Zanima me kam se poravna končni CDS (`NW_025235126`) v gen `AB982279.1` . Glej 3. Exercise: Construction of CRISPR/Cas9 plasmid.

In [ ]:
```
from Bio import Align
import Bio.SeqIO as SeqIO
aligner = Align.PairwiseAligner()
```

In [71]:
```
aligner.mode = 'global'
print(aligner)
```

```
Pairwise sequence aligner with parameters
  wildcard: None
  match_score: 1.000000
  mismatch_score: 0.000000
  open_internal_insertion_score: -1.000000
  extend_internal_insertion_score: -1.000000
  open_left_insertion_score: -1.000000
  extend_left_insertion_score: -1.000000
  open_right_insertion_score: -1.000000
  extend_right_insertion_score: -1.000000
  open_internal_deletion_score: -1.000000
  extend_internal_deletion_score: -1.000000
  open_left_deletion_score: -1.000000
  extend_left_deletion_score: -1.000000
  open_right_deletion_score: -1.000000
  extend_right_deletion_score: -1.000000
  mode: global
```

In [26]:
```
s1 = "AB982279.1.fasta"
s2 = "NW_025235126.1[4419..5271].fasta"
```

In [40]:
```
with open(s1, 'r') as fs1, open(s2, 'r') as fs2:
    cs1 = SeqIO.parse(fs1, "fasta")
    cs1 = next(cs1)
    cs2 = SeqIO.parse(fs2, "fasta")
    cs2 = next(cs2)
```

In [62]:
```
print(cs1)
print("seq len:", len(cs1.seq))
```

```
ID: AB982279.1
Name: AB982279.1
Description: AB982279.1 Triticum aestivum gene for alpha/beta-gliadin, complete c
ds, cultivar: Chinese Spring, clone: clone46
Number of features: 0
Seq('GCTAATAACTTCTAGAACACTACAACACTTGACATGTAAAAGGAATTTGATGAG...CCG')
seq len: 1570
```

In [51]:
```
print(cs2)
print("seq len:", len(cs2.seq))
```

```
ID: ref|NW_025235126.1|:4419-5271
Name: ref|NW_025235126.1|:4419-5271
Description: ref|NW_025235126.1|:4419-5271 Triticum aestivum cultivar Chinese Spr
ing unplaced genomic scaffold, IWGSC CS RefSeq v2.1 scaffold73444, whole genome s
hotgun sequence
Number of features: 0
Seq('TGCCCTCCTTGCTATTGTAGCAACCACCGCCACAATTGCAGTTAGAGTTCCAGT...TAT')
seq len: 853
```

In [72]:
```
alignment = aligner.align(cs2, cs1)
# scores = aligner.score(cs2, cs1)
```

In [ ]:
```
# print(scores)
```

```
127.0
```

In [73]:
```
a = alignment[0]
print("score:", a.score)
print("length:", a.length)
print(a)
```

```
score: 127.0
length: 1573
ref|NW_02         0 --T-----------G--C-C--C-----T---C----------------------C----
                  0 --|-----------|--|-|--|-----|---|----------------------|----
AB982279.         0 GCTAATAACTTCTAGAACACTACAACACTTGACATGTAAAAGGAATTTGATGAGTCAGCG

ref|NW_02         8 T--T---GC---T-AT-T-----GT---A------------G-C-A-AC-------CA--
                 60 |--|---||---|-||-|-----||---|------------|-|-|-||-------||--
AB982279.        60 TACTAAAGCAAGTTATATTACTAGTCTTATCTACCTTAACAGGCCACACAAGATTACAAA

ref|NW_02        26 CC--G--C----C-A-C-A-----AT-T-G------C--A-------G-------TTA--
                120 ||--|--|----|-|-|-|-----||-|-|------|--|-------|-------|||--
AB982279.       120 CCAAGTTCTGTGCCAGCCATGCTTATCTAGATTATGCGTAACAATTTGCAGAAAATTACA

ref|NW_02        44 ------G----AG-----T------TC----------------C---A--GTG-------
                180 ------|----||-----|------||----------------|---|--|||-------
AB982279.       180 AACTTAGTTTCAGAAAAATAGGCAATCTAGATTAGTGTTTGAGCTGTAAAGTGAGTAAAA

ref|NW_02        55 -----C---C-----A-CA-A--TT---G-----------------CA-------G---C
                240 -----|---|-----|-||-|--||---|-----------------||-------|---|
AB982279.       240 TGAGTCATGCGTGTTATCACACCTTTTTGGTTGTGGAATGATAGTGCAACAACAAGGAAC

ref|NW_02        68 --------C-A---CAA-AAT-C-C----A--T----C--------------T-C-A---
                300 --------|-|---|||-|||-|-|----|--|----|--------------|-|-|---
AB982279.       300 TTTAATGACCAGTTCAAGAATACACTTGTAAGTAGTGCCACCGAACACAACATACCAAAT

ref|NW_02        84 ---G-----------CA---A--CA-----AC-CA--CAA--G--A---G----------
                360 ---|-----------||---|--||-----||-||--|||--|--|---|----------
AB982279.       360 TATGATTTTGAGAAGCATCCAAGCACTTCCACACAAGCAAATGCCAATTGTGAAAGAGAT

ref|NW_02       100 CA----A-G-----T-T-------CC---A---T-----T--------------------
                420 ||----|-|-----|-|-------||---|---|-----|--------------------
AB982279.       420 CATGCCATGGCAGCTATAAATAGGCCCGTAGCATGACGATCATCCTTCCTCATCCATCAT

ref|NW_02       111 --------G--G------T-A------C-AA-CAA-CA------A---CAA-T------T
                480 --------|--|------|-|------|-||-|||-||------|---|||-|------|
AB982279.       480 TCTCATAAGTAGAGCGCATCATTTAAGCCAAGCAAGCAGTGCTCAATACAAATCCACCAT

ref|NW_02       129 -------T--C-CA-----G--------G-----G---CA------GC-A-A----CA--
                540 -------|--|-||-----|--------|-----|---||------||-|-|----||--
AB982279.       540 GAAGACCTTTCTCATCCTTGCCCTCCTTGCTATTGTAGCAACCACCGCCACAATTGCAGT

ref|NW_02       144 -A-----CCA-T-------TT-C--C-AC-----C-A-C--A--A-CA-------G--C-
                600 -|-----|||-|-------||-|--|-||-----|-|-|--|--|-||-------|--|-
AB982279.       600 TAGAGTTCCAGTGCCACAATTGCAGCCACAAAATCCATCTCAGCAACAACCACAAGAGCA

ref|NW_02       164 ----C-AT----A-----------T--CC-G--CAGC--C------------------
                660 ----|-||----|-----------|--||-|--||||--|------------------
AB982279.       660 AGTTCCATTGGTACAACAACAACAATTTCCAGGGCAGCAACAACCATTTCCACCACAACA

ref|NW_02       177 GC-A-A-CC--A----------TTTCCATCACAACAACCATATCTGCAGCTGCAACCATT
                720 ||-|-|-||--|----------||||||||||||||||||||||||||||||||||||
AB982279.       720 GCCATATCCGCAGCCGCAACCATTTCCATCACAACAACCATATCTGCAGCTGCAACCATT

ref|NW_02       222 TCCGCAGCCGCAACTACCATATCCGCAGCCGCAACTACCATATCCGCAGCCGCAACTACC
                780 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AB982279.       780 TCCGCAGCCGCAACTACCATATCCGCAGCCGCAACTACCATATCCGCAGCCGCAACTACC

ref|NW_02       282 ATATCCGCAGCCGCAACCATTTCGACCACAACAACCATATCCACAATCGCAACCACAGTA
                840 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
AB982279.       840 ATATCCGCAGCCGCAACCATTTCGACCACAACAACCATATCCACAATCGCAACCACAGTA

ref|NW_02       342 TTCGCAACCACAACAACCAATTTCGCAGCAGCAGCAGCAGCAACAACAACAACAACAACA
                900 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AB982279.       900 TTCGCAACCACAACAACCAATTTCGCAGCAGCAGCAGCAGCAACAACAACAACAACAACA

ref|NW_02       402 AAAACAACAACAACAACAACAACAACAGATCCTTCAACAAATTTTGCAACAACAACTGAT
                960 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AB982279.       960 AAAACAACAACAACAACAACAACAACAGATCCTTCAACAAATTTTGCAACAACAACTGAT

ref|NW_02       462 TCCATGCAGGGATGTTGTATTGCAACAACACAGCATAGCGTATGGAAGCTCACAAGTTTT
               1020 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AB982279.      1020 TCCATGCAGGGATGTTGTATTGCAACAACACAGCATAGCGTATGGAAGCTCACAAGTTTT

ref|NW_02       522 GCAACAAAGTACTTACCAGCTGGTGCAACAATTGTGTTGTCAGCAGCTGTGGCAGATCCC
               1080 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AB982279.      1080 GCAACAAAGTACTTACCAGCTGGTGCAACAATTGTGTTGTCAGCAGCTGTGGCAGATCCC

ref|NW_02       582 CGAGCAGTCGCGGTGCCAAGCCATCCACAATGTTGTTCATGCTATTATTCTGCATCAACA
               1140 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AB982279.      1140 CGAGCAGTCGCGGTGCCAAGCCATCCACAATGTTGTTCATGCTATTATTCTGCATCAACA

ref|NW_02       642 GCAACAACAACAACAACAACAACAACAACAACCGTTGAGCCAGGTCTCCTTCCAACAGCC
               1200 ||||||||||||||||||||||||||||||||--|-||||||||||||||||||||||||
AB982279.      1200 GCAACAACAACAACAACAACAACAACAAC--C-GTTGAGCCAGGTCTCCTTCCAACAGCC

ref|NW_02       702 TCAACAACAATATCCATCAGGCCAGGGCTCCTTCCAGCCATCTCAGCAAAACCCACAGGC
               1260 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AB982279.      1257 TCAACAACAATATCCATCAGGCCAGGGCTCCTTCCAGCCATCTCAGCAAAACCCACAGGC

ref|NW_02       762 CCAGGGCTCTGTCCAGCCTCAACAACTGCCCCAGTTTGAGGAAATAAGGAACCTAGCGCT
               1320 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AB982279.      1317 CCAGGGCTCTGTCCAGCCTCAACAACTGCCCCAGTTTGAGGAAATAAGGAACCTAGCGCT

ref|NW_02       822 AGAGACGCTACCTGCAATGTGCAATGTCTAT----------------------------
               1380 |||||||||||||||||||||||||||||||----------------------------
AB982279.      1377 AGAGACGCTACCTGCAATGTGCAATGTCTATATCCCTCCATATTGCACCATTGCTCCAGT

ref|NW_02       853 ------------------------------------------------------------
               1440 ------------------------------------------------------------
AB982279.      1437 TGGCATCTTCGGTACTAACTGAGAAGAGAAGAACTCTAGTAGTAGATATATGATACACCG

ref|NW_02       853 ------------------------------------------------------------
               1500 ------------------------------------------------------------
AB982279.      1497 TTTTCTTAGTCCATGGTTTGGTCGTTGTAGCGGTGAAAAAATAAAGTGACATGCACTATC

ref|NW_02       853 -------------  853
               1560 ------------- 1573
AB982279.      1557 ATGTAAGAACCCG 1570
```