BMC Genomics

RESEARCH

Open Access

Genome diversity, population structure and MALDI-TOF MS profiling of *Aspergillus oryzae/flavus* strains from fermentation and wild environments

Dong-Hyun Kim[1,2], Dong-Chan Kim[3], Donggun Seo[3], Ki-Tae Kim[4], Sang-Han Lee[2*] and Seung-Beom Hong[1*]

Abstract

Various strains of *Aspergillus oryzae*, regarded as a domesticated variant of aflatoxigenic *Aspergillus flavus*, are utilized in the soybean fermentation industry of Korea. This study compared *A. oryzae/flavus* strains isolated from various environments in Korea including industrial settings, Meju (brick of dried fermented soybeans), and wild conditions with globally reported strains using genomic analysis to determine their taxonomic positions and risk of mycotoxicity. Using population genomics, five distinct groups (A to E) were identified, with all aflatoxigenic Korean strains in Group C and non-aflatoxigenic Korean strains in Groups A, B, and E. Korean strains from Meju and wild conditions are distributed across Groups A and B, and most of the Korean industrial strains form a sub-cluster with Japanese industrial strains in Group A. Comparing secondary metabolite gene cluster mutation pattern, three gene clusters (Aflatoxin, Cyclopiazonic acid and Ditryptophenaline) were revealed as group specific ones. In aflatoxin and cyclopiazonic acid clusters, most of the Group C strains had intact regions compared to strains in other groups. Since most of the Group C strains produce aflatoxin and have intact Aflatoxin and Cyclopiazonic acid gene clusters, we considered that this group represent *A. flavus*. Profiling using MALDI-TOF MS analysis also distinguished Group C from Groups A, B and E by specific three proteomic peaks. Among the three peaks, those around 12,700 to 12,900 m/z (Da) are expected to correspond to AflF (nor B), an enzyme involved in Aflatoxin metabolism. These results showed taxonomic positions of Korean strains of *A. oryzae/flavus* from various environments and also showed possibility to differentiate between *A. oryzae* and *A. flavus* with genome and MALDI-TOF MS analysis.

Keywords Genomics, *Aspergillus oryzae/flavus* complex, Structure analysis, Aflatoxin, MALDI-TOF

*Correspondence:
Sang-Han Lee
sang@knu.ac.kr
Seung-Beom Hong
funguy@korea.kr
[1] Korean Agricultural Culture Collection, Agricultural Microbiology Div., National Institute of Agricultural Sciences, RDA, Wanju 55365, Republic of Korea
[2] Department of Food Science and Biotechnology, Kyungpook National University, Daegu 41566, Republic of Korea
[3] Division of Medical Device R&D, NQ-Lab, Inc, Yongin 16827, Republic of Korea
[4] Department of Agricultural Life Science, Sunchon National University, Suncheon 57922, Republic of Korea

Introduction

The filamentous fungal species *Aspergillus oryzae* is classified as a GRAS (Generally Recognized as Safe) microorganism, and plays an indispensable role in the fermentation process of various Asian foods and beverages, including doenjang, soy sauce, and sake [1–5]. In contrast, some species within the *Aspergillus* genus, such as *Aspergillus fumigatus*, are known to cause opportunistic infections in immunocompromised individuals, a condition referred to as aspergillosis [6, 7]. More recently, cases of opportunistic infection involving *Aspergillus flavus* have also been reported [8]. However, in the context of food fermentation, the primary safety concern lies in their potential to produce harmful mycotoxins such as aflatoxin and cyclopiazonic acid.

A. oryzae is considered as domesticated variant of *A. flavus*, which produce these mycotoxins. The two species are members of the section *Flavi* and are so closely related that they cannot be reliably distinguished using secondary marker genes such as BenA or CaM [9]. However, they are known to exhibit high genetic diversity, particularly in sub-telomeric regions where several mycotoxin biosynthetic gene clusters are located [3, 10].

A previous study demonstrated that the *A. oryzae* strain RIB40 is incapable of producing toxic secondary metabolites such as aflatoxin and cyclopiazonic acid, but rather produces isomeric metabolites such as aflatrem, miyakamides and ditryptophenaline compared to *A. flavus* strain NRRL 3357 [11]. Despite their significant biochemical differences, these two species exhibit a close genomic relationship. Moreover, different strains of *A. flavus* exhibit variability in aflatoxin production, influenced by temperature and humidity [12]. Due to these characteristics, some atoxigenic strains with polymorphisms in the aflatoxin biosynthetic pathway were misclassified as *A. flavus* instead of *A. oryzae* [13]. This reflects the low reliability of *A. oryzae* as a distinct species. Therefore, comparing whole genomes is required to differentiate these closely related species precisely.

A previous whole genomic analysis of *A. oryzae* classified industrial strains used in the Japanese fermentation industry into eight clades [5]. Further efforts aimed at distinguishing *A. oryzae* from *A. flavus* through comparative genomics, with a particular focus on aflatoxin gene cluster and the composition of Carbohydrate-Active enzymes (CAZymes), have not conclusively differentiated the two species [14].

While extensive genomic studies have been conducted on industrial strains, relatively little attention has been given to non-aflatoxigenic strains derived from traditional meju and wild conditions, which are expected to have fewer events of genetic improvement. In a study on *Penicillium roqueforti*, traditional strains used in Roquefort cheese production were found to be intermediate between wild and industrial strains. Thus, strains are also presumed to be intermediate between industrial *A. oryzae* and aflatoxigenic *A. flavus* [15]. Therefore, understanding their population structure and aflatoxin production risk at the genomic level is necessary to monitor their distribution.

In recent microbial identification and classification studies, the MALDI-TOF MS (Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry) method has emerged as a powerful tool. For example, *Weissella confusa* and *Weissella cibaria* are genetically close and had different roles, with *W. confusa* contributing to health benefits while *W. cibaria* having a potential pathogenic role. This relationship is similar to that shared between *A. oryzae* and *A. flavus*. To distinguish these closely related species, recent studies have developed markers using machine learning to analyze MALDI-TOF MS patterns [16]. There was also attempt to use MALDI-TOF MS to distinguish *A. oryzae/flavus* complex, but the study concluded that this method is not suitable for accurate differentiation due to the high overlapping protein profiles of the two species [17].

Our study aimed to elucidate the distribution of diverse Korean *A. oryzae/flavus* strains from various environments by comparing their genomes with those of strains from previous studies originated from diverse geographical regions, such as Japan, the United States, and China.

Furthermore, variations of mutations within gene clusters associated with mycotoxins and secondary metabolites, related to safety in the food industry, were analyzed. In this study, a machine learning model was applied to the MALDI-TOF MS data, enabling more accurate proteomics pattern profiling, successfully identifying distinct patterns that differentiate the clustered groups. Based on the findings, we propose a group distribution of *A. oryzae* and *A. flavus* strains categorized by their genomic and proteomic differences. We particularly focused on their originated source, variations in mycotoxin production, and mutations in biosynthetic gene clusters.

Materials & methods

Collection of genomes of Korean and global *A. oryzae/flavus* strains

Korean strains from 4 diverse sources, Korean Industrial non-aflatoxigenic strains (KRI, 5 strains), Korean non-aflatoxigenic strains from meju (KRM, 12 strains), Korean non-aflatoxigenic strains from wild conditions (KRWO, 7 strains), Korean aflatoxigenic strains from wild conditions (KRWF, 15 strains) were collected (Table 1). All 39 Korean strains were previously designated as *A. oryzae* or *A. flavus* based on their aflatoxin production and the presence or absence of key aflatoxin biosynthetic genes

Table 1 Korean *A. oryzae/flavus* strains from various condition

Strain name		KACC no	Discription	Region	Aflatoxin	*norB/cypA*[a]	*omtA*[b]
MWA1	KRM	K46457	Meju, Goisan, omtA A type (Hong et al. (2013)	Korea	NO	Type I	A
MWA2		K46810	Meju, Sunchang, omtA A type (Hong et al. (2013)	Korea	NO	X	A
MWA3		K46811	Meju, Jeju, omtA A type (Hong et al. (2013)	Korea	NO	X	A
MWB1		K46455	Meju, Gyeongsan, omtA B type (Hong et al. (2013)	Korea	NO	Type I	B
MWB2		K46470	Meju, Yongin, omtA B type (Hong et al. (2013)	Korea	NO	Type I	B
MWB3		K46471	Meju, Incheon, omtA B type (Hong et al. (2013)	Korea	NO	X	B
MWC1		K46456	Meju, Haenam, omtA C type (Hong et al. (2013)	Korea	NO	Type I	C
MWC2		K46474	Meju, Damyang omtA C type (Hong et al. (2013)	Korea	NO	Type I	C
MWC3		K46469	Meju, Jeju, omtA C type (Hong et al. (2013)	Korea	NO	Type I	C
MWX1		K46465	Meju, Buan, no omtA(Hong et al. (2013)	Korea	NO	Type I	X
MWX2		K46466	Meju, Gongju, no omtA(Hong et al. (2013)	Korea	NO	Type I	X
K93210		K93210	Meju, patent strain	Korea	NO	Type I	C
KRI1	KRI	K46640	Korean industrial strain	Korea	NO	Type I	C
KRI2		K46641	Korean industrial strain	Korea	NO	Type I	B
KRI3		K46642	Korean industrial strain	Korea	NO	Type I	C
KRI4		K47488	Korean industrial strain	Korea	NO	Type I	C
KRI5		K47843	Korean industrial strain	Korea	NO	Type I	C
Aor- 06	KRWO	K46923	Rice straw, Yangyang	Korea	NO	Type I	ND
Aor- 17		K46924	Rice straw, Gongju	Korea	NO	Type I	ND
Aor- 38		K46909	Soybean, Incheon	Korea	NO	X	ND
SD045		K48145	Peanut, Jinyang	Korea	NO	X	ND
SL041		K46918	Soil, Suwon	Korea	NO	Type I	ND
SL046		K46920	Soybean farm soil, Namhae	Korea	NO	Type I	ND
SL055		K46922	Greentea farm soil, Jeju	Korea	NO	X	ND
Aor- 34	KRWF	K46927	Rice straw, Yangpyeong	Korea	YES (B)	Type II	C
AR018		K46892	Indoor air, Seoul	Korea	YES (B)	Type II	C
AR028		K46894	Outdoor air, Suwon	Korea	YES (B/G)	Type III	Out group
SD016		K46897	Corn, Hongcheon	Korea	YES (B/G)	Type III	Out group
SD022		K46898	Corn, Hongcheon	Korea	YES (B)	Type II	C
SD024		K46899	Corn, Hongcheon	Korea	YES (B)	Type II	C
SD039		K46902	Peanut, Seosan	Korea	YES (B/G)	Type II	C
SD059		K46903	Peanut, Jangseong	Korea	YES (B)	Type II	C
SD061		K46904	Peanut, Danyang	Korea	YES (B/G)	Type II	C
SL001		K46928	Soil, Gwangju	Korea	YES (B)	Type II	C
SL005		K46929	Soil, Gwangju	Korea	YES (B)	Type II	C
SL008		K46912	Soil, Deagu	Korea	YES (B)	Type II	C
SL015		K46913	Soil, Gwangju	Korea	YES (B/G)	Type III	Out group
SL034		K46916	Soil, Suwon	Korea	YES (B/G)	Type II	C
SL044		K46919	Garlic farm soil, Namhae	Korea	YES (B/G)	Type II	C

KRM are non-aflatoxigenic strains from traditional meju. KRI are non- aflatoxigenic strains used in Korean fermentation Industry. KRWO are non- aflatoxigenic strains from Korean wild condition. KRWF are aflatoxigenic strains from Korean wild condition. *norB/cypA* and *omtA* are marker genes used to distinguish *A. oryzae/flavus,* and part of the aflatoxin biosynthesis gene cluster

[a] Type I, II, III: Different product sizes determined by PCR amplification, with I being the smallest and III the largest. "X"indicates no band detected

[b] A, B, C: Phylogenetic groups based on omtA gene analysis

such as omtA and norB/cypA [9]. Strains were deposited in Korean Agricultural Culture Collection (KACC), and stored at − 80 °C. For genomic DNA extraction, 10 µL of spore suspension from each stock was inoculated at three points onto a Malt Extract Agar (MEA) plate and incubated at 25 °C for seven days. Subculturing under the same condition was performed once, prior to DNA extraction.

Genomic DNA (gDNA) of Korean strains was extracted from mycelia grown on MEA plate, using the DNeasy Plant Mini kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. The quality and concentration of the extracted DNA were assessed using an Analytik Jena ScanDrop 2 spectrophotometer by measuring absorbance at 260 nm. PCR-free 150-bp paired-end libraries were constructed and sequenced by Macrogen Inc (Seoul) on NovaSeq 6000 platform (Illumina, Details). Raw reads underwent quality control using Trimmomatic (v0.38) to remove adapter sequences and low-quality reads [18] and FastQC was performed to check the read quality (https://www.bioinformatics.babraham.ac.uk/projects/fastqc).

For the comparative genomic analysis, sequence reads from strains originated from 4 different sources were collected. The sources include Japanese industrial strains (JPI, 17 strains), NCBI non-aflatoxigenic strains (NO, 13 strains), NCBI aflatoxigenic strains (NF, 9 strains), and Afla-guard non-aflatoxigenic strains (AG, 2 strains). All sequence reads were obtained as raw files from the National Center for Biotechnology Information (NCBI) Sequence Reads Archive (Table 2). JPI strains, previously categorized into eight clades (Clade A to H), were chosen with two representatives from each clades and an extra strain that did not fit into any clade (TK- 24) were selected [5]. The distinction between the NO and NF strains was not based on the existing nomenclature. Instead, it was determined by previous research studies that examined the aflatoxin production capabilities of these strains [3–5, 13, 19–25].

Single Nucleotide Polymorphisms (SNPs) calling and population genetic analysis

For SNP calling, filtered DNA sequencing reads of the 80 individuals were aligned to the reference genome, NRRL 3357 (GenBank ac no.: GCA_009017415.1), using BWA software v0.7.17 MEM module [29]. SAM alignment files were sorted and converted to BAM files with SAMtools v1.3 [30]. Picard was used to delete duplicate reads and create index of reference genome. Next, various programs in GATK packages (HaplotypeCaller for variant calling, CombineGVCFs for merging GVCF files, GenotypeGVCFs for converting GVCF into VCF file, SelectVariants, and VariantFiltration for selecting and filtering SNPs) were used [31]. The filtering parameters were as follows: QD < 2.0, FS > 80.0, MQ < 20.0, SOR > 3.0, MQRankSum < − 12.5, Read- PosRankSum < − 8.0, and QUAL < 40.0. VCF files were converted into fasta files using vcf2fasta in order to construct a phylogenetic tree (https://github.com/santiagosnchez/vcf2fasta). The vcf files of KACC strains were deposited in the European Molecular Biology Laboratory European Variation Archive (EMBL EVA, https://www.ebi.ac.uk/eva/) under accession number PRJEB79400 (https://www.ebi.ac.uk/ena/browser/view/PRJEB79400).

The phylogenetic tree was drawn using the whole-genome-based phylogenetic reconstruction program, SANS serif [32]. PCA with population-scale SNPs was performed using PLINK [33] and GCTA [34]. A Bayesian population structure assessment was performed using ADMIXTURE v1.3.0 [35] with block relaxation algorithm and maximum likelihood estimation. Pre-defined genetic clusters were increased from K = 2 to K = 15 (assumed number of ancestral populations). To identify the most optimal K number, we calculated the cross-validation (CV) error values, where a lower CV value indicates a better estimation. This calculation was performed systematically for K values ranging from 1 to 15 [36]. Using this approach, K = 11 was identified as the most likely number of populations. Prior to running ADMIXTURE, SNPs in high linkage disequilibrium (LD) were removed. Specifically, PLINK was used to prune SNPs employing a windowed approach, with an R2 threshold of 0.8.

Genotypic analysis of secondary metabolites gene cluster and design of diagnostic markers

Significant phenotypic differences influenced by *A. oryzae/flavus* domestication are prominently reflected in the mutations of secondary metabolite biosynthetic gene clusters, with substantial variations in the secretion levels of numerous toxic metabolites. The presence and locations of the aflatoxin gene cluster [37–39], the CPA gene cluster [40], and the ditryptophenaline gene cluster [41] in the genome of *Aspergillus flavus* NRRL 3357 have been elucidated in previous studies. SNPs within these clusters, compared to the reference genome NRRL 3357, were visualized using the Integrated Genomics Viewer (IGV) to observe the mutation patterns. This visualization allows the identification of specific mutations or deletions occurring within these gene clusters of each strain.

MALDI TOF metabolites pattern analysis

Mass spectrometry data from MALDI-TOF/MS on KACC strains containing Groups A, B, C and E were analyzed to compare the similarity patterns among groups classified based on genomic analysis results. The group specific mass-spectral peaks patterns were explored by comparing presence or absence of mass peak at certain m/z values, which correlate with the protein mass.

Each strain underwent individual mass analysis using MicroIDSys (ASTA Inc.). For each strain, a single colony grown on MEA media at 25 °C for seven days, was used for each analysis. Using the smear method, one colony was smeared onto an individual spot on the μID plate (ASTA Inc.), which contains 20 spots in total

Table 2 Global *A. oryzae/flavus* strains from NCBI database

Strain name		NCBI No	Discription	Region	Aflatoxin
TK- 1	JPI	DRX154124	Sake, Clade A in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019) [5]
TK- 2		DRX154126	Sake, Clade A in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019)
TK- 20		DRX154161	Sake, Clade B in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019)
TK- 26		DRX154172	Sake, Clade B in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019)
TK- 24		DRX154168	Outer Clade in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019)
TK- 5		DRX154132	Sake, Clade C in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019)
TK- 7		DRX154136	Miso, Clade C in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019)
TK- 59		DRX154211	Soyou, Clade D in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019)
TK- 60		DRX154212	Soyou, Clade D in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019)
TK- 9		DRX154140	Soyou, Clade E in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019)
TK- 13		DRX154148	Soyou, Clade E in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019)
TK- 4		DRX154130	Sake, Clade F in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019)
TK- 11		DRX154144	Miso, Clade F in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019)
TK- 10		DRX154142	Miso, Clade G in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019)
TK- 15		DRX154151	Miso, Clade G in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019)
TK- 12		DRX154146	Soyou, Clade H in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019)
TK- 14		DRX154149	Soyou, Clade H in Watarai et al. (2019)	Japan	NO, Watarai et al. (2019)
14,160	NO	SRX013842	Xinyang City, Chacón-Vargas et al. (2021),	China	NO, Chacón-Vargas et al. (2021) [20]
BP2 - 1		SRX6074494	Korean strain in Watarai et al. (2019)	Korea	NO, Watarai et al. (2019)
RIB537		SRX147127	Gibbons et al. (2012)	America	NO, Gibbons et al. (2012) [3]
RIB949		SRX147131	Gibbons et al. (2012)	America	NO, Gibbons et al. (2012)
M2040[a]		SRX4479827	Biocontroller in Alshannaq et al. (2018)	Korea	NO, Alshannaq et al. (2018) [4]
NRRL35739		SRX5329434	Biocontroller in Pennerman et-al. (2019)	America	NO, Pennerman et-al. (2019) [21]
SU- 16		SRX8635871	Suzhou winery	China	NO, Sun, Liu et al. (2022) [13]
WRRL1519		SRX3067799	almond nuts	America	NO, Yin et-al. (2018) [26]
2017 Washington T2		SRX5358295	Biocontrol strain, SRR8556566	America	NO, Weaver et al. (2019) [27]
A1		SRX8062526	Louisiana State University	America	NO, Abbas et al. (2011) [28]
AF36		SRX8062528	Biocontrol strain	America	NO, Abbas et al. (2011)
NRRL30797 (K49)		SRX8062530	Biocontrol strain	America	NO, Fountain et al. (2020) [19]
VCG1		SRX8062533	Louisiana State University	America	NO, Fountain et al. (2020)
2017 Washington T5	NF	SRX5358298	USDA-ARS	America	YES, Weaver et al. (2019)
A9		SRX8062527	Louisiana State University	America	YES, Fountain et al. (2020)
CA14		SRX6432310	Pistachio, Hua et al.(2012}	America	YES, Hua et al. (2012) [25]
E1445		SRX8621731	Ethiopia peanut	Ethiopia	YES, Arias et al. (2020) [24]
E1404		SRX8617991	Ethiopia peanut	Ethiopia	YES, Arias et al. (2020)
E1402		SRX8617924	Ethiopia peanut	Ethiopia	YES, Arias et al. (2020)
K54 A		SRX8062531	Louisiana State University	America	YES, Abbas et al. (2011)
MRI19		SRX10945235	Tiger nuts, Schamann et al. (2022)	Spain	YES, Schamann et al. (2022) [23]
Tox4		SRX8062532	Louisiana State University	America	YES, Fountain et al. (2020)
Afla-Guard	AG	SRX4479828	University of Georgia	America	NO, Fountain et al. (2020)
Yazoo S2		SRX5356519	USDA-ARS	America	NO, Weaver et al. (2019)

JPI are atoxigenic strains used in Japanese industry. Two strains for each clade, 8 clades strains are gathered. TK- 24 is not included in any clade. NO are global strains known as atoxigenic. NF are global strains known as toxigenic. AG are strains known as Afla-Guard

[a] Non-aflatoxigenic strains from traditional meju; KRM strain

using a toothpick or a picking tool, with one strain per plate. The spot was then dried, and 1.5 µL of CHCA matrix was applied. The raw spectrum data generated by MicroIDSys ranges from 2,000 to 20,000 daltons (m/z), with an identification cutoff set at ≥ 140 as per the manufacturer's recommendation.

The raw data from MicroIDSys undergoes normalization to adjust peak height relative to m/z using proprietary software from ASTA Inc. and NQ-Lab. Co.,Ltd. Subsequently, the data is processed with a binning parameter set at 5. An averaging procedure calculates the mean peak heights for each m/z range within the binning. The processed data is then interpreted using a Python-based code, resulting in a combined dataset.

The prepared data is analyzed using a machine learning model developed by NQ-Lab., Co., Ltd. Logistic Regression, is the chosen machine learning algorithm tailored with hyperparameters optimized for MALDI-TOF/MS characteristics. The analysis accounts for various factors influencing MALDI-TOF/MS results, such as temporal, physical, and biochemical attributes. To ensure a comprehensive analysis, multiple statistical tests are conducted. The ANOVA test is utilized to examine mean differences between groups, considering the continuity of m/z values. Additionally, the Chi-squared (χ^2) test is employed to assess the independence of categories, assuming that the two groups are distinct entities.

Results

Phylogenetic relationship of *A. oryzae/flavus* strains

This comprehensive analysis classified the non-aflatoxigenic and aflatoxigenic *A. oryzae/flavus* strains into five distinct groups, designated as Groups A to E (Fig. 1).

Group A predominantly comprises non-aflatoxigenic strains, with a significant presence of KRI and JPI strains. Notably, most KRI strains were found to cluster within Clade C. Additionally, among the KRM strains, MWC2, MWC3, and M2040 also aligned with Clade C, while K93210 aligned with Clade D. Other KRM and KRWO strains were also classified under Group A, although they did not align with any of the specific clades. Group B primarily includes non-industrial strains such as KRM, KRWO, and NO, apart from a singular KRI strain, KRI2, indicating a diverse genetic background within non-aflatoxigenic populations.

Group C encapsulates all aflatoxigenic strains, including KRWF and NF. Intriguingly, this group also contains five non-aflatoxigenic strains from the NO category (2017 Washington-T2, A1, AF36, NRRL30797, and VCG1), highlighting the complex genetic landscape that does not strictly correlate with aflatoxin production. Group D is characterized by its genetic divergence from Groups A, B, and C, comprising mainly JPI strains. with the exception of one NO strain, 14,160. Group E, distinguished as the most genetically distinct cluster, includes biocontrol strains Afla-guard, and Yazoo S2, alongside two unique KRM strains (MWA1 and MWA3).

Genome and population structure analysis of *A. oryzae/flavus*

Principal Component Analysis (PCA) of the *A. oryzae/flavus* population structure revealed that Group D closely aligns with Groups A and B, demonstrating their closer genetic relationship, while Group E exhibits distinct genetic separateness from all other groups (Fig. 2A). This distinct positioning of Group E highlights its unique genetic makeup. Admixture analysis further elucidates the genetic diversity and interrelation among the groups. At K = 11, part of the Group A strains shared population with Group B and D, while part of the Group C strains shared small population with Group E. Phylogenetic network analysis further supports the presence of five main populations and individual population assignment into these populations (Fig. 2B).

Secondary metabolite gene cluster analysis

Among the various secondary metabolite gene clusters, three (aflatoxin, CPA and ditryptophenaline) gene clusters showed group-specific characteristics. In genome of *A. flavus* NRRL3357, the aflatoxin gene cluster region is located on chromosome 3, 4,940,000 to 5,010,000 (Fig. 3). CPA gene cluster region is located proximate to the aflatoxin gene cluster on chromosome 3, 5,010,000 to 5,034,262 (Fig. 4). The ditryptophenaline gene cluster region is located on chromosome 4, 3,181,000 to 3,195,000 (Fig. 5).

Group A strains exhibited a shared pattern of deletions and mutations within their gene clusters. The pattern of the CPA gene cluster appeared to align with the deletion pattern observed in the aflatoxin gene cluster. For example, strains with a large deletion in the aflatoxin gene cluster also exhibited a large deletion in the CPA gene cluster, while strains with intact aflatoxin clusters maintained more complete CPA gene cluster regions. These similarities were visually assessed based on gene cluster genotype comparison plots.

For the ditryptophenaline cluster, Group A strains shared a similar mutation pattern. Group B showed similar mutation patterns with part of the Group A strains in the aflatoxin and CPA gene clusters, but they had intact ditryptophenaline gene clusters.

Most of the Group C strains had intact three gene clusters. Conversely, three aflatoxigenic strains (AR028, SD016, and SL015) exhibited unique mutation patterns in the aflatoxin and CPA gene clusters. This mutation pattern is odd compared to Group A and B mutation patterns.

Five non-aflatoxigenic strains from the NO category (2017 Washington-T2, A1, AF36, NRRL30797, and VCG1) commonly had an intact aflatoxin gene cluster,

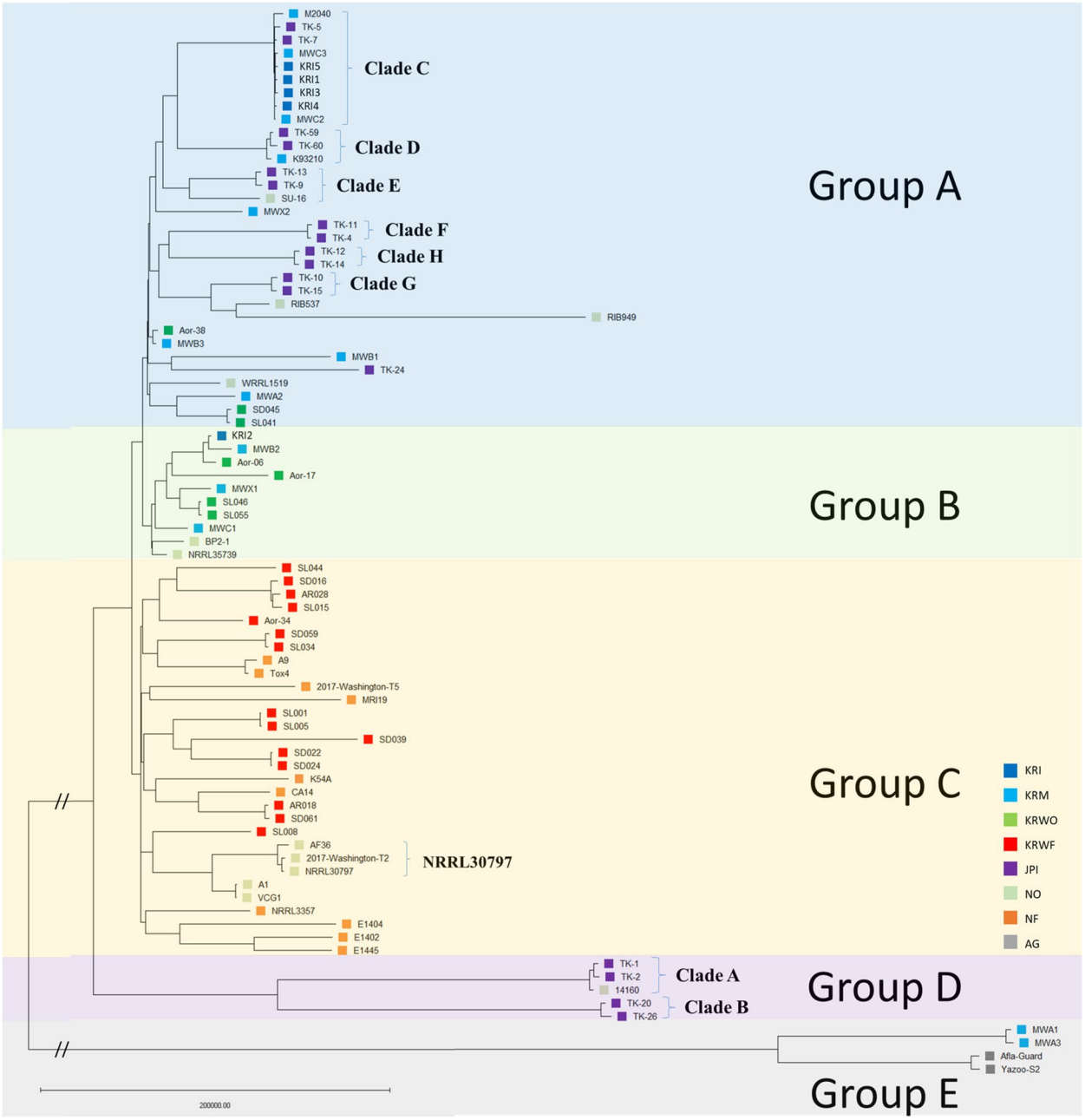

Fig. 1 Phylogenetic tree of 80 *A. oryzae/flavus* complex strains from eight different sources, categorized into five distinct groups (A-E). Color coding denotes the different sources: KRI (Korean Industrial non-aflatoxigenic strains, Blue), KRM (Korean non-aflatoxigenic strains from meju, Cyan), KRWO (Korean non-aflatoxigenic strains from wild conditions, Green), KRWF (Korean aflatoxigenic strains from wild conditions, Red), JPI (Japanese industrial strains, Purple), NO (NCBI non-aflatoxigenic strains, Pale green), NF (NCBI aflatoxigenic strains, Orange), and AG (Afla-guard non-aflatoxigenic strains, Grey). Branch lengths represent genetic distances

but exhibited a small deletion in the *cypA* gene at the same location as observed in Group A and B strains.

Group D strains presented a distinctive pattern with large deletions in the aflatoxin cluster and complete deletion in the CPA cluster, maintaining an intact Ditryptophenaline gene cluster.

In Group E, KRM strains and Afla-guard strains showed distinct features. In contrast to other strains, where the CPA gene cluster pattern depends on the aflatoxin gene cluster's state, KRM strains uniquely exhibited a deleted aflatoxin gene cluster while retaining the CPA gene cluster. Afla-guard strains had a

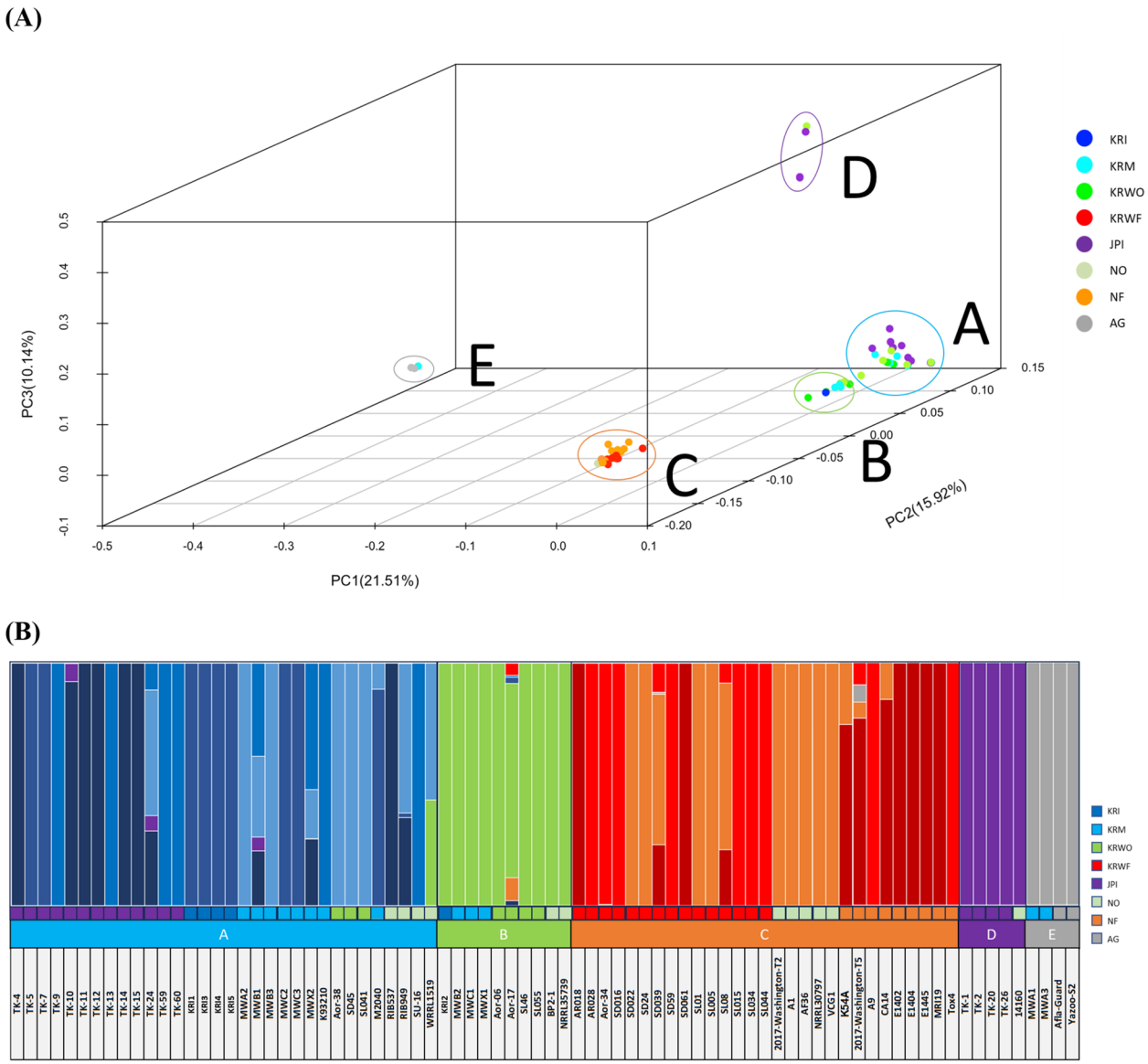

Fig. 2 (A) Principal component analysis (PCA) plot of 80 *A. oryzae/flavus* strains. Each dot represents an individual strain, with colors corresponding to the groups identified in the phylogenetic analysis: KRI (Blue), KRM (Cyan), KRWO (Green), KRWF (Red), JPI (Purple), NO (Pale green), NF (Orange), and AG (Grey). The first two principal components (PC1 and PC2) capture the majority of the genetic variance among the strains, highlighting the closer genetic relationships between certain groups and the distinctiveness of others. (B) Admixture plot depicting the estimated population structure and genetic admixture of *A. flavus* and *A. oryzae* strains across varying numbers of ancestral populations (K values). Each vertical bar represents an individual strain, partitioned into colored segments that reflect the strain's estimated proportion of membership in each of the K-inferred genetic clusters

whole deletion in the aflatoxin and CPA gene clusters. However, these Group E strains had an intact ditryptophenaline gene cluster similar to the others. No group-specific characteristics were observed among the groups concerning Aspergillic acid and Aflatrem gene clusters.

MALDI TOF/MS patterning

Comparing patterns of the peaks of Group C and the non-C groups (A, B and E; *A. oryzae* groups), the two groups were distinctly separated, while non-C groups were not distinctly separated from each other (Fig. 6A). Group C and the non-C groups exhibited several

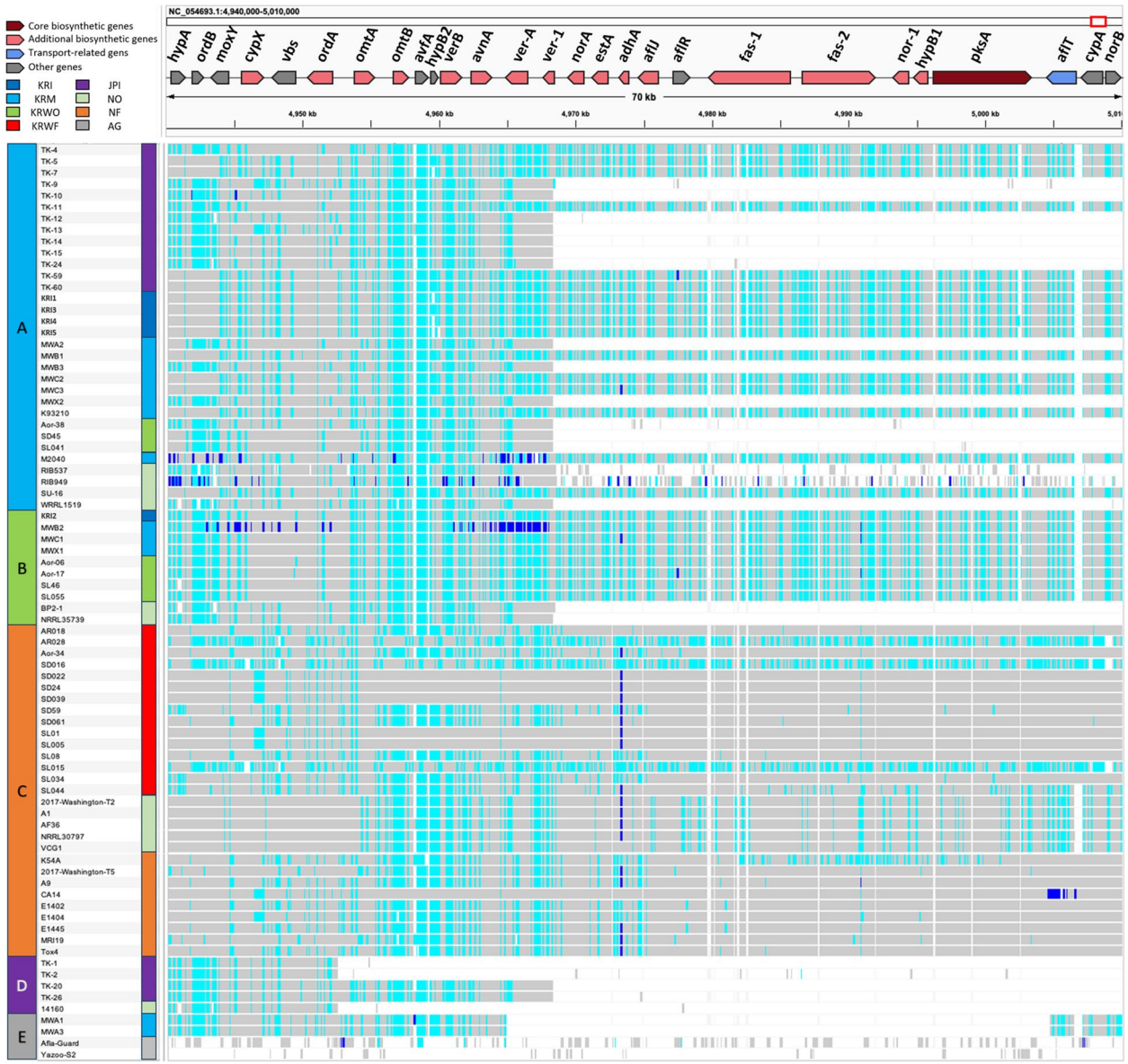

Fig. 3 Aflatoxin biosynthesis gene cluster genotype comparison. Grey bars indicate genomic matches to the *A. flavus* reference genome (NRRL 3357), showing no mutations. Cyan bars depict homozygous mutations differing from the reference. Blue bars mark heterozygous mutations. Absent bars signal deletions, where genomic regions are missing. The gene clusters identified through antiSMASH were overlaid on the corresponding genomic regions

significantly different loci; 3100–3200 m/z, 6200–6500 m/z, and 12,700–12900 m/z showed significant differences (Fig. 6B). When comparing the non-C groups, no distinct loci were found that significantly differentiated between the groups in overall ranges (Supplementary Fig. 2). Most of the Group C strains had significantly higher peaks in the 6355 to 6385 m/z and 12,795 to 12,810 m/z ranges, while most of the non-C groups had high peaks at 3145 m/z, 6285 to 6295 m/z, and 12,730 m/z.

Discussion

The phylogenetic analysis and genome structure investigation of *A. oryzae/flavus* strains from fermentation and wild environment in Korea, compared with globally reported strains, provided significant insights into the genetic diversity and evolutionary pathways of these fungi. The phylogenetic tree, PCA and ADMIXTURE analysis offer an illustration of the genetic distances and relationships among the strains, revealing a broad spectrum of evolutionary divergence within five distinct

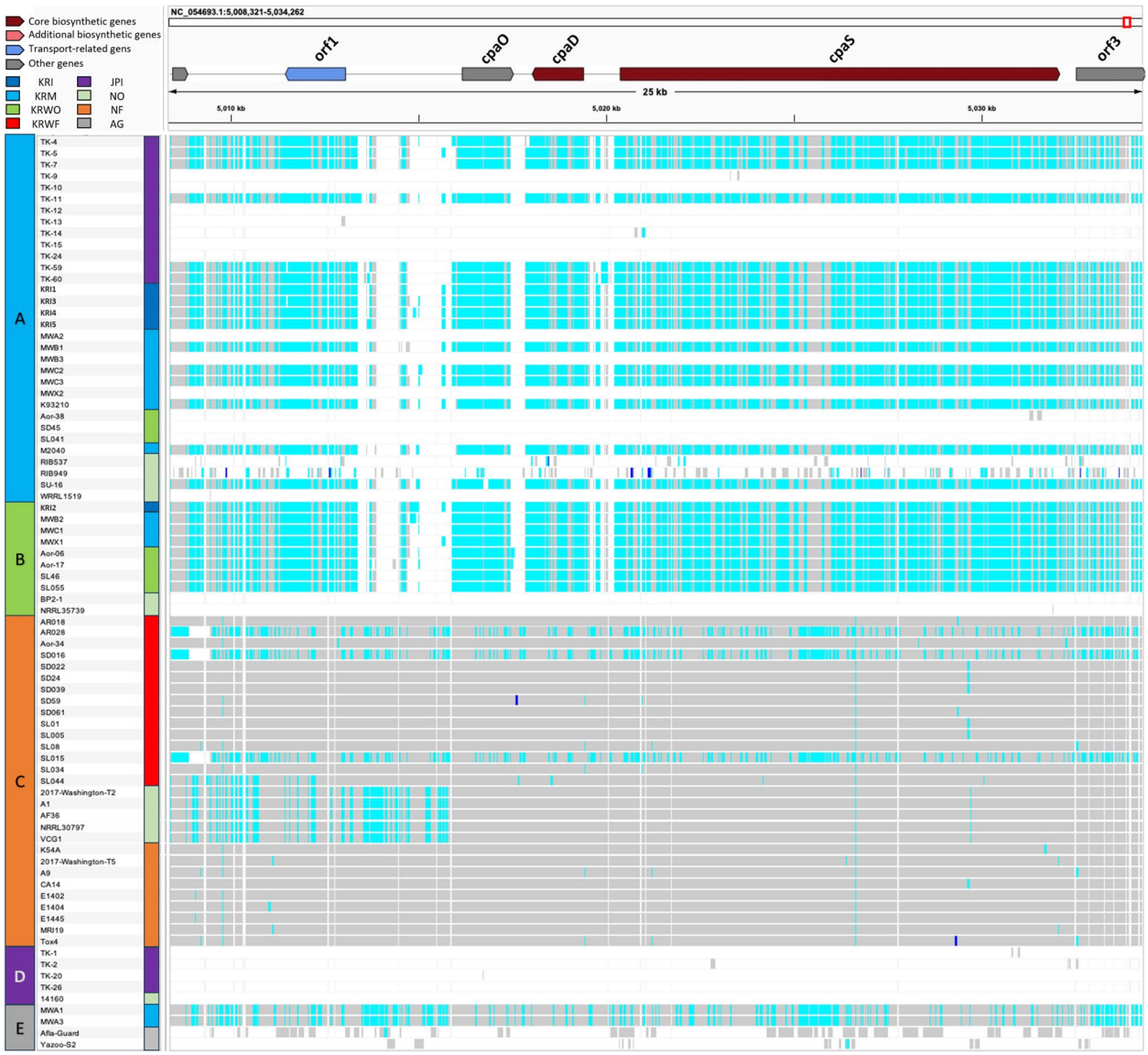

Fig. 4 Cyclopiazonic acid biosynthesis gene cluster genotype comparison. Grey bars indicate genomic matches to the *A. flavus* reference genome (NRRL 3357), showing no mutations. Cyan bars depict homozygous mutations differing from the reference. Blue bars mark heterozygous mutations. Absent bars signal deletions, where genomic regions are missing. The gene clusters identified through antiSMASH were overlaid on the corresponding genomic regions

groups (A to E). Groups A, B, D and E are clustered with putative *A. oryzae* strains while Group C is clustered with putative *A. flavus* strains. Groups A and D consist mainly of industrial strains, Group B is dominated by wild Korean strains, and Group E contains a smaller group including the biocontrol strain Afla-guard.

In particular, Groups A and B demonstrate closer genetic affiliations, suggesting a shared evolutionary pathway that may be rooted in their non-aflatoxigenic nature and potential adaptations to industrial or natural environments. Conversely, Group E, characterized by its distinct genetic makeup, stands out as the most genetically divergent cluster. This may indicate unique evolutionary pressures or historical genetic isolation that have shaped their current genomic constitution. Group D, consisted with JPI strains that fall into the A and B clades as defined by Watarai et al., exhibits an interesting feature since it has a greater genetic distance from Groups A and B than from Group C, yet PCA analysis showed that it aligns more closely with Groups A and B.

Previous research by Watarai et al. [5] categorized JPI strains into eight clades (Clade A to H), and strains

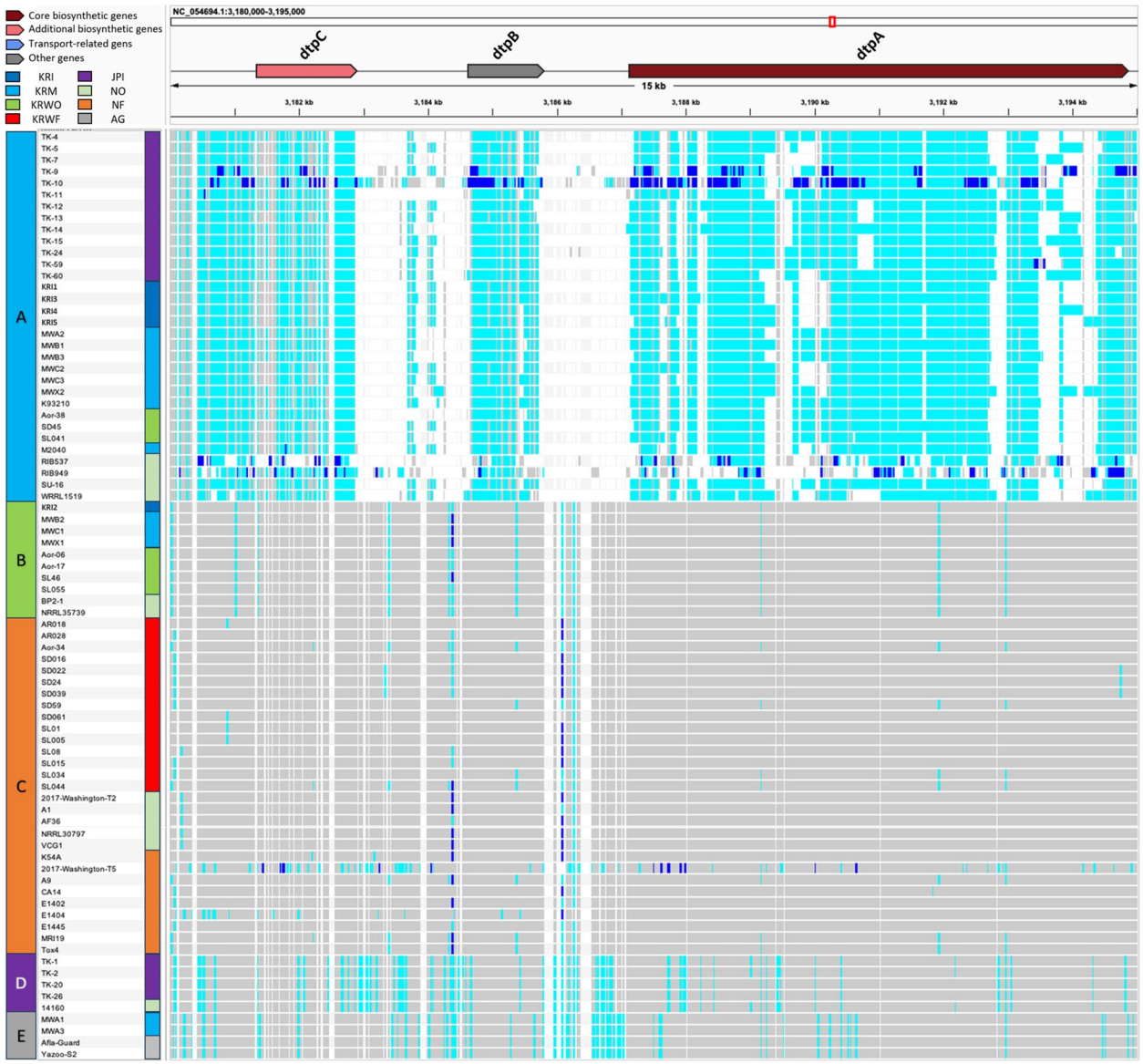

Fig. 5 Ditryptophenaline biosynthesis gene cluster genotype comparison. Grey bars indicate genomic matches to the *A. flavus* reference genome (NRRL 3357), showing no mutations. Cyan bars depict homozygous mutations differing from the reference. Blue bars mark heterozygous mutations. Absent bars signal deletions, where genomic regions are missing. The gene clusters identified through antiSMASH were overlaid on the corresponding genomic regions

belonging to Clades C to H were integrated into Group A. Considering distribution of Korean strains, industrial strains (KRI), excluding KRI2, clustered within Group A, specifically aligning with JPI Clade C strains. The narrow genetic distances among these industrial strains suggest recent differentiation driven by fermentation functionalities, resulting in their separation as individual strains [5].

Korean Meju-originated strains (KRM) exhibited broader distribution across Groups A, B, and E. Some strains were closely matched to industrial strains, likely due to spore dispersal from industrial environments to traditional fermentation settings, as seen in the clustering of MWC2 and MWC3 with KRI strains [42]. Other Meju strains in Group A, B and E showed distinct genetic positions with industrial strains. Notably, two Meju strains in Group E showed genetic similarity with the biocontrol agent, Afla-guard but differed in aflatoxin and CPA gene clusters, suggesting potential genetic recombination events with wild strains. The distribution of Meju strains (KRM) was similar with that of non-aflatoxigenic

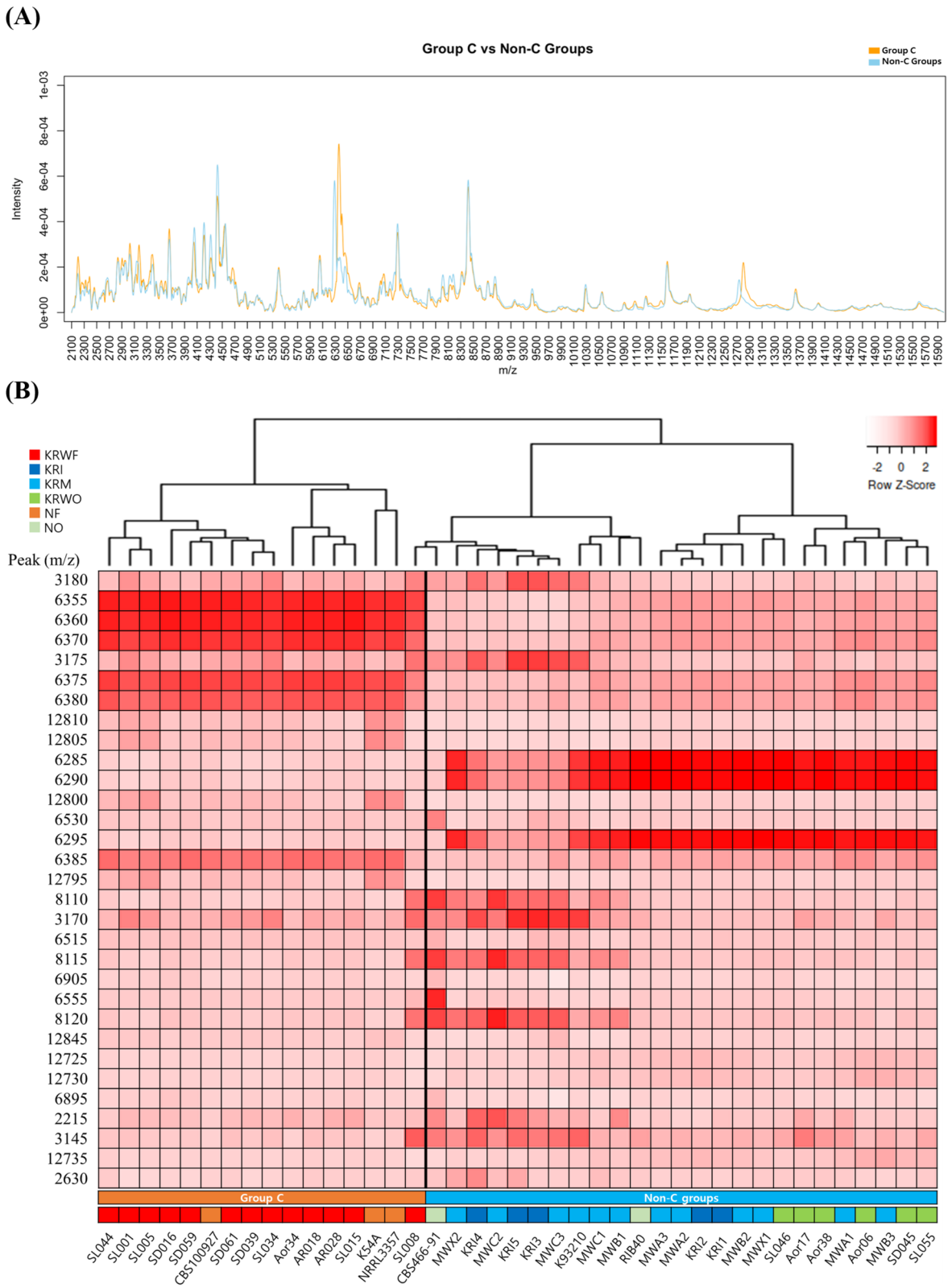

Fig. 6 **A** Average intensity of Group C and Non-C group strains in 2100 to 16,000 m/z. **B** The matrix of peak intensity of mass-spectral loci among KACC strains. The mass-spectral loci according to the difference ($p < 0.001$) between Group C and non-C groups are listed on the right side of the matrix

strain from wild conditions (KRWO). This wide distribution compared to industrial strains clusters supports the notion that traditional Meju fermentation without artificial starter inoculation, resulting in weak selective pressure. A similar observation was reported in previous research by Dumas et al. [15], which investigated the blue cheese fermentation starter, *P. roqueforti*. In this case, genomic comparison between industrially developed non-roquefort group and naturally domesticated roquefort group revealed that the roquefort strains exhibited weaker selective pressure compared to the non-roquefort strains. This suggests that the natural domestication process of *P. roqueforti*, like that of Meju strains, involved environmental influences with minimal artificial intervention, leading to broader genetic diversity [15].

Non-aflatoxigenic strains from Korean wild environments (KRWO) were found in Groups A and B. Group A wild strains exhibited genetic distinctions from JPI strains, suggesting different evolutionary pressures. These wild strains hold intermediate genetic positions between industrial cluster and aflatoxigenic cluster, suggesting they may represent an evolutionary link between these groups.

All aflatoxigenic Korean strains (KRWF) were categorized in Group C. Especially strains AR028, SD016, and SL015 had notable genetic characteristics highlighted in previous studies [9]. PCR analysis patterns of the norB/cypA cluster in these strains were similar to that of *A. parasiticus*. Genome-wide clustering confirmed their placement within Group C, yet their aflatoxin cluster mutation patterns were notably distinct from NRRL 3357 and other Group C strains. Despite these genomic differences, MALDI-TOF MS results closely matched NRRL 3357 and other Group C strains. These genomic variations are likely to be neutral mutations that do not substantially affect protein expression or structure since their proteomic profiles remained similar. This proteomic resemblance may also suggest similar metabolomics characteristics among these strains [43].

Aflatoxin mutation patterns in this study have some points of concordance with previous research, Han et al. [14], as well as advancements in understanding strains mutation. Comparing the strains, Afla-guard (NRRL 21882), which had a complete deletion of the aflatoxin gene cluster, and strain 14,160, which exhibited deletion starting from the *omtA* gene showed the same results. Strains BP2 - 1, TK- 10, TK24, WRRL1519, and NRRL35739 showed deletions beginning from the *norA* gene. These results are also consistent with previous findings [14].

Additionally, strains 3.042, TK- 5, TK- 59, and SU- 16 were described as having mostly intact gene clusters with partial deletions starting from the *AflT* gene. Group C strains (A9, CA14, NRRL 3357, VCG1, Washington T5, E1404, E1445) were reported to have partial deletions starting from the *cypA* gene [14]. In this study, non-aflatoxigenic strains, including those in Group C (2017 Washington-T2, A1, AF36, NRRL30797, and VCG1), frequently exhibited additional deletions of the *cypA* gene compared to NRRL 3357. This suggests that the additional deletion or mutation in *cypA* may significantly impact aflatoxin production. This finding also aligns with previous research that classified *norB-cypA* PCR patterns into type 1 (short read type) and type 2 (long read type), with aflatoxigenic strains being type 2 and non-aflatoxigenic strains being type 1 [9]. There were also three type 3 (longest read type, *A. parasiticus* type) strains, AR028, SD016, and SL015, and these three strains had specific mutation pattern in *norB-cypA* region. Previous study revealed that *A. parasiticus* has more intact *norB-cypA* gene region compared to *A. flavus*, which allow it to produce G-type aflatoxins in addition to B-type aflatoxins [44].

Across Groups A and B, strains with high mutation rates or high deletion rates are mixed and each of them shares a similar mutation pattern, but they don't share similar mutation patterns with Group C strains. This suggests significant genetic exchange through interbreeding between Groups A and B during domestication, while the two groups appear to be reproductively isolated from the aflatoxigenic group, indicating speciation. Therefore, these strains with high mutation rates or high deletion rates are considered to have less potential for mycotoxin production, and unlikely to form hybrids with aflatoxigenic strains.

Group C mostly consists of aflatoxigenic strains (KRWF and NF), but there are some exceptional non-aflatoxigenic strains (2017 Washington-T2, A1, AF36 and K54 A). Among them, AF36 is even known as biocontrol agent to reduce aflatoxin contamination [45]. This suggests that the genetic landscape may not always strictly correlate with aflatoxin production, and the potential for cyclopiazonic acid (CPA) production should also be considered. Furthermore, the presence of largely intact aflatoxin and CPA gene clusters in these non-aflatoxigenic strains indicates a latent potential for mycotoxin production under specific conditions. This raises important concerns for risk of mycotoxin contamination in food industry applications, even though CPA is not currently a regulated mycotoxin in most regions.

The evolutionary direction of Group A, characterized by unique mutations in the ditryptophenaline gene cluster, indicates a potential adaptive response to industrial fermentation processes, possibly due to selective pressures to mitigate negative impacts on food products. Although the toxicity or side effects of

ditryptophenaline have not been thoroughly studied, the uniform occurrence of mutations in this cluster among Group A strains implies that such mutations may confer a selective advantage in the context of fermented food production, possibly due to reduced detrimental effects on product quality [11].

By incorporating MALDI-TOF MS data, we observe significant differentiation, particularly between Group C and the non-C groups (A, B, and E), with unique proteomic features corresponding to their genomic distinctions. Group C exhibited the unique features in the ranges of 3100–3200 m/z, 6200–6500 m/z, and 12,700–12900 m/z. A database of various proteins specific for *A. flavus* and *A. oryzae* in UniProtKB (www.uniprot.org) suggests that these peaks may correspond to significant fungal proteins. For instance, peaks within the 12,700 to 12,900 m/z range may correspond to proteins such as AflF, also known as norB (primary accession number: A0 A7U2MNK6, Mass: 12,855 Da). This protein is one of the enzymes included in the aflatoxin gene cluster. No characterized proteins within the 3100 to 3200 m/z range were identified in the database. These specific proteins play crucial roles in cellular metabolism, suggesting their potential as reliable biomarkers for differentiating between groups.

Previous attempts to differentiate *A. oryzae* from *A. flavus* have been numerous. A previous study showed that targeting the *Cyp51 A* gene could provide differentiation, but the limited strain diversity hindered its broad applicability [46]. There were also attempts to analyze aflatoxin and cyclopiazonic acid gene clusters to see the difference between two species, and although some differentiation was achieved, it was not definitive [3, 47]. There have also been attempts to distinguish the two species through profiling of CAZyme (Carbohydrate-Active Enzyme) genes and secondary metabolite biosynthesis gene clusters, but the ambiguous similarities between the two species disturbed clear differentiation [14].

In this study, a comprehensive approach was employed, combining whole-genome SNP-based population structure analysis, detailed secondary metabolite gene cluster variation analysis, and MALDI-TOF MS profiling.

Group C formed a distinct cluster in genomic analysis, separating it from other groups. Similarly, MALDI-TOF MS analysis distinguished Group C from other groups, displaying unique peaks. Furthermore, group C strains have comparatively intact gene clusters of aflatoxin and cyclopiazonic acid and produced aflatoxin B. Given these characteristics, Group C aligns with traits traditionally associated with *A. flavus*. Therefore, this study proposes classifying Group C as *A. flavus* and the non-C groups as *A. oryzae*.

MALDI-TOF MS is a highly efficient and cost-effective tool, providing results within minutes from a single colony [48, 49]. In this study, MALDI-TOF MS identified specific markers for the *A. oryzae/flavus* complex, with unique peaks at 6200–6500 m/z and 12,700–12900 m/z. These proteomic patterns serve as reliable markers to rapidly distinguish *A. flavus* from *A. oryzae*. Combined with genomic analysis, this method offers a practical advantage for routine microbial identification and differentiation, improving both accuracy and efficiency in species classification.

Conclusion

This study provides a comprehensive genomic and proteomic characterization of Korean *A. oryzae/flavus* complex strains compared to global strains. The analysis highlights the genetic diversity and distinct evolutionary paths of these strains, with particular attention to their mycotoxin contamination for food industry applications. Genomic analysis divided the *A. oryzae*/*flavus* strains into five groups, with Group C showing relatively intact aflatoxin and cyclopiazonic acid (CPA) gene clusters, suggesting a high potential for producing both metabolites. Furthermore, unique proteomic patterns identified through MALDI-TOF MS provide reliable biomarkers for distinguishing Group C aflatoxigenic strains. Characteristics of Group C were well matched with those of *A. flavus* and the other groups with *A. oryzae*. These findings establish refined criteria for differentiating *A. oryzae* from *A. flavus*, contributing their safer utilization in terms of mycotoxin management and better understanding of their genetic dynamics.

Abbreviations

CPA	Cyclopiazonic acid
KRI	Korean Industrial non-aflatoxigenic strains
KRM	Korean non-aflatoxigenic strains from meju
KRWO	Korean non-aflatoxigenic strains from wild conditions
KRWF	Korean aflatoxigenic strains from wild conditions
JPI	Japanese Industrial strains
NO	NCBI non-aflatoxigenic strains
NF	NCBI aflatoxigenic strains
AG	Afla-guard non-aflatoxigenic strains
SNP	Single nucleotide polymorphism
VCF	Variant calling format

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11596-9.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements
Not applicable.

Authors' contributions
D.H. Kim wrote the main manuscript text and prepared figures 1-5. S.B. Hong and K.T. Kim provided revisions to the contents of main manuscript. D.C. Kim and D.G. Seo prepared figure 6. S.H. Lee reviewed and corrected the manuscript's grammar and language.

Funding
This work is supported by the grant from the National Institute of Agricultural Sciences, Rural Development Administration, Republic of Korea (PJ017286), and the Ministry of Science, ICT (MSIT) of Korea (2021M3H9A1081101).

Data availability
The vcf files data for genetic polymorphisms of KACC strains were deposited on European Molecular Biology Laboratory European Variation Archive (EMBL EVA, https://www.ebi.ac.uk/eva/) as PRJEB79400 (https://www.ebi.ac.uk/ena/browser/view/PRJEB79400).'

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 6 August 2024 Accepted: 10 April 2025
Published online: 30 April 2025

References

1. Machida M, et al. Genome sequencing and analysis of Aspergillus oryzae. Nature. 2005;438(7071):1157–61.
2. Machida M, Yamada O, Gomi K. Genomics of Aspergillus oryzae: learning from the history of Koji mold and exploration of its future. DNA Res. 2008;15(4):173–83.
3. Gibbons JG, et al. The evolutionary imprint of domestication on genome variation and function of the filamentous fungus Aspergillus oryzae. Curr Biol. 2012;22(15):1403–9.
4. Alshannaq AF, et al. Controlling aflatoxin contamination and propagation of Aspergillus flavus by a soy-fermenting Aspergillus oryzae strain. Sci Rep. 2018;8(1):16871.
5. Watarai N, et al. Evolution of Aspergillus oryzae before and after domestication inferred by large-scale comparative genomic analysis. DNA Res. 2019;26(6):465–72.
6. Latgé J-P. Aspergillus fumigatus and aspergillosis. Clin Microbiol Rev. 1999;12(2):310–50.
7. Kosmidis C, Denning DW. The clinical spectrum of pulmonary aspergillosis. Thorax. 2015;70(3):270–7.
8. Rudramurthy SM, et al. Invasive aspergillosis by Aspergillus flavus: epidemiology, diagnosis, antifungal resistance, and management. Journal of Fungi. 2019;5(3):55.
9. Hong S-B, et al. The proportion of non-aflatoxigenic strains of the Aspergillus flavus/oryzae complex from meju by analyses of the aflatoxin biosynthetic genes. J Microbiol. 2013;51:766–72.
10. Kjærbølling I, et al. A comparative genomics study of 23 Aspergillus species from section Flavi. Nat Commun. 2020;11(1):1106.
11. Rank C, et al. Comparative chemistry of Aspergillus oryzae (RIB40) and A. flavus. Metabolites. 2012;2(1):39–56.
12. Medina A, Rodriguez A, Magan N. Effect of climate change on Aspergillus flavus and aflatoxin B1 production. Front Microbiol. 2014;5:348.
13. Sun H, et al. Safety evaluation and comparative genomics analysis of the industrial strain Aspergillus flavus SU-16 used for huangjiu brewing. Int J Food Microbiol. 2022;380: 109859.
14. Han DM, et al. Comparative pangenome analysis of Aspergillus flavus and Aspergillus oryzae reveals their phylogenetic, genomic, and metabolic homogeneity. Food Microbiol. 2024;119: 104435.
15. Dumas E, et al. Independent domestication events in the blue-cheese fungus Penicillium roqueforti. Mol Ecol. 2020;29(14):2639–60.
16. Kim E, et al. Differentiation between Weissella cibaria and Weissella confusa using machine-learning-combined MALDI-TOF MS. Int J Mol Sci. 2023;24(13):11009.
17. Hedayati MT, et al. Discrimination of aspergillus flavus from Aspergillus oryzae by matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry. Mycoses. 2019;62(12):1182–8.
18. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
19. Fountain JC, et al. Draft genome sequences of one Aspergillus parasiticus isolate and nine Aspergillus flavus isolates with varying stress tolerance and aflatoxin production. Microbiology Resource Announcements. 2020;9(37). https://doi.org/10.1128/mra00478–20.
20. Chacón-Vargas K, et al. Comparison of two Aspergillus oryzae genomes from different clades reveals independent evolution of alpha-amylase duplication, variation in secondary metabolism genes, and differences in primary metabolism. Front Microbiol. 2021;12: 691296.
21. Pennerman KK, et al. Aspergillus flavus NRRL 35739, a poor biocontrol agent, may have increased relative expression of stress response genes. Journal of Fungi. 2019;5(2):53.
22. Chang PK. Genome-wide nucleotide variation distinguishes Aspergillus flavus from Aspergillus oryzae and helps to reveal origins of atoxigenic A. flavus biocontrol strains. J Appl Microbiol. 2019;127(5):1511–20.
23. Schamann A, Geisen R, Schmidt-Heydt M. Draft genome sequence of an aflatoxin-producing Aspergillus flavus strain isolated from food. Microbiology Resource Announcements. 2022;11(2):e00894-e921.
24. Arias RS, et al., Sixteen draft genome sequences representing the genetic diversity of Aspergillus flavus and Aspergillus parasiticus colonizing peanut seeds in Ethiopia. Microbiology Resource Announcements. 2020. 9(30). https://doi.org/10.1128/mra00591–20.
25. Hua SST, et al. Characterization of aflatoxigenic and non-aflatoxigenic Aspergillus flavus isolates from pistachio. Mycotoxin Res. 2012;28:67–75.
26. Yin G, et al. Genome sequence and comparative analyses of atoxigenic Aspergillus flavus WRRL 1519. Mycologia. 2018;110(3):482–93.
27. Weaver MA, Mack BM, Gilber MK. Genome sequences of 20 georeferenced Aspergillus flavus isolates. Microbiology Resource Announcements. 2019;8(11). https://doi.org/10.1128/mra.01718–18.
28. Abbas H, et al. Comparison of major biocontrol strains of non-aflatoxigenic Aspergillus flavus for the reduction of aflatoxins and cyclopiazonic acid in maize. Food Addit Contam. 2011;28(2):198–208.
29. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589–95.
30. Li H, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
31. Van der Auwera GA, et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43(1):11.10-11.10. 33.
32. Rempel A, Wittler R. SANS serif: alignment-free, whole-genome-based phylogenetic reconstruction. Bioinformatics. 2021;37(24):4868–70.
33. Chang CC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4(1):s13742-015-0047–8.
34. Yang J, et al. GCTA: a tool for genome-wide complex trait analysis. The American Journal of Human Genetics. 2011;88(1):76–82.
35. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655–64.
36. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics. 2011;12:1–6.
37. Yu J, et al. Clustered pathway genes in aflatoxin biosynthesis. Appl Environ Microbiol. 2004;70(3):1253–62.
38. Amaike S, Keller NP. Aspergillus flavus. Annu Rev Phytopathol. 2011;49:107–33.
39. Caceres I, et al. Aflatoxin biosynthesis and genetic regulation: A review. Toxins. 2020;12(3):150.
40. Chang P-K, Ehrlich KC, Fujii I. Cyclopiazonic acid biosynthesis of Aspergillus flavus and Aspergillus oryzae. Toxins. 2009;1(2):74–99.
41. Kishimoto S, et al. Evaluation of biosynthetic pathway and engineered biosynthesis of alkaloids. Molecules. 2016;21(8):1078.
42. Kim D-H, et al. The mycobiota of air inside and outside the meju fermentation room and the origin of meju fungi. Mycobiology. 2015;43(3):258–65.

43. Clark CM, et al. Coupling MALDI-TOF mass spectrometry protein and specialized metabolite analyses to rapidly discriminate bacterial function. Proc Natl Acad Sci. 2018;115(19):4981–6.
44. Ehrlich KC, et al. Aflatoxin biosynthesis cluster gene cypA is required for G aflatoxin formation. Appl Environ Microbiol. 2004;70(11):6518–24.
45. Doster MA, Cotty PJ, Michailides TJ. Evaluation of the atoxigenic Aspergillus flavus strain AF36 in pistachio orchards. Plant Dis. 2014;98(7):948–56.
46. Nargesi S, et al. Differentiation of Aspergillus flavus from Aspergillus oryzae targeting the cyp51A gene. Pathogens. 2021;10(10):1279.
47. Toyotome T, et al. Comparative genome analysis of Aspergillus flavus clinically isolated in Japan. DNA Res. 2019;26(1):95–103.
48. Kim J-M, et al. Rapid discrimination of methicillin-resistant Staphylococcus aureus by MALDI-TOF MS. Pathogens. 2019;8(4):214.
49. Sogawa K, et al. Use of the MALDI BioTyper system with MALDI–TOF mass spectrometry for rapid identification of microorganisms. Anal Bioanal Chem. 2011;400:1905–11.

Publisher's Note