UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
INSTITUTO DE BIOCIÊNCIAS - RIO CLARO

CIÊNCIAS BIOLOGICAS

MATHEUS CUSTODIO SARKIS CARDOZO

DIVERSIDADE VIRAL EM ANÁLISES METAGENÔMICAS DE FUNGOS CULTIVADOS POR FORMIGAS

VIRAL DIVERSITY IN METAGENOMIC ANALYSES OF FUNGI CULTIVATED BY ANTS

Rio Claro - SP
2024

MATHEUS CUSTODIO SARKIS CARDOZO

DIVERSIDADE VIRAL EM ANÁLISES METAGENÔMICAS DE FUNGOS CULTIVADOS POR FORMIGAS

VIRAL DIVERSITY IN METAGENOMIC ANALYSES OF FUNGI CULTIVATED BY ANTS

Trabalho de Conclusão de Curso apresentado ao Instituto de Biociências – Câmpus de Rio Claro, da Universidade Estadual Paulista “Júlio de Mesquita Filho”, para obtenção do grau de Bacharel em Ciências Biológicas

Orientador: Pepijn Wilhelmus Kooij

Supervisora: Milene Ferro

Rio Claro - SP
2024

C268d

Cardozo, Matheus Custodio Sarkis
Diversidade viral em análises metagenômicas de fungos cultivados por formigas / Matheus Custodio Sarkis Cardozo. -- Rio Claro, 2024
32 p.

Trabalho de conclusão de curso (Bacharelado - Ciências Biológicas) - Universidade Estadual Paulista (UNESP), Instituto de Biociências, Rio Claro
Orientador: Pepijn Wilhelmus Kooij
Coorientadora: Milene Ferro

1. Metagenômica. 2. de novo. 3. Formigas Cultivadoras de Fungos. 4. Fungos Mutualistas. 5. Viroma. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

MATHEUS CUSTODIO SARKIS CARDOZO

DIVERSIDADE VIRAL EM ANÁLISES METAGENÔMICAS DE FUNGOS CULTIVADOS POR FORMIGAS

VIRAL DIVERSITY IN METAGENOMIC ANALYSES OF FUNGI CULTIVATED BY ANTS

Trabalho de Conclusão de Curso apresentado ao Instituto de Biociências – Câmpus de Rio Claro, da Universidade Estadual Paulista "Júlio de Mesquita Filho", para obtenção do grau de Bacharel em Ciências Biológicas.

BANCA EXAMINADORA:

Prof. Dr. Pepijn Wilhelmus Kooij (orientador)

Profa. Dra. Milene Ferro (supervisora)

Prof. Dr. Andre Rogrigues

Aprovado em: 11 de Novembro de 2024

gov.br Documento assinado digitalmente
MATHEUS CUSTODIO SARKIS CARDOZO
Data: 11/12/2024 20:41:19-0300
Verifique em https://validar.iti.gov.br

Assinatura do discente

gov.br Documento assinado digitalmente
PEPIJN WILHELMUS KOOIJ
Data: 11/12/2024 19:16:17-0300
Verifique em https://validar.iti.gov.br

Assinatura do orientador

gov.br Documento assinado digitalmente
MILENE FERRO
Data: 11/12/2024 18:17:21-0300
Verifique em https://validar.iti.gov.br

Assinatura da supervisora

Dedico este trabalho à minha mãe, Danila, que sempre me apoiou em todas as minha jornada.

AGRADECIMENTOS

Este trabalho de conclusão de curso, assim como toda a minha jornada até então dentro na Unesp de Rio Claro é resultado do apoio de pessoas muito importantes na minha vida nos mais diversos âmbitos. Portanto, dedico algumas palavras de agradecimento por tornar meu desenvolvimento possível:

A minha mãe, Danila, por toda a perseverança, amor e carinho com que sempre cuidou de mim, independente de todas as dificuldades que enfrentamos pelo caminho, assim como sempre me incentivar e me dar condições para buscar os estudos.

Ao meu pai, Marcelo, por se fazer presente em minha vida mesmo com tantas dificuldades, e sempre demonstrar muito carinho e atenção e me inspirar a sempre buscar novos recomeços.

Ao meu orientador, Prof. Dr. Pepijn, por ter me concedido os dados e meios necessários para a realização deste trabalho, assim como uma incrível oportunidade ao me aceitar como aluno, sempre estar disponível para me ajudar com todos os desafios que enfrentei no decorrer deste trabalho e muitos outros, por ter acreditado em mim e estimulando o meu pensamento científico.

A Profa. Dra. Milene, por me apresentar ao horizonte da bioinformática e me ajudar com o possível, sempre com bom humor, atenção e disposição.

A Unesp de Rio Claro e todos os meus professores, que me acolheram como aluno e me proporcionaram infinitas oportunidades de conhecimento, nunca passei um dia sequer sem um aprendizado novo.

RESUMO

A relação simbiótica entre fungos e formigas cultivadoras de fungos vem sendo amplamente estudada como um sistema-modelo para relações mutualisticas em geral, devido a complexidade de tanto as interações entre o fungo e as formigas que o cultivam, quanto com outros organismos, como bactérias. Embora sejam principalmente associados com doenças e outras enfermidades, vírus poderiam potencialmente estarem envolvidos em uma relação de mutualismo em certas circunstâncias. Vírus que impedem o desenvolvimento de estruturas reprodutivas em fungos seriam prejudiciais, mas poderiam ser benéficos em um sistema onde o fungo seja transmitido verticalmente e não precisa mais se reproduzir sexualmente. Neste estudo, analisamos a diversidade viral associada a uma amostra de micélio de um fungo mutualista cultivado por formigas atíneas por meio de dados metagenômicos de amostras virais purificadas obtidas do fungo simbionte. Detectamos a presença de sequências de bacteriófagos na amostra, mas o viroma analisado não correspondeu a nenhum micovírus de sequência conhecida em bancos de dados públicos, como o NCBI. No entanto, muitas das sequências obtidas não foram identificadas, podendo representar uma diversidade viral ainda não catalogada, incluindo micovírus associados aos fungos mutualistas.

Palavras-chave: Análise Metagenômica; *de novo*; Formigas-Cultivadoras de Fungos; Fungos Mutualistas; Viroma.

ABSTRACT

The symbiotic relationship between fungi and fungus-growing ants has been widely studied as a model system for mutualism in general because of the complexity of both interactions between the fungus and the ants that cultivate it, as well as with other organisms, such as bacteria. While mostly associated with diseases and illnesses, viruses could potentially be involved in a mutualistic relationship under the right circumstances. Fungal castrating viruses would be detrimental but could be beneficial in a system where the fungus is vertically transmitted and should no longer have to reproduce sexually. In this study, we analyzed the viral diversity associated with a mycelium sample obtained from a mutualistic fungus cultivated by *Attini* ants through metagenomic analysis from a purified sample of the fungus. We detected the presence of bacteriophage sequences in the samples, but the virome analyzed didn't correspond to any known mycovírus available at public databases like NCBI. However, most of the sequences obtained weren't properly identified, possibly corresponding to a viral diversity not yet described, including potential mycoviruses associated with the mutualistic fungus.

Keywords: Metagenomic Analysis; *de novo*; Fungus-Growing ants; Mutualistic Fungi; Virome.

Title in english: Viral Diversity in Metagenomic analyses of Fungi Cultivated by Ants

IMAGE LIST

TABLES LIST

ABBREVIATIONS LIST

MYA	Million-Years Ago
dsRNA	Double-stranded RNA
ssRNA	Single-stranded RNA
ssDNA	Single-stranded DNA
NGS	Next Generation Sequencing
MEYB	Malt Extract Yeast Bacto
cDNA	Complementary DNA
contig	Contiguous Sequence
NCBI	National Center for Biotechnology Information
EDTA	Ethylenediamine tetraacetic acid
PEG	Polyethylene glycol
TE	Tris+EDTA buffer

TABLE OF CONTENTS

1 INTRODUCTION

The Kingdom Fungi is not only a vastly diverse group of species in the natural world, but it also represents many important assets for the global economy, eg., gastronomy, healthcare products, bioremediation, environmental research, and many other fields of study (CORBU *et al*. 2023). As such, the analyses of organisms that affect fungi in general, such as bacteria, archaebacteria, viruses, or even other fungi, are of great interest to many fields.

Bacteria at times form mutualistic relationships with fungi (Scott *et al*., 2008; Martiarena *et al*., 2023). For example, lichens are well-studied as cases of mutualism between filamentous fungi and photosynthetic individuals such as algae and cyanobacteria (Lutzoni and Miadlikowska 2009). Also, viruses can affect their fungal hosts in many ways, varying from detrimental for their hosts to even mutualistic interactions, although predominantly the outcome of such infections is a reduced growth rate (Myers *et al*. 2022).

Mutualistic relations are not only frequent in nature but also key drivers for maintaining biodiversity. However, for such a relationship to remain beneficial along the evolutionary history for both species, it is necessary to maintain a high degree of partner commitment and functional complementarity. (Janzen 1985; Keeler 1985; Kooij 2013).

Originating around 66 MYA, the mutualistic symbiosis between fungi and fungus-growing ants from the subtribe Attina (Formicidae: Myrmicinae: Attini) offers an extremely rich system in biodiversity and complexity, making it ideal for studying how the specialization process behind mutualisms develops (Schultz *et al*. 2024). The ants grow the fungus on plant material and use it as their main food source (Möller 1893; Weber 1955).

The fungal symbiont is passed down to new colonies vertically by the new queens themselves, which transport small pellets of the fungus within pouches present in their mouth cavities (infrabuccal pockets) to begin the nest's fungus garden (Mueller *et al*., 2005; Wheeler, 1907). This shows a high degree of mutual commitment of both species since the ant colony only grows that specific strain while providing it with growth substrates, protection, and mode of dispersal. In turn, the fungus won't reproduce sexually because it is transferred vertically and clonally and

functions as a stable food source for the ants. This relation however does not end with just the cultivated fungus and the attini ants, also bacteria and yeasts are known to enrich the microbiome in these fungal gardens (Mendes *et al.,* 2012; Martiarena *et al*., 2023) showing how factors outside the two main individuals can be pivotal for a mutualistic relation.

Mycoviruses are very common in some groups of fungi, like the endophytic fungi, potentially presenting mutualistic roles in complex interactions between organisms (Ahn and Lee 2001; Márquez *et al*., 2007; Pearson *et al*., 2009), which led to mycovirus research to be stimulated by the idea that they could be an effective tool for the biocontrol of fungal pathogens (Pearson, 2009). Also, some play important roles in pharmaceutics, eg., *Penicillium* which can have viruses that are potent stimulators of interferon production in animals (Ghabrial *et al*., 2015) or even an effective mechanism of control for yeast production (Meyers *et al*., 2022). Some yeasts like *Saccharomyces cerevisiae* are also known to be in relation with killer viruses that produce specific toxins against other fungi (Schmitt & Breinig, 2006).

There is the possibility that mycoviruses could be playing a vital role in the evolution of the Fungi. For example, it is not uncommon for mycoviruses to affect the growth rate and/or the development of sexual structures of their hosts (Pearson *et al*., 2009; Ghabrial *et al*., 2015). This could allow fungus-growing ants to control the sexual reproduction of their fungal partner and, in doing so, maintain it genetically the same.

Mycoviruses are primarily grouped into nine viral families of dsRNA genomes that are approved by the International Committee on Taxonomy of Viruses. However, some mycoviruses can have negative-sense ssRNA genomes and a few are known to have ssDNA genomes. RNA mycovirus genomes range in size from 2.5-23kb and encode 1-12 genes on one to several genomic segments (Myers *et al*., 2022).

The study of viruses, especially when it concerns the possible role they could play in a complex relation such as a mutualistic one, faces many challenges, especially the difficulty of the cultivation of viruses in the laboratory while trying to emulate the natural environment. This can be one of the factors for the lack of data on mycoviruses present in fungi grown by ants, or even the lack of reference material on mycoviruses in general at databases such as NCBI (Knight *et al.,* 2012*;* Villan *et al.* 2023). The use of metagenomic data can be a relevant tool for the proper study of mycoviruses in such cases since this kind of data has been used more in recent

years because of the advancement of tools such as NGS (Knight *et al*., 2012) like Oxford Nanopore (Oxford Nanopore Technologies, Oxford, UK), which can produce high-quality data.

With metagenomic data obtained with a next-generation sequencer Oxford Nanopore, it is possible to quantify and study the viral diversity found in a purified environmental sample of a region containing mutualistic fungi grown by fungus-growing ants. For virome data (viral metagenomic genomes), *de novo* approaches are often necessary, building the assembly from scratch without the need for reference materials, which for mycoviruses are often underrepresented in databases (Sutton *et al*., 2019).

2 OBJECTIVES

2.1 General Objectives

The objective was to quantify and study the viral diversity found in a purified sample of two mutualistic fungus from an ant garden grown by fungus-growing ants, *Leucoagaricus* sp., grown by *Paratrachymyrmex cornetzy* and *Leucoagaricus gongylophorus*, grown by *Atta colombica* by using the *de novo* approach to process the virome data while aiming to deepen further the knowledge of mycoviruses for future investigations on the possible role the mycoviruses might have concerning the asexuality of the fungus.

2.2 Specific Objectives

a) Process the metagenomic data obtained from a mutualistic fungus grown by ants using an NGS sequencer Oxford Nanopore Technologies MinION sequencer with the *de novo* approach;
b) Determine if the virome present in the data comes from closely related viral families that generally infect fungi.

3 MATERIAL AND METHODS

For a better understanding of the diversity present in the sample, the metagenomic data obtained by the NGS tool MinION by Oxford Nanopore Technologies was analyzed with the de novo strategy in mind, given that such data need to be created from scratch, without any reference material.

3.1 Sample Collection, Isolation and Sequencing

3.1.1 Sample Collection

The samples were obtained from the mycelium of fungal gardens grown by fungus-growing ants nests, a *Leucoagaricus* sp. grown by *Paratrachymyrmex cornetzy* and a *Leucoagaricus gongylophorus* grown by *Atta colombica* and then grown in MEYB medium, with streptomycin, tetracyclin, and a protein mix.

3.1.2 Virus Isolation and Purification Process

For the crude virus preparation, the samples were rinsed with ddH_2O on a vacuum filter, and 2 volumes of extraction buffer consisting of 50mM Tris-Cl and 1mM EDTA pH 8.0 were added. The mixture was then homogenized with a blender for 3 minutes at high speed, and centrifuged at 10.000g for 20 minutes, with the supernatant being separated. PEG-6000 and NaCl were added to the sample to a final concentration of 10% and 0.6 M respectively, and then the mixture was centrifuged at 10.000g for 20 minutes, with the supernatant removed, the pellet was resuspended in TE buffer. The suspension was clarified by centrifugation at 10.000g for 20 minutes and then ultracentrifugated at 40.000 rpm for 38 minutes. The supernatant was removed and the pellet was resuspended in 1mL TE buffer, then the suspension was clarified by centrifugation at low for 20 minutes.

For purification, 1mL of the crude virus preparation was layered onto 1,585 g/cm CsCl and ultracentrifuged at 34.700 rpm for more than 14 hours. The gradients were fractionated in 2 parts of 4mL, and then the fractions were dialyzed 24-36 hours against 3 L TE buffer, with 2 changes. The fractions were then ultracentrifuged at 40.000 rpm for 47 minutes, the supernatant was removed and the pellet was resuspended in TE buffer and stored at -20°C. The genetic material was then

converted into cDNA from RNA, and special barcodes were assigned depending on the fraction it came from to further differentiate each virome in the analysis.

The rough metagenomic data was then obtained through the NGS tool MinION Oxford Nanopore and provided by Dr. Kooij for the development of this study.

3.2 Bioinformatic Analysis

3.2.1 De novo genome assembly

Traditionally, genome assembly is achieved through comparative analysis using already published data, based on the assumption that the researcher knows which organism they are studying. However, this approach is not always feasible, especially for metagenomic data. Metagenomic data, obtained from environmental samples rather than isolated organisms, includes genetic material from many species, often unrelated to each other. To efficiently assemble genomes from such complex data, it is crucial to use sequencing technology that provides high-quality reads. Advances in NGS technologies, such as Oxford Nanopore's MinION, have been offering more precise and detailed data as time goes on, which makes the assembly of genomes from metagenomic samples increasingly feasible and common in the scientific community. Additionally, Oxford Nanopore's MinION generates long reads, which make it possible to assemble the genomic material fully without gaps, especially for shorter genomes like those from viruses.

3.2.2 Basecalling

Using the Dorado basecaller software version 0.6.2 (Oxford Nanopore Technologies), which specializes in working with data output from the Oxford Nanopore sequencer, the reads obtained by the equipment were converted into nucleotide sequences, which were then submitted to additional processes for noise correction, removal of low-quality reads and experimental artifacts generated in the first step to raise the accuracy of the sequences.

3.2.3 Correction and Assembly

The assembly was first done with two software specialized in long-reads, Canu (version 2.2) and MetaFlye (version 2.9), both employed to also correct the outputs directly out of Dorado, dealing with eventual errors that could have passed through

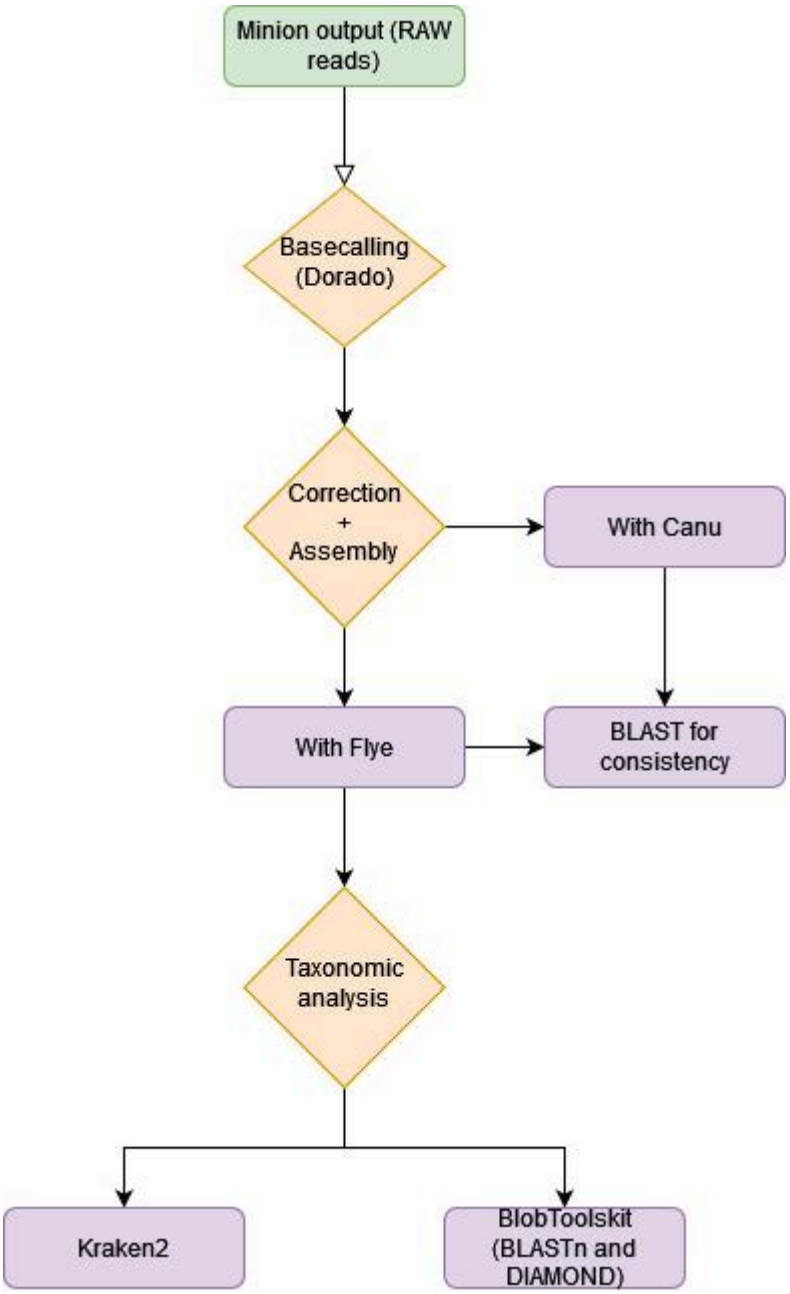
the basecaller's quality control, since they were designed for high-noise long-reads sequences (Koren *et al*., 2017; Kolmogorov *et al.*, 2020). Canu was used for being relatively good when handling high-noise sequences (Koren *et al*., 2017), while MetaFlye was used simultaneously for the assembly for its use of repeat graphs instead of de Bruijn graphs since the latter requires exact k-mer matches, which makes it more difficult to process higher noise reads compared to the approximate sequence matches of repeat graphs (Kolmogorov *et al*., 2020).

3.2.4 Taxonomic Analysis

After the assembly was properly done and processed, both the Canu and the MetaFlye sequences were used for BLAST searches under the NCBI's core nucleotide database (*core_nt*) (Altschul *et al*., 1990) to check the consistency of the correction. Then, the files generated with MetaFlye were processed through Blobtoolskit's (version 4.3) pipeline (Challis, R. *et al*., 2020) to aid in the taxonomic classification and analysis of the assembled contigs from the sample, as well as for the use of DIAMOND for BLASTX of the contigs (Buchfink, B. *et al.,* 2014) with its pipeline.

For further investigation, the contigs generated with MetaFlye were then processed with Kraken2 (version 2.1) for assigning taxonomic labels and confirming the viral diversity present in the contigs by using the exact alignment of k-mers (Wood & Salzburg, 2014; Wood, Lu & Langmead, 2019), using a custom database specifically made with only the standard viral reference library provided by the developers while using the taxonomic data (maps, name and tree information) from NCBI.

Image 1 – Workflow of the bioinformatic analysis



Sources: Developed by the author.

4 RESULTS AND DISCUSSION

Basecalling of the data provided by Dr. Kooij with the Oxford Nanopore NGS using Dorado outputted 38 batch files in ".fastq" format, distributed across 8 special identification barcodes as previously described (depending on the fraction they were collected in the gradients). After assembly, each set of sequences under a barcode was submitted to the Canu and MetaFlye pipelines. The Canu pipeline with the "Raw" setting, which means the software treats the reads as low-quality, noisy reads, resulted in 12 contigs for all the sequences, while the MetaFlye pipeline with the same configurations resulted in 43 contigs for the same sequences. When using the "HQ" setting, which means the software treats the reads as high-quality ones instead, the Canu assembly resulted in 12 contigs, while MetaFlye resulted in 28, almost double the amount compared to the other assembler. The number of contigs generated for the "RAW" setting are shown in Tables 1 and 2 for MetaFlye and Canu respectively, while the number of contigs generated for the "HQ" settings are shown in Tables 3 and 4.

When comparing the characteristics of the contigs generated by the different assemblers, as can be observed in Tables 1 to 4, both were very consistent when treating the sequences as low-quality, noisy reads (the RAW setting), with very similar N50 (when 50% of the contigs are in a sequence of a set length) and overall length. The "HQ" generated contigs, on the other hand, varied greatly both in number and characteristics. Tables 1 to 4 show a comparison of the general statistics for the assemblies in both "RAW" and "HQ" settings for both MetaFlye and Canu. No gaps were present in any of the contigs.

Table 1 – Assembly statistics of the contigs derived from MetaFlye “RAW”

barcode	#contigs	N50	Max contig	Min contig	Total length	Average contig
2	5	5074	5074	522	11169	2234
3	8	3501	3606	507	14542	1818
4	6	3601	3601	1037	11285	1881
5	4	3679	3679	471	7450	1863
6	5	3673	3673	1091	8582	1716
7	6	3665	3665	583	9951	1659
8	3	533	3602	533	6750	2250
9	6	2698	2698	585	7468	1245

Sources: Developed by the Author. Data generated with MetaFlye (Kolmogorov *et al.*, 2020).

Table 2 – Assembly statistics of the contigs derived from Canu “RAW”

barcode	#contigs	N50	Max contig	Min contig	Total length	Average contig
2	5	5074	5074	522	11169	2234
3	8	3501	3606	507	14542	1818
4	6	3601	3601	1037	11285	1881
5	6	3679	3679	627	11969	1995
6	5	3673	3673	1091	8582	1716
7	6	3665	3665	583	9951	1659
8	3	533	3602	533	6750	2250
9	6	2698	2698	585	7468	1245

Sources: Developed by the Author. Data generated with Canu (Koren *et al.*, 2017).

Table 3 – Assembly statistics of the contigs derived from MetaFlye "HQ"

barcode	#contigs	N50	Max contig	Min contig	Total length	Average contig
2	2	637	3721	637	4358	2179
3	3	620	4222	620	5916	1972
4	5	3905	3905	773	11558	2312
5	2	3679	6842	3679	10521	5261
6	4	3518	3518	400	8074	2019
7	5	3665	3665	683	7819	1564
8	5	3602	3602	1294	11567	2313
9	2	2697	4338	2697	7035	3518

Sources: Developed by the Author. Data generated with MetaFlye (Kolmogorov *et al.*, 2020).

Table 4 – Assembly statistics of the contigs derived from Canu "HQ"

barcode	#contigs	N50	Max contig	Min contig	Total length	Average contig
2	1	1924	1924	1924	1924	1924
3	2	580	3559	580	4139	2070
4	1	2738	2738	2738	2738	2738
5	2	1716	3636	1716	5352	2676
6	1	3513	3513	3513	3513	3513
7	1	3567	3567	3567	3567	3567
8	2	1664	3455	1664	5119	2560
9	2	631	3564	631	4195	2098

Sources: Developed by the Author. Data generated with Canu (Koren *et al.*, 2017).

Interestingly, the BLASTn run of the contigs remained consistent between Cany and MetaFlye. Almost every match found in the NCBI core nucleotide database with relevant similarity corresponded to plasmids (especially *Acinetobacter baumannii*, *Escherichia coli,* and *Synechocystis* sp. plasmids). Notably, many contigs found no matches whatsoever, which could correspond to unidentified viral species not present in the NCBI database. For the MetaFlye assembler, 31 of the 43 contigs generated with the "RAW" setting didn't match with any sequence in the NCBI core nucleotide database, while 16 of the 28 contigs generated with the "HQ" setting also didn't match with any sequence. No contigs matched with reliable sequence

coverage and similarity with any mycovirus, the vast majority of the “successful” matches being with bacteriophages which is not necessarily unheard of, once virome data is often underrepresented and incorrectly identified even in the big databases available, especially for mycoviruses (Sutton *et al*., 2019). The BLASTn results for the contigs generated with the MetaFlye assembler, both with the “RAW” and “HQ” settings can be found in Tables 5 and 6 respectively.

Table 5 – BLAST results for MetaFlye contigs with the “RAW” setting

Contig barcode	Match name	Query Cover	E value	Perc. ID
Contig1 Barcode2	*Mycolicibacterium novocastrense*	6.00%	2.00E-23	100.00%
	Bacillus phage FI KG-Lek	5.00%	9.00E-23	100.00%
	Saccharomyces cerevisiae	4.00%	2.00E-14	100.00%
Contig6 Barcode2	*Acinetobacter baumannii plasmid*	96.00%	0	99.42%
	E. coli	94.00%	0	99.91%
	Synechocystis sp. *Plasmid*	94.00%	0	99.91%
	Klebsiella quasipneumoniae	94.00%	0	99.80%
Contig8 Barcode3	*Acinetobacter baumannii plasmid*	100.00%	0	99.80%
	Synechocystis sp. *Plasmid*	98.00%	0	99.94%
	E. coli	98.00%	0	99.89%
	Staphylococcus aureus	98.00%	0	99.86%
	Klebsiella quasipneumoniae plasmid	98.00%	0	99.80%
Contig1 Barcode4	*Acinetobacter baumannii plasmid*	99.00%	0	99.33%
	E. coli plasmid	98.00%	0	99.92%
	Synechocystis sp. *plasmid*	98.00%	0	99.92%
	Straphylococcus aureus	98.00%	0	99.83%
	Klebsiella quasipneumoniae	98.00%	0	99.77%

Contig1 Barcode5	*Acinetobacter baumannii plasmid*	98.00%	0	99.42%
	Synechocystis sp. *plasmid*	96.00%	0	99.89%
	Staphylococcus aureus	96.00%	0	99.86%
	E. coli	96.00%	0	99.83%
	Klebsiella quasipneumoniae	96.00%	0	99.80%
Contig1 Barcode6	*Acinetobacter baumannii plasmid*	98.00%	0	99.39%
	E. coli	96.00%	0	99.97%
	Synechocystis sp. *plasmid*	96.00%	0	99.97%
	Staphylococcus aureus	96.00%	0	99.83%
	Klebsiella quasipneumoniae	96.00%	0	99.78%
Contig5 Barcode6	*Botrytis cinerea*	5.00%	3.00E-22	100.00%
	White spot syndrome virus	5.00%	2.00E-19	98.44%
	Mycolicibacterium novocastrense	5.00%	3.00E-17	94.20%
	Salmonella enterica	4.00%	3.00E-17	100.00%
	Helicoverpa zea	4.00%	4.00E-16	100.00%
	Acinetobacter baumannii plasmid	3.00%	3.00E-12	100.00%
	Trichoderma virens	3.00%	5.00E-05	97.37%
Contig5 Barcode7	*Staphylococcus phage*	2.00%	2.00E-10	97.92%
	Acinetobacter phage	2.00%	2.00E-10	97.92%
	Mucor sp.	2.00%	2.00E-10	97.92%
Contig7 Barcode7	*E. coli*	96.00%	0	99.94%
	Synechocystis sp. Plasmid	96.00%	0	99.94%
	Acinetobacter baumannii plasmid	98.00%	0	99.34%
Contig1 Barcode8	*E. coli*	98.00%	0	99.94%
	Synechocystis sp. *Plasmid*	98.00%	0	99.94%
	Acinetobacter baumannii plasmid	98.00%	0	99.75%

Contig1 Barcode9	*Mycolicibacterium novocastrense*	7.00%	8.00E-17	95.31%
	Amoebozoa sp.	5.00%	1.00E-15	100.00%
	Unculturate Ureaplasma sp.	4.00%	5.00E-09	100.00%
	Unculturate Parasutterella sp.	4.00%	7.00E-08	100.00%
	Mucor sp.	3.00%	3.00E-06	100.00%
	Plasmodium berghei	3.00%	1.00E-05	100.00%
Contig6 Barcode9	*Acinetobacter baumannii plasmid*	98.00%	0	99.89%
	E. coli	97.00%	0	100.00%
	Synechocystis sp. *Plasmid*	97.00%	0	100.00%
	Staphylococcus aureus	97.00%	0	99.89%
	Klebsiella quasipneumoniae	97.00%	0	99.89%

Source: Developed by the author. Data generated with BLASTn (Altschul *et al.*, 1990) through the Blobtoolskit pipeline (Challis, R. *et al*., 2020).

Table 6 – BLAST results for MetaFlye contigs with the "HQ" setting

Contig barcode	match name	Query Cover	E value	Perc. ID
Contig1 Barcode2	*Acinetobacter baumannii plasmid*	96.00%	0	99.41%
	E. coli	94.00%	0	99.89%
	Synechocystis sp. *Plasmid*	94.00%	0	99.89%
	Staphylococcus aureus	94.00%	0	99.83%
	Klebsiella quasipneumoniae	94.00%	0	99.77%
Contig1 Barcode3	*Acinetobacter baumannii plasmid*	86.00%	0	99.78%
	Synechocystis sp. *Plasmid*	84.00%	0	99.92%
	E. coli	84.00%	0	99.89%
	Staphylococcus aureus	84.00%	0	99.83%
	Klebsiella quasipneumoniae	84.00%	0	99.80%

Contig1 Barcode4	*Acinetobacter baumannii plasmid*	99.00%	0	99.33%
	E. coli	98.00%	0	99.92%
	Synechocystis sp. *Plasmid*	98.00%	0	99.92%
	Staphylococcus aureus	98.00%	0	99.83%
	Klebsiella quasipneumoniae	98.00%	0	99.77%
Contig1 Barcode5	*Acinetobacter baumannii plasmid*	98.00%	0	99.42%
	Synechocystis sp. *Plasmid*	96.00%	0	99.89%
	Staphylococcus aureus	96.00%	0	99.86%
	E. coli	96.00%	0	99.83%
	Klebsiella quasipneumoniae	96.00%	0	99.80%
Contig2 Barcode5	*Mucor* sp.	1.00%	5.00E-07	100.00%
	Mycolicibacterium novocastrense	1.00%	6.00E-06	90.57%
	Turicimonas sp. *Clone*	1.00%	2.00E-05	97.50%
	Klebsiella phage	1.00%	3.00E-04	97.30%
	Klebsiella pneumoniae	1.00%	1	97.22%
Contig4 Barcode6	*E. coli*	100.00%	0	99.94%
	Synechocystis sp. *Plasmid*	100.00%	0	99.94%
	Xanthomonas arboricola	100.00%	0	99.86%
	Staphylococcus aureus	100.00%	0	99.80%
	Arcanobacterium hippocoleae plasmid	100.00%	0	99.86%
Contig1 Barcode7	*Mucor* sp.	3.00%	1.00E-22	100.00%
	Staphylococcus phage	3.00%	2.00E-15	96.67%

	Acinetobacter phage	2.00%	1.00E-17	100.00%
	Acinetobacter baumannii plasmid	2.00%	8.00E-05	95.00%
Contig6 Barcode7	*Acinetobacter baumannii plasmid*	98.00%	0	99.34%
	E. coli	96.00%	0	99.94%
	Synechocystis sp. *Plasmid*	96.00%	0	99.94%
	Staphylococcus aureus	96.00%	0	99.80%
	Arcanobacterium hippocoleae	96.00%	0	99.86%
	Klebsiella quasipneumoniae	96.00%	0	99.75%
Contig4 Barcode8	*Plasmodium berghei ANKA*	2.00%	1.00E-14	98.18%
	E. coli	2.00%	2.00E-12	96.30%
Contig5 Barcode8	*Acinetobacter baumannii plasmid*	99.00%	0	99.75%
	E. coli	98.00%	0	99.94%
	Synechocystis sp. *Plasmid*	98.00%	0	99.94%
	Staphylococcus aureus	98.00%	0	99.80%
	Arcanobacterium hippocoleae plasmid	98.00%	0	99.86%
	Klebsiella quasipneumoniae	98.00%	0	99.75%
Contig1 Barcode9	*Acinetobacter baumannii plasmid*	98.00%	0	99.89%
	E. coli	97.00%	0	100.00%
	Synechocystis sp. *Plasmid*	97.00%	0	100.00%
	Staphylococcus aureus	97.00%	0	99.89%
	Klebsiella quasipneumoniae	97.00%	0	99.85%
	Arcanobacterium hippocoleae plasmid	97.00%	0	99.85%

	Mycolicibacterium novocastrense	3.00%	1.00E-15	94.03%
	Amoebozoa sp.	2.00%	5.00E-15	100.00%
Contig2	*Unculturated Ureaplasma clone*	1.00%	2.00E-08	100.00%
Barcode9	*Unculturated Parasutterella clone*	1.00%	3.00E-07	100.00%
	Unculturated Pseudomonas clone	1.00%	4.00E-06	100.00%
	Mucor sp.	1.00%	1.00E-05	100.00%

Source: Developed by the author. Data generated with BLASTn (Altschul *et al.*, 1990) through the Blobtoolskit pipeline (Challis, R. *et al*., 2020).

Further analysis using the contigs generated by MetaFlye with the kraken2 software provided some information about the taxonomy of the contigs present in the database provided for the pipeline. For the contigs generated with the “HQ” setting, at least one sequence from each group of barcodes (from 2 to 9) was classified as *Escherichia* phage Lambda, a bacteriophage. For the contigs generated with the “RAW” setting, a similar situation was observed, except the 7th barcode, which in addition to one contig classified as *E.* phage Lambda, also had a contig classified as the Sida golden mottle virus, a Begomovirus known to infect *Sida santaremensis,* a plant native to South America from the *Malvaceae* family.

Overall, the metagenomic analysis classified much of the samples as genetic material from bacteriophages, especially Lambda viruses that commonly infect *E. coli*, although many other lineages also have a high percentage identity and coverage, such as *Staphylococcus* and *Klebsiella*. Many sequences matching plasmids were also found among the matches with the highest coverage and identity. However, the majority of the contigs across the barcode groups were once more branded “Unclassified” (more than 60% of the contigs), which could represent novel viral sequences, or even known viral sequences that are underrepresented in public databases, as has been the case in past studies (Sutton *et al*., 2019), though it is important to also ponder the possibility of these being the results of sequencing errors.

5 CONCLUSION

The metagenomic data analyzed was partially identified with no known mycovirus detected among the sequences available on the public databases. The distinct presence of bacteriophage sequences in the virome was unexpected, and the possibility that those sequences could be derived from contamination is not null. Still, the similarity and coverage, as well as the precision in the sequence matches with known sequences could justify further studies on the matter. However, most contigs generated were deemed unclassified, which is not necessarily uncommon for *de novo* assemblies. The number of unidentified contigs (which showed to be more than 60% of the total number of contigs generated) serves as an example of the lack of information on the diversity of environment samples, thus the necessity of more studies taking the *de novo* approaches for further confirmation of novel species, as well as correct identification of their taxonomy.

REFERENCES

AHN, I. P.; LEE, Y. H. A viral double-stranded RNA up-regulates the fungal virulence of Nectria radicicola. **Molecular plant-microbe interactions: MPMI**, [*s. l.*], v. 14, n. 4, p. 496–507, 2001.

AL-AQEEL, H. A.; IQBAL, Z.; POLSTON, J. E. Characterization of sida golden mottle virus isolated from Sida santaremensis Monteiro in Florida. **Archives of Virology**, [*s. l.*], v. 163, n. 10, p. 2907–2911, 2018.

ALTSCHUL, S. F. *et al.* Basic local alignment search tool. **Journal of Molecular Biology**, [*s. l.*], v. 215, n. 3, p. 403–410, 1990.

BETTS, M. G. *et al.* When are hypotheses useful in ecology and evolution?. **Ecology and Evolution**, [*s. l.*], v. 11, n. 11, p. 5762–5776, 2021.

BORODYNKO, N. *et al.* La France disease of the cultivated mushroom Agaricus bisporus in Poland. **Acta Virologica**, [*s. l.*], v. 54, n. 3, p. 217–219, 2010.

BUCHFINK, B.; XIE, C.; HUSON, D. H. Fast and sensitive protein alignment using DIAMOND. **Nature Methods**, [*s. l.*], v. 12, n. 1, p. 59–60, 2015.

CHALLIS, R. *et al.* BlobToolKit – Interactive Quality Assessment of Genome Assemblies. **G3 Genes|Genomes|Genetics**, [*s. l.*], v. 10, n. 4, p. 1361–1374, 2020.

CORBU, V. M.; *et al.* Current Insights in Fungal Importance-A Comprehensive Review. **Microorganisms**, v. 11, n. 6, 2023.

DOUGLAS, A. E. Conflict, cheats and the persistence of symbioses. **New Phytologist**, [*s. l.*], v. 177, n. 4, p. 849–858, 2008.

DOUGLAS, A. E. Host benefit and the evolution of specialization in symbiosis. **Heredity**, [*s. l.*], v. 81, n. 6, p. 599–603, 1998.

FRANK, S. A. Host-symbiont conflict over the mixing of symbiotic lineages. **Proceedings. Biological Sciences**, [*s. l.*], v. 263, n. 1368, p. 339–344, 1996.

GENERATION OF NUTRIENTS AND DETOXIFICATION: POSSIBLE ROLES OF YEASTS IN LEAF-CUTTING ANT NESTS. [*S. l.*], [*s. d.*]. Disponível em: https://www.mdpi.com/2075-4450/3/1/228. Acesso em: 4 out. 2024.

GHABRIAL, S. A. *et al.* 50-plus years of fungal viruses. **Virology**, [*s. l.*], v. 479–480, p. 356–368, 2015.

JANZEN, D. The natural history of mutualisms. [*s. l.*], Disponível em: https://www.academia.edu/30073403/The_natural_history_of_mutualisms. Acesso em: 4 out. 2024.

KEELER, KH. Cost-benefit models of mutualism. *In*: THE BIOLOGY OF MUTUALISM. New York, USA: Bucher DH, 1985. p. 100–127.

KNIGHT, R. *et al.* Unlocking the potential of metagenomics through replicated experimental design. **Nature Biotechnology**, [*s. l.*], v. 30, n. 6, p. 513–520, 2012.

KOLMOGOROV, M. *et al.* metaFlye: scalable long-read metagenome assembly using repeat graphs. **Nature Methods**, [*s. l.*], v. 17, n. 11, p. 1103–1110, 2020.

KOOIJ, P. W. **Fungal adaptations to mutualistic life with ants**. 2013. Tese de Doutorado - University of Copenhagen, Faculty of Science, Department of Biology, Centre for Social Evolution, 2013.

KOREN, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. **Genome Research**, [*s. l.*], v. 27, n. 5, p. 722–736, 2017.

KWON, Y. C. *et al.* Curing viruses in *Pleurotus ostreatus* by growth on a limited nutrient medium containing cAMP and rifamycin. **Journal of Virological Methods**, [*s. l.*], v. 185, n. 1, p. 156–159, 2012.

LAUBER, C.; SEITZ, S. Opportunities and Challenges of Data-Driven Virus Discovery. **Biomolecules**, [*s. l.*], v. 12, n. 8, p. 1073, 2022.

LAW, R.; LEWIS, D. H. Biotic environments and the maintenance of sex–some evidence from mutualistic symbioses. **Biological Journal of the Linnean Society**, [*s. l.*], v. 20, n. 3, p. 249–276, 1983.

LEIGH JR, E. G. The evolution of mutualism. **Journal of Evolutionary Biology**, [*s. l.*], v. 23, n. 12, p. 2507–2528, 2010.

LUTZONI, F.; MIADLIKOWSKA, J. Lichens. **Current biology: CB**, [*s. l.*], v. 19, n. 13, p. R502-503, 2009.

MÁRQUEZ, L. M. *et al.* A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance. **Science (New York, N.Y.)**, [*s. l.*], v. 315, n. 5811, p. 513–515, 2007.

MARTIARENA, M. J. S. *et al.* The Hyphosphere of Leaf-Cutting Ant Cultivars Is Enriched with Helper Bacteria. **Microbial Ecology**, [*s. l.*], v. 86, n. 3, p. 1773–1788, 2023.

MÖLLER, A. **Botanische Mittheilungen aus den Tropen**. 6. ed. Oxford University: [*s. n.*], 1898.

MUELLER, U. G. *et al.* The Evolution of Agriculture in Insects. **Annual Review of Ecology, Evolution, and Systematics**, [*s. l.*], v. 36, n. Volume 36, 2005, p. 563–595, 2005.

MYERS, J. M.; JAMES, T. Y. Mycoviruses. **Current Biology**, [*s. l.*], v. 32, n. 4, p. R150–R155, 2022.

PARTIDA-MARTINEZ, L. P. *et al.* Endosymbiont-Dependent Host Reproduction Maintains Bacterial-Fungal Mutualism. **Current Biology**, [*s. l.*], v. 17, n. 9, p. 773–777, 2007a.

PARTIDA-MARTINEZ, L. P. *et al.* Endosymbiont-Dependent Host Reproduction Maintains Bacterial-Fungal Mutualism. **Current Biology**, [*s. l.*], v. 17, n. 9, p. 773–777, 2007b.

PAYSAN-LAFOSSE, T. *et al.* InterPro in 2022. **Nucleic Acids Research**, [*s. l.*], v. 51, n. D1, p. D418–D427, 2023.

PEARSON, M. N. *et al.* Mycoviruses of filamentous fungi and their relevance to plant pathology. **Molecular Plant Pathology**, [*s. l.*], v. 10, n. 1, p. 115–128, 2009.

SCHMITT, M. J.; BREINIG, F. Yeast viral killer toxins: lethality and self-protection. **Nature reviews microbiology**. p. 212-221. 2006.

SCHULTZ, T. R. *et al.* The coevolution of fungus-ant agriculture. **Science**. v. 386, n. 6717. p. 105-110. 2024.

SCOTT, J. J. *et al.* Bacterial Protection of Beetle-Fungus Mutualism. **Science**, [*s. l.*], v. 322, n. 5898, p. 63–63, 2008.

SUTTON, T. D. S. *et al.* Choice of assembly software has a critical impact on virome characterization. **Microbiome**, [*s. l.*], v. 7, n. 1, p. 12, 2019.

THE GEOGRAPHIC MOSAIC OF COEVOLUTION IN MUTUALISTIC NETWORKS | PNAS. [*S. l.*], [*s. d.*]. Disponível em: https://www.pnas.org/doi/full/10.1073/pnas.1809088115. Acesso em: 4 out. 2024.

VILLAN, D. C. *et al.* Exploring the Mycovirus Universe: Identification, Diversity, and Biotechnological Applications. **Journal of Fungi**. v. 9, n. 3. 2023

WEBER, N. A. Pure Cultures of Fungi Produced by Ants. **Science (New York, N.Y.)**, [*s. l.*], v. 121, n. 3134, p. 109, 1955.

WHEELER, W. M. **The fungus-growing ants of North America**. American Museum of Natural History, 1907.

WOOD, D. E.; LU, J.; LANGMEAD, B. Improved metagenomic analysis with Kraken 2. **Genome Biology**, [*s. l.*], v. 20, n. 1, p. 257, 2019a.

WOOD, D. E.; LU, J.; LANGMEAD, B. Improved metagenomic analysis with Kraken 2. **Genome Biology**, [*s. l.*], v. 20, n. 1, p. 257, 2019b.

WOOD, D. E.; SALZBERG, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. **Genome Biology**, [*s. l.*], v. 15, n. 3, p. R46, 2014.

YU, H. J.; LIM, D.; LEE, H.-S. Characterization of a novel single-stranded RNA mycovirus in Pleurotus ostreatus. **Virology**, [*s. l.*], v. 314, n. 1, p. 9–15, 2003.