LecPCR

Why was written and what does "LecPCR"

The 'LecPCR' program has been developed to transform matrices of qualitative traits such as fragment size matrices into tabular binary matrices.

The program was used to transform DNA profiling data (i.e. results of AFLP analyses of bacterial strains) in large (200 rows, 1200 columns) binary matrices for the purpose of subsequent genomic studies using the program 'DistAFLP' (Cf. Mougel *et al*. 2001, 2002, Jarraud *et al*. 2002). In addition, 'LecPCR' can provide ouput files in the ADE-4 format suitable for multivariate analysis methods (Thioulouse *et al*. 1997, Jarraud *et al*. 2002).

Format of input and output files

Input files

The input file consists in a succession of rows corresponding OTUs (Operational Taxonomic Units).
Each row contains:
1) the OTU name on a maximum of 10 characters followed by spaces up to 10 characters,
2) the names of OTU traits separated by spaces. Use sequential not interleaved data.

```
ex:
name1      47 53 63 67 75 91 106 110 114 117 281 294 331 342 357 370 376 394 400
455 501 503 548 602
name2      47 53 67 82 91 97 98 110 112 114 117 119 341 393 400 424 452 465 499 555
559 602
```

Rows are separated by "return". Mind that the number of rows in the input file determines the number of OTUs in the output file: thus, there must be only one "return" between rows and one "return" at the end of file.

Input files must be saved in "text only" format.

Output files

The main output file is the binary matrix provided in PHYLIP sequential format (Felsenstein 1993), with the first line containing the number of OTUs and the total number of characters (i.e. qualitative traits) separated by spaces. The following rows correspond to OTUs with OTU names on 10 characters as in the input file and binary data not separated by spaces. This format is suitable for the calculation of genome divergence with the program 'DistAFLP'.

Three additional output files are provided :
- the binary matrix in ADE-4 format convenient for PCO (filename plus suffix "b")
- the list of OTU names (filename plus extension ".labs")
- the list of trait names(filename plus extension ".labf")

ex: if testfile01 is the outputfile name chosen in the dialog box, the various output files will be : testfile01, testfile01b, testfile01.labs, testfile01.labf, respectively.

Running LecPCR

At opening, a dialog box is open asking for the input file. The input file must be in the current directory, else the directory can be selected using "Preference" in the "Edit" menu.
The dialog box asks for a name for output file.

Example

Input file :   ex: testfile

```
------------------------------------------------------------------------------------------------
name1     47 53 63 67 75 91 106 110 114 117 281 294 331 342 357 370 376 394 400
name2     47 53 67 82 91 97 98 110 112 114 117 119 341 393 400 424 452 465 499
name3     53 63 68 75 91 106 110 114 117 119 281 283 294 331 394 400 424
name4     53 67 82 91 94 97 106 110 112 114 281 294 296 314 336 341 394 400 424
name5     47 53 68 110 112 114 117 119 263 271 274 281 294 296 334 341 375 394
name6     47 53 91 94 97 114 117 119 133 137 281 294 296 334 341 375 394 400
name7     52 53 67 68 79 91 98 110 114 117 134 341 358 376 394 400 405 424 437
name8     47 53 67 85 91 94 106 110 112 114 271 274 283 294 341 394 400 412 424
name9     36 47 53 67 85 91 101 110 117 119 271 274 283 294 341 406 408 424 437
------------------------------------------------------------------------------------------------
```

Output files: ex: testfile01 (PHYLIP format)

```
------------------------------------------------------------------------------------------------
9 55
name1     0101110100010000110110000000010100100011010101011000000000
name2     0101010001010110011111000000000000001000001010100001011111
name3     0001101100010000110111000000111001000000000011000010000
name4     0001010001011100111100000000101110011000000011000010000
name5     0101001000000000011111000111101100101000010010000000000
name6     0101000000011100000111101000101100101000010011000000000
name7     0011011010010010010110010000000000001001001011100011000
name8     0101010000111000111100000011011000001000000011000110000
name9     1101010000110001010011000011011000001000000000011011000
------------------------------------------------------------------------------------------------
```
ex: testfile01b (ADE-4 format)
```
------------------------------------------------------------------------------------------------
```
Cf. Thioulouse *et al.* (1997).
```
------------------------------------------------------------------------------------------------
```
ex: testfile01.labs
```
------------------------------------------------------------------------------------------------
name1
name2
name3
name4
name5
name6
name7
name8
name9
------------------------------------------------------------------------------------------------
```
ex: testfile01.labf
```
36
47
52
53
63
67
```

```
68
75
79
82
85
91
94
97
98
101
106
110
112
114
117
119
133
134
137
263
271
274
281
283
294
296
314
331
334
336
341
342
357
358
370
375
376
393
394
400
405
406
408
412
424
437
452
465
499
```

References

Jarraud S., Mougel C., Thioulouse J., Lina G., Meugnier H., Forey F., Nesme X., Etienne J., Vandenesch F. 2002. Relationships between *Staphylococcus aureus* genetic background, virulence factors, *agr* type (alleles) and human disease type. Infect. Immun. 70:631-641.

Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) version 3.5c, Distributed by the author, Department of genetics, University of Whashington, Seattle, USA.

Mougel C., Teyssier S., d'Angelo C., Groud K., Neyra M., Sidi-Boumedine K., Cloeckaert A., Peloille M., Baucheron S., Chaslus-Dancla E., Jarraud S., Meugnier H., Forey F, Vandenesch F., Lina G., Etienne J, Thioulouse J., Manceau C., Robbe P., Nalin R., Briolay J., Nesme X. 2001. Experimental and theoretical evaluation of typing methods based upon random amplification of genomic restriction fragments (AFLP) for bacterial population genetics. Genet. Sel. Evol. 33 (Suppl. 1):S319-S338.

Mougel C., Thioulouse J., Perrière G., Nesme X. 2002. A mathematical method for determining genome divergence and species delineation using AFLP. Int. J. Syst. Evol. Microbiol. 52:573-586.

Thioulouse, J., Chessel, D., Dolédec, S. & Olivier, J.M. (1997). ADE-4: a multivariate analysis and graphical display software. Statistics and Computing 7:75-83.