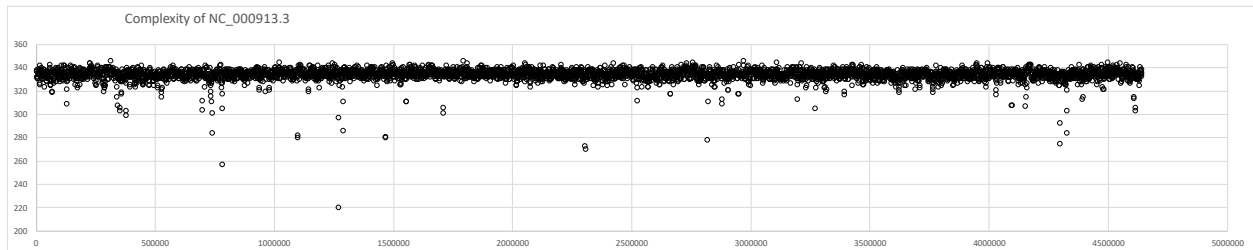
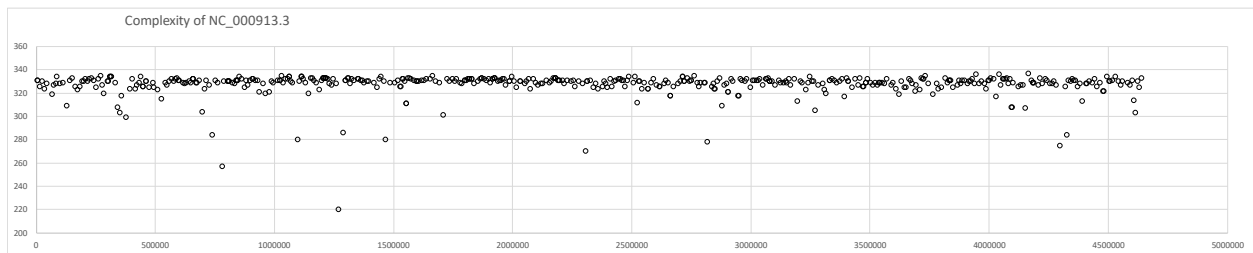


Looking for low-complexity sequences in a genome

Plots of results from the *E. coli* genome NC_000913.3



All windowed compression ratio values



Windowed compression ratio values *that represent the local minimum over an 11-kilobase range*

Python code and example of calling it

```
#!/usr/bin/python3
import fileinput
import re
import zlib

'''
Computes compression ratio for a fasta-format
sequence in overlapping 2-K blocks
'''

# grab the sequence lines, ignoring annotation lines
sequence = ''

with fileinput.input() as stdin:
    for line in stdin:
        line = line.rstrip('\r\n')
        p = re.compile('^>')
        m = p.match(line)
        if not(m):
            sequence += line
            # done

# OK, now I have just the sequence data.

length = len(sequence)
print ('Sequence length: ' + str(length)) # python is strict...

# Go through the sequence in chunks
position = 1
while position < (length - 1000):
    endpos = position + 2001
    substring = sequence[position:endpos]
    compressed = zlib.compress(substring.encode('utf-8'))
    # Note: we must convert native Python string object
    # into bytes. Python2 way was just use the bytes()
    # function, but explicitly specifying the encoding
    # is considered better practice with modern Python
    ratio = len(compressed) / len(substring)
    ratio = int(1000 * ratio + 0.5)
    ncbi = '>NbCiAcCno'
    # + str(position-1) +
    # '-' + str(endpos-2)
    print (str(ratio) + '\t' +
           ncbi + str(position-1) +
           '-' + str(endpos-2) + '\t' +
           substring)
    position = position + 1000
```

And here is how I ran this code in a Virtual Linux machine:

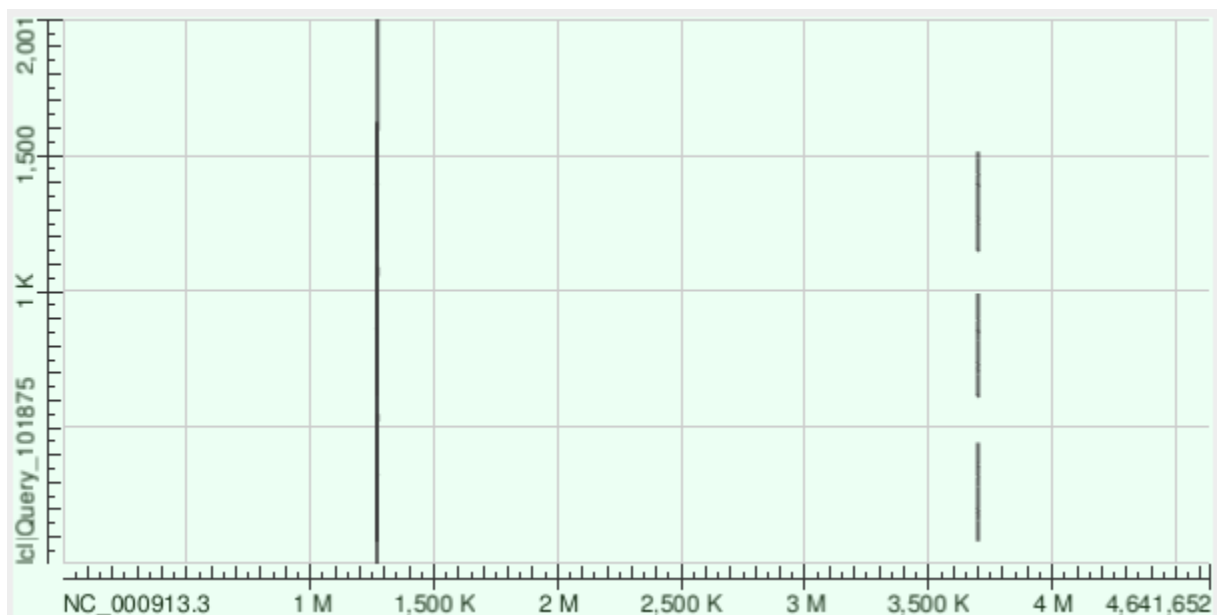
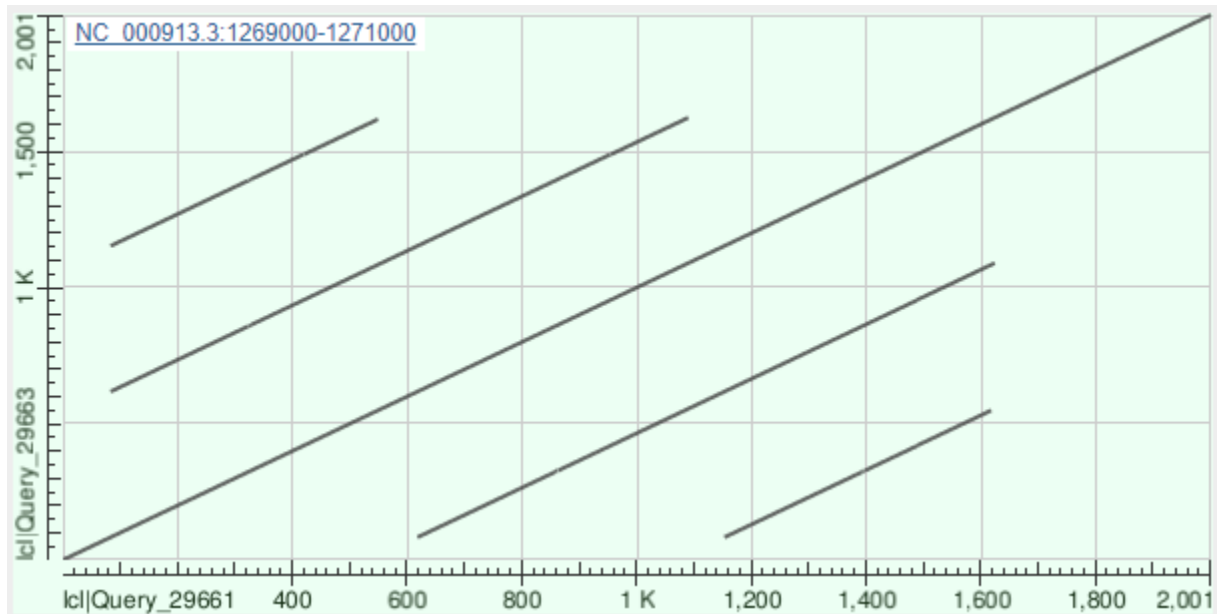
```
manager@bl8vbox[LowComplexityTool]
manager@bl8vbox[LowComplexityTool] ./WindowedCompressionRatios.py <
NC_000913.3_Escherichia_coli_K-12_MG1655.fasta|sort -n | perl -pe
's/$/\n/g;s/\t/\n/g;s/NbCiAcCno/NC_000913.3:/' >
NC_000913.3_ChunksByComplexity.txt
manager@bl8vbox[LowComplexityTool]
```

I used a couple of one-liner regexes to pull out the compression ratio values and the positions, and loaded those into MS Excel for interactive plotting and filtering.

To make the dot matrix plots, I simply pasted sequence chunks into the NCBI Blast Web Page, with the Align two or more sequences box checked. In the standard output format from the NCBI Blast Web Page, there is an expandable section with the Dot Matrix View.

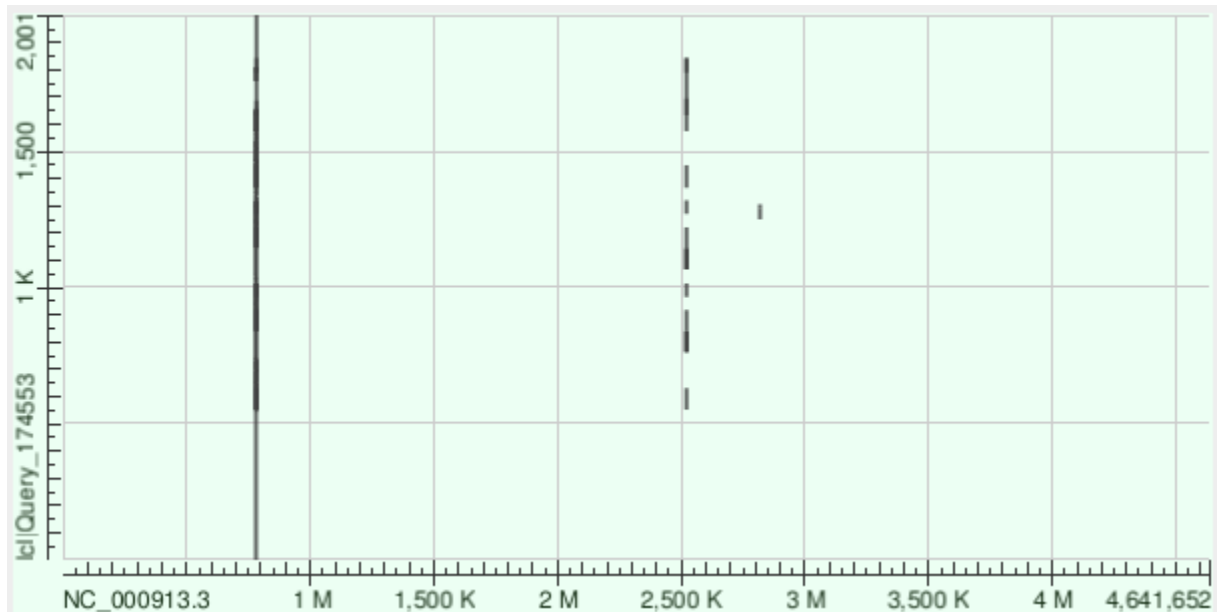
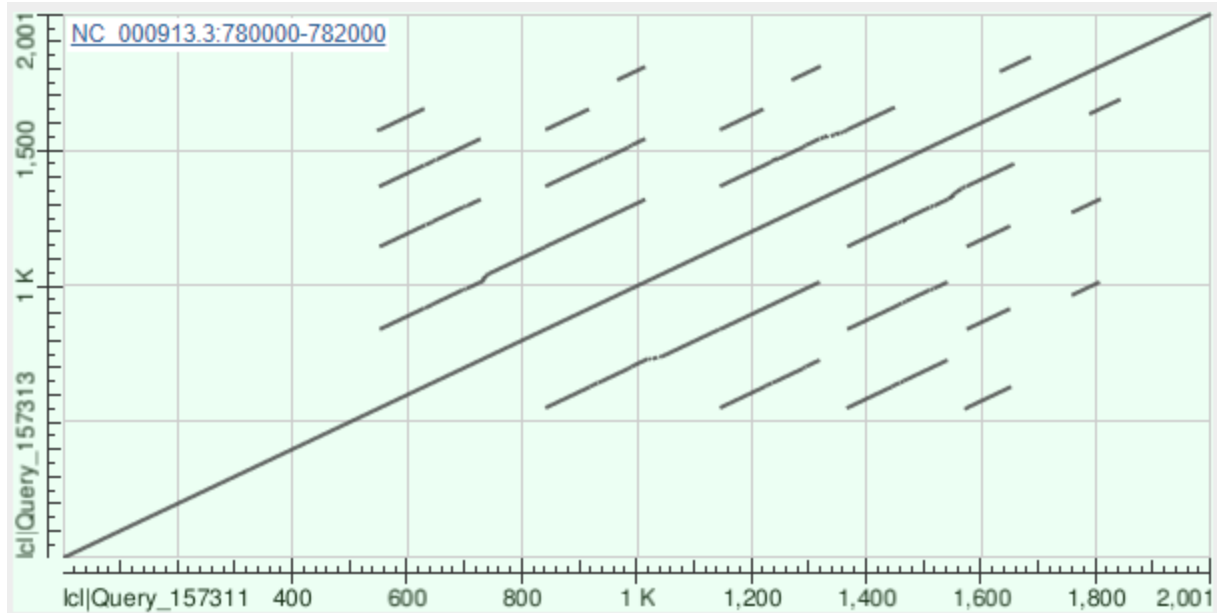
For each example below, I show two Dot Matrix Plots. The first is the subsequence compared to itself. The second is the subsequence compared to the entire genome. You can see that some of these low-complexity sequences appear just one place in the genome, and no other region is similar, while others have multiple full or partial repeats in the genome.

Example results from the *E. coli* genome NC_000913.3



```
>NC_000913.3:1269000-1271000
CTGGATACCAGCAAGTAATCTTTTTTGCTTGAAAAATAAAGTATTAGCGTTCTGCGTTAAGACTTTTTTCATGGGTGCCGGATACAAAAAGGCCGC
AGGCTGTTACCCCTGCGGCCGGTTTCGGGCGCATATTGCCATCACGGCAGCCTGACGCCCCTTTTCACCTTACTTCCGGTTACGCCACAGCTGACA
ATCGCTGCGGTAATAATTCGCCAGGATCGGTGCTGCCAGGTCGTGCCAGAAATCATGGCAAACGCGCGAGCGTCATATAGCCGCTTGTGTGA
ATGACAACATTTTGC GGCTATTCTTGAATTGTTCTGGTTCAAGATTAGCCCCGTTCTGTTGTCAGGTTGTACCTCTCAACGTGCGGGGGTTTTCTC
TTTCCAGCAACCAATGCCACCAGGGATAAAGCCCCCGCAACATTGCGCCTCACCGGATAACGCGCGGCTTGGTGTGGATACTACGACTCAATTCATCT
TCACTTCAATCCCTGAAATGTTTGCAATGAAGAGTGCAATCCGGTTTTTCAACAGCTGTACAGTCATTTTCATGAGTGCTCTGGATGAGGCTTCCAGC
TCGGGTTGCCAATATTTACTTGTGGAAGTGATAAAGACAAAAATGGCCGCGAGGCTGTTACCCCTGCGGCCGGTTTCGGGCGCATATTGCCATCACGG
CAGCCTGACGCCCCTTTTCACCTTACTTCCGGTTACGCCACAGCCGACAATCGCTGCGGTAATAATTCGCCAGGATCGGTGCCGCCAGGTCGTG
CCAGAAAGTCATGGCAAACGCGCGAGCGTCATATAGCCGCTTGTGTAATGACAACATTTTGC GGCTATTCTTGAAGTGCTGTTTCAAGATTA
GCCCCGTTCTGTTGTCAGGTTTTACCTCTCAACGTGCGGGGGTTTTCTCTTTCCAGCAACCAATGCCACCAGGGATAAAGCCCCCGCAACATTGCG
CCTCACCGGATAATTCGGCTTGGTGTGGATACTACGTCTCAATTCATCTTCACTTCAATCCCTGAAATGTTTGCAATAAAGAGTACATTCGGCTTT
```

TCAACAGCTGTTGCAGTCGTTTCATGAGTGCTCTGGATGATGCTTCCAGCTCGGGTTGCCAATATTTACTTGTGGAAGAGATAAAGACAAAAATGGC
CGCAGGCTGTTACCCCTGCGGCCGGTTTCGGGCGCATATTGCCATCACGGCAGCCTGACGCCCGTTTTTCACCTTACTTCCGGTTACGCCACCAGCTG
ACAATCGCTGCGGTAATAATTCCCGCCAGGATCGGTGCTGCCAGGTCTGCCAGAAAATCATGGCAAAC TGCGGAGCGTCATATAGCCGCCTTGTT
GTAATGACAACATTTTGCGGCTATTCTTGAAGTGTCTGGTTTCAAGATTAGCCCCGTTTTTGTGTCAGGTTTTACCTCTCAACGTGCGGGGGTTTT
CTCTTTCCAGCAACCAATGCCACCAGGGATAAAGCCCCGCAACATTGCGCCTCACCGGATAACGCCGGCTTG GTGTGGATACTACGTGCAATTCAT
CTTCACTTCATCCCTGAAATGTTTGCAATGAAGAGTACATTCCGGCTTTTCAACAGCTGTTGCAGTGGTTTCGTGCAGGGTGTGGTACAGGCTCGCA
ATTCTGATTAAACGACCTGTCCAGGTAGTATGAAGAAATACAATAAGCAGGAATTTATATTATCTCCCCGATTACCGTCATCGGAGGAGATATACT
GTCAGGCAAATATCGTCATCAAATAGGCGGCAAACAGTGCCAGATGCGCTGCGCCATTGAGCACGTTAGTACGTCCGGTGGAGAAGGAGATATGGCA
CAGCACTAAAGAGGCCACCATCACCACATTTCTGGCGCACCAGTGCAAAC TGCAATTCGTTACCCGTCATAAAGGCAATTAGCGTGACGACAGGT
ACGGTAAGCGAAATGGTTGCTAACACTGAACCAAAGAACAGATTCATCGCGCGCTGAACCT



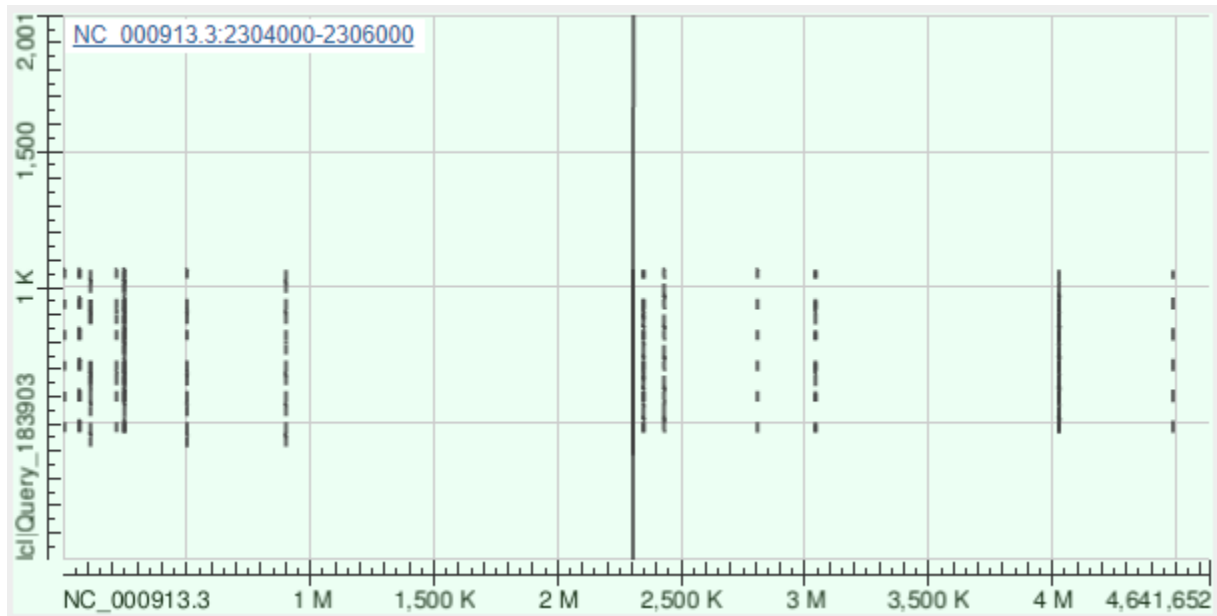
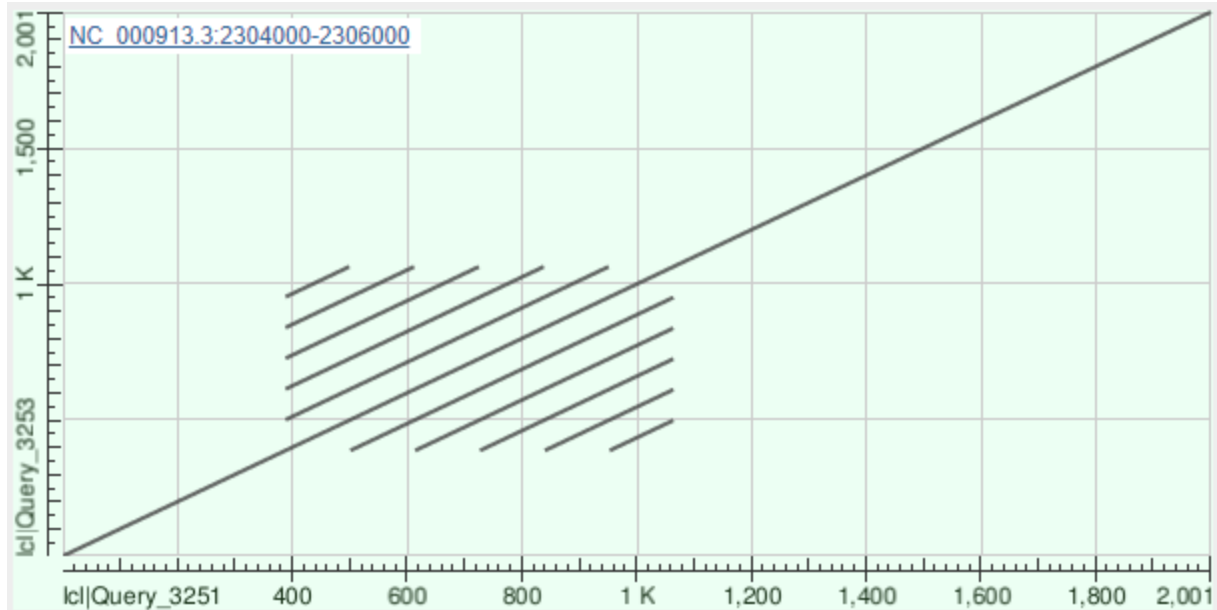
>NC_000913.3:780000-782000

```

GCCGGTGAAAAGCGGTAATGCAACACGGATTACAATGCAGCTATTGCGCTGGTGCAGGATAAATCCCGCCAGGATGACGCAATGGTGGCATTTCAG
AATTTTCATCAAAAATTACCTTGATTCAACTTACCTGCCAAACGCCAATTATTGGCTGGGTGAGTTAACTACAACAAGGGTAAAAAGATGATGCGG
CGTACTATTTTGCTTCGGTAGTGAAAACTATCCGAAGTCACCAAAGGCTGCAGATGCGATGTTTAAAGTCGGCGTCATCATGCAGGACAAAGGTGA
CACCGCAAAAAGCGAAAGCCGTATACCAGCAGGTATCAGTAAATACCTGGTACTGATGGCGCTAAACAGGCACAAAAACGTCTGAACGCGATGTAA
TGCATAACACACGACCAGAAGTCGCATTATTTCTGGTCGTGTCGTGCGAATCATAAGCAGTTGAGTGATCTACATCGAAATTTTGTGCGCTCAAG
TCTGAAATCAGTAATATATGCCGCCGTGCCACGGGATATCAAACAACCCGAAAGCAACGAAAAAGTGGGTCGTTAGCTCAGTTGGTAGAGCAGTTG
ACTTTTAATCAATTGGTCGCAGGTTTCAATCCTGCACGACCCACCAATCGCTAAGGTGGAAGCGGTAGTAAACGTGAAGGATAACGTTGCATGAGC
AACGGCCCGAAGGGCGAGACGAAGTCGAGTCATCCTGCACGACCCACCACTAACATAGTTAGTTGTAGTATCCAGCGTAGTATCGGGTGATTAGCTC
AGCTGGGAGAGCACCTCCCTTACAAGGAGGGGGTCGGCGGTTTCGATCCCGTCATCACCACCCAGGGTCGTTAGCTCAGTTGGTAGAGCAGTTGAC
TTTGAATCAATTGGTCGCAGGTTTCAATCCTGCACGACCCACCACTTTTAAACATCAAAGACAGATGTTAAGCGTGTAGGATAACGTTGCGTCAGCAA
CGGCCCGTAGGGCGAGCGAAGCGAGTCATCCTGCACGACCCACCACTAATGACGGTGGGTTTCGGTGGAAGTAGTTTGTAGTATCCAGCGCAGTATCG
GGTGATTAGCTCAGCTGGGAGAGCACCTCCCTTACAAGGAGGGGGTCGGCGGTTTCGATCCCGTCATCACCACCACTCGGGTCGTTAGCTCAGTTGG
TAGAGCAGTTGACTTTTAATCAATTGGTCGCAGGTTTCAATCCTGCACGACCCACCACTTTTAAACATCAAAGTCAAGTGTAAAGCGTGAAGGATAAC
GTTGCGCGCAGCAACGGCCGTAGGGCGAGCGAAGCGAGTCATCCTGCACGACCCACCACTTTAAAGATTGGCCCCGAGTAAAAATCTTTCAGGTAA
CACCCGTATGGGTCGTTAGCTCAGTTGGTAGAGCAGTTGACTTTTAATCAATTGGTCGCAGGTTTCAATCCTGCACGACCCACCACTTCAAGGTGG
TTACTGGTAGAGAACGTGAAGGATAACGTTGCGTTAGCAACGGCCGAGGGCGAGACGAAGTCGAGTCATCCTGCACGACCCACCACTCCTGAATGA
TTAAGGCAGCATAATCCCGCAAGGGGTCGTTAGCTCAGTTGGTAGAGCAGTTGACTTTTAATCAATTGGTCGCAGGTTTCAATCCTGCACGACCCAC

```

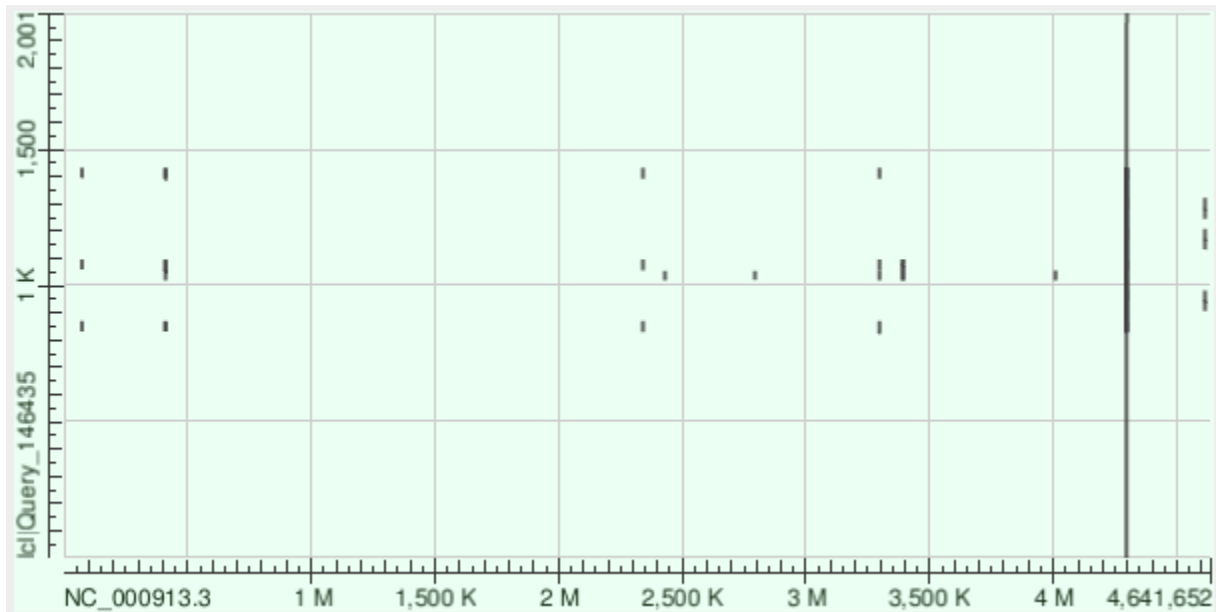
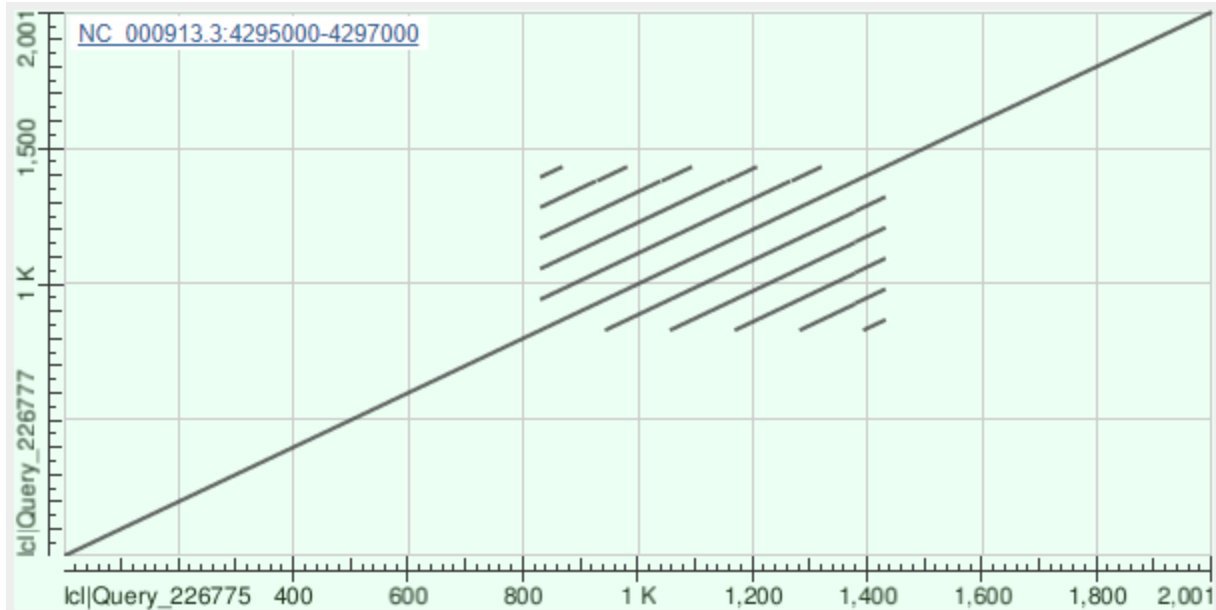
CAATGTAAAAAGCGCCCTAAAGGCGCTTTTTTGCTATCTGCGATACTCAAAGATTGGAACCTGCAGCAGGTTGAGTTGAGCGCAGCGAAACAACG
GAGCCGCTCGCGGCGACGGCCCGAAGGGCGAGCGAAGCGAGTCATCCTGCACGACCCACCAATGTAAAAAGCGCCCTAAAGGCGCTTTTTTGCTAT
TCAGGCATCCTCAATTTCACTTTGTAAACCTGATGACATCGTCAGAGCTTACTGTGCAAGCAACTCTATGTCGGTGGAATTAGGCGTAAATGACGC
ATCCTGCACATTAGGCGTAATTCGAGTGACTTTTCCCCACCATTCGACTATCTTGTTTAGC



>NC_000913.3:2304000-2306000

```
ATCCACAAGCTGAAAAAGGGATGAAGCGTCAGGTGATTGAGTTAACCCCGCAAGAAGATGAATCTACCTGAAAGTAGAACTGTTAATCGGTCAGAC
GCTGGAAGTCGATTGCAATTTGCATCGTCTCGGCGGGAAGCTGGAACAAAACGCTGGAAGGCTGGGGCTATGATTATTATGTCCTTTGATAAAGTC
AGTCCCCCGGTTTCAACGATGATGGCCTGCCCGGATGGCAAGAAAGAGAAGAAATTTGTACCGCGTATCTGGGCGATGCTGGAATGCTGCGTTACA
ACAGCAAGCTGCCGATCGTGGGTATACGCCAGACAATGTAGATGTGAAGTACCGCGTCTGGAAGGCGGAAGAGAAAATTGACAACCGCGGTAGTTTCG
CTAAACTGCCGTGAAGTGCAGCACCCCGTAGGTGAGTACAGCAAGGCGGTACGCGCGCATCCGACATCCAACGCCCCGAGCCGGTTGCCTGATGCGACGCTG
GCGGTCTTATCAGGCCTACACCGCTGTGAAGTGCAGCACCCCGTAGGTGAGTACAGCAAGGCGGTACGCGCGCATCCGACATCCAACGCCCCGAGCCGGTT
GCCTGATGCGACGCTGGCGCGTCTTATCAGGCCTACACCGCTGTGAAGTGTCTCCACCCCGTAGGTGCGGATAAGGCGGTACGCGCGCATCCGACATCT
AACGCCCCAAGCCGGTTGCCTGATGCGACGCTGGCGCGTCTTATCAGGCCTACACCGCTGTGAAGTGTCTCCACCCCGTAGGTGGATAAGGCGGTAC
GCCGATCCGACATCCAACGTCGAGCCGGTTGCCTGATGCGACGCTGGCGCGTCTTATCAGGCCTACACCGCTGTGAAGTGTCTCCACCCCGTAGGT
CGGATAAGGCGGTACGCGCATCCGACATTAACGCCCCAAGCCGGTTGCCTGATGCGACGCTGGCGCGTCTTATCAGGCCTACACCGCTGTGAAGT
GCTTACCCCGTAGGTGGATAAGGCGGTACGCGCATCCAACATCTAACGCCCCGAGCCGGTTGCCTGATGCGACGCTGGCGCGTCTTATCAGCAT
ACGCCACATCCGCGATACCATGCCGATGTGGCGTATCATTACAACGCAATATCCGCCACTTCTTTTTGAACGGGTTGCGGTTTCAACTGCGGTTTC
GGCGTACTATCTGCTGCTGCGGCTTGTCTAGTTTCAACCCAGCACTTCGCTGGTGTACTGCAACTCGCGTCTGTTGCCGCTACATCACCGTTCA
GTTGCGTCCATAAGACGGAACGATCGCTTTCAACGTAGCCTGCCATTGCGGGCTGGAACACGATCGCCAAATACTTTTCCAGCAGATTCAACAT
AATCGGCGCGGCGGTTGACGCCCCCTGGCGATGCCCCAGGAGCGCGGCAATGGTTTCTTGGTGGTCACTGACGACTTCAGTACCCAGACGCGATACG
CCACCTTTCTCGGCATCAGCTTGATAATCTGCACGCGCTGCCCGCTTGCCACAAACGCCAGTCTCTTTTTTCGCTGCGGATAGTACTCTTTCA
ACGCTTCAAAACGATCCTCTTCACTCAACATCACTGACTCACCAGATATTTACCAGATCGAAATTATCCAGCCGACGTGCATCATCGGCATCAC
```

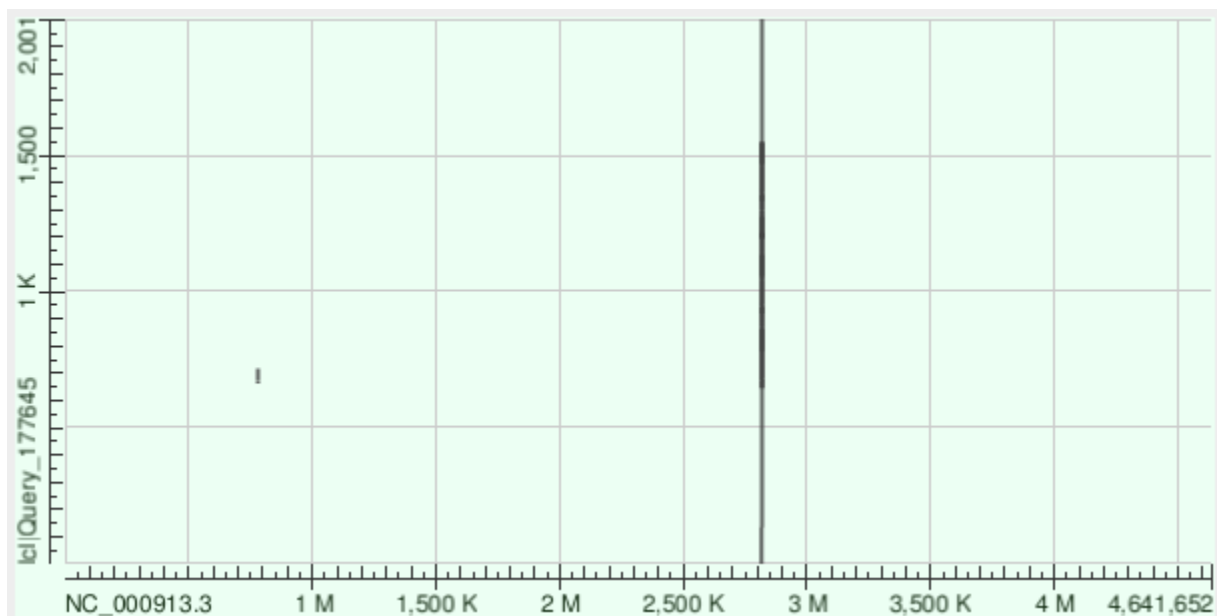
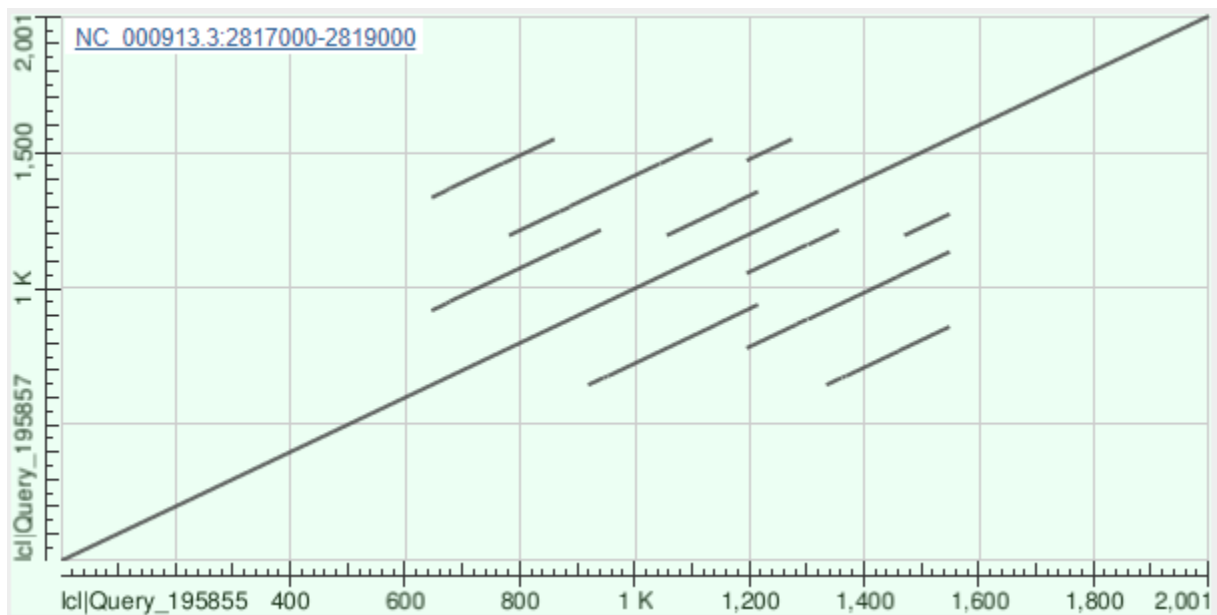

GTAGAGGTGGTGGTGGAACTCATTAGATCCCACAATGAACCGTTTTTGAGGAATTTGGTTGAGAAGGTGGCAAATGGCCCAAACAGCACTACGCGT
TTACCGTCCAGAACGCGGGTATCGATATGCGGAACCGACATCGGTGGTGCGCCAACGGATGCTTTACCGTAAACCTTCGCCAGATGGTGATTAACCA
CGTCCGGGTTTTCCGAAACAAGGAAGTGTCCGCCACCGGGAACCGGCGTAGTCTTTCGCTTCCGGAATCCCGATTCTGTAAACAGCTTCAGCGC
CGCACCGCCCGCGCCGATAAAGACAAATTTGGCACGAATGTTCTGTGCAGTGCCATTTTC



>NC_000913.3:4295000-4297000

TTCTTCCTGCCGCGACGCATCGTGAACCGCTGGTGACCGTGTTCCGCTCCATCTCTGAAACCATGTTTAAAGTGACTCACATGGTGATGCGTTAT
GCACCGTGGGTGTGTTTGCCTGATTGCGGTGACGGTGGCTAACTTTGGTTTCTCGTCTCTGTGGCCACTGGCGAAACTGGTGCTGCTGGTGCAATT
TCGCCATTCTGTTCTTCGCGCTGGTAGTGCTGGGAATTGTGGCGCGCTGTGCGGGTTAAGCGTCTGGATCCTGATTCTGTAAGATGAGCT
GATTCGGCGTACTCCACTGCCAGCTCTGAAAGCGTGCTGCCGCAATTATTGAGAAGATGGAAGCCTACGGAGCACCGGTGTCGATCACCAGTTTC
GTGGTGCCGACCGGTTACTCTTTTAACCTTGATGGTTCGACGCTGTATCAAAGTATTGCCGCTATCTTCATCGCGCAGTTGTATGGCATTGACCTGT
CCATCTGGCAGGAAATCATTCTGGTGCTGACGCTGATGGTGACCTCGAAAGGATTGCTGGCGTGCTGGCGTGTCTGTTGTGGTGTGGTGGCAAC
GCTGGGTAGCGTAGGTATTCCGCTGGAAGGTCTGGCGTTTATTGCTGGTGTGACCGTATCCTCGACATGGCGCGTACTGCGCTGAACGTGGTGGGT
AATGCGCTGGCGGTGCTGGTGATTGCCAAGTGGGAACACAAATTTGACCGTAAGAAAGCGCTGGCTTATGAGCGTGAAGTGCTGGGCAAAATTTGATA
AACTGCGGATCAATAATTGAAGATTGCCGGGGATATCCACCCGGCAATGTGTGAATGCCTGATGCGACGCTTGCCGCGTCTTATCAGGCCTACGCC
AGACAGCGCAATAGCCTGATTTAGCGTGATTTGTAGGTCGGATAAGGCGTTTATGCCGCATCCGACATCAACGCCTGATGCGACGCTTAACGCGTC
TTATCAGGCCTACGCCAGACAGCGCAATAGCCTGATTTAGCGTGATTTGTAGGTCGGATAAGGCGTTTACGCCGCATCCGACATCAACGCCTGATG
CGACGCTTGCCGCGTCTTATCAGGCCTACGCCAGACAGCGCAATAGCCTGATTTAGCGTGATTTGTAGGTCGGATAAGGCGTTTATGCCGCATCCG
ACATCAACGCCTGATGCGACGCTTAACGCGTCTTATCAGGCCTACGCCAGACAGCGCAATAGCCTGATTTAGCGTGATTTGTAGGTCGGATAAGGC
GTTTATGCCGCATCCGACATCAACGCCTGATGCGACGCTTAACGCGTCTTATCAGGCCTACGCCAGACAGCGCAATAGCCTGATTTAGCGTGATTTT
GTAGGTCGGATAAGGCGTTTACCGCATCCGACATCAATGCCTGATGCGACGCTTGCCGCGTCTTATCAGGCCTATCTTAACCGTTGGTTAATTTTTC
AACTCTTCACACCCGGTATCAACCCCTCCATACAGCTCAGGTTCAACCAGTGACGCGCCTTTGCTTGTTCTTCTCGATAAACCTTCTTCACCG
TTTAAGAACATCATTCCGCGCCAGTACTCGGAATAACCGGTTCGGGAAATTGCAGAGCTGCGTTTGAAATACCAGGTTGCCTTGTCATCGTCCGCT

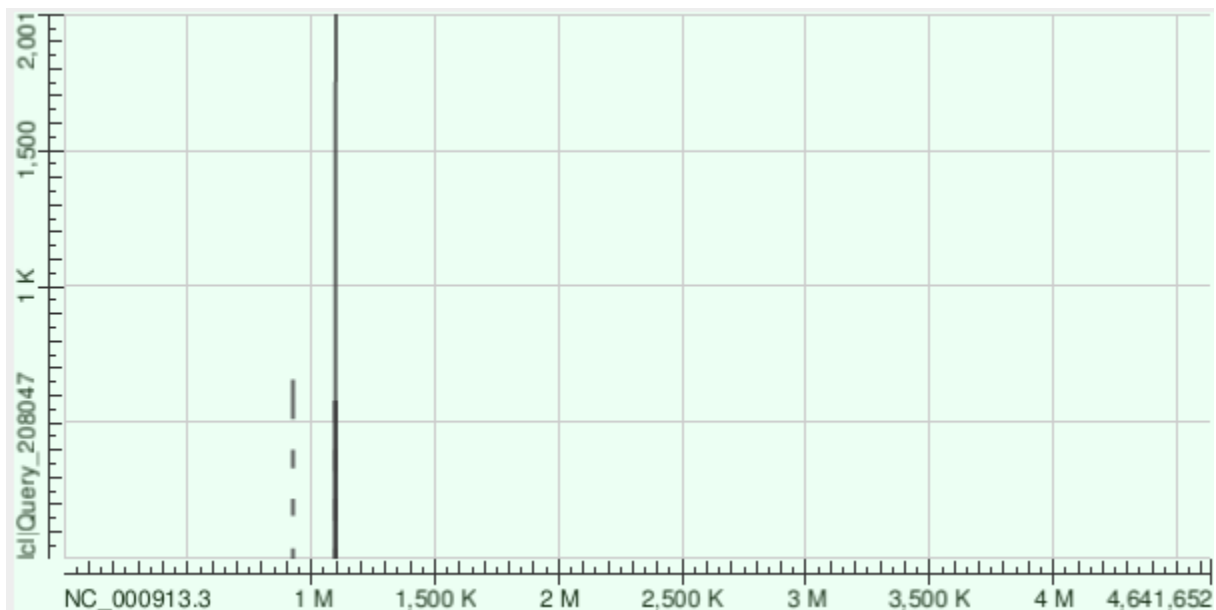
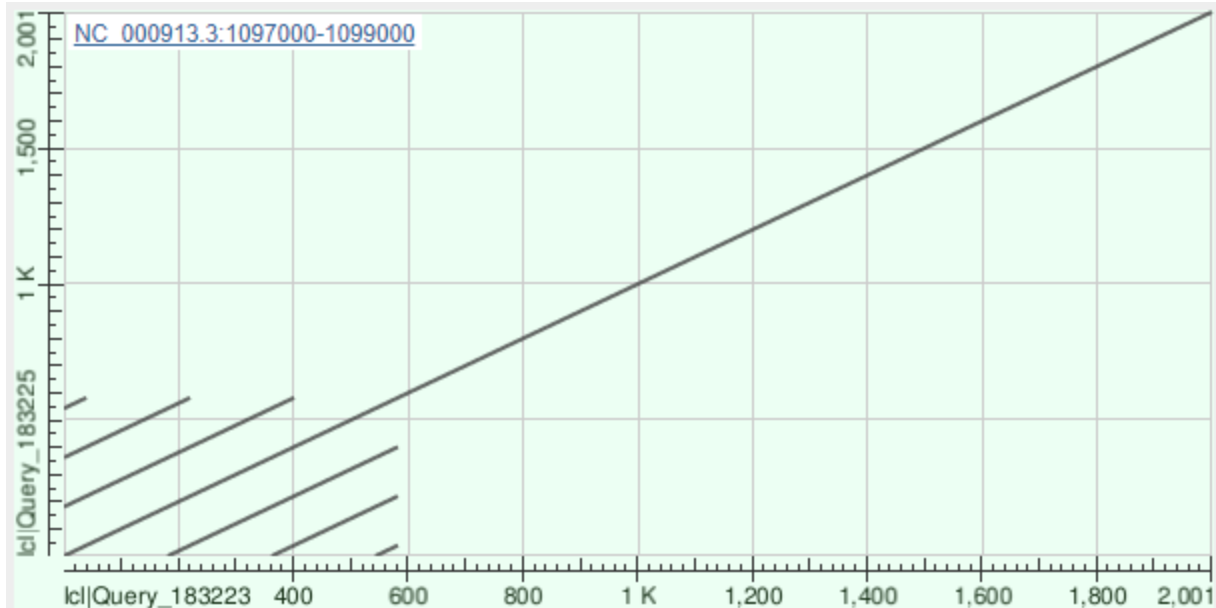
TAATGCCACGCCGTTGGCGTAAATCAAACCAAGCAGCATTTGGGCATCGACGGCAGAGTCGTTCTCCAGATCTTCCGAGGCGTTTTCTAACAGCGA
AATTGCCTTTGGATAATCCGGTTTACCCGCTTGAGTATTTACCAGAATATGCGCCAGCGTTACTTCACCTTCTTTACTGCCCCGCTTGCGCCGCTTTC
TCTGCAAGAACTTTTGTGTGGATAGTCCAGACTGACCGGATTGGTGATTTTGATCTGCGCCAGCAGCGCGCAGGCATCAGCATCACCGCTGTCTGG
CGGCTTCTGTGCCCAATATTGGCTTTGTCTCAAATCGCCGGAGCTAAACCAGCTGTCTGGC



>NC_000913.3:2817000-2819000

CATCTTCAAAGACCACACACTGCGTCGGTTGCACGCCCATACGCTGCGCGCACAAACAAAATGTGTCTGGCGCGGGTTTATGGTGTTTGACGTGATC
GGCAGCGCAGCAGCGCGTCAAATAATGGCGTAATCCCAGGTGCGCCAGCAATGCCTCAGCGATGGCGCTTTCACTCCCCGTTCCTACAGCCATTGGG
CGACGACCATGCCAACTTTTACCACATCAACAAGAGGAAGCGGTTTCGACGCTATCCAGCAGCATACTTCTTACTGCTTCTGTTTTTTCACGCGCTA
ACGCATGCGGGTGCAGATCGGCCTGATTCAGCTCAATAATTGCCTGAGCAATACGCCAGGTGGGCGATCCATTAAAGCGCAATCATCGCCTGAATATC
GTACTGAAGACCGTAGTGCCCTAATACTTCGCGCCACGCTTTACGGTGCGTAGGCTCCGTATCCAGGATTGTGCCATCCATATCAAAAATTAACCT
GCATAACGCTCGTACATGGTCTTCTCGCAAGTCGAAAAATATGACGCTACTTTAGCGTAATTGCTTGATTTTGTGCGTGATGGCGAAGGAGAAGAAA
ATTCAGGAAATGGATAAAGGAGTTATTTAGAAATGAGATATTTTGAAGGAAATTTTGGAGAAGATGGTGCATCCGGGAGGATGACTCACTGACG
TTCGCCCTTCGGGCCGTTGTCTGGCGCAACGTTATCCTTCTGGTGCTGCGGTTAATACCGCTATCTTTCGATACCTTCATTGCTGAAGGTTACGG

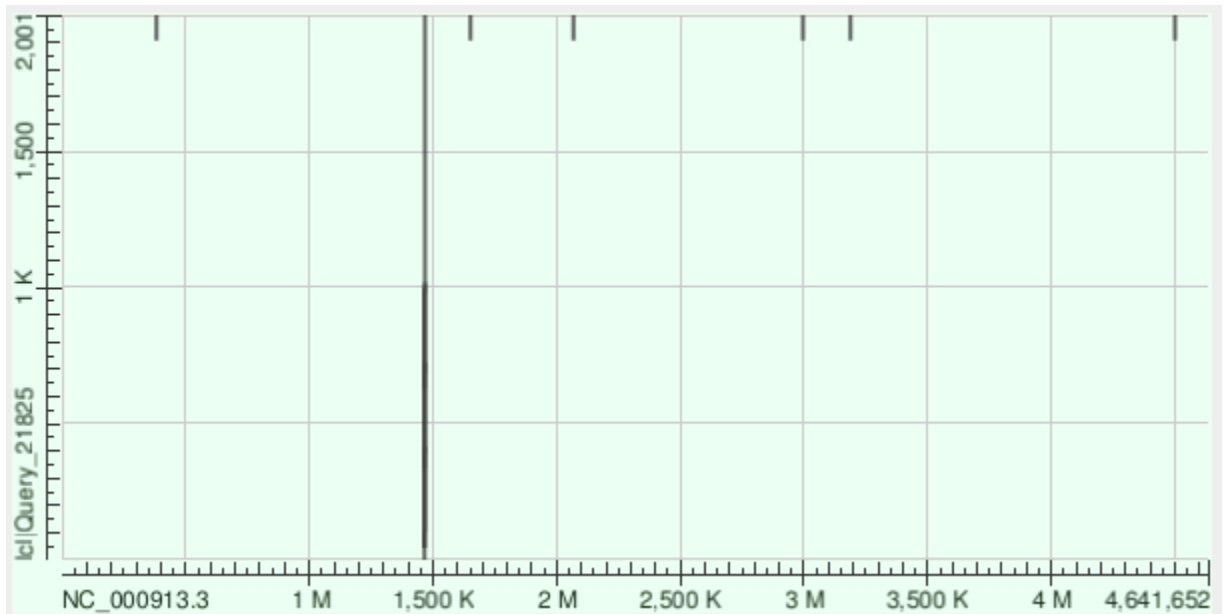
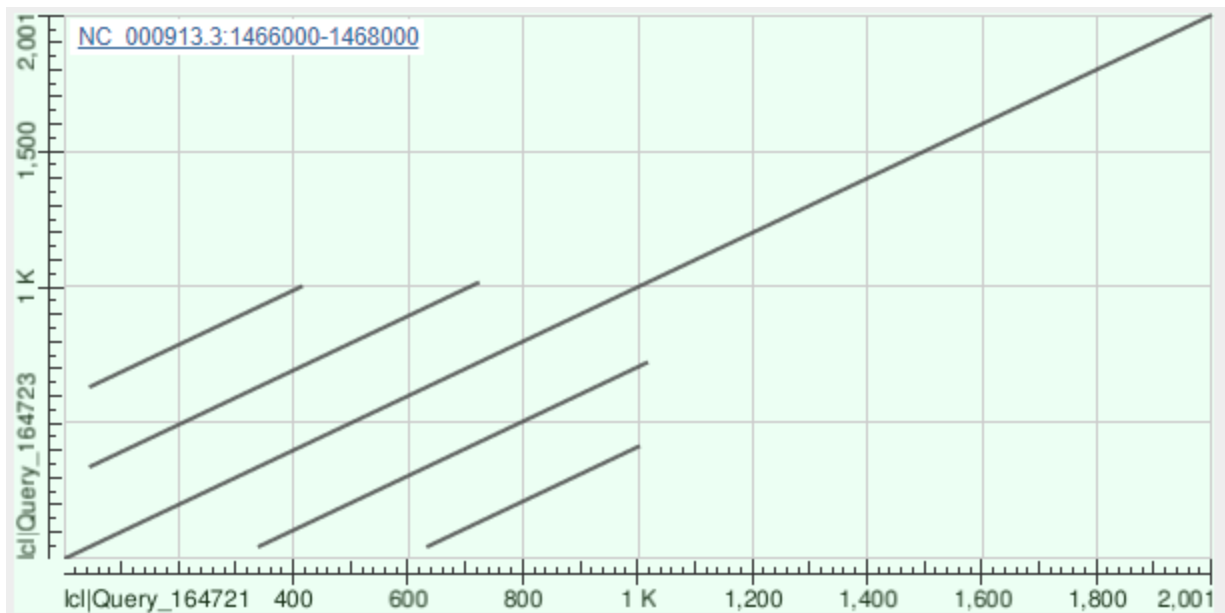
AGAATATGGTGCATCCGGGAGGATTCTGAACCTCCGACCGCTCGGTTTCGTAGCCGAGTACTCTATCCAGCTGAGCTACGGATGCATCGGGAAAACCTTA
TTCTACTGCTGATATTGATACCGCTACTAAAGCCATATCAATTAAGAGATGGTGCATCCGGGAGGATTACTCGCCTTTCGGGCTCGCCCTTCGGGCCG
TTGCTAAAGCAACGTTATCATCCCTAGTGCCTGCGGTTAATTCCGCTACCTCTTCGATACCTTCATTGCTGAAGATTTTCGGAGAATATGGTGCATCC
GGGAGGATTCTGAACCTCCGACCGCTCGGTTTCGTAGCCGAGTACTCTATCCAGCTGAGCTACGGATGCATCGGGAAATTTATTTTACTGCTGATATTG
ATACCGCTACTAAAGCCATATCAAGTAAGAGATGGTGCATCCGGGAGGATTCTGAACCTCCGACCGCTCGGTTTCGTAGCCGAGTACTCTATCCAGCTG
AGCTACGGATGCATCGGGAACTTACTTTACTGCAGATTTTTTGATACCGCTACTAAAGCCGTATCAAGTAAGAGATGGTGCATCCGGGAGGATTAC
TCGCCTTTCGGGCTCGCCCTTCGGGCCGTTGCTAAAGCAACGTTATCCTCCCTGGTGCCTGCGGTTATTACCGCCACCTCTCCATACCTTCATCGCT
GAAAGTACGTAGAAATATGGTGCATCCGGGAGGATTCTGAACCTCCGACCGCTCGGTTTCGTAGCCGAGTACTCTATCCAGCTGAGCTACGGATGCAAAAT
GGCGGTGAGGCGGGGATTCTGAACCCCGATGCAGCTTTTGACCGCATACTCCCTTAGCAGGGGAGCGCCTTCAGCCTCTCGGCCACCTCACCACACG
CCTCTTACGAGTGCTTCTGAAGAACTTGTTTATCGCTCATCGTCGCTGCGTGGCGCACATATTACTTTCTGAGACTTATAAGTCAAACAATTTTTCCT
ACACTTTTATCGTTTGCACACTTCACGTTCAATTAGTCTGTAAAAACGTCAAAAAGAGTGTTTTATCAACAGAAGAATGGAGGTCTGACAGATAGTA
GTAATGCAAAAAAATGGAGACTTAAGTTGAATGAACGGGAGTAAAGCGAAAAGACTATAGAGTGAAGGAGAAATTTTGAGGGTGCGTCTCACCGATA
AAGATGAGACGCGGAAAGATTAGTAACCTGGACTGCTGGGATTTTTTCAGCCTGGATACGCTG



>NC_000913.3:1097000-1099000

```
TACGAGCTCTTCTTTAAATATGGCGGTGAGGGGGGATTGACTCGCTTCGCTCGCCCTGCGGGCAGCCCGCTCACTGCGTTACGGTCTGTCCAAC
GGCTGTCGCCAGTTGTCGACCCCGGTGCGGGCTTCTCATCCCCCGGTGTGTGCAATACACGAAAAAAGCCCGTACTTTCGTACGAGCTCTTCT
TTAAATATGGCGGTGAGGGGGGATTGACTCGCTTCGCTCGCCCTGCGGGCAGCCCGCTCACTGCGTTACGGTCTGTCCAACGGTGTGCGCCAGT
TGTCGAACCCCGGTGCGGGCTTCTCATCCCCCGGTGTGTGCAATATACGAAAAAAGCCCGTACTTTCGTACGAGCTCTTCTTTAAATATGGCGG
TGAGGGGGGGATTGACTCGCTTCGCTCGCCCTGCGGGCAGCCCGCTCACTGCGTTACGGTCTGTCCAACGGTGTGCGCCAGTTGTGCAACCCCGG
TCGGGGCTTCTCATCCCCCGGTGTGTGCAATATACGAAAAAAGCCCGTACTTTCGTACGAGCTCTTCTTTAAATATGGCGGTGAGGGGGGGATT
CGAACCCCGGATACGTTGCCGTATACACACTTCCAGGCGTGTCTCTCAGCCACTCGGACACCTCACCAAATGTGCTTCTGTCTGTGGAACG
GGCGCTAATTTAGGGAATCATGACCTGAGGTCAACAACTTTTGA AAAAATCGCGCGTTTATTCAAACCTCAATCAATGTGTGGTTTAAATAAGC
GAAATCTGCTTTTTTTGCCACCGACACGGATTGTTATGCTGGTGGCCTTTGTAGATCATAACGATAAGTGCGAATAAATTCGCACAACGCTTTT
CGGGAGTCAGTATGGATATCATCTTTTATACCCAACGTTTCGATACCAATGGTGGATTGAGGCACCTGCGCAAAGCTATTCTCAGGCAAGAGTCAG
AGCATGGAAAAGCGGAGATAATGACTGTGCTGATTATGCTTTAGTCTGGCATCTCTGTTGAAATGTGGCAGGGCGCGATCTTAAAGCGGTGTTTC
GCACTCGGGGCCGGTGTGATTCTATTTTGTGCAAGCTACAGGCACACCCTGAAATGCTGAACCCCTCTGTTCACCTTTTTCGCTGGAAGATACCG
GTATGGGCGAGCAAAATGCAGGAATATGCTGTCAGTCAGGTGCTGCATTGGTTTCGACGTTTTTGACGATTATCGCATCCAGCAAAAATAGTTCGCATTG
GCAACCGTGCCTGAATATCATCGGGAAGATTTTACCATCGGCATTTTGGGCGCAGGCGTACTGGGCAGTAAAGTTGCTCAGAGTCTGCAACCTGG
CGCTTTCCGCTGCGTTGTGAGTTCGAACCCGTAAATCGTGGCTGGCGTGCAAGCTTTGCCGGACGGGAAGAACTGTCTGCATTTCTGAGCCAAT
GTCGGGTATTGATTAATTTGTTACCGAATACCCCTGAAACCGTCGGCATTATTAATCAACAATTACTCGAAAAAATTACCGGATGGCGCGTATCTCCT
CAACCTGGCGCGTGGTGTTCATGTTGTGGAAGATGACCTGCTCGCGCGCTGGATAGCGGCAAGTTAAAGGCGCAATGTTGGATGTTTTTAATCGT
GAACCCCTTACCGCTGAAAGTCCGCTCTGGCAACATCCACGCGTGACGATAACACCACATGTGCGCGGATTACCCGTCCCGCTGAAGCTGTGGAGT
```

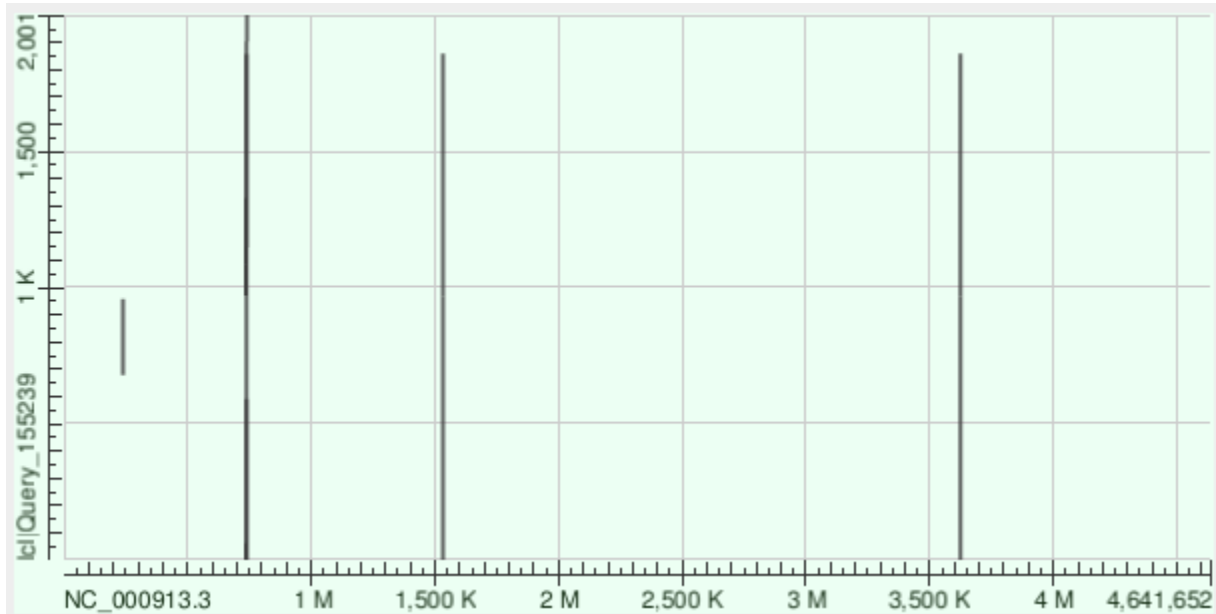
ACATTTCTCGCACCATTGCCAGCTCGAAAAAGGGGAGAGGGTCTGCGGGCAAGTCGACCGCGCACGCGGCTACTAATAAAGCATCAGGATTCCTGC
TATCCTTGGCGGGAATTGAATACAGGAGAGAGTTATGTATCCCGTCGACCTTCATATGCATACCGTTGCCAGCACACATGCATATAGCACATTAAGT
GATTACATTGCCAGGCCAAACAAAAGGCATTAACTTTTTGCGATCACCGATCATGGCC



>NC_000913.3:1466000-1468000

```
ATGGCGCAACGAAGTGGCAATATAATCACAACGGTGAGCTTGTATTACGGGTGATAATGCCACAGTAAACAACAATGGCAAAACCACGGTTGACGG
TAAAGATTCCACCGGTACGGAATCAACGGTAATAACGGGAAAGTGATTCAGGACGGCGATCTGGATGTCAGCGGCGGCGGTACGGTATTGATATC
ACCGGTGACAGCGCGACGGTGGATAACAAGGGCACCATGACCGTCACCGATCCGGAGTCCATGGGTATCCAGATCGACGGTGACAAGGCCATCGTCA
ATAACGAAGGCGAGAGCACCATCACCAACGGTGGCACCGGCACCCAGATTAATGGTGATGACGCCACGGCAAAACAACACGGCAAAACCACGGTGA
CGGCAAGGATTCCACCGGTACGGAATCAACGGTAATAACGGGAAAGTTATTAGGACGGCGATCTGGATGTCAGCGGCGGCGGTACGGTATTGAT
ATCACCGCGACAGCGCAACGGTGGATAACAAGGGCACCATGACCGTCACCGATCCGGAGTCCATCGGTATCCAGGTTGACGGCGACACGGCGGTTG
TTAACAACGAAGGCGAGAGCGCCATCACCAACGGTGGCACCGGCACGCAGATTAACGGTGATGACGCCACGGCAAAACAACACGGCAAAACCACCGT
TGACGGCAAGGATTCCACCGGTACGGAATTTGCTGGCAATAACGGGAAAGTGATTCAGGACGGCGATCTGGATGTCAGCGGCGGCGGTACGGTATT
GATATCACCGGCGACAGCGCAACGGTGGATAACAAGGGCACCATGACCGTCACCGATCCGGAGTCCATCGGTATCCAGATTGACGGCGACACGGCCA
TCGTCAATAACGAAGGCGAGAGCACTATCACCAATGGCGGCACCGGCACTCAGATCAACGGTAACGACGCCACCGCGAATAACAGTGGGAAAAACCA
TGTTGATGGAAAAGATTCCACCGGTACCAAAATCGCGGCAATATCGGCATTGTAATCTGGATGGTAGCCTGACTGTTACAGGCGGTGCGCATGGT
GTTGAGAACATTGGTGACAACGGCACGGTTAACAACAAAGGAGATATTGTTGTTTCCGATACTGGATCGATTGGCGTGCTCATCAACGGTGAGGGGG
CAACAGTATCCAATACGGGTGATGTTAACGTTAGCAATGAAGCGACAGGGTTCAGCATCACACCAACAGTGGGAAGGTTTCGCTGGCAGGCAGTAT
GCAGGTTGGCGATTTCTCGACCGGGGTAGATCTTAATGGCAACAATAACAGCGTGACGCTGGCGGCAAAAGATCTAAAAGTGGTCGGGCAGAAAGCG
ACGGGCATAACGTTTCTGGCGATGCGAATACAGTGAATATCACTGGTAACGTTCTGGTTGATAAGGATAAAACCGCAGACAAATGCGGCGGAATATT
TCTTCGATCCATCCGTGGGTATCAACGTTTACGGCAGTGATAATAACGTGACGCTGGATGGAAAGTTAACTGTTGTATCAGACAGTGAGGTTACTTC
TCGTGAGAGTAATTTATTTGATGGCAGCGCAGAGAAAACGTGAGGTCGTTGTTGATTGGCGATGGCAATACCGTTAATATGAATGGTGGACTTGAA
```

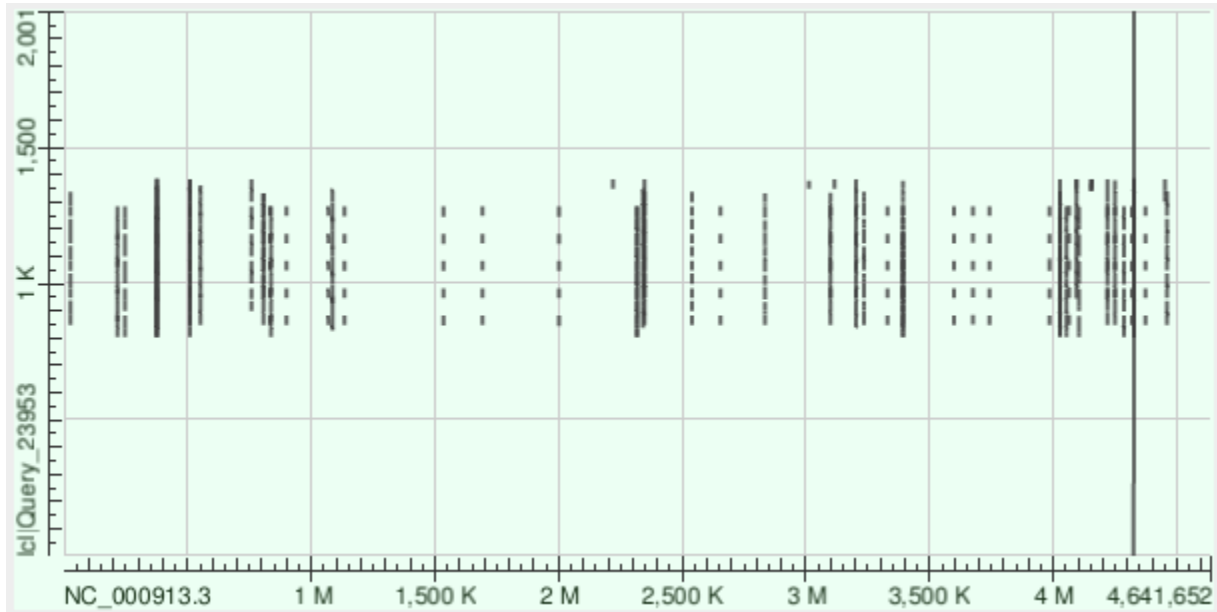
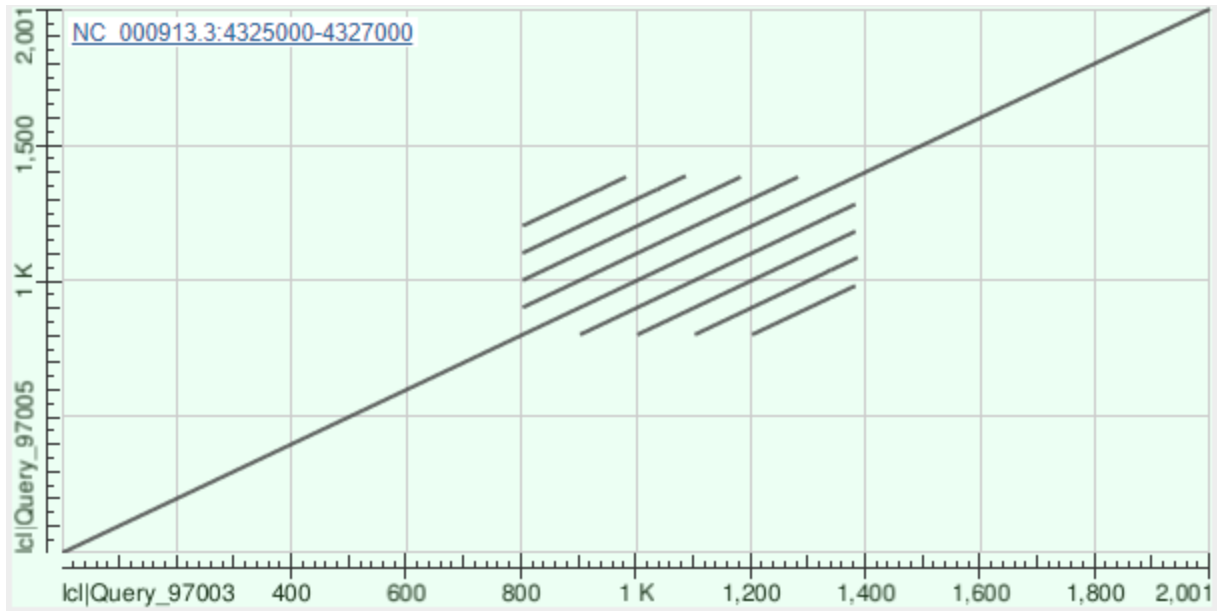
CTGATTGGAGAGAAAAACGCGCTTGCAGATGGGTCGCAGGTTACTTCCTTGCGCACAGGATATAGTTATACCAGCGTTATTGTCGTTAGTGGTGAGT
CGTCGGTATATCTGAATGGAGATACGACAATCAGCGGAGAATTCCCTCTGGGGTTTGCCGGGGTTATTCGGGTACAGGATAAAGCTTTGCTGGAAAT
TGGCAGTGGCGCTACGCTAACAAATGCAGGATATTGACAGTTTTGAACATCATGGGACAAGAACCCTGGATTTGCCCCATATTTCAGACATCTGTT
ATCACTTAACCCATTACAAGCCCGCTGCCGCAGATATTCCCGTGGCGAGCGATAACCCAGC



>NC_000913.3:737000-739000

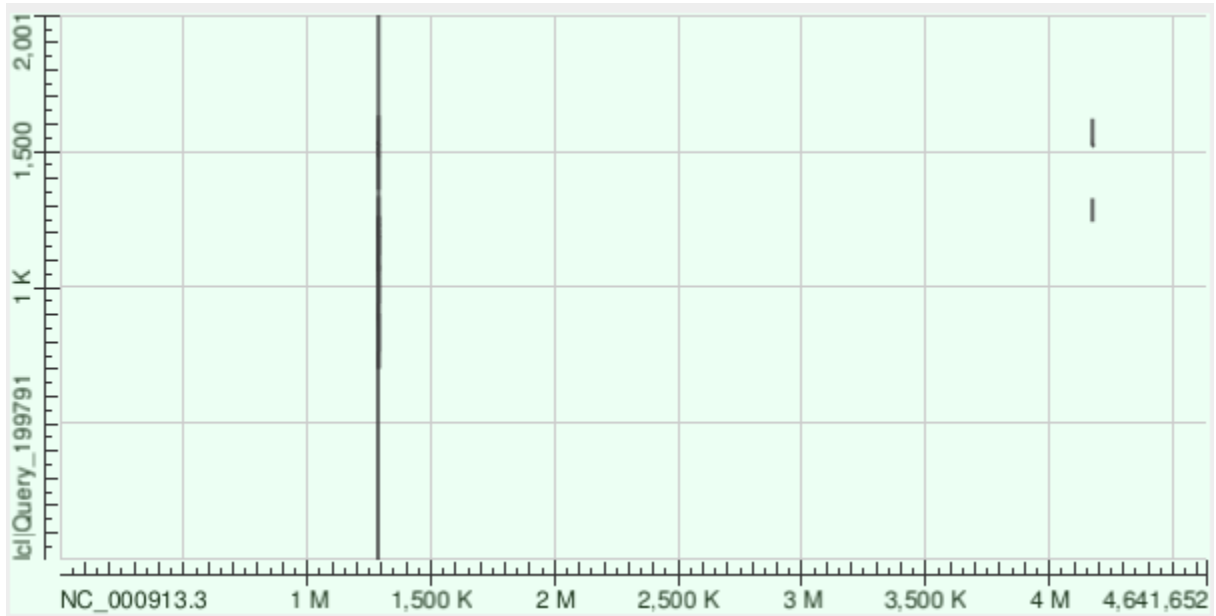
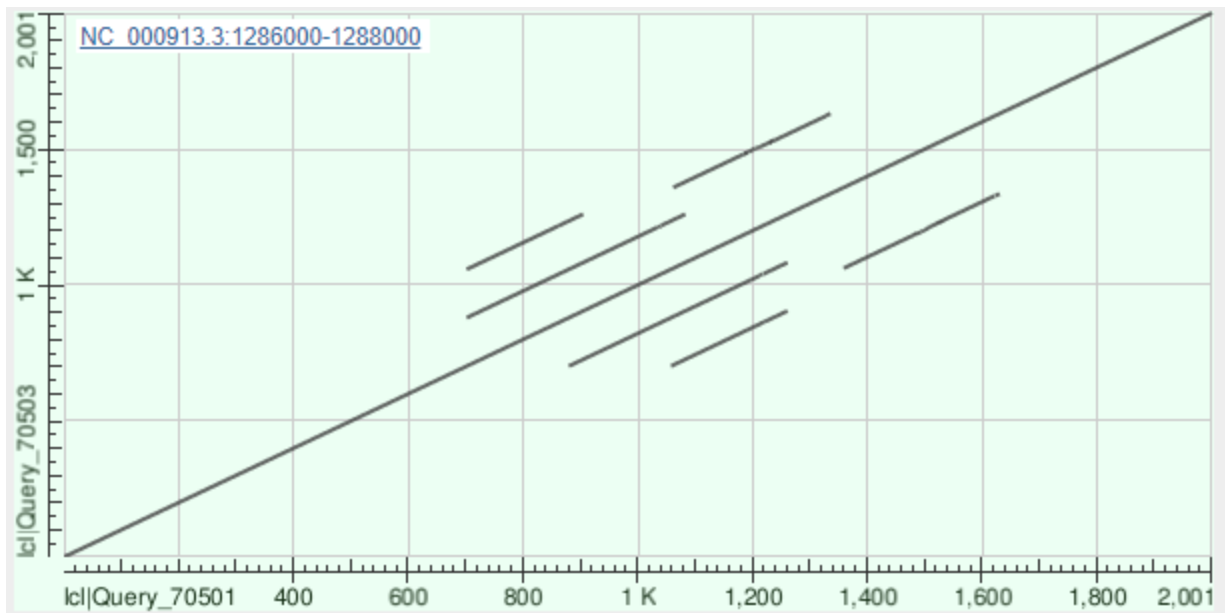
CAATATGGTGATTTTGAAAATGGTATTCTGTTACGATACTATTGCCAGAGTTGTATCTGTATCAGTCCTGCAAAATTTACGAGTGCTTTATTA
 ACTGGATGCGTGATTGCCATTCATCAAAATGATAAAGACGTCATTGCAATTGATGGAAAAACGCTCCGGCACTCTTATGACAAGAGTCGCCGAGGGG
 AGCGATTTCATGTCATTAGTGCGTTCTCAACAATGCACAGTCTGGTTCATCGGGCAGATCAAGACGGATAAGAAATCCAATGAGATCACAGCTATCCCT
 GAACTTCTTAACATGCTGGATATTAAGGAAAAATCATCACAACCTGATGCGATGGGTTGCCAGAAAGATATTGCAGAGAAGATACAAAAACAGGGAG
 GTGATTATTTATTCGCTGTAAAAGGAAACAGGGGCGGCTTAATAAAGCCTTTGAGGAAAAATTTCCGCTGAAAGAATTAAATAATCCAGAGCATGA
 CAGTTACGCAATGAGTGAAGAGTACGGCAGAGAAGAAATCCGTCATATTGTTGCGATGTCCTGATGAACCTATTGATTTACGTTTGA
 TGGAAAGGACTGAAGAAATTATGCGTGGCAGTCTCCTTTCCGTTCAATAATAGCAGAACAAAAGAAAGCCAGAAATGACGGTCAGATATTATATCA
 GTCTGTGATTAAACCGCAGAAAAGTTCGCCACAGCAATCCGAAACCACTGGCACGTTGGAGAATAAGCTGCACTGGCGTCTGGACGTTGGTAATGAA
 TGAAGACGACTGCAAAATAAGAAGAGGAAACGCCGAGCAATATTTTCAGGGATACGGCACATCGCTATTAATATTTTAACGAATGATAAGGTATTC
 AAGGCAGGTTAAGACGTAAGATGCGAAAAGCAGCCATGGATAGAACTATCTCGCGTCAGTCCTTGCAGGAGCGGGCTTTGTAATCTTTCCCTG
 CTTTTGTACTAAATAATTCGCATTTTATGTTTAAAAATTGAGATATTCCTTATTACCTGAAGCTGTTTTTTATTGCTTATACATGATCAAACTCT
 CTACATAATTAAGGAGAACAAAATGGAACCTAAAAAATTGATGGAACATATTCTATTATCCCCGATTACAGACAAGCCTGGAAAGTGGAAACATAA
 ATTATCGGATATTTCTACTGTTGACTATTTGTGCCGTTATTTCTGGTGCAGAAGGCTGGGAAGATATAGAGGATTTTGGGGAACACATCCCGATTTT
 TTGAAGCAATATGGTGATTTTGGAAAATGGTATTCCTGTTACGATAACCATGCCAGAGTTGTATCCTGTATCTGTCTGCGAAATTTTCATGAGAGCT
 TTATTAACGAGTGTGTTGACTACCATTTCTTCAGATGATAAAGACGTCATCGCAATTGATGGAAAAATACACCGGCATTCTTATGACAAGAGTCGCCG
 TAAGGAGCGATTTCATGTCATTAGTGCGTTCTCAACATGCACAGTCTGGTTCATCGGACAGATCAAGACGGATAAGAAATCCAATGAGATCACAGCT
 ATCCCTGAACTTCTTAACATGCTGGATATTAAGGAAAAATCATCAAAACCGATGCGATGGGTTGCCAGAAAGATATTGCAGAGAAGATACAAAAAC

AGGGAGGTGATTATTTATTCGCTGTAAAAGGAAACCAGGGGCGGCTAAATAAAGCCTTCGAGGAAAAATTTCCGCTGAAAGAATTAAATAATCCCAA
GCATGACAGTTACGCAATTAGTGAAAAGAGTCACGGCAGAGAAGAAACCCGTCTTCATATTGTTGCGATGTCCCTGATGAGCTTATTGATTTACG
TTTGAATAGAAAGGGCAGCTATGGTTAGAAACTACCTGACGTCAGTCCTTGCGGGGAGCAGGCTTCGTAAATTTGTCCTGCTACAACAGGATTAC
TTCACAAATATCATTCTCAACGCTACACTTACTCCTGTAAACCGCTCAGGAGCAGTAAT



```
>NC_000913.3:4325000-4327000
GTAATCAAACCGCTTAAGTGACGTAAAAGGGTGGATTTCGGAACCCGACGGCCCAAGCAGAGCCACCATTTCACCGTGATGAATGTTTCAGATCAA
CCGCATGCAGCGCTGATGCTGATTGAAGGTTTGGCGAGCTTCTCGACACGGATAATCGTTTGCATGATGCAGCCTCCCTAAAAAGTGGCTCCAT
CGTGGCGGATTATGTGACGATTGGTTAATTAAGTGACTAACAGATGAAGAGTTAACGGGAATTCGATGACAGTGCGGGAGGGCGGGCCCTCCCG
GCGGGTTATTGCGTTGGTTGTTGTTGACGACATTAATCATCCACGGTACGCCAAATTTATCGGTGACTTTGCCAAAGCCATGCGCCCAAGAAAGTTT
CCTGCCAGGCCATTCGATTTTCCGTTAGCGGCAAGATTGTCAAACACAGCGTTTTCCTTCTTCGACCTGTTGCGAATCGAGCACCGGTAAAGCC
GGAGTAGTGGCTTTTCTGACGGCATGGCATCGCTCATCATGATGTGCGTTCCGGCAATGCGCACATTTGGCGTGGGCGATGCGGGTATCGGGAAAT
TGCATTCCGGAAGGGCAGTTCTCGGCGCTGTCTGCGCTGATTTTGGCATTTCGCGCAAGCTGATTTTATAGAGCAGTTCCGCGCCCAACGTACGTT
GATAATAGGCAATCGCGTCGGAACAGTTACCGGCAAGAGAGGTAGGGAATTAACGGCATGATGGTGACCTCAGTTAAGAGAGCCAGTTAAGTGT
AGTTCGAAATTTATACAGATGAGAGGCGGCTGATAAGACGCGCGAGCGTCGCATCAGGCAGTCGGCACTGTTGCCGATGCGGCGTAAACGCCTTA
TCCGGCCTACGGAGGGTTCGGGAATTTGTAGGCCTGATAAGACGCGCAAGCGTCGCATCAGGCAGTCGGCACCATTTGCCGGATGCGGCGTAAACGCC
TTATCCGGCCTACGGAGGGTTCGGGAATTTGTAGGCCTGATAAGACGCGCAAGCGTCGCATCAGGCAGTCGGCACCGTTGCCGGATGCGGCGTAAAC
GCCTTATCCGGCCTACGGATGGTTCGGGAATTTGTAGGCCTGATAAGACGCGCAAGCGTCGCATCAGGCAGTTGGCACTGTTGCCGGATGCGGCGTA
AACGCCTTATCCGGCCTACGGAGGGTTCGGGAATTTGTAGGCCTGATAAGACGCGCAAGCGTCGCATCAGGCAGTCGGCACCGTTGCCGGATGCGGCGT
GTAAACGCCTTATCCGGCCTACGGAGGGTTCGGGAATTTGTAGGCCTGATAAGACGCGCAAGCGTCGCATCAGGCAGTCGGCACCGTTGCCGGATGCG
GGCGGAGCGCCTTATCCGGCCTACCGATCACAATACAATCAGTTCCTTTTTCACAACTCAGATTTTCAGTTTCATCGGACCAAAACCGTCGATTTTG
CAATCGATGTTTATGGTTCGCTTCAACCAGGCGGATGTTTTCACTTTGGTGCCAATTTTCAGCATCGAAGAGCTACCTTTACCTTCAGATCTTTAA
TGATGGTAACGCTGTGCGCGTCAGCCAGCAGATTGCCGTTAGCATCTTTAACGATCAGCTCGTCGCTTCTGTGCAGGTTCTGCGTCGTTCCATTTC
```

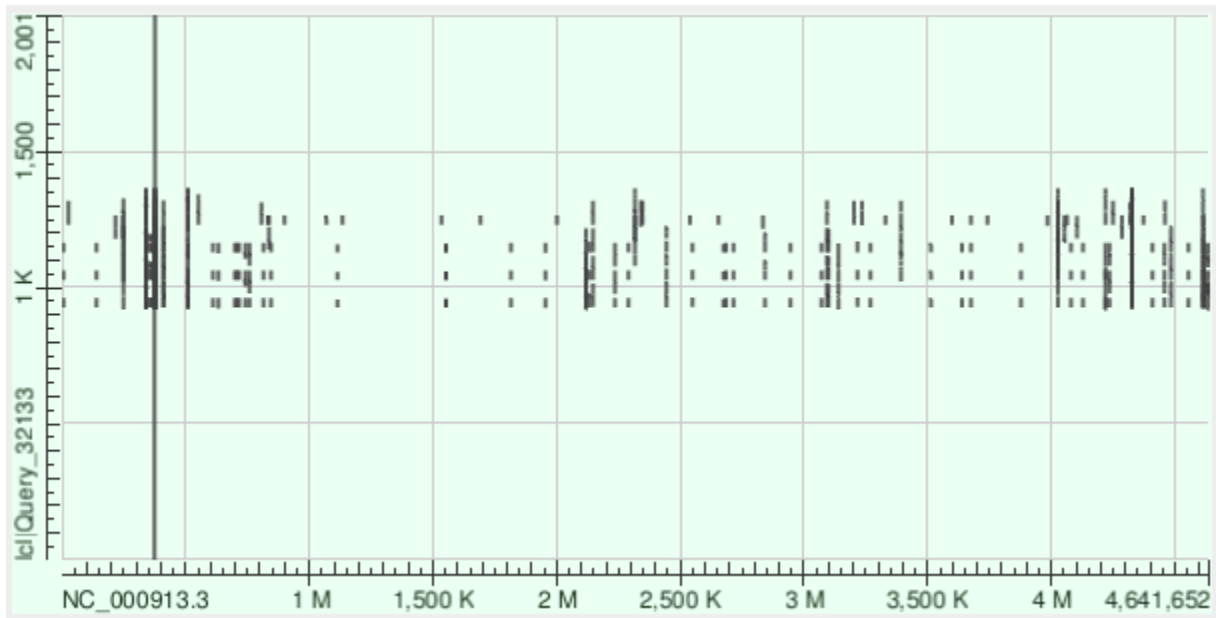
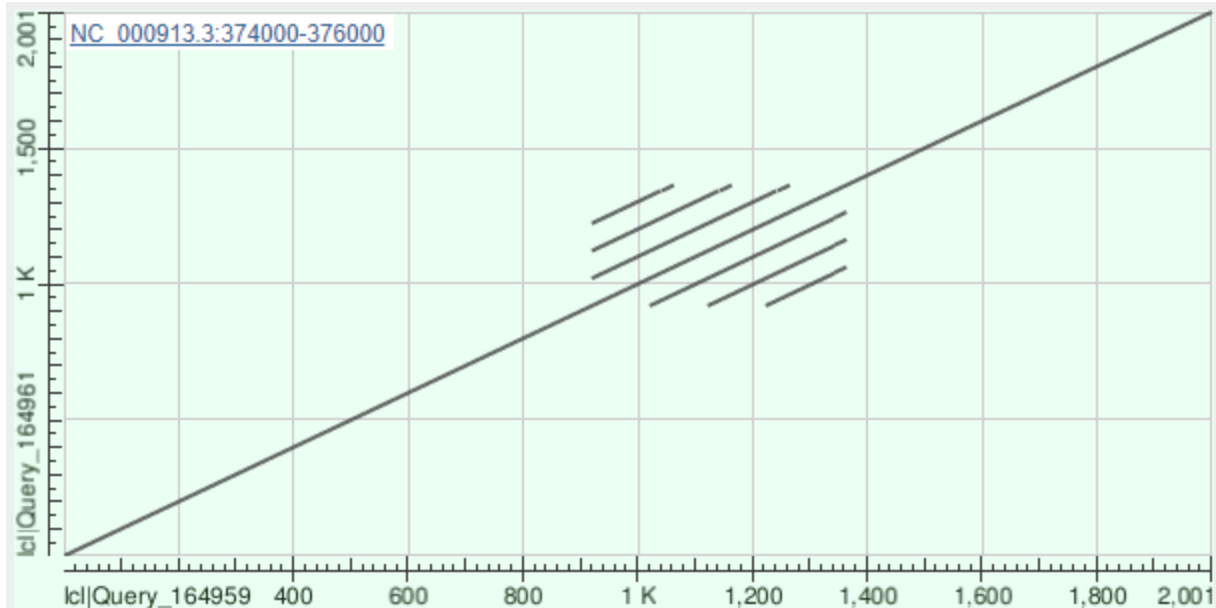
GTAGGCACATTCCGGGCAGATGTACATGCCGTTATCTTCGTAGGTGTATTCGGAGTTGCATTTGGGCAGTGTGGTAATGACATGTGGATTTCTCA
AAAGTCAGCAGGGCAAAGCGCCAAAAACGGCAGATTGCCGAAAAAGCCGCAATTATACACAAAATCCCTGAGTTTGTCTGGGGGTACTGCTGATT
TTAGACGACGATCGTTTTTCATTTGAAAGGAAACGCAAAAAGAAAATAATCTTTTGCCTCTATTTATTATTTAAGATAATGTTAATGCTCTACTTGT
TCTTTTTTTAATTCACATGATATTTTTATCTTTCTGACGTTTTAAACGTATCAGAAGACGT



>NC_000913.3:1286000-1288000

```
TCGAACCTGTTCCATATCGGGATTCTGGGGATTTTGTTCGGTCACTTCTTCGGTATGCTGACGCCGCACT
GGATGTATGAAGCCTGGCTGCCGATTGAAGTGAAACAGAAAAATGGCAATGTTTGCTGGTGGTGCCAGCGG
CGTGCTGTGTCTGATTGGCGGCGTGCTGTTGTCTGAAACGTCGCTGTTCAGCCACGCGTGCGTGCAACC
ACTACCGGAGCGGATATCCTGATCCTGTCGCTGCTCGTTATCCAGTGC GCGCTGGGCCTGTTGACCATTC
CGTTCTCCGCTCAGCATATGGACGGTAGCGAGATGATGAACTGGTTGGCTGGGCGCAGTCGTTGGTGAC
CTTCCACGGTGGCGCTTCTCAACACCTCGATGGTGTGGCGTTTATCTCCGCTGCACTGGTGGTGGG
ATGACGTTATTCCTGCTGTTCCCGTTTCTCGCTCTGATACACATCTGGAGCGTACCGTGAGTATCTGA
CACGTAAGTACCAGCTGGTGCGCGCTCGTCACTAAGCGAATTTAGTTCACATAGACCCTGCTTCGGCGG
GGTTTTTTTATGGGCACGGTGCGGGGTGAGTTGTGGATGCGCTTCGCTTATCCGACCTACAGGGGAGGA
TATCCAGGCGCGCTTAAACGTGGTGCCATCGGGCAAAATTAGCTCGGAGTAACAGGTTTGATTATTAA
GGAAGCGATGGTGGTGGGGGAAGGATTACTCAGCGCTGCGCGCTTCGCCCTTCGGGTCGTTGCCTGCGGC
AACGCTCTCTCGCTGGCGCTCAAGTCGAACCTTGGTTCGAAGCTTCTCATCCTTCCCCGCTTGGGCAGAAT
ATTTGATTGCGGATTCTGTTGAGAATTCGGGGGCTTTTGAAAGTGATGGTGGTGGGGGAAGGATTACTCA
GCGCTGCGCGCTTCGCCCTTCGGGTGCTTGCTGCGGCAACGCTCTCTCGCTGACGCTCGAGCCGAACCT
TAGTCGAAGCTTCTCATCCTTCCCCGCATGGGCAGAATATTTGATTGCGGATTCTGCTGAGAGTTCAGGG
ACTTTTGAAAGTGATGGTGGTGGGGGAAGGATTACTCAGCGCTGCGCGCTTCGCCCTTCGGGTCGTTGCC
TGCGGCAACGCTCTCTCGCTGGCGCTCGAGTCGAACCTTGGTTCGAAGCTTCTCATCCTTCCCCGCATGGG
```

CAGAATATTTAATTGCGGATTTCGTTGGGAAGTTCAGGGACTTTTGAAAGTGATGGTGGTGGGGGAAGGAT
TCGAACCTTCGAAGTCGATGACGGCAGATTTACAGTCTGCTCCCTTTGGCCGCTCGGGAACCCACCAGG
GGTAATTCAAATTTTGAGGTAATGCTTGAGATGGTGGTGGGGGAAGGATTATTCGTCGCTTCGCTCCTCA
CCCTTCGGGCCGTTGCCTGTGGCAACGTTCTCTCGCTTCGCTCGAATCGAACCTTAGTCGAAGGTTCTC
ACCCTTCCCCGATGAGTGCAAACCTTTCACAATCTCACCGAAGTTACCACATCGCTGTGGTGAATTATGGT
GGTGGGGGAAGGATTCGAACCTTCGAAGTCGATGACGGCAGATTTACAGTCTGCTCCCTTTGGCCGCTCG
GGAACCCACCACGGGGTAATGCTTTTTACTGGCCTGCTCCCTTATCGGGAAGCGGGGCGCATCATATCA
AATGACGCGCCGCTGTAAAGTGTACGTTGAGAAAAATGAACTGGTTGCGTAATTTTCATCCGTAAACGGA
TTAAAGGTAACCAGTTATTTTTGCTGGCGATTAAAGAATAATCGTTCGATTACCGTAAACAAAGACGCGC
TGTGCCAGTACTTTGTATAGTGACGACTTAAGACGTTTTTCTCGACGTCACGACCTGCGCGCATCATAT
CTTCAGCTGTGTAGGTATGATCGACATGAATAACGTCCTGCATGATGATTGGGCCTTCGTCCAGATTGTC
ATTCACATAGTGAGCGGTTGCGCCAATAATCTTCACACCAC



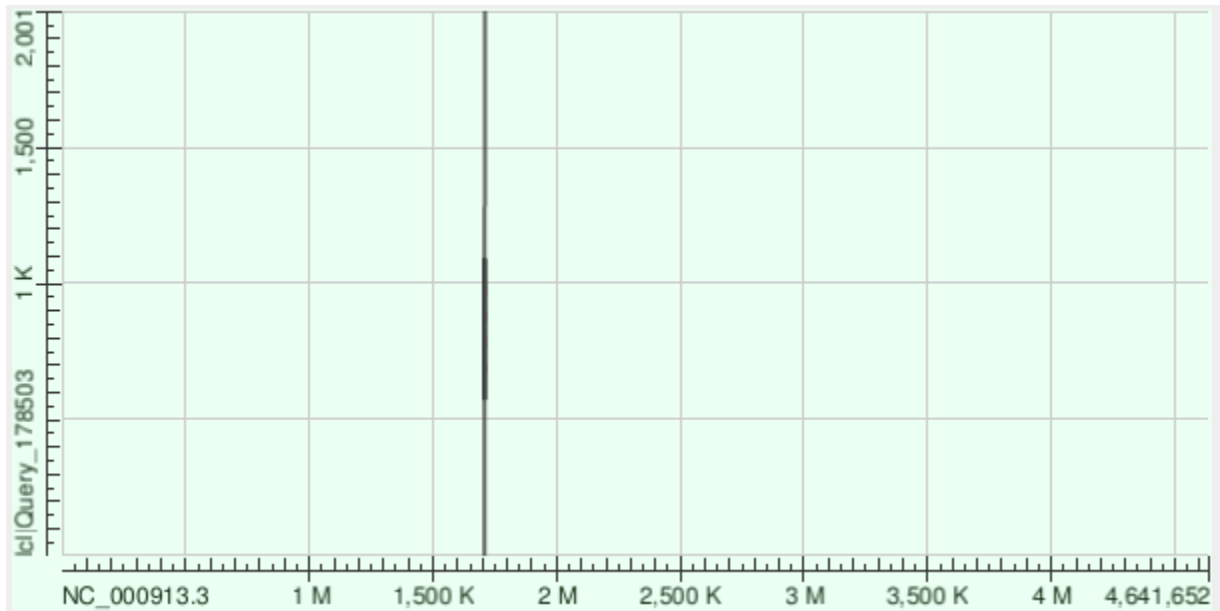
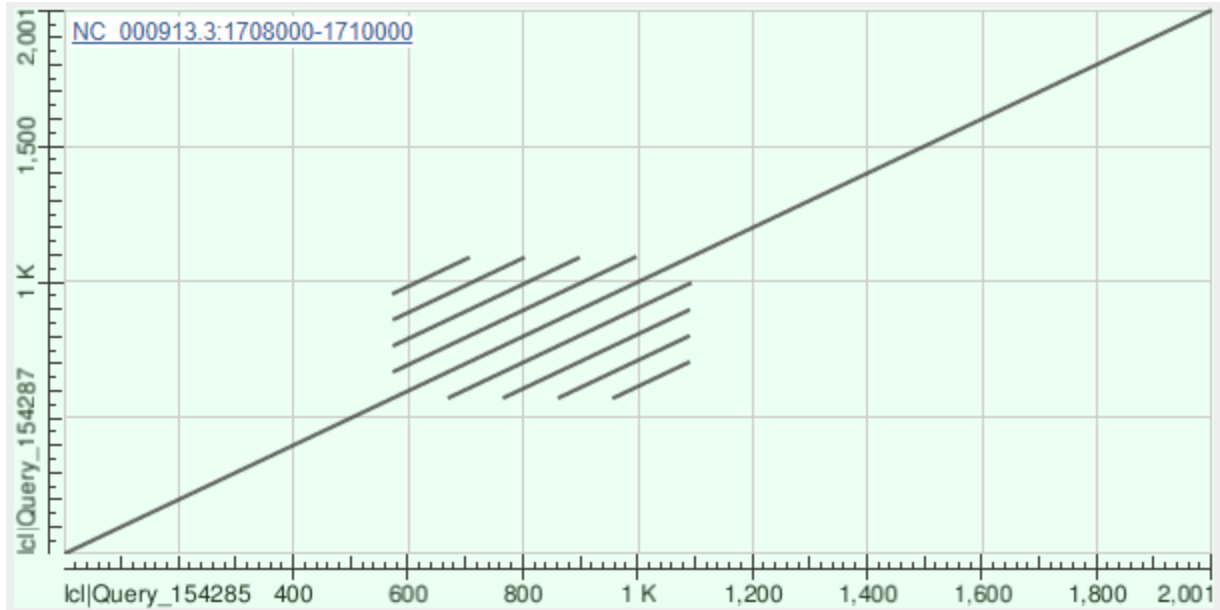
>NC_000913.3:374000-376000

```

AGTGGCCACGGCGACGGTTTGCAAGGTTCCAGCTTTAACTATGGTTTCGGCGCACATAGCGACCTTGAA
TGGATTGAAGCGGCGGCGGATGTGGTGAAGCACGCCAAAATCGCGACGTTGTTGCTGCCAGGAATCGGCA
CTATTCACGATCTGAAAAATGCCTGGCAGGCTGGCGCGCGGGTGGTTTCGTGTGGCAACGCACTGTACCGA
AGCTGATGTTTCCGCCCAGCATATTCAGTATGCCCAGCTCGGAATGGACACCGTTGGTTTCTGATG
ATGAGCCATATGACCACGCCGAGAATCTCGCCAAGCAGGCAAAGCTGATGGAAGGCTACGGTGGCAGCT
GTATTTATGTGGTGGATCTGGCGGTGCGATGAACATGAGCGATATCCGTGACCGTTCCGCGCCCCTGAA
AGCAGAGCTGAAACCAGAAACGCAAACTGGCATGCACGCTCACCATAACCTGAGTCTTGGCGTGGCGAAC
TCTATCGCGGCGGTGGAAGAGGGCTGCGACCGAATCGACGCCAGCCTCGCGGGAATGGGCGCGGCGCAG
GTAACGCACCGCTGGAAGTGTATTATGCCCAGCGGATAAACTGGGCTGGCAGCATGGGACCGATCTCTA
TGCCTTAATGGATGCCGCCGACGACCTGGTGCGTCCGTTGCGAGATCGACCGTACGAGTCGATCGCGAA
ACGCTGGCGCTGGGATACGCTGGTGTCTACTCGAGCTTCCTGCGTCACTGTGAAACGGCGCGGCGCGT
ATGGCTTAAGTGGTGGATATCTCGTTGAGCTGGGCAAACGCCGATGGTTGGCGGCCAGGAGGATAT
GATCGTTGACGTGGCGCTGGATCTGCGCAACAACAAATAATGATGACTGCCGAGAACGTGCATTTGTAT
GCCATCCCAGCACGACTGCCGGATGCGGCGTGAACGCCTTATCCGGCTACGGATGGCGTGAGAATTG
TAGGTCTGATAAGACGCGTTAGCGTCGCATCAGGCATCTGCGCACGACTGCCGGATGCGGCGTGAACGCC
TTATCCGGCTACGGATGGCGCGGGAATTTGTAGGCTGATAAGACGCGTTAGCGTCGCATCAGGCATCT
GCGCACGACTGCCGGATGCGGCGTGAACGCCTTATCCGGCTACGGGTGGCGCGAGAATTTGTAGGCTG

```

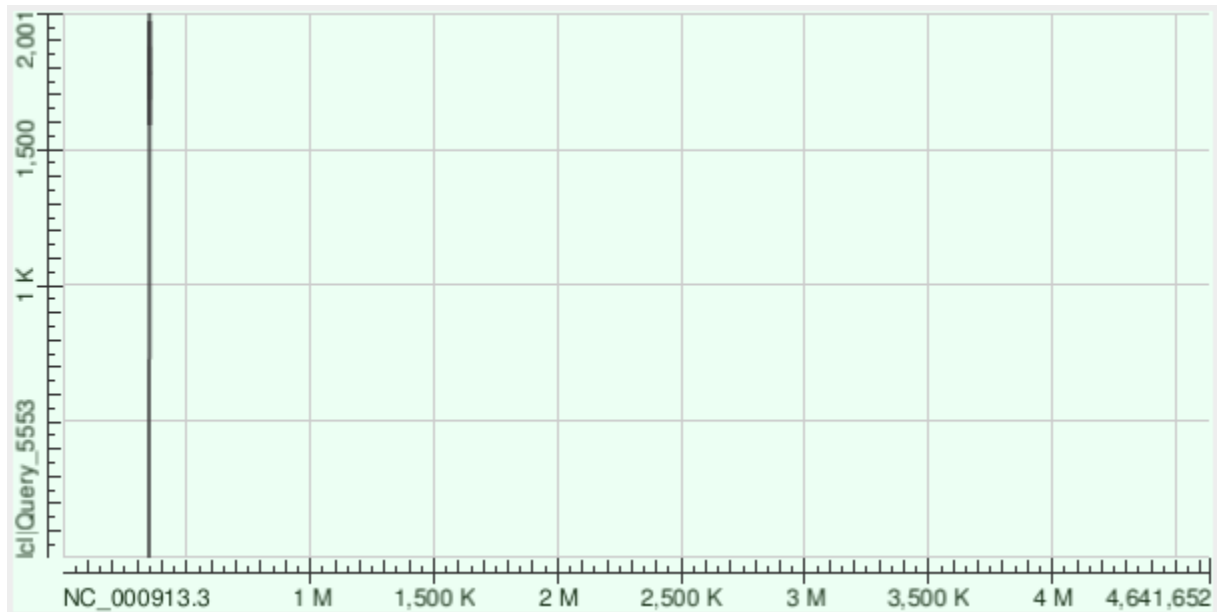
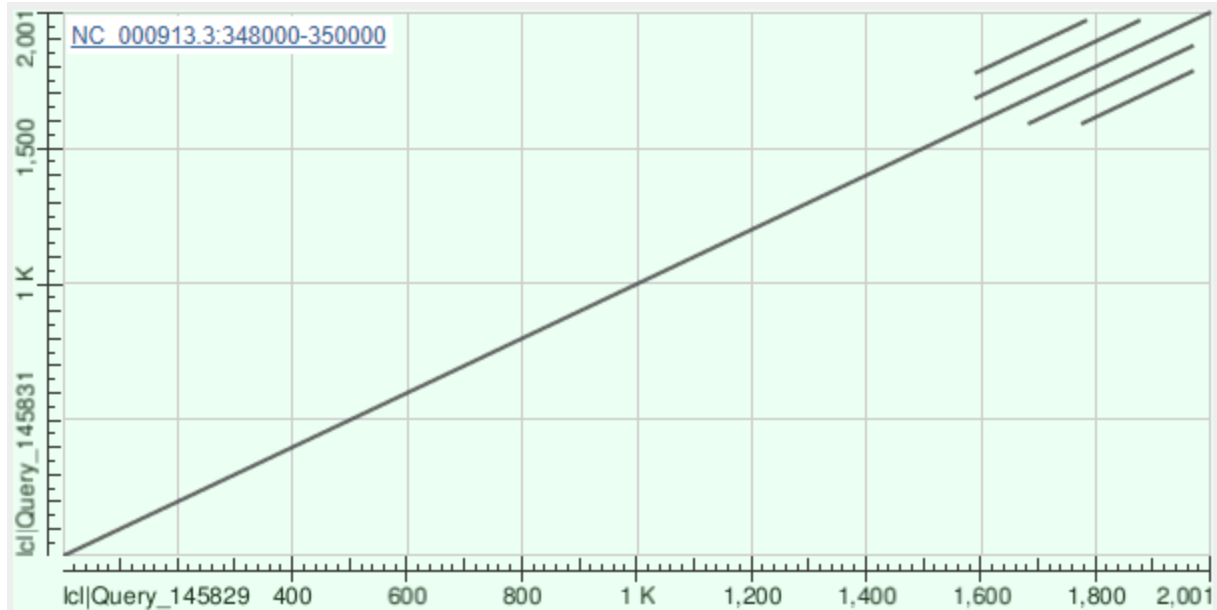
ATAAGACGCGTTAGCGTCGCATCAGGCATCTGCGCACGACTGCCGGATGCGGCGTAAACGCCTTATCCGG
CCTACGGATGGCGCGGAATTTGTAGGCCTGATAAGACGCGTTAGCGTCGCATCAGGCATCTGCACACGA
CTGCCGGATGCGATAAACGTCTTGTCCGGCCTACATTTGCCCCGTAGGCAGTCATTAAATAGTTCTGATT
ACGCGGCTGGCAATGTATCAGTCGCGATTACCTGCACTCGCAACGAGGTTTATCTTATGTCGACTCGTA
CCCCTTTCATCATCTTCATCCCGCTGATGCTGACCATCGGGCTTTGTTTTTTGGTCGCTCTGATGGAAGG
GCTGGATCTTCAGGCGGCTGGCATTGCGGCGGGTGGCATCGCCCAGGCTTTCGCACTCGATAAAATGCAA
ATGGGCTGGATATTTAGCGCCGGAATACTCGGTTTGCTACCCGGCGCGTTGGTTGGCGGAATGCTGGCGG
ACCGTTATGGTCGCAAGCGCATTTTGATTTGGCTCAGTTGCGCTGTTTGGTTTGTCTCACTGGCAACGGC
GATTGCCCTGGGATTTCCCTCACTGGTCTTTGCGCGGCTGATGACCGGTGTCGGGCTGGGGCGGCGTTG
CCGAATCTTATCGCCCTGACGTCTGAAGCCGCGGGTCCACGTTTTTCGTGGGACGGCAGTGAGCCTGATGT
ATTGCGGTGTTCCCATTTGGCGCGGCGCTGGCGGCGACACTGGGTTTCGCGGGGGCAAACCTTAGCATGGCA
AACGGTGTTTTGGGTAGGTGGTGTGGTGCCGTTGATTCTGG



>NC_000913.3:1708000-1710000

GCTCCCTCTGCCAATGAACCTGGCGAACCACAGGAAGAACAAGCTGCATCCGGTGTAGCGCCTGTGCTGACGCCTGCCCTGCGGATCTTTTGCCGC
AACAGTTGTAAGTTCAGCAAGGTCAGCAACACGATAAAGCTACCACGCATAACATTGCTGATTGATTGAATGTGGGGCTTGCGCGTGGGTTTG
CCCGAGCAATATTTCCCTTGGTGAATATTTCCGTCAGGAAAAAGCTGAAATTCGGCTATTTCGTCAGGAAGAAAAGCGCGCCGAGCAAGCCAAAGCG
CGTTTCGAAGCGCGCCAGGCTCGTCTGGAGCGCGAAAAAGCGGCTCGCCTTGAACGACATAAGAGCGCAGCCGTTCAACCTGCAGCCAAAGATAAAG
ATGCGATTGCTGCCGCTCTGGCGCGGGTGAAAGAGAAACAGGCCCAGGCTACACAGCCTATTGTGATTAAAGCGGGCGAACCGCCGGATAACAGTGC
AATTATTGCAGCACGGGAAGCCCGTAAAGCGCAAGCCAGAGCGAAACAGGCAGAACTGCAGCAAACTAACGACGCAGCAACCGTTGCTGATCCACGT
AAAACCTGCCGTTGAAGCAGCTATCGCCCGCGCCAAAGCGCGCAAGCTGGAACAGCAACAGGCTAATGCGGAACCAGAAACAACAGGTCGATCCGCGCA
AAGCCCGCTCGAAGCCGCTATTGCCCGTGCCAAAGCGCGCAAGCTGGAACAGCAACAGGCTAATGCGGAACCAGAAACAACAGGTCGATCCGCGCAA
AGCCCGCTCGAAGCCGCTATTGCCCGTGCCAAAGCACGCAAGCTGGAACAGCAACAGGCTAATGCGGAGCCAGAAACAACAGGTCGATCCGCGCAA
GCCGCGCTCGAAGCCGCTATTGCCCGAGCCAAAGCGCGCAAAACGGGAACAGCAACCGGCTAATGCGGAGCCAGAAACAACAGGTTGATCCGCGCAAAG
CTGCCGCTCGAAGCGGCTATTGCACGCGCCAAAGCACGCAAGCTGGAACAGCAACAGGCTAATGCGGTACCAGAAACAACAGGTTGATCCGCGCAAAGC
GGCAGTTGCCGCGGCTATTGCCCGCGCTCAGGCAAAAAAGCGCCAGCAGAAAGTTGTAACGAGGACTAAATGGTATTTCAGAAATAGCTAGCTCC
CCTTATACCCATAACAGCGCCAGACATCGCGCATTATGCTGTTGGTGTGCTCGCAGCCGTGCCAGGAATCGCAGCGCAACTGTGGTTTTTTGGTT
GGGGTACTCTCGTTCAGATCCTGTTGGCATCGGTTAGTGCTCTGTTAGCCGAAGCTCTCGTACTCAAACCTACGCAAGCAGTCGGTAGCCGCAACGTT
GAAAGATAACTCAGCATTGCTGACAGGCTTATTGCTGGCGGTAAGTATTCCTCCCTCGCGCCATGGTGGATGGTTCGTGCTGGGTACGGTGTGTTGCG
GTGATCATCGCTAAACAGTTGTATGGCGGTCTGGGACAAAACCCGTTTAATCCGGCAATGATTGGTTATGTGGTCTTACTGATCTCCTTCCCGGTGC
AGATGACCAGCTGGTTACCGCCACATGAAATTCGGTCAACATCCCTGGTTTTATCGACGCCATCCAGGTTATTTTAGCGGTCATACCGCCAGTGG

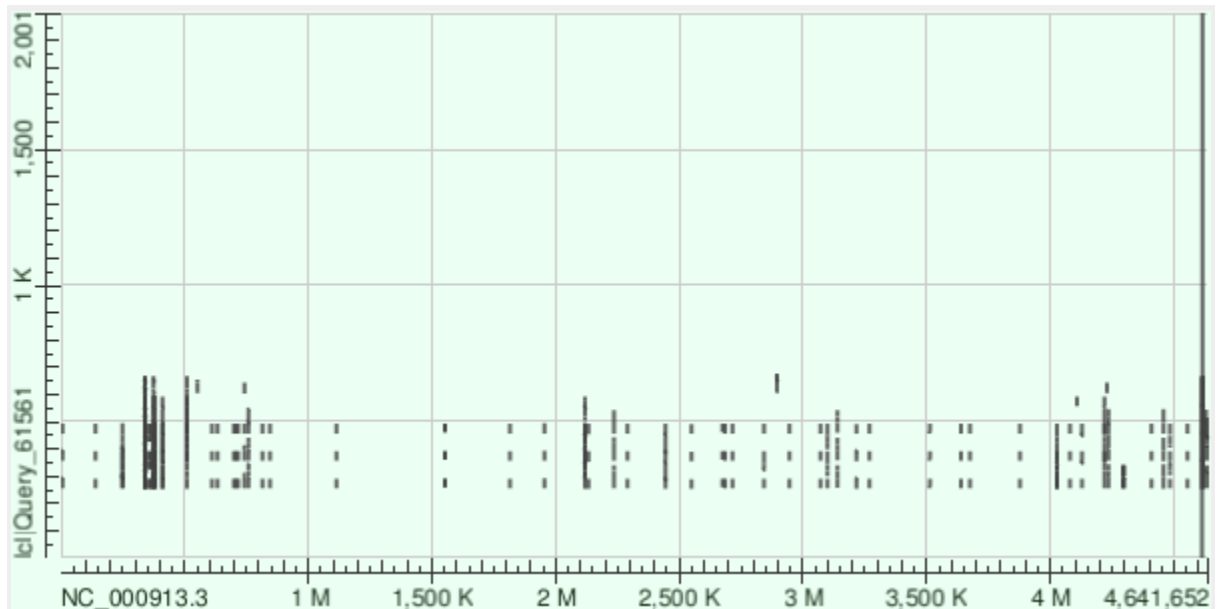
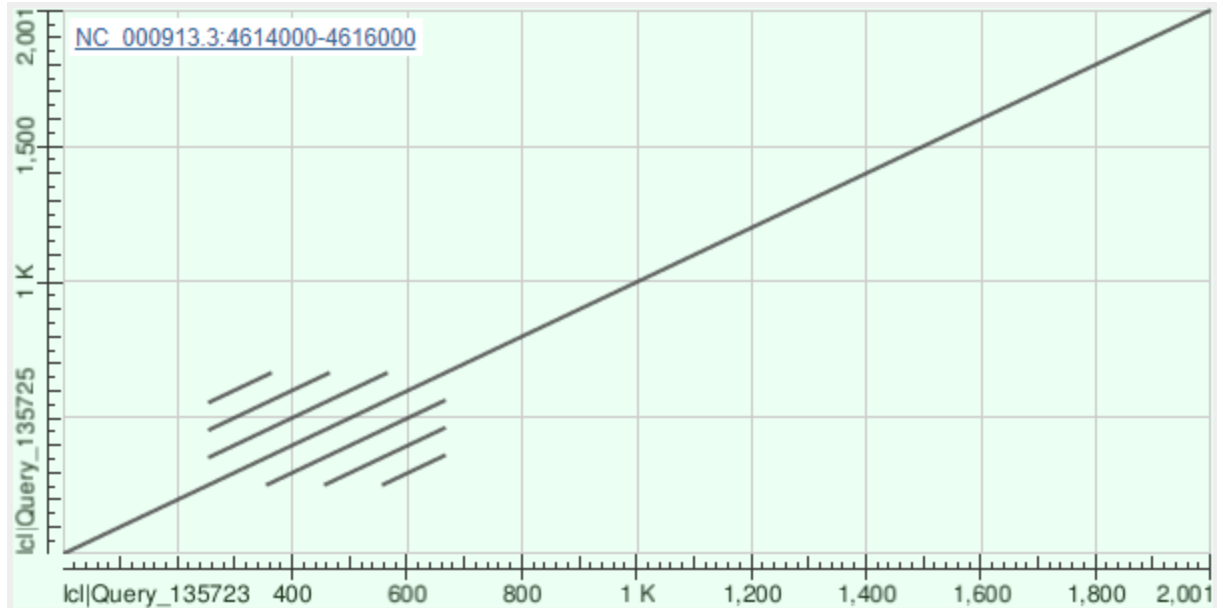
TGGTGATATGAACACACTACGCTTAGGTATTGATGGCATTAGTCAGGCGACACCGCTGGATACATTTAAAACCTCTGTCCGTGCCGGTCATT CGGT
GAACAGATTATGCAATATCCGATCTACAGCGGTATTCTGGCGGGCGCTGGTTGGCAATGGGTAAATCTCGCCTGGCTGGCTGGCGGCGTATGGTTGC
TATGGCAGAAAGCGATTGCTGGCATATCCCCTCAGCTTCTTAGTAACGCTGGCGTTATGCGCAATGTTGGGCTGGTTGTTCTCACCAGAAACACT
GGCAGCACCGCAAATTCATCTGCTGCTGAGCGACCATGCTCGGCGCATTCTTTATTTTG



>NC_000913.3:348000-350000

```
GTCCTGTTAATCTGCCCCGCTGCGTCTTCCTCGGTAATGTAGCTACGTTGGTCGAGGCGCAAATTAAAGGT
TTTTTGAAACGCCACCAGCGCTGGAATGGTTTTCTGATAGGTGACAACGCCGATAGAAGAGGTGAGTTT
CCGGCTTTTGCCAGTGCTGTAAACATCGTAGCCGCTCGGTTTAATCAAAATAACTGGCACTGACAGGC
GGCTTTTCAGGTACGCGCCGTTAGAGCCAGCCGCGATGATGGCGTCACAGCGTTCGTTTGCCAGTTTCTT
GCGGATGTAGGTCACTGCTTTTTCAAAGCCAAGCTGAATAGGGGTAATGTTTCGCCAGGTGATCAAACCTCG
AGGCTGATATCGCGAAACAGCTCGAACAGGCGCGTTACAGATACCGTCCAGATAACCGGTTTGTCTCAT
TAAGCCGTGGTGGATGTGCCATAGCGCACCGCAAAGTTAAGAAACCGAATATTGGGTTTAGTCTTGTTTC
ATAATTGTTGCAATGAAACGCGGTGAAACATTGCCTGAAACGTTAACTGAAACGCATATTTGCGGATAG
TTCATGACTTTATCTCTAACAATTGAAATTAAACATTTAATTTTATTAAGGCAATTGTGGCACACCCCT
TGCTTTGCTCTTATCAACGCAATAACAAGTTGATAACAAAGGATGGGCTATGTCTCTACACTCTCCAGG
TAAAGCGTTTCGCGCTGCACTGACTAAAGAAAATCCATTGCAGATTGTTGGCACCATCAACGCTAATCAT
GCGCTGTTGGCGCAGCGTGCCGATATCAGGCAATTTATCTTCTGGCGGTGGCGTGGCGGCAGGTTGCG
TGGGGCTGCCGATCTCGGTATTTCTACCCTTGATGATGTGCTGACCGACATTCGCCGTATCACCGACGT
TTGTTGCTGCGCTGCTGGTGATGCGGATATCGGTTTGGTTCTTCGGCCTTTAACGTGGCGCGCACCC
GTGAAATCGATGATTAAGCCGGTGCGGCAGGATTGCATATTGAAGATCAGGTTGGTGCGAAACGCTGCG
GTATCGTCCGATAAAGCGATCGTCTCGAAAGAAGAGATGGTGGATCGGATCCGCGCGGCGGTGGATGC
GAAACCGATCCTGATTTTGTGATCATGGCGCGCACCGATGCTCTGGCGGTAGAGGGGCTGGATGCGGCG
```

ATCGAGCGTGCGCAGGCCTATGTTGAAGCGGGTGCCGAGATGTTGTTCCCGGAGGCGATTACCGAACTCG
CCATGTACCGCCAGTTTGGCCGATGCGGTGCAGGTGCCGATCCTCGCCAACATCACCGAATTTGGTGCCAC
GCCGCTGTTTACCACCGACGAATTACGCAGCGCCCATGTCGCAATGGCGCTGTACCCACTTTTCAGCGTTC
CGCGCCATGAACCGCGCCGCTGAACATGTCTACAACGTCCTGCGCCAGGAAGGCACGCAGAAAAGCGTCA
TCGACACCATGCAGACCCGCAACGAGCTGTACGAAAGCATCAACTACTACCAGTACGAAGAGAAGCTCGA
CAACCTGTTTGCCCGTAGCCAGGTGAAATAAAAAACGCCCCTTGATTGTATTCGACAGCCGATGCCTGAT
GCGTCGCTGACGCGACTTATCAGGCCTACGAGGTGCACTGAACTGTAGGTCGGATAAGACGGATGGCGTC
GCATCCGACAACCGATGCCTGATGCGCCGCTGACGTGACTTATCAGGCCTACGGGGTGCACTGAACTGTA
GGTCGGATAAGACGCATAGCGTCGCATCCGACAACCGATGCCTGATGCGCCGCTGACGCGACTTATCAGG
CCTACGGGGTGCACTGAACTGTAGGTCGGATAAGACGCATAGCGTCGCATCCGACAACCGATGCCTGATG
CGCCGCTGACGCGACTTATCAGGCCTACGGGGTGAACGAACTGTAGGTCGGATAAGACGCATAGCGTCG
CATCCGACAACAATCTCGACCCTACAAATGATAACAATGAC



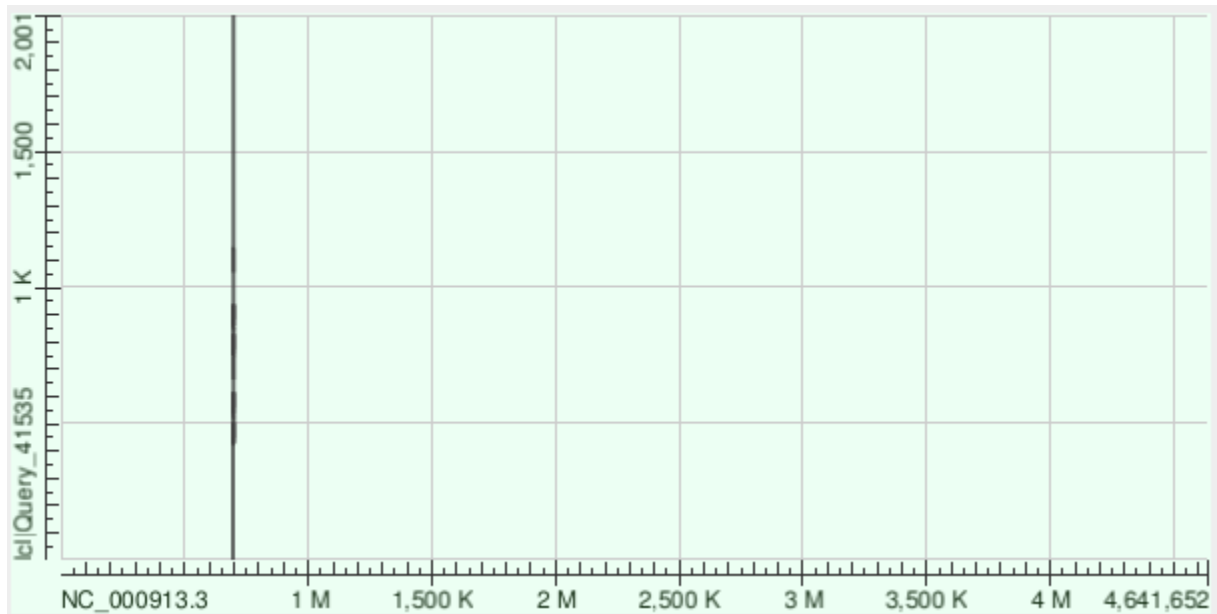
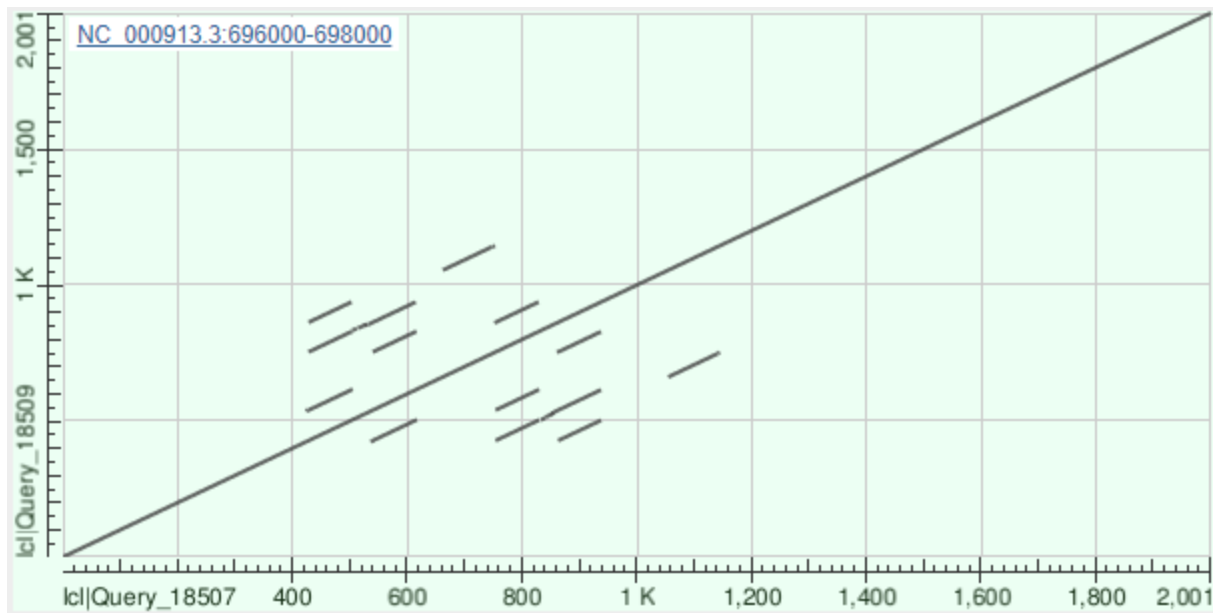
>NC_000913.3:4614000-4616000

```

CAAAATTGGCGTAGGCGGTACTATCACCTATCCACGCGCCAGTAAACCCGCGATGTCATCGCAAAATTACCGCTGGCATCGTTATTGCTGGAACC
GACGCGCCGGATATGCCGCTCAACGGTTTTTCAGGGGCAGCCTAACCGCCGGAGCAGGCTGCCCGTGTGTTCCGCCGTGCTTTGCGAGTTGCGCCGGG
AACC GGCGGATGAGATTGCGCAAGCGTTGCTTAATAACACGTATACGTTGTTAACGTGCCGTAGGCCGGATAAGGCGTTACGCGCGCATCCGGCAG
TTGGCGCACAAATGCCTGATGCGACGCTTAACGCGTCTTATCATGCCTACAGGTTTGTGCCGAACCGTAGGCCGGATAAGGCGTTCACGCCGCATCCG
GCAGTTGGCGCACAAATGCCTGATGCGACGCTTGTGCGGCTTATCATGCCTACAAGTCTGTGCCGAACCGTAGGCCGGATAAGGCGTTCACGCCGCA
TCCGGCAGTCCGGCGCATAATGCCTGATGCGACGCTTGTGCGGCTTATCATGCCTACAGGTTTGTGCCGAACCGTAGGCCGGATAAGGCGTTCGCGC
CGCATCCGGCAGTTGGCGCACAAATGCCTGATGCGACGCTTACGCGCTTATCAGGCCTACAAGTCTGTGCCGAACCGTAGGCCGTATCCGGCATGT
CACAAATAGAGCGCCGGAATATCAACCGGCTCACCCGCGCACCTTTAACGCATCAGCCAACGGCTCAACGCTCTCCGGCGTGGCGCTCGCCAGC
TTTGGCGCTCGCCATACACGCCGTGGCGATGAAACGCGTTACGGCGTACCCGGAACATCGCCGAGTCCCTTGATAAACGCCGAGTTCTTCGATGTG
TTGCAAAATCCACCTGGCCAGGGATCACCAGCAAACGCAGTTCCGCCAGCTTCCGCCGCTCTGCCAGCAAATAGATGCTGCGCTTAATCTGCTGA
TTATCGCTCCCGTGAGTTGTTGATGACATTCCGCTCCCCACGCTTTGAGATCGAGCATTGCCCGTCGCACACCGGGAGCAATTTTCCAGCCGG
TTTCGCTCAACATGCCGTTACTGTCCACCAGACAGGTGAGATGGCGCAGTTGCGGATCGTTTTTGATAGCAGTAAACAGCGCCACCACAAACGGCAG
CTGGGTCGTGGCTTCACCGCCACTCACCGTTATCCCTTCGATAAACAGCACTGCTTTGCGGACATGGCTAAGCACTTCGTCCACGCTCATGGATTGC
GCCATGGGCGTGGCATGTTGCGGACACCTCTTCAGGCAGGTATCACACTGCTCGCAAACCACAGCGTTCACACCCTTTGCCGTCAACAATCTGCA
ACGCTGATGCGGACACTGTGGCAGCACTCCCCACAGTCATTGCAACGTCCCATCGTCCACGGATTGTGACAGTTTTTGACGCGCAGATTGCAGCC
CTGCAAAAACAGAGCCAGACGACTGCCTGGCCCGTCAACGCAGGAGAAGGGGATAATCTTACTGACTAAAGCGCATCTGCTGTTTATGGCTTATCAC
GCGCGCTGGCGTTCCGAATACGAGTGTGCGTGGCGCTTCTTCGCCCAGCCAGGTGGTGTGGTGCCTGACCTTCGGCGCGATATTTTCTAAA
TCCGACAAACGCACCATATAACCGGTAACGCGAACCAGATCGTTACCGCTGACATTGGCGGTAATTCACGCATTCCGGCTTTAAAGGCACCGAGGC

```

AAAGCTGTACCAGTGCCTGCGGGTTACGTTTGATGGTTTCGTCGAGCGTCAGAATGTCGCTGATGCCGGAATAATAATAAGCATGATGCGGGGCGAC
AGTTTGCAGATGGGTGATCGGATCTGGCTCATCGCCATACGGCAAACGCGCGCCCGGCGTGGTGCCGATATCGGAACTGATCCCCGACTGTGCGTGT
AACATGGCGCGTTTTTGCCAGCCATATTCA
CGGGGGTATTGGCGACAAACTCCGCCAGTT



>NC_000913.3:696000-698000

```
TGAATCTTGGTTATCGTGATGTCGATGCCCTGATTGATGTTCTGGTCAACGCCCGCAGCTACGGCGAAGCGTGGGCCAGTTATCCTGTCTCAAGCG
TTACCAGATGCGGCGCATGGCGGATAAATTCTATTATGCAAAGCGGTATGGATCTGTTTATGCGGATTTCAGCAATAATCTGCCACCACTGCGTTTT
ATGCGTAATCTCGGGTTAATGGCGGCGGAGCGTGCTGGCGTGTGAAACGTGAGGCGCTGAAATATGCGTTAGGGTTGTAGCCTTACAACATTGCCG
GATGCGTGCCAAACCGTAGGTCCGATAAGACGCGGCAGCGTCGCATCCGACATTGAAGGATAAGACGTGTCAACATCGCATTTCGACATTGAATGAACG
CAGAAAAGCAAAAAGCTCGCCGAAGCGAGCTTTTTTAATGTGGCTGGGGTACGAGGATTGCAACCTCGGAATGCCGAATCAGAATCCGGTGCCTTA
CCGCTTGGCGATACCCCAAGTGGTGCACCTTACAAGGTAAGCGCTCTTGAATAAATTGGCTGGGGTACGAGGATTGCAACCTCGGAATGCCGAATCA
GAATCCGGTGCCTTACCGCTTGGCGATACCCCAACAAATTGGTTTTGAATTTGCCGAACATATTCGATACATTGCAATTTGGTGGCTACGACGGGA
TTCGAACCTGTGACCCCATCATATGAGTGATGTGCTCTAACCAACTGAGCTACGTAGCCAGATTGTTCTTCGATGGCTGGGGTACCTGGATTTCGA
ACCAGGGAATGCCGGTATCAAAAACCGGTGCCTTACCGCTTGGCGATACCCCAATAACCGGGCGGTGAACCGCTTACTCGAAGAAGATGGCTGGGGT
ACCTGGATTTCGAACCAAGGAATGCCGTATCAAAAACCGGTGCCTTACCGCTTGGCGATACCCCATCCGTACAACGCTTCTCTGGTGAATGGTGGCGG
AGGCGAGACTTGAACTCGCACACCTTGGCGCGCCAGAACCTAAATCTGGTGGCTCTACCAATTTGCCCACTCCCGCAAAAAAAGATGGTGGCTACGA
CGGGATTGCAACCTGTGACCCCATCATATGAGTGATGTGCTCTAACCAACTGAGCTACGTAGCCATCTTTTTTTTCGCGATACCTTATCGGCGTTG
CGGGGCGCATTATGCGTATAGAGCCTTGCAGCGTCAACCTCTTTTTCAAGGAAATTTGCTCGAAAGTGACTGTTTGGTTAGGTTGCGAACAGCGTGG
CGTATATTCGTCAATTATTGTTTACTTTGTGTTTTTCCACCCCTACAGCCATTCTTTTGTATACAGGATGAAATTCGGAATTTAACAATAGTGG
TGGGAAATTAATCTATGAAATACTGGCTACAGTGATGAGTTGTCAAACAGTGATGTGGCAAACCCGGAACATTTCTTACTGCATATCAGAATCAA
CAAGCACCTCAATAACTGAAACAGCCCGGATTCACCGGGGCTGTTTCGCATTTCTTACTTATACGCCGACTGGTGAAACACCAACCGCGCGACCAG
ACGGATCGTCCATTTTCTGAACGCTTCATCCCATTCGATCGCTTTAGCGGAAGAACAAGCGACGGAAGGACCGCCCGGCACGCACTCAGCGGCGCT
```

CGGAAGCGGGAATAGTTCTTCAAAGATCTCCCGATACAAGTACGCTTCTTTAGAGGTTGGCGTGTGTACGGGAAGCGGAAGCGGGCAGTTTCCAGT
TGCTGATCAGAAACCTGCTGCGCAGCCACTTCTTTCAGGGTGTGATCCAACGTAAACGACGCCATCGGAGAACTGCTCTTCTGCCGCCAGGCCA
CGCTTGCAGGCAGATACGCTTCAAACATTACGCAGGATGTGTTTTCCATTTGCCGTTACCGCACATTTTATCCTGTGGGTTAATACGCATCGC
CACATCAAGGAATTTTTTGTGAGGAACGGAACGCGTGCTTCCACGCCCCAGGCTGACATC