

Genome mining and characterisation of biosynthetic clusters in *Aspergillus japonicus* isolated from the Amazon rainforest

Josy Caldas Rodrigues^{1/+}, Maria Eduarda Grisolia¹, Alice da Silva Queiroz¹, Ana Luísa Rodrigues Lima², Clarice Virginia Santos Goiabeira¹, Leila de Mendonça Lima³, Ormeizinda Celeste Cristo Fernandes¹

¹Fundação Oswaldo Cruz-Fiocruz, Instituto Leônidas & Maria Deane, Manaus, AM, Brasil

²Universidade Federal do Amazonas, Manaus, AM, Brasil

³Fundação Oswaldo Cruz-Fiocruz, Instituto Oswaldo Cruz, Rio de Janeiro, RJ, Brasil

BACKGROUND In light of the biotechnological potential demonstrated by *Aspergillus* species and, considering the great need for further research into the search for new sources of active molecules and the biodiversity of these microorganisms occurring in the Amazon region.

OBJECTIVES This research aimed to investigate the biotechnological potential of the fungus *Aspergillus japonicus* Amazon Fungi Collection (CFAM) 0234, a fungal strain isolated from Amazonian soil and stored in the CFAM.

METHODS For this purpose, the *Aspergillus* species was investigated through comparative genomic analysis and antimicrobial activity assays.

FINDINGS Genome sequencing revealed a fragmented assembly (72.67 Mbp, N50 = 152 kbp) containing 106 biosynthetic clusters (BGCs), surpassing the reference strain CBS 114.51 (57 BGCs). Among the clusters identified, NRPS, PKS type I and hybrid NRPS-PKS systems stood out, including clusters exclusive to betalactones and isocyanides, potentially involved in the synthesis of β -lactam antibiotics and innovative metabolites. BiG-SCAPE analysis identified 63 BGC families unique to CFAM 0234, suggesting evolutionary adaptations to the competitive environment of the Amazon. Biological assays demonstrated selective antimicrobial activity of the ethyl acetate extract against *Escherichia coli*, *Shigella sonnei* and *Staphylococcus aureus* (MRSA), with inhibition halos ranging from 8 mm to 6 mm in diameter, pathogens classified as priorities for research into new antibiotics. The correlation between predicted BGCs and antimicrobial activity reinforces the strain's biotechnological potential. Despite the fragmentation of the genome, the high completeness assessed by BUSCO (98.5%) confirms the quality of the assembly, while the detection of single nucleotide polymorphisms (SNPs) in regulatory regions and rearrangements close to BGCs suggests evolutionary pressure for metabolic diversification. The lack of correspondence with the minimum information about a biosynthetic gene cluster (MIBiG) bank and the limitations of crude extracts highlight the need for complementary techniques, such as long-read sequencing (Oxford Nanopore) and metabolomic analysis [liquid chromatography-mass spectrometry (LC-MS)], to link clusters to active metabolites.

MAIN CONCLUSIONS *Aspergillus japonicus* CFAM 0234 represents a promising microbial resource for bioprospecting in the Amazon, offering relevant genomic and chemical insights for the development of new antimicrobial agents. Future studies will focus on the purification of compounds and activation of silent BGCs, aiming at sustainable pharmaceutical applications.

Key words: bioprospecting - Amazon - *Aspergillus japonicus* - biosynthetic clusters - antimicrobial resistance

One of the great desires related to the Amazon region is associated with the possibility of curing various diseases through the development of medicines sourced from the region's biodiversity.⁽¹⁾

In this sense, studies involving the bioprospecting of fungi of the genus *Aspergillus* from Amazonian soil have shown that these microorganisms are an alternative source in the search for new active molecules capable of inhibiting microorganisms such as *Candida albicans*,

Escherichia coli and *Staphylococcus aureus* that are recognised for causing hospital infections and presenting resistance to available drugs.^(2,3,4)

However, elucidating the secondary metabolites responsible for these desired activities is still a challenge - many studies are carried out using organic extracts and fractions due to the financial and technical difficulties involved in isolating and identifying these metabolites, thus hampering the development of innovative products.⁽⁵⁾

doi: 10.1590/0074-02760250075

Financial support: FAPEAM (PRODOC - Post-Doctorate Support Program), ILMD/FIOCRUZ.

CR and MEG contributed equally to this work.

+ Corresponding author: josy.silva@fiocruz.br |  <https://orcid.org/0000-0002-7886-7929>

Received 01 April 2025

Accepted 14 November 2025

How to cite: Rodrigues JC, Grisolia ME, Queiroz AS, Lima ALR, Goiabeira CVS, Lima LM, et al. Genome mining and characterisation of biosynthetic clusters in *Aspergillus japonicus* isolated from the Amazon rainforest. Mem Inst Oswaldo Cruz. 2025; 120: e250075.

Handling editor: Adeilton Alves Brandão |  <https://orcid.org/0000-0001-5877-607X>



Sequencing combined with genome mining using bioinformatics algorithms enables the identification of groups of secondary metabolism genes. As an example, citric acid fermentation by *Aspergillus niger* has been studied for almost 100 years, but the elucidation of the central carbon metabolism involved in the process was only possible with the availability of genome data.⁽⁶⁾

Therefore, analysis of the genome of microorganisms provides a better understanding of their fundamental cellular processes and metabolic potential, contributing to solving urgent and important challenges in health, enzyme biotechnology, bioenergy and ecological diversity. The availability of whole genomes allows the search for bioactives with enhanced properties and provides invaluable help to improve the production of biotechnological products.⁽⁷⁾

Based on the above considerations and the great need for new sources of active molecules, the aim of this work was to investigate the biotechnological potential, based on comparative genomic analysis, of the fungus *A. japonicus* stored in the Amazon Fungi Collection (CFAM).

MATERIALS AND METHODS

Acquisition of strains, DNA extraction and sequencing - The fungal strain used in this study was selected at the Multiuser Health/Mycology Laboratory of the Leônidas and Maria Deane Institute (ILMD/FIOCRUZ Amazônia). Genomic sequencing was conducted at the Genomics Technology Platform, linked to the Technological Development Program in Health Supplies (PD-TIS), under the coordination of the Vice-presidency of Research and Reference Laboratories at FIOCRUZ.

The microorganisms used as tests were supplied by the Amazon Bacteria Collection (CBAM), the CFAM and the INPA Collection of Microorganisms of Medical Interest. For this study, the *A. japonicus* strain (CFAM 0234), isolated from the soil of the Amazon biome and stored in the CFAM-ILMD (FIOCRUZ), was selected. The strain was reactivated in Petri dishes containing malt extract agar (MEA) culture medium and incubated at 28°C for seven days in a BOD incubator.⁽⁸⁾

Genomic DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen), which is widely used to obtain high-quality DNA from fungi. The quality of the extracted DNA was assessed using 1.0% agarose gel electrophoresis, stained with GelRed, and quantification was carried out using spectrophotometry. The genomic DNA was prepared and sequenced on the Next Generation Nucleic Acid Sequencing Platform (FIOCRUZ/IOC - RJ).

The genomic libraries were prepared using the Illumina DNA Prep protocol, with a fragmentation time of 15 to 20 minutes at 55°C, in order to optimise the yield due to the characteristics of the fungal DNA.

Sequencing was carried out on the Illumina NextSeq2000 platform using the P3 300-cycle kit, which generates approximately 1.2 billion reads per run. A negative control (H₂O) was included to monitor the quality of the library construction process.

Genome assembly, quality assessment and annotation - Raw reads were assessed for quality with FastQC v0.11.8 and processed with Trimmomatic (parameters:

SLIDINGWINDOW:4:20 MINLEN:50) to remove adapters, low-quality sequences and truncated reads.

⁽⁹⁾ The assembly was carried out with SPAdes v3.13,⁽¹⁰⁾ using a hybrid strategy that combined *trusted* contigs, based on the incomplete genome of *A. japonicus* CBS 114.51 (scaffold level, N50 = 412 kbp; NCBI accession: GCA_003184785.1), and *untrusted* contigs, generated by an initial *novel* assembly with corrected reads. The quality of the assembly was assessed with QUAST v5.0.2,⁽¹¹⁾ for analysis of fragmentation metrics (N50, number of contigs), and compared with other *A. japonicus* strains (strains Y4009A, accession GCA_023625335.1, and PSFR, accession GCA_016808025.1). Genomic completeness was checked with BUSCO v5.4.3,⁽¹²⁾ using the *aspergillus_odb12* database.

To identify orthologous genes, the analysis was conducted with OrthoFinder v2.5.4,⁽¹³⁾ using predicted proteomes from *A. japonicus* CFAM 0234 (annotated with Funannotate v1.8.1) and CBS 114.51 (NCBI public annotation). The aim was to identify groups of conserved and unique genes in order to explore functional and evolutionary differences between the strains.

Analysis of biosynthetic clusters (BGCs) and genomic comparison - The prediction of BGCs was carried out with antiSMASH v7.0,⁽¹⁴⁾ using the detection *strictness* parameter to identify enzyme systems associated with the production of secondary metabolites, including polyketide synthases (T1PKS, T3PKS), non-ribosomal peptides (NRPS) and hybrid systems (NRPS-PKS, terpenes, siderophores). To group the BGCs into gene cluster families (GCFs) and compare profiles between the strains, the GenBank files of the predicted BGCs (CFAM0234 and CBS 114.51) were analysed with BiG-SCAPE v1.1.5,⁽¹⁵⁾ the process is now being scaled up to mine entire genera, strain collections and microbiomes. However, no bioinformatic framework is currently available for effectively analyzing datasets of this size and complexity. In the present study, a streamlined computational workflow is provided, consisting of two new software tools: the 'biosynthetic gene similarity clustering and prospecting engine' (BiG-SCAPE using a similarity threshold of 30% (default value for conservative clustering) and Jaccard's distance to calculate similarity.

The genomic alignment between *A. japonicus* CFAM 0234 and the reference strain CBS114.51 was performed with MUMmer4.⁽¹⁶⁾ Parameters such as *-maxmatch* (for identifying all exact matches) and *-c 1000* (filtering out short alignments) were used. The raw results (delta files) were refined with the *delta-filter* tool, retaining only regions with identity $\geq 99\%$ and length ≥ 500 kb. The coordinates of the alignments were extracted with *show-coords*, generating a tabular file with genomic positions, identity and orientation of the sequences. To detect single nucleotide polymorphisms (SNPs), *show-snps* was used with the *-C* parameter (exclusion of repetitive regions), followed by manual filtering to remove variants in regions of low complexity or close to gaps.

The distribution of biosynthetic clusters was analysed quantitatively, comparing the relative abundance of each class among the strains. For integrated visual-

isation, BGCs, SNPs and structural rearrangements were mapped on the genome with Circos v0.69-9.⁽¹⁷⁾ The alignment data was processed in R, converting coordinates to the appropriate format for the Circos *karyotype*, and plotted in circular graphs highlighting the location of regions of metabolic interest and genomic variations.

Experimental validation: biological tests - To confirm the production of the metabolites with antimicrobial action, biological activity tests were carried out.

Extraction of secondary metabolites - The fungal culture was grown in 500 mL Erlenmeyer flasks containing 200 mL of YES culture medium [Czapeck yeast extract 2% (w/v) and sucrose 20% (w/v)], incubated at 28°C for seven days. After growth, the secondary metabolites were cold extracted with 200 mL of P.A. grade ethyl acetate (AcOEt) (SYNTH). The AcOEt extract was kept in contact with the culture for 48 h, filtered on Whatman No. 30 paper and concentrated in an exhaust hood (SCIOLOGEX RE 100-Pro) to assess antimicrobial activity.⁽¹⁸⁾

Confirmation of antibacterial activity - The antibacterial activity of the extract of *A. japonicus* CFAM 0234 was evaluated by the agar diffusion technique per well, using pathogenic bacteria of clinical origin frequently associated with hospital infections (Table I). The bacterial strains selected included *Acinetobacter nosocomialis*, *Enterococcus faecium*, *E. coli*, *Klebsiella pneumoniae*, *S. aureus* (MRSA) and others (Table I).

The bacteria were seeded on Petri dishes (10 mm × 90 mm) containing Mueller-Hinton agar and incubated at 37°C for 24 h. From the cultures, a cell suspension adjusted to turbidity equivalent to the McFarland 0.5 scale was prepared. Next, 150 µL of the suspension was sown evenly on the surface of Mueller-Hinton agar. Three 5 mm di-

ameter wells were drilled in each plate, into which 150 µL of the fungal extract (1 mg/mL) were added, solubilised in sterile distilled water and 4% (v/v) dimethyl sulfoxide (DMSO). As controls, vancomycin (1 mg/mL; Merck, Germany) was used as a positive and a mixture of sterilised distilled water/DMSO 4% (v/v) as a negative. The plates were incubated at 37°C for 24 h and antibacterial activity was determined by measuring the inhibition halo (in millimetres). All tests were carried out in triplicate.⁽¹⁸⁾

RESULTS

Genome sequencing - Sequencing the genome of *A. japonicus* CFAM 0234 resulted in an assembly of 962 contigs, with a total size of 72.67 Mbp and an N50 value of 152 kbp. The GC content was 50.45%, and the largest contig reached 600 kbp. The *CBS 114.51* reference genome, available on NCBI, is also not complete, being at scaffold level, with a total size of 36.1 Mbp and an N50 of 412 kbp.

The comparative analysis with CBS 114.51 should be interpreted with caution, since both assemblies (CFAM 0234 and CBS 114.51) are fragmented. The apparent doubling of the size of the CFAM 0234 genome (72.67 Mbp vs. 36.1 Mbp) may reflect both biological differences and assembly artifacts, since regions missing from the reference could not be validated. On the other hand, BUSCO's assessment of genomic integrity indicated that the assembled genome shows a high level of completeness, with 4959 complete genes identified (339 single-copy and 4620 duplicated), as well as only 10 fragmented and 65 missing genes, out of a total of 5034 genes assessed.

Orthologue analysis using *OrthoFinder* compared the study strain with the CBS 114.51 reference genome. The total number of genes identified was 962 in the sequenced strain and 163 in the reference, with the majority grouped into orthologs (926 and 161, respectively).

TABLE I
Pathogenic bacteria of clinical origin used to evaluate antibacterial activity

Bacteria/CBAM	Antibiotic resistance	Clinical origin
<i>Acinetobacter nosocomialis</i> 0725	Cephalothin, oxacillin, penicillin and vancomycin	Oropharynx
<i>Enterococcus faecium</i> 0639	Amikacin, cephalothin, penicillin and oxacillin	Oropharynx
<i>Enterococcus faecalis</i> 0735	Cephalothin and oxacillin	Oral cavity
<i>Escherichia coli</i> 0001	Ampicillin, cephalothin, oxocillin, penicillin, tetracycline and vancomycin	Diarrheal stools
<i>Escherichia coli</i> 0002	Amikacin, ampicillin, cephalothin, cefotaxime, oxacillin, penicillin, piperacillin+tazobactam, tetracycline and vancomycin	Diarrheal stools
<i>Klebsiella aerogenes</i> 0717	Amoxicillin, cephalothin, oxacillin, penicillin and vancomycin	
<i>Klebsiella pneumoniae</i> 0462	Amoxicillin, oxacillin, penicillin and vancomycin	Oropharynx
<i>Proteus mirabilis</i> 0686	Amikacin, cephalothin, oxacillin and vancomycin	Diarrheal stools
<i>Pseudomonas aeruginosa</i> 0024	Amoxicillin, cephalothin, oxacillin and vancomycin	Diarrheal stools
<i>Salmonella enterica</i> 0047	Oxacillin and vancomycin	
<i>Salmonella typhi</i> 0009	Oxacillin and vancomycin	Skin lesion
<i>Shigella flexneri</i> 0046	Amoxicillin, cephalothin, oxacillin and vancomycin	Diarrheal stools
<i>Shigella sonnei</i> 0036	Amoxicillin, cephalothin, oxacillin and vancomycin	Diarrheal stools
<i>Staphylococcus aureus</i> ATCC 4330 (MRSA)	Methicillin	-

CBAM: Amazon Bacteria Collection.

Only 36 genes in the lineage studied and two in the reference were not assigned to orthogroups. Several contigs in the lineage analysed were associated with specific orthogroups, with genes distributed in contigs such as *Contig_591*, *Contig_606*, *Contig_681*, and others standing out, which may indicate regions of functional interest that have not yet been fully characterised.

Comparative analysis of BGCs - The CFAM 0234 strain revealed an expanded repertoire of BGCs, with 106 BGCs identified compared to 59 in the CBS 114.51 reference strain. BiG-SCAPE analysis using a 30% similarity threshold showed that 63 BGC families (GCFs) are unique to CFAM 0234, while 33 are unique to CBS 114.51 (Table II). The lack of correspondence with clusters from the minimum information about a biosynthetic

gene cluster (MIBiG) bank reinforces the unique biotechnological potential of these strains.

The comparative map generated by Circos (Figure A) highlights the association between structural rearrangements and the location of BGCs, particularly NRPS and T1PKS clusters. These regions showed the highest density of SNPs, suggesting evolutionary pressure for metabolic diversification. CFAM 0234 exhibits greater BGC complexity, with 34 NRPS clusters (vs. 15 in the reference) and 12 NRPS-T1PKS hybrid clusters (vs. 7), including unique systems such as betalactones and isocyanides (Figure B).

Among the 96 shared GCFs, clusters with conserved enzyme profiles stand out, such as FAM_00001 (T1PKS) and FAM_00009 (NRPS-T1PKS), which may represent essential metabolic pathways for the fungus' ecological niche (Table III). Exclusive GCFs, such as FAM_00096 (T1PKS-isocyanide) in CFAM 0234, are priority candidates for functional characterisation. Table III summarises the disparity in the distribution of BGC classes between the strains, highlighting exclusive clusters such as FAM_00096 (T1PKS + Isocyanide) in CFAM 0234 and shared ones such as FAM_00009 (NRPS + T1PKS), with potential for antifungal hybrid metabolites.

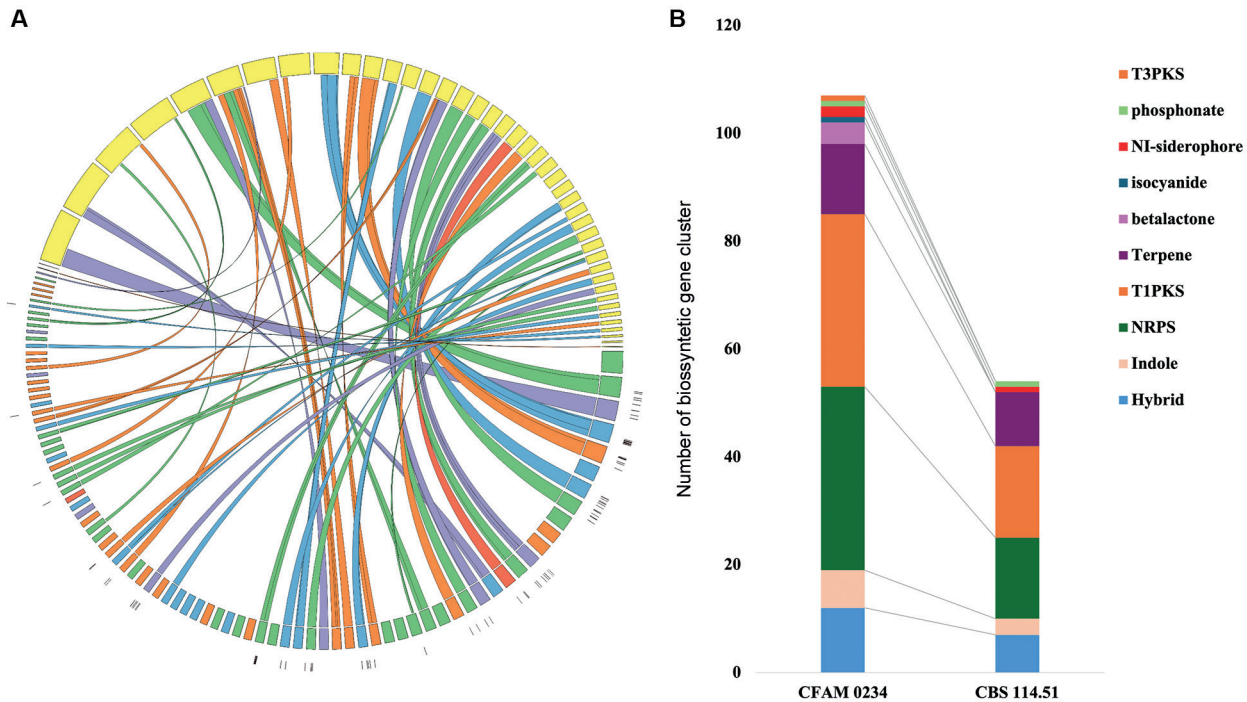
Antibacterial activity of the CFAM 0234 strain extract - The ethyl acetate extract (AcOEt) of CFAM 0234 selectively inhibited *E. coli* CBAM 0001, *Shigella sonnei* CBAM 0036 and *Staphylococcus aureus* ATCC 4330 (MRSA), with inhibition halos of 8 mm, 8 mm and 6 mm in diameter, respectively (Table IV).

TABLE II

Overview of biosynthetic gene clusters families (BGCs)

Métric	CFAM 0234	CBS 114.51	Both
Total GCFs	89	59	125
Exclusive GCFs*	63 (71%)	33 (56%)	-
Shared GCFs**	-	-	96 (77%)
Similarity to MIBiG	0%	0%	0%

*GCFs containing only BGCs from one strain; **families with BGCs from both strains (similarity > 30%); CFAM: Amazon Fungi Collection.



Comparative cluster analysis of biosynthesis genes in *Aspergillus japonicus*. (A) Circos representation of the alignment between strain Amazon Fungi Collection (CFAM) 0234 (coloured internal contigs) and reference strain CBS 114.51 (yellow external contigs). Colours in the CFAM 0234 contigs indicate biosynthetic clusters (BGCs) classes: green [non-ribosomal peptides (NRPS)], orange (T1PKS), light blue (NRPS+T1PKS hybrids), purple (terpenes), and red (siderophores). Links connect homologous regions; single nucleotide polymorphisms (SNPs) are highlighted in black. (B) Stacked bar graph comparing the abundance of BGCs between strains. CFAM 0234 shows greater diversity, with unique clusters (e.g., betalactones).

TABLE III

Synthesis of relevant classes of biosynthetic gene clusters (BGCs) and gene clusters families (GCFs) in *Aspergillus japonicus*

BGC class	Strain	Distribution (%)	Relevant GCFs	Features/Potential
NRPS	CFAM 0234	45% (48/106)	FAM_00002 (Shared)	Synthesis of non-ribosomal peptides
	CBS 114.51	38% (22/57)	FAM_00125 (CBS Exclusive)	Unusual enzyme domains
TIPKS	CFAM 0234	30% (32/106)	FAM_00001 (Shared)	Biosynthesis of polyketides (e.g. pigments)
	CBS 114.51	35% (20/57)	FAM_00096 (CFAM Exclusive)	TIPKS + Isocyanide hybrid cluster
Terpenes	CFAM 0234	12% (13/106)	FAM_00057 (CFAM Exclusive)	Production of uncharacterized terpenoids
	CBS 114.51	15% (9/57)	FAM_00125 (CBS Exclusive)	Atypical modular architecture
Hybrids (NRPS+TIPKS)	CFAM 0234	8% (9/106)	FAM_00009 (Shared)	Antifungal hybrid metabolites
	CBS 114.51	7% (4/57)	FAM_00125 (CBS Exclusive)	Rare addition domains (ex.: epoxidase)
Betalactones	CFAM 0234	4 clusters (Exclusive)	FAM_00024 (Shared)	Potential for β -lactam antibiotics
Indole/Other	CFAM 0234	5% (5/106)	FAM_00010 (CBS Exclusive)	Unmapped secondary roads
	CBS 114.51	5% (3/57)	FAM_00055 (CFAM Exclusive)	Absent in databases (evolutionary novelty)

TABLE IV

Determination of antibacterial activity using the agar diffusion technique per well

Extract	Bacteria tested (halo mm)													
<i>Aspergillus japonicus</i> 0234	<i>A. n</i> 0725	<i>E. f</i> 0639	<i>E. f</i> 0735	<i>E. c</i> 0001	<i>E. c</i> 0002	<i>K. a</i> 0717	<i>K. p</i> 0462	<i>P. m</i> 0686	<i>P. a</i> 0024	<i>S. e</i> 0047	<i>S. t</i> 0009	<i>S. f</i> 0046	<i>S. s</i> 0036	MRSA 4330
	-	-	-	8	-	-	-	-	-	-	-	-	8	6

A. n: *Acinetobacter nosocomialis*; *E. f*: *Enterococcus faecium*; *E. c*: *Escherichia coli*; *K. a*: *Klebsiella aerogenes*; *K. p*: *Klebsiella pneumoniae*; *P. m*: *Proteus mirabilis*; *P. a*: *Pseudomonas aeruginosa*; *S. e*: *Salmonella enterica*; *S. t*: *Salmonella typhi*; *S. f*: *Shigella flexneri*; *S. s*: *Shigella sonnei*; MRSA: *Staphylococcus aureus*.

DISCUSSION

The production of secondary metabolites is strongly influenced by culture conditions, and the present work was focused on exploring the genomic potential of *A. japonicus* CFAM 0234. The antimicrobial activity observed with the crude extract supports the hypothesis that some predicted biosynthetic clusters are active. Future work will aim to validate these predictions through metabolomic analyses under different growth conditions and extraction methodologies.

The relevance of the antimicrobial activity of *A. japonicus* CFAM 0234 against *E. coli*, *S. sonnei* and *S. aureus* (MRSA). Inhibition zones between 6 - 10 mm were considered indicative of moderate activity when compared with the reference antibiotic control. Although the agar disk diffusion assay is qualitative and does not allow precise quantification of the antimicrobial agent,⁽¹⁹⁾ it provides preliminary evidence of bioactivity. Determination of the minimum inhibitory concentration (MIC) will be necessary in future work to accurately quantify antimicrobial potency.

Recently, the World Health Organisation⁽²⁰⁾ classified these bacteria among the priority pathogens for research into new antibiotics, reinforcing the importance of exploring therapeutic alternatives. A study by Naddaf⁽²¹⁾ reveals that deaths linked to resistant gram-negative bacteria, such as *E. coli*, have increased by almost 150%, from 50,900 cases in 1990 to 127,000 cases in 2021.

Shigellosis, which causes around 700,000 deaths annually — mainly in children in low- and middle-income countries⁽²²⁾ — and the prevalence of *S. sonnei* in outbreaks in Brazil,⁽²³⁾ emphasise the urgent need for new approaches.

According to data reported to Agência Nacional de Vigilância Sanitária (ANVISA) in 2019, MRSA accounted for 59.1% of *S. aureus* samples isolated from hospitalised adult patients and was one of the most lethal pathogen-drug combinations, resulting in approximately 121,000 deaths attributable to antimicrobial resistance, according to a study by the Institute for Health Metrics and Evaluation.^(24,25,26,27)

The experimental results show that the ethyl acetate extract of CFAM 0234 inhibits *E. coli*, *S. sonnei* and *S. aureus* (MRSA), demonstrating antimicrobial activity. This finding is corroborated by the literature which documents the resistance of these pathogens to antibiotics such as amoxicillin and cephalothin,^(28,29) indicating that the metabolites produced by CFAM0234 can act by alternative mechanisms and bypass the resistance mechanisms.

Genomic analysis revealed important information that corroborates the phenotypic findings. The genome of CFAM 0234 was assembled into 962 contigs (total size of 72.67 Mbp, N50 = 152 kbp), although the assembly is fragmented - a fact also seen in the reference strain CBS 114.51 (36.1 Mbp, N50 = 412 kbp). The comparison between the two strains, carried out using OrthoFind-

er, showed significant differences in the repertoire of genes and suggested the presence of specific genomic regions with functional potential, as observed in specific contigs (e.g., Contig_591, Contig_606 and Contig_681).

In terms of biosynthetic potential, CFAM 0234 presented a significantly expanded repertoire, with 106 BGCs identified compared to 57 in strain CBS 114.51. Analysis with BiG-SCAPE revealed that 63 BGC families (GCFs) are unique to CFAM 0234, while 33 are unique to the reference, demonstrating a disparity that may reflect evolutionary adaptations to Amazonian soil. The comparative map (Figure A) shows the association between structural rearrangements and the location of BGCs — especially those of NRPS and TPKS —, with a high density of SNPs indicating selective pressure for metabolic diversification.⁽³⁰⁾

Additionally, the correlation between the antimicrobial activity of the crude extract and the detection of 14 BGCs, including five NRPS clusters with homology to gliotoxin-producing systems, two type I PKS clusters similar to those involved in lovastatin synthesis and a hybrid NRPS-PKS cluster containing chloramphenicol acetyltransferase domains, reinforces the potential of novel compounds. Although the activity of the extract is moderate (8 mm halo), this evidence motivates the application of purification techniques, such as liquid chromatography, to isolate the active compounds.

In addition, the abundance and diversity of BGCs, especially the NRPS, PKS and hybrid clusters, suggest promising biotechnological potential. While NRPS clusters are historically related to the synthesis of non-ribosomal peptides (e.g., penicillin), NRPS-PKS hybrid systems can produce molecules with unique structures, such as rapamycin, already recognised for its immunosuppressive properties.⁽³¹⁾ The identification of unique clusters, such as those of betalactones — potential precursors of β -lactam antibiotics — further highlights the possibility of discovering new bioactive compounds.⁽³²⁾

Integrating the genomic data with the results of the biological tests strengthens the hypothesis that the antimicrobial activity observed is not fortuitous, but rather the result of the expression of secondary metabolites derived from the BGCs identified. The concentration of SNPs in regulatory regions of BGCs (such as NRPS promoters) suggests variations in gene regulation between strains, possibly influenced by specific environmental pressures.^(33,34) In addition, structural rearrangements near PKS clusters may be promoting genetic recombination that diversifies metabolic products, an evolutionary strategy already demonstrated in fungal studies.⁽³⁵⁾

Finally, the fragmentation of the genome (N50 = 152 kbp) and the lack of correspondence with the MIBiG database highlight limitations that can be overcome with the use of long-read sequencing technologies (Oxford Nanopore) and metabolomic profiling techniques [liquid chromatography-mass spectrometry (LC-MS)]. These approaches would be fundamental for the functional validation of the predicted BGCs and for the isolation of the metabolites responsible for antimicrobial activity, paving the way for the development of new drugs.

In conclusion - This study has shown that *A. japonicus* CFAM 0234 has a broad and diverse repertoire of biosynthetic clusters, which is reflected in its antimicrobial activity against *E. coli*, *S. sonnei* and *S. aureus*. The genomic and phenotypic data indicate that the secondary metabolites produced by this strain may offer promising alternatives to combat resistant pathogens, in line with the priorities established by the WHO. Although the fragmentation of the genome and the lack of correspondence with the MIBiG bank represent methodological limitations, these issues open up avenues for the application of advanced techniques, such as long-read sequencing and integration with metabolomic analyses, in order to validate and fully exploit the biotechnological potential identified.

Therefore, the results presented not only contribute to the understanding of the resistance mechanisms of these bacteria and the bioactive factors in Amazonian fungi, but also establish a solid basis for future investigations that could lead to the development of new antimicrobial agents. The continuation of this work, with the improvement of analytical approaches and the functional validation of the predicted clusters, promises to significantly advance the discovery of innovative compounds to combat emerging infections.

ACKNOWLEDGEMENTS

To Amazon Bacteria Collection - CBAM and Amazon Fungi Collection - CFAM

AUTHORS' CONTRIBUTION

JCR - participation in the antimicrobial activity test against the clinical test microorganisms. Furthermore, analysed the results obtained for data dissemination and participated in the writing of the manuscript; MEG - performance of the bioinformatics analyses, analysed the results obtained for data dissemination and participated in the writing of the manuscript; ASQ and ALRL - participation in the antimicrobial activity test against the clinical test microorganisms; CVSG - reactivation of the selected fungal culture for the performance of the activities and also in the writing of the manuscript in the English language; LML - collaboration with the involved researchers to ensure that the research was conducted with quality. Additionally, they analysed the results obtained for data dissemination; OCCF - collaboration with the involved researchers to ensure that the research was conducted with quality. The authors declare no conflicts of interest.

DATA AVAILABILITY

Raw genomic data were deposited in the NIH's genetic sequence database - GenBank.

REFERENCES

1. Simonetti PAC. Evaluation System for Biomolecular Innovation - SINBIOMOL: from Amazon biodiversity to drugs [Internet]. 2019. Available from: <https://tede.ufam.edu.br/handle/tede/7057>.
2. Lima RMD. Bioprospecção de antimicrobianos produzidos por fungos do solo amazônico com ação frente às principais bactérias multirresistentes [Dissertação de Mestrado]. Manaus: Universidade Federal do Amazonas; 2016.
3. Prado FB, Rocha WC, Martim SR, Alecrim MM, Silva LP, Silva LSC, et al. Produção de compostos bioativos por *Aspergillus*

- mantidos sob duas condições de preservação. Boletim Do Museu Paraense Emílio Goeldi - Ciências Naturais. 2017; 12(1): 37-47.
4. Rodrigues JC, Lima da Silva W, Ribeiro da Silva D, Maia CR, Goiabeira CVS, Chagas HDF, et al. Antimicrobial activity of *Aspergillus* sp. from the Amazon biome: isolation of Kojic acid. Int J Microbiol. 2022; 2022: 4010018.
 5. Sales-Campos C, Chevreuil LR. Macromicetos amazônicos: potenciais biotecnológicos de modo sustentável. 2019. Available from: <https://repositorio.inpa.gov.br/handle/1/33882>.
 6. Pel HJ, de Winde JH, Archer DB, Dyer PS, Hofmann G, Schaap PJ, et al. Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. Nat Biotechnol. 2007; 25(2): 221-31.
 7. Meyer V, Basenko EY, Benz JP, Braus GH, Caddick MX, Csukai M, et al. Growing a circular economy with fungal biotechnology: a white paper. Fungal Biol Biotechnol. 2020; 7: 5.
 8. Klich MA, Klich Maren A, Pitt JI. A laboratory guide to the common *Aspergillus* species and their teleomorphs. North Ryde, NSW. 1988; 116 pp.
 9. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30(15): 2114-20.
 10. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012; 19(5): 455-77.
 11. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013; 29(8): 1072-5.
 12. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015; 31(19): 3210-2.
 13. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019; 20(1): 238.
 14. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res. 2023; 51(1): 46-50.
 15. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, et al. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol. 2020; 16(1): 60-8.
 16. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. PLoS Comput Biol. 2018; 14(1): e1005944.
 17. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009; 19(9): 1639-45.
 18. Teixeira MFS, Silva TA, Carneiro ALB, Palheta R. Fungos da Amazônia: uma riqueza inexplorada (aplicações biotecnológicas). 1a ed. Manaus: EDUA - Universidade Federal do Amazonas; 2011. p. 230-6.
 19. Balouiri M, Sadiki M, Ibsouda SK. Methods for in vitro evaluating antimicrobial activity: a review. J Pharm Anal. 2016; 6(2): 71-9.
 20. WHO - World Health Organization - Global antibiotic resistance surveillance report. 2025. Available from: <https://iris.who.int/server/api/core/bitstreams/139d068a-1086-43f7-a47d-575d0118d034/content>.
 21. Naghavi M, Vollset SE, Ikuta K, Swetschinski LR, Gray AP, Wool EE, et al. Global burden of bacterial antimicrobial resistance 1990-2021: a systematic analysis with forecasts to 2050. Lancet. 2024; 404(10459): 1199-226.
 22. WHO - World Health Organization. Shigellosis fact sheet. 2022. Available from: <https://www.who.int/teams/immunization-vaccines-and-biologicals/diseases/shigella>.
 23. Madela NK, Alves EC, Marques DF, Peresi JTM, de Almeida IAZC. Ocorrência da shigelose no Brasil: um problema de saúde pública. BEPA. 2017; 14(164): 1-19.
 24. ANVISA - Agência Nacional de Vigilância Sanitária. Boletim Segurança do Paciente e Qualidade em Serviços de Saúde no. 22. Brasília: ANVISA; 2019 [Internet]. 2019. Available from: <https://www.gov.br/anvisa/pt-br/centraisdeconteudo/publicacoes/servicosdesaude/boletins-e-relatorios-das-notificacoes-de-irras-e-outros-eventos-adversos-1/boletins-e-relatorios-das-notificacoes-de-irras-e-outros-eventos-adversos>.
 25. Ferreira ES, Gómez ASP, Almeida TVR, Frank CHM, Melo SA, Marinho EPM, et al. Microbiological profile of bloodstream infections and antimicrobial resistance patterns at a tertiary referral hospital in Amazon, Brazil. Rev Soc Bras Med Trop. 2023; 56: e03822023.
 26. IHME - Institute for Health Metrics and Evaluation. Among superbugs, MRSA is at the forefront of antimicrobial resistance | Institute for Health Metrics and Evaluation, 2019 [Internet]. 2019. Available from: <https://www.healthdata.org/news-events/insights-blog/acting-data/among-superbugs-mrsa-forefront-antimicrobial-resistance>.
 27. Martins MAV. *Staphylococcus aureus* resistente à meticilina associado a serviços de saúde: patógeno persistente e ou epidemiologia em transição? [PHD Thesis]. Botucatu: Faculdade de Medicina de Botucatu, Universidade Estadual Paulista. 2024.
 28. Bastos FC, Loureiro ECB. Antimicrobial resistance of *Shigella* spp. isolated in the State of Pará, Brazil. Rev Soc Bras Med Trop. 2011; 44(5): 607-10.
 29. Shad AA, Ahmed WAW. *Shigella sonnei*: virulence and antibiotic resistance. Arch Microbiol. 2021; 203(1): 45-58.
 30. Keller NP. Fungal secondary metabolism: regulation, function and drug discovery. Nat Rev Microbiol. 2019; 17(3): 167-80.
 31. Boettger D, Hertweck C. Molecular diversity sculpted by fungal PKS-NRPS hybrids. Chembiochem. 2013; 14(1): 28-42.
 32. de Oliveira ACFM, Vieira BD, de Félício R, Silva LSE, Veras AAO, Graças DA, et al. A metabologenomics approach reveals the unexplored biosynthetic potential of bacteria isolated from an Amazon Conservation Unit. Microbiol Spectr. 2025; 13(1): e0099624.
 33. Findley AS, Monziani A, Richards AL, Rhodes K, Ward MC, Kalita CA, et al. Functional dynamic genetic effects on gene regulation are specific to particular cell types and environmental conditions. Elife. 2021; 10: e67077.
 34. Jin HJ, Jung S, DebRoy AR, Davuluri RV. Identification and validation of regulatory SNPs that modulate transcription factor chromatin binding and gene expression in prostate cancer. Oncotarget. 2016; 7(34): 54616-26.
 35. Rokas A, Mead ME, Steenwyk JL, Raja HA, Oberlies NH. Biosynthetic gene clusters and the evolution of fungal chemodiversity. Nat Prod Rep. 2020; 37(7): 868-78.

OPEN PEER REVIEW

Memórias do IOC thanks the anonymous reviewers for their contribution to the peer review of this work.

FIRST REVIEW ROUND

REVIEWERS' COMMENTS

REVIEWER #1

The article is interesting; however, I believe that, to make it more compelling, the authors should demonstrate that the substances identified in the genome are indeed being produced by the fungus. To achieve this, different methodologies for secondary metabolite production should be employed, including variations in culture media and extraction methods, in order to determine the best approach for producing the metabolites of interest.

1) The culture medium used for fungal growth for DNA and secondary metabolite extraction should be the same, since, depending on the culture medium used, different metabolites may be produced. Therefore, the fungus grown in YES medium may not produce the same metabolites that are encoded in the genome.

2) What would be considered moderate antimicrobial activity, as you mentioned in the discussion and conclusion (line 258 and 310)? What parameter was used for this classification?

The disk diffusion method for extracts and non-standardized substances is only a qualitative method, and therefore it is not possible to quantify the amount of antimicrobial activity. I suggest performing the minimum inhibitory concentration (MIC) test to determine more precisely whether the antimicrobial activity of the extract is noteworthy.

3) I missed having a table containing the disk diffusion test results with all the bacteria tested.

AUTHORS' RESPONSE TO THE REVIEWERS

Question 1) The culture medium used for fungal growth for DNA and secondary metabolite extraction should be the same, since, depending on the culture medium used, different metabolites may be produced. Therefore, the fungus grown in YES medium may not produce the same metabolites that are encoded in the genome.

Response: We thank the reviewer for this important observation. We fully agree that the expression of biosynthetic gene clusters (BGCs) and the production of secondary metabolites are highly dependent on environmental and nutritional factors. The main purpose of the present study was to explore the genomic potential of *Aspergillus japonicus* CFAM 0234, identifying biosynthetic clusters with biotechnological relevance through comparative genome mining.

Although we did not perform metabolomic validation under different culture conditions, the detection of selective antimicrobial activity in the crude extract provides preliminary experimental evidence that some biosynthetic pathways may indeed be active. We have clarified this in the revised Discussion (lines 240 - 245) and explicitly stated that further experiments involving optimization of culture media, extraction protocols and LC-MS metabolomic profiling will be carried out to link BGCs to specific metabolites.

Question 2) What would be considered moderate antimicrobial activity, as you mentioned in the discussion and conclusion (line 258 and 310)? What parameter was used for this classification?

The disk diffusion method for extracts and non-standardized substances is only a qualitative method, and therefore it is not possible to quantify the amount of antimicrobial activity. I suggest performing the minimum inhibitory concentration (MIC) test to determine more precisely whether the antimicrobial activity of the extract is noteworthy.

Response: We thank the reviewer for pointing this out. We have now clarified in the Materials and Methods and Discussion sections that the classification of "moderate activity" was based on halo diameters between 6 and 10 mm, following comparative evaluation with positive (vancomycin, 1 mg/mL) and negative controls (DMSO 4%).

We fully agree that the disk diffusion test is a qualitative approach and that it does not allow precise quantification of antimicrobial activity, as the amount of compound diffused into the agar cannot be accurately determined. As highlighted in the literature (Balouiri et al, 2015), disk diffusion is widely used for preliminary antimicrobial screening of extracts and other compounds due to its simplicity, low cost, and ability to test multiple microorganisms simultaneously, but MIC determination is required for accurate assessment of potency. We have added a note acknowledging this limitation and emphasizing that MIC and purification assays will be addressed in future studies. (changes in line 247-253)

Question 3) I missed having a table containing the disk diffusion test results with all the bacteria tested.

Response: We appreciate the suggestion. The revised manuscript now includes a summary table (Table 4) listing the inhibition halo diameters for all tested bacterial strains, as obtained in the disk diffusion assays. This addition allows a clearer overview of the selectivity profile of the crude extract.

SECOND REVIEW ROUND

REVIEWERS' COMMENTS

REVIEWER #1

No comments.

DECISION: ACCEPT SUBMISSION | RECEIVED 01 APRIL 2025 | ACCEPTED 14 NOVEMBER 2025