REPORT

Alignment-free tools for metagenomics-data analysis

Robert Deibel

Abstract

TODO

Keywords: alignment-free; report; metagenome

Introduction

Metagenomics

A puddle of mud The metagenome is the whole set of genes, of a population, of microorganisms as found in a sample of a microbiome. As such metagenomics is the study and analysis of these metagenomes.[1]
A microbiome is the "home" of countless bacteria, archea and viruses; like all microorganisms >90% of those found in microbioms are uncultured, leaving researchers with the problem of how to study those organisms.

Accumulated data from microbiome samples Choosing a sample is the easiest part of the analysis of a microbiome; the following steps are:

1. DNA isolation from samples
2. construction of DNA libraries (typically in *E. Coli* as host)
3. Mining for clones and DNA sequences of interest
4. Accumulation of desired clones and DNA sequences

as stated in Streit et al [2], to obtain a metagenomic library, which is the base of analysis.

NGS – Next Generation Sequencing The sheer amount of data gathered through such samples – Kakirde et al[3] states 10000 Gb of DNA in a soil sample – leaves researches with the problem of sequencing.
While Sanger sequencing is an accurate and proven method for sequencing it is dated for the scale of metagenomics. Nowadays new high throughput methods – also Next Generation Sequencing or NGS for short – are used to handle this problem. NGS is a conglomerate of methods used in bioinformatics for rapid parallelized sequencing, producing thousands or millions of sequences concurrently.

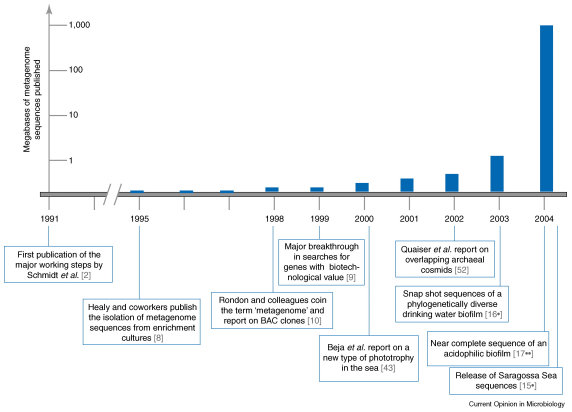

Figure 1 Timescale of metagenomic-derived and published DNA sequences. The timescale ranges from 1991, the initial outline of the major working steps, to the first mapping of archaeal comids in 2002 and the snap shot sequence analysis of the Sargasso Sea published earlier this year.–taken from Streit et al [2]

What do we want to achieve? Researches use the information gained through metagenomic-data analysis to design antibiotics and medicine or to analyze the metabolism of microorganisms and its hosts. Due to the rising number of identified genes using metagenomics-data analysis (Figure 1) and the >90% uncultured microorganisms, metagenomics is a field of vast research.
I want to briefly summarize two approaches of data analysis and showcase one of those in more detail.

The "classical" approach

Alignment-based method

Correspondence: robert.deibel@student.uni-tuebingen.de
Eberhard-Karls Universität, Tübingen, DE
Full list of author information is available at the end of the article

The good

The bad – Too much data, too little time The analysis of such metagenomes is heavy on computation and time resources, due to the amount of data collected; this results in the pursuit of faster and more effective methods for data analysis

The alternative approach
Alignment-free method

The ugly

Methods
Statistics
The power of statistics

k-tupel approach – Song et al
What is a k-tupel

D_2

Nucleotide bias

Visualization approach
The idea behind

non-linear dimension reduction – Laczny et al
Weiss noch nicht hier

Results
Application of tools on data set
hier kommt was hin

Content
Section title
Sub-heading for section
Sub-sub heading for section

Sub-sub-sub heading for section

Competing interests
Author's contributions
Acknowledgements
References

1. Handelsman, J.: Metagenomics: application of genomics to uncultured microorganisms. Microbiology and molecular biology reviews **68**(4), 669–685 (2004)
2. Streit, W.R., Schmitz, R.A.: Metagenomics – the key to the uncultured microbes. Current Opinion in Microbiology **7**(5), 492–498 (2004). doi:10.1016/j.mib.2004.08.002
3. Kakirde, K.S., Parsley, L.C., Liles, M.R.: Size does matter: Application-driven approaches for soil metagenomics. Soil Biology and Biochemistry **42**(11), 1911–1923 (2010). doi:10.1016/j.soilbio.2010.07.021

Figures
Tables
Additional Files
Additional file 2 — Sample additional file title