CS-576 HW2
Sparsh agarwal
9075905142

A2. 1. The substitutions are:
('C', 'A', 1093)
('G', 'A', 9025)
A2. 2. SInce the mutation is within the base range of gene that cleaves into functional proteins. The mutation may make it more dangerous

A3.1. No this will not result in true genome because AGTCG and TCGTG have maximum overlap and if we combine these two by greedy approach, the third fragment(CGATC) will have no overlap with the resulting fragment and will be left alone.
A3.2. If k=4 and genome is AGTCGATCGTG
The kmers will be AGTC, GTCG, CGAT, GATC, TCGT, CGTG
Using de Bruijn approach, graph with (k-1)mers will be
AGT, GTC, GTC, TCG, CGA, GAT, GAT, ATC, TCG, CGT, CGT, GTG
AGT, ATC, CGA, CGT, CGT, GAT, GAT, GTC, GTC, GTG, TCG, TCG
The edges will be:
GTC->TCG
CGA->GAT
GAT->ATC
TCG->CGT
CGT->GTG
The maximum connected graph obtained will be:
AGT->GTC->TCG->CGT->GTG
Since there are no edges to connect CGA, GAT and ATC, the final genome cannot be obtained given the above reads using spectral method, if k is 4.

A4.The SBH graph of circular genome will have one extra edge connecting the last (k-1)mer and first one, when compared to that of linear genome(even if there are no repetitive kmers)