Genome Sequencing and Evolutionary Analysis of Marine Gut Fungus *Aspergillus* sp. Z5 from *Ligia oceanica*

Xue Li[1], Jin-Zhong Xu[1], Wen-Jie Wang[1], Yi-Wang Chen[2], Dao-Qiong Zheng[1], Ya-Nan Di[1], Ping Li[2], Pin-Mei Wang[1] and Yu-Dong Li[2]

[1]Department of Marine Sciences, Ocean College, Zhejiang University, Hangzhou, China. [2]Department of Bioengineering, School of Food Science and Biotechnology, Zhejiang Gongshang University, Hangzhou, China.

Supplementary Issue: Bioinformatics Methods and Applications for Big Metagenomics Data

ABSTRACT: *Aspergillus* sp. Z5, isolated from the gut of marine isopods, produces prolific secondary metabolites with new structure and bioactivity. Here, we report the draft sequence of the approximately 33.8-Mbp genome of this strain. To the best of our knowledge, this is the first genome sequence of *Aspergillus* strain isolated from marine isopod *Ligia oceanica*. The phylogenetic analysis supported that this strain was closely related to *A. versicolor*, and genomic analysis revealed that *Aspergillus* sp. Z5 shared a high degree of colinearity with the genome of *A. sydowii*. Our results may facilitate studies on discovering the biosynthetic pathways of secondary metabolites and elucidating their evolution in this species.

KEYWORDS: marine isopod, *Aspergillus*, genome, secondary metabolites

SUPPLEMENT: Bioinformatics Methods and Applications for Big Metagenomics Data

CITATION: Li et al. Genome Sequencing and Evolutionary Analysis of Marine Gut Fungus *Aspergillus* sp. Z5 from *Ligia oceanica*. *Evolutionary Bioinformatics* 2016:12(S1) 1–4 doi: 10.4137/EBO.S37532.

TYPE: Short Report

RECEIVED: November 30, 2015. **RESUBMITTED:** February 15, 2016. **ACCEPTED FOR PUBLICATION:** February 21, 2016.

ACADEMIC EDITOR: Jike Cui, Associate Editor

PEER REVIEW: Three peer reviewers contributed to the peer review report. Reviewers' reports totaled 641 words, excluding any confidential comments to the academic editor.

FUNDING: The study was financially supported by grants from the National Natural Science Foundation of China (NSFC No. 41406141), the Scientific Research Fund of Zhejiang Provincial Education Department (Y201327852), and the Doctoral Fund of Ministry of Education of China (20130101120155). The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

CORRESPONDENCE: wangpinmei@zju.edu.cn; lyd@zjsu.edu.cn

Marine-derived metabolites continue to be a prolific source of bioactive natural products with a high tendency to become drug candidates.[1] Research interests in secondary metabolites (SMs) from marine-derived fungi have increased in recent years, as many of them are structurally unique and possess interesting biological and pharmacological properties.[2] Various microbes, including fungi, are found in the gut of marine isopods[3–6] and may serve as food or defense for marine isopods by producing SMs.[7] While studies have shown that gut microbes exist in the worldwide distributed genus *Ligia*,[8] knowledge of their SMs is scanty. Gut-derived fungi from *Ligia* may be a prolific source of bioactive marine natural products. In this study, a fungus was obtained from the gut of the marine isopod *Ligia oceanica* and was identified as *Aspergillus* sp. based on the analysis of the ribosomal internal transcribed spacers and the 5.8S rRNA gene sequence. This strain of *Aspergillus* sp. Z5 has been deposited in the China Center for Type Culture Collection with accession number CCTCC M 2015238.

The genus *Aspergillus* is the major contributor to new bioactive SMs of marine fungal origin.[2] Therefore, they are recognized as a prolific source of pharmacologically valuable new compounds.[9] We also obtained many compounds with new structure and bioactivity from *Aspergillus* sp. Z5 (unpublished data). The genome sequencing of *Aspergillus* sp. Z5 from the gut of marine isopod *L. oceanica* may provide fundamental molecular information on elucidating the metabolic pathway of SMs in this species.

For genome sequencing, *Aspergillus* sp. Z5 was cultured in glucose minimal media with 30 g/L sea salt and incubated at 30 °C, 200 rpm, for 24 hours. Isolation of genomic DNA isolation was performed as described by Green and Sambrook.[10] The genome of *Aspergillus* sp. Z5 was sequenced using the Illumina HiSeq 2000 technology at the Shanghai Majorbio Bio-pharm Technology Co., Ltd. A library with a fragment length of 300 bp was constructed, and a total of 24,591,619 paired-end reads were generated. Approximately 23,727,930 high-quality reads, which provided a 137.5-fold depth of coverage, were assembled with SOAPdenovo version 1.05.[11] Based on the assembly, the genome size was estimated to be 33.8 Mbp with a GC content of 50.5% (Table 1). The assembly was organized in 1,089 contigs, which are linked by paired-end reads into 246 scaffolds (>1,000 bp). The average base is found in a scaffold of N_{50} size 1,256,305 bp and a contig of N_{50} size 195,844 bp (Table 1). This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. LDZW00000000. The version described in this paper is LDZW01000000.

Table 1. Comparison of genome characteristics.

ATTRIBUTE	*ASPERGILLUS* SP. Z5	*A. NIDULANS* FGSC A4*	*A. ORYZAE* RIB40*	*A. FUMIGATUS* AF293*
Assembly size (bp)	33.8 M	30.5 M	37.9 M	29.4 M
Predicted ORFs	11791	10687	12090	9840
GC content (%)	50.5%	50.0%	46.3%	50.0%
Gene/genome (%)	55%	50%	45%	49%
Contigs	1089			
Contig N50 (bp)	195,844			
Contig N90 (bp)	37,806			
Scaffolds (>1000 bp)	246			
Scaffold N50 (bp)	1,256,305			
Scaffold N90 (bp)	381,266			

Note: *Data are from the calculation based on the database of AspGD (http://www.aspgd.org/).

Protein-coding sequences were predicted by Augustus 2.5.5 and annotated using BLAST searches of nonredundant protein from the NCBI database. The genome of *Aspergillus* sp. Z5 encodes 11,791 predicted proteins, and the total length of genes was 18.7 Mbp, which makes up 55.3% of the genome (Table 1). The ratio of gene/genome in *Aspergillus* sp. Z5 is higher than that in three known *Aspergillus* species (Table 1). The rest of the noncoding genomic sequences are made up of intergenic sequences (including introns, promoters, and terminators), noncoding RNAs (rRNAs and tRNAs), origins of DNA replication, centromeres, and telomeres. Of the total predicted genes, 11,178 genes (94.8%) encode known function proteins, 210 genes (1.8%) are considered to encode hypothetical proteins, and 403 (3.4%) genes have no match in the database.

Recently, molecular analyses of the genus *Aspergillus* have offered a better insight into their taxonomic and phylogenetic relations.[12,13] Phylogenetic relationships between filamentous fungi have often been based on ribosomal DNA sequences or single-gene families,[14,15] such as the genes encoding β-tubulin (*benA*), small rRNA subunit (*rns*), cytochrome oxidase subunit I (*cox1*), and RNA polymerase II second largest subunit gene (*rpb2*). Besides the ribosomal sequence analysis, we further performed multilocus sequence analysis of the four housekeeping genes (*benA*, *rns*, *cox1*, and *rpb2*)[13] from *Aspergillus* sp. Z5 and 23 *Aspergillus* species. The DNA sequences of these genes in each *Aspergillus* species were obtained from GenBank database and aligned with ClustalW for further analysis. The maximum-likelihood and neighbor-joining trees were generated using the MEGA 6 software with default settings. The phylogenetic analysis supported that *Aspergillus* sp. Z5 falls within the *A. versicolor* clade (Fig. 1A). Compared with the genome sequences available in the AspGD (http://www.aspgd.org/), the longest scaffold in the genome of *Aspergillus* sp. Z5 was 3.7 Mb in length and shared a high degree of colinearity with the reference chromosome of *A. sydowii* (Fig. 1B), a closely related species to *A. versicolor*.[16,17]

In order to predict the clusters involved in SM biosynthesis, the reported genome was analyzed using antiSMASH pipeline.[18] As shown in Figure 2, 89 SM biosynthetic gene clusters and partial clusters were predicted in *Aspergillus* sp. Z5, including 10 nonribosomal peptide synthetase (NRPS) clusters, 10 polyketide synthetase (PKS) clusters, 8 terpene clusters, 2 PKS/NRPS hybrid clusters, 1 terpene/PKS cluster, and 58 clusters designated *putative* or *other* (Fig. 2). In total, there are 2,018 genes related to SM production (Supplementary Table 1, Fig. 2). Although the types of SM clusters could be accurately predicted according to the core SM biosynthetic genes encoding backbone enzymes, it is still not possible to predict with accuracy the boundaries of SM gene clusters or the functions of some clusters without backbone enzymes.[19] This is due to the fact that many of the genes surrounding the core SM biosynthetic genes often have unknown functions, making predictions of their involvement in the biosynthetic process of the SM almost impossible.[19] Based on the number of similar cluster genes and the identity of cluster genes (cutoff criteria: identity of 40% and query coverage of 50%) with other SM biosynthetic gene clusters, the types of most putative clusters without backbone enzymes were predicted in this study (Supplementary Table 1). However, the elucidation of these SM biosynthetic gene clusters is mainly dependent on experimental verification, followed by identification and characterization of SMs produced by the deletion strains.

The increasing availability of *Aspergillus* genomes has led to a rapid identification of secondary metabolism biosynthetic pathways (SMBPs) in recent years. However, only a small part of SMBPs are conserved between even closely related species.[20] In this work, comparative analysis of SM cluster genes encoding backbone enzymes was performed between *Aspergillus* sp. Z5 and the other three well-annotated *Aspergillus* species (*Aspergillus nidulans*, *Aspergillus oryzae,* and *Aspergillus fumigatus*), revealing that some backbone enzymes are highly conserved among all these four *Aspergillus* species, including cluster 11 and 28 (siderophore), cluster 30, 32, 42, 64,

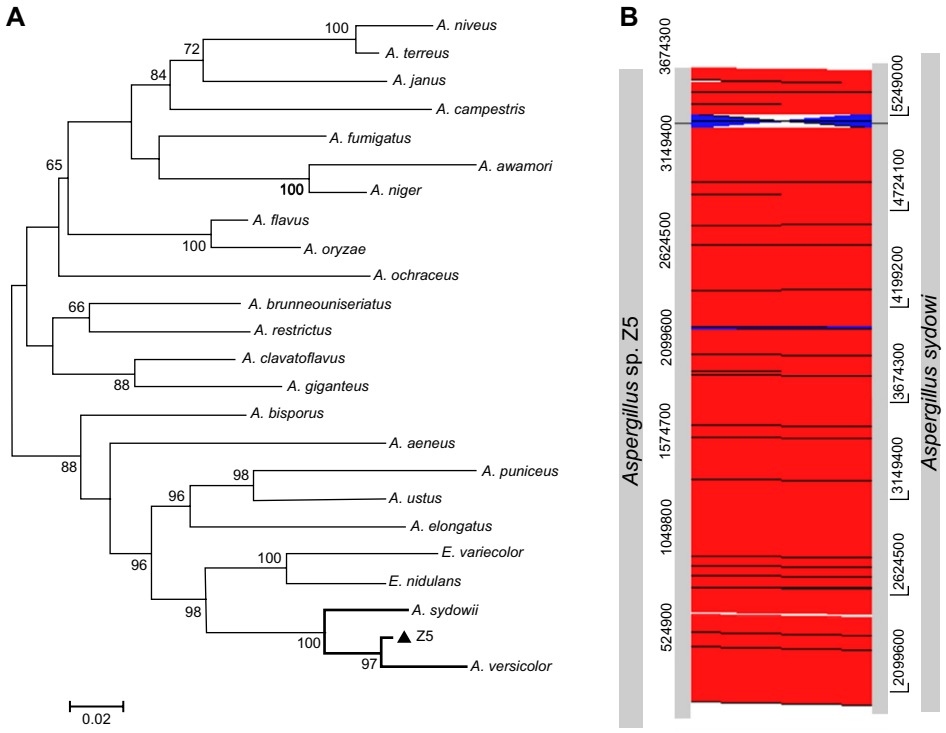

Figure 1. The evolutionary analysis of the genome of *Aspergillus* sp. Z5. (**A**) The phylogenetic position of *Aspergillus* sp. Z5 based on multilocus sequence analysis of the four housekeeping genes (*benA*, *rns*, *cox1*, and *rpb2*). The maximum-likelihood tree was generated using the MEGA6 software, and bootstrap values are shown as percentages (values below 60% are omitted). (**B**) The alignment of the largest scaffold in the *Aspergillus* sp. Z5 assembly was aligned against the genome of *A. sydowii* using BLASTN and visualized using the Artemis Comparison Tool.
Notes: Red lines between genomes indicate orthologous genes in the same orientation. Blue lines indicate orthologous genes in reverse orientation.

and 86 (unknown product), and cluster 35 (neosartoricin and fumicycline A). Besides these conserved backbone enzymes among the four species, *Aspergillus* sp. Z5 shared the greatest number of backbone enzymes with *A. nidulans*, including cluster 2 (emericellin), cluster 15 (orsellinic acid/F9975/violaceols), cluster 17 (asperthecin), cluster 63 (asperfuranone), cluster 79 (terrequinone), and other clusters with unknown products (Supplementary Table 1). Much less backbone enzymes are conserved in *A. oryzae* or *A. fumigatus* (Supplementary Table 1). Although the products of some clusters in *Aspergillus* sp. Z5

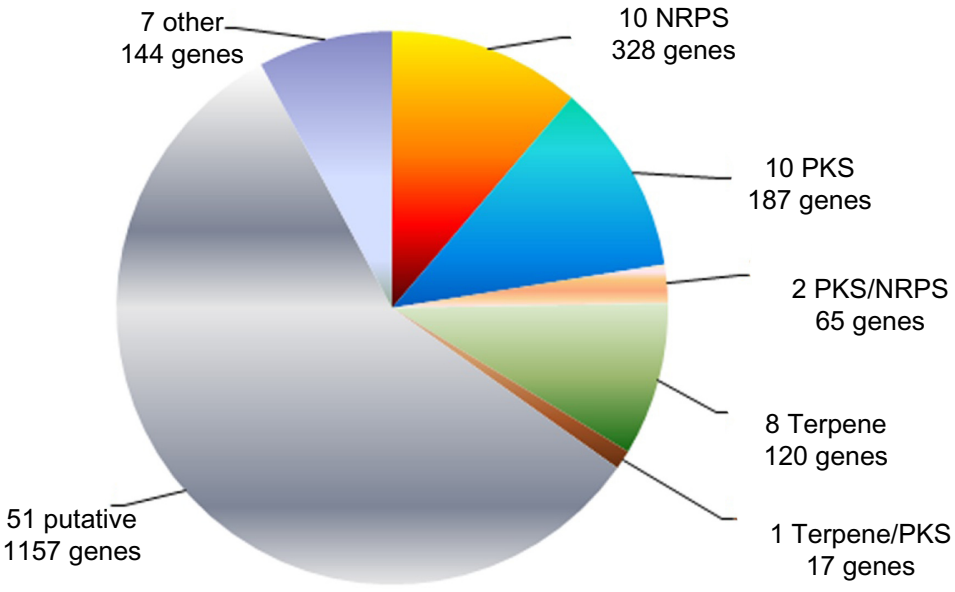

Figure 2. Graphical map of the putative SM gene clusters and the related gene numbers in the genome of *Aspergillus* sp. Z5.

were predicted based on the conserved sequences of SMBP backbone enzymes in genus *Aspergillus*, the functions of the rest of these 89 clusters remain unknown. In addition, some backbone enzymes with unknown functions were found in distantly related fungi such as *Penicillium*, *Metarhizium,* and *Tolypocladium*, but not in other *Aspergillus* species (cutoff criteria: identity of 40% and query coverage of 50%), ie, clusters 8, 14, 31, and 48, which are highly possible to be new compounds in *Aspergillus* species (Supplementary Table 1).

To the best of our knowledge, this is the first genome sequence of *Aspergillus* species isolated from the gut of marine isopod *L. oceanica*. Due to the limited knowledge and data about the marine isopod, it is currently impossible to explore the relationship between SMs from the *Aspergillus* species and its host. However, availability of this genome sequence presents a basis for exploring new bioactive compounds and elucidating the evolution of the metabolic pathway of SMs in marine-derived *Aspergillus* species.

Author Contributions

Conceived and designed the experiments: PMW, YDL. Analyzed the data: DQZ, YND. Wrote the first draft of the manuscript: XL, JZX. Contributed to the writing of the manuscript: WJW, YWC. Agree with manuscript results and conclusions: XL, JZX, WJW, YWC, DQZ, YND, PL, PMW, YDL. Jointly developed the structure and arguments for the paper: YND, PL. Made critical revisions and approved final version: XL, JZX, WJW, YWC, DQZ, YND, PL, PMW, YDL. All authors reviewed and approved of the final manuscript.

Supplementary Material

Supplementary Table 1. Secondary metabolites biosynthetic gene clusters in *Aspergillus* sp. Z5 and the backbone enzyme homologs in *A. nidulans*, *A. oryzae* and *A. fumigatus*. (XLS format).

REFERENCES

1. Ebada SS, Proksch P. Marine-Derived Fungal Metabolites in *Hb25_Springer Handbook of Marine Biotechnology*. Springer; NewYork, 2015.
2. Rateb ME, Ebel R. Secondary metabolites of fungi from marine habitats. *Nat Prod Rep*. 2011;28:290–344.
3. Xu J, Zhao S, Yang X. A new cyclopeptide metabolite of marine gut fungus from *Ligia oceanica*. *Nat Prod Res*. 2014;28:994–7.
4. Hernandez Roa JJ, Virella CR, Cafaro MJ. First survey of arthropod gut fungi and associates from Vieques, Puerto Rico. *Mycologia*. 2009;101:896–903.
5. Liu Y, Zhao S, Ding W, Wang P, Yang X, Xu J. Methylthio-aspochalasins from a marine-derived fungus *Aspergillus* sp. *Mar Drugs*. 2014;12:5124–31.
6. Raghukumar C. Marine fungal biotechnology: an ecological perspective. *Fungal Divers*. 2008;31:19–35.
7. Lindquist N, Barber PH, Weisz JB. Episymbiotic microbes as food and defence for marine isopods: unique symbioses in a hostile environment. *Proc Biol Sci*. 2005;272:1209–16.
8. Zimmer M, Danko J, Pennings S, et al. Hepatopancreatic endosymbionts in coastal isopods (Crustacea: Isopoda), and their contribution to digestion. *Mar Biol*. 2001;138:955–63.
9. Lee YM, Kim MJ, Li H, et al. Marine-derived *Aspergillus* species as a source of bioactive secondary metabolites. *Mar Biotechnol (NY)*. 2013;15:499–519.
10. Green MR, Sambrook J. *Molecular Cloning: A laboratory Manual*. Cold Spring Harbor Laboratory Press; NewYork, 2012.
11. Li R, Li Y, Kristiansen K, Wang J. SOAP: short oligonucleotide alignment program. *Bioinformatics*. 2008;24:713–4.
12. Samson R, Varga J. What is a species in *Aspergillus*? *Med Mycol*. 2009;47:13–20.
13. Krimitzas A, Pyrri I, Kouvelis VN, Kapsanaki-Gotsi E, Typas MA. A phylogenetic analysis of Greek isolates of *Aspergillus* species based on morphology and nuclear and mitochondrial gene sequences. *Biomed Res Int*. 2013;2013:1–18.
14. Samson RA. Current taxonomic schemes of the genus *Aspergillus* and its teleomorphs. *Biotechnology*. 1992;23:355–90.
15. Pel HJ, de Winde JH, Archer DB, et al. Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nat Biotechnol*. 2007;25:221–31.
16. Chiu YL, Liaw SJ, Wu VC, Hsueh PR. Peritonitis caused by *Aspergillus sydowii* in a patient undergoing continuous ambulatory peritoneal dialysis. *J Infect*. 2005;51:159–61.
17. Patterson TF, Kirkpatrick WR, White M, et al. Invasive aspergillosis. Disease spectrum, treatment practices, and outcomes. I3 *Aspergillus* study group. *Medicine (Baltimore)*. 2000;79:250–60.
18. Weber T, Blin K, Duddela S, et al. antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res*. 2015;43(W1):W237–43.
19. Yaegashi J, Oakley BR, Wang CC. Recent advances in genome mining of secondary metabolite biosynthetic gene clusters and the development of heterologous expression systems in *Aspergillus nidulans*. *J Ind Microbiol Biotechnol*. 2014;41:433–42.
20. Inglis DO, Binkley J, Skrzypek MS, et al. Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of *Aspergillus nidulans*, *A. fumigatus*, *A. niger* and *A. oryzae*. *BMC Microbiol*. 2013;13:1–21.