Modeling transition/transversion bias of nucleotide substitution over time

Case study 4

AUTHOR

Stefan Schmutz

AFFILIATION

ZHAW

PUBLISHED

Dec. 15, 2019

TABLE OF CONTENTS

Estimating nucleotide frequencies

> *Estimate nucleotide frequencies from the pairwise alignment of human and mouse cytochrome b gene as given in the file "mt-cyb-human- mouse_cDNAalignment.fasta". Use these values to parameterize the model.*

One way to estimate the nucleotide frequencies is to count them.
Since we're working with a pairwise alignment without indels, the total length of both sequences is the same (1140 nt). The detailed composition is listed in Table 1.

Table 1: Nucleotide frequencies

nucleotide	human	mouse
T	286	327
C	391	312
A	326	361
G	137	140

Estimating transition transversion rate ratio

> *Propose a simple way of estimating transition transversion rate ratio from the dataset and use this estimate for the parameterization of the model.*

Table 2: Nucleotide comparisons

human-mouse	n
a-a	264
a-c	27
a-g	15
a-t	20
c-a	59
c-c	249
c-g	11
c-t	72
g-a	19
g-c	4
g-g	111
g-t	3
t-a	19
t-c	32
t-g	3
t-t	232