Fungi population metabolomics and molecular network study reveal novel biomarkers for early detection of aflatoxigenic Aspergillus species

Journal of Hazardous Materials

Xie, Huali; Wang, Xiupin; Hooft, Justin J.J.; Medema, Marnix H.; Chen, Zhi Y. et al

https://doi.org/10.1016/j.jhazmat.2021.127173

Journal of Hazardous Materials 424 (2022) 127173

Contents lists available at ScienceDirect

Journal of Hazardous Materials

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

Fungi population metabolomics and molecular network study reveal novel biomarkers for early detection of aflatoxigenic *Aspergillus* species

Huali Xie [a,b,c,f], Xiupin Wang [a,c,d], Justin JJ van der Hooft [f], Marnix H. Medema [f], Zhi-Yuan Chen [g], Xiaofeng Yue [a,c], Qi Zhang [a,b,c,d,e,*], Peiwu Li [a,b,c,d,e,*]

[a] Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430061, China
[b] Key laboratory of Detection for Aflatoxins, Ministry of Agriculture, Wuhan, China
[c] Laboratory of Risk Assessment for Oilseeds Products (Wuhan), Ministry of Agriculture, Wuhan 430061, China
[d] Quality Inspection and Test Center for Oilseeds Products, Ministry of Agriculture, Wuhan 430061, China
[e] Hubei Hongshan Laboratory, Wuhan, China
[f] Bioinformatics Group, Wageningen University, 6708PB Wageningen, The Netherlands
[g] Department of Plant Pathology and Crop Physiology, Louisiana State University Agricultural Center, Baton Rouge, LA 70803, USA

ARTICLE INFO

Editor: Teresa A.P. Rocha-Santos

Keywords:
Aflatoxins
Biomarkers
Early detection
Metabolomics
Machine learning

ABSTRACT

Mycotoxins threaten global food safety, public health and cause huge socioeconomic losses. Early detection is an effective preventive strategy, yet efficient biomarkers for early detection of aflatoxigenic *Aspergillus* species are lacking. Here, we proposed to use untargeted metabolomics and machine learning to mine biomarkers of aflatoxigenic *Aspergillus* species. We systematically delineated metabolic differences across 568 extensive field sampling *A. flavus* and performed biomarker analysis. Versicolorin B, 11-hydroxy-O-methylsterigmatocystin et.al metabolites shown a high correlation (from 0.71 to 0.95) with strains aflatoxin-producing capacity. Molecular networking analysis deciphered the connection of aflatoxins and biomarkers as well as potential emerging mycotoxins. We then developed a model using the biomarkers as variables to discern aflatoxigenic *Aspergillus* species with 97.8% accuracy. A validation dataset and metabolome from other 16 fungal isolates confirmed the robustness and specificity of these biomarkers. We further demonstrated the solution feasibility in agricultural products by early detection of biomarkers, which predicted aflatoxin contamination risk 35–47 days in advance. A developed operable decision rule by the XGBoost algorithm help regulators to intuitively assess the risk prioritization with 87.2% accuracy. Our research provides novel insights into global food safety risk assessment which will be crucial for early prevention and control of mycotoxins.

1. Introduction

Mycotoxins are toxic secondary metabolites produced by fungi during infection of peanuts, corn, cotton, tree nuts, and other susceptible crops both pre- and post-harvest (storage and processing) (T. Wang et al., 2017; B. Wang et al., 2017). Mycotoxin contamination of food and feed is not only high in incidence and prevalence worldwide, but also extremely harmful, as these mycotoxins cause a variety of health problems for humans and domestic animals (Roze et al., 2013). For instance, they possess carcinogenic (Huang et al., 2017), immunosuppressive, hepatotoxic, nephrotoxic, and neurotoxic properties (Chawanthayatham et al., 2017). They also cause huge economic losses annually (Mitchell et al., 2016). The most serious producer of carcinogenic mycotoxins is *Aspergillus flavus*, which is deemed one of the world's ten most feared fungi (Hyde et al., 2018). This fungus overwinters in the soil or plants debris as conidia or sclerotia, which germinate to produce additional hyphae and conidia to initiate new food infections in the spring (Hedayati et al., 2007). As a result, many countries set strict standards for the maximum allowable levels of aflatoxin in food and feed. A series of management strategies to mitigate aflatoxin contamination are being developed, which include efforts in reducing pre-harvest aflatoxin contamination through enhancing host resistance (Sharma et al., 2018), optimizing cultural practices (Ndemera et al., 2018), applying atoxigenic biocontrol strains of *A. flavus* (Atehnkeng

Abbreviations: APC, aflatoxin-producing capacity; AFB_1, Aflatoxin B_1; VerB, Versicolorin B; UPLC-HRMS, Ultra Performance Liquid Chromatography Tandem High-resolution Mass Spectrometry; PCA, Principal Component Analysis; RF, Random Forest algorithm.

* Corresponding authors at: Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430061, China.
E-mail addresses: zhangqi01@caas.cn (Q. Zhang), lipeiwu@caas.cn (P. Li).

https://doi.org/10.1016/j.jhazmat.2021.127173
Received 13 June 2021; Received in revised form 4 September 2021; Accepted 6 September 2021
Available online 10 September 2021

et al., 2014), or improving storage conditions (Pietsch et al., 2020), and detoxifying contaminated grains to reduce post-harvest contamination (Adebo et al., 2017).

It is of major importance to develop a reliable prediction method, which contributes to preemptively prevent mycotoxins contamination. Currently, mycotoxin risk prediction includes macro- and micro-risk prediction. Macro-risk prediction is applied in modeling approaches to predict mycotoxin contamination using climate, soil, and crop physiological (in the pre-harvest) (Dovenyi-Nagy et al., 2020) or moisture content, the temperature, and carbon dioxide content (in the storage) (Jiang et al., 2019) as input variables. For example, the Battilani's group developed the AFLA-Maize model to predict aflatoxin contamination risk in maize, and wheat in Europe by hourly temperature, relative humidity, and precipitation et.al climate factors. They predicted that Aflatoxin B_1 will become a food safety issue in Europe under climate change expected for the next years (Battilani et al., 2016). In the storage, Garcia-Cela reported that carbon dioxide production was an indicator of *A.flavus* colonization and aflatoxins/cyclopiazonic acid contamination in peanuts stored (Garcia-Cela et al., 2020). However, these models have certain limitations due to regional climatic differences, soil types, and cultivars of crops and are not sufficiently accurate to predict specific batches of samples. Micro risk prediction focuses on the development of sensitive and rapid methods for the early detection of pathogenic fungi to indicate mycotoxins contamination. The presence of toxin-producing fungi will cause the production of mycotoxins once the moisture content, temperature, and humidity are suitable for microbial growth. Micro-risk predictions complement macro-risk predictions, providing more accurate and direct risk assessment results for specific batches of samples. The micro-risk predictions strategy was mainly used in the post-harvest stage. Some methods for early detecting *A. flavus* in agricultural have been developed, such as detection of the expression of aflatoxin biosynthetic pathway genes (Peromingo et al., 2017), proteins (B. Wang et al., 2017; T. Wang et al., 2017), or chemical markers (Saldan et al., 2018). There are also disadvantages to the current micro-risk prediction method. For example, the prediction accuracy is much influenced by the representativeness of the sampling. At this point, macro-risk prediction provides complementary evidence for micro-risk prediction.

Furthermore, aflatoxin biosynthesis is affected by genetic and environmental factors and, as a result, a lot of *A. flavus* strains do not produce aflatoxin under natural conditions (Keller, 2019). Therefore, it is not sufficient to merely detect the presence of the fungus-based on *Aspergillus*-specific biomarkers, which do not always provide a good indication of aflatoxin-producing capacity (APC). Currently, biomarkers for effective discrimination of the high- and low-virulent *A. flavus* are not available. Therefore, the major objectives of our study were to: (1) develop biomarkers, for identifying aflatoxigenic strains with high aflatoxins-producing capacity (APC) from agricultural products and (2) monitor these biomarkers during food storage to predict the severity of aflatoxin contamination.

We hypothesize that (i) there are metabolites whose presence or abundance can distinguish between high- and low-virulent strains, which can be called aflatoxins-producing capacity (APC)-associated metabolites. The presence of a higher concentration of these metabolites can indicate the presence of aflatoxigenic *Aspergillus* species in agricultural products; and (ii) the onset of aflatoxin contamination can be predicted by aflatoxin biosynthetic pathway precursors or other metabolites, which can be called aflatoxin contamination severity (ACS)-associated metabolites. To test these hypotheses, here, we systematically compared the metabolic profiles from a large *A. flavus* population to discover biomarkers.

2. Materials and methods

2.1. A. flavus isolates and study design

Rhizosphere soil and peanut isolates of *A. flavus* were gathered from extensive field sampling samples, ranging from southern Zhanjiang, Guangdong province, to northern Tailai, Heilongjiang province, and from eastern Lianyungang, Jiangsu province, through Hongan, Hubei province, to western Chayu Tibet, from 2013 to 2018. A library of over 2000 *A. flavus* isolates was created by isolating strains, single-spore culturing, and species identification by ITS sequencing coupled with morphological identification. The overall experimental workflow for this study was constructed, which is shown in Fig. 1a. Firstly, 568 isolates of *A. flavus* from the above library random sampling to obtain a representative subset were used to comparative metabolomics studies. Specifically, 568 representative strains isolated from 16 peanut-producing provinces. The detailed strain information was listed in the supplementary sheet(strains-metadata). Among them, 54 strains were selected from the northern areas, 395 strains were selected from the central areas, and 119 strains were selected from the southern areas (Figs. 2a and S1).

2.2. Fungal culture and metabolites extraction

A. flavus was incubated into a Potato Dextrose Agar (PDA) medium in the dark at 29 ± 1 °C for 8–10 days. Conidia were suspended in a 0.1% tween-80 solution and were quantified by a hemocytometer. 50 mL of autoclaved liquid medium, containing 0.25% yeast extract, 0.1% K_2HPO_4, 0.05% $MgSO_4$–$7H_2O$, and 10% glucose (*p*H = 6.0) were inoculated with 2.5×10^5 conidia/mL. We set up three biological replicates for the training dataset samples. Conidia of *A. flavus* was cultured with shaking (180 rpm) at 29 ± 1 °C for 5 days in the dark. Subsequently, we collected the mycelium at 5 days by filtering the culture from the flasks with four layers of cheesecloth and washed it with 10 mL of 4 °C saline solution (0.9% NaCl). Then, mycelium was quickly transferred into a 50 mL centrifuge tube and quenched with liquid nitrogen. The mycelium samples were freeze-dried, and then 50 mg per sample was extracted by vortexing vigorously for 1 min with 1 mL of extraction solutions 1 (methanol: acetonitrile: water 2:2:1 v/v/v), which contained the internal standard of 2-chlorophenylalanine (20.0 μg/mL) and camphanic acid (25.0 μg/mL). The mycelia were homogenized in a ball mill for 4 min at 45 Hz, then it was treated for 5 min with ultrasound of 700 W (incubated in ice water). This process was repeated one more time using solution 2 (methanol: dichloromethane: ethyl acetate 1:1:1 v/v/v). The proteins were precipitated by incubating the homogenate for 1 h at −20 ℃ followed by centrifugation at 20,000 g/min for 10 min at −20 °C. 1 mL of the resulting supernatant was transferred into a fresh 2 mL LC-MS/MS glass vial for the UPLC-HRMS (Ultra performance liquid chromatography-tandem high-resolution mass spectrometry) analysis. Quality control (QC) samples were prepared by pipetting 10 μL from each sample extracts into a QC injection vial using a 100 μL pipette. Blank samples were prepared by pipetting 1 mL of pure methanol solvent (99.99%) into the injection vial named "blank".

2.3. Metabolomics data acquisition by UPLC-HRMS

Raw metabolomics data for all the isolates were acquired via a standardized metabolomics platform based on UPLC-HRMS. An Ultimate 3000 system (Dionex, Sunnyvale, CA, USA) and a C18 column (Hypersil Gold, 100 mm × 2.1 mm (i.d.)) with 3 μm pore size (Thermo Fisher Scientific, USA) was used for LC separation. Mobile phase A is H_2O: MeOH (95:5, v/v) containing 0.1% HCOOH, and 10 mM $HCOONH_4$ and mobile phase B is composed of MeOH: H_2O (95:5, v/v) containing 0.1% HCOOH and 10 mM $HCOONH_4$. They were mixed following the program below at a flow rate of 300 μL/min as a gradient mobile phase in the separation by UPLC. The elution gradient was

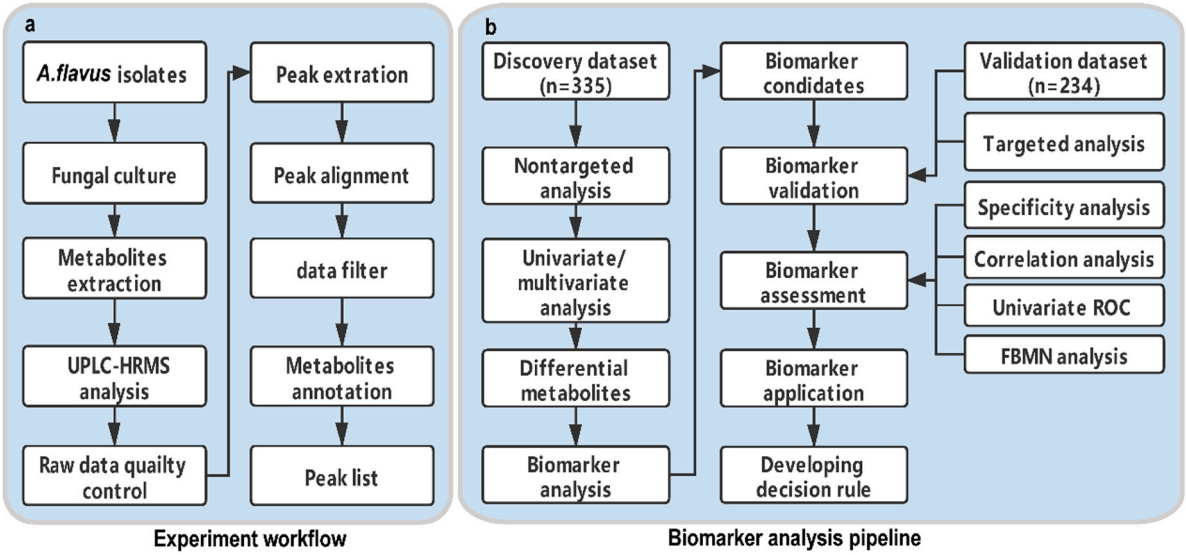

Fig. 1. (a) Experiment workflow of this study. (b) biomarker analysis pipeline.

programmed as: 0–1 min: 85% A (mobile phase A), 1–3 min: 85–50% A, 3–5 min: 50–30% A, 5–10 min: 30–0% A, 10–13 min: 0% A, 15 min: 0–85%A, and from 15 to 20 min: 85% A. Data were acquired by 2 μL of the extracted metabolites was injected into Orbitrap Fusion mass spectrometer (Thermo, USA).

We used a full-scan MS coupled with data-dependent fragmentation (DDA) mode to acquire the mass spectrometry data(Xie et al., 2018). The positive and negative ionization mode data were separately acquired. UPLC-HRMS was equipped with the HESI (Heated Electron Spray Ionization) ion source. The capillary voltage was set at −1.9 kV and the ion transfer tube temperature was set at 320 °C. Ion source working parameter settings were as follows: heater temperature 320 °C, spray voltage: static, positive ion: 3.5 kV, negative ion: 3.0 kV, sheath gas (Arb): 40, aux gas (Arb): 5, and sweep gas (Arb): 0. The main parameter settings for the full-scan MS experiment were identical as for the positive model and negative model except for polarity. The main parameters were as follow: detector type: Orbitrap, Orbitrap resolution: 12,000, scan range (*m/z*): 100–1200, RF lens (%): 60, AGC target: $3.0e^5$, maximum injection time (ms): 100, and source fragmentation: disabled. Filter parameters include setting intensity threshold at $1.0e^4$, charge state: 1–2, and dynamic exclusion after 1 time. The exclusion duration was the 60 s. Top speed was selected in the data dependent mode and the number of scan event types was set at 1 s. The main parameter settings for the DDA experiment were identical as for positive mode and negative mode except for the high-energy collision-induced dissociation (HCD). In positive mode, a type stepped collision energy in HCD activation was set at 40 ± 5 eV and 33% for fragmentation of the isolated precursor ions, and HCD cell pressure was 8 mTor. However, in negative mode, the value of HCD was set at 30 ± 5 eV. Other parameters included detector type: Orbitrap, a mass resolution of 30000 FWHM, AGC target: $5.0e^4$, and maximum injection time was 100. Ions for MS2 were isolated using the quadrupole with a narrow isolation window of ±1 Th. Instrument control and data processing were carried out by Xcalibur 4.0 software (Thermo Fisher Scientific).

2.4. Metabolomics data preprocessing

2.4.1. Data quality control. Three methods used for data quality control. Firstly, two internal standards were added to the extraction solvent to monitor experimental deviations in the sample pre-treatment process to ensure the quality of the data. Meanwhile, a widely used assessment criterion was employed to assess the quality of metabolomic data. Namely, If the data from quality control (QC) samples were all closely clustered to the origin in the PCA scores plot, indicating that the quality of data was suitable for subsequent analysis (Hu et al., 2019). Furthermore, we inserted a sample of blank solvent per 10 samples to monitor the cross-contamination between samples. A diagram of the sample run sequence shown in Fig. S2h.

2.4.2. Peak extraction, alignment, filter and gap-filled. Metabolomics raw data files were converted from the.raw data format to.mzXML format by the ProteoWizard software (Chambers et al., 2012). All.mzXML file were then imported into MZmine 2v53(Linux version) (Pluskal et al., 2010) and processed using the following procedure: ① We extracted the MS1 and MS2 mass information by setting the noise level at 1E5 and 0, respectively. ② In the ADAP chromatogram builder, a minimum group size of scans set 5, a minimum highest intensity set 2E5, and an *m/z* tolerance set 0.01 Da. ③ The baseline cut-off deconvolution algorithm was selected with the following settings: min peak height: 2.0E5, peak duration range(min): 0.05–3.0, Baseline level: 2E5, *m/z* range for MS2 scan pairing(Da): 0.01, RT range for MS2 scan pairing(min): 0.2. ④ We set the isotopic peaks grouper algorithm with an *m/z* tolerance of 0.01 Da (or 10 ppm) and an RT tolerance of 0.2 min to group isotopes. ⑤ Peak alignment used the join aligner module (*m/z* tolerance = 0.01 Da, weight for *m/z* = 2, weight for RT = 2, absolute retention time tolerance = 0.2 min). ⑥ Feature filtering was performed by feature list rows filter selection, minimum peaks in a row = 2, minimum peaks in an isotope pattern = 2. ⑦ The peak gap was filled with the same RT and *m/z* range gap filler module (*m/z* tolerance of 0.01 Da). ⑧ Eventually, a. csv quantitative file was exported using the option of "Export".

2.4.3. Metabolomics data matrix preprocessing. Metabolomics data matrix preprocessing included missing value processing and normalization. Missing value processing involved removing features with > 50% missing values, estimating the missing values of remaining features via replacing them with a small value (half of the minimum positive value in the original data). The normalization was performed in three steps: sample normalization, data transformation, and data scaling. The sample normalization allows general-purpose adjustment for differences among samples. We selected internal standard intensity as a normalization factor to remove the batch effect and mitigate the measurement deviation of the UPLC-HRMS across samples. Data transformation and scaling are two different approaches to make individual features more comparable. Log transformation and Pareto scaling were selected in this metabolomics study. All the data preprocessing was done in MetaboAnalyst 4.0 (Chong et al., 2018).

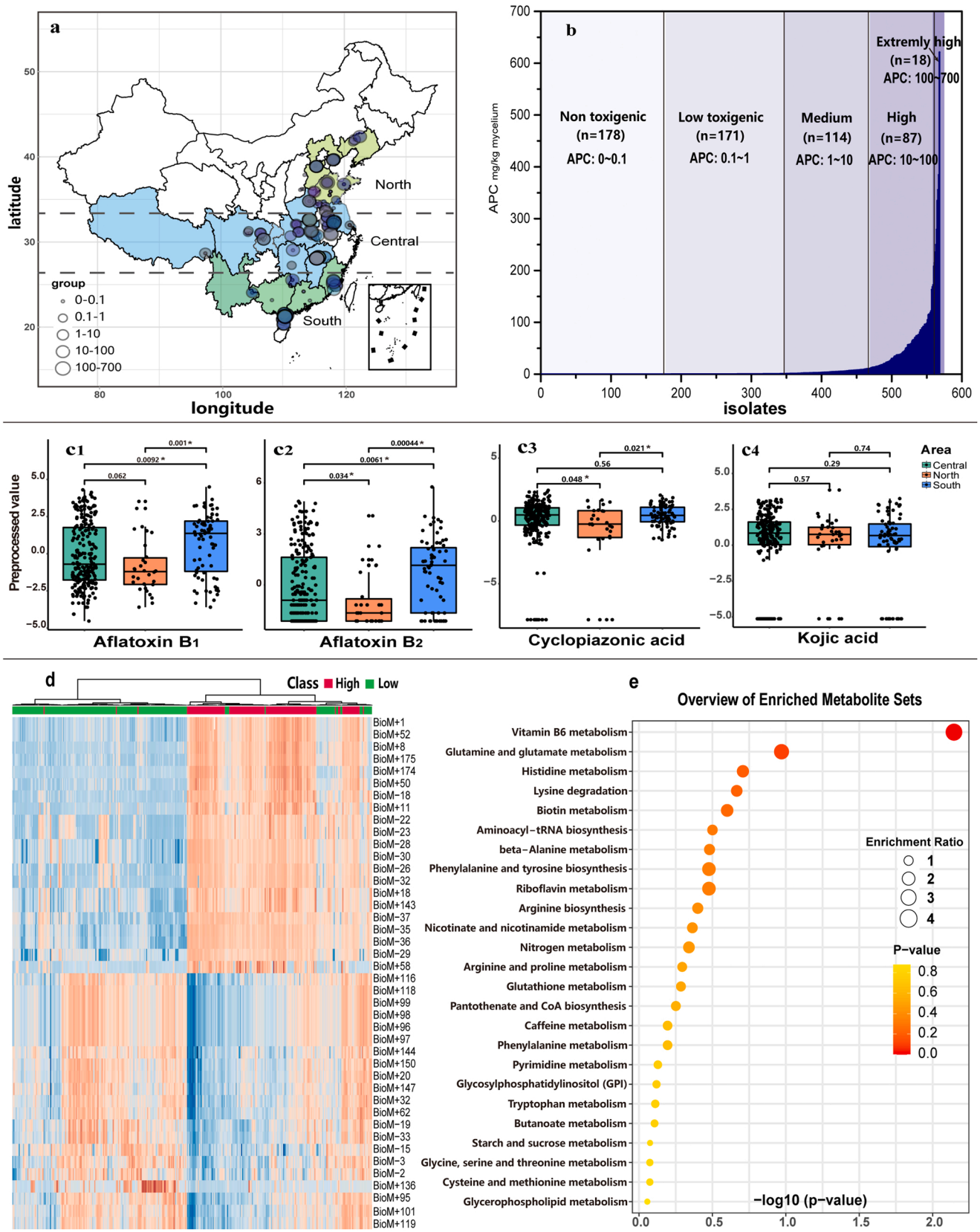

Fig. 2. (a) Mapping the aflatoxins-producing capacity of aflatoxin B_1(AFB_1) on the map; (b) Aflatoxins-producing capacity and proportions of *A. flavus* isolates belonging to non-, low-, middle-, high-, or extremely high-aflatoxic groups; (c) Boxplot visualizing the regional differences of Aflatoxin B_1(AFB_1), Aflatoxin B_2(AFB_2), cyclopiazonic acid (CPA), kojic acid (KA); (d) High/low aflatoxin-producing capacity strains metabolites abundance heatmap. (e) Overview of *Aspergillus flavus* population metabolites pathway enrichment analysis.

2.5. Univariate and multivariate statistical analysis for screening differential metabolites

The pipeline for biomarker analysis was shown in Fig. 1b. Univariate data analysis, (such as fold change, *t*-tests) was used to perform differential metabolites statistical tests. $P < 0.05$ and $0.05 < P < 0.10$ was considered as statistical significance. Multiple testing corrections were checked based on a significance threshold FDR-adjusted P values < 0.05. Fold change, *t*-tests, and principal component analysis (PCA) was carried out in different analysis modules within MetaboAnalyst 4.0 (Chong et al., 2018). The partial least squares discriminant analysis (PLS-DA) model, a multivariate statistical analysis method, was employed to screen differential metabolites (Xie et al., 2021) by Simca 14.1 (Umetrics, Umea, Sweden) software. A variable influence on projection (VIP) value ≥ 1.0 was used as a reference threshold to select the differential metabolites. Finally, we compared the results of univariate and multivariate statistical tests for comparison of their screening effectiveness and then we took the intersection of the metabolites screened by the different methods as biomarker analysis input dataset.

2.6. Metabolites annotation

Metabolomics raw data were imported into the software of Compound Discovery (CD) 2.1(Thermo Fisher Scientific, USA) to generate a data matrix that consists of the retention time (RT), mass-to-charge ratio (m/z) values, peak intensity, and annotation of metabolites. Metabolite annotation consists of the following steps: ① Unknown peaks were aligned and detected with the parameters of RT tolerance of 0.8 min and 5 ppm mass deviation. Minimum peak intensity set to 10,000 and S/N threshold = 3. ② The compounds were annotated with different types of databases. Among them, the mzCloud database was used to identify compounds on MS/MS level with a mass tolerance of 10 ppm. Compound class set to All, Match ion activation type = False. Chemspider, BioCyc, and KEGG database were used to annotate features based on exact mass (MS1) with a mass tolerance of 5 ppm. An endogenous metabolites database of 4400 compounds embedded in the CD internal database also was used to annotate metabolites. In addition, an in-house mass spectral library of > 3384 microbial natural products and mycotoxin reference standards was used to search by the search mass lists module in CD. ③mass tolerance set to 5 ppm, RT tolerance(min) = 0.05, S/N threshold = 1.5 infill gaps (missing values) module. We used Xcaliber 4.0 (Thermo Fisher Scientific, USA) to manually check and identify the metabolites identified above. Genesis peak detection method was selected to detect the peak area. Minimum peak height(S/N) set to 3. Metabolic features(m/z) with peak intensity greater than 1E4 in QC raw data files were retained. The metabolites that did not meet the requirements were discarded. Finally, compounds were identified and selected by integrating positive and negative ion mode to build a peak list for subsequent biomarker analysis.

2.7. Biomarker analysis by Random Forests (RF) model

Univariate and multivariate statistical analysis exists in poorly screening ability. There was a need to build more powerful models based on these differential metabolites to screen for the most taxonomically effective metabolites as biomarkers. For this reason, a supervised Random Forests (RF) classifier was trained to screen the potential biomarkers. Specifically, biomarker analysis was performed by Multivariate Exploratory ROC Analysis Module in MetaboAnalyst 4.0 (Chong et al., 2018). Firstly, we selected the Random Forest classifier and Random Forest feature ranking method to construct the Random Forests (RF) model. Then, confusion matrix (cross-validation) as a measure for internal evaluation of the models was also conducted using the biomarker analysis function in MetaboAnalyst 4.0. The exploration and validation of biomarkers stemmed from metabolic profiling comprised of three stages: (i) generating signatures (machine learning models) based on two thirds (2/3) training set of the samples that were only used to evaluate the feature importance; (ii) using the top 2, 3, 5, 10.100 (max) important features to build classification models, which were validated on the 1/3 the samples that were left out; and (iii) further validating the potential biomarkers using independent data sets. Finally, the ROC curve was generated in MetaboAnalyst 4.0 for evaluating the model performance by Monte-Carlo cross-validation (MCCV) using a balanced sub-sampling. We evaluated the model's performance by calculating the classification precision (ratio of true positive samples overall positive samples) and accuracy (ratio of true predictions overall predictions). The precision indicates the proportion of examples classified as actually positive samples overall positive samples. which is defined as

$$\textbf{Precision} = \frac{\textbf{TPi}}{(\textbf{TPi} + \textbf{FPi})}$$

and the accuracy denotes the number of correctly classified samples divided by the number of all samples, which is calculated as

$$\textbf{Accuracy} = \frac{(\textbf{TP} + \textbf{TN})}{(\textbf{TP} + \textbf{TN} + \textbf{FP} + \textbf{FN})}$$

where N is the number of samples. TP, TN, FP and FN stand for true positive, true negative, false positive and false negative rates, respectively.

2.8. Designing an early detection solution by pooled testing and accelerated growth incubator

Pooling subsamples and testing them in groups proved a useful strategy to reduce the cost and minimize test time (Mercer and Salit, 2021). Here we proposed a pooled testing strategy for different risk level samples to identify individuals contaminated with aflatoxigenic *Aspergillus* species in a small number of tests and few rounds of testing, which is shown in Fig. 5a. Once aflatoxins contamination risk was predicted by the macro-risk prediction model, geographic areas will be classified as regions at low-, medium-, or high-risk areas. Samples from high-risk areas were tested individually. In the medium- and low-risk areas, a 5:1 and 10:1 pooled sampling method was used to reduce the number of tests (Li et al., 2021). The first step was to group all samples. Then the samples within the group were mixed into a sample by taking equal parts. The first round of testing was performed on these pooled samples. For pooled samples that test positive, went back to the corresponding group, and tested the members in the group one by one. As we know, the *Aspergillus* species spores that were dormant in agro-products for days even months before the mycotoxin contamination. Minimizing the latent period will buy time for taking mycotoxin control measures. For this purpose, we designed an early detection of aflatoxigenic *Aspergillus* species workflow by detection of biomarkers and accelerated growth incubator. Firstly, we selected different tests strategy according to the risk level of the sample source (Fig. 5a). Before performing the test, the sample was incubated in an incubator until the filamentous fungi was observed. Sample preparation and biomarker assays were then performed according to Methods 2.2 and 2.3. The workflow is shown in Fig. 5b.

2.9. Development of an operable decision rule by XGBoost model

Despite the high predictive power of Random Forests, However, Random Forests model usually does not explain their predictions process. For this reason, we used the XGBoost software library (Chen and Carlos, 2016) couple with 180 actual peanut samples as a training dataset to train an XGBoost model (a state-of-the-art decision trees algorithms) and then obtained some single-XGBoost tree by simplifying the complexity of the model. XGBoost model was trained with the following parameters as described previously (Chen and Carlos, 2016; Yan et al., 2020): number of tree estimators set to 1, values of the regularization parameters α and β both set to 0. The detailed run process, code of the XGBoost model were available in (https://github.com/jeep3/AFbiosignature-file).

3. Results

3.1. Aflatoxins-producing capacity (APC) and metabolic diversity across A. flavus isolates

Out of the 568 *A. flavus* isolates in the dataset, 349 isolates had nontoxigenic or low aflatoxins-producing capacity (with APC of 0–1.0 mg/kg mycelia). Another 114 isolates were attributed to the medium APC group, whereas the remaining 87 and 18 isolates belonged to the high-APC and extremely high-APC groups, respectively (Fig. 2b). The medium and high APC strains account for 38.5%, which is the main source of aflatoxin menace in China. APC data of isolates from different regions were significantly different. 395 strains originated from the central areas, with 31.4% being high APC strains and 68.6% being low-APC or atoxigenic strains. 119 strains with varying virulence levels were selected from southern areas, with 52.1% and 47.9% of them being high and low-APC strains, respectively. The average APC in the southern region was significantly higher than that in the central region ($p = 0.0092$) and significantly higher than in the northern region ($p = 0.001$). The average APC of AFB_1 was significantly higher in the central region than in the northern region ($p = 0.00044$) (Fig. 2c1). This trend was also observed for aflatoxin B_2 (AFB_2) production by isolates from different regions (Fig. 2 c2). For other toxins simultaneously produced by *Aspergillus flavus*, the regional differences were relatively small for cyclopiazonic acid (CPA), a neurotoxin alkaloid produced by many *Aspergillus* or *Penicillium* species (Chang et al., 2009). The isolates from the northern region had the lowest levels of CPA, which were significantly different from those in the central and southern regions ($p = 0.041$ and 0.021, respectively) (Fig. 2c3). No clear regional differences in the level of kojic acid (KA) (Terabayashi et al., 2010) produced among the isolates were observed (Fig. 2c4).

We also found that the *A. flavus* population possesses rich secondary metabolic diversity within the species. The presence of subpopulation-specific metabolites can be observed from the heatmap (Fig. 2d). Enrichment analysis revealed that most primary metabolites are enriched in a variety of amino acid biosynthetic and metabolic pathways, such as D−glutamine and D−glutamate metabolism, histidine metabolism, lysine degradation, beta−alanine metabolism and phenylalanine, tyrosine, and tryptophan biosynthesis et.al pathways (Fig. 2e).

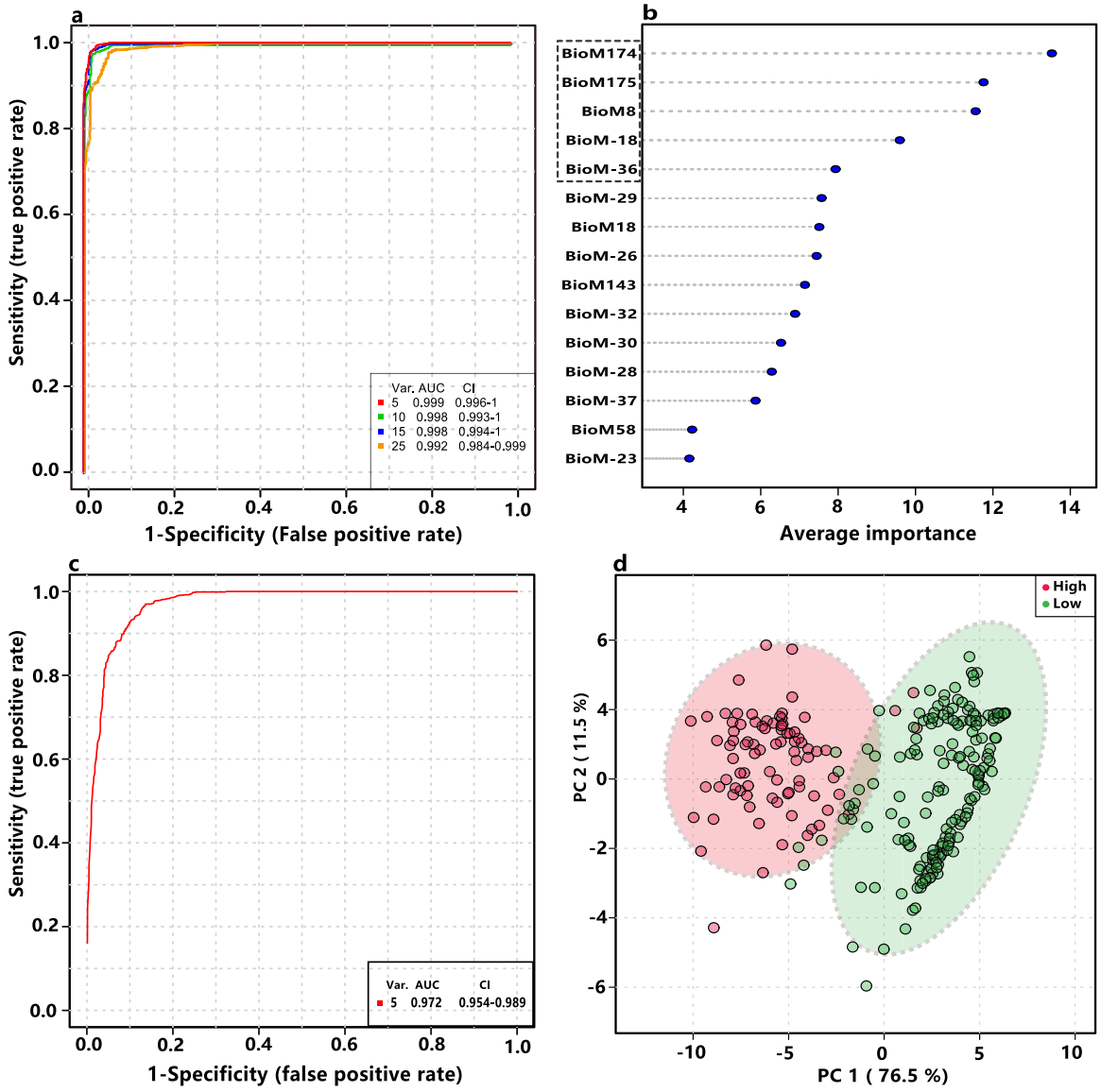

Fig. 3. Screening and validating biomarker of aflatoxigenic *A.flavus* via Random Forests algorithm. (a) Total receiver operating characteristic curve (ROC) of random forests algorithms in training dataset; (b) The list of top important metabolites identified in training dataset by Random Forests model; (c) ROC curve of random forests model was constructed by five biomarkers panel; d. Principal Component Analysis (PCA) of *A. flavus* strains belonging to the high and low aflatoxin-producing capacity groups.

Active primary metabolisms provided a rich substrates flux for secondary metabolism differentiation. The metabolic diversity underlies the basis for screening biomarkers of aflatoxigenic *Aspergillus* species.

3.2. Biomarkers analysis of aflatoxigenic Aspergillus species

Fig. 3a shown that the best classifier (based on AUC score) was the Random Forests model that was generated with the top five classification variables (AUC: 0.999, Cl: 0.996–1, Cl indicates the confidence intervals). Metabolites were ranked according to their importance. Fig. 3b shown that top five variables stemmed from BioM174 (unknown21), BioM8 (11-hydroxy O-methylsterigmatocystin (HOMST)), BioM175 (unknown22), BioM-18 (Versiconol), BioM-36 (Versicolorin_B). The predictive accuracy and precision with these top 5 biomarkers amounted to 97.6% and 97.1%, respectively. The performance of the model did not improve significantly in the area under the curve (AUC) scores when the number of metabolites was increased to one hundred (Fig. S5a). Hence, the top five biomarkers were selected as potential biomarkers panels in this study. The performance of this model was cross-validated (Fig. S5b), which revealed that an excellent classification performance by predicting class probabilities. Fig. S5b points out that the classification accuracy for the group of high-APC isolates was up to 97.6% (326/334). Therefore, this model was selected to perform the automated selection of important metabolites. To better evaluate the predictive power of the single biomarkers, classical univariate ROC curve analyses for individually selected biomarkers were also analyzed. The AUC score of five biomarkers ranged from 0.999 to 0.958 (Fig. S6).

3.3. Validation of potential biomarkers

Two hundred and thirty-four isolates were randomly selected from the library as an independent validation dataset to confirm the robustness of the potential biomarkers. Fig. 3c revealed that the Random Forest model achieved similar performance (AUC: 0.972, Cl: 0.954–0.989) using the validation dataset. This model was then used to again generate a list of important biomarkers (Fig. S5c), which appeared the same top 5 metabolites compare with the one produced with the training dataset, suggesting that this model captures the key biomarkers. The principal component analysis (PCA) demonstrated a clear separability for discerning the high-APC and low-APC strains, which is shown in Fig. 3d. The total contribution rate of the first and second principal components reached 88%. The above results had shown that these biomarkers possess high robustness and repeatability. Fig. S5d shown that the predictive accuracy and precision of the Random Forests classifiers reached 97.8% and 95.4%, respectively. Finally, we demonstrated the reliability of biomarkers structural annotation by the mass spectrometry information of the biomarkers (Table 1), the mirror plot of MS/MS mass spectra by matching standards, and samples (Fig. S7) and the MS/MS fragmentation trees (Fig. S8).

3.4. The sensitivity and specificity of biomarkers

The sensitivity and specificity of the biomarkers are two important aspects that must be examined for biomarker analysis methods. The sensitivity of biomarkers was measured via the parameters of linear range, the limit of detection (LOD), the limit of quantification (LOQ), precision, and specificity. The results from Table S1 shown that LOD and LOQ ranged from 0.003 to 0.20 and 0.012−0.50 μg/mg(mycelia), respectively. Good linearity was observed with correlation coefficients (R2) higher than 0.9993 for all biomarkers, which is shown in Fig. S9. The intra- and inter-day precisions were 0.02−0.56 and 0.06−0.44 (%), respectively. The results demonstrated that this method was robust and highly sensitive for detecting biomarkers. The specificity of these biomarkers was confirmed by checking the presence/absence of biomarkers in different rhizosphere soil fungal species metabolome. As we expected, the biomarkers were found in *Aspergillus parasiticus* and *Aspergillus tamarii* (two known aflatoxin-producing fungi), but not found in non-aflatoxigenic *A. flavus* or other rhizosphere soil fungal species, which is shown in Table 2. The above results revealed that these biomarkers are specific for the aflatoxigenic *Aspergillus* species and had highly sensitive.

3.5. The correlation between biomarkers and APC

Fig. 4a revealed that positively correlated compounds with APC had higher correlation coefficients and lower correlation coefficients for negatively correlated metabolites. Speradine A is derivatives of cyclopiazonic acid (CPA), which was shown a negative correlation with APC (Pearson's $r = -0.35$), which was shown in Fig. 4b. Chrysophanic acid, an anthraquinone metabolites, shown a negative correlation coefficient with APC (Pearson's $r = -0.48$), which was shown in Fig. 4c. Conversely, the correlation coefficient of unkown21 with the APC was 0.95 and those for the other four biomarkers ranged from 0.92 to 0.71, which was shown in Fig. 4d-i, respectively. The correlation coefficient between the average value of the five biomarkers and the APC reached 0.87 (Fig. 4i). We finally selected BioM174 (unknown21), BioM175 (unknown22), BioM8 (11-hydroxy O-methylsterigmatocystin (HOMST)), BioM-18 (Versiconol), BioM-36 (Versicolorin_B) as APC-associated biomarkers panel based on the result above. BioM-18 (Versiconol) and BioM-36 (Versicolorin_B) are early intermediates in aflatoxin biosynthesis and are thus considered to be aflatoxins contamination severity (ACS)-associated metabolites.

3.6. The structural relationship between biomarkers and aflatoxins

A global molecular networking (Wang et al., 2016) map of the detectable metabolites visualized in Fig. S10a and a partially enlarged schematic diagram (Fig. S10b) shows the aflatoxins molecular family network that contains 26 metabolite features. We found that the structural similarity of aflatoxin B_1(node N1), aflatoxin B_2(N2), aflatoxin G_1(N3), aflatoxin G_2(N4), sterigmatocystin(N5), and 11-hydroxy O-methylsterigmatocystin (HOMST)(N6) was reflected by a high spectral similarity (cosine value > 0.8). The actual structures of aflatoxin

Table 1
UPLC-HRMS information of biomarkers and aflatoxin B_1.

Name	Molecular formula	Ion mode	Accurate mass		Mass deviation(ppm)
			Theoretical	Experimental	
Aflatoxin B_1	$C_{17}H_{12}O_6$	$[M+H]^+$	313.07066	313.07012	1.73
HOMST	$C_{19}H_{14}O_7$	$[M+H]^+$	355.08122	355.08068	1.52
unknown21	$C_{34}H_{24}O_{12}$	$[M+Na]^+$	647.11147	647.11479	5.13
unknown22	$C_{21}H_{12}O_7$	$[M+H]^+$	377.06289	377.06218	-1.88
Versicolorin B	$C_{18}H_{12}O_7$	$[M-H]^-$	339.05102	339.05096	0.17
Versiconol	$C_{18}H_{16}O_8$	$[M-H]^-$	359.07724	359.07733	0.25

HOMST: 11-hydroxy O-methylsterigmatocystin.

Table 2
Specificity assessment by 16 different fungal strains isolated from rhizosphere soil.

Strains	Species	HOMST	Unknown21	Unknown22	Versiconol	Versicolorin B
AnHHF-33	*Aspergillus oryzae*	–	–	–	–	–
HeBHD-1	*Aspergillus oryzae*	–	–	–	–	–
HuBLT-3	*Aspergillus oryzae*	–	–	–	–	–
CJ-3–3	*Aspergillus ochraceus*	–	–	–	–	–
BNCC336184	*Aspergillus ochraceus*	–	–	–	–	–
HuBzhx-43	*Aspergillus fumigatus*	–	–	–	–	–
LNCT-4	*Aspergillus parasiticus*	+	+	+	+	+
AHTL-15	*Aspergillus tamarii*	+	+	+	+	+
BNCC340687	*Fusarium moniliforme*	–	–	–	–	–
FJQ2H-4	*Rhizopus oryzae*	–	–	–	–	–
GDHZ-1	*Pichia guilliermondii*	–	–	–	–	–
GXWM-1	*Penicillium janthinellum*	–	–	–	–	–
D83	*Trichoderma* spp.	–	–	–	–	–
XZ-2–9	*Trichoderma* spp.	–	–	–	–	–
XZ-11–5	*Trichoderma* spp.	–	–	–	–	–
BNCC143078	*Fusarium oxysporum*	–	–	–	–	–

- Absence + Presence.

B_1(node N1), aflatoxin B_2(N2), aflatoxin G_1(N3), aflatoxin G_2(N4), sterigmatocystin(N5), 11-hydroxy O-methylsterigmatocystin (HOMST) (N6) were shown in Fig. S10b. Interestingly, biomarkers of $m/z = 647.11147$ (unknown21) and $m/z = 377.06289$ (unknown22), unidentified metabolites, were also members of this molecular family. Unexpectedly, some precursors in the aflatoxin biosynthesis pathway, such as versicolorin B (m/z 359.0767), did not appear in this molecular family, most likely because these compounds had a low response intensity in positive ion mode and were more suitable for mass spectrometry data acquisition in negative ion mode. In brief, the structural relationships between aflatoxins and biomarkers were to some extent explained by feature-based molecular networking analysis (Nothias et al., 2020).

3.7. An early detection solution for predicting aflatoxin contamination risk

We designed an early detection solution by pooled testing and accelerated growth incubator, which is shown in Fig. 5a,b. 426 peanut samples sampling from nationwide mycotoxins monitoring programs were checked by a thermostatic incubator. We found that out of 86 suspected samples being contaminated by fungi. Of these, we discovered that 39 suspected samples were contaminated with aflatoxigenic *Aspergillus* species and the other 47 samples were infected by non-aflatoxigenic fungi, which is shown in Fig. 5c. 97.4% positive samples were accurately classified by early detecting the APC-associated biomarkers. To quantify the early-warning time window, peanut samples contaminated with aflatoxigenic *Aspergillus* species and non-contaminated samples were cultured in an incubator (1–13 days) and warehouse (1–60 days) to obtain a series of time-course samples. Phenotype change over time of aflatoxigenic *Aspergillus* species contaminated and non-contaminated samples in an incubator were shown in Fig. S11. Fig. 5d shown that the timeline of aflatoxins contamination risk in the warehouse (upper section) and the thermostatic incubators (bottom section). In the incubator environment, the biomarker of Versicolorin B was detected on day 5, when the concentration of aflatoxin B_1 was still very low and the peanut samples appeared clean without visible fungal mycelia (Fig. S11). On day 7, the aflatoxin concentration exceeded the maximum allowable threshold of 5 μg/kg, and the Versicolorin B concentration at this moment was 13.8 times that of aflatoxin B_1(Fig. S12). The early warning window is longer in the warehouse, which provides 35–47 days to take management actions in advance if high levels of versicolorin B are found in early detection (Fig. 5d).

3.8. An operable decision rule intuitively discerns aflatoxigenic Aspergillus species

The resulting performance of the XGBoost model is shown, respectively, in Table S2 and Figs. S13–S19. 11-hydroxy O-methylsterigmatocystin (HOMST) and versicolorin_B were selected to generate a single-tree XGBoost model based on the performance of various single-tree XGBoost models generated. These two trees (Figs. S14 and S15) were further trimmed to simplify the decision-making process to generate a practical operable decision tree (Fig. 6) with a slight reduction in accuracy from 95.4% to 87.2%. The decision rule used two key features, and their thresholds of 11-hydroxy O-methylsterigmatocystin (HOMST) > 34.7 μg/kg and versicolorin_B > 96.35 μg/kg were computationally selected and validated using the training and test datasets, which is shown in Fig. 6. The overall accuracy rate of using this decision tree reached 87.2%. We divided the data into subcategories, the prediction accuracy of low-APC aflatoxigenic *Aspergillus* species samples reached 94.9% and the prediction rate of high-APC aflatoxigenic *Aspergillus* species samples reached 100%. However, the prediction accuracy rate of medium-APC aflatoxigenic *Aspergillus* species was only 39.5%. Therefore, in the actual application process, the prediction accuracy of low- and high-APC samples were very high, however, further assessment is still required for medium-APC samples.

4. Discussion

Mycotoxins pose a global food safety and public health threat, with a worsening trend under global warming (Battilani et al., 2016). Early detection proved an effective strategy to control and mitigate this issue (Jiang et al., 2019). We presented a roadmap from discovering biomarkers to design early detection solutions using population metabolomics and machine learning. Novel biomarkers of aflatoxigenic *Aspergillus* species performed better than previously reported (Fig. S20a, b). The early detection solution designed by pooled testing and accelerated cultivating reduced the cost, test time, and maximized warning time compared to previous researchers who directly detected samples in the warehouse (Jiang et al., 2019). Our study confers regulatory bodies with the ability to smart govern the mycotoxin contamination risk and to timely take measures to inhibit fungal growth or discard samples to safeguard consumer health, while reducing economic losses.

The survey data revealed that 38.5% of all isolates harbor medium or high aflatoxin-producing capacity, which is the main source of aflatoxin menace, especially in the south and central area of China (Fig. 2b,c). Therefore, early detection of aflatoxigenic *Aspergillus* species together with effective monitoring of moisture content (e.g. <15% maize; <7.5% peanuts) et al. measures before entering the warehouse, that plays a crucial role in mycotoxin risk assessment. The Liu's group predicted the risk of aflatoxin contamination 18 days (mean value) in advance by detection of versicolorin A concentration in maize at warehouse condition (Jiang et al., 2019). It was worth noting that versicolorin A

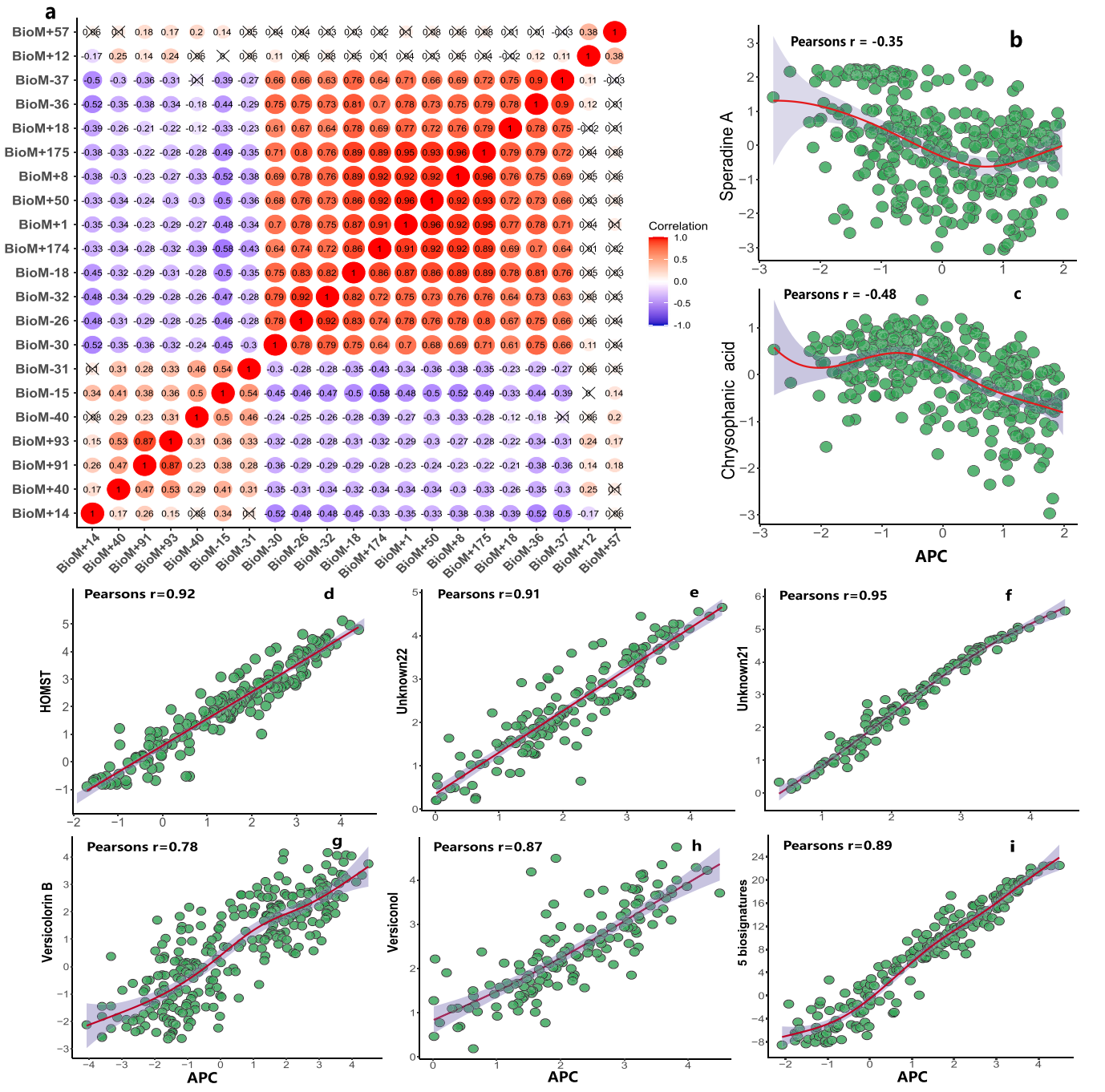

Fig. 4. Correlation analysis of the identified biomarkers and the aflatoxins-producing capacity (APC) of the individual isolates. (a) Pearson correlation analysis results of different compounds. (b-i) visualizing the correlation of speradine A, chrysophanic_acid, 11-hydroxy O-methylsterigmatocystin (HOMST), unknown22, unknown21, versicolorin_B, versiconol, and 5 biomarkers average values with the APC by scatter plot, respectively.

(bioM-37) also had predictive power in this study (Fig. 3b), which proved the powerfulness of our approach. However, Versicolorin A is the first intermediate metabolite in the aflatoxin biosynthetic pathway with a furofuran (bisfuran) structure, whose chemical structure shares many features with AFB_1. Recent research reported that Ver A was a food contaminant toxic to human intestinal cells (Gauthier et al., 2020). VerB and versiconol lack the furofuran (bisfuran) C=C double bond structure and are more upstream in aflatoxin biosynthesis, which implied that their toxicity was lower (Budin et al., 2021). In addition, we found that the correlation between VerB and APC was 0.78, which was higher than that between VerA and APC (0.71) (Fig. S20a,b). These biomarker assessment results suggested that VerB and versiconol are more suitable as (ACS)-associated biomarkers. 11-hydroxy O-methylsterigmatocystin (HOMST) was structurally closely related to sterigmatocystin (ST). Metabolomics data revealed that the abundance of HOMST in the majority of *A.flavus* isolates was higher than that of ST. We also observed a higher correlation between HOMST (Pearson's = 0.92) and APC than for ST (Pearson's = 0.43) (Fig. S20c, d). In addition, Feature-based molecular network analysis revealed that some new metabolites, such as $m/z = 341.0663$ and $m/z = 657.1323$, were aflatoxin-like metabolites, as they grouped in the same molecular family with them. The results elucidated that those of aflatoxin-like metabolites as emerging co-occurrence mycotoxins should be brought to the attention of risk regulators.

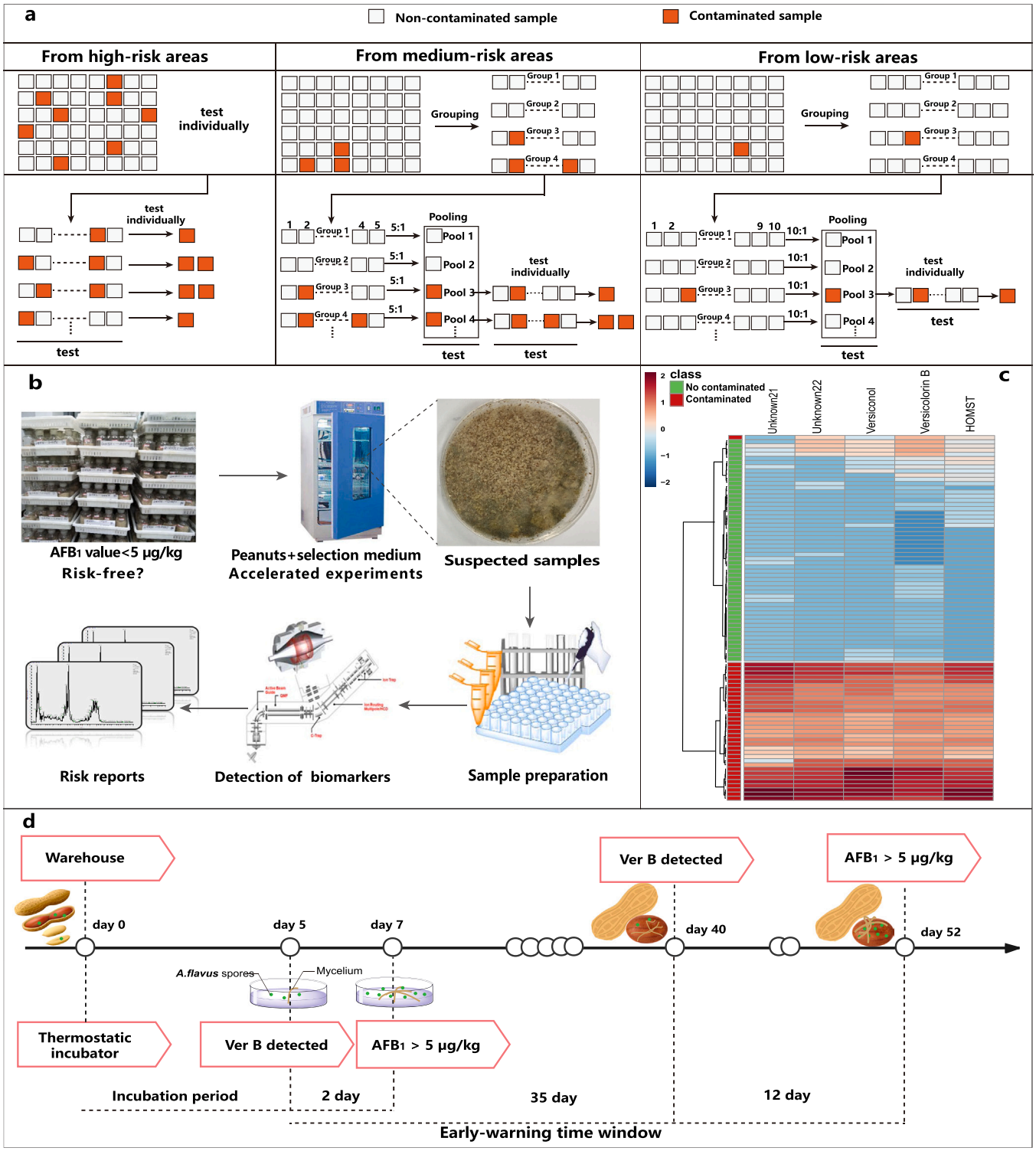

Fig. 5. (a) A pooled testing strategy for identifying contaminated samples by aflatoxigenic *Aspergillus* species. (b) A recommended workflow of early detection by incubator accelerated growth and detection of biomarkers. (c) A heatmaps that show contaminated and non-contaminated samples by aflatoxigenic *Aspergillus* species in 86 suspected samples. (d) Timeline of aflatoxins occurrence risk in warehouse (upper section) and timeline of early detection of aflatoxigenic *Aspergillus* species in thermostatic incubators (bottom section).

To quantify the warning time, we compared the time course difference of aflatoxins and biomarkers production in warehouse and incubator conditions based on a proposed early detection solution. The entire early detection process was completed within 6–8 days from hidden fungi grow in an incubator (5–7 days), to sample pre-treatment and UPLC-HRMS analysis time(6–8 h) per 10 samples. Ver B was detected two days earlier than AFB_1 under incubator conditions. In contrast, Ver B was first detected under warehouse conditions with a lag of 35 days. After 12 days, we found that AFB_1 exceeded the maximum acceptable level. In terms of time spent on analysis methods, our method takes 6–8 h from metabolites extraction to mass spectrometry analysis per 10 samples. Gene marker-based approach takes 4–5 h from DNA

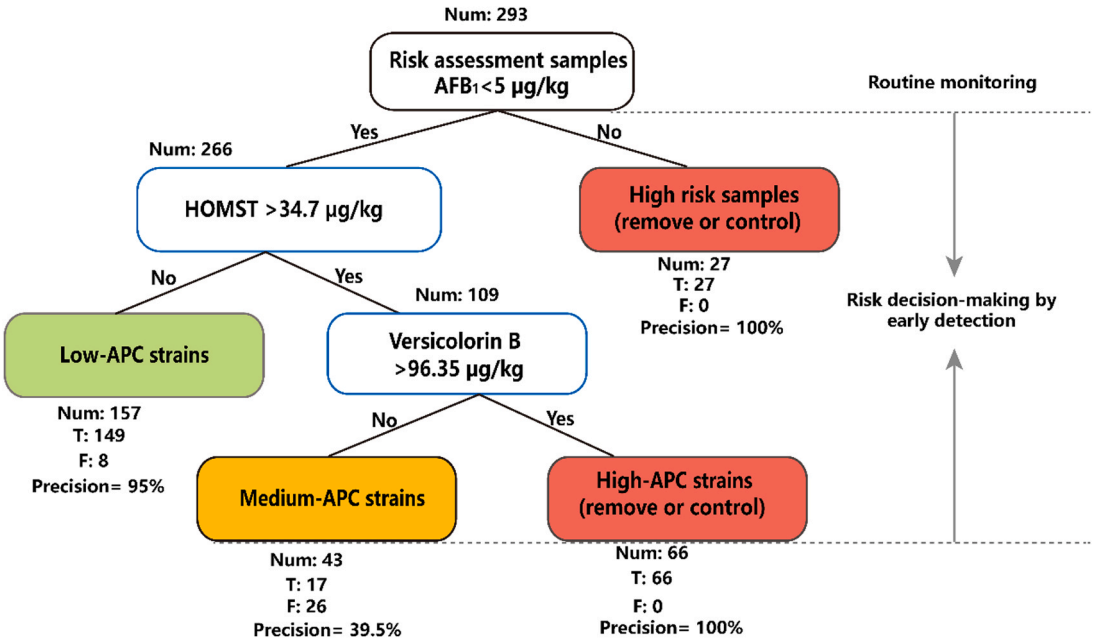

Fig. 6. A decision rule using two key features and their thresholds. Num, the number of samples; T, the number of correctly classified samples; F, the number of misclassified samples.

extraction to PCR test per 10 samples (Ren et al., 2020). The protein marker-based early detection method takes 18–24 h from protein extraction to immunoassay test per 10 samples (B. Wang et al., 2017; T. Wang et al., 2017). Despite the speed of nucleic acid sequence-based detection methods, no aflatoxigenic *Aspergillus* species-specific sequences at the subspecies level are currently available. The good news is that the development of rapid detection methods for testing small molecules at a comparable speed is not difficult to achieve. In terms of warning time windows, the incubator accelerated the germination and growth of aflatoxigenic *Aspergillus* species spores that are dormant in the agro-products. If we use the incubator as an accelerator, this warning time window can be increased to more than 35 days (Fig. 5d). Liu's group reported that the time interval (warning time windows) between detection of VerA and aflatoxin contamination was 18 days (mean value) in the warehouse (Jiang et al., 2019). Hence, our proposed solution possesses a longer warning time window and saves more time on mycotoxin risk assessment compared to directly test biomarkers of aflatoxigenic *Aspergillus* species in warehouses. Because the accelerated incubator accelerates the production of mycotoxins. It should be noted that the warning time window was influenced by the moisture content of the agro-product, the temperature of the warehouse, water activity et al. environmental factors.

Some limitations in this study warrant a discussion. The current use of high-resolution mass spectrometry is expensive, which hinders the widespread use of this method. For widely applying this risk assessment strategy, it should be necessary to develop fast, simpler, and cheaper analysis tools to monitor the biomarkers, for example, rapid immunoassay methods (Matabaro et al., 2017), biosensors (Xu et al., 2018), and low-resolution mass spectrometry methods (Medina et al., 2021). In addition, climate change has profoundly affected many aspects of agricultural production, including food safety (Medina et al., 2015). It was difficult to predict the risk of mycotoxin contamination with high accuracy using a technique. We envisage a future where we can integrate macroclimate data and micro biomarkers of aflatoxigenic *Aspergillus* species data, mycotoxins level, moisture content of agricultural products et al. factors as variables to construct more accurate early warning models using artificial intelligence.

5. Conclusion

In summary, metabolic diversities of a natural collection of the *A. flavus* population were systematically delineated using high-resolution mass spectrometry. We discovered five novel biomarkers associated with aflatoxigenic *Aspergillus* species by population metabolomics and machine learning. We demonstrated that these biomarkers discern aflatoxigenic *Aspergillus* species with 97.8% accuracy in the early stages of aflatoxin contamination and predicted the onset of aflatoxin contamination severity 35–47 days in advance using our proposed solution. An operable decision workflow developed by a state-of-the-art XGBoost algorithm provided an intuitive test to quantify the risk prioritization. The study provided a novel insight for risk assessment of the agricultural product to ensure food safety and public health.

Author contributions

The idea was initiated and designed experiments by Prof. Qi Zhang and Huali Xie, who prepared the initial draft and final version. Huali Xie conducted experiments and analyzed the data. Prof. Xiupin Wang provided the resources for mass spectrometry analysis platform; Prof. Justin JJ van der Hooft and Prof. Marnix H Medema provided guidance for molecular network analysis and revised the manuscript. Prof. Zhi-Yuan Chen revised the manuscript and we also appreciate the helpful discussions with him. Dr. Xiaofeng Yue contributed materials. Prof. Qi Zhang and Prof. Peiwu Li proof read and edited the manuscript. All authors approved the final manuscript.

Declaration of Competing Interest

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

Data Availability

The *A. flavus isolates metadata, metabolomics data matrix, including the training dataset and the independent validation dataset for screening and validating the biomarkers were summarized in a Supplementary sheet.*

Acknowledgments

We thank Qiuyu Yu and Tao Li for assistance with storage of the strains used in this work, as well as for helpful discussions in data analysis. This work was supported by the project of National Key Research and Development Program of China (2017YFC1601200), China, the project of National Natural Sciences Foundation of China (32030085), China, the project of Agricultural Science and Technology Innovation Program of CAAS (CAAS-ZDRW202011), China, and the project of Major Project of Hubei Provincial Technical Innovation (2018ABA081), China.

Supporting information

Supporting Information is available from the Supplementary information or Supplementary sheet.

References

Adebo, O.A., Njobeh, P.B., Gbashi, S., Nwinyi, O.C., Mavumengwana, V., 2017. Review on microbial degradation of aflatoxins. Crit. Rev. Food Sci. Nutr. 57, 3208–3217.

Atehnkeng, J., Ojiambo, P.S., Cotty, P.J., Bandyopadhyay, R., 2014. Field efficacy of a mixture of atoxigenic *Aspergillus flavus* Link:Fr vegetative compatibility groups in preventing aflatoxin contamination in maize (Zea mays L.). Biol. Control 72, 62–70.

Battilani, P., Toscano, P., Van der Fels-Klerx, H.J., Moretti, A., Camardo Leggieri, M., Brera, C., Rortais, A., Goumperis, T., Robinson, T., 2016. Aflatoxin B1 contamination in maize in Europe increases due to climate change. Sci. Rep. 6, 24328.

Budin, C., Man, H.Y., Al-Ayoubi, C., Puel, S., van Vugt-Lussenburg, B.M.A., Brouwer, A., Oswald, I.P., van der Burg, B., Soler, L., 2021. Versicolorin A enhances the genotoxicity of Aflatoxin B1 in human liver cells by inducing the transactivation of the Ah-Receptor. Food Chem. Toxicol., 112258

Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T.A., Brusniak, M.Y., Paulse, C., Creasy, D., Flashner, L., Kani, K., Moulding, C., Seymour, S.L., Nuwaysir, L.M., Lefebvre, B., Kuhlmann, F., Roark, J., Rainer, P., Detlev, S., Hemenway, T., Huhmer, A., Langridge, J., Connolly, B., Chadick, T., Holly, K., Eckels, J., Deutsch, E.W., Moritz, R.L., Katz, J.E., Agus, D.B., MacCoss, M., Tabb, D.L., Mallick, P., 2012. A cross-platform toolkit for mass spectrometry and proteomics. Nat. Biotechnol. 30, 918–920.

Chang, P.K., Ehrlich, K.C., Fujii, I., 2009. Cyclopiazonic acid biosynthesis of *Aspergillus flavus* and *Aspergillus oryzae*. Toxins 1, 74–99.

Chawanthayatham, S., Valentine 3rd, C.C., Fedeles, B.I., Fox, E.J., Loeb, L.A., Levine, S. S., Slocum, S.L., Wogan, G.N., Croy, R.G., Essigmann, J.M., 2017. Mutational spectra of aflatoxin B1 in vivo establish biomarkers of exposure for human hepatocellular carcinoma. Proc. Natl. Acad. Sci. U. S. A. 114, E3101–E3109.

Chen, T., Carlos G., 2016, XGBoost: a scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D.S., Xia, J., 2018. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res. 46, W486–W494.

Dovenyi-Nagy, T., Racz, C., Molnar, K., Bako, K., Szlama, Z., Jozwiak, A., Farkas, Z., Pocsi, I., Dobos, A.C., 2020. Pre-harvest modelling and mitigation of aflatoxins in maize in a changing climatic environment – a review. Toxins 12.

Garcia-Cela, E., Sanchez, F.J.G., Sulyok, M., Verheecke-Vaessen, C., Medina, A., Krska, R., Magan, N., 2020. Carbon dioxide production as an indicator of *Aspergillus flavus* colonisation and aflatoxins/cyclopiazonic acid contamination in shelled peanuts stored under different interacting abiotic factors. Fungal Biol. -Uk 124, 1–7.

Gauthier, T., Duarte-Hospital, C., Vignard, J., Boutet-Robinet, E., Sulyok, M., Snini, S.P., Alassane-Kpembi, I., Lippi, Y., Puel, S., Oswald, I.P., Puel, O., Versicolorin, A., 2020. A precursor in aflatoxins biosynthesis, is a food contaminant toxic for human intestinal cells. Environ. Int. 137, 105568.

Hedayati, M.T., Pasqualotto, A.C., Warn, P.A., Bowyer, P., Denning, D.W., 2007. *Aspergillus flavus*: human pathogen, allergen and mycotoxin producer. Microbiology (Reading) 153, 1677–1692.

Hu, Y.X., Cai, B., Huan, T., 2019. Enhancing metabolome coverage in data-dependent LC-MS/MS analysis through an integrated feature extraction strategy. Anal. Chem. 91, 14433–14441.

Huang, M.N., Yu, W., Teoh, W.W., Ardin, M., Jusakul, A., Ng, A.W.T., Boot, A., Abedi-Ardekani, B., Villar, S., Myint, S.S., Othman, R., Poon, S.L., Heguy, A., Olivier, M., Hollstein, M., Tan, P., Teh, B.T., Sabapathy, K., Zavadil, J., Rozen, S.G., 2017. Genome-scale mutational signatures of aflatoxin in cells, mice, and human tumors. Genome Res. 27, 1475–1486.

Hyde, K.D., Al-Hatmi, A.M.S., Andersen, B., Boekhout, T., Buzina, W., Dawson, T.L., Eastwood, D.C., Jones, E.B.G., de Hoog, S., Kang, Y., Longcore, J.E., McKenzie, E.H. C., Meis, J.F., Pinson-Gadais, L., Rathnayaka, A.R., Richard-Forget, F., Stadler, M., Theelen, B., Thongbai, B., Tsui, C.K.M., 2018. The world's ten most feared fungi. Fungal Divers. 93, 161–194.

Jiang, M.P., Zheng, S.Y., Wang, H., Zhang, S.Y., Yao, D.S., Xie, C.F., Liu, D.L., 2019. Predictive model of aflatoxin contamination risk associated with granary-stored corn with versicolorin A monitoring and logistic regression. Food Addit. Contam. Part A Chem. Anal. Control Expo. Risk Assess. 36, 308–319.

Keller, N.P., 2019. Fungal secondary metabolism: regulation, function and drug discovery. Nat. Rev. Microbiol. 17, 167–180.

Li, Z.J., Liu, F.F., Cui, J.Z., Peng, Z.B., Chang, Z.R., Lai, S.J., Chen, Q.L., Wang, L.P., Gao, G.F., Feng, Z.J., 2021. Comprehensive large-scale nucleic acid-testing strategies support China's sustained containment of COVID-19. Nat. Med. 27, 740–742.

Matabaro, E., Ishimwe, N., Uwimbabazi, E., Lee, B.H., 2017. Current immunoassay methods for the rapid detection of aflatoxin in milk and dairy products. Compr. Rev. Food Sci. F 16, 808–820.

Medina, Á., Rodríguez, A., Magan, N., 2015. Climate change and mycotoxigenic fungi: impacts on mycotoxin production. Curr. Opin. Food Sci. 5, 99–104.

Medina, D.A.V., Borsatto, J.V.B., Maciel, E.V.S., Lancas, F.M., 2021. Current role of modern chromatography and mass spectrometry in the analysis of mycotoxins in food. Trac-Trend Anal. Chem. 135.

Mercer, T.R., Salit, M., 2021. Testing at scale during the COVID-19 pandemic. Nat. Rev. Genet. 22, 415–426.

Mitchell, N.J., Bowers, E., Hurburgh, C., Wu, F., 2016. Potential economic losses to the US corn industry from aflatoxin contamination. Food Addit. Contam. A 33, 540–550.

Ndemera, M., De Boevre, M., De, S., 2018. Saeger, Mycotoxin management in a developing country context: a critical review of strategies aimed at decreasing dietary exposure to mycotoxins in Zimbabwe. Crit. Rev. Food Sci. Nutr. 60, 529–540.

Nothias, L.F., Petras, D., Schmid, R., Duhrkop, K., Rainer, J., Sarvepalli, A., Protsyuk, I., Ernst, M., Tsugawa, H., Fleischauer, M., Aicheler, F., Aksenov, A.A., Alka, O., Allard, P.M., Barsch, A., Cachet, X., Caraballo-Rodriguez, A.M., Da Silva, R.R., Dang, T., Garg, N., Gauglitz, J.M., Gurevich, A., Isaac, G., Jarmusch, A.K., Kamenik, Z., Kang, K.B., Kessler, N., Koester, I., Korf, A., Le Gouellec, A., Ludwig, M., Martin, H.C., McCall, L.I., McSayles, J., Meyer, S.W., Mohimani, H., Morsy, M., Moyne, O., Neumann, S., Neuweger, H., Nguyen, N.H., Nothias-Esposito, M., Paolini, J., Phelan, V.V., Pluskal, T., Quinn, R.A., Rogers, S., Shrestha, B., Tripathi, A., van der Hooft, J.J.J., Vargas, F., Weldon, K.C., Witting, M., Yang, H.J., Zhang, Z., Zubeil, F., Kohlbacher, O., Bocker, S., Alexandrov, T., Bandeira, N., Wang, M.X., Dorrestein, P.C., 2020. Feature-based molecular networking in the GNPS analysis environment. Nature Methods 17, 905–908.

Peromingo, B., Rodriguez, M., Delgado, J., Andrade, M.J., Rodriguez, A., 2017. Gene expression as a good indicator of aflatoxin contamination in dry-cured ham. Food Microbiol. 67, 31–40.

Pietsch, C., Müller, G., Mourabit, S., Carnal, S., Bandara, K., 2020. Occurrence of fungi and fungal toxins in fish feed during storage. Toxins 12, 171.

Pluskal, T., Castillo, S., Villar-Briones, A., Oresic, M., 2010. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. Bmc Bioinforma. 11.

Ren, X.F., Yue, X.F., Mwakinyali, S.E., Zhang, W., Zhang, Q., Li, P.W., 2020. Small molecular contaminant and microorganism can be simultaneously detected based on nanobody-phage: using carcinogen aflatoxin and its main fungal *Aspergillus* Section Flavi spp. in stored maize for demonstration. Front. Microbiol. 10.

Roze, L.V., Hong, S.Y., Linz, J.E., 2013. Aflatoxin biosynthesis: current frontiers. Annu. Rev. Food Sci. Technol. 4, 293–311.

Saldan, N.C., Almeida, R.T.R., Avincola, A., Porto, C., Galuch, M.B., Magon, T.F.S., Pilau, E.J., Svidzinski, T.I.E., Oliveira, C.C., 2018. Development of an analytical method for identification of *Aspergillus flavus* based on chemical markers using HPLC-MS. Food Chem. 241, 113–121.

Sharma, K.K., Pothana, A., Prasad, K., Shah, D., Kaur, J., Bhatnagar, D., Chen, Z.-Y., Raruang, Y., Cary, J.W., Rajasekaran, K., Sudini, H.K., Bhatnagar-Mathur, P., 2018. Peanuts that keep aflatoxin at bay: a threshold that matters. Plant Biotechnol. J. 16, 1024–1033.

Terabayashi, Y., Sano, M., Yamane, N., Marui, J., Tamano, K., Sagara, J., Dohmoto, M., Oda, K., Ohshima, E., Tachibana, K., Higa, Y., Ohashi, S., Koike, H., Machida, M., 2010. Identification and characterization of genes responsible for biosynthesis of kojic acid, an industrially important compound from *Aspergillus oryzae*. Fungal Genet. Biol. 47, 953–961.

Wang, M.X., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D., Watrous, J., Kapono, C.A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M.J., Liu, W.T., Criisemann, M., Boudreau, P.D., Esquenazi, E., Sandoval-Calderon, M., Kersten, R.D., Pace, L.A., Quinn, R.A., Duncan, K.R., Hsu, C. C., Floros, D.J., Gavilan, R.G., Kleigrewe, K., Northen, T., Dutton, R.J., Parrot, D., Carlson, E.E., Aigle, B., Michelsen, C.F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B.T., Gerwick, L., Liaw, C.C., Yang, Y.L., Humpf, H.U., Maansson, M., Keyzers, R.A., Sims, A.C., Johnson, A.R., Sidebottom, A. M., Sedio, B.E., Klitgaard, A., Larson, C.B., Boya, C.A., Torres-Mendoza, D., Gonzalez, D.J., Silva, D.B., Marques, L.M., Demarque, D.P., Pociute, E., O'Neill, E.C., Briand, E., Helfrich, E.J.N., Granatosky, E.A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J.J., Zeng, Y., Vorholt, J.A., Kurita, K.L., Charusanti, P., McPhail, K.L., Nielsen, K.F., Vuong, L., Elfeki, M., Traxler, M.F., Engene, N., Koyama, N., Vining, O.B., Baric, R., Silva, R.R., Mascuch, S.J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P.G., Dai, J.Q., Neupane, R., Gurr, J., Rodriguez, A.M.C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B.M., Almaliti, J., Allard, P.M., Phapale, P., Nothias, L.F., Alexandrovr, T., Litaudon, M., Wolfender, J.L., Kyle, J.E., Metz, T.O., Peryea, T., Nguyen, D.T., VanLeer, D., Shinn, P., Jadhav, A., Muller, R., Waters, K.M., Shi, W.Y., Liu, X.T., Zhang, L.X., Knight, R., Jensen, P.R., Palsson, B.O., Pogliano, K., Linington, R.G., Gutierrez, M., Lopes, N.P., Gerwick, W.H., Moore, B.S., Dorrestein, P.C., Bandeira, N., 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol 34, 828–837.

Wang, B., Han, X.Y., Bai, Y.H., Lin, Z.G., Qiu, M.G., Nie, X.Y., Wang, S., Zhang, F., Zhuang, Z.H., Yuan, J., Wang, S.H., 2017. Effects of nitrogen metabolism on growth and aflatoxin biosynthesis in *Aspergillus flavus*. J. Hazard. Mater. 324, 691–700.

Wang, T., Li, P., Zhang, Q., Zhang, W., Zhang, Z., Wang, T., He, T., 2017. Determination of *Aspergillus* pathogens in agricultural products by a specific nanobody-polyclonal antibody sandwich ELISA. Sci. Rep. 7, 4348.

Xie, H.L., Wang, X.P., Zhang, L.X., Wang, T., Zhang, W., Jiang, J., Chang, P.K., Chen, Z. Y., Bhatnagar, D., Zhang, Q., Li, P.W., 2018. Monitoring metabolite production of aflatoxin biosynthesis by orbitrap fusion mass spectrometry and a D-optimal mixture design method. Anal. Chem. 90, 14331–14338.

Xie, H.L., Jallow, A., Yue, X.F., Wang, X.P., Fu, J.Y., Mwakinyali, S.E., Zhang, Q., Li, P. W., 2021. *Aspergillus* flavus's response to antagonism bacterial stress sheds light on a regulation and metabolic trade-off mechanism for adversity survival. J. Agric. Food Chem. 69, 4840–4848.

Xu, L., Zhang, H.Q., Yan, X.W., Peng, H.Y., Wang, Z.X., Zhang, Q., Li, P.W., Zhang, Z.W., Le, X.C., 2018. Binding-induced DNA dissociation assay for small molecules: sensing aflatoxin B1. ACS Sens. 3, 2590–2596.

Yan, L., Zhang, H.-T., Goncalves, J., Xiao, Y., Wang, M., Guo, Y., Sun, C., Tang, X., Jing, L., Zhang, M., Huang, X., Xiao, Y., Cao, H., Chen, Y., Ren, T., Wang, F., Xiao, Y., Huang, S., Tan, X., Huang, N., Jiao, B., Cheng, C., Zhang, Y., Luo, A., Mombaerts, L., Jin, J., Cao, Z., Li, S., Xu, H., Yuan, Y., 2020. An interpretable mortality prediction model for COVID-19 patients. Nat. Mach. Intell. 2, 283–288.