

Bird museum data (BMD)

Alexander Florez Rodriguez
Center for Macroecology, Evolution and Climate

4/3/2017

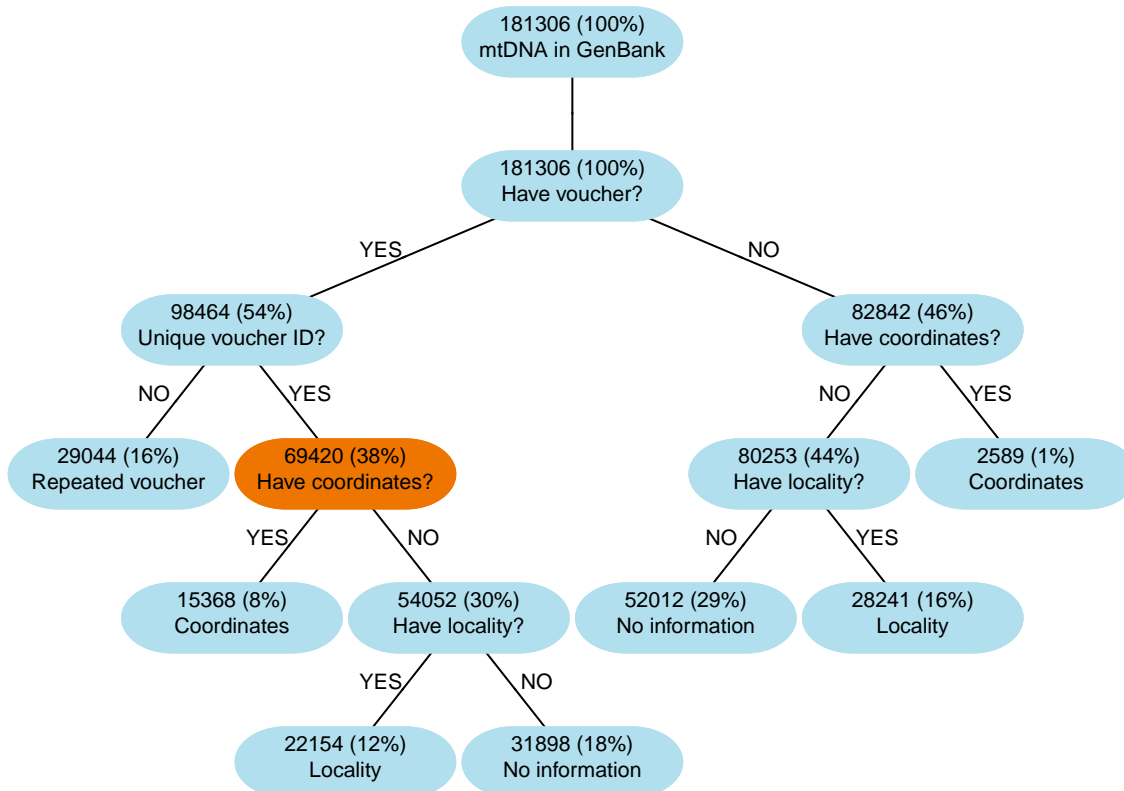
Map of Genetic diversity of Aves

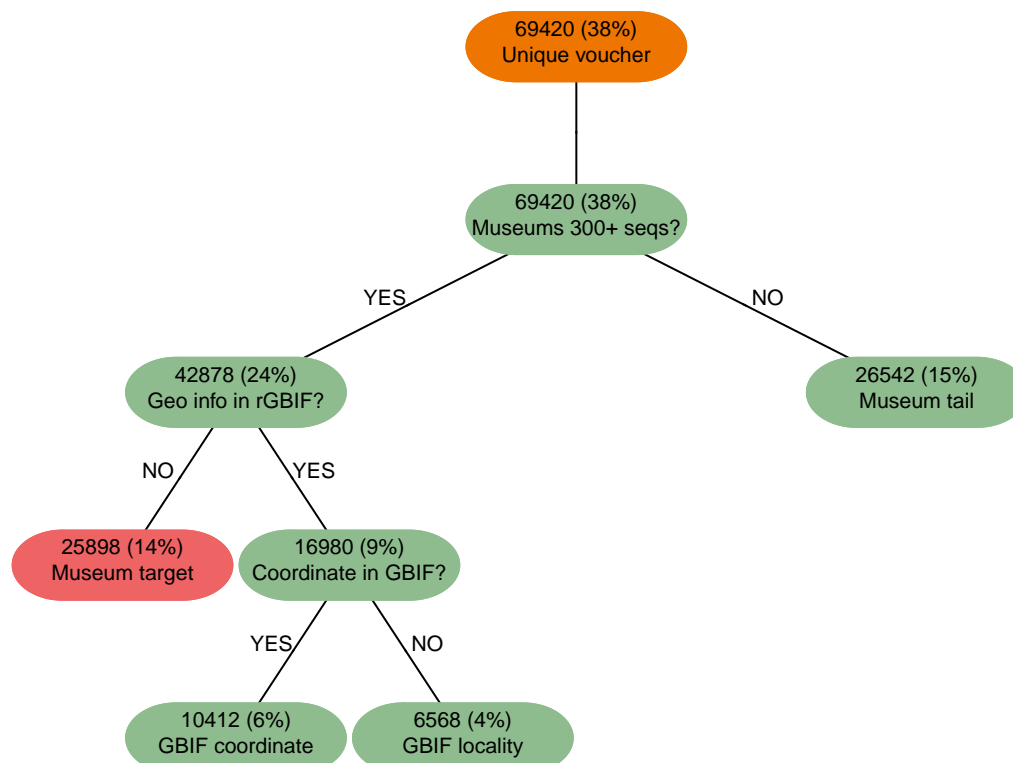
What data we need?

We need georeferenced DNA sequences to map the genetic diversity of any taxa. There are several ways to obtain this data. 1) directly from GenBank (or other genetic repositories), 2) using geonames to estimate the coordinates of the localities reported in GenBank, 3) Use other information in the GenBank records (e.g. voucher number, collection id) as links for cross-referencing with other databases (e.g. museum's and Gbif). Here I describe the dataset of mitochondrial DNA sequences for all Aves available in GenBank and how to obtain the voucher information for the sequences in GenBank.

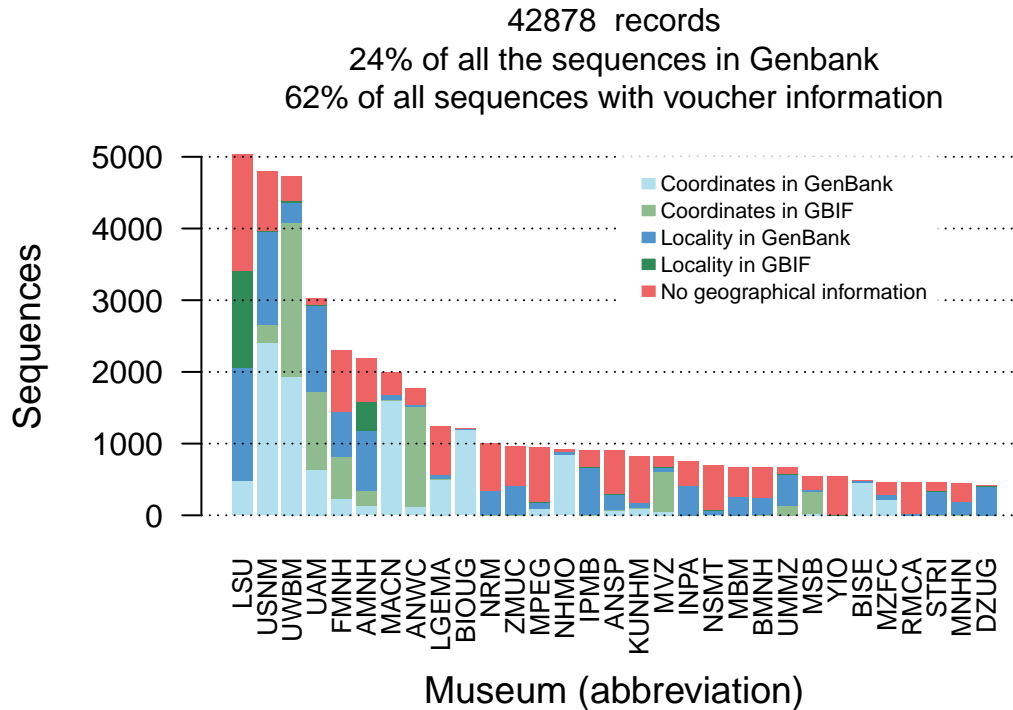
We obtained all mitochondrial sequences for Aves in Genbank, in total there are 181306 sequences. The binary tree below shows how the data is divided by category. Each one of the nodes correspond to a subset of the total dataset. For each node the number of sequences and the percentage relative to the total database is shown. Also, for each node a question (?) or a descriptor is shown (below the numbers). If it is a question, this node will be divided (subseted) in both, YES and NO branches.

Loading required package: shape





Museums with more than 300 sequences in GenBank



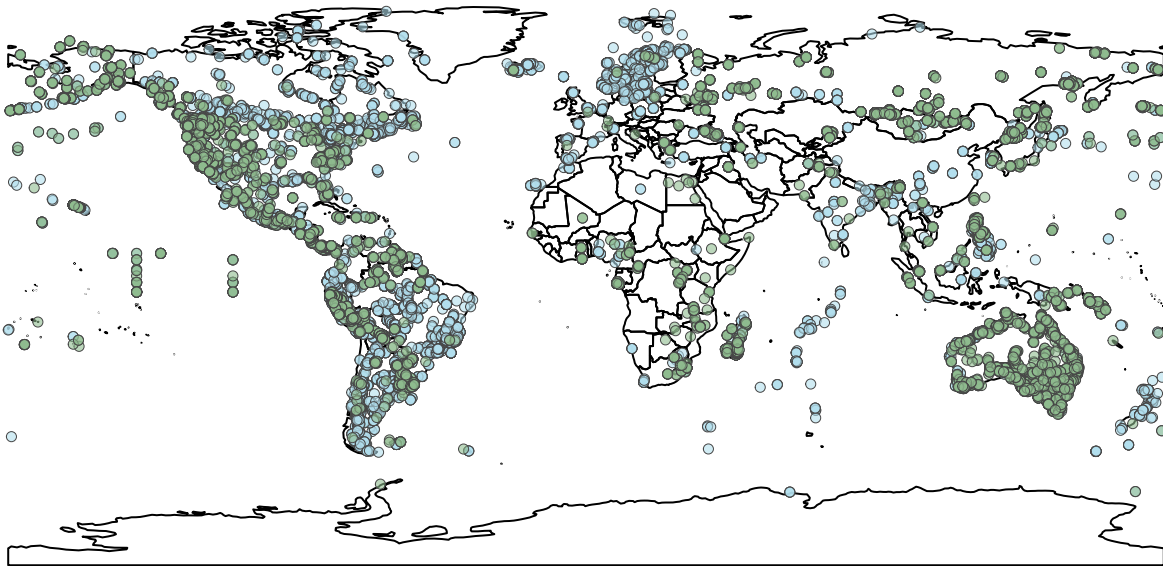
```
## Loading required package: sp
```

```
##
```

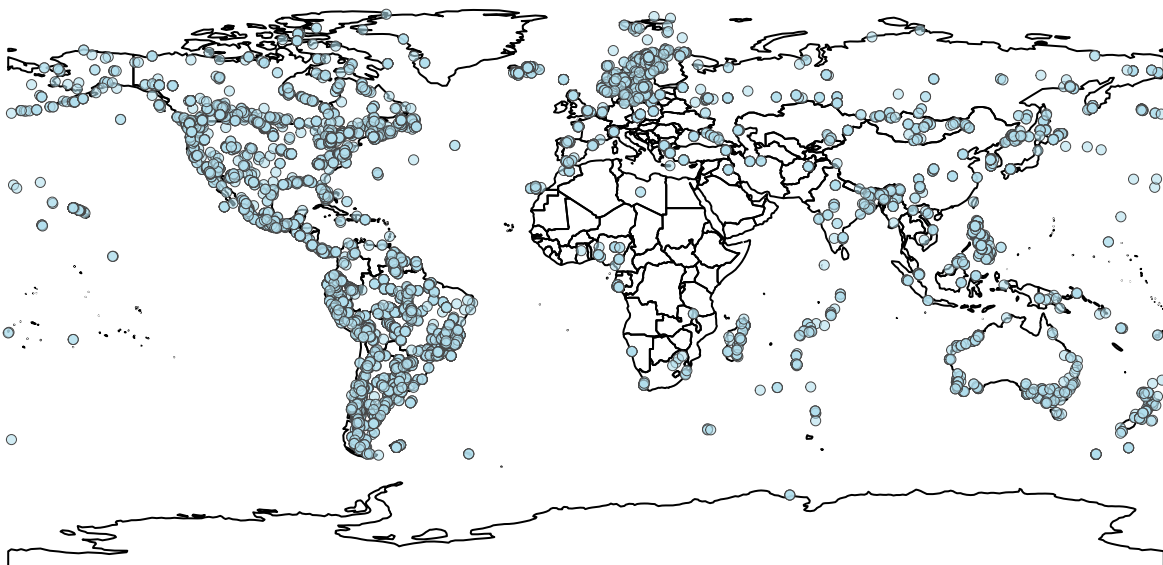
```
## Attaching package: 'sp'
```

```
## The following object is masked from 'package:diagram':
##
##   coordinates
## ### Welcome to rworldmap ###
## For a short introduction type :  vignette('rworldmap')
```

Sequences with coordinates in Genbank and GBIF 21995 sequences



Sequences with coordinates ONLY in Genbank 11597 sequences



Sequences with coordinates ONLY in GBIF
6712 sequences



Sequences with coordinates in BOTH Genbank and GBIF
3693 sequences

