International Journal of Food Microbiology 380 (2022) 109859

Contents lists available at ScienceDirect

International Journal of Food Microbiology

journal homepage: www.elsevier.com/locate/ijfoodmicro

Safety evaluation and comparative genomics analysis of the industrial strain *Aspergillus flavus* SU-16 used for *huangjiu* brewing

Hailong Sun [a,b,1], Shuangping Liu [a,b,c,*,1], Jing Zhang [a,b], Songjing Zhang [b], Jieqi Mao [d], Yuezheng Xu [c], Jiandi Zhou [c], Jian Mao [a,b,c,*]

[a] National Engineering Research Center of Cereal Fermentation and Food Biomanufacturing, School of Food Science and Technology, Jiangnan University, Wuxi, Jiangsu 214122, China
[b] Shaoxing Key Laboratory of Traditional Fermentation Food and Human Health, Jiangnan University (Shaoxing) Industrial Technology Research Institute, Shaoxing, Zhejiang 312000, China
[c] National Engineering Research Center of Huangjiu, Zhejiang Guyuelongshan Shaoxing Wine Co., Ltd, Shaoxing, Zhejiang 31200, China
[d] Department of Food Science and Technology, National University of Singapore, Science Drive 2, 117542, Singapore

ARTICLE INFO

Keywords:
Huangjiu
Aspergillus flavus
Non-aflatoxigenic
Comparative genomics

ABSTRACT

Huangjiu is a popular Chinese traditional alcoholic beverage, while its brewing processes have rarely been explored. We herein report the first gapless, near-finished genome assembly of the industrial strain *Aspergillus flavus* SU-16 for *huangjiu* brewing. This work provides insights and supports for the further industrial applications of *A. flavus* isolates by comprehensively studying of the safety and genomic variations of SU-16. We demonstrated that SU-16 is a non-aflatoxigenic *A. flavus* at both molecular and metabolic levels. Using of nanopore sequencing technology resulted in a complete genome sequence for all 8 *A. flavus* chromosomes, as well as the mitochondrion. Genome comparisons of SU-16 with reference strains identified the chromosomal rearrangements, revealed the adaption mechanism of SU-16 to *huangjiu* ecological niche, and found that SU-16 is a good repository for CAZymes and some bioactive secondary metabolites. The results will help to develop more scientific *huangjiu* fermentation processes, and explore metabolism pathways of desired or harmful components in *huangjiu* to improve its quality.

1. Introduction

As one of the world's three ancient wines, *huangjiu* (Chinese rice wine) has been brewed and consumed for >5000 years (McGovern et al., 2004). *Huangjiu* is popular in southeast China due to its unique aroma, subtle flavor and low alcohol content (Sun et al., 2020). With thousands of years of historical and cultural accumulation, *huangjiu* has not only become an important part of the consumption of alcoholic beverages in China, but it also occupies an important position in traditional Chinese culture (Chen et al., 2013; Sun et al., 2020). *Huangjiu* is produced by a multispecies (such as molds, yeast, lactic acid bacteria and *saccharopolyspora*) fermentation with simultaneous saccharification and fermentation by using glutinous rice, water, and wheat qu as raw materials (Liu et al., 2019; Zhang et al., 2012). The saccharification and more generally proteolytic and metabolic activities of microorganisms in wheat qu not only fuel the yeast but also contribute metabolites that influence the flavor and aroma of *huangjiu* (Gibbons et al., 2012; Liu et al., 2019).

Traditional wheat qu, cooked wheat qu and inoculated raw wheat qu are the three main types of wheat qu used in *huangjiu* brewing, and they play key roles in substrate catabolism and biosynthesis of flavor compounds during *huangjiu* fermentation (Liu et al., 2020; Zhang et al., 2012). Traditional wheat qu is made by the spontaneous fermentation of crushed wheat and contains complex microbial communities (Liu et al., 2020; Wang et al., 2020). Cooked wheat qu (used for the fermentation of mechanized *huangjiu*) and inoculated raw wheat qu are produced by inoculating with a pure culture of strain SU-16 (*A. flavus* or *A. oryzae*) that produces a high level of hydrolase activity (Liu et al., 2020). SU-16 was isolated from Suzhou winery (China), and which has been used in *huangjiu* brewing for >60 years. However, there is no definite report on

* Corresponding authors at: National Engineering Research Center of Cereal Fermentation and Food Biomanufacturing, Jiangnan University, Wuxi, Jiangsu 214122, China.
E-mail addresses: liushuangping668@126.com (S. Liu), maojian@jiangnan.edu.cn (J. Mao).
[1] Hailong Sun and Shuangping Liu contributed equally to this work.

https://doi.org/10.1016/j.ijfoodmicro.2022.109859
Received 20 January 2022; Received in revised form 8 July 2022; Accepted 26 July 2022
Available online 29 July 2022

the taxonomic information and safety of strain SU-16 used in *huangjiu* brewing.

As is well-known, aspergilli can be found in almost all ecosystems have significant commercial relevance in industrial biotechnology (Cleveland et al., 2009; Kjaerbolling et al., 2020). Some species (*A. oryzae* and *A. sojae*) are usually used in the industrial production of sake, soy sauce, and other fermented foods (Gibbons et al., 2012; Machida et al., 2005; Zhao et al., 2013). Moreover, some species (*A. tamarii*, *A. niger* and *A. avenaceus*) are used industrially for production of enzymes (amylases, proteases and cellulase) and bioactive compounds (anti-insectant, antibiotic) (Kjaerbolling et al., 2020). However, species like *A. flavus* and *A. parasiticus* are notorious for the production of highly carcinogenic aflatoxins, and the contamination and damage to crops, food and feed (Cleveland et al., 2009; Frisvad et al., 2019), which have severely restricted their applications for the industrial production. Previously studies have shown polymorphisms in aflatoxin production for different *A. flavus* strains (Frisvad et al., 2019; Gibbons et al., 2012). There is evidence that certain non-aflatoxigenic *A. flavus* isolates obtained from the field have characteristics of *A. oryzae* (Chang and Ehrlich, 2010; Kjaerbolling et al., 2020). *A. oryzae* is atoxigenic and has been used as a source of industrial enzymes and as fermentation strains for some traditional fermented foods, such as sake, miso and soy sauce (Gibbons et al., 2012; Machida et al., 2005). Studies showed that *A. oryzae* is the domesticated forms of *A. flavus*, and the genome of *A. oryzae* strains are quite similar with those of *A. flavus* strains obtained from other environments (Frisvad et al., 2019; Gibbons et al., 2012). However, though SU-16 has been used in *huangjiu* brewing for a long time, the genome structure variants between SU-16 and the wild *A. flavus* and domesticated *A. oryzae* strains have never been characterized, as well as the functional difference to flavor formation of *huangjiu* and sake. Thus, safety assessment and the complete genome sequence of SU-16 is fundamental for understanding the genetic and regulatory systems of *huangjiu* strains.

To address these questions, taxonomic identification and safety assessment of SU-16 were carried out from multiple aspects in this study. Achievement of the gapless, near-finished genome sequence of SU-16 is a milestone in *huangjiu* industry. Comparative genomics analysis provides insights into the adaptation mechanism and functional differentiation of SU-16 to *huangjiu* ecology. This will help us to understand the safety and evolutionary history of beverage brewing strains, and provide insights into the improvement of the complicated brewing processes from the traditional and intuitive way to a modern and scientific way.

2. Materials and methods

2.1. Strains and morphological characteristics

Strain SU-16 was isolated from Suzhou Brewery (China Center of Industrial Culture Collection number CICC 2226/40336). The aflatoxigenic strain *A. flavus* NRRL3357 (American Type Culture Collection number ATCC 200026) and an *A. oryzae* strain MQ (China Center for Type Culture Collection number CCTCC M2015201) for sake brewing were used as the reference strains in genotypic analysis.

For macromorphological observations, Czapek Yeast Autolysate (CYA), Malt Extract Autolysate (MEA) agar, Yeast Extract Sucrose Agar (YES), Creatine Agar (CREA), and *A. flavus/parasiticus* Agar (AFPA) were used (Frisvad et al., 2019). Strains were inoculated at three points on each plate and incubated at 25 °C and 37 °C in the dark for 7 days (Varga et al., 2011).

2.2. Genotypic analysis of strain SU-16

DNA was extracted from 5 days-old colonies using the CTAB method (Liu et al., 2020). The ITS region and parts of the β-tubulin (*benA*) and calmodulin (*caM*) genes were amplified and sequenced as described previously (Luo et al., 2019; Samson et al., 2014; Varga et al., 2011). Sequencing of amplicons was performed by Sangon Biotech (Shanghai, China). Neighbor-joining trees were calculated using MEGA version 7.0 (Luo et al., 2019; Samson et al., 2014). To obtain the species-specific compatibility profiles of the strains, minisatellite-primed polymerase chain reaction (MSP-PCR) fingerprinting was carried out according to Luo et al. (2019) by using the synthetic oligonucleotides $(GTG)_5$ as single primer. NRRL3357 and MQ were used as the reference strains. Amplification products were separated by electrophoresis in 1.4 % (w/v) agarose gels in 0.5 × TAE (Tris Acetate EDTA). DNA-banding patterns were acquired with the Vilber Lourmat (Quantum CX5, Paris, France) and strains with identical DNA-banding patterns were considered to be more similar on the genetic level, and belong to the same species.

2.3. Detection of enzyme activity and aflatoxins

SU-16 was grown on sterilized wheat to model the industrial production of cooked wheat qu (Zhang et al., 2012). Wheat was crushed into 3 to 4 pieces, mixed with 40 % water (w/w), and then sterilized at 121 °C for 30 min. After cooling to 30 °C, SU-16 was inoculated to the sterilized wheat and which was then incubated at 30–35 °C for 3–4 days. After cultivation, surface of the wheat was covered with fungal mycelium. Enzyme activities including amylase, glucoamylase and protease present in cooked wheat qu were analyzed according to Liu et al. (2020). Aflatoxins were extracted by mixing vigorously in methanol solution as previously published (Frisvad et al., 2019; Toyotome et al., 2019). The extracts were purified by an immunoaffinity column (pribofast® M260, Pribolab Pte. Ltd., Singapore). The aflatoxin content was detected as described previously by using a Waters 2695 high performance liquid chromatograph (HPLC, Waters Co., Milford, USA) (Toyotome et al., 2019; Varga et al., 2011).

2.4. Whole-genome sequencing

Strain SU-16 was inoculated in potato dextrose broth medium and shake cultivation at 28 °C for 3 days. Mycelial pellets were washed and then lyophilized in a Martin Christ Freeze Dryers (Martin Christ, Osterode, Germany). High quality DNA from the lyophilized mycelial pellets was extracted using a QIAGEN® Genomic DNA kit (QIAGEN GMBH, Hilden, Germany). The extracted DNA was detected by NanoDrop™ One UV–Vis spectrophotometer (Thermo Fisher Scientific, USA) for DNA purity, then Qubit® 3.0 Fluorometer (Invitrogen, USA) was used to quantify DNA accurately. For the next generation sequencing, 1.5 μg DNA was used as input material for the DNA sample preparation (Zhang et al., 2018). The libraries were generated using Truseq Nano DNA HT Sample preparation Kit (Illumina, CA, USA) (Vesth et al., 2018). Next, the constructed libraries were sequenced on the Illumina Novaseq platform (Illumina, CA, USA). For the Nanopore sequencing, libraries were constructed using approximately 10 μg of genomic DNA. The DNA fragments were size selected from the qualified sample with a minimum cutoff of 20 kb using the BluePippin system (Sage Science, USA). Genomic libraries were prepared by ligation of hairpin adaptors at both ends of the selected DNA fragments using the Nanopore Sequencing Kit SQK-LSK109 (Oxford Nanopore Technologies, UK) according to manufacturer's instructions. Qubit® 3.0 Fluorometer (Invitrogen, USA) was used to quantify the size of library fragments, and then the DNA library was eluted and loaded onto the flow cell for sequencing at the GridION X5 instrument (Oxford Nanopore Technologies, UK).

2.5. De novo genome assembly and annotation

2.5.1. De novo assembly

To make sure reads reliable and without artificial bias in the following analyses, raw reads were firstly quality control-filtered to remove artifacts/process contamination (Jenjaroenpun et al., 2018; Payne et al., 2021). *De novo* genome assembly was performed for the Nanopore datasets by using Canu v1.8 (Koren et al., 2017; Senol Cali

et al., 2019). Nanopolish v0.8.4 and Pilon v1.22 were further used to correct and polish the assembled contigs with Illumina reads to obtain a high quality of genome (Zhang et al., 2018). Finally, the order and orientation of contigs from the *de novo* assembly were determined by aligning to the reference genome (*A. flavus* NRRL3357 and *A. oryzae* RIB40) using CLC Genomics Workbench v8.0.

2.5.2. Gene prediction and functional annotation

Protein-coding region identification and gene prediction were conducted through a combination of *ab initio* gene prediction, transcriptome-based prediction, and homology-based prediction methods (Jenjaroenpun et al., 2018; Zhang et al., 2018). All gene models predicted from the above three approaches were combined into a nonredundant set of gene structures. The gene models were further filtered based on their C-score, peptide coverage and its coding sequences overlapping with transposons using TransposonPSI software (Urasaki et al., 2017). The veracity and integrity of the assembly genome was evaluated by comparing the sequences of the genes predicted from the assembled genome with the conserved single-copy homologous genes in Orthodb database (https://www.orthodb.org/ using) by using BUSCO 3.0.1 (Simao et al., 2015). All predicted proteins were analyzed by BLAST against the databases of SwissProt, Cluster of Orthologous Groups of proteins (COGs), Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGGs) and Carbohydrate-Active enzymes (CAZy).

2.6. Analysis of the genome variations between SU-16 and reference strains

Illumina reads were mapped to reference genomes to estimate the genetic distance by using BWA-MEM with the default parameters (Zhang et al., 2018). SAMtools was used for sorting and the duplicates were removed using Picard (Li et al., 2009). Genetic variants identified above (SNPs and indels) were further annotated using SnpEff. To detect the fast-evolving genes, substitution rate of nonsynonymous to synonymous (*Ka/Ks*) relative to reference strains was assessed (Zhang et al., 2018). Genes with *Ka/Ks* value > 1 were classified as under positive selection. Reads were counted using a sliding window (1 kb) and used to find copy number variations (CNVs). CNVs were identified using HMMcopy based on the ReadDepth method (Ha et al., 2012). The public software BLASTN and Tbtools (Chen et al., 2020) were used for the genome comparisons. The results of these two analyses were checked manually and combined to identify the synteny, inversion, and translocation events of the genome. All results derived from the genome features and comparisons were summarized and visualized using R and TBtools.

2.7. Prediction of secondary metabolite biosynthesis gene clusters

For secondary metabolite biosynthesis gene clusters (SMGCs) prediction, combination of the anti-SMASH 5.0 program with manual correction was carried out (Blin et al., 2019). The complete genome assembly and predicted coding proteins were supplied to antiSMASH with default parameters. Each of the known cluster entries from the MIBiG database was used as a query against the genomes. The entry with the most genes was chosen for clusters that had redundant entries. Aflatoxin biosynthesis gene clusters were corrected by comparison of tested strains to the known cluster from the reference genomes. Most of the gene clusters were validated by aligning to the reference genomes using local BLASTN. Conservation of clusters was calculated as the percentage of genes in the query cluster conserved in the syntenic counterpart. TBtools, R package RIdeogram and EasyFig were used for visualization of cluster synteny and similarity (Chen et al., 2020; Kjaerbolling et al., 2020).

3. Results and discussion

3.1. Morphology and physiology of strain SU-16

Morphology forms an important part of the species concept of *Aspergillus* (Chen et al., 2017; Samson et al., 2014). Fig. 1 shows the colonies of strain SU-16 on the five mediums. SU-16 grows well on these mediums and the colonies mostly reach a diam of 4–6 cm within 7 days. It is more similar to *A. flavus* NRRL3357 (Fig. S1a). Colony characteristics of SU-16 on CYA at 25 °C are velutinous with dense sporulation and yellow mycelium. However, when at 37 °C, colonies are velvety, plain and low at margins, mycelium yellow-green. Colonies on MEA at 25 °C are slightly sulcate, velvety, mycelium yellow-green, sporulation dense. The mycelium color of SU-16 and NRRL3357 on YES is strontium yellow (25 °C) and dark yellow (37 °C) (Fig. 1 and S1a, b), whereas the mycelium color of MQ is white all the time (Fig. S1c, d). Aspergillic acid, which is a typical metabolite of *A. flavus* forms a ferrous iron complex that is readily expressed on AFPA as an orange reverse (Frisvad et al., 2019). SU-16 and NRRL3357 have orange reverse on AFPA, and the color at 25 °C is deeper than that at 37 °C, whereas MQ has a dark yellow color (Fig. S1). Acid production is often a useful character and this is observed on the purple medium CREA (Samson et al., 2014), which turns yellow when acid is produced by colonies. Different from MQ, colonies of SU-16 and NRRL3357 on CREA are velutinous, white or pale yellow (25 °C) to yellow green (37 °C). NRRL3357 is capable of acid production at both 25 and 37 °C, but SU-16 and MQ shows the opposite results when at 25 °C and 37 °C, respectively (Fig. S1). The results showed morphological diversity of the tested strains, and it is difficult to distinguish the differences between related species using cultivation. Whereas, orange reverse on the AFPA medium provided obvious results for distinguishing *A. flavus* from *A. oryzae* strains based on the characteristics of aspergillic acid production of *A. flavus* strains. Thus, SU-16 was preliminarily identified as an *A. flavus* strain according to its morphological characteristics.

3.2. Genotypic analysis

3.2.1. Sequencing and phylogenetic analysis of ITS, benA and caM gene

The phylogenetic analysis of SU-16 with *A. flavus* NRRL3357 and *A. oryzae* MQ was examined based on the sequence analysis of three loci: ITS, *benA* and *caM* gene (Frisvad et al., 2019; Luo et al., 2019). ITS, *benA* and *caM* gene sequence of SU-16 exhibited very high sequence identities (99–100 %) to the species of *A. flavus* and *A. oryzae*, which suggests a close relatedness between SU-16 and reference species in *Aspergillus* section *Flavi*. Strain SU-16 together with already-described *A. flavus* and *A. oryzae* strains formed a distinct *A. flavus/A. oryzae* clade in all analyses (Fig. 2a and S2). Although the *caM* gene sequence was more diverse than those of ITS and *benA* between SU-16 and reference strains, this diversity is limited to the substitution of a few single bases. Thus, phylogenetic analysis based on the single loci sequence (Fig. S2) and concatenated sequence data (Fig. 2a) of three loci of ITS, *benA* and *caM* gene sequence could not effectively distinguish SU-16 from *A. flavus* or *A. oryzae*.

3.2.2. MSP-PCR fingerprinting

Several studies have reported successful use of the DNA fingerprint technology to differentiate the interspecific and intraspecific differences of microorganisms, such as yeasts, molds and lactic acid bacteria (Luo et al., 2019). In this study, MSP-PCR fingerprint method was carried out for the further analysis of genetic differences between SU-16 and reference strains. As shown in Fig. 2b, $(GTG)_5$ -PCR fingerprints of strain SU-16, NRRL3357 and MQ had 4 to 5 bands respectively, and which were mainly concentrated in the range of 200–800 bp. The band profiles of SU-16 and NRRL3357 were almost identical, while MQ showed additional band in comparison to SU-16 and NRRL3357 at about 750 bp. The results clearly reflected much higher similarity in genetic level for

Fig. 1. Colonies of strain SU-16 incubated at 25 °C (a) and 37 °C (b) on CYA, MEA, YES, AFPA (reverse) and CREA medium for 7 days. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

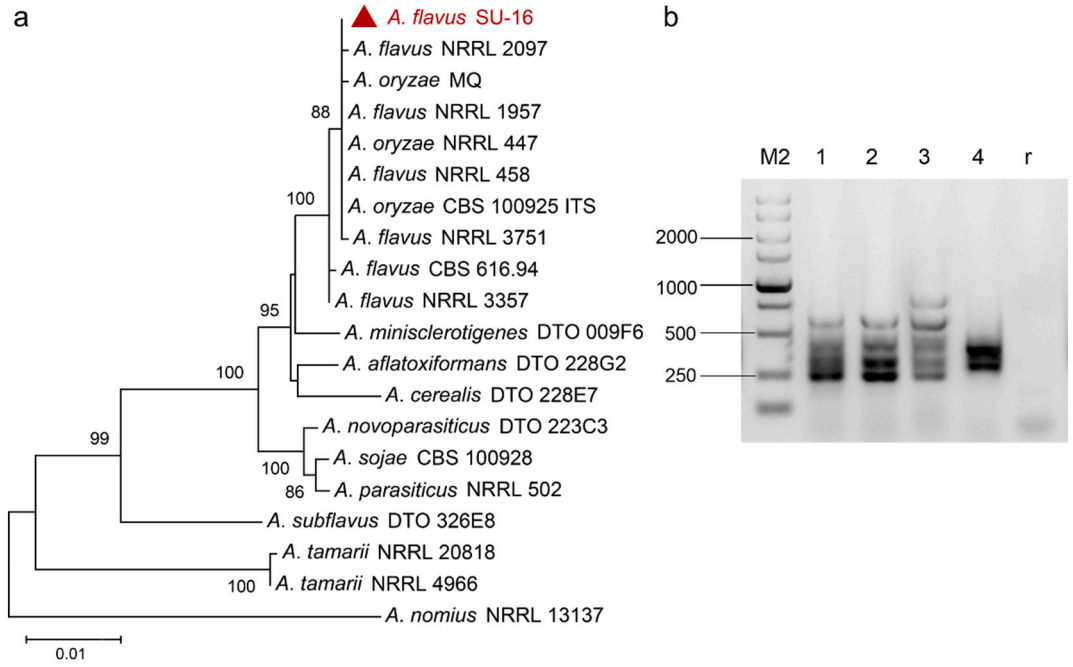

Fig. 2. Genotypic analysis of strain SU-16. (a) Phylogenetic tree based on the combined dataset of ITS, β-tubulin (*benA*) and calmodulin (*caM*) gene of strain SU-16 and other reference strains; (b) MSP-PCR fingerprinting pattern. M2: DL5000 DNA Marker, 1–4: *A. flavus* NRRL3357, SU-16, *A. oryzae* MQ and *Saccharomyces cerevisiae*, r: Blank control of PCR.

SU-16 and *A. flavus* NRRL3357 than *A. oryzae* MQ.

3.3. SU-16 is a non-aflatoxigenic Aspergillus flavus

Based on the polyphasic-analysis of phenotypic characteristics and genotype comparisons above, SU-16 was identified as an *A. flavus* strain. To determine the safety of *A. flavus* SU-16 in industrial applications, its ability to produce aflatoxin was further investigated. SU-16 was used to model the industrial production of cooked wheat qu, one of the main fermentation starters for *huangjiu* brewing. The hydrolytic enzyme activities of cooked wheat qu were analyzed firstly and which showed ideal glucoamylase, amylase and protease activities (Table S1). The aflatoxins (B_1, B_2, G_1 and G_2) in cooked wheat qu were examined using the HPLC system. Results showed that none of the compounds were detected (Table S2). Thus, we further concluded that SU-16 is a non-aflatoxigenic *A. flavus* strain according to the results of aflatoxins detection.

3.4. Nanopore sequencing enables near-complete genome assembly of SU-16

To obtain a complete chromosome level *de novo* assembly of *A. flavus* SU-16, long-read sequencing on the nanopore sequencing platform was performed. 570-Fold coverage of reads from nanopore sequencing platform were produced and the details of the third-generation sequencing reads after quality control are provided in Table S3. *De novo* assembly using the Canu software was performed and the assembly was then corrected and polished using Illumina reads (Senol Cali et al., 2019). This yielded a genome assembly of 37.56 Mb (9 contigs) with a contig N50 size of 4.77 Mb, and the longest contig is 6.57 Mb (Table S4). For the structural accuracy of the assembly, 99.91 % of the mapped Illumina reads of SU-16 could be mapped with the correct orientation and estimated insert size. The assembly was aligned against the *A. oryzae* RIB40 genome sequence to determine the order and orientation of contigs in the SU-16 genome. The resulting *de novo* assembly produced full-length, contiguous DNA sequences for 8 chromosomes and 1

mitochondrion, and it showed high collinearity and structural conservation by comparison of SU-16 genome with NRRL3357 and RIB40 chromosomes (Fig. S3). Finally, the gapless and near-complete genome sequence of SU-16 was generated (Fig. 3a, Table S4), and which has been deposited in the GenBank (CP047249–CP047257).

High-quality gene annotation of SU-16 genome based on the combination of *ab initio* prediction, transcriptome-based prediction and homology-based prediction methods was carried out. A total of 12,332 protein-coding genes (Table S4) were retrieved from the genome (Fig. 3a). On average, SU-16 genes encode transcripts of 1647.1 bp and contain 3.27 exons with a length of 444.87 bp, comparable to *A. flavus* NRRL21882 and *A. oryzae* 3.042 (Zhao et al., 2013), but longer than *A. flavus* NRRL3357 and *A. oryzae* RIB40 (Machida et al., 2005) (Table S4). The annotated genes covered 98.97 % of the conserved single-copy homologous genes in BUSCO database (Table S5) and 99.72 % of which were annotated with known proteins and/or domains (Table S6), which indicated high quality of genome assembly and annotation.

3.5. Genome variations of strain SU-16 compared with NRRL3357 and RIB40

Intergenomic analyses between SU-16, NRRL3357 and RIB40 revealed highly conserved collinearity (Fig. 3b), which supports a close evolutionary relationship among these strains (Gibbons et al., 2012). Notwithstanding, comparative genomics has also revealed a number of significant differences in genome configuration. There are 13 and 21 large regions identified as synteny, translocation, or inversion of the chromosomes when compared to NRRL3357 and RIB40, respectively as illustrated in Fig. 3b. The main differences relative to NRRL3357 are one large inversion of 373 kb on chromosomes 3, the translocation of a 11.6 kb fragment from the right arm of chromosome 6 to the left arm of chromosome 7, and inversion of the complete chromosome 8. However, more structural variations were examined between SU-16 and RIB40 despite the much closer relationship in function and geography (Gibbons et al., 2012; Machida et al., 2005). Such as two large translocation regions of 951.2 kb and 969.7 kb on chromosome 2 and 6 respectively, the near complete inversion of chromosome 1 and 7, and the large fragment translocation between chromosome 1, 4, 6 and 8 (Fig. 3).

As shown in Table S7, when compared with NRRL3357, a total of 179,322 SNPs and 12,625 small indels (<100 bp) were identified, whereas only about half numbers of SNPs and indels (93,001 and 6876) were identified when comparing to RIB40. These polymorphisms are not uniformly distributed (Fig. 3a) and most of the indels are short (≈55 % are 1 bp in length). Even though >50 % of the SU-16 genome is non-coding, <20 % of the detected SNPs were intergenic. More than 38 % of the detected SNPs in the protein-coding regions resulted in non-synonymous substitutions, and some of the genes have gained or lost stop codons. In addition, fast-evolving genes were identified by assessment of the substitution rate of nonsynonymous to synonymous (*Ka/Ks*) relative to the wild *A. flavus* NRRL3357 (Zhang et al., 2018). Results showed that 1937 genes (*Ka/Ks* > 1) in NRRL3357 were identified as the positive selection genes (Table S8). Furthermore, CNVs were also characterized, 366 and 286 duplication-deletion events were detected when compared with NRRL3357 and RIB40 respectively (Table S9). CNVs are mainly enriched in subtelomeric and centromeric regions, whereas internal chromosomal regions are largely copy number stable.

Domestication is driven by adaptation to environmental niches, as a consequence humans use these traits for our benefit (Gallone et al., 2016; Gibbons et al., 2012). Fermentation experiment results showed excellence in enzyme production and non-aflatoxigenic characteristics of SU-16. Previous studies showed that *A. oryzae* is the domesticated form of *A. flavus* (Frisvad et al., 2019; Gibbons et al., 2012). The genome of *A. oryzae* strains are quite similar with those of *A. flavus* strains obtained from other environments. However, some genomic variations have explained the differences between them to some extent. Function analysis revealed that most of the genes associated with these genomic variations are involved in primary and secondary metabolism, such as the hydrolysis of carbohydrates and amino acid transportation, which play roles in both substrate degradation and production of flavor compounds. The variations between these strains are largely associated with artificial domestication and adaptation to specific niches (Duan et al., 2018; Gibbons et al., 2012). In addition, metabolic differences caused by the genome variations could be the main driving force for the functional differentiation of *A. flavus* and *A. oryzae* strains during domestication (Gibbons et al., 2012; Machida et al., 2005).

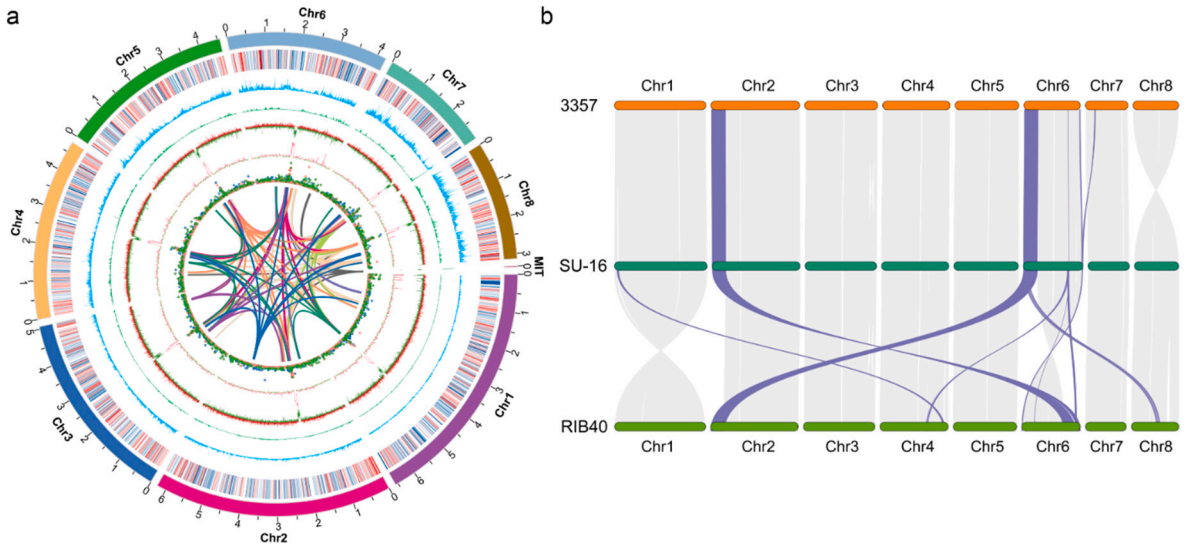

Fig. 3. Genomic features of *A. flavus* SU-16. (a) Basic characteristics of the eight chromosomes and landscape of the genome variations of strain SU-16. Circles represent, from outermost to innermost. Circle 1: The 8 chromosomes of SU-16; Circle 2: Gene density along each chromosome; Circle 3 and 4: Density of SNPs in SU-16 compared with NRRL3357 and RIB40; Circle 5 and 6: GC content and GC skew; Circle 7: Density of the simple sequence repeats (blue), Tandem repeats (pink) and transposable elements (green). The innermost: intra-genome collinear blocks connected by curved lines. (b) Genomic synteny between SU-16 with NRRL3357 and RIB40. The highlighted lines indicate the large structure variations (inversion and translocation). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.6. *A. flavus* SU-16 is a rich source of carbohydrate-active enzymes

Rice and wheat are the main raw materials for *huangjiu* brewing, which contain a large content of starch, cellulose, pectin and other plant polysaccharides. *Saccharomyces cerevisiae*, the core microorganism in *huangjiu* fermentation process, plays a key role in the quality and formation of the flavor of *huangjiu* (Duan et al., 2018; Liu et al., 2019). However, because of the absence of several enzymes (such as α-amylase) associated with primary carbon metabolism in *S. cerevisiae* (Aguilar-Pontes et al., 2018; Liu et al., 2019), the normal fermentation of *huangjiu* requires a large number of glycosyl hydrolases secreted by other species to digest the insoluble cellulose, starch and other macromolecules. In the previous study, the CAZymes/carbon utilization is mainly described for *A. oryzae* (Gibbons et al., 2012; Ichishima, 2016; Machida et al., 2005), and to a lesser extent for *A. flavus* (Cleveland et al., 2009). To better understand the carbohydrate degradation capacity of SU-16 in the *huangjiu* brewing, genes encoding CAZymes were annotated.

It is interesting to note that though most of the gene-encoding CAZymes in *A. flavus* and *A. oryzae* have a high degree of conserved synteny, some variations in the gene copy numbers were found, as well as chromosomal rearrangement caused by inversions and translocations (Fig. S4). SU-16 possesses a moderate number of CAZymes compared with NRRL3357 and RIB40 (Fig. 4b, Table S10). SU-16 contains 603 CAZyme-encoding gene counts distributed unequally between glycoside hydrolases (53.07 %), glycosyl transferases (21.06 %), polysaccharide lyases (4.31 %), carbohydrate esterases (4.31 %), auxiliary activities (10.12 %) and carbohydrate binding modules (7.13 %). Although SU-16 shows excellent α-amylase activity (Table S1), only one copy of the α-amylase gene is found in the SU-16 genome (Fig. 4a). Compared to NRRL3357 and RIB40, a higher number of cellulases, pectinases and xyloglucanases, and lower number of xylanases, mannanases and genes related to starch digestion are found for SU-16 (Fig. 4c, Table S10). All three strains have the identical number of genes related to the degradation of inulin and chitin. When comparing specific enzymes of SU-16 with these two strains, variable gene numbers were observed for *exo*-β-glucosaminidase, β-galactosidase and β-glucuronidase, α-L-rhamnosidase, Beta-L-arabinobiosidase, acetyl xylan esterase and acetylesterase. Glycogen debranching enzyme, α-arabinofuranosidase and amylo-α-1,6-glucosidase are only found in NRRL3357 and RIB40, whereas rhamnogalacturonan exolyase, L-rhamnose-α-1,4-D-glucuronate lyase and lytic cellulose monooxygenase are the SU-16 specific enzymes.

Overall, the variation in CAZymes within SU-16 genome is relatively low. Though there is no significant difference in the CAZyme content of these strains, they show different phenotypes in the growth and the application of fermentation industry (Frisvad et al., 2019; Gibbons et al., 2012). These might be mainly attributed to the artificial domestication, which had driven the adaptive microevolution of SU-16 from the saprotrophic lifestyle to fermentation environment. It is therefore likely that variation in the CAZyme content does not reflect the differences in phenotypes intraspecies, the strain-specific differences could be largely driven by the regulatory level (Gibbons et al., 2012; Kjaerbolling et al., 2020).

3.7. Secondary metabolite gene clusters in *A. flavus* SU-16

Secondary metabolism is considered to be an important component of chemical defense, virulence, toxicity, and communication in fungi (Cleveland et al., 2009; Kjaerbolling et al., 2020). SMGCs are generally recognized as the main difference in fungal species at genetic level (Frisvad et al., 2019; Machida et al., 2005). Thus, to gain insights into the potential for SM production and the strain specific genetic variation, SMGCs in genome of SU-16, NRRL3357 and RIB40 were compared. Within the three strains, there are a total of 233 predicted SMGCs (Fig. 5a, Table S11a–c). SMGCs distributed unequally on the 8 chromosomes and the core genes in SMGCs are found less often at the subtelomeric regions (Fig. 5a). About half of the identified SMGCs are conserved, and the number of the 7 types of the backbone genes in inter-strains SMGCs is found to vary inconspicuously. 70 of the predicted SMGCs are partially conserved synteny (50–90 %) or completely collinear, and these clusters are present in all three strains. This suggests the high diversity of SMs in *A. flavus* and *A. oryzae*, and the homogeneity

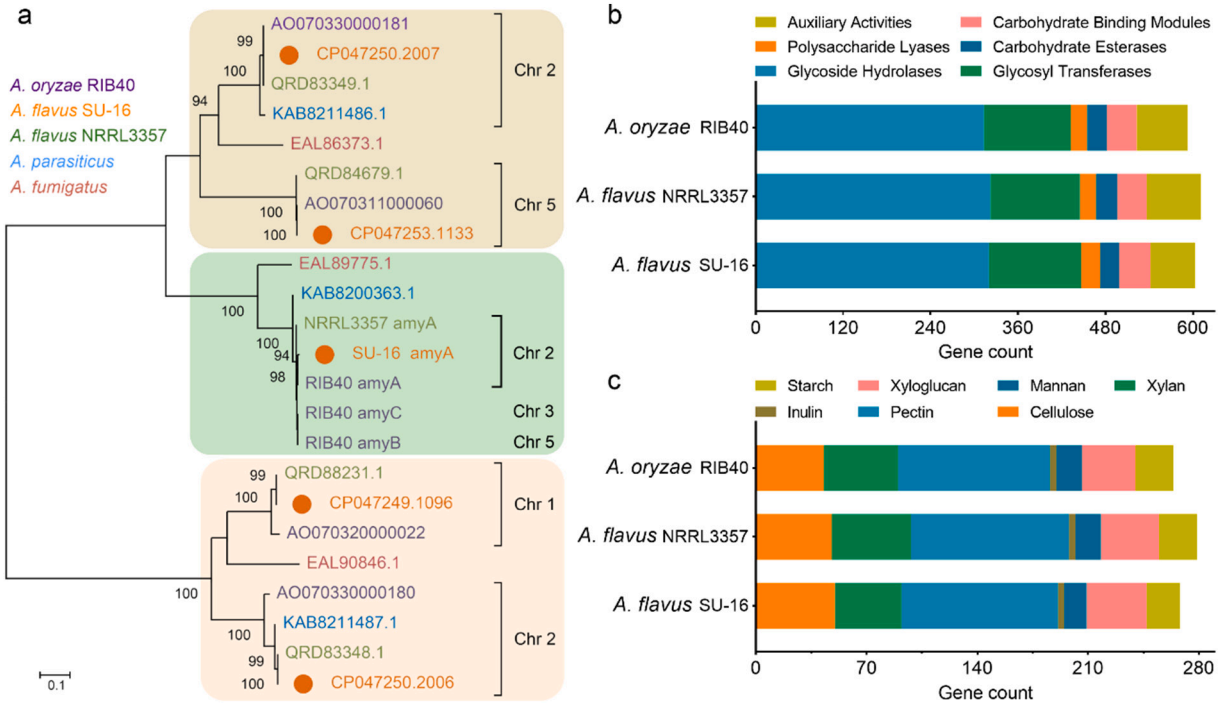

Fig. 4. Comparison of CAZyme-encoding genes between *A. flavus* SU-16, NRRL3357 and *A. oryzae* RIB40. (a) Phylogenetic analysis of α-amylase. The phylogenetic relationship of α-amylase homologues from the five strains was analyzed using MEGA 7.0. (b) The total number of CAZymes in each species distributed on six categories of enzyme activity. (c) Gene numbers related to degradation of different target polysaccharides; details on CAZy families are available in Table S10.

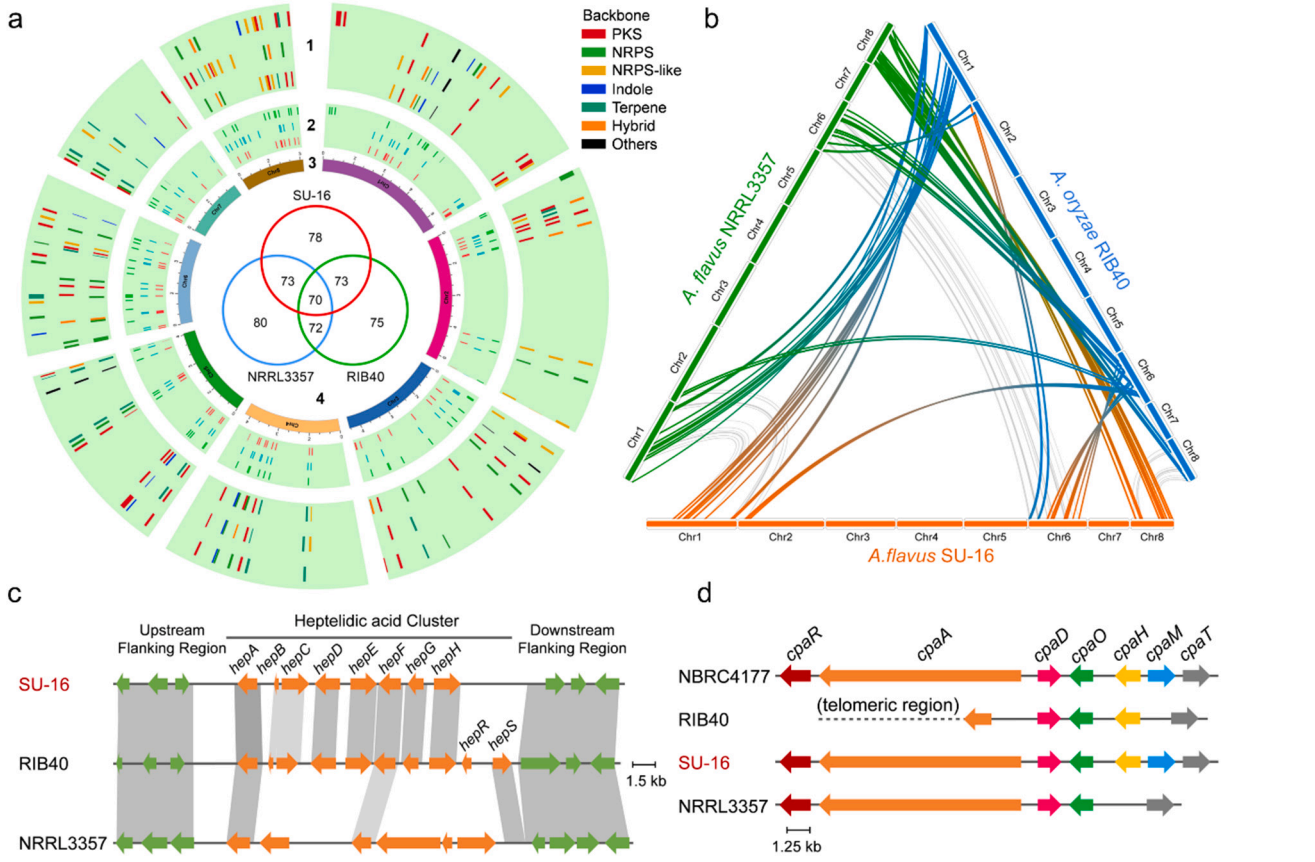

Fig. 5. Comparison of SMGCs and types of cluster backbones between *A. flavus* SU-16, NRRL3357 and *A. oryzae* RIB40. (a) Circle 1 and 2: Distribution of the SMGCs and cluster backbones in each chromosome. From outermost to innermost: RIB40, SU-16 and NRRL3357. Circle 3: The 8 chromosomes of SU-16. Circle 4: The Venn diagram compares the shared and unique SMGCs between the genome of SU-16, NRRL3357 and RIB40. (b) Synteny analysis of SMGCs in SU-16, NRRL3357 and RIB40. (c) Microsynteny of the locus harboring the HA biosynthesis gene cluster and its flanking regions in SU-16, NRRL3357 and RIB40. (d) Comparison of the CPA biosynthesis gene cluster of SU-16 with *A. flavus* NRRL3357, *A. oryzae* RIB40 and NBRC4177.

of SMGCs within isolates from the same species. However, it is interesting that synteny analysis of SMGCs also reveals the same pattern as CAZymes. Chromosomal rearrangements caused by inversions, insertions and translocations of the SMGCs are observed on the chromosome 1, 2, 4, 6 and 8, and RIB40 shows more variations than SU-16 compared with NRRL3357 genome (Fig. 5b). Meanwhile, compared with the other two strains, eight SMGCs are unique to SU-16 genome and nine are unique to NRRL3357, whereas only 4 clusters are exclusive to RIB40 (Table S11d).

Twenty-two SMGCs show >50 % of similarity to a known compound cluster (Table S11e). Ten of the SMGCs present in three genomes are completely identical (100 %) to the aflavarin, clavaric acid, naphthopyrone, alternariol, aspirochlorine, leporin B, 6-methylsalicyclic acid and oryzine A clusters. In addition, some clusters of SU-16 with partially conserved synteny to the verified characterized clusters are found (50–90 %; Table S11e). These clusters are likely to be inactive, or the products of which are similar but not identical. Meanwhile, genome data also confirmed the difference in the capability of producing particular SMs. For instance, astellolide A, produced by *A. oryzae* RIB40, displays various biological activities (Shinohara et al., 2016). It is observed that nine of ten astellolide A cluster genes in RIB40 are syntenic in SU-16. Whereas only one gene (*astK*) belonging to that is found in NRRL3357 (Fig. S5a). Aspergillic acid is a typical SM of *A. flavus* and its ferric ion complex is readily expressed on AFPA as an orange reverse (Lebar et al., 2018). SU-16 and NRRL3357 produce aspergillic acid, whereas the RIB40 does not, despite the high conservation and synteny of the corresponding cluster between NRRL3357 and RIB40 (Fig. 1 and S1). This can be explained by the lack of *asaF* gene and the significant difference in genes of *asaR* and *asaB* between RIB40 and NRRL3357 (Fig. S5b). Heptelidic acid (HA), a sesquiterpene antibiotic, was identified in *A. oryzae* RIB40 (Shinohara et al., 2019). The genome analysis reveals conservation for HA cluster in most of the *A. oryzae* strains, but polymorphism in almost all *A. flavus* strains (Gibbons et al., 2012; Kjaerbolling et al., 2020). In the NRRL3357 genome, there is a six-gene cluster comprised by *hepA* and *hepF* from the HA cluster together with four other unrelated genes. Whereas SU-16 possesses a cluster that comprised by the minimum required genes for the biosynthetic pathway of HA (Fig. 5c). Previous studies have shown that RIB40 does not produce cyclopiazonic acid (CPA), due to deletion of *cpaR* and *cpaA* gene (Kato et al., 2011). Although both *A. oryzae* NBRC4177 and *A. flavus* NRRL3357 produces CPA, which can be converted into the less toxic 2-oxocyclopiazonic acid due to the presence of *cpaH* gene in NBRC4177 (Kato et al., 2011). While the CPA cluster is perfectly conserved between SU-16 and NBRC4177 genome (Fig. 5d), thus 2-oxocyclopiazonic acid is likely to be one of the products of SU-16.

All above suggest intraspecies gene transfer or loss may be relatively common. These patterns not only suggest that SMGCs are the species-unique genes for fungi at the interspecific genetic level (Frisvad et al., 2019; Kjaerbolling et al., 2020), which may also indicate the critical differences at intraspecies genetic level, and this difference may depend on the lifestyle or requirements in different ecological niches (Gibbons et al., 2012; Vesth et al., 2018).

3.8. *A. flavus* SU-16 possess an inactive aflatoxin biosynthesis gene cluster

Evidence has shown the polymorphisms in aflatoxin production for different *A. flavus* strains (Frisvad et al., 2019; Okoth et al., 2018). Since none of the aflatoxins were detected in cooked wheat qu fermented by inoculation of SU-16, we suspected that SU-16 may possess an inactive aflatoxin biosynthesis gene cluster. Comparative genomics results showed that the cluster is extremely well conserved with no rearrangements and a high alignment identity for the aflatoxin genes (Fig. 6a, Table S12). Most of the genes, especially six key genes *aflJ*, *aflR*, *ver-1*, *verA*, *otmB* and *avfA* in the cluster of SU-16 are more similar to *A. oryzae* RIB40. Although *A. flavus* AF13, AF70 and AF36 are capable of producing aflatoxin (Gibbons et al., 2012), the clusters in these strains are clustered in one clade with non-aflatoxigenic strain *A. oryzae* RIB40. Genes *hypA*, *ordB* and *aflT* in five *A. flavus* isolates show high identity to NRRL3357.

AflJ is one of the two regulators in aflatoxin pathway that binds to *aflR* to cooperatively regulate genes in the aflatoxin biosynthesis cluster (Cleveland et al., 2009; Yu et al., 2004). Previous study reported that four unique amino acid substitutions in *aflJ* gene in RIB40 induced the inactivation of *aflJ*. As a result, the interaction between *aflJ* and *aflR* was inhibited to protect against aflatoxin biosynthesis (Kiyota et al., 2011; Yu et al., 2004). In this study, SU-16 shares the four amino acids substitutions with RIB40 at the 8th, 22nd, 268th, and 354th nucleotide residues of *aflJ* (Fig. S6), suggesting that *aflJ* is inactivated in SU-16. *aflT*, encoding a membrane-bound protein with homology to antibiotic efflux genes, was reported to have a large deletion of the C-terminal region due to a TC nucleotide insertion in RIB40 (Toyotome et al., 2019; Yu et al., 2004). Large segment deletions are found at both ends of the *aflT* gene in SU-16 genome. Meanwhile, deletions of the C-terminal region occurred to almost all strains except for NRRL3357 and AF13. *Ver-1* and *verA* are involved in the conversion of versicolorin A to demethylsterigmatocystin (Samson et al., 2014; Varga et al., 2011). However, deletions of 40 bp and 108 bp in the coding region occurred to gene *ver-1* and *verA* respectively in SU-16 aflatoxin biosynthesis cluster. *OmtB* is required for the conversion from demethylsterigmatocystin to sterigmatocystin in aflatoxin biosynthesis (Varga et al., 2011; Yu et al., 2004). A 54 bp deletion occurred for this gene in the SU-16 genome compared with other *A. flavus* and *A. oryzae* isolates. Studies (Chang and Ehrlich, 2010; Ehrlich et al., 2004) showed that deletion of the intergenic region and portions of the 5′ ends of the coding sequences of *norB* and *cypA* removed the promoter regions and translational start sites of these two genes is the main cause for *A. flavus* isolates incapable of producing G aflatoxins. Results showed that the *cypA* gene was absent and most sequence of the *norB* gene (>800 bp) was deleted in *A. oryzae* RIB40. However, though partial sequences of these two genes were detected, a much longer deletion of the sequence (1910 bp) than other *A. flavus* isolates (854–1516 bp) (Ehrlich et al., 2004) was found in *norB-cypA* region of SU-16 genome (Fig. 6b, Table S12). Thus, similarity of gene sequences does not imply the functional consistency. Though SU-16 appears to have a seemingly complete aflatoxin biosynthesis cluster, none of the aflatoxins and key intermediate products in the aflatoxin biosynthesis process could be produced by SU-16, because of the inactivation of some key genes caused by partial sequence deletions and mutations.

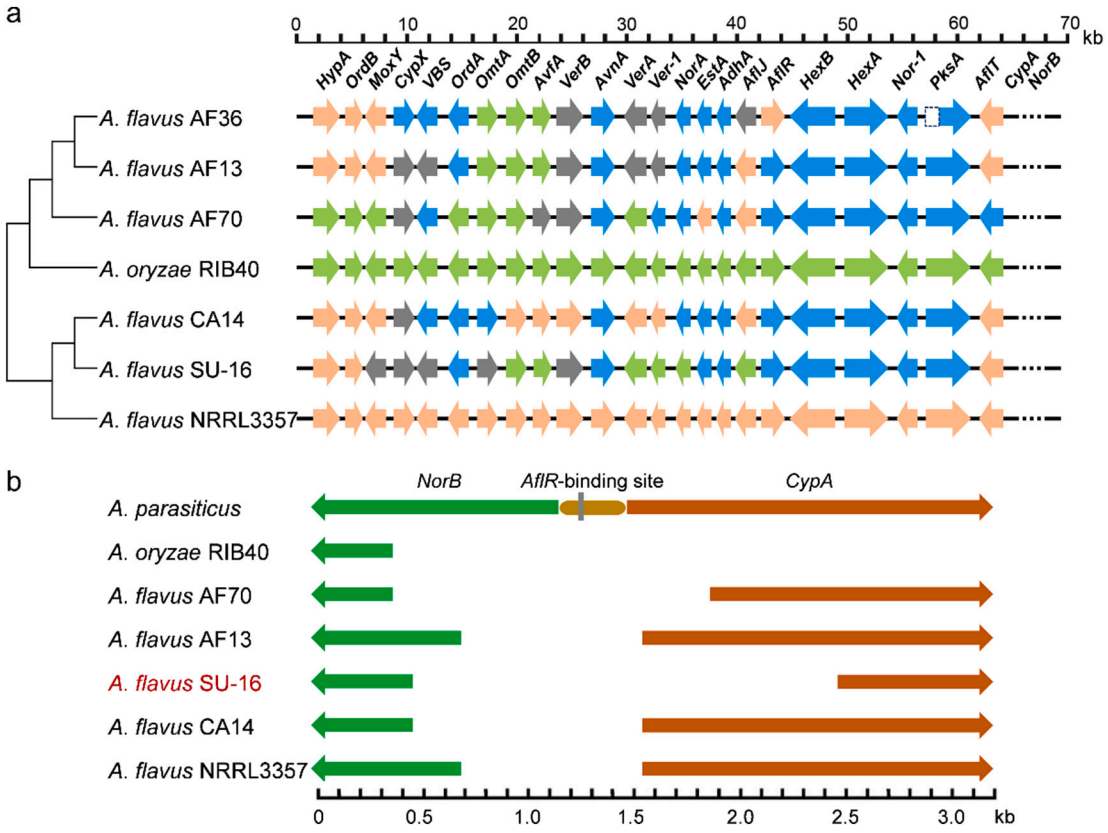

Fig. 6. Comparison of aflatoxin biosynthesis gene clusters. (a) Orange arrows indicate genes showing higher similarity to NRRL3357, while green arrows indicate those more similar to RIB 40. Grey arrows indicate genes that share <99 % identity with those in NRRL3357 and RIB40. Blue arrows indicate ≥99 % identity with NRRL3357 and RIB40. Dotted line blocks indicate the deletion of the genes. Dotted lines corresponding to the deletions of the promoter regions and translational start sites of *cypA* and *norB*. (b) Deletions in *cypA-norB* region in aflatoxin biosynthesis gene clusters of SU-16 and other strains. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Conclusion

For the further guidance of *huangjiu* brewing scientifically and guarantee of the *huangjiu* quality, safety assessment and genomic function of *huangjiu* brewing strain SU-16 based on the combination of molecular, metabolic and whole genome sequencing technology were carried out in this study. SU-16 was identified as a non-aflatoxigenic *A. flavus* strain and a near-complete genome assembly of *A. flavus* SU-16 was obtained for the first time. Comparative genomics showed that SU-16 might have been domesticated for industrial use from a non-toxigenic *A. flavus* strain from the beginning. Meanwhile, SU-16 was found to be the good repository for CAZymes and some bioactive SMs. This work will not only provide a rich genetic resource for the *A. flavus* fundamental and applied research communities, but also help to develop more scientific *huangjiu* fermentation processes, and explore metabolism pathways of desired or harmful components in *huangjiu* to improve its quality. In addition, our findings may also provide insights and supports for the further application of *A. flavus* in the fields of other fermented food (such as cooking wine, soy sauce and fermented soybean), environmental protection and enzyme preparation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was financially supported by the National Natural Science Foundation of China (32072205, 22138004), the Special Project of Science and Technology Plan of Shaoxing (2019B11001).

Data available

Assembly data for the complete genome of *A. flavus* SU-16 can be found at the NCBI database (https://www.ncbi.nlm.nih.gov/assembly/GCA_009856665.1).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijfoodmicro.2022.109859.

References

Aguilar-Pontes, M.V., Brandl, J., McDonnell, E., Strasser, K., Nguyen, T.T.M., Riley, R., Mondo, S., Salamov, A., Nybo, J.L., Vesth, T.C., Grigoriev, I.V., Andersen, M.R., Tsang, A., de Vries, R.P., 2018. The gold-standard genome of *Aspergillus niger* NRRL 3 enables a detailed view of the diversity of sugar catabolism in fungi. Stud. Mycol. 91, 61–78.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., Weber, T., 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 47, W81–W87.

Chang, P.K., Ehrlich, K.C., 2010. What does genetic diversity of *Aspergillus flavus* tell us about *Aspergillus oryzae*? Int. J. Food Microbiol. 138, 189–199.

Chen, S., Xu, Y., Qian, M.C., 2013. Aroma characterization of Chinese rice wine by gas chromatography-olfactometry, chemical quantitative analysis, and aroma reconstitution. J. Agric. Food Chem. 61, 11295–11302.

Chen, A.J., Hubka, V., Frisvad, J.C., Visagie, C.M., Houbraken, J., Meijer, M., Varga, J., Demirel, R., Jurjevic, Z., Kubatova, A., Sklenar, F., Zhou, Y.G., Samson, R.A., 2017. Polyphasic taxonomy of Aspergillus section *Aspergillus* (formerly *Eurotium*), and its occurrence in indoor environments and food. Stud. Mycol. 88, 37–135.

Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y., Xia, R., 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol. Plant 13, 1194–1202.

Cleveland, T.E., Yu, J., Fedorova, N., Bhatnagar, D., Payne, G.A., Nierman, W.C., Bennett, J.W., 2009. Potential of *Aspergillus flavus* genomics for applications in biotechnology. Trends Biotechnol. 27, 151–157.

Duan, S.F., Han, P.J., Wang, Q.M., Liu, W.Q., Shi, J.Y., Li, K., Zhang, X.L., Bai, F.Y., 2018. The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. Nat. Commun. 9, 2690.

Ehrlich, K.C., Chang, P.K., Yu, J., Cotty, P.J., 2004. Aflatoxin biosynthesis cluster gene cypA is required for G aflatoxin formation. Appl. Environ. Microbiol. 70, 6518–6524.

Frisvad, J.C., Hubka, V., Ezekiel, C.N., Hong, S.B., Novakova, A., Chen, A.J., Arzanlou, M., Larsen, T.O., Sklenar, F., Mahakarnchanakul, W., Samson, R.A., Houbraken, J., 2019. Taxonomy of *Aspergillus* section Flavi and their production of aflatoxins, ochratoxins and other mycotoxins. Stud. Mycol. 93, 1–63.

Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., Teiling, C., Steffy, B., Taylor, M., Schwartz, A., Richardson, T., White, C., Baele, G., Maere, S., Verstrepen, K.J., 2016. Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. Cell 166 (1397–1410), e1316.

Gibbons, J.G., Salichos, L., Slot, J.C., Rinker, D.C., McGary, K.L., King, J.G., Klich, M.A., Tabb, D.L., McDonald, W.H., Rokas, A., 2012. The evolutionary imprint of domestication on genome variation and function of the filamentous fungus *Aspergillus oryzae*. Curr. Biol. 22, 1403–1409.

Ha, G., Roth, A., Lai, D., Bashashati, A., Ding, J., Goya, R., Giuliany, R., Rosner, J., Oloumi, A., Shumansky, K., Chin, S.F., Turashvili, G., Hirst, M., Caldas, C., Marra, M. A., Aparicio, S., Shah, S.P., 2012. Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. Genome Res. 22, 1995–2007.

Ichishima, E., 2016. Development of enzyme technology for *Aspergillus oryzae*, *A. sojae*, and *A. luchuensis*, the national microorganisms of Japan. Biosci. Biotechnol. Biochem. 80, 1681–1692.

Jenjaroenpun, P., Wongsurawat, T., Pereira, R., Patumcharoenpol, P., Ussery, D.W., Nielsen, J., Nookaew, I., 2018. Complete genomic and transcriptional landscape analysis using third-generation sequencing: a case study of *Saccharomyces cerevisiae* CEN.PK113-7D. Nucleic Acids Res. 46, e38.

Kato, N., Tokuoka, M., Shinohara, Y., Kawatani, M., Uramoto, M., Seshime, Y., Fujii, I., Kitamoto, K., Takahashi, T., Takahashi, S., Koyama, Y., Osada, H., 2011. Genetic safeguard against mycotoxin cyclopiazonic acid production in *Aspergillus oryzae*. Chembiochem 12, 1376–1382.

Kiyota, T., Hamada, R., Sakamoto, K., Iwashita, K., Yamada, O., Mikami, S., 2011. Aflatoxin non-productivity of *Aspergillus oryzae* caused by loss of function in the aflJ gene product. J. Biosci. Bioeng. 111, 512–517.

Kjaerbolling, I., Vesth, T., Frisvad, J.C., Nybo, J.L., Theobald, S., Kildgaard, S., Petersen, T.I., Kuo, A., Sato, A., Lyhne, E.K., Kogle, M.E., Wiebenga, A., Kun, R.S., Lubbers, R.J.M., Makela, M.R., Barry, K., Chovatia, M., Clum, A., Daum, C., Haridas, S., He, G., LaButti, K., Lipzen, A., Mondo, S., Pangilinan, J., Riley, R., Salamov, A., Simmons, B.A., Magnuson, J.K., Henrissat, B., Mortensen, U.H., Larsen, T.O., de Vries, R.P., Grigoriev, I.V., Machida, M., Baker, S.E., Andersen, M.R., 2020. A comparative genomics study of 23 *Aspergillus* species from section Flavi. Nat. Commun. 11, 1106.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27, 722–736.

Lebar, M.D., Cary, J.W., Majumdar, R., Carter-Wientjes, C.H., Mack, B.M., Wei, Q., Uka, V., De Saeger, S., Diana Di Mavungu, J., 2018. Identification and functional analysis of the aspergillic acid gene cluster in *Aspergillus flavus*. Fungal Genet. Biol. 116, 14–23.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., <collab>Genome Project Data Processing, S.</collab>, 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Liu, S., Chen, Q., Zou, H., Yu, Y., Zhou, Z., Mao, J., Zhang, S., 2019. A metagenomic analysis of the relationship between microorganisms and flavor development in Shaoxing mechanized huangjiu fermentation mashes. Int. J. Food Microbiol. 303, 9–18.

Liu, S., Hu, J., Xu, Y., Xue, J., Zhou, J., Han, X., Ji, Z., Mao, J., 2020. Combined use of single molecule real-time DNA sequencing technology and culture-dependent methods to analyze the functional microorganisms in inoculated raw wheat Qu. Food Res. Int. 132, 109062.

Luo, B., Sun, H., Zhang, Y., Gu, Y., Yan, W., Zhang, R., Ni, Y., 2019. Habitat-specificity and diversity of culturable cold-adapted yeasts of a cold-based glacier in the Tianshan Mountains, northwestern China. Appl. Microbiol. Biotechnol. 103, 2311–2327.

Machida, M., Asai, K., Sano, M., Tanaka, T., Kumagai, T., Terai, G., Kusumoto, K., Arima, T., Akita, O., Kashiwagi, Y., Abe, K., Gomi, K., Horiuchi, H., Kitamoto, K., Kobayashi, T., Takeuchi, M., Denning, D.W., Galagan, J.E., Nierman, W.C., Yu, J., Archer, D.B., Bennett, J.W., Bhatnagar, D., Cleveland, T.E., Fedorova, N.D., Gotoh, O., Horikawa, H., Hosoyama, A., Ichinomiya, M., Igarashi, R., Iwashita, K., Juvvadi, P.R., Kato, M., Kato, Y., Kin, T., Kokubun, A., Maeda, H., Maeyama, N., Maruyama, J., Nagasaki, H., Nakajima, T., Oda, K., Okada, K., Paulsen, I., Sakamoto, K., Sawano, T., Takahashi, M., Takase, K., Terabayashi, Y., Wortman, J. R., Yamada, O., Yamagata, Y., Anazawa, H., Hata, Y., Koide, Y., Komori, T., Koyama, Y., Minetoki, T., Suharnan, S., Tanaka, A., Isono, K., Kuhara, S., Ogasawara, N., Kikuchi, H., 2005. Genome sequencing and analysis of *Aspergillus oryzae*. Nature 438, 1157–1161.

McGovern, P.E., Zhang, J., Tang, J., Zhang, Z., Hall, G.R., Moreau, R.A., Nunez, A., Butrym, E.D., Richards, M.P., Wang, C.S., Cheng, G., Zhao, Z., Wang, C., 2004. Fermented beverages of pre- and proto-historic China. Proc. Natl. Acad. Sci. U. S. A. 101, 17593–17598.

Okoth, S., De Boevre, M., Vidal, A., Diana Di Mavungu, J., Landschoot, S., Kyallo, M., Njuguna, J., Harvey, J., De Saeger, S., 2018. Genetic and toxigenic variability within *Aspergillus flavus* population isolated from maize in two diverse environments in Kenya. Front. Microbiol. 9, 57.

Payne, A., Holmes, N., Clarke, T., Munro, R., Debebe, B.J., Loose, M., 2021. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. Nat. Biotechnol. 39, 442–450.

Samson, R.A., Visagie, C.M., Houbraken, J., Hong, S.B., Hubka, V., Klaassen, C.H., Perrone, G., Seifert, K.A., Susca, A., Tanney, J.B., Varga, J., Kocsube, S., Szigeti, G., Yaguchi, T., Frisvad, J.C., 2014. Phylogeny, identification and nomenclature of the genus *Aspergillus*. Stud. Mycol. 78, 141–173.

Senol Cali, D., Kim, J.S., Ghose, S., Alkan, C., Mutlu, O., 2019. Nanopore sequencing technology and tools for genome assembly: computational analysis of the current state, bottlenecks and future directions. Brief. Bioinform. 20, 1542–1559.

Shinohara, Y., Takahashi, S., Osada, H., Koyama, Y., 2016. Identification of a novel sesquiterpene biosynthetic machinery involved in astellolide biosynthesis. Sci. Rep. 6, 32865.

Shinohara, Y., Nishimura, I., Koyama, Y., 2019. Identification of a gene cluster for biosynthesis of the sesquiterpene antibiotic, heptelidic acid, in *Aspergillus oryzae*. Biosci. Biotechnol. Biochem. 83, 1506–1513.

Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.

Sun, H., Liu, S., Mao, J., Yu, Z., Lin, Z., Mao, J., 2020. New insights into the impacts of huangjiu compontents on intoxication. Food Chem. 317, 126420.

Toyotome, T., Hamada, S., Yamaguchi, S., Takahashi, H., Kondoh, D., Takino, M., Kanesaki, Y., Kamei, K., 2019. Comparative genome analysis of *Aspergillus flavus* clinically isolated in Japan. DNA Res. 26, 95–103.

Urasaki, N., Takagi, H., Natsume, S., Uemura, A., Taniai, N., Miyagi, N., Fukushima, M., Suzuki, S., Tarora, K., Tamaki, M., Sakamoto, M., Terauchi, R., Matsumura, H., 2017. Draft genome sequence of bitter gourd (*Momordica charantia*), a vegetable and medicinal plant in tropical and subtropical regions. DNA Res. 24, 51–58.

Varga, J., Frisvad, J.C., Samson, R.A., 2011. Two new aflatoxin producing species, and an overview of *Aspergillus* section Flavi. Stud. Mycol. 69, 57–80.

Vesth, T.C., Nybo, J.L., Theobald, S., Frisvad, J.C., Larsen, T.O., Nielsen, K.F., Hoof, J.B., Brandl, J., Salamov, A., Riley, R., Gladden, J.M., Phatale, P., Nielsen, M.T., Lyhne, E. K., Kogle, M.E., Strasser, K., McDonnell, E., Barry, K., Clum, A., Chen, C., LaButti, K., Haridas, S., Nolan, M., Sandor, L., Kuo, A., Lipzen, A., Hainaut, M., Drula, E., Tsang, A., Magnuson, J.K., Henrissat, B., Wiebenga, A., Simmons, B.A., Makela, M. R., de Vries, R.P., Grigoriev, I.V., Mortensen, U.H., Baker, S.E., Andersen, M.R., 2018. Investigation of inter- and intraspecies variation through genome sequencing of *Aspergillus* section Nigri. Nat. Genet. 50, 1688–1695.

Wang, Z.M., Wang, C.T., Shen, C.H., Wang, S.T., Mao, J.Q., Li, Z., Ganzle, M., Mao, J., 2020. Microbiota stratification and succession of amylase-producing *Bacillus* in traditional Chinese Jiuqu (fermentation starters). J. Sci. Food Agric. 100, 3544–3553.

Yu, J., Chang, P.K., Ehrlich, K.C., Cary, J.W., Bhatnagar, D., Cleveland, T.E., Payne, G.A., Linz, J.E., Woloshuk, C.P., Bennett, J.W., 2004. Clustered pathway genes in aflatoxin biosynthesis. Appl. Environ. Microbiol. 70, 1253–1262.

Zhang, B., Guan, Z.B., Cao, Y., Xie, G.F., Lu, J., 2012. Secretome of *Aspergillus oryzae* in Shaoxing rice wine koji. Int. J. Food Microbiol. 155, 113–119.

Zhang, W., Li, Y., Chen, Y., Xu, S., Du, G., Shi, H., Zhou, J., Chen, J., 2018. Complete genome sequence and analysis of the industrial *Saccharomyces cerevisiae* strain N85 used in Chinese rice wine production. DNA Res. 25 (3), 297–306.

Zhao, G., Yao, Y., Wang, C., Hou, L., Cao, X., 2013. Comparative genomic analysis of *Aspergillus oryzae* strains 3.042 and RIB40 for soy sauce fermentation. Int. J. Food Microbiol. 164, 148–154.