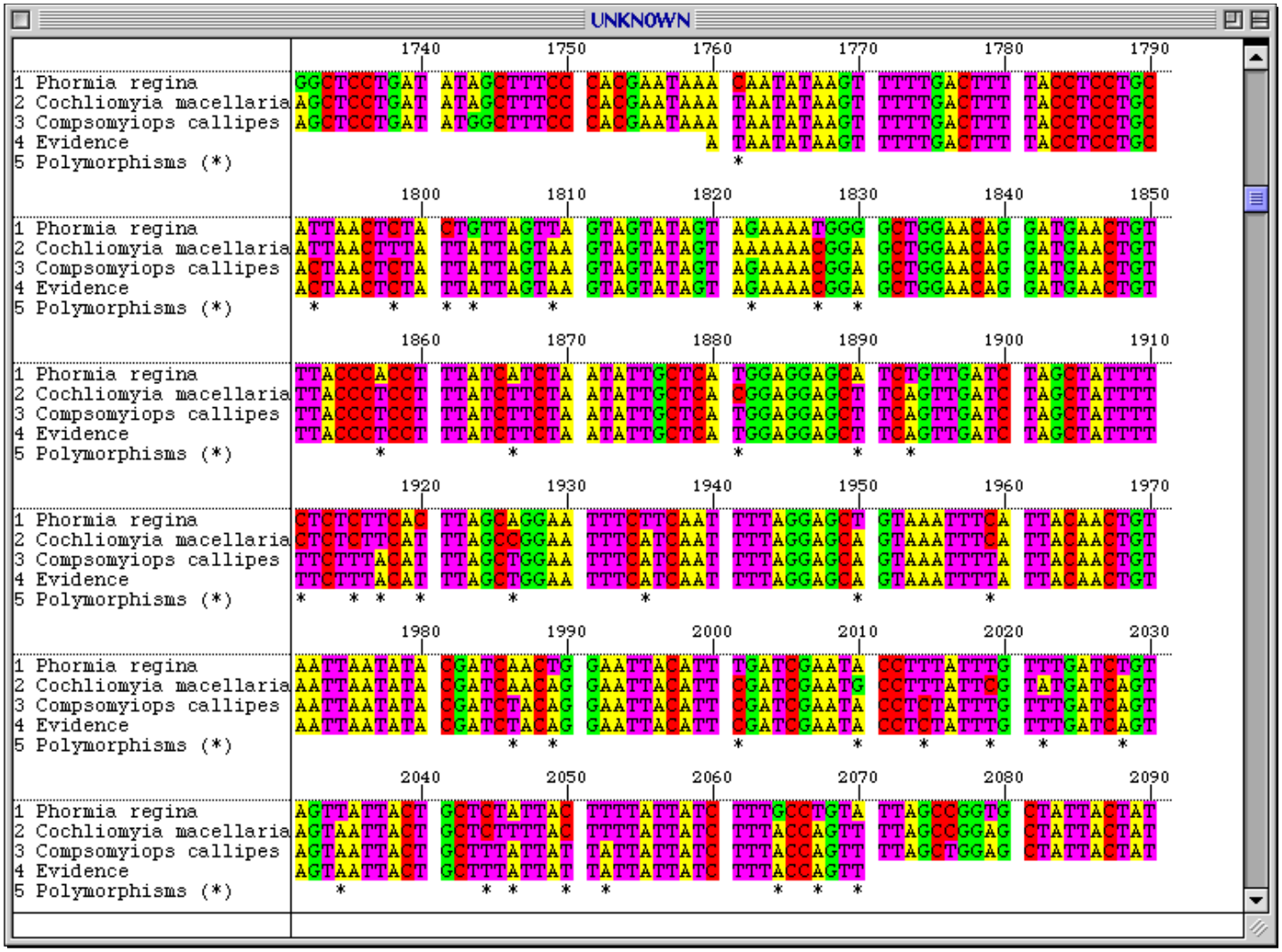


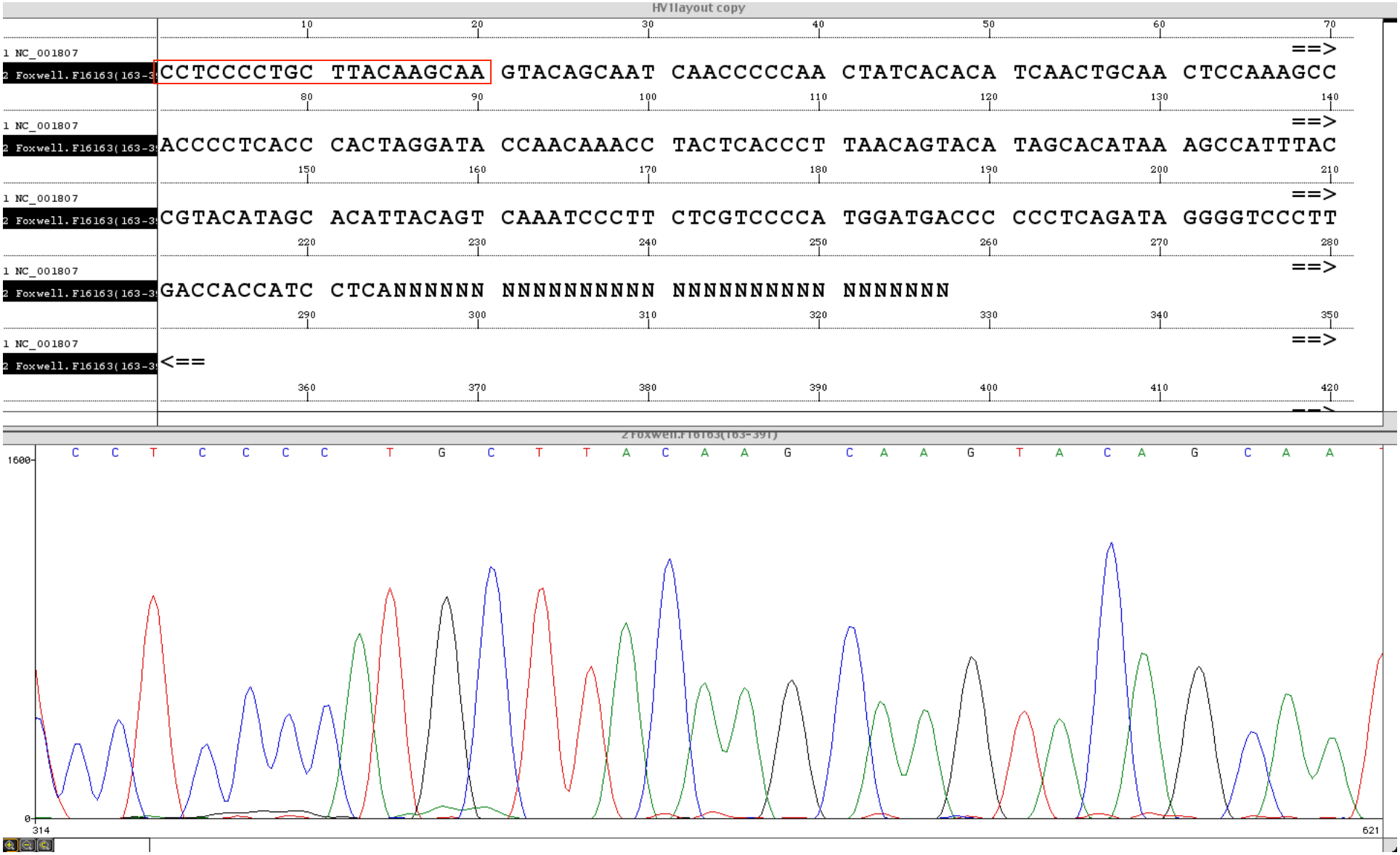
Sequence alignment.

Two general human i.d. alignment situations

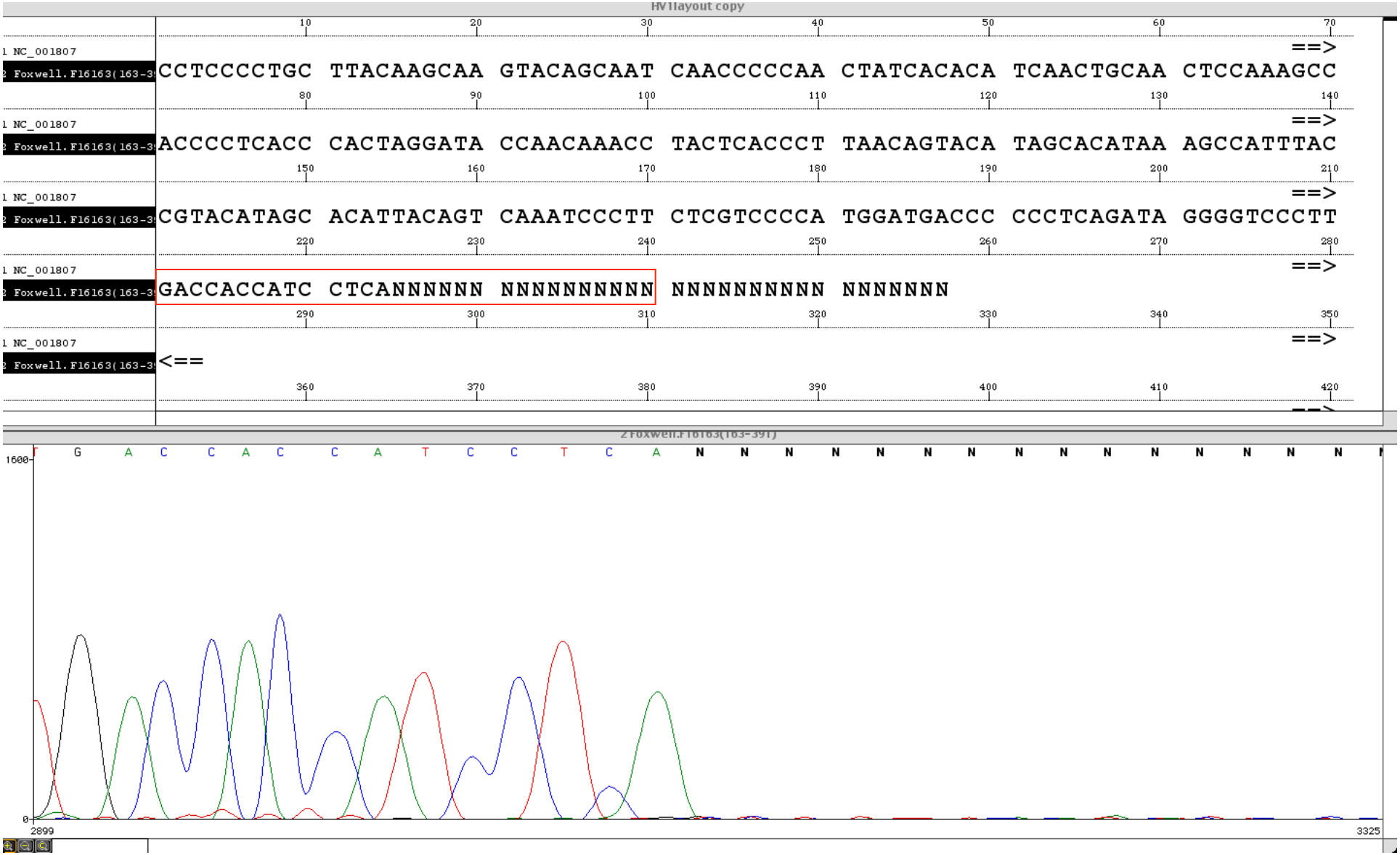
- Aligning overlapping data from the same specimen
- Aligning your sample sequence with the CRS.

Sequence editing and alignment can be done in a variety of ways. This can be mostly manual or very automated. Forensic analysis tends to be more manual.

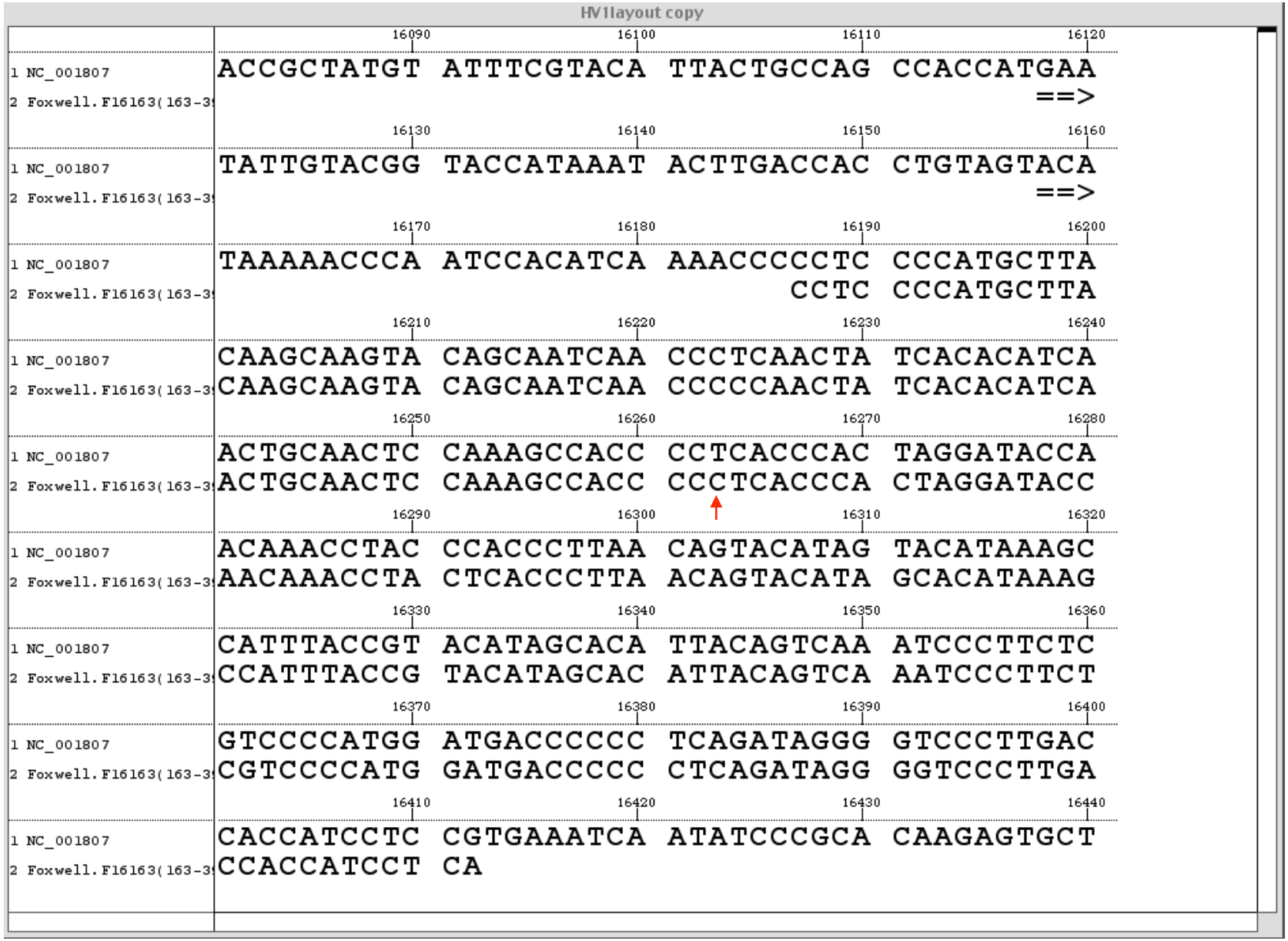
Initial manual editing. The letters above show the sequence as automatically called by the analytical software. The actual data, the electropherogram, for a selected region is shown below. A text file of the CRS is out of view in this layout.



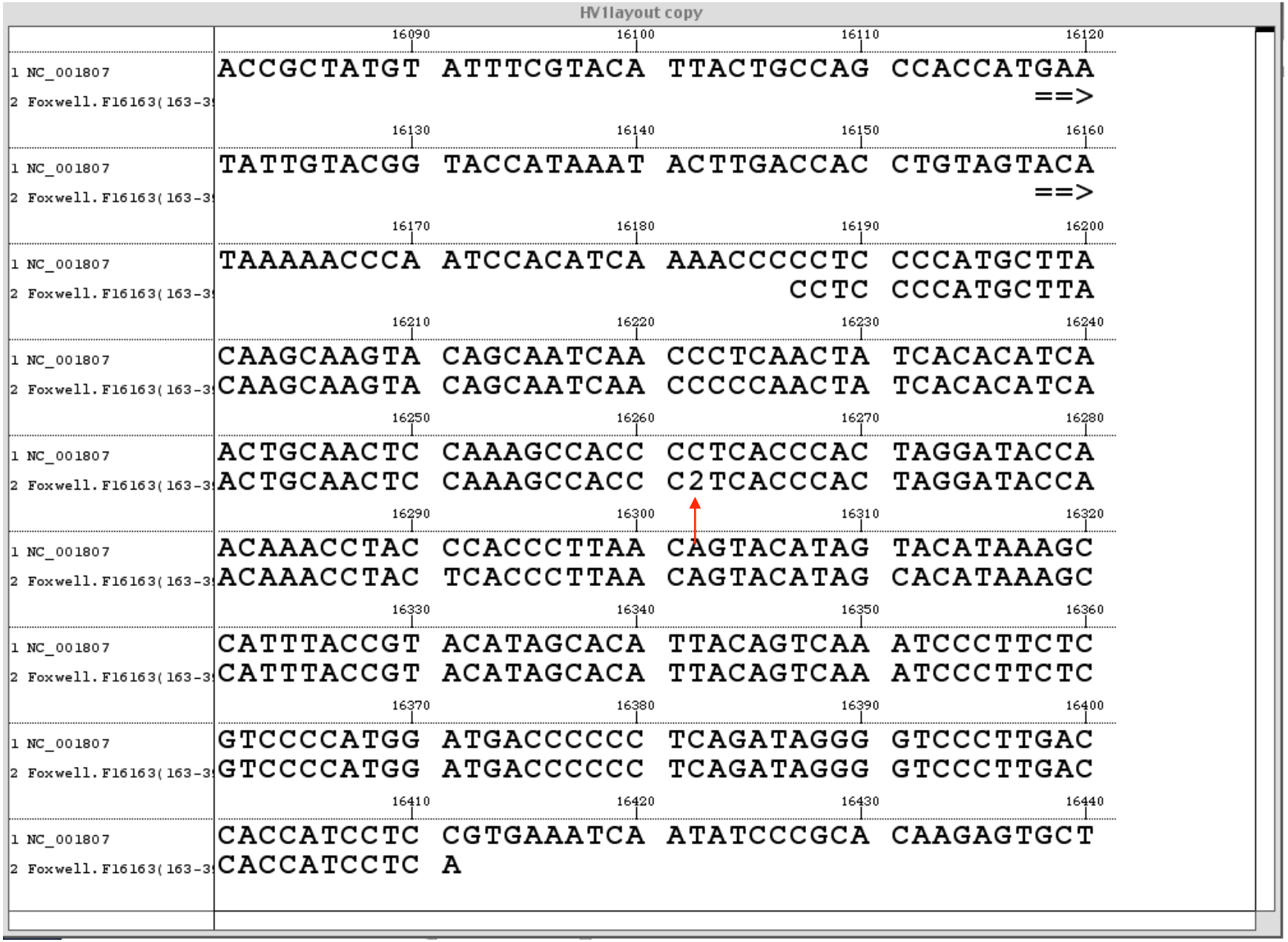
After cropping 5’ end. What else is wrong with the sequence as called by the computer?



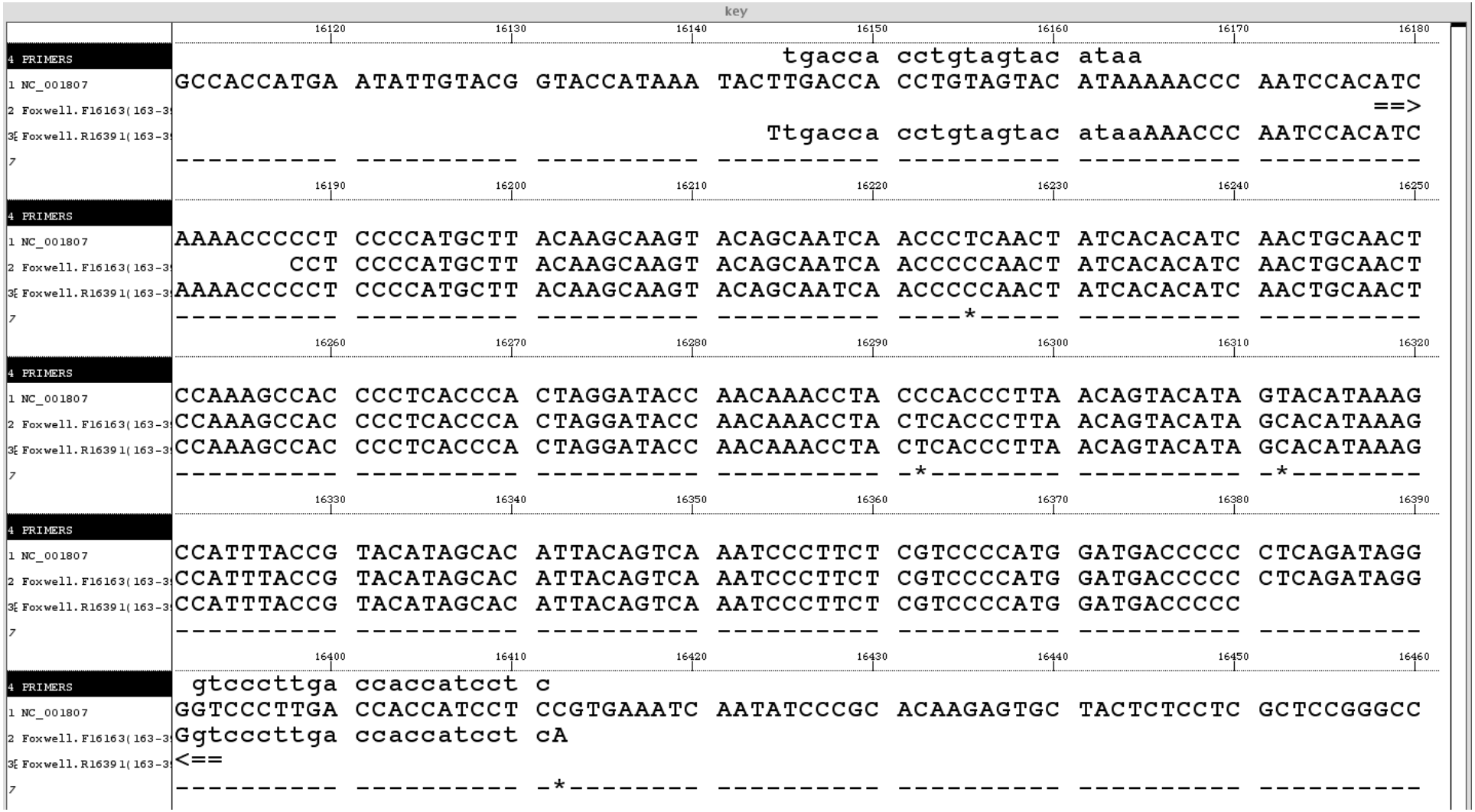
Although the electropherogram does tend to get messier at the 3’ end. The string of Ns simply reflects the fact that the genetic analyzer was set to collect data after the last fragments passed the detector.



The edited data file is then moved to align it with the reference sequence. This establishes base position # and allows comparison between sequences. However an indel can throw off alignment.

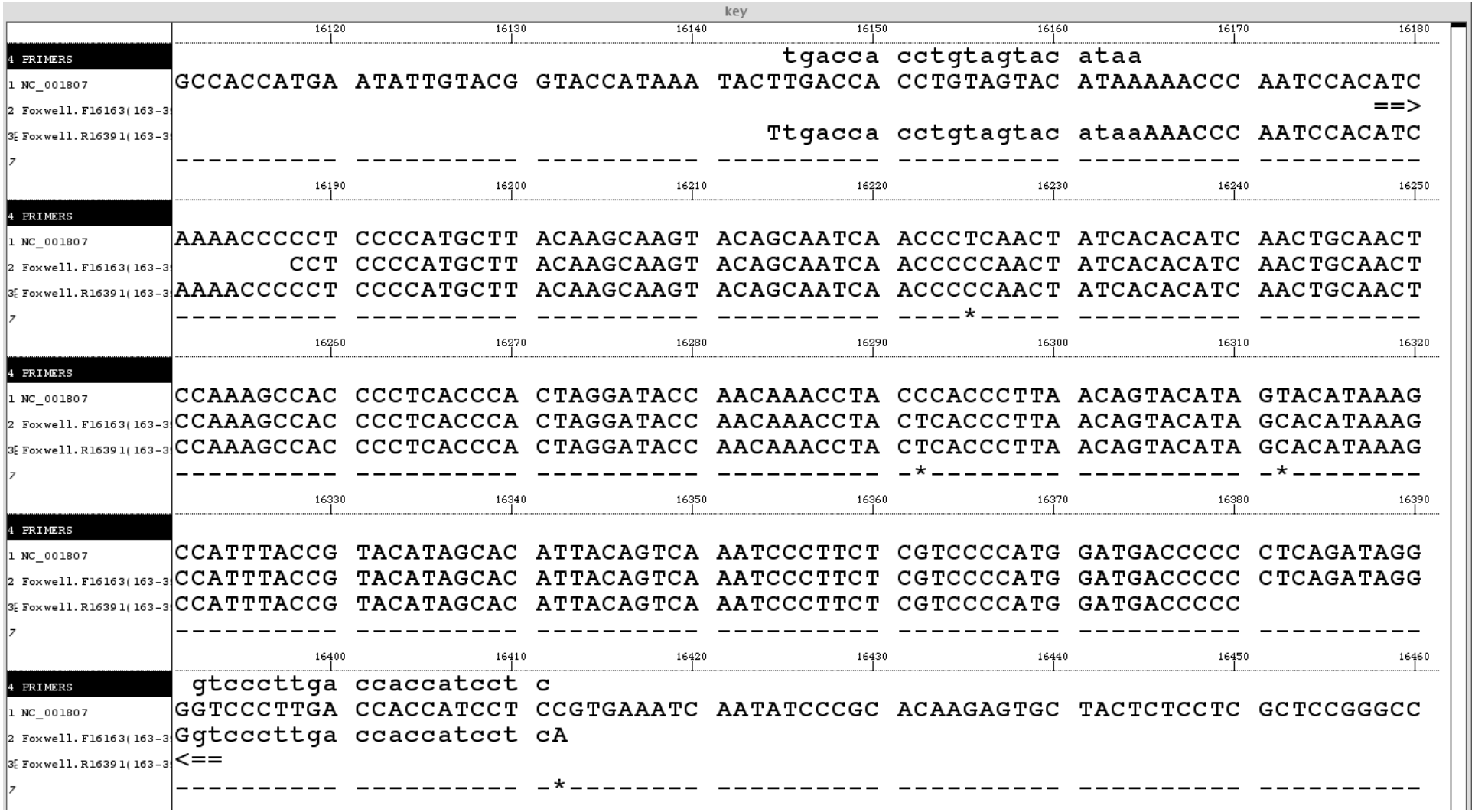


The Wells system for maintaining alignment faced with an insertion relative to the CRS. However you do it, the insertion is recorded with an extra decimal for the position. This sequence has a C at position 16262.1. By convention, it is located as far 3' as possible.



One way to finish the layout. The forward sequence and the reverse complement of the reverse sequence are edited and aligned. Primer sequences have been identified. The program has automatically found differences with the CRS.

There should be no conflict between overlapping sequences generated from the same DNA extraction.



-Which sequence is forward and which is reverse?

-What is the explanation for that most 3’ polymorphism?

Length heteroplasmy relative to the Cambridge Reference Sequence has led to database entry discordance. The FBI has advocated rules for promoting consistency.

The common problem occurs when the sample has a deletion relative to the CRS. See Appendix 1 in Budowle et al.

A standard system of alignment with the CRS is necessary for a database search to be meaningful.

How should these be aligned?

Evidence: ATTGATGTC

CRS: ATTGAATGTC

Evidence: ATTG–ATGTC
CRS: ATTGAATGTC

or

Evidence: ATTGA–TGTC
CRS: ATTGAATGTC

Example: Two alternative gap placements. Each implies an indel, but no other mutation hypothesis required.

Recommended standards.

1. Use the least possible number of hypothesized mutations (gap at one site or a point mutation).
2. When choosing between two scenarios with the same number of mutations, prefer a gap over a transition over a transversion.
3. An indel should be located as far 3’ as possible on the light strand, and if possible indels should be combined.
4. Gaps should be combined only if this doesn’t increase the total number of mutations.

Evidence: ATTG–ATGTC
CRS: ATTGAATGTC

or

Evidence: ATTGA–TGTC ☆
CRS: ATTGAATGTC

☆=the preferred alignment

```
Evidence: AAAACCTCCCCCCATGCT
CRS:      AAAACCCCCTCCCCATGCT

Evidence: AAAACCTCC—CCCCATGCT
CRS:      AAAACCCCCTCCCCATGCT
          gap + one transition

Evidence: AAAACC—TCCCCCCATGCT
CRS:      AAAACCCCCTCCCCATGCT
          gap + two transitions

Evidence: AAAACCTCCCCCC—ATGCT
CRS:      AAAACCCCCTCCCCATGCT
          gap + two transitions
```