# Folder matrix:

**Excel\_results**: including IGCnum, posterior in excel files.

**Reconstruction likelihood:**

*arg\_paralog1\_paralog2\_MG94\_IGC\_node\_i.txt*: The first ten composite states that have biggest likelihood. The arg number of state can be parsed by divide 62. Python program is like this:

bases = 'tcag'.upper()

codons = [a+b+c for a in bases for b in bases for c in bases]

amino\_acids = 'FFLLSSSSYY\*\*CC\*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG'

codon\_table = dict(zip(codons, amino\_acids))

codon\_nonstop = [a for a in codon\_table.keys() if not codon\_table[a]=='\*']

codon\_to\_state = {a.upper() : i for (i, a) in enumerate(codon\_nonstop)}

state\_to\_codon = {i : a.upper() for (i, a) in enumerate(codon\_nonstop)}

state\_to\_codon[61] = 'xxx'

def state\_to\_compositecodons(state):

state\_1, state\_2 = divmod(state, 62)

state\_1 = int(state\_1)

state\_2 = int(state\_2)

return (state\_to\_codon[state\_1],state\_to\_codon[state\_2])

In which state\_to\_codon[0]~state\_to\_codon[60] means Codons, state\_to\_codon[61] means absorb. We just need to look those state\_to\_codon[0]~state\_to\_codon[60]

I believe similar R can be written.

*likelihood\_paralog1\_paralog2\_MG94\_IGC\_node\_i.txt*: The likelihood of the state indicated by arg.

**sitewise\_IGC\_statmatrix:**

the results’ are indicated by the folder. It is sorted by paralog pairs, every matrix is site\*node size.

# Folder PAML fasta:

The reconstructed alignment of PAML

# Desired\_result:

It is preliminary result, it shows the sites that are difference in our IGC or tau=0, it is not needed now.

# Model\_likelihood:

The first row is likelihood while the second row is AIC.

# Parameters:

Branch\_length: it is the same order with the branch length save txt defined by Xiang. So it is the same order with the order in our branch wise analysis data.

Tau: indicated by itself

Transition model: the first four is the rate of AGCT

The fifth is kappa, the sixth is omega.

# PAML likelihood

Ancestral\_reconstruction\PAML\output\summary

This is where I save PAML likelihood and AIC

# IGCevent\_result.txt

Is the result of multi-nucleotide change and can be result by IGC.

When select all and copy to an excel it can be seen clearly.

The excel is the excel file with same name.