command	in	out	option						operation	comment
txtr	(general explanation)								This program converts text files by various modifications.	
txtr	infile	outfile	1						Delete spaces from the text	
	infile	outfile	2						Delete spaces and isolated numbers from the text.	
	infile	outfile	3						Delete a user-defined character from the text.	
	infile	outfile	4						Uppercase to lowercase.	
	infile	outfile	5						Lowercase to uppercase.	
	infile	outfile	6						Delete empty lines.	
	infile	outfile	7						Copy file.	
	infile	outfile	CR						change LF or CR for the current system.	
	infile	outfile	cr						change LF or CR for the current system.	
	infile	outfile	J						Right-justify lines to a user-defined length.	
	infile	outfile	j						Right-justify lines to a user-defined length.	
	infile	outfile	C						Copy lines from xxx to yyy.	Use cxxx-yyy in command line.
	infile	outfile	c						Copy lines from xxx to yyy.	Use cxxx-yyy in command line.
	infile	outfile	exit						Exit program.	
	help								Show help	
getseq2	infile	outfile	startpos	endpos	n/c				This program extracts a part of sequence from a database entry	+ or − can be used in startpos and endpos. last can be used as endpo
	help								Show help	
toprot	(general explanation)								This program converts a DNA sequence into various formats with translation	
toprot	infile	outfile	1	[codon_table]					a single-frame translation in the fasta file	
	infile	outfile	6	[codon_table]					a 6-frame translation in the fasta file	
	infile	outfile	c	[codon_table]					a composite of 6-frame translation and DNA sequence	
	infile	outfile	n	[codon_table]					a composite of 3-frame translation and DNA sequence	
	infile	outfile	p	[codon_table]					a composite of exon translation and DNA sequence	add exon_file after the p character without space.
	infile	outfile	a	[codon_table]					a list of possible open reading frames	add MinLimit after the a character without space. max can be used to extract the longest ORF. longests followed by number (without space) can be used to extract the longest of overlapping ORFs.
	infile	outfile	s	[codon_table]					a set of possible open reading frames	add MinLimit after the a character without space. max can be used to extract the longest ORF. longests followed by number (without space) can be used to extract the longest of overlapping ORFs.
	help								Show help	

tofast	(general explanation)									This program converts a multiple (catenated) sequence file in the raw (single sequence only), GenBank and EMBL formats into a DNA,RNA or protein sequence file in the complementary or normal sequences in a FASTA file. A CLUSTAL alignment file is also converted to a FASTA file. Also compatible with a catenated or single PIR and SwissProt file.	
tofast	infile	outfile (stdout)	d	[n/c]						to DNA	In [n,c], n: normal strand c: complementary strand.In options d and r, 'n' can be omitted.
	infile	outfile (stdout)	r	[n/c]						to RNA	In [n,c], n: normal strand c: complementary strand. In options d and r, 'n' can be omitted.
	infile	outfile (stdout)	p	[frame]	[codon file]					to protein	codon table file can be specified.
	infile	outfile (stdout)	c							to complementary (no change of data type)	
	infile	outfile (stdout)	I							print sequence information.	
	infile	outfile (stdout)	I	s						print sequence (only)	
	infile	outfile (stdout)	I	l						print sequence length	
	infile	outfile (stdout)	I	n						print sequence name	
	infile	outfile (stdout)	I	#						print number of sequences	
	infile	outfile (stdout)	I	c						print codon usage	
	infile	outfile (stdout)	b							print information about variation in base composition etc.	window size and slide size should be added.
	infile	outfile (stdout)	s							sort multiple sequences according to sequence name.	
	infile	outfile (stdout)	min							select sequences longer than the minimum value.	
	infile	outfile (stdout)	max							select sequences shorter than the maximum value.	
	infile	outfile (stdout)	ms							MS analysis.	name of definition file should be added
	infile	outfile (stdout)								just converts to FASTA <clear mode>	
	help										
seqcat	file1	file2	outfile							This program adds sequences in file2 to the sequences in file1 and writes the combined sequences in outfile.	All file formats that can be read in the siseq package are allowed as a single sequence file. If file1 and file2 are in the CLUSTAL format, the output will be in the CLUSTAL format. Otherwise, output will be in the FASTA format.
	help									Show help	
extcds	infile	outfile	[cds name]							This program extracts CDS from GenBank or EMBL data file. Output is a multiple fasta file.	If you add <cds_name>, only one sequence is extracted. By default, all CDS are extracted.
	infile	outfile_root	t							as above	Table file outfile_root.p.table is also made.
	infile	outfile	exoninf							as above	Detailed information on exons, which were shown. in older versions of SISEQ, is normally omitted. If exoninf option is enabled, exon information is printed as before.
	help									Show help	

										description	notes
cdsnuc	infile	outfile	s <number>	e <number>	[cds name]				This program extracts the nucleic acid sequence for CDS from GenBank or EMBL data file. Output is a multiple fasta file. start and end should be S xx or E xx.	S xx means xx nucleotides from the CDS start position. E xx means xx nucleotides from the CDS end position. <cds_name> is the name of CDS to be extracted (optional). If ADDSEQG is set true, introns are included . Partial truncated sequences are also shown if ADDSEQG is true. Usage: cdsnuc infile outfile split <cds_name>.
	infile	outfile	split	[cds name]					as above	exons are printed individually with xxxxx interleaved.
	help								Show help	
extrna	infile	outfile	s <number>	e <number>	[RNA name]				This program extracts the nucleic acid sequence for RNA from GenBank or EMBL data file. Output is a multiple fasta file. start and end should be S xx or E xx.	S xx means xx nucleotides from the CDS start position. E xx means xx nucleotides from the CDS end position. <RNA_name> is the name of RNA to be extracted (optional). By default, all sequences are extracted.If ADDSEQG is set true, introns are included. Partial truncated sequences are also shown if ADDSEQG is true.
	help									
genlist	infile	[outfile]							This program shows CDS names in GenBank or EMBL file. It also shows information on FASTA sequences.	Default output is the console unless <outfile> is specified.
	infile	outfile	t						table output,	outfile is necessary
	help								Show help	
getent	dbfile	outfile	ID=···.						This program extracts sequences from the database file (dbfile), according to upto 10 keywords, including ID, Ac, OS, and DE.	Wild card (#, not *) can be used at the end of ID and AC. For a multi-FASTA file, only ID can be used.
	dbfile	outfile	AC=···						as above	as above
	dbfile	outfile	OS=···						as above	as above
	dbfile	outfile	DE=···						as above	as above
	dbfile	outfile	uniID=···						Retrieves only the first found entry.	
	dbfile	outfile	NO=xxx,yyy						Retrieves xxx-th entry to yyy-th entry.	xxx and yyy are numbers (separated by a comma
	dbfile	outfile	File=···						Retrieves sequences listed in the file.	
	dbfile	outdir	Group=···						Retrieves sequences listed in the file and save them as Group files in the directory outdir.	
	dbfile	outfile	Split=xx						Splits the sequence file into designated number of files.	xx is a number
	help								Show help	

extint	infile	outfile	s \<number\>	e \<number\>	[cds name]				This program extracts the intron sequences for CDS from GenBank or EMBL data file. Output is a multiple fasta file. start and end should be S xx or E xx.		S xx means xx nucleotides from the intron start position. E xx means xx nucleotides from the intron end position. \<cds_name\> is the name of CDS to be extracted (optional).
	help										
noncod	infile	outfile	0						This program extracts non-coding regions from an annotated file in the GenBank or EMBL format.		introns are not included in the output
	infile	outfile	1						as above		introns are included
	help								Show help		
getclu	infile	outfile	startpos	endpos					This program extracts a part of clustal sequence alignment file. + or - can be used in startpos and endpos. last can be used as endpos.		
	infile	outfile	ph						File conversion to Phylip format		Nexus file cannot be read
	infile	outfile	cl						File conversion to Clustal format		Nexus file cannot be read
	infile	outfile	nex						File conversion to Nexus format		
	infile	outfile	sort						Sequences in a Clustal file are sorted according to sequence name		
	infile	outfile	gap	level/range					sites containing gaps at a level higher than the given level are removed.	range should be min-max without space around hyphen.
	infile	outfile	inf	level/range					sites containing information content at a level between the given range in fraction of maximum value are selected.		range should be min-max without space around hyphen. Information content varies depending on data type and composition. Range should therefore be in relative value.
	help								Show help		
chname	infile	outfile	[old name]	[new name]					This program changes sequence name (ID) of FASTA and CLUSTAL files.		If old_name and new_name are omitted, it opens a dialog for changing name of sequences. Use - for an unspecified old_name.
	help								Show help		
simtbl	infile	outfile							This program calculates similarity score between each pair of sequences in a FASTA or CLUSTAL files. The sequence length of all sequences in a FASTA file should be identical.		
	infile	outfile	list						This program prints a list of identical sequences.		
	infile	outfile	nr						Identical sequences are removed, leaving non-redundant sequences.		
	help								Show help		
splcod	infile								This program splits the three codon positions of a CLUSTAL files. New three files will be created.		
	help								Show help		

									description	notes
nucaln	aa.aln	nuc.fa	out.aln	[editing 1/0]					This program creates a nucleic acid sequence alignment that matches exactly an amino acid sequence alignment. aa.aln should be an amino acid sequence alignment in Clustal format. nuc.fa should be a multi-fasta file of nucleic acid sequences.	Output file is a nucleic acid alignment file in Clustal format. All names of sequences should be identical between aa.aln and nuc.fa. <editing> switch: if 1, editing of C to T is allowed.
	help								Show help	
sites	infile	outfile	0	motif	[matching level]				This program shows location of a sequence motif within a database entry. Output is a short summary only	Input file formats should be: raw, Fasta, GenBank, EMBL, SwissProt, or PRF.
			1	motif	[matching level]				detailed output (default)	
			2	motif	[matching level]				table output	
			3	xx.yy					Automatic search for palindrome sequences.	Min (xx) and max (yy) lengths should be given as motif. These two numbers should be separated by a dot.
	help								Show help	
Internal commands									These are only used within the siseq command. Namely, after 'siseq' or within the siseq.cf file.	
copy	file1	file2							copies <file1> to <file2>	
cp	file1	file2							copies <file1> to <file2>	
remove	file1	file2	…						removes all files <file1> <file2> …	
rm	file1	file2	…						removes all files <file1> <file2> …	
delete	file1	file2	…						removes all files <file1> <file2> …	
del	file1	file2	…						removes all files <file1> <file2> …	
fcat	file1	file2							Add contents of <file2> at the end of <file1>.	
setvar	form	circular							environmental variable FORM is set to circular	
	form	default							use setting as described in the database file	
	printline	<number>							sets line length to the number	
	printline	default							default line length is 75 characters	
	addseqg	true/TRUE/1							If ADDSEQG is set to true, introns are included in the output in the commands, cdsnuc and extrna	
	addseqg	false							Introns are not included in cdsnuc and extrna	
	seq_import	true/TRUE/1							This enables import of external sequences, if accession number is described within the GenBank file	
	seq_import	false							No import of external sequences	
system	<system command>								This executes UNIX system command	