

1 **Evaluation of Postharvest Senescence in Broccoli Via Hyperspectral Imaging**

2

3

4 Xiaolei Guo^{1,¶}, Yogesh Ahlawat^{2,¶}, Alina Zare^{1*}, Tie Liu^{2*}

5

6

7

8 ¹ University of Florida, Department of Electrical and Computer Engineering, Gainesville, Florida

9 United States of America

10 ² University of Florida, Horticultural Science Department, Gainesville, Florida, United States of
11 America

12

13

14 * Corresponding author

15 Email: azare@ece.ufl.edu (AZ) and tieliu@ufl.edu (TL) (“AZ and TL are Joint Senior Authors”)

16

17

18

19 ¶ These authors contributed equally to this work

20

21

22

23

24 **ABSTRACT**

25

26 Fresh fruits and vegetables are invaluable for human health, but their quality deteriorates
27 before reaching consumers due to ongoing biochemical processes and compositional changes.
28 The current lack of any objective indices for defining “*freshness*” of fruits or vegetables limits
29 our capacity to control product quality leading to food loss and waste. It has been hypothesized
30 that certain proteins and compounds such as glucosinolates can be used as an indicator to
31 monitor the freshness of vegetables and fruits. However, it is challenging to “visualize” the
32 proteins and bioactive compounds during the senescence processes. In this work, we
33 propose machine learning hyperspectral image analysis approaches for estimating glucosinolates
34 levels to detect postharvest senescence in broccoli. Therefore, we set out the research to quantify
35 glucosinolates as “freshness-indicators” which aid in the development of an innovative and
36 accessible tool to precisely estimate the freshness of produce. Such a tool would allow for
37 significant advancement in postharvest logistics and supporting the availability for high-quality
38 and nutritious fresh produce.

39

40 **Introduction**

41

42 Broccoli (*Brassica oleracea* L. var. *italica*) is a nutritious vegetable that is well enriched in
43 anti-cancerous chemical compounds like glucosinolates [1]. Broccoli is highly perishable and
44 senesces quickly after harvest. Broccoli is usually harvested at a developing stage of

45 inflorescence therefore causes stress-induced senescence. The stress-induced senescence leads to
46 faster chlorosis and increases in proteases resulting in the dismantling of chloroplast number and
47 component [2]. Thus, during the postharvest storage and transportation, broccoli florets starts to
48 turn yellow accompanied with a decrease in nutritional quality [1].

49 Senescence is a developmental process that can be tracked by monitoring physiological
50 and biochemical changes in transcripts, proteins, and metabolites in broccoli. One particularly
51 interesting class of phytochemicals are glucosinolates, given their importance not only for plant
52 protection, but also for their dietary significance as chemo-preventative compounds that are
53 found in edible plants, (i.e., cruciferous crops). However, the actual quality, storability and
54 overall “freshness” of broccoli postharvest is quite uncertain unless the changes are visible to the
55 human eye. Detecting any deteriorative physiological signs and symptoms before any
56 irreversible damage occurs could allow for the development of freshness indicators, which can
57 be used to identify the best postharvest handling, process and storage practices [3]. Therefore,
58 sensitive indicators of potential deterioration are essential for extending optimal postharvest
59 quality.

60 Prior work to detect symptoms of degradation in quality include using color
61 measurements at an early postharvest stages [4]. Chlorophyll fluorescence and RGB (red, green,
62 blue) color imagery analysis were used to measure the pigment change in broccoli using
63 fluorescence and inverse red channel for color measurements [5]. However, objectively measuring
64 the progressive loss of freshness after harvest has been a heretofore intractable problem in
65 postharvest handling of fresh produce. Until now, determining freshness has been mostly based
66 on external visual criteria like wilting, shriveling and color changes, which is laborious, time-
67 consuming, and subjective [5].

68 The rapid advancement of optical sensors and imaging technologies has significantly
69 impacted agriculture and brought about more automation [6]. Initially, imaging techniques used
70 red-green-blue color systems for the identification of color change and defects in the food and
71 agricultural products [7]. Since, multispectral fluorescence imaging has been used with maize,
72 peas, soybean for measuring color change [8].

73 Hyperspectral Imaging (HSI) has also evolved over the years and is being explored as
74 technique for nondestructive food analysis. HSI provides both spatial and spectral information
75 about an object. HSI consists of many thousands of pixels in a two-dimensional array, with each
76 pixel containing a spectrum corresponding to a specific region on the surface of the sample.
77 These spectra vary according to unique material and chemical compositions. Interrogation of
78 these spectra makes possible the development of mathematical models to estimate the chemical
79 composition or functional class of a sample associated with each pixel. Results reported in
80 several studies have indicated that hyperspectral imaging is able to predict a number of food
81 components and quality parameters in a wide range of biological matrices [5]. Previous research
82 has shown that HSI was used in a plethora of applications in agriculture and food industry to
83 measure the textural changes including bruise, chilling injury, firmness [9,10,11], and
84 biochemical detection such as moisture content, soluble solids content, acidity, and phenolics
85 [12, 13,14], biosafety measurement in bacterial or fungal infection and fruit-fly infestation
86 [15,16]. In addition, photosynthetic rate in mangroves was studied in relation to salinity stress
87 using HSI technology [5]. More importantly, Hernandez et al. [17] reported that hyperspectral
88 imaging can quantify the localization and total glucoraphanin in florets and stalks of broccoli.
89 Moreover, the low instrument cost and fast-detecting properties of HSI have enabled the

90 development of powerful diagnostic tools for detecting, classifying, and visualizing quality and
91 safety attributes of fresh produce.

92 This study was conducted to evaluate the potential significance of hyperspectral imaging as
93 a tool to determine the freshness of broccoli during postharvest storage. Through High
94 Performance Liquid Chromatography (HPLC) quantitation, we showed that there is a linear
95 correlation between the total glucosinolates concentration and post-harvest senescence in
96 broccoli. Moreover, we performed Real-Time Polymerase Chain Reaction (PCR) for expression
97 studies on 13 enzymes are involved in the biochemical pathway producing glucosinolates. We
98 believe that combination of HSI and glucosinolates analysis can define the freshness signature in
99 postharvest broccoli.

100

101 **Materials and methods**

102

103 **Tissue collection and Preparation**

104

105 Freshly grown broccoli (cultivar, Emerald Crown) florets were manually harvested from
106 local farms in Hastings, Florida to avoid any mechanical damage. All the broccoli florets were
107 selected with the same shape and size in this study. The florets were then stored in either a cold
108 room (4–5°C, darkness) for the cold treatment, or in a plant growth chamber at 25°C (RT) with
109 16 hours of light and 8 hours of dark. Four biological replicate of broccoli florets were used for
110 the experimental sampling for tissue collections. Tissue samples were collected from the broccoli
111 florets every other day during a twelve-day period. The samples were wrapped in aluminum foil,

112 immediately frozen in liquid nitrogen and then stored in -80°C for quantitative PCR and freeze-
113 dried for HPLC.

114

115 **Hyperspectral imaging setup**

116

117 Broccoli hyperspectral images were collected with a HinaLea 4200 hyperspectral camera
118 and converted to reflectance spectra. The camera covers the wavelengths ranging from 400nm to
119 1000nm with a resolution at 4nm (based on full width at half maximum), resulting in 300
120 wavebands. Halogen lamps were used as illuminators in an imaging chamber. Within the
121 chamber, each broccoli sample was placed on a black plate with matte black siding to absorb
122 redundant light in order to minimize scattering. Hyperspectral measurements of three biological
123 replicates were carried out at every time point and the experiment was repeated three times.

124

125 **Pre-processing Reflectance Spectra**

126

127 Pre-processing of the measured spectra for noise and illumination variation was required
128 prior to analysis. The measured reflectance spectra consistently contained higher levels of noise
129 at the two ends of the wavelength range. In addition, since the illuminator used was a point light
130 source and did not evenly cover the entire imaging surface. As a result, the center of the imaging
131 plane was brighter than the outer edge. Fig. 1 (d) shows examples for the issues mentioned
132 above. To mitigate these issues, first, a median filter of length 5 along the wavelength axis was
133 applied to the reflectance spectra. Next, responses below 500 nm and above 900 nm were
134 removed due to the high noise levels. Finally, in order to reduce the magnitude difference caused

135 by point light source, we applied l_2 normalization in which each spectral signature is normalized
136 to have unit norm.

137

138 **Segmenting Floret from Background**

139

140 After pre-processing spectra, the regions of florets were segmented from the remainder of
141 the hyperspectral cube. This segmentation was accomplished in two steps: (1) segmenting the
142 broccoli sample from the black background; (2) segmenting the floret from the stalk. For the
143 first stage, Fig. 1 (a) and (d)-(e) illustrates the spectral differences between the broccoli and the
144 black background. Namely, the broccoli spectra have a bump around 550nm wavelength (visible
145 bands of green) and a sharp increase around 700nm wavelength (Near infrared/Red edge whereas
146 the spectra for the black background are nearly flat up to 800nm and then increases rapidly.
147 Given these significant spectral differences, the k-means clustering algorithms was applied to
148 cluster the pixel spectra into two groups, broccoli and background. After clustering, a
149 morphological image closing operation was applied to connect any disconnected points. An
150 example of the segmented results is shown in Fig. 1 (b).

151

152 **Fig 1. RGB image and spectra.** (a) Broccoli sample placed on a black plate. (b) Segmented
153 broccoli sample. (c) Segmented broccoli floret. (d) Reflectance measured by HinaLea 4200
154 hyperspectral camera. (e) Preprocessed spectra. The color of spectra in the (d)-(e) is
155 corresponding to the points in (a).

156

157 Since the glucosinolates concentration was measured on broccoli spores, we hypothesized
158 that focusing on the spectra of broccoli flower will generate stronger correlation than analyzing
159 the spectra of entire broccoli. In order to segment the floret from the stalk, we applied the
160 GraphCut algorithm [18] of the image segmentation toolbox in Matlab 2019b [19]. The
161 algorithm was seeded by the user providing a marking that denotes the broccoli flower and the
162 background including broccoli pedicle. The segmentation took around 5 to 10 seconds for each
163 image. The segmented results are shown in Fig. 1 (c).

164

165 **Unmixing and correlation with glucosinolates information**

166

167 The hyperspectral image collects a high-dimensional image cube that describes each pixel
168 as the radiance or the reflectance at a range of wavelengths across the electromagnetic spectrum.
169 The spectrum of a pixel is usually determined by the material of the object surface. With the
170 assumption that the measured spectrum is consists of a set of constituent spectra, also known as
171 endmembers, spectral unmixing is defined as decomposing the mixed spectrum into a collection
172 of endmembers and their corresponding proportions, also known as abundances [20]. A well-
173 known spectral model (and the most commonly applied to perform hyperspectral unmixing) is
174 the linear mixing model (LMM) which represents each measured spectrum as a convex
175 combination of endmembers as shown in Equation 1 [20],

176

$$177 \quad s_i = \sum_{k=1}^M a_{ik}e_k + \varepsilon_i \text{ such that } \sum_{k=1}^M a_{ik} = 1, a_{ik} \in [0,1] \quad (1)$$

178

179 where s_i is the spectrum of pixel i , ϵ_i is the noise vector, M is the number of endmembers, e_k is
180 k^{th} endmember, and a_{ik} is the corresponding abundance value. The objective of each unmixing
181 broccoli sample is to find a set of endmembers and abundances that can represent the freshness
182 of broccoli. Thus, in this work, we attempt to estimate endmembers that represent the range of
183 “freshness” levels in the samples. Then the associated abundances for the endmembers
184 corresponding to “fresh” samples can use viewed as a freshness indicator for the sample. The
185 endmembers and abundances were estimated using two approaches averaging (as described
186 below in Section A) or the application of the SPICE algorithm (as described in Section B).

187 **A. Unmixing with averaging spectra as endmembers**

188 In the averaging approach, the endmembers were extracted by averaging the broccoli spectrum
189 from two extreme storage conditions. Specifically, e_1 is the average from broccoli on day 1,
190 representing the best fresh level. e_2 is the average from broccoli kept under room temperature
191 for 12 days, representing the least fresh level. The greater a_{i1} indicates more fresh level of pixel
192 i in the broccoli sample.

193

194 The abundance values were then estimated by optimizing the fully constrained least squares of
195 each pixel with the above two endmembers as shown in Equation 2 [22].

196

$$197 \operatorname{argmin}_{a_{ik}} \sum_{i=1}^N |s_i - \sum_{k=1}^M a_{ik} e_k|^2 \text{ such that } \sum_{k=1}^M a_{ik} = 1, a_{ik} \in [0,1] \quad (2)$$

198

199 where $M = 2$ in this case.

200

201 **B. Unmixing with SPICE algorithm**

202

203 Instead of extracting average spectra as endmembers, the sparsity promoting iterated
204 constrained endmember (SPICE) algorithm benefits from simultaneous estimating the shape and
205 number of endmembers as well as the abundances [21]. The algorithm is initialized with a large
206 number of endmembers and iteratively updates the estimated endmembers and abundances by
207 optimizing Equation 3 using an alternating optimization approach,

208

$$209 \operatorname{argmin}_{\mathbf{a}_{ik}, \mathbf{e}_k} \left[\frac{1-\mu}{N} \sum_{i=1}^N \|\mathbf{s}_i - \sum_{k=1}^M \mathbf{a}_{ik} \mathbf{e}_k\|^2 + \mu V + SPT \right], \quad SPT = \sum_{k=1}^M \frac{\Gamma}{\sum_{i=1}^N \mathbf{a}_{ik}} \sum_{i=1}^N \mathbf{a}_{ik} \quad (3)$$

210

211 where V is the sum of variances among estimated endmembers, μ is the regularization parameter
212 to balance the reconstruction error and variance, Γ is a constant that decide the proportion values
213 are driven to zero, and \mathbf{a}'_{ik} is the abundance value for the i th pixel in the k th endmember from
214 the previous iteration. After each iteration, endmembers which are not being used to represent
215 the data are pruned from the overall endmember set¹.

216

217 **C. Correlating abundance with glucosinolates concentration level**

218

219 The estimated abundance vectors $\hat{\mathbf{a}} = \{\hat{a}_1, \hat{a}_2, \dots, \hat{a}_k\} \in \mathbb{R}^{1 \times M}$ for each broccoli sample (where
220 the value \hat{a}_k is the average abundance of all pixels over the region of interest as $\hat{a}_k = \frac{1}{N} \sum_{i=1}^N \mathbf{a}_{ik}$,
221 where N denotes the number of pixels) were used to attempt to predict measured glucosinolates
222 concentration values. Specifically, a multi-variable linear regression (MLR) model as Equation

¹ The Matlab and Python implementation for SPICE can be found here: github.com/GatorSense

223 4 was fit using least squares estimation approaches to predict the glucosinolates concentration
224 value,

225

$$226 \text{ glucosinolate} \sim b + \sum_{k=1}^M w_k \hat{a}_k \quad (4)$$

227

228 where b is the bias, w_k is the coefficient for \hat{a}_k , M is the number of estimated endmembers.

229

230

231 **Extraction of total glucosinolates for HPLC quantification**

232

233 HPLC-UV analysis of total glucosinolates was extracted according to previously defined
234 methods with some modifications [22]. Raw materials from previously harvested broccoli
235 florets were stored at -80°C and further samples were taken out to be dispersed in liquid
236 nitrogen. 100 mg of samples was weighed and crushed to fine powder using mortar and pestle.
237 Subsequently, crushed tissue powder was dissolved in 1 ml of 50 % methanol contained in the
238 1.5 ml eppendorf tube. The tubes were further kept at 65°C for 1 hour in the water heating bath.
239 Samples were then centrifuged at 15000 g for 10 minute. The supernatant was filtered through a
240 0.22um hydrophilic PTFE syringe filter (Sigma Aldrich, USA). HPLC analysis was done with
241 the flow rate of 0.4 ml/min at a column temperature of 40°C with a wavelength of 227 nm. The
242 solvent used were water and 100 % acetonitrile. The individual glucosinolates were estimated by
243 their HPLC peak area with reference to a desulfo-sinigrin method [22]. Total peak area was
244 calculated from broccoli florets from day 1, 3, 5, 7, 9, 11 when stored at growth chamber (25°C)

245 and cold temperature (4°C). Four biological replicate were used for each time point during the
246 study.

247

248 **RNA extraction and gene expression studies using Quantitative real-** 249 **time PCR**

250

251 Florets tissue samples from day 1, 3, 5 were chosen for glucosinolates pathway expression
252 analysis. Total RNA was isolated from broccoli floret tissue stored in liquid nitrogen using
253 TRIZOL (Ambion, life Technologies) method and DNase treatment (Turbo DNA free, Thermo-
254 fisher). First strand cDNA synthesis with 1µg of total RNA was performed using reverse
255 transcription kit (Applied Biosystems). For quantitative real-time RT-PCR, primers were
256 designed using Primer Quest, Integrated DNA Technologies (IDT) software. The primers of
257 glucosinolates pathway genes were listed in Table 1. Real time PCR reaction was performed in
258 Applied Biosystems qPCR machine (Thermofisher). Total reaction was 10 µl for each gene in
259 triplicates with thermocycler conditions as: 95°C for 10 min, 45 cycles for 95°C for 30 sec, 60°C
260 for 30 sec. Relative gene expression was calculated by Δ CT method. Actin 2 was used as an
261 internal control. This experiment was repeated twice.

262

263 **Table 1 : List of primers for quantitative PCR performed for glucosinolates pathway in**
264 **Broccoli**

Glucosinolates biosynthetic genes	Sequence
BO_ACTIN2-FORWARD	TGGTCGTGACCTTACTGACTAT

BO_ACTIN2-REVERSE	TCACTTGTCGTCGGGTAAT
BO_ST5B2-FORWARD	CCCATATACCCAACGGGTCG
BO_ST5B2-REVERSE	CCCATGAACTCAGCCAACCT
BO_MAM1-FORWARD	GGAATTATCCCTACCACCAGTTC
BO_MAM1-REVERSE	CAGAGGAGCAACATGAGATGAG
BO_CYP79F1-FORWARD	GTTAGGACAAGCGGAGAAAGA
BO_CYP79F1-REVERSE	CCATCAATGTTCCAACCTCTAAAC
BO_AOP2-FORWARD	GTGAGGAGTGATGTCCGTAAAG
BO_AOP2-REVERSE	GCCTCAACTGGTAACTCGAAA
BO_ESM1-FORWARD	CCGGAAGTAGCGTTGTTTACT
BO_ESM1-REVERSE	GTTAGGGTCGTCAAGGGATTT
BO_MAM3-FORWARD	ATCGTCCGTACAACAAGTCATC
BO_MAM3-REVERSE	GTATGTACTCTGGCCACCTTTC
BO_ESP-FORWARD	AGGACGATCGAGGCCTATAA
BO_ESP-REVERSE	GAATCCAGCTCCACCTCTTT
BO_FMOGSOX1-FORWARD	GGATTAATAGCGGCCAGAGAG
BO_FMOGSOX1-REVERSE	GCGGGTCCGATTTCAGATTTA

265

266

267 **Results**

268

269 **Use HSI as tool to detect postharvest senescence in broccoli**

270
271 First, spectra were downsampled to 5,000 per segmented broccoli sample via k-means
272 clustering to accelerate computing. Downsampling was applied to the two segmentation
273 scenarios being considered: (1) entire broccoli, and (2) broccoli florets, respectively. Next, the
274 entire dataset was split into training, validation and testing sets. To be more specific, 48 samples
275 under 12 conditions (2 storage conditions over 6 time stages) were randomly divided into 4 folds,
276 one for testing, and the other 3 for training and validation. Each fold contains 12 samples, 1
277 replication from each condition. The training and validation dataset were shuffled in every
278 repetition.

279
280 In the first step, two endmembers were calculated by averaging spectra from the most and least
281 fresh replication in the training folds. The glucosinolates concentration and derived abundance
282 feature for all training replications were applied to fit the MLR model. The training process was
283 repeated for 10 times over 3 folds. In each repetition, we tested the trained model on the testing
284 fold, calculated the mean and standard deviation of the testing and training prediction error in
285 Table 2. In addition, a model was selected according to the root means square error (RMSE) and
286 R-squared value from training and validation folds and was applied to the testing fold to generate
287 the result shown in Fig. 2.

288

289 **Table 2. Comparison of glucosinolates prediction error from endmember abundances**

		Average Spectra		SPICE	
		RMSE	R2	RMSE	R2
Entire	Training&Validation	48.12 ± 1.50	0.68 ± 0.02	44.93 ± 1.37	0.72 ± 0.02

Broccoli	Testing	28.81 ± 3.13	0.82 ± 0.04	26.14 ± 4.87	0.85 ± 0.06
Broccoli	Training&Validation	47.36 ± 1.52	0.69 ± 0.02	44.78 ± 1.23	0.72 ± 0.02
Florets	Testing	26.72 ± 1.82	0.85 ± 0.03	21.34 ± 3.26	0.90 ± 0.03

290

291 **Fig 2. Predicted glucosinolates on testing fold.** x-axis indicates the real glucosinolates
 292 concentration, y-axis indicates the prediction. Markers in various shape and color denotes
 293 prediction with different methods. Markers that closer to the “x = y” line indicates more accurate
 294 prediction.

295

296 Next, endmembers were estimated from training folds via SPICE. Since the estimated
 297 endmembers highly depends on the parameter Γ , a various range of Γ , starting from 10 to 150
 298 with stepsize 10, were explored. We conducted 10 repetitions over 3 folds for 15 Γ values.
 299 Similar to the first stage, with the estimated endmembers, the abundance feature can be derived
 300 from the training folds to fit the MLR model. Fig. 3 (a-b), (f-g) shows greater prediction error on
 301 validation folds with a smaller Γ , which indicates over-fitting of the approach. Namely, Since Γ
 302 determines the proportion of estimated endmembers to be eliminated, a smaller Γ results in a
 303 greater number of endmembers and more parameters that need to be estimated (and provide
 304 opportunity for overfitting). Fig. 3 (c-d), (h-i) illustrate the tendency of over-fitting with
 305 increasing number of endmembers M . In addition, according to Fig. 3 (e), (j), $M = 3$ was
 306 determined for both segmentation, since it is the most number over 450 replications. We tested
 307 the trained MLR models with estimated endmembers on the test fold, generating the prediction
 308 errors shown in Table 2. A model was selected via the same criterion from repetitions and was
 309 applied to the same test fold as the first stage. The prediction results are shown in Fig. 2.

310

311 **Fig 3. Exploration of SPICE parameters.** (a-e) The training and validation error across
312 various parameter setting for broccoli florets. (f-j) The training and validation error across
313 various parameter setting for entire broccoli replicate.

314

315 The visualization of unmixing result is shown in Fig. 4, where (a-b) plot the estimated
316 endmembers with average and SPICE on broccoli florets. Correspondingly, (c-d) visualize the
317 estimated abundance map for testing samples in day 1, day 5, and day 12, and (e-f) plot the
318 histogram of abundance value. The distribution of abundance and the number of pixels
319 associated with each endmember is informative of freshness over days. In addition, the
320 abundance map for the 3rd endmember in SPICE shows a relatively high concentration in day 5.
321 It would be worth to explore whether the 3rd endmember can reveal a transition status during the
322 progress of decay.

323

324 **Fig 4. Estimated endmembers and unmixing results of testing samples in day 1, 5 and 12.**

325 (a-b) Estimated endmembers with average and SPICE on broccoli florets. (c-d) Visualization of
326 estimated abundance map for testing samples in day 1, day 5, and day 12. (e-f) Histogram of
327 abundance value.

328

329 **Indo-glucosinolate peaks increased during the postharvest**

330 **senescence**

331

332 To access the possibility that the glucosinolates can be applied as senescence indicator, we
333 performed HPLC analysis to measure the total glucosinolates concentration during the
334 postharvest stored broccoli. In HPLC analysis, we monitored total peak area for indole-
335 glucosinolates, the most widely distributed glucosinolates, at six time points during a 12-day of
336 period for the broccoli florets stored at room temperature (25°C) and cold temperature (4°C).
337 Four biological replicates were analyzed at each timepoint. We found that there is a linear
338 increase in indo-glucosinlates concentration throughout a 12-day period in broccoli when stored
339 at room temperature (Fig 5). However, this changes were not observed during the cold storage
340 treatment in broccoli. Thus, there is a strong correlation observed between the indo-
341 glucosinolates levels and progression of postharvest senescence in broccoli. This data suggested
342 that indo-glucosinlates can potentially serve as an ‘freshness indicator’ to define a freshness
343 signature.

344

345 **Fig 5 . Quantification of the indo-glucosinolates peak area by HPLC.** Graph displayed the
346 changes in the indo-glucosinolates level in the broccoli florets on day 1, 3, 5, 7, 9, 11. Data
347 represented means \pm SE bars (n=4 for each day).

348

349 **Glucosinolates transcriptional levels were increased during the**
350 **postharvest senescence**

351

352 To further validate this correlation and examine how the glucosinolates biosynthetic
353 pathway is affected during the postharvest senescence process, we carried out quantitative gene
354 expression of key genes in the glucosinolates biosynthetic pathway and observed that the

355 expression of all genes involved in glucosinolates biosynthetic pathway were up-regulated when
356 stored at room temperature. Mostly, all the key candidates' genes showed significant increase in
357 transcript expression on day 5 when the broccoli florets were stored at room temperature
358 however, the increase in glucosinolates profile during the postharvest day 5 was significant but
359 less drastic than the senescent conditions (Fig 6). This data showed that glucosinolates levels
360 increased rapidly in room temperature when stored at the room temperature. In case of pathway
361 intermediate methylthioalkylmalate synthases (MAM1 and MAM3), there was 4.3 fold and 11
362 fold significant increase in transcript levels from day 1 to day 3 and day 3 to day 5, respectively
363 at room temperature. However, in cold conditions, MAM1 levels were undetectable on day 3 but
364 increased on day 5 by 5.3 fold (Fig 6). This implied that MAM1 levels were increased in higher
365 proportions under room temperature condition. Similar patterns were observed for MAM3,
366 *epithiospecifier modifier 1 (ESM1)*, α -ketoglutarate-dependent dioxygenase (*AOP2*),
367 epithiospecifier protein (*ESP2*), CYP79ST5B2 as their transcripts were significantly higher from
368 day 0 to day 5. However, under cold conditions, the increase in flavin-monooxygenases
369 (FMOGSOX2) transcript levels from day 0 to day 5 was not significant. This observation
370 provided evidence that the gene expression changes in glucosinolates pathway were associated
371 with postharvest storage conditions. Transcript levels for all genes were significantly higher at
372 25°C indicating that the cold temperature was inhibited indo-glucosinolate production in
373 postharvest broccoli. Our results suggested that there is correlation between senescence and
374 indo-glucosinolates concentration in postharvest broccoli.

375

376 **Fig 6. A flowsheet of Indo-glucosinolates biosynthetic pathway displaying the transcript**
377 **levels on day 1, 3, 5 for key enzymatic intermediates catalyzing the biosynthetic pathway.**

378 All the key genes in glucosinolates biosynthetic pathway are highlighted in yellow. Red bars in
379 the graph at each step shows the transcript level of specific candidate genes at 25°C on day 1, 3,
380 5 whereas blue bars show the expression level of the genes at 4°C on day 1, 3, 5. Expression of
381 each gene was normalized using actin as an internal control. Data represents means \pm SE bars
382 (n=3). Asterisks (*) indicate statistically significant differences from day 1 (control) to day 3,
383 day 5 (storage temperature conditions) ($p < 0.05$).

384

385 **Outliers**

386

387 The smaller prediction error for testing fold compared with training and validation folds in
388 Table 2 can be explained by the observed outliers S1 Fig. showed the prediction performance of
389 the training fold with SPICE on the broccoli florets. The marker size and color are related to their
390 prediction error. Apparently, the three circled outliers generated greater prediction error
391 compared with others. The S1 Table listed the glucosinolates concentration of 4 replications kept
392 in room temperature over 6 time points. Three bold numbers are corresponding to the circled
393 outliers in S1 Fig. In rep 1, the glucosinolates concentration was increasing along days, while in
394 rep 2-4, the bold number showed "abnormal" performance. An additional experiment was
395 conducted, where the outliers were moved to the testing fold. S2 Fig. and S2 Table showed the
396 prediction performance and error on the additional testing fold.

397

398 **S1 Table. The glucosinolates concentration under 25°C over days**

	Rep1	Rep2	Rep3	Rep4

Day1	11.9544	11.9405	14.2901	21.099
Day3	53.858	31.6877	75.4149	25.0054
Day5	129.0532	113.1505	122.6086	122.941
Day8	135.2934	151.8804	348.884	82.8482
Day10	182.7224	80.5022	192.528	261.347
Day12	200.8352	269.094	215.7358	213.3986

399
400
401

S2 Table. Comparison of prediction error on additional testing fold

		Average spectra		SPICE	
		RMSE	R2	RMSE	R2
Entire	Training & Validation	31.81 ± 0.64	0.82 ± 0.01	29.24 ± 1.89	0.85 ± 0.02
Broccoli	Testing	65.84 ± 0.32	0.50 ± 0.01	61.81 ± 2.71	0.58 ± 0.05
Broccoli	Training & Validation	29.19 ± 0.36	0.85 ± 0.01	27.54 ± 0.42	0.86 ± 0.01
florest	Testing	66.22 ± 0.22	0.49 ± 0.01	62.47 ± 0.70	0.58 ± 0.01

402
403
404

Discussion

405

406

407 The above experiments were conducted on both the entire broccoli sample and the
408 segmented broccoli florets. Fig. 4 and Table 2 compared the prediction performance and error.
409 Overall, results depicted that the estimated abundances can indicate change in the glucosinolates

410 concentration values. The prediction error can be explained by the fact that the measurement of
411 hyperspectral data and glucosinolates concentration were conducted across different sample
412 scales. Namely, the abundance values are derived from imaging across the entire surface of one
413 side of a broccoli sample, whereas the glucosinolates value is measured using only one small
414 component of the broccoli tissue. The RMSE values show that unmixing with broccoli florets
415 only has slightly less error than when using the entire broccoli sample. In addition, SPICE
416 outperforms the simple averaging. However, when considering the computing and operation
417 complexity, averaging spectra is a more straightforward approach to estimate endmembers as
418 compared with SPICE and does not require parameter selection.

419 In summary, hyperspectral imaging holds promising strength in demonstrating state of art
420 performance in the area of crop sciences through the modulation of imaging with spectroscopy.
421 As shown in this effort, HSI has the potential to provide quantitative parameters in understanding
422 postharvest senescence.

423

424 **Acknowledgements**

425

426 We thank Dr. Diane Rowland for providing the HinaLea 4200 hyperspectral camera and
427 converted to reflectance spectra for supporting in sample imaging. We thank Dr. Ru Dai and Dr.
428 Jeongim Kim for the HPLC analysis and troubleshooting. This work was supported by UF Seed
429 Fund (#P0175583 to A.Z., T.L.).

430

431 **References**

432

- 433 1. Angelino D, Jeffery E. Glucosinolate hydrolysis and bioavailability of resulting
434 isothiocyanates: Focus on glucoraphanin. *Journal of Functional Foods*. 2014 Mar ;7:67–
435 76.
- 436 2. Coupe SA, Sinclair BK, Watson LM, Heyes JA, Eason JR. Identification of dehydration-
437 responsive cysteine proteases during post-harvest senescence of broccoli florets. *Journal*
438 *of Experimental Botany*. 2003 Mar 1;54(384):1045–56.
- 439 3. Karimi Y, Maftoonazad N, Ramaswamy HS, Prasher SO, Marcotte M. Application of
440 Hyperspectral Technique for Color Classification Avocados Subjected to Different
441 Treatments. *Food Bioprocess Technol*. 2012 Jan;5(1):252–64.
- 442 4. Shewfelt RL, Heaton EK, Batal KM. Nondestructive Color Measurement of Fresh
443 Broccoli. *J Food Science*. 1984 Nov;49(6):1612–3.
- 444 5. Feng Y-Z, Sun D-W. 2012. Application of hyperspectral imaging in food safety inspection
445 and control: a review. *Crit Rev Food Sci Nutr* 52:1039–58.
- 446 6. Gowen A, Odonnell C, Cullen P, Downey G, Frias J. Hyperspectral imaging – an emerging
447 process analytical tool for food quality and safety control. *Trends in Food Science &*
448 *Technology*. 2007 Dec;18(12):590–8.
- 449 7. Singh CB, Jayas DS, Paliwal J, White NDG. Fungal Damage Detection in Wheat Using
450 Short-Wave Near-Infrared Hyperspectral and Digital Colour Imaging. *International*
451 *Journal of Food Properties*. 2012 Jan;15(1):11–24.
- 452 8. Chen Y-R, Chao K, Kim MS. Machine vision technology for agricultural applications.
453 *Computers and Electronics in Agriculture*. 2002 Nov;36(2–3):173–91.
- 454 9. Ariana DP, Lu R, Guyer DE. Near-infrared hyperspectral reflectance imaging for
455 detection of bruises on pickling cucumbers. *Computers and Electronics in Agriculture*.
456 2006 Aug;53(1):60–70.
- 457 10. Cheng X, Chen YR, Tao Y, Wang CY, Kim MS, Lefcourt AM. 2004. A novel integrated
458 PCA and FLD method on hyperspectral image feature extraction for cucumber chilling
459 damage inspection. *Trans ASAE* 47:1313–20
- 460 11. Noh HK, Lu R. Hyperspectral laser-induced fluorescence imaging for assessing apple
461 fruit quality. *Postharvest Biology and Technology*. 2007 Feb;43(2):193–201.
- 462 12. ElMasry G, Wang N, Vigneault C. 2009. Detecting chilling injury in Red Delicious apple
463 using hyperspectral imaging and neural networks. *Postharvest Biol Technol* 52:1–8.
- 464 13. Li X, He Y. Non-destructive measurement of acidity of Chinese bayberry using Vis/NIRS
465 techniques. *Eur Food Res Technol*. 2006 Oct;223(6):731–6.
- 466 14. Zhu Q, Huang M, Zhao X, Wang S. Wavelength Selection of Hyperspectral Scattering
467 Image Using New Semi-supervised Affinity Propagation for Prediction of Firmness and
468 Soluble Solid Content in Apples. *Food Anal Methods*. 2013 Feb;6(1):334–42
- 469 15. Cernadas RA, Benedetti CE. Role of auxin and gibberellin in citrus canker development
470 and in the transcriptional control of cell-wall remodeling genes modulated by
471 *Xanthomonas axonopodis* pv. *citri*. *Plant Science*. 2009 Sep;177(3):190–5.
- 472 16. Xing J, Guyer D, Ariana D, Lu R. Determining optimal wavebands using genetic
473 algorithm for detection of internal insect infestation in tart cherry. *Sens & Instrumen*
474 *Food Qual*. 2008 Sep;2(3):161–7.
- 475 17. Hernandez-Hierro JM, Esquerre C, Valverde J, Villacreces S, Reilly K, Gaffney M,

- 476 Gonzalez-Miret ML, Heredia FJ, O'Donnell CP, Downey G. 2014. Preliminary study
477 on the use of near infrared hyperspectral imaging for quantitation and localisation of total
478 glucosinolates in freeze-dried broccoli. *J Food Eng* 126:107–12.
- 479 18. Kwatra V, Schödl A, Essa I, Turk G, Bobick A. Graphcut textures: image and video
480 synthesis using graph cuts. *ACM Transactions on Graphics (ToG)*. 2003 Jul 1;22(3):277-
481 86.
- 482 19. Gilat A. *MATLAB: An introduction with applications*. New York:: Wiley; 2008.
- 483 20. Keshava N, Mustard JF. Spectral unmixing. *IEEE Signal Process Mag*. 2002
484 Jan;19(1):44–57.
- 485 21. Zare A, Gader P. Sparsity Promoting Iterated Constrained Endmember Detection in
486 Hyperspectral Imagery. *IEEE Geosci Remote Sensing Lett*. 2007 Jul;4(3):446–50.
- 487 22. Lee JG, Lim S, Kim J, Lee EJ. The mechanism of deterioration of the glucosinolate-
488 myrosinase system in radish roots during cold storage after harvest. *Food Chemistry*.
489 2017 Oct;233:60–8.

490

491 **Supporting information**

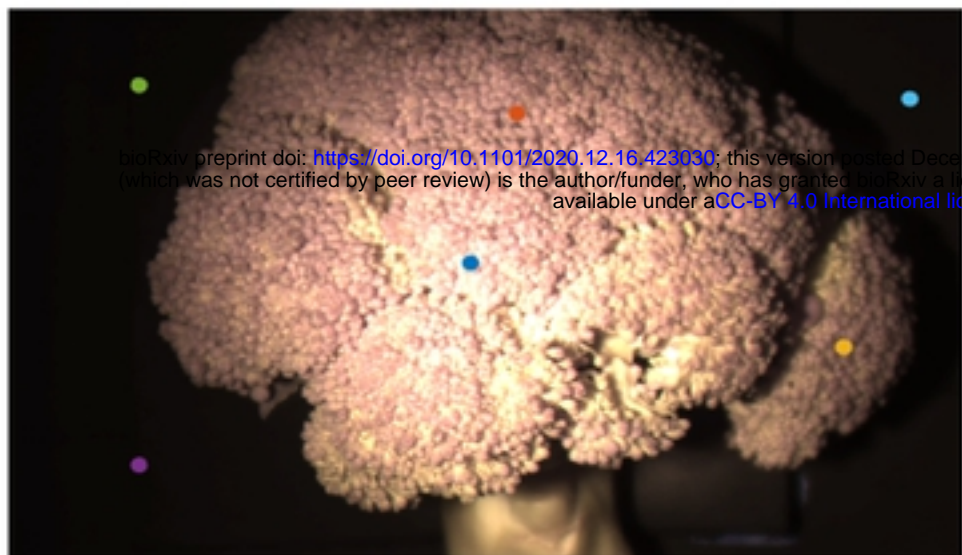
492 **S1 Fig. Predicted glucosinolates on training folds.** x-axis indicates the real glucosinolates
493 concentration, y-axis indicates the prediction. Markers that closer to the “x = y” line indicates
494 more accurate prediction. The marker size and color is corresponding to the prediction error, the
495 bigger and brighter markers indicate greater error.

496

497 **S2 Fig. Predicted glucosinolates on additional testing fold.** x-axis indicates the real
498 glucosinolates concentration, y-axis indicates the prediction. Markers in various shape and color
499 denotes prediction with different methods. Markers that closer to the “x = y” line indicates more
500 accurate prediction.

501

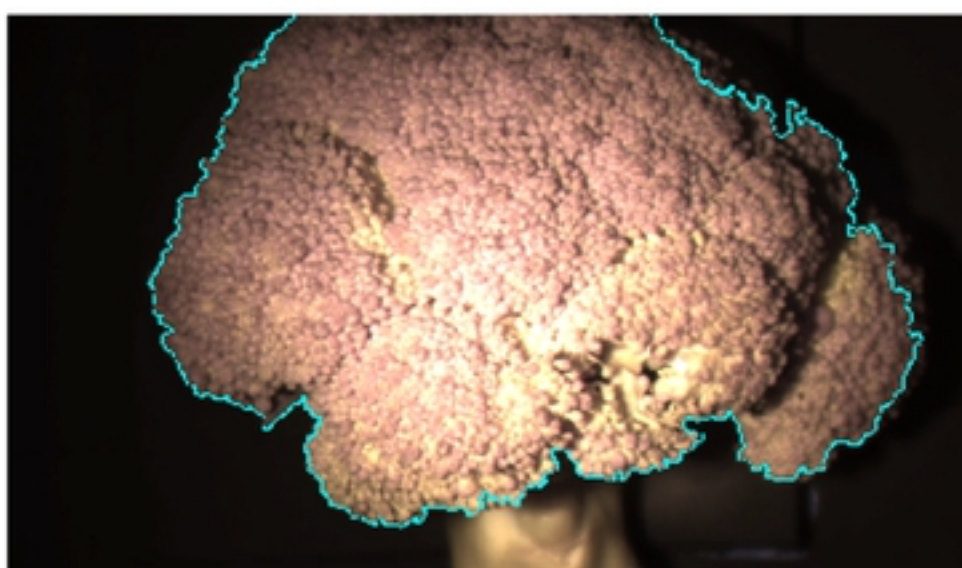
502



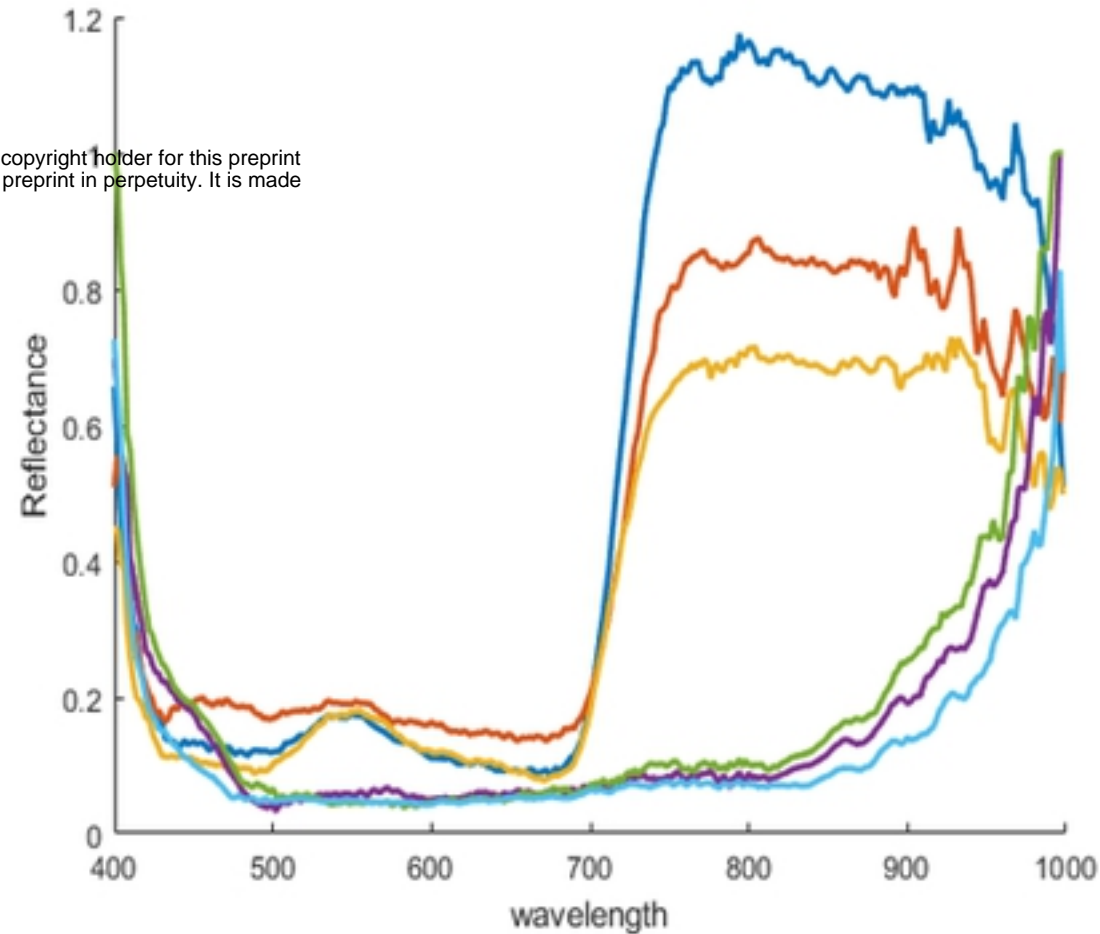
(a) RGB image of broccoli



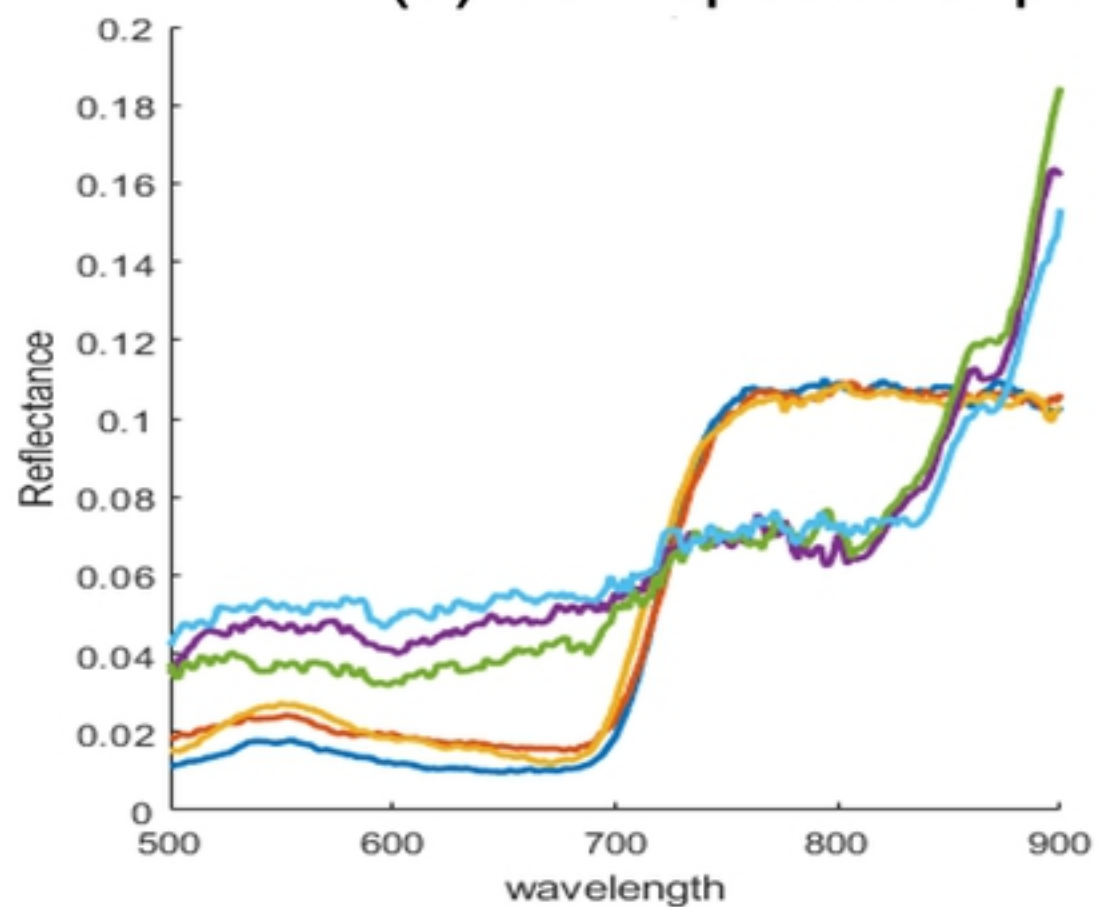
(b) Segmented broccoli



(c) Segmented broccoli flower



(d) Raw spectra of points



(e) Preprocessed spectra

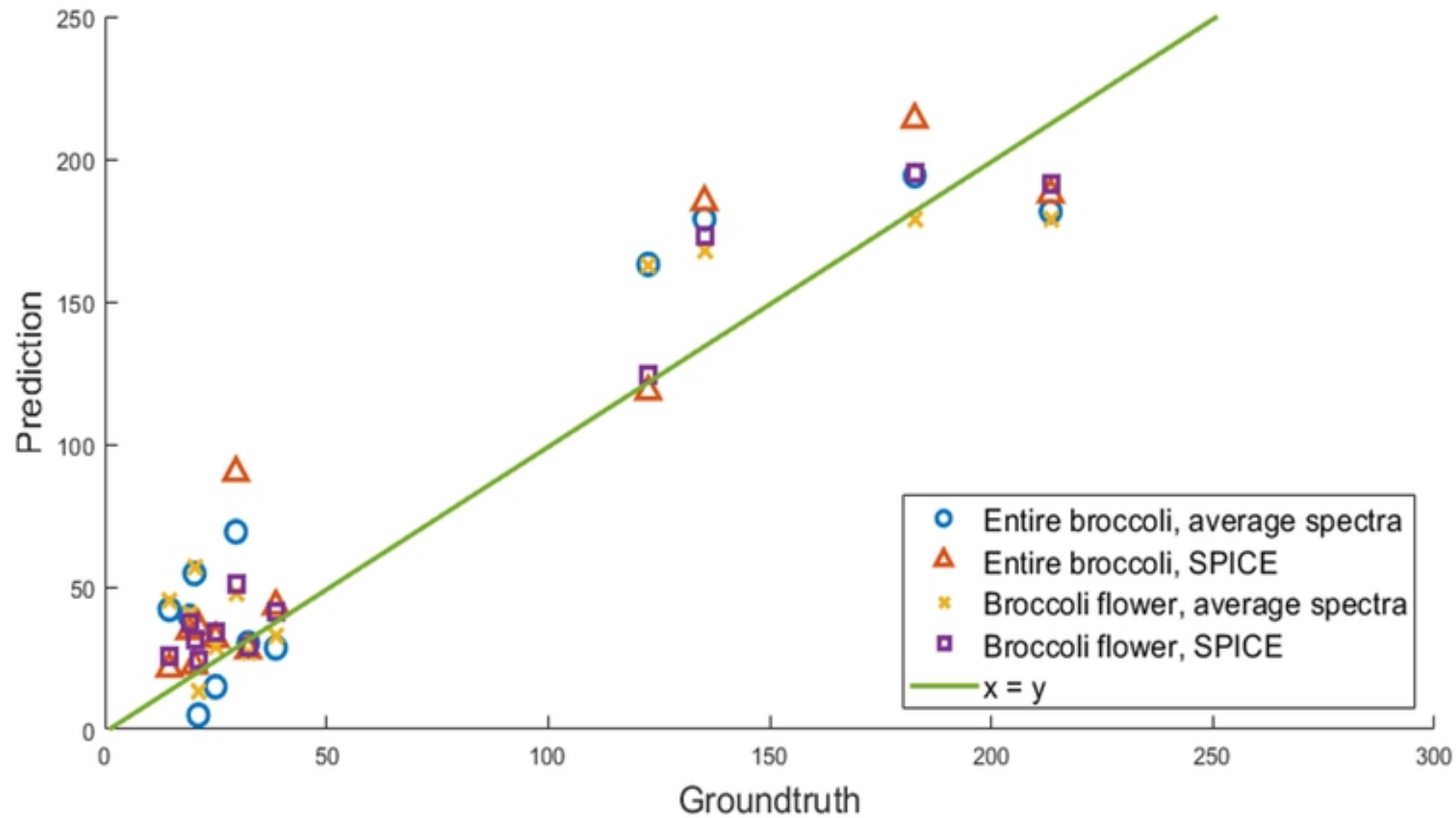
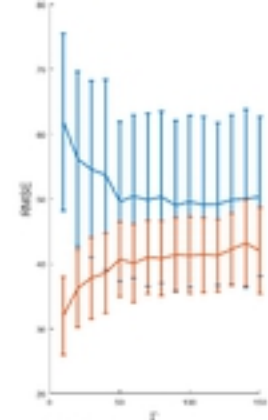
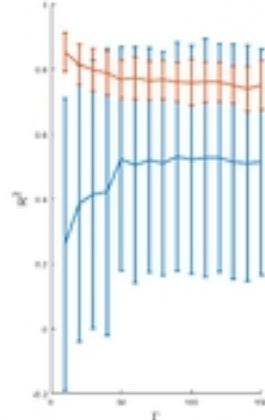


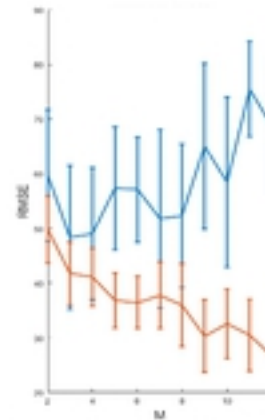
Fig 2



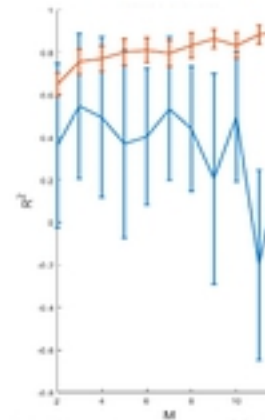
(a) RMSE over Γ (entire broccoli)



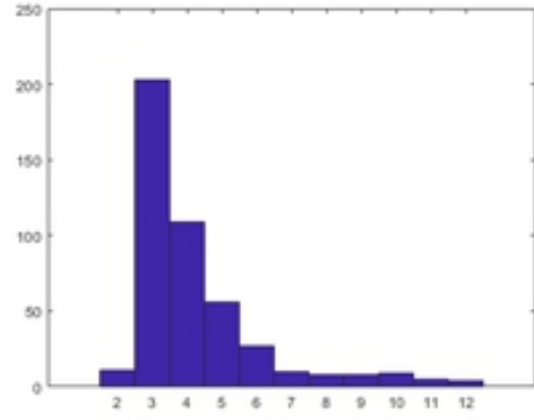
(b) R-squared over Γ (entire broccoli)



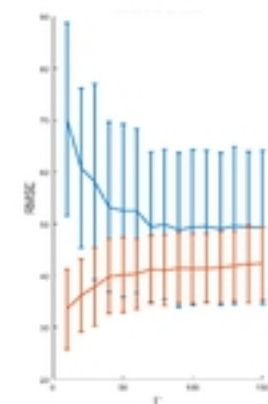
(c) RMSE over M (entire broccoli)



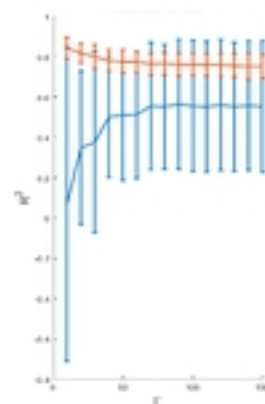
(d) R-squared over M (entire broccoli)



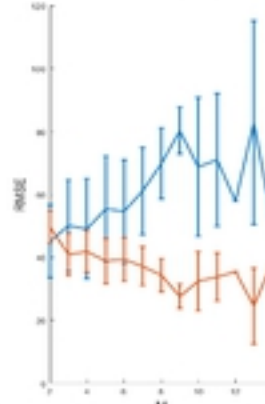
(e) Histogram of M (entire broccoli)



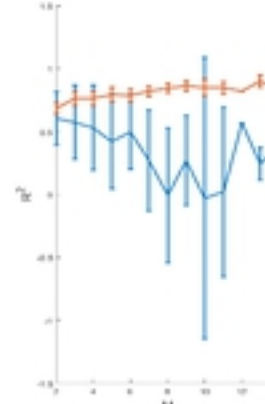
(f) RMSE over Γ (broccoli forest)



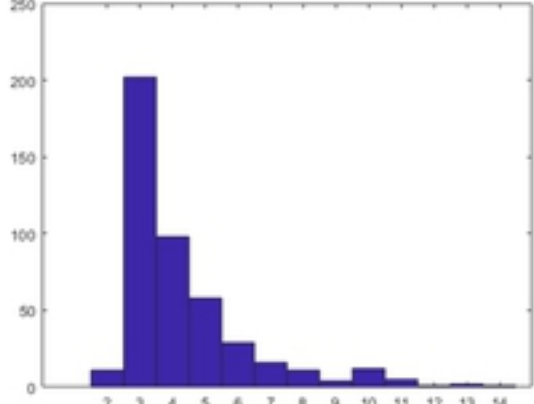
(g) R-squared over Γ (broccoli forest)



(h) RMSE over M (broccoli forest)



(i) R-squared over M (broccoli forest)



(j) Histogram of M (broccoli forest)

Fig 3

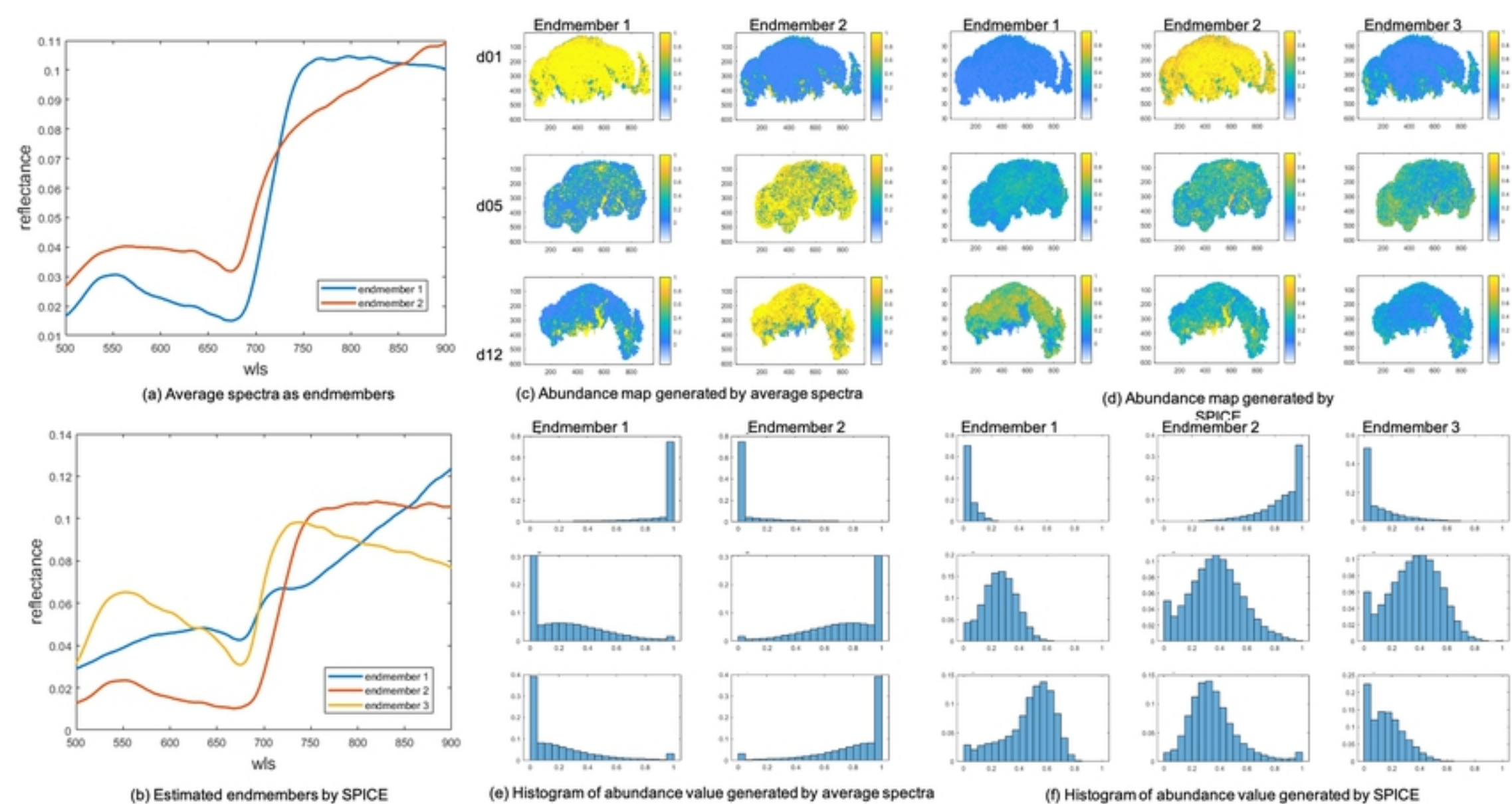


Fig 4

Indo-glucosinolate content (HPLC)

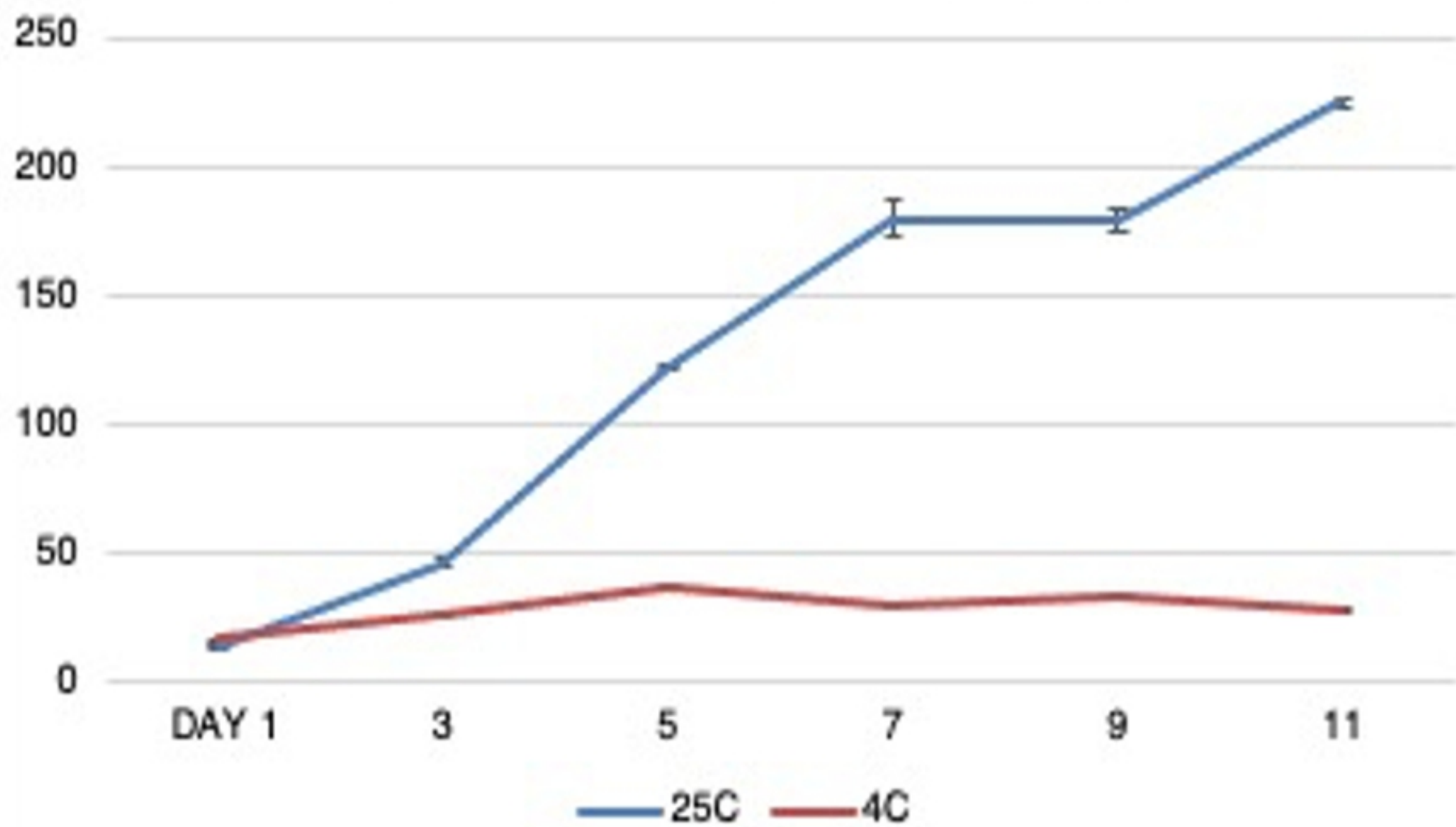


Fig 5

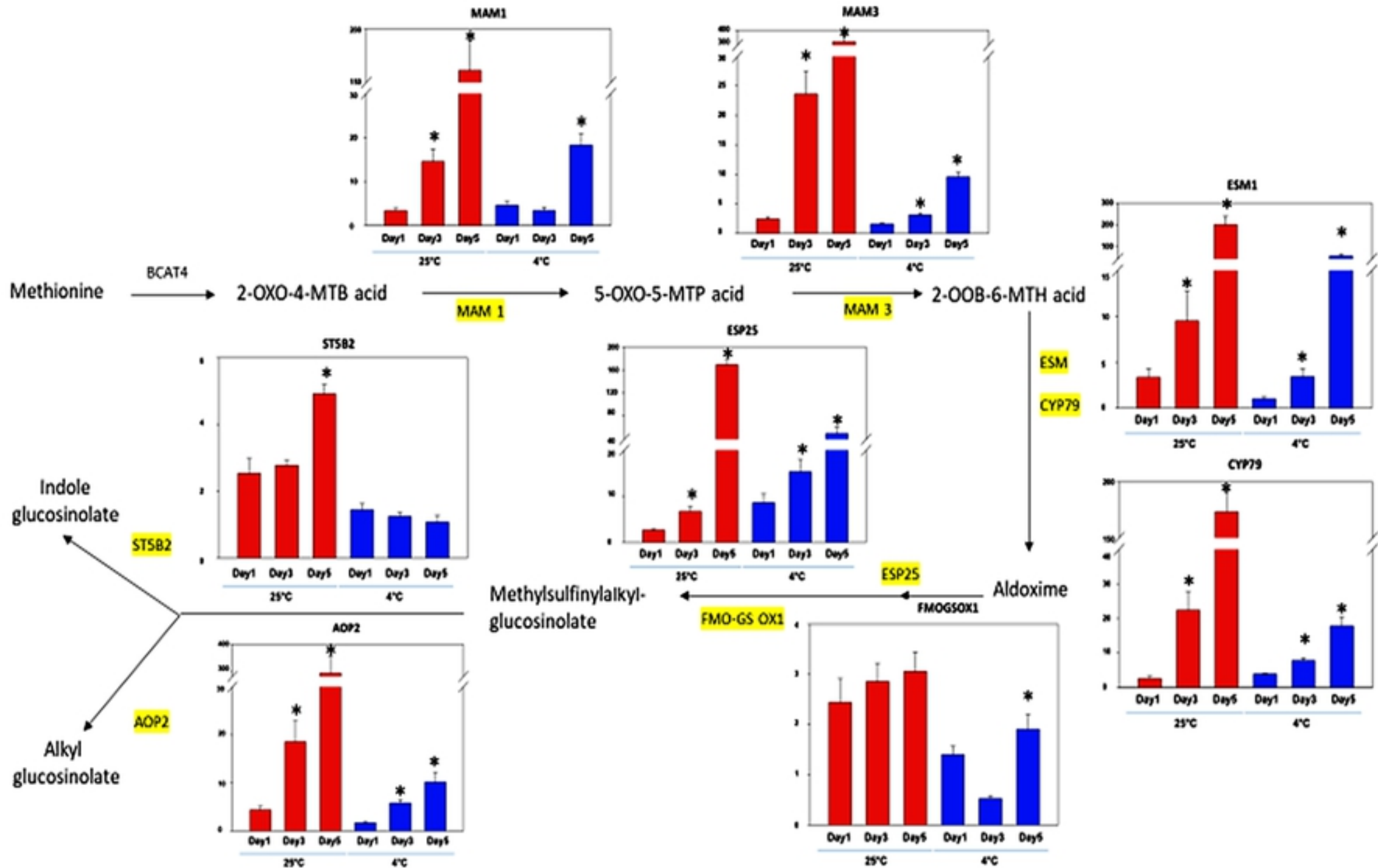


Fig 6