Metagenomic Binning Pipelines - the State of the Art

3 Abstract

- 4 New generations of sequencing platforms coupled with numerous bioinformatics tools have led
- 5 to rapid technological progress in metagenomics and metatranscriptomics to investigate complex
- 6 microorganism communities. Nevertheless, a combination of different bioinformatic tools remains
- 7 necessary to draw conclusions out of microbiota studies. As sequencing costs have dropped at
- 8 a rate above 'Moore's law', a greater number of large data sets are being produced than ever
- 9 before. Newer algorithms that take advantage of the size of these datasets are continually being
- developed. Binning algorithms are defined as the grouping of assembled metagenomic contigs by
- their genome of origin. Selecting the most appropriate binning algorithm can be a daunting task
- and is influenced by many factors This review serves as a guide to direct the researcher to the
- binning algorithm that best suits their needs.

$_{14}$ Index

16

17

21

- Abstract
 - Background/Introduction
 - Binning problem definition (recover biological entities from metagenomic sequencing)
- problem relevance (Explosion in metagenomics, reduction in sequencing cost, increased computer capacity)
- Review objectives (Brief summary on popular tools, innovations overview of recent tools)
 - Popular/Previous Binning software
- Proposed solutions (bin contings into bins(MAG if good quality) based on their kmer composition and abundance/coabundance)

- Tools available (Cite recent benchmark)
 - Overview of recent metagenomic binning tools
 - Inovations in binning tools
 - * Inovations in proposed solutions/strategy innovations Read binning gene-abundance binning (CAG, MGS, MSPi) Integrate new experimental data
 - * Software/algorithms innovations machine-learning/deep-learning implementation
- Inovations in specific biological questions Viral genomes and viral strains; Endosymbionts
 - Choosing a binning algorithm
- Identify start point variables
 - * Sample origin (Host contamination, diversity)
- * Number of samples (some tools require many samples to perform well)
- * Sequencing technology (Most tools employ illumina, LongReads are increasing)
- * $Computational\ resources\ available$
- Identify endpoint
 - * organism of interest viral(ref viral catalogue), bacteria, all
- Tools are complementary MSP/Metabat
- Conclusions

32

- Current limitations and future directions Do not perform well on multiple strains, on the same sample
- Figure. Binning software historical citations barplot Figure. Decision tree, overview of metagenomic binning Table. List of binning software since 2017

46 Background

- 47 The explosion in popularity and success in the field of metagenomics over the last 25 years can
- 48 be largely attributed to the advances in computing tecnologies. An example of the outcomes of
- 49 this can be found in the Human Microbiome Project; a project that has been greatly imrpoved
- the understanding of the microbila flora involved in human health and disease. These advances

have brought with them greater demands for storage, CPU time, and consequently more efficient algorithms. The main function of binning tools is to reconstruct species/biological entities from metagenomic samples. Compared to amplicon, shotgun metagenome can provide functional gene profiles directly and reach a much higher resolution of taxonomic annotation. However, due to the high demands on computational resources, cost, and expertise necessary to perform this analysis, researchers have historically been limited in their capacity to collect and analyse sequencing data. As the cost of sequencing is rapidly falling, this burden has been significantly lessened At the time of writing, shotgun metagenomic sequeng costs on average three times as much as 16S sequencing comparitavely. The objectives of this review is for the reader to be better informed about the latest algorithms (since 2017) for binning metagenomic samples. The second part of this review is for the reader to be informed about distinguishing factors between the methods. The last part is for the reader to make an informed decision based on those factors for their needs. This review will be broken down into the following sections:

- List of the binning algorithms
- Classify binning algorithms based on their objectives, guideline for algorithm choice, subsection msp mag
 - Current limitations and Future directions

Recent methods for metagenomic binning

A metagenomic sample is comformed of many organisms and the standard procedure is to retrieve the sequences from the mixture of organisms. The final goal of binning is to reconstruct the sequences from each organism present in the original sample. Among the binning tools developed in recent years we can distinguish a subset dedicated to cluster reads (read-binning) (MetaBBC-LR, BioBloom Tools, CLAME, LVQ-KKN, Meta VW, HirBin, MEGAN-LR). The main purpose of read-binning tools is to preprocess reads into clusters for a posterior targeted assembly, here we find reference-free and non-reference-free tools, and tools designed for short-read or long-read sequencing technologies. The majority of binning tools we can find are oriented toward clustering contigs (contig-binning) into bins, which may represent the genome from a single biological entity/organism. Contig-binning tools normally rely on coverage information and sequence composition. Progress in contig-binning algorithms can be seen in the proposals to integrate new sources of information (for example, from scaffold-graphs(Binnacle), paired-end reads(COCACOLA), or

- 3D contact information(MetaTOR)) and state of the art algorithms in machine learning (CoCoNet,
- Variational Autoencoders for Metagenomic Binning (VAMB)).

83 Metagenome Assembled Genomes

- ⁸⁴ A Metagenome-Assembled Genome (MAG) is a single-taxon assembly based on one or more binned
- ₈₅ metagenomes that has been asserted to be a close representation to an actual individual genome
- 86 (that could match an already existing isolate or represent a novel isolate).

87 Binning microbial genomes with deep learning

- The integration of deep learning techniques into the field of metagenomics has revolutionised the
- ₈₉ field of metagenomics. The VAMB pipeline was developed to take advantage of variational au-
- $_{90}$ to encoders; a generative machine learning model that uses a combination. Improved metagenome
- 91 binning and assembly using deep variational autoencoders. Nature biotechnology the VAMB
- 92 pipeline (?, ?).

93 Binning for viral genomes

94 New insights from uncultivated genomes of the global human gut microbiome (?, ?).

95 Chosing the most appropriate binning algorithm (Classifica-

stion by output)

- 97 A review on the benchmarking binning algorithms was done by ?, ?. Resource management is
- 98 an important factor in the choice of binning algorithm. The tradeoff between number of Central
- 99 Processing Units (CPUs), memory, and time are important considerations. Newer advances in
- pipeline technologies have ameliorated these costs. Alignment based or alignment free. An analysis
- pipeline is defined as a program that combines several programs in a defined order to complete
- a complex analysis. Improperly developed, validated, and/or monitored pipelines may generate
- 103 inaccurate results.

MSPs, binning co-abundant genes

- Binning of co-abuntant genes represents an alternative proposal to reconstruct species/biological
- entities from a set of metagenomic samples. Co-abundant gene binning methods assume each

Table 1: Comparison of binning algorithms

Software / Algorithm Year Description/propose Common/Highlight Dos CoCoNet 202 Despitation/propose 10.1009/Abusinformatics/Phab213 10.1009/Abusinformatics/Phab213 VANIE 202 Metagenome Banning and Alexa sentiation and and an analysis of Metagenome banning and Metagenome Contige and Metagenome Continuing Deptagenome Continuing Deptagenome Continuing Deptagenome Continuing Deptagenome Continuing Deptagenome Continuing Deptagenome Continuing Continuing Deptagenome Continuing Deptagenome Continuing Deptagenome Continuing Deptagenome Continuing Continuing Deptagenome Continuing Continuing Deptagenome Continuing Continuing Continuing Continuing Continuing Continuing Continuing Continuing Continuin						
2021 Using soaffide to improve Metagenomic bin quality Metagenomic bining and MAG assembly, alignment Suitable for Long-Read bining and MAG assembly, alignment Solitable for Long-Read bining and MAG assembly alignment and seembled contrigs using as an incorporate as seasmbly graphs information and properties of the preparation of the properties of the proper	Software/Algorithm	Year		${ m Comment/Highlight}$	Doi	PubmedID
222 Using scaffolds to improve Mategagnome Binning Reconstructs varied at the processing Activities and MAG assembly alterated and Machagagnome Binning using deep variational autoc 222 Metagenomic binning study of Congress and Mategagnome Congress assembly and MAG assembly. Surface assembly, alterated and Mategagnomic congress assembly and MAG assembly alterated assembly alterated binning of mateganomic congress assembly and mateganomic attains and mateganomic congress assembly and mateganomic attains and mateganomic congress assembly and mateganomic attains and mateganomic congress assembly and and annotation. 2011 Improvement binning and mateganome binning and annotation. 2012 Metagenomic binning and mateganome complets and annotation. 2013 Metagenomic binning and association of plantane and annotation. 2014 Metagenomic binning and association of plantane and annotation. 2015 Improve decembly and association of plantane and annotation. 2016 Improved contains agreegation and annotation. 2017 Intercruce-free learning sequence classific 2018 Interpret of Metagenome Binners and annotation. 2019 Interpret of Metagenome Binners and annotation and annotat	7	000			5 5 6 6 6 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7	
2021 Meragenome binning stating the transcription and the control of the greated seembly alignment. 2020 Meragenomic binning stating as a serifical profit in and And Ca assembly alignment. 2020 Meragenomic binning for targeted assembly, alignment. 2020 Reade binning for targeted assembly, alignment. 2020 Simulating metagenomic eabble isotope probing d. 2020 Simulating metagenomic eabble isotope probing d. 2020 Simulating metagenomic eabble focus. 2020 Simulating metagenomic exploration makeds. 2020 Simulating metagenomic exploration for mind analysis of MAGs from manualian gets using meta. 2020 Simulating metagenomic planting with semi-supervise. 2020 Signal processing method for alignment free met. 2020 Signal processing method for alignment free method seem for method seem for se	CoCoNet	2021	Deep learning tool for Viral Metagenome Binning	Reconstucts viral genomes	10.1093/bioinformatics/btab213	33822891
2020 Metagenome brinning using deep variational autoe 2020 Reads binning sind deep variational autoe 2020 Metagenome brinning suits for targeted assembly. 2020 Reads binning for targeted assembly, alignment 2020 Simulating metagenomic stable isotope probling assembled course. 2020 Stable binning algorithm for large scale mean 2021 Appropriate brinning algorithm for large scale mean 2022 Stable binning algorithm for large scale mean 2023 Alignment based algorithm allowed description of microbial genomes from individual 2024 Alignment based Reportant metagenomic sequence fragments between the search of microbial parameters and the search of the searc	Binnacle	2021	Using scaffolds to improve Metagenomic bin quality	Incorporates scaffold information	10.3389/fmicb.2021.638561	33717033
2020 SRINA profiling and MAG assembly 2020 Metagenomic binning for Long-Reads 2020 Metagenomic binning for Long-Reads 2020 Refined binning for targeted assembly, alignment. 2020 Simulating metagenomic contigs using assembly graphs information 2020 Simulating metagenomic contigs using extended contigs 2020 Simulating metagenomic contigs using content in the properation for large scale metage. 2020 Simulating metagenomic contigs using content in mixed s. 2020 Simulating metagenomic contigs using content in mixed s. 2020 Simulating metagenomic state in an analysis of MACs 2021 Grant manualian gits using from mixed s. 2021 Grant manualian gits using from mixed s. 2022 Simulating metagenomic scale metagenome in dividual. 2023 Grant metagenome binning disputed description of mirrobal parameter through the properties of population for genome reconstruction of mirrobal parameter description of mirrobal parameter description of mirrobal parameter comp. 2023 Grant mirrobal mining disputed description of mirrobal parameter description of mirrobal parameter comp. 2024 Simulation metagenomic binning disputed description of mirrobal parameter description of	VAMB	2021	Metagenome binning using deep variational autoe	Autoencoder algorithm, fast processing	10.1038/s41587-020-00777-4	33398153
2020 Reads binning for totageted assembly, alignment 2020 Reads binning for totageted assembly, alignment 2020 Reads binning for totageted assembly, alignment 2020 Reads binning for targeted assembly, alignment 2020 Reads binning of netagenomic contage using as 2020 Simulating metagenomic stable isotope probligs and analysis of MAGs 2020 Binning viral haplotypes from assembled contage. 2020 The proportion and analysis of MAGs 2020 Enone separation of Endosymbionts from mixed s 2021 Ashaptative binning algorithm for large scale mear 2021 Ashaptative binning algorithm for large scale mear 2022 Alignment based algorithm individual 2023 Alignment based algorithm individual 2024 Alignment based algorithm individual 2025 Alignment based constitution of mitrobial parases and aspectation of mitrobial general algorithm of metagenomic sequence classific 2026 Angegonemic binning trough low description of mitrobial parases mention of mitrobial parases mention of mitrobial parases and aspectation of mitrobial parases and parases and aspectation of mitrobial parases a	phyloFlash	2020	ssrRNA profiling and MAG assembly	incorporates ssrRNA profiling info into MAG ass	10.1128/mSystems.00920-20	33109753
2020 Reads binning for targeted assembly, alignment 2020 Simulating mategenomic contigs using as 2020 Simulating mategenomic contigs using as 2020 Simulating mategenomic contigs using as 2020 Simulating mategenomic stable isatope probing d 2021 Genome separation of Endosymbionts from marks 2022 Genome separation of Endosymbionts from marks 2023 Genome separation of Endosymbionts from marks 2024 Adaptative binning algorithm for lage scale meta 2025 Method for partitioning polypioid sub genomes b 2026 Method for partitioning polypioid sub genomes b 2027 Method for partitioning with remeta 2028 Simulation sequence fragment free met 2029 Signal processing method for alignment free met 2020 Signal processing method for alignment free met 2021 Signal processing method for alignment free met 2022 Alignment based algorithm allowed description of microbial pan 2023 Alignment based algorithm allowed description of microbial pan 2024 Alignment based algorithm allowed description of microbial pan 2025 Alignment based algorithm allowed marks sequence comp. 2026 Alignment based algorithm allowed marks sequence comp. 2027 Alternation algorithm allowed description of microbial pan 2028 Abundance based reconstitution of microbial pan 2029 Alignment based marks and all pances and algorithm marks and algorithm algorithm marks and algorithm algo	MetaBCC-LR	2020	Metagenomic binning for Long-Reads	Suitable for Long Reads sequencing technology	10.1093/bioinformatics/btaa441	32657364
2020 Refined binning of netagenomic contigs as: 2020 Simulating metagenomic stable isotope probing a content binning viral haplotypes from assembled contigs 2020 Simulating metagenomic stable isotope probing a pipeline and search of Endosymbions from mixed s 2021 Genome separation of Endosymbions from mixed s 2022 Applicative binning algorithm for genome recons 2023 Adugment black of partitioning polypic did up genomes b 2024 Mignant based algorithm allowed description of microbial genomes from individual 2025 Anguant based and content free met 2026 Anguant based angorithm allowed description of microbial genome from individual 2027 Alignment based angorithm allowed description of microbial genome resolved metagenome sequence chagners 2028 Anguant based angorithm allowed description of microbial genome econstitution of differential 2017 Metagenomic binning and acoring strategy and contractive visualisation tool for verification. Single sample, microbial genome propriation genome binning using fuzzy C-Means Method of mining using fuzzy C-Means Method of contractive visualisation tool	BioBloom Tools	2020	Reads binning for targeted assembly, alignment	Data preparation for targeted assembly, using s	10.1073/pnas.1903436117	32641514
2019 Ususpervised binning kenners and coverage, focus. 2019 Binning viral haplotypes from analysis of MAG as the production and analysis of MAG as the processing mention of microbial genomes from individual. 2019 Adaptative binning algorithm for genome reconstruction of microbial genomes from individual. 2019 Adaptative binning algorithm for genome reconstruction of microbial genomes from individual. 2019 Adaptative binning algorithm for genome reconstruction of microbial genomes from individual. 2019 Adaptative binning algorithm for genome reconstruction of microbial genomes from individual. 2019 Adaptative binning algorithm allowed description of microbial genomes from individual. 2019 Adaptative binning algorithm allowed description of microbial genome based reconstitution of microbial pipeline for genome resolved metagenome. 2018 Aligement based algorithm allowed description of sequence capsification aggrenates of the production algorithm of microbial pipeline for genome resolved metagenome. 2018 Aligement based algorithm and scoring stategy. 2019 Annalysis through the combination of microbial pipeline for genome resolved metagenome. 2018 Aligement of Meragenome Binning and association of plasmida. 2019 Acres and algorithm and sociation of plasmida. 2010 Acres and annotation of differential annotation of differential compactive searetiplost supervised binning and association of plasmida. 2010 Acres and Adaptation recovery of population	GraphBin	2020	Refined binning of metagenomic contigs using as	Incorporates assembly graphs information	10.1093/bioinformatics/btaa180	32167528
2019 Gramework for Production and analysis of MAGs 2019 Framework for Production and analysis of MAGs 2019 Gramework for Production and analysis of MAGs 2019 Gramework for Production and analysis of MAGs 2019 High quality MAGs from manualian guts of meta- 2019 Adatoptative binning algorithm for genome recons 2019 Gramework for participing polyploid sub genomes b 2019 Method for participing polyploid sub genomes b 2019 Adatoptation of microbial genomes from individual 2019 Adatoptation of microbial genome reconstitution of microbial genome from individual 2018 Abundance based reconstitution of microbial panametric production of microbial panametric	MetaSIPSim	2020	Simulating metagenomic stable isotope probing d	Augment binning resolution with extra experimen	10.1186/s12859-020-3372-6	32000676
Pinning viral haplotypes from analysis of MAGs	MetaCon	2019	Unsupervised binning k-mers and coverage, focus	Focus different lengths contigs	10.1186/s12859-019-2904-4	31757198
Pramework for Production and analysis of MA Gss	VirBin	2019	Binning viral haplotypes from assembled contigs	Viral haplotypes MAGs	10.1186/s12859-019-3138-1	31684876
2019 Genome separation of Endosymbionts from mixed s 2019 Adap patity MAGs from mammalian guts using met 2019 Adap patity by many against method for partitioning algorithm for genome recons 2019 Adap patitive binning algorithm for genome recons 2019 Scalable binning algorithm for genome reconstruction of microbial genomes from individual 2019 Signal processing method for alignment free met 2018 Fuzzy binning of metagenomic sequence fragments from microbial genomes from individual 2018 Fuzzy binning of metagenomic sequence fragments for microbial pan 2018 Fuzzy binning of metagenome for microbial pan 2019 Metagenomic binning active or microbial genome resolved metagenom 2018 Metagenomic binning and sage sequence classific 2019 Metagenomic binning and association of plasmids 2018 Binning contigs using coverage and additional panels and annotation 2019 Metagenomic binning and association of plasmids 2010 Metagenomic binning and association of plasmids 2011 Metagenomic binning and association of differential 2012 Metagenomic binning and association of differential 2013 Metagenomic binning and annotation 2014 High resolution identification of differential 2015 Metagenomic binning and association of differential 2016 Metagenomic binning and association of differential 2017 Metagenomic binning and association decome 2018 Metagenomic properties of contrage and addition tool for verification 2019 Metagenomic binning and association decome 2010 Metagenomic properties of contrage and addition tool for verification 2011 Metagenomic binning using Pruzzy C-Means Method 2012 Metagenome bins through the combination of differential 2013 Metagenome bins through the combination general decompanity of differential 2014 Thorove genome bins through the combination ge	MAGO (*only tool pipeline)	2019	Framework for Production and analysis of MAGs	pipeline	10.1093/molbev/msz237	31633780
4 dataptative binning algorithm for genome recons 2019 Adaptative binning algorithm for genome recons 2019 Adaptative binning algorithm for large scale meta 2019 Adaptative binning algorithm for large scale meta 2019 Method for partitioning with semi-supervis 2019 Extraction of microbial genomes from individual 2018 Fuzzy binning of metagenomic binnerselved for alignment free met 2018 Fuzzy binning of metagenomic sequence fragements composition based RNA or DNA binning of short s 2018 Fuzzy binning of metagenomic sequence fragements composition based RNA or DNA binning of short s 2018 Fuzzy binning of metagenomic sequence classific 2018 Fuzzy binning of metagenomic sequence classific 2018 Adaptative binning through low density binning of short s 2019 Adaptative binning with semi-supervise 2018 Fuzzy binning of metagenomic sequence classific 2018 Adaptative binning with sequence flassific 2019 Adaptative binning with semi-supervise. 2018 In ming contigs using coden usage sequence composition 2019 Adaptative binning and annotation interactive visualisation of differential 2010 Metagenomic binning and association of plasmids 2011 Improved binning using Fuzzy C-Means Method 2012 Inproved binning using Fuzzy C-Means Method 2014 Figh teapluative or manufacture and manufacture and parameter turing prevised binning complete and an expension of differential 2019 Adaptation and secondary and annotation and annotation are an expension to form the combination of combeting and annotation are an expension of differential 2017 Adaptation and secondary and an experiment of the compination of differential 2018 Adaptation density binning worked binning using Fuzzy C-Means Method 2019 Adaptation general and quantification of differential 2010 Fuzzy C-Means Method 2011 Fuzzy C-Means Method 2012 Fuzzy C-Means Method 2013 Fuzzy C-Means Method 2014 Fuzzy C-Means Method 2015 Fuzzy C-Means Method 2016 Fuzzy C-Means Method 2017 Fuzzy C-Means Method 2018 Fuzzy C-Means Met	SeqDex	2019	Genome separation of Endosymbionts from mixed s	Identification of endosymbiont	10.3389/fgene.2019.00853	31608107
2019 Adatptative binning algorithm for large scale meta 2019 Method for partitioning polyploid sub genomes becomes b 2019 Method for partitioning polyploid sub genomes b 2019 Signal processing method for alignment free met 2019 Signal processing method for alignment free met 2018 Fuzzy binning of metagenomic sequence fragments 2018 Alignment based algorithm allowed description of 2018 Fuzzy binning of metagenomic sequence fragments 2018 Animal passed reconstitution of microbial pan 2019 Fuzzy binning of metagenomic sequence classific 2018 Fuzzy binning of metagenomic sequence classific 2018 Animal passed reconstitution of microbial pan 2019 Fuzzy binning of metagenomic sequence classific 2018 Metagenomic binning through low density binning 2018 Binning cortigs using coden usage sequence comp 2018 Assessment of Metagenome Binners 2018 Assessment of Metagenome binning and association of plasmids 2017 Metagenomic binning and association of differential 2018 Metagenomic binning and association of differential 2017 Metagenomic binning and annotation 2018 Metagenomic binning and annotation 2019 Metagenomic binning and annotation 2010 Metagenomic binning and annotation 2011 Metagenomic binning and annotation 2012 Metagenomic binning and annotation 2013 Metagenomic binning surjucy expense and differential and annotation 2014 Timproved genome bins through the combination of of differential 2017 Metagenomic binning using Fuzzy C-Means Method 2018 Metagenomic binning using Pruzy C-Means Method 2019 Metagenomic binning using Pruzy C-Means Method 2010 Metagenomic binning contrage and an description of population general description general description general 2011 Metagenomic binning using Pruzy C-Means Method 2012 Metagenomic binning using Pruzy C-Means Method 2013 Metagenomic binning using Pruzy C-Means Method 2014 Tool for automatic recovery of population general description general description general description general description general descript	MetaTOR	2019	High quality MAGs from mammalian guts using met	Incorporates 3D contact information	10.3389/fgene.2019.00753	31481973
Scalable binning algorithm for large scale meta Soluble binning and grithm for large scale meta Method for partitioning polyploid sup genomes b Method for partitioning polyploid supprevois Method for partitioning with semi-supervis Soluble pineling processing method for alignment free met Method for partitioning with semi-supervis Migment based algorithm allowed description of Alignent based favorithm allowed description of Alignent based algorithm allowed binning using Puzzy C-Means Method Alignent of long sequence classific Alignent of long sequence sequences Alignent based for highly allowed binning using Puzzy C-Means Method Add codon usage information of differential Plasmid control for verification of differential Alignent based for highly allowed binning using Puzzy C-Means Method Add sequence endesigned content or competition, or	MetaBAT 2	2019	Adatptative binning algorithm for genome recons	Eliminates manual parameter tuning from previou	10.7717/peerj.7359	31388474
2019 Method for partitioning with semi-supervis 2019 Extraction of microbial genomes from individual 2019 Signate processing method for alignment free met 2019 Signate processing method for alignment free met 2018 Fuzzy binning of metagement free met 2018 Animatic based algorithm allowed description of 2018 Animatic based reconstitution of microbial pan 2019 Fuzzy binning of metagement free met 2018 Abundance based reconstitution of microbial pan 2019 Fuzzy binning of metagement free met 2018 Abundance based reconstitution of microbial pan 2019 Fuzzy binning of metagement free met 2019 Fuzzy binning of metagement free met 2018 Abundance based reconstitution of microbial pan 2019 Fuzzy binning of metagement free met 2018 Abundance based reconstitution of microbial pan 2019 Fuzzy binning of metagement free met 2018 Abundance based reconstitution of microbial pan 2019 Fuzzy binning of metagement free met 2018 Abundance based reconstitution of microbial pan 2019 Metagenomic binning acquere classific 2018 Abundance based reconstitution of microbial pan 2019 Metagenomic binning and secriting strategy. 2010 Metagenomic binning and secriting strategy. 2011 Metagenomic binning and association of plasmids 2012 Metagenomic binning and annotation. 2013 Metagenomic binning and annotation. 2014 Tight resolution identification of differential 2017 Puzzy Chaens Method 2018 Fuzzy Chaens Method 2019 Metagenome bins through her calculate oligonucleotide 2010 Fuzzy Chaens Method 2011 Metagenomic binning and annotation. 2012 Metagement of oligenucleotide 2013 Metagement of oligenucleotide 2014 Metagenomic pan. Method 2015 Metagement of differential 2016 Fuzzy Chaens Method 2017 Metagement of oligenucleotide 2017 Metagement of differential 2018 Method formula for calculate oligenucleotide 2019 Metagement of oligenucleotide 2010 Metagement of oligenucleotide 2011 Metagement of oligenucleotide	MetaBMF	2019	Scalable binning algorithm for large scale meta	Employs sample X contigs of mapped read counts	10.1093/bioinformatics/btz577	31347687
2019 Improving metageanome binning with semi-supervise 2019 Signal processing method for alignment free met 2018 Alignment based algorithm allowed description of 2018 Anguent based algorithm allowed description of 2018 Puzzy binning of metagenomic sequence fragments. 2018 Composition based RNA or DNA binning of short s 2018 Abundance based reconstitution of microbial pan 2018 Abundance based reconstitution of microbial pan 2018 Metagenomic binning through low density binning carties using coverage and comming corties are coverage and comming corties and expension of plasmids. 2018 Dinning corties using coverage and comming corties are coverage and affilement of long reads against reference sequence comp 2017 Improved founding using prize genome bins through the combination of microbial pan 2018 Improved comming using Puzzy C-Means Method 2019 Signal processing method for alignment free met 2018 Abundance based Roconstitution of microbial pan 2018 Metagenomic binning and association of plasmids 2019 Metagenomic binning and association of differential 2017 Improved conting strategy and composition, rederence-free learning with multiple metagenom 2017 Improved binning using Puzzy C-Means Method 2018 Improved second to differential 2019 Metagenome bins through the combination of microbial pan 2017 Improved proved plasmids are composition, read comming contrage and added participation of opportation genome. 2017 Improved planning using Puzzy C-Means Method 2018 Metagenome plass through the combination of read genome lengths 2019 Metagenome forms through the combination of read genome lengths 2019 Metagenome forms through the combination of read propervised clustering and annotation. 2017 Metagenome for differential 2018 Metagenome for differential 2019 Metagenome for differential 2011 Metagenome for differential 2012 Metagenome	PolyCRACKER	2019	Method for partitioning polyploid sub genomes b	Haplotypes for polyploid genomes	10.1186/s12864-019-5828-5	31299888
2019 extraction of microbial genomes free met 2018 Alignment based algorithm allowed description of 2018 Alignment based algorithm allowed description of 2018 Fuzzy binning of metageanomic sequence fragaments 2018 Composition based RNA or DNA binning of short s 2018 Flexible pipeline for genome resolved metagenom 2018 Evizy bening of metageanomic operation of sequence classiffer 2018 Assessment of Metagenome Binners comp 2018 Encreptication aggregation and scoring strategy 2018 Encreptication aggregation and scoring strategy 2019 Encreptication aggregation and scoring strategy 2010 Encreptication aggregation and scoring strategy 2011 Encreative visualisation of plasmids 2012 Metagenomic binning and association of plasmids 2013 Encreative elearning with multiple metagenom 2014 Tipp recounting suing coverage and aff fire-rental to of ordiferental in ming contiguation deliation deliation identification of differential 2019 Assistance of the straing with multiple metagenom 2010 Encreaved formula for calculate oligonuclectide 2011 Improved commits using Fuzzy C-Means Method 2012 Improve denome bins through the combination of differential coverage and aff fire prepared on the properties of the properties of the properties of the properties of population genome 2019 Assistance of the properties of the properties of contractic seconds of the properties of the properties of properties of contractic seconds of the properties of	SolidBin	2019	Improving metagenome binning with semi-supervis	NaN	10.1093/bioinformatics/btz253	30977806
2019 Signal processing method for alignment free met 2018 Alignment based algorithm allowed description of 2018 Fuzzy binning of metagenomic sequence fragments 2018 Composition based RAA or DNA binning of short s 2018 Composition based RAA or DNA binning of short s 2018 Composition based RAA or DNA binning of short s 2018 Abundance based reconstitution of microbial pan 2018 Abundance based reconstitution of microbial pan 2018 Metagenomic binning through low density binning 2018 Assessment of Metagenome Binners 2018 Binning contrigs using codon usage sequence comp 2018 Assessment of Metagenome Binners 2018 Assessment of Metagenome Binners 2018 Assessment of Metagenome Binners 2019 Complexation and scoring strategy 2017 Improved formula for calculate oligoned codied 2017 Improved binning using preversive clustering using converge and affi 2017 Improved binning using Puzzy C-Means Method 2018 Improved binning using Puzzy C-Means Method 2019 Improved binning using Puzzy C-Means Method 2010 Simple position identification of proposement and describing the composition, rand coverage 2011 Alignment properties of complexition of predation of genome bins through the combination of complexition of propulation genome 2017 Adde pulpeding and annotation and affirential and annotation and affirential annotation and addes single complexition of predational and composition, rand coverage 2019 Assessment of Metagenome Binning contriguence classific 2017 Proproved contriguence comp 2018 Assessment of Metagenome Binning and annotation and addes single samples 2019 Assessment of Metagenome Binning and annotation and addes single samples 2010 Assessment of all association of differential annotation and addes single	Autometa	2019	extraction of microbial genomes from individual	Handles eukaryotic contamination	10.1093/nar/gkz148	30838416
2018 Alignent based algorithm allowed description of 2018 Composition based RNA or DNA binning of shorts 2019 Abundance based reconstitution of microbial pan 2019 Abundance based reconstitution of microbial pan 2018 Abundance based reconstitution of microbial pan 2018 Flexible pipeline for genome resolved metagenom: 2018 Large scale Machine Learning Sequence classific 2018 Metagenomic binning through low density binning and scoring strategy. 2018 Binning vorkflow using contain coverage and com 2017 Metagenomic binning and association of plasmids 2017 Interactive visualisation tool for verification 2017 Interactive clustering using coverage and affire 2017 Interactive clustering using coverage and affire 2017 Interactive clustering using coverage and affire 2018 Binning contigs taxonomic binning and amortation 2017 Interactive clustering using coverage and affire 2018 Computervised clustering and amortation 2019 Inprove denoming using Fuzzy C-Means Method 2010 Inprove domining using Puzzy C-Means Method 2011 Improve domining using Puzzy C-Means Method 2012 Inprove domining using Puzzy C-Means Method 2013 Abundance based reconstitution of microbial pan 2014 Tool of page 100 of microbial pan 2015 According taxonomic binning contigs using contain coverage and affine microbial page provised page 100 of microbial page 100 of m	MLBP MrGBP (Algorithm)	2019	Signal processing method for alignment free met	Alternative description of sequences designed f	10.1038/s41598-018-38197-9	30770850
2018 Fuzzy binning of metagements 2018 Composition based RNA or DNA binning of short s 2018 Abundance based reconstitution of microbial pan 2018 Abundance based reconstitution of microbial pan 2018 Fuzzibe pipplicite for genome resolved metagenome 2018 Fuzzibe pipplicite for genome resolved metagenome 2018 Matagenomic binning through low density binning 2018 Binning contigs using coden usage sequence comp 2018 Matagenomic binning and secritation aggregation and scoring strategy. 2019 Derreplication aggregation and scoring strategy. 2017 Metagenomic binning and association of plasmids 2017 Metagenomic binning and association of plasmids 2017 Metagenomic binning and annotation 2018 Improved clearning with multiple metagenom 2017 Improved clearning with multiple metagenom 2017 High resolution identification of differential 2018 Improve genome bins through the combination of 2019 Improved binning using Fuzzy C-Means Method 2016 Imming contigs using coverage and aff 2017 Combines several binning and serverage and aff 2018 Particular of severage and aff 2019 Particular of metagenome in general sequence of lassim 2017 Requires multiple samples to sequence dissim 2018 Particular of differential 2017 Requires multiple samples supervised combestion, unsupervised clustering using coverage and aff 2017 Combines several binning at strain level using methylati 2018 Parting workflow using coverage and aff 2019 Particular of differential 2010 Particular of differential 2011 Requires multiple samples supervised binning august strain level using methylati 2012 Reduce bias for high/low abundance combestion, unsupervised clustering using coverage and aff 2014 Tool of our submatic recovery of population genema 2015 Abundance comparage to composition, rand coverage 2016 Abundance comparage and combestion 2017 Adds patier dead of combestion 2018 Abundance comparage of population genemal and consignment op	CLAME	2018	Aligment based algorithm allowed description of	Aligment based for reads	10.1186/s12864-018-5191-y	30537931
2018 Composition based RAO or DNA binning of short s 2018 Interpret the pipeline for genome resolved metagenom 2018 Flexible pipeline for genome resolved metagenom 2018 Interpret the pipeline for genome resolved metagenom 2018 Metagenomic binning through low density binning 2018 Binning contigs using codon usage sequence comp 2018 Assessment of Metagenome Binners 2018 Derreplication aggregation and scoring strategy 2018 Interpret the conting staxonomic binning and association of plasmids 2017 Metagenomic binning and anotation of differentiall 2017 Metagenomic binning and anotation densitie oligonucleoide 2017 Metagenomic binning using coverage and aff flat resolution identification of differentiall 2017 Improved formula for calculate oligonucleoide 2018 Metagenome binning using converage and aff flat resolution identification of differentiall 2017 Improve genome bins through the combination of 2018 Metagenome dinning using coverage and aff flat resolution identification of differentiall 2017 Improve genome bins through the combination of 2018 Metagenome plant density binning using coverage 2019 Metagenome bins through the combination of 2010 Metagenome density binning using coverage 2011 Improved binning using composition, read coverage 2012 Metagenome binning using contiguation generally and another density and another density binning algorithms density density density and another density densi	3D BH SNE (Algorithm)	2018	Fuzzy binning of metagenomic sequence fragments	Horizontal gene transfer and regions of uncerta	10.1109/EMBC.2018.8512529	30440633
2018 Abudance based reconstitution of microbial pan 2018 Abudance based reconstitution of microbial pan 2018 Large scale Machine Learning Sequence classific 2018 Margaeomoic binning through low density binning 2018 Binning contigs using codon usage sequence classific 2018 Assessment of Margaeomoic Binning 2018 Binning contigs using codon usage sequence comp 2018 Assessment of Margaeomoic Binning 2018 Long Read / contigs taxonomic binning 2018 Metagenomic binning and association of plasmids 2017 Reference-free learning with multiple metageomo 2017 Interactive visualisation tool for verification 2017 Map resolution identification of differential 2017 Interactive visualisation tool for verification 2017 Improved clustering using coverage and aff 2018 Manuals for calculate oligomedecide 2017 Repended clustering using coverage and aff 2018 Manuals for calculate oligomedecide 2019 Math formula to realculate oligomedecide 2017 Repart taxonomic levels, discove Plasming deciden aggregation and scoring strategy Alignent of long reads against reference sequences simple include gc content and dance fre Plasmid baning at strain level using methylati Plasmid baning at strain level using methylati Adal paire annotation of differential 2017 Interactive visualisation tool for verification of differential 2018 Manuals for calculate oligomedecide 2019 Math formula to calculate oligomedecide 2010 Interactive scatterplots supervised binning 2011 Improve contiguation and aff 2011 Combination of differential 2012 Manuals for calculate oligomedecide 2013 Math formula of differential 2014 Math resolution identification of differential 2015 Math resolution identification of differential 2016 Improve genome bins through the combination of differential overage or competition 2017 Math spatial or composition, and competition designed to composition of differential or composition 2018 Math spatial designation des	LVQ-KNN	2018	Composition based RNA or DNA binning of short s	Classify into DNA or RNA sequence	10.1016/j.virusres.2018.10.002	30291874
2018 Flexible pipeline for genome resolved metagenom 2018 Metagenomic binning through low density binning 2018 Metagenomic binning through low density binning 2018 Binning contrigs using codon usage sequence comp 2018 Binning contrigs using codon usage sequence comp 2018 Binning contrigor using codon usage sequence comp 2018 Binning contrigor and second strategy 2018 Binning workflow using contain coverage and com 2017 Improved formula for exclusion of plasmids 2017 Improved contrigor strategy 2017 Improved converge and association of plasmids 2017 Improve genome bins through the combination of improved competition identification of differentiall 2017 Improve genome bins through the combination of improved competition identification of differentiall 2016 Improved companies of differentiall 2017 Improve genome bins through the combination of improved competition, and annotation in the reactive standard and condition of differentiall 2016 Improved competition, read coverage 2017 Improve genome bins through the combination of improved binning using Fuzzy C-Means Method 2014 Add setimated distribution of real genome lengths 2015 Improved composition, read coverage 2016 Improved planning using Fuzzy C-Means Method 2017 Improve genome bins through the combination of adds estimated distribution of real genome lengths 2016 Improved planning using Fuzzy C-Means Method 2017 Improve genome bins through the combination of adds estimated distribution of real genome lengths 2018 Improved composition 2019 Improved composition, read coverage 2010 Improved planning using Fuzzy C-Means Method 2011 Improved planning using Fuzzy C-Means Method 2012 Improved planning using Fuzzy C-Means Method 2014 Indianal Coverage and and configuration of real genome lengths	MSPminer	2018	Abundance based reconstitution of microbial pan	Pan genome reconstitution	10.1093/bioinformatics/bty830	30252023
2018 Large scale Machine Learning Sequence classific 2018 Metagenomic binning through low density binning 2018 Binning contigs using codon usage sequence comp 2018 Assessment of Metagenome Binners 2018 Derreplication aggregation and scoring strategy 2018 Long Read/ contigs taxonomic binning and association of plasmids 2017 Metagenomic binning and anotation. 2017 Interactive visualisation tool for verification. 2017 Mish resolution identification of differential 2017 Improved clustering using coverage and aff 2018 Metagenomic binning using Puzzy C-Means Method 2016 binning contigs using composition, read coverage 2017 Association of differential 2018 Metagenomic binning and association of differential 2019 Metagenomic binning using Puzzy C-Means Method 2014 Tool for automatic recovery of population general 2015 Association of differential 2016 Association of differential 2017 Add estimated distribution of read genome lengths 2016 Improved binning using Puzzy C-Means Method 2017 Add estimated coverage or and	MetaWRAP*	2018	Flexible pipeline for genome resolved metagenom	Hybrid bin extraction algorithm	10.1186/s40168-018-0541-1	30219103
2018 Metagenomic binning through low density binning 2018 Binning contigs using codon usage sequence comp 2018 Assessment of Metagenome Binners 2018 Assessment of Metagenome Binners 2018 Interplication aggregation and scoring strategy 2018 Metagenomic binning and association of plasmids 2017 Metagenomic binning and association of plasmids 2017 Improved formula for calculate oligonuclectide 2017 Migh resolution identification of differentiall 2017 Improve genome bins through the combination of 2017 Improve genome bins through the combination of 2018 Metagenomic binning and annotation 2019 Metagenomic binning and annotation 2017 Improve denome bins through the combination of 2018 Metagenomic binning and annotation 2019 Metagenomic binning and annotation 2017 Improve genome bins through the combination of 2018 Metagenomic binning using Fuzzy C-Means Method 2019 Metagenomic binning using Fuzzy C-Means Method 2016 Improved composition 2017 Adds spatient at dame reactions of population general and complement compose	MetaVW	2018	Large scale Machine Learning Sequence classific	Machine learning for reads based on Khmer profile	10.1007/978-1-4939-8561-6_2	30030800
2018 Binning contige using coden usage sequence comp 2018 Derreplication aggregation and scoring strategy 2018 Lore Read Counties taxonomic binning 2018 In proved formula for calculate oligonacleoide 2017 Metagenomic binning and association of plasmids 2017 Improved formula for calculate oligonacleoide 2017 Interactive visualisation tool for verification 2017 Interactive visualisation tool for verification 2017 Improved clustering using coverage and aff fifter and annotation improved clustering using coverage and aff fifter annotation in the combination of a fifter and annotation in the combination of binning using composition, read coverage 2016 Improved binning using composition, read coverage 2017 All the resolution identification of differentiall 2017 Improved composition, read coverage 2018 Lore Read Conting and annotation of a fifteential and a fifteential and annotation of a fifteential and a fifteent	Opal (algorithm*)	2018	Metagenomic binning through low density binning	Improvement at higher taxonomic levels, discove	10.1093/bioinformatics/bty611	30010790
2018 Derreplication aggregation and scoring strategy 2018 Independent of contigs taxonomic binning 2018 Independent of contigs taxonomic binning 2018 Independent of contigs taxonomic binning 2017 Metagenomic binning and association of plasmids 2017 Independent of contiguent of contiguent of plasmids 2017 Independent of contiguent of plasmids 2017 Independent of contiguent of plasmids 2017 Independent of contiguent of of contiguent of of contiguent of plasmids 2018 Independent of contiguent of of contiguent of of plasmids 2019 Independent of of contiguent of of plasmids 2010 Independent of of contiguent of of plasmids 2011 Independent of of contiguent of of plasmids 2012 Independent of of contiguent of of plasmids 2013 Independent of of contiguent of of plasmids 2014 Independent of of contiguent of of plasmids 2015 Independent of of contiguent of of plasmids 2016 Independent of of contiguent of of plasmids 2016 Independent of of contiguent of of plasmids 2017 Independent of of contiguent of of plasmids 2018 Independent of of contiguent of of plasmids 2019 Independent of of contiguent of of plasmids 2010 Independent of of contiguent of of plasmids 2011 Independent of of contiguent of of plasmids 2012 Independent of of contiguent of of population of of differential occupance of of plasmids 2014 Independent of of population genome 2015 Independent of of population genome 2016 Independent of of population genome 2017 Independent of of population of of plasmids 2018 Independent of of population genome 2019 Independent of of population genome 2010 Independent of of population genome 2011 Independent of of population genome 2012 Independent of of population genome 2013 Independent of of population genome 2014 Independent of plasmids 2015 Independent of plasmids 2016 Independent of plasmids 2016 Independent of plasmids 2017 Independent of plasmids 2018 Independent of plasmids 2018 Independent of plasmids 201	BMC3C	2018	Binning contigs using codon usage sequence comp	Add codon usage information	10.1093/bioinformatics/bty519	29947757
2018 Derreplication aggregation and scoring strategy 2018 Long Read / contigs taxonomic binning 2018 Eliming workflow using contain coverage and com 2017 Metagenomic binning and association of plasmids 2017 Improved formula for calculate oligonucleotide 2017 Interactive visualisation tool for verification 2017 High resolution identification of differentiall 2017 Improve canome blus through the combination of 2017 Improve enome blus through the combination of 2017 Improve genome blus through the combination of 2018 Combination of different bluning algorithms to content and draw free 2017 Improve genome blus through the combination of 2016 Improved binning using coverage and offine 2017 Add estimated distribution of real genome lengths 2016 Improved binning using coverage 2017 Add estimated distribution of real genome lengths 2016 Improved binning using coverage 2017 Add estimated distribution of real genome lengths 2016 Improved binning using coverage 2017 Add estimated distribution of real genome lengths 2018 Add estimated distribution of real genome lengths 2019 Add estimated coverage 2017 Add estimated coverage on an expension of different simplemant composition.	AMBER tool	2018	Assessment of Metagenome Binners	NaN	10.1093/gigascience/giy069	29893851
2018 Long Read, contigs taxonomic binning 2016 Metagenomic binning and association of plasmids 2017 Metagenomic binning and association of plasmids 2017 Metagenomic binning and association of plasmids 2017 Improved formula for calculate oligonucleotide 2017 Botstrapped supervises binning and annotation 2017 Interactive visualisation tool for verification of differentiall 2017 Improved clustering using coverage and aff aff 2017 Improved clustering using coverage and aff aff 2017 Improve genome bins through the combination of 2016 Improved oblining using composition, read coverage 2016 Improved optiming using composition, read coverage 2017 Add estimated distribution of real genome lengths 2016 Improved optiming using composition genome 2017 Add estimated distribution of real genome lengths 2018 Add estimated distribution of real genome lengths 2019 Add estimated coverage 2011 Add estimated distribution of real genome lengths 2012 Add estimated differential information sequences 2013 Add estimated coverage and configuration of optimation genome lengths and another sequences 2014 Add estimated contigues using composition genome	DAS Tool	2018	Derreplication aggregation and scoring strategy	Combines several binning algorithm results	10.1038/s41564-018-0171-1	29807988
2017 Metagenomic binning and association of plasmids 2017 Metagenomic binning and association of plasmids 2017 Improved formula for calculate oligometocidie 2017 Improved formula for calculate oligometocidie 2017 Improved commonication tool for verification 2017 Improved clusterial subjects of the combination of improved clusterial subjects of the combination of influence of different binning algorithms. 2017 Improve genome bins through the combination of composition, read coverage 2018 Improved committee of composition, and associated binning using Fuzzy C-Means Method Adds estimated distribution of real genome lengths and configuration of population genome 2018 Improved composition, read coverage 2019 Improved binning using Fuzzy C-Means Method Adds estimated distribution of real genome lengths and configuration for material and coverage composition 2014 Adds paper and configuration for material and and configurant information of the complement compose 2015 Improved binning using Fuzzy C-Means Method Adds differential coverage to complement compose	MEGAN-LR	2018	Long Read/ contigs taxonomic binning	Aligment of long reads against reference sequences	10.1186/s13062-018-0208-7	29678199
2017 Metagenomic binning and association of plasmids 2017 reference-free learning with multiple metagenom 2017 Improved formula for calculate oligomedovide 2017 Improved supervises binning and annotation and annotation and annotation interactive visualisation tool for verification 2017 High resolution identification of differentiall 2017 Improve genome blus through the combination of 2016 Improved binning using coverage and aff 2016 Improved binning using composition, read coverage 2016 Inning contigs using composition, read coverage 2017 Add estimated distribution of real genome lengths 2016 Inning contigs using composition, read coverage 2017 Add estimated distribution of real genome lengths 2018 Add estimated distribution of call genome lengths 2019 Add estimated distribution of call genome lengths 2016 Inning contigs using composition, read coverage 2017 Add estimated distribution of call genome lengths 2018 Add estimated distribution of call genome lengths 2019 Add estimated coverage and conplement compose	CoMet	2018	Binning workflow using contain coverage and com	Single sample, include gc content and 4mer fre	10.1186/s12859-017-1967-3	29297295
2017 reference-free learning with multiple metagenom 2017 Improved formula for calculate oligomedecide 2017 Bootstrapped supervises binning and annotation 2017 High resolution identification of differentiall 2017 Unsupervised clustering using coverage and aff 2017 Improve genome bins through the combination of 2016 Improved composition, read coverage 2016 Improved composition, read coverage 2017 Add estimated distribution of real genome lengths 2016 Improved pinning using recovery of population genome 2016 Add estimated and coalignment information 2017 Add estimated distribution of real genome lengths 2016 Improved pinning using recovery of population genome 2017 Add estimated distribution of real genome lengths 2018 Add estimated and coalignment information 2019 Add estimated and coalignment information 2019 Add estimated coverage or complement compos	٥.	2017	Metagenomic binning and association of plasmids	Plasmid banning at strain level using methylati	10.1038/nbt.4037	29227468
2017 Improved formula for calculate oligonucleotide 2017 Bootstrapped supervises binning and annotation 2017 Interactive visualisation tool for verification 2017 High resolution identification of differential 2017 Improve genome blus through the combination of a limptove genome blus through the combination of a limptove binning using coverage and affi 2016 Improved binning using rowerdy of population genome lengths 2016 binning contigs using coverage 2017 Add estimated distribution of real genome lengths 2016 binning contigs using coverage 2018 Add spaired end read and coaligment information and a delighment composition.	MetaGen	2017	reference-free learning with multiple metagenom	Requires multiple samples	10.1186/s13059-017-1323-y	28974263
2017 Bootstrapped supervises binning and annotation 2017 Interactive visualisation tool for verification. 2017 Interactive visualisation tool for verification. 2017 High resolution identification of differentiall 2017 Unsupervised clustering using coverage and aff 2017 Improve genome bins through the combination of combination of different binning algorithms 2016 Improved binning using Foracy Chansan Method 2016 binning contigs using coverage, and Adds patiented end read and coaligment information ground and coaligment information. 2017 Adds patient and a composition, read coverage. 2018 Add spatient and coverage or composition genome.	d2sBin add onn	2017	Improved formula for calculate oligonucleotide	Math formula to calculate oligo sequence dissim	10.1186/s12859-017-1835-1	28931373
2017 Interactive visualisation tool for verification 2017 High resolution identification of differentiall 2017 Unsupervised clustering using coverage and affi 2017 Improve genome bins through the combination of 2016 Improved principle prazy C-Means Method 2016 Improved binning using composition, read coverage 2016 Inining contigs using composition, read coverage 2016 Adds paired end read and coaligment information 2016 Inining counties using compulation genome 2017 Adds differential coverage to complement compose	BusyBee Web	2017	Bootstrapped supervises binning and annotation	2d interactive scatterplots supervised binning	10.1093/nar/gkx348	28472498
2017 High resolution identification of differentiall 2017 Unsupervised clustering using coverage and affi 2017 Unsupervised clustering using coverage and affi 2017 Improve genome bins through the combination of afficent binning algorithms 2016 Improved binning using composition, read coverage 2016 Improved binning contigs using composition, read coverage 2016 Adds plained and coaligment information 2017 High resolution destriction of information composition, read coverage 2018 Add estimated distribution of real genome lengths 2019 Add estimated and coaligment information 2014 Tool for automatic recovery of population genome Add estimated differential coverage to complement composition.	ICoVer	2017	Interactive visualisation tool for verification	Interactive visualisation tool	10.1186/s12859-017-1653-5"	28464793
2017 Unsupervised clustering using coverage and affi Reduce blas for high./low abundance 2017 Improve genome bins through the combination of Combination of different binning algorithms 2016 Improved binning using Fuzzy C-Means Method 2016 Improved binning configuration for composition, read coverage Adds spatied end read and configurent information 2014 Tool for automatic recovery of population genom Adds differential coverage to complement compos	HirBin*	2017	High resolution identification of differentiall	Supervised annotation, unsupervised clustering	10.1186/s12864-017-3686-6	28431529
2017 Improve genome bins through the combination of Combination of different binning algorithms 2016 Improved binning using Fuzzy C-Means Method 2016 Improved binning contigs using composition, read coverag 2014 Tool for automatic recovery of population genome Adds differential coverage to complement compos	BinSanity	2017	Unsupervised clustering using coverage and affi	Reduce bias for high/low abundance	10.7717/peerj.3035	28289564
2016 Improved binning using Fluzy C.Means Method Add estimated distribution of real genome lengths 2016 binning contigs using composition, read coverage Adds paired end read and coaligment information 2014 Tool for automatic recovery of population genom Adds differential coverage to complement compos	Binning_refinner	2017	Improve genome bins through the combination of	Combination of different binning algorithms	10.1093/bioinformatics/btx086	28186226
2016 binning contigs using composition, read coverag Adds paired end read and coaligment information 2014 Tool for automatic recovery of population genom Adds differential coverage to complement compos	IFCM add on	2016	Improved binning using Fuzzy C-Means Method	Add estimated distribution of real genome lengths	10.1109/TCBB.2016.2576452	27295684
2014 1001 for automatic recovery of population genom Adds differential coverage to complement compos	COCACOLA	2016	binning contigs using composition, read coverag	Adds paired end read and coaligment information	10.1093/bioinformatics/btw290	27256312
	GroopM (2)	2014	Tool for automatic recovery of population genom	Adds differential coverage to complement compos	10.7717/peerj.603	25289188

gene coming from a shared chromosome will display proportional abundances across samples, if
you have enough samples from a common environment you can identify the sets of genes from
a common organism of origin (MLGs Chameleon-clust 2012, CAGs and MGSs Canopy 2014,
Markovclust-MGCs Karlsson 2013, MSPs MSPminner 2018). IIIIIII HEAD

To the extent of our knwoledge, in the past few years MSPminer is the only available Software exploiting this approach. MSPminer introduced a robust proportionality measure detecting co abundant but no necessarily co ocurring. This tools groups co-abundant genes into Metagenomic Species Pan-genomes or MSPs and classify genes within an MSP as core, accessory and shared.

The factors that impact directly on MSPs quality include the sample composition, the sequencing depth, the previos bioinforamtic steps to build the reference gene dataset and to map the reads. A high number of samples with varying phenotipes improve the quality of MSPs.

MSPs can be employed for taxonomic profiles of new samples from similar ecosystems, to compare strains between samples building a presence/absence table of accessory genes and for biomarker discovery. By binning contigs carrying genes from the same MSP it is also possible to build a MAG.

Co-abundant gene binning methods perform better in large sample datasets

123 0.1 Metagenomic Species Pan-genomes

Microbial pan-genomes are gene repertoires composed of core genes present in all strains and accessory genes present in only some of them (Medini et al., 2005). In a shotgun metagenomic sequencing context, we define as shared the genes detected in some samples where the species is not present.

A strain found in a sample is an instance of the species pan-genome: it is made of all the species (shared) core genes and of a subset of (shared) accessory genes. Core genes are suitable for taxonomic profiling at species-level while accessory genes can be used to compare strains across samples. Genes tagged as shared should be used carefully as they contain false positives counts or are subject to horizontal transfer.

133 1 Conclusion

111

112

113

114

115

116

117

118

120

121

122

129

131

• New and open areas of research in which the application of metagenomic pipelines are relevant

- HMP and other
- The increased impact of machine learning in analysis
- Short section just for past-present-future completeness
- Future developments for metagenomic analysis

1.1 Weaknesses and future developments

141 Until now binning methods perform poorly in samples containing similar strains.

===== To the extent of our knwoledge, in the past few years MSPminer is the only avail-142 able Software exploiting this approach. MSPminer introduced a robust proportionality measure 143 detecting co abundant but no necesarily co ocurring. This tools groups co-abundant genes into Metagenomic Species Pan-genomes or Metagenomic Species Pan-genomess (MSPs) and classify 145 genes within an MSP as core, accessory and shared. The factors that impact directly on MSP quality include the sample composition, the sequencing depth, the previous bioinforamtic steps to 147 build the reference gene dataset and to map the reads. A high number of samples with varying phenotipes improve the quality of MSPs. MSPs can be employed for taxonomic profiles of new 149 samples from similar ecosystems, to compare strains between samples building a presence/absence 150 table of accesory genes and for biomarker discovery. By binning contigs carrying genes from the 151 same MSP it is also possible to build a MAG. 152

153 Metagenomic Species Pan-genomes

Microbial pan-genomes are gene repertoires composed of core genes present in all strains and accessory genes present in only some of them (?, ?). In a shotgun metagenomic sequencing context, we define as shared the genes detected in some samples where the species is not present. A strain found in a sample is an instance of the species pan-genome: it is made of all the species (shared) core genes and of a subset of (shared) accessory genes. Core genes are suitable for taxonomic profiling at species-level while accessory genes can be used to compare strains across samples. Genes tagged as shared should be used carefully as they contain false positives counts or are subject to horizontal transfer.

62 Conclusion

- 163 New and open areas of research in which the application of metagenomic pipelines are relevant
- The increased impact of machine learning in analysis Short section just for past-present-future
- 165 completeness Future developments for metagenomic analysis

References

- Nayfach, S., Shi, Z. J., Seshadri, R., Pollard, K. S., & Kyrpides, N. C. (2019). New insights from uncultivated genomes of the global human gut microbiome. *Nature*, 568(7753), 505–510.
- Nissen, J. N., Johansen, J., Allesøe, R. L., Sønderby, C. K., Armenteros, J. J. A., Grønbech,
- $C.\ H.,\ldots$ others (n.d.). Improved metagenome binning and assembly using deep variational
- autoencoders. Nature Biotechnology, 1–6.
- 173 Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., . . . others
- (2005). Genome analysis of multiple pathogenic isolates of streptococcus agalactiae: impli-
- cations for the microbial "pan-genome". Proceedings of the National Academy of Sciences,
- 102(39), 13950–13955.
- 177 Yue, Y., Huang, H., Qi, Z., Dou, H.-M., Liu, X.-Y., Han, T.-F., ... Tu, J. (2020). Evaluating
- metagenomics tools for genome binning with real metagenomic datasets and cami datasets.
- BMC bioinformatics, 21(1), 1–15.