Bio-informatica aantekeningen

Plan van aanpak, parameters van BLAST

Pairwise alignment: the process of lining up two sequences to achieve maximal levels of identity (and conservation, in the case of amno acid sequences) for the purpose of assessing the degree of similarity and the possibility of homology.

Gaps toevoegen voor betere score

Structural analysis has shown that fewer indels occur in sequences of structural importance and that insertion tend to be several residues long rather than just a single residue long.

Waarom kan een pariwise alignement met proteïne sequenties informatiever zijn dan een aligment met DNA sequenties?

Eiwitten zijn geconserveerd, silent mutaties in het DNA.

Similarity

The extent to which nucleotide or protein sequences are related.

Identity

The extent to which two sequences are invariant.

Homoloog

2 sequenties zijn voor 30% identity (>100 a.a.) is een veilige grens. Lager dan 30% kan nog steeds homoloog zijn aan elkaar. Lage conservatie van letters, maar vorm blijft dan behouden.

Orthologs

Homologe eiwitten in verschillende soorten, ontstaan door soortvorming. Evolutionaire druk op 1 versie is groter, mag niet muteren.

Paralogen

Verschillende vormen van een eiwit in 1 organisme, ontstaan door gen-duplicatie. Minder evolutionaire druk op de genen, er is ruimte voor mutaties.

Door gen-duplicatie wordt de druk om te muteren kleiner omdat er meerdere duplicaties van een gen zijn.

Bacterien hebben alleen orthologen, ze hebben kompacte genomen.

Quality of an alignment is measured by giving it a quantitive score.

For eacht alignment a score is calculated

The alignment that gives the best score is referred to as the optimal alignment

There is no scoring scheme that perfectly models the evolutionary process: therefore an optimal alignment will not necessarily be the biologically correct one .

Substitution matrix:

For alignments of protein sequences, the score is assigned to each aligned pair of amino acids using a substitution matrix

What matrix schores represent

Positive scores = larger probability of meaningful match

Negative scores = larger probability of random paring

Score of 0 = equally likely that the pairing is meaningful as that it is random.

PAM accepted point mutation score matrix

PAM1 score matrix calculated from a set of very similar sequences (1% divergence)

From the alignment of these sequences the probability was calculated that any one amino acid would mutate into any other over a given period of evolutionary time

Probabilities are converted to logarithms so that the final alignment score can be calculated by summation of the individual scores from aligned pairs of amino acids.

PAM1: 1 mutation per 100 residues

PAM10: 10 mutations per 100 residues, 10 mutaties per 100 aminozuren.

PAM250: 250 mutations per 100 residues, op dezelfde plek kunnen vaker mutaties plaats vinden.

The cells in a log odds matrix consist of an ‘odds ratio’

The probability that an alignment is authentic, the probability that the alignment was random.

PAM 80, voor beide soorten sequenties een goede matrix

BLOSUM Matrices are based on local alignments

BLOSUM stands for blocks substitution matrix

BLOSUM 62 is a matrix calculated from comparisons of sequences with at least 62% identity. Is vergelijkbaar met PAM70/PAM80

Blast is een local alignment.

Aantekeningen bio-informatica 18-5-18

Geen resultaten?

* Lage quality score
* Blastx, nr database 🡪 geen coderende regio. Intron
  + blastn, nr database 🡪 proberen, op dna niveau.
* Organisme
* Functie eiwit
* Domeinen/ moleculaire functie
* Locatie in de cel / topologie
* Biologisch proces.

Score: een hoogte van de score is een maat van de kwaliteit van de alignment.

Evalue: het aantal random sequenties uit de database die eenzelfde of hogere bitscore geven. Statistische betrouwbaarheid. Evalue hangt af van:

* Groote van je database
* Groote van je query
* Hoogte van je score

Positves = Identitys en simmularitys

Blast algoritme:

* Breekt sequentie in ‘woorden’
* Gaat de ‘woorden’ muteren, om te kijken wat er nog meer zou kunnen
* Met die woorden gaat hij dan zoeken in de database.
* Wanneer 2 woorden van 1 woord matched, gaat hij de match uitbreiden.
* Wanneer de match minder wordt stopt blast met blasten en returnt de hit.

Doe van algoritmes:

* Zo hoog mogelijke score verkrijgen.

2 soorten gap penalties

* Openen van een gap
* Gap verlenging 🡪 gap elongation
* Gebruikt om alignments en de verdeling van gaps daarin homogener te maken. Op deze manier is de alignment representatiever evolutionair gezien.

Hydrofobe 20-25 aminozuren helix ben je door de membraan heen

Verslag 🡪 accessiecode van eiwit van uniprot.

