**Hierarchical Clustering**

**Team members:**

2015A7PS0080H: Navneet Kumar

2015A7PS0107H: Rajat Biswas

2015A7PS0111H: Abhijeet Bajaj

2015AAPS0239H: Vatsal Joshi

**Dataset used** = Human Gene DNA sequences

**Preprocessing:-**

Calculation of the proximity matrix to determine dissimilarity between DNA sequences.

**Formulas:-**

Average formula for agglomerative:

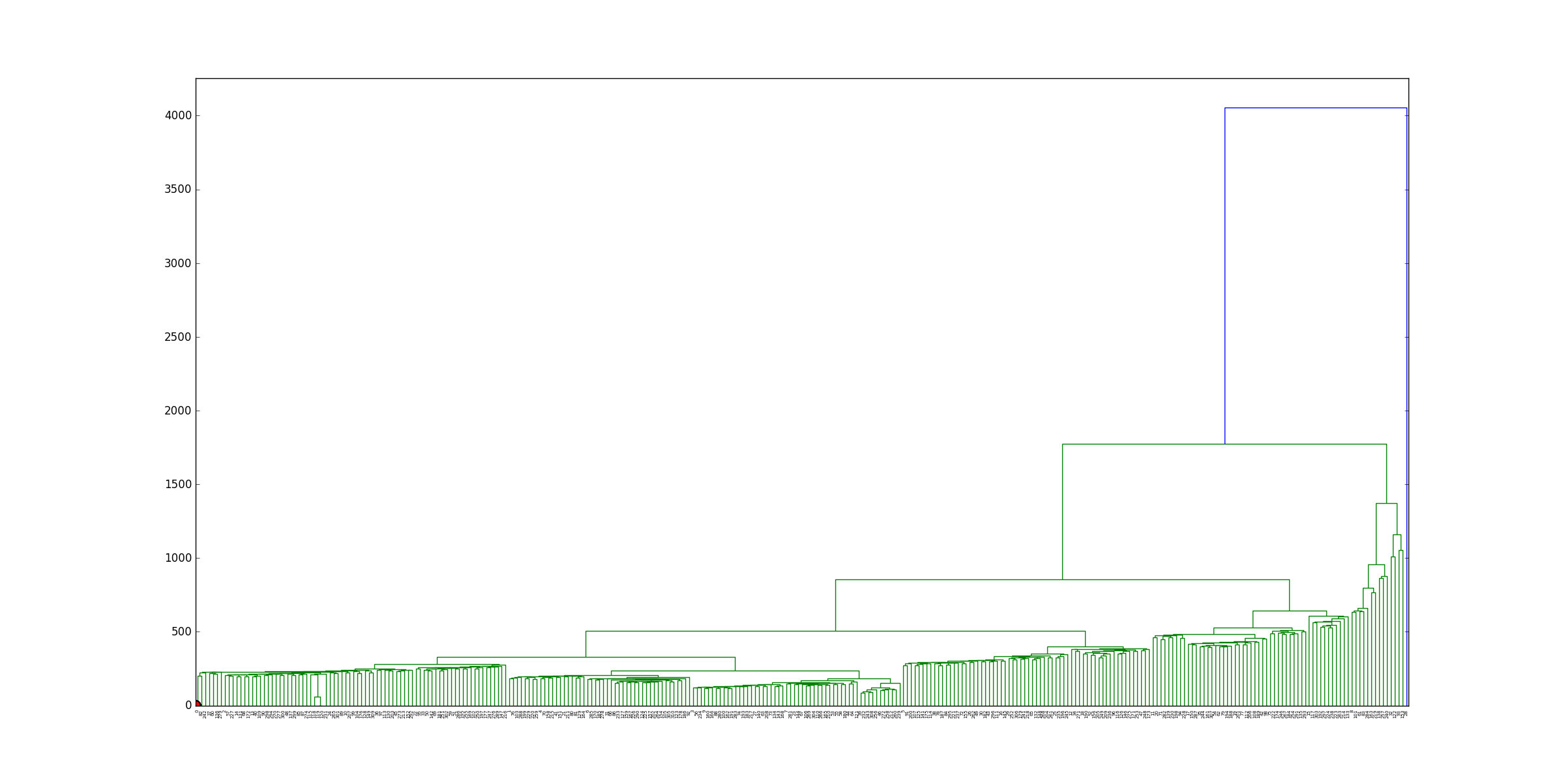
Max formula for divisive:

Diameter(Clusteri) = max p\_i,p\_j∈ Cluster\_i (proximity(p\_i,p\_j))

Linkage: For agglomerative is average and for divisive is complete(max).

Distance metric:-A variation of longest common subsequence where each mismatch has penalty of 1, match has penalty of 0 and gap has penalty of 2.

Dendrogram of Agglomerative:



Dendrogram of Divisive:

