

| | |
|---|----------|
| 1: Introduction | 3 |
| 1.1: Version | 3 |
| 1.2: Copyright, License and Warranty | 3 |
| 1.3: Using this Manual | 3 |
| 1.4: Getting Help | 4 |
| 1.5: Why use (and what is) RJE_SEQ?..... | 4 |
| 1.6: Installation..... | 4 |
| 1.6.1: Files Required for RJE_SEQ..... | 4 |
| 1.6.2: Programs Used by RJE_SEQ..... | 4 |
| 1.6.3: Setting up the INI File | 5 |
| 1.6.4: Reducing Memory Requirements | 5 |

H 5_?Sm?vmFFHF5PSvNTTfmHe5NSTvfPBHc5_mSqNPvfHt5qSPBFbTnHi5_qSmv??FHR5vSNB?THg5_FS?P?q
(5)-2.9084()-2.59001()-1.16765()-2.59005()84260739()-2.71739()-2.71739()-2.71739()-2.71739()-2.71739()-2.71739()-2.71739()
SFBBBNHs5FPSNPBPhn5_FSFNTNmHi5_qSmv??Hi5_qSmv??FHn5vST?FvPvmH.5qNNHEFHn5vST?FvPHy5_
y??Sm?vmHi5_qSmv??FHn5_FSFNTNo15_?Sm?vmfHs5fSTTBbThT F5 PS?ffvH.5_qSTFTfvH.5_qSTFTfvH.5

| | |
|--|-----------|
| 6.2: Appendix II: Distributed Python Modules | 16 |
| 6.3: Appendix III: Log Files..... | 16 |
| 6.4: Appendix IV: Species Codes for IPI & EnSEMBL | 17 |
| 6.5: Appendix V: Troubleshooting | 18 |
| 6.6: Appendix VI: Glossary | 18 |
| 6.7: Appendix VII: References..... | 19 |

1: Introduction

Appendices

Version



1.1: Version

1.2: Copyright, License and Warranty

1.3: Using this Manual

in Courier New

written

written in bold Courier New

red

dark red

italicised Courier New

[]

bold Times New Roman

1.4: Getting Help

```
python rje_seq.py help
```

Output

1.5: Why use (and what is) RJE_SEQ?

1.6: Installation

2.1: Running RJE_SEQ

www.python.org
Python 2.4

1.6.1: Files Required for RJE_SEQ

Python Modules (*.py): `rje`, `rje_blast`, `rje_dismatrix`, `rje_pam`, `rje_seq`, `rje_sequence`, `rje_uniprot`

Additional Files: `jones.pam`

1.6.2: Programs Used by RJE_SEQ

ALIGN:

NB.

BLAST:

CLUSTALW:

MUSCLE:

1.6.3: Setting up the INI File

`rje_seq.py` `rje.ini` `rje_seq.ini`

`blastpath=PATH`

`fastapath=PATH`

`clustalw=PATH`

`muscle=PATH`

`PATH`

Replacing Components with Other Programs

`haquesaq.ini`
`win32=T` `*.ini`

NB. `PATH`

`/`
`path="example path"`

1.6.4: Reducing Memory Requirements

`autoload=T`

`memsaver=T`

`cwcut=X`

`X`

`X`

`X`

`cwcut=1000`

2: Fundamentals

2.1: Running RJE_SEQ

2.1.1: The Basics

Installation

```
python rje_seq.py seqin=FILENAME
```

rje_seq.py

1.6:

i=-1

Note:

```
python rje_seq.py seqin="example with spaces.fas"
```

2.1.2: Interactivity and Verbosity settings

```

                                x
                                i=X
                                v=X
i=-1    v=-1

```

Note:

2.1.3: Other Options

Appendix I: Command-line Options

6.1:

Files

*.ini

6.1.3: INI

```

                                option=X
                                option=FILE
                                option=T/F
option=F    option=F
option option=T    option=True

                                option=X, Y
option=X, Y, ..., Z    ..
option=LIST            option=X    option=A, B, C, D
                                LIST

```

2.2: Input Formats

2.2.1: Sequence formats

seqin=FILE

2.2.3: Fasta format details

fastacmd

fasdb=FILE

formatdb

blastpath=PATH

acclist=LIST

LIST

acclist=acc1,acc2,acc3

3: Sequence Filters

2.2.2: Mapping sequences from one file onto sequence details from another

seqin=FILE

mapseq=FILE

mapseq=FILE

seqin=FILE

mapseq=FILE

seqin=FILE

2.2.3: Fasta format details

```

>
>Seq1 And its description
SEQUENCE-ONE-GOES-HERE
>Seq2
---GAPS--ARE--ALLOWED-
AS-ARE-MULTIPLE-LINES

```

2.2.4: GnSpAcc Format (and the gnspacc=T/F option)

gnspacc=T

```
>ENSP000000223233 blah blah blah
```

```
>ens_HUMAN__ENSP000000223233 blah blah blah
```

gnspacc=F

```
>GN_SP__ACC
```

GN

SP

AC

2.2.5: PAM Matrix

jones.pam
pamfile=FILE

```

A  R  N  D  C  Q  E  G  H  I  L  K  M  F  P  S  T  W  Y  V
Ala 0.98754 0.00030 0.00023 0.00042 0.00011 0.00023 0.00065 ...
Arg 0.00044 0.98974 0.00019 0.00008 0.00022 0.00125 0.00018 ...
Asn 0.00042 0.00023 0.98720 0.00269 0.00007 0.00035 0.00036 ...
...
Val 0.00226 0.00009 0.00007 0.00016 0.00012 0.00008 0.00027 ...

```


2.3: Output

| Format | Description | Extension | Reformat |
|--------|-------------|-----------|----------|
|--------|-------------|-----------|----------|

gnspacc=F

| Format: | Description: | Extension: |
|---------|-----------------------------|----------------------------|
| | Reformat: reformat=X | |
| | 0Sequence Utilities | 3: Sequence Filters |
| | | reformat=X X |
| | | 0Sequence Utilities |

3: Sequence Filters

```

filterout=FILE      filterout=FILE  seqout=FILE
                    format=X          seqout=FILE

```

3.1: *Sequence/Database Feature Filters*

3.1.1: Inherent Sequence Feature Filters

```

minlen=10 maxlen=100      minlen=X    maxlen=X

maxgap=X      maxgap=0.5      maxgap=X
                                gapfilter=F

                                dbonly=T

dblist=X,Y,...,Z      X,Y,...,Z
                                unkspec=F

```

3.1.2: Filters from Lists or Files

```

goodX=LIST      badX=LIST      X

➤
➤
➤
➤
➤

LIST      X,Y,...,Z
goodspec=HUMAN,MOUSE,RAT  baddesc=bad_desc.txt
                        bad_desc.txt

```

3.2: Redundancy Checking

accnr=T

seqnr=T

nrid=X nrsim=X

dblist=X, Y, . . . , Z

specnr=T

3.3: Sequence Filter Options

| Option | Description | Default |
|-------------|-------------------------|---------|
| minlen=X | | |
| maxlen=X | | |
| maxgap=X | | |
| dbonly=T/F | | |
| dblist=LIST | | |
| unspec=T/F | | |
| goodX=LIST | | LIST |
| | X, Y, . . . , Z FILE | |
| | ➤ goodacc | |
| | ➤ goodseq | |
| | ➤ goodspec | |
| | ➤ gooddb | |
| | ➤ gooddesc | |
| badX=LIST | goodX | |
| accnr=T/F | | |
| seqnr=T/F | | |
| specnr=T/F | | |
| nrid=X | | |
| nrsim=X | | |

4: Sequence Utilities

4.1: *Sequence Reformatting/Splitting*

2.3: Output

seqout=FILE

format=X

split=X

X

seqout=mysplit.fas format=fasta split=1000

mysplit.1.fas

mysplit.2.fas

4.2: *Fasta Files from BLAST*

blast2fas=FILE1, FILE2, ..., FILEn

AccNum.blast.fas

AccNum

rje_blast.py

python rje_blast.py help

4.3: *PAM Distance Matrix*

pamdis

5: Module Classes

5.1: *The SeqList Class*

5.2: *The Sequence Class*

5.3: *The DisMatrix Class*

6: Appendices

6.1: Appendix I: Command-line Options

6.1.1: How to Use this Section

```

                                in brackets
                                help
                                __doc__

print rje_seq.__doc__
python rje_seq.py help

```

6.1.2: Option Types

```

                                option=X
                                option=FILE
                                option=T/F
                                option=F    option=F
                                option    option=T    option=True

                                option=X, Y    ..
                                option=X, Y, .., Z    option=X    option=A, B, C, D
                                option=LIST                                LIST

```

6.1.3: INI Files

```

                                ini=FILE
                                hagesac.ini    rje.ini

```

6.1.4: Option Precedence

```

                                rje.ini

                                eg.ini    i=1

rje_seq.py ini=eg.ini i=1
rje_seq.py i=1 ini=eg.ini
i=1
                                eg.ini

```

6.1.5: Command-line Options

| Option | Description | Default | Module |
|--|-------------|---------|----------|
| <u>General Dataset Input/Output</u> | | | |
| seqin=FILE | | | rje_seq |
| fasdb=FILE | | | rje_seq |
| query=X | | | rje_seq |
| basefile=X | | | rje_seq |
| acclist=LIST | | | rje_seq |
| mapseq=FILE | | | rje_seq |
| seqout=FILE | | | rje_seq |
| filterout=FILE | | | rje_seq |
| reformat=X | | | rje_seq |
| gnspacc=T/F | | | rje_seq |
| autoload=T/F | | | rje_seq |
| v=X | | | rje |
| i=X | | | rje |
| d=X | | | haquesac |
| log=FILE | | | rje |
| newlog=T/F | | | rje |
| <u>Sequence Filtering</u> | | | |
| minlen=X | | | rje_seq |
| maxlen=X | | | rje_seq |
| maxgap=X | | | rje_seq |
| dbonly=T/F | | | rje_seq |
| dblist=LIST | | | rje_seq |

| Option | Description | Default | Module |
|---------------------------|----------------------------|---------|------------------|
| unspec=T/F | | | rje_seq |
| goodX=LIST | LIST X,Y,...,Z FILE | | rje_seq |
| | ➤ goodacc | | |
| | ➤ goodseq | | |
| | ➤ goodspec | | |
| | ➤ gooddb | | |
| | ➤ gooddesc | | |
| badX=LIST | goodX | | rje_seq |
| accnr=T/F | | | rje_seq |
| seqnr=T/F | | | rje_seq |
| specnr=T/F | | | rje_seq |
| nrid=X | | | rje_seq |
| nrsim=X | | | rje_seq |
| autofilter=T/F | | | rje_seq |
| <u>System Info</u> | | | |
| blastpath=PATH | | | rje_blast |
| fastapath=PATH | | | rje_seq |
| clustalw=PATH | | | rje_seq |
| muscle=PATH | | | rje_seq |
| win32=T/F | | | rje |
| memsaver=T/F | | | rje |

6.2: Appendix II: Distributed Python Modules

| Module | Description | Classes |
|----------------------------|-------------|---------|
| <code>rje_seq</code> | | |
| <code>rje</code> | | |
| <code>rje_blast</code> | | |
| <code>rje_dismatrix</code> | | |
| <code>rje_pam</code> | | |
| <code>rje_sequence</code> | | |
| <code>rje_uniprot</code> | | |

6.3: Appendix III: Log Files

6.4: Appendix IV: Species Codes for IPI & EnsEMBL

[rje_sequence.py](#)

| Species | Common Name | Species Code |
|---------|-------------|--------------|
|---------|-------------|--------------|

extractDetails()

Sequence

6.5: Appendix V: Troubleshooting

- ## 2.2: Input

6.6: Appendix VI: Glossary

- **ML.**
- **PAM.**
- **NSF.**
- **MSA.**

6.7: Appendix VII: References

215

5

5

73

8

24

132

22