NAME

qanalys – application program package for sequence analysis

SYNOPSIS

qanalys exit

DESCRIPTION

Qanalys is a application program package for DNA/protein sequence analysis maintained by Dr T.Miyata and co-workers at the Kyushu University in Japan. This package has been translated to unix system and is maintained by Hidenori Hayashida at the National Institute of Genetics in Japan. At present, the following programs are available. To use these programs, type

```
% qanalys
```

and to exit, type

```
% exit
```

APPLICATION PROGRAMS

antiseq < seqfile
– production of antistrand sequences

acont < seqfile
– calculation of amino acid contents

bcont < seqfile
– calculation of nucleotide contents

codonusg < seqfile
– calculation of codon usage

dnatoaa [-myf] < seqfile
– translation from DNA to amino acids by some code table

hmatrix seqfile [metafile]
– homology matrix for DNA or amino acid sequences

hypath seqfile [metafile]
– hydropathy profiles for amino acid sequences

qformat < seqfile
– general formatter for QANALYS

seqcut < seqfile
– cut out and joint of sequences

seqlist < seqfile
– display of nucleotide sequences

SEE ALSO

gredit(1)

AUTHERS

Programmed by
Dr. Miyata and co-workers
Department of Biology, Faculty of Science Kyushu University Fukuoka 812, Japan
Maintained by
Hidenori Hayashida (hhayashi@niguts.nig.junet@relay.cs.net)
Laboratory of Genetic Information Analysis
Center for genetic Information Research
National Institute of Genetics
Mishima, Shizuoka 411
Japan

BUGS

NAME

acont – calculation of amino acid content

SYNOPSIS

acont

DESCRIPTION

This program calculates the contants and frequences of inputted protein sequences.

This program is included the QANALYS, which is a program package for sequence analysis developed by Dr. Miyata and co-workers at Kyushu University.

acont reads sequences reformed to QANALYS format from the standard input and writes on the standard output. Start and stop positions of calculation can be given (default; hole sequence) and lower and upper case characters are allowed in the protein sequences.
The characters of "-" in the output list mean non amino acid characters of original sequences.

INPUT FILE FORMAT

```
number of the sequences  (I5 : right fixed integer per 5 columns in record)
title line (1 record and < 81 characters)
sequence length [, start, stop ]  ((3)I5)
protein sequence  (60 characters per record)
  .
  .
  .
next title line
next sequence length . . .
  .
  .
  .
```

SEE ALSO

qformat(1)

AUTHERS

Programmed in 1979

Dr. Miyata and co-workers
Department of Biology, Faculty of Science
Kyushu University
Fukuoka 812, Japan

BUGS

sequence length must be shorter than 100000.

NAME

antiseq – production of antistrand sequences

SYNOPSIS

antiseq

DESCRIPTION

This program produces the antistrand DNA sequences against the inputted DNA (RNA) sequences.

This program is included the QANALYS, which is a program package for sequence analysis developed by Dr. Miyata and co-workers at Kyushu University.

antiseq reads sequences reformed to QANALYS format from the standard input and writes on the standard output. Lower and upper case characters are allowed in the DNA (RNA) sequences. The characters of "*" in DNA sequences are judged as insertion to DNA seqence by this program and are not transformed.
Antistrand sequences from output are written by lower case character of DNA. The characters of "-" in the antistrand sequences mean non-nucleotide character of original sequences.

INPUT FILE FORMAT

number of the sequences (I5 : right fixed integer per 5 columns in record)
title line (1 record and < 73 characters)
sequence length (I5)
either DNA or RNA sequence (60 characters per record)
.
.
.
next title line
next sequence length . . .
.
.
.

SEE ALSO

qformat(1)

AUTHERS

Programmed in 1980

Dr. Miyata and co-workers
Department of Biology, Faculty of Science
Kyushu University
Fukuoka 812, Japan

BUGS

sequence length must be shorter than 100000.

NAME

bcont – calculation of base content

SYNOPSIS

bcont

DESCRIPTION

This program calculates the contants and frequences of inputted DNA or RNA sequences.

This program is included the QANALYS, which is a program package for sequence analysis developed by Dr. Miyata and co-workers at Kyushu University.

bcont reads sequences reformed to QANALYS format from the standard input and writes on the standard output. Start and stop positions of calculation can be given (default; hole sequence) and lower and upper case characters are allowed in the DNA or RNA sequences.
The characters of "-" in the output list mean non-nucleotide characters of original sequences.

INPUT FILE FORMAT

number of the sequences (I5 : right fixed integer per 5 columns in record)
title line (1 record and < 81 characters)
sequence length [, start, stop] ((3)I5)
either DNA or RNA sequence (60 characters per record)
.
.
.
next title line
next sequence length . . .
.
.
.

SEE ALSO

qformat(1)

AUTHERS

Programmed in 1979

Dr. Miyata and co-workers
Department of Biology, Faculty of Science
Kyushu University
Fukuoka 812, Japan

BUGS

sequence length must be shorter than 100000.

NAME

codonusg – calculation of codon usage

SYNOPSIS

codonusg

DESCRIPTION

This program calculates the codon usage per coding segments (exon) and total coding regions.

This program is included the QANALYS, which is a program package for sequence analysis developed by Dr. Miyata and co-workers at Kyushu University.

codonusg reads sequences reformed to QANALYS format from the standard input and writes on the standard output. The output image of this program is 135 characters per record.

INPUT FILE FORMAT

```
number of the sequences    (I5 : right fixed integer per 5 columns in record)
"    0"  (this record needed for the compatibility to the other programs; four blanks and "0")
number of segments (exon)    (I5)
start and stop positions of segment    (2I5)
  .
  .
"    1"  (this record needed for the compatibility to the other programs; four blanks and "1")
title line    (1 record and < 81 characters)
sequence length  (I5)
either DNA or RNA sequence    (60 characters per record)
  .
  .
  .
"    0"
number of segments
start and stop
  .
  .
"    1"
next title line
next sequence length . . .
  .
  .
  .
```

SEE ALSO

qformat(1)

AUTHERS

Programmed in 1979

Dr. Miyata and co-workers
Department of Biology, Faculty of Science
Kyushu University
Fukuoka 812, Japan

BUGS

sequence length must be shorter than 100000.

NAME

dnatoaa – translation from DNA to amino acids

SYNOPSIS

dnatoaa [options]

DESCRIPTION

This program translates from DNA or RNA to amino acids by selected codon table.

This program is included the QANALYS, which is a program package for sequence analysis developed by Dr. Miyata and co-workers at Kyushu University.

dnatoaa reads sequences reformed to QANALYS format from the standard input and writes on the standard output. Lower and upper case characters are allowed in the DNA sequences. The characters of "*" in DNA sequences are judged as insertion to DNA seqence by this program.
Amino acid sequences from output are written by upper case one letter code of amino acids. The characters of "*", "-" and "." in the amino acid sequence mean an amino acid insertion (all DNA positions of a codon are "*".), a codon including some non-nucleotide character(s) and stop codon, respectively.

OPTIONS

It is now possible to specify the codon table on the command line. No option means the translation by using the universal codon table. The command line options are preceeded by a dash; the following options are available:

–f	Fungi mitochondria codon table
–m	Mammals mitochondria codon table
–y	Yeast mitochondria codon table; codons "CTX" are translated to threonine.

INPUT FILE FORMAT

Sequence length record is allowed to contain the positions of start and stop of tlanslation.

number of the sequences (I5 : right fixed integer per 5 columns in record)
title line (1 record)
sequence length [start position] [stop position] (3I5)
DNA sequence (60 characters per record)
.
.
. next title line next sequence length . . .
.
.
.

SEE ALSO

qformat(1)

AUTHERS

Programmed in 1979
Dr. Miyata and co-workers
Department of Biology, Faculty of Science Kyushu University Fukuoka 812, Japan

BUGS

sequence length must be shorter than 100000.

NAME

gredit – graphic editor for GKS *metafiles*

SYNOPSIS

gredit *metafile* [*wktype*]

DESCRIPTION

This program displays the graphics interprted from the GKS *metafile* on a graphic terminal and edits these graphics. GKS is the international standard for graphics by International Standards Organization.

If the GKS *metafile* contains the item type number (identifier of graphic primitive) greater than 30000, gredit interprets this item as a separater of segment, which is a named collection of graphic primitives, and can edit the each unit of segments.

WKTYPE

This parameter is the identifier of graphic terminal; refer to *TEKTRONIX PLOT 10 GKS* manuals for detail. Default is 401400 and this is the identifier of *TEKTRONIX 4014* type graphic terminal.

COMMANDS

gredit reads a command from the keyboard.

no	display the segment numbered *no*
ex	close GKS and exit
pp *x,y*	set the pipot point for rotation and zooming; *x,y* = [0,1]
rt *phi*	rotation of *phi* degree
zm *x,y*	zooming of *x* and *y* times each along *x*- and *y*-axis; *x,y* = [0.1,5.0]; if *y* omitted, then *y* = *x*
tr *x,y*	move to the direction of *x*- or *y*-axis; *x,y* = [-1.0,1.0]
cr	clear the scaling parameters and initialize
sv [*file*]	save graph to *file*; default *file* name is "./GRAFIC.META"; Once a *file* name is setted, followed save command can not change it. So, next save command may be "sv" and the graphic images to save are added to *file* with the item of segment separater.
cm	display or undisplay the command list
sg *no*	read the segment numbered *no*

SEE ALSO

hmatrix(1)

AUTHERS

Programmed in May 27, 1988

Hidenori Hayashida (hhayashi@niguts.nig.junet)
National Institute of Genetics
Mishima, Shizuoka 411, Japan

REFERENCES

PLOT10 GKS user's manual

NAME

hmatrix – homology matrix for DNA or amino acid sequences

SYNOPSIS

hmatrix file [metafile]

DESCRIPTION

This program produces the graphic image metafile of the homology matrix by GKS, which is graphic tools package documented by International Standards Organization (ISO).

This program is included the QANALYS, which is a program package for sequence analysis developed by Dr. Miyata and co-workers at Kyushu University.

hmatrix can calculate some set of comparisons of one sequence versus the other sequences. The algorithm of this program is as follows: A pair of continuous sequence segments are selected from a pair of sequences to compare, respectively. This pair of segments have the same sequence length called *cell length.* Then, the homology between them is calculated and if this homology is higher than a threshold level called *level*, a line is drawn on the square matrix corresponding to their sequence positions. This procedure is repeated against all the conbinations of a pair of segments in a pair of sequences.
In the case of amino acid sequences, the level of homology ($score(i,j)$) is calculated from the parameters ($d(i,j)$) of physicochemical difference matrix of amino acids described in Ref.3. For example, the $score(i,j)$ of a pair of amino acid i and j is calculated as follows: $score(i,j) = 100$ for $d(i,j) = 0$ $score(i,j) = 100 \times (1 - d(i,j) / 3.5)$ for a range $0 < d(i,j) < 3.5$; $score(i,j) = 0$ for $d(i,j) = 3.5$; $score(i,j) = 0$ for $d(i,j) > 3.5$;

INPUT FILE FORMAT

Sequence format is QANALYS format. Sequence length record is allowed to contain the positions of start and stop of comparison.

number of the sets of comparison, option (I5,4X,A1)
window size, cell length (2I5)
red level, white level (2I5)
title line of ordinate sequence (1 record)
sequence length [start position] [stop position] (3I5)
either protein or DNA sequence (60 characters per record)
.
.
.

number of the abscissa sequence(s) (I5)
title line of the first abscissa sequence
sequence length [start position] [stop position]
either protein or DNA sequence
sequence length . . .
.
.
.
next title line
.
.

I5 :	right fixed integer per 5 columns in record
option :	if the 10th column of first record is "D" or "d", all of sequences are judged as DNA sequence, otherwise protein.
window size :	zooming parameter; This parameter must be larger than the abscissa sequence length.
cell length :	unit length of compared segment (default = 20).

red and white level : threshold levels of homology; red and white lines are drawn on the color graphic display corresponding to their homology level.

OUTPUT FILE

Default of metafile is ./HOMOLOGY.MAT.
This metafile contains some non-standard records of GKS graphic metafile. These records are used as separater of metafile and recognized as separater of homology matrices by gredit(1).

SEE ALSO

qformat(1),dnatoaa(1),gredit(1)

AUTHERS

Programmed in October 27, 1982
Revised in Sep 12, 1984
Revised in Jul 7, 1987
Hidenori Hayashida. (hhayashi@niguts.nig.junet)
National Institute of Genetics
Mishima, Shizuoka 411, Japan

REFERENCES

1. Toh et al., (1983) Nature 305:827-829.
2. Miyata and Hayashida, (1986) Saibo Kogaku 5:166-174. (in Japanese)
3. Miyata et al., (1979) J. Mol. Evol. 12:219-236.

BUGS

each sequence length must be shorter than 15000.

NAME

hypath – hydropathy profiles for amino acid sequences

SYNOPSIS

hypath file [metafile]

DESCRIPTION

This program produces the graphic image metafile of the hydropathy profiles developed by Kyte and Doolittle (ref.1). Used graphic tools are GKS, which is graphic tools package documented by International Standards Organization (ISO).

This program is included the QANALYS, which is a program package for sequence analysis developed by Dr. Miyata and co-workers at Kyushu University.

hypath calculates the average of hydropathy indices of amino acids from *i* to *i + cell length - 1* for every *i* from *i*=1 to *i*=sequence length + 1 along the amino acid sequence. The hydropathy indices of individual amino acids have been taken from ref.1.

INPUT FILE FORMAT

number of the sequences, *cell length* (2I5)
title line (1 record < 81 characters)
sequence length (I5)
amino acis sequence (60 characters per record)
.
.
.
next title line
sequence length
.
.
.

I5	right fixed integer per 5 columns in record
cell length	unit length of the average of hydropathy indices (default = 9).

OUTPUT FILE

Default of metafile is ./HYDROPATH.META.
This metafile contains some non-standard records of GKS graphic metafile. These records are used as separater of metafile and recognized as separater of hydropathy profiles by gredit(1).

SEE ALSO

qformat(1),dnatoaa(1),gredit(1)

AUTHERS

Programmed in 1982

Dr. Miyata and co-workers
Department of Biology, Faculty of Science
Kyushu University
Fukuoka 812, Japan

REFERENCES

1. Kyte and Doolittle, (1982) J. Mol. Biol. 157:105-132.
2. PLOT10 GKS user's manual

BUGS

sequence length must be shorter than 100000.

NAME

qformat – general formatter for QANALYS

SYNOPSIS

qformat

DESCRIPTION

This program reforms the format of sequences for QANALYS, which is a program package for sequence analysis developed by Dr. Miyata and co-workers at Kyushu University. The output from this program is the general format of the programs included in the QANALYS package.

qformat reads sequences from the standard input and writes reformatted sequences on the standard output.

INPUT FILE FORMAT

This program ignores characters more than 81th column in each record and unprinted characters in the sequence record. The end of sequence is indicated by "//" at the head of record.

title line (1 record)
either protein or DNA sequence
.
.
.
//
next title line
.
.
.

OUTPUT FILE FORMAT

number of the sequences (I5 : right fixed integer per 5 columns in record)
title line (1 record)
sequence length (I5)
either protein or DNA sequence (60 characters per record)
.
.
. next title line next sequence length
.
.
.

SEE ALSO

dnatoaa(1),hmatrix(1)

AUTHERS

Programmed in 1978

Dr. Miyata and co-workers
Department of Biology, Faculty of Science Kyushu University Fukuoka 812, Japan

BUGS

All records must be shorter than 80 characters.
Sequence length must be shorter than 100000.

NAME

seqcut – cut out and joint of sequences

SYNOPSIS

seqcut

DESCRIPTION

This program cuts out some segments from the inputted sequences and joints them.

This program is included the QANALYS, which is a program package for sequence analysis developed by Dr. Miyata and co-workers at Kyushu University.

seqcut reads sequences reformed to QANALYS format from the standard input and writes on the standard output.

INPUT FILE FORMAT

```
number of the sequences   (I5 : right fixed integer per 5 columns in record)
"    0"  (this record needed for the compatibility to the other programs; four blanks and "0")
number of segments to cut  (I5)
start and stop positions of segment   (2I5)
  .
  .
"    1"  (this record needed for the compatibility to the other programs; four blanks and "1")
title line    (1 record and < 81 characters)
sequence length  (I5)
sequence    (60 characters per record)
  .
  .
  .
"    0"
number of segments
start and stop
  .
  .
"    1"
next title line
next sequence length . . .
  .
  .
  .
```

SEE ALSO

qformat(1)

AUTHERS

Programmed in 1980

Dr. Miyata and co-workers
Department of Biology, Faculty of Science
Kyushu University
Fukuoka 812, Japan

BUGS

sequence length must be shorter than 100000.

NAME

seqlist – display of nucleotide sequences

SYNOPSIS

seqlist

DESCRIPTION

This program displays the nucleotide sequences and amino acid sequences translated from nucleotide sequences.

This program is included the QANALYS, which is a program package for sequence analysis developed by Dr. Miyata and co-workers at Kyushu University.

seqlist reads sequences reformed to QANALYS format from the standard input and writes on the standard output.

INPUT FILE FORMAT

```
number of the sequences    (I5 : right fixed integer per 5 columns in record)
"    0"  (this record needed for the compatibility to the other programs; four blanks and "0")
number of segments (exon)    (I5)
start and stop positions of segment    (2I5)
  .
  .
"    1"  (this record needed for the compatibility to the other programs; four blanks and "1")
title line    (1 record and < 81 characters)
sequence length   (I5)
either DNA or RNA sequence     (60 characters per record)
  .
  .
  .
"    0"
number of segments
start and stop
   .
   .
"    1"
next title line
next sequence length . . .
   .
   .
   .
```

SEE ALSO

qformat(1)

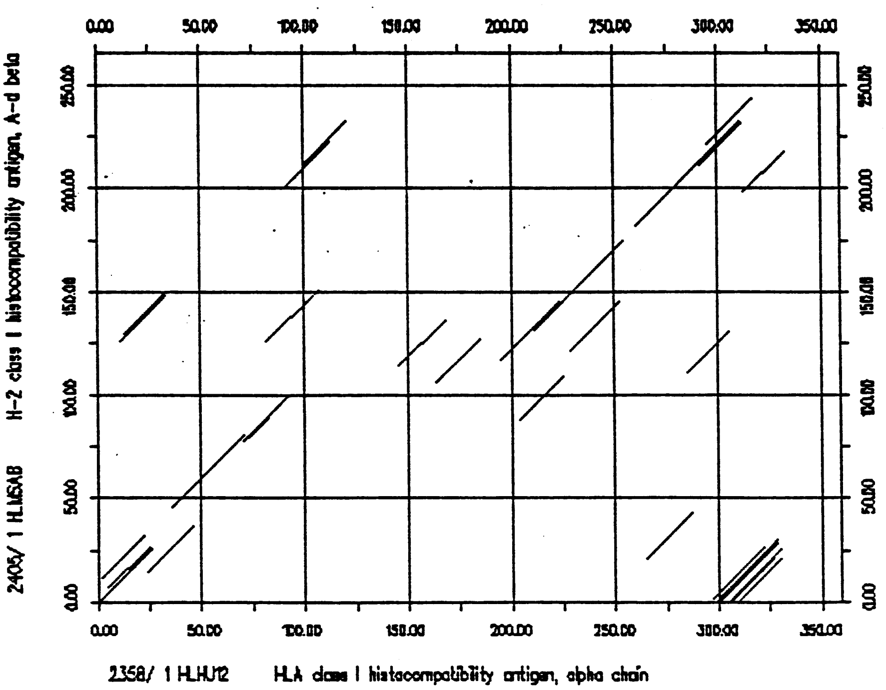
AUTHERS

Programmed in 1979

Dr. Miyata and co-workers
Department of Biology, Faculty of Science
Kyushu University
Fukuoka 812, Japan

BUGS

sequence length must be shorter than 100000.

0.00
50.00
100.00
150.00
200.00
250.00
300.00
350.00
0.00
50.00
100.00
150.00
200.00
250.00
2405/ 1 HLMSAB H-2 class I histocompatibility antigen, A-d beta
2358/ 1 HLHU12 HLA class I histocompatibility antigen, alpha chain