

*total.fasta*

29,233 RABV  
Sequences NCBI



*concat\_seq\_genes.fasta*

19,329 RABV  
Sequences NCBI

---

Less than 200bp in coding region  
Collection date pre 1971  
No date, country\*



*Africa\_aln.fasta*

4,632 Sequences

---

5 sample from non-African clades  
African origin in metadata  
Within African clades  
Canine



*Africa\_RABV.nwk*

4,574 Sequences

---

LSD2 Outlier Removal

\*Sequences from Kissy, 1999 and Talbi, 2010 did not have country data in NCBI, but it was extracted through isolate country code

# Subsampling Ideas and Trial

4574 Sequences

## **Subsampling protocol:**

Remove sequences from countries with the most over representation, removing sequences with shorted branch lengths (waiting for Sam's code to remove monophyletic sequences)

# Results

4000 Sequence Target

Country	Leaves	Equal Dist.	Subsampled
South Africa	947	33.05785	799
Morocco	454	33.05785	454
Algeria	375	33.05785	375
Nigeria	283	33.05785	283
Tanzania	282	33.05785	282
Ethiopia	224	33.05785	224
CAR	215	33.05785	215
Zambia	199	33.05785	199
...			

3000 Sequence Target

Country	Leaves	Equal Dist.	Subsampled
South Africa	947	24.79339	239
Morocco	454	24.79339	239
Algeria	375	24.79339	239
Nigeria	283	24.79339	239
Tanzania	282	24.79339	239
Ethiopia	224	24.79339	224
CAR	215	24.79339	215
Zambia	199	24.79339	199
...			