

# 1 Datasets

## 1.1 MetaPhyler

	Sequenzen	durchschnittliche Basenpaare
Original	73086	300
Rechercheergebnis	40039	644,51
Bazinet	73086	300

### Anmerkungen zu unserer Recherche:

Link: <http://metaphyler.cbc.umd.edu/#download>

MetaPhylerSRV0.115.tar.gz (30MB)

Markers.fna wurde als Dataset verwendet.

## 1.2 PhymmBL

	Sequenzen	durchschnittliche Basenpaare
Original	80215	243
Rechercheergebnis	73252	204,46
Bazinet	80215	243

### Anmerkungen zu unserer Recherche:

Link: [www.cbc.umd.edu/software/phymm/exp\\_datasets/](http://www.cbc.umd.edu/software/phymm/exp_datasets/)

Besteht aus mehreren Textdateien aufgeteilt in 100bp, 200bp, 400bp, 800bp und 1000bp.

## 1.3 RAIphy

	Sequenzen	durchschnittliche Basenpaare
Original	477000	238
Rechercheergebnis	477000	233,33
Bazinet	477000	233,33

### Anmerkungen zu unserer Recherche:

Link: <http://bioinfo.unl.edu/raiphy.php>

Genutzt Simple Test (36,6MB).

Besteht aus drei Fastadateien aufgeteilt in 100bp, 200bp und 400bp.

## 1.4 CARMA

	Sequenzen	durchschnittliche Basenpaare
Original	25000	265
Rechercheergebnis	25000	265

### Anmerkungen zu unserer Recherche:

Link: [http://www.cebitec.uni-bielefeld.de/webcarma.cebitec.uni-bielefeld.de/download/simulated\\_metagenome\\_454\\_265bp.fna](http://www.cebitec.uni-bielefeld.de/webcarma.cebitec.uni-bielefeld.de/download/simulated_metagenome_454_265bp.fna)

## 1.5 PhyloPythia

	Sequenzen	durchschnittliche Basenpaare
Original	124941	961
Rechercheergebnis	114457	969,09
Bazinet:		
SIMLC	97497	951,96
SIMMC	wie Recherche	
SIMHC	116771	949,51

### Anmerkungen zu unserer Recherche:

Link: [http://fames.jgi-psf.org/Retrieve\\_data.html](http://fames.jgi-psf.org/Retrieve_data.html)

SimMC Dataset sequence file runtergeladen.

Besteht aus 113 Fastadateien.

## 1.6 FACS

	Sequenzen	durchschnittliche Basenpaare
Original	100000	269
Rechercheergebnis		
Bazinet	100000	269

### Anmerkungen zu unserer Recherche:

Nicht gefunden.

Email an Autoren geschrieben, warten auf Antwort. -> Antwort erhalten (11.11.2015) mit dem Link [http://molevol.umiacs.umd.edu/labfiles/sequence\\_classification\\_data\\_sets.zip](http://molevol.umiacs.umd.edu/labfiles/sequence_classification_data_sets.zip)

Folgende Datensätze nun vollständig:

- CARMA
- FACS
- MetaPhyler
- PHymmBL

Bei RAIphy ist Anzahl der Sequenzen gleich, durchschnittliche Länge unterscheidet sich.

Bei PhyloPythia stimmen die Längen ungefähr überein. Die Anzahl der Sequenzen jedoch nicht.

-> Mail an Autor geschrieben. Noch keine Antwort.