GENOME STATISTICS

GENOME ASSEMBLY

NULL MODELS:  **a pattern-generating model that is based on randomization of eco-** **logical**. **data or random sampling from a known or imagined distribution**.

EXAMPLE :

* random DNA-sequence

SEQUENCE ANALYSIS:

* framework for mutational process inference
* mutation is heavily affected by the neighbor bases
* context : where point mutation is in the sequence

counting gives you different answer

• evaluate nucleotide fractions by dividing the counts by total counts:

{0.23913, 0.217391, 0.210145, 0.333333}

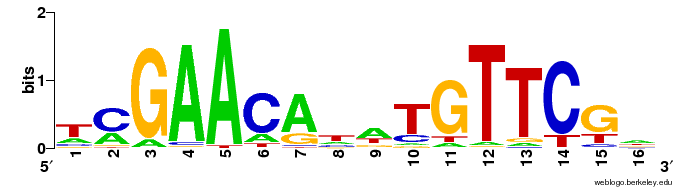
🡪 can be used to create random sequence (presents)

* changing null model 🡪 should still explain the real data (then the unknown is interesting )
* creating the null model (random sequence) so that you choose what kind of algorithm

64 codons 🡪 20 amino acids + stop codon

* mutation in second codon 🡪 different amino acid
* amino acid change doesn’t always make a difference

transcription factor can recognize somewhat mutated binding sites



position 16 🡪 similar to null model

vertaillaan null model vs. POSITION WEIGHT MATRIX

* mutation annotation 🡪 what mutation creates

CONTEXT OF THE TF and DNA binding sites

* more things to take into consideration that the sequence itself
* human binding sites are shorter than procaryotic

SUMMARY:

* comparing
* 1 PROBLEM
* 2 PROBLEM BINDING SITE STATISTICS