

1 **Integration of hyperspectral imaging, non-targeted metabolomics**  
2 **and machine learning for vigour prediction of naturally and**  
3 **accelerated aged sweetcorn seeds**

4

5

6 Tingting Zhang <sup>a, 1, \*</sup>, Long Lu <sup>a, 1</sup>, Ni Yang <sup>c, 1</sup>, Ian D. Fisk <sup>c, d</sup>, Wensong Wei <sup>e</sup>, Li Wang <sup>a</sup>, Jing  
7 Li <sup>a</sup>, Qun Sun <sup>b, \*</sup>, Rensen Zeng <sup>a, \*</sup>

8

9

10

11

12

13 <sup>a</sup> Key Laboratory of Ministry of Education for Genetics, Breeding and Multiple Utilization of  
14 Crops, College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou, 350002,  
15 PR China.

16 <sup>b</sup> Department of Plant Genetics & Breeding and Seed Science, College of Agronomy and  
17 Biotechnology, China Agricultural University/Beijing Key Laboratory of Crop Genetic  
18 Improvement/The Innovation Center (Beijing) of Crop Seed Sciences Ministry of  
19 Agriculture, Beijing 100193, PR China.

20 <sup>c</sup> International Flavour Research Centre, Division of Food, Nutrition and Dietetics, University  
21 of Nottingham, Sutton Bonington Campus, Loughborough LE12 5RD, United Kingdom.

22 <sup>d</sup> International Flavour Research Centre (Adelaide), School of Agriculture, Food and Wine  
23 and Waite Research Institute, The University of Adelaide, PMB 1, Glen Osmond, South  
24 Australia 5064, Australia.

25 <sup>e</sup> Institute of Food Science and Technology, Chinese Academy of Agricultural Sciences,  
26 (CAAS), Beijing, China.

27 <sup>1</sup> These authors contributed equally to this work.

28

29 **\* Corresponding Authors**

30 Tingting Zhang (Email: [ztt\\_fafu@163.com](mailto:ztt_fafu@163.com))

31 Qun Sun (Email: [sunqun@cau.edu.cn](mailto:sunqun@cau.edu.cn))

32 Rensen Zeng (Email: [rszeng@fafu.edu.cn](mailto:rszeng@fafu.edu.cn))

33

34 **ABSTRACT**

35 Understanding and predicting the storage stability of sweetcorn seeds is critical for  
36 effective supply chain management, however, prediction ability relies heavily on  
37 accelerated ageing (AA) studies and this is not always directly applicable to natural  
38 ageing (NA). In this study, hyperspectral imaging (HSI) and non-targeted  
39 metabolomics (LC-MS/MS) were integrated using PLS-R, SVM-R and OPLS-DA to  
40 predict loss of seed vigour in NA seeds, using data based on AA seeds. The  
41 inconsistencies in the pattern of spectral variation between seeds undergoing AA and  
42 NA were first identified. AA-based vigour prediction models were then built using all  
43 wavelengths and effective wavelengths (EWs) selected by regression coefficients.  
44 These models were externally validated by independent AA and NA seed datasets,  
45 respectively. The results yielded satisfactory predictions for AA seeds ( $R^2 \geq 0.814$ ),  
46 but low precision for NA seeds ( $R^2 \leq 0.696$ ). Metabolome analysis identified 54  
47 differential metabolites, containing a large proportion of amino acids, dipeptides and  
48 their derivatives, which were important substances reflecting discrepancies between  
49 the ageing mechanisms of AA and NA seeds. Subsequently, N-H bond-related  
50 wavebands were deemed to be a possible interference factor in the models'  
51 practicability. After removing the N-H bond-related EWs, the AA-based models  
52 achieved better performance on NA seeds, with  $R^2_{v-2}$  value increasing from 0.696 to  
53 0.720 for Lvsechaoren and from 0.668 to 0.727 for Zhongtian 300. In summary,  
54 coupling HSI, LC-MS/MS and machine learning was shown as an appropriate  
55 approach for non-destructive monitoring and predicting the vigour of stored  
56 sweetcorn seeds.

57

58 **Keywords:** Hyperspectral imaging; Non-targeted metabolomics; Machine learning;  
59 Seed vigour; Natural ageing; Accelerated ageing.

60

61

62



## 64 **1. Introduction**

65 Sweetcorn (*Zea mays* L. *saccharata* Sturt) is an important vegetable and  
66 becoming increasingly popular in international trade (Chauhan et al., 2022). The high  
67 economic value can be attributed to its good taste, unique flavour and high nutritional  
68 value content such as protein, amino acids, vitamins and dietary fibre (Bai et al.,  
69 2021). Sweetcorn is naturally developed by recessive mutations in characteristic  
70 genes that control the conversion of sugar into starch (Singh et al., 2014). The  
71 property of high soluble sugar in sweetcorn kernels has resulted in increased fungi  
72 infection. Meanwhile, the absence of substantial starch led to the rapid reduction in  
73 seed quality (vigour) during storage. Poor seed vigour mainly exhibits a negative  
74 influence on the germination ability, uniform emergence, yield potential, tolerance to  
75 abiotic stress and sowing cost (Rodo & Marcos-Filho, 2003). Consequently, real-time  
76 monitoring of changes in vigour parameters and better predictive ability of seed  
77 longevity during storage can provide guidance on when and which seed stocks should  
78 be placed onto the seed market. For this reason, it is crucial to have reliable methods  
79 of determining sweetcorn seed vigour accurately, quickly and consistently.

80 Conventional methods for assessing seed vigour or longevity potentials, such as  
81 tetrazolium staining (Hosomi et al., 2012), conductivity tests (Rahman & Cho, 2016),  
82 and germination tests (Merritt et al., 2014), have a number of limitations, including  
83 being destructive, labour-intensive, highly subjective and time-consuming. Given  
84 these drawbacks, there is an increasing interest worldwide to investigate new rapid  
85 and non-destructive analytical technologies for discriminating highly-viable seeds  
86 based on the development of computer and optical sensor technology.

87 Among the numerous emerging technologies, the hyperspectral imaging (HSI)  
88 system provides a three-dimensional hyperspectral data cube containing spatial  
89 information of the first and second dimensions (as regular imaging systems), and  
90 spectral data of the third dimension (spectral information for each pixel in an image)  
91 (Fan et al., 2019; Zhang et al., 2014; Zhang et al., 2023). The HSI has been confirmed  
92 to be a powerful tool with a great potential for seed vigour evaluation as demonstrated

93 in wheat (Zhang et al., 2020) and corn (Ambrose et al., 2016; Wakholi et al., 2018),  
94 tomato (Shrestha et al., 2016), Japanese mustard spinach (Ma et al., 2020),  
95 muskmelon (Kandpal et al., 2016) and sugar beet (Yang et al., 2021). However, most  
96 of the reported studies relied on seeds treated by accelerated ageing (AA) to establish  
97 generic prediction models, but the limitations of the AA approach have not been fully  
98 considered.

99 Generally, AA is used to simulate natural ageing (NA) to shorten timescales by  
100 combining high ambient temperature and high relative humidity conditions. However,  
101 the consistency of metabolic mechanisms in seeds between these two different ageing  
102 processes remains controversial. Priestley and Leopold (1983) observed that the  
103 proportion of polyunsaturated fatty acids in soybean seeds decreased under NA  
104 condition, whereas it did not change under AA (Priestley & Leopold, 1983).  
105 Subsequently, the activity of acid phosphatase in NA rice seeds was proved to be  
106 lower than that in AA seeds (Freitas & Dias, 2006). Recent data on the volatile  
107 metabolites of sweetcorn seeds during NA and AA (Zhang et al., 2021) revealed that  
108 14 of 33 identified compounds exhibited a significant difference in relative content.  
109 The types and concentration of chemical substances in seeds are closely related to the  
110 spectral characteristics because the amplitude and waveform variations of HSI system  
111 can record the molecular overtones and combinations of the fundamental vibrations  
112 caused by the stretching and bending of C-H, O-H and N-H groups (Liu et al., 2017;  
113 Wang et al., 2021).

114 In this study, we proposed that spectral information could reflect discrepancies in  
115 metabolites between NA and AA seeds, which may affect AA-based predictive model  
116 application to predict the vigour of NA seeds. Despite the practicality of AA-based  
117 models for predicting the vigour of sweetcorn seeds based on the volatile metabolites  
118 detected by gas chromatography-ion mobility spectrometry has been reported (Zhang  
119 et al., 2021), there is no published work researching this in spectral information of the  
120 HSI system. Additionally, up to now, there is a lack of study to investigate the  
121 underlying reasons related to the problems in the model application. LC-MS-based  
122 metabolomics was a typical method to investigate the wide arrays of non-volatile

123 metabolites in biological samples. Therefore, the hypothesis of this study is that  
124 linking the metabolites data with spectral information would help explain the  
125 similarities and differences between seeds undergoing NA and AA. To the best of our  
126 knowledge, this will be the first study to assess the effectivity of AA predictive  
127 models for NA seeds and extend the models by integrating HSI and metabolomics  
128 data.

129 Therefore, the objectives of this study were: i) to investigate the similarities and  
130 differences between spectral information of NA and AA seeds; ii) to evaluate the  
131 application discrepancies of the optimal vigour prediction models on NA and AA  
132 seeds; iii) to explain the discrepancies in model application by analysing the  
133 metabolome of NA and AA seeds; iv) to optimise the performance of AA-based  
134 vigour predictive model on stored seeds by coupling mathematical prediction models  
135 and metabolomics.

## 136 **2. Materials and methods**

### 137 **2.1. Seed sample collection**

138 Two commercial sweetcorn cultivars widely promoted in China, called  
139 Lvsechaoren and Zhongtian 300, were purchased from Hezhiyuan Seed Company  
140 (Weifang, Shandong, China). The initial moisture contents (MCs) were 11.0%  
141 (Lvsechaoren) and 12.5% (Zhongtian 300) wet weight basis. The corresponding raw  
142 germination percentages (GP, reflecting seed vigour) were 85.5% and 92.0%,  
143 respectively. Since the sampling frequency could not be determined until the target  
144 GP (Lvsechaoren: 75%, 65% and 45%; Zhongtian 300: 83%, 75% and 65%) was  
145 reached, 37.5 kg seeds for each cultivar were collected to ensure sufficient samples.  
146 Seeds were then randomly allocated to three groups: control group, natural ageing  
147 (NA) group and accelerated ageing (AA) group.

148 For the control group, a 1.5 kg portion of seeds from each cultivar was taken as a  
149 no-retreatment control set, immediately sealed in a polyethene bag and stored at -  
150 20°C. For the NA group, 12 equal portions (1.5 kg) of each cultivar were stored in  
151 cotton bags at ambient room temperature (17 ~ 28°C and 30 ~ 60% RH). One portion

152 of each cultivar was taken for GP assessment at a monthly frequency until the GP of  
153 the seed sample was reduced to approximately 75%. The assessments were then  
154 performed every half month once the GP was below 75%. For the AA group, 12 x 1.5  
155 kg portions of each cultivar were stored in nylon mesh bags, suspended over distilled  
156 water inside sealed glass containers in an electric oven (45°C, 100% RH) (Muasya et  
157 al., 2009). One portion of each cultivar was assessed for GP every 16 hours before it  
158 dropped to approximately 75% and then sampled every 8 hours after the GP dropped  
159 below 75%.

160 The isolated samples were transferred into an incubator ( $23 \pm 1^\circ\text{C}$ ) and dried  
161 back to their original MCs. Damaged seeds were removed during the natural drying  
162 processes. Once the seeds dried to their original MCs, 200 seeds were randomly  
163 selected from every portion for GP assessment (Zhang et al., 2020), and the residual  
164 seeds were sealed up in a polyethene bag and placed at  $-20^\circ\text{C}$ . When the NA and AA  
165 seeds for each cultivar reached the target GP, the residual seeds of corresponding  
166 portion were reserved for further experiments.

## 167 **2.2. Hyperspectral image acquisition**

168 All seed samples were scanned by an assembled visible-near infrared (Vis-NIR)  
169 HSI system (**Fig. 1**) (Zhang et al., 2020). The details of the HSI system were  
170 described in Method S1. Each seed was positioned on the platform with the embryo  
171 side facing upwards for scanning. The raw hyperspectral image with three dimensions  
172 ( $x, y, \lambda$ ) was obtained with a scanning speed of 1.3 mm/s and an exposure time of 15  
173 ms. In this study, the images contained 1004 pixels in the x-direction and 1002  
174 wavebands in the  $\lambda$  direction with 0.31 nm intervals. The number of pixels in the y-  
175 direction depended on the scanning length for samples. During spectral data  
176 acquisition, the relative humidity and the temperature in the laboratory were 30% and  
177  $23 \pm 1^\circ\text{C}$ , respectively.

## 178 **2.3. Hyperspectral image correction**

179 Details of the hyperspectral image correction were described in Method S1.

## 180 **2.4. Spectral Data Extraction**

181 The seed sample in each corrected hyperspectral image was identified as a region  
182 of interest (ROI) and segmented from the background. Spectral data of the ROI were  
183 then extracted and averaged based on the reflectance value of each pixel in it. The  
184 beginning 133 and ending 16 wavebands were eliminated due to the low signal-to-  
185 noise ratio. Ultimately, 853 bands from 400.2 nm to 1069.9 nm for the sweetcorn  
186 seeds remained for future data analysis. The processes of spectral data segmentation  
187 and extraction were conducted by using the software ENVI 5.1 (ITT Visual  
188 Information Solutions, Boulder, CO, USA)

## 189 **2.5. Non-targeted metabolite profiling**

190 Six biological replicates were used for each target GP, and within each replicate,  
191 50 seeds were included. The seeds were firstly ground into a powder with liquid  
192 nitrogen, 60 mg of each replicate removed and homogenized in 200  $\mu$ L deionized  
193 water. After adding 800  $\mu$ L methanol/acetonitrile (1:1, v/v) and vortexing for 30 s,  
194 samples were dissolved and decomposed by ultrasonication in an ice bath for 30 min  
195 at 4°C. Subsequently, to precipitate and remove the protein, samples were incubated  
196 for 1 hour at -20°C and centrifuged for 15 min (13000 rpm, 4°C). The supernatant  
197 was collected, dried in a vacuum centrifuge, and stored at -80°C. Each sample was  
198 re-dissolved in acetonitrile/water (1:1, v/v) solvent prior to the non-targeted  
199 metabolite analysis. Samples were analysed using LC-MS/MS, and details were  
200 described in Method S1.

## 201 **2.6. Statistical analysis**

### 202 *2.6.1 HSI data analysis*

203 A total of 272 seed samples, each containing 50 seeds of uniform size without  
204 physical damage, were randomly taken out for HSI modelling and analysis (**Fig. 2**).  
205 For each cultivar, AA seed samples (n=64), taken across the GP range were used to  
206 construct predictive regression models for vigour. Separately, two additional seed sets  
207 (n=36), one from the AA trial and another from the NA trial, were used to validate the

208 predictive accuracy and practicability of the models created. Distribution of seed  
209 vigour for each cultivar in calibration set, validation set-1 and validation set-2 was  
210 shown in **Table 1**.

#### 211 (1) Spectra Pre-processing

212 Spectral pre-processing techniques are utilised to maximise the quality of  
213 hyperspectral measurement and minimise the interference information existing in  
214 spectral data (He et al., 2023). Standard normal variate (SNV) is applied to reduce  
215 additive and multiplicative effects by spectral normalisation approach. First derivative  
216 (FD) is a common method for getting rid of baseline offsets. Smoothing is regularly  
217 served to alleviate high-frequency noise produced by instruments. In this study,  
218 standard normal variate (SNV), autoscale (AS), and Savitzky–Golay first derivative  
219 (FD, second-order polynomial) with a smoothing gap of 9 and their combinations  
220 were investigated.

#### 221 (2) Effective Wavelengths (EWs) Selection

222 Given that the high-resolution hyperspectral image data contain large amounts of  
223 redundant information, it was necessary to select EWs that carried the maximum  
224 spectral information to reduce irrelevant variables as well as to enhance the  
225 computational efficiency of prediction models. Regression coefficients (RC), a  
226 model-based EWs selection method, has been proved in our previous work to be an  
227 excellent method to effectively extract the spectral characteristics and sensitive  
228 spectrum bands that are closely related to the change in vigour levels during seed  
229 ageing (Zhang et al., 2020). Generally, RC in PLS-R models is implemented to  
230 distinguish the variables with measurable influences on the dependent ( $Y$ ) variables.  
231 The significance of the influences for predicting the  $Y$ -variables is expressed in  
232 absolute values (He et al., 2013). Thus, wavelengths with large absolute values of beta  
233 coefficients can be considered the EWs.

234 Following pre-processing, AA-based vigour prediction models (PLS-R and  
235 SVM-R) were established, cross-validated/internal-validated and external-validated

236 using the data sets collected from control and AA seeds. Leave-one-out was  
237 implemented as the cross-validation (CV) method. Subsequently, optimal AA-based  
238 PLS-R models were obtained based on representative wavelengths screened by RC.  
239 Separate data sets from control and NA seeds were used to externally validate the  
240 predictive ability of the optimal PLS-R models. The performance of these models was  
241 evaluated by the coefficient of determination ( $R^2$ ) of calibration ( $R^2_c$ ), cross-validation  
242 ( $R^2_{cv}$ ) and validation ( $R^2_v$ ), and the root mean square error (RMSE) of calibration  
243 (RMSEC), cross-validation (RMSECV) and validation (RMSEV). The procedure was  
244 carried out in MATLAB R2014a (The MathWorks, Natick, MA, USA).

### 245 2.6.2 Metabolite analysis

246 Raw data collected from LC-MS/MS were converted to mzXML files by  
247 ProteoWizard and then processed by the XCMS program, which includes peak  
248 alignment, retention time correction and peak area extraction. Metabolite annotations  
249 were conducted based on mass accuracy (<25 ppm) and their secondary spectral  
250 pattern, and then matched with the in-house database. The metabolite data were pre-  
251 processed by Pareto-scaling algorithm, and then processed by principal component  
252 analysis (PCA) and orthogonal partial least squares discriminant analysis (OPLS-DA)  
253 models. The model fitting and predictability were evaluated by  $R^2$  and  $Q^2$ ,  
254 respectively. In this study, metabolites with variable importance in projection (VIP>1)  
255 algorithm and ANOVA analysis ( $p<0.05$ ) were identified as differential metabolites.  
256 These processes were conducted in SIMCA-P 14.1 software (Umetrics, Umea,  
257 Sweden).

## 258 3. Results and discussion

### 259 3.1. Spectral characteristics of naturally and accelerated aged seeds

260 The spectra of Lvsechaoren and Zhongtian 300 sweetcorn seeds are presented in  
261 **Fig. 3**. For each cultivar, the spectral profiles of seed samples in the same spectral  
262 range followed a similar trend, but varied in the vibration magnitude of reflectivity  
263 (**Fig. 3A and D**). Subsequently, the spectra of seeds with the same vigour level of

264 each cultivar were averaged to clearly observe the amplitude and waveform variations  
265 across vigour levels (**Fig. 3B and E**). The spectral curves of the two varieties  
266 increased to a plateau before decreasing, and the curves' fluctuation in the near-  
267 infrared region (NIR, 760~1070 nm) was greater than that in the visible region (Vis,  
268 400~760 nm).

269 The difference in spectral reflectance between different vigour levels increased  
270 gradually above 600 nm. Hence, the spectra in the 600~900 nm range were extracted  
271 for plotting to zoom in on the differences in spectral features (**Fig. 3C and F**). They  
272 showed that the mean-reflectance of accelerated aged (AA) seeds of both cultivars  
273 was higher than that of naturally aged (NA) and RAW seeds in the NIR and most of  
274 the Vis region. Moreover, the reflection spectra exhibited an inverse relationship with  
275 germination% among AA groups, i.e., reduced relative reflectance with increasing  
276 germination% (AA 45%/65% > AA 65%/75% > AA 75%/83% RAW > 85%/92%).  
277 However, this pattern was not observed in NA groups. Lysechaoren seeds  
278 demonstrated lower relative reflectance at NA 75% (early ageing) compared to NA  
279 65% and NA 45%, and for Zhongtian 300 seeds, the relative reflectance of NA 65%  
280 was lower than NA 83% but higher than NA 75%. The inconsistent spectral patterns  
281 between seeds undergoing AA and NA may be due to some discrepancies in the  
282 physicochemical properties of sweetcorn seeds induced by AA compared to those  
283 induced by NA. These inconsistencies were indicative of the potential challenges in  
284 NA seed vigour prediction based on AA-based models.

### 285 **3.2. Development and validation of AA-based predictive models using full** 286 **wavelengths**

287 Various pre-processing algorithms (including AS, FD, SNV and FD+SNV) were  
288 first utilised to maximize the usefulness of spectrum data. PLS-R and SVM-R  
289 prediction models for seed vigour were established based on the raw and pre-  
290 processed full spectra from the calibration set (**Table 2 and Table S1**). Vigour  
291 predictive models generated from AA seed data sets, were not only validated by  
292 separate AA seed data (validation set-1), but more importantly, were assessed for

293 applicability to stored seeds using NA validation seed data (validation set-2). The  
294 tables showed the performance of the PLS-R and SVM-R models for the vigour of AA  
295 and NA seeds, including the coefficient of determination ( $R^2$ ) and the root mean  
296 square error (RMSE) for the calibration, cross-validation (internal validation,  
297 reflecting models' precision), validation set-1 (external validation, reflecting models'  
298 stability) and validation set-2 (external validation, reflecting models' practicability).

299 In general terms, the PLS-R models based on full wavelengths for Lvsechaoren  
300 and Zhongtian 300 seeds both showed higher stability than the SVM-R models.  
301 Hence, the PLS-R algorithm was used in the successor investigation. The PLS-R  
302 models for the two cultivars presented good performance in cross-validation set ( $R^2_{cv}$   
303  $> 0.830$  and  $RMSE \leq 0.060$ ). Among them, two highlighted models (SNV-PLS-R  
304 models for Lvsechaoren and FD-SNV-PLS-R model for Zhongtian 300) exhibited the  
305 optimal predictive ability in cross-validation set ( $R^2_{cv} = 0.871$ ,  $RMSECV = 0.054$ ;  
306  $R^2_{cv} = 0.846$ ,  $RMSECV = 0.037$ , respectively), and showed the excellent prediction  
307 results in validation set-1 ( $R^2_{v-1} = 0.905$ ,  $RMSEC-1 = 0.049$ ;  $R^2_{v-1} = 0.838$ ,  $RMSEC-1$   
308  $= 0.040$ ). Despite good performance on AA data of calibration sets, cross-validation  
309 sets and validation set-1, the models' performance on NA data as validation set-2 was  
310 poor ( $R^2_{v-2} = 0.513$  for Lvsechaoren and  $R^2_{v-2} = 0.326$  for Zhongtian 300).

### 311 **3.3. Effective wavelengths (EWs) selection based on optimal PLS regression** 312 **models**

313 Regression coefficients (RC) method was then utilized to extract the effective  
314 wavelengths (EWs) with the most information of seed vigour, based on the two  
315 optimal AA-PLS regression models (bold) in Table 2. The RC of each variable in the  
316 models were calculated and the corresponding RC curves were presented in **Fig. 4**.  
317 EWs were then selected at the peaks and valleys of the curves. Ultimately, 33 EWs for  
318 Lvsechaoren seeds and 31 for Zhongtian 300 seeds were summarised, respectively  
319 (**Table S2**). The results revealed an interesting phenomenon, that is, although the EWs  
320 of these two cultivars screened by RC method were different, they both lied on the  
321 vicinity of some specific spectral regions. In the visible region (400~760 nm),

322 wavebands at around 430 nm are associated with chlorophyll *a*, those around 471 nm  
323 represent carotenoids (Nansen et al., 2015), and those in 520~560 nm could be  
324 derived from anthocyanins (Yokoi & Sait, 1973). EWs in the NIR region (760~1070  
325 nm) mainly originated from the overtones and combinations of fundamental  
326 vibrations of O-H, N-H and C-H functional groups, which are the essential  
327 components of seed molecules (i.e. carbohydrate, fat, water and protein, etc.).  
328 Intriguingly, a large proportion of the selected EWs for both cultivars was  
329 concentrated in the NIR region, which indicated that the application precision of the  
330 AA-based vigour predictive models could be greatly affected by the consistency of  
331 structure/type and concentration of chemical composition in seeds between AA and  
332 NA. Besides, compared with wheat seeds (Zhang et al., 2020), fewer EWs in the  
333 visible region and more EWs in the NIR region were selected for sweetcorn seeds,  
334 which may be due to the thin seed coat and less pigment content of sweetcorn seeds.

#### 335 **3.4. Development and validation of AA-based predictive models using EWs**

336 EWs of Lvsechaoren and Zhongtian 300 AA seeds were respectively applied as  
337 input variables to construct simplified PLS-R models (**Table 3**). For Lvsechaoren, the  
338 optimal simplified RC-SNV-PLS-R model ( $R^2_{v-1} = 0.909$ , RMSEV-1 = 0.050) based  
339 on only 33 characteristic bands exhibited similar prediction power to the Full-SNV-  
340 PLS-R model ( $R^2_{v-1} = 0.905$ , RMSEV-1 = 0.049) for AA seeds. The precision of the  
341 corresponding models on NA data showed an  $R^2_{v-2}$  value of 0.586 (RC-SNV-PLS-R)  
342 and an  $R^2_{v-2}$  value of 0.513 (Full-SNV-PLS-R). For Zhongtian 300, the optimal  
343 simplified RC-FD-SNV-PLS-R model ( $R^2_{v-1} = 0.866$ , RMSEV-1 = 0.036) was  
344 slightly better than the Full-FD-SNV-PLS-R model ( $R^2_{v-1} = 0.838$ , RMSEV-1 =  
345 0.040) for AA seeds. The RC-FD-SNV-PLS-R model ( $R^2_{v-1} = 0.561$ , RMSEV-1 =  
346 0.076) as well performed better than the Full-FD-SNV model ( $R^2_{v-1} = 0.326$ ,  
347 RMSEV-1 = 0.114) in predicting NA seeds of the same cultivar. Noteworthy, the  
348 simplified RC-FD-PLS-R model showed better performance in NA seeds compared to  
349 the full-band FD-PLS-R model: the corresponding  $R^2_{v-2}$  value increased from 0.279 to  
350 0.553 and the RMSEV-2 value decreased from 0.117 to 0.078. These indicated that

351 the combination of optimal EWs and a suitable algorithm could improve the  
352 application performance of the AA-PLS regression models on stored seeds in this  
353 case.

354 Additionally, the simplified PLS-R models built by AA seeds performed well on  
355 the vigour prediction of seeds treated by the same process ( $R^2_{v-1} \geq 0.820$ ), but they  
356 did not have satisfactory performance on NA seeds ( $R^2_{v-2} \leq 0.696$ )(**Table 3**).  
357 Specifically, for Lvsechaoren, the  $R^2_{v-2}$  of the optimal RC-SNV-PLS-R model  
358 decreased by 0.323 when compared to  $R^2_{v-1}$ . Moreover, the RC-SNV-PLS-R model of  
359 Lvsechaoren presented a better prediction than that of Zhongtian 300 on AA seeds  
360 ( $R^2_{v-1}$ : 0.909 vs. 0.846), but performed slightly worse than that of Zhongtian 300 on  
361 NA seeds ( $R^2_{v-2}$ : 0.586 vs. 0.668). These results indicated the optimal PLS regression  
362 models for predicting seed vigour, generated using data from AA seeds may not be  
363 able to accurately predict the vigour changes of seeds during NA.

### 364 **3.5. Age-related changes in metabolite levels in the AA and NA seeds**

365 To investigate the specific differences in the ageing mechanism underlying the  
366 prediction performance of the PLS-R models on AA and NA seeds, the age-related  
367 changes in metabolite levels of Lvsechaoren seeds from RAW 85% and critical node  
368 (AA 75% and NA 75%) samples were identified by LC-MS/MS. The large complex  
369 datasets generated from positive ion modes (PIM) and negative ion modes (NIM)  
370 were subjected to non-supervised principal component analysis (PCA), which  
371 displayed the difference of chemical substances in seeds from different vigour and  
372 treatment seeds. The 2D score plots of PCA (**Fig. 5A and 5B**) showed the seed  
373 samples were grouped into three clusters in PIM and NIM. Although there were large  
374 overlap areas among these clusters, some degree of differences in variation trends still  
375 existed, which aroused our interest on exploring the underlying metabolic mechanism  
376 and also provided a statistical basis for further data mining.

377 A supervised multivariate analysis method, orthogonal partial least squares  
378 discriminant analysis (OPLS-DA) (**Fig. 6**), and univariate methods, fold change (FC)  
379 and T-test, were combined to analyse the metabolites between every two groups.

380 Chemical compounds that met the criteria (i.e., the variable importance in the project  
381 (VIP) value  $\geq 1$ , FC  $\geq 1.2$  or FC  $\leq 0.8$  and  $p$ -value  $< 0.05$ ) were selected as significant  
382 differential metabolites. Most of them are involved in the primary metabolism, such  
383 as sucrose, choline, linolenic acid, and ferulic acid (**Table S3 and S4**). Venn diagrams  
384 were then applied to further depict the shared differential metabolites among the  
385 pairwise comparisons (**Fig. 5C and 5D**). Overlaps of 7 differential metabolites in the  
386 PIM (yellow patch) and 7 differential metabolites in the NIM (yellow patch) were  
387 revealed to change simultaneously (up-regulated or down-regulated) between  
388 sweetcorn seeds during AA and NA. Moreover, overlapping differential metabolites of  
389 28 (22+4+2) in the PIM (green patch) and 26 (23+1+2) in the NIM (green patch) were  
390 identified to change asynchronously between AA and NA. These metabolites were  
391 possibly key age-related differentiators of seeds responding to NA and AA, and some  
392 of them may greatly contribute to building the AA-based models. An in-depth analysis  
393 of these metabolites is helpful to study the differences in metabolic mechanisms in  
394 seeds with high sugar content under different ageing processes.

395 The overlapping metabolites with synchronous and asynchronous changes  
396 between seeds during NA and AA were visualised in **Fig. 5E & 5F and Table S3 &**  
397 **S4**. After AA treatment, the number of up-regulated metabolites (30 in PIM and 26 in  
398 NIM) in seeds is higher than that of down-regulated metabolites (3 in PIM and 5 in  
399 NIM). This phenomenon was also found in the seeds treated by NA. These results  
400 manifested that some key physiological and metabolic activities contributing to the  
401 reduction in seed vigour might be activated during AA and storage. Furthermore, the  
402 relative content of 28 substances (PIM) and 25 substances (NIM) in AA seeds were  
403 significantly higher than those in NA and RAW seeds, indicating that the inner  
404 physiological metabolism of sweetcorn seeds under AA may be faster than that under  
405 NA, which is consistent with the findings obtained from our previous study (Zhang et  
406 al., 2021).

407 Additionally, results from this current work revealed an interesting phenomenon,  
408 which has not been reported before. There were differences in the types of key age-  
409 related metabolites that changed asynchronously and synchronously in seeds between

410 AA and NA. Among the asynchronous metabolites (**Fig. 5F, Table S3 & S4**), amino  
411 acids and their derivatives (such as sarcosine and 4-aminobutyric acid) account for a  
412 large proportion. In the synchronous substances (**Fig. 5E, Table S3 & S4**), however,  
413 sugars and their derivatives (such as sucrose, raffinose, and glucosamine) are  
414 predominant. These important results can be interpreted in two different ways. Firstly,  
415 in terms of inner physiological metabolism of seed ageing, variations in the content of  
416 many soluble sugars presumably play an important role in the vigour of sweetcorn  
417 seeds, which is consistent with the findings in maize and legume seeds (Bernal-Lugo  
418 & Leopold, 1995; Obendorf & Górecki, 2012). The second is a more interesting  
419 perspective, which starts with the principle of spectroscopy. As is well known, the  
420 abundant spectral features result from photons absorption by light interacting with  
421 molecules. Specifically, the overtone of molecules and combinations of the  
422 fundamental vibrations caused by the stretching and bending of C-H, O-H and N-H  
423 bonds can be recorded by the variations of spectral reflectivity. As a result, the  
424 types/structures and concentration of molecules in seeds have a great effect on  
425 spectral information. Given this, the key differential molecules between NA and AA  
426 seeds contain a large proportion of amino acid, dipeptides and their derivatives, most  
427 of which have N-H bonds (**Fig. 7**). This may result in a certain degree of difference in  
428 spectral characteristics between NA and AA seeds, which adversely influenced the  
429 practicality of the AA-based model on sweetcorn seeds during storage.

### 430 **3.6. Vigour prediction of NA seeds based on EWs without N–H bonds**

431 To minimise the interference of age-related differential metabolites to the  
432 mathematical models' utility, the EWs are associated with the N–H bonds were  
433 identified and eliminated. Among the Vis-NIR (400~1070 nm) range, the spectral  
434 bands at 760~820 nm were associated with the third overtone of N-H stretching  
435 (Cheng & Sun, 2015), and the wavebands between 1040~1050 nm were attributed to  
436 the N-H bond second overtone of protein (Fan et al., 2021). Besides, wavelengths at  
437 around 430 nm were derived from chlorophyll *a*, and its molecular structure contains  
438 the N–H bond (**Fig. 4**). After deleting the corresponding EWs (bold), the remaining

439 EWs were further used to establish the new AA-based PLS-R model and tested the  
440 practicality with AA and NA seeds (**Table 4**). The precision of the RC-PLS-R-models  
441 on NA seeds is generally improved to a certain extent after deleting the EWs  
442 associated with the N–H bonds. Specifically, the highest  $R^2_{v-2}$  value increased from  
443 0.696 of RC-FD-SNV-PLS-R model to 0.720 of RC-AS-PLS-R model without N–H  
444 bond-related bands. For Zhongtian 300, the maximum value of  $R^2_{v-2}$  elevated from  
445 0.668 in RC-SNV-PLS-R model to 0.727 in RC-SNV-PLS-R model without N–H  
446 bond-related bands. Whereas, most of the RC-PLS-R models without N–H bond-  
447 related bands were found to be less powerful than RC-PLS-R models on the vigour  
448 prediction of AA seeds. This may be because the removed N-H bond-related bands,  
449 while not conducive to the vigour prediction of NA seeds, are highly important bands  
450 that achieve accurate prediction of the AA seed vigour.

451 Seeds under AA contained much higher levels of fatty acids and their derivatives  
452 (such as cis-9-Palmitoleic acid and Jasmonic acid), amino acids and their derivatives  
453 (such as sarcosine and 4-aminobutyric acid), compared to NA seeds (**Table S3 & S4**).  
454 These demonstrated that more fatty acids and amino acids changed in sweetcorn seeds  
455 during AA. The increased free fatty acids may be generated due to lipid peroxidation.  
456 Amino acids can be produced from the degradation of proteins and serve as  
457 precursors of many secondary metabolites under biotic and abiotic stresses. Such a  
458 rapid change of amino acids derived from the glycolytic pathway (such as valine and  
459 leucine) was also found in wheat in response to drought stress (Roessner, 2012).

460 Sweetcorn seeds are rich in carbohydrates, such as sucrose and raffinose-related  
461 oligosaccharides. Sucrose and oligosaccharides can be hydrolysed to reduce sugars,  
462 which are involved in the initial stages of the Maillard-Amadori reaction (MR). The  
463 free amino acids and dipeptides in seeds also have strong reactivity as substrates of  
464 MR. MR, as a non-enzymatic reaction involving reducing sugars and amino groups of  
465 amino acids or proteins, can initiate in seeds with low temperatures and very low  
466 moisture content and has been observed to exist a correlation with seed deterioration  
467 (Sun & Leopold, 2010). The increased reducing sugar and amino acids in sweetcorn  
468 seeds might act in concert to drive the MR in dried sweetcorn seeds during AA and

469 NA. This is consistent with the abundance of volatile MR products (e.g.,  
470 benzaldehyde and 2-heptanone) observed in AA and NA seeds from our previous  
471 studies (Zhang et al., 2021).

472 Additionally, raffinose family oligosaccharides, especially raffinose and inositol  
473 galactoside (a precursor of raffinose oligosaccharide synthesis), were suggested to be  
474 related to seed vigour in tomato, Arabidopsis and rice (Deborah et al., 2016; Yan et  
475 al., 2018). Raffinose and saccharose, together with other oligosaccharides and  
476 cycloalcohols, are capable of binding free radicals to protect cytomembranes from  
477 deterioration during prolonged storage (Zalewski & Lahuta, 1998). An earlier study  
478 declared the depletion of raffinose amounts is positively correlated with the declined  
479 vigour in the stored maize seeds ( Bernal-Lugo & Leopold, 1992). However, Yan et al.  
480 (2018) recently found that the variation of raffinose content was not significant in rice  
481 seeds under 24-month natural ageing (Yan et al., 2018). Interestingly, data in our work  
482 exhibited that the levels of raffinose increased in desiccated sweetcorn seeds after  
483 both AA and NA treatments (**Fig. 5E and 5F**). Whether this difference is caused by  
484 species/cultivar diversity or by ageing stages (Phase I/Phase II/Phase III) needs  
485 further investigation.

486 Notably, sucrose, a main form of soluble sugar stored in sweetcorn seeds, has  
487 been reported to decrease significantly in aged maize axes and rice embryos (Kataki  
488 et al., 1997), and its concentration exhibited strong consistency in the extent of the  
489 vigour decrease in sweetcorn seeds treated by AA and NA (**Fig. 5E and 5F**).  
490 Furthermore, sucrose can be converted by sucrose synthase (EC 2.4.1.13), encoded by  
491 the *Sus1* gene of maize, into UDP-glucose and fructose (Chen & Chourey, 1989).  
492 Fructose and glucose could then be synthesised to sorbitol under the action of  
493 aldoketone reductases. Sorbitol, as an acyclic polyol, was detected earlier in  
494 developing maize kernels (Shaw & Dickinson, 1984) and germinating soybean seeds  
495 (Kuo et al., 1990). The synthesis of sorbitol in aged seed embryos is important to  
496 arrest the increase of reducing sugar concentration and reduce the non-enzymatic  
497 attack of reducing sugar on protein amino acids and nucleic acid/protein complexes  
498 (Lahuta et al., 2007). Kataki et al. (1997) observed that sorbitol significantly

499 accumulates in rice, soybean, cotton, and lettuce maize seeds during rapid ageing  
500 treatment (45°C and 75% RH) and considered it to be a good common biomarker for  
501 seed deterioration (Kataki et al., 1997). Remarkably, our results not only support this  
502 idea in sweetcorn seeds, and further found the accumulation of sorbitol levels in NA  
503 seeds was consistent with that in AA seeds (**Fig. 5E and 5F**). On the whole, we infer  
504 that soluble sugars and their derivatives, as well as amino acids and their derivatives,  
505 could potentially constitute ideal starting points for studying the metabolic  
506 mechanisms and longevity prediction of seeds with high sugar content under different  
507 ageing processes. Meanwhile, amino acids and their derivatives can be a potential  
508 entry point to optimise the application ability of vigour prediction model based on  
509 spectral information from sweetcorn seeds.

#### 510 **4. Conclusion**

511 In the present study, Vis-NIR HSI, non-targeted metabolomics and intelligent  
512 data mining approaches were combined for the first time to investigate the differences  
513 in metabolomic and predicted vigour between sweetcorn seeds undergoing  
514 accelerated ageing (AA) and natural ageing (NA). Specifically, the difference in  
515 ageing treatments resulted in a significant influence on the spectral reflectance of  
516 sweetcorn seeds across the two cultivars (Lvsechaoren and Zhongtian 300). The  
517 spectral reflectance of both cultivars increased linearly during AA (RAW 85% /92% <  
518 AA 75% /83% < AA 65% /75% < AA 45% /65%), but increased irregularly during  
519 NA. Recognition that Full-PLS-R and RC-PLS-R models based on AA seed data  
520 showed high prediction precision ( $R^2 \geq 0.814$ ) for AA seeds, but reduced predictive  
521 ability ( $R^2 \leq 0.696$ ) for stored seeds (NA). Moreover, use of metabolomics to identify  
522 ageing-related differential metabolites between AA and NA seeds, especially amino  
523 acids and their derivatives, enabled the identification of factors responsible for  
524 negatively impacting the accuracy of predictive models. After analysing the molecular  
525 structure of these key differential metabolites, N–H bond-related EWs were  
526 considered to adversely influence the application of AA-based predictive models.

527 Ultimately, by getting rid of the interference EWs, the AA-based RC-PLS-R models  
528 were further adapted to increase predictive power and applicability for NA seeds.

529 The combination of a mathematical predictive model, generated using  
530 hyperspectral imaging data, with differential data captured on ageing-related  
531 metabolites was demonstrated to be an appropriate approach to interpret and optimise  
532 the application of the AA-based vigour predictive model for NA seeds. These findings  
533 will, therefore, provide the theoretical basis for longevity/storability prediction in  
534 seeds and enable the provision of preferentially high-vigour seeds, which ultimately  
535 help to reduce post-harvest losses for farmers and improve food quality for  
536 consumers.

537

538 **Funding**

539 This work was supported by the National Natural Science Foundation of China  
540 (grant number U2005208, 31971833, 32101676), the Natural Science Foundation of  
541 Fujian Province (2021J02024) and the Biotechnology and Biological Sciences  
542 Research Council (grant number BB/W006979/1).

543

544

545 **Acknowledgement**

546 The authors are grateful for helpful comments of bioinformatics analyses from  
547 Shanghai Applied Protein Technology Co., Ltd. (Shanghai, China).

## 548 References

- 549 Ambrose, A., Kandpal, L. M., Kim, M. S., Lee, W. H., & Cho, B. K. (2016). High speed measurement  
550 of corn seed viability using hyperspectral imaging. *Infrared Physics & Technology*, *75*, 173-179.  
551 <https://doi.org/10.1016/j.infrared.2015.12.008>
- 552 Bai, J., Hao, F., Cheng, G., & Li, C. (2021). Machine vision-based supplemental seeding device for  
553 plug seedling of sweet corn. *Computers and Electronics in Agriculture*, *188*, 106345.  
554 <https://doi.org/10.1016/j.compag.2021.106345>
- 555 Bernal-Lugo, I., & Leopold, A. C. (1995). Seed stability during storage: Raffinose content and seed  
556 glassy state. *Seed Science Research*, *5*(2), 75-80. <https://doi.org/10.1017/s096025850002646>
- 557 Chauhan, H. S., Chhabra, R., Rashmi, T., Muthusamy, V., Zunjare, R. U., Mishra, S. J., Gain, N.,  
558 Mehta, B. K., Singh, A. K., Gupta, H. S., & Hossain, F. (2022). Impact of *vte4* and *crtRB1* genes  
559 on composition of vitamin-E and provitamin-A carotenoids during kernel-stages in sweet corn.  
560 *Journal of Food Composition and Analysis*, *105*, 104264.  
561 <https://doi.org/10.1016/j.jfca.2021.104264>
- 562 Chen, Y. C., & Chourey, P. S. (1989). Spatial and temporal expression of the two sucrose synthase  
563 genes in maize: immunohistological evidence. *Theoretical & Applied Genetics*, *78*(4), 553-559.  
564 <https://doi.org/10.1007/BF00290842>
- 565 Cheng, J. H., & Sun, D. W. (2015). Data fusion and hyperspectral imaging in tandem with least  
566 squares-support vector machine for prediction of sensory quality index scores of fish fillet. *LWT -*  
567 *Food Science and Technology*, *63*(2), 892-898. <https://doi.org/10.1016/j.lwt.2015.04.039>
- 568 Deborah, D., Willems, L., Arkel, J. V., Dekkers, B., Hilhorst, H., & Bentsink, L. (2016). Galactinol as  
569 marker for seed longevity. *Plant Science An International Journal of Experimental Plant Biology*,  
570 *246*, 112-118. <https://doi.org/10.1016/j.plantsci.2016.02.015>
- 571 Fan, N., Ma, X., Liu, G., Ban, J., Yuan, R., & Sun, Y. (2021). Rapid determination of TBARS content  
572 by hyperspectral imaging for evaluating lipid oxidation in mutton. *Journal of Food Composition*  
573 *and Analysis*, *103*, 104110. <https://doi.org/10.1016/j.jfca.2021.104110>
- 574 Fan, S., Li, J., Xia, Y., Tian, X., Guo, Z., & Huang, W. (2019). Long-term evaluation of soluble solids  
575 content of apples with biological variability by using near-infrared spectroscopy and calibration  
576 transfer method. *Postharvest Biology and Technology*, *151*, 79-87.  
577 <https://doi.org/10.1016/j.postharvbio.2019.02.001>
- 578 Freitas, R. A., Dias, D. C. F. S., Oliveira, G. A., Dias, L. A. S., & Jose, I. C. (2006). Physiological and  
579 biochemical changes in naturally and artificially aged cotton seeds. *Seed Science and Technology*,  
580 *34*(2), 253-264. <https://doi.org/10.15258/sst.2006.34.2.01>
- 581 He, H. J., Wu, D., & Sun, D. W. (2013). Non-destructive and rapid analysis of moisture distribution in  
582 farmed Atlantic salmon (*Salmo salar*) fillets using visible and near-infrared hyperspectral  
583 imaging. *Innovative Food Science & Emerging Technologies*, *18*, 237-245.  
584 <https://doi.org/10.1016/j.ifset.2013.02.009>
- 585 He, H. J., Chen, Y., Li, G., Wang, Y., Ou, X., & Guo, J. (2023). Hyperspectral imaging combined with  
586 chemometrics for rapid detection of talcum powder adulterated in wheat flour. *Food Control*, *144*,  
587 109378. <https://doi.org/10.1016/j.foodcont.2022.109378>
- 588 Hosomi, S. T., Custódio, C. C., Seaton, P. T., Marks, T. R., & Machado-Neto, N. B. (2012). Improved  
589 assessment of viability and germination of *Cattleya* (Orchidaceae) seeds following storage. *In*  
590 *Vitro Cellular & Developmental Biology - Plant*, *48*, 127-136. [2](https://doi.org/10.1007/s11627-011-</a></p></div><div data-bbox=)

591 9404-1

592 Kandpal, L. M., Lohumi, S., Kim, M. S., Kang, J. S., & Cho, B. K. (2016). Near-infrared hyperspectral  
593 imaging system coupled with multivariate methods to predict viability and vigor in muskmelon  
594 seeds. *Sensors and Actuators B: Chemical*, 229, 534-544.  
595 <https://doi.org/10.1016/j.snb.2016.02.015>

596 Kataki, P. K., Horbowicz, M., Taylor, A. G., & Obendorf, R. L. (1997). Changes in sucrose, cyclitols  
597 and their galactosyl derivatives with seed ageing. *Basic and Applied Aspects of Seed Biology*, 30,  
598 515-522. [https://doi.org/10.1007/978-94-011-5716-2\\_56](https://doi.org/10.1007/978-94-011-5716-2_56)

599 Kuo, T. M., Doehlert, D. C., & Crawford, C. G. (1990). Sugar metabolism in germinating soybean  
600 seeds: evidence for the sorbitol pathway in soybean axes. *Plant Physiology*, 93(4), 1514-1520.  
601 <https://doi.org/10.1104/pp.93.4.1514>

602 Lahuta, L. B., Górecki, R., Zalewski, K., & Hedley, C. L. (2007). Sorbitol accumulation during natural  
603 and accelerated ageing of pea (*Pisum sativum* L.) seeds. *Acta Physiologiae Plantarum*, 29(6), 527-  
604 534. <https://doi.org/10.1007/s11738-007-0063-0>

605 Bernal-Lugo, I., & Leopold, A. C. (1992). Changes in soluble carbohydrates during seed storage. *Plant*  
606 *Physiology*, 98(3), 1207-1210. <https://doi.org/10.1104/pp.98.3.1207>

607 Liu, Y., Pu, H., & Sun, D. W. (2017). Hyperspectral imaging technique for evaluating food quality and  
608 safety during various processes: A review of recent applications. *Trends in Food Science &*  
609 *Technology*, 69, 25-35. <https://doi.org/10.1016/j.tifs.2017.08.013>

610 Ma, T., Tsuchikawa, S., & Inagaki, T. (2020). Rapid and non-destructive seed viability prediction using  
611 near-infrared hyperspectral imaging coupled with a deep learning approach. *Computers and*  
612 *Electronics in Agriculture*, 177, 105683. <https://doi.org/10.1016/j.compag.2020.105683>

613 Merritt, D. J., Martyn, A. J., Ainsley, P., Young, R. E., Seed, L. U., Thorpe, M., Hay, F. R., Commander,  
614 L. E., Shackelford, N., Offord, C. A., Dixon, K. W., & Probert, R. J. (2014). A continental-scale  
615 study of seed lifespan in experimental storage examining seed, plant, and environmental traits  
616 associated with longevity. *Biodiversity and Conservation*, 23, 1081-1104.  
617 <https://doi.org/10.1007/s10531-014-0641-6>

618 Nansen, C., Zhao, G., Dakin, N., Zhao, C., & Turner, S. R. (2015). Using hyperspectral imaging to  
619 determine germination of native Australian plant seeds. *Journal of Photochemistry &*  
620 *Photobiology B Biology*, 145, 19-24. <https://doi.org/10.1016/j.jphotobiol.2015.02.015>

621 Obendorf, R. L., & Górecki, R. (2012). Soluble carbohydrates in legume seeds. *Seed Science Research*,  
622 22(04), 219-242. <https://doi.org/10.1017/S0960258512000104>

623 Priestley, D. A., & Leopold, A. C. (1983). Lipid changes during natural aging of soybean seeds.  
624 *Physiologia Plantarum*, 59(3), 467-470. <https://doi.org/10.1111/j.1399-3054.1983.tb04231.x>

625 Rahman, A., & Cho, B. K. (2016). Assessment of seed quality using non-destructive measurement  
626 techniques: a review. *Seed Science Research*, 26(04), 285-305.  
627 <https://doi.org/10.1017/S0960258516000234>

628 Rodo, A. B., & Marcos-Filho, J. (2003). Onion seed vigor in relation to plant growth and yield.  
629 *Horticultura Brasileira*, 21(2), 220-226. <https://doi.org/10.1590/S0102-05362003000200020>

630 Bowne, J. B., Erwin, T. A., Juttner, J., Schnurbusch, T., Langridge, P., Bacic, A., & Roessner, U.  
631 (2012). Drought responses of leaf tissues from wheat cultivars of differing drought tolerance at the  
632 metabolite level. *Molecular Plant*, 5(2), 418-429. <https://doi.org/10.1093/mp/5.2.418>

633 Shaw, J. R., & Dickinson, D. B. (1984). Studies of sugars and sorbitol in developing corn kernels.  
634 *Plant Physiology*, 75(1), 207-211. <https://doi.org/10.1104/pp.75.1.207>

635 Shrestha, S., Knapič, M., Žibrat, U., Deleuran, L. C., & Gislum, R. (2016). Single seed near-infrared  
636 hyperspectral imaging in determining tomato (*Solanum lycopersicum* L.) seed quality in  
637 association with multivariate data analysis. *Sensors & Actuators B Chemical*, *237*, 1027-1034.  
638 <https://doi.org/10.1016/j.snb.2016.08.170>

639 Singh, I., Langyan, S., & Yadava, P. (2014). Sweet corn and corn-based sweeteners. *Sugar Tech*, *16*,  
640 144-149. <https://doi.org/10.1007/s12355-014-0305-6>

641 Sun, W. Q., & Leopold, A. C. (1995). The Maillard reaction and oxidative stress during aging of  
642 soybean seeds. *Physiologia Plantarum*, *94*(1), 94-104. <https://doi.org/10.1111/j.1399-3054.1995.tb00789.x>

644 Wakholi, C., Kandpal, L. M., Lee, H., Bae, H., Park, E., Kim, M. S., Mo, C., Lee, W., & Cho, B.  
645 (2018). Rapid assessment of corn seed viability using short wave infrared line-scan hyperspectral  
646 imaging and chemometrics. *Sensors and Actuators B Chemical*, *255*, 498-507.  
647 <https://doi.org/10.1016/j.snb.2017.08.036>

648 Wang, C., Liu, B., Liu, L., Zhu, Y., & Li, X. (2021). A review of deep learning used in the  
649 hyperspectral image analysis for agriculture. *Artificial Intelligence Review*, *54*, 5205-5253.  
650 <https://doi.org/10.1007/s10462-021-10018-y>

651 Yan, S., Huang, W., Gao, J., Fu, H., & Liu, J. (2018). Comparative metabolomic analysis of seed  
652 metabolites associated with seed storability in rice (*Oryza sativa* L.) during natural aging. *Plant  
653 Physiology & Biochemistry*, *127*, 590-598. <https://doi.org/10.1016/j.plaphy.2018.04.020>

654 Yang, J., Sun, L., Xing, W., Feng, G., & Wang, J. (2021). Hyperspectral prediction of sugarbeet seed  
655 germination based on gauss kernel SVM. *Spectrochimica Acta Part A: Molecular and  
656 Biomolecular Spectroscopy*, *253*, 119585. <https://doi.org/10.1016/j.saa.2021.119585>

657 Yokoi, M., & Sait, N. (1973). Light absorption patterns of intact Rosa flowers in relation to the flower  
658 colour. *Phytochemistry*, *12*(7), 1783-1786. [https://doi.org/10.1016/0031-9422\(73\)80402-9](https://doi.org/10.1016/0031-9422(73)80402-9)

659 Zalewski, K., & Lahuta, L. B. (1998). The metabolism of ageing seeds: changes in the raffinose family  
660 oligosaccharides during storage of field bean (*Vicia faba* var. *minor Harz*) seeds. *Acta Societatis  
661 Botanicorum Poloniae*, *67*(2), 193-196. <https://doi.org/10.5586/asbp.1998.022>

662 Zhang, B., Huang, W., Li, J., Zhao, C., Fan, S., Wu, J., & Liu, C. (2014). Principles, developments and  
663 applications of computer vision for external quality inspection of fruits and vegetables: a review.  
664 *Food Research International*, *62*, 326-343. <https://doi.org/10.1016/j.foodres.2014.03.012>

665 Zhang, T., Ayed, C., Fisk, D. I., Pan, T., Wang, J., Yang, N., & Sun, Q. (2022). Evaluation of volatile  
666 metabolites as potential markers to predict naturally-aged seed vigour by coupling rapid analytical  
667 profiling techniques with chemometrics. *Food Chemistry*, *367*, 130760.  
668 <https://doi.org/10.1016/j.foodchem.2021.130760>

669 Zhang, T., Fan, S., Xiang, Y., Zhang, S., Wang, J., & Sun, Q. (2020). Non-destructive analysis of  
670 germination percentage, germination energy and simple vigour index on wheat seeds during  
671 storage by Vis/NIR and SWIR hyperspectral imaging. *Spectrochimica Acta Part A: Molecular and  
672 Biomolecular Spectroscopy*, *239*, 118488. <https://doi.org/10.1016/j.saa.2020.118488>

673 Zhang, W., Pan, L., & Lu, L. (2023). Prediction of TVB-N content in beef with packaging films using  
674 visible-near infrared hyperspectral imaging. *Food Control*, *147*, 109562.  
675 <https://doi.org/10.1016/j.foodcont.2022.109562>

676  
677  
678

679 **Figure Captions**

680 **Fig. 1. Schematic diagram of hyperspectral imaging system.**

681 **Fig. 2. Flow diagram of data sets constructed by sweetcorn seeds during AA and**  
682 **NA.** GP: Germination percentage; AA: Artificial ageing; NA: Natural ageing.

683 **Fig. 3. Reflectance spectra of Lvsechaoren (A-C) and Zhongtian 300 (D-F)**  
684 **sweetcorn seeds with various vigour levels during AA and NA.** (A) Spectral curves  
685 of all seeds from Lvsechaoren. (B) Average spectral curves of each vigour level for  
686 Lvsechaoren seeds. (C) Spectral curves in the 600~900 nm range of (B). (D) Spectral  
687 curves of all seeds from Zhongtian300. (E) Average spectral curves of each vigour  
688 levels for Zhongtian300 seeds. (F) Spectral curves in the 600~900 nm range of (E).  
689 For Lvsechaoren, Raw 85%: raw seeds with 85% germination percentage; AA 75%,  
690 AA 65% and AA 45%: accelerated aged seeds with 75%, 65% and 45% germination  
691 percentage; NA 75%, NA 65% and NA 45%: naturally aged seeds with 75%, 65% and  
692 45% germination percentage. For Zhongtian 300, Raw 92%: raw seeds with 92%  
693 germination percentage; AA 83%, AA 75% and AA 65%: accelerated aged seeds with  
694 83%, 75% and 65% germination percentage; NA 83%, NA 75% and NA 65%:  
695 naturally aged seeds with 83%, 75% and 65% germination percentage.

696 **Fig. 4. Regression coefficients curves from SNV-PLS-R model of Lvsecchaoren**  
697 **(A) and FD-SNV-PLS-R model of Zhongtian 300 (B)**

698 **Fig. 5. Metabolomics analysis of sweetcorn seeds during AA and NA.** (A) Scores  
699 scatter plot of principal component analysis (PCA) in the positive ionization mode  
700 (PIM). (B) Scores scatter plot of PCA in the negative ionization mode (NIM). (C)  
701 Venn diagrams of the identified differential metabolites in all pairwise comparisons in  
702 the PIM. (D) Venn diagrams of the identified differential metabolites in all pairwise  
703 comparisons in the NIM. (E) Cluster heatmap analysis of the overlapping metabolites  
704 with synchronous change in NA and AA seeds. (F) Cluster heatmap analysis of the  
705 overlapping metabolites with asynchronous change in NA and AA seeds. Raw: raw  
706 seeds with 85% germination percentage; NA: naturally aged seeds with 75%

707 germination percentage; AA: accelerated aged seeds with 75% germination  
708 percentage. The solid lines in PCA plots encircling the points define the 95%  
709 confidence intervals for all groups.

710 **Fig. 6. OPLS-DA analysis of seed samples from RAW 85% vs. AA 75% (A, D),**  
711 **RAW 85% vs. NA 75% (B, E) and AA 75% vs. NA 75% (C, F) in the PIM and**  
712 **NIM.**

713 **Fig. 7. Structural formulas of typical differential metabolites containing N-H**  
714 **bonds.**  
715

716 **Table 1 Germination percentage distribution of sweetcorn seeds in calibration set**  
 717 **and validation set**

Varieties	Data sets	No.	Minimum (%)	Maximum (%)	Mean (%)	Standard deviation (%)
	Calibration set	64	40	92	67.9	15.0
Lvsechaoren	Validation set -1	36	42	90	67.6	14.7
	Validation set -2	36	42	90	68.1	14.0
Zhongtian 300	Calibration set	64	60	96	78.3	9.5
	Validation set -1	36	62	96	78.6	9.8
	Validation set -2	36	62	96	80.6	9.9

718 Abbreviations: Number of samples (No.)

719

720 **Table 2 Prediction performance of PLS-R models for sweetcorn seeds of**  
721 **Lvsecchaoren and Zhongtian 300 based on full wavelengths in validation set-1**  
722 **and validation set-2**

Cultivars	Pre-treatments	LV	Calibration set (AA)		Cross-Validation		Validation set-1 (AA)		Validation set-2 (NA)	
			$R^2_c$	RMSE <sub>C</sub>	$R^2_{cv}$	RMSE <sub>CV</sub>	$R^2_{v-1}$	RMSE <sub>V-1</sub>	$R^2_{v-2}$	RMSEV <sub>-2</sub>
Lvsechaoren	None	8	0.903	0.046	0.837	0.060	0.908	0.047	0.486	0.156
	AS	9	0.921	0.042	0.849	0.058	0.904	0.048	0.576	0.126
	FD	8	0.950	0.033	0.860	0.057	0.889	0.049	0.679	0.145
	<b>SNV</b>	<b>10</b>	<b>0.939</b>	<b>0.037</b>	<b>0.871</b>	<b>0.054</b>	<b>0.905</b>	<b>0.049</b>	<b>0.513</b>	<b>0.127</b>
	FD+SNV	9	0.954	0.032	0.844	0.060	0.898	0.047	0.676	0.141
Zhongtian 300	None	10	0.906	0.029	0.832	0.039	0.850	0.038	0.332	0.109
	AS	10	0.922	0.026	0.841	0.038	0.814	0.043	0.285	0.109
	FD	7	0.906	0.029	0.841	0.038	0.851	0.038	0.279	0.117
	SNV	9	0.900	0.030	0.831	0.039	0.856	0.038	0.546	0.098
	<b>FD+SNV</b>	<b>7</b>	<b>0.908</b>	<b>0.029</b>	<b>0.846</b>	<b>0.037</b>	<b>0.838</b>	<b>0.040</b>	<b>0.326</b>	<b>0.114</b>

723 Abbreviations: Natural ageing (NA); Artificial ageing (AA); Latent variable (LV); Partial  
724 least squares regression (PLS-R); Root mean square error of calibration (RMSEC); Root  
725 mean square error of cross validation (RMSECV); Root mean square error of validation  
726 (RMSEV); Autoscaling (AS); Savitzky–Golay first derivative (FD); Standard normal variate  
727 (SNV).

728

729 **Table 3 Prediction performance of PLS-R models for sweetcorn seeds of**  
730 **Lvsechaoren and Zhongtian 300 based on optimal wavelengths in validation set-**  
731 **1 and validation set-2**

Cultivars	Pre-treatments	LV	Calibration set (AA)		Cross-Validation		Validation set-1 (AA)		Validation set-2 (NA)	
			$R^2_c$	RMSEC	$R^2_{cv}$	RMSECV	$R^2_{v-1}$	RMSEV-1	$R^2_{v-2}$	RMSEV-2
Lvsechaoren	None	7	0.920	0.042	0.877	0.052	0.939	0.043	0.367	0.165
	AS	8	0.931	0.039	0.872	0.053	0.921	0.047	0.461	0.158
	FD	10	0.921	0.042	0.861	0.056	0.929	0.043	0.682	0.122
	<b>SNV</b>	<b>9</b>	<b>0.942</b>	<b>0.036</b>	<b>0.882</b>	<b>0.052</b>	<b>0.909</b>	<b>0.050</b>	<b>0.586</b>	<b>0.140</b>
	FD+SNV	9	0.917	0.043	0.865	0.055	0.920	0.048	0.696	0.108
Zhongtian 300	None	10	0.870	0.034	0.787	0.044	0.823	0.041	0.431	0.091
	<b>AS</b>	10	0.874	0.034	0.780	0.045	0.820	0.042	0.393	0.087
	FD	11	0.850	0.037	0.780	0.045	0.875	0.036	0.553	0.078
	SNV	9	0.863	0.035	0.788	0.044	0.846	0.039	0.668	0.073
	<b>FD+SNV</b>	<b>10</b>	<b>0.852</b>	<b>0.037</b>	<b>0.796</b>	<b>0.043</b>	<b>0.866</b>	<b>0.036</b>	<b>0.561</b>	<b>0.076</b>

732 Abbreviations: As shown in Table 2.

733 **Table 4 Prediction performance of Lvsechaoren and Zhongtian 300 PLS-R**  
 734 **models based on EWs without N–H bonds in validation set-1 and validation set-2**

Cultivars	Pre-treatments	LV	Calibration set (AA)		Cross-Validation		Validation set-1 (AA)		Validation set-2 (NA)	
			$R^2_c$	RMSEC	$R^2_{cv}$	RMSECV	$R^2_{v-1}$	RMSEV-1	$R^2_{v-2}$	RMSEV-2
Lvsechaoren	None	7	0.911	0.044	0.869	0.054	0.896	0.051	0.560	0.160
	AS	8	0.915	0.043	0.874	0.053	0.919	0.049	0.720	0.117
	FD	8	0.895	0.048	0.837	0.060	0.912	0.048	0.599	0.147
	SNV	7	0.911	0.044	0.855	0.057	0.914	0.047	0.644	0.148
	FD+SNV	8	0.885	0.050	0.830	0.062	0.908	0.049	0.653	0.124
Zhongtian 300	None	9	0.856	0.036	0.783	0.044	0.848	0.039	0.414	0.087
	AS	11	0.875	0.034	0.782	0.045	0.824	0.041	0.571	0.077
	FD	11	0.849	0.037	0.772	0.046	0.850	0.038	0.600	0.073
	SNV	9	0.858	0.036	0.781	0.045	0.842	0.039	0.727	0.071
	FD+SNV	10	0.847	0.037	0.786	0.044	0.838	0.040	0.570	0.073

735 Abbreviations: As shown in Table 2.

736