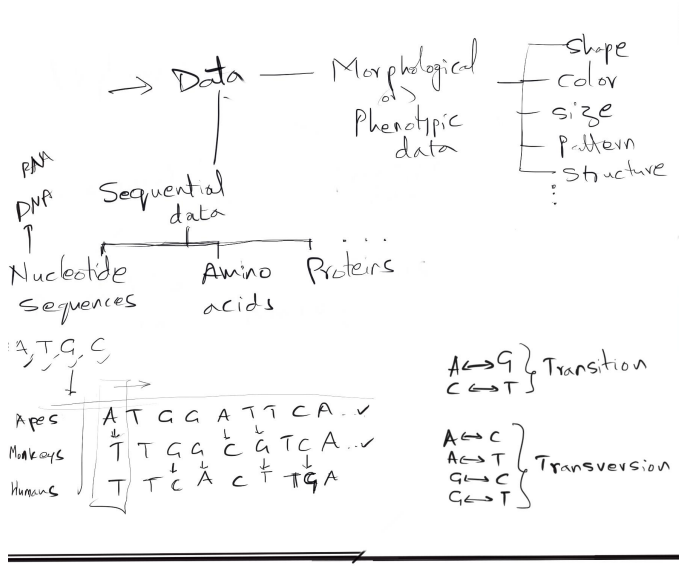
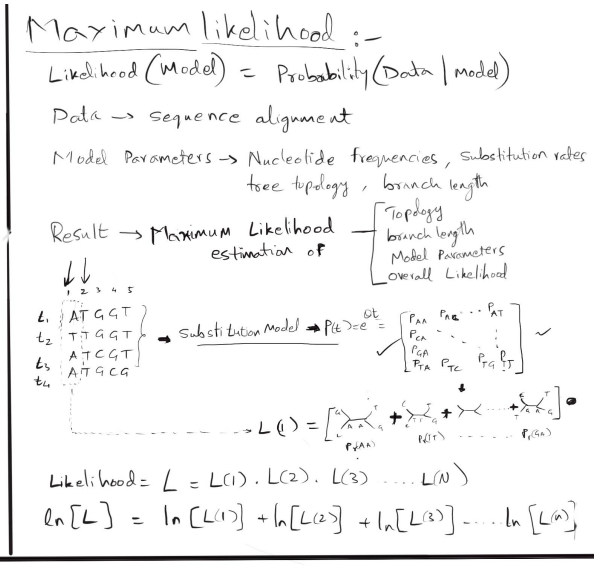
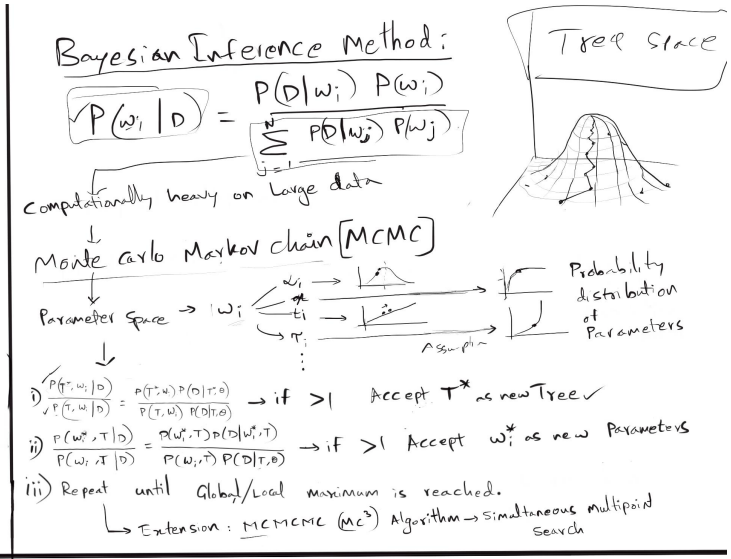
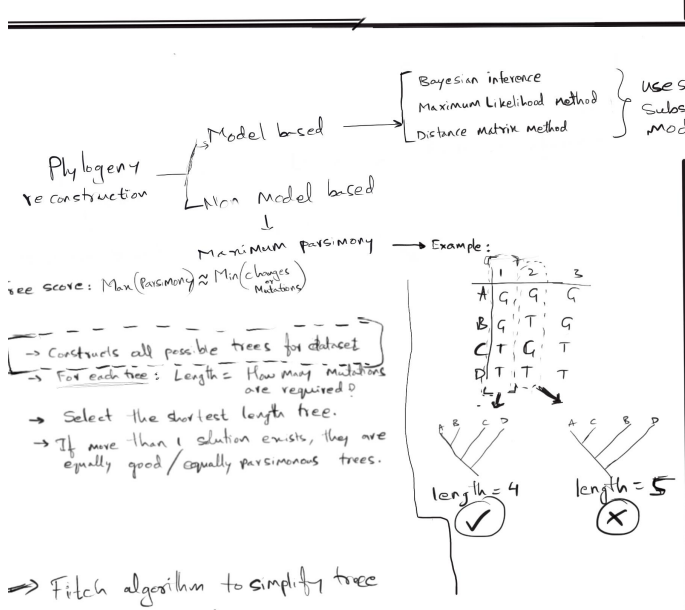
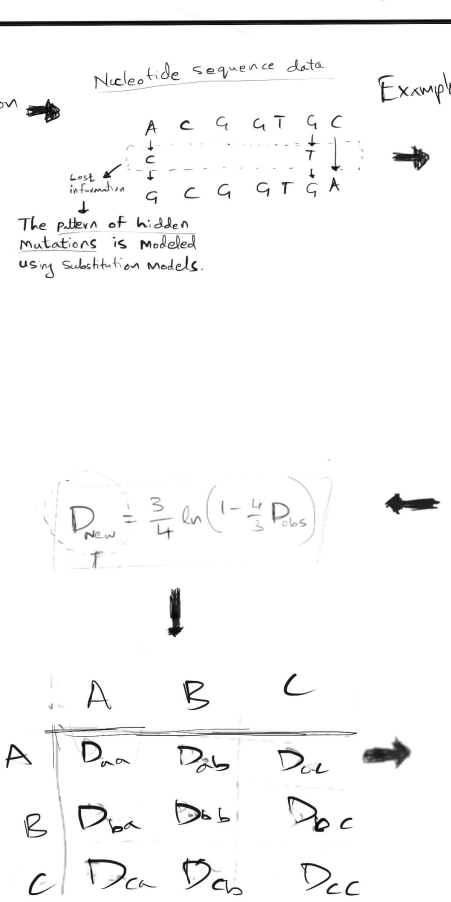
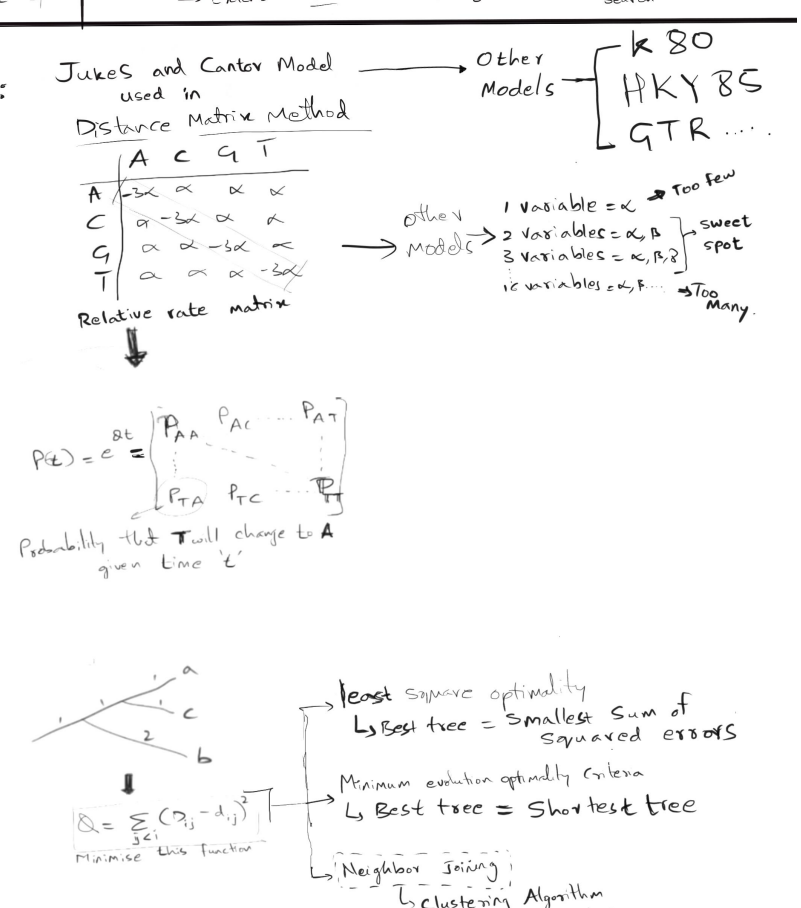

Data
Morphological or Phenotypic data
Shape
Color
size
Pattern
Structure
RNA
DNA
Sequential data
Nucleotide sequences
Amino acids
Proteins
A T G C
Apes
Monkeys
Humans
A T G G A T T C A
T T G G C G T C A
T T C A C T T G A
A ⟷ G } Transition
C ⟷ T
A ⟷ C
A ⟷ T
G ⟷ C
G ⟷ T } Transversion


Maximum likelihood :-
Likelihood (Model) = Probability (Data | model)
Data → sequence alignment
Model Parameters → Nucleotide frequencies, substitution rates
tree topology, branch length
Result → Maximum Likelihood estimation of
Topology
branch length
Model Parameters
overall Likelihood
Substitution Model
Likelihood = L = L(1) . L(2) . L(3) ..... L(N)
ln[L] = ln[L(1)] + ln[L(2)] + ln[L(3)] ..... ln[L(n)]


Bayesian Inference Method:
P(wi | D) = P(D|wi) P(wi) / Σ P(D|wj) P(wj)
computationally heavy on Large data
Monte carlo Markov chain [MCMC]
Tree Space
Parameter Space → wi
Probability distribution of Parameters
Assumption
i) P(T*, wi | D) / P(T, wi | D) = P(T*, w) P(D|T*, θ) / P(T, w) P(D|T, θ) → if >1 Accept T* as new Tree
ii) P(wi*, T | D) / P(wi, T | D) = P(wi*, T) P(D|wi*, T) / P(wi, T) P(D|T, θ) → if >1 Accept wi* as new Parameters
iii) Repeat until Global/Local Maximum is reached.
Extension : MCMCMC (MC3) Algorithm → Simultaneous multipoint search


Phylogeny reconstruction
Model based
Non Model based
Bayesian inference
Maximum Likelihood method
Distance matrix method
uses Substitution Models
Maximum parsimony → Example:
tree score: Max(Parsimony) ≈ Min(changes / Mutations)
→ Constructs all possible trees for dataset
→ For each tree: Length = How many mutations are required?
→ Select the shortest length tree.
→ If more than 1 solution exists, they are equally good / equally parsimonous trees.
length = 4
length = 5
→ Fitch algorithm to simplify tree space search
→ other extensions
↳ PAUP
↳ MEGA (Branch and Bound)
↳ TNT


Nucleotide sequence data
Example:
Lost information
The pattern of hidden mutations is modeled using substitution models.
D_New = 3/4 ln(1 − 4/3 D_obs)
T
A B C
A
B
C


Jukes and Cantor Model used in Distance Matrix Method
Other Models
K 80
HKY 85
GTR ....
Relative rate matrix
other Models
1 variable = α → Too few
2 variables = α, β
3 variables = α, β, δ
sweet spot
i6 variables = α, β, ... → Too Many
P(t) = e^Qt
Probability that T will change to A given time 't'
a
b
c
z
Q = Σ_{j<i} (z_ij − d_ij)^2
Minimise this function
least square optimality
↳ Best tree = Smallest Sum of squared errors
Minimum evolution optimality criteria
↳ Best tree = Shortest tree
Neighbor Joining
↳ Clustering Algorithm
↳ Unique resultant tree