BMC Genomics

RESEARCH

Open Access

Genomic insights into the biosynthetic capacity of the sponge-associated fungus *Aspergillus puulaauensis* Hmp-F48

Yue Yan[1], Xiao Wang[1], Qiang Ma[4], Yijie Lou[3], Zhe Chen[3], Ning Li[1*] and Nan Wang[2*]

Abstract

Background Marine-derived fungi are prolific producers of structurally diverse secondary metabolites with significant pharmaceutical potential. The discovery of natural products has been unprecedentedly accelerated by the prediction of biosynthetic gene clusters (BGCs) within the whole-genome context. This study presents the whole-genome sequencing, comprehensive annotation, and biosynthetic potential predictions of the sponge-associated fungus *Aspergillus puulaauensis* Hmp-F48.

Results Genome sequencing of *A. puulaauensis* Hmp-F48 generated a high-quality draft assembly of 35.86 Mb. Structural annotation revealed a complex genomic architecture, comprising 10,611 protein-coding genes, 210 non-coding RNAs, and 155 tRNAs. Functional annotation using NR, Swiss-Prot, GO, KEGG, and eggNOG databases highlighted significant enrichment in biosynthetic, metabolic, and transport processes. AntiSMASH analysis identified 78 putative BGCs, including 16 type I polyketide synthase (T1PKS), 27 nonribosomal peptide synthetase (NRPS), 7 hybrid PKS-NRPS, 9 terpene-related clusters, 6 RiPP-related clusters and 13 clusters associated with other secondary metabolites. Several clusters exhibited high homology to known BGCs responsible for bioactive secondary metabolites, including asperthecin, sterigmatocystin, calbistrins, F-9775 A/B, nidulanin A, aspercryptins, fellutamide B, acetylaranotin, burnettramic acid A, equisetin, and pyranonigrin E. Experimental isolation confirmed the presence of PKS-derived metabolites, including sterigmatocystin, averantin, and decumbenones — one of which represents a previously undescribed congener.

Conclusions This study highlights the extensive biosynthetic potential of *A. puulaauensis* Hmp-F48, offering valuable insights into its capacity for secondary metabolite production.

Keywords Marine fungi, Aspergillus puulaauensis, Genome sequencing, Biosynthetic gene cluster, Secondary metabolites

*Correspondence:
Ning Li
liningsypharm@163.com
Nan Wang
n_wang@zju.edu.cn
[1]School of Traditional Chinese Materia Medica, Shenyang Pharmaceutical University, Shenyang 110016, China
[2]Ocean College, Zhejiang University, Zhoushan 316021, China
[3]The First Affiliated Hospital of Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Chinese Medicine), Hangzhou 310053, China
[4]School of Food and Pharmacy, Zhejiang Ocean University, Zhoushan 316022, China

Background

Marine ecosystems, characterized by unique physicochemical conditions, drive the evolution of distinctive metabolic pathways in marine microorganisms, resulting in structurally diverse secondary metabolites with pharmaceutical potential [1–4]. Marine fungi, particularly those associated with marine sponges are recognized as prolific producers of bioactive secondary metabolites [5]. Among these, marine-derived *Aspergillus* species have attracted significant interest due to their ability to synthesize diverse bioactive compounds, including polyketides, peptides, alkaloids, and terpenoids with potent antimicrobial, anticancer, anti-inflammatory, and antioxidant activities [6–8]. A notable example is plinabulin, a synthetic derivative of halimide from marine-derived *Aspergillus ustus*, is currently undergoing Phase III clinical trials as a second-line chemotherapy agent in combination with paclitaxel for the treatment of advanced or metastatic non-small cell lung cancer (NSCLC) and for the prevention of chemotherapy-induced neutropenia (CIN) caused by docetaxel, underscoring the therapeutic promise of marine *Aspergillus* metabolites [9].

Recent advancements in genome sequencing technologies have established genome-based strategies as powerful tools for natural product discovery [10, 11]. Meanwhile, bioinformatic tools for the prediction of biosynthetic gene clusters (BGCs) efficiently bridge genomic dataset to chemically diversity, offering unprecedented opportunities for targeted natural product discovery [12, 13].

In this study, *Aspergillus puulaauensis* Hmp-F48 was prioritized for whole-genome sequencing from the *Hymeniacidon perleve*-associated fungal community due to its distinctively rich metabolic profile and significant bioactivity observed in preliminary screenings [14]. Elucidating its genomic architecture is therefore essential to decipher the biosynthetic machinery driving these traits and to facilitate the discovery of novel natural products. The whole-genome of Hmp-F48 was sequenced using Illumina HiSeq and PacBio platforms. Comprehensive genome annotation, including prediction of protein-coding genes and non-coding RNAs, was performed. For functional annotation, we utilized databases such as Nr, Swiss-Prot, GO, KEGG, eggNOG, and CAZy. BGCs were predicted and analyzed using antiSMASH. Finally, we cultivated *A. puulaauensis* Hmp-F48, isolated its secondary metabolites, and established direct links between its predicted BGCs and chemical profiles through structural characterization.

Materials and methods

Fungal strain

The fungal strain *A. puulaauensis* Hmp-F48 was isolated from a marine sponge (*Hymeniacidon perleve*) collected from the Bohai Sea at Heishi Reef Bay (38°51′52.4′′N121°33′05.0′′E) in Dalian, China and is preserved at the China Center for Type Culture Collection (CCTCC NO: AF 2025006). The Hmp-F48 strain was maintained on Czapek–Dox Medium (0.2% $NaNO_3$, 3% Sucrose, 0.1% K_2HPO_4, 0.05% KCl, 0.05% $MgSO_4{\cdot}7H_2O$, 0.001% $FeSO_4{\cdot}7H_2O$) at 28 °C with shaking at 180 rpm for 5 days. Large-scale fermentation (20 L) for metabolite isolation was performed in PDB Medium (200 g/L Potato, 2% Sucrose) at 28 °C with shaking at 180 rpm for 7 days. Spore morphology was observed after 4–5 days of cultivation on PDA Medium (200 g/L potato, 2% sucrose, 2% agar). The morphology of Hmp-F48 was examined using scanning electron microscopy (SEM, Phenom™ ProX, Phenom World BV, Netherlands).

Genome sequencing and assembly

Genomic DNA was extracted from mycelia using the OMEGA Fungal Genomic DNA Isolation Kit following the manufacturer's protocol. DNA was assessed via agarose gel electrophoresis. The genome of the Hmp-F48 strain was sequenced using the Whole Genome Shotgun (WGS) strategy combining Illumina and PacBio platforms. For PacBio sequencing, SMRTbell libraries with an average insert size of ~ 10 kb were prepared and sequenced on the PacBio Sequel II platform, producing approximately 8.37 Gb of raw long-read data, corresponding to an estimated ~ 233× genome coverage. For Illumina sequencing, paired-end libraries (2 × 151 bp) with an average insert size of ~ 400 bp were prepared and sequenced on an Illumina HiSeq instrument, yielding approximately 5.41 Gb of clean reads, corresponding to an estimated ~ 147× genome coverage. Raw Illumina reads were quality-checked using FastQC v0.11.9, adapter contamination was removed with AdapterRemoval v1.5.4, and base quality correction was performed with SOAPec v2.0 (K-mer = 17) before being used for polishing. PacBio long reads were first assembled *de novo* using Canu v2.2 to generate draft contigs and scaffolds [15]. The corrected Illumina short reads were then aligned to these PacBio scaffolds using BWA-MEM, and two rounds of error correction were performed using Pilon v1.23 to correct base substitutions, small insertions and deletions, and local misassemblies [16]. The completeness of the genome assembly was assessed using BUSCO (v3.0.2, https://busco.ezlab.org) [17].

Genome annotation

Gene model prediction was carried out using three *de novo* prediction tools: Augustus (v3.03), GlimmerHMM (v3.0.1), and SNAP (v2006-07–28). Homology-based gene prediction was conducted using Exonerate (v2.2.0) with protein sequences from closely related species as references. The results from *de novo* predictions and

homology-based approaches were integrated using EVidenceModeler (vr2012-06–25), producing a comprehensive gene annotation. Non-coding RNA (ncRNA) annotation included identifying tRNAs with tRNAscan-SE (v1.3.1) and rRNAs with RNAmmer (v1.2). Additional ncRNA elements were identified through comparative analysis with the Rfam database (http://rfam.xfam.org/), allowing for the identification of conserved RNA families and regulatory RNAs. Functional gene annotation was performed using five databases, including Nr (https://ftp.ncbi.nlm.nih.gov), eggNOG (https://www.ncbi.nlm.nih.gov/COG/), Swiss-Prot (https://www.expasy.org/resources/uniprotkb-swiss-prot), KEGG (https://www.kegg.jp/kegg/) and GO (http://geneontology.org/), to gain insights into gene functions and associated metabolic pathways. Additionally, carbohydrate-active enzyme (CAZy) genes (http://www.cazy.org) within the genome were predicted using hmmscan (v3.1b2). The analysis focused on open reading frames (ORFs) exceeding 80 amino acids in length, with an E-value threshold of 1e-5 to ensure high-confidence alignments.

Phylogenetic analysis

For phylogenomic analysis with whole genome sequence, the final annotated data were processed to retain the protein-coding genes and the longest transcript. OrthoFinder v2.5.5 was used to perform gene family analysis using Hmp-F48 and 17 other *Aspergillus* strains, with *Penicillium chrysogenum* IBT 35,668 included as an outgroup (Table S1 lists all genomes and NCBI accession numbers), resulting in a total of 25,144 clusters of orthologous proteins (COPs) across all genomes analyzed. Multiple sequence alignments of these COPs were generated with ParaAT v1.0, and the resulting sequences were concatenated into a supergene using seqkit v2.7.0. Phylogenomic trees were constructed using RAxML-NG v1.2.1 under the GTR + G + I model with 1000 bootstrap replications. All resulting trees were visualized using FigTree v1.4.4 or ITOL (Interactive Tree of Life), with final layout adjustments made in Adobe Illustrator CC 2019. In addition, Average Nucleotide Identity (ANI) values between Hmp-F48 and related strains were calculated using FastANI v1.33 with default parameters, providing quantitative measures of genome similarity to support phylogenomic relationships.

Table 1 Statistics of genome assembly results of Hmp-F48

Property	Contig
Shortest (bp)	15,037
Longest (bp)	5,691,020
Total sequence number	30
N20 (bp)	5,140,981
N50 (bp)	2,260,453
N90 (bp)	688,255
N number	0
N rate	0
Total sequence length (bp)	35,864,449
GC content (%)	49.72%
Sequences greater than 1 kb	30

Total sequence number: The number of contiguous sequences generated during assembly; Total sequence length (bp): The assembly length of the genome; Shortest sequence length (bp): The length of the shortest contig; Longest sequence length (bp): The length of the longest contig; GC content: The percentage of guanine and cytosine bases in the genome; N number: The total number of undetermined nucleotides (N) across all contigs; N rate: The percentage of ambiguous bases relative to the total assembly length; Sequences greater than 1 kb: The number of contigs longer than 1 kb in length; N20 (bp): The weighted median contig length at 20% of the total assembly size, calculated by summing the lengths of contigs from longest to shortest until 20% of the total assembly size is reached; The N50 and N90 algorithm is the same as N20

Prediction of biosynthetic gene clusters

The BGCs responsible for secondary metabolite production in the Hmp-F48 genome were predicted using the antiSMASH (fungal version 8.0.0) web server (http://antismash.secondarymetabolites.org/). Each identified gene cluster were extracted from the genome and subjected to BLAST analysis to identify homologous sequences and predict potential product of the cluster.

Fungal metabolite extraction and separation

The Hmp-F48 strain was cultured in PDB medium at 28 °C with shaking at 200 rpm for 7 days. Following fermentation, the culture broth was extracted three times with ethyl acetate. The combined organic phases were concentrated under reduced pressure to yield a crude extract (2.63 g). Initial fractionation of the crude extract was performed using silica gel column chromatography with dichloromethane-methanol gradient (90:10 to 80:20, v/v), resulting in two subfractions (Fr.1 and Fr.2). Subfraction Fr.2 was further purified using an ODS column with a methanol-water gradient (20:80 to 40:60, v/v), yielding four subfractions (Fr.1-Fr.4). Finally, semi-preparative HPLC was performed on a SHIMAZU LC-20 equipped with a DAD detector and a YMC-ODS-A C18 column (250 × 4.6 mm, 5 μm) to isolate compounds 1–4 from these subfractions. Nuclear magnetic resonance (NMR) spectra were recorded on a Bruker AV-600 spectrometer (Billerica, MA, USA) using tetramethylsilane (TMS) as the internal standard. HR-ESI-MS analysis were performed on a Bruker micrOTOF-Q and a Agilent.

Results

Genome sequencing and assembly

The genome of *A. puulaauensis* Hmp-F48 consists of 30 contigs with a GC content of 49.72% and a total assembly length of 35.86 Mb. The N50 and N90 values are 2,260,453 bp and 688,255 bp, respectively, indicating high assembly contiguity (Table 1). BUSCO analysis (290 total groups) of the assembly revealed that 99.3% of the

genes are complete, confirming the genome's high completeness and reliability (Table S2). These findings confirm that the Hmp-F48 genome is highly contiguous and nearly complete, supporting further genomic and functional investigations.

Morphology, classification and phylogenetic analysis of strain Hmp-F48

Strain Hmp-F48 was grown on PDA medium at 28 °C for 4–5 days, forming colonies with diameters ranging from 2 to 8 mm. The colonies displayed a yellow-green coloration with a green central region and yellow edges, while the reverse side exhibited a yellow-brown hue without distinct striations. The conidia were predominantly spherical, with a granular texture and yellow-green coloration (Fig. 1A). Scanning electron microscopy (SEM) revealed globose conidia with granular and ridged surface ornamentation, borne in dense radiate clusters on smooth-walled hyphae (Fig. 1B). Phylogenomic analysis based on whole-genome protein sequences positioned Hmp-F48 within the *Aspergillus versicolor* clade, showing its closest relationship to *Aspergillus puulaauensis* MK2 [18]. ANI analysis using FastANI v1.33 supported this placement, with an ANI score of 96% between Hmp-F48 and *A. puulaauensis* MK2 (Table S3), providing robust genomic evidence for its taxonomic assignment (Fig. 1C).

Prediction of genetic structure

The structural prediction analysis of Hmp-F48 identified 10,611 protein-coding genes, with a total coding sequence (CDS) length of 16,857,201 bp, accounting for 47.00% of the genome. The average gene length was calculated to be 1588.6 bp, with an average of 2.9 exons per gene. With a total length of 15,135,382 bp, exons represent 42.20% of the genome. Introns, in contrast, have an average length of 81.4 bp. Notably, the total length of CDSs matches that of exons, also constituting 42.20% of the genome, with an average CDS length of 1426.3 bp

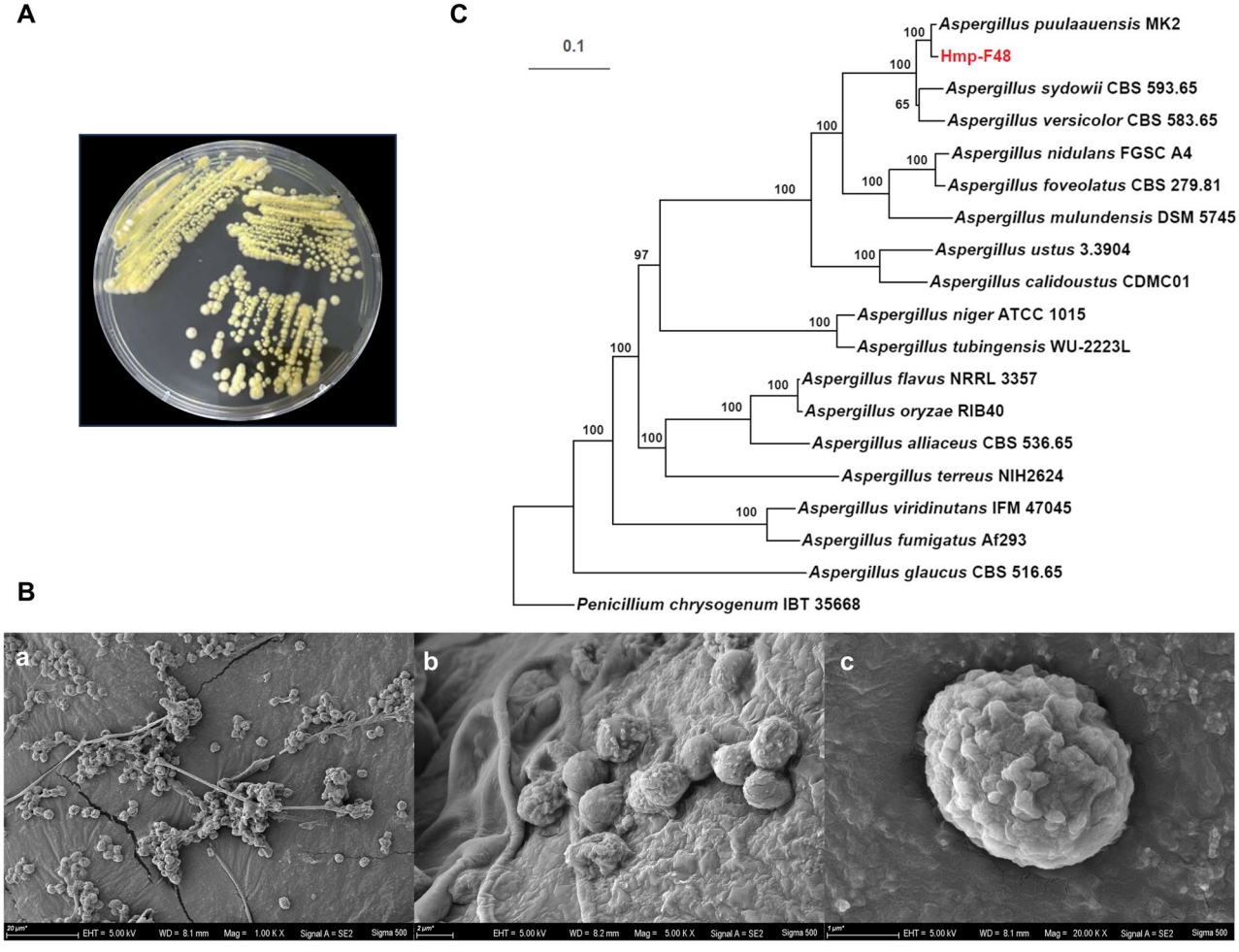

Fig. 1 Morphological and molecular taxonomy of Hmp-F48. **A** The colony of strain Hmp-F48 on PDA Medium. **B** Scanning electron microscopy (SEM) images of Hmp-F48 conidia and hyphae. a: 1000× b: 5000× c: 20,000×. Scale bars: shown in panels. Imaging parameters: EHT = 5.00 kV, WD = 8.1–8.2 mm, SE2 detector. **C** Whole-genome protein sequence-based phylogenetic tree of strain Hmp-F48 and related *Aspergillus* species, with *P. chrysogenum* IBT 35,668 as the outgroup

(Table S4). The genome contains 30 copies of 5S rRNA gene, along with 5 copies each of 5.8S and 28S rRNA genes, 7 copies of 18S rRNA gene, 155 tRNA genes, and 210 ncRNA genes (Table S5), indicating a complex genomic structure with diverse functional capacities.

Annotation of gene functions

To obtain a comprehensive overview of the functional repertoire of Hmp-F48, protein-coding genes were annotated against multiple complementary databases, each providing distinct insights into taxonomy, biochemical function, and metabolic potential. The 10,611 protein-coding genes of Hmp-F48 were annotated across multiple databases, including NR, Swiss-Prot, GO [19], KEGG [20], and eggNOG [21]. Annotation success rates were 99.35% (10543 genes) for NR, 73.36% (7784 genes) for Swiss-Prot, 72.52% (7695 genes) for GO, 33.91% (3598 genes) for KEGG, and 93.42% (9913 genes) for eggNOG (Table S6). The NR database analysis revealed that most homologous sequences belonged to *Aspergillus* species, particularly *A. sydowii* CBS 593.65 and *A. versicolor* CBS 583.65 (Fig. 2A). GO annotation categorized the genes into biological processes, cellular components, and molecular functions. A significant proportion of genes associated with biological processes are involved in biosynthesis (1752 genes), nitrogen compound metabolism (1822 genes), transport (1307 genes), and small molecule metabolism (1088 genes). Molecular function annotations highlight roles in catalytic activities like oxidoreductase activity (1657 genes), methyltransferase activity (537 genes), and hydrolase activity targeting non-peptide or glycosidic bonds (402 genes). Additionally, a notable number of genes were involved in binding functions, particularly iron binding (2654 genes). In the cellular components category, genes contribute prevalently to cellular structure (2638 genes), intracellular components (2501 genes), organelles (1980 genes), and cytoplasmic composition (1303 genes) (Fig. 2B). KEGG pathway mapping revealed that 36.9% of genes (2,424 genes) were involved in metabolic pathways, suggesting substantial metabolic potential and capacity for diverse metabolite production. (Fig. 2C). The eggNOG classification showed 36.95% of genes (3,663 genes) with unknown functions, indicating potential for discovering novel chemistry. Among characterized genes, those related to carbohydrate transport and metabolism (956 genes, 9.64%), secondary metabolite biosynthesis, transport, and catabolism (704 genes, 7.10%), and amino acid transport and metabolism (539 genes, 5.44%) were prominent (Fig. 2D). Further functional annotations using the CAZy database identified six categories of carbohydrate-active enzymes [22]: glycosyl transferases (GT), polysaccharide lyases (PL), carbohydrate esterases (CE), auxiliary activities (AA), carbohydrate-binding modules (CBM), and glycoside hydrolases (GH) (Fig. 2E). This classification emphasizes the strain's enzymatic diversity and potential applications in carbohydrate metabolism and biotechnological.

Annotation of secondary metabolites biosynthetic gene clusters

The biosynthetic potential of Hmp-F48 for secondary metabolites was predicted using antiSMASH [23], identifying 78 putative biosynthetic gene clusters (BGCs). These BGCs mainly include 16 type I polyketide synthase (T1PKS) clusters, 27 NRPS clusters, and 7 hybrid PKS-NRPS clusters. Additionally, 9 terpene-related clusters, 6 RiPP-related clusters and 13 clusters encoding other classes of secondary metabolites were identified (Fig. 3A, Table S7). Based on their completeness and homology to known pathways, four PKS, four NRPS, and three PKS-NRPS clusters were prioritized for further functional annotation and product prediction.

PKS gene clusters

Cluster 1 (T1PKS)

Shares 66–100% sequence identity with the asperthecin BGC in *A. nidulans* FGSC A4 [24]. Asperthecin, a polyphenolic compound, facilitates ascus formation and confers UV resistance in fungi [25]. Notably, its inhibition of tau protein aggregation highlights therapeutic potential in neurodegenerative diseases [26]. The cluster harbors homologs of *aptA* (PKS), *aptB* (thioesterase), and *aptC* (monooxygenase), suggesting a conserved role in asperthecin-like polyketide biosynthesis (Figs. 3B and 4A).

Cluster 2 (T1PKS)

Displays 66–100% homology to the sterigmatocystin cluster in *A. ochraceoroseus* [27]. Sterigmatocystin is a potent carcinogen, mutagen and teratogen produced by several species of *Aspergillus* [28]. With a xanthone nucleus attached to a bifuran structure, sterigmatocystin is a biosynthetic precursor to aflatoxin. Notably, this BGC harbors additional transporter genes compared with the known reference cluster (Figs. 3B and 4B).

Cluster 3 (T1PKS)

Shows 71–100% homology to the calbistrin/decumbenone biosynthetic cluster *cal* from *A. stellatus* [29, 30]. Calbistrins are antifungal and cytotoxic natural products structurally analogous to lovastatins due to their decalin core and polyketide side chain [31, 32] (Figs. 3B and 4C).

Cluster 4 (T1PKS)

Shares 68–100% identity with the cysteine protease cathepsin K inhibitor F-9775 A/B biosynthetic cluster in *A. nidulans* FGSC A4 [33]. In addition to the homologous

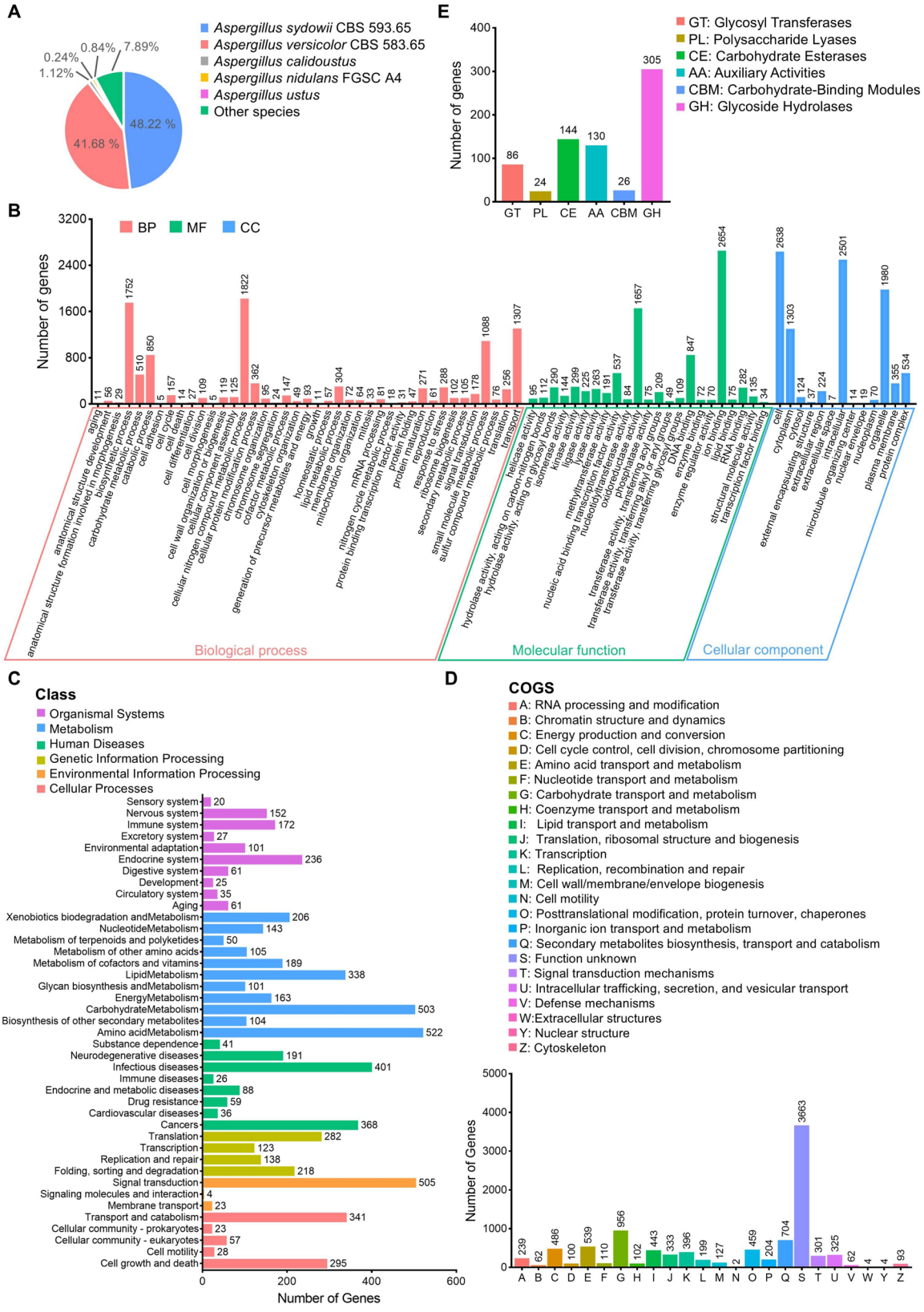

Fig. 2 Gene Function Annotation. **A** Distribution of species based on matches from the non-redundant protein sequence (Nr) database. **B** Classification statistics based on Gene Ontology (GO) annotations. **C** Functional classification of pathways based on the Kyoto Encyclopedia of Genes and Genomes (KEGG). **D** Classification statistics based on Cluster of Orthologous Groups (COG) annotations. **E** Functional classification of Carbohydrate-Active Enzymes (CAZy)

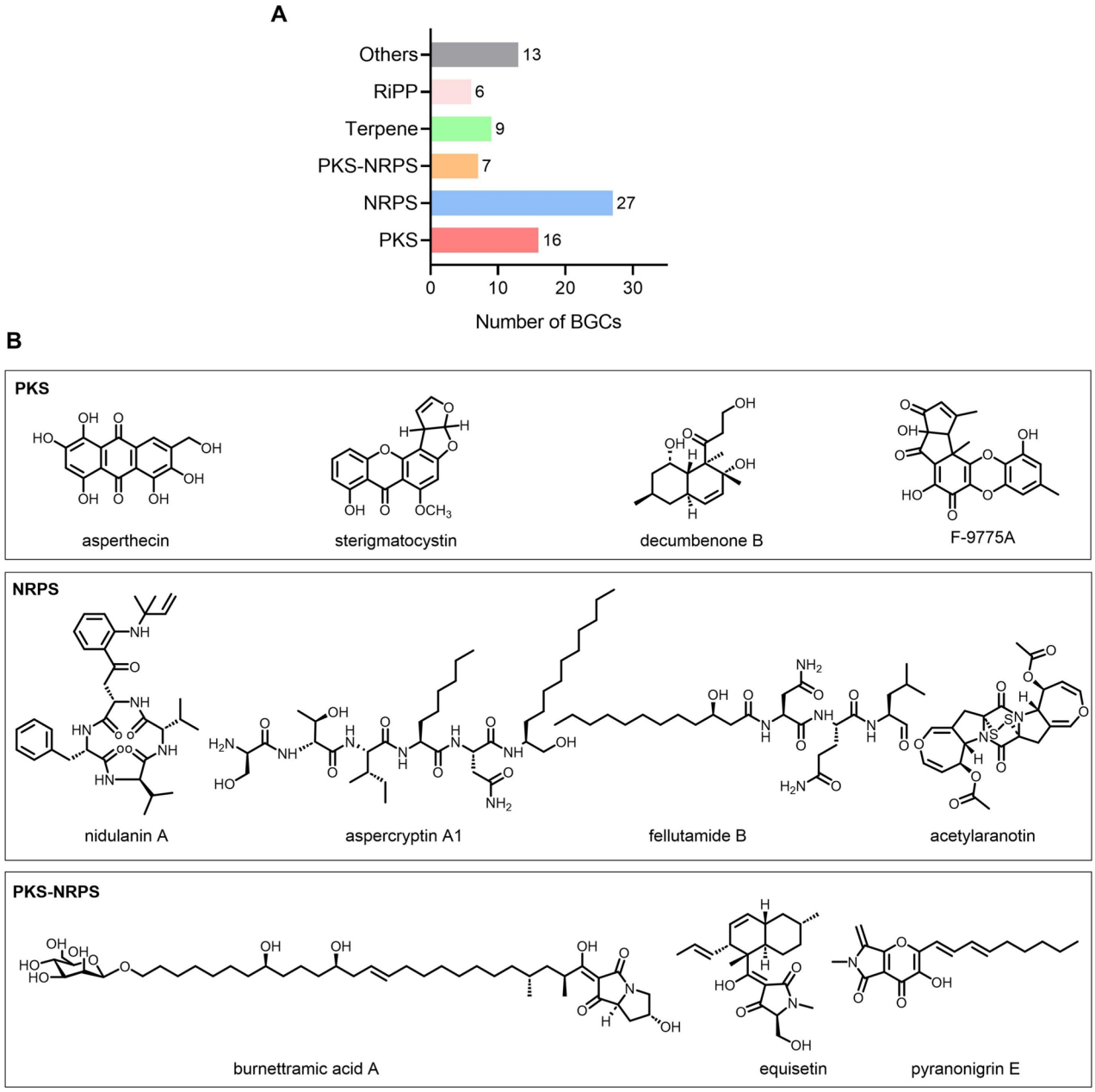

Fig. 3 **A** Types and abundance of biosynthetic gene clusters (BGCs) in strain Hmp-F48. **B** Chemical structures of products from homologous gene clusters corresponding to BGCs 1–11 in strain Hmp-F48

genes AN7909-AN7914, this cluster contains several other genes with unknown functions (Figs. 3B and 4D).

NRPS gene clusters

Cluster 5

Homologous (66–100%) to the nidulanin A biosynthetic cluster in *A. nidulans* FGSC A4 [34, 35] which is known for producing tetracyclopeptide/isoprene hybrid metabolites (Figs. 3B and 4E).

Cluster 6

Shares 69–100% identity with the *atn* gene cluster producing aspercryptin A1 in the *A. nidulans* FGSC A4 [36]. The core biosynthetic genes for aspercryptin A1 include an NRPS gene (*atnA*) and two fatty acid synthase genes (*atnF* and *atnM*). Notably, antiSMASH predicts the absence of aminotransferase genes critical for α-amino fatty acid synthesis in this gene cluster, suggesting either truncation of the cluster or that these genes lie outside its boundaries (Figs. 3B and 4F). Therefore, this cluster likely

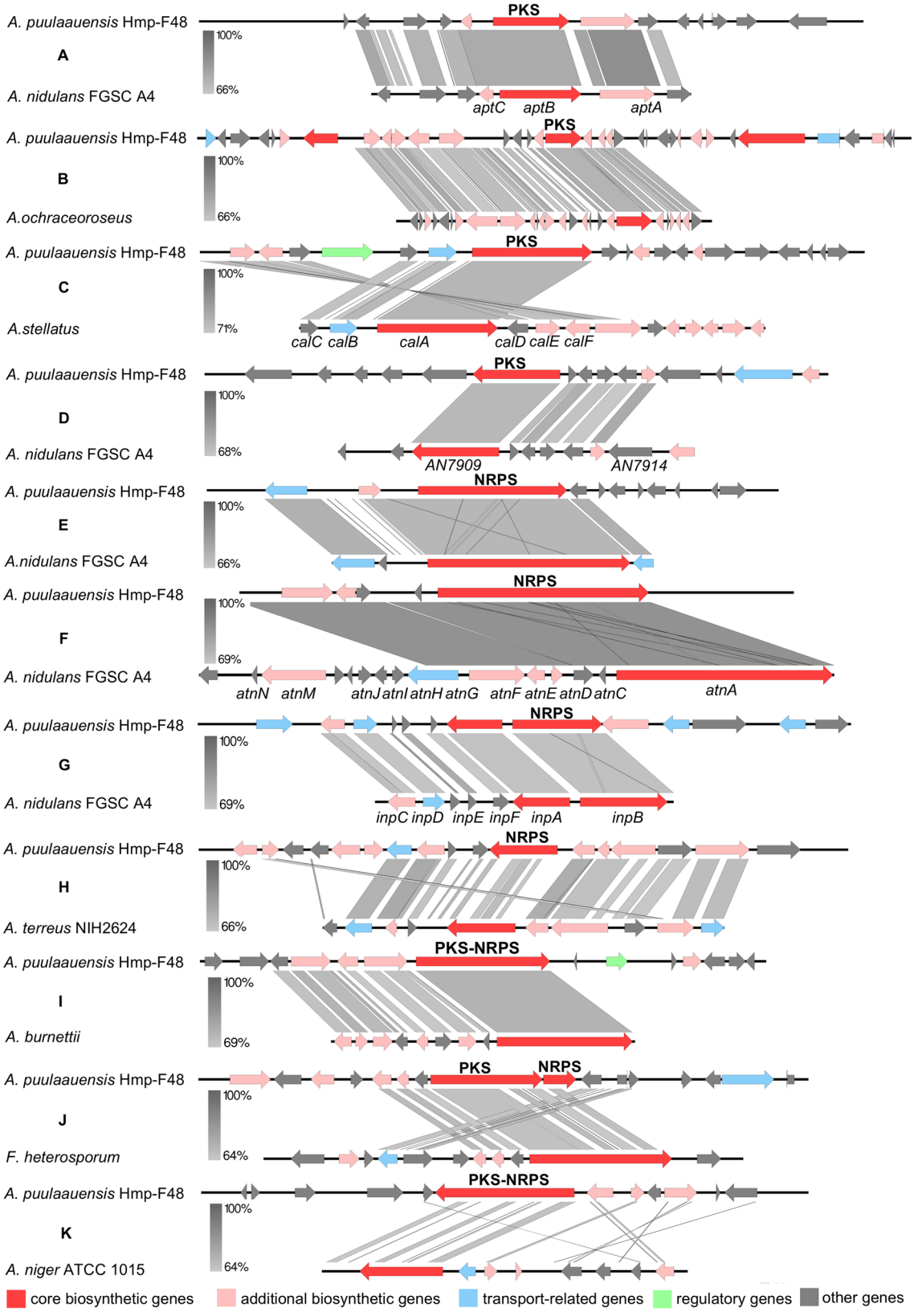

Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Synteny plots of BGCs 1–11 and their homologous clusters, generated using Easyfig (BLASTn). **A** Syntenic comparison between the predicted asperthecin cluster in Hmp-F48 and the identified candidate cluster in *A. nidulans* FGSC A4. **B** Syntenic comparison between the predicted sterigmatocystin cluster in Hmp-F48 and the identified candidate cluster in *A. ochraceoroseus*. **C** Syntenic comparison between the predicted calbistrins cluster in Hmp-F48 and the identified candidate cluster in *A. stellatus*. **D** Syntenic comparison between the predicted F-9775 A/B cluster in Hmp-F48 and the identified candidate cluster in *A. nidulans* FGSC A4. **E** Syntenic comparison between the predicted nidulanin A cluster in Hmp-F48 and the identified candidate cluster in *A. nidulans* FGSC A4. **F** Syntenic comparison between the predicted aspercryptins cluster in Hmp-F48 and the identified candidate cluster in *A. nidulans* FGSC A4. **G** Syntenic comparison between the predicted fellutamide B cluster in Hmp-F48 and the identified candidate cluster in *A. nidulans* FGSC A4. **H** Syntenic comparison between the predicted acetylaranotin cluster in Hmp-F48 and the identified candidate cluster in *A. terreus* NIH2624. **I** Syntenic comparison between the predicted burnettramic acid A cluster in Hmp-F48 and the identified candidate cluster in *A. burnettii*. **J** Syntenic comparison between the predicted equisetin cluster in Hmp-F48 and the identified candidate cluster in *F. heterosporum*. **K** Syntenic comparison between the predicted pyranonigrin E cluster in Hmp-F48 and the identified candidate cluster in *A. niger* ATCC 1015

encodes a noncanonical or an incomplete aspercryptin biosynthetic machinery.

Cluster 7

Shows significant alignment (69–100% similarity) with the fellutamide B BGC in *A. nidulans* FGSC A4. The proteasome subunit InpE in this cluster [37] is consistent with fellutamide B's role as a proteasome inhibitor with antitumor potential [38] (Figs. 3B and 4G).

Cluster 8

Exhibiting 69–100% homology to the acetylaranotin cluster in *A. terreus* [39], associated with producing cytotoxic diketopiperazine natural products, epipolythiodioxopiperazines [40] (Figs. 3B and 4H).

PKS-NRPS hybrid clusters

Cluster 9

Highly homologous with the burnettramic acid A BGC in *A. burnettii*, comprising a PKS-NRPS core gene *buaA* and tailoring genes *buaB-G* [41]. In previously studies, burnettramic acid A exhibited potent antimicrobial activity, particularly against *Candida albicans* [42] (Figs. 3B and 4I).

Cluster 10

With 64–100% sequence similarity to the equisetin cluster from *Fusarium heterosporum* [43], this BGC is propose to synthesize decalin-containing compounds structurally related to lovastatins, known for their lipid-lowering, antimicrobial, and anti-inflammatory activities [44, 45] (Figs. 3B and 4J).

Cluster 11

Exhibits partial homology to the pyranonigrin E cluster from *A. niger* [46], suggesting a potential for producing pyranonigrin-like metabolites. As is typical for all in silico predicted BGCs within this study, its associated product profiles require future experimental validation (Figs. 3B and 4K).

Isolation of secondary metabolites

To functionally annotate the predicted BGCs in Hmp-F48, we systematically isolated and structural elucidated its secondary metabolites. Notably, averantin and sterigmatocystin, previously isolated from this strain, are likely products of the PKS gene clusters BGC1 and BGC2, respectively [14]. Our recent chemical investigations revealed a series of decumbenone-type derivatives, including the known compounds decumbenone B (1), craterellone F (2), versiol (3) and an undescribed congener designated as 1*R*, 3*S*, 5*S*, 8*S*, 9*S*, 10*R*–9,10-dihydroversiol (4) (Fig. 5). The NMR spectroscopic data for these compounds are provided in Figures S1-S12 and Tables S8-S9. While compound 4 shares an identical planar structure to isoversiol A, it differs in the stereochemistry of C-8 position [47]. In the NOESY spectrum of compound 4, H-13a showed a cross-peak with 14-CH_3, and 14-CH_3 correlated with 15-CH_3. In addition, 5-H exhibited NOE interactions with both 15-CH_3 and H-3. These correlations suggest that H-13a, 14-CH_3, 15-CH_3, 5-H, and H-3 are α-oriented. Conversely, NOE cross-peaks between 10-H and H-1, H-2a, and H-4a indicate that these protons are β-oriented (Fig. S1). When the modified Mosher's method was applied to the absolute configuration at position 1, the results were inconclusive, likely due to steric hindrance from the pyran ring that limited the applicability of this method. Based on the typical configuration observed in compounds isolated from *A. versicolor* is 1*R*, 3*S*, 5*S*, 9*S*, 10*R*, we propose that compound 4 has an absolute configuration of 1*R*, 3*S*, 5*S*, 8*S*, 9*S*, 10*R* [48].

Although in silico analyses suggest diverse biosynthetic machinery including NRPS, terpene, RiPP, and hybrid clusters, all experimentally characterized metabolites from Hmp-F48 thus far are derived from PKS pathways. This is likely due to the limitation in current chemical profiling approaches. On the other hand, many gene clusters are silent under regular culture conditions. To fully understand their products, culture condition optimization, gene regulation and heterologous expression may be required to activate these cryptic BGCs.

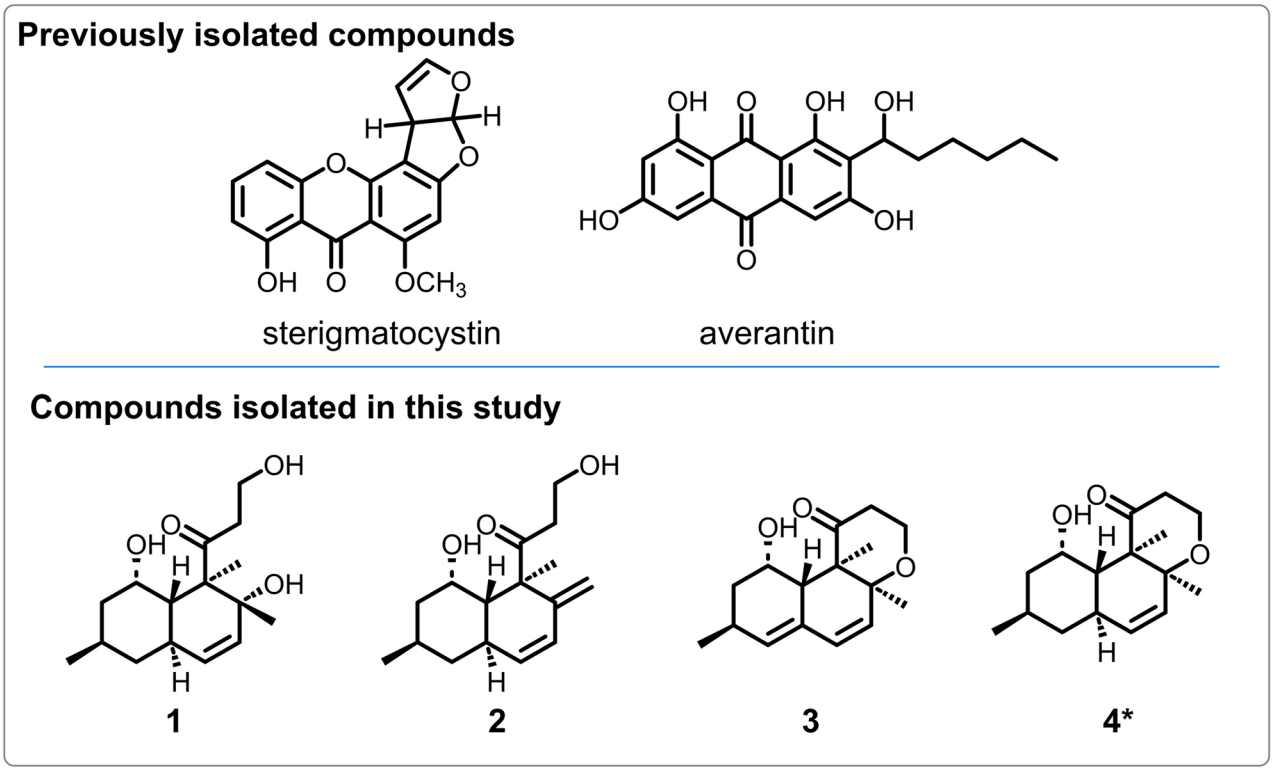

Fig. 5 Chemical structures of the compounds isolated from strain Hmp-F48

Discussion

The comprehensive genomic analysis of *A. puulaauensis* Hmp-F48 revealed a high-quality assembly (35.86 Mb, N50 2.26 Mb), comparable to those of the reference strains such as *A. versicolor* CBS 583.65 (33.1 Mb, N50 2.5 Mb) and *A. sydowii* CBS 593.65 (34.4 Mb, N50 2.3 Mb). Notably, Hmp-F48 contained no ambiguous bases and was assembled into only 30 scaffolds, surpassing the completeness of CBS 583.65 (51 scaffolds) and CBS 593.65 (97 scaffolds). This enhanced continuity provides a robust foundation for functional annotation.

AntiSMASH analysis of the *A. puulaauensis* Hmp-F48 genome revealed 78 biosynthetic gene clusters (BGCs). This number is comparable to, and slightly higher than, those found in closely related *Aspergillus* species, such as *A. puulaauensis* MK2 (69 BGCs) [49], *A. nidulans* FGSC A4 (61 BGCs) [50], *A. sydowii* CBS 593.65 (59 BGCs) [51], and *A. versicolor* CBS 583.65 (66 BGCs) [52]. These observations indicate that Hmp-F48 has a predicted BGC repertoire at the upper end of the range typically observed in *Aspergillus* species. The diversity of BGC types further underscores the strain's metabolic versatility, encompassing 27 NRPS, 16 T1PKS, 7 hybrid clusters, 9 terpene-related clusters, 6 RiPP-related clusters and 13 clusters associated with other secondary metabolites. Particularly noteworthy is the presence of 6 RiPP-related BGCs, while NRPS and PKS type BGCs are commonly observed in *Aspergillus* species [53]. The relatively high number of RiPP clusters in Hmp-F48 suggests a significant capacity for RiPP biosynthesis and merits further investigation. Several of the most homologous BGCs (1–11) also show partial loss of synteny with their closest references and display gene content variation, such as additional or missing genes (e.g., Cluster 2 and 4). Meanwhile, a considerable number of BGCs with low homology to known clusters incomplete clusters, or clusters lacking core genes remain uncharacterized. This suggests the possibility of unique, novel metabolites and pathways awaiting discovery. Future research should therefore not only characterize these cryptic clusters but also assess the biological activities of selected metabolites to clarify their ecological and biotechnological relevance.

Our experimental isolation efforts primarily led to structure determination of polyketides, such as averantin, sterigmatocystin, and decumbenone derivatives. These findings likely result from the inherent bias of conventional organic solvent extraction and column chromatography, which favors moderately polar PKS products while potentially overlooking hydrophilic peptides or volatile terpenoids. Expanding the range of extraction techniques — including advanced liquid-liquid partitioning, solid-phase extraction, and headspace trapping — may enhance the recovery of diverse metabolites. Additionally, optimizing culture conditions (e.g., altering nutrient

composition, co-cultivation, or stress induction) and exploring heterologous expression or epigenetic manipulation could help activate cryptic BGCs [54].

Although Hmp-F48 exhibits remarkable potential for secondary metabolite production, it should be noted that the current isolation and identification efforts were restricted to major metabolites under specific culture conditions. Consequently, these findings cannot reflect the strain's comprehensive biosynthetic repertoire. Future studies employing an integrated multi-omics approach—encompassing transcriptomics, proteomics, and metabolomics—across diverse culture conditions are expected to address this limitation. This strategy will enable us to link predicted gene clusters with detected metabolites and elucidate the strain's unique biosynthetic machinery [55, 56], ultimately facilitating the discovery of novel chemical scaffolds and biocatalysts.

Conclusion

This study presents a comprehensive genomic analysis of the sponge-associated fungus *A. puulaauensis* Hmp-F48, offering valuable insights into its genetic architecture and biosynthetic capacities. The high-quality genome assembly (35.86 Mb) and the annotation of 10,611 protein-coding genes illustrate the strain's extensive metabolic potential. Functional annotation across multiple databases (Nr, Swiss-Prot, GO, KEGG, eggNOG, and CAZy) revealed a broad spectrum of metabolic pathways and a substantial proportion of uncharacterized genes, indicating the possibility of discovering novel bioactive compounds. AntiSMASH predictions identified 78 BGCs, predominantly involved in the biosynthesis of polyketides and nonribosomal peptides. The detailed analysis of conserved clusters confirmed Hmp-F48's ability to produce notable bioactive metabolites, including the proteasome inhibitor fellutamide B and the antimicrobial burnettramic acid A. The isolation of a potentially novel decumbenone derivative (compound 4) further supports the strain's ability to generate unique secondary metabolites. Despite the promising genomic data, the limited diversity of isolated metabolites highlights the need for deeper exploration of the strain's secondary metabolism. Targeted activation of cryptic BGCs, and heterologous expression techniques should be employed to maximize the recovery of structurally diverse and biologically significant metabolites. Ultimately, this research underscores the potential of Hmp-F48 as a promising resource for natural product discovery, with implications for biotechnology and drug development.

Abbreviations

NSCLC	Non-small cell lung cancer
CIN	Chemotherapy-induced neutropenia
BGCs	Biosynthetic gene clusters
PDA	Potato dextrose agar
PDB	Potato dextrose broth
ITS	Internal transcribed spacer region
SEM	Scanning electron microscopy
CDS	Protein coding sequence
Nr	Non-redundant protein sequence
COG	Cluster of orthologous group of proteins
KEGG	Kyoto encyclopedia of gene and genome
GO	Gene ontology
CAZy	Carbohydrate-active enzymes
GT	Glycosyl transferases
PL	Polysaccharide lyases
CE	Carbohydrate esterases
AA	Auxiliary activities
CBM	Carbohydrate-binding modules
GH	Glycoside hydrolases
PKS	Polyketide synthase
NRPS	Nonribosomal peptide synthetase
NOESY	Nuclear overhauser effect spectroscopy
MTPA	α-Methoxy-α-(trifluoromethyl) phenylacetic acid

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12569-2.

Supplementary Material 1.

Acknowledgements

Not applicable.

Authors' contributions

N.W. and N.L. conceived the project; N.W., N.L., Y.Y., and X.W. designed the experiments; Y.Y. performed bioinformatics, X.W. and Y.Y. performed natural product isolation and structure determination; N.W., Y.Y., X.W., N.L., Q.M. Y.L. and Z.C. analyzed the results; Y.Y. wrote the first draft with input from all co-authors; N.W. and Y.Y. revised and finalized the manuscript. All authors reviewed the manuscript.

Funding

This work was financially supported by the Key Research and Development Program of Guangxi (AB24010108) and the National Natural Science Foundation of China (22077113).

Data availability

Sequence data that support the findings of this study have been deposited in the NCBI database under the BioProject PRJNA1249370, with the accession number JBNBLO010000000.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 12 May 2025 / Accepted: 16 January 2026
Published online: 26 January 2026

References

1. Carroll AR, Copp BR, Davis RA, Keyzers RA, Prinsep MR. Marine natural products. Nat Prod Rep. 2019;36(1):122–73.
2. Carroll AR, Copp BR, Davis RA, Keyzers RA, Prinsep MR. Marine natural products. Nat Prod Rep. 2022;39(6):1122–71.

3. Liu Y, Zhou Z, Sun S. Prospects of marine-derived compounds as potential therapeutic agents for glioma. Pharm Biol. 2024;62(1):513–26.
4. Bharathi D, Lee J. Recent advances in marine-derived compounds as potent antibacterial and antifungal agents: a comprehensive review. Mar Drugs. 2024;22(8):348.
5. Varijakzhan D, Loh JY, Yap WS, Yusoff K, Seboussi R, Lim SHE, et al. Bioactive compounds from marine sponges: fundamentals and applications. Mar Drugs. 2021;19(5):246.
6. Qiu Y, Chen S, Yu M, Shi J, Liu J, Li X, et al. Natural products from marine-derived fungi with anti-inflammatory activity. Mar Drugs. 2024;22(10):433.
7. Li H, Fu Y, Song F. Marine *Aspergillus*: a treasure trove of antimicrobial compounds. Mar Drugs. 2023;21(5):277.
8. Sun L, Wang H, Yan M, Sai C, Zhang Z. Research advances of bioactive sesquiterpenoids isolated from marine-derived *Aspergillus* Sp. Molecules. 2022;27(21):7376.
9. Saeed AF, Su J, Ouyang S. Marine-derived drugs: recent advances in cancer therapy and immune signaling. Biomed Pharmacother. 2021;134:111091.
10. Qian Y, Cheng B, Zhijun T, Wen L. Applications and prospects of genome mining in the discovery of natural products. Synth Biol J. 2021;2(5):657–715.
11. Albarano L, Esposito R, Ruocco N, Costantini M. Genome mining as new challenge in natural products discovery. Mar Drugs. 2020;18(4):199.
12. Alam K, Islam MM, Gong K, Abbasi MN, Li R, Zhang Y, et al. In silico genome mining of potential novel biosynthetic gene clusters for drug discovery from *Burkholderia* bacteria. Comput Biol Med. 2022;140:105046.
13. Bauman KD, Butler KS, Moore BS, Chekan JR. Genome mining methods to discover bioactive natural products. Nat Prod Rep. 2021;38(11):2100–29.
14. Wang X, Mou Y, Hu J, Wang N, Zhao L, Liu L, et al. Cytotoxic polyphenols from a sponge-associated fungus *Aspergillus versicolor* Hmp-48. Chem Biodivers. 2014;11(1):133–9.
15. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5):722–36.
16. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE. 2014;9(11):e112963.
17. Seppey M, Manni M, Zdobnov EM. Busco: assessing genome assembly and annotation completeness. Methods Mol Biol. 2019;1962:227–45.
18. Steenwyk JL, Balamurugan C, Raja HA, Gonçalves C, Li N, Martin F, et al. Phylogenomics reveals extensive misidentification of fungal strains from the genus *Aspergillus*. Microbiol Spectr. 2024;12(4):e0398023.
19. Consortium GO. The gene ontology resource: 20 years and still going strong. Nucleic Acids Res. 2019;47(D1):D330–38.
20. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45(D1):D353–61.
21. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. Eggnog-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021;38(12):5825–29.
22. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42(D1):D490–95.
23. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, Van Wezel GP, Medema MH, et al. AntiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 2021;49(W1):W29–35.
24. Szewczyk E, Chiang YM, Oakley CE, Davidson AD, Wang CC, Oakley BR. Identification and characterization of the asperthecin gene cluster of *Aspergillus Nidulans*. Appl Environ Microbiol. 2008;74(24):7607–12.
25. Palmer JM, Wiemann P, Greco C, Chiang YM, Wang CC, Lindner DL, et al. The sexual spore pigment asperthecin is required for normal ascospore production and protection from UV light in *Aspergillus nidulans*. J Ind Microbiol Biotechnol. 2021;48(9–10):kuab055.
26. Paranjape SR, Chiang YM, Sanchez JF, Entwistle R, Wang CC, Oakley BR, et al. Inhibition of Tau aggregation by three *Aspergillus nidulans* secondary metabolites: 2, ω-dihydroxyemodin, asperthecin, and asperbenzaldehyde. Planta Med. 2014;80(01):77–85.
27. Cary J, Ehrlich K, Beltz S, Harris-Coward P, Klich M. Characterization of the *Aspergillus ochraceoroseus* aflatoxin/sterigmatocystin biosynthetic gene cluster. Mycologia. 2009;101(3):352–62.
28. Zingales V, Fernández-Franzón M, Ruiz MJ. Sterigmatocystin: occurrence, toxicity and molecular mechanisms of action–a review. Food Chem Toxicol. 2020;146:111802.
29. Grijseels S, Pohl C, Nielsen JC, Wasil Z, Nygård Y, Nielsen J, et al. Identification of the decumbenone biosynthetic gene cluster in *Penicillium decumbens* and the importance for production of calbistrin. Fungal Biol Biotechnol. 2018;5:1–17.
30. Tao H, Mori T, Wei X, Matsuda Y, Abe I. One polyketide synthase, two distinct products: trans-acting enzyme-controlled product divergence in calbistrin biosynthesis. Angew Chem Int Ed. 2021;60(16):8851–58.
31. Brill GM, Chen RH, Rasmussen RR, Whittern DN, Mcalpine JB. Calbistrins, novel antifungal agents produced by *Penicillium restrictum* II. isolation and elucidation of structure. J Antibiot. 1993;46(1):39–47.
32. Zhuravleva OI, Afiyatullov SS, Vishchuk OS, Denisenko VA, Slinkina NN, Smetanina OF. Decumbenone c, a new cytotoxic decaline derivative from the marine fungus *Aspergillus sulphureus* KMM 4640. Arch Pharm Res. 2012;35:1757–62.
33. Sanchez JF, Chiang YM, Szewczyk E, Davidson AD, Ahuja M, Elizabeth Oakley C, et al. Molecular genetic analysis of the Orsellinic acid/F9775 gene cluster of *Aspergillus nidulans*. Mol Biosyst. 2010;6(3):587–93.
34. Klitgaard A, Nielsen JB, Frandsen RJ, Andersen MR, Nielsen KF. Combining stable isotope labeling and molecular networking for biosynthetic pathway characterization. Anal Chem. 2015;87(13):6520–26.
35. Raffa N, Keller NP. A call to arms: mustering secondary metabolites for success and survival of an opportunistic pathogen. PLoS Pathog. 2019;15(4):e1007606.
36. Henke MT, Soukup AA, Goering AW, Mcclure RA, Thomson RJ, Keller NP, et al. New aspercryptins, lipopeptide natural products, revealed by HDAC inhibition in *Aspergillus nidulans*. ACS Chem Biol. 2016;11(8):2117–23.
37. Yeh HH, Ahuja M, Chiang YM, Oakley CE, Moore S, Yoon O, et al. Resistance gene-guided genome mining: serial promoter exchanges in *Aspergillus nidulans* reveal the biosynthetic pathway for fellutamide B, a proteasome inhibitor. ACS Chem Biol. 2016;11(8):2275–84.
38. Lin G, Li D, Chidawanyika T, Nathan C, Li H. Fellutamide B is a potent inhibitor of the *Mycobacterium tuberculosis* proteasome. Arch Biochem Biophys. 2010;501(2):214–20.
39. Guo CJ, Yeh HH, Chiang YM, Sanchez JF, Chang SL, Bruno KS, et al. Biosynthetic pathway for the epipolythiodioxopiperazine Acetylaranotin in *Aspergillus terreus* revealed by genome-based deletion analysis. J Am Chem Soc. 2013;135(19):7205–13.
40. Huber EM. Epipolythiodioxopiperazine-based natural products: building blocks, biosynthesis and biological activities. ChemBioChem. 2022;23(23):e202200341.
41. Li H, Gilchrist CL, Lacey HJ, Crombie A, Vuong D, Pitt JI, et al. Discovery and heterologous biosynthesis of the burnettramic acids: rare PKS-NRPS-derived bolaamphiphilic pyrrolizidinediones from an Australian fungus, *Aspergillus burnettii*. Org Lett. 2019;21(5):1287–91.
42. Aldholmi M, Marchand P, Ourliac-Garnier I, Le Pape P, Ganesan A. A decade of antifungal leads from natural products: 2010–2019. Pharmaceuticals. 2019;12(4):182.
43. Sims JW, Fillmore JP, Warner DD, Schmidt EW. Equisetin biosynthesis in *Fusarium heterosporum*. Chem Commun. 2005;2:186–88.
44. Xu Z, Liu D, Liu D, Ren X, Liu H, Qi G, et al. Equisetin is an anti-obesity candidate through targeting 11β-HSD1. Acta Pharm Sin B. 2022;12(5):2358–73.
45. Wei W, Chen P, Khan B, Tian K, Feng Y, Lv B, et al. Evaluation of equisetin as an anti-microbial and herbicidal agent from endophytic fungus *Fusarium* sp. JDJR1. Agronomy. 2023;14(1):31.
46. Yamamoto T, Tsunematsu Y, Noguchi H, Hotta K, Watanabe K. Elucidation of pyranonigrin biosynthetic pathway reveals a mode of tetramic acid, fused γ-pyrone, and exo-methylene formation. Org Lett. 2015;17(20):4992–95.
47. Cho N, Ransom TT, Sigmund J, Tran T, Cichewicz RH, Goetz M, et al. Growth inhibition of colon cancer and melanoma cells by versiol derivatives from a *Paraconiothyrium* species. J Nat Prod. 2017;80(7):2037–44.
48. Fu Y, Wu P, Xue J, Wei X, Li H. Versicorin, a new lovastatin analogue from the fungus *Aspergillus versicolor* SC0156. Nat Prod Res. 2015;29(14):1363–68.
49. Futagami T, Mori K, Kadooka C, Niihara H, Tashiro K, Tamaki H, et al. Chromosome-level genome sequence of *Aspergillus puulaauensis* MK2, a fungus isolated from a dead hard tick. Microbiol Resour Announc. 2021;10(36):e00372-21.
50. Wortman JR, Gilsenan JM, Joardar V, Deegan J, Clutterbuck J, Andersen MR, et al. The 2008 update of the *Aspergillus nidulans* genome annotation: a community effort. Fungal Genet Biol. 2009;46(Suppl 1):S2–13.
51. Brandt SC, Ellinger B, Nguyen TV, Harder S, Schlüter H, Hahnke RL, et al. *Aspergillus sydowii*: genome analysis and characterization of two heterologous expressed, non-redundant xylanases. Front Microbiol. 2020;11:2154.

52. Gupta SK, Srivastava M, Osmanoglu Ö, Xu Z, Brakhage AA, Dandekar T. *Aspergillus fumigatus* versus genus *Aspergillus*: conservation, adaptive evolution and specific virulence genes. Microorganisms. 2021;9(10):2014.
53. Uka V, Cary JW, Lebar MD, Puel O, De Saeger S, Diana Di Mavungu J. Chemical repertoire and biosynthetic machinery of the *Aspergillus flavus* secondary metabolome: a review. Compr Rev Food Sci Food Saf. 2020;19(6):2797–842.
54. Zhao F, Sun C, Liu Z, Cabrera A, Escobar M, Huang S, et al. Multiplex base-editing enables combinatorial epigenetic regulation for genome mining of fungal natural products. J Am Chem Soc. 2022;145(1):413–21.
55. Li Y, Zhan G, Tu M, Wang Y, Cao J, Sun S. A chromosome-scale genome and proteome draft of *Tremella fuciformis*. Int J Biol Macromol. 2023;247:125749.
56. Kang Z, Yuan X, Zhang C, Wang Y, Li L, Zheng Y. Genomic and multi-omics analysis of *Phlebopus portentosus*: effects of cultivation on secondary metabolites. J Fungi. 2025;11(4):323–54.

Publisher's note