# Genome Annotation Submission Cover Sheet

**Preliminary Annotation Review Checklist 5-15-2018**

| | |
|---|---|
| Phage Name: | JeTaime |
| 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.

PECAAN was used by College of Charleston students to annotate the adopted mycobacteriophage JeTaime. 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 and TopCons (when TopCons didn't work on PECAAN, we checked it manually using the TopCons website). The absence of tRNAs and tmRNAs was confirmed using online ARAGORN v.2.0 and online tRNAscan-SE. Full annotation was exported from PECAAN and re-BLASTED. A copy of the JeTaime *M. smegmatis* GeneMark file is attached. Thanks for reviewing this!

QUESTIONS ABOUT FUNCTIONAL ASSIGNMENTS:
(DNA Master gene product numbers shown)

gp18: 10851 – 11231 (NKF)
Thought to have been head-to-tail adaptor, but no HHPRED alignment to SPP1 15, HK97 gp6, or the *Bacillus* protein yqbG crystal structure was found. There is the presence of an alignment to HK97 gp10. Should this be called NKF?

gp27: 22235 – 23068 (NKF)
Based on the location of this sequence, this may be a minor tail protein however there was no evidence of homology to minor tail protein in HHpred or the CDD. Were we correct to call it NKF?

gp32: 29951-29766 (helix-turn-helix DNA binding domain)
This gene product has strong homology to Sigma factor. Should we call it this?

gp64: 44183 – 44515 (NKF)
Based on HHpred results, may be a secreted protein. Should we call it NKF?

gp69: 46918 – 47229 (NKF)
Based on HHpred results, this may contain a DNA binding domain. Should we call it NKF?

gp101: 60970 – 61260 (NKF)
HHpred suggested several functions not present in the official functions list. Should we have called this NKF?

gp102: 61265 – 61576 (helix-turn-helix DNA binding domain)
Didn`t call this a terminase because we had already identified 2 terminases in the genome. Is next to a HNH endonuclease as expected for a terminase. Should this be categorized as a terminase?

gp111: 64317 – 65102 (DNA polymerase)
Appears to be a DNA polymerase II, but that function was not listed in official functions list. Should that function be added to the list?

gp123:  68113 – 68244 (NKF)
Predicted to be carbamoyl-phosphate synthase small chain by HHPRED. Should we have called this NKF?

gp140:  73808-73644 (NKF)
HHpred suggests pyruvate kinase. Should we have called this NKF?

EVIDENCE FOR FUNCTIONAL ASSIGNMENTS (TopCons was checked manually):

gp11:  4996 – 5217 (RNA binding protein)
The archaeal Lsm protein binds to small RNAs. Also the sm domain is an RNA-binding motif. Because of this, we called this a RNA binding protein.

gp13:  6763 – 8055 (capsid maturation protease)
Caudovirus prohead serine protease and phage serine protease XkdF are orthologous groups highly similar to capsid maturation protease according to the PhROG database. Based on this and other evidence, we categorized this as capsid maturation protease.

gp16:  9938 – 10504 (head-to-tail adaptor)
Evidence found included alignment to SPP1 15, HK97 gp6, and Bacillus protein yqbG.

gp17:  10519 – 10854 (head-to-tail stopper)
Evidence found included alignment to SPP1 16.

gp19:  11209 – 11613 (tail terminator)
Evidence found included alignment to SPP1 17 and Lambda U (3FZ2_chains A-F).

gp 22 and 23:  13268 – 13729 and 13268-14139 (tail assembly chaperones)
Ribosomal slippage
-1 frameshift
CDS join (13268..13720;13720..14139)

gp35:  31948 – 32331 (holin)
Confirmed presence of two transmembrane domains in TOPCONS and TMHMM as expected in a holin.

gp97:  58624 – 59160 (helix-turn-helix DNA binding domain)
Functional searches identified zinc ribbon domain hits. We called this helix-turn-helix DNA binding domain because zinc ribbon domains are helix-turn-helix domains.

GENES ASSIGNED AS NKF (FORMERLY CALLED MEMBRANE PROTEINS) THAT HAVE TRANSMEMBRANE DOMAINS IDENTIFIED BY BOTH TMHMM AND TOPCONS:

gp33: 30084 – 30368 (NKF)
Both TOPCONS and TMHMM confirmed presence of one transmembrane domain.

gp88: 54667 – 54846 (NKF)
Two transmembrane domains detected in both TOPCONS and TMHMM.

gp89: 54843 – 54941 (NKF)
Kept even though too short because one transmembrane domain detected in both TOPCONS and TMHMM.

# GeneMark

*Version 2.5p (09.08.06)* 

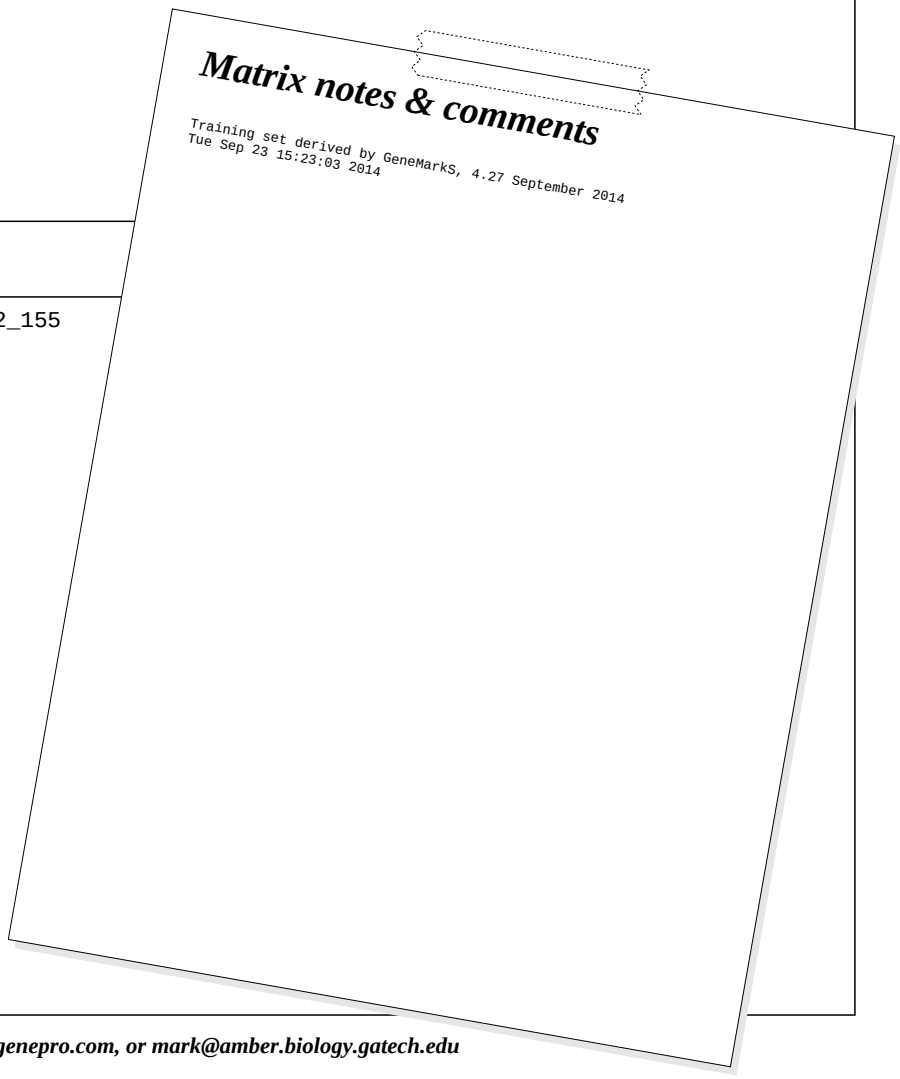
## PROGRAM INFORMATION

```
Sequence         :  Mycobacterium phage JeTaime complete sequence, 75099 bp including 9-base 3' overhang (CGCTT
Analysis Date    :  1/27/21 at 6:46:28
Pages            :  37
Sequence Length  :  75099 bp
GC Content       :   62.94%

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

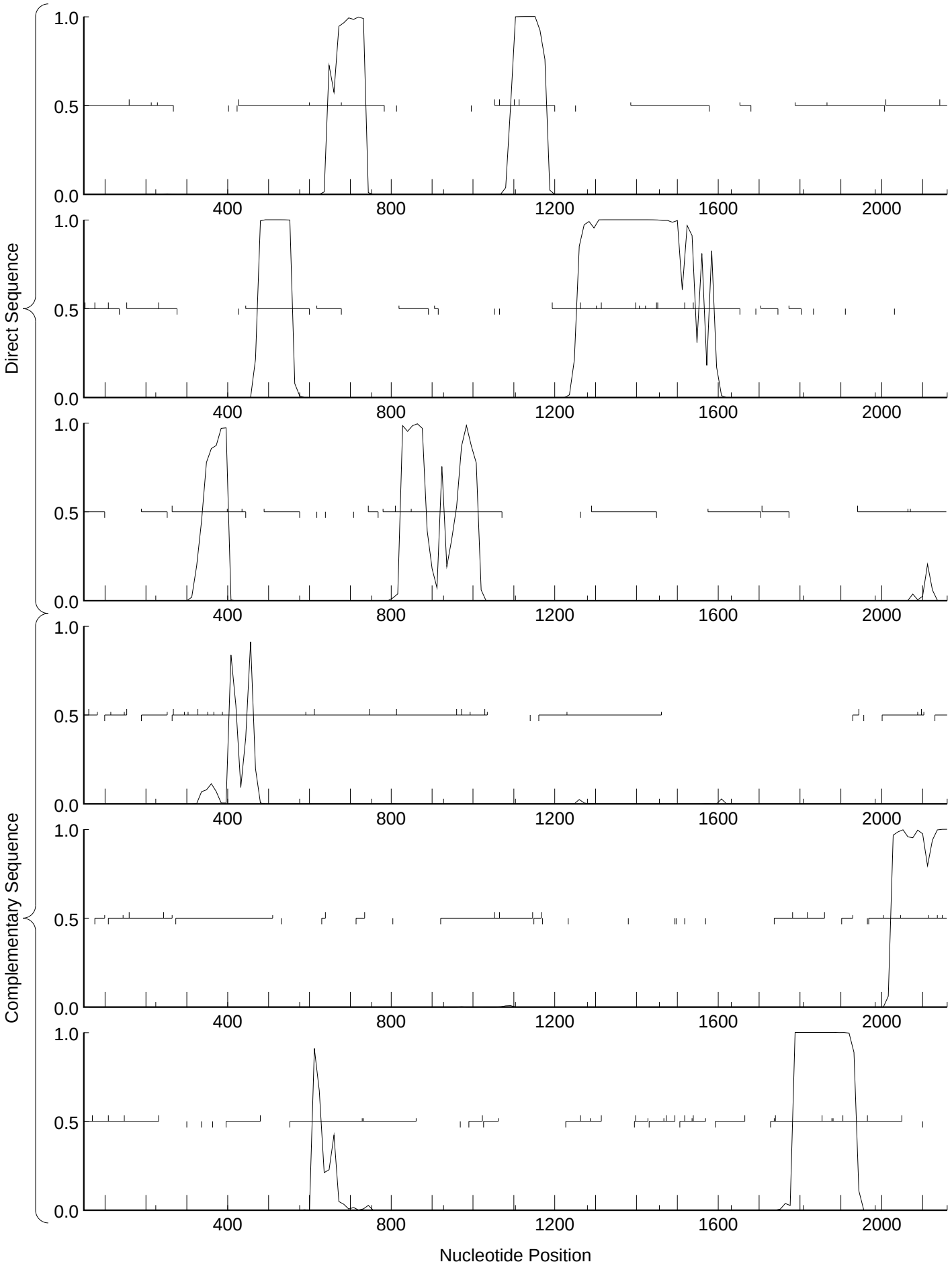
```
Matrix     :  Mycobacterium_smegmatis_MC2_155
Author     :  -

Order      :  4
```

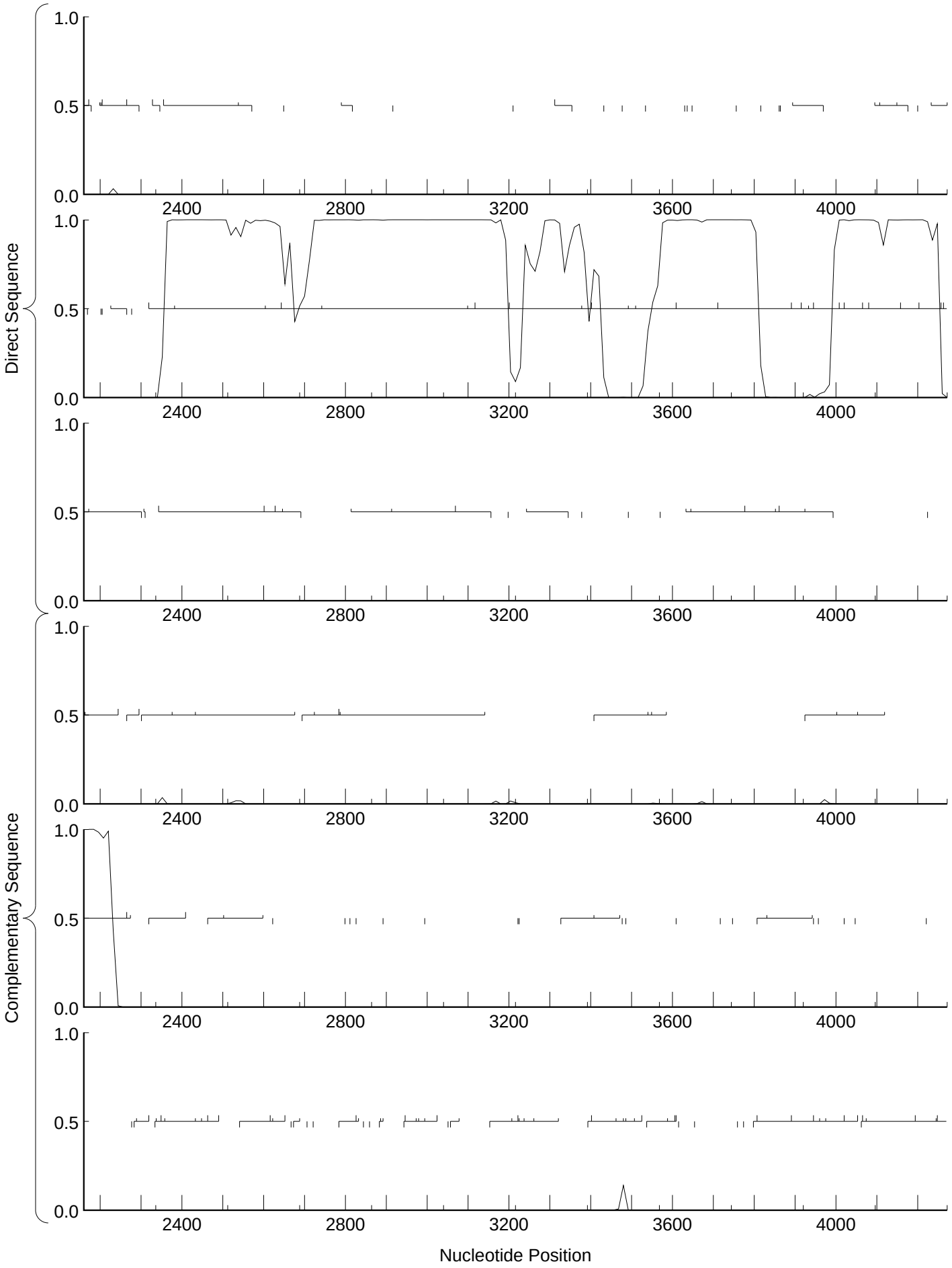

```
  Send questions / comments to:
      Dr. M. Borodovsky
      Georgia Institute of Technology
      School of Biology
      Atlanta, GA  30332-0230
```


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

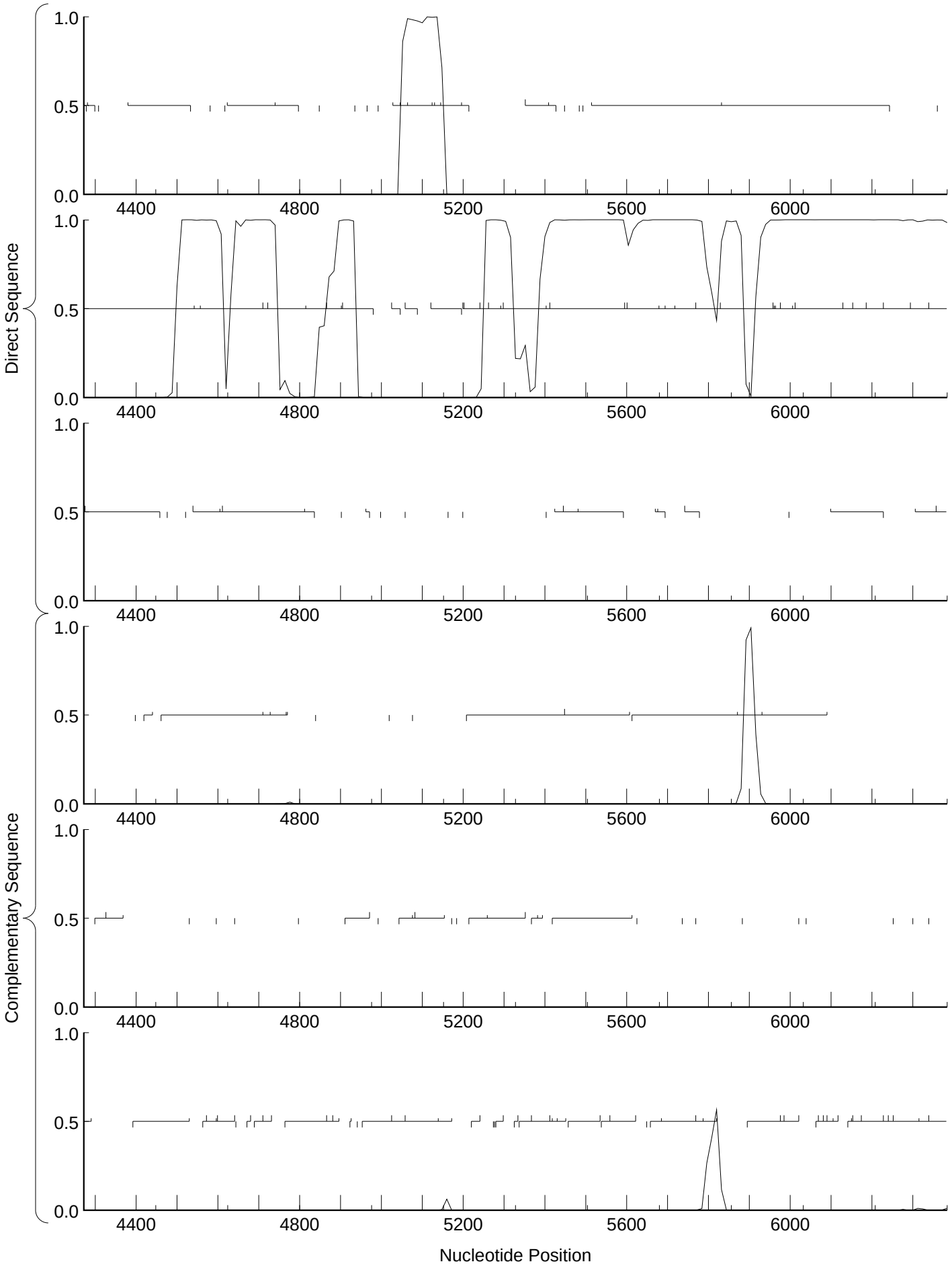
Taime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 2/37



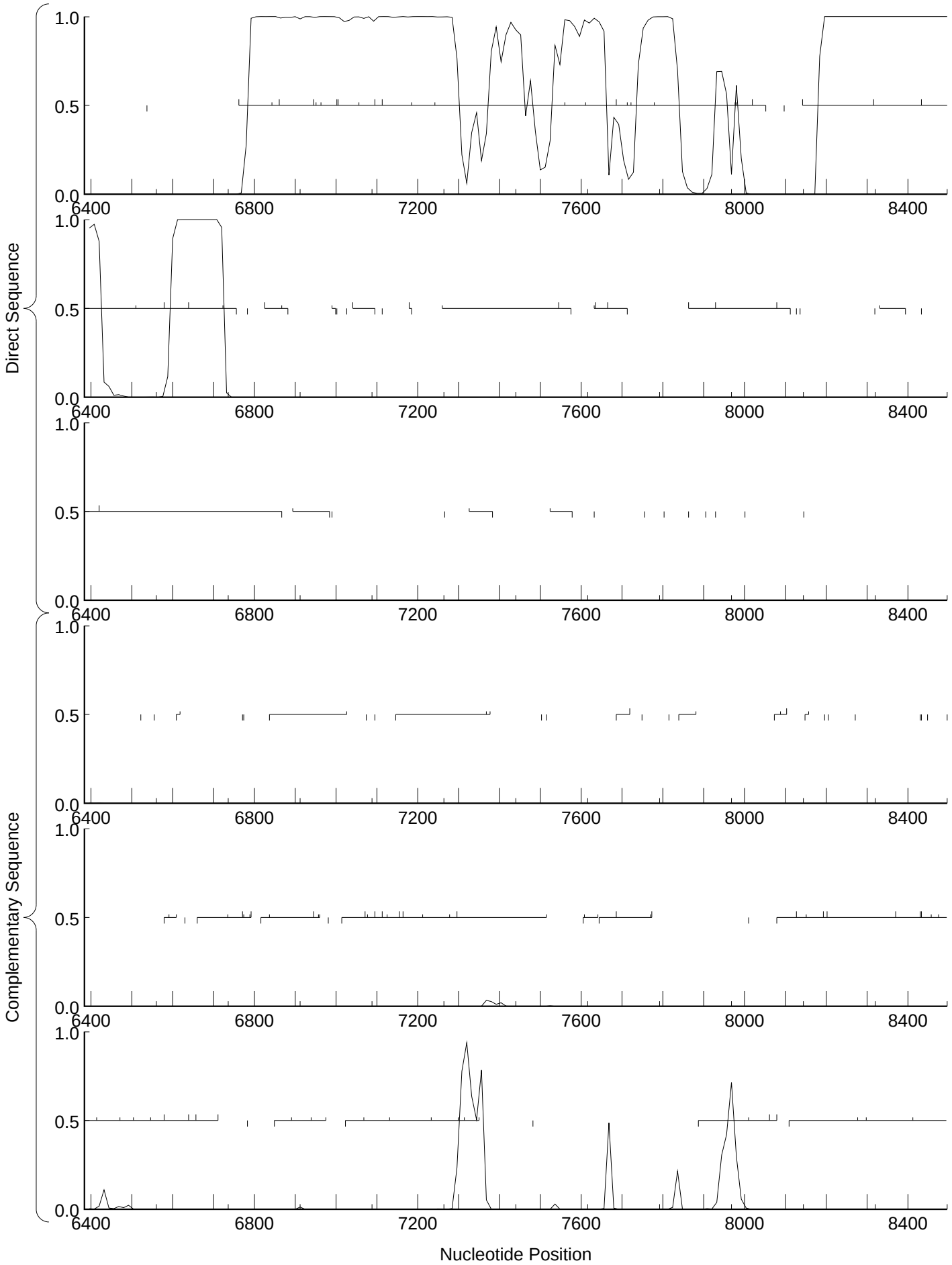
Taime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 3/37



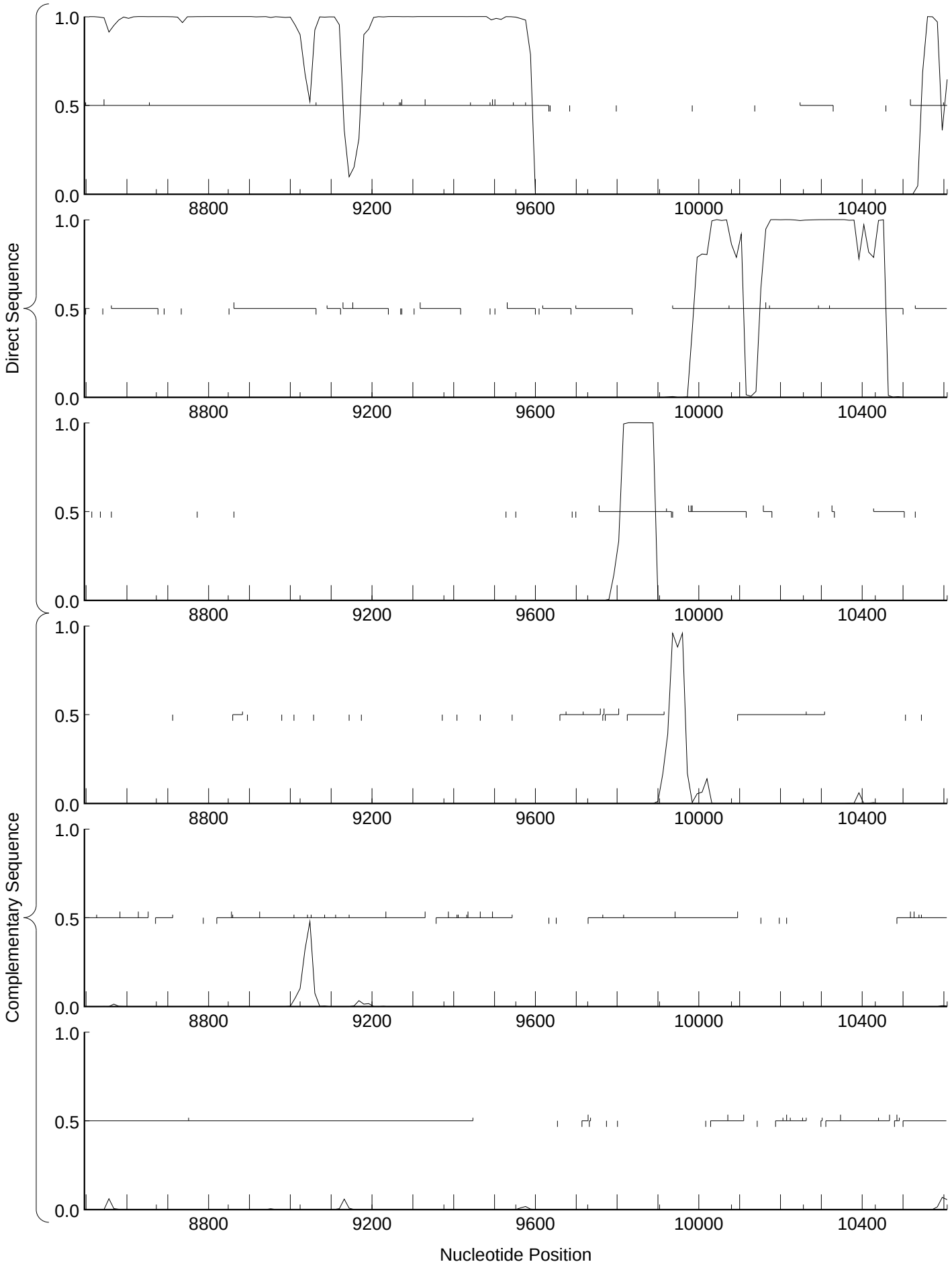
Taime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 4/37



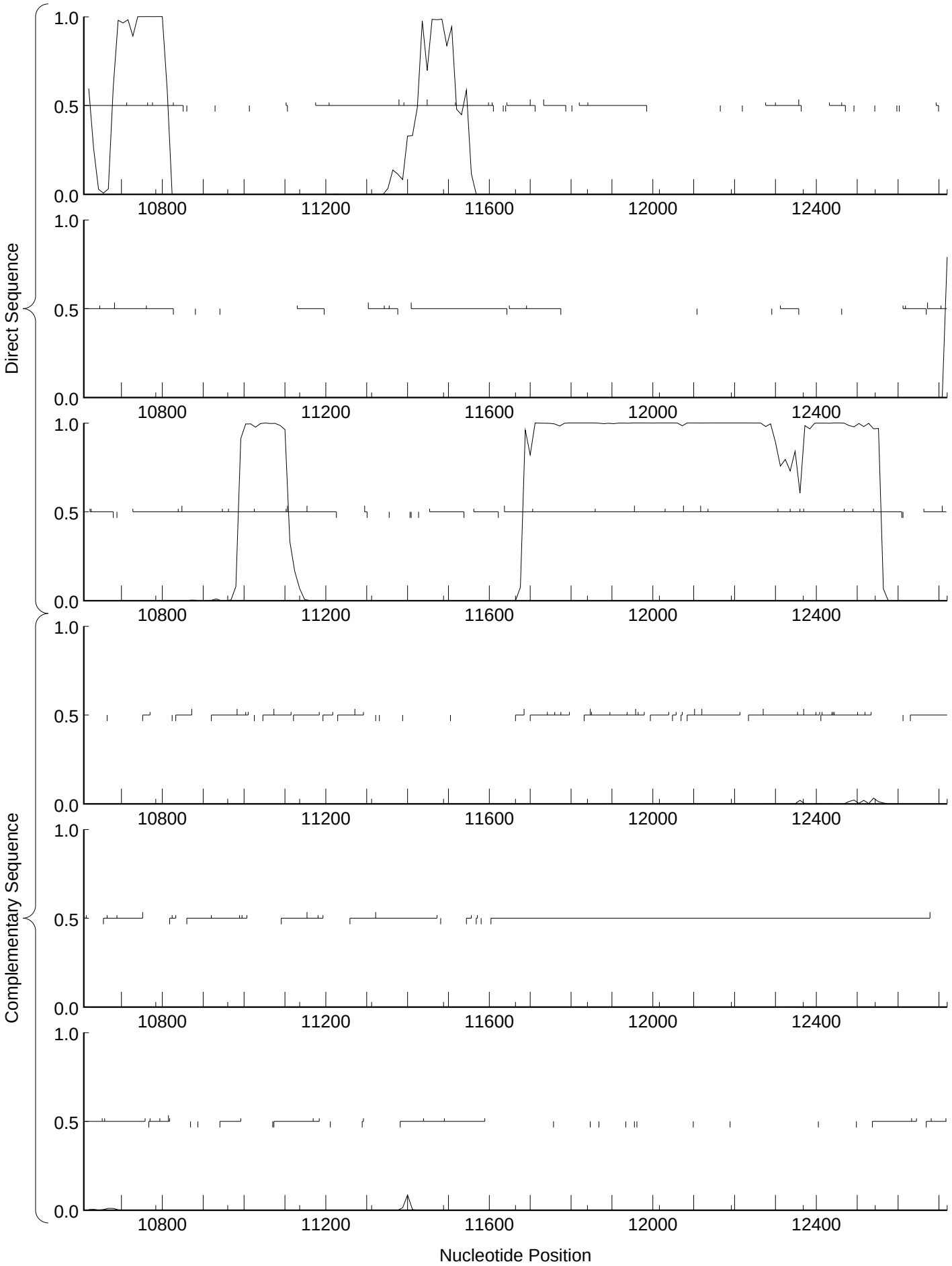
Taime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 5/37



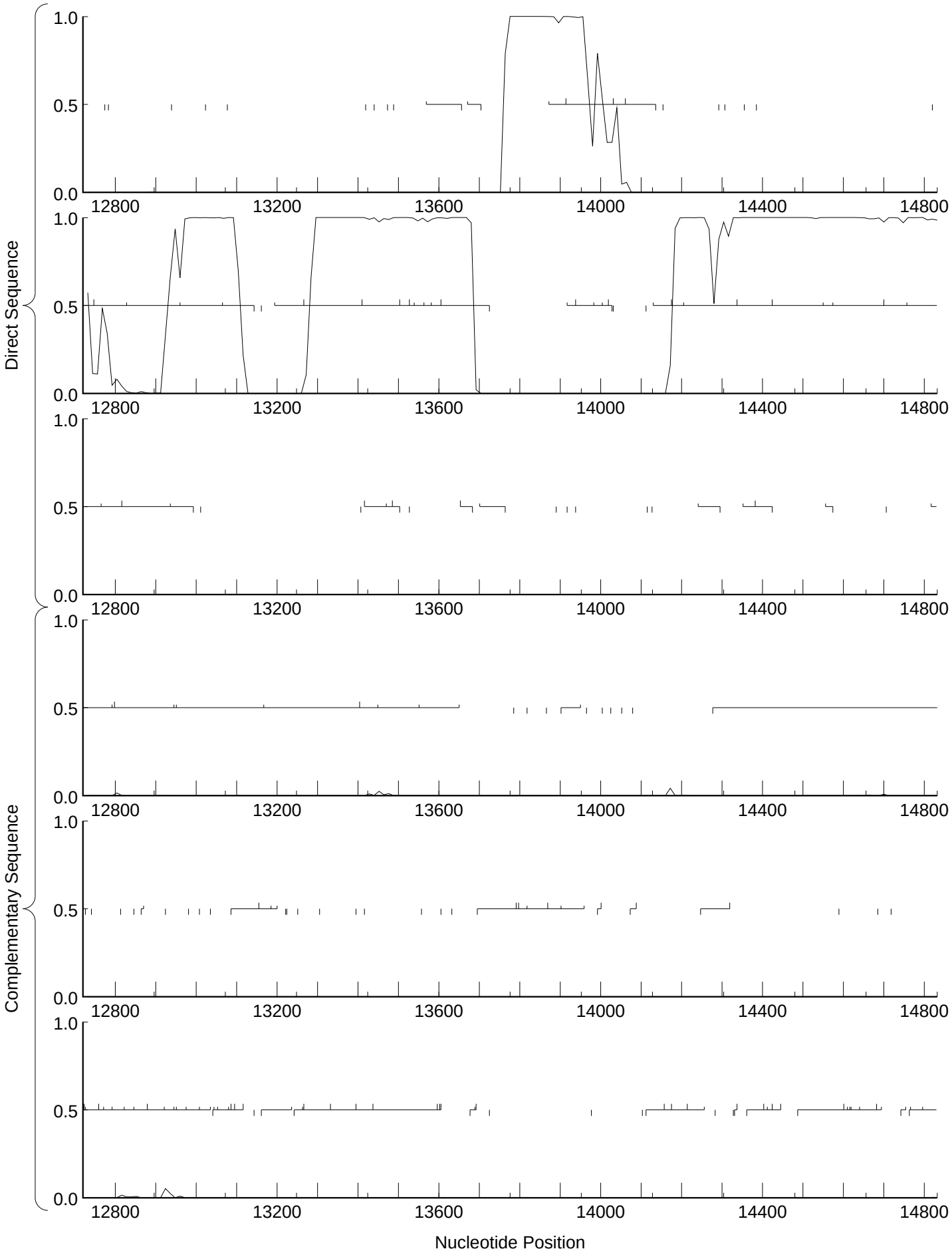
Taime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 6/37



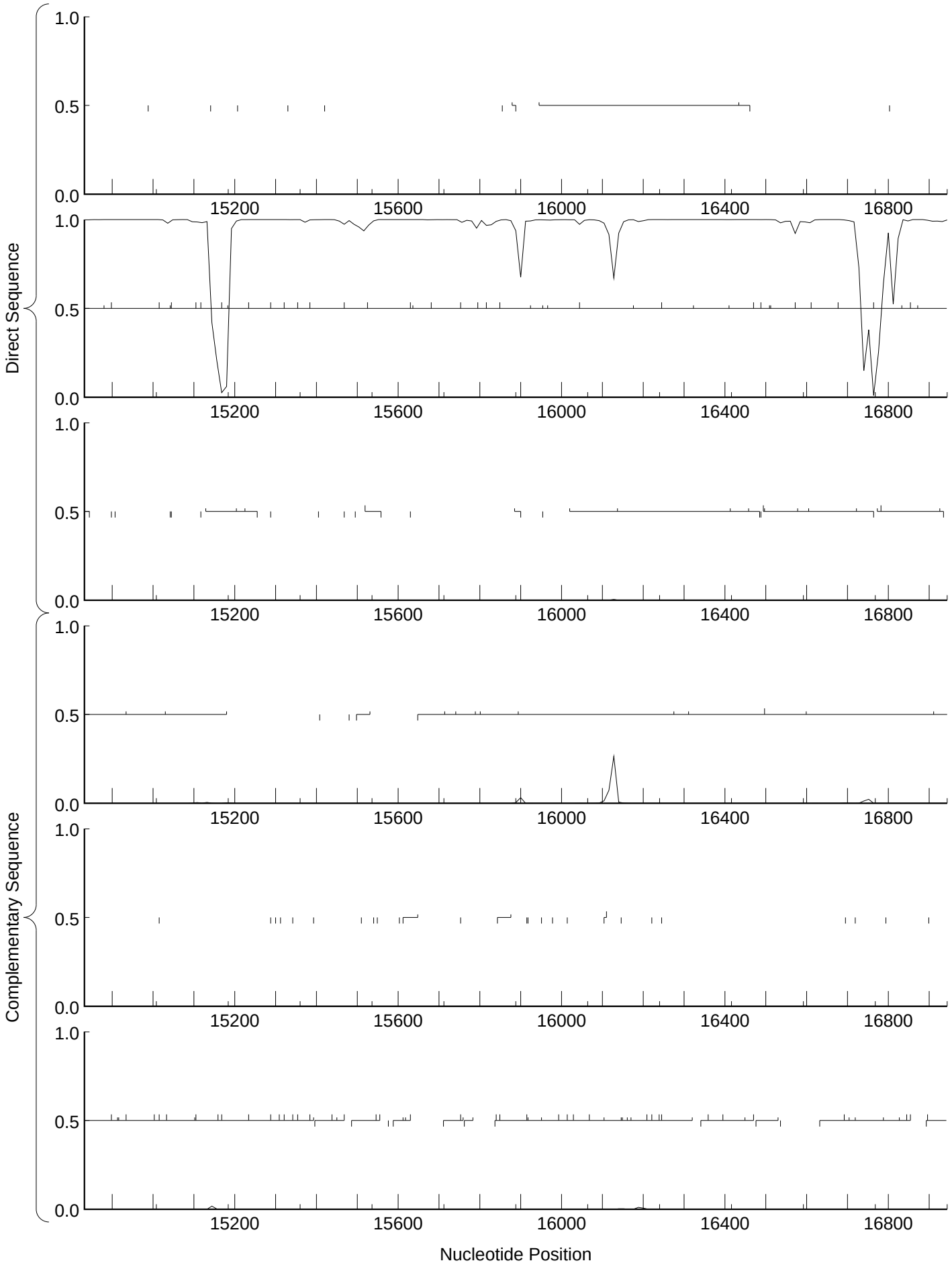
Taime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 7/37



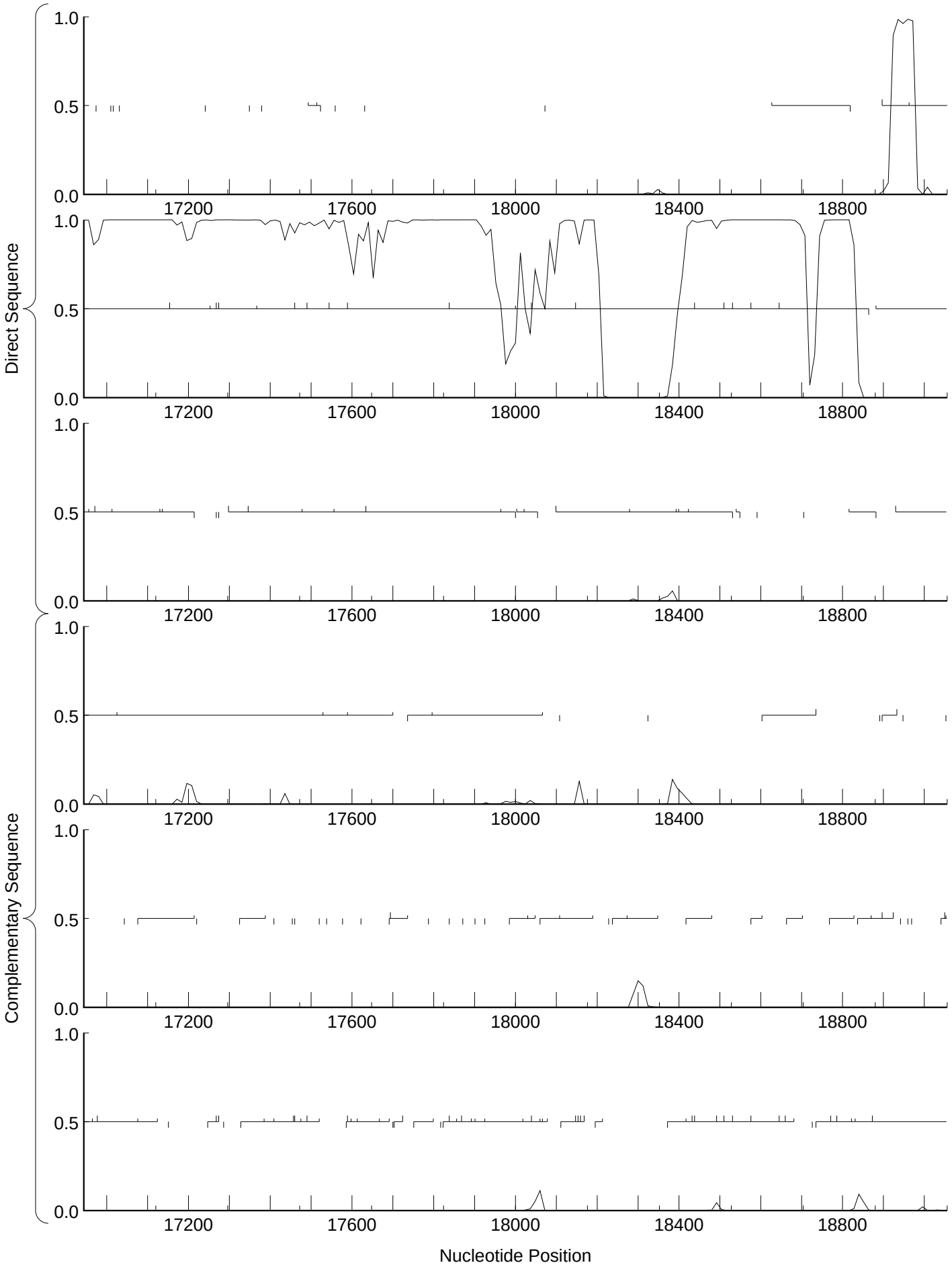
Taime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 8/37



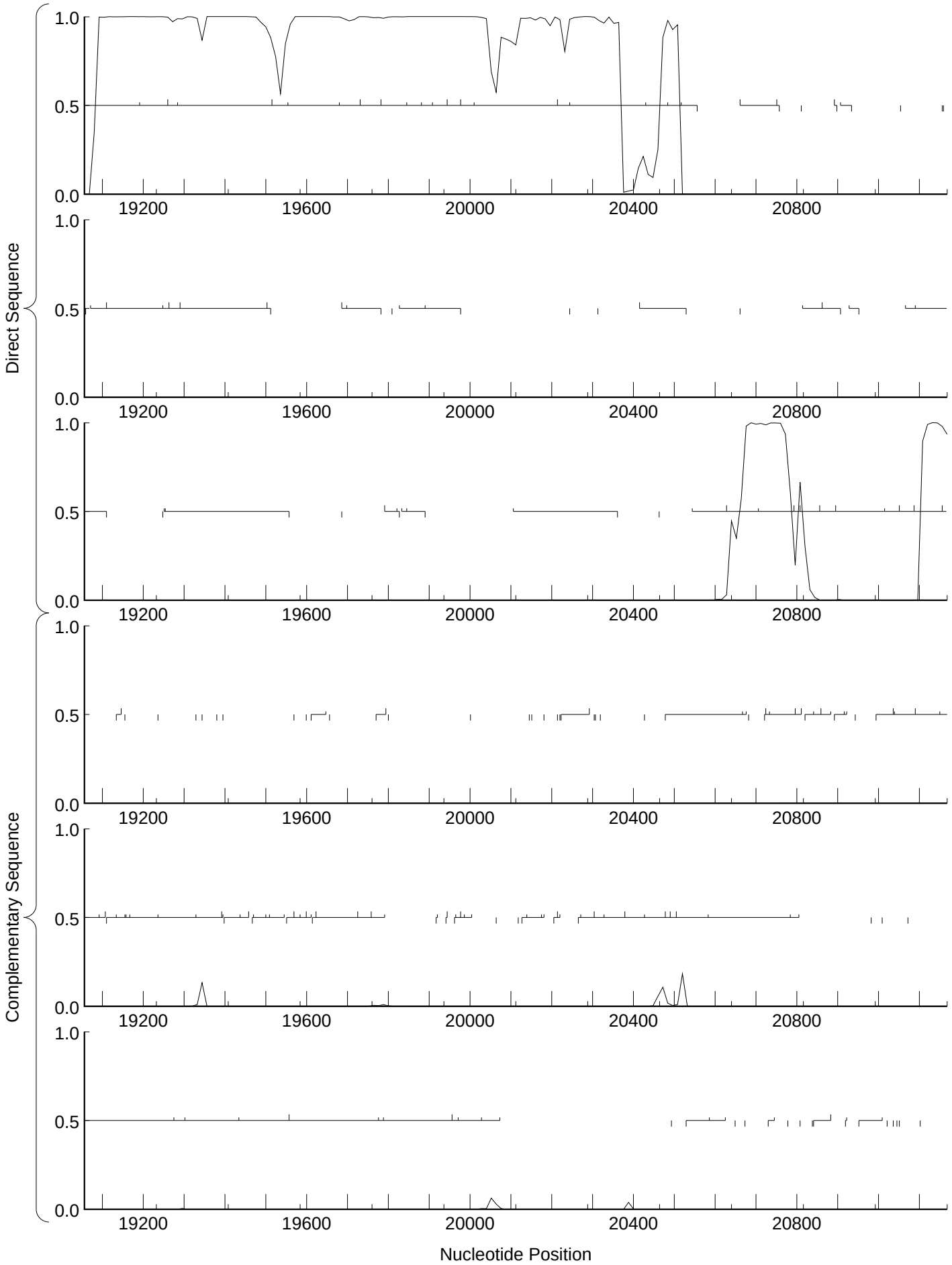
Taime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 9/37



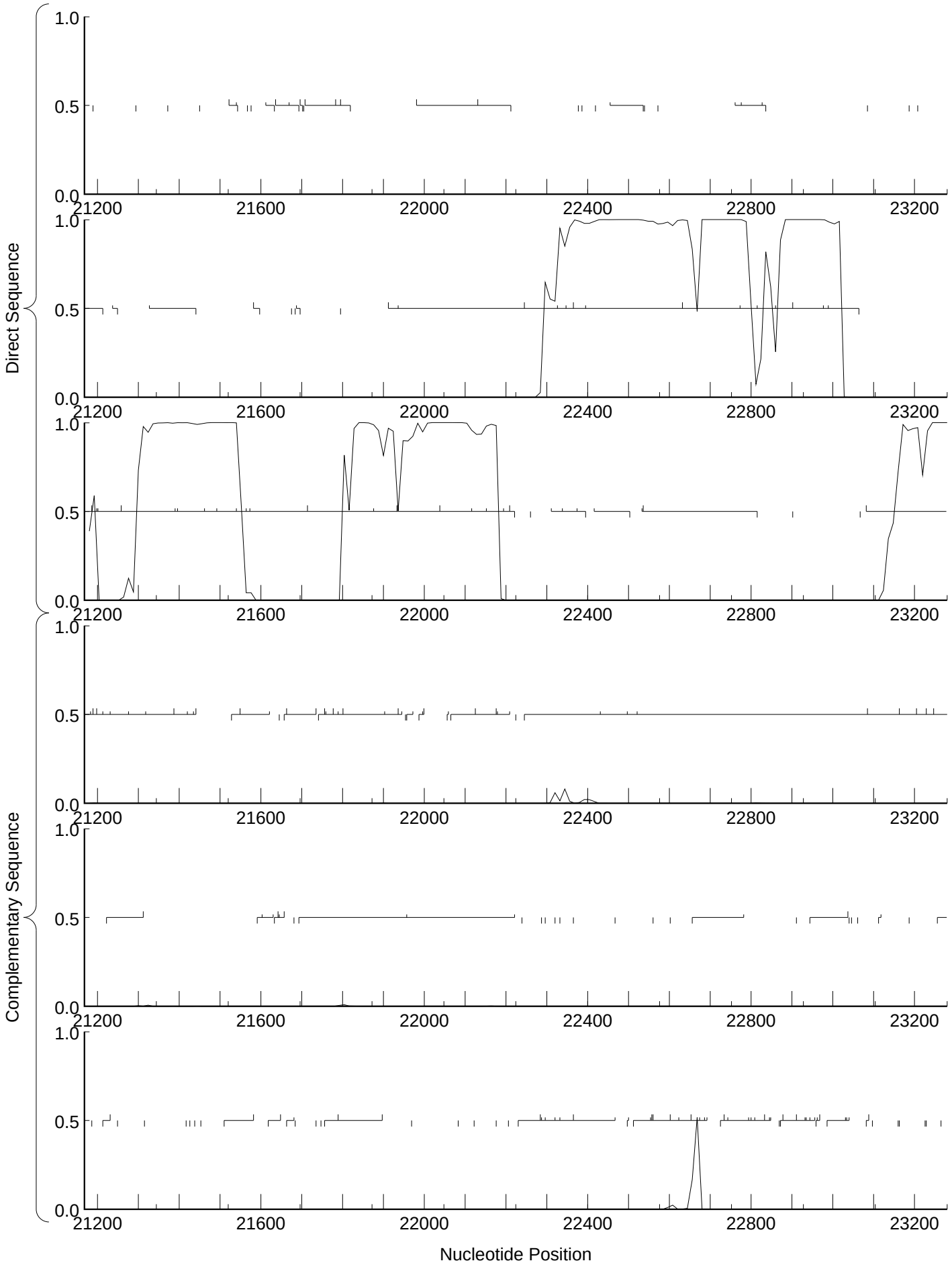
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 10/37



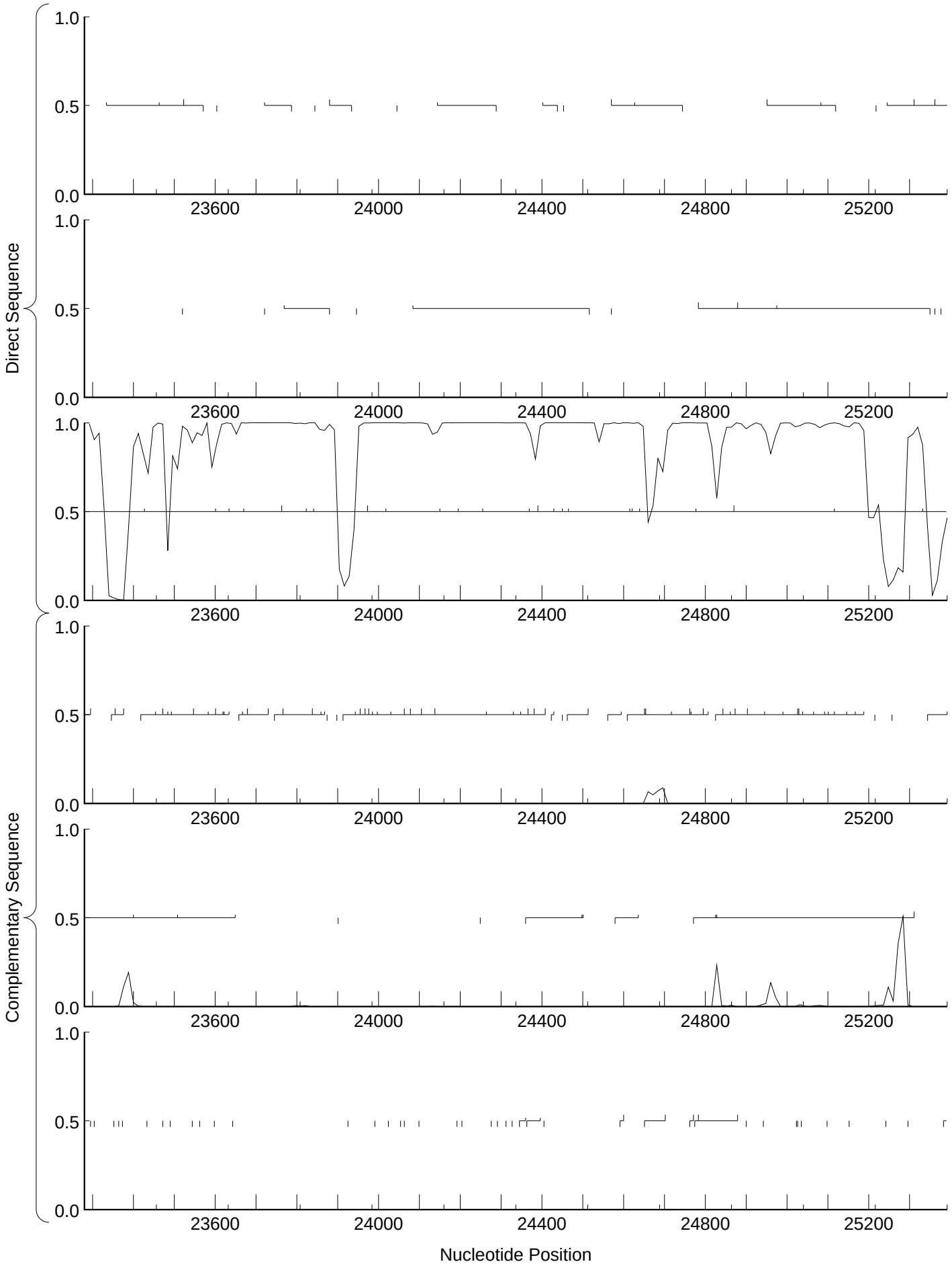
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 11/37



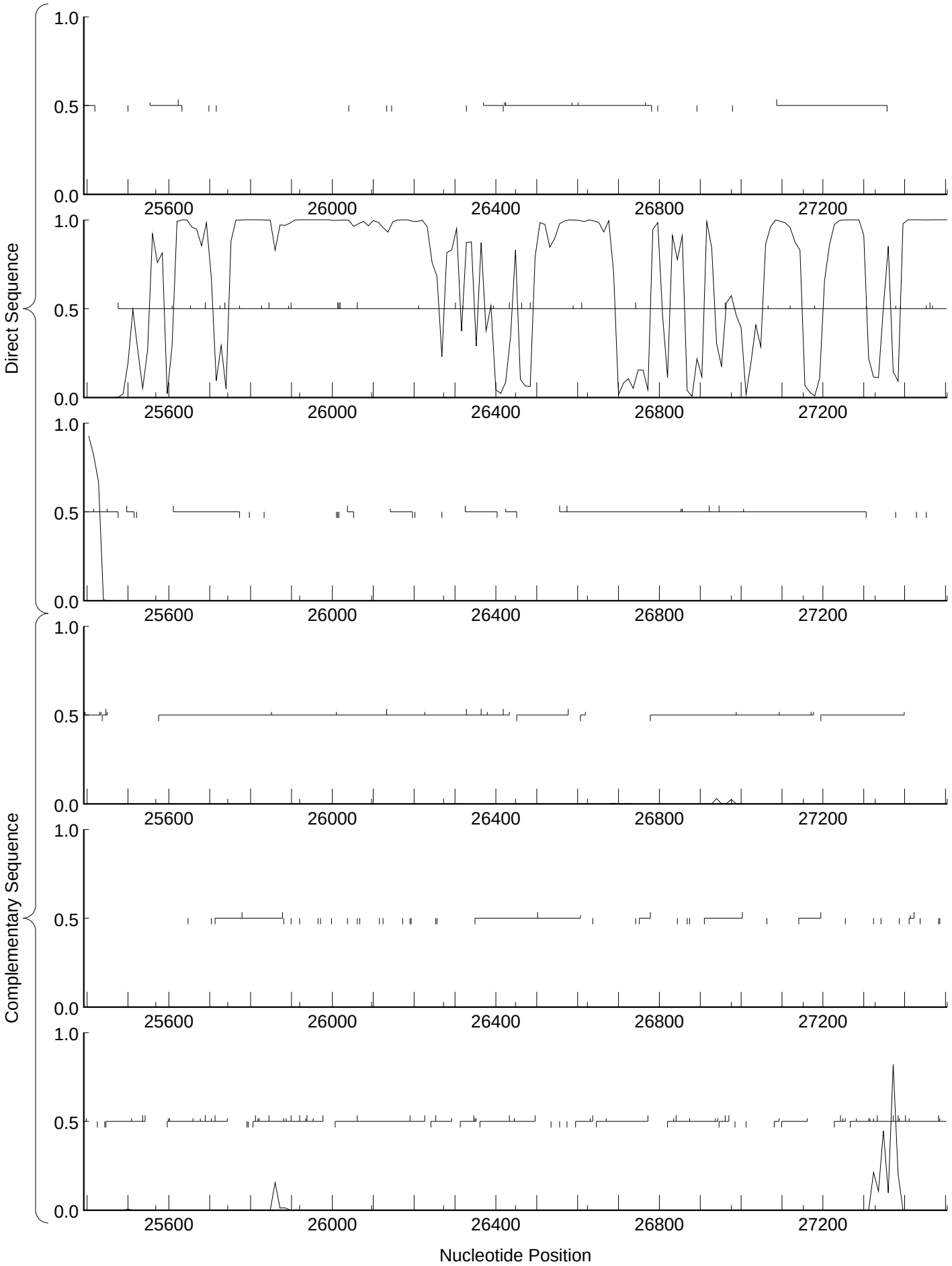
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 12/37



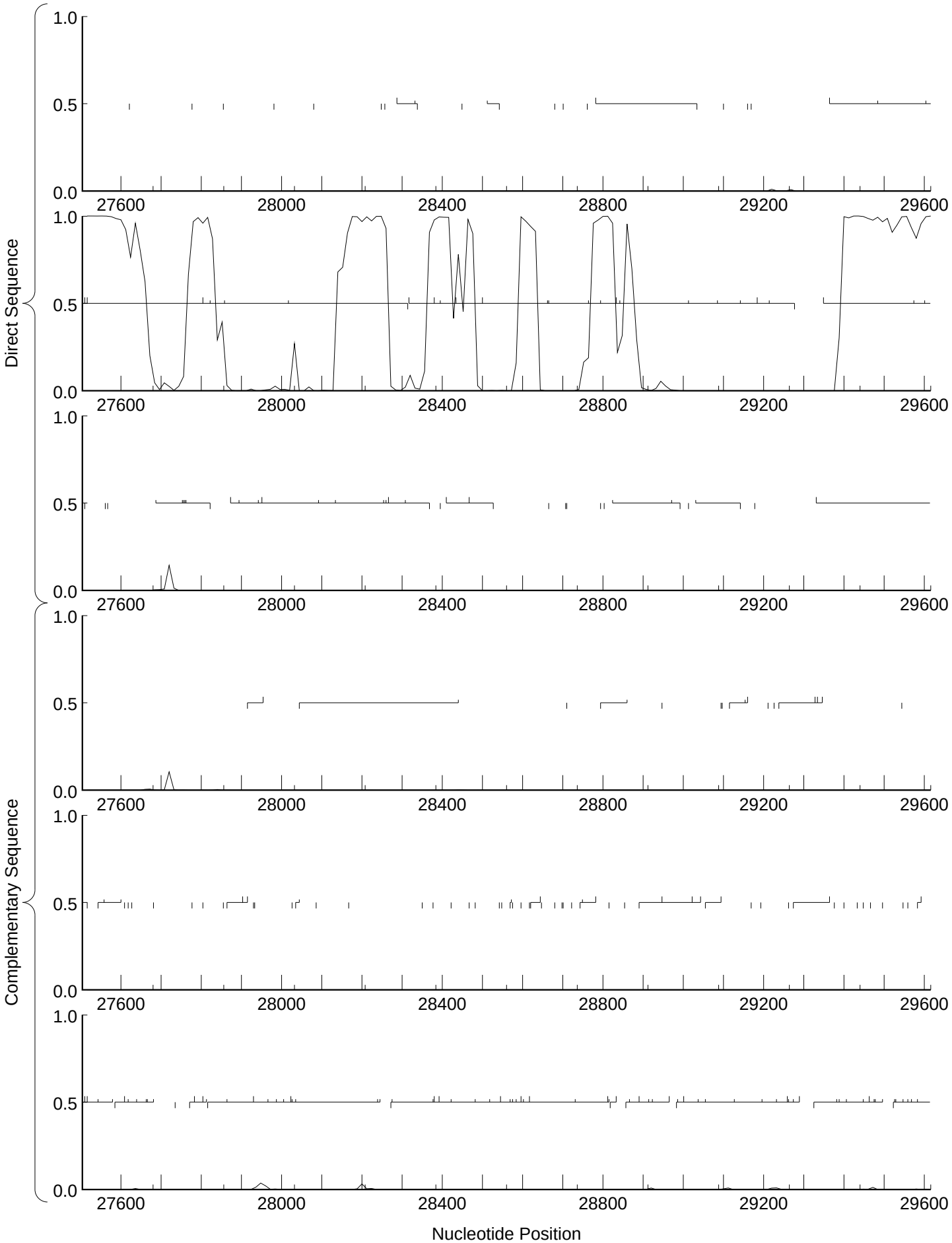
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 13/37



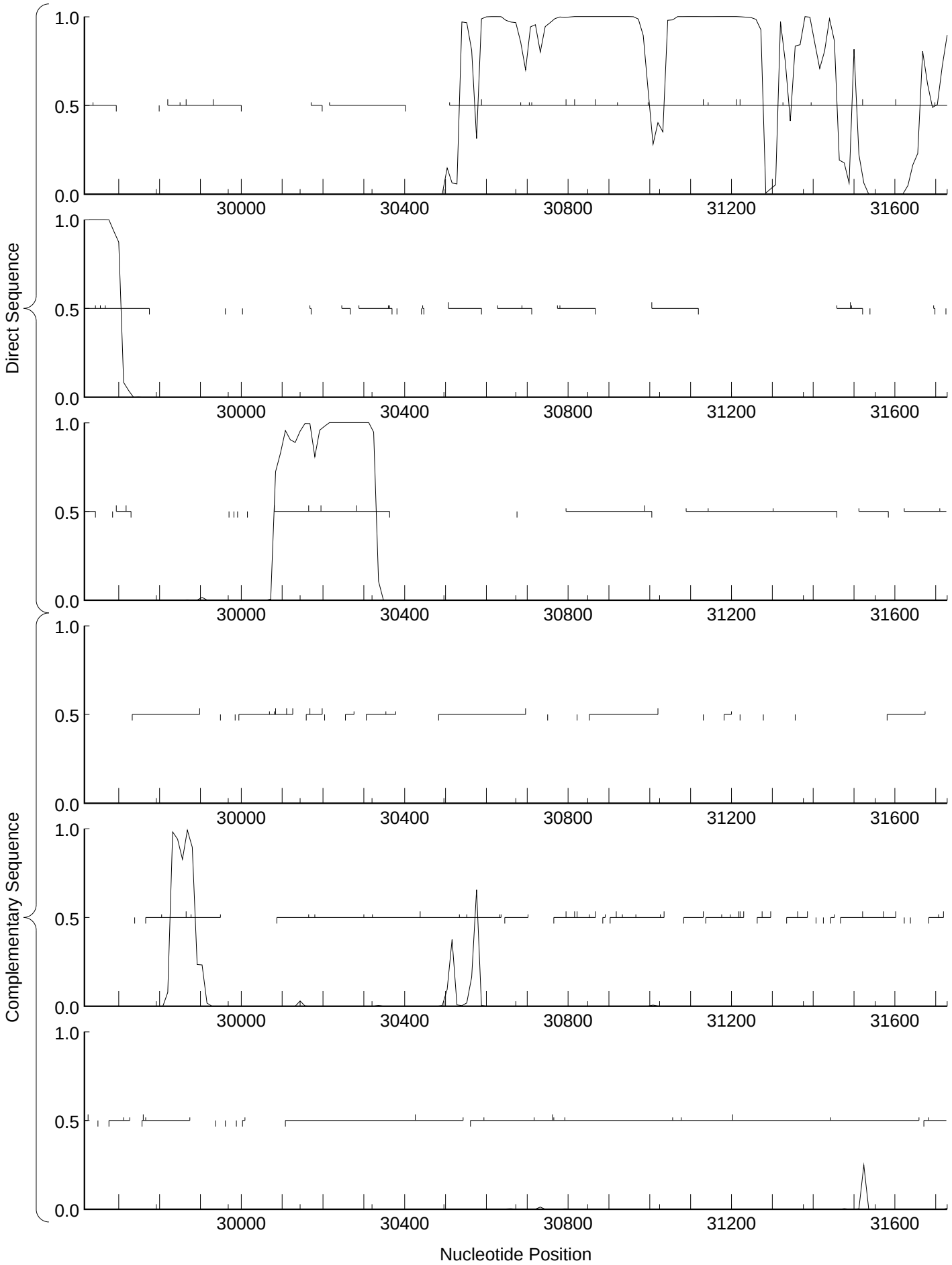
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 14/37



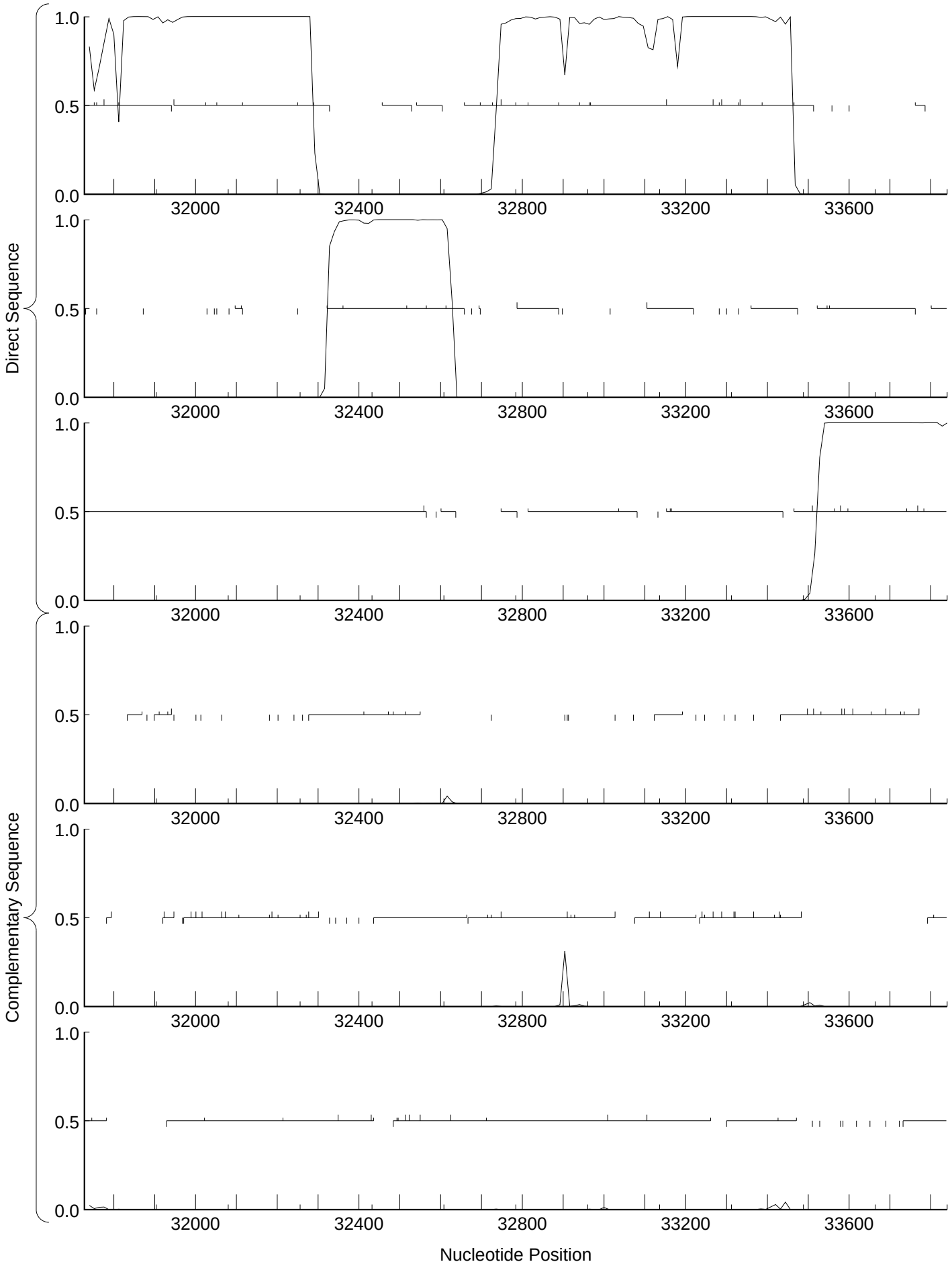
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 15/37



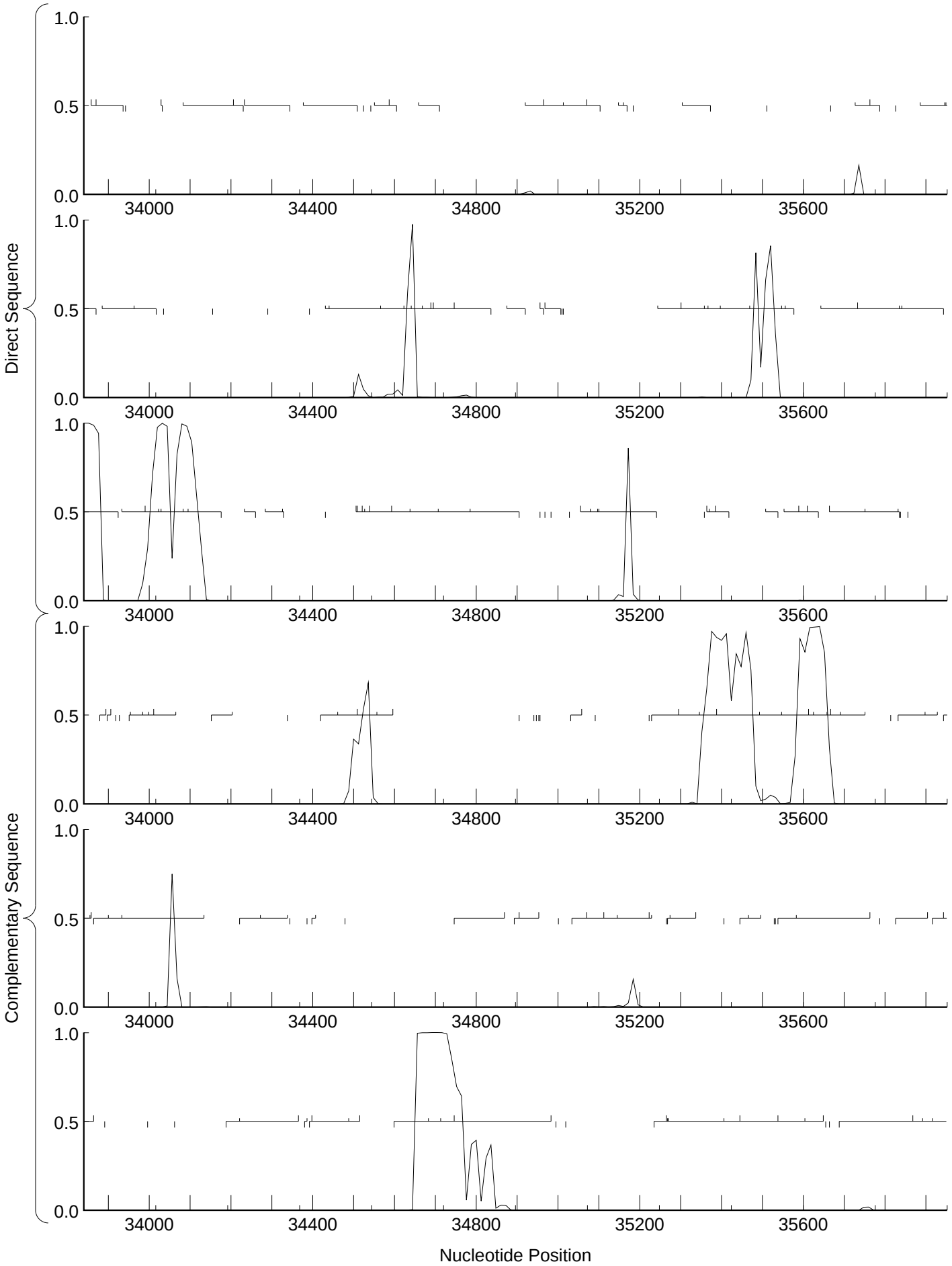
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 16/37



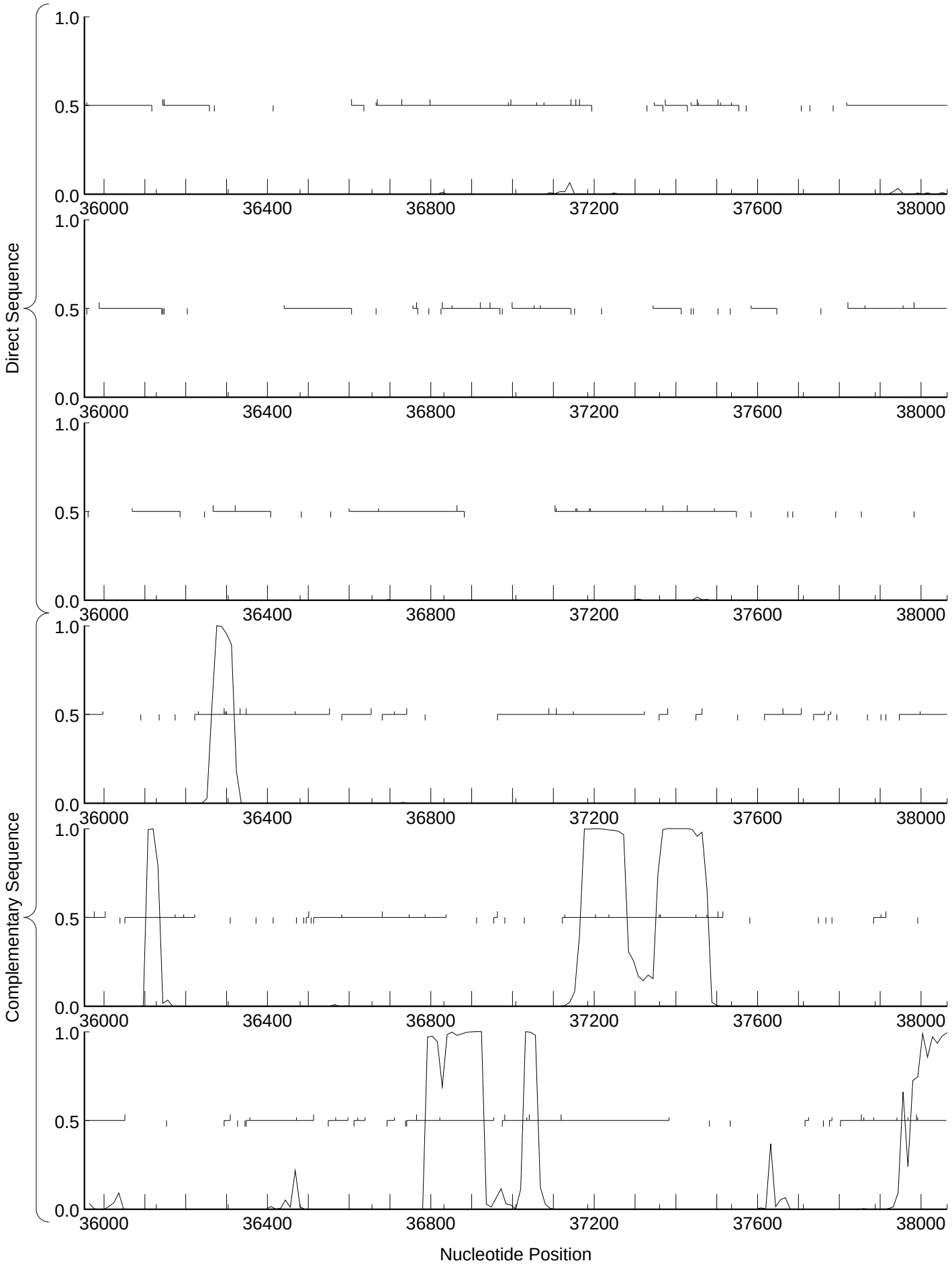
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 17/37



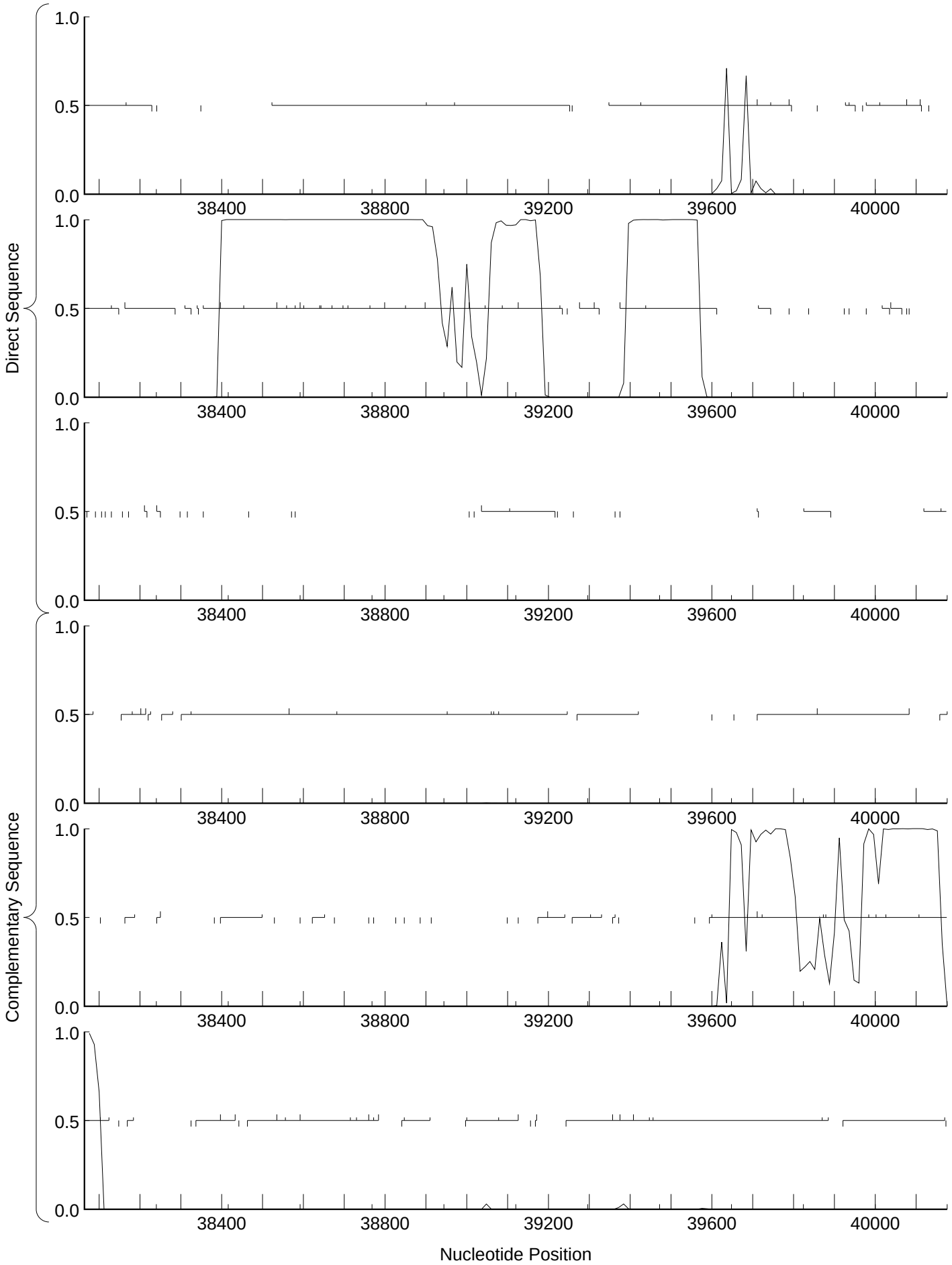
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 18/37



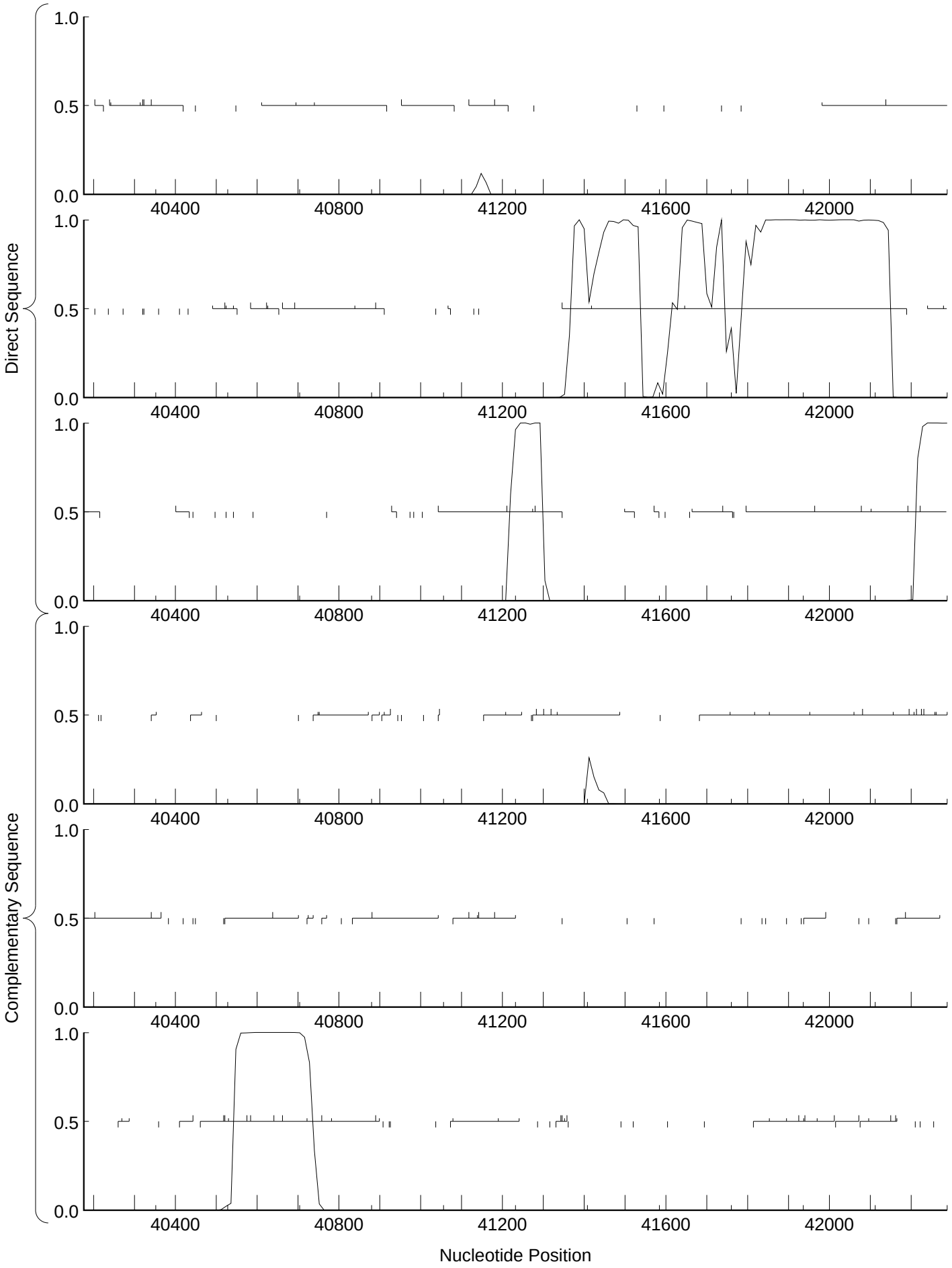
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 19/37



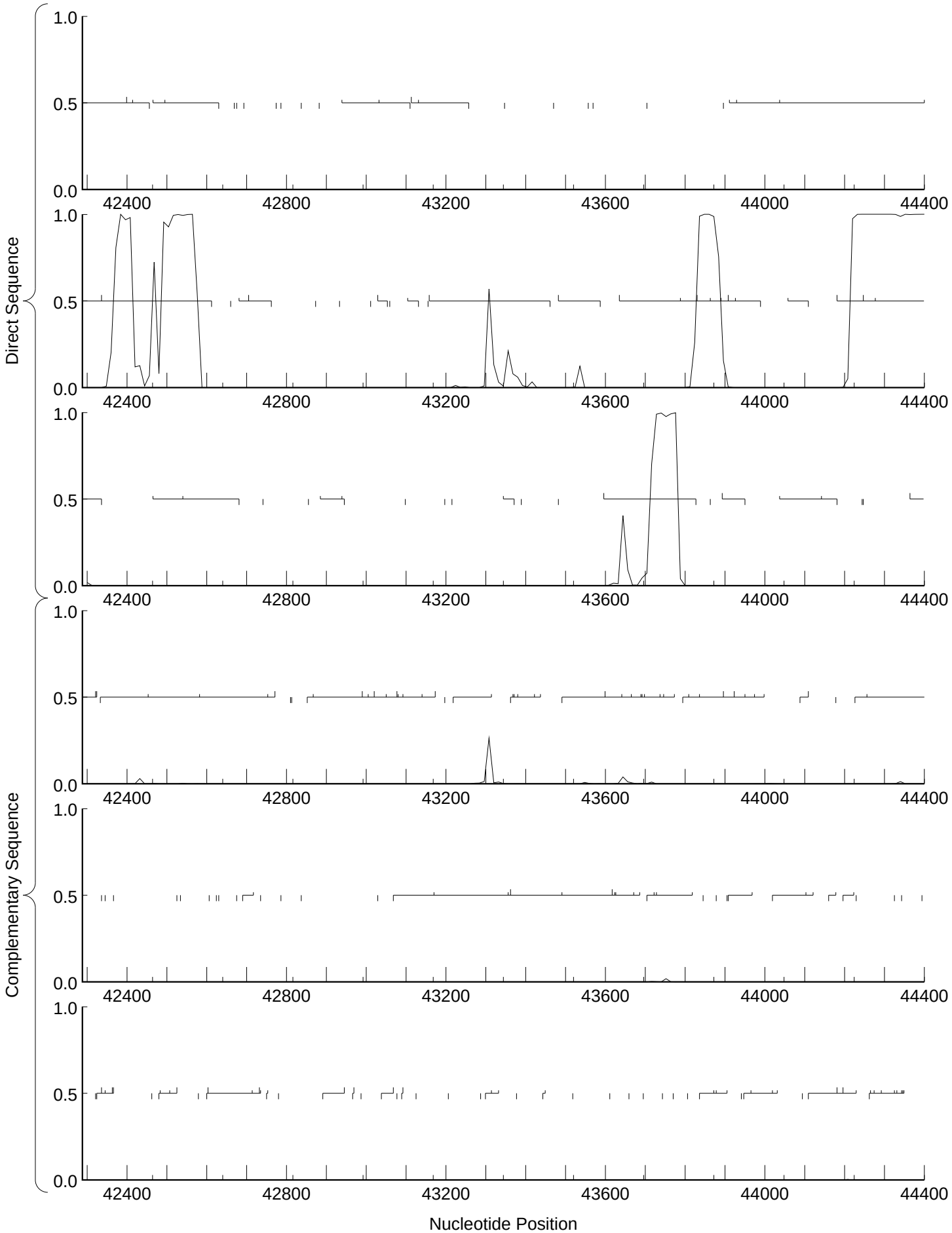
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 20/37



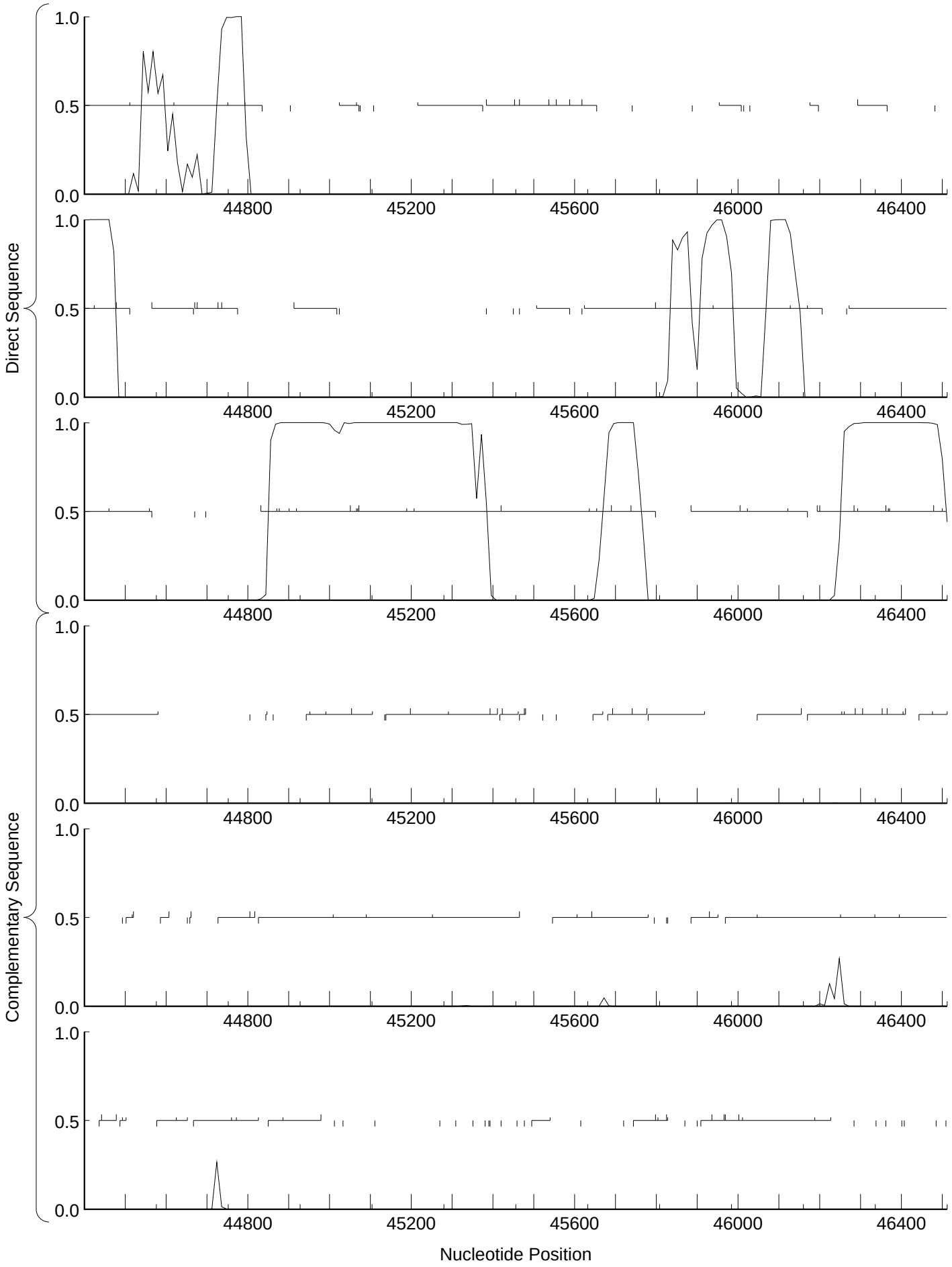
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 21/37



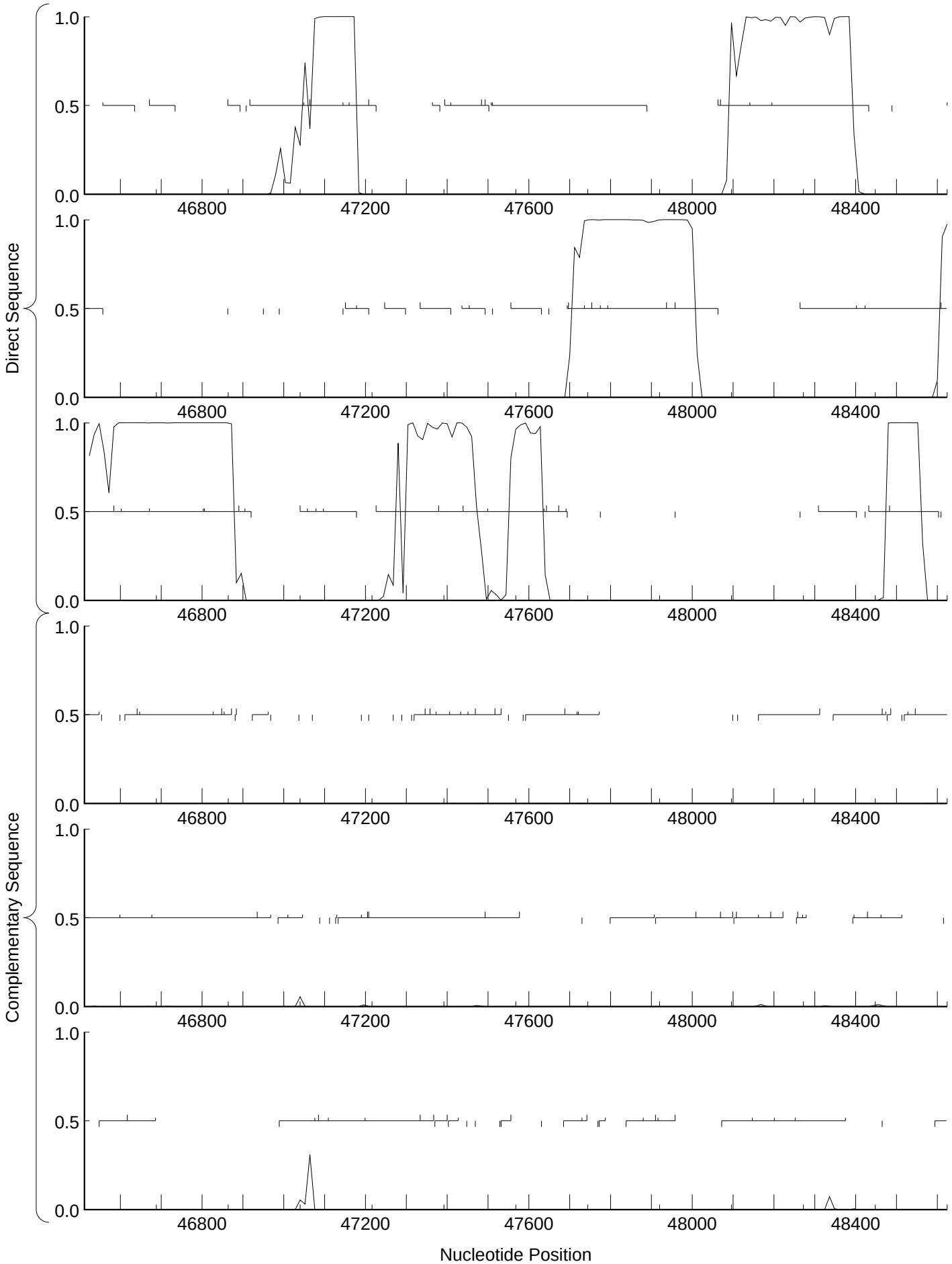
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 22/37



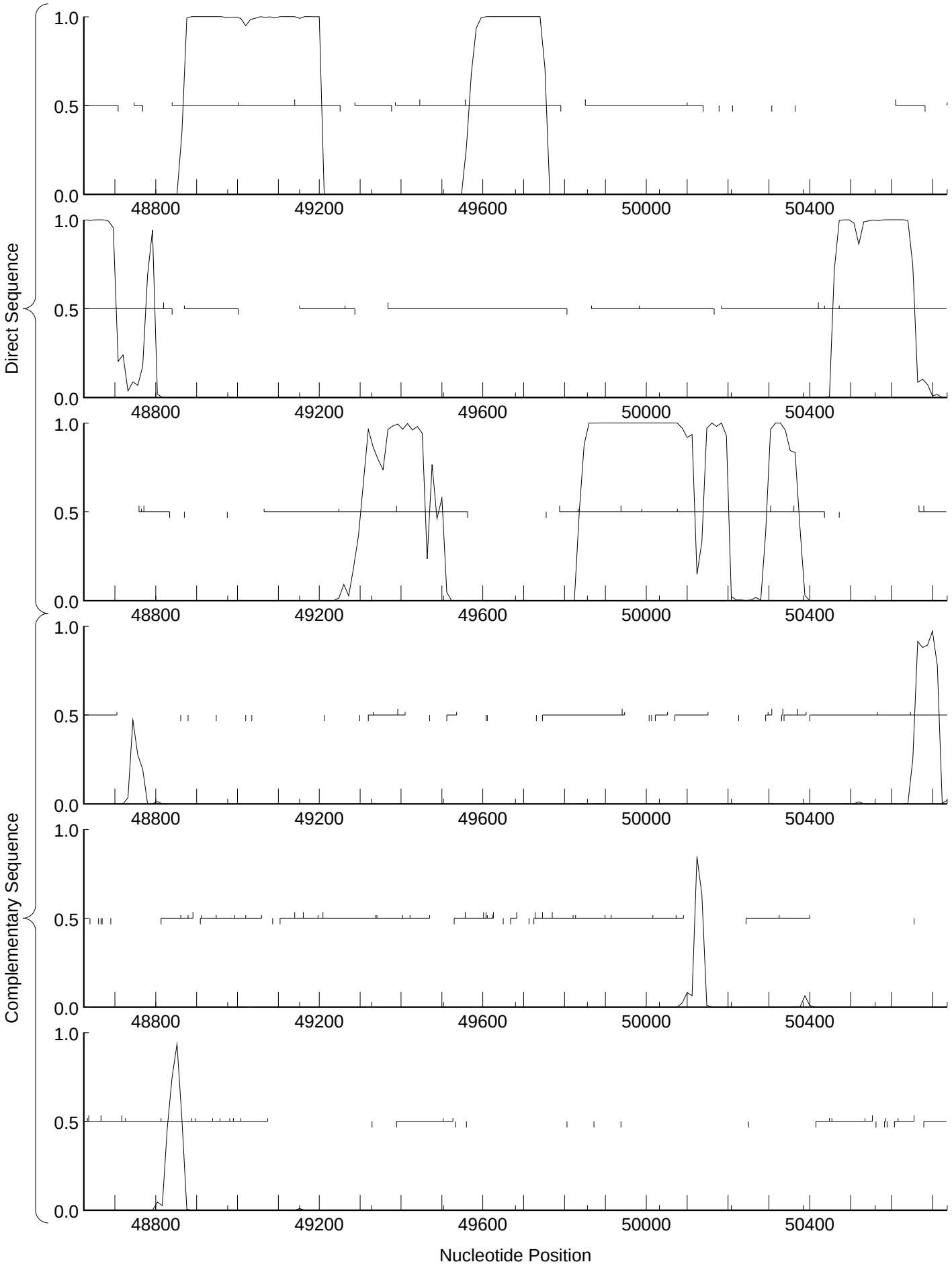
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 23/37



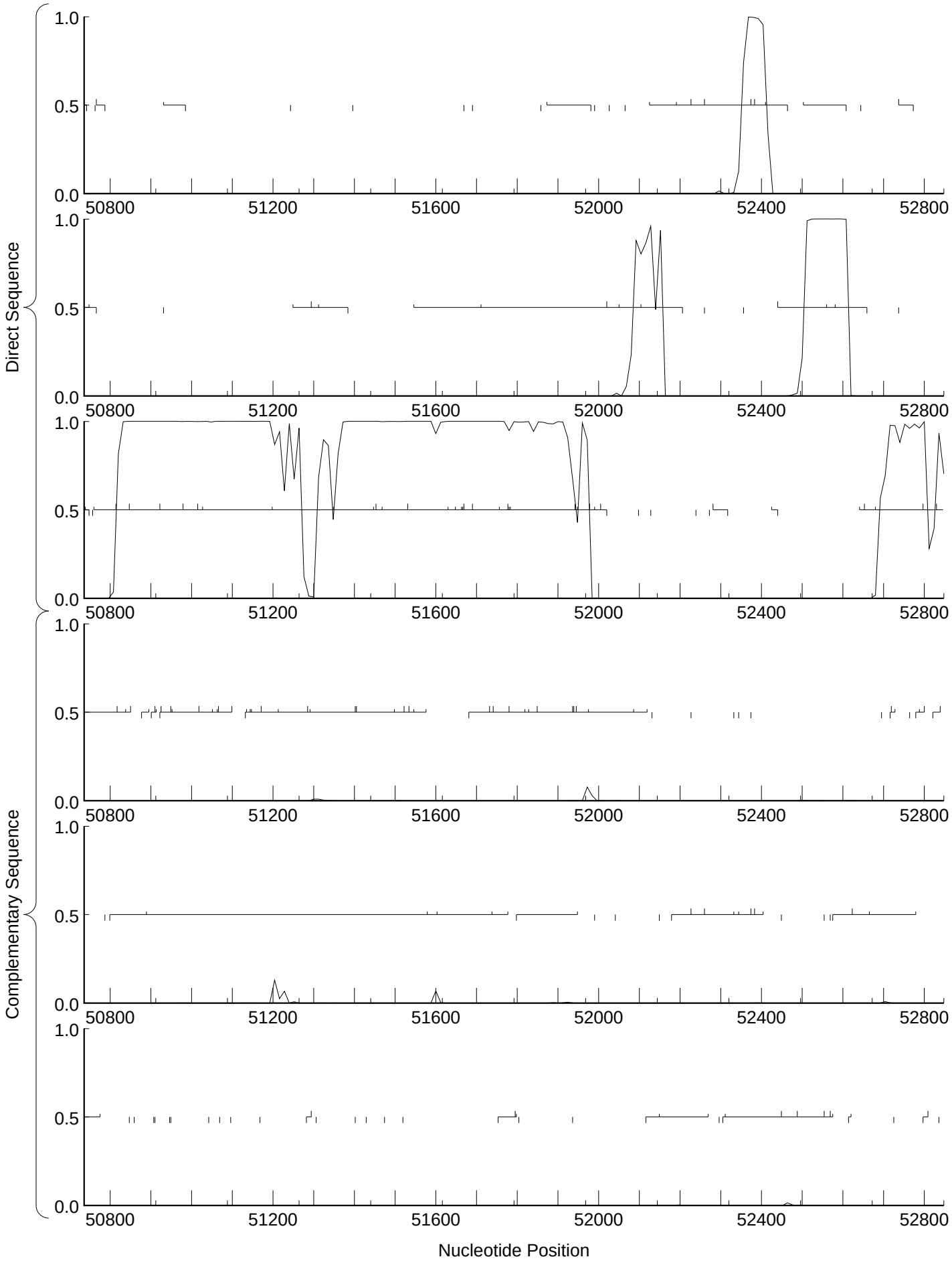
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 24/37



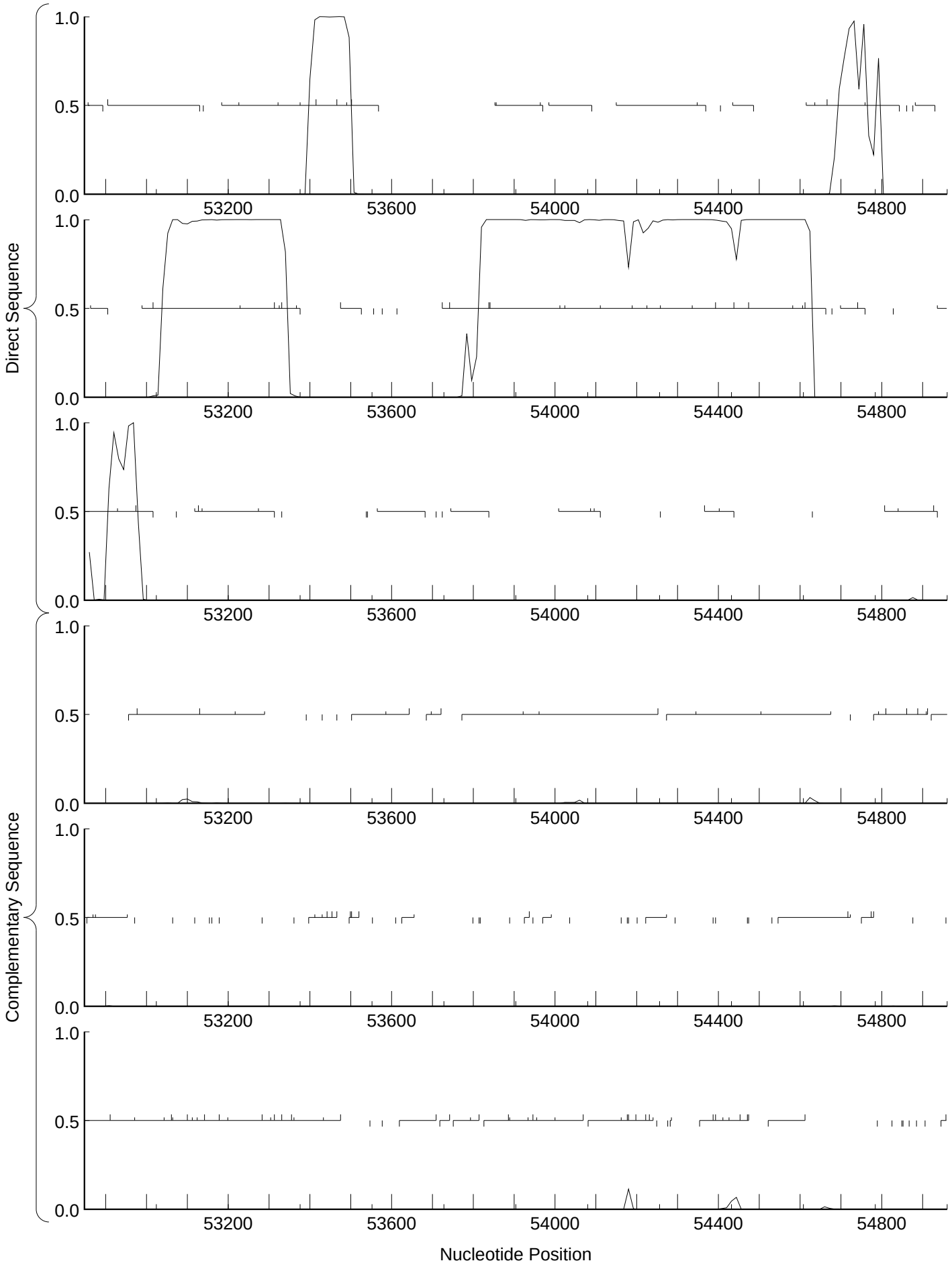
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 25/37



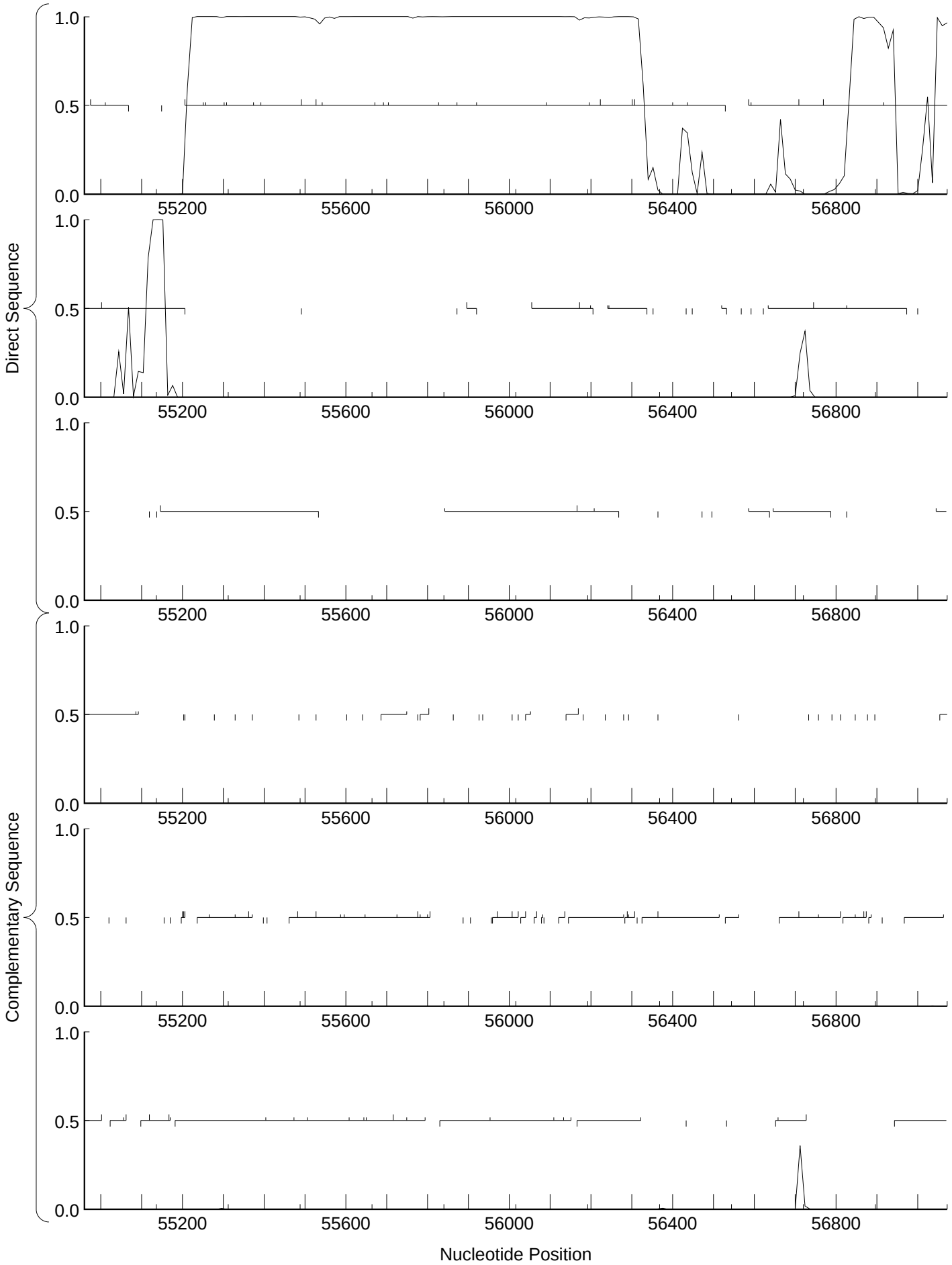
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 26/37



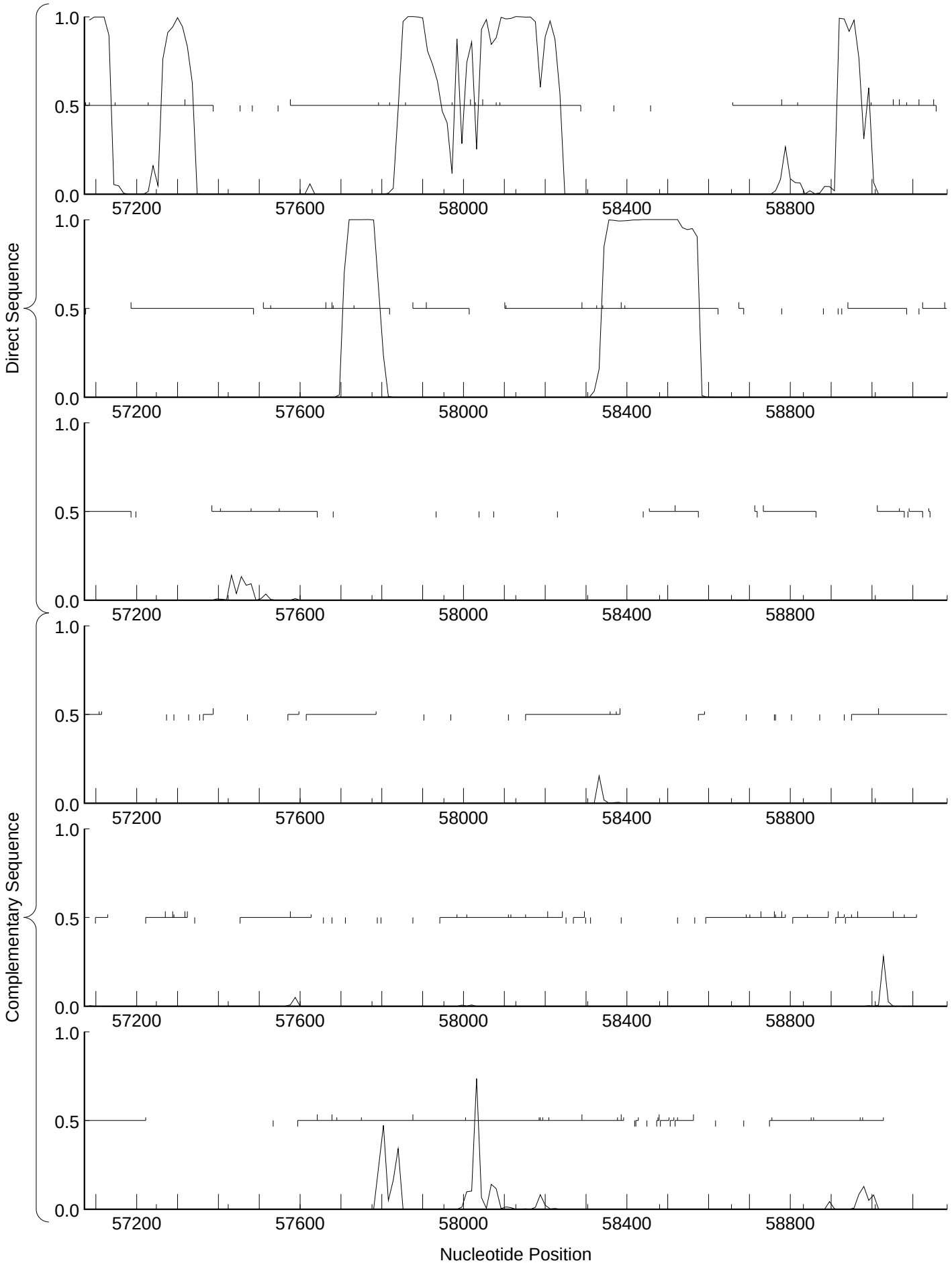
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 27/37



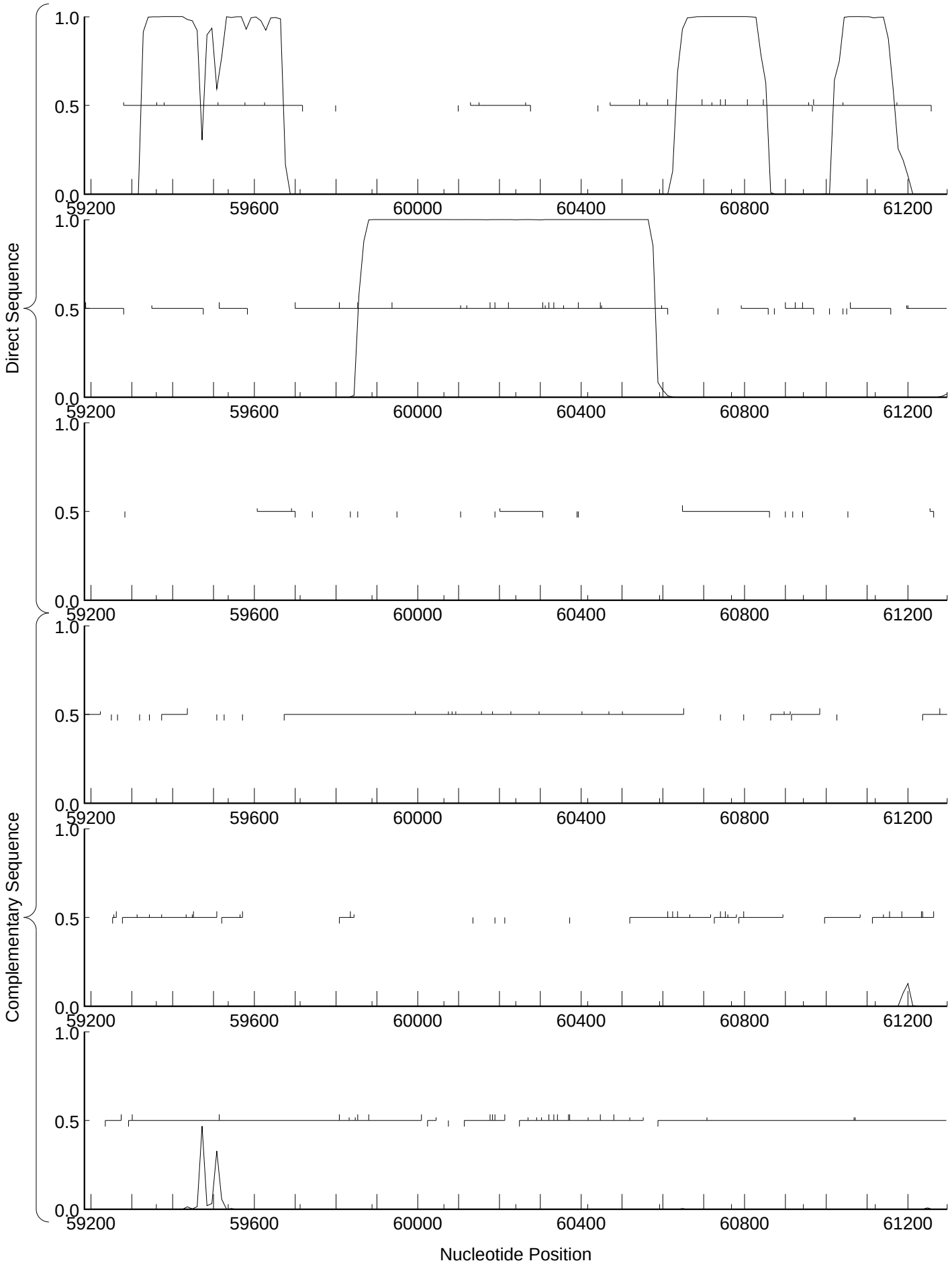
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 28/37



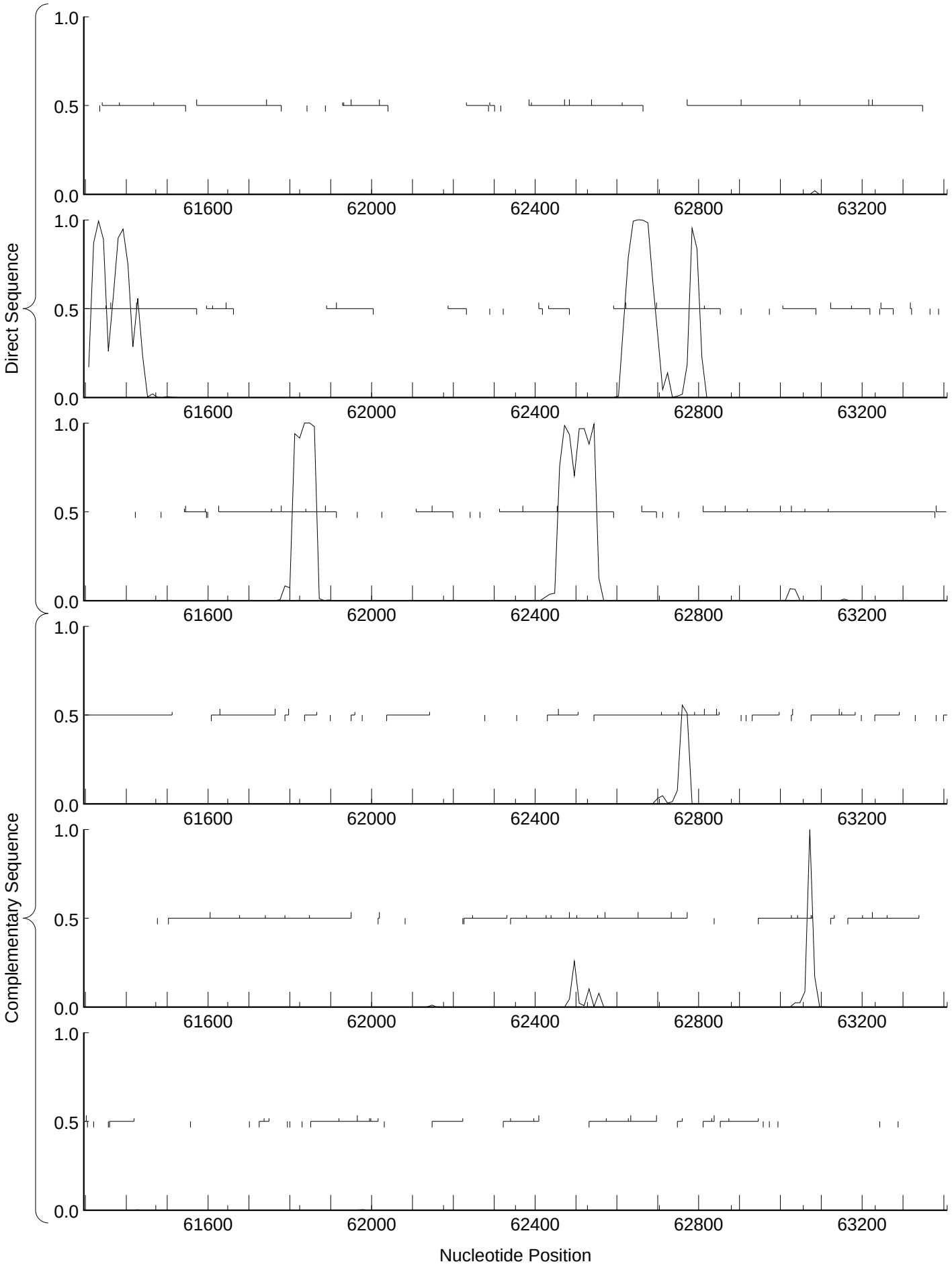
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 29/37



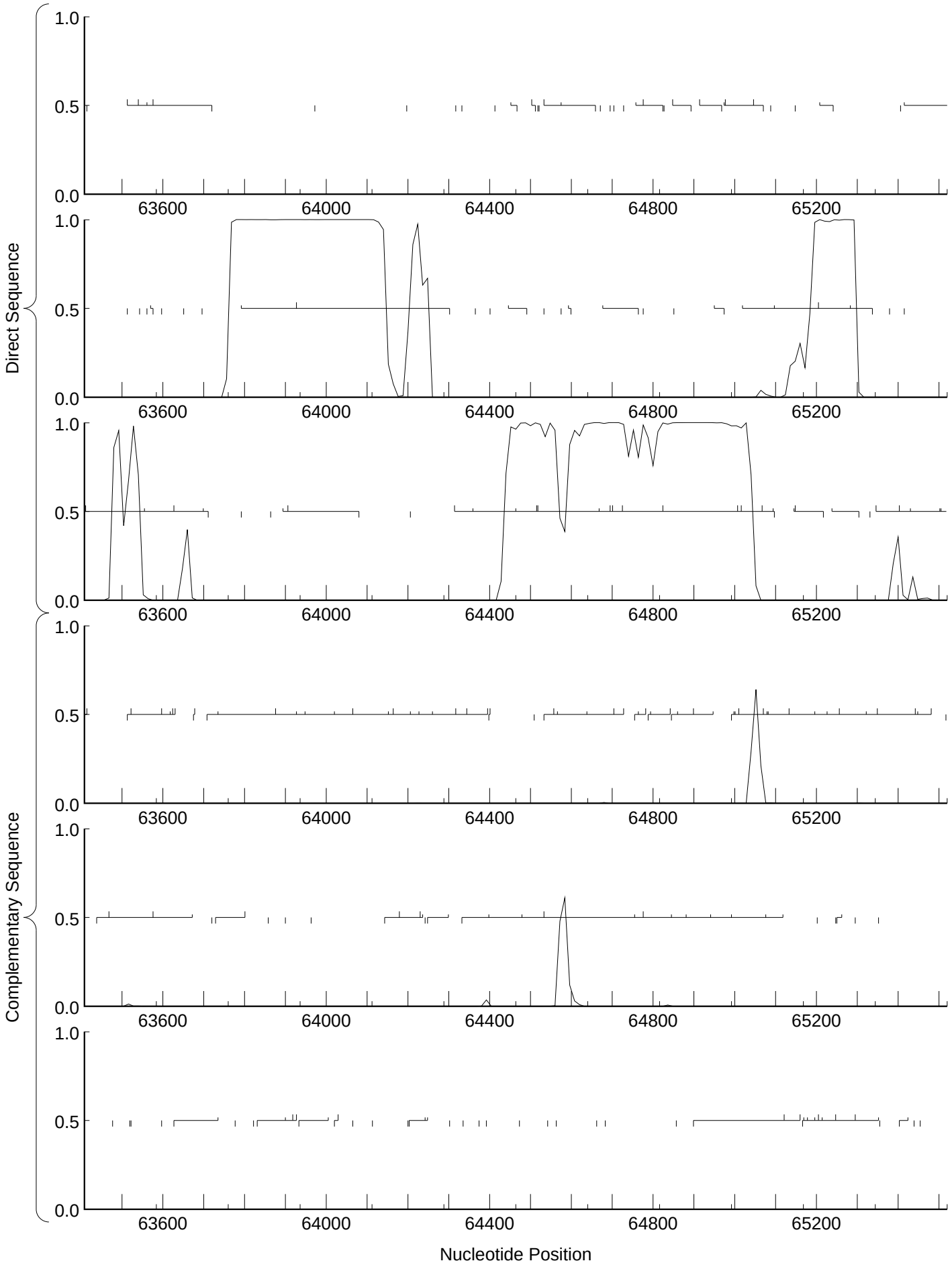
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 30/37



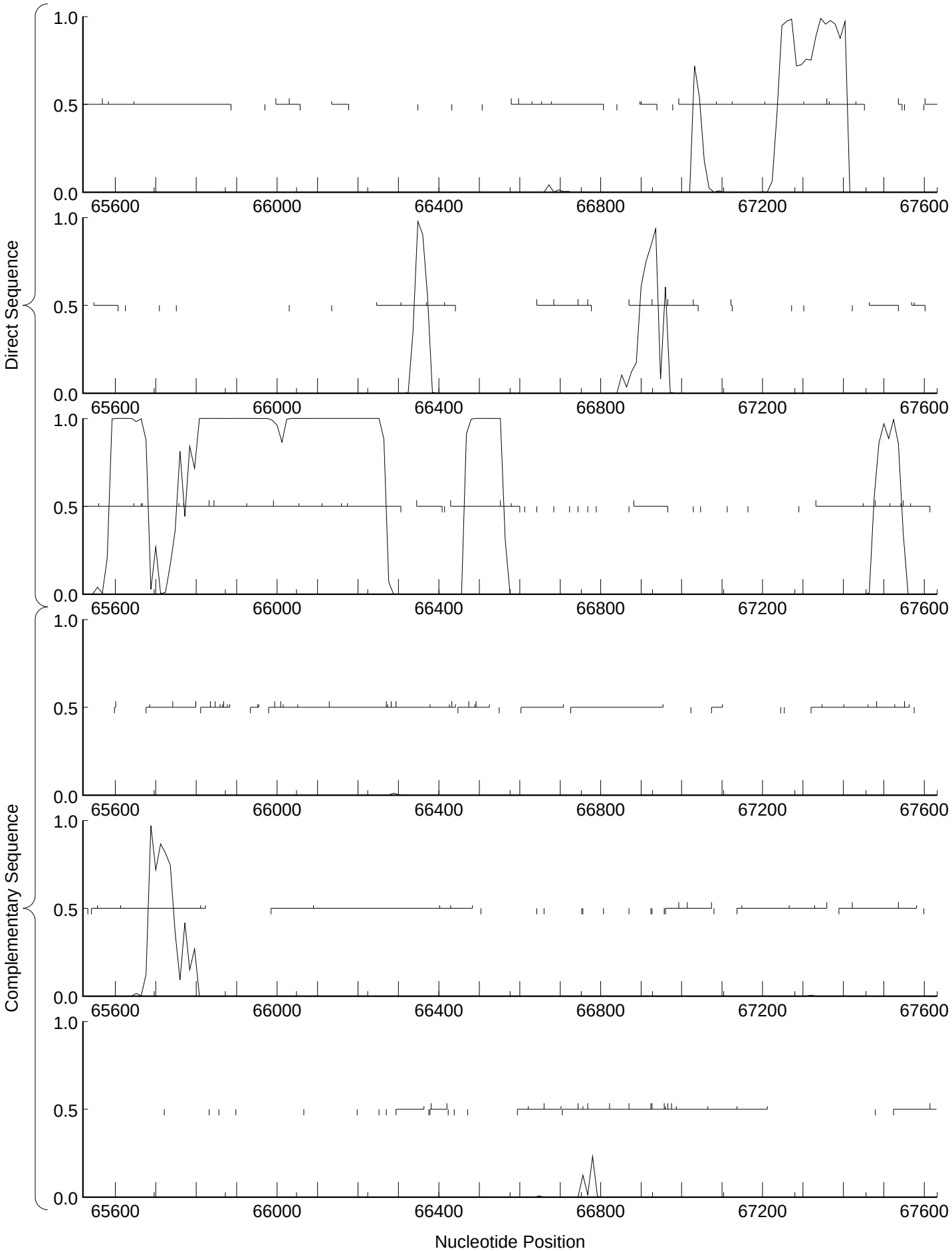
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 31/37



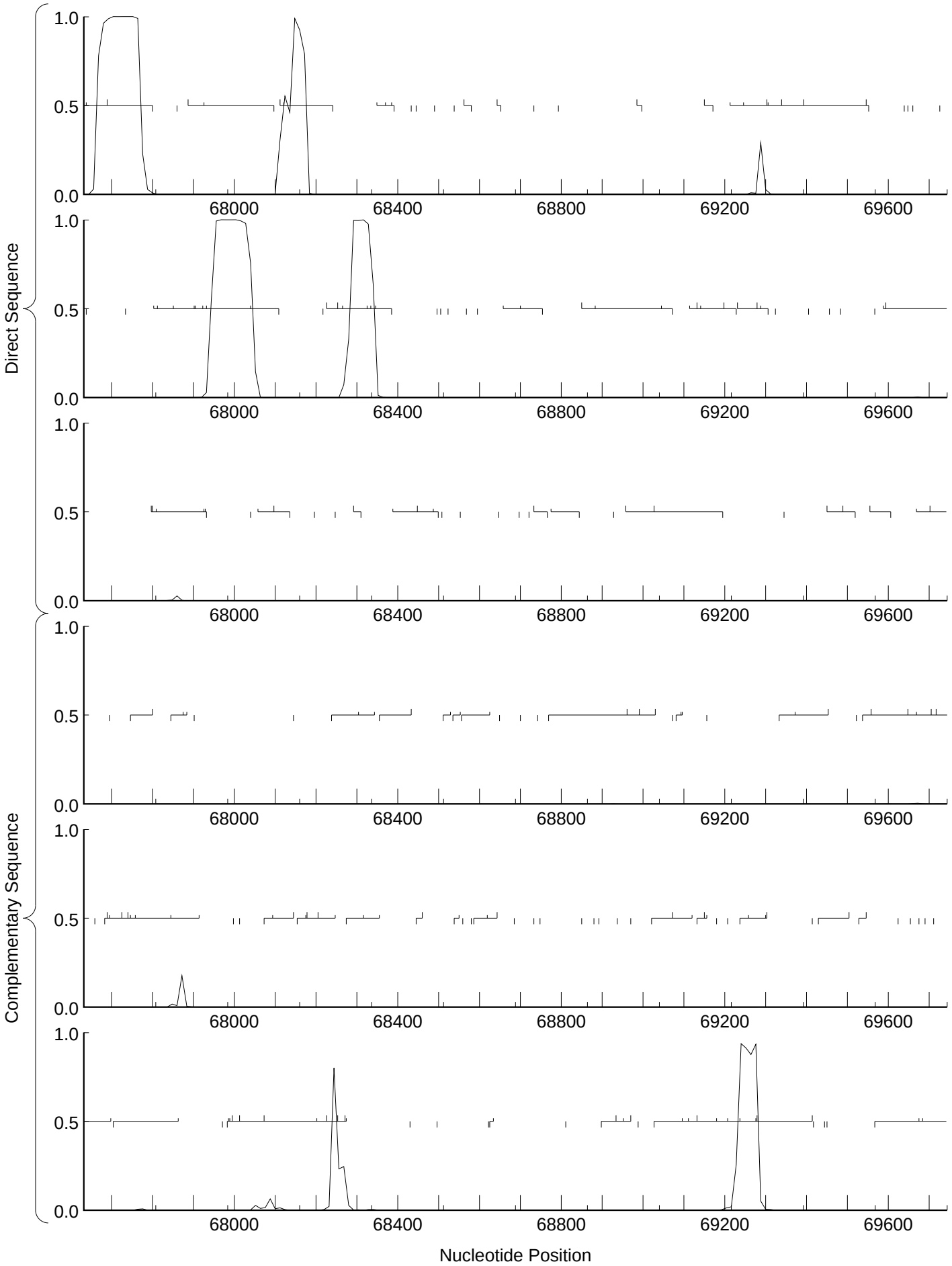
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 32/37



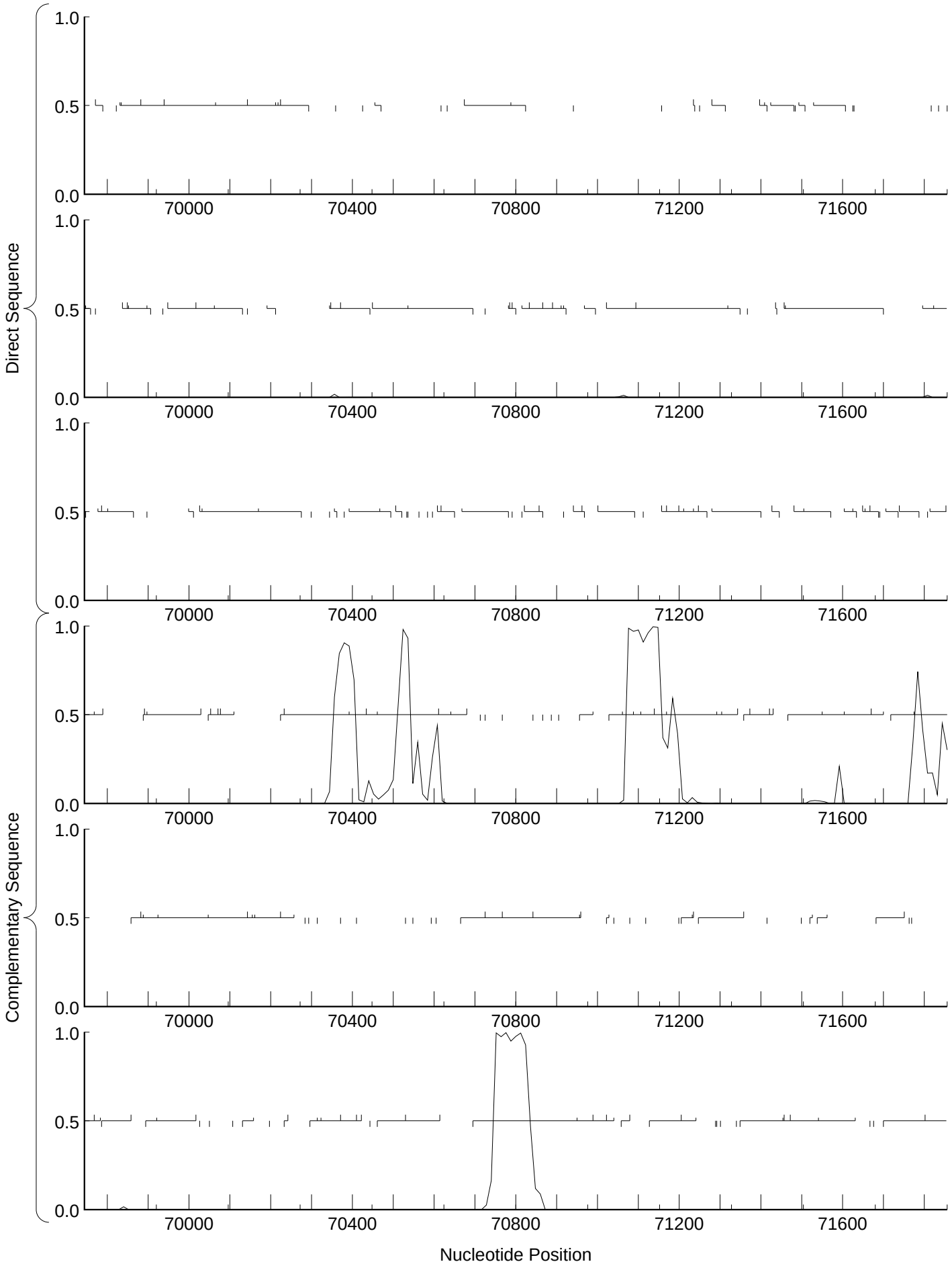
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 33/37



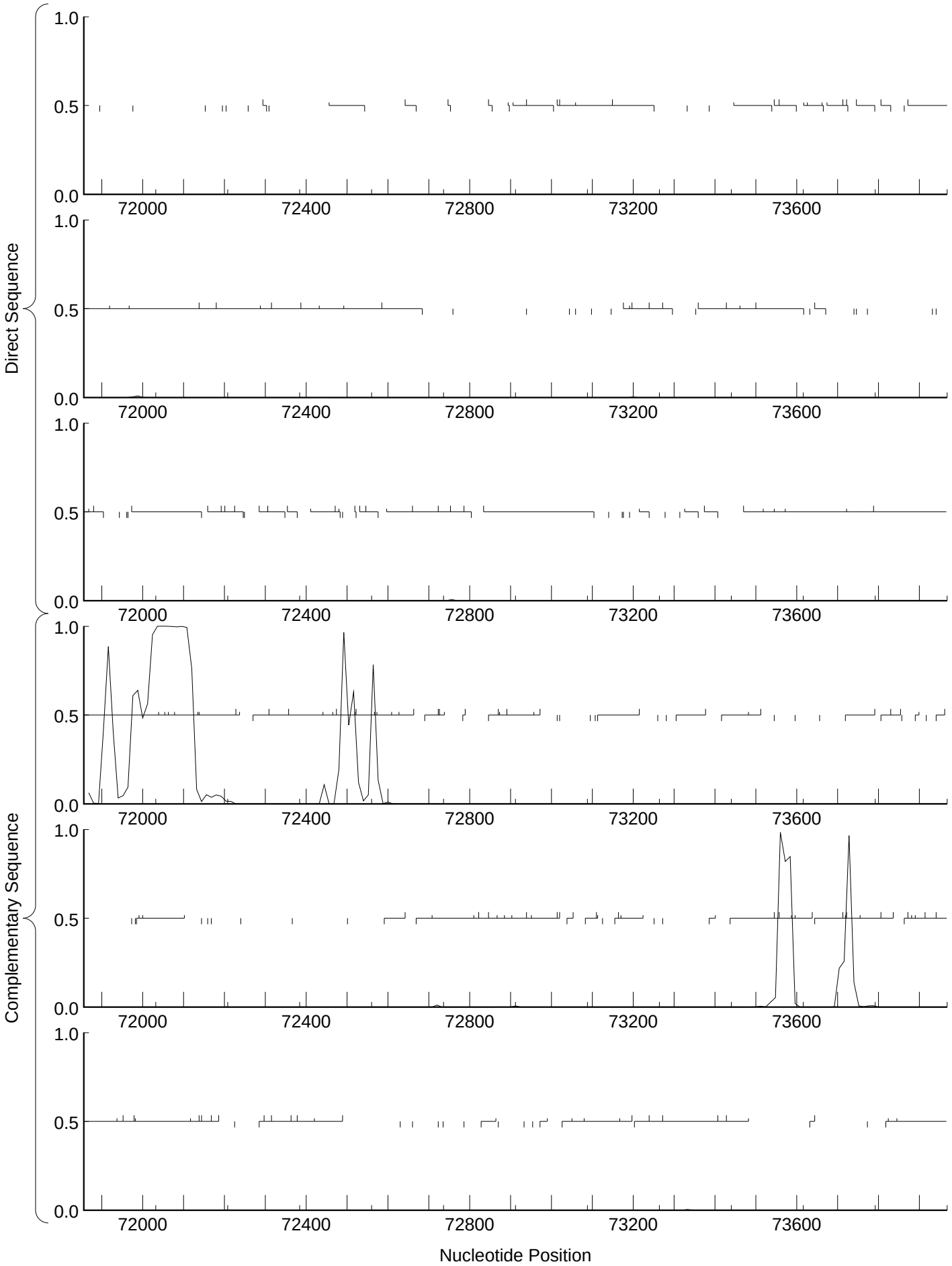
aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 34/37



aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 35/37



aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 36/37



aime complete sequence, 75099 bp including 9-base 3’ overhang (CGCTTGTCA), Cluster E, Order 4, Window 96, Step 12, 37/37