Genome Annotation Submission Cover Sheet

Preliminary Annotation Review Checklist 5-15-2018

Phage Name: KingMidas

Your Name: Christine Byrum

Your Institution: College of Charleston

Your email: byrumc@cofc.edu

Additional emails:
(For correspondence)

Please check each box indicating completion of each task. If you are not sure how to do something, please see the Online Bioinformatics manual page "How to Pass Preliminary Review".

- ☑ 1. Does the genome sequence in your final contain the same number of bases and is it the same as the posted sequence on phagesdb.org?
- ☑ 2. Are all the genes "valid" when you click the "validate" button?
- ☑ 3. Have the genes been renumbered such that they go sequentially from 1 to the highest number?
- ☑ 4. Have all old BLAST hits been cleared, and all gene features reBLASTed?
- ☑ 5. Are the locus tags the"SEA_ PHAGENAME"?
- ☑ 6. Has the Documentation been recreated to match the information in the feature table?
- ☑ 7. Have tRNA ends been adjusted with web-based Aragorn and/or tRNAscan SE?
- ☑ 8. Has the frameshift in the tail assembly chaperone been annotated (where applicable?)

9. For the items below, generate a genome profile, and review the following. For the

YourPhageName_CompleteNotes.dnam5 file:

- ☑ a. Have any duplicate genes (or any with the same stop coordinate?) been removed?
- ☑ b. Does every gene have **one and only one** complete set of Notes
- ☑ c. Do the functions in the Notes match the official function list?
- ☑ d. Are all three lines of functional evidence described for EVERY gene?
- ☑ e. Do the notes contain the initial Glimmer/GeneMark data from the autoannotation?

For the YourPhageName .dnam5 file:

- ☑ a. Have any duplicate genes (or any with the same stop coordinate?) been removed?
- ☑ b. Is the Notes field empty (including hidden marks?)
- ☑ c. Do the function names in the Product field either match the official function list or say "Hypothetical Protein"?
- ☑ d. Is the Function field empty (including hidden marks?)

☑ 10. Did you use PECAAN to annotate your phage?

If, so please describe how in the text field after question 11.

11. Describe any issues or specific genes that you were unable to satisfactorily resolve, and warrant further inspection in the Quality Control review.

This is my first annotated sequence. I hope everything is as it should be. I think the GeneMark/Glimmer data is in my DNA Master file, but have also attached a pdf copy of the KingMidas GeneMark data in case it is not there. Thank you for checking my files!

PECAAN was used to annotate. Start sites and locations of uncalled genes were determined using GeneMark, Glimmer, Starterator, and Phamerator. Functions of genes were determined using PhamBLAST, NCBI BLAST, HHPred, CDD, and Phamerator. Transmembrane domains were located based on TMHMM results and confirmed using SOSUI. The lack of valid tRNA and tmRNA was confirmed using online ARAGORN v.2.0 and online tRNAscan-SE. Full annotation was exported, edited to denote hypothetical protein sequences and the

GeneMark

Version 2.5p (09.08.06) *Copyright 1993 - M. Borodovsky, J. McIninch*

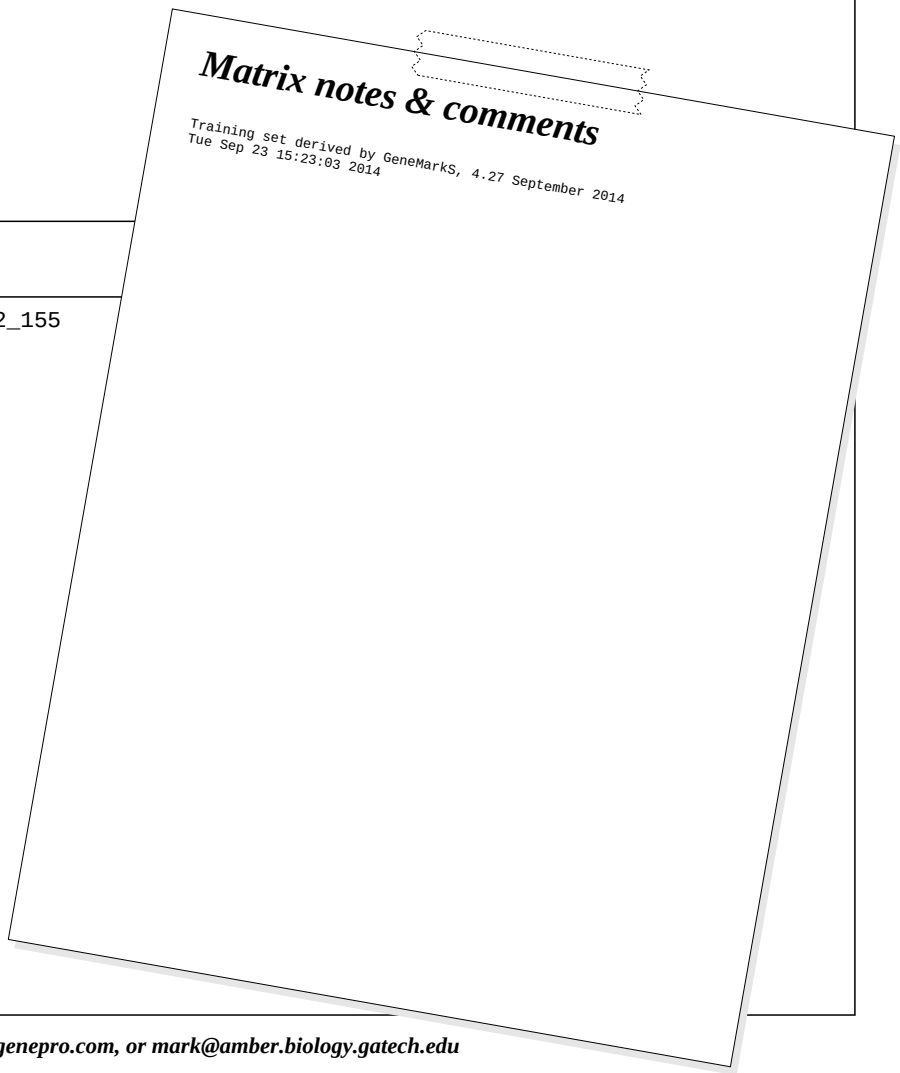
PROGRAM INFORMATION

```
Sequence          :  Mycobacterium phage KingMidas complete sequence, 57386 bp including 10-base 3' overhang (CG
Analysis Date     :  3/19/19 at 22:56:54
Pages             :  29
Sequence Length   :  57386 bp
GC Content        :   61.41%

Window Length     :  96 bp
Window Step       :  12 bp
Threshold Value   :  0.500

PS-Version        :  1.2

GeneMark Options  :  PostScript graph,
                     Mark ORFs / splice sites,
                     List ORFs,
                     List regions and/or splice sites,
```

MATRIX INFORMATION

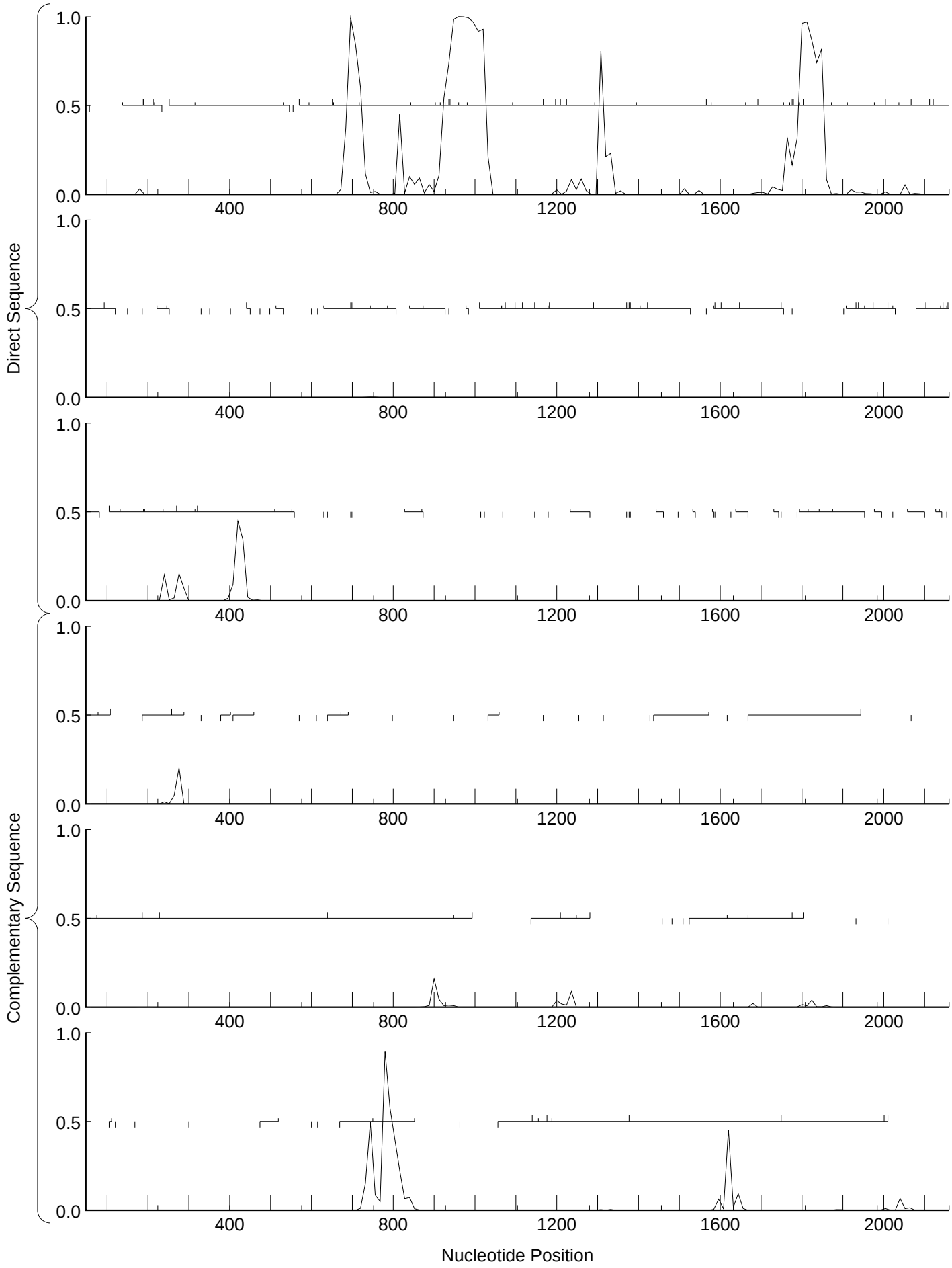
```
Matrix      :  Mycobacterium_smegmatis_MC2_155
Author      :  -

Order       :  4
```


```
  Send questions / comments to:
      Dr. M. Borodovsky
      Georgia Institute of Technology
      School of Biology
      Atlanta, GA  30332-0230
```

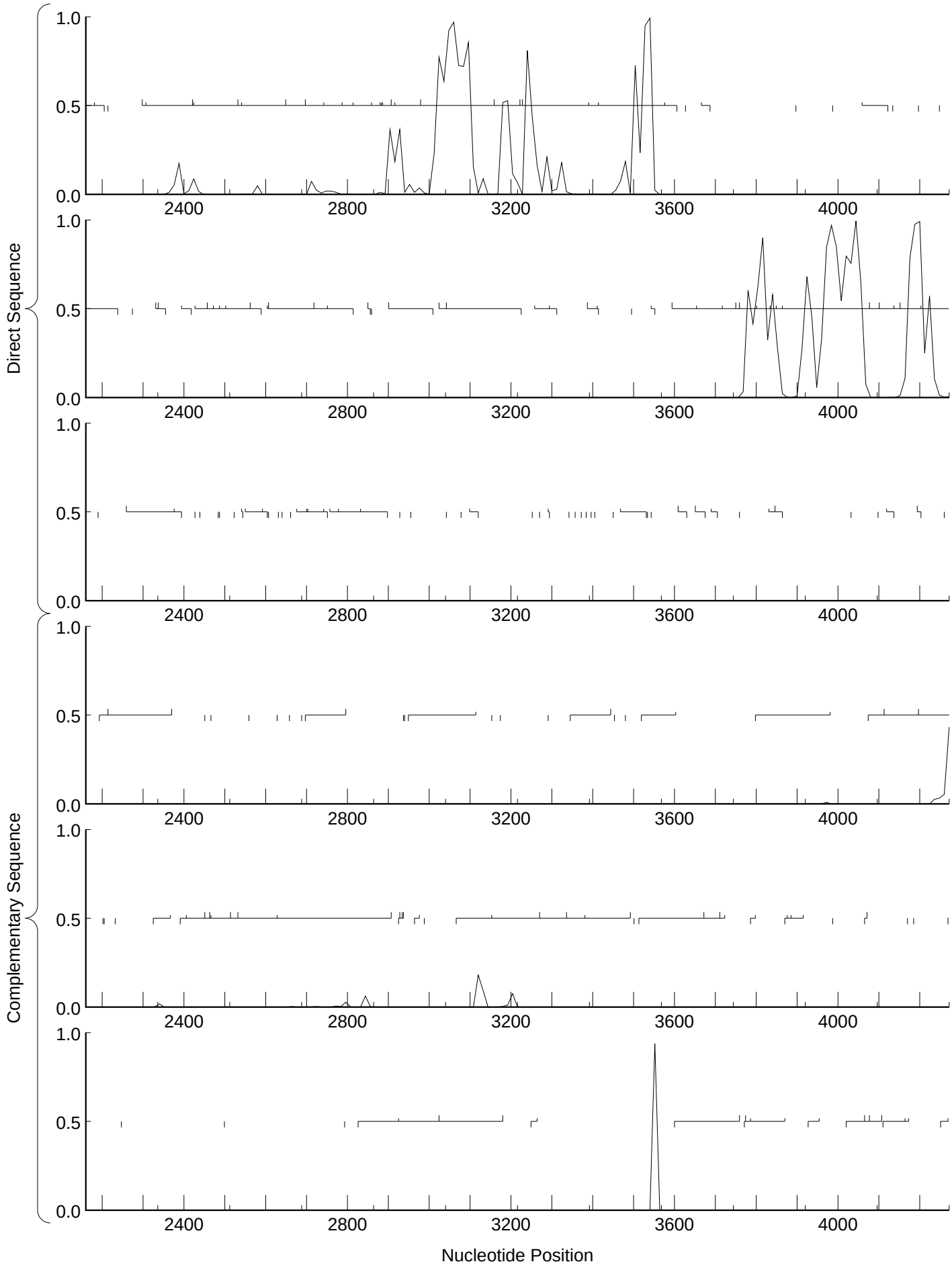



For help or information, send electronic mail to: custserv@genepro.com, or mark@amber.biology.gatech.edu

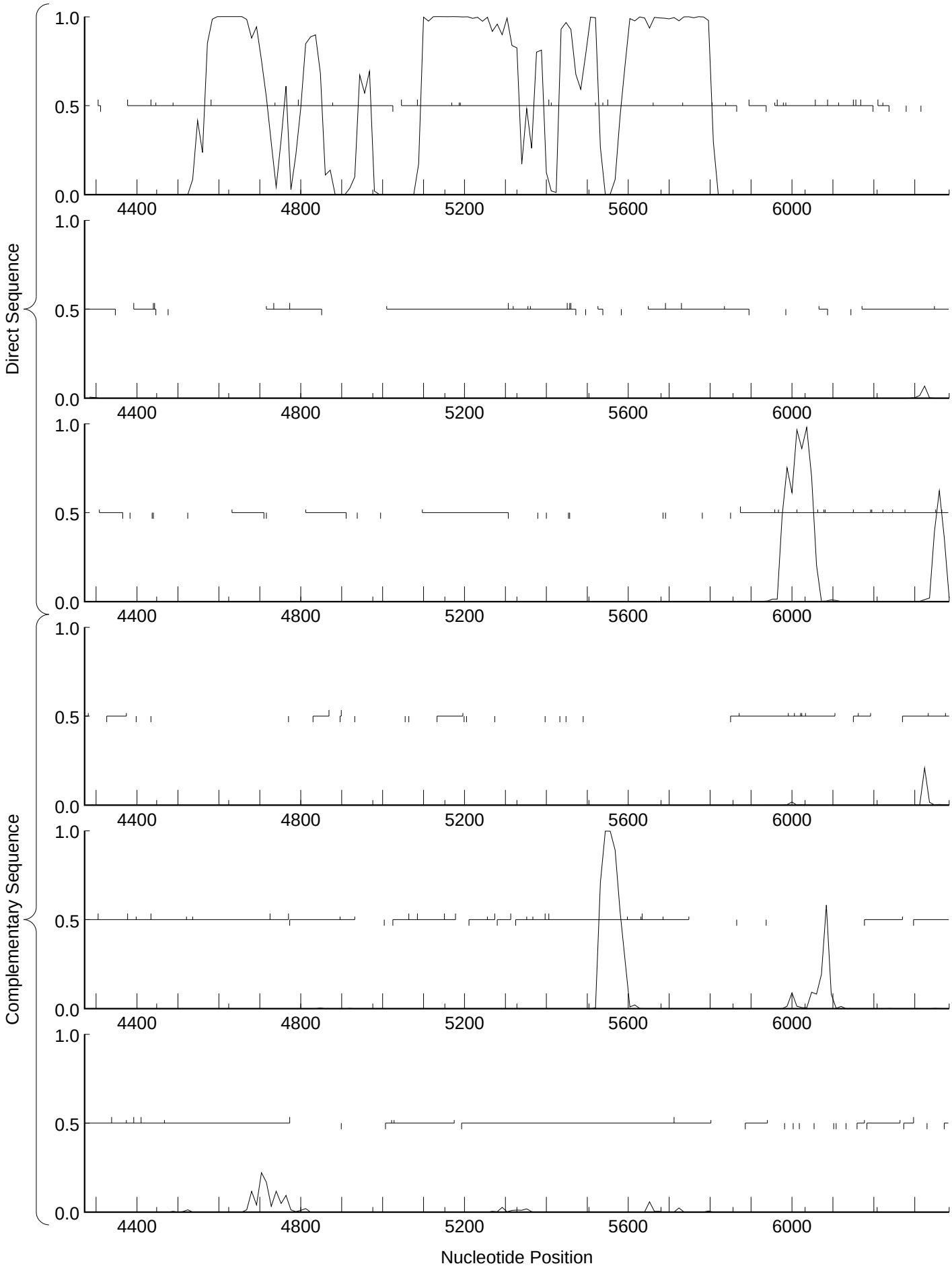
s complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 2/29



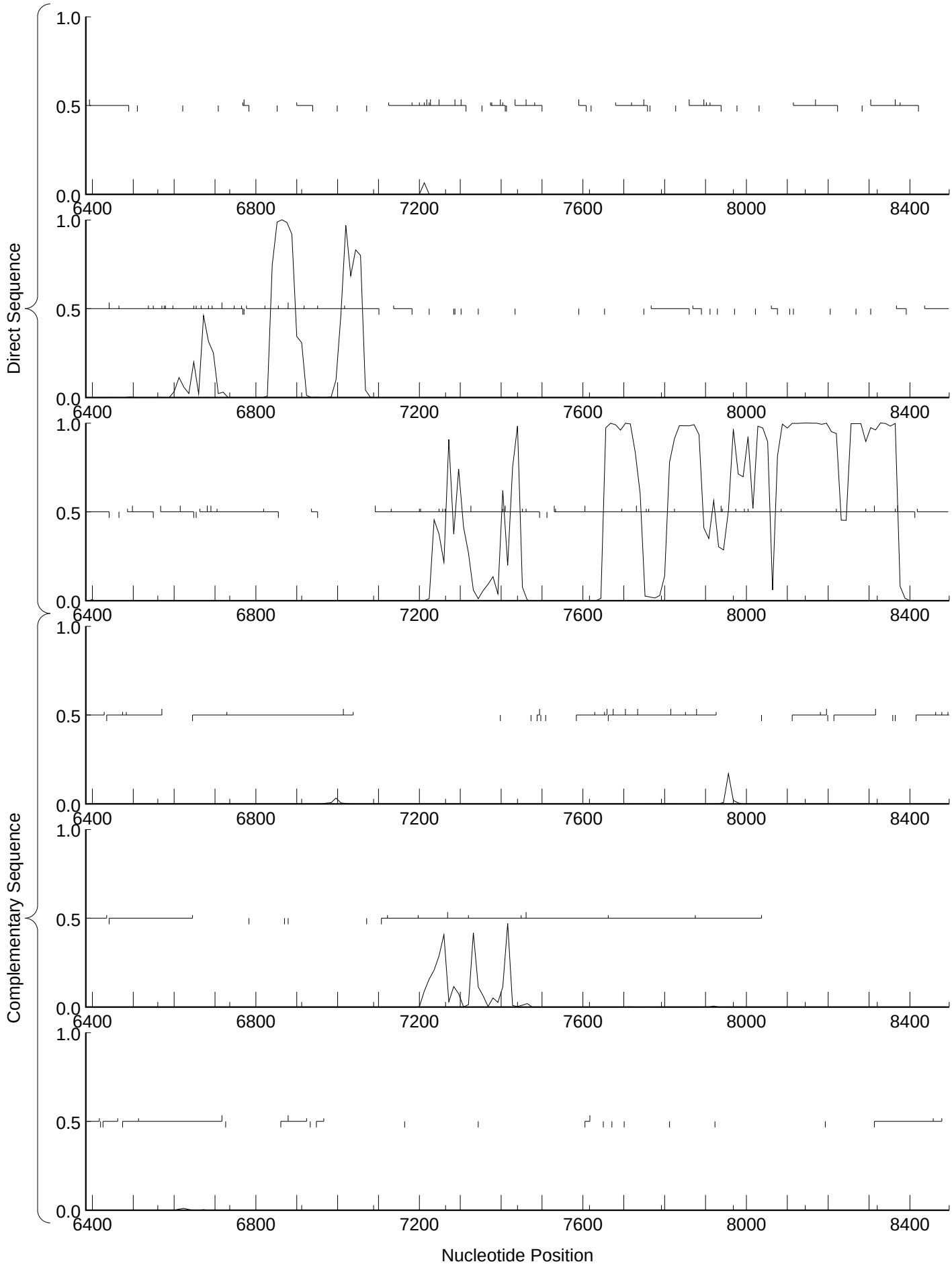
s complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 3/29



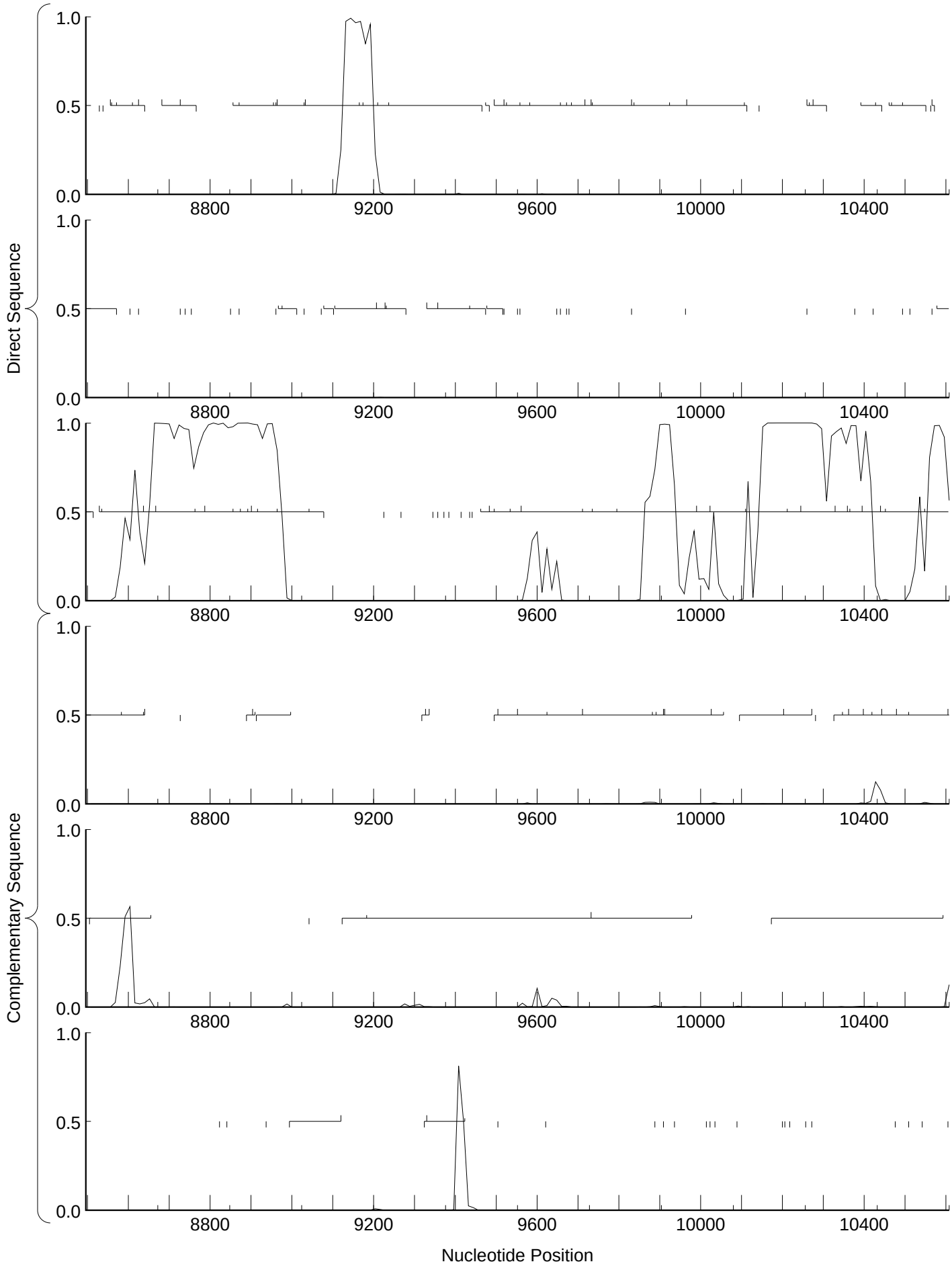
s complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 4/29



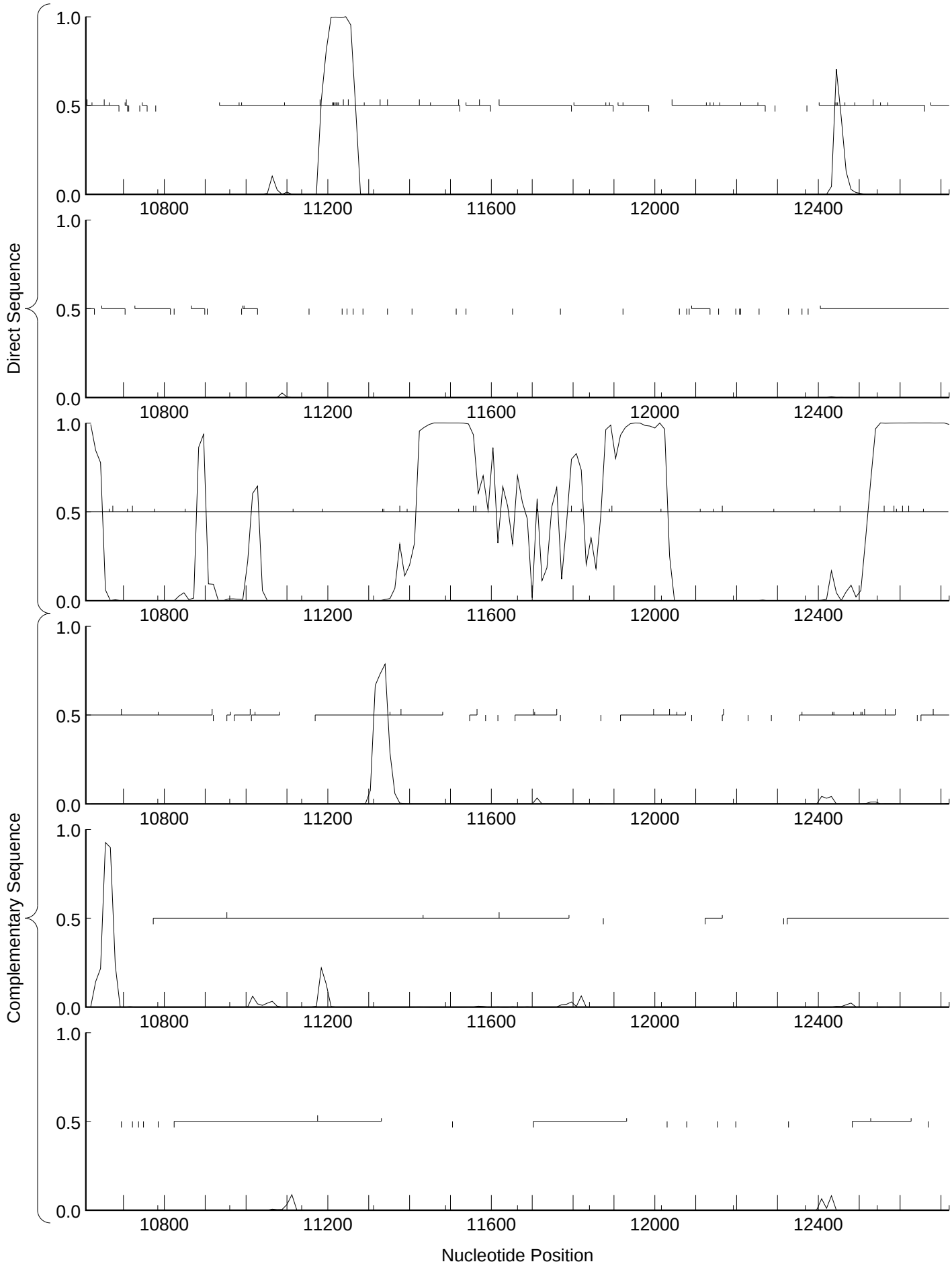
s complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 5/29



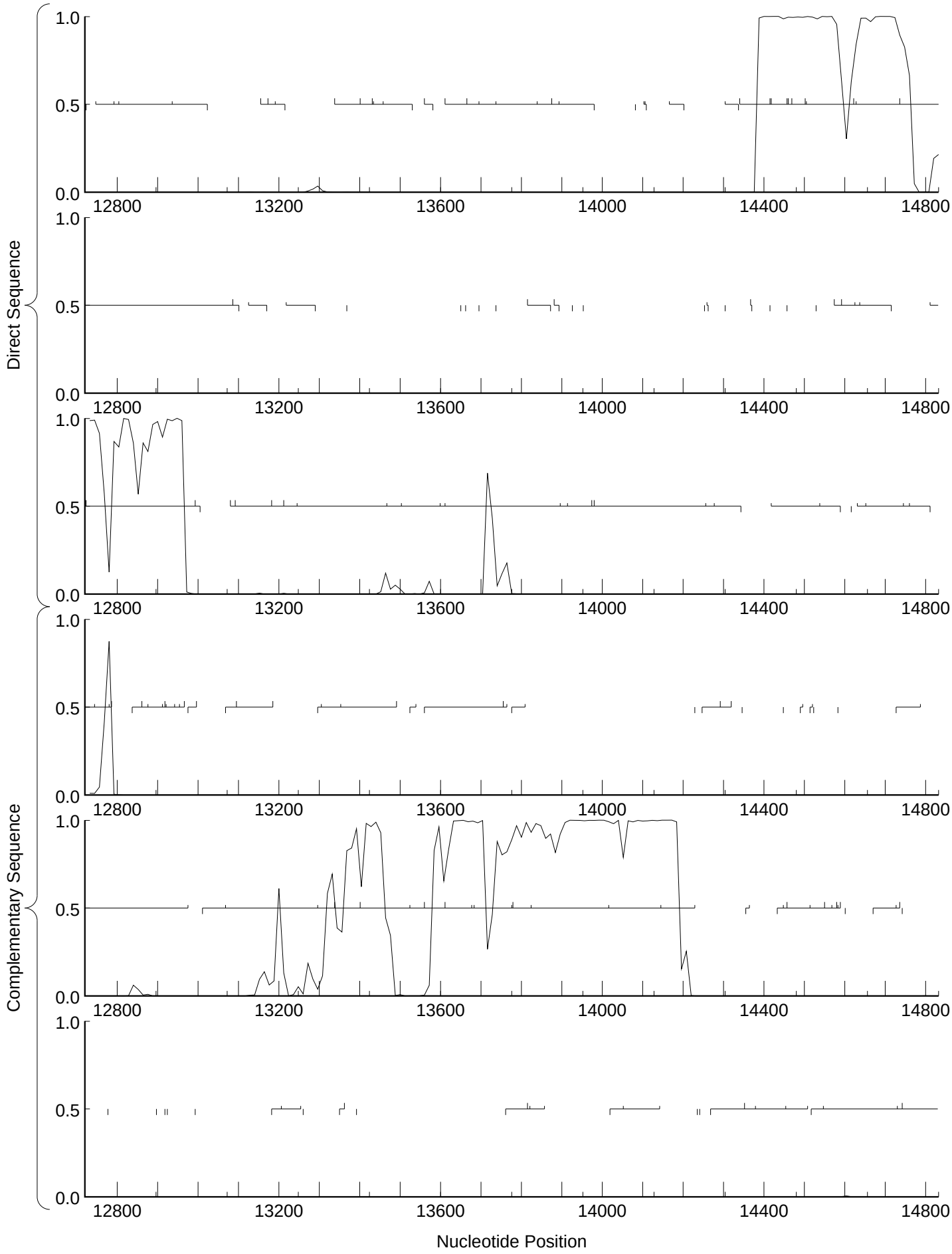
s complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 6/29



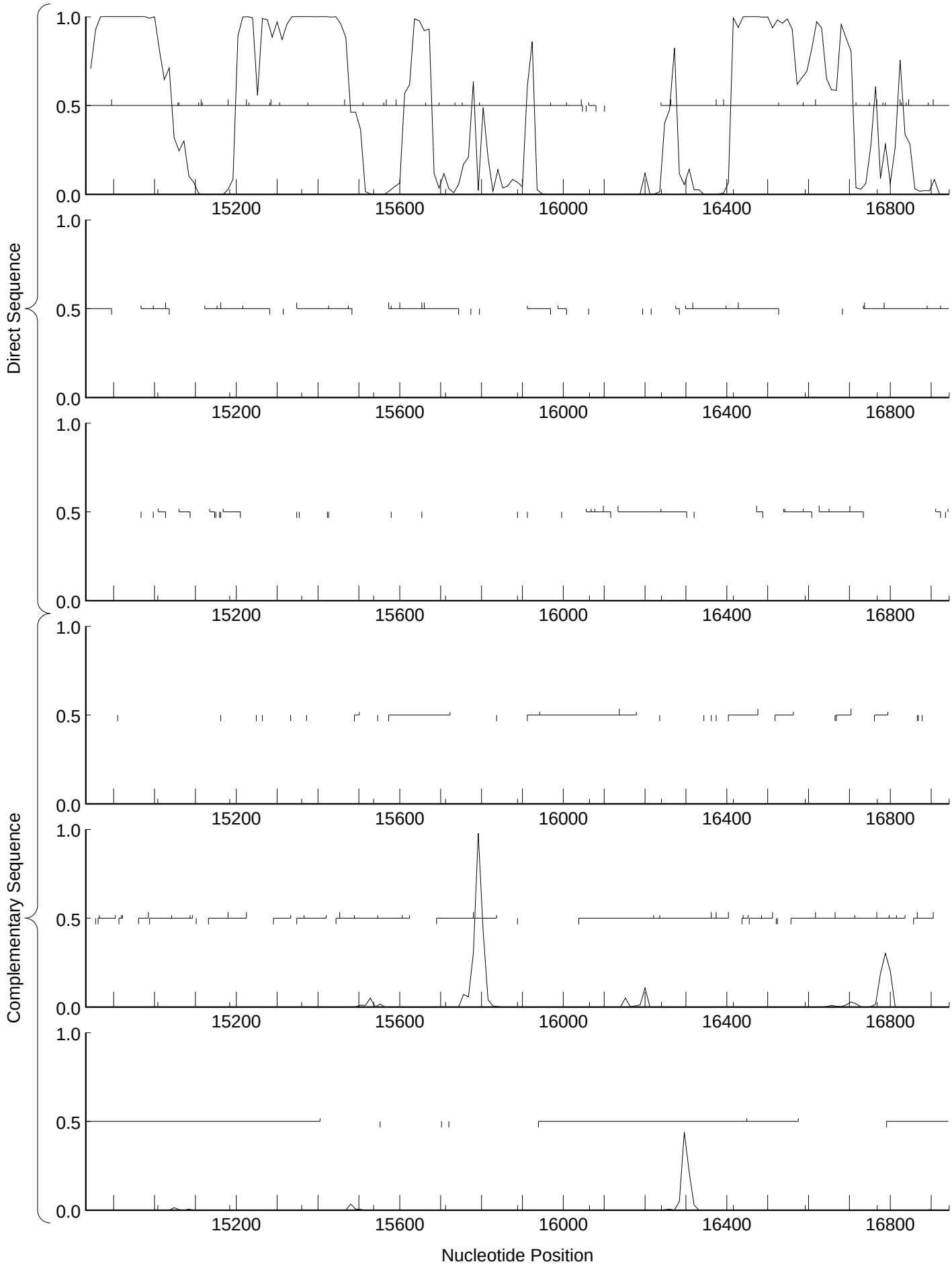
s complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 7/29



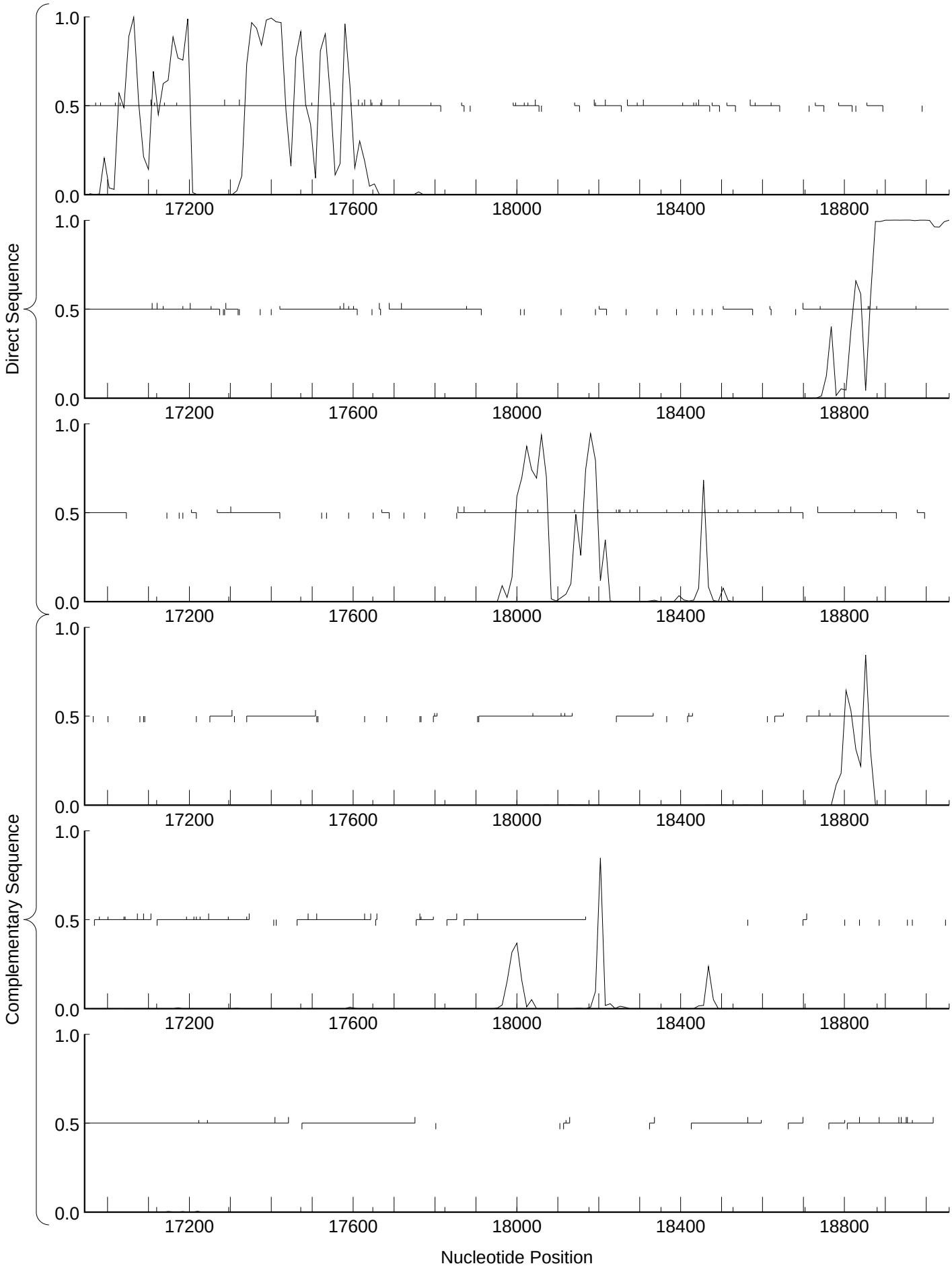
s complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 8/29



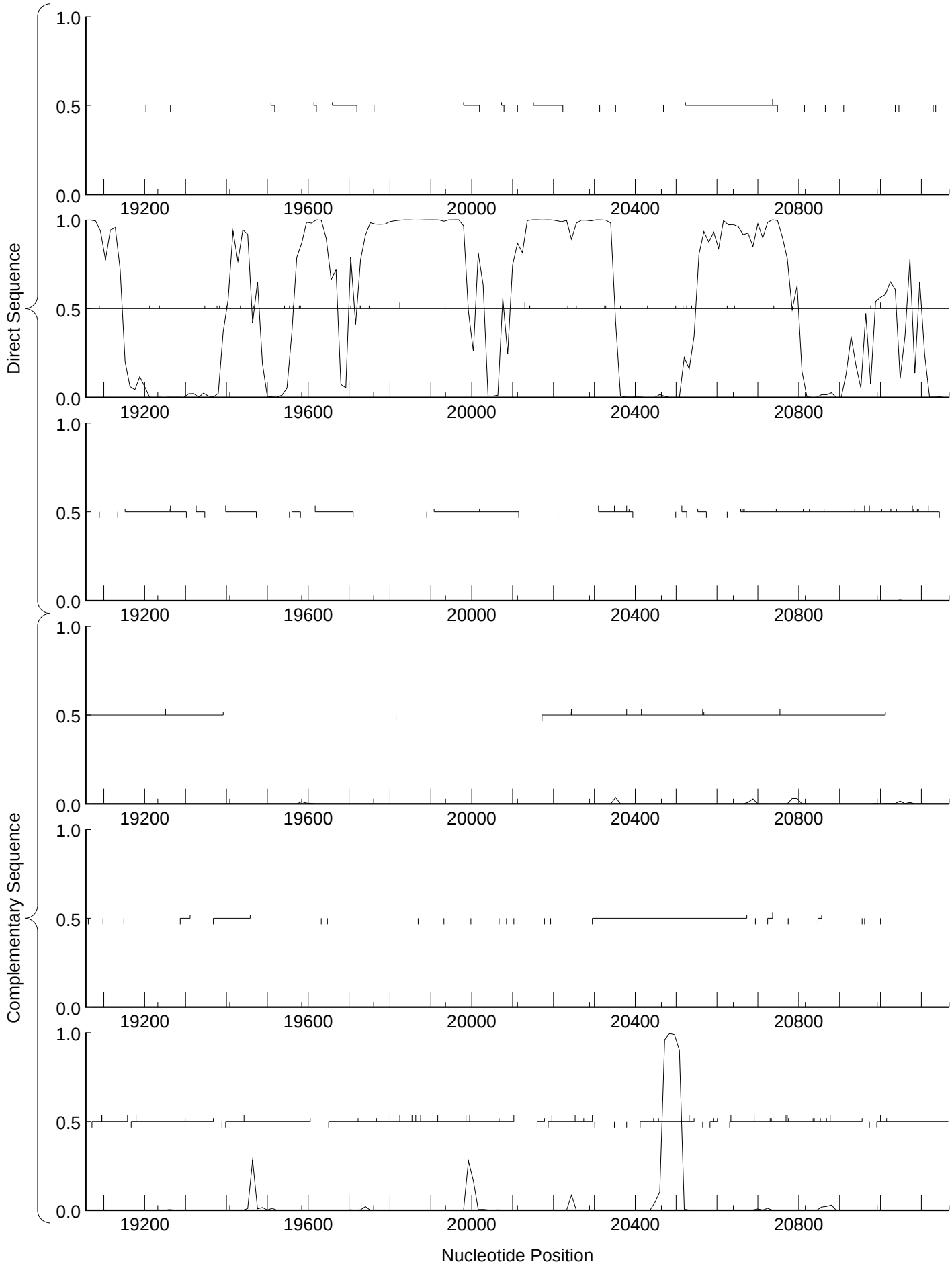
s complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 9/29



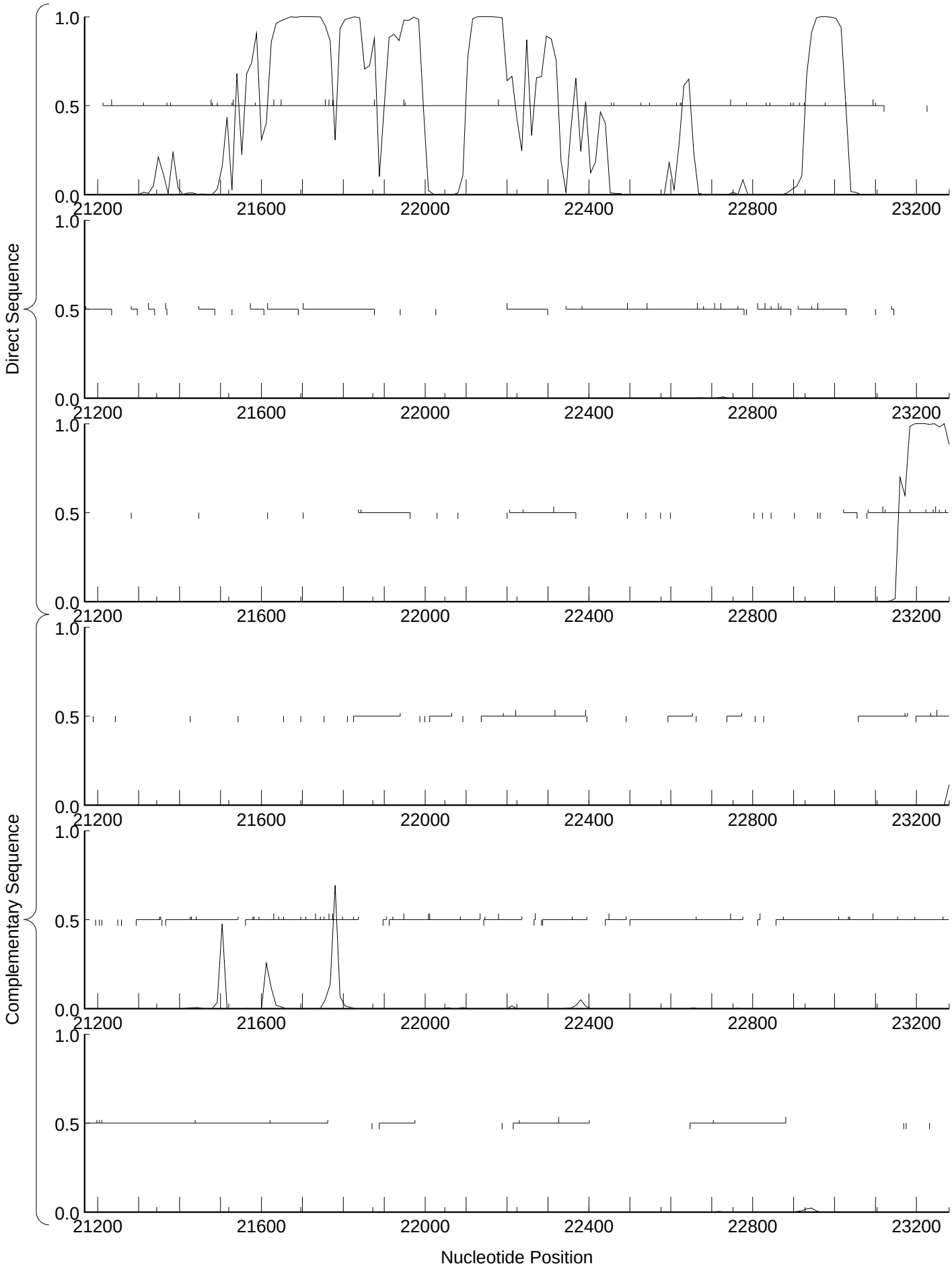
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 10/29



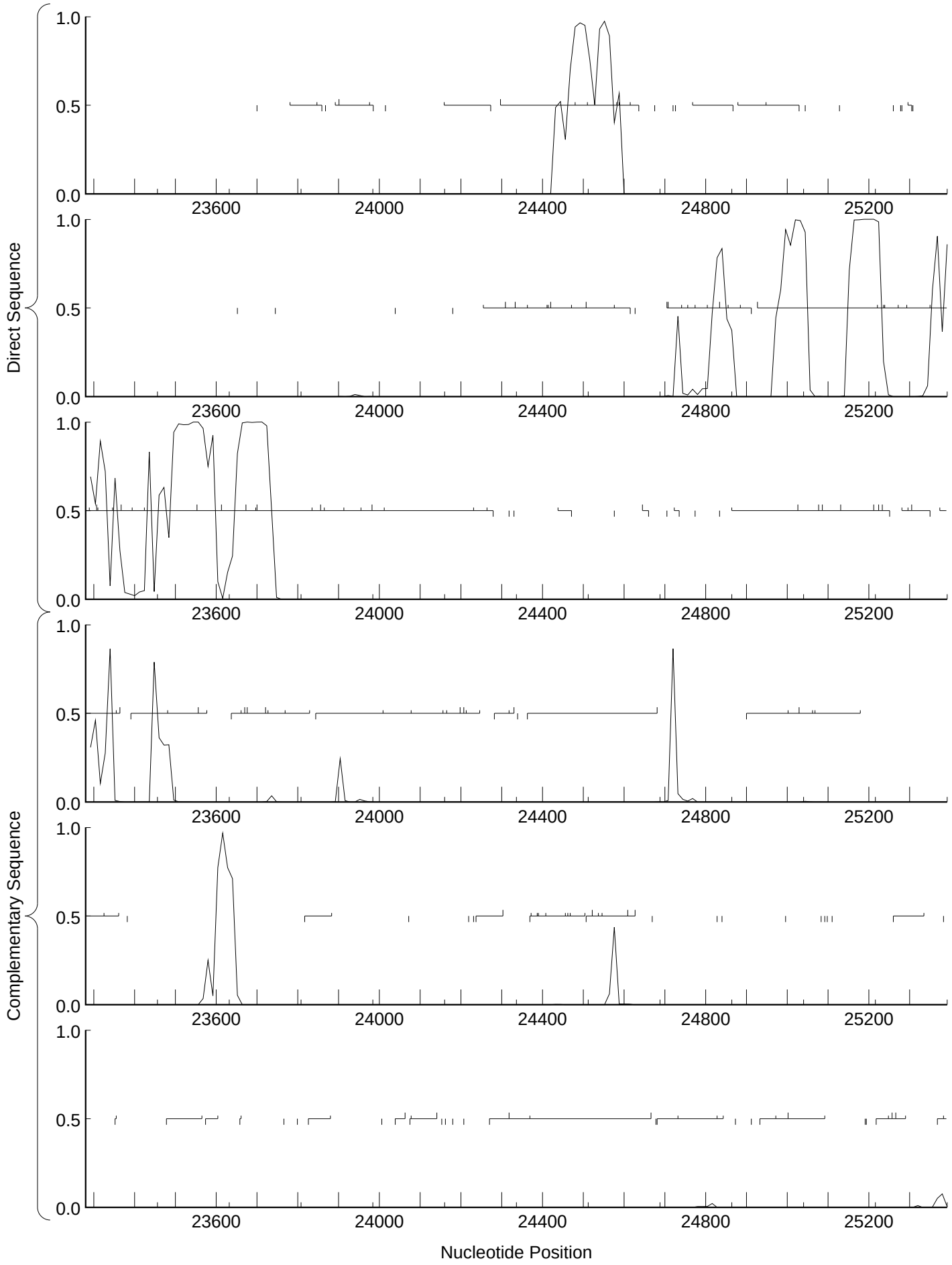
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 11/29



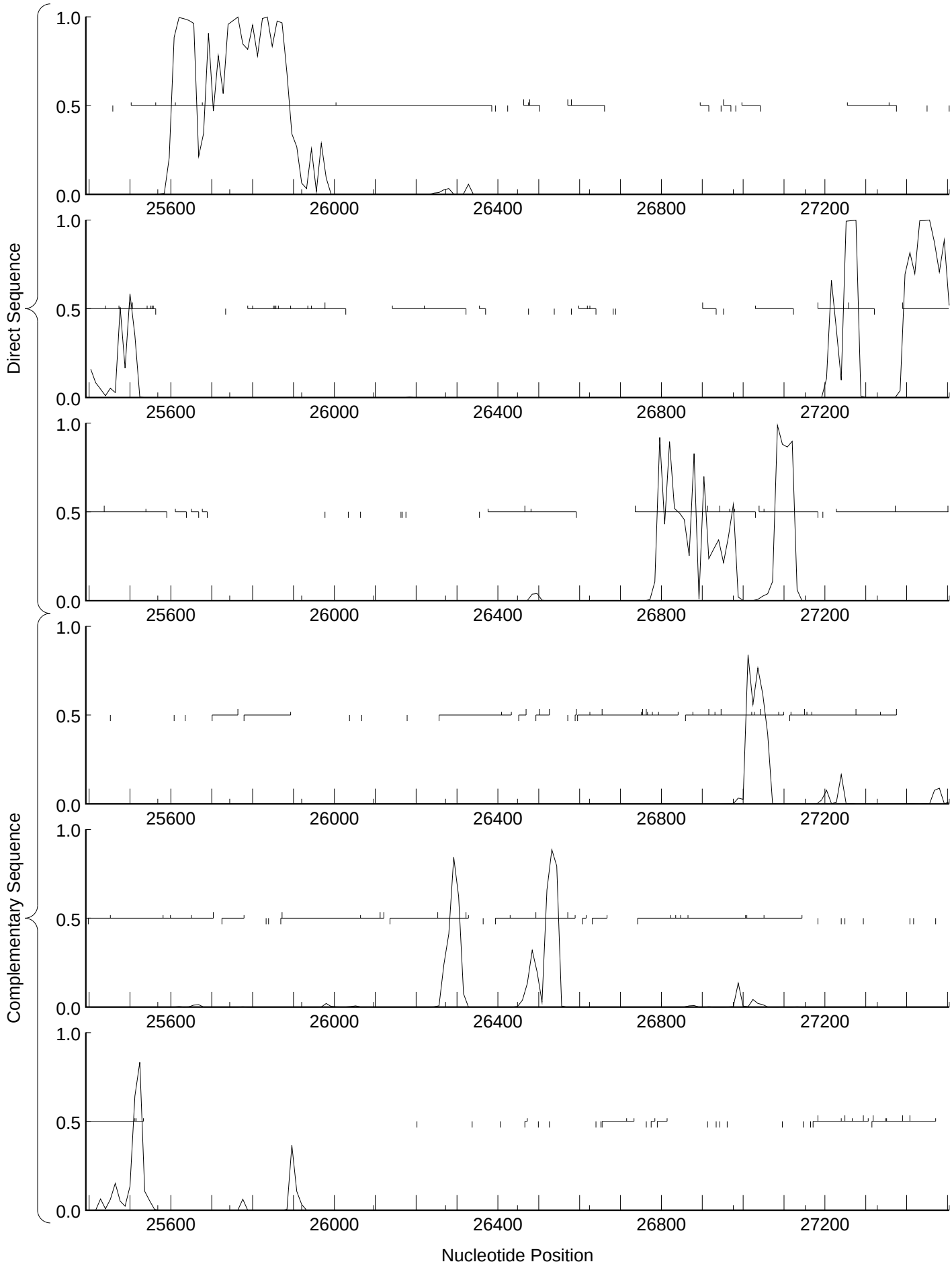
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 12/29



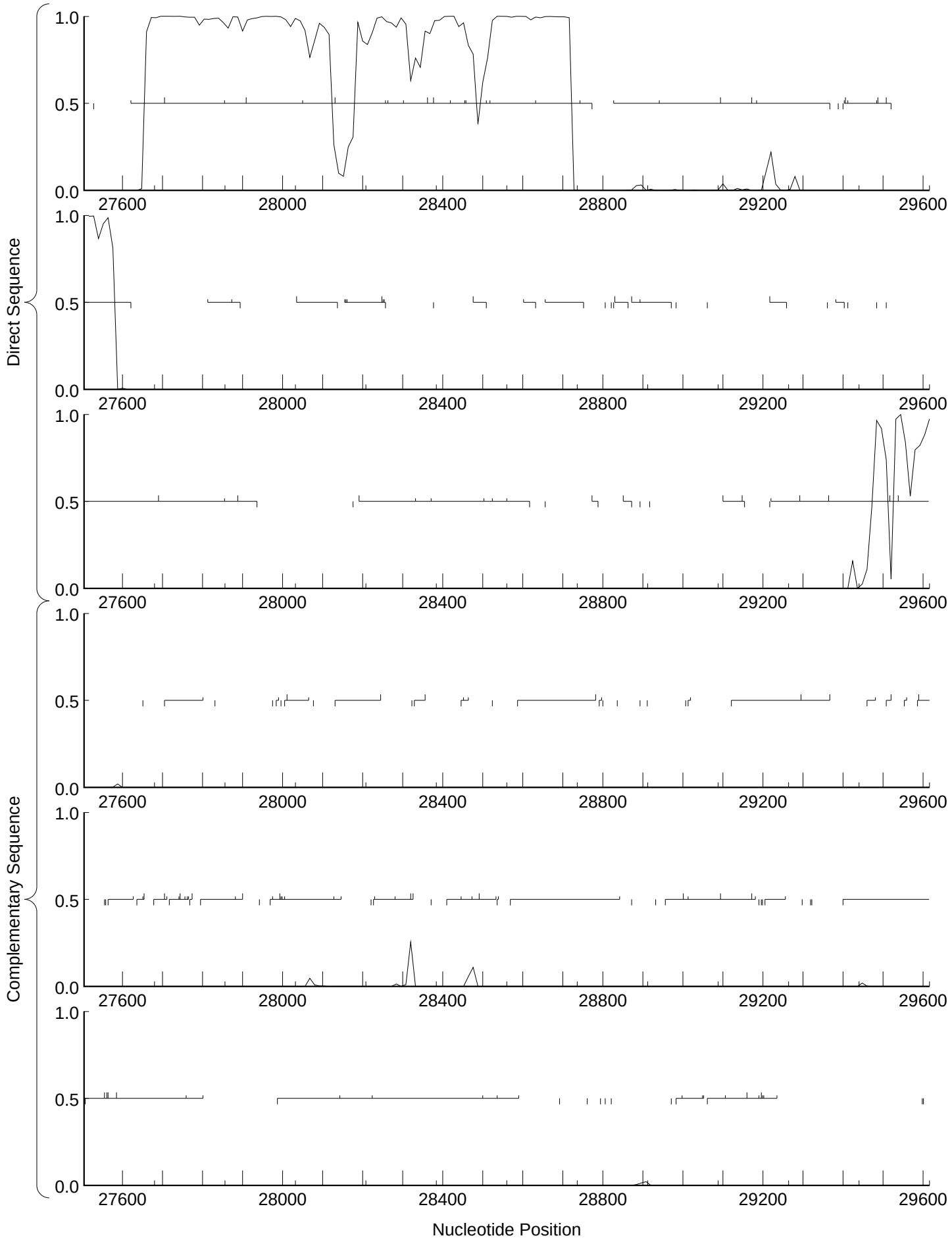
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 13/29



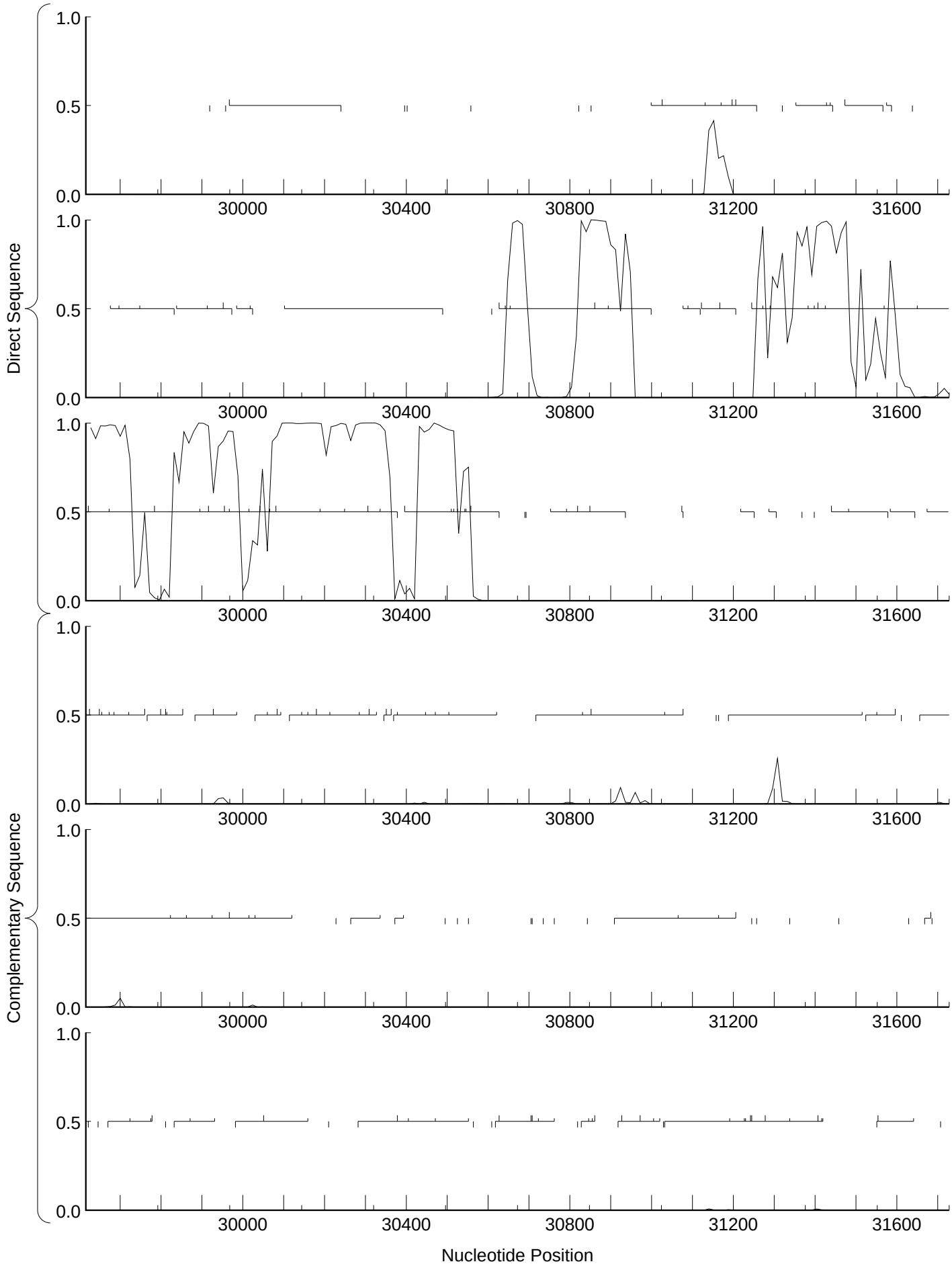
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 14/29



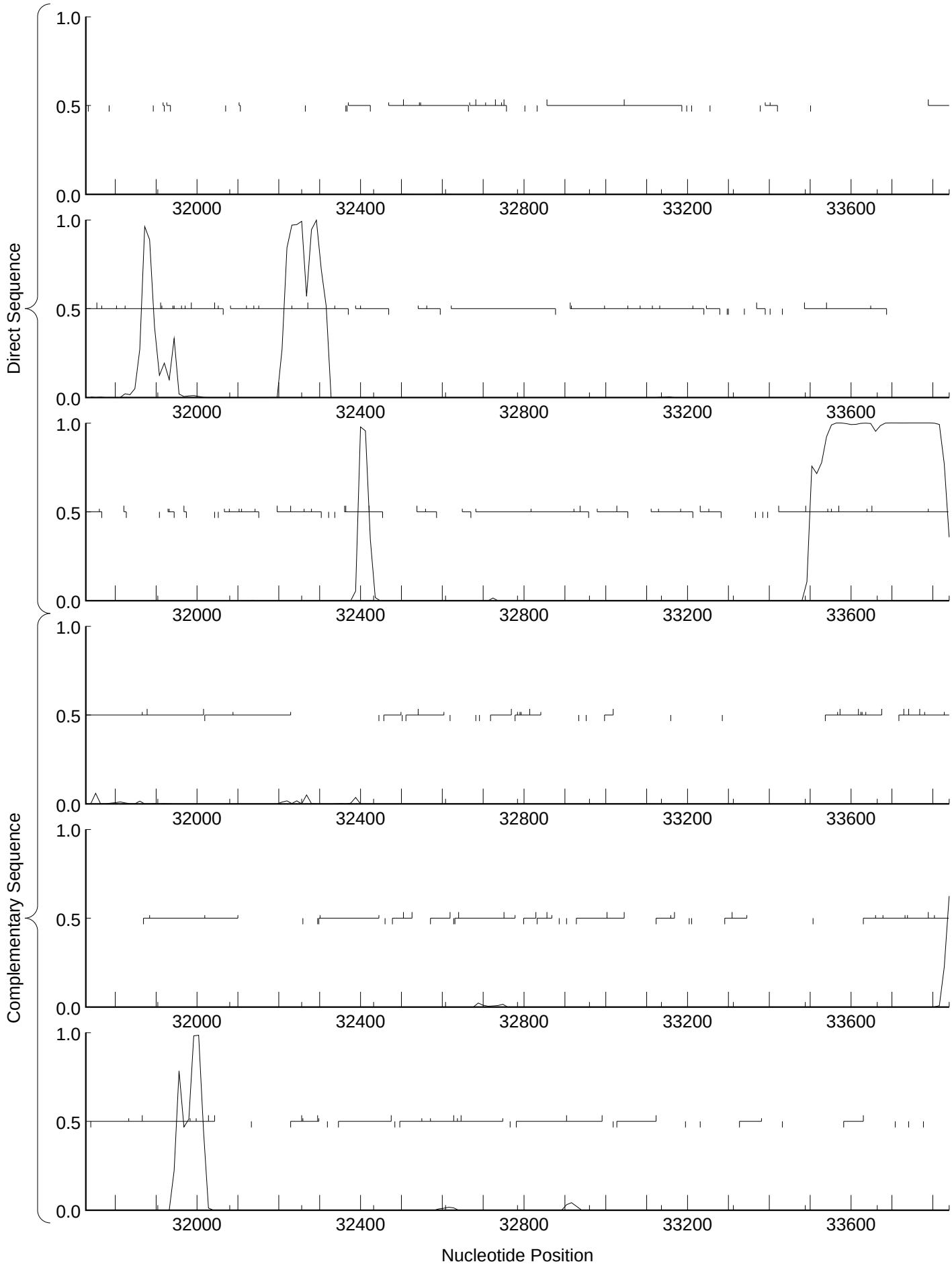
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 15/29



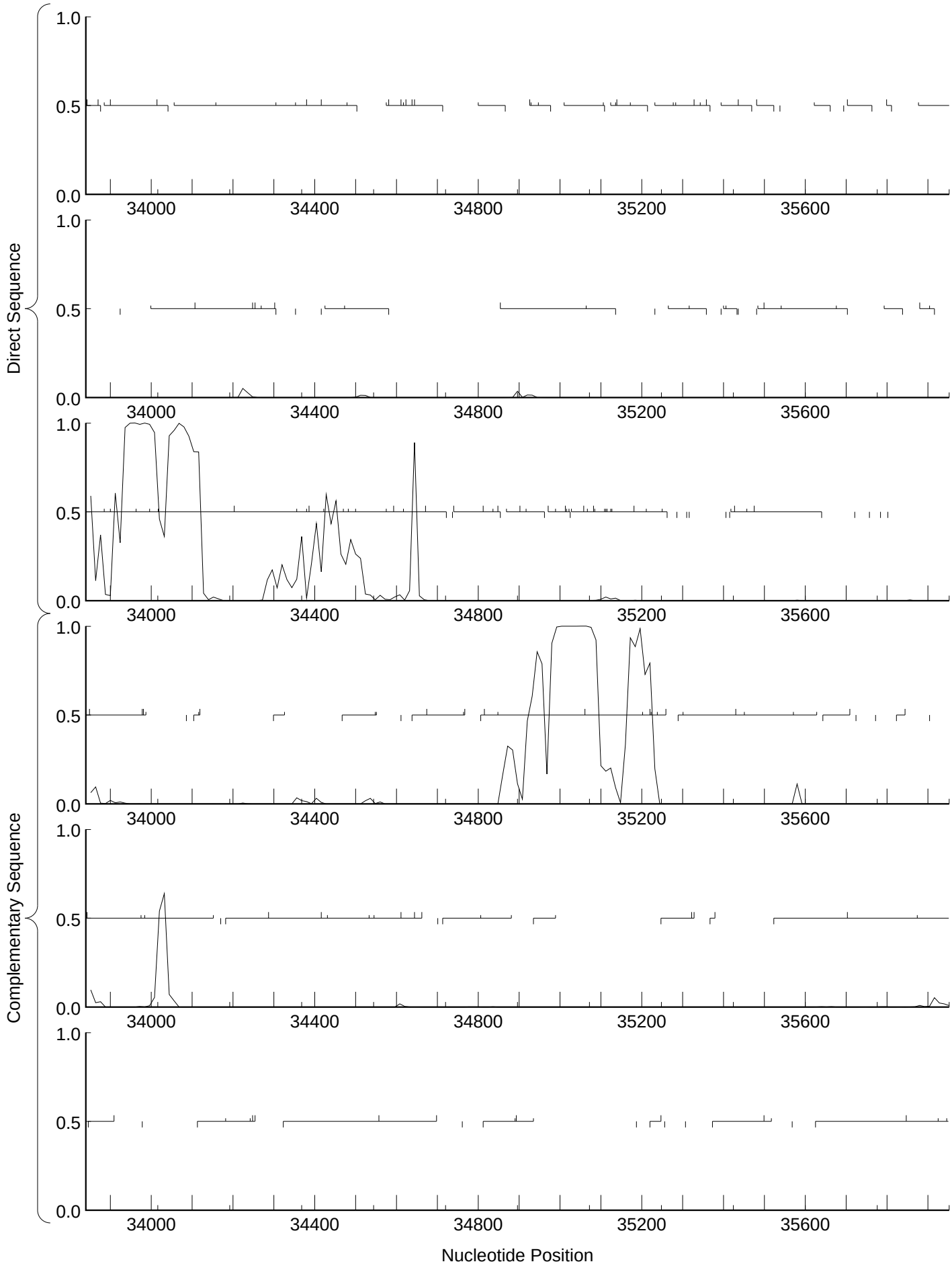
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 16/29



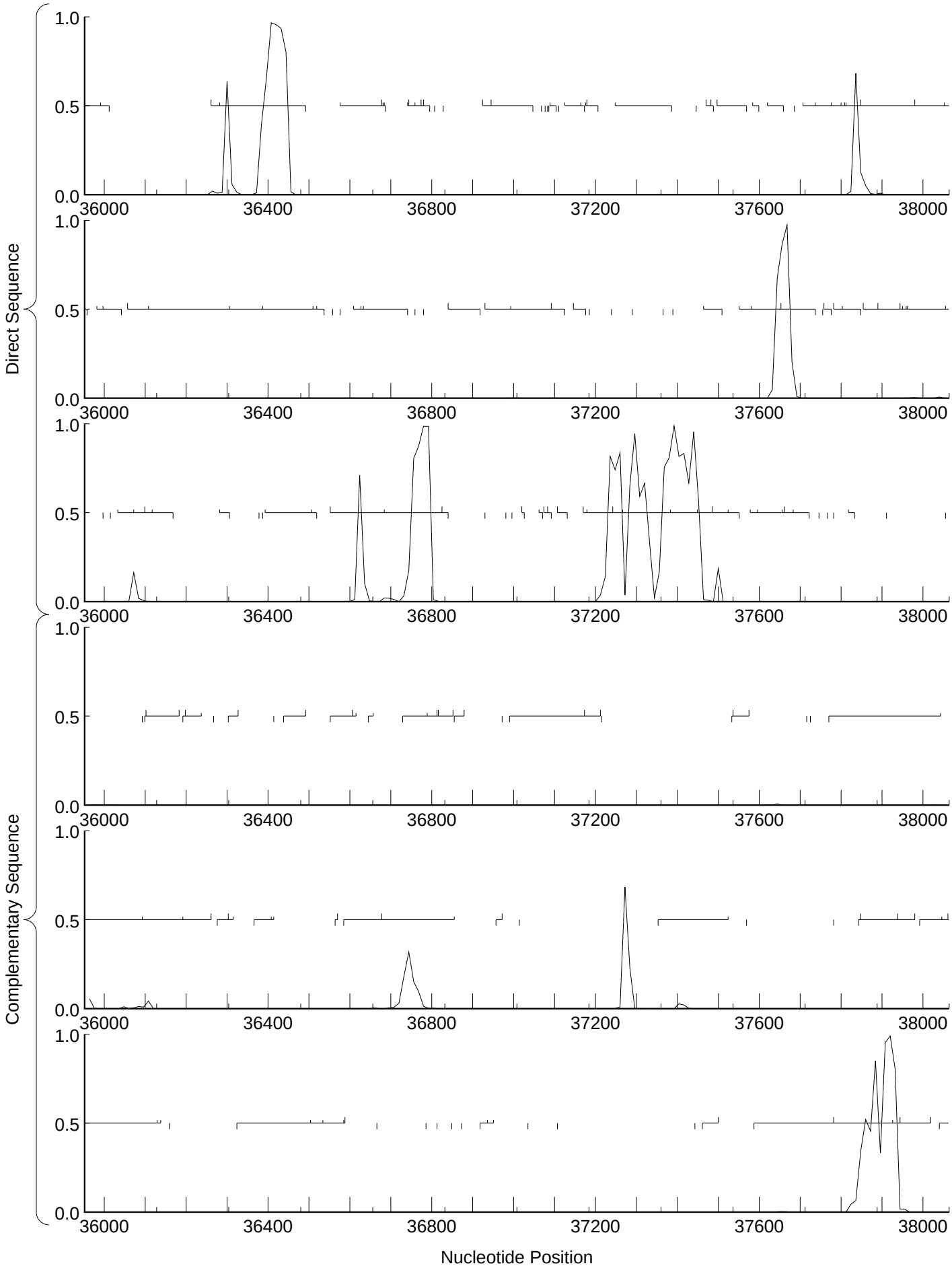
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 17/29



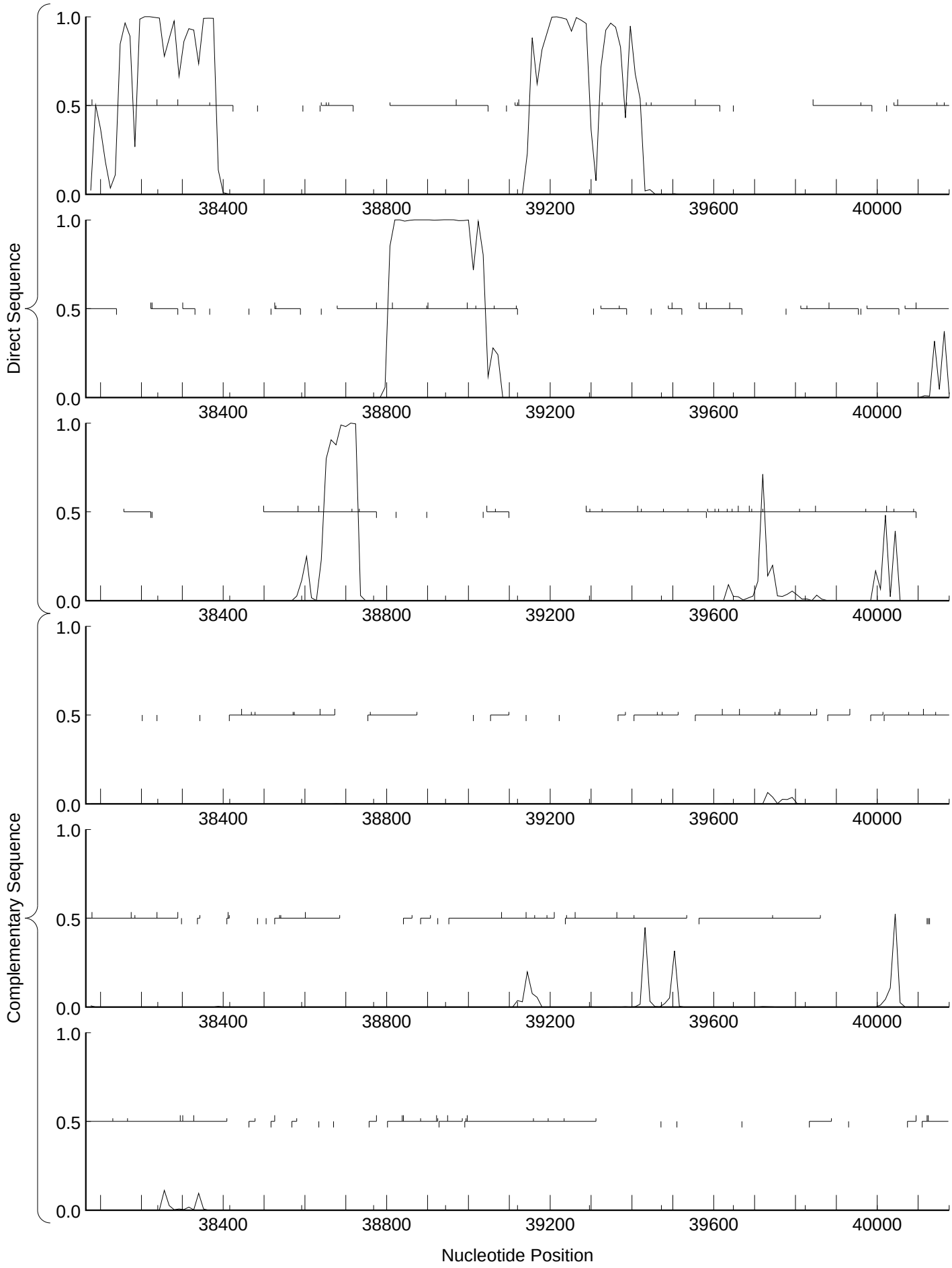
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 18/29



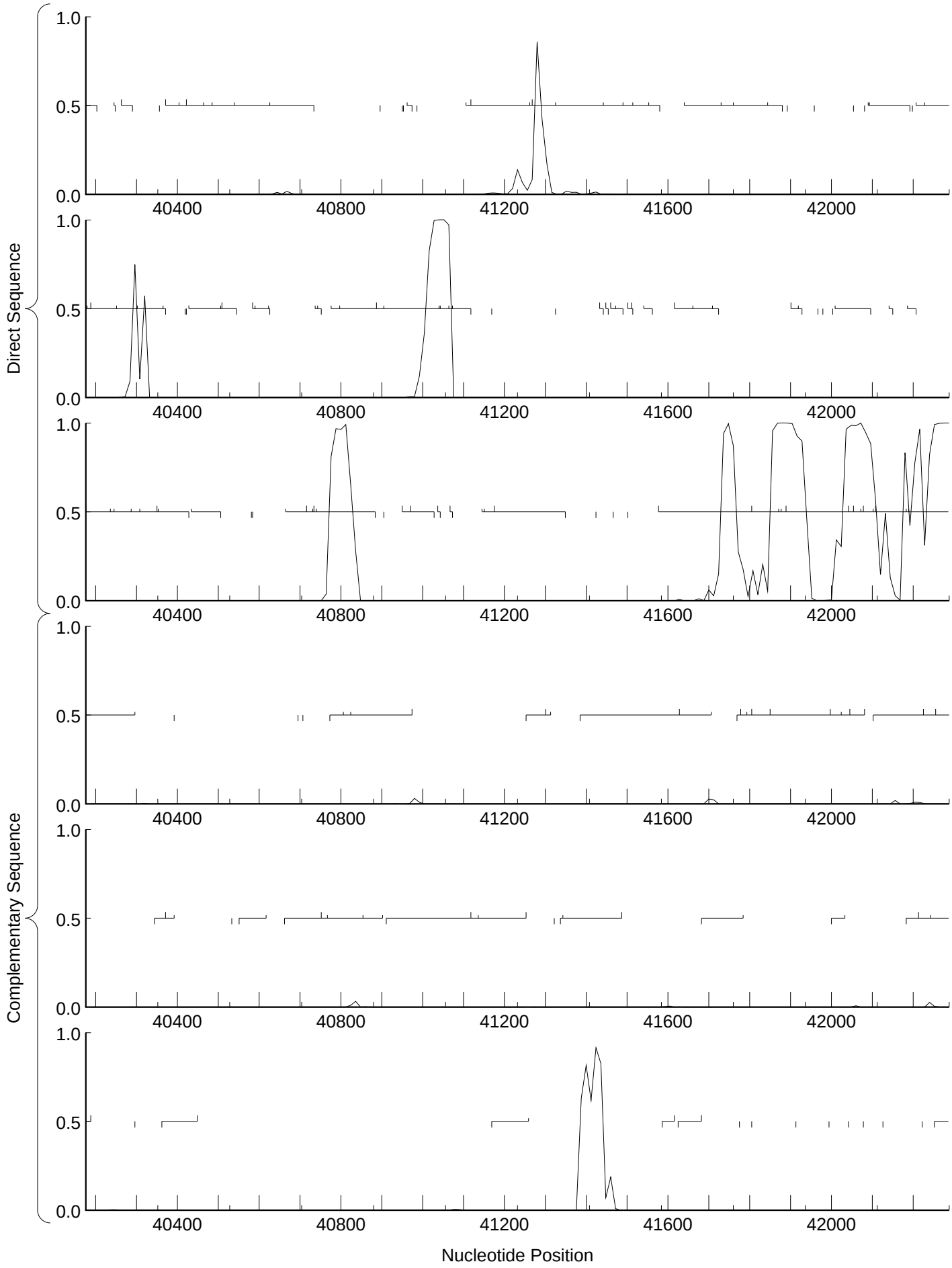
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 19/29



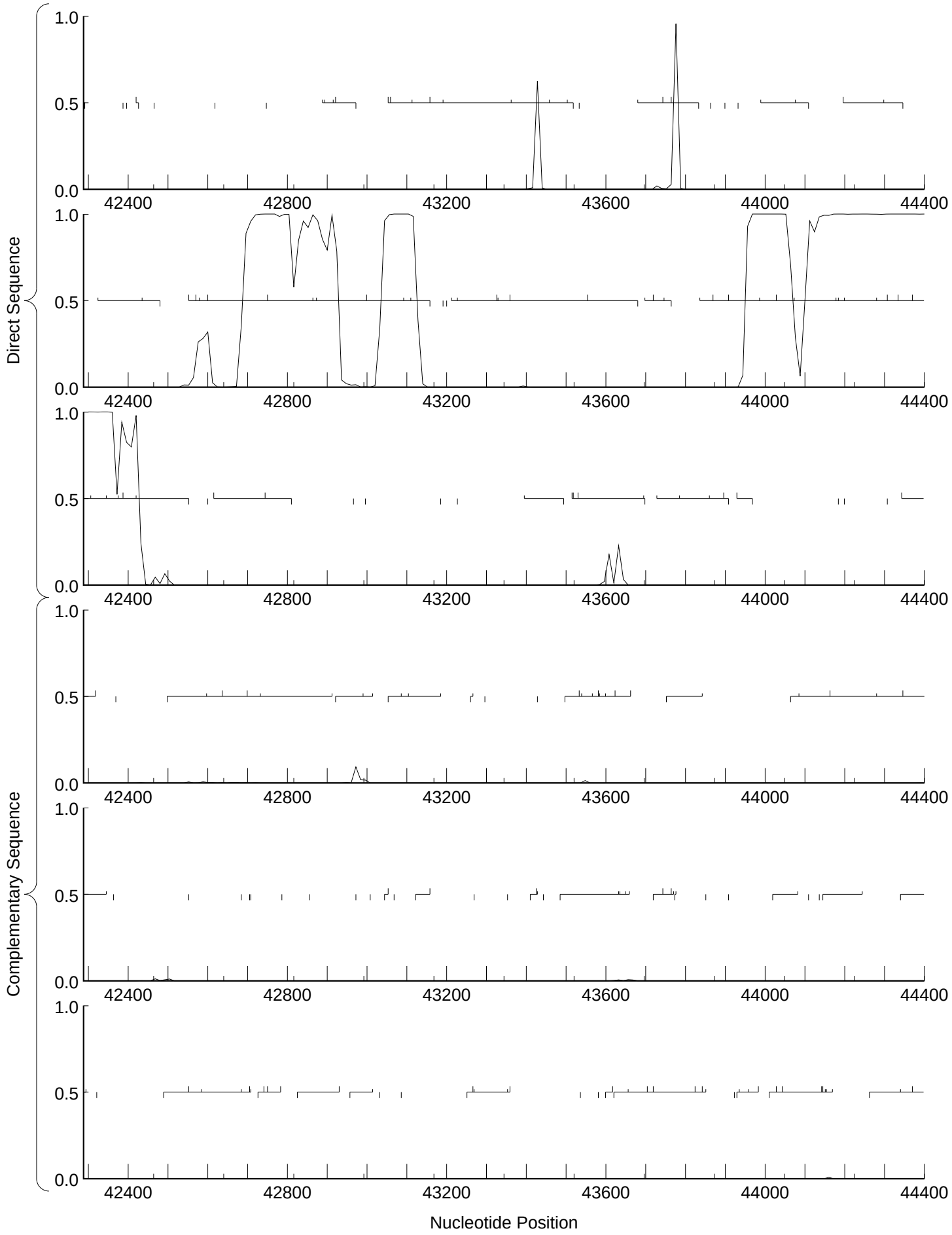
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 20/29



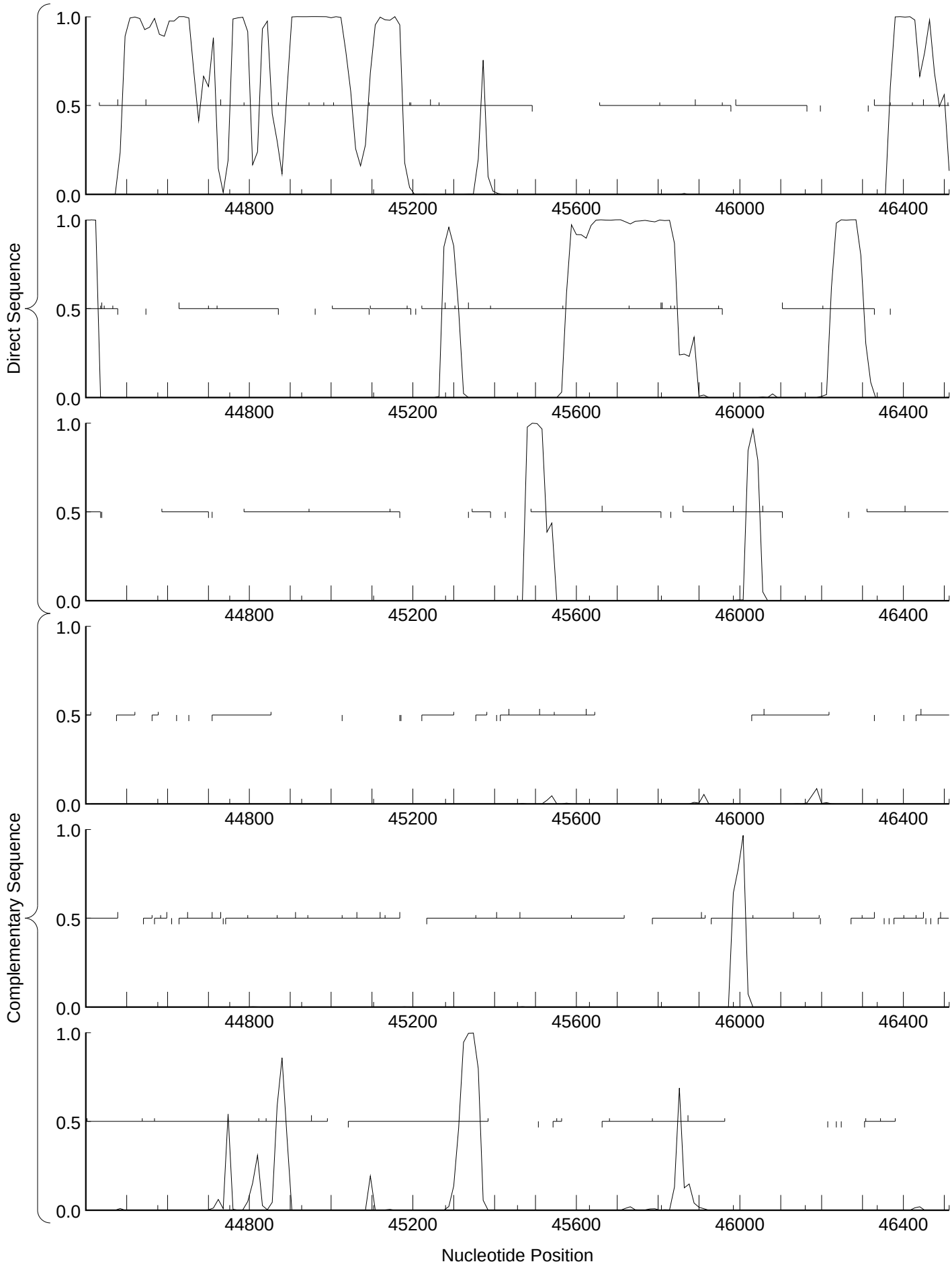
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 21/29



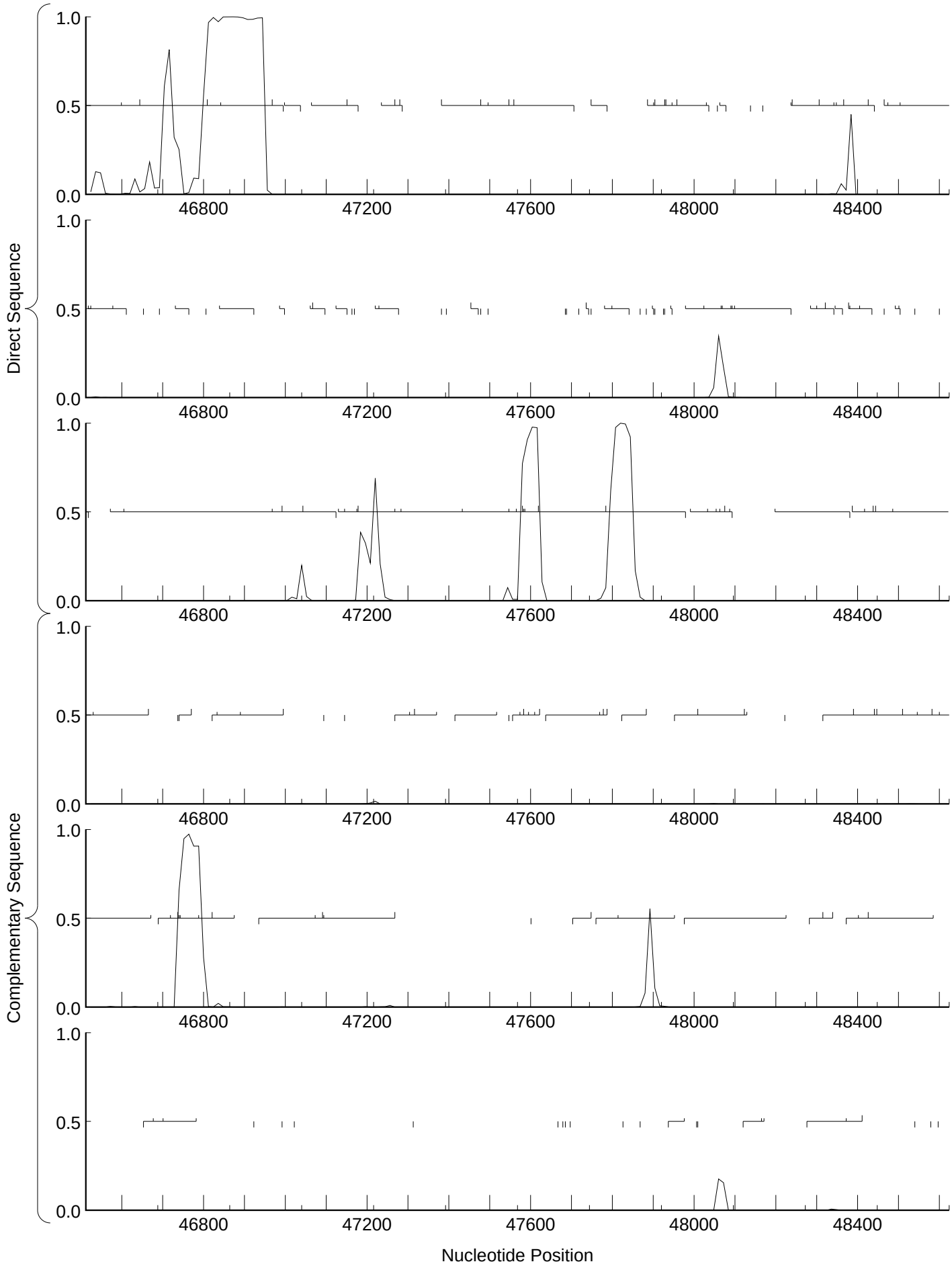
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 22/29



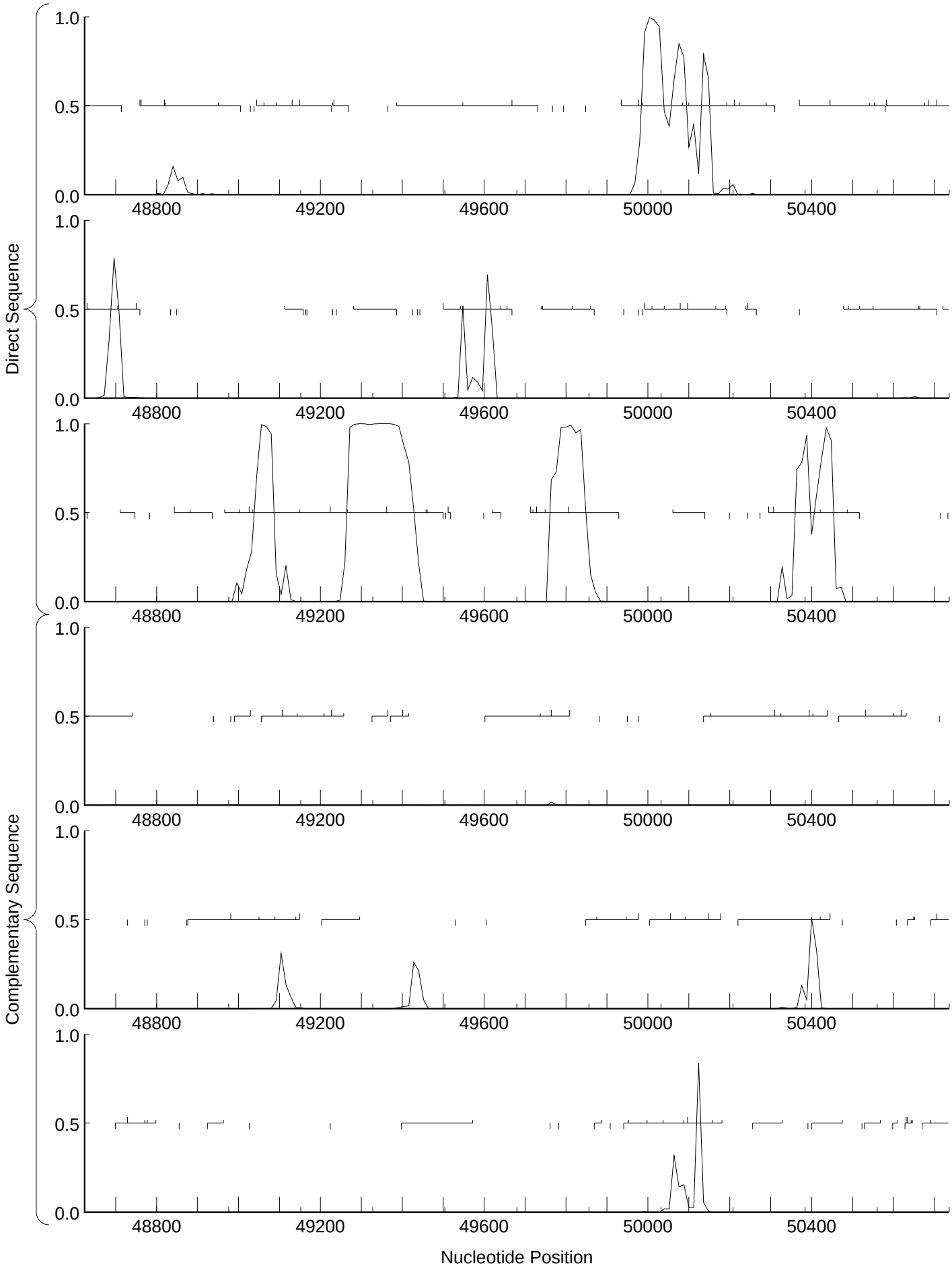
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 23/29



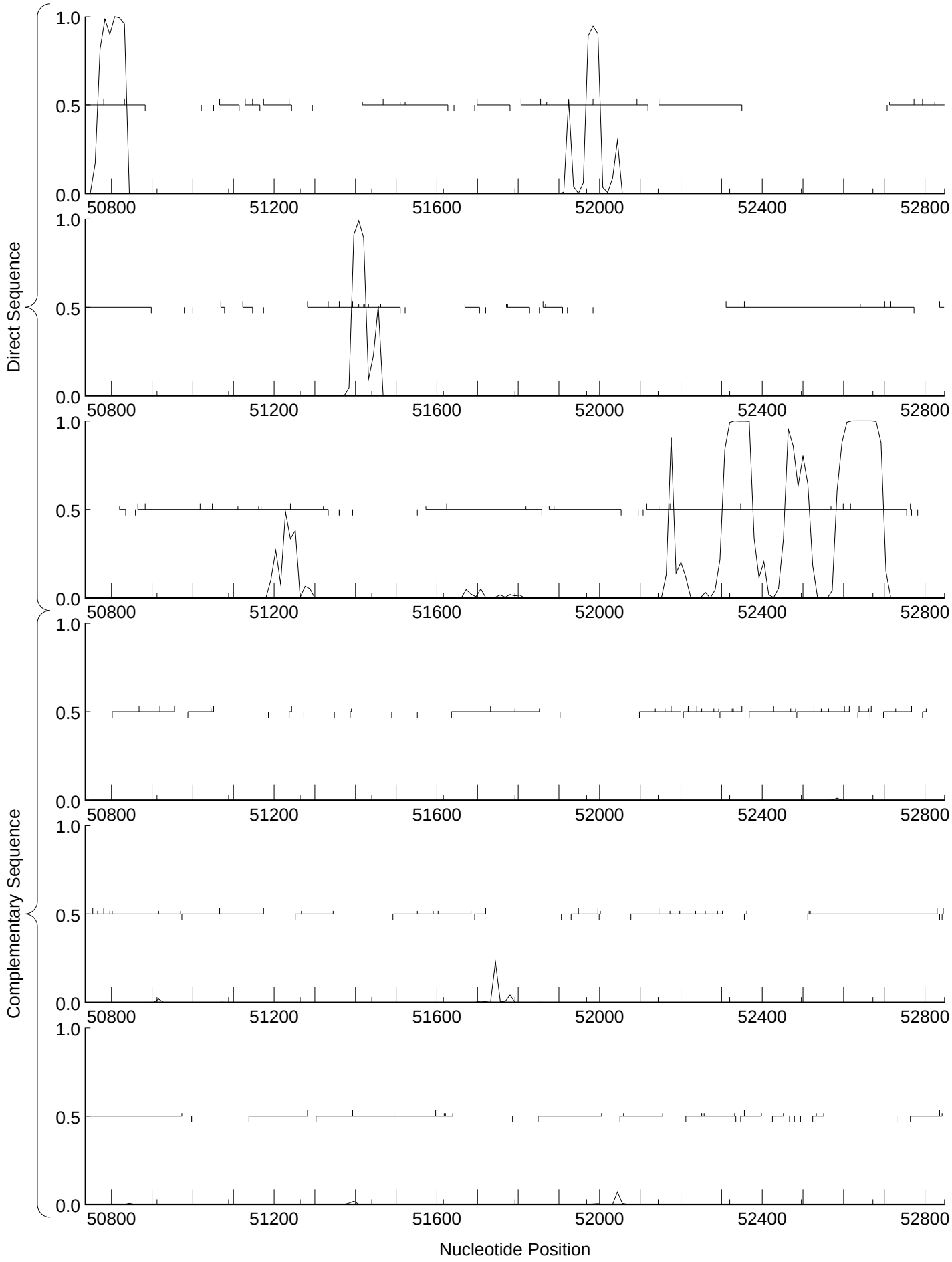
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 24/29



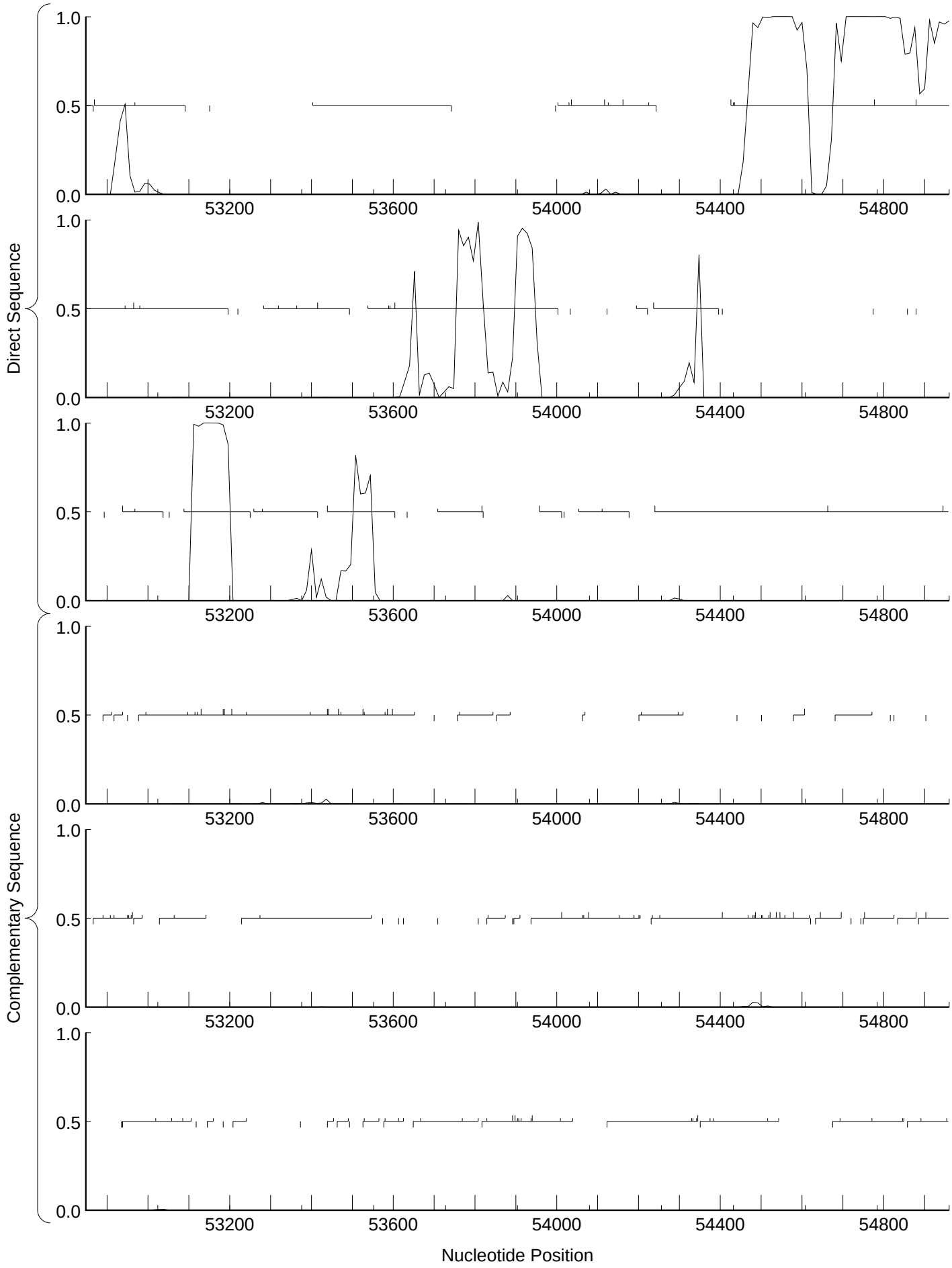
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 25/29



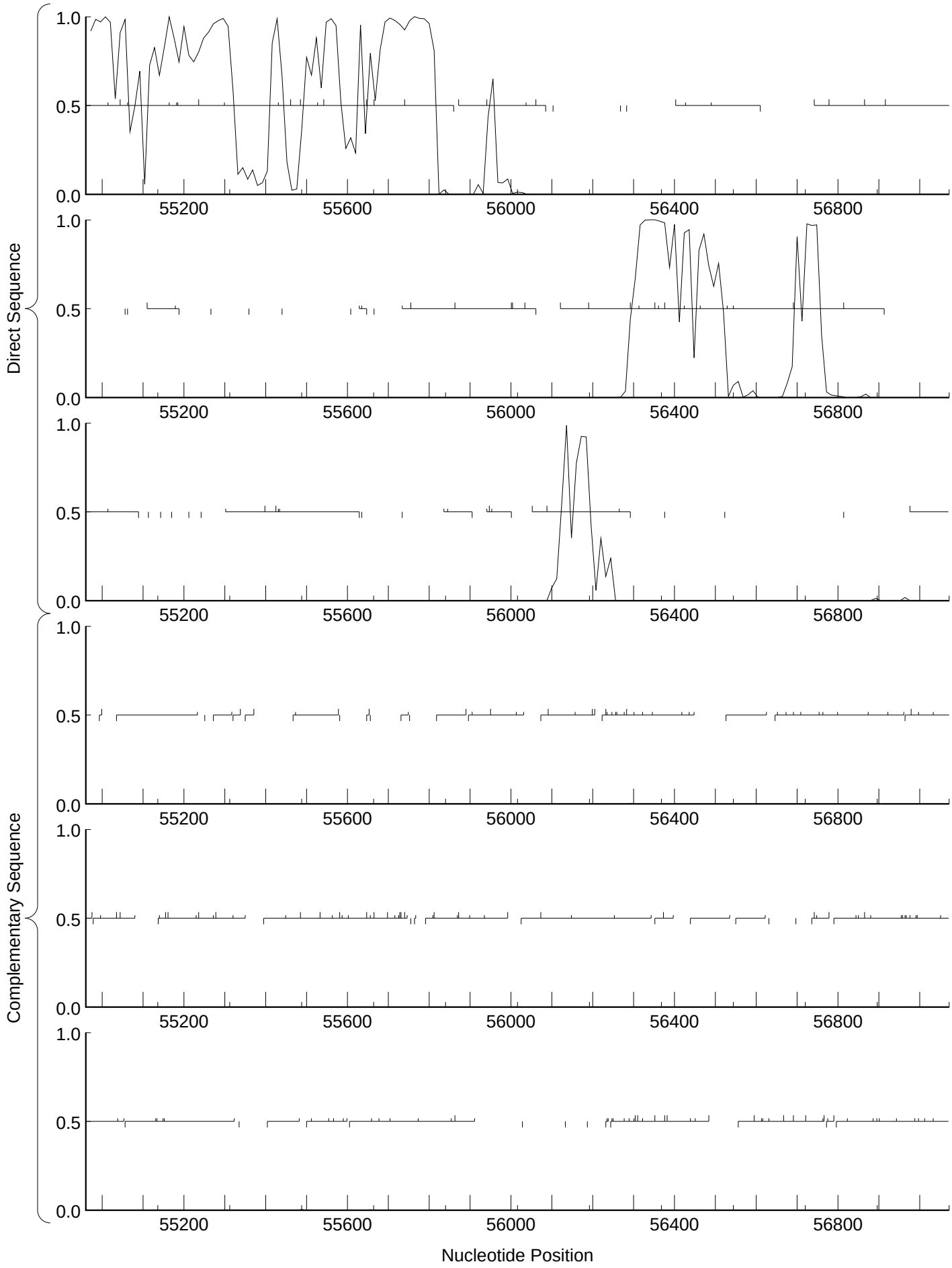
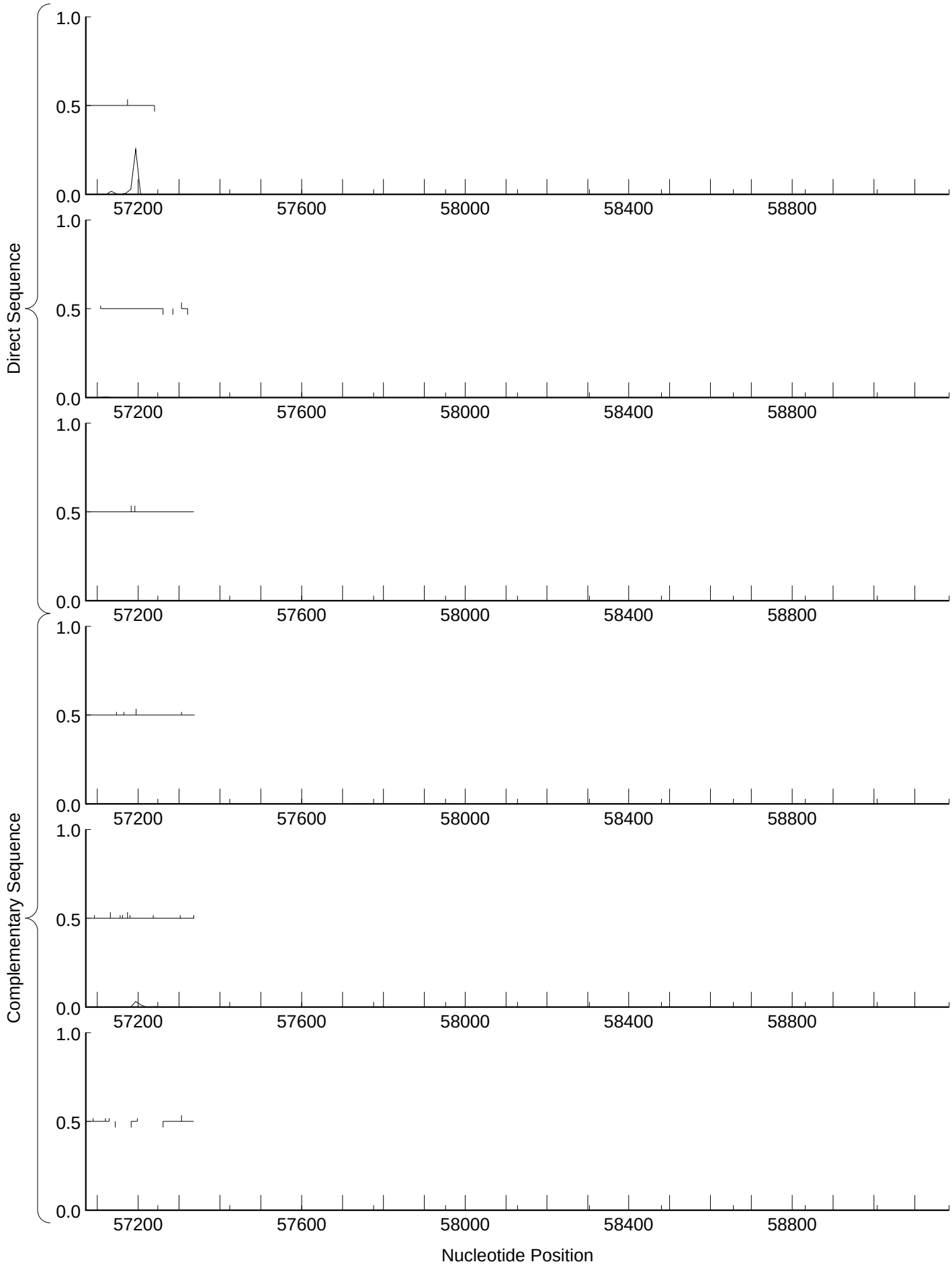
complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 26/29



complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 27/29



complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 28/29



complete sequence, 57386 bp including 10-base 3' overhang (CGGACGGCGC), Cluster F1, Order 4, Window 96, Step 12, 29/29