Les techniques de séquençages se basent sur la découpe de polymère et l’analyse en petits fragments. Les

1. Séquençages
2. Assembleur
3. Alignement

C=NL/GC=NL/G.

G taille totale

L taille des fragments

# Séquençage

Parameter important dans le séquençage.

Read lengths:

Error rates and types of errors:

# Assembleur

Principale difficulté est la redondance de motif.

1. There are errors in the reads, and therefore looking for exact alignment is challenging.

2. The genome is quite repetitive, and therefore a read may have several places it can align to. In fact, there are regions of the genome (e.g. Alu repeats) that are repeated millions of times.

3. Recall above that we need to align N=N= 900,000,000 reads to a length 3B genome. Therefore the naive process of scanning the entire genome for each read’s match is too slow.

Quantifier l’ARN expectation–maximization (EM) algorithm

Mutation ponctuelle accepté mutation d’aa qui n’ont pas affecté la fonction de la protéine.

# Alignement

% de similitude d’un alignement = nbre d’acides aminées bien alignée.

Les alignements globaux sont plus fréquemment utilisés lorsque les séquences mises en jeu sont identiques et de taille égale. Une technique générale, nommée algorithme de Needleman-Wunsch est basée sur la programmation dynamique.

Les alignements locaux sont plus fréquemment utilisés lorsque deux séquences dissemblables sont soupçonnées de posséder des motifs identiques malgré l'environnement. L'algorithme de Smith-Waterman est une méthode d'alignement local générale basée aussi sur la programmation dynamique.

Avec des séquences suffisamment semblables, il n'y aucune différence dans les résultats.

Des méthodes hybrides, des méthodes semi-locales, s'avèrent utiles lorsque...

Match value — Value assigned for matching characters

Mismatch value — Value assigned for mismatching characters

Gap penalty — Value assigned for spaces

Global Needleman-Wunsch algorithm.

Local Smith-Waterman algorithm

Score pour un match, mismatch, gap, extension de gap