Practical on genetic drift, mutation and divergence

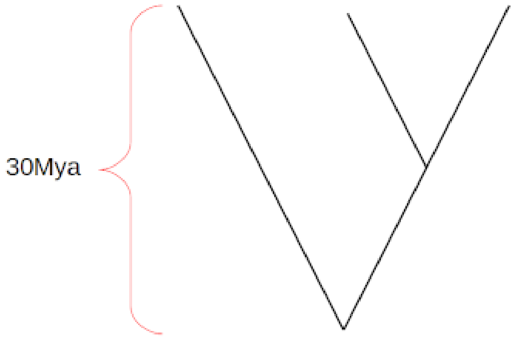
You are interested in the divergence time of three species of gecko.
You obtained some samples and sequenced a fraction of their genome.
Specifically you have access to 20kbp for 10 individuals from each species of:

- western banded gecko
- bent-toed gecko
- leopard gecko

Therefore, you have data for 30 diploid individuals, 10 for each species.
The genomic data is stored in three .csv files, each one named after its species:
`western_banded_gecko.csv`, `bent-toed_gecko.csv`, `leopard_gecko.csv`.

Obtain an estimate of the divergence time between bent-toed and western banded geckos assuming that:

- these three species have a most recent common ancestor 30 million years ago;
- the topology of the species tree is:



A suggestion would be to calculate the genetic divergence for each pair of species to assign leaves to the proposed topology.
What is the genetic divergence? It is the proportion of sites which are fixed for different alleles between populations/species.
Consider this example of 4 DNA sequences from two different species:

Species:1	A A G C A
Species:1	A G G C A
Species:2	G G G T G
Species:2	G C G C G

In this case, positions 1 and 5 are the only diverged sites, as they are fixed for different alleles in different species. Sites 2 and 4 cannot be considered as they are polymorphic within species. Site 3 is a non-diverged position. Thus, the genetic divergence here is 2/3. In fact, we need to exclude (ignore) polymorphic sites for this analysis, and as such we have only 3 sites analysed here, of which 2 are subsitutions (divergent sites).