UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

VINICIUS SOUSA FLORES

Recuperação e caracterização de genomas de bacteriófagos a partir de metagenomas de compostagem

Versão Corrigida
São Paulo
2023

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

VINICIUS SOUSA FLORES

Recuperação e caracterização de genomas de bacteriófagos a partir de metagenomas de compostagem

Dissertação apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática da Universidade de São Paulo para obtenção do título de Mestre em Ciências.

Área de Concentração: Bioinformática.

Orientadora: Profa. Dra. Aline Maria da Silva
Co-Orientador: Prof. Dr. João Carlos Setubal

São Paulo
2023

UNIVERSITY OF SAO PAULO
INSTITUTE OF CHEMISTRY
BIOINFORMATICS GRADUATE PROGRAM

VINICIUS SOUSA FLORES

Recovery and characterization of bacteriophage genomes from composting metagenomes

Dissertation presented to the Bioinformatics Graduate Program at the University of Sao Paulo as part of the requirements to obtain the Master of Science degree.

Concentration area: Bioinformatics

Advisor: Prof. Aline Maria da Silva
Co-Advisor: Prof. João Carlos Setubal

Sao Paulo
2023

Ficha catalográfica

Candidato: Vinicius Sousa Flores

Título: Recuperação e caracterização de genomas de bacteriófagos a partir de metagenomas de compostagem

Dissertação apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática da Universidade de São Paulo para obtenção do título de Mestre em Ciências

Aprovado em: 30/08/2023

Banca Examinadora

Examinador(a)	Nome	Alessandro de Mello Varani
	Instituição	UNESP
Examinador(a)	Nome	Prof. Dr. Henrique Ferreira
	Instituição	UNESP
Examinador(a)	Nome	Prof. Dr. Robson Francisco de Souza
	Instituição	ICB-USP
Presidente	Nome	Aline Maria da Silva
	Instituição	IQ-USP

AGRADECIMENTOS

Ante a todos que fizeram parte dessa jornada, devo agradecer primeiro minha companheira, Amanda Cerqueira. Boa parte desse trabalho é fruto de uma solidão científica que passei durante a pandemia de coronavírus, tendo conhecido bem pouco meus colegas de laboratório devido a isso. Amanda foi minha companheira nessa jornada, estando ao meu lado em todas as conquistas, dificuldades e reclamações. Meu amor, te amo muito e sou eternamente agradecido por ter você ao meu lado.

Aproveito também para agradecer meus orientadores: Profa. Dra. Aline Maria da Silva e Prof. Dr. João Carlos Setubal. Primeiro por terem me aceito como aluno no auge da pandemia e por toda a assistência, confiança e paciência que tiveram comigo durante esse percurso.

Agradeço especialmente ao Dr. Deyvid Amgarten por ter de fato oferecido esse projeto como oportunidade para meu mestrado, foi de fundamental importância para eu ter dado o passo inicial na Bioinformática. Conhecer o Deyvid também foi mérito do Henrique Iglesias e do Luís Gustavo, meus amigos e companheiros na primeira jornada que tive em um laboratório. Layla Martins, Suzana Guima, Fernando Rossi, Bruno Iha e Guillermo Uceda-Campos, também sou muito grato a vocês por toda a atenção que tiveram comigo. Também agradeço ao Carlos Piroupo pelo apoio técnico e gerenciamento dos recursos computacionais.

Não posso deixar de incluir também minha família. Meus irmãos - Juliana e Matheus - e meus pais - Joelço e Renilda - por todos os momentos que passamos juntos e a paciência e empenho de vocês por me ouvirem ao falar do meu projeto e o suporte emocional incondicional ao longo da minha graduação e pós-graduação.

Sobra espaço também para agradecer meus tijolos do RPG: Flowers, Junior, Ocarina, Shark, Nemo e Aloha. Por todas as mesas que (possivelmente) nunca vamos finalizar, meus eternos agradecimentos por estarem lá.

Por fim, agradeço a todos que fizeram parte dessa jornada. Meus mais sinceros e eternos agradecimentos.

Este trabalho foi apoiado por auxílios à pesquisa da CAPES (Projeto 3385/2013) e da FAPESP (Processo 2011/508706) que financiaram a infraestrutura computacional instalada no Laboratório de Bioinformática do IQ-USP.

“Gostaria de caminhar contigo para averiguar se olhamos as coisas da mesma forma”

A Theo van Gogh, Antuérpia, 28 de Novembro de 1885.

RESUMO

FLORES, V.S. Recuperação e caracterização de genomas de bacteriófagos a partir de metagenomas de compostagem. 2023. 60 páginas. Dissertação de Mestrado. Programa de Pós-Graduação Interunidades em Bioinformática, Universidade de São Paulo, São Paulo.

Bacteriófagos ou fagos são vírus que infectam bactérias. Esses vírus causam grande impacto nas comunidades bacterianas dos mais diversos ambientes, sendo também centrais para a regulação de ciclos biogeoquímicos dos ambientes em que se encontram. Por muito tempo a sua diversidade e papel ecológico foi subestimada. O surgimento do sequenciamento de alto desempenho acompanhado do aprimoramento das técnicas de metagenômica permitiram um maior aprofundamento no estudo da diversidade e ecologia desses vírus. A recuperação de genomas montados a partir de metagenomas (ou *Metagenome-Assembled Genomes* - MAGs) permite a inferência do possível impacto que os genomas recuperados exercem sobre o ambiente. Nos últimos anos, compêndios de genomas de fagos têm sido construídos a partir de metagenomas de variados ambientes e suas implicações ecológicas têm sido exploradas. Entretanto, pouco ainda se sabe sobre o componente viral na compostagem, um ambiente altamente dinâmico em que as comunidades de microrganismos são definidas e reguladas pelo tipo de material compostado, disponibilidade de oxigênio, umidade e temperatura. O objetivo deste trabalho foi a construção de um catálogo de genomas de fagos recuperados de metagenomas de amostras de compostagem do Parque Zoológico de São Paulo. Para tal, realizamos a reconstrução de MAGs virais (vMAGs) com foco em fagos da classe *Caudoviricetes*. Foram recuperados 401 vMAGs de metagenomas de amostras coletadas em 6 composteiras distintas. A média de tamanho destes vMAGs foi 39 kbp, variando entre 10 kbp a 600 kbp. Um vMAG foi classificado como tendo seu genoma completo e 35 vMAGs foram classificados com alto grau de completude. Cerca de 50% dos vMAGs foram atribuídos a fagos com estilo de vida virulento, porém entre os vMAGs de maior completude o estilo de vida temperado foi predominante. A análise de clusterização apontou 77% dos clusters formados integralmente por vMAGs recuperados da compostagem, sugerindo que boa parte dos genomas recuperados pertencem a novos gêneros. Genes Metabólicos Auxiliares (*Auxiliary Metabolic Genes – AMGs*) foram encontrados em 63 vMAGs, sendo ~6% de AMGs ligados ao metabolismo de carboidratos encontrados em plantas. Quinze vMAGs com AMGs foram ligados a MAGs bacterianos previamente recuperados e que possuem um papel central no metabolismo de carboidratos complexos. A análise do perfil de abundância dos vMAGs recuperados de duas das composteiras analisadas revelam um aumento gradual na abundância de fagos ao longo dos dias de compostagem.

Palavras-chave: fagos, metagenoma, compostagem, genomas recuperados de metagenomas.

ABSTRACT

FLORES, V.S. Recovery and characterization of bacteriophage genomes from composting metagenomes 2023. 60 pages. Master´s Dissertation. Bioinformatics Graduate Program, Universidade de São Paulo, São Paulo.

Bacteriophages or phages are viruses that infect bacteria. These viruses have a major impact on bacterial communities in a wide range of environments and are also central to the regulation of biogeochemical cycles in the environments in which they are found. For a long time, their diversity and ecological role were underestimated. The emergence of high-throughput sequencing, and the improvement of metagenomics techniques, have allowed a deeper study of the diversity and ecology of these viruses. The recovery of Metagenome-Assembled Genomes (MAGs) allows inferring the possible impact that the recovered genomes have on the environment. In recent years, compendia of phage genomes have been constructed from metagenomes from varied environments, and their ecological implications have been explored. However, little is known about the viral component in composting, a highly dynamic environment in which communities of microorganisms are defined and regulated by the composted material type, oxygen availability, moisture, and temperature. The aim of this work was the construction of a catalog of phage genomes retrieved from metagenomes of composting samples of the Sao Paulo Zoo Park. For that, we performed reconstruction of viral MAGs (vMAGs), focusing on phages of the class Caudoviricetes. We recovered 401 vMAGs from metagenomes of samples collected in 6 distinct composting piles. The vMAGs average size was 39 kbp ranging from 10 kbp to 600 kbp. One vMAG was classified as having its genome complete, and 35 vMAGs were classified with high degree of completeness. About 50% of the vMAGs were assigned to phages with virulent lifestyle, but among the vMAGs with the highest completeness, the temperate lifestyle was predominant. Clustering analysis indicated that 77% of the clusters were formed entirely by vMAGs recovered from composting, suggesting that most of the recovered genomes belong to new genera. Auxiliary Metabolic Genes (AMGs) were found in 63 vMAGs, with about 6% of AMGs linked to carbohydrate metabolism found in plants. Fifteen vMAGs with AMGs were linked to previously recovered bacterial MAGs that have a central role in complex carbohydrate metabolism. Analysis of the abundance profile of the vMAGs recovered from two of the analyzed composting piles reveals a gradual increase in phage abundance over the composting days.

Keywords: phage, metagenome, composting, metagenome-assembled genomes.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Viruses of Bacteria

Bacteriophages are the group of viruses which infects bacteria (HATFULL, 2021). Commonly referred to as phages, these viruses are the most abundant entities in the biosphere (DION et al., 2020) and are found in virtually every ecosystem where bacteria can thrive (WEINBAUER, 2004; CLOKIE et al., 2011). These viruses have a profound impact on bacterial communities, deeply influencing the evolutionary trajectory of bacteria through several mechanisms such as the regulation of bacterial abundance and the spread of genes (WEINBAUER, 2004; KNOWLES et al., 2016; HOWARD-VARONA et al., 2017). This impact is not restricted to bacterial communities alone. Given that bacteria play major roles in ecosystem regulation, such as carbon cycles, bacteriophages can also participate as key players in this regulation through the modulation of the structure of bacterial communities (WILHELM & SUTTLE, 1999; WEINBAUER et al., 2004; LIAO et al., 2023).

The most frequently isolated bacteriophages present double-stranded DNA (dsDNA) enclosed in an icosahedral capsid connected with a tail (DION et al., 2020). These phages belong to the class *Caudoviricetes* and have a wide range of genome sizes, varying between 10 to 800 kbp with a mean size of 40-50 kbp (MAHMOUDABADI & PHILLIPS, 2018; AL-SHAYEB et al., 2020; TURNER et al., 2023). Other classes of phages have non-tailed capsids with dsDNA genome or non-tailed capsids with single-stranded DNA (ssDNA) or RNA genomes (OFIR & SOREK, 2018).

Mosaicism is the rule of thumb for bacteriophage genome architecture, meaning that horizontal genetic exchange plays an important role in the evolution of their genomes. Consequently, different segments of bacteriophage genomes have followed different evolutionary paths (HATFULL & HENDRIX, 2011; DION et al., 2020). However, despite this pervasive mosaicism, core and non-core genes are identified in bacteriophage genomes. The difference between core and non-core genes lies in their mobility, with core genes having a lower degree of exchange within themselves compared to non-core genes. Generally, core genes are associated with important biological functions such as phage structure (i.e., virion head, tail, and fibers), replication, and lysis of the bacterial host. In contrast, non-core genes are not essential for the phage life cycle, and many of them have no assigned function. However, they can provide benefits for phage growth and multiplication, and their alteration can create

new ecological niches to be explored (HATFULL & HENDRIX, 2011; HARPER et al., 2021)

The life cycle of bacteriophages can be seen as a continuum between mandatory lytic bacteriophages (virion-productive) and obligately lysogenic ones, in which either beneficial or antagonistic interactions are possible (CORREA et al., 2021). Lytic phages hijack the bacterial cell molecular machinery, replicate their genome, create new virions, and then lyse the cell membrane to release the progeny into the environment (WEINBAUER, 2004; CLOKIE et al., 2011; CORREA et al., 2021). Temperate bacteriophages can either enter the lytic path presented before or integrate into the host genome through the lysogenic path. The lysogenic path occurs when temperate bacteriophages integrate directly into the host genome or act like a plasmid. Bacteriophages in this state are called prophages. Prophages are replicated jointly with the host genome through an undefined number of generations until environmental conditions induce them to follow a lytic path (HOWARD-VARONA et al., 2017)

Through lytic or lysogenic life cycles, bacteriophages can shape bacterial communities and the environment around them. The models 'kill-the-winner' and 'piggyback-the-winner' are two frameworks proposed to describe the ecology of bacteriophages (SILVEIRA & ROHWER, 2016; DION et al., 2020). The first model denotes the importance of lytic phages infecting the most abundant bacteria in the environment, releasing tons of organic matter fixed by the 'winner,' and opening new niches to be explored (WILHELM & SUTTLE, 1999; KNOWLES et al., 2016; CHEVALLEREAU et al., 2022). The second model is linked to the prevalence of lysogeny at high bacterial densities. Lysogenic bacteriophages can benefit the fitness of their hosts either by carrying Auxiliary Metabolic Genes (AMGs) to enhance host metabolism or Virulence Factors (VFs) that help in the competition against other bacteria (ABEDON & LEJEUNE, 2005; HURWITZ & U'REN, 2016), as well as in many other ways (DION et al., 2020).

1.2 Metagenomics and Recovery of Phage Genomes

For a long time, the study of bacteriophages relied upon culture-dependent approaches, which caused an underestimation of their biodiversity and importance in the environment (BREITBART & ROHWER, 2005). The rise of culture-independent methods such as metagenomics and metatranscriptomics coupled to robust bioinformatics approaches has largely boosted the study of the phage 'dark matter'

(CLOKIE et al., 2011; DION et al., 2020; KHOT et al., 2020; BAJIYA et al., 2022; SCHACKART et al., 2023). The culture-independent methods led to the creation of catalogues of phage genomes for a number of environments, such as marine, soil, invertebrate, and vertebrate-associated microbiomes (PAEZ-ESPINO et al., 2016; ROUX et al., 2016; GREGORY et al., 2019; DEBOUTTE et al., 2020; CAMARILLO-GUERRERO et al., 2021; NAYFACH et al., 2021b; TISZA & BUCK, 2021; CAMARGO et al., 2023). The ever-growing number of studies describing the phage 'dark matter' are unraveling the predicted phage diversity, their related evolutionary processes, and their role in the environment (DION et al., 2020; LIAO et al., 2023; SANTIAGO-RODRIGUEZ & HOLLISTER, 2022).

A strategy for recovering bacterial and phage genomes from metagenomic datasets is through binning, which results in the formation of Metagenome-Assembled Genomes (MAGs). Binning is defined as the process of grouping reads/contigs of the same biological taxon into a bin, with each bin corresponding to a MAG (PÉREZ-COBAS et al., 2020). The MAGs can be further analyzed for taxonomic classification, functional characterization, and abundance quantification, providing insights into the relationships within microbial populations in an environment and how these relationships impact the environment itself (SETUBAL, 2021; YANG et al., 2021).

1.3 Composting Process

Aerobic composting is a self-heating process that utilizes microbial metabolism to degrade organic matter. With temperatures ranging from 40 to 80 °C due to microbial activity, the composting process is divided into four phases: initial mesophilic phase, thermophilic phase, second mesophilic phase and maturation phase (RYCKEBOER, 2003). Each phase presents a different microbial composition, in which the breakdown of organic matter by a group of decomposers creates a new physicochemical environment that favors the growth of another group of decomposers (RYCKEBOER, 2003; JURADO et al., 2014). This dynamic occurs even within and between composting phases and is primarily driven by bacteria in the first three phases of composting (MARTINS et al., 2013; JURADO et al., 2014; ANTUNES et al., 2016). However, during the maturation phase, fungi emerge as the main degraders of the remaining carbon source (RYCKEBOER, 2003; PARTANEN et al., 2010).

The bacterial communities involved in mesophilic and thermophilic composting phases have been extensively studied. Their crucial role as the primary drivers of

organic matter decay during these phases has been established (LÓPEZ-GONZÁLEZ et al., 2015; ANTUNES et al., 2016; BRAGA et al., 2021a). The bacterial diversity of composting is influenced by various physical characteristics of the composting pile, including oxygen availability, moisture, pH, and temperature. Additionally, the nature of the composted material creates a multitude of niches that can be explored by different bacterial groups, and thus influencing the bacterial diversity of composting (RYCKEBOER, 2003). Therefore, as a consequence of the significant bacterial diversity resulting from a dynamic environment, composting serves as a hot spot for exploring genes with biotechnological applications (JURADO et al., 2014; SIU-RODAS et al., 2018; BUZÓN-DURÁN et al., 2020), such as lignocellulose degradation for bioethanol production (DI DONATO et al., 2019).

Despite the extensive exploration of bacteria and fungi in composting (MARTINS et al., 2013; LÓPEZ-GONZÁLEZ et al., 2015; JURADO et al., 2020; BRAGA et al., 2021a), little is known about the diversity and ecology of viruses in this environment (CHEEPUDOM et al., 2015; LIMA-JUNIOR et al., 2016; AMGARTEN et al., 2017). Recently, evidence of the significant impact that viruses have on the composting process has been reported (LIAO et al., 2023).

The São Paulo Zoo Park utilizes semi-industrial composting piles to biodegrade the organic matter generated by its facilities. ANTUNES et al (2016) conducted an extensive investigation using metagenomics and metatranscriptomics in two different composting piles (ZC3 and ZC4) from the São Paulo Zoo Park. Their research unveiled the composition of bacterial communities, characterized their ecological succession profiles, and identified their relationship with organic matter decay. BRAGA et al (2021a) further explored the same composting piles by retrieving bacterial metagenome-assembled genomes (bMAGs), deepening the work carried out by ANTUNES et al (2016) to specifically examine the bacteria involved in the composting process in these two piles. Additionally, MARTINS et al (2013) conducted a study on different composting piles (ZC1 and ZC2) at the São Paulo Zoo Park, investigating bacterial diversity. Recently, GUIMA, 2021 investigated the bacterial composition of another composting samples (ZCI and ZCM). However, the viral counterpart of these composting processes has been poorly explored (LIMA-JUNIOR et al., 2016; AMGARTEN et al., 2017), leaving a knowledge gap regarding the diversity and impact of these entities in composting environments.

2. OBJECTIVES

As mentioned above, the viral counterpart of the composting processes has been poorly explored and, until now, little is known regarding the phage diversity and their impact in composting environments. Thus, we set as goals of this project:

1. Construction of a catalog of Metagenome-Assembled Genomes from bacterial viruses (vMAGs) focused on the *Caudoviricetes* class recovered from metagenomic datasets of composting samples collected at the São Paulo Zoo Park;
2. Evaluation of the impact of phages in the composting process.

2.1 SPECIFIC OBJECTIVES

1. Characterization of composting vMAGs:
    a. Quantification of completeness;
    b. Quantification of the degree of novelty in relation to reference repositories;
    c. Functional annotation focusing on Auxiliary Metabolic Genes (AMGs), Virulence Factors (VF), structural proteins and Antibiotic Resistance Genes (ARGs);
    d. Prediction of probable family and genus;
    e. Prediction of putative host;
    f. Lifestyle categorization (virulent [lytic], temperate, prophage).
2. Impact of phages in the composting process:
    a. Identify genes related to carbohydrate degradation and virulence factors present in the recovered vMAGs;
    b. Quantify the abundance of phage genomes over the days of composting and segment by presence of auxiliary metabolic genes (AMGs), virulence factors (VF) and putative lifestyle.

3. MATERIAL AND METHODS

The methods applied in this work are summarized in Supplementary Figure S3.

3.1 Metagenomic datasets from composting samples

The metagenomic datasets used in this work derived from samples collected at the composting facility of São Paulo Zoo Park, São Paulo, Brazil. The metagenomes from samples collected from Zoo Composting piles named ZC1, ZC2, ZC3 and ZC4 were previously explored for their bacterial composition, and recovery of bacterial MAGs (MARTINS et. al., 2013; ANTUNES et. al., 2016; BRAGA et al., 2021a). The metagenomes from ZCI (Zoo Composting Inoculum) and ZCM (Zoo Composting Mature) samples have been also previously analyzed regarding their bacterial composition, including recovery of bacterial MAGs (GUIMA, 2021).

The compost piles ZC1 and ZC2 were established on 01/26/2011 and 07/27/2009, respectively and samples were collected on the day 8 (ZC1) and day 60 (ZC2) day after beginning the composting process. DNA extracted from these samples were sequenced using Roche 454 GS FLX Titanium platform (MARTINS et. al., 2013).

The compost piles ZC3 and ZC4 were established on 06/27/2011 and 05/08/2013, respectively. After the composting piles assembly, 5 samples were collected from ZC3 on days 1, 30, 64, 78 and 99. For ZC4, 9 samples were collected on days 1, 3, 7, 15, 30, 64, 67, 78 and 99. A turning procedure was carried out on day 65 for ZC3 and on day 63 for ZC4. DNA extracted from ZC3 samples was sequenced using Roche 454 GS FLX Titanium and Illumina MiSeq platforms, while ZC4 samples were sequenced with Illumina MiSeq platform.

The compost inoculum is a portion of a consolidated compost pile used to start a new pile. Inoculum samples were collected from five piles and named ZCI1, ZCI2, ZCI3, ZCI4, and ZCI5. Samples for mature compost were collected from mature piles named ZCM1 and ZCM2. Metagenomic sequencing of these samples was performed using Illumina MiSeq platform.

Supplementary Table S1 summarizes the metagenomic datasets used in this work.

3.2 Recovery of viral Metagenome-Assembled Genomes (vMAGs)

The workflow to recover vMAGs is composed of five steps: contigs assembly, multifasta dereplication, binning, phage prediction and MAGs dereplication.

3.2.1 Contigs assembly

Paired-end reads of ZC3, ZC4, ZCI and ZCM samples were assembled using SPAdes v3.15.4 (NURK et al., 2017) with the flag '–meta'. The scaffolds of ZC3 and ZC4 time-series samples were concatenated to a multifasta. Single-end reads of ZC1 and ZC2 samples were assembled using MIRA v4.9.6 (CHEVREUX, 2007) with default parameters.

3.2.2 Multifasta dereplication

Dereplication of multifasta files generated after contigs assembly were performed with dedupe from BBTools v37.96 (BUSHNELL, 2014). Contigs with the same nucleotide sequence were removed including matches in the reverse-complement strand.

3.2.3 Binning

The reads coverage was calculated using BWA-MEM v0.7.17 (LI & DURBIN, 2010). For composting sample ZC3 both paired- (Illumina) and single-end (Roche 454) reads were used. The SAM files generated were converted to BAM file format and sorted using SAMTools v1.10 (LI et al., 2009). Binning (a.k.a MAGs generation) was performed with MetaBat2 2.11.3 (KANG et al., 2019), which relies on normalized tetra-nucleotide frequency (TNF) scores and an iterative graph structure for clustering contigs and assembling MAGs. To improve binning of bacteriophage sequences, we set a minimum of 1500 bp contig length during the binning step, and bins greater than 10000 bp as the output. A seed was used to guarantee experimental replication.

3.2.4 Phage Prediction

To identify MAGs of bacteriophages (vMAGs), we retrieved the consensus between the prediction tools MARVEL v0.2 (AMGARTEN et al, 2018) and VirSorter2 v2.2.3 (GUO et al., 2021). Both tools use machine learning strategies to assign a score that determines the likelihood of the analyzed sequence being that of a bacteriophage or not. In both cases, only sequences that exceeded the cutoff score set by the tools themselves were considered. MARVEL was used with default parameters and models. The parameters '–include-groups "dsDNAphage" –provirus-off' were used with VirSorter2. The provirus prediction of VirSorter2 was deactivated for performance improvement.

3.2.5 vMAGs dereplication

Putative vMAGs generated from all the composting samples were joined and a new dereplication step was carried out. To dereplicate vMAGs a clustering approach with Average Nucleotide Identity (ANI) was applied. The ANI alignments between vMAGs were calculated using OrthoANIu v1.2 (LEE et al., 2016) and hits with less than 95% ANI identity and 90% coverage were removed. A similarity matrix was built using the ANI value of each alignment and Python3 (v3.8.10) implementation of the Markov Clustering algorithm (MCL) v0.0.6 (VAN DONGEN, 2000) was used to solve the clusters. vMAGs within a cluster were considered to be the same MAG. The vMAG with the greater length within each cluster was retrieved.

3.3 Completeness assessment

We used CheckV v0.8.1 (NAYFACH et al., 2021a) to estimate the completeness of each vMAG. This program calculates completeness based on a comparison with reference phage databases. Contigs of each vMAG were concatenated using N nucleotides between each contig before applying CheckV. The classification tiers of CheckV are consistent with the standards proposed by Minimum Information about an Uncultivated Virus Genomes - MIUViG (ROUX et al., 2019). CheckV was used with default parameters and models.

3.4 Gene prediction and functional annotation

Gene prediction and functional annotation of the vMAGs were performed using the MultiPhate2 v2.0.2 pipeline (ECALE ZHOU et al., 2019). Only consensus genes called with Phanotate (MCNAIR et al., 2019) (the primary gene caller), Prodigal (HYATT et al., 2010) and Glimmer (DELCHER et al., 2007) were assessed. Functional annotation was made using blastp v2.9.0+ with databases pVOGs, Phantome and NCBI Virus Protein, and HMMScan with pVOGs and VOGs profiles.

Structural phage proteins (i.e, capsid and tail proteins) were assessed with the software DeepCapTail (ABID & ZHANG, 2018). The amino acid sequences of genes predicted with MultiPhate2 pipeline were submitted to DeepCapTail, and sequences with more than 95% probability of being structural capsid or tail proteins were annotated.

Antibiotic resistance genes (ARGs) were assessed with the CARDS database as reference. Blastp v2.9.0+ (ALTSCHUL et al., 1990) was used to compare the amino acid sequences of vMAGs genes and proteins in CARDS (ALCOCK et al., 2023)

database (collected in February 2023). Hits with more than 80% of identity and 85% of coverage were considered as significant.

The dbCAN3 (ZHENG et al., 2023) server was used to annotate genes linked with carbohydrate metabolism. A job was submitted to dbCAN3 server giving as input the amino acid sequences of vMAGs predicted genes. Hits in concordance between 2 of the 3 tools used by the server were considered as significant.

Virulence genes were assessed with the Virulence Factor Database (VFDB; LIU et al., 2022; collected in February 2023). Hits were retrieved using blastp v2.9.0+ (ALTSCHUL et al., 1990) comparing the amino acid sequences of vMAGs genes with the protein sequences deposited in VFDB. Only hits with more than 40% identity, 70% coverage and less than 10^{-5} e-value were considered significant.

3.5 Abundance Estimation

We assessed the abundance of each vMAG within the composting piles using the module *quant_bins* of MetaWRAP v1.3.2 (URITSKIY et al., 2018). This module quantifies the average abundance of the submitted genomes using Salmon (PATRO et al., 2017) to build coverage tables. Default parameters were used to estimate the abundance of phages from composting samples ZC3, ZC4, ZCI and ZCM. The algorithm of *quant_bins* was modified to accept single-end reads in order to estimate the abundances from composting samples ZC1 and ZC2.

3.6 vMAGs taxonomic classification

The vMAGs were clustered using vContact2 v0.9.19 (BIN JANG et al., 2019). This tool utilizes gene sharing networks to group phages into clusters. Genomes within the same cluster are highly likely to belong to phages of the same genus. Furthermore, the relationships between clusters can also reflect broader taxonomic levels, such as families. The amino acid sequences of genes predicted with MultiPhate2 pipeline were submitted to vContact2 with default parameters and the reference database of Millard Lab Phage Genomes database (February, 2023) was used for clustering. All the overlaps and outlier clusters were removed from the final network. The viral clusters were analyzed with Cytoscape (SHANNON et al., 2003). vMAGs in the same cluster of reference phages were considered to share the same genus.

3.7 Comparison of vMAGs with reference phage genomes

We carried out average amino acid identity (AAI) to estimate the degree of novelty between vMAGs. CompareM v0.1.2 (PARKS, 2018) was used with default parameters to calculate AAI between vMAGs and reference phage genomes from IMG/VR (complete phage genomes only), NCBI Virus RefSeq phages, NCBI GenBank phages and Millard Lab Phage Genomes database. All databases were collected in February 2023. Hits were filtered for values greater than 80% AAI identity and 50% of orthologous fraction (OF).

3.8 vMAGs Host assignment

Three different strategies were used to predict vMAGs putative hosts. The software vHULK v2.0.0 (AMGARTEN et al., 2022) and RaFAH v0.3 (COUTINHO et al., 2021) were used with their default parameters and models to estimate host genera (both) and host species (vHULK only). vHULK uses a neural network model trained to estimate the probable host genera and species of a bacteriophage through gene annotation against the pVOGs database. RaFAH's approach is based on a random forest classifier that looks into the protein content of bacteriophages to predict their probable bacterial host genera. Also, CRISPR spacers alignments were carried out to predict host species for each vMAG. NCBI blastn v2.9.0+ was used to align all spacers in CRISPRCas++ database (collected in February 2021; POURCEL et al., 2020) with the retrieved vMAGs. The following cut-offs were used for each approach: vHULK score ≥ 0.3, entropy < 2 and energy < 5; RaFAH winner score ≥ 0.4; CRISPR spacers alignments with 100% identity (perfect matches). vMAGs contigs were concatenated using N nucleotides between each contig only for RaFAH analysis.

To link the bacterial MAGs (bMAGs) recovered from ZC3 and ZC4 composting samples (BRAGA et al., 2021a) with vMAGs recovered from these samples, three different approaches were used. First, CRISPR spacers were retrieved from bMAGs with CRT (version 1.2; BLAND et al., 2007), then their nucleotide sequences were aligned against the vMAGs with blastn v2.9.0+ (100% identity, mismatch ≤ 1, and e-value ≤ 10^{-5}). Second, shared genomic content between bMAGs and vMAGs were searched using blastn v2.9.0+ and considering the thresholds: bitscore ≥ 50, e-value ≤ 10^{-3}, identity ≥ 70% and matching length ≥ 2500 bp. Finally, for the third approach tRNA genes were searched in vMAGs using tRNAscan (version 2.0.9; CHAN et al., 2021) with bacterial/archaeal models, then the sequence similarity between the tRNA nucleotide

sequences and the bMAGs were searched using blastn v2.9.0+ (coverage ≥ 95%, identity ≥ 90% and e-value ≤ 10^{-5}).

3.9 vMAGs lifestyle prediction

Two approaches were used to predict the lifestyles of the retrieved vMAGs. CheckV (NAYFACH et al., 2021a) was used to classify vMAGs into the prophage category. CheckV classifies a high-quality (more than 90% completeness) bacteriophage into the prophage category if the viral region detected is flanked by host genes on both sides. Bacphlip (HOCKENBERRY & WILKE, 2021) was used to classify vMAGs into virulent (lytic) or temperate (lysogenic) lifestyles. This software assumes that the input genome sustains a virulent (lytic) lifestyle and changes this classification to a temperate (lysogenic) lifestyle if a pattern of specific protein domains is found. Both software were used with their default parameters and models. vMAGs classified as prophage by CheckV and virulent by Bachplip were assigned as 'uncertain'.

4. RESULTS

4.1 Recovered vMAGs

We were able to recover 1679 MAGs from contigs assembled from the metagenomes of the six composting piles (Supplementary Table S1). After applying MARVEL (AMGARTEN et al, 2018) and VirSorter2 (GUO et al, 2021), 411 vMAGs were identified. Then, after applying a final dereplication step, we retrieved 401 vMAGs. A mean of 10 contigs per vMAG, with lengths spanning from 10 to 600 kbp, were identified. Additionally, nearly 70% of them have a total length below the mean value (39 kbp; Figure 1A). However, we identified 5 vMAGs with lengths above 200 kbp, classifying these vMAGs into the categories of jumbo and mega phages (AL-SHAYEB et al., 2020).

The greatest number of vMAGs retrieved (226) came from composting pile ZC4 (Figure 1B). Since this composting pile yielded the greatest number of contigs per day sampled (Supplementary Table S1), that was expected. The second greatest source of vMAGs was the composting inoculum (ZCI) with 77 (Figure 1B), followed by 55 vMAGs recovered from composting pile ZC3. We found 1 vMAG classified as complete, and 35 vMAGs classified with 'high-quality' completeness into CheckV tiers (NAYFACH et al., 2021a). However, nearly 67% (226 vMAGs) were classified with 'low-quality' completeness (Figure 1C), suggesting that the majority of the recovered vMAGs are genome fragments.

Nearly 50% of the recovered vMAGs were assigned with virulent (i.e. lytic) lifestyle. Between the remaining vMAGs, 28% were predicted with temperate lifestyle and 4.5% (18 MAGs) were predicted as temperate but in prophage state. Looking into each composting pile, we saw a similar pattern in which virulent vMAGs was more common, followed by temperate vMAGs and, if present, lysogenic ones (Figure 1D).

The majority of vMAGs classified into ‘high-quality’ completeness (MiUViG standard) were predicted either as temperate or as integrated prophage. 28 out of 36 vMAGs with high completeness had a lifestyle prediction assigned. Out of these 28 vMAGs, 19 were assigned with temperate lifestyle and 3 were reported as prophages. It is worth mentioning that all 5 vMAGs in the category of huge phages were classified with high completeness, and 3 of them were assigned with a temperate or prophage lifestyle (Supplementary Table S2). The vMAG classified as 'Complete' in CheckV did not have a lifestyle assigned.

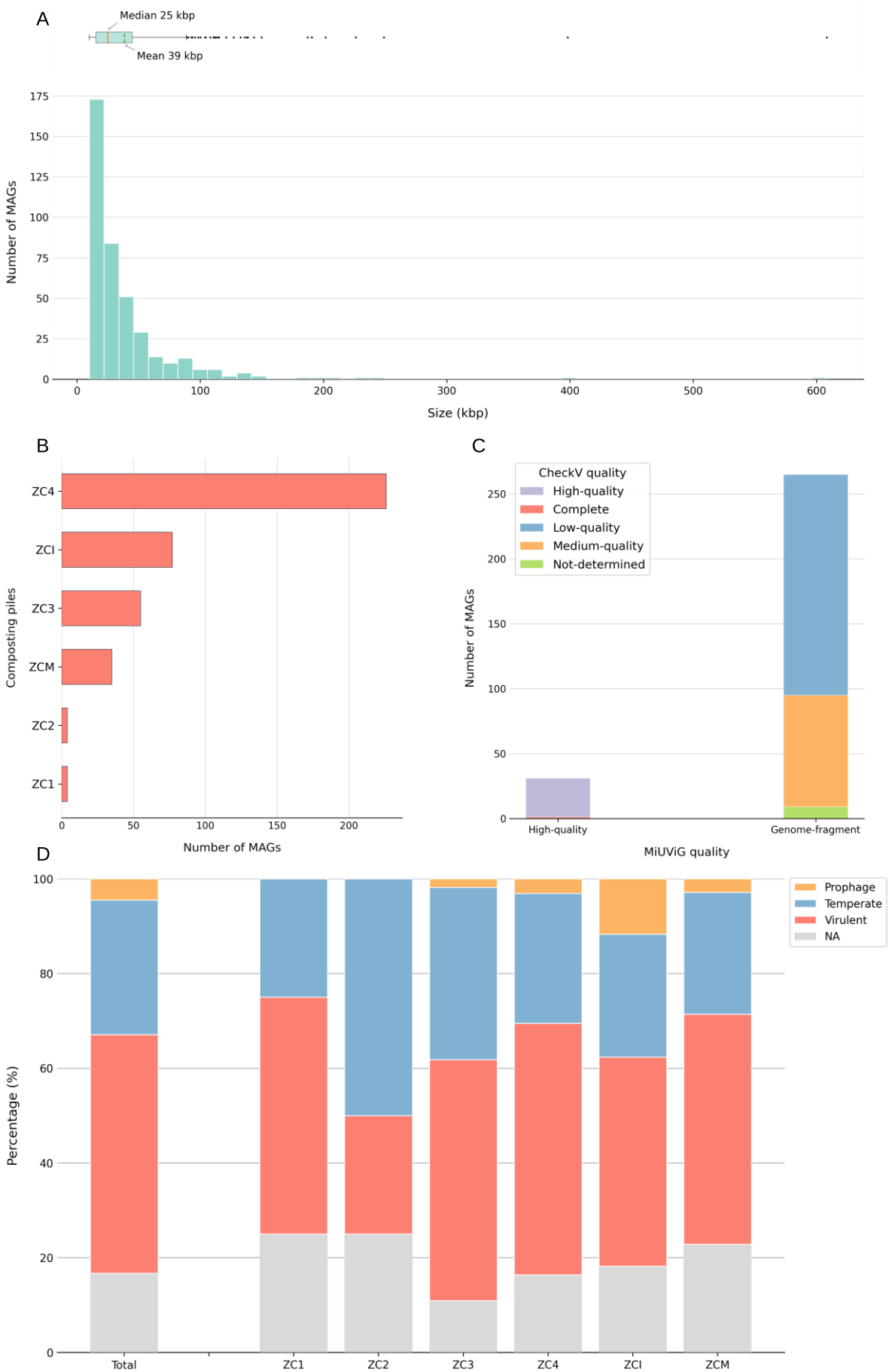

A
Median 25 kbp
Mean 39 kbp
Number of MAGs
Size (kbp)
B
Composting piles
ZC4
ZCI
ZC3
ZCM
ZC2
ZC1
Number of MAGs
C
CheckV quality
High-quality
Complete
Low-quality
Medium-quality
Not-determined
Number of MAGs
High-quality
Genome-fragment
MiUViG quality
D
Percentage (%)
Prophage
Temperate
Virulent
NA
Total
ZC1
ZC2
ZC3
ZC4
ZCI
ZCM

Figure 1: Composting phage MAGs retrieved from six different composting piles. (A) Distribution of vMAGs genome size. The length of each contig within a vMAG were summed to quantify the total length of that vMAG. Mean and median were calculated using the total length of each vMAG. (B) Number of vMAGs isolated per composting pile. (C) CheckV estimated completeness quality tiers. (D) Predicted lifestyles for recovered vMAGs**.** The lifestyles were predicted as described in Material and Methods 'vMAGs lifestyle prediction'. The first bar shows the predictions for all composting piles grouped.

4.2 Composting vMAGs novelty degree

The composting vMAGs appear to exhibit great novelty when compared with reference phage databases. Carrying out average amino acid identity alignments (AAI) comparisons between the retrieved vMAGs and reference phages, we identified only 14 vMAGs with significant alignments (Table 1). This result suggests that nearly 96% of the retrieved vMAGs are novel phage genomes.

Table 1. Composting vMAGs novelty degree. The AAI alignments were carried out as described in 'Comparison of vMAGs with reference phage genomes' in Materials and Methods section.

Databases	# of aligned MAGs	# of distinct subject sequence aligned	# of distinct aligned MAGs (%)
IMG/ VR	0	0	0 (0%)
NCBI Virus RefSeq	5	5	3 (0.75%)
NCBI GenBank	29	20	13 (3.24%)
MillardDB	29	30	13 (3.24%)
Total	77	59	14 (3.50%)

4.3 Host assignment for composting vMAGs

A total of 67 vMAGs had a putative host assigned on at least one of the approaches used (see Material and Methods 'vMAGs Host prediction'), comprising nearly 17% of the total number of recovered vMAGs. *Streptomyces* (14 vMAGs) and *Bacillus* (10 vMAGs) were the genera with the greatest number of predictions (Figure 2A). At species level, 25 vMAGs yield a putative host assignment. *Streptomyces lividans* (4 vMAGs) and *Rhodothermus marinus* (3 vMAGs) were the top two predictions at this

level (Figure 2B). Only one prediction generated consensus between genus and species level in the three approaches used, that was for ZC4vMAG372 with *Klebsiella pneumoniae* as putative host.

Bacteria of *Streptomyces* and *Bacillus* genera were reported as abundant in the composting piles ZC3 and ZC4 as in composting piles in general (RYCKEBOER et al., 2003; ANTUNES et al., 2016). In addition, *Rhodothermus marinus* was reported as the most abundant bacteria in composting ZC4 (ANTUNES et al., 2016), and 2 out of 3 predictions for these bacteria as putative hosts were assigned for vMAGs of composting pile ZC4.

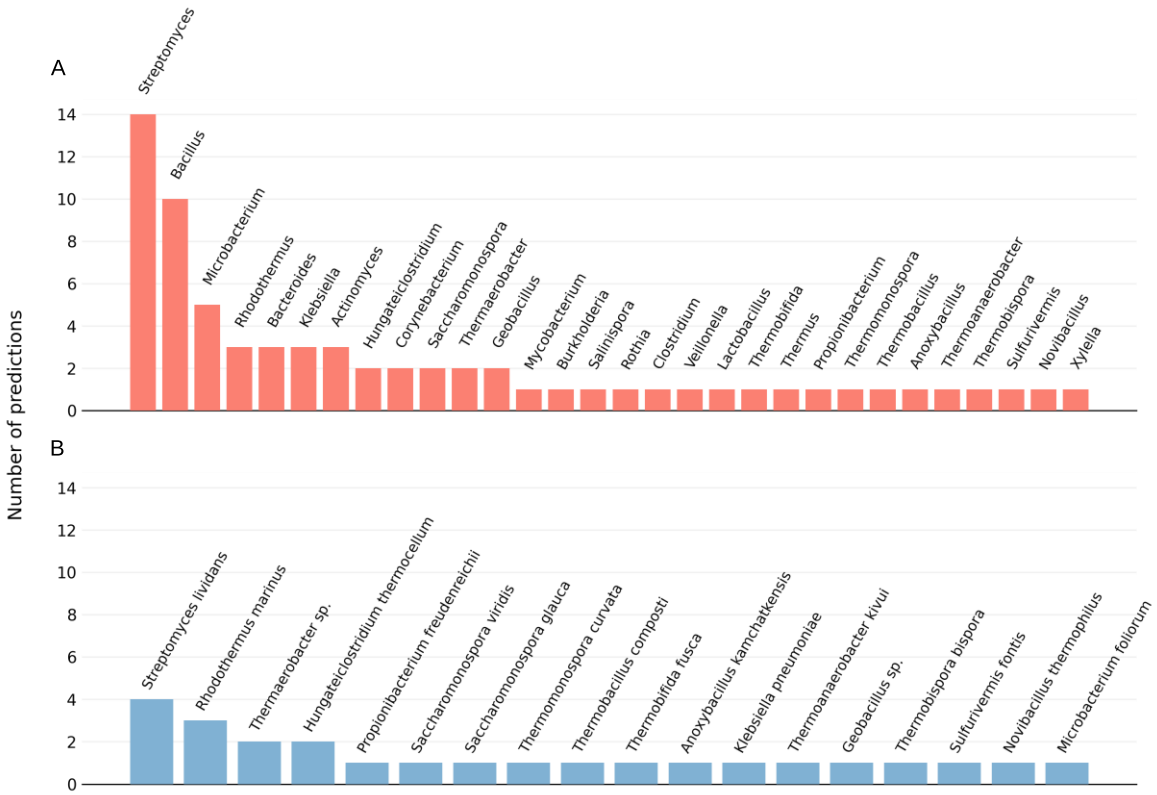


Figure 2. Host assignment for composting phage MAGs. (A) Host assignment at genus level. (B) Host assignment at species level. The predictions on this level were assigned only by vHULK and CRISPR approaches, since RaFAH classify only at genus level.

A total of 60 bacterial MAGs (bMAGs) were recovered from composting piles ZC3 and ZC4 and were stated as key players in the composting process (BRAGA et al., 2021a). We linked 65 vMAGs recovered from these composting piles with 44 of the bMAGs previously recovered. The bMAGs ZC3RG05 and ZC3RG21 were classified as

strains of *Thermobifida fusca* and yielded the greatest number (12) of assignments. We also associated vMAGs for *Thermobispora bispora* (6) and *Rhodothermus marinus* (1) bacterial MAGs. Also, we identified 15 vMAGs assigned for bMAGs of different orders, suggesting that these vMAGs may have a broad host range (Supplementary Table S3).

4.4 Taxonomic classification of viral MAGs

We retrieved 72 clusters, in which 16 clusters had at least 1 reference phage and the remaining were formed only with vMAGs. Within the 16 clusters with reference phages, only 1 was assigned with a bacteriophage family (*Drexlerviridae*). The 15 remaining clusters did not have family assignment for reference phages. A similar pattern was observed at genus level, 12 clusters had reference phages with unassigned genus, 3 clusters had reference phages from different genera, causing uncertainty to assign a genus, and only 1 cluster had reference phages from one genus (*Roufvirus*; Supplementary Figure S1).

We observed an overlap between the potential hosts predicted for vMAGs and the hosts of clustered reference phages. The predicted host for ZC1vMAG12 was *Thermobifida fusca*, and this vMAG clustered with the reference phage P1312, which indeed infects *T. fusca* (CHEEPUDOM et al., 2015). Also, the vMAGs identified as hosts of bMAGs classified as *Thermobifida fusca* (ZC3RG05 and ZC4RG21) were in the same cluster of reference phage P1312. The vMAG ZC4vMAG372 was assigned as *Klebsiella pneumoniae* as host, and the majority of reference phages in its cluster has a bacterium of *Klebsiella* genus as host. This result suggests that phages sharing the same cluster can also infect the same host.

We recovered 56 clusters formed only with composting vMAGs. Since a cluster tends to group bacteriophages from the same genus, we probably recovered at least 56 new genera of bacteriophages. Probably this number could be even greater, whereas a total of 212 vMAGs clustered, 167 distributed in the 56 clusters only with vMAGs and 45 in the 16 clusters with reference phages, leaving 189 composting vMAGs without any clue about their taxonomy.

4.5 AMGs in composting viral MAGs

A total of 118 significant alignments were found for auxiliary metabolic genes (AMGs) in a subset of 63 vMAGs (Supplementary Table S6). The great majority of these hits were for glycoside hydrolases (or GHs; 58%) and glycosyltransferases (or GTs; 33%), wherein composting pile ZC4 concentrated nearly 66% of all hits found (Figure 3A). Although the majority (85%) of AMGs did not have a putative substrate assigned, we found AMGs linked to metabolism of arabinan, xylan, beta-glucan and cellulose (6% in total), which are common carbohydrates found in plants (BUCKERIDGE, 2023). Also, 4% of hits were for alginate as substrate, which is a key component of bacterial biofilm (Figure 3B; MORADALI & BERND, 2019). In addition, we identified that between the vMAGs with AMGs, nearly 44% were assigned with a lysogenic (i.e. temperate or prophage) lifestyle and 34% having a virulent lifestyle.

We identified that 15 out of the 65 vMAGs linked to bMAGs as carriers of AMG genes. Between them we noticed vMAGs assigned to the bMAGs ZC4RG21 (2 vMAGs), ZC4RG20 (1 vMAG), ZC4RG32 (1 vMAG), ZC4RG13 (1 vMAG), ZC4RG28 (1 vMAG) and ZC4RG04 (1 vMAG), which were reported as important key-players in lignin degradation (BRAGA et al., 2021a). Alginate and beta-glucan, xylan or cellulose were 10% and 6% of the AMGs found in these 15 vMAGs, respectively. Looking into the lifestyle of these vMAGs, we observed that 5 were classified as temperate, 2 as putative prophage and only 2 as virulent. Suggesting that these viruses can potentially have an impact on the key bacterial players of composting, primarily through lysogeny.

4.6 Virulence genes in composting vMAGs

Looking for virulence factors (VFs), we found 27 significant hits distributed through 13 vMAGs (Supplementary Table S7). We identified 14 different virulence factors, in which nearly 30% of them are linked to bacterial motility (Figure 3C). We found 1 VF linked to biofilm formation through alginate production. However, since bacterial motility is an important factor in biofilm formation, probably the hits of that category are also linked to biofilm development (LIAQAT et al., 2019). In addition, we identified that 55% of the vMAGs with VF were assigned with virulent lifestyle, and the last 45% assigned as lysogenic ones. Noteworthy, we found 4 vMAGs with VFs in the categories Regulation (2 vMAGs), Stress survival (1 vMAG), Immune modulation (1 vMAG) and Adherence (1 vMAG) linked to bMAGs, and 3 of them were assigned with temperate lifestyle.

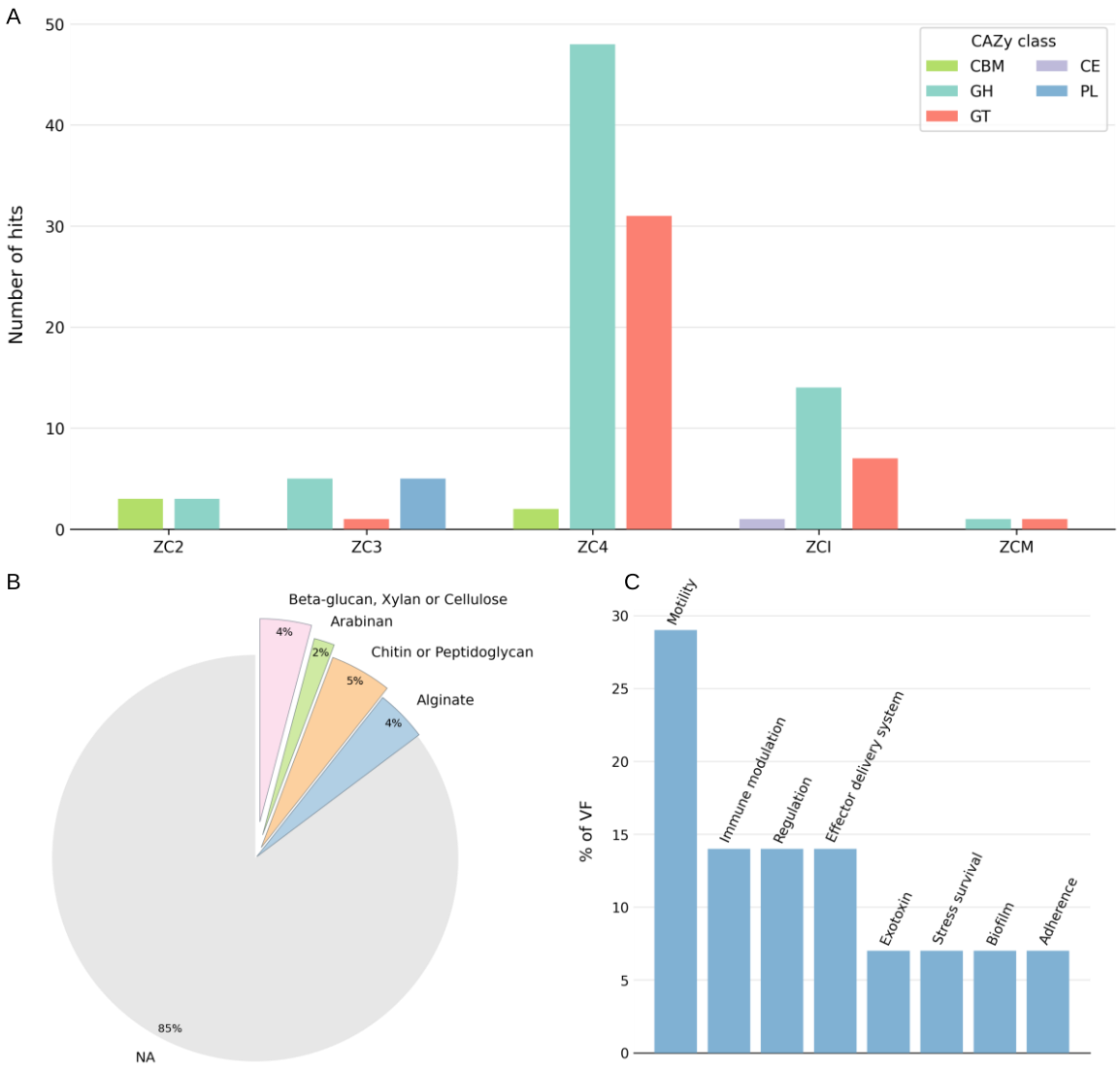


Figure 3. Auxiliary metabolic genes and virulence factors exploration. (A) Distribution of CAZymes classes through composting piles. (B) Substrate target for identified CAZymes. (C) Virulence factors category.

4.7 vMAGs Structural Proteins

We identified that more than a half of the vMAGs (242 of 401) encode a proportion of structural proteins ranging from 30% to 60% (Supplementary Figure S2) which is aligned with the proportion of structural proteins described for *Caudovirales* viruses (MAHMOUDABADI & PHILLIPS, 2018). Only 37 vMAGs showed more than 80% of their genomes composed of structural proteins.

4.8 Antibiotic resistance genes in vMAGs

No ARGs were found in the composting vMAGs with the thresholds and databases used.

4.9 Variation in vMAGs abundance over time

An increase of composting vMAGs throughout composting days was a pattern observed both in ZC3 and ZC4. Despite the differences of sampling days and number of recovered vMAGs, we observed an increase in the number of viral MAGs through the composting process for both ZC3 and ZC4, reaching the maximum after day 64 for both composting piles (Figure 4A). In addition, looking into the vMAGs assigned with a putative lifestyle, we noticed that virulent (i.e. lytic bacteriophages) were more abundant than those classified as temperate or prophage for the majority of days sampled (Figure 4B). This was not observed for ZC1, ZC2, ZCI and ZCM (Supplementary Table S4).

We noticed a variation in the abundance of vMAGs with auxiliary metabolic genes (AMGs) and Virulence Factors (VFs) throughout the composting days of ZC4. vMAGs with AMGs maintain a similar abundance through composting days, except for days 30 and 99 in which a decrease was observed. vMAGs with VF displayed low abundance across all composting days, except for day 30, during which ZC4vMAG696 exhibited relative abundance of nearly 27% (Figure 4C-D). Abundant vMAGs with AMGs and VF were identified in composting ZC3 and ZCI samples. Only vMAGs with AMGs were identified in ZC2 and ZCM samples (Supplementary Table S5).

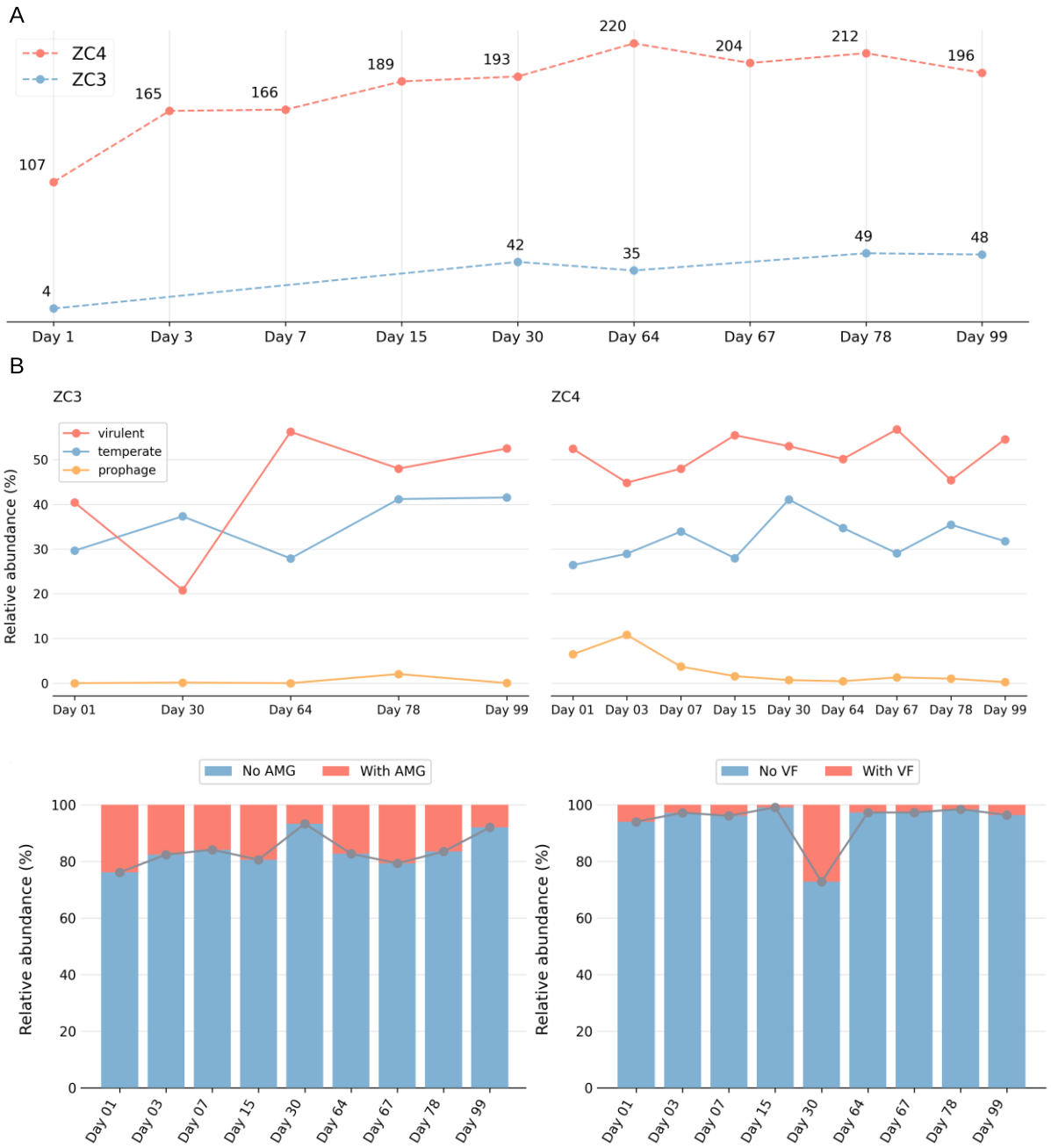


Figure 4: Relative abundance of phages genomes throughout composting process in ZC3 and ZC4 piles. (A) Number of vMAGs observed in each composting day. (B) Relative abundance of vMAGs stratified by putative lifestyle. (C) Relative abundance of ZC4 vMAGs with and without auxiliary metabolic genes (AMG). (D) Relative abundance of ZC4 vMAGs with and without virulence factors (VF).

5. DISCUSSION

The primary focus of previous studies conducted on the metagenomes of composting samples from the São Paulo Zoo Park has been on the bacterial component while the viral counterpart was poorly explored (MARTINS et al., 2013; ANTUNES et al., 2016; LIMA-JUNIOR et al., 2016; AMGARTEN et al., 2017; BRAGA et al., 2021a; GUIMA, 2021). Although the importance of viruses on the composting process has been recently reported (LIAO et al., 2023), still there a significant gap in our understanding of the viral diversity and ecological role within composting processes.

Here we investigated the phage composition on the metagenomes datasets from the São Paulo Zoo Park composting, using viral metagenome-assembled genomes (vMAGs), with special focus on viruses of the class *Caudoviricetes*. A total of 401 vMAGs were recovered from the six composting piles investigated, with nearly 9% of the vMAGs exhibiting high genome completeness. In line with recent studies where the vast majority of phage genomes recovered from environmental samples exhibited poor completeness (CAMARILLO-GUERRERO et al., 2021; NAYFACH et al., 2021b; LIAO et al., 2023), almost 67% of the retrieved vMAGs displayed 'low-quality' completeness. Despite the lack of completeness, the recovered vMAGs displayed a high degree of novelty, with nearly 96% of them lacking a similar phage genome in the reference datasets searched, which is consistent with the degree of novelty reported in other studies (CAMARILLO-GUERRERO et al., 2021; NAYFACH et al., 2021b; DANKO et al., 2021). In addition, we identified a possible set of 56 new phage genera based on vContact2 clustering approach. Therefore, we assert that the vMAG dataset presented in this study serves as a significant resource for exploring the viral biodiversity of composting.

Virulent lifestyle (i.e. lytic) were identified as the most common strategy (50.37%) between all the vMAGs recovered. However, investigating the vMAGs with high completeness and assigned lifestyle, we observed that approximately 78% of them were classified as having a lysogenic lifestyle (i.e. temperate or prophage). Considering that the majority of the vMAGs exhibited a low degree of completeness and the tools used relies upon the identification of important genome features of lysogenic phages to classify them (NAYFACH et al., 2021a; HOCKENBERRY & WILKE, 2021), these results indicate a possible underestimation of lysogenic lifestyle between all vMAGs recovered. However, assuming we have identified almost all lysogenic vMAGs their abundance in the composting piles is not so low. This is evident in composting piles ZC3 and ZC4,

where we observed a consistent presence of temperate phages throughout the composting process (Figure 4B). Additionally, in the remaining composting piles, lysogenic phages were also found to be abundant (Supplementary Table S5). Furthermore, lysogeny is known to be favored under conditions of high temperature, pH, and nutrient depletion (HOWARD-VARONA et al., 2017; DION et al., 2020), like those observed in the composting samples examined (ANTUNES et al., 2016). In addition, a recent study on phages from high-temperature composting revealed the prevalence of virulent bacteriophages and, similar to our observations, the majority of the recovered phages exhibited a low degree of completeness and were classified as virulent due to the lack of information about lysogeny genes (LIAO et al., 2023). Thus, there is a possibility that the number of lysogenic bacteriophages in composting is currently being underestimated due to the lack of completeness. However, despite the potential misclassification, our observations indicate a significant abundance of lysogenic vMAGs. This suggests the possibility that bacteriophages in composting exert an influence on the bacterial community through various lysogeny mechanisms, as observed in other environments (ABEDON & LEJEUNE, 2005; HURWITZ & U'REN, 2016; DION et al., 2020).

The genera *Streptomyces* and *Bacillus* are commonly found in composting environments (RYCKEBOER et al., 2003; PARTANEN et al., 2010) and were identified as abundant in the same set of composting samples analyzed in this study (ANTUNES et al., 2016; BRAGA et al., 2021a; GUIMA, 2021). Furthermore, both groups have species that are capable of degrading lignin and other complex polysaccharides (SIU-RODAS et al., 2018; BUZÓN-DURÁN et al., 2020; HEMATI et al., 2021). We recovered vMAGs assigned with putative host bacteria from these genera, and more specifically, we also identified vMAGs infecting *Streptomyces lividans* and *Rhodothermus marinus*. The latter is particularly abundant in composting ZC4 (ANTUNES et al., 2016) and reported as an important player in decaying organic matter environments (LYKIDIS et al., 2007). Bacterial MAGs (bMAGs) were recovered from composting piles ZC3 and ZC4, and their importance in the lignocellulose breakdown were identified (BRAGA et al., 2021a). We identified vMAGs recovered from composting piles ZC3 and ZC4 as putative hosts of the bMAGs previously recovered, including bMAGs associated to lignocellulose decay such as ZC4RG20 (Gammaproteobacteria), ZC4RG21 (*Thermobifida fusca*) and ZC4RG32 (*Caldicoprobacter oshimai*) (BRAGA et al 2021a).

We found AMGs linked to different CAZyme classes (DRULA et al., 2022) within the vMAGs, and the most abundant class found was Glycosyl-Hydrolases (GH), which is linked to carbohydrate catabolism. Digging into the putative substrate of the identified AMGs, we found AMGs associated with carbohydrates commonly found in plants, such as arabinan, xylan, beta-glucan, and cellulose (BUCKERIDGE, 2023). Moreover, as mentioned above, we discovered vMAGs whose putative hosts were identified as the primary degraders of lignocellulose, and these vMAGs carry AMGs linked to these polysaccharides. Finally, looking into the putative lifestyle of all vMAGs carrying AMGs, we observed that nearly 44% of them were classified as temperate or prophage. Since it was shown that bacteria are the major degraders in the analyzed composting piles, particularly in the metabolism of complex carbohydrates such as lignin and cellulose (ANTUNES et al., 2016; BRAGA et al., 2021a), and AMGs encoded by phages can enhance host metabolism and impact the environmental dynamics (THOMPSON et al., 2011; ROSENWASSER et al., 2016; BRAGA et al., 2021b), the role of bacteriophages in the composting process may be even greater than expected. This role can still be mediated through lysogenic phages in a similar way to the 'piggy-back the winner' model (KNOWLES et al., 2016; DION et al., 2020), as the majority of vMAGs with AMGs were classified as temperate or prophage. Additionally, as reported, it is likely that we are underestimating the numbers of lysogenic phages, and finally, vMAGs with AMGs maintain high abundance throughout the composting days, as demonstrated in the case of composting pile ZC4.

In summary, this study has successfully created the first comprehensive catalogue of viral MAGs in different composting phases, with a specific focus on viruses belonging to the class *Caudoviricetes*. The results shown starts to shed light on the biodiversity and ecology of bacteriophages in the composting process, and the dataset produced serves as a valuable resource to mine genes with biotechnological and medical application. The findings presented in this work contribute significantly to our understanding of viral dynamics in composting systems and pave the way for further research in this field.

6. REFERENCES

1. ABEDON, Stephen T.; LEJEUNE, Jeffrey T. Why bacteriophage encode exotoxins and other virulence factors. **Evolutionary Bioinformatics**, v. 1, p. 117693430500100001, 2005.
2. ABID, Dhoha; ZHANG, Liqing. DeepCapTail: a deep learning framework to predict capsid and tail proteins of phage genomes. **BioRxiv**, p. 477885, 2018.
3. AL-SHAYEB, Basem et al. Clades of huge phages from across Earth's ecosystems. **Nature**, v. 578, n. 7795, p. 425-431, 2020.
4. ALCOCK, Brian P. et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. **Nucleic Acids Research**, v. 51, n. D1, p. D690-D699, 2023.
5. ALTSCHUL, Stephen F. et al. Basic local alignment search tool. **Journal of molecular biology**, v. 215, n. 3, p. 403-410, 1990.
6. AMGARTEN, Deyvid et al. MARVEL, a tool for prediction of bacteriophage sequences in metagenomic bins. **Frontiers in genetics**, v. 9, p. 304, 2018.
7. AMGARTEN, Deyvid et al. Three novel Pseudomonas phages isolated from composting provide insights into the evolution and diversity of tailed phages. **BMC genomics**, v. 18, p. 1-18, 2017.
8. AMGARTEN, Deyvid et al. vHULK, a New Tool for Bacteriophage Host Prediction Based on Annotated Genomic Features and Neural Networks. **PHAGE**, v. 3, n. 4, p. 204-212, 2022.
9. ANTUNES, Luciana Principal et al. Microbial community structure and dynamics in thermophilic composting viewed through metagenomics and metatranscriptomics. **Scientific reports**, v. 6, n. 1, p. 38915, 2016.
10. BAJIYA, Nisha et al. Advances in the field of phage-based therapy with special emphasis on computational resources. **Briefings in Bioinformatics**, v. 24, n. 1, p. bbac574, 2023.
11. BIN JANG, Ho et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. **Nature biotechnology**, v. 37, n. 6, p. 632-639, 2019.
12. BITENCOURT, Ana Luisa V. et al. Core sampling test in large-scale compost cells for microorganism isolation. **African Journal of Microbiology Research**, v. 4, n. 15, p. 1631-1634, 2010.

13. BLAND, Charles et al. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. **BMC bioinformatics**, v. 8, n. 1, p. 1-8, 2007.
14. BRAGA, Lucas Palma Perez et al. Genome-resolved metagenome and metatranscriptome analyses of thermophilic composting reveal key bacterial players and their metabolic interactions. **BMC genomics**, v. 22, p. 1-19, 2021a.
15. BRAGA, Lucas Palma Perez et al. Novel virocell metabolic potential revealed in agricultural soils by virus-enriched soil metagenome analysis. **Environmental microbiology reports**, v. 13, n. 3, p. 348–354, jun. 2021b.
16. BREITBART, Mya; ROHWER, Forest. Here a virus, there a virus, everywhere the same virus?. **Trends in microbiology**, v. 13, n. 6, p. 278-284, 2005.
17. BUCKERIDGE, Marcos S. The diversity of plant carbohydrate hydrolysis in nature and technology. **Polysaccharide Degrading Biocatalysts**, p. 55-74, 2023.
18. BUSHNELL, Brian. BBTools Software Package. SourceForge, http://sourceforge.net/projects/bbmap, 2014.
19. BUZÓN-DURÁN, Laura et al. Applications of Streptomyces spp. enhanced compost in sustainable agriculture. **Biology of composts**, p. 257-291, 2020.
20. CAMARGO, Antonio Pedro et al. IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. **Nucleic acids research**, v. 51, n. D1, p. D733-D743, 2023.
21. CAMARILLO-GUERRERO, Luis F. et al. Massive expansion of human gut bacteriophage diversity. **Cell**, v. 184, n. 4, p. 1098-1 109. e9, 2021.
22. CHAN, Patricia P. et al. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. **Nucleic acids research**, v. 49, n. 16, p. 9077-9096, 2021.
23. CHEEPUDOM, Jatuporn et al. Isolation, characterization, and complete genome analysis of P1312, a thermostable bacteriophage that infects Thermobifida fusca. **Frontiers in microbiology**, v. 6, p. 959, 2015.
24. CHEVALLEREAU, Anne et al. Interactions between bacterial and phage communities in natural environments. **Nature Reviews Microbiology**, v. 20, n. 1, p. 49-62, 2022.
25. CHEVREUX, Bastien. MIRA: an automated genome and EST assembler. 2007.

26. CLOKIE, Martha RJ et al. Phages in nature. **Bacteriophage**, v. 1, n. 1, p. 31-45, 2011.
27. CORREA, Adrienne MS et al. Revisiting the rules of life for viruses of microorganisms. **Nature Reviews Microbiology**, v. 19, n. 8, p. 501-513, 2021.
28. COUTINHO, Felipe Hernandes et al. RaFAH: Host prediction for viruses of Bacteria and Archaea based on protein content. **Patterns**, v. 2, n. 7, p. 100274, 2021.
29. DANKO, David et al. A global metagenomic map of urban microbiomes and antimicrobial resistance. **Cell**, v. 184, n. 13, p. 3376-3393. e17, 2021.
30. DEBOUTTE, Ward et al. Honey-bee–associated prokaryotic viral communities reveal wide viral diversity and a profound metabolic coding potential. **Proceedings of the National Academy of Sciences**, v. 117, n. 19, p. 10511-10519, 2020.
31. DELCHER, Arthur L. et al. Identifying bacterial genes and endosymbiont DNA with Glimmer. **Bioinformatics**, v. 23, n. 6, p. 673-679, 2007.
32. DI DONATO, Paola et al. The production of second generation bioethanol: The biotechnology potential of thermophilic bacteria. **Journal of cleaner production**, v. 233, p. 1410-1417, 2019.
33. DION, Moïra B.; OECHSLIN, Frank; MOINEAU, Sylvain. Phage diversity, genomics and phylogeny. **Nature Reviews Microbiology**, v. 18, n. 3, p. 125-138, 2020.
34. ECALE ZHOU, Carol L. et al. multiPhATE: bioinformatics pipeline for functional annotation of phage isolates. **Bioinformatics**, v. 35, n. 21, p. 4402-4404, 2019.
35. GUIMA, Suzana Eiko Sato. **Microbial composition of inoculum and mature compost in the São Paulo Zoo composting process assessed through metagenomics**. 2021. Dissertação (Mestrado em Bioinformática) - Bioinformática, Universidade de São Paulo, São Paulo, 2021. doi:10.11606/D.95.2021.tde-16022021-122522. Acesso em: 2023-08-09.
36. GUO, Jiarong et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. **Microbiome**, v. 9, p. 1-13, 2021.
37. GREGORY, Ann C. et al. Marine DNA viral macro-and microdiversity from pole to pole. **Cell**, v. 177, n. 5, p. 1109-1123. e14, 2019.
38. HARPER, David R. et al. (Ed.). Bacteriophages: biology, technology, therapy. **Springer Nature**, 2021.

39. HATFULL, Graham F.; HENDRIX, Roger W. Bacteriophages and their genomes. **Current opinion in virology**, v. 1, n. 4, p. 298-303, 2011.
40. HATFULL, Graham F. Bacteriophage discovery and genomics. **Bacteriophages: Biology, Technology, Therapy**, p. 219-230, 2021.
41. HEMATI, Arash et al. Role of lignin and thermophilic lignocellulolytic bacteria in the evolution of humification indices and enzymatic activities during compost production. **Waste Management**, v. 119, p. 122-134, 2021.
42. HOCKENBERRY, Adam J.; WILKE, Claus O. BACPHLIP: predicting bacteriophage lifestyle from conserved protein domains. **PeerJ**, v. 9, p. e11396, 2021.
43. HOWARD-VARONA, Cristina et al. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. **The ISME journal**, v. 11, n. 7, p. 1511-1520, 2017.
44. HURWITZ, Bonnie L.; U'REN, Jana M. Viral metabolic reprogramming in marine ecosystems. **Current opinion in microbiology**, v. 31, p. 161-168, 2016.
45. HYATT, Doug et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. **BMC bioinformatics**, v. 11, n. 1, p. 1-11, 2010.
46. JURADO, Macarena et al. Exploiting composting biodiversity: study of the persistent and biotechnologically relevant microorganisms from lignocellulose-based composting. **Bioresource technology**, v. 162, p. 283-293, 2014.
47. JURADO, Macarena M. et al. Integral approach using bacterial microbiome to stabilize municipal solid waste. **Journal of environmental management**, v. 265, p. 110528, 2020.
48. KANG, Dongwan D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. **PeerJ**, v. 7, p. e7359, 2019.
49. KHOT, Varada; STROUS, Marc; HAWLEY, Alyse K. Computational approaches in viral ecology. **Computational and Structural Biotechnology Journal**, v. 18, p. 1605-1612, 2020.
50. KNOWLES, B. et al. Lytic to temperate switching of viral communities. **Nature**, v. 531, n. 7595, p. 466-470, 2016.
51. LEE, Imchang et al. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. **International journal of systematic and evolutionary microbiology**, v. 66, n. 2, p. 1100-1103, 2016.

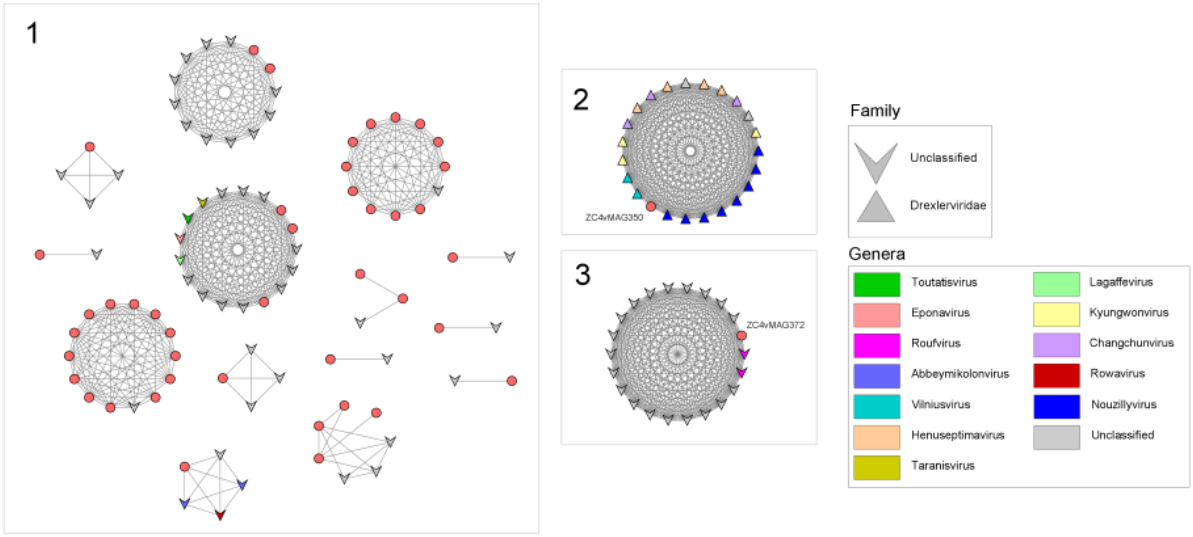
52. LI, Heng; DURBIN, Richard. Fast and accurate short read alignment with Burrows–Wheeler transform. **bioinformatics**, v. 25, n. 14, p. 1754-1760, 2009.
53. LI, Heng et al. The sequence alignment/map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078-2079, 2009.
54. LIAO, Hanpeng et al. Mesophilic and thermophilic viruses are associated with nutrient cycling during hyperthermophilic composting. **The ISME Journal**, p. 1-15, 2023.
55. LIAQAT, I. et al. Biofilm formation, maturation and prevention: a review. **J Bacteriol Mycol**, v. 6, n. 1, p. 1092, 2019.
56. LIMA-JUNIOR, James Daltro et al. Characterization of mycobacteria and mycobacteriophages isolated from compost at the São Paulo Zoo Park Foundation in Brazil and creation of the new mycobacteriophage Cluster U. **BMC microbiology**, v. 16, p. 1-15, 2016.
57. LIU, Bo et al. VFDB 2022: a general classification scheme for bacterial virulence factors. **Nucleic acids research**, v. 50, n. D1, p. D912-D917, 2022.
58. LÓPEZ-GONZÁLEZ, J. A. et al. Dynamics of bacterial microbiota during lignocellulosic waste composting: studies upon its structure, functionality and biodiversity. **Bioresource technology**, v. 175, p. 406-416, 2015.
59. MAHMOUDABADI, Gita; PHILLIPS, Rob. A comprehensive and quantitative exploration of thousands of viral genomes. **Elife**, v. 7, p. e31955, 2018.
60. MARTINS, Layla Farage et al. Metagenomic analysis of a tropical composting operation at the São Paulo Zoo Park reveals diversity of biomass degradation functions and organisms. **PloS one**, v. 8, n. 4, p. e61928, 2013.
61. MCNAIR, Katelyn et al. PHANOTATE: a novel approach to gene identification in phage genomes. **Bioinformatics**, v. 35, n. 22, p. 4537-4542, 2019.
62. MORADALI, M. Fata; REHM, Bernd HA. The role of alginate in bacterial biofilm formation. **Extracellular sugar-based biopolymers matrices**, p. 517-537, 2019.
63. NAYFACH, Stephen et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. **Nature biotechnology**, v. 39, n. 5, p. 578-585, 2021a.
64. NAYFACH, Stephen et al. Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. **Nature microbiology**, v. 6, n. 7, p. 960-970, 2021b.

65. NURK, Sergey et al. metaSPAdes: a new versatile metagenomic assembler. **Genome research**, v. 27, n. 5, p. 824-834, 2017.
66. OFIR, Gal; SOREK, Rotem. Contemporary phage biology: from classic models to new insights. **Cell**, v. 172, n. 6, p. 1260-1270, 2018.
67. PAEZ-ESPINO, David et al. Uncovering Earth's virome. **Nature**, v. 536, n. 7617, p. 425-430, 2016.
68. PARKS, Donovan. "CompareM." GitHub, github.com/dparks1134/CompareM, 2018.
69. POURCEL, Christine et al. CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers. **Nucleic acids research**, v. 48, n. D1, p. D535-D544, 2020.
70. PARTANEN, Pasi et al. Bacterial diversity at different stages of the composting process. **BMC microbiology**, v. 10, n. 1, p. 1-11, 2010.
71. PATRO, Rob et al. Salmon provides fast and bias-aware quantification of transcript expression. **Nature methods**, v. 14, n. 4, p. 417-419, 2017.
72. PÉREZ-COBAS, Ana Elena; GOMEZ-VALERO, Laura; BUCHRIESER, Carmen. Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. **Microbial genomics**, v. 6, n. 8, 2020.
73. ROUX, Simon et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. **Nature**, v. 537, n. 7622, p. 689-693, 2016.
74. ROUX, Simon et al. Minimum information about an uncultivated virus genome (MIUViG). **Nature biotechnology**, v. 37, n. 1, p. 29-37, 2019.
75. ROSENWASSER, Shilo et al. Virocell metabolism: metabolic innovations during host–virus interactions in the ocean. **Trends in microbiology**, v. 24, n. 10, p. 821-832, 2016.
76. RYCKEBOER, Jaak et al. A survey of bacteria and fungi occurring during composting and self-heating processes. **Annals of microbiology**, v. 53, n. 4, p. 349-410, 2003.
77. SANTIAGO-RODRIGUEZ, Tasha M.; HOLLISTER, Emily B. Unraveling the viral dark matter through viral metagenomics. **Frontiers in Immunology**, v. 13, p. 1005107, 2022.

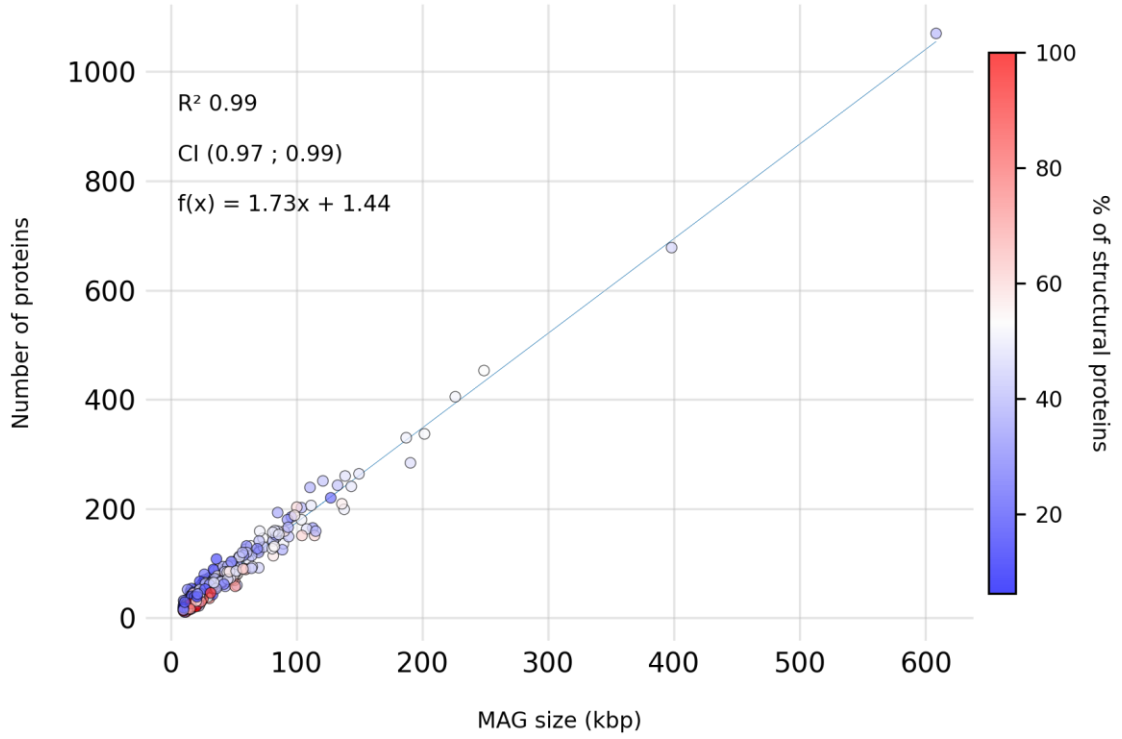
78. SCHACKART III, Kenneth E. et al. Evaluation of computational phage detection tools for metagenomic datasets. **Frontiers in Microbiology**, v. 14, p. 1078760, 2023.
79. SETUBAL, João C. Metagenome-assembled genomes: concepts, analogies, and challenges. **Biophysical Reviews**, v. 13, n. 6, p. 905-909, 2021.
80. SHANNON, Paul et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. **Genome research**, v. 13, n. 11, p. 2498-2504, 2003.
81. SILVEIRA, Cynthia B.; ROHWER, Forest L. Piggyback-the-Winner in host-associated microbial communities. **npj Biofilms and Microbiomes**, v. 2, n. 1, p. 1-5, 2016.
82. SIU-RODAS, Yadira et al. Bacillus subtilis with endocellulase and exocellulase activities isolated in the thermophilic phase from composting with coffee residues. **Revista argentina de microbiología**, v. 50, n. 3, p. 234-243, 2018.
83. TURNER, Dann et al. Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee. **Archives of Virology**, v. 168, n. 2, p. 74, 2023.
84. THOMPSON, Luke R. et al. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. **Proceedings of the National Academy of Sciences**, v. 108, n. 39, p. E757-E764, 2011.
85. TISZA, Michael J.; BUCK, Christopher B. A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases. **Proceedings of the National Academy of Sciences**, v. 118, n. 23, p. e2023202118, 2021.
86. URITSKIY, Gherman V.; DIRUGGIERO, Jocelyne; TAYLOR, James. MetaWRAP - a flexible pipeline for genome-resolved metagenomic data analysis. **Microbiome**, v. 6, n. 1, p. 1-13, 2018.
87. VAN DONGEN, Stijn Marinus. **Graph clustering by flow simulation**. 2000. Tese de Doutorado.
88. WEINBAUER, Markus G. Ecology of prokaryotic viruses. **FEMS microbiology reviews**, v. 28, n. 2, p. 127-181, 2004.
89. WILHELM, Steven W.; SUTTLE, Curtis A. Viruses and nutrient cycles in the sea: viruses play critical roles in the structure and function of aquatic food webs. **Bioscience**, v. 49, n. 10, p. 781-788, 1999.

90. YANG, Chao et al. A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data. **Computational and Structural Biotechnology Journal**, v. 19, p. 6301-6314, 2021.
91. ZHENG, Jinfang et al. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. **Nucleic Acids Research**, p. gkad328, 2023.

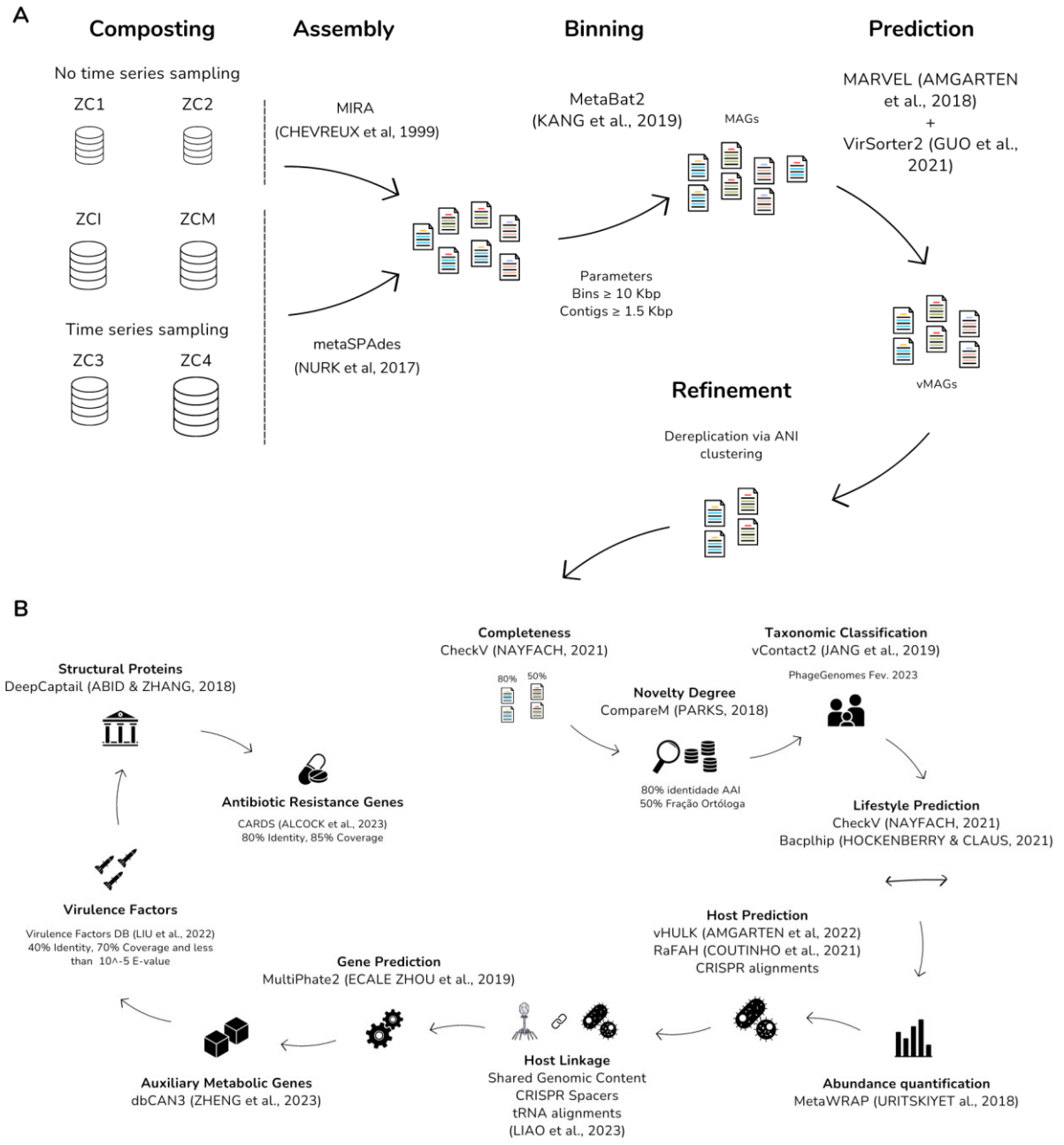
7. SUPPLEMENTARY INFORMATION



Supplementary Figure S1. Clusters of vMAGs with reference phages from Millard Lab dataset. The clusters between vMAGs and reference phages were assembled as described in section MAGs taxonomic classification in Material and Methods. (1) Clusters between vMAGs (circles) and reference phages with unassigned family. (2) Cluster with reference phages from *Drexlerviridae* family. (3) Cluster with reference phages from genus *Roufvirus*.



Supplementary Figure S2. Relationship between the percentage of structural proteins and size of vMAGs. Each circle in the figure represents a vMAG and the color gradient reports the proportion of structural proteins. The regression line, its equation, confidence interval and correlation (R^2) are presented.



Supplementary Figure S3. Summary of the methods and characterizations applied. (A) Pipeline used to recover vMAGs from metagenome samples. (B) Methods applied to characterize the vMAGs recovered. The methods applied are described in Material and Methods.

Supplementary Table S1. Number of reads and assembled contigs from samples of each composting pile. Column ‘Sampled period’ indicates the day or month of collection after compost pile assembly (Day 0). The total number of reads and reads mean length after quality control are shown in columns 'Total reads' and 'Reads mean length'. * Samples with Illumina / Roche 454 reads.

Composting pile	Sampled period	Total reads	Reads mean length	Contigs assembled
ZC1	Day 8	3,167,044	276	52,953
ZC2	Day 60	2,966,244	299	52,182
ZC3*	Day 01	2,270,264 / 520,074	186 / 452	122,695 / 9,913
	Day 30	2,145,650 / 737,772	186 / 467	99,499 / 13,506
	Day 64	1,265,240 / 771,427	225 / 484	114,638 / 16,933
	Day 78	2,169,689 / 1,063,197	225 / 471	250,005 / 20,220
	Day 99	1,509,658 / 711,081	232 / 466	172,812 / 15,796
ZC4	Day 01	8,213,864	212	369,035
	Day 03	9,340,402	226	428,524
	Day 07	9,113,124	178	234,902
	Day 15	14,438,720	170	348,551
	Day 30	9,653,864	193	357,696
	Day 64	14,469,070	204	602,051

	Day 67	8,369,838	185	253,626
	Day 78	22,490,204	148	433,352
	Day 99	16,647,048	167	494,831
ZCI	Day 85	2,438,268	181	66,317
	Day 87	2,818,406	186	132,474
	Day 64	2,051,854	206	55,375
	Day 84	2,966,140	205	97,977
	Day 82	1496682	76	802
ZCM	4 months	7,359,702	196	144,694
	2 months	6,238,624	191	140,984

Supplementary Table S2. Predicted lifestyle of vMAGs with high-quality completeness

vMAG	Lifestyle	Composting pile
ZC3vMAG114	temperate	ZC3
ZC3vMAG177	temperate	ZC3
ZC3vMAG198	temperate	ZC3
ZC3vMAG38	temperate	ZC3
ZC3vMAG78	temperate	ZC3

ZC3vMAG82	temperate	ZC3
ZC4vMAG132	virulent	ZC4
ZC4vMAG143	temperate	ZC4
ZC4vMAG190	temperate	ZC4
ZC4vMAG326	temperate	ZC4
ZC4vMAG358	virulent	ZC4
ZC4vMAG414	prophage	ZC4
ZC4vMAG417	temperate	ZC4
ZC4vMAG486	temperate	ZC4
ZC4vMAG490	temperate	ZC4
ZC4vMAG515	temperate	ZC4
ZC4vMAG548	temperate	ZC4
ZC4vMAG603	virulent	ZC4
ZC4vMAG612	virulent	ZC4
ZC4vMAG721	virulent	ZC4
ZC4vMAG737	temperate	ZC4
ZC4vMAG803	temperate	ZC4
ZC4vMAG837	temperate	ZC4
ZC4vMAG884	prophage	ZC4

ZC4vMAG9	temperate	ZC4
ZCIvMAG211	temperate	ZCI
ZCIvMAG57	prophage	ZCI
ZCMvMAG167	virulent	ZCM

Supplementary Table S3. Assignment of vMAGs with bMAGs retrieved from composting piles ZC3 and ZC4. The column 'bMAG taxon' shows either the species level (if reached) or order of the bMAG. Column 'association type' identifies the kind of method that retrieved the predicted association.

vMAG	bMAG	bMAG taxon	Association type	bMAG total taxonomy
ZC3vMAG11	ZC4RG05	*Limnochordales*	crispr	d Bacteria:p Firmicutes Gic Limnochordia;o Limnochordales; ZCTH02-B6;g
ZC3vMAG13	ZC4RG37	*Thermaerobacterales*	shared	d Bacteriacp Firmicutes Ex Thermaerobacteria:co Thermaerobacterales;f Thermaerobacteraceac:g a
ZC3vMAG135	ZC3RG05	*Thermobifida fusca*	crispr	d Bacteria:p Actinobacteriota;c Actinobacteria;o (Streptosporangiales;f_Streptosporangiaceaeg Thermobifida
ZC3vMAG143	ZC3RG08	*Planifilum fulgidum*	crispr	d Bacteria:p Firmicutes;c Bacilli A;o Thermoactinomycetales; Thermoactinomycetaceaeg Planifilum
ZC3vMAG167	ZC3RG08	*Planifilum fulgidum*	crispr	d Bacteria:p Firmicutes;c Bacilli A;o Thermoactinomycetales; Thermoactinomycetaceaeg Planifilum
ZC3vMAG172	ZC4RG48	*Chloroflexales*	shared	d Bacteria:pp Chloroflexota;c Chloroflexia;o Chloroflexales;f Roseiflexaceae:g a
ZC3vMAG177	ZC3RG08	*Planifilum fulgidum*	crispr	d Bacteria:p Firmicutes;c Bacilli A;o Thermoactinomycetales; Thermoactinomycetaceaeg Planifilum
ZC3vMAG181	ZC4RG43	*Mycobacterium hassiacum*	shared	d Bacteriacp Actinobacteriotacc Actinobacteriaco Corynebacteriales;f Corynebacteriaceae:g Mycobacterium
ZC3vMAG191	ZC3RG05	*Thermobifida fusca*	crispr	d Bacteria:p Actinobacteriota;c Actinobacteria;o

				(Streptosporangiales;f_Streptosporangiacea eg Thermobifida
ZC3vMAG191	ZC4RG21	*Thermobifida fusca*	crispr	d Bacteriacp Actinobacteriota;c Actinobacteria; Streptosporangiales;f Streptosporangiaccae;g Thermobifida
ZC3vMAG192	ZC3RG05	*Thermobifida fusca*	crispr	d Bacteria:p Actinobacteriota;c Actinobacteria;o (Streptosporangiales;f_Streptosporangiacea eg Thermobifida
ZC3vMAG192	ZC4RG21	*Thermobifida fusca*	crispr	d Bacteriacp Actinobacteriota;c Actinobacteria; Streptosporangiales;f Streptosporangiaccae;g Thermobifida
ZC3vMAG21	ZC4RG45	*Thermocrispu m agreste*	shared	d Bacteriacp Actinobacteriotacc Actinobacteriaco Corymebacteriales;f Pseudonocardiacene:g Thermocrispum
ZC3vMAG38	ZC3RG03	*Christensenell ales*	trna	d Bacteria:pp Firmicutes Ax Clostridiaco Christensenellales:f CAG-74;g 28
ZC3vMAG38	ZC3RG08	*Planifilum fulgidum*	trna	d Bacteria:p Firmicutes;c Bacilli A;o Thermoactinomycetales; Thermoactinomycetaceaeg Planifilum
ZC3vMAG38	ZC3RG09	*Limnochordal es*	trna	d Bacteria:pp Firmicutes Gue Limnochordia;o Limnochordales;f ZCTH02-B6;g ZCTH02-B6;s
ZC3vMAG38	ZC4RG03	*Calditerricolal es*	trna	d Bacteria:pp Firmicutes;c Bacilli A:o Calditerricolales:f Calditerricolaccac;g Calditerricola:s
ZC3vMAG38	ZC4RG06	*Limnochordal es*	trna	d Bacteria:pp Firmicutes Gue Limnochordia:co Limnochordales;f ZCTH02-B6;g ZCTH02-B6;
ZC3vMAG38	ZC4RG07	*Bacillales*	trna	d Bacteria:pp Firmicutes;c Bacilli A;o Bacillales F:f Bacillacene Mag Bacillus BB:s GCA 001884825.1
ZC3vMAG38	ZC4RG09	*Planifilum fulgidum*	trna	d Bacteria:pp Firmicutes:c Bacilli A:o Thermoactinomycetales;f Thermoactinomycetaceac;g Planifilum
ZC3vMAG38	ZC4RG14	*Steroidobacte rales*	trna	d Bacteriacp Proteobacteria;c Gammaproteobacteria;e Steroidobactersles:f Steroidobacteracesece a
ZC3vMAG38	ZC4RG16	*Gemmatimon adetes SG8-23*	trna	d Bacteriacp Gemmatimonadota; Gemmatimonadetes;o SG8-23;f
ZC3vMAG38	ZC4RG32	*Caldicoprobac ter ashimat*	trna	d Bacteriacp Firmicutes Age Mahelliaco Caldicoprobacterales,f Caldicoprobacteraceae:g Caldicoprobacter

ZC3vMAG38	ZC4RG34	*Thermovenabulales*	trna	d Bacteriacp Firmicutes Ax Thermovenabulia:o Thermovenabulales:f Thermovenabulaceac:g a
ZC3vMAG67	ZC3RG08	*Planifilum fulgidum*	crispr	d Bacteria:p Firmicutes;c Bacilli A;o Thermoactinomycetales; Thermoactinomycetaceaeg Planifilum
ZC3vMAG67	ZC4RG48	*Chloroflexales*	shared	d Bacteria:pp Chloroflexota;c Chloroflexia;o Chloroflexales;f Roseiflexaceae:g a
ZC3vMAG74	ZC3RG05	*Thermobifida fusca*	crispr	d Bacteria:p Actinobacteriota;c Actinobacteria;o (Streptosporangiales;f_Streptosporangiaceaeg Thermobifida
ZC3vMAG74	ZC4RG21	*Thermobifida fusca*	crispr	d Bacteriacp Actinobacteriota;c Actinobacteria; Streptosporangiales;f Streptosporangiaccae;g Thermobifida
ZC3vMAG82	ZC4RG21	*Thermobifida fusca*	crispr	d Bacteriacp Actinobacteriota;c Actinobacteria; Streptosporangiales;f Streptosporangiaccae;g Thermobifida
ZC3vMAG82	ZC4RG26	*Sphaerobacter thermophilus*	shared	d Bacteriacp Chloroflexota;c Chloroflexia;o Thermomicrobiales;f Thermomicrobiaceacg Sphaerobacter
ZC3vMAG96	ZC3RG01	*Bacteroidales*	shared_trna	d Bacteria:pp Bacteroidota:c Bacteroidia:e Bacteroidales: F082:g F082:s
ZC4vMAG116	ZC4RG32	*Caldicoprobacter ashimat*	shared	d Bacteriacp Firmicutes Age Mahelliaco Caldicoprobacterales,f Caldicoprobacteraceae:g Caldicoprobacter
ZC4vMAG116	ZC4RG49	*[Clostridium] cellulosi*	shared	d Bacteriacp Firmicutes Ax Clestridia:e Oscillospirales; Ethanoligenenacese:g Ruminiclostridium D
ZC4vMAG122	ZC4RG25	*Rhizobiales*	trna	d Bacteriacp Proteobacteria;e Alphaproteobacteria:o Rhizobiales:f Hyphomicrobiaceac:g
ZC4vMAG122	ZC4RG31	*Rhizobiales*	shared	d Bacteriacp Proteobacteria:c Alphaproteobacteria:co Rhizobiales;f Hyphomicrobiaceac:g is
ZC4vMAG14	ZC4RG20	*Gammaproteobacteria UBA6522*	shared	d Bacteriacp Proteobacteria:c Gammaproteobacteria;o UBA6522;f UBA6522;g is
ZC4vMAG166	ZC4RG04	*Thermobispora hispora*	crispr	d Bacteria:pp Actinobacteriota;c Actinobacteria; Streptosporangiales;f Streptosporangiaccac;g Thermobispora
ZC4vMAG166	ZC4RG05	*Limnochordales*	crispr	d Bacteria:p Firmicutes Gic Limnochordia;o Limnochordales; ZCTH02-B6;g
ZC4vMAG178	ZC3RG08	*Planifilum fulgidum*	crispr	d Bacteria:p Firmicutes;c Bacilli A;o Thermoactinomycetales;

				Thermoactinomycetaceaeg Planifilum
ZC4vMAG178	ZC4RG48	*Chloroflexales*	shared	d Bacteria:pp Chloroflexota;c Chloroflexia;o Chloroflexales;f Roseiflexaceae:g a
ZC4vMAG188	ZC3RG05	*Thermobifida fusca*	crispr	d Bacteria:p Actinobacteriota;c Actinobacteria;o (Streptosporangiales;f_Streptosporangiacea eg Thermobifida
ZC4vMAG188	ZC4RG21	*Thermobifida fusca*	crispr	d Bacteriacp Actinobacteriota;c Actinobacteria; Streptosporangiales;f Streptosporangiaccae;g Thermobifida
ZC4vMAG194	ZC4RG01	*Caldibacillus debilis*	trna	d Bacteria:pp Firmicutes:c Bacilli;o Bacillales: Caldibacillaceae;g Caldibacillas
ZC4vMAG199	ZC4RG05	*Limnochordal es*	crispr	d Bacteria:p Firmicutes Gic Limnochordia;o Limnochordales; ZCTH02-B6;g
ZC4vMAG22	ZC4RG26	*Sphaerobacte r thermophilus*	shared	d Bacteriacp Chloroflexota;c Chloroflexia;o Thermomicrobiales;f Thermomicrobiaceacg Sphaerobacter
ZC4vMAG264	ZC4RG28	*Streptosporan giales*	trna	d Bacteriacp Actinobacteriotace Actinobacteriaco Streptosporangiales;f Streptosporangiaccacig is
ZC4vMAG264	ZC4RG47	*Corynebacteri ales*	trna	d Bacteriacp Actinobacteriotace Actinobacteriaco Corynebacteriales;f Micromonosporaceseg Micromonospora;s
ZC4vMAG271	ZC4RG05	*Limnochordal es*	crispr	d Bacteria:p Firmicutes Gic Limnochordia;o Limnochordales; ZCTH02-B6;g
ZC4vMAG286	ZC4RG05	*Limnochordal es*	crispr	d Bacteria:p Firmicutes Gic Limnochordia;o Limnochordales; ZCTH02-B6;g
ZC4vMAG292	ZC4RG25	*Rhizobiales*	shared	d Bacteriacp Proteobacteria;e Alphaproteobacteria:o Rhizobiales:f Hyphomicrobiaceac:g
ZC4vMAG336	ZC4RG04	*Thermobispor a hispora*	crispr	d Bacteria:pp Actinobacteriota;c Actinobacteria; Streptosporangiales;f Streptosporangiaccac;g Thermobispora
ZC4vMAG347	ZC4RG05	*Limnochordal es*	crispr	d Bacteria:p Firmicutes Gic Limnochordia;o Limnochordales; ZCTH02-B6;g
ZC4vMAG414	ZC3RG05	*Thermobifida fusca*	crispr	d Bacteria:p Actinobacteriota;c Actinobacteria;o (Streptosporangiales;f_Streptosporangiacea eg Thermobifida
ZC4vMAG414	ZC4RG04	*Thermobispor a hispora*	crispr	d Bacteria:pp Actinobacteriota;c Actinobacteria; Streptosporangiales;f Streptosporangiaccac;g Thermobispora
ZC4vMAG414	ZC4RG28	*Streptosporan*	crispr	d Bacteriacp Actinobacteriotace

		giales		Actinobacteriaco Streptosporangiales;f Streptosporangiaccacig is
ZC4vMAG414	ZC4RG43	*Mycobacterium hassiacum*	shared	d Bacteriacp Actinobacteriotacc Actinobacteriaco Corynebacteriales;f Corynebacteriaceae:g Mycobacterium
ZC4vMAG451	ZC3RG08	*Planifilum fulgidum*	trna	d Bacteria:p Firmicutes;c Bacilli A;o Thermoactinomycetales; Thermoactinomycetaceaeg Planifilum
ZC4vMAG451	ZC4RG03	*Calditerricolales*	trna	d Bacteria:pp Firmicutes;c Bacilli A:o Calditerricolales:f Calditerricolaccac;g Calditerricola:s
ZC4vMAG451	ZC4RG07	*Bacillales*	trna	d Bacteria:pp Firmicutes;c Bacilli A;o Bacillales F:f Bacillacene Mag Bacillus BB:s GCA 001884825.1
ZC4vMAG451	ZC4RG09	*Planifilum fulgidum*	trna	d Bacteria:pp Firmicutes:c Bacilli A:o Thermoactinomycetales;f Thermoactinomycetaceac;g Planifilum
ZC4vMAG451	ZC4RG15	*Bacillales*	trna	d Bacteria:pp Firmicutes:c Bacillizo Bacillales A;f Planococcaccae;g Ureibacillus;s
ZC4vMAG451	ZC4RG19	*Microtrichales*	trna	d Bacteriacp Actinobacteriota;co Acidimicrobiia; Microtrichales:f UBA9382:g is
ZC4vMAG451	ZC4RG34	*Thermovenabulales*	trna	d Bacteriacp Firmicutes Ax Thermovenabulia:o Thermovenabulales:f Thermovenabulaceac:g a
ZC4vMAG451	ZC4RG41	*Actinomycetales*	trna	d Bacteriacp Actinobacteriotaçc Actinobacteriaço Actinomycetales;f Demequinacenc; is
ZC4vMAG451	ZC4RG45	*Thermocrispum agreste*	trna	d Bacteriacp Actinobacteriotacc Actinobacteriaco Corymebacteriales;f Pseudonocardiacene:g Thermocrispum
ZC4vMAG451	ZC4RG47	*Corynebacteriales*	trna	d Bacteriacp Actinobacteriotace Actinobacteriaco Corynebacteriales;f Micromonosporaceseg Micromonospora;s
ZC4vMAG468	ZC3RG05	*Thermobifida fusca*	crispr	d Bacteria:p Actinobacteriota;c Actinobacteria;o (Streptosporangiales;f_Streptosporangiaceaeg Thermobifida
ZC4vMAG468	ZC4RG05	*Limnochordales*	crispr	d Bacteria:p Firmicutes Gic Limnochordia;o Limnochordales; ZCTH02-B6;g
ZC4vMAG468	ZC4RG21	*Thermobifida fusca*	crispr	d Bacteriacp Actinobacteriota;c Actinobacteria; Streptosporangiales;f Streptosporangiaccae;g Thermobifida

ZC4vMAG473	ZC3RG05	*Thermobifida fusca*	crispr	d Bacteria:p Actinobacteriota;c Actinobacteria;o (Streptosporangiales;f_Streptosporangiacea eg Thermobifida
ZC4vMAG473	ZC4RG21	*Thermobifida fusca*	crispr	d Bacteriacp Actinobacteriota;c Actinobacteria; Streptosporangiales;f Streptosporangiaccae;g Thermobifida
ZC4vMAG486	ZC4RG26	*Sphaerobacte r thermophilus*	crispr	d Bacteriacp Chloroflexota;c Chloroflexia;o Thermomicrobiales;f Thermomicrobiaceacg Sphaerobacter
ZC4vMAG490	ZC3RG11	*Thermaeroba cterales*	shared	d Bacteria:pp Firmicutes Ex Thermaerobacteriaco Thermaerobacterales;f if is
ZC4vMAG490	ZC4RG05	*Limnochordal es*	crispr	d Bacteria:p Firmicutes Gic Limnochordia;o Limnochordales; ZCTH02-B6;g
ZC4vMAG502	ZC4RG47	*Corynebacteri ales*	crispr	d Bacteriacp Actinobacteriotace Actinobacteriaco Corynebacteriales;f Micromonosporaceseg Micromonospora;s
ZC4vMAG520	ZC4RG06	*Limnochordal es*	shared	d Bacteria:pp Firmicutes Gue Limnochordia:co Limnochordales;f ZCTH02-B6;g ZCTH02-B6;
ZC4vMAG541	ZC4RG49	*[Clostridium] cellulosi*	shared	d Bacteriacp Firmicutes Ax Clestridia:e Oscillospirales; Ethanoligenenacese:g Ruminiclostridium D
ZC4vMAG542	ZC4RG05	*Limnochordal es*	crispr	d Bacteria:p Firmicutes Gic Limnochordia;o Limnochordales; ZCTH02-B6;g
ZC4vMAG548	ZC4RG09	*Planifilum fulgidum*	crispr	d Bacteria:pp Firmicutes:c Bacilli A:o Thermoactinomycetales;f Thermoactinomycetaceac;g Planifilum
ZC4vMAG550	ZC3RG11	*Thermaeroba cterales*	shared	d Bacteria:pp Firmicutes Ex Thermaerobacteriaco Thermaerobacterales;f if is
ZC4vMAG565	ZC3RG10	*Limnochordal es*	trna	d Bacteria:pp Firmicutes Gxc Limnochordia:c Limnochordales: ZCTH02-B6;g 3
ZC4vMAG565	ZC4RG07	*Bacillales*	trna	d Bacteria:pp Firmicutes;c Bacilli A;o Bacillales F:f Bacillacene Mag Bacillus BB:s GCA 001884825.1
ZC4vMAG57	ZC4RG14	*Steroidobacte rales*	crispr	d Bacteriacp Proteobacteria;c Gammaproteobacteria;e Steroidobactersles:f Steroidobacteracesece a
ZC4vMAG60	ZC3RG09	*Limnochordal es*	crispr	d Bacteria:pp Firmicutes Gue Limnochordia;o Limnochordales;f ZCTH02-B6;g ZCTH02-B6;s

ZC4vMAG632	ZC3RG07	*Limnochordales*	trna	d Bacteria:pp Firmicutes Goe Limnochordia;a Limnochordales;f Limnochordaceac:g Limnochordacs
ZC4vMAG632	ZC3RG08	*Planifilum fulgidum*	trna	d Bacteria:p Firmicutes;c Bacilli A;o Thermoactinomycetales; Thermoactinomycetaceaeg Planifilum
ZC4vMAG632	ZC3RG09	*Limnochordales*	shared_trna	d Bacteria:pp Firmicutes Gue Limnochordia;o Limnochordales;f ZCTH02-B6;g ZCTH02-B6;s
ZC4vMAG632	ZC3RG10	*Limnochordales*	trna	d Bacteria:pp Firmicutes Gxc Limnochordia:c Limnochordales: ZCTH02-B6;g 3
ZC4vMAG632	ZC4RG05	*Limnochordales*	crispr_trna	d Bacteria:p Firmicutes Gic Limnochordia;o Limnochordales; ZCTH02-B6;g
ZC4vMAG632	ZC4RG06	*Limnochordales*	shared_trna	d Bacteria:pp Firmicutes Gue Limnochordia:co Limnochordales;f ZCTH02-B6;g ZCTH02-B6;
ZC4vMAG632	ZC4RG09	*Planifilum fulgidum*	trna	d Bacteria:pp Firmicutes:c Bacilli A:o Thermoactinomycetales;f Thermoactinomycetaceac;g Planifilum
ZC4vMAG632	ZC4RG11	*Limnochordales*	trna	d Bacteria:pp Firmicutes Gic Limnochordia:o Limnochordales:f Limnochordaceac;g Limnochorda;s
ZC4vMAG642	ZC4RG03	*Calditerricolales*	shared	d Bacteria:pp Firmicutes;c Bacilli A:o Calditerricolales:f Calditerricolaccac;g Calditerricola:s
ZC4vMAG660	ZC4RG04	*Thermobispora hispora*	crispr	d Bacteria:pp Actinobacteriota;c Actinobacteria; Streptosporangiales;f Streptosporangiaccac;g Thermobispora
ZC4vMAG660	ZC4RG47	*Corynebacteriales*	crispr	d Bacteriacp Actinobacteriotace Actinobacteriaco Corynebacteriales;f Micromonosporaceseg Micromonospora;s
ZC4vMAG694	ZC3RG05	*Thermobifida fusca*	crispr	d Bacteria:p Actinobacteriota;c Actinobacteria;o (Streptosporangiales;f_Streptosporangiaceaeg Thermobifida
ZC4vMAG694	ZC4RG21	*Thermobifida fusca*	crispr	d Bacteriacp Actinobacteriota;c Actinobacteria; Streptosporangiales;f Streptosporangiaccae;g Thermobifida
ZC4vMAG696	ZC3RG08	*Planifilum fulgidum*	crispr	d Bacteria:p Firmicutes;c Bacilli A;o Thermoactinomycetales; Thermoactinomycetaceaeg Planifilum
ZC4vMAG709	ZC3RG10	*Limnochordales*	trna	d Bacteria:pp Firmicutes Gxc Limnochordia:c Limnochordales: ZCTH02-B6;g 3

ZC4vMAG709	ZC4RG05	*Limnochordales*	crispr_trna	d Bacteria:p Firmicutes Gic Limnochordia;o Limnochordales; ZCTH02-B6;g
ZC4vMAG709	ZC4RG10	*Limnochordia DTU080*	trna	d Bacteria:pp Firmicutes Gxc Limnochordia:co DTU080;f is 3
ZC4vMAG724	ZC4RG27	*Rhizobiales*	trna	d Bacteria:p Proteobacteria:c Alphaproteobacteria:o Rhizobiales:f Beijerinckinceae;g Chelatococcuss
ZC4vMAG724	ZC4RG31	*Rhizobiales*	trna	d Bacteriacp Proteobacteria:c Alphaproteobacteria:co Rhizobiales;f Hyphomicrobiaceac:g is
ZC4vMAG724	ZC4RG33	*Rhizobiales*	trna	d Bacteriacp Protephacteria:c Alphaprotoobacteria:o Rhizobiales:f Rhizobiaccae;g Aquamicrobium;s
ZC4vMAG737	ZC3RG05	*Thermobifida fusca*	crispr	d Bacteria:p Actinobacteriota;c Actinobacteria;o (Streptosporangiales;f_Streptosporangiacea eg Thermobifida
ZC4vMAG737	ZC4RG21	*Thermobifida fusca*	crispr	d Bacteriacp Actinobacteriota;c Actinobacteria; Streptosporangiales;f Streptosporangiaccae;g Thermobifida
ZC4vMAG793	ZC3RG08	*Planifilum fulgidum*	crispr	d Bacteria:p Firmicutes;c Bacilli A;o Thermoactinomycetales; Thermoactinomycetaceaeg Planifilum
ZC4vMAG793	ZC4RG48	*Chloroflexales*	shared	d Bacteria:pp Chloroflexota;c Chloroflexia;o Chloroflexales;f Roseiflexaceae:g a
ZC4vMAG798	ZC4RG13	*Rhodothermus marinus*	crispr	d Bacteria:pp Bacteroidota;c Rhodothermia;o Rhodothermales:f Rhodothermaceac.g Rhodothermus
ZC4vMAG817	ZC4RG04	*Thermobispora hispora*	crispr	d Bacteria:pp Actinobacteriota;c Actinobacteria; Streptosporangiales;f Streptosporangiaccac;g Thermobispora
ZC4vMAG824	ZC3RG07	*Limnochordales*	trna	d Bacteria:pp Firmicutes Goe Limnochordia;a Limnochordales;f Limnochordaceac:g Limnochordacs
ZC4vMAG824	ZC3RG09	*Limnochordales*	trna	d Bacteria:pp Firmicutes Gue Limnochordia;o Limnochordales;f ZCTH02-B6;g ZCTH02-B6;s
ZC4vMAG824	ZC3RG10	*Limnochordales*	trna	d Bacteria:pp Firmicutes Gxc Limnochordia:c Limnochordales: ZCTH02-B6;g 3
ZC4vMAG824	ZC3RG11	*Thermaerobacterales*	trna	d Bacteria:pp Firmicutes Ex Thermaerobacteriaco Thermaerobacterales;f if is
ZC4vMAG824	ZC4RG05	*Limnochordales*	trna	d Bacteria:p Firmicutes Gic Limnochordia;o Limnochordales; ZCTH02-B6;g

ZC4vMAG824	ZC4RG06	*Limnochordales*	trna	d Bacteria:pp Firmicutes Gue Limnochordia:co Limnochordales;f ZCTH02-B6;g ZCTH02-B6;
ZC4vMAG824	ZC4RG11	*Limnochordales*	trna	d Bacteria:pp Firmicutes Gic Limnochordia:o Limnochordales:f Limnochordaceac;g Limnochorda;s
ZC4vMAG824	ZC4RG14	*Steroidobacterales*	shared	d Bacteriacp Proteobacteria;c Gammaproteobacteria;e Steroidobactersles:f Steroidobacteracesecea
ZC4vMAG824	ZC4RG18	*Thermaerobacterales*	trna	d Bacteria:pp Firmicutes Ex Thermaerobacteria:co Thermaerobacterales:f ig_is
ZC4vMAG824	ZC4RG23	*Symbiobacteriales*	trna	d Bacteriacp Firmicutes Ex Symbiobacteria:co Symbiobacteriales;f Symbiobacteriaccac:g Symbiobacterium:sa
ZC4vMAG824	ZC4RG25	*Rhizobiales*	trna	d Bacteriacp Proteobacteria;e Alphaproteobacteria:o Rhizobiales:f Hyphomicrobiaceac:g
ZC4vMAG824	ZC4RG31	*Rhizobiales*	trna	d Bacteriacp Proteobacteria:c Alphaproteobacteria:co Rhizobiales;f Hyphomicrobiaceac:g is
ZC4vMAG824	ZC4RG38	*Symbiobacteriales*	trna	d Bacteriacp Firmicutes Ex Symbiobacteriia;o Symbiobacteriales:f g is
ZC4vMAG837	ZC3RG05	*Thermobifida fusca*	crispr	d Bacteria:p Actinobacteriota;c Actinobacteria;o (Streptosporangiales;f_Streptosporangiaceaeg Thermobifida
ZC4vMAG837	ZC4RG21	*Thermobifida fusca*	crispr	d Bacteriacp Actinobacteriota;c Actinobacteria; Streptosporangiales;f Streptosporangiaccae;g Thermobifida
ZC4vMAG837	ZC4RG47	*Corynebacteriales*	shared	d Bacteriacp Actinobacteriotace Actinobacteriaco Corynebacteriales;f Micromonosporaceseg Micromonospora;s
ZC4vMAG846	ZC4RG09	*Planifilum fulgidum*	crispr	d Bacteria:pp Firmicutes:c Bacilli A:o Thermoactinomycetales;f Thermoactinomycetaceac;g Planifilum
ZC4vMAG851	ZC4RG39	*Steroidobacterales*	shared	d Bacteria:p Proteobacteria;c Gammaproteobacteriae Steroidobacterales; Steroidobacteraceacig a
ZC4vMAG878	ZC3RG07	*Limnochordales*	crispr	d Bacteria:pp Firmicutes Goe Limnochordia;a Limnochordales;f Limnochordaceac:g Limnochordacs
ZC4vMAG883	ZC3RG06	*Limnochordia DTU080*	trna	d Bacteria:pp Firmicutes Gx Limnochordia;o DTU080:f ig a

ZC4vMAG883	ZC3RG07	*Limnochordales*	trna	d Bacteria:pp Firmicutes Goe Limnochordia;a Limnochordales;f Limnochordaceac:g Limnochordacs
ZC4vMAG883	ZC3RG09	*Limnochordales*	trna	d Bacteria:pp Firmicutes Gue Limnochordia;o Limnochordales;f ZCTH02-B6;g ZCTH02-B6;s
ZC4vMAG883	ZC3RG10	*Limnochordales*	trna	d Bacteria:pp Firmicutes Gxc Limnochordia:c Limnochordales: ZCTH02-B6;g 3
ZC4vMAG883	ZC4RG05	*Limnochordales*	trna	d Bacteria:p Firmicutes Gic Limnochordia;o Limnochordales; ZCTH02-B6;g
ZC4vMAG883	ZC4RG06	*Limnochordales*	trna	d Bacteria:pp Firmicutes Gue Limnochordia:co Limnochordales;f ZCTH02-B6;g ZCTH02-B6;
ZC4vMAG883	ZC4RG10	*Limnochordia DTU080*	trna	d Bacteria:pp Firmicutes Gxc Limnochordia:co DTU080;f is 3
ZC4vMAG883	ZC4RG11	*Limnochordales*	trna	d Bacteria:pp Firmicutes Gic Limnochordia:o Limnochordales:f Limnochordaceac;g Limnochorda;s
ZC4vMAG884	ZC4RG04	*Thermobispora hispora*	crispr	d Bacteria:pp Actinobacteriota;c Actinobacteria; Streptosporangiales;f Streptosporangiaccac;g Thermobispora
ZC4vMAG92	ZC4RG33	*Rhizobiales*	shared	d Bacteriacp Protephacteria:c Alphaprotoobacteria:o Rhizobiales:f Rhizobiaccae;g Aquamicrobium;s

Supplementary Table S4. Abundance of MAGs per lifestyle in composting piles ZC1, ZC2, ZCI and ZCM.

Composting	Virulent	Temperate	Prophage	NA
ZC1	52,25%	25,79%	0	21,96%
ZC2	24,14%	64,2%	0	11,66%
ZCI	42,83%	31,96%	6,72%	18,49%
ZCM	67,22%	20,68%	0,65%	11,45%

Supplementary Table S5. Abundance of MAGs carrying AMG and VF in the distinct composting piles. The values are shown in proportion (%). Each column presents the sum of MAGs identified with AMGs, VFs or none of these gene classes (NA).

Composting	Days	vMAGs with AMGs (%)	vMAGs with VFs (%)	NA (%)
ZC1	Day 08	0	0	100
ZC2	Day 60	35,79	0	64,21
ZC3	Day 01	0	29,91	70,09
	Day 30	3,43	0,51	96,06
	Day 64	12	0,56	87,44
	Day 78	8,43	3,29	88,28
	Day 99	9,6	2,86	87,54

	Day 01	23,85	5,97	70,18
	Day 03	17,56	2,78	79,66
	Day 07	15,87	3,88	80,25
	Day 15	19,36	0,89	79,75
ZC4	Day 30	6,67	27,16	66,17
	Day 64	17,25	2,71	80,04
	Day 67	20,66	2,69	76,65
	Day 78	16,51	1,52	81,97
	Day 99	7,99	3,58	88,43
	Day 85	10,43	1	88,57
	Day 87	45,69	0,05	54,26
ZCI	Day 64	13,16	1,11	85,73
	Day 84	24,68	3,43	71,89
	Day 82	14,18	0,26	85,56
	4 months	0,17	0	99,83
ZCM	2 months	4,83	0	95,17

Supplementary Table S6. vMAGs genes identified as possible AMGs. The dbCAN families, subfamilies, and putative substrates are shown. See section 3.4 of Material and Methods for a detailed description.

dbCAN subfam	Subfam Composition	Subfam EC	Substrate	Profile Length	Gene ID	E Value	Coverage
GT101_e12	GT101:67|GT2:1	-	-	224	Contig_12_ZC4_bin_118_phanotate_70_geneCall	2,5E-18	67%
GH108_e39	GH108:10	-	-	77	Contig_49_ZC4_bin_490_phanotate_197_geneCall	7,8E-18	96%
GH108_e39	GH108:10	-	-	77	Contig_5_ZC4_bin_286_phanotate_31_geneCall	9,7E-31	96%
GH23_e533	GH23:7	-	-	121	Contig_6_ZC4_bin_788_phanotate_21_geneCall	1E-15	90%
GH108_e39	GH108:10	-	-	77	Contig_72_ZC4_bin_490_phanotate_299_geneCall	9,7E-31	96%
GH24_e192	GH24:6	-	-	121	Contig_10_ZC4_bin_798_phanotate_105_geneCall	5,1E-49	98%
GH108_e29	GH108:8	-	-	85	Contig_11_ZCI_bin_82_phanotate_46_geneCall	3,4E-15	98%
GT4_e695	GT4:13	-	-	85	Contig_135_ZC4_bin_414_phanotate_912_geneCall	3,3E-39	96%
GT101_e12	GT101:67|GT2:1	-	-	224	Contig_17_ZC4_bin_603_phanotate_190_geneCall	1,3E-24	88%
PL7_e3	PL7_3:74|CBM32:18|GH16:7|PL6_2:6|PL7:3|CBM16:1	4,2,2,11:4	alginate	212	Contig_1_ZC3_bin_21_phanotate_2_geneCall	2E-26	87%
GH108_e39	GH108:10	-	-	77	Contig_2_ZC4_bin_166_phanotate_10_geneCall	1,3E-16	96%
GT2	GT2:84	-	-	246	Contig_33_ZC4_bin_709_phanotate_128_geneCall	3,6E-32	80%
GT101_e12	GT101:67|GT2:1	-	-	224	Contig_40_ZC4_bin_118_phanotate_249_geneCall	3,8E-19	92%
GH23_e335	GH23:1146|CBM50:10	3,2,1,17:12|4,2,2,n1:6	peptidoglycan, peptidoglycan	121	Contig_5_ZC3_bin_185_phanotate_22_geneCall	1,2E-30	97%
PL7_e3	PL7_3:74|CBM32:18|GH16:7|PL6_2:6|PL7:3|CBM16:1	4,2,2,11:4	alginate	212	Contig_5_ZC3_bin_21_phanotate_19_geneCall	3,1E-41	87%
GH1_e0	GH1:9150	3,2,1,86:26|3,2,1,23:2|3,2,1,21:1	beta-glucan, beta-galactan, beta-glucan	472	Contig_7_ZC2_bin_44_phanotate_26_geneCall	1,9E-124	57%
GT4_e695	GT4:13	-	-	85	Contig_12_ZC4_bin_414_phanotate_88_geneCall	1,3E-39	96%
GH24_e18	GH24:97	-	-	133	Contig_1_ZC4_bin_350_phanotate_4_geneCall	5,8E-59	99%

GH23_e480	GH23:8|CBM50:3	-	-	133	Contig_4_ZC4_bin_523_phanotate_18_geneCall	2,7E-56	98%
GH18_e308	GH18:56	-	-	245	Contig_4_ZC4_bin_647_phanotate_14_geneCall	1,3E-51	91%
GT4_e695	GT4:13	-	-	85	Contig_51_ZC4_bin_414_phanotate_360_geneCall	1,3E-39	96%
GH23_e958	GH23:159	-	-	115	Contig_6_ZC4_bin_217_phanotate_34_geneCall	1E-17	93%
GH108_e29	GH108:8	-	-	85	Contig_7_ZC4_bin_102_phanotate_26_geneCall	3,4E-15	98%
GH108_e71	GH108:7	-	-	84	Contig_8_ZC4_bin_603_phanotate_114_geneCall	2,6E-16	98%
GH24_e192	GH24:6	-	-	121	Contig_13_ZC4_bin_798_phanotate_140_geneCall	5,1E-49	98%
GT4_e695	GT4:13	-	-	85	Contig_23_ZC4_bin_414_phanotate_199_geneCall	1,3E-39	96%
GT101_e12	GT101:67|GT2:1	-	-	224	Contig_28_ZC4_bin_118_phanotate_182_geneCall	4E-35	92%
GH19_e101	GH19:179|CBM50:25	-	-	160	Contig_2_ZCI_bin_198_phanotate_14_geneCall	1E-64	94%
PL6_e21	PL6_2:28|CBM32:9|GH16:6|PL7_3:6|PL6:2|PL38:1	4,2,2,-:1	alginate	363	Contig_5_ZC3_bin_21_phanotate_20_geneCall	6,4E-64	99%
GT4_e695	GT4:13	-	-	85	Contig_88_ZC4_bin_414_phanotate_675_geneCall	3,4E-39	96%
GH23_e480	GH23:8|CBM50:3	-	-	133	Contig_1_ZCI_bin_4_phanotate_36_geneCall	1,2E-55	98%
GT4_e3882	GT4:31	-	-	152	Contig_23_ZCI_bin_211_phanotate_132_geneCall	9,1E-43	95%
GH19_e34	GH19:662|CBM50:1	-	-	156	Contig_2_ZCI_bin_42_phanotate_69_geneCall	8,3E-51	99%
GH18_e308	GH18:56	-	-	245	Contig_3_ZC4_bin_272_phanotate_10_geneCall	2,8E-69	92%
GT2	GT2:10	-	-	146	Contig_11_ZC4_bin_838_phanotate_50_geneCall	4,7E-20	83%
GT4_e695	GT4:13	-	-	85	Contig_125_ZC4_bin_414_phanotate_863_geneCall	1,6E-39	96%
GH24_e192	GH24:6	-	-	121	Contig_2_ZC4_bin_798_phanotate_22_geneCall	5,1E-49	98%
GT4_e3136	GT4:15	-	-	154	Contig_3_ZCI_bin_213_phanotate_20_geneCall	9,4E-34	97%
GH108_e34	GH108:9	-	-	82	Contig_96_ZC4_bin_490_phanotate_429_geneCall	1,4E-24	98%

GT4_e695	GT4:13	-	-	85	Contig_114_ZC4_bin_414_phanotate_821_geneCall	3E-24	96%
GT4_e695	GT4:13	-	-	85	Contig_12_ZC4_bin_414_phanotate_90_geneCall	3E-24	96%
GH25_e97	GH25:9|CBM50:8	-	-	188	Contig_1_ZC2_bin_montagem_00002_phanotate_16_geneCall	4,2E-75	99%
CBM50_e511	CBM50:8|GH25:8	-	chitin, peptidoglycan	43	Contig_1_ZC2_bin_montagem_00002_phanotate_16_geneCall	2,4E-20	95%
CBM50_e758	CBM50:712|GH25:677|GH23:10|GH19:1	3,2,1,17:4	chitin, peptidoglycan	43	Contig_1_ZC2_bin_montagem_00002_phanotate_16_geneCall	2,4E-15	98%
GT2	GT2:136	-	-	278	Contig_23_ZCI_bin_211_phanotate_133_geneCall	1,6E-39	97%
GH24_e192	GH24:6	-	-	121	Contig_2_ZCI_bin_234_phanotate_14_geneCall	5,1E-49	98%
GT4_e695	GT4:13	-	-	85	Contig_51_ZC4_bin_414_phanotate_362_geneCall	6,8E-39	96%
GH24_e192	GH24:6	-	-	121	Contig_16_ZC4_bin_798_phanotate_210_geneCall	5,1E-49	98%
GT2	GT2:20	-	-	124	Contig_16_ZCI_bin_5_phanotate_79_geneCall	2,1E-48	99%
GH108_e6	GH108:25	-	-	77	Contig_1_ZC4_bin_256_phanotate_3_geneCall	6,6E-32	97%
GT4_e1208	GT4:9	-	-	91	Contig_1_ZCM_bin_22_phanotate_10_geneCall	2,3E-35	92%
GH108_e39	GH108:10	-	-	77	Contig_23_ZC4_bin_490_phanotate_107_geneCall	3,1E-32	97%
GT101_e12	GT101:67|GT2:1	-	-	224	Contig_44_ZC4_bin_118_phanotate_303_geneCall	6,5E-34	92%
GH108_e71	GH108:7	-	-	84	Contig_4_ZC4_bin_603_phanotate_57_geneCall	2,6E-16	98%
GH108_e34	GH108:9	-	-	82	Contig_6_ZC4_bin_632_phanotate_66_geneCall	6,7E-23	96%
GT4_e695	GT4:13	-	-	85	Contig_88_ZC4_bin_414_phanotate_677_geneCall	7,8E-39	96%
GH24_e282	GH24:279	-	-	129	Contig_11_ZCI_bin_64_phanotate_64_geneCall	3,8E-29	91%
GH108_e29	GH108:8	-	-	85	Contig_18_ZC4_bin_102_phanotate_114_geneCall	3,4E-15	98%
GH24_e96	GH24:29	-	-	136	Contig_1_ZC4_bin_372_phanotate_18_geneCall	6,9E-51	99%
GH3_e88	GH3:482	3,2,1,37:15|3,2,1,55:3|3,2,1,21:1	xylan, arabinan, beta-glucan	228	Contig_25_ZCI_bin_57_phanotate_138_geneCall	2,1E-106	99%

GH23_e533	GH23:7	-	-	121	Contig_2_ZC4_bin_788_phanotate_8_geneCall	1,3E-14	90%
GH19_e101	GH19:179|CBM50:25	-	-	160	Contig_2_ZC4_bin_857_phanotate_17_geneCall	1E-64	94%
GH24_e46	GH24:14	-	-	138	Contig_30_ZC4_bin_860_phanotate_117_geneCall	8,9E-41	99%
GH23_e405	GH23:17	-	-	175	Contig_3_ZC4_bin_125_phanotate_27_geneCall	1,4E-103	99%
GT4_e695	GT4:13	-	-	85	Contig_125_ZC4_bin_414_phanotate_865_geneCall	3,4E-39	96%
GH39_e57	GH39:36	3,2,1,37:3	xylan	424	Contig_2_ZC4_bin_116_phanotate_6_geneCall	1,1E-203	99%
GH108_e6	GH108:25	-	-	77	Contig_2_ZC4_bin_759_phanotate_82_geneCall	7,6E-24	95%
GH23_e480	GH23:8|CBM50:3	-	-	133	Contig_3_ZC4_bin_523_phanotate_13_geneCall	2,7E-56	98%
GT101_e12	GT101:67|GT2:1	-	-	224	Contig_3_ZC4_bin_603_phanotate_24_geneCall	1,3E-24	88%
GT32_e150	GT32:5	-	-	78	Contig_4_ZCI_bin_172_phanotate_58_geneCall	2,7E-19	95%
CBM50_e485	CBM50:25	-	chitin, peptidoglycan	46	Contig_6_ZC4_bin_565_phanotate_21_geneCall	1,8E-24	93%
GH23_e533	GH23:7	-	-	121	Contig_6_ZCI_bin_85_phanotate_41_geneCall	9,7E-30	90%
GH51_e0	GH51:2271	3,2,1,55:37	arabinan	492	Contig_13_ZC3_bin_67_phanotate_77_geneCall	1,2E-249	99%
GT4_e695	GT4:13	-	-	85	Contig_77_ZC4_bin_414_phanotate_585_geneCall	1,4E-39	96%
GT32_e151	GT32:18|GT62:13|GT2:11	-	-	79	Contig_7_ZC4_bin_473_phanotate_35_geneCall	5E-17	96%
GT101_e12	GT101:67|GT2:1	-	-	224	Contig_7_ZC4_bin_603_phanotate_101_geneCall	3E-27	88%
GH23_e533	GH23:7	-	-	121	Contig_7_ZC4_bin_788_phanotate_26_geneCall	9,5E-16	90%
GH16_e103	GH16_3:123|CBM13:53|GH16_5:2|GH16:2|CBM6:1	3,2,1,39:7	beta-glucan	233	Contig_11_ZC3_bin_21_phanotate_41_geneCall	1,6E-62	99%
GT4_e695	GT4:13	-	-	85	Contig_165_ZC4_bin_414_phanotate_1053_geneCall	2,6E-40	96%
GH108_e39	GH108:10	-	-	77	Contig_16_ZC4_bin_490_phanotate_75_geneCall	7,4E-19	96%
GH24_e192	GH24:6	-	-	121	Contig_17_ZCI_bin_35_phanotate_63_geneCall	2,5E-49	98%

GH18_e308	GH18:56	-	-	245	Contig_18_ZC4_bin_565_phanotate_73_geneCall	2,1E-129	100%
GH108_e29	GH108:8	-	-	85	Contig_1_ZC4_bin_92_phanotate_2_geneCall	1,1E-37	98%
GT51_e171	GT51:341	-	-	185	Contig_20_ZCI_bin_15_phanotate_82_geneCall	7,9E-13	41%
GT101_e12	GT101:67|GT2:1	-	-	224	Contig_14_ZC4_bin_118_phanotate_93_geneCall	4,9E-20	92%
GH108_e71	GH108:7	-	-	84	Contig_14_ZC4_bin_603_phanotate_170_geneCall	2,6E-16	98%
GH108_e6	GH108:25	-	-	77	Contig_14_ZCI_bin_188_phanotate_83_geneCall	8,4E-21	97%
GT32_e151	GT32:18|GT62:13|GT2:11	-	-	79	Contig_16_ZC4_bin_473_phanotate_79_geneCall	2E-17	96%
GH108_e71	GH108:7	-	-	84	Contig_22_ZC4_bin_603_phanotate_264_geneCall	2,6E-16	98%
GH18_e308	GH18:56	-	-	245	Contig_2_ZCI_bin_26_phanotate_51_geneCall	7,2E-54	92%
CE9_e10	CE9:120	-	-	375	Contig_39_ZCI_bin_35_phanotate_149_geneCall	3,6E-101	69%
GH24_e282	GH24:279	-	-	129	Contig_3_ZC3_bin_9_phanotate_25_geneCall	1,1E-33	95%
GH18_e308	GH18:56	-	-	245	Contig_3_ZC4_bin_647_phanotate_11_geneCall	1,8E-66	91%
GH108_e34	GH108:9	-	-	82	Contig_3_ZC4_bin_682_phanotate_8_geneCall	0,00000002	96%
GH108_e39	GH108:10	-	-	77	Contig_3_ZCI_bin_44_phanotate_43_geneCall	6,2E-32	95%
GH23_e210	GH23:12	-	-	104	Contig_4_ZC4_bin_482_phanotate_22_geneCall	1E-38	97%
GH28_e80	GH28:246|GH105:15|CE8:1	-	-	368	Contig_7_ZC4_bin_306_phanotate_21_geneCall	7,3E-160	99%
GT2	GT2:136	-	-	278	Contig_9_ZCI_bin_15_phanotate_32_geneCall	3,7E-62	97%
GT4_e695	GT4:13	-	-	85	Contig_135_ZC4_bin_414_phanotate_910_geneCall	3,5E-39	96%
GT20_e8	GT20:314	2,4,1,213:1	-	487	Contig_26_ZC4_bin_856_phanotate_116_geneCall	3,9E-131	71%
GT4_e3610	GT4:133	-	-	164	Contig_4_ZC4_bin_614_phanotate_11_geneCall	2,1E-54	98%
GT4_e695	GT4:13	-	-	85	Contig_77_ZC4_bin_414_phanotate_587_geneCall	1,8E-39	96%

GH23_e533	GH23:7	-	-	121	Contig_9_ZC4_bin_788_phanotate_45_geneCall	1E-15	90%
GH108_e29	GH108:8	-	-	85	Contig_10_ZC4_bin_853_phanotate_55_geneCall	1,7E-15	98%
GT32_e151	GT32:18|GT62:13|GT2:11	-	-	79	Contig_13_ZC3_bin_192_phanotate_100_geneCall	2E-17	96%
GH24_e192	GH24:6	-	-	121	Contig_14_ZC4_bin_798_phanotate_163_geneCall	5,1E-49	98%
GT101_e12	GT101:67|GT2:1	-	-	224	Contig_21_ZC4_bin_603_phanotate_248_geneCall	5,2E-26	88%
GH23_e796	GH23:14	-	-	153	Contig_28_ZCl_bin_15_phanotate_118_geneCall	2,1E-28	96%
GH24_e235	GH24:17	-	-	121	Contig_2_ZC4_bin_584_phanotate_14_geneCall	1,8E-19	97%
GH24_e192	GH24:6	-	-	121	Contig_3_ZC4_bin_798_phanotate_61_geneCall	5,1E-49	98%
GH43_e128	GH43_10:29	-	-	272	Contig_3_ZCl_bin_35_phanotate_14_geneCall	1,4E-65	50%
GH51_e19	GH51:1621	3,2,1,55:27	arabinan	489	Contig_5_ZC4_bin_116_phanotate_43_geneCall	2,1E-209	75%
GH108_e39	GH108:10	-	-	77	Contig_5_ZC4_bin_199_phanotate_36_geneCall	3,5E-32	96%
GH12_e4	GH12:7	-	-	139	Contig_5_ZC4_bin_51_phanotate_59_geneCall	1,9E-39	98%
PL7_e3	PL7_3:74|CBM32:18|GH16:7|PL6_2:6|PL7:3|CBM16:1	4,2,2,11:4	alginate	212	Contig_8_ZC3_bin_21_phanotate_26_geneCall	8,9E-27	81%
GH16_e264	GH16:16	-	-	221	Contig_8_ZC3_bin_21_phanotate_26_geneCall	8,5E-78	100%
PL6_e21	PL6_2:28|CBM32:9|GH16:6|PL7_3:6|PL6:2|PL38:1	4,2,2,-:1	alginate	363	Contig_1_ZC3_bin_21_phanotate_1_geneCall	2,2E-59	87%
GH26_e101	GH26:7	-	-	148	Contig_1_ZCM_bin_80_phanotate_7_geneCall	5,9E-15	89%
GT4_e695	GT4:13	-	-	85	Contig_23_ZC4_bin_414_phanotate_197_geneCall	3,4E-39	96%
GH25_e144	GH25:840|CBM50:678	3,2,1,17:4	peptidoglycan	181	Contig_4_ZC2_bin_35_phanotate_22_geneCall	1,3E-62	97%
CBM50_e758	CBM50:712|GH25:677|GH23:10|GH19:1	3,2,1,17:4	chitin, peptidoglycan	43	Contig_4_ZC2_bin_35_phanotate_22_geneCall	1,3E-20	98%
CBM6_e6	CBM6:625|GH43_16:614|CBM22:56|CBM36:43|CBM13:18|GH43_10:14|CBM2:8|GH10:5|CBM3:3|CE0:2|CBM35:1|GH43:1|CBM5:1	3,2,1,55:13|3,2,1,8:3|3,2,1,-:2|3,2,1,37:1	beta-glucan, xylan, cellulose	127	Contig_9_ZC4_bin_116_phanotate_52_geneCall	2,3E-45	98%

Supplementary Table S7. vMAGs genes identified as possible VF. The hits presented have already been screened for identity and e-value below the thresholds. Only coverage is shown. See section 3.4 of Material and Methods for a detailed description.

Query	Subject	Query length	Subject length	Hit length	Coverage
Contig_10_ZC4_bin_490_phanotate_43_geneCall	VFG000077(gb|NP_465991)	202	198	187	93%
Contig_11_ZC3_bin_82_phanotate_70_geneCall	VFG014099(gb|NP_251849)	400	436	404	93%
Contig_11_ZC4_bin_116_phanotate_61_geneCall	VFG000077(gb|NP_465991)	193	198	193	97%
Contig_22_ZC4_bin_856_phanotate_105_geneCall	VFG001214(gb|NP_253237)	387	445	379	85%
Contig_22_ZC4_bin_856_phanotate_105_geneCall	VFG001248(gb|NP_249788)	387	490	376	77%
Contig_22_ZC4_bin_856_phanotate_105_geneCall	VFG007496(gb|WP_001190986)	387	488	376	77%
Contig_22_ZC4_bin_856_phanotate_105_geneCall	VFG011946(gb|YP_002344419)	387	433	394	91%
Contig_22_ZC4_bin_856_phanotate_105_geneCall	VFG043381(gb|WP_000684509)	387	381	385	99%
Contig_39_ZC3_bin_47_phanotate_137_geneCall	VFG000130(gb|NP_252238)	386	520	375	72%
Contig_3_ZC4_bin_329_phanotate_26_geneCall	VFG045470(gb|AAM75247)	62	66	58	88%
Contig_3_ZC4_bin_696_phanotate_45_geneCall	VFG001866(gb|WP_015444568)	284	341	274	80%
Contig_3_ZC4_bin_696_phanotate_45_geneCall	VFG043648(gb|NP_252312)	284	334	270	81%
Contig_45_ZC4_bin_490_phanotate_181_geneCall	VFG000477(gb|NP_461845)	371	330	294	79%
Contig_47_ZC3_bin_47_phanotate_166_geneCall	VFG041304(gb|WP_010947678)	131	188	135	72%
Contig_4_ZC3_bin_147_phanotate_22_geneCall	VFG000477(gb|NP_461845)	272	330	256	78%
Contig_4_ZC3_bin_147_phanotate_22_geneCall	VFG001866(gb|WP_015444568)	272	341	257	75%
Contig_4_ZC3_bin_147_phanotate_22_geneCall	VFG043648(gb|NP_252312)	272	334	263	79%
Contig_4_ZCl_bin_186_phanotate_63_geneCall	VFG000477(gb|NP_461845)	272	330	256	78%

Contig_4_ZCI_bin_186_phanotate_63_geneCall	VFG001866(gb|WP_015444568)	272	341	257	75%
Contig_4_ZCI_bin_186_phanotate_63_geneCall	VFG043648(gb|NP_252312)	272	334	263	79%
Contig_50_ZCI_bin_5_phanotate_225_geneCall	VFG013265(gb|WP_005694045)	216	195	192	89%
Contig_5_ZC4_bin_358_phanotate_61_geneCall	VFG000477(gb|NP_461845)	272	330	256	78%
Contig_5_ZC4_bin_358_phanotate_61_geneCall	VFG001866(gb|WP_015444568)	272	341	257	75%
Contig_5_ZC4_bin_358_phanotate_61_geneCall	VFG043648(gb|NP_252312)	272	334	263	79%
Contig_8_ZC4_bin_184_phanotate_51_geneCall	VFG001866(gb|WP_015444568)	285	341	274	80%
Contig_8_ZC4_bin_184_phanotate_51_geneCall	VFG043648(gb|NP_252312)	285	334	270	81%
Contig_8_ZC4_bin_19_phanotate_40_geneCall	VFG003997(gb|NP_461706)	168	350	272	78%