

Genome Annotation Submission Cover Sheet

Preliminary Annotation Review Checklist 5-15-2018

Phage Name:

Your Name:

Your Institution:

Your email:

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 Pinkcreek. 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, TopCons, and SOSUI. Twenty-three tRNAs and 1 tmRNA were verified using online ARAGORN v.2.0 and online tRNAscan-SE. Full annotation was exported from PECAAN and re-BLASTED. A copy of the Pinkcreek *M. smegmatis* GeneMark file is attached as well as a copy of the Pinkcreek Self GeneMark file. Thanks for reviewing this!

QUESTIONS ABOUT FUNCTIONAL ASSIGNMENTS/ADDED GENES/START SITES: (DNA Master gene product numbers shown)

gp3 Forward 477 - 2348

Most called this nucleotidyl transferase based on BLAST hits, Phages Function Frequency, and the CDD predictions, but HHPRED predicts homing endonuclease. Should this be called nucleotidyl transferase?

gp54 Forward 19855 - 20118

Added this sequence: Product was detectable on alternative GeneMark files as seen in C1 cluster phage Koguma.

gp95 Forward 55444 - 55644

This sequence has been added. Same sequence is seen in C1 cluster phage Koguma.

gp120 Forward 73636 - 76482

This function was assigned as minor tail protein due to the gene being located downstream of tape measure protein and the fact that HHpred reported hydrolase activity which commonly occurs in some minor tail proteins.

gp149 Forward 94190 - 94390

This gene prediction was added because it was detected on GeneMark TB and Self files and this sequence is also reported in C1 cluster members Cane17, Ronan, etc.

gp166 Forward 101635 - 101724

Did not omit for gene product for being too small (90 bp) because it has a -4 gap (part of an operon) and was included in other C1 phages genomes.

gp213 Forward 132523 - 132627

Product is small (105bp), but was left in the genome because it has a -4 gap and other C1 phage genomes contain this.

gp220 Forward 141243 - 141653

Likely holin homologue, based on location, transmembrane domains, and weak matches in HHpred/BLAST. Called membrane protein to be conservative.

GENES THAT COULD BE CALLED MEMBRANE PROTEINS BECAUSE TMHMM, TOPCONS, AND SOSUI ALL PREDICTED PRESENCE OF 1 OR MORE TRANSMEMBRANE DOMAINS (SIGNAL PEPTIDE PREDICTIONS NOT INCLUDED).

gp2 Forward 252 – 473

One transmembrane domain detected.

gp5 Forward 2558 - 2764

Two transmembrane domains detected.

gp13 Forward 5502 - 5729

One transmembrane domain detected.

gp67 Forward 26334 - 26822

One transmembrane domain detected.

gp76 Forward 36360 - 36515

One transmembrane domain detected.

gp92 Forward 54589 – 54960

Two transmembrane domains detected

gp127 Forward 81382 - 81525

One transmembrane domain detected.

gp202 Forward 126,001 – 126,354

Two transmembrane domains detected.

gp223 Forward 145777 - 145905

One transmembrane domain detected by all three.

FRAMESHIFT INFORMATION FOR CHAPERONE PROTEINS:

gp114 Forward 69017 – 69544 and 69607 – 69915 (tail assembly chaperones)

Ribosomal slippage

-1 frameshift

CDS join(69017..69529;69529..69915)

GeneMark

Version 2.5p (09.08.06)

Copyright 1993 - M. Borodovsky, J. McIninch

PROGRAM INFORMATION

Sequence : Mycobacterium phage Pinkcreek complete sequence, 153184 bp, circularly permuted, Cluster C1
Analysis Date : 7/22/21 at 15:30:32
Pages : 74
Sequence Length : 153184 bp
GC Content : 64.65%

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,

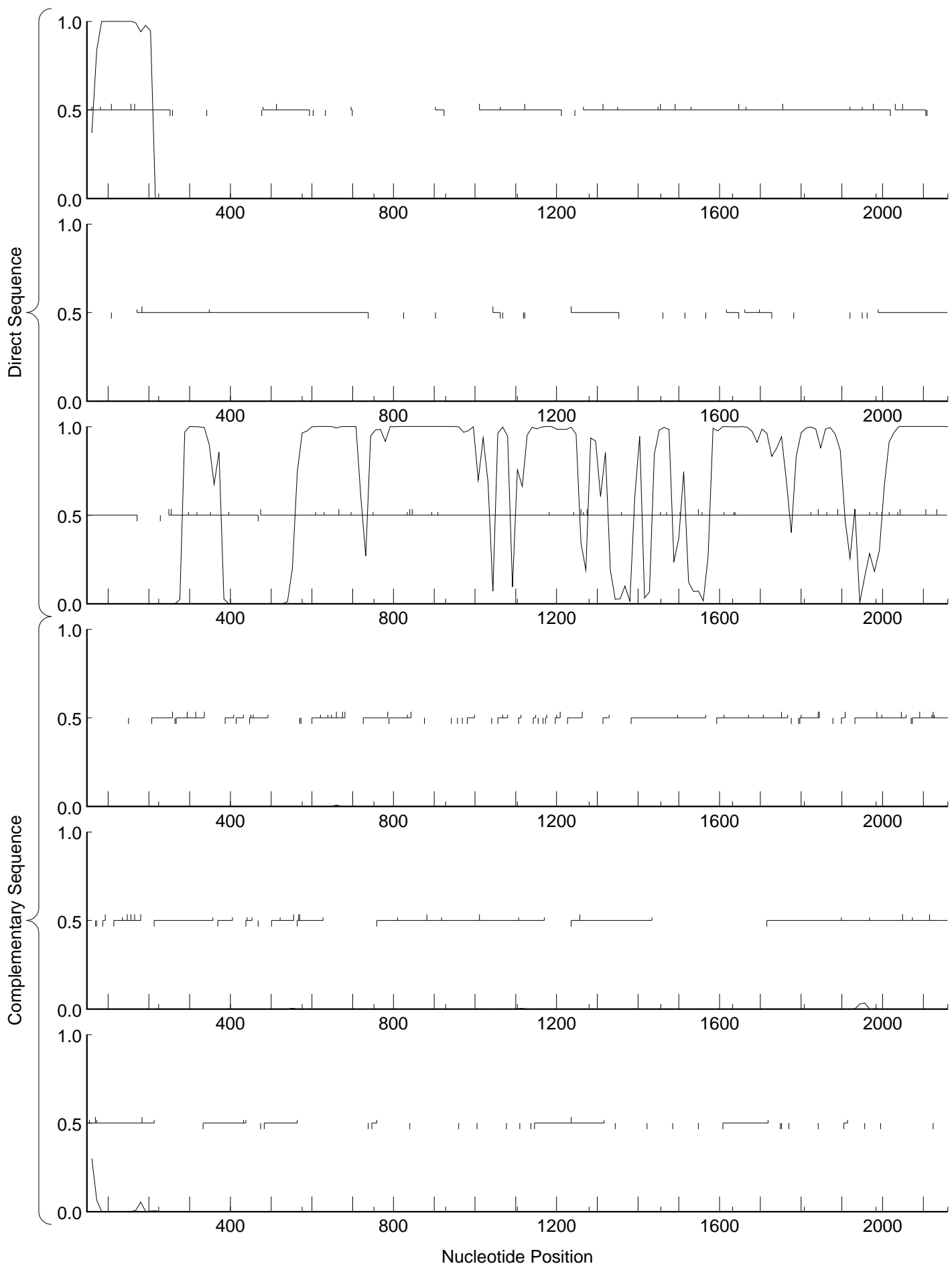
Matrix notes & comments

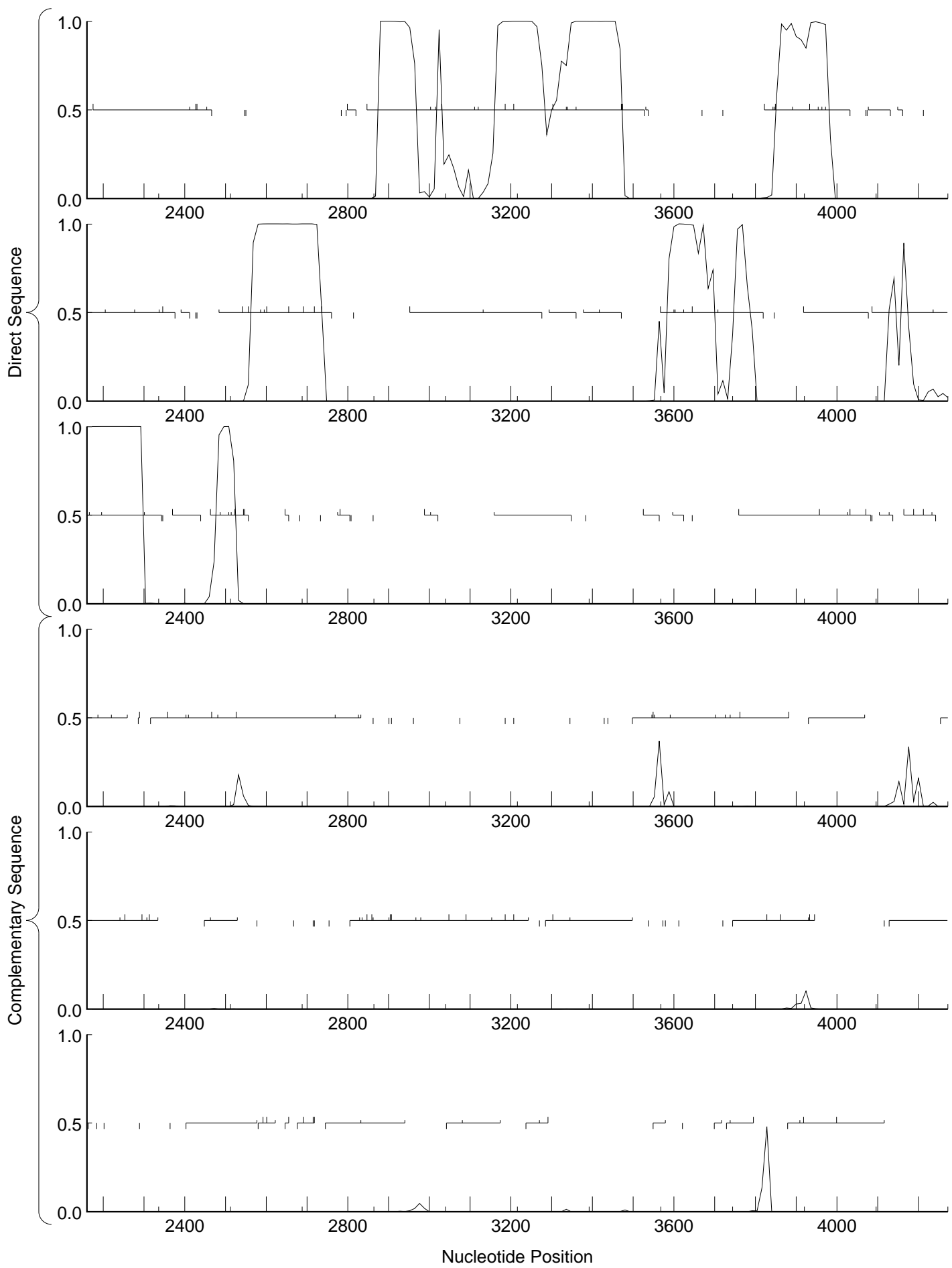
Training set derived by GeneMarkS, 4.27 September 2014
Tue Sep 23 15:23:03 2014

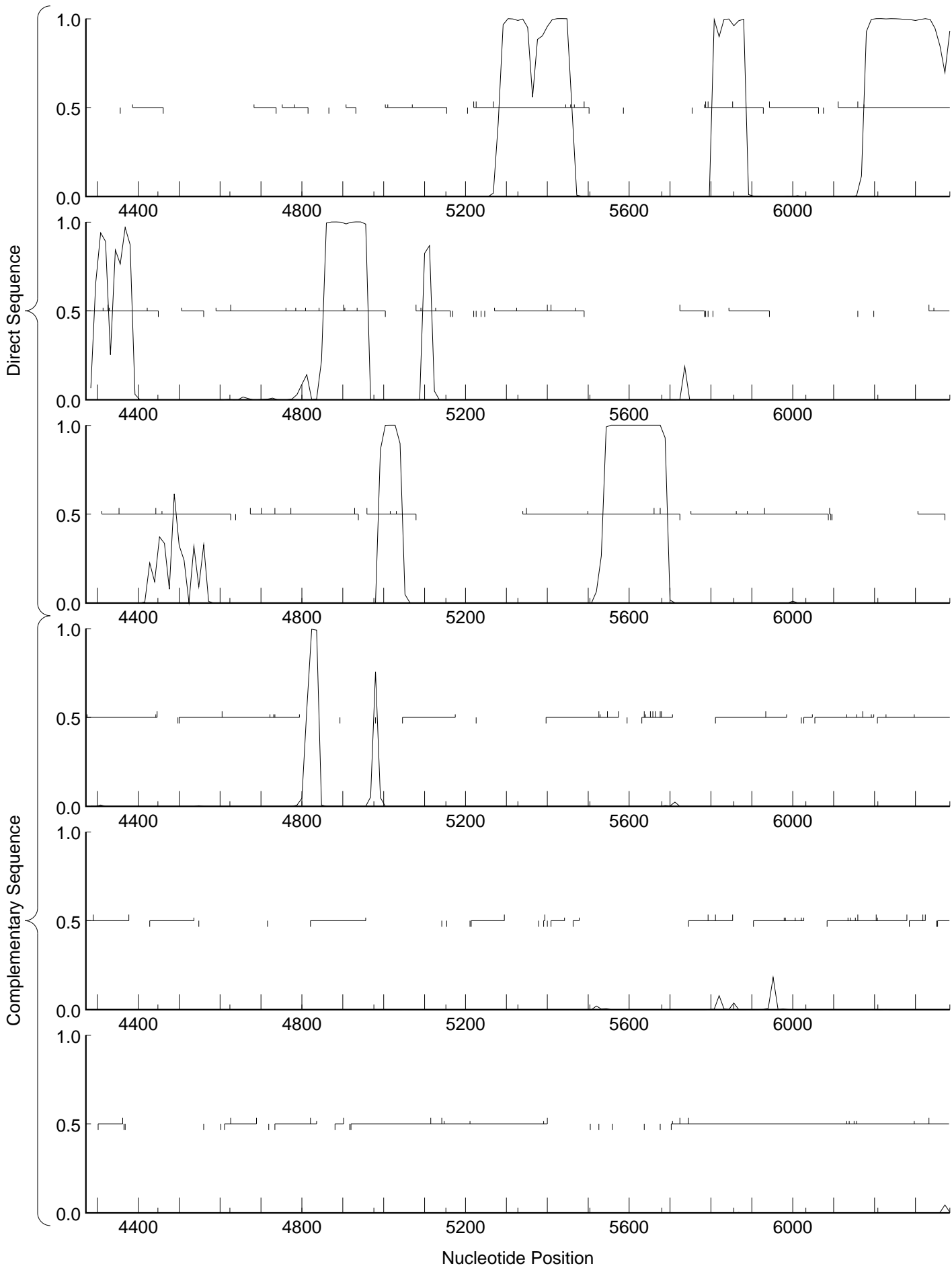
MATRIX INFORMATION

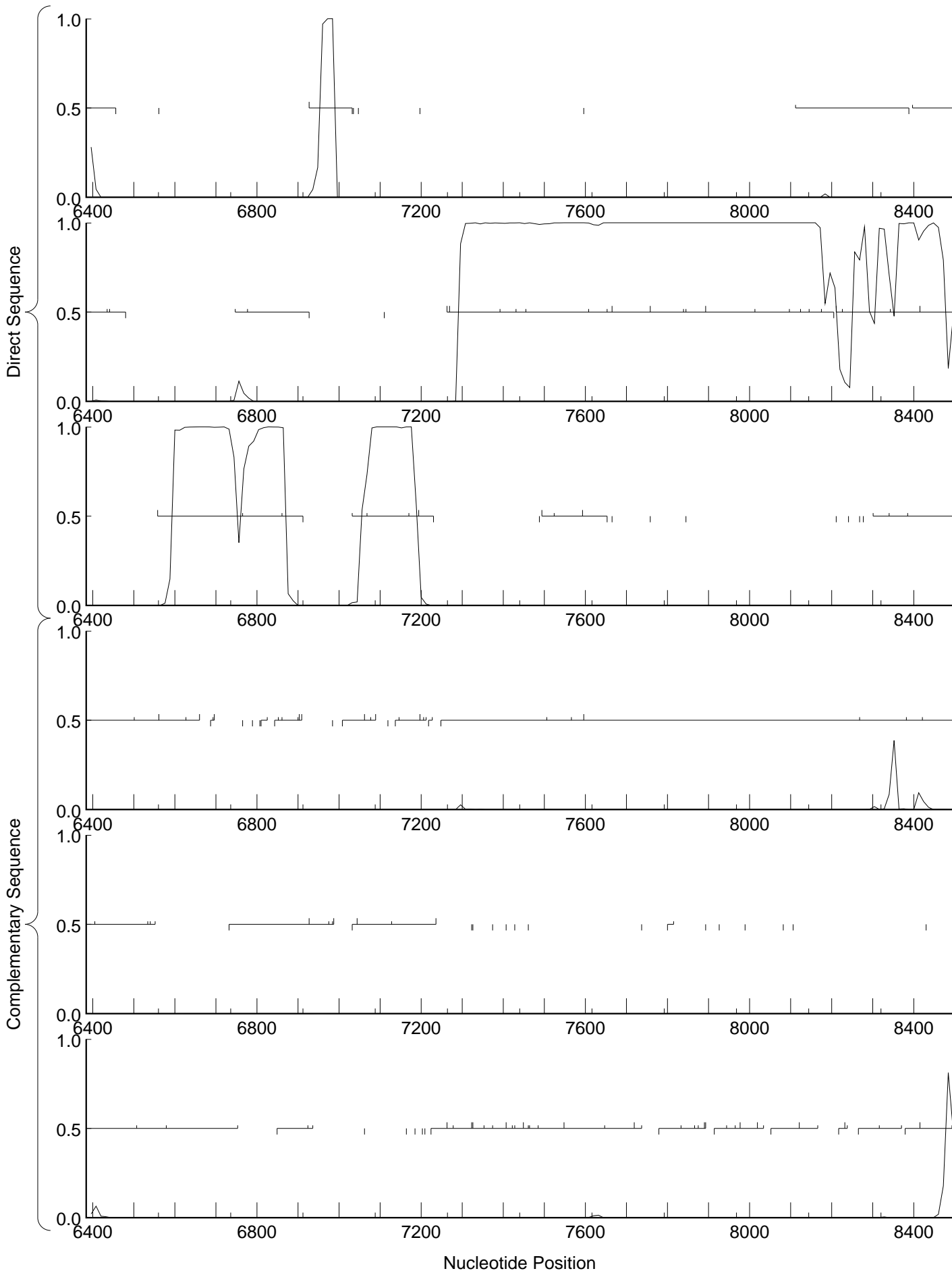
Matrix : Mycobacterium_smegmatis_MC2_155
Author : -
Order : 4

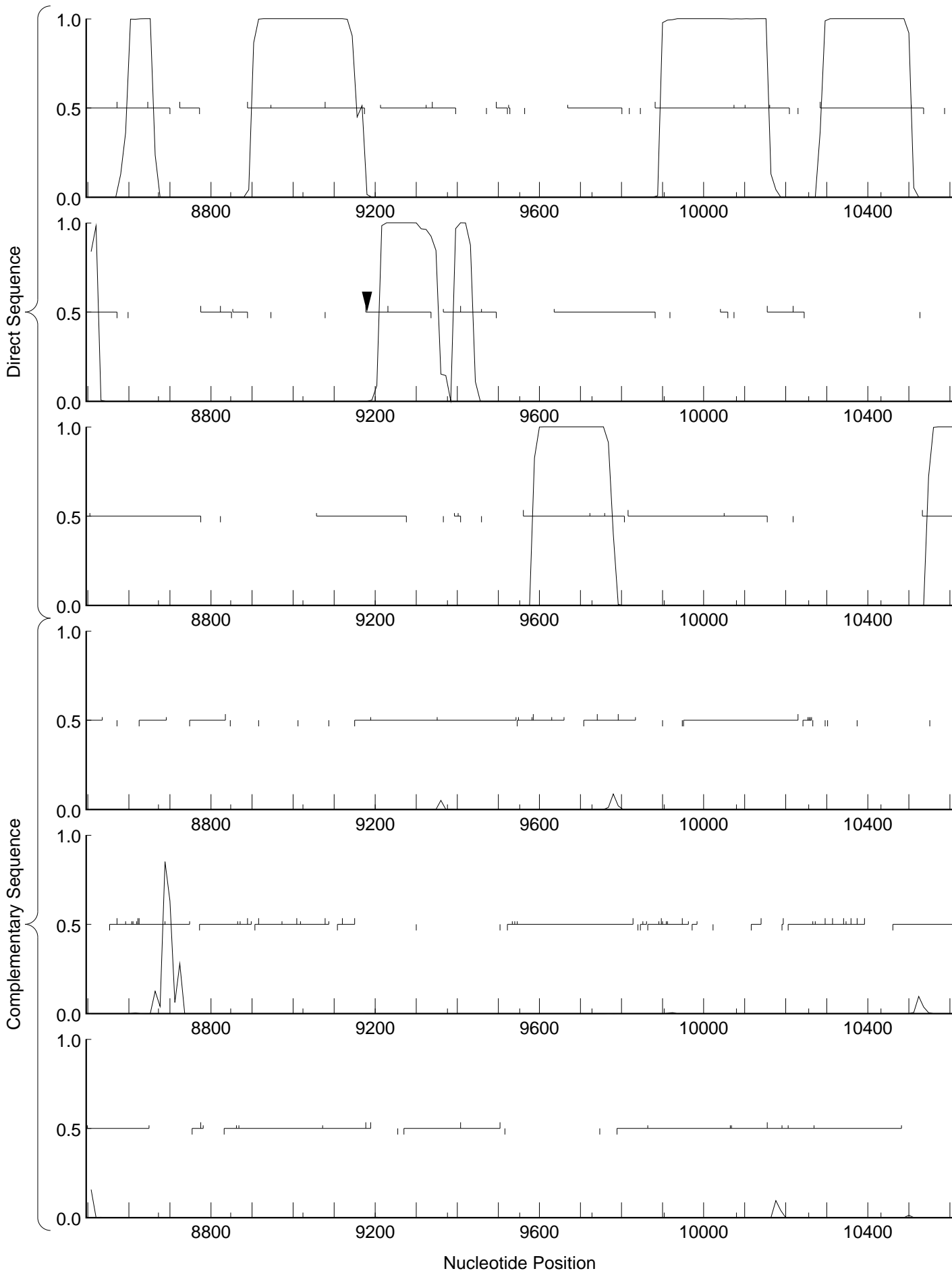
Send questions / comments to:
Dr. M. Borodovsky
Georgia Institute of Technology
School of Biology
Atlanta, GA 30332-0230

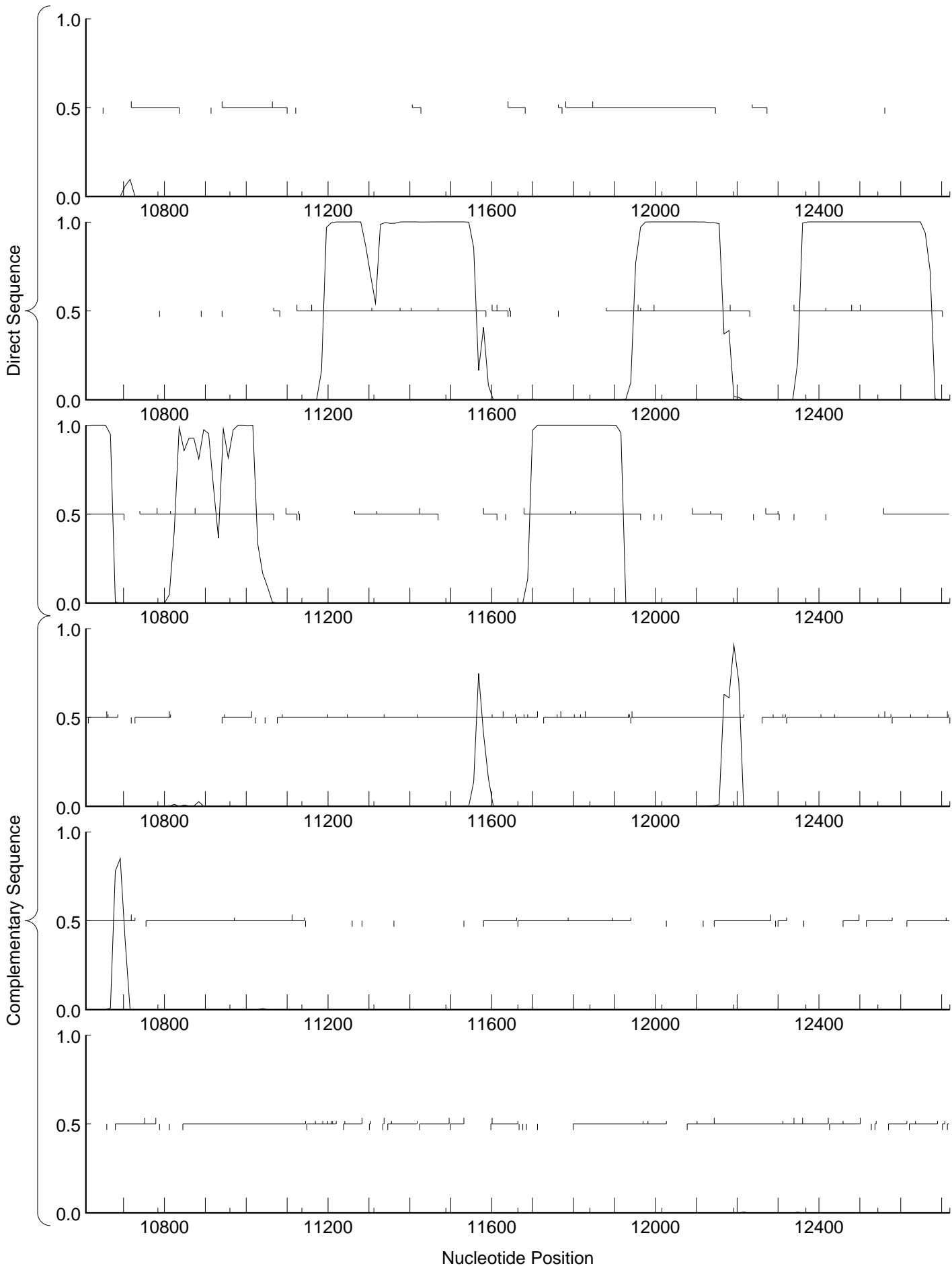


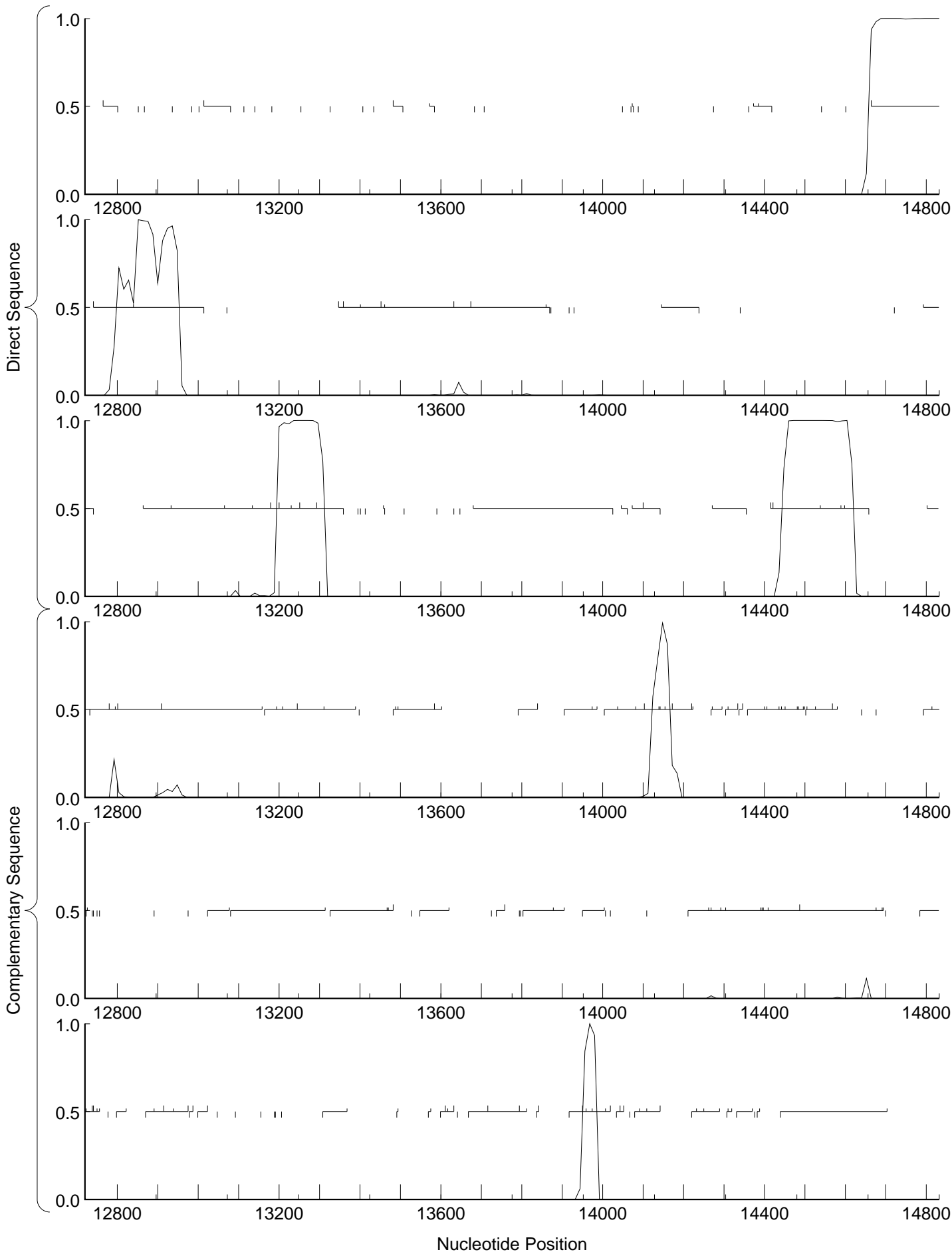


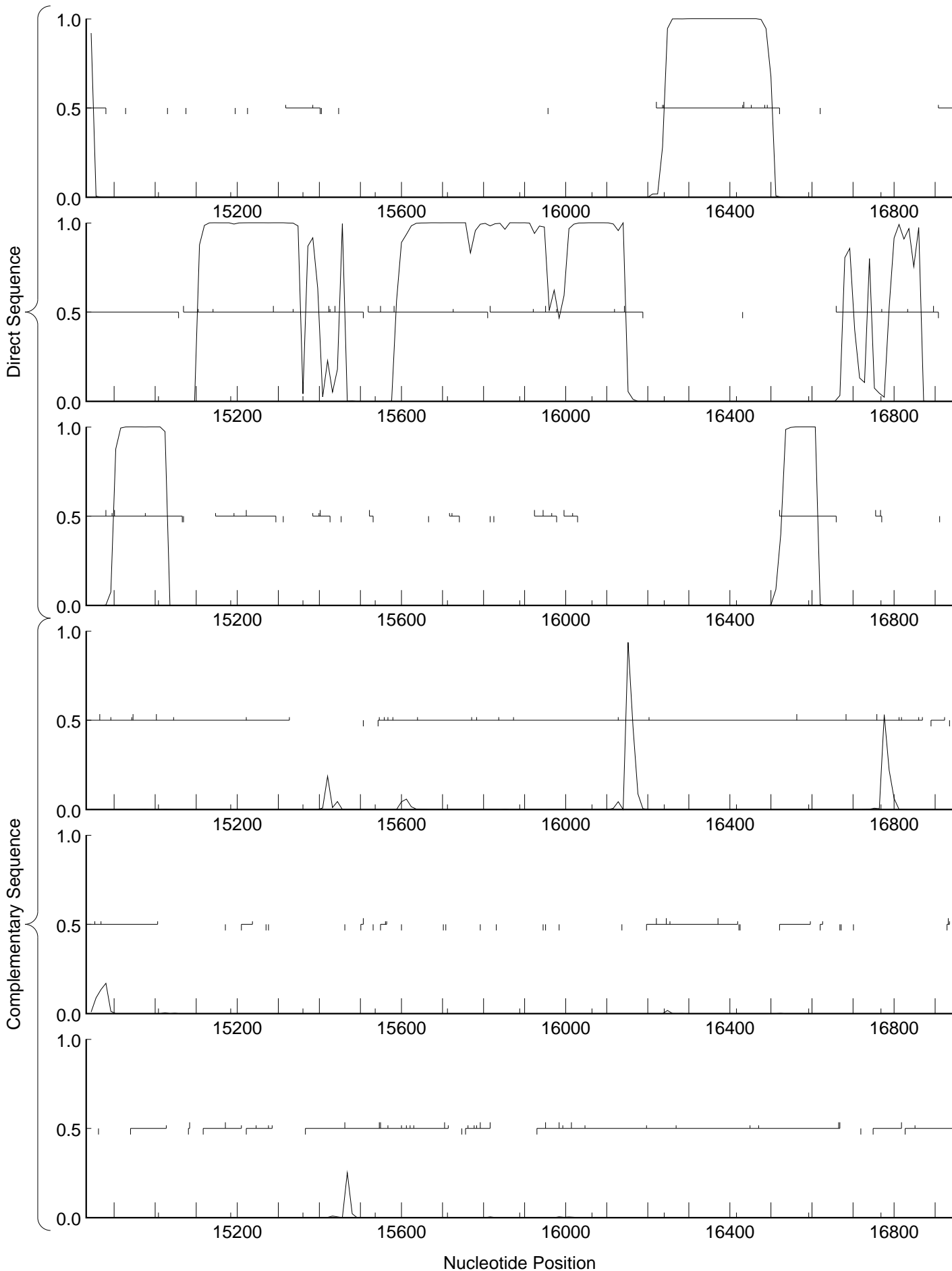


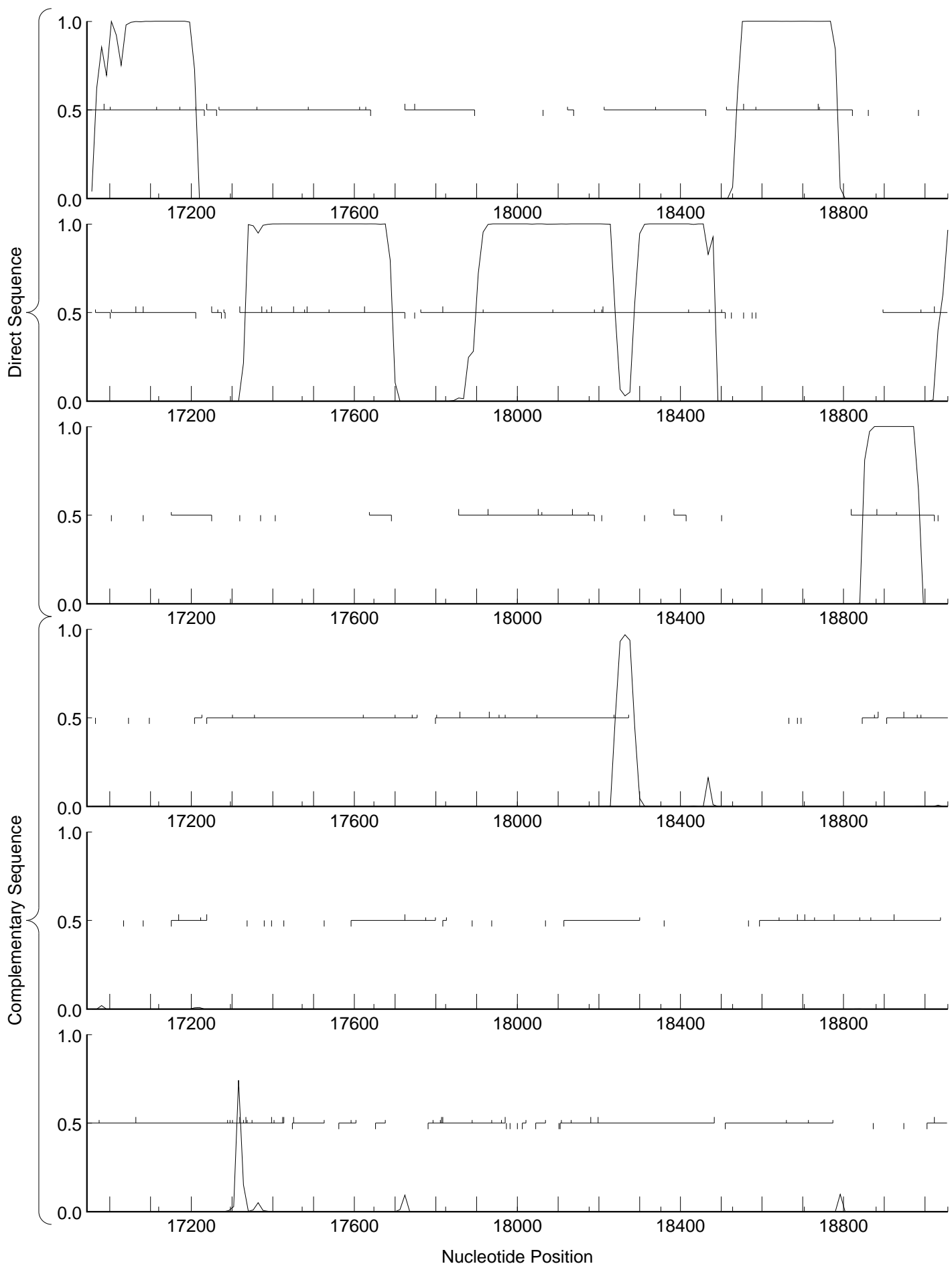


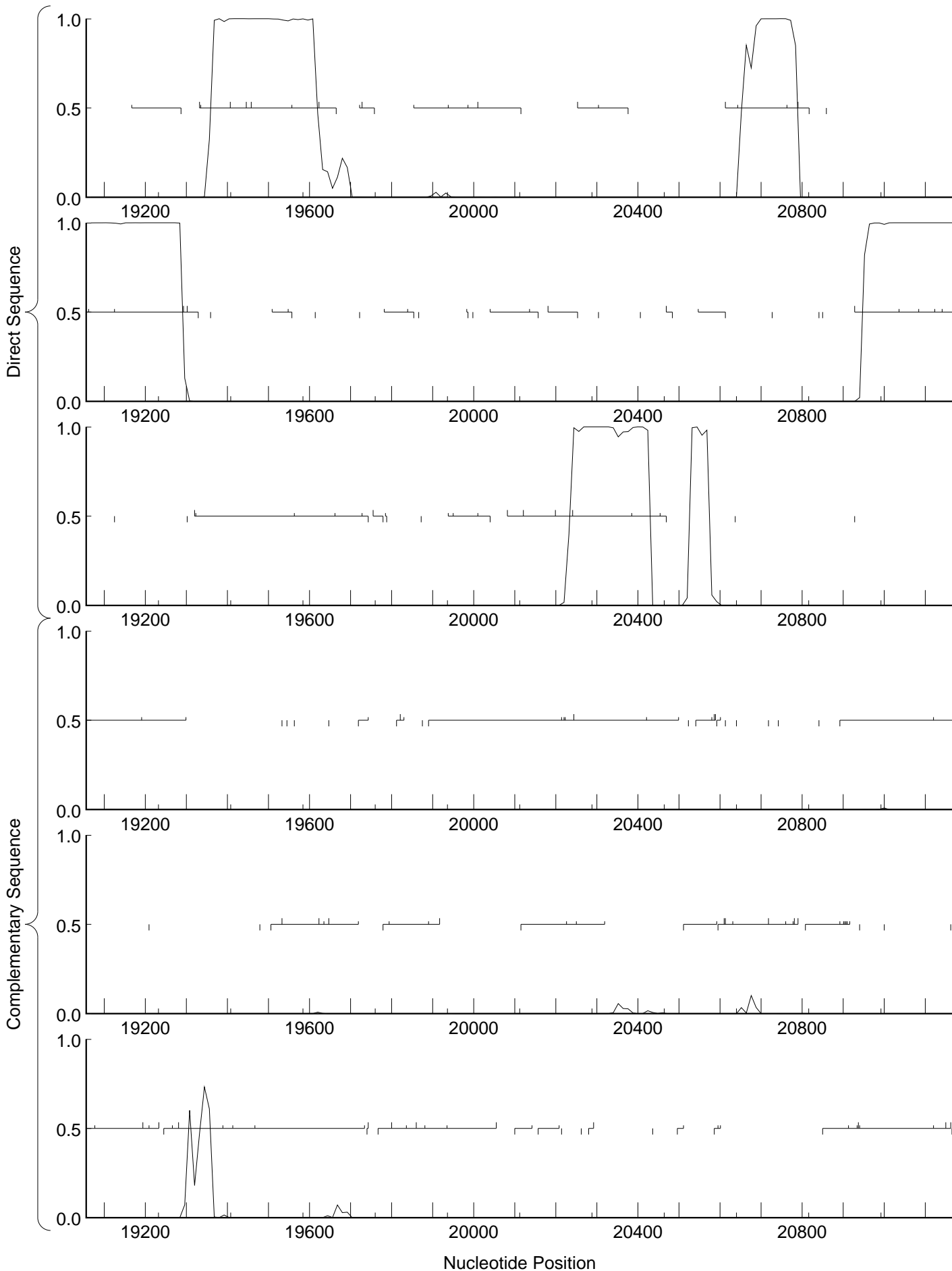


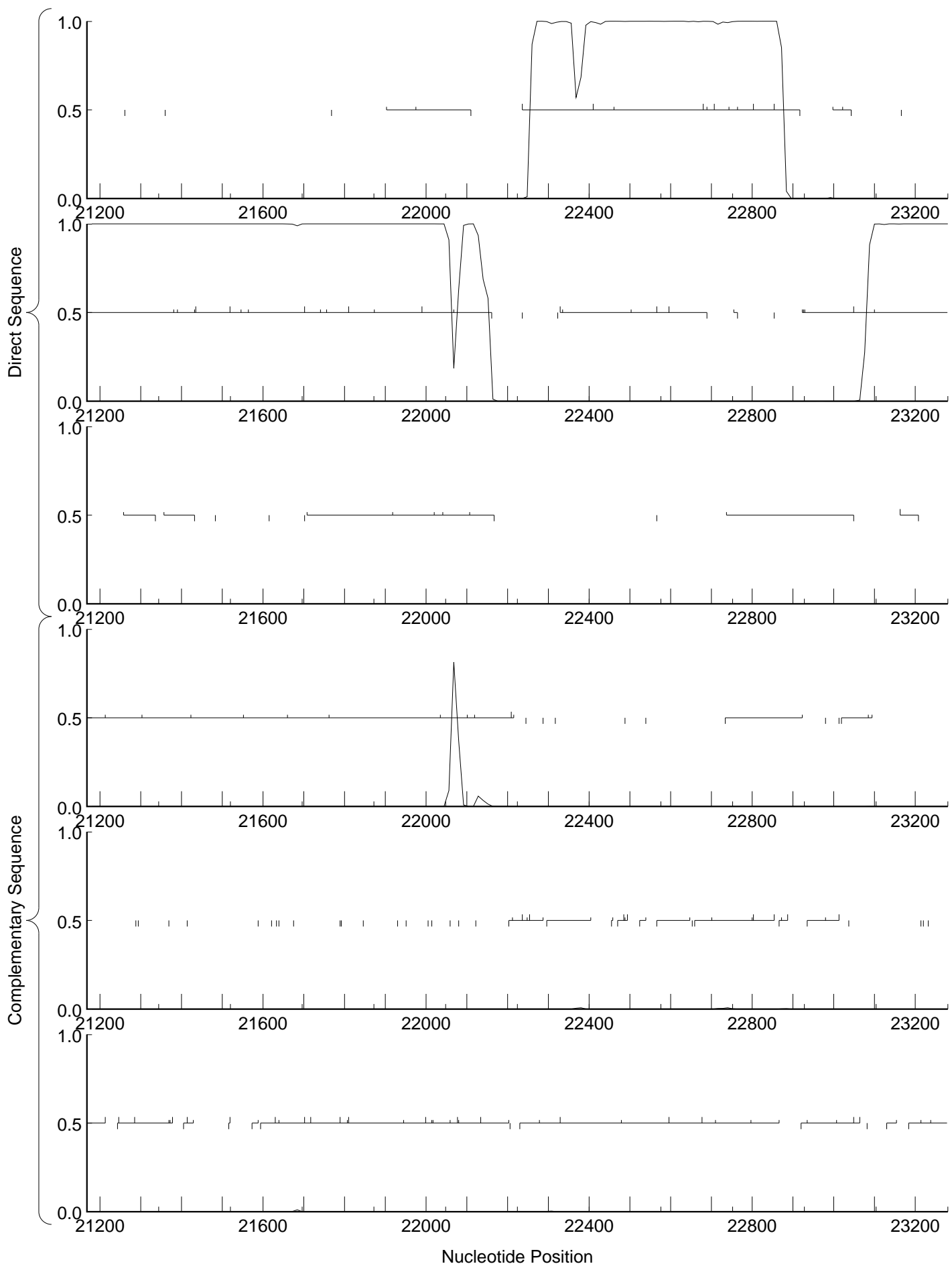


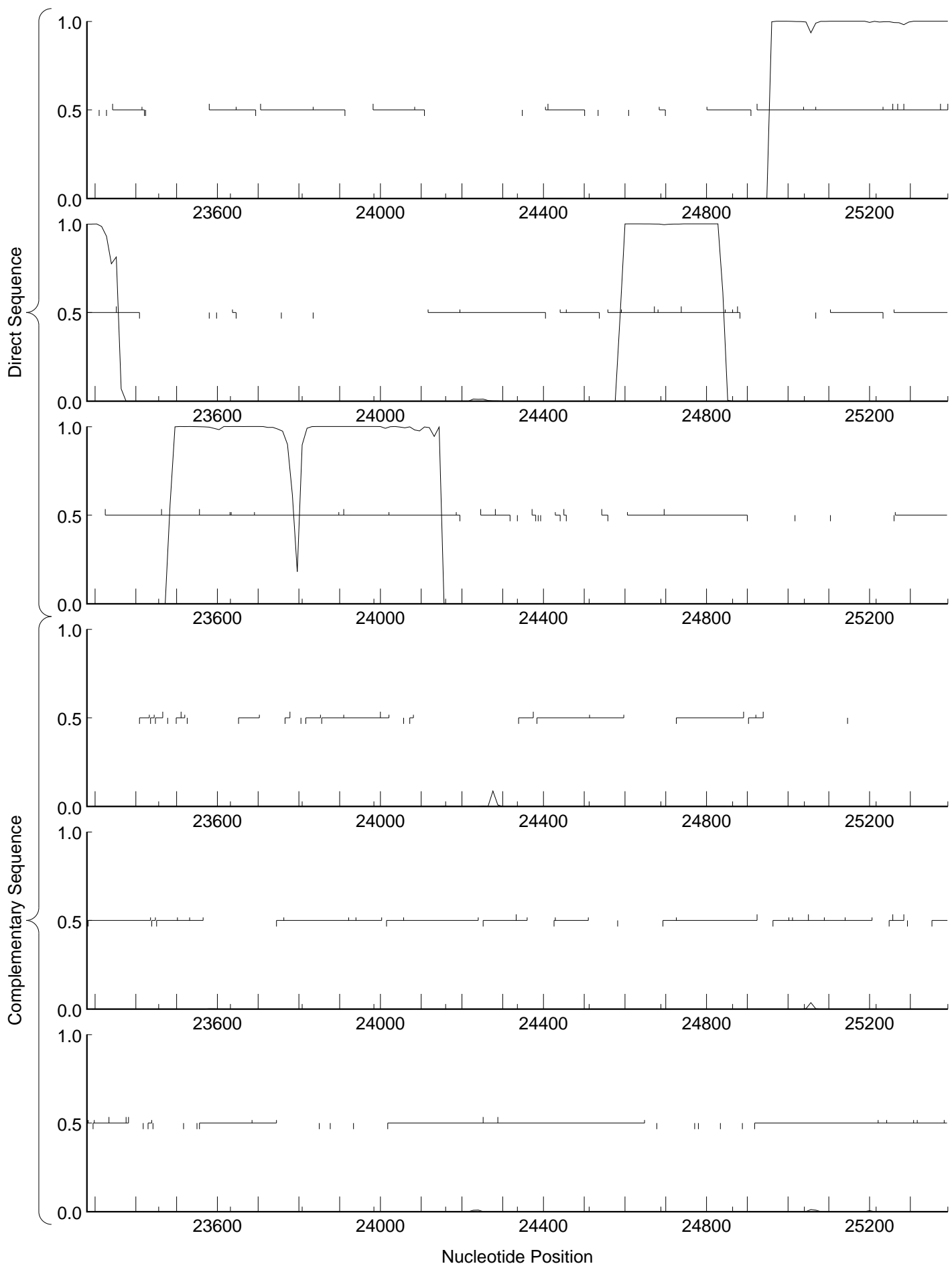


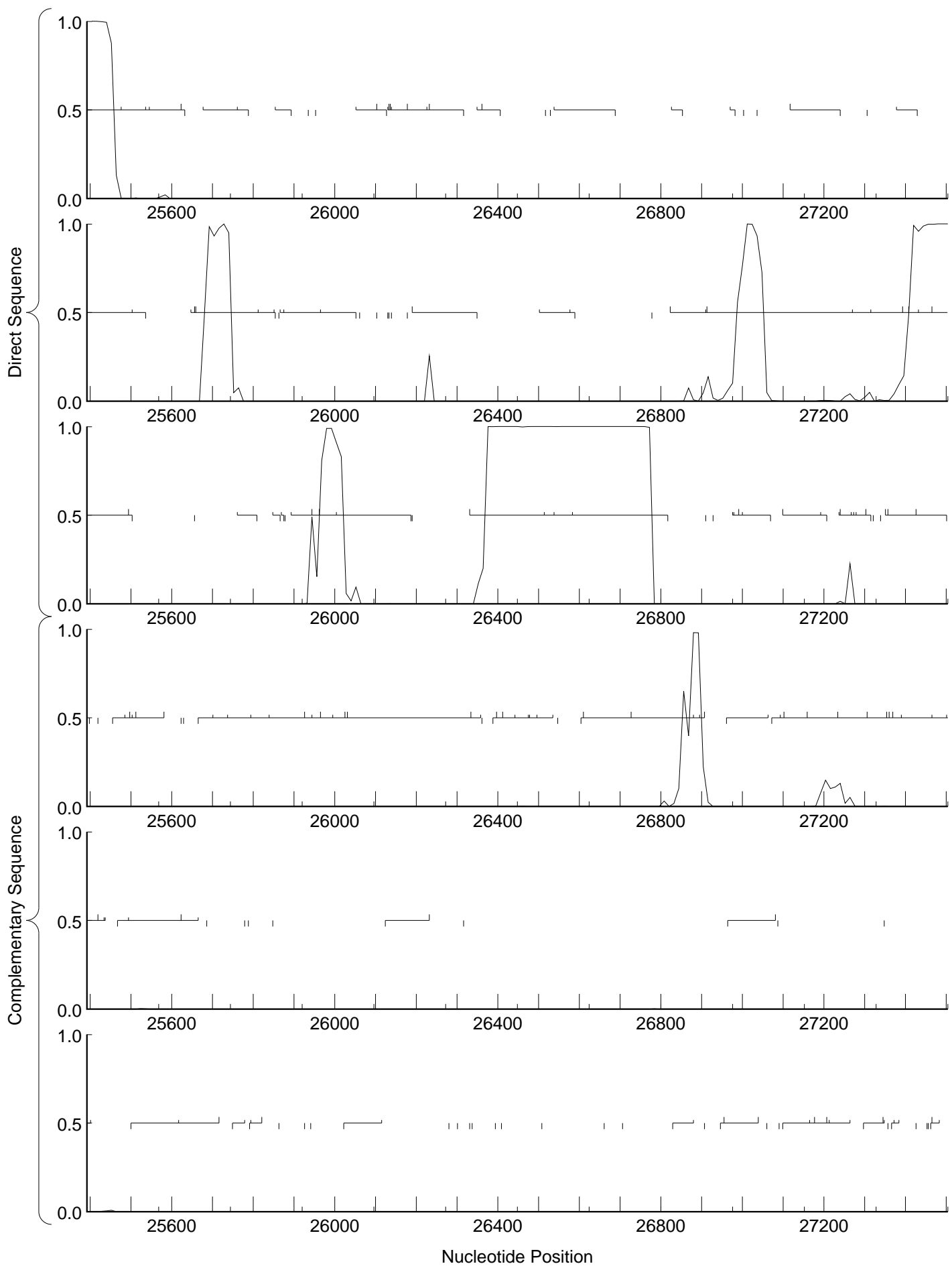


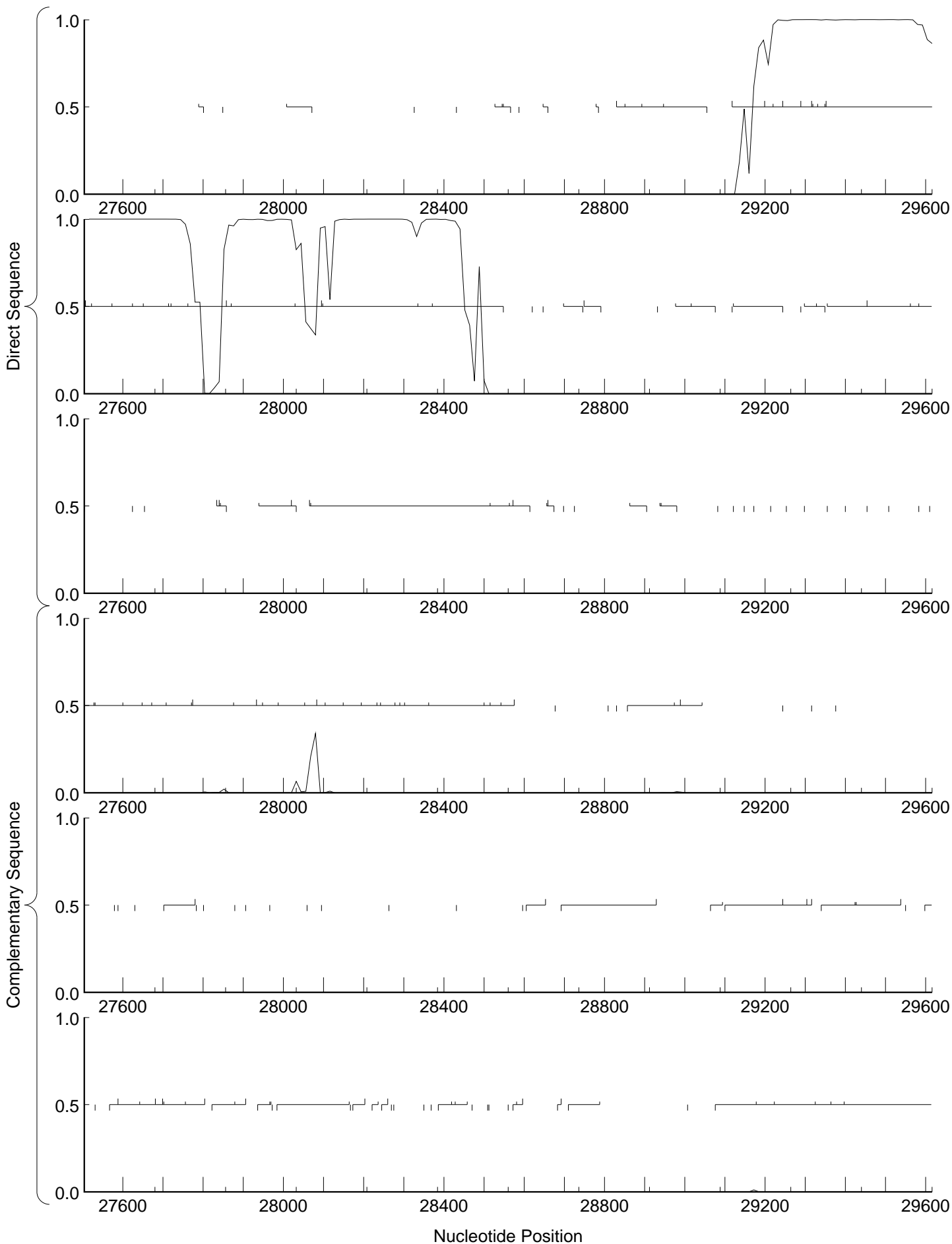


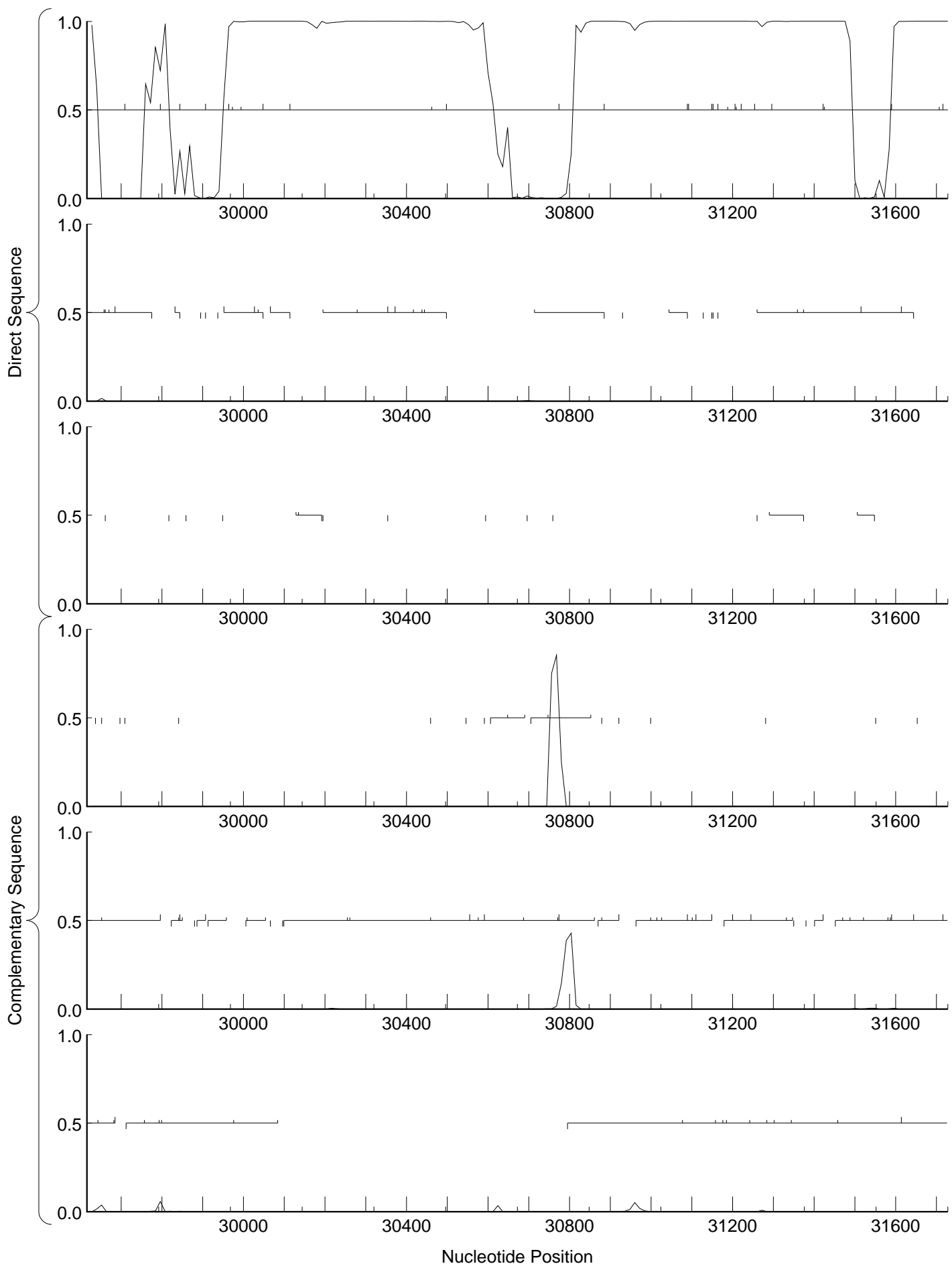


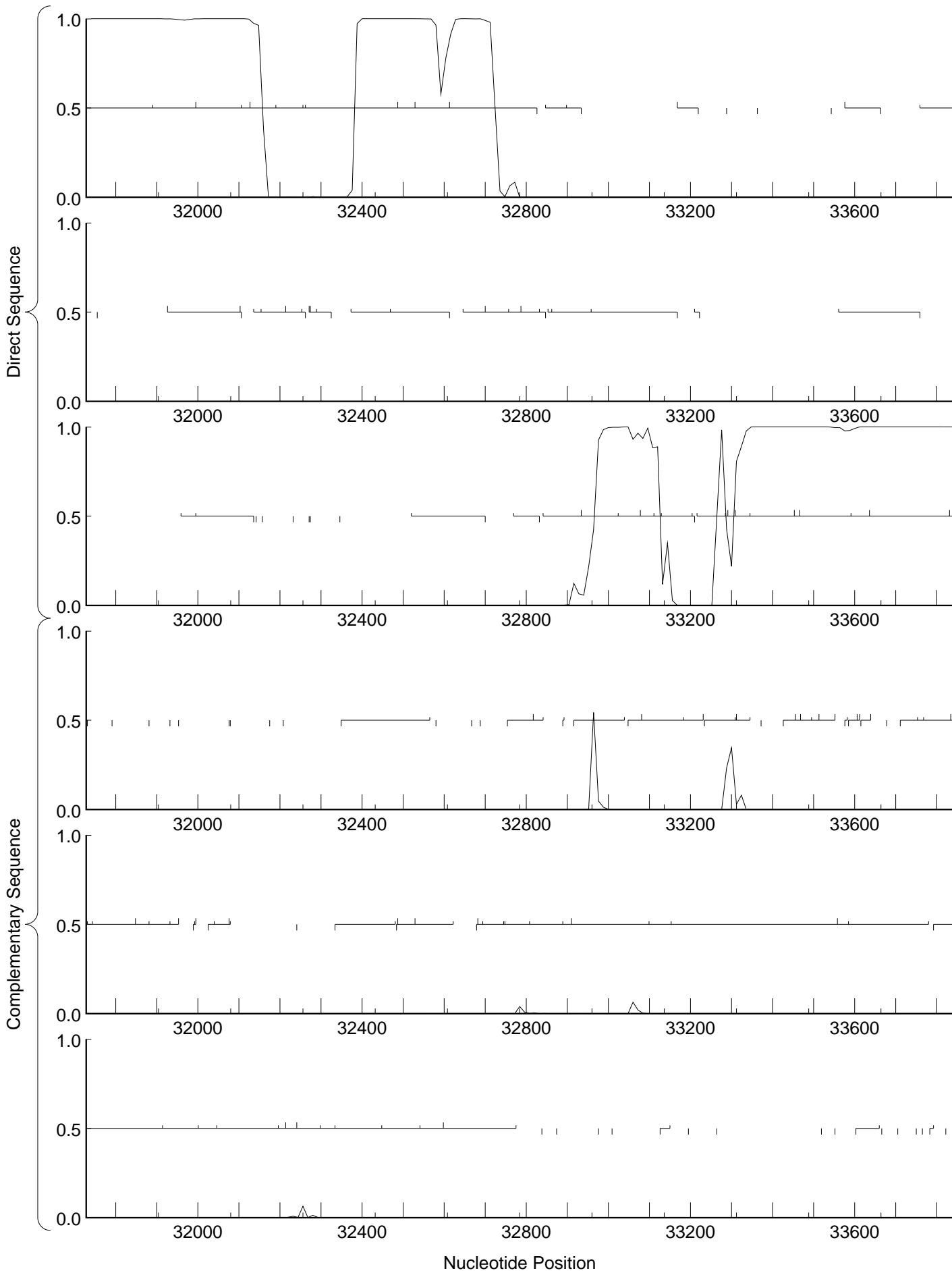


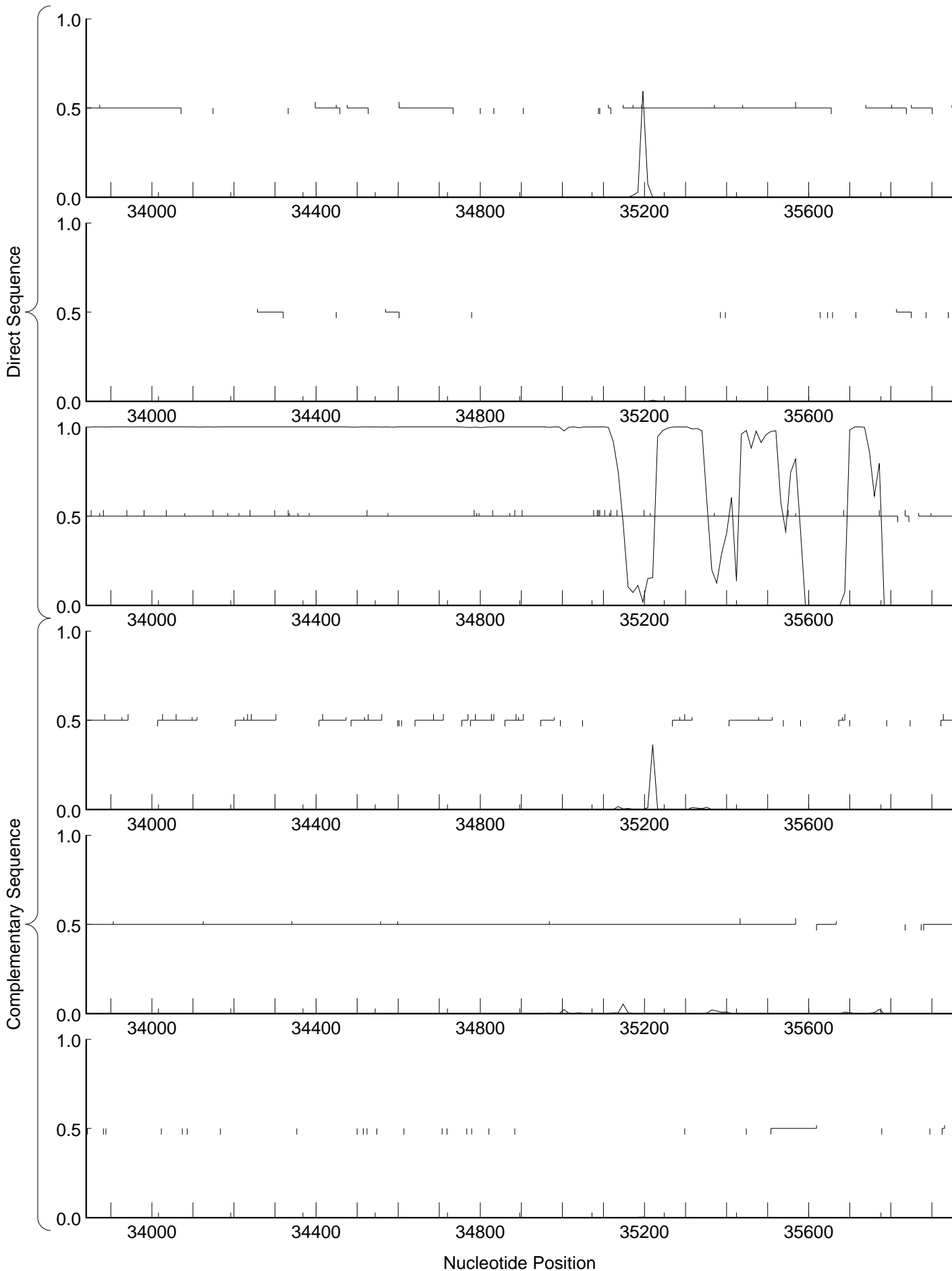


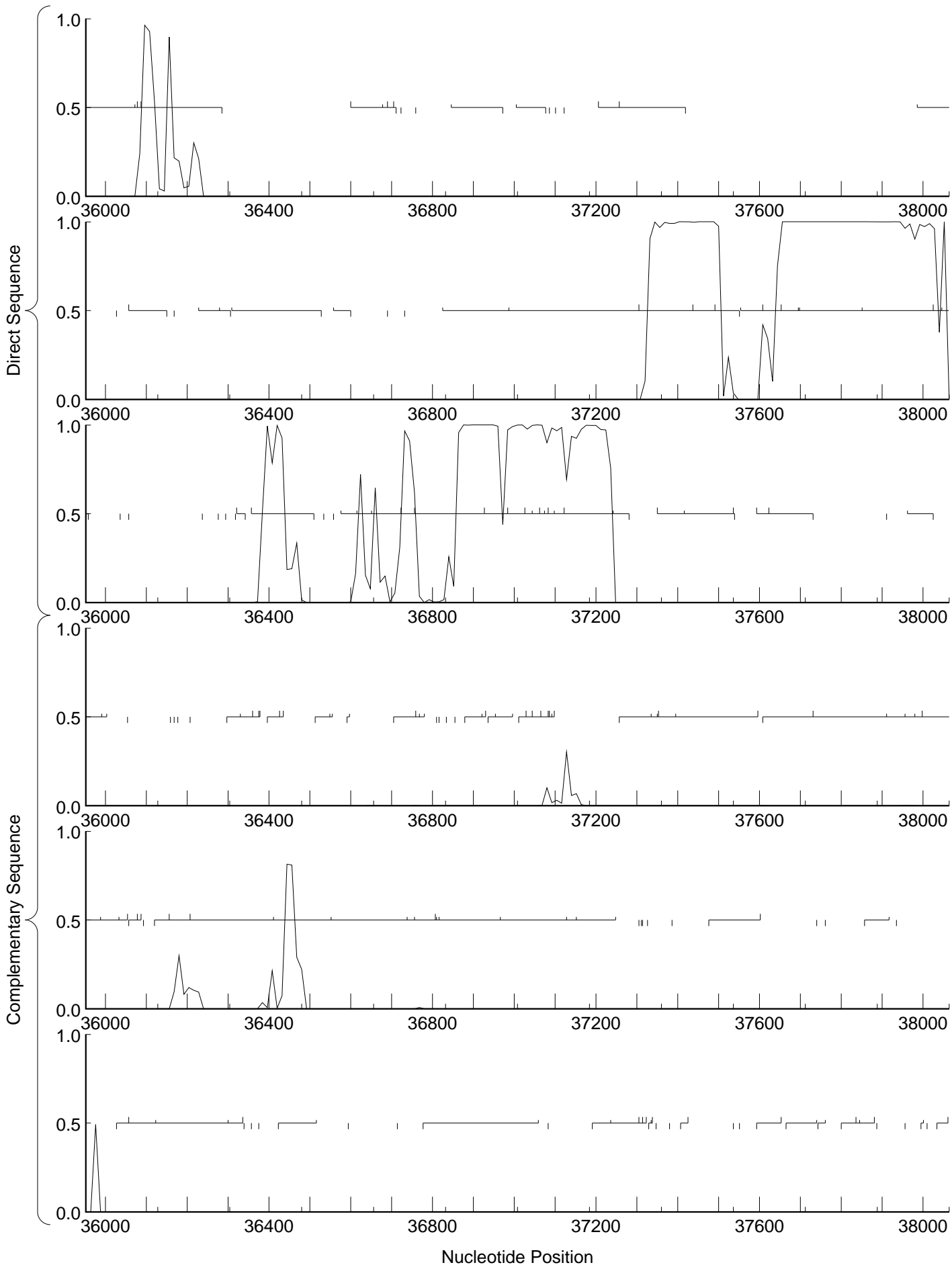


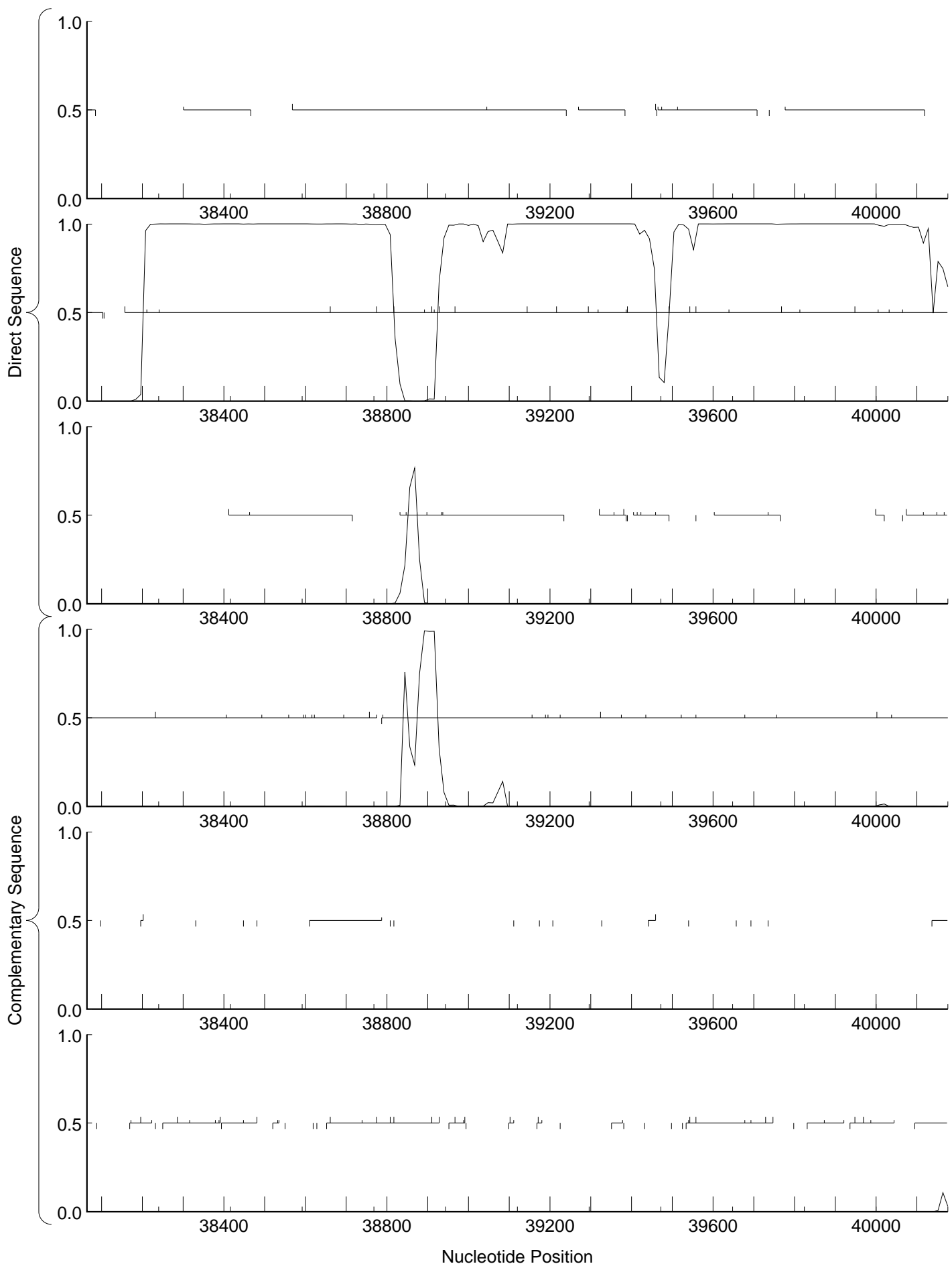


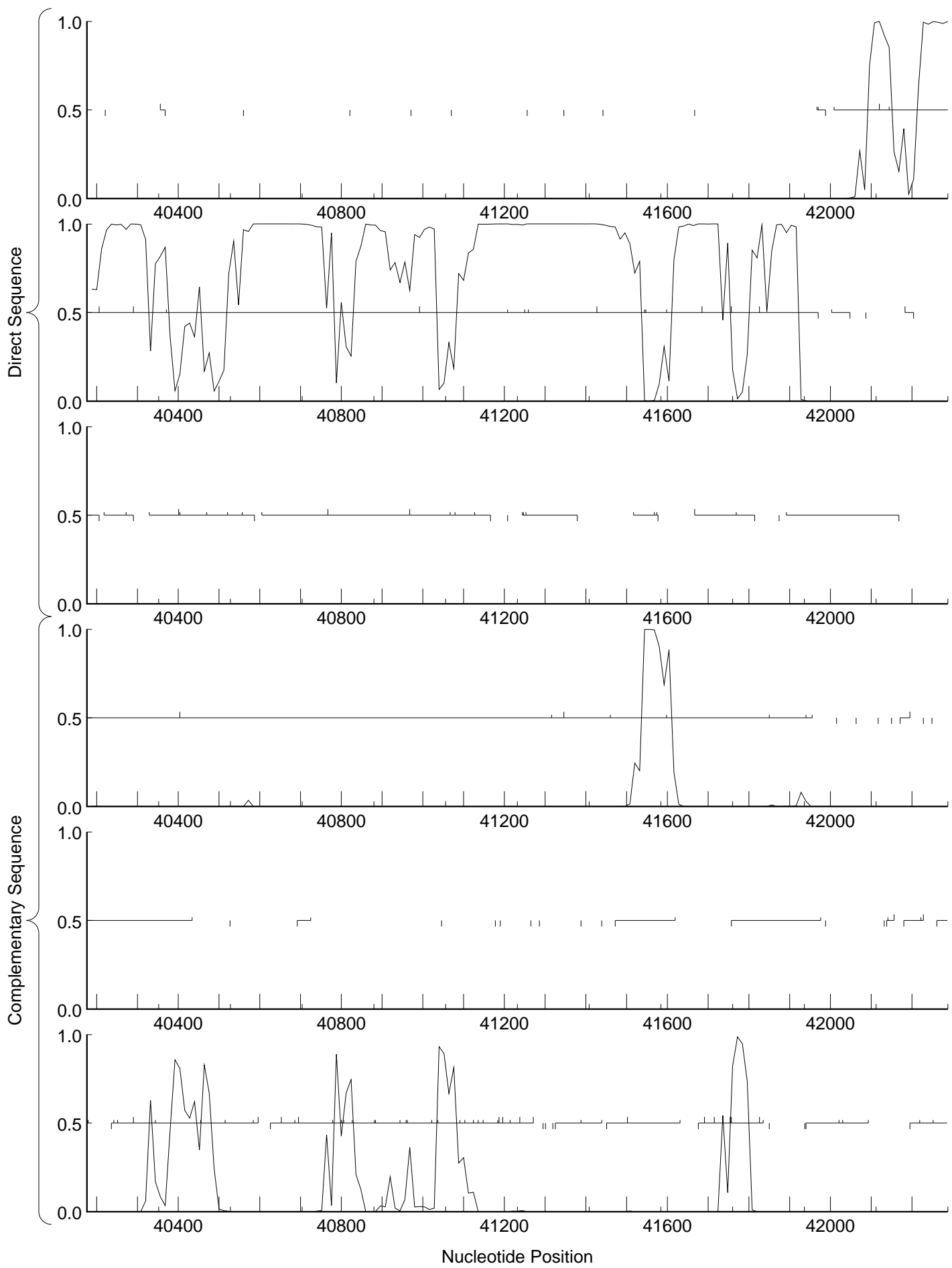


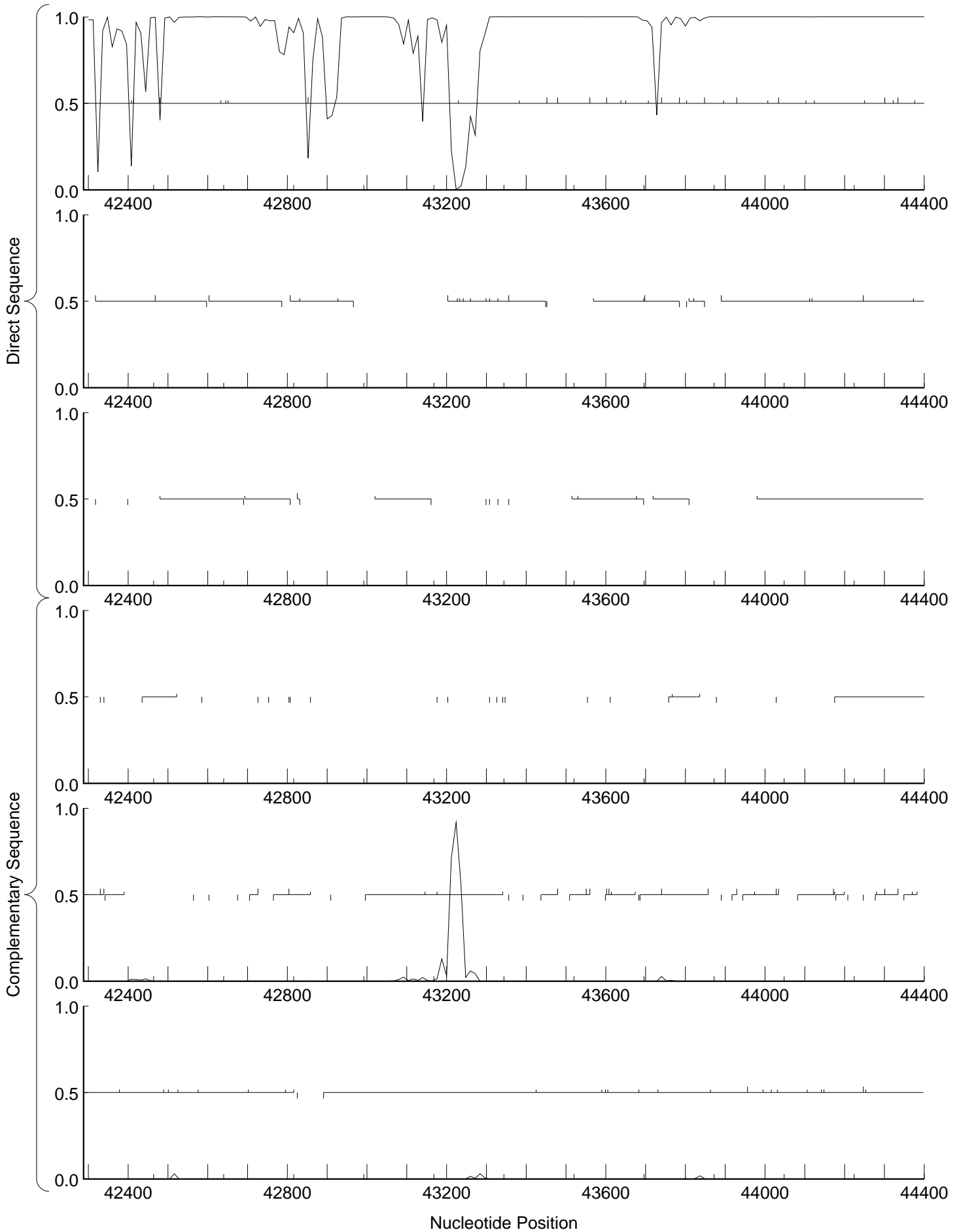


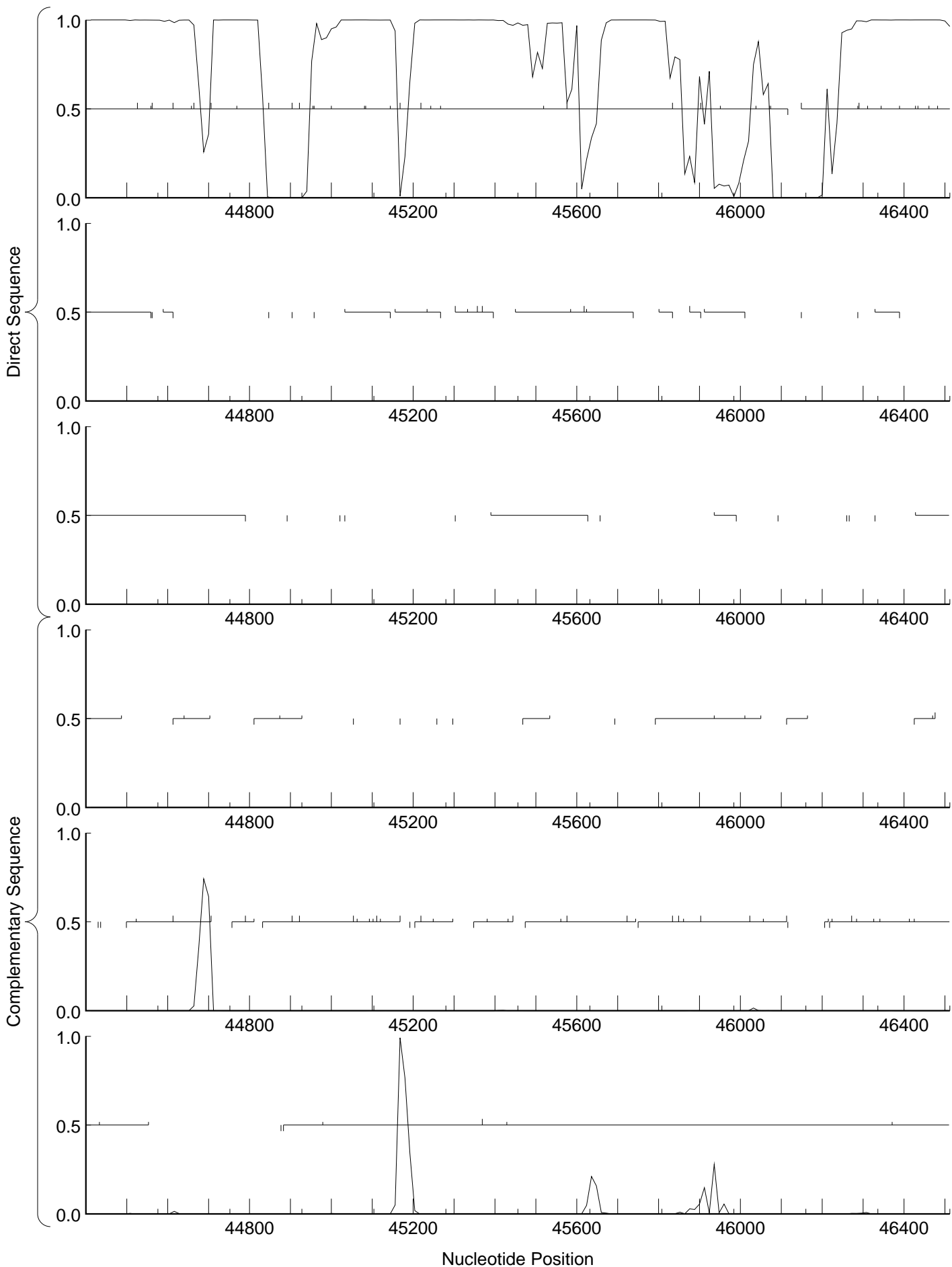


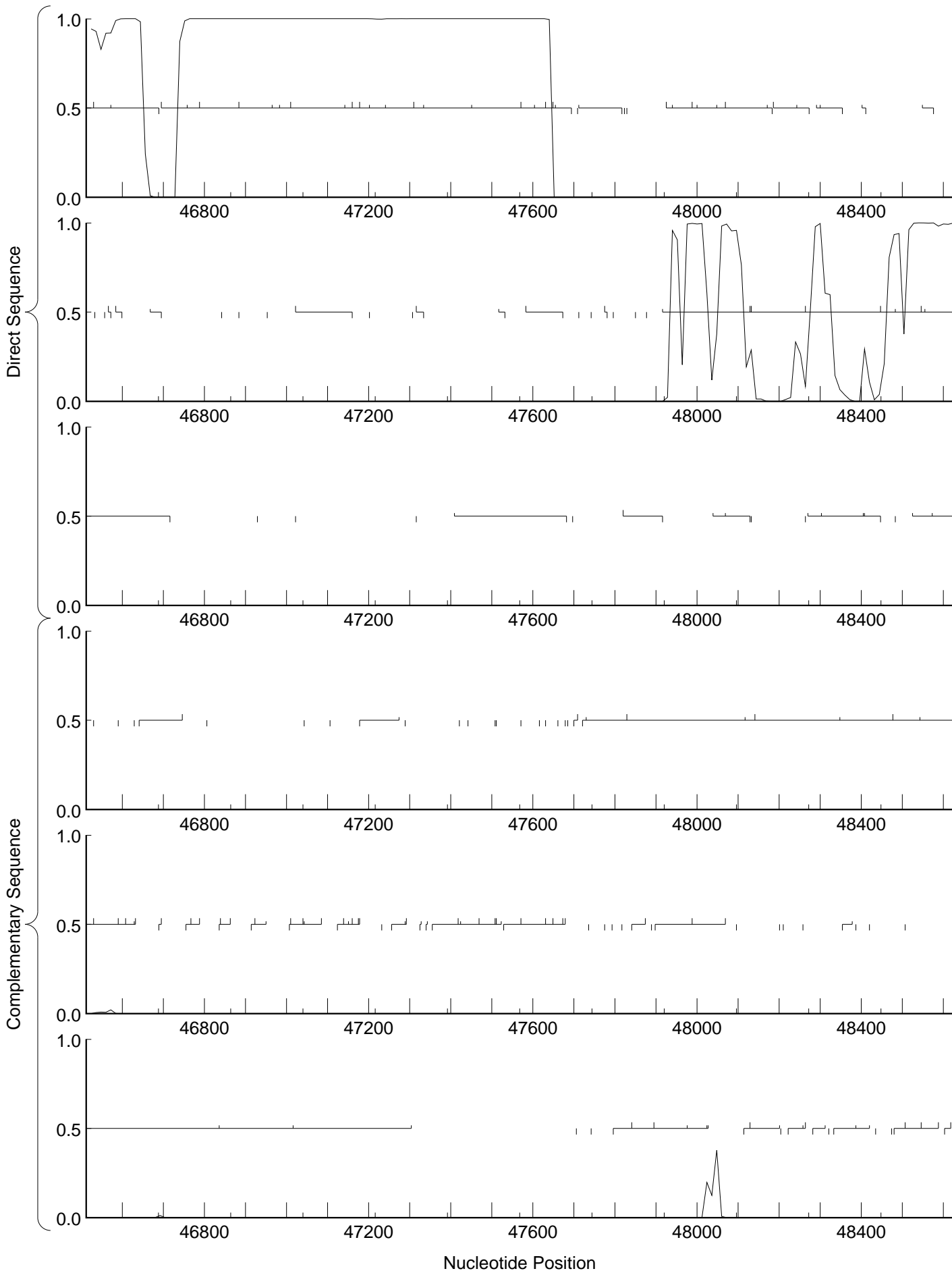


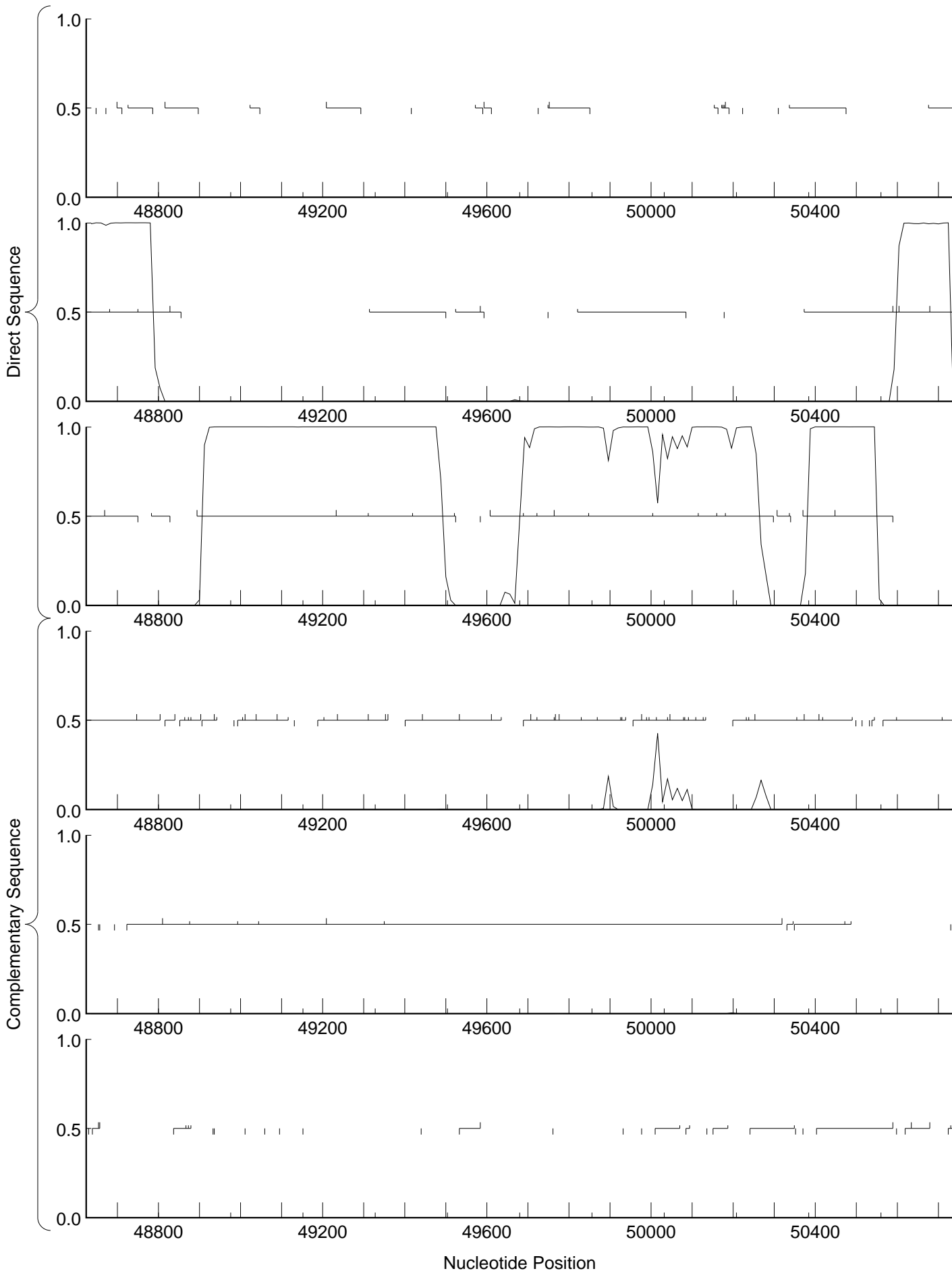


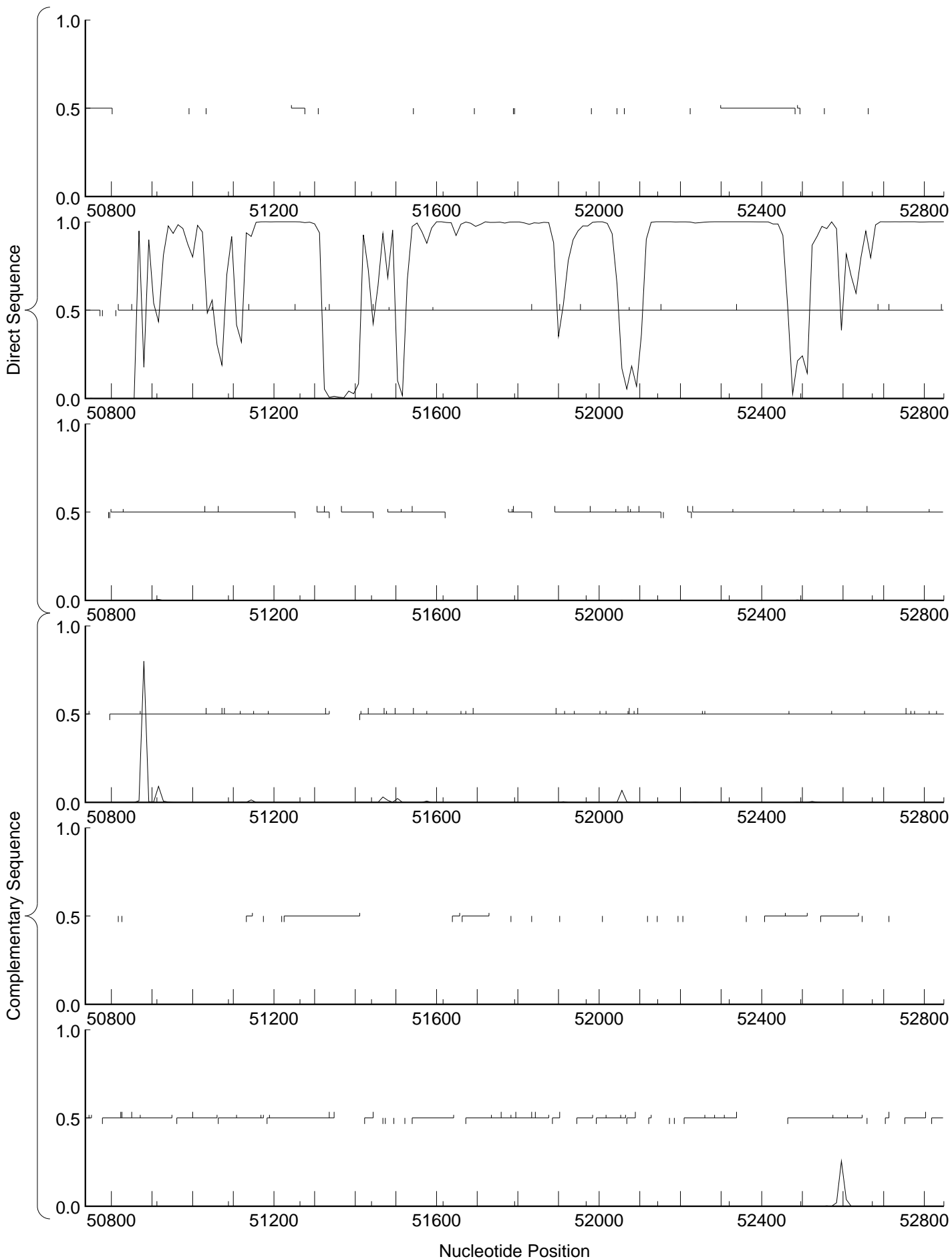


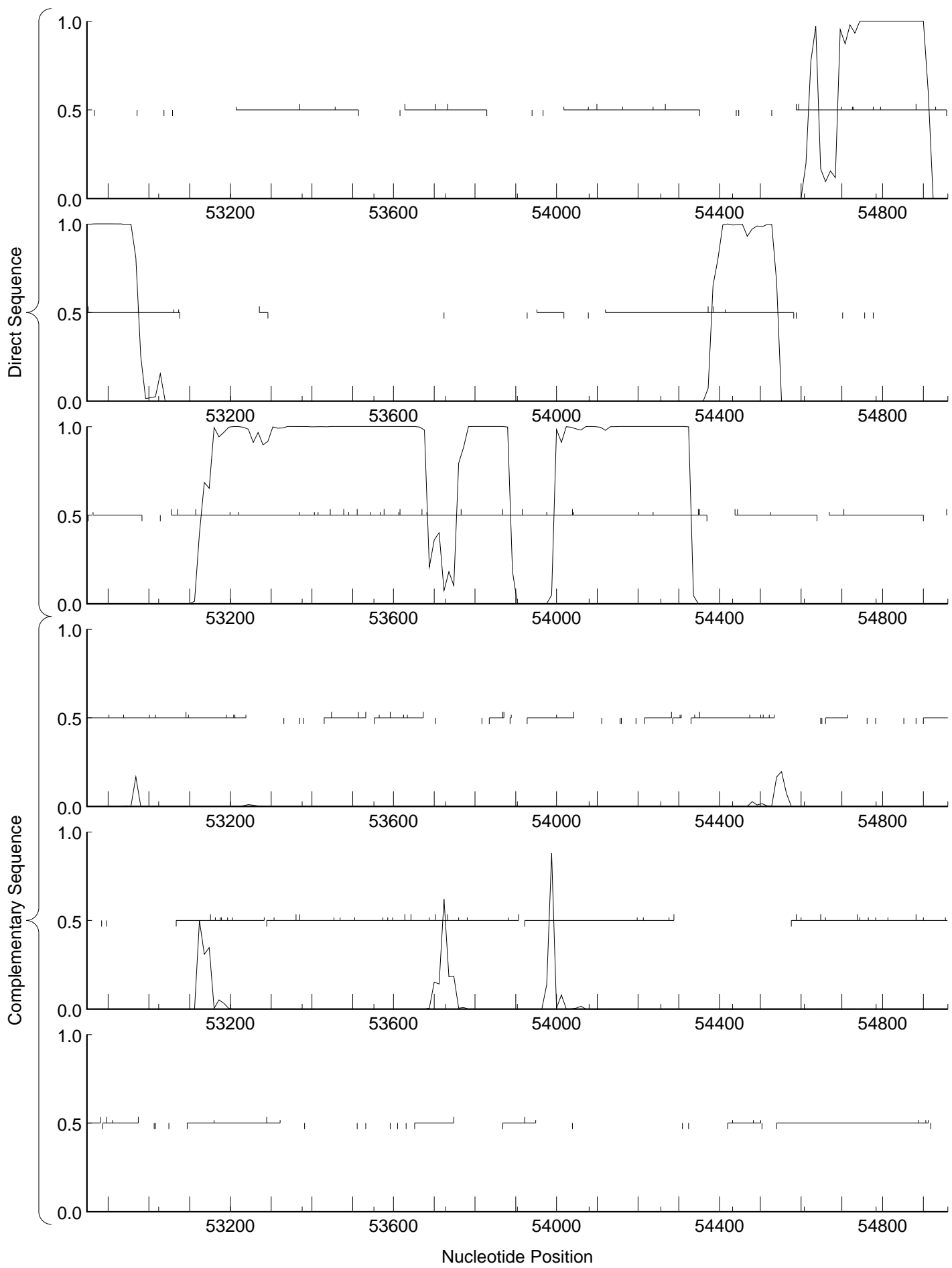


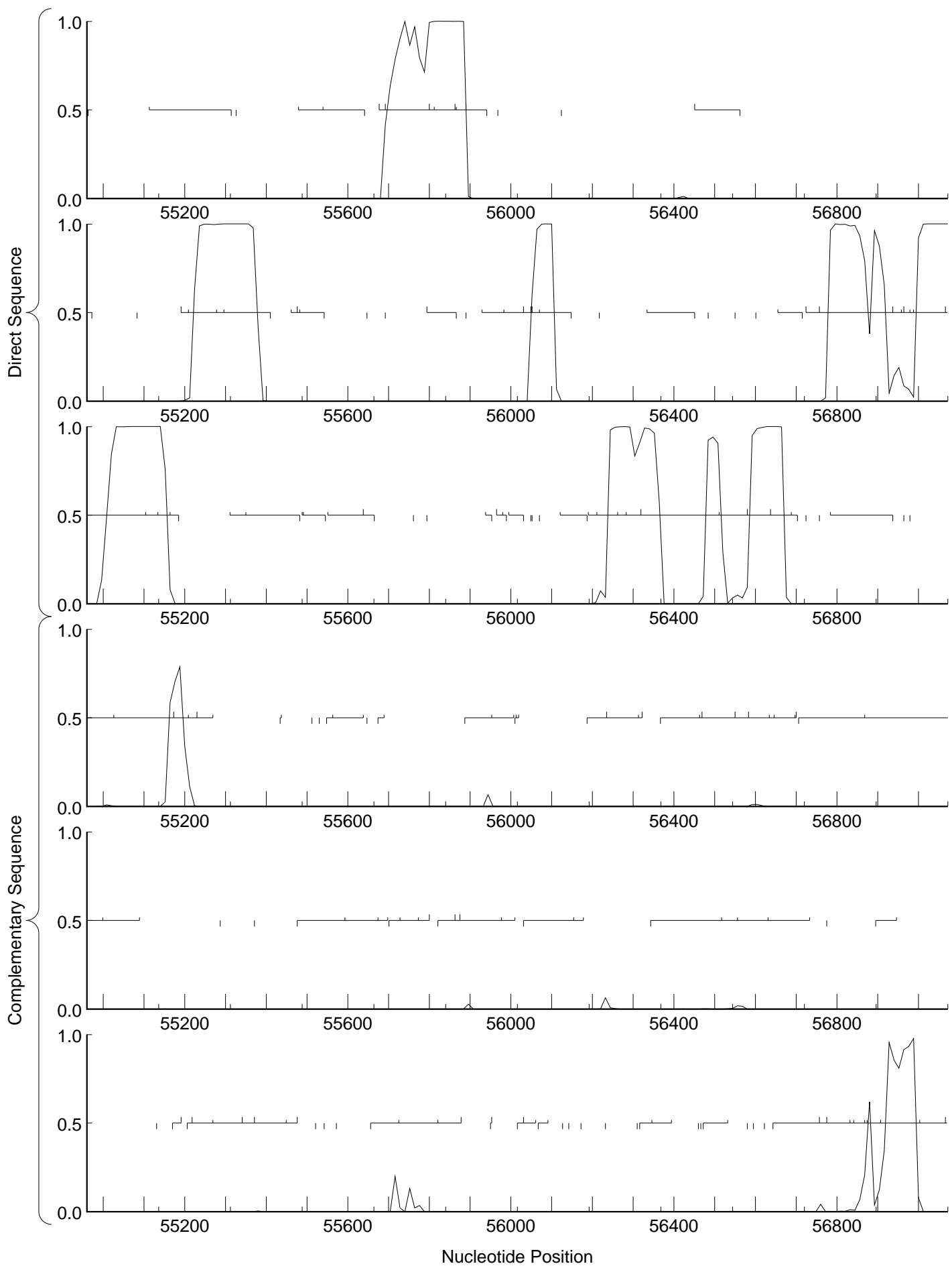


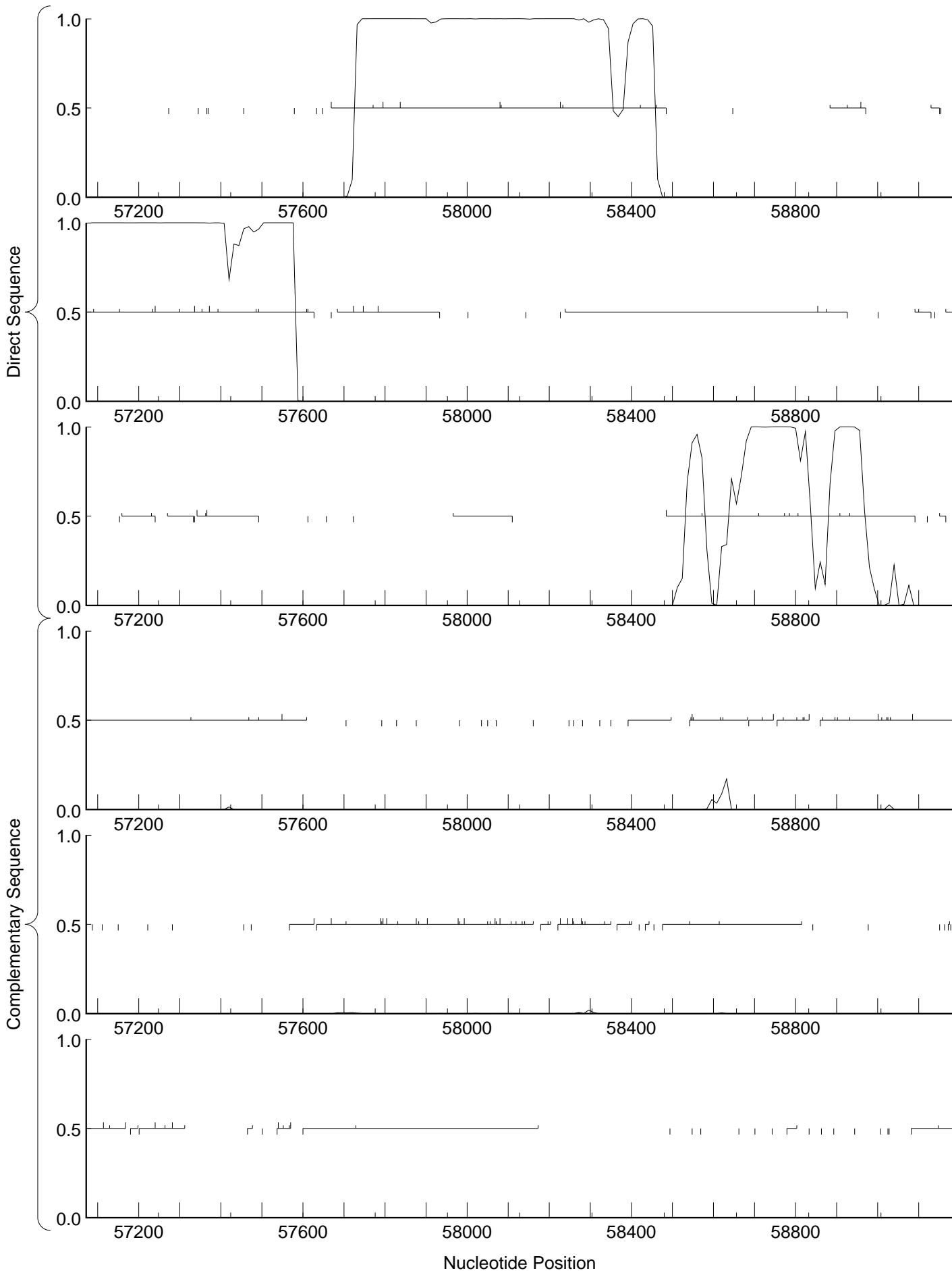


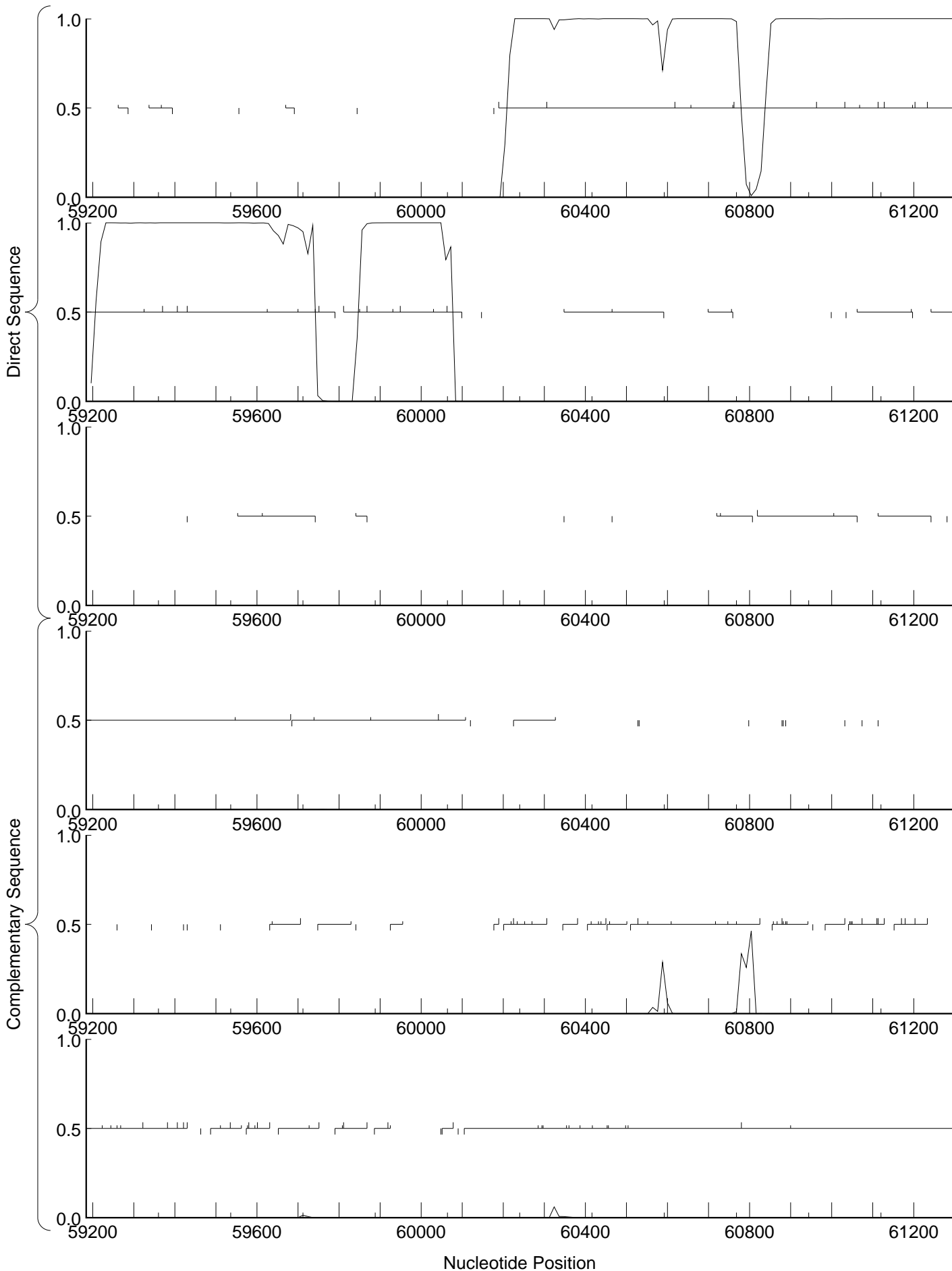


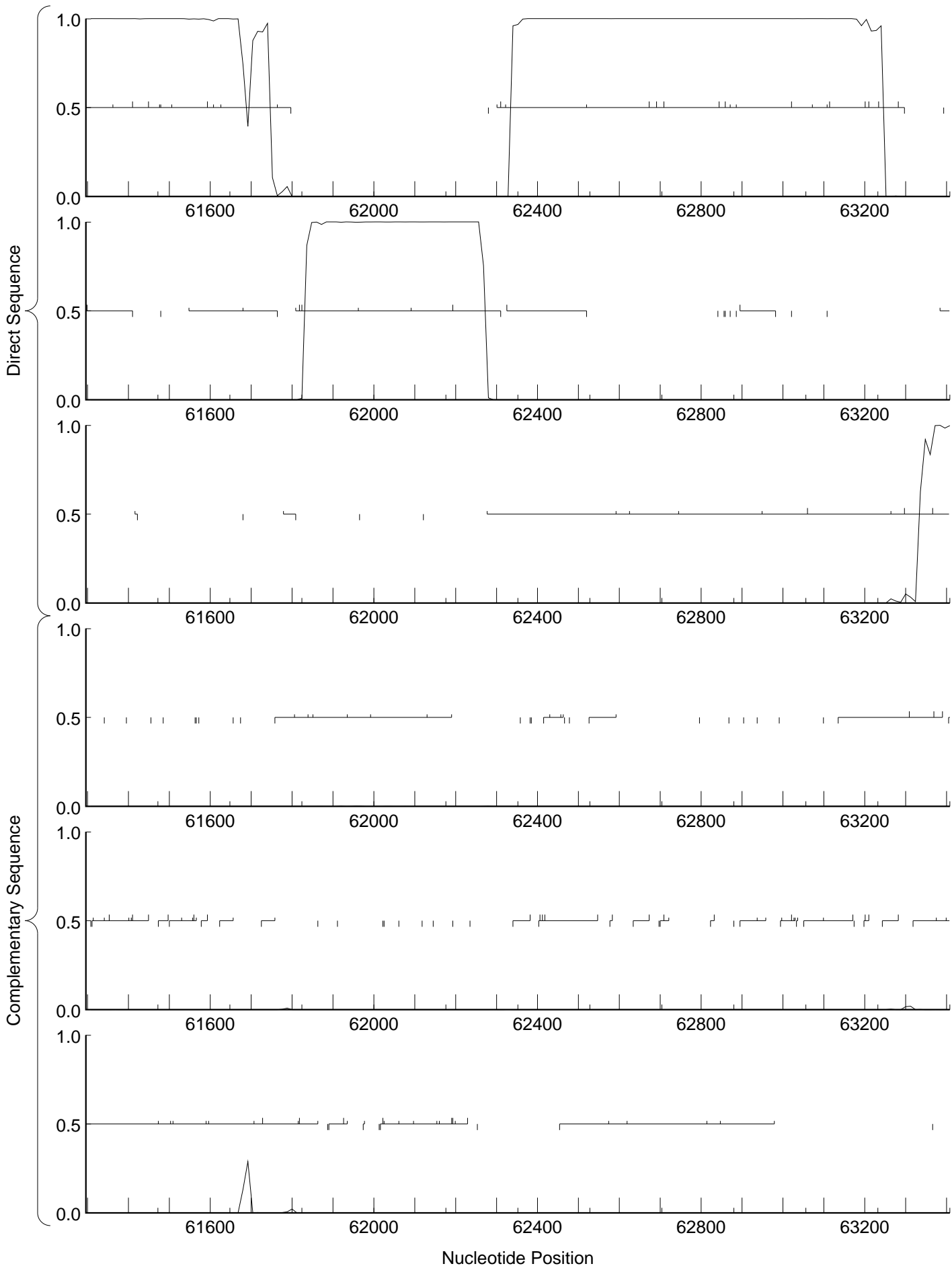


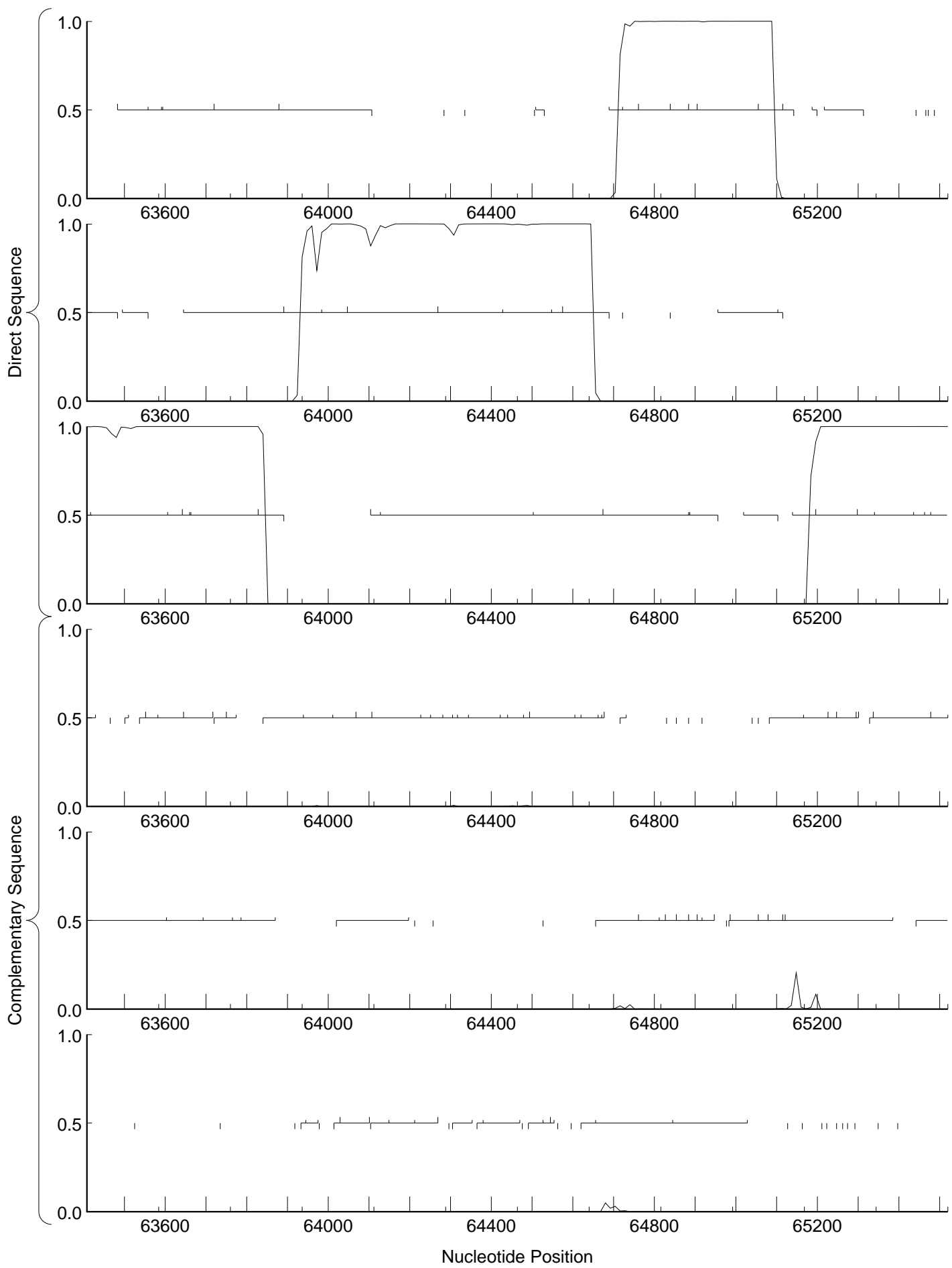


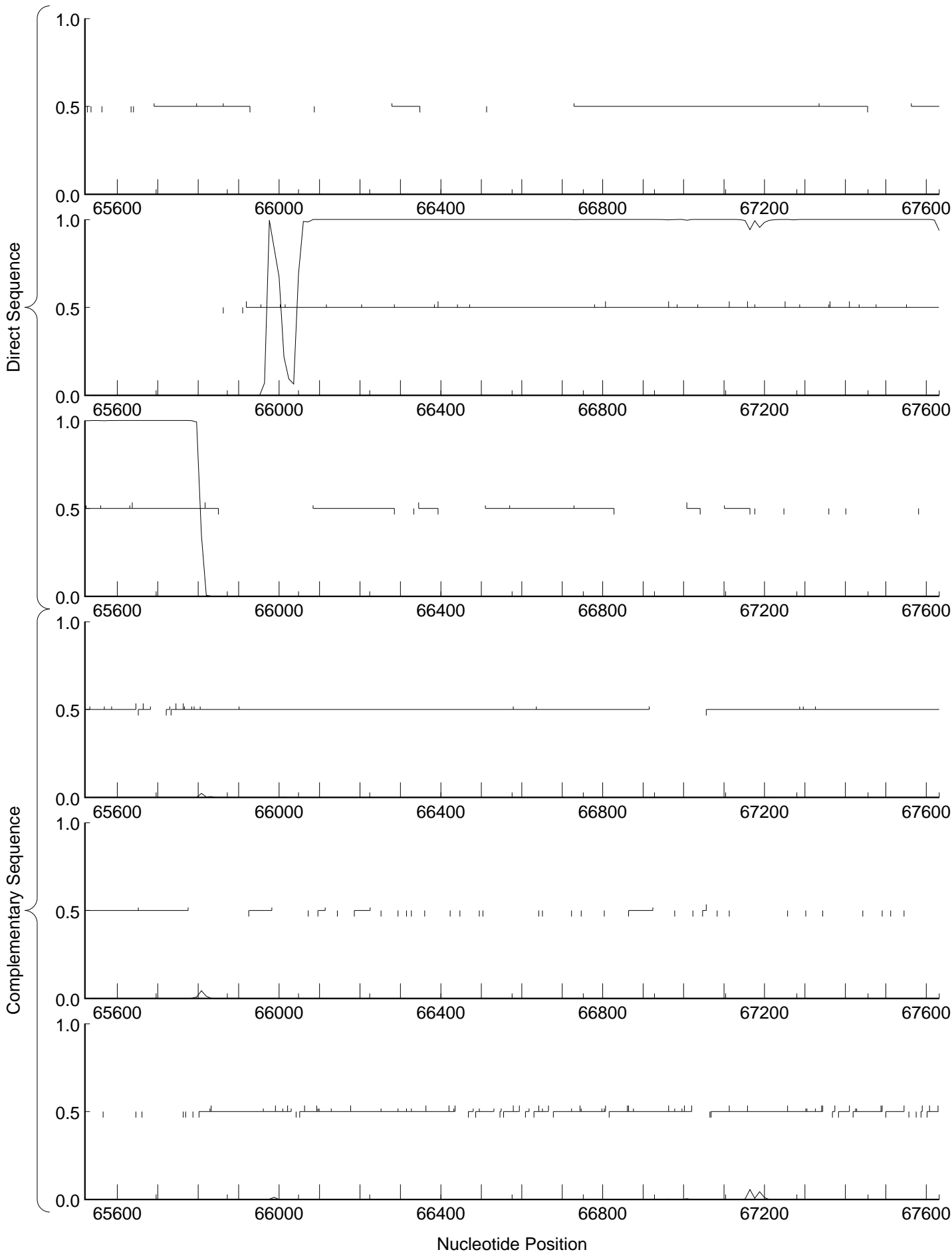


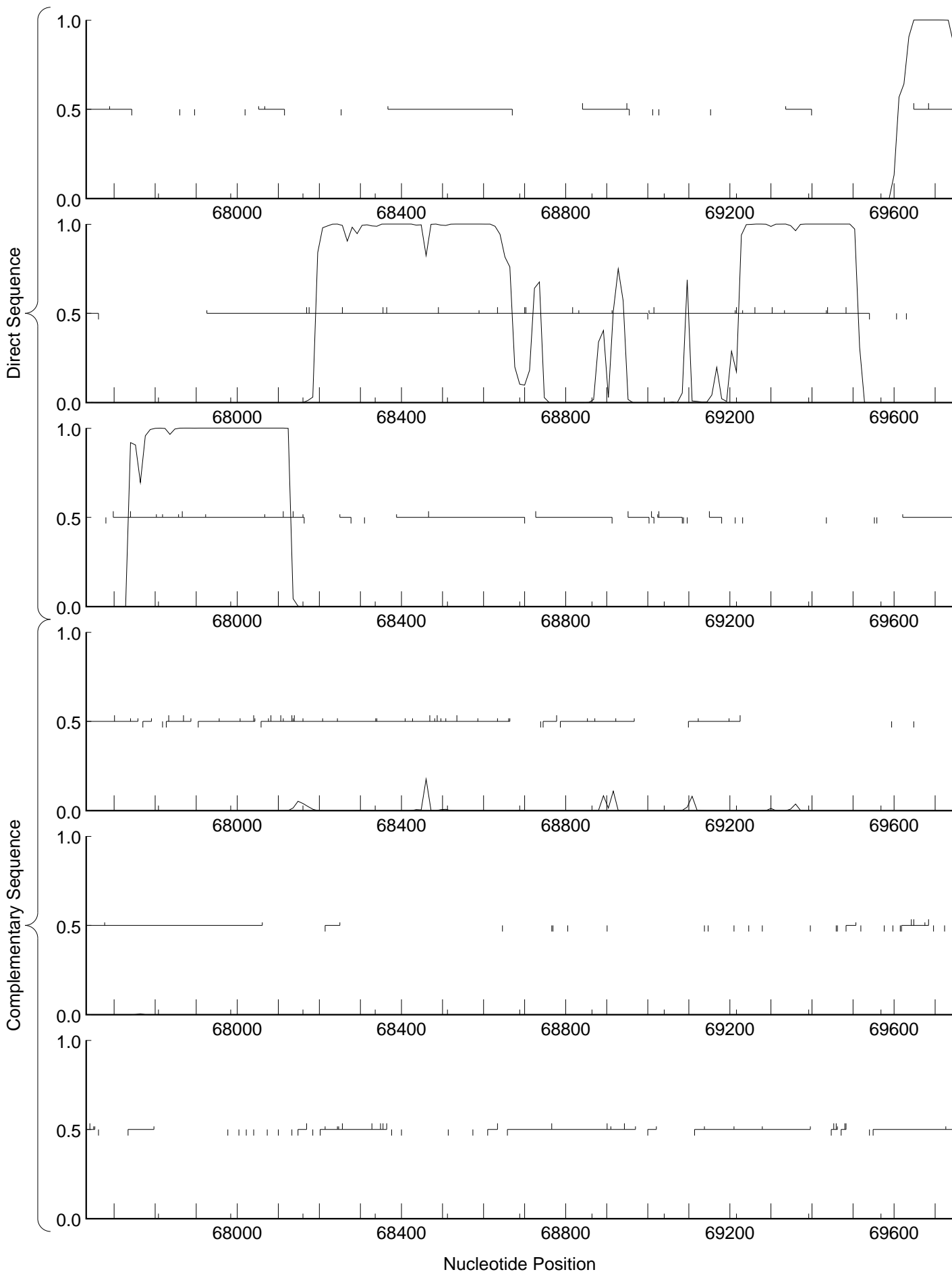


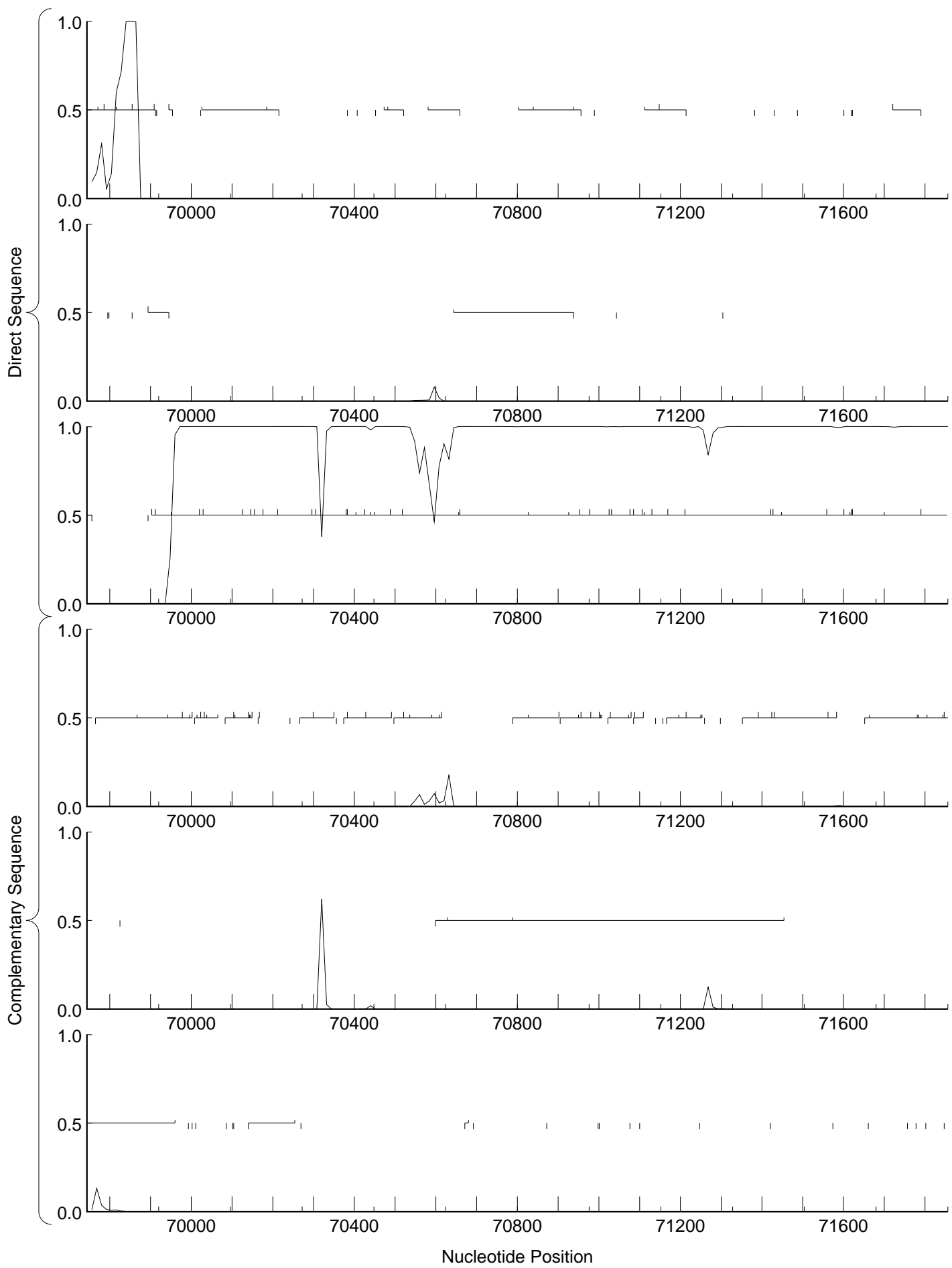


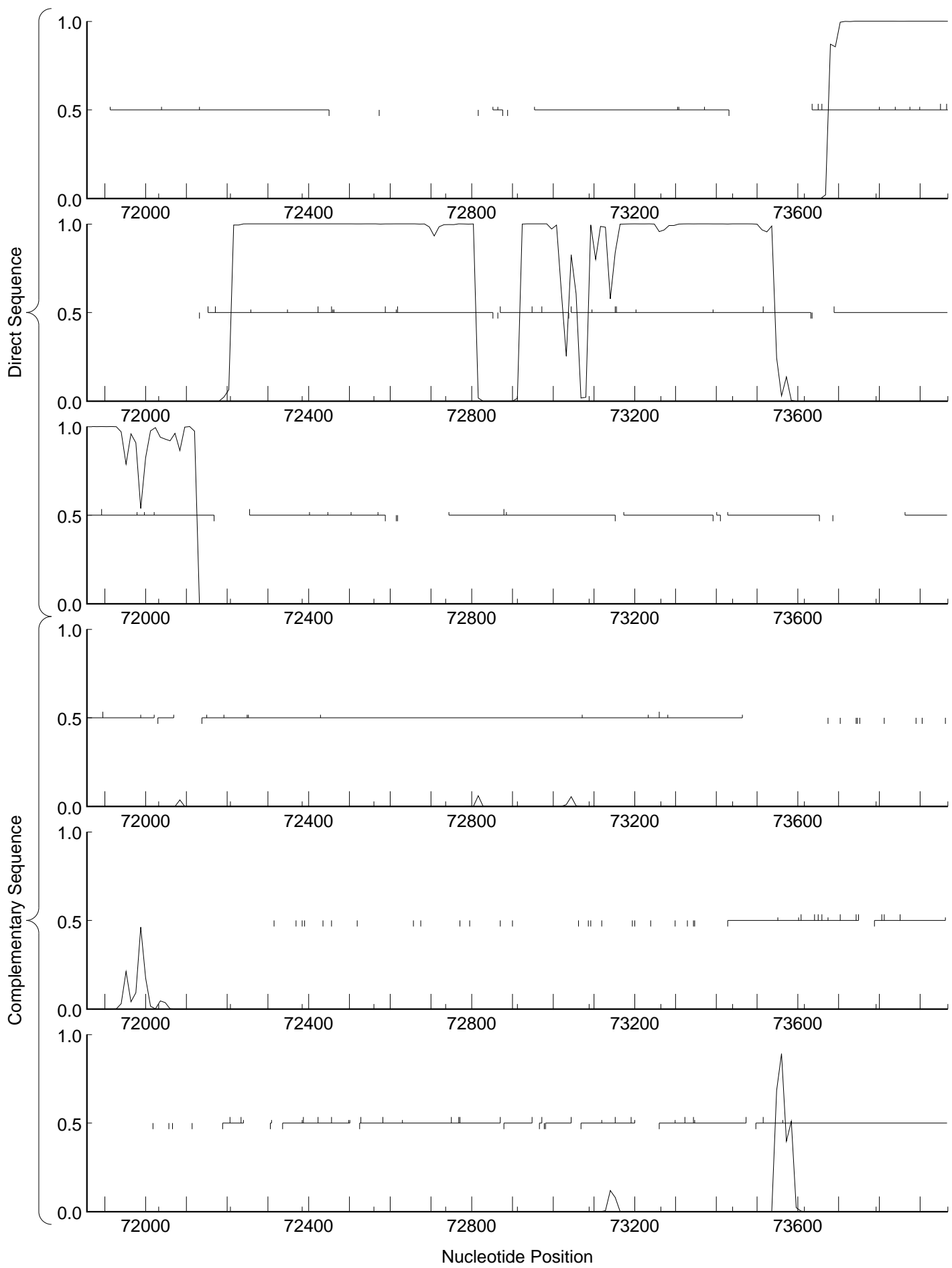


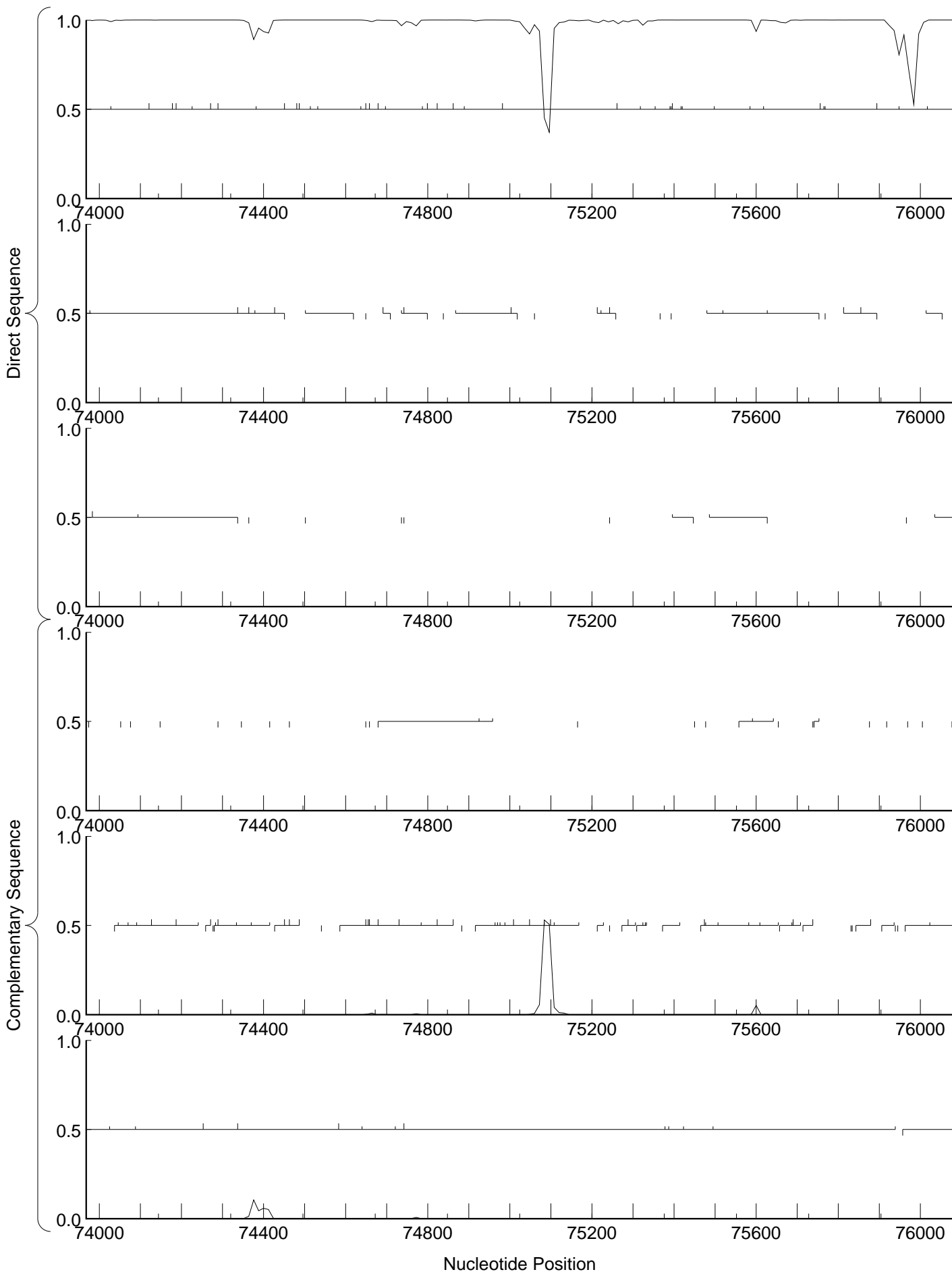


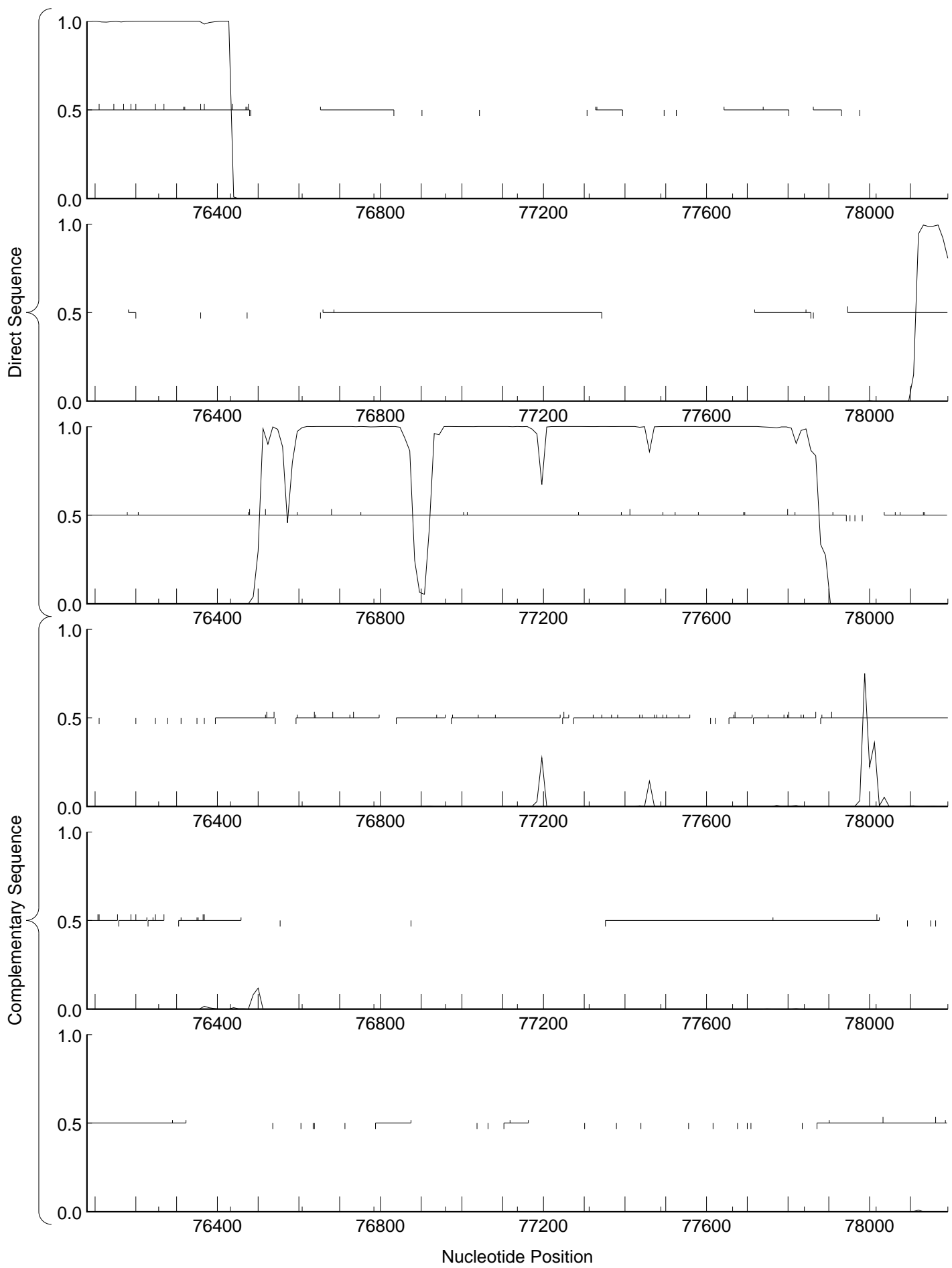


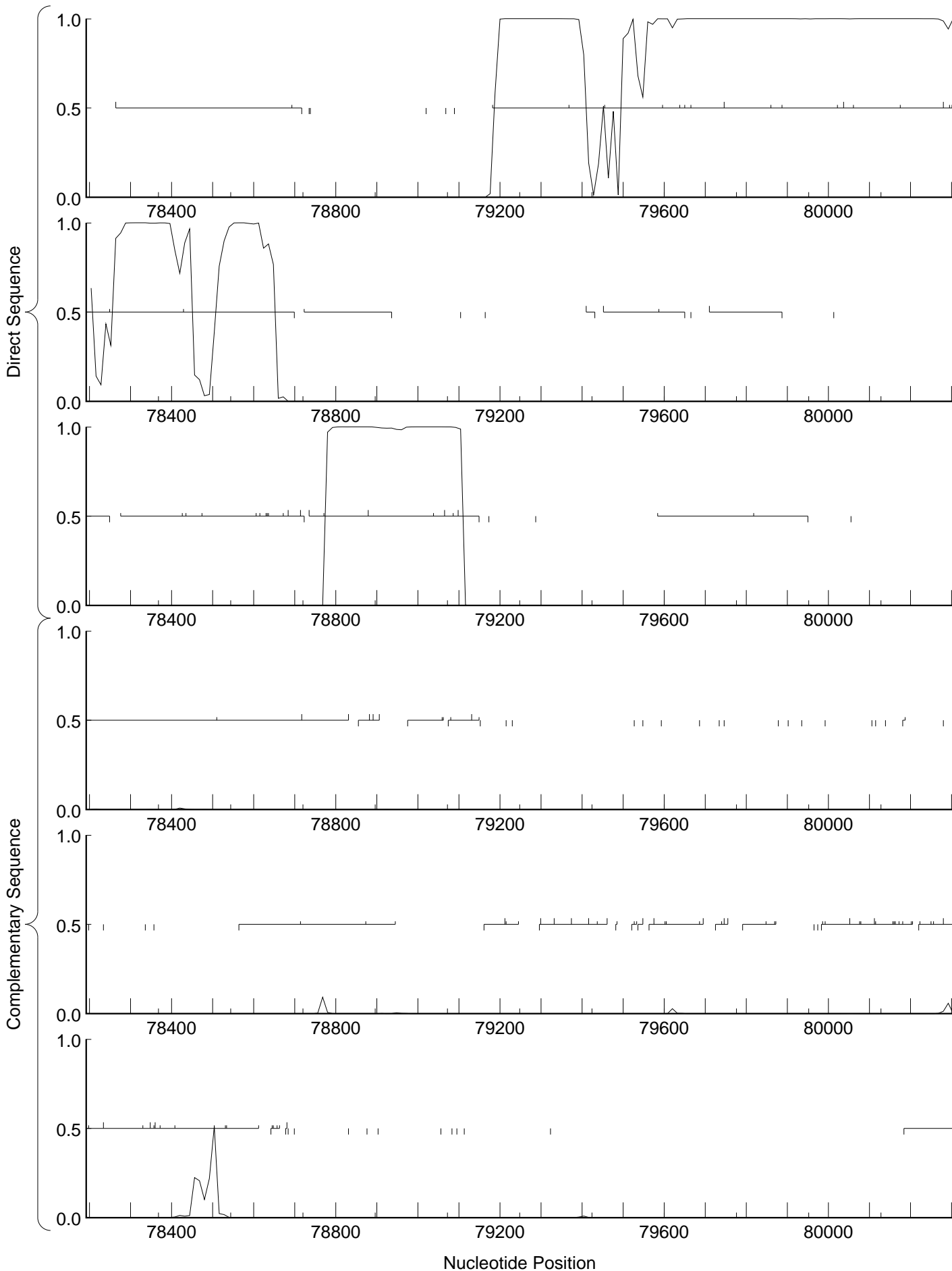


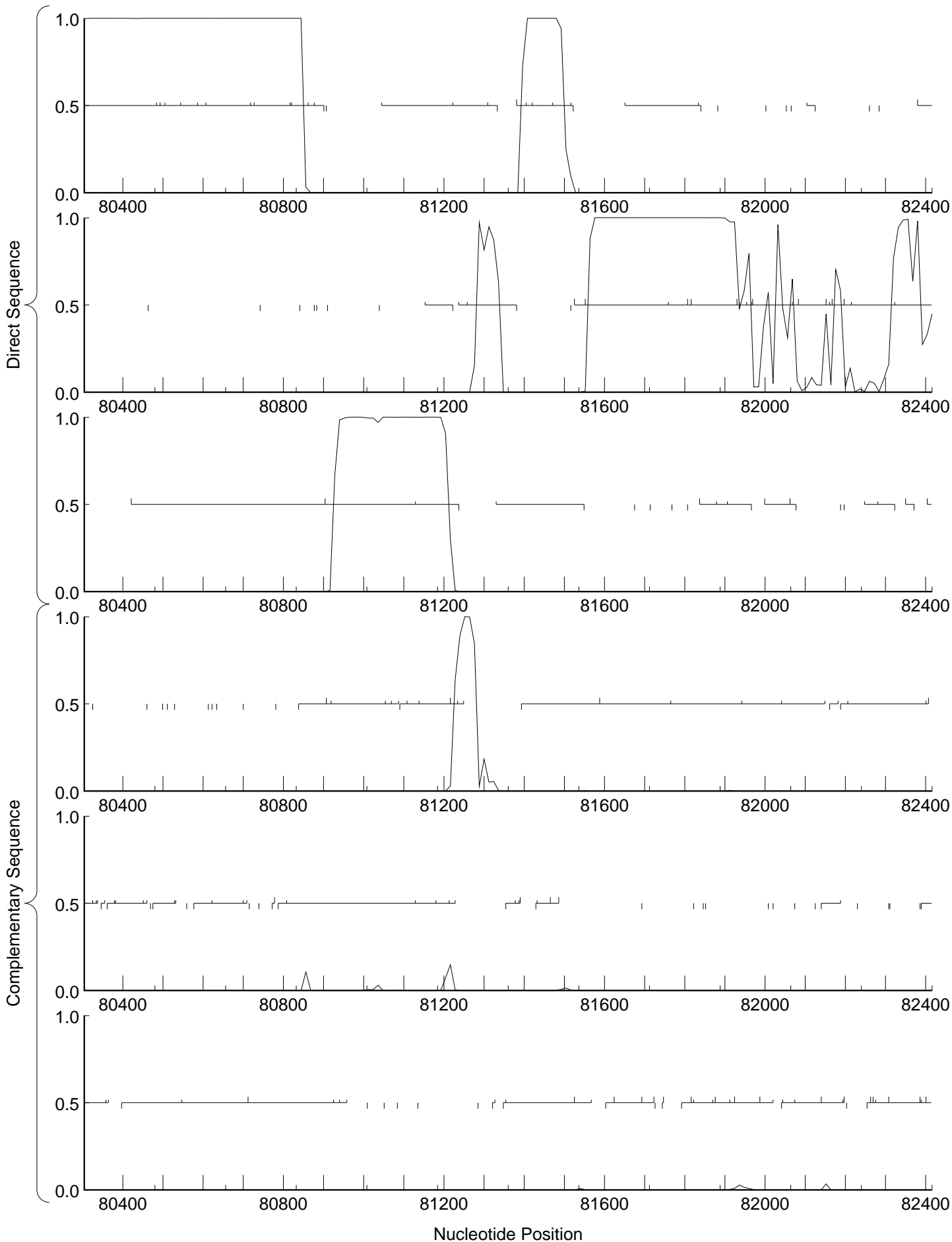


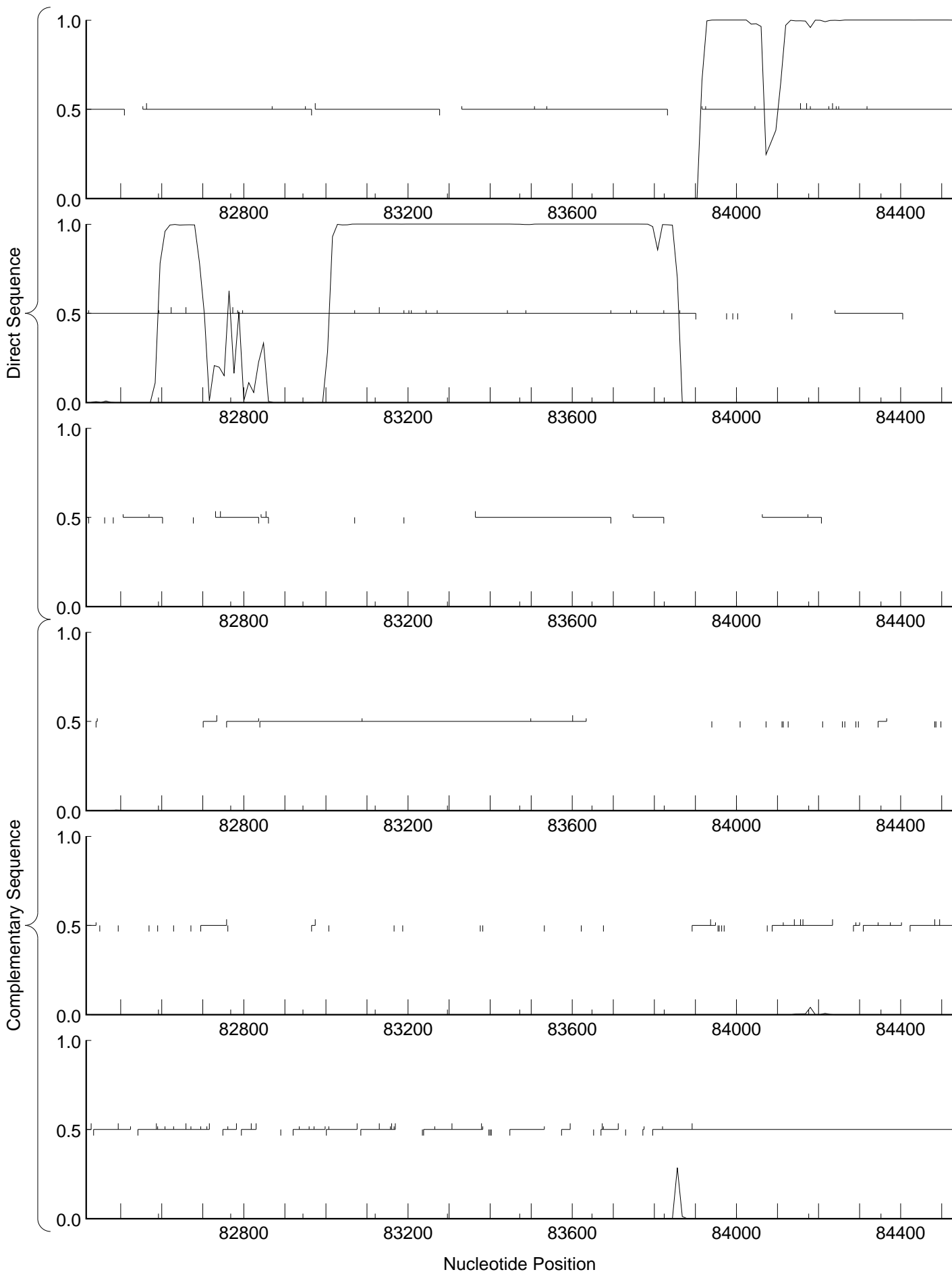


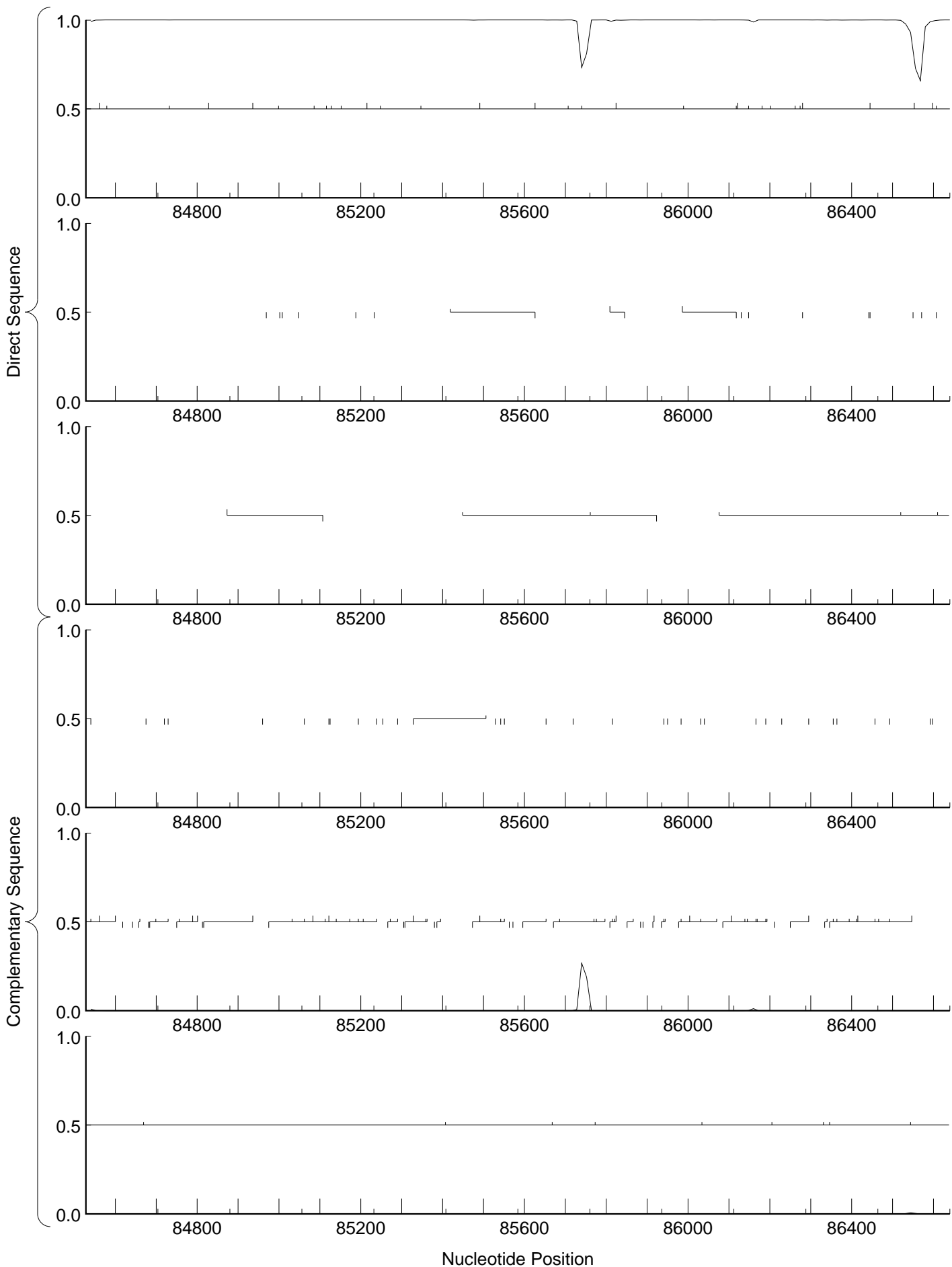


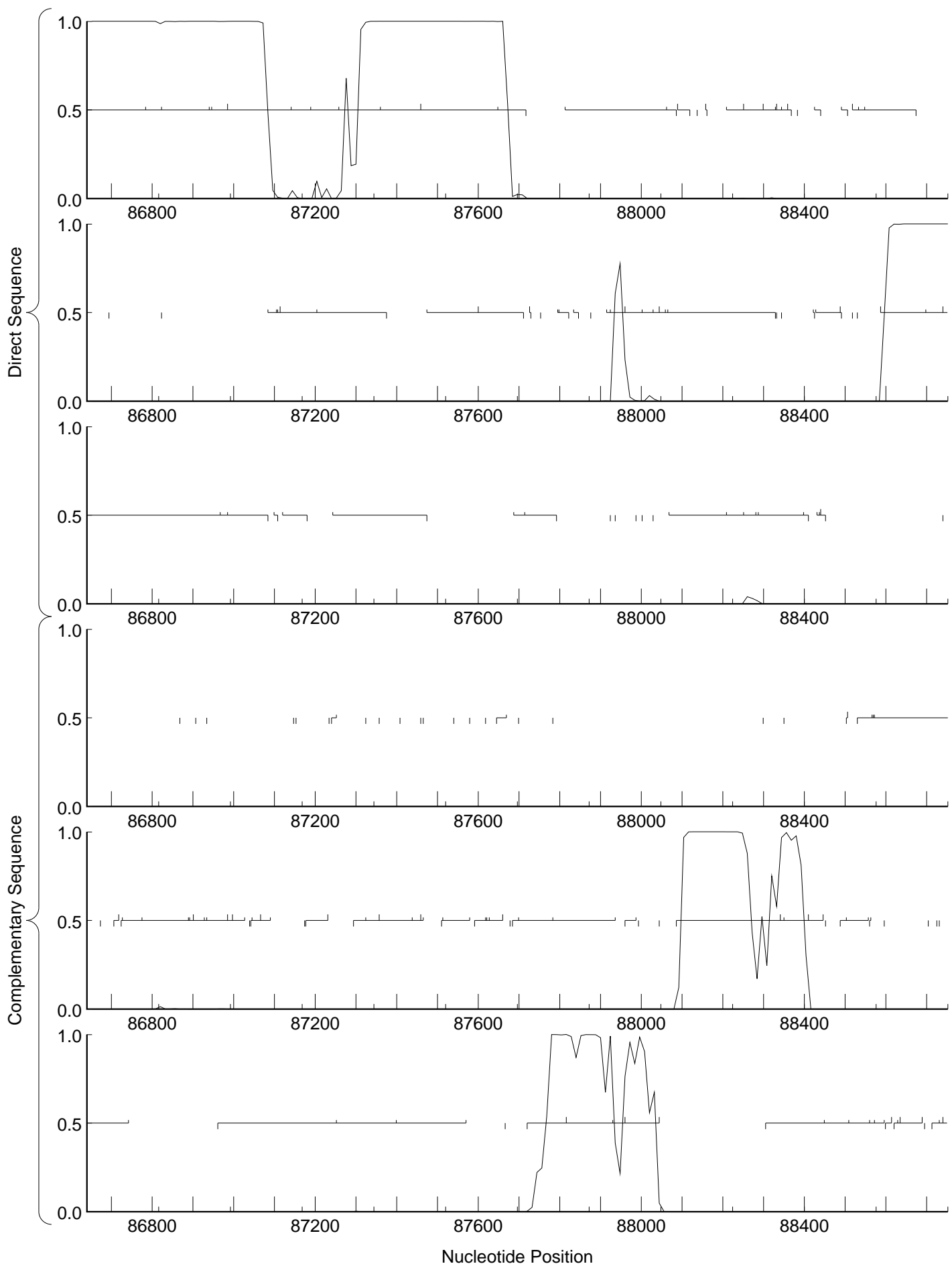


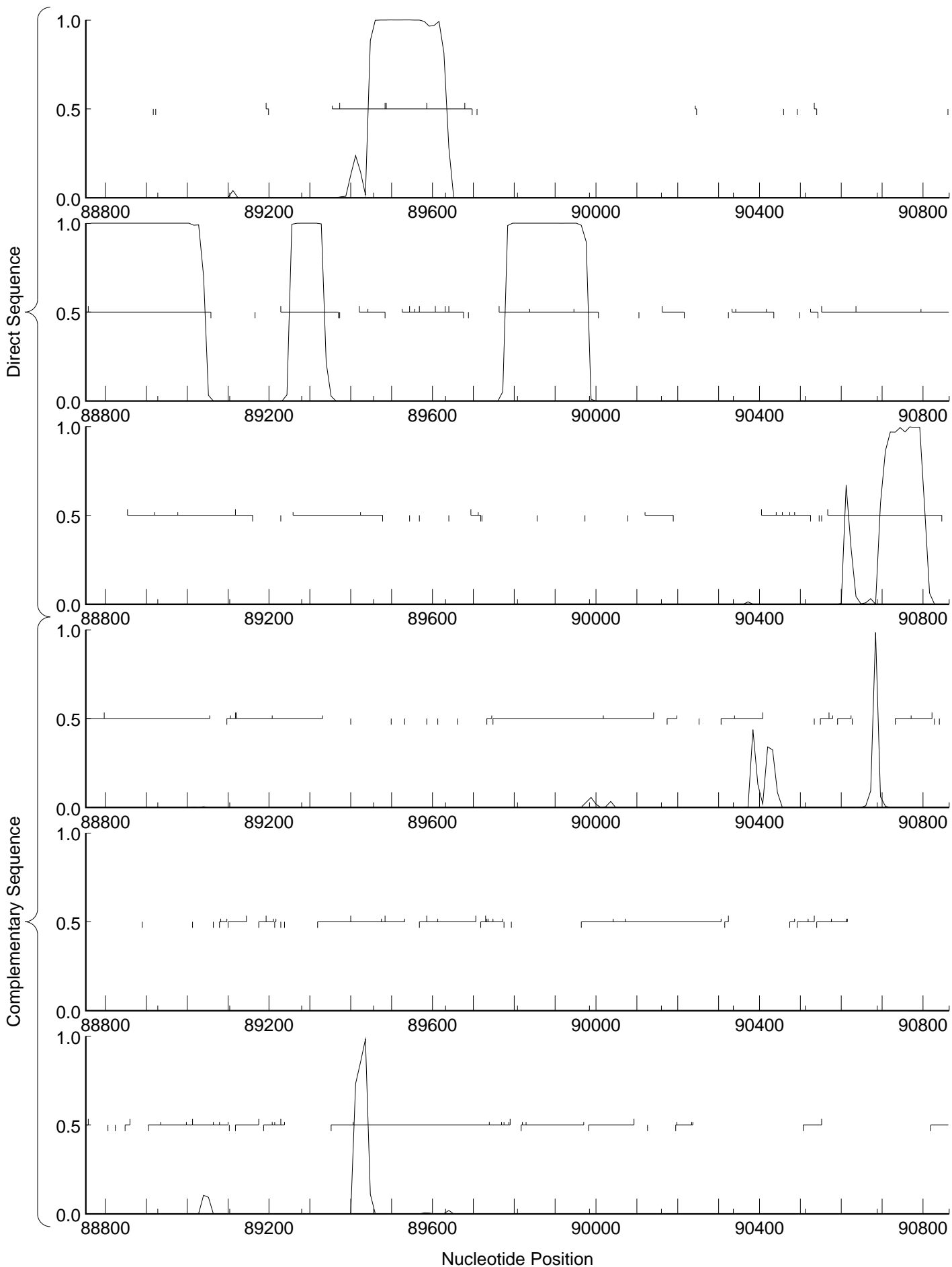


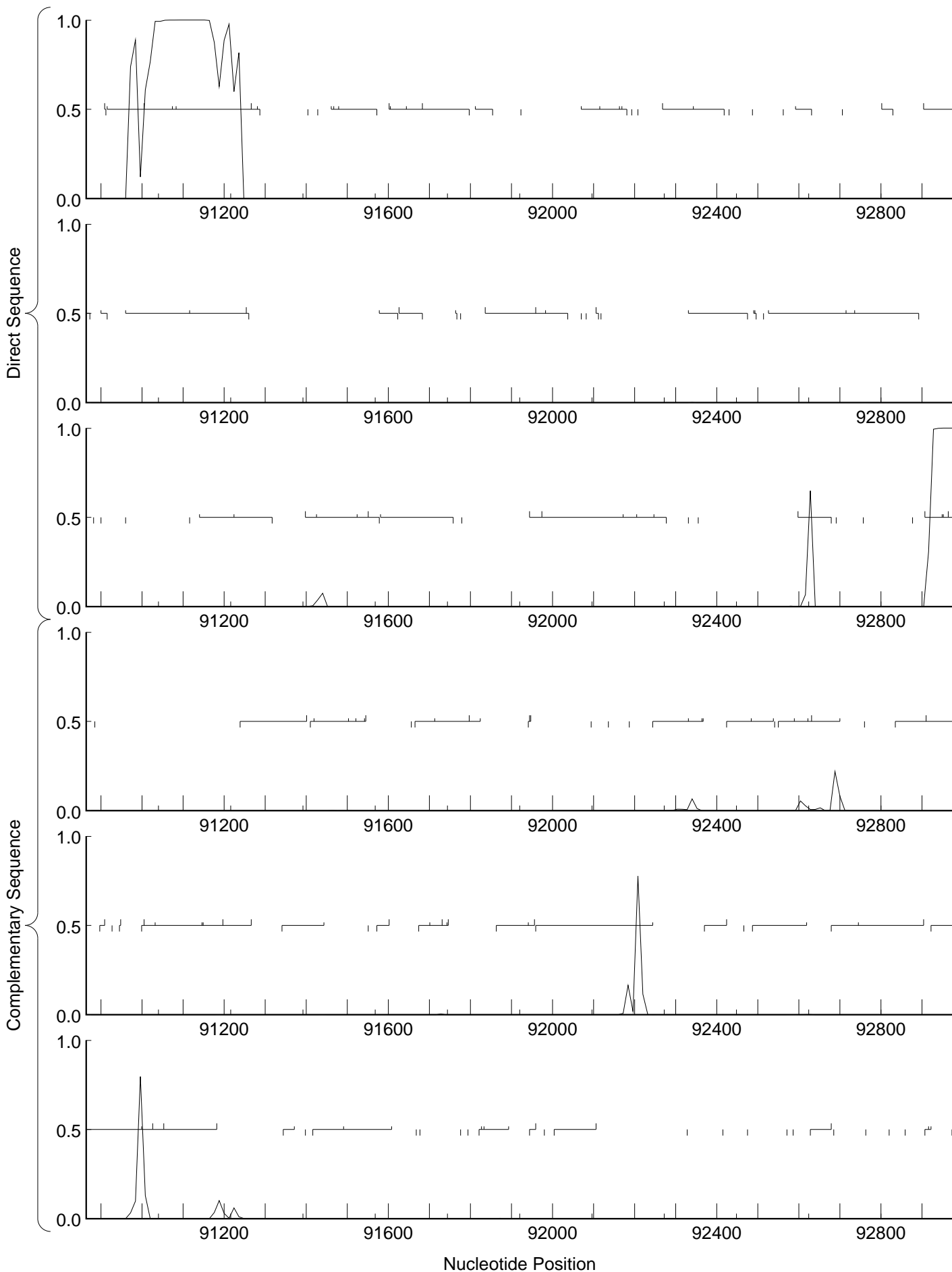


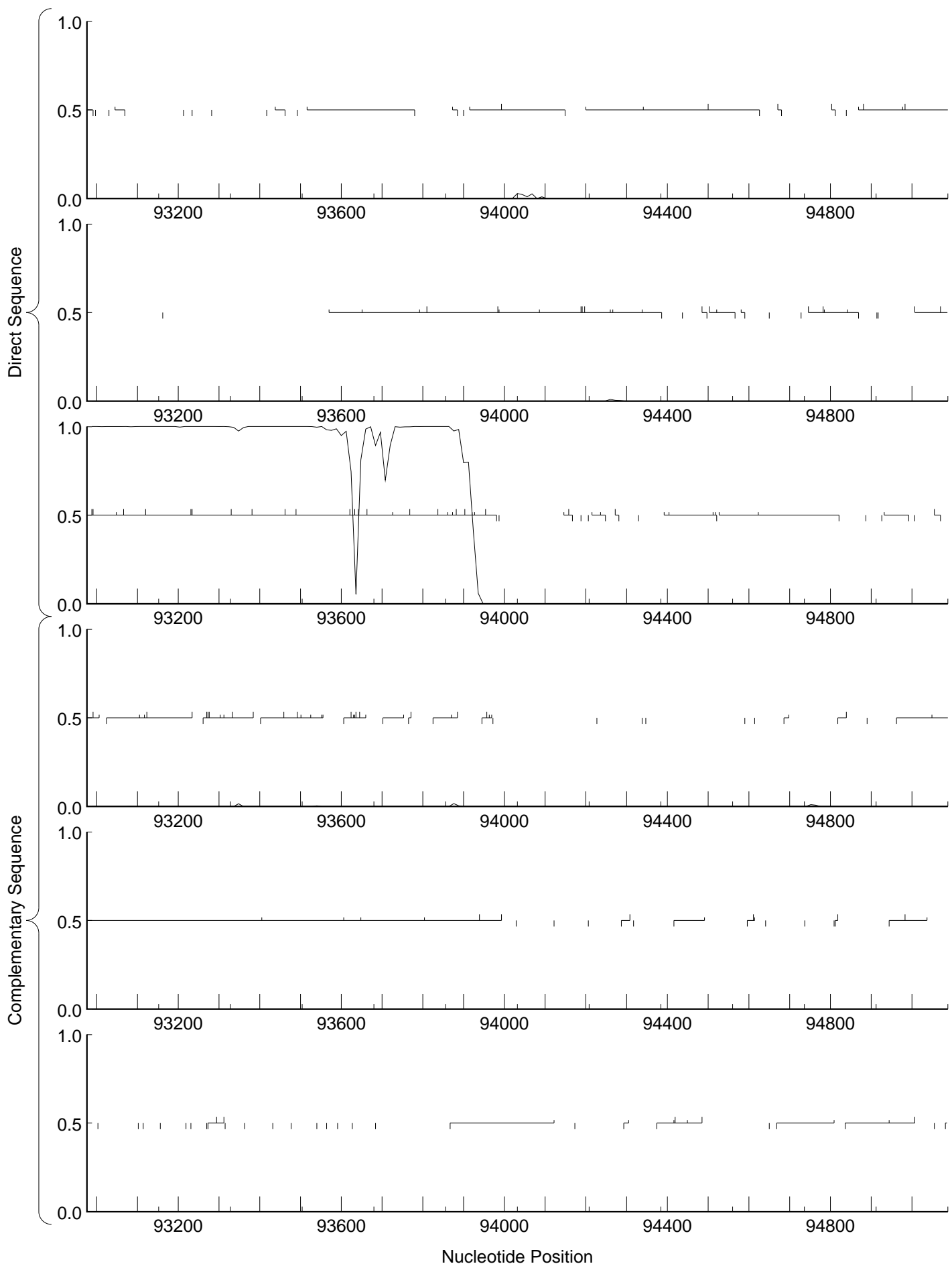




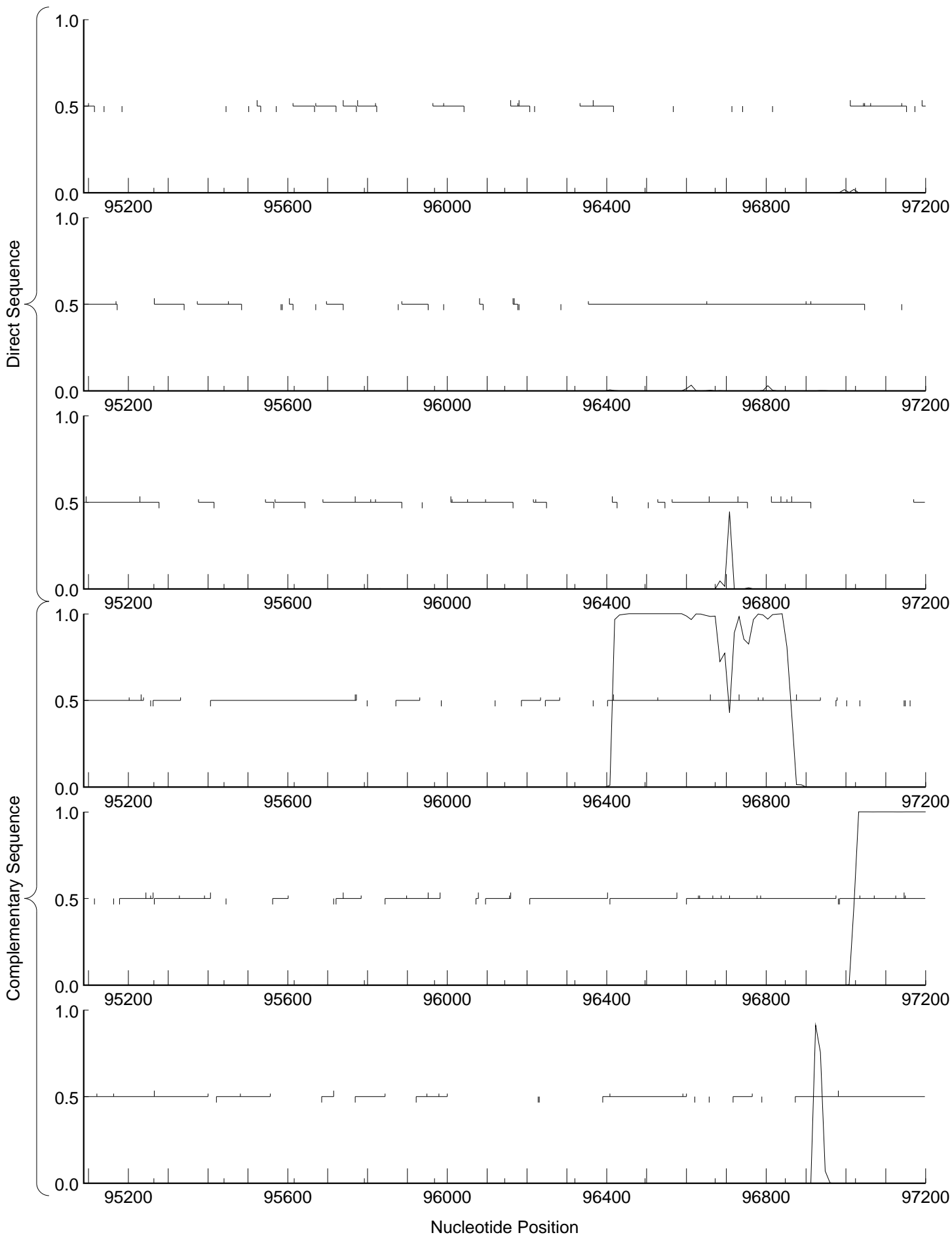


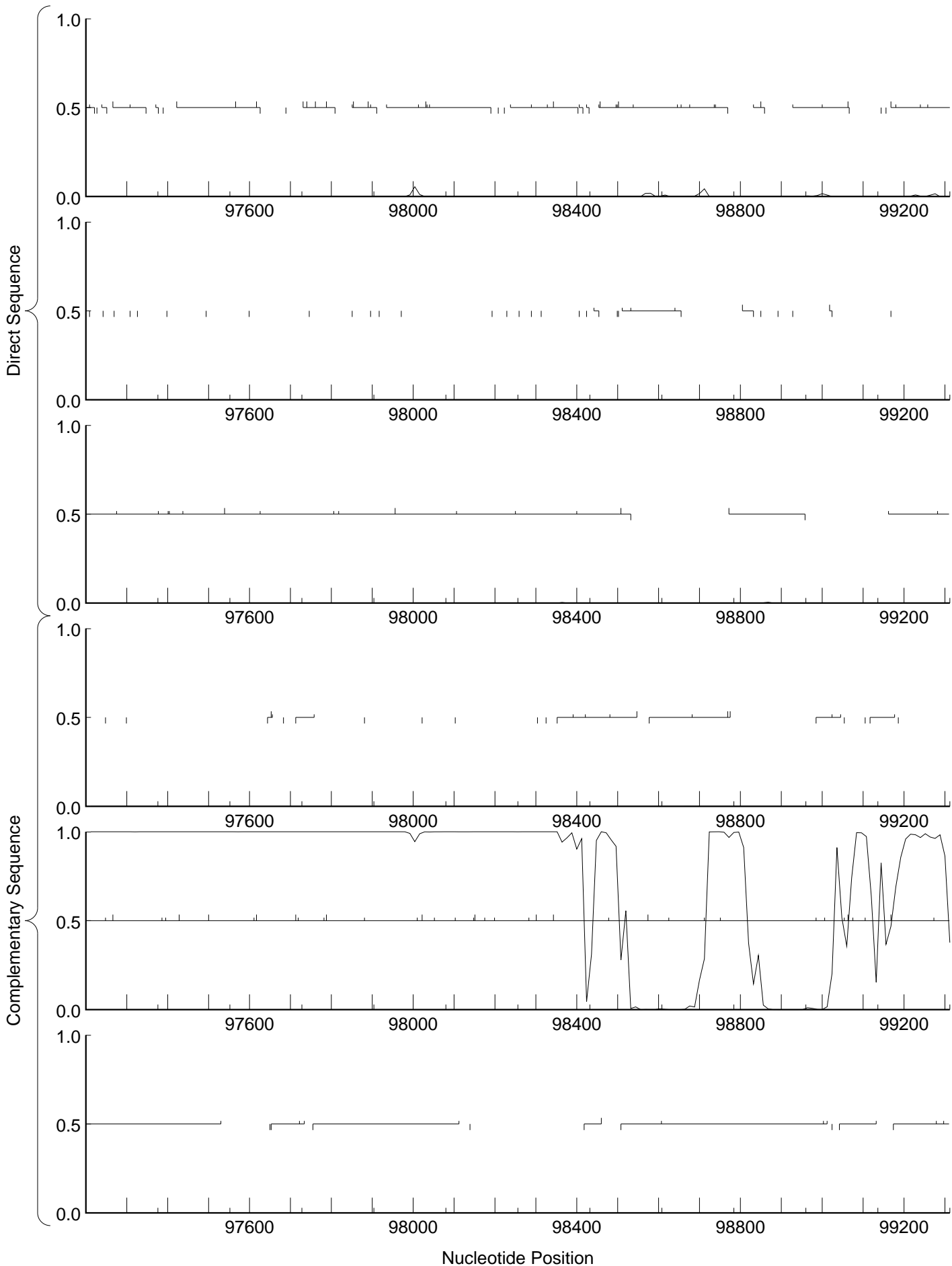


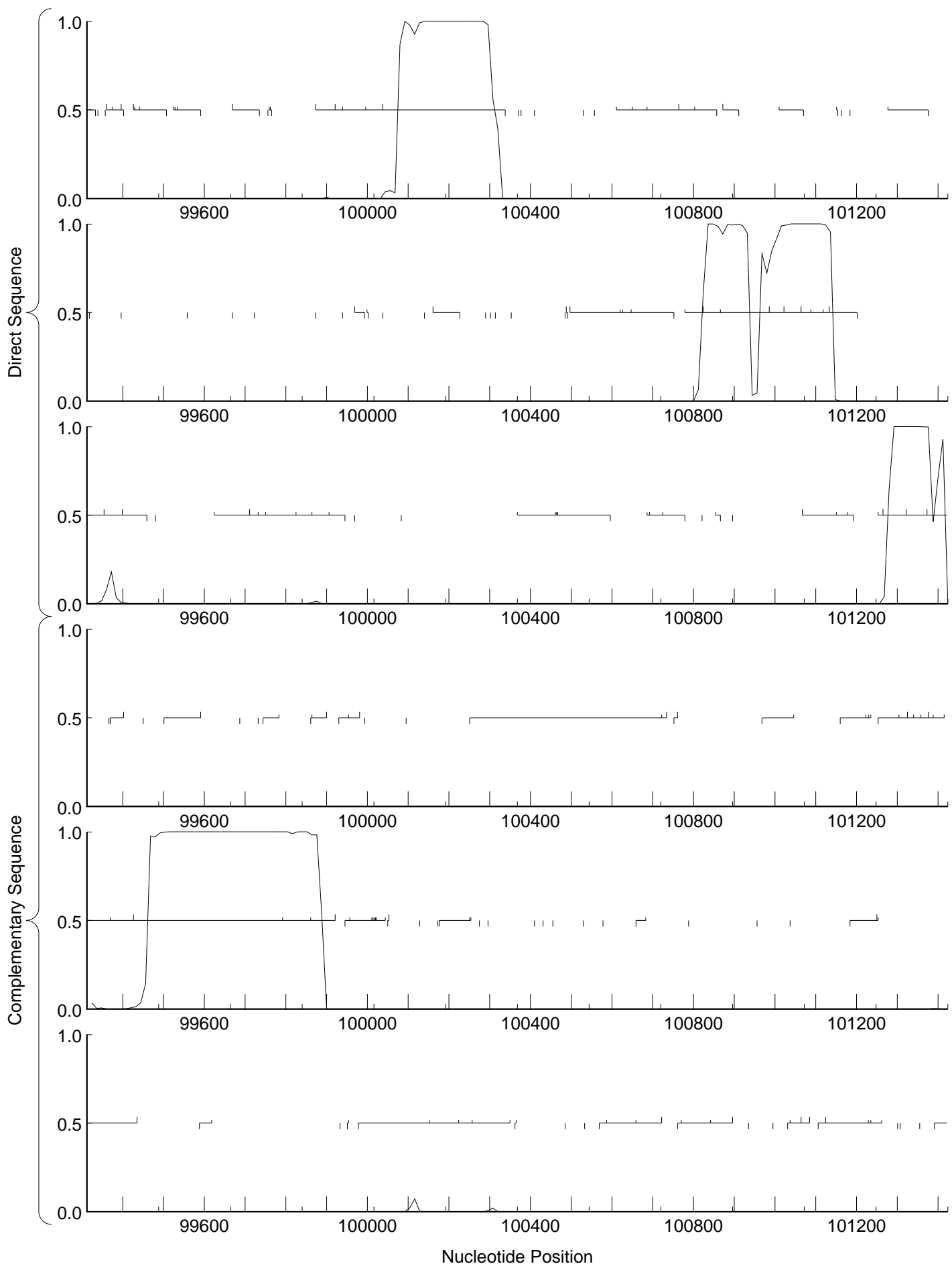


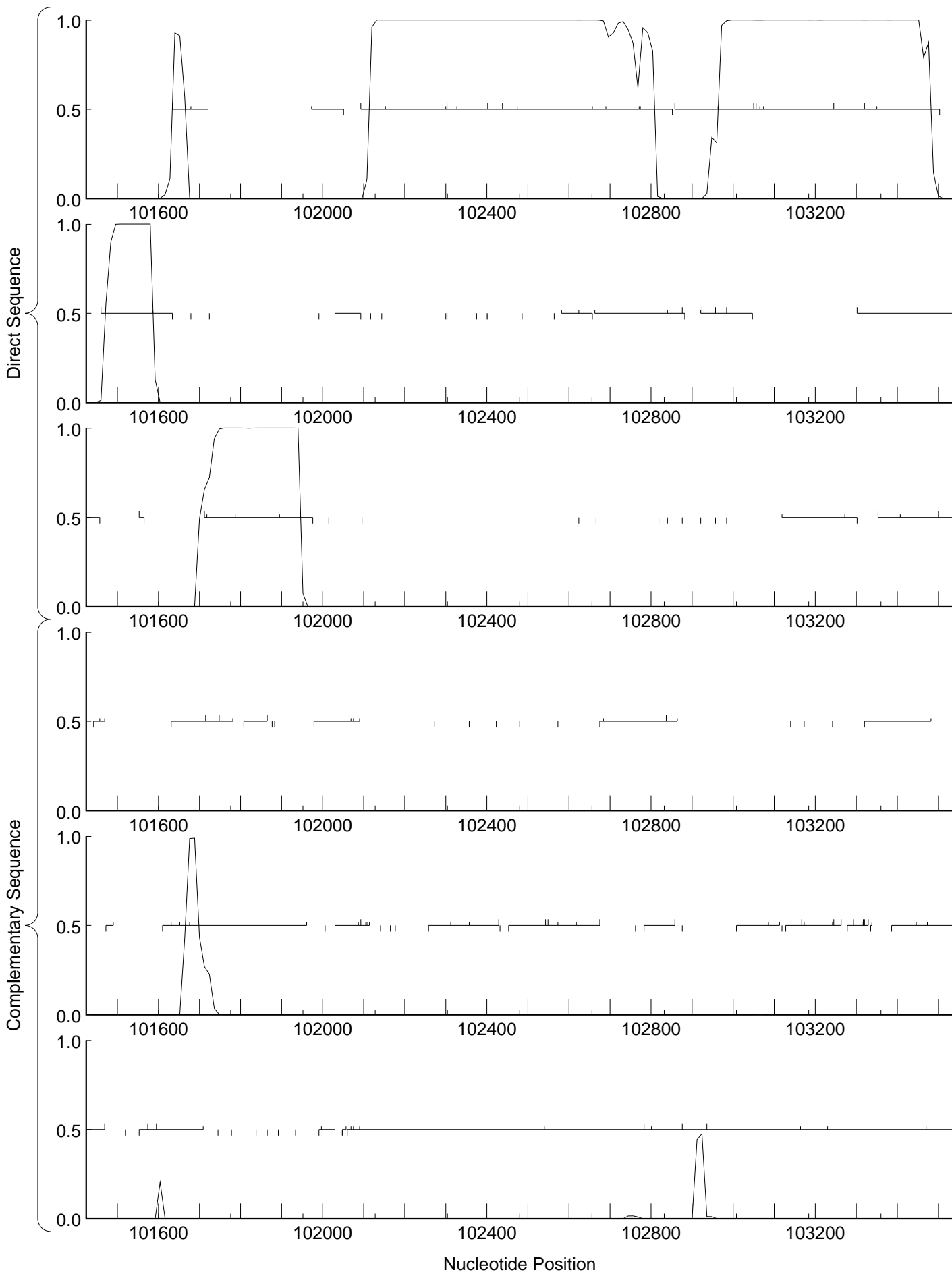


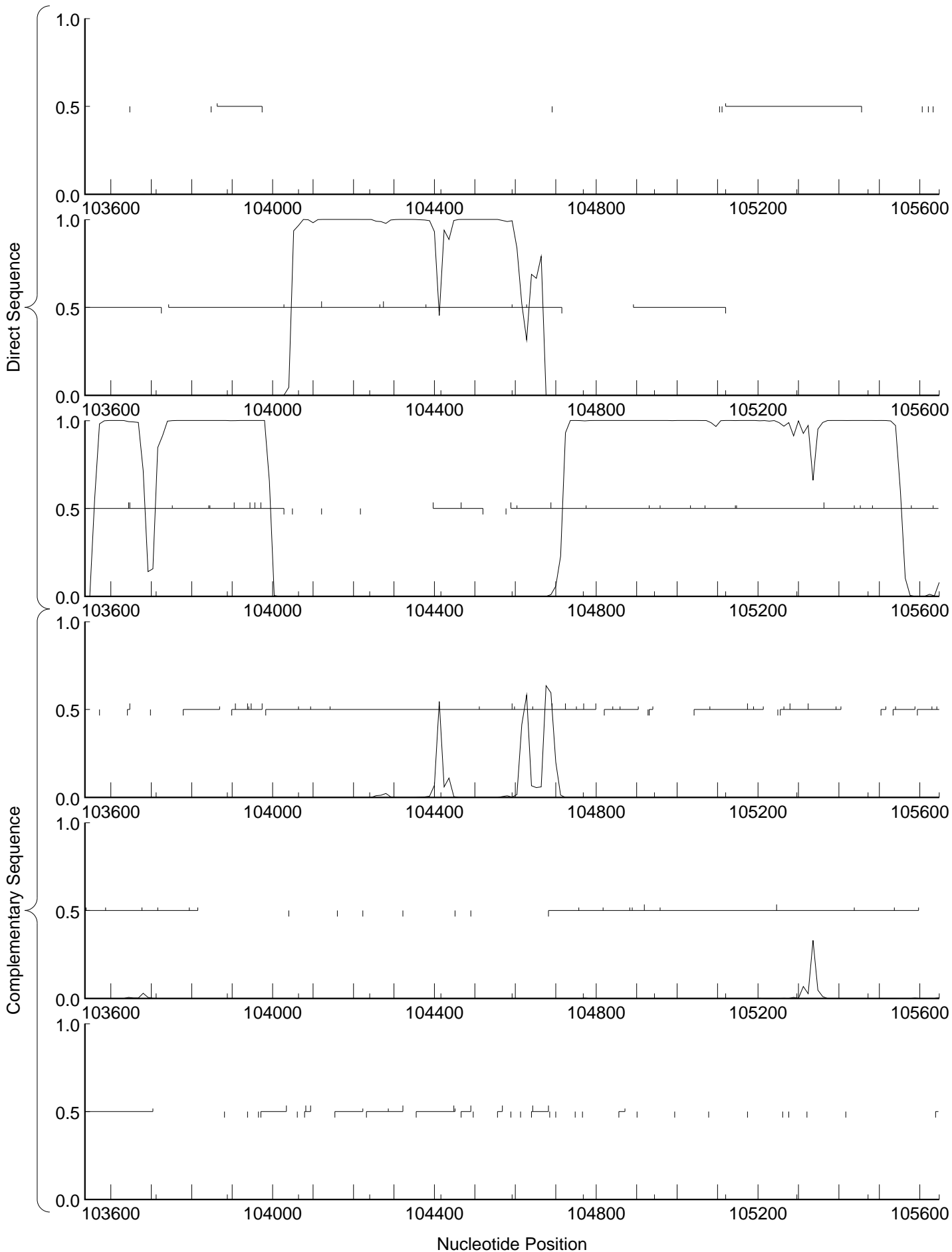
Nucleotide Position

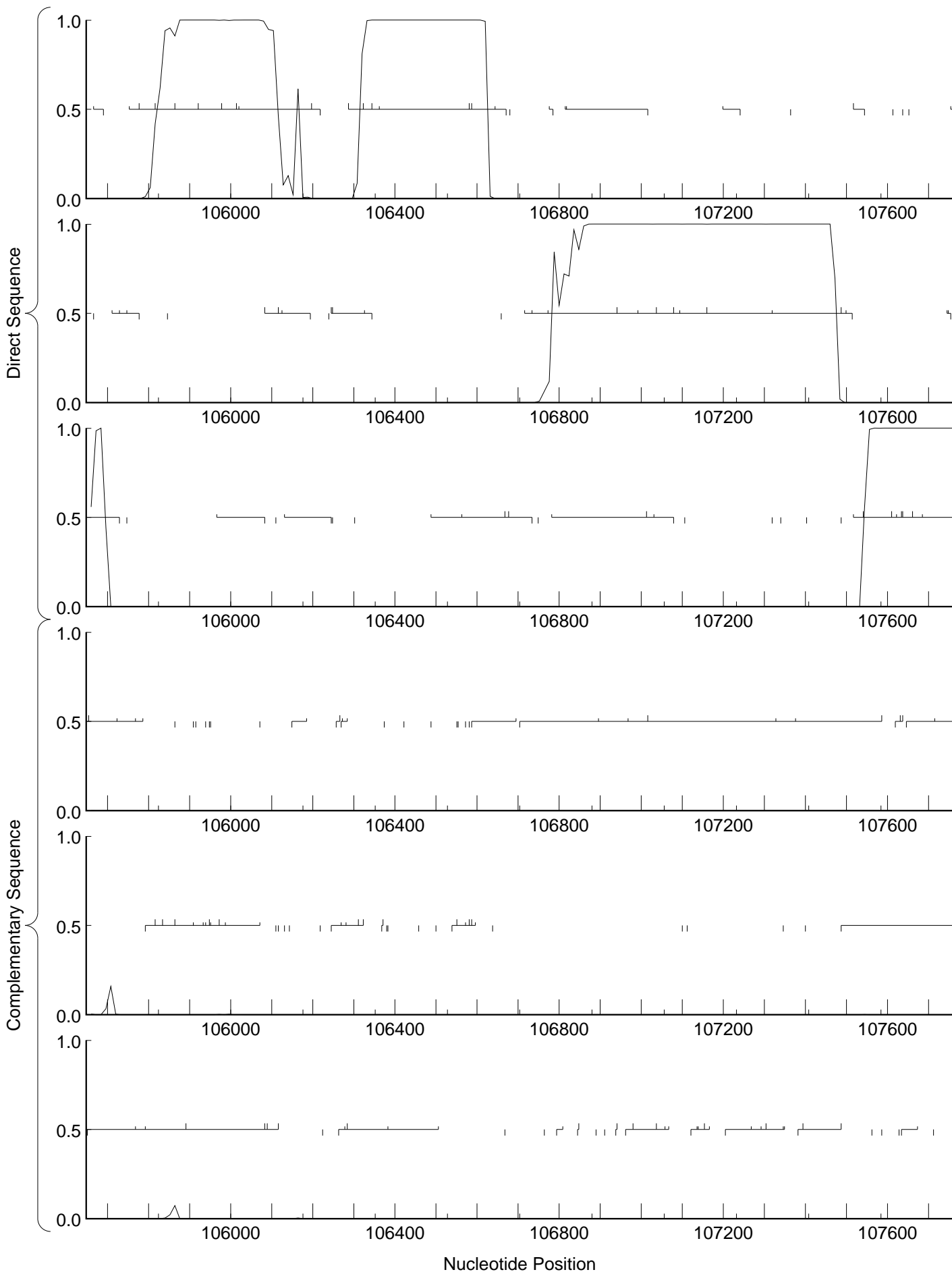


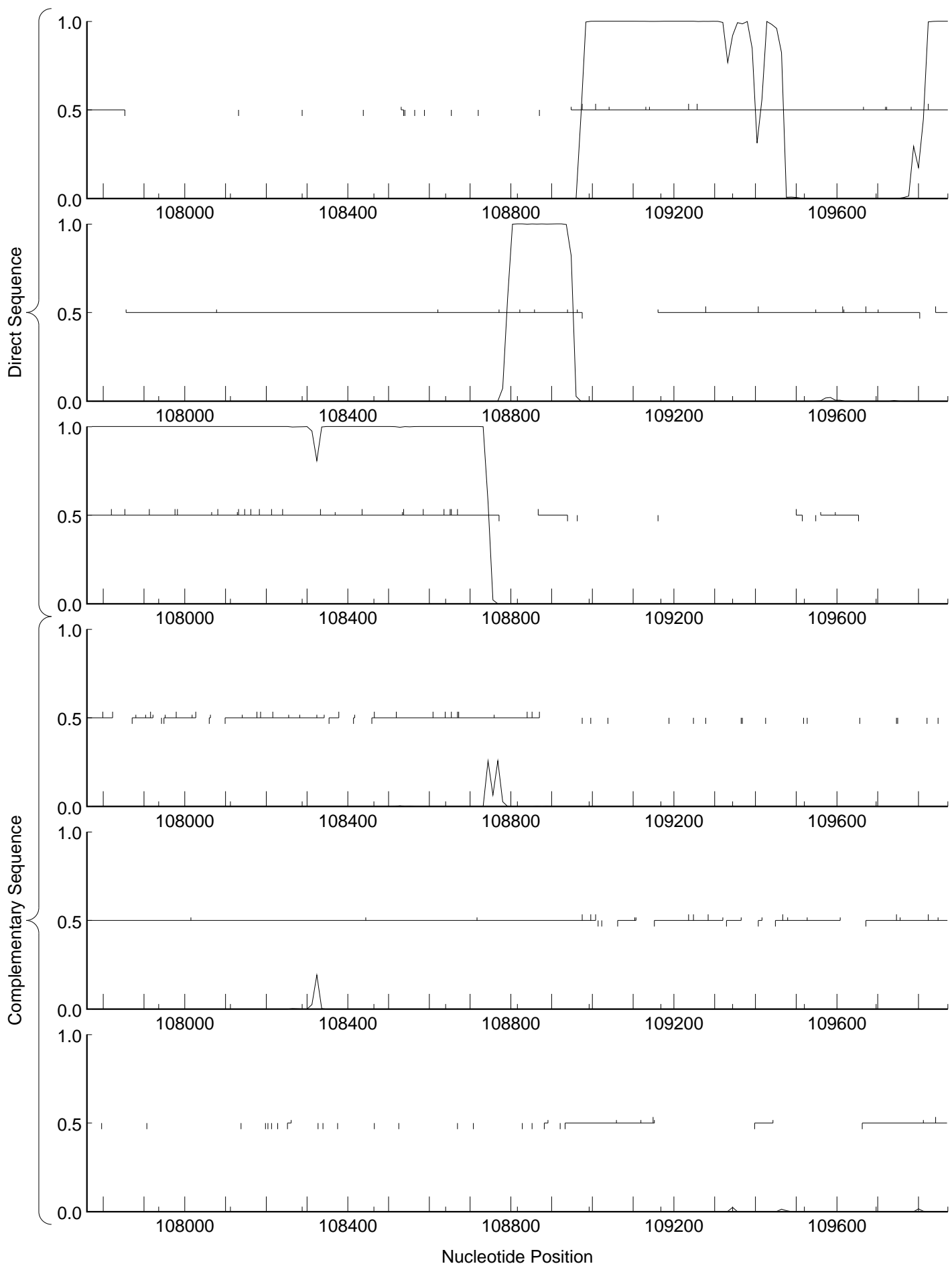


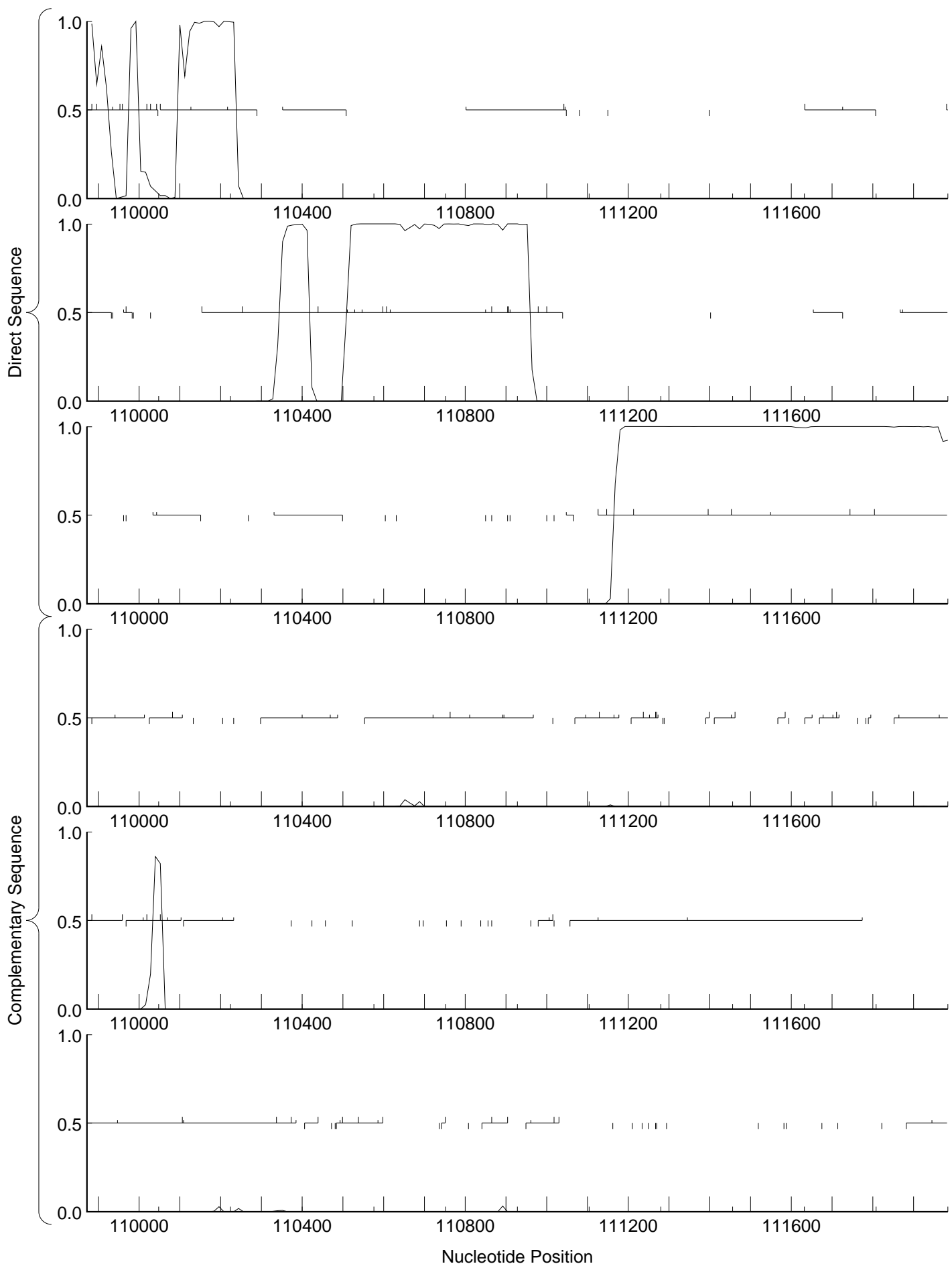


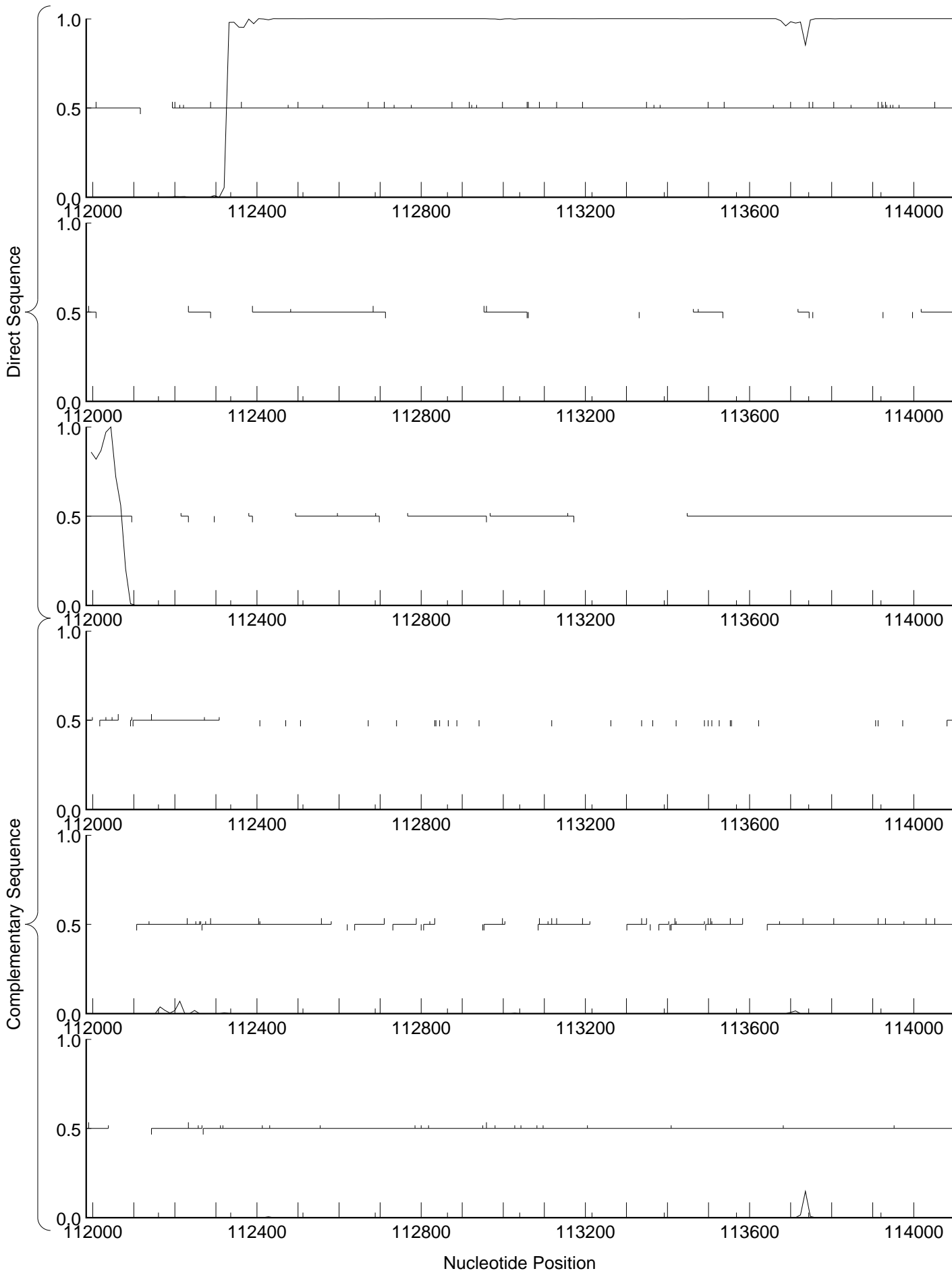


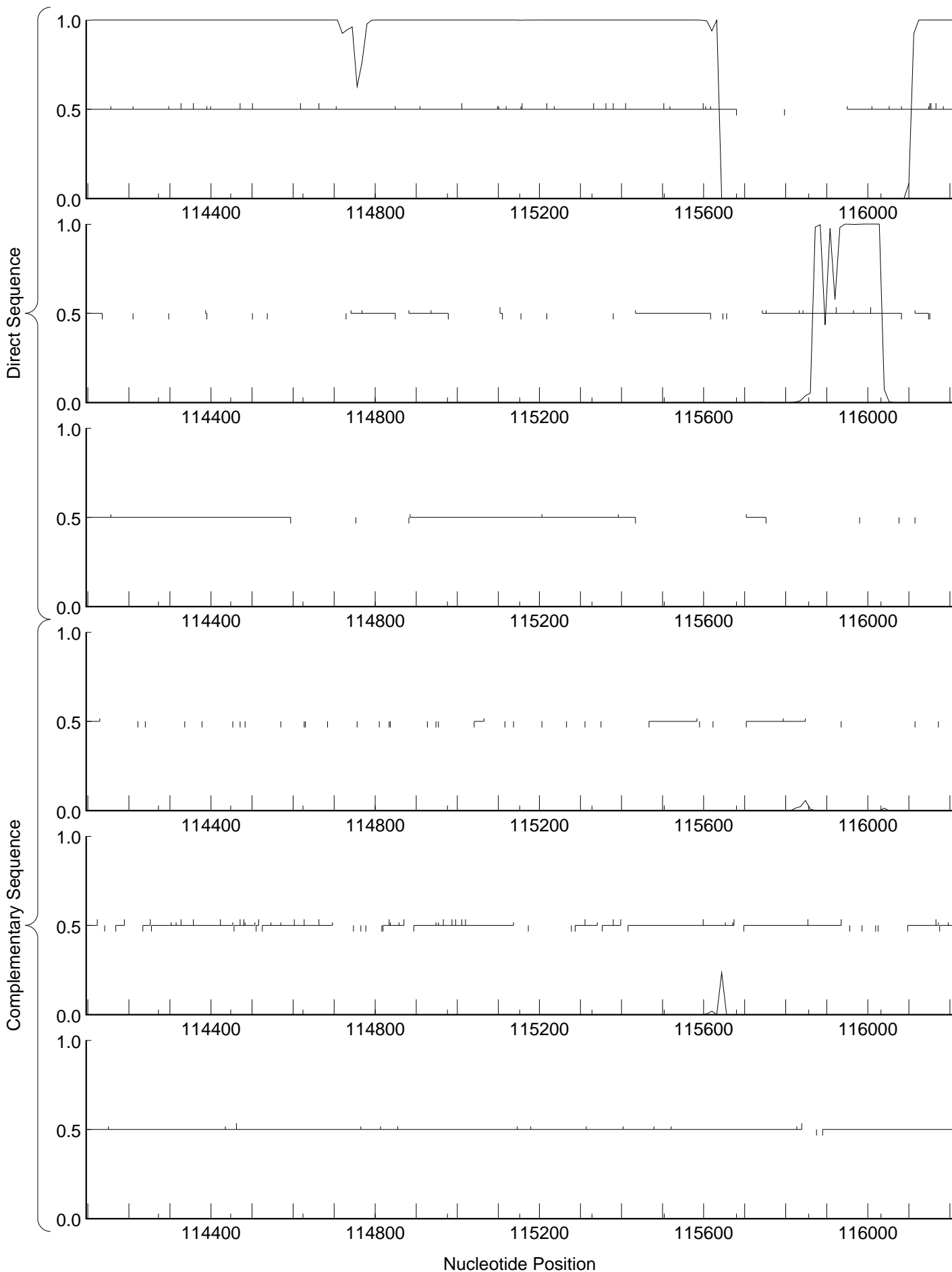


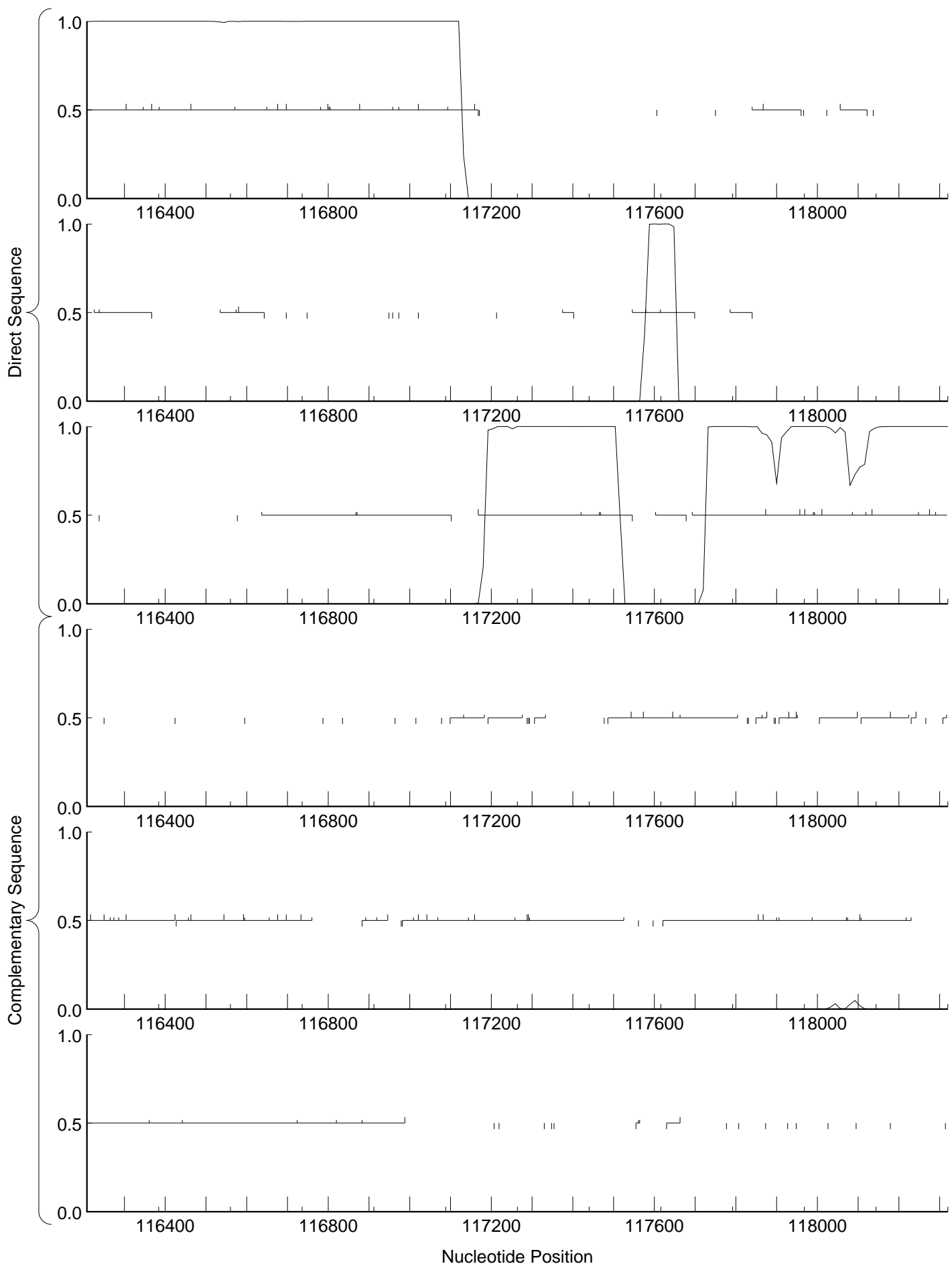


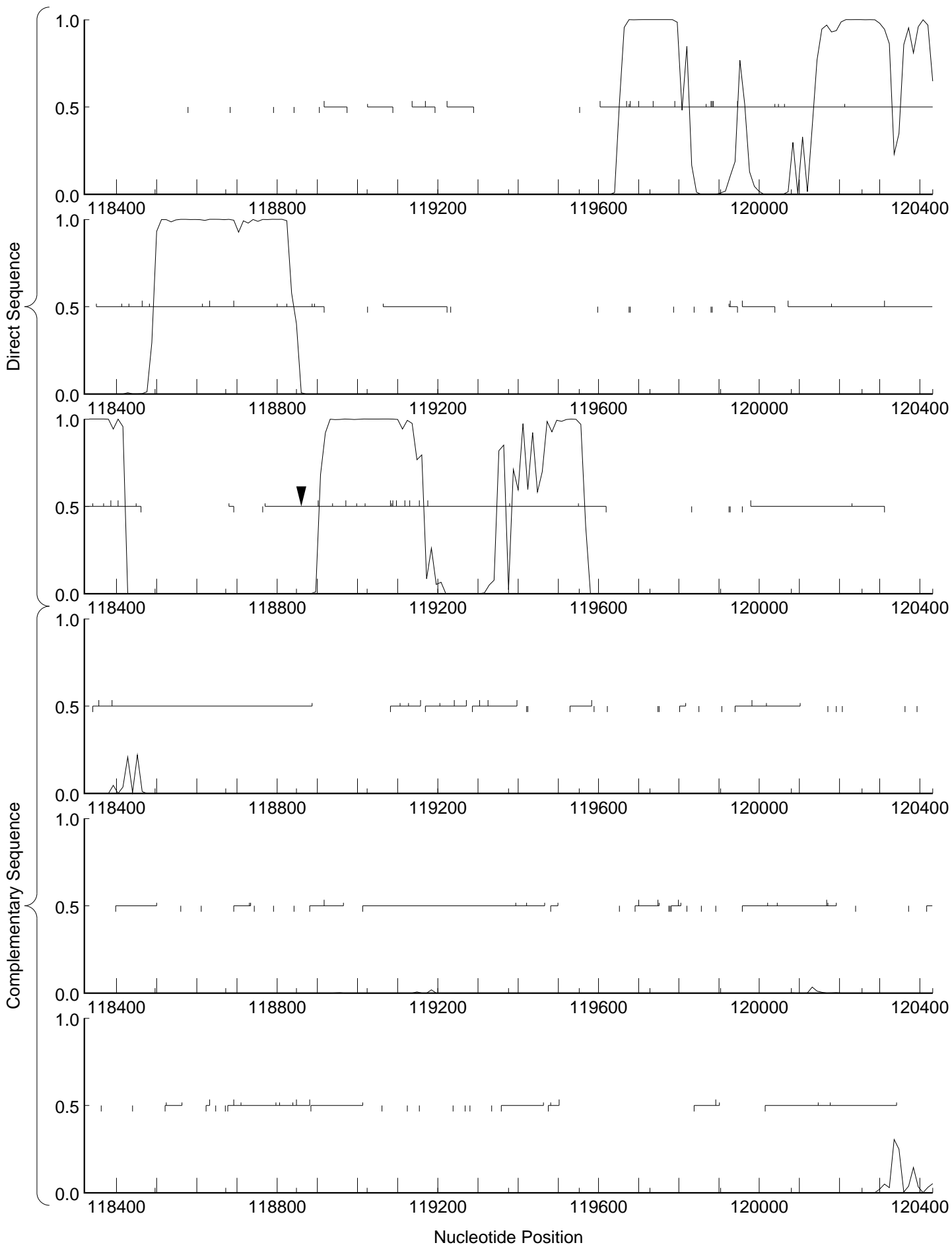


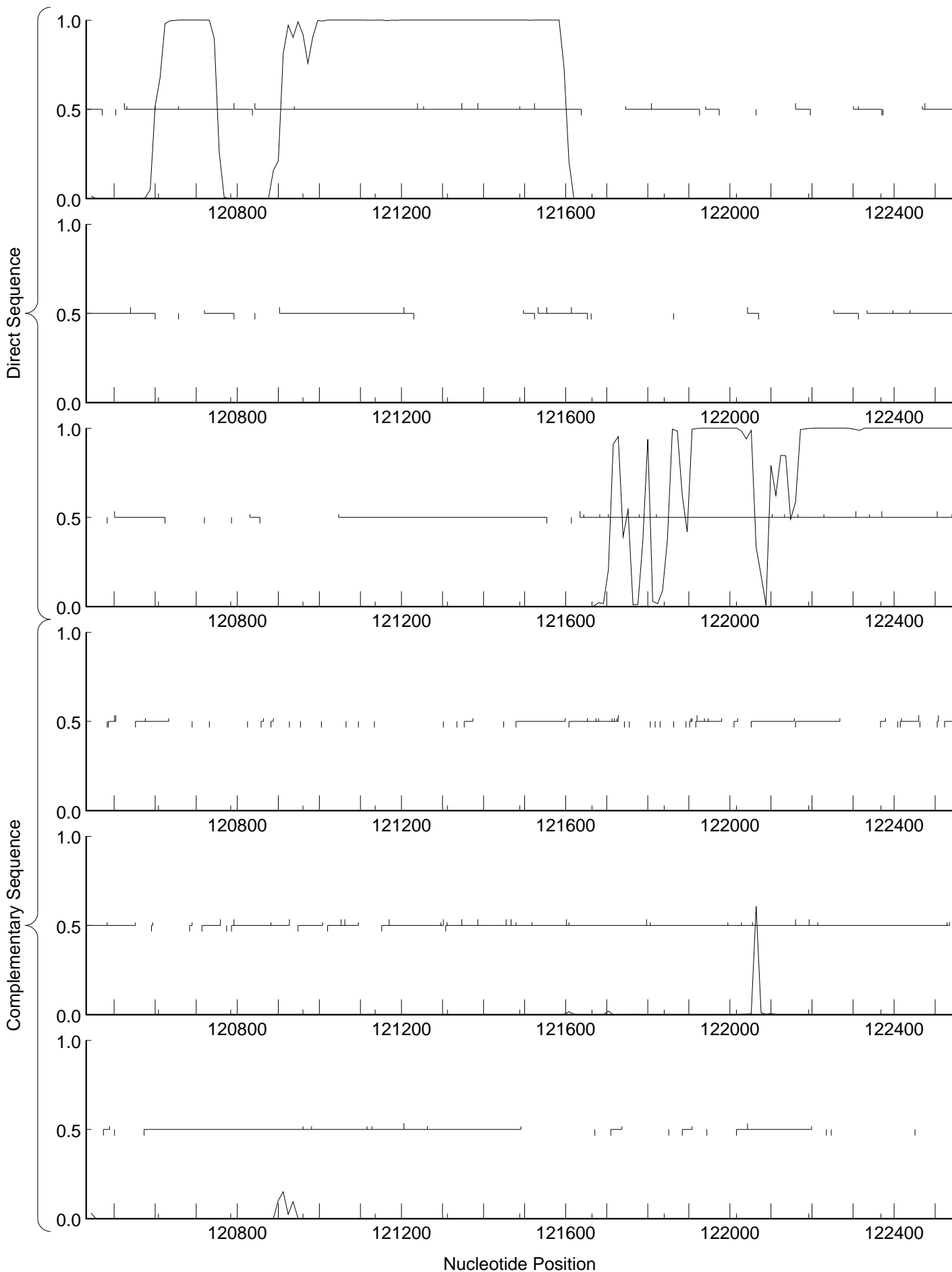


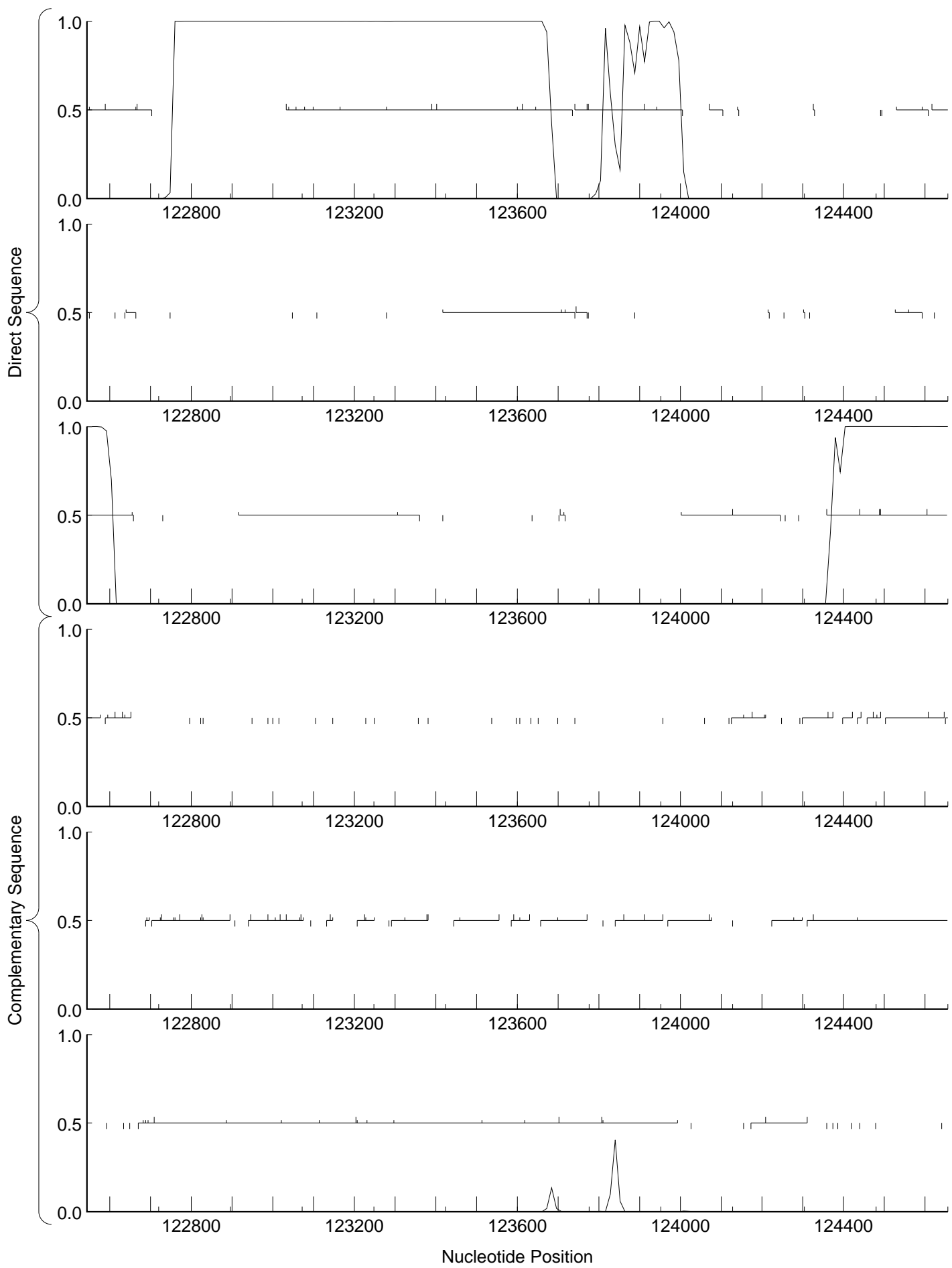


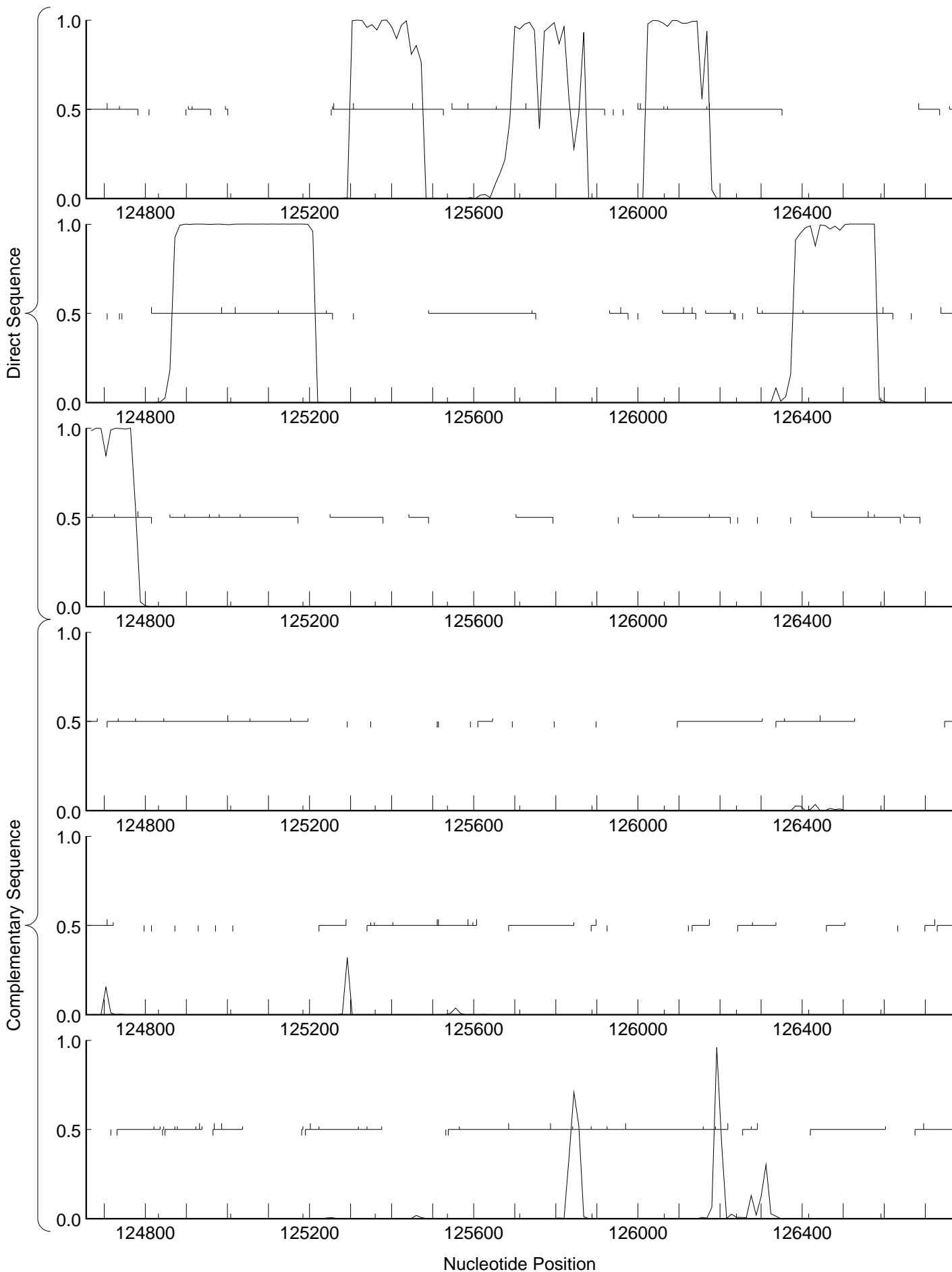


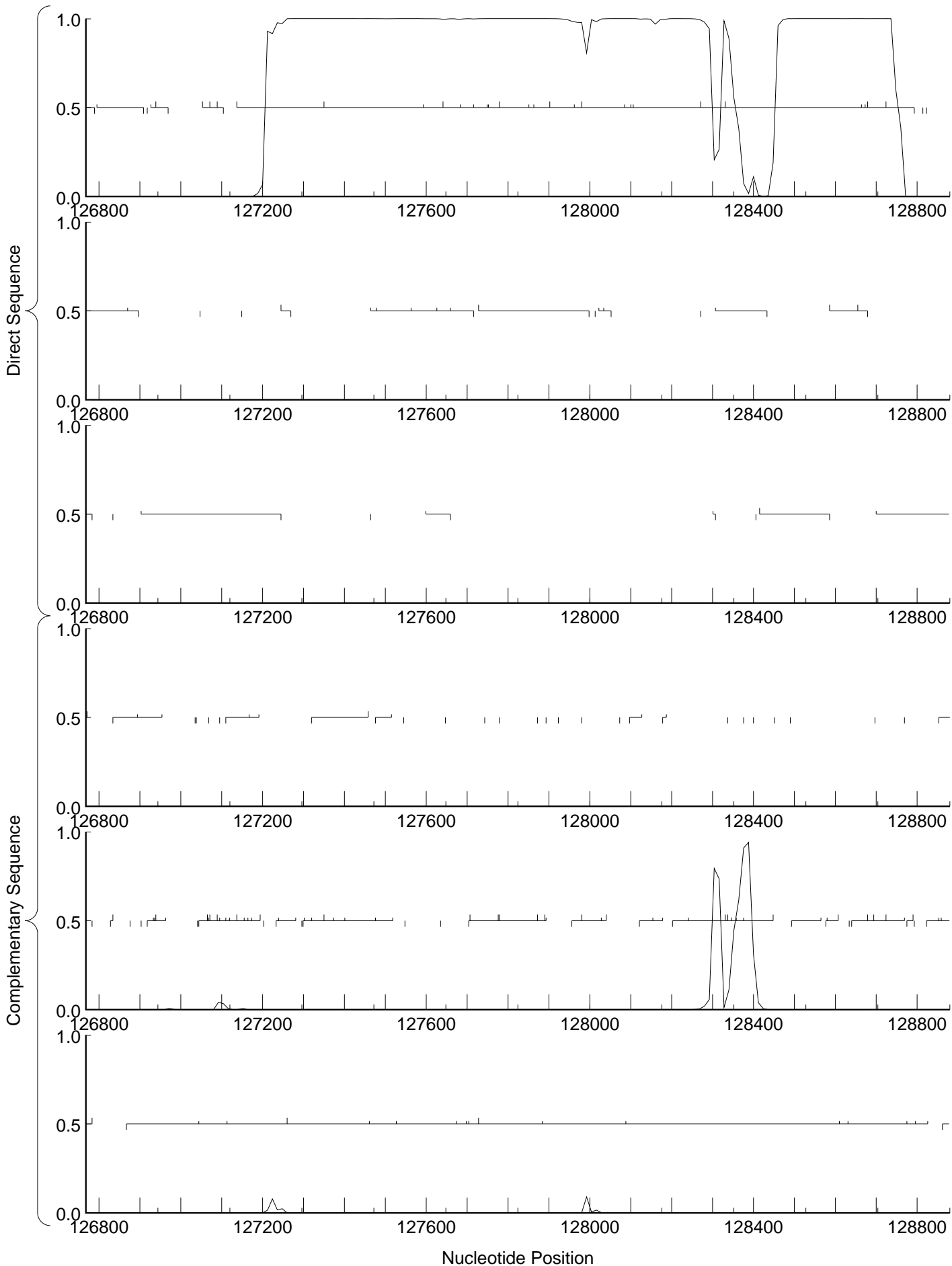


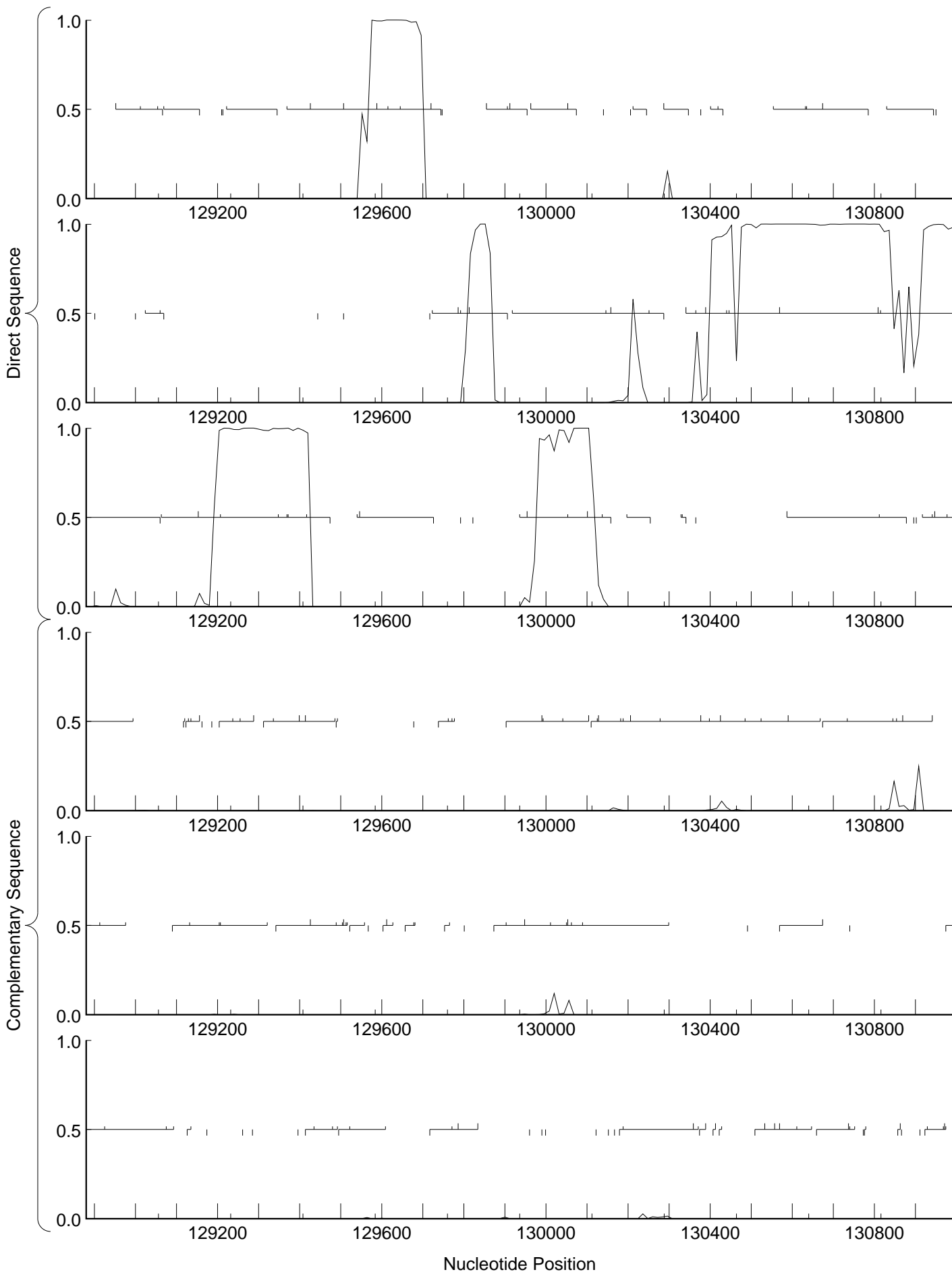


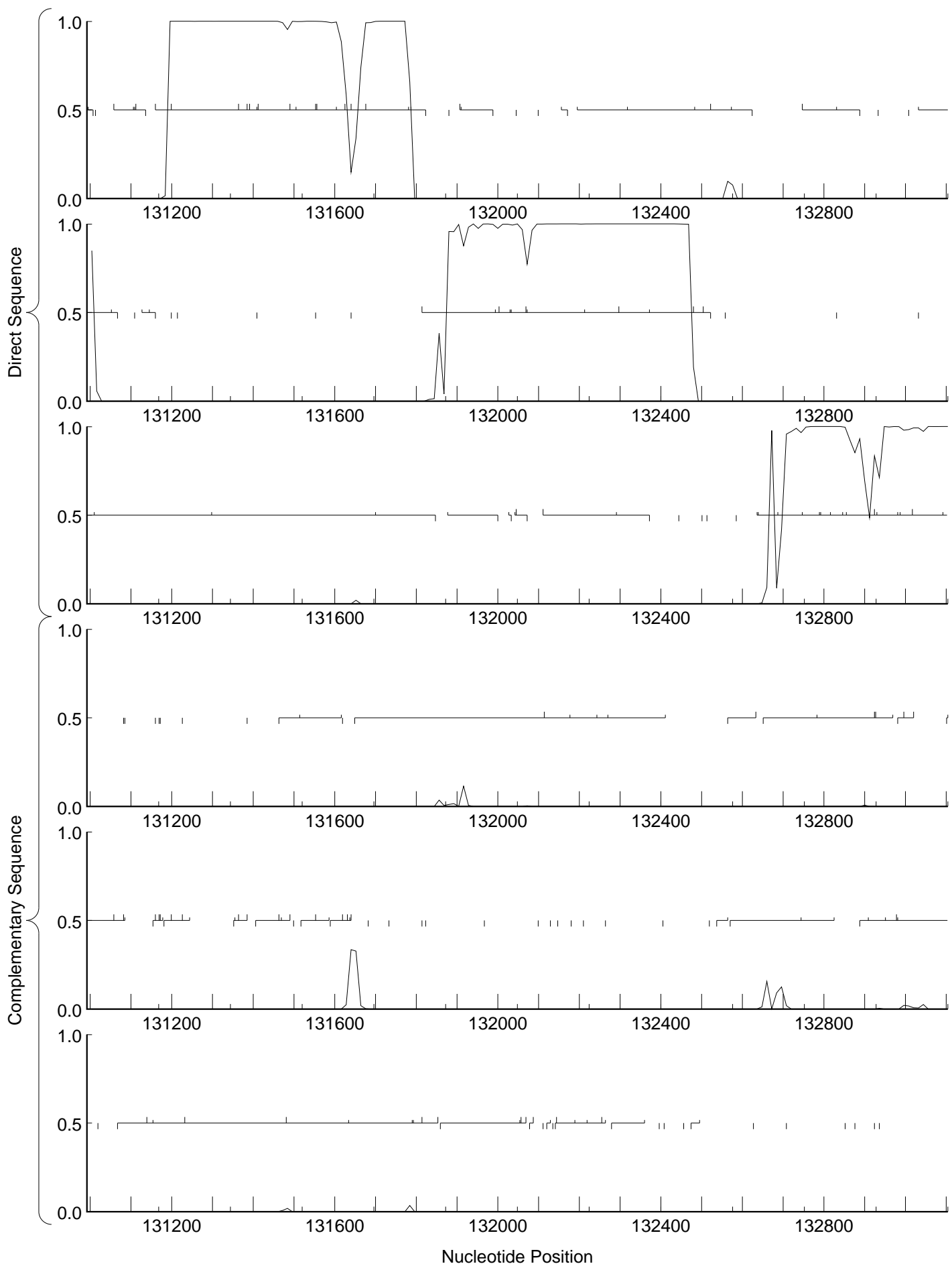


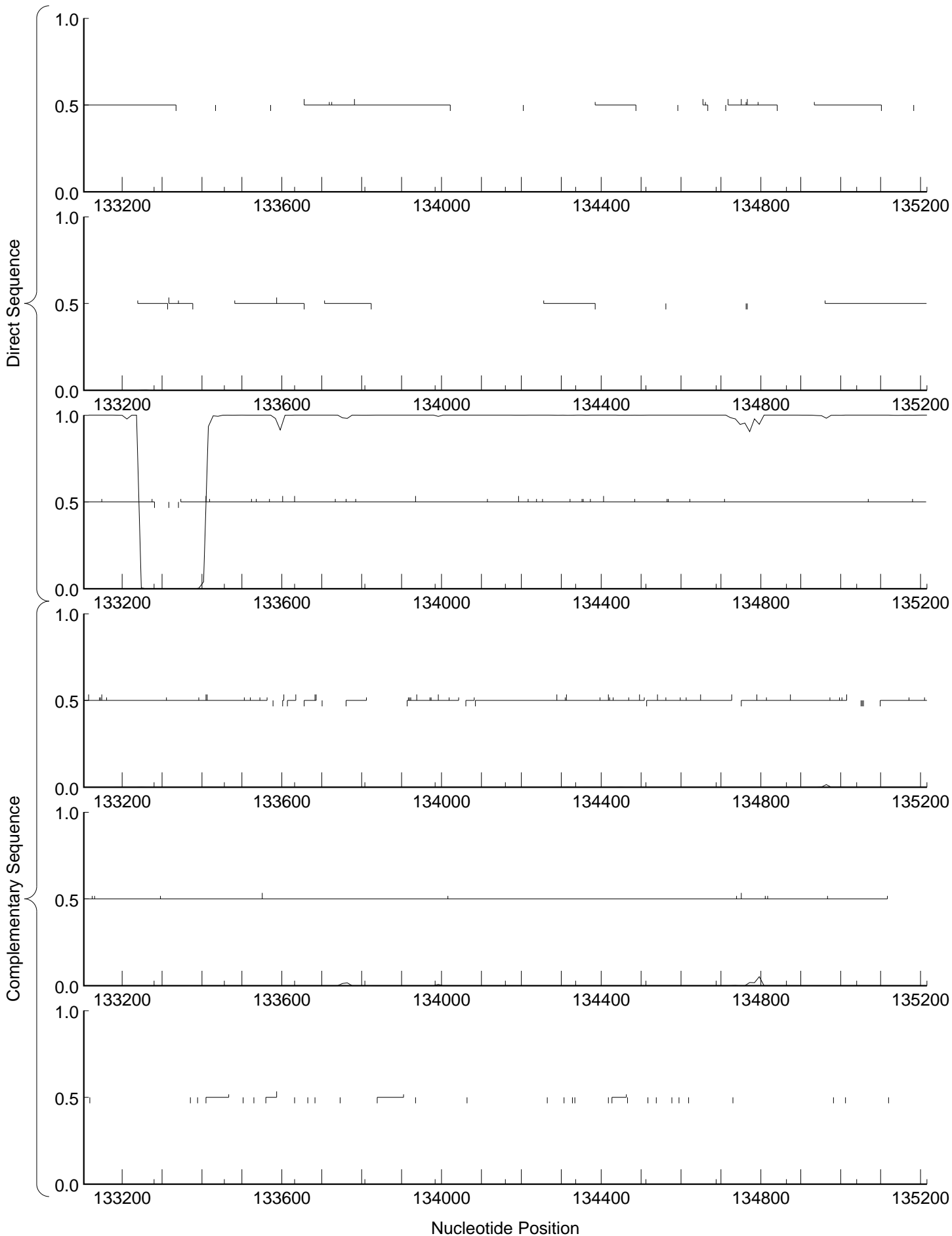


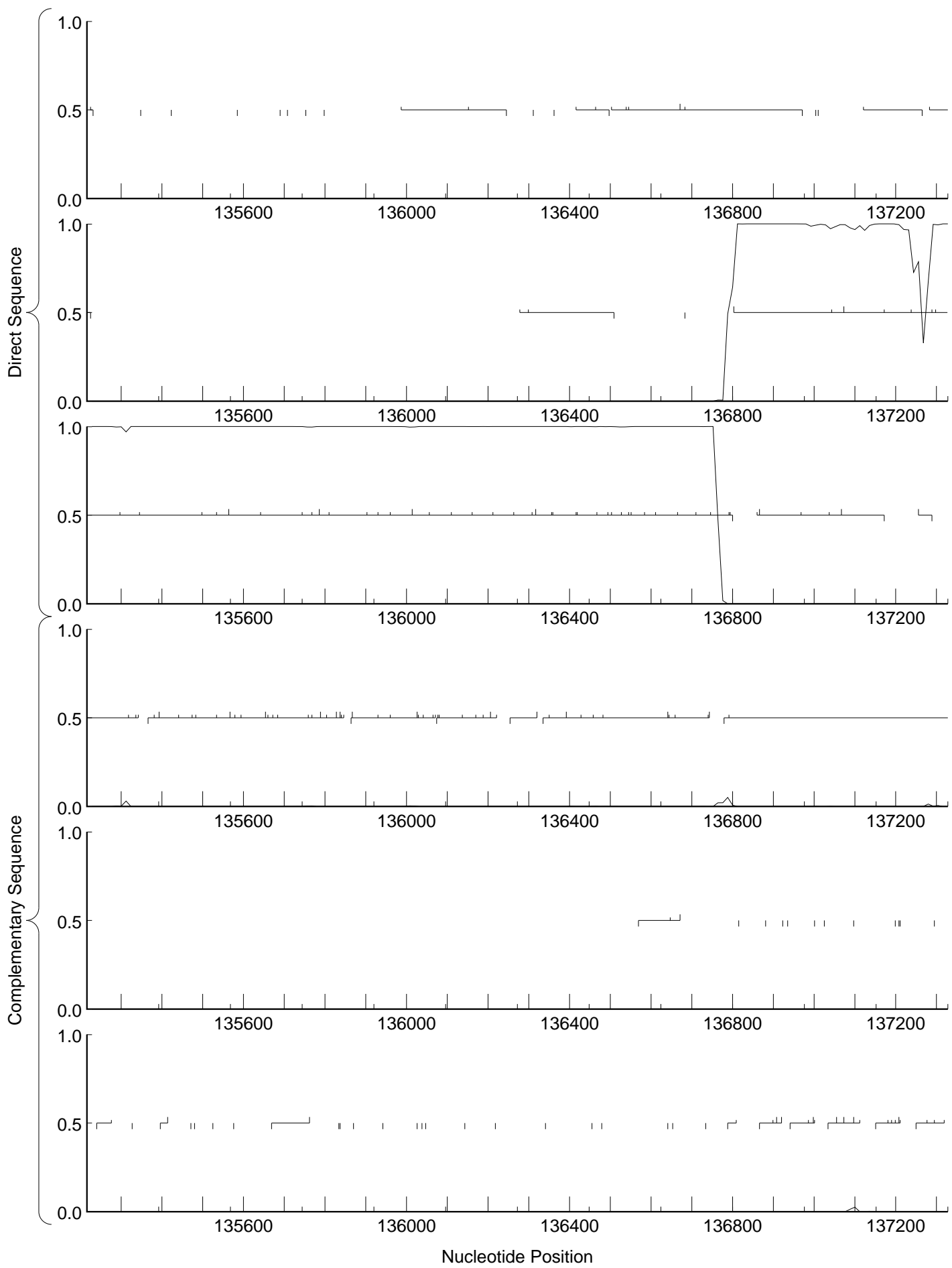


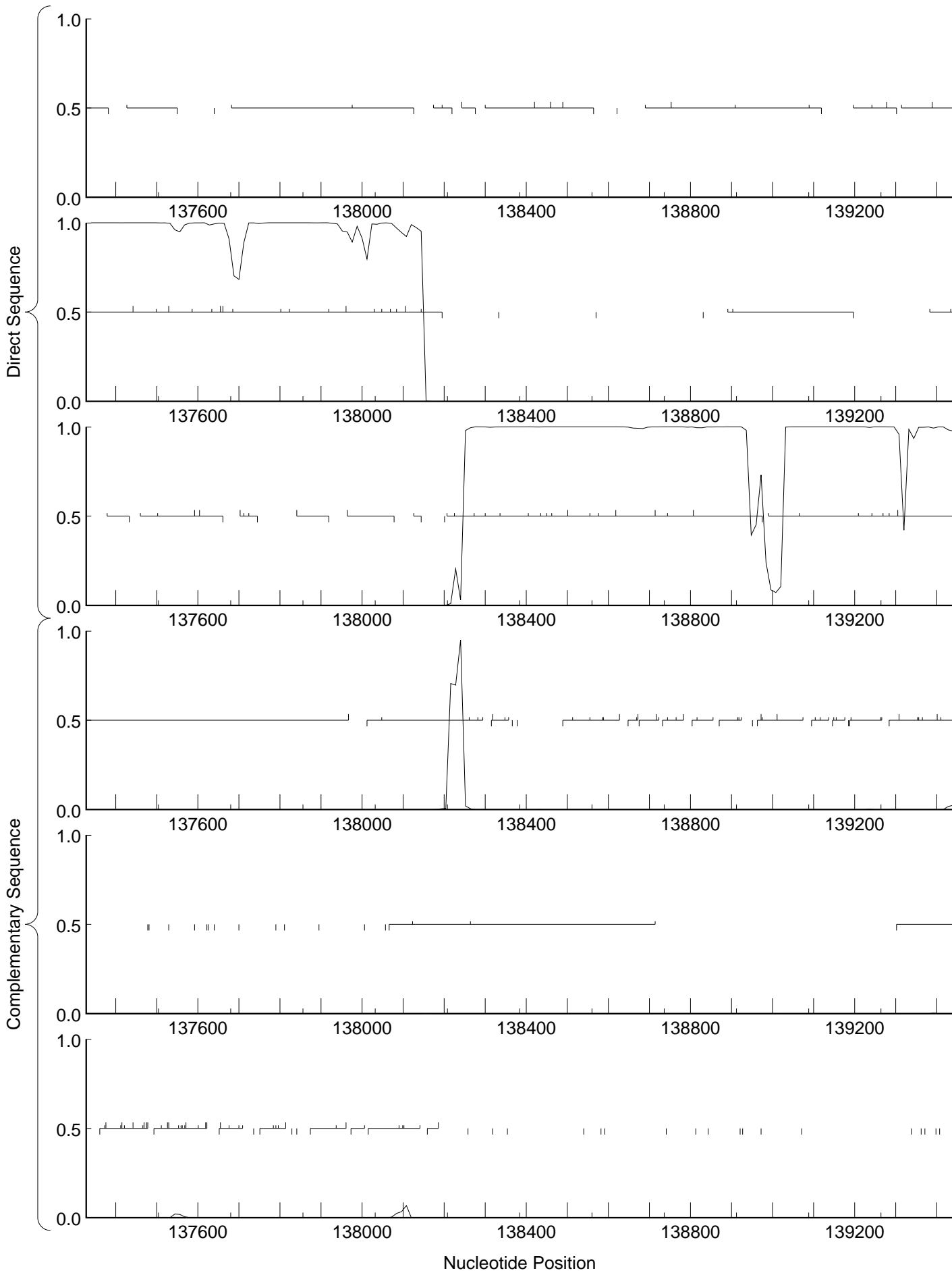


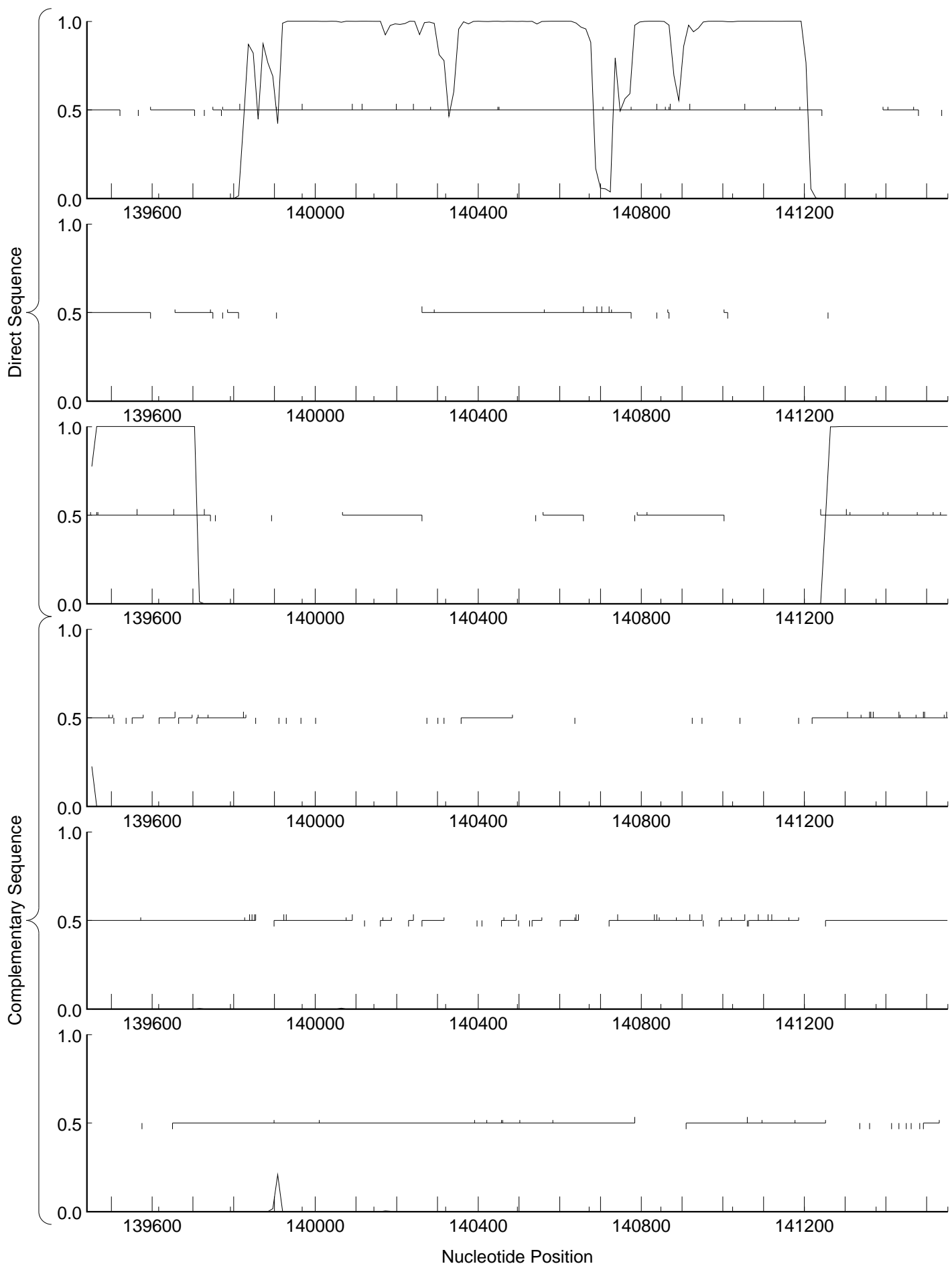


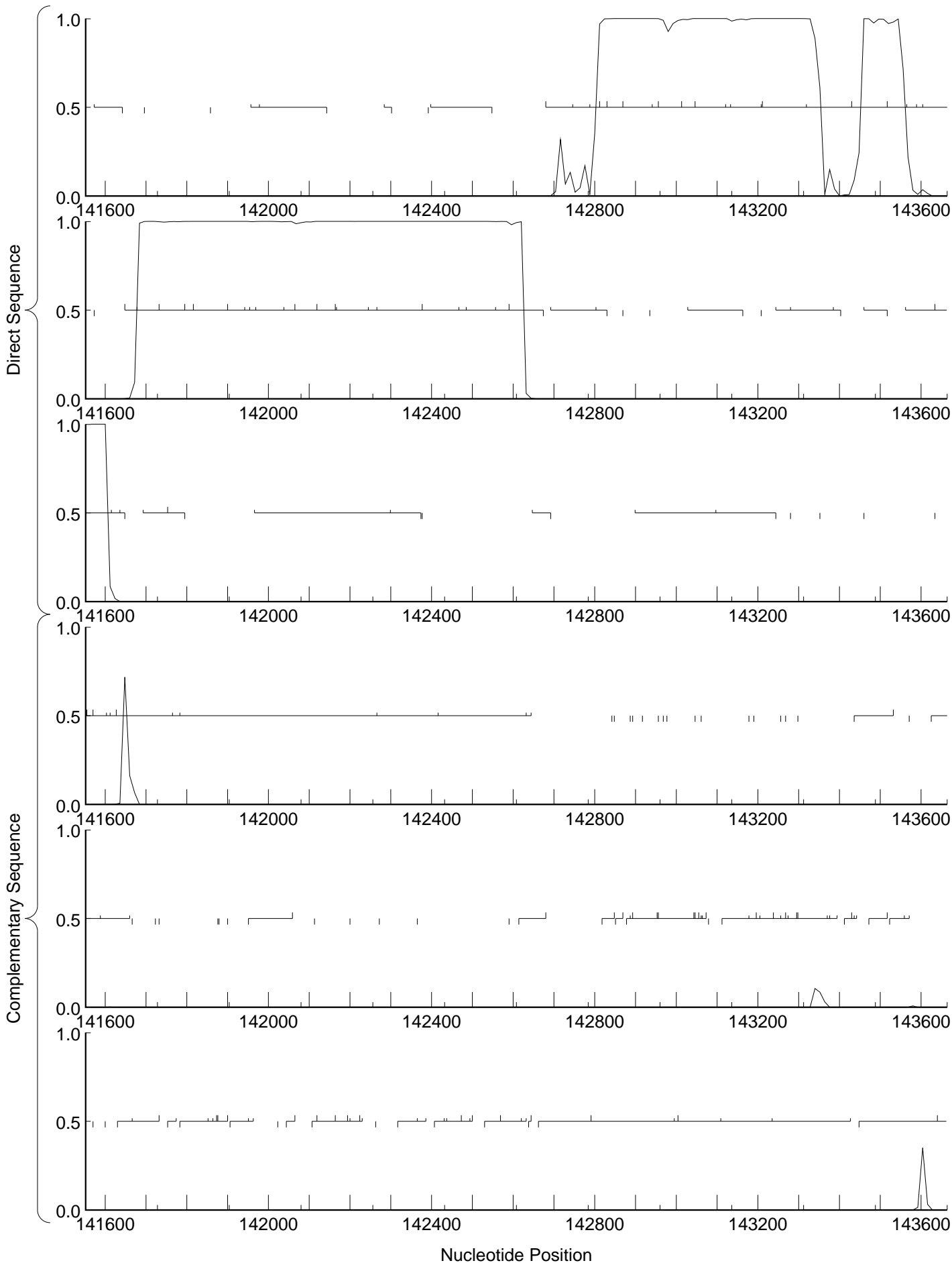


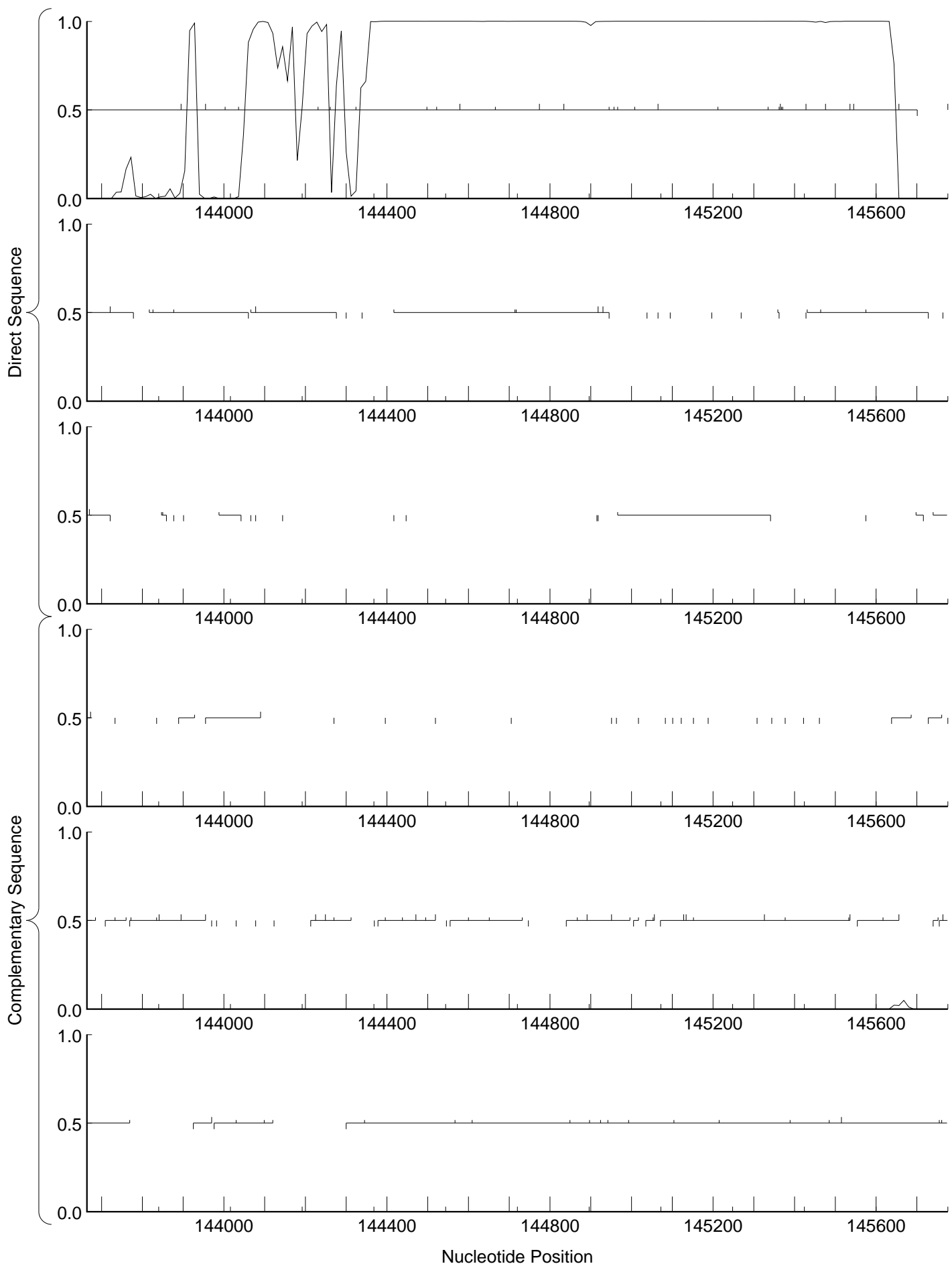


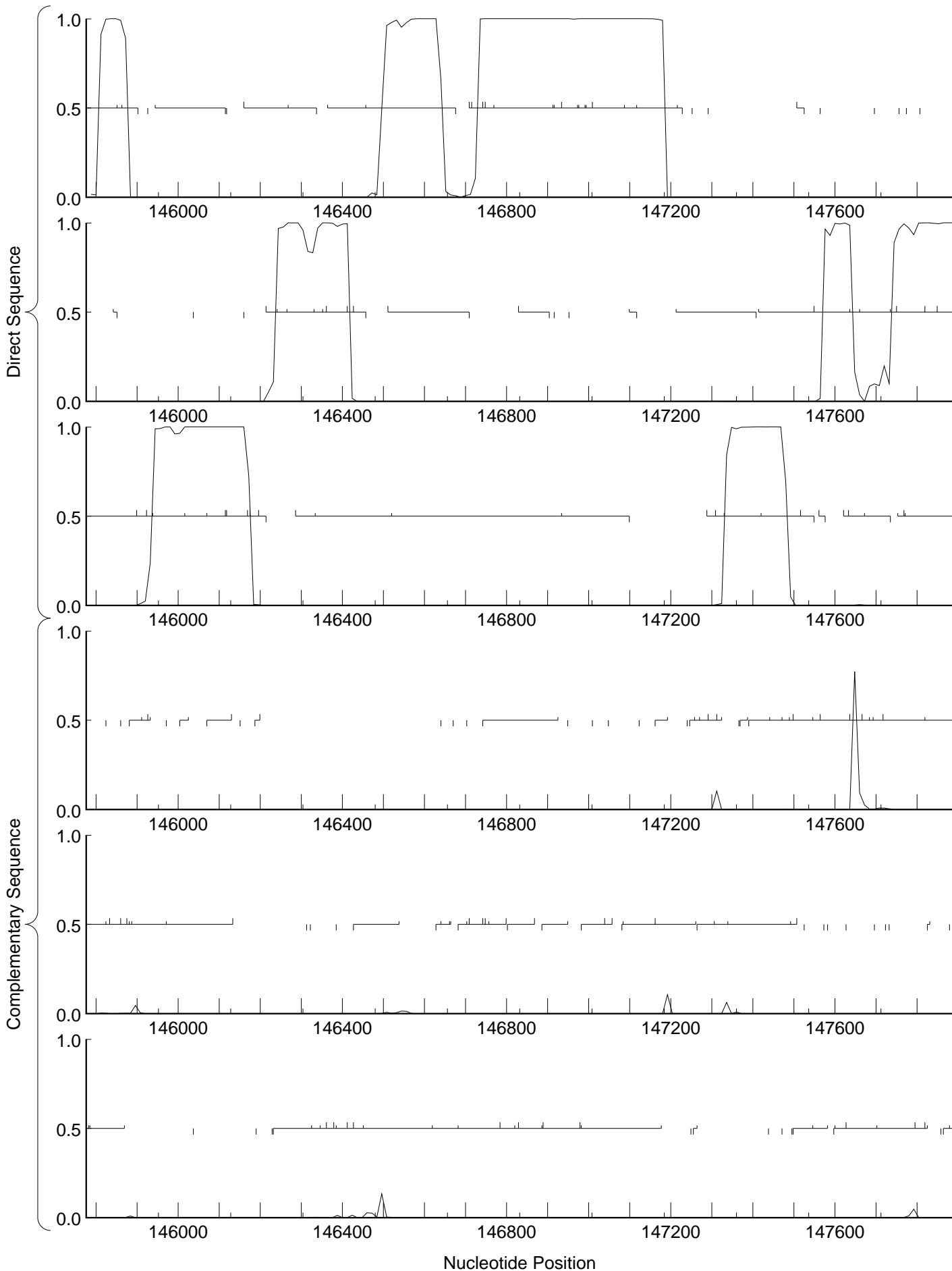


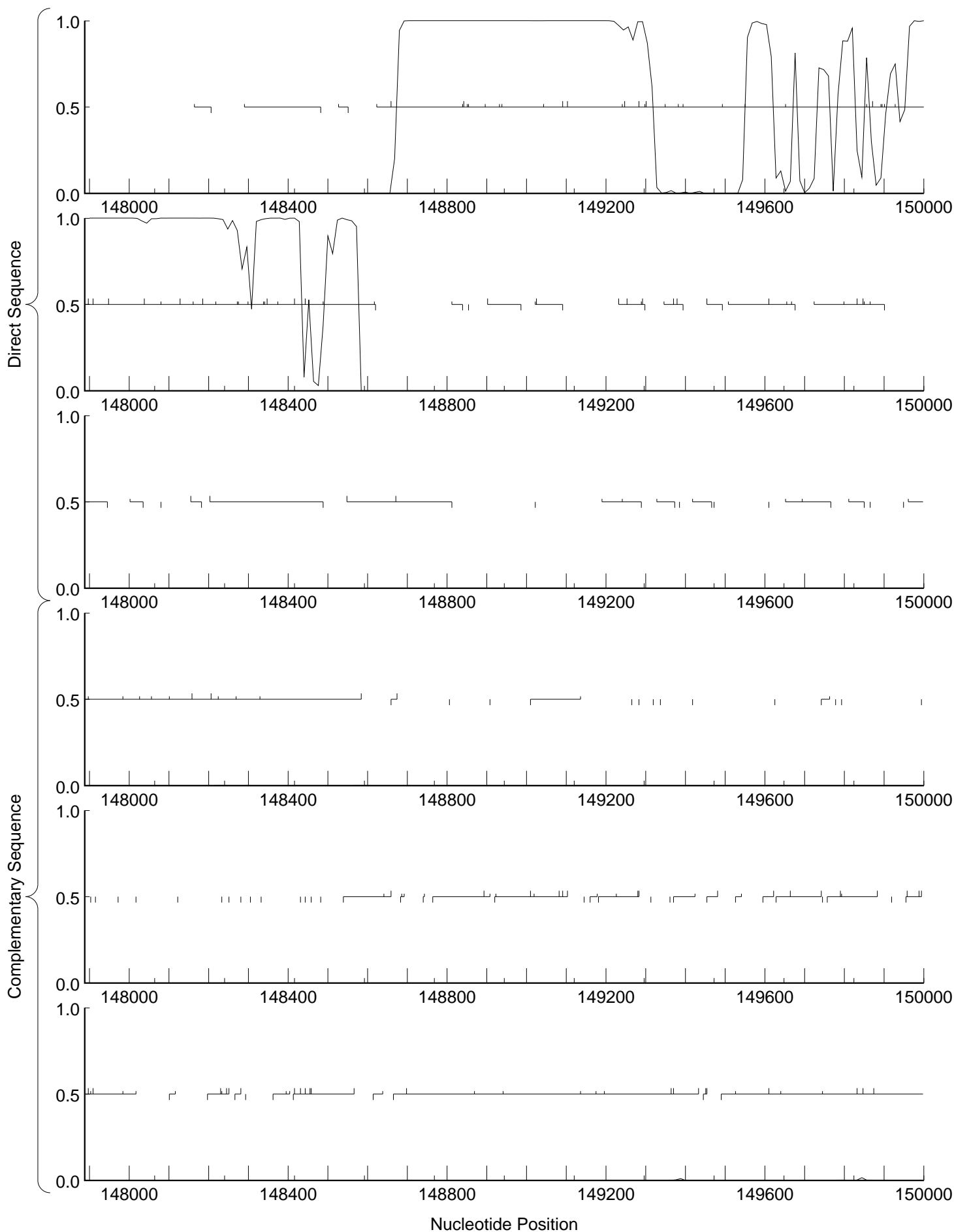


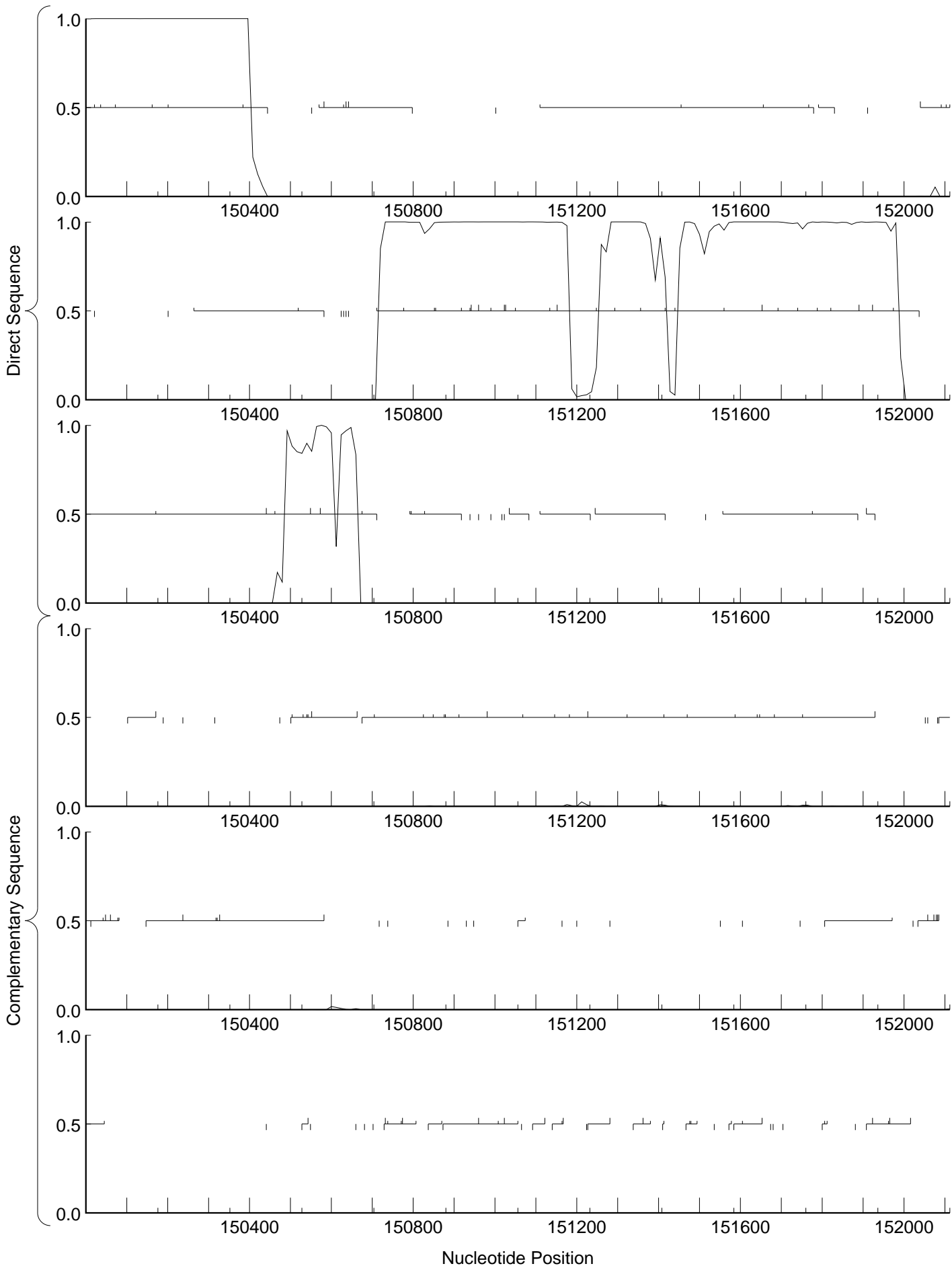


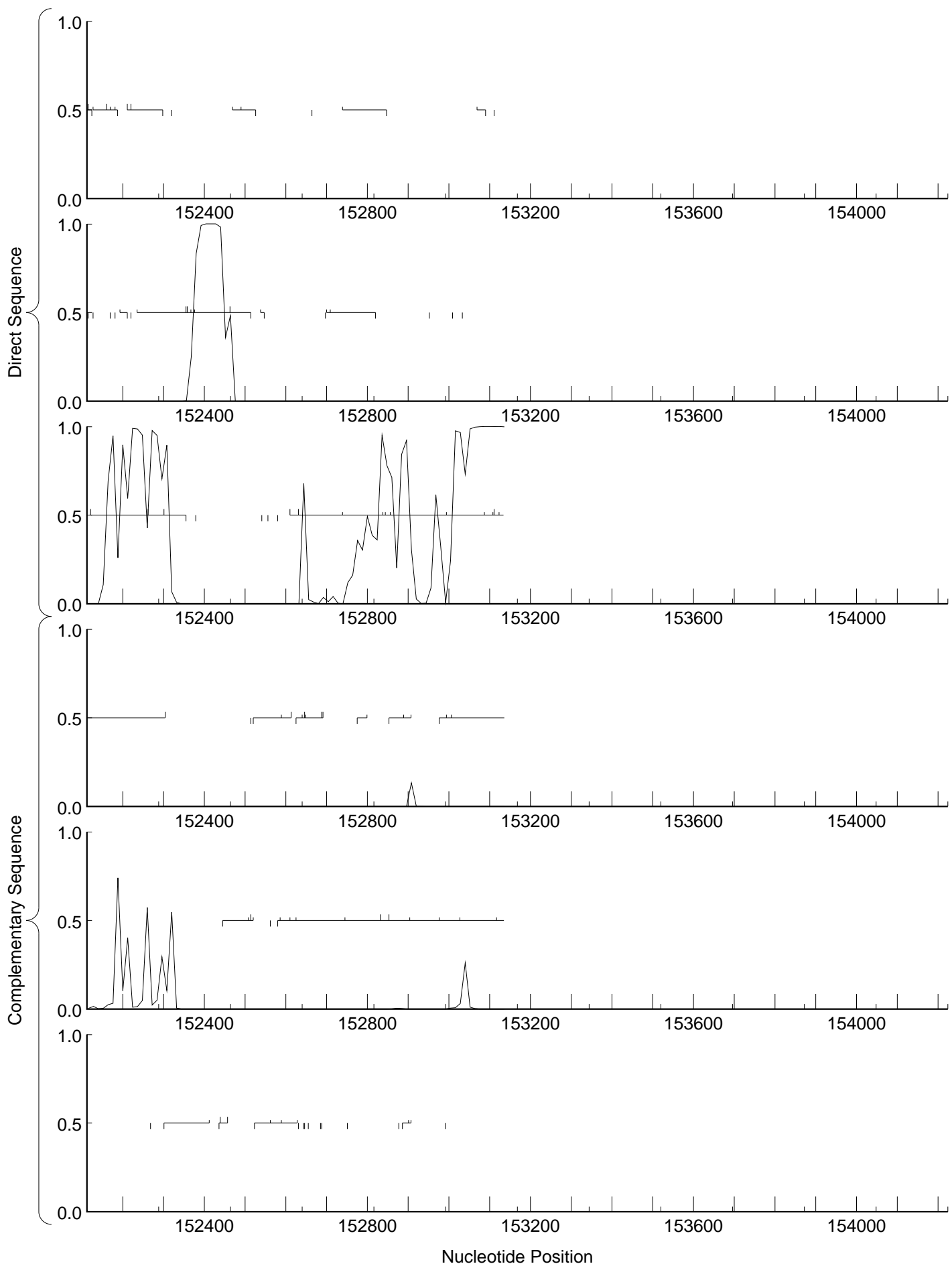












GeneMark

Version 2.5p (09.08.06)

Copyright 1993 - M. Borodovsky, J. McIninch

PROGRAM INFORMATION

Sequence : Mycobacterium phage Pinkcreek complete sequence, 153184 bp, circularly permuted, Cluster C1
Analysis Date : 7/1/21 at 11:35:50
Pages : 74
Sequence Length : 153184 bp
GC Content : 64.65%

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,

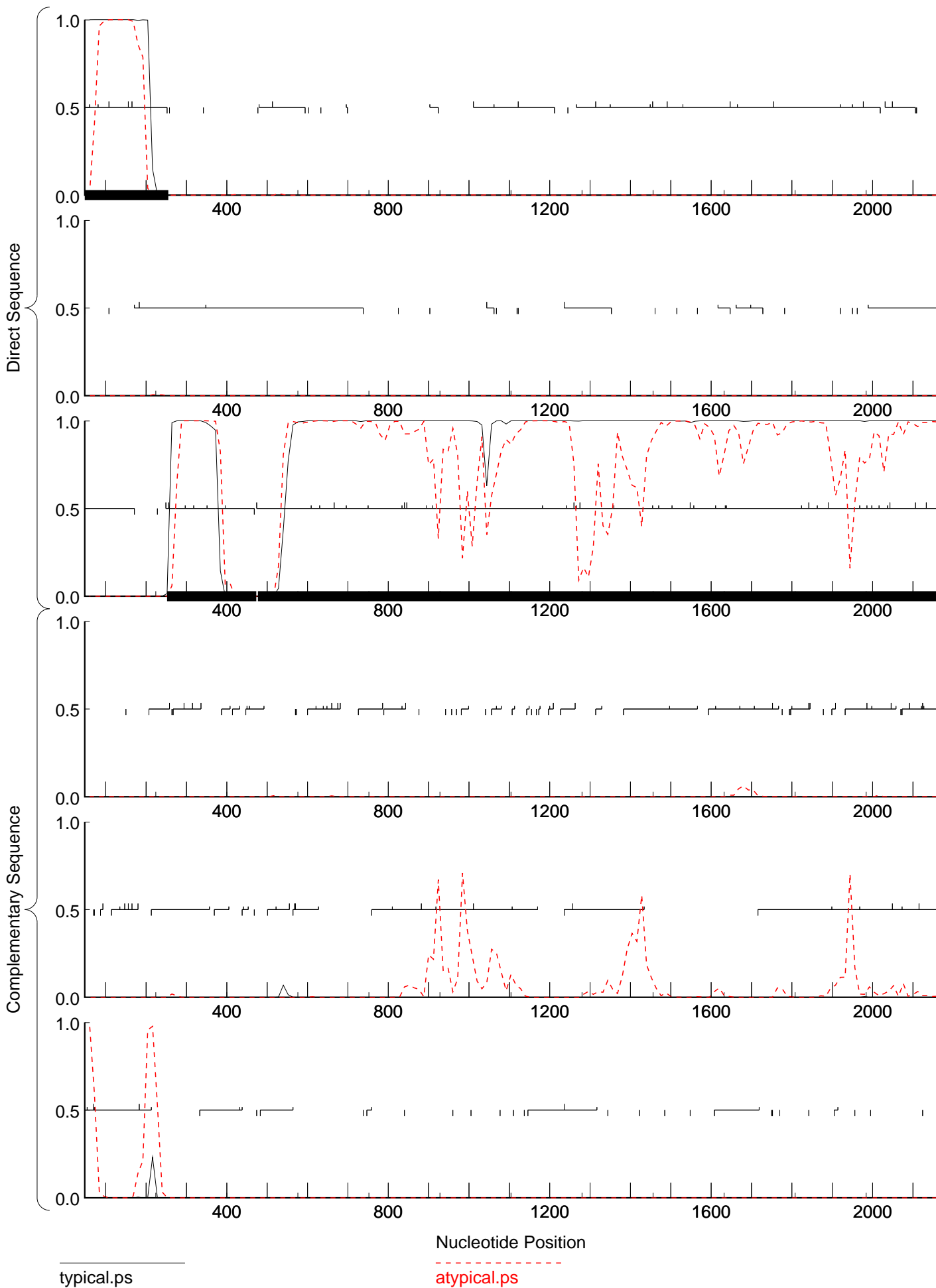
Matrix notes & comments

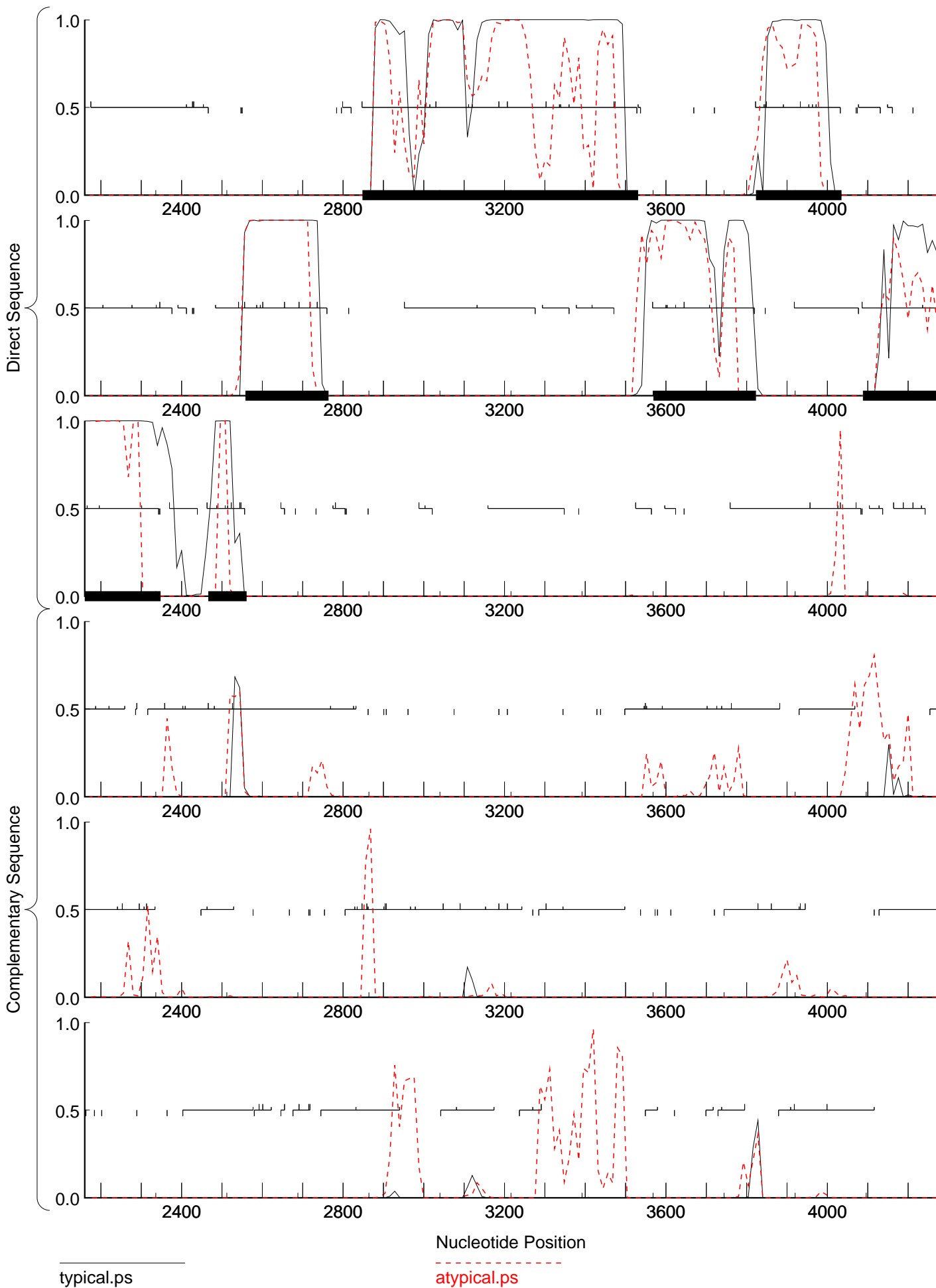
Training set derived by GeneMarkS, 4.30 October 2014
Thu Jul 1 11:35:50 2021

MATRIX INFORMATION

Matrix : -
Author : -
Order : 4

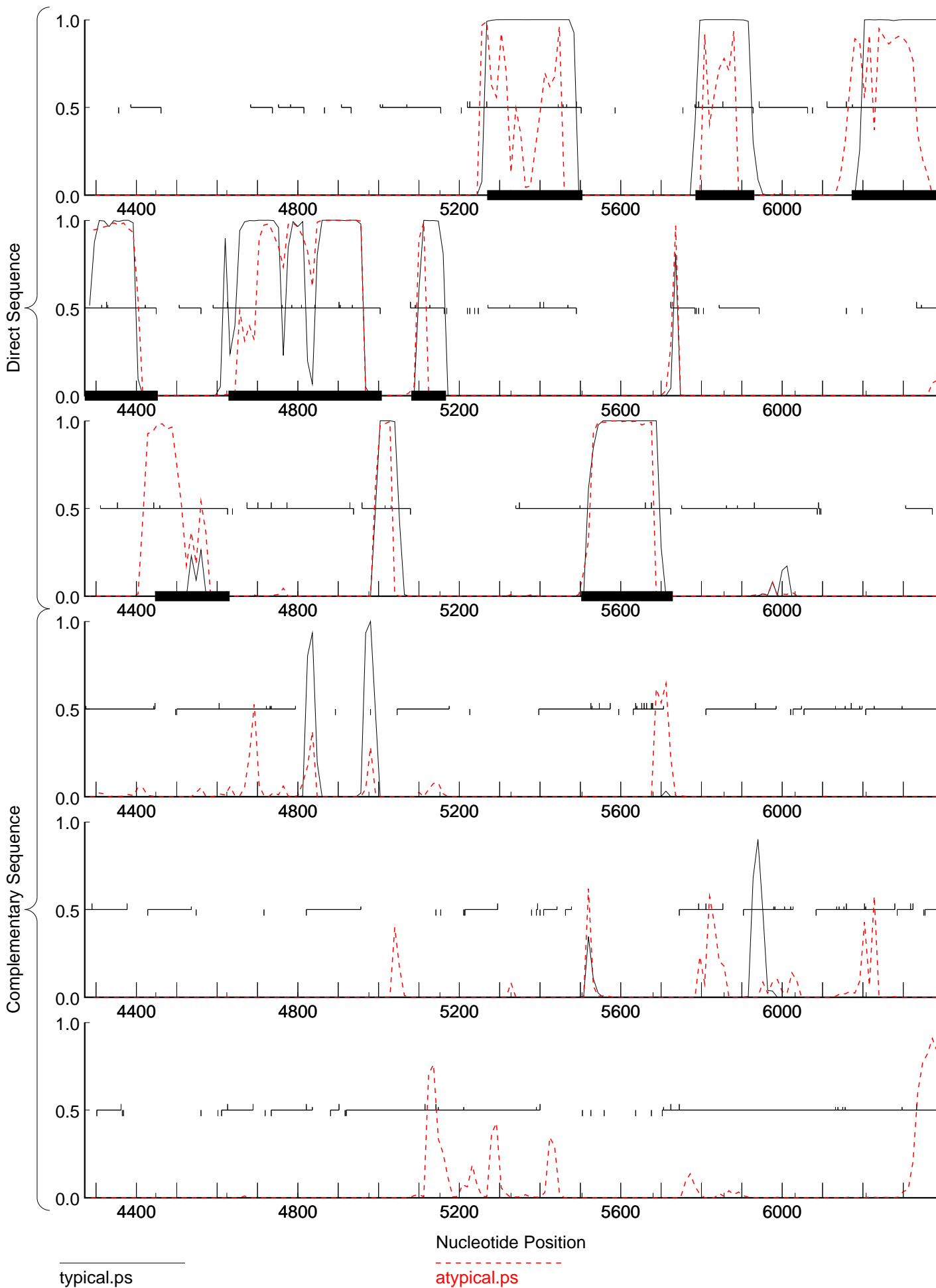
Send questions / comments to:
Dr. M. Borodovsky
Georgia Institute of Technology
School of Biology
Atlanta, GA 30332-0230

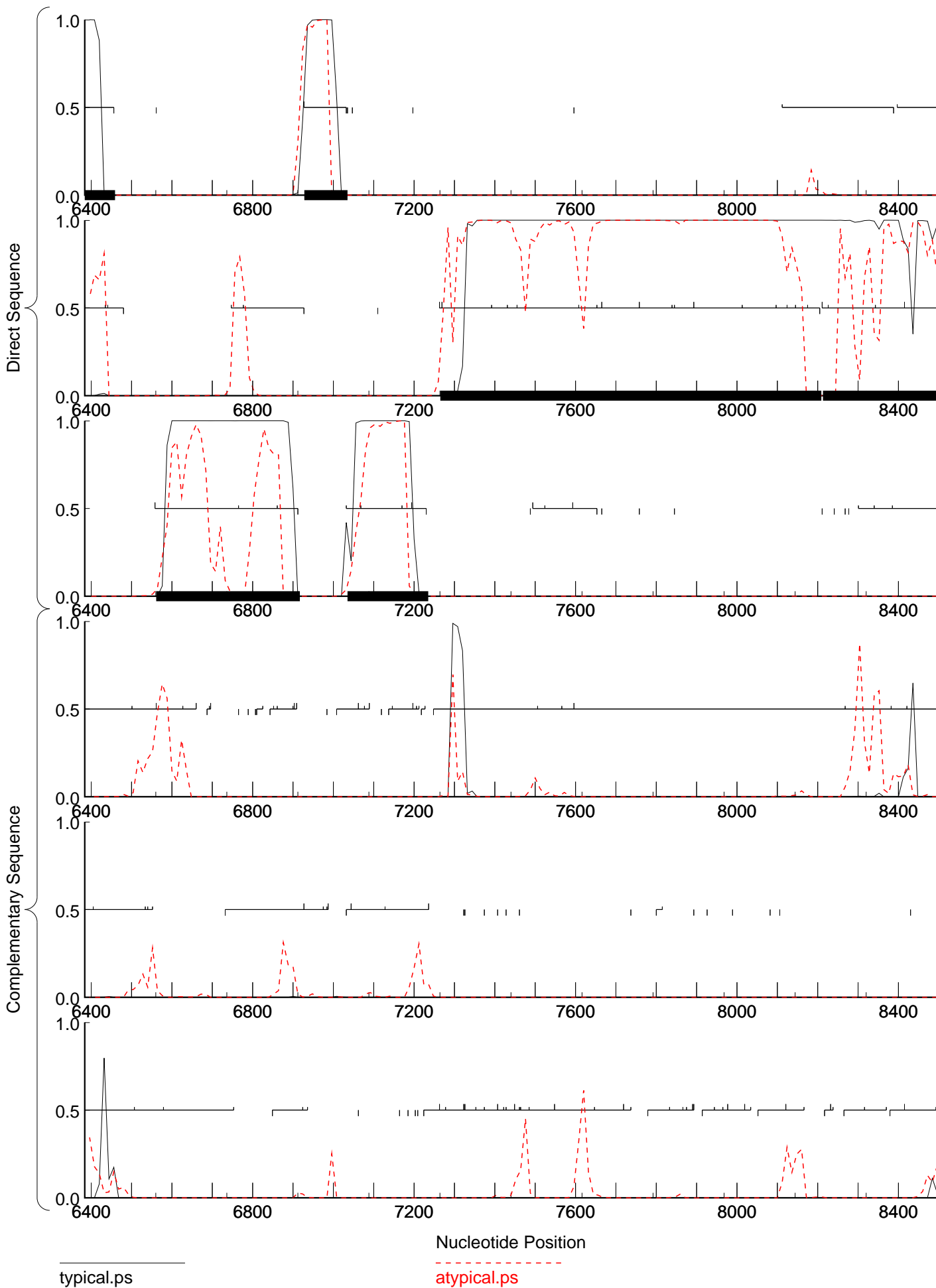


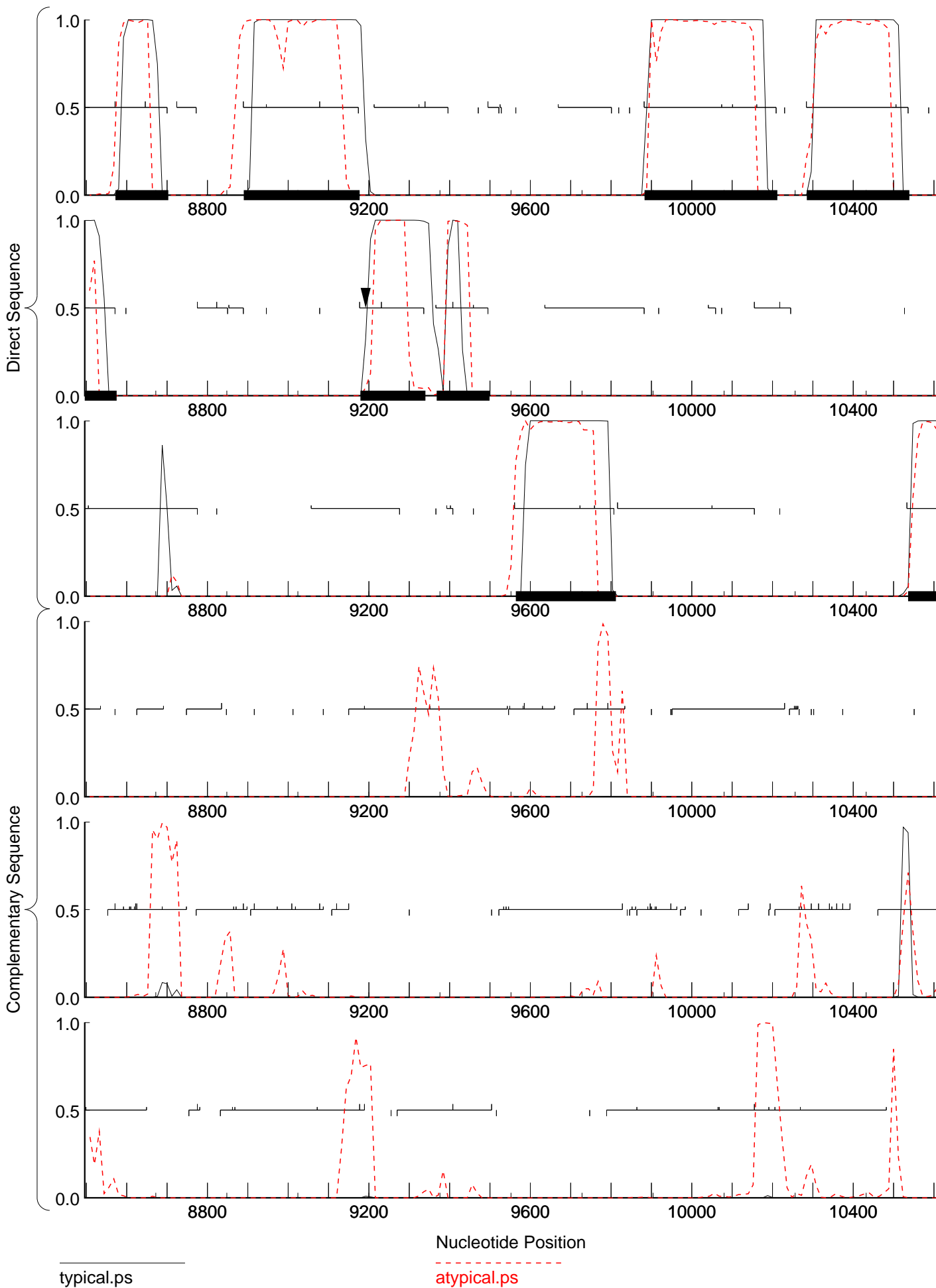


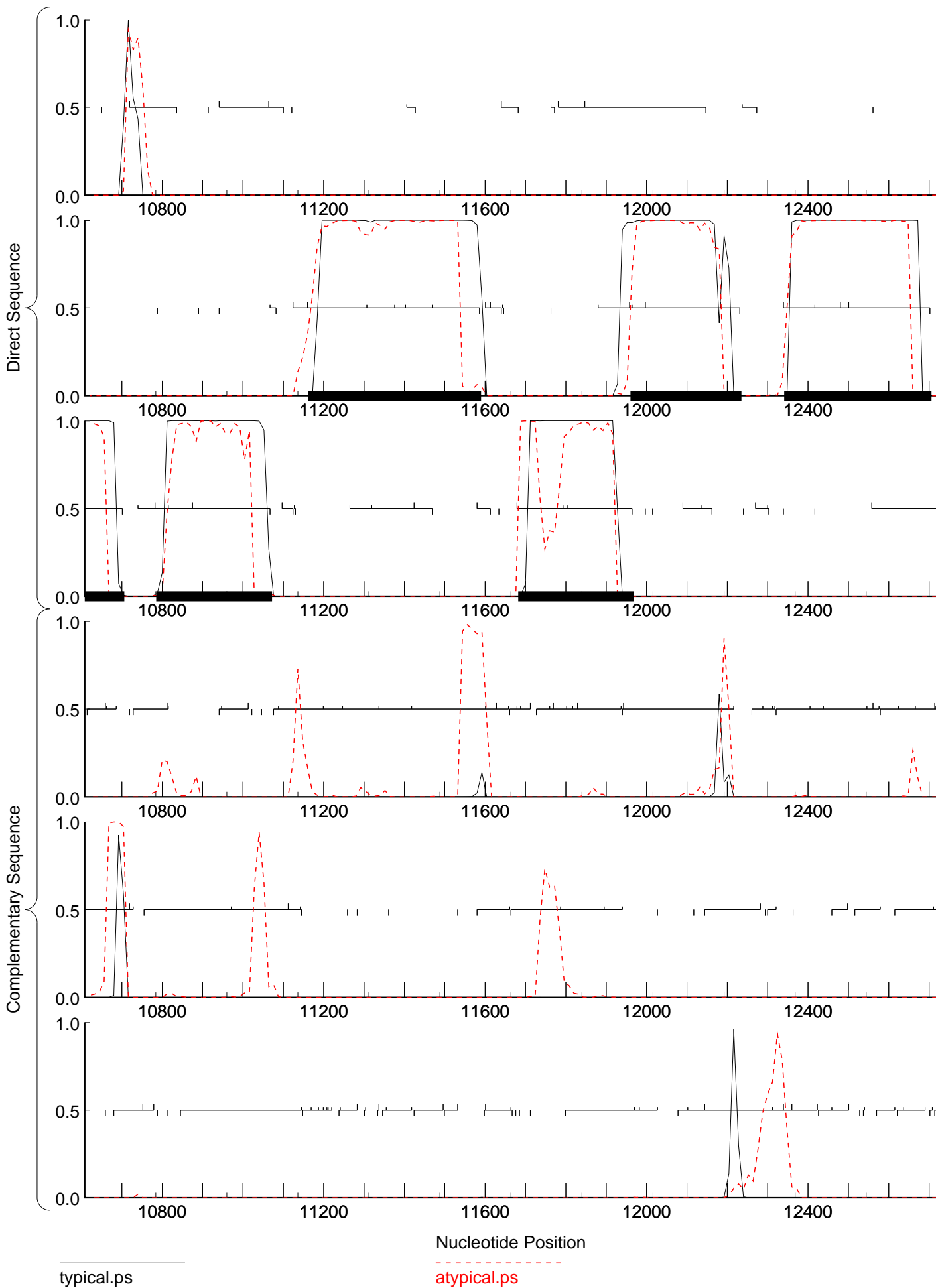
typical.ps

atypical.ps



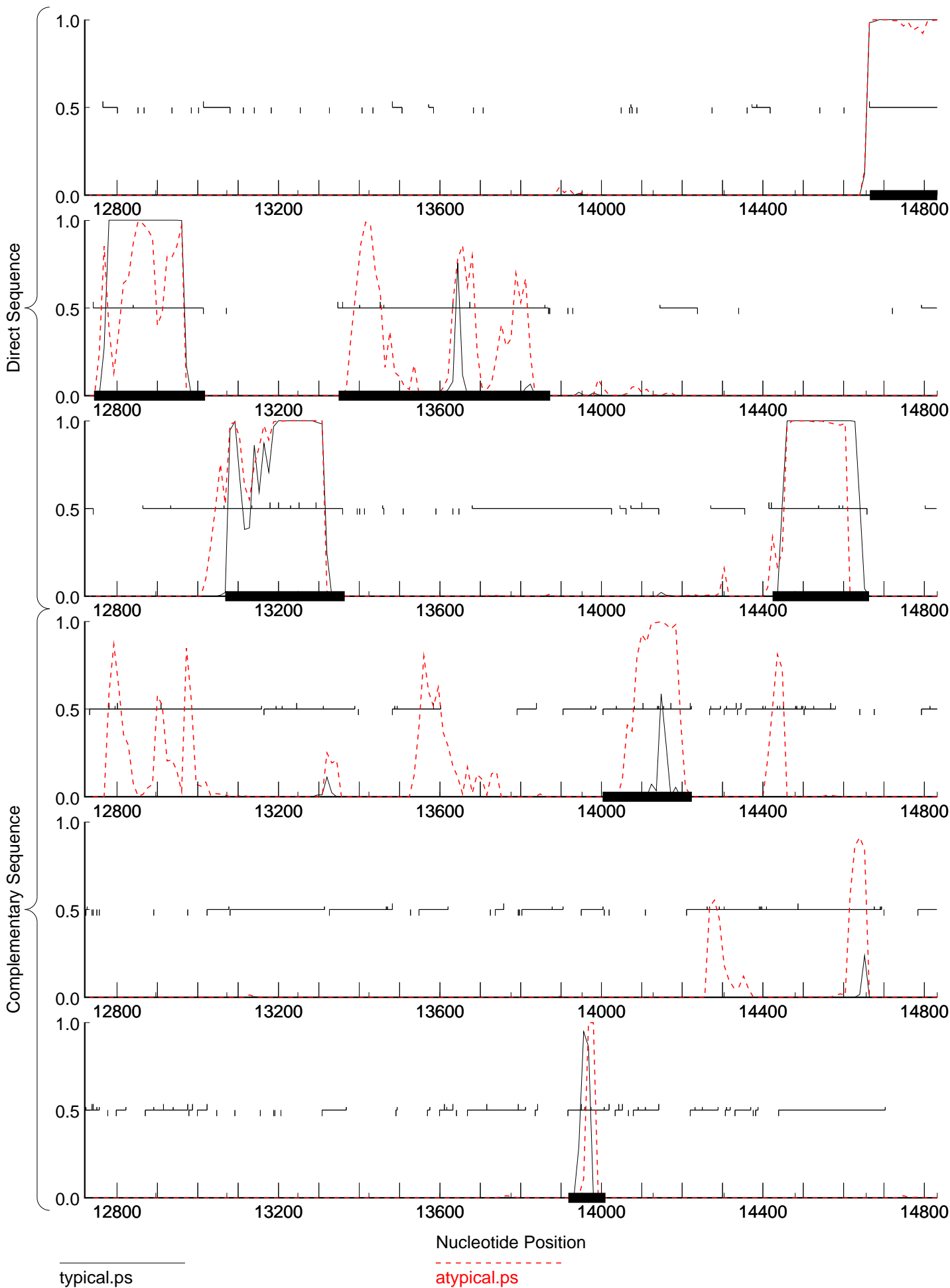






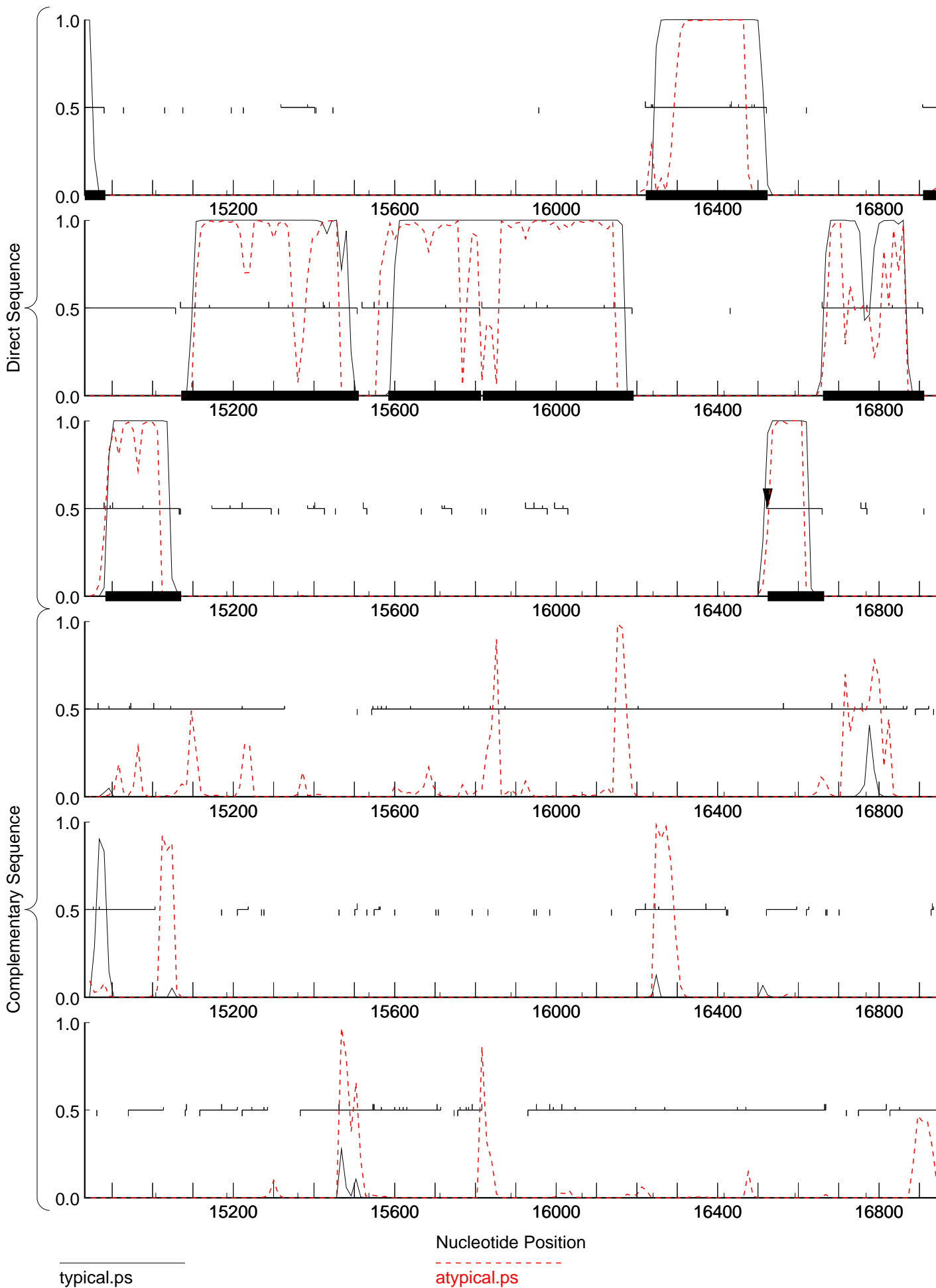
typical.ps

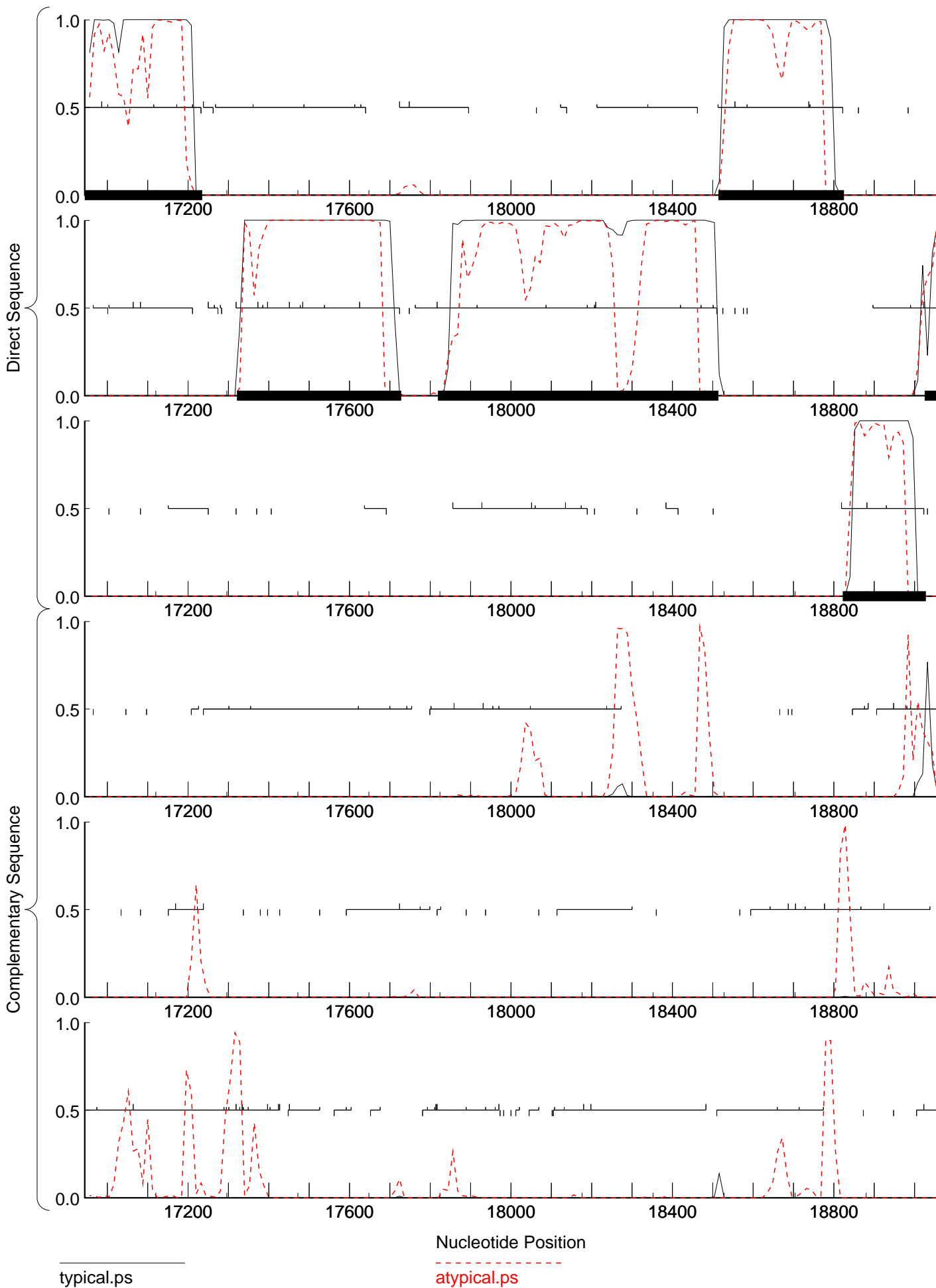
atypical.ps



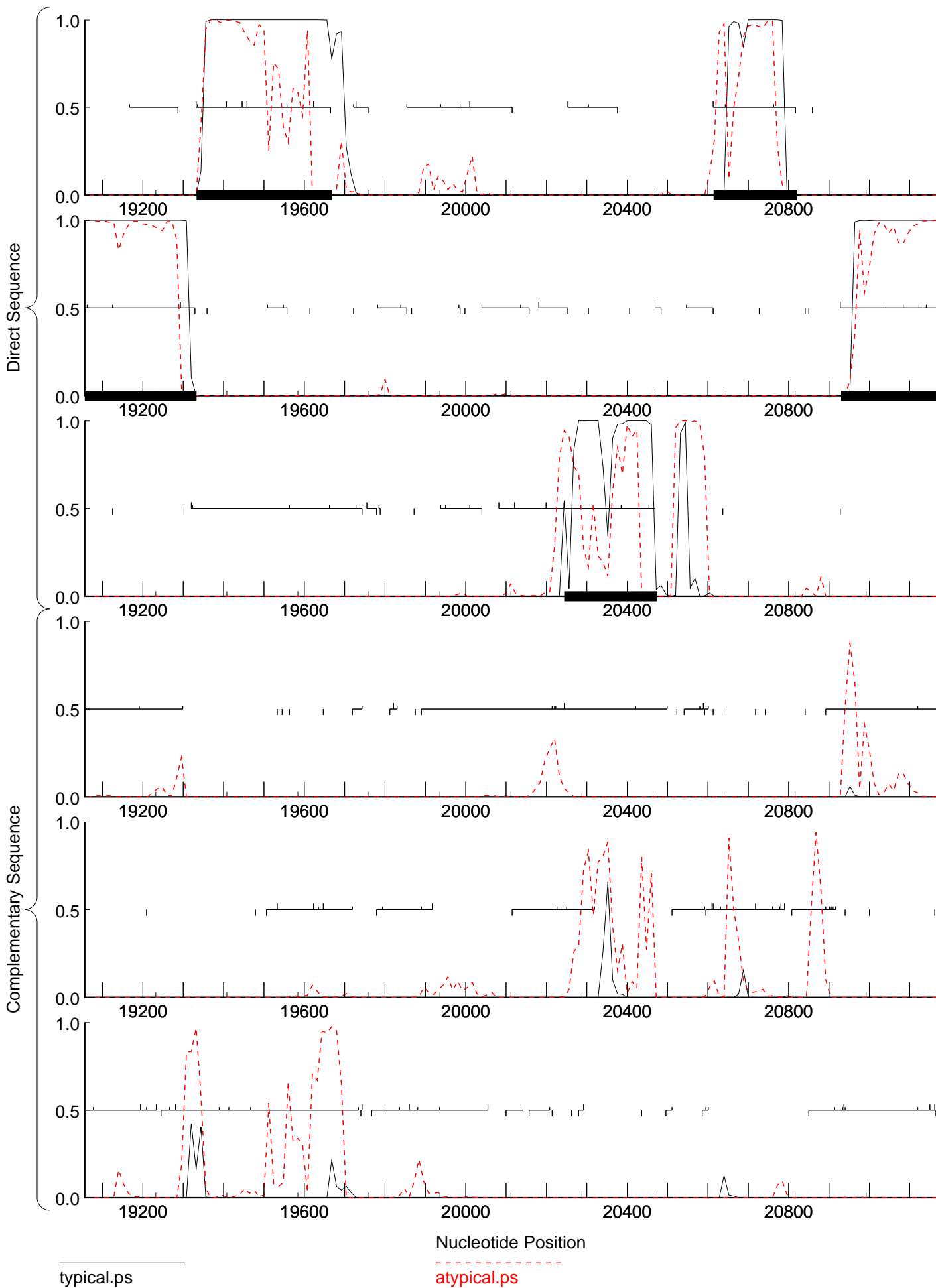
typical.ps

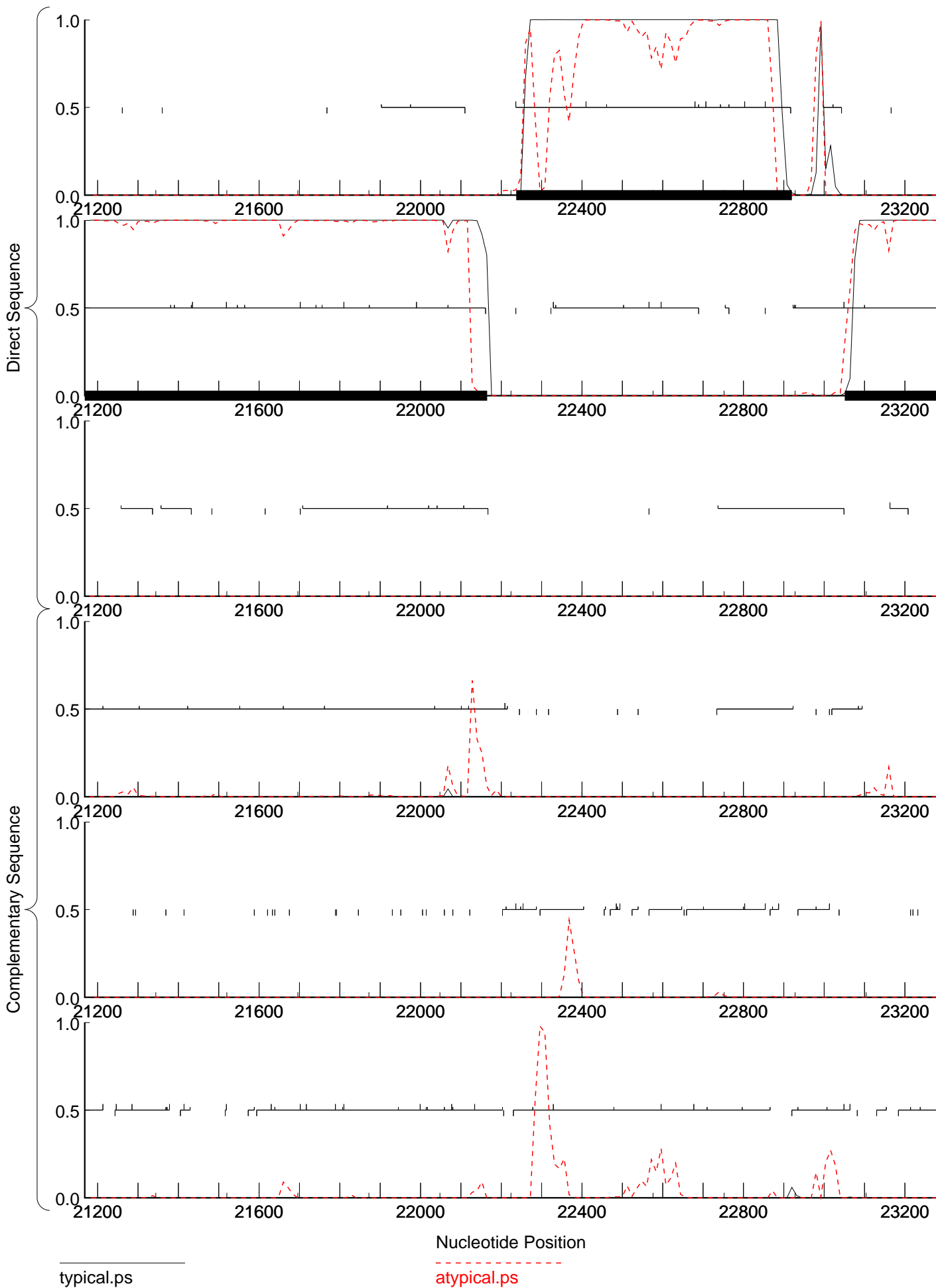
atypical.ps





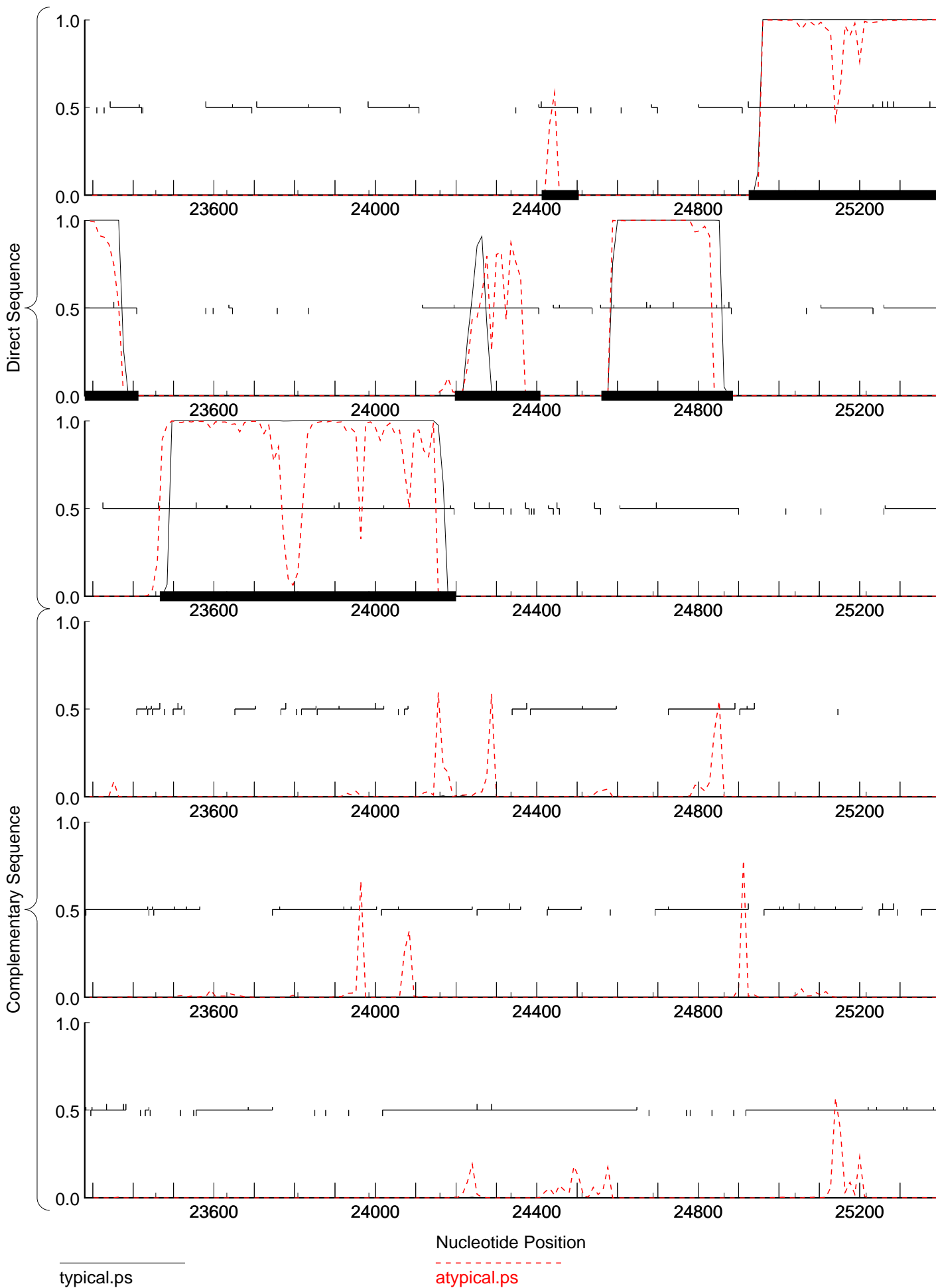
GeneMark.hmm prediction

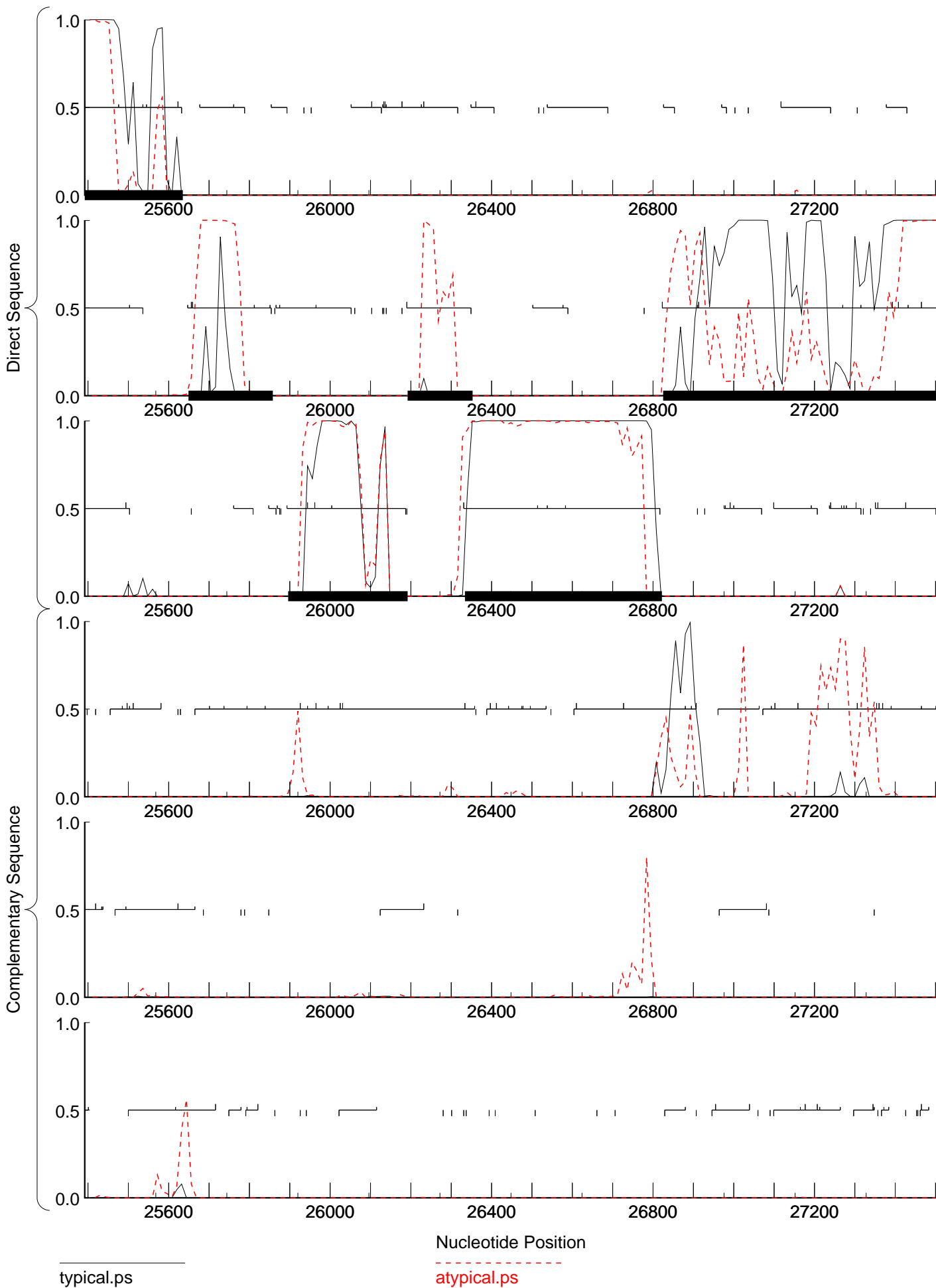


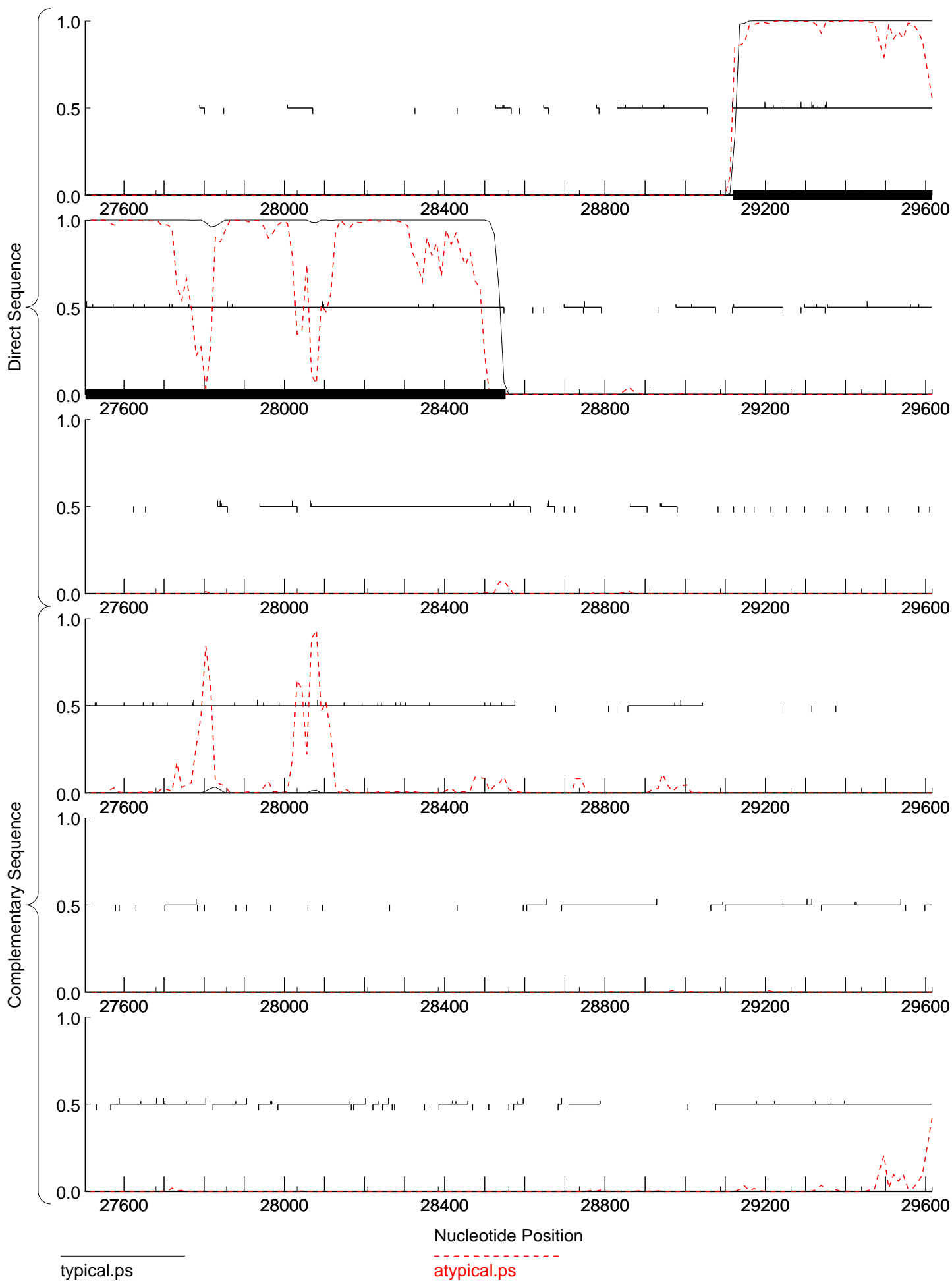


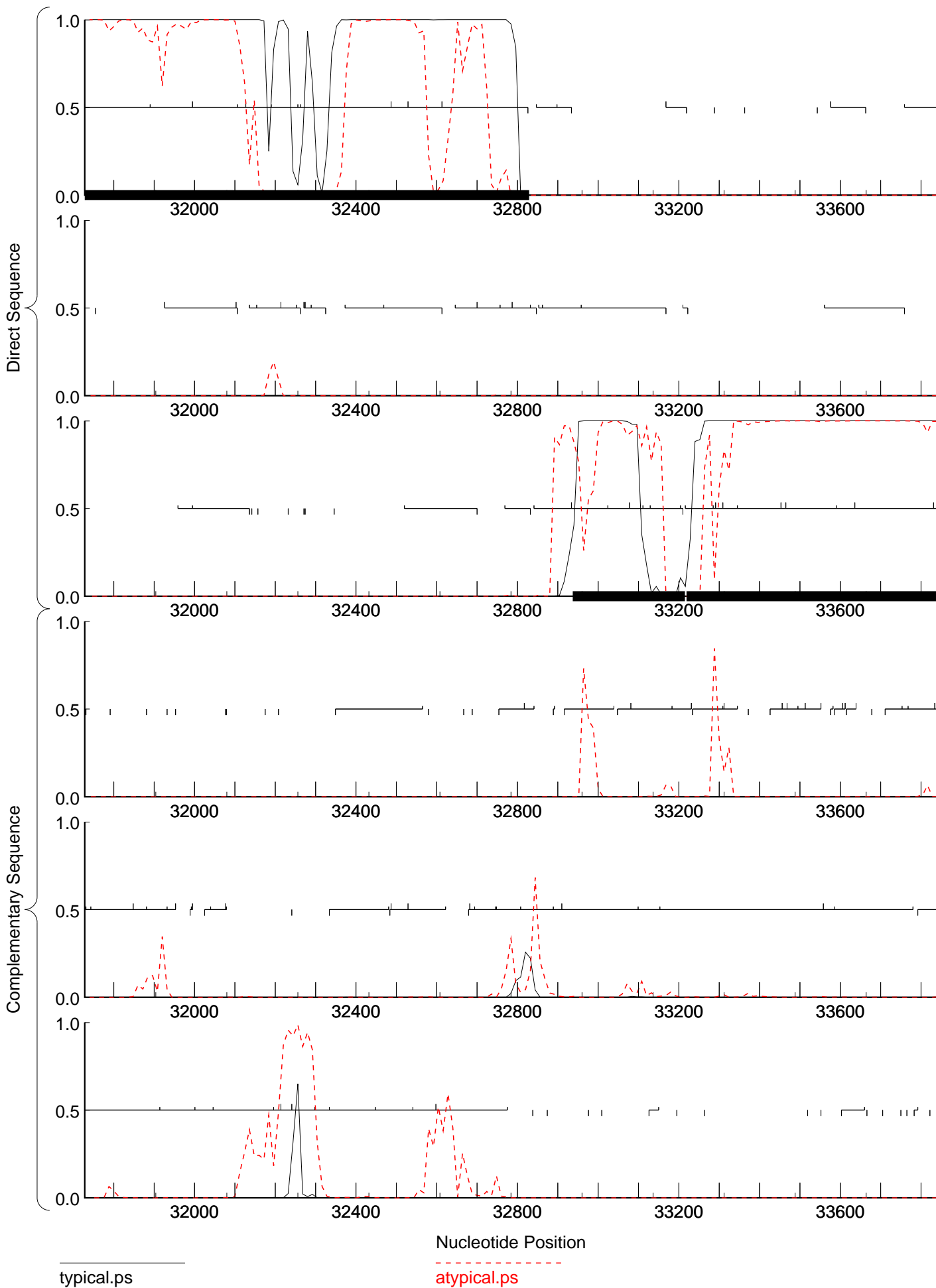
typical.ps

atypical.ps



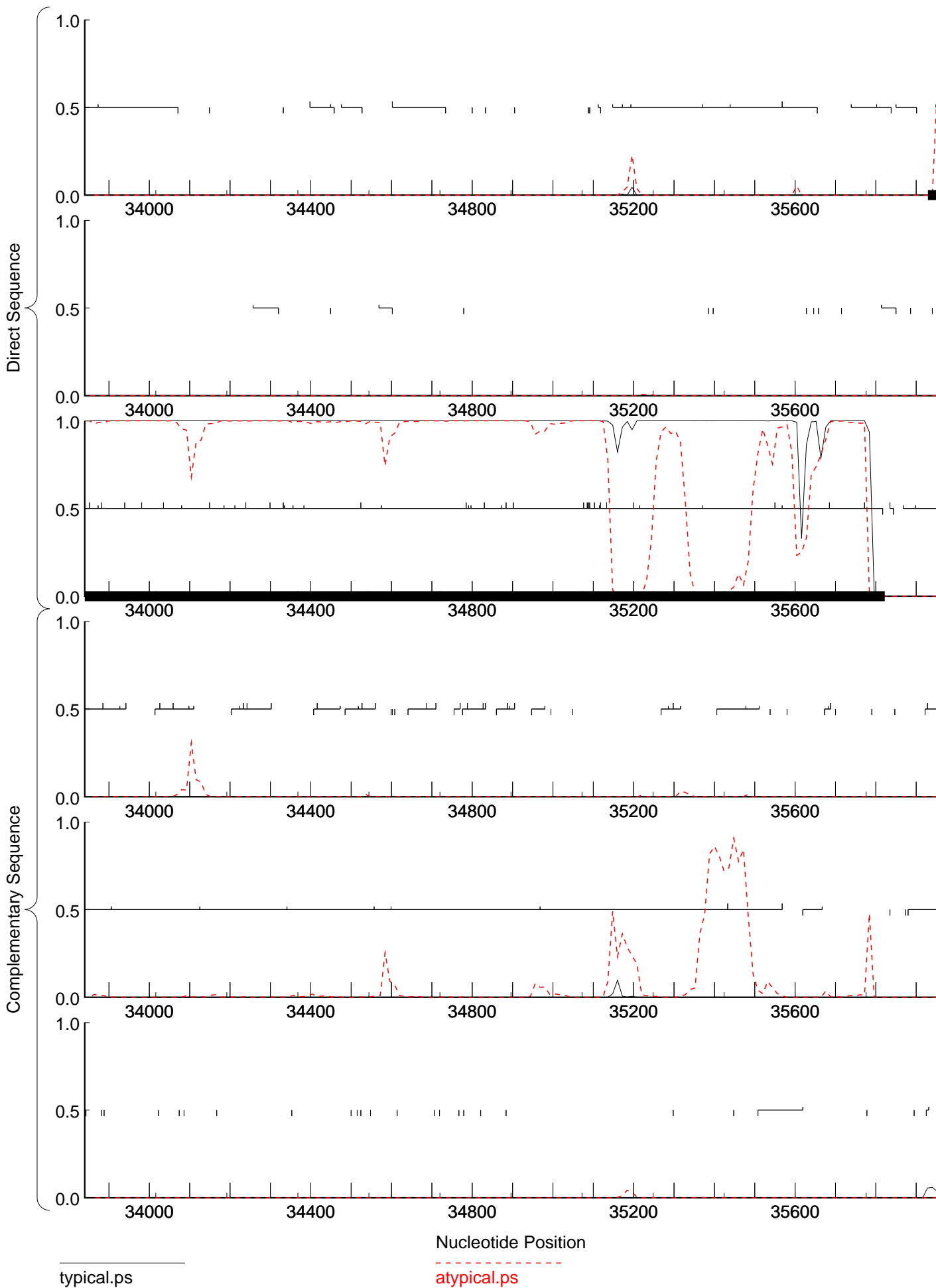






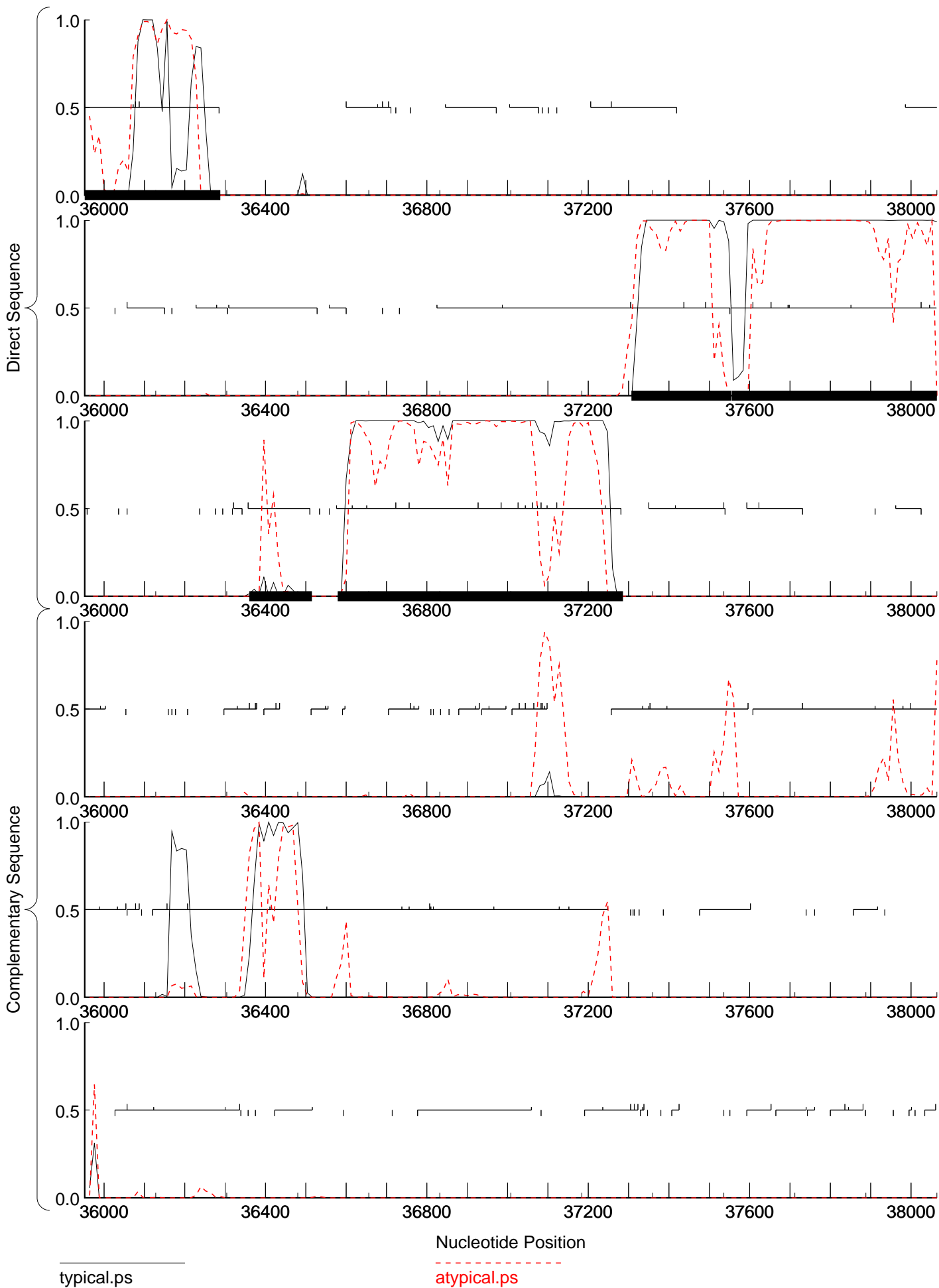
typical.ps

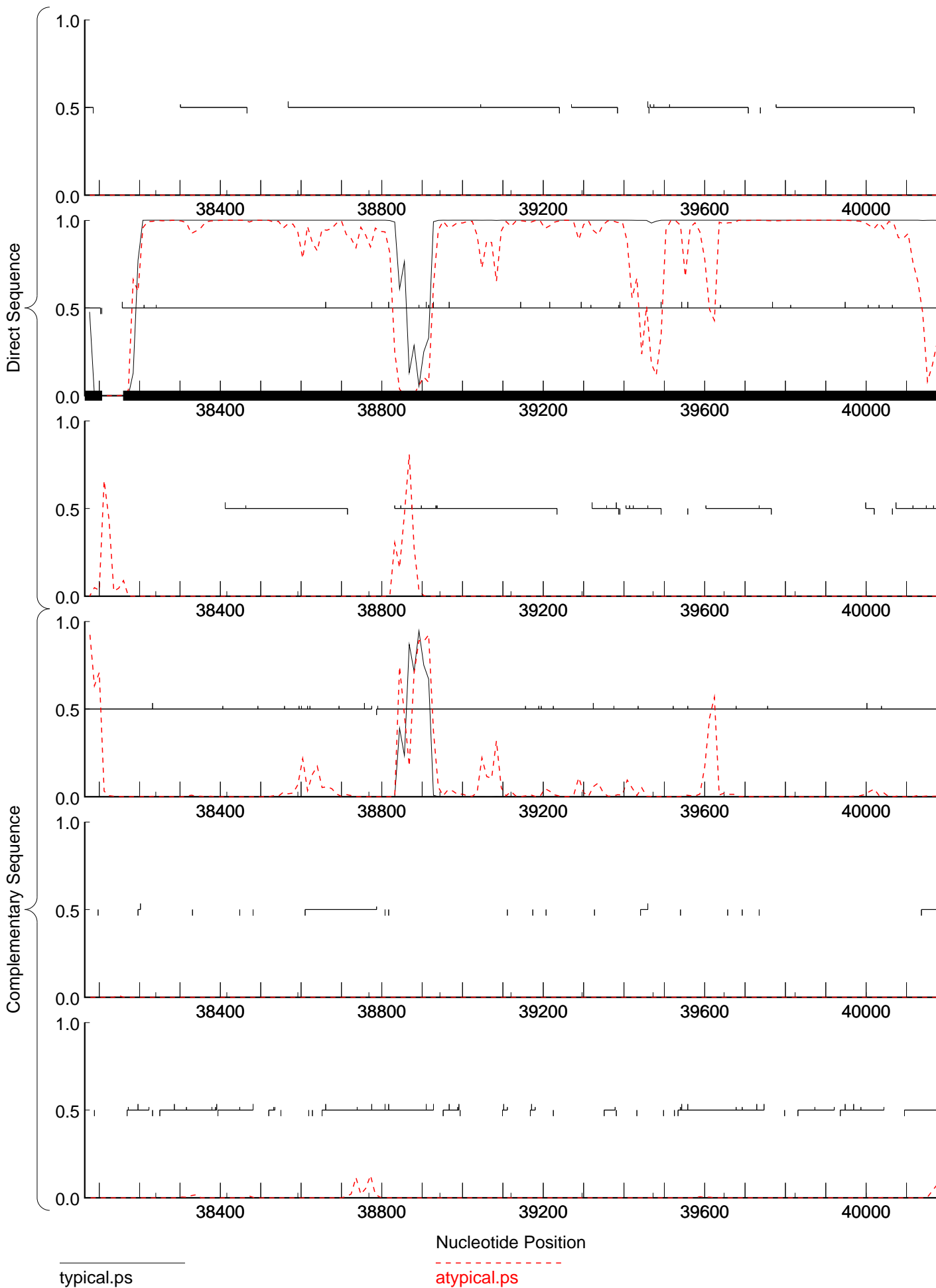
atypical.ps

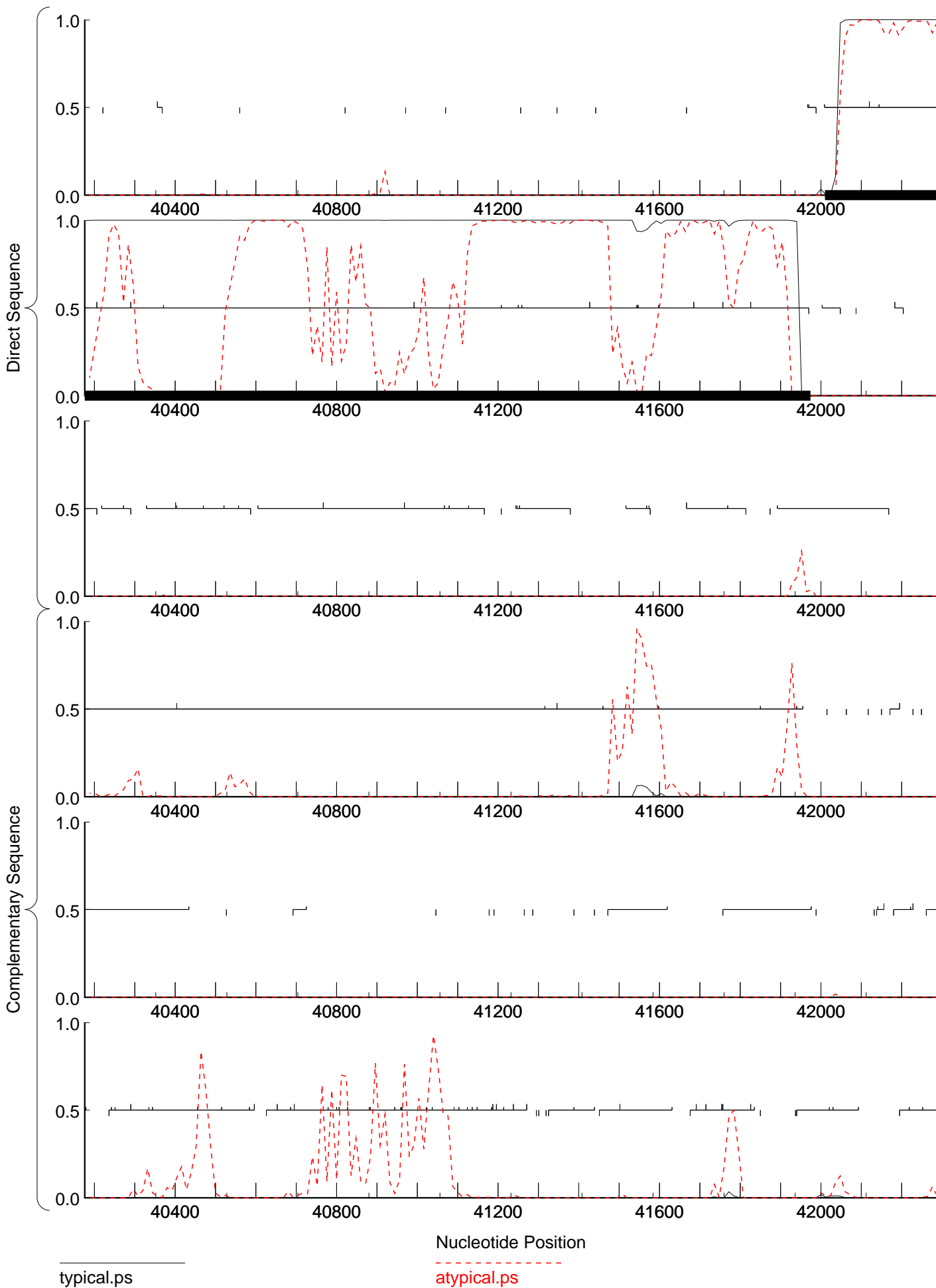


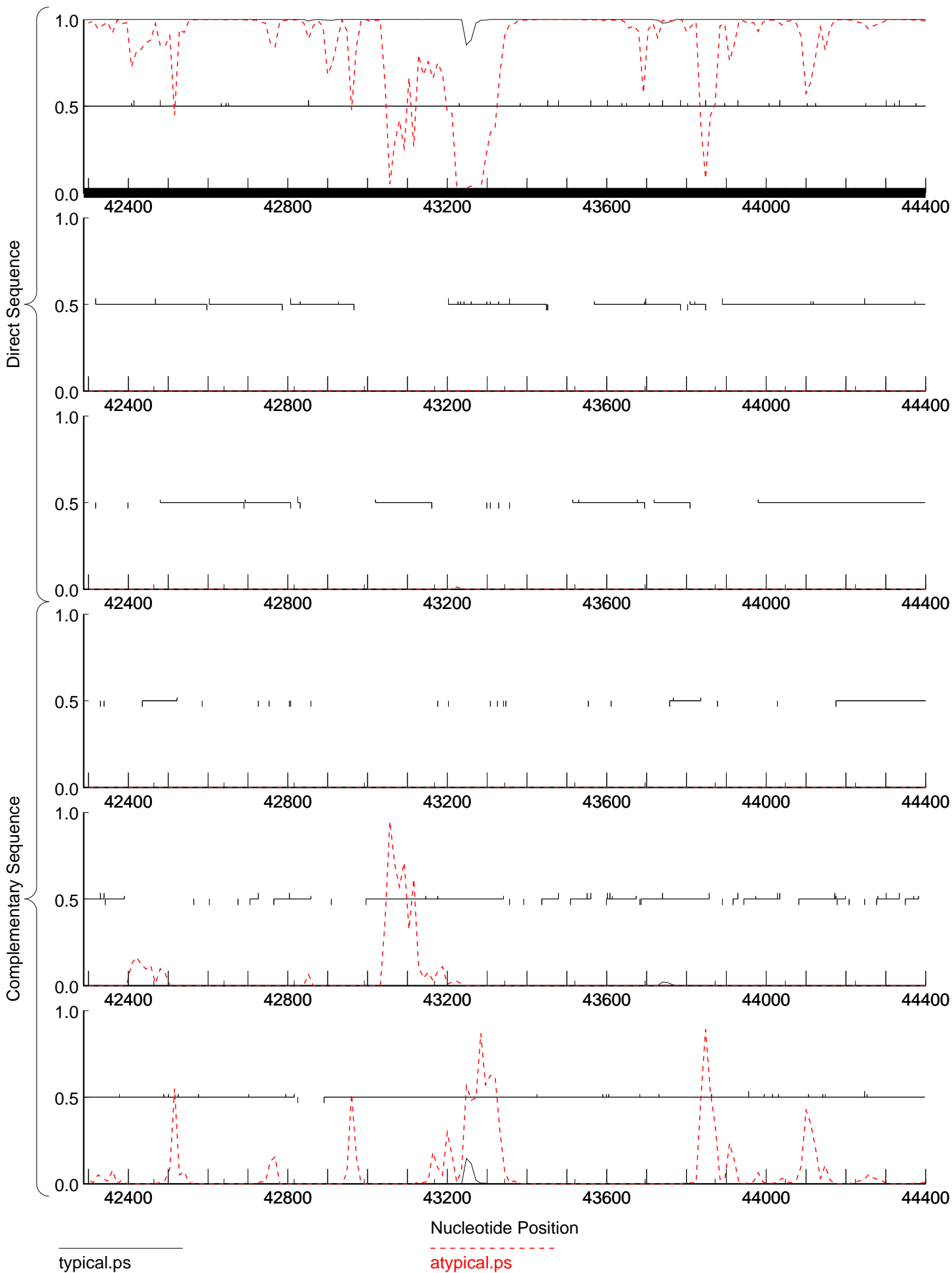
typical.ps

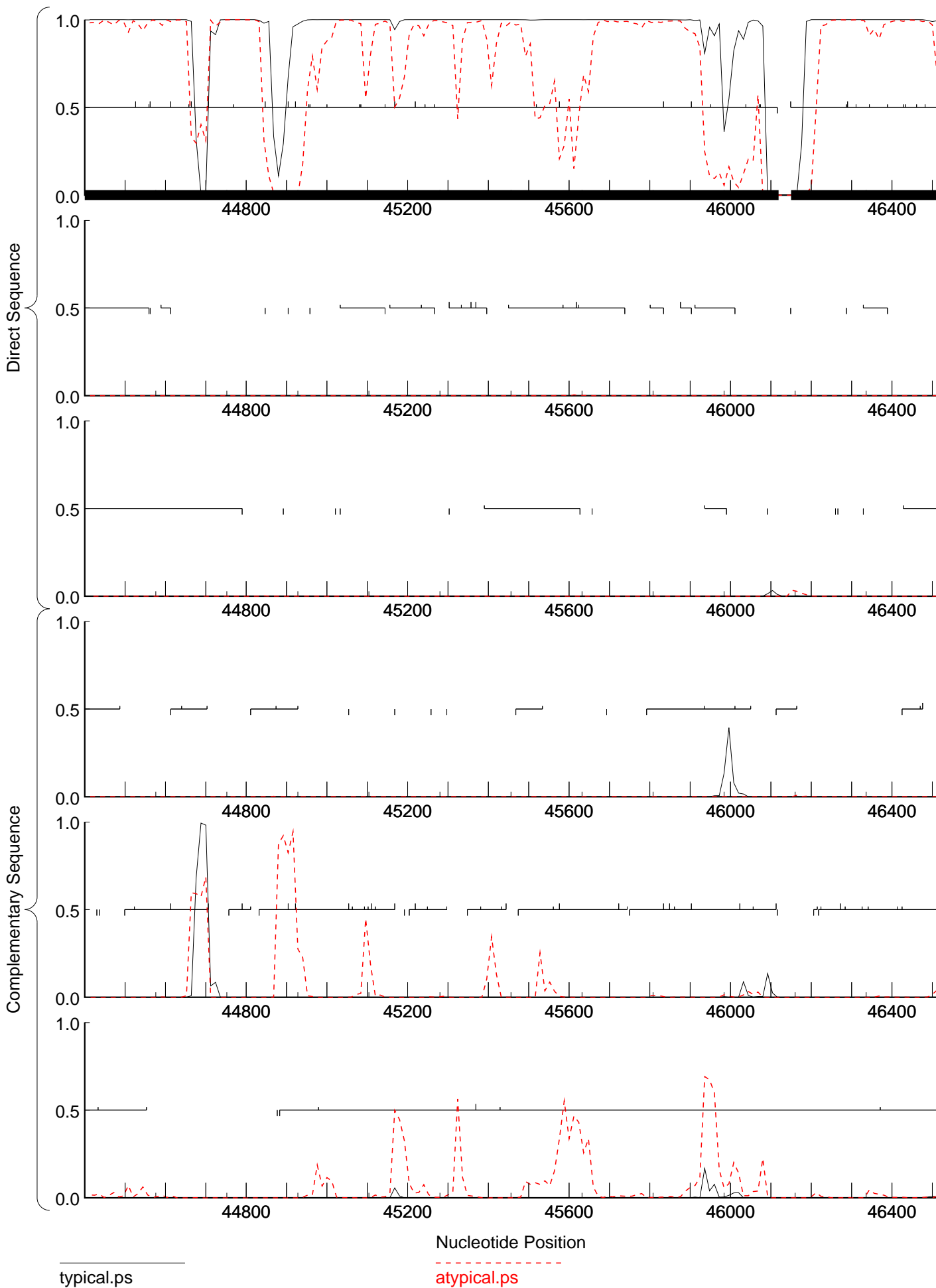
atypical.ps

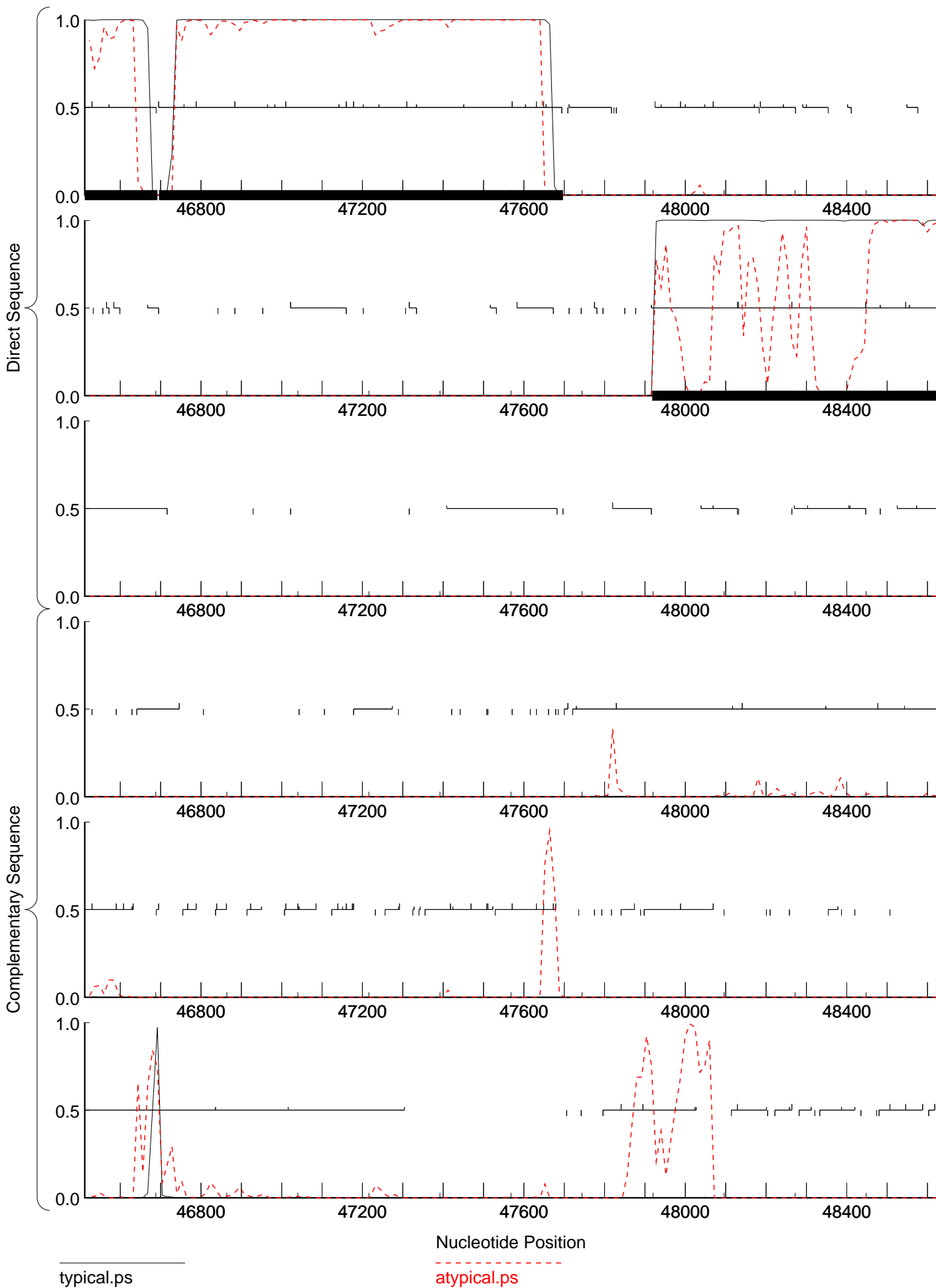




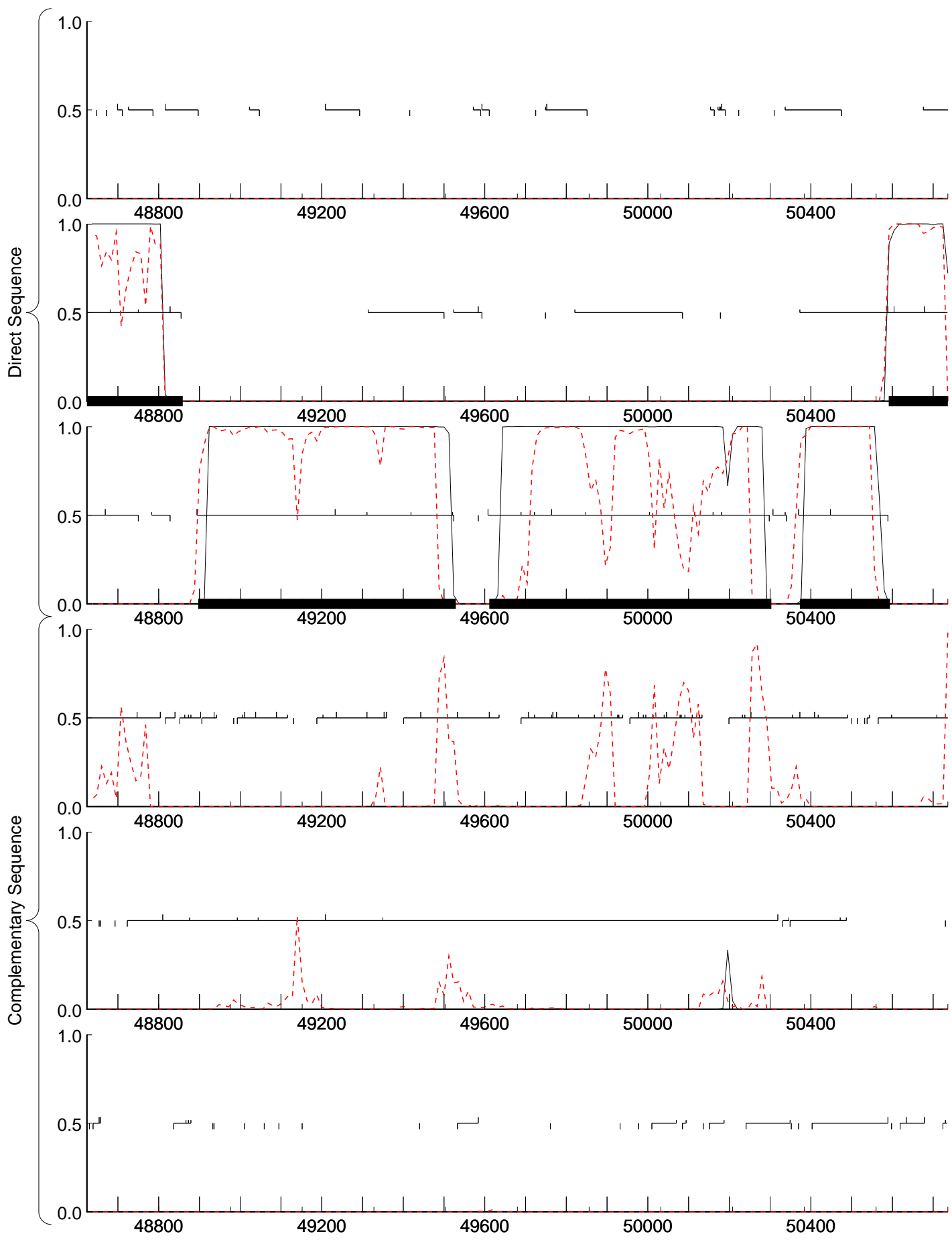






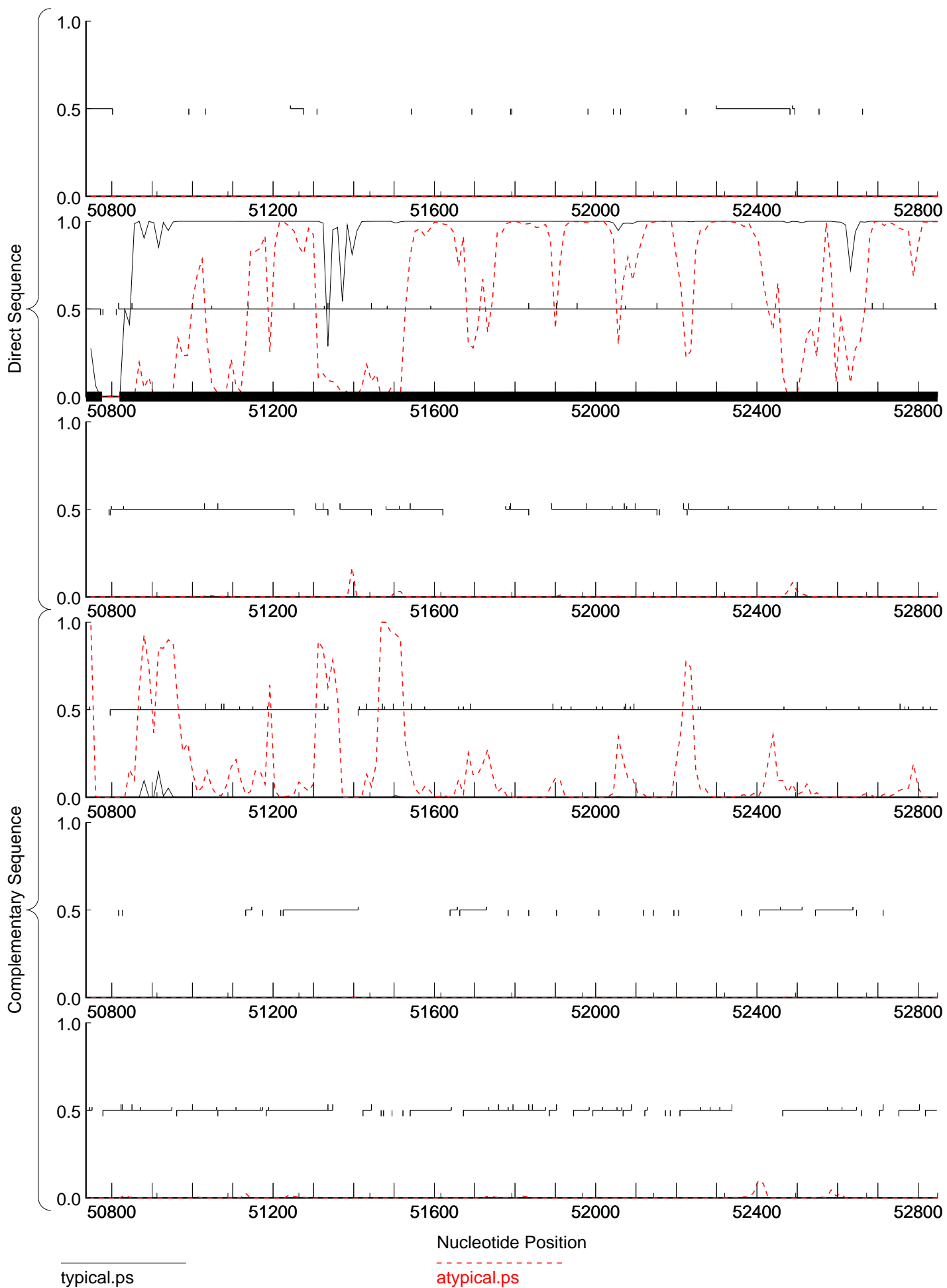


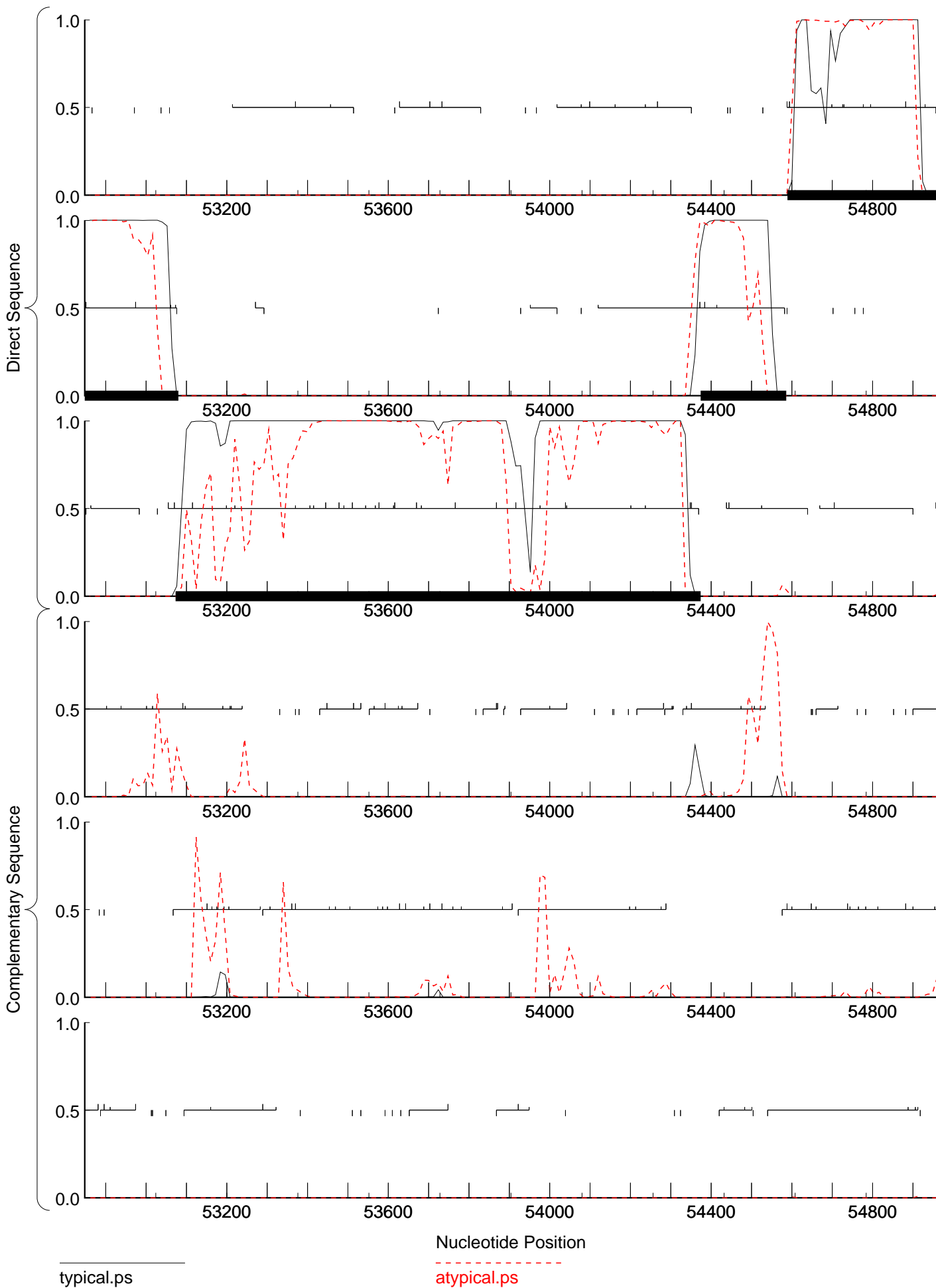
GeneMark.hmm prediction

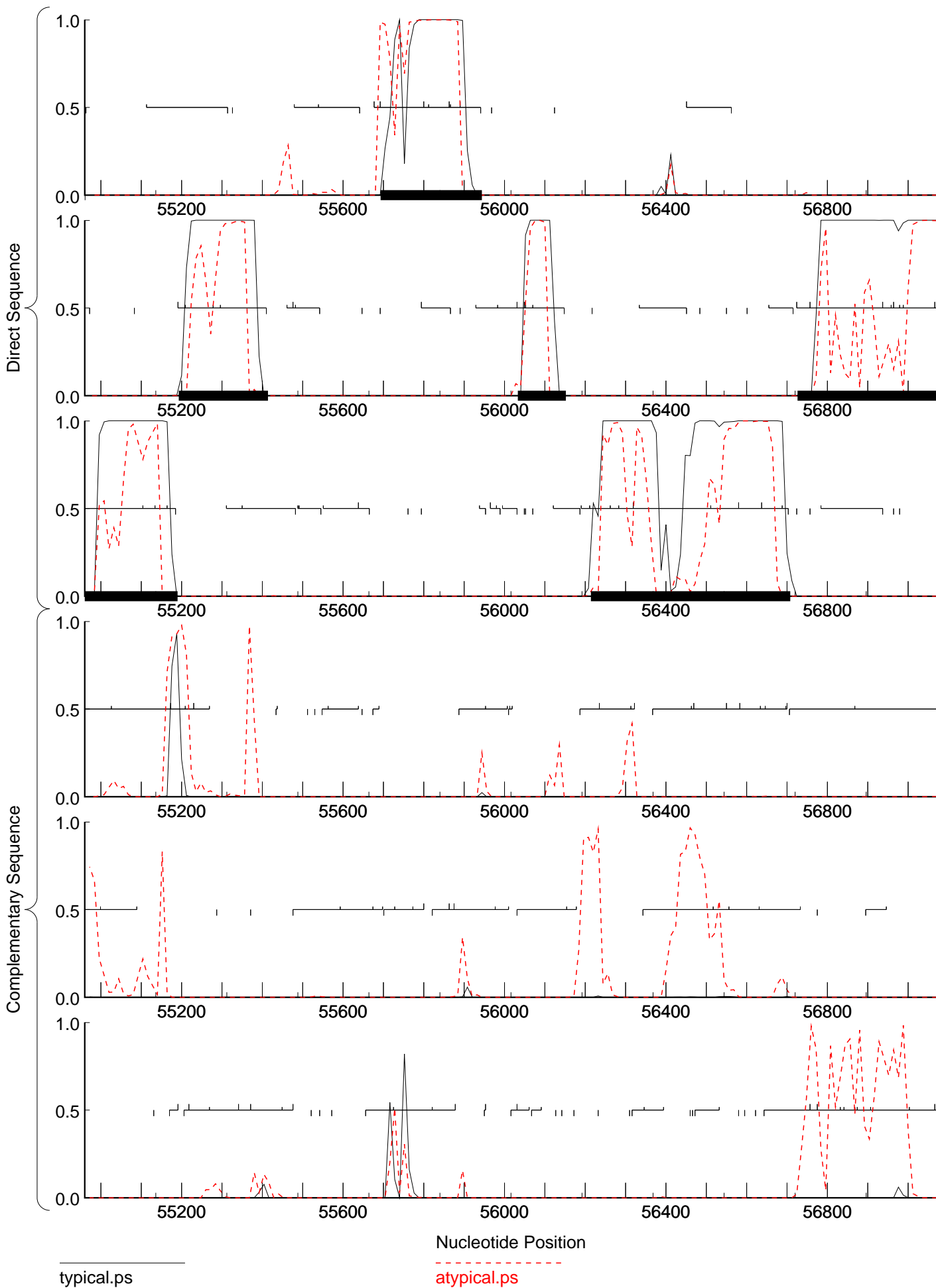


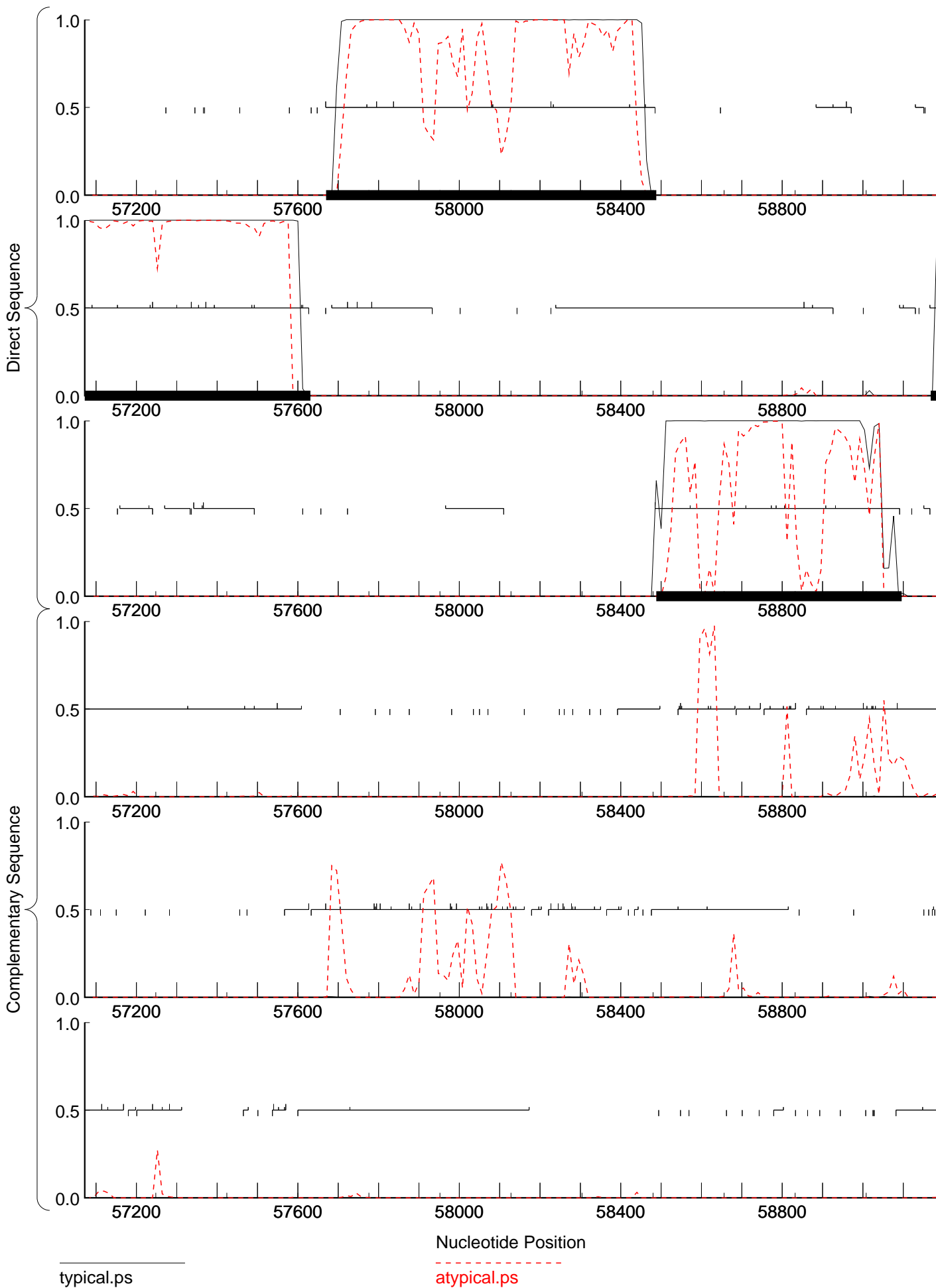
typical.ps

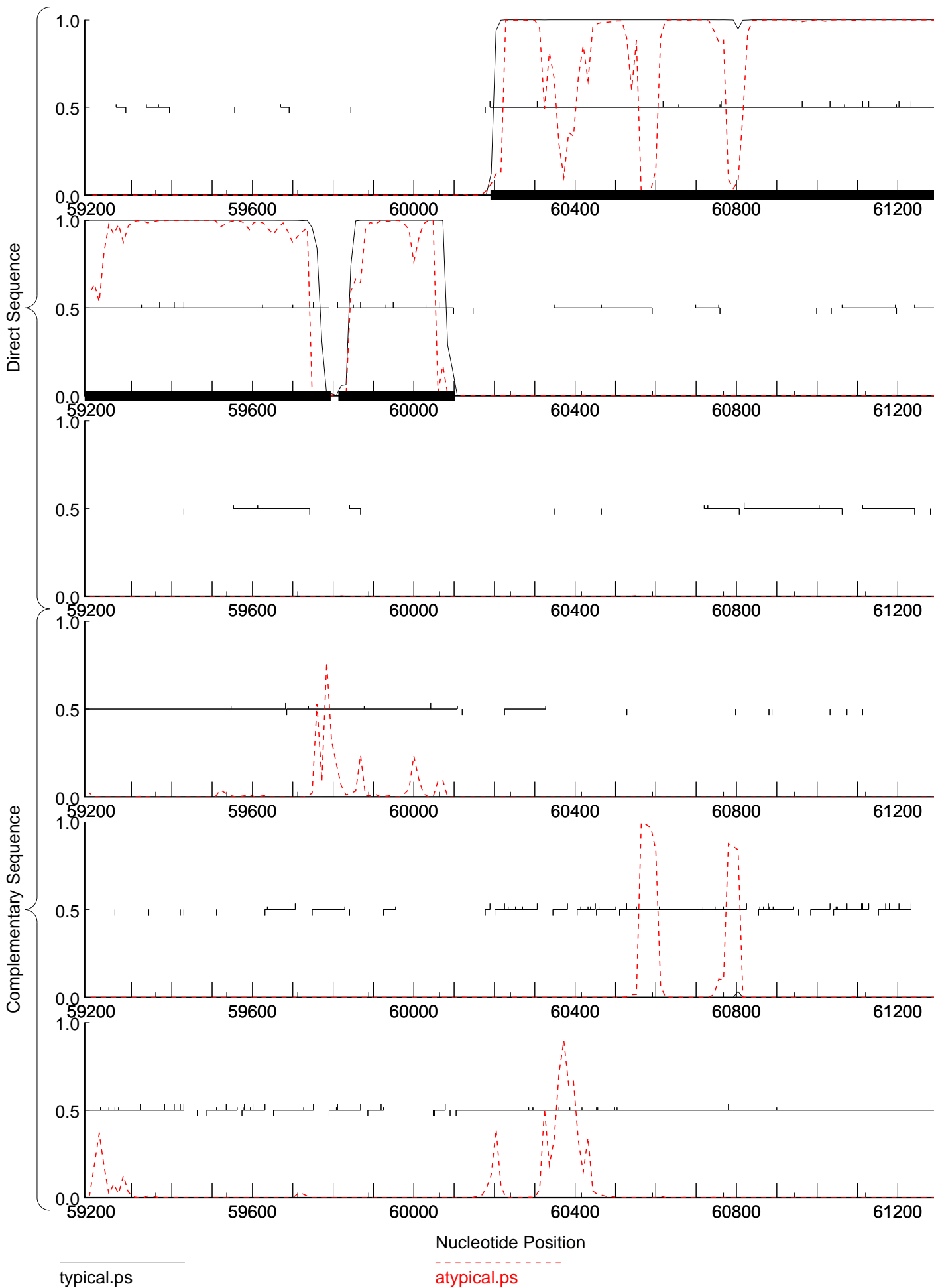
atypical.ps

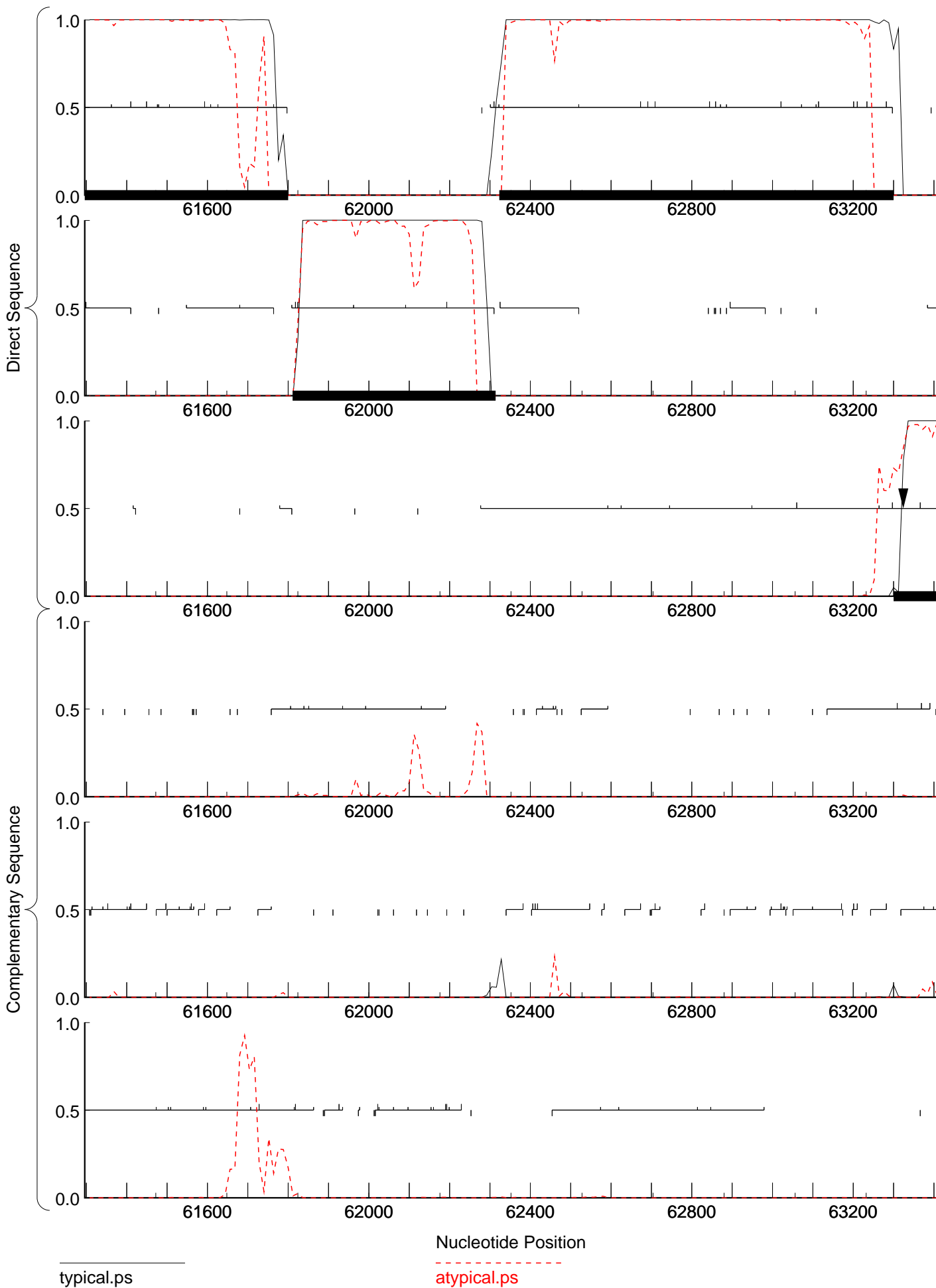


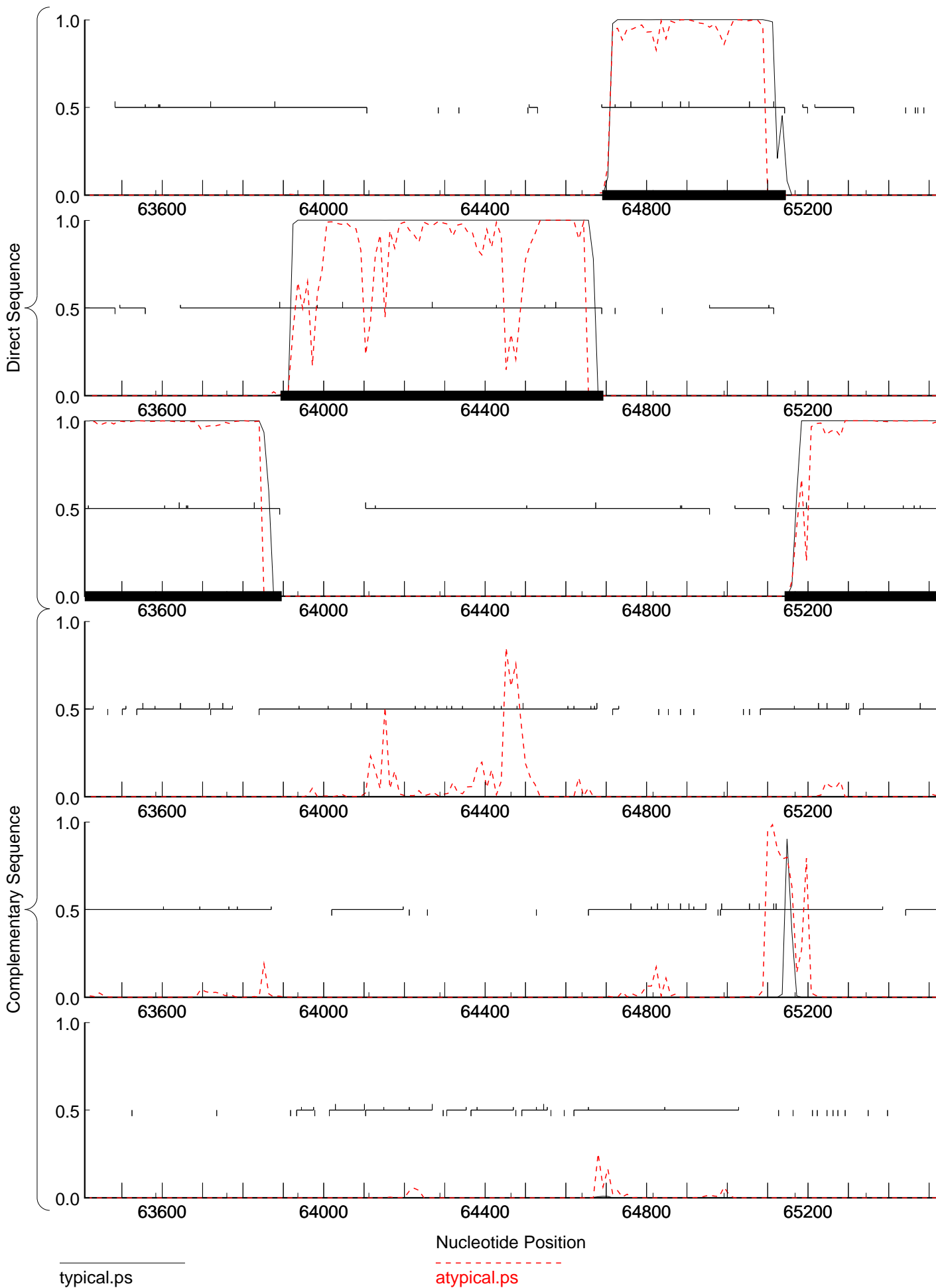


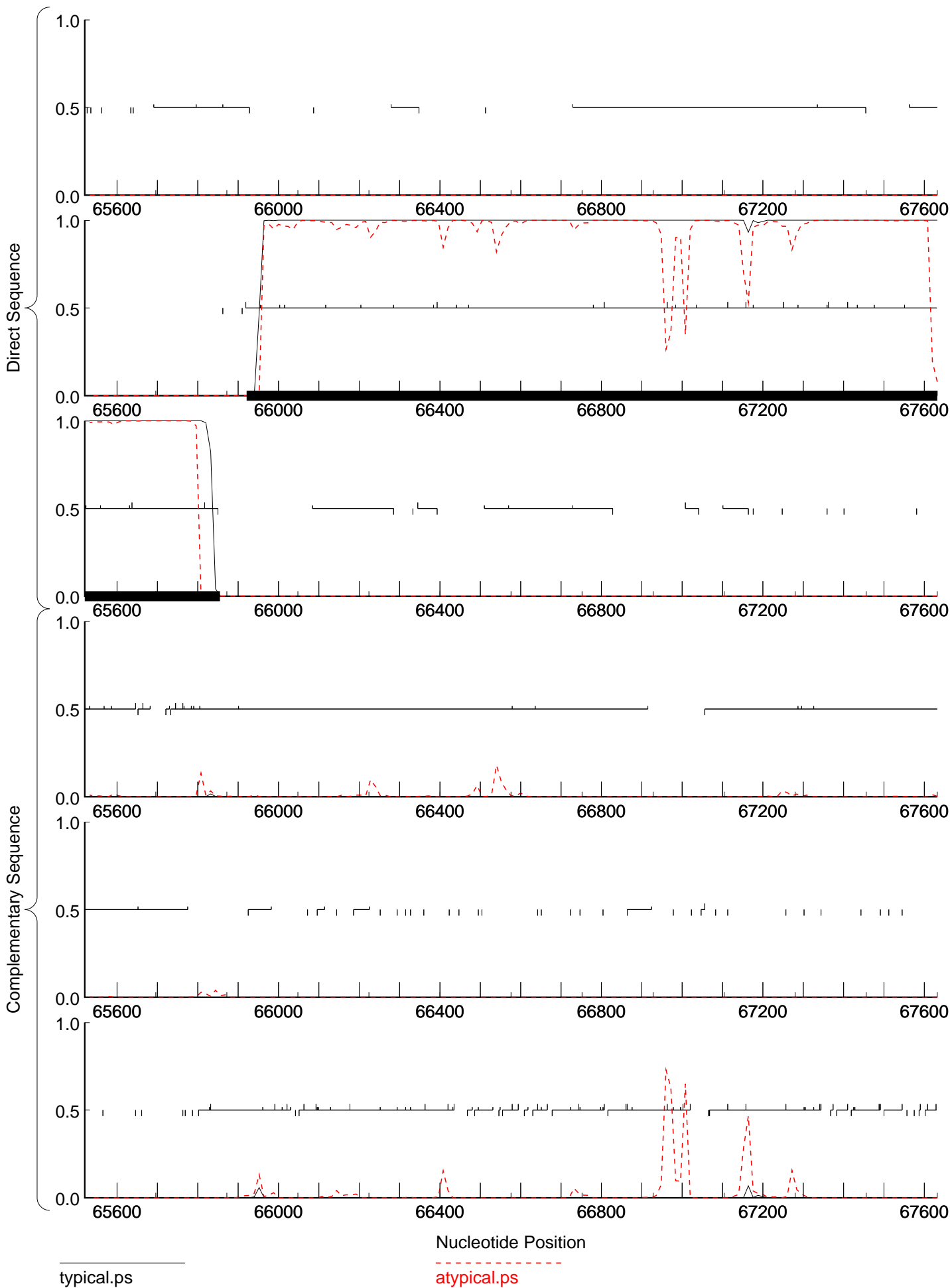


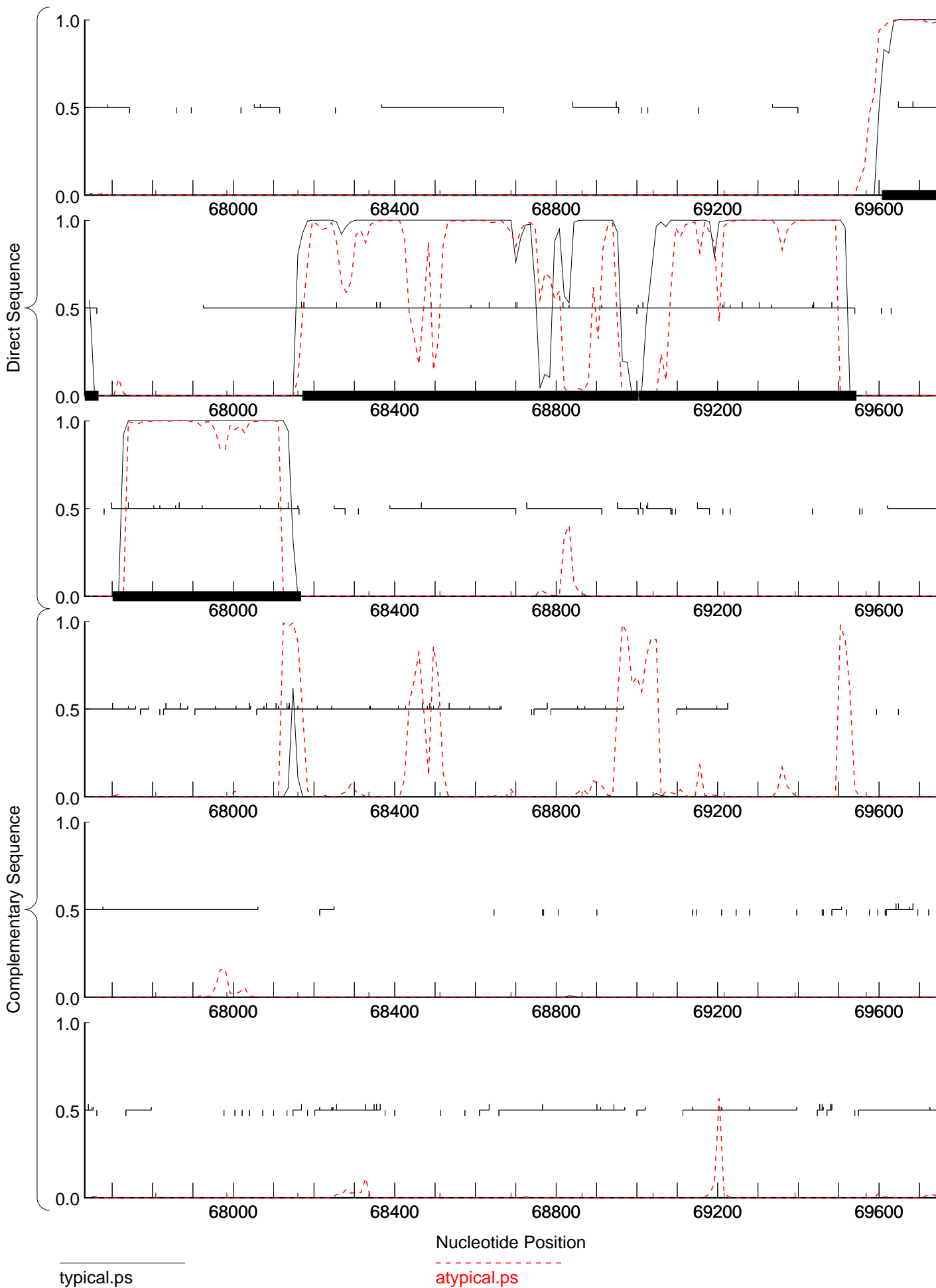






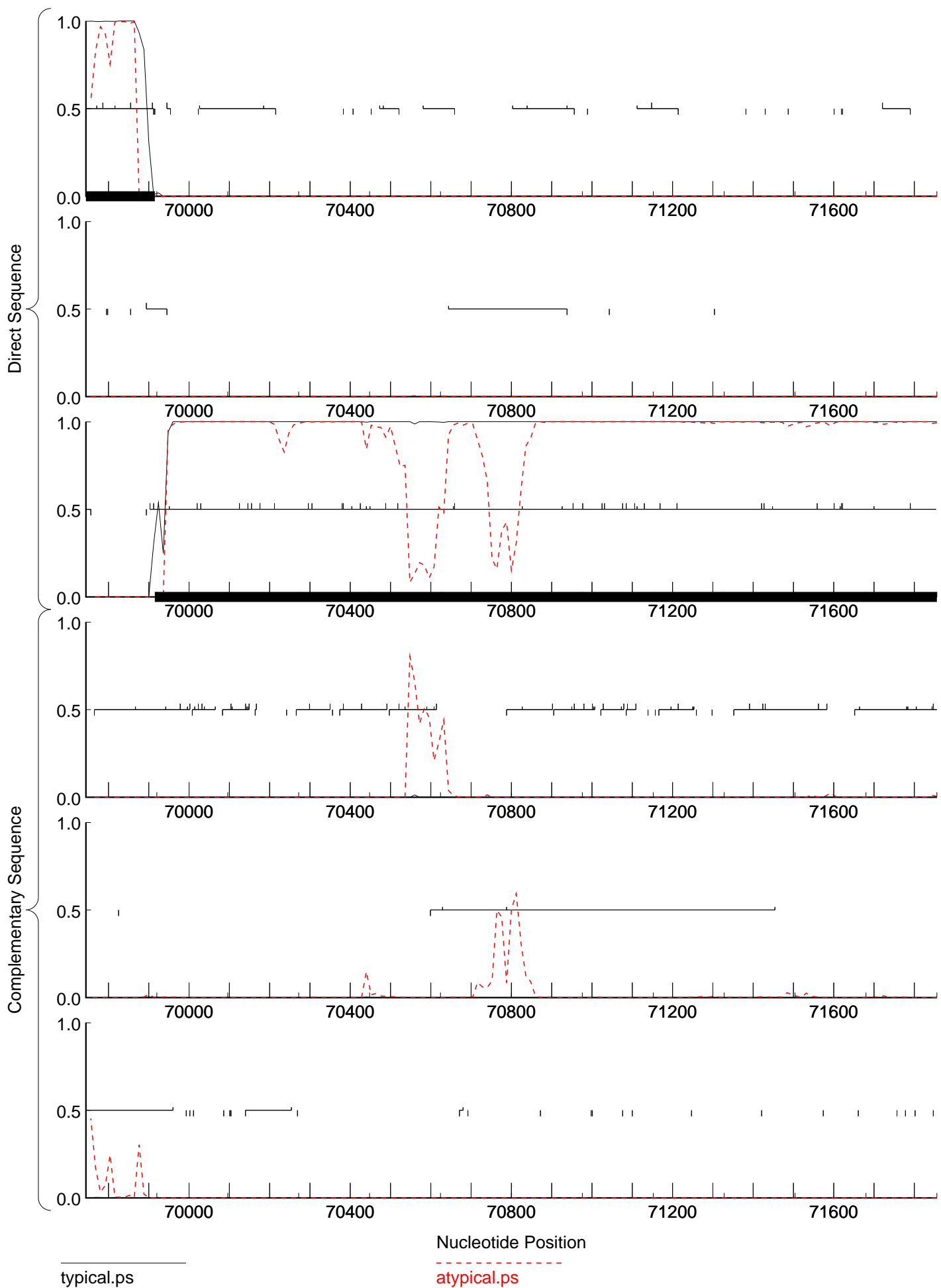


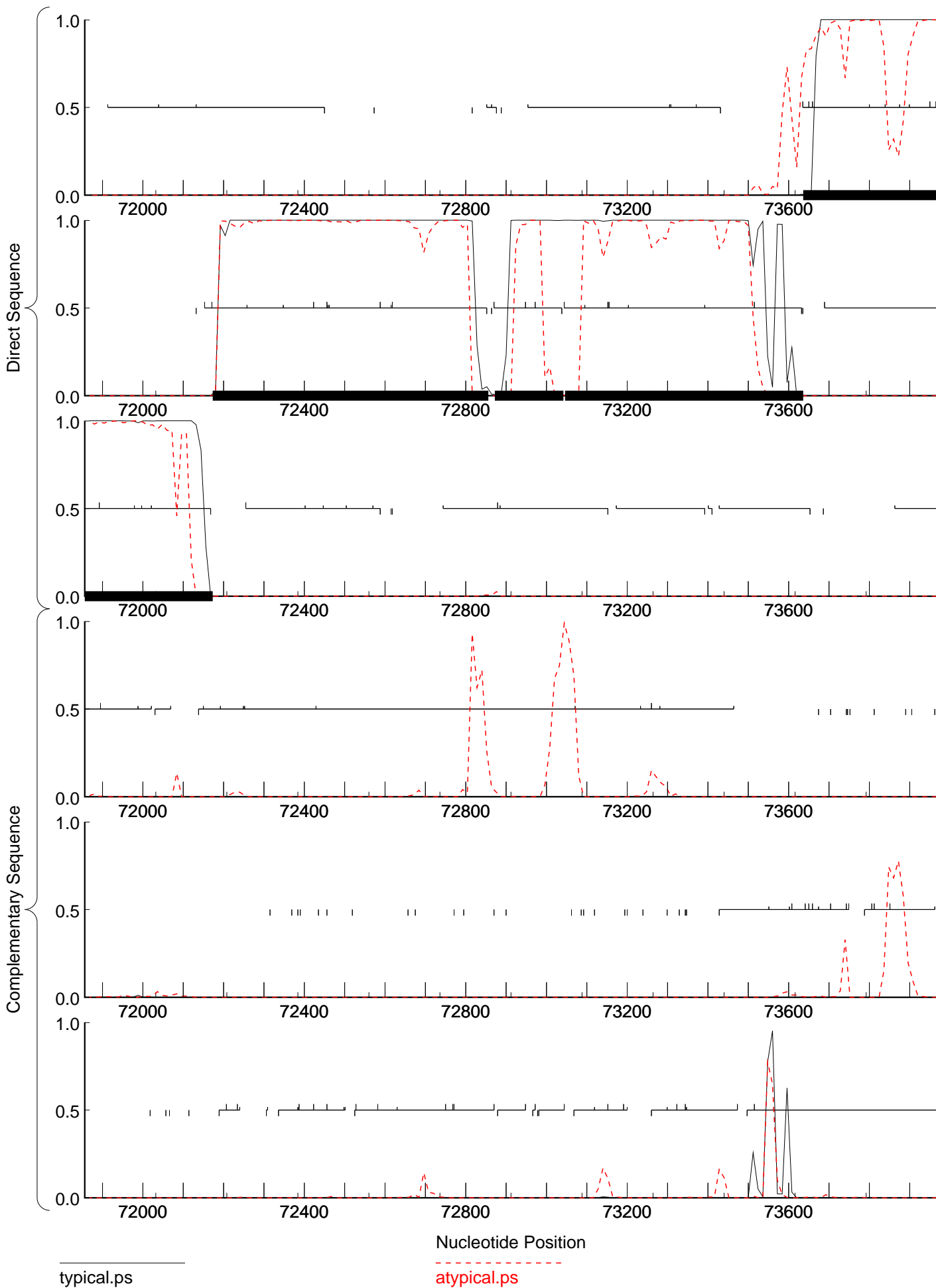


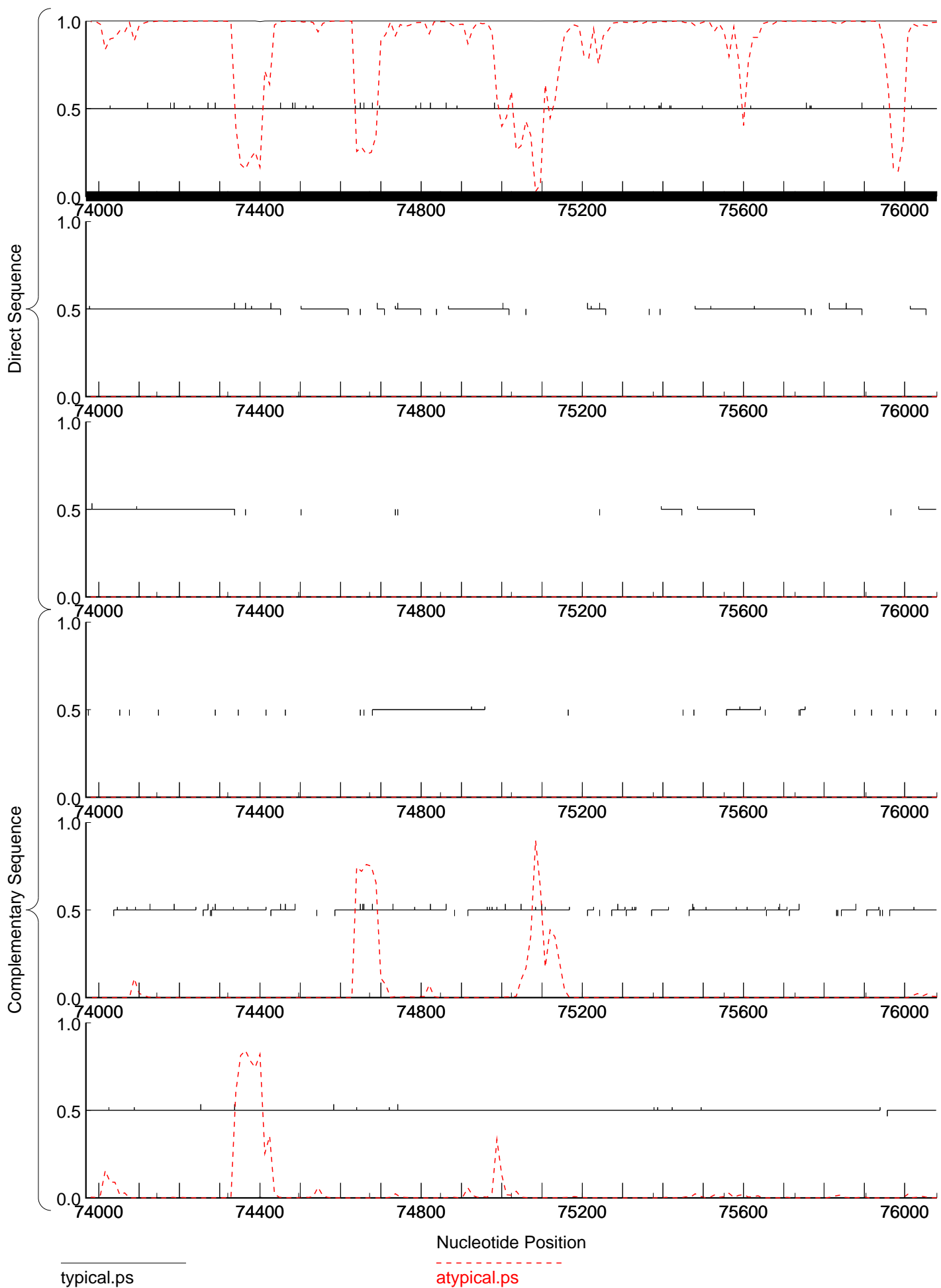


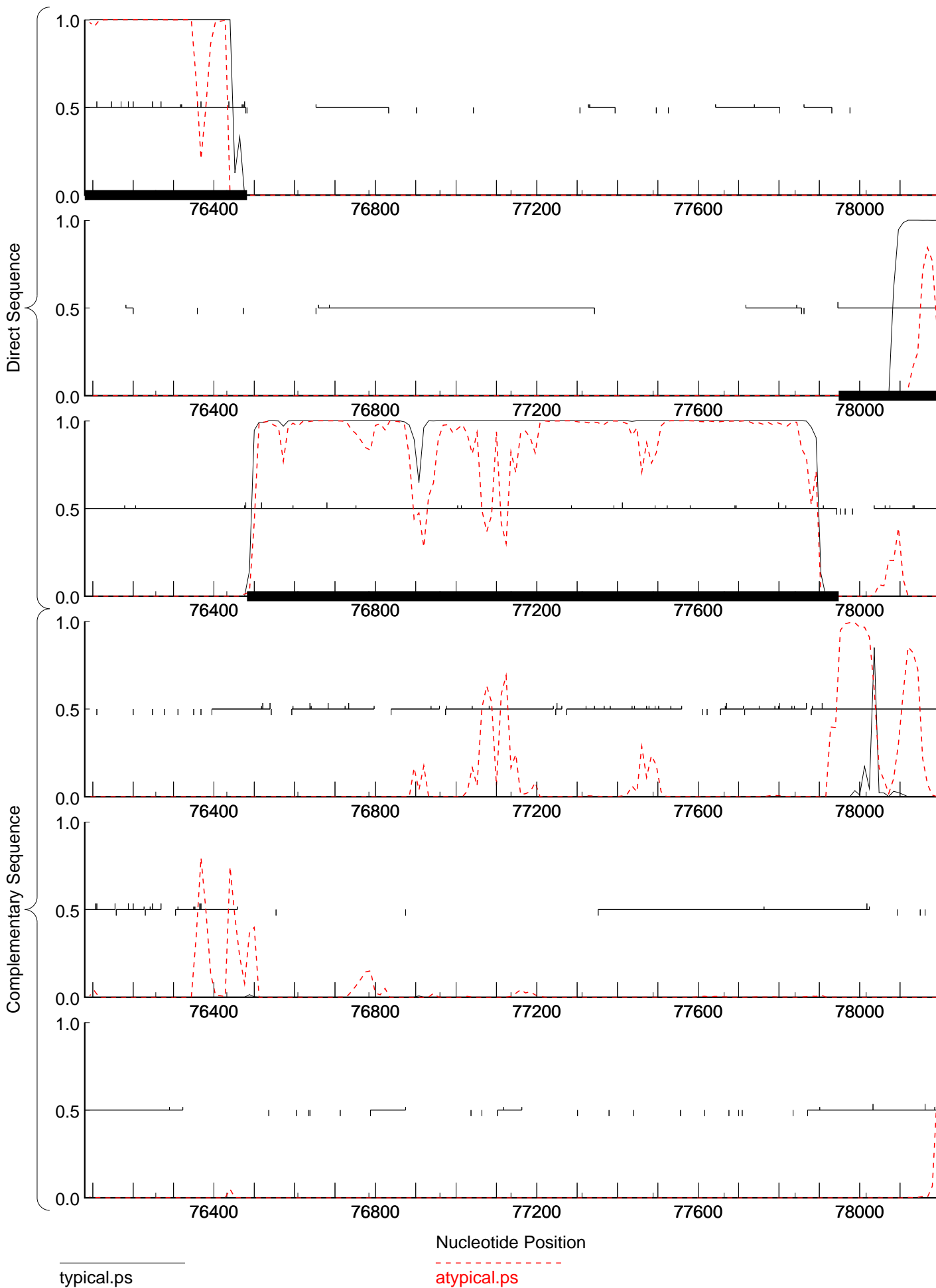
typical.ps

atypical.ps



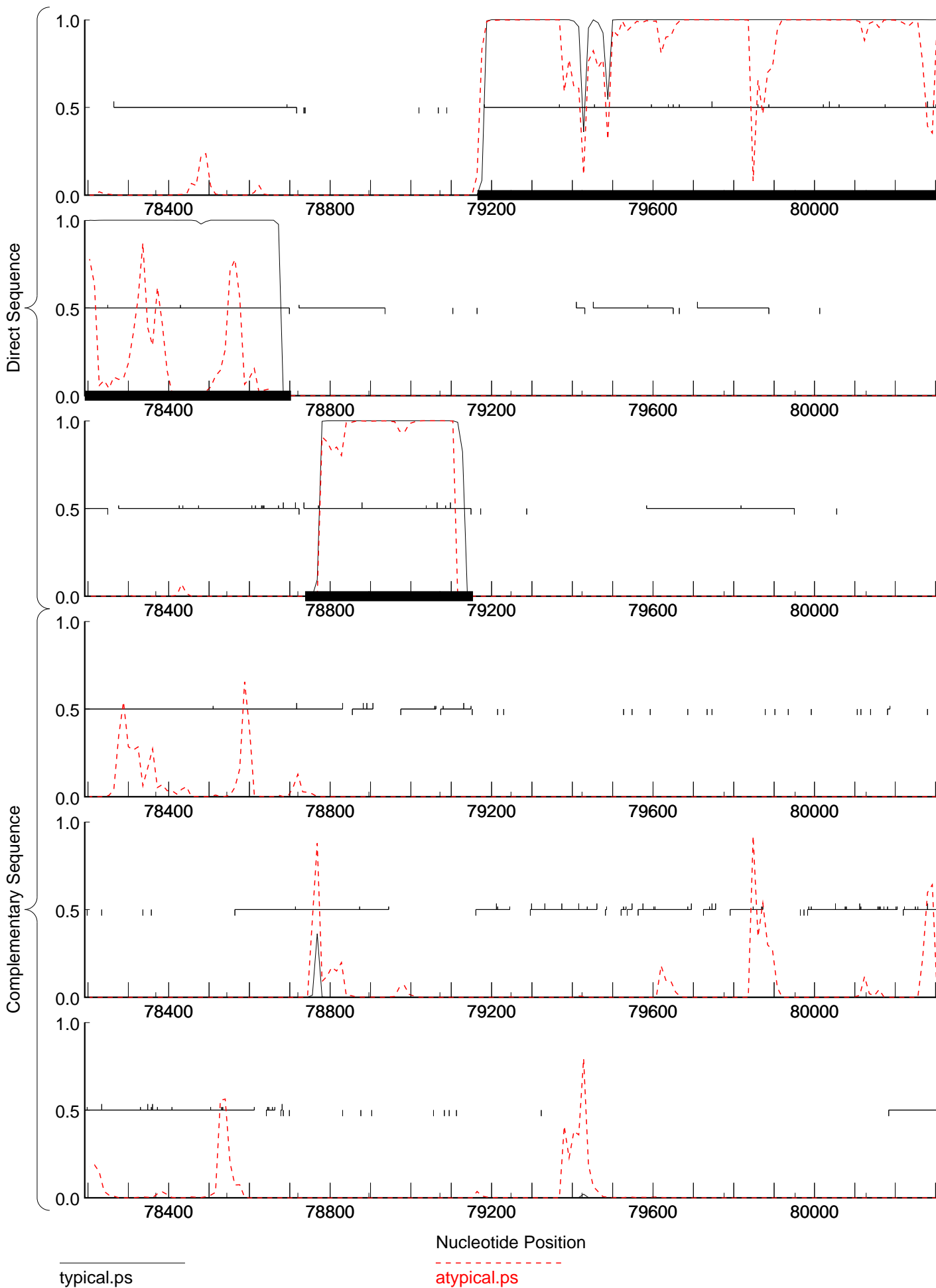


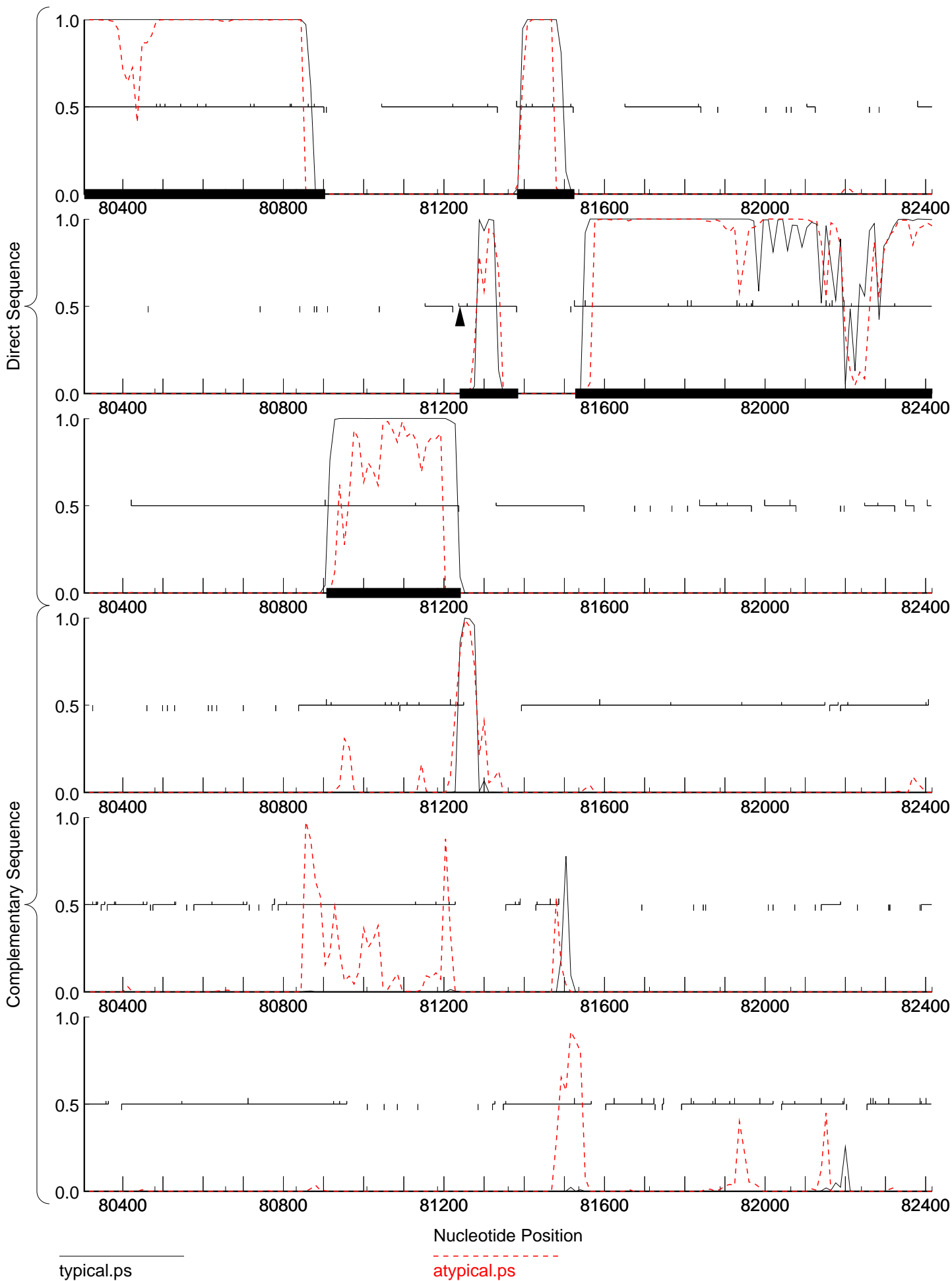


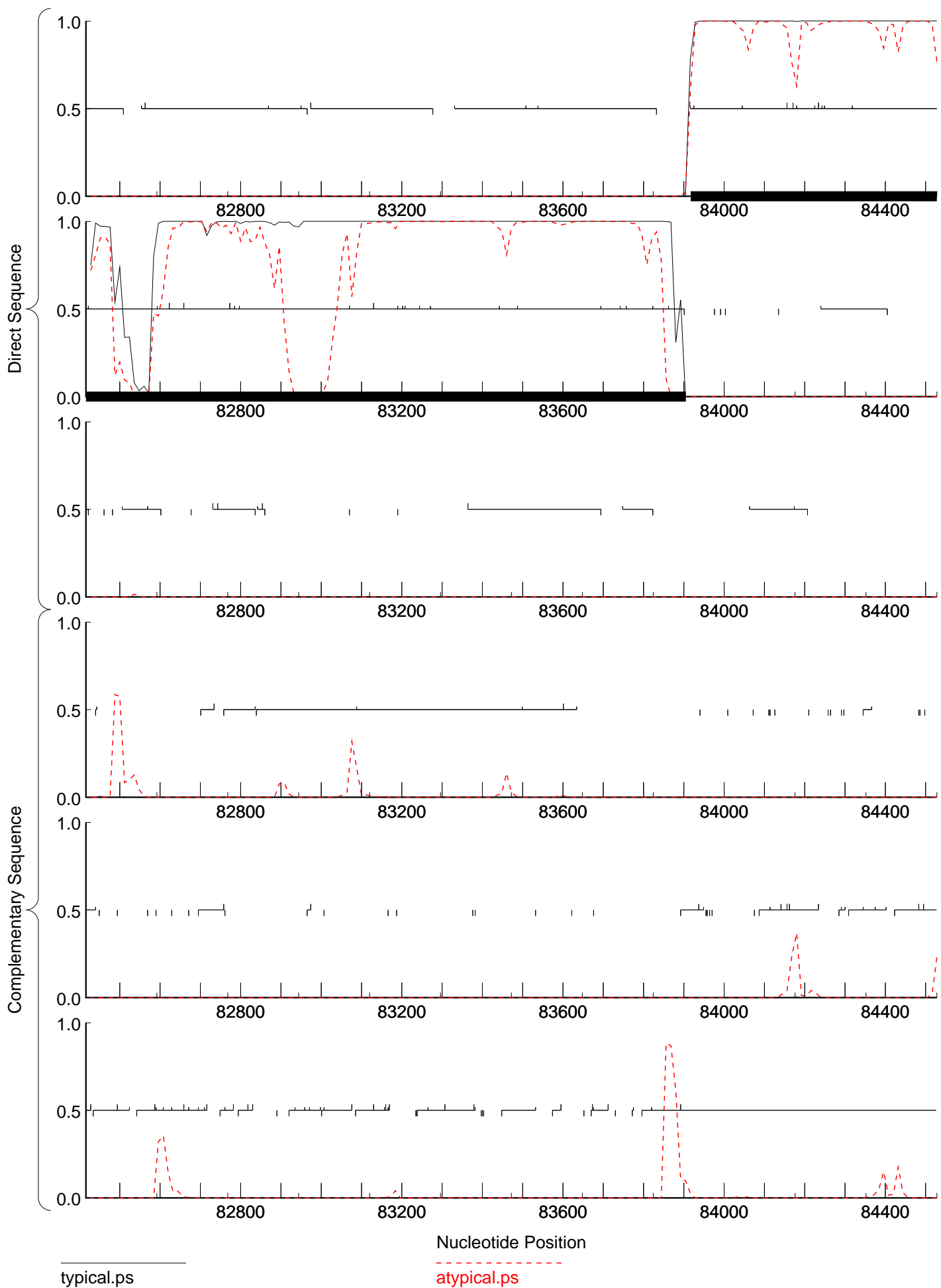


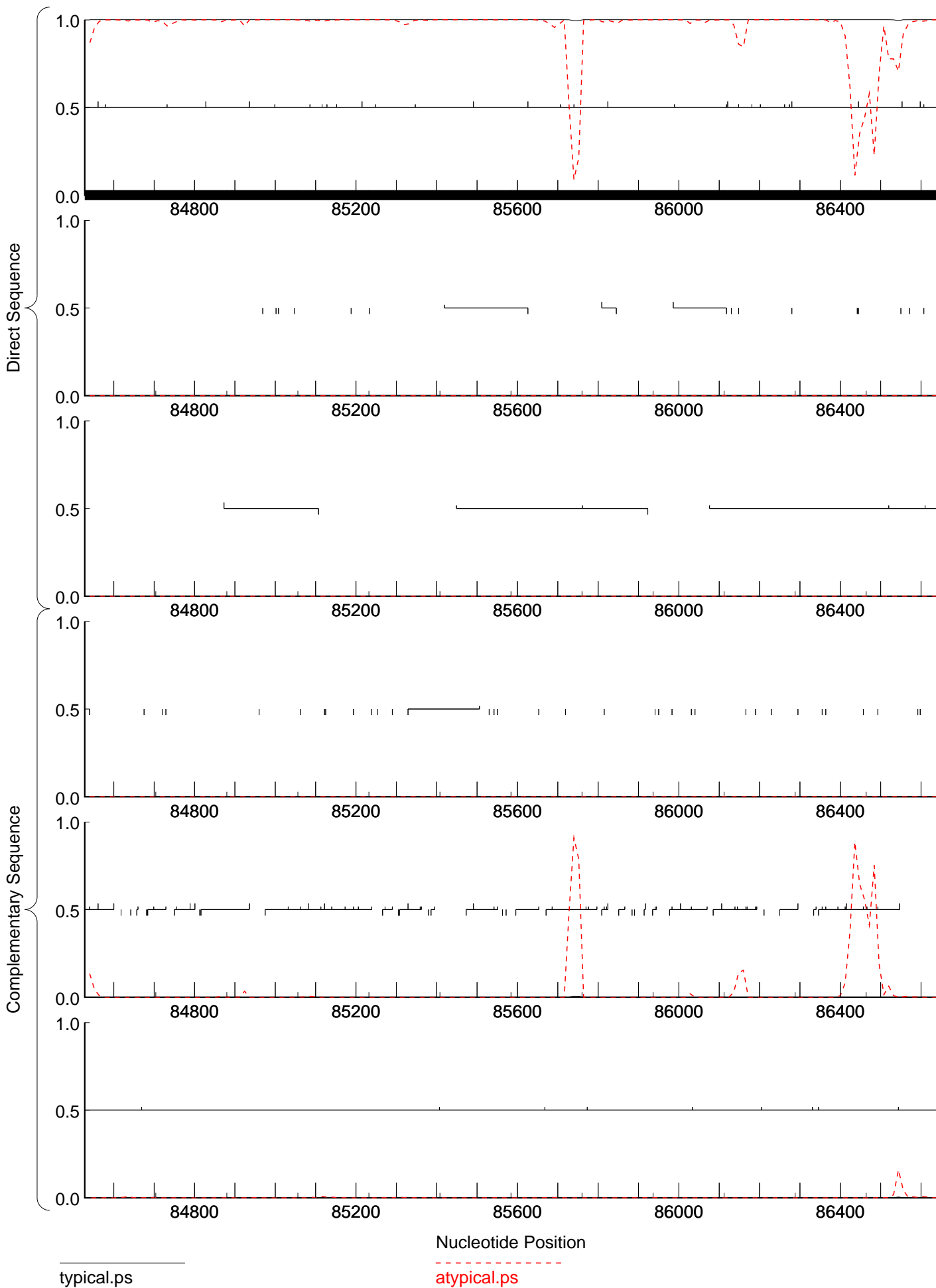
typical.ps

atypical.ps



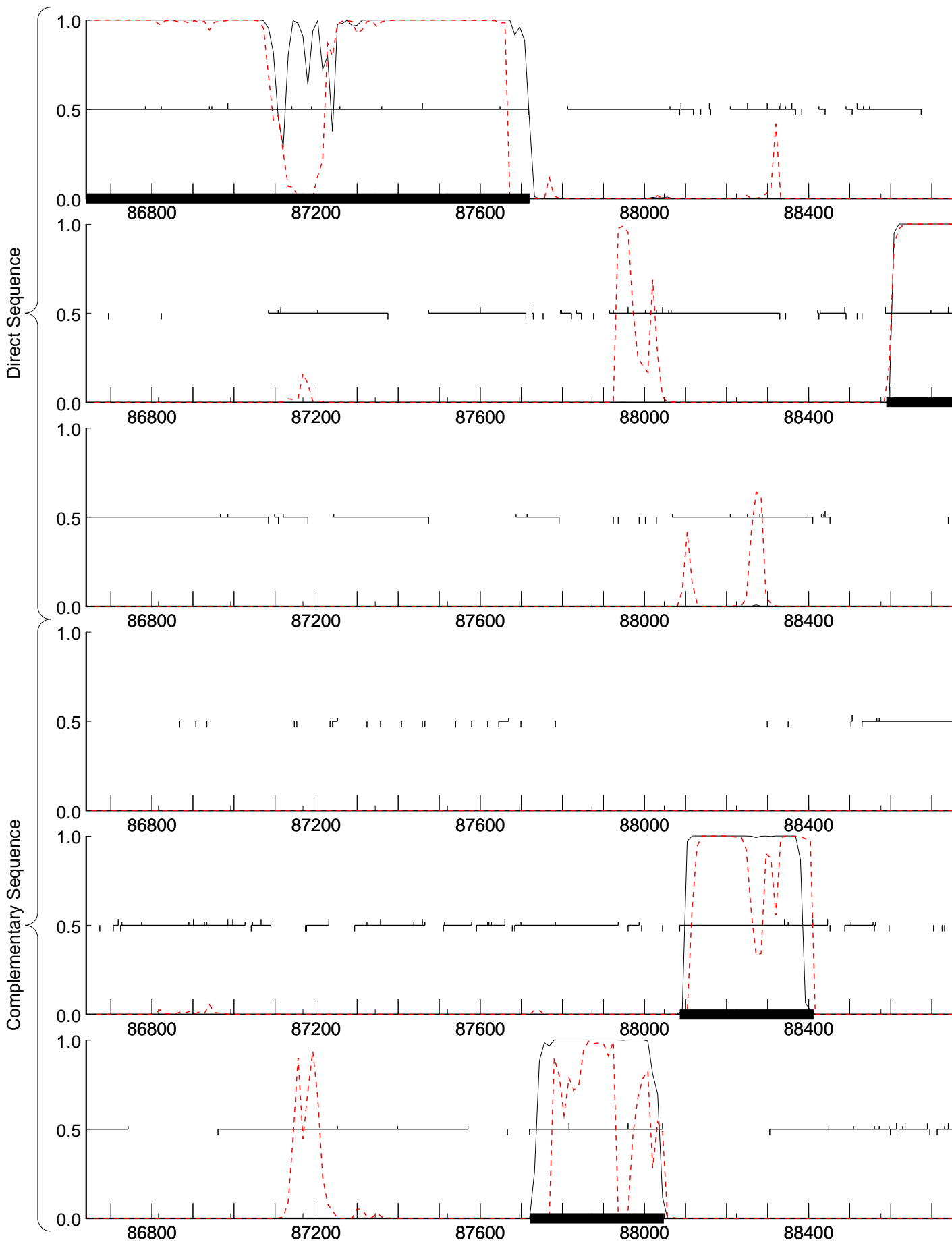






typical.ps

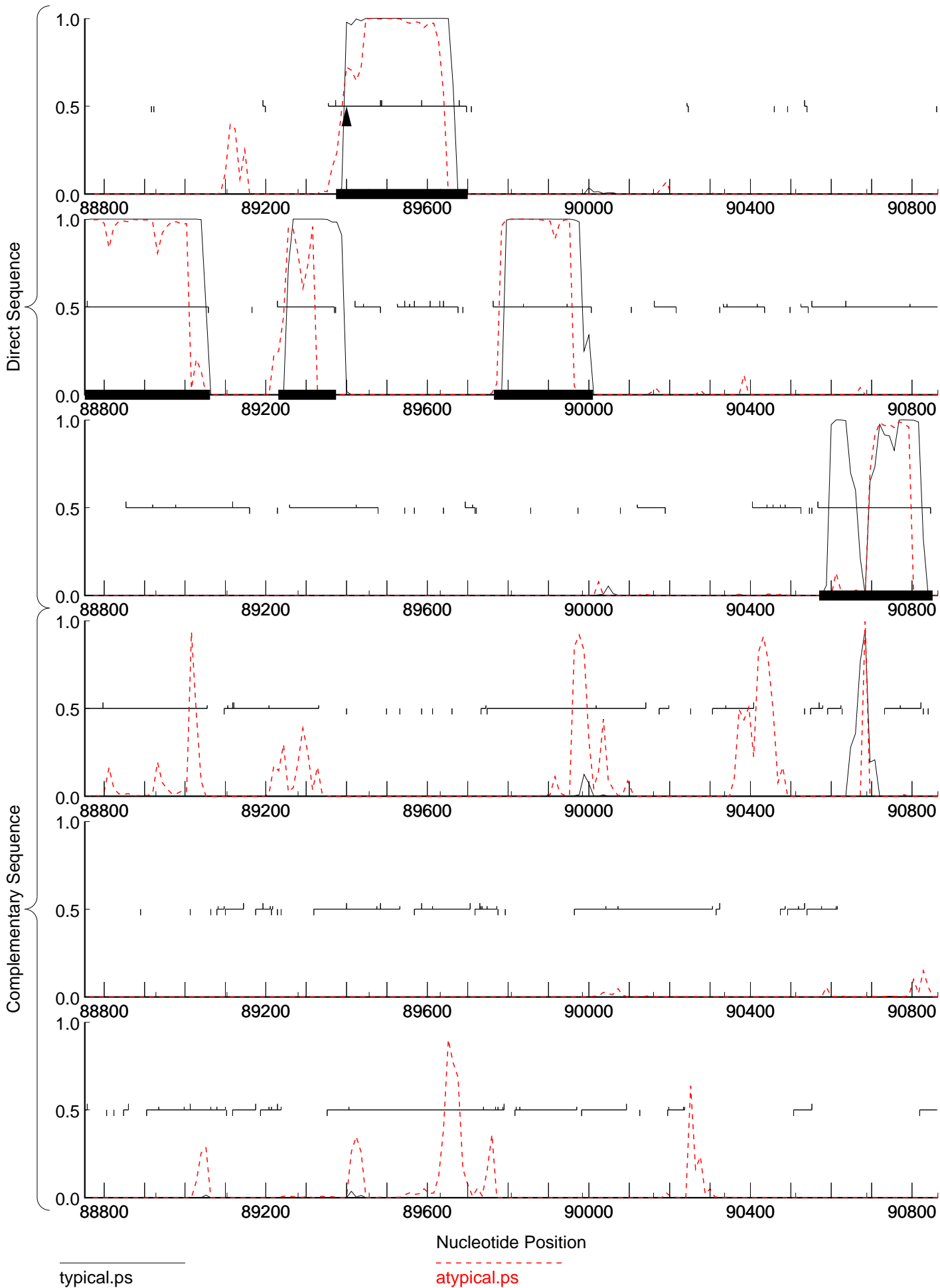
atypical.ps

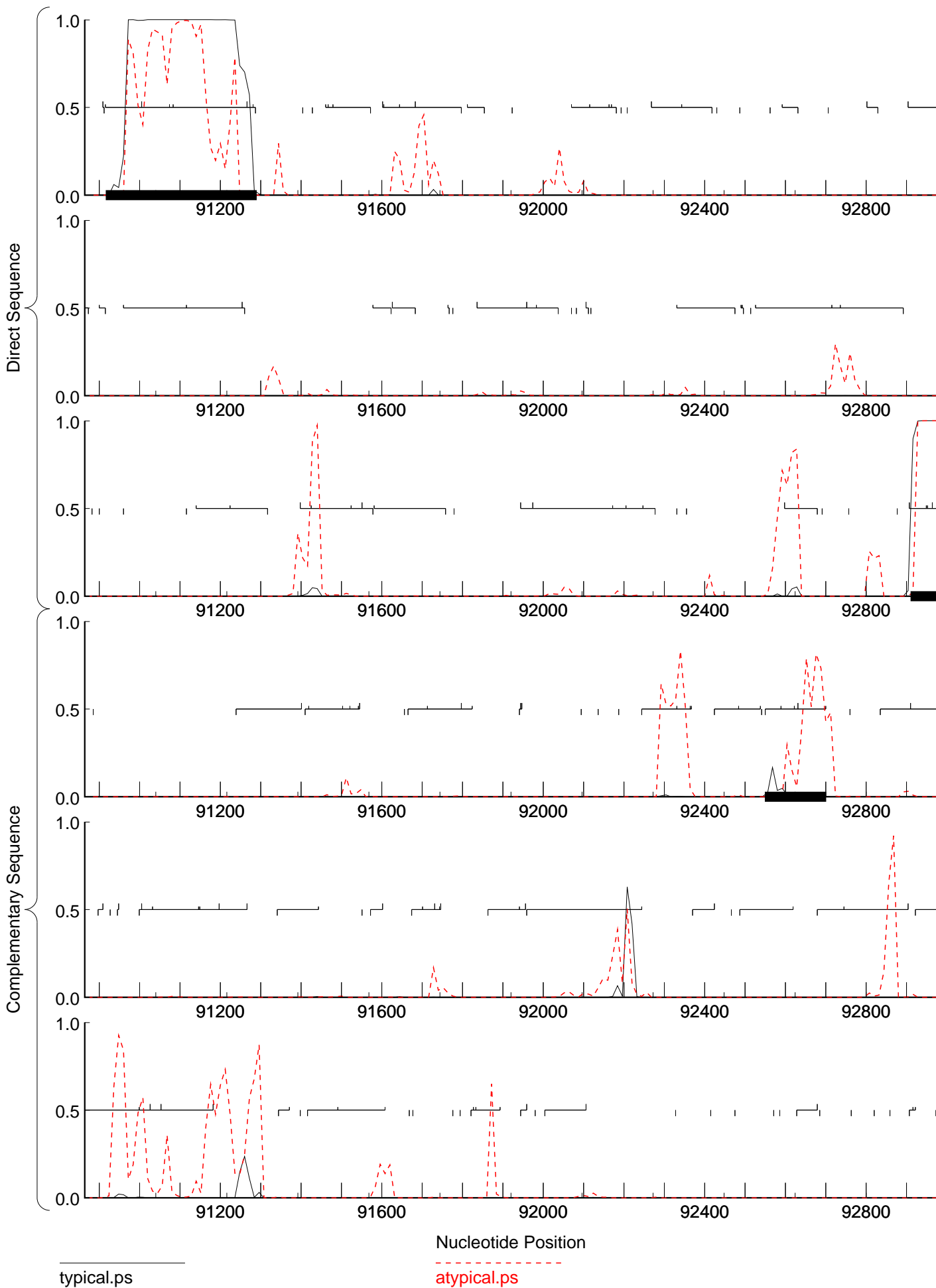


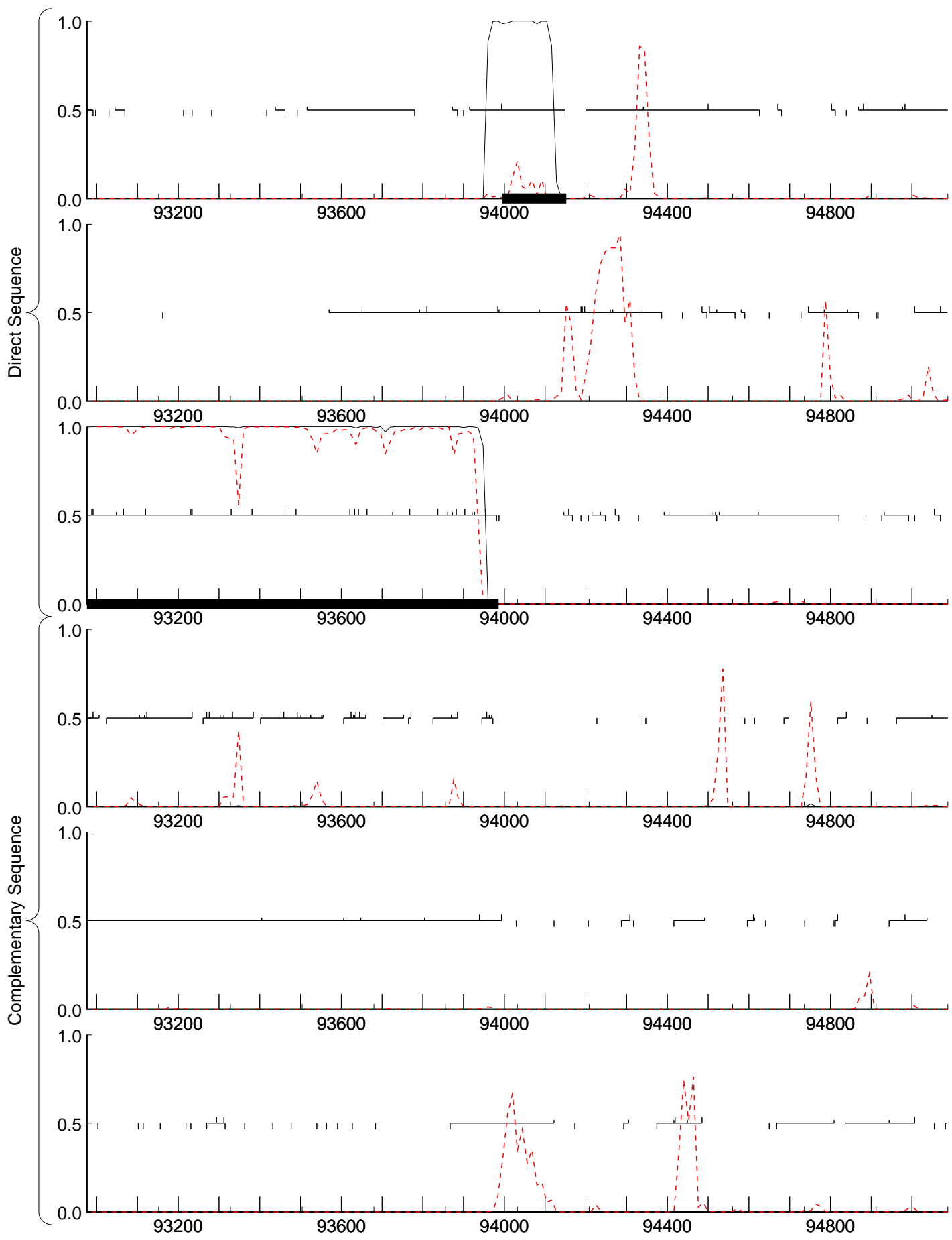
typical.ps

atypical.ps

GeneMark.hmm prediction

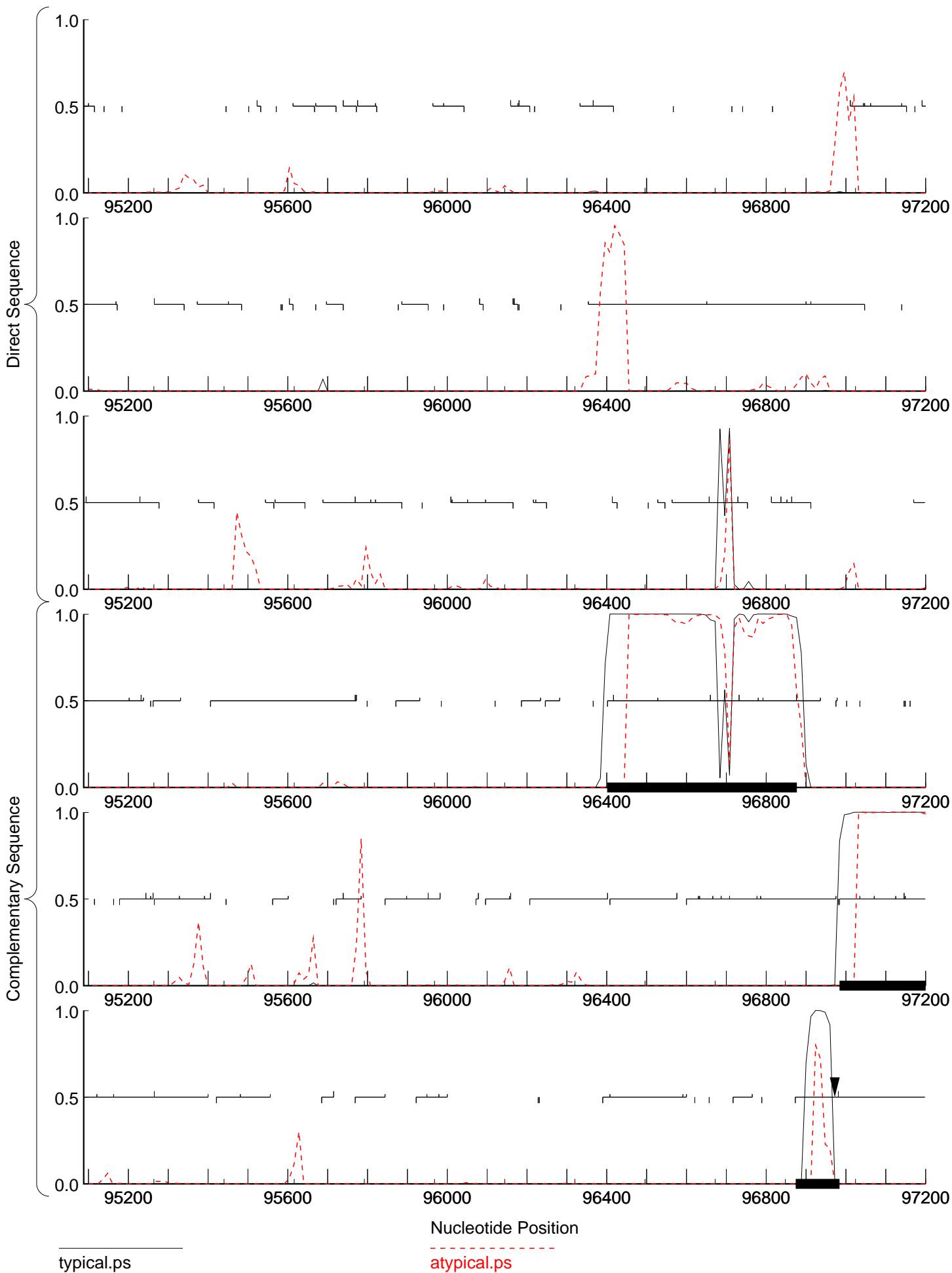


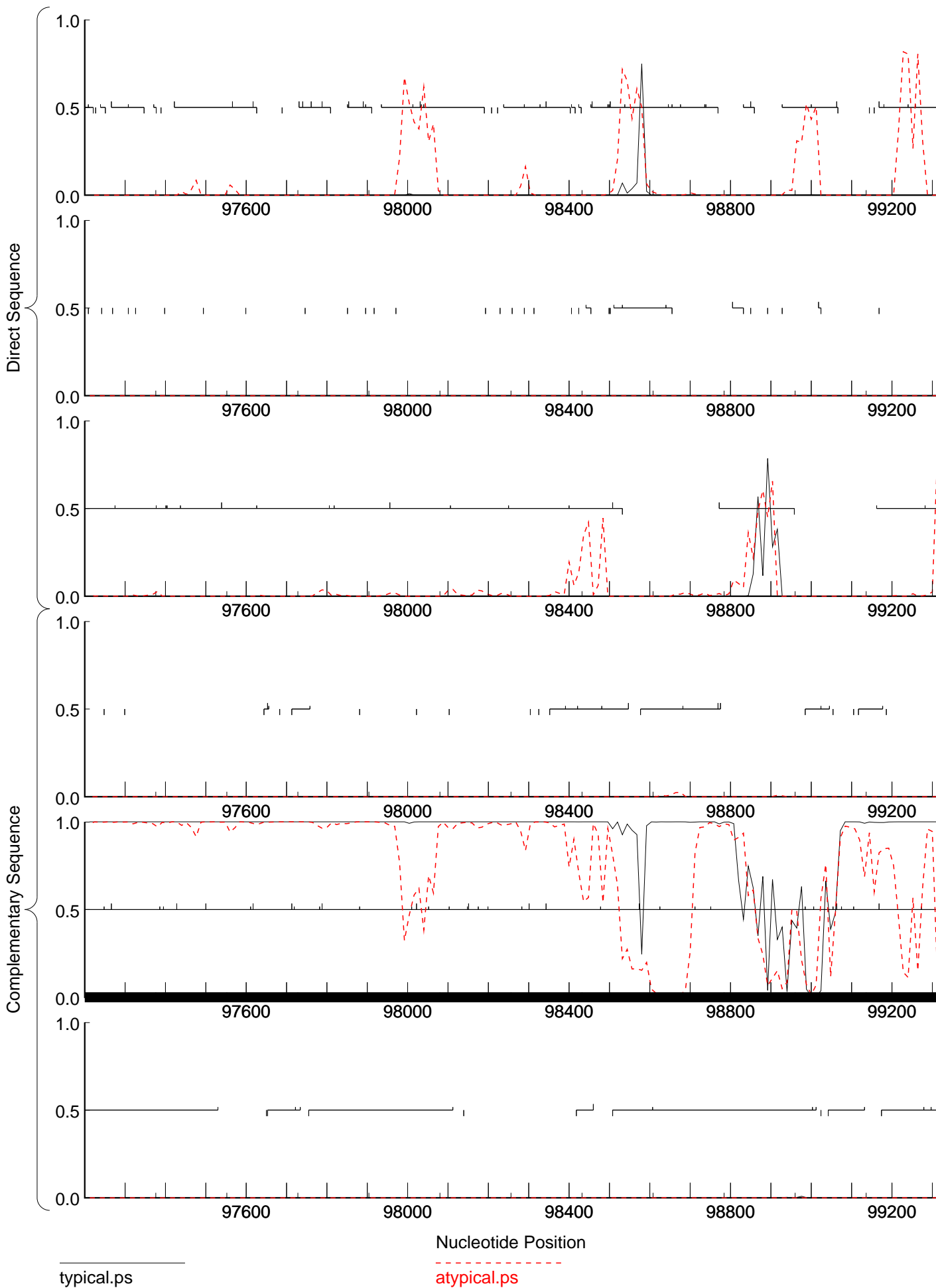


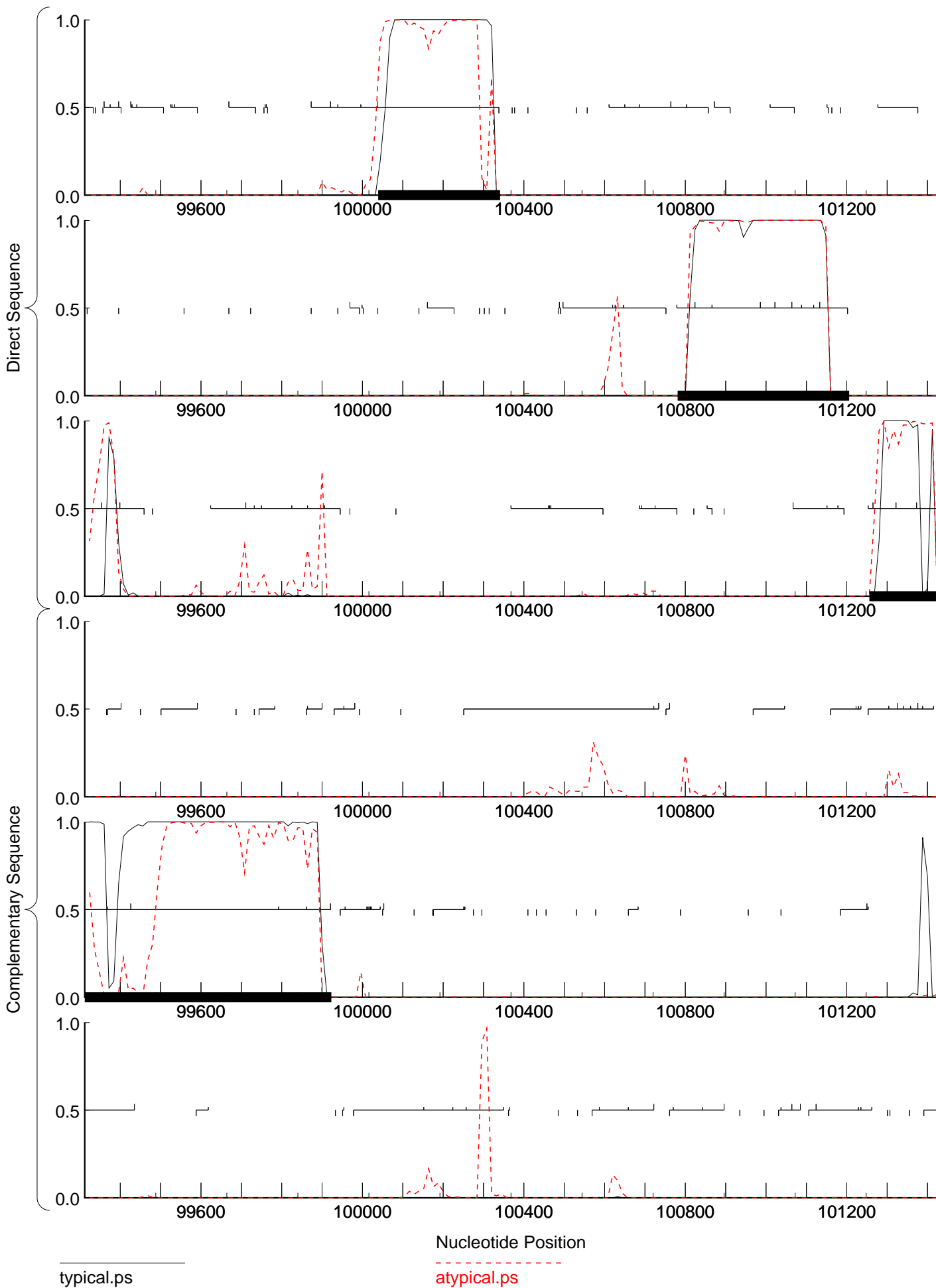


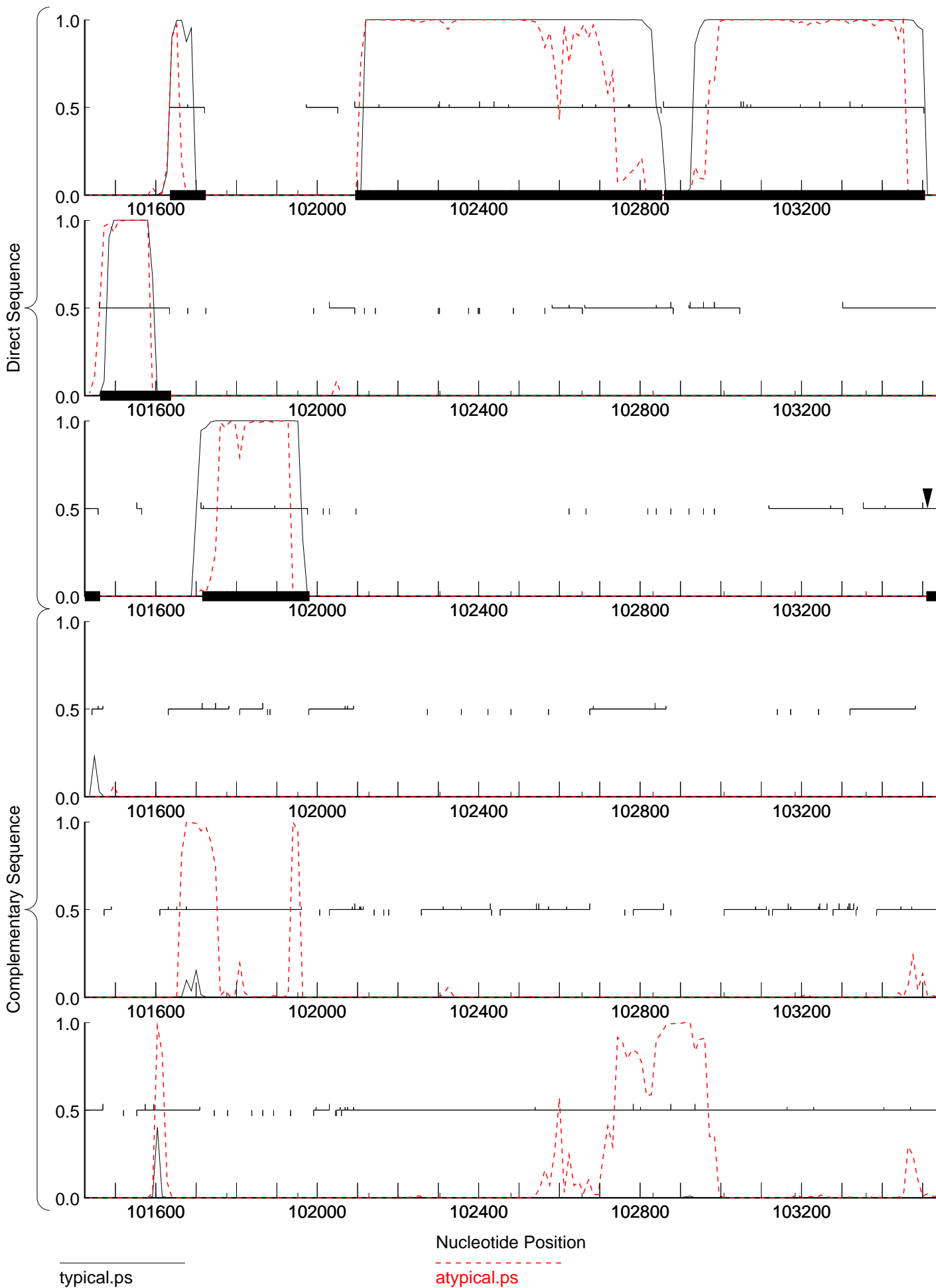
typical.ps

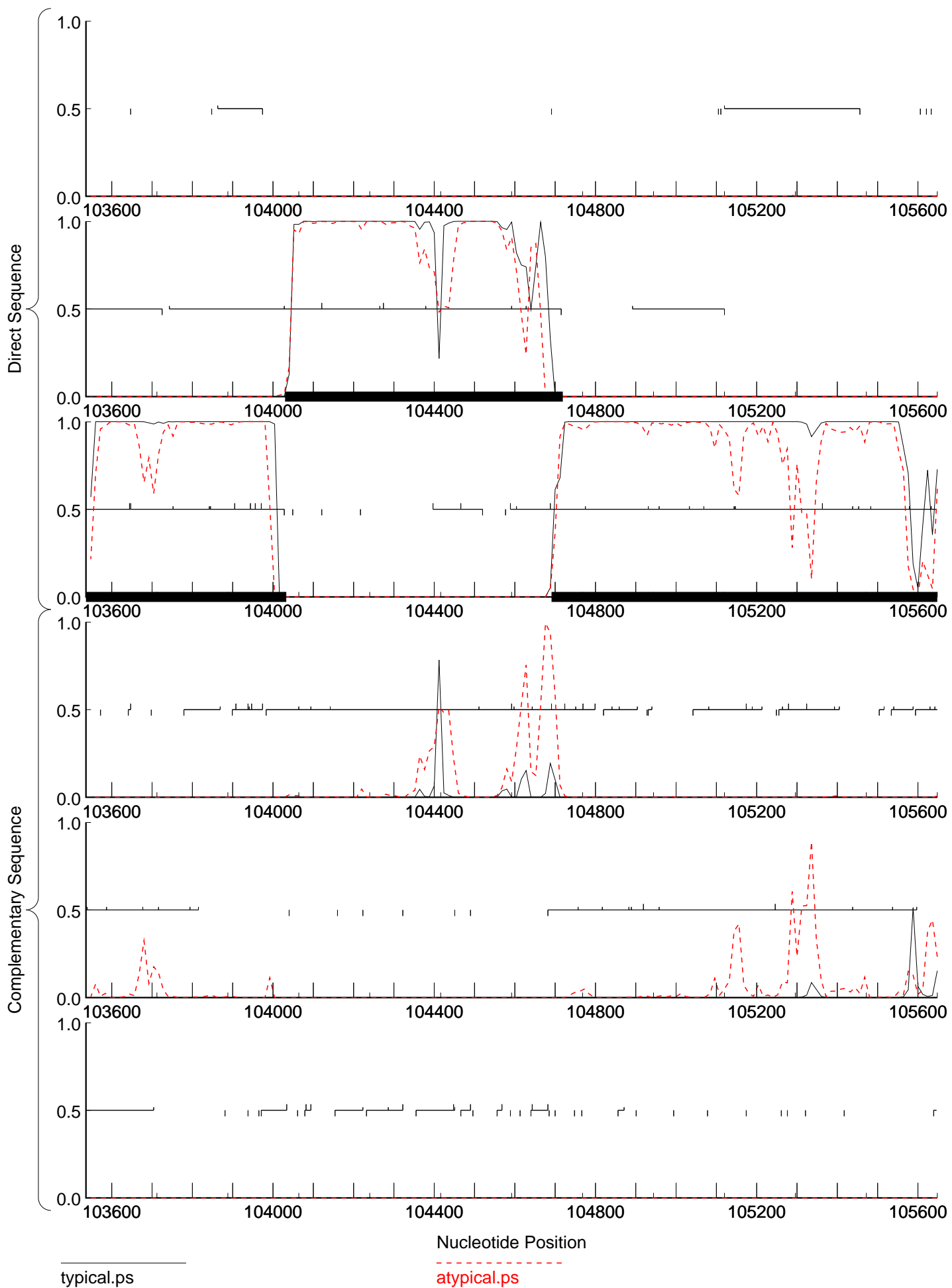
atypical.ps

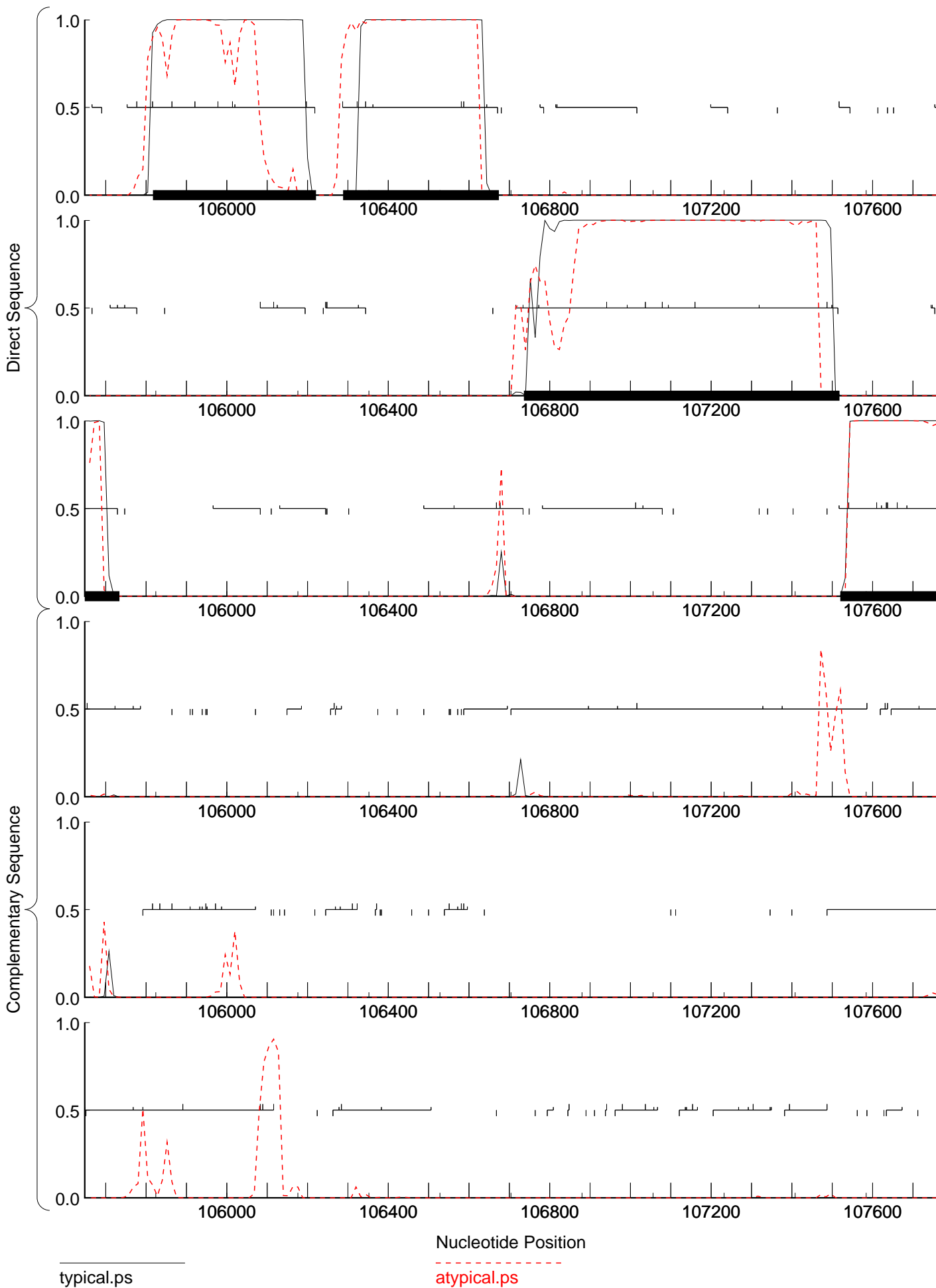


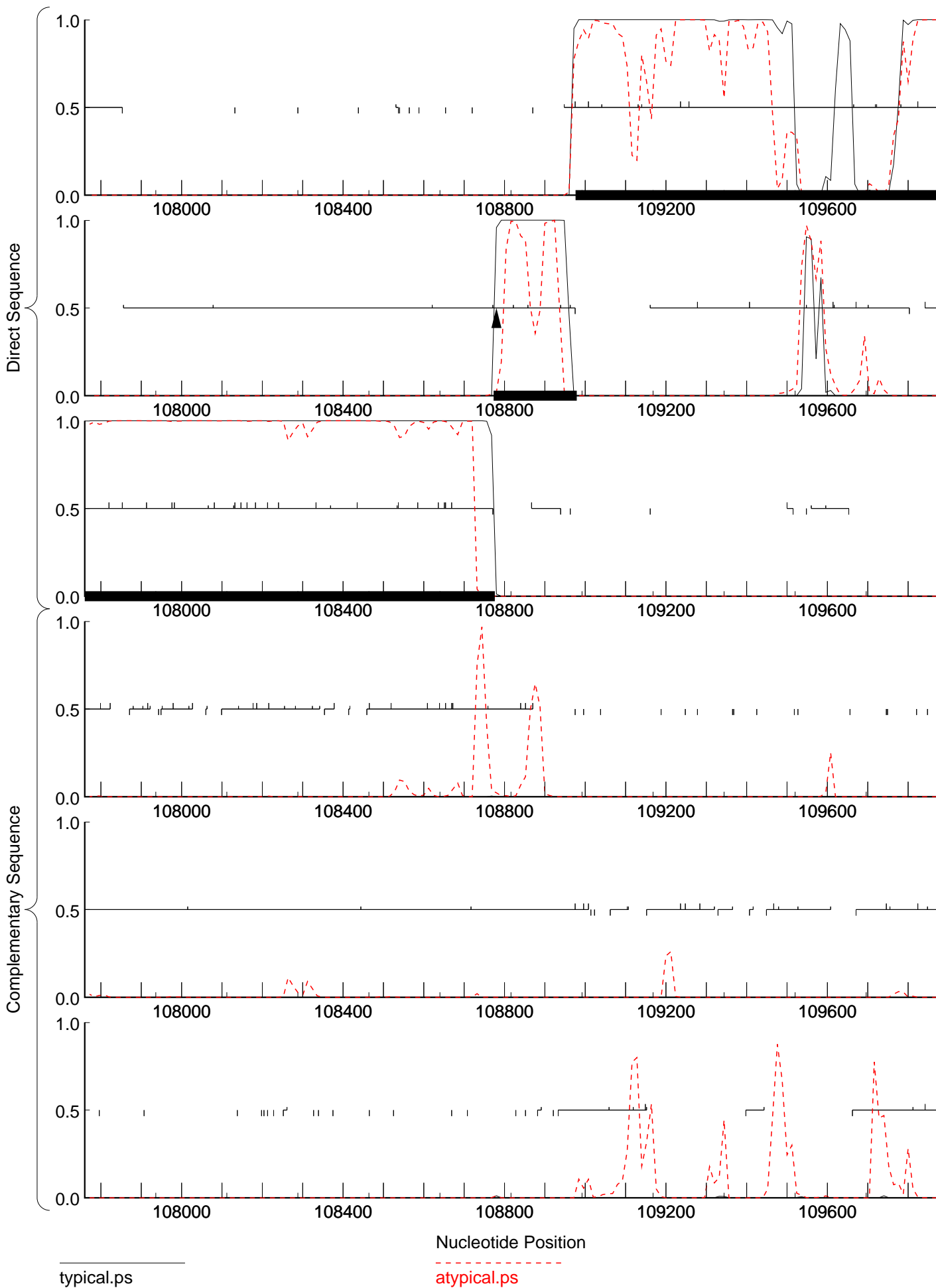






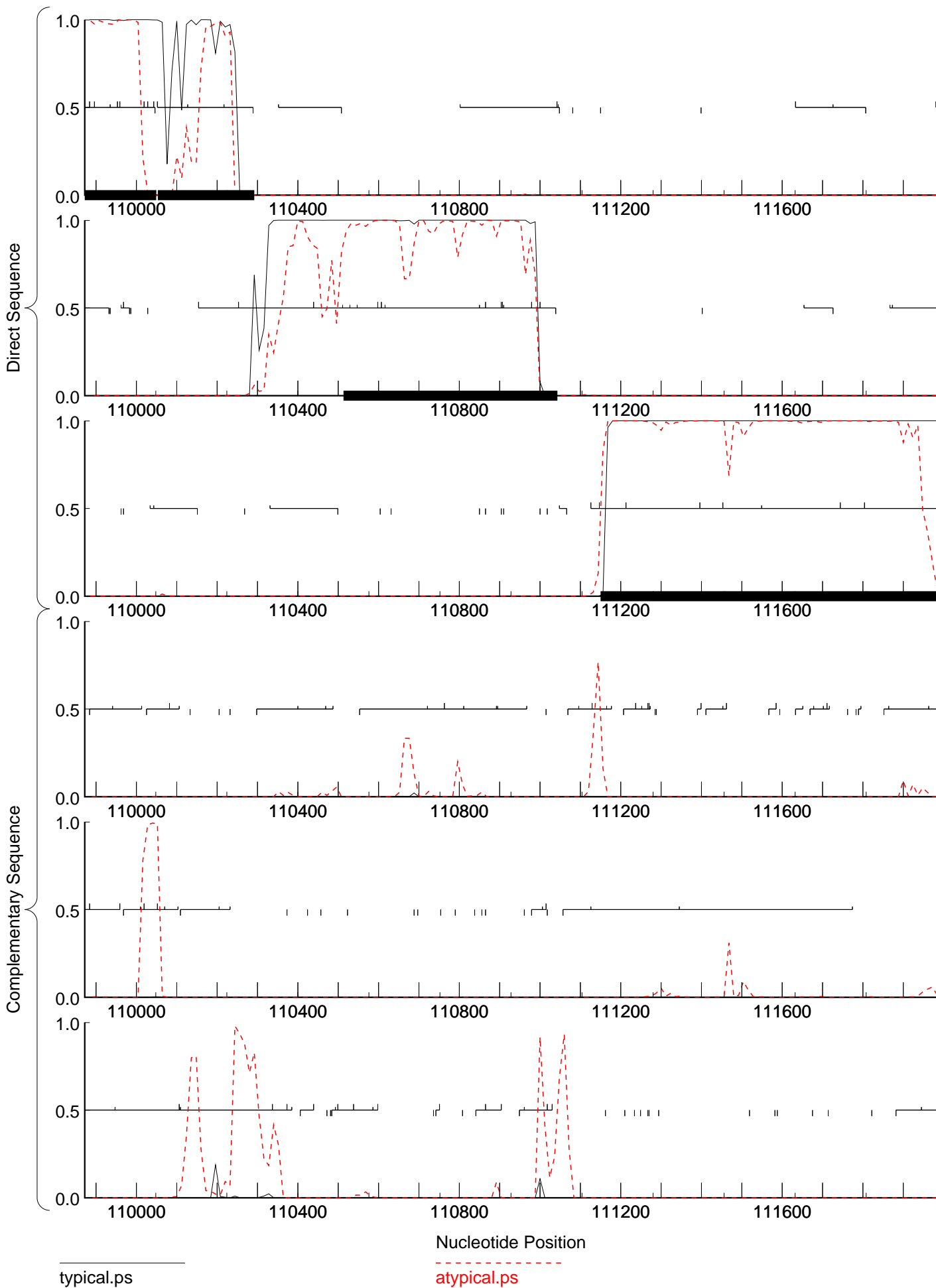






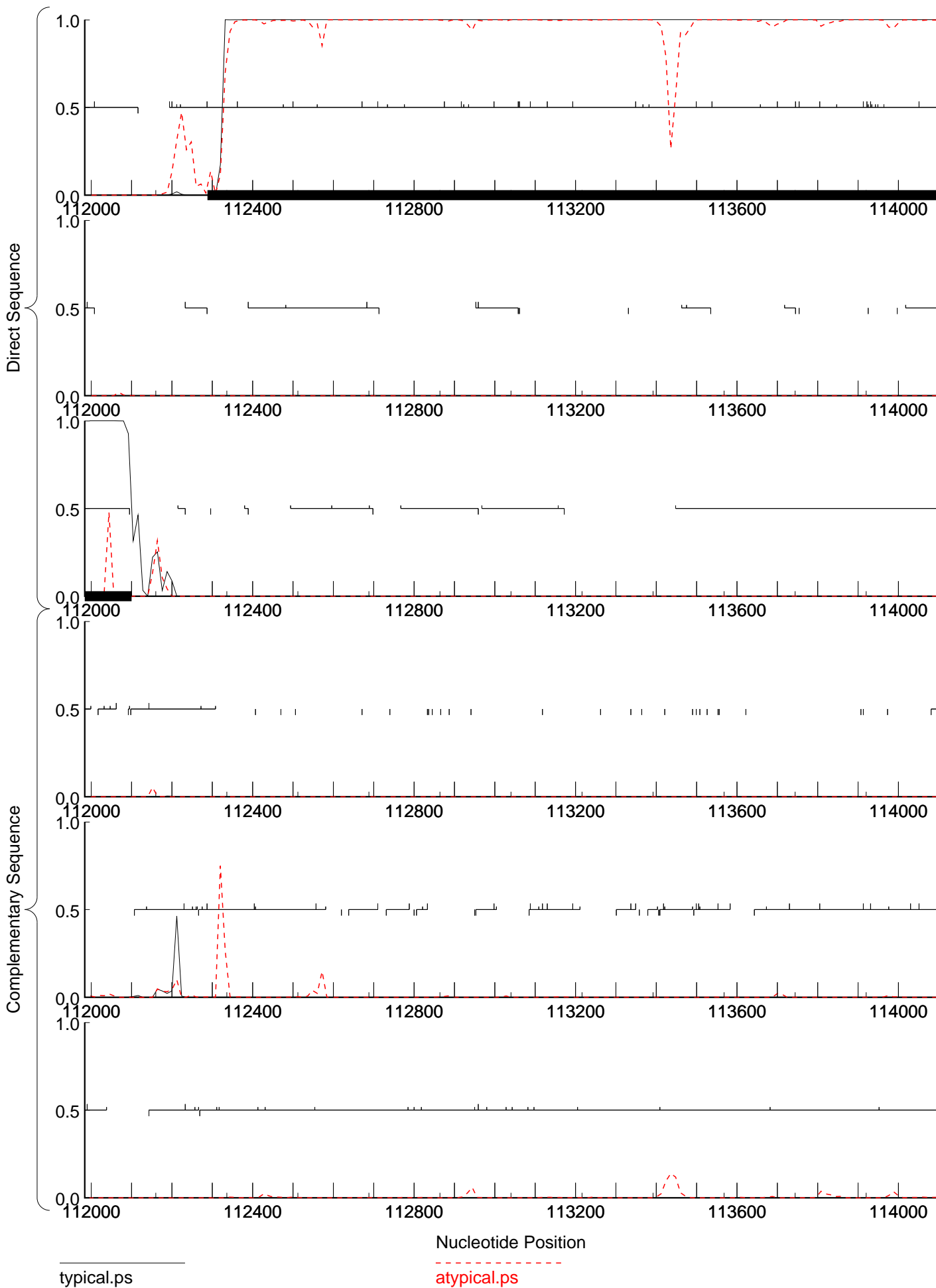
typical.ps

atypical.ps



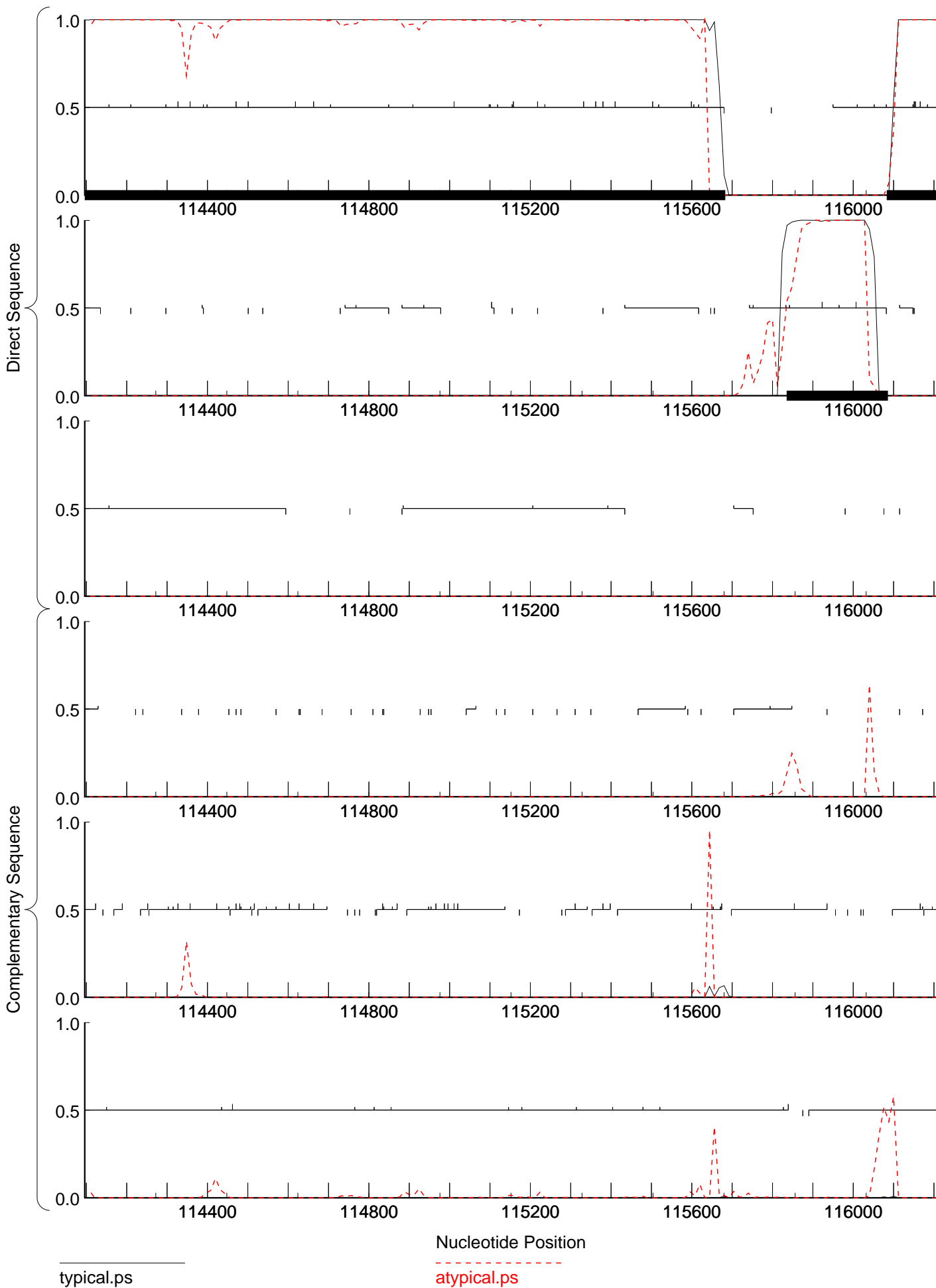
typical.ps

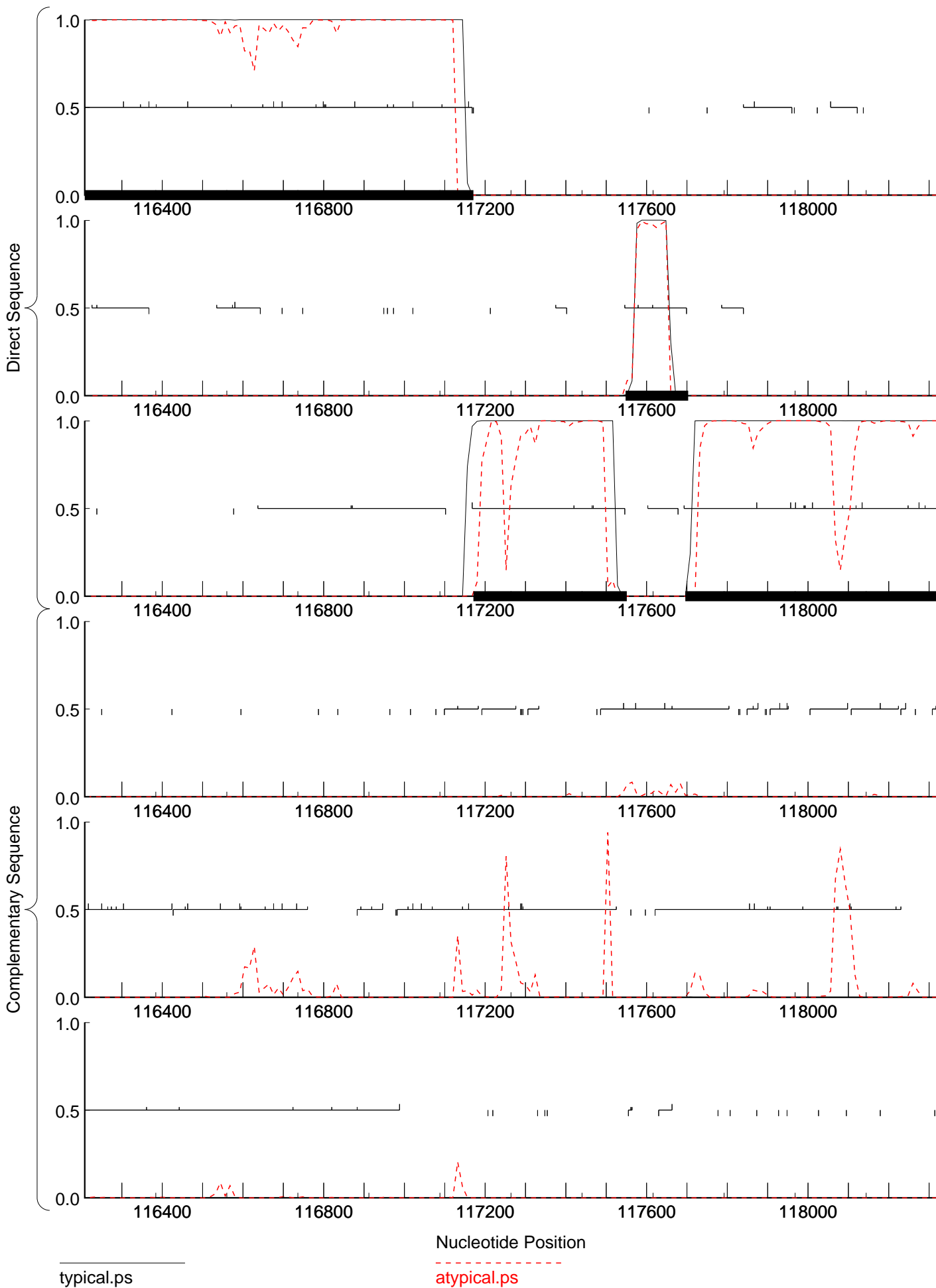
atypical.ps

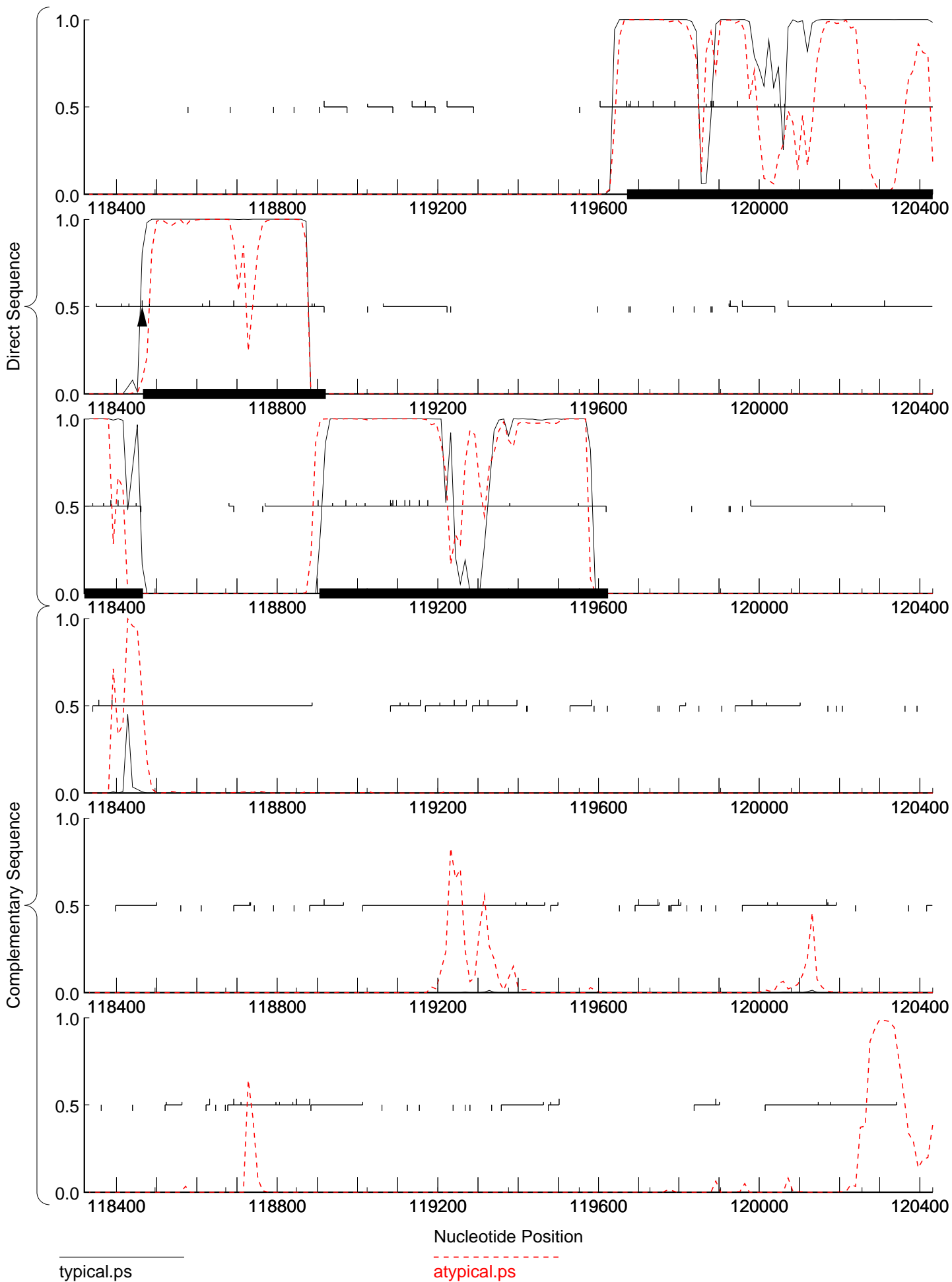


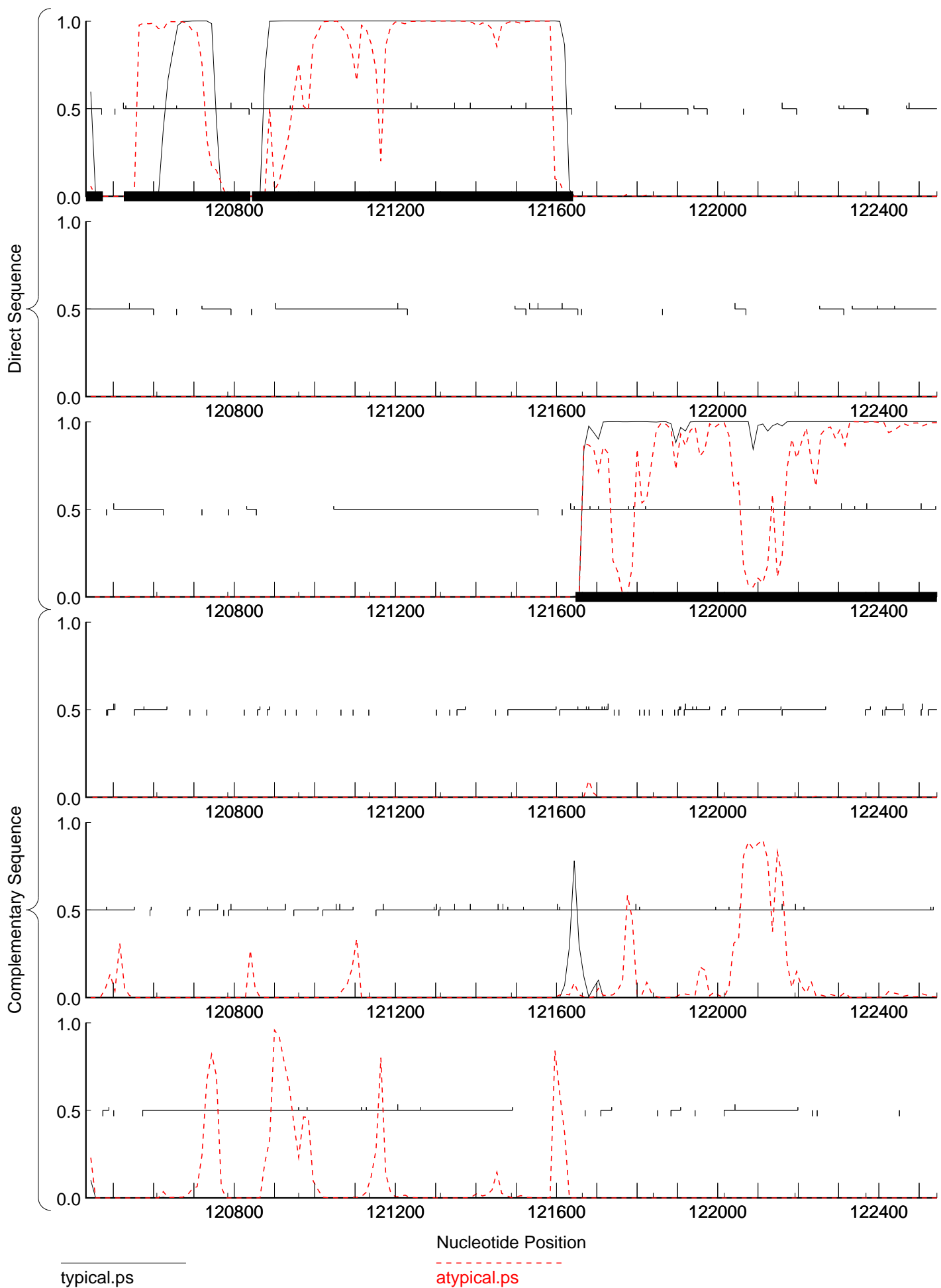
typical.ps

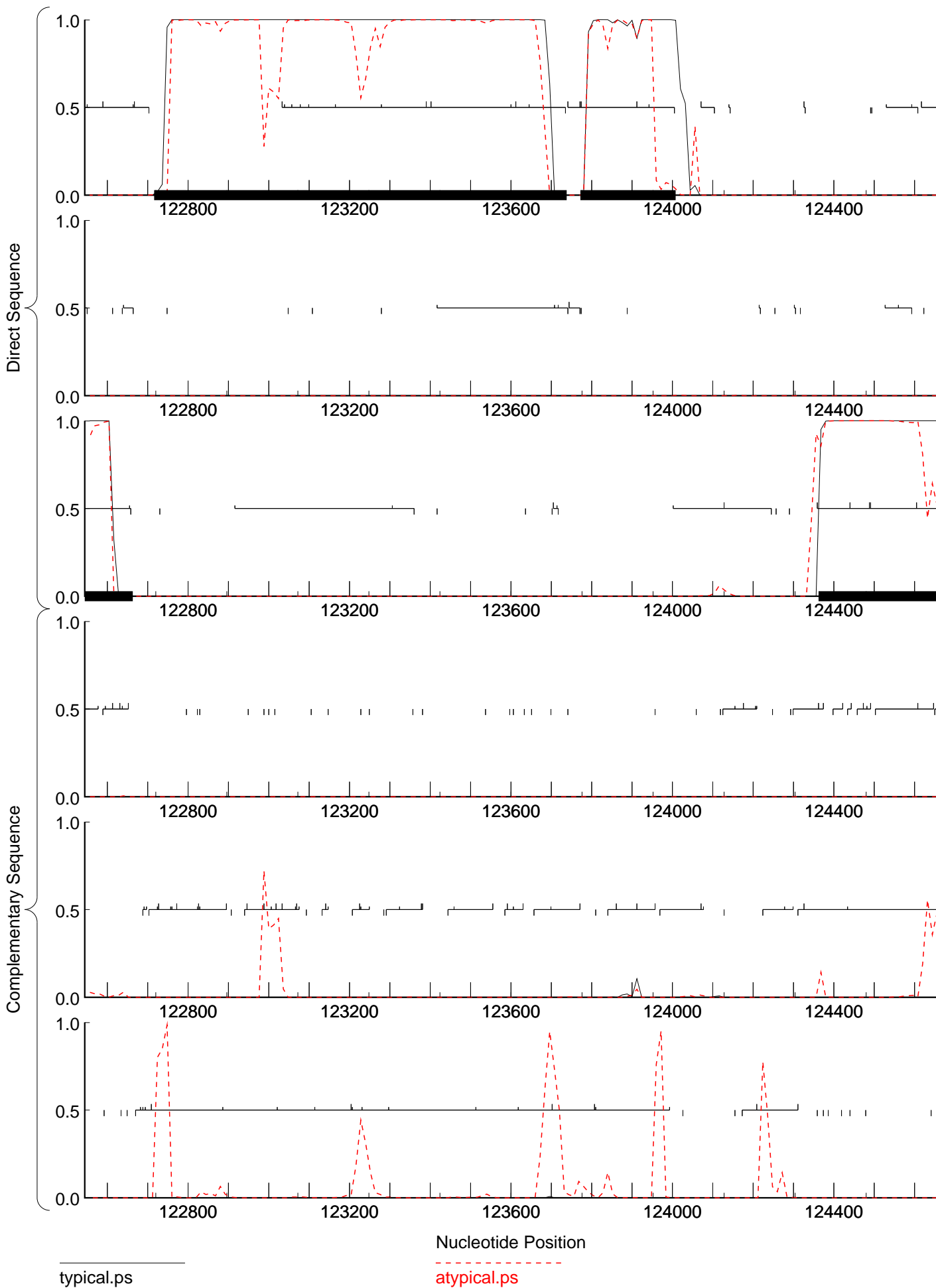
atypical.ps





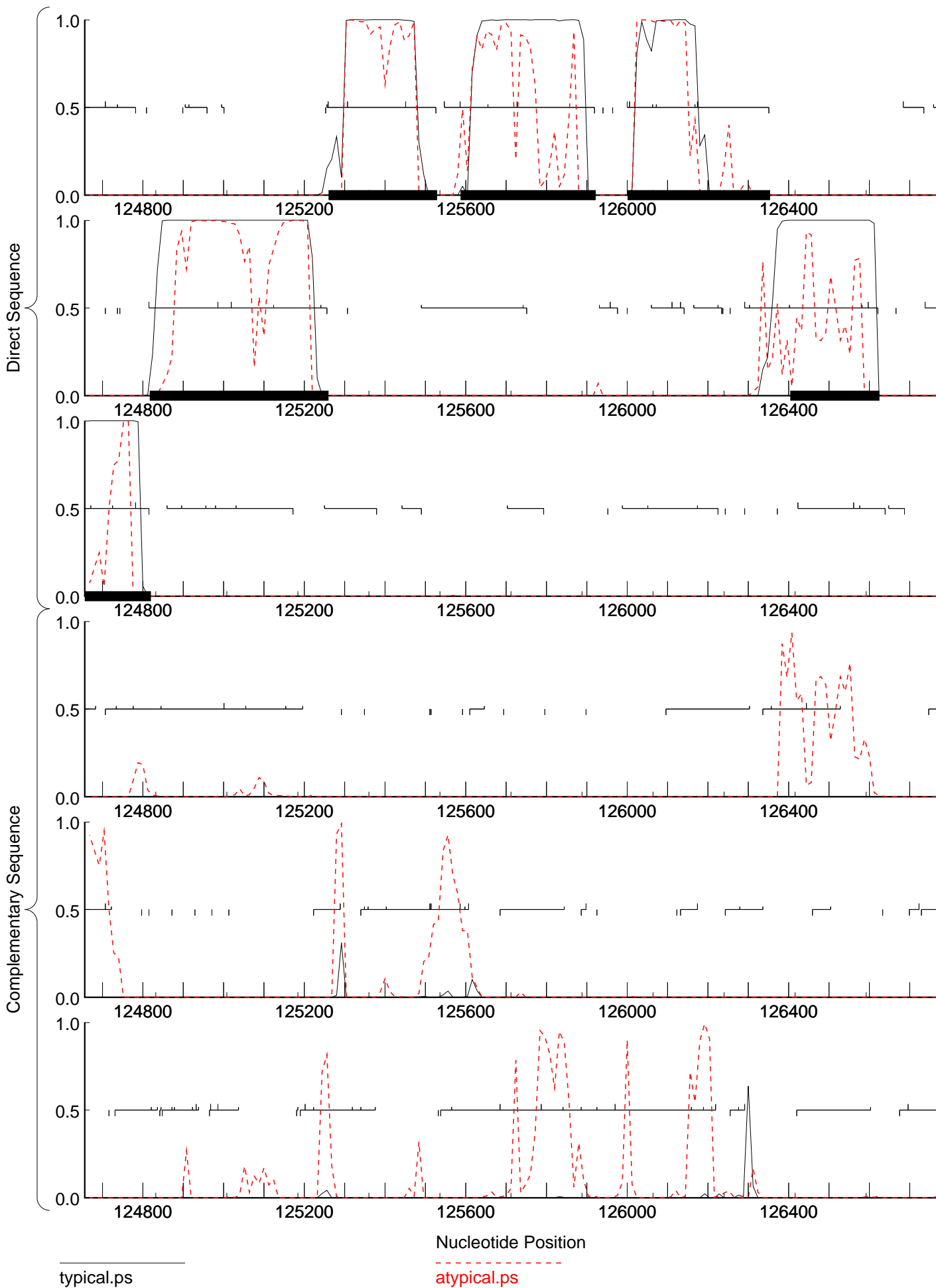






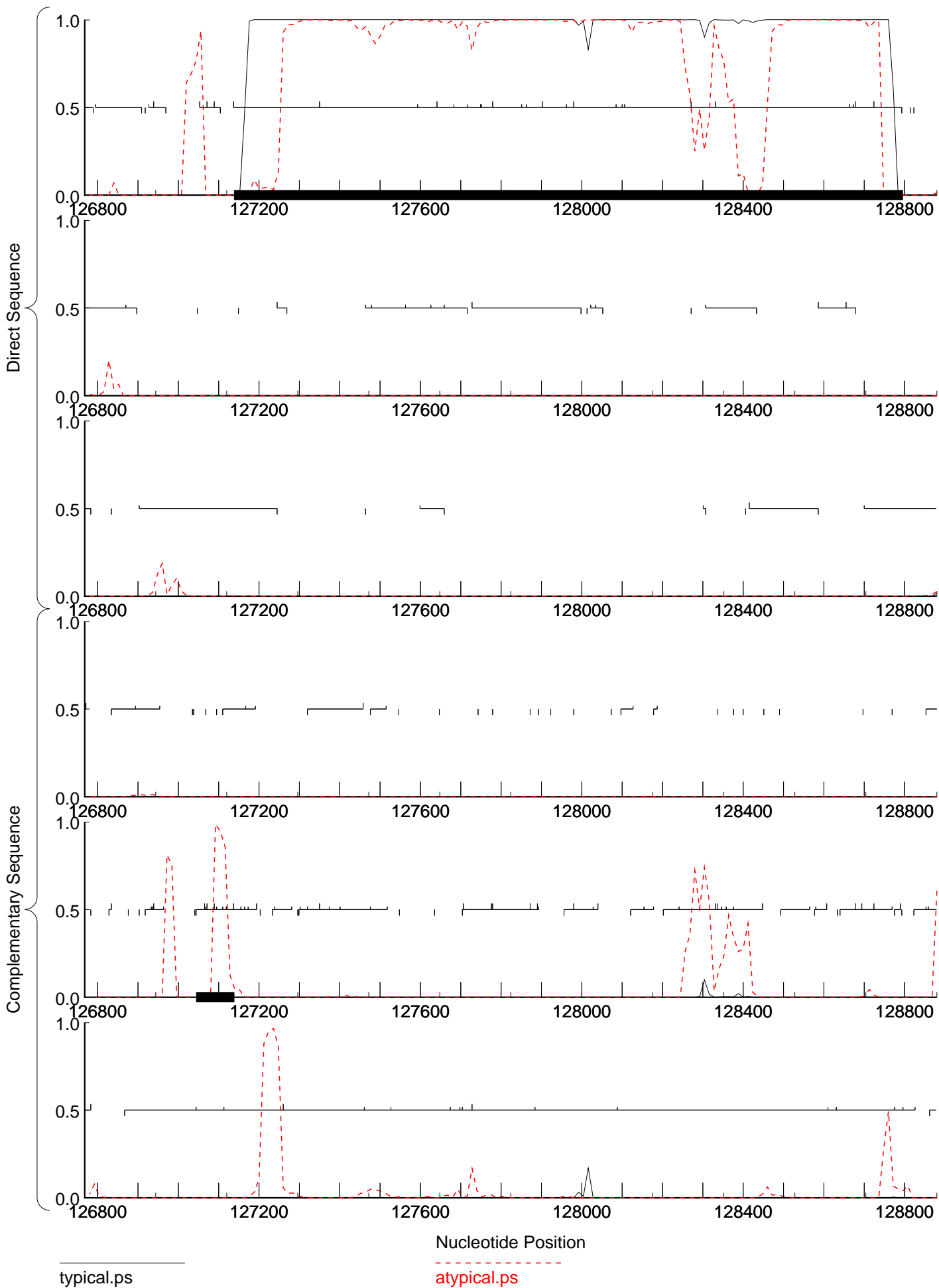
typical.ps

atypical.ps



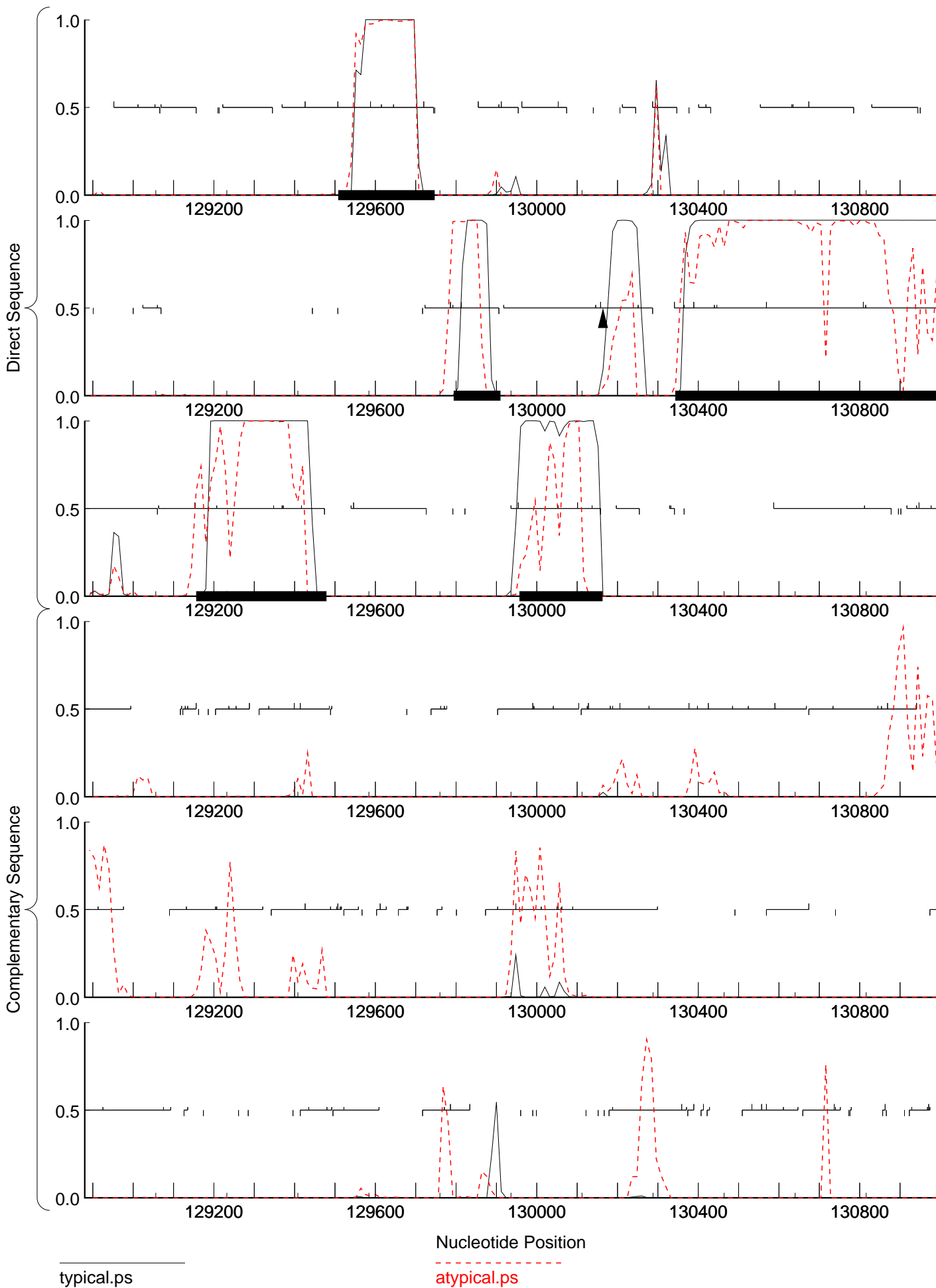
typical.ps

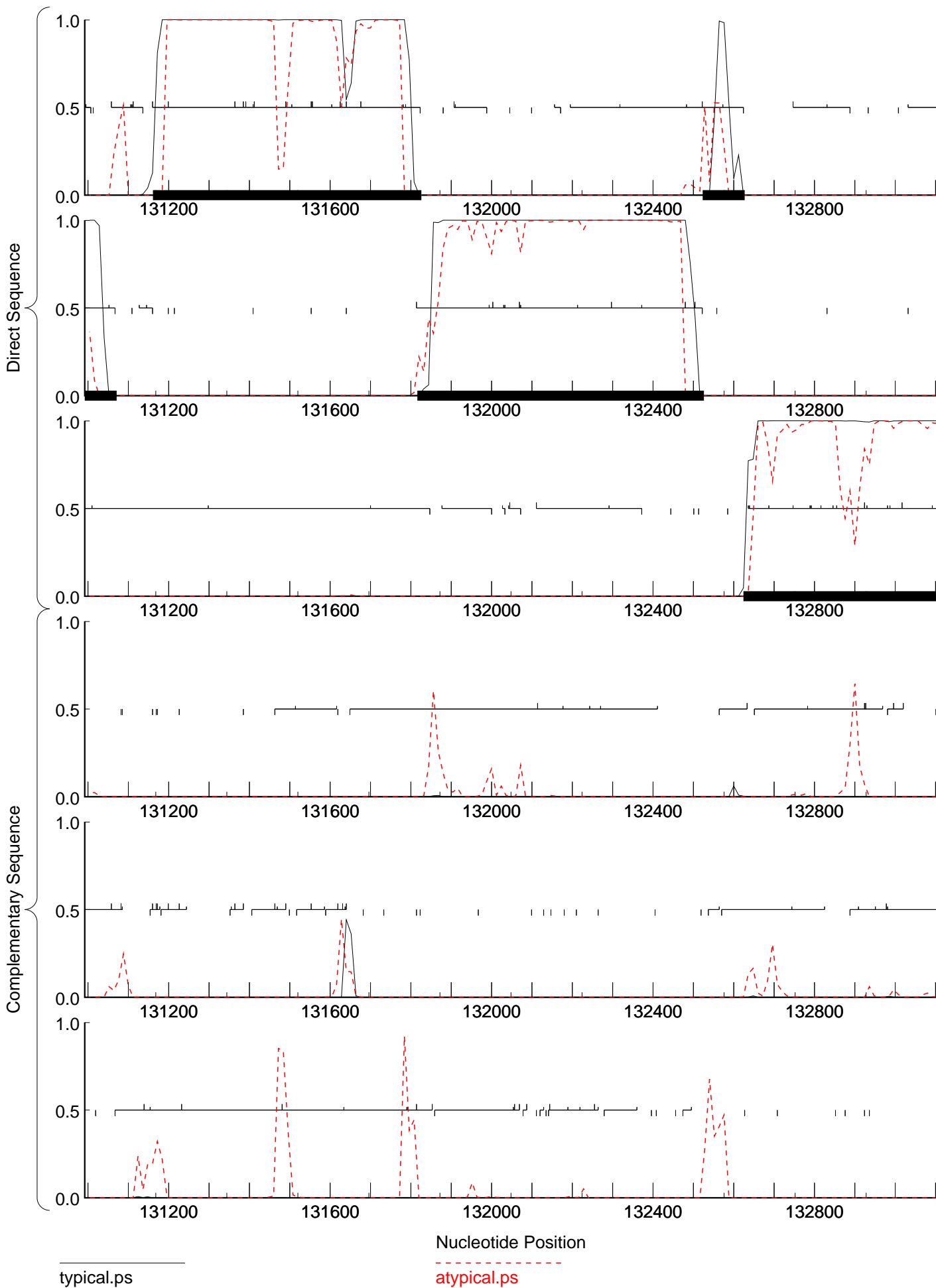
atypical.ps

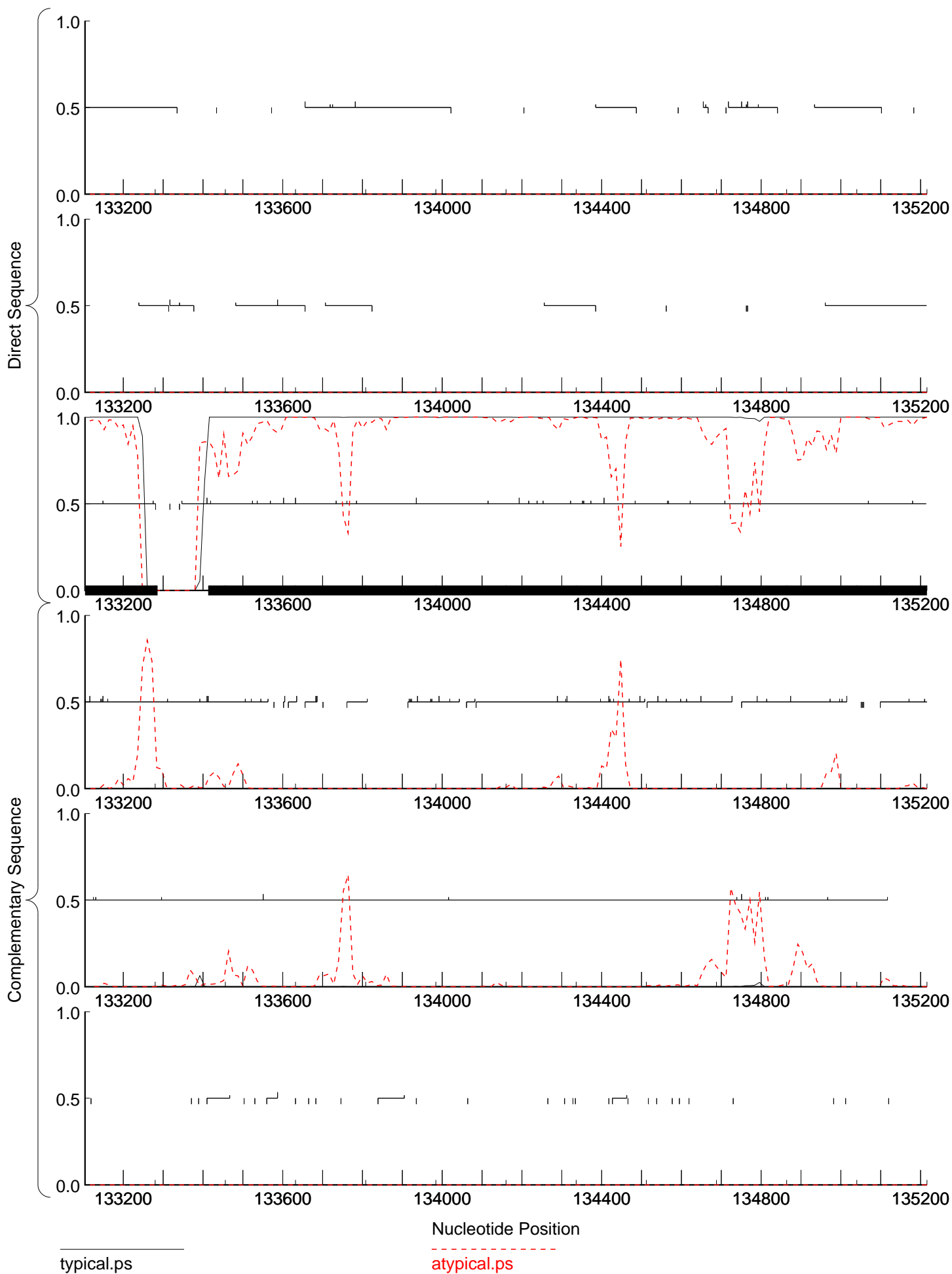


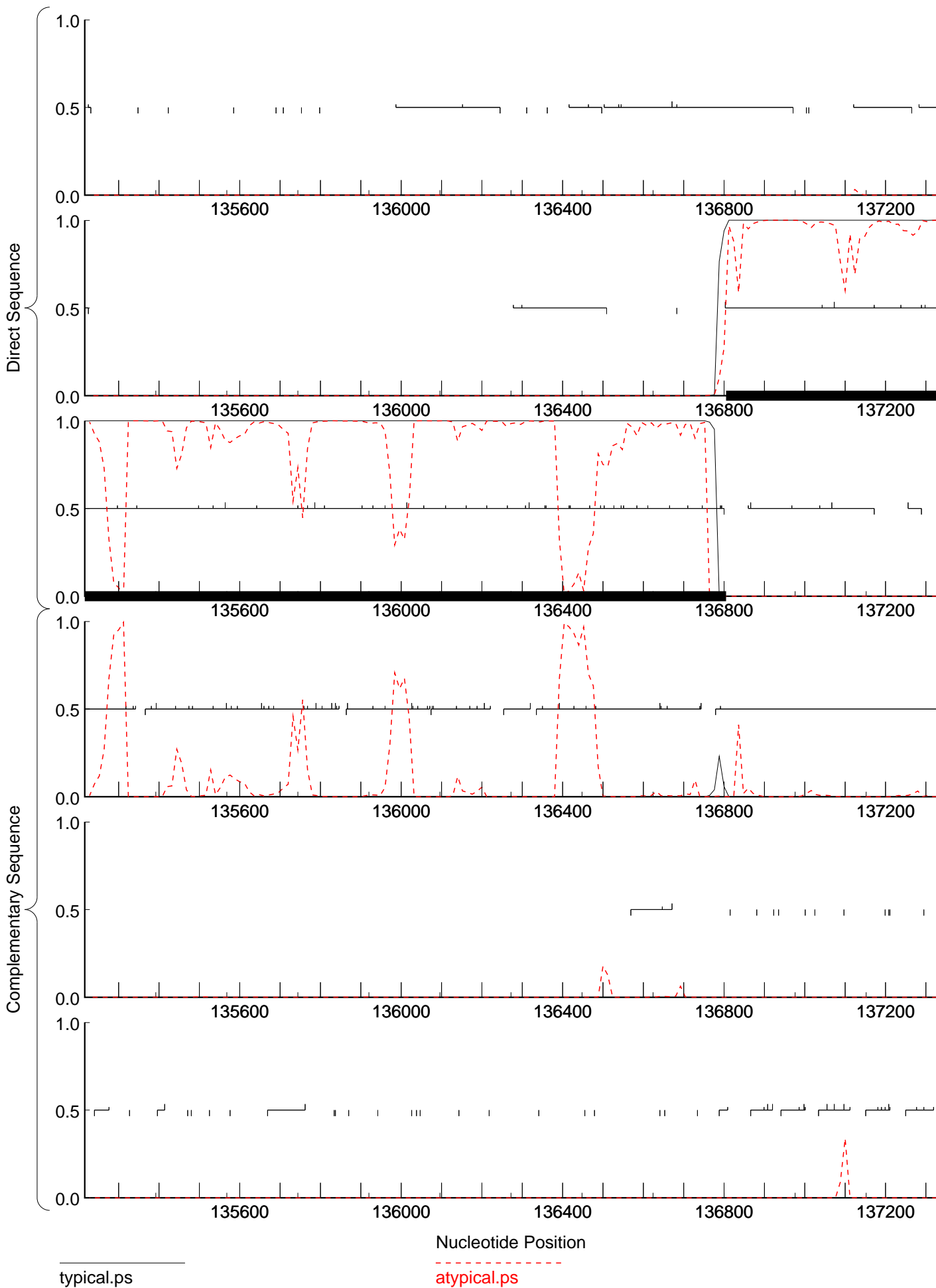
typical.ps

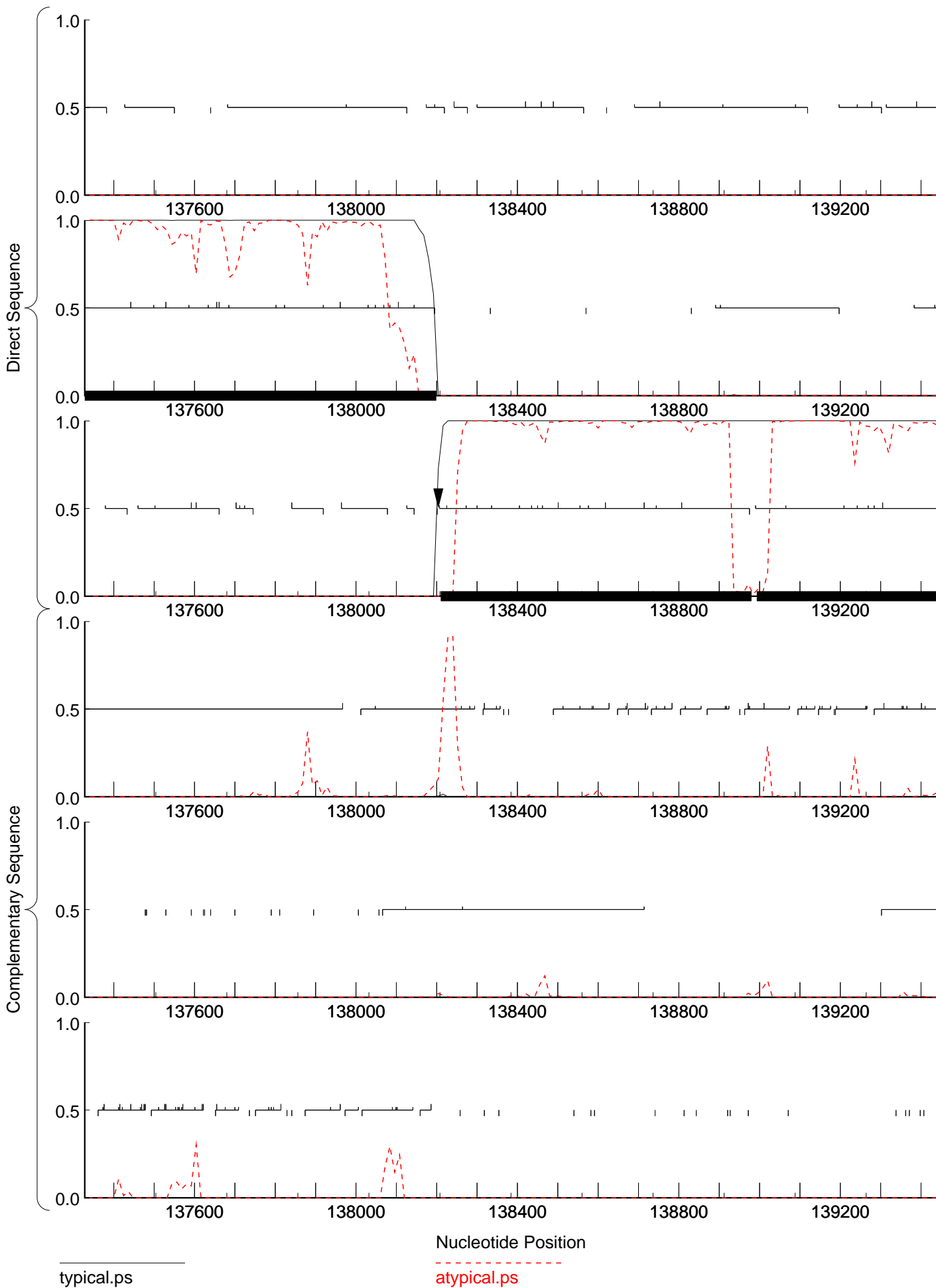
atypical.ps





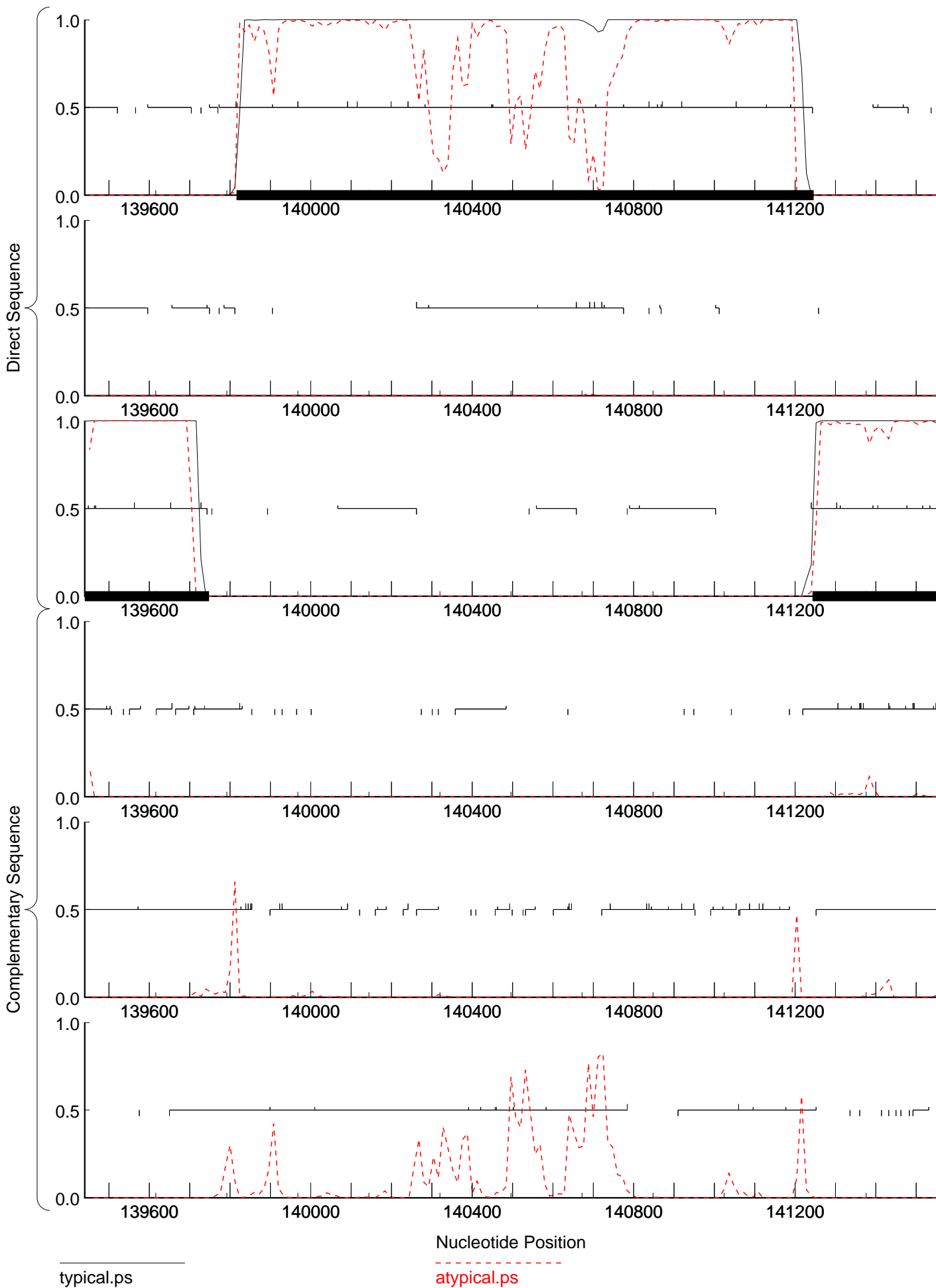


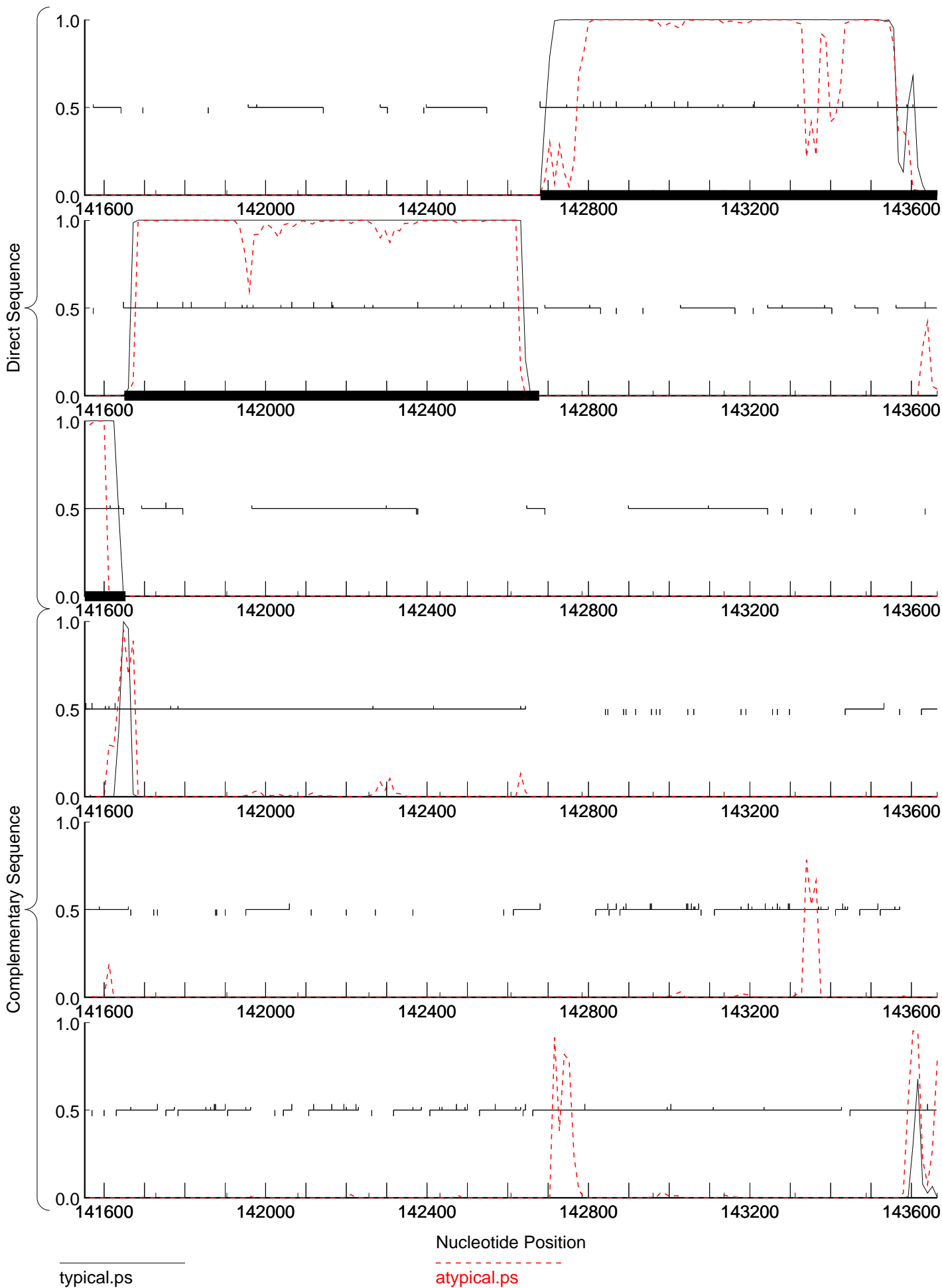


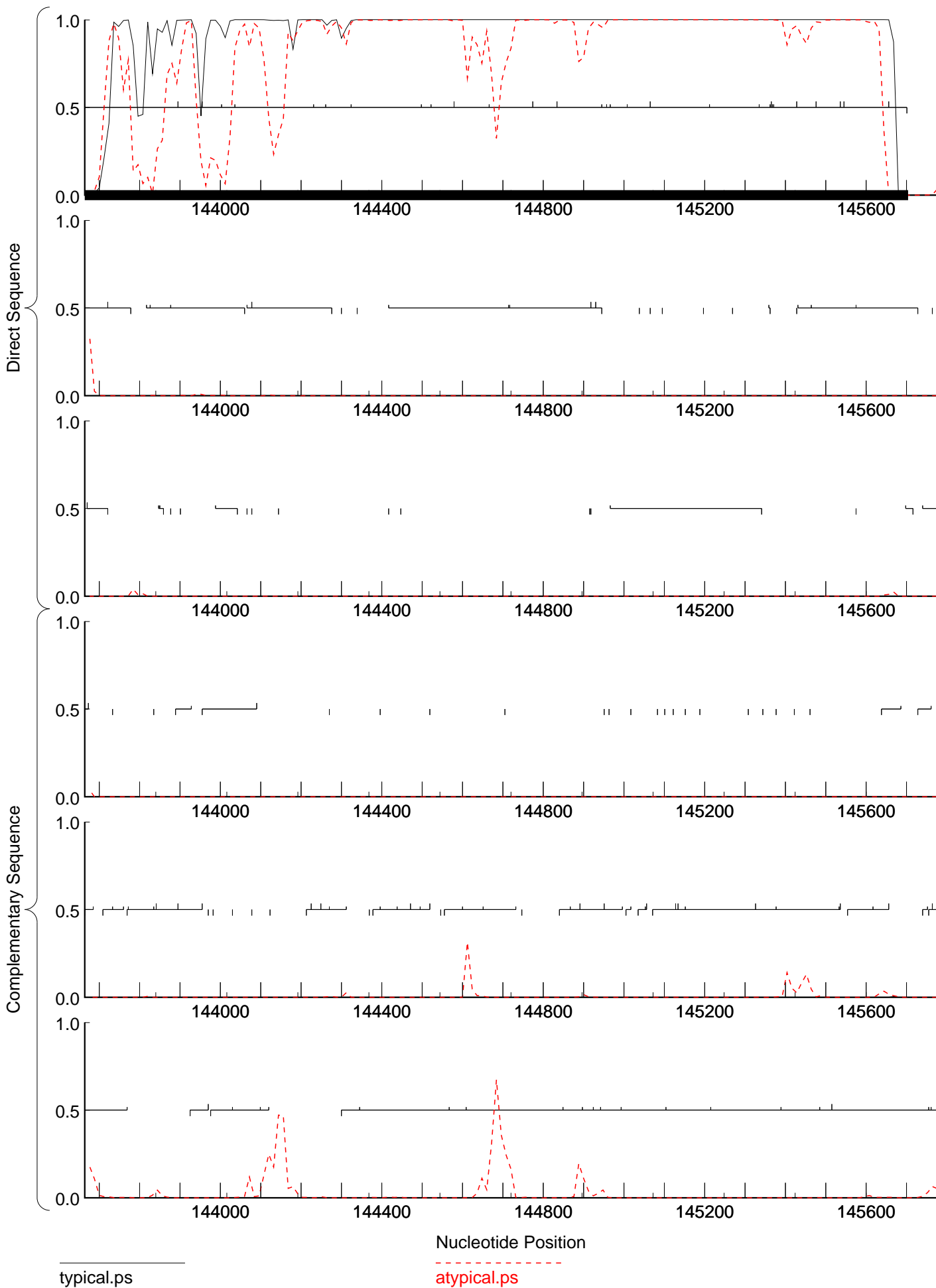


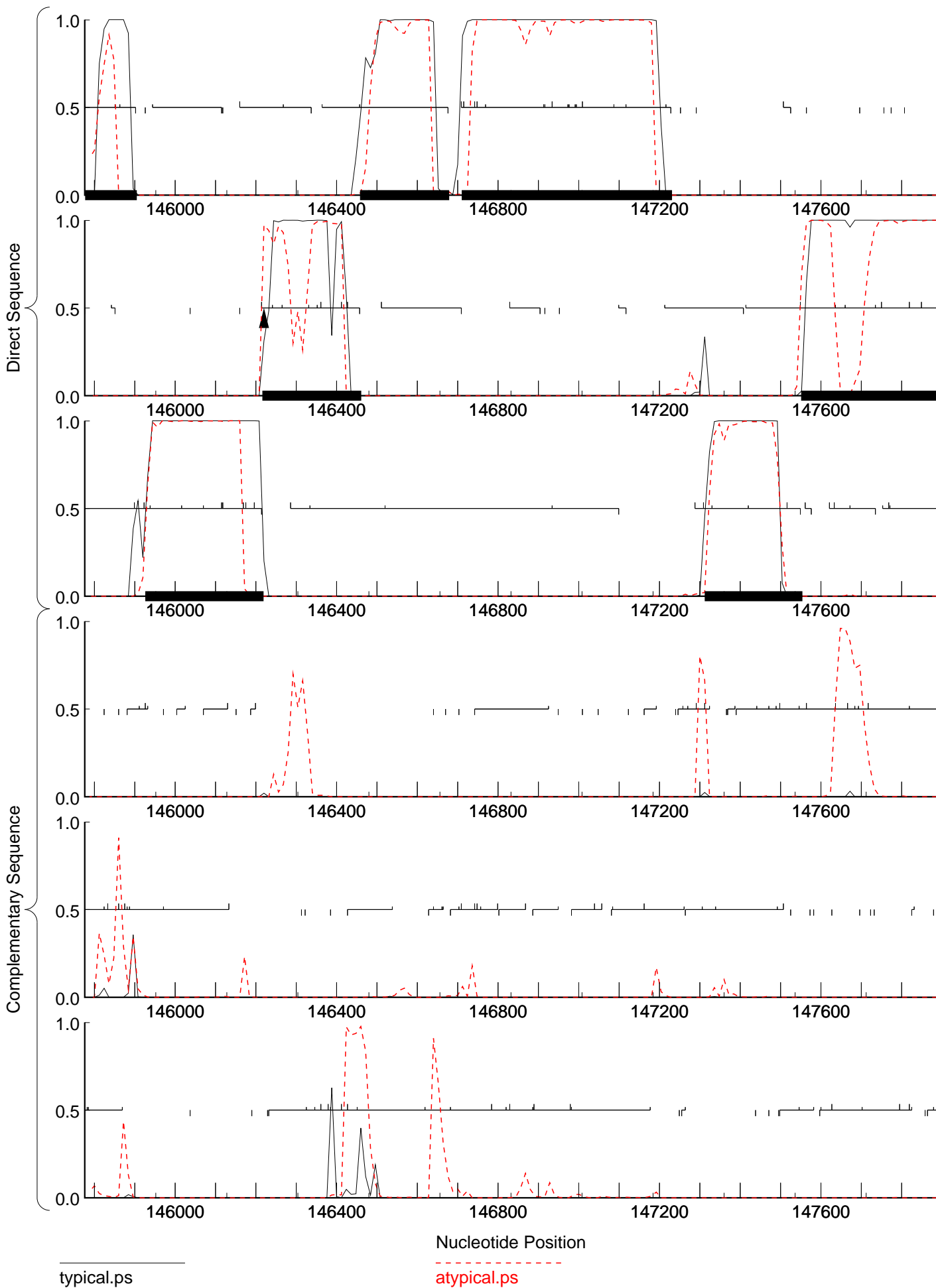
typical.ps

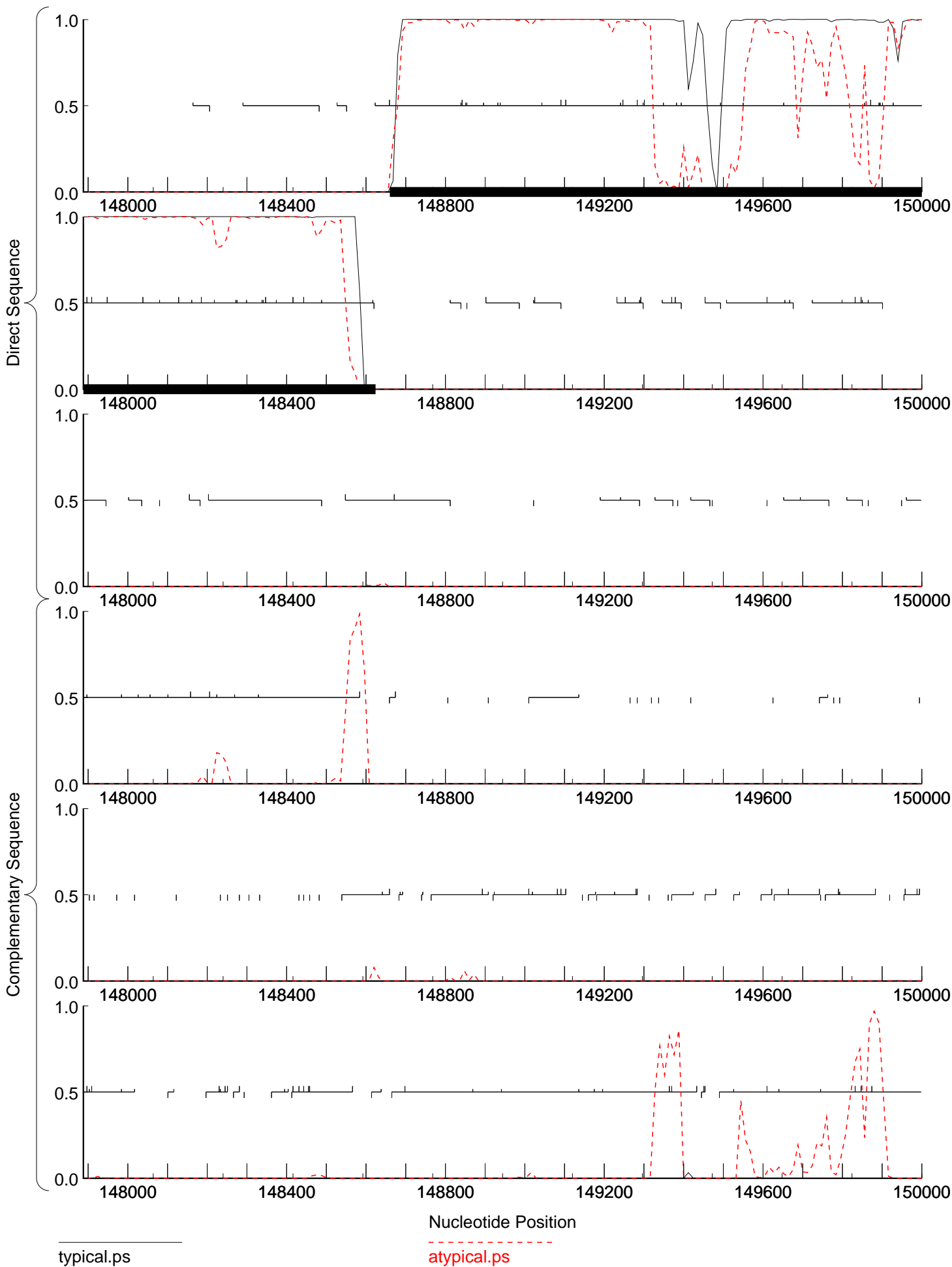
atypical.ps











typical.ps

atypical.ps

