Ami Cooper

n ca2 ᵐ 2 M G *

1. Data were taken from the ncbi website. no manual trimming was done; just took the full genome file in fasta format as it was. Species is Sulfolobus solfataricus.

2. **Part 1: Most Frequent Kmer in forward and rev complement**
   Input: string of genomic data
   Output: reverse complement of string and most frequent 9 and 13mers and the frequency of said kmers.

   Part 2: computing GC skew
   Input: string of genomic data (same as before)
   Output: list of values of GC skew ((C-G)/(C+G)) at each index and a plot of that list of values against the base pair number.

   Part 3*: sign changes**
   Input: list of values of GC skew
   Output: indices at which the GC skew changes from positive to negative
   ****** this part was unsuccessful compared to the known value of where the paper says the***

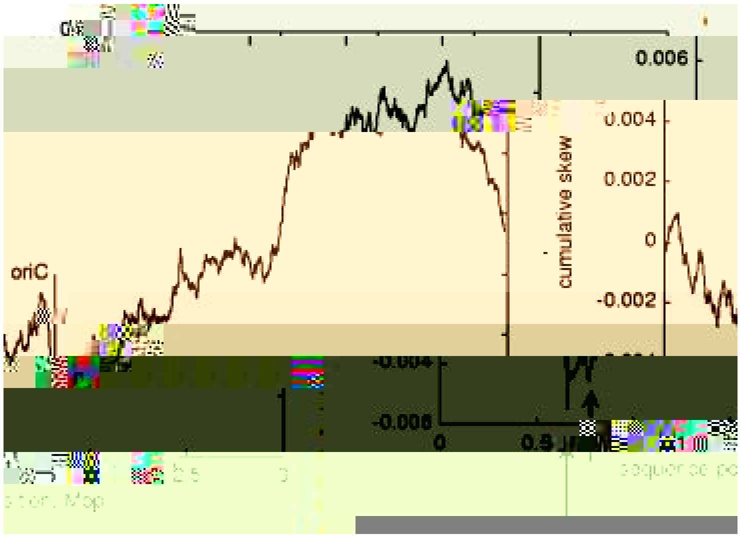

Figure 1.1 borrowed from Contrussi et al.’s Identification and autonomous replication capability of a chromosomal replication origin from the archaeon Sulfolobus solfataricus. Solid arrow indicates the origin of replication. However, another paper identifies two additional origins of replication that are not annotated by base pair location in that paper and therefore are not annotated here.

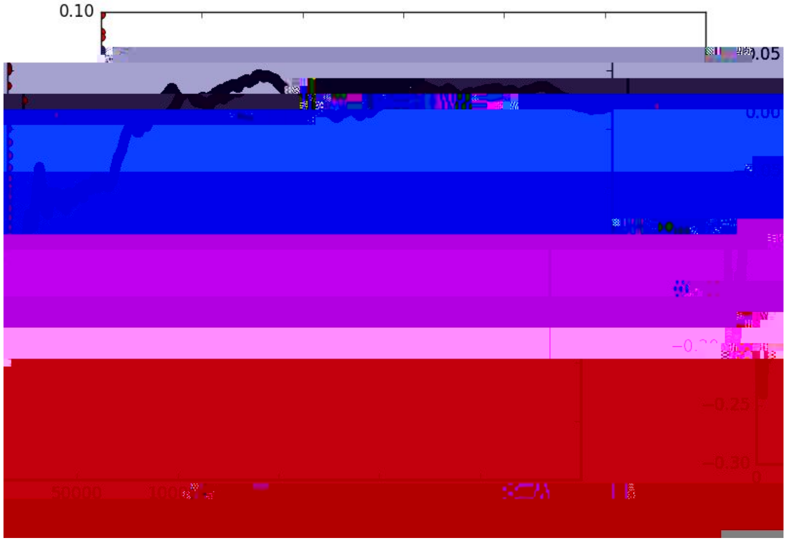

Figure 1.2 Plotted output of my

genome all of the kmers were located along the strand because their distribution might give a