



# Web

( )

E-mail : [kjpark@smallsoft.co.kr](mailto:kjpark@smallsoft.co.kr)





2000

(post-genome)

I.

2000

1.

6 1 가

1990

가

가

가

DNA  
( G, T )

( A, C, 4가 )

Celera

(10

4  
(genome;  
DNA

, DNA

2000

가 4.5 x 10<sup>6</sup>,

3 x 10<sup>9</sup>

(

(post-genome)

)

20

가

130

(

가

가

),

DNA

가

.( 가 )

DNA

2.

( , )

(data)

, DNA 가

가

1960

21

가가

, 1980

가

가

1990

가가21

2001

가

Science Nature

가

3

10

(bioinformatics)

4

10

Silico Biology

( In

가  
가

programs)

(analysis tools; analysis

(byte)

3 GB 가

가

가

, DNA

가

가

, 가

가

가

2가



# 과학기술종합정보시스템

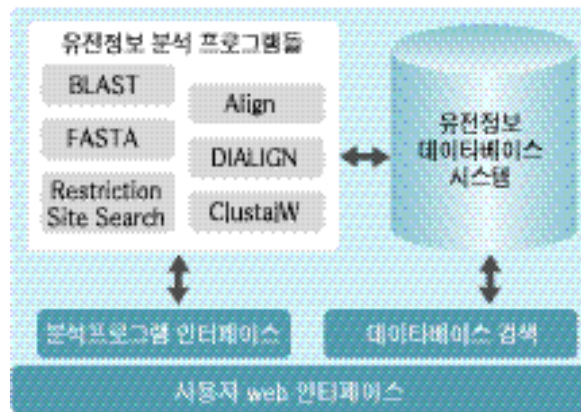
가 , web

, web

1.

가

가



6 DB

가

web

(a)



- DB

1

< 1>

6

8

(b)

web

- CGI - hyperlink

CGI

site

DBMS

Oracle

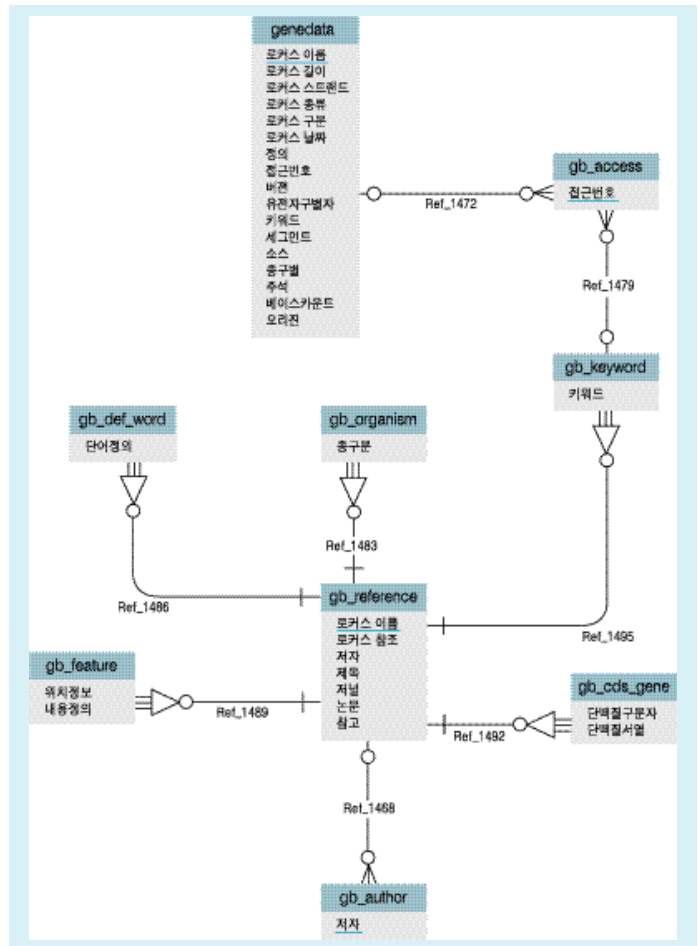
. CGI , C,

upgrad라

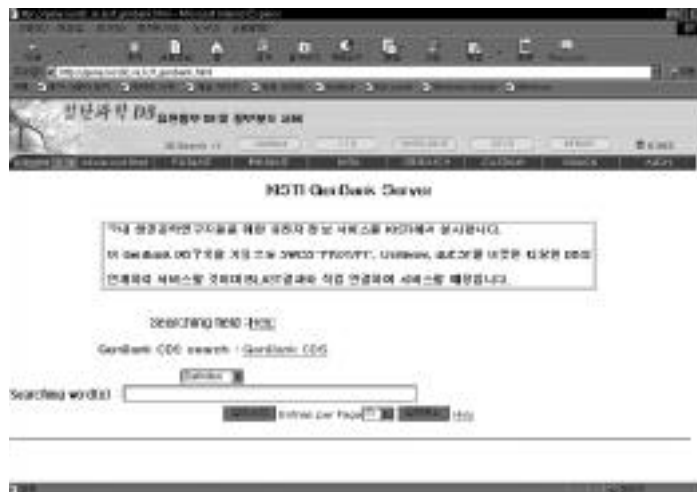
MySQL, Perl, PHP



가 ,  
 .  
 1) DB ORF/CDS  
 GenBank  
 , DB DNA DB  
 가 DNA  
 가 2001 2  
 1000 ,  
 , 가  
 < 2-(a)> ,  
 가 10  
 가 2 가 가  
 ,  
 가  
 GenBank 가  
 GenBank daily update scheme update  
 ,  
 update  
 author, definition, locus  
 < 2-(b)> ,  
 BLAST , format



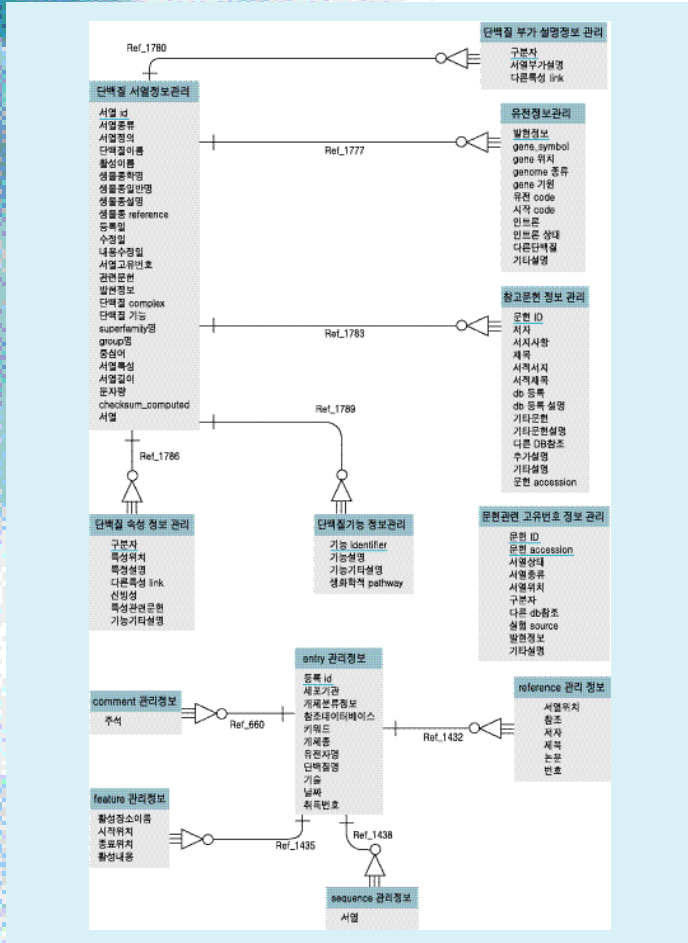
(a) GenBank



(b) GenBank

2) DB PIR SwissProt  
 DB가 SwissProt , PIR  
 가  
 DB ,  
 < 2> DB

# 과학기술종합정보시스템



(a) PIR

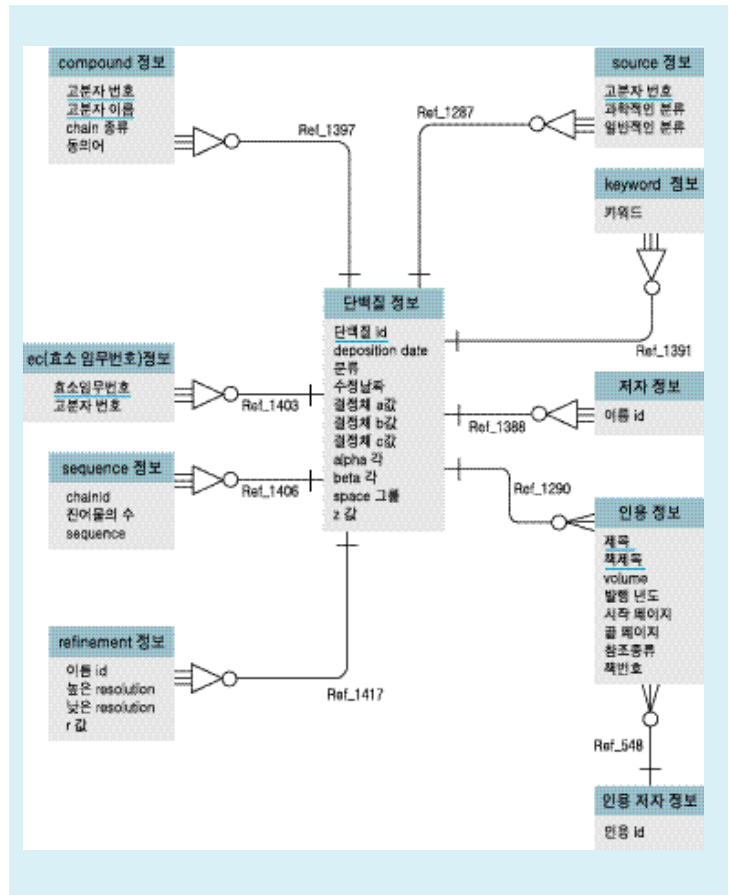


(b) PIR

DB  
 , DB가  
 < 3-(b)>. DB <  
 3-(a)> DB  
 , PIR  
 CGI  
 web  
 3) DB  
 PDB . PDB  
 X- NMR  
 3  
 가  
 가  
 ID ,  
 5가  
 < 4-(a)> CGI  
 < 4-(b)>. 3  
 가  
 가 VRML 가  
 가  
 web CGI  
 4) DB

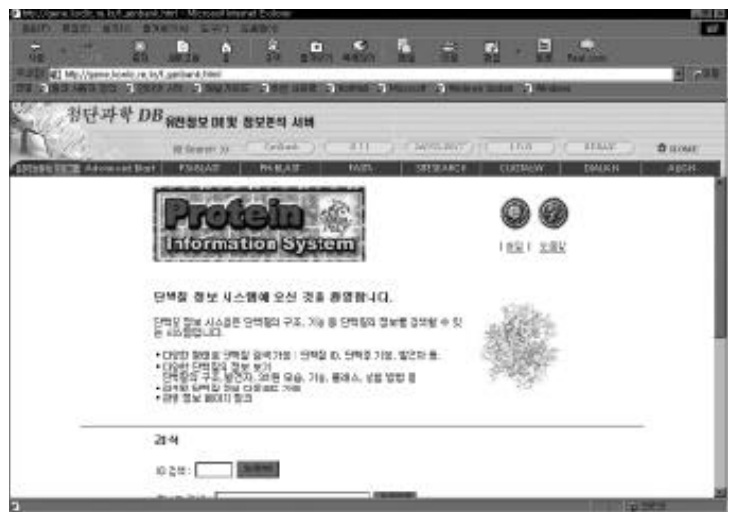


REBASE  
가  
REBASE  
38  
, 15  
, 5-(b).  
4  
-  
-  
-  
web-link  
web-link  
4  
5-(a). 4  
web-link  
CGI  
web CGI



(a) PDB

2.  
1>  
가  
가  
6>



(b) PDB

1) BLAST  
FASTA

< 4 > DB

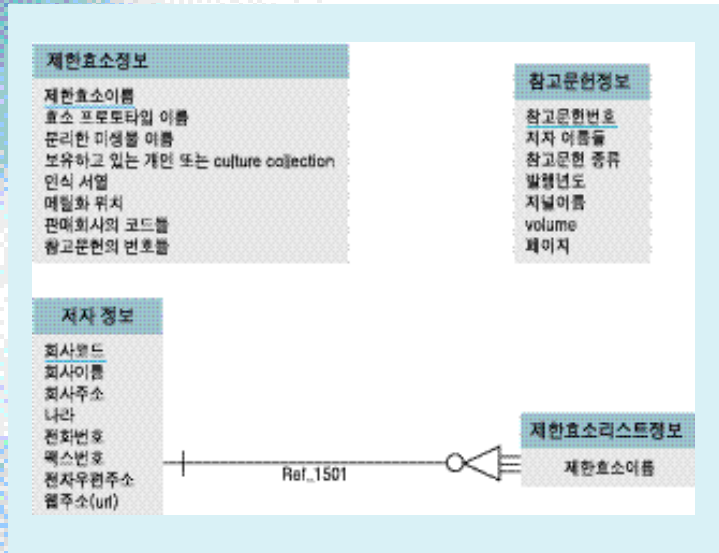
# 과학기술종합정보시스템

1990 가  
 homology local  
 homology ,  
 translation  
 score 가

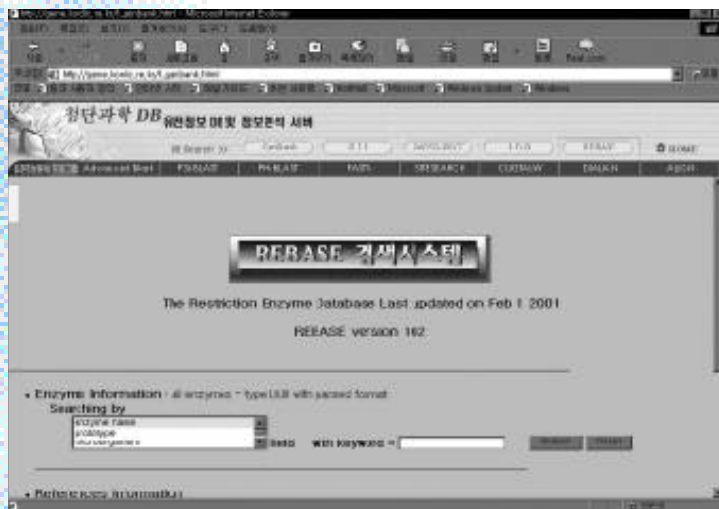
GenBank PIR  
 binary format  
 가 ,  
 conversion  
 GenBank NCBI 2 GenBank  
 format update , PIR  
 Internet update  
 conversion  
 BLAST . BLAST

BLAST2.0 BLAST  
 BLAST PHI-BLAST, PSI-  
 가 .  
 Advanced BLAST  
 , NCBI GenBank

2) PSI-BLAST  
 PSI-BLAST(Position Specific Iterated  
 BLAST) local 가



(a) REBASE



(b) REBASE

< 5> DB

. blast  
 Query  
 Scoring matrix ,  
 Weak  
 homology  
 Advanced BLAST

3) PHI-BLAST  
 PHI-BLAST (Patter Hit Initiated  
 BLAST) local 가





DB

Advanced BLAST

4) FASTA  
Lipman  
homology

GenBank PIR

homology

FASTP

FASTN

global

hology alignment

format FASTA

format , GenBank

PIR format

Internet bioinformatics  
server

FASTA33

< 6>

FASTA33

BLAST

가

. blast

motif 가

fasta33, fastf33, fasts33, fastx33, fasty33,  
tfasta33, tfastf33, tfasts33, tfastx33, tfasty33

local

Motif

EBI

PHI-BLAST

. Motif

Prosite

Motif

AdvancedBLAST

5) site  
site search

DNA

REBASE

8) Align

Align 2

(DNA

global

(restriction site)

automata

site search

)

gap

GOP

(Gap Opening Penalty)

GEP (Gap Extension

Penalty)

L

gap

gap penalty

$GOP + GEP * (L-1)$

6) ClustalW

III.

multialignment

ClustalV

가

ClustalW

가

Windows

ClustalX

PC

, UNIX

porting

EMBL

ClustalW

web

- pairwise/multialignment

- Phylogenetic analysis

- phylogenetic tree

가

가  
(functional  
genomics)

(functional

7) DIALIGN

ClustalW

가

가

multialignment

가

multialignment

multialignment

가

가



가



- [1] Altschul, S.F. et. al., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410.
- [2] Altschul, S.F., et. al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389-3402.
- [3] Higgins, D.G., A.J. Bleasby, and R. Fuchs, 1992. CLUSTAL V: improved software for multiple sequence alignment. *Comp. Appl. Biosci.* 8, 189-191
- [4] Karlin, S., and S.F. Altschul, 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. USA* 87, 2264-2268.
- [5] S. Karlin, S., A. Dembo, and T. Kawabata, 1990. Statistical composition of high-scoring segments from molecular sequences. *The Annals. of Statistics* 18, 571-581.
- [6] Karlin, S., and S.F. Altschul, 1993. Applications and statistics for multiple high-scoring segments in molecular sequences. *Proc. Natl. Acad. Sci. USA* 90, 5873-5877.
- [7] Lipman, D. J., and W.R. Pearson, 1985. Rapid and sensitive protein similarity searches. *Science* 227, 1435-1441.
- [8] Needleman, S.B., and C.D. Wunsch, 1970. A general method applicable to the search for similarities in the amino acid sequences of two proteins. *J. Mol. Biol.* 48, 443-453.
- [9] Pearson, W.R., and D.J. Lipman, 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* 85, 2444-2448.
- [10] Smith, T.F., and M.S. Waterman, 1981. Identification of common molecular subsequences. *J. Mol. Biol.* 147, 195-197.
- [11] Tompson, J.D., D.G. Higgins, and T.J. Gibson, 1994. Improved sensitivity of profile searches through the use of sequence weights and gap excision. *Comput. Applic. Biosci.* 10, 19-29.
- [12] Wilbur, W.J., and D.J. Lipman, 1983. Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA* 80, 726-730.

