B = blue   Y = yellow
R = red   G = green

a) ACG   CGG   GGA   GAC   ACG   CGT   GTA   TAA   AAA

GAG   AGT

GCT   CTA

b) GGA CGT

GGA   GAC   ACG   CGT

B   B   B   B   B
R   Y   Y   Y   =   Y
Y
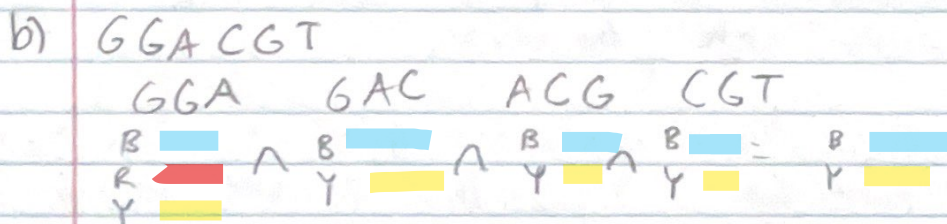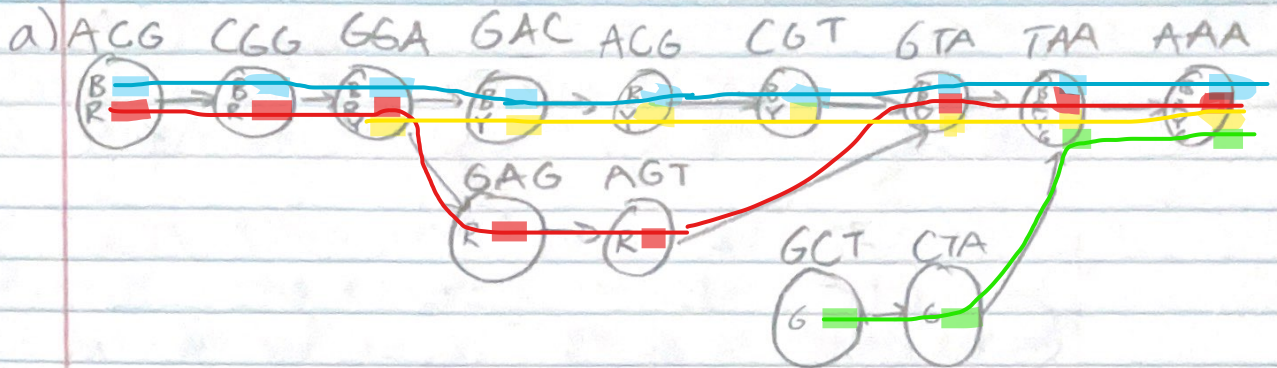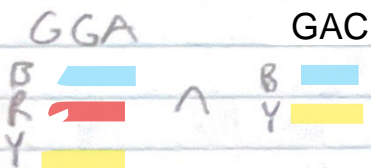
With skips

GGA   GAC   skip to next infomative k-mer (node that
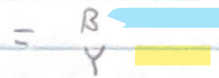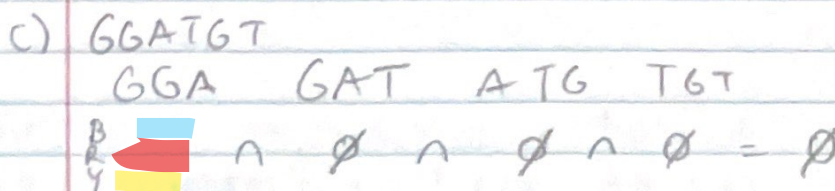   redefines equivalence class)

B   B   B
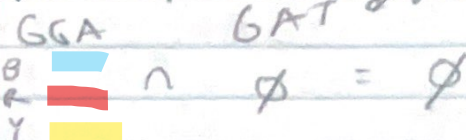R   Y   =   Y   we don't need to skip to CGT
Y   since it does not redefine the
                                         equivalence class

c) GGATGT

GGA   GAT   ATG   TGT

B
R   ∧   ∅   ∧   ∅   ∧   ∅ = ∅
Y

With skips   next informative k-mer

GGA   GAT

B
R   ∧   ∅   =   ∅
Y

→

d) We can backtrack:
   Option 1: When we intersect and find a k-mer with the null set, report the last non-empty set

   Option 2: When we intersect and find a k-mer with the null set, we know the error is contained in the next k bases (3), so skip the next 3 k-mers and continue psuedoalignment
   (nodes)

   Option 1 is the fastest but loses data and has false alignments if the equivalence set intersection gets smaller after the error.

   Option 2 is slower and loses the 3 k-mers that are skipped, but as long as the skipped nodes are not breakpoints (nodes with a different equivalent set compared to the last), we still get the correct alignment.

   GGATGT

   For the previous read error, since the error 'T' happens at a breakpoint node, we would end up misaligning the read.

e) TTTACG. This read is a reverse complement of the last 6 bases of gene 1; we psuedoalign it as follows

Last 4 nodes of graph

CGT   GTA   TAA   AAA   There are 2 strands
                                in DNA: the forward
Reverse   strand and the
complement   ACG   TAC   TTA   TTT   reverse complement strand
the nodes