Chemical Compositions of Edamame Beans and Valorization of Edamame Shells

Dajun Yu

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Food Science and Technology

Haibo Huang, Chair
Yin Yun
Bo Zhang
Sean F. O'Keefe
Susan E. Duncan

December 14th, 2022
Blacksburg, Virginia

Keywords: Edamame Beans, Physical and Chemical Properties, Edamame Genotypes, Planting Locations, Harvest Time, Spectroscopy-based Machine Learning, Edamame Shells, Soluble Dietary Fiber

Chemical Compositions of Edamame Beans and Valorization of Edamame Shells

Dajun Yu

ABSTRACT

Edamame is becoming more popular in the U.S. due to its high nutritional value and potential health benefits. However, more than 70% of edamame is imported from outside of the U.S. Therefore, developing elite edamame genotypes is critically desirable to increase the domestic production of edamame in the U.S. Genotype, planting location, and harvest time play essential roles in the chemical composition of edamame, which further decide edamame's nutritional value and sensory characteristics. Therefore, the first goal of this study is to comprehensively evaluate the chemical composition of edamame genotypes grown in different locations. Ten selected edamame genotypes were grown in three locations in the U.S. - Whitethorne, Virginia (VA), Little Rock, Arkansas (AR) and Painter, VA. Sugars, alanine, protein, oil, neutral detergent fiber (NDF), starch, ash, and moisture contents, were comprehensively analyzed. The results showed that location had significant effects on all chemical components of edamame with $p < 0.05$. Compared to Painter and Little Rock, genotypes planted in Whitethorne had higher averaged free sucrose, fructose, glucose, raffinose, stachyose, and starch contents and total sweetness. The highest crude protein and oil contents were found on edamame planted in Painter, while Little Rock produced edamame with the highest free alanine, ash, and moisture contents. Genotype significantly affected chemical compositions except for NDF and raffinose. Therefore, planting location and edamame genotype should be considered when producing elite edamame for the U.S. market.

Chemical composition changes with the development of edamame; therefore, harvest time is essential for harvesting high-quality edamame. The second objective of this study is to quantify the changes in both physical and chemical properties of edamame over bean development and apply a combined spectroscopy and machine learning (ML) technique to help determine the optimal harvest time. Physical and chemical properties were analyzed for edamame harvested at R5 (beginning seed), R6 (full seed), and R7 (beginning maturity) growth stages, and the spectral reflectance (360 – 740 nm) of edamame pods was measured using a handheld spectrophotometer. The samples harvested at different stages were labeled as 'early,' 'ready,' and 'late.' At R6, pod/bean weight and pod thickness reached the peak and then stayed stable, while sugar, alanine, starch, and glycine also peaked at R6 but declined afterward. The spectra-based ML method had high accuracy (0.95) when classifying 'early' and 'late' edamame, and the accuracy was 0.87 for classifying 'early' and 'ready' edamame. These results indicated that this spectra-based ML method could determine the optimal harvest time of edamame.

Food waste and loss not only lead to economic loss but also significant greenhouse gas emissions. With edamame food/snack production increasing, edamame shells, the low-value byproduct from this processing, will potentially threaten the environment. Similar to other food processing byproducts, edamame shell is rich in dietary fiber (DF). However, the high concentration of insoluble dietary fiber (IDF) limits its application as a food additive. Therefore, extraction/modification processes are needed to convert IDF to soluble dietary fiber (SDF) and improve the properties of edamame shell-derived DF. Ball milling is one of the most efficient techniques to break down biomaterials into sub-micro-level particles. Citric acid, as a natural and safe food additive, can help break down cell walls and improve the dissolution of SDF by ionizing the hydrogen ions with carboxyl groups. Therefore, the third objective of this study is to

develop a process that combines ball milling and citric acid treatments to produce SDF from edamame shells. We investigated different treatment parameters, including different citric acid concentrations, treatment temperatures and time, and the application of ball milling. To determine if the combined treatment can potentially improve the properties of the produced SDFs, we characterized the physicochemical, morphological, structural, rheological, thermal, and functional properties of SDFs produced at different conditions. The results showed that the highest SDF yield (19.5%) was found when the edamame shells were pretreated by a ball mill. In addition, the combined citric acid and ball milling treatment altered several properties of the produced SDFs, including particle size, morphology, and crystallinity. Moreover, ball milling treatment led to a higher exothermic temperature peak of SDF indicating better thermal stability. All produced SDFs significantly elevated the production of short-chain fatty acids during *in vitro* fermentation (compared to the control fermentation) which indicated their potential benefits of promoting gut health. Overall, we demonstrated that ball-milling-assisted citric acid processing can be an effective green technique to produce SDF from edamame shells. The SDF produced from edamame shells can be regarded as a promising and novel ingredient with great potential to be used in foods.

Chemical Compositions of Edamame Beans and Valorization of Edamame Shells

Dajun Yu

GENERAL AUDIENCE ABSTRACT

Edamame is becoming increasingly popular among consumers in the U.S. because it is nutritious and good for health. However, more than 70% of edamame in the U.S. market is imported from other countries. Therefore, having more edamame genotypes that adapt to the growing environment in the U.S. will help increase the domestic production of edamame. Genotype and planting location are essential in deciding edamame's nutritional value and taste. Therefore, the first objective of this study is to comprehensively understand the nutritional value of different edamame genotypes grown in three planting locations. The results showed that both location and genotype affected the nutritional values of edamame, indicating that planting location and edamame genotype should be considered when developing better edamame for the U.S. market.

Nutritional value and sweetness change with the growth of edamame beans. Therefore, harvest time is crucial for harvesting edamame with better nutrition and taste. This study's second objective is to observe edamame's nutritional factors and sweetness over bean development and develop a method using a handheld colorimeter to help determine the optimal harvest time. The results showed that the edamame harvested at the full seed stage (called R6) is the sweetest compared to the other two stages. In addition, the handheld colorimeter combined with the machine learning technique showed high accuracy in separating ‘early’ and ‘late’ harvested edamame and ‘early’ and ‘ready’ harvested samples. These results indicated that the

combination of colorimeter and machine learning could help determine the optimal harvest time of edamame.

Food waste and loss not only lead to economic loss but also significant greenhouse gas emissions. Edamame shells, the low-value byproduct from edamame snack/food processing, will potentially threaten the environment if edamame consumption keeps increasing. Like other food waste, edamame shell is rich in dietary fiber (DF). Therefore, it is vital to find a way to recover the DF in edamame for other applications. Ball milling is a green technology that can efficiently break down big particles. Citric acid is a natural and safe food additive and can help break down insoluble cell walls. Therefore, this study aims to produce soluble dietary fiber (SDF) from edamame shells using ball milling and citric acid. We proved that ball-milling assisted acid processing can be an environmentally friendly method to produce edamame shell SDF which can potentially be used as a suitable food ingredient.

Acknowledgments

First, I would like to thank the USDA Specialty Crop, Virginia Agricultural Experiment Station and the Hatch Program of the National Institute of Food and Agriculture (NIFA), USDA, and the Water INTERface-Interdisciplinary Graduate Education Program for supporting and funding my research and study.

I would like to express my deepest appreciation to my advisor, Dr. Haibo Huang, for taking me to graduate school and giving me the opportunity for advanced education at Virginia Tech. Throughout my Ph.D. study, you provided countless support, invaluable advice, and guidance for all my projects and experiments. You taught me how to become an independent researcher with critical thinking and the ability to solve research problems. When I joined graduate school, I had zero experience in writing a manuscript. However, you are patient in revising each of my manuscripts and showed me how to write a good scientific paper. I am fortunate to have such a responsible and knowledgeable advisor.

I would also like to extend my deepest gratitude to the rest of my committee: Dr. Yun Yin, Dr. Bo Zhang, Dr. Sean F. O'Keefe, and Dr. Susan E. Duncan for your willingness to serve on my committee. During my Ph.D. study, you gave me relentless support, invaluable advice, and feedback. I am lucky to have such a knowledgeable and involved committee for my Ph.D. study. Here, I want to express special thanks to Dr. Duncan, even though when I was writing the acknowledgments, you are not in the world anymore. I still remember how you inspired me the first time we met in our 15-min interview. Those words are the greatest motivations for my graduate study. R.I.P.

I also had great pleasure working with everyone on the edamame team. We always had great discussions and generated great research ideas in every meeting. Everyone would love to help each other and offer potential collaboration opportunities. In this project, I learned so much about analytical chemistry, contributing to my current career as an analytical chemistry scientist. I am always proud to tell people that I have analyzed over 1000 edamame samples using HPLC, so I had chances to encounter different problems and become an expert in HPLC. Thanks to everyone who planted, harvested, and cooperated with us to ship the edamame samples based on our requirements.

Thanks should also go to everyone that provided help with my experiments: Dr. Hengjian Wang for helping me freeze dry and analyze countless edamame samples. Whenever I have questions or need help, I know you are always there and willing to help us with your immense knowledge and experiences. Dr. Jian Wu for helping me take care of the edamame post-harvest processing, which saved me lots of time so I could work on the other part of my experiments; Brian Wiersema for helping us design and setup the edamame post-harvest processing and coordinate the usage of the pilot plant; Brett Driver for facilitating us with the usage of the pilot plant; Kim Waterman for helping me set up the autoclave whenever needed; Dr. Ken Hurley for providing advice on instrumental analysis; all faculty, staff, and students in FST for bringing the happiness and being willing to help all the time. I am also extremely grateful to all my undergraduate research assistants. You helped me so much with the post-harvesting processing of the edamame. We processed thousands of edamame together and all of you are extremely supportive and hard-working. I cannot get so many samples done without you.

Thanks to Justin Polk and Nick Lord for helping me plan the edamame planting and harvesting for the harvest time project and taking care of the harvesting and physical properties

measurements. I also appreciate Dr. Song Li and his student Kshitiz Dhakal for helping build the machine learning model in this project. At the same time, I very much appreciate the companionship and help from all my lab mates, Dr. Qing Jin, Dr. Yiming Feng, Dr. Josh O'Hair, Dr. Yanhong He, Jordan Allen, Nicholas Poe, Andrea Johnson and Yilin Li. It is a happy time to work with all of you and thank you for being supportive lab mates.

My father, Zhongquan Yu, and my mother, Yexiang Wang, are the most importantly emotional support during my graduate study and throughout my life. Thank you for respecting and supporting every decision I have made, even though I will not be around you for so many years. A simple thank you is not enough to express how grateful I am. I love you so deeply.

My next sincere and special thanks go to my husband-to-be Zhijie Yang. Thank you for bringing brightness to every dark corner of my life. I have never been so happy until I met you. You let me believe I am a brilliant person, so I become more confident. You always treat me as your princess and take responsibility for lots of things in our life. I cannot imagine a life without you, and it will be the happiest thing to spend the rest of my life with you. Thank you for your endless love and company. Words cannot express how much you mean to me. I love you so much.

Finally, I want to say thanks to my cutest family member, Curry. You are the angel that brings lots of happiness and comfort to my life. You are so sweet and always stick to me and follow me everywhere I go. Lots of the happy memories in Blacksburg are contributed by you. My second cat, Rice came into my life when I moved to California. This baby girl is also a beautiful angel accompanying me so that Curry and I won't feel lonely. Thank you, my kitties, Curry and Rice, for your company.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Edamame, called "*mao dou*" in China, vegetable soybean or edible soybean in the U.S., belongs to the same species as the grain soybean [*Glycine max* (L.) Merr.]. Their ancestor is Glycine soja Sieb & Zucc, originating in eastern China (Li et al., 2022; Zeipiņa et al., 2017). It is an important specialty soybean in Asia and had a long history in China for more than 2000 years (Miles et al., 2018). Different from traditional grain soybean, edamame is harvested before full maturity when its pods and beans are still fresh and green, and beans are still tender. It has been widely consumed in China and Japan for centuries as a snack or side dish (Zeipin¸a et al., 2017; Lara et al., 2019). Due to its high nutrition and potential health benefits, the consumption of edamame in the U.S. has been rising quickly, and it has become the second-largest soy food with around 30,000 tons of annual consumption (Xu et al., 2016; Zeipin¸a et al., 2017). More than 70% of the edamame in the U.S. market is imported from China and Taiwan (Li et al., 2022). Edamame is imported as frozen products, and such frozen processing might lower the quality and affect the taste of edamame (Li et al., 2022; Nolen et al., 2016). In addition, foodborne bacteria such as *Listeria monocytogenes* and *Escherichia coli* have received increasing concerns. Due to these two reasons, consumers start shopping for fresh edamame grown domestically. However, the U.S. has limited edamame genotypes that are best suited for growing domestically in terms of nutritional composition and consumer acceptance (Shurtleff and Akiko, 2014). Therefore, selecting and breeding elite U.S. edamame genotypes is the primary task to promote domestic edamame production.

The nutritional values and health benefits of edamame are driving its growing popularity. The nutritional values are mainly influenced by its chemical compositions, which include protein, fiber, starch, sugars, and minerals. Aside from nutrients, sensory characteristics are other critical factors affecting consumer acceptance and sweetness is one of the most essential sensory characteristics (Johnson et al., 1999). These two factors, chemical composition and sweetness, are potentially affected by edamame genotype, planting location, and harvest time. Song et al. reported that different contents of sugars, free amino acids, and organic acids were found in different edamame genotypes (Song et al., 2013). Based on our preliminary data, the soluble sugar contents were significantly different among different edamame genotypes planted in different locations. Therefore, identifying a better genotype and planting location is crucial to obtain high-quality edamame. Additionally, Xu et al. (2016) reported that the nutritional properties of edamame change over the plant development, indicating that an appropriate harvest time is crucial for producing edamame of optimal quality and taste. Nevertheless, the edamame harvest window is short (only one week), and a better understanding of the changes in edamame over bean development will facilitate determining the harvest time efficiently, quickly, and accurately. Also, technology is highly demanded to uniformize the harvest time decision process and promote harvest consistency.

Food loss and waste constantly threaten the sustainability of the food system. Approximately 40% of the food is wasted, most of which ends up in landfills, resulting in massive economic losses and significant greenhouse gas emissions (Poe et al., 2020; Yu et al., 2022). Edamame shell, a low-value byproduct generated by edamame food/snack manufacturing, is a rich source of dietary fiber, similar to other food processing byproducts (e.g., tomato pomace, corn bran, pomegranate peel, and lemon peel). However, the high content of insoluble

dietary fiber (IDF) limits its potential to be used as a food additive because it affects not only the sensory quality but also the processibility of food (Li et al., 2022). To overcome this disadvantage, techniques are needed to convert IDF to soluble dietary fiber (SDF) and extract SDF from edamame shells. Different processes have been reported in previous papers including physical, chemical and biological methods. Ball milling is a popular physical treatment because of its effectiveness in particle size reduction and increasing the extraction yield of SDF (Jiang et al., 2022). Furthermore, citric acid is a common organic acid found in many fruits, such as oranges, grapefruits, and pineapples (Behera et al., 2021). Therefore, it is frequently used as a safe food additive/ingredient in food industries (Yan et al., 2019). However, few or no research has been conducted on understanding how the combined ball milling and citric acid process affects the SDF extraction from edamame shells. More importantly, it is critical to understand how this combined treatment modifies the properties of the produced SDF.

Therefore, the goal of the dissertation research is to thoroughly investigate the effects of genotype, location, and harvest time on the chemical composition and sweetness of edamame and use the obtained information to guide edamame breeding and harvesting. Additionally, this study aims to develop a consistent, rapid, and accurate spectroscopy-based machine learning (ML) technique for identifying the optimum harvest time of edamame and an efficient combined ball milling and citric acid process to produce SDF from edamame shells.

Chapter 2

Literature Review

2.1 Chemical compositions and nutritional values of edamame

The increasing popularity of edamame is due to its nutritional value and health benefits. Chemical composition of edamame is important to its nutrition. It was reported that edamame was highly nutritious and valuable with a high content of high-quality protein with isoflavones (Zeipin¸a et al., 2017). Such secondary metabolites, including isoflavones, phytic acids, phenolic compounds, and saponins, can be healthy supplements for a healthy diet (Zhang et al., 2017). Based on our preliminary data, the protein content in edamame is around 40% which is the second most abundant component after water. Moreover, edamame has a 56% higher protein content compared to green peas (*P. sativum*) (Masuda, 1991). Based on the above-mentioned information, edamame has both good quality and a good quantity of proteins and could offer plant-based proteins and 18 amino acids to vegetarians and vegans (Jiang et al., 2018).

Edamame is also rich in carbohydrates (Johnson et al., 1999; Xu et al., 2012a), which could offer energy when supplemented with diet. However, lower amounts of oligosaccharides, such as raffinose and stachyose are expected because they may potentially lead to digestive diseases (Choct et al., 2010; Saldivar et al., 2011; Kumar et al., 2010). Moreover, edamame contains a significant amount of polyunsaturated fatty acids (linoleic acid and α-linolenic acid) (Kumar et al., 2006a), which makes edamame a nutritious and healthy snack. Edamame is also a good source of vitamins (C, E, and thiamin), minerals, phosphorus, phytochemicals, and other active compounds (Johnson et al., 1999; Song et al., 2003; Kaiser and Ernst, 2013; Zeipin¸a et

al., 2017). In addition, compared to soybean, edamame has a lower trypsin-inhibitor level and a higher vitamin C content (Cai et al., 2013). With these superior nutrition compositions, edamame has the potential health benefits to help prevent many diseases, such as cardiovascular disease, cancer, or osteoporosis (Sirtori, 2001).

Other than that, edamame is different from soybean because the primary use of edamame is human food, while most of the soybean is used to produce soybean oil or feed animals. Therefore, a better edamame cultivar should have better consumer acceptance. Regarding sensory attributes, edamame is slightly sweeter, has a softer texture, and has less beany and nutty flavor than soybean (Konovsky et al., 1994). Sweetness is one of the most important eating qualities of edamame and is also one of the three primary concerns of edamame quality measured in Japan (Johnson et al., 1999). Sucrose, a digestible sugar, is crucial for the taste of edamame because it is responsible for its sweetness (Kumar et al., 2010; Song et al., 2013). It was reported that the sweetness of edamame can be estimated by the soluble sugars, including sucrose, glucose, and fructose (Konovsky et al., 1994; Song et al., 2013b; Zeipin¸a et al., 2017). It was reported that the edible quality score of edamame had a positive relationship with sucrose content (Zhang et al., 2015). Another two easily digested sugars (i.e., fructose and glucose), together with the amino acid alanine, should also be explored since they also contribute to the sweetness of edamame (Song et al., 2013).

2.2 Factors related to the chemical compositions of edamame

2.2.1 Genotypes

The chemical composition and nutritional values of edamame are highly related to genotypes/cultivars. Therefore, breeding and genetic research on analyzing edamame of different genetic characteristics is important to screen and select adapted edamame with better nutritional

values and consumer acceptance for the U.S. market. Rao et al. spent four years evaluating the fresh bean compositions of four large-seeded plant introductions and six edamame cultivars from Japan, two vegetable soybean cultivars from China, and two adapted U.S. cultivars (Rao et al., 2002). The oil and protein contents range from 13.1 to 15.6% and 33.3 to 38.6%, respectively. The average glucose content was 6.7%, while the mean phytate content was 1.3%. They identified some genotypes that might be used in the breeding program for Georgia and the southeastern USA. Jiang et al. analyzed protein, oil, starch, dietary fiber, sucrose, stachyose, ash, raffinose, and total sugar contents of 86 U.S.-developed breeding lines and cultivars grown in Virginia in both 2015 and 2016 (Jiang et al., 2018). This study found that genotypic differences were significant for most chemical components except for ash in both years and raffinose in the year 2016. Another study by Carson et al. (2011) determined the protein and oil contents of five edamame cultivars in the mid-Atlantic United States. The results showed differences in both protein and oil contents among cultivars. In the study of Guo et al. (2020), 19 different genotypes were selected and analyzed for protein and oil contents. Results showed that the oil and protein contents of different edamame genotypes ranged from 1.9 to 5.7% and 38.4 to 42.6%, respectively. However, limited research has conducted a complete macronutrient and total sugar analysis on edamame of different genotypes. Such comprehensive studies are required for selecting better edamame breeding lines for the U.S. market.

2.2.2 Planting location

Planting location is another essential factor to consider for producing edamame of better quality. Previous studies have shown that the planting location had a more significant influence on the chemical compositions of crops [e.g., peanut (*Arachis hypogaea L.*), amaranth (*Amaranthus cruentus*)] than genotype (Eheart et al., 1955; Berganza et al., 2003). Sakla et al.

(1988) evaluated the environmental effects on the chemical composition of three soybean varieties. The results found significant differences in moisture, oil, carbohydrates, sucrose, and protein contents of soybean planted in different locations. No significant difference was observed in fiber and ash contents among different varieties and locations. The effects of environmental conditions on the chemical composition of soybean seeds were also reported in other literature (Cartter et al., 1942; McClure et al., 2017). However, Cartter et al. (1942) reported that varieties played a more critical role in affecting the chemical composition of soybean seed in their study compared to locations. Guo et al. (2020) planted 19 different edamame genotypes in two locations in Florida and analyzed the protein and oil contents of edamame beans. They found that most genotypes showed different protein and oil contents between these two locations.

Environmental factors in different planting locations, including temperatures, soil types and properties, daylight intensity, altitude, and precipitation, can potentially affect the chemical compositions of edamame. Wolf et al. (1982) and Kumar et al. (2010) reported that increasing the temperature during seed development led to decreased sucrose content of soybean seeds. Wolf et al. (1982) discovered that when the day temperature exceeded 30 ℃, the protein content of soybean seeds increased. A second study also observed increased protein content at higher temperatures (Dornbos and Mullen, 1992). Different soils in different planting locations might have different nutrients and pH levels, thus affecting sucrose synthesis in crops. According to Zhao-Hui et al. (2008), kidney beans have a higher soluble sugar level when grown in soil with more potassium (K). Increased sugar content in sugar beet was observed in the study of McEnroe and Coulter (1964) when the soil pH increased from 6 to 7. The rate of photosynthesis and, consequently, the sugar content in vegetable crops are both impacted by light intensity. According to Xu et al. (2009), soluble sugar concentrations in non-heading Chinese cabbage are

positively correlated with light intensity. The study by Yang et al. (2009) showed a similar result that soluble sugars in pepper fruits reduced with decreased light intensity. Precipitation can also cause water stress or flooding, impacting the crop sugar content. Plants prefer accumulating soluble sugars in response to water stress to regulate their osmotic pressure. Okunlola et al. (2016) evaluated the total sugar accumulation of three pepper types under water stress and discovered that pepper with less irrigation had a higher total sugar level. Flooding depletes soil nutrients, which then affects the sugar content of crops (Clark, 2020). Moreover, it is crucial to identify genotypes that can perform well at different locations.

2.2.3 Harvest time

Development and maturation of soybean

The development and maturation of soybean plants are divided into two stages, vegetative and reproductive, as listed in Table 2.1 (Fehr and Caviness, 1977). The vegetative phase starts from stage VE when emergence begins, to stage V6 when the sixth node appears. Their corresponding descriptions are included in Table 2.1. The reproductive phase has eight stages, from R1 to R8. For the determinate type of soybean, R1 and R2 might occur simultaneously since flowering starts from the upper nodes of the main stem. There is around a three-day difference between R1 and R2 for indeterminate soybean genotypes because the flowering begins from the lower part of the main stem and gradually develops upward. At R3, the pods start to develop and reach their full size at R4. Afterward, the seeds (beans) begin to enlarge at stage R5 and gradually reach full size during the R6 stage. The sign of stage R7 is that at least one normal pod on the main stem has turned to a mature color. At R8, 95% of pods have reached their mature color. Chemical composition of edamame changes with different development stages of the soybean plant. Yazdi-Samadi et al. (1977) and Saldivar et al. (2011)

investigated the changes in different chemical components, including oil, protein, sugars, starch, organic acids, and amino acids, during the development of soybean seeds. The study by Xu et al. (2016) showed the changes in the anti-nutritional properties of edamame beans over plant development.

The optimal harvest time of edamame

Identifying the optimal harvest time is vital for agro-food chains since it directly determines the quality of the products and shelf-life potential (Bertone et al., 2012; Costa et al., 2010). Ideally, edamame should be harvested between R6 and R7 stages, when moisture and bean weight reach the peak values and before the pods start to turn yellow (Moseley et al., 2021). Harvesting at the optimal harvest time ensures that the harvested edamame has the peak morphological and eating quality (Konovsky et al., 2020, Zeipiņa et al., 2017). However, because of the dynamic nature of bean growth between the R6 and R7 phases, harvesting edamame outside the appropriate harvest window may compromise its marketability. Harvesting too early, for example, can result in lower production, sweetness, and bean size, whereas harvesting too late results in fibrous and yellow beans (Carson, 2010). Additionally, the narrow harvest window (only one week) further complicates the edamame harvest. Currently, the determination of edamame optimal harvest time depends on the experience of edamame producers. They can discern these changes by taste, touch, or visual. However, this practice is exceptionally subjective, posing considerable challenges for inexperienced edamame producers. What is worse is the lower quality of edamame that might result in significant economic losses.

Table 2.1 Development and maturity stages of soybean.

Phase	Stage	Description
Vegetative phase	VE	Emergence of primary root or radical
	V.C.	Cotyledons fully exposed
	V1	First node (trifoliate)
	V2	Second node
	V3	Third node
	V6	Sixth node
Reproductive Phase	R1	Beginning bloom
	R2	Full bloom
	R3	Beginning pod
	R4	Full pod
	R5	Beginning seed
	R6	Full seed
	R7	Beginning maturity
	R8	Full maturity

2.3 The application of spectroscopic technique and machine learning in optimal harvest time determination

Because the spectroscopic technique can detect minor color changes that are typically invisible to the naked eye, recently, this technique has been investigated for its accuracy in different agricultural purposes, including determining the optimal harvest time of fruits. However, before it can be used for harvest time determination, this spectroscopy-based method requires calibration against a reference method. Calibration is typically performed using multivariate regression analysis; however, due to the complexity of the spectra, it does not always produce satisfactory results (Cortés et al., 2019). The recent developments in machine learning provide an efficient tool to analyze the spectroscopic dataset and provide accurate and reliable calibration (Singh et al., 2016, Singh et al., 2018). Gao et al. (2020) investigated the application of hyperspectral imaging and convolutional neural networking (in laboratory and field conditions) to estimate the ripeness of strawberries. They obtained a high accuracy of 98.6% when classifying the strawberries in the early ripe and ripe stages. These results showed a

promising real-time hyperspectral image system for determining the ripeness of the strawberry. Bertone et al. (2012) applied UV-VIS and NIR spectroscopy together with the Partial Least Squares algorithm in monitoring the changes of chlorophyll content in red apple skin throughout the tree ripening. This color change is difficult to observe during the ripening progress. When associated with chemometric elaboration, the UV-VIS and NIR spectroscopic techniques can be a solid way of predicting the best harvest time of red apples. Yang (2011) researched using VIS-NIR spectroscopy and Partial Least Squares Regression with Leave-one-out Cross Validation to determine the optimal harvest time of cherry tomatoes. The built prediction model reached a high determination coefficient (R^2) of 0.9. In the study of Lv et al. (2009), the VIS-NIR spectroscopy technique was combined with Partial Least Squares/Linear Discrimination Analysis or Principal Component Analysis/Linear Discrimination Analysis to classify less-ripen, ripen, and over-ripen grapes with high accuracy. This technique was also used to determine the optimum harvest time of other agricultural products such as oil pal fresh fruit (Cherie et al., 2019), wine grapes (Larraín et al., 2008), Verbena Officinalis (Pezzei et al., 2017) and nectarines (Eccher Zerbini et al., 2004).

Another algorithm, Random Forest (RF), is an ensemble learning technique that has become more popular due to its high classification accuracy and processing speed (Belgiu & Drăguţ, 2016). RF has frequently been used to classify foods based on obtained hyperspectral and multispectral data. In de Santana et al. (2019), RF was used to process the infrared spectroscopy dataset and successfully classified the authentic and adulterated nutmeg. It outperformed other algorithms, such as Partial Least Squares Discriminant Analysis (PLS-DA) and Soft Independent Modeling of Class Analogy (SIMCA). In another study by Peidad et al. (2018), three commonly used machine learning algorithms, including RF, Artificial Neural

Network, and Support Vector Machines, were investigated and used to classify bananas into several categories based on the banana size and color. According to the results, RF achieved the highest classification accuracy among the three machine learning algorithms.

All these studies showed that this non-destructive, clean, and affordable spectroscopy-based machine learning technology has great potential to be used by producers to improve the harvest of different agricultural products. Thus, it can also be applied to determine the optimal harvest time of edamame more rapidly, consistently, and conveniently.

2.4 Soluble dietary fiber production from agricultural food byproducts

Dietary fiber is classified into insoluble dietary fiber (IDF) and soluble dietary fiber (SDF) according to its solubility in water. SDF comprises oligosaccharides and various indigestible polysaccharides including gum, inulin, soluble hemicellulose, β-glucan, and pectin (Li et al., 2022; Dai and Chau, 2017;). Different studies found benefits of SDF intake: it can help minimize the risk of metabolic syndrome, reduce gained body weight, and control elevated plasma cholesterol (Wang et al., 2021; Artiss et al., 2006; Haskell, et al., 1992). SDF is superior to IDF because it can be metabolized by the gut microbiome more easily and the produced favorable metabolites can significantly affect the abundance and diversity of the gut microbiota (Guan et al., 2021). It can provide broader physiological activities and applications in foods such as modulating food fluidity, viscosity, and emulsification due to its unique physicochemical properties (Li et al., 2022; Anderson et al., 2009). Low-value food byproducts generated by food processing, such as tomato pomace, corn bran, pomegranate peel, and lemon peel, are rich sources of dietary fiber. However, the high IDF contents of these byproducts lead to undesirable sensory and food processing properties (Li et al., 2022). Therefore, many researchers seek extraction techniques that can effectively extract SDF and increase the SDF extraction yield from

food byproducts. Various extraction approaches have been investigated for this purpose and to improve the products' functional and nutritional properties.

Physical modification methods include single screw extruder, twin screw extrusion, blasting extrusion, high-hydrostatic pressure (HHP), high-pressure homogenization (HPH), cavitation jet processing, ultrafine grinding, and ball milling. Some representative methods to increase the SDF content are listed in Table 2.2. Ball milling stands out among other physical processing methods and received much attention because it is not only efficient at reducing particle size but also modifying the properties of dietary fiber. Different studies showed the potential of ball milling for increasing the SDF content in food byproducts and producing fiber with better physicochemical, functional, and nutritional properties. These properties include water holding and swelling capacity, oil and cholesterol binding capacity, emulsifying ability and stability, foaming ability and stability, antioxidant activity, total phenolic, and flavonoid contents, etc. In the study of Chen et al. (2020), ball milling was investigated to process okara (soybean residue). They found that SDF content in okara was increased from 3.7 to 34.9%, and the treated okara showed significantly smaller particle size, and higher phenolic and flavonoid contents compared to the untreated okara. Chitrakar et al. (2020) treated asparagus leaf byproduct using a ball mill, and the results showed that SDF content was increased by 9.2%. They also found that ball milling increased the total phenolic and flavonoid contents and antioxidant activity. Bender et al. (2020) reported that ball milling significantly increased the SDF contents and decreased the contents of lignin and cellulose in grape pomace. These studies proved that ball milling is a promising technique that can produce promising nutritional dietary fiber supplements for food industries.

Table 2.2 Representative physical processing methods for modifying IDF and producing SDF from food byproducts.

Method	Material	Output	Reference
Single screw extruder	Orange pomace	SDF content increased from 17.3 to 30.3%	Huang and Ma (2016)
Twin screw extrusion	Garlic skin	SDF content increased from 5.3 to 15.9%	Guo et al. (2018)
Blasting extrusion	Soybean residue	SDF content increased by 11.6 times	Chen et al. (2014)
High hydrostatic pressure (HHP)	Okara	SDF content increased from 4.6 to 30.3%	Mateos-Aparicio et al. (2010)
HHP	Pear pomace	SDF content increased from 10.0 to 16.0%	Yan et al. (2019)
High-pressure homogenization (HPH)	Orange peel	SDF content increased from 13.3 to 15.9% and 9.8 to 12.8%, respectively	Zhang et al. (2020)
HPH	Purple-fleshed potatoes	SDF content increased from 17.9 to 39.3%	Xie et al. (2017)
Cavitation jet processing	Okara	SDF content increased by 5 times	Wu et al. (2020)
Ultrafine grinding	Buckwheat hulls	SDF content increased from 16.0 to 26.6%	Zhu et al. (2014)
Low-temperature ball milling	Asparagus leaf byproduct	SDF content increased from 10.3 to 19.5%	Chitrakar et al. (2020)
Planetary ball milling	Okara	SDF content increased from 3.7 to 34.9%	Chen et al. (2020)

Alkali and acid are also used for modifying or extracting SDF from food byproducts and some studies are listed in Table 2.2. Fent et al. (2017) treated black bean coats with alkaline hydrogen peroxide (AHP) and found that SDF content increased from 7.8 to 16.9%. The modified SDF not only showed stronger gelation capacity but also good *in vitro* bile acid binding ability. The SDF extraction yield from sweet potato residue was increased from 1.7 to 3.3% after AHP treatment and the produced SDF held better gelling, water holding and swelling, and oil

holding capacities (Liu et al., 2021). Šoronja-Simović et al. (2016) also reported a higher SDF/IDF ratio and better water binding and swelling capacities of AHP-modified fiber in sugar beet pulp. The reason could be that AHP can partially solubilize lignin in the cell wall by breaking the hydrogen bonds of the molecular chains. This reaction opened more internal structures of the cell wall fiber, thus leading to the increase of the SDF content. Other alkaline chemicals were also investigated for extracting SDF from food byproducts. Huang et al. (2015) used Na_2HPO_4 to process okara, and their results showed that the SDF yield reached as high as 57.2%. This could be due to the breakage of glycosidic bonds or other weak bonds of polysaccharides. However, the Na_2HPO_4 treatment didn't lead to better cholesterol- and bile acid-binding capacities and even decreased the antioxidant activity. Moczkowska et al. (2019) investigated the effects of Na_2CO_3 treatment on SDF extraction from flaxseed. They found a high yield, but it was due to the high content of protein and other impurities. The results of this study indicated that the alkaline method tends to extract more protein and impurities instead of the SDF.

Certain acids were investigated for extracting SDF from food byproducts. Dong et al. (2020) conducted acid treatment on coffee peel using hydrochloride acid to extract the SDF. Around 9.2% of SDF was extracted due to the breakage of glycosidic linkages of the cell wall fiber. However, they mentioned that a strong acid condition might cause the loss of SDF, hemicellulose, and cellulose. Meanwhile, the SDF extracted by pure acid treatment didn't show superior properties compared to other combined methods which will be discussed in the later session. Sulfuric acid was used by Li et al. (2014) to extract SDF from apple pomace and 10.3% of SDF could be successfully extracted. However, SDF produced by conventional acid extraction

didn't show competitive properties (oil retention, water retention, swelling capacities) compared to other tested integrated methods.

Table 2.3 Representative chemical processing methods for modifying IDF and producing SDF from food byproducts.

Method	Material	Output	Reference
Alkaline hydrogen peroxide (AHP)	Black bean coat	SDF content increased from 7.8 to 16.9%	Feng et al. (2017)
AHP	Sweet potato residue	SDF yield increased from 1.7 to 3.3%	Liu et al. (2021)
AHP	Sugar beet pulp	SDF content increased by 3.5%	Šoronja-Simović et al. (2016)
Alkali (Na_2HPO_4)	Okara	SDF yield increased to 57.2%	Huang et al. (2015)
Alkali (Na_2CO_3)	Flaxseed	SDF yield was 17.7%	Moczkowska et al. (2019)
Subcritical water	Wheat bran	SDF yield was 12.9%	Yan et al. (2019)
Acid (HCl)	Coffee peel	SDF yield was 9.2%	Dong et al. (2020)
Acid (H_2SO_4)	Apple pomace	SDF yield was 10.3%	Li et al. (2014)

Enzymatic and fermentation are the other two processing methods that have been investigated to modify or extract SDF from food processing byproducts, and some representative studies were listed in Table 2.4. Three common enzymes used are cellulase, xylanase, and multi-enzymatic complex (Viscozyme L). Cellulase was tested on treating rice bran in the study conducted by Wen et al. (2017) and SDF contents increased from 1.5 to 7.2%. In this study, xylanase was also used, and the amount of SDF was further increased to 8.6%. The authors also found that enzymatic processing modified the functional properties of the rice bran dietary fiber. Water and oil holding capacities decreased while cholesterol absorption and sodium taurocholate absorption capacities increased. After treating the tomato peels with Viscozyme L, its SDF extraction yield increased by 72.3% (Gu et al., 2020). The results also showed that enzyme-

treated SDF had a better gelling ability, hydration properties, and glucose-adsorption capacity than original tomato peels. Ma et al. (2022) used cellulose and xylanase to treat potato residue and successfully increased its SDF contents from 17.5 to 26.8%. Cellulase and xylanase modification also improved the functional properties of potato residue dietary fiber such as swelling, cation exchange, water and oil holding, and cholesterol and glucose absorption capacities. However, the high enzyme price constantly jeopardizes its extended usage compared to the physical and chemical processing methods.

Table 2.4 summarized some fermentation processing methods. Jia et al. (2019) increased the SDF yield from defatted rice bran using *Trichoderma viride* fermentation and the produced SDF had better water and oil holding capacity, water solubility, and cholesterol absorption capacity. SDF was extracted from tea residues by *Trichoderma viride* fermentation and the extraction yield increased by eight folds. The produced SDF possessed higher uronic acid content, better thermal stability, and higher heavy metal ion binding capacity. *Monascus anka*, a type of fungus, increased the SDF content in okara by 1.8 times and improved the functional properties of dietary fiber in okara (Sun et al. 2020). Chu et al. (2019) fermented millet bran using *Bacillus natto*, increased the SDF contents from 2.3 to 13.2%, and enhanced both the physicochemical and functional properties of the produced dietary fiber. These processing methods showed promising performances in increasing SDF content and improving the properties of dietary fiber from different food byproducts. However, they focus more on modifying the DF instead of extracting the SDF.

Table 2.4 Representative enzymatic and fermentation methods for modifying IDF and producing SDF from food byproducts.

Method type	Method	Material	Output	Reference
Enzymatic	Viscozyme L	Tomato peels	SDF yield increased by 72.3%	Gu et al. (2020)
	Cellulase	Rice bran	SDF content increased from 1.5 to 7.2%	Wen et al. (2017)
	Xylanase	Rice bran	SDF content increased from 1.5 to 8.6%	Wen et al. (2017)
	Cellulase/xylanase	Potato residue	SDF content increased from 17.5% to 26.8%	Ma et al. (2022)
Fermentation	*Trichoderma viride* fermentation	Defatted rice bran	SDF yield increased from 10.5 to 33.4%	Jia et al. (2019)
	Trichoderma viride fermentation	Tea residues	SDF yield increased from 4.3 to 31.6%	Chen et al. (2020)
	Monascus anka fermentation	Okara	SDF content increased by 1.8 times	Sun et al. (2020)
	Bacillus natto fermentation	Millet bran	SDF content increased from 2.3 to 13.2%	Chu et al. (2019)

Integrated processing methods were investigated by many researchers to extract SDF from food processing byproducts (Table 2.5). The combined processing methods can lead to not only high SDF yield but also desirable SDF properties. One of the most common methods is the ultrasound-assisted method where the ultrasound treatment is combined with alkaline, acid, or enzyme to further improve the process effectiveness. The ultrasound-assisted alkaline method has been investigated in extracting SDF from papaya peel by Zhang et al. (2017) with an SDF yield of 37%. The produced SDF had higher thermal stability, water-holding, oil-holding, and swelling capacities than the SDF extracted by conventional alkaline treatment. Li et al. (2014) applied ultrasound-assisted acid treatment on apple pomace and achieved an SDF yield of 16.4%. Compared to conventional acid extraction, ultrasound-assisted acid extraction improved the water retention, swelling, and oil retention capacities. Ultrasound was also combined with

enzymes for extracting SDF from coffee peel and 13.0% of SDF was successfully extracted. The produced SDF showed better thermal stability and glucose adsorption capacity than SDF extracted by pure enzyme or acid treatment. Microwave is another common method that can be combined with other treatments. Gan et al. (2020) extracted SDF from grapefruit peel using microwave-assisted alkaline, microwave-assisted enzymatic and microwave-assisted ultrasonic treatments. Among these three different treatments, the microwave-assisted alkaline method led to the highest SDF yield. However, SDF extracted by microwave-assisted ultrasonic treatment possessed the best water-holding, oil-holding, glucose-adsorption, and cholesterol-adsorption capacities. The physical method (ball milling) was combined with the enzymatic method by Song et al. (2021) to modify IDF in lemon peel pomace. The SDF content was increased from 10.3 to 26.0% and they observed better water-holding, oil-retention, and water-swelling capacities on the produced SDF.

Table 2.5 Representative integrated processing methods for producing SDF from food byproducts.

Method	Material	Output	Reference
Ultrasound-assisted alkaline (NaOH)	Papaya peel	SDF yield was 37.0%	Zhang et al. (2017)
Ultrasound-assisted acid (sulfuric acid)	Apple pomace	SDF yield was 16.4%	Li et al. (2014)
Ultrasound-assisted enzymatic	Coffee peel	SDF yield was 13.0%	Dong et al. (2020)
Chemical (NaOH) & enzymatic	Defatted coconut flour	SDF yield was 12.0%	Du et al. (2021)
Chemical (hydrochloric acid) & enzymatic	Coffee peel	SDF yield was 11.4%	Dong et al. (2020)
Microwave-assisted chemical (NaOH)	Grapefruit peel	SDF yield increased from 3.5 to 17.2%	Gan et al. (2020)
Microwave-assisted enzymatic	Grapefruit peel	SDF yield increased from 3.5 to 9.1%	Gan et al. (2020)
Microwave-assisted ultrasound	Grapefruit peel	SDF yield increased from 3.5 to 8.4%	Gan et al. (2020)
Water media ball milling and enzymatic	Lemon peel pomace	SDF content increased from 10.3 to 26.0%	Song et al. (2021)

All the above-mentioned processing methods have their advantages and disadvantages. Physical methods are simple, cheap, and environmentally friendly but they are considered high-risk work and require large operating space (Gan et al., 2021). Chemical methods are efficient and practical but many times induce harmful chemical compounds to both SDF products and the environment. Biological methods are specific and efficient, but enzyme purification and strain breeding are usually expensive. The integrated processing methods stand out among all different treatments because they are not only efficient in extracting SDF from food byproducts but also can improve certain properties of produced SDF. Additionally, the integrated methods can meet various modification purposes within a shorter time and with less energy consumption. However, choosing the treatments that can be combined for extracting SDF from food byproducts effectively will be challenging. Moreover, the processing optimization will be more complicated

compared to the single treatment. Among different physical treatments, ball milling stands out because it is efficient at decreasing particle size as well as modifying the properties of dietary fiber. Compared to alkali, acid is more appropriate for extracting SDF since alkali solubilizes lignin, affecting the purity of the extracted SDF. Among different acids, citric acid, a commonly found organic acid, exists in many fruits, such as grapefruits, oranges, and pineapples (Behera, Mishra, and Mohapatra, 2021). It is commonly used as food ingredients/additives by food industries and can serve as a green chemical for extracting SDF from edamame shells. However, no study has been conducted previously on investigating the combined ball milling and citric acid treatment for extracting SDF from food byproducts. Therefore, it will be of great interest to see if the combined physical (ball milling) and chemical (citric acid) treatments can serve as a green method to extract SDF from edamame shells effectively and how it will affect the properties of produced SDF.

2.5 References

Anderson, J.W., Baird, P., Davis, R.H., Ferreri, S., Knudtson, M., Koraym, A., Waters, V. and Williams, C.L., 2009. Health benefits of dietary fiber. Nutrition Reviews, 67(4), pp.188-205.

Artiss, J.D., Brogan, K., Brucal, M., Moghaddam, M. and Jen, K.L.C., 2006. The effects of a new soluble dietary fiber on weight gain and selected blood parameters in rats. Metabolism, 55(2), pp.195-202.

Behera, B.C., Mishra, R. and Mohapatra, S., 2021. Microbial citric acid: production, properties, application, and future perspectives. Food Frontiers, 2(1), pp.62-76.

Belgiu, M. and Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, pp.24-31.

Berganza, B.E., Moran, A.W., Rodríguez, G.M., Coto, N.M., Santamaría, M. and Bressani, R., 2003. Effect of variety and location on the total fat, fatty acids and squalene content of amaranth. Plant Foods for Human Nutrition, 58(3), pp.1-6.

Bertone, E., Venturello, A., Leardi, R. and Geobaldo, F., 2012. Prediction of the optimum harvest time of 'Scarlet' apples using DR-UV–Vis and NIR spectroscopy. Postharvest Biology and Technology, 69, pp.15-23.

Cai, M.S., Li, J.B., Lin, J.Q., Chen, M.T., Liu, Y.H. and Li, Y.C., 2013. Determination and nutritional value evaluation of protein and vitamin C in edamame, soybeans and bean sprouts. Mod. Agric. Sci. Technol., 14, p.180.

Carson, L.C., Freeman, J.H., Zhou, K., Welbaum, G. and Reiter, M., 2011. Cultivar evaluation and lipid and protein contents of Virginia-grown edamame. HortTechnology, 21(1), pp.131-135.

Chen, J., Huang, H., Chen, Y., Xie, J., Song, Y., Chang, X., Liu, S., Wang, Z., Hu, X. and Yu, Q., 2020. Effects of fermentation on the structural characteristics and in vitro binding capacity of soluble dietary fiber from tea residues. LWT, 131, p.109818.

Chen, P.C., Lin, C., Chen, M.H. and Chiang, P.Y., 2020. The micronization process for improving the dietary value of okara (soybean residue) by planetary ball milling. LWT, 132, p.109848.

Chen, Y., Ye, R., Yin, L. and Zhang, N., 2014. Novel blasting extrusion processing improved the physicochemical properties of soluble dietary fiber from soybean residue and in vivo evaluation. Journal of Food Engineering, 120, pp.1-8.

Cherie, D., Rini, R. and Makky, M., 2019, September. Determination of the optimum harvest window and quality attributes of oil palm fresh fruit bunch using non-destructive shortwave infrared spectroscopy. In AIP Conference Proceedings (Vol. 2155, No. 1, p. 020034). AIP Publishing LLC.

Chitrakar, B., Zhang, M., Zhang, X. and Devahastin, S., 2020. Bioactive dietary fiber powder from asparagus leaf byproduct: Effect of low-temperature ball milling on physico-chemical, functional and microstructural characteristics. Powder Technology, 366, pp.275-282.

Clark, J., 2020. Managing soil and soil fertility after flooding. SDSU Extension. Available online at: https://extension.sdstate.edu/managing-soil-and-soil-fertility-after-flooding (accessed on March 21, 2020).

Cortés, V., Blasco, J., Aleixos, N., Cubero, S. and Talens, P., 2019. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. Trends in Food Science & Technology, 85, pp.138-148.

Chu, J., Zhao, H., Lu, Z., Lu, F., Bie, X. and Zhang, C., 2019. Improved physicochemical and functional properties of dietary fiber from millet bran fermented by *Bacillus* natto. Food Chemistry, 294, pp.79-86.

Choct, M., Dersjant-Li, Y., McLeish, J. and Peisker, M., 2010. Soy oligosaccharides and soluble non-starch polysaccharides: A review of digestion, nutritive and anti-nutritive effects in pigs and poultry. Asian-Australasian Journal of Animal Sciences, 23(10), pp.1386-1398.

Costa, G., Noferini, M., Fiori, G., Ziosi, V., Berthod, N. and Rossier, J., 2010. Establishment of the optimal harvest time in apricot ('Orangered' and 'Bergarouge') by means of a new index based on vis spectroscopy. Acta Hortic, 862, pp.533-538.

Dai, F.J. and Chau, C.F., 2017. Classification and regulatory perspectives of dietary fiber. Journal of Food and Drug Analysis, 25(1), pp.37-42.

Dong, W., Wang, D., Hu, R., Long, Y. and Lv, L., 2020. Chemical composition, structural and functional properties of soluble dietary fiber obtained from coffee peel using different extraction methods. Food Research International, 136, p.109497.

Dornbos, D.L. and Mullen, R.E., 1992. Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. Journal of the American Oil Chemists Society, 69(3), pp.228-231.

Du, X., Wang, L., Huang, X., Jing, H., Ye, X., Gao, W., Bai, X. and Wang, H., 2021. Effects of different extraction methods on structure and properties of soluble dietary fiber from defatted coconut flour. LWT, 143, p.111031.

de Santana, F.B., Neto, W.B. and Poppi, R.J., 2019. Random forest as one-class classifier and infrared spectroscopy for food adulteration detection. Food Chemistry, 293, pp.323-332.

Eccher Zerbini, P., Vanoli, M., Grassi, M., Rizzolo, A., Fibiani, M., Biscotti, G., Pifferi, A., Torricelli, A. and Cubeddu, R., 2004, June. Time-resolved reflectance spectroscopy as a non-destructive tool to assess the maturity at harvest and to model the softening of nectarines. In V International Postharvest Symposium 682 (pp. 1459-1464).

Eheart, J.F., Young, R.W. and Allison, A.H., 1955. Variety, type, year, and location effects on the chemical composition of peanuts. Food Research, 20, pp.497-505.

Fehr, W.R. and Caviness, C.E., 1977. Stages of soybean development (Vol. 80, p. 11). Ames: Iowa State University of Science and Technology.

Feng, Z., Dou, W., Alaxi, S., Niu, Y. and Yu, L.L., 2017. Modified soluble dietary fiber from black bean coats with its rheological and bile acid binding properties. Food Hydrocolloids, 62, pp.94-101.

Gao, Z., Shao, Y., Xuan, G., Wang, Y., Liu, Y. and Han, X., 2020. Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning. Artificial Intelligence in Agriculture, 4, pp.31-38.

Gan, J., Xie, L., Peng, G., Xie, J., Chen, Y. and Yu, Q., 2021. Systematic review on modification methods of dietary fiber. Food Hydrocolloids, 119, p.106872.

Gan, J., Huang, Z., Yu, Q., Peng, G., Chen, Y., Xie, J., Nie, S. and Xie, M., 2020. Microwave assisted extraction with three modifications on structural and functional properties of soluble dietary fibers from grapefruit peel. Food Hydrocolloids, 101, p.105549.

Ge, B.G., Wu, M.Y., He, F.T. and Zhao, Y., 2007. Study on the modification apple pomace dietary fiber. Food Technology, 10, pp.234-237.

Gideon, O.O., Richard, O.A. and Adekunle, A.A., 2016. Proline and soluble sugars accumulation in three pepper species (Capsicum spp) in response to water stress imposed at different stages of growth.

Guo, J., Rahman, A., Mulvaney, M.J., Hossain, M.M., Basso, K., Fethiere, R. and Babar, M.A., 2020. Evaluation of edamame genotypes suitable for growing in Florida. Agronomy Journal, 112(2), pp.693-707.

Gu, M., Fang, H., Gao, Y., Su, T., Niu, Y. and Yu, L.L., 2020. Characterization of enzymatic modified soluble dietary fiber from tomato peels with high release of lycopene. Food Hydrocolloids, 99, p.105321.

Guo, Y., Liu, W., Wu, B., Wu, P., Duan, Y., Yang, Q. and Ma, H., 2018. Modification of garlic skin dietary fiber with twin-screw extrusion process and in vivo evaluation of Pb binding. Food Chemistry, 268, pp.550-557.

Haskell, W.L., Spiller, G.A., Jensen, C.D., Ellis, B.K. and Gates, J.E., 1992. Role of water-soluble dietary fiber in the management of elevated plasma cholesterol in healthy subjects. The American Journal of Cardiology, 69(5), pp.433-439.

Huang, S., He, Y., Zou, Y. and Liu, Z., 2015. Modification of insoluble dietary fibres in soya bean okara and their physicochemical properties. International Journal of Food Science & Technology, 50(12), pp.2606-2613.

Huang, Y.L. and Ma, Y.S., 2016. The effect of extrusion processing on the physiochemical properties of extruded orange pomace. Food Chemistry, 192, pp.363-369.

Jia, M., Chen, J., Liu, X., Xie, M., Nie, S., Chen, Y., Xie, J. and Yu, Q., 2019. Structural characteristics and functional properties of soluble dietary fiber from defatted rice bran obtained through *Trichoderma viride* fermentation. Food Hydrocolloids, 94, pp.468-474.

Jiang, G.L., Rutto, L.K. and Ren, S., 2018. Evaluation of soybean lines for edamame yield traits and trait genetic correlation. HortScience, 53(12), pp.1732-1736.

Jiang, Z., Mu, S., Ma, C., Liu, Y., Ma, Y., Zhang, M., Li, H., Liu, X., Hou, J. and Tian, B., 2022. Consequences of ball milling combined with high-pressure homogenization on structure, physicochemical and rheological properties of citrus fiber. Food Hydrocolloids, 127, p.107515.

Jing, Y. and Chi, Y.J., 2013. Effects of twin-screw extrusion on soluble dietary fibre and physicochemical properties of soybean residue. Food Chemistry, 138(2-3), pp.884-889.

Johnson, D., Wang, S., & Suzuki, A., 1999. Edamame: a vegetable soybean for Colorado. Perspectives on New Crops and New Uses, 385–387.

Kaiser, C., and Ernst, M., 2013. Edamame.

Konovsky, J., Lumpkin, T.A. and McClary, D., 2020. Edamame: the vegetable soybean. In Understanding the Japanese Food and Agrimarket (pp. 173-181). CRC Press.

Kumar, V., Rani, A., Billore, S.D. and Chauhan, G.S., 2006. Physico-chemical properties of immature pods of Japanese soybean cultivars. International Journal of Food Properties, 9(1), pp.51-59.

Kumar, V., Rani, A., Goyal, L., Dixit, A.K., Manjaya, J.G., Dev, J. and Swamy, M., 2010. Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. Journal of Agricultural and Food Chemistry, 58(8), pp.5081-5085.

Larraín, M., Guesalaga, A.R. and Agosín, E., 2008. A multipurpose portable instrument for determining ripeness in wine grapes using NIR spectroscopy. IEEE Transactions on Instrumentation and Measurement, 57(2), pp.294-302.

Lara, L.M., Wilson, S.A., Chen, P. and Atungulu, G.G., 2019. The effects of infrared treatment on physicochemical characteristics of vegetable soybean. Heliyon, 5(1), p.e01148.

Li, S., Hu, N., Zhu, J., Zheng, M., Liu, H. and Liu, J., 2022. Influence of modification methods on physicochemical and structural properties of soluble dietary fiber from corn bran. Food Chemistry: X, 14, p.100298.

Li, X., He, X., Lv, Y. and He, Q., 2014. Extraction and functional properties of water‐soluble dietary fiber from apple pomace. Journal of Food Process Engineering, 37(3), pp.293-298.

Li, X., Welbaum, G.E., Rideout, S.L., Singer, W. and Zhang, B., 2022. Vegetable Soybean and Its Seedling Emergence in the United States.

Liu, M., Zhou, S., Li, Y., Tian, J. and Zhang, C., 2021. Structure, physicochemical properties and effects on nutrients digestion of modified soluble dietary fiber extracted from sweet potato residue. Food Research International, 150, p.110761.

Liu, Z., Jiang, L., Li, X., Härdter, R., Zhang, W, Zhang, Y., and Zheng, D., 2008. Effect of N and K fertilizers on yield and quality of greenhouse vegetable crops. Pedosphere, 18(4), pp.496-502.

Lv, G., Yang, H., Xu, N. and Mouazen, A.M., 2012, April. Identification of less-ripen, ripen, and over-ripen grapes during harvest time based on visible and near-infrared (Vis-NIR) spectroscopy. In 2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet) (pp. 1067-1070). IEEE.

Ma, Q., Ma, Z., Wang, W., Mu, J., Liu, Y., Wang, J., Stipkovits, L., Hui, X., Wu, G. and Sun, J., 2022. The effects of enzymatic modification on the functional ingredient-dietary fiber extracted from potato residue. LWT, 153, p.112511.

Masuda, R., 1991. Quality requirement and improvement of vegetable soybean. Vegetable Soybean Research Needs for Production and Quality Improvement, pp.92-102.

Mateos-Aparicio, I., Mateos-Peinado, C. and Rupérez, P., 2010. High hydrostatic pressure improves the functionality of dietary fibre in okara byproduct from soybean. Innovative Food Science & Emerging Technologies, 11(3), pp.445-450.

McEnroe, P. and Coulter, B., 1964. Effect of soil pH on sugar content and yield of sugar beet. Irish Journal of Agricultural Research, pp.63-69.

Miles, C.A., Daniels, C.H., Zenz, L. and King, J., 2018. Edamame. PNW (Series), 525, Pullman, Washington: Washington State University Extension

Moczkowska, M., Karp, S., Niu, Y. and Kurek, M.A., 2019. Enzymatic, enzymatic-ultrasonic and alkaline extraction of soluble dietary fibre from flaxseed–A physicochemical approach. Food Hydrocolloids, 90, pp.105-112.

Moseley, D., Da Silva, M.P., Mozzoni, L., Orazaly, M., Florez-Palacios, L., Acuna, A., Wu, C. and Chen, P., 2021. Effect of planting date and cultivar maturity in edamame quality and harvest window. Frontiers in Plant Science, 11, p.585856.

Nolen, S., Zhang, B. and Kering, M.K., 2016. Increasing fresh edamame bean supply through season extension techniques. Journal of Horticulture, pp.1-5.

Pezzei, C.K., Schönbichler, S.A., Kirchler, C.G., Schmelzer, J., Hussain, S., Huck-Pezzei, V.A., Popp, M., Krolitzek, J., Bonn, G.K. and Huck, C.W., 2017. Application of benchtop and portable near-infrared spectrometers for predicting the optimum harvest time of *Verbena officinalis*. Talanta, 169, pp.70-76.

Poe, N.E., Yu, D., Jin, Q., Ponder, M.A., Stewart, A.C., Ogejo, J.A., Wang, H. and Huang, H., 2020. Compositional variability of food wastes and its effects on acetone-butanol-ethanol fermentation. Waste Management, 107, pp.150-158.

Rao, M.S.S., Bhagsari, A.S. and Mohamed, A.I., 2002. Fresh green seed yield and seed nutritional traits of vegetable soybean genotypes. Crop Science, 42(6), pp.1950-1958.

Sakla, A.B., Ghali, Y., El-Farra, A. and Rizk, L.F., 1988. The effect of environmental conditions on the chemical composition of soybean seeds: Relationship between the protein, oil, carbohydrate and trypsin inhibitor content. Food Chemistry, 29(3), pp.221-231.

Saldivar, X., Wang, Y.J., Chen, P. and Hou, A., 2011. Changes in chemical composition during soybean seed development. Food Chemistry, 124(4), pp.1369-1375.

Shen, F., Zhang, B., Cao, C. and Jiang, X., 2018. On-line discrimination of storage shelf-life and prediction of post-harvest quality for strawberry fruit by visible and near infrared spectroscopy. Journal of Food Process Engineering, 41(7), p.e12866.

Shurtleff, W. and Aoyagi, A., 2014. History of soybeans and soyfoods in Japan, and in Japanese cookbooks and restaurants outside Japan (701 CE to 2014). Soyinfo Center.

Sirtori, C.R., 2001. Risks and benefits of soy phytoestrogens in cardiovascular diseases, cancer, climacteric symptoms and osteoporosis. Drug Safety, 24(9), pp.665-682.

Song, J.Y., An, G.H. and Kim, C.J., 2003. Color, texture, nutrient contents, and sensory values of vegetable soybeans [*Glycine max* (L.) Merrill] as affected by blanching. Food Chemistry, 83(1), pp.69-74.

Song, J., Liu, C., Li, D. and Gu, Z., 2013. Evaluation of sugar, free amino acid, and organic acid compositions of different varieties of vegetable soybean (*Glycine* max [L.] Merr). Industrial Crops and Products, 50, pp.743-749.

Song, L.W., Qi, J.R., Liao, J.S. and Yang, X.Q., 2021. Enzymatic and enzyme-physical modification of citrus fiber by xylanase and planetary ball milling treatment. Food Hydrocolloids, 121, p.107015.

Sun, C., Wu, X., Chen, X., Li, X., Zheng, Z. and Jiang, S., 2020. Production and characterization of okara dietary fiber produced by fermentation with *Monascus anka*. Food Chemistry, 316, p.126243.

Šoronja-Simović, D., Šereš, Z., Maravić, N., Djordjević, M., Djordjević, M., Luković, J. and Tepić, A., 2016. Enhancement of physicochemical properties of sugar beet fibres affected by chemical modification and vacuum drying. Food and Bioproducts Processing, 100, pp.432-439.

Wang, G., Zhong, D., Liu, H., Yang, T., Liang, Q., Wang, J., Zhang, R. and Zhang, Y., 2021. Water soluble dietary fiber from walnut meal as a prebiotic in preventing metabolic syndrome. Journal of Functional Foods, 78, p.104358.

Wen, L., Zhang, Z., Zhao, M., Senthamaraikannan, R., Padamati, R.B., Sun, D.W. and Tiwari, B.K., 2020. Green extraction of soluble dietary fibre from coffee silverskin: Impact of ultrasound/microwave‐assisted extraction. International Journal of Food Science & Technology, 55(5), pp.2242-2250.

Wen, Y., Niu, M., Zhang, B., Zhao, S. and Xiong, S., 2017. Structural characteristics and functional properties of rice bran dietary fiber modified by enzymatic and enzyme-micronization treatments. LWT, 75, pp.344-351.

Wolf, R.B., Cavins, J.F., Kleiman, R. and Black, L.T., 1982. Effect of temperature on soybean seed constituents: oil, protein, moisture, fatty acids, amino acids and sugars. Journal of the American Oil Chemists' Society, 59(5), pp.230-232.

Wu, C., Teng, F., McClements, D.J., Zhang, S., Li, Y. and Wang, Z., 2020. Effect of cavitation jet processing on the physicochemical properties and structural characteristics of okara dietary fiber. Food Research International, 134, p.109251.

Xie, F., Li, M., Lan, X., Zhang, W., Gong, S., Wu, J. and Wang, Z., 2017. Modification of dietary fibers from purple-fleshed potatoes (Heimeiren) with high hydrostatic pressure and high-pressure homogenization processing: A comparative study. Innovative Food Science & Emerging Technologies, 42, pp.157-164.

Xu, L., Jiang, F., Wu, Z., Wang, J. and Zhou, X., 2009. Effects of substrate water content and light intensity on growth and quality of non-heading Chinese cabbage. Jiangsu Journal of Agricultural Sciences, 25(4), pp.865-870.

Xu, Y., Cartier, A., Kibet, D., Jordan, K., Hakala, I., Davis, S., Sismour, E., Kering, M. and Rutto, L., 2016. Physical and nutritional properties of edamame seeds as influenced by stage of development. Journal of Food Measurement and Characterization, 10(2), pp.193-200.

Xu, Y., Sismour, E., Pao, S., Rutto, L., Grizzard, C. and Ren, S., 2012a. Textural and microbiological qualities of vegetable soybean (edamame) affected by blanching and storage conditions. Journal of Food Processing & Technology, 3(6), pp.1-6.

Yan, J.K., Wu, L.X., Cai, W.D., Xiao, G.S., Duan, Y. and Zhang, H., 2019. Subcritical water extraction-based methods affect the physicochemical and functional properties of soluble dietary fibers from wheat bran. Food Chemistry, *298*, p.124987.

Yu, D., O'Hair, J., Poe, N., Jin, Q., Pinton, S., He, Y. and Huang, H., 2022. Conversion of food waste into 2, 3-butanediol via thermophilic fermentation: effects of carbohydrate content and nutrient supplementation. Foods, 11(2), p.169.

Yan, L., Li, T., Liu, C. and Zheng, L., 2019. Effects of high hydrostatic pressure and superfine grinding treatment on physicochemical/functional properties of pear pomace and chemical composition of its soluble dietary fibre. LWT, 107, pp.171-177.

Yang, H.Q., 2011. Non-destructive prediction of optimal harvest time of cherry tomatoes using VIS-NIR spectroscopy and PLSR calibration. In Advanced Engineering Forum (Vol. 1, pp. 92-96). Trans Tech Publications Ltd.

Yang, S.Y., Li, J.H., and Zhu, L.L., 2009. Effects of light intensity on quality of pepper fruits. Northern Horticulture. 2, 65–67.

Yazdi‐Samadi, B., Rinne, R.W. and Seif, R.D., 1977. Components of developing soybean seeds: Oil, protein, sugars, starch, organic acids, and amino acids[1]. Agronomy Journal, 69(3), pp.481-486.

Zeipiņa, S., Alsiņa, I. and Lepse, L., 2017. Insight in edamame yield and quality parameters: A review. Research for Rural Development, 2, pp.40-45.

Zhang, Q., Li, Y., Chin, K.L. and Qi, Y., 2017. Vegetable soybean: Seed composition and production research. Italian Journal of Agronomy, 12(3).

Zhang, Q.Y., Li, Y.S. and Liu, X.B., 2015. Edible quality and its regulation in vegetable soybean (*Glycine max* (L.) Merr.). In ASA, CSSA, SSSA International Annual Meetings Nov (pp. 15-18).

Zhang, Y., Qi, J., Zeng, W., Huang, Y. and Yang, X., 2020. Properties of dietary fiber from citrus obtained through alkaline hydrogen peroxide treatment and homogenization treatment. Food Chemistry, 311, p.125873.

Chapter 3

Chemical Compositions of Edamame Genotypes Grown in Different Locations in the U.S.

3.1 Introduction

Edamame, more commonly referred to as "vegetable or edible soybean" in the U.S. and "*mao dou*" in China, belongs to the same species as grain soybean [*Glycine max* (L.) Merr.]. It is an important vegetable in Asia and has been widely consumed in China and Japan for centuries as a snack or side dish (Zeipina et al., 2017; Lara et al., 2019). In the U.S., the consumption of frozen edamame has been increasing, and it has become the second most highly consumed soy-food after soymilk because it is both nutritious and has potential health benefits (Zeipina et al., 2017). However, more than 70% of the edamame consumed in the U.S. market is imported from overseas. Although there are several edamame lines, such as UA Kirksey and UA Mulberry, developed for U.S. domestic production (Chen et al., 2017), it is still critically needed to continuously breed elite US edamame genotypes with high nutritional value and consumer acceptance to meet the ever-increasing market demand.

The increasing popularity of edamame is primarily due to its nutritional value and health benefits. The nutritional value of edamame is mainly determined by its chemical constituents, such as protein, fiber, starch, and sugars. It was reported that edamame is highly nutritious because of the content of high-quality protein with isoflavones (Zeipina et al., 2017). Edamame is also a good source of dietary fiber when supplemented with the diet (Johnson et al., 1999; Xu et al., 2012). Dietary fiber could help reduce blood cholesterol levels because of its viscosity, solubility, and ability to adsorb/bind molecules (Lin et al., 2020). Moreover, edamame contains a

significant amount of health-promoting polyunsaturated fatty acids, such as linoleic acid and α-linolenic acid (Kumar et al., 2006a). In addition, edamame is a good source of vitamins (C, E, and thiamin), minerals, phytochemicals, and other active compounds (Johnson et al., 1999; Song et al., 2003; Kaiser and Ernst, 2013; Zeipina et al., 2017). With these superior nutritional constituents, edamame has potential health benefits to help reduce the risk of many diseases, such as cardiovascular disease, cancer, and osteoporosis (Sirtori, 2001).

Besides nutrition, sensory attributes are also important for consumer acceptance of edamame. Compared to soybean, edamame is slightly sweeter, has a softer texture and a less beany and nutty flavor (Konovsky et al., 1994). Sweetness is one of the most important sensory attributes of edamame and is a primary indicator of edamame quality (Johnson et al., 1999). Carneiro et al. (2020) reported that sweetness is considered a major sensory attribute that leads to higher consumer acceptability. The sweetness of edamame is mainly determined by soluble sugars including sucrose, glucose, and fructose (Konovsky et al., 1994; Song et al., 2013; Zeipina et al., 2017). Alanine, which has a sweet taste, also contributes to sweetness (Kirimura et al., 1969). Therefore, identification of edamame with higher soluble sugars and alanine levels may increase the popularity of domestically produced edamame.

Many studies have been conducted to analyze the seed composition of soybeans but very few studies have specifically investigated the chemical composition of edamame. Recently, Jiang et al. (2020) evaluated the chemical composition (e.g., protein, oil, and sugar) of different edamame genotypes, which is essential for edamame research and breeding. However, planting location, another important factor to consider for producing edamame, was not considered in the Jiang et al. (2020) study. Previous studies have shown that the planting location often had a greater influence on the chemical composition of crops [e.g., peanut (Arachis hypogaea L.),

amaranth (Amaranthus cruentus)] than genotype (Eheart et al., 1955; Berganza et al., 2003). Sakla et al. (1988) evaluated the environmental effects on the chemical composition of three soybean varieties. The results showed that significantly different moisture, oil, carbohydrates, sucrose, and protein contents were due to the localities. No significant difference was observed in fiber and ash contents among different varieties and locations. The effects of environmental conditions on the chemical composition of soybean seeds were also reported in other literature (Cartter et al., 1942; McClure et al., 2017). However, Cartter et al. (1942) reported that varieties played a more important role in affecting the chemical composition of soybean seed in their studies compared to locations. Moreover, it is important to identify genotypes that can perform well at different locations. Besides chemical composition, it is also important to investigate the sweetness of edamame in order to breed edamame with high sensory attributes and thus increased consumer acceptance. Therefore, the objective of this study is to systematically evaluate the chemical composition and sweetness of 10 selected edamame genotypes grown in three locations in the U.S. Three soluble sugars (sucrose, fructose, and glucose) and free alanine were quantified to estimate edamame sweetness. Comprehensive chemical compounds including oligosaccharides (raffinose and stachyose), crude protein, oil, starch, moisture content of fresh beans, neutral detergent fiber (NDF), and ash content were also determined to build a complete chemical composition profile of the edamame and the data was used to observe the effect of planting location and genotype.

3.2 Materials and methods

3.2.1 Plant materials, sample pre-treatments, chemicals, and reagents

One commercial check (cultivar UA-Kirksey) and nine edamame breeding lines were planted in late May 2018 at three locations—Virginia Tech's Kentland Farm in Whitethorne VA,

University of Arkansas Research Farm in Little Rock AR, and Virginia Tech's Eastern Shore Agricultural Research and Extension Center in Painter VA. All genotypes are from the maturity group V. The names of the selected genotypes are R13-5029, R14-6450, V10-3653, V13-0329, V13-0339, V13-1644, V15-0396, V16-0523, and V16-0547. Each genotype was grown in two replications and the plots were arranged in a randomized complete block design (RCBD). Each plot was 6.10 m long, 0.75 m inter-row spacing and a seeding rate of 20 seeds per m. Half pound of edamame pods was manually harvested at the R6 stage of soybean development when beans filled 80–90% of the pod cavity. Edamame samples were stored in coolers filled with ice bags and transferred to the food processing pilot plant at Virginia Tech (Whitethorne, VA, USA) for processing. The pods were blanched in boiling water (98.3 ± 0.1 °C) for 1 min to inactivate enzymes and decrease the microbial load (Pao et al., 2008; Xu et al., 2012). Blanched samples were immediately cooled in an ice-water bath for 2 min to avoid overcooking (4.5 ± 0.5 °C) and then dried with a paper towel until no flowing water was observed. Afterward, the beans were shelled out manually and stored at −80 °C. Frozen beans were freeze-dried and then milled by an IKA MF 10 Basic Microfine Grinder (IKA®-Werke GmbH & Co. KG, Germany) and passed through a 500 μm sieve. All chemical standards (sugars and alanine) and sodium tetraborate decahydrate ($Na_2B_4O_7 \cdot 10H_2O$) were purchased from Sigma-Aldrich (St. Louis, MO, USA). AdvanceBio AAA standards and reagents kit was purchased from Agilent for alanine analysis. The rest of the chemicals used in this study were purchased from Fisher Scientific (Hampton, NH, USA).

3.2.2 Free sugars, alanine, and sweetness

Free sugars and free alanine

The extraction of free sugars (sucrose, fructose, glucose, raffinose, and stachyose) and free alanine in edamame was conducted based on the method reported by Yu et al. (2016) and Machado et al. (2020), with some adjustments. Briefly, 1.5 mL of deionized water and 0.15 g of dry powdered edamame were mixed in a 2 mL centrifuge tube. The tube was placed on a tube revolver (Thermo ScientificTM, Waltham, MA, USA) and shaken for 2 h at room temperature, followed by centrifugation at 13,500 × g for 10 min. Afterward, 750 µL acetonitrile was added to 750 µL of the supernatant for purification. The mixture was shaken at room temperature for 10 min, and then centrifugated at 13,500 × g for 10 min. After centrifugation, 750 µL of the supernatant was filtered through a 0.2-µm membrane disc filter into a 1.5-mL HPLC vial and sugar content was determined using high-performance liquid chromatography (HPLC, 1260 Infinity II, Agilent Technologies, Santa Clara, CA, USA) equipped with a refractive index detector (RID). The Luna Omega 3 µm SUGAR column (150 × 4.6 mm, Phenomenex, Torrance, CA, USA) was used to separate different sugars. The column temperature was set at 40 °C with a flow rate of 1.0 mL/min. The mobile phase was a mixed solution of acetonitrile/water (75:25 v: v) and the injection volume was 5 µL. The measurement of alanine content was conducted according to an application developed by Agilent (Palaniswamy, 2017) using the HPLC. Derivatization was done automatically by an autosampler using chemicals in the Agilent AdvanceBio AAA standards and reagents kit (p/n 5190-9426). The injection volume was 1.0 µL and alanine was separated in an Agilent AdvanceBio AAA C18 column, 4.6 × 100 mm, 2.7 µm, (40 °C) at a flow rate of 1.5 mL/min with a gradient program. Mobile phase A contained 10 mM Na_2HPO_4, and 10 mM $Na_2B_4O_7$ at pH of 8.2, while the mobile phase B was the mixed solution

of acetonitrile, methanol, and water (45:45:10, v: v: v). Eluted alanine from the column was detected by a diode array detector at λ = 338 nm.

Sweetness

Sucrose, glucose, fructose, and alanine all contribute to the sweetness of edamame; therefore, all of them should be taken into consideration when determining edamame sweetness (Monteiro et al., 2007; Saldivar et al., 2010). Because sucrose, glucose, fructose, and alanine have different sweetness intensities, sucrose is commonly used as a reference to compare the sweetness intensities of different sugars and its sweetness is set at 1.00 (Brady, 2013). The relative sweetness (RS) of glucose, fructose, and alanine to the same concentration of sucrose is 0.40 – 0.79 (Brady, 2013), 1.00 – 1.75 (Brady, 2013), and 0.93 – 1.70 (Cameron, 1945) respectively. Thus, edamame sweetness was calculated using the following equation based on the concentrations and averaged relative sweetness of different sugars

$$Sweetness = C_{sucrose} \times RS_{sucrose} + C_{glucose} \times RS_{glucose} + C_{fructose} \times C_{fructose} + C_{alanine} \times RS_{alanine} \quad (3.1)$$

where C represented the concentration (mg/g) and RS was the average of the highest RS and the lowest RS of each sugar and alanine.

Protein, oil, neutral detergent fiber (NDF), starch, ash, and moisture

The Kjeldahl method was used to measure the nitrogen content in edamame and the protein content was calculated by multiplying with a conversion factor of 6.25 (AOAC, 2005a). The oil in the edamame was extracted using petroleum ether, and its content was measured according to AOAC 2003.05 (AOAC, 2005b). The NDF in edamame was measured by the ANKOM fiber analyzer (ANKOM Technology, Macedon, NY, USA). In brief, the non-fiber component in 0.5 g of dry sample powder in a filter bag was washed out by a neutral detergent

solution in the fiber analyzer. The dry weight after digestion was used to calculate the NDF content (Yu et al., 2020 and Ohair et al., 2020). To measure the ash content, edamame powder was placed in a muffle furnace at 550 °C and burned for 12 h according to AOAC 942.05 (AOAC, 2005c; He et al., 2019). Starch content was determined using the method described by Vidal et al. (2009) by measuring hydrolyzed glucose by HPLC with RID using Bio-Rad Aminex HPX-87H (Bio-Rad Laboratories, Hercules, CA, USA). Except for the moisture content measurement, chemical composition analysis was conducted on powdered edamame samples prepared in section Plant Materials, Sample Pre-treatments, Chemicals, and Reagents and reported on a dry matter basis. To measure the moisture content, fresh beans were dried in an oven at 105 °C until the weight was constant. Moisture content was calculated based on the weight difference before and after drying and reported on a wet matter basis.

Statistical analysis

All measurements were conducted on biological replicates of edamame and results were presented as means ± standard deviation (n = 2). Pearson's correlation analysis between each of the free sugars and total sweetness was conducted using GraphPad Prism (8.3.0, GraphPad Software, San Diego, CA, USA). Two-way ANOVA was performed to test the significant effects of genotype, location, and genotype × location on each of the chemical constituents, followed by a post-hoc Tukey's HSD (Honestly Significant Difference) for a pair-wise comparison of means using the statistical software SPSS (22.0.0.0, IBM Corporation, Armonk, NY, USA). The statistical significance level was 0.05 ($p < 0.05$). Principle component analysis (PCA) was conducted by MATLAB (R2020a, MathWorks, Natick, MA, USA).

3.3 Results and discussions

3.3.1 Free sugars, alanine, and total sweetness

Free sugars

The free sugars, including sucrose, fructose, and glucose, predominately decide the sweetness of edamame (Song et al., 2013). Sucrose was the most abundant soluble sugar in edamame and there was considerable variation in the sucrose content among samples (Table 3.2). Significant effects of genotype, location, and their interaction (genotype × location) were observed (Table 3.1). The edamame planted in Whitethorne had an average sucrose content of 59.29 mg/g, which was much higher than the samples planted in Little Rock (42.82 mg/g) and Painter (40.60 mg/g). Different temperatures could be a reason for the higher sucrose content of edamame planted in Whitethorne. Wolf et al. (1982) and Kumar et al. (2010) reported that increased temperature during seed development led to decreased sucrose content of soybean seeds. The average temperatures in Whitethorne from June to October were 8–9 °C lower than the average temperatures in Little Rock and 3–6 °C lower than the average temperatures in Painter. Besides temperature, soil types and properties, daylight intensity, and precipitation could also be reasons for the variation in sucrose content. Nutrients and pH differ among different soils and they might lead to sucrose variations in crops. Zhao-Hui et al. (2008) reported that soil containing a higher level of potassium (K) led to a higher soluble sugar content in kidney beans. In the study of McEnroe and Coulter (1964), increasing soil pH from 6 to 7 leads to increased sugar content in sugar beet. The light intensity affects the photosynthesis rate and thus the sugar content of vegetable crops. Xu et al. (2009) reported a positive correlation between light intensity and soluble sugar contents in non-heading Chinese cabbage. The same trend was also reported by Yang et al. (2009) that soluble sugars in pepper fruits decreased with decreasing light

intensity. Precipitation can also lead to water stress or flooding which potentially affects sugar contents in crops. Under appropriate water stress, plants tend to accumulate soluble sugars to adjust their osmotic pressure. Okunlola et al. (2016) investigated the total sugar accumulation of three pepper varieties under water stress and observed higher total sugar contents in pepper with less irrigation. Flooding will cause soil nutrient loss and then affect the sugar contents in the vegetables (Clark, 2020).

Table 3.1 *p*-values of the two-way ANOVA for all compositions.

Source of variance	***p*-value**						
	Sucrose	**Fructose**	**Glucose**	**Alanine**	**Total Sweetness**	**Raffinose**	**Stachyose**
Genotype	<0.0001	0.0210	0.0310	0.0150	0.0180	0.4440	0.0000
Location	<0.0001	0.0020	< 0.0001	<0.0001	0.0000	<0.0001	0.0000
Genotype * Location	<0.0010	0.1250	0.4460	0.1540	0.1360	0.1510	0.0000
	Protein	**Oil**	**NDF**	**Starch**	**Ash**	**Moisture Content**	
Genotype	<0.0001	<0.0001	0.3750	0.0010	<0.0001	0.0310	
Location	<0.0001	<0.0001	<0.0001	0.0050	<0.0001	<0.0001	
Genotype * Location	0.0010	0.1100	0.3290	0.0270	<0.0001	0.3000	

Table 3.2 Sucrose, fructose and glucose contents (mg/g dry matter) of 10 edamame genotypes planted in three locations.

Genotype	Whitethorne, VA	Little Rock, AR	Painter, VA	Mean (across locations)	Ranking
Sucrose					
R13-5029	60.78 ± 7.03	51.93 ± 4.60	46.71 ± 0.35	**53.14 ± 7.39^{a}**	**1**
R14-6450	33.17 ± 2.23	35.00 ± 3.83	41.23 ± 1.22	**36.46 ± 4.30^{c}**	**10**
UA-Kirksey	72.25 ± 2.51	47.29 ± 6.02	39.77 ± 1.79	**53.10 ± 15.50^{a}**	**2**
V10-3653	54.82 ± 18.39	37.00 ± 2.44	32.53 ± 3.13	**41.45 ± 13.50^{bc}**	**9**
V13-0329	62.56 ± 7.38	45.78 ± 3.75	46.47 ± 0.06	**51.60 ± 9.26^{ab}**	**3**
V13-0339	69.20 ± 4.72	40.97 ± 0.56	39.35 ± 1.91	**49.84 ± 15.19^{ab}**	**5**
V13-1644	70.62 ± 5.29	35.57 ± 5.29	33.38 ± 3.32	**46.52± 19.04^{abc}**	**7**
V15-0396	63.18 ± 2.17	45.15 ± 5.02	43.91 ± 1.25	**50.75 ± 9.97^{ab}**	**4**
V16-0523	51.71 ± 4.44	42.50 ± 0.54	39.09 ± 2.10	**44.43 ± 6.24^{abc}**	**8**
V16-0547	54.56 ± 5.41	47.03 ± 10.45	43.57 ± 3.13	**48.39 ± 7.41^{ab}**	**6**
Mean (across genotypes)	**59.29 ± 12.50^{A}**	**42.82 ± 6.60^{B}**	**40.60 ± 4.99^{B}**		
Fructose					
R13-5029	17.01 ± 5.79	8.67 ± 0.64	9.69 ± 1.13	**11.79 ± 4.86**	**3**
R14-6450	12.32 ± 1.59	14.44 ± 5.42	8.69 ± 2.42	**11.82 ± 3.78**	**2**
UA-Kirksey	9.72 ± 0.23	9.69 ± 0.62	8.26 ± 1.39	**9.22 ± 1.01**	**8**
V10-3653	12.82 ± 3.89	12.07 ± 1.45	8.54 ± 0.82	**11.14 ± 2.79**	**4**
V13-0329	7.24 ± 1.66	7.98 ± 1.03	8.36 ± 0.01	**7.86 ± 1.01**	**10**
V13-0339	11.70 ± 1.46	11.34 ± 0.44	9.45 ± 0.15	**10.83 ± 1.28**	**5**
V13-1644	7.43 ± 0.03	9.52 ± 0.56	9.74 ± 2.83	**8.90 ± 1.72**	**9**
V15-0396	10.73 ± 0.91	9.58 ± 1.50	7.45 ± 2.30	**9.25 ± 1.97**	**7**
V16-0523	10.62 ± 0.39	8.99 ± 0.29	9.05 ± 2.48	**9.55 ± 1.40**	**6**
V16-0547	14.63 ± 4.30	12.87 ± 1.11	8.41 ± 0.29	**11.97 ± 3.49**	**1**
Mean (across genotypes)	**11.42 ± 3.55^{A}**	**10.51 ± 2.46^{A}**	**8.76 ± 1.43^{B}**		
Glucose					
R13-5029	6.48 ± 2.24	2.74 ± 0.26	4.09 ± 1.44	**4.44 ± 2.08^{ab}**	**3**
R14-6450	5.97 ± 0.80	3.84 ± 1.31	3.38 ± 1.76	**4.40 ± 1.62^{ab}**	**4**
UA-Kirksey	3.99 ± 0.04	3.16 ± 0.63	3.04 ± 0.88	**3.40 ± 0.67^{ab}**	**8**
V10-3653	6.34 ± 1.86	2.87 ± 0.14	2.06 ± 0.43	**3.75 ± 2.21^{ab}**	**6**
V13-0329	2.84 ± 0.43	2.39 ± 0.08	2.65 ± 0.67	**2.63 ± 0.41^{b}**	**10**
V13-0339	4.42 ± 0.41	3.12 ± 0.31	3.36 ± 0.16	**3.64 ± 0.66^{ab}**	**7**
V13-1644	6.06 ± 0.36	3.87 ± 0.27	3.79 ± 2.30	**4.57 ± 1.56^{ab}**	**2**
V15-0396	4.76 ± 0.51	2.66 ± 0.03	2.54 ± 1.27	**3.32 ± 1.27^{ab}**	**9**
V16-0523	6.07 ± 0.21	2.77 ± 0.33	2.58 ± 0.36	**3.81 ± 1.77^{ab}**	**5**
V16-0547	6.90 ± 2.44	4.23 ± 0.42	3.14 ± 0.58	**4.75 ± 2.07^{a}**	**1**
Mean (across genotypes)	**5.38 ± 1.57^{A}**	**3.17 ± 0.70^{B}**	**3.06 ± 1.05^{B}**		

Different letters (abc & AB) indicate a significant difference based on the two-way ANOVA with Tukey's HSD test ($p < 0.05$).

On average, R13-5029 had the highest sucrose content (53.14 mg/g) while R14-6450 and V10-3653 had the lowest sucrose content among all genotypes (36.46 and 41.45 mg/g, respectively). The highly significant effect of genotype × location indicated that genotypes ranked differently among themselves from location to location for sucrose content. In Whitethorne, the check cultivar UA-Kirksey had the highest sucrose content (72.25 mg/g) followed by genotype V13-1644 (70.62 mg/g) and V13-0339 (69.20 mg/g). In Little Rock, R13-5029 had the highest sucrose content of 51.93 mg/g and R14-6450 had the lowest sucrose content of 35.00 mg/g. In Painter, genotype R13-5029 had the highest sucrose content of 46.71 mg/g while V10-3653 had the lowest sucrose content (32.53 mg/g). Overall, genotype, location, and their interaction all had a significant influence on sucrose content.

The contents of fructose and glucose were lower than the content of sucrose in edamame (Table 3.2). Fructose and glucose contents were affected by location. Edamame planted in Whitethorne had the highest average fructose and glucose contents (11.42 and 5.38 mg/g, respectively) followed by the samples planted in Little Rock (10.51 and 3.17 mg/g) and Painter (8.76 and 3.06 mg/g). Although significant differences in fructose and glucose contents were observed among the three locations, these differences were relatively small compared to those in sucrose.

Alanine

Alanine also tastes sweet and contributes to the sweetness of edamame. The alanine contents of different edamame genotypes planted in different locations are listed in Table 3.3. The alanine contents were much lower than the free sugar contents in edamame. The overall range of alanine was 0.15 – 3.47 mg/g. This result was consistent with a previous study in which the authors reported a low alanine content range of 0.21 – 1.76 mg/g in edamame (Song et al.,

2013). The variation in alanine content was mainly due to location—the edamame planted in Little Rock had a high average alanine content of 2.67 mg/g across genotypes, whereas the edamame planted in Painter and Whitethorne had relatively low average alanine contents of 1.87 and 0.43 mg/g, respectively. A significant effect of genotype on alanine was observed (Table 3.1) but no significant difference was found among different genotypes (Table 3.3).

Table 3.3 Alanine contents (mg/g dry matter) of 10 edamame genotypes planted in three locations.

Genotype	Whitethorne, VA	Little Rock, AR	Painter, VA	Mean (across locations)	Ranking
R13-5029	0.42 ± 0.23	3.47 ± 0.06	2.12 ± 0.41	**2.00 ± 1.38**	**1**
R14-6450	0.18 ± 0.04	3.45 ± 1.09	1.99 ± 0.02	**1.87 ± 1.55**	**4**
UA-Kirksey	0.87 ± 0.24	2.97 ± 0.55	2.14 ± 0.38	**1.99 ± 0.99**	**2**
V10-3653	0.30 ± 0.14	2.37 ± 0.19	1.85 ± 0.41	**1.51 ± 0.98**	**7**
V13-0329	0.29 ± 0.06	1.88 ± 0.28	1.59 ± 0.29	**1.25 ± 0.78**	**10**
V13-0339	0.57 ± 0.03	1.96 ± 0.13	1.62 ± 0.30	**1.38 ± 0.66**	**8**
V13-1644	0.41 ± 0.02	2.23 ± 0.56	1.51 ± 0.30	**1.38 ± 0.87**	**8**
V15-0396	0.79 ± 0.81	2.59 ± 0.34	1.83 ± 0.07	**1.74 ± 0.90**	**5**
V16-0523	0.32 ± 0.03	2.67 ± 0.26	1.58 ± 0.04	**1.52 ± 1.06**	**6**
V16-0547	0.15 ± 0.03	3.13 ± 1.06	2.49 ± 0.48	**1.93 ± 1.50**	**3**
Mean (across genotypes)	**0.43 ± 0.31^{C}**	**2.67 ± 0.70^{A}**	**1.87 ±0.38^{B}**		

Different letters (ABC) indicate a significant difference based on the two-way ANOVA with Tukey's HSD test ($p < 0.05$).

Total sweetness

The total sweetness of edamame was calculated based on the concentrations of sugars and alanine and their relative sweetness to sucrose. The total sweetness of the 10 edamame genotypes at three locations is shown in Table 3.4. Location has a profound and significant effect on the total sweetness. Edamame samples planted in Whitethorne had the highest sweetness among the three locations. Therefore, locations having similar weather and soil type to Whitethorne could be considered for producing sweet edamame. It is worth noting that the

genotype R13-5029 performed consistently well in all three locations—it ranks third, second, and first in total sweetness in Whitethorne, Little Rock, and Painter, respectively. The mean sweetness of R13-5029 across the three locations was 73.74 mg/g dry matter, higher than that of check cultivar UA-Kirksey (69.52 mg/g). Meanwhile, the genotype V16-0547 and V15-0396 also performed relatively well in all three locations and had the mean sweetness of 69.35 and 66.95 mg/g dry matter across the three locations. Therefore, these three top genotypes (R13-5029, V16-0547, and V15-0396) are the potential candidates for developing edamame with a high sweet taste. Overall, to breed sweeter edamame, both location and genotype need to be considered. Moreover, Pearson's correlation analysis between each of the free sugars (and alanine) and total sweetness showed that the sucrose content is strongly correlated with the total sweetness with a high correlation coefficient (r) of 0.94, whereas the r values between the fructose, glucose, and alanine content and the total sweetness were 0.44, 0.62, and −0.42, respectively. Therefore, in the future, it is recommended that studies with limited resources could infer total sweetness just based on the sucrose content measurement.

Table 3.4 Total sweetness (equivalent mg sucrose/g dry matter) of 10 different edamame genotypes planted in different locations.

Genotype	Whitethorne, VA	Little Rock, AR	Painter, VA	Mean (across locations)	Ranking
R13-5029	88.48 ± 16.18	68.42 ± 3.60	64.30 ± 2.41	**73.74 ± 12.94^{a}**	**1**
R14-6450	53.90 ± 0.48	60.10 ± 5.35	56.90 ± 5.59	**56.96 ± 3.10^{b}**	**10**
UA-Kirksey	88.78 ± 2.37	65.02 ± 6.94	54.77 ± 3.92	**69.52 ± 17.44^{ab}**	**2**
V10-3653	76.57 ± 24.76	57.34 ± 0.21	47.08 ± 2.60	**60.33 ± 14.97^{ab}**	**9**
V13-0329	74.50 ± 9.97	59.79 ± 5.45	60.92 ± 0.11	**65.07 ± 8.18^{ab}**	**6**
V13-0339	88.48 ± 6.96	60.11 ± 0.34	55.75 ± 1.56	**68.12 ± 17.77^{ab}**	**4**
V13-1644	84.85 ± 5.03	52.88 ± 6.37	50.35 ± 8.86	**62.69 ± 19.23^{ab}**	**7**
V15-0396	81.50 ± 4.40	62.11 ± 3.22	57.24 ± 2.64	**66.95 ± 12.83^{ab}**	**5**
V16-0523	70.27 ± 3.79	58.79 ± 0.93	54.43 ± 5.71	**61.16 ± 8.19^{ab}**	**8**
V16-0547	79.01 ± 12.80	69.92 ± 9.54	59.12 ± 4.27	**69.35 ± 9.96^{ab}**	**3**
Mean (across genotypes)	**78.63 ± 10.79^{A}**	**61.45 ± 5.13^{B}**	**56.09 ± 4.95^{B}**		

Different letters (ab & AB) indicate a significant difference based on the two-way ANOVA with Tukey's HSD test ($p < 0.05$).

Raffinose and stachyose

Raffinose and stachyose belong to the raffinose family oligosaccharides (RFOs) and they are also important free sugars in edamame (Kumar et al., 2010). RFOs cannot be digested in the human gastrointestinal tract. When passed to the lower gut, their fermentation by intestinal microflora creates flatulence-inducing gases and leads to abdominal discomfort or even diarrhea (Kumar et al., 2010). Therefore, edamame genotypes with low levels of RFOs are usually desired. In this study, the raffinose contents of edamame were at the same level as glucose, and the stachyose content was closer to alanine. The variance analysis showed that location had a significant effect on both the raffinose and stachyose contents while genotype only affected the stachyose content (Table 3.1). Most of the samples in Little Rock and Painter did not have detectable stachyose (Table 3.5). Samples planted in Whitethorne had the highest raffinose and stachyose contents (5.32 and 2.34 mg/g, respectively) among the three locations. Compared to the study of Xu et al. (2016), the genotypes in this study had similar raffinose and stachyose

contents with the genotype Asmara (4.6 and 1.1 mg/g) but higher raffinose and stachyose contents than the genotype Mooncake (1.6 and 0.7 mg/g). It is interesting to note that the raffinose and stachyose contents are positively correlated with sucrose, fructose and glucose contents (further illustrated in Figure 3.1). For example, the genotype R13-5029 had higher sucrose, fructose and glucose contents, and its raffinose and stachyose contents were also high. V15-0396 and V16-0547 also had relatively high raffinose content and detectable stachyose compared to UA-Kirksey. Therefore, if breeders want to select genotypes with high fructose and glucose contents, these genotypes will potentially have high raffinose and stachyose contents. It was reported that a person weighing 60 kg can consume as much as 38.4 g (male) or 57.6 g (female) of soybean oligosaccharides without any gastrointestinal troubles (Hata et al., 1991). Therefore, the genotypes used in this study are unlikely to cause diarrhea due to the low content of oligosaccharides. It would be meaningful to investigate which of raffinose and stachyose leads to more severe abdominal discomfort, however, this information is not readily available. Future research might be needed to provide this information.

Table 3.5 Raffinose and stachyose contents (mg/g dry matter) of 10 edamame genotypes planted in different locations.

Genotype	Whitethorne, VA	Little Rock, AR	Painter, VA	Mean (across locations)	Ranking
Raffinose					
R13-5029	9.57 ± 2.24	0.78 ± 0.60	3.12 ± 0.09	**4.49 ± 4.20**	**1**
R14-6450	5.29 ± 1.51	1.29 ± 1.44	1.93 ± 1.20	**2.84 ± 2.20**	**6**
UA-Kirksey	4.35 ± 1.02	1.23 ± 1.25	3.12 ± 1.10	**2.90 ± 1.65**	**4**
V10-3653	4.58 ± 2.57	1.01 ± 1.21	2.88 ± 0.48	**2.82 ± 2.05**	**7**
V13-0329	6.89 ± 2.94	0.52 ± 0.03	2.62 ± 0.42	**3.34 ± 3.19**	**2**
V13-0339	4.49 ± 0.53	0.95 ± 0.47	2.88 ± 1.47	**2.77 ± 1.75**	**8**
V13-1644	4.97 ± 0.36	0.11 ± 0.15	2.54 ± 0.88	**2.54 ± 2.22**	**10**
V15-0396	5.36 ± 0.30	0.80 ± 1.13	2.43 ± 0.30	**2.86 ± 2.13**	**5**
V16-0523	3.40 ± 0.16	1.99 ± 2.82	2.79 ± 1.23	**2.73 ± 1.52**	**9**
V16-0547	4.35 ± 1.66	0.68 ± 0.32	4.09 ± 2.02	**3.04 ± 2.18**	**3**
Mean (across genotypes)	**5.32 ± 2.07^{A}**	**0.94 ± 1.02^{C}**	**2.84 ± 0.96^{B}**		
Stachyose					
R13-5029	13.11 ± 1.50	0.00 ± 0.00*	0.00 ± 0.00*	**4.37 ± 6.80^{a}**	**1**
R14-6450	0.00 ± 0.00*	0.00 ± 0.00*	0.00 ± 0.00*	**0.00 ± 0.00^{*b}**	**8**
UA-Kirksey	0.00 ± 0.00*	0.00 ± 0.00*	0.32 ± 0.45	**0.11 ± 0.26^{b}**	**7**
V10-3653	1.38 ± 1.95	0.28 ± 0.40	0.48 ± 0.68	**0.71 ± 1.08^{b}**	**5**
V13-0329	3.01 ± 1.30	0.00 ± 0.00*	0.00 ± 0.00*	**1.00 ± 1.66^{b}**	**2**
V13-0339	0.00 ± 0.00	0.00 ± 0.00*	0.00 ± 0.00*	**0.00 ± 0.00^{*b}**	**8**
V13-1644	0.00 ± 0.00	0.00 ± 0.00*	0.00 ± 0.00*	**0.00 ± 0.00^{*b}**	**8**
V15-0396	1.83 ± 0.77	0.00 ± 0.00*	0.00 ± 0.00*	**0.61 ± 1.00^{b}**	**6**
V16-0523	1.75 ± 0.66	0.00 ± 0.00*	0.54 ± 0.77*	**0.76 ± 0.92^{b}**	**3**
V16-0547	2.29 ± 0.65	0.00 ± 0.00*	0.00 ± 0.00*	**0.76 ± 1.22^{b}**	**3**
Mean (across genotypes)	**2.34 ± 3.90^{A}**	**0.03 ± 0.13^{B}**	**0.13 ± 0.34^{B}**		

Different letters (ab & ABC) indicate a significant difference based on the two-way ANOVA with Tukey's HSD test ($p < 0.05$). 0.00* means the concentration cannot be detected by the HPLC but not absolutely zero.

3.3.2 Other chemical compositions

Crude protein

Protein is one of the most important constituents in edamame. The high-quality protein in edamame makes it a good alternative diet for vegans and vegetarians, thus, higher protein content is usually desired (Zeipina et al., 2017). In this study, the protein content ranged from 38.77 to 45.57% based on dry mass. This result is consistent with the study of Guo et al. (2020),

in which the protein content of edamame ranged from 36 to 45%. The protein content in edamame was significantly affected by location (Tables 3.1 and 3.6). The edamame planted in Painter and Little Rock had higher average protein content (43.11% and 42.47%, respectively) than the samples planted in Whitethorne (41.01%). The temperature could be a potential reason for protein variation. According to a previous study conducted by Wolf et al. (1982), protein content in soybean seeds increased when the day temperature was higher than 30 °C. Increased protein content under higher temperatures has also been observed in a second study (Dornbos and Mullen, 1992). In our study, there were 81 days in Little Rock and 77 days in Painter between June and October had the highest temperature over 30 °C compared to 25 days in Whitethorne, which could partially explain the high protein contents of edamame planted in Little Rock and Painter. Genotype also significantly influenced protein content. Among the 10 genotypes, UA-Kirksey had the highest mean protein content across locations (44.31%) while V13-0339 had the lowest protein content (40.26%). Regarding the top genotypes which had high sweetness, R13-5029 and V16-0527 had relatively high protein contents of 42.1 and 42.5%, respectively, while V15-0396 had a slightly lower protein content (41.0%) than R13-5029 and V16-0527. Genotype × location interaction also had a significant effect on protein content indicating that these genotypes ranked differently in protein content across the three locations.

Table 3.6 Crude protein contents (%, dry basis (d.b.)) of 10 edamame genotypes planted in different locations.

Genotype	Whitethorne, VA	Little Rock, AR	Painter, VA	Mean (across locations)	Ranking
R13-5029	42.08 ± 0.07	41.27 ± 0.24	42.92 ± 0.53	**42.09 ± 0.78^{bc}**	**6**
R14-6450	42.60 ± 1.35	43.38 ± 0.42	42.49 ± 0.08	**42.82 ± 0.77^{ab}**	**3**
UA-Kirksey	43.76 ± 0.16	44.08 ± 0.44	45.07 ± 0.03	**44.31 ± 0.65^{a}**	**1**
V10-3653	40.96 ± 0.84	43.40 ± 0.20	45.57 ± 0.78	**43.31 ± 2.13^{ab}**	**2**
V13-0329	41.04 ± 1.20	40.57 ± 0.53	43.72 ± 1.05	**41.78 ± 1.70^{bcd}**	**7**
V13-0339	38.77 ± 1.25	40.66 ± 1.29	41.34 ± 1.12	**40.26 ± 1.52^{d}**	**10**
V13-1644	39.52 ± 1.39	41.99 ± 0.92	43.46 ± 0.56	**41.66 ± 1.95^{bcd}**	**8**
V15-0396	40.95 ± 1.36	42.15 ± 0.39	39.92 ± 0.84	**41.01 ± 1.24^{cd}**	**9**
V16-0523	40.43 ± 0.03	43.55 ± 0.99	42.81 ± 1.61	**42.26 ± 1.68^{bc}**	**5**
V16-0547	40.03 ± 0.74	43.67 ± 0.47	43.80 ± 0.78	**42.50 ± 1.99^{bc}**	**4**
Mean (across genotypes)	**41.01 ± 1.61^{B}**	**42.47 ± 1.37^{A}**	**43.11 ± 1.73^{A}**		

Different letters (abcd & ABC) indicate a significant difference based on the two-way ANOVA with Tukey's HSD test ($p < 0.05$).

Oil

The oil content of edamame ranged from 17.37 to 21.61% depending on different genotypes (Table 3.7). In a previous study, Xu et al. (2016) reported that the oil contents of two edamame varieties (Asmara and Mooncake) were between 17.2 and 18.9% when they were harvested at R6 stage. Unlike soybeans with their desirable higher oil content for vegetable oil and soy-diesel, a lower oil content is usually preferred for soy food products including edamame (Zhang et al., 2017). Location also had a significant effect on oil content (Table 3.1). The highest oil content was observed on V13-0339 in Painter and the lowest oil content was found on UA-Kirksey in Little Rock. Averaged across all genotypes, the edamame planted in Painter had the highest oil content (20.33%), and V13-1644 had the highest oil content when averaged across all three locations (20.91%). For the three top genotypes based on sweetness, R13-5029 had a mean oil content of 18.76%, which is comparable to the mean oil content (18.66%) of the cultivar check UA-Kirksey. Genotypes V16-0547 and V15-0396 had mean oil contents of 19.15 and

18.99%, slightly higher than that of UA-Kirksey. The significant effects of planting environment and genotype and their interaction on the oil content of soybean seeds were observed in the studies of Kumar et al. (2006b) and Arslanoglu et al. (2011). However, in our study, no significant effect of genotype × location interaction was observed on the oil content, indicating that genotypes ranked similarly among themselves from location to location (Table 3.1).

Table 3.7 Oil contents (%, d.b.) of 10 edamame genotypes planted in different locations.

Genotype	Whitethorne, VA	Little Rock, AR	Painter, VA	Mean (across locations)	Ranking
R13-5029	18.00 ± 0.68	18.34 ± 1.03	19.95 ± 0.66	**18.76 ± 1.12^{bc}**	**9**
R14-6450	20.27 ± 1.02	19.15 ± 0.54	20.41 ± 0.23	**19.94 ± 0.81^{ab}**	**5**
UA-Kirksey	18.70 ± 0.90	17.37 ± 0.14	19.89 ± 0.52	**18.66 ± 1.22^{c}**	**10**
V10-3653	20.46 ± 0.16	20.30 ± 0.80	20.81 ± 0.83	**20.52 ± 0.57^{a}**	**4**
V13-0329	20.33 ± 0.09	21.03 ± 0.04	21.25 ± 0.47	**20.87 ± 0.48^{a}**	**2**
V13-0339	20.19 ± 0.14	20.57 ± 0.38	21.61 ± 0.72	**20.79 ± 0.76^{a}**	**3**
V13-1644	20.30 ± 0.74	20.86 ± 1.13	21.56 ± 0.82	**20.91 ± 0.91^{a}**	**1**
V15-0396	19.25 ± 0.17	18.31 ± 0.66	19.41 ± 0.64	**18.99 ± 0.67^{bc}**	**8**
V16-0523	19.39 ± 0.07	19.16 ± 1.07	19.05 ± 0.20	**19.20 ± 0.51^{bc}**	**6**
V16-0547	19.93 ± 0.07	18.13 ± 0.25	19.38 ± 0.90	**19.15 ± 0.92^{bc}**	**7**
Mean (across genotypes)	**19.68 ± 0.89^{B}**	**19.32 ± 1.36^{B}**	**20.33 ± 1.03^{A}**		

Different letters (abc & AB) indicate a significant difference based on the two-way ANOVA with Tukey's HSD test ($p < 0.05$).

Starch

Starch is the main carbohydrate in plant storage organs (Stevenson et al., 2006), but it is not well-studied in soybean or edamame because of its relatively low content and because soybean is usually considered as a protein and oil crop. Immature soybean usually has a starch content of 10–15% on a dry matter basis. The starch content of edamame in this study ranged from 11.99 to 16.58%, with the genotype V13-0339 planted in Whitethorne having the highest and V10-3653 planted in Little Rock the lowest (Table 3.8). Averaged across all genotypes, edamame planted in Whitethorne had the highest starch content (15.14%) and those planted in

Painter had the lowest (14.15%). This observation is similar to the observation that edamame planted in Whitethorne had the highest sugar (sucrose, glucose, and fructose) content and those in Painter had the lowest sugar content. Averaged across all locations, genotype V13-0339 had the highest (15.60%) and V10-3653 had the lowest (13.06%) starch content. The two top genotypes R13-5029 and V16-0547 had a lower starch content of 13.42 and 14.59% than that of the check cultivar (14.92%). V15-0396 had a higher starch content (15.17%) than the check cultivar.

Table 3.8 Starch contents (%, d.b.) of 10 edamame genotypes planted in different locations.

Genotype	Whitethorne, VA	Little Rock, AR	Painter, VA	Mean (across locations)	Ranking
R13-5029	12.14 ± 0.69	15.28 ± 2.27	12.84 ± 1.17	**13.42 ± 1.89^{bc}**	**9**
R14-6450	13.79 ± 0.28	12.92 ± 0.70	15.36 ± 0.61	**14.02 ± 1.19^{abc}**	**8**
UA-Kirksey	15.50 ± 0.17	14.96 ± 1.20	14.31 ± 0.53	**14.92 ± 0.80^{ab}**	**5**
V10-3653	15.14 ± 2.85	11.99 ± 0.20	12.07 ± 0.21	**13.06 ± 2.06^{c}**	**10**
V13-0329	16.27 ± 1.63	14.80 ± 0.55	14.08 ± 0.10	**15.05 ± 1.26^{ab}**	**3**
V13-0339	16.58 ± 0.55	14.98 ± 0.45	15.26 ± 1.18	**15.60 ± 0.98^{a}**	**1**
V13-1644	15.60 ± 0.34	14.56 ± 0.39	14.66 ± 0.77	**14.94 ± 0.66^{ab}**	**4**
V15-0396	15.54 ± 0.33	15.41 ± 0.27	14.56 ± 0.23	**15.17 ± 0.52^{ab}**	**2**
V16-0523	15.55 ± 0.26	14.33 ± 0.31	14.09 ± 0.78	**14.66 ± 0.80^{abc}**	**6**
V16-0547	15.34 ± 1.14	14.16 ± 0.46	14.28 ± 0.04	**14.59 ± 0.80^{abc}**	**7**
Mean (across genotypes)	**15.14 ± 1.51^{A}**	**14.34 ± 1.25^{B}**	**14.15 ± 1.11^{B}**		

Different letters (abc & AB) indicate a significant difference based on the two-way ANOVA with Tukey's HSD test ($p < 0.05$).

Neutral detergent fiber (NDF)

Besides protein, oil, and starch, edamame also offers a good source of fiber. The NDF content of edamame in this study ranged from 7.14 to 9.81% (Table 3.9). This range was consistent with one previous study in which the authors reported a fiber content range of 6.7–10.7% in soybean seeds (Ciabotti et al., 2016). The fiber content of soybean seeds in another study was slightly lower, between 5.53 and 8.04% (Jiang et al., 2018). The highest NDF content

was observed on genotype V13-1644 planted in Little Rock while the lowest NDF content was found on genotype V16-0547 planted in Whitethorne. Location had a significant effect on edamame NDF content (Table 3.1). Averaged across the genotypes, edamame planted in Little Rock had significantly higher NDF contents (9.00%) than samples harvested in Whitethorne (7.96%) and Painter (8.39%). The significant effect of genotype on the NDF content was not observed, probably because the genotypes chosen in this study were not genetically distant enough. In the future, more genotypes with a wide genetic distance should be included for identifying edamame with significantly higher fiber content. With the ever-increasing attention to the health benefits of fiber intake, edamame with a high fiber content is desired in the market as a health-promoting food. According to FDA regulations, a food product must contain over 10% of the Recommended Daily Intake (RDI) for fiber per serving to be claimed as a "good source" of fiber. Considering that the RDI for fiber is 25 grams per day and a typical serving size of edamame is 100 g (wet weight), edamame can be claimed as a good source of fiber if it contains at least 2.5 g fiber per 100 g of wet weight. In this study, all three top genotypes (R13-0529, V15-0396, and V16-0547) can be claimed as a "good source" of fiber according to the calculations of the fiber content on a wet basis.

Table 3.9 Neutral detergent fiber contents (%,d.b.) of 10 edamame genotypes planted in different locations.

Genotype	Whitethorne, VA	Little Rock, AR	Painter, VA	Mean (across locations)	Ranking
R13-5029	8.37 ± 0.51	9.49 ± 0.53	8.43 ± 0.48	**8.76 ± 0.68**	**2**
R14-6450	7.77 ± 0.03	9.48 ± 1.43	7.95 ± 0.34	**8.40 ± 1.06**	**6**
UA-Kirksey	8.37 ± 0.56	8.40 ± 0.78	8.09 ± 1.12	**8.29 ± 0.68**	**8**
V10-3653	8.07 ± 0.32	9.01 ± 0.13	8.91 ± 0.47	**8.66 ± 0.53**	**3**
V13-0329	8.24 ± 0.98	8.81 ± 1.36	7.91 ± 0.32	**8.32 ± 0.87**	**7**
V13-0339	7.95 ± 0.21	7.97 ± 0.55	7.90 ± 0.16	**7.94 ± 0.27**	**10**
V13-1644	7.64 ± 0.58	9.81 ± 0.56	8.53 ± 0.32	**8.66 ± 1.05**	**3**
V15-0396	8.10 ± 0.06	9.34 ± 0.47	8.95 ± 1.08	**8.80 ± 0.77**	**1**
V16-0523	7.90 ± 0.63	8.25 ± 0.37	8.51 ± 0.37	**8.22 ± 0.45**	**9**
V16-0547	7.14 ± 0.49	9.47 ± 0.02	8.71 ± 0.01	**8.44 ± 1.09**	**5**
Mean (across genotypes)	**7.96 ± 0.52^{B}**	**9.00 ± 0.82^{A}**	**8.39 ± 0.56^{B}**		

Different letters (AB) indicate a significant difference based on the two-way ANOVA with Tukey's HSD test ($p < 0.05$).

Ash

Ash has the lowest chemical composition except for oligosaccharides and alanine. Xu et al. (2016) analyzed the ash contents of edamame samples over bean development and the ash content range was between 3.58 and 5.49%. In this study, the ash content range was from 4.29 to 5.83% (Table 3.10). The highest ash content was observed on V10-3653 in Little Rock and the lowest was found on V13-0339 planted in Whitethorne. Genotype, location, and their interaction all had a significant effect on the ash content of edamame (Table 3.1). Averaged across genotypes, edamame planted in Little Rock had the highest ash content (5.60%) while edamame planted in Whitethorne had the lowest ash content (4.67%). Averaged across three different locations, V13-1644 had the highest ash content (5.39%) while V16-0547 had the lowest ash content (5.08%).

Table 3.10 Ash contents (%, d.b.) of 10 edamame genotypes planted in different locations.

Genotype	Whitethorne, VA	Little Rock, AR	Painter, VA	Mean (across locations)	Ranking
R13-5029	5.13 ± 0.03	5.63 ± 0.09	4.97 ± 0.20	**5.25 ± 0.32^{abc}**	**3**
R14-6450	4.65 ± 0.14	5.76 ± 0.34	5.35 ± 0.00	**5.25 ± 0.53^{abc}**	**3**
UA-Kirksey	4.67 ± 0.06	5.59 ± 0.15	5.48 ± 0.00	**5.24 ± 0.46^{abc}**	**5**
V10-3653	4.57 ± 0.25	5.83 ± 0.04	5.62 ± 0.13	**5.34 ± 0.62^{ab}**	**2**
V13-0329	4.73 ± 0.08	5.38 ± 0.10	4.96 ± 0.19	**5.02 ± 0.31^{cd}**	**9**
V13-0339	4.29 ± 0.12	5.55 ± 0.08	5.44 ± 0.12	**5.09 ± 0.63^{bcd}**	**7**
V13-1644	4.85 ± 0.14	5.57 ± 0.10	5.74 ± 0.14	**5.39 ± 0.44^{a}**	**1**
V15-0396	4.68 ± 0.04	5.56 ± 0.11	5.13 ± 0.03	**5.12 ± 0.40^{bcd}**	**6**
V16-0523	4.49 ± 0.06	5.53 ± 0.09	4.71 ± 0.08	**4.91 ± 0.50^{d}**	**10**
V16-0547	4.69 ± 0.02	5.65 ± 0.04	4.90 ± 0.13	**5.08 ± 0.46^{cd}**	**8**
Mean (across genotypes)	**4.67 ± 0.23^{C}**	**5.60 ± 0.16^{A}**	**5.23 ± 0.35^{B}**		

Different letters (abcd & ABC) indicate a significant difference based on the two-way ANOVA with Tukey's HSD test ($p < 0.05$).

Moisture content of fresh beans

Water takes up most of the wet weight of edamame and makes it easier to prepare and cook than soybean. In this study, the moisture content of fresh beans was in the range of 63.10–73.22% (Table 3.11). The highest moisture content was observed on V16-0547 planted in Little Rock and the lowest moisture content was found on V15-0396 planted in Whitethorne. Location had a significant effect on the moisture content of edamame. Averaged across all genotypes, edamame planted in Little Rock contained significantly more moisture than edamame planted in Whitethorne and Painter. Significant effect of genotype on the moisture content of edamame was found based on the ANOVA test. Averaged across all locations, the potential top genotypes R13-5029 and V16-0547 had higher moisture content (69.80 and 69.62%, respectively) while V15-0396 had relatively low moisture contents (66.08%). No significant effect of genotype × location interaction was found on the moisture content of edamame. Moisture content decreases gradually during bean development, thus harvest time is another potential reason leading to differences in

moisture contents (Xu et al., 2016). The optimal harvest time of edamame is when the beans fill 80–90% of the pod cavity and the pods are still green and immature. Harvesting edamame at the optimal harvest time can help ensure its optimal quality.

Table 3.11 Moisture contents (%, wet basis (w.b.)) of fresh beans of 10 edamame genotypes planted in different locations.

Genotype	Whitethorne, VA	Little Rock, AR	Painter, VA	Mean (across locations)	Ranking
R13-5029	66.63 ± 1.25	72.13 ± 1.97	70.63 ± 0.04	**69.80 ± 2.75**	**1**
R14-6450	63.29 ± 0.97	72.80 ± 7.04	65.41 ± 3.92	**67.17 ± 5.75**	**8**
UA-Kirksey	67.39 ± 0.29	69.18 ± 0.35	66.01 ± 1.09	**67.53 ± 1.52**	**5**
V10-3653	65.73 ± 1.74	68.92 ± 0.74	67.04 ± 0.89	**67.23 ± 1.71**	**7**
V13-0329	65.13 ± 0.56	67.44 ± 1.71	64.96 ± 1.33	**65.85 ± 1.59**	**10**
V13-0339	67.15 ± 3.10	68.26 ± 2.65	67.26 ± 0.70	**67.56 ± 1.93**	**4**
V13-1644	66.27 ± 1.10	72.49 ± 2.47	66.33 ± 3.33	**68.36 ± 3.73**	**3**
V15-0396	63.10 ± 0.08	69.68 ± 0.76	65.47 ± 0.43	**66.08 ± 3.00**	**9**
V16-0523	64.67 ± 0.04	69.50 ± 0.59	68.30 ± 1.32	**67.49 ± 2.34**	**6**
V16-0547	66.19 ± 1.23	73.22 ± 0.69	69.46 ± 0.79	**69.62 ± 3.23**	**2**
Mean (across genotypes)	**65.56 ± 1.76^{B}**	**70.36 ± 2.81^{A}**	**67.09 ± 2.25^{B}**		

Different letters (AB) indicate a significant difference based on the two-way ANOVA with Tukey's HSD test ($p < 0.05$).

3.3.3 Principal component analysis

In this study, 12 chemical constituents were determined among 10 edamame genotypes planted in three locations. To better understand the relationships among these chemical constituents and how location affected the chemical composition of edamame, principal component analysis (PCA) was conducted. The first three principal components (PC) were able to explain 68.90% of the total variance. The first principal component (PC1) and the second principal component (PC2) accounted for 38.28 and 18.62% of the variance, respectively. Figure 3.1 consists of the PCA score plot and component loading plot of PC1 vs PC2 and it provided better visualization of the edamame samples. PC1 mainly consisted of sucrose, glucose, raffinose, alanine, crude protein, moisture, NDF, and ash, while PC2 mainly consisted of

glucose, fructose, stachyose, protein, oil, and starch. In the component loading plot, each constituent is illustrated with a vector and the angle of any two vectors indicates the correlation between those two constituents. When the angle is exactly 90°, there is no correlation. With angles being between 0–90° and 90–180°, there is a positive or negative correlation, respectively (Bi et al., 2017). The component loading plot shows that the 12 chemical constituents were divided into three groups. The first group consisted of free sugars including sucrose, raffinose, glucose, stachyose, and fructose which indicates that free sugar contents were all positively correlated with each other. The second group contained crude protein, NDF, moisture, ash and alanine contents while starch and oil were in the last group. All chemical constituents in the second group were negatively correlated with sucrose, indicating that the samples having higher sucrose contents tend to have lower protein, NDF, moisture, ash, and alanine contents. Sucrose is positively correlated with starch but does not correlate with oil. This information is useful to breeders to balance the sweetness (sugars) and nutritional values (proteins, fibers) when breeding consumer-preferred edamame varieties.

The PCA score plot shows the distributions of edamame planted in three locations. The sample points close to each other have similar chemical compositions. The samples were divided into two clusters—one cluster mainly consisted of the samples planted in Whitethorne and the other cluster consisted of the samples planted in Painter and Little Rock. This indicates that location affected the chemical compositions of edamame. All samples planted in Whitethorne fell into the cluster where free sugars dominated, indicating that free sugars are the primary reason for separation of edamame planted in Whitethorne from the samples planted in Little Rock and Painter. The points of Little Rock's samples fell into the area which was dominated by crude protein, NDF, moisture, ash, and alanine contents. The chemical composition data showed

that edamame planted in Little Rock had significantly higher crude protein, NDF, moisture, ash, and alanine contents. The samples planted in Painter were located between the samples planted in Whitethorne and Little Rock, indicating that the samples planted in Painter had a middle level of chemical compositions. Overall, the PCA analysis showed the pronounced effect of planting locations on the chemical composition of edamame. The information could provide the breeders with the guidance about how to choose locations for selecting/breeding for cultivars with specific traits.

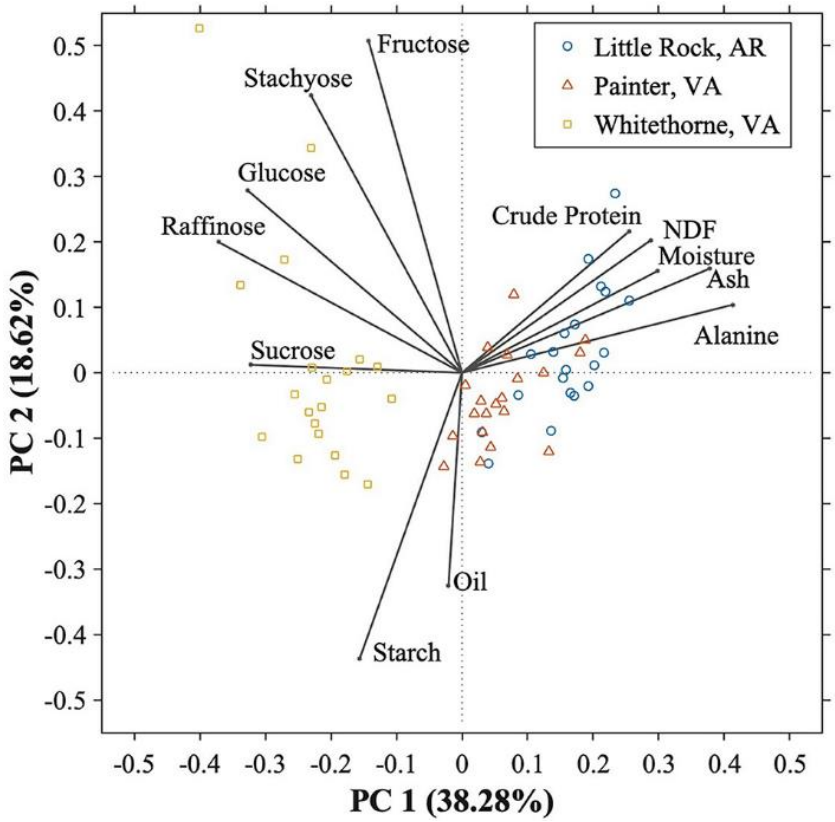

Figure 3.1 Principal component analysis (PCA) score plot and the component loading plot of PC1 vs. PC2 of the 10 different edamame genotypes planted in three different locations.

3.4 Conclusions

The effect of genotype and planting location on chemical compositions was investigated on 10 edamame genotypes planted in three locations—Whitethorne, VA; Little Rock, AR; and Painter, VA. The determined chemical constituents included soluble sugars (sucrose, glucose, fructose, raffinose, and stachyose), alanine, crude protein, oil, starch, moisture, neutral detergent fiber, and ash. The main findings are listed as follows:

a) Planting location had significant effects on all chemical constituents of edamame ($p < 0.05$). The edamame planted in Whitethorne had the highest soluble sugars, total sweetness, and starch contents while the edamame planted in Little Rock had the highest soluble alanine, NDF, moisture and ash contents. The edamame planted in Painter had the highest crude protein and oil contents.

b) Genotype had significant effects on all chemical constituents except for NDF and raffinose. Among the 10 genotypes, R13-5029 had the highest sucrose content, sweetness, and moisture contents while UA-Kirksey had the highest crude protein content. The highest oil content was observed on V13-1644 and the highest starch content was found on V13-0339.

c) R13-5029 consistently had high total sweetness across the three locations, meanwhile it had relatively high protein and fiber contents but low oil content. Therefore, it is identified as a potential genotype for future edamame production in the US.

d) The significant effect of genotype × location interaction was only observed on sucrose, stachyose, protein, and starch contents.

e) PCA analysis showed that there is a negative correlation between sucrose content and protein and fiber contents in edamame.

f) To breed better edamame genotypes for the US market, both genotype and location should be taken into consideration.

3.5 References

Agilent Technologies, Inc. (2017). Determination of amino acid composition of cell culture media and protein hydrosylate standard.

AOAC (2005a). Official methods of analysis of AOAC INTERNATIONAL (2005) 18th Ed., AOAC INTERNATIONAL, Gaithersburg, MD, USA, Official Method 2001.11. Available at: https://www.scribd.com/document/184788392/AOAC-2001-11-PROTEINA [Accessed March 21, 2020].

AOAC (2005b). Official methods of analysis of AOAC INTERNATIONAL (2005) 18th Ed., AOAC INTERNATIONAL, Gaithersburg, MD, USA, Official Method 2003.05. Available at: https://www.scribd.com/doc/128237003/Aoac-920-39-2003-05-Extracto-Etereo-Dietileter [Accessed March 21, 2020].

AOAC (2005c). Official methods of analysis of AOAC INTERNATIONAL (2005) 18th Ed., AOAC INTERNATIONAL, Gaithersburg, MD, USA, Official Method 942.05. Available at: https://www.scribd.com/doc/53398720/Aoac942-05-Ash [Accessed March 21, 2020].

Arslanoglu, F., Aytac, S. and Oner, E.K., 2011. Effect of genotype and environment interaction on oil and protein content of soybean (Glycine max (L.) Merrill) seed. African Journal of Biotechnology, 10(80), pp.18409-18417.

Berganza, B.E., Moran, A.W., Rodríguez, G.M., Coto, N.M., Santamaría, M. and Bressani, R., 2003. Effect of variety and location on the total fat, fatty acids and squalene content of amaranth. Plant Foods for Human Nutrition, 58(3), pp.1-6.

Bi, Y., Zhang, Y., Jiang, H., Hong, Y., Gu, Z., Cheng, L., Li, Z. and Li, C., 2017. Molecular structure and digestibility of banana flour and starch. Food Hydrocolloids, 72, pp.219-227.

Brady, J. W. (2013). Introductory food chemistry. Comstock Pub. Associates.

Cameron, A.T., 1945. The relative sweetness of various sweet compounds and of their mixtures. Canadian Journal of Research, 23(5), pp.139-166.

Carneiro, R., Duncan, S., O'Keefe, S., Yu, D., Huang, H., Yin, Y., Neill, C., Zhang, B., Kuhar, T., Rideout, S. and Reiter, M., 2021. Utilizing consumer perception of edamame to guide new variety development. Frontiers in Sustainable Food Systems, 4, p.556580.

Carson, L. C., Freeman, J. H., Harris, J. R., Welbaum, G. E., and Carson, L. C., 2010. Cultivation and nutritional constituents of Virginia grown edamame by cultivation and nutritional constituents of Virginia grown edamame. (Doctoral dissertation, Virginia Tech).

Cartter, J.L. and Hopper, T.H., 1942. Influence of variety, environment, and fertility level on the chemical composition of soybean seed (No. 1488-2016-123875). US Department of Agriculture.

Ciabotti, S., Silva, A.C.B.B., Juhasz, A.C.P., Mendonça, C.D., Tavano, O.L., Mandarino, J.M.G. and ConÇAlves, C.A.A., 2016. Chemical composition, protein profile, and isoflavones content in soybean genotypes with different seed coat colors.

Chen, P., Orazaly, M., Bacon, R., Florez-Palacios, L., Moseley, D., Lancaster, S., Hedge, J., McCoy, J. and Hayes, S., 2017. Breeding new soybean cultivars with high yield and disease resistance. Soybean Research Studies 2016, p.20.

Clark, J., 2020. Managing soil and soil fertility after flooding. SDSU Extension. Available at: https://extension.sdstate.edu/managing-soil-and-soil-fertility-after-flooding. [Accessed on March 21, 2020].

Dornbos, D.L. and Mullen, R.E., 1992. Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. Journal of the American Oil Chemists Society, 69(3), pp.228-231.

Eheart, J.F., Young, R.W. and Allison, A.H., 1955. Variety, type, year, and location effects on the chemical composition of peanuts. Food Research, 20, pp.497-505.

Guo, J., Rahman, A., Mulvaney, M.J., Hossain, M.M., Basso, K., Fethiere, R. and Babar, M.A., 2020. Evaluation of edamame genotypes suitable for growing in Florida. Agronomy Journal, 112(2), pp.693-707.

HATA, Y., YAMAMOTO, M. and NAKAJIMA, K., 1991. Effects of soybean oligosaccharides on human digestive organs estimation of fifty percent effective dose and maximum non-effective dose based on diarrhea. Journal of Clinical Biochemistry and Nutrition, 10(2), pp.135-144.

He, Y., Kuhn, D.D., Ogejo, J.A., O'Keefe, S.F., Fraguas, C.F., Wiersema, B.D., Jin, Q., Yu, D. and Huang, H., 2019. Wet fractionation process to produce high protein and high fiber products from brewer's spent grain. Food and Bioproducts Processing, 117, pp.266-274.

Jiang, G.L., Katuuramu, D.N., Xu, Y., Ren, S. and Rutto, L.K., 2020. Analysis and comparison of seed protein, oil, and sugars in edamame dried using two oven‐drying methods and mature soybeans. Journal of the Science of Food and Agriculture, 100(10), pp.3987-3994.

Jiang, G.L., Chen, P., Zhang, J., Florez-Palacios, L., Zeng, A., Wang, X., Bowen, R.A., Miller, A. and Berry, H., 2018. Genetic analysis of sugar composition and its relationship with protein, oil, and fiber in soybean. Crop Science, 58(6), pp.2413-2421.

Johnson, D., Wang, S., & Suzuki, A., 1999. Edamame: A vegetable soybean for Colorado. Perspectives on New Crops and New Uses, 385–387.

Kaiser, C., and Ernst, M., 2013. Edamame. Center for Crop Diversification Crop Profile, University of Kentucky Cooperative Extension Service. 1-4.

Kirimura, J., Shimizu, A., Kimizuka, A., Ninomiya, T. and Katsuya, N., 1969. Contribution of peptides and amino acids to the taste of foods. Journal of Agricultural and Food Chemistry, 17(4), pp.689-695.

Konovsky, J., Lumpkin, T. A., and Mcclary, D., 1994. Edamame: The vegetable soybean. Understand. Japan. Food Agrimarket 1988, 173–181.

Kumar, V., Rani, A., Billore, S.D. and Chauhan, G.S., 2006. Physico-chemical properties of immature pods of Japanese soybean cultivars. International Journal of Food Properties, 9(1), pp.51-59.

Kumar, V., Rani, A., Solanki, S. and Hussain, S.M., 2006. Influence of growing environment on the biochemical composition and physical characteristics of soybean seed. Journal of Food Composition and Analysis, 19(2-3), pp.188-195.

Kumar, V., Rani, A., Solanki, S. and Hussain, S.M., 2006b. Influence of growing environment on the biochemical composition and physical characteristics of soybean seed. Journal of Food Composition and Analysis, 19(2-3), pp.188-195.

Lara, L.M., Wilson, S.A., Chen, P. and Atungulu, G.G., 2019. The effects of infrared treatment on physicochemical characteristics of vegetable soybean. Heliyon, 5(1), p.e01148.

Lin, T., O'Keefe, S., Duncan, S. and Fernández-Fraguas, C., 2020. Manipulation of the dry bean (*Phaseolus vulgaris* L.) matrix by hydrothermal and high-pressure treatments: Impact on in vitro bile salt-binding ability. Food Chemistry, 310, p.125699.

Machado, S., Costa, A.S., Pimentel, F., Oliveira, M.B.P. and Alves, R.C., 2020. A study on the protein fraction of coffee silverskin: Protein/non-protein nitrogen and free and total amino acid profiles. Food Chemistry, 326, p.126940.

McClure, T., Cocuron, J.C., Osmark, V., McHale, L.K. and Alonso, A.P., 2017. Impact of environment on the biomass composition of soybean (*Glycine max*) seeds. Journal of Agricultural and Food Chemistry, 65(32), pp.6753-6761.

McEnroe, P. and Coulter, B., 1964. Effect of soil pH on sugar content and yield of sugar beet. Irish Journal of Agricultural Research, pp.63-69.

Monteiro, S.T., Minekawa, Y., Kosugi, Y., Akazawa, T. and Oda, K., 2007. Prediction of sweetness and amino acid content in soybean crops from hyperspectral imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 62(1), pp.2-12.

OHair, J., Jin, Q., Yu, D., Wu, J., Wang, H., Zhou, S. and Huang, H., 2021. Non-sterile fermentation of food waste using thermophilic and alkaliphilic *Bacillus licheniformis* YNP5-TSU for 2, 3-butanediol production. Waste Management, 120, pp.248-256.

Okunlola, G.O., Akinwale, R.O. and Adelusi, A.A., 2016. Proline and soluble sugars accumulation in three pepper species (*Capsicum* spp) in response to water stress imposed at different stages of growth. Sciences in Cold and Arid Regions, 8(3), pp.205-211.

Pao, S., Ettinger, M.R., Khalid, M.F., Mebrahtu, T. and Mullins, C., 2008. Microbiological quality of frozen “edamame” (vegetable soybean). Journal of Food Safety, 28(2), pp.300-313.

Sakla, A.B., Ghali, Y., El-Farra, A. and Rizk, L.F., 1988. The effect of environmental conditions on the chemical composition of soybean seeds: Relationship between the protein, oil, carbohydrate and trypsin inhibitor content. Food Chemistry, 29(3), pp.221-231.

Saldivar, X.E.A., Wang, Y.J., Chen, P. and Mauromoustakos, A., 2010. Effects of blanching and storage conditions on soluble sugar contents in vegetable soybean. LWT-Food Science and Technology, 43(9), pp.1368-1372.

Sirtori, C.R., 2001. Risks and benefits of soy phytoestrogens in cardiovascular diseases, cancer, climacteric symptoms and osteoporosis. Drug Safety, 24(9), pp.665-682.

Song, J.Y., An, G.H. and Kim, C.J., 2003. Color, texture, nutrient contents, and sensory values of vegetable soybeans [*Glycine max* (L.) Merrill] as affected by blanching. Food Chemistry, 83(1), pp.69-74.

Song, J., Liu, C., Li, D. and Gu, Z., 2013. Evaluation of sugar, free amino acid, and organic acid compositions of different varieties of vegetable soybean (*Glycine max* [L.] Merr). Industrial Crops and Products, 50, pp.743-749.

Stevenson, D.G., Doorenbos, R.K., Jane, J.L. and Inglett, G.E., 2006. Structures and functional properties of starch from seeds of three soybean (*Glycine max* (L.) Merr.) varieties. Starch-Stärke, 58(10), pp.509-519.

Vidal, B.C., Rausch, K.D., Tumbleson, M.E. and Singh, V., 2009. Determining corn germ and pericarp residual starch by acid hydrolysis. Cereal Chem, 86(2), pp.133-135.

Wolf, R.B., Cavins, J.F., Kleiman, R. and Black, L.T., 1982. Effect of temperature on soybean seed constituents: oil, protein, moisture, fatty acids, amino acids and sugars. Journal of the American Oil Chemists' Society, 59(5), pp.230-232.

Xu, L., Jiang, F., Wu, Z., Wang, J. and Zhou, X., 2009. Effects of substrate water content and light intensity on growth and quality of non-heading Chinese cabbage. Jiangsu Journal of Agricultural Sciences, 25(4), pp.865-870.

Xu, Y., Cartier, A., Kibet, D., Jordan, K., Hakala, I., Davis, S., Sismour, E., Kering, M. and Rutto, L., 2016. Physical and nutritional properties of edamame seeds as influenced by stage of development. Journal of Food Measurement and Characterization, 10(2), pp.193-200.

Xu, Y., Sismour, E., Pao, S., Rutto, L., Grizzard, C. and Ren, S., 2012. Textural and microbiological qualities of vegetable soybean (edamame) affected by blanching and storage conditions. Journal of Food Processing & Technology, 3(6), pp.1-6.

Yang, S.Y., Li, J.H. and Zhu, L.L., 2009. Effects of light intensity on quality of pepper fruits [J]. Northern Horticulture, 2, 65-67.

Yu, D., Sun, Y., Wang, W., O'Keefe, S.F., Neilson, A.P., Feng, H., Wang, Z. and Huang, H., 2020. Recovery of protein hydrolysates from brewer's spent grain using enzyme and ultrasonication. International Journal of Food Science & Technology, 55(1), pp.357-368.

Yu, X., Yuan, F., Fu, X. and Zhu, D., 2016. Profiling and relationship of water-soluble sugar and protein compositions in soybean seeds. Food Chemistry, 196, pp.776-782.

Zeipiņa, S., Alsiņa, I. and Lepse, L., 2017. Insight in edamame yield and quality parameters: a review. Research for Rural Development, 2, pp.40-45.

Zhang, Q., Li, Y., Chin, K.L. and Qi, Y., 2017. Vegetable soybean: Seed composition and production research. Italian Journal of Agronomy, 12(3).

Zhao-Hui, L.I.U., Jiang, L.H., Xiao-Lin, L.I., Härdter, R., Zhang, W.J., Zhang, Y.L. and Zheng, D.F., 2008. Effect of N and K fertilizers on yield and quality of greenhouse vegetable crops. Pedosphere, 18(4), pp.496-502.

Chapter 4

Physical and Chemical Properties of Edamame during Bean Development and Application of Spectroscopy-Based Machine Learning Methods to Predict Optimal Harvest Time

4.1 Introduction

Edamame, also called vegetable soybean [*Glycine max* (L.) Merr.], has been widely consumed in East Asia for centuries and is a trending soy product in the U.S. It is highly nutritious due to the content of high-quality protein with isoflavones, vitamins (C and E), monounsaturated fatty acids, minerals, and dietary fiber (Johnson et al., 1999; Mentreddy et al., 2002; Zeipiņa et al., 2017). The quality of edamame is mainly determined by morphological quality (pod size, color, and weight), eating quality, and nutrition (Zeipiņa et al., 2017) and these quality parameters change over bean development.

Harvesting edamame at an appropriate time ensures peak morphological and eating quality, which offers the edamame high marketability and consumer acceptability (Konovsky et al., 2020; Zeipiņa et al., 2017). Moreover, edamame with consistent quality also eases post-harvest processing. Edamame should ideally be harvested sometime between R6 and R7 growth stages, just before the pods begin to turn yellow and when moisture and bean weight approach their maximum levels (Moseley et al., 2020; Yu et al., 2021). Given the dynamic nature between the R6 and R7 stages of bean growth, harvesting edamame outside of the optimal harvest window can potentially jeopardize its marketability. For example, harvesting too early can lead to reduced yield, sweetness, and size of beans, while harvesting too late leads to fibrous and

yellow beans (Carson, 2010). Further complicating edamame harvest is the narrow harvest window (about one week) for growers once they reach their optimal harvest time (Carson et al., 2011). A plant at the R6 growth stage can be distinguished by pods with beans that have completely filled the pod cavity, while plants at R7 can be distinguished by at least one pod having reached mature pod color (Licht, 2014). Accompanying the transition from R6 to R7 is a series of physiological changes that signal the culmination of reproductive growth and the subsequent initiation of senescence towards full maturity (R8). These changes include growing beans occupying approximately 85–90% of pod space, the color of leaves, pods, and beans changing from green to yellow, and sugars and other chemical constituents accumulating in the immature beans.

Studies have been conducted to investigate the physical properties and chemical compositions of edamame or soybean at different bean growth stages. Xu et al. (2016) studied the effects of edamame development on the physical, chemical, and anti-nutritional properties of edamame beans. Saldivar et al. (2011) investigated chemical composition changes including protein, oil, starch, and soluble saccharides, and the seed length changes during soybean development from R1 to R8. Yazdi-Samadi et al. (1977) investigated the oil, protein, sugars, starch, organic acids, and amino acid changes in developing soybean seeds. Lowell and Kuo (1989) studied the metabolism and accumulation of oligosaccharides during the development of soybeans. All these studies provided a clear picture of how the physical and chemical attributes of edamame beans or soybean seeds change throughout reproductive development. However, exploiting these differences to predict optimal harvest time was rarely studied on edamame.

Current methods for determining optimal harvest time of edamame rely on the ability of experienced edamame growers to detect these changes visually, by touch, or by taste. These

determination methods can be quite subjective; they can pose a major obstacle for relatively inexperienced or new edamame growers and cause significant economic losses due to the reduced quality of edamame harvested outside of the optimal window. Therefore, more rapid, consistent, and standardized methods for determining optimal harvest time are desired. Recently, spectroscopic techniques have been used to determine the optimal harvest time of strawberries (Shen et al., 2018; Gao et al., 2020), cherry tomatoes (Yang, 2011), and apples (Bertone et al., 2012). However, to our best knowledge, no literature has reported the application of spectroscopic methods to identify the optimal harvest time of edamame. Moreover, using a handheld spectroscopy instrument is promising because it offers a fast way (usually a few seconds) for in-field determining the optimal harvest time of edamame compared to the lengthy chemical analysis. In addition, the spectroscopic analysis can identify small changes of pod color over edamame development, which usually cannot be captured by the naked eye. However, the spectroscopy-based analysis is a secondary method requiring calibration against a reference method for identifying the optimal harvest time of edamame. Calibration is usually conducted using multivariate regression analysis; nevertheless, it sometimes cannot deliver satisfactory results due to the complexity of the spectra (Cortés, et al., 2019). Fortunately, the recent developments of machine learning techniques provide an opportunity to analyze the complex spectroscopic dataset and provide accurate and reliable calibration (Singh et al., 2016; Singh et al., 2018). Random Forest (RF) is an ensemble learning technique and it has received increasing attention due to the excellent classification results and the speed of processing (Belgiu & Drăguţ, 2016). RF has been widely applied to classify different types of food using multispectral and hyperspectral data. For example, RF was successfully applied to classify the adulterated and authentic nutmeg using infrared spectroscopy and the RF presented superior performance than

other classification methods including Partial Least-Squares Discriminant Analysis (PLS-DA) and Soft independent modeling of class analogy (SIMCA) (de Santana et al., 2019). In another study, Peidad et al. (2018) applied three widely-used machine learning methods (i.e., artificial neural network, RF, support vector machines) to classify bananas into different categories based on their measured qualities. The results demonstrated that RF achieved the highest classification accuracy among the three machine learning methods.

The objective of this study is to investigate the changes in the physical and chemical properties of edamame during bean development and apply the spectroscopy-based machine learning technique to determine the appropriate harvest time. In this study, physical properties and chemical compositions were quantified for the edamame harvested from R5 to R7 stages. Physical properties included pod weight, 20-bean weight, pod dimensions (width, length, and thickness), and color. Chemical compositions included soluble sugars and free amino acids (sucrose, fructose, glucose, alanine, and glycine), oligosaccharides (raffinose and stachyose), the moisture of fresh beans, protein, starch, fat, neutral detergent fiber (NDF), and ash. The quantified physical and chemical properties of harvested beans were used to determine the optimal stage for edamame harvesting as well as identifying the beans that were harvested 'too early' and 'too late'. Meanwhile, the spectral reflectance between 360 and 740 nm was measured on the harvested edamame pods using a handheld, portable spectrophotometer. Using the measured spectra, a machine learning approach was used to determine the readiness of edamame harvesting based on the collected spectral reflectance. This study is significant and novel because it is the first to develop a spectroscopy-based technique for rapidly and accurately identifying the optimum harvest time of edamame, which is essential to produce consistent and high-quality edamame for the market.

4.2 Materials and methods

4.2.1 Plant materials

Three edamame genotypes (R15-10280, V16-0547, UA-Kirksey) were planted on May 22nd, 2019 at Kentland farm, Whitethorne VA. Plots were arranged in a randomized complete block design (RCBD) and each plot consisted of 7 m long rows with 76 cm spacing between rows. The planting rate is about 18 seeds per meter, so in total 126 seeds per row. With an average emergence rate of 75%, there are about 95 plants per row. In total, three replications were used for the study. Once flowering began, three dates were selected to tag individual nodes along the first nine feet of each entry. In each replication, between 25 and 30 nodes were tagged, depending on the flower availability of the individual genotype. Because all genotypes were planted in the same field on the same date, they have had the same growing environment and all of them are maturity group V; therefore, the duration stages of all the genotypes are similar. In addition, the dates of tagging flowers of the three genotypes were only two days apart. The flowering tag dates were as follows, 4 August 2019, 6 August 2019, and 8 August 2019, respectively for the genotype R15-10280, V16-0547, UA-Kirksey. The nodes were cleaned of all younger flowers when necessary to allow for easier harvest and to ensure that the flowering date of each node was known. Pods were hand-harvested at six different time points. The six harvests were made on 17 September, 22 September, 28 September, 4 October, 10 October, and 11 October in 2019 respectively, corresponding to R5-1, R5-2, R6-1, R6-2, R7-1, and R7-2 of the growth stages. It is worth noting that there was only a one-day difference between R7-1 and R7-2 growth stages, as the morphological and chemical properties of pods and beans changed quickly due to the dry weather and a drought field. Details of the soybean planting and harvesting time were listed in Table A1 in the Appendix. During each of the six harvest times, 10

pods were selected from each genotype in all three replications totaling 90 pods for each harvest date. In total, 540 pods were collected in six harvests. The pods were brushed off during collection to remove any dirt and debris on the pods, but they were not extensively cleaned. This was done to preserve as much of the intact pubescence as possible and not alter the pod color due to possible injury. Once harvested, they were put in Ziplock bags, placed in a cooler containing an ice pack, and transported to the laboratory for measuring physical properties.

4.2.2 Physical properties of edamame pods and beans

In the laboratory, every pod was measured using a digital fractional caliper (Husky Tools) for length, width, and thickness. Width and thickness were taken at the thickest and widest points on each pod and length was measured from top to bottom. Pod weight was also measured using an analytical balance (Mettler-Toledo, LLC, Columbus, OH, USA). Afterward, randomly selected pods from each genotype and replication were opened until 20 beans were available and a 20-bean weight was recorded. The CIE L*, a*, and b* values were recorded to compare the lightness (+)/darkness (−), redness (+)/greenness (−), and yellowness (+)/blueness (−) of samples using a handheld, portable Konica Minolta CM-700d Spectrophotometer (Konica Minolta Sensing Americas, Inc, NJ, USA). Once all data were collected, all pods were put back in individually labeled Ziploc bags and placed in the −80 °C ultra-low freezer until further chemical composition analysis.

4.2.3 Chemical compositions of edamame beans

4.2.3.1 Free sugars, glycine, and alanine

Edamame beans were hand-shelled out of the pods and freeze-dried. Subsequently, the dry beans were milled by a grinder to pass through a 500 µm sieve (IKA®-Werke GmbH & Co. KG, Germany) and the powder was used for chemical composition analysis. Sucrose, fructose,

glucose, glycine, alanine, and oligosaccharides (raffinose and stachyose) were extracted by the method described by Yu et al., 2016, Poe et al., 2020 and Machado et al. (2020) with minor modifications. Briefly, 0.15 g of dry samples were measured and mixed with 1.5 mL of deionized water (DI water) in a 2 mL centrifuge tube. The mixture was shaken at room temperature for 2 hrs and then centrifuged at 13,500 × g for 10 min. Afterward, 750 µL of supernatant was collected and mixed with 750 µL of acetonitrile. The mixture was shaken for 10 min at room temperature followed by centrifugation for another 10 min. Finally, 750 µL of supernatant was filtered through a 0.20 µm membrane filter for sugar analysis using high-performance liquid chromatography (HPLC) equipped with a refractive index detector (RID) (Agilent Technologies, Santa Clara, CA, USA). A Luna Omega 3 µm SUGAR column (150 × 4.6 mm, Phenomenex, Torrance, CA, USA) was used for sugar separation at 40 °C with acetonitrile/water (75:25 v: v) as the mobile phase. The flow rate was set at 1 mL/min and the injection volume was 5 µL.

Glycine and alanine are two amino acids that contribute to sweetness (Birch and Kemp, 1989; Schiffman et al., 1981; Solms et al., 1965). They were determined by HPLC with a 1200 diode array detector (DAD) at $\lambda = 338$ nm based on the application method from Agilent. Online derivatization of glycine and alanine were conducted using o-Phthalaldehyde (OPA) to form highly fluorescent products that could be detected by DAD. After derivatization, 1 µL of the sample was injected into an Agilent AdvanceBio AAA C18 column at 40 °C. Mobile phase A contained 10 mM Na2HPO4, and 10 mM $Na_2B_4O_7$ with a pH of 8.2 while mobile phase B was the mixture of acetonitrile, methanol, and water (45:45:10, v: v: v). The flow rate was set at 1.5 mL/min with a gradient program for separation.

4.2.3.2 Total sweetness

Different free sugars and amino acids have different sweetness intensities. To calculate the total sweetness, sucrose was used as a reference and the sweetness of other sugars and compounds could be described by relative sweetness (RS) to sucrose which is listed in Table A2. Based on the RS, the total sweetness was calculated by the following equation:

$$Sweetness = C_{sucrose} \times RS_{sucrose} + C_{glucose} \times RS_{glucose} + C_{fructose} \times C_{fructose} + C_{glycine} \times RS_{glycine} + C_{alanine} \times RS_{alanine} \quad (4.1)$$

where C was the concentration of different compounds (mg/g), RS was their relative sweetness to sucrose. The RS used was the average of the highest RS and the lowest RS.

4.2.3.3 Determination of moisture, protein, fat, neutral detergent fiber (NDF), starch, and ash

The moisture content of fresh beans was determined by oven drying at 105 °C until the weight was constant (Yu et al., 2021). The protein content was determined by the Kjeldahl method to get the total nitrogen content, followed by multiplying by a protein conversion factor of 6.25 (AOAC, 2001.11). The fat content was extracted by petroleum ether and determined by AOAC 2003.05. The NDF content was determined by the ANKOM fiber analyzer (ANKOM Technology, Macedon, NY, USA). The non-fiber part was washed out by a neutral detergent solution and the NDF content was calculated based on the dry weight after digestion (OHair et al., 2021). The ash content was determined at 550 °C for 12 hrs in a muffle furnace according to the AOAC 942.05 in which the weight difference before and after incineration was calculated (Yu et al., 2020). Starch content was determined using the method described by Vidal et al. (2009) by measuring hydrolyzed glucose using the HPLC with RID using Bio-Rad Aminex HPX-87H (Bio-Rad Laboratories, Hercules, CA, USA).

4.2.4 Spectra-based machine learning approach for predicting harvest time

4.2.4.1 Spectral reflectance measurement on edamame pods

Before the weight of 20 beans was measured, ten pods of each edamame sample were also measured for spectral reflectance between 360 and 740 nm using the portable Konica Minolta CM-700d spectrophotometer (Konica Minolta Sensing Americas, Inc, NJ, USA) with a measurement area of 8 mm, equipped with a pulsed xenon lamp and a diffuse illumination/8° viewing system. Three images were taken of each pod and the spectral reflectance data were averaged using the Konica Minolta software.

4.2.4.2 Preprocessing of the data

Based on the physical and chemical data obtained, the condition of the edamame samples was evaluated, and all 54 samples were classified into three categories and labeled by “early class”, “ready class” and “late class”. To match with the spectral reflectance dataset (n = 10 for each sample), the same class label was assigned to each of the 10 pods for every sample. Therefore, the dataset consists of a 540 spectral reflectance of edamame pods. Among these data sets, 220 observations were from the late class, 180 were from the early class, and 140 were from the ready class. Both the raw spectral data (primary spectral) and the first-order derivatives (FOD) of each spectral data were calculated and used as a feature matrix for training and testing the RF classifier. FOD transformations of the spectral curve are a commonly applied technique used to increase classification quality by enhancing spectral features and minimizing random noise.

4.2.4.3 Random forest classification

The RF analysis method was adapted from Heim et al. (2018) using the R programming language and VSURF and caret packages. The dataset was split 80:20 into training and test data

subsets, and 10-fold repeated cross-validation was applied to the training data. This process was repeated 100 times and the mean accuracy over these repetitions was calculated. RF classifiers were trained to assign each spectrum to one of the three classes (i.e. early, ready and late) or two classes (e.g. early vs late) and the model performance as measured by classifier accuracy were compared based on cross-validation results. Accuracy was calculated as the following equation:

$$Accuracy\ (\%) = \frac{True\ Positive+True\ Negative}{True\ Positive+False\ Positive+True\ Negative+Fals\ Negative} \times 100 \qquad (4.2)$$

All three classes were classified based on 39 predictor variables (waveband at 10 nm resolution). The final model parameters were tuned to mtry = 30 and n-tree = 2000 after the best classifier was identified using the training data. Feature selection was performed using the VSURF package, and selected spectral bands were used to determine the prediction accuracy.

4.2.5 Statistical analysis

All measurements were performed on biological triplicates and results were presented as means ± standard deviation (n = 3). One-way ANOVA was performed to observe the significant effect of harvest time on physical and chemical properties of edamame beans and pods, followed by Tukey's Honestly Significant Different (HSD) to compare differences among groups using SPSS (22.0.0.0, IBM Corporation, Armonk, NY, USA). Principle component analysis (PCA) with trajectory analysis was also conducted by SPSS.

4.3 Results and discussion

4.3.1 Physical properties of edamame beans and pods

Physical properties, including pod weight, 20-bean weight, pod width, length, thickness, and color, were recorded during edamame development from R5 to R7 (Table 4.1). To harvest edamame of high quality, immature pods with fully developed green beans are preferred and the beans should not start to dry out at the harvest time (Born, 2006). For all three genotypes, pod

and 20-bean weight increased from R5 to R6 significantly and peaked at stage R6-1 or R6-2. A slight but not significant decrease was found on genotype V16-0547 and UA-Kirksey through stage R7 while a slight increase was found from R7-1 to R7-2 for genotype R15-10280. This increase was due to the random selection of edamame pods of larger size which was shown by the results of the pod width and pod length. Based on the soybean growth and development, seed filling begins at stage R5 and the dry mass accumulates at the same time (Purcell et al., 2014). When the R5 ends and R6 begins, the rate of seed growth and dry mass accumulation slows down and finally reaches the maximum value. Once they entered stage R7, the beans started to dry out and lead to a slight but not significant decrease in the pod and bean weights (Wang et al., 2006). Changes in pod width and length were not as significant as those in weight because the pods already reached the maximal size and had the fixed pod width and pod length at the end of R4. However, the pod thickness increased significantly from R5 to R6 due to the enlargement of the seeds. After the seed filling ended, changes in pod thickness over R7 were not significant.

Table 4.1 Measured physical properties of edamame in different harvest stages.

Genotype	Harvest Stage	Pod Weight/g	20-bean weight/g	Pod Width/mm	Pod Length/mm	Pod Thickness/mm
R15-10280	R5-1	0.87 ± 0.12^{c}	3.73 ± 1.07^{c}	10.10 ± 0.32^{b}	41.69 ± 0.36^{b}	5.22 ± 0.57^{b}
	R5-2	1.37 ± 0.24^{bc}	6.99 ± 1.35^{b}	10.89 ± 0.83^{ab}	46.05 ± 3.16^{ab}	6.58 ± 0.40^{a}
	R6-1	1.94 ± 0.34^{a}	10.26 ± 1.73^{a}	11.53 ± 0.57^{ab}	48.88 ± 3.17^{a}	7.39 ± 0.50^{a}
	R6-2	1.82 ± 0.15^{ab}	10.62 ± 0.50^{a}	11.72 ± 0.21^{a}	45.18 ± 0.65^{ab}	7.69 ± 0.40^{a}
	R7-1	1.57 ± 0.05^{ab}	8.04 ± 0.72^{ab}	11.41 ± 0.77^{ab}	45.32 ± 1.85^{ab}	7.60 ± 0.20^{a}
	R7-2	1.84 ± 0.20^{ab}	9.94 ± 1.05^{ab}	11.51 ± 0.33^{ab}	47.30 ± 2.86^{ab}	7.63 ± 0.48^{a}
V16-0547	R5-1	0.94 ± 0.30^{c}	3.73 ± 1.64^{c}	11.59 ± 0.73^{a}	42.21 ± 2.03^{a}	5.78 ± 1.12^{c}
	R5-2	1.29 ± 0.13^{bc}	6.17 ± 0.91^{bc}	12.39 ± 0.17^{a}	42.90 ± 2.44^{a}	6.80 ± 0.39^{bc}
	R6-1	1.95 ± 0.34^{ab}	11.66 ± 2.13^{a}	12.03 ± 0.48^{a}	43.40 ± 2.48^{a}	8.13 ± 0.55^{ab}
	R6-2	2.13 ± 0.29^{a}	12.59 ± 1.34^{a}	12.11 ± 0.40^{a}	43.79 ± 2.35^{a}	8.71 ± 0.49^{a}
	R7-1	1.80 ± 0.21^{ab}	10.61 ± 1.33^{a}	11.77 ± 0.96^{a}	41.56 ± 4.59^{a}	8.45 ± 0.26^{a}
	R7-2	1.67 ± 0.22^{ab}	10.01 ± 1.46^{ab}	11.20 ± 0.40^{a}	39.93 ± 4.42^{a}	8.22 ± 0.36^{ab}
UA-Kirksey	R5-1	0.84 ± 0.14^{c}	3.32 ± 0.72^{b}	10.78 ± 0.32^{ab}	40.69 ± 1.66^{a}	4.88 ± 0.32^{b}
	R5-2	1.10 ± 0.16^{bc}	5.42 ± 1.20^{b}	10.96 ± 0.14^{ab}	42.79 ± 1.30^{a}	5.59 ± 0.18^{b}
	R6-1	1.74 ± 0.26^{a}	9.78 ± 1.21^{a}	11.24 ± 0.55^{ab}	44.99 ± 3.69^{a}	7.20 ± 0.18^{a}
	R6-2	1.62 ± 0.26^{a}	8.88 ± 0.74^{a}	12.46 ± 1.28^{a}	44.32 ± 1.36^{a}	7.36 ± 0.44^{a}
	R7-1	1.53 ± 0.16^{ab}	8.59 ± 0.60^{a}	10.65 ± 0.58^{b}	43.46 ± 2.22^{a}	7.43 ± 0.54^{a}
	R7-2	1.43 ± 0.02^{ab}	8.24 ± 0.52^{a}	10.63 ± 0.17^{b}	41.99 ± 2.78^{a}	7.43 ± 0.25^{a}

Different letters (abc) from each column indicate a significant difference based on one-way ANOVA with Tukey's HSD test ($p < 0.05$).

The pod color is crucial for consumer acceptance, and brighter, greener pods are preferred in the market (Zeipiņa et al., 2017; Born, 2006). Major changes were found on L*, a*, and b* values from stage R5-1 to R6-2 but not stage R7 (Table A3). For all three genotypes, the values of L* increased rapidly from R5-1 to R6-2 followed by a slight increase to R7-2. The increase of the L* value indicates the pods became lighter with growth. The values of a* were all negative indicating that the pods were green; a* increased gradually but not significantly from R5-1 to R6-2. Increased a* was found on all three genotypes from R6-2 to R7-2 but only

significant for genotype R15-10280 and UA-Kirksey. This indicates the green color was maintained until R7-1 and then became less green. The overall increase of b* values was observed over pod development. However, the increase was not significant for genotype R15-10280 (34.77 to 37.44) and UA-Kirksey (32.62 to 35.98) but significant for genotype V16-0547 (32.88 to 39.43). The increased b* values, the degree of yellowness, indicates the pods became more yellow along with the edamame development. Overall, the color of the pods became lighter, less green, and more yellow during the development. This is probably due to the decrease of greenish chlorophyll catabolites (chlorophyll a and chlorophyll b) in edamame from stage R6 to R8 (Borrmann et al., 2009).

4.3.2 Free sugars, alanine, and glycine

4.3.2.1 Free sugars and amino acids that contribute to sweetness

The results of free sugars that contributed to sweetness (sucrose, fructose, and glucose) were shown in Figure 4.1. Sucrose was the most abundant sugar in edamame compared to fructose and glucose (Yu et al. 2021). The sucrose content increased until R6-1 and decreased thereafter. At R5-1, the sucrose contents of genotype R15-10280 and V16-0547 were both around 82.0 mg/g and increased significantly to 125.3 mg/g and 111.54 mg/g respectively at R6-1. Although it started to decrease after R6-1, the sucrose content of samples harvested at R6-2 had no significant difference from those harvested at R6-1. After R6, the sucrose content decreased significantly. UA-Kirksey had the same trend of sucrose changes with R15-10280 and V16-0547 which started from 68.7 mg/g at R5-1, increased to 76.3 mg/g at R6-1, and decreased to 54.2 mg/g finally at R7-2. However, these changes were less intense compared to the other two genotypes (R15-10280 and V16-0547). The same sucrose change trend was also found in studies by Xu et al., 2016 and Kuo et al., 1997 in which the highest sucrose content occurred at

early R6 and decreased thereafter. Sucrose is the predominant photosynthate (carbon source) that is transported from leaf tissues to the developing bean embryos and accumulates in the beans during pod filling (Du et al., 2020; Thorne, 1980). Its metabolism and accumulation are closely related to the changes in different enzymes during seed development (Sitthiwong et al., 2005; Lowell and Kuo, 1989). Sitthiwong et al. (2005) reported that the sucrose content of edamame beans was positively correlated with sucrose synthase activities. Lowell & Kuo (1989) studied the change of sucrose synthase activities from stage R4 to R8 and they found that the activities remained low before seed filling (R5) and increased by 5 times from R5 to early R6 after which the activities keep decreasing until the full maturity of the beans. The increasing activities of sucrose synthase led to the increased sucrose content from R5 to R6. The decrease of the sucrose content over bean development is usually associated with its transformation to other storage sugars, such as raffinose and stachyose. The changes of galactinol synthase activity could further explain the decrease of the sucrose content, this would be discussed later in the results of oligosaccharides contents.

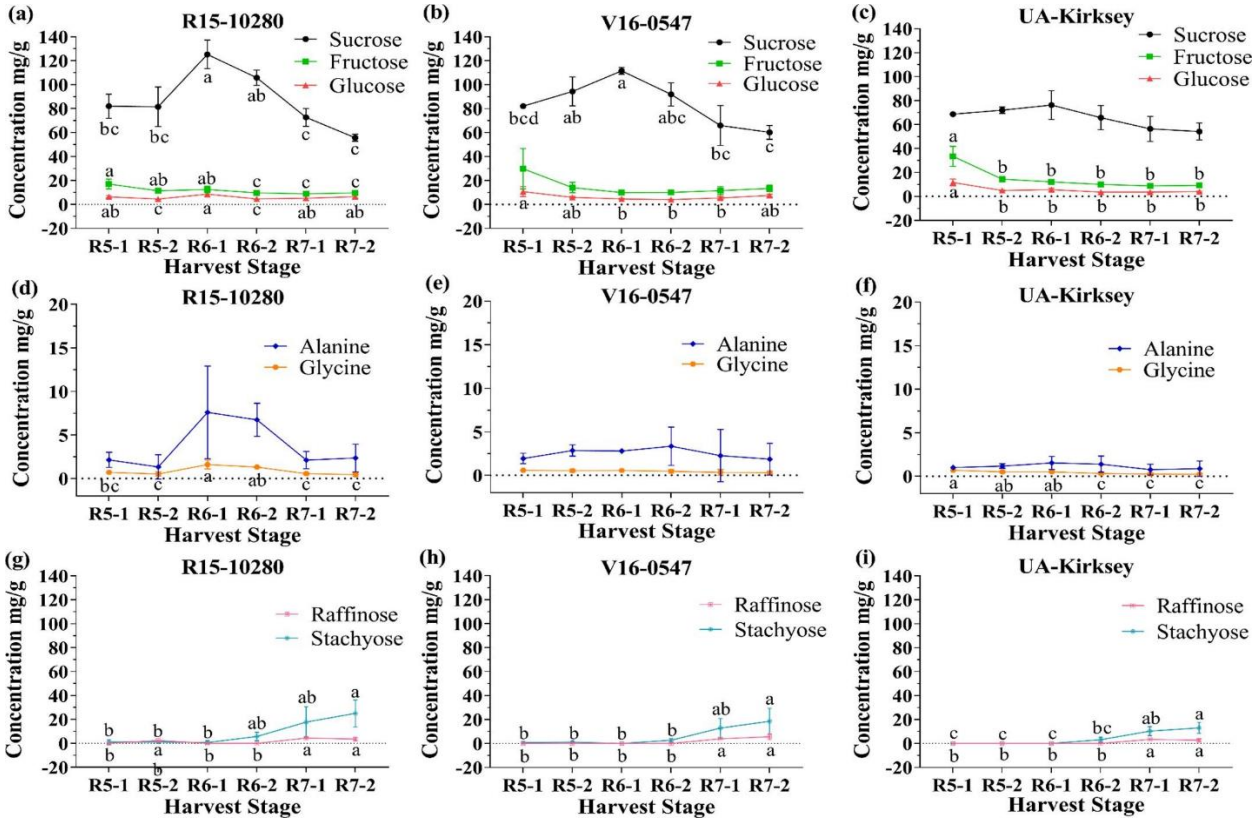

Figure 4.1 Changes in free sucrose, fructose, glucose, alanine, glycine, raffinose, and stachyose concentration (mg/g dry mass) of edamame in different harvest stages. Different letters (abcd) from each line indicate a significant difference based on the one-way ANOVA with Tukey's HSD test ($p < 0.05$). Points on the same line without significant levels have no significant levels.

Compared to sucrose, fructose and glucose contents were much lower in all three genotypes. The fructose and glucose contents at R5 were 17.0 mg/g and 6.4 mg/g, 29.8 mg/g and 10.7 mg/g, and 33.5 mg/g and 11.6 mg/g for genotype R15-10280, V16-0547, and UA-Kirksey, respectively. A significant decrease of fructose and glucose was found at different stages of these three genotypes which were R6-2 for R15-10280, R6-1 for V16-0547, and R5-2 for UA-Kirksey. After the decrease at the respective stage, these two sugars remained stable until R7-2.

Free alanine and glycine both have sweet tastes and can contribute to the sweetness of edamame. Figure 4.1(d)–(f) show the changes in free alanine and glycine from R5 to R7. Compared to free sugars, the content of free alanine and glycine were low. R15-10280 had the

highest alanine and glycine contents which were 7.59 mg/g and 1.60 mg/g at stage R6-1, respectively. For all genotypes, alanine content had no significant change from R5 to R7 indicating that alanine accumulation was not significantly affected by the bean development stage. For genotype R15-10280, glycine content increased significantly from R5 to R6 and then decreased significantly by R7, while for genotype UA-Kirksey, it decreased significantly from R6-1 to R6-2 and then remained stable. However, glycine content did not change over bean development for genotype V16-0547. Since the contents of alanine and glycine were extremely low, the changes in them have little effect on the sweetness of edamame.

4.3.2.2 Total sweetness

It was reported that edamame with higher sweetness tended to have higher consumer acceptability (Carneiro et al., 2021). The changes in total sweetness from R5-1 to R7-2 were shown in Figure A1. For genotype R15-10280, edamame harvested at R6-1 and R6-2 had the highest total sweetness than those harvested at R5 and R7. For genotype V16-0547, samples with the highest total sweetness were those harvested at R5-2 to R6-2. For UA-Kirksey, no significant difference was observed in the total sweetness of samples harvested at different developmental stages. The overall trend of the total sweetness was similar to the changes in sucrose concentration, especially for genotype R15-10280 and V16-0547 even though all sweet compounds in edamame were considered. This was due to the high amount of sucrose presented in edamame compared to other compounds, making the change of sucrose a predominant factor influencing the total sweetness. In a previous study, Yu et al. (2021) also showed that there was a high correlation coefficient (r) between the sucrose content and total sweetness.

4.3.2.3 Free raffinose and stachyose

Raffinose and stachyose are the galactosyl derivatives of sucrose and belong to the raffinose family oligosaccharides (RFOs) (Kumar et al., 2010). In legume seeds, the accumulated oligosaccharides during maturity are stored as energy sources for future germination (Kuo et al., 1988). Raffinose is composed of one galactose linked to one sucrose and stachyose is composed of two galactose molecules linked to one sucrose. Different from sucrose, fructose, and glucose, raffinose and stachyose are not desired in edamame because they cannot be digested by humans and may cause flatulence or more severe GI tract disease (Lowell & Kuo, 1989). The changes of raffinose and stachyose from harvest stage R5 to R7 are shown in Figure 4.1(g–i). For all three genotypes, no raffinose or trace amounts of raffinose were detected before R6-2, when the significant accumulation of raffinose started. Stachyose started to accumulate at R6-1 as evidenced by its trace content in beans harvested at this stage, and its content rapidly increased after R6-2. The results were in accordance with the studies of Yazdi-Samadi et al. in 1977, Dornbos and McDonald in 1986, Lowell and Kuo in 1989, and Saldivar et al. in 2011, in which they found the same pattern of accumulation of raffinose and stachyose. The final synthesis of RFOs by raffinose synthase needs both sucrose and galactinol but the first step is to synthesize the galactinol. Based on Lowell and Kuo's study in 1989, the galactinol content in maturing soybean from R6 to R7 increased from almost 0 to 3.0 mg/seed and the galactinol synthase activity stayed at a high level which initiated the accumulation of galactinol. When the beans start to mature, the raffinose synthase starts the synthesis of raffinose by using the sucrose and galactinol found in the beans. The results in Figure 4.1 also show that the stachyose content was higher than the raffinose content. The possible reason could be that the raffinose was used to synthesize the stachyose; stachyose synthesis needs raffinose as the basis and adds one more

galactinol (Jing et al., 2018). Additionally, the results show that the stachyose accumulation started at R6-1 and accelerated after R6-2. This trend was commensurate with the change in sucrose. The sucrose content decreased after R6-1, a more significant decrease was found after R6-2. The accumulation of raffinose and stachyose could help explain the decrease of sucrose after R6-1.

4.3.3 Macronutrient compositions

Besides appearance and sweetness, macronutrients in edamame are also of great importance to consumers. For this reason, starch, protein, fat, fiber, ash, and water contents were quantified (Table 4.2). Water was the most abundant component of edamame beans and made up more than 60% of the bean's wet weight. High moisture content is one of the most important characteristics of edamame. Overall, all genotypes had a high moisture content of fresh beans (69.5% to 75.6%) at R5-1 and the moisture content decreased gradually from R5-1 to R7-2. The decrease from R5-1 to R6-1 was significant for genotype V16-0547 and UA-Kirksey, but not for R15-10280. Starch is the main energy source stored in seeds and it is an important component contributing to the hardness and chewiness of edamame (Stevenson et al., 2007; Xu et al., 2012). The initial starch content of all three genotypes was around 11% and it increased slightly from R5-1 to R6-1 for genotype R15-10280 and V16-0547 while for UA-Kirksey, the starch content remained stable from R5-1 to R6-1. Starch content in all three genotypes decreased significantly from R6-1 to R7-2. This change was also reported by Saldivar et al. (2001). In their study, the starch accumulated at the beginning of bean development and then decreased drastically after the bean almost filled the pods. Starch content is related to the physical properties of edamame during post-harvest processing. Blanching is used to deactivate enzymes and increase the shelf life of edamame. It causes starch gelatinization and pectin solubilization which leads to the

decreased hardness of edamame. To get soft edamame after blanching, higher starch is usually preferred.

Table 4.2 Chemical compositions of edamame in different harvest stages.

Genotype	Harvest Stage	Moisture Content of Fresh Beans %	Protein %	Starch %	Fat %	NDF %	Ash %
R15-10280	R5-1	69.45 ± 7.22^{a}	39.10 ± 0.61^{b}	11.82 ± 0.13^{a}	13.80 ± 0.22^{c}	7.14 ± 0.51^{a}	5.85 ± 0.29^{a}
	R5-2	66.86 ± 2.53^{a}	41.40 ± 1.76^{ab}	10.99 ± 1.71^{ab}	15.18 ± 0.28^{bc}	6.24 ± 0.25^{a}	5.62 ± 0.68^{ab}
	R6-1	64.88 ± 0.65^{a}	40.27 ± 2.12^{ab}	12.58 ± 0.50^{a}	14.49 ± 0.58^{bc}	6.42 ± 0.14^{a}	4.00 ± 0.16^{c}
	R6-2	61.72 ± 1.51^{a}	42.51 ± 1.67^{ab}	9.63 ± 2.62^{ab}	16.05 ± 0.59^{ab}	6.46 ± 0.30^{a}	4.59 ± 0.31^{bc}
	R7-1	61.38 ± 3.43^{a}	42.26 ± 0.53^{ab}	6.39 ± 2.47^{bc}	16.81 ± 1.02^{a}	6.65 ± 0.81^{a}	5.57 ± 0.53^{ab}
	R7-2	62.00 ± 4.50^{a}	43.02 ± 0.97^{a}	4.75 ± 1.62^{c}	16.82 ± 0.43^{a}	6.56 ± 0.19^{a}	5.70 ± 0.34^{ab}
V16-0547	R5-1	75.56 ± 6.20^{a}	41.28 ± 1.88^{a}	10.65 ± 0.51^{ab}	12.65 ± 1.52^{b}	7.98 ± 0.24^{a}	5.98 ± 0.19^{a}
	R5-2	67.05 ± 0.75^{ab}	41.10 ± 0.56^{a}	11.57 ± 0.42^{a}	14.08 ± 1.32^{ab}	6.38 ± 0.40^{b}	5.43 ± 0.27^{a}
	R6-1	65.62 ± 0.41^{b}	40.22 ± 1.19^{a}	13.50 ± 0.72^{a}	15.12 ± 0.76^{ab}	5.29 ± 0.12^{c}	3.98 ± 0.23^{b}
	R6-2	63.55 ± 2.19^{b}	42.94 ± 1.06^{a}	10.49 ± 1.00^{ab}	17.15 ± 1.54^{ab}	5.45 ± 0.23^{bc}	4.26 ± 0.21^{b}
	R7-1	64.37 ± 4.39^{b}	40.92 ± 1.53^{a}	7.63 ± 2.06^{bc}	18.17 ± 3.02^{a}	5.86 ± 0.31^{bcd}	5.64 ± 0.44^{a}
	R7-2	59.06 ± 3.72^{b}	43.57 ± 0.98^{a}	6.67 ± 2.24^{c}	16.56 ± 1.69^{ab}	6.23 ± 0.35^{bc}	5.53 ± 0.21^{a}
UA-Kirksey	R5-1	71.91 ± 4.03^{a}	40.08 ± 1.62^{a}	10.58 ± 0.82^{a}	11.67 ± 1.19^{c}	7.06 ± 1.03^{a}	6.05 ± 0.19^{a}
	R5-2	65.93 ± 2.53^{ab}	42.36 ± 1.59^{a}	10.76 ± 1.75^{a}	16.37 ± 0.70^{b}	6.31 ± 0.29^{ab}	5.36 ± 0.47^{ab}
	R6-1	62.94 ± 1.59^{bc}	42.88 ± 0.97^{a}	10.56 ± 0.23^{a}	16.81 ± 0.76^{ab}	5.65 ± 0.21^{b}	4.38 ± 0.17^{b}
	R6-2	61.40 ± 0.57^{bc}	43.89 ± 1.34^{a}	8.26 ± 0.58^{ab}	18.90 ± 0.32^{a}	5.80 ± 0.26^{ab}	4.38 ± 0.40^{b}
	R7-1	60.45 ± 1.28^{bc}	42.55 ± 2.10^{a}	7.27 ± 0.77^{b}	18.28 ± 1.23^{ab}	6.40 ± 0.31^{ab}	5.62 ± 0.59^{a}
	R7-2	58.56 ± 1.82^{c}	42.93 ± 1.93^{a}	7.13 ± 0.69^{b}	17.64 ± 0.97^{ab}	6.43 ± 0.04^{ab}	5.11 ± 0.39^{ab}

The results of the moisture content of fresh beans were calculated based on wet weight and other compositions were calculated based on the dry weight. Different letters (abc) from each column indicate a significant difference based on the one-way ANOVA with Tukey's HSD test ($p < 0.05$).

Edamame had around 40% protein based on dry weight and this makes edamame a good source of vegetable protein. The protein contents of all harvested edamame samples ranged between 39.1% and 43.9% with a slight increase (2–3%) from R5-1 to R7-2. This slight increase agreed with the studies of Dornbos and McDonald, 1986 and Saldivar et al., 2011. In their studies, the protein content of edamame increased by 2–5% from R5 to R7. Overall, the protein content was stable from stage R5 to R7. The fat content of R15-10280, V16-0547, and UA-Kirksey was 13.8, 12.7, and 11.7%, respectively, at the R5-1 stage. It accumulated from R5 to R6-2 and no significant change was found afterward. The initial NDF content at R5-1 was around 7 to 8% and the initial ash content was around 6%. The NDF contents decreased significantly from R5-1 to R6-1 for genotype V16-0547 and UA-Kirksey and increased slightly afterward. For R15-10280, the change of NDF has the same trend with the other two genotypes but the change was not significant overall. A significant decrease from R5 to R6 and a significant increase from R6 to R7 of ash content were found in all three genotypes.

Overall, based on the physical and chemical properties, the early stage of R6 (R6-1) is the optimal time to harvest edamame. At this stage, edamame has a large bean size, favorable light green color, high sweetness, and high protein and starch contents, as well as low raffinose and stachyose contents. To have a longer harvest window, R6-2 could also be an acceptable harvest window because no significant difference was found in both the physical and chemical properties compared to edamame harvested at R6-1.

4.3.4 Spectroscopy-based machine learning classifier

4.3.4.1 Spectral reflectance curves

The spectral data from all pods in early, ready, and late stages were averaged and plotted in Figure 4.2. Overall, the spectral reflectance curves increase gradually from 360 to 490 nm,

then start to increase rapidly and reach the first peak at approximately 565 nm. Afterward, the reflectance curves decrease and reach the local minimum between 670 and 690 nm before increasing again. The color spectrometer used in this study has a maximum detection wavelength at approximately 700 nm. The shape of the reflectance curves follows the typical observation of vegetation, in particular the green leaves, which are consistent with the observation of the green color of edamame pods when they are ready to be harvested. Although the overall trends are the same for all three categories of edamame pods, there is a clear separation between the average spectral curves among these categories. The pods that are too early to harvest have the lowest average reflectance compared to the pods that are ready and too late to harvest. The increase of reflectance in this spectral range is similar to those observed on senescent leaves (Carter, 1993).

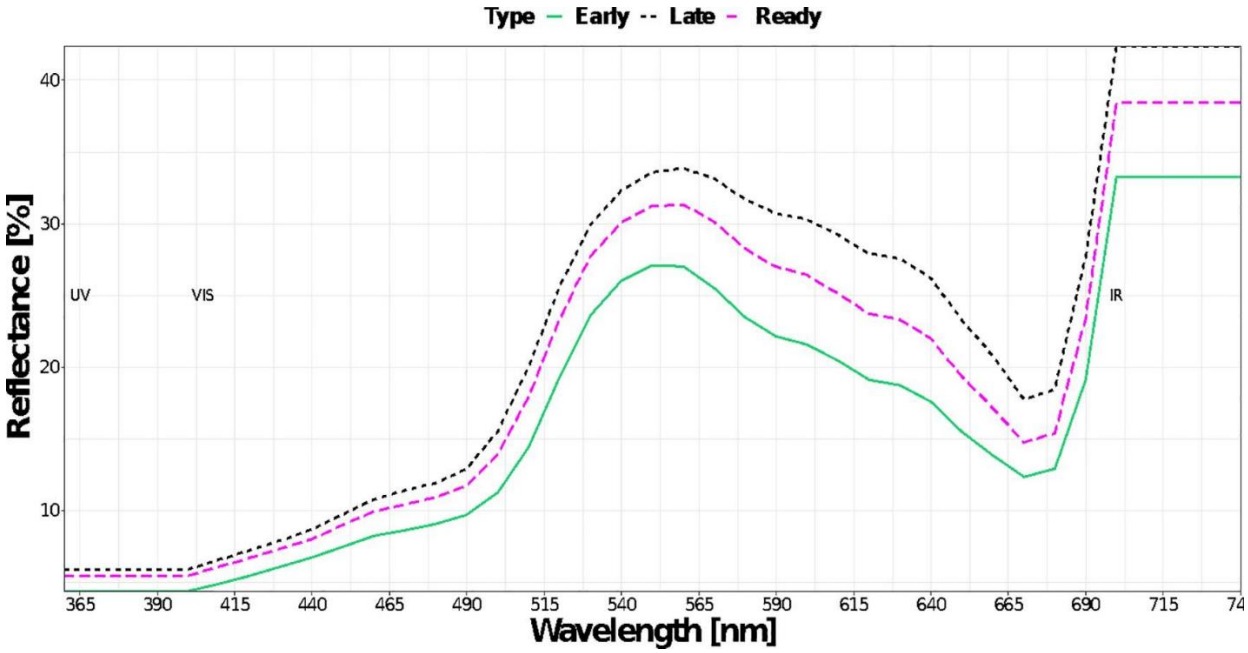

Figure 4.2 Analysis of spectral reflectance using machine learning. Average spectral reflectance of edamame pods that are in the three categories: early, late, and ready to harvest.

4.3.4.2 Three-class classification using full wavelengths

To determine, quantitatively, whether we can classify edamame pods from different categories using spectral data, we performed the classification using the RF algorithm (Figure 4.3). At first, three-class RF classification models were trained using the primary (Prim) spectral data and first-order derivative (FOD) spectral data at full wavelengths, respectively (Figure 4.3a). The trained models were tested with a new set of data and the accuracies were 0.69 (Prim) and 0.71 (FOD). Although the accuracies of 0.69 – 0.71 were not ideal, they indicate that the classification of 'early', 'ready', and 'late' pod stages may be realized using reflective spectra coupled with the machine learning technique, and a further improvement in accuracy is needed. A previous study used similar spectroscopy-based techniques to predict harvest time for apples (Bertone et al., 2012). In their study, UV–Vis and near-infrared spectroscopies coupled with partial least square regression were used to monitor the chlorophyll content (the green color) of the skin of the red apple, which indicates the apple ripeness. The major difference is that the

response variable in Bertone et al.'s study is continuous whereas, in our situation, it is categorical. Therefore, classification using machine learning methods such as RF is feasible for predicting the optimal harvest time for edamame and other vegetables/fruits.

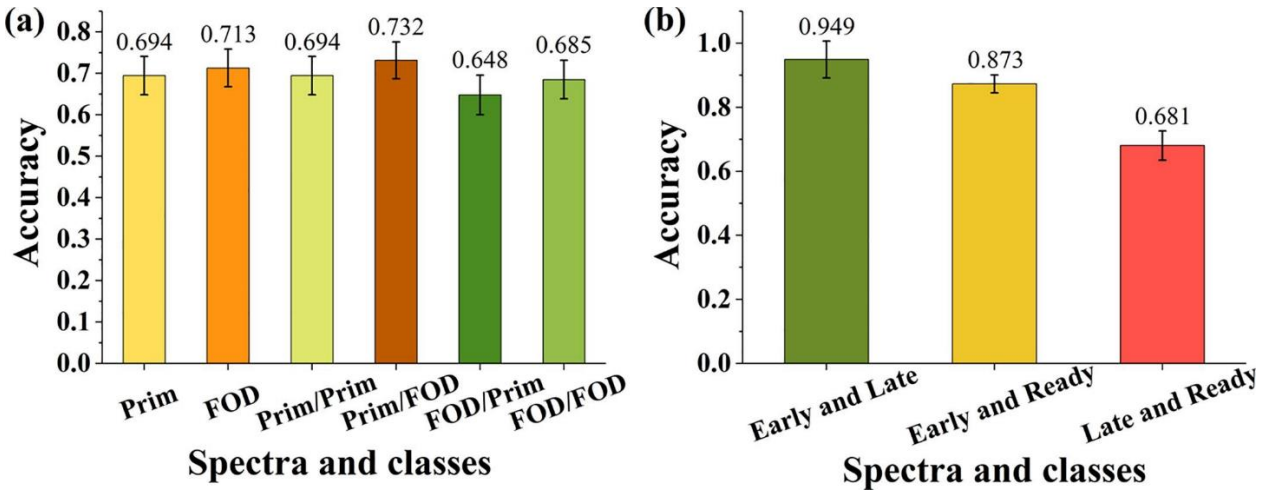

Figure 4.3 Comparison of model accuracy among different classification methods. (A) Classification accuracy of three categories. (B) Classification accuracy of two categories. Prim: primary spectral data. FOD: first-order derivative of spectral data. Prim/Prim: use Prim selected spectral wavelengths and Prim spectral data for classification. Prim/FOD: use Prim selected spectral wavelengths and FOD spectral data for classification. FOD/Prim: use FOD selected spectral wavelengths and Prim spectral data for classification. FOD/FOD: use FOD selected wavelengths and FOD data for classification.

4.3.4.3 Three-class classification using selected wavelengths

Using a large number of spectral bands might create challenges for collecting data in the field and further lead to redundant information. Massive data generated by spectroscopic technique is also a challenge for data analysis. Therefore, selecting important wavelengths that carry the most useful information with minimal redundancy can help improve both the data collection and the data analysis efficiency. To determine the performances of models using selected important spectral bands, 12 wavelengths were selected from the analysis using the Prim spectral data and 9 wavelengths were selected from the analysis using the FOD spectral data. The number of wavelengths (12 and 9) was automatically selected through a machine learning

process called feature selection in the RF method. Three-class RF classification models were trained again using the Prim and FOD spectral data of these selected wavelengths and their performances were shown in Figure 4.3a, Prim/Prim, FOD/FOD, FOD/Prim, and Prim/FOD. Interestingly, the models built from selected important wavelengths showed similar accuracies (0.65 to 0.73) compared to the models built from full wavelengths. A previous study has compared the performances of support vector machine (SVM) models built on full spectra and selected optimal spectra of hyperspectral image systems on evaluating the ripeness of strawberries (Guo et al., 2016). They found that SVM models built on selected optimal spectra showed acceptable results compared to models built on full spectra. The difference is that in their study, the selected optimal spectra were from two different spectral ranges. Satisfactory results were observed on selected optimal wavelengths from 441.1 to 1013.97 nm and worse results were found on selected optimal wavelengths of 941.46 to 1578.13 nm. Thus, our study, together with others, indicates that the reduction of wavelength number can be realized without compromising the model accuracies for classifying edamame with different maturity stages.

Overall, the model trained using FOD spectral data performs better than using the Prim spectral data. This finding was not surprising because FOD spectra are generally better at resolving overlapping wavebands and reducing random noise. In a detailed investigation of spectral classification techniques, Ghiyamat et al. (2013) showed that FOD-based approaches showed the least improvement (over primary spectra) in very complex datasets and the most improvement in less complex datasets. In this study, it was found that classification accuracy increased when using an RF classifier combined with the FOD spectra. Classification accuracy using primary spectra and a random forest classifier could still be considered substantial. Taken together with the results from other studies, it seems that both the classification method and the

number of classes used in the classification influence whether the FOD spectra can improve the classification accuracy. However, visual inspection of FOD spectral data does not show a clear separation of the curves (Figure 4.2).

4.3.4.4 Two-class classification

After analyzing the confusion matrix, we noticed that the model performed poorly when classifying spectral data of the "ready" category; however, it performed better when separating "early" with "late" categories (Figure 4.4). With this observation, the data was further analyzed using two-class classification between every two categories separately with only primary spectral data. We did not use the FOD because the visual inspection of FOD does not suggest a clear separation of "early" and "late" categories (data not shown). The model accuracy increases substantially to 0.95 for separating "early" and "late" categories and to 0.87 for separating "early" and "ready" categories (Figure 4.3b). The model accuracy for separating "late" and "ready" remains as low as 0.68. Feature selection was performed, and 12 wavelengths were selected as the important ones (430, 450, 490, 500, 540, 550, 560, 600, 630, 640, 680, and 700 nm). The model using selected wavelengths provides a similar accuracy for the primary spectral data of edamame pods. The higher model accuracies for two-class classifications were in agreement with Cen et al. (2016)'s study, which applied a hyperspectral imaging technique on the detection of chilling injury in cucumber fruit. The overall accuracies for the two-class classifications (i.e., normal and chilling) were 100%, while the overall accuracies for the three-class classifications (i.e., normal, lightly chilling, and severely chilling) were lower at 91.6%.

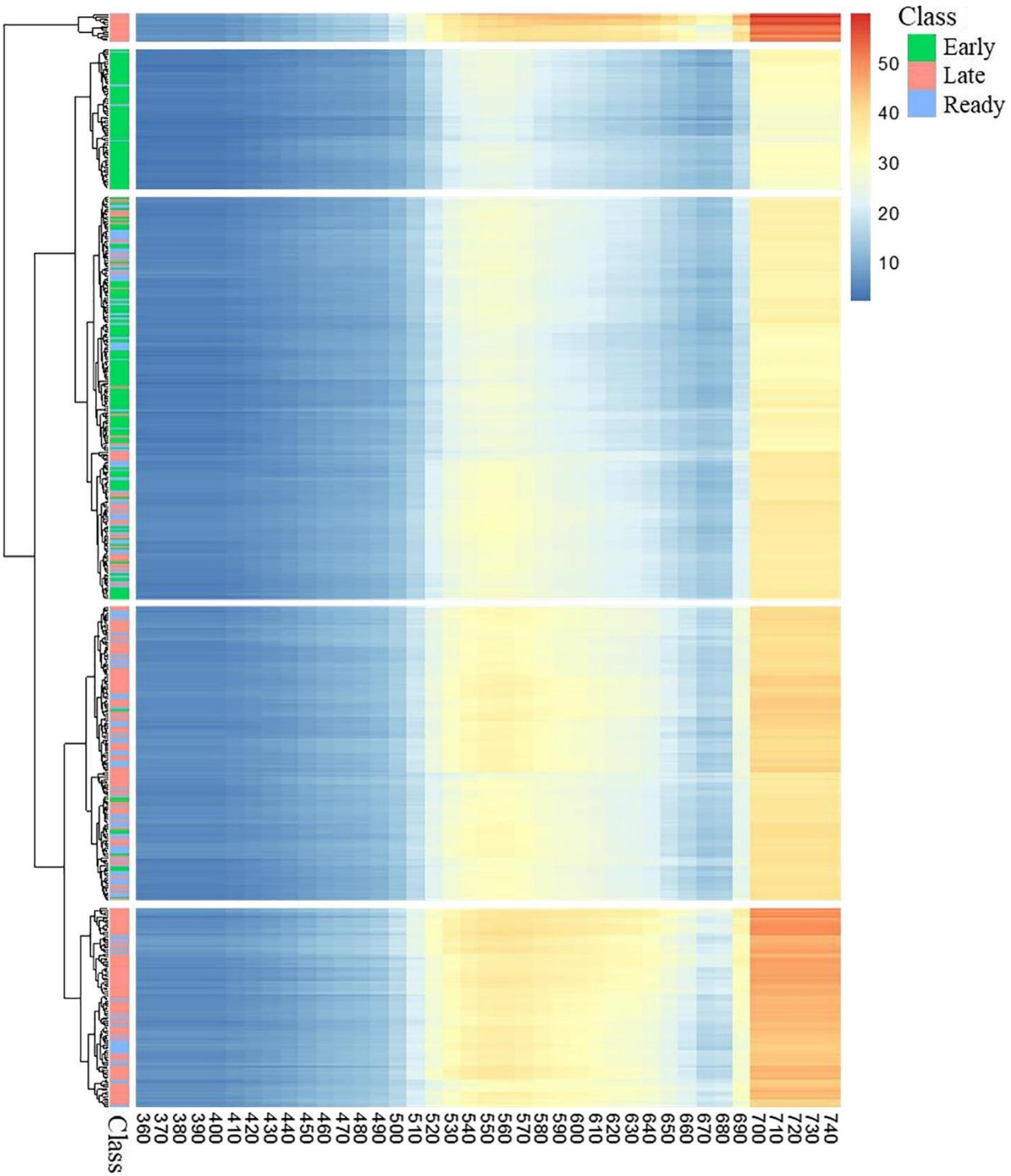

Figure 4.4 Hierarchical clustering of spectral reflectance across all spectral data of three categories. The clustering results are separated into five groups based on the branch height in the dendrogram. More or fewer groups can be generated but the main observation remains the same.

Overall, it was demonstrated that the RF classification method can reasonably identify the early, late, and ready stages of edamame harvest using the spectra collected by a handheld, portable spectrometer. This is practically important because edamame has a short harvest window (only about one week) for producing edamame with high marketability and consumer acceptability. The portable spectrometer coupled with the machine learning technique will allow for rapidly and accurately determining optimal harvest time in the field, ensuring peak morphological and eating quality of edamame, and mitigating the heavy reliance on experienced edamame growers through touch, taste, and observation to determine the harvest time. In real edamame production, it often happens that those pods of different maturity stages are on the same plant at the same time. For the small-scale edamame production where edamame is manually harvested, the technique will provide critical information about which pods are ready to harvest. For the large-scale edamame production where edamame is harvested by large combines, it is not realistic to only harvest mature pods and leave young ones. However, the technique can still help decide the best harvesting time when the majority of pods in plants are "ready to harvest".

In the future, it may be necessary to refine spectral sets of data down to some level where a single waveband can be considered unique for the system under investigation. Further investigation would be needed to confirm or disprove this suggestion. The successful discrimination between spectral signatures is only the first step towards using spectral approaches to determine the proper harvesting stage in the edamame industry. The results of the present study represent a proof of concept for incorporating a spectral approach into a precision farming tool used for the edamame.

4.4 Conclusions

This study investigated the changes in physical and chemical properties of edamame over bean development and applied a spectroscopy-based machine learning method to identify the optimal harvest time of edamame. Pod weight, bean weight, and pod thickness reach the peak values at stage R6. The color of edamame becomes lighter, more yellow, and less green as the beans develop. All genotypes have similar chemical composition changes from R5 to R7. The sucrose, alanine, glycine, and starch contents are highest at R6 when the edamame has the highest sweetness. Oppositely, the fat, NDF, and ash contents are relatively low at this stage. Considering all physical properties and chemical composition changes over the bean development, the early R6 (R6-1) stage was determined as the optimal time to harvest edamame. However, if a longer harvest window is needed, R6-2 is acceptable and better than R5 and R7. The machine learning method based on the pods' spectral reflectance had a high accuracy of 0.95 for classifying "early" and "late" samples and 0.87 for classifying "early" and "ready" samples. However, a relatively low accuracy of 0.68 was obtained for classifying "late" and "ready" samples. Overall, this study demonstrated that the machine learning method based on the pods' spectra reflectance can identify the optimal harvest time of edamame.

4.5 References

Belgiu, M. and Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, pp.24-31.

Bertone, E., Venturello, A., Leardi, R. and Geobaldo, F., 2012. Prediction of the optimum harvest time of 'Scarlet' apples using DR-UV–Vis and NIR spectroscopy. Postharvest Biology and Technology, 69, pp.15-23.

Birch, G.G. and Kemp, S.E., 1989. Apparent specific volumes and tastes of amino acids. Chemical Senses, 14(2), pp.249-258.

Born, H., 2006. Edamame: vegetable soybean. https://attra.ncat.org/product/edamame-vegetable-soybean/ [Accessed on August 4, 2020].

Borrmann, D., de Andrade, J.C. and Lanfer-Marquez, U.M., 2009. Chlorophyll degradation and formation of colorless chlorophyll derivatives during soybean (*Glycine* max L. Merill) seed maturation. Journal of Agricultural and Food Chemistry, 57(5), pp.2030-2034.

Carneiro, R., Duncan, S., O'Keefe, S., Yu, D., Huang, H., Yin, Y., Neill, C., Zhang, B., Kuhar, T., Rideout, S. and Reiter, M., 2021. Utilizing consumer perception of edamame to guide new variety development. Frontiers in Sustainable Food Systems, 4, p.556580.

Carson, L.C., Freeman, J.H., Zhou, K., Welbaum, G. and Reiter, M., 2011. Cultivar evaluation and lipid and protein contents of Virginia-grown edamame. HortTechnology, 21(1), pp.131-135.

Carson, L.C., 2010. Cultivation and nutritional constituents of Virginia grown edamame (Doctoral dissertation, Virginia Tech).

Carter, G.A., 1993. Responses of leaf spectral reflectance to plant stress. American Journal of Botany, 80(3), pp.239-243.

Cen, H., Lu, R., Zhu, Q. and Mendoza, F., 2016. Nondestructive detection of chilling injury in cucumber fruit using hyperspectral imaging with feature selection and supervised classification. Postharvest Biology and Technology, 111, pp.352-361.

Cortés, V., Blasco, J., Aleixos, N., Cubero, S. and Talens, P., 2019. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. Trends in Food Science & Technology, 85, pp.138-148.

de Santana, F.B., Neto, W.B. and Poppi, R.J., 2019. Random forest as one-class classifier and infrared spectroscopy for food adulteration detection. Food Chemistry, 293, pp.323-332.

Dornbos Jr, D.L. and McDonald Jr, M.B., 1986. Mass and composition of developing soybean seeds at five reproductive growth stages 1. Crop Science, 26(3), pp.624-630.

Du, Y., Zhao, Q., Chen, L., Yao, X., Zhang, H., Wu, J. and Xie, F., 2020. Effect of drought stress during soybean R2–R6 growth stages on sucrose metabolism in leaf and seed. International Journal of Molecular Sciences, 21(2), p.618.

Gao, Z., Shao, Y., Xuan, G., Wang, Y., Liu, Y. and Han, X., 2020. Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning. Artificial Intelligence in Agriculture, 4, pp.31-38.

Ghiyamat, A., Shafri, H.Z.M., Mahdiraji, G.A., Shariff, A.R.M. and Mansor, S., 2013. Hyperspectral discrimination of tree species with different classifications using single-and multiple-endmember. International Journal of Applied Earth Observation and Geoinformation, 23, pp.177-191.

Guo, C., Liu, F., Kong, W., He, Y. and Lou, B., 2016. Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine. Journal of Food Engineering, 179, pp.11-18.

Heim, R.H.J., Wright, I.J., Chang, H.C., Carnegie, A.J., Pegg, G.S., Lancaster, E.K., Falster, D.S. and Oldeland, J., 2018. Detecting myrtle rust (*Austropuccinia psidii*) on lemon myrtle trees using spectral signatures and machine learning. Plant Pathology, 67(5), pp.1114-1121.

Jing, Y., Lang, S., Wang, D., Xue, H. and Wang, X.F., 2018. Functional characterization of galactinol synthase and raffinose synthase in desiccation tolerance acquisition in developing Arabidopsis seeds. Journal of Plant Physiology, 230, pp.109-121.

Johnson, D., Wang, S., & Suzuki, A., 1999. Edamame: a vegetable soybean for Colorado. Perspectives on New Crops and New Uses, 385–387.

Konovsky, J., Lumpkin, T.A. and McClary, D., 2020. Edamame: the vegetable soybean. In Understanding the Japanese Food and Agrimarket (pp. 173-181). CRC Press.

Kumar, V., Rani, A., Goyal, L., Dixit, A.K., Manjaya, J.G., Dev, J. and Swamy, M., 2010. Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. Journal of Agricultural and Food Chemistry, 58(8), pp.5081-5085.

Kuo, T.M., Lowell, C.A. and Smith, P.T., 1997. Changes in soluble carbohydrates and enzymic activities in maturing soybean seed tissues. Plant Science, 125(1), pp.1-11.

Kuo, T.M., VanMiddlesworth, J.F. and Wolf, W.J., 1988. Content of raffinose oligosaccharides and sucrose in various plant seeds. Journal of Agricultural and Food Chemistry, 36(1), pp.32-36.

Licht, M., 2014. Soybean growth and development. Iowa State University Extension and Outreach, Ames, Iowa, USA

Lowell, C.A. and Kuo, T.M., 1989. Oligosaccharide metabolism and accumulation in developing soybean seeds. Crop Science, 29(2), pp.459-465.

Machado, S., Costa, A.S., Pimentel, F., Oliveira, M.B.P. and Alves, R.C., 2020. A study on the protein fraction of coffee silverskin: Protein/non-protein nitrogen and free and total amino acid profiles. Food Chemistry, 326, p.126940.

Mentreddy, S.R., Mohamed, A.I., Joshee, N. and Yadav, A.K., 2002. Edamame: a nutritious vegetable crop Edamame: a nutritious vegetable crop. Trends in new crops and new uses, ASHS Press, Atlanta, Georgia, USA (2002), pp. 432-438.

Moseley, D., Da Silva, M.P., Mozzoni, L., Orazaly, M., Florez-Palacios, L., Acuna, A., Wu, C. and Chen, P., 2021. Effect of planting date and cultivar maturity in edamame quality and harvest window. Frontiers in Plant Science, 11, p.585856.

OHair, J., Jin, Q., Yu, D., Wu, J., Wang, H., Zhou, S. and Huang, H., 2021. Non-sterile fermentation of food waste using thermophilic and alkaliphilic *Bacillus licheniformis* YNP5-TSU for 2, 3-butanediol production. Waste Management, 120, pp.248-256.

Peidad, E.J., Larada, J.I., Pojas, G.J. and Ferrer, L.V.V., 2018. Postharvest classification of banana (Musa acuminata) using tier-based machine learning. Postharvest biology and technology, 145, pp.93-100.

Poe, N.E., Yu, D., Jin, Q., Ponder, M.A., Stewart, A.C., Ogejo, J.A., Wang, H. and Huang, H., 2020. Compositional variability of food wastes and its effects on acetone-butanol-ethanol fermentation. Waste Management, 107, pp.150-158.

Purcell, L.C., Salmeron, M. and Ashlock, L., 2014. Soybean growth and development. Arkansas Soybean Production Handbook, MP197 (2014), pp. 1-8

Saldivar, X., Wang, Y.J., Chen, P. and Hou, A., 2011. Changes in chemical composition during soybean seed development. Food Chemistry, 124(4), pp.1369-1375.

Schiffman, S.S., Sennewald, K. and Gagnon, J., 1981. Comparison of taste qualities and thresholds of D-and L-amino acids. Physiology & behavior, 27(1), pp.51-59.

Shen, F., Zhang, B., Cao, C. and Jiang, X., 2018. On-line discrimination of storage shelf-life and prediction of post-harvest quality for strawberry fruit by visible and near infrared spectroscopy. Journal of Food Process Engineering, 41(7), p.e12866.

Singh, A.K., Ganapathysubramanian, B., Sarkar, S. and Singh, A., 2018. Deep learning for plant stress phenotyping: trends and future perspectives. Trends in Plant Science, 23(10), pp.883-898.

Singh, A., Ganapathysubramanian, B., Singh, A.K. and Sarkar, S., 2016. Machine learning for high-throughput stress phenotyping in plants. Trends in Plant Science, 21(2), pp.110-124.

Sitthiwong, K., Matsui, T., Okuda, N. and Suzuki, H., 2005. Changes in carbohydrate content and the activities of acid invertase, sucrose synthase and sucrose phosphate synthase in vegetable soybean during fruit development. Asian Journal of Plant Sciences, 4 (6), pp. 684-690,

Solms, J., Vuataz, L. and Egli, R.H., 1965. The taste of L-and D-amino acids. Experientia, 21(12), pp.692-694.

Stevenson, D.G., Jane, J.L. and Inglett, G.E., 2007. Structures and physicochemical properties of starch from immature seeds of soybean varieties (*Glycine* max (L.) Merr.) exhibiting normal, low-linolenic or low-saturated fatty acid oil profiles at maturity. Carbohydrate Polymers, 70(2), pp.149-159.

Thorne, J.H., 1980. Kinetics of 14C-photosynthate uptake by developing soybean fruit. Plant Physiology, 65(5), pp.975-979.

Vidal, B.C., Rausch, K.D., Tumbleson, M.E. and Singh, V., 2009. Determining corn germ and pericarp residual starch by acid hydrolysis. Cereal Chemistry, 86(2), pp.133-135.

Wang, L., Wang, T. and Fehr, W.R., 2006. Effect of seed development stage on sphingolipid and phospholipid contents in soybean seeds. Journal of Agricultural and Food Chemistry, 54(20), pp.7812-7816.

Xu, Y., Cartier, A., Kibet, D., Jordan, K., Hakala, I., Davis, S., Sismour, E., Kering, M. and Rutto, L., 2016. Physical and nutritional properties of edamame seeds as influenced by stage of development. Journal of Food Measurement and Characterization, 10(2), pp.193-200.

Xu, Y., Sismour, E., Pao, S., Rutto, L., Grizzard, C. and Ren, S., 2012. Textural and microbiological qualities of vegetable soybean (edamame) affected by blanching and storage conditions. Journal of Food Processing & Technology, 3(6), pp.1-6.

Yang, H.Q., 2011. Nondestructive prediction of optimal harvest time of cherry tomatoes using VIS-NIR spectroscopy and PLSR calibration. In Advanced Engineering Forum (Vol. 1, pp. 92-96). Trans Tech Publications Ltd.

Yazdi-Samadi, B., Rinne, R.W. and Seif, R.D., 1977. Components of developing soybean seeds: Oil, protein, sugars, starch, organic acids, and amino acids 1. Agronomy Journal, 69(3), pp.481-486.

Yu, D., Lin, T., Sutton, K., Lord, N., Carneiro, R., Jin, Q., Zhang, B., Kuhar, T., Rideout, S., Ross, J. and Duncan, S., 2021. Chemical compositions of edamame genotypes grown in different locations in the US. Frontiers in Sustainable Food Systems, 5, p.620426.

Yu, D., Sun, Y., Wang, W., O'Keefe, S.F., Neilson, A.P., Feng, H., Wang, Z. and Huang, H., 2020. Recovery of protein hydrolysates from brewer's spent grain using enzyme and ultrasonication. International Journal of Food Science & Technology, 55(1), pp.357-368.

Yu, X., Yuan, F., Fu, X. and Zhu, D., 2016. Profiling and relationship of water-soluble sugar and protein compositions in soybean seeds. Food Chemistry, 196, pp.776-782.

Zeipiņa, S., Alsiņa, I. and Lepse, L., 2017. Insight in edamame yield and quality parameters: A review. Research for Rural Development, 2, pp.40-45.

Chapter 5

Production and Properties of Soluble Dietary Fiber from Edamame Shell using Combined Physical and Chemical Treatments

5.1 Introduction

Dietary fiber is considered one of the most important components of the human diet due to its health benefits of preventing certain chronic diseases such as hemorrhagic and ischaemic stroke, cardiovascular diseases, rectal cancer, and diabetes (Gill et al., 2021). Based on the solubility in water, dietary fiber is classified into insoluble dietary fiber (IDF) and soluble dietary fiber (SDF). SDF is a non-cellulosic polysaccharide composed of indigestible polysaccharides (e.g., pectin, gum, inulin, β-glucan and soluble hemicellulose) and oligosaccharides (Dai and Chau, 2017; Li et al., 2022). Various studies found that SDF intake can help manage elevated plasma cholesterol (Haskell, et al., 1992), reduce body weight gain (Artiss et al., 2006), and prevent metabolic syndrome (Wang et al., 2021). Additionally, compared to IDF, SDF is easier to be metabolized by gut bacteria, promotes the growth of health-promoting gut microorganisms, and increases the production of beneficial metabolites such as short-chain fatty acids (Guan et al., 2021). More importantly, SDF has superior physicochemical properties to modulate food fluidity, viscosity, and emulsification (Anderson et al., 2009; Li et al., 2022). Studies have reported the addition of SDF significantly increased the water absorption of dough, resulting in dumpling wrappers with higher hardness and cooking yield and decreased cooking loss (Wu et al., 2014). SDF can also be combined with gelatin to form composite gels as a fat substitution in meat manufacturing to provide a healthy diet (Essa and Elsebaie, 2022).

Like other food processing byproducts (e.g., tomato pomace, pomegranate peel, and lemon peel), edamame shell, a low-value byproduct generated by edamame food/snack processing, is a rich source of dietary fiber. However, the tough and hard texture and the high IDF amount of edamame shell make it not suitable to be directly used as a food additive since it affects not only the sensory properties but also the processibility of foods (Li et al., 2022). In order to better utilize the dietary fiber in edamame shells, one promising way is to convert the IDF in edamame shell to SDF. Current techniques including physical, chemical, and biological methods have been investigated to partially convert the IDF to SDF with desirable functional and nutritional properties of the produced SDF (Bader Ul Ain et al., 2019). Among many available physical treatments, milling is one of the important operations in the processing of food by-products (Chitrakar et al., 2020). Ball milling receives particular attention due to its high effectiveness in particle size reduction and unique capacity to modify the physicochemical, functional, and nutritional properties of IDF (Chitrakar et al., 2020; Song et al., 2021). Chen et al. (2020) used ball milling to improve the dietary value of okara (soybean residue) and found that IDF was efficiently converted to SDF with a conversion rate of 43g/100g and ball-milled okara possesses a higher dietary value such as the higher total flavonoid contents and bioavailability. It was also reported that micronization by ball milling significantly reduced the lignin and cellulose contents while increasing the content of SDF in grape pomace (Bender et al., 2020). Ball milling was also combined with high-pressure homogenization to improve the physicochemical and rheological properties of citrus fiber (Jiang et al., 2022). Chemical processing using sulfurous acid is environmentally unfriendly due to the SO_2 release and generation of toxic substances in the process of dissolving sulfur dioxide in water (Liu et al., 2021). Citric acid, a commonly found organic acid that exists in many fruits, such as oranges,

pineapples, and grapefruits (Behera et al., 2021), has been widely used in the food industry as a safe food ingredient/additive (Yan et al., 2019). On the other hand, citric acid can facilitate SDF extraction by ionizing the hydrogen ions with carboxyl groups and breaking up the complex cell walls. Yan et al. (2019) found that the combination of citric acid and ultrasound treatment not only increased the extraction yield of SDF from wheat bran but also elevated the antioxidant and α-amylase inhibitory activities of the extracted SDF.

The objective of this study is to develop a combined ball milling and citric acid process to extract SDF from edamame shells. To the best of our knowledge, few or no studies have been conducted for producing SDF from edamame shells by the combined ball milling and citric acid process. This study also investigated how different processing parameters, including citric acid concentrations, treatment temperatures and time, and the application of ball milling, affect the yield, physicochemical, morphological and structural, and functional properties of SDF from edamame shells. The outcome of this study not only develops an environmentally friendly process to convert edamame processing waste into high-value SDF but also provides a new source of food ingredients in the food industry, thus supporting the sustainability of the food supply chain.

5.2 Materials and methods

5.2.1 Materials & reagents

Edamame shell was collected from the food processing pilot plant at Virginia Tech (Blacksburg, VA, USA) and stored at -20 °C until use. Citric acid monohydrate, trifluoroacetic acid (TFA), meta-hydroxydiphenyl, sulfuric acid, methanol, sodium carbonate, ethanol, potassium peroxydisulfate, Trolox, 2,2-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt (ABTS), sodium dodecyl sulfate (SDS), D-glucuronic acid and 3-phenylphenol

(90%) were purchased from Thermo Fisher Scientific (Hampton, NH, USA). Sodium tetraborate, sugar standards, gallic acid, Folin & Ciocalteu's phenol reagent, and formic acid were purchased from Sigma-Aldrich (St. Louis, MO, USA).

5.2.2 Composition analysis

The moisture content of fresh edamame shells was measured under 105 °C in an oven until the weight was constant (Yu et al., 2021). Crude protein, fat, and ash contents of edamame shell powder were determined according to AOAC 990.03, 2003.05, and 942.05, respectively. Soluble and insoluble dietary fiber were analyzed using the Integrated Total Dietary Fiber Assay Kit (K=INTDF) according to AOAC 2011.25. The composition analysis showed that the edamame shell contained 75.8% moisture, 51.3% IDF, 10.3% SDF, 13.7% protein, 7.0% ash, and 0.8% fat.

5.2.3 Sample pretreatment

The frozen edamame shells were dried at 50 °C in a convection oven until the weight was constant. Afterward, the dried shells were milled by an IKA MF 10 Basic Microfine Grinder (IKA®-Werke GmbH & Co. KG, Germany) equipped with a 500-μm sieve to get the edamame shell powder. Protein was removed from the edamame shell powder using enzymatic hydrolysis with Alcalase 2.4 L at a loading of 0.8% w/w. The solid loading of the edamame-shell powder suspension for enzymatic hydrolysis is 10% w/v. The pH was adjusted to 7.5 by 4M NaOH and the sample was incubated in a 60 °C water bath with a shaking speed of 115 rpm. After 4 h of incubation, the sample was centrifuged at 13,800 × g for 10 min at 4 °C. The solid was collected and freeze-dried to get the pretreated edamame shell powder.

5.2.4 Ball mill treatment

The pretreated edamame shell powder was further treated by a ball mill (Ball Mill Emax, Retsch GmbH, Germany) with 50-mL grinding jars and 5-mm stainless grinding balls. Two grams of pretreated edamame shell powder were placed into each jar with 200 grinding balls (a total volume of 25 mL). The speed of the ball mill was set at 1,000 rpm and the processing time was 36 min with an interval of 15 min and a break of 5 min. The particle size of the edamame powder before and after the ball milling was measured by a laser diffraction particle size analyzer (Model LS13320XR, Beckman Coulter, Inc., USA).

5.2.5 Citric acid treatment

The acid treatment was carried out with different concentrations of citric acid (1% and 2%), treatment temperatures (90 and 130 °C), and treatment time (30 or 60 min). To be detailed, 1 g of pretreated edamame shell powder was mixed with 20 mL citric acid in a 50 mL media bottle, which was incubated in a water bath (90 °C) or autoclave (130 °C). For the control group, deionized water was used instead of citric acid, and the incubation lasted for 30 min under 90 °C. Afterward, the incubated mixture was cooled down in an ice water bath and centrifuged at 13,800 × g for 10 min. The supernatant was collected, mixed with 4 volumes of 95% ethanol, and then stored under 4 °C overnight to precipitate SDF. The precipitated SDF was collected after centrifugation at 13,800 × g for 20 min and dried in a convection oven until constant weight. Nine different treatments (including the control) were conducted in total and the yield from each treatment was calculated using the following equation:

$$Yield\ (\%) = \frac{Weight\ of\ SDF/g}{Dry\ mass\ of\ pretreated\ edamame\ shell/g} \times 100\% \quad (5.1)$$

For the ball-milling treated sample, only the temperature of 130 °C and treatment time of 30 min was tested according to the performances of different treatment parameters. The

following procedures of SDF production from the ball-milling treated sample were the same as the above-mentioned procedures.

5.2.6 Characterization of properties

5.2.6.1 Sample preparation

To investigate how different processing parameters affect the physical and chemical properties of the produced SDF, four representative SDF samples were chosen for comprehensive property characterization. These four treatments were listed here: 1) 90 °C, 30 min, 1% acid (SDF-I); 2) 130 °C, 30 min, 1% acid (SDF-II); 3) 130 °C, 30 min, 2% acid (SDF-III); 4) 130 °C, 30 min, 1% acid and ball milling (SDF-IV). The samples were prepared using the same procedures described in sections 5.2.3 - 5.2.5. The only difference is that the precipitated SDF was redissolved in DI water and freeze-dried instead of oven-dried for the best maintenance of the SDF properties.

5.2.6.2 Physicochemical properties

Uronic acid content

Uronic acid was determined according to the method described by Blumenkrantz and Asboe-Hansen (1973) with minor modifications. The produced SDF samples were dissolved in DI water at a concentration of 0.1 mg/g and shaken for 1 h at 55 °C, followed by cooling down to room temperature. 200 μL of each sample was mixed with 1.2 mL sulfuric acid/tetraborate solution (0.0125 M tetraborate in concentrated sulfuric acid) in pre-refrigerated tubes. The mixture was then vortexed and put in a boiling water bath for 5 min followed by being placed into an ice-water bath. Afterward, 20 μL of the m-hydroxydiphenyl reagent was added to the mixture and the tubes were vortexed. Absorbance was measured at 520 nm by a UV/VIS

spectrophotometer (Thermo Fisher Scientific, Hampton, NH, USA) within 5 min, and glucuronic acid was used as a standard to quantify the amount of uronic acid.

Monosaccharide composition

The SDF samples were hydrolyzed by acid based on a method reported by Yan et al. (2019) with minor modifications. To be detailed, 2 mg of sample was mixed with 3 mL 2M TFA and incubated in an oil bath under 110 ℃ for 4 h. Then, the water and TFA in the hydrolysate were evaporated by a rotary evaporator with reduced pressure and followed by washing with 3 mL methanol 4 times. The washed hydrolysate was resuspended in 5 mL DI water for analyzing the concentrations of glucose, xylose, and arabinose using an Agilent 1260 high-performance liquid chromatography (HPLC) system with a refractive index detector (RID) (Agilent Technologies, Santa Clara, CA, USA). Bio-Rad Aminex® HPX-87H column (Bio-Rad Laboratories, Hercules, CA, USA) was used with 0.005 M H_2SO_4 as the mobile phase (0.6 mL/min) at 50 °C and the injection volume of 20 μL (Yu et al., 2022).

Protein content, total phenolic content (TPC), and particle size

The protein content was measured by the Kjeldahl method, followed by multiplying by a factor of 6.25 (AOAC, 2005d). TPC was analyzed by the Folin-Ciocalteu method described by Jin et al. (2019) using gallic acid as the standard. Both gallic acid and SDF solutions (500 μL) were diluted with 900 μL DI water and mixed with 2.5 mL of 0.2 N Folin–Ciocalteu reagent and 2 mL saturated sodium carbonate. The mixtures were incubated at room temperature for 2 h. The absorbance at 765 nm was read by the UV/VIS spectrophotometer and TPC was expressed as mg gallic acid/g SDF. To measure the particle size, SDF samples were mixed with DI water to reach appropriate concentrations and the particle size was analyzed by the above-mentioned laser diffraction particle size analyzer.

5.2.6.3 Structural characteristics

Scanning electron microscopy (SEM)

The morphological properties of the SDF samples were observed by SEM. The SDF samples were fixed into Leica ACE600 Sputter and sputtered with 10 nm Pt/Pd. The SEM images were taken by FEI Quanta 600 FEG (FEI Company, Hillsboro, OR, USA) with an accelerating voltage of 20 kV and an image magnification of 500X.

Fourier-transformed infrared analysis (FTIR)

Chemical bonds of different SDF samples were analyzed by FTIR (Hong et al., 2022). The instrument was calibrated through background scanning. Afterward, samples were directly placed on the ZnSe flat top plate. The FTIR spectrum was analyzed using a 400 FT-IR/FT-NIR Spectrometer (PerkinElmer Corp., Shelton, CT, USA) with a wavenumber range of 400-4000 cm^{-1}, resolution of 4 cm^{-1}, and scan of 32 times.

X-ray diffraction (XRD)

The crystalline structure was measured by the Rigaku MiniFlex II diffractometer (Rigaku Corporation, Tokyo, Japan). SDF samples were scanned at the speed of 1°/min with the diffraction angle (2θ) ranging from 5° to 60° (Liu et al., 2019; Qiao et al., 2021). Both the crystalline and amorphous peak areas were analyzed by OriginPro 9.0 (OriginLab®, Northampton, MA, USA) and the crystallinity index (CI) was calculated based on the following equation:

$$CI(\%) = \frac{A_C}{A_a + A_C} \times 100\% \quad (5.2)$$

where A_C is the crystalline peak area and A_a is the amorphous peak area.

5.2.6.4 Rheological properties

The rheological properties of the SDF solutions were measured using the DHR3 rheometer (TA Instruments, New Castle, NE, USA) equipped with a plate-plate geometry system (parallel plate, 40 mm diameter, 1 mm gap). SDF solution samples were prepared at concentrations of 10 and 40 mg/mL. The steady shear measurements (apparent viscosity vs. shear rate) were measured at the shear rate ranging from 0.1 to 1000 s^{-1} (Yan et al., 2019). In brief, a certain amount of the sample solution that is enough to cover the plate was loaded on the plate and soaked for 120 s. The temperature plate was controlled at 25 ℃ using a water circulating system. Apparent viscosity and shear stress were recorded by the manufacturer-supplied software (Rheology Advantage, TA Instruments, New Castle, NE, USA). The data were fitted into five different rheological models, namely Newtonian, Bingham, Casson, Power Law, and Herschel-Bulkley.

Before the dynamic oscillatory measurements, amplitude sweeps were performed under a constant angular frequency of 1 rad/s and temperature of 25 ℃ to determine the linear viscoelastic region of different SDF samples. The tested shear strain range was from 0.01 to 1000% (Lin & Fernandez-Fraguas, 2019). Under the constant shear strain within the linear viscoelastic region, a frequency sweep was carried out with the angular frequency ranging from 0.1 to 100 rad/s. The storage modulus (G', Pa) and loss modulus (G'', Pa) were recorded respectively.

5.2.6.5 Thermal properties

Differential scanning calorimetry

The thermal properties were measured using a DSC-Q20 Differential Scanning Calorimeter (TA Instruments, New Castle, NE, USA). Around 3 mg of each sample was

measured into an aluminum pan and tightly sealed with a paired aluminum lid. Heating was carried out from 40 ℃ to 300 ℃ with a scan rate of 10 ℃/min under nitrogen flow (Yan et al., 2019). A sealed empty pan was used as a reference.

Thermogravimetric analysis (TGA)

TGA was performed using a Q50 Thermogravimetric Analyzer (TA Instruments, New Castle, NE, USA). Briefly, 3-6 mg of the samples were placed in the crucibles and heated from room temperature to 800 ℃ with a ramping rate of 10 ℃/min under a nitrogen atmosphere (Xue et al., 2019).

5.2.6.6 Functional properties

***In-vitro* fermentation**

Fecal samples were collected from three adult pigs in the Animal and Poultry Sciences Department at Virginia Tech. The pigs were fed with digestible starch and fish meal 10 days before fecal collection. After collection, the fecal samples were immediately transferred to the lab in an anaerobic container inside an iced cooler. After arriving in the lab, all fecal samples were pooled and mixed well with pre-autoclaved glycerol in the anaerobic chamber (Coy Laboratory Products, Inc., Grass Lake, MI, USA) and then split into multiple centrifuge tubes and stored under -80 °C until use. Before use, the frozen fecal sample was fully thawed at room temperature. The fecal slurry preparation and the in vitro fermentation was performed inside the vinyl anaerobic chamber based on the method described by Long et al. (2015). To make 10% (w/v) fecal slurry, the fecal sample was diluted in a sterile PBS medium and then thoroughly mixed. Then, the slurry was filtered via two layers of cheesecloth to remove large particles. The *in-vitro* fermentation started by thoroughly mixing 250 mg of each SDF sample and 20 mL of fecal slurry in 50 mL centrifuge tubes. Control fermentation was conducted by only adding 20

mL of the fecal slurry without SDF samples. All tubes were incubated inside the anaerobic chamber at 37 °C with the cap loose and without stirring.

Fermentation samples (1 mL) were collected in a 1.5-mL centrifuge vial at 24 h and immediately centrifuged under 16,639 × g. The supernatant was filtered using a 0.2 μm syringe filter into HPLC vials. Three short-chain fatty acids (SCFAs) including acetic acid, propionic acid, and butyric acid were measured using the Agilent 1260 HPLC system with a refractive index detector (RID) and UV detector. The instrument setup and column were the same as described in the monosaccharide analysis.

ABTS scavenging abilities

The 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) radical scavenging activity of SDF samples was measured based on the method reported by Re et al. (1999). SDF solution (200 μL) was mixed with 7.6 mL of ABTS solution and reacted for 6 min in the dark cabinet. Absorbance was read at the wavelength of 734 nm using the UV/VIS spectrophotometer and Trolox was used as the standard.

Emulsifying properties

Emulsifying properties were analyzed by the turbidimetric method described by Lin et al. (2020) and Shen et al. (2017). Emulsifying activity index (ESI, m^2/g) was used to evaluate the capacity of SDF to form an emulsion and emulsifying stability index (EAI, min) was used to measure the stability of the emulsion after storing a certain time. In detail, 100 mg of SDF products were dissolved in 20 mL of 10 mM phosphate buffer (pH 7.0). For full solubilization, the mixture was incubated under 55 °C for 1 h with a shaking speed of 143 rpm. Afterward, 3 g of canola oil was added to 9 mL of SDF solution followed by homogenization using the ultrasonic processer (Model 505, Fisher Scientific Inc., Waltham, MA, USA) equipped with a

3.0-mm ultrasound probe. The amplitude was set at 30% with 10 s on/ 10 s off for total processing of 10 min. Then, 20 µL of the emulsion was immediately sampled and diluted with 5 mL of 0.1% sodium dodecyl sulfate solution. Another 20 µL was sampled 10 min after the ultrasound processing and followed the same dilution procedures. Absorbance was measured at the wavelength of 500 nm by the UV/VIS spectrophotometer. The EAI and ESI were calculated by the following equations:

$$EAI\ \left(\frac{m^2}{g}\right) = \frac{A_0 \times 2 \times 2.203 \times DF}{C \times \varphi \times \theta \times 1{,}000} \quad (5.3)$$

$$ESI\ (min) = \frac{A_0}{A_0 - A_{10}} \times 10 \quad (5.4)$$

where A_0 and A_{10} are the absorbance of the emulsions at 0 and 10 min, respectively. DF is the dilution factor (250 in this study), C is the SDF concentration (mg/mL) before forming the emulsion, φ is the optical path (0.01 m) and θ is the ratio of oil fraction to emulsion (0.25).

5.2.7 Statistical analysis

All procedures and analyses were performed in duplicates and the results were expressed as mean ± standard deviation (SD). One-way analysis of variance (ANOVA) with Tukey's HSD (Honestly Significant Difference) test was conducted using the statistical software SPSS (22.0.0.0, IBM Corporation, Armonk, NY, USA) to compare differences among different SDF samples with a significance level of 0.05.

5.3 Results and discussion

5.3.1 Effects of treatment parameters on SDF yield

The application of citric acid treatment greatly enhanced SDF yield. In the control treatment where no citric acid was added, the SDF yield from edamame shell was only 5.6 ± 0.7% (Figure 5.1). The SDF yield significantly increased to 8.1% when 1% citric acid was added at the processing temperature of 90 °C and processing time of 30 min. Citric acid has been

reported by Yan et al. (2019) to increase the SDF production from wheat bran. It is also used to improve the pectin production from apple pomace (Canteri-Schemin et al., 2005). The increase in SDF production by using citric acid is mainly because citric acid has carboxyl (COOH) groups to enhance the ionization of hydrogen ions (H+), thus can promote the solubilization of hemicellulose (Yan et al., 2019). Further increasing the citric acid concentration from 1% to 2% increased the SDF yield by 2.6% with treatment under 90 ℃ and 60 min. However, no significant increase in SDF yield was observed on other treatments at 90 ℃, which could be due to the full depolymerization of hemicellulose to mono sugars, furans, and other undesirable organic acids at higher citric acid concentrations (Chandel et al., 2012; Gomes et al., 2022). Under the same treatment temperature and citric acid concentration, increasing the treatment time from 30 to 60 min did not improve the SDF yield. This indicated that treatment time is not as sensitive as citric acid concentration or processing temperature. Moreover, the increase in processing temperature from 90 to 130 ℃ greatly increased the SDF yield. When the treatment temperature was 130 ℃ with 2% acid concentration and 30 min treatment time, the SDF yield significantly increased by almost two folds to 17.8% compared to the treatments at 90 ℃. Thermal treatment was reported to increase the soluble dietary fiber ratio in barley (Bader Ul Ain et al., 2019), orange peels (Tejada-Ortigoza et al., 2018), and wheat (Căpriţă et al., 2011). Autoclaving treatment was reported to increase the SDF yield from soybean curd residue by 12% under the combined effects of both high temperature and pressure (Li et al., 2019). This could be due to the breakdown of the glycosidic linkages that lead to the partial degradation of cellulose, hemicellulose, or pectic polysaccharides into shorter chain carbohydrates and results in the solubilization of insoluble dietary fiber (Benítez et al. 2011).

Figure 5.1 also shows that the application of ball milling increased the SDF yield, and the increase was significant for the treatment at 130 °C, with 1% acid addition, and 30 min which led to the highest SDF yield (19.5%) among all treatments. Ball milling has been reported to increase the SDF contents in different materials. Bender et al. (2020) reported that the SDF content in grape pomace increased from 4.06 to 13.76 g/100 g DW after ball milling. Ge et al. (2021) found that ball milling can increase the SDF yield (up to 6.03%) from bamboo powder. Ball milling provides various crushing directions that heterogeneously break down the fibers into small pieces by creating the impact and shear forces (Wang et al., 2019; Chen et al., 2020). The loosened structure with a more exposed amorphous area not only promotes the solubilization of fibers but also provides more surface area for acid hydrolysis. In our study, the particle size (D(3,2)) of the edamame powder significantly decreased from 152.2 ± 4.6 μm to 18.7 ± 0.3 μm after ball milling which manifested the increased surface area of the edamame powder.

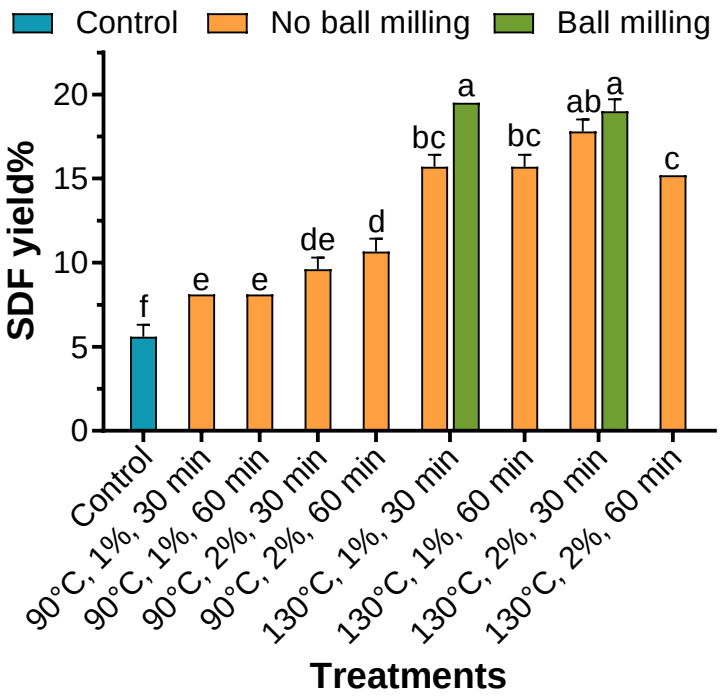

Figure 5.1 Soluble dietary fiber yield of different combinations of ball milling, citric acid concentrations, and thermal treatments.

5.3.2 Physicochemical properties

Different treatments lead to different chemical compositions of the produced SDFs. Table 5.1 shows the uronic acid, mono sugar, protein, and total phenolic contents of SDFs extracted by four different treatments, namely SDF-I, SDF-II, SDF-III, and SDF-IV. All four SDFs contain a high uronic acid content of 45 to 59% which indicates a high pectic polysaccharide content. The high uronic acid content is in accordance with the previous study that soybean hulls were a good source of pectin (Kim et al. 2015), as edamame belongs to the same species as grain soybean. Among the four SDFs, SDF-III has the highest uronic acid content which indicated that a higher citric acid concentration and treatment temperature tend to extract more soluble pectic polysaccharides from edamame shells. Citric acid was reported to be an effective chemical for extracting pectin from different raw materials such as passion fruit peel (Klieman et al., 2009),

apple peel waste (Virk and Sogi, 2004), and yellow passion fruit rind (Yapo, 2009). It provides hydrogen ions (H+) for hydrolyzing the cross-link networks of cell walls, thus releasing the pectin from the matrix (Cui et al., 2021). Vriesmann et al. used hot citric acid to extract pectin from cacao pod husks and their final product contained 54.4 to 68.9 % uronic acid which is consistent with our results. The uronic acid content in SDF-IV (extracted from ball-mill treated edamame shell) significantly decreased to 45.9% while it has the highest mono sugar contents (18.6%) compared to the other three SDFs. This further demonstrated that ball milling can promote the break-down of fiber (especially xylose-rich hemicellulose) into shorter-chain polysaccharides which are soluble in water (Gao et al., 2017). No protein was determined in SDF-I while SDF-II, III and IV had a small amount of protein contents of 1.50, 1.31 and 1.86%, respectively, with no significant difference. This indicates that citric acid promoted the dissolution of protein which resulted in the conjugated proteins in the produced SDFs. All SDF products had a certain amount of phenolic compound as shown in Table 5.1. SDF-IV contains the significantly highest TPC (4.1%) followed by SDF-III of 3.7%, SDF-II of 3.0%, and SDF-I of 3.1%. The relatively high content of TPC can make the produced SDF as a potential source of antioxidant dietary fiber. Many studies have reported that ball milling increased the TPC contents. For example, Chitrakar et al. (2020) revealed that ball milling treatment increased the TPC in asparagus leaves. Bender et al. (2020) reported that TPC in grape pomace increased from 0.22 to 0.91 g gallic acid/g DW after ball milling. This improvement could be due to the breakdown of the lignin during the milling process and the release of the embedded phenolic compounds (Bender et al., 2020).

Particle size has a profound effect on the functionalities (such as gel strength) and the applications of dietary fiber (Huang et al., 2020), thus the particle size of SDFs were measured

using a laser diffraction particle size analyzer. SDF-I had the largest particle size (137.5 μm) followed by SDF-III (34.9 μm), SDF-II (26.5 μm), and SDF-II (17.7 μm). Increased treatment temperature from 90 to 130 °C significantly decreased the particle size of the SDF, which is likely due to the more severe decomposition of fiber at a higher processing temperature. On the other hand, increasing citric acid concentration from 1 to 2% and the application of ball milling increased the particle size slightly, which might be due to the aggregations of small particles.

Table 5.1 Uronic acid, mono sugars, protein, total phenolic content and mean particle size of SDF extracted by different treatments.

Sample	SDF-I	SDF-II	SDF-III	SDF-IV
Uronic acid (% wt)	51.81 ± 0.67 b	54.31 ± 1.08 b	58.57 ± 0.05 a	45.89 ± 0.21 c
Mono sugars (% wt)				
Glucose (Glu)	3.50 ± 0.19 ab	3.02 ± 0.21 b	3.51 ± 0.56 ab	4.41 ± 0.10 a
Xylose (Xyl)	4.53 ± 0.11 c	8.71 ±0.40 b	6.67 ± 0.63 bc	12.28 ± 1.04 a
Arabinose (Ara)	2.22 ± 0.10 a	1.42 ± 0.11 bc	1.03 ± 0.01 c	1.90 ± 0.29 ab
Glu + Xly + Ara	10.25 ± 0.01 c	13.15 ± 0.72 b	11.21 ± 0.07 bc	18.58 ± 0.85 a
Protein (% wt)	0.00 b	1.50 ± 0.38 a	1.31 ± 0.29 a	1.86 ± 0.00 a
Total phenolic content (mg gallic acid/g SDF)	3.05 ± 0.03 c	2.97 ± 0.01 c	3.70 ± 0.04 b	4.08 ± 0.03 a
Mean particle size (μm)	137.48 ± 12.44 a	17.65 ± 10.98 c	34.93 ± 9.27 b	26.52 ± 3.51 b

5.3.3 Morphological and structural properties

5.3.3.1 Scanning electron microscopy (SEM)

The morphology of SDFs was observed by SEM (Figure 5.2). All SDFs had an overall sheet-like structure, but the detailed structure was significantly affected by different extraction methods. The image of SDF-I showed significantly larger particles compared to the other SDF products which is coherent with the particle size results reported in Table 5.1. SDF-I possesses a thin-sheet-like structure and the surface is smooth and intact without pores or cracks. Even with different extents of folding, its structure still stays integrated. SDF-II, SDF-III, and SDF-IV had

smaller pieces and looser structures than SDF-I due to the more severe treatment conditions. SDF-II and SDF-III also had a smooth surface (Figure 5.2) but lower rigidity (easier to be broken down into small pieces) compared to SDF-I. The sheet-like structure of SDF-IV looks thicker and several cracks and a large number of aggregates were observed on its surface. The aggregates were the smaller SDF particles since the ball mill provided powerful mechanical shearing and tearing to break down fiber from more long-chain polysaccharides to shorter-chain soluble polysaccharides (Huang et al., 2021). The same phenomenon was also observed by Liu et al. (2016) on IDF from orange peel. In their study, the surface of regularly milled IDF (using QE-100g high-speed multi-function crusher) was smooth while many small particles were found on the surface of the ball-milled IDF. Additionally, not many foldings of the sheet were observed on SDF-IV and it can be easily broken down into small pieces.

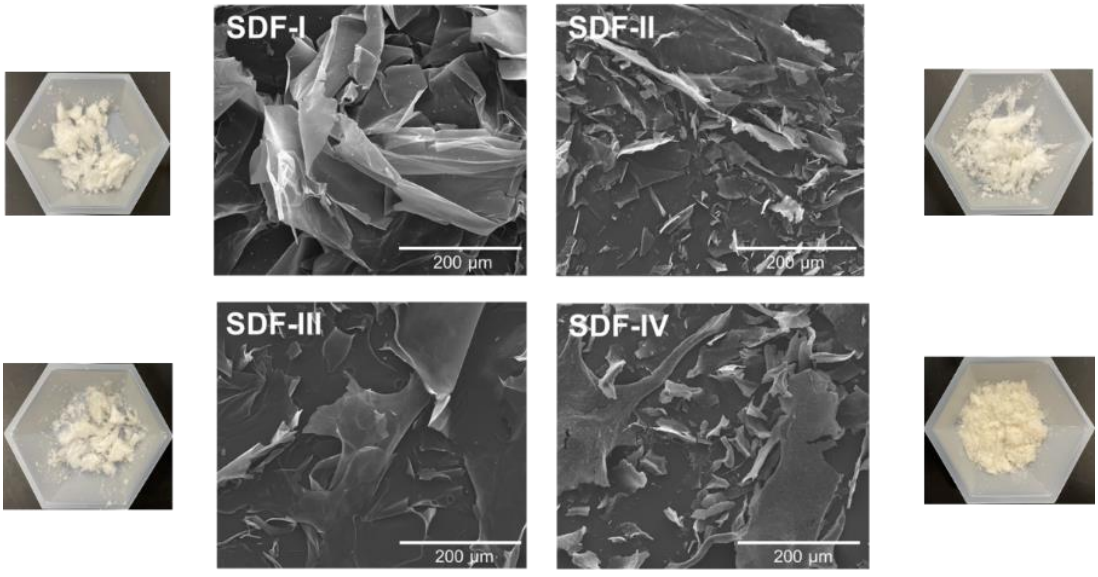

Figure 5.2 Actual photos and scanning electron microscopy images (500X) of SDF-I, SDF-II, SDF-III, and SDF-IV.

5.3.3.2 FTIR

All SDF samples had similar surface functional groups. FTIR spectra of SDFs prepared by four different treatments are shown in Figure 5.3 (a). The broad absorption peak at 3360 cm^{-1} and the small peak at 2923 cm^{-1} was assigned to the stretching vibration of the hydroxyl group (-OH) and methyl and methylene (C-H), respectively. These are the typical functional groups of polysaccharides (Yan et al., 2019; Liu et al., 2016) which indicated the presence of short-chain compounds derived from hemicellulose or cellulose. The absorption peak at 1737 cm^{-1} corresponded to the carbonyl group (C=O) and indicated the existence of uronic acid. Stretching bands at 1608 cm^{-1} revealed the existence of the benzene ring in lignin, which could be associated with phenolic structures (Bender et al., 2019). The stretching vibrations at 1145, 1070, and 1016 cm^{-1} were associated with the C-O of the pyranose ring which evidenced the presence of sugar rings (Bender et al., 2019; Chen et al., 2018). The FTIR spectra of the SDF products in our study are similar to the SDF extracted from other materials, such as wheat bran (Yan et al., 2019), tomato peels (Li et al., 2018), and nodes of lotus root (Chen et al., 2018).

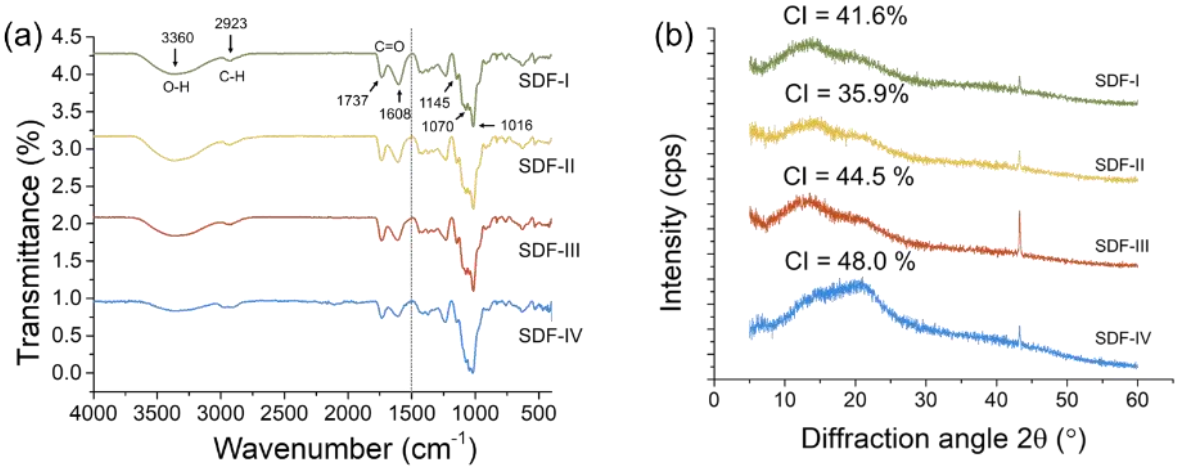

Figure 5.3 (a) FTIR and (b) XRD spectra of SDF-I, SDF-II, SDF-III, and SDF-IV from edamame shell.

5.3.3.3 XRD

XRD was performed to investigate the crystal structure differences among the four SDFs. The X-ray diffractogram (Figure 5.3 (b)) shows that all SDF had broad bands with two characteristic diffraction peaks, one at around 14.0° 2θ and the other one at 20.7° 2θ, suggesting the typical amorphous structure (Jia et al., 2019; Gao et al., 2018). No other peak was observed after treatments indicating that different treatments do not significantly affect the crystalline type. However, the first peak of SDF-III shifted to a smaller angle (13.5° 2θ) while SDF-IV had a stronger peak at 20.7° 2θ compared to SDF-I and SDF-II. This indicated that higher citric acid concentrations and the application of ball milling changed the crystalline structure of the SDFs. This change could be due to chemical composition changes of the SDFs extracted by different treatments. SDF-I, II and III had significantly higher uronic contents which indicated potentially higher pectin contents. Their XRD patterns are similar to the XRD pattern of pectin reported by Nesic et al. (2014). SDF-IV had lower pectin content but higher mono-sugar content which might come from the hydrolyzed hemicellulose. Accordingly, the XRD pattern of SDF-IV is consistent with the XRD pattern of hemicellulose as shown in the study of Wang and Liang (2017). Increasing the treatment temperature from 90 to 130°C decreased the crystalline index of SDF from 41.6% to 35.9%, suggesting the elevated temperature can partially damage the crystal structure of SDF. However, the crystalline index increased to 44.5 and 48.0%, respectively for SDF-III and SDF-IV. These results indicated that higher citric acid concentration and ball milling treatment didn't destroy the crystalline region but resulted in a more ordered and regular crystal structure of SDF.

5.3.4 Thermal properties (DSC and TGA)

Thermal treatments such as pasteurization and sterilization are the most crucial steps in food processing for ensuring safety and extending shelf life, producing desirable flavor, and increasing texture (Yousefi and Abbasi, 2022), and they are inevitable during food processing. Therefore, thermal stability of SDF is an important property for SDF to be used as a food ingredient (Wang et al., 2021). DSC was performed to provide thermodynamic properties of the SDF products and the curves between 40 and 300°C are shown in Figure 5.4 (a). An endothermic and an exothermic peak were observed on the DSC curves of all SDF products. Similar DSC curves were also observed on SDF produced from corn bran (Li et al., 2022). All of the SDFs in our study had similar endothermic temperature peaks occurring at around 112 °C which attributed to the water loss, hydrogen bonding, and conformational change of polysaccharides (Qiao et al., 2020; Li et al., 2022). Differences were observed in the exothermic peaks which correspond to the pyrolysis of polysaccharides through thermal and oxidative degradation (Zhang et al., 2017; Li et al., 2022). SDF-I, SDF-II, and SDF-III had similar exothermic temperature peaks at around 260 °C while SDF-IV had a higher exothermic peak at 287 °C. This upshift of the peak temperature indicated that SDF-IV had better thermal stability compared to other SDFs.

Figure 5.4 (b) shows the TGA curve of four different SDF products and provides the weight loss and degradation information within a chosen temperature range. Overall, all SDF products showed a similar three-stage weight loss when the temperature increased from room temperature to 800 °C. The first stage was a slight but fast weight decrease observed between room temperature and 110 °C which was mainly attributed to the loss of free water (Gu et al., 2020). A rapid weight loss was found between 208 and 345 °C which was due to the degradation

of dietary fiber through dehydroxylation, deoxylation, or decarboxylation (Xue et al., 2019). The temperature range of the decomposition of polysaccharides in TGA matched that in DSC. The slower weight loss observed between 345 and 800 °C was caused by the thermal decomposition of char (Gu et al., 2020). The residual mass for SDF-I, SDF-II, SDF-III, and SDF-IV were 14.7, 15.4, 14.8, and 16.2%, respectively. The application of ball milling increased the thermal stability of the final SDF product slightly. The potential reason could be that the application of ball milling increased the crystallinity of the SDF. The crystallized zone has more intramolecular and intermolecular hydrogen bonds which contribute to stability (Pace et al., 2014; Liu et al., 2021). Both DSC and TGA results demonstrated that ball milling improved the thermal stability of the final SDF products and makes SDF-IV a better potential candidate for thermally processed foods as a functional food additive. For example, it can be used as a dietary fiber supplement in bread, juices, nutrition shakes, etc.

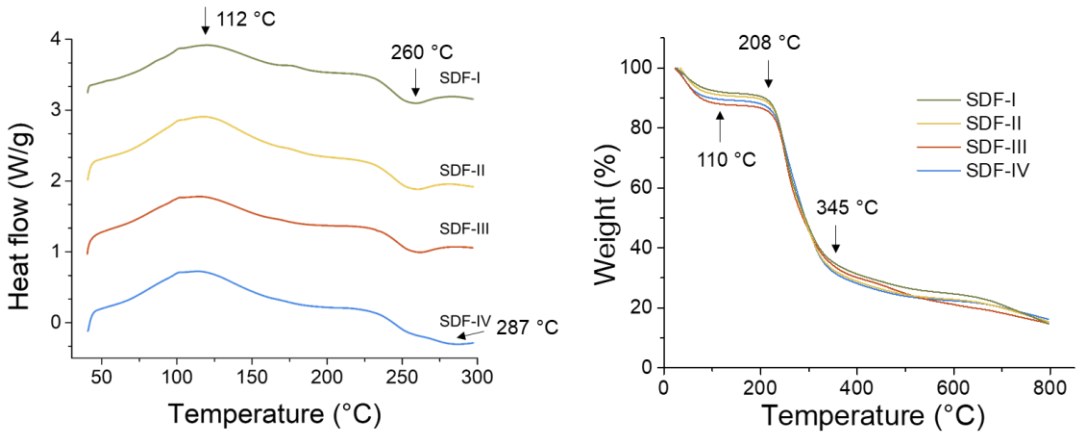

Figure 5.4 (a) DSC and (b) TGA curves of SDF-I, SDF-II, SDF-III, and SDF-IV from edamame shell.

5.3.5 Rheological properties

5.3.5.1 Dependence of apparent viscosity on shear rate

The rheological behavior of food ingredients is crucial because it will decide the final application of SDF and will further affect the food processing, food texture, and final consumer acceptability. The change of apparent viscosity with the shear rate of SDF solutions at 10 and 40 mg/mL concentrations is shown in Figure 5.5 (a) and (b), respectively. Overall, the data were fitted into five different rheological models including Casson, Newtonian, Bingham, Power Law, and Herschel-Bulkley. Herschel-Bulkley displayed the best fitting with the highest R^2 which demonstrated that all samples are non-Newtonian fluid. To be detailed, the apparent viscosity of all SDF solutions decreased sharply with the increased shear rate in the low shear rate range (0-400 s^{-1}). This indicates that all SDF possessed a shear-thinning behavior when dissolved in water at both low and high concentrations (10 and 40 mg/mL). At the high shear rate, the apparent viscosity became stable gradually and showed the Newtonian flow behavior. A similar flow behavior was also found in SDF extracted from wheat bran (Yan et al., 2019) and tomato peels (Li et al., 2018). Even though the overall trends were similar, the apparent viscosity of 40 mg/mL SDF solutions was higher than 10 mg/mL SDF samples. The reason might be because, under high concentrations, stronger intra-molecular and inter-molecular interactions provide more molecular chain entanglement and aggregates, thus increasing the apparent viscosity (Yan et al., 2019). Additionally, for both concentrations, the apparent viscosity of SDF-I decreased more slowly compared to SDF-II, SDF-III, and SDF-IV. The flow behaviors were similar for 10 mg/mL SDF-II, III, and IV solutions. However, differences in the shear thinning decrease were observed when SDF was dissolved at a higher concentration of 40 mg/mL. Figure 5.5 (b) showed that SDF-III had the second slow decrease compared to SDF-I, and then followed by SDF-II and

SDF-IV. It is worth noting that the shear thinning behavior of 40 mg/mL SDF solutions was consistent with the particle size in Table 5.5. SDF-I had a significantly larger particle size followed by SDF-III, then SDF-II and SDF-IV had similar particle sizes with no significant difference. Compared to a longer chain, the entanglement formed within the shorter chains was easier to be disrupted by the increased shear rate (Yan et al., 2019). In summary, these results demonstrated that different acid and thermal treatments and ball milling showed observable effects on the rheological properties of the SDF extracted from edamame shells.

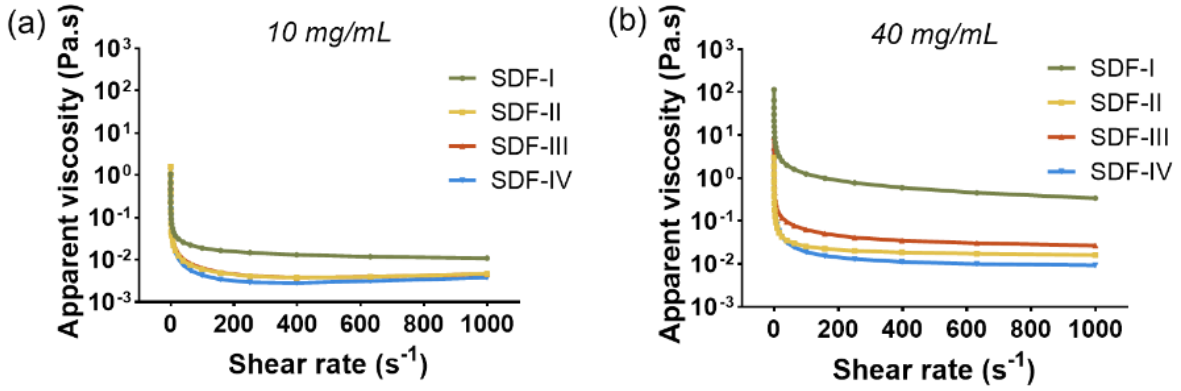

Figure 5.5 Dependence of apparent viscosity on the shear rate for SDF-I, SDF-II, SDF-III, and SDF-IV at different concentrations: (a) 10 mg/mL and (b) 40 mg/mL.

5.3.5.2 Viscoelastic behavior - dependence of G' and G'' on angular frequency (ω)

The trends of storage modulus (G') and loss modulus (G'') with increased angular frequency (ω) were plotted in Figure 5.6 (a) – (d). Except for the low-concentration solution of SDF-I, all other samples possessed a predominant G' which was always above G'' throughout the tested ω range. Both G' and G'' increased with increased ω but no crossover occurred in the tested range, indicating that all SDF samples possessed a typical gel-like behavior (Niu et al., 2018). For the low-concentration solution of SDF-I, G'' became higher than G' under higher ω

which means that the sample transitioned from gel-like to fluid-like behavior. The high-concentration SDF-I solution had a much higher G' and G'' compared to the other SDF solutions. This might be attributed to the larger particle size of SDF-I and high concentrations, making it easier to form stronger intra- and inter-molecular interactions (Yan et al., 2019). However, overall, different treatments didn't profoundly affect the dynamic behavior of SDF-II, SDF-III, and SDF-IV.

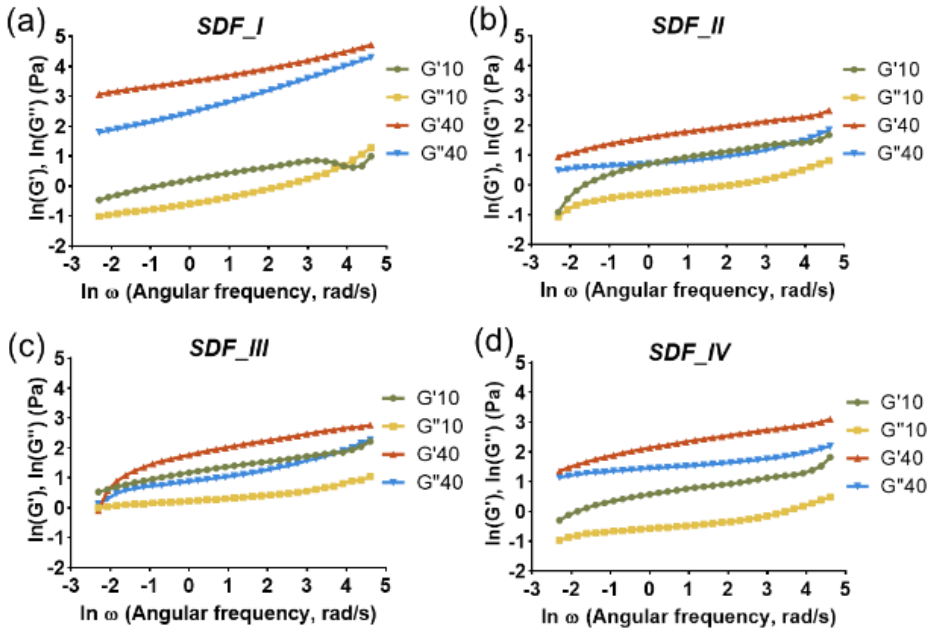

Figure 5.6 Dependence of storage modulus (G') and loss modulus (G'') on the angular frequency (ω) for (a) SDF-I, (b) SDF-II, (c) SDF-III, and (d) SDF-IV at different concentrations: 10 mg/mL and 40 mg/mL.

5.3.6 Functional properties

5.3.6.1 SCFA formation during *in vitro* fermentation

Fermentation of dietary fiber by the microbiome in the gut results in the generation of SCFAs, which have a considerable beneficial impact on numerous physiological processes, including improved gut barrier integrity and regulation of the immune system. Therefore, the four SDF products were subjected to *in vitro* fermentations for SCFA production, and the results are shown in Figure 5.7 (a) to (d). Acetic acid, propionic acid, and butyric acid were the three main acids produced during fermentation, thus the total SCFAs were calculated by adding all three acids together. After 24-hr fermentation, the samples supplemented with SDF-I, II, III, and IV produced 5.55 ± 0.60, 5.31 ± 0.06, 5.42 ± 0.15, and 5.26 ± 0.03 mg/mL of the total SCFAs, respectively with no significant difference observed. Compared to the control, all SDF products significantly elevated the total SCFAs production which indicated their potential to promote gut health (Silva et al., 2020). To look at individual acid, SDF-I produced the highest acetic and butyric acid while SDF-IV produced the highest propionic acid. This might be due to the different chemical compositions of SDF products. Bai et al. (2021) conducted *in vitro* fermentation using different soluble dietary fiber including xylan, xylooligosaccharide, citrus pectin, resistant starch, inulin, and oat β-glucan as supplements. They found that different soluble dietary fibers led to different metabolic profiles in SCFAs. In our study, SDF containing less uronic acid led to lower acetic and higher propionic acid. In summary, different extraction methods in our study led to slightly different SCFAs production in *in vitro* fermentation. In the future, it would be necessary to further investigate the dietary fiber effect on the composition of the microbial community in the gut.

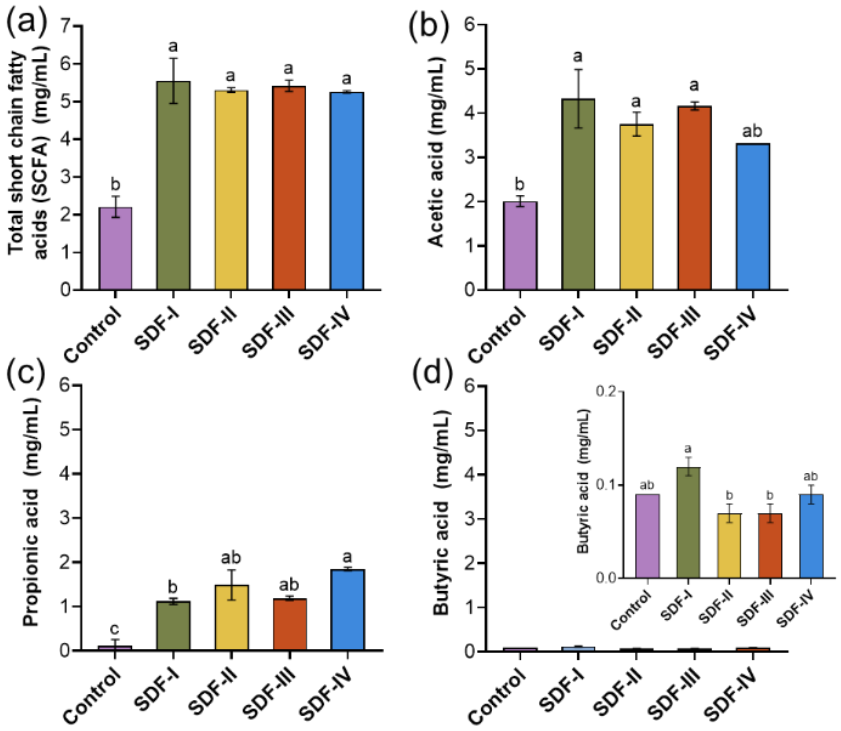

Figure 5.7 SCFAs (total SCFA, acetic acid, propionic acid, and butyric acid) profiles produced during *in vitro* fermentation of (a) SDF-I, (b) SDF-II, (c) SDF-III and (d) SDF-IV.

5.3.6.2 ABTS free radical scavenging capacity of SDF

ABTS is a common free radical cation that can be used to do the antioxidant capacity test (Chen et al., 2019) and the ABTS scavenging activity of different SDF are shown in Figure 5.8 (a). SDF-IV showed the highest ABTS scavenging activity of 3.78 ± 0.41 µmol Trolox/ g SDF followed by SDF-III of 3.00 ± 1.23 µmol Trolox/ g SDF, SDF-I of 2.52 ± 0.55 µmol Trolox/ g SDF and SDF-II of 2.23 ± 0.14 µmol Trolox/ g SDF. However, the difference was not significant. It is worth noting that the trend of the ABTS scavenging activity was consistent with the TPC as shown in Table 5.1 which indicated the TPC is the main contributor to the antioxidant activity of the SDF products. These results showed that different treatments had effects on the ABTS scavenging activity, but it is not significant.

5.3.6.3 Emulsifying properties

Emulsifying properties evaluated the ability of all SDF samples to promote the solubilization of two immiscible liquids and the results were shown in Figure 5.8 (b). The overall trend of EAI was the same as ESI which indicated the sample having higher EAI tended to have higher ESI. The results showed that SDF-III had the highest EAI of 87.8 m^2/g and ESI of 72.1 m^2/g followed by SDF-II and SDF-I. SDF-IV had significantly lower EAI and ESI compared to the other SDF product. The trends of both EAI and ESI were consistent with the uronic acid (pectin) contents of SDF as shown in Table 5.1. Pectin has been reported to have superior emulsifying properties, such as potato pectin (Yang et al., 2018), persimmon peel pectin (Jiang et al., 2020), and sugar beet pump pectin (Ma et al., 2013). Due to the high pectin contents of SDF products in this study, pectin might be the main contributor to the difference in the emulsifying properties, and different treatments led to these observed differences.

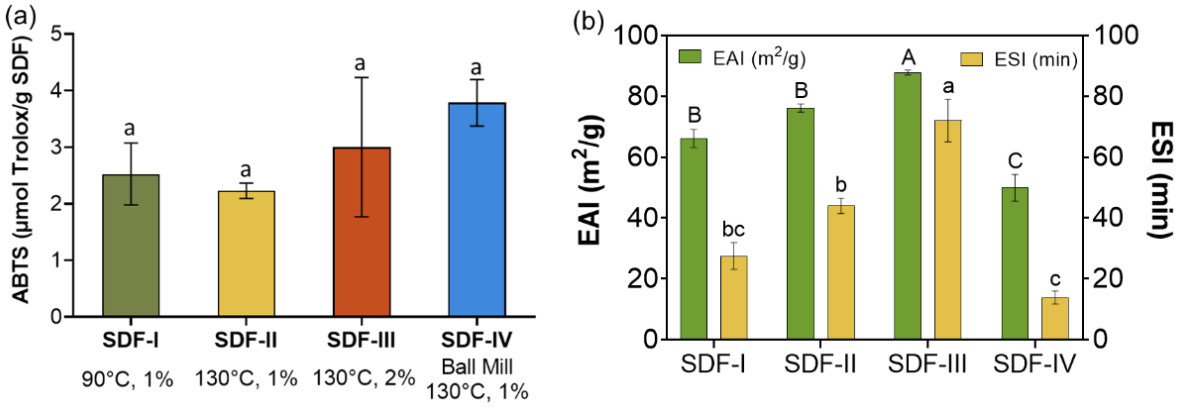

Figure 5.8 (a) ABTS scavenging properties and (b) emulsifying activity index and emulsifying stability index for SDF-I, SDF-II, SDF-III, and SDF-IV.

5.4 Conclusions

The present study developed a ball milling-assisted chemical processing to produce SDF from edamame shells. The effects of key processing conditions, i.e., citric acid concentration,

processing temperature, processing duration, and ball milling, on the yield and properties of SDF from edamame shells were thoroughly investigated. The application of ball milling helps achieve the optimal yield and the produced SDF-IV had the lowest pectin and the highest short-chain carbohydrate and total phenolic contents. The combined citric acid and ball milling methods changed the particle size, morphology, and crystalline properties of the produced SDF. Additionally, the thermal stability of the produced SDF was improved by ball milling and the shearing thinning behavior was also enhanced, making the SDF easier to be applied in food processing. SDF-IV had a competitive *in vitro* fermentation performance and antioxidant properties which can potentially provide health benefits to the human body and contribute to the protection of functional components in foods. In summary, ball milling-assisted citric acid processing could be used as an environmentally friendly method to produce SDF from edamame shells. In addition, the produced SDF by this method can be considered a promising and novel food ingredient that might be applied to foods. In the future, comprehensive investigations on the application of SDF in real food systems need to be conducted in the future. Moreover, a rigorous techno-economic analysis will be conducted to evaluate the economic feasibility of producing SDF from edamame shells at a commercial scale.

5.5 References

Artiss, J.D., Brogan, K., Brucal, M., Moghaddam, M. and Jen, K.L.C., 2006. The effects of a new soluble dietary fiber on weight gain and selected blood parameters in rats. Metabolism, 55(2), pp.195-202.

Bader Ul Ain, H., Saeed, F., Ahmed, A., Asif Khan, M., Niaz, B. and Tufail, T., 2019. Improving the physicochemical properties of partially enhanced soluble dietary fiber through innovative techniques: A coherent review. Journal of Food Processing and Preservation, 43(4), p.e13917.

Bader Ul Ain, H., Saeed, F., Khan, M.A., Niaz, B., Rohi, M., Nasir, M.A., Tufail, T., Anbreen, F. and Anjum, F.M., 2019. Modification of barley dietary fiber through thermal treatments. Food Science & Nutrition, 7(5), pp.1816-1820.

Bai, J., Li, Y., Li, T., Zhang, W., Fan, M., Zhang, K., Qian, H., Zhang, H., Qi, X. and Wang, L., 2021. Comparison of different soluble dietary fibers during the *in vitro* fermentation process. Journal of Agricultural and Food Chemistry, 69(26), pp.7446-7457.

Behera, B.C., Mishra, R. and Mohapatra, S., 2021. Microbial citric acid: Production, properties, application, and future perspectives. Food Frontiers, 2(1), pp.62-76.

Bender, A.B.B., Speroni, C.S., Moro, K.I.B., Morisso, F.D.P., dos Santos, D.R., da Silva, L.P. and Penna, N.G., 2020. Effects of micronization on dietary fiber composition, physicochemical properties, phenolic compounds, and antioxidant capacity of grape pomace and its dietary fiber concentrate. LWT, 117, p.108652.

Benítez, V., Mollá, E., Martín-Cabrejas, M.A., Aguilera, Y., López-Andréu, F.J. and Esteban, R.M., 2011. Effect of sterilisation on dietary fibre and physicochemical properties of onion by-products. Food Chemistry, 127(2), pp.501-507.

Blumenkrantz, N. and Asboe-Hansen, G., 1973. New method for quantitative determination of uronic acids. Analytical Biochemistry, 54(2), pp.484-489.

Canteri-Schemin, M.H., Fertonani, H.C.R., Waszczynskyj, N. and Wosiacki, G., 2005. Extraction of pectin from apple pomace. Brazilian Archives of Biology and Technology, 48, pp.259-266.

Căpriţă, A., Căpriţă, R., Simulescu, V.O. and Drehe, R.M., 2011. The effect of temperature on soluble dietary fiber fraction in cereals. Journal of Agroalimentary Processes and Technologies, 17, pp.214-217.

Chandel, A.K., Antunes, F.A., Arruda, P.V.D., Milessi, T.S., Silva, S.S.D. and Almeida Felipe, M.D.G.D., 2012. Dilute acid hydrolysis of agro-residues for the depolymerization of hemicellulose: State-of-the-art. d-Xylitol, pp.39-61.

Chen, H., Zhao, C., Li, J., Hussain, S., Yan, S. and Wang, Q., 2018. Effects of extrusion on structural and physicochemical properties of soluble dietary fiber from nodes of lotus root. Lwt, 93, pp.204-211.

Chen, P.C., Lin, C., Chen, M.H. and Chiang, P.Y., 2020. The micronization process for improving the dietary value of okara (soybean residue) by planetary ball milling. LWT, 132, p.109848.

Chen, X., Huang, C., Jian, R., Zhang, Y., Wu, W., Tang, X., Tan, Q., Sheng, Z. and Zhang, K., 2019. Optimization of microwave‐assisted extraction of soluble dietary fiber from potato pulp and its properties. Journal of Food Processing and Preservation, 43(10), p.e14204.

Chitrakar, B., Zhang, M., Zhang, X. and Devahastin, S., 2020. Bioactive dietary Fiber powder from asparagus leaf by-product: Effect of low-temperature ball milling on physico-chemical, functional and microstructural characteristics. Powder Technology, 366, pp.275-282.

Cui, J., Zhao, C., Feng, L., Han, Y., Du, H., Xiao, H. and Zheng, J., 2021. Pectins from fruits: Relationships between extraction methods, structural characteristics, and functional properties. Trends in Food Science & Technology, 110, pp.39-54.

Dai, F.J. and Chau, C.F., 2017. Classification and regulatory perspectives of dietary fiber. Journal of Food and Drug Analysis, 25(1), pp.37-42.

Essa, R.Y. and Elsebaie, E.M., 2022. New fat replacement agent comprised of gelatin and soluble dietary fibers derived from date seed powder in beef burger preparation. LWT, 156, p.113051.

Gao, A., Dong, S., Wang, X., Li, S. and Chen, Y., 2018. Preparation, characterization and calcium release evaluation in vitro of casein phosphopeptides-soluble dietary fibers copolymers as calcium delivery system. Food Chemistry, 245, pp.262-269.

Gao, C., Xiao, W., Ji, G., Zhang, Y., Cao, Y. and Han, L., 2017. Regularity and mechanism of wheat straw properties change in ball milling process at cellular scale. Bioresource Technology, 241, pp.214-219.

Ge, Q., Li, P., Li, C., Wu, P.T., Li, H.Q., Xiao, Z.Q. and Mao, J.W., 2021. Optimization of soluble dietary fiber from ultrafine bamboo powder by ball milling and adsorption capacity of heavy metals. Journal of Biobased Materials and Bioenergy, 15(2), pp.244-252.

Gill, S.K., Rossi, M., Bajka, B. and Whelan, K., 2021. Dietary fibre in gastrointestinal health and disease. Nature Reviews Gastroenterology & Hepatology, 18(2), pp.101-116.

Gomes, M.G., de Oliveira Paranhos, A.G., Camargos, A.B., Baêta, B.E.L., Baffi, M.A., Gurgel, L.V.A. and Pasquini, D., 2022. Pretreatment of sugarcane bagasse with dilute citric acid and enzymatic hydrolysis: Use of black liquor and solid fraction for biogas production. Renewable Energy, 191, pp.428-438.

Gu, M., Fang, H., Gao, Y., Su, T., Niu, Y. and Yu, L.L., 2020. Characterization of enzymatic modified soluble dietary fiber from tomato peels with high release of lycopene. Food Hydrocolloids, 99, p.105321.

Guan, Z.W., Yu, E.Z. and Feng, Q., 2021. Soluble dietary fiber, one of the most important nutrients for the gut microbiota. Molecules, 26(22), p.6802.

Haskell, W.L., Spiller, G.A., Jensen, C.D., Ellis, B.K. and Gates, J.E., 1992. Role of water-soluble dietary fiber in the management of elevated plasma cholesterol in healthy subjects. The American Journal of Cardiology, 69(5), pp.433-439.

He, Y., Wang, B., Wen, L., Wang, F., Yu, H., Chen, D., Su, X. and Zhang, C., 2022. Effects of dietary fiber on human health. Food Science and Human Wellness, 11(1), pp.1-10.

Hong, S., Dia, V.P., Baek, S.J. and Zhong, Q., 2022. Nanoencapsulation of apigenin with whey protein isolate: Physicochemical properties, in vitro activity against colorectal cancer cells, and bioavailability. LWT, 154, p.112751.

Huang, J.Y., Liao, J.S., Qi, J.R., Jiang, W.X. and Yang, X.Q., 2021. Structural and physicochemical properties of pectin-rich dietary fiber prepared from citrus peel. Food Hydrocolloids, 110, p.106140.

Jia, M., Chen, J., Liu, X., Xie, M., Nie, S., Chen, Y., Xie, J. and Yu, Q., 2019. Structural characteristics and functional properties of soluble dietary fiber from defatted rice bran obtained through *Trichoderma viride* fermentation. Food Hydrocolloids, 94, pp.468-474.

Jiang, Y., Xu, Y., Li, F., Li, D. and Huang, Q., 2020. Pectin extracted from persimmon peel: A physicochemical characterization and emulsifying properties evaluation. Food Hydrocolloids, 101, p.105561.

Jiang, Z., Mu, S., Ma, C., Liu, Y., Ma, Y., Zhang, M., Li, H., Liu, X., Hou, J. and Tian, B., 2022. Consequences of ball milling combined with high-pressure homogenization on structure, physicochemical and rheological properties of citrus fiber. Food Hydrocolloids, 127, p.107515.

Jin, Q., O'Hair, J., Stewart, A.C., O'Keefe, S.F., Neilson, A.P., Kim, Y.T., McGuire, M., Lee, A., Wilder, G. and Huang, H., 2019. Compositional characterization of different industrial white and red grape pomaces in Virginia and the potential valorization of the major components. Foods, 8(12), p.667.

Kim, H.W., Lee, Y.J. and Kim, Y.H.B., 2015. Efficacy of pectin and insoluble fiber extracted from soy hulls as a functional non-meat ingredient. LWT-Food Science and Technology, 64(2), pp.1071-1077.

Kliemann, E., De Simas, K.N., Amante, E.R., Prudêncio, E.S., Teófilo, R.F., Ferreira, M.M. and Amboni, R.D., 2009. Optimisation of pectin acid extraction from passion fruit peel (*Passiflora edulis flavicarpa*) using response surface methodology. International Journal of Food Science & Technology, 44(3), pp.476-483.

Lapasin, R., 2012. Rheology of industrial polysaccharides: Theory and applications. Springer Science & Business Media.

Liimatainen, H., Sirviö, J., Haapala, A., Hormi, O. and Niinimäki, J., 2011. Characterization of highly accessible cellulose microfibers generated by wet stirred media milling. Carbohydrate Polymers, 83(4), pp.2005-2010.

Li, N., Feng, Z., Niu, Y. and Yu, L., 2018. Structural, rheological and functional properties of modified soluble dietary fiber from tomato peels. Food Hydrocolloids, 77, pp.557-565.

Li, S., Chen, G., Qiang, S., Tang, D., Chen, Y., Zhang, Z., Lei, Z. and Chen, Y., 2019. Intensifying soluble dietary fiber production and properties of soybean curd residue via autoclaving treatment. Bioresource Technology Reports, 7, p.100203.

Li, S., Hu, N., Zhu, J., Zheng, M., Liu, H. and Liu, J., 2022. Influence of modification methods on physicochemical and structural properties of soluble dietary fiber from corn bran. Food Chemistry: X, 14, p.100298.

Liang, T. and Wang, L., 2017. Thermal treatment of poplar hemicelluloses at 180 to 220 °C under nitrogen atmosphere. BioResources, 12(1), pp.1128-1135.

Lin, T. and Fernández-Fraguas, C., 2020. Effect of thermal and high-pressure processing on the thermo-rheological and functional properties of common bean (*Phaseolus vulgaris* L.) flours. LWT, 127, p.109325.

Liu, M., Zhou, S., Li, Y., Tian, J. and Zhang, C., 2021. Structure, physicochemical properties and effects on nutrients digestion of modified soluble dietary fiber extracted from sweet potato residue. Food Research International, 150, p.110761.

Liu, X., Liu, S., Xi, H., Xu, J., Deng, D. and Huang, G., 2019. Effects of soluble dietary fiber on the crystallinity, pasting, rheological, and morphological properties of corn resistant starch. LWT, 111, pp.632-639.

Liu, Y., Wang, L., Liu, F. and Pan, S., 2016. Effect of grinding methods on structural, physicochemical, and functional properties of insoluble dietary fiber from orange peel. International Journal of Polymer Science.

Liu, Y., Zhang, H., Yi, C., Quan, K. and Lin, B., 2021. Chemical composition, structure, physicochemical and functional properties of rice bran dietary fiber modified by cellulase treatment. Food Chemistry, 342, p.128352.

Long, W., Xue, Z., Zhang, Q., Feng, Z., Bridgewater, L., Wang, L., Zhao, L. and Pang, X., 2015. Differential responses of gut microbiota to the same prebiotic formula in oligotrophic and eutrophic batch fermentation systems. Scientific Reports, 5(1), pp.1-11.

Ma, S., Yu, S.J., Zheng, X.L., Wang, X.X., Bao, Q.D. and Guo, X.M., 2013. Extraction, characterization and spontaneous emulsifying properties of pectin from sugar beet pulp. Carbohydrate Polymers, 98(1), pp.750-753.

Nesic, A.R., Kokunesoski, M.J., Ilic, S.M., Gordic, M.V., Ostojic, S.B., Micic, D.M. and Velickovic, S.J., 2014. Biocomposite membranes of highly methylated pectin and mesoporous silica SBA-15. Composites Part B: Engineering, 64, pp.162-167.

Niu, Y., Li, N., Xia, Q., Hou, Y. and Xu, G., 2018. Comparisons of three modifications on structural, rheological and functional properties of soluble dietary fibers from tomato peels. LWT, 88, pp.56-63.

Pace, C.N., Fu, H., Lee Fryar, K., Landua, J., Trevino, S.R., Schell, D., Thurlkill, R.L., Imura, S., Scholtz, J.M., Gajiwala, K. and Sevcik, J., 2014. Contribution of hydrogen bonds to protein stability. Protein Science, 23(5), pp.652-661.

Qiao, H., Shao, H., Zheng, X., Liu, J., Liu, J., Huang, J., Zhang, C., Liu, Z., Wang, J. and Guan, W., 2021. Modification of sweet potato (*Ipomoea batatas* Lam.) residues soluble dietary fiber following twin-screw extrusion. Food Chemistry, 335, p.127522.

Re, R., Pellegrini, N., Proteggente, A., Pannala, A., Yang, M. and Rice-Evans, C., 1999. Antioxidant activity applying an improved ABTS radical cation decolorization assay. Free Radical Biology and Medicine, 26(9-10), pp.1231-1237.

Shen, X., Fang, T., Gao, F. and Guo, M., 2017. Effects of ultrasound treatment on physicochemical and emulsifying properties of whey proteins pre-and post-thermal aggregation. Food Hydrocolloids, 63, pp.668-676.

Silva, Y.P., Bernardi, A. and Frozza, R.L., 2020. The role of short-chain fatty acids from gut microbiota in gut-brain communication. Frontiers in Endocrinology, 11, p.25.

Song, L.W., Qi, J.R., Liao, J.S. and Yang, X.Q., 2021. Enzymatic and enzyme-physical modification of citrus fiber by xylanase and planetary ball milling treatment. Food Hydrocolloids, 121, p.107015.

Tejada-Ortigoza, V., Garcia-Amezquita, L.E., Serna-Saldívar, S.O., Martín-Belloso, O. and Welti-Chanes, J., 2018. High hydrostatic pressure and mild heat treatments for the modification of orange peel dietary fiber: Effects on hygroscopic properties and functionality. Food and Bioprocess Technology, 11(1), pp.110-121.

Virk, B.S. and Sogi, D.S., 2004. Extraction and characterization of pectin from apple (*Malus Pumila*. Cv Amri) peel waste. International Journal of Food Properties, 7(3), pp.693-703.

Vriesmann, L.C., Teofilo, R.F. and de Oliveira Petkowicz, C.L., 2012. Extraction and characterization of pectin from cacao pod husks (Theobroma cacao L.) with citric acid. LWT, 49(1), pp.108-116.

Wang, G., Zhong, D., Liu, H., Yang, T., Liang, Q., Wang, J., Zhang, R. and Zhang, Y., 2021. Water soluble dietary fiber from walnut meal as a prebiotic in preventing metabolic syndrome. Journal of Functional Foods, 78, p.104358.

Wang, H., Sun, F., Qu, Z., Wang, K., Wang, L., Pi, X., Gao, J. and Zhao, G., 2019. Oxygen functional group modification of cellulose-derived hard carbon for enhanced sodium ion storage. ACS Sustainable Chemistry & Engineering, 7(22), pp.18554-18565.

Wang, K., Li, M., Wang, Y., Liu, Z. and Ni, Y., 2021. Effects of extraction methods on the structural characteristics and functional properties of dietary fiber extracted from kiwifruit (*Actinidia deliciosa*). Food Hydrocolloids, 110, p.106162.

Wen, Y., Niu, M., Zhang, B., Zhao, S. and Xiong, S., 2017. Structural characteristics and functional properties of rice bran dietary fiber modified by enzymatic and enzyme-micronization treatments. LWT, 75, pp.344-351.

Wu, L.G., Wang, A.N., Li, X.L. and Qu, L.B., 2014. Effect of extruded corn pericarp dietary fiber on dough rheology and dumpling wrapper quality. Food Science and Technology Research, 20(2), pp.235-240.

Xue, Z., Chen, Y., Jia, Y., Wang, Y., Lu, Y., Chen, H. and Zhang, M., 2019. Structure, thermal and rheological properties of different soluble dietary fiber fractions from mushroom *Lentinula edodes* (Berk.) Pegler residues. Food Hydrocolloids, 95, pp.10-18.

Yang, B., Wu, Q., Song, X., Yang, Q. and Kan, J., 2019. Physicochemical properties and bioactive function of Japanese grape (*Hovenia dulci*s) pomace insoluble dietary fibre modified by ball milling and complex enzyme treatment. International Journal of Food Science & Technology, 54(7), pp.2363-2373.

Yang, J.S., Mu, T.H. and Ma, M.M., 2018. Extraction, structure, and emulsifying properties of pectin from potato pulp. Food Chemistry, 244, pp.197-205.

Yapo, B.M., 2009. Biochemical characteristics and gelling capacity of pectin from yellow passion fruit rind as affected by acid extractant nature. Journal of Agricultural and Food Chemistry, 57(4), pp.1572-1578.

Yan, J.K., Wu, L.X., Cai, W.D., Xiao, G.S., Duan, Y. and Zhang, H., 2019. Subcritical water extraction-based methods affect the physicochemical and functional properties of soluble dietary fibers from wheat bran. Food Chemistry, 298, p.124987.

Yousefi, N. and Abbasi, S., 2022. Food proteins: Solubility & thermal stability improvement techniques. Food Chemistry Advances, p.100090.

Yu, D., Lin, T., Sutton, K., Lord, N., Carneiro, R., Jin, Q., Zhang, B., Kuhar, T., Rideout, S., Ross, J. and Duncan, S., 2021. Chemical compositions of edamame genotypes grown in different locations in the US. Frontiers in Sustainable Food Systems, 5, p.620426.

Yu, D., Sun, Y., Wang, W., O'Keefe, S.F., Neilson, A.P., Feng, H., Wang, Z. and Huang, H., 2020. Recovery of protein hydrolysates from brewer's spent grain using enzyme and ultrasonication. International Journal of Food Science & Technology, 55(1), pp.357-368.

Zhang, W., Zeng, G., Pan, Y., Chen, W., Huang, W., Chen, H. and Li, Y., 2017. Properties of soluble dietary fiber-polysaccharide from papaya peel obtained through alkaline or ultrasound-assisted alkaline extraction. Carbohydrate Polymers, 172, pp.102-112.

Chapter 6

Conclusions and Future Works

6.1 Conclusions

This study aimed to thoroughly understand how genotype, planting location and harvest time affect the edamame chemical composition and use the obtained information as guidance for edamame breeding and harvesting. This research also aimed to utilize edamame shells to generate a value-added product, i.e., soluble dietary fiber, using combined physical and chemical treatments. The study was accomplished through the detailed literature review of factors that potentially affect the chemical composition and sweetness of edamame during plant development and current techniques used for extracting SDF from food processing byproducts (Chapter 2); comprehensively analyzing the chemical compositions of different genotypes of edamame planted in different locations (Chapter 3); observing the change of physical properties and chemical compositions of edamame harvested at different development stages and applying spectroscopy-based ML to determine the optimal harvest time (Chapter 4), and investigating how different processing parameters of the combined ball milling and citric acid process affect the yield and the properties of the extracted SDFs (Chapter 5).

In Chapter 3, ten edamame genotypes were harvested from three different locations and analyzed for chemical compositions, including soluble sugars (sucrose, glucose, fructose, raffinose, and stachyose), alanine, crude protein, NDF, starch, oil, moisture, and ash. The results showed that genotype had significant effects on all chemical compositions excluding NDF and raffinose while planting location significantly affected all chemical compositions of edamame ($p < 0.05$). Therefore, both genotype and planting location need to be considered for breeding better

edamame cultivars for the U.S. market. Among ten genotypes, R13-5029 was identified as a candidate genotype for breeding in the U.S. due to its consistently high sweetness across all planting locations.

In Chapter 4, three genotypes of edamame were harvested at six different development stages and their physical and chemical properties were analyzed and compared. A spectroscopy-based ML technique was applied to determine the edamame's optimal harvest time and the accuracies were collected and compared. The results showed that at stage R6, pod and bean weight, and pod thickness reached their highest values. Edamame also had the highest sucrose, glycine, alanine, and starch content as well as the highest sweetness. At this stage, the NDF, fat, and ash contents are relatively low. The R6-1 stage was determined to be the best time to harvest high-quality edamame if taking all physical and chemical properties changes during bean development into account. The spectroscopy-based ML method showed the highest accuracy of 0.95 for classifying "early" and "late" harvested edamame and 0.87 for classifying "early" and "ready" harvested edamame. Overall, our results indicate that the spectroscopy-based ML technique can determine the best harvest time for edamame.

In Chapter 5, we investigated the effects of citric acid concentrations, extraction temperatures, extraction time and ball milling on the extraction yield and properties of SDFs from edamame shells. The optimal SDF yield was found on the SDF extracted from ball-mill treated edamame shells. The produced SDF (SDF-IV) had the lowest pectin content and the highest total phenolic and short-chain carbohydrate contents. Particle size, morphology and crystalline characteristics of SDF changed after the combined citric acid and ball milling treatment. Moreover, ball milling improved the thermal stability and the shearing thinning behavior. SDF-IV is also promising regarding the *in vitro* fermentation performance and

antioxidant properties when compared to other SDF products, which makes it a good functional ingredient in foods with health benefits. In summary, SDF can be promisingly produced from edamame shells by ball mill-assisted citric acid treatment. The produced SDF can be considered a novel food ingredient.

6.2 Future works

First, in Chapter 3, we collected 1-year data to investigate the effect of planting location and genotype on the chemical compositions of edamame, which provided valuable guidance on edamame breeding. However, it should be noted that year-to-year variation cannot be ignored since the differences in the environment might also affect the chemical compositions of edamame. Therefore, to further confirm the effects of planting location, multiple years of study needs to be conducted on the same genotypes planted in the three locations investigated in this study.

Second, as mentioned in Chapter 4, a consistent, rapid, and accurate technique for determining the optimal harvest time is crucial for harvesting edamame with consistently high quality. In this study, our results proved the capability of the spectroscopy-based ML method on identifying the early, late and ready stages of the harvested edamame, representing a proof of concept for applying the precision farming tool for edamame harvesting. However, in the current model, multiple wavelengths were used for building the classification model. Therefore, it is of great importance to investigate if the spectral sets of data can be refined to a unique wavelength which will simplify the determination process.

Third, in Chapter 5, we proved that the combined ball milling and citric acid processing was feasible to produce SDF from edamame shells with improved certain properties which

makes it a potential candidate for novel food ingredients. However, the benefits of the SDF produced by this process to the actual foods are still unknown. Therefore, it will be interesting to apply the SDF products to different food systems (such as paste, drinks, or bread) and investigate the changes it brings to the food matrix such as food texture and potential health benefits.

Appendix A

Appendix of Chapter 4

Table A1 Dates of planting and six different harvest stages.

	Date	Days from planting date
Planting	2019.05.22	0
1st Harvest (R5-1)	2019.09.17	118
2nd Harvest (R5-2)	2019.09.22	123
3rd Harvest (R6-1)	2019.09.28	129
4th Harvest (R6-2)	2019.10.04	135
5th Harvest (R7-1)	2019.10.10	141
6th Harvest (R7-2)	2019.10.11	142

R5~R7 are three of the reproductive stages of soybean plants.
R5: beginning seed; R6: full seed; R7: beginning maturity.

Table A2 The relative sweetness of fructose, glucose, glycine, and alanine to sucrose.

Sugar/amino acid	Relative sweetness	References
Sucrose	1.00	Brady, 2013
Fructose	1.00 – 1.75	Brady, 2013
Glucose	0.40 – 0.79	Brady, 2013
Glycine	0.46 – 1.20	Cameron, 1945
Alanine	0.93 – 1.70	Cameron, 1945

Table A3 Measured color (L* a* b* value) of edamame in different harvest stages.

Genotype	Harvest Stage	L*	a*	b*
R15-10280	R5-1	53.50 ± 0.35^{c}	-7.35 ± 0.36^{c}	34.77 ± 0.93^{a}
	R5-2	55.58 ± 0.58^{bc}	-6.80 ± 0.42^{bc}	35.14 ± 0.50^{a}
	R6-1	56.86 ± 2.13^{abc}	-6.73 ± 0.35^{bc}	35.40 ± 1.26^{a}
	R6-2	59.71 ± 1.25^{ab}	-5.71 ± 0.16^{bc}	36.15 ± 1.07^{a}
	R7-1	60.01 ± 2.19^{ab}	-4.89 ± 0.74^{ab}	36.70 ± 1.44^{a}
	R7-2	61.27 ± 2.14^{a}	-3.10 ± 1.41^{a}	37.44 ± 2.22^{a}
V16-0547	R5-1	50.80 ± 0.26^{c}	-7.26 ± 0.83^{b}	32.88 ± 1.53^{c}
	R5-2	52.50 ± 0.59^{bc}	-6.89 ± 0.26^{b}	33.59 ± 0.52^{bc}
	R6-1	56.39 ± 1.81^{ab}	-6.83 ± 0.30^{b}	37.09 ± 1.63^{abc}
	R6-2	59.31 ± 1.49^{a}	-5.91 ± 0.35^{ab}	38.09 ± 1.27^{ab}
	R7-1	58.12 ± 1.88^{a}	-5.16 ± 1.35^{ab}	38.52 ± 2.82^{abc}
	R7-2	60.86 ± 3.22^{a}	-3.68 ± 2.23^{a}	39.43 ± 1.33^{a}
UA-Kirksey	R5-1	51.27 ± 0.31^{c}	-7.71 ± 0.45^{c}	32.62 ± 1.12^{ab}
	R5-2	52.23 ± 1.32^{c}	-7.30 ± 0.07^{bc}	31.34 ± 1.25^{b}
	R6-1	53.62 ± 0.19^{bc}	-7.44 ± 0.24^{bc}	31.82 ± 0.35^{b}
	R6-2	56.67 ± 0.40^{ab}	-6.93 ± 0.14^{bc}	33.13 ± 0.76^{ab}
	R7-1	56.87 ± 1.50^{ab}	-6.37 ± 0.76^{b}	34.50 ± 0.50^{ab}
	R7-2	59.44 ± 2.34^{a}	-5.23 ± 0.35^{a}	35.98 ± 2.33^{a}

Different letters (abc) from each column indicate a significant difference based on the one-way ANOVA with Tukey's HSD test ($p < 0.05$).

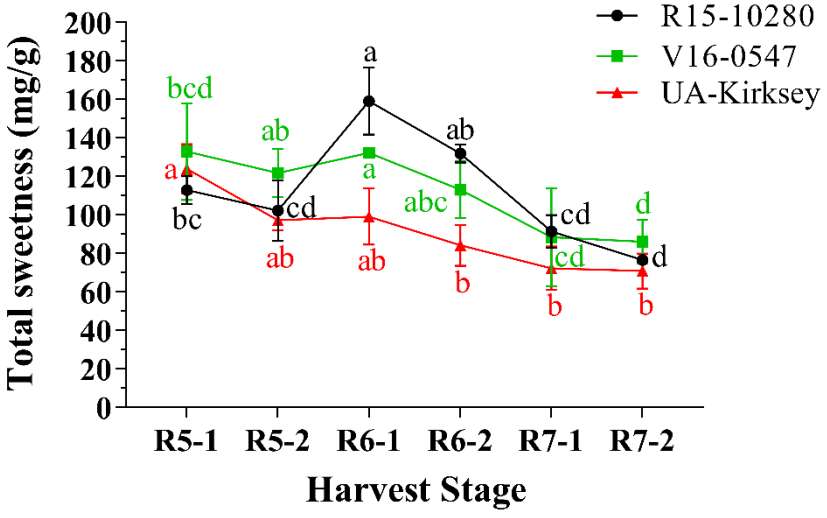

Figure A1 Total sweetness of edamame in different harvest stages. Different letters (abcd) from each line indicate a significant difference based on the one-way ANOVA with Tukey's HSD test ($p < 0.05$).

A.1 Results and discussion

Principal component analysis

In total, 13 different compositions were measured and there were inner relationships between them, thus PCA was conducted to further elucidate the relationships among sucrose, fructose, glucose, raffinose, stachyose, alanine, glycine, moisture content, protein, fat, starch, NDF, and ash content of edamame harvested from different stages. Besides, the trajectory analysis of PCA score plots could reflect the differences among different harvest stages.

Figure A2 shows the combination of PCA score plots and their corresponding trajectory analysis along with the component plots (loading plots) of three genotypes. The first two principal components (PC1 and PC2) could explain 69, 76, and 82% of the variance of R15-10280, V16-0547, and UA-Kirksey, respectively. Overall, the PCA score plots of different genotypes show the same behavior of the distributions which were widely scattered. The three score plots show that most of the points of R5 fell into the area where moisture content, fructose, glucose, NDF, and ash dominated and these were in accordance with the previous results that R5 samples had relatively higher moisture content, fructose, glucose, NDF, and ash content compared to R6 and R7. Most of the points of R6 fell into the area in which sucrose, alanine, and glycine were dominating and this agreed with the fact that R6 samples had the relatively higher sucrose, alanine, and glycine content. Regarding the points of R7, they fell into the area dominated by stachyose, raffinose, fat, and protein because these compositions increased gradually and reached a peak at R7.

The trajectory analysis of PCA was used to group the points of each genotype based on their chemical characteristics and to show the moving track of points at different harvest stages. Generally, it shows the differences between points at different harvest stages but there were also intersections. For R15-10280, the points of R5, R6, and R7 were separated into 3 clusters

separately and clearly but one point from R5-2 was grouped to cluster 3. This point sat in the middle of the area of R6 and R7 points which indicated that its chemical compositions might be more like the samples of R6 and R7. This could be due to the replication difference or experimental error. The outside arrows show the track of points at different harvest stages and it also clearly shows the harvest stage-dependent changes in the chemical compositions of edamame. For genotype V16-0547, R5-1 points were grouped into cluster 1 while R5-2 points were grouped to cluster 2 with R6 points which indicated that the samples harvested at R5-2 had more similar chemical compositions to samples harvested at R6. Points of R7 were grouped to cluster 3 while only one of them was also grouped to cluster 2. The outside arrows show the overall time-dependent changes while the inside black arrows show more detailed changes that show the trackable differences from R5-1 to R5-2 to R6, finally to R7. For genotype UA-Kirksey, all R5-1 points were grouped while R5-2 points were grouped with most of the R6 points, one R7-1, and one R7-2 point. One R6-2 point was grouped with most of the R7 points together. The changes were still time-dependent and similar to the previous two genotypes. Overall, the trajectory analysis of PCA further proved that edamame harvested from different stages had different chemical characteristics and these differences were harvest time-dependent.

To better understand the relationships among different chemical compositions, all three component plots were also shown in Figure A2. In the component plots, the lines that connect the composition points and the origin point of the coordinate system were called vectors (not shown in the figures) and each composition had one vector. The angle of the two vectors could tell us the correlations between these two compositions. The smaller the angles are, the more positively correlated the vectors are. With angles being 90º, the vectors are not likely to be correlated whereas with angles being larger (close to 180º), they are more likely negatively correlated. In all three

component plots, sucrose content was found to be always positively correlated to alanine, glycine, and starch content and negatively correlated to stachyose and raffinose content. This indicates that the accumulations and consumptions of sucrose, alanine, glycine and starch in edamame were simultaneous. Sucrose was also negatively correlated to protein and fat content. Starch and NDF contents were negatively correlated to fat and protein content which demonstrated the possibility that the carbohydrates in the edamame could be used by the plants for the synthesis of fat and protein over bean development. The moisture content was negatively correlated to the fat and protein content which indicated that over the bean development, the dry mass accumulation was mainly from the fat and protein along with the loss of the moisture.

The PCA analysis with the component plots and the trajectory analysis of the PCA gave a comprehensive perspective of the changes in chemical compositions of edamame during the bean development. A better understanding of the chemical composition changes would facilitate the harvest of edamame of better quality.

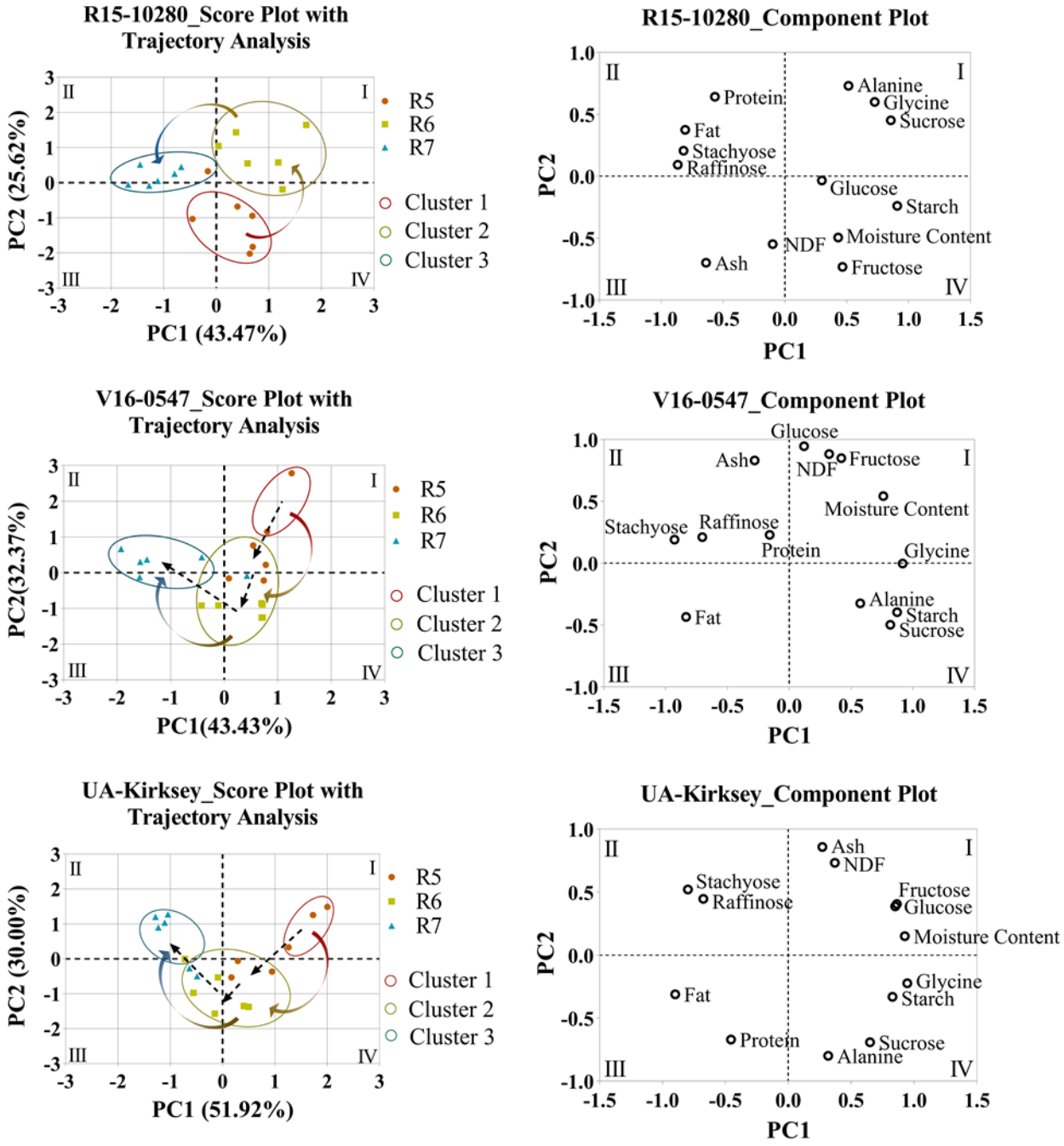

Figure A2 Trajectory analysis of principal component analysis (PCA) score plots and component plot of PC1 and PC2 of edamame harvest from different stages.

Appendix B

Copyright Release

Chapter 3: Published in Frontiers in Sustainable Food Systems

Chapter 4: Published in Food Chemistry