

IT

1.

(genome) 'chromosome' (Adenine), (Thymine), (Cytosine), (Guanine)

'gene' DNA DNA DNA DNA

15 35 35 10 30 가

6 1000 DNA 30 가

3 97%가 9000 DNA (SNP: Single Nucleotide Polymorphism) 3

가 가 가

DNA

DNA

가

. (.)

가

(local alignment score)

가

IT

2.

. 가

1:1

(pairwise

alignment)

(local alignment)

(global alignment)

(scoring matrix)

(gap cost)

가

PAM

BLOSUM

. PAM

1969 M. Dayhoff 가

BLOSUM 1992 S. Henikoff가

BLAST

FASTA

1 FASTA

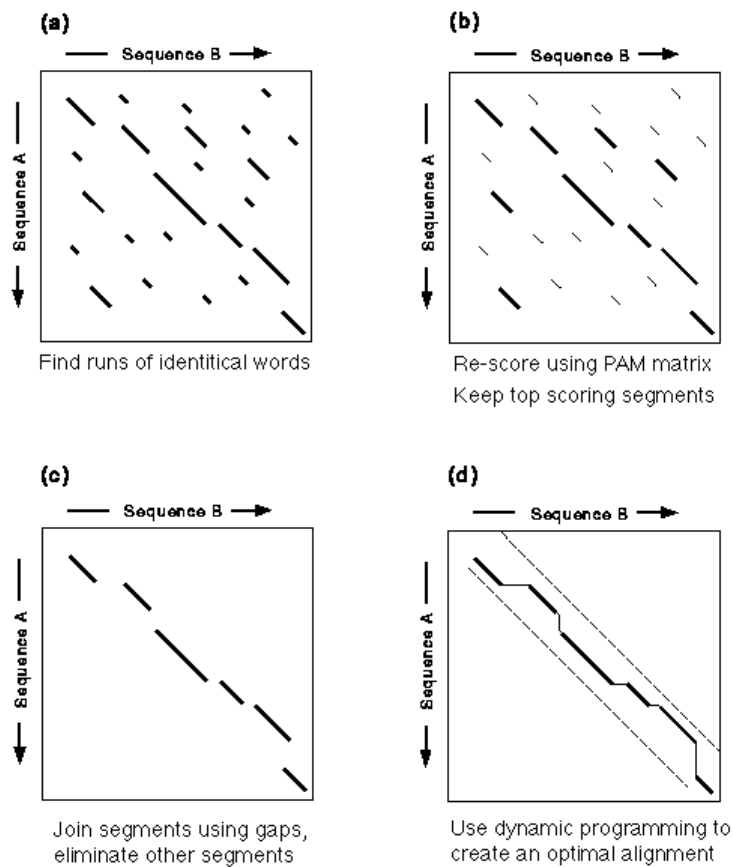
(:

가

)

가 가
가 가
FASTA European
Bioinformatics Institute <http://www.ebi.ac.uk/fasta33>

FASTA Algorithm



1 FASTA

가 1990 Karlin Altschul 가
BLAST(Basic Local Alignment Search Tool) 가
BLAST (local alignment)
FASTA BLAST NCBI(National Center for Biotechnology
Information) <http://www.ncbi.nlm.nih.gov/BLAST/>
가
InterPro (<http://www.ebi.ac.uk/interpro/>)

3. IT

“ ”
 (full-text)
 가
 가 CAFE
 DNA
 KISTI N-gram
 ()
 가 (inverted file) DNA
 가 가 n-gram N-gram
 가 n-gram
 가 N-gram k-tuple w-mer
 ACEPITCH n 4
 n-gram ACEP, CEPI, EPIT, PITC, ITCH가
 n-gram DB
 가

BLOCKS, PRINTS, PFAM, METAFAM, COG

iProClass, INTERPRO,

SwissProt

. PIR-NREF

2004 8 30 178

SwissProt

15

KISTI

KRISTAL-2002

2



2 (ProSeS)

ProSeS(Protein SequenSe Search)

1.

가 (가)

2.

가

3.

가

4.

3 FASTA



3

4

ID

BLAST

가

가

BLAST

가

번호	Accession No.	Description	Score
1	AF040404	Proteinase K (Kluyveromyces fragilis)	100.0
2	AF040405	Proteinase K (Kluyveromyces fragilis)	100.0
3	AF040406	Proteinase K (Kluyveromyces fragilis)	100.0
4	AF040407	Proteinase K (Kluyveromyces fragilis)	100.0
5	AF040408	Proteinase K (Kluyveromyces fragilis)	100.0
6	AF040409	Proteinase K (Kluyveromyces fragilis)	100.0
7	AF040410	Proteinase K (Kluyveromyces fragilis)	100.0
8	AF040411	Proteinase K (Kluyveromyces fragilis)	100.0
9	AF040412	Proteinase K (Kluyveromyces fragilis)	100.0
10	AF040413	Proteinase K (Kluyveromyces fragilis)	100.0

4: (a)

(b) NF00720858

가

5

ProSLP(Protein Subcelluarl Localization Prediction)

가 50

500

ProSeS가 BLAST

6.5

Seq	Label	Location	Weight
1	T10	cytoplasmic protein	30.94%
2	T11	mitochondrial protein	3.12%
3	T12	membrane protein	3.12%
4	T13	cytoplasmic protein	10.24%
5	T14	cytoplasmic protein	10.24%
6	T15	cytoplasmic protein	10.24%
7	T16	cytoplasmic protein	10.24%
8	T17	cytoplasmic protein	10.24%
9	T18	cytoplasmic protein	10.24%
10	T19	cytoplasmic protein	10.24%
11	T20	cytoplasmic protein	10.24%

(a)

5: (a)

Class	Family	Family name	Weight
Cytoplasm	C10000	Cytoplasmic protein	10.24%
	C10001	Cytoplasmic protein	10.24%
	C10002	Cytoplasmic protein	10.24%
	C10003	Cytoplasmic protein	10.24%
Mitochondrion	M10000	Mitochondrial protein	3.12%
	M10001	Mitochondrial protein	3.12%
	M10002	Mitochondrial protein	3.12%
	M10003	Mitochondrial protein	3.12%
Membrane	Me10000	Membrane protein	3.12%
	Me10001	Membrane protein	3.12%
	Me10002	Membrane protein	3.12%
	Me10003	Membrane protein	3.12%
Cytoplasmic	Cy10000	Cytoplasmic protein	10.24%
	Cy10001	Cytoplasmic protein	10.24%
	Cy10002	Cytoplasmic protein	10.24%
	Cy10003	Cytoplasmic protein	10.24%
Cytoplasmic	Cy20000	Cytoplasmic protein	10.24%
	Cy20001	Cytoplasmic protein	10.24%
	Cy20002	Cytoplasmic protein	10.24%
	Cy20003	Cytoplasmic protein	10.24%
Cytoplasmic	Cy30000	Cytoplasmic protein	10.24%
	Cy30001	Cytoplasmic protein	10.24%
	Cy30002	Cytoplasmic protein	10.24%
	Cy30003	Cytoplasmic protein	10.24%
Cytoplasmic	Cy40000	Cytoplasmic protein	10.24%
	Cy40001	Cytoplasmic protein	10.24%
	Cy40002	Cytoplasmic protein	10.24%
	Cy40003	Cytoplasmic protein	10.24%

(b)

(b)

4.

IT

ProSeS

CCBB

가 CCBB
PIR-NREF

ProSeS

BLAST

가

IT

ProSeS가