Simka: fast kmer-based method for estimating the similarity between numerous metagenomic datasets
Gaëtan Benoit, Pierre Peterlongo, Dominique Lavenier, Claire Lemaitre

To cite this version:
Gaëtan Benoit, Pierre Peterlongo, Dominique Lavenier, Claire Lemaitre. Simka: fast kmer-based method for estimating the similarity between numerous metagenomic datasets. JOBIM 2015, Jul 2015, Clermont-Ferrand, France. , 10.1093/Bioinformatics/btu406. hal-01180603

HAL Id: hal-01180603
https://inria.hal.science/hal-01180603
Submitted on 27 Jul 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Simka: fast kmer-based method for estimating the similarity between numerous metagenomic datasets

Gaëtan Benoit1, Pierre Peterlongo1, Dominique Lavenier2, Claire Lemaître1

1 IRISA/RISA Gwéndale, Campus de Beaulieu, 35042 Rennes cedex, France.
2 UNIV. DE RENNES 1

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity:

Simka was tested and compared to Commet [1] (state of the art) on 21 Tara Oceans samples:
- 3G reads
- 400 GB of data
- 219G 31-mers
- 123G distinct 31-mers

Datasets

Criteria for the method:
- Filtering input reads
- Provide similarity confidence intervals with bootstrap
- Selecting discriminative kmers

Relative similarity:

Absolute Similarity: