTTCTATCCACAATAATATTAATGGCTCTCGTATTAATAATAGGTATAGCTCCCCTCCACTTCTGATTG
CCAGAAGTAACCCAAGGCATTCCCCTAATTCCCCTATACTTATCCTTACATGACAAAACTAGCCCCCATGTCAATTAT
TATTCAAATCTTCCCATCTATCAACTCAAGTGCCCTACTAATAATCTCACTTATATCAATCATAATTGGCAGCTGAGGAG
GACTTAACCAAAACACAACACTACGAAAAGCCTAGCCTATACCTCAATCACCCTATAGGGTGAATAATGGCAGTACTATTC
TATAACCAAAACATTACCGCCTTAAGTTTACTTATCTACATCTTCTAACAATTTCCCACTTATAACTTTCTACTCAAAA
CTCAACACAACAACCTTATCATTATCCCACACTTGAAATAAACTCACATGAGTTATACCAATAATCCCAATAATAATAA
TATCCTTAGGAGGCTCCCCCACTAACAGGCTTCTCCCCCAATGAGCTATTATACAAGAAATTACAAAAACAACAGC
CTTATTCTCCCCCTCACCATAACAATACTTACACTAATAAACTTATATTTCTACATACGCTAACATACTCAATCTCGAT
AACAAATTTTCTTACATCCAACAACAAAAATTAGCTGACAACTAAAACATCTAAAACCAATACCCTTCTACCTCCCC

TCATAATTTCTTCTCCTCCTACTACCAATCACACCACTAATGCTAATGACCTAGAAAATTTAGGTTAATAAGACCAAGA
GCCTTCAAAGCCCCTAGTAAGTAAATTTTACTTAATTTCTGCACCGCTCTAAGGACTGCAAACTATACCTTGCATCAATT
GAACGCAAATCAACTACTTTAATTAAGCTAAGCCCTCCCTAGACTGATGGGACTTTAACCACAAAAATTTAGTTAACAG
CTAAATAACCTAATCAACTGGCTTCAATCTAATTCTCCCGCTTTAAGGGGAAAAAAGGCGGGAGAAGCCCCGGCAGCA
TTGAAGCTGCTTCTTTGAATTTGCAGTTCAACATGACAGTTACCTCAGGGCTGGTAAAAAGAGGGGGTCACCTCTCTGT
CTTTAGATTTACAGTCTAATGCTTACTCAGCCATTTTACCTTCTACTTATGTTTATAAATCGCTGACTATTTTCAACTAA
CCATAAAGATATTGGTACACTGTATTTAATATTTGGTGCATGAGCCGGAGCAACAGGAACAGCCTTAAGTCTTCTAATTC
GAGCTGAGCTGGGCCAACAGGAAGCCTAATAGAAGACGACCATGTTTACAATGTTATTGTTACCTCTCACGCATTTATT
ATAATCTTCTTCATGGTCATACCAATTATAATTGGCGGCTTCGGGAATTGATTAGTGCTCTAATAATCGGCGCCCCGA
TATAGCTTTTCTCGTATAAATAATATAAGCTTCTGACTCCTCCCCCATCCCTTCTTCTCCTACTTGCCTCTTCAACTC
TAGAGGCTGGTGTGGAACTGGCTGAACAGTCTACCCCTCCCCTGGCAGGAAATATATCACACCCTGGAGCCTCTGTAGAC
TTAACTATCTTTTCACTGCATCTAGCGGTATTTCTCTATTCTAGGGCTATTAACCTTTATTACAACAATTATTAATAT
AAAACCACAGCCACAACCCAATATCAAACACCCCTATTTGTATGATCCGTACTCATTACAGCAATCCTTCTACTTCTCT
CCCTCCCAGTCTAGCTGCTGGAATTACTATACTATTAACCGACCGTAACCTTAACACCACTTTTCTTCGACCCTGCTGGT
GGTGGTGACCCCATTTCTATATCAACACCTATTTTGATTTTTTGGTCACCCCGAAGTTTATATCCTTATTTTACCAGGGTT
TGGGATAATCTCACATATTGTAACATATTACTCTAACAAAAAAGAGCCATTTGGGTATATAGGAATAGTTTGAGCCATAA
TATCCATTGGCTTCTAGGCTTTATCGTATGAGCTCACCATATATTCACAGTAGGAATAGATGTGGATACACGCGCGTAT
TTTACATCAGCTACCATAATCATTGCCATCCCCACTGGGGTCAAAGTATTTAGCTGGTTGGCTACACTACATGGAGGCAA
CATCAAGTGATCTCCTGCAATACTATGAGCCCTGGGTTTTATTTTTCTTCTTCACTGTGGGTGGATTGACAGGAGTTGTAT
TAGCTAACTCATCATTAGATATTGTCTACATGATACATACTACGTGGTAGCCCATTTCCATTACGTCTTATCAATAGGA
GCAGTGTGGCCATTATAGGAGGCTTCATTCACTGATTTCCATTATTTTTCAGGCTATACACTCGACCAAACCTTACGCTAA
AATCCACTTTACTATTATATTTGTAGGTGTAAACATAACCTTCTTTCCACAACACTTTCTCGGCCTATCTGGAATACCTC
GACGGTACTCAGACTACCCCGATGCATATACTGCATGAAATATTATCTCATCCGTAGGCTCATTTATTTCACTGGCAGCA
GTCATTCTAATAATCTTCATAATTTGAGAAGCCTTCTCTTCAAAACGAAAAGTTCTAACCATCGAACAAATATCTACCAA
CCTAGAGTGGTTATATGGCTGCCCCCTCCTTATCACACATTTGAAGAGGCTACCTACGGTAAAACTCCTAAACGAAAAA
GGAAGGATTTGAACCCCCAAAAATTGGTTTCAAGCCAATCCCATAGACCCTATGACTTTTTCTATGAGATATTAGTAAAA
TAAATTACATAACTTTGTCAAAGTTAAATTATAAGACCAAAACCTATATATCTTAGTGGCAACACCAGCTCAACTAGGCC
TACAAAACGCTACATCCCCATTATAGAAGAACTCATTGCCTTCCACGACCACACCCTTATAATTATTTTCTAATTAGT
TCACTAGTCTTATACATTATCTCTTTAATGCTTACCACAAAACCTAATCATACAGCACTATAAATGCCCAAGAAATCGA
AACAATTTGAACTATCCTACCCGCAATTATTCTTATTATAATTGCCCTTCCATCCCTACGCATTTTATATATAACAGATG
AATTTAATAAGCCCTACTTAACCCCTTAAGCAATTGGCCACCAATGGTATTGAAGCTATGAGTACTCAGACTATGAGGAC
CTATTCTTTGATTCTACATTATGCCAACCTACTACCTCCAACCAGGCGAATTCCGACTTCTTGAAGTGGAACAATCGAAC
AACCTTACCAATAGAAGCAGATATTCGATATCTAATCTCATCACAAGATGTATTACACTCATGAACCGTCCCATCATTAG
GCGTTAAAGCAGATGCAATTCAGGCCGATTGAACCAAGCCATACTAGCCTCAATACGACCAGGACTATTTTACGGGCAA
TGCTCAGAAATTTGCGGGTCAAATCACAGCTTTATACCTATTGTCTAGAAATTTATCTATTTCCAAGATTTTCAAGTATG
GGCTTCATACTTATATATTGTATCACTGTAAAGCTACTTAGCATTAACCTTTTAAAGTTAAAGACAGAGAGAATCTCTCTA
CAGTGAGTGCCACAACCTAGACATCTCACCATGACCAATGGTGACTTTATCAATGATTTTAAACCCTGTTTTATGCTATACA
ACTAAAAATATTAATTTTATTTTCCACACTACCCCACTATCAAAATTAACAAAAATCAAAACCAAAAAACAACCTGAG
AACTAAAAATGAACCAAAATCTATTTGCCTCTTTCAATATAACCAATAATCCTGGGAATCCCCCTAGTAGCACTATTTATTT
TATTTCCCACTATACTAATTACACCCTCTAACAACCTAAACAATAACCGATCCTCCTCCCTTCAACAATGACTAATCCAA
CTTGTAATAAAACAAATAATGACAAACCATACCACAAAGGACAAACCTGATCCCTTATACTTTTAAACCCTAATTACCTT
TATTACCCTGAACAAACCTTCTCGGAATTACACCCTACGCATTTACACCTACCACACAACCTATCATTAAATCTAGGCATGG
CAATTTCCCTATGAATAGCAACCGTACTTATAGGACTTCGATTTAAACAAAAGCAACCCCTTGCTCACCTCTTACCTCAA
GGAACACCTATCCCACTCATCCCTATACTTATTATTGAGACTATCAGCCTCTTCATTCAACCTATGGCACTAGCCGT
ACGATTAACAGCCAACATCACAGCAGGCCACCTATTAATGCATCTACTAGGAGACACTATATTAATTCTTCTTTCTATTT
ACCTCTCTTCTCCGTAAATCACCGTTATCGTTATTATTCTATTAATTACCCTAGAACTAGGCGTAGCCCTAATCCAAGCC
TATGTATTCACACTCCTAGTAAGTCTCTATTTGCACAATAACTCATAATAATTTCCACCCAATTATCTATAATGACACA
CCAAACACACGCTTATCATATAGTAAATCCAAGCCCCTGACCACTAACAGGAGCTCTATCAGCTTTCCTATTAACCTCCG

GCCTAGTTATATGGTTCCATTTCTACTACACCCCCCTATTTAATGCAGGCTTACTAGCCAGTGCAATAACAATAGCTCAA
TGATGACGCGACGTAGTGCGAGAAAGTACATATCAAGGCCACCACACTATGCCCGTTCAAAAAGGCCTTCGGTATGGAAT
AATCCTATTTATTATTTTTCAGAAGTTTTCTTCTCGCCGGATTCTTCTGGGCATTCTATCACTCAAGCCTAGCTCCAACCC
CTCAAACAGGAGGACATTGACCTCCCACAGGTATCTTTCCCTAAACCCAATAGAAGTACCACTTCTAAATACAGCTGTC
TTACTCGCATCAGGAGTCACAATTACCTGAGCACATCATAGCCTAATAGAAGCCAACCAAGAAAATTCAACCCAAGCATT
GTCCTTAACTATCGCACTAGGAATTTACTTTACCTATCTTCAAGTATCAGAGTACTCTGAAACACCCTTCTCCATCTCCG
ACGGAATTTACGGCTCCACATTCTTTATAGCCACAGGCTTCCATGGTCTTCATGTTATTATCGGAACCACCTTCCCTGCC
ACATGCTATATTCGCCAACAATTATATCACTTTACACCTAATCACCCTTTGGATTGGAAGCCGCCGCATGATACTGACA
TTTTGTAGATGTGGTATGACTATTCCTCTACATCTCTATCTACTGGTGAGGATCCTATTCTCTTAGTATAAACAGTACAA
TTGACTTCCAATTAATAGGCCTTGATAAACCCAAGAGAGAATAATTAACCTTAATATTAACCCTAGCAACCAGTACCCTAC
TAGCCCTACTACTAATTACTATTACATTCTGAATACCACAATTAATAAATACACAGAAAAACACAACCCCTATGAATGC
GGATTTGATCCTACAACCTCCGCCACCTACCCTTTTCTATAAAATTTTCCCTAGTAGCTATCACATTCCCTCCTATTTGA
CCTAGAAATTGCCCTACTCCTACCCTTACCCTGAGCTACCCAAACAGATAACTTAATACTGATAACTAACATAGTCTTTA
CTCTACTTATTATTCTAGCACTAGGACTAGCCTACGAGTGAGCCAAAAGGGATTAGACTGAGTTGATTGTTGATATAGT
TTAAACAAAATAAGTGATTTTCGACTCATTAGACTATGGTAAACCGTATATACCAAAATGCCTTTTATCTATATCAACACC
GCACTGGCATACTCTATATCTCTACTAGGACTGTTAATTTACCGATCCACCTAATATCGTCCCTATTATGCTTAGAAGG
TATGATACTATCACTATTTATTATAATAACAACCATAACCTTAAATATACACCTAATATTAATATACATACTACCCATTA
TTCTACTGGTATTCGCCGCATGTGAAGCTGCAGTAGGCCTAGCCTTATTAATTTAATCTCCAACCTTATATGGCTTAGAC
TACGTACAAAACCTAACTTACTCCAATGCTAAAAATTATTCTCCCGACAATCATAATAATCCCAACAATATATCTATCA
AAAAACCATATAATATGAATCAATACAATAATCTGCAGCCTATCAATTAGCGCCCTAACCCCTTATAATTCTACATATACC
CAACAACCCATGCAACTTATCGCTAACTTTCTTCTCAGACCCATTAACATCACCCTTTTAGCTCTAACACCTGACTAC
TGCCACTAATAATCTTAGCAACACAACAACATATTTACAGCAACTCCCTTCCCCGAAAAAAATTATACACTTCAATACTA
ATTATCCTACAAAATTTCCCTAATTATAACATTGCGAGCTACAGAACTAATCCTATTCTACATTTTATTTGAACTACTCT
AATTCACCCCTAATTATTATTACCCGCTGAGGATACCAGCCAGAGCGCTCAACGCCGGCTCGTACTTTCTATTCTACA
CACTAGCTGGATCACTTCCACTACTCATTACACTCCTTTACTGCCTCAACAACCTTAGGAACCTAAATATCCTAACATG
ACAATTAACACCAAAGAATTACTAACATCTTGAACCTAATAATATTATATGACTAGGATGCACGATAGCCTTCATAGTAAA
AATACCTCTGTACGGCCTTCACCTATGACTCCCCAAAGCCCATGTAGAAGCCCCAATTGCCGGCTCAATAGTACTTGCAG
CAATCCTGTAAAACTAGGTAGCTATGGCATAATACGAATCATTCCCTACCCTCAACCCCTTAACAGAAAAAATAAGCTAC
CCCTTCCTCATCCTATCCCTATGAGGCATGTTAATAACAAGCTCCATCTGCTTACGACAGGCCGACCTAAAATCATTAAAT
TGCCTACTCTTCCATCAGCCATATGGCGCTTATTACCCTGGCTATTCTTATCCAAACCCCTGAAGCCTTACCGGTGCAA
TACTACTAATAGTCGCACATGCATTTACCTCATCCCTACTATTCTGTCTAGCAAACTCCAATTACGAACGTATTCACAGC
CGAACTATAATATTTACCCGAGGCCTCCAAGCACTATTTCCACTCCTAACTCTCTGATGACTTCTAGCAACCTTGCCAA
TCTTGCTCTACCTCCAACCTATTAATCTACTAGGAGAATTATCTACAATTTTAGCCGCCTTCTCCTGATCTAATTTTACCA
TCGCATTTACAGGATTCAATATACTTATCACAGCACTATACTCACTACATATATTACCTCAACACAACGAGGACCATTA
ACACACAGCACCAGCAATGTAAAACCCCTATTTACAGGAGAAAAACACTAATACTAATAACATATAGCACCATCCTTCT
TCTCACTTTAAACCCCAAAAATAATGATAGGCCTAATACCATGTAGCTATAGTTTAATTAACCAATTAGATTGTGAATCTA
ATAATAGAAGACTACAACCTTCTTAACTACCGAGAAAGTATGCAAGAAGTCTAACTCCTGCCTCCAGACTTAATATCCTG
GCCTTCTCAGCTTTTAAAGGATAGTAGTTATCCATTGGTCTTAGGAGCCAAAAATATTGGTGCAACTCCAAATAAAGCA
AAATGCACTTTTCTATTACTCTAATAACACTAATTCCTTACTAGTACCTATCATAACTACCCTAGTTAAATCTAACAAA
AATCCCCTATATCCACACTATGTAAACTAGCCATTATTTATGCCCTTATTACTAGCACCTTGTCTATAACAATATTTAT
TCTCACAGGCCAAGAATCAATAATTTCAAACCTGACATTGAGCGACAATCCAACTATCAAACCTATCACTAACTTTAAAC
TAGACTTCTTCCATAATATTCACGCCTGTAGCACTATTCGTTACCTGATCAATTGTAGAATTCTCAATATGATATATA
AACTCAGACCCAAATATTAATCAATTCCTCAAGTACTTACTTATATTCCTCATCACAATATTAATCTTAATTACCGCTAA
TAACTTATTTCCAACTATTTATCGGGTGAGAAGGAATAGGTATTATATCCTTTCTACTAATTAGCTGGTGGTATGGCCGAA
CAGAAGCTAACACAGCAGCCTTACAAGCAATCTTATATAACCGTATTGGAGATATTGGCCTTATCCTTGCAATAGCATGA
TTTTTTTTTACTCTCAACTCATGAGATTTCCAGCAAATATTTATTCTTAACCCCACTACAACTTTTTTCCCTGATAAG
CCTTCTCCTAGCAGCAACAGGAAAAATCAGCCCAATTTGGTCTCCACCCATGACTTCCCTCCGCCATAGAAGGCCACCC
CAGTCTCAGCACTACTTCACTCCAGTACAATAGTCGTGCGAGGTGTCTTCCCTAATTATCCGCTTTTCATCCTCTAATAGAA

AACAACCTTACTCATCCAAACACTTACCTTATCGCTAGGGGCCATTACCACCCTATTACATCAATCTGTGCTCTAACACA
AAATGACATAAAAAAAATTGTGGCCTTCTCAACCTCAAGCCAACTAGGCCTTATAACAGTAACAGTCGGTATTAACCAAC
CACATTTAGCCTTTCTACACATCTGTACTCATGCCTTTTTCAAAGCTATATTATTTCTATCCGCAGGATCCATTATCCAC
AATCTCAACAACGAACAAGACATTCGAAAAATAGGAGGCCTATTTAAAACCATACCTTTCACTGCCTCCTCCCTCATTAT
TGGCAATCTTGCCTTATAGGAACACCATTTCCTTACAGGTTTCTACTCAAAAGATCTAATTATCGAAACCATCAACACGT
CATATACCAACGCCTGAGCCCTTACAACCTACCCTCGTAGCCACCTCCCTTACAGCAATATACAGTGGCCGCATTATATTT
TTTACCTTAACAGGATACCTTCGCTTTACAACCTCCACCCTAATCAACGAAAATAACAATCTCCTAATAAATCCAATTAG
CCGCCTAGCAGTAGGTAGTATCCTTGCCGGGTTTCTTATTTCTAATTGCTTACCCCTACTATATACCCCAAATAACCA
TACCTTTCCACCTTAAGCTTACAGCTCTAAGTGTAAACCACCCTAGGACTCCTTATAGCAATAGAACTTAACTCTTTAACT
AACAACATAAAACTATATACCCAGTAAAAATTTACTACTTCTCTAATATACTAGGCTTCTACTCAATTACTACTCACCG
CCTCAACCCCATTTAAACCTAACCACAAGTCAAAATTTACCTCCACTTTACTAGACCTATTCTGACTAGAAAAATCTA
TACCAAAATTAACAACACAAACACAAATTTCAATAGCTACAATTACATCAACTCAAAAAGGCCTAATTAACTCTACTTC
TTTACCTTCTTTATTCCACTCATCCTAACGCTTTTTCTTACAATTTAACCCCTCCCGAGTTAACTCAATTGCAATATG
TATACCCATAAAATAACGCCCAACAAGTAACATAAAACAACCTCAAAACACCATAATCATATAAAGCAGCAGCACCTGCAGGAT
CCTCACGAATCAAACCCGGCCCTCACCTCATAAATCATTCAACCTGCCACAGTATCATAATTAACAGCAATCTCCACC
GTTTTATTAGGACTACCCCCAGTAATAATACCACCACTACCTCCATTGCTAAACCCAATACAAGCATCCCTAAAATATC
AACACTCGATACCCATGTCTCAGGGTATTCATCGACTGCTATCGCTGCAGTATAACCAAAAAACAACCATATACCACCCA
AATAAATTAAAAAAACTATAAGCCCTATATAAGATCCACCAAGATACAGTATAAGTACACAACCCACAGCACCACTAAAA
ACCAACACCAACCCCCCATAAATAGGTGAAGGTTTAGAGGAAAATCCCACAAACCCTGTTACCAATATAACTTTAGCAA
AAATAAAATATACATCACTATTCCCACATGGATTATAACCATGACTAATGATATGAAAAGCCATCGTTGTATTTCAACTA
TAAGAATACTAATGACCTCTCCCGCAAAACACCCCATTAATAAAAATTATTAATAGTTTCATTTATTGATCTGCCACA
CCATCCAACATCTCCTCCTGATGAACTTCGGATCACTTCTAGGCGCCTGCCTAATAATTCAAATTACCACAGGCCTATT
CTTAGCGATACACTATACGCCAGACACCTCAACCGCCTTCTCCTCAGTAGCACATATTACCCGAGATATTAATTACGGTT
GAATAATCCGCCTCTACACGCCAATGGTGCCTCATATTTTTGTGTGCTTATTTCTCCACACTGGCCGAGGCCTCTAC
TACGGATCTTTTCTCTTTCTAAACACCTGAAATATTGGTACAATCCTATTATTAATAACAATAGCCACAGCCTTTATAGG
CTATGTCTTACCGTGAGGCCAAATATCATTCTGAGGAGCCACAGTTATTACAAATCTTCTATCAGCCATCCCTTATACCG
GACATAACCTTGTACAATGAATCTGAGGTGGCTTTTCAGTAGATAAACCACCCCTCACACGATTCTTTACCTTTCACTTT
ATTTTACCTTTTATTATCACAGCCCTAGCAACTATTACCTTTTTATTTCTACATGAAACAGGCTCAAATAATCCATCAGG
AATAGCATCTAGCCCGATAAAATTATATTCCATCCCTACTACACAACCAAAGATATTTTTGGATTAAACCTTCTTCTCC
TACTCCTTACAAGCCTAACCCTATTTACCCCGACCTTTTAACTGACCCAGATAACTACACACTAGCCAACCCCTTAAT
ACTCCACCCCATATTAAGCCAGAGTGATACTTTCTATTTCGCATACACAATTTTACGATCTATTCCAAATAAACTAGGAGG
TGTTCTAGCTCTTCTATTATCTATTATAATCCTAACAATTATCCCTGCCACTCACCTATCCAAACAACAAAGTATAATAT
TCCGACCAATCACCCAAATCCTATTCTGAACCTAGCAGCCGATCTACTTACACTTACATGAATTGGAGGCCAACCGTA
GAATACCCCTTTGAAGCCATTGGCCAAACCGCATCTATTGCTTACTTCCTTATTATTACTCTAATTCCTCTATCAGCCCT
AACTGAAAATAAGCTACTTAAATGATAACGTCCCTGTAGTATATCCAATTACCCCGCCTTGTAACCCGAAAAGGAGGC
ACGCTAACTCCCCAGGACAATCAGGAAAAGAATAATTAATTCTACCATCAACACCCAAAGCTGAAATTCTACATTTAACT
ATTCCCTGAGCATAGTAAATAATAGATATATATACTATTTGATGGCTTACCCATAAAGTACTCTAAAGGCATACACAATT
TTTTTTCCCTATGTAATTAGTGCATTATTGCTTATCCCCATGAATAATACATAGTACTACACATGCTTAATTATACATA
GCATATAGAGTCAAAACGTGCATACTACGTCACAGAACATGCTTACAAGCAGGAACTATTATTTTACAGCGGACTATAGC
GCATTAACCTCAAAACCCACGCTCCAATTACAGTCACCACGGATATCGTACAGCCCATAGAATCATTGATAGTACATCTGC
ACATCAAAATGATTAAACGGACATAGCGCATCCTATTTCTTCAGTCCTTCTAACCATGGATATCCCCTTGATATTTGGTC
TCTTAATCTACCAACCTCCGTGAAACCACCAACCCGCCCACTTCTGCGGCTCTTCTCGCTCCGGGCCCATATAGACAGGG
CTTGTTTATACTGAACTATACCTGGCATTGTTTCTACCTCAGGGCCATCTCACTAAGACCGTGTCCACGTTCCCTCTT
AAATAAGACATCACGATGGTGTGGCGCTATCACCTCTTAACCGCGTCACTGGATGCATAGGGTGCCTGGGTGGGGAAAG
GGGGGGGAGGGAATCCTCAGCATTGCCGTAGGCTCCGGTAGAAGTCCCGCCCCCGTCTGTGGGACCTGTCTGTGTA
CTGCCATGCCTTATGCTATCATCGCACCTAAATTGAATGTCTTGCCCCCAACCCGCCCACTAGGTGTTATTAGTCAAT
GGTCTCAGGACATAATAGATAATTACAGCACTATACCCAAAAATTTAACACACCAACCCCTTCAACCTTATCCCATATTTT
CTGGGGCGAGCGTCGTAAAGAGATATTTTACCATTAGCAACTAAATTTTATAAATCAACAAATTAATAGAAATTTTACAC

CCCATAGTACACACCTATTTAAATAATACTAGATACAATTACCTACCGCATACCTCAATCTCTATCCCTCAGAAAACAGA
TCTATACCACATAATCGCCCTGACCTCAAATCTATATAACCTACTGCTATATCCTTATACTGGGCCCTAATTAA

Or, with a little more effort, we could get the reverse sequence in fasta format:

```
[123]: for i in SeqIO.parse("Cebus_albifrons_NC_002763.1.gb", "genbank"):
        header = i.description
        print((i.reverse_complement(header, "", "").format("fasta")))
```

```
>Cebus albifrons mitochondrion, complete genome
TTAATTAGGGCCAGTATAAGGATATAGCAGTAGGTTATATAGATTTGAGGTCAGGGCGA
TTATGTGGTATAGATCTGTTTTCTGAGGGATAGAGATTGAGGTATGCGGTAGGTAATTGT
ATCTAGTATTATTTTAATAGGTGTGTACTATGGGGTGTAATAATTCTATTTAATTTGTTGA
TTTATAAAATTTAGTTGCTAATGGTAAATATCTCTTACGACGCTCGCCCCAGAAAATA
TGGGATAAGGTTGAAGGGTTGGTGTGTTAAATTTTGGGTATAGTGCTGTAATTATCTAT
TATGTCCTGAGACCATGACTGAATAACACCTAGTGGGCGGGTTGGGGGCAAGACATTC
AATTTAGGTGCGATGATAGCATAAGGCATGGCAGTACACAGACAGGTCCCACAGGACGGG
GGGCGGGACTTCTACCGGAGCCTACGGCAATGCTGAGGATTTCCCTCCCCCCCCCTTCC
CCACCCAGGCACCCTATGCATCCAGTGACGCGGTTAAGAGGGTGATAGCGCCACACCATC
GTGATGTCTTATTTAAGAGGAACGTGGACACGGTCTTAGTGAGATGGCCCTGAGGTAGGA
ACCAAATGCCAGGTATAGTTTCAGTATAACCAAGCCCTGTCTATATGGGCCCGAGCGAG
AAGAGCCGCAGAAGTGGGCGGGTTGGTGGTTTCACGGAGGTTGGTAGATTAAGAGACCAA
ATATACAAGGGGATATCCATGGTTAGAAGGACTGAAGAAATAGGATGCGCTATGTCCGGT
TAATCATTTGATGTGCAGATGTACTATCAATGATTCTATGGGCTGTACGATATCCGTGGT
GACTGTAATTGGAGCGTGGGTTTTGAGTTTAATGCGCTATAGTCCGCTGTGAAATAATAGT
TCCTGCTTGTAAAGCATGTTCTGTGACGTAGTATGCACGTTTTGACTCTATGTGCTATGTA
TAATTAAGCATGTGTAGTACTATGTATTATTCATGGGGATAAGCAATAATGCACTAATTA
CATAGGGGAAAAAAATTTGTGTATGCCTTTAGAGTACTTTATGGGTAAGCCATCAAATAG
TATATATATCTATTATTTACTATGCTCAGGGAATAGTTTAATGTAGAATTTACAGCTTTGG
GTGTTGATGGTAGAATTAATTATTTCTTTCCCTGATTGTCCTGGGGAGTTAGCGTGCCTCC
TTTTCCGGTTTACAAGGCCGGGTAATTGGATATACTACAAGGACGTTATCATTTAAGTA
GCTTATTTTCAGTTAGGGCTGATAGAGGAATTAGAGTAATAATAAGGAAGTAAGCAATAG
ATGCGGTTTGGCCAATGGCTTCAAAGGGGTATTCTACTGGTTGGCCTCCAATTCATGTAA
GTGTAAGTAGATCGGCTGCTAGGGTTCAGAATAGGATTTGGGTGATTGGTCGGAATATTA
TACTTTGTTGTTTGGATAGGTGAGTGGCAGGGATAATTGTTAGGATTATAATAGATAATA
GAAGAGCTAGAACACCTCCTAGTTTATTTGGAATAGATCGTAAAATTGTGTATGCGAATA
GAAAGTATCACTCTGGCTTAATATGGGGTGGAGTATTAAGGGGGTTGGCTAGTGTGTAGT
TATCTGGGTCAGTTAAAAGGTCGGGGGTAAATAGGGTTAGGCTTGTAAGGAGTAGGAGAA
GAAGGGTTAATCCAAAAATATCTTTGGTTGTGTAGTAGGGATGGAATATAATTTTATCGG
GGCTAGATGCTATTCTGATGGATTATTTGAGCCTGTTTCATGTAGAAATAAAAAGGTGAA
TAGTTGCTAGGGCTGTGATAATAAAAAGGTAAAAATAAAGTGAAAGGTAAAGAATCGTGTGA
GGGTGGGTTTATCTACTGAAAAGCCACCTCAGATTCATTGTACAAGGTTATGTCCGGTAT
AGGGGATGGCTGATAGAAGATTTGTAATAACTGTGGCTCCTCAGAATGATATTTGGCCTC
ACGGTAAGACATAGCCTATAAAGGCTGTGGCTATTGTTATTAATAATAGGATTGTACCAA
TATTTACAGGTGTTTAGAAAGAGAAAAAGATCCGTAGTAGAGGCTCGGCCAGTGTGGAGAA
ATAAGCACACAAAAAATATGGAGGCACCATTTGGCGTGTAGGAGGCGGATTATTCAACCGT
AATTAATATCTCGGGTAATATGTGCTACTGAGGAGAAGCGGTTGAGGTGTCTGGCGTAT
AGTGTATCGCTAAGAATAGGCCTGTGGTAATTTGAATTATTAGGCAGGCGCCTAGAAGTG
```

ATCCGAAGTTTCATCAGGAGGAGATGTTGGATGGTGTGGGCAGATCAATAAATGAACTAT
TAATAATTTTTATTAATGGGTGTGTTTTGCGGGGAGAGGTCATTAGTATTCTTATAGTTG
AAATACAACGATGGCTTTTCATATCATTAGTCATGGTTATAATCCATGTGGGAATAGTGA
TGTATATTTTATTTTGTCTAAGTATTATATTGGTAACAGGGTTTGTGGGATTTTCCTCTA
AACCTTCACCTATTTATGGGGGGTGGTGTGGTTTTTAGTGGTGCTGTGGGTTGTGTAC
TTATACTGTATCTTGGTGGATCTTATATAGGGCTTATAGTTTTTTTAATTTATTTGGGTG
GTATAATGGTTGTTTTTGGTTATACTGCAGCGATAGCAGTCGATGAATACCCTGAGACAT
GGGTATCGAGTGTGATATTTTAGGGATGCTTGTATTGGGTTTAGCAATGGAGGTAGTGG
TGGTATTATTACTGGGGGGTAGTCCTAATAAAACGGTGAGATTGCTGTAAATTATGATA
CTGTGGCAGGTTGAATGATTTATGAGGGTGAGGGGCCGGGTTTGATTTCGTGAGGATCCTG
CAGGTGCTGCTGCTTTATATGATTATGGTGTGTTGAGTTGTTTTAGTTACTTGTGGGCGT
TATTTATGGGTATACATATTGCAATTGAGTTAACTCGGGGAGGGGTTAAATTGTAAGAA
AAAGCGTTAGGATGAGTGGAATAAAGAAGGTAAGAAGTAGAGTTTAATTAGGCCTTTTT
GAGTTGATGTAATTGTAGCTATTGAAATTTGTGTTTGTGTTGTTAATTTTGGTATAGATT
TTTCTAGTCAGAAATAGGTCTAGTAAAGTGGAGGTGAAATTTTGACTTGTGGTTAGGTTTA
AATGGGGGTTGAGGCGGTGAGTAGTAATTGAGTAGAAGCCTAGTATATTAGAGAAGTAGT
AAATTTTTACTGGGGTATATAGTTTTATGTTGTTAGTTAAAGAGTTAAGTTCTATTGCTA
TAAGGAGTCCTAGGGTGGTTACACTTAGAGCTGTAAGCTTAAGGTGGAAGGTATGGTTA
TTTGGGGGTATATAGTAGGGGGTAAGCAATTAGAAATAAGAAACCCGGCAAGGATACTAC
CTACTGCTAGGCGGCTAATTGGATTTATTAGGAGATTGTTATTTTCGTTGATTAGGGTGG
AAGTTGTAAAGCGAGGGTATCCTGTTAAGGTAAAAAATATAATGCGGCCACTGTATATTG
CTGTAAGGGAGGTGGCTACGAGGGTAGTTGTAAGGGCTCAGGCGTTGGTATATGACGTGT
TGATGGTTTTCGATAAATTAGATCTTTTGAGTAGAAACCTGTAAGAAATGGTGTTCCTATAA
GTGCAAGATTGCCAATAATGAGGGAGGAGGCAGTGAAAGGTATGGTTTTAAATAGGCCTC
CTATTTTTCGAATGTCTTGTTCGTTGTTGAGATTGTGGATAATGGATCCTGCGGATAGAA
ATAATATAGCTTTGAAAAAGGCATGAGTACAGATGTGTAGAAAGGCTAAATGTGGTTGGT
TAATACCGACTGTTACTGTTATAAGGCCTAGTTGGCTTGAGGTTGAGAAGGCCACAATTT
TTTTTATGTCATTTTGTGTTAGAGCACAGATTGATGTGAATAGGGTGGTAATGGCCCCTA
GCGATAAGGTAAGTGTGTTGGATGAGTAAGTTGTTTTCTATTAGAGGATGAAAGCGGATAA
TTAGGAAGACACCTGCGACGACTATTGTACTGGAGTGAAGTAGTGCTGAGACTGGGGTGG
GTCCTTCTATGGCGAGGGAAGTCATGGGTGGAGACCAAATTGGGCTGATTTTCCTGTTG
CTGCTAGGAGAAGGCTTATCAGGGGAAAAAGTTTGTAGTGGGGTTAAGAATAAATATTT
GCTGGAATCTCATGAGTTGGAGTGTAAAAAAATCATGCTATTGCAAGGATAAGGCCAA
TATCTCCAATACGGTTATATAAGATTGCTTGTAAAGCTGCTGTGTTAGCTTCTGTTGGC
CATACCACAGCTAATTAGTAGAAAGGATATAATACCTATTCCTTCTACCCGATAAATA
GTTGGAATAAGTTATTAGCGGTAATTAAGATTAATATTGTGATGAGGAATATAAGTAAGT
ACTTGAGGAATTGATTAATATTTGGGTCTGAGTTTATATATCATATTGAGAATTCTACAA
TTGATCAGGTAACGAATAGTGCTACAGGCGTGAATATTATGGAGAAGAAGTCTAGTTTAA
AGTTTAGTGATAGTTTGATAGTTTGGATTGTCGCTCAATGTCAGTTTGAAATTATTGATT
CTTGGCCTGTGAGAATAAATATTGTTATAGACAAGGTGCTAGTAATAAGGGCATAAATAA
TGGCTAGTTTTACATAGTGTGGATATAGGGGATTTTTGTTAGATTTAACTAGGGTAGTTA
TGATAGGTACTAGTAAGGGAATTAGTGTTATTAGAGTAATAGAAAAGTGCATTTTGCTTT
TATTTGGAGTTGCACCAATATTTTGGCTCCTAAGACCAATGGATAACTACTATCCTTTA
AAAGCTGAGAAGGCCAGGATATTAAGTCTGGAGGCAGGAGTTAGCAGTTCTTGCACTT
TCTCGGTAGTTAAGAAGTTGTAGTCTTCTATTATTAGATTCAATCTAATGTTTAAATT
AAACTATAGCTACATGGTATTAGGCCTATCATTATTTGGGGTTTAAAGTGAGAAGAAGG
ATTGGTGCTATATGTATTAGTATTAGTGTATTTTCTCGTGAAATAGGGGTTTTACATTG
CTGGTGCTGTGTGTTAATGGTCCTCGTTGTGTTGAGGTAAATATATGTAGTGAGTATAGT

GCTGTGATAAGTATATTGAATCCTGTAAATGCGATGGTAAAAATTAGATCAGGAGAAGGCG
GCTAAAAATTGTAGATAATTCTCCTAGTAGATTAATAGTTGGAGGTAGAGCAAGATTGGCA
AGGTTTGTCTAGAAATCATCAGAGAGTTAGGAGTGAAATAGTGCTTGGAGGCCTCGGGTA
AATATTATAGTTTCGGCTGTGAATACGTTTCGTAATTGGAGTTTGCTAGACAGAATAGTAGG
GATGAGGTAATATGCATGTGCGACTATTAGTAGTATTGCACCGGTAAGGCTTCAGGGGGTT
TGGATAAGAATAGCCAGGGTAATAAGCGCCATATGGCTGATGGAAGAGTAGGCAATTAAT
GATTTTAGGTGCGCCTGTCTGAAGCAGATGGAGCTTGTTATTACCATGCCTCATAGGGAT
AGGATGAGGAAGGGGTAGCTTATTTTTCTGTAAAGGGTTGAGGGTAGGAATGATTCGT
ATTATGCCATAGCTACCTAGTTTTAGCAGGATTGCTGCAAGTACTATTGAGCCGGCAATT
GGGGCTTCTACATGGGCTTTGGGGAGTCATAGGTGAAGGCCGTACAGAGGTATTTTTACT
ATGAAGGCTATCGTGCATCCTAGTCATATAATATTATTAGTTCAAGATGTTAGTAATTCT
TTGGTGTTAATTGTCATTGTTAGGATATTTAGGGTTCCTAAGTTGTTGAGGCAGTAAAGG
AGTGTAAATGAGTAGTGGAAGTGATCCAGCTAGTGTGTAGAATAGAAAGTACGAGCCGGCG
TTGAGGCGCTCTGGCTGGTATCCTCAGCGGTAATAATAATTAGGGTTGGAATTAGAGTA
GTTTCAAATAAAAAGTAGAATAGGATTAGTTCTGTAGCTGCAATGTTATAATTAGGGAA
ATTTGTAGGATAATTAGTATTGAAGTGATAATTTTTTCGGGAAGGGAGTTGCTGTAA
ATATGTTGTTGTGTTGCTAAGATTATTAGTGGCAGTAGTCAGGTTGTTAGAGCTAAAAAGT
GGTGATGTTAATGGGTCTGAGAAGAAAGTTAGCGATAAGTTGCATGGGTTGTTGGGTATA
TGTAAGATTATAAGGGTTAGGGCGCTAATTGATAGGCTGCAGATTATTGTATTGATTCAT
ATTATATGGTTTTTGATAGATATATTGTTGGGATTATTATGATTGTCGGGAGAATAATT
TTTAGCATTGGAGTAAGTTTAGGTTTTGTACGTAGTCTAAGCCATATAAGTTGGAGATTA
AAATTAATAAGGCTAGGCCTACTGCAGCTTCACATGCGGCGAATACCAGTAGAATAATGG
GTAGTATGTATATTAATATTAGGTGTATATTTAAGGTTATGGTTGTTATTATAATAAATA
GTGATAGTATCATACCTTCTAAGCATAATAGGGACGATATTAGGTGGGATCGGTAAATTA
ACAGTCCTAGTAGAGATATAGAGTATGCCAGTGCGGTGTTGATATAGATAAAAAGGCATTT
TGGTATATACGGTTTACCATAGTCTAATGAGTCGAAATCACTTATTTTTGTTTAACTATA
TACCAAATCAACTCAGTCTAATCCCTTTTGGGTCCACTCGTAGGCTAGTCCTAGTGCTAG
AATAATAAGTAGAGTAAAGACTATGTTAGTTATCAGTATTAAGTTATCTGTTTGGGTAGC
TCACGGTAAGGGTAGGAGTAGGGCAATTTCTAGGTCAAATAGGAGGAATGTGATAGCTAC
TAGGAAAAATTTATAGAAAAGGGTAGGTGGGCGGAGGTTGTAGGATCAAATCCGCATTC
ATAGGGGTTGTGTTTTCTGTGTATTTATTTAATTGTGGTATTCAGAATGTAATAGTAAT
TAGTAGTAGGGCTAGTAGGGTACTGGTTGCTAGGGTTAATATTAAGTTAATTATTCTCTC
TTGGGTTTATCAAGGCCTATTAATTGGAAGTCAATTGTACTGTTTATACTAAGAGAATAG
GATCCTCACCAGTAGATAGAGATGTAGAGGAATAGTCATACCACATCTACAAAATGTCAG
TATCATGCGGCGGCTTCAAATCCAAAGTGGTGATTAGGTGTAAAGTGATATAATTGTTGG
CGAATATAGCATGTGGCAAGGAAGGTGGTTCCGATAATAACATGAAGACCATGGAAGCCT
GTGGCTATAAAGAATGTGGAGCCGTAAATTCCGTGCGAGATGGAGAAGGGTGTTCAGAG
TACTCTGATACTTGAAGATAGGTAAAGTAAATTCCTAGTGCAGATAGTTAAGGACAATGCT
TGGGTTGAATTTCTTGGTTGGCTTCTATTAGGCTATGATGTGCTCAGGTAATTGTGACT
CCTGATGCGAGTAAGACAGCTGTATTTAGAAGTGGTACTTCTATTGGGTTTAGGGGAAAG
ATACCTGTGGGAGGTCAATGTCCTCCTGTTTGAAGGGTTGGAGCTAGGCTTGAGTGATAG
AATGCCCAGAAGAATCCGGCGAAGAAGAAACTTCTGAAATAATAAATAGGATTATTCCA
TACCGAAGGCCTTTTTGAACGGGCATAGTGTGGTGGCCTTGATATGTACTTTCTCGCACT
ACGTCGCGTCATCATTGAGCTATTGTTATTGCACTGGCTAGTAAGCCTGCATTAAATAGG
GGGGTGTAGTAGAAATGGAACCATATAACTAGGCCGGAGGTTAATAGGAAAGCTGATAGA
GCTCCTGTTAGTGGTCAGGGGCTTGGAATTTACTATATGATAAGCGTGTGTTTGGTGTGTC
ATTATAGATAAATTGGGTGGGAAATTATTATGAGTTATTGTGCAAATAGAGACTTACTAGG
AGTGTGAATACATAGGCTTGGATTAGGGCTACGCCTAGTTCTAGGGTAATTAATAGAATA

ATAACGATAACGGTGATTACGGAGGAAGAGAGGTAAATAGAAAAGAAGAATTAATATAGTG
TCTCCTAGTAGATGCATTAATAGGTGGCCTGCTGTGATGTTGGCTGTTAATCGTACGGCT
AGTGCCATAGGTTGAATGAAGAGGCTGATAGTCTCAATAATAAAGTATAGGGATGAGT
GGGATAGGTGTTCTTGAGGTAAGAGGTGAGCAAGGGTTGCTTTTGTTTTAAATCGAAGT
CCTATAAGTACGGTTGCTATTCATAGGGGAATTGCCATGCCTAGATTTAATGATAGTTGT
GTGGTAGGTGTAAATGCGTAGGGTGTAATCCGAGAAGGTTGTTTCAGGGTAATAAAGGTA
ATTAGGGTTAAAAGTATAAGGGATCAGGTTTGTCTTTGGTGGTATGGTTTGTCAATTATT
TGTTTTAGTACAAGTTGGATTAGTCATTGTTGAAGGGAGGAGGATCGGTTATTGTTTAGG
TTGTTAGAGGGTGTAATTAGTATAGTGGGAAATAAAATAAATAGTGCTACTAGGGGGATT
CCCAGGATTATTGGTATATTGAAAGAGGCCAAATAGATTTTGGTTCATTTTAGTTCTCAGG
TTGTTTTTTGGTTTGAATTTTAGTTAATTTTGATAGTGGGGTAGTGTGGAATAAATT
TTAATATTTTAGTTGTATAGCATAAAACAGGGTTAAAATCATTGATAAAGTCACCATTG
GTCATGGTGAGATGTCTAGTTGTGGCACTCACTGTAGAGAGATTCTCTGTCTTTAACT
TAAAAGGTTAATGCTAAGTAGCTTTACAGTGATACAATATATAAGTATGAAGCCCATACT
TCGAAATCTTGAAAATAGATAAATTCTAGGACAATAGGTATAAAGCTGTGATTTGACCCG
CAAATTTCTGAGCATTGCCCGTAAAATAGTCTGGTCGTATTGAGGCTAGTATGGCTTGG
TTCAATCGGCCCTGGAATTGCATCTGCTTTAACGCCTAATGATGGGACGGTTCATGAGTGT
AATACATCTTGTGATGAGATTAGATATCGAATATCTGCTTCTATTGGTAAGGTTGTTTGA
TTGTCCACTTCAAGAAGTCGGAATTCGCCTGGTTGGAGGTAGTAGGTTGGCATAATGTAG
GAATCAAAGAATAGGTCCTCATAGTCTGAGTACTCATAGCTTCAATACCATTGGTGGCCA
ATTGCTTTAAGGGTTAAGTAGGGCTTATTAATTCATCTGTTATATATAAAATGCGTAGG
GATGGAAGGGCAATTATAATAAGAATAATTGCGGGTAGGATAGTTCAAATTGTTTCGATT
TCTTGGGCATTTATAGTCTGGTATGAGTTAGTTTGTGGTAAGCATTAAAGAGATAATG
TATAGGACTAGTGAACATAATTAGGAAAATAATTATAAGGGTGTGGTCGTGGAAGGCAATG
AGTTCTTCTATAATGGGGGATGTAGCGTTTTGTAGGCCTAGTTGAGCTGGTGTGGCACT
AAGATATATAGGTTTTGGTCTTATAATTTAACTTTGACAAAGTTATGTAATTTATTTTAC
TAATATCTCATAGAAAAAGTCATAGGGTCTATGGGATTGGCTTGAACCAATTTTGGGG
GTTCAAATCCTTCCTTTTTTCGTTTAGGAGTTTTACCGTAGGTAGCCTCTTCAAATGTGTG
ATAAGGAGGGGGCAGCCATATAACCACTCTAGGTTGGTAGATATTTGTTTCGATGGTTAG
AACTTTTCGTTTTGAAGAGAAGGCTTCTCAAATTATGAAGATTATTAGAATGACTGCTGC
CAGTGAAATAAATGAGCCTACGGATGAGATAATATTTTCATGCAGTATATGCATCGGGGTA
GTCTGAGTACCGTCGAGGTATTCCAGATAGGCCGAGAAAGTGTGTGGAAGAAGGTTAT
GTTTACACCTACAAATATAATAGTAAAGTGGATTTTAGCGTAAGTTTGGTCGAGTGTATA
GCCTGAAAATAATGGAAATCAGTGAATGAAGCCTCCTATAATGGCAAACACTGCTCCTAT
TGATAAGACGTAATGGAAATGGGCTACCACGTAGTATGTATCATGTAGGACAATATCTAA
TGATGAGTTAGCTAATACAACTCCTGTCAATCCACCCACAGTGAAGAGAAAAATAAAACC
CAGGGCTCATAGTATTGCAGGAGATCACTTGATGTTGCCTCCATGTAGTGTAGCCAACCA
GCTAAATACTTTGACCCAGTGGGGATGGCAATGATTATGGTAGCTGATGTAAAAACGC
GCGTGTATCCACATCTATTCCCTACTGTGAATATATGGTGAGCTCATACGATAAAGCCTAG
GAAGCCAATGGATATTATGGCTCAAACCTATTCCTATATACCCAAATGGCTCTTTTTTGT
AGAGTAATATGTTACAATATGTGAGATTATCCCAAACCTGGTAAAATAAGGATATAAAC
TTCGGGGTGACCAAAAAATCAAAATAGGTGTTGATATAGAATGGGGTCACCACCACCAGC
AGGGTCAAGAAAAGTGGTGTAAAGGTTACGGTCGGTTAATAGTATAGTAATTCAGCAGC
TAGGACTGGGAGGGAGAGAAGTAGAAGGATTGCTGTAATGAGTACGGATCATACAAATAG
GGGTGTTTGATATTGGGTTGTGGCTGGTGGTTTTATATTAATAATTGTTGTAATAAAGTT
AATAGCCCCTAGAATAGAGGAAATACCCGCTAGATGCAGTGAAAAGATAGTTAAGTCTAC
AGAGGCTCCAGGGTGTGATATATTTCTGCCAGGGGAGGGTAGACTGTTTCAGCCAGTTCC
AACACCAGCCTCTAGAGTTGAAGAGGCAAGTAGGAGAAGAAGGGATGGGGGAGGAGTCA

GAAGCTTATATTATTTATACGAGGAAAAGCTATATCGGGGGCGCCGATTATTAGAGGCAC
TAATCAATTCCCGAAGCCGCCAATTATAATTGGTATGACCATGAAGAAGATTATAATAAA
TGCGTGAGAGGTAACAATAACATTGTAAACATGGTCGTCTTCTATTAGGCTTCCTGGTTG
GCCCAGCTCAGCTCGAATTAGAAGACTTAAGGCTGTTCTGTTGCTCCGGCTCATGCACC
AAATATTAATAACAGTGTAACCAATATCTTTATGGTTAGTTGAAAATAGTCAGCGATTTAT
GAACATAAGTAGAAGGTAATAATGGCTGAGTAAGCATTAGACTGTAAATCTAAAGACAGAG
AAGTGACCCCTCTTTTTACCAGCCCTGAGGTGAAGTGTGATGTTGAAGTGCATAATCAA
AGAAGCAGCTTCAATGCTGCCGGGGCTTCTCCCGCCTTTTTTTCCCTTAAAGGCGGGAG
AATTAGATTGAAGCCAGTTGATTAGGTTATTTAGCTGTAACTAAATTTTTGTGGGTTAA
AGTCCCATCAGTCTAGGGAGGGCTTAGCTTAATTAAGTAGTTGATTTGCGTTCAATTGA
TGCAAGGTATAGTTTGCAGTCCTTAGAGCGGTGCAGAAATTAAGTAAAATTTACTTACTA
GGGGCTTTGAAGGCTCTTGGTCTTATTAACCTAAATTTCTAGGTCATTAGCATTAGTGGT
GTGATTGGTAGTAGGAGGGAGGAAGAAATTATGAGGGGAGGTAGAAGTGGTATTGGTTTT
AGATGTTTTAGTTGTCAGCTAATTTTTGTGTTGTTGGATGTAGGAAATATTGTTATCGAG
ATTGAGTATGTTAGGCGTATGTAGAAATATAAGTTTATTAGTGTAAGTATTGTTATGGTG
AGGGGGAGAATAAGGCTGTTGTTTTTTGTAATTTCTTGATAATAGCTCATTGCGGGGAG
AAGCCTGTTAGTGGGGGAGGCCTCCTAAGGATATTATTATTATTGGGATTATTGGTATA
ACTCATGTGAGTTTATTTCAAGTGTGGGATAATGATAAGGTTGTTGTGTTTGAGTTTGAG
TAGAAAGTTATAAGTGTGGAATTTGTTAGGAAGATGTAGATAAGTAACTTAAGGCGGTA
ATGTTTGGGTTATAGAATAGTACTGCCATTATTCACCCTATATGGGTGATTGAGGTATAG
GCTAGGCTTTTGCGTAGTTGTGTTTGGTTAAGTCCTCCTCAGCTGCCAATTATGATTGAT
ATAAGTGAGATTATTAGTAGGGCACTTGAGTTGATAGATGGGAAGATTTGAATAATAATT
GACATGGGGGCTAGTTTTTGTGTCATGTAAGGATAAGTATAGTGGGAATTAGGGGAATGCCT
TGGGTTACTTCTGGAATCAGAAGTGGAGGGGAGCTATACCTAATTTTAATACGAGAGCC
ATTAATATTATTGTGGATAGAAGTTGATTAGAAGGTGGATTAATTGTTCACTGGCCGTAG
AGTAAATTGTTGAGAAAAATAGTTATTAGAAGGATTATGGACGCGGATGCTTGTGTTAGG
AAATACTTAGTGGATGCTTCTGTGGAGCGGGGATTATATTTTTAGCGAGTATGGGTATA
ATAGCTAATATGTTCAAGTTCTAAGCCTGTTTACTAGGAATCAGTGTGAGCTTAGGATT
GTGATTATGGTTCTTATTGGAATGGTTAGGGAGATAATTAGGTGGGCTAGGGGATTAATA
TTAGTACGGGAAGGATTGGACCAACATTTTCGGGCTATGGGCCGATAGCTTATTTAGCT
GACCTTACTGTTTAGAGTATGGTGTAAAGGTAGCACGGAGGATTTTGAGTTCTTAGGAA
TAGGTTGAGTCCCTAATAATTCTAGAAATAAGAGGATTTAAACCTCTATAGTTTACTCTAT
CAAAGTAATTCTTTTGTGACACATATTTCTTATGTTTGGGGTGGGATGCCGATAATAAT
ATAGGTATTGAAATATATCATATACATAGTGCCAGTGTGAGTGGTAGGAATTTTTTCAT
AGGAGATACATTAATTGGTCGTAGCGAAAACGGGGGTATGCCGTACGGATTCATAAAAAAT
AGGGTGGTTAGTAGAAGTGTGTTTAGTTGTGAAGTTTAAAGTATAAAGTTCTGTTGTGATT
ATATTGTAGGATGTTGCTGTAAAGATGGTAGTAGTTAGAGCATTTATTATAATAATGTTT
ATATATTCTGCCATAAAGAACAGGGCGAATGAACCTGCGGCATATTCAATATTGAAGCCT
GAGACTAGTTCTGATTACCTTCTGTTAGGTGCAAGGGGGCTCGATTGGTTTCTGCTAGT
GTGGAGATAAATCATATTATCGTTAGGGGTCATGATGGGAGTAGAAGTCAAGATTGTTCT
TGTGTTGTAATTAGTGATTGTAGATTAAAGGAGCCACTTATTAGTAGTGTGATAGTAGA
ATAATGGCTAGGGTAACTTCGTATGAAATGGTTTGGGCTACGGCTCGTAGTGCGCCAATT
AATGCGTAGTTAGAGTTTGATGCTCATCCGGATCATAGGGTTGAGTAAACGGCTAGACTT
GATGTTGCCAGTATAAATAGGAGGCCTAGATTGAAGTTAACTAGGGAGTACGGTATAGGG
AGGGGACTCCATAGAAGAAGGGAAATGGTTAGGGCTAGGGTTGGGGCGATTATATATAGG
GTTGTGGTAGATGTAGGTAGTAGAGGTTCTTTGTAAAGAGTTTTATTGCATCGGCG
ATTGTTGGAGTACCCCATATGGGCCTACAATGTTGGGGCCTTTGCGGAGTTGTATATAG
CCTAAGATTTTTCGTTCTGTAAAGTGTAAAAATGCTATAGCAACTAGGGCTGGGGTAATT

AATAAGAGTAGTTGATTATAAACATATTTGTTAAGAAGAGGAGGAGTGAACCTCTGATTGTGA
AAGTTTTAAGTTTTATGCAATTGCCGGGCTCTGCCATCTTAACTAAGCCCTGTTTTTGGG
CTATGCATATTAATTTGTAAGGTAAAGATATTAGTCATCTGACATTATGAGGGCGCTTTG
TAAAGTGGGCTCTATTTCTCTTGCTCTTCGTAAGGAGGAATCTTTAATAGATAGAAA
CCGACCTGGATTGCTCCGGTCTGAACTCAGATCACGTAAGACTTTAATCGTTGAACAAAC
GAACCCCTTAATAGCGGCTACACCATTGGGATGTCTTGATCCAACATCGAGGTCGTAAACC
CTATTGTCGATATGGACTCTAGAATAGGATTGCGCTGTTATCCTAGGGTAACTTAGTCCG
TTGATCAAGTTATTGGGTCAATGAGTGTAATTTACTTAGACTAGTTAGGTCATGGCTTAT
GTTTTCGGAGGTTGTAAGTGTACTCCGAGGTACACCCCAACCGAAATTTATAATACATTGA
GGTGTGTTAGTGCTGTTGGTTTTAGTGTGCTAGTTTGTATTATTAATTGAAGCTCCA
TAGGGTCTTCTCGTCTTGTGTTGTATATCCGCCTCCTCACGGATAGGTCAATTTCACTG
ATTAAGTAAGAGACAGTTAAGCCCTCGTGTTGCCGTCATACAAGTCCCTATTTAGAG
ACAAGTGATTATGCTACCTTGGCACGGTCAGGGTACCGCGCCGTTAAACATGTGTCAC
CGGGCAGGCAGTGCTCTAATACTATTAATGCTAGAGGTGATGTTTTTGGTAAACAGGCG
GGGTAAAGTTTGCCGAGTTCCTTTTACTTTTTTAACTTTTCTCGGGTGCATGCCTGTG
TTGGGTAAACAGTTTAGTTAATGGGTATTAAGTTGTGGGTGTAATGATTGTCCTGTTA
ACTAGCAGTAGTTGTTTCGATCTGTAATAAGCTTATGCGGGGAGAATAGTTTTCATGTT
ACTTATATTAACATTATTGCTTCTATTAGATAATAGATTAGTCCAATGTGTTTTAGGAGT
TCAGTGAAGTTAATGGAATTAGAGGGTGTGTTGGATATTGAGCTTGAACGCTTCTTAATT
GATGGCTGCTTTTAGGCCAACTATGGTGGTGAAGTGTTTTACTCACTATAGTAAGGTTTT
TTCCTAGGGTCTAAAGGGCTGCCCCCTTTAGACTAACAATTAATTTACGAGGGGATTA
AATGGTCTGTAAAGTAAATATAAAGTTGAACTAAGACTCTGTCTTGGAACAACAGCTATC
ACCAGGCTCGGTAGGTTTGTGCGCACTACCCATAGATTTTCCCACTATTTGCCACATAG
ACGGGTGCGCTCTTTTAGCTATCTTTGGGTAGCTCGTCTGTTTTCGGGGAGCTTTATTTA
AATTCTTTGTATAAAGTTATTTCTAGTTAATTCATTATGCGCAAGGTATAAGGGCTTGTC
TTTGCTTTTTATGCTTTGTTGGGGTTTCTTTCTTACGGTACTTCATCTATTGCGCCC
GGATATAATTTCTATCGCCTATACTTTTATAGTAGGTAAATGATTTGGTTATTGAATTTA
GGTATTAGTATAGAGAGATTGTTTGGGCTAGAGTTAGCTCAAAGTGATCGGTTGTTGTGA
GATCTTCTGGGTGTAGGCCAGATGCTTTAATTTAAGCTACACTTTGATTGTCCTCAAGTAC
ACTTTCCAGTACACTTACCATGTTACGACTTATCCCTCTGCATCAATATGAAGCGGATT
AAATTGTTAATATATTTTCTACATGATGGTTGAGGAGGGTGACGGGCGGTGTGTGCGTGC
TTAATGGCCTTGCTTCAATCAAGCTCTCTATTCTTGATTTACTGCTAAATCCACCTTGG
CCTTTAAATTTTATAAAGGCTATCGTGTAATTTCTAGATTAGAAAATGTAGCCCATTTT
TTTCCACTTCATTGGCTGCACCTTGACCTAACGTTTTTATGGTAATACTTTTGCTTACTT
TACGATCTTTAGGGAGTTTGCTGAGGATGGCGGTATACAGGCTGGGGGCAAGAGGTGGTA
AGGTTTATCGGGGTATATCGATTACAGAACAGGCTCCTCTAGACGGATGTAAAGCACCGC
CAGGTCCTTTGAGTTTCAAGCTATTGCTTGTAGTGTTCTGGCGAGTAATTTTGTTGGTAG
AGTTATTGGAGTTTAGGGCTAGGCATAGTGGGGTATCTAATCCAGTTTGTGTCTTAGCT
ATGGTGTATTTCAGGAATATTAAGCCACTTTCTGATGGGTATTTCACTATAACTGGAGTTT
TCCACAACTTAGTTATAGGTTAGCTTTATTGAGGGTGGGGCCTTAAACACTCTTTACGCC
GGACTTTATTAACCTTGAGTTAATCGTATGGCCGCGGTGGCTGGCACGAAATTGACCAACT
CTATTGTCAGTGATGCTTAGTTAACTTTCTGTTTATTGCTAAAAGTTTGTCACTGCTGTT
TCCCGTGGGGGTGTGGCTAAGCAAAGTGTTTTGAGCTGCATATATGCGTGCTTGATACT
GCTCCTCATAATTTGGTGTCTAGAGGGCATTCTCACAGGATAGTAGATGCTTGCATGTG
TAATCTCACTGAGGGCTAATAGAAAGGCTAGGACCAAACCTGTGTGTTTATGGGGTAGTA
AATACCCGCTCTAGGCATTTTCACTGCTTGTCTTTGAGTATTAAGCTACATTAAC

Also, you could look for specific features within the genbank file with the method "features" (which is a list, by the way):

```
[130]: for i in SeqIO.parse("Cebus_albifrons_NC_002763.1.gb", "genbank"):
        print(type(i.features))
        print(i.features)
```

```
<class 'list'>
[SeqFeature(Location(ExactPosition(0), ExactPosition(16554), strand=1),
type='source'), SeqFeature(Location(ExactPosition(0), ExactPosition(70),
strand=1), type='tRNA'), SeqFeature(Location(ExactPosition(70),
ExactPosition(1028), strand=1), type='rRNA'),
SeqFeature(Location(ExactPosition(1028), ExactPosition(1086), strand=1),
type='tRNA'), SeqFeature(Location(ExactPosition(1086),
ExactPosition(2651), strand=1), type='rRNA'),
SeqFeature(Location(ExactPosition(2230), ExactPosition(2366), strand=1),
type='STS'), SeqFeature(Location(ExactPosition(2327),
ExactPosition(2520), strand=1), type='STS'),
SeqFeature(Location(ExactPosition(2651), ExactPosition(2726), strand=1),
type='tRNA'), SeqFeature(Location(ExactPosition(2728),
ExactPosition(3685), strand=1), type='gene'),
SeqFeature(Location(ExactPosition(2728), ExactPosition(3685), strand=1),
type='CDS'), SeqFeature(Location(ExactPosition(3684),
ExactPosition(3753), strand=1), type='tRNA'),
SeqFeature(Location(ExactPosition(3750), ExactPosition(3822), strand=-1),
type='tRNA'), SeqFeature(Location(ExactPosition(3825),
ExactPosition(3893), strand=1), type='tRNA'),
SeqFeature(Location(ExactPosition(3895), ExactPosition(4936), strand=1),
type='gene'), SeqFeature(Location(ExactPosition(3895),
ExactPosition(4936), strand=1), type='CDS'),
SeqFeature(Location(ExactPosition(4934), ExactPosition(5001), strand=1),
type='tRNA'), SeqFeature(Location(ExactPosition(5009),
ExactPosition(5077), strand=-1), type='tRNA'),
SeqFeature(Location(ExactPosition(5078), ExactPosition(5151), strand=-1),
type='tRNA'), SeqFeature(Location(ExactPosition(5152),
ExactPosition(5188), strand=-1), type='rep_origin'),
SeqFeature(Location(ExactPosition(5186), ExactPosition(5253), strand=-1),
type='tRNA'), SeqFeature(Location(ExactPosition(5252),
ExactPosition(5320), strand=-1), type='tRNA'),
SeqFeature(Location(ExactPosition(5328), ExactPosition(6874), strand=1),
type='gene'), SeqFeature(Location(ExactPosition(5328),
ExactPosition(6874), strand=1), type='CDS'),
SeqFeature(Location(ExactPosition(6873), ExactPosition(6942), strand=-1),
type='tRNA'), SeqFeature(Location(ExactPosition(6945),
ExactPosition(7015), strand=1), type='tRNA'),
SeqFeature(Location(ExactPosition(7015), ExactPosition(7703), strand=1),
type='gene'), SeqFeature(Location(ExactPosition(7015),
ExactPosition(7703), strand=1), type='CDS'),
```

```

SeqFeature(FeatureLocation(ExactPosition(7703), ExactPosition(7766), strand=1),
type='tRNA'), SeqFeature(FeatureLocation(ExactPosition(7766),
ExactPosition(7967), strand=1), type='gene'),
SeqFeature(FeatureLocation(ExactPosition(7766), ExactPosition(7967), strand=1),
type='CDS'), SeqFeature(FeatureLocation(ExactPosition(7927),
ExactPosition(8608), strand=1), type='gene'),
SeqFeature(FeatureLocation(ExactPosition(7927), ExactPosition(8608), strand=1),
type='CDS'), SeqFeature(FeatureLocation(ExactPosition(8632),
ExactPosition(9416), strand=1), type='gene'),
SeqFeature(FeatureLocation(ExactPosition(8632), ExactPosition(9416), strand=1),
type='CDS'), SeqFeature(FeatureLocation(ExactPosition(9416),
ExactPosition(9483), strand=1), type='tRNA'),
SeqFeature(FeatureLocation(ExactPosition(9483), ExactPosition(9829), strand=1),
type='gene'), SeqFeature(FeatureLocation(ExactPosition(9483),
ExactPosition(9829), strand=1), type='CDS'),
SeqFeature(FeatureLocation(ExactPosition(9829), ExactPosition(9895), strand=1),
type='tRNA'), SeqFeature(FeatureLocation(ExactPosition(9896),
ExactPosition(10193), strand=1), type='gene'),
SeqFeature(FeatureLocation(ExactPosition(9896), ExactPosition(10193), strand=1),
type='CDS'), SeqFeature(FeatureLocation(ExactPosition(10186),
ExactPosition(11561), strand=1), type='gene'),
SeqFeature(FeatureLocation(ExactPosition(10186), ExactPosition(11561),
strand=1), type='CDS'), SeqFeature(FeatureLocation(ExactPosition(11561),
ExactPosition(11630), strand=1), type='tRNA'),
SeqFeature(FeatureLocation(ExactPosition(11630), ExactPosition(11689),
strand=1), type='tRNA'), SeqFeature(FeatureLocation(ExactPosition(11689),
ExactPosition(11760), strand=1), type='tRNA'),
SeqFeature(FeatureLocation(ExactPosition(11762), ExactPosition(13568),
strand=1), type='gene'), SeqFeature(FeatureLocation(ExactPosition(11762),
ExactPosition(13568), strand=1), type='CDS'),
SeqFeature(FeatureLocation(ExactPosition(13564), ExactPosition(14098),
strand=-1), type='gene'), SeqFeature(FeatureLocation(ExactPosition(13564),
ExactPosition(14098), strand=-1), type='CDS'),
SeqFeature(FeatureLocation(ExactPosition(14098), ExactPosition(14167),
strand=-1), type='tRNA'), SeqFeature(FeatureLocation(ExactPosition(14171),
ExactPosition(15308), strand=1), type='gene'),
SeqFeature(FeatureLocation(ExactPosition(14171), ExactPosition(15308),
strand=1), type='CDS'), SeqFeature(FeatureLocation(ExactPosition(15309),
ExactPosition(15379), strand=1), type='tRNA'),
SeqFeature(FeatureLocation(ExactPosition(15380), ExactPosition(15448),
strand=-1), type='tRNA'), SeqFeature(FeatureLocation(ExactPosition(15448),
ExactPosition(16554), strand=1), type='D-loop')]

```

We can, for instance, generate a multifasta containing all tRNA features from the genbank:

```

[15]: for i in SeqIO.parse("Cebus_albifrons_NC_002763.1.gb", "genbank"):
      for feature in i.features:
          if feature.type == "rRNA":

```

```

        print(feature)
    #     header = feature.qualifiers.get('product')[0]
    #     print(header)

```

```

type: rRNA
location: [70:1028](+)
qualifiers:
    Key: product, Value: ['12S ribosomal RNA']

```

```

type: rRNA
location: [1086:2651](+)
qualifiers:
    Key: product, Value: ['16S ribosomal RNA']

```

```

[176]: for i in SeqIO.parse("Cebus_albifrons_NC_002763.1.gb", "genbank"):
        for feature in i.features:
            if feature.type == "tRNA":
                header = feature.qualifiers.get('product')[0]
                seq = feature.location.extract(i.seq)
                print(">{}\n{}".format(header, seq))

```

```

>tRNA-Phe
GTTAATGTAGCTTAATACTCAAAGCAAGGCACTGAAAATGCCTAGACGGGTATTTACTACCCCATAAACA
>tRNA-Val
CAAAGTGTAGCTTAAATTAAAGCATCTGGCCTACACCCAGAAGATCTCACAACAACCG
>tRNA-Leu
GTTAAGATGGCAGAGCCCGCAATTGCATAAACTTAAACTTTACAATCAGAGGTTCAACTCCTCTTCTTAACA
>tRNA-Ile
AGAAATATGTCTGACAAAAGAATTACTTTGATAGAGTAACTATAGAGGTTTAAATCCTCTTATTTCTA
>tRNA-Gln
TAGAGTATGGTGTAAATAGGTAGCACGGAGGATTTTGAGTTCTTAGGAATAGGTTGAGTCTATAATTCTAG
>tRNA-Met
AGTAAGGTCAGCTAAATAAGCTATCGGGCCCATACCCCGAAAATGTTGGTCCAATCCTTCCCGTACTA
>tRNA-Trp
AGAAATTTAGGTTAATAAGACCAAGAGCCTTCAAAGCCCCTAGTAAGTAAATTTTACTTAATTTCTG
>tRNA-Ala
GAGGGCTTAGCTTAATTAAGTAGTTGATTTGCGTTCAATTGATGCAAGGTATAGTTGCAGTCCTTA
>tRNA-Asn
TAGATTGAAGCCAGTTGATTAGGTTATTTAGCTGTAACTAAATTTTTGTGGGTTAAAGTCCCATCAGTCTAG
>tRNA-Cys
AGCCCTGAGGTGAACTGTCATGTTGAACTGCAAATTCAAAGAAGCAGCTTCAATGCTGCCGGGGCTT
>tRNA-Tyr
GGTAAATGGCTGAGTAAGCATTAGACTGTAAATCTAAAGACAGAGAAGTGACCCCTCTTTTTACCA
>tRNA-Ser
GAAAAAGTCATAGGGTCTATGGGATTGGCTTGAAACCAATTTTTGGGGTTCAAATCCTTCCTTTTTTCG
>tRNA-Asp
GAGATATTAGTAAAAATAAATTACATAACTTTGTCAAAGTTAAATTATAAGACCAAAACCTATATATCTTA

```

```

>tRNA-Lys
CACTGTAAAGCTACTTAGCATTAAACCTTTTAAGTTAAAGACAGAGAGAATCTCTCTACAGTGA
>tRNA-Gly
ATTCTCTTAGTATAAACAGTACAATTGACTTCCAATTAATAGGCCTTGATAAACCCAAGAGAGAATA
>tRNA-Arg
TGGTATATAGTTTAAACAAAATAAGTGATTTTCGACTCATTAGACTATGGTAAACCGTATATACCAA
>tRNA-His
GTAGCTATAGTTTAATTAATAAACATTAGATTGTGAATCTAATAATAGAAGACTACAACCTTCTTAACCTACC
>tRNA-Ser
GAGAAAGTATGCAAGAAGCTGCTAACTCCTGCCTCCAGACTTAATATCCTGGCCTTCTCA
>tRNA-Leu
GCTTTTAAAGGATAGTAGTTATCCATTGGTCTTAGGAGCCAAAAATATTGGTGCAACTCCAAATAAAAGCA
>tRNA-Glu
ATTCTTATAGTTGAAATACAACGATGGCTTTTCATATCATTAGTCATGGTTATAATCCATGTGGGAATA
>tRNA-Thr
GTCCTTGTAAGTATATCCAATTACCCGGCCTTGTAACCGGAAAAGGAGGCACGCTAACTCCCCAGGACA
>tRNA-Pro
CAGGGAATAGTTTAATGTAGAATTTTCAGCTTTGGGTGTTGATGGTAGAATTAATTATTCTTTTCCTGA

```

2 Handling fasta files:

The fasta files have way less options, but one of them is the seq.description, that gets the whole header line.... (*FINISH THIS LATER...*)

3 Handling ace files:

Ace files can be treated as sequences or as alignments, as can be seen [here](#)

These are only a few examples of what SeqIO (and the Biopython module in general) can do in only a few lines of code. Live long and prosper and happy parsing!