

About This Summary:

What this summary is about?

This is a summary of CoCoNet tool article which is an efficient deep learning tool for viral metagenome binning.

[link of original article](#)

About original article:

Authors:

[Cédric G Arisdakessian](#), [Olivia D Nigro](#), [Grieg F Steward](#), [Guylaine Poisson](#), [Mahdi Belcaid](#)

Publishing Date:

05 April 2021

Article History:

Received:

18 June 2020

Revision received:

24 March 2021

Editorial decision:

25 March 2021

Accepted:

02 April 2021

Published:

05 April 2021 (**Bioinformatics** journal)

1.Abstract:

Why metagenomics?!

Characterizing microbial communities and unraveling the intricate link between the microbiome and biological processes are two important things that metagenomics approaches provide.

Metagenomics and Assembly:

Assembly is a critical step in metagenomics experiments. It transforms overlapping **DNA sequencing** reads into sufficiently accurate representations of the community's genomes.

Why Composition and Coverage Network (CoCoNet)?!

Process of assembly is computationally difficult and results in fragments across many contigs in genome. There are Computational binning methods are used to mitigate fragmentation, they have been principally tuned for bacterial genomes and do not perform favorably on viral metagenomes. So, this result in Composition and Coverage Network (CoCoNet) deep learning tool, a new binning method for viral metagenomes. CoCoNet substantially outperforms existing binning methods on viral datasets.

2.Introduction:

Importance of metagenomics:

Metagenomics plays important role to know composition and function of microbial communities in different environments from human gut to deep sea.

Metagenomics Assembly challenge:

Short contigs is a challenge in Metagenomic assembly, which leading to poor characterization of the underlying species diversity, richness and functional capacity.

Assembly against reference genome vs Assembly de novo:

Reference genome sequences assembly can be achieved by clustering contigs that align with high-confidence against the same reference genome. Sequences without reference genome assembly is a complicate process since contigs need to be clustered *de novo*.

Binning contigs methods types:

There are three types of existing contig binning methods:

1) Sequence composition methods:

This type of methods depends on distributions of K-Mers. It bins together contigs with similar K-Mer distributions.

Advantage:

It works best on species with sufficiently divergent K-Mer distributions.

Disadvantage:

It is less effective for resolving closely related species for which the K-Mer distributions may be indistinguishable, particularly over short contigs.

2) Coverage correlation methods:

This type uses the sequencing coverage, the number of reads aligning to each contig, and reports contigs that consistently share similar coverage values across multiple samples as belonging to the same genome.

Advantage:

It is accurate.

Disadvantage:

It works best when many samples are available.

3)Combination of sequence correlation and coverage correlation:

This type leverages both K-Mer and coverage profiles.

Advantage:

It combines the advantages of composition- and coverage-based solutions.

Disadvantage:

It is computationally more complex.

CoCoNet new binning method:

It is for binning contigs assembled from viral metagenomic data.

It leverages **deep learning** to model the K-Mer composition (first type of methods) and the coverage (second type of methods) for binning contigs assembled from viral metagenomic data.

It uses a neural network trained using contigs to learn flexible function for predicting the probability that any pair of contigs originated from the same genome.

Performance of CoCoNet:

CoCoNet was on both simulated and experimental viral metagenome data, and results show that CoCoNet outperforms existing tools optimized for binning bacterial data.

[Source Code and Documentation Of CoCoNet tool on git hub](#)

Algorithms:

CoCoNet algorithm:

-It uses **deep learning** in conjunction with **clustering**.

-works in two phases:

- 1) Trains a deep neural network to estimate the probability that two fragments belong to the same genome, given their composition and coverage information.
- 2) Uses a computationally tractable heuristic to bin the contigs using the co-occurrence probabilities inferred in the previous phase.

Other algorithms:

1) CONCOCT's latest version 1.1.0.

2) Metabat2 v2.15.

