Squangle.r

This text documents the functions available in the R package "squangle20080303.r". This package infers phylogenetic quartets using the three group invariant functions known as the "squangles" (http://arxiv.org/abs/0711.3503).

Author: Jeremy Sumner
jsumner@utas.edu.au or jsumner@it.usyd.edu.au

Depends: R package: "tensor" (http://cran.r-project.org/).
Files: "sqg1.txt", "sqg2.txt", "sqg3.txt".

The polynomial form of the degree d=5 squangles is encoded in the files sqg1.txt", "sqg2.txt", "sqg3.txt". Noting that these polynomials are in a basis as specified in Appendix B of http://arxiv.org/abs/0711.3503, reading the .txt files by row, the coefficients of each term occur as every 6th integer and the variables themselves are represented by the interleaving 5 integers with the correspondence:

p_{ijkl} ~ 4^3*(i-1)+4^2*(j-1)+4^1*(k-1)+4^0*(l-1)+1.

For example, the first 6 integers in "sqg1.txt" are (24, 12, 127, 161, 18, 197) and this corresponds to the term 24*p_{0023}*p_{1332}*p_{2200}*p_{0101}*p_{3010}.

read.seq

Description: Uses the R function "scan" to read multiple sets of m-aligned sequences of length N. Assumes the "phylip" format is followed.

Usage: read.seq(seq.file)

Arguments:

seqfile: The file containing multiple sets of m aligned DNA sequences of length N. Format must be that used in the "phylip" program with each sequence appearing as single, uninterrupted character string.
eg.

```
        m    N
SeqA        CACCCGG…
SeqB        CGCCCGC…
…
SeqM        CAAACCC…
```

Value: A character vector.
eg.

```
[1]    "m"
[2]    "N"
[3]    "SeqA"
```

[4] “CACCCGG…”
[5] “SeqB”
…
[1+2M] “SeqM”
[2+2M] “CAAACCC…”

squangle

Description: Computes the value of the three squangle polynomials {Q1,Q2,Q3} evaluated on the data set of a quartet of aligned DNA sequences. It will only work for aligned quartets arranged sequentially. A termination message is given if $m \neq 4$.

Usage: squangle(allseq,quart=c(1,2,3,4))

Arguments:

allseq: A character vector of multiple quartets of aligned sequences as computed using “read.seq”.

quart: Optional: default=c(1,2,3,4)
Specifies which quartet to analyse.

Value: A vector; the numerical values of the squangles.

tree

Description: Returns the maximum likelihood tree for the three possibilities,

tree1=(12,34),
tree2=(13,24),
tree3=(14,23).

For the three possible trees the expectation values of the squangles are known to be,

tree1: E[Q1]=0 and E[Q2]=-E[Q3]>0
tree2: E[Q2]=0 and E[Q1]=E[Q3]>0
tree3: E[Q3]=0 and E[Q1]=E[Q2]<0

Assuming the squangles are independent and normally distributed with identical variance=v, and mean=u>0 as above, the likelihood of each quartet is

L1~nd(Q1,0,v)*nd(Q2,u,v)*nd(Q3,-u,v)
L2~nd(Q1,u,v)*nd(Q2,0,v)*nd(Q3,u,v)
L3~nd(Q1,-u,v)*nd(Q2,-u,v)*nd(Q3,0,v).

Under these conditions the MLE of u is independent of v and corresponds to the least squares estimator. Recalling the constraint u>0, we have

MLE[u|tree1]=max[0, (Q2-Q3)/2]
MLE[u|tree2]=max[0, (Q1+Q3)/2]
MLE[u|tree3]=max[0, -(Q1+Q2)/2].

Note: u=0 corresponds to the star tree.

Usage: tree(squang)

Arguments:

squang: A vector of squangles. ie. c(Q1,Q2,Q3)

Value: Character vector.

write.squang
read.squang

Description: Convenience functions for reading and writing to file a list of multiple squangle vectors.

Usage: write.squang(squangs,file.sqg)
read.squang(file.sqg)

Arguments:
squangs: A list of squangle vectors.
file.sqg: File to write/read to/from.

Values: file.sqg
A list of squangle vectors.