

# FASconCAT

Convenient handling of data matrices  
Kück & Meusemann 2010 (MPE acc.)

Patrick Kück

Forschungsmuseum Koenig, Bonn

## FASconCAT

- concatenate sequence alignment files
- considers. . .
  - standard nucleotide sequences

## FASconCAT

- concatenate sequence alignment files
- considers. . .
  - standard nucleotide sequences
  - recoded nucleotide sequences

## FASconCAT

- concatenate sequence alignment files
- considers. . .
  - standard nucleotide sequences
  - recoded nucleotide sequences
  - amino acid sequences

## FASconCAT

- concatenate sequence alignment files
- considers. . .
  - standard nucleotide sequences
  - recoded nucleotide sequences
  - amino acid sequences
  - dot-bracket structure strings

## FASconCAT

- concatenate sequence alignment files
- considers. . .
  - standard nucleotide sequences
  - recoded nucleotide sequences
  - amino acid sequences
  - dot-bracket structure strings
- missing sequences replaced by. . .
  - 'N'  $\leftrightarrow$  nucleotide sequences

## FASconCAT

- concatenate sequence alignment files
- considers. . .
  - standard nucleotide sequences
  - recoded nucleotide sequences
  - amino acid sequences
  - dot-bracket structure strings
- missing sequences replaced by. . .
  - 'N'  $\leftrightarrow$  nucleotide sequences
  - 'X'  $\leftrightarrow$  amino acid sequences

## FASconCAT

- concatenate sequence alignment files
- considers. . .
  - standard nucleotide sequences
  - recoded nucleotide sequences
  - amino acid sequences
  - dot-bracket structure strings
- missing sequences replaced by. . .
  - 'N'  $\hookrightarrow$  nucleotide sequences
  - 'X'  $\hookrightarrow$  amino acid sequences
  - '.'  $\hookrightarrow$  structure sequences



## FASconCAT

- concatenate sequence alignment files
- considers. . .
  - standard nucleotide sequences
  - recoded nucleotide sequences
  - amino acid sequences
  - dot-bracket structure strings
- missing sequences replaced by. . .
  - 'N'  $\hookrightarrow$  nucleotide sequences
  - 'X'  $\hookrightarrow$  amino acid sequences
  - '.'  $\hookrightarrow$  structure sequences
- concatenation of amino acid AND nucleotide files possible

## FASconCAT

- concatenate sequence alignment files
- considers. . .
  - standard nucleotide sequences
  - recoded nucleotide sequences
  - amino acid sequences
  - dot-bracket structure strings
- missing sequences replaced by. . .
  - 'N'  $\hookrightarrow$  nucleotide sequences
  - 'X'  $\hookrightarrow$  amino acid sequences
  - '.'  $\hookrightarrow$  structure sequences
- concatenation of amino acid AND nucleotide files possible

## FASconCAT

- handles different input formats in one run
- different output formats selectable

## FASconCAT

- handles different input formats in one run
- different output formats selectable
- provides additional information about. . .
  - supermatrix file

## FASconCAT

- handles different input formats in one run
- different output formats selectable
- provides additional information about. . .
  - supermatrix file
  - each single input file

## FASconCAT

- handles different input formats in one run
- different output formats selectable
- provides additional information about. . .
  - supermatrix file
  - each single input file
  - structure sequence

## FASconCAT

- handles different input formats in one run
- different output formats selectable
- provides additional information about. . .
  - supermatrix file
  - each single input file
  - structure sequence
- written on Linux in Perl

## FASconCAT

- handles different input formats in one run
- different output formats selectable
- provides additional information about. . .
  - supermatrix file
  - each single input file
  - structure sequence
- written on Linux in Perl
- works on WindowsPCs, Mac OS and Linux running systems



## FASconCAT

- handles different input formats in one run
- different output formats selectable
- provides additional information about. . .
  - supermatrix file
  - each single input file
  - structure sequence
- written on Linux in Perl
- works on WindowsPCs, Mac OS and Linux running systems
- Perl interpreter has to be extra installed for WindowsPCs

## FASconCAT

- handles different input formats in one run
- different output formats selectable
- provides additional information about. . .
  - supermatrix file
  - each single input file
  - structure sequence
- written on Linux in Perl
- works on WindowsPCs, Mac OS and Linux running systems
- Perl interpreter has to be extra installed for WindowsPCs

## Possible input/output formats

Input format	Ending
FASTA	.fas/.fasta
PHYLIP	.phy
CLUSTAL	.aln
Output files	Content
FcC_smatrix.fas	Supermatrix in FASTA format
FcC_smatrix.phy	Supermatrix in PHYLIP format
FcC_smatrix.nex	Supermatrix in NEXUS format
FcC_info.xls	Concatenation info
FcC_structure.txt	Structure information *

(\* if structure strings are present)

## Prearrangement

- input files must be located in the program folder
- open a terminal window

## Prearrangement

- input files must be located in the program folder
- open a terminal window
- navigate to the FASconCAT located folder

## Prearrangement

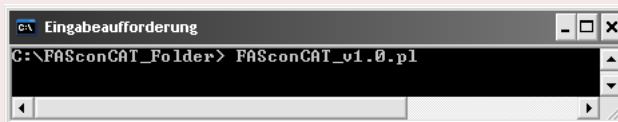
- input files must be located in the program folder
- open a terminal window
- navigate to the FASconCAT located folder
- open the FASconCAT menu

## Prearrangement

- input files must be located in the program folder
- open a terminal window
- navigate to the FASconCAT located folder
- open the FASconCAT menu

## Open the menu (Windows)

- FASconCAT\_v1.0.pl <enter>





## Open the menu (Linux/Mac)

- `perl FASconCAT_v1.0.pl <enter>`



A terminal window titled `~/FASconCAT_Folder` with standard window controls. The prompt is `Patrick@PREACHER ~/FASconCAT_Folder`. The command `$ perl FASconCAT_v1.0.pl` has been entered and is highlighted in yellow.

## FASconCAT menu

```
~#perlFASconCAT_tk/2.1
Patrick@PREACHER ~/Perl/FASconCAT_tk/2.1
$ perl FASconCAT_v1.0.pl

-----
Welcome to FASconCAT v1.0 !
A perlscript for sequence concatenation
written by Patrick Kueck (ZFMK Bonn, 2010)
-----

START  FASconCAT      :                type <a> <enter>
INPUT  ALL/SINGLE     :                type <f> <enter>
INFO   ALL/SMALL      :                type <l> <enter>
NEXUS  BLOCK/MrBAYES :                type <n> <enter>
PHYLIP NO/YES         :                type <p> <enter>
HELP   FASconCAT      :                type <h> <enter>
QUIT   FASconCAT      :                type <q> <enter>
PREFACE FASconCAT      :                type <a> <enter>
-----

FASTA/PHYLIP-INPUT
Concatenate ALL files : YES
Concatenate SINGLE files : NO

OUTPUT
Supernmatrix + ALL info : NO
Supernmatrix : YES
NEXUS-Block : NO
PHYLIP : NO
-----

COMMAND: -
```

- configure a specified setup or use the defaults
- type 's <enter>' to start FASconCAT

## FASconCAT menu

```

~#./PerUFASconCAT_tk/2.1
Patrick@PREACHER ~/Perl/FASconCAT_tk/2.1
$ perl FASconCAT_v1.0.pl

-----
Welcome to FASconCAT v1.0 !
A perlscript for sequence concatenation
written by Patrick Kueck (ZFMK Bonn, 2010)
-----

START  FASconCAT      :          type <a> <enter>
INPUT  ALL/SINGLE      :          type <f> <enter>
INFO   ALL/SMALL      :          type <l> <enter>
NEXUS  BLOCK/MrBAYES  :          type <n> <enter>
PHYLIP NO/YES         :          type <p> <enter>
HELP   FASconCAT      :          type <h> <enter>
QUIT   FASconCAT      :          type <q> <enter>
PREFACE FASconCAT      :          type <a> <enter>
-----

FASTA/PHYLIP-INPUT
Concatenate ALL files : YES
Concatenate SINGLE files : NO

OUTPUT
Supernmatrix + ALL info : NO
Supernmatrix           : YES
NEXUS-Block           : NO
PHYLIP                 : NO
-----

COMMAND:  -

```

- configure a specified setup or use the defaults
- type 's <enter>' to start FASconCAT

## Prearrangement & examples

- navigate to the FASconCAT & input file located folder
- type the program name, followed by a blank and the demand options with a minus (-) sign in front of each

## Prearrangement & examples

- navigate to the FASconCAT & input file located folder
- type the program name, followed by a blank and the demand options with a minus (-) sign in front of each
- press <enter>, e.g.:

## Prearrangement & examples

- navigate to the FASconCAT & input file located folder
- type the program name, followed by a blank and the demand options with a minus (-) sign in front of each
- press <enter>, e.g.:
  - 'FASconCAT\_v1.0.pl -s <enter>' (default start)

## Prearrangement & examples

- navigate to the FASconCAT & input file located folder
- type the program name, followed by a blank and the demand options with a minus (-) sign in front of each
- press <enter>, e.g.:
  - 'FASconCAT\_v1.0.pl -s <enter>' (default start)
  - 'FASconCAT\_v1.0.pl -help <enter>' (help menu)

## Prearrangement & examples

- navigate to the FASconCAT & input file located folder
- type the program name, followed by a blank and the demand options with a minus (-) sign in front of each
- press <enter>, e.g.:
  - 'FASconCAT\_v1.0.pl -s <enter>' (default start)
  - 'FASconCAT\_v1.0.pl -help <enter>' (help menu)



## Options via command line

### General options

Help menu

-help

Preface

-a

Start FASconCAT

-s

### Parameter options

### Command

### Default

Defined input files

-f

none

Dispense all infos

-i

none

PHYLIP output (strict)

-p

none

PHYLIP output (relaxed)

-p -p

none

NEXUS output (blank)

-n

none

NEXUS output (MrBayes)

-n -n

none

## specified concatenation & file converting

- user defined input files
- single files convert to a specified output format

## specified concatenation & file converting

- user defined input files
- single files convert to a specified output format
- input files are chosen via listnumber

## specified concatenation & file converting

- user defined input files
- single files convert to a specified output format
- input files are chosen via listnumber

Listnumber	Filename
1	example_file_1.fas
2	example_file_2.phy
3	example_file_3.aln
4	example_file_4.aln
5	example_file_5.aln

## specified concatenation & file converting

- user defined input files
- single files convert to a specified output format
- input files are chosen via listnumber

Listnumber	Filename
1	example_file_1.fas
2	example_file_2.phy
3	example_file_3.aln
4	example_file_4.aln
5	example_file_5.aln

## List of default & additional info

- provides additional information about. . .
  - each single input file

## List of default & additional info

- provides additional information about. . .
  - each single input file
  - supermatrix file

## List of default & additional info

- provides additional information about. . .
  - each single input file
  - supermatrix file
  - structure string



## List of default & additional info

- provides additional information about. . .
  - each single input file
  - supermatrix file
  - structure string
- additional computation time data depending

## List of default & additional info

- provides additional information about. . .
  - each single input file
  - supermatrix file
  - structure string
- additional computation time data depending
- not included as default setting

## List of default & additional info

- provides additional information about. . .
  - each single input file
  - supermatrix file
  - structure string
- additional computation time data depending
- not included as default setting

## List of default & additional info

Default information	Supermatrix file	input files
Single fragment ranges	yes	no
Number of concatenated sequences per taxon	yes	no
<b>Additional information</b>		
Number of taxa	yes	yes
Number of sequence characters	yes	yes
Data type (nucleotide/amino acid)	yes	yes
Number of single nucleotide characters	yes	yes
Number of gaps	yes	yes
Number of ambiguity characters	yes	yes
Number of inserted replacement characters	yes	yes
Number of missing taxa per fragment	no	yes
Number of inserted replacement strings	yes	yes
Number of characters in total	yes	no
Number of amino acid characters	no	yes
Percent & total number of nucleotides	yes	no
Percent & total number of gaps	yes	no
Percent & total number of ambiguities	yes	no
Percent & total number of inserted replacements	yes	no
Percent & total number of loop characters	yes	yes
Percent & total number of stem characters	yes	yes
Percent & total number of missing data (?)	yes	yes
List of loop positions	yes	no
List of stem pairing positions	yes	no

## -n option

- additional NEXUS file will be generated
- directly loadable into PAUP, MrBayes or other NEXUS file using programs

## -n option

- additional NEXUS file will be generated
- directly loadable into PAUP, MrBayes or other NEXUS file using programs

## -n -n option

- executable file for MrBayes

### -n option

- additional NEXUS file will be generated
- directly loadable into PAUP, MrBayes or other NEXUS file using programs

### -n -n option

- executable file for MrBayes
- integrated presetting as start point for Bayesian analyses

## -n option

- additional NEXUS file will be generated
- directly loadable into PAUP, MrBayes or other NEXUS file using programs

## -n -n option

- executable file for MrBayes
- integrated presetting as start point for Bayesian analyses
- compiles a partition set for MrBayes with separate charset for stem and loop regions



## -n option

- additional NEXUS file will be generated
- directly loadable into PAUP, MrBayes or other NEXUS file using programs

## -n -n option

- executable file for MrBayes
- integrated presetting as start point for Bayesian analyses
- compiles a partition set for MrBayes with separate charset for stem and loop regions
- can be changed manually by any text editor

## -n option

- additional NEXUS file will be generated
- directly loadable into PAUP, MrBayes or other NEXUS file using programs

## -n -n option

- executable file for MrBayes
- integrated presetting as start point for Bayesian analyses
- compiles a partition set for MrBayes with separate charset for stem and loop regions
- can be changed manually by any text editor

## MrBayes setup parameters

MrBayes commands	Setup
Number of generations	2000
Print frequency	100
Sample frequency	100
Number of chains	4
Save branch lengths	yes
Set autoclose	yes
No warnings	yes
Unlink statefrequency	all
tratio	all
Shape	all
Number of substitution	6
Rates	gamma
Sump burnin	20
Number of sump runs	2
Sumt burnin	20
Number of sumt runs	2
Inputfilename	FcC_smatrix.nex
<b>Structure partition</b>	
Set partition	looms
partition looms	2: loops, stems
lset 1	nucmodel= 4by4
lset 2	nucmodel= dublet

## Computation time & costs

- computation time in acceptable manner
- even for phylogenetic purposes

## Computation time & costs

- computation time in acceptable manner
- even for phylogenetic purposes
- computation time depends on...

## Computation time & costs

- computation time in acceptable manner
- even for phylogenetic purposes
- computation time depends on...
  - the amount of data

## Computation time & costs

- computation time in acceptable manner
- even for phylogenetic purposes
- computation time depends on...
  - the amount of data
  - chosen options

## Computation time & costs

- computation time in acceptable manner
- even for phylogenetic purposes
- computation time depends on...
  - the amount of data
  - chosen options
- NEXUS files are most time consuming



## Computation time & costs

- computation time in acceptable manner
- even for phylogenetic purposes
- computation time depends on...
  - the amount of data
  - chosen options
- NEXUS files are most time consuming
- Additional info little more time consuming

## Computation time & costs

- computation time in acceptable manner
- even for phylogenetic purposes
- computation time depends on...
  - the amount of data
  - chosen options
- NEXUS files are most time consuming
- Additional info little more time consuming
- PHYLIP files do not prolong computation time

## Computation time & costs

- computation time in acceptable manner
- even for phylogenetic purposes
- computation time depends on...
  - the amount of data
  - chosen options
- NEXUS files are most time consuming
- Additional info little more time consuming
- PHYLIP files do not prolong computation time

## INDELIBLE simulations (26 sequences)

	Distinct concatenation processes						
	10	10	10	10	10	10	10
N data sets	10	10	10	10	10	10	10
Single lengths [bp]	100	500	1,000	10,000	25,000	50,000	100,000
Supermatrix [bp]	1,000	5,000	10,000	100,000	250,000	500,000	1,000,000
Output options	Computation time [sec]						
Default	0.2	0.1	0.1	0.5	1.2	2.4	4.8
PHYLIP	0.1	0.1	0.2	0.6	1.2	2.4	4.9
Default + all info	0.2	0.3	0.5	4	9.7	19.7	40.1
PHYLIP + all info	0.1	0.3	0.5	3.9	9.7	19.8	40.1
NEXUS	0.2	0.4	0.9	16.1	75.8	281.9	1321.6

## INDELIBLE simulations (108 sequences)

	Distinct concatenation processes						
	10	10	10	10	10	10	10
N data sets	100	500	1,000	10,000	25,000	50,000	100,000
Single lengths [bp]	1,000	5,000	10,000	100,000	250,000	500,000	1,000,000
Supermatrix [bp]							
Output options	Computation time [sec]						
Default	0.3	0.4	0.5	2.3	5.5	11.3	21.4
PHYLIP	0.3	0.4	0.5	2.3	5.5	11	21.9
Default + all info	0.5	1.1	1.9	16.6	42.8	89.1	180.5
PHYLIP + all info	0.4	1.1	1.9	16.8	43.2	88.7	156.8
NEXUS	0.5	1.7	3.4	69.3	320.5	1172.5	5583.4

## FASconCAT links

- FASconCAT and manual downloadable under <http://fasconcat.zfmk.de>
- FASconCAT help desk:  
[fasconcat@web.de](mailto:fasconcat@web.de)

## FASconCAT links

- FASconCAT and manual downloadable under <http://fasconcat.zfmk.de>
- FASconCAT help desk:  
[fasconcat@web.de](mailto:fasconcat@web.de)