**Markov Model:**

Gives probability of any nucleotide following a given nucleotide

Markov transition matrix:

Both markov chains and mutinomial models are used

**GC Content:**

GC Content = (G + C / A + T + G + C) \* 100

It is important because it differs greatly between certain organisms

Can be used to identify horizontal gene transfer

Coding regions have a higher GC content

**k-mer Frequency Analysis:**

analyzing the frequency of “words” in the genome

1 mer words: A, T, C, G

2 mer words: AA, AT, AG, AC, …

3 mer words: AAA, AAT, …

Words that are highly frequent/infrequent have biological meaning

CTAG kinks DNA

Palindromic k-mers: restriction sites: EcoRI / GAATTC

Often they are singals for restriction enzymes (cut DNA in middle of palindrome)

**Odds Rations:**

Observed/expected

Example for AT 2-mer

P(AT) / (P(A) \* P(T))

Odds ratio well above/below 1 indicates something unusual