

# Definition of the Class "bitsplits"

Andrei-Alin Popescu & Emmanuel Paradis

July 30, 2013

## 1 Overview

The class "bitsplits" is used to model a set of weighted splits in R. The splits themselves are modelled using an array of bytes (mode "raw" in R) in the following way. Suppose we have  $n$  taxa, and that we choose an arbitrary ordering on these taxa. Then we model a split on these taxa by using a bitmask of size  $n$ , where if bit  $i$  is set, taxon  $i$  belongs to one half of the split, and if it is not set it belongs to the other half of the split. Since often we will have more taxa than could be stored in a single byte, we will use a vector of size  $\lceil n/8 \rceil$  of bytes, which is basically a breakdown into "byte-sized chunks" of the entire set of bits which needs to be used. In the event that we have trailing bits which are not used, user code should make no assumptions of the values of these bits.

## 2 Structure

The class has the following fields:

1. **matsplit**: a matrix of mode "raw". This is used to model the actual splits in the set. This matrix has  $\lceil n/8 \rceil$  rows and  $m$  columns. Column  $i$  is a vector which models a split (as described in the overview);
2. **labels**: a vector of mode character of length  $n$  with the taxon labels;
3. **freq**: a numeric vector of length  $m$  of split weights.