

Bancos de Dados Públicos

- ⇒ Primários: nucleotídeos e aminoácidos
- ⇒ Secundários: proteínas
- ⇒ Principais primários: GenBank, EBI, DDBJ (DNA data bank of Japan e PDB (repositórios de seqs. e interligados para atualização, exigidos para publicação)
- ⇒ Principais secundários (usam info. dos primários): Swiss-Prot, PIR (protein information resource)

Bioinformática – Visão Geral

Bancos de Dados Públicos

- ⇒ Swiss-Prot: função, domínios funcionais, proteínas homólogas
- ⇒ Bancos estruturais: estrutura de proteínas
- ⇒ Bancos funcionais: KEGG (Kyoto Encyclopedia of Genes and Genomes) mapas metabólicos
- ⇒ Lista de todos os bancos (Baxevanis):
<http://www3.oup.co.uk/nar/database>

Genbank: www.ncbi.nlm.nih.gov

EBI: www.ebi.ac.uk

PDB: www.rcsb.org/pdb

Swiss-Prot: <http://us.expasy.org/sprot/>

KEGG: www.genome.ad.jp/kegg

Bioinformática – Visão Geral

Alinhamento de Sequências

- ⇒ Uma das operações mais importantes em bioinformática
- ⇒ Busca similaridade entre 2 ou mais seqs. ou entre partes delas
- ⇒ Alinhamento indica grau de similaridade o que é diferente de homologia que indica relação evolutiva!
- ⇒ Programas (on-line): ClustalW(X), Multialin, FASTA, BLAST2sequences, etc.

Bioinformática – Visão Geral

Alinhamento de Sequências

(a)

```
HBA_HUMAN  GSAQVKGHGKQVADALTNVAHVDDMPNALSALSDDLHAHKL
              G+ +VK+HGKKV  A++++AH+D++ +++++LS+LH  KL
HBB_HUMAN  GNPVKAHGKQVLGAFSDGLAHLNLLKGTTFATLSELHCDKL
```

(b)

```
HBA_HUMAN  GSAQVKGHGKQVADALTNVAHV---D---DMPNALSALSDDLHAHKL
              ++ +++++H+ KV  + +A  ++                +L+ L+++H+ K
LGB2_LUPLU NNPELQAHAGKQVFKLVYEAAIQLOQVTGVVVTDATLKNLGSVHVSKG
```

(c)

```
HBA_HUMAN  GSAQVKGHGKQVADALTNVAHVDDMPNALSALSDDL---LHAHKL
              GS+ + G +    +D L  ++ H+ D+  A +AL D    ++AH+
F11G11.2   GSGYLVGDSLTFVDLL--VAQHTADLLAANAALLDEFFPQFKAHQE
```

Figure 2.1 Three sequence alignments to a fragment of human alpha globin. (a) Clear similarity to human beta globin. (b) A structurally plausible alignment to leghaemoglobin from yellow lupin. (c) A spurious high-scoring alignment to a nematode glutathione S-transferase homologue named F11G11.2.

Usa-se sistema de pontuação (gaps, matches, mismatches) ou matrizes de Substituição no alinh. de proteínas

Bioinformática – Visão Geral

Alinhamento de Sequências

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	5	-2	-1	-2	-1	-1	-1	0	-2	-1	-2	-1	-1	-3	-1	1	0	-3	-2	0
R	-2	7	-1	-2	-4	1	0	-3	0	-4	-3	3	-2	-3	-3	-1	-1	-3	-1	-3
N	-1	-1	7	2	-2	0	0	0	1	-3	-4	0	-2	-4	-2	1	0	-4	-2	-3
D	-2	-2	2	8	-4	0	2	-1	-1	-4	-4	-1	-4	-5	-1	0	-1	-5	-3	-4
C	-1	-4	-2	-4	13	-3	-3	-3	-3	-2	-2	-3	-2	-2	-4	-1	-1	-5	-3	-1
Q	-1	1	0	0	-3	7	2	-2	1	-3	-2	2	0	-4	-1	0	-1	-1	-1	-3
E	-1	0	0	2	-3	2	6	-3	0	-4	-3	1	-2	-3	-1	-1	-1	-3	-2	-3
G	0	-3	0	-1	-3	-2	-3	8	-2	-4	-4	-2	-3	-4	-2	0	-2	-3	-3	-4
H	-2	0	1	-1	-3	1	0	-2	10	-4	-3	0	-1	-1	-2	-1	-2	-3	2	-4
I	-1	-4	-3	-4	-2	-3	-4	-4	-4	5	2	-3	2	0	-3	-3	-1	-3	-1	4
L	-2	-3	-4	-4	-2	-2	-3	-4	-3	2	5	-3	3	1	-4	-3	-1	-2	-1	1
K	-1	3	0	-1	-3	2	1	-2	0	-3	-3	6	-2	-4	-1	0	-1	-3	-2	-3
M	-1	-2	-2	-4	-2	0	-2	-3	-1	2	3	-2	7	0	-3	-2	-1	-1	0	1
F	-3	-3	-4	-5	-2	-4	-3	-4	-1	0	1	-4	0	8	-4	-3	-2	1	4	-1
P	-1	-3	-2	-1	-4	-1	-1	-2	-2	-3	-4	-1	-3	-4	10	-1	-1	-4	-3	-3
S	1	-1	1	0	-1	0	-1	0	-1	-3	-3	0	-2	-3	-1	5	2	-4	-2	-2
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	2	5	-3	-2	0
W	-3	-3	-4	-5	-5	-1	-3	-3	-3	-3	-2	-3	-1	1	-4	-4	-3	15	2	-3
Y	-2	-1	-2	-3	-3	-1	-2	-3	2	-1	-1	-2	0	4	-3	-2	-2	2	8	-1
V	0	-3	-3	-4	-1	-3	-3	-4	-4	4	1	-3	1	-1	-3	-2	0	-3	-1	5

Matriz BLOSUM50
 Números indicam a pontuação na ocorrência de pares de aminoácidos, a diagonal é + provável

Figure 2.2 The BLOSUM50 substitution matrix. The log-odds values have been scaled and rounded to the nearest integer for purposes of computational efficiency. Entries on the main diagonal for identical residue pairs are highlighted in bold.

Alinhamento de Sequências

- ⇒ PAM e BLOSUM + utilizadas
- ⇒ PAM (point accepted mutation): matrizes construídas a partir da análise de mutações entre proteínas com 1% de aminoácidos diferentes (divergência)
- ⇒ BLOCKS (block substitution matrices): análise na substituição das seqs. que possuem 50% de similaridade.

Bioinformática – Visão Geral

Alinhamento de Sequências

- ⇒ Alinhamento global e local
- ⇒ Global: as seqs. São alinhadas de um extremo a outro (com apenas um resultado)
- ⇒ Local: alinha-se as regiões mais conservadas que podem ser mais de uma.
- ⇒ Softwares:
- ⇒ BLAST (basic local alignment search tool)
www.ncbi.nlm.nih.gov/BLAST
- ClustalW: www.ebi.ac.uk/clustalw/index.html
- Multialin, Fasta etc.

Bioinformática – Visão Geral

Projetos Genoma e Transcriptoma

⇒ GENOMA

- ⇒ **Genoma shotgun** (pequenos): DNA é fragmentado em pequenos pedaços (seqüenciadores podem ler no máximo 1000 pares de bases)
- ⇒ Clonagem (cópia) em vetores plasmidiais p/ seqüenciamento das extremidades
- ⇒ Montagem e fechamento de gaps
- ⇒ P/ genomas grandes usa-se shotgun hierárquico com BAC's ou YAC's

Bioinformática – Visão Geral

Alinhamento de Sequências

⇒ TRANSCRIPTOMA

- ⇒ **mRNA**: coleta-se moléculas de mRNA do organismo e com a transcriptase reversa cria-se uma biblioteca de cDNA
- ⇒ Sequencia-se as extremidades do cDNA que identificam genes expressos EST's (Expressed Sequence Tags)
- ⇒ Identifica genes em diferentes estágios do desenvolvimento
- ⇒ A maior parte dos projetos genômicos no Brasil usam esta técnica: cana-de-açúcar, boi, vassoura de bruxa, camarão, eucalipto, s. mansoni e outros
- ⇒ As técnicas de shotgun e EST são complementares!

Bioinformática – Visão Geral

Anotação de Genomas

- ⇒ Nucleotídeos → Proteínas → Processos
- ⇒ Onde Genes? → Qual Função? → Interações?
- ⇒ Tipos de seqs: genes, RNA transportador, RNA ribossomal, intergênicas, repetitivas
- ⇒ GenScan <http://genes.mit.edu/GENSCAN.html>
- ⇒ tRNAscan-SE www.genetics.wustl.edu/eddy/tRNAscan-SE
- ⇒ BLAST
- ⇒ RepeatMasker