EVERYTHING EDAMAME: BIOLOGY, PRODUCTION, NUTRITION, SENSORY AND ECONOMICS

EDITED BY: Bo Zhang, Xujun Fu, William Jeremy Ross and Martin Williams

PUBLISHED IN: Frontiers in Sustainable Food Systems and Frontiers in Plant Science

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-880-6
DOI 10.3389/978-2-88976-880-6

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

EVERYTHING EDAMAME: BIOLOGY, PRODUCTION, NUTRITION, SENSORY AND ECONOMICS

Topic Editors:
Bo Zhang, Virginia Tech, United States
Xujun Fu, Zhejiang Academy of Agricultural Sciences, China
William Jeremy Ross, University of Arkansas System, United States
Martin Williams, Agricultural Research Service, United States Department of Agriculture, United States

Citation: Zhang, B., Fu, X., Ross, W. J., Williams, M., eds. (2022). Everything Edamame: Biology, Production, Nutrition, Sensory and Economics. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-880-6

Table of Contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 27 July 2022
DOI 10.3389/fpls.2022.976008

Check for updates

Editorial: Everything edamame: Biology, production, nutrition, sensory and economics

Martin M. Williams II[1]*, Bo Zhang[2], Xujun Fu[3] and Jeremy Ross[4]

[1]Global Change and Photosynthesis Research Unit, USDA-Agricultural Research Service, Urbana, IL, United States, [2]School of Plant and Environmental Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States, [3]Institute of Crop and Nuclear Technology Utilization, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, [4]Crop, Soil, and Environmental Sciences, University of Arkansas System Division of Agriculture, Little Rock, AR, United States

KEYWORDS

beer bean, green soybean, immature soybean seed, mukimame, vegetable soybean

OPEN ACCESS

EDITED AND REVIEWED BY
Leo Marcelis,
Wageningen University and Research, Netherlands

*CORRESPONDENCE
Martin M. Williams II
martin.williams@usda.gov

SPECIALTY SECTION
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

RECEIVED 22 June 2022
ACCEPTED 14 July 2022
PUBLISHED 27 July 2022

CITATION
Williams MM II, Zhang B, Fu X and Ross J (2022) Editorial: Everything edamame: Biology, production, nutrition, sensory and economics. *Front. Plant Sci.* 13:976008.
doi: 10.3389/fpls.2022.976008

Editorial on the Research Topic
Everything edamame: Biology, production, nutrition, sensory and economics

Globally, demand for edamame, also called vegetable soybean, is on the rise. As a crop grown and consumed in East Asian countries for centuries, in recent decades edamame has been consumed with a rising trend in new parts of the world, including the U.S. Edamame is recognized as a healthy plant-based protein which is also rich in vitamins, dietary fiber, and isoflavones. Most commonly, edamame soybean is consumed as a snack or added to salads, soups, stews, or dips.

In some countries, such as China and Taiwan, a well-developed edamame industry exists. In other countries, such as the U.S., edamame remains a crop in its infancy. In between the small, niche market and the large national market, there exists a void where the market is poorly defined and hurdles to growing an industry are both varied and numerous. Regardless of the maturity of the edamame value chain in a particular country, addressing consumer expectations and improving the sustainability of food production in the face of global change requires innovation.

This Research Topic aims to provide the latest achievements of edamame research, in multiple disciplines, to identify the status of the edamame value chain from field to fork. We are honored to receive submissions of many manuscripts on edamame. After a vigorous review and revision procedure, 12 articles were collected in this Research Topic, covering consumer preferences, crop physiology and production, economics, marketing, food processing, and plant breeding.

Despite the U.S. being a leading producer of grain-type soybean, multiple attempts over the last century to develop a substantial domestically produced edamame industry has gained minimal traction. Production of specialty crops such as edamame are generally high-risk, high-reward endeavors. In this Research Topic, Neill and Morgan conduct a risk assessment of U.S. edamame, accounting for production, finance, regulatory, price, and human resource risks. They note that numerous assumptions made about edamame production are due to lack of data. While the authors conclude edamame production in the U.S. has great potential, the authors call for private-public partnerships to shed light on potential risks and facilitate the realization of domestic production.

Growing demand for edamame in the U.S. presents a unique challenge because the food products may be minimally familiar to consumers. A greater understanding of U.S. consumer preferences for edamame was a major thrust in this Research Topic. Carneiro, Adie et al. investigated the role of overall appearance and color characteristics in consumer's acceptability of edamame beans. Using beans from 10 edamame genotypes grown in Virginia, the authors used a sensory panel to compare "dark" vs. "light" green edamame beans. They found consumers favored dark green edamame beans. Lord et al. evaluated consumer preferences for edamame marketed as fresh, local, organic, or beans on-the-stalk. They found consumers were willing to pay more for fresh, local, and organic edamame, while they were willing to pay less for beans on-the-stalk. Carneiro, Drape et al. investigated consumers' preferences and intentions to buy edamame grown in the U.S. They found domestically grown, in-shell edamame products were preferred compared to shelled edamame or imported products. In addition, survey respondents exhibited higher intention to buy fresh edamame relative to frozen edamame. Unfortunately, information on product characteristics desirable to consumers is often obtained in the late stages of edamame variety development due to the cost and complexity of traditional sensory methodologies. The review paper by Carneiro, Duncan, O'Keefe, Yin et al. argues the importance of integrating sensory attributes of edamame with germplasm improvement. They call for the development of alternative sensory methods that are simple, fast, and effective to obtain consumer data. Indeed, one paper in the Research Topic directly linked sensory evaluations with edamame breeding. Carneiro, Duncan, O'Keefe, Yu et al. used a sensory evaluation to identify edamame genotypes and sensory attributes preferred by consumers to support breeding selection criteria. They found traits described as "bitter," "sour," or "starchy" appeared less acceptable, while "salty" and "sweet" appeared more acceptable.

The need for improved edamame cultivars through plant breeding was the subject of three additional papers in this Research Topic. Yuan et al. evaluated the volatile compounds of 30 edamame core cultivars from a breeding program in Hangzhou, China. They found that the composition and concentration of volatile compounds from the cultivars examined varied with cultivar ecotype, objectives of the breeding selection, and geographic origin. They conclude volatile fingerprints of lines in a breeding program can be used to improve the desirable aroma of future cultivars. Sensory attributes are not the only traits important in edamame breeding. Effectiveness of mechanical edamame harvest could be improved with an understanding of plant architectural traits. Dhakal et al. used digital imaging technology and computer vision algorithms to characterize plant architecture and identify genetic control of these traits. They found a combination of multiple topological features that contribute to the overall pod numbers on a plant. They also identified potential candidate genes associated with pod number. Finally, Kao et al. describe a core collection of edamame accessions in Taiwan, which has created a successful edamame industry in the last half-century. The authors developed an algorithm to select 30 accessions with maximum pairwise genetic distance from a collection of 200 unique accessions. They conclude the core collection will benefit future research and breeding efforts since it retains diversity and genetic variability of edamame suitable for Taiwan.

Potassium is essential to plant growth and potassium efficiency has been the study of previous research in edamame. Liu et al. examine root potassium affinity-associated drivers and photosynthesis in edamame with different potassium efficiency. They conclude that stronger root potassium affinity drivers associated with photosynthetic adaptability to low potassium stress were key factors in determining the potassium high efficiency of edamame.

Even with an edamame cultivar capable of delivering sensory traits important to the consumer, determining the optimal time to harvest the crop is currently a combination of art and science. The size, color, and uniformity of both seeds and pods, as well as protein, oil, and starch components within the seed, change dynamically during reproductive plant growth. Moseley, Paulo da Silva et al. developed an Edamame Harvest Quality Index and demonstrate its responsiveness to planting date and cultivar. They conclude the research will help define a planting and harvesting strategy for edamame production in the U.S. Mid-South.

The current market for edamame consists of either fresh, frozen, roasted, or freeze-dried products; however, producing a high moisture content product that is shelf stable necessitates an improved pasteurization technique. Moseley, Mozzoni et al. compared acid-treatment to boiling on edamame texture and color as well as the effect of additions and cultivars on acid-treatment. They found the acid-treatment, including the addition of turmeric, helped maintain quality of canned edamame seeds. This work identifies a promising path to new edamame products not currently available to many consumers.

Collectively, the scope of papers on this Research Topic illustrates a diversity of research efforts on improving the edamame value chain from field to fork. The collection offers new insights for countries where edamame is mainstream, such as China and Taiwan, as well as places such as the U.S. where the market is growing and domestic production appears viable.

Author contributions

MW drafted the manuscript. BZ, XF, and JR reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the U.S. Department of Agriculture. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

frontiers in Sustainable Food Systems

REVIEW
published: 19 August 2020
doi: 10.3389/fsufs.2020.00124

Sensory and Consumer Studies in Plant Breeding: A Guidance for Edamame Development in the U.S.

Renata C. V. Carneiro[1], Susan E. Duncan[1], Sean F. O'Keefe[1], Yun Yin[1], Clinton L. Neill[2] and Bo Zhang[3]*

[1] Department of Food Science and Technology, Virginia Tech, Blacksburg, VA, United States, [2] Department of Agricultural and Applied Economics, Virginia Tech, Blacksburg, VA, United States, [3] School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, United States

Plant breeding is an important discipline to develop food products and improve overall quality, chemical composition, and nutritional value of crops, vegetables, fruits, and nuts, which can be important allies in health promotion. Apples, blueberries, wine grapes, tomatoes, and peanuts are a few examples of food products that were improved in past decades through plant breeding programs in the United States. Recently, edamame (vegetable soybean) has gained special attention from breeders, non-breeder researchers, growers, and consumers, and new edamame varieties are currently being developed for domestic production. As a popular nutritious crop in Asian countries, edamame is increasing in sales and consumption in the United States. Therefore, edamame has great potential to be a profitable alternative crop to replace tobacco farming, whose production and market value have been declining. Until the present date, most published reviews on edamame have focused on its agronomic characteristics. However, understanding consumer expectations, needs, and acceptability for new and improved crops like edamame is vital to guide and sustain their production. It is important that researchers working on plant breeding programs understand and consider the aspects that are relevant for both growers and consumers (e.g., crop productivity, pest and disease resistance, nutritional properties, and sensory attributes). Thus, this review paper aims to integrate available information on sensory quality of edamame and to support its development and production in the United States. This review presents an overview of how sensory evaluation and consumer studies have been used to support plant breeding programs in the development of alternative crops, such as edamame.

Keywords: consumer studies, sensory, plant breeding, food development, edamame, vegetable soybean, *Glycine max* (L.) Merr.

OPEN ACCESS

Edited by:
Kathleen L. Hefferon, Cornell University, United States

Reviewed by:
Victor Castro-Alves, Man-Technology-Environment Research Centre, Örebro University, Sweden
Jayashree Arcot, University of New South Wales, Australia

***Correspondence:**
Susan E. Duncan
duncans@vt.edu

Specialty section:
This article was submitted to Nutrition and Sustainable Diets, a section of the journal Frontiers in Sustainable Food Systems

Received: *27 April 2020*
Accepted: *10 July 2020*
Published: *19 August 2020*

Citation:
Carneiro RCV, Duncan SE, O'Keefe SF, Yin Y, Neill CL and Zhang B (2020) Sensory and Consumer Studies in Plant Breeding: A Guidance for Edamame Development in the U.S. Front. Sustain. Food Syst. 4:124. doi: 10.3389/fsufs.2020.00124

INTRODUCTION

Breeding programs elevate crop productivity and adaptation, which are crucial to expand and sustain the market and industry growth (Gallardo et al., 2018; Padikasan et al., 2018). The development of new food products through plant breeding programs is also an important way to promote human health and dietary improvement as farmers are able to sustainably produce the plant-based products consumers desire and are willing to purchase (Hansson et al., 2018). Breeders' contributions can help improve the nutritional quality of plant-based food products, for example, by increasing the amount of micronutrients (minerals and vitamins) and bioactive compounds (flavonoids, organic acids), modifying fat and oil composition (fatty acid composition), and improving carbohydrate quality (dietary fiber and sugar profile) and protein quality (amino acid profile) (Welch and Graham, 2004; Sands et al., 2009; Patil et al., 2014; Hansson et al., 2018; Padikasan et al., 2018). Over the last decades several food crops have been improved in the United States and Canada through breeding, such as apples (Hampson et al., 2000), blueberries (Gilbert et al., 2015; Gallardo et al., 2018), edamame (Jiang et al., 2018; Carneiro et al., 2020), peanuts (Pattee et al., 2001), and wine grapes (Reynolds et al., 2004). However, in order to succeed, plant breeders must consider both producers' and consumers' current and emerging needs and desires to determine priority traits (Gallardo et al., 2018; Morris and Taylor, 2019). Understanding consumer behavior, needs, and expectations is vital to direct investments and research, and promote sustainable production of new food products. Therefore, interdisciplinary approaches and collaboration among breeders, other agricultural researchers, public health officials, nutritionists, food scientists and technologists, economists, and social and political scientists are important and needed over the next decades (National Academies of Sciences, 2019).

Edamame or vegetable soybean [*Glycine max* (L.) Merr.] is a very popular food in East Asia and is increasing in popularity in the United States (Zhang and Kyei-Boahen, 2007; Carson et al., 2011). In addition to its Japanese name "edamame," vegetable soybeans are also known as "maodou" in China and "poot kong" in Korea (Kumar et al., 2011). Although often available in pods, only edamame beans are edible; they are mostly consumed as a snack, after being cooked in salted boiling water for a short time or roasted like peanuts, and may also be consumed as additions to salads, soups, and stews, stir-fried, or processed sweets and desserts (Sirisomboon et al., 2007; Mebrahtu and Devine, 2008; Sujith Kumar et al., 2011). Due to the need for import, seasonal production, and short harvest period, frozen edamame (in pods or shelled) is more common than fresh edamame in the U.S. market (Montri et al., 2006; Saldivar et al., 2010; Nolen et al., 2016; Wolfe et al., 2018). Asian countries are still the major suppliers of the edamame consumed in the U.S. and since the 1990s China has replaced Taiwan as the major edamame exporter (Wang, 2018; Flores et al., 2019). However, the growing demand for edamame in the U.S. has aroused the interest of breeders, growers, and food processors to produce this vegetable domestically (Xu et al., 2016). Hence, edamame has been suggested as an alternative crop to replace the decreasing tobacco production, for example, in Virginia and Kentucky (Xu et al., 2012; Ogles et al., 2016). Although genetically modified (GM) soybeans are predominant in the U.S. market for feed and oil production, only non-GM edamame has been sold for food consumption. In addition, consumers in the U.S. have reported they are willing to pay significantly more for non-GM edamame (Lee et al., 2018; Wolfe et al., 2018), which emphasizes the importance of breeding programs to increase domestic production and consumption of edamame in the United States.

Edamame quality is comprised of its agronomic characteristics, sensory attributes, and nutritional value. Characteristics for high quality edamame pods are commonly described as bright green with a light pubescence (white to gray), intact, without external defects, a spotless surface, good shape, and must contain two or more beans per pod to be acceptable for sale (Wszelaki et al., 2005; Williams, 2015). Edamame are harvested when the plant is still immature (between reproductive growth stages R6 and R7), seeds have filled 80–90% of the green pod width and still retain around 65% moisture content, with Brix readings (total soluble solids) between 8.5 and 12 (Johnson, 2000; Sujith Kumar et al., 2011; Nolen et al., 2016). Harvesting edamame at the R6 stage brings the benefits of having desired quality attributes, such as intense green color, low concentrations of oligosaccharide and anti-nutrients, and both sucrose content and seed weight at their peak (Xu et al., 2016).

Although the Japanese market is still the largest consumer of edamame (Wang, 2018), the number of consumers who are interested in improving their health by following a better diet has been increasing in the U.S., as well as the demand for soy products and alternative sources of protein (Ogles et al., 2016). Edamame is a nutritious, high-value and easy-to-grow specialty crop, and an appealing product for consumers interested in natural foods, especially when coming from organic production (Montri et al., 2006; Zhang and Kyei-Boahen, 2007). Soy foods like edamame are healthy dietary options for most consumers and are premier choices to versatile vegetarian and vegan diets, because they are rich sources of protein and many other nutrients (Rizzo and Baroni, 2018). The major isoflavones present in edamame (genistein and daidzein), for example, are known for their potent antioxidant property that is associated with the health benefits of soy products (Mebrahtu et al., 2004; Roland et al., 2011). However, the soybean isoflavones are associated with astringency and bitterness, two undesired sensory attributes that can impact acceptability (Roland et al., 2011).

In terms of food composition, the carbohydrate, fat and oil composition, energy density, and micronutrients (minerals and vitamins) contents are important aspects to consider for health-promoting breeding (Hansson et al., 2018). Breeders in the U.S. have worked on the development of new edamame varieties better adapted to the U.S. soil and climate through crossing between adapted U.S. grain varieties and Asian large-seeded varieties (Zhang and Kyei-Boahen, 2007). Asmara, Randolph, and Owens are examples of North American edamame cultivars and their compositions are shown in **Table 1**. Over the last decades, agronomic research studies have been performed in universities across the U.S. to introduce and improve new edamame varieties. However, most of these studies did not include information regarding consumers' perceptions and buying attitudes toward edamame (Flores et al., 2019). Likewise, the few reviews of edamame published until recently have focused exclusively on the agronomic characteristics, despite the importance of understanding consumer data to answer breeding and food production questions. As new breeding programs have started to focus on the improvement of edamame to push for a competitive production in the U.S., this review aimed (1) to combine and summarize the available information regarding consumer preferences and sensory quality of edamame for the U.S. market, and (2) to understand and describe how sensory

TABLE 1 | Edamame bean composition of three North American cultivars on a dry weight basis averaged.

	Asmara[a]	Randolph[b]	Owens[c]
Sucrose	39.6 g g^{-1}	51.8 g g^{-1}	63.0 mg g^{-1}
Protein	430 g kg^{-1}	445 g kg^{-1}	350 g kg^{-1}
Oil	92 g kg^{-1}	(Not informed)	139 g kg^{-1}
Oleic acid	43.3 % of total oil	39.3 % of total oil	45.3 % of total oil

[a]Mebrahtu et al. (2005a); [b]Mebrahtu et al. (2005b); [c]Mebrahtu et al. (2007).

evaluation methods and consumer studies have been used to assist plant breeding programs and to help the development of edamame in the U.S. We also identify some of the challenges and benefits associated with providing sensory and consumer data in the early stages of new food crop development.

SENSORY EVALUATION AND CONSUMER STUDIES IN PLANT BREEDING: IMPORTANCE, METHODS AND CHALLENGES

In plant breeding, the genetic pattern of plants can be modified to address economic importance (Padikasan et al., 2018). For example, Gallardo et al. (2018) reported that quality traits such as firmness, flavor, and shelf life can influence price (premium products), consumer demand and acceptability, machine harvestability, and economic viability of the blueberries industry (Gallardo et al., 2018). Similar association can be made to other fruit and vegetable crops. For edamame, the volatile compound 2-acetyl-1-pyrroline (2AP), characterized by a "popcorn-like" aroma, is an important aroma discriminator for premium characteristics and higher price of edamame and influence its acceptability and consumer preference (Arikit et al., 2011a,b). However, instrumental methods are still not able to completely mimic human sensory responses and perceive food products as humans do (Lawless and Heymann, 2010). Thus, sensory and consumer studies can be used to investigate quality attributes and preferences for different plant cultivars and can be valuable tools to support plant breeders in parent selection as well as selection of new breeding lines and cultivars (Hampson et al., 2000; Suwonsichon, 2019).

Although sensory evaluation has already been employed in plant breeding research worldwide, it is still common that breeding programs often are limited to the tasting results, experiences, and perspectives from only a few experts (frequently the plant breeders) to assist varietal and traits selection (Hampson et al., 2000; Bowen et al., 2019). Application of sensory techniques and/or consumer studies for guiding plant parent selection, or selection of breeding lines and cultivars requires consideration of the intended use of the information. Sensory evaluation methods and consumer studies can be used to understand and/or measure sensory attributes (e.g., appearance, aroma, flavor, texture, mouthfeel) of food products, or how consumers perceive and respond to them. Discrimination tests and descriptive sensory methods typically focus on the characteristics and differences in products, while affective tests (acceptance and preference) focus on consumer response to the product characteristics (Civille and Oftedal, 2012). Descriptive methods [e.g., quantitative descriptive analysis (QDA)] may use fewer people to complete an assessment, but participants must be trained to recognize, identify, and quantify the characteristics they perceive in the varieties. Thus, the low number of participants in descriptive panels, typically between 5 and 20, is justified by their level of calibration (Lawless and Heymann, 2010). By using this approach, differentiation of produce or crop product attributes may be assessed early in the breeding process, possibly with correlation to instrumental analyses of compositional or quality parameters (Morris and Taylor, 2019). This contribution can be highly informative to breeders and assist in guiding the development of breeding lines and cultivars. However, the time investment for training and the need to retain the trained panel members over the study duration or across multiple years of breeding development increases complexity and can be unaffordable (Morris and Taylor, 2019).

Although descriptive sensory analyses can help differentiate varieties, the data-derived information does not suggest that the varieties will be well received by consumers. Wang and Kays (2003) cited the examples of improved strawberry (*Fragaria* × *ananassa* Duch.) and tomato (*Lycopersicon esculentum* Mill.) cultivars that were released to the market without previous validation of consumer acceptability; they have bigger size, longer shelf-life, but do not meet consumer expectations in terms of flavor. Consumer testing is used to estimate the response by untrained product users, purchasers, or those interested in the broad class of products. In order to estimate the public acceptance, a large number of participants are needed and a typical guidance is 75–150 responses per product (Lawless and Heymann, 2010). Product limitations in early stage of breeding programs may limit the use of consumer testing to advanced breeding lines. Therefore, the first step to obtain acceptability information for varietal development is to assure sufficient sample availability for the sensory studies.

Plant breeding programs have commonly applied sensory evaluation and/or consumer testing to investigate which cultivars (or new genotypes) are mostly preferred by consumers and which major sensory characteristics drive these preferences (Hampson et al., 2000). The list of major challenges regarding the development and evaluation of novel fruit and vegetable cultivars includes natural variability, the fact that products cannot be stored for a long time, and different cultivars are not always available at the same period due to different optimal maturation and harvest dates (Jaeger et al., 2003; Bowen et al., 2019). These factors can complicate, for example, the selection and training of panelists for descriptive analysis, which is a sensory method often applied in later stages of breeding programs (to assess flavor of selected varieties, for example). Consumer acceptance and preference mapping are additional sensory analyses that have been applied to support plant breeding programs. These approaches were used, for example, to support development of kiwifruit and apple cultivars in New Zealand and Canada, respectively (Jaeger et al., 2003; Bowen et al., 2019). However, traditional acceptability studies are often difficult to perform or are avoided for routine screening selection in breeding programs due to limitations in resources and sample availability (Hampson et al., 2000). In the two fruit development studies cited above, researchers conducted descriptive sensory analysis prior to investigating product acceptability (Jaeger et al., 2003; Bowen et al., 2019). Statistical approaches (regression and correlation) can also be used to relate instrumental measures and consumer acceptance with descriptive analysis information (Lawless and Heymann, 2010).

The development of a lexicon, a standardized list of descriptors that characterize a food product, is another important application of sensory evaluation methods to support the development of new products, such as new plant cultivars (Suwonsichon, 2019). Lexicons, also called word lists, are important tools applied in descriptive analysis and enable clear and effective communication among different audiences (e.g., scientists, researchers, consumers, breeders, product developers, producers, industry etc.) (Suwonsichon, 2019). In the last decades, lexicons have been developed and used to support improvement and development of food crops worldwide. Talavera-Bianchi et al. (2010), for example, used samples of beet greens, swiss chard, spinach, endive, radicchio, lettuce, mustard greens, pak (bok) choy, turnip greens, cabbage, collard greens, kale, arugula, watercress, cilantro, and parsley to develop a lexicon (32 terms) to describe flavor of fresh leafy vegetables. Likewise, Belisle et al. (2017) developed a lexicon (29 terms) to describe appearance, aroma/flavor, texture, and feeling factors of fresh peach, and Griffin et al. (2017) developed a lexicon (29 terms) to characterize sensory attributes of cashew nuts (feeling factors, flavor, and texture terms). These validated sensory descriptors can be applied, for example, in check-all-the-apply (CATA) lists and surveys (e.g., economic, marketing, and behavioral surveys), which can be used to support varietal screening and selection in breeding programs. For instance, the edamame lexicon developed by Krinsky et al. (2006) was used as the major reference for the CATA list used by consumers in our recent study (Carneiro et al., 2020).

Overall, optimized sensory evaluation methods are desired to evaluate product quality (e.g., flavor and texture) and support breeding selection in early stages of the breeding schemes (Morris and Taylor, 2019). CATA is a fast and simple descriptive methodology that does not require trained panelists. Nevertheless, it requests a minimum of 60–80 participants, who are instructed to select from a list of descriptors the ones in their opinion that best describe the products (Qannari, 2017; Alexi et al., 2018). CATA typically does not measure intensity of attributes, but can show differences in sensory profile (Alexi et al., 2018). It has been combined with traditional hedonic preference tests to support the development of food products such as strawberries in Uruguay (Lado et al., 2010), *Amaranthus* in South Africa (Hiscock et al., 2018), and tomatoes in Mexico (Vela-Hinojosa et al., 2018). Those researchers suggested CATA as a simple and less time-consuming method to evaluate differences among new genotypes and support the breeding selection. The use of CATA and other sensory approaches to access consumer's preferences and acceptability of edamame in the U.S. market are discussed next.

CONSUMER PERCEPTION OF EDAMAME SENSORY QUALITY IN THE U.S.

While breeders focus on in-field appearance criteria (number of beans per pod, color, shape, defects at time of harvest), from the global perspective, consumers, and distributors evaluate edamame quality by its desirable sensory attributes, including appearance (pod and bean), aroma, flavor, and texture (firmness) (Kelley and Sánchez, 2005; Williams, 2015). Additionally, when breeding to develop heathier produce and crop food products and improve nutritional profile of fruits and vegetables, it is important to consider the impact that bioactive compounds might have on flavor (e.g., increase bitterness or astringency) (Civille and Oftedal, 2012; Patil et al., 2014). Nutritional and sensory properties are key motivators for consumers to purchase edamame, which means commercial varieties must have high-quality sensory characteristics in order to be accepted (Kelley and Sánchez, 2005). Nevertheless, only a few research papers published in the last decades show how descriptive sensory and/or consumer methods (affective tests) have been used by researchers in the U.S. to investigate consumer perception of edamame sensory quality (**Table 2**). In a collective manner, these studies investigated a large set of edamame genotypes (cultivars and/or varieties in development) that were grown in the U.S., as well as commercially processed edamame products (in-pod or shelled beans) available in the U.S. market. Understanding which sensory characteristics of edamame are important to growers, processors, and consumers is vital to develop a sustainable domestic production. Therefore, their findings are summarized in the next paragraphs.

In the earliest study presented in **Table 2**, researchers in Colorado reported that American consumers seem to prefer more mature beans, with "buttery" texture and flavor, while Japanese consumers prefer sweeter beans, with crisper texture and flower-like flavor (Johnson et al., 1999). According to the researchers, this preferred "buttery" texture could be obtained through a delay in harvest, but no further information was given about it, nor about how the sensory panels were conducted. Next, researchers in Virginia reported that untrained consumers evaluated sensory attributes (texture, color, beaniness, nuttiness, sweetness, oiliness, and aftertaste) and overall eating quality of 31 edamame genotypes (maturity groups III–VI), and most samples were characterized as color "light green" to "green," texture "slightly resistant" to "resistant," relatively low sweetness and nuttiness, "slightly beany" to "beany," not oily, and pleasant aftertaste (Young et al., 2000). Their color results suggested panelists evaluated the genotypes differently, which illustrated a limitation of the study: the use of untrained panelists to perform a descriptive test (not calibrated before the study). In addition, despite the small number of panelists in the study, which was another limitation, results suggested flavor, and nutrient attributes as the major motivators to purchase edamame. The importance of considering sensory attributes when selecting

TABLE 2 | Sensory studies performed with edamame samples in the United States.

Location	Sensory evaluation	Panelists	Edamame samples	Sensory attributes analyzed	References
Virginia	*Acceptability*; 9-point hedonic scales (1 = "dislike extremely" and 9 = "like extremely")	Screening study $n = 182$ (50–53 per test); validation study $n = 171$ (90 per test) (untrained)	Screening study: 20 edamame genotypes (2 cultivars, Asmara and UA-Kirksey, and 18 advanced breeding lines) grown in Little Rock, AR, Blacksburg and Painter, VA	Overall liking, appearance, aroma, flavor, texture	Carneiro et al., 2020
	Descriptive; 5-point intensity scale (1 = "not sweet," 5= "extremely sweet")	Participants were allowed to participate in one or more tests (screening study: up to 10; validation study: up to 4)	Validation study: 10 edamame genotypes (1 cultivar, UA-Kirksey, and 9 advanced breeding lines) grown in Blacksburg and Painter, VA, Portageville, MO, and Stoneville, MS	Sweetness intensity	
	Descriptive; check-all-that-apply (CATA) question using a list of 15 descriptors			Aroma, flavor, texture	
California	*Acceptability*; 11-point hedonic scales (0 = "Do not like at all" and 10 = "Like extremely")	$n = 74$ (untrained)	Giant Midori, ButterBean, and Kuroshinja varieties, organically grown in Northern California. Same samples were evaluated in both sensory studies	Flavor, texture, appearance, and overall-liking	Flores et al., 2019
	Descriptive; Free choice profiling (FCP) methodology, scale: 0 = "None" to 10 = "Extreme"	$n = 25$ (flavor), $n = 24$ (texture), and $n = 37$ (appearance) (untrained)		Flavor, texture and appearance	
Arkansas	*Acceptability*; 9-point hedonic scales (1 = "dislike extremely" and 9 = "like extremely")	$n = 117$ (untrained)	A genetically modified (GM) and a non-GM soybean cultivar intended for feed and oil production harvested at the edamame stage	Appearance, aroma, flavor, texture, and overall impression	Wolfe et al., 2018
Illinois	Not described; study performed by a vegetable processor	Not described	Fresh pods and seeds of selected edamame genotypes grown at the University of Illinois Vegetable Crop Farm	Appearance (pod size and color, seed color and blemishes), texture, and flavor	Williams, 2015
North Carolina	*Descriptive*; Lexicon development, 0–15-point intensity scale	$n = 12$ (untrained)	Twenty commercial frozen edamame products (brands from China, Taiwan or Japan; shelled and in-pod options) obtained in U.S. grocery stores	Flavor	Krinsky et al., 2006
	Descriptive; Lexicon verification, 0–15-point intensity scale	$n = 12$ (trained)	(1) A subset of the commercial samples used for the lexicon development. (2) Mojo Green variety grown at North Carolina State University research farm (Goldsboro, NC). Samples from a blanching study (100°C for 0, 30, 60, 90, 120, and 180 s)		
Pennsylvania	*Acceptability*; Overall-liking: 9-point hedonic scale (1 = "extremely dislike" and 9 = "like extremely"). Firmness: 7-point "just about right scale" (1 = "much too soft" and 7 = "much too firm")	$n = 113$ (untrained)	Early Hakucho, Green Legend, and Kenko cultivars grown at the Horticulture Research Farm, Russell E. Larson Research Center (Rock Springs, PA)	Overall-liking and firmness	Kelley and Sánchez, 2005
	Preference; Ranking (preference order from "most liked" to "least liked)			Liking	

(Continued)

TABLE 2 | Continued

Location	Sensory evaluation	Panelists	Edamame samples	Sensory attributes analyzed	References
Ohio	*Descriptive*; 9 cm horizontal line scales (from less to more)	$n = 10$ (trained)	Six commercial varieties (Sapporo Midori, White Lion, Early Hakucho, Sayamusume, Misono Green, and Kenko) organically grown at the Ohio Agricultural Research and Development Center (Wooster, OH)	Flavor and texture	Wszelaki et al., 2005
	Acceptability; 9-point hedonic scales (1 = "dislike extremely" and 9 = "like extremely")	$n = 54$ (untrained)		Appearance (pods and beans), aroma, taste, texture, aftertaste, and overall acceptability	
Virginia	*Descriptive*; 5-point scales. Color range: 1 = "yellow green" to 5 = "dark green." Texture range: 1 = "not resistant" to 5 = "extremely resistant." Intensity scale ranges (flavor attributes): from 1 = "not intense" to 5 = "extremely intense." Aftertaste: 1 = "extremely unpleasant "to 5 = "extremely pleasant, "plus a sixth category labeled 6 = "no aftertaste"	$n = 22$ (total) (untrained) Panelists were grouped into 3 panels, A, B, and C, which had 8, 6, and 10 participants, respectively	31 maturity groups III-VI genotypes grown at Randolph Research Farm of Virginia State University (Petersburg, VA)	Color, texture, sweetness, nuttiness, beaniness, oiliness, and aftertaste	Young et al., 2000
	Acceptability; Overall eating quality: 5-point scale (poor, fair, good, very good, and excellent)			Overall eating quality	
Colorado	*Not described*; 10-point scale (1 = "poor, "10 = "excellent")	Not described	Five Japanese edamame cultivars (SE1–SE5) provided by Seedex, Inc. (Longmont, CO), which were grown in Rocky Ford and Ft. Collins, CO, between 1994 and 1998	Texture	Johnson et al., 1999

genotypes for production was acknowledged by the authors, but they did not report how sensory data was used to guide breeders throughout the breeding process (for example, selection criteria or decision tree).

Sensory studies conducted a few years later in Ohio and Pennsylvania used commercial cultivars to assess U.S. consumer acceptability of edamame. Wszelaki et al. (2005) investigated acceptability and sensory characteristics of six commercial cultivars already available to growers and overall consumer acceptability (mean scores) of the edamame cultivars were reported as following: Misono Green = 5.5, Early Hakucho = 5.9, Kenko = 6.1, Sapporo Midori = 6.1, White Lion = 6.1, and Sayamusume = 6.3 (9-point hedonic scale; 9 = "like extremely"). Researchers reported that significant differences in acceptability were only observed in pod appearance and taste of edamame beans. Consumers liked better the pod appearance of Kenko, White Lion, and Sayamusume cultivars and the taste of Sayamusume, Kenko and Sapporo Midori beans. In sequence, sweetness and chewiness were suggested by consumers as the most important attributes to differentiate edamame varieties (Wszelaki et al., 2005). The cultivar Kenko was reported as the sweetest edamame evaluated, but the intensity of its sweetness attribute was not significantly different from cultivar Sapporo Midori. Kelley and Sánchez (2005) also reported a high overall acceptability score for the edamame cultivar Kenko (mean score = 6.84; 9-point hedonic scale). In their study, consumers evaluated overall-liking and firmness of edamame beans of cultivars Kenko, Early Hakucho and Green, then ranked the three cultivars in order of preference. Researchers reported that Kenko was ranked the most preferred edamame and its firmness was rated as just about right. Although the studies of Wszelaki et al. (2005) and Kelley and Sánchez (2005) investigated acceptability and sensory attributes of commercial edamame cultivars not necessarily developed to be grown in the U.S., they offer valuable information that can support parent selection in edamame breeding programs. They also provide initial information to breeders about quality traits of edamame that are desired by consumers in the U.S. and can drive purchase intent.

Sensory and consumer studies that investigated edamame attributes and acceptability in the U.S. mostly used samples of processed edamame products instead of raw edamame. Krinsky et al. (2006) used up to 20 commercial processed (frozen) edamame products for the development and validation of a lexicon that contained 14 terms to describe edamame flavor: "raw bean," "cooked bean," "green complex," "fruity complex," "nutty/almond," "brothy," "sulfur," "salty," "sweet," "sour," "bitter," "astringent," "umami," and "metallic." Samples for the lexicon development consisted of edamame products (in-pod and shelled options) processed in Asian countries and obtained in U.S. grocery stores. For the lexicon verification, participants evaluated two sets of samples: a subset of commercial samples, and a set of shelled edamame (variety Mojo Green) from a blanching study conducted by the researchers. The importance of lexicons to support breeding programs was discussed in the previous section of this review. Although the study performed by Krinsky et al. (2006) was not directly associated to a breeding program, it is the only lexicon for edamame found in the literature. As the lexicon was focused on flavor descriptors, there is still a need for a more complete standardized list of descriptors that includes other edamame attributes, such as texture, mouthfeel and appearance.

Recently, the acceptability of three edamame cultivars (Giant Midori, Kuroshinja, and ButterBean) was investigated as part of a broader organic vegetable research project in California (Flores et al., 2019). Participants rated overall liking, appearance, flavor, and texture of each sample, and Giant Midori was the most liked edamame (overall liking and all sensory attributes), while ButterBean had the lowest sensory scores. Next, a free choice profiling (FCP) descriptive analysis was performed after the acceptability testing; participants (untrained) created their own descriptors to describe appearance, flavor, and texture, then rated each of their descriptors using an intensity scale. Similar descriptors were grouped by researchers for analysis. For appearance, all cultivars were mostly described as small, green, and fuzzy. However, researchers did not clearly describe if the appearance descriptors were associated with edamame pods, beans, or both. Two factors explained flavor variability; the first factor (sweet with minor notes of strong and fresh) was mostly identified with the Giant Midori and Kuroshinja cultivars, and the second one (bland, earthy, and grass) was mostly linked to the ButterBean edamame. Lastly, three factors explained texture variability; the first factor (crunchy, with some firmness and wetness) was mostly associated with Giant Midori edamame; the second factor (chewy, firm, slippery, smooth, squishy) best explained the Kuroshinja cultivar, and the third factor (bumpy, dry, fuzzy, hard, mealy, soft, stringy) was mostly linked to the ButterBean cultivar. As FCP does not request trained sensory panelists, it is a less expensive and quicker descriptive method that can be used, for example, to provide breeders with information about how consumers perceive the sensory attributes of improved varieties, especially in earlier stages.

CATA is another example of a quick descriptive method that does not request trained participants and could be associated with acceptability tests, as previously mentioned in this review. The use of a CATA question to investigate sensory profile of edamame was only reported by Carneiro et al. (2020). In Virginia, consumer studies and sensory evaluation are currently being used to support breeding decisions in a multistate plant-breeding program focused on developing varieties for domestic production (Carneiro et al., 2020). The authors divided their sensory study in two parts: screening study (first year) and validation study (second year). First, 20 edamame genotypes (breeding lines and cultivars) were evaluated by untrained consumers who participated in one or multiple sensory panels. Then, first-year consumer data (overall-liking, appearance, aroma, taste, and texture liking, and CATA descriptor selection) led to the development of a decision tree to assist breeding selection criteria, and the following selection of 10 genotypes for further sensory evaluation (validation study). Researchers also reported the use of penalty analysis to understand the impact of each of the 15 CATA descriptors in edamame acceptability. They suggested "salty" and "sweet" as the main natural sensory attributes of edamame associated with high acceptability scores, while "bitter" was the main attribute associated with lower acceptability scores.

The decision tree developed by Carneiro et al. (2020) illustrated how sensory studies can be used to support breeding selection criteria. The authors reported the tool was developed based on the literature and acceptability scores of selected edamame cultivars (checks). For example, genotypes suggested to continue in the breeding programs should have at least a 5.9 (rounded up) overall-liking mean score and/or at least a 1.8 sweetness intensity mean score. This decision tree included approval of edamame breeding lines whose profile was characterized by at least 4 sensory attributes with high acceptability scores. It supported the selection of most breeding lines chosen for the validation study, as well as the identification of varieties that are strong candidates to be released. Likewise, a previous study performed in Illinois reported how sensory studies helped identify edamame genotypes that were promising genotypes to be grown in the North Central United States (Williams, 2015). Sensory data was obtained by a vegetable processor and the following sensory attributes were evaluated: pod color and size, and seed color, blemishes, texture, and flavor. Details about how the sensory study was performed were reported. Sensory evaluation criteria was an acceptable threshold to the vegetable processor, and the basis for this threshold was only described as acceptable R6 pods and seeds, meaning "two- to three-seed pods, green pods and seeds, seed free of blemishes, a smooth seed texture, and seed with a sweet and/or nutty flavor." Besides an acceptable sensory profile, their selection criteria included emergence (>36%), height at R6 (<66 cm), and seed mass (>20 g 100 seed^{-1}). Both agronomic and sensory criteria supported the selection of 12 edamame genotypes from an initial set of 136 genotypes (Williams, 2015). As there is no standard way to use sensory results to make breeding decisions, the studies above can be used as a reference for future plant breeding programs.

CONSUMER BEHAVIOR AND PURCHASE INTENTION OF EDAMAME IN THE U.S. MARKET

In the mid-Atlantic and Southeast U.S., edamame has been promoted to growers as a profitable alternative or new crop, for example, to replace tobacco (Xu et al., 2012; Ogles et al., 2016). However, when selecting the best cultivar, growers are suggested to consider agronomic aspects of the cultivars, such as yield characteristics, but also consumers' preferences (Ogles et al., 2016). One requirement for varietal success (vegetables, fruits, nuts) is having a market for the new developed crop. Seed and crop producers, as well as food processors, want to know that there will be economic value for growing, distributing, and selling the new varieties in a competitive food market. Estimating consumer interest in and motivation for edamame products through consumer willingness-to-pay (WTP) studies helps provide that information. For instance, Flores et al. (2019) reported that consumers in California showed the highest purchase intent for the edamame cultivar Giant Midori and the lowest for the cultivar ButterBean, which were, respectively, the most and least liked edamame evaluated.

U.S. consumers are willing to pay price premium for non-genetically modified edamame (Wolfe et al., 2018), which suggests breeding programs are vital to develop and sustain the edamame production in the U.S. Wolfe et al. (2018) reported that although no significant difference was observed between preference scores of genetically modified (GM) and non-GM edamame samples, consumers valued more on non-GM-labeled edamame and were willing to pay at least $0.42 more per unit for that information. Based on WTP, unlabeled or GM-labeled products did not provide similar additional value to edamame. In addition, Lee et al. (2018) reported that negative information about GM products affects consumer WTP for edamame more than positive information. The authors suggested it would be difficult to introduce GM edamame in the U.S. market, which reinforced the importance of breeding programs for increasing domestic production of this vegetable.

Furthermore, a deeper understanding of factors that drive consumer purchase intent is important to build a sustainable domestic production. Recently, Carneiro et al. (2020) reported that in both years their sensory study was conducted, ∼50% of the participants answered they consume edamame few times per year, and chose "like the taste" and "for heath reason," respectively, as their main motivations to consume soy products. Previous consumer studies conducted in Pennsylvania also investigated behaviors and attitudes toward edamame to have a better picture of the U.S. market preferences and needs. Kelley and Sánchez (2005) investigated the potential demand for edamame through a telephone survey in the Metro-Philadelphia area. The majority of the participants belonged to the group of potential edamame purchasers and they were more likely to care about the nutritional profile of the products they purchased and consumed. This group was also characterized by the largest number of participants who reported they have included soy or soy-based products in past purchases, and had heard about edamame before the survey. After these participants were informed about edamame origin, health benefits associated with its consumption, and some ways to prepare it, most of them reported their potential to eat edamame as "very likely" or "likely."

A subsequent consumer study performed in supermarkets located in the metropolitan Philadelphia area investigated consumer interest in fresh and in-shell edamame (Montri et al., 2006). Although fresh edamame can be occasionally purchased in farmers markets or groceries stores, most edamame available in the U.S. market is sold as a processed product, typically blanched and frozen stored. Consumers reported same preference to purchase in the future either fresh edamame in-shell only, or both in-shell and shelled. In addition, more than half of the participants reported they were more likely to buy Pennsylvania grown edamame, mostly because they were farmed without the application of pesticides. Among the factors that could possibly affect their decision to purchase a new product, participants ranked friend's recommendation, in-store promotions (sample of product at the supermarket), price, outside advertising (magazine or news article), and product packaging (health benefits stated on package), in this order, as the main factors. In summary, U.S. consumers have expressed that they value the nutritional and sensory profile of edamame and it can drive their purchase decisions. For edamame breeders, the consumer studies presented in this review reinforce the importance of breeding to improve nutritional quality of new varieties for the U.S. market. For domestic edamame growers and processors, this knowledge is important to guide production decisions, such as type of products (in-shell or shelled) and packaging information that are mostly appreciated.

CONCLUSION

Sensory attributes and nutritional value, as well as agronomic characteristics, are important factors to be considered when breeding new cultivars to develop and improve food products, such as edamame. Sensory evaluation and consumer studies provide valuable information to support plant breeders in the selection of genotypes that have more potential for market success, such as desirable product characteristics (e.g., sensory and nutritional profile, shelf life, organic production) for which consumers would be willing to pay more. However, this information is often obtained only in late stages of breeding schemes due to sample limitations, cost, and complexity of traditional sensory methodologies. Releasing improved cultivars of fruits and vegetables without understanding consumers' preferences and expectations increases the risk of market failure. Understanding the available sensory methods (discrimination, descriptive, acceptance, and preference tests) is essential since early planning phases to manage resources and mitigate problems related, for example, to sample quality and limitations. Additionally, we acknowledge the importance of seeking alternatives that can contribute to reduce the cost and complexity to obtain consumer data. Future efforts are needed and encouraged to develop and validate new simple, fast, and optimized sensory methods to support breeding programs. Likewise, future studies aiming to understand the relationship between sensory and analytical data are suggested to support the development of quality evaluation methodologies. Lastly, an interdisciplinary approach that integrates breeders and non-breeder researchers, such as crop scientists, nutritionists, food scientists, and economists, has been proven to be of great value for the development and success of new food crops and is suggested for future breeding programs.

AUTHOR CONTRIBUTIONS

RC drafted the manuscript and edited based on co-author and reviewers' comments. All authors reviewed, edited, and approved draft and submitted versions of the manuscript.

FUNDING

This work was funded, in part, by USDA-NIFA, Grant No. 2018-51181-28384, Accession No. 1016465, and the Virginia Agricultural Experiment Station.

ACKNOWLEDGMENTS

The authors thank USDA-NIFA and Virginia Agricultural Experiment Station for the financial support.

REFERENCES

Alexi, N., Nanou, E., Lazo, O., Guerrero, L., Grigorakis, K., and Byrne, D. V. (2018). Check-All-That-Apply (CATA) with semi-trained assessors: sensory profiles closer to descriptive analysis or consumer elicited data? *Food Qual. Prefer.* 64, 11–20. doi: 10.1016/j.foodqual.2017.10.009

Arikit, S., Yoshihashi, T., Wanchana, S., Tanya, P., Juwattanasomran, R., Srinives, P., et al. (2011a). A PCR-based marker for a locus conferring aroma in vegetable soybean (*Glycine max* L.). *Theor. Appl. Genet.* 122, 311–316. doi: 10.1007/s00122-010-1446-y

Arikit, S., Yoshihashi, T., Wanchana, S., Uyen, T. T., Huong, N. T. T., Wongpornchai, S., et al. (2011b). Deficiency in the amino aldehyde dehydrogenase encoded by GmAMADH 2, the homologue of rice Os2AP, enhances 2-acetyl-1-pyrroline biosynthesis in soybeans (*Glycine max* L.). *Plant Biotechnol. J.* 9, 75–87. doi: 10.1111/j.1467-7652.2010.00533.x

Belisle, C., Adhikari, K., Chavez, D., and Phan, U. T. X. (2017). Development of a lexicon for flavor and texture of fresh peach cultivars. *J. Sens. Stud.* 32, 1–12. doi: 10.1111/joss.12276

Bowen, A. J., Blake, A., Tureček, J., and Amyotte, B. (2019). External preference mapping: a guide for a consumer-driven approach to apple breeding. *J. Sens. Stud.* 34. doi: 10.1111/joss.12472

Carneiro, R. C. V., Duncan, S. E., O'Keefe, S. F., Yu, D., Huang, H., Yin, Y., et al. (2020). Utilizing consumer perception of edamame to guide new variety development. *Front. Sustain. Food Syst.* [submitted for review in this same special edition]

Carson, L., Freeman, J., Zhou, K., Welbaum, G., and Reiter, M. (2011). Cultivar evaluation and lipid and protein contents of Virginia-grown edamame. *Horttechnology* 1, 131–135. doi: 10.21273/HORTTECH.21.1.131

Civille, G. V., and Oftedal, K. N. (2012). Sensory evaluation techniques—make "good for you" taste "good." *Physiol. Behav.* 107, 598–605. doi: 10.1016/j.physbeh.2012.04.015

Flores, D., Giovanni, M., Kirk, L., and Liles, G. (2019). Capturing and explaining sensory differences among organically grown vegetable-soybean varieties grown in Northern California. *J. Food Sci.* 84, 613–622. doi: 10.1111/1750-3841.14443

Gallardo, R. K., Zhang, Q., Dosset, M., Polashock, J., Rodriguez-Saona, C., Vorsa, N., et al. (2018). Breeding trait priorities of the blueberry industry in the United States and Canada. *HortScience* 53, 1021–1028. doi: 10.21273/HORTSCI12964-18

Gilbert, J. L., Guthart, M. J., Gezan, S. A., De Carvalho, M. P., Schwieterman, M. L., Colquhoun, T. A., et al. (2015). Identifying breeding priorities for blueberry flavor using biochemical, sensory, and genotype by environment analyses. *PLoS ONE* 10:e0138494. doi: 10.1371/journal.pone.0138494

Griffin, L. E., Dean, L. L., and Drake, M. A. (2017). The development of a lexicon for cashew nuts. *J. Sens. Stud.* 32, 1–10. doi: 10.1111/joss.12244

Hampson, C. R., Quamme, H. A., Hall, J. W., Macdonald, R. A., King, M. C., and Cliff, M. A. (2000). Sensory evaluation as a selection tool in apple breeding. *Euphytica* 111, 79–90. doi: 10.1023/A:1003769304778

Hansson, S. O., Åman, P., Becker, W., De Koning, D. J., Lagerkvist, C. J., Larsson, I., et al. (2018). Breeding for public health: a strategy. *Trends Food Sci. Technol.* 80, 131–140. doi: 10.1016/j.tifs.2018.07.023

Hiscock, L., Bothma, C., Hugo, A., Van Biljon, A., and Van Rensburg, W. S. J. (2018). Overall liking and sensory profiling of boiled Amaranthus leaves using the check-all-that-apply question. *CYTA J. Food* 16, 822–830. doi: 10.1080/19476337.2018.1464521

Jaeger, S. R., Rossiter, K. L., Wismer, W. V., and Harker, F. R. (2003). Consumer-driven product development in the kiwifruit industry. *Food Qual. Prefer.* 14, 187–198. doi: 10.1016/S0950-3293(02)00053-8

Jiang, G.-L., Rutto, L. K., Ren, S., Bowen, R. A., Berry, H., and Epps, K. (2018). Genetic analysis of edamame seed composition and trait relationships in soybean lines. *Euphytica* 214, 1–10. doi: 10.1007/s10681-018-2237-9

Johnson, D. (2000). Edamame: westerners develop a taste for Japanese soybean. *Eng. Technol. Sustain. World* 7, 11–12.

Johnson, D., Wang, S., and Suzuki, A. (1999). "Edamame: a vegetable soybean for Colorado," in *Perspectives on New Crops and New Uses*, ed J. Janick (Alexandria, VA: ASHS Press), 385–387. Available online at: https://hort.purdue.edu/newcrop/proceedings1999/pdf/v4-385.pdf (accessed September 3, 2018).

Kelley, K. M., and Sánchez, E. S. (2005). Accessing and understanding consumer awareness of and potential demand for edamame. *HortScience* 40, 1347–1353. doi: 10.21273/HORTSCI.40.5.1347

Krinsky, B. F., Drake, M. A., Civille, G. V., Dean, L. L., Hendrix, K. W., and Sanders, T. H. (2006). The development of a lexicon for frozen vegetable soybeans (edamame). *J. Sens. Stud.* 21, 644–653. doi: 10.1111/j.1745-459X.2006.00088.x

Kumar, V., Rani, A., Goyal, L., Pratap, D., Billore, S. D., and Chauhan, G. S. (2011). Evaluation of vegetable-type soybean for sucrose, taste-related amino acids, and isoflavones contents. *Int. J. Food Prop.* 14, 1142–1151. doi: 10.1080/10942911003592761

Lado, J., Vicente, E., Manzzioni, A., and Ares, G. (2010). Application of a check-all-that-apply question for the evaluation of strawberry cultivars from a breeding program. *J. Sci. Food Agric.* 90, 2268–2275. doi: 10.1002/jsfa.4081

Lawless, H. T., and Heymann, H. (2010). *Sensory Evaluation of Food Principles and Practices, 2nd Edn*. New York, NY: Springer.

Lee, J. Y., Popp, M. P., Wolfe, E. J., Nayga, R. M. Jr., Popp, J. S., et al. (2018). Information and order of information effects on consumers' acceptance and valuation for genetically modified edamame soybean. *PLoS ONE* 13:e0206300. doi: 10.1371/journal.pone.0206300

Mebrahtu, T., and Devine, T. E. (2008). Diallel analysis of sugar composition of 10 vegetable soybean lines. *Plant Breed.* 128, 249–252. doi: 10.1111/j.1439-0523.2008.01561.x

Mebrahtu, T., Devine, T. E., Donald, P., and Abney, T. S. (2005a). Registration of "Asmara" vegetable soybean. *Crop Sci.* 45, 408–409. doi: 10.2135/cropsci2005.0408

Mebrahtu, T., Devine, T. E., Donald, P., and Abney, T. S. (2005b). Registration of "Randolph" vegetable Soybean. *Crop Sci.* 45, 2644–2945. doi: 10.2135/cropsci2005.007

Mebrahtu, T., Devine, T. E., Donald, P. A., and Abney, T. S. (2007). Registration of "Owens" vegetable soybean. *J. Plant Regist.* 1, 95–96. doi: 10.3198/jpr2006.09.0570crc

Mebrahtu, T., Mohamed, A., Wang, C. Y., and Andebrhan, T. (2004). Analysis of isoflavone contents in vegetable soybeans. *Plant Foods Hum. Nutr.* 59, 55–61. doi: 10.1007/s11130-004-0023-4

Montri, D. N., Kelley, K. M., and Sánchez, E. S. (2006). Consumer interest in fresh, in-shell edamame and acceptance of edamame-based patties. *HortScience* 41, 1616–1622. doi: 10.21273/HORTSCI.41.7.1616

Morris, W. L., and Taylor, M. A. (2019). Improving flavor to increase consumption. *Am. J. Potato Res.* 96, 195–200. doi: 10.1007/s12230-018-09702-7

National Academies of Sciences, Engineering and Medicine. (2019). *Science Breakthroughs to Advance Food and Agricultural Research by 2030*. Washington, DC: The National Academies Press.

Nolen, S., Zhang, B., and Kering, M. K. (2016). Increasing fresh edamame bean supply through season extension techniques. *J. Hortic.* 03, 1–5. doi: 10.4172/2376-0354.1000170

Ogles, C. Z., Guertal, E. A., and Weaver, D. B. (2016). Edamame cultivar evaluation in Central Alabama. *Agron. J.* 108, 2371–2378. doi: 10.2134/agronj2016.04.0218

Padikasan, I. A., Chinnannan, K., Kumar, S., and Subramaniyan, G. (2018). "Agricultural biotechnology: engineering plants for improved productivity and quality," in *Omics Technologies and Bio-Engineering. Volume 2: Towards Improving Quality of Life*, eds D. Barh and V. Azevedo (Cambridge, MA: Academic Press), 87–104.

Patil, B. S., Crosby, K., Byrne, D., and Hirschi, K. (2014). The intersection of plant breeding, human health, and nutritional security: lessons learned and future perspectives. *HortScience* 49, 116–127. doi: 10.21273/HORTSCI.49.2.116

Pattee, H. E., Isleib, T. G., Corbet, D. W., Ciesbrecht, F. G., and Cui, Z. (2001). Parent selection in breeding for roasted peanut flavor quality. *Peanut Sci.* 28, 51–58. doi: 10.3146/i0095-3679-28-2-3

Qannari, E. M. (2017). Sensometrics approaches in sensory and consumer research. *Curr. Opin. Food Sci.* 15, 8–13. doi: 10.1016/j.cofs.2017.04.001

Reynolds, A. G., Cliff, M., Wardle, D. A., and King, M. (2004). Evaluation of winegrapes in British Columbia: new cultivars and selections from Germany and Hungary. *Horttechnology* 14, 420–436. doi: 10.21273/HORTTECH.14.3.0420

Rizzo, G., and Baroni, L. (2018). Soy, soy foods and their role in vegetarian diets. *Nutrients* 10:43. doi: 10.3390/nu10010043

Roland, W. S. U., Vincken, J.-P., Gouka, R. J., Van Buren, L., Gruppen, H., and Smit, G. (2011). Soy isoflavones and other isoflavonoids activate the human bitter taste receptors hTAS2R14 and hTAS2R39. *J. Agric. Food Chem.* 59, 11764–11771. doi: 10.1021/jf202816u

Saldivar, X., Wang, Y.-J., Chen, P., and Mauromoustakos, A. (2010). Effects of blanching and storage conditions on soluble sugar contents in vegetable soybean. *LWT Food Sci. Technol.* 43, 1368–1372. doi: 10.1016/j.lwt.2010.04.017

Sands, D. C., Morris, C. E., Dratz, E. A., and Pilgeram, A. L. (2009). Elevating optimal human nutrition to a central goal of plant breeding and production of plant-based foods. *Plant Sci.* 177, 377–389. doi: 10.1016/j.plantsci.2009.07.011

Sirisomboon, P., Pornchaloempong, P., and Romphophak, T. (2007). Physical properties of green soybean: criteria for sorting. *J. Food Eng.* 79, 18–22. doi: 10.1016/j.jfoodeng.2006.01.022

Sujith Kumar, P. V., Basheer, S., Ravi, R., and Thakur, M. S. (2011). Comparative assessment of tea quality by various analytical and sensory methods with emphasis on tea polyphenols. *J. Food Sci. Technol.* 48, 440–446. doi: 10.1007/s13197-010-0178-y

Suwonsichon, S. (2019). The importance of sensory lexicons for research and development of food products. *Foods* 8:27. doi: 10.3390/foods8010027

Talavera-Bianchi, M., Chambers, I. V., E., and Chambers, D. H. (2010). Lexicon to describe flavor of fresh leafy vegetables. *J. Sens. Stud.* 25, 163–183. doi: 10.1111/j.1745-459X.2009.00249.x

Vela-Hinojosa, C., Escalona-Buendía, H. B., Mendoza-Espinoza, J. A., Díaz de León-Sánchez, F., Lobato-Ortíz, R., Rodríguez-Pérez, J. E., et al. (2018). Chemical and sensory analysis of native genotypes and experimental lines of tomato (*Solanum lycopersicum* L.). *Fruits* 73, 60–71. doi: 10.17660/th2018/73.1.7

Wang, K. (2018). East Asian food regimes: Agrarian warriors, edamame beans and spatial topologies of food regimes in East Asia. *J. Peasant Stud.* 45, 739–756. doi: 10.1080/03066150.2017.1324427

Wang, Y., and Kays, S. J. (2003). Analytically directed flavor selection in breeding food crops. *J. Am. Soc. Hortic. Sci.* 128, 711–720. doi: 10.21273/jashs.128.5.0711

Welch, R. M., and Graham, R. D. (2004). Breeding for micronutrients in staple food crops from a human nutrition perspective. *J. Exp. Bot.* 55, 353–364. doi: 10.1093/jxb/erh064

Williams, M. M. (2015). Phenomorphological characterization of vegetable soybean germplasm lines for commercial production. *Crop Sci.* 55, 1274–1279. doi: 10.2135/cropsci2014.10.0690

Wolfe, E., Popp, M., Bazzani, C., Nayga, R. M., Danforth, D., Popp, J., et al. (2018). Consumers' willingness to pay for edamame with a genetically modified label. *Agribusiness* 34, 283–299. doi: 10.1002/agr.21505

Wszelaki, A. L., Delwiche, J. F., Walker, S. D., Liggett, R. E., Miller, S. A., and Kleinhenz, M. D. (2005). Consumer liking and descriptive analysis of six varieties of organically grown edamame-type soybean. *Food Qual. Prefer.* 16, 651–658. doi: 10.1016/J.FOODQUAL.2005.02.001

Xu, Y., Cartier, A., Kibet, D., Jordan, K., Ivy, H., Davis, S., et al. (2016). Physical and nutritional properties of edamame seeds as influenced by stage of development. *J. Food Meas. Charact.* 10, 193–200. doi: 10.1007/s11694-015-9293-9

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. *J. Food Process. Technol.* 3, 1–6. doi: 10.4172/2157-7110.1000165

Young, G., Mebrahtu, T., and Johnson, J. (2000). Acceptability of green soybeans as a vegetable entity. *Plant Foods Hum. Nutr.* 55, 323–333. doi: 10.1023/A:1008164925103

Zhang, L., and Kyei-Boahen, S. (2007). Growth and yield of vegetable soybean (edamame) in Mississippi. *Horttechnology* 17, 26–31. doi: 10.21273/HORTTECH.17.1.26

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Sustainable Food Systems

ORIGINAL RESEARCH
published: 29 September 2020
doi: 10.3389/fsufs.2020.569625

Quality of Acid-Preserved Edamame Soybean at Immature and Mature Stages

David Moseley[1], **Leandro Mozzoni**[2]*, **Moldir Orazaly**[2], **Liliana Florez-Palacios**[2] and **Pengyin Chen**[3]

[1] Dean Lee Research and Extension Center, LSU AgCenter, Alexandria, LA, United States, [2] Department of Crop, Soil and Environmental Sciences, University of Arkansas, Fayetteville, AR, United States, [3] Fisher Delta Research Center, University of Missouri, Portageville, MO, United States

Edamame is a food-grade soybean that is harvested at the green-immature stage (R6) and sold fresh or frozen for consumption after steaming or boiling. Limited studies have been conducted on high-temperature sterilization of edamame in cans and on acid preservation of edamame. The objectives of this study were to evaluate the color retention and texture of edamame when pasteurized in acidic brine as compared to boiling, to assess the effect of sucrose and turmeric addition on acid-treated edamame, and to characterize varietal differences when edamame lines were pasteurized in an acidic brine at either R6 or R8 stage. All studies were conducted using industry-standard processing conditions for acid-preservation of food in glass containers. The results of this research indicated that acid processing caused losses in intensity of green color and hue, but much smaller than those reported when heat-processing edamame in cans. We also observed a small and borderline significant increase in texture ($p = 0.0790$) in acid-preserved samples. In addition, we found that green color is positively affected by the addition of turmeric to the brine, but not of sucrose. Finally, the varieties R07-589 (red-brown seed coat) and R09-345 (black seed coat) acid-processed at R8 had color and texture similar to canned in-kind substitute products. In conclusion, acid-preservation, and addition of turmeric to immature edamame, helped maintain an acceptable quality of the processed products.

Keywords: edamame, acidic-pasteurization, turmeric, color, texture

OPEN ACCESS

Edited by:
Martin Williams, United States Department of Agriculture, United States

Reviewed by:
Jacek Słupski, University of Agriculture in Krakow, Poland
Qiuying Zhang, Northeast Institute of Geography and Agroecology (CAS), China

***Correspondence:**
Leandro Mozzoni
lmozzon@uark.edu

Specialty section:
This article was submitted to Nutrition and Sustainable Diets, a section of the journal Frontiers in Sustainable Food Systems

Received: 04 June 2020
Accepted: 26 August 2020
Published: 29 September 2020

Citation:
Moseley D, Mozzoni L, Orazaly M, Florez-Palacios L and Chen P (2020) Quality of Acid-Preserved Edamame Soybean at Immature and Mature Stages. Front. Sustain. Food Syst. 4:569625. doi: 10.3389/fsufs.2020.569625

INTRODUCTION

Edamame is a food-grade soybean (*Glycine max* (L.) Merrill) that is high in protein and phytochemicals and low in saturated fats, making it a good food product for addition to soups and salads, or as healthy snack alternative to chips and candy (Masuda, 1991; Rayaprolu et al., 2015). Edamame is normally harvested by picking pods at the R6 physiological stage when the seeds are still green and fill 80–90% of the pods (Fehr et al., 1971; Konovsky et al., 1994; Shanmugasundaram and Yan, 2004). However, edamame can also be harvested when the seed is mature and dry at the R8 reproductive stage. While the seed color at R6 is green, it will either stay green or turn yellow, black, red-brown, or brown at the R7 stage, depending on the genetics of seed-coat and cotyledon color for the particular variety (Kiuchi et al., 1987). While all commodity-soybean have yellow seed coat, some of the edamame lines are selected to have green cotyledons to facilitate harvest, for these lines do not turn yellow if harvest is slightly delayed. Very few food-grade soybean lines are available with colored coats for commercialization. However, the pigments responsible for the various seed coat colors at the R8 reproductive stage can have health benefits, such as inflammation-preventing anthocyanins and procyanidins in the black, red-brown and brown seed coats (Takahata et al., 2001; Kim et al., 2006; Nizamutdinova et al., 2009).

Edamame is preserved in several ways. Currently, the market primarily consists of either fresh, frozen, roasted (Mentreddy et al., 2002), or freeze-dried products (Rayaprolu et al., 2015). While roasting and freeze-drying products have become commercial options, improved pasteurization techniques to create a shelf-stable product with high moisture content are necessary. Mozzoni et al. (2009a) and Czaikoski et al. (2013, 2018) attempted to conduct high-temperature sterilization using retorts; however, quality attributes of the end product, including texture and color are lower in thermal-processed edamame than in fresh or frozen products (Mozzoni et al., 2009a; Czaikoski et al., 2018). The standard for an edamame product is to have a firm texture with high green/low yellow intensity (Mozzoni et al., 2009a; Czaikoski et al., 2013; Rayaprolu et al., 2015). Czaikoski et al. (2013) estimated that adding 34.3 g mL^{-1} of sucrose can increase hue of pasteurized edamame. In addition, turmeric has been used in food preservation as a healthy alternative to synthetic food coloring of chicken breast filets (Abdeldaiem, 2014). Mozzoni et al. (2009a) found that adding $CaCl_2$ to the brine will increase firmness of the canned product, and that increasing pH of the brine will decrease intensity of green color (IGC), the later an undesirable attribute. The negative correlation between pH and IGC (or hue) is due to the acidity of the brine and heat process that subtracts a Mg^{2+} ion converting the chlorophyll pigment into pheophytin (von Elbe and Schwartz, 1996). Reducing the thermal processing time can limit this reaction, resulting in a higher hue value and intensity of green color (Czaikoski et al., 2013); however, this must be balanced with bromatological safety requirements. In contrast, if the pH of the brine is < 4.5, the edamame can be pasteurized instead of sterilized, requiring less thermal processing (Abbatemarco and Ramaswamy, 1994). Reduced thermal processing will result in a firmer texture (Czaikoski et al., 2013) and less break down of the chlorophyll (von Elbe and Schwartz, 1996). The objective of this research was, therefore, to study the color preservation and texture of edamame harvested at the R6 or R8 reproductive growth stages when processed using industry-standard acid-preservation methodologies.

MATERIALS AND METHODS

Edamame Varieties and Cultural Conditions

Three varieties (fixed factor), "R07-10397," "R09-345," and "R07-589," were grown in 2016 at the University of Arkansas Vegetable Research Station in Kibler, AR on a very fine sandy loam, coarse-silty, mixed, superactive, thermic, non-acid, typic udifluvents soil (Roxana series) (Soil Survey Staff, 2017) under standard agronomic practices as reported by Mozzoni et al. (2009b). Harvesting occurred once the plots reached the R6 or R8 physiological stage. In addition, a frozen sample of commercial edamame variety "8080," harvested at the R6 stage, was obtained from American Vegetable Soybean and Edamame, Inc. (Mulberry, AR).

Immature Edamame Harvest and Blanching

Once the plots were at R6 stage, entire plants were harvested and the pods were stripped using an edamame motive–power threshing machine (KE-6) (Doubletreasure Enterprise Inc., Plano, TX). The edamame was shelled from the pods using a "Little Sheller" (Taylor mfg. Co, Inc., Moultrie, GA), and samples were placed in a refrigerator at 4°C and subjected to blanching within 24 h. Blanching was conducted using a 100°C water bath for 90 s to reduce 99% initial lipoxygenase activity as reported by Mozzoni et al. (2009b) and confirmed by Xu et al. (2012). Since acid preservation was conducted once after all samples (R6 and R8) were harvested, the R6 samples were flash frozen with liquid nitrogen immediately after blanching to maintain cell structure during freezer storage (Luyet, 1968). Prior to the acid preservation, the frozen edamame samples were thawed by placing in an 82.2°C water bath consisting of 0.47 L of 5% distilled white all-purpose vinegar for 30 s and immediately cooled to ambient temperature by placing in cool water for 1 min.

Mature Edamame Seed Production

The same edamame varieties were also harvested at the mature (R8) reproductive stage. The varieties were chosen to represent different mature seed coat and cotyledon colors, namely "R07-10397" had green, "R07-589" red-brown, and "R09-345" black seed color. After harvesting, the samples were stored in a cool, dry place in cloth bags until processing.

Base Brine and Standard Acidic Processing Conditions

Previous research from this group (Mozzoni et al., 2009a) and others (Czaikoski et al., 2013) involved the execution of Central Composite Rotatable Designs to identify optimum brine composition for pasteurization of edamame. Based on these findings, our base brine consisted of a solution containing 0.56 L water, 0.4 L of 5% distilled white all-purpose vinegar, 60.0 g of NaCl, and 2.6 g of $CaCl_2$. The purpose of the vinegar was to bring the pH below 4.5 (Czaikoski et al., 2013) and the $CaCl_2$ was used to maintain a firm texture of the edamame after the thermal processing (Mozzoni et al., 2009a). Glass jars (236.6 ml) were filled with 148.8 g of shelled and blanched edamame and 88.7 ml of brine. The closed jars were placed three-quarters of the way into boiling water for 6 min for thermal processing. To ensure a commercially-sterile pasteurized product, the jars were tested for temperature and pH. Two test jars were opened immediately after the thermal processing to check for a minimum temperature of 85°C in the cold spot (McGlynn, 2000), located between 1/3 and 1/2 of the jar's height, following standard industry processing conditions for the selected glass jars, brine volume, and fresh weight of product utilized (Fellows, 2000; Mozzoni et al., 2009a). Brine pH was measured using a Symphony SP79P pH Meter (VWR, Radnor, PA) 2 weeks after processing. After thermal processing, the jars were immediately cooled to ambient temperature using tap water for 10 min. Thermal processing was done at the pilot plant of Bryant Preserving Co. in Alma, AR on August 31, 2016. Each treatment was subjected to three replications (random factor).

Acidic Preservation of Edamame at the R6 Reproductive Stage

Effect of Acidic Preservation on Edamame Quality

To test the effect of acidic preservation on edamame color and texture, a sample of edamame variety "8080" was subjected to acidic processing as described in section Base Brine and Standard Acidic Processing Conditions, and compared to a non-processed sample of the same variety that was cooked in boiling water for 6 min (the water remained boiling after the addition of the sample) on a stove top at the University of Arkansas' test kitchen.

Effect of Turmeric and Sucrose Addition on Post-processing Quality

The effect of sucrose and turmeric on color and texture was investigated using a commercial variety "8080" with a two-factor factorial experimental design with three levels for each factor. The factors and levels were sucrose (0, 29.5, and 59 g L^{-1}), and oleoresin turmeric (0, 0.26, 0.53 ml L^{-1}) added to the brine. Brine's pH was readjusted prior to the addition of the edamame sample, and acid processing was conducted as described in section Base Brine and Standard Acidic Processing Conditions. As control, a 60-g sample of frozen edamame "8080" was cooked in boiling water for 6 min.

Variation of Immature Edamame Varieties After Acidic Preservation

Three edamame varieties, "R07-10397," "R09-345," and "R07-589" were harvested and handled as described in section Immature Edamame Harvest and Blanching, and subjected to acid processing in jars as described in section Base Brine and Standard Acidic Processing Conditions. These acid-preserved samples were compared for differences in texture and color to a 60-g sample of edamame variety "8080" cooked in boiling water for 6 min.

Variation of Mature Edamame Varieties After Acidic Preservation

Prior to processing, the dry samples were soaked in distilled water for 24 h to promote uniform texture and expansion during the thermal process (Nordstrom and Sistrunk, 1977). Acidic preservation was conducted as described in section Base Brine and Standard Acidic Processing Conditions. The acid-processed product of the three colored varieties were compared to a non-processed control, where the variety was soaked in water for 24 h but not acid-treated, and to an in-kind commercial product. For the latter, R07-10397 (green seed) variety was compared to thawed and cooked samples of the commercial "8080" edamame variety, R07-589 (red-brown seed coat) was compared to canned samples of commercially-available Pinto and Kidney beans, and R09-345 (black seed coat) was compared to canned samples of commercially-available black beans. After processing, the cold spot temperature and the pH of the brine was examined to ensure successful preservation. The cold spot temperature was confirmed above 85°C, and the average pH of the jars were 4.28 with a standard deviation of 0.05.

Traits Assessed and Statistical Analysis

Texture

A single-bite test on a TMS 2000 texture analyzer (Food Technology Corp., Sterling VA, USA) with an Allo Kramer shear cell (10 blades) was used to measure texture. The settings of the instrument were max force at 50 kg, return distance at 40 mm, return speed at 3 mm sec^{-1}, and contact force at 500 g. The texture was reported as the peak force in Newtons (N) the blades required to penetrate the sample (Mozzoni et al., 2009a). Twenty grams of edamame were used for the texture analysis of R6 materials, and 10 gram-samples were used in the R8 stage materials due to an increase in firmness of the mature edamame. Texture analysis was not conducted on water-soaked treatment from 2.4 because the material was too hard for the load cell of our equipment.

Color

The color of the samples was measured with a HunterLab Color Flex (Hunter Associates Laboratory Inc., Reston, VA, U.S.A.). Three values were recorded, L^*, a^*, and b^* which represent the brightness/darkness, redness/greenness, and the blueness/yellowness of the sample, respectively. An increasing L^* value indicates a brighter sample, a smaller a^* value indicates a greener sample, and larger b^* value indicates a more yellow sample (Hunter Associates Laboratory Inc., Reston, VA, U.S.A). The instrument was calibrated with a black glass tile first, then with a white standard tile. The white tile had L^*, a^*, and b^* values of 93.76, −0.93, and 1.02, respectively. Prior to sampling, the calibration was validated with a green standard tile with values $L^* = 52.96$, $a^* = -25.30$, and $b^* = 13.71$. The intensity of green color (IGC) was calculated as ($-a^*/b^*$) (Mozzoni et al., 2009a), and Hue was calculated as (degrees(ATAN2)(a^*,b^*)) (Rayaprolu et al., 2015), respectively. The intensity of green color indicates the ratio of green to yellow, and the hue, measured in degrees, indicates how close the color is to pure red (0°), yellow (90°), green (180°), or blue (270°) (Lawless and Heymann, 1998).

Statistical Analysis

Experimental factors were analyzed with SAS v.9.4 (SAS Institute, 2014) using the PROC MIXED procedure. Least square means (LSM) of the main effects and their interactions were estimated with the Type 3 method and the means were separated by interpreting the *p*-values generated by the DIFF option.

RESULTS

Acidic Preservation of Edamame at the R6 Reproductive Stage

Effect of Acidic Preservation on Edamame Quality

When comparing the effect of processing on acidified brine to the samples cooked and non-further processed of variety

TABLE 1 | Color and texture of edamame cultivar "8080" when blanched and processed in acidic brine (Acid-treated) vs. blanched without further processing (Control).

Treatment	L†	a†	b†	IGC‡	Hue(°)	Force (N)
"8080" Acid-treated	51.05^{a}	1.63^{b}	37.26^{b}	-0.04^{b}	87.49^{b}	478.82^{a}
"8080" Blanched-control-	47.19^{a}	-10.25^{a}	26.22^{a}	0.39^{a}	111.36^{a}	435.44^{a}

Means followed by the same letters are not significantly different (p < 0.05) according to Fisher's protected least significant differences (LSD).
†L, a, and b values represent L, a*, and b* values.*
‡intensity of green color (–a/b*).*

TABLE 2 | Average color and texture for edamame variety "8080" after pasteurization in an acidic brine with different levels of the turmeric effect, and a 6-min boiled control not subject to acidic brine treatment that simulates the standard blanched-frozen-cooked product.

Turmeric level (ml L^{-1})	L†	a†	b†	IGC‡	Hue (°)	Force (N)
0	51.25a	1.96c	37.57b	−0.05c	87.01c	472.28a
0.26	51.01a	1.15b	47.02c	−0.03b	88.56b	452.46a
0.53	51.52a	0.94b	51.36d	−0.02b	88.88b	465.77a
Unprocessed control	42.94a	−10.62a	26.38a	0.40a	111.95a	436.03a

Means followed by the same letters are not significantly different (p < 0.05) according to Fisher's protected least significant differences (LSD).
†L, a, and b values represent L, a*, and b* values.*
‡intensity of green color (–a/b*).*

"8080," we observed that the acid treatment caused a borderline non-significant effect in reducing L*-value ($p = 0.1302$), significant effects on a*-value ($p = 0.0001$), b*-value ($p = 0.0050$), intensity of green ($p = 0.0001$), and hue ($p = 0.0001$), and borderline non-significant effect on texture ($p = 0.0790$) (**Table 1**). The processing in acid caused an undesirable decrease in hue and intensity of green color, but also caused a desirable increase of the texture of the processed samples.

Effect of Turmeric and Sucrose Addition on Post-processing Quality

Adding sucrose to the brine did not have a significant effect on color or texture of edamame "8080." Brine with 0.0, 29.5 or 59.0 g L^{-1} of sucrose resulted in Hue values of 88.68, 88.41, and 88.34, respectively, which are not significant at $\alpha = 0.05$. Similarly, the aforementioned sucrose treatments resulted in Force readings of 470.40, 454.40, and 465.70 N, respectively, also not-statistically different at $\alpha = 0.05$.

In contrast, addition of various levels of turmeric caused significant differences ($p < 0.05$) on a*, b*, hue, and intensity of green color, but texture remained unaffected (**Table 2**). Adding 0.53 or 0.26 ml L^{-1} of turmeric to the brine resulted in the lowest a* value, and the largest intensity of green color and hue value. The b* value was significantly smaller when adding turmeric at the level of 0.26 vs. 0.53 ml L^{-1}. When no turmeric was added, the edamame had a significantly lower intensity of green color and hue value ($p < 0.05$) (**Table 2**). The brine-processed edamame had L* value similar to the frozen and boiled check control, irrespective of turmeric levels; however, the frozen and boiled check control had a hue value that was closer to a pure green color than the brine-processed samples regardless of turmeric levels.

Variation of Immature Edamame Varieties After Acidic Preservation

Three varieties (R07-10397, R07-589, and R09-345) were harvested at the R6 reproductive stage while the pods were still green when processed on an acidified brine, and compared against edamame check 8080 cooked in boiling water. We observed that R07-589 had the lowest brightness (L*) value and greatest a* value, significantly different than the check variety 8080 and the other two edamame lines. Concomitantly, Hue and IGC values for R07-589 were lower than any of the other treatments. It is worth noting that the frozen and boiled edamame check 8080 had the greatest IGC and hue closer to true green (111.07° for 8080 vs. 120.00° for true green) than any of the acid-preserved edamame varieties (**Table 3**).

Variation of Mature Edamame Varieties After Acidic Preservation

R07-10397 (Green Seed Coat)

The hue value (91.69°) and IGC (0.029) for the processed product of R07-10397 at R8 was lower than the frozen sample of 8080 and an unprocessed sample of R07-10397, indicating less green color (**Table 4**). However, the hue value of R07-10397 after processing was similar to the processed edamame sample reported by Czaikoski et al. (2013) (93.50°), and was higher than 8080 (88.56°) after processing with sucrose and turmeric (**Table 4**). Furthermore, the texture of processed R07-10397 was not significantly different than "8080" (408.94 vs. 442.12 N, respectively).

R07-589 (Red-Brown Coat)

As the varieties R07-589 approach the R8 stage, the seed cotyledons and embryo gradually change from green to yellow, and the seed coats turn red-brown. The L* value of R07-589 (17.26) prior to processing was significantly lower ($p < 0.05$)

TABLE 3 | Color values and texture of three edamame varieties harvested at the R6 stage and acid-processed after pasteurization, and an unprocessed sample of a commercial variety blanched used as check.

Variety	L†	a†	b†	IGC‡	Hue(°)	Force (N)
"R07-10397"	46.75[a]	1.26[b]	38.72[c]	−0.04[b]	87.57[b]	310.08[c]
"R07-589"	34.27[b]	9.37[d]	19.86[a]	−0.47[c]	64.78[c]	355.11[b]
"R09-345"	45.94[a]	4.78[c]	23.95[ab]	−0.20[b]	78.70[b]	345.70[b]
"8080" boiled -control-	47.86[a]	−10.09[a]	26.04[b]	0.39[a]	111.07[a]	446.76[a]

Means followed by the same letters are not significantly different ($p < 0.05$) according to Fisher's protected least significant differences (LSD).
†L, a, and b values represent L, a*, and b* values.*
‡intensity of green color (–a/b*).*

TABLE 4 | Color and texture (force) values of three edamame varieties with colored seed coats harvested at the mature (R8) growth stage.

Effect	L†	a†	b†	IGC‡	Hue (°)	Force (N)
"R07-10397" (green coat)						
Non-processed§	46.18[a]	−7.01[b]	29.13[ab]	0.25[b]	103.96[b]	.
Acid preserved¶	49.08[a]	−0.69[c]	30.30[b]	0.03[c]	91.69[c]	408.94[a]
"8080" blanched	49.54[a]	−10.16[a]	26.92[a]	0.38[a]	110.92[a]	442.12[a]
"R07-589" (red-brown coat)						
Non-processed§	17.26[b]	13.67[a]	13.08[a]	−1.02[ab]	44.44[ab]	.
Acid preserved¶	21.81[ab]	19.16[c]	18.61[ab]	−1.02[ab]	44.38[ab]	461.73[a]
Pinto beans canned#	33.44[a]	17.21[b]	21.83[b]	−0.83[b]	50.39[b]	117.01[b]
Kidney beans canned#	28.51[a]	20.19[c]	18.83[b]	−1.11[a]	41.88[a]	112.15[b]
"8080" blanched	.	.	.	.	.	435.37[a]
"R09-345" (black coat)						
Non-processed§	10.3[a]	0.30[a]	−0.45[a]	−36.50[b]	−42.92[b]	.
Acid preserved¶	8.9[a]	6.52[b]	−0.19[a]	48.96[a]	2.07[ab]	476.75[a]
Black beans canned#	13.4[a]	9.53[c]	5.45[b]	−1.87[ab]	31.08[a]	77.81[c]
"8080" blanched	.	.	.	.	.	421.63[b]

Means followed by the same letters are not significantly different ($p < 0.05$) according to Fisher's protected least significant differences (LSD).
†L, a, and b values represent L, a*, and b* values.*
‡intensity of green color (–a/b*).*
§Dry seeds soaked in distilled water for 24 h.
¶Dry seeds, soaked in distilled water for 24 h, and subsequently acid-processed in brine consisting of 0.4 L of 5% distilled white all-purpose vinegar, 60.0 g of NaCl, and 2.6 g of $CaCl_2$.
#Commercially-available canned Pinto, Kidney, or Black beans.

than the canned product of pinto (33.44) and kidney beans (28.51); however, after processing, the L* value of R07-589 (21.81) and the pinto and kidney beans (33.44 and 28.51, respectively) were not significantly different. The hue value of the processed sample of R07-589 (44.38°) was not significantly different than the unprocessed sample and the canned samples of pinto and kidney beans (**Table 4**), but we observed the texture of acid-preserved R07-589 to be not-statistically different than "8080" check and significantly harder than that of canned Pinto and Kidney beans.

R09-345 (Black Coat)

As for the case of R07-589, as the mature embryo and cotyledons of R09-345 change color as the seed mature, and seed coats are fully black at R8 stage. We observed no differences in L* value for R09-345 between water-soaked or acid-preserved samples, nor the L* value was different than that of canned black beans (**Table 4**). However, texture of R09-345 preserved in acidified brine was significantly harder than that of "8080" edamame check, and about 6-fold greater than that of canned black beans (**Table 4**).

DISCUSSION

Mozzoni et al. (2009a) established a protocol to blanch edamame before sterilization to deactivate lipoxygenase activity; and developed a base brine consisting of NaCl and $CaCl_2$. Czaikoski et al. (2013) adapted the protocol set by Mozzoni et al. (2009a) by pasteurizing in an acidic brine and evaluating levels of sucrose to retain green color. Czaikoski et al. (2013) concluded that beans processed with sucrose were significantly greener than without; however, the processed product was significantly less green than the beans *in natura*. Although McGlynn et al. (1993) reported a brine below pH 4.5 can result in a firmer texture after thermal processing, Czaikoski et al. (2013) observed a product that was less firm than beans *in natura*.

The overall objective of this research was to improve the methodologies established by Mozzoni et al. (2009a) and Czaikoski et al. (2013), resulting in a product that would be commercially acceptable. Similar to Czaikoski et al. (2013, 2018), we observed a loss in green color of acid-processed edamame as compared to boiled samples; however, the color retention of samples is much larger than that of the color of heat-processed canned edamame reported by Mozzoni et al. (2009a). In Mozzoni's trial, intensity of green color dropped from 0.45 to −5.33 for blanching vs. canned product, or 5.78 absolute points; whereas in our current experiment the intensity of color is reduced from 0.39 to −0.04, or 0.43 absolute points. The samples processed under our acid conditions had greater color retention as compared to those processed in a retort by Mozzoni et al. (2009a).

Czaikoski et al. (2013) estimated adding 34.3 g L^{-1} of sucrose can increase the hue value by 1.17°. Although our research did not find sucrose to have a significant effect on color, the hue value dropped (less green) by 0.34° after adding 59 g L^{-1} of sucrose. The discrepancy between the two results may be due to the fact the suggestion of 34.3 g L^{-1} reported by Czaikoski et al. (2013) was a projection as it was outside of the central composite design to evaluate the effects of added sucrose. Abdeldaiem (2014) suggested adding turmeric can be a healthy alternative to artificial dyes, as it can have health benefits such as antioxidant and antimicrobial activities. Furthermore, Cleary and McFeeters (2006) inferred using turmeric in a pasteurization process can minimize off-flavors due to oxidation. In our research, we found the addition of 0.26 ml L^{-1} of turmeric would result in a hue closer to green than that without turmeric addition.

The variety R07-10397 (green seed coat and cotyledon), harvested mature at the R8 growth stage had the highest hue (most green); therefore, it may give the best chance to preserve an edamame product with acceptable color. This variety matures with a green cotyledon in addition to a green seed coat, which may increase the seed's ability to retain a higher green color after processing. The brightness (L*) value of R07-589 harvested mature and dry (R8 growth stage) improved after processing, in agreement with the report of Rayaprolu et al. (2015) that indicate an increase in brightness after processing. Furthermore, the hue value of R07-589 after processing was similar to that of the soaked sample, canned pinto beans, and canned kidney beans. These results indicate thermal processing in an acidic brine did not significantly alter the red-brown color of R07-589. Similarly, the L* value of the soaked and of the processed sample of R09-345 was not statistically different than the canned black beans, indicating that the thermal process in an acidic brine did not alter the black color of the beans. The retention of red-brown and black color for R07-589 and R09-345, respectively, after processing indicates the pigments causing the colors did not break down due to the thermal process or the acidic brine, leaving a product that would be aesthetically acceptable with healthy attributes. Lau et al. (2000) reported vegetables will soften during thermal processing. In this study, however, the texture of the processed edamame in the sucrose by turmeric test (R6 growth stage) and the three mature (R8) varieties were similar to the commercial check ("8080"). Maintaining the texture can be attributed to the addition of $CaCl_2$ (Mozzoni et al., 2009a) and decrease in duration of thermal processing (Czaikoski et al., 2013).

In conclusion, acid-preservation of edamame and addition of turmeric to the brine maintained the quality characteristics of color and texture of the processed product; traits that were otherwise further degraded when edamame was canned in retorts. Therefore, acid-preservation of edamame either at immature or mature stages could be a commercially viable alternative, provided bromatological studies are conducted to confirm sterility of end product.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

DM: conceptualization, methodology, investigation, formal analysis, writing—original draft, writing—review, and editing. LM: formal analysis, supervision, writing—review, and editing. MO: investigation and methodology. LF-P: writing—review and editing. PC: conceptualization, methodology, formal analysis, project administration, and supervision. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors thank Bryant Preserving Co in Alma, AR for their collaboration in this research, and the Arkansas Soybean Promotion Board for their sponsorship of this project.

REFERENCES

Abbatemarco, C., and Ramaswamy, H. S. (1994). End-over-end thermal processing of canned vegetables: effect on texture and color. *Food Res. Int.* 27, 327–334. doi: 10.1016/0963-9969(94)90188-0

Abdeldaiem, M. H. (2014). Use of yellow pigment extracted from turmeric (*Curcuma Longa*) rhizomes powder as natural food preservative. *Am. J. Food Sci. Technol.* 2, 36–47. doi: 10.12691/ajfst-2-1-6

Cleary, K., and McFeeters, R. F. (2006). Effects of oxygen and turmeric on the formation of oxidative aldehydes in fresh-pack dill pickles. *J. Agric. Food Chem.* 54, 3421–3427. doi: 10.1021/jf052868k

Czaikoski, K., Leite, R. S., Mandarino, J. M. G., Carrao-Panizzi, M. C., da Silva, J. B., and Ida, E. I. (2013). Canning of vegetable-type soybean in acidified brine: effect of the addition of sucrose and pasteurization time on color and other characteristics. *Ind. Crops Prod.* 45, 472–476. doi: 10.1016/j.indcrop.2012.09.009

Czaikoski, K., Leite, R. S., Mandarino, J. M. G., Carrao-Panizzi, M. C., da Silva, J. B., and Ida, E. I. (2018). Physicochemical characteristics of canned vegetable-type soybean processed with zinc at different pasteurization times. *Pesqui. Agropecu. Bras.* 53, 840–848. doi: 10.1590/s0100-204x2018000700008

Fehr, W. R., Caviness, C. E., Burmood, D. T., and Pennington, J. S. (1971). Stage of development descriptions for soybeans, *Glycine Max* (L.)

Merrill. *Crop Sci.* 11, 929–931. doi: 10.2135/cropsci1971.0011183X001100060051x

Fellows, P. (2000). "Heat Sterilization," in *Food Processing Technology: Principles and Practice, 2nd Edn*, ed P. Fellows (Boca Raton, FL: CRC Press LLC), 250–277. doi: 10.1201/NOE0849308871

Kim, E. H., Kim, S. H., Chung, J. I., Chi, H. Y., Kim, J. A., and Chung, I. M. (2006). Analysis of phenolic compounds and isoflavones in soybean seeds (*Glycine max* (L.) Merrill) and sprouts grown under different conditions. *Eur. Food Res. Technol.* 222, 201–208. doi: 10.1007/s00217-005-0153-4

Kiuchi, Y., Ishikawa, H., Nitta, N., and Sasaki, T. (1987). Collection of crop genetic resources and their characteristics in Iwate Prefecture In: characteristics of local soybean varieties. *Tohoku Nogyo Kenkyu* 40, 139–140.

Konovsky, J., Lumpkin, T. A., and McClary, D. (1994). "Edamame: the vegetable soybean," in *Understanding the Japanese Food and Agrimarket: a Multifaceted Opportunity*, ed A. D. O'Rourke (Binghamton, NY: Haworth Press), 173–181. doi: 10.1201/9781003075172-15

Lau, M. H., Tang, J., and Swanson, B. G. (2000). Kinetics of textural and color changes in green asparagus during thermal treatments. *J. Food Eng.* 45, 231–236. doi: 10.1016/S0260-8774(00)00069-8

Lawless, H. T., and Heymann, H. (1998). *Sensory Evaluation of Food.* New York, NY: Chapman and Hall. doi: 10.1007/978-1-4615-7843-7

Luyet, B. J. (1968). *Investigations on Freezing and Freeze-Drying of Selected Fruits and Vegetables.* Natick, MA: Technical report 69-8-FL. U.S. Army Natick Laboratories.

Masuda, R. (1991). "Quality requirement and improvement of vegetable-type soybean," in *Vegetable-Type Soybean: Research Needs for Production and Quality Improvement*, ed S. Shanmugasundaram (Taiwan: Asian Vegetable Research and Development Center), 92–102.

McGlynn, W. (2000). *The Importance of food pH in commercial canning operations. Division of Agricultural Sciences and Natural Resources. Oklahoma State University Cooperative Extension Service, Web.* Available online at: http://pods.dasnr.okstate.edu/docushare/dsweb/Get/Document-603/FAPC-118pod.pdf (acccessed November 14, 2017).

McGlynn, W. G., Davis, D. R., and Honarmand, F. (1993). Gluconic acid influences texture and color of canned asparagus. *J. Food Sci.* 58, 614–615. doi: 10.1111/j.1365-2621.1993.tb04338.x

Mentreddy, S. R., Mohamed, A. I., Joshee, N., and Yadav, A. K. (2002). "Edamame: a nutritious vegetable crop," in *Trends in New Crops and New Uses,* eds J. Janick and A. Whipkey (Alexandria, VA: ASHS Press), 432–438.

Mozzoni, L. A., Chen, P., Morawicki, R. O., Hettiarachchy, N. S., Brye, K. R., and Mauromoustakos, A. (2009b). Quality attributes of vegetable soybean as a function of boiling time and condition. *Int. J. Food Sci. Technol.* 44, 2089–2099. doi: 10.1111/j.1365-2621.2009.02038.x

Mozzoni, L. A., Morawicki, R. O., and Chen, P. (2009a). Canning of vegetable soybean: procedures and quality evaluations. *Int. J. Food Sci. Technol.* 44, 1125–1130. doi: 10.1111/j.1365-2621.2009.01929.x

Nizamutdinova, I. T., Kim, Y. M., Chung, J. I., Shin, S. C., Jeong, Y. K., Seo, H. G., et al. (2009). Anthocyanins from black soybean seed coats stimulate wound healing in fibroblasts and keratinocytes and prevent inflammation in endothelial cells. *Food Chem. Toxicol.* 47, 2806–2812. doi: 10.1016/j.fct.2009.08.016

Nordstrom, C. L., and Sistrunk, W. A. (1977). Effect of type of bean, soak time, canning media and storage time on quality attributes and nutritional value of canned dry beans. *J. Food Sci.* 42, 795–798. doi: 10.1111/j.1365-2621.1977.tb12605.x

Rayaprolu, S., Hettiarachchy, N., Aldoury, M., Cho, S., Moseley, D., and Chen, P. (2015). Physical and textural attributes of freeze-dried genetically modified and non-genetically modified soy beans. *J. Food Nutr. Sci.* 3, 119–125. doi: 10.11648/j.jfns.20150303.17

SAS Institute (2014). *The SAS System for Windows.* Cary, NC: Version 9.4. SAS Inst. Inc.

Shanmugasundaram, S., and Yan, M. R. (2004). "Global expansion of high value vegetable soybean," in *VII World Soybean Conference/IV International Soybean Processing and Utilization Conference/III Brazilian Soybean Congress* (Foz do Iguassu), 915–920.

Soil Survey Staff (2017). *Natural Resources Conservation Service, United States Department of Agriculture. Web soil survey.* Available online at: https://websoilsurvey.sc.egov.usda.gov/ (accessed December 11, 2017).

Takahata, Y., Ohnishi-Kameyama, M., Furuta, S., Takahashi, M., and Suda, A. (2001). Highly polymerized procyanidins in brown soybean seed coat with a high radical-scavenging activity. *J. Agric. Food Chem.* 49, 5843–5847. doi: 10.1021/jf010307x

von Elbe, J. H., and Schwartz, S. J. (1996). "Colorants," in: *Food Chemistry, 3rd Edn,* ed O. Fennema (New York, NY: Marcel Dekker), 651–722.

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. *J. Food Process. Technol.* 3:165. doi: 10.4172/2157-7110.1000165

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Sustainable Food Systems

ORIGINAL RESEARCH
published: 11 January 2021
doi: 10.3389/fsufs.2020.582834

Beyond Scale and Scope: Exploring Economic Drivers of U.S. Specialty Crop Production With an Application to Edamame

Clinton L. Neill[1]* *and Kimberly L. Morgan*[2]

[1] Department of Population Medicine and Diagnostic Services, Cornell University, Ithaca, NY, United States, [2] Department of Food and Resource Economics, University of Florida, Gainesville, FL, United States

Specialty crops are considered high-risk, high-reward, yet growers face differing, and relatively larger risk exposure when compared to traditional row crops. With traditional row crops, economies of scale and scope are key factors to increasing economic profitability. However, increasing economic profit for specialty crop operations present challenges which limit grower ability to easily take advantage of scale and scope economies. The authors discuss production, finance, regulatory, price, and human resource risks unique to U.S.-grown specialty crops. We apply our economic risk assessment framework to analyze U.S. edamame and present strategies to manage and mitigate risks faced by growers. We conclude that edamame may represent a profitable alternative crop in the U.S., and suggest future research topics are needed to optimize yields and meet market demand.

Keywords: risk assessment, economics, edamame, enterprise analysis, profitability

OPEN ACCESS

Edited by:
Xujun Fu,
Zhejiang Academy of Agricultural Sciences, China

Reviewed by:
Maria Ewa Rembialkowska,
Warsaw University of Life Sciences, Poland
David Moseley,
Louisiana State University Agricultural Center, United States

Correspondence:
Clinton L. Neill
cln64@cornell.edu

Specialty section:
This article was submitted to Nutrition and Sustainable Diets, a section of the journal Frontiers in Sustainable Food Systems

Received: *13 July 2020*
Accepted: *04 December 2020*
Published: *11 January 2021*

Citation:
Neill CL and Morgan KL (2021) Beyond Scale and Scope: Exploring Economic Drivers of U.S. Specialty Crop Production With an Application to Edamame. Front. Sustain. Food Syst. 4:582834. doi: 10.3389/fsufs.2020.582834

INTRODUCTION

Across the last two centuries, global farm operations have taken advantage of two economic principles—the economies of scale and scope. Combined with advances in research and grower adoption practices, scale, and scope economies allow growers to be more efficient while supplying food and fiber to meet consumer demand for product value and variety. However, specialty crop economics are fundamentally different than traditional row crops (i.e., corn, cotton, tobacco, grain soybeans, etc.) in many critical areas. Primarily, gains in specialty crop efficiencies and product diversity have larger opportunity costs—that is, the intrinsic value of the next best alternative is relatively greater—for specialty crops. We discuss these differences, and propose that increases in specialty crop profitability are driven by other economic factors that go beyond scale and scope.

The principle of scale economies is defined by the different types of production costs incurred, and a firm can cover those costs more efficiently by increasing the number of units produced. By reducing the marginal cost (cost per unit) of production, a firm can increase the amount of profit earned per unit (Stigler, 1958). An efficient producer aims to find ways to share fixed costs across additional units, or in the traditional row crop producer's case—acres (Heady and Ball, 1972). Fixed costs do not vary with output. For example, buildings/barns, planting and harvesting equipment, and land payments are costs that must be paid regardless of number of units produced. In contrast, variable costs are incurred once a decision is made to produce a unit/crop (Bressler, 1945).

Examples of variable costs are seed, fertilizer, and irrigation/water expenses. Total costs are the sum of fixed and variable costs across all crops produced on the farm. Traditional row crops are typically annual crops, that follow a cycle of planting, maintaining, and harvesting every year, allowing for some flexibility in the timing of these variable and fixed costs over the lifetime of the farm operation. Moreover, most row crops are processed or fed to livestock rather than for direct human consumption.

For specialty crops, calculating total costs depends on several complicating factors. Many of the specialty crops require relatively higher costs to establish orchards, vineyards, and bushes (called perennial crops), and most are marketed for direct human consumption. So, a grower's ability to take advantage of scale economies is inherently more expensive as compared to row crops. In the perennial crop case, multi-year investments are required, which economists refer to as sunk costs, that are unrecoverable costs once expended (Sutton, 1991). Many annual specialty crops (i.e., strawberries, onions, lettuce, etc.) require more sophisticated irrigation systems, access to added labor in a narrow timeframe, and additional processing facilities, which are not needed for row crops. Therefore, any decision to expand production to achieve scale economies by harvesting additional acres results in more of a financial burden on growers.

Economies of scope is based on the principle that goods/crops are easily interchanged within the production process (Fernandez-Cornejo et al., 1992). In the case of traditional row crop production, this could mean changing between varieties of the same crop or changing to an entirely different crop. To a large extent, this is a relatively easy transition for growers to make from year to year. Many growers already rotate corn and soybean production or cotton and peanuts, among many other alternatives, to achieve scope economies. Within limits, this rotation between crops involves lower costs and are accepted practices adopted by most traditional row crop operations. Similarly, changing between varieties is relatively cost-effective, as varieties are predominately bred for genetic resistance to chemicals. For specialty crops, this decision to switch between alternate crops is much more convoluted. Changing to a new crop when perennial crops are established requires significant financial capital to remove existing crops, prepare planting sites, purchase new crops, learn optimal management techniques, and seek out new markets. This may even result in a loss of expected future returns, if this decision is made while existing orchards/vineyards/bushes are in the middle of prime production years. With annual and perennial specialty crops, the infrastructure to properly plant, irrigate, and harvest are highly specialized for each type of crop, further complicating and restricting the decision to switch between crops. Annual specialty crops substitution decisions hinge heavily on the demand side of a market, and growers are faced with determining whether any buyers exist for a different variety. As for production, they must know if the growing requirements are different, which may require extensive research or hiring someone with specific crop experience. These issues plague specialty crop growers far more than traditional row crop growers.

While it is well-known that specialty crops are considered high-value crop production, growers also face relatively greater risk exposures, and increased production costs. In the next section, the authors discuss why and how economic principles that affect specialty crop growers are fundamentally different than traditional row crop operations. The discussion is framed within the five types of agricultural risk: production, financial, human resources, marketing, and legal/regulatory. We present an application of our agricultural risk assessment framework to U.S. edamame production. Finally, we discuss key economic questions that still need to be answered to provide U.S. edamame growers, and the specialty crop industry, with profitable alternatives to management and mitigation of these risks.

ECONOMIC DRIVERS AND RISK MANAGEMENT IN SPECIALTY CROP PRODUCTION

While the principles of scale and scope economies are well-understood by specialty crop growers, other economic issues are equally important. Factors such as technology (including the efficiency of that technology), training to ensure proper field use of equipment, ability and capacity for on-farm storage/processing, produce form (fresh vs. frozen), timing and geography of harvest, etc., play interrelated roles in impacting the economic profitability of specialty crop growers. Moreover, these factors may affect the level of risk growers face at any given time within the production process. Most conversations related to agricultural-related risks are centered around financial details. However, financial risk is interwoven within all five risks, and as we have outlined in our introduction, our goal is to provide a more comprehensive understanding of the unique portfolio of risks facing specialty crop growers.

Production Risk

Production risk refers to the uncertainties affecting the natural process of crop growth. Weather, disease, weed competition, pest infestations, introduction of new technology (bio-tech, robots, processing, etc.) and crop damage are common risks within the production process (Traxler et al., 1995). Production risk increases year-to-year variability in harvest amounts which, in turn, affects profit potential. There are three ways for growers to reduce risk in this area: (1) control or minimize risk through irrigation practices, regular machinery maintenance, and close monitoring for pests/disease; (2) reduce production variability by diversifying farm enterprises, creating flexibility to evolving economic conditions, integrating multiple enterprises, improving technology, and/ preparing contingency plans; and (3) transfer the risk to someone else through contracting or insurance (Drollette, 2009b). Specialty crops are especially vulnerable to production risk as many are grown and marketed either fresh or minimally processed to end users. Any defects in the final product means economic losses as accountability for market quality is placed on individual growers and/or packing houses by retailers. While production risk can never be eliminated, it may be mitigated to enhance the probability of higher profits.

Financial Risk

As mentioned earlier, financial risk is the most referenced type of risk, largely due to the fact that all other types of risk are intimately related to farm finances. Finance risk primarily occurs when a farm business borrows money or, more generally, creates an obligation to repay some form of debt. This type of risk is primarily a function of the timing of cash flows, to meet the daily operational needs of the farm, and the capital-intensive nature of the business (Kay and Edwards, 1994). Specialty crops growers balance their need for capital and labor expenses as both are significant financial factors for production and harvesting. Moreover, local and global conditions play a large role in the amount of financial risk farms face. For example, the cost of capital (the interest rate on loans), availability of funds, land rents, and leasing agreements are predominately local concerns. On a global scale, worldwide supply and demand, trade conditions, and agricultural subsidies across all countries impact the extent of financial risks facing the farm business. Optimal management of financial risks requires a structured, well-documented financial tracking system to optimize the timing of cash needs, expenses, and revenues (Morgan et al., 2016).

Regulatory/Legal Risk

Uncertainty that deals with governmental actions or legal rights and regulations is referred to as regulatory or legal risk. This type of risk involves topics ranging from business structures to insurance standards to tort liability (property and personal injury) to environmental legislation specific to resource use. Growers may successfully mitigate regulatory risk by implementing and following the practice of "due diligence," a technique that refers to taking reasonable actions to obtain all pertinent information (Harl, 2005). In choosing a business structure, consulting with a certified public accountant and a tax lawyer may ensure growers personal and farm business assets are identified accurately and legally secured. Growers may contact a local environmental regulatory agency to confirm that farm operational procedures are meeting, or exceeding, all current environmental standards and best management practices. Growers may research the latest legal rules and regulations specific to the geographic location of the farm operation, in conjunction with appropriate expert consultation, to discover ways and means to minimize legal risk. In 2019, the Federal Crop Insurance Coverage available for specialty crops exceeded $18 billion (USDA, 2020), yet the majority of specialty crops are not covered by the program and depend on the particular crop, the product form, and the production region (The National Agricultural Law Center, 2020).

Market/Price Risk

Market risk exposure depends on the type of agricultural commodity produced, but is focused on uncertainties about prices received for products and/or price of inputs. For most farm commodities, prices are dictated by national, or global market demands trends. An individual grower is a "price-taker" and cannot set product prices, instead receiving the price offered by the market on that day. In order to mitigate grower exposure to relatively lower market prices for homogenous products, many use hedging or storage to offset low prices (Drollette, 2009a; Broll et al., 2013).

Growers may decide to sell direct-to-consumers or retail outlets which gives an individual power over the market price through differentiation of the product offering (Nartea and Morgan, 2015). These "price-setters" manage price risks related to ever-changing market demand by connecting with individual customers. Direct customer tastes and preferences define the connections forged with individual growers and products, yet these are subject to variation over time. To mitigate price risk unique to direct market outlets, growers invest time and resources in a marketing plan, to include analysis of sales and market data, and create and cultivate market awareness of their unique product offerings. Growers who act as price-setters may mitigate financial risks by setting their prices above total costs of operations (Holdren, 1965) and earning marginal gains in profit relative to competitors. There are risks to both and it is up to the grower to decide which option best suits the short- and long-term goals of the farm operation.

Human Resources/Labor Risk

Every agricultural operation needs people in order to be successful. However, some agricultural operations require a larger labor force than others—i.e., specialty crop production. The share of variable costs attributed to labor for specialty crops is 2–3 times greater than many traditional row crops (**Figure 1**). As mentioned earlier, production risk is of utmost importance for specialty crop growers as the majority of these crops are intended for direct human consumption. In order to maintain integrity, many specialty crop operations employ a large labor force to maintain and harvest the crops. Machine harvesting of specialty crops often damages the product and increases losses to growers along the supply chain. Specialty crop growers are exposed to a relatively greater amount of labor risk exposure when compared to traditional row crops because it is heavily dependent on hand-harvesting the final product within a limited timeframe.

Labor risk also refers to the human health, personal relationship, and labor availability problems that directly affect the farm business (Billikopf, 2003). From family relationships to personnel management to health and communication, labor risk management requires development, and implementation of a plan to ensure access to a healthy and efficient workforce. Nearly all specialty crop growers rely on access to migrant laborers, whose availability is subject to the ever-changing political environment and extensive regulatory and legal paperwork (Mapes, 2010). With continued advances in agricultural mechanization, required numbers of hand laborers has reduced over time, particularly for specialty crops sold into processed or frozen product forms. However, hand harvesting is required for most fresh marketed fruits and vegetables, which earn greater market prices and improve overall grower profitability.

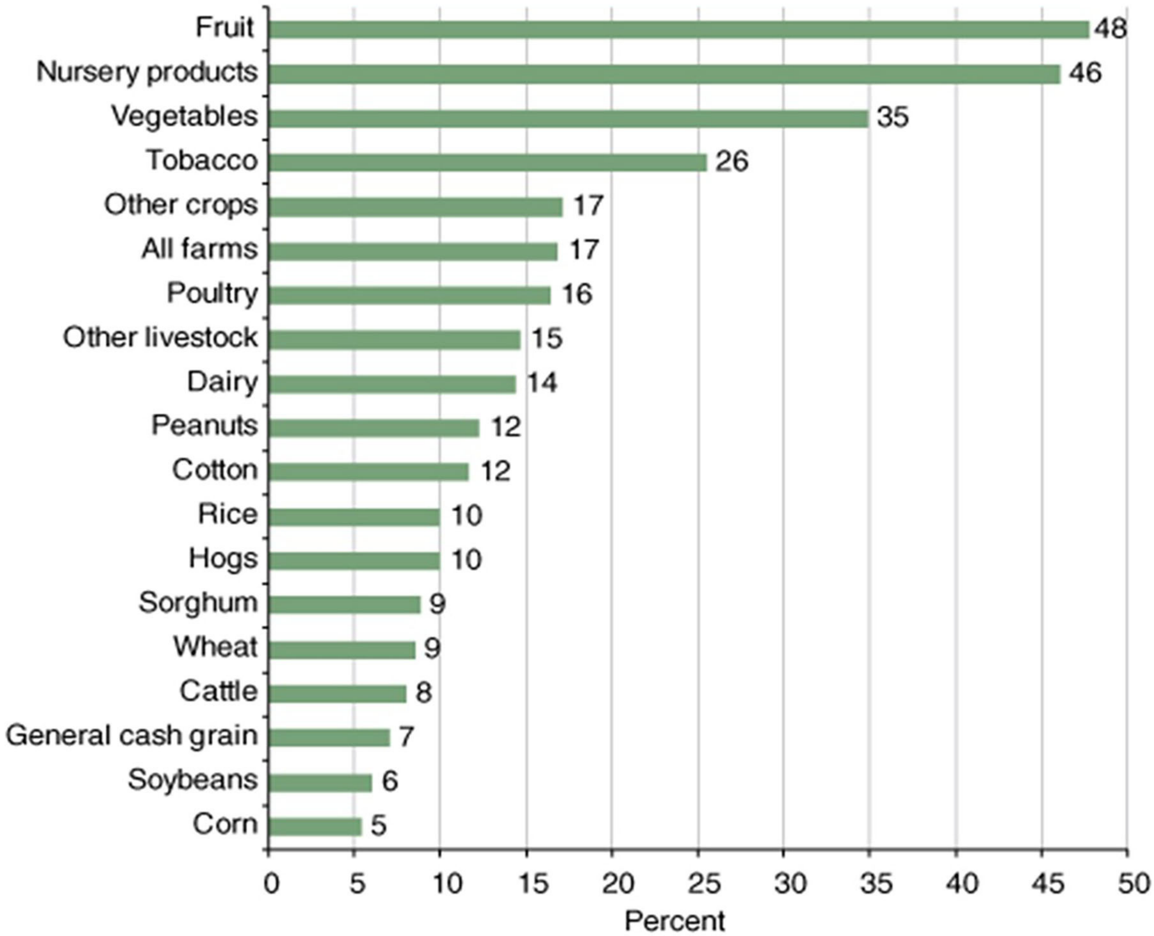

FIGURE 1 | Conmparison of hired labor costs as a share of production costs by farm sector.

ECONOMIC RISK ASSESSMENT OF U.S. EDAMAME PRODUCTION

Applying our risk assessment framework, we examined edamame, a relatively new U.S. specialty crop, to demonstrate the economic factors that distinguish specialty crop production. Edamame is harvested from the same plant as conventional grain soybeans, though there are some varietal differences. By harvesting beans/pods at an earlier growth stage (R6 and R7), edamame is bright green and has a higher moisture content. Edamame also has a larger seed size and higher simple sugar content that lends itself to a sweeter taste and improved digestibility. More importantly, edamame is classified as a specialty crop by the United States Department of Agriculture (USDA), while grain soybeans—harvested at a later growth stage of the same plant—is a traditional row crop.

Consumer demand for edamame has been growing at an estimated rate of 12–15% annually (Bernick, 2009; Edamame Production Facts, 2012). This growth in demand is largely driven from consumer desires for healthier, plant-based protein alternatives. Edamame is known for its high protein and essential amino acids content. In addition, it is a rich source of dietary fiber, minerals, and vitamins (Zhang et al., 2013). In recent times, edamame has become the second largest soyfood consumed in the U.S. at 25,000–30,000 tons annually (Soyfoods, 2015). All of these demand factors have influenced the heightened interest in domestic edamame production and influced the potential profitably of such an enterprise. As will be pointed out below, the keys to sustaninable production of edamame production to meet this growing demand are predominately around minimizing production costs and developing a clear market for fresh edamame.

The production risks associated with edamame that differ from conventional grain soybeans are predominately due to the lack of genetically modified edamame varieties available to growers, and edamame is marketed for direct human consumption. There are few approved herbicides for fresh market edamame post-emergence, resulting in increased need for hand labor to control for weeds which increases the financial risk to the farm business. The planting of non-genetically modified edamame requires that pest and weed resistance are bred through traditional plant breeding techniques. At the same time, a non-genetically modified edamame production has benefits to environmental sustainability of production that could mitigate this financial risk. Fresh market edamame is harvested when the beans have a much higher moisture content, which requires precise timing for harvest and staggered harvesting to avoid overwhelming harvest laborers. Moreover, traditional soybean harvesters cannot be used as they damage the edamame beans, thus modified equipment or hand harvesting must be employed. Once harvested, edamame must be cooled to maintain edible quality of the beans.

Mitigating each of these production risks requires additional equipment and storage facilities. Thus, these additional costs incurred when switching from traditional grain soybeans to edamame introduce financial risk. Additional equipment or buildings are capital investments that create financial obligations in the long term. With a potential increase in labor cost, edamame growers must consider how a change in cost allocation will affect economic profitability. Under a purely hand-harvested operation, labor costs would encompass over 62% of the operations expenses (Garber et al., 2019). Mechanical harvesting is an option as it could reduce labor costs to as low as 25% of total expenses. However, this increases equipment costs and maintenance while also increasing pod damage. Under current market varieties this increase in damage would lower potential revenue. If harvest damage can remain lower than 20% of total yield, then mechanical harvest is potentially profitable (Garber et al., 2019).

Enterprise budget analysis is a useful tool to estimate the potential profitability of various enterprises and variations within enterprises. A one-acre edamame operation in Virginia, with several assumptions about average inputs, that employs hand harvesting is presented in **Table 1**. Lord et al. (2019) found that hand harvesting is not profitable, and any potential revenue is not enough to cover variable costs. Moreover, the breakeven price received by the grower to cover total costs is $1.03 per pound, which exceeds current fresh market edamame market prices. We investigated machine harvesting of edamame pods, and found the operation is potentially profitable with a breakeven grower price of $0.53/pound, which is more realistic based on current market demand (**Table 2**). Enterprise budgets provide growers with the ability to understand the opportunity cost of choosing to grow one crop relative to the next best alternative use of available time and resources. This type of tool is a powerful planning tool for growers interested in edamame as a potential crop in their operation, and may be adjusted to capture the specific farm characteristics (Morgan et al., 2016).

The edamame example demonstrates the unique financial risks facing all specialty crop growers. The issue is exacerbated because of the additional regulatory constraints and the lack of a government safety net. Regulatory risk for specialty crop growers arises as products are typically used in human consumption, which requires growers and processors meet strict food safety guidelines both on and off farm. Following rules set out by legislation such as the Food Safety Modernization Act and Good Agricultural Practices are critical for market access and may significantly increase the cost of producing specialty crops. Traditional row crop growers have a more robust safety net from crop insurance subsidized by the federal government. While crop insurance coverage has grown over the last decade (see **Figure 2**), many specialty crops have little to no subsidization of insurance coverage for a specific crop in a specific geographic location. In many cases, the only option is Whole Farm Revue Protection, which may provide insufficient coverage or is too expensive for smaller operations. This lack of a safety net increases the risk exposure of every specialty crop producer when unexpected weather, disease, and pest pressures occur. Since edamame is a relatively new crop for domestic production, there is a long way to go before targeted insurance products may be available. Management and mitigation of these production risks unique to specialty crops requires relatively larger cash reserves to prevent long-term financial losses.

Market/Price risk is a key issue for all specialty crop growers. Edamame has only recently been produced in the United States for fresh markets. Traditionally, edamame was imported in the form of frozen product from Asian countries, where it originated, such as Taiwan and China (Born, 2006). Over the past 20 years, domestic U.S. production has been steadily increasing due to an increase in consumer demand for fresh edamame products (Lord et al., 2019). With the advent of fresh market opportunities comes the risk of increased price fluctuations. Early season edamame garners higher prices as there are fewer competitors in the market. As the season progresses, edamame market prices decrease at an increasing rate, and growers must decide to stop harvesting when the market price falls below their operational costs. To mitigate this financial risk exposure, growers may stagger production and consider early varieties/plantings to take advantage of the higher early-season prices. Additionally, the establishment of a local processing plant designed for edamame would alleviate the market risks. In Arkansas, a processing plant was established in 2012 which allowed for some growers to have access to a dedicated buyer which reduced the uncertainty of selling the product after harvest (CBS News, 2014).

While hand-harvesting is not generally a profitable way to produce edamame, there are likely situations that indicate when this is the best option for some growers. To determine the risks associated with hand harvesting, growers need answers to several key questions. Is there enough of a qualified labor force available that could be employed to harvest one's edamame crop? Are they available for the entire season? Can family labor be substituted for hired labor? One way to offset the issue of labor force availability is to pursue migrant/immigrant labor, but this creates another layer of risk. Constantly changing political climates apply pressure to the feasibility of migrant labor. In addition, migrant labor may not be cheaper than other alternatives. Small to medium sized edamame operations may benefit from hand-harvesting as the labor force needed is much smaller, but availability of labor fluctuates from year to year. Mitigating this risk requires adopting varieties that are more suited to mechanical harvesting or, hiring and training a smaller labor force that plays a more permanent role in an operation rather than seasonal/temporary workers. Training is the key to creating and maintaining a viable work force

TABLE 1 | Enterprise Budget for Average Farm in Virginia—Edamame, hand-harvested (Source: Garber et al., 2019).

			Quantity	Unit	Price/unit	Total
			Gross Receipts			
Fresh Edamame pods			1,875	lb	$ 0.75	$ 1,406.25
Fresh Edamame beans			5,625	lb	$ 0.50	$ 2,812.50
Total gross receipts						**$ 4,218.75**
			Preharvest costs			
Seed						
	Foodgrade seed		50	lb	$ 1.00	$ 50.00
Fertilizer						
	15-5-80					$ 43.65
Herbicides (Pre-emergent)						
	First Rate		0.6	oz	$ 38.50	$ 23.10
	Metalachlor		1.7	pt	$ 4.38	$ 7.45
Machinery variable costs						
	Fuel		10.8	gal	$ 1.75	$ 18.90
Labor						
	Spot spray for weeds		3.4	hr	$ 11.46	$ 38.96
Total preharvest costs						**$ 182.06**
			Harvest and marketing costs			
Harvest labor						
	Tractor operator labor		3.11	hr	$ 11.46	$ 35.64
	Tractor fuel		11.99	gal	$ 1.75	$ 20.98
Postharvest handling						
	Special labor		416	hr	$ 11.46	$ 4,767.36
	Packaging		7500	lb	$ 0.10	$ 750.00
Marketing charge			10%		of Gross Receipts	$ 421.88
Variable refrigeration cost			1	acre	$ 450.00	$ 450.00
Harvest damage risk			5%		of Gross Receipts	$ 210.94
Total harvest and marketing costs						**$ 6,656.80**
Total variable costs						**$ 6,838.86**
Return above variable costs						**$ (2,620.11)**
Breakeven price (variable)						**$ 0.89**
			Fixed costs			
Land rent						
	County	State Average	1	acre	$ 105.00	$ 105.00
Machinery and equipment			1	acre	$ 19.80	$ 19.80
General overhead				3%	of Variable Costs	$ 205.17
Refrigeration			1	acre	$ 657.00	$ 657.00
Total fixed costs						**$ 881.97**
Total costs						**$ 7,720.82**
Returns to land, capital, and management						**$ (3,502.07)**
Breakeven price				lb		**$ 1.03**

TABLE 2 | Enterprise budget for average farm in Virginia – Edamame, Machine-Harvested (Source: Garber et al., 2019).

			Quantity	Unit	Price/unit	Total
			Gross Receipts			
Fresh Edamame pods			1,500	lb	$ 0.75	$ 1,125.00
Fresh Edamame beans			6,000	lb	$ 0.50	$ 3,000.00
Total Gross Receipts						**$ 4,125.00**
			Preharvest Costs			
Seed						
	Foodgrade seed		50	lb	$ 1.00	$ 50.00
Fertilizer						
	15-5-80					$ 43.65
Herbicides (Pre-emergent)						
	First rate		0.6	oz	$ 38.50	$ 23.10
	Metalachlor		1.7	pt	$ 4.38	$ 7.45
Machinery variable costs						
	Fuel		10.8	gal	$ 1.75	$ 18.90
Labor						
	Spot spray for weeds		3.4	hr	$ 11.46	$ 38.96
Total preharvest costs						**$ 182.06**
			Harvest and marketing costs			
Harvest labor						
	Tractor operator labor		10.36	hr	$ 11.46	$ 118.73
	Tractor fuel		39.98	gal	$ 1.75	$ 69.97
Postharvest handling						
	Special labor		20	hr	$ 11.46	$ 229.20
	Packaging		7,500	lb	$ 0.10	$ 750.00
Marketing charge			10%		of Gross Receipts	$ 412.50
Variable refrigeration cost			1	acre	$ 450.00	$ 450.00
Harvest damage risk			20%		of Gross Receipts	$ 825.00
Total harvest and marketing costs						**$ 2,855.39**
Total variable costs						**$ 3,037.45**
Return above variable costs						**$ 1,087.55**
Breakeven price (Variable)						**$ 0.38**
			Fixed costs			
Land rent						
	County	State Average	1	acre	$ 105.00	$ 105.00
Machinery and equipment		1	acre	$ 61.72	$ 61.72	
General overhead		8%			of Variable costs	$ 243.00
Refrigeration			1	acre	$ 657.00	$ 657.00
Total fixed costs						**$ 961.72**
Total costs						**$ 3,999.17**
Returns to land, capital, and management						**$ 125.83**
Breakeven price				lb		**$ 0.53**

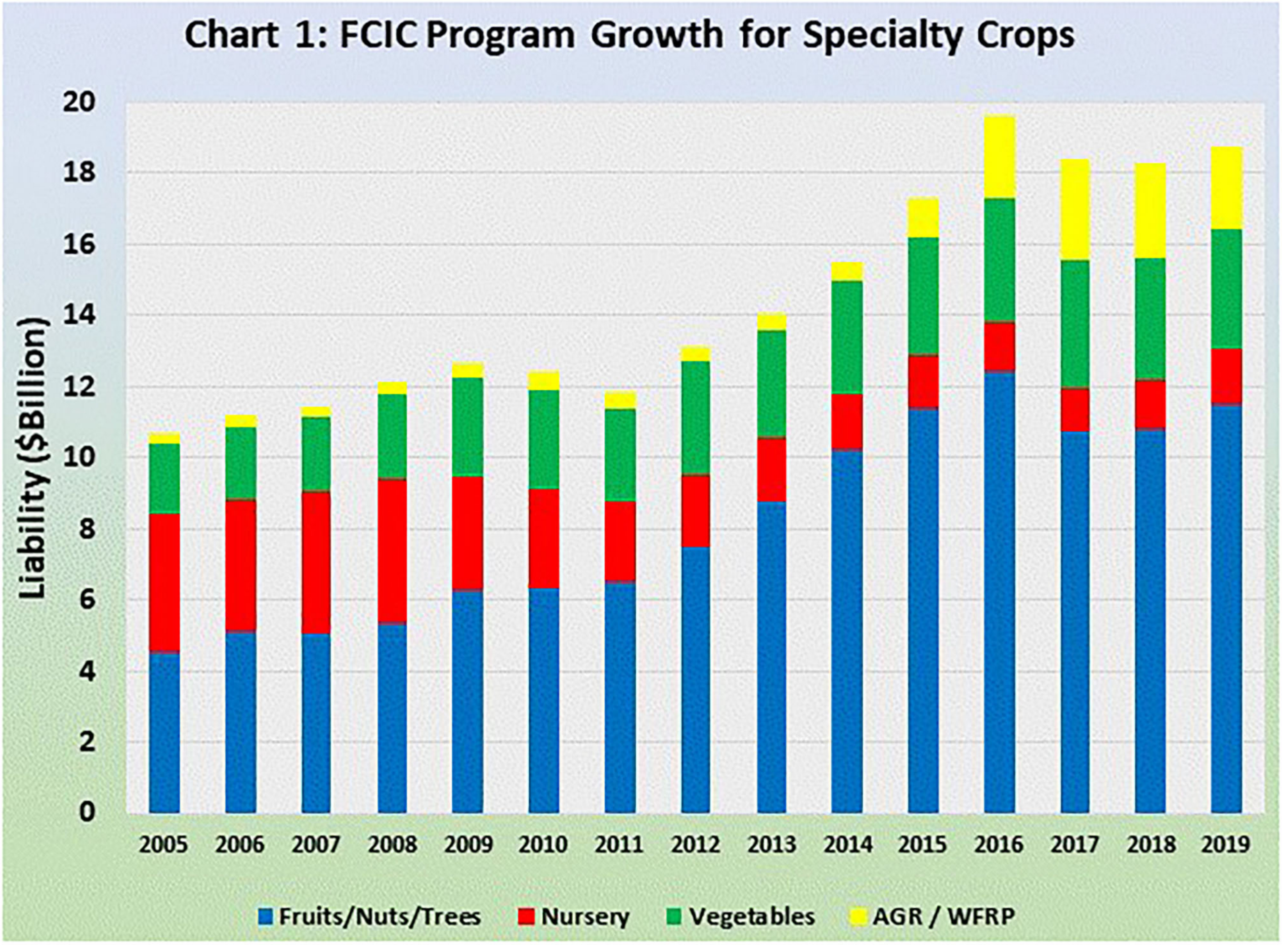

FIGURE 2 | Federal Crop Insurance Coverage Program Growth for specialty crops (Sources USDA -Risk Management Agencvy, 2020).

on farm, but those workers must be retained in order to be effective. Overall, labor risk will continue to be a critical issue in U.S. edamame production, as in all of specialty crop enterprises.

DISCUSSION AND CONCLUSIONS

We present an economic risk assessment framework and an application to U.S. edamame production. Overall, edamame is a interesting case study as the growing demand for the crop is driving the need to better understand risks associated with specialty crop production. Specifically, U.S. production of edamame can meet growing consumer preferences for fresh, healthy, plant-based protein alternatives. However, there are several limitations to sustainable domestic production. To assist in promoting sustainable production we identify five areas of agricultural risk based on underlying economic concepts, along with potential strategies to manage and mitigate risks behind and beyond the farm gate. What is apparent about U.S. edamame production is the many unknowns about the long-term viability of such an enterprise. As with all businesses, profitability of an enterprise is of utmost importance, but a lack of existing data on new enterprises is always a challenging issue. To better answer the numerous questions posed by our findings, we encourage improved data collection specific to each risk category to inform future research. This includes grower-informed surveys to better understand the key issues that limit expansion of U.S. edamame production, and public availability of market price trends within production regions.

Industry partnerships may provide an opportunity to enhance the data and knowledge of the potential profitability of edamame production. This requires industry to "buy-in" and provide needed transparency and information sharing. Creation of public-private partnerships is crucial to ensuring that specialty crops continue to be profitable to all supply chain participants. These partnerships would ensure that the research conducted within public institutions is grounded in application and focused on improving grower profitability and long-term business survival. Moreover, partnering provides better decision and management tools, which are especially relevant when considering farm enterprise diversification with new crops like edamame. As noted in the enterprise analysis, many assumptions are made about the production of edamame in Virginia. Any number of those assumptions may be incorrect and/or require adjustments as edamame production becomes more widespread. Industry partnerships would ensure that those assumptions are reduced so that potential edamame growers better predict the expected benefits of producing this crop on their farms. This is especially relevant for the small to medium sized growers with tighter operating budgets and limited access to human resources.

As mentioned in the regulatory risk assessment, there are no comprehensive government-supported or private insurance policies available to all specialty crop growers. This exposes most specialty crop growers to more risk as compared to traditional row crop growers. By coordinating efforts across different specialty crop grower associations, lobbying for more effective policies may be achieved. To reiterate, we encourage a collective lobbying effort to benefit all specialty crop growers. Without collective effort, new specialty crop growers—like U.S. edamame growers—will not have the economic incentives to explore these novel enterprises.

As we have shown, profitable specialty crop production goes far beyond the traditional thoughts about scale and scope economies. Edamame production in the U.S. has great potential, but incentivizing farmers to grow the crop requires a deep understanding of the added economic risks. Mitigating each of the risks presented in this analysis are vital to the success of U.S. grown edamame. By analyzing these issues in future research, we expect U.S. edamame production to represent a profitable enterprise for growers.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

CN drafted the manuscript and edited based on co-author and reviewers' comments. All authors reviewed, edited, and approved draft and submitted versions of the manuscript.

FUNDING

This work was funded, in part, by USDA-NIFA, Grant No. 2018-51181-28384, Accession No. 1016465 and the Virginia Agricultural Experiment Station.

ACKNOWLEDGMENTS

The authors thank USDA-NIFA and Virginia Agricultural Experiment Station for the financial support.

REFERENCES

Bernick, K. (2009). *Edamame Takes Root in U.S. Corn and Soybean Digest.* Available online at: http://www.cornandsoybeandigest.com/edamame-takes-root-us (accessed on April 18, 2020).

Billikopf, G. E. (2003). *Labor Management in Agriculture: Cultivating Personnel Productivity*. University of California, Division of Agriculture and Natural Resources, Agricultural Issues Center.

Born, H. (2006). *Edamame: Vegetable Soybean. National Sustainable Agriculture Information Service*. Available online at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.669.8757andrep=rep1andtype=pdf (accessed on April 18, 2020).

Bressler, R. G. (1945). Research determination of economies of scale. *J. Farm Econom.* 27, 526–539. doi: 10.2307/1232647

Broll, U., Welzel, P., and Wong, K., Kit, P. (2013). Price risk and risk management in agriculture. *Contempor. Econom.* 7, 17–20.

CBS News (2014). *Edamame Production Goes From Asia to Arkansas*. Available online at: https://www.cbsnews.com/news/asias-edamame-business-is-coming-to-us-soil/ (accessed on November 30, 2020).

Drollette, S. (2009a). *Managing Marketing Risk in Agriculture*. Logan, UT: Department of Applied Economics Utah State University

Drollette, S. A. (2009b). *Managing Production Risk in Agriculture*. Logan, UT: Department of Applied Economics Utah State University.

Edamame Production Facts (2012). Available at: https://www.uaex.edu/farmranch/cropscommercial-horticulture/edamame/docs/EdamameFacts1.pdf (accessed on November 18, 2020).

Fernandez-Cornejo, J., Gempesaw, C. M., Elterich, J. G., and Stefanou, S. E. (1992). Dynamic measures of scope and scale economies: an application to German agriculture. *Am. J. Agricult. Econom.* 74, 329–342. doi: 10.2307/1242487

Garber, B., Neill, C., and Lord, N. (2019). *Edamame: Costs, Revenues, and Profitability*. Virginia Cooperative Extension AAEC-189P.

Harl, N. E. (2005). *Agricultural Law Manual*. Eugene, OR: Agricultural Law Press.

Heady, E. O., and Ball, A. G. (1972). *Public Policy Means and Alternatives*. Ball, A; Gordon Size, Structure, and Future of Farms.

Holdren, B. R. (1965). Competition in food retailing. *J. Farm Econom.* 47, 1323–1331. doi: 10.2307/1236390

Kay, R. D., and W. M. Edwards. (1994). *Farm Management*. 3rd ed. New York, NY: McGraw-Hill.

Lord, N., Neill, C., and Zhang, B. (2019). Production and economic considerations for fresh market Edamame in Southwest Virginia. *Virginia Cooperative Extension AAEC-188P*.

Mapes, K. (2010). *Sweet Tyranny: Migrant Labor, Industrial Agriculture, and Imperial Politics*. Urbana, IL: University of Illinois Press.

Morgan, K. L., Callan, P. L., Mark, A., Niewolny, K., Nartea, T. J., Scott, K. H. (2016). *Farm Financial Risk Management Series. Part I, Overview of Financial Systems for New and Beginning Farmers*. Virginia Cooperative Extension AAEC-114P.

Nartea, T. J., and Morgan, K. L. (2015). *Selling Directly to Buyers: How to Price Your Products.*

Soyfoods (2015). *Soy Products Sales and Trends*. Available online at: http://www.soyfoods.org/soyproducts/sales-and-trends (accessed on November 18, 2020).

Stigler, G. J. (1958). The economies of scale. *J. Law and Econom.* 1, 54–71. doi: 10.1086/466541

Sutton, J. (1991). *Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration*. MIT press.

The National Agricultural Law Center, N.A. Specialty Crops (2020). *An Overview.* University of Arkansas. Available online at: https://nationalaglawcenter.org/overview/specialty-crops/ (accessed on April 18, 2020).

Traxler, G., Falck-Zepeda, J., Ortiz-Monasterio R, J. I., and Sayre, K. (1995). Production risk and the evolution of varietal technology. *Am. J. Agricult. Econom.* 77, 1–7.

USDA, Risk Management Agency. (2020). *Specialty Crops. Risk Management Agency Fact Sheet*. Available online at: https://rma.usda.gov/en/Topics/Specialty-Crops.

Zhang, B., Zheng, C., Zeng, A., and Chen, P. (2013). *Vegetable Soybean (edamame): Production, Processing, Consumption and Health Benefits*. In: Agricultural research updates. P. Gorawala and S. Mandhatri (ed.). Nova Science Publishers, Inc., Hauppauge, NY.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 12 January 2021
doi: 10.3389/fpls.2020.612106

A Modified Roger's Distance Algorithm for Mixed Quantitative–Qualitative Phenotypes to Establish a Core Collection for Taiwanese Vegetable Soybeans

Chung-Feng Kao[1,2], Shan-Syue He[1†], Chang-Sheng Wang[1,2†], Zheng-Yuan Lai[1], Da-Gin Lin[3] and Shu Chen[4]*

[1] Department of Agronomy, College of Agriculture and Natural Resources, National Chung Hsing University, Taichung, Taiwan, [2] Advanced Plant Biotechnology Center, National Chung Hsing University, Taichung, Taiwan, [3] Biotechnology Division, Taiwan Agricultural Research Institute, Taichung, Taiwan, [4] Plant Germplasm Division, Taiwan Agricultural Research Institute, Taichung, Taiwan

OPEN ACCESS

Edited by:
Xujun Fu,
Institute of Crop and Nuclear Technology Utilization, Zhejiang Academy of Agricultural Sciences, China

Reviewed by:
Yingpeng Han,
Northeast Agricultural University, China
Henrik Schumann,
Universität Bonn, Germany

***Correspondence:**
Chung-Feng Kao
kaoc@nchu.edu.tw;
gcf6@hotmail.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: *30 September 2020*
Accepted: *08 December 2020*
Published: *12 January 2021*

Citation:
Kao C-F, He S-S, Wang C-S, Lai Z-Y, Lin D-G and Chen S (2021) A Modified Roger's Distance Algorithm for Mixed Quantitative–Qualitative Phenotypes to Establish a Core Collection for Taiwanese Vegetable Soybeans. Front. Plant Sci. 11:612106. doi: 10.3389/fpls.2020.612106

Vegetable soybeans [*Glycine max* (L.) Merr.] have characteristics of larger seeds, less beany flavor, tender texture, and green-colored pods and seeds. Rich in nutrients, vegetable soybeans are conducive to preventing neurological disease. Due to the change of dietary habits and increasing health awareness, the demand for vegetable soybeans has increased. To conserve vegetable soybean germplasms in Taiwan, we built a core collection of vegetable soybeans, with minimum accessions, minimum redundancy, and maximum representation. Initially, a total of 213 vegetable soybean germplasms and 29 morphological traits were used to construct the core collection. After redundant accessions were removed, 200 accessions were retained as the entire collection, which was grouped into nine clusters. Here, we developed a modified Roger's distance for mixed quantitative–qualitative phenotypes to select 30 accessions (denoted as the core collection) that had a maximum pairwise genetic distance. No significant differences were observed in all phenotypic traits (p-values > 0.05) between the entire and the core collections, except plant height. Compared to the entire collection, we found that most traits retained diversities, but seven traits were slightly lost (ranged from 2 to 9%) in the core collection. The core collection demonstrated a small percentage of significant mean difference (3.45%) and a large coincidence rate (97.70%), indicating representativeness of the entire collection. Furthermore, large values in variable rate (149.80%) and coverage (92.5%) were in line with high diversity retained in the core collection. The results suggested that phenotype-based core collection can retain diversity and genetic variability of vegetable soybeans, providing a basis for further research and breeding programs.

Keywords: vegetable soybean, edamame, germplasm, modified Roger's distance, mixed quantitative-qualitative phenotypes, core collection, diversity, multiple imputation

INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] is a legume crop that is rich in protein, oil, and other nutrients such as lecithin and isoflavones. It is one of the most important economic crops worldwide (Liu, 1997). Owing to the high nutritional value, soybeans are good plant-based protein foods. The soy protein is regarded as a complete protein because it contains essential amino acids in animal proteins (Velásquez and Bhathena, 2007). Mainly, soybeans are classified into grain soybeans and vegetable soybeans according to their harvest period. Briefly, grain soybeans harvested at reproductive stage 8 (R8) are processed as oil products, animal feeds, and soy products. On the other hand, vegetable soybeans, also known as edamame, that were harvested at reproductive stage 6 to 7 (R6–R7) are taken as snacks and vegetables (Fehr, 1971; Young et al., 2000; Reddy et al., 2019).

In contrast to grain soybeans, vegetable soybeans are characteristic of the green-colored pods and seeds, larger seeds (over 30 g/100 seeds), smooth texture, less beany flavor, and higher contents of vitamin A, vitamin C, sucrose, and starch (Saldivar et al., 2010; Zhang et al., 2013). The demand for vegetable soybeans has increased as a result of changes in eating habits nowadays. Vegetable soybeans are consumed as frozen edamame or vegetables in Taiwan, Japan, China, and South Korea (Konovsky et al., 1994; Hu et al., 2007). This kind of soybean tastes sweet because it was picked at a high sugar level, and it contains high calcium and approximately 13% of protein (Dwevedi and Kayastha, 2011). Due to the abundant phytochemicals in edamame, it is known as a functional food (Mebrahtu, 2008). In recent years, the awareness of food nutritive value and health benefit brings a steady increase in the demand and also the planted area of vegetable soybeans in the United States (Jiang et al., 2018). Therefore, vegetable soybeans are good nutritional crops with high commercial value.

Soybeans are rich in isoflavones and anthocyanins, which prevent or inhibit the progression of cancer and other chronic diseases (Magee et al., 2004; Messina, 2016; Ziaei and Halaby, 2017). Evidence showed that breast cancer incidence is much lower in Asians than other populations resulting from the intake of isoflavones in soybeans in their daily diet (Adlercreutz, 2002; Mense et al., 2008). Anthocyanins are neuroprotective agents in the central nervous system, participating in mechanisms of suppression of neuroinflammation and oxidative stress (Zafra-Stone et al., 2007; Zhang et al., 2019). High isoflavone and anthocyanin contents of soybeans are value-added nutrition, as well as soy products. It was pointed out that isoflavone and anthocyanin contents in soybeans can be influenced by genotypes, food processing, and cultivated environment (Kim et al., 2014; Miladinović et al., 2019). On top of that, anthocyanins are abundant in black soybean seed coat, and isoflavones are positively correlated to plant height, effective branches, number of pods per plant, and number of seeds per plant (Choung et al., 2001; Zhang et al., 2014). Our vegetable soybean germplasms provide detailed information of those phenotypic traits, which could be quality data for isoflavones and anthocyanins, for establishing a core collection (CC).

Apart from the amount of nutrient composition in vegetable soybeans, the quality of vegetable soybeans depends on the external appearance (Shanmugasundaram, 1991). To market, a good-quality vegetable soybean consists of large seeds, bright green pods, light color pubescence (white to gray), and two seeds per pod at least (Johnson et al., 1999). The pod color is affected by genotype, planting density, and processing procedure (Zeipiòa et al., 2017). Seed size is a measure of seed length, width, and thickness (Xu et al., 2011). One hundred-seed weight, affected by seed size, is significantly correlated with pod length and pod width (Xu et al., 2011). Besides, a narrow leaflet is related to a greater number of three- and four-seeded pods (Bravo et al., 1980; Frank and Fehr, 1981; Shanmugasundaram, 1991). Additionally, soybean is a chilling-sensitive crop, and it reduces the growth potential when the temperature is below 20°C (Hofstra, 1972). The chilling temperatures (about 15°C) are an unfavorable condition that would affect the growth and the development of soybeans, in particular, the early stage of the growth. In the flowering stage, it may cause flower abscission and pod setting failure. In the pod and grain filling stage, grains would fill poorly (Kurosaki and Yumoto, 2003). Low temperatures result in the reduction of total seed weight per plant and of the pod number per plant in soybeans (Ikeda et al., 2009). Low temperatures in the flowering stage cause browning and cracking of the seed coat in soybeans, leading to lower market value due to poor appearance quality (Sunada and Ito, 1982; Githiri et al., 2007). In the past, during 1960–1985, farmers in Taiwan grew grain soybeans mostly, and breeders under the Council of Agriculture (COA) conducted crossbreeding programs resulting in several soybean varieties; therefore, most of our previous germplasms were for grain soybeans. However, the cheap imported soybean replaced most of the Taiwanese self-production and provided the most requirement of the soybean industry in Taiwan. Since the 1980s, our soybean production has turned to the goal of producing high-quality vegetable soybeans, and that is the reason why few collections mainly aimed for breeding vegetable soybean varieties by the 1990s. A large-scale soybean germplasm collection and renewal was started in 1986 led by National Chung-Hsing University with soybean breeders from other institutions, including the Asian Vegetable Research and Development Center (AVRDC), the Taiwan Agricultural Research Institute (TARI), National Chiayi University, and many local agricultural experiment stations under COA. The collections were from around Taiwan island including wild species and introducing those from our technological teams around the world. These results enriched our germplasm collection in various ways, genetic variation and quantity included. Additionally, for some of them, their origin can be recognized using the accession number (Lin, 1997). To make it usable for practical plant breeding, we decided to generate a CC from germplasms of vegetable soybean.

The concept of a CC, defined as a limited set of accessions, was first introduced by Frankel (1984). The CC is representative of the whole accessions, with minimum repetitiveness and maximum diversity. There are many studies on CC for soybean (Wang et al., 2008; Qiu et al., 2013; Guo et al., 2014), but no CC for vegetable soybean. This is because vegetable soybean accessions are quite hard to acquire. The major reason is that vegetable soybean is harvested at an immature stage and thus requires specific planting to collect seeds for a special purpose. In particular, after vegetable soybean is harvested, there will be no grain soybeans. Also, a small consumer market and less attention are other reasons. To date, there are only a few studies on vegetable soybean germplasm (Jiang et al., 2018). The first genetic diversity investigation of vegetable soybean was by Mimura et al. (2007). They used 122 alleles detected by 17 simple sequence repeats (SSRs) to investigate the genetic diversity among 130 accessions (107 from Japan, 10 from China, and 13 from the United States) of vegetable soybeans and found that Japanese edamame have a narrow diversity. Later, Zhang et al. (2013) used 22 expressed sequence tag-derived SSRs selected from grain soybean to access the genetic diversity of 48 vegetable soybean accessions (43 from China, 3 from Japan, and 2 from the United States) and found a narrow genetic base in Chinese vegetable soybean accessions (Zhang et al., 2013). Both studies only calculated genetic diversity of vegetable soybean accessions without establishing a CC. A possible reason may be due to the small collection of accessions and a narrow genetic base. A narrow genetic base may narrow variability coverage and further limit vegetable soybean breeding and variety improvement, as vegetable soybean requires a broad and diverse collection of accessions to produce high-quality products for the global market. To enrich the genetic diversity, it is necessary to introduce more foreign soybean varieties. With the nutritional value and benefit for neurological health in vegetable soybeans, it is vital for us to raise attention to the preservation of vegetable soybean resources.

The establishment of a CC is based on similarity among accessions. Similarity is generally defined as a distance-based measure in a multidimensional space. The similarity measures for quantitative traits are computed by using *Euclidean* distance and *Manhattan* distance, which are well-defined (Aggarwal et al., 2001). However, it is not straightforward for qualitative traits. To do this, a frequent-based approach (Rogers, 1972) can be applied for the purpose of calculating genetic distance among accessions for qualitative traits. A larger distance between two accessions represents dissimilarity between two accessions. It is important to extract information, for given accessions, accurately from phenotypic traits using correct similarity metrics, so that the interpretation of findings from trait-based studies is on a theoretically sound basis.

k-means has long been used for clustering. However, there is some weakness for *k*-means. First, the original version of *k*-means had the difficulty to handle qualitative data. Different initial seeds may lead to distinct results. It is also difficult to determine the number of clusters due to its nature of a supervised algorithm (Sajidha et al., 2020). By applying the dummy variable for qualitative traits, *k*-means can be modified as weighted *k*-means clustering (Huang et al., 2005; Foss and Markatou, 2018), so that it is able to deal with qualitative and quantitative traits at the same time. In addition, the weight of the noise can be reduced by determining the weight of qualitative and quantitative traits. The present study used weighted *k*-means to perform clustering for individual phenotype and for the overall phenotypes, in a base of unsupervised data learning, to determine a proper number of clusters automatically by the data learning.

Missing data in phenotypes are often seen in many germplasms due to problems with cultivation or negligence in investigation (Singh et al., 2019). Deletion of accessions with missing phenotypes and the utilization of simple imputation methods (e.g., mean or median substitution, regression imputation) should not be used to calculate trait-based diversity, because both methods do not take into account differences of functional plant traits between accessions (Taugourdeau et al., 2014). Multiple imputation estimated missing phenotypes by chained equations and produced a lower error and bias on the estimation of missing phenotypes as uncertainty is taken into account during computational process (Penone et al., 2014). In particular, it is superior and robust to deal with no more than 30% threshold of missing phenotypes (Taugourdeau et al., 2014). Multiple imputation requires high performance and a parallel computing system, especially for categorical variables.

In the present study, we proposed a modified Roger's distance algorithm for mixed quantitative–qualitative phenotypes to establish a CC for breeders using 213 Taiwanese vegetable soybean germplasms. To downsize germplasms without losing much diversity (Frankel and Arber, 1984), we first constructed multiple imputation to predict missing phenotypes for constructing a completed (i.e., observed plus imputed values) dataset of phenotypes (i.e., observed plus imputed values), so that the weighted *k*-means clustering can be applied to search for the patterns and the genetic structure of germplasms in vegetable soybean. This provides an opportunity for a better understanding of genetic diversity among complex characteristics in accessions (Yun et al., 2015). We then performed a modified Roger's distance algorithm using observed phenotypes of vegetable soybean germplasms to calculate pairwise genetic distance separately for quantitative and qualitative phenotypes among accessions to construct the CC of Taiwanese vegetable soybeans, followed by diversity calculation of individual phenotype and CC assessment.

MATERIALS AND METHODS

Vegetable Soybean Germplasm

The germplasm of vegetable soybean composed of 213 accessions used in the present study was collected from the National Plant Genetic Resources Center (NPGRC) in Taiwan. The countries of origin of the accessions were grouped into Taiwan, Japan, United States, China, Hong Kong, and Philippines. Phenotypic traits were investigated and recorded in the field at Kaohsiung District Agricultural Research and Extension Station, COA, and followed the guidelines of distinctness, uniformity, and stability (DUS) test for vegetable soybean in four consecutive seasons (1995–1998 autumn). Each accession was characterized for 47 phenotypic traits in relation to morphology (38 traits), growth (5 traits), phenology (2 traits), and production (2 traits). For detailed agronomic descriptors, please refer to **Supplementary Table 1**. The phenotype data were in stacked format that cannot be directly used for statistical analysis. We first unstacked the phenotype data and found that no accessions have missing values for all 47 phenotypic traits. Thirteen pairs of accessions were identically redundant within and among all the phenotypic traits; hence, 13 accessions were randomly selected and excluded from each pair of accessions. In the present study, a total of 200 accessions with 29 phenotypic traits were retained for analysis, based on a threshold of slight to mild missing rate ($<30\%$) in phenotype. There are 15 quantitative traits [seed length (mm), seed width (mm), seed thickness (mm), 100 seed weight (g), plant height (cm), leaflet length (cm), leaflet width (cm), pod length (cm), pod width (cm), stem length to first pod (cm), shelling rate (%), immature seed length (mm), immature seed width (mm), immature seed thickness (mm), and 100 immature seed weight (g)] and 14 qualitative traits (seed shape, seed coat color, hilum color, hypocotyl coloration, stem color, number of branches, leaflet size, leaflet shape, leaf color, plant type, pubescence density, pubescence color, corolla color, and pod set capacity). We denoted the 200 germplasm accessions as the entire collection (EC), where 29 phenotypic traits were used to construct a CC.

Meteorology Data

The meteorology data, including temperature (°C), humidity (%), duration of sunshine (hours), precipitation (mm), and days with precipitation (days), were collected daily during 1995–1998 from the Kaohsiung District Agricultural Research and Extension Station and the Kaohsiung Weather Station of the Central Weather Bureau in Taiwan. One-way ANOVA was conducted to test for differences among four autumn seasons (October–November) and among 4 years, followed by paired samples t-test only when ANOVA p-value reached significant difference. The results suggested that no significantly environmental effects were observed on the phenotypes investigated in the field at Kaohsiung District Agricultural Research and Extension Station (**Supplementary Table 2**).

Multiple Imputation of Incomplete Phenotypes

Multiple imputation is a mathematically computational method to deal with missing values in phenotypes (Lee and Simpson, 2014). There were three steps, consisting of imputation, estimation, and pooling, in multiple imputation. In the imputation step, missing values of a phenotypic trait were imputed by using Bayesian linear regression based on maximum completed data from the remaining traits to generate a completed dataset (i.e., observed plus imputed values). We repeated this procedure several times (say 30 times) to create multiple completed datasets in order to capture the uncertainty during the procedure. The model treated all variables (i.e., traits) as fixed and all error terms as random to maintain the natural variability in the data. In the estimation step, standard statistical analysis was performed for each of the completed datasets. In the pooling step, we aggregated the results derived from each of the completed data analyses. The three steps guarantee a more accurate and reliable data to obtain valid conclusions for downstream analysis (Royston, 2004). The mi package in R was used in the analysis. The values of qualitative traits were estimated from posterior predictive distribution, while quantitative traits were from predictive mean matching (Su et al., 2011; Kropko et al., 2017). Compared to other packages, Bayesian linear regression was applied in the mi package, which sorts out the issue of separation for qualitative traits. It can examine the collinearity in the data automatically. Four independent chains (i.e., chained equations) were used to account for the uncertainty between and within the completed datasets during missing data imputation. An $\hat{R}$ statistic was calculated to evaluate the iterative convergence (default is smaller than 1.1) of the multiple imputation (Su et al., 2011).

Weighted *k*-Means Clustering

To examine the diversity of vegetable soybean germplasms, weighted k-means approach by kamila package in R (Lloyd, 1982; Huang et al., 2005; Foss et al., 2016; Foss and Markatou, 2018) was used to explore the feature of the germplasms. The weighted k-means clustering can handle both qualitative and quantitative data. Let $X = \{X_1, X_2, \ldots, X_n\}$ be a set of n vegetable soybean accessions. Assume that each accession has m phenotypes. The objective function $P(A, W, C)$ is calculated as

$$P(A, W, C) = \sum_{l=1}^{k}\sum_{i=1}^{n}\sum_{j=1}^{m} a_{il} w_j d(x_{ij}, c_{lj})$$

where a_{il} is an $n \times k$ assignment matrix, which subjects to

$$a_{il} \in \{0, 1\}, \quad 1 \le i \le n, 1 \le l \le k$$

$$\sum_{l=1}^{k} a_{il} = 1, \quad \forall i$$

$a_{il} = 1 \Leftrightarrow x_i$ belongs to the lth cluster

w_j is the weight parameter, which is defined by

$$w_j = \begin{cases} \text{w}, & \text{if j} \in \text{qualitative trait} \\ 1 - \text{w}, & \text{if j} \in \text{quantitative trait} \end{cases},$$

where w is an adjustable parameter, and we selected weight as 0.5 by default. $d(x_{ij}, c_{lj})$ represents the distance between x_{ij} and c_{lj}. Let x_{ij} be the value of the jth phenotype in the ith accession, and c_{lj}be the centroid of the lth cluster in the jth phenotype. The distance of the quantitative trait is measured by Euclidean distance $d(x_{ij}, c_{lj})$

$$d(x_{ij}, c_{lj}) = (x_{ij} - c_{lj})^2.$$

The qualitative trait is converted to dummy variable (0 or 1). For instance, a qualitative phenotype $Y = \{Y_1, Y_2, \ldots, Y_n\}$ with p different trait levels is expanded to p indicator variables, denoted as $I_A(y)$

$$I_A(y) = \begin{cases} 1, \text{ if } y_i \text{ equals to the specific level} \\ 0, \text{ otherwise} \end{cases}.$$

Thus, Euclidean distance can be applied to measure the distance. The aim is to minimize the objective function $P(A, W, C)$

during the procedure of iteration. To select the optimal number of clusters, we set criteria as follows: Shannon diversity index is above 90%, Nei's diversity index is above 85%, and the proportion of variance explained is at least 70%.

Modified Roger's Distance

We proposed modified Roger's distance to construct a CC, which is a small collection with minimal redundancy and maximal representative of the ES. The modified Roger's distance is a suitable algorithm for mixed quantitative–qualitative phenotypes. The distance (d^2_{Eul}) for quantitative phenotype j ($\in C$)can be expressed as

$$d^2_{Eul} = \sum\nolimits_{j \in C} \sum\nolimits_{i \neq i^*} (x_{ij} - x_{i^*j})^2,$$

where C is the quantitative phenotype, and x_{ij} and x_{i^*j} represent the value of the ith and i^*th accession in the jth trait, respectively. The distance (d^2_{mR}) for qualitative phenotype j ($\in Q$) is defined as,

$$d^2_{mR} = \begin{cases} 0, & \text{if the same trait level between i and i*} \\ \sum\limits_{j \in Q} \sum\limits_{i \neq i^*} (\log f_{ij} - \log f_{i^*j})^2, & \text{if different trait level between i and i*} \end{cases}$$

Where Q is the qualitative phenotype, and f_{ij} and f_{i^*j} represent the frequency of ith and i^*th accession in jth trait, respectively. Thus, the modified Roger's distance can be of the form $d^2 = d^2_{Eul} + d^2_{mR}$.

Phenotypic Diversity

Phenotypic diversity is calculated using the Shannon–Weaver diversity index (H') and Nei's diversity index (Shannon and Weaver, 1962; Nei, 1973). The Shannon–Weaver diversity index (H') is defined as

$$H' = \frac{\sum_{i=1}^{S} p_i \ln(p_i)}{\ln(S)},$$

and Nei's diversity index is defined as

$$1 - \sum_{i=1}^{S} p_i^2,$$

where p_i is the proportion of accessions in the ith cluster to the total number of germplasms and S is the total number of clusters. The value of H' reflects the degree of evenness of germplasms. The higher H', the more evenness. $H' \in [0,1]$. The value of Nei's diversity index is bounded between 0 and $(1 - \frac{1}{S})$, which really depends on the number of clusters.

Assessment of the CC

To evaluate whether the CC is representative of the EC, five properties of the CC can be used to assess the characteristics and diversity structure of the traits. These properties include (1) the mean difference percentage ($MD\% = \frac{1}{m} \sum_{i=1}^{m} \frac{|M_e - M_c|}{M_c} \times 100\%$), (2) the variance difference percentage ($VD\% = \frac{1}{m} \sum_{i=1}^{m} \frac{|V_e - V_c|}{V_c} \times 100\%$), (3) the coincidence rate ($CR\% = \frac{1}{m} \sum_{i=1}^{m} \frac{R_c}{R_e} \times 100\%$), (4) the variable rate ($VR\% = \frac{1}{m} \sum_{i=1}^{m} \frac{CV_c}{CV_e} \times 100\%$), and (5) coverage ($Coverage\% = \frac{1}{m} \sum_{i=1}^{m} \frac{D_c}{D_e} \times 100\%$), where M, V, R, CV, D, and m are the mean, variance, range, coefficient of variation, the number of clusters, and the number of traits, respectively. As for the subscript, e is short for the EC while c is short for the CC. We considered the CC to be representative of the EC if (1) MD% is small and the percentage of significant mean difference ($\alpha = 0.05$) between the CC and the EC in all 29 phenotypic traits is no more than 20%, and (2) the CR% is greater than 80% (Hu et al., 2000; Kim et al., 2007).

Levene's test was conducted, with 1,000 bootstrap iterations, to check homogeneity of variance among groups (Levene, 1960; Joseph et al., 2020). After testing for homogeneity of variance, we then checked the difference between the EC and the CC. For quantitative traits, traits with equal and unequal variances were, respectively, performed by Student's t-test and Welch's t-test (Delacre et al., 2017). For quantitative traits, Chi-squared test of homogeneity was performed to see whether germplasms of each phenotype was significant difference in distribution between the EC and the CC (Koehler and Wilson, 1986) (p-values < 0.05).

RESULTS

A total of 213 vegetable soybean germplasms were used to establish the CC for effective utilization. Thirteen redundant accessions were removed from the analysis, because these accessions, with different names or designators, have identical phenotypes in all traits to other varieties. There is no accession having missing values in all traits. After germplasm quality control, 200 accessions remained and regarded as entire collection (EC).

Our data suffered, to a certain extent, from missing rates, ranging from 0.5 to 26.5% (**Supplementary Table 3**). These missing phenotypes were estimated by chained equations in multiple imputation. The value of $\hat{R}$ statistic for each of the traits was smaller than 1.1 on the second-stage multiple imputation, indicating converged estimates on imputed phenotypes (**Supplementary Table 4**). **Tables 1**, **2** demonstrate behavioral patterns of central tendency and variability for quantitative traits and frequency distributions of trait characteristics for qualitative traits, respectively, for observed and imputed phenotypic traits. All imputed phenotypes, except stem color and plant type having slight differences, showed non-significant differences (p-values > 0.05) among observed and imputed phenotypic traits, suggesting reliable and accurate imputed values.

We conducted weighted k-means clustering method, using completed phenotypes (i.e., observed plus imputed values on 29 mixed quantitative–qualitative phenotypic traits), to search for the population structure of the whole picture on the EC. **Figure 1** demonstrates that our vegetable germplasms can be grouped into nine clusters, using weighted k-means clustering algorithm for 29 mixed-type phenotypic traits. We calculated modified Roger's

TABLE 1 | Difference tests of quantitative traits between observed phenotypes and imputed phenotypes.

Phenotypic trait	Observed phenotypes		Imputed phenotypes[a]		Difference test (*p*-value)[b]
	N	**Mean ± s.d.**	***N***	**Mean ± s.d.**	
Seed length (mm)	200	8.96 ± 0.78	200	8.96 ± 0.78	1.00
Seed width (mm)	200	8.28 ± 0.59	200	8.28 ± 0.59	1.00
Seed thickness (mm)	200	7.09 ± 0.67	200	7.09 ± 0.67	1.00
100 seed weight (g)	200	33.41 ± 7.51	200	33.41 ± 7.51	1.00
Plant height (cm)	198	37.09 ± 10.71	200	37.01 ± 10.69	0.94
Leaflet length (cm)	200	10.6 ± 7.61	200	10.6 ± 7.61	1.00
Leaflet width (cm)	200	7.22 ± 1.18	200	7.22 ± 1.18	1.00
Pod length (cm)	149	4.59 ± 0.53	200	4.56 ± 0.5	0.62
Pod width (cm)	149	1.22 ± 0.28	200	1.21 ± 0.24	0.71
Stem length to first pod (cm)	150	11.85 ± 4.41	200	11.4 ± 4.3	0.34
Shelling rate (%)	150	56.79 ± 6.97	200	57.02 ± 6.48	0.75
Immature seed length (mm)	149	15.49 ± 1.63	200	15.61 ± 1.61	0.51
Immature seed width (mm)	150	11.04 ± 3.13	200	11.03 ± 3.49	0.96
Immature seed thickness (mm)	150	8.07 ± 0.82	200	8.11 ± 0.85	0.67
100 immature seed weight (g)	149	69.05 ± 14.87	200	69.57 ± 13.74	0.74

N, number of germplasms; s.d., standard deviation. [a]Multiple imputation was used to estimate missing phenotypes. [b]Student's t-test was conducted to test difference among observed and imputed phenotypes.

distance for mixed quantitative–qualitative phenotypic traits to capture pairwise genetic distance among accessions in the EC. The larger genetic distance, the much dissimilar. As a result, 30 accessions (**Table 3** and **Supplementary Material 1**) were selected from the EC and regarded as CC, showing maximal genetic distance among accessions. Levene's test was first applied to examine homogeneity of variances for each of phenotypic traits, followed by the difference test to compare whether there is difference between the CC and the EC. Our results showed that no significant difference (p-values > 0.05) was observed between the CC and the EC in all traits, except plant height (**Tables 4**, **5**), indicating that the central tendency and variability of phenotypic traits in the CC were consistent with the EC.

Table 6 showed cluster and diversity comparison between the CC and the EC for individual traits in vegetable soybean. The Shannon–Weaver diversity index of individual phenotypic traits in the CC and the EC ranged from 0.68 to 0.98 and from 0.64 to 0.97 in quantitative traits and ranged from 0.41 to 1.00 and from 0.27 to 0.99 in qualitative traits, with an overall average of 0.88 and 0.83, respectively. Similarly, Nei's diversity index in the CC and the EC ranged from 0.46–0.89 and 0.50–0.89 in quantitative traits and ranged from 0.15 to 0.75 and from 0.09 to 0.74, with an overall average of 0.66 and 0.66, respectively. Compared to the EC, our CC suggested that approximately 79.31% (23 out of 29) of phenotypic traits demonstrated higher or equal diversities on Shannon–Weaver diversity and/or Nei's diversity indices, indicating that about 80% phenotypic diversities were retained. It is worth noticing that three traits (100 seed weight, pod width, and immature seed width) in the CC demonstrated more evenness than those in the EC, resulting in increase in diversities in spite of losing clusters. About 21.69% (6 out of 29) of phenotypic traits demonstrated a mild degree (2–9%) of diversity lost in the CC due to the loss of clusters and/or reduced evenness among clusters. Taking all 29 phenotypic traits together, a high overall diversity (Shannon–Weaver diversity index is 0.91, Nei's diversity index is 0.85) was observed in the EC, and a reasonably high overall diversity (Shannon–Weaver diversity index is 0.83, Nei's diversity index is 0.81) was retained in the CC (data not shown), showing a mild degree (4–8%) of overall diversity lost in the CC. This suggested that our selected CC is representative of diversity from the EC.

The CC selected by modified Roger's distance gave a small value of MD% (6.35%), an intermediate value of VD% (52.90%), a large value of CR% (97.70%), a very high value of VR% (149.80%), and a very high percentage of coverage (94.76% for qualitative traits, 90.38% for quantitative traits, and 92.5% for the combined) (**Table 7**). These properties of the CC suggested that the CC retained valuable characteristics and large variability for phenotypic traits in the EC. In addition, the percentage of significant difference was 3.45%, which is less than the threshold of 20%, indicating a very low significant difference between the EC and the CC in traits. This suggested that the CC selected by the modified Roger's distance was representative of the EC.

DISCUSSION

To the best of our knowledge, our vegetable soybean germplasms consisted of 213 accessions, which are the largest collection worldwide. Vegetable soybean becomes popular just for more than one decade and mostly in East Asia, e.g., Japan, Taiwan, China, and South Korea; therefore, other countries take merely a small amount of production; the majority of vegetable soybeans are imported from Asia (Wang, 2018; Carneiro et al., 2020). This is why it did not attract the attention of breeders until this century. Hence, there are only limited accessions available

TABLE 2 | Difference tests of qualitative traits between observed phenotypes and imputed phenotypes.

Phenotypic trait	Observed phenotypes	Imputed phenotypes[a]	Difference test (p-value)[b]
	***N* (%)**	***N* (%)**	
Seed shape			1.00
Round	107 (53.5)	107 (53.5)	
Oblate	21 (10.5)	21 (10.5)	
Oval	58 (29)	58 (29)	
Flat	13 (6.5)	14 (7)	
Seed coat color			1.00
Yellowish white	26 (13)	28 (14)	
Yellow	69 (34.5)	73 (36.5)	
Green	92 (46)	97 (48.5)	
Pale brown	1 (0.5)	1 (0.5)	
Reddish brown	1 (0.5)	1 (0.5)	
Hilum color			0.37
Light yellow	13 (6.5)	22 (11)	
Yellow	58 (29)	60 (30)	
Brown	102 (51)	109 (54.5)	
Green	4 (2)	9 (4.5)	
Hypocotyl coloration			1.00
Green	131 (65.5)	132 (66)	
Purple	68 (34)	68 (34)	
Stem color			0.01
Light green	106 (53)	106 (53)	
Green	68 (34)	68 (34)	
Dark green	8 (4)	26 (13)	
Number of branches			1.00
Low	58 (29)	58 (29)	
Medium	79 (39.5)	80 (40)	
High	62 (31)	62 (31)	
Leaflet size			0.66
Small	81 (40.5)	102 (51)	
Medium	58 (29)	78 (39)	
Large	11 (5.5)	20 (10)	
Leaflet shape			0.98
Lanceolate	1 (0.5)	1 (0.5)	
Lanceolate to oblong	37 (18.5)	53 (26.5)	
Rhomboid	24 (12)	36 (18)	
Oval	46 (23)	62 (31)	
Elliptic	39 (19.5)	48 (24)	
Leaf color			0.52
Green	46 (23)	69 (34.5)	
Dark green	104 (52)	131 (65.5)	
Plant type			**0.04**
Determinate	142 (71)	177 (88.5)	
Semi-determinate	7 (3.5)	23 (11.5)	
Pubescence density			1.00
Absent	4 (2)	5 (2.5)	
Rare	20 (10)	20 (10)	
Sparse	33 (16.5)	33 (16.5)	
Medium	86 (43)	86 (43)	
Dense	56 (28)	56 (28)	
Pubescence color			1.00
Grayish white	102 (51)	104 (52)	
Pale brown	63 (31.5)	64 (32)	
Brown	32 (16)	32 (16)	
Corolla color			1.00
White	138 (69)	139 (69.5)	
Purple throat	34 (17)	34 (17)	
Purple	27 (13.5)	27 (13.5)	
Pod set capacity			0.95
Low	13 (6.5)	17 (8.5)	
Medium	54 (27)	76 (38)	
High	82 (41)	107 (53.5)	

N, number of germplasms;%, percentage. [a]Multiple imputation was used to estimate missing phenotypes. [b]Chi-squared test was conducted to test difference among observed and imputed phenotypes. Bold value (p-value < 0.05) represents significant difference between two sets.

in the seedbank. The same results are also shown in this study. Therefore, without special attention and care, the vegetable germplasm may not be focused. The market determines the importance of vegetable soybean and breeder's attention.

In this study, a total of 29 traits (15 quantitative and 14 qualitative traits) were selected from 47 phenotypic traits, based on at most 30% threshold of missing rate. Among them, 5 traits had no missing rates, 10 traits had low missing rates (0.5–11.5%), and the remaining traits had moderate missing rates (25.5–26.5%) (**Supplementary Table 3**). Generally, missing phenotypes are very often seen in many germplasms. Fortunately, our phenotypes did not seriously suffer from missing data. However, missing values in phenotypes, even at a low or moderate level, still leave space for uncertainty in search for the behavior of traits in vegetable soybean germplasms.

Phenotype completeness is the key to access in developing a core set of germplasm for efficient exploration and effective utilization in breeding programs. In fact, the impact of missing data on phenotypic traits research can be serious, and the utilization of phenotypes is often limited by incomplete phenotypes of interest (Haupt and Schmid, 2020). In particular, trait-based analysis like CC really relies on the completeness of phenotypes. Hence, a model-based imputation algorithm, such as multiple imputation by chained equations, can fill the gap in the trait-based germplasm database. As discussed by Taugourdeau et al. (2014), multiple imputation generally has accurate estimates on missing phenotypes and stable convergence statistic $\hat{R}s$, which were smaller than 1.1 (**Supplementary Table 4**), suggesting that multiple imputation can produce stable and reliable estimates for missing phenotypes. It is worth noticing that accessions for diverse landraces or genotypes from different countries or continents may demonstrate diverse phenotypes in traits, indicating that some of these traits may have unbalanced distributions. This means that the distributions

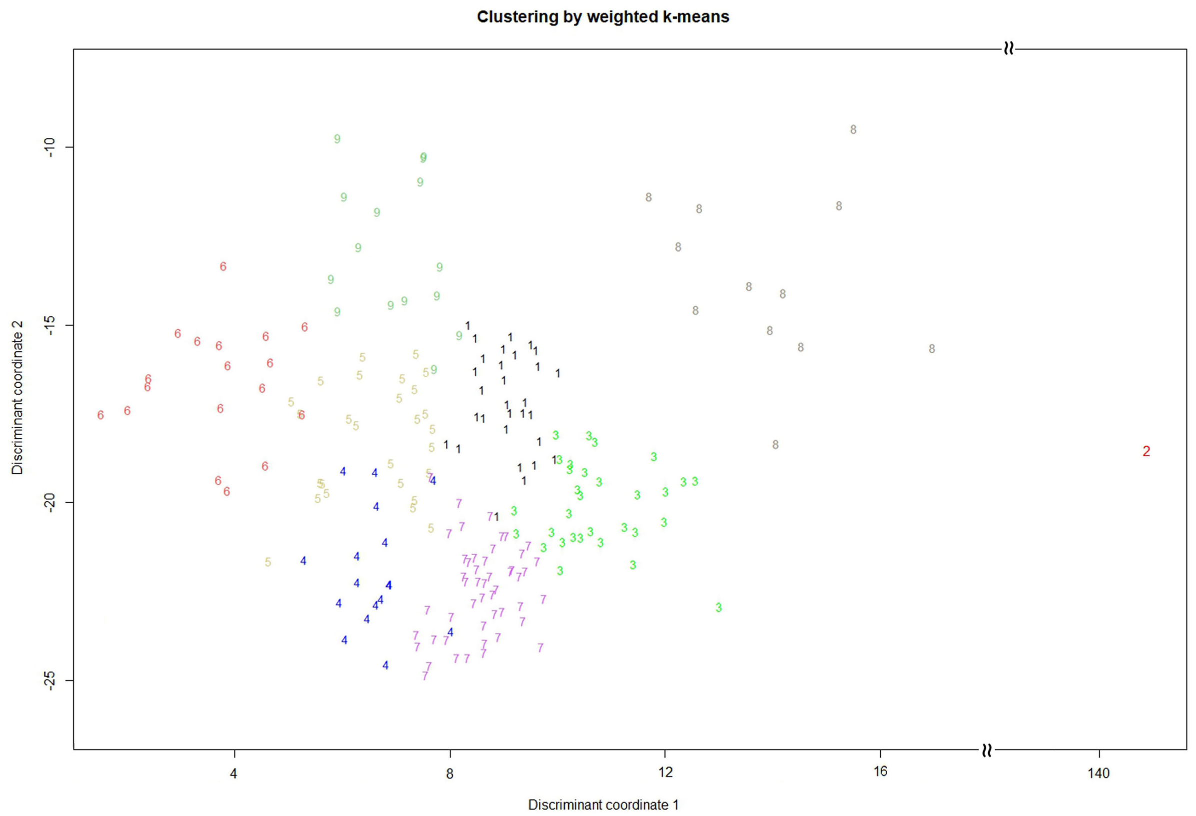

FIGURE 1 | Cluster analysis by weighted *k*-means for mixed quantitative–qualitative phenotypic traits. The vegetable soybean germplasms (200 accessions) are separately grouped into nine clusters.

TABLE 3 | Selected core collection of vegetable soybean.

Germplasm ID	Accession name[a]	Origin	Germplasm ID	Accession name[a]	Origin
KG0164	Nuli-6-G2657	Unknown	KG0127	207	China
KG0038	G10502	Japan	KG0180	Hsueh Chih Hsia-28	Japan
KG0031	G10493	Japan	KG0169	KS1822	Taiwan
KG0009	ESB-67-9 [KG0009]	Taiwan	KG0175_1	Kahori	Japan
KG0153	KVS39	Taiwan	KG0163	D-62-7815	United States
KG0137	G7332	United States	KG0181	Hsueh Chih Hsia-34	Japan
KG0011	ESB-67-14	Taiwan	KG0179	Hsueh Tou	Unknown
KG0159	Kaorihime	Japan	KG0177	GC 84136-P-4-1-8	Taiwan
KG0101	Ryokukou [KG0101]	Japan	KG0199	Goku Wase Osuro [KG0199]	Japan
KG0125	Chen Hsiang	Taiwan	KG0172	Lu Kuang-75	Japan
KG0185	AGS186	Taiwan	KG0171	Lu Kuang-74	Japan
KG0192	Sakata Kairyo Mikowashima	Japan	KG0113	Taiuasu Dare	Japan
KG0106	Kamui	Japan	KG0132	Mainland China	Hong Kong
KG0196	Tzuraroko Daizu	Japan	KG0010	ESB-67-10	Taiwan
KG0050	Fubaye	Unknown	KG0162	G10137	Philippines

[a]*The 30 accessions were selected, using modified Roger's distance, as the core collection.*

of the phenotypes for some traits are unbalanced (i.e., skewed distributions). For instance, for some traits such as 100 seed weight (g), 100 immature seed weight (g), and shelling rate (%), most values were high, but with few extreme low values; similarly, for plant height (cm), leaflet length (cm), and immature seed width (mm), most values were low, but with few extreme high values. Hence, multiple imputation in chained equations is an appropriate method to produce accurate and robust results for the unbalanced traits (Taugourdeau et al., 2014). Multiple imputation has addressed some imputation problems,

TABLE 4 | Difference test of quantitative traits between the core collection and the entire collection in vegetable soybean.

Phenotypic trait	Entire collection (EC)							Core collection (CC)[a]							Difference test[b]	
	N	Min	Max	Range	Mean	*SD*	CV (%)	*N*	Min	Max	Range	Mean	*SD*	CV (%)	p.homo	p.diff
Seed length (mm)	200	7.1	11.2	4.1	9.0	0.8	8.7	30	7.2	10.8	3.6	9.2	1.2	12.8	<0.001	0.40
Seed width (mm)	200	6.2	9.7	3.5	8.3	0.6	7.1	30	6.3	9.7	3.4	8.3	0.8	9.3	0.047	0.88
Seed thickness (mm)	200	5.1	8.8	3.7	7.1	0.7	9.5	30	5.1	8.8	3.7	6.9	0.8	11.4	0.25	0.22
100 seed weight (g)	200	4.2	51.2	47.0	33.4	7.5	22.5	30	4.2	51.2	47.0	32.4	12.6	38.9	<0.001	0.68
Plant height (cm)	198	17.3	70.7	53.4	37.1	10.7	28.9	30	17.3	70.7	53.4	45.2	14.4	31.9	0.007	**0.01**
Leaflet length (cm)	200	7.0	116.0	109.0	10.6	7.6	71.8	30	7.4	116.0	108.6	14.0	19.3	138.3	0.11	0.08
Leaflet width (cm)	200	1.4	10.4	9.0	7.2	1.2	16.3	30	1.4	8.8	7.4	7.3	1.5	20.9	0.52	0.69
Pod length (cm)	149	1.3	5.8	4.5	4.6	0.5	11.6	30	1.3	5.8	4.5	4.5	0.8	18.7	0.01	0.55
Pod width (cm)	149	0.9	4.3	3.4	1.2	0.3	23.0	30	0.9	4.3	3.4	1.3	0.6	44.3	0.10	0.19
Stem length to first pod (cm)	150	4.2	22.3	18.1	11.9	4.4	37.2	30	4.3	22.3	18.0	13.6	5.0	36.9	0.24	0.052
Shelling rate (%)	150	4.8	75.0	70.2	56.8	7.0	12.3	30	4.8	75.0	70.2	54.2	11.5	21.1	0.10	0.10
Immature seed length (mm)	149	6.0	18.1	12.1	15.5	1.6	10.5	29	6.0	18.0	12.0	15.4	2.4	15.7	0.20	0.75
Immature seed width (mm)	150	1.1	40.9	39.8	11.0	3.1	28.4	30	1.1	40.9	39.8	12.2	6.2	51.2	0.04	0.34
Immature seed thickness (mm)	150	5.8	9.7	3.9	8.1	0.8	10.2	30	5.8	9.7	3.9	8.0	0.9	11.3	0.91	0.83
100 immature seed weight (g)	149	6.2	100.0	93.8	69.1	14.9	21.5	30	6.2	100.0	93.8	64.7	24.7	38.1	< 0.001	0.36

N, number of germplasms; SD, standard deviation; CV, coefficient of variation; p.homo, p-value of homogeneity test for variance (Levene's test); p.diff, p-value of difference test (Student's t-test or Welch's t-test).
[a]Core collection was identified using modified Roger's distance for mixed quantitative–qualitative phenotypic traits. [b]Student's t-test (if assumption of homogeneity of variance is met) and Welch's t-test (if assumption of homogeneity of variance is not met) were used to conduct mean difference among two collections. Bold value (p-value < 0.05) represents significant difference between two sets.

TABLE 5 | Difference test of qualitative traits between the core collection and the entire collection in vegetable soybean.

Phenotypic trait	Entire collection (EC)	Core collection (CC)[a]	Difference test (*p*-value)[b]
	***N* (%)**	***N* (%)**	
Seed shape			0.91
Round	107 (53.5)	15 (50.0)	
Oblate	21 (10.5)	3 (10.0)	
Oval	58 (29.0)	9 (30.0)	
Flat	13 (6.5)	3 (10.0)	
Seed coat color			0.97
Yellowish white	26 (13.0)	4 (14.3)	
Yellow	69 (34.5)	9 (32.1)	
Green	92 (46.0)	15 (53.6)	
Pale brown	1 (0.5)	0 (0)	
Reddish brown	1 (0.5)	0 (0)	
Hilum color			0.28
Light yellow	13 (6.5)	2 (7.7)	
Yellow	58 (29.0)	11 (42.3)	
Brown	102 (51.0)	11 (42.3)	
Green	4 (2.0)	2 (7.7)	
Hypocotyl coloration			0.68
Green	131 (65.5)	18 (60.0)	
Purple	68 (34.0)	12 (40.0)	
Stem color			0.49
Light green	106 (53.0)	19 (63.3)	
Green	68 (34.0)	11 (36.7)	
Dark green	8 (4.0)	0 (0)	
Number of branches			0.77
Low	58 (29.0)	9 (30.0)	
Medium	79 (39.5)	10 (33.3)	
High	62 (31.0)	11 (36.7)	
Leaflet size			0.64
Small	81 (40.5)	10 (43.5)	
Medium	58 (29.0)	11 (47.8)	
Large	11 (5.5)	2 (8.7)	
Leaflet shape			0.17
Lanceolate	1 (0.5)	1 (4.2)	
Lanceolate to oblong	37 (18.5)	7 (29.2)	
Rhomboid	24 (12.0)	1 (4.2)	
Oval	46 (23.0)	11 (45.8)	
Elliptic	39 (19.5)	4 (16.6)	
Leaf color			0.75
Green	46 (23.0)	6 (25.0)	
Dark green	104 (52.0)	18 (75.0)	
Plant type			0.80
Determinate	142 (71.0)	22 (91.7)	
Semi-determinate	7 (3.5)	2 (8.3)	
Pubescence density			0.10
Absent	4 (2.0)	2 (6.7)	
Rare	20 (10.0)	6 (20.0)	
Sparse	33 (16.5)	4 (13.3)	
Medium	86 (43.0)	7 (23.3)	
Dense	56 (28.0)	11 (36.7)	
Pubescence color			0.46
Grayish white	102 (51.0)	11 (39.3)	
Pale brown	63 (31.5)	11 (39.3)	
Brown	32 (16.0)	6 (21.4)	
Corolla color			0.65
White	138 (69.0)	20 (66.7)	
Purple throat	34 (17.0)	7 (23.3)	
Purple	27 (13.5)	3 (10.0)	
Pod set capacity			0.68
Low	13 (6.5)	1 (4.2)	
Medium	54 (27.0)	8 (33.3)	
High	82 (41.0)	15 (62.5)	

N, number of germplasms; %, percentage. [a]Core collection was identified using modified Roger's distance for mixed quantitative–qualitative phenotypic traits. [b]Chi-squared test was conducted to test for difference among two collections.

including structural (multi)collinearity, estimation convergence, semi-continuous data analysis, and data separation, which were neglected by other approaches (Su et al., 2011). Therefore, the multiple imputation method is one of the best methods for traits imputation of missing data, with a maximum tolerance of 30% missing rate. Fortunately, our vegetable germplasm has a low level of missing data, and our imputed missing data exhibited converged estimates as well as a lower error on the estimation of missing phenotypic trait values.

The selection of sampling strategies does affect the results of the selection of the CC, as well as its diversity. To have a better illustration, we compared the proposed modified Roger's distance algorithm with other sampling schemes, including stratified proportional sampling (SPS), simple random sampling (SRS), and advanced M strategy (using PowerCore software). As the difference test, there were only two traits, seed width by advanced M strategy and number of branches by SRS, showing significant mean difference between the EC and the CC (**Table 5** and **Supplementary Table 5**). Most of the tests were found to be non-significant between the EC and the CC, indicating that four sampling methods had similar abilities to collect the CC. **Supplementary Table 6** demonstrates the phenotypic diversity using three other sampling strategies. Among the four strategies, the CC from modified Roger's distance algorithm showed the highest phenotypic diversity.

The CC collected from distinct sampling strategies was evaluated. We evaluated the percentage of the trait differences between the CC and the EC, using PowerCore, SPS, and SRS methods, compared to our developed modified Roger's distance (please refer to **Table 7** and **Supplementary Table 7**). All methods showed a very small mean difference and a very low percentage of significant difference (MD% < 20%) between the CC and the EC, demonstrating equal central tendency properties. Comparing coincidence rates, the modified Roger's

TABLE 6 | Diversity comparison between the core collection and the entire collection in vegetable soybean.

Phenotypic trait	Clusters and diversity					
	Entire collection (EC)			Core collection (CC)[a]		
	k_{EC}[b]	H'	Nei's	k_{CC}	H'	Nei's
Quantitative traits						
Seed length (mm)	6	0.91	0.78	6	0.97	0.82
Seed width (mm)	9	0.94	0.86	9	0.95	0.86
Seed thickness (mm)	5	0.95	0.77	5	0.98	0.79
100 seed weight (g)	12	0.93	0.89	11	0.95	0.89
Plant height (cm)	11	0.95	0.89	9	0.89	0.83
Leaflet length (cm)	3	0.65	0.50	3	0.68	0.46
Leaflet width (cm)	6	0.91	0.79	6	0.85	0.74
Pod length (cm)	10	0.86	0.84	7	0.89	0.79
Pod width (cm)	5	0.64	0.54	4	0.86	0.66
Stem length to first pod (cm)	5	0.96	0.77	4	0.99	0.74
Shelling rate (%)	5	0.88	0.75	4	0.90	0.69
Immature seed length (mm)	6	0.89	0.78	6	0.91	0.78
Immature seed width (mm)	9	0.86	0.83	8	0.97	0.86
Immature seed thickness (mm)	6	0.97	0.82	5	0.95	0.77
100 immature seed weight (g)	6	0.84	0.75	6	0.95	0.81
Qualitative traits						
Seed shape	4	0.80	0.61	4	0.84	0.64
Seed coat color	5	0.65	0.61	3	0.89	0.59
Hilum color	4	0.69	0.55	4	0.81	0.63
Hypocotyl coloration	2	0.93	0.45	2	0.97	0.48
Stem color	3	0.75	0.52	2	0.95	0.46
Number of branches	3	0.99	0.66	3	1.00	0.66
Leaflet size	3	0.81	0.55	3	0.84	0.57
Leaflet shape	5	0.87	0.74	5	0.80	0.67
Leaf color	2	0.89	0.43	2	0.81	0.38
Plant type	2	0.27	0.09	2	0.41	0.15
Pubescence density	5	0.82	0.70	5	0.92	0.75
Pubescence color	3	0.91	0.60	3	0.97	0.65
Corolla color	3	0.75	0.47	3	0.76	0.49
Pod set capacity	3	0.83	0.56	3	0.72	0.50

k_{EC}, number of clusters in the entire collection; k_{CC}, number of clusters in the core collection; H′, Shannon–Weaver diversity index; Nei's, Nei's diversity index. [a]Core collection was identified using modified Roger's distance for mixed quantitative–qualitative phenotypic traits. [b]Clustering analyses were conducted using weighted k-means clustering algorithm.

distance had the highest CR% = 97.7% (>80%), followed by Advanced M strategy by PowerCore (CR% = 96.56%), but SRS (CR% = 69.18) and SPS (CR% = 58.38%) had poor genetic variability for traits. Our modified Roger's distance (VD% = 52.90%, VR% = 149.80%) and PowerCore (VD% = 52.16%, VR% = 151.68%) showed equally slightly larger variance difference percentage and larger variable rate, suggesting to provide a good representation of the genetic diversity of the EC. However, SRS had a small VD% (41.94%) and an intermediate value of VR% (117.41%); SPS demonstrated an extremely large VD% (356.85%) and extremely low VR% (89.67%). This indicated that the properties of the CC constructed by both SPS and SRS were not representative of the EC. Furthermore, we found that the modified Roger's distance demonstrated the highest coverage% in quantitative (90.38%), qualitative (94.76%), and combined traits (92.50%), compared to SPS and SRS. Notice that PowerCore had 100% coverage percentage in qualitative traits, but not available in quantitative traits and the combined traits. As discussed above, different sampling strategies can affect the properties of the CC. Therefore, the CC constructed by the modified Roger's distance can capture accessions with valuable and/or special characteristics and larger variability from the EC.

In this study, a total of 30 accessions were selected as our CC. Investigating origins and major features of the CC indicate that these accessions are excellent germplasms with abundant genetic diversity, including several germplasm test lines from the vegetable soybean breeding programs in Taiwan, plenty of germplasms introduced from Japan, some Japanese varieties that have long been planted in Taiwan, and other germplasms that come from United States, China, Hong Kong, Philippines, and other unknown countries (**Tables 3–7**). As for the important traits of vegetable soybean (e.g., seed size and weight, cold tolerance, isoflavone content, and yield), most accessions in CC cover all the range of these traits to be the same as or approximately similar to that of the EC, suggesting that our CC can be representative of the diversity of the EC. For example, immature seed size and weight were found to be associated with yield, quality, and appearance for vegetable soybean (Krisnawati and Adie, 2015). The two most important vegetable soybean cultivars in Taiwan are Kaohsiung 9 and Kaohsiung 12, which are characterized by 100 immature seed weight (81 g for #9 and 75 g for #12) and shattering rate (61% for #9 and 48% for #12). Our CC revealed that 100 immature seed weight and shattering rate ranged between 37 and 100 g and between 41

TABLE 7 | Evaluation in percentage of the trait differences between the core collection and the entire collection in vegetable soybean.

Modified Roger's distance	MD%	VD%	CR%	VR%	Coverage %[a]		
					Quantitative traits	Qualitative traits	Combined traits
The property of the CC	6.35	52.90	97.70	149.80	90.38	94.76	92.50
Percentage of significant difference[b]	3.45						

MD%, mean difference percentage; VD%, variance difference percentage; CR%, coincidence rate; VR%, variable rate. [a]Coverages were computed according to quantitative traits, qualitative traits, and combined traits. [b]The core collection is considered to be representative of the entire collection if (1) MD% is small and percentage of significant mean difference (α = 0.05) between the core collection and the entire collection in all 29 phenotypic traits is no more than 20%, and (2) the CR% is greater than 80%.

and 75%, respectively. In addition, our CC includes two small seed varieties (KG0050: 6.8 g, KG0127: 6.2 g). These suggested that our CC provides potential for breeding programs. On the other hand, pod set capacity affects the production of vegetable soybean at low temperature (Kurosaki et al., 2003). Among the CC, all three levels of pod set capacity were covered and most accessions have high pod set capacity (**Table 5**), which reveals that our accessions in CC may have potential for cold tolerance. It indicated that the CC contained market-oriented phenotypes, and it can be used as a priority resource for vegetable soybean breeding applications. In the future, if there are any requirements for the introductions of the novel disease/pest resistances or stress tolerances, this CC can be applied first for the breeding programs. Taken together, the CC determined by the new algorithm not only retained most diversity from the EC but also will be useful for the breeders and plant scientists for breeding program.

From diversity investigation and assessment of the CC (please refer to **Tables 4–7**), the CC retained valuable characteristics and large variability for phenotypic traits in the EC, suggesting representativeness of the whole germplasms. We take the edamame traits of the immature seeds as an example. There are four traits (immature seed length, immature seed width, immature seed thickness, and 100 immature seed weight) related to immature seeds. From **Table 4**, we can see that the ranges (includes minimal and maximum) of the four traits in the CC fully covered those in the EC, with quite similar means. From the difference test, we can also obtain non-significant difference among the CC and the EC, suggesting that the CC has high coverage to the EC.

Among the CC, there is one germplasm called Mikawashima. It is a local variety in Japan, which was bred through a variety of improvements using a landrace by the Sakata Seed Company in Japan (Mimura et al., 2007). It might contain several common and popular traits of vegetable soybean in Japan but had not been applied by Taiwanese breeders. Therefore, Mikawashima became an outlier in the clustering analysis (**Figure 1**).

There are some limitations of this study. First, many of the quantitative phenotypic traits are influenced by environmental effects. To minimize environmental effects, we only focused on vegetable soybean accessions planted in autumn to remove possible effects from environmental factors. In addition, our collected meteorology data showed that environmental factors did not significantly (p-values > 0.05) affect our trait values measured during 1995–1998 autumn seasons (**Supplementary Table 2**). Second, the CC was established by using trait-based data. However, some studies suggested that the choice of data type (phenotypes or genotypes) is not the main cause to alter the diversity (Grenier et al., 2000; Yun et al., 2015). The quality of data control is the first thing taken into consideration regardless of the data used for developing the CC. In the current study, we have conducted data quality control, including focusing on autumn crops and redundant exclusion, to remove possible noises and biases and produce reliable results.

The CC accounted for 15% (30/200) of the entire collection. Moreover, the high diversity of the CC indicated that the germplasms were simplified without losing information. The result showed that the modified Roger's distance algorithm was a good approach to construct the CC. This study provides a detailed list of the CC, which can be a basis for future research of vegetable soybeans and breeding practices.

CONCLUSION

To the best of our knowledge, this study reported results of the first CC for vegetable soybean (or edamame) worldwide. This study consisted of the largest vegetable soybean germplasms worldwide at present, which provided valuable and/or special information on Taiwanese vegetable soybean germplasms. We proposed a modified Roger's distance algorithm, which is a mixed-type sampling strategy, to construct a CC composed of a total of 30 accessions based on morphological traits. It was evident that the properties of the CC constructed by the modified Roger's distance algorithm were reliable, stable, and feasible. With the representativeness of the whole germplasms, the high-quality CC possessed small mean difference, high coincidence rate, variable rate, and high coverage. The CC preserved the high phenotypic diversity in germplasms of vegetable soybean. The removal of redundancy enables us to make full use of the vegetable soybean germplasms for overall phenotype clustering. With the support of the material, the CC can be helpful for further breeding of commercial varieties.

DATA AVAILABILITY STATEMENT

The original contributions generated for this study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

C-FK: study conception and design and acquisition and analysis of data. C-FK, S-SH, C-SW, D-GL, and SC: interpretation of data. C-FK and S-SH: drafting the manuscript. C-FK and Z-YL: manuscript revision. All authors read and approved the final manuscript.

FUNDING

This work was supported by grants MOST 107-2313-B-005-046 and MOST 108-2313-B-005-017 from the Taiwan Ministry of Science and Technology and grant 109AS-1.1.5-ST-aE from the Council of Agriculture, Executive Yuan, Taiwan. This work was financially supported (in part) by the Advanced Plant Biotechnology Center from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

ACKNOWLEDGMENTS

We thank Yi-An Chang and Ya-Syuan Lai for English editing and data management.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.612106/full#supplementary-material

Supplementary Table 1 | Forty-seven phenotypic traits of vegetable soybean in Taiwan.

Supplementary Table 2 | Meteorology data during 2006–2009 in Kaohsiung, Taiwan.

Supplementary Table 3 | Summary information of missing rate in phenotypic traits.

Supplementary Table 4 | Summary results of imputed phenotypic traits by chained equation in multiple imputation.

Supplementary Table 5 | Difference test of phenotypic traits between the core collection (using PowerCore, SPS, and SRS methods) and the entire collection in vegetable soybean.

Supplementary Table 6 | Diversity between the core collection (using PowerCore, SPS and SRS methods) and the entire collection in vegetable soybean.

Supplementary Table 7 | Evaluation of the core collection (using PowerCore, SPS, and SRS methods) in percentage of the trait differences with the entire collection in vegetable soybean.

Supplementary Material 1 | The raw data of the core collection in vegetable soybean.

REFERENCES

Adlercreutz, H. (2002). Phyto-oestrogens and cancer. *Lancet Oncol.* 3, 364–373. doi: 10.1016/S1470-2045(02)00777-5

Aggarwal, C. C., Hinneburg, A., and Keim, D. A. (2001). "On the surprising behavior of distance metrics in high dimensional space," in *Database Theory — ICDT 2001. ICDT 2001. Lecture Notes in Computer Science*, Vol. 1973, eds J. Van den Bussche and V. Vianu (Berlin: Springer).

Bravo, J. A., Fehr, W., and de Cianzio, S. R. (1980). Use of pod width for indirect selection of seed weight in soybeans. *Crop Sci.* 20, 507–510. doi: 10.2135/cropsci1980.0011183X002000040022x

Carneiro, R. C. V., Duncan, S. E., O'Keefe, S. F., Yin, Y., Neill, C. L., and Zhang, B. (2020). Sensory and consumer studies in plant breeding: a guidance for edamame development in the U.S. *Front. Sustain. Food Syst.* 4:124. doi: 10.3389/fsufs.2020.00124

Choung, M.-G., Baek, I.-Y., Kang, S.-T., Han, W.-Y., Shin, D.-C., Moon, H.-P., et al. (2001). Isolation and determination of anthocyanins in seed coats of black soybean (*Glycine max* (L.) Merr.). *J. Agric. Food Chem.* 49, 5848–5851. doi: 10.1021/jf010550w

Delacre, M., Lakens, D., and Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of student's t-test. *Int. Rev. Soc. Psychol.* 30, 92–101. doi: 10.5334/irsp.82

Dwevedi, A., and Kayastha, A. M. (2011). "Soybean: a multifaceted legume with enormous economic capabilities," in *Soybean: Biochemistry, Chemistry and Physiology*, ed. N. Tzi-Bun (Rijeka: IntechOpen).

Fehr, W. R. (1971). Stage of development descriptions for soybeans, *Glycine max* (L.). *Merrill. Crop Sci.* 11, 929–931. doi: 10.2135/cropsci1971.0011183X001100060051x

Foss, A., Markatou, M., Ray, B., and Heching, A. (2016). A semiparametric method for clustering mixed data. *Mach. Learn.* 105, 419–458. doi: 10.1007/s10994-016-5575-7

Foss, A. H., and Markatou, M. (2018). kamila: clustering mixed-type data in R and Hadoop. *J. Stat. Softw.* 83, 1–44.

Frank, S. J., and Fehr, W. R. (1981). Associations among pod dimensions and seed weight in soybeans. *Crop Sci.* 21, 547–550. doi: 10.2135/cropsci1981.0011183X002100040018x

Frankel, O., and Arber, W. (1984). "Genetic perspectives of germplasm conservation," in *Genetic Manipulation: Impact on Man and Society*, eds W. Arber, K. Illemensee, W. J. Peacock, and P. Starlinger (Cambridge, MA: Cambridge University Press).

Frankel, O. H. (1984). *Genetic Perspectives of Germplasm Conservation. Genetic Manipulation: Impact on Man and Society*. Cambridge, MA: Cambridge University Press.

Githiri, S. M., Yang, D., Khan, N. A., Xu, D., Komatsuda, T., and Takahashi, R. (2007). QTL analysis of low temperature–induced browning in soybean seed coats. *J. Hered.* 98, 360–366. doi: 10.1093/jhered/esm042

Grenier, C., Deu, M., Kresovich, S., Bramel-Cox, P. J., and Hamon, P. (2000). Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs non-random sampling procedures : B. Using molecular markers. *Theor. Appl. Genet.* 101, 197–202. doi: 10.1007/s001220051469

Guo, Y., Li, Y., Hong, H., and Qiu, L.-J. (2014). Establishment of the integrated applied core collection and its comparison with mini core collection in soybean (*Glycine max*). *Crop J.* 2, 38–45. doi: 10.1016/j.cj.2013.11.001

Haupt, M., and Schmid, K. (2020). Combining focused identification of germplasm and core collection strategies to identify genebank accessions for central European soybean breeding. *Plant Cell Environ.* 43, 1421–1436. doi: 10.1111/pce.13761

Hofstra, G. (1972). Response of soybeans to temperature under high light intensities. *Can. J. Plant Sci.* 52, 535–543. doi: 10.4141/cjps72-084

Hu, J., Zhu, J., and Xu, H. M. (2000). Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theor. Appl. Genet.* 101, 264–268. doi: 10.1007/s001220051478

Hu, Q.-G., Zhang, M., Mujumdar, A. S., Xiao, G.-N., and Sun, J.-C. (2007). Performance evaluation of vacuum microwave drying of edamame in deep-bed drying. *Dry. Technol.* 25, 731–736. doi: 10.1080/07373930701291199

Huang, J. Z., Ng, M. K., Rong, H., and Li, Z. (2005). Automated variable weighting in k-means type clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 657–668. doi: 10.1109/tpami.2005.95

Ikeda, T., Ohnishi, S., Senda, M., Miyoshi, T., Ishimoto, M., Kitamura, K., et al. (2009). A novel major quantitative trait locus controlling seed development at low temperature in soybean (*Glycine max*). *Theor. Appl. Genet.* 118, 1477–1488. doi: 10.1007/s00122-009-0996-3

Jiang, G.-L., Rutto, L. K., Ren, S., Bowen, R. A., Berry, H., and Epps, K. (2018). Genetic analysis of edamame seed composition and trait relationships in soybean lines. *Euphytica* 214:158. doi: 10.1007/s10681-018-2237-9

Johnson, D., Wang, S., and Suzuki, A. (1999). "Edamame: a vegetable soybean for Colorado," in *Perspectives on New Crops and New Uses*, ed. J. Janick (Alexandria, VA: ASHS Press), 385–387.

Joseph, L. G., Yulia, R. G., Hui, W. L. W., Vyacheslav, L., Weiwen, M., and Kimihiro, N. (2020). *lawstat: Tools for Biostatistics, Public Policy, and Law*. Available online at: https://CRAN.R-project.org/package=lawstat (accessed Augeust 15, 2020).

Kim, E.-H., Lee, O.-K., Kim, J. K., Kim, S.-L., Lee, J., Kim, S.-H., et al. (2014). Isoflavones and anthocyanins analysis in soybean (*Glycine max* (L.) Merill) from three different planting locations in Korea. *Field Crops Res.* 156, 76–83. doi: 10.1016/j.fcr.2013.10.020

Kim, K.-W., Chung, H.-K., Cho, G.-T., Ma, K.-H., Chandrabalan, D., Gwag, J.-G., et al. (2007). PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics* 23, 2155–2162. doi: 10.1093/bioinformatics/btm313

Koehler, K. J., and Wilson, J. R. (1986). Chi–square tests for comparing vectors of proportions for several cluster samples. *Commun. Stat. Theory Methods* 15, 2977–2990. doi: 10.1080/03610928608829290

Konovsky, J., Lumpkin, T. A., and McClary, D. (1994). "Edamame: the vegetable soybean," in *Understanding the Japanese Food and Agrimarket: A Multifaceted Opportunity*, ed. A. D. O'Rourke (Binghamton, NY: Haworth Press), 173–181. doi: 10.1201/9781003075172-15

Krisnawati, A., and Adie, M. M. (2015). Selection of soybean genotypes by seed size and its prospects for industrial raw material in Indonesia. *Proc. Food Sci.* 3, 355–363. doi: 10.1016/j.profoo.2015.01.039

Kropko, J., Goodrich, B., Gelman, A., and Hill, J. (2017). Multiple imputation for continuous and categorical data: comparing joint multivariate normal and conditional approaches. *Polit. Anal.* 22, 497–519. doi: 10.1093/pan/mpu007

Kurosaki, H., and Yumoto, S. (2003). Effects of low temperature and shading during flowering on the yield components in soybeans. *Plant Prod. Sci.* 6, 17–23. doi: 10.1626/pps.6.17

Kurosaki, H., Yumoto, S., and Matsukawa, I. (2003). Pod setting pattern during and after low temperature and the mechanism of cold-weather tolerance at the flowering stage in soybeans. *Plant Prod. Sci.* 6, 247–254. doi: 10.1626/pps.6.247

Lee, K. J., and Simpson, J. A. (2014). Introduction to multiple imputation for dealing with missing data. *Respirology* 19, 162–167. doi: 10.1111/resp.12226

Levene, H. (1960). "Robust Tests for Equality of Variances," in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, ed. I. Olkin (Palo Alto, CA: Stanford University Press), 278–292.

Lin, S. F. (1997). Soybean germplasm in Taiwan. *Plant Genet. Resour. Newsl.* 2, 4–6.

Liu, K. (ed.) (1997). "Chemistry and nutritional value of soybean components," in *Soybeans: Chemistry, Technology, and Utilization*, (Boston, MA: Springer), 25–113. doi: 10.1007/978-1-4615-1763-4_2

Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Trans. Inf. Theory* 28, 129–136. doi: 10.1109/tit.1982.1056489

Magee, P. J., McGlynn, H., and Rowland, I. R. (2004). Differential effects of isoflavones and lignans on invasiveness of MDA-MB-231 breast cancer cells in vitro. *Cancer Lett* 208, 35–41. doi: 10.1016/j.canlet.2003.11.012

Mebrahtu, T. (2008). Analysis of nutritional contents in vegetable soybeans. *J. Crop Improv.* 21, 157–170. doi: 10.1080/15427520701885675

Mense, S. M., Hei, T. K., Ganju, R. K., and Bhat, H. K. (2008). Phytoestrogens and breast cancer prevention: possible mechanisms of action. *Environ. Health Perspect.* 116, 426–433. doi: 10.1289/ehp.10538

Messina, M. (2016). Soy and health update: evaluation of the clinical and epidemiologic literature. *Nutrients* 8:754. doi: 10.3390/nu8120754

Miladinović, J., Đorđević, V., Balešević-Tubić, S., Petrović, K., Ćeran, M., Cvejić, J., et al. (2019). Increase of isoflavones in the aglycone form in soybeans by targeted crossings of cultivated breeding material. *Sci. Rep.* 9:10341. doi: 10.1038/s41598-019-46817-1

Mimura, M., Coyne, C. J., Bambuck, M. W., and Lumpkin, T. A. (2007). SSR diversity of vegetable soybean [*Glycine max* (L.) Merr.]. *Genet. Resour. Crop Evol.* 54, 497–508. doi: 10.1007/s10722-006-0006-4

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *PNAS* 70, 3321–3323. doi: 10.1073/pnas.70.12.3321

Penone, C., Davidson, A. D., Shoemaker, K. T., DiMarco, M., Rondinini, C., Brooks, T., et al. (2014). Imputation of missing data in life-history trait datasets: which approach performs the best? *MEE* 5, 961–970. doi: 10.1111/2041-210X.12232

Qiu, L.-J., Xing, L.-L., Guo, Y., Wang, J., Jackson, S. A., and Chang, R.-Z. (2013). A platform for soybean molecular breeding: the utilization of core collections for food security. *Plant Mol. Biol.* 83, 41–50. doi: 10.1007/s11103-013-0076-6

Reddy, M. R., Poshadri, A., Chauhan, S., Prashanth, Y., and Reddy, R. U. (2019). Evaluation of soybean lines for edamame (*Glycine max* (L.) merrill) as a potential vegetable for Telangana state of India. *Int. J. Curr. Microbiol. Appl. Sci.* 8, 552–560. doi: 10.20546/ijcmas.2019.803.067

Rogers, J. S. (1972). "Measures of genetic similarity and genetic distance," in *Studies in Genetics VII*, ed. M. R. Wheeler (Austin: University of Texas Publication), 7145–7153.

Royston, P. (2004). Multiple imputation of missing values. *Stata J.* 4, 227–241. doi: 10.1177/1536867x0400400301

Sajidha, S. A., Desikan, K., and Chodnekar, S. P. (2020). Initial seed selection for mixed data using modified k-means clustering algorithm. *Arab. J. Sci. Eng.* 45, 2685–2703. doi: 10.1007/s13369-019-04121-0

Saldivar, X., Wang, Y. J., Chen, P., and Mauromoustakos, A. (2010). Effects of blanching and storage conditions on soluble sugar contents in vegetable soybean. *LWT Food Sci. Technol.* 43, 1368–1372. doi: 10.1016/j.lwt.2010.04.017

Shanmugasundaram, S. (1991). "Vegetable soybean: research needs for production and quality improvement," in *Proceedings of the Workshop Held at Kenting*, Taiwan.

Shannon, C. E., and Weaver, W. (1962). *The Mathematical Theory of Communication*. Champaign, IL: The University of Illinois Press.

Singh, N., Wu, S., Raupp, W. J., Sehgal, S., Arora, S., Tiwari, V., et al. (2019). Efficient curation of genebanks using next generation sequencing reveals substantial duplication of germplasm accessions. *Sci. Rep.* 9:650. doi: 10.1038/s41598-018-37269-0

Su, Y.-S., Gelman, A. E., Hill, J., and Yajima, M. (2011). Multiple imputation with diagnostics (mi) in R: opening windows into the black box. *J. Stat. Softw.* 45, 1–31. doi: 10.18637/jss.v045.i02

Sunada, K., and Ito, T. (1982). Soybean grain quality as affected by low temperature treatments in plants (color of hilum, seed coat cracking). Report of the Hokkaido Branch, the Japanese Society of Breeding and Hokkaido Branch. *Crop Sci. Soc. Jpn.* 22:34. doi: 10.20751/hdanwakai.22.0_34

Taugourdeau, S., Villerd, J., Plantureux, S., Huguenin-Elie, O., and Amiaud, B. (2014). Filling the gap in functional trait databases: use of ecological hypotheses to replace missing data. *Ecol. Evol.* 4, 944–958. doi: 10.1002/ece3.989

Velásquez, M. R. T., and Bhathena, S. J. (2007). Role of dietary soy protein in obesity. *Int. J. Med. Sci.* 4, 72–82. doi: 10.7150/ijms.4.72

Wang, K.-C. (2018). East Asian food regimes: agrarian warriors, edamame beans and spatial topologies of food regimes in East Asia. *J. Peasant Stud.* 45, 739–756. doi: 10.1080/03066150.2017.1324427

Wang, K.-J., Li, X.-H., and Li, F.-S. (2008). Phenotypic diversity of the big seed type subcollection of wild soybean (*Glycine soja* Sieb. et Zucc.) in China. *Genet. Resour. Crop Evol.* 55, 1335–1346. doi: 10.1007/s10722-008-9332-z

Xu, Y., Li, H.-N., Li, G.-J., Wang, X., Cheng, L.-G., and Zhang, Y.-M. (2011). Mapping quantitative trait loci for seed size traits in soybean (*Glycine max* L. Merr.). *Theor. Appl. Genet.* 122, 581–594. doi: 10.1007/s00122-010-1471-x

Young, G., Mebrahtu, T., and Johnson, J. (2000). Acceptability of green soybeans as a vegetable entity. *Plant Foods Hum. Nutr.* 55, 323–333. doi: 10.1023/A:1008164925103

Yun, W. H., Ban, S. H., Kim, G. H., Kim, J. H., Kwon, S. I., and Choi, C. (2015). Assessment of apple core collections constructed using phenotypic and genotypic data. *Genet. Mol. Res.* 14, 6453–6464. doi: 10.4238/2015.June.11.21

Zafra-Stone, S., Yasmin, T., Bagchi, M., Chatterjee, A., Vinson, J. A., and Bagchi, D. (2007). Berry anthocyanins as novel antioxidants in human health and disease prevention. *Mol. Nutr. Food Res.* 51, 675–683. doi: 10.1002/mnfr.200700002

Zeipiòa, S., Alsiòa, I., and Lepse, L. (2017). Insight in edamame yield and quality parameters: a review. *Res. Rural. Dev.* 2, 40–45. doi: 10.22616/RRD.23.2017.047

Zhang, G. W., Xu, S. C., Mao, W. H., Hu, Q. Z., and Gong, Y. M. (2013). Determination of the genetic diversity of vegetable soybean [*Glycine max* (L.) Merr.] using EST-SSR markers. *J. Zhejiang Univ. Sci. B.* 14, 279–288. doi: 10.1631/jzus.B1200243

Zhang, J., Ge, Y., Han, F., Li, B., Yan, S., Sun, J., et al. (2014). Isoflavone content of soybean cultivars from maturity group 0 to VI grown in Northern and Southern China. *J. Am. Oil Chem. Soc.* 91, 1019–1028. doi: 10.1007/s11746-014-2440-3

Zhang, J., Wu, J., Liu, F., Tong, L., Chen, Z., Chen, J., et al. (2019). Neuroprotective effects of anthocyanins and its major component cyanidin-3-O-glucoside (C3G) in the central nervous system: an outlined review. *Eur. J. Pharmacol.* 858:172500. doi: 10.1016/j.ejphar.2019.172500

Ziaei, S., and Halaby, R. (2017). Dietary isoflavones and breast cancer risk. *Medicines* 4:18. doi: 10.3390/medicines4020018

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Sustainable Food Systems

ORIGINAL RESEARCH
published: 18 January 2021
doi: 10.3389/fsufs.2020.556580

Check for updates

Utilizing Consumer Perception of Edamame to Guide New Variety Development

Renata Carneiro[1], Susan Duncan[1*], Sean O'Keefe[1], Dajun Yu[1], Haibo Huang[1], Yun Yin[1], Clinton Neill[2†], Bo Zhang[3], Thomas Kuhar[4], Steve Rideout[5], Mark Reiter[6], Jeremy Ross[7], Pengyin Chen[8] and Anne Gillen[9]

[1] Department of Food Science and Technology, Virginia Tech, Blacksburg, VA, United States, [2] Department of Agricultural and Applied Economics, Virginia Tech, Blacksburg, VA, United States, [3] School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, United States, [4] Department of Entomology, Virginia Tech, Blacksburg, VA, United States, [5] Department of Plant Pathology, Physiology, and Weed Science, Virginia Tech, Painter, VA, United States, [6] Eastern Shore Agricultural Research and Extension Center, School of Plant and Environmental Sciences, Virginia Tech, Painter, VA, United States, [7] Cooperative Extension Service, Department of Crop, Soil, and Environmental Sciences, University of Arkansas, Little Rock, AR, United States, [8] Fisher Delta Research Center, University of Missouri, Portageville, MO, United States, [9] USDA, Agricultural Research Service, Crop Genetics Research Unit, Stoneville, MS, United States

OPEN ACCESS

Edited by:
Rakesh Bhardwaj,
National Bureau of Plant Genetic Resources (ICAR), India

Reviewed by:
Maria Ewa Rembiałkowska,
Warsaw University of Life Sciences, Poland
Maria Giovanni,
California State University, Chico, United States

****Correspondence:***
Susan Duncan
duncans@vt.edu

†Present address:
Clinton Neill,
Dyson School of Applied Economics and Management, Cornell University, Ithaca, NY, United States

Specialty section:
This article was submitted to Nutrition and Sustainable Diets, a section of the journal Frontiers in Sustainable Food Systems

Received: *28 April 2020*
Accepted: *15 December 2020*
Published: *18 January 2021*

Citation:
Carneiro R, Duncan S, O'Keefe S, Yu D, Huang H, Yin Y, Neill C, Zhang B, Kuhar T, Rideout S, Reiter M, Ross J, Chen P and Gillen A (2021) Utilizing Consumer Perception of Edamame to Guide New Variety Development. Front. Sustain. Food Syst. 4:556580. doi: 10.3389/fsufs.2020.556580

Consumption of edamame (vegetable soybeans) has increased significantly in the U.S. over the last 20 years. Although market demand has been increasing, most edamame is still imported from Asian countries. A team of multistate plant-breeding programs in the mid-Atlantic and Southeast U.S. has focused on developing new breeding lines that grow well in the U.S. and deliver what domestic growers, processors and consumers need and expect from their edamame. In our study, sensory evaluation was used to identify edamame genotypes and sensory attributes preferred by consumers to support breeding selection criteria. In the first year (reported as our "screening study"), 20 edamame genotypes were grown in three locations: Newport, AR, and Blacksburg and Painter, VA. In the second year (reported as our "validation study"), 10 edamame genotypes selected after our screening study were grown in Blacksburg and Painter, VA, Portageville, MO, and Stoneville, MS. In both years of research, untrained participants (adults; vegetable consumers not allergic to soy; $N \geq 50$) used a traditional 9-point acceptability (hedonic) scale (1 = "dislike extremely"; 9 = "like extremely") to evaluate overall-liking, aroma, appearance, taste, and texture, and a 5-point scale (1 = "not sweet," 5 = "extremely sweet") to evaluate sweetness intensity. Next, participants used a check-all-that-apply (CATA) list of selected sensory terms to describe the sensory characteristics of each edamame sample. Overall acceptability of edamame genotypes was significantly different among all genotypes ($p < 0.05$). Samples described as "bitter," "sour" (flavor) or "starchy" (texture) were associated with lower acceptability scores while "salty" and "sweet" (flavor) were correlated with higher acceptability. Sensory data from the screening study were used to select the best genotypes by use of a defined decision process based on the consumer data. The validation study tested the selection decisions and further supported the genotype choices. Sensory evaluation is a powerful tool to direct breeders to improve market acceptability and develop new edamame genotypes. Both screening and validation studies illustrate the significant role of consumer sensory data in support of genotypes targeted for domestic (U.S.) production.

Keywords: edamame, vegetable soybean, *Glycine max* (L.) Merr., sensory, consumer, plant breeding, food development

INTRODUCTION

Edamame, vegetable soybean [*Glycine max* (L.) Merr.], is a nutritious and high-value specialty crop that has been consumed in Asian countries for centuries and its consumption has increased in the U.S. in the past 20 years (Johnson et al., 1999; Wolfe et al., 2018). Edamame is rich in vitamins, dietary fibers, isoflavones, proteins, and essential amino acids (Masuda, 1991; Mebrahtu et al., 2004; Song et al., 2013; Ntatsi et al., 2018). Generally, edamame beans are consumed as a snack or added in salads, stews, and soups (Masuda, 1991; Konovsky et al., 1994; Mebrahtu and Devine, 2008; Kumar et al., 2011). Edamame has also gained popularity among growers and processors in the mid-Atlantic and Southeast U.S as it has been promoted as a potential alternative crop to replace some traditional crops whose production and value have been decreasing, such as tobacco (Ernst and Woods, 2001; Carson et al., 2011; Xu et al., 2012; Ogles et al., 2016). In order to increase the domestic production of edamame, improved seeds that grow well in the U.S. and deliver what U.S. consumers expect from their edamame products are needed. Therefore, in addition to improving nutritional profile and agronomic characteristics of edamame, breeders working on the improvement of this specialty crop for domestic production should also consider the sensory attributes of edamame, such as appearance, flavor, and texture, and consumer preferences (Carneiro et al., 2020).

The current quality and grading standards of edamame are mostly defined by the Japanese, who are still the major consumers and importers of this vegetable in the world (Sirisomboon et al., 2007; Wang, 2018). Traditionally, Japanese consumers prefer sweeter edamame, crisper beans, and flowery flavor, while American consumers prefer more mature beans that have a buttery flavor (Johnson et al., 1999). In a recent review article, Carneiro et al. (2020) summarized the limited information on U.S. consumer preferences for edamame. Despite the increasing consumption of edamame in the U.S over the last decades, only a few studies have investigated consumers' acceptability and preferences for edamame with focus on the U.S. market (Johnson et al., 1999; Young et al., 2000; Kelley and Sánchez, 2005; Wszelaki et al., 2005; Krinsky et al., 2006; Williams, 2015; Wolfe et al., 2018; Flores et al., 2019). Pod appearance, bean taste, aroma, and texture were reported as sensory attributes that significantly affect edamame acceptability among U.S. consumers, who have also shown higher preference for a moderate chewy texture and balanced sweet and nutty flavor (Wszelaki et al., 2005; Flores et al., 2019). Nevertheless, a broader understanding of sensory attributes that can contribute to increase consumer acceptability or may lead to consumer rejection of new edamame genotypes in the U.S. market is still needed to guide breeding programs in the country. This knowledge is also of great value to guide growers and processors in their business decisions. Sensory evaluation can provide essential information for breeding selections and development of new food crops (Hampson et al., 2000). However, there is a lack of information regarding how sensory data could be used in such decision processes. The objectives of this study were to evaluate acceptability of new edamame breeding lines, identify genotypes preferred by consumers in the U.S., and develop a protocol for using consumer sensory data to support breeding selection for edamame genotypes with high market potential.

MATERIALS AND METHODS

In order to support edamame breeders in the process of selecting edamame genotypes for U.S. production, consumer studies were conducted in two sequential years. The material included edamame breeding lines developed by the breeding programs of Virginia Tech (V lines) and University of Arkansas (R lines), and released edamame cultivars ("Asmara" and "UA-Kirksey") as references, commonly called "checks" by plant breeders. Edamame genotypes evaluated in each year are further detailed in sections Edamame Samples and Edamame Samples.

In both years of research, affective evaluation (acceptability) of overall and targeted attributes was completed by untrained consumers of vegetables; participants were adults (18+ years old) and not allergic to soy ($N \geq 50$). Traditional descriptive analyses such as QDA (quantitative descriptive analysis) were not chosen for this project as they do not reflect consumers' acceptability of the samples, and require longer time investment and commitment of trained panelists, which can also increase costs (Carneiro et al., 2020). The faster check-all-that-apply (CATA) descriptive method, which has already been used with affective tests to support breeding programs around the world (Lado et al., 2010; Vela-Hinojosa et al., 2018), was chosen to help explain our acceptability data. Sweetness has been reported as one of the most important sensory attributes to differentiate edamame genotypes (Wszelaki et al., 2005; Flores et al., 2019); in this study sweetness intensity also was included to estimate how consumers in the U.S. may perceive the sweetness of our edamame breeding lines. Consumers were instructed to rate the intensity of the sweet taste compared to other cooked vegetables they typically included in their diet. Although the lack of calibration among panelists is a limitation of our study, the use of untrained panelists to evaluate intensity of edamame flavor characteristics (sweetness, beaniness, nuttiness, and oiliness) and support the selection of genotypes for production was previously reported by Young et al. (2000).

The first year of our research (summer 2018 to spring 2019) is further presented and discussed as the "screening study" [section Screening Study (Year 1)]; it was an exploratory study and was the first year of edamame genotype screening supported by consumer sensory testing. Consumer data obtained in the first year helped us to determine criteria standards for genotype selection of edamame. It also guided the development of a decision tree, which contributed to reducing the total number of edamame genotypes to be analyzed in the sequential year of research.

The second year of our research (summer 2019 to spring 2020) is further presented and discussed as the "validation study" [section Validation Study (Year 2)]; it served as a validation for the selected genotypes and the screening criteria suggested in the first year (screening study). Therefore, year 2 was not a replication of year 1. As described throughout the next sections, the methodology was partially changed from the first year to the second in order to address study limitations, challenges, and points of improvement identified after concluding the first year (e.g., drying step of edamame processing, sample preparation for sensory evaluation, and statistical design of sensory tests).

Screening Study (Year 1)

Edamame Samples

A total of 24 edamame genotypes were planted in the summer 2018 from the following growing locations: Newport, AR, and Blacksburg and Painter, VA. Twenty edamame genotypes were tested in the sensory panels: 2 edamame cultivars ("Asmara" and "UA-Kirksey"), and 18 advanced breeding lines (R08-4004, R13-5029, R14-16195, R14-6238, R14-6450, V09-4192, V10-3653, V13-0329, V13-0339, V13-1644, V15-0344, V15-0396, V15-0398, V15-0411, V16-0523, V16-0524, V16-0528, and V16-0547). Four genotypes (edamame cultivar "LV75" and 3 advanced breeding lines, R07-589, R15-10280, and R16-5336) presented color variability and/or low amount of beans from one or more growing locations, so their incomplete datasets were not included in this report.

Edamame pods were harvested between September and October, 2018, and stored in cold before processed at the Virginia Tech Food Processing Pilot Plant (Blacksburg, VA, USA) within 48 h after harvesting. Samples from more distant growing locations (Newport and Painter) were placed in coolers or Styrofoam boxes containing ice bags and shipped overnight in order to minimize quality changes; samples were processed the next day, upon receipt, at Virginia Tech facility. Distilled water was used in all processing steps, except in the initial washing, when Blacksburg tap water (pH 7.5, 78 mg/L total dissolved solids, disinfected with chloramines Yao et al., 2019), was used to remove most soil and dirt. Fresh edamame pods were placed in metal baskets, blanched (boiling water, 98 ± 1°C) for 1 min in a steam kettle (Legion Utensils Co., Long Island City, NY, USA), and then immersed in a cooling bath (4 ± 1°C) for 2 min to avoid over blanching. Then, samples were manually dried with paper towels, packed in re-sealable plastic storage bags, labeled, and stored in a −20°C walk-in freezer. Blanching was performed following processing conditions suggested by Sheu and Chen (1991), which were determined based on remaining activity of the lipoxygenase and peroxidase enzymes. The same processing conditions were also tested for microbial quality and safety by Pao et al. (2008), who reported that blanching edamame in pod at 98°C for 1 min eliminated inoculated *Escherichia coli* and *Listeria monocytogenes* (≈ 6 log cfu/g), and naturally occurring coliform, yeast, or mold counts were below detection levels. In January and February 2019, samples were thawed for ∼8 h at 4°C in a walk-in fridge and were manually shelled. During this step, yellow beans were separated and discarded, as well as unhealthy beans with visible signs of pest or plant diseases (e.g., brown spots on pods). Shelled beans were stored in re-sealable plastic bags at ∼20°C until the day before sensory evaluation.

Sensory Sample Preparation

Edamame beans were thawed overnight in a refrigerator (4°C), then cooked in polyethylene plastic bags for 4 min in microwave on high power (SHARP household microwave oven, model n° R-2W38, serial n°34328, 120 VAC, 60 Hz, Thailand, 1996) the day before each sensory panel. After cooled to room temperature (≈ 21°C), cooked edamame beans were placed in 2-ounce black and clear plastic cups with clear lids. Each black cup contained one bean (evaluated for overall acceptability) and each clear cup contained two beans (evaluated for additional attribute acceptability and characterization). A random 3-digit code was assigned to each edamame sample for identification and to reduce risk of bias. All cups were labeled with the 3-digit code equivalent to the sample they contained and were kept refrigerated until served. Participants were not given any information about the samples that could influence their decisions.

Sensory Evaluation

The 20 edamame genotypes that had enough beans from all three growing locations (section Edamame Samples) were randomly analyzed within a total of 10 sensory panels, which took place at the Virginia Tech Sensory Evaluation Lab (Blacksburg, VA, USA) in February and March, 2019. The study was approved by the Virginia Tech Institutional Review Board (IRB) for Research Involving Human Subjects (IRB 18-310). Participants were untrained volunteers and consisted of students, faculty, staff, and the general public from Virginia Tech and its surrounding area. Study recruitment was accomplished through physical and online distribution of advertisement flyers, posts on social media and VT Daily News, emails to individuals and listservs, and word-of-mouth. Participants were 18 years or older and vegetable consumers not allergic to soy; they were asked to review and consent to the study parameters in the first session in which they participated. In each sensory panel, a set of 6 edamame samples (samples from all 3 growing locations of 2 edamame genotypes chosen at random) was served to each participant to limit sensory fatigue. Samples were served in a monadic random order and each sample was analyzed by a minimum of 50 people. As the edamame sample set was always different in each one of the 10 sensory panels, volunteers were allowed to participate in one or more panels (up to 10 times) in a period of 6 weeks, but no more than once on a single day. Participants did not receive financial compensation; they were compensated with a snack and/or non-alcoholic beverage (valued at $1 or less per unit) after completing a session of the study.

Participants were seated in individual sensory booths with white daylight lighting and equipped with touchscreen monitor. Data were collected using Compusense® Cloud version 20.0.7373.25578 (Compusense Inc., Guelph, Ontario, Canada) on the monitor. Participants were instructed to first ensure the code on the sample cup was the same code as listed on the monitor, then remove the lid from the sample cup, taste the edamame bean presented in the coded black cup, and evaluate overall-liking of the sample using a 9-point hedonic scale (1 = "dislike extremely," 9 = "like extremely"). In sequence, they were asked to use the same 9-point scale to evaluate acceptability of sensory characteristics of two edamame beans from the same sample (coded clear cup). Participants were directed to smell the beans to evaluate aroma and then visually assess appearance. Next, they were instructed to chew one bean and assess taste and texture liking, evaluate intensity of sweet taste using a 5-point scale (1 = "not sweet," 5 = "extremely sweet"), and answer a check-all-that-apply (CATA) question using a list of sensory attributes to describe the sensory profile of each sample. Panelists could consume the second bean if needed to complete the assessment. The 15 descriptive terms included in the CATA list were chosen from literature and are shown in **Table 1** (Johnsen et al., 1988; Day N'Kouka et al., 2004; Krinsky et al., 2006; Wang et al., 2017). Terms and definitions were available to participants at the beginning of each CATA question and, at the end of the question, the single words that could be selected were presented in a fixed order. Subjects received unsalted saltine crackers and a glass of water for refreshing their palate and rinsing their mouth between samples. In addition to the sensory test, volunteers participating for the first time were asked to first answer a demographic and behavioral survey, then answer an economics survey after the sensory evaluation (data not presented).

TABLE 1 | Check-All-That-Apply (CATA) list of sensory attributes for edamame.

Attribute	Definition	References
Chewy	Food texture that requests long chewing time in order to obtain a satisfactory consistency for swallowing	Wang et al., 2017
Starchy	Aroma, flavor, and aftertaste associated with raw wheat flour	Day N'Kouka et al., 2004
Raw bean	Aroma, flavor, and aftertaste associated with raw soybeans, legumes	Day N'Kouka et al., 2004; Krinsky et al., 2006
Cooked bean	Aroma, flavor, and aftertaste associated with cooked soybeans, legumes	Day N'Kouka et al., 2004; Krinsky et al., 2006
Green/Grassy	Aroma associated with fresh green beans and freshly cut twigs, grass.	Day N'Kouka et al., 2004; Krinsky et al., 2006
Fruity	Aroma and flavor associated with a mixture apple/pear/tropical	Krinsky et al., 2006
Nutty	Aroma and flavor associated with nuts and having legume-like character	Krinsky et al., 2006
Sulfury/ Rotten Egg	Aroma and flavor associated with hydrogen sulfide, rotten egg	Krinsky et al., 2006
Brothy/ Umami	Aroma and flavor associated with boiled meat, soup, stock. Feeling factor associated with glutamate, aspartate, ribonucleotides	Krinsky et al., 2006
Sweet	Basic taste on tongue associated with sugars and high potency sweeteners	Johnsen et al., 1988; Day N'Kouka et al., 2004; Krinsky et al., 2006
Salty	Basic taste on tongue associated with salts and sodium ions	Johnsen et al., 1988; Day N'Kouka et al., 2004; Krinsky et al., 2006
Sour	Basic taste on tongue and aftertaste associated with acids	(Johnsen et al., 1988; Day N'Kouka et al., 2004
Bitter	Basic taste on tongue and aftertaste associated with caffeine, quinine, alkaloids	Johnsen et al., 1988; Day N'Kouka et al., 2004; Krinsky et al., 2006
Astringent	Dry, puckering, chemical feeling factor and aftertaste associated with pure cranberry juice, tannins, alum	Johnsen et al., 1988; Day N'Kouka et al., 2004; Krinsky et al., 2006
Metallic	Metallic, flat chemical feeling factor associated with iron and copper	Johnsen et al., 1988

Validation Study (Year 2)

Edamame Samples

Ten edamame genotypes (1 cultivar, UA-Kirksey, and 9 advanced breeding lines, R14-16195, R14-6238, R14-6450, R15-10280, R16-5336, V10-3653, V16-0524, V16-0528, and V16-0547) were grown during summer 2019 in Blacksburg and Painter, VA, Portageville, MO, and Stoneville, MS. Eleven edamame genotypes were originally selected by breeders after the first year of research, but genotype R07-589 was not included in the sensory evaluation panels due to notable differences in color (reddish beans) in comparison to the other genotypes (greenish beans). Only edamame cultivar UA-Kirksey was used as a reference in the validation study because there were not enough seeds of Asmara cultivar to be planted in the second year. Harvest and processing occurred between August and October 2019. In order to reduce the variability that may be associated with shipping, samples from all four growing locations were placed in coolers containing ice bags and delivered within 24 h after harvest to be processed at Virginia Tech Food Processing Pilot Plant. Processing flow was similar to the first year (section Edamame Samples), but salad spinners were used for the drying step instead of manual drying with paper towels. Additionally, samples were manually shelled right after blanching and edamame beans were stored at −20°C until use.

Sensory Sample Preparation

The day before each sensory test, bags containing frozen edamame beans were placed in a refrigerator (4°C) for 4–6 h before cooking. Edamame beans were placed in a glass microwave-safe container, covered with a paper towel sheet and cooked on high power for 1.5 min in a R-2W38 Sharp microwave oven. After cooking, edamame beans remained in the glass container covered with a paper towel for another 1 min. Similar to the previous year (section Sensory Sample Preparation), beans were placed in 2-ounce black and clear plastic cups with clear lids after cooled to room temperature (21°C). However, in this study, the number of beans was increased compared to the previous year: overall acceptability (black cup) contained two beans instead of only one, and characterization (clear cups) contained three beans instead of two. As in the first year of research, samples were anonymized and all cups were labeled with random 3-digit codes, then kept refrigerated until served.

Sensory Evaluation

The research protocol was updated by researchers and approved by the Virginia Tech Institutional Review Board for Research Involving Human Subjects (IRB 18-310). Consumers were recruited to participate in up to four sensory panels (one per growing location), which took place at the Virginia Tech Sensory Evaluation Lab (Blacksburg, VA, USA) in February 2020. Recruitment method and selection criteria were the same as the previous year (session Sensory Evaluation). Participants were adults (18 years or older) who were vegetable consumers and not allergic to soy, and were students, faculty, staff, or the general public from Virginia Tech and its surrounding area. As the edamame sample set was different in each sensory panel (same 10 genotypes, but a different growing location every week), subjects who agreed with study parameters and signed the consent forms were allowed to participate in one or multiple panels in the period of 4 weeks, but no more than once a week. After completing a session of this study, participants were compensated with a snack and/or non-alcoholic beverage (valued at $1 or less per unit) for their time and efforts. In the first panel, in which volunteers participated, they were instructed to answer a demographic and behavioral survey before the sensory test, then answer a new economics survey (different then first year's survey) after the sensory evaluation (data not presented). Sensory studies were performed as described in section Sensory Evaluation (Year 1), but the experimental design was changed to a balanced incomplete block design (BIBD) that required a total of 90 participants per sensory panel. The BIBD allowed each edamame breeding line to be analyzed by 50 people, and the control (cultivar UA-Kirksey) to be analyzed by all 90 participants of the panel. Therefore, the new experimental design and the smaller number of genotypes allowed a direct comparison between the control and all breeding lines, which was an improvement in comparison with our screening study. In each panel, 6 edamame samples (5 randomly selected breeding lines and the control) were served in a monadic random order to each one of the 90 participants along with a glass of water and unsalted saltine crackers. Participants were asked to evaluate overall-liking and characteristics (aroma, appearance, taste, and texture liking, sweetness intensity, and the CATA), as described in section Sensory Evaluation.

Statistical Analysis

Hedonic and intensity scores were analyzed using analysis of variance (ANOVA) and multiple comparison of means were conducted using Tukey's Honestly Significant Difference (HSD) tests. Unchecked and checked terms in the CATA were assigned codes 0 and 1, respectively. Correspondence analysis (CA) was performed to investigate relationships between sensory attributes, genotypes, and growing locations. Penalty analysis (PA) was conducted to identify which sensory terms contributed to higher or lower hedonic and intensity scores. Statistical analysis was performed in R, RStudio, and JMP Pro® and a 5% significance level was considered for all tests.

RESULTS AND DISCUSSION

Screening Study (Year 1)

Participants

A total of 182 untrained volunteers participated in the screening study. About 58% of the participants were female, 63% were 20–29 years old, 60% were white, Caucasian, and 56% had at least a bachelor's degree. Participants reported they intentionally consume vegetables at least once a day (71%), soy-based products a few times per month (39%), and edamame a few times per year (48%). Additionally, "like the taste" (65%) and "for heath reason" (45%) were reported as their main motivators to consume soy products.

Edamame Selection for Breeding Programs

The quality of edamame can be affected by environmental factors like soil, weather and climate (e.g., temperature, light intensity, rainfall), and season (Masuda, 1991). The growing regions have different soil types and weather conditions (**Table 2**).

The decision tree shown in **Figure 1** was developed and used to assist breeders of Virginia Tech (V lines) and University of Arkansas (R lines) breeding programs in their selection criteria. In the screening study, 10 experimental edamame genotypes with the best sensory attributes were identified and suggested to continue in the next steps of the breeding programs mentioned above: V16-0547, V13-0329, R14-16195, V16−0524, R13-5029, V15-0396, V13-0339, R14-6238, V13-1644, and R14-6450. Previous studies reported overall acceptability for edamame cultivars varied between 5.5 and 6.84 based on a 9-point hedonic scale (1 = "dislike extremely"; 5 = "neither like nor dislike" 7 = "like moderately"; 9 = "like extremely") (Kelley and Sánchez, 2005; Wszelaki et al., 2005). Thus, an overall acceptability mean score of 6.0 ("like slightly") or higher on a 9-point hedonic scale was considered a good target for this study. Due to the fact sweetness plays a relevant role to differentiate edamame genotypes (Wszelaki et al., 2005; Flores et al., 2019), sweetness intensity mean score was considered an important score to be considered in our decision tree.

TABLE 2 | Environmental characteristics of growing locations.

	Blacksburg, VA		Painter, VA		Newport, AR		Portageville, MO		Stoneville, MS	
	Average temperature (°C)	Total percipitation (mm)	Average temperature (°C)	Total percipitation (mm)	Average temperature (°C)	Total percipitation (mm)	Average temperature (°C)	Total percipitation (mm)	Average temperature (°C)	Total percipitation (mm)
2018										
May	19.2	107.2	21.6	127.3	24.2	79.8				
June	21.6	28.2	24.3	114.8	26.8	43.4				
July	22.2	72.4	24.8	153.9	27.9	34.6				
August	23.9	103.6	26.1	67.3	26.1	144.2				
September	22.5	185.7	24.6	239.8	23.5	163.6				
October	16.7	91.9	17.2	108.5	16.3	113.1				
2019										
May	19.9	66.8	21.8	45.2			21.7	218.7	23.9	320.8
June	21.8	84.3	24.2	105.7			24.4	198.4	26.1	193.8
July	22.8	65.3	26.4	220.2			26.6	128.3	27.8	139.2
August	23.4	61.7	25.3	34.3			25.9	172.5	27.8	102.6
September	22.6	11.7	23.2	43.7			25.8	26.4	28.3	10.2
October	16.7	160.8	18.3	171.2			15.5	203.5	18.3	287.3
Soil Type	Hayter loam: fine-loamy, mixed, active, mesic Ultic Hapludalfs		Bojac sandy loam: Coarse-loamy, mixed, semiactive, thermic Typic Hapludults		Dexter silt loam: fine-silty, mixed, active, thermic Ultic Hapludalfs		Dundee silt loam, fine-silty, mixed, active, thermic Typic Endoapualfs		Sharkey clay. Very-fine, smectitic, thermic Chromic Epiaquerts	

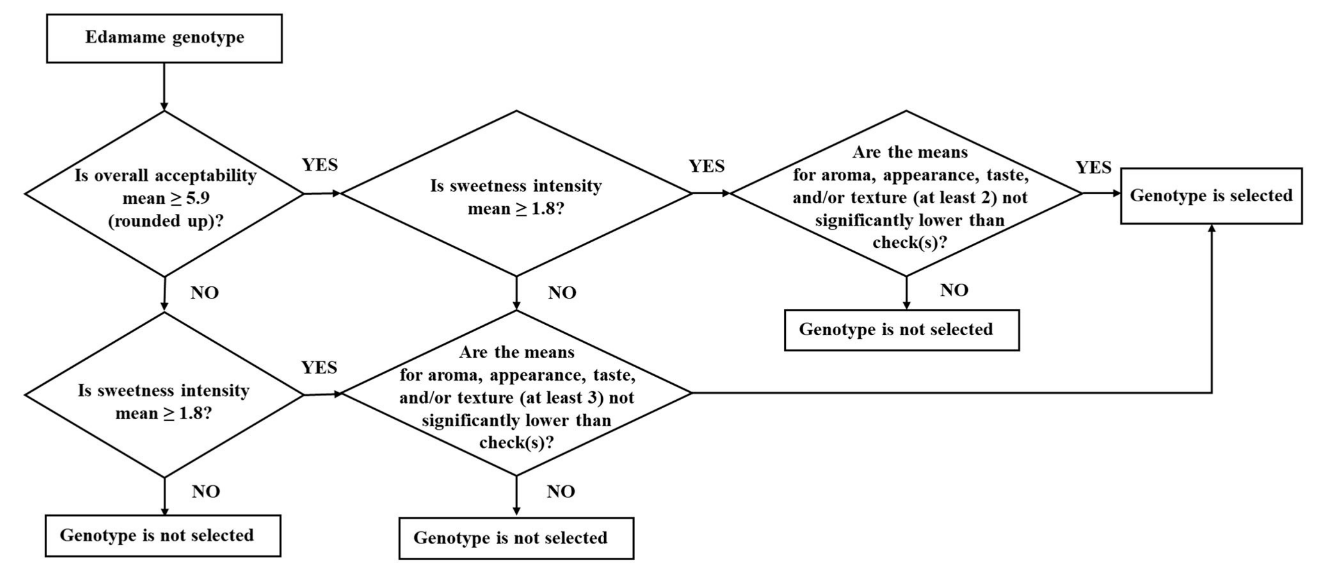

FIGURE 1 | Decision tree for selection of edamame genotypes based on sensory evaluation scores.

As a reference, Young et al. (2000) reported the sweetness intensity mean score for a group of 31 edamame genotypes evaluated by untrained panelists was 1.85 (5-point intensity scale, 1 = "not intense," 5 = "extremely intense"). In our decision process, we also considered that the combination of sensory characteristics that were within the same range of our references (cultivars Asmara and UA-Kirksey) would suggest potential commercial value.

In summary, our selected genotypes had a mean of 5.9 (rounded up) or higher for overall acceptability (OA) and/or mean of 1.8 or higher (1 = "not sweet"; 2 = "slightly sweet"; 5 = "extremely sweet") for sweetness intensity (SI). The criteria for selection included a combination of at least four attributes, which necessarily included overall acceptability and/or sweetness intensity as described above, plus mean scores that were not significantly lower than the references for aroma, appearance, taste, and/or texture means (**Table 3**). Mean scores for each edamame genotype represent the average of all three growing locations scores. A large number of samples with limited amount of beans was a limiting factor for this study and required several days of sensory panels and a large number of volunteers. Ideally, all samples should be evaluated by all the participants to increase the statistical power of the study, but this was not viable. Thus, a smaller number of genotypes with larger amount of beans and a new experimental design were chosen for the validation study (year 2).

Edamame Acceptance

Panelists had very different opinions about edamame overall-liking. The average acceptance scores (mean value ± standard deviation; average of 3 growing locations) for edamame genotypes ranged from 5.1 ± 2.0 (V15-0344) to 6.3 ± 1.7 (UA-Kirksey, reference). Growing location did not have a significant effect in overall acceptability ($p > 0.05$). However, the interaction between genotypes and location was significant, as well as the genotype factor ($p < 0.05$), which means the changes in genotype scores were different depending on growing location. Wszelaki et al. (2005) investigated appearance of edamame pods and beans, aroma, taste, texture, and aftertaste of beans, and overall acceptability of six commercial edamame genotypes (Sapporo Midori, White Lion, Early Hakucho, Sayamusume, Misono Green, and Kenko), which were grown organically in Ohio. The hedonic scores reported in **Table 3** were close to aroma, taste, and texture liking scores of edamame beans reported by Wszelaki et al. (2005). In their study, sweetness intensity was not investigated and panelists ($N = 54$, 46% females, ages between 20 and 60 years old) may have considered sensory attributes of edamame pods (not edible) to evaluate overall acceptability of their samples. In our study, both taste and texture mean scores ranged between 5 ("neither like nor dislike") and 7 ("like moderately"), which was similar to the scores reported by Wszelaki et al. (2005). Appearance mean scores reported by Wszelaki et al. (2005) for edamame beans were between 6 ("like slightly") and 7 ("like moderately"), which were higher than the mean scores obtained in this study (range: 4.9–6.6; 4 = "dislike slightly" and 7 = "like moderately"). However, aroma scores reported in their study were slightly lower (range: 5.3–5.7) than mean scores shown in **Table 3**.

The frequencies (percentages) at which sensory descriptors from the CATA question were used by our panelists to describe each edamame genotype are shown in **Table 4**. The most used descriptors were "cooked bean," "green/grassy," "chewy," "nutty," "raw bean," "starchy," and "sweet" (frequency >25% for most genotypes). Penalty analysis (PA) is a traditional tool used to analyze Just-About-Right (JAR) data; it has been used in the analysis of CATA data to understand how sensory attributes that are not in their optimal levels can cause drop in overall-liking scores (Qannari, 2017). Results of the PA performed using

TABLE 3 | Screening Study: Sensory scores (overall acceptability, sweetness intensity, aroma, appearance, taste, and texture) of edamame genotypes for breeding selection.

Genotype	Overall acceptability[1]	Sweetness intensity[2]	Aroma[1]	Appearance[1]	Taste[1]	Texture[1]	Suggestion[3]
UA-Kirksey	6.3 ± 1.7^{a}	1.9 ± 0.9	6.3 ± 1.4^{a,b,c}	6.6 ± 1.6^{a}	6.2 ± 1.8^{a}	6.4 ± 1.5^{a}	Reference
Asmara	6.1 ± 1.8^{a,b}	1.9 ± 1.0	6.4 ± 1.4^{a,b}	6.0 ± 1.7^{a,b,c,d}	5.9 ± 1.7^{a,b}	6.0 ± 1.7^{a,b,c,d}	Reference
V16-0547	**6.1 ± 2.0**^{a,b}	**2.0 ± 1.0**	**6.3 ± 1.7**^{a,b,c,d}	**6.2 ± 1.7**^{a,b}	**5.8 ± 1.8**^{a,b}	**6.0 ± 1.6**^{a,b,c,d}	Include
V13-0329	**6.0 ± 1.9**^{a,b}	**1.9 ± 0.8**	**6.0 ± 1.6**^{a,b,c,d,e}	5.9 ± 1.6^{b,c,d}	**5.9 ± 1.7**^{a,b}	**5.9 ± 1.7**^{a,b,c,d}	Include
R14-16195	**6.0 ± 1.8**^{a,b}	**1.9 ± 0.9**	**5.8 ± 1.7**^{b,c,d,e}	5.7 ± 1.7^{b,c,d}	**5.8 ± 1.7**^{a,b}	**6.0 ± 1.5**^{a,b,c,d}	Include
V16-0524	**6.0 ± 1.7**^{a,b}	**1.8 ± 0.8**	**6.0 ± 1.8**^{a,b,c,d,e}	**6.4 ± 1.7**^{a,b}	**6.0 ± 1.6**^{a}	**6.0 ± 1.8**^{a,b,c,d}	Include
R13-5029	**6.0 ± 1.9**^{a,b}	**2.1 ± 1.0**	5.7 ± 1.8^{c,d,e}	5.9 ± 1.6^{b,c,d}	**6.0 ± 1.9**^{a}	**6.2 ± 1.6**^{a,b}	Include
V15-0396	**5.9 ± 1.6**^{a,b}	1.8 ± 0.8	**5.8 ± 1.6**^{a,b,c,d,e}	5.5 ± 1.6^{c,d,e}	**6.0 ± 1.5**^{a,b}	**5.8 ± 1.7**^{a,b,c,d}	Include
V13-0339	**5.9 ± 1.8**^{a,b}	**1.8 ± 0.8**	5.6 ± 1.7^{d,e}	5.5 ± 1.7^{c,d,e}	**5.9 ± 1.6**^{a,b}	**6.1 ± 1.7**^{a,b,c}	Include
R14-6238	**5.9 ± 1.6**^{a,b}	**1.8 ± 0.9**	**6.5 ± 1.4**^{a}	**6.0 ± 1.6**^{a,b,c,d}	**6.0 ± 1.6**^{a}	**5.9 ± 1.7**^{a,b,c,d}	Include
V13-1644	**5.9 ± 1.9**^{a,b}	**2.0 ± 1.0**	**5.8 ± 1.7**^{b,c,d,e}	**6.2 ± 1.7**^{a,b}	**5.7 ± 1.9**^{a,b}	**6.1 ± 1.7**^{a,b,c,d}	Include
R14-6450	**5.9 ± 1.8**^{a,b}	**1.9 ± 0.9**	**6.0 ± 1.5**^{a,b,c,d,e}	5.4 ± 1.9^{c,d,e}	**5.8 ± 1.6**^{a,b}	5.7 ± 1.9^{b,c,d}	Include
R08-4004	5.8 ± 1.8^{a,b,c}	1.8 ± 0.8	**6.1 ± 1.8**^{a,b,c,d,e}	**6.1 ± 1.8**^{a,b,c}	**5.8 ± 1.7**^{a,b}	**6.1 ± 1.7**^{a,b,c,d}	Exclude
V16-0523	5.8 ± 2.0^{a,b,c}	**1.9 ± 0.9**	5.6 ± 1.7^{e}	4.9 ± 2.0^{e}	**5.7 ± 1.8**^{a,b}	5.5 ± 1.9^{c,d}	Exclude
V15-0398	5.8 ± 1.9^{a,b,c}	1.8 ± 0.8	**6.0 ± 1.6**^{a,b,c,d,e}	5.8 ± 1.7^{b,c,d}	**5.9 ± 1.7**^{a,b}	**5.9 ± 1.9**^{a,b,c,d}	Exclude
V10-3653	5.7 ± 1.8^{a,b,c}	1.7 ± 0.8	5.7 ± 1.7^{c,d,e}	5.3 ± 2.0^{d,e}	**5.8 ± 1.7**^{a,b}	5.7 ± 1.9^{b,c,d}	Exclude
V15-0411	5.6 ± 1.7^{a,b,c}	**1.8 ± 0.8**	5.6 ± 1.7^{e}	4.9 ± 1.8^{e}	**5.6 ± 1.6**^{a,b}	5.4 ± 1.7^{c,d}	Exclude
V09-4192	5.5 ± 1.9^{b,c}	**1.8 ± 0.8**	5.7 ± 1.6^{c,d,e}	5.9 ± 1.7^{b,c,d}	**5.6 ± 1.8**^{a,b}	**5.9 ± 1.6**^{a,b,c,d}	Exclude
V16-0528	5.5 ± 1.8^{b,c}	**1.9** ± 0.9	**5.8 ± 1.5**^{b,c,d,e}	4.9 ± 1.8^{e}	**5.6 ± 1.7**^{a,b}	5.5 ± 1.7^{c,d}	Exclude
V15-0344	5.1 ± 2.0^{c}	**1.8 ± 0.9**	5.6 ± 1.7^{e}	5.4 ± 1.8^{d,e}	5.3 ± 1.7^{b}	5.4 ± 1.8^{d}	Exclude

[a–e]Means ± SD followed by a letter in common are not significantly different ($p > 0.05$).

[1]Scale: 1 = "dislike extremely"; 5 = "neither like nor dislike"; 9 = "like extremely"; [2]Scale: 1 = "not sweet"; 5 = "extremely sweet."

[3]Criteria for selection include (1) overall acceptability: mean of 5.9 (rounded up) or higher; (2) sweetness intensity: mean of 1.8 or higher; (3) aroma, appearance, taste, and texture: no significant lower than both references; (4) combination of at least 4 attributes, which must include overall acceptability and/or sweetness intensity. Bolded means indicate a match to criteria. "Reference" refers to a selected cultivar (control genotype).

TABLE 4 | Citation frequency (%) for sensory descriptors from Check-All-That-Apply (CATA) list used to describe edamame genotypes in the screening study.

Genotype	Chewy	Starchy	Raw bean	Cooked bean	Green/ Grassy	Fruity	Nutty	Sulfury/ Rotten egg	Brothy/ Umami	Sweet	Salty	Sour	Bitter	Astringent	Metallic
UA-Kirksey*	**42.0%**	**26.0%**	**34.7%**	**58.7%**	**40.0%**	15.3%	**40.0%**	2.7%	9.3%	**34.0%**	6.7%	0.7%	6.0%	8.7%	6.7%
Asmara*	**37.2%**	**27.6%**	**37.8%**	**47.4%**	**47.4%**	11.5%	**41.7%**	3.8%	6.4%	**30.8%**	8.3%	3.8%	9.6%	5.8%	5.8%
V16-0547**	**46.0%**	**26.7%**	**33.3%**	**54.0%**	**39.3%**	15.3%	**39.3%**	4.0%	11.3%	**31.3%**	4.0%	4.0%	10.7%	9.3%	10.0%
V13-0329**	**37.3%**	**35.3%**	**50.0%**	**40.0%**	**46.0%**	12.0%	**47.3%**	9.3%	10.7%	**28.0%**	8.7%	4.0%	8.0%	3.3%	8.7%
R14-16195**	**42.9%**	**27.6%**	**38.5%**	**51.3%**	**41.7%**	12.8%	**36.5%**	9.0%	9.0%	**28.8%**	7.1%	7.1%	13.5%	5.8%	17.9%
V16-0524**	**44.0%**	**38.0%**	**35.3%**	**50.0%**	**49.3%**	14.7%	**42.0%**	8.7%	12.0%	**28.7%**	3.3%	4.0%	10.0%	1.3%	6.7%
R13-5029**	**43.4%**	**30.2%**	**41.5%**	**49.1%**	**49.1%**	11.3%	**35.2%**	5.7%	11.3%	**32.7%**	6.3%	5.0%	7.5%	3.8%	5.7%
V15-0396**	**39.9%**	**36.6%**	**44.4%**	**50.3%**	**55.6%**	11.1%	**36.6%**	3.9%	10.5%	20.3%	5.9%	1.3%	9.8%	2.6%	4.6%
V13-0339**	**40.5%**	**31.4%**	**37.9%**	**49.0%**	**41.2%**	9.8%	**45.1%**	5.2%	9.2%	**34.0%**	2.6%	3.3%	8.5%	3.9%	9.8%
R14-6238**	**44.9%**	**32.7%**	**31.4%**	**50.6%**	**47.4%**	9.0%	**43.6%**	4.5%	7.7%	**30.1%**	6.4%	2.6%	5.8%	3.8%	10.9%
V13-1644**	**45.1%**	**32.7%**	**50.3%**	**37.9%**	**53.6%**	9.8%	**33.3%**	7.2%	7.2%	**28.8%**	7.8%	0.7%	8.5%	5.2%	3.3%
R14-6450**	**41.7%**	**29.5%**	**38.5%**	**42.9%**	**40.4%**	14.7%	**35.3%**	9.6%	11.5%	**30.8%**	5.1%	4.5%	8.3%	5.1%	12.8%
R08-4004	**39.3%**	**34.7%**	**39.3%**	**47.3%**	**61.3%**	13.3%	**45.3%**	7.3%	8.0%	**32.7%**	4.7%	3.3%	10.7%	3.3%	9.3%
V16-0523	**42.5%**	**33.3%**	**43.1%**	**43.8%**	**33.3%**	9.2%	**35.9%**	5.2%	13.1%	24.8%	9.2%	2.0%	9.2%	3.3%	7.2%
V15-0398	**36.7%**	**38.7%**	**49.3%**	**42.0%**	**44.0%**	9.3%	**46.7%**	8.7%	14.0%	24.0%	5.3%	2.0%	13.3%	4.7%	8.0%
V10-3653	**43.1%**	**34.6%**	**34.6%**	**42.5%**	**42.5%**	14.4%	**41.2%**	7.8%	7.2%	20.9%	6.5%	2.0%	11.1%	2.6%	13.7%
V15-0411	**38.6%**	**34.6%**	**41.8%**	**37.9%**	**41.2%**	7.2%	**39.9%**	3.3%	11.8%	22.9%	9.2%	0.7%	9.2%	0.7%	7.2%
V09-4192	**40.9%**	**29.6%**	**47.2%**	**45.3%**	**57.2%**	8.2%	**42.1%**	6.3%	9.4%	18.2%	6.3%	1.9%	8.8%	3.8%	8.2%
V16-0528	**38.0%**	**28.7%**	**30.7%**	**52.0%**	**36.0%**	8.7%	**32.0%**	4.0%	5.3%	**29.3%**	5.3%	2.7%	6.0%	6.0%	5.3%
V15-0344	**42.7%**	**32.0%**	**33.3%**	**47.3%**	**34.7%**	7.3%	**37.3%**	4.7%	5.3%	**27.3%**	4.7%	2.0%	13.3%	4.0%	8.7%

*Edamame genotypes from each location were evaluated by 50–53 untrained participants. This table combines data from the 3 locations (Newport, AR, and Blacksburg and Painter, VA). High responses for an attribute (≥25% frequency) are bolded. *"Check" (selected cultivar; control genotype). **Suggested genotypes to continue in the breeding selection.*

the overall acceptability scores showed the CATA attributes "salty," "sweet," "brothy/umami," "nutty," "cooked bean," "fruity," and "chewy" contributed positively to the acceptability of edamame genotypes when they were present, but liking scores dropped when the attributes "starchy," "raw bean," "metallic," "green/grassy," "astringent," "sour," "sulfury/rotten egg," and "bitter" were used to describe the samples (**Figure 2A**). The CATA terms that were associated with higher acceptability of edamame samples were the same ones associated with higher taste scores. "Chewy" was the major sensory attribute associated with texture in our CATA list, followed by "starchy." Blanching and cooking parameters (time and temperature) can affect edamame texture (Konovsky et al., 1994). However, as all samples were blanched and cooked following the same parameters, differences among hedonic scores were most likely associated with differences in genotypes and harvesting conditions. PA performed using the texture hedonic scores showed that lower texture scores were obtained when "chewy" was associated with the samples, but the opposite was observed for overall acceptability. "Starchy" had a negative impact on texture scores as well as observed for overall acceptability. Flores et al. (2019) reported increasing hardness was positively correlated to texture liking scores (based on an 11-point hedonic scale; 0 = "Do not like at all," 10 = "Like extremely"). Hardness of edamame seeds was not measured in this study and is suggested for further research. Our results are in accordance with the study of Wszelaki et al. (2005) which suggested an equilibrium between sweet and nutty attributes, plus a moderate chewy texture would be preferred characteristics for edamame consumer in U.S.

The relationship between genotypes and sensory attributes was verified through correspondence analysis (CA) (**Figure 3A**). The first two dimensions explained 51% of the variance of the data, which was a relatively small amount. Both references, Asmara and UA-Kirksey, and most genotypes that were suggested to continue in the breeding programs were associated with the desired attribute "sweet." Savory flavor in edamame can be associated with amino acids content, and sucrose content is the major contributor to the sweet flavor (Konovsky et al., 1994). The suggested genotypes R14-16195 (OA = 6.0) and R14-6450 (OA = 5.9) were associated with the undesired attributes "bitter," "sour," and "metallic," which can be an important aspect to be considered by the breeders. Bitter taste can be associated with the enzyme lipoxygenase, which concentration increases as beans mature, and/or the presence of saponins (Masuda, 1991; Konovsky et al., 1994; Young et al., 2000). Furthermore, the off-flavors "bitter," "astringent,", and "metallic" that contributed to lower taste scores are associated with undesired dry-mouth feeling (Masuda, 1991). In our study, lipoxygenase activity and saponin contents were not measured. It was assumed that processing conditions effectively inactivated oxidation enzymes present in the samples as the blanching method and time/temperature were suggested by others (Sheu and Chen, 1991). Thus, further investigation would be needed to understand whether bitterness of samples was associated with the genotypes, late harvest, or ineffective

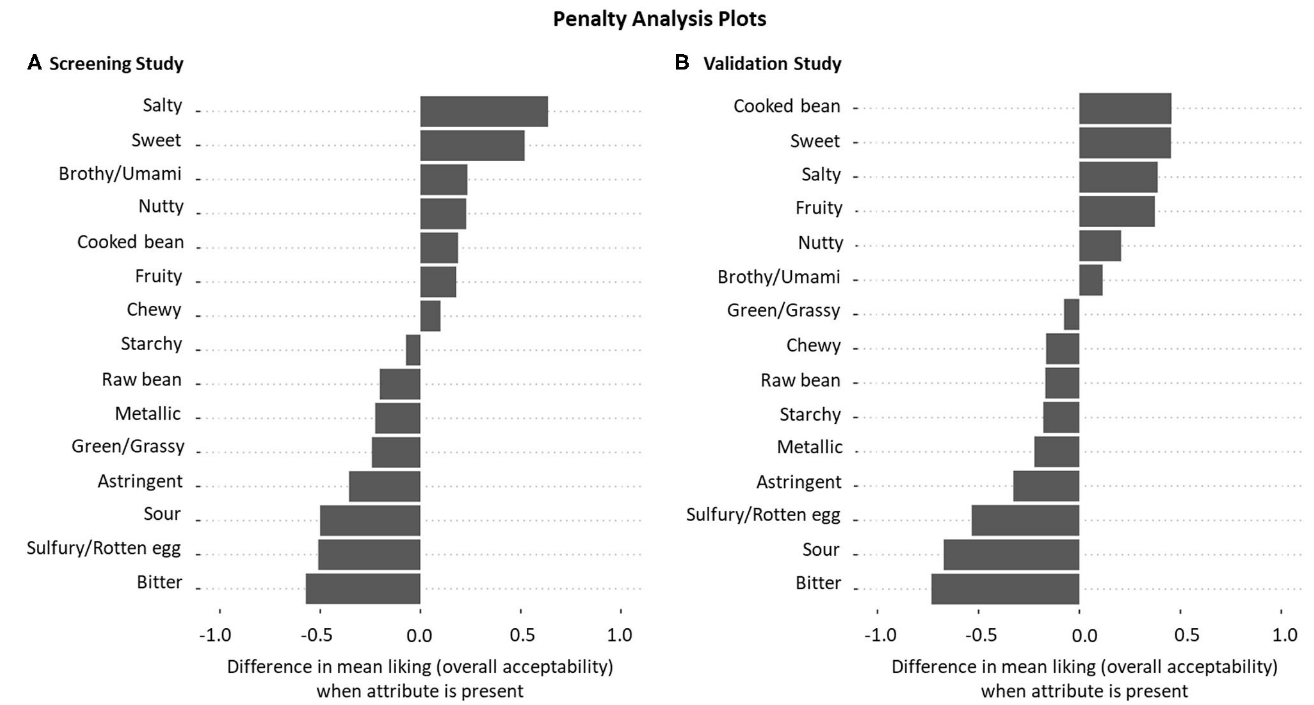

FIGURE 2 | Penalty analysis plots representing the impact of sensory descriptors from the Check-All-That-Apply (CATA) list in the overall acceptability scores of edamame genotypes. **(A)** corresponds to the screening study and **(B)** corresponds to the validation study.

blanching. Our experimental blanching was the same for all genotypes we studied, so it is not likely that bitterness differences between genotypes were related to blanching effectiveness.

Environmental factors can affect sensory traits and affect quality and value of vegetables (Ferreira et al., 2012). Thus, the relationship between growing location and sensory attributes was also verified and the first two dimensions of the CA explained 100% of the sensory space (**Figure 2B**). The CA factor map for location showed that samples from Newport, AR, were discriminated from samples from the other two growing locations in Virginia (**Table 2**). As explained by the first dimension of the CA map (83%), genotypes grown in Newport were more associated with the negative terms "sour," "astringent," "bitter," "green/grassy," and "raw beans," while genotypes grown in Blacksburg and Painter, VA were more associated with the attributes "sweet," "nutty," and "chewy." In addition, Blacksburg and Newport samples were also more associated to the terms "salty" (positive, as shown by PA plot) and "sulfury/rotten egg" (negative). This analysis may suggest Painter, VA, on the Eastern Shore and a region known for growing produce and vegetables, as the location with most potential to grow high-quality edamame; however, a deeper understanding of other factors like agronomic or climatic conditions, and harvest standards is needed to confirm this potential. Some genotypes may perform better (or lower) in one location than others, and in our study interactions between genotypes and locations were not controlled, which was a limitation of the study.

Validation Study (Year 2)

Participants

A total of 171 untrained volunteers participated in the validation study. Their demographic and behavior profiles were similar to the profiles reported in the screening study (section Participants). Participants were mostly female (51%), age between 20 and 29 years old (56%), white, Caucasian (46%), and had at least a bachelor's degree (68%). Most participants reported they intentionally consume vegetables one or more times per day (67%), soy-based products a few times per month (34%), and edamame a few times per year (50%). Lastly, their major motivations to consume soy products were also "like the taste" (48%) and "for health reason" (40%).

Edamame Selection for Breeding Programs

The decision tree developed in the screening study (**Figure 1**; explained in section Edamame Selection for Breeding Programs) was used again to suggest whether edamame genotypes should continue or not in the breeding programs after our validation study (**Table 5**). Regarding the edamame genotypes that were analyzed in the screening study, V10-3653 and V16-0528 were suggested to be excluded from the 2019 field trials, but they were selected by breeders to continue due to agronomic traits such as high fresh yield. As in the first year, V10-3653 (OA = 5.8) did not meet the selection criteria and was suggested to be excluded from the breeding programs, but V16-0528 (OA = 5.8) had a different performance and was suggested to continue in the program. Moreover, we observed that the overall acceptability

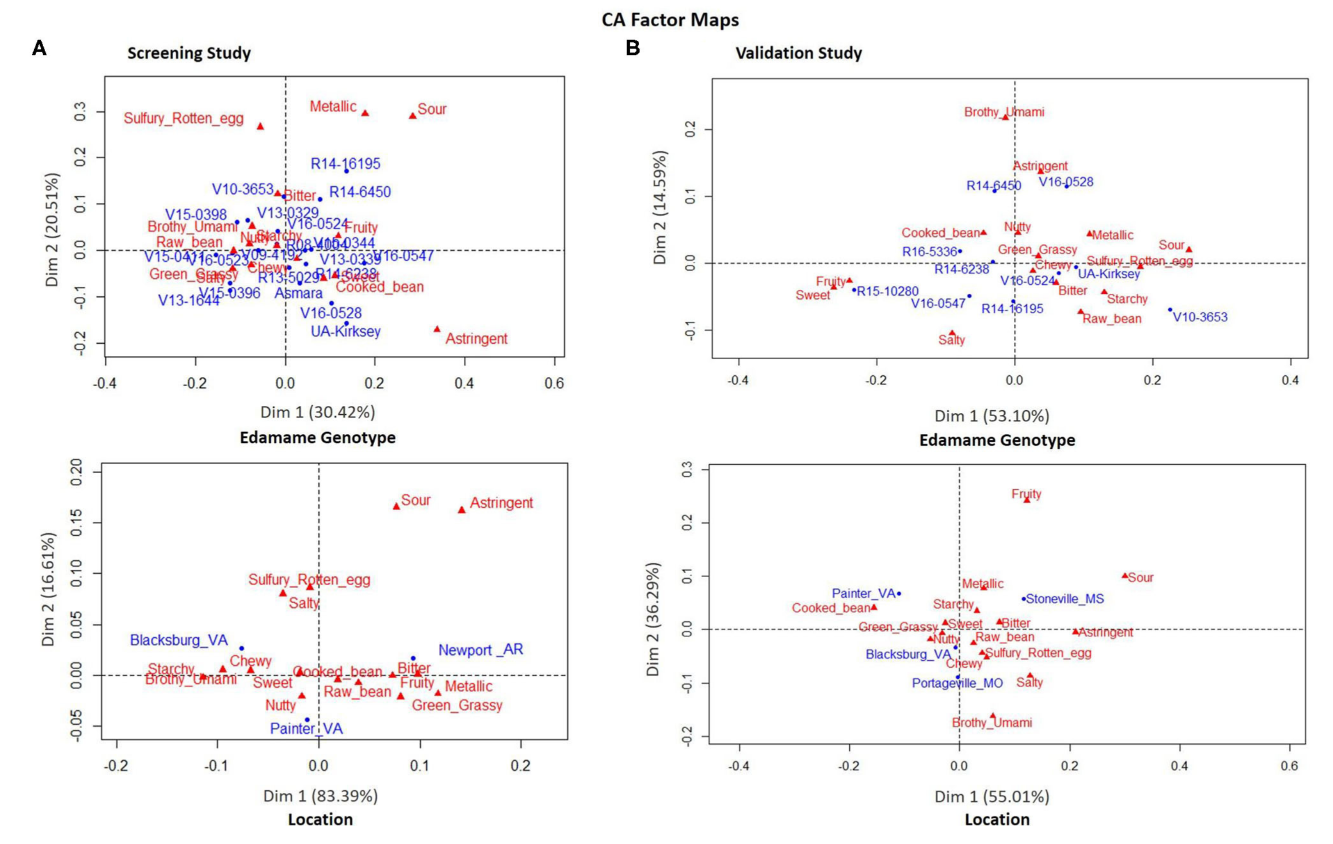

FIGURE 3 | Correspondence analysis (CA) factor maps for edamame genotypes and growing locations. **(A)** correspond to the screening study and **(B)** correspond to the validation study.

of the reference UA-Kirksey in the validation study (OA = 5.9) was lower than previous year (OA = 6.3), but this variability could be due to the variable nature of consumer data. It was not possible to assure which factors mainly contributed to the reduced acceptability of this cultivar in the second year. However, the BIBD used to collect sensory data in our validation study allowed consumers to make better direct comparisons among multiple genotypes, which may have led to the changes observed in the mean scores. The breeding line R15-10280, which was not tested in the screening study due to low amount of beans from one growing location, had the highest overall acceptability (OA = 6.3) and sweetness intensity score (SI = 2.4) among all genotypes; it was also more preferred than UA-Kirksey in taste and texture. Therefore, R15-10280 was suggested as the genotype with the most preferred sensory profile and great market potential.

In both years, our decision tree (**Figure 1**) was used to suggest which edamame genotype should be included or excluded from the breeding programs based on our sensory data (**Tables 2, 4**). Overall, most suggestions (include or exclude) made after analyzing our validation study data confirmed the suggestions made in the first year, which suggests that our criteria were appropriate. Williams (2015) previously reported a set of criteria used to select edamame genotypes that included both agronomic and sensory characteristics of edamame, but almost no information was given about how sensory data was obtained or analyzed. The author only reported that sensory evaluation criteria was based on edamame (pods and seeds) being acceptable to a vegetable processor, which meant “two- to three-seed pods, green pods and seeds, seed free of blemishes, a smooth seed texture, and seed with a sweet and/or nutty flavor.” However, the references used to determine whether an edamame bean had a smooth texture, or a sweet and/or nutty flavor were not presented, which makes the criteria difficult to be replicated or adapted. On the other hand, the detailed sensory evaluation tests and decision process used in our study could be easily replicated in future edamame studies or adapted to guide other plant breeding programs in the development of improved vegetables, fruits and nuts.

Edamame Acceptance

Genotype, location and their interaction significantly affected overall acceptability scores ($p < 0.05$). As reported in **Table 5**, overall acceptability (mean value ± standard deviation; average of 4 growing locations) of edamame genotypes ranged from 5.8 ± 1.7 (V16-0528) to 6.4 ± 1.7 (R15-10280). The frequencies (percentage) of which sensory terms from the CATA list were used by our panelists to describe each edamame genotype are shown in **Table 6**. Similar to the screening study results, the terms

TABLE 5 | Validation study: sensory scores (overall acceptability, sweetness intensity, aroma, appearance, taste, and texture) of edamame genotypes for breeding selection.

Genotype	Overall acceptability[1]	Sweetness intensity[2]	Aroma[1]	Appearance[1]	Taste[1]	Texture[1]	Suggestion[3]
UA-Kirksey	5.9 ± 1.7^{b}	$1.8 \pm 0.9^{c,d}$	5.7 ± 1.5	$6.0 \pm 1.6^{a,b,c}$	$5.7 \pm 1.7^{b,c}$	$5.8 \pm 1.7^{b,c}$	Reference
R15-10280*	**6.4 ± 1.7^{a}**	**2.4 ± 1.2^{a}**	**5.9 ± 1.7**	**$6.1 \pm 1.6^{a,b,c}$**	**6.3 ± 1.8^{a}**	**6.3 ± 1.7^{a}**	Include
R14-16195	**$6.3 \pm 1.8^{a,b}$**	**$1.9 \pm 0.9^{b,c,d}$**	**5.8 ± 1.7**	**$6.3 \pm 1.6^{a,b}$**	**$6.0 \pm 1.8^{a,b,c}$**	**$6.3 \pm 1.6^{a,b}$**	Include
R16-5336*	**$6.2 \pm 1.6^{a,b}$**	**$2.0 \pm 1.0^{b,c}$**	**5.8 ± 1.6**	**$5.9 \pm 1.7^{b,c}$**	**$6.1 \pm 1.5^{a,b,c}$**	**$6.2 \pm 1.6^{a,b,c}$**	Include
V16-0547	**$6.2 \pm 1.7^{a,b}$**	**2.1 ± 0.9^{b}**	**5.8 ± 1.6**	**$6.1 \pm 1.6^{a,b,c}$**	**$6.1 \pm 1.6^{a,b}$**	**$6.3 \pm 1.5^{a,b,c}$**	Include
R14-6238	**$6.2 \pm 1.7^{a,b}$**	**$2.0 \pm 1.0^{b,c}$**	**5.7 ± 1.6**	**6.5 ± 1.7^{a}**	**$6.1 \pm 1.7^{a,b}$**	**$6.2 \pm 1.8^{a,b,c}$**	Include
V16-0524	**$6.1 \pm 1.7^{a,b}$**	**$2.0 \pm 0.9^{b,c}$**	**5.7 ± 1.5**	**$6.1 \pm 1.5^{a,b,c}$**	**$5.9 \pm 1.6^{a,b,c}$**	**$6.3 \pm 1.6^{a,b}$**	Include
R14-6450	**$6.1 \pm 1.6^{a,b}$**	**$2.1 \pm 0.9^{b,c}$**	**5.6 ± 1.7**	**$5.8 \pm 1.7^{b,c}$**	**$5.9 \pm 1.6^{a,b,c}$**	**$6.3 \pm 1.6^{a,b,c}$**	Include
V10-3653**	5.8 ± 1.6^{b}	1.6 ± 0.8^{d}	**5.8 ± 1.5**	**$6.0 \pm 1.6^{b,c}$**	**5.5 ± 1.6^{c}**	5.8 ± 1.6^{c}	Exclude
V16-0528**	5.8 ± 1.7^{b}	**$1.8 \pm 0.8^{b,c,d}$**	**5.6 ± 1.6**	**5.6 ± 1.7^{c}**	**$5.6 \pm 1.7^{b,c}$**	**$5.9 \pm 1.7^{a,b,c}$**	Include

[a–d]Means ± SD followed by a letter in common are not significantly different ($p > 0.05$).
[1]Scale: 1 = "dislike extremely"; 5 = "neither like nor dislike"; 9 = "like extremely"; [2]Scale: 1 = "not sweet"; 5 = "extremely sweet."
[3]Criteria for selection include (1) overall acceptability: mean of 5.9 (rounded up) or higher; (2) sweetness intensity: mean of 1.8 or higher; (3) aroma, appearance, taste, and texture: no significant lower than reference; (4) combination of at least 4 attributes, which must include overall acceptability and/or sweetness intensity. Bolded means indicate a match to criteria. "Reference" refers to a selected cultivar (control genotype).
**Genotypes not tested in the screening study. **Genotypes suggested to be excluded in the screening study.*

TABLE 6 | Citation frequency (%) for sensory descriptors from Check-All-That-Apply (CATA) list used to describe edamame genotypes in the validation study.

Genotype	Chewy	Starchy	Raw bean	Cooked bean	Green/Grassy	Fruity	Nutty	Sulfury/Rotten egg	Brothy/Umami	Sweet	Salty	Sour	Bitter	Astringent	Metallic
UA-Kirksey*	**38.6%**	**31.1%**	**41.4%**	**31.1%**	**46.9%**	10.6%	**43.3%**	5.0%	10.3%	24.4%	9.2%	3.6%	11.7%	5.0%	11.7%
R15-10280**	**35.0%**	23.5%	**36.5%**	**35.0%**	**41.5%**	16.0%	**48.5%**	3.0%	9.5%	**52.5%**	12.0%	2.0%	8.5%	5.5%	5.5%
R14-16195**	**37.5%**	**27.5%**	**43.5%**	**39.0%**	**40.5%**	10.0%	**43.0%**	3.0%	7.0%	**32.0%**	10.0%	3.0%	11.0%	3.5%	11.0%
R16-5336**	**31.0%**	**28.0%**	**38.0%**	**35.0%**	**44.5%**	18.0%	**51.0%**	5.0%	11.5%	**36.5%**	8.0%	3.0%	9.0%	3.0%	11.5%
V16-0547**	**40.0%**	**25.0%**	**46.0%**	**36.5%**	**55.5%**	14.0%	**41.5%**	3.5%	9.5%	**39.5%**	10.5%	1.0%	10.5%	3.5%	13.5%
R14-6238**	**32.0%**	**29.0%**	**44.5%**	**37.0%**	**49.5%**	14.5%	**43.5%**	4.5%	11.5%	**35.0%**	9.0%	3.5%	7.0%	5.5%	10.0%
V16-0524**	**34.5%**	**28.0%**	**48.5%**	**30.0%**	**43.5%**	8.5%	**44.5%**	6.0%	12.0%	**28.0%**	8.0%	1.5%	7.0%	3.0%	11.0%
R14-6450**	**32.0%**	20.0%	**36.5%**	**36.0%**	**48.5%**	11.0%	**43.0%**	3.5%	15.0%	**30.5%**	8.0%	2.5%	9.0%	5.0%	13.5%
V10-3653	**35.0%**	**37.0%**	**54.5%**	**28.0%**	**50.5%**	7.0%	**46.0%**	6.0%	7.5%	18.0%	8.5%	5.0%	10.5%	5.0%	12.0%
V16-0528**	**38.5%**	**30.5%**	**38.5%**	**38.5%**	**47.0%**	7.5%	**54.5%**	4.5%	14.0%	**26.0%**	6.5%	3.5%	9.5%	7.0%	10.5%

Edamame genotypes from each location were evaluated by 50 untrained participants, except UA-Kirksey, which was evaluated by 90 participants. This table combines data from the 4 locations (Blacksburg and Painter, VA, Portageville, MO, and Stoneville, MS). High responses for an attribute (≥25% frequency) are bolded.
**"Reference" (selected cultivar; control genotype). **Suggested genotypes to continue in the breeding selection.*

"cooked bean," "raw bean," "green/grassy," "chewy," "starchy," "nutty," and "sweet" were the most used terms (frequency >25% for most genotypes). Penalty analysis (PA) confirmed the positive association between the sensory attributes "cooked bean," "sweet," "salty," "fruity," "nutty," and "brothy/umami" and higher acceptability of edamame genotypes (**Figure 2B**). Therefore, these six flavor attributes were confirmed as desired sensory characteristics to be considered when breeding to develop edamame genotypes for the U.S. market. The descriptors "green/grassy," "raw bean," "starchy," "metallic," "astringent," "sulfury/rotten egg," "sour," and "bitter," once more, were associated with lower acceptability scores and were confirmed as undesired edamame sensory attributes for consumers in the U.S. According to Vara-Ubol et al. (2004), some sensory attributes like "brown" "green/pea pod," "musty/dusty," "musty/earthy," "nutty," and "starchy" flavors, sour aromatics, and a powdery feel (texture) can be associated with the undesired "beany" characteristic, frequently associated with soybean products. The PA of our validation study did not confirm the texture attribute "chewy" as a sensory characteristic associated with higher acceptability scores. Wszelaki et al. (2005) reported chewiness is an important sensory attribute to differentiate edamame genotypes and suggested it is a desired attribute that increases with maturity of pods and beans. Cooking can affect texture of edamame (Young et al., 2000); it is possible that our changes in microwave cooking conditions (sensory sample preparation) affected texture of the edamame beans. However, texture and appearance mean scores obtained in our validation study were, in general, higher than the mean scores obtained in our screening study, which suggests changes in microwave cooking were positive (edamame beans were placed in a glass microwave-safe container instead of being cooked in polyethylene plastic bags and microwave cooking time was reduced from 4 to 1.5 min). Although texture acceptability was evaluated in our study, further sensory and analytical studies are suggested to better understand which edamame texture characteristics are desired or preferred by consumers.

Correspondence analyses (CA) were performed to verify relationships among sensory attributes and genotypes, and sensory attributes and growing locations (**Figure 3B**). The first two dimensions of the CA map for edamame genotype explained 68% of the sensory space. R15-10280 was the breeding line that obtained the highest scores for overall acceptability, sweetness intensity, aroma, taste and texture in our validation study. This genotype was highly associated with the sensory attributes "sweet," "fruity," and "salty" (third quadrant of the CA map). As opposed to that, V10-3653 was suggested both years to be excluded from the breeding programs and was mostly associated with the sensory descriptors "starchy," "raw bean," and "bitter" (fourth quadrant of the CA map). In our validation study, higher overall acceptability was observed for Painter, VA, but it was not significantly different from Portageville, MO, and Blacksburg, VA for the edamame grown in the 2019 season. Stoneville, MS, had the lowest overall acceptability, but it did not differ significantly from Blacksburg, VA. Environmental characteristics of all four locations are presented in **Table 2**. In the CA map for location (**Figure 3B**), Stoneville, MS was separated from the other three growing locations by the first dimension of the map. Samples from this location were more associated with the most rejected sensory attributes "bitter" and "sour." The poor taste of the Stoneville samples may be partially explained by the historically high temperature and lack of rain immediately preceding harvest (**Table 2**). The two dimensions of the CA map for location explained 91% of the variance in the data. However, it was not possible to assure whether the differences shown by the sensory data were due to environment or harvest conditions. In fact, the 2019 season in Stoneville, MS was considered very hot, but irrigation was done only earlier in the season in this location, while in Painter, VA irrigation was done throughout the whole season. The different irrigation practices and climatic conditions (**Table 2**) may have contributed to the different edamame sensory profiles observed among growing locations, but might not be the only factors. Further investigation is suggested to identify the best practices to grow high-quality edamame in the U.S, and to better understand differences among growing locations, interactions between genotypes and locations, and identify best genotypes for the different growing locations (suitability).

CONCLUSION

Consumer studies performed in two sequential years successfully supported the selection of edamame genotypes and allowed the development of a decision tree that can be adapted to provide guidance to present and future plant breeding programs for incorporating consumer acceptability data to support breeding decisions. References and standards used for breeding selection criteria depend on the food crop (vegetable, fruit, nut) targeted for development or improvement. In our study, for example, it was chosen to exclude edamame genotypes with acceptability (hedonic) scores lower than 5.9, even when the scores were not significantly different than this standard. Additional criteria (sweetness; acceptability of additional attributes) were included to create a holistic evaluation. Thus, breeding selection results may differ among research groups that choose different references and standards for a same crop. Although in the first year of our research some statistical analyses were limited by the experimental design, the validation study (year 2) sustained the results of the previous screening study (year 1). A reduced number of samples (genotypes/growing locations) and larger availability of plant material for analysis (in this study, edamame beans) are recommended for future affective testing aiming to support breeding decisions because they allow the use of a more stringent experimental design, such as the BIBD used in the validation study. Overall, sensory data provided a better understanding of U.S. consumers' perception and acceptability of edamame; "salty" and "sweet" were confirmed as the major natural sensory attributes that drove higher acceptability scores, while "bitter" was highly associated with lower acceptability. Future work would benefit by the use of trained panels and descriptive sensory evaluation methods, such as QDA. Descriptive methods could be used to quantify valuable sensory attributes of selected breeding lines. As our edamame breeding programs continue, this valuable consumer information contributes to the improvement of edamame genotypes to be tested in the next years. It has also been contributing to the identification of advanced breeding lines with market potential in the U.S., such as R15-10280. This breeding line had consumer acceptability scores significantly higher than reference cultivar UA-Kirksey and was suggested as a strong candidate to be grown at the locations reported in this study and to be commercially released. However, different/new genotypes may have a better performance in different growing locations, which justifies breeding efforts to continue in order to find a genotype that best suits to those potential growing locations. Moreover, as edamame samples were discriminated by growing location, the fact that genotype interactions with environmental conditions (e.g., soil, weather) and agricultural practices (e.g., irrigation, pest, and weed control) were not controlled was identified as a limitation of our study. Further research is recommended to develop a better picture of each location potential to grow high-quality edamame and understand which best practices should be followed by growers. Our data for the edamame genotypes investigated in this study suggested Painter, VA as a good potential location to grow this specialty crop, but it was not possible to assure which main factor(s) contributed to the differences observed in acceptability (e.g., environmental and/or harvest conditions). In summary, consumer studies and sensory evaluation are valuable tools to guide breeders, growers, and processors in the development, selection, and production of high-quality edamame with sensory attributes desired by consumers in the U.S.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because dataset is still needed to support further investigations and publications. Requests to access the datasets should be directed to Dr. Susan Duncan (duncans@vt.edu).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Virginia Tech Institutional Review Board (IRB 18-310). The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RC, CN, and SD contributed to the conception and design of the sensory study. RC and DY worked on edamame processing. RC collected and analyzed sensory data, wrote first draft, and edited manuscript. BZ, JR, PC, and AG developed the breeding lines and managed the experimental plots that produced the edamame for this study. TK, SR, and MR provided agronomic support for the edamame throughout the growing seasons, identified harvest time, and harvested the edamame for this study. All authors reviewed, edited, and approved final version of the manuscript.

FUNDING

This work was funded by USDA-NIFA (Grant No. 2018-51181-28384; Accession No. 1016465), and, in part, by the Virginia Agricultural Experiment Station. This project was partially supported by the ARS Project Number 6066-21220-012-00D. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

ACKNOWLEDGMENTS

The authors thank all members of our edamame team (faculty, staff, and students) for their dedication and hard work. The authors also thank all volunteers who took part in our consumer studies, Drs. Katherine Phetxumphou and Jacob Lahne for their assistance during the sensory tests and support with statistical analysis, and the Virginia Tech Statistical Applications & Innovations Group (SAIG) for their support with data analysis.

REFERENCES

Carneiro, R. C. V., Duncan, S. E., O'Keefe, S. F., Yin, Y., Neill, C. L., and Zhang, B. (2020). Sensory and consumer studies in plant breeding: a guidance for edamame development in the U.S. *Front. Sustain. Food Syst*. 4:124. doi: 10.3389/fsufs.2020.00124

Carson, L., Freeman, J., Zhou, K., Welbaum, G., and Reiter, M. (2011). Cultivar evaluation and lipid and protein contents of Virginia-grown edamame. *Horttechnology* 1, 131–135. doi: 10.21273/HORTTECH.21.1.131

Day N'Kouka, K., Klein, B. P., and Lee, S. Y. (2004). Developing a lexicon for descriptive analysis of soymilks. *J. Food Sci*. 69, 259–263. doi: 10.1111/j.1365-2621.2004.tb13625.x

Ernst, M., and Woods, T. (2001). "Marketing challenges for emerging crops in Kentucky: vegetable soybeans," in *Second International Vegetable Soybean Conference* (Tacoma, WA).

Ferreira, J. J., del Castillo, R. R., Perez-Vega, E., Plans, M., Simó, J., and Casañas, F. (2012). Sensory changes related to breeding for plant architecture and resistance to viruses and anthracnose in bean market class Fabada (*Phaseolus vulgaris L.*). *Euphytica* 186, 687–696. doi: 10.1007/s10681-011-0540-9

Flores, D., Giovanni, M., Kirk, L., and Liles, G. (2019). Capturing and explaining sensory differences among organically grown vegetable-soybean varieties grown in Northern California. *J. Food Sci*. 84, 613–622. doi: 10.1111/1750-3841.14443

Hampson, C. R., Quamme, H. A., Hall, J. W., Macdonald, R. A., King, M. C., and Cliff, M. A. (2000). Sensory evaluation as a selection tool in apple breeding. *Euphytica* 111, 79–90. doi: 10.1023/A:1003769304778

Johnsen, P. B., Civille, G. V., Vercellotti, J. R., Sanders, T. H., and Dus, C. A. (1988). Development of a lexicon for the description of peanut flavor. *J. Sens. Stud*. 3, 9–17. doi: 10.1111/j.1745-459X.1988.tb00426.x

Johnson, D., Wang, S., and Suzuki, A. (1999). "Edamame: a vegetable soybean for Colorado," in *Perspectives on New Crops and New Uses*, ed. J. Janick (Alexandria, VA: ASHS Press), 385–387.

Kelley, K. M., and Sánchez, E. S. (2005). Accessing and understanding consumer awareness of and potential demand for edamame. *HortScience* 40, 1347–1353. doi: 10.21273/HORTSCI.40.5.1347

Konovsky, J., Lumpkin, T. A., and McClary, D. (1994). "Edamame: the vegetable soybean," in *Understanding the Japanese Food and Agrimarket: A Multifaceted Opportunity*, ed A. D. O'Rourke (Binghamton, NY: Haworth Press), 173–181

Krinsky, B. F., Drake, M. A., Civille, G. V., Dean, L. L., Hendrix, K. W., and Sanders, T. H. (2006). The development of a lexicon for frozen vegetable soybeans (edamame). *J. Sens. Stud*. 21, 644–653. doi: 10.1111/j.1745-459X.2006.00088.x

Kumar, V., Rani, A., Goyal, L., Pratap, D., Billore, S. D., and Chauhan, G. S. (2011). Evaluation of vegetable-type soybean for sucrose, taste-related amino acids, and isoflavones contents. *Int. J. Food Prop*. 14, 1142–1151. doi: 10.1080/10942911003592761

Lado, J., Vicente, E., Manzzioni, A., and Ares, G. (2010). Application of a check-all-that-apply question for the evaluation of strawberry cultivars from a breeding program. *J. Sci. Food Agric*. 90, 2268–2275. doi: 10.1002/jsfa.4081

Masuda, R. (1991). "Quality requirement and improvement of vegetable soybean," in *Vegetable Soybean: Research Needs for Production and Quality Improvement*, ed. S. *Shanmugasundaram* (Taiwan: Asian vegetable research and Development Center), 92−102.

Mebrahtu, T., and Devine, T. E. (2008). Combining ability analysis for selected green pod yield components of vegetable soybean genotypes (Glycine max). *New Zeal. J. Crop Hortic. Sci*. 36, 97–105. doi: 10.0014–0671/08/3602-0097

Mebrahtu, T., Mohamed, A., Wang, C. Y., and Andebrhan, T. (2004). Analysis of isoflavone contents in vegetable soybeans. *Plant Foods Hum. Nutr*. 59, 55–61. doi: 10.1007/s11130-004-0023-4

Ntatsi, G., Egea Gutiérrez-Cortines, M., Karapanos, I., Barros, A., Weiss, J., Balliu, A., et al. (2018). The quality of leguminous vegetables as influenced by preharvest factors. *Sci. Hortic.* 232, 191–205. doi: 10.1016/j.scienta.2017.12.058

Ogles, C. Z., Guertal, E. A., and Weaver, D. B. (2016). Edamame cultivar evaluation in Central Alabama. *Agron. J*. 108, 2371–2378. doi: 10.2134/agronj2016.04.0218

Pao, S., Ettinger, M. R., Khalid, M. F., Mebrahtu, T., and Mullins, C. (2008). Microbiological quality of frozen "edamame" (vegetable soybean). *J. Food Saf*. 28, 300–313. doi: 10.1111/j.1745-4565.2008.00121.x

Qannari, E. M. (2017). Sensometrics approaches in sensory and consumer research. *Curr. Opin. Food Sci*. 15, 8–13. doi: 10.1016/j.cofs.2017.04.001

Sheu, S. C., and Chen, A. O. (1991). Lipoxygenase as blanching index for frozen vegetable soybeans. *J. Food Sci*. 56, 448–451. doi: 10.1111/j.1365-2621.1991.tb05300.x

Sirisomboon, P., Pornchaloempong, P., and Romphophak, T. (2007). Physical properties of green soybean: criteria for sorting. *J. Food Eng*. 79, 18–22. doi: 10.1016/j.jfoodeng.2006.01.022

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). *Ind. Crops Prod*. 50, 743–749. doi: 10.1016/J.INDCROP.2013.08.064

Vara-Ubol, S., Chambers Ivz, E., and Chambers, D. H. (2004). Sensory characteristics of chemical compounds potentially associated with beany aroma in foods. *J. Sens. Stud*. 19, 15–26. doi: 10.1111/j.1745-459X.2004.tb00133.x

Vela-Hinojosa, C., Escalona-Buendía, H. B., Mendoza-Espinoza, J. A., Díaz de León-Sánchez, F., Lobato-Ortíz, R., Rodríguez-Pérez, J. E., et al. (2018). Chemical and sensory analysis of native genotypes and experimental lines of tomato (*Solanum lycopersicum L.*). *Fruits* 73, 60–71. doi: 10.17660/th2018/73.1.7

Wang, K. (2018). East Asian food regimes: agrarian warriors, edamame beans and spatial topologies of food regimes in East Asia. *J. Peasant Stud*. 45, 739–756. doi: 10.1080/03066150.2017.1324427

Wang, S., Adhikari, K., and Hung, Y.-C. (2017). Acceptability and preference drivers of freshly roasted peanuts. *J. Food Sci*. 82, 174–184. doi: 10.1111/1750-3841.13561

Williams, M. M. (2015). Phenomorphological characterization of vegetable soybean germplasm lines for commercial production. *Crop Sci*. 55, 1274–1279. doi: 10.2135/cropsci2014.10.0690

Wolfe, E., Popp, M., Bazzani, C., Nayga, R. M., Danforth, D., Popp, J., et al. (2018). Consumers' willingness to pay for edamame with a genetically modified label. *Agribusiness* 34, 283–299. doi: 10.1002/agr.21505

Wszelaki, A. L., Delwiche, J. F., Walker, S. D., Liggett, R. E., Miller, S. A., and Kleinhenz, M. D. (2005). Consumer liking and descriptive analysis of six varieties of organically grown edamame-type soybean. *Food Qual. Prefer*. 16, 651–658. doi: 10.1016/J.FOODQUAL.2005.02.001

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. *J. Food Process. Technol*. 3:165. doi: 10.4172/2157-7110.1000165

Yao, W., Gallagher, D. L., Marr, L. C., and Dietrich, A. M. (2019). Emission of iron and aluminum oxide particles from ultrasonic humidifiers and potential for inhalation. *Water Res*. 164:114899. doi: 10.1016/j.watres.2019.114899

Young, G., Mebrahtu, T., and Johnson, J. (2000). Acceptability of green soybeans as a vegetable entity. *Plant Foods Hum. Nutr*. 55, 323–333. doi: 10.1023/A:1008164925103

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORIGINAL RESEARCH
published: 18 January 2021
doi: 10.3389/fpls.2020.585856

Effect of Planting Date and Cultivar Maturity in Edamame Quality and Harvest Window

David Moseley[1], Marcos Paulo da Silva[2], Leandro Mozzoni[2]*, Moldir Orazaly[2], Liliana Florez-Palacios[2], Andrea Acuña[2], Chengjun Wu[2] and Pengyin Chen[3]

[1] LSU AgCenter, Alexandria, LA, United States, [2] Department of Crop, Soil, and Environmental Sciences, University of Arkansas, Fayetteville, AR, United States, [3] Division of Plant Sciences, University of Missouri, Columbia, MO, United States

Edited by:
Martin Williams,
Agricultural Research Service,
United States Department of Agriculture, United States

Reviewed by:
Silvia Pampana,
University of Pisa, Italy
Charles Bowen,
Soybean/Maize Germplasm, Pathology, and Genetics Research, Agricultural Research Service (USDA), United States
Daljeet Dhaliwal,
University of Illinois at Urbana-Champaign, United States

***Correspondence:**
Leandro Mozzoni
lmozzon@uark.edu

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: 21 July 2020
Accepted: 15 December 2020
Published: 18 January 2021

Citation:
Moseley D, da Silva MP, Mozzoni L, Orazaly M, Florez-Palacios L, Acuña A, Wu C and Chen P (2021) Effect of Planting Date and Cultivar Maturity in Edamame Quality and Harvest Window. Front. Plant Sci. 11:585856. doi: 10.3389/fpls.2020.585856

Edamame is a food-grade soybean [*Glycine max* (L.) Merr.] that is harvested immature between the R6 and R7 reproductive stages. To be labeled as a premium product, the edamame market demands large pod size and intense green color. A staggered harvest season is critical for the commercial industry to post-harvest process the crop in a timely manner. Currently, there is little information to assist in predicting the optimum time to harvest edamame when the pods are at their collective largest size and greenest color. The objectives of this study were to assess the impact of cultivar, planting date, and harvest date on edamame color, pod weight, and a newly minted Edamame Harvest Quality Index combining both aforementioned factors. And to predict edamame harvest quality based on phenological stages, thermal units, and planting dates. We observed that pod color and weight depended on the cultivar, planting date, and harvest date combination. Our results also indicated that edamame quality is increased with delayed planting dates and that quality was dependent on harvest date with a quadratic negative response to delaying harvest. Maximum quality depended on cultivar and planting and harvest dates, but it remained stable for an interval of 18–27 days around the peak. Finally, we observed that the number of days between R1 and harvest was consistently identified as a key factor driving edamame quality by both stepwise regression and neural network analysis. These research results will help define a planting and harvest strategy for edamame production in Arkansas and the United States Mid-South.

Keywords: edamame, quality, harvest date, planting date, color, pod weight

INTRODUCTION

Edamame (vegetable soybean) is a food-grade soybean [*Glycine max* (L.) Merr.], which is harvested immature between the reproductive stages of R6 and R7, when the beans fill 80–90% of the pod (Konovsky et al., 1994; Shanmugasundaram and Yan, 2004). As a vegetable product, the appearance of the pod and bean must be acceptable for end consumers. The main physical attributes of edamame include large seed weight (>30 g per 100 seeds) and large and green crescent shaped pods with two or three seeds (Mentreddy et al., 2002; Shanmugasundaram and Yan, 2004). Production of edamame in the United States is thought to have started in the 1950s including home gardens and food processors. Demand for edamame in the United States has seen a dramatic increase since the early 2000s (Mentreddy et al., 2002). Nuss (2013) reported that between 22,600 and 27,000 Mg of edamame per year was consumed in the United States, estimated to be a $175–$200 million market. The United States is one of the top soybean-producing countries; therefore, soybean growers have the potential to produce edamame competitively, since commodity soybean and edamame share requirements of photoperiod sensitivity, fertilization practices, disease management, and irrigation techniques (Nuss, 2013; Ross, 2013; Ogles et al., 2016).

Soybean development and maturation are divided into vegetative and reproductive physiological stages (Fehr et al., 1971). Reproductive stages are characterized by blooming (R1 and R2), pod development (R3 and R4), seed filling (R5 and R6), and plant maturity (R7 and R8) stages (Fehr et al., 1971). Pod and seed appearance and seed composition change during these reproductive stages. Previous work from our research team, working on food-grade soybean cultivars including edamame materials, demonstrated that protein initially decreases for 3–5 weeks after flowering but then begins to accumulate, contrarily to oil that is accumulated steadily during the early reproductive stages (Saldivar et al., 2011). Also, starch and sucrose contents steadily decrease with seed development, while oligosaccharides remain low in seed until 3 weeks prior to R8 stage (Saldivar et al., 2011). In addition, Xu et al. (2016), working on two edamame cultivars, reported that seed weight peaks at the R6 stage, observed a continuous decrease in seed green color from R5 to R8 stage, and confirmed the report by Saldivar et al. (2011) on seed protein, oil, and carbohydrate accumulation patterns. Such drastic changes in soybean seed composition with stage of development highlight the importance of a timely harvest of edamame to ensure both maximum seed size and an optimal seed composition.

Edamame under commercial production is typically harvested using a modified green bean picker. To spread out crop risks and to even the flow of materials entering post-harvest processing facilities, the edamame crop is typically stagger-planted through various dates and maturity-group combinations. Nolen et al. (2016) reported that techniques such as these can extend the harvesting season to several months and that a staggered harvest is critical due to the short window a cultivar will have acceptable pod size and color. It has been reported that the range from reproductive stages R5.8–R7.0 can be 18–20 days (Purcell et al., 2014); however, Nolen et al. (2016) suggested that the harvest window for an acceptable edamame product can be less than 18 days.

Soybeans will mature faster as the nights become longer (Garner and Allard, 1920). Garner and Allard (1920) added that photoperiodism is a major factor in soybean yield. Johnson et al. (1960) indicated photoperiodism can affect later stages of reproductive development, not just triggering flowering. In addition, some soybean cultivars are less sensitive than others to delayed planting and changes in photoperiod (Johnson et al., 1960), whereas very early cultivars [maturity group (MG) 00 and 0] have been reported not to be sensitive (Polson, 1972). In addition, as the relative maturity increases, the soybean reproductive growth stages become increasingly more sensitive to long nights (Johnson et al., 1960; Major et al., 1975b).

The ability to predict the harvest date of many horticulture crops is based on accumulated thermal units (Tu) above a crop-specific base temperature throughout the crop's growing season (Oliver and Annandale, 1998; Miller et al., 2001). The base temperature below which growth and development of soybean stop is 7°C (Boote et al., 1998). Previous research has suggested that it is possible to use temperature in correlation with growth (Major et al., 1975b), but it has also been reported that predicting soybean growth stages using thermal units may be no more accurate than using calendar days (Major et al., 1975a). Therefore, the objectives of this study were to first assess the impact of planting date and harvest date on edamame pod color and pod weight on three edamame cultivars of contrasting maturity and growth habit; second, to identify the effect of planting date and harvest date on a newly defined Edamame Harvest Quality Index (EHQI) for each of the three aforementioned cultivars; and, third, to predict edamame harvest quality based on phenological stages, thermal units, and planting dates using Stepwise Regression and Artificial Neural Network Analysis.

MATERIALS AND METHODS

Field Experimental Design

The experiment was designed as a split-split plot with three replications. The whole plot was planting date (three levels), the split-plot was edamame cultivar ("8080," "R08-4002," and "R09-345"), and the split-split plot was harvest date (eight levels) nested within planting date by cultivar. Harvest was initiated when a cultivar within a planted date reached R5.8 stage (Fehr et al., 1971) on the plot assigned for the first Harvest Date and continued approximately every 5 days on each of the subsequent Harvest Date plots. Harvesting was discontinued when the crop reached R7 (yellowing of pods); therefore, not all harvest date plots were used for every cultivar and planting date, as the total number of harvests depended on speed of crop progression to R7 stage. It is noteworthy that because the three cultivars in this study represent different MGs and because of the variation in photoperiod and temperature across planting dates, a given cultivar could be at different physiological stages across planting dates even if harvested the same number of days after R5.8.

Of the cultivars used, 8080 was an indeterminate MG3 cultivar, whereas R08-4002 and R09-345 were determinate MG5 and MG6 breeding cultivars, respectively. The experiment was grown over 2 years (2014 and 2015) in two locations, the Arkansas Agricultural Research and Extension Center (AAREC) in Fayetteville, AR, United States and the Vegetable Research Station in Kibler, AR, United States. Soils of the former are silt loam (Johnsburg Series; fine-silty, mixed, active, mesic Aquic Fragiudults), while for the latter were very fine sandy loam (Roxana Series; coarse-silty, mixed, superactive, nonacid, thermic Typic Udifluvents) (Soil Survey Staff, 2017). The three planting dates were mid to late May (PD1), mid to late June (PD2), and mid of July (PD3), representing planting dates typically used for edamame production in Arkansas. Each plot consisted of four rows 10.7 m long and 0.91 m wide. The seeding rate was 33 seeds per meter row, resulting in

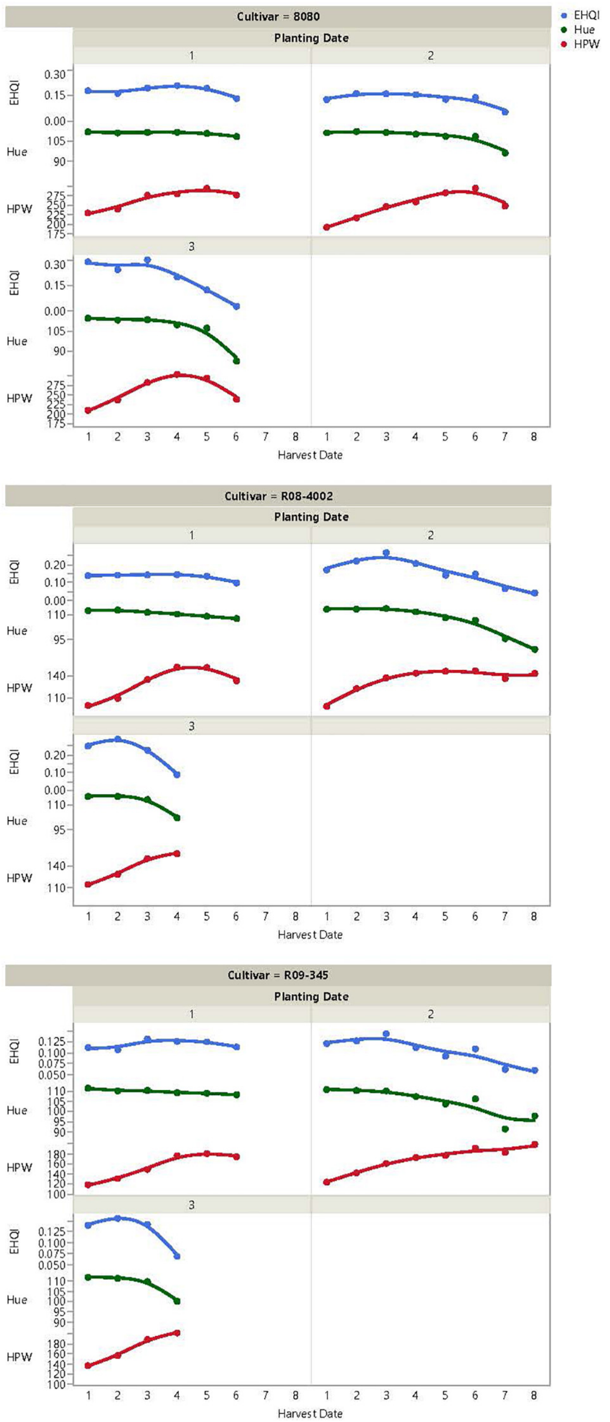

FIGURE 1 | Edamame Harvest Quality Index (EHQI), Hue, and Hundred-Pod Weight (HPW) of three soybean cultivars when planted in three planting dates and subjected to weekly harvest dates, beginning at R5.8 stage and discontinuing at R7 physiological stage.

approximately 16 seeds per meter row at emergence. The plots were managed using standard agricultural practices for irrigation, fertilizer, and pesticides. At harvest, a total of 100 pods were randomly picked by hand throughout the canopy within the middle two rows of each four-row plot. The pods were immediately sealed in plastic bags and placed on ice. Then, the pod samples were blanched in a 100°C water bath (Mozzoni et al., 2009) and stored in a refrigerator at 1.6°C to maintain freshness until color determination (Tsay and Sheu, 1991).

Traits Assessed

A description of the traits assessed is as follows:

PlantingDate_DOY: Day of year when planting occurred.
Ve (Day of emergence): Calendar date when the cotyledons have pulled through the soil surface.
VeDate_DOY: Day of year when Ve occurred.
R1 (Day of first flower): Calendar date when the first flowers emerged in at least one plant of the plot.
R1Date_DOY: Day of year when R1 occurred.
#Days Ve-R1: Number of days elapsed between Ve and R1 for a given plot.
Harvest date: Calendar date when plot was harvested.
HarvestDate_DOY: Day of year when harvest occurred.
#Days R1-Harvest: Number of days elapsed between R1 and harvest for a given plot.
#Days Ve-Harvest: Number of days elapsed between Ve and harvest for a given plot.
GDD Ve-R1 (Growing Degree Days to R1): The Tu was calculated as described by Miller et al. (2001) with a base temperature suggested by Boote et al. (1998) for the days between Ve and R1, with data generated by weather station nearby the trial location.
GDD Ve-Harvest: Calculated as described by Miller et al. (2001) using observed thermal units between Ve and Harvest.
GDD R1-Harvest: Calculated as described by Miller et al. (2001) using observed thermal units between R1 and day of Harvest.
IGC (Intensity of Green Color): Pod color was measured with a HunterLab ColorFlex (Hunter Associates Laboratory Inc., Reston, VA, United States). The instrument was calibrated with a black glass tile and a white standard tile with values of a* (−0.93) and b* (1.02). A green standard tile with values of a* (−25.30) and b* (13.71) was used to validate the calibration. IGC was calculated as: $IGC = \frac{-a}{b}$
Hue: Describes how close a color is to pure red, yellow, green, or blue (values of 0°, 90°, 180°, or 270°, respectively). Pod hue was calculated, as reported by Rayaprolu et al. (2015), using the equation:
$Hue = arc\ \ Tan\ \ (\frac{a}{b})$
HPW (Hundred-pod weight): Weight in grams of 100 pods prior to blanching.
EHQI: An index was developed to represent into a single trait the maximization values for IGC, HUE, and HPW. With this index, the greater the value, the greater the

TABLE 1 | Stepwise regression coefficients of cultivar response to Edamame Harvest Quality Index (*EHQI*) based on planting date, harvest date, and their squared values for three edamame cultivars subjected to three planting dates and up to eight harvest dates between stages R5.8 and R7 planted in two Arkansas locations over 2 years.

Cultivar	Intercept	*PlantingDate_DOY*	*PlantingDate_DOY²*	*HarvestDate_DOY*	*HarvestDate_DOY²*	_RMSE_	Adjusted_R²_
8080	−5.21578	−0.02442	0.00008	0.06083	−0.00013	0.04861	0.50740
R08-4002	−22.67342	–	0.00001	0.17232	−0.00033	0.05957	0.42270
R09-345	−9.80640	−0.00970	0.00003	0.08111	−0.00015	0.02823	0.34650

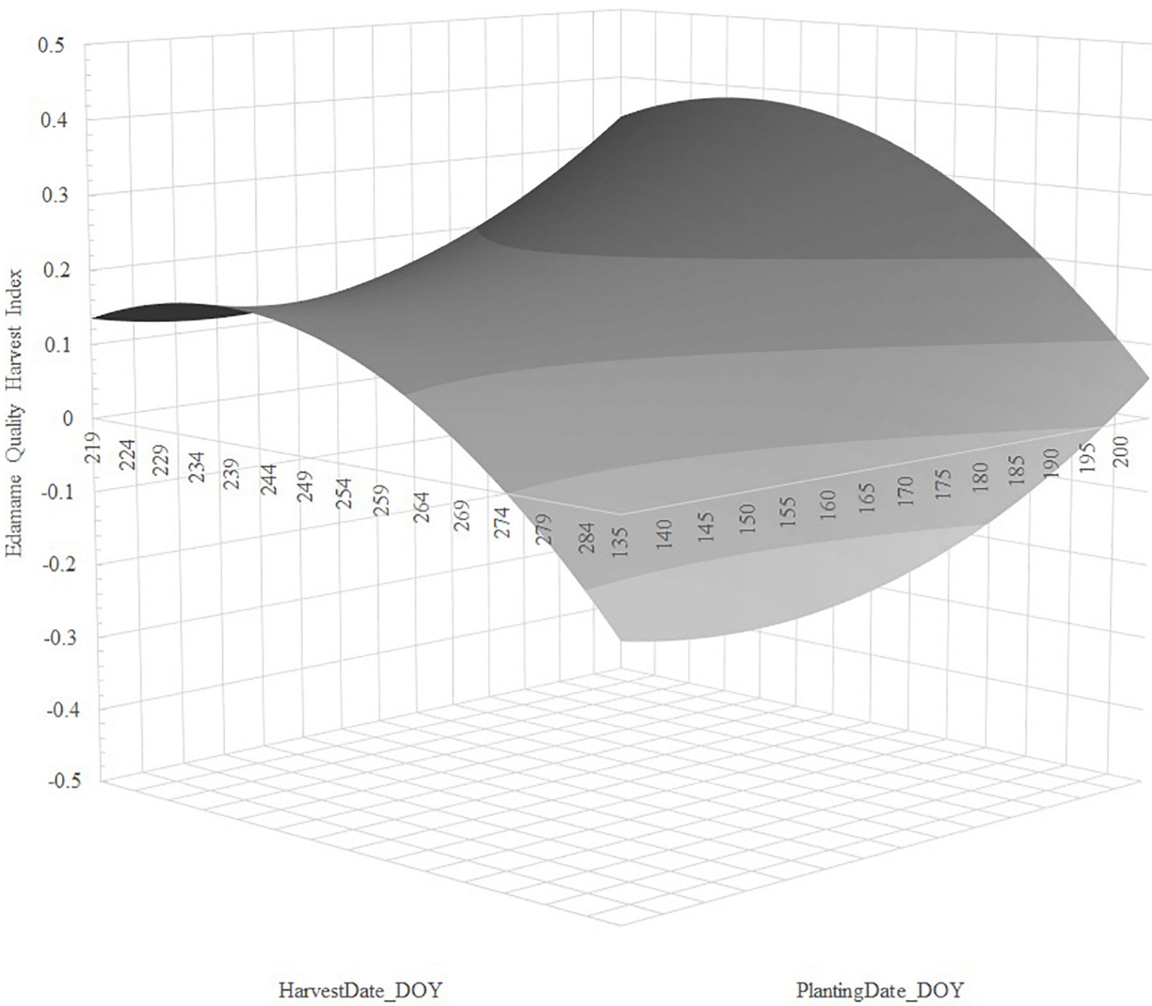

FIGURE 2 | Predicted response of Edamame Harvest Quality Index (*EHQI*) for cultivar 8080, as a function of the regression of *EHQI* on *PlantingDate_DOY*, *PlantingDate_DOY*^{2}, *HarvestDate_DOY*, and *HarvestDate_DOY*^{2}. Stepwise regression calculated from the dataset of edamame cultivar 8080 planted on two Arkansas locations over 2 years in three planting dates and subjected to up to eight harvests dates, approximately 5 days apart, between R5.8 and R7.

quality of the edamame pods. EHQI was calculated as:

$$EHQI = \frac{\left(\frac{HPW}{HPWmax^{\dagger}}\right)}{(120 - HUE) * (1 - IGC)}$$

where HPWmax† was calculated as the maximum hundred pod weight for a given cultivar, planting date, and location and year combination.

This index was built considering two key factors in edamame quality, namely, pod color, and pod size. As reported by Wibowo et al. (2020), #1 grade edamame (standard quality) is determined by the number of pods per 500 g (equivalent to our proposed *HPW* measurement), by the appearance of pods that are not too old and yellow, and by the pod color that must be uniformly green, among other factors. The other parameters in Wibowo et al. (2020) edamame grading system are either under heavy genetic control or under environmental effect but not necessarily affected by the crop developmental stage at harvest time (such as damage by pests and diseases or pod shape). Since

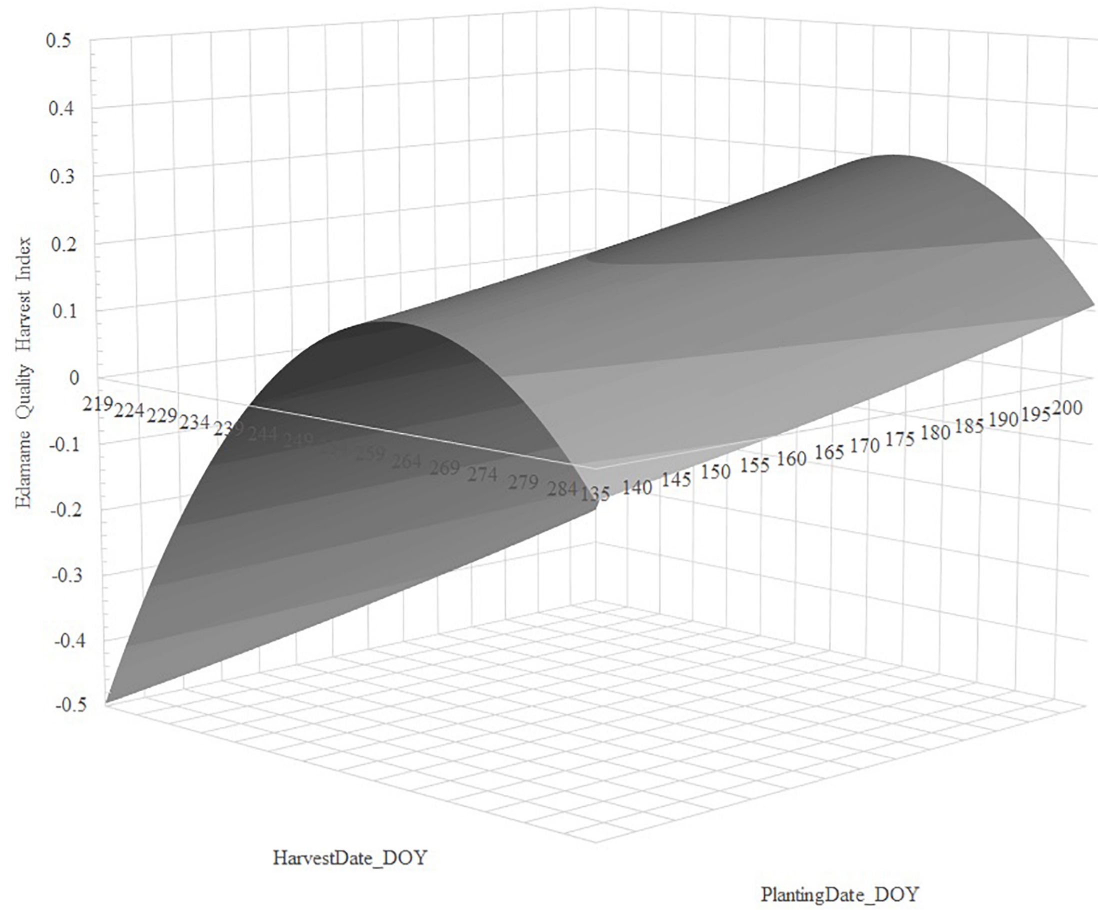

FIGURE 3 | Predicted response of Edamame Harvest Quality Index (*EHQI*) for cultivar R08-4002 as a function of the regression of *EHQI* on *PlantingDate_DOY*^{2}, *HarvestDate_DOY*, and *HarvestDate_DOY*^{2}. Stepwise regression calculated from the dataset of edamame cultivar R08-4002 planted on two Arkansas locations over 2 years in three planting dates and subjected to up to eight harvests dates, approximately 5 days apart, between R5.8 and R7. *PlantingDate_DOY* was not significant in the stepwise analysis and was not used to build this response surface.

seed size, and concomitantly *HPW*, is under genotypic control (Xu et al., 2016), in our index, we used a ratio of *HPW* to the maximum *HPW* observed for a given cultivar across all its harvest dates. Also, the denominator of *EHQI* includes a measurement of *HUE* and *IGC* as a means to counter the effect of pod size in the index; *HUE* and *IGC* are multiplicative and placed in the denominator of the index to highlight the greater importance of green color in overall edamame quality as demonstrated by the inclusion of two elements of color in edamame grading systems (Wibowo et al., 2020). *HUE* and *IGC* are highly correlated traits, and even though an alternative option would have been to build *EHQI* index using only one of the traits and weighed it using a square power, the authors decided to utilize both *IGC* and *HUE* in the original building of *EHQI*.

Statistical Analysis

Each Year and Location combination was aggregated into an "Environment" variable that was considered a random factor in all ANOVA, except when predicting the least-square means to be used for the stepwise and neural network analysis of weather variables, under which Environment had to be assumed a fixed effect.

Objective 1. Impact of Cultivar, Planting Date, and Harvest Date on Edamame Pod Color and Pod Weight

The PROC GLIMMIX procedure of SAS 9.4 software (SAS Institute, Cary, NC, United States) was used to analyze *HPW*, *IGC*, and *Hue*, with a model with the following fixed effects. Planting Date was the whole plot, Cultivar was the split-plot, and Harvest Date nested within Planting Date and Cultivar was the split-split plot. The random effects were Environment, Block nested within Environment, and Planting Date by Block nested within Environment. A beta distribution with logit link was used for *IGC* analysis, whereas a normal distribution with identity link was used for the models of *HPW* and *Hue*.

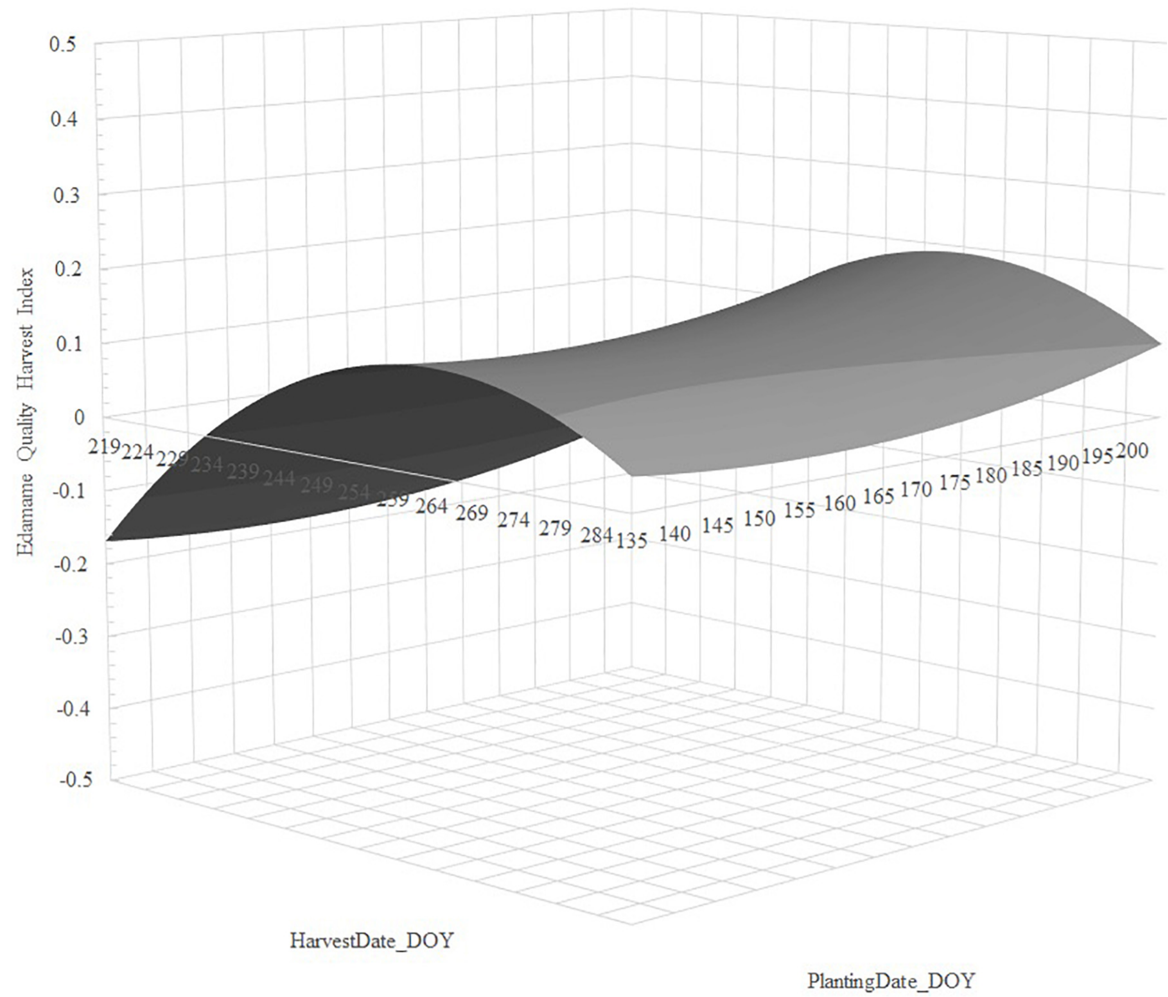

FIGURE 4 | Predicted response of Edamame Harvest Quality Index (*EHQI*) for cultivar R09-345 as a function of the regression of *EHQI* on *PlantingDate_DOY*, *PlantingDate_DOY*^{2}, *HarvestDate_DOY*, and *HarvestDate_DOY*^{2}. Stepwise regression calculated from dataset of edamame cultivar R09-345 planted on two Arkansas locations over 2 years in three planting dates and subjected to up to eight harvest dates, approximately 5 days apart, between R5.8 and R7.

Objective 2. Impact of Planting and Harvest Dates on EHQI of Three Soybean Cultivars

Since each cultivar is expected to have its own *HPWmax* because of the genetic control of seed size (Xu et al., 2016), ANOVA for *EHQI* was conducted by cultivar using a model in PROC GLIMMIX of SAS 9.4 with a beta distribution with logit link. For this analysis, Planting Date was whole plot and Harvest Date nested within Planting Date was split plot. The random effects were Environment, Block nested within Environment, and Planting Date by Block nested within Environment.

In addition, to characterize the change in edamame quality over time, regression analysis was conducted for *EHQI* as response of *PlantingDate_DOY*, *HarvestDate_DOY*, and their squared values. A stepwise regression was conducted independently for each cultivar using PROC REG in SAS 9.4, with significance level of 0.15 to enter or remove variables from the linear model and minimum Akaike information criterion (AIC) selection criteria. Modeled parameter estimates were then used to build response surfaces to predict *EHQI* for all days within the planting and harvesting day-of-year used.

Objective 3. Prediction of EHQI Based on Phenological Stages and Thermal Units Using Stepwise Regression and Artificial Neural Network Analysis

A PROC GLIMMIX procedure for *EHQI* by cultivar, and with Environment as fixed factor, was used to derive least-square means of *EHQI* by cultivar, environment, planting, and harvest date combinations. Harvest Date nested within Planting Date was the split plot in the analysis, and Block nested within Environment, and Planting Date by Block nested within Environment were random terms. The model was run assuming a beta distribution with logit link. Subsequently, a stepwise regression model in PROC REG in SAS 9.4 was used to predict *EHQI* by Cultivar, with *PlantingDate_DOY*, *PlantingDate_DOY*^{2}, *VeDate_DOY*, *R1Date_DOY*, *HarvestDate_DOY*, *HarvestDate_DOY*^{2}, *#Days Ve-R1*, *GDD Ve-R1*, *#Days R1-Harvest*, *GDD R1-Harvest*, *#Days Ve-Harvest*, and *GDD Ve-Harvest* as factors entering and leaving the model. Significance level 0.15 was used to enter or remove variables and estimate the linear model with lowest AIC.

Finally, because of a risk of collinearity and/or non-linear responses to some of the variables entering the stepwise procedure, a neural network analysis was conducted using JMP 15.1, executing a random holdback validation and testing multiple different hidden layer structures of TanH, Linear, or Gaussian activation types and two, three, or 10 first and secondary layers. Absolute penalty was implemented, as it was assumed that a few of the variables contribute more than others to the predictive model. The number of tours was set to 1,000 and random seed to 0.5. Factors included in the analysis were *PlantingDate_DOY*, *VeDate_DOY*, *R1Date_DOY*, *HarvestDate_DOY*, *#Days Ve-R1*, *GDD Ve-R1*, *#Days R1-Harvest*, *GDD R1-Harvest*, *#Days Ve-Harvest*, and *GDD Ve-Harvest*, and the response variable was *EHQI*.

RESULTS

Impact of Cultivar, Planting Date, and Harvest Date on Edamame Pod Color and Pod Weight

A split-split-plot analysis for edamame *HPW* indicated a non-significant Planting Date effect (0.2235) or Planting Date by Cultivar interaction (p = 0.2040) but significant Cultivar (p < 0.0001) and Harvest-Date-by-Planting-Date-by-Cultivar interactions (p < 0.0001). Similarly, for *Hue* and *IGC*, the Harvest-Date-by-Planting-Date-by-Cultivar interactions were highly significant (p < 0.0001). Those models also indicated significant main effects of planting date (p < 0.0001 and p = 0.0003 for *Hue* and *IGC*, respectively) and Cultivar (p = 0.0036 and p < 0.0001 for *Hue* and *IGC*, respectively) and a highly significant interaction with Planting Date-by-Cultivar (p < 0.0001 for *Hue* and p < 0.0001 for *IGC*). All these results indicated that the responses of *HPW*, *Hue*, and *IGC* must be explored independently by planting date, harvest date, and cultivar combination (**Figure 1**). **Supplementary Tables 1–3** present the least square mean estimates, standard error, and conservative T-grouping for *HPW*, *Hue*, and *IGC*, respectively. In general, it was observed that *HPW* increased over the first four harvest dates. The earliest-maturity cultivar (8080) presented a significant decrease in *HPW* for the second and third planting dates when harvesting extended past Harvest Date 6; such drop in *HPW* was not observed for the later maturity cultivars (**Figure 1** and **Supplementary Table 1**). On the other hand, *Hue* (**Figure 1** and **Supplementary Table 2**) and *IGC* (**Supplementary Table 3**) showed a decrease with soybean physiological development, and the maximum values were observed at R5.8, corresponding to the first Harvest Date. It is noteworthy that *Hue* and *IGC* are highly correlated traits (r = 0.99, p < 0.001), and they behaved similarly for all cultivars and planting dates. Future research efforts in edamame may focus on just *IGC* instead of measuring both traits because *IGC* is easier to interpret since the objective is to maximize *IGC* for a dark-green edamame pod product. Finally, Planting Date 2 resulted in the greatest number of weekly harvests possible between R5.8 and R7 for all cultivars, while delayed planting (Planting Date 3) resulted in a rapid reduction in green color (*Hue* and *IGC*) and the crop reached R7 faster for all three cultivars compared to the other planting dates (**Figure 1**).

Impact of Planting and Harvest Dates on Edamame Harvest Quality Index of Thee Soybean Cultivars

ANOVA of *EHQI* for the three soybean cultivars showed significant effects of Planting Date (p = 0.0003, p < 0.0001, and p = 0.0356 for 8080, R08-4002, and R09-345, respectively) and highly significant Harvest-Date-by-Planting-Date effects (all three cultivars had p < 0.0001). **Supplementary Table 4** reports the mean *EHQI* for the interaction of Harvest-Date-by-Planting-Date. Initial inspection of plots of *EHQI* over Harvest-Date-by-Planting-Date (**Figure 1** and **Supplementary Figures 1–3**) showed that the third planting date consistently resulted in the greatest *EHQI* on all three cultivars and that *EHQI* began to decay over harvest date for all planting dates and cultivars. In addition, *EHQI* was highly influenced by pod color, whereby *EHQI* only increased because of an increase in *HPW* if *Hue* (or *IGC*) was at a maximum level. For instance, for Cultivar 8080 and Planting Date 1, it can be observed how *EHQI* increased on the third and fourth Harvest Date as *HPW* increased while *Hue* remained relatively flat (**Figure 1**). However, when intensity of green pod color decreased, *EHQI* decreased concomitantly, regardless of a potential increase in *HPW*. Such situation can be observed for Cultivar 8080 in Planting Date 3, between Harvest Dates 3 and 4, where there was an increase in *HPW*, a decrease in *Hue*, and a concomitant decrease in *EHQI* (**Figure 1**).

Stepwise regression analysis retained linear and quadratic terms for *PlantingDate_DOY* and *HarvestDate_DOY* for cultivars 8080 and R09-345, but the linear *PlantingDate_DOY* was non-significant for R08-4002. We observed that regression of *EHQI* using *PlantingDate_DOY*, *HarvestDate_DOY*, *PlantingDate_DOY*^{2}, and *HarvestDate_DOY*^{2} fitted the data well, with adjusted R^2 values ranging from 34.7 to 50.7% and very low root mean square error (RMSE) (ranging from 0.028 to 0.059), indicating low standard deviations for the unexplained variance in the models (**Table 1**). Regression parameter estimates showed a quadratic increase of edamame quality with delayed planting; on the contrary, we observed a quadratic decrease of quality with delayed harvesting for all cultivars (**Table 1**).

For cultivar 8080, our data indicated delayed planting increased *EHQI*, and that *EHQI* decreased with delayed harvest; however, near the peak, *EHQI* remained fairly stable (within 0.024 units *EHQI*, or one standard error) for 27 days (**Figure 2**). For cultivar R08-4002, we also observed that delayed planting resulted in greater *EHQI* and that the quality decreased quadratic with delayed harvest; however, this cultivar did not retain *EHQI* well with delayed harvest, and *EHQI* remained within one standard error (0.026 *EHQI* units) from peak for 18 days (**Figure 3**). Finally, cultivar R09-345 showed the least total *EHQI*

TABLE 2 | Stepwise regression coefficients of cultivar response to Edamame Harvest Quality Index (*EHQI*) based on phenological and thermal functions for three edamame cultivars subjected to three planting dates and up to eight harvest dates between stages R5.8 and R7 planted in two Arkansas locations over 2 years.

Cultivar	Intercept	*Planting Date_ DOY*	*Ve Date_ DOY*	*Harvest Date_ DOY²*	*#Days Ve-R1*	*GDD Ve-R1*	*#Days R1-Harvest*	*GDD R1-Harvest*	*#Days Ve-Harvest*	_ RMSE_	Adjusted_ R^2_
8080	−0.51394	−0.01766	0.02647	−0.00001		−0.00078	−0.02018	0.00131		0.4454	0.5864
R08-4002	0.36379						−0.00765		0.00247	0.0611	0.3933
R09-345	0.11650				0.00205		−0.00171			0.0305	0.2390

from all cultivars, yet we still observed a quadratic improvement with delayed planting and a quadratic decrease of *EHQI* with delayed harvest (**Figure 4**). R09-345 had a harvest window around the peak of EHQI that spanned for approximately 20 days, where *EHQI* was within one standard error (or 0.015 *EHQI* units) from peak *EHQI*.

TABLE 3 | Neural Network Analysis model with NTanH(10) summary of training and validation model and variable importance assuming dependent resampled inputs for the prediction of response variable Edamame Harvest Quality Index (*EHQI*) based on phenological (Ve, emergence; R1, first flower) day-of-year (DOY), and thermal functions (GDD, growing degree days) for soybean cultivar 8080.

Model summary	Training	Validation
R^2	0.842	0.735
RMSE	0.029	0.032
Variable importance	**Main effect**	**Total effect**
#Days R1-Harvest	0.119	0.503
#Days Ve-Harvest	0.162	0.162
GDD Ve-Harvest	0.106	0.114
HarvestDate_DOY	0.109	0.109
GDD Ve-R1	0.102	0.102
GDD R1-Harvest	0.084	0.084
PlantingDate_DOY	0.082	0.082
VeDate_DOY	0.082	0.082
R1Date_DOY	0.081	0.081
#Days Ve-R1	0.073	0.073

Prediction of Edamame Harvest Quality Index Based on Phenological Stages and Thermal Units Using Stepwise Regression and Artificial Neural Network Analysis

A stepwise regression and an Artificial Neural Network Analysis were conducted for *EHQI* using variables that included days to various phenological stages and thermal units accumulated to key phenological stages. Stepwise regression models found *#Days R1-Harvest* to be significant across all three cultivars (**Table 2**), with a negative impact on quality, as the longer period between initiation of flowering and harvest resulted in lower quality. Other significant terms in the model included *#Days Ve-R1* for R09-345 and *#Days Ve-Harvest* for R08-4002. Interestingly enough, the only cultivar that responded to planting date and thermal units was 8080, which is an indeterminate, early cultivar (MG3) and had significant *PlantingDate_DOY*, *VeDate_DOY*, *HarvestDate_DOY*^{2}, *GDD Ve-R1*, and *GDD R1-Harvest* (**Table 2**).

TABLE 4 | Neural Network Analysis model with NTanH(10) summary of training and validation model and variable importance assuming dependent resampled inputs for the prediction of response variable Edamame Harvest Quality Index (*EHQI*) based on phenological (Ve, emergence; R1, first flower) day-of-year (DOY), and thermal functions (GDD, growing degree days) for soybean cultivar R08-4002.

Model summary	Training	Validation
R^2	0.703	0.591
RMSE	0.046	0.039
Variable importance	**Main effect**	**Total effect**
#Days R1-Harvest	0.171	0.314
GDD R1-Harvest	0.122	0.122
R1Date_DOY	0.108	0.108
VeDate_DOY	0.104	0.104
PlantingDate_DOY	0.103	0.103
GDD Ve-Harvest	0.098	0.098
GDD Ve-R1	0.080	0.080
HarvestDate_DOY	0.073	0.073
#Days Ve-Harvest	0.074	0.074
#Days Ve-R1	0.066	0.066

TABLE 5 | Neural Network Analysis model with NTanH(10) summary of training and validation model and variable importance assuming dependent resampled inputs for the prediction of response variable Edamame Harvest Quality Index (*EHQI*) based on phenological (Ve, emergence; R1, first flower) day-of-year (DOY), and thermal functions (GDD, growing degree days) for soybean cultivar R09-345.

Model summary	Training	Validation
R^2	0.668	0.783
RMSE	0.017	0.020
Variable importance	**Main effect**	**Total effect**
#Days Ve-R1	0.069	0.201
HarvestDate_DOY	0.155	0.155
#Days Ve-Harvest	0.135	0.135
#Days R1-Harvest	0.132	0.132
GDD Ve-Harvest	0.082	0.109
GDD R1-Harvest	0.103	0.103
R1Date_DOY	0.100	0.100
PlantingDate_DOY	0.082	0.082
VeDate_DOY	0.078	0.078
GDD Ve-R1	0.064	0.064

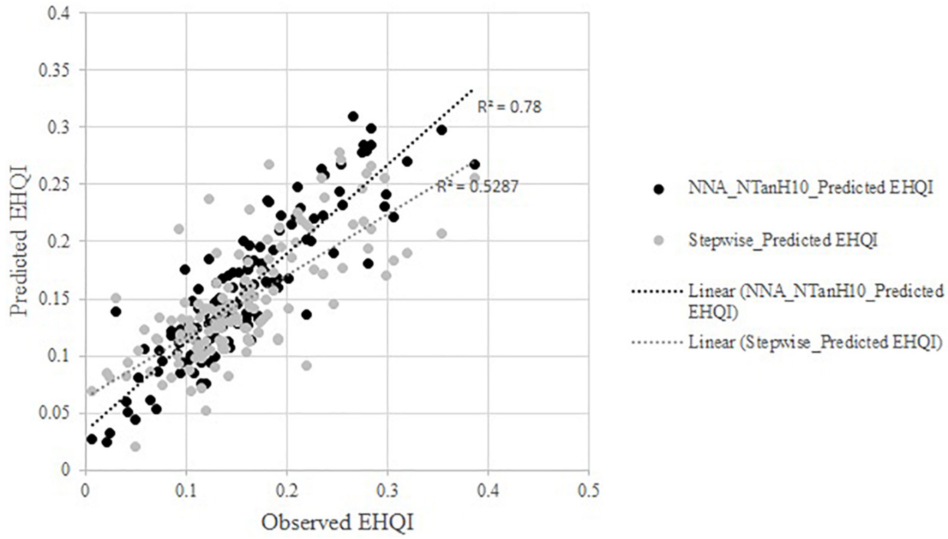

FIGURE 5 | Plot of predicted Edamame Harvest Quality Index (*EHQI*) by Neural Network Analysis and stepwise analysis as a function of the observed *EHQI*. Neural network model with NTanH(10) assuming dependent resampled inputs. Both models utilized phenological and thermal functions for the prediction of *EHQI*.

A Neural Network model with NTanH parameter of 10 nodes resulted in the lowest RMSE for all three cultivars as compared to models with multiple combinations of TanH, Linear, or Gaussian activation types and two, three, or 10 first and secondary layers (data not shown). For cultivar 8080, we observed that *#Days R1-Harvest* was the most important variable in predicting *EHQI*, with a total effect of 0.503 that was three times larger than the next variable in total effect. Also, for cultivar 8080, the variables with main effect greater than 0.100 included *#Days Ve-Harvest*, *#Days R1-Harvest*, *GDD Ve-Harvest*, *HarvestDate_DOY*, and *GDD Ve-R1* (**Table 3**). For soybean cultivar R08-4002, we observed that #Days R1-Harvest also had the largest contribution to the predictive model for *EHQI* (total effect 0.314). Additionally, for R08-4002, we observed that the variables with main effects greater than 0.100 included *#Days R1-Harvest*, *GDD R1-Harvest*, *R1Date_DOY*, *VeDate_DOY*, and *PlantingDate_DOY* (**Table 4**). Lastly, for the prediction of *EHQI* of R09-345, we observed that *#Days Ve-R1* had the largest total effect (0.201) and that the following variables each had main effects greater than 0.100: *HarvestDate_DOY*, *#Days Ve-Harvest*, *#Days R1-Harvest*, *GDD R1-Harvest*, and *R1Date_DOY* (**Table 5**).

DISCUSSION

High edamame quality is characterized by large pod size and intense and uniform green pod color (Wibowo et al., 2020). Panthee et al. (2004) suggested that seed size has high heritability, indicating that the trait should be controlled more by genetic than environmental variances. A study by Beatty et al. (1982) found that seed weight was not significantly different from April 15 to May 15 planting date but dropped significantly each month from a May 15 to July 15 planting date. Similarly, in our study, we found that *HPW* depended on the interaction between planting date and harvest date and that *HPW* increased during the first four Harvest Dates for all cultivars and planting dates.

The second component of edamame quality is intensity of pod green color. Of the cultivars chosen for our study, R09-345 has green seeds at R6 but develops black seed coat color at maturity. Such cultivar showed significantly lower pod *Hue* and *IGC* and low overall *EHQI* than the other two cultivars (**Figure 4** and **Supplementary Tables 2, 3**), suggesting that cultivars whose seed turn black or brown at maturity will not have the same pod color at R6 compared to cultivars that either stay green or turn yellow at the R8 reproductive stage. Thus, cultivars with dark seed coat at maturity may not be appropriate for fresh/frozen market edamame production.

We observed that, in general, delayed planting maximized *EHQI*. Mean separations for *EHQI* analyzed as a split-plot design (**Supplementary Table 4**) showed that the first harvest dates were usually not statistically different from each other, but that as harvest was delayed, *EHQI* dropped. We also observed a quadratic decrease of *EHQI* with delayed harvest in all cultivars from our regression analysis. The harvest window seemed planting date and cultivar dependent. The window for maximum *EHQI* was shorter in late plantings than in earlier planting dates (**Supplementary Figures 1–3**). Near the cultivar optimum, harvest window for *EHQI* ranged from 18 to 27 days. The low end of the spectrum agrees with the 18-day window reported by Nolen et al. (2016), but cultivar 8080 showed a much larger window where quality did not drop. Therefore, edamame companies aiming for high quality must procure late planting and ensure logistics are in place for earlier and timely harvests 10–15 days after R5.8 stage is observed. This must be achieved by carefully planning the logistics of field equipment availability, field access under unfavorable weather/road conditions, and processing house turnaround times. On the contrary, early plantings maintain quality for longer periods of time, albeit not maximizing *EHQI*; therefore, edamame-growing companies could target earlier plantings to marginal grounds, farms with difficult access, or periods where the processing plant is at its peak, all while managing the maturity of the cultivar to spread out flowering (and harvest) timelines.

The early maturing indeterminate cultivar 8080 showed a significant response to thermal units (*GDD Ve-R1* and *GDD R1-Harvest*) in stepwise regression analysis, while the late-maturity cultivars did not. This agrees with expected soybean response where the temperature is considered a modifier to the effect of photoperiod whereby short days enhance reproductive development rate (Setiyono et al., 2007) and with the expected insensitivity of earlier maturity to photoperiod (Salmeron et al., 2014), thus enhancing the opportunity for temperature responses.

Both stepwise regression and Neural Network Analysis identified *#Days R1-Harvest* as a key variable affecting *EHQI* for all three soybean cultivars. Additionally, when looking at the prediction values for *EHQI* based on Neural Network and stepwise analyses (**Figure 5**), we observed that formulas from stepwise analysis tended to overpredict lower and underpredict higher *EHQI* values, whereas Neural Network prediction was more consistent over the range of data. However, the simplicity of stepwise predictions involving a simple model with few parameters for cultivars of maturities adapted to Arkansas could make field assessments easier. Future research is needed to test these prediction models on other field-grown edamame cultivars to explore their applicability in forecasting harvest decisions.

Finally, and even though our research did not study pod or seed yield, it is a very important criterion for edamame farmers as they need to balance yield and quality of their end product. Therefore, agronomic practices must be used to balance increased seed and pod yield resulting from earlier planting dates (De Bruin and Pedersen, 2008; Mourtzinis et al., 2017) and enhanced seed and pod quality observed from later planting dates.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

DM was responsible for the investigation, methodology, and writing of the original draft. LM was responsible for supervision, data curation, data analysis, and editing and reviewing the final document. MS, MO, LF-P, AA, and CW were responsible for the investigation and editing and reviewing the final document. PC acquired funding and conceptualized the project and edited and reviewed the final document. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors thank the Arkansas Soybean Promotion Board for its sponsorship of this project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.585856/full#supplementary-material

Supplementary Figure 1 | Edamame Harvest Quality Index (*EHQI*) as function of harvest date treatment by planting date treatment for soybean cultivar 8080. Error bars represent standard errors of least-square means for *EHQI*.

Supplementary Figure 2 | Edamame Harvest Quality Index (*EHQI*) as function of harvest date by planting date for soybean breeding cultivar R08-4002. Error bars represent standard errors of least-square means for *EHQI*.

Supplementary Figure 3 | Edamame Harvest Quality Index (*EHQI*) as function of harvest date by planting date for soybean breeding cultivar R09-345. Error bars represent standard errors of least-square means for *EHQI*.

Supplementary Table 1 | Least-square means, Standard Error, and Conservative T-grouping of Hundred Pod Weight (*HPW*) per cultivar, planting and harvest date combination analyzed on a split-split block design with block and environment as random factors. Levels not connected by same letter are significantly different at $\alpha = 0.05$.

Supplementary Table 2 | Least-square means, Standard Error, and Conservative T-grouping of *Hue* per cultivar, planting and harvest date combination analyzed on a split-split block design with block and environment as random factors. Levels not connected by same letter are significantly different at $\alpha = 0.05$.

Supplementary Table 3 | Least-square means, Standard Error, and Conservative T-grouping of Intensity of Green Color (*IGC*) per cultivar, planting and harvest date combination analyzed on a split-split block design with block and environment as random factors. Levels not connected by same letter are significantly different at $\alpha = 0.05$.

Supplementary Table 4 | Least-square mean, Standard Error, and Conservative T-grouping of Edamame Harvest Quality Index (*EHQI*) by cultivar, per planting and harvest date combination, analyzed as a split block design with block and environment as random factors. Levels not connected by same letter are significantly different at $\alpha = 0.05$. **(A)** R08-4002, **(B)** 8080, and **(C)** R09-345.

REFERENCES

Beatty, K. D., Eldridge, I. L., and Simpson, A. M. Jr. (1982). Soybean response to different planting patterns and dates. *Agron. J.* 74, 859–862. doi: 10.2134/agronj1982.00021962007400050021x

Boote, K. J., Jones, J. W., and Hoogenboom, G. (1998). "Simulation of crop growth: CROPGRO model," in *Agricultural Systems Modeling and Simulation*, eds R. M. Peart, and R. B. Curry, (New York, NY: Marcel Dekker), 651–692. doi: 10.1201/9781482269765-18

De Bruin, J., and Pedersen, P. (2008). Soybean seed yield response to planting date and seeding rate in the upper Midwest. *Agron. J.* 100, 696–703. doi: 10.2134/agronj2007.0115

Fehr, W. R., Caviness, C. E., Burmood, D. T., and Pennington, J. S. (1971). Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. *Crop Sci.* 11, 929–931. doi: 10.2135/cropsci1971.0011183x001100060051x

Garner, W. W., and Allard, H. A. (1920). Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. *J. Agric. Res.* 18, 553–606.

Johnson, H. W., Borthwick, H. A., and Leffel, R. C. (1960). Effects of photoperiod and time of planting on rates of development of the soybean in various stages of the life cycle. *Bot. Gaz.* 122, 77–95. doi: 10.1086/336090

Konovsky, J., Lumpkin, T. A., and McClary, D. (1994). "Edamame: the vegetable soybean," in *Understanding the Japanese Food and Agrimarket: A Multifaceted Opportunity*, ed. A. D. O'Rourke, (Binghamton, NY: Haworth Press), 173–181. doi: 10.1201/9781003075172-15

Major, D. J., Johnson, D. R., and Luedders, V. D. (1975a). Evaluation of eleven thermal unit methods for predicting soybean development. *Crop Sci.* 15, 172–174. doi: 10.2135/cropsci1975.0011183x001500020008x

Major, D. J., Johnson, D. R., Tanner, J. W., and Anderson, I. C. (1975b). Effects of day length and temperature on soybean development. *Crop Sci.* 15, 174–179.

Mentreddy, S. R., Mohamed, A. I., Joshee, N., and Yadav, A. K. (2002). "Edamame: a nutritious vegetable crop," in *Trends in New Crops and New Uses*, eds J. Janick, and A. Whipkey, (Alexandria, VA: ASHS Press), 432–438.

Miller, P., Lanier, W., and Brandt, S. (2001). *Using Growing Degree Days to Predict Plant Stages*. Bozeman, MT: Montana State University Extension Service.

Mourtzinis, S., Gaspar, A., Naeva, S., and Conley, S. (2017). Planting date, maturity, and temperature effects on soybean seed yield and composition. *Agron. J.* 109, 2040–2049. doi: 10.2134/agronj2017.05.0247

Mozzoni, L. A., Morawicki, R. O., and Chen, P. (2009). Canning of vegetable soybean: procedures and quality evaluations. *Int. J. Food Sci. Technol.* 44, 1125–1130. doi: 10.1111/j.1365-2621.2009.01929.x

Nolen, S., Zhang, B., and Kering, M. K. (2016). Increasing fresh edamame bean supply through season extension techniques. *J. Hortic.* 3, 170. doi: 10.4172/2376-0354.1000170

Nuss, J. (2013). *US Soybean Farmers See Growth Potential in Edamame. CNSnews.com*. Available online at: http://www.cnsnews.com/news/article/us-soybean-farmers-see-growth-potential-edamame (accessed January 11, 2017).

Ogles, C. Z., Guertral, E. A., and Weaver, D. B. (2016). Edamame cultivar evaluation in central Alabama. *Agron. J.* 108, 2371–2378. doi: 10.2134/agronj2016.04.0218

Oliver, F. C., and Annandale, J. G. (1998). Thermal time requirements for the development of green pea (*Pisum sativum* L.). *Field Crops Res.* 56, 301–307. doi: 10.1016/s0378-4290(97)00097-x

Panthee, D. R., Pantalone, V. R., West, D. R., Saxton, A. M., and Sams, C. E. (2004). Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. *Crop Sci.* 45, 2015–2022. doi: 10.2135/cropsci2004.0720

Polson, D. E. (1972). Day-neutrality in soybeans. *Crop Sci.* 12, 773–776. doi: 10.2135/cropsci1972.0011183x001200060017x

Purcell, L. C., Salmeron, M., and Ashlock, L. (2014). "Soybean growth and development," in *Arkansas Soybean Handbook, MP 197*, (Little Rock, AR: University of Arkansas Cooperative Extension Service), 1–8.

Rayaprolu, S., Hettiarachchy, N., Aldoury, M., Cho, S., Moseley, D., and Chen, P. (2015). Physical and textural attributes of freeze-dried genetically modified and non-genetically modified soy beans. *J. Food Nutr. Sci.* 3, 119–125. doi: 10.11648/j.jfns.20150303.17

Ross, J. (2013). *Edamame Production Practices: What We Know Now*. Fayetteville, AR: University of Arkansas.

Saldivar, X., Wang, Y.-J., Chen, P., and Hou, A. (2011). Changes in chemical composition during soybean seed development. *Food Chem.* 124, 1369–1375. doi: 10.1016/j.foodchem.2010.07.091

Salmeron, M., Gbur, E. E., Bourland, F. M., Buehring, N. W., Earnest, L., Fritschi, F. B., et al. (2014). Soybean maturity group choices for early and late plantings in the midsouth. *Agron J.* 106, 1893–1901. doi: 10.2134/agronj14.0222

Setiyono, T. D., Weiss, A., Specht, J., Bastidas, A. M., Cassman, K. G., and Dobermann, A. (2007). Understanding and modeling the effect of temperature and day length on soybean phenology under high-yield conditions. *Field Crops Res.* 100, 257–271. doi: 10.1016/j.fcr.2006.07.011

Shanmugasundaram, S., and Yan, M. R. (2004). "Global expansion of high value vegetable soybean," in *Proceedings of the VII World Soybean Conference/IV International Soybean Processing and Utilization Conference/III Brazilian Soybean Congress, 29 Feb–5 Mar 2004*, (Foz do Iguassu: Brazilian Agricultural Research Corporation, National Soybean Research Center), 915–920.

Soil Survey Staff, (2017). *Natural Resources Conservation Service, United States Department of Agriculture, Web Soil Survey*. Available online at: https://websoilsurvey.sc.egov.usda.gov/ (accessed November 12, 2017).

Tsay, L., and Sheu, S. (1991). "Studies on the effects of cold storage and precooling on the quality of vegetable soybean," in *Vegetable Soybean: Research Needs for Production and Quality Improvement. Proceedings of the Workshop, Kenting, Taiwan*, ed. S. Shanmungasundaram, (Taipei: Asian Vegetable Research and Development Center).

Wibowo, Y., Mahardika, N. S., Afrizal, R., and Udayana, I. G. B. (2020). Design of the expert system for edamame grading using forward chaining method. *Sustain. Environ. Agric. Sci.* 4, 26–37. doi: 10.22225/seas.4.1.1686.26-37

Xu, Y., Cartier, A., Kibet, D., Jordan, K., Hakala, I., Davis, S., et al. (2016). Physical and nutritional properties of edamame sedes as influenced by stage of development. *Food Meas.* 10, 193–200. doi: 10.1007/s11694-015-9293-9

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 03 March 2021
doi: 10.3389/fpls.2021.614926

Check for updates

Analysis of Shoot Architecture Traits in Edamame Reveals Potential Strategies to Improve Harvest Efficiency

Kshitiz Dhakal[1], Qian Zhu[1], Bo Zhang[1], Mao Li[2]* and Song Li[1]*

[1] School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, United States, [2] Donald Danforth Plant Science Center, St. Louis, MO, United States

OPEN ACCESS

Edited by:
Claudio Bonghi,
University of Padua, Italy

Reviewed by:
Lin Xi,
University of Hohenheim, Germany
Agnieszka Klimek-Kopyra,
University of Agriculture in Krakow, Poland
Víctor Arturo González-Hernández,
Colegio de Postgraduados (COLPOS), Mexico

***Correspondence:**
Mao Li
mli@danforthcenter.org
Song Li
songli@vt.edu

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: *07 October 2020*
Accepted: *01 February 2021*
Published: *03 March 2021*

Citation:
Dhakal K, Zhu Q, Zhang B, Li M and Li S (2021) Analysis of Shoot Architecture Traits in Edamame Reveals Potential Strategies to Improve Harvest Efficiency. Front. Plant Sci. 12:614926. doi: 10.3389/fpls.2021.614926

Edamame is a type of green, vegetable soybean and improving shoot architecture traits for edamame is important for breeding of high-yield varieties by decreasing potential loss due to harvesting. In this study, we use digital imaging technology and computer vision algorithms to characterize major traits of shoot architecture for edamame. Using a population of edamame PIs, we seek to identify underlying genetic control of different shoot architecture traits. We found significant variations in the shoot architecture of the edamame lines including long-skinny and candle stick-like structures. To quantify the similarity and differences of branching patterns between these edamame varieties, we applied a topological measurement called persistent homology. Persistent homology uses algebraic geometry algorithms to measure the structural similarities between complex shapes. We found intriguing relationships between the topological features of branching networks and pod numbers in our plant population, suggesting combination of multiple topological features contribute to the overall pod numbers on a plant. We also identified potential candidate genes including a lateral organ boundary gene family protein and a MADS-box gene that are associated with the pod numbers. This research provides insight into the genetic regulation of shoot architecture traits and can be used to further develop edamame varieties that are better adapted to mechanical harvesting.

Keywords: phenotyping, shoot architecture, edamame, breeding, persistent homology

INTRODUCTION

Edamame is a type of green, vegetable soybean which has become a popular food ingredient in many countries because it is a nutritious food source of protein, isoflavones, and vitamins (Mentreddy et al., 2002; Lee et al., 2018; Mahoussi et al., 2020). Edamame has been cultivated in east Asian countries for more than 2,000 years and documented edamame varieties have been mainly

Abbreviations: APBL, average primary branch length; FIL, first internode length; FNH, first node height; FPH, first pod height; MBL, main branch length; MDS1, multidimensional scaling 1; MDS2, multidimensional scaling 2; MDS3, multidimensional scaling 3; NPB, number of primary branch; PH, plant height; PN, pod number; PN10, pod height 10 cm above ground; SIL, Second internode length; SIN, short internode; TBL, total branch length; TIL, third internode length.

originated from this area (William and Aoyagi, 2009). In recent years, production and breeding of locally adapted edamame varieties have been reported in North and South America, Europe, and Africa (Konovsky et al., 1994). The yield components of soybeans have been studied and include plant density, number of pods and number of seeds per pod and seed size (Liu et al., 2010a; Ulloa et al., 2010). However, little is known about how these components affect edamame yield, because the yield is evaluated when the seeds are at an immature stage.

There are several major differences between edamame and grain soybeans. First, edamame is harvested when the pods are fully filled while beans are still green with high level of moisture and sugar content (Shanmugasundaram et al., 1991). In contrast, grain soybeans for feed and oil are typically harvested when the pods and beans are dry. Second, due to consumer preference, edamame seeds are much larger than grain soybean seeds (Carson, 2010). Because of these key differences, grain soybean varieties cannot be directly used for edamame production and optimization of additional traits are needed to produce new edamame varieties that are better accepted by the producers and the consumers. In the United States, despite being a major producer of grain soybeans, most frozen edamame products have been imported from Asia. The main obstacles for commercial production of edamame in the United States are the efficiency of mechanical harvest and the cost of hand harvesting where manual harvesting is still a common practice for small farmer (Tadesse and Chris, 2007; Lord et al., 2019).

A number of studies have been performed to test commercial harvesters on edamame. For example, a common bean harvester, Oxbo BH100 was tested to harvest edamame and the results were compared to hand harvesting (Tadesse and Chris, 2007). It was found that hand harvesting generated twice as much pods as compared to mechanical harvest. However, mechanical harvest has generated cleaner products. Mechanical harvest was found to give best results for plant with 55–66 cm in height. Harvest efficiency of the same type of harvester was tested on three edamame varieties and the harvest efficiency is between 54 and 85% (Zandonadi et al., 2010). The speed of harvester does not affect the harvest efficiency when it was below 2 miles per hour. In a more recent study, four cultivars of edamame were used to studied the optimal plant density of edamame (Dhaliwal and Williams, 2020) and these varieties were harvested by the same bean harvester. This research showed, with higher plant density, the number of branches and pod mass/vegetative mass ratio decrease whereas height and leaf area index increase for all varieties tested. In particular, for the same variety, the main stem branch changes from 6 to 1 with increasing plant density. Using the machine harvester, it was found that 86–95% marketable pods can be harvested mechanically Dhaliwal and Williams (2020).

A number of environmental factors are known to affect pod numbers and plant architecture in soybeans and edamame. In a comparison of determinant and indeterminant varieties (Egli and Bruening, 2006), it was found that 85% of pods were initiated before stage R5. R5 stage is one of the reproductive stages of the edamame when seeds begin to develop (Fehr and Caviness, 1977). At this stage, seed is 3 mm in size, which develops inside a pod at one of the four uppermost nodes on the main stem with a fully developed leaf. Indeterminant varieties have longer pod production period for approximately 50 days. In a test of maturity of soybeans, late mature groups have more nods and more pods per plant as compared to early mature groups (Zhang and Kyei-Boahen, 2007). Photoperiod is a major factor that affect number of pods in soybeans where long photoperiod mainly affects pod number during the R3-R6 stage (Kantolic and Slafer, 2007). Long day also delays flower to pod transition and seed filling, but it does not affect pod elongation (Nico et al., 2016). In addition to photoperiod, higher temperature can also contribute to higher number of flowers and pods, but these flowers may fail to produce mature pods and cause a reduction of yield (Kim et al., 2020). A multi-year study of edamame breeding lines show that there are significant trait variations between years, including changes in pod yield and plant height, suggesting environmental variation is an important factor for edamame development (Jiang et al., 2018).

At molecular and genetic level, many key genes in soybeans related to the shoot architecture traits have been identified. Soybeans can be categorized into three types of stem growth habits: determinate, indeterminate and semi-determinate growth. Two genes, Dt1 and Dt2, are known to regulate this process in soybean (Liu et al., 2010b; Tian et al., 2010; Ping et al., 2014; Zhang et al., 2019). GmDt1 is a homolog to Arabidopsis terminal flower 1 (TFL1) and GmDt1 in cultivated soybeans confers determinate growth habit. GmDt2 is a MADS-box transcription factor which represses GmDt1 expression and confers semi-determinate growth (Ping et al., 2014). In addition to transcription factors, microRNAs, in particular, gmmiR156b has been shown to regulate soybean shoot architecture. Over expression of this microRNA lead to a 100% increase of branches without changing plant height. Pod per plant is also increased more than 30% without affecting seed protein and oil content (Sun et al., 2019). Using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), double mutant of GmSPL9a/b, the target gene of gmmiR156b, showed similar phenotype of increased branch numbers as found in gmmiR156b over expression lines (Bao et al., 2019). Besides these well characterized molecular pathways that regulate soybean shoot architecture, genome wide association studies have also identified many Genome-Wide Association Studies (GWAS), Quantitative Trait Locus (QTLs) that are associated with shoot architecture traits such as plant heights, branch numbers and pod numbers (Hao et al., 2012; Zhang H. et al., 2015; Fang et al., 2017). Using pod number as one example, these published studies have identified 26 QTL loci and 9 candidate genes that are associated with pod number variations in soybeans.

The soybean research and breeding community have accumulated large amount of genomics resources including more than 20,000 plant introductions (PIs) that were genotyped by a 50K SNP array (Song et al., 2013) and over 3,000 PIs with full genome sequences available (Liu et al., 2020). To leverage this genetic diversity and genomic resources in edamame breeding to improve harvest efficiency, the major challenge lies in phenotyping. High throughput phenotyping in soybean have been used to study leaf shape (Chen and Nelson, 2004), root architecture (Fenta et al., 2014), and canopy cover (Xavier et al., 2017). In this work, we develop a phenotyping pipeline to collect images for edamame at harvest stage of R6 to R7 and to quantify major shoot architecture traits related to harvest efficiency including plant height, branching patterns, pod numbers and pod locations. We also collected canopy cover data over the growth season to quantify and correlation of canopy cover with other shoot architecture traits. We applied a topological approach called persistent homology (Li et al., 2017, 2019) to quantify the shoot architecture in topological space. Using a mini-core collection of edamame, we explore the correlation between geometric traits and topological traits and test whether known markers for pod numbers are associated with the traits observed in our data. Our results provide a scalable pipeline of shoot architecture phenotyping and provides novel candidate markers and genes for improving shoot architecture traits in edamame.

MATERIALS AND METHODS

Plant Materials and Shoot Image Collection

A total of 151 soybean PIs with > 20 g/100 seeds that are potential parental lines for developing edamame varieties (referred as edamame PIs) were sown in 3 m row and 0.75 m row spacing (with a seeding rate of ∼70,000 plants per hectare) arranged in a complete randomized block design with two to four replications in Kentland farm at Blacksburg, VA in 2019. We selected these 151 PIs in our collection and two to four replicates (plants) per PI (540 plants) were harvested by cutting them from the soil line using a bypass looper (large secateur). The leaves and petioles were taken off of the plants before they were taken to the imaging station. The imaging station consisted of a black background, inch tapes at the borders, a camera tripod, and a digital single-lens reflex (DSLR) camera. The entry names and sample numbers of the plants were printed as a barcode on an iPad and captured by the camera. Images were captured from both sides of the plants. We have generated 1,202 images for all plants that were harvested. Based on a preliminary analysis of all images, we selected 178 images from 24 edamame PIs for this study because these images showed diverse phenotypic traits such as plant heights and branching patterns and all 24 varieties have been genotyped using the 50 K SNP array. These 24 PIs showed diverse heights from dwarf to tall. The branching pattern was diverse ranging from one branch to several branches and the shape was also varying from candle shaped to one straight branch. A list of 24 selected PIs and traits measured in our analysis is provided as **Supplementary Table 1**.

Drone Image Collection and Analysis

A DJI Phantom 4-Advanced was used for the canopy cover study during the 2019 growth season from May 2019 to September 2019. A total of 1,853 images were collected during this growth season with an average of 120 drone images were collected for each flight day. Drone flight waypoints were generated using an iPad app, DroneDeploy. Flight height was set to 30.5 m (∼100 ft) above ground level. Side overlap and front overlap were set at 75% with padding. Ground control points (GCPs) were used and the precise GPS locations of GCPs were determined using a Real-time kinematic GPS. Orthorectified drone maps were generated using AgiSoft Metashape professional edition (Version 1.6) and subplot extraction were done manually using ArcGIS pro software. Canopy cover was extracted and averaged across replicates to generate a growth curve for each variety and the results were compared and correlated with other shoot architecture traits.

Image Analysis and Characterization of Topological and Geometric Traits

Two to eight images from each PI were used for image annotation using ImageJ (Rueden et al., 2017) and ImgLabel[1] software. For each plant two images were taken from both side of the plant. Because of the bilateral symmetry of edamame plants, for each PI, a plant is placed on a black background with branches laying on a flat surface to take a first photo and the plant is flipped to take another photo (**Supplementary Figure 1A**). These images were analyzed using ImageJ program with a custom plugin script to label all the branches (**Supplementary Figure 1B**). The branches were then analyzed using script developed in Matlab to convert the labeled images into a network of branches with vertices representing the locations of landmarks used in the labeling process. The main branch and side branches were labeled separately which allows post processing to calculate the branch length separately and to identify internodes in the branch networks. The correlation between geometric parameters measured in the photos was tested using cor.test function in R to test for Pearson's product moment correlation coefficient based on fisher's Z transformation. Pods in each image were also labeled manually using ImgLabel, which generated an Extensible Markup Language (XML) file for each labeled image and the XML file contains all the x and y coordinates for the labeled pod locations (**Supplementary Figure 1C**). The top and bottom of each plant were also labeled using ImgLabel program. A python script was developed to process the XML files to extract traits including pod numbers, pod locations, plant height and pixel per centi meter from the images. Each plant was imaged and labeled twice and the results were averaged for the final analysis. The primary branches and the main branch for each plant were also detected using Matlab script and manual curation (**Supplementary Figure 1D**). Each primary branch was represented by a path (a sequence of edges which join a sequence of vertices). The primary branch length was calculated by adding up the length of all the edges of this path. Density plots were generated using plot density function from ggplot2 package in R.

To calculate the topological similarities between different branching networks, we first calculated the geodesic distance from all the vertices on the branches to the bottom of the main branch. A persistence barcode was generated for each

[1] https://github.com/tzutalin/labelImg

image following a published approach (Li et al., 2017, 2019). Pairwise distance between different barcodes were calculated using bottleneck distance (Cohen-Steiner et al., 2007) and multi-dimensional scaling (MDS) was used to perform dimension reduction in this pair-wise similarity space to obtain the coordinates of the first three MDS dimensions. Only first three dimensions were used in our analysis and other dimensions were ignored in this analysis because these lower rank dimensions provide limited variation regarding the overall similarity between different branching networks. We used Euclidean MDS-PCA space to approximate the non-linear topological space. The percentage numbers calculated from variation of PCA from the MDS results are the estimation of the variation. Correlations between different traits and heatmap were generated using R programming language and pheatmap package[2]. Matlab codes are provided in our github repository[3].

Terminology Used for Shoot Architecture Analysis

Although a few excellent review papers have described the shoot architecture of many plant species (Benlloch et al., 2015; Teichmann and Muhr, 2015; Wang et al., 2018), there is no commonly accepted terminologies for edamame and soybean plants in order to provide detailed description of the branching patterns that we are aiming to study. Therefore, we provide a schematic diagram to illustrate the terminologies used in our analysis (**Supplementary Figure 2**). Those terminologies are as follows: plant height (PH), pod number (PN), first pod height (FPH), multidimensional scaling 1 (MDS1), multidimensional scaling 2 (MDS2), multidimensional scaling 3 (MDS3), average primary branch length (APBL), main branch length (MBL), total branch length (TBL), first node height (FNH), pod number above 10 cm from ground (PN10), height above ground for 5% pods (P5H), height above ground for 1% pods (P1H), first internode length (FIL), second internode length (SIL), third internode length (TIL), number of primary branches (NPB). In the diagram, lines with arrowheads represent branches and circles represent flowers/pods (**Supplementary Figure 2A**). The main branch in our terminology is sometime called main stem in other publications. All edamame varieties in our study first generated several primary branches on the main branch before producing pods on the main branch. Primary branches are the side branches that directly emerged from the main branch and secondary branches are those initiated from the primary branches. In our data, only a few varieties generated secondary branches such that we did not include secondary branches in our analysis. Detailed descriptions of these terminologies are included in the schematic diagram in **Supplementary Figure 2**.

Genetic Data Analysis

GWAS QTL markers were downloaded from soybase.org (Grant et al., 2010). Three published GWAS studies (Hao et al., 2012; Zhang H. et al., 2015; Fang et al., 2017) analyzed the shoot architecture traits in soybeans and these publications provided the marker names or candidate genes. Only statistically significant markers from these publications were used for our analysis. Because different studies used different genotyping approaches, we try to match the markers used in our study (50 K SNP array) to markers used in other publications by determining the genomic locations of these markers on the same reference genome. For the known markers that are associated with pod numbers, we first identified their locations in a recent soybean genome release (Wm82.a2.v1). We then identified those 50 K SNP array markers that are most close to these published markers (within 50 kb). In most cases, we can find associated markers within 10 kb from the published markers and in some cases, there are multiple markers located within our predefined genomic range. Marker data were downloaded from soybase.org as a Variant Call Format (VCF) file and the genotypes were summarized per plant based on whether the plant was having pod numbers higher than average or lower than average (**Table 1**). The association of markers with the pod number trait was tested using fisher's exact test ($p < 0.05$). Candidate genes were identified as those genes that are more close to the significant SNP markers. In case the marker is located in a gene dense region, the gene functions were manually selected based on their homology to other plant genes that are more likely to be involved in regulating pod numbers.

[2]https://cran.r-project.org/web/packages/pheatmap/

[3]https://github.com/maoli0923/Edmame-Shoot-Architecture

RESULTS

Overview of Phenomics Analysis Pipeline

To understand the genetic control of shoot architecture in edamame plants, we used a mini-core collection of 151 edamame PIs with maturity group (MG) IV and V that are adapted to the growth conditions in Virginia as our model population (**Figure 1**). Maturity group zones are defined as the areas where a cultivar is best adapted. MG IV and V are best adapted to the growth conditions of most of the southern states and in Virginia

TABLE 1 | Heritability of plant traits.

Traits	Heritability
Pod number (PN)	0.91
First pod height (FPH)	0.91
MDS2	0.88
Average primary branch length (APBL)	0.85
Main branch length (MBL)	0.85
Total branch length (TBL)	0.85
MDS1	0.84
Plant height (PH)	0.84
First node height (FNH)	0.83
Pod number above 10 cm (PN10)	0.83
MDS3	0.75
Number of primary branch (NPB)	0.73
Second internode length (SIL)	0.70
Short internode (SIN)	0.56
Third internode length (TIL)	0.48
First internode length (FIL)	0.30

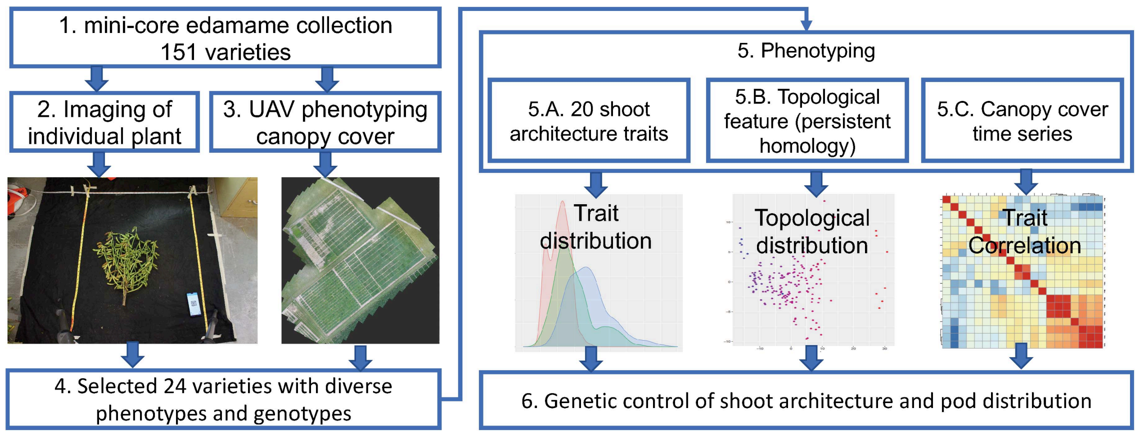

FIGURE 1 | Workflow of phenomics analysis of edamame shoot architecture and canopy cover. Step 1. Data from a mini-core collection of edamame varieties were used for this study. Step 2. Individual plant was imaged twice on a black background. Step 3. Unmanned aerial vehicle was used to collect canopy cover data over the growth season. Step 4. Selected varieties were used for detailed characterization of the shoot architecture. Step 5. Phenotypic data analyses were performed. Step 6. Potential genetic control points of shoot architecture were analyzed.

(Egli, 1993; Mourtzinis and Conley, 2017). For each of these 151 PIs, we have collected two types of image data. To study the shoot architecture and pod locations, we harvested 2–4 plants per PI and removed all leaves and petioles before imaging (**Figure 1**, step 2). Because the shoot of edamame is bilateral-symmetrically distributed, we only need two photos for the "front" and "back" of each plant to capture the variations in the branching patterns. We also collected drone images to study edamame canopy coverage over the growth season (**Figure 1**, step 3). From these 151 PIs, we selected 24 varieties for detailed characterization because of their diversity in the genomic sequences as well as phenotypic variations. These images were manually labeled to identify the location of each pod with ImgLabel software, and also to trace the branches using a modified ImageJ plugin (Rueden et al., 2017). For each plant, 20 phenotypic traits, including the length of the main branch, the number of primary branches and the number of pods were measured. The terminologies used here are described in details in section "Materials and Methods" and in **Supplementary Figure 2**. We further translated the branching patterns using a topological approach called persistent homology and projected the topological pattern into lower dimensional space (Li et al., 2017). Finally, we studied the trait correlations between the visible traits measured on images with topological traits (**Figure 1**, step 5). To understand the potential genetic control of these traits, we analyzed the published SNP map of these 24 varieties and studied whether some known major QTLs control shoot architectures are candidate regulatory regions in these varieties.

Correlation of Shoot Architecture Parameters Between Technical Replicates

The shoot architecture and phyllotaxis of edamame has not been extensively documented before. Our approach of phenotyping involved taking plant images on a flat surface from both sides. We hypothesize that this approach can capture most variations of branching patterns. To test this hypothesis, we manually analyzed 178 photos (89 pairs of images) form 24 varieties of edamame and measured 12 geometric traits related to the branching patterns and shoot architecture (**Figure 2A**). Among these parameters, we found that five parameters showed high correlation between the images taken on both sides of the same plant. These parameters include plant height (PH, **Figure 2B**), main branch length (MBL, length of longest branch in cm), pod numbers (PN, total number of pods), total branch length (TBL, the sum of lengths of all branches) and average primary branch length (APBL, the sum of lengths of all branches divided by the total number of branches). The high correlations of these parameters between two images of the same plant are expected and suggested that only taking one image on one side of the plant is sufficient to capture the variations of these parameters in our plant population. Here technical replicates represent the two images of the same plant taken from two different sides. First pod height (FPH) is a parameter that is related to harvester efficiency (Tadesse and Chris, 2007), and this parameter showed lower correlation (0.86) than plant height. This is likely due to the fact that branches are flexible and when flipping the plants while taking the images, some branches can change their position. Parameters such as branch length will not be changed but the location of a pod relative to the bottom of a plant will be affected.

Some other parameters, such as first internode length (FIL, **Figure 2C**) and second internode length (SIL) showed lower correlation of 0.63 and 0.60, respectively, but the correlations are still statistically significant. The only parameter that is not significantly correlated between the two images of the same plant is the third internode length (TIL, **Figure 2A**). Interestingly, these parameters are not affected by the position of the branches. There are two reasons that might explain these lower correlations. First, some outlier observations (**Figure 2C**) could reduce the overall correlation. Second, some internodes are very short and

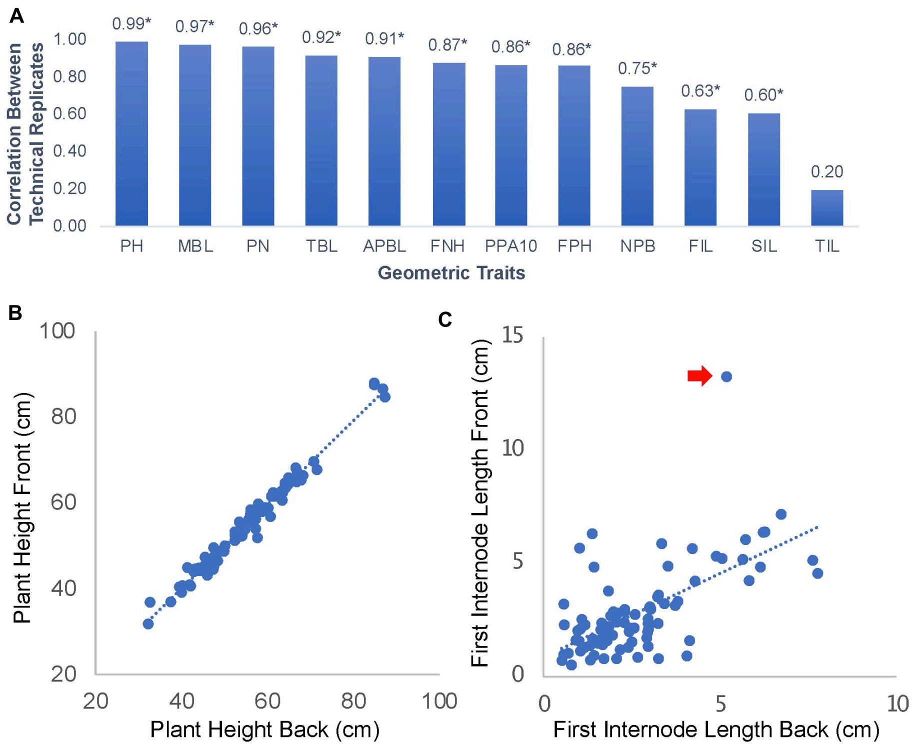

FIGURE 2 | Parameter correlations for images of edamame shoot architecture. **(A)** 12 geometric traits were measured in this study. PH, plant height; MBL, main branch length; PN, pod number; TBL, total branch length; APBL, average primary branch length; FNH, first node height; PPA10, percent of pod above 10 cm from ground; FPH, first pod height; NPB, number of primary branches; FIL, first internode length; SIL, second internode length; TIL, third internode length. Numbers on top of the bars are Pearson Correlation Coefficient (PCC). * indicates statistical significance ($p < 0.01$). **(B)** Scatter plot of plant height measurement between technical replications (front and back images of the same plant). **(C)** Scatter plot of first internode length. Red arrow indicates an outlier data point.

the precise location where the primary branches connected to the main branch could be blocked on one side of the plant but more visible on the other side of the plant. These results suggest that for majority of shoot architecture parameters, our approach of image analysis can provide accurate measurements. Cautions should be taken when trying to interpret results from FIL, SIL, and TIL which showed variable results that were affected by the image analysis process.

Distribution of Shoot Architecture Parameters

To understand the shoot architecture of edamame plants, we analyzed the distribution of shoot architecture parameters of our plant population (**Figure 3**). We first focused on parameters related to plant height and branch length (**Figure 3A**), and we found that the distribution of PH is almost identical to the MBL. There is a small shift toward longer length when measuring MBL as compared to PH. This is expected because we measure plant height by measuring the distance from the ground to the top of the main branch. For some edamame varieties, main branches may have small angles at each internode, therefore overall length of the main branch is highly similar to the overall height of the plants with some varieties have longer MBL than PH. The average MBL and PH are 55.7 and 54.4 cm, respectively. The average PH is shorter than data generated by measuring PH (68–81 cm) in the field conditions (Zhang H. et al., 2015; Jiang et al., 2018), and this difference is likely due to the fact that we removed all petioles from the plants before the measurement. Petioles of edamame are very long (∼30 cm) and contribute significantly to the height of the plant if the measurement is taken at field when leaves are still green and before the plant reaches full maturity. Another possibility is that different soybean varieties were used in published studies. The average primary branch length is approximately half of the main branch length (**Figure 3A**, see section "Materials and Methods"). This result shows that the major contributor of the plant height for edamame is the main branch length and other primary branches are shorter than the main branches on average.

To understand where primary branches are emerged from the main branch, we have measured four parameters: first node height, first internode length, second internode length and third internode length (**Figure 3B**). First node height is the length from ground to node where first primary branch meets the main branch. First internode length is the distance from where the first primary branch meets the main branch to where the second primary branch meets the main branch. The second and the third internode length are similarly defined. There are cases where we found very short internode and such short internodes were ignored in our analysis but were included as one feature

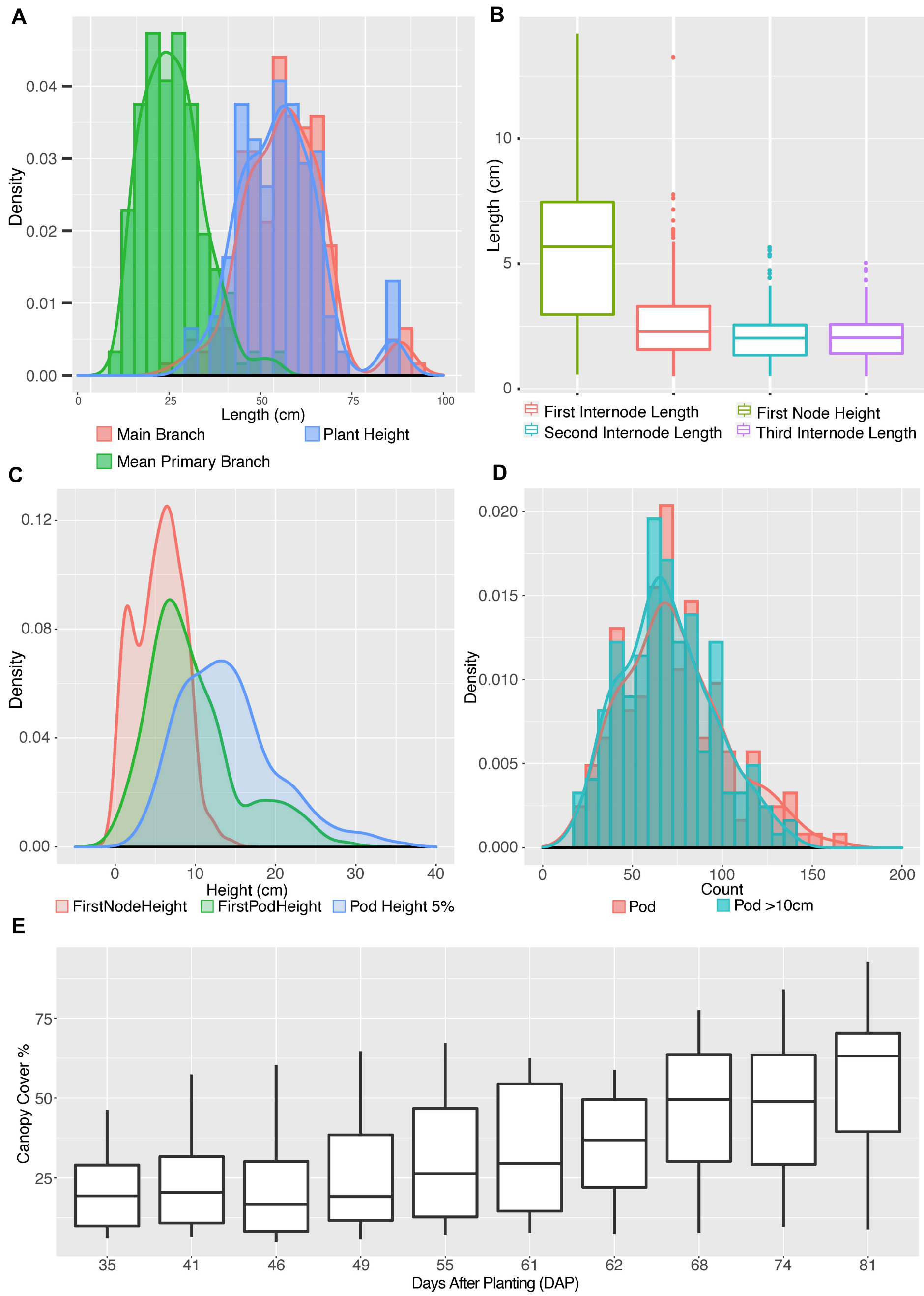

FIGURE 3 | Distribution of shoot architecture parameters in edamame plants. **(A)** Histograms and density plots for main branch length, plant height and average primary branch length. **(B)** Boxplot shows the distribution of first pod height, first, second and third internode lengths. **(C)** Density plots compare first node height with first pod height and 5% pod height. **(D)** Histograms and density plots for number of pods per plant and number of pods above 10 cm from ground. **(E)** Boxplot of canopy cover changes in the growth season.

in our phenotypic analysis (**Supplementary Table 1**). Such a short internode is a known feature for some soybean plants and whether such a short internode is genetically controlled is still not well understood (Yoshikawa et al., 2013).

Our results (**Figure 3B**) show that the average first node height is 5.47 cm, which is almost twice the length of average first internode length (2.75 cm). The second and third internodes are relatively shorter than the first internode but are similar to each other with average lengths of 2.08 and 2.10 cm, respectively. Once the first primary branch has emerged, the second and third primary branches will emerge subsequently after similar intervals. Note that there is a large variation in the height of first node, with 2.97 and 7.47 cm at first quartile and third quartile, respectively. This interquartile length is 5.8 times larger than the interquartile length of first internode, showing a large variation on the development of main branch before transition into producing primary branches.

To understand how pod production is correlated with length of plant branches, we generated the density distributions of three parameters: first node height, first pod height and 5% pod height. The 5% pod height is defined as the distance from ground when 5% of all pods were observed on a plant. We chose to measure 5% pod height because the lower 5% of pods are more likely to be lost during the harvesting process than other pods. The average first pod height is 9.72 cm whereas the average height of the first internode is 5.47 cm. In fact, the average first pod height is higher than the sum of first node and first internode (7.55 cm) and slightly lower than the second internode (10.30 cm). Interestingly, the density distribution of the first node height and first pod height both seem to show two peaks (red and green curves in **Figure 3C**). The separation of two peaks in the first pod height distribution is very clear with one summit of the distribution at ∼7 cm and the second summit at ∼19 cm. Although these two distinct peaks were not clearly visible in the 5% pod height distribution but a wide distribution of this parameter is noticeable. These results suggest that plants in this study can be approximately categorized into two types according to where the first pods were produced.

We also analyzed the distribution of total pod numbers and compared to the pod numbers above 10 cm from ground (**Figure 3D**). As expected, two distributions are highly overlapping, with the average pod number above 10 cm from ground (green histogram) slightly lower than total pod number (red histogram), suggesting some varieties have pod located below 10 cm from ground level. These close-to-ground pods are difficult to be picked up by mechanical harvesting. Finally, we analyzed the change of canopy cover of these edamame varieties during the growth season with drone images from 35 days after planting (DAP) to 81 days after planting. The average canopy cover showed a steady increase over the growth season as expected. We selected 61DAP and 81DAP data from canopy cover data for downstream analysis. These dates were selected because 61DAP represents one of the early dates of canopy expansion and 81DAP represents one of the late dates of canopy expansion, respectively.

Persistent Homology of Shoot Architecture Uncovers Hidden Connection Between Branching Patterns and Plant Productivity

Because different varieties of edamame have different numbers of branches and the directions of these branches can also vary, comprehensive comparison of the branching patterns between different varieties are challenging. To solve this problem, we applied a mathematical approach called persistent homology (Verri et al., 1993; Carlsson, 2009; Edelsbrunner and Harer, 2010) to convert the complex patterns on a 2D image into a topological space where branching patterns are comparable between different samples. Images of edamame plants were manually skeletonized and labeled as main branch and primary branches (see section "Materials and Methods"). Secondary branches were not included in this analysis. For each plant image, the branching skeleton was translated into a network representation and the geodesic distance (also the shortest path length) from vertices on each branch to the ground was calculated. A persistence barcode summarizing the branching topological information was generated (Li et al., 2017, 2019) for each plant and the distance between different plants were calculated using bottleneck distance (Cohen-Steiner et al., 2007). Multidimensional scaling was used to project the distance between different branching patterns into low dimensional space and only the first three dimensions were included in this analysis (MDS1, MDS2, and MDS3, see section "Materials and Methods"). These three dimensions explained around 54.3% of total variations. The top 30 dimensions explained around 85% of total variations, however, the fourth dimension and above each explains very small fraction of the total variations such that they were not included in our analysis.

To understand the relationship among the low dimension projection of the persistent homology and other traits, we performed pair-wise correlation analysis with hierarchical clustering (**Figure 4A**). The clustering heatmap shows that the first pod height (FPH) is highly correlated with the height of 1 and 5% of total pods (P1H PCC = 0.981 and P5H PCC = 0.921, $p < 2.2e\text{-}16$, **Supplementary Figures 3A,B**). We also found that the canopy cover at 61DAP is highly correlated (Pearson correlation 0.87, $p < 2.2e\text{-}16$, **Supplementary Figure 3C**) with canopy cover at 81DAP, which is consistent with field observations. However, canopy cover traits do not show strong correlation with any other single trait, suggesting combination of multiple shoot architecture traits or other traits (petiole length or leaf surface area) that are not measured in our study may contribute to the canopy cover. Another pair of high correlation was found between total branch length (TBL) and average primary branch length (APBL, PCC = 0.550, $p < 1.72e\text{-}15$, **Supplementary Figure 3D**), because the APBL equals TBL divided by number of branches.

With regard to topological traits (MDS1, MDS2 and MDS3), we found strong correlation between MDS1, plant height (PCC = 0.936, $p < 2.2e\text{-}16$, **Supplementary Figure 3E**) and main branch length (PCC = 0.940, $p < 2.2e\text{-}16$), suggesting that the major variation in the topological space is related to plant height. To confirm this correlation, we plotted the plant height on the two-dimension MDS plot (**Figure 4B**) and we indeed found that higher MDS1 corresponds to taller plants and lower MDS1 corresponds to shorter plants. Surprisingly, we found that MDS2 is positively correlated with pod number (PN, PCC = 0.293, $p < 7.03e\text{-}05$, **Supplementary Figure 3F**). This is an intriguing observation because in our data processing pipeline, pods and branches are labeled separately (with different software, ImgLabel for pods and ImageJ for branches). In another word, the data were analyzed independently, but the analysis showed a positive correlation between these two traits. To confirm this relationship, we

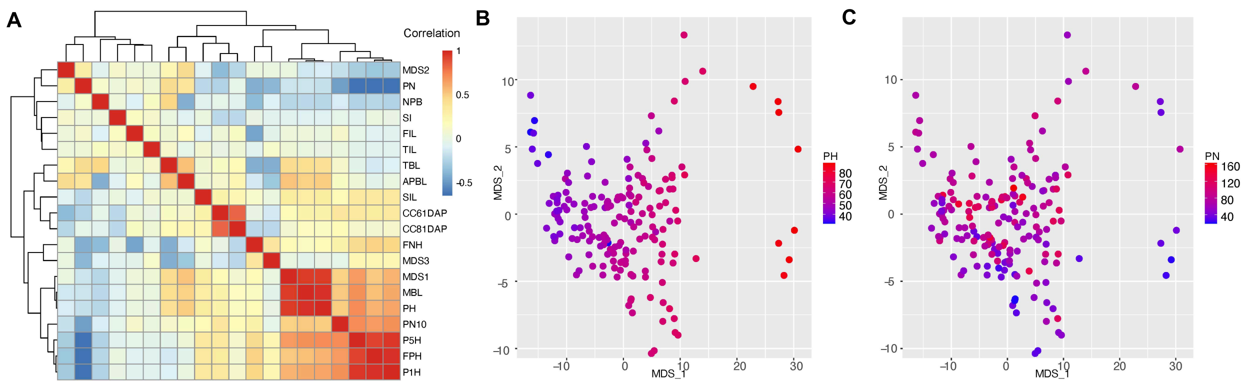

FIGURE 4 | Correlation analysis of traits characterized in this study. **(A)** correlation between edamame shoot architecture traits with topological traits (MDS1, 2, and 3). Trait names are described in **Figure 2** and main text. PN, pod number; NPB, number of primary branches; SI, short internode; FIL, first internode length; SIL, second internode length; TIL, third internode length; TBL, total branch length; APBL, average primary branch length; CC61DAP, canopy cover 61 days after planting; CC81DAP, canopy cover 81 days after planting; FNH, first node height; MBL, main branch length; PH, plant height; PN10, pod number above 10 cm from ground; P5H, height above ground for 5% pods; P1H, height above ground for 1% pods; FPH, first pod height. **(B)** Comparison of MDS1 and MDS2 with plant height. **(C)** Comparison of MDS1 and MDS2 with pod number.

plotted the pod number on the two-dimension MDS plot (**Figure 4C**), and we found that small MDS2 indeed correlated with low number of pods and large MDS2 tends to have higher number of pods. However, some plants with highest number of pods (*PN* = 140–160) do not have high MDS2, suggesting that additional variation cannot be explained by MDS2. Further analysis of the correlation map shows that MDS3 does not seem to explain this additional variation, but MDS3 is positively correlated with first node height (FNH, PCC = 0.363, $p < 6.42e\text{-}07$, **Supplementary Figure 3G**), which is another branch length related trait.

A central question of this work is to investigate the distribution of pods on the plant branches. This trait has been studied in soybeans (Illipronti et al., 2000; Liu et al., 2010a; Ning et al., 2018) and several other crop species (Kigel et al., 1991; Decoteau and Graham, 1994). There are five traits directly related to pod distribution and yield, which include pod number (PN), pod number above 10 cm (PN10), first pod height (FPH), P1H and P5H. Interestingly, pod number is negatively correlated with other four traits related to pod distribution on the plant. Pod distribution is directly related to the harvest efficiency, and specifically, the first pod height is positively correlated with harvest efficiency using combine harvesters (Rajkumar et al., 2012; Beikúfner et al., 2019). The negative correlation (average PCC = −0.601, **Supplementary Figure 3H**) between pod number and first pod height indicates that the lower the pods are produced, the more pods a plant can produce. However, because of this negative correlation, a challenge is to increase first pod height (thus to improve harvest efficiency), without reducing total pod number per plant.

Pods that are close to the ground are more likely to be lost due to harvest than those that are away from the ground. PN also has negative correlations with plant height (PCC = −0.18, $p < 0.021$, **Supplementary Figure 3I**), first node height (PCC = −0.417, $p < 6.68e\text{-}09$) and several other traits that are related to plant statues. These results show there is a trend where taller plants tend not to produce as many pods as shorter plants in this edamame population under our experimental condition such as plant density and local climate.

Genetic Control for Edamame Shoot Architecture

To investigate whether the shoot architecture traits are genetically controlled, we first calculated the heritability using a linear mixed effect model (Nyquist and Baker, 1991) with lme4 (Bates et al., 2015) package (**Table 1**) with correction for unbalanced number of replicates. We found high heritability (>70%) for most traits measured in our study. For example, plant height is a trait that has been studied in many prior reports and the estimated heritability of plant height is 84%, which is similar to what has been reported in other soybeans (85%) and edamame (79%) populations (Chang et al., 2018; Jiang et al., 2018). Other traits not in other published work but analyzed in our study also showed high heritability. In particular, pod number and first pod height have the highest heritability of 91%. Interestingly, the topological traits such as MDS1 and MDS2 also have high heritability. MDS1 is highly correlated with plant height and have the same heritability as plant height. MDS2 is positively correlated with pod number and have slightly lower heritability as compared to pod number. There results suggest, under our experimental environment and plant density, there is evidence of genetic control of pod numbers and first pod heights in our selected varieties of edamame.

Because producing large numbers of fresh pod is a major goal for edamame breeding, to further investigate the candidate genes that control the pod numbers, we collected known GWAS QTLs that are associated with pod numbers in soybeans. These known QTLs were downloaded from the Soybase and the SNPs that are associated with pod numbers were provided from three studies (**Supplementary Table 2**). There are 26 SNPs/genomic locations that are associated with pod number from chromosome 1, 2, 5, 6, 9, 11, 13, 15, 17, 18, and 19. Nine candidate genes were provided by one of the publications and others publications did not provide candidate genes. Because different studies have different SNP markers and we used the 50 K SNP array data for our selected edamame lines, we identified the SNP markers in our marker data that are closely localized to the published SNP markers. We found 114 SNP markers in our genotyping data that are within 50 Kb from these published markers. Using fisher's exact test, we determine the association of these SNP markers to the pod number trait (**Table 2**).

We found eight SNP markers in our population showed statistically significant association with pod numbers. For example, ss715609881 appears as the same as the reference allele in 12 varieties with low pod numbers (**Table 2**, marked by $), and in one variety that has high pod number (**Table 2**, marked by #). Low and high pod numbers are defined as pod numbers below or above average, respectively. A candidate gene, Glyma.11G164800, is located within 5 kb from this SNP marker, and this gene encodes a LOB (lateral organ boundary) protein. Although the function of this particular gene has not been characterized, members of this gene family have been shown to be related to flower and embryo development in Arabidopsis (Chalfun-Junior et al., 2005; Borghi et al., 2007), maize (Evans, 2007), and rice (Li et al., 2008). These results support a potential role of this candidate gene in regulating pod numbers in edamame.

Another example marker is ss715582578, which appears as the same as the non-reference alleles in 12 varieties with low pod number and in six varieties with high pod number ($p = 0.014$). There are two candidate genes that are located within 10 kb of this SNP marker (**Table 2**). One of the candidate gene downstream of this SNP marker (Glyma.02G216600) encodes homologous gene to AGAMOUS-like 16. Genes in the AGAMOUS gene family are well known for their functional role in floral development in plants including Arabidopsis (Mizukami and Ma, 1997) and soybeans (Chi et al., 2017). These results suggest that Glyma.02G216600 might be a shoot architecture-related gene. The upstream gene (Glyma.02G216500) is a TRAF-type zinc finger-related transcription factor which is poorly characterized, but can also be considered as a candidate gene because its potential role of expression regulation. We analyzed a published gene expression data comparing shoot apical meristems with leaves in soybeans (Wong et al., 2013). We found six genes are highly expressed in shoot apical meristems with $\log_2$ fold change higher than 1 (**Supplementary Table 3**), which indicates more than twofold up regulation of these genes in a tissue type that is directly related to pod formation. These results further support the potential roles of these genes in regulating pod number in edamame.

DISCUSSION

With the decreasing cost of sequencing, many soybean varieties have been either genotyped using SNP arrays (Song et al., 2013) or whole-genome re-sequencing (Zhou et al., 2015), which provide a rich resource of genetic markers and potential functional genetic variations. Therefore, to fully utilize such genetic resources, one approach is to perform association studies to identify SNP markers for traits of interest. A bottleneck for such association study is the ability to collected phenotypic data for a large population with different genetic makeup at field scale. In soybean research, a large number of GWAS studies have been published including those studies canopy cover (Xavier et al., 2017) and shoot architecture traits (Zhang J. et al., 2015; Fang et al., 2017). With the increasing use of drones in field research, measuring canopy cover and plant height from drone images has become a preferred approach and have led to the discovery of many GWAS QTL that are associated with these traits (Mogili and Deepak, 2018).

However, in our study, we are interested in the distribution of pods on the plant and how this trait is related to other shoot architecture traits. The phenotyping task for this study is challenging because the shoot structure is covered by leaves when edamame were harvested. Manually removing all leaves

TABLE 2 | Candidate genes associated with pod number.

Pod number			High		Low					
Published QTL marker	**Ref.**	**SNP id**	**NR**	**Ref**	**NR**	**Ref**	***p*-value**	**Candidate gene**	**Fold change**	**Gene function**
Chr11:15649090	1	ss715609881	11	1#	0	12$	9.60E-06	Glyma.11G164800	3.17*	LOB domain-containing protein 4
Chr15:5252046	1	ss715622826-7	3	9	11	1	2.80E-03	Glyma.15G068500	8.6*	60S ribosomal protein L26-1-like
Chr18:55626229	2	ss715632229	5	7	12	0	4.60E-03	Glyma.18G274000	16.83*	RING finger protein 165-like
Chr11:5338661	3	ss715610851-6	8	4	1	11	9.40E-03	Glyma.11G071000	1.28*	Auxilin-like protein 1-like
Chr02:40368201	3	ss715582578	6	6	12	0	1.40E-02	Glyma.02G216500	2.44*	TRAF-type zinc finger-related
Chr02:40368201	3	ss715582578	6	6	12	0	1.40E-02	Glyma.02G216600	1.39*	AGAMOUS-like 16
Chr11:33902439	1	ss715610584	8	4	2	10	3.60E-02	Glyma.11G245300	0.72	63 kDa inner membrane protein
Chr18:55662445	2	ss715632230	7	5	12	0	3.70E-02	Glyma.18G274200	0.08	Pollen Ole e1 extensin family protein

NR, non-reference allele; Ref, reference allele. The numbers of NR and Ref indicate the number of varieties out of 24 PIs. Low and High are the pod numbers and are defined as pod numbers below or above average, respectively. The p-values were calculated using fisher's exact test. Candidate genes were selected with 50 Kb window of enriched SNP based on published 50K SNP data. References: 1. Zhang J. et al. (2015). 2. Fang et al. (2017). 3. Hao et al. (2012). SNP id ss715622826-7 means ss715622826 and ss715622827. SNP id ss715610851-6 means ss715610851, ss715610852, ss715610853, ss715610854, ss715610855, and ss715610856. The genes marked with asterisk (*) have $\log_2$fold change > 1 indicating the candidate genes are expressed more than two-fold higher in shoot apical meristem than in leaves. # and $ mark the numbers described in the main text.

is the major time limiting step in our analysis. One alternative approach is to collect image data after all leaves are dropped off (for example see Sun et al., 2019), which will be explored in a future study. Another time-consuming step in our analysis is to cut the plants and lay the plants on a flat surface for imaging. An alternative approach would be to use 3D imaging or LiDAR to collect data in the field without cutting down the plants (Sun et al., 2018; Dhami et al., 2019). Another major obstacle to extend this study is that manually label all the branches and pod in each image is also a time and labor-intensive process. How to automatically detect the pod locations using machine learning such as YOLO algorithms (Redmon and Farhadi, 2018) will be tested using data collected in this work. Finally, with the image data collected in this study, we can test whether machine learning methods can generate semantic labels (Barth et al., 2019; Adams et al., 2020) such as the location of first internode or to determine the main branch and primary branches. Unlike the object detection task, such as detecting pods in an image, determine the difference between main branch and primary branches require the computer algorithms to understand how branches are organized in the image. This is a more challenging task for machine learning than simply detecting objects such as pods in an image.

Although the population used in our study has been genotyped using 50 K SNP array, some of the known markers published by other studies are not represented in our SNP data. For example, the well-known Dt1 alleles (Liu et al., 2010b; Tian et al., 2010) that regulate stem growth habits are not represented in our SNP data and we can only use a SNP that is closely linked to Dt1 locus as a proxy to estimate the genotype of the individual plants in our population. Although all of our selected individuals (PIs) are homozygous at both Dt1 (and Dt2) neighboring loci, these data cannot rule out the possibility that there are mutations in Dt1 and Dt2 loci that are associated with the observed variations in phenotypes in our population. Several recent studies have shown that genomic variations at regulatory sequences for major QTL genes can be important in explaining the missing heritability problem that is commonly observed in GWAS studies (Zhou et al., 2019; Alonge et al., 2020; Liu et al., 2020). Therefore, a potential next step for our work is to generate the genomic sequences of the key genes or resequencing the entire genome to determine whether additional genetic variations existed in these key regulators of plant architectures and whether those variations are functionally connected to the observed phenotypes.

In our phenotypic analysis, we used a topological approach to understand plant shoot architecture. As compared to simple geometric features, such as branch length and internode length, topological features take into account the overall structure of the branching patterns and have been shown to provide more informative features for trait association studies (Li et al., 2019). In our case we found that first dimension of the topological distance (MDS1) is highly correlated with plant height, which suggests that topological analysis did capture important plant architecture parameters. More interestingly, we found that the MDS2, but not other branch geometric parameters, is correlated with pod numbers. This is an important insight from the topological data analysis where the hidden features of branch patterns can be associated with the number of pods produced on the whole plant. Additional analysis could help to dissect the connection between MDS2 and pod number. For example, from the clustering dendrogram in **Figure 4A**, we found that MDS2 and pod number are in the same clade as four other geometric parameters including number of primary branches, short internode, first internode length and third internode length. Other fourteen traits are clustered in a separate clade as compared to these six traits. However, the correlations between each of the four geometric traits with pod number are close to zero. This result suggests that combinations or interactions of the four geometric traits could be related to total pod numbers. This can be observed in our data, for example, more primary branches can lead to more pod formation for a plant. However, when there is short internode in a plant, the two branches sometimes are underdeveloped and do not produce as many pods as those found in typical branches. Additional quantitative analysis can further help to elucidate such relationships between different traits.

In conclusion, we performed phenotypic analysis of a small collection of edamame varieties and identified intriguing correlations between geometric traits and topological traits in these varieties. Using known genetic markers and genes that are associated with pod numbers, we found several novel candidate/putative genes that might be related to the pod numbers. We found a negative correlation between pod number and first pod height, which suggests that breeding for new varieties of edamame to optimize the distribution of pods on plants will require a balance between these two traits. In the future, a larger population of edamame varieties would be required to perform GWAS study to identify more markers and candidate genes using our analytical pipelines developed in this work.

DATA AVAILABILITY STATEMENT

The original data presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

SL and BZ designed the experiment. BZ developed the edamame population and provided plant samples. KD and QZ performed field phenotyping experiments. ML performed topological analysis for the branching patterns. KD, SL, and ML wrote the manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.614926/full#supplementary-material

REFERENCES

Adams, J., Qiu, Y., Xu, Y., and Schnable, J. C. (2020). Plant segmentation by supervised machine learning methods. *Plant Phenome J.* 3:e20001. doi: 10.1002/ppj2.20001

Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., et al. (2020). Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* 182, 145–161.e23. doi: 10.1016/j.cell.2020.05.021

Bao, A., Chen, H., Chen, L., Chen, S., Hao, Q., Guo, W., et al. (2019). CRISPR/Cas9-mediated targeted mutagenesis of GmSPL9 genes alters plant architecture in soybean. *BMC Plant Biol.* 19:131. doi: 10.1186/s12870-019-1746-6

Barth, R., IJsselmuiden, J., Hemming, J., and Van Henten, E. J. (2019). Synthetic bootstrapping of convolutional neural networks for semantic plant part segmentation. *Comput. Electron. Agric.* 161, 291–304. doi: 10.1016/j.compag.2017.11.040

Bates, D., Machler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Beiküfner, M., Hüsing, B., Trautz, D., and Kühling, I. (2019). Comparative harvest efficiency of soybeans between cropping systems affected by first pod height and plant length. *Org. Farming* 5, 3–13. doi: 10.12924/of2019.05010003

Benlloch, R., Berbel, A., Ali, L., Gohari, G., Millán, T., and Madueño, F. (2015). Genetic control of inflorescence architecture in legumes. *Front. Plant Sci.* 6:543. doi: 10.3389/fpls.2015.00543

Borghi, L., Bureau, M., and Simon, R. (2007). Arabidopsis JAGGED LATERAL ORGANS is expressed in boundaries and coordinates KNOX and PIN activity. *Plant Cell* 19, 1795–1808. doi: 10.1105/tpc.106.047159

Carlsson, G. (2009). Topology and data. *Bull. Am. Math. Soc.* 46, 255–308. doi: 10.1090/S0273-0979-09-01249-X

Carson, L. C. (2010). *Cultivation and Nutritional Constituents of Virginia Grown Edamame*. Master thesis, Virginia Tech University, Blacksburg, VA.

Chalfun-Junior, A., Franken, J., Mes, J. J., Marsch-Martinez, N., Pereira, A., and Angenent, G. C. (2005). ASYMMETRIC LEAVES2-LIKE1gene, a member of the AS2/LOB family, controls proximal?distal patterning in *Arabidopsis* petals. *Plant Mol. Biol.* 57, 559–575. doi: 10.1007/s11103-005-0698-4

Chang, F., Guo, C., Sun, F., Zhang, J., Wang, Z., Kong, J., et al. (2018). Genome-Wide association studies for dynamic plant height and number of nodes on the main stem in summer sowing soybeans. *Front. Plant Sci.* 9:1184. doi: 10.3389/fpls.2018.01184

Chen, Y., and Nelson, R. L. (2004). Evaluation and classification of leaflet shape and size in wild soybean. *Crop Sci.* 44, 671–677. doi: 10.2135/cropsci2004.6710

Chi, Y., Wang, T., Xu, G., Yang, H., Zeng, X., Shen, Y., et al. (2017). GmAGL1, a MADS-Box gene from soybean, is involved in floral organ identity and fruit dehiscence. *Front. Plant Sci.* 8:175. doi: 10.3389/fpls.2017.00175

Cohen-Steiner, D., Edelsbrunner, H., and Harer, J. (2007). Stability of persistence diagrams. *Discrete Comput. Geom.* 37, 103–120. doi: 10.1007/s00454-006-1276-5

Decoteau, D. R., and Graham, H. A. H. (1994). Plant spatial arrangement affects growth, yield, and pod distribution of cayenne peppers. *HortScience* 29, 149–151. doi: 10.21273/HORTSCI.29.3.149

Dhaliwal, D. S., and Williams, M. M. (2020). Economically optimal plant density for machine-harvested edamame. *HortScience* 55, 368–373. doi: 10.21273/HORTSCI14642-19

Dhami, H., Yu, K., Xu, T., Zhu, Q., Dhakal, K., Friel, J., et al. (2019). Crop height and plot estimation from unmanned aerial vehicles using 3D LiDAR. *arXiv* [preprint].

Edelsbrunner, H., and Harer, J. L. (2010). *Computational Topology: An Introduction*. Providence RI: American Mathematical Society.

Egli, D. B. (1993). Cultivar maturity and potential yield of soybean. *Field Crop Res.* 32, 147–158. doi: 10.1016/0378-4290(93)90027-K

Egli, D. B., and Bruening, W. P. (2006). Temporal profiles of pod production and pod set in soybean. *Eur. J. Agron.* 24, 11–18. doi: 10.1016/j.eja.2005.04.006

Evans, M. M. S. (2007). The indeterminate gametophyte1 gene of maize encodes a LOB domain protein required for embryo sac and leaf development. *Plant Cell* 19, 46–62. doi: 10.1105/tpc.106.047506

Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z., Yang, R., et al. (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* 18:161. doi: 10.1186/s13059-017-1289-9

Fehr, W. R., and Caviness, C. E. (1977). Stages of soybean development. *Spec. Rep.* 87, 7–8.

Fenta, B. A., Beebe, S. E., Kunert, K. J., Burridge, J. D., Barlow, K. M., Lynch, J. P., et al. (2014). Field phenotyping of soybean roots for drought stress tolerance. *Agronomy* 4, 418–435. doi: 10.3390/agronomy4030418

Grant, D., Nelson, R. T., Cannon, S. B., and Shoemaker, R. C. (2010). SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* 38, D843–D846. doi: 10.1093/nar/gkp798

Hao, D., Cheng, H., Yin, Z., Cui, S., Zhang, D., Wang, H., et al. (2012). Identification of single nucleotide polymorphisms and haplotypes associated with yield and yield components in soybean (*Glycine max*) landraces across multiple environments. *Theor. Appl. Genet.* 124, 447–458. doi: 10.1007/s00122-011-1719-0

Illipronti, R. A., Lommen, W. J. M., Langerak, C. J., and Struik, P. C. (2000). Time of pod set and seed position on the plant contribute to variation in quality of seeds within soybean seed lots. *NJAS Wageningen J. Life Sci.* 48, 165–180. doi: 10.1016/S1573-5214(00)80012-3

Jiang, G.-L., Rutto, L. K., and Ren, S. (2018). Evaluation of soybean lines for edamame yield traits and trait genetic correlation. *HortScience* 53, 1732–1736. doi: 10.21273/HORTSCI13448-18

Kantolic, A. G., and Slafer, G. A. (2007). Development and seed number in indeterminate soybean as affected by timing and duration of exposure to long photoperiods after flowering. *Ann. Bot.* 99, 925–933. doi: 10.1093/aob/mcm033

Kigel, J., Konsens, I., and Ofir, M. (1991). Branching, flowering and pod-set patterns in snap-bean (*Phaseolus vulgaris* L.) as affected by temperature. *Can. J. Plant Sci.* 71, 1233–1242. doi: 10.4141/cjps91-171

Kim, Y.-U., Choi, D.-H., Ban, H.-Y., Seo, B., Kim, J., and Lee, B.-W. (2020). Temporal patterns of flowering and pod set of determinate soybean in response to high temperature. *Agronomy* 10:414. doi: 10.3390/agronomy10030414

Konovsky, J., Lumpkin, T. A., and McClary, D. (1994). "Edamame: the vegetable soybean," in *Understanding the Japanese Food and Agrimarket*, ed. A. D. O'Rourke (Boca Raton, FL: CRC Press), 173–181.

Lee, J. Y., Popp, M. P., Wolfe, E. J., Nayga, R. M. Jr., Popp, J. S., Chen, P., et al. (2018). Information and order of information effects on consumers' acceptance and valuation for genetically modified edamame soybean. *PLoS One* 13:e0206300. doi: 10.1371/journal.pone.0206300

Li, A., Zhang, Y., Wu, X., Tang, W., Wu, R., Dai, Z., et al. (2008). DH1, a LOB domain-like protein required for glume formation in rice. *Plant Mol. Biol.* 66, 491–502. doi: 10.1007/s11103-007-9283-3

Li, M., Duncan, K., Topp, C. N., and Chitwood, D. H. (2017). Persistent homology and the branching topologies of plants. *Am. J. Bot.* 104, 349–353. doi: 10.3732/ajb.1700046

Li, M., Klein, L. L., Duncan, K. E., Jiang, N., Chitwood, D. H., Londo, J. P., et al. (2019). Characterizing 3D inflorescence architecture in grapevine using X-ray imaging and advanced morphometrics: implications for understanding cluster density. *J. Exp. Bot.* 70, 6261–6276. doi: 10.1093/jxb/erz394

Liu, B., Liu, X. B., Wang, C., Li, Y. S., Jin, J., Herbert, S. J. (2010a) Soybean yield and yield component distribution across the main axis in response to light enrichment and shading under different densities. *Plant Soil Environ.* 56, 384–392

Liu, B., Watanabe, S., Uchiyama, T., Kong, F., Kanazawa, A., Xia, Z., et al. (2010b). The soybean stem growth habit gene Dt1 is an ortholog of *Arabidopsis* TERMINAL FLOWER1. *Plant Physiol.* 153, 198–210. doi: 10.1104/pp.109.150607

Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., et al. (2020). Pan-Genome of Wild and Cultivated Soybeans. *Cell* 182, 162–176.e13. doi: 10.1016/j.cell.2020.05.023

Lord, N., Neill, C., and Zhang, B. (2019). *Production and Economic Considerations for Fresh Market Edamame in Southwest Virginia*. Blacksburg, VA: Virginia Cooperative Extension.

Mahoussi, K. A. D., Eric, E. A., Symphorien, A., Florent, J.-B. Q., Flora, J. C., Achille, E. A., et al. (2020). Vegetable soybean, edamame: research, production, utilization and analysis of its adoption in sub-saharan Africa. *J. Hortic. For.* 12, 1–12. doi: 10.5897/JHF2019.0604

Mentreddy, S. R., Mohamed, A. I., Joshee, N., and Yadav, A. K. (2002). "Edamame: a nutritious vegetable crop," in *Trends New Crop New Uses*, eds J. Janick and A. Whipkey (Alexandria, VA: ASHS Press), 432–438.

Mizukami, Y., and Ma, H. (1997). Determination of *Arabidopsis* floral meristem identity by AGAMOUS. *Plant Cell* 9, 393–408. doi: 10.1105/tpc.9.3.393

Mogili, U. R., and Deepak, B. B. V. L. (2018). Review on application of drone systems in precision agriculture. *Procedia Comput. Sci.* 133, 502–509. doi: 10.1016/j.procs.2018.07.063

Mourtzinis, S., and Conley, S. P. (2017). Delineating soybean maturity groups across the United States. *Agron. J.* 109, 1397–1403. doi: 10.2134/agronj2016.10.0581

Nico, M., Mantese, A. I., Miralles, D. J., and Kantolic, A. G. (2016). Soybean fruit development and set at the node level under combined photoperiod and radiation conditions. *J. Exp. Bot.* 67, 365–377. doi: 10.1093/jxb/erv475

Ning, H., Yuan, J., Dong, Q., Li, W., Xue, H., Wang, Y., et al. (2018). Identification of QTLs related to the vertical distribution and seed-set of pod number in soybean [*Glycine max* (L.) Merri]. *PLoS One* 13:e0195830. doi: 10.1371/journal.pone.0195830

Nyquist, W. E., and Baker, R. J. (1991). Estimation of heritability and prediction of selection response in plant populations. *CRC Crit. Rev. Plant Sci.* 10, 235–322. doi: 10.1080/07352689109382313

Ping, J., Liu, Y., Sun, L., Zhao, M., Li, Y., She, M., et al. (2014). Dt2 is a gain-of-function MADS-Domain factor gene that specifies semideterminacy in soybean. *Plant Cell* 26, 2831–2842. doi: 10.1105/tpc.114.126938

Rajkumar, R., Devvrat, S., and Pooja, M. (2012). Selecting soybean (*Glycine max*) genotypes for insertion height of the lowest pod, the useful trait for combine harvester. *Indian J. Agric. Sci.* 82, 511–515.

Redmon, J., and Farhadi, A. (2018). YOLOv3: an incremental improvement. *arXiv* [preprint].

Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., et al. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* 18:529. doi: 10.1186/s12859-017-1934-z

Shanmugasundaram, S., Cheng, S.-T., Huang, M.-T., and Yan, M.-R. (1991). "Varietal improvement of vegetable soybean in Taiwan," in *Proceedings of the Workshop on Vegetable Soybean: Research Needs for Production and Quality Improvement*, (Shanhua: AVRDC), 30–42.

Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., et al. (2013). Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS One* 8:e54985. doi: 10.1371/journal.pone.0054985

Sun, S., Li, C., Paterson, A. H., Jiang, Y., Xu, R., Robertson, J. S., et al. (2018). In-field high throughput phenotyping and cotton plant growth analysis using LiDAR. *Front. Plant Sci.* 9:16. doi: 10.3389/fpls.2018.00016

Sun, Z., Su, C., Yun, J., Jiang, Q., Wang, L., Wang, Y., et al. (2019). Genetic improvement of the shoot architecture and yield in soya bean plants via the manipulation of GmmiR156b. *Plant Biotechnol. J.* 17, 50–62. doi: 10.1111/pbi.12946

Tadesse, M., and Chris, M. (2007). Efficiency of mechanical harvest for immature vegetable soybean pods. *Va. J. Sci.* 58:2. doi: 10.25778/qej4-5k62

Teichmann, T., and Muhr, M. (2015). Shaping plant architecture. *Front. Plant Sci.* 6:233. doi: 10.3389/fpls.2015.00233

Tian, Z., Wang, X., Lee, R., Li, Y., Specht, J. E., Nelson, R. L., et al. (2010). Artificial selection for determinate growth habit in soybean. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8563–8568. doi: 10.1073/pnas.1000088107

Ulloa, S. M., Datta, A., Malidza, G., Leskovsekc, R., and Knezevic, S. Z. (2010). Yield and yield components of soybean [*Glycine max* (L.) Merr.] are influenced by the timing of broadcast flaming. *Field Crop Res.* 119, 348–354. doi: 10.1016/j.fcr.2010.08.006

Verri, A., Uras, C., Frosini, P., and Ferri, M. (1993). On the use of size functions for shape analysis. *Biol. Cybern.* 70, 99–107. doi: 10.1007/BF00200823

Wang, B., Smith, S. M., and Li, J. (2018). Genetic regulation of shoot architecture. *Annu. Rev. Plant Biol.* 69, 437–468. doi: 10.1146/annurev-arplant-042817-040422

William, S., and Aoyagi, A. (2009). *History of Edamame, Green Vegetable Soybeans, and Vegetable-Type Soybeans*. Lafayette, CA: Soyinfo Center.

Wong, C. E., Singh, M. B., and Bhalla, P. L. (2013). The dynamics of soybean leaf and shoot apical meristem transcriptome undergoing floral initiation process. *PLoS One* 8:e65319. doi: 10.1371/journal.pone.0065319

Xavier, A., Hall, B., Hearst, A. A., Cherkauer, K. A., and Rainey, K. M. (2017). Genetic architecture of phenomic-enabled canopy coverage in *Glycine max*. *Genetics* 206, 1081–1089. doi: 10.1534/genetics.116.198713

Yoshikawa, T., Ozawa, S., Sentoku, N., Itoh, J., Nagato, Y., and Yokoi, S. (2013). Change of shoot architecture during juvenile-to-adult phase transition in soybean. *Planta* 238, 229–237. doi: 10.1007/s00425-013-1895-z

Zandonadi, R., Stombaugh, T., Coolong, T., and Pfeiffer, T. (2010). *Mechanical Harvesting of Edamame. SARE Proj Final Rep*. Available online at: https://www.uky.edu/ccd/sites/www.uky.edu.ccd/files/edamame_mechanical_harvest.pdf (accessed February 7, 2021).

Zhang, D., Wang, X., Li, S., Wang, C., Gosney, M. J., Mickelbart, M. V., et al. (2019). A post-domestication mutation, Dt2, triggers systemic modification of divergent and convergent pathways modulating multiple agronomic traits in soybean. *Mol. Plant* 12, 1366–1382. doi: 10.1016/j.molp.2019.05.010

Zhang, H., Hao, D., Sitoe, H. M., Yin, Z., Hu, Z., Zhang, G., et al. (2015). Genetic dissection of the relationship between plant architecture and yield component traits in soybean (*Glycine max*) by association analysis across multiple environments. *Plant Breed.* 134, 564–572. doi: 10.1111/pbr.12305

Zhang, J., Song, Q., Cregan, P. B., Nelson, R. L., Wang, X., Wu, J., et al. (2015). Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics* 16:217. doi: 10.1186/s12864-015-1441-4

Zhang, L., and Kyei-Boahen, S. (2007). Growth and yield of vegetable soybean (Edamame) in Mississippi. *Horttechnology* 17, 26–31. doi: 10.21273/HORTTECH.17.1.26

Zhou, P., Hirsch, C. N., Briggs, S. P., and Springer, N. M. (2019). Dynamic patterns of gene expression additivity and regulatory variation throughout maize development. *Mol. Plant* 12, 410–425. doi: 10.1016/j.molp.2018.12.015

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33, 408–414. doi: 10.1038/nbt.3096

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Sustainable Food Systems

ORIGINAL RESEARCH
published: 31 August 2021
doi: 10.3389/fsufs.2021.651505

Investigating Consumer Demand and Willingness to Pay for Fresh, Local, Organic, and "On-the-Stalk" Edamame

Nick Lord[1], *Bo Zhang*[1] *and Clinton L. Neill*[2]*

[1] *School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, United States,* [2] *Department of Population Medicine and Diagnostic Sciences, Cornell University, Ithaca, NY, United States*

Introduction of locally adapted, commercially viable edamame varieties can allow it to be marketed as fresh, local, organic, or on the stalk. Here, we utilized a one-and-one-half bounded (OOHB) elicitation format to estimate mean willingness to pay (WTP) for these external attributes in relation to a vector of explanatory variables. Results showed 84-, 85-, and 28-cent premiums for fresh, local, and organic edamame (10 oz). Pro-environmental attitudes drove WTP for all three of these attributes, while shopping location significantly increased mean WTP for fresh and organic attributes. A 40-cent price discount was observed for the "on-the-stalk" attribute, suggesting that convenience also plays an important role in marketing edamame. The results suggest that more research regarding edamame demand is warranted.

Keywords: consumer preference, economics, willingness to pay, edamame, organic, local, one-and-one-half bound dichotomous choice

OPEN ACCESS

Edited by:
Giuseppe Di Vita,
University of Turin, Italy

Reviewed by:
Daniela Spina,
University of Catania, Italy
Timothy Woods,
University of Kentucky, United States

Correspondence:
Clinton L. Neill
cln64@cornell.edu

Specialty section:
This article was submitted to Nutrition and Sustainable Diets, a section of the journal Frontiers in Sustainable Food Systems

Received: *10 January 2021*
Accepted: *31 July 2021*
Published: *31 August 2021*

Citation:
Lord N, Zhang B and Neill CL (2021) Investigating Consumer Demand and Willingness to Pay for Fresh, Local, Organic, and "On-the-Stalk" Edamame. Front. Sustain. Food Syst. 5:651505. doi: 10.3389/fsufs.2021.651505

INTRODUCTION

As Americans become more health and/or environmentally conscious, there is a growing desire among consumers to reduce meat consumption, which has led to noticeable growth in demand for alternative, plant-based sources of protein (Bashi et al., 2019). In addition, the disproportionately large and rapid growth of Asian populations in the United States promises to create new market opportunities for specialty Asian produce (Govindasamy et al., 2010). Vegetable soybean, more commonly referred to as edamame (pronounced "eh-duh-MAH-may"), has quickly emerged as a prime candidate to capitalize on both of these domestic trends, appearing more and more frequently in salad bars and sushi restaurants nationwide.

Edamame has a long history of consumption in East Asia (Shurtleff and Aoyagi, 2009) where the industry has become well-established. It was not until the turn of the 21st century that it began to be imported to U.S. markets. In order to meet domestic demand, roughly 25,000 to 30,000 tons of edamame is consumed annually (CBS News, 2013), which is predominantly met through frozen imports sold to consumers year-round in grocery stores. In the United States, edamame is most commonly supplied as pods, where consumers suck the beans out of the pod. Edamame can also be supplied as shelled beans that have already been removed from the pods.

Soybean proteins contain all essential amino acids that are required for human growth and development, which earned them the designation of being a high-quality protein source (Michelfelder, 2009). In addition, several studies have suggested that soy consumption can reduce levels of low-density lipoprotein (LDL) cholesterol, or "bad" cholesterol, in the body, which can build up in arteries and increase risk of cardiovascular disease (Hasler, 2002; Taku et al., 2007). As legumes, soybeans also exploit biological nitrogen fixation, which reduces the need for synthetic nitrogen application and is beneficial for soil health (Beyan et al., 2018). This unique combination of nutritional and environmentally friendly characteristics has allowed edamame to carve out a distinguished position in the domestic marketplace for vegetables.

Edamame's growing popularity has sparked interest in bolstering U.S. edamame production to reduce reliance on imports and to allow domestic growers to capitalize on edamame's premium market position. In the coming years, plant breeders are set to release seed inputs adapted to U.S. production regions with higher yield potential, enhanced sensory quality, and increased suitability for mechanized harvest in hopes of catalyzing domestic production (Lord et al., 2019b). The provision of locally adapted seed inputs can lead to many new edamame product options in the domestic marketplace.

With more edamame grown on U.S. soil, edamame producers and distributors can seek to access premiums associated with quality-differentiated external attributes such as being freshly supplied, locally grown, and USDA certified organic or market it as an alternative end product. As such, this study seeks to understand what marketing advantage, if any, these attributes hold over the frozen, imported, and non-GMO products that are currently available to consumers today.

LITERATURE REVIEW

Consumption of fresh produce in the United States has considerably grown as selection and quality of fresh vegetables in the marketplace have increased (Pollack, 2001). Consumer preference for fresh produce can be largely attributed to perceived losses in nutritional quality from blanching and freezing that is associated with frozen produce (Heinrichs, 2016). In the case of edamame, several studies have suggested relatively low production costs and the potential profitability of fresh market edamame (Shockley et al., 2011; Sharma, 2013; Garber and Neill, 2019; Lord et al., 2019a). However, it is unclear how much more consumers would value a fresh, seasonal edamame product over the frozen edamame available in most grocery stores.

The local food market continues to occupy an increasingly important share of agricultural sales in the United States. Though "local food" remains a loosely defined concept, it generally refers to food produced within close geographical proximity to consumers and which also contains certain social and supply chain distinctions related to production practice, environmental impact, food safety, fair labor practices, and animal welfare (Martinez, 2010; Holcomb et al., 2018; Neill et al., 2020). Numerous studies have demonstrated that consumers are willing to pay price premiums for food labeled as locally grown, which can be primarily attributed to perceived improvements in quality, nutrition, and value for price, as well as support for local economies, concern over environmental impact, and demographic characteristics (Brown, 2003; Carpio and Isengildina-Massa, 2009; Martinez, 2010; Wang et al., 2010; Feldmann and Hamm, 2015; Neill and Williams, 2016). Moreover, a recent study by Fan et al. (2019) suggests that these premiums may be irrespective of actual quality or flavor. Edamame is different than most of the produce items that have been previously studied for the local attribute, given its origins in East Asia and the predominance of imports used to meet U.S. demand. To date, no studies have yet been conducted to understand how consumers would value a locally grown edamame product compared to the non-local, imported edamame products that are already available to consumers.

At the same time, the purchase of organic vegetables has drastically risen in the United States over the past few decades, with its consumer base becoming increasingly diverse (Stevens-Garmon et al., 2007; Dettmann and Dimitri, 2009). In 2003, the USDA established the National Organic Program (NOP), which regulates organic marketing by certifying that products are produced in accordance with a set of approved inputs and substances that are consistent with regenerative and sustainable agricultural principles. As a result of the myriad provisions guaranteed by the NOP and the additional cost to obtain certification, organic produce often commands substantial premiums in the marketplace. Interestingly, evidence suggests that many consumers may not necessarily value all provisions of the USDA Organic label equally. Some, for example, are only willing to pay more for the guaranteed use of non-GMO seed inputs but are not necessarily concerned about which synthetic fertilizers and pesticides are used during production. To test this hypothesis, Bernard et al. (2006) conducted a study intended to mimic market conditions by observing subjects in an experimental auction setting where they were asked to indicate their willingness to pay (WTP) for various food items that fall under conventional, organic, and non-GMO categories. While overall results showed the highest WTP estimates for food items in the organic category, further analysis revealed that beyond the non-GMO attribute, subjects did not appear willing to pay significantly extra for the remaining attributes of the organic category. No studies have yet explored consumer WTP for USDA-certified organic edamame; however, a recent study by Wolfe et al. (2018) demonstrated that non-GMO edamame may already hold some appeal to consumers with risk aversion to GMO products and that this may even compensate for shortcomings in sensory quality. To this end, a gap exists in the understanding of how consumer WTP for organic edamame is affected by the presence of non-GMO edamame, which is already common in the marketplace. Such findings are crucial to more completely understand the potential of USDA-certified organic edamame in the United States.

Growing consumer inclination for convenience has become increasingly important in consumer food purchasing decisions in recent years (Pollack, 2001; Brunner et al., 2010). The term convenience encompasses time, physical, and mental effort associated with meal preparation and has been shown to be positively correlated with a number of socioeconomic and attitudinal variables including household size, working status, and cooking enjoyment among others (Candel, 2001; Brunner et al., 2010). Interestingly, Lockie et al. (2004) found that concern for convenience was negatively correlated with consumption of organic food, of which naturalness of the product was a particularly strong predictor. These findings were later corroborated by Brunner et al. (2010), who found that naturalness was a consistently strong, negative predictor for consumption of highly processed, minimally processed, and single-component food categories. "On-the-stalk" edamame is an ideal candidate to study the competing influence of naturalness with convenience on consumer WTP. In whole plant or "on-the-stalk" end product, pods are still attached to the branches or stalks of the plant, making for an end product that appears considerably more natural (**Figure 1**). Marketing edamame as an "on-the-stalk" end product is more common in Japan, where many consumers actually prefer it supplied in this way for its perceived added freshness (Born, 2006). By observing consumer WTP for "on-the-stalk" edamame, which decreases convenience by requiring consumers to pick the pods off of the stalks themselves, in comparison to pre-stripped edamame pods that require much less preparation time and are already common in U.S. markets, we can also gain insight into the role that convenience plays in consumer WTP for edamame.

Other studies have shown that social–altruistic and biospheric value bases have translated to higher WTP for regional and organic products (Umberger et al., 2009; Rahman and Reynolds, 2017; Shin et al., 2017). In addition, Onozaka et al. (2011) contend that shopping location may also influence consumer valuation for various labeling campaigns by indirectly sorting consumers based on the different sustainability claims made in-store or within the market venue. Moreover, the hierarchical nature of consumer preferences is an important aspect of the demand issue that often leads to clustering of consumers (Di Vita et al., 2021). Information on how these factors, as well as many other explanatory variables relating to demographic information, dietary habits, and personal beliefs, affect consumer WTP also remains poorly understood. As such, the goal of this work was to use contingent valuation (CV) methodology to (1) estimate mean WTP for edamame marketed as fresh, local, USDA Organic, and "on the stalk" and (2) identify significant explanatory variables that influence these estimates in order to better understand the potential of alternatively marketed edamame in the United States.

DATA AND METHODS

CV is a flexible survey technique used to estimate WTP for non-market goods and services (Lopez-Feldman, 2012). It has been frequently exploited in the literature to forecast success or demand for various agricultural products, ecolabels, and marketing trends (Carpio and Isengildina-Massa, 2009; Haghiri et al., 2009; Owusu and Owusu Anifori, 2013; Neill and Williams, 2016). CV methodology can be particularly useful in exploring price premiums and WTP for products or product attributes that have yet to appear in the marketplace. This is achieved by presenting participants with a hypothetical scenario through which to observe their purchasing decisions when both the "status-quo" and newly proposed product or attribute of interest are present. CV methods are also ideal for consumer intercept surveys, like this study, to keep questionnaires short when presented on paper. While this does risk some hypothetical bias, Penn and Hu (2018) noted that the bias is not as large as that present in auction methods. However, a choice experiment or referendum experiment would likely have reduced hypothetical bias further.

While edamame is already in the market, the specific attributes inquired about via our study are not overtly available. In particular, edamame is commonly found as frozen, in pods or shelled bean form, and often imported. As much as 80% of edamame is imported to meet a growing domestic demand, which means that the majority is frozen to remain edible (Bernick, 2009; Wolfe et al., 2018). This also means that it is rarely found "on the stalk," fresh, or local. While some edamame is sold as USDA certified organic, it is unknown if this marketing label is preferred over the alternative of non-GMO. All edamame is produced as non-GMO as there are no commercial GMO varieties. In addition, domestic edamame production is on the rise, particularly in Arkansas, given the proximity to proper processing facilities (Jaeger, 2019).

During the month of October 2018, consumer intercept CV surveys (on paper as consumers entered the store) were distributed at three primary locations in Blacksburg, VA—Oasis International Supermarket, the Blacksburg Farmer's Market,

FIGURE 1 | Pictures of edamame end products. Edamame supplied as pre-stripped pods is predominantly consumed in the United States today, while edamame supplied as "on the stalk" (right) is more commonly preferred in East Asia for its added freshness.

1. What is your current age? Under 18 18 to 30 30 to 45 46 to 50 Older than 60

2. What gender do you identify with? Male Female Other

3. What is your ethnicity?

Caucasian African American Asian Hispanic Other

4. Are you the main shopper for your household? Yes No

5. What is your current household family size, including yourself?

1 2 3 4 More than 4

6. How many meals do you eat at home per week?

1 to 5 6 to 10 11 to 15 16 or more

7. What is your current level of education?

Some school High school diploma Some college Bachelor's degree Advanced degree

8. What is your household income?

Less than 20,000 20,000-35,000 35,001-50,000 50,001-70,000 70,001-100,000 100,000+

9. How many days a week do you consume meat (including fish)?

0 1 2 3 4 5 6 7

10. Do you have any desire to reduce your meat consumption? Yes No

10a. If you answered "yes" to question 10, please check which reasons you desire to reduce meat consumption: ___ health ___ religious reasons ___reduce environmental impact ___other

11. Are you full/partially vegan or vegetarian? Yes No

12. On a scale of 1 to 5 with 1 being never and 5 being very often/always, how often do you purchase food that is labeled organic, free range/cage free, Rainforest Alliance, Wild Caught, natural, or non-GMO? 1 2 3 4 5

FIGURE 2 | Example of demographic questions on survey instrument.

and the Virginia Tech Squires Student Center. Other survey locations were also utilized but in much smaller quantities and are classified as "other." Surveys consisted of 20 questions each. The first 16 questions pertained to demographic information and characteristics of participants relating to purchasing habits, environmental attitudes, dietary habits, and familiarity with edamame (**Figure 2**). The location of survey was included as a categorical variable in order to also observe the effect of market venue. Responses to these questions were used to identify important explanatory variables of mean WTP.

The remaining four questions were used to elicit WTP using dichotomous-choice (DC) questions. Specifically, consumers evaluated fresh vs. frozen, local vs. imported, organic vs. non-GMO, and "on the stalk" vs. pods of edamame in the four DC questions. This can also be seen in **Table 1**. Each DC question was asked independently of each other. More importantly, as previously mentioned, each one had a specific and status quo option relevant to the attribute of interest—fresh vs. frozen, local vs. non-local,[1] organic vs. non-GMO, and on the stalk vs. pods. An example of the one-and-one-half bounded (OOHB)-DC questions can be seen in **Figure 3**. In these questions, respondents must choose between purchasing the "status quo" option, which is always listed at the "fair" price (in this case \$3.50 per 10 oz of pods) or the new or alternatively marketed product of the same quantity when it is offered at a premium. In order to get information on multiple premiums, three survey versions were randomly administered, each only differing in the extent of the premium for the alternative product. One of the three versions

[1]Note that we did not define local for the participants. This could have caused additional hypothetical bias, but we did not want to define a geographical boundary given that all of the participants were from one particular town.

TABLE 1 | Alternatives in each of the OOHB-DC CV questions.

Question	Attribute of interest	Status quo option
1	Fresh	Frozen
2	Local	Imported
3	Organic	Non-GMO
4	On the stalk	Pods

also offered the alternative product at a lower price than the status quo to see if participants actually discounted the alternative product. In total, premiums of 0.25, 0.50, and 1.00 and a discount of 0.25 were observed. For each survey version, all four DC questions had the same premium/discount associated with the attribute. For example, the same participant was asked to choose between fresh and frozen edamame at a premium of $0.50 and was asked their choice between non-GMO and organic edamame also at a premium of $0.50.

DC questions are regarded as the most adequate, reliable, and heavily utilized CV elicitation technique used in the literature, given their increased statistical efficiency and reduced response bias over other CV elicitation methods such as open-ended questions and payment cards (Venkatachalam, 2004; Lopez-Feldman, 2012). As opposed to DC questions, single-bounded questions consisting of a single "take-it-or-leave-it" question simply ask the individual whether they would accept ($y_i = 1$) or reject ($y_i = 0$) a bid for the product/attribute of interest. To increase the statistical efficiency of the single-bounded estimation, Hanemann et al. (1991) later proposed the addition of a follow-up question dependent on the response to the initial question. Responses to initial and follow-up questions can be captured by dichotomous variables y_i^1 and y_i^2, respectively. If the individual selects "no" to the initial bid ($y_i^1 = 0$), a lower bid is offered, whereas if the individual selects "yes" ($y_i^1 = 1$), then a higher bid is offered. Despite its increased statistical efficiency, discrepancies observed between the initial and follow-up questions of the DB format have caused concern over its consistency and reproducibility. The OOHB question format was thus developed to address this response bias while maintaining statistical efficiency (Cooper et al., 2002). In OOHB-DC questions, respondents are only asked to answer a follow-up if they select "yes" for the bid listed in the initial question. Responses to OOHB-DC questions can therefore result in one of the following three scenarios: no–no, yes–no, or yes–yes.

An interval data model was used to estimate individual WTP as a function of explanatory variables and error. The probability that the WTP falls between specific minimum and maximum bounds can be estimated using a log-likelihood function. To understand the econometric estimation behind the probabilities for each of the aforementioned scenarios, let us consider the following set of equations, taken from Neill and Williams (2016):

$$WTP_i\left(z_i,\, u_i\right) = z_i\beta + u_i$$
$$\Pr\left(y_i^1 = 1,\, y_i^2 = 0 \middle| z_i\right) = Pr(s, n)$$

In equation 1, we assume WTP is a function of an individual's characteristics (demographics, environmental attitudes, diet, etc.) plus error, where z_i represents a vector of explanatory variables, β represents a vector of coefficients for the explanatory variables, and u_i represents an error term. Equation (2) represents a simplified notation for the probability of an individual answering "yes" to the initial question and "no" to the follow-up, dependent on the vector of explanatory variables. Under the assumptions of Equation 1 and that our data are normally distributed, we can obtain the following:

$$\begin{aligned}\Pr(s,n) &= \Pr\left(t^1 \leq WTP < t^2\right)\\ &= \Pr\left(t^1 \leq z_i^{'}\beta + u_i < t^2\right)\\ &= Pr\left(\frac{t^1 - z_i^{'}\beta}{\sigma} \leq \frac{u_i}{\sigma} < \frac{t^2 - z_i^{'}\beta}{\sigma}\right)\\ &= \varphi\left(\frac{t^2 - z_i^{'}\beta}{\sigma}\right) - \varphi\left(\frac{t^1 - z_i^{'}\beta}{\sigma}\right)\end{aligned}$$

where t^1 represents the suggested bid for the initial question and t^2 represents the suggested bid for the follow-up question. Given that Equation 6 follows $\Pr\left(a \leq X < b\right) = F\left(b\right) - F(a)$, using symmetry of the normal distribution, we can rearrange it once more to get the following:

$$\Pr(s,n) = \varphi\left(z_i^{'}\frac{\beta}{\sigma} - \frac{t^1}{\sigma}\right) - \varphi\left(z_i^{'}\frac{\beta}{\sigma} - \frac{t^2}{\sigma}\right)$$

The other two conditions can be similarly derived by replacing $t^1 \leq WTP < t^2$ with $t^2 \leq WTP < \infty$ for the yes–yes scenario and $0 \leq WTP < t^1$ for the no scenario. Taken together, these scenarios comprise the censored likelihood function shown below:

$$\sum_{i=1}^{N}\left[d_i^{sn}\ln\left(\varphi\left(z_i^{'}\frac{\beta}{\sigma} - \frac{t^1}{\sigma}\right) - \varphi\left(z_i^{'}\frac{\beta}{\sigma} - \frac{t^2}{\sigma}\right)\right)\right.$$
$$\left. + d_i^{ss}\ln\left(\varphi\left(z_i^{'}\frac{\beta}{\sigma} - \frac{t^2}{\sigma}\right)\right) + d_i^{nn}\ln\left(1 - \varphi\left(z_i^{'}\frac{\beta}{\sigma} - \frac{t^2}{\sigma}\right)\right)\right]$$

where d_i^{sn}, d_i^{ss}, and d_i^{nn} represent indicator variables for yes–no, yes–yes, and no–no, respectively. Depending on how the individual responds to the DC questions, the non-relevant indicator variables will take the value of zero, allowing only the relevant case to contribute to the likelihood function at any given time.

Censored regression analysis to estimate mean WTP for each of the four external attributes was conducted in R using the DCchoice package. The DCchoice package contains an oohbchoice function, which was specifically developed to execute maximum likelihood estimation on OOHB-DC data based on a number of required and optional arguments that the researcher specifies, such as the formula, distribution, and omission of missing data (Nakatani et al., 2016). The output is similar to that which is generated from the LIFEREG procedure on SAS v. 9.3, where parameter estimates can be directly interpreted as

17. Under the following prices, which edamame product would you choose?

Non-GMO edamame pods at $3.50 (10 oz. bag) USDA organic edamame pods at $3.75 (10 oz. bag)

17a. If you answered USDA organic edamame pods, choose which edamame product you would rather purchase:

Non-GMO edamame pods at $3.50 (10 oz. bag)

USDA organic edamame pods at $4.00 (10 oz. bag)

17b. If you answered Non-GMO edamame pods, would you be willing to pay anything above $3.50 for USDA Organic edamame pods? Yes No

FIGURE 3 | Example of OOHB-DC questions used in survey instrument.

changes in the marginal WTP (Neill and Williams, 2016). The three primary survey locations were coded into the model, while the other locations were used as the reference point. Mean WTP is calculated from each of the regressions based on the normal density function and calculated using marginal effects as in Neill and Holcomb (2019):

$$E\left(WTP\right)=\Phi\left(\frac{\boldsymbol{x}^{'}\beta}{\sigma}\right)\left[\boldsymbol{x}^{'}\beta+\sigma\left(\frac{\phi\left(\boldsymbol{x}^{'}\beta\right)}{\Phi\left(\boldsymbol{x}^{'}\beta\right)}\right)\right].$$

A total of 222 surveys were collected, with 188 completed responses. While this is a small sample size as compared to many other consumer WTP studies, our sample has significant variation in education, income, and gender. For the OOHB-DC questions, only complete sets of responses (yes–no, yes–yes, or no–no) were used for the interval regression. A complete table of summary statistics can be seen in **Table 2**. Approximately 69% of the respondents were between the ages of 18 and 30, and 34% of the respondents reported a household income below $20,000, which can be attributed to the survey being distributed in a small college town where students are more prevalent in the community. The vast majority of respondents had received at least some level of secondary schooling, with 42% having taken some school, 25% having received a bachelor's degree, and 26% having received some sort of advanced or postgraduate degree.[2] About 48% of the population were female. The majority of respondents were either Caucasian (57%) or Asian (33%). Approximately 44% of respondents indicated that they had a desire to reduce their meat consumption, of which 78% indicated health as a reason why and 56% indicated environment as a reason why. In regard to familiarity with edamame, ∼55% indicated that they were mostly (20%) to extremely familiar (35%) while another 42% indicated that they consume edamame at least once per week.

[2]Note that we test restricted versions of the model without education to determine if there is a high correlation between the variable and income. This was done via likelihood ratio tests, and a significant reduction in explanatory power was found.

RESULTS AND DISCUSSION

WTP estimates derived from the censored regression for all four attributes can be seen in **Table 3**. The interval regression estimated a mean WTP of ∼$4.34 per 10 oz for the fresh attribute. In other words, survey respondents indicated that they were willing to pay up to 84 cents more for edamame pods available fresh as opposed to frozen on average. The variables *Female* and *Likelihood to shop local*, significant at the 1% and 5% levels, respectively, both exhibited a negative effect on mean WTP for the fresh attribute. If the survey respondent was female, the average discount associated with fresh edamame is about 71 cents. This observation may be related to public uncertainty surrounding the safety of phytohormones in soy products for women's health (White et al., 2000; Duffy et al., 2007; Bar-El and Reifen, 2010; Cederroth et al., 2012), which are present in lower levels in frozen edamame products (Simonne et al., 2000). Meanwhile, the more likely a respondent was to shop local, the less they were willing to pay for a fresh edamame product. In other words, for a one-point increase in likelihood to shop local as indicated by respondents via a Likert scale, consumers discount the fresh edamame product by ∼34 cents.

A considerable price discount was observed for edamame supplied as an on-the-stalk product. When given the choice between edamame supplied as pods and those supplied on the stalk, survey respondents were not willing to pay any more than $3.10 per 10 oz. None of the covariates were significant in this case. For the certified organic attribute, only a small premium was observed; survey respondents indicated a maximum premium of 28 cents more for an edamame product labeled as USDA certified organic if an equivalent non-GMO product was also available. That is to say, beyond the non-GMO guarantee of their edamame product, respondents were not willing to pay appreciably more for the other provisions encompassed under the USDA-certified label. The only statistically significant variable was the participants' likelihood to purchase eco-labeled products.

The mean WTP estimate for the local attribute was similar to the estimate for the fresh attribute. When given a choice between local and non-local edamame pods, survey respondents indicated that they would be willing to pay an average of up to 85 cents

TABLE 2 | Summary statistics for survey respondents.

Variable	Description	Percentage of occurrence	Mean	Standard deviation
Location	Location of the survey			
	1 = Farmer's market	31.08%		
	2 = Supermarket	16.67%		
	3 = Student center	18.02%		
	4 = Other	34.23%		
Age	Age of participant			
	1 = Under 18	0.90%	2.6347	1.213
	2 = 18–30	67.57%		
	3 = 30–45	13.06%		
	4 = 46–50	4.50%		
	5 = 51–60	6.31%		
	6 = Older than 60	5.86%		
Gender	Dummy variable			
	0 = Male	50.45%	0.4886	0.501
	1 = Female	48.20%		
Ethnicity	Categorical variable:			
	1 = Caucasian	56.68%		
	2 = African American	2.76%		
	3 = Asian	33.18%		
	4 = Hispanic	4.15%		
	5 = Other	3.23%		
Main shopper for household	Dummy variable			
	0 = No	36.92%	0.6308	0.4837
	1 = Yes	63.08%		
Household size	Number of people currently living in household:			
	1 = 1	32.57%	2.6284	1.4669
	2 = 2	20.18%		
	3 = 3	13.76%		
	4 = 4	18.81%		
	5 = More than 4	14.68%		
Home food consumption	Number of meals consumed at home per week:			
	1 = 1–5	16.89%	2.5707	0.9992
	2 = 6–10	29.68%		
	3 = 11–15	32.88%		
	4 = 16 or more	20.55%		
Education	Highest level of education completed:			
	1 = Some school	1.39%	3.6497	1.0031
	2 = High school diploma	6.48%		
	3 = Some college	41.67%		
	4 = Bachelor's degree	25.00%		
	5 = Advanced degree	25.46%		
Household income	Household income levels:			
	1 = <$20,000	34.91%	3.1173	2.0375
	2 = $20,000–$35,000	16.04%		
	3 = $35,001–$50,000	9.43%		
	4 = $50,001–$70,000	6.13%		
	5 = $70,001–$100,000	10.85%		
	6 = More than $100,000	22.64%		

(Continued)

TABLE 2 | Continued

Variable	Description	Percentage of occurrence	Mean	Standard deviation
Meat consumption	How many days per week meat is consumed:			
	0 = 0	10.60%	4.7981	4.4932
	1 = 1	2.76%		
	2 = 2	8.29%		
	3 = 3	10.14%		
	4 = 4	13.36%		
	5 = 5	10.14%		
	6 = 6	12.90%		
	7 = 7	31.80%		
Desire to reduce meat consumption	Dummy variable			
	0 = No	56.19%	0.238	0.4364
	1 = Yes	43.81%		
Reason for desire to reduce meat consumption	Health	78.26%		
	Religion	3.26%		
	Environment	56.52%		
	Other	8.70%		
Fully/partially vegan or vegetarian	Dummy variable			
	0 = No	79.53%	0.2047	0.4044
	1 = Yes	20.47%		
Environmental value	How often the consumer purchases eco-labeled products (Likert scale 1–5):			
	1 = Never	9.30%	3.1116	1.1983
	2 = Rarely	22.79%		
	3 = Some times	31.16%		
	4 = Somewhat frequently	20.93%		
	5 = Always	15.81%		
Edamame familiarity	Familiarity with edamame (Likert scale 1–5):			
	1 = Not familiar at all	14.02%	3.4672	1.4716
	2 = Slightly familiar	17.29%		
	3 = Somewhat familiar	12.62%		
	4 = Mostly familiar	20.09%		
	5 = Extremely familiar	35.98%		
Edamame consumption	Frequency of edamame consumption (Likert scale 1–5)			
	1 = Never	27.44%	2.8372	1.2367
	2 = Once a month	3.26%		
	3 = Twice a month	27.44%		
	4 = Once a week	41.86%		
	5 = Multiple times per week	0.00%		
Buying local	Likeliness to purchase local (Likert scale 1–5):			
	1 = Never	11.57%	2.9167	1.1743
	2 = Rarely	27.31%		
	3 = Somewhat often	30.09%		
	4 = Very often	19.91%		
	5 = Always	11.11%		

more for the local option. It should be noted that no formal definition of "local" regarding vicinity of production, state, or region was provided as part of the study. Nevertheless, a high mean WTP was observed. The *Likelihood to purchase eco-labeled products* variable appeared to have a significant effect on mean WTP for the local attribute, suggesting relatively broad interest in a local edamame product among our survey sample. There was also a negative WTP for both White and Asian participants as

TABLE 3 | WTP estimates for edamame marketed as fresh, local, organic, and on the stalk ($N = 188$).

Parameter	Fresh	Local	Organic	On-the-stalk
Mean WTP (per 10 oz)	4.34	4.35	3.78	3.10
Intercept	4.118**	6.192**	4.445**	3.167**
Location—Farmer's market	0.856**	0.319	−0.051	−0.067
Location—Supermarket	1.221**	0.289	0.608*	0.105
Location—Student center	0.292	−0.368	−0.338	−0.180
Age	0.007	0.223	0.081	−0.006
Female	−0.708**	0.054	−0.105	−0.344
White	−0.103	−1.011**	0.388	0.620
Asian	0.043	−1.672**	−0.102	0.612
Shopper	−0.345	0.106	0.308	−0.007
Household size	0.041	0.036	0.062	0.058
Meals consumed at home	0.070	−0.107	0.155	0.096
Education	0.079	−0.029	−0.012	−0.078
Household income	−0.083	0.009	−0.040	−0.061
Meat consumption frequency	0.009	−0.005	−0.003	−0.043
Desire to decrease meat consumption	0.002	0.204	0.044	−0.079
Vegan/vegetarian	−0.013	0.206	−0.389	−0.570
Purchase of eco-labeled products	0.336**	0.285**	0.320**	0.128
Edamame consumption frequency	−0.133	−0.103	0.061	0.078
Likelihood to shop local	−0.260*	0.082	0.024	−0.123
BID	−0.990**	−1.487**	−1.737**	−1.070**
Log likelihood	−149.37	−124.37	−167.85	−142.95

*Asterisks "**" and "*" indicate statistical significance at 1% and 5% levels, respectively.*

compared to the "other" category. Asian participants discounted local more heavily as they likely have a higher preference for products imported from an Asian country where edamame originates and is commonly imported. This could also be due to consumers questioning the quality of locally produced edamame. Since they were not tasting local vs. imported edamame, there was likely some preconceived notion that domestically produced edamame does not meet the same taste/nutritional quality.

The variable *Likelihood to purchase eco-labeled products* was used in this study to gauge pro-environmental attitudes of the survey respondents. By including it as a parameter, our intention is to determine if pro-environmental attitudes have an impact on consumer WTP for the various attributes. In our results, we observe that *Likelihood to purchase eco-labeled products* consistently appeared to be a significant driver of mean WTP for three out of the four attributes studied. For each unit increase in pro-environmental attitudes indicated on the Likert scale, mean WTP increased by 32 cents for fresh edamame, 27 cents for local edamame, and 33 cents for USDA-certified organic edamame. The *Likelihood to purchase eco-labeled products* variable did not show statistical significance for the on-the-stalk attribute, which may be due to the fact that the on-the-stalk attribute is more related to physical appearance and, as a result, does not hold as many environmental implications as the other three attributes. These results follow a stream of literature suggesting that pro-environmental attitudes and beliefs continue to play an increasingly important role in consumer WTP for products (Umberger et al., 2009; Lusk et al., 2014; Neill and Williams, 2016; Rahman and Reynolds, 2017; Shin et al., 2017).

Survey location was a notable predictor of mean WTP observed for the fresh attribute in this study. In total, all surveys collected fell into one of four categories: those collected at the farmer's market, those collected at the local supermarket, those collected at the student center, and those collected from miscellaneous individuals around campus (henceforth referred to as "Misc. group"). We included the two purchasing locations (farmer's market and supermarket) and one non-purchasing location (student center) in our analysis and interpreted the results as they related to the Misc. group. Results showed that if a respondent completed the survey at the farmer's market, the mean WTP for the fresh attribute was 86 cents higher as compared to that in the Misc. group. Meanwhile, the mean WTP for surveys completed at the international supermarket was $1.22 higher for the fresh attribute. For all attributes studied, the mean WTP for surveys collected at the student center—where no produce is sold and no sustainability claims are therefore made—failed to show significance. Here, we can clearly see that whether or not the survey was collected from a purchasing location largely influenced the premiums observed for the fresh attribute. Furthermore, the type of purchasing location (farmer's market vs. local supermarket) appeared to drastically impact the extent of the premiums observed for each of these attributes. While it is a bit surprising that we did not find location to be significant when it came to WTP for local edamame, this could be due to the fact that we did not define the geography to which being local represented. As Holcomb et al. (2018) noted, consumer preferences for local products are heavily dependent on how "local" is defined. A positive premium for organic edamame was found at the supermarket location.

CONCLUSIONS

With the hope of increased domestic production of edamame slowly becoming a reality, it is important to understand what factors can favor successful marketing efforts in the United States. In this study, we used the CV methodology to estimate the mean WTP for edamame marketed as fresh, local, and certified organic or as an on-the-stalk end product and then related these estimates to information on demographic information, dietary habits, and personal beliefs. Our results showed that fresh and local edamame held a significant marketing advantage over frozen products and non-local products, respectively. We also observed that despite its myriad of environmental provisions, edamame labeled as USDA certified organic may offer only limited marketing benefit to distributors if consumers already know that the edamame product that they are purchasing is non-GMO. This smaller premium must then be assessed at the cost of obtaining the organic certification. If the marginal cost of obtaining the organic certification is lower than the premium, then it would benefit the producer to pursue certification. In regard to end products, the convenience of pre-stripped edamame pods appears to take priority over the naturalness and freshness offered by an on-the-stalk or whole plant edamame product.

Several factors are important for understanding the domestic potential of alternatively marketed edamame. First, WTP for fresh edamame appears to be particularly influenced by demographic factors, especially gender. Meanwhile, personal beliefs such as pro-environmental attitudes appear to more consistently drive premiums for fresh, local, and organic edamame. This study also provides major evidence for the sorting effect that market venues have on consumer valuation of external attributes, especially for fresh edamame. These findings also highlight the importance of accounting for potential bias that can occur when collecting consumer intercept surveys from purchasing and non-purchasing locations.

Results from this study present important *preliminary* findings regarding the potential of alternatively marketed edamame in the domestic marketplace. Overall, these results support the continued growth of the domestic edamame industry. As with many WTP studies, our assertions are based on a single geographic location using a limited sample population, which may not be generally true in other states or regions across the United States. Therefore, future research should aim to substantiate these findings in other locations using different populations before producers and marketers make decisions. Given the hypothetical bias inherent in the CV methodology, non-hypothetical valuation methods may also serve to improve upon the results of this study by more accurately modeling consumer valuation of edamame based on real consequences. Another limitation of this study is that those who participated in the survey may have been predisposed to have a preference for edamame. While a large portion of our sample did not regularly consume edamame, it is still a concern that can limit the generalizability of the results. Future work could also focus on the combination of the attributes in this study. The goal of this study was to identify the individual attributes of interest, but the combination of attributes may have non-linear effects on WTP. As edamame variety development continues, there may be additional opportunities to study consumer WTP for edamame on the basis of sensory characteristics such as quality, flavor profile, and appearance as well. While further work is needed to validate the WTP values extracted from our sample, the information garnered from this study does serve as a reference for future studies on domestic edamame demand.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Virginia Tech Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NL drafted the manuscript and edited based on co-author comments. All authors reviewed, edited, and approved the draft and submitted versions of the manuscript.

FUNDING

This work was funded, in part, by USDA-NIFA, Grant No. 2018-51181-28384, Accession No. 1016465, and the Virginia Agricultural Experiment Station.

ACKNOWLEDGMENTS

The authors thank USDA-NIFA and Virginia Agricultural Experiment Station for the financial support.

REFERENCES

Bar-El, D. S., and Reifen, R. (2010). Soy as an endocrine disruptor: cause for caution? *J. Pediatr. Endocrinol. Metab.* 23, 855–861. doi: 10.1515/jpem.2010.138

Bashi, Z., McCullough, R., Ong, L., and Ramirez, M. (2019). *The market for alternative protein: Pea protein, cultured meat, and more | McKinsey*. New York, NY: McKinsey and Company. Available online at: https://www.mckinsey.com/industries/agriculture/our-insights/alternative-proteins-the-race-for-market-share-is-on (accessed January 5, 2021).

Bernard, J. C., Zhang, C., and Gifford, K. (2006). An experimental investigation of consumer willingness to pay for non-gm foods when an organic option is present. *Agric. Resour. Econ. Rev.* 35, 1–12. doi: 10.1017/S1068280500006808

Bernick, K. (2009). *Edamame Takes Root in U.S. Corn and Soybean Digest. 24 Feb. 2018*. Available online at: http://cornandsoybeandigest.com/edamame-takes-root-us (accessed January 5, 2021).

Beyan, S. M., Wolde-meskel, E., and Dakora, F. D. (2018). An assessment of plant growth and N2 fixation in soybean genotypes grown in uninoculated soils collected from different locations in Ethiopia. *Symbiosis (Philadelphia, PA)* 75, 189–203. doi: 10.1007/s13199-018-0540-9

Born, H. (2006). *Edamame: Vegetable soybean. National Center for Appropriate Technology (NCAT) a Publication of ATTRA-National Sustainable Agriculture Information Service*. Paul Driscoll (Ed.) Available online at: http://Attra.Ncat.Org/Attrapub/PDF/Edamame.Pdf (accessed June 25, 2009).

Brown, C. (2003). Consumers' preferences for locally produced food: a study in southeast Missouri. *Am. J. Altern. Agric.* 18, 213–224. doi: 10.1079/AJAA200353

Brunner, T. A., van der Horst, K., and Siegrist, M. (2010). Convenience food products. Drivers for consumption. *Appetite* 55, 498–506. doi: 10.1016/j.appet.2010.08.017

Candel, M. J. (2001). Consumers' convenience orientation towards meal preparation: conceptualization and measurement. *Appetite* 36, 15–28. doi: 10.1006/appe.2000.0364

Carpio, C. E., and Isengildina-Massa, O. (2009). Consumer willingness to pay for locally grown products: the case of South Carolina. *Agribusiness* 25, 412–426. doi: 10.1002/agr.20210

CBS News (2013, March 29). Edamame: is the future of American soy farmers' profits in edible soy beans for people? Available online at: https://www.cbsnews.com/news/edamame-is-the-future-of-american-soy-farmers-profits-in-edible-soy-beans-for-people/ (accessed January 5, 2021).

Cederroth, C. R., Zimmermann, C., and Nef, S. (2012). Soy, phytoestrogens and their impact on reproductive health. *Mol. Cell. Endocrinol.* 355, 192–200. doi: 10.1016/j.mce.2011.05.049

Cooper, J. C., Hanemann, M., and Signorello, G. (2002). One-and-one-half-bound dichotomous choice contingent valuation. *Rev. Econ. Stat.* 84, 742–750. doi: 10.1162/003465302760556549

Dettmann, R. L., and Dimitri, C. (2009). Who's Buying organic vegetables? demographic characteristics of U.S. consumers. *J. Food Prod. Mark.* 16, 79–91. doi: 10.1080/10454440903415709

Di Vita, G., Zanchini, R., Falcone, G., D'Amico, M., Brun, F., and Gulisano, G. (2021). Local, organic or protected? detecting the role of different quality signals among Italian olive oil consumers through a hierarchical cluster analysis. *J. Clean. Prod.* 290:125795. doi: 10.1016/j.jclepro.2021.125795

Duffy, C., Perez, K., and Partridge, A. (2007). Implications of Phytoestrogen Intake for Breast Cancer. *CA Cancer J. Clin.* 57, 260–277. doi: 10.3322/CA.57.5.260

Fan, X., Gómez, M. I., and Coles, P. S. (2019). Willingness to pay, quality perception, and local foods: the case of broccoli. *Agric. Resour. Econ. Rev.* 48, 414–432. doi: 10.1017/age.2019.21

Feldmann, C., and Hamm, U. (2015). Consumers' perceptions and preferences for local food: a review. *Food Qual. Prefer.* 40, 152–164. doi: 10.1016/j.foodqual.2014.09.014

Garber, B., and Neill, C. (2019). *Edamame: Costs, Revenues, and Profitability*. Blacksburg, VA: Virginia Cooperative Extension.

Govindasamy, R., Van Vranken, R., Sciarappa, W., Ayeni, A., Puduri, V. S., Pappas, K., et al. (2010). Ethnic crop opportunities for growers on the east coast: a demand assessment. *J. Ext.* 48, 1–9.

Haghiri, M., Hobbs, J. E., and McNamara, M. L. (2009). Assessing consumer preferences for organically grown fresh fruit and vegetables in eastern new brunswick. *Int. Food Agribus. Manag. Rev.* 12, 1–20. doi: 10.22004/ag.econ.92556

Hanemann, M., Loomis, J., and Kanninen, B. (1991). Statistical efficiency of double-bounded dichotomous choice contingent valuation. *Am. J. Agric. Econ.* 73, 1255–1263. doi: 10.2307/1242453

Hasler, C. M. (2002). The cardiovascular effects of soy products. *J. Cardiovasc. Nurs.* 16, 50–63. doi: 10.1097/00005082-200207000-00006

Heinrichs, P. A. (2016). *Consumer preferences and willingness to pay for fresh and frozen vegetables*. Champaign, IL: University of Illinois at Urbana-Champaign. Available online at: https://pdfs.semanticscholar.org/02fa/c09206b0db3db090197ed217d2ba47842c5c.pdf (accessed January 5, 2021).

Holcomb, R. B., Neill, C. L., Lelekacs, J., Velandia, M., Woods, T. A., Goodwin, J.r, H. L., et al. (2018). A local food system glossary: a rose by any other name. *Choices* 33, 1–8. doi: 10.22004/ag.econ.276057

Jaeger, A. (2019). *Arkansas Has Unsuspected Might in Edamame Market*. Available online at: https://www.thv11.com/article/news/edamame/91-56781308-85d2-4e7d-9242-6afef0712996 (accessed January 5, 2021).

Lockie, S., Lyons, K., Lawrence, G., and Grice, J. (2004). Choosing organics: a path analysis of factors underlying the selection of organic food among Australian consumers. *Appetite* 43, 135–146. doi: 10.1016/j.appet.2004.02.004

Lopez-Feldman, A. (2012). *Introduction to Contingent Valuation Using Stata. 17*. College Station, TX: Stata Press; StataCorp.

Lord, N., Neill, C., and Zhang, B. (2019a). *Production and Economic Considerations for Fresh Market Edamame in Southwest Virginia*. Blacksburg, VA: Virginia Tech.

Lord, N., Zhang, B., Kuhar, T., Carneiro, R., Sutton, K., and Yu, D. (2019b). *USDA Edamame Project*. Blacksburg, VA: Virginia Cooperative Extension. Available online at: https://www.pubs.ext.vt.edu/content/dam/pubs_ext_vt_edu/spes/SPES-104/SPES-104.pdf (accessed January 5, 2021).

Lusk, J. L., Schroeder, T. C., and Tonsor, G. T. (2014). Distinguishing beliefs from preferences in food choice. *Eur. Rev. Agric. Econ.* 41, 627–655. doi: 10.1093/erae/jbt035

Martinez, S. (2010). *Local Food Systems; Concepts, Impacts, and Issues*. Collingdale, PA: DIANE Publishing.

Michelfelder, A. J. (2009). Soy: a complete source of protein. *Am. Fam. Physician* 79, 43–47.

Nakatani, T., Aizaki, H., and Sato, K. (2016). *oohbchoice: Parametric Approach to Analyze One And-One-Half-Bound... in DCchoice: Analyzing Dichotomous Choice Contingent Valuation Data*. Available online at: https://rdrr.io/cran/DCchoice/man/oohbchoice.html (accessed January 5, 2021).

Neill, C. L., and Holcomb, R. B. (2019). Does a food safety label matter? consumer heterogeneity and fresh produce risk perceptions under the food safety modernization act. *Food Policy* 85, 7–14. doi: 10.1016/j.foodpol.2019.04.001

Neill, C. L., Holcomb, R. B., and Lusk, J. L. (2020). Potential beggar-thy-neighbor effects of state branding programs. *Agribus. Int. J.* 36, 3–19. doi: 10.1002/agr.21625

Neill, C. L., and Williams, R. B. (2016). Consumer preference for alternative milk packaging: the case of an inferred environmental attribute. *J. Agric. Appl. Econ.* 48, 241–256. doi: 10.1017/aae.2016.17

Onozaka, Y., Nurse, G., and McFadden, D. T. (2011). Defining sustainable food market segments: do motivations and values vary by shopping locale? *Am. J. Agric. Econ.* 93, 583–589. doi: 10.1093/ajae/aaq152

Owusu, V., and Owusu Anifori, M. (2013). Consumer willingness to pay a premium for organic fruit and vegetable in Ghana. *Int. Food Agribus. Manag. Rev.* 16, 1–20, doi: 10.22004/ag.econ.144649

Penn, J., and Hu, W. (2018). Understanding hypothetical bias: An enhanced meta-analysis. *Am. J. Agric. Econ.* 100, 1186–1206. doi: 10.1093/ajae/aay021

Pollack, S. L. (2001). Consumer demand for fruit and vegetables: the US example. *Chang. Struct. Glob. Food Consum. Trade* 6, 49–54.

Rahman, I., and Reynolds, D. (2017). Organic wine: the influence of biospheric, altruistic, and egoistic values on purchase intention, willingness to pay more, and willingness to sacrifice. *Int. J. Hosp. Beverage Manag.* 1:1. doi: 10.34051/j/2019.1

Sharma, K. P. (2013). *Varietal Yield Stability and Appropriate Management for Quality Organic Edamame Production*. Vernon, BC: Organic BC.

Shin, Y. H., Moon, H., Jung, S. E., and Severt, K. (2017). The effect of environmental values and attitudes on consumer willingness to pay more for organic menus: a value-attitude-behavior approach. *J. Hosp. Tour. Manag.* 33, 113–121. doi: 10.1016/j.jhtm.2017.10.010

Shockley, J. M., Dillon, C. R., and Woods, T. A. (2011). *Estimating the Economic Viability of a New Crop Alternative for the US Organic Market: Edamame-a Vegetable Soybean*. Pittsburgh, PA: Agricultural and Applied Economics Association.

Shurtleff, W., and Aoyagi, A. (2009). *History of Edamame, Green Vegetable Soybeans, and Vegetable-Type Soybeans (1275-2009): Extensively Annotated Bibliography and Sourcebook*. Lafayette, CA: Soyinfo Center.

Simonne, A. H., Smith, M., Weaver, D. B., Vail, T., Barnes, S., and Wei, C. I. (2000). Retention and changes of soy isoflavones and carotenoids in immature Soybean Seeds (Edamame) during processing. *J. Agric. Food Chem.* 48, 6061–6069. doi: 10.1021/jf000247f

Stevens-Garmon, J., Huang, C. L., and Lin, B.-H. (2007). Organic demand: a profile of consumers in the fresh produce market. *Choices* 22, 109–115.

Taku, K., Umegaki, K., Sato, Y., Taki, Y., Endoh, K., and Watanabe, S. (2007). Soy isoflavones lower serum total and LDL cholesterol in humans: a meta-analysis of 11 randomized controlled trials. *Am. J. Clin. Nutr.* 85, 1148–1156. doi: 10.1093/ajcn/85.4.1148

Umberger, W. J., McFadden, D. D. T., and Smith, A. R. (2009). Does altruism play a role in determining U.S. consumer preferences and willingness to pay for natural and regionally produced beef? *Agribusiness* 25, 268–285. doi: 10.1002/agr.20194

Venkatachalam, L. (2004). The contingent valuation method: a review. *Environ. Impact Assess. Rev.* 24, 89–124. doi: 10.1016/S0195-9255(03)00138-0

Wang, Q., Sun, J., and Parsons, R. (2010). Consumer preferences and willingness to pay for locally grown organic apples: evidence from a conjoint study. *HortScience* 45, 376–381. doi: 10.21273/HORTSCI.45.3.376

White, L. R., Petrovitch, H., Ross, G. W., Masaki, K., Hardman, J., Nelson, J., et al. (2000). Brain aging and midlife tofu consumption. *J. Am. Coll. Nutr.* 19, 242–255. doi: 10.1080/07315724.2000.10718923

Wolfe, E., Popp, M., Bazzani, C., Nayga Jr, R. M., Danforth, D., Popp, J., et al. (2018). Consumers' willingness to pay for edamame with a genetically modified label. *Agribusiness* 34, 283–299. doi: 10.1002/agr.21505

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

ORIGINAL RESEARCH
published: 21 October 2021
doi: 10.3389/fpls.2021.732164

Root K Affinity Drivers and Photosynthetic Characteristics in Response to Low Potassium Stress in K High-Efficiency Vegetable Soybean

Changkai Liu[1,2], Xue Wang[1,2,3], Bingjie Tu[1,2,3], Yansheng Li[1,2], Heng Chen[1,2,3], Qiuying Zhang[1,2] and Xiaobing Liu[1]*

[1] Key Laboratory of Mollisols Agroecology, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin, China, [2] Innovation Academy for Seed Design, Chinese Academy of Sciences, Harbin, China, [3] University of the Chinese Academy of Sciences, Beijing, China

Edited by:
*Bo Zhang,
Virginia Polytechnic Institute and State University, United States*

Reviewed by:
*Hadi Pirasteh-Anosheh,
National Salinity Research Center, Agricultural Research, Education and Extension Organization, Iran
Agnieszka Klimek-Kopyra,
University of Agriculture in Krakow, Poland*

***Correspondence:**
*Qiuying Zhang
zhangqiuying@iga.ac.cn*

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: *29 June 2021*
Accepted: *27 September 2021*
Published: *21 October 2021*

Citation:
Liu C, Wang X, Tu B, Li Y, Chen H, Zhang Q and Liu X (2021) Root K Affinity Drivers and Photosynthetic Characteristics in Response to Low Potassium Stress in K High-Efficiency Vegetable Soybean. Front. Plant Sci. 12:732164. doi: 10.3389/fpls.2021.732164

Significant variations of potassium absorption and utilization exist in vegetable soybean. Pot and hydroponic experiments were carried out to examine the characteristics of root potassium (K) affinity-associated drivers and photosynthesis in vegetable soybean (edamame) [*Glycine max* (L.) Merr.] with different K efficiency. Two K high-efficiency vegetable soybean genotypes (Line 19 and Line 20) and two K low-efficiency genotypes (Line 7 and Line 36) were investigated in low K and normal K conditions. The root of K high-efficiency genotypes had a higher K^+ affinity associated with a higher maximum K^+ uptake rate (*I*max), but lower Michaelis constant for K^+ absorption (*K*m) and lower compensation concentration for K^+ uptake (*C*min). Seedlings of K high-efficiency genotypes also had higher root vigor [triphenyl tetrazolium chloride (TTC) reduction method] and greater absorbing activity (methylene blue method), especially in the low K condition. Furthermore, the root bleeding-sap rate of K high-efficiency genotypes in low K stress was 9.9–24.3% greater than that of normal K conditions, which was accompanied by a relatively higher K concentration of root bleeding-sap in contributing to K^+ upward flux. The root of K high-efficiency vegetable soybean genotypes exhibited K^+ high-affinity and driving advantages. Photosynthetic parameters of K high-efficiency vegetable soybean genotypes were less affected by low K stress. Low K stress decreased the net photosynthetic rate of K high-efficiency genotypes by 6.1–6.9%, while that of K low-efficiency genotypes decreased by 10.9–15.7%. The higher chlorophyll (Chl) a/b ratio with enhanced relative content of Chl a in response to low K stress might be an adapted mechanism for K high-efficiency genotypes to maintain photosynthetic capacity. Stronger root K affinity drivers associated with photosynthetic adaptability to low K stress are the key factors in determining the K high-efficiency of vegetable soybeans.

Keywords: potassium, potassium efficiency, bleeding sap, chlorophyll a/b ratio, vegetable soybean

INTRODUCTION

Potassium application benefits vegetable soybean yield and quality (Liu et al., 2017), while the direct absorption and utilization of available potassium by plants in cultivated soil are always essential (Singh and Reddy, 2017; Chen et al., 2020; Dev et al., 2021). Selecting and breeding potassium (K) efficient varieties of vegetable soybean is an important biological means in making full use of K resources (Rengel and Damon, 2008). Previous studies have shown that there are great differences in K efficiency among different genotypes. For instance, intraspecific variations in K efficiency have been reported in many crops including rice (Yang et al., 2003), wheat (Zhang et al., 1999; Damon and Rengel, 2007), sweet potato (Wang et al., 2015), tomato (Chen and Gabelman, 1995), and soybean (Sale and Campell, 1987; Wang C. et al., 2012; Liu et al., 2019b).

Differences in the K efficiency of crops can be understood from two main aspects, such as (1) the difference in K uptake efficiency and (2) the difference in K utilization efficiency (Rengel and Damon, 2008). The utilization efficiency of K refers to the ability of the crop to convert unit K into dry matter yield (Wang et al., 2018). K high-efficiency vegetable soybean genotypes are good at redistributing K and dry matters with higher harvest index (HI) and higher K harvest index (KHI) (Liu et al., 2019b). K uptake efficiency emphasizes the capacity of root K absorption (Tsialtas et al., 2017). The higher specific K uptake rate (total K content/total root length) in K high-efficiency vegetable soybean genotypes ensures the supply of K to the whole plant. Besides, K high-efficiency genotypes also have a strong ability to regulate their root architecture to adapt to low K conditions (Liu et al., 2019a).

Except for root morphology, K uptake kinetic parameters, root bleeding-sap, and root vigor are also demonstrated as effective parameters in evaluating K uptake efficiency (Teo et al., 1992; Hao et al., 2015; Cui et al., 2016; Zhang et al., 2017). Sufficient K can increase plant hydraulic conductance and transpiration (Pettigrew, 2008). The vigorous root can increase nutrient and water uptake, promoting the whole plant growth. For instance, the root bleeding-sap and the upward fluxes of K are higher in the cotton cultivars with high K efficiency (Yang, 2011). As substrates for photosynthesis, water and mineral elements absorbed by plant roots are transported upward by transpiration (Liu et al., 2016). Improved root characteristics may contribute to plant-water status, enhanced photosynthesis, biomass, and yield of soybean cultivars (Cui et al., 2016). In cotton, K efficient genotype 103 has more suitable K absorption kinetic parameters and more efficient photosynthate transport (Hao et al., 2015, 2016). Thus, the uptake power of the root system determines the supply of nutrients in the above-ground part of the plant.

Less photosynthetic assimilates and reduced assimilate transport out of the leaves to the developing fruit greatly contribute to the negative consequences that deficiencies of K have on yield and quality production (Pettigrew, 2008). Genotypic variations in photosynthetic decline caused by K deficiency have been reported. For instance, compared with the K-inefficient cotton cultivar, the K-efficient cultivar has a higher net photosynthetic rate (Pn) associated with higher biomass products (Wang N. et al., 2012).

Hence, understanding how plants take up and use K is of scientific and practical importance. Research into this mechanism has become increasingly urgent as a result of major ecological and agricultural issues. In our previous studies, K high-efficiency vegetable soybean genotypes were demonstrated with a strong K uptake and redistribution ability (Liu et al., 2019a,b). However, the root uptake dynamics and photosynthetic characteristics of K high-efficiency vegetable soybean are still not clear. Therefore, this study compared the differences of root activity, root bleeding-sap, the upward fluxes of K, K kinetics parameters, photosynthetic parameters, and chlorophyll content between K high-efficiency and K low-efficiency vegetable soybean genotypes. The data obtained revealed the mechanisms underlying K absorption and utilization of high-efficiency in vegetable soybean.

MATERIALS AND METHODS

Plant Material

Based on previous K efficiency selection (Liu et al., 2019a), two K high-efficiency vegetable soybean genotypes (Line 19 and Line 20) and two K low-efficiency genotypes (Line 7 and Line 36) were used in this study. Compared with K low-efficiency genotypes, the K high-efficiency genotypes have higher K agronomic efficiency (KAE), recovery, internal utilization-efficiency rate (KIUE), and specific K uptake rate (Liu et al., 2019a,b). Greater reductions in K concentration of vegetative organs were found in K high-efficiency genotypes than K low-efficiency genotypes in low K conditions (Liu et al., 2019b). All genotypes were released by the Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin, China.

Experimental Design

A pot experiment was conducted at the agronomy farm of Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin, China (45°732N, 126°612E; altitude 128 m a.s.l.) in 2017. The soil used was a typical Mollisol (Black soil) with the following properties: soil pH 6.6, organic matter 28.9 g kg^{-1}, total N 2.3 g kg^{-1}, total P 1.3 g kg^{-1}, total K 18.9 g kg^{-1}, available N 159 mg kg^{-1}, available P 57.0 mg kg^{-1}, and available K 85 mg kg^{-1} (insufficient for vegetable soybean yield and quality; Liu et al., 2017). The experiment was conducted in a completely random design with three replicates and five pots per replicate. About 3 seeds per pot were sown in May 2017 with regular manual pest and weed control. The pot size was 32 cm in diameter by 27 cm tall. A uniform fertilizer application at seeding included 70 kg ha^{-1} diammonium phosphate and 98 kg ha^{-1} urea. Treatments consisted of two K-fertilizer rates at seeding, K_2SO_4 0 kg ha^{-1} (K0) and K_2SO_4 120 kg ha^{-1} (K120). Plants were randomly harvested at fourth-node (V4), full bloom (R2), beginning seed (R5), and full seed (R6) stages (Fehr et al., 1971) for root bleeding-sap and K upward fluxes measurement.

A hydroponic experiment was used for K^+ absorption kinetic parameters, root vigor, and root absorbing activity test. Seeds of the four selected genotypes were sterilized and germinated on moistened filter paper in a plant growth chamber at 60% humidity and 28°C, under a 16 h light, 8 h dark cycle for 3–4 days. After that, the seedlings were transferred into light-proof glass boxes (volume 15 cm^3 × 20 cm^3 × 20 cm^3) with half-strength modified Hoagland nutrient solution, as described by Wang C. et al. (2012). There were six replications in each box. The cotyledons of all samples were excised to eliminate any additional supply of nutrients. When the plants grew to the first trifoliolate stage (V1), they were treated with half-strength Hoagland nutrient solution with different concentrations of K^+ (0.5 and 3 mmol L^{-1}). The K source was K_2SO_4. The initial pH value of the nutrient solution was 6, which was adjusted by 0.1 mol L^{-1} NaOH or 0.1 mol L^{-1} HCl. After transplanting, the samples were continuously cultured for 9 days. The nutrient solution was changed every 3 days and kept enough O_2 by regular ventilation with an air pump. The pH value of the solution was adjusted to 6 when the nutrient solution was changed.

Measurements

Collection of root bleeding-sap: root bleeding-sap collection was based on the method used by Cui et al. (2016) with slight improvements in the pot experiments. The root bleeding-sap was conducted from 9:30 a.m. to 1:30 p.m. at V4, R2, R5, and R6 stages. The plants were cut with branch scissors at cotyledons position, and the cross-section was cleaned with a small amount of absorbent cotton and then immediately connected to the root bleeding-sap collection device and the joint was sealed. The collected root bleeding-sap was temporarily stored in a refrigerator at 4°C for measurement. The formula was as follows: root bleeding-sap rate (g h^{-1} per plant) = bleeding-sap weight/4 h.

Where K upward fluxes = root bleeding-sap rate × K concentration, K concentration was determined by flame spectrophotometer (INESA FP6400A, Shanghai, China) (Liu et al., 2017; Nguyen et al., 2017).

K^+ absorption kinetic parameters: A measurement of a depletion curve of a single plant. Five replicates per genotype were conducted. After K deficiency and starvation for 48 h, samples under the treatment of K3 at 9 days were immersed in 0.2 mmol L^{-1} $CaSO_4$ solutions three times, dried, and put into a 200 ml absorption solution. The composition of the absorption solution was 0.05 mmol L^{-1} potassium chloride (KCl) + 0.2 mmol L^{-1} $CaSO_4$. The absorption solution was placed in a dark conical flask, the culture temperature was (25 ± 2)°C, and the light intensity was 210–260 mol m^{-2} s^{-1}. Parameter calculation including the maximum K^+ uptake rate (*I*max), Michaelis constant for K^+ absorption (Km), compensation concentration for K^+ uptake (*C*min), and statistical analysis refers to the modified method of Epstein and Hagen (1952) and Claassen and Barber (1977). *C*min represents the concentration of K^+ in the solution when the concentration remains constant. Ion consumption curve equation: $Y = a + bX + cX^2$ (1), take the negative derivative of the equation: Y' = b + 2cX (2). Hence, if X goes to 0, then Y' = b = Km.

Root vigor, as determined by enzymatic reduction assay of triphenyl tetrazolium chloride (TTC) was conducted according to the modified protocol of Duncan and Widholm (2004).

Root absorption activity was determined by methylene blue adsorption (Zhang and Yan, 2003; Song and Wang, 2005). First, the root volume is converted to 1 ml = 1 g. The root was put into a solution with a known concentration of methylene blue for 1.5 min. Removed and placed the root into the second beaker, repeated twice. Colorimetric determination of methylene blue was conducted in the remaining solution in the three beakers at 660 nm using the UV-visible spectrophotometer (PERSEE T6, Beijing, China).

$$\text{Total absorbing area } (m^2) = [(C1 - C1') \times V1] \times 1.1 + [(C2 - C2') \times V2] \times 1.1$$

$$\text{Actively absorbing area } (m^2) = [(C3 - C3') \times V3] \times 1.1$$

$$\text{The ratio of actively absorbing area to total absorbing area } (\%) = 100 \times \text{actively absorbing area/total absorbing area}$$

$$\text{The ratio of the total absorbing area to root volume } (m^2\ cm^{-3}) = \text{Total absorbing area/Root volume}$$

Where C, the original concentration of the solution (mg ml^{-1}); C′, the concentration of the solution after leaching (mg ml^{-1}); 1, 2, 3 is the beaker number; V, Volume of the solution; When 1 mg methylene blue forms a monolayer, it covers an area of 1.1 m^2 (Zhang and Yan, 2003).

Photosynthetic properties of the youngest fully expanded main-stem leaf (the third leaf from the apex) were determined at 10:00–12:00 am at V4, R2, R4 (full pod), R5, and R6 stages with a Li-6800 (Li-COR, Lincoln, NE, United States) at 25°C, 60% relative humidity, 500 μmol mol^{-1} CO_2 concentration, and 1,200 μmol m^{-2} s^{-1} quantum flux.

Determination of chlorophyll (Chl) content (Wang et al., 2021) was conducted using 0.2 g fresh leaves which were cut into pieces and put into a 50 ml centrifuge tube. Then, 20 ml of 80% ethanol solution was added and then soaked in a dark place for 8 h. Absorbance was determined at 665 and 649 nm, respectively, by the UV-visible spectrophotometer.

$$\text{Chl a} = 13.95A_{665} - 6.88A_{649}$$
$$\text{Chl b} = 2.96A_{649} - 7.32A_{665}$$
$$\text{Total chlorophyll concentration} = \text{Chl a} + \text{Chl b}.$$

Statistical Analyses

Differences among treatments and genotypes were examined by ANOVA using SPSS 17, and the means were separated by the LSD test at the 5% level. Linear regression analysis between root bleeding-sap rate per root length, K upward fluxes rate per root length, K concentration of root bleeding-sap, and plant K concentration were conducted after Pearson correlation analysis using a two-tailed test in SPSS 17. The figures were created using SigmaPlot 12.

RESULTS

Comparison of K^+ Absorption Kinetic Parameters Between Two K Efficiency Genotypes

Table 1 shows the K^+ kinetics absorption parameters of the four vegetable soybean genotypes in a hydroponic experiment. The maximum K^+ uptake rate (*I*max) in K high-efficiency genotypes was around 58.2–65.5 μmol g^{-1} min^{-1} FW, which was significantly higher than that of K low-efficiency genotypes around 42.4–54.1 μmol g^{-1} min^{-1} FW ($P < 0.05$). The compensation concentration for K^+ uptake (*C*min) in K high-efficiency genotypes was 0.83–1.32 μmol L^{-1}, which was lower than that of K low-efficiency genotypes (3.16–3.22 μmol L^{-1}) ($P < 0.05$). The Michaelis constant for K^+ absorption (Km) is a parameter evaluating the affinity between root and K^+. The greater the value, the smaller the affinity. Higher affinity was found in K high-efficiency vegetable soybean genotypes with a lower Km of 32.8–35.0 μmol L^{-1} than that of 48.6–49 μmol L^{-1} in K low-efficiency genotypes ($P < 0.05$).

Root Bleeding-Sap and Upward Fluxes of K^+ in the Pot Experiment

As shown in **Figure 1**, root bleeding-sap rate per plant showed a trend of increasing first and then decreasing from the fourth-node (V4) to full seed (R6) stage with the maximum at full bloom (R2) stage. The root bleeding-sap rate per plant of Line 20 was the highest among the four genotypes with 1.8 ml h^{-1} $root^{-1}$ in K0 and 1.6 ml h^{-1} $root^{-1}$ in K120, respectively. Compared with K120 treatment, K0 treatment increased the average of the root bleeding-sap rate per plant over the four stages by 9.9–24.3% in K high-efficiency genotypes, but by –2.2–5.1% in K low-efficiency genotypes. On the other hand, the root bleeding-sap rate per root length in four genotypes was found higher at the V4–R2 stage, dropped down at the R5–R6 stage. Interestingly, K high-efficiency genotypes had a higher root bleeding-sap rate per root length compared with K low-efficiency genotypes at the beginning seed stage ($P < 0.05$). At this time, root bleeding-sap rate per root length ranged 0.09–0.13 ml h^{-1} cm^{-1} in K0 and 0.10–0.15 ml h^{-1} cm^{-1} in K120 treatment, while in K low-efficiency genotypes kept around 0.05 ml h^{-1} cm^{-1} in K0 and 0.05–0.07 ml h^{-1} cm^{-1} in K120 treatment.

Higher K upward fluxes were found in K120 treatment over the four genotypes, which was accompanied consistently by a higher concentration of K in the root bleeding-sap (**Figure 2**).

TABLE 1 | Comparison of K^+ absorption kinetic parameters of root systems between two potassium (K) efficiency types under low K stress.

	Line 19	Line 20	Line 7	Line 36
Imax (μmol g^{-1} min^{-1} root FW^{-1})	65.0 a	58.2 b	42.4 d	54.1 c
Km (μmol L^{-1})	35.0 b	32.8 c	49.0 a	48.6 a
*C*min (μmol L^{-1})	1.32 b	0.83 c	3.16 a	3.22 a

Different letters within the same row indicate statistical significance at the $P < 0.05$ level.

The K flux rate per root length was highest at the fourth-node stage, with plant growth K flux rate per root length decreased. At the beginning seed stage, K high-efficiency genotypes had a higher K flux rate per root length and K concentration of root bleeding sap compared with K low-efficiency genotypes ($P < 0.05$). The K flux rate per root length of K high-efficiency genotypes was 19.4–32.4 μg h^{-1} cm^{-1} in K0 and 24.1–37.2 μg h^{-1} cm^{-1} in K120 treatment, that of 7.6–8.7 μg h^{-1} cm^{-1} in K0 and 9–15.2 μg h^{-1} cm^{-1} in K120 treatment in K low-efficiency genotypes. Meanwhile, K concentration of root bleeding-sap in K high-efficiency genotypes reached 51.1–63.2 μg ml^{-1} in K0 and 60.2–64 μg ml^{-1} in K120 treatment at the beginning seed stage, which was significantly higher than that of K low-efficiency genotypes with 41.1–48 μg ml^{-1} in K0 and 46.1–52.4 μg ml^{-1} in K120 treatment ($P < 0.05$).

Correlation analysis revealed that K upward fluxes rate per root length and K concentration of root bleeding-sap were positively correlated with K concentration per plant ($P < 0.05$) (**Table 2**).

Root Vigor and Absorbing Activity

The root vigor was tested in a hydroponic experiment using the measurement of respiratory activity with triphenyl tetrazolium chloride (TTC) (**Figure 3**). At the seedling stage (9 days after treatment), the root vigor of K high-efficiency genotypes was induced by low K stress. Low K stress (K0.5) increased the root vigor of K high-efficiency genotypes by 46–85% compared with normal K treatment (K3) ($P < 0.05$). Although no consistent trend was observed in K low-efficiency genotypes, low K stress decreased root vigor by 21% in Line 36 ($P < 0.05$). The highest root vigor of 516 μg g^{-1} FW h^{-1} was found in Line 20 under low K conditions.

The root absorbing activity parameters are shown in **Table 3**. Higher total absorbing area and actively absorbing area were found in K high-efficiency genotypes. Meanwhile, in K high-efficiency genotypes, low K stress increased the actively absorbing area by 21.4–30.6% and the total absorbing area by 9.6–19% ($P < 0.05$). In contrast, an opposite trend was found in K low-efficiency genotypes. Under low K stress, the actively absorbing area decreased by 6.1–10.3% and the total absorbing area decreased by 6.6–15.7% ($P > 0.05$). Furthermore, K high-efficiency genotypes also had a higher ratio of total absorbing area to root volume, especially in Line 19 with 62.3 m^2 cm^{-3} (K0.5) and 60.2 m^2 cm^{-3} (K3). Low K stress increased the ratio of actively absorbing area to total absorbing area by 9.3–9.4% in K high-efficiency genotype ($P < 0.05$), but by 0.6–5% in K low-efficiency genotypes ($P > 0.05$).

Comparison of Photosynthetic Parameters Between Two K Efficiency Genotypes

A pot experiment was conducted to determine photosynthetic parameters of two K efficiency vegetable soybean genotypes affected by K deficiency (K0) are shown in **Figure 4**. K deficiency decreased the photosynthetic rate (Pn) of the four vegetable soybean genotypes. During the whole growth stages, the Pn of

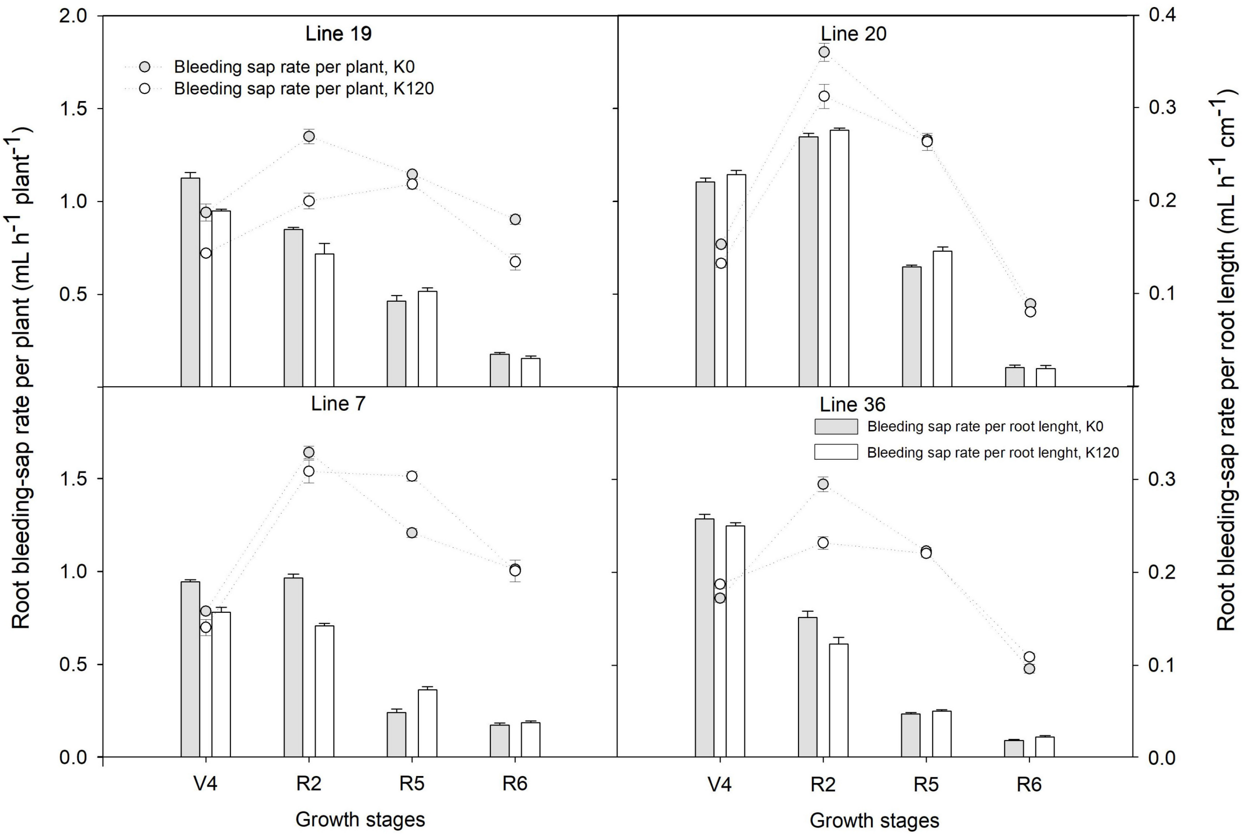

FIGURE 1 | Root bleeding-sap of distinct potassium (K) efficiency genotypes under low K application in a pot experiment. Values are means of three replicates ± SE.

K high-efficiency genotypes decreased by 6.1–6.9%, while that of low-efficiency genotypes decreased by 10.9–15.7%. From full bloom to full seed stage, the Pn increased first and then decreased. The maximum value occurred at the full pod (R4) stage of the K low-efficiency genotype Line 36, which reached 28.8 μmol CO_2 m^{-2} s^{-1} and 33.8 μmol CO_2 m^{-2} s^{-1} under K0 and K120 treatments, respectively. Meanwhile, from full pod to full seed stage, the dramatic decline of the Pn was found in K low-efficiency genotypes. The Pn of Line 36 decreased by 87.2% in K0 and 86.2% in K120 treatment ($P < 0.05$), and that of Line 7 decreased by 54.3% in K0 and 59% in K120 treatment. Whereas the decline of the Pn in K high-efficiency genotypes was around 33.5–51.8% in K0 and 36.7–41.1% in K120 treatment, which was not as much as that of K low-efficiency genotypes.

Potassium deficiency also decreased the stomatal conductance (Gs) in all genotypes, but K high-efficiency genotypes were less affected. The decrease of Gs by K deficiency was 7.5–7.8% in K high-efficiency genotypes, but 9.4–12.5% in K low-efficiency genotypes. Low K treatment had the most obvious effect on Gs at the beginning seed stage.

As opposed to Pn and Gs, K deficiency increased the intercellular CO_2 concentration (Ci), especially in K low-efficiency genotypes. The increase of Ci by K deficiency was 7.3–7.9% in K low-efficiency genotypes, but 2–6% in K high-efficiency genotypes.

Potassium deficiency decreased the transpiration rate (Tr) of all genotypes. Across the five stages, Tr of K high-efficiency genotypes was around 8.8–9.8 mmol H_2O m^{-2} s^{-1} in K0 and 9–10.1 mmol H_2O m^{-2} s^{-1} in K120 treatment, which was higher than that of 7.1–8.2 mmol H_2O m^{-2} s^{-1} in K0 and 7.1–8.6 mmol H_2O m^{-2} s^{-1} in K120 treatment in K low-efficiency genotypes. At the full seed stage, higher Tr was found in K high-efficiency genotypes with 4–4.5 mmol H_2O m^{-2} s^{-1} in K0 and 4.2–4.7 mmol H_2O m^{-2} s^{-1} in K120 treatment, but only 1–1.2 mmol H_2O m^{-2} s^{-1} and 1.4 –1.6 mmol H_2O m^{-2} s^{-1} in K low-efficiency genotypes.

Chlorophyll

Pot experiment indicated that K deficiency significantly decreased leaf chlorophyll content of the four vegetable soybean genotypes (**Table 4**). The average total Chl content in the four growth stages of K high-efficiency genotypes was 2.29–2.53 mg g^{-1} in K0 and 2.89–2.91 mg g^{-1} in K120 treatment, which was lower than that of K low-efficiency genotypes with 2.57–3.2 mg g^{-1} in K0 and 3.08–3.5 mg g^{-1} in K120 treatment. However, K deficiency increased the ratio of Chl a to Chl b, especially in K high-efficiency genotypes. The ratio of Chl a to Chl b was increased by 36.2–38.1% in K high-efficiency genotypes ($P < 0.05$), but only by 6–9.8% in K low-efficiency genotypes.

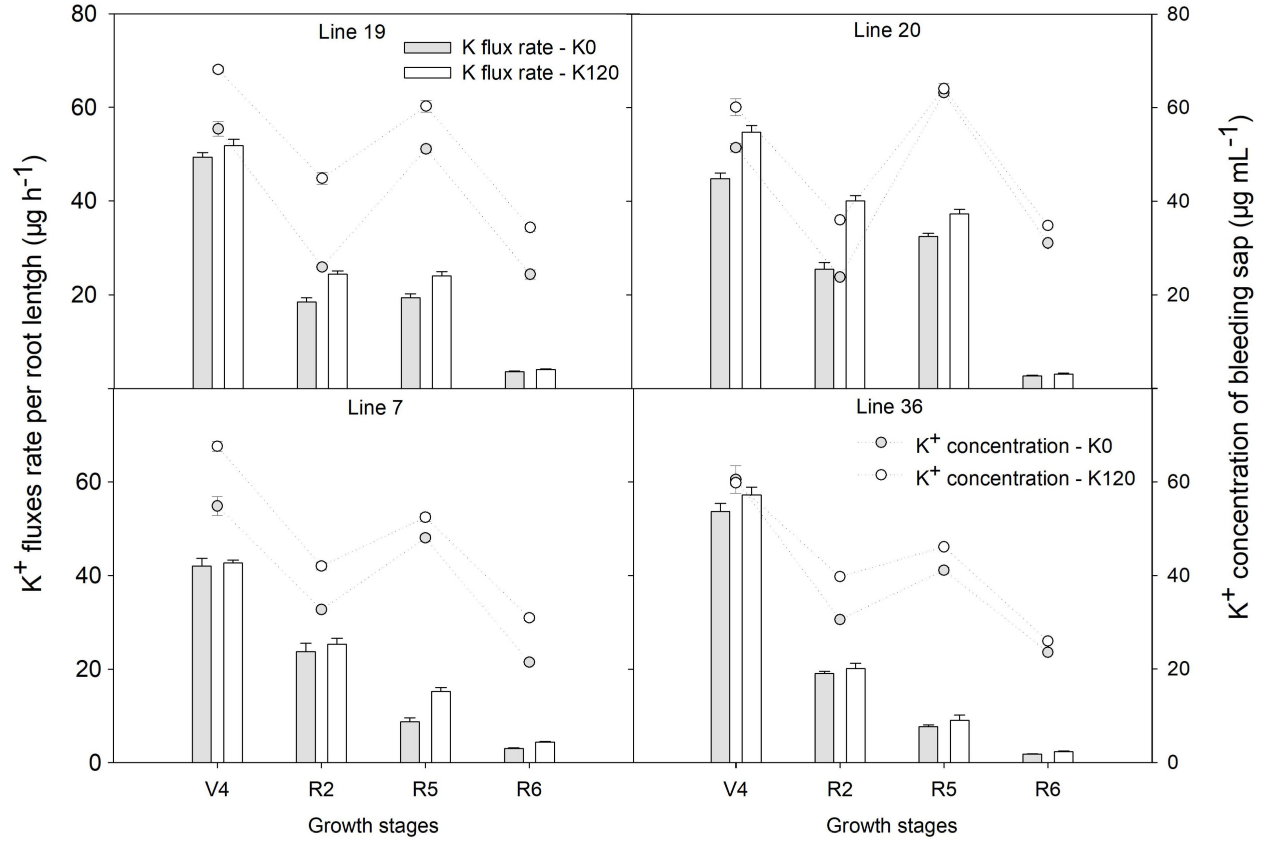

FIGURE 2 | K upward fluxes and concentration in root bleeding sap between two K efficiency genotypes in a pot experiment. Values are means of three replicates ± SE.

TABLE 2 | Correlation analysis of root bleeding-sap rate per root length, K upward fluxes rate per root length, K concentration of root bleeding-sap, and plant K concentration in vegetable soybean.

	K upward fluxes rate per root length	K concentration of root bleeding-sap	Plant K concentration
Root bleeding-sap rate per root length	0.884**	0.415*	0.707**
K upward fluxes rate per root length		0.745**	0.831**
K concentration of root bleeding-sap			0.598**

$^{*}P < 0.05$, $^{**}P < 0.01$ for significance of correlations (Pearson). Plant K concentration data based on the study of Liu et al. (2019a).

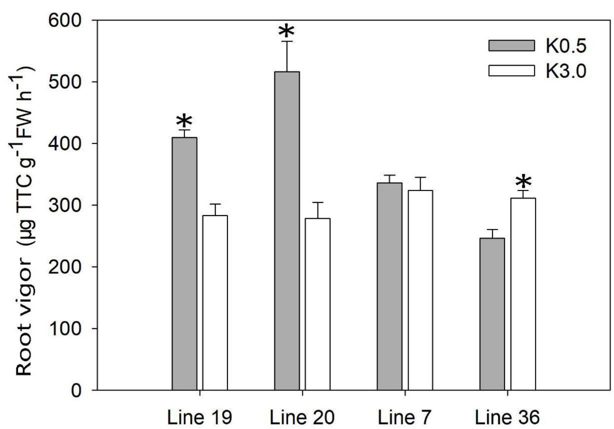

FIGURE 3 | Effect of low K treatment on root vigor of vegetable soybean with different K efficiency in a hydroponic experiment. Values are means of three replicates ± SE. *means significant differences at $P < 0.05$ level.

DISCUSSION

Potassium high-efficiency vegetable soybean genotypes are more efficient in root architecture adjustment associated with higher specific root K uptake rate (total K accumulation/total root length) to adapt to low K conditions, which ensures an adequate supply of K (Liu et al., 2019a). Based on these, this study investigated the root bleeding-sap rate and K upward fluxes rate of K high and low-efficiency vegetable soybean genotypes, which are important indicators assessing root pressure, root activity, and K uptake abilities (Doussan et al., 2006;

TABLE 3 | Root adsorption activity of vegetable soybean genotypes with different K efficiency.

		Total absorbing area (m^2 $plant^{-1}$)	Actively absorbing area (m^2 $plant^{-1}$)	Ratio of actively absorbing area to total absorbing area (%)	Ratio of total absorbing area to root volume (m^2 cm^{-3})
Line 19	K0.5	0.103 a	0.051 a	49.5 a	62.3 a
	K3.0	0.094 b	0.042 bc	45.3 bc	60.2 a
Line 20	K0.5	0.094 b	0.047 ab	49.9 a	52.9 b
	K3.0	0.079 c	0.036 cd	45.6 bc	52.5 b
Line 7	K0.5	0.070 d	0.035 de	50.1 a	52.1 b
	K3.0	0.083 c	0.039 cd	47.7 ab	44.2 c
Line 36	K0.5	0.071 d	0.031 e	43.9 c	40.6 cd
	K3.0	0.076 cd	0.033 de	43.6 c	38.8 d

Different letters within the same column indicate statistical significance at the $P < 0.05$ level.

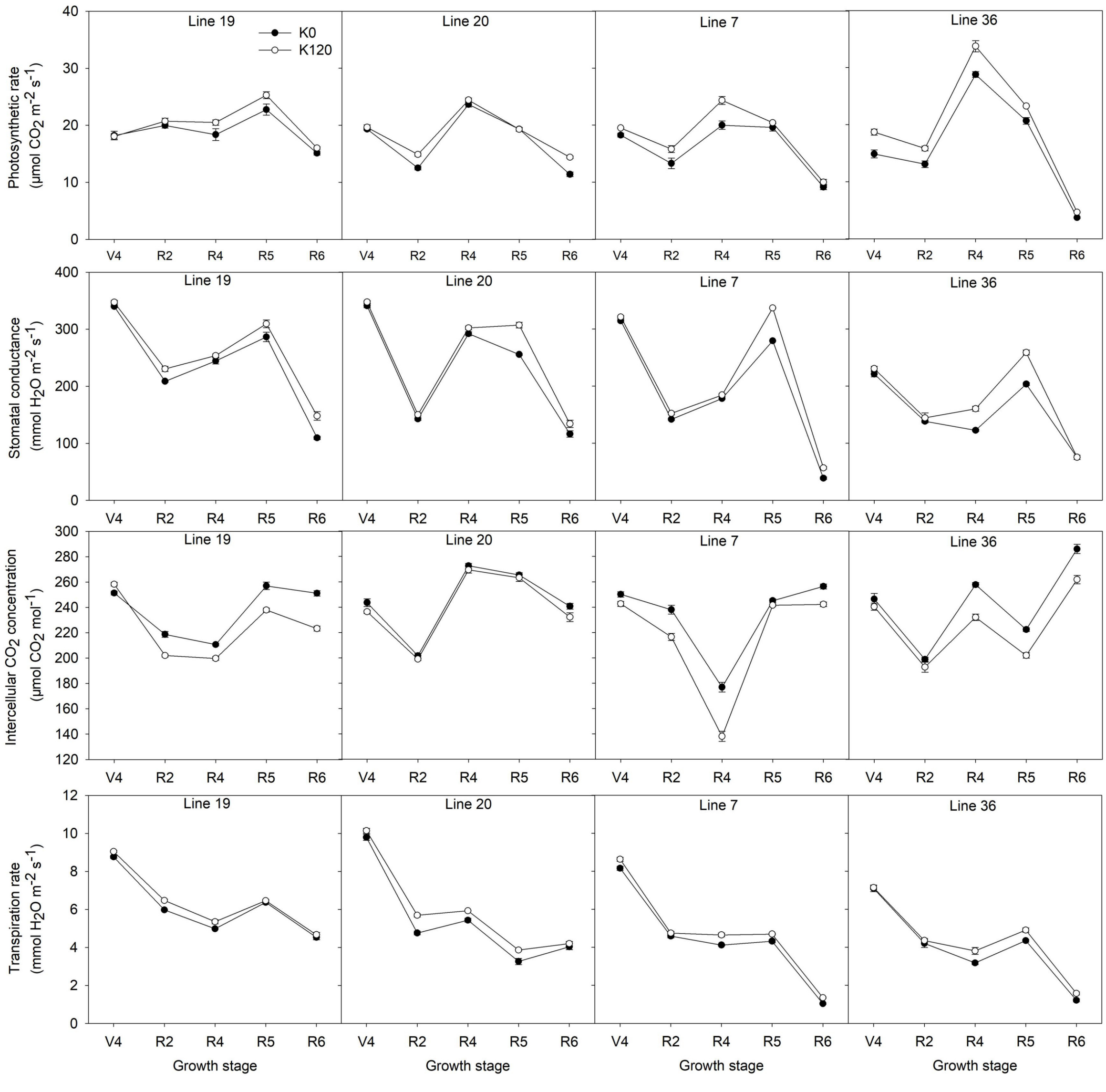

FIGURE 4 | Changes of the photosynthetic parameters of vegetable soybean genotypes with different K efficiency in a pot experiment. Values are means of three replicates ± SE.

TABLE 4 | Changes of chlorophyll a, chlorophyll b, and total chlorophyll concentrations (mg g^{-1}) of vegetable soybean with different K efficiency in a pot experiment.

Stage		Line 19		Line 20		Line 7		Line 36	
		K0	K120	K0	K120	K0	K120	K0	K120
V4	Chl a	1.66 d	1.99 ab	1.36 f	1.86 c	1.83 c	1.94 b	1.46 e	2.06 a
	Chl b	0.26 d	0.59 b	0.25 d	0.68 a	0.48 c	0.54 bc	0.31 d	0.47 c
	Total	1.92 c	2.58 a	1.61 e	2.54 a	2.31 b	2.48 a	1.77 d	2.53 a
	Chl a/b ratio	6.36 a	3.36 f	5.33 b	2.73 g	3.78 e	3.61 ef	4.77 c	4.41 d
R2	Chl a	1.90 d	1.97 d	2.11 c	2.22 b	2.30 b	2.53 a	2.08 c	2.21 b
	Chl b	0.95 d	1.05 c	1.12 b	1.28 a	1.03 c	1.28 a	1.12 b	1.25 a
	Total	2.84 e	3.02 d	3.23 c	3.50 b	3.34 bc	3.81 a	3.21 c	3.46 b
	Chl a/b ratio	2.00 b	1.86 bc	1.88 bc	1.74 c	2.23 a	1.98 b	1.86 bc	1.76 c
R5	Chl a	2.03 c	2.26 b	1.68 d	2.11 c	2.31 b	2.48 a	2.18 bc	2.56 a
	Chl b	1.01 c	1.17 b	0.86 d	1.15 b	1.17 b	1.33 a	0.97 c	1.20 b
	Total	3.04 c	3.44 b	2.53 d	3.26 bc	3.48 b	3.80 a	3.15 c	3.76 a
	Chl a/b ratio	2.01 bc	1.93 bc	1.96 bc	1.83 c	1.97 bc	1.87 c	2.24 a	2.13 ab
R6	Chl a	1.52 de	1.63 d	1.17 f	1.47 e	2.53 b	2.68 a	1.59 de	1.80 c
	Chl b	0.81 cd	0.88 c	0.63 de	0.87 c	1.13 b	1.23 a	0.57 e	0.78 cd
	Total	2.33 de	2.51 cd	1.80 f	2.34 de	3.67 b	3.92 a	2.16 e	2.58 c
	Chl a/b ratio	1.87 cd	1.84 cd	1.87 cd	1.69 d	2.23 b	2.17 bc	2.77 a	2.31 b

Different letters within the same row indicate statistical significance at the $P < 0.05$ level.

Wang P. et al., 2020). The results indicated that K high-efficiency vegetable soybean genotypes against low K condition by increasing root bleeding-sap rate per plant and maintaining higher root bleeding-sap rate per root length at the beginning seed stage. The root bleeding-sap rate of K high-efficiency genotypes in low K stress was 9.9–24.3% greater than that of normal K conditions, which was accompanied by the relatively higher K concentration of root bleeding-sap in contributing to K^+ upward flux. In K high-efficiency cotton cultivars, higher root bleeding-sap and K upward fluxes could also be induced by low K stress (Yang, 2011). Suitable grafting would help watermelon seedlings accumulate more K by increasing root bleeding-sap volume and the total K in the root bleeding-sap (Huang et al., 2013). Due to the rate of root bleeding-sap is closely related to plant nutrient supply, water transport, and even photosynthesis (Guan et al., 2014; Jia et al., 2018; He et al., 2019), increased rate of root bleeding-sap under K deficiency might be an important regulatory mechanism for vegetable soybean efficient uptake K. Besides, beginning seed stage is a most important period for seed establishment, the higher root bleeding-sap rate per root length in K high-efficiency vegetable soybean genotypes accompanied by higher K upward fluxes rate per root length and higher K concentration of root bleeding-sap is positively correlated with plant K concentration ($P < 0.05$). This is another evidence demonstrating that the root of K high-efficiency genotypes has a stronger affinity for K^+.

Potassium kinetic parameters and root activities are also important factors assessing K efficiency when plants suffering low K stress (Silberbush and Barber, 1983; Cui et al., 2016; White et al., 2018). The present study revealed a lower Michaelis constant (Km) and compensation concentration for K^+ uptake (Cmin), and a higher maximum K^+ uptake rate (Imax) from K high-efficiency genotypes, compared with K low-efficiency genotypes. Lower Km and Cmin indicate higher affinity between the roots and K^+ and stronger ability to use the low-concentration K^+, while higher Imax ensures a faster K uptake rate (Glass, 1980; Teo et al., 1992). The absorption kinetic parameters of ion uptake are useful indexes of the level of adaptation of the genotype to the nutrient condition in the soil (Crowley, 1975; Daniel et al., 2020). The study of Hao et al. (2015) also recognized that K kinetic parameters could be used to test crop low-K adaptability. Therefore, it was reasonable to say that K high-efficiency genotypes had better K adaptability than K low-efficiency genotypes. On the other hand, root activity is another important heritable trait in evaluating root absorption ability, which also influences nutrients acquisition and initial canopy cover and, thereby, crop yields (Liu et al., 2015; Cui et al., 2016; White et al., 2018). In the present study, both root vigor and absorbing activity (including actively absorbing area, total absorbing area, and the ratio of actively absorbing area to total absorbing area) were consistently enhanced by low K stress in K high-efficiency genotypes, while those of K low-efficiency genotypes are inhibited. This kind of difference can be regulated by plant phytohormones (Yang, 2011), which were controlled by specific genes and pathways (Liu et al., 2020). Therefore, K high-efficiency genotypes are more adapted to low K stress through regulating root K affinity drivers, which is beneficial for the upward flux of nutrients (Liu et al., 2015).

Photosynthetic parameters affected by K levels are direct references to characterize the photosynthetic capacity of crops (Wang Y. et al., 2020). Hence, it is important to examine the photosynthetic characteristics of vegetable soybean with different K efficiency types. In the present study, photosynthetic parameters are less affected by low K in K high-efficiency genotypes. The *Pn* of K high-efficiency genotypes decreased by 6.1–6.9% and K low-efficiency genotypes by 10.9–15.7%. Many investigations also indicate that crops with high K efficiency should be more efficient in photosynthesis or that photosynthetic capacity is less affected by low K stress (Wang et al., 2014). For instance, under K

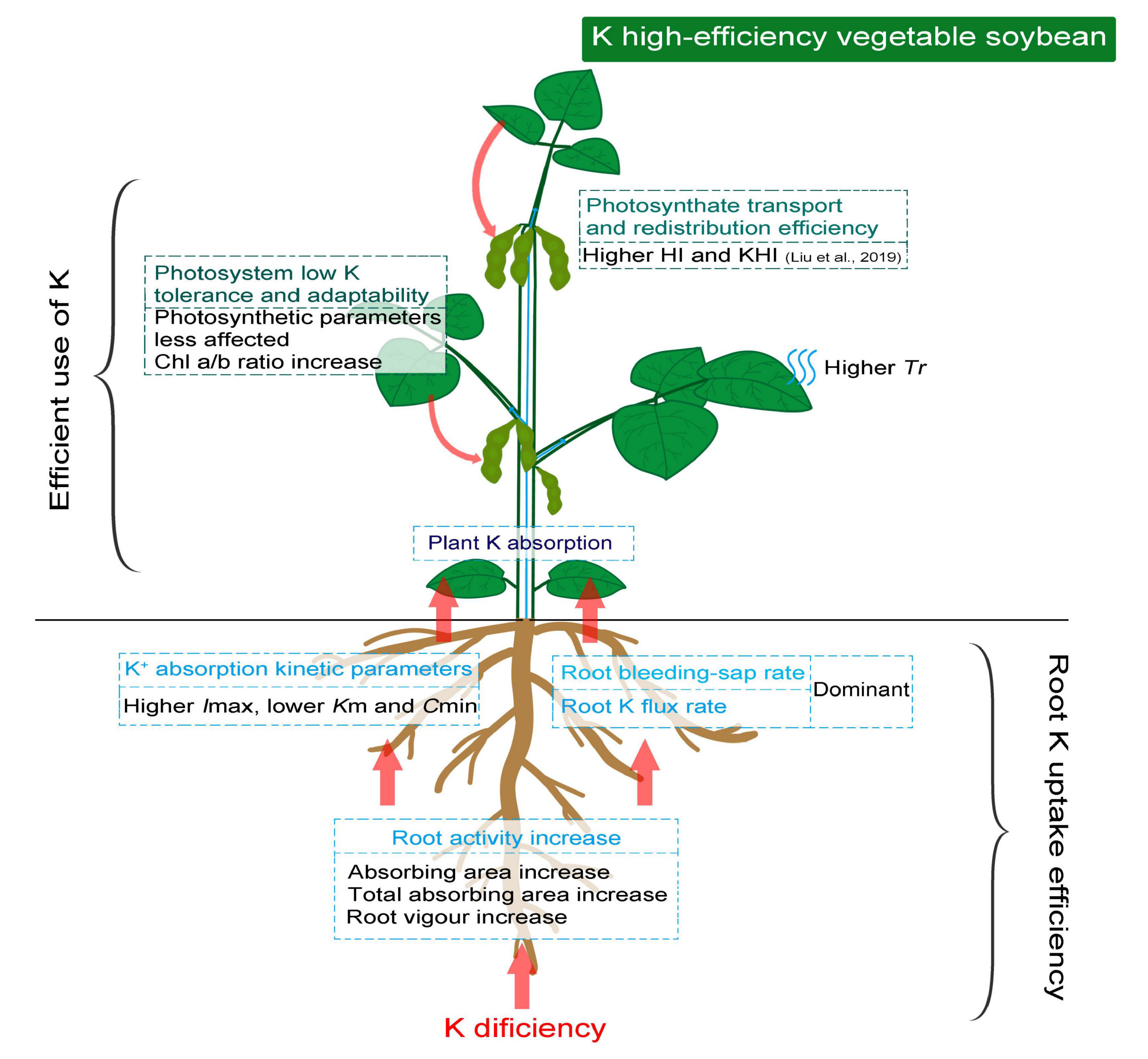

FIGURE 5 | Schematic diagram demonstrating efficient K absorption and photosynthetic assimilation in K high-efficiency vegetable soybean under low K condition.

deficiency, *Pn* of K-efficient cotton cultivar Liaomian 18 is 19.4% higher than that of K-inefficient cultivar NuCOTN99^{B}. Besides, photosynthetic parameters of Liaomian 18 were less affected by low K stress (Wang N. et al., 2012). However, the present study did not find absolute superiority of *Pn* in K high-efficiency genotypes, but they have a higher HI and KHI [the relevant results have been published by Liu et al. (2019b)]. A high HI is fundamental to efficient utilization of all resources taken up by the plant, and the photosynthate transport and distribution rather than photosynthesis rate are critical for low K adaptation in K high-efficiency genotypes (Rengel and Damon, 2008; Hao et al., 2016).

Consistent with photosynthetic parameters, Chl content also reflects the photosynthetic activity in leaves (Szafrańska et al., 2017; Choudhury et al., 2019). In the present study, K high-efficiency vegetable soybean genotypes exhibit lower total Chl content but greater Chl a/b ratio when suffering low K stress. The effect of photosynthetic photon flux density on the leaf Chl a/b ratio is one of the most characteristic differences between sun and shade leaves (Abtahi et al., 2019). Typically, total Chl content per unit leaf area is lower and the Chl a/b ratio is greater in sun compared with shade soybean leaves (Anderson, 1986; Fritschi and Ray, 2007). The higher Chl a/b ratio with enhanced relative content of Chl a in response to low K stress might be an adapted mechanism for K high-efficiency genotypes to maintain photosynthetic capacity because Chl a is the main pigment in leaves that absorbs light energy, which ensures light absorption as much as possible (Abtahi et al., 2019). Similar results were also revealed in the research of Wang et al. (2008). Besides, low K stress decreased leaf K^+ concentration of vegetable soybean, but K high-efficiency genotypes were less affected (Liu et al., 2019a). This could be an important internal factor affecting the chlorophyll regulation ability of K high-efficiency vegetable soybean. The present study recognized that the high efficiency of photosynthesis, including more adaptable photosynthetic parameters and Chl proportion, was essential for K utilization efficiency in K high-efficiency vegetable soybean genotypes.

It is root uptake power that drives up nutrients transport, while photosynthetic capacity and assimilate redistribution capacity are the key to crop yield (He et al., 2019). In the present study, K high-efficiency vegetable soybean was found to have obvious advantages in root K affinity drivers, which ensured the upward supply of K and other nutrients. Thus, the photosynthetic system of K high-efficiency genotypes was less susceptible to low K conditions and has a stronger regulation ability, ensuring the efficiency of K utilization.

CONCLUSION

The higher affinity of root to K^+ associated with root activity under low K stress is essential to promote root K absorption. The strong drivers, represented with higher root bleeding-sap rate and Tr induced by low K stress, ensure the upward flux of K^+ and other essential nutrients as much as possible. The photosynthetic system of K high-efficiency vegetable soybean genotypes is less affected and reasonably regulated by low K stress to maintain photosynthates. Therefore, the ability to redistribute photosynthetic products seemed more important than photosynthetic capacity in K high-efficiency genotypes (**Figure 5**). Overall, crop K high-efficiency should be holistic, and the factors involved are not isolated. Stronger root K affinity drivers associated with photosynthetic adaptability to low K stress were the key factors in determining the K high-efficiency of vegetable soybeans.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

XL and QZ designed the experiments, supervised the study, and revised the manuscript. CL and XW performed the research and wrote the manuscript. HC, BT, and YL helped in planting and data analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This research was partially funded by the National Natural Science Foundation of China (Grant No. 41977096) and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA28070400).

REFERENCES

Abtahi, M., Majidi, M. M., Saeidnia, F., Bahrami, S., and Mirlohi, A. (2019). Genetic and physiological aspects of drought tolerance in smooth bromegrass. *Crop Sci.* 59, 2601–2607. doi: 10.2135/cropsci2019.05.0288

Anderson, J. M. (1986). Photoregulation of the composition, function, and structure of thylakoid membranes. *Annu. Rev. Plant Physiol.* 37, 93–136. doi: 10.1146/annurev.pp.37.060186.000521

Chen, J. J., and Gabelman, W. H. (1995). Isolation of tomato strains varying in potassium acquisition using a sand-zeolite culture system. *Plant Soil* 176, 65–70. doi: 10.1007/BF00017676

Chen, J., Guo, Z., Chen, H., Yang, X., and Geng, J. (2020). Effects of different potassium fertilizer types and dosages on cotton yield, soil available potassium and leaf photosynthesis. *Arch. Agron. Soil Sci.* 67, 1–13. doi: 10.1080/03650340.2020.1817902

Choudhury, S., Larkin, P., Meinke, H., Hasanuzzaman, M. D., Johnson, P., and Zhou, M. (2019). Barley yellow dwarf virus infection affects physiology, morphology, grain yield and flour pasting properties of wheat. *Crop Pasture Sci.* 70, 16–25. doi: 10.1071/CP18364

Claassen, N., and Barber, S. A. (1977). Potassium influx characteristics of corn roots and interaction with N, P, Ca, and Mg influx[1]. *Agronomy J.* 69, 860–864. doi: 10.2134/agronj1977.00021962006900050034x

Crowley, P. H. (1975). Natural selection and the Michaelis constant. *J. Theor. Biol.* 50, 461–475. doi: 10.1016/0022-5193(75)90093-4

Cui, X., Dong, Y., Gi, P., Wang, H., Xu, K., and Zhang, Z. (2016). Relationship between root vigor, photosynthesis and biomass in soybean cultivars during 87 years of genetic improvement in the northern China. *Photosynthetica* 54, 81–86. doi: 10.1007/s11099-015-0160-z

Damon, P. M., and Rengel, Z. (2007). Wheat genotypes differ in potassium efficiency under glasshouse and field conditions. *Aust. J. Agric. Res.* 58, 816–825. doi: 10.1071/AR06402

Daniel, G. C., Rafael, D. S. M., Stelamaris, D. O. P.-M., Humberto, H. D. C., José, T. P., and Enéas, G.-F. (2020). Ammonium nutrition modulates K^+ and N uptake, transport and accumulation during salt stress acclimation of sorghum plants. *Arch. Agron. Soil Sci.* 66, 1991–2004. doi: 10.1080/03650340.2019.1704736

Dev, T. B., Devrim, C., and Herbert, J. K. (2021). Potassium physiology from Archean to Holocene: a higher-plant perspective. *J. Plant Physiol.* 262:153432. doi: 10.1016/j.jplph.2021.153432

Doussan, C., Pierret, A., Garrigues, E., and Loïc, P. (2006). Water uptake by plant roots: II – modelling of water transfer in the soil root-system with explicit account of flow within the root system – comparison with experiments. *Plant Soil* 283, 99–117. doi: 10.1007/s11104-004-7904-z

Duncan, D. R., and Widholm, J. M. (2004). Osmotic induced stimulation of the reduction of the viability dye 2,3,5-triphenyltetrazolium chloride by maize roots and callus cultures. *J. Plant Physiol.* 161, 397–403. doi: 10.1078/0176-1617-01237

Epstein, E., and Hagen, C. E. (1952). A kinetic study of the absorption of alkali cations by barley roots. *Plant Physiol.* 27, 457–474. doi: 10.1104/pp.27.3.457

Fehr, W. R., Caviness, C. E., Burmood, D. T., and Pennington, J. S. (1971). Stage of development descriptions for soybeans, Glycine-max (L.) Merrill. *Crop Sci.* 11, 929–931. doi: 10.2135/cropsci1971.0011183X001100060051x

Fritschi, F. B., and Ray, J. D. (2007). Soybean leaf nitrogen, chlorophyll content, and chlorophyll a/b ratio. *Photosynthetica* 45, 92–98. doi: 10.1007/s11099-007-0014-4

Glass, A. D. (1980). Varietal differences in potassium uptake by barley. *Plant Physiol.* 65, 160–164. doi: 10.1104/pp.65.1.160

Guan, D., Al-Kaisi, M. M., Zhang, Y., Duan, L., Tan, W., Zhang, M., et al. (2014). Tillage practices affect biomass and grain yield through regulating root growth, root-bleeding sap and nutrients uptake in summer maize. *Field Crops Res.* 157, 89–97. doi: 10.1016/j.fcr.2013.12.015

Hao, Y., Lei, J., Wang, Q., Wu, L., and Jiang, C. (2015). Two typical K-efficiency cotton genotypes differ in potassium absorption kinetic parameters and patterns. *Acta Agric. Scand. B Soil Plant Sci.* 65, 45–53. doi: 10.1080/09064710.2014.965194

Hao, Y., Lei, J., Wu, X., Wu, L., and Jiang, C. (2016). Photosynthate transport rather than photosynthesis rate is critical for low potassium adaptation of two cotton genotypes. *Acta Agric. Scand. B Soil Plant Sci.* 66, 170–177. doi: 10.1080/09064710.2015.1069887

He, Z., Xu, C., Liu, B., Yao, B., Wang, H., Chen, Z. Y., et al. (2019). Relationship between photosynthesis, bleeding-sap mass, and bleeding components in maize hybrids and corresponding parents in northern China. *Photosynthetica* 57, 698–704. doi: 10.32615/ps.2019.096

Huang, Y., Li, J., Hua, B., Liu, Z., Fan, M., and Bie, Z. (2013). Grafting onto different rootstocks as a means to improve watermelon tolerance to low potassium stress. *Sci. Hortic.* 149, 80–85. doi: 10.1016/j.scienta.2012.02.009

Jia, Q., Chen, K., Chen, Y., Ali, S., Manzoor, Sohail, A., et al. (2018). Mulch covered ridges affect grain yield of maize through regulating root growth and root-bleeding sap under simulated rainfall conditions. *Soil Tillage Res.* 175, 101–111. doi: 10.1016/j.still.2017.08.017

Liu, C., Tu, B., Li, Y., Tian, B., Zhang, Q., Liu, X., et al. (2017). Potassium application affects key enzyme activities of sucrose metabolism during seed filling in vegetable soybean. *Crop Sci.* 57, 1–11. doi: 10.2135/cropsci2016.08.0648

Liu, C., Tu, B., Wang, X., Li, Y., and Liu, X. (2020). Transcript profile in vegetable soybean roots reveals potential gene patterns regulating K uptake efficiency. *Agronomy* 10:1796. doi: 10.3390/agronomy10111796

Liu, C., Wang, X., Tu, B., Li, Y., Liu, X., Zhang, Q., et al. (2019b). Dry matter partitioning and K distribution of vegetable soybean genotypes with higher potassium efficiency. *Arch. Agron. Soil Sci.* 66, 717–729. doi: 10.1080/03650340.2019.1638508

Liu, C., Tu, B., Wang, X., Jin, J., Li, Y., Zhang, Q., et al. (2019a). Potassium translocation combined with specific root uptake is responsible for the high potassium efficiency in vegetable soybean. *Crop Pasture Sci.* 70, 516–525. doi: 10.1071/CP19042

Liu, R., Yang, C., Zhang, G., Zhang, L., Yang, F., and Guo, W. (2015). Root recovery development and activity of cotton plants after waterlogging. *Agron. J.* 107:2038. doi: 10.2134/agronj14.0567

Liu, X., Nie, Y., Luo, T., Yu, J., Shen, W., and Zhang, L. (2016). Seasonal shift in climatic limiting factors on tree transpiration: evidence from sap flow observations at alpine treelines in southeast Tibet. *Front. Plant Sci.* 7:1018. doi: 10.3389/fpls.2016.01018

Nguyen, T. H., Huang, S., Meynard, D., Chaine, C., Michel, R., Roelfsema, M. R. G., et al. (2017). A dual role for the OsK5.2 ion channel in stomatal movements and K^+ loading into xylem sap. *Plant Physiol.* 174, 2409–2418. doi: 10.1104/pp.17.00691

Pettigrew, W. T. (2008). Potassium influences on yield and quality production for maize, wheat, soybean and cotton. *Physiol. Plant.* 133, 670–681. doi: 10.1111/j.1399-3054.2008.01073.x

Rengel, Z., and Damon, P. M. (2008). Crops and genotypes differ in efficiency of potassium uptake and use. *Physiol. Plant.* 133, 624–636. doi: 10.1111/j.1399-3054.2008.01079.x

Sale, P. W. G., and Campell, L. C. (1987). Differential response to K deficiency among soybean cultivars. *Plant Soil* 104, 183–190. doi: 10.1007/BF02372531

Silberbush, M., and Barber, S. A. (1983). Prediction of phosphorus and potassium uptake by soybeans with a mechanistic mathematical model. *Soil Sci. Soc. Am. J.* 47, 262–265. doi: 10.2136/sssaj1983.03615995004700020018x

Singh, S. K., and Reddy, V. R. (2017). Potassium starvation limits soybean growth more than the photosynthetic processes across CO_2 levels. *Front. Plant Sci.* 8:991. doi: 10.3389/fpls.2017.00991

Song, H. X., and Wang, X. L. (2005). The space distribution of the maize root activity and its absorbing area. *Acta Agric. Boreali-occidentalis Sin.* 14, 137–141. (in Chinese with English abstract),

Szafrańska, K., Reiter, R. J., and Posmyk, M. M. (2017). Melatonin improves the photosynthetic apparatus in pea leaves stressed by paraquat via chlorophyll breakdown regulation and its accelerated de novo synthesis. *Front. Plant Sci.* 8:878. doi: 10.3389/fpls.2017.00878

Teo, Y. H., Beyrouty, C. A., and Gbur, E. E. (1992). Nitrogen, phosphorus, and potassium influx kinetic parameters of three rice cultivars. *J. Plant Nutr.* 15, 435–444. doi: 10.1080/01904169209364331

Tsialtas, I. T., Shabala, S., Baxevanos, D., and Matsi, T. (2017). Cation selectivity in cotton (*Gossypium hirsutum* L.) grown on calcareous soil as affected by potassium fertilization, cultivar and growth stage. *Plant Soil* 415, 331–346. doi: 10.1007/s11104-016-3164-y

Wang, C., Chen, H. F., Hao, Q. N., Sha, A. H., Shan, Z. H., and Chen, L. M. (2012). Transcript profile of the response of two soybean genotypes to potassium deficiency. *PLoS One* 7:e39856. doi: 10.1371/journal.pone.0039856

Wang, J., Zhu, G., Dong, Y., Zhang, H., Rengel, Z., Ai, Y., et al. (2018). Potassium starvation affects biomass partitioning and sink-source responses in three sweet potato genotypes with contrasting potassium-use efficiency. *Crop Pasture Sci.* 69, 506–514. doi: 10.1071/CP17328

Wang, J. D., Wang, H., Zhang, Y., Zhou, J., and Chen, X. (2015). Intraspecific variation in potassium uptake and utilization among sweet potato (*Ipomoea batatas L.*) genotypes. *Field Crop. Res.* 170, 76–82.

Wang, N., Hua, H., Eneji, A. E., Li, Z., Duan, L., and Tian, X. (2012). Genotypic variations in photosynthetic and physiological adjustment to potassium deficiency in cotton (*Gossypium hirsutum*). *J. Photochem. Photobiol. B Biol.* 110, 1–8. doi: 10.1016/j.jphotobiol.2012.02.002

Wang, P., Wang, C., Gao, L., Cui, Y. N., Yang, H. L., Nayana, D. G., et al. (2020). Aliphatic suberin confers salt tolerance to *Arabidopsis* by limiting Na^+ influx, K^+ efflux and water backflow. *Plant Soil* 448, 603–620. doi: 10.1007/s11104-020-04464-w

Wang, W., Li, X. T., Qi, Z. Y., Qu, T. T., and Cao, M. J. (2008). Effect of low potassium stress on photosynthetic and chlorophyll fluorescence parameters of two soybean varieties with different K-efficiency. *Soybean Sci.* 27, 451–455. (in Chinese with English abstract),

Wang, X., Liu, C. K., Tu, B. J., Li, Y. S., Zhang, Q. Y., and Liu, X. B. (2021). Effects of carbon ion beam irradiation on phenotypic variations and biochemical parameters in early generations of soybean plants. *Agriculture* 11:98. doi: 10.3390/agriculture11020098

Wang, X., Zhao, X., Jiang, C., Li, C., Cong, S., Wu, D., et al. (2014). Effects of potassium deficiency on photosynthesis and photo-protection mechanisms in soybean (*Glycine max* (L.) merr.). *J. Integr. Agric.* 14, 856–863. doi: 10.1016/S2095-3119(14)60848-0

Wang, Y., Zhang, Z., Liang, Y., Han, Y., Han, Y., and Tan, J. (2020). High potassium application rate increased grain yield of shading-stressed winter wheat by improving photosynthesis and photosynthate translocation. *Front. Plant Sci.* 11:134. doi: 10.3389/fpls.2020.00134

White, P. J., Bradsha, J. E., Brown, L. K., Dale, M. F. B., Dupuy, L. X., George, T. S., et al. (2018). Juvenile root vigour improves phosphorus use efficiency of potato. *Plant Soil* 432, 45–63. doi: 10.1007/s11104-018-3776-5

Yang, F. (2011). *Potassium Absorption and Use Efficiency and The Physiological Mechanism for Different Type of Cultivars in Cotton*. Yangzhou: Yangzhou University. (in Chinese with English abstract).

Yang, X. E., Liu, J. X., Wang, W. M., Li, H., Luo, A. C., Ye, Z. Q., et al. (2003). Genotypic differences and some associated plant traits in potassium internal use efficiency of lowland rice (*Oryza sativa L.*). *Nutr. Cycl. Agroecosys.* 67, 273–282. doi: 10.1023/B:FRES.0000003665.90952.0c

Zhang, G. P., Chen, J. X., and Eshetu, A. T. (1999). Genotypic variation for potassium uptake and utilization efficiency in wheat. *Nutr. Cycl. Agroecosys.* 54, 41–48.

Zhang, J., Wang, Y., Wang, P., Zhang, Q., Yan, C., Yu, F., et al. (2017). Effect of different levels of nitrogen, phosphorus, and potassium on root activity and chlorophyll content in leaves of *Brassica oleracea* seedlings grown in vegetable nursery substrate. *Hortic. Environ. Biotechnol.* 58, 5–11. doi: 10.1007/s13580-017-0177-2

Zhang, Z., and Yan, W. (2003). *Experimental Guidance of Plant Physiology*. Beijing: Higher Education Press.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 14 December 2021
doi: 10.3389/fpls.2021.768675

Comparative Analysis and Development of a Flavor Fingerprint for Volatile Compounds of Vegetable Soybean Seeds Based on Headspace-Gas Chromatography-Ion Mobility Spectrometry

Fengjie Yuan*, Xujun Fu, Xiaomin Yu, Qinghua Yang, Hangxia Jin and Longming Zhu

Hangzhou Sub-Center of National Soybean Improvement, Institute of Crop and Nuclear Technology Utilization, Zhejiang Academy of Agricultural Sciences, Hangzhou, China

OPEN ACCESS

Edited by:
María Serrano,
Miguel Hernández University of Elche, Spain

Reviewed by:
Mingquan Guo,
Wuhan Botanical Garden, Chinese Academy of Sciences (CAS), China
Baoguo Sun,
Beijing Technology and Business University, China

****Correspondence:***
Fengjie Yuan
fjyuanhz@126.com

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: *01 September 2021*
Accepted: *29 October 2021*
Published: *14 December 2021*

Citation:
Yuan F, Fu X, Yu X, Yang Q, Jin H and Zhu L (2021) Comparative Analysis and Development of a Flavor Fingerprint for Volatile Compounds of Vegetable Soybean Seeds Based on Headspace-Gas Chromatography-Ion Mobility Spectrometry. Front. Plant Sci. 12:768675. doi: 10.3389/fpls.2021.768675

Evaluating the volatile compounds and characteristic fingerprints of the core cultivars of vegetable soybean would provide useful data for improving their aroma in the breeding programs. The present study used headspace-gas chromatography-ion mobility spectrometry (HS-GC-IMS) to evaluate the volatile compounds of vegetable soybean seeds at a specific growth stage. In total, 93 signal peaks were identified, 63 compounds qualitatively, with 14 volatile flavor compounds providing multiple signals. The 63 volatile compounds consisted of 15 esters, 15 aldehydes, 13 alcohols, 15 ketones, one acid, and four other compounds. The peak intensity of most of the volatile compounds varied greatly between the core cultivars. The alcohols and aldehydes determined the basic volatile flavor of the vegetable soybean seeds. Volatile flavors were determined by their respective esters, ketones, or other components. Characteristic fingerprints were found in some core vegetable soybean cultivars. Four cultivars (Xiangdou, ZHE1754, Zhexian 65018-33, and Qvxian No. 1) had pleasant aromas, because of their higher content of 2-acetyl-1-pyrroline (2-AP). A principal component analysis (PCA) was used to distinguish the samples based on the signal intensity of their volatile components. The results showed that the composition and concentration of volatile compounds differed greatly between the core cultivars, with the volatile flavor compounds of soybeans being determined by the ecotype of the cultivar, the direction of breeding selection, and their geographical origin. Characteristic fingerprints of the cultivars were established by HS-GC-IMS, enabling them to be used to describe and distinguish cultivars and their offspring in future breeding studies.

Keywords: vegetable soybean, volatile compounds, fingerprint, HS-GC-IMS, cultivals

INTRODUCTION

Soybean (*Glycine max* L. Merr.) is the most important crop cultivated worldwide. It is a major source of protein and vegetable oil for human and animal consumption and contains several phytochemicals, such as isoflavones and phenolic compounds. Because of its high nutritional value, the soybean is processed into many different products, such as soybean flour, soybean milk, soy sauce, tofu, natto, and other snacks. The vegetable soybean (*Glycine max* L. Merr., also known as edamame) is a food-grade soybean variety that is generally harvested when the pods are fully filled and still green (Zhang and Kyei-Boahen, 2007). Vegetable soybeans are consumed widely in China, Japan, and south-east Asia as a snack food. The main edible part of the vegetable soybean is the seed, which is rich in carbohydrates, proteins, vitamins, minerals, and phytochemicals. Apart from its macronutrients and micronutrients, the dark green color of the vegetable soybean at maturity, its large seed size, soft texture, sweetness, and less beany flavor differentiate it from the regular soybean (Saldivar et al., 2011). Of these attributes, flavor often has the greatest influence on consumer acceptance and behavior. The aromatic vegetable soybean has now become more popular and gained wider acceptance in Japan, the United States, and Europe than the regular soybean (Saldivar et al., 2011). The aromatic type of vegetable soybean commands a higher price than the non-aromatic varieties in international markets, mainly because of its characteristic flavor (Wu et al., 2009). Flavor is perceived primarily by the sense of taste and olfaction (aromatics/aroma) (Glanz et al., 1998), with a unique flavor being associated with a complex mixture of compounds belonging to the different chemical classes (Bravo, 1998). Determining the diversity of these flavor compounds and their contribution to the volatile flavor of vegetable soybean seeds is invaluable for assessing the quality of the soybean at the edible stage (Castada et al., 2019).

The diversity of volatile compounds has been studied in many vegetable and fruit crops, such as rice (Monsoor and Proctor, 2004), soybean (Ramasamy et al., 2019), sorghum (Zanan et al., 2016), melon (Shi et al., 2020), pyrus (Qin et al., 2012), peppers (Ge et al., 2020), and mushroom (Li et al., 2019). The nutritional quality attributes of the vegetable soybean are mainly investigated at specific stages of seed maturity, with its aroma and overall acceptability usually evaluated by its taste (Xu et al., 2016; Jadhav et al., 2018). However, little quantitative information is available for comparing the volatile compounds of a large number of vegetable soybean cultivars at specific stages, with no information available on the composition and ratio of the volatile flavor components in vegetable soybean seeds. The present study will detect the volatile flavor compounds in boiled vegetable soybean seeds, harvested at the R6-R7 growth stage, using headspace-gas chromatography-ion mobility spectrometry (HS-GC-IMS). The effectiveness of ion mobility spectrometry (IMS) is reported as suitable for characterizing the volatile compounds because it can rapidly analyze the samples with low detection limits without pretreatment, is highly sensitive to the compounds with high electronegativity and high proton affinity, and can detect many chemically diverse compounds, such as alcohols, aldehydes, esters, and ketones (Márquezsillero et al., 2011) and has therefore become widely used in the food analysis (Arroyo-Manzanares et al., 2017).

Therefore, the present study aimed to evaluate the volatile compounds of the vegetable soybean at the core cultivar level using HS-GC-IMS. The results will provide a reference for identifying the cultivars and their offspring for improving the taste and flavor of vegetable soybeans as part of a breeding program.

MATERIALS AND METHODS

Plant Materials

In this study, 30 vegetable soybean core cultivars from a breeding program were used. These consisted of eight spring vegetable soybean cultivars with either a high protein content or good flavor released by the Zhejiang local government; three autumn local cultivars with large-sized seeds, two of them with a green cotyledon; eight autumn vegetable soybean varieties with a different seed color and size and good flavor released in Zhejiang province; three imported vegetable soybean cultivars, representing the typical vegetable soybean flavor, to be used as core parents in the breeding program; and eight vegetable soybean breeding lines with different flavors.

All thirty cultivars (**Table 1**) were preserved in the Hangzhou Sub-Center of National Soybean Improvement and sown in the plots following a completely randomized block design with three replications in the experimental field of Zhejiang Academy of Agricultural Sciences, Hangzhou, Zhejiang province, China, in 2019, with each block measuring 1.5 m wide by 10 m long. The orchard management procedures, such as fertilization and irrigation, were same for all the cultivars. The fresh soybean pods were harvested at the R6-R7 stage, with 50 pods collected from each vegetable cultivar block and combined for each of the three replicates. The fresh soybean pod samples were wrapped in aluminum foil then stored at −80°C until the subsequent analyses.

HS-GC-IMS Data Acquisition

The soybean pod samples were first boiled in water at 100°C for 5–8 min, then saved for subsequent analyses using a FlavourSpec gas chromatograph (G.A.S. Gesellschaft für Analytische Sensorsysteme mbH, Dortmund, Germany), equipped with a CombiPal GC autosampler (CTC Analytics AG, Zwingen, Switzerland).

The soybean seed samples (3 g) were placed in 20-ml headspace vials, incubated at 60°C for 15 min spinning at 500 rpm and then, a headspace volume of 400 μl was automatically injected by a syringe at 65°C into an MXT-5 capillary column (15 m, i.d. 0.53). Nitrogen (99.99% purity) was used as the carrier gas programmed as follows: initial flow rate of 2 ml/min, maintained for 2 min, increased to 100 ml/min over 18 min, then maintained at 100 ml/min for 2 min before stopping. The analytes were separated in the column at 60°C then ionized in the IMS ionization chamber at 45°C, with a constant gas flow of 150 ml/min. Furthermore, 2-ketones were used to standardize the instrument as the IMS was not responsive to the alkanes. The

TABLE 1 | The ecotype and origin of the 30 vegetable soybean cultivars used in the present study.

No	Cultivars	Ecotype/origin	No	Cultivars	Ecotype/origin
1	Zheqiudou No. 5	Autumn/Zhejiang	16	ZH716	Spring/Zhejiang
2	Zheqiudou No. 2	Autumn/Zhejiang	17	ZK1754	Spring/Zhejiang
3	Zhechun No. 3	Spring/Zhejiang	18	Taiwan 75	Spring/Taiwan
4	Zhechun No. 4	Spring/Zhejiang	19	TMD	Spring/Zhejiang
5	Zhechun No. 8	Spring/Zhejiang	20	Zhexian No. 12	Spring/Zhejiang
6	Xiangdou	Autumn/Shanghai	21	Zhexian No. 21	Spring/Zhejiang
7	Danbo black	Autumn/Japan	22	Zhexian No. 9	Spring/Zhejiang
8	Qingpiqingren	Autumn/Zhejiang	23	Zhexian 19	Spring/Zhejiang
9	Lvpiqingren	Autumn/Zhejiang	24	Zhexian 2013	Spring/Zhejiang
10	Kaixinlv	Summer/Liaoning	25	Zhexian 77	Spring/Zhejiang
11	Xiaonongqiuyan	Autumn/Zhejiang	26	Zhexian 76004	Spring/Zhejiang
12	Qvxian No. 1	Autumn/Zhejiang	27	Zhexian 6-12	Spring/Zhejiang
13	Zhexian No. 84	Autumn/Zhejiang	28	Zhexian 65018-18	Spring/Zhejiang
14	Zhexian No. 85	Autumn/Zhejiang	29	Zhexian 65018-32	Spring/Zhejiang
15	Huning95-1	Spring/Shanghai	30	Zhexian 65018-33	Spring/Zhejiang

2-ketones (C4-C9) were used to calculate the retention index (RI) of the volatile compounds as an external reference. The volatile compounds were identified by comparing their RI values using the Gas Chromatographic Retention Database in NIST and the Drift Time Database (self-established). The total analysis time was about 20 min.

Data Analysis

All the experiments were performed three times. All data were acquired in the positive ion mode, with each spectrum formed from an average of 12 scans. Three software programs developed by G.A.S were used to view the analytical spectrum and for quantitative analysis. During the first step, the Laboratory Analytical Viewer (v.2.2.1, G.A.S.) and Reporter analysis (v.1.2.12, G.A.S.) were used to compare the 2D top view, 3D spectrogram, and the spectral differences among the samples. In addition, a GC × IMS Library Search (v.1.0.3, G.A.S.) was used for the qualitative analysis of the volatile of compounds based on their retention time in the GC column and drift time (time of flight in the drift tube). The reference RI data were supplied by NIST 2014, and the drift time data by G.A.S. The quantitative analysis was based on the peak height of the selected signal peak using the gallery plot analysis (v.1.0.7, G.A.S.). A principal component analysis (PCA) was used to visualize the differences between the soybean cultivars with Dynamic PCA software (G.A.S.).

A standard curve was established between the peak height and a 2-acetyl-1-pyrroline (2-AP) standard sample (99%). The content of 2-AP in the vegetable soybean cultivars was calculated by using the external standard method.

RESULTS

Analysis of HS-GC-IMS Spectra in Vegetable Soybean Seeds

Headspace-gas chromatography-ion mobility spectrometry three-dimensional (3D) and two-dimensional (2D) spectra were used to analyze the changes and diversity of the volatile flavor compounds of the vegetable soybean samples. The data are represented using a 3D topographical visualization and 2D topographic plot. The differences between the different cultivars were obvious (**Figures 1A,B**). In **Figure 1A**, the X-axis represents the ion drift time of the volatile flavor compounds, the Y-axis represents the gas phase retention time, and the Z-axis represents the peak intensity. The peak signal distributions of the different samples were very similar, but the signal intensity varied, indicating that the content of volatile flavor compounds differed among the samples.

In a further comparison using 2D spectra (**Figure 1B**), the reactive ion peak (RIP) is represented by the red vertical line at the horizontal coordinate of 1.0, with each point on the right side of RIP representing the type of volatile compound, with the retention time of most signals appearing between 100 and 600 s. To compare the differences among the samples, sample ZH716 was used as a reference, with the spectral background of the other samples being subtracted from this reference. After subtraction, the background was white, with the red area indicating that the content of the compound was higher than that of the reference sample, and the blue area indicating that it was lower. **Figure 1B** shows the diversity of volatile flavor compounds among the different samples directly.

Qualitative Analysis of Volatile Flavor Compounds in Vegetable Soybean Seeds by HS-GC-IMS

The spectral topographic plots of the Huning 95-1 cultivar were selected for the qualitative analysis, because all the samples had similar volatile flavor compounds. In **Figure 2**, each dot represents the type of volatile flavor compound. Most of the dots were concentrated in the range of retention times between 100 and 400 s and abscissae between 1 and 2. The compounds identified were numbered, with unmarked dots denoting unidentified compounds. From all the samples, 93 signal peaks were identified and 63 compounds were qualitatively identified using the built-in NIST database and IMS database in the GC-IMS library search (**Table 2**). Fourteen volatile flavor compounds provided multiple signals, such as monomers and dimers. These included methyl octanoate, ethyl hexanoate, ethyl 2-methylpropanoate, ethyl propanoate, (E)-2-octenal, (E)-hept-2-enal, (E)-2-hexen-1-ol, oct-1-en-3-ol, pentan-1-ol, *trans*-2-pentenal, 6-methyl-hept-5-en-2-ol, 3-octanone, 1-octen-3-one, and 2-heptanone. These compounds exhibited similar retention times, but different migration times, related to their content. In the ionization region, the protonated molecules and neutral molecules combined to form dimers, whose quantity could be enhanced by a high content of the compounds (Ewing et al., 1999; Arroyo-Manzanares et al., 2017; Rodríguez-Maecker et al., 2017).

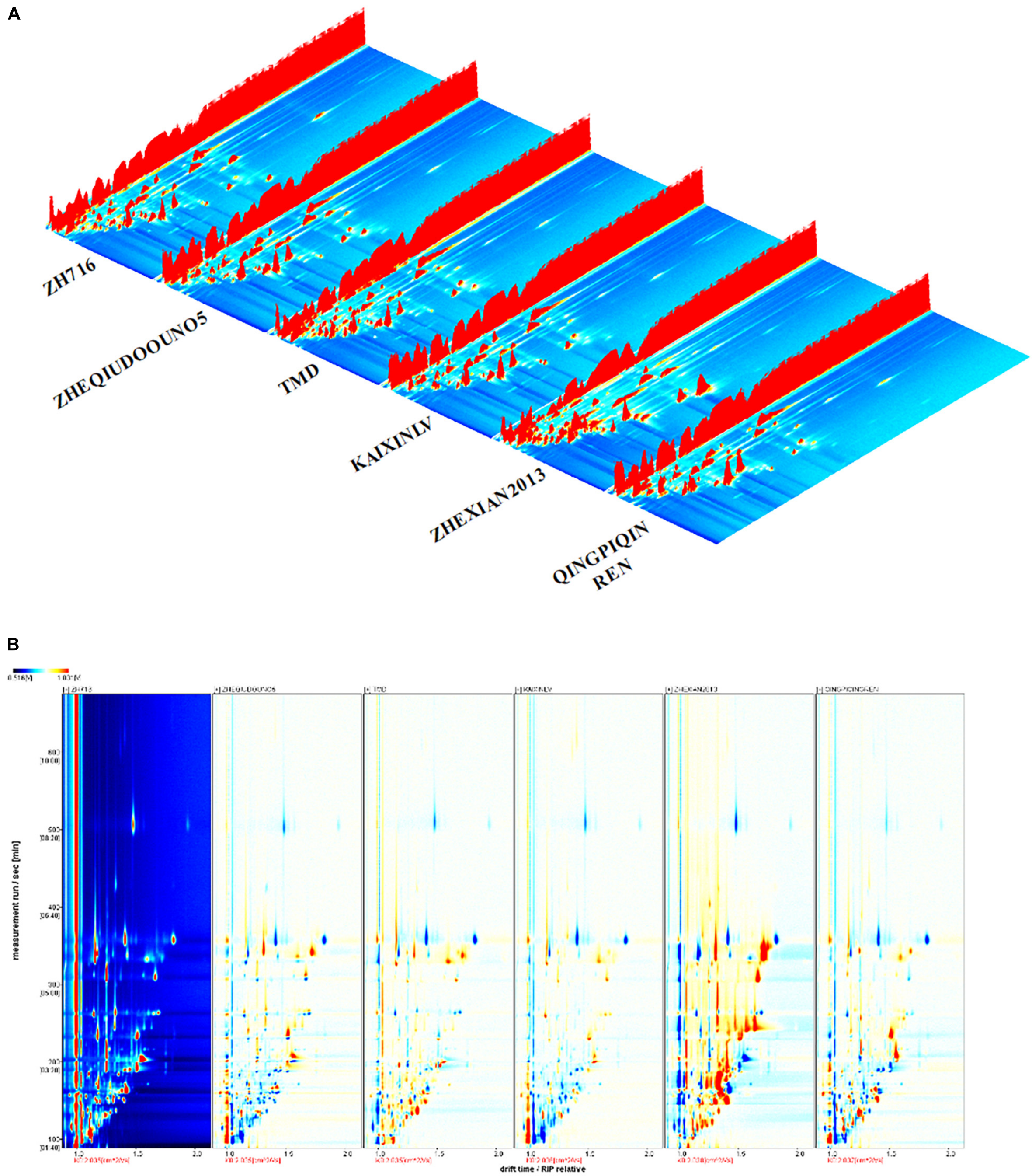

FIGURE 1 | Changes in the volatile compounds from the seeds of six representative vegetable soybean cultivars shown in three-dimensional (3D) **(A)** and two-dimensional (2D) **(B)** topographic forms.

Relative Abundance of Major Volatile Flavor Compounds in the Core Vegetable Soybean Cultivars

A total of 63 diverse volatile compounds were identified by using the GC-IMS, consisting of esters (15), aldehydes (15), alcohols (13), ketones (15), acid (1), and other volatiles (4). The relative peak signal volumes of the volatile compounds from the different cultivars are given in a **Supplementary Table**. The results showed that the peak volumes of most of the volatile compounds varied greatly between the different cultivars.

Of the 15 ester compounds, acetyl esters (9/15), such as ethyl acetate, ethyl hexanoate, ethyl hex-3-enoate, ethyl 2-methylpropanoate, ethyl 2-methylbutanoate, and ethyl nonanoate were predominant. Among the ester compounds, the

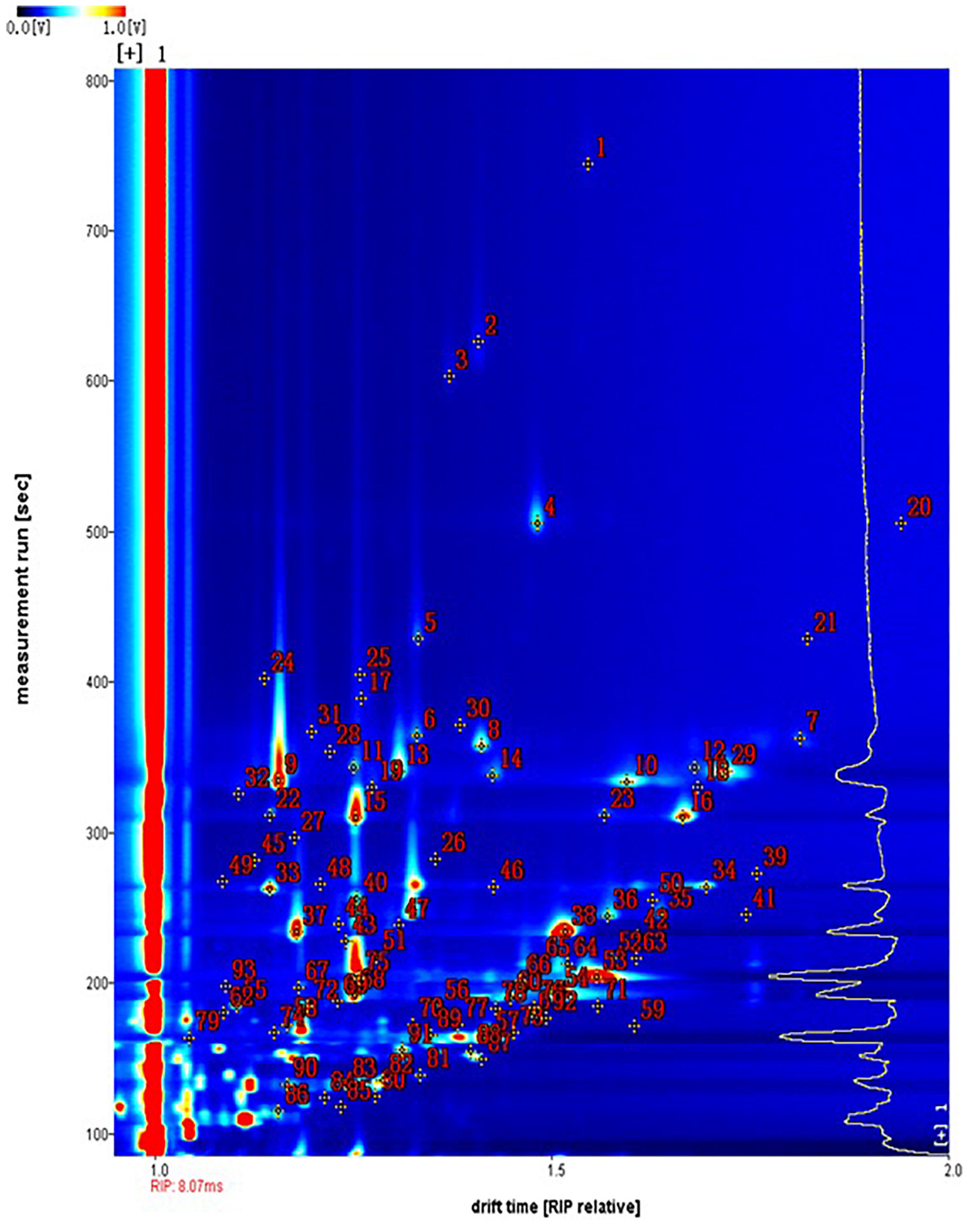

FIGURE 2 | The topographic plots of gas chromatography-ion mobility spectrometry (GC-IMS) spectra obtained from 30 vegetable soybean cultivars. The numbers 1–93 correspond to the detected signals, with those identified listed in **Table 2** and the other numbers representing unidentified signals.

TABLE 2 | Gas chromatography-ion mobility spectrometry (GC-IMS) integration parameters of the volatile compounds detected in the 30 vegetable soybean cultivars used in the present study.

No.	Compounds	CAS N0	Molecule formula	RT	RI	DT
1	Ethyl nonanoate	C123295	C11H22O2	744.318	1,279.2	1.54734
2	2-Heptylfuran	C3777717	C11H18O	625.734	1,195.8	1.40878
3	(E,Z)-2,6-Non-adienal	C557482	C9H14O	603.054	1,179.9	1.373
4	Methyl octanoate-M	C111115	C9H18O2	505.221	1,111.0	1.48412
5	(E)-2-Octenal-M	C2548870	C8H14O	428.289	1,057	1.33344
6	Ethyl hexanoate-M	C123660	C8H16O2	364.378	1,012.1	1.33189
7	Ethyl hexanoate-D	C123660	C8H16O2	362.127	1,010.5	1.8143
9	Oct-1-en-3-ol-M	C3391864	C8H16O	334.082	986.8	1.15717
10	Oct-1-en-3-ol-D	C3391864	C8H16O	333.043	985.5	1.59629
11	6-Methyl-hept-5-en-2-ol-M	C1569604	C8H16O	343.538	997.4	1.25248
12	6-Methyl-hept-5-en-2-ol-D	C1569604	C8H16O	342.717	996.9	1.6822
13	3-Octanone-M	C106683	C8H16O	340.256	994.2	1.30858
15	(E)-Hept-2-enal-M	C18829555	C7H12O	309.572	957.3	1.25503
16	(E)-Hept-2-enal-D	C18829555	C7H12O	309.244	956.9	1.66689
18	1-Octen-3-one-D	C4312996	C8H14O	330.305	982.2	1.68612
19	1-Octen-3-one-M	C4312996	C8H14O	330.125	982	1.27522
20	Methyl octanoate-D	C111115	C9H18O2	505.549	1,111.3	1.94189
21	(E)-2-Octenal-D	C2548870	C8H14O	428.453	1,057.1	1.82428
25	Benzene acetaldehyde	C122781	C8H8O	404.794	1,040.5	1.26018
29	3-Octanone-D	C106683	C8H16O	340.656	994.7	1.72139
30	Ethyl hex-3-enoate-M	C2396830	C8H14O2	371.285	1,017	1.38579
31	(E, E)-2,4-Hheptadienal	C4313035	C7H10O	367.182	1,014.1	1.19904
32	3-Furanmethanol	C4412913	C5H6O2	325.076	957.9	1.10743
33	3,4-Dimethylthiophene	C632155	C6H8S	262.175	900.3	1.14686
34	Heptanal	C111717	C7H14O	263.459	901.8	1.69677
35	n-Hexanol	C111273	C7H14O	245.481	873.7	1.64108
37	(E)-2-Hexen-1-ol-M	C928950	C6H12O	233.41	851.4	1.18027
38	(E)-2-Hexen-1-ol-D	C928950	C6H12O	233.41	851.4	1.51857
39	Amyl acetate	C628637	C7H14O2	272.705	913	1.76081
40	2-Hheptanone-M	C110430	C7H14O	255.754	892.6	1.25684
41	Isoamyl acetate	C123922	C7H14O2	245.224	873.3	1.74689
43	Ethyl 2-methylbutanoate	C7452791	C7H14O2	227.76	840.9	1.24292
44	(Z)-3-hexen-1-ol	C928961	C6H12O	239.528	862.7	1.23289
45	2-Acetyl-1-pyrroline	C85213225	C6H9NO	281.577	923.6	1.12706
46	Allylacetic acid	C591800	C5H8O2	263.539	901.9	1.42771
49	3-Methylthiopropanal	C3268493	C4H8OS	267.253	906.4	1.08724
50	2-Heptanone-D	C110430	C7H14O	255.376	892.1	1.62874
53	Hexanal-D	C66251	C6H12O	203.58	796.1	1.558
54	Pentan-1-ol-D	C71410	C5H12O	192.428	772.3	1.51051
55	*Trans*-2-pentenal-M	C1576870	C5H8O	184.531	752.4	1.10452
56	*Trans*-2-pentenal-D	C1576870	C5H8O	184.472	752.3	1.3591
57	Pentanal	C110623	C5H10O	163.531	699.5	1.42268
61	2-Methylbutan-1-ol	C137356	C5H12O	177.233	734	1.47685
67	2-Hexanone	C591786	C6H12O	196.843	783.4	1.18353
68	Pentan-1-ol-M	C71410	C5H12O	191.273	769.4	1.25306
69	Isobutyl acetate	C110190	C6H12O2	188.204	761.7	1.23193
70	3-Hydroxybutan-2-one	C513860	C4H8O2	171.939	720.7	1.32652
71	Ethyl 2-methylpropanoate-D	C97621	C6H12O2	184.199	751.6	1.56015
72	Ethyl 2-methylpropanoate-M	C97621	C6H12O2	184.812	753.1	1.19454
73	Ethyl propanoate-D	C105373	C5H10O2	166.611	707.3	1.45352
74	Ethyl propanoate-M	C105373	C5H10O2	167.122	708.6	1.15153
75	Hexanal-M	C66251	C6H12O	203.055	795.1	1.25879
79	1-Propene-3-methylthio	C10152768	C4H8S	163.218	698.7	1.04449
80	Butanal	C123728	C4H8O	124.157	566.5	1.27936
81	Ethyl Acetate	C141786	C4H8O2	139.053	618.4	1.33626
83	2-Butanone	C78933	C4H8O	132.102	594.2	1.24322
87	3-Methylbutanal	C590863	C5H10O	149.341	654.2	1.41376
88	2-Methylbutanal	C96173	C5H10O	154.609	672.6	1.40052
89	3-Pentanone	C96220	C5H10O	165.145	703.6	1.35003
90	2,3-Butanedione	C431038	C4H6O2	132.369	595.1	1.16847
91	1-Penten-3-one	C1629589	C5H8O	154.995	674	1.31331
92	3-Methylbutan-1-ol	C123513	C5H12O	176.127	731.2	1.49496
93	2,3-Hexanedione	C3848246	C6H10O2	197.37	784.6	1.09121

-D, *dimer; -M, monomer.*

peak signal of ethyl acetate with its rich content ranged most widely, from 121.98 to 8,003.02 a.u., followed by amyl acetate, methyl octanoate-M and ethyl hexanoate-D, with peak volumes ranging from 13.59 to 1,268.59 a.u., 176.96 to 1,946.78 a.u., and 88.03 to 1,056.93 a.u., respectively. These four ester compounds might contribute to the diversity of the ester notes in the aromas of the 30 vegetable cultivars. Ethyl nonanoate exhibited the narrowest range of peak volumes of all the ester compounds from 52.93 to 136.94 a.u.

Hexanal-D and oct-1-en-3-ol-M had the richest content, with a range of peak volumes of 1,344.21–8,611.99 a.u. and 4,223.95–8,436.49 a.u., respectively.

Ketones, which contribute to the flavor of food, were detected in the vegetable soybean seeds. Compared with the alcohols and aldehydes, the peak volumes of most ketone compounds (10/15) were less than 1,000 a.u. The highest peak volumes of the ketone compounds were for 3-octanone-D and 3-octanone-M, followed by 3-pentanone and 2-butanone. As ester compounds, the ketones in the vegetable soybean seeds provided the characteristic volatile flavor of the different vegetable soybean cultivars.

Five other compounds were evaluated in the present study: allylacetic acid, 1-propene-3-methylthio, 3,4-dimethylthiophene, 2-heptylfuran, and 2-acetyl-1-pyrroline. Then, 2-acetyl-1-pyrroline was detected and quantified by the external reference method. This showed that four cultivars had a higher content of 2-AP: Xiangdou, ZHE1754, Zhexian 65018-33, and Qvxian No. 1 (**Table 3**).

Fingerprints of Cultivars Based on Volatile Substances Using HS-GC-IMS

The volatile compounds in the different vegetable soybean cultivars were analyzed by using the HS-GC-IMS. Fingerprint imaging (**Figure 3**) shows the gallery plot of each sample and their color differences so that the content of volatile compounds can be approximated by the color of each square, with a brighter color representing a higher content of the compound. Each column indicates a signal peak, and each row represents a sample, with three sample repeats. Ninety-three compounds were detected in the 30 samples, with 63 compounds being analyzed qualitatively and quantitatively. Of the 63 compounds, pentan-1-ol, (E)-2-hexen-1-ol, (E)-hept-2-enal, oct-1-en-3-ol, and hexanal were detected and had a comparatively high content in all 30 samples. In contrast, ethyl hex-3-enoate, amyl acetate, isoamyl acetate, and ethyl 2-methylpropanoate-D were not detected in the 30 samples except for Zhexian No. 19. The seeds of this cultivar exhibited the nine strongest signal peaks for ester compounds: ethyl acetate, amyl acetate, ethyl hex-3-enoate-M, ethyl 2-methypropanoate-D, ethyl 2-methypropanoate-M, ethyl 2-methylbutanoate, ethyl propanoate-D, isoamyl acetate, and ethyl propanoate-M. Therefore, these peak signals constituted a very particular fingerprint feature for the Zhexian No. 19 cultivar (**Figure 3**).

The strongest peak signals for ethyl nonanoate, methyl octanoate-D, and methyl octanoate-D were provided by the Huning 95-1 cultivar, therefore, these three volatile flavor compounds could be taken as marker signals for the Huning 95-1 cultivar. The characteristic fingerprint of the Lvpiqingren cultivar consisted of strong peak signals for (E)-2-octenal-D, (E)-2-octenal-M, heptanal, and *trans*-2-pentenal-M. Of the 13 alcohol compounds, the peak signals of (Z)-3-hexen-1-ol, pentan-1-ol-D, and n-hexanol provided a special fingerprint for the Zhechun No. 3, Zhechun No. 4, and Zhechun No. 8 cultivars, which all have a high protein content. The Lvpiqingren and Qingpiqingren cultivars exhibited stronger peak signals for 6-methyl-hept-5-en-2-ol-D and 6-methyl-hept-5-en-2-ol-M than the other cultivars. These two cultivars were summer ecotypes and local germplasms with green cotyledons, an exceptional phenotype compared with the other cultivars.

Ten ketones and one acid were detected in the present study, with the peak signals of 2,3-butanedione, 2-hexanone, and allylacetic acid endowing the characteristic fingerprints for the Zhexian76004 and TMD cultivars. These three compounds exhibited strong peak signals in the Zhexian 65018-33 cultivar, as well as the strongest peak signal for 3-hydroxybutan-2-one. The Lvpiqingren cultivar exhibited a special peak signal for 2-heptanone-D, 2-heptanone-M, and 2-butanone. Of the five other volatile compounds, the peak signal for 2-acetyl-1-pyrroline showed the brightest color from the Zhexian 65018-33 cultivar.

TABLE 3 | 2-acetyl-1-pyrroline (2-AP) content in 30 different vegetable soybean cultivars.

No.	Cultivals	Content of 2-AP (μg/g)	No.	Cultivals	Content of 2-AP (μg/g)
1	Zheqiudou No. 5	–	16	ZH716	–
2	Zheqiudou No. 2	–	17	ZK1754	7.21 ± 0.62
3	Zhechun No. 3	–	18	Taiwan 75	–
4	Zhechun No. 4	–	19	TMD	–
5	Zhechun No. 8	–	20	Zhexian No. 12	–
6	Xiangdou	6.04 ± 0.08	21	Zhexian No. 21	–
7	Danbo black	–	22	Zhexian No. 9	–
8	Qingpiqingren	–	23	Zhexian 19	–
9	Lvpiqingren	–	24	Zhexian 2013	–
10	Kaixinlv	–	25	Zhexian 77	–
11	Xiaonongqiuyan	–	26	Zhexian 76004	–
12	Qvxian No. 1	6.33 ± 1.97	27	Zhexian 6-12	2.02 ± 0.62
13	Zhexian No. 84	–	28	Zhexian 65018-18	–
14	Zhexian No. 85	–	29	Zhexian 65018-32	2.07 ± 0.14
15	Huning95-1	–	30	Zhexian 65018-33	10.76 ± 1.85

PCA Analysis Based on Volatile Substances Detected by HS-GC-IMS

Principal component analysis can reduce the number of dimensions and classify the original data. **Figure 4** shows that the two principal components explained 78% of the total variance: PC1 54% and PC2 24%. The data from the present study were separated into four groups. Group I, consisting of five high-protein soybean cultivars, were poorly correlated with the other vegetable soybean cultivars. The autumn and summer ecotype vegetable soybean cultivars were gathered in the same region to form group II. The spring ecotype vegetable soybean cultivars divided into the groups III and IV, with the flavors between these two groups varying greatly. The six cultivars in group III included Taiwan 75 and Huning 95-1, which were introduced from the Shanghai City and Taiwan regions and were one of the hybrid parents of the other four cultivars. In group IV, six cultivars were bred by crossing local soybean cultivars with cultivars from Japan. Four cultivars, ZK1754, TMD, Zhexian No. 19, and Zhexian 76004, belonged to a single category, because of their special volatile flavor components, which was consistent with the results on the special fingerprints of volatile flavor.

DISCUSSION

HS-GC-IMS Analytical Approach for the Measurement of Volatile Compounds in Plant-Based Products

Several analytical approaches are developed for measuring the volatile compounds in plant-based products. GC-MS has been considered a powerful analytical instrument to identify the chemical. However, time-consuming, complex sample pretreatment, and a significant constraint in the distinguishing of isomeric molecules, particularly ring-isomeric compounds (Kranenburg et al., 2020) limit its application for the plant-based products screening. GC-O-MS is a powerful tool for extracting

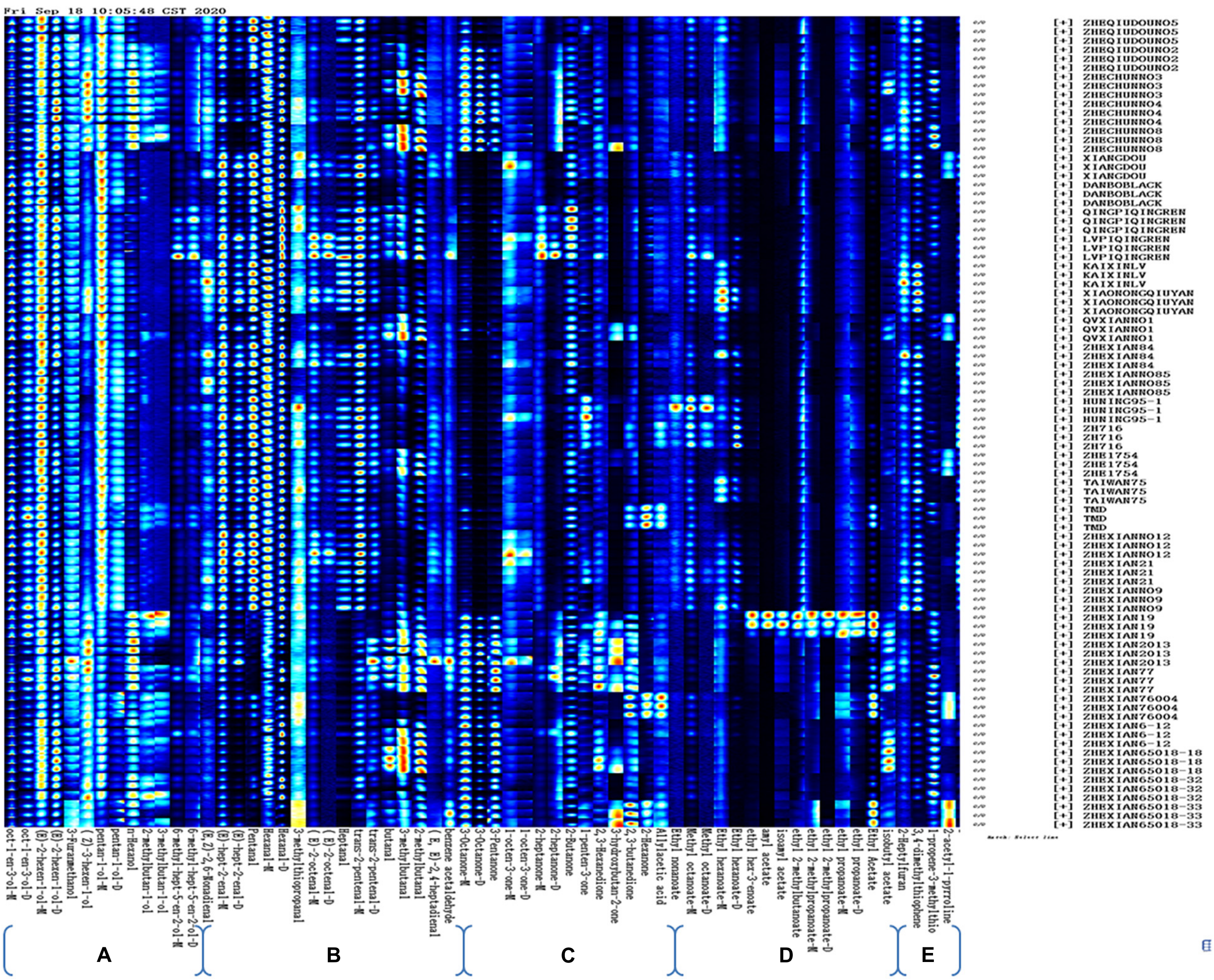

FIGURE 3 | Gallery plot of the signal peak areas obtained from 30 vegetable soybean cultivars. **(A)** Alcohol compounds; **(B)** aldehyde compounds; **(C)** ketone compounds; **(D)** ester compounds; and **(E)** other compounds.

aroma-active compounds from the complex mixtures, because of repetitive time-consuming labor, this method is not ideal for the rapid detection of volatile organic compounds (VOCs) in the plant-based products (Wang et al., 2020). Despite the fact that gas sensors are sensitive at room temperature and have strong selectivity and low detection limits, their detection performances are greatly influenced by moisture and air fluctuation in the environment, resulting in unstable results and poor instrument repeatability (Chen et al., 2018, 2020).

Ion mobility spectrometry is an emerging technique for detecting the trace gases and for characterizing chemical ionic substances based on the differences in the rate of migration of gas phase ions in an electric field (Li et al., 2019). This technique has many advantages over other methods, such as a rapid speed of detection, high sensitivity, easy preparation, and simple sample preparation steps. However, there are also some limitations for IMS, especially for the complex samples (Arce et al., 2014). Combining IMS with GC column could provide a better method for detecting the volatile compounds and has already proved to be useful for analyzing and characterizing the volatile compounds. HS-GC-IMS has been widely and usefully applied for analyzing wine, eggs, jujube fruits, and honey (Garrido-Delgado et al., 2011; Cavanna et al., 2019; Sun et al., 2019; Wang et al., 2019). More importantly, HS-GC-IMS can also be used to establish the visual fingerprints of volatile compounds which show changes in their variety and progress during a food process (Ge et al., 2020). This could be the most powerful function of GC-IMS compared with other analysis methods.

Volatile Compounds in Vegetable Soybean

Various volatile compounds may serve as indicators of the maturity of soybeans and biochemical markers to evaluate seed quality. Several classes of compounds, such as alcohols, aldehydes, esters, and ketones, were the main volatile compounds

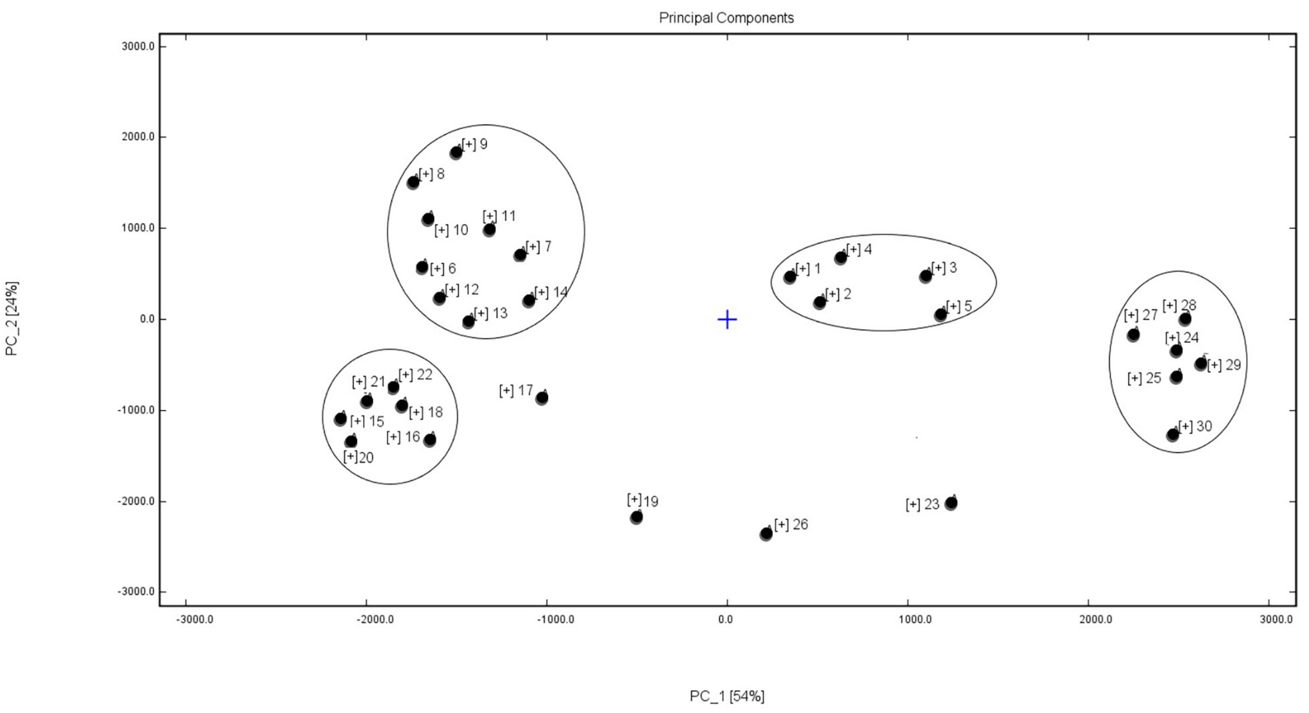

FIGURE 4 | Principal component analysis (PCA) results for volatile flavor compounds from vegetable soybean seeds from 30 different cultivars. The numbers, 1–30, indicate the 30 soybean cultivars shown in **Table 1**.

identified in the present study. Of these compounds, acetyl esters were the main characteristic components, agreeing with the previous studies reporting that esters with fruity notes were the major aromatic components found in the ripe fruit (Lara et al., 2003; Moyaleon et al., 2006; Qin et al., 2012). Aldehydes and alcohols are C6 volatile compounds, often referred to as "green leaf" volatiles because of their green grass note (Yang et al., 2009), and were found to be important background volatiles of vegetable soybeans in the present study. Some previous studies have reported that the light and water characteristics and growth temperatures can significantly affect the formation of volatile metabolites, particularly alcohols and aldehydes in the plants (Bertrand et al., 2012; Benelli et al., 2015). In addition, the present study has found that some alcohols and aldehydes in vegetable soybean seeds exhibited significantly different peak volumes, even though these cultivars were produced in the same environment. These results indicated that the genetic background of the cultivar was another major factor determining the volatile flavor metabolites.

The previous studies have reported that the major volatiles of grain soybeans were ethanol, 1-octen-3-ol, phenylethyl alcohol, hexanal, octanal, 2-propanone, and r-butyrolactone (Lee and Shibamoto, 2000; Boué et al., 2003; Dings et al., 2005; Kim et al., 2020). However, most of these compounds were not detected in the present study except for oct-1-en-3-ol and hexanal, possibly because of the different sampling stages and different soybean processing methods. In the present research, the vegetable soybean seed samples were harvested at the R6 growth stage then the fresh pods were boiled in water. The specific harvesting stage and treatment method also contributed to the presence of some particular volatile compounds. In contrast, the different extraction methods used by HS-GC-IMS could be another factor affecting the composition of volatile flavor compounds. Two other important compounds were found in the present study, furans and 2-acetyl-1-1-pyrroline (2-AP). Furans usually exhibit sweet, burned, and baking odors formed through Maillard reactions (Fischer et al., 2017), and showed a stable peak volume among the cultivars so could have contributed to the basic flavor of the vegetable soybean. 2-acetyl-1-1-pyrroline (2-AP), a 5-membered *N*-heterocyclic ring compound, is identified as the most important compound contributing to the aroma of soybean (Wu et al., 2009; Arikit et al., 2011a,b). It was also detected in the six cultivars where its content was significantly higher than that in the other cultivars, a result consistent with their taste and aroma.

Overall, the soybean cultivars had stronger alcohol and aldehyde component peak signals, but also possessed special ester and ketone signals. In other words, the alcohol and aldehyde compounds determined the basic volatile flavor of the vegetable soybean seeds, with particular volatile flavors being determined by their respective compositions of esters, ketones, or other compounds.

The most abundant volatile flavor compound signals were detected in the spring vegetable soybean cultivars, the Zhexian serial cultivars from Zhexian 12 to Zhexian 65018-33 (**Figure 4**). The autumn soybean cultivars, Zheqiudou No. 2 and Zheqiudou No. 5, exhibited the most similar volatile flavor compound signals. A summer vegetable soybean cultivar, such as Kaixinlv, and some imported vegetable soybean cultivars, such as Taiwan 75, exhibited fairly abundant flavor compound signals. These results could contribute in forming the breeding objectives, and select the direction of breeding and growing environment. The quality objectives for breeding spring vegetable soybean varieties were good taste, flavor, and shape, with the better aromas provided by individual or lines of cultivars being selected by breeders during the selection procedure. In contrast, Zheqiudou No. 2 and Zheqiudou No. 5 were bred for producing the soy food products, such as tofu, soy sauce, and soybean milk, with little attention being paid to the flavor of the fresh seed. Most summer and autumn cultivars were intended for use as vegetables and for producing soy foods, so the flavor of the fresh soybean seeds was given only moderate consideration. However, the growing environment for all the cultivars would have some effect on their volatile flavor compounds.

The use of volatile-compound imaging and determining the markers based on HS-GC-IMS for discriminating among the vegetable soybean cultivars was a non-targeted approach to analysis. This could play an important role in screening for the specific markers, and for extracting reliable, unbiased, and visual information from a large amount of data (Wang et al., 2019). The present study found that some specific fingerprints belonged to different cultivars, that the content of volatile flavor compounds in the different soybean cultivars could be affected by their cultivation method or growing environment, but that some particular volatile flavors (specific fingerprints) could be determined by special genes that could be passed on to their offspring. Therefore, these types of data visualization combined with the results of sensory evaluation could be applied for selecting the volatile flavors and breeding new vegetable varieties with pleasant and acceptable flavors.

PCA Analysis of Volatile Compounds in Vegetable Soybean

A PCA is a multivariate statistical analysis method that linearly transforms several variables to select a smaller number of significant variables (Li et al., 2019). The PCA results on volatile compounds from the 30 vegetable soybean cultivars tended to show a clear separation. The soybean cultivars with a high seed protein content were less correlated with the other vegetable soybean cultivars. The volatile flavor compounds of these high-protein soybean cultivars were different to the middle or lower protein content soybean cultivars. The autumn and summer ecotype vegetable soybean cultivars were located in the same PCA region, thus reflecting the higher correlation of volatile flavor components of these samples. The flavors of the spring ecotype vegetable soybean cultivars varied greatly.

The particular volatile flavor compounds were determined by the ecotype of the cultivars, the direction of breeding selection, and the original cultivation area. The data from HS-GC-IMS contained useful information and could be a reasonably useful tool for distinguishing among the vegetable soybean cultivars. These non-targeted characteristic markers offer the potential for selecting the new vegetable soybean lines with good volatile flavors in the future breeding programs.

In conclusion, 93 volatile compounds were detected in the seeds from 30 vegetable soybean cultivars, of which 63 compounds were detected qualitatively. Alcohol and aldehydes were the predominant volatile compounds followed by esters and ketones. The composition and concentration of volatile compounds differed greatly between the cultivars.

Some characteristic fingerprints of vegetable cultivars, established using HS-GC-IMS, could be used to describe and distinguish cultivars and their offspring in the future studies. Four vegetable cultivars that exhibited an aromatic flavor because of their high 2-AP content would be valuable in the vegetable soybean breeding programs. Based on the PCA analysis, the volatile flavor compounds of the soybean seeds were determined by using the ecotype of the cultivars, the direction of breeding selection, and the geographical origin.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

XF participated in planting vegetable soybean cultivars. XY performed the statistical analysis and helped to draft the manuscript. HJ collected the materials. QY helped for the analysis of the data. LZ helped to perform the HS-GC-IMS experiment. FY designed the study, carried out the HS-GC-IMS experiment and helped in drafting the manuscript. All authors have read and approved this manuscript.

FUNDING

This study was supported financially by the program of the Major Science and Technology Special Program for Crop Breeding in Zhejiang of China (No.2021C02064-5) and the China Agriculture Research System (CARS-04-CES26).

ACKNOWLEDGMENTS

We would like to thank Philip Creed, from Liwen Bianji (Edanz) (www.liwenbianji.cn/), for editing the English text of a draft of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.768675/full#supplementary-material

REFERENCES

Arce, L., Gallegos, J., Garrido, R., Medina, L. M., Sielemann, S., and Wortelmann, T. (2014). Ion mobility spectrometry a versatile analytical tool for metabolomics application in food science. *Curr. Metabol.* 2, 264–271. doi: 10.2174/2213235x03999150212102944

Arikit, S., Yoshihashi, T., Wanchana, S., Tanya, P., Juwattanasomran, R., Srinives, P., et al. (2011a). A PCR-based marker for a locus conferring aroma in vegetable soybean (Glycine max L.). *Theor. Appl. Genet.* 122, 311–316. doi: 10.1007/s00122-010-1446-y

Arikit, S., Yoshihashi, T., Wanchana, S., Uyen, T. T., Huong, N. T. T., Wongpornchai, S., et al. (2011b). Deficiency in the amino aldehyde dehydrogenase encoded by GmAMADH2, the homologue of rice Os2AP, enhances 2-acetyl-1-pyrroline biosynthesis in soybeans (Glycine max L.). *Plant Biotechnol. J.* 9, 75–87. doi: 10.1111/j.1467-7652.2010.00533.x

Arroyo-Manzanares, N., Martín-Gómez, A., Jurado-Campos, N., Garrido-Delgado, R., Arce, C., and Arce, L. (2017). Target vs spectral fingerprint data analysis of Iberian ham samples for avoiding labelling fraud using headspace-gas chromatography–ion mobility spectrometry. *Food Chem.* 246, 65–73. doi: 10.1016/j.foodchem.2017.11.008

Benelli, G., Caruso, G., Giunti, G., Cuzzola, A., Saba, A., Raffaelli, A., et al. (2015). Changes in Olive Oil Volatile Organic Compounds Induced by Water Status and Light Environment in Canopies of Olea Europaea. L. Trees. *J. Sci. Food Agric.* 95, 2473–2481. doi: 10.1002/jsfa.6977

Bertrand, B., Boulanger, R., Dussert, S., Ribeyre, F., Berthiot, L., Descroix, F., et al. (2012). Climatic Factors Directly Impact the Volatile Organic Compound Fingerprint in Green Arabica Coffee Bean as Well as Coffee Beverage Quality. *Food Chem.* 135, 2575–2583. doi: 10.1016/j.foodchem.2012.06.060

Boué, S. M., Shih, B. Y., Carter-Wientjes, C. H., and Cleveland, T. E. (2003). Identification of Volatile Compounds in Soybean at various Developmental Stages using Solid Phase Microextraction. *J. Agric. Food Chem.* 51, 4873–4876. doi: 10.1021/jf030051q

Bravo, L. (1998). Polyphenols: chemistry, dietary sources, metabolism, and nutritional significance. *Nutr. Rev.* 56, 317–333. doi: 10.1111/j.1753-4887.1998.tb01670.x

Castada, H. Z., Hanas, K., and Barringer, S. A. (2019). Swiss cheese flavor variability based on correlations of volatile flavor compounds, descriptive sensory attributes, and consumer preference. *Foods* 8:78. doi: 10.3390/foods8020078

Cavanna, D., Zanardi, S., Dall'Ssta, C., and Suman, M. (2019). Ion mobility spectrometry coupled to gas chromatography: a rapid tool to assess eggs freshness. *Food Chem.* 271, 691–696. doi: 10.1016/j.foodchem.2018.07.204

Chen, W., Deng, F., Xu, M., Wang, J., Wei, Z., and Wang, Y. (2018). GO/Cu2O nanocomposite based QCM gas sensor for trimethylamine detection under low concentrations. *Sensor. Actuator. B Chem.* 273, 498–504. doi: 10.1016/j.snb.2018.06.062

Chen, W., Wang, Z., Gu, S., Wang, J., Wang, Y., and Wei, Z. (2020). Hydrophobic aminofunctionalized graphene oxide nanocomposite for aldehydes detection in fish fillets. *Sensor. Actuator. B Chem* 306:127579.

Dings, L., Aprea, E., and Odake, S. (2005). Comparison of Volatile Flavour Profiles of Kidney Beans and Soybeans by GC-MS and PTR-MS. *Food Sci. Technol. Int.* 11, 63–70. doi: 10.3136/fstr.11.63

Ewing, R. G., Eiceman, G. A., and Stone, J. A. (1999). Proton-bound cluster ions in ion mobility spectrometry. *Int. J. Mass Spectrom* 193, 57–68.

Fischer, M., Wohlfahrt, S., Varga, J., Matuschek, G., Walte, A., and Saraji-Bozorgzad, M. R. (2017). Evolution of volatile flavor compounds during roasting of nut seeds by thermogravimetry coupled to fast cycling optical heating gas chromatography-mass spectrometry with electron and photoionization. *Food Anal. Methods* 10, 1–14.

Garrido-Delgado, R., Arce, L., Guaman, A. V., Pardo, S., Marco, S., and Valcarce, M. (2011). Direct coupling of a gas-liquid separator to an ion mobility spectrometer for the classification of different white wines using chemometrics tools. *Talanta* 84, 471–479. doi: 10.1016/j.talanta.2011.01.044

Ge, S., Chen, Y. Y., Ding, S. H., Zhou, H., Jiang, L. W., Yi, Y. J., et al. (2020). Changes in volatile flavor compounds of peppers during hot air drying process based on headspace-gas chromatography-ion mobility spectrometry. *J. Sci. Food Agric.* 100, 3087–3098. doi: 10.1002/jsfa.10341

Glanz, K., Basil, M., Maibach, E., Goldberg, J., and Snyder, D. (1998). Why Americans eat what they do: taste, nutrition, cost, convenience, and weight control concerns as influences on food consumption. *J. Am. Diet Assoc.* 98, 1118–1126. doi: 10.1016/S0002-8223(98)00260-0

Jadhav, P., Pathan, S., Ingle, K., Moharil, M., Dudhare, M. S., Kale, P., et al. (2018). Stage Specific Seed Nutritional Quality Attributes and Genotyping of Vegetable Soybean Genotypes through DNA-SCoT Profiling. *Int. J. Curr. Microbiol. Appl. Sci.* 1719–1732.

Kim, S. Y., Lee, S. M., Lee, D. Y., Shin, B. K., Kang, D. J., Choi, H. K., et al. (2020). Discrimination of Cultivated Regions of Soybeans (Glycine max) Based on Multivariate Data Analysis of Volatile Metabolite Profiles. *Molecules* 25, 763–779. doi: 10.3390/molecules25030763

Kranenburg, R. F., Verduin, J., Stuyver, L. I., Ridder, R., Beek, A., van, C. E., et al. (2020). Benefits of derivatization in GC-MS-based identification of new psychoactive substances. *Forens. Chem.* 20:100273. doi: 10.1016/j.forc.2020.100273

Lara, I., Miró, R. M., Fuentes, T., Sayez, G., Graell, J., and López, M. L. (2003). Biosynthesis of volatile aroma compounds in pear fruit stored under long-term controlled-atmosphere conditions. *Postharvest. Biol. Tec.* 29, 29–39. doi: 10.1016/s0925-5214(02)00230-2

Lee, K. G., and Shibamoto, T. (2000). Antioxidant Properties of Aroma Compounds Isolated from Soybeans and Mung Beans. *J. Agric. Food Chem.* 48, 4290–4293. doi: 10.1021/jf000442u

Li, M. Q., Yang, R. W., Zhang, H., Wang, S. L., Chen, D., and Lin, S. Y. (2019). Development of a flavor fingerprint by HS-GC-IMS with PCA for volatile compounds of Tricholoma mastrtake Singer. *Food Chem.* 290, 32–39.

Márquezsillero, I., Aguileraherrador, E., Cárdenas, S., and Valcárcel, M. (2011). Determination of 2,4,6-tricholoroanisole in water and wine samples by ionic liquid-based single-drop microextraction and ion mobility spectrometry. *Anal. Chim. Acta* 702, 199–204. doi: 10.1016/j.aca.2011.06.046

Monsoor, M. S., and Proctor, A. (2004). Volatile component analysis of commercially milled head and broken rice. *J. Food Sci.* 69, C632–C636.

Moyaleon, M., Vergara, M., Bravo, C., Montes, M., and Moggia, C. (2006). 1-MCP treatment preserves aroma quality of 'Packham's Triumph' pears during long-term storage. *Postharvest. Biol. Tec.* 42, 85–197.

Qin, G. H., Tao, S. T., Cao, Y. F., Wu, J. Y., Zhang, H. P., and Huang, W. J. (2012). Evaluation of the volatile profile of 33 pyrus ussuriensis cultivars by HS-SPME with GC-MS. *Food Chem.* 134, 2367–2382. doi: 10.1016/j.foodchem.2012.04.053

Ramasamy, R., Ali, T., Durga, K., and Reneth, M. (2019). Rapid profiling of soybean aromatic compounds using electronic nose. *Biosensors* 9, 66–79. doi: 10.3390/bios9020066

Rodríguez-Maecker, R., Vyhmeister, E., Meisen, S., Rosales-Martinez, A., Kuklya, A., and Telgheder, U. (2017). Identification of terpenes and essential oils by means of static headspace gas chromatography-ion mobility spectrometry. *Anal. Bioanal. Chem.* 409, 6595–6603. doi: 10.1007/s00216-017-0613-2

Saldivar, X., Wang, Y. J., Chen, P., and Hou, A. (2011). Changes in chemical composition during soybean seed development. *Food Chem.* 124, 1369–1375. doi: 10.1016/j.foodchem.2010.07.091

Shi, J. D., Wu, H. B., Xiong, M., Chen, Y. J., Chen, J. H., Zhou, B., et al. (2020). Comparative analysis of volatile compounds in thirty nine melon cultivars by head space solid-phase microextraction and gas chromatography-mass spectrometry. *Food Chem.* 316:12642. doi: 10.1016/j.foodchem.2020.126342

Sun, X., Gu, D. Y., Fu, Q. B., Gao, L., Shi, C. H., and Zhang, T. R. (2019). Content variations in compositions and volatile component in jujube fruits during the blacking process. *Food Sci. Nutr.* 7, 1387–1395. doi: 10.1002/fsn3.973

Wang, S., Chen, H., and Sun, B. (2020). Recent progress in food flavor analysis using gas chromatography-ion mobility spectrometry (GC-IMS). *Food Chem.* 315:126158. doi: 10.1016/j.foodchem.2019.126158

Wang, X. R., Yang, S. P., He, J. N., Chen, L. Z., Zhang, J. Z., and Jin, Y. (2019). A green triple-locked strategy based on volatile-compound imaging, chemometrics, and markers to discriminate winter honey and sapium honey using headspace gas chromatography-ion mobility spectrometry. *Food Res. Int.* 119, 960–967. doi: 10.1016/j.foodres.2019.01.004

Wu, M. L., Chou, K. L., Wu, C. R., Chen, J. K., and Huang, T. C. (2009). Characterization and the possible formation mechanism of 2-acetyl-1-pyrroline in aromatic vegetable soybean (Glycine max L.). *J. Food Sci.* 74, 192–197. doi: 10.1111/j.1750-3841.2009.01166.x

Xu, Y. X., Cartier, A., Kibet, D., Jordan, K., Hakala, I., Davis, S., et al. (2016). Physical and nutritional properties of edamame seeds as influenced by stage of development. *Food Measure* 10, 193–200. doi: 10.1007/s11694-015-9293-9

Yang, C. X., Wang, Y. J., Liang, Z. C., Fan, P. G., Wu, B. H., and Yang, L. (2009). Volatiles of grape berries evaluated at the germplasm level by headspace-SPME with GC-MS. *Food Chem.* 114, 1106–1114. doi: 10.1016/j.foodchem.2008.10.061

Zanan, R., Khandagale, K., Hinge, V., Elangovan, M., Robert, J. H., and Nadaf, A. (2016). Characterization of fragrance in sorghum grain and development of a gene-based marker for selection in breeding. *Mol. Breed.* 36, 146–158.

Zhang, L., and Kyei-Boahen, S. (2007). Growth and yield of vegetable soybean (Edamame) in Mississippi. *Horttechnology* 17, 26–31. doi: 10.21273/horttech.17.1.26

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Sustainable Food Systems

ORIGINAL RESEARCH
published: 18 January 2022
doi: 10.3389/fsufs.2021.736247

Assessing Consumer Preferences and Intentions to Buy Edamame Produced in the U.S.

Renata C. V. Carneiro[1†], Tiffany A. Drape[2], Clinton L. Neill[3†], Bo Zhang[4], Sean F. O'Keefe[1] and Susan E. Duncan[1*]

[1] Department of Food Science and Technology, Virginia Tech, Blacksburg, VA, United States, [2] Department of Agricultural, Leadership, and Community Education, Virginia Tech, Blacksburg, VA, United States, [3] Department of Agricultural and Applied Economics, Virginia Tech, Blacksburg, VA, United States, [4] School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, United States

OPEN ACCESS

Edited by:
Uma Tiwari,
Technological University Dublin, Ireland

Reviewed by:
George Lazaroiu,
Spiru Haret University, Romania
Qiuying Zhang,
Northeast Institute of Geography and Agroecology, Chinese Academy of Science (CAS), China

***Correspondence:**
Susan E. Duncan
duncans@vt.edu

†Present addresses:
Renata C. V. Carneiro,
Department of Chemistry and Fermentation Sciences, Appalachian State University, Boone, NC, United States
Clinton L. Neill,
Dyson School of Applied Economics and Management, Cornell University, Ithaca, NY, United States

Specialty section:
This article was submitted to Nutrition and Sustainable Diets, a section of the journal Frontiers in Sustainable Food Systems

Received: 04 July 2021
Accepted: 27 December 2021
Published: 18 January 2022

Citation:
Carneiro RCV, Drape TA, Neill CL, Zhang B, O'Keefe SF and Duncan SE (2022) Assessing Consumer Preferences and Intentions to Buy Edamame Produced in the U.S.. Front. Sustain. Food Syst. 5:736247. doi: 10.3389/fsufs.2021.736247

Due to the growing consumer demand for edamame (vegetable soybean) in the U.S., the domestic production of this specialty crop has been promoted in several Mid-Atlantic and Southeast states as an economically attractive alternative to replace the decreasing tobacco production. For the edamame agrobusiness to be successful in the U.S., consumer studies are as needed as new commercial cultivars that are developed for the U.S. environment. Thus, in this exploratory study, we investigated consumers' preferences and intentions to buy edamame products in the U.S., especially domestic products. Data was collected through a web-based survey distributed through Qualtrics^{XM} and a convenience sampling method was chosen. Volunteers who completed the survey ($N = 309$) were 82% female, 57% residents of the South Atlantic area, and 79% daily consumers of vegetables. Survey respondents had a positive attitude toward domestically produced vegetables and valued supporting U.S. producers. Overall, domestically grown, in-shell edamame products were preferred compared to shelled edamame or imported products. Regarding future purchasing, respondents exhibited higher intention to buy fresh edamame relative to frozen edamame. Additionally, respondents considered price, availability, and familiarity with the vegetable brand, respectively, as the most important factors in their decision-making process to buy edamame products. Our study confirmed there is a market potential for domestically produced edamame and it also provides valuable information to support future studies, production decisions, and the growth of the edamame agrobusiness in the U.S.

Keywords: ***Glycine max*** **(L.) Merr., vegetable soybean, specialty crop, domestic production, survey**

INTRODUCTION

Edamame is the Japanese name for vegetable soybean (*Glycine max* (L.) Merr.), a nutritious vegetable crop widely consumed in Asian countries, mostly as a snack (Mebrahtu and Devine, 2008; Carneiro et al., 2020). In the past two decades, edamame sales and consumption have been increasing in the U.S. (Zhang and Kyei-Boahen, 2007; Wolfe et al., 2018; Neill and Morgan, 2021), which has aroused the interest of breeders, growers, and food processors to produce this specialty crop in the country (Ogles et al., 2016). In the Mid-Atlantic and Southeast areas of the U.S., for example, edamame has been suggested as a promising alternative crop to substitute traditional row crops, such as tobacco (Carson et al., 2011; Neill and Morgan, 2021). Additionally, edamame can be an important ally to increase average consumption of fruit and vegetables in the U.S., which remains lower than recommended by the Dietary Guidelines for Americans (Storey and Anderson, 2018). Nevertheless, several challenges still need to be addressed to reduce the increasing need for imports and promote domestic production of edamame in the U.S. For instance, the development of improved seeds and machinery that reduces the exhaustive nature of labor during production and harvest of edamame, improvements in weed management, the absence of processing facilities, and restricted consumer base and firmly established marketing channels are some major challenges that have been identified in the literature (Zhang and Kyei-Boahen, 2007; Zhang et al., 2017). Although the U.S. is one of the largest soybean producers in the world, most domestic soybeans are targeted to animal feed and food ingredients. Seasonal production and short harvest period also hamper the supply of fresh edamame in the country and increase the demand for imports. As a result, most edamame beans and pods sold in the U.S. are in their frozen form and there is very limited availability of fresh edamame for the domestic market (Montri et al., 2006; Nolen et al., 2016; Wolfe et al., 2018).

Only a few studies have gathered information about preferences, motivations, and purchase intentions of edamame consumers in the U.S. market (Carneiro et al., 2020), which results in a lack of recent information regarding which factors may drive edamame consumption and sales in the country. Relevant studies on consumption and purchase intention of edamame were performed in the metropolitan Philadelphia area, PA in the early 2000s (Kelley and Sánchez, 2005; Montri et al., 2006). However, consumer profile and food trends have significantly changed in the last two decades, which motivated our research group to perform this exploratory consumer study. Recent data can help construct a more accurate picture of current and potential edamame consumers and this information can be used, for example, to support the development of a sustainable edamame agrobusiness in the U.S. Understanding current needs and preferences of domestic vegetable consumers, as well as their motivational factors to introduce, include, and sustain edamame in their shopping and diet (retail, food service, and in-home) is vital to prepare key messages to farmers, processors, health specialists, and consumers at all stages of the agriculture and food systems, including the development of new edamame cultivars through plant-breeding efforts and production decisions (Nardi et al., 2019; Carneiro et al., 2020). In this study, we hypothesized consumers in the U.S. have a positive attitude toward domestically produced edamame and investigated consumers' intention to purchase different types of edamame products. Our goal was to identify consumer preferences that could support business decisions, future research, and potentially promote the domestic production and consumption of edamame.

TABLE 1 | Theories, models, and disciplines for food consumer science.

Theory or model	Discipline/knowledge field
Asymmetry of information	Economics
Economic household models	
Economy of quality	
Food safety economics	
Institutional economics	
Bio-psychological approach	Food science/nutrition
Health belief model	
Theory of reasoned action (TRA)	Psychology
Theory of planned behavior (TPB)	
Food choice	Sociology
Process model	
Theory of linear knowledge transfer or demand driven	Communication
Food supply chain management	Marketing
Corporate social responsibility	

Adapted from Barjolle et al. (2013).

MATERIALS AND METHODS

Survey

Several theories and models have been used as framework to guide food consumer studies (**Table 1**). In this exploratory study, the constructs of the Theory of Planned Behavior (TPB) (Ajzen, 1991), attitudes, subjective norms, and perceived behavior control, were the basis of the questionnaire designed to investigate preferences and intention of consumers in the U.S. to buy and consume edamame products, especially domestically produced products. The TPB states that it is possible to predict the intention one has to perform different behaviors (e.g., food consumption behavior) by considering an individual's attitudes toward behavior (e.g., individual's beliefs about consequences of consuming a specific food product), subjective norms (e.g., beliefs about other people's approval or disapproval to a specific food consumption—social pressure), and perceived behavioral control (e.g., factors that individuals believe would make it easier or harder to consume a specific food). Also, according to Ajzen (1991), a behavior is more likely to be performed when there is a strong intention to perform it (**Figure 1**).

Our survey contained an initial consent-in survey question (question 0), which was followed by 33 exploratory questions. Survey questions were mostly adapted from Vabø and Hansen (2016) and grouped as follows: current food (vegetable/edamame) consumption behavior (1–2), attitudes (3–15), subjective norms (16–21), perceived behavioral control (22–28), purchase intentions (29), and demographics (30–33). Participants were asked to answer most questions (3–24 and 26–28) using a 6-point Likert scale (1 = "strongly disagree"; 6 = "strongly agree"; no neutral point). Questions 1–2, 30–31, and 33 were closed-ended multiple-choice questions, 25 and 29 were ranking questions, and question 32 was an open-ended question.

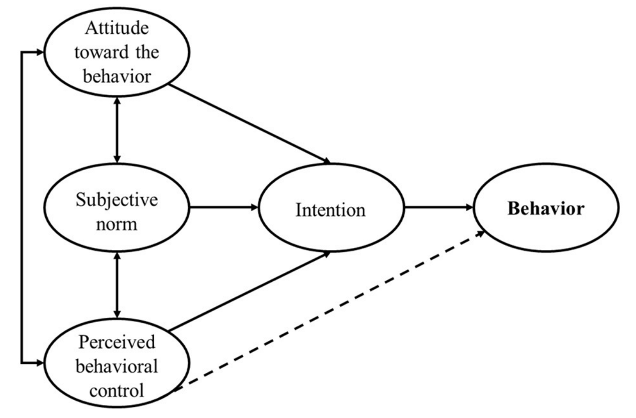

FIGURE 1 | Theory of planned behavior (TPB). Adapted from Ajzen (1991).

Sample

The non-probability convenience sampling method was chosen and participation was based on the volunteer's willingness to take part in the study. Before beginning our investigation, this study obtained ethical approval from the Virginia Tech Institutional Review Board (IRB) for research involving human subjects (IRB 20-023). Then, data collection was open for 7.5 weeks through Qualtrics^{XM} (Qualtrics, Provo, UT), a web-based survey platform. Recruitment of participants occurred through direct posts in the Virginia Tech Daily News and social media, listservs, and direct emails that contained the survey link or QR code; no compensation was provided. A total of 415 volunteers ($\geq$18 years old) gave informed consent before taking part on this study, which occurred online by answering "I agree to participate" in the initial consent-in survey question. However, only 314 participants (75.7%) filled out the whole survey. Although the consent question stated the survey was directed to adults living in the U.S. at the time the survey was taken, 5 volunteers answered on question 32 ("In which state of the United States do you currently live?") they were not living in the U.S. and their results were not considered for statistical analysis. Therefore, a total of 309 complete surveys answered by U.S. residents were considered for statistical analysis.

Data Analysis

The frequencies of participants for demographic questions and behavioral questions were calculated and chi-square test of independence was performed to investigate relationship between two categorical variables (for example, vegetable or edamame consumption and gender). Cronbach's alpha was used to evaluate internal consistency of the TPB questions (attitude, subjective norms, perceived behavioral control). Spearman's rank-order correlation tests were performed to determine the relationship between consumers' choices in questions 25 and 29. A 5% significance level ($\alpha = 0.05$) was considered for statistical analysis, which was performed using JMP® Pro 15.0.0.

RESULTS AND DISCUSSION

Consumer's Self-Identity and Vegetable/Edamame Consumption

The demographic profile (self-identity) of the participants who completed the questionnaire and answered they were living in the U.S. at the time the survey was taken ($n = 309$) is shown in **Table 2**. Most volunteers were self-identified as female (82%), between 21 and 29 years old (36%), residents of the South Atlantic area (57%), and their household income was <$100,000 (64%; <$50,000 = 32%). On average, our participants were more likely to consume vegetables one or more times per day (79%) and consume edamame a few times per year (52%) as part of their diet. A very similar consumption behavior was reported by the participants of our sensory studies performed over 2 consecutive years in Blacksburg, VA (Carneiro et al., 2021); considering the average of both years, our sensory participants, who were mostly female (55%), reported their intentional vegetable consumption was at least once a day (69%) and consumption of edamame was a few times per year (49%). Even though the demand for edamame in the U.S. has been increasing 12–15% annually (Neill and Morgan, 2021), our studies confirmed the frequency in which consumers include edamame in their diets is still low and it suggests the edamame market has potential to continue growing in the country.

The relationship (dependence) between edamame consumption or vegetable consumption, and gender identity, age range, total household income, or U.S. region were tested. Chi-square test only suggested a positive relationship between edamame consumption and gender identity ($p < 0.05$), as well as between vegetable consumption and gender identity ($p < 0.05$). However, our dataset was skewed toward female respondents, which was likely due to the limitations of the convenience sampling method chosen for this study. Previous consumer studies on fruit and vegetable consumption (Emanuel et al., 2012) and on edamame (Kelley and Sánchez, 2005; Montri et al., 2006) that were conducted in the U.S. also showed female volunteers were more likely to participate in the studies. Nevertheless, we suggest that the relationships between gender identity and vegetable or edamame consumption that were reported above should be reviewed with a more balanced dataset. Cultural factors are also recognized as dominant factors in food choice (Steptoe et al., 1995; Pocol et al., 2021) and a previous consumer study on edamame considered reported ethnicity for data analysis, due to the fact edamame is traditionally consumed in Asian countries (Kelley and Sánchez, 2005). In this study, ethnic/racial background of participants was not investigated and it is noted as a potential limitation of the work. Additionally, it is known that consumers consider other food-related associations beyond exclusively satisfying their nutritional needs; lifestyle, healthiness, convenience, and sensory appeal are just some examples of other factors that can be considered when making dietary choices (Barjolle et al., 2013). Our study did not identify psychological or individual traits in consumers, such as lifestyle and healthiness. Thus, we suggest that further studies investigate whether consumers identify themselves as vegetarians or vegans, have concerns regarding their health, or follow a particular

TABLE 2 | Demographic profile of participants.

Characteristic	Answer	Participants*	
		%	Count
Current gender identity	Male	14.6	45
	Female	81.6	252
	Transgender male/trans man/female-to-male (FTM)	0.0	0
	Transgender female/trans woman/male-to-female (MTF)	0.0	0
	Genderqueer, neither exclusively male nor female	1.3	4
	Additional gender category (or other)	1.0	3
	Prefer not to answer	1.6	5
Age range	Under 20 years old	2.9	9
	21–29 years old	36.2	112
	30–39 years old	26.5	82
	40–49 years old	12.3	38
	50–59 years old	12.3	38
	60 years or older	8.1	25
	Prefer not to answer	1.6	5
Geographic region	New England	3.2	10
	Middle Atlantic	6.5	20
	East North Central	9.1	28
	West North Central	3.2	10
	South Atlantic	57.0	176
	East South Central	1.0	3
	West South Central	3.2	10
	Mountain	6.1	19
	Pacific	10.4	32
	Other - Military Okinawa	0.3	1
Total household income	<$50,000	32.4	100
	$50,000–99,999	31.7	98
	$100,000–149,999	14.9	46
	$150,000 or more	10.7	33
	Prefer not to answer	10.4	32

Only completed surveys answered by consumers living in the U.S. at the time the survey was taken were considered (n = 309).

diet. Another limitation that we identified is self-reporting bias. Even though the reliance on self-reported behavior measures is common in this type of consumer studies, it can be seen as a limitation (Carfora et al., 2015).

TPB Constructs: Attitude, Subjective Norms, and Perceived Behavioral Control

Descriptive statistics (means and standard deviations) for TPB constructs (attitude, subjective norms, and perceived behavioral control) are shown in **Table 3**. Cronbach's alpha scores showed "questionable" internal consistency for attitudes (0.68), "acceptable" internal consistency for subjective norms (0.79), and "poor" internal consistency for perceived behavioral control (0.58) (Gliem and Gliem, 2003). Results shown in **Table 3** suggest consumers in the U.S. have a positive impression of domestic vegetables (produced in the U.S.) (mean = 4.8 ± 1.1; Likert scale: 4 = "Somewhat agree" and 5 = "Agree"), as hypothesized. Our survey respondents also had a positive impression of the vegetable seller when choosing edamame (mean = 4.3 ± 0.9). Regarding their attitude toward edamame products, they strongly agreed that it is important to them that edamame is safe for consumption, which means the product does not offer risk of food poisoning or foodborne illness (mean = 5.7 ± 0.6; Likert scale: 6 = "Strongly agree"). A somewhat neutral opinion was observed when participants were asked whether domestically produced edamame is safer for consumption than imported edamame (mean = 3.5 ± 1.2; Likert scale: 3 = "Somewhat disagree" and 4 = "Somewhat agree"). Nevertheless, participants showed highest agreement with the subjective norms that by choosing domestically produced edamame they support domestic producers (mean = 5.1 ± 0.9) and supporting domestic edamame producers is important to them (mean = 4.3 ± 1.1). These results are of great importance because they suggest consumer acceptability of domestic edamame in the U.S. market and justify breeding efforts to develop improved seeds to be grown in the country. However, it is relevant to remember that other factors can also drive edamame acceptability in the U.S., such as sensory characteristics (Carneiro et al., 2021). Recently,

TABLE 3 | Mean likeness and standard deviation of theory of panned behavior (TPB) variables.

TPB variable		Question	Mean*	Standard deviation
Attitudes	3	I have a positive impression of the vegetable seller when choosing edamame (vegetable soybean).	4.3	0.9
	4	I have a positive impression of certified organic vegetables (vegetables produced following USDA organic standards, without genetic modification, ionizing radiation, synthetic herbicides, pesticides, or fertilizers).	4.2	1.3
	5	I have a positive impression of non-GMO vegetables (vegetables without any genetic modification/bioengineering).	3.9	1.5
	6	I have a positive impression of domestic vegetables (vegetables produced in the United States).	4.8	1.1
	7	A positive impression of the vegetable seller is important to me when choosing edamame (vegetable soybean).	4.0	1.2
	8	I have a negative impression of foreign (imported) vegetables in relation to GMO and non-GMO vegetables.	3.0	1.2
	9	It is important to me that edamame (vegetable soybean) is safe for consumption (does not offer risk of food poisoning or foodborne illness).	5.7	0.6
	10	Domestically produced edamame (vegetable soybean) is safer for consumption than imported edamame.	3.5	1.2
	11	A consistent sensory quality (e.g., appearance, flavor, taste, and texture) is important to me to continue to buy edamame (vegetable soybean).	4.9	0.9
	12	Domestically produced edamame (vegetable soybean) has a more consistent sensory quality (e.g., appearance, flavor, taste, and texture) than imported alternatives.	3.3	1.0
	13	I have a negative impression of frozen edamame (vegetable soybeans).	2.3	1.3
	14	Frozen domestically produced edamame (vegetable soybean) is fresher than frozen imported edamame.	3.2	1.2
	15	Edamame (vegetable soybean) beans sold encased in their pods are better than shelled edamame beans.	4.0	1.4
Subjective norms	16	By choosing domestically produced edamame (vegetable soybean) we support domestic producers.	5.1	0.9
	17	People important to me are concerned with upholding domestic food traditions.	3.9	1.4
	18	My family and friends think we should all buy domestic products when possible.	3.9	1.3
	19	Supporting domestic edamame (vegetable soybean) producers is important to me.	4.3	1.1
	20	When buying edamame (vegetable soybean), I take into consideration the support of domestic food traditions.	3.4	1.3
	21	My choice of buying domestically produced edamame (vegetable soybean) will be influenced by the opinions of my family and friends.	2.9	1.3
Perceived behavioral control	22	There is wide range of domestically produced edamame (vegetable soybean) available in the places where I do my grocery shopping.	2.7	1.2
	23	I think that frozen edamame (vegetable soybean) is reasonably priced (in pod = $1.69/10 oz package and shelled = $1.69/8 oz package).	4.6	0.9
	24	Domestic edamame (vegetable soybean) is available at the grocery store where I most frequently purchase vegetables.	3.6	1.5
	26	I would pay more for domestically produced edamame (vegetable soybean) relative to non-domestically produced edamame.	3.7	1.2
	27	I would pay more for certified organic edamame (vegetable soybean) relative to non-certified organic edamame.	3.3	1.5
	28	I would pay more for certified non-GMO edamame (vegetable soybean) relative to genetically modified edamame.	3.0	1.5

**6-point Likert scale (1 = "strongly disagree"; 3 = "slightly disagree"; 4 = "slightly agree";6 = "strongly agree"; no neutral point). Only completed surveys answered by consumers living in the U.S. at the time the survey was taken were considered (n = 309).*

Flores et al. (2019) reported that higher scores in overall liking, texture and appearance of different edamame cultivars increased the odds of consumers answering "yes" in their question about purchase intention. In this study, respondents agreed that a consistent sensory quality (e.g., appearance, flavor, taste, and texture) is important to them to continue to buy edamame (mean = 4.9 ± 0.9), but slightly disagreed that domestically produced edamame has a more consistent sensory quality than imported alternatives (mean = 3.3 ± 1.0). These results reinforce the importance of applying sensory evaluation methods to guide edamame development, which was discussed in previous review article (Carneiro et al., 2020) and is currently being done by

TABLE 4 | Important factors and product characteristics considered by consumers during decision process to buy edamame.

Factor or product characteristic	Importance in purchase decision process—rank*								Total count	Factor rank
	1		2		3		4			
	%	Count	%	Count	%	Count	%	Count		
Price	39.1	106	30.6	83	14.4	39	15.9	43	271	1
Availability	38.3	95	27.8	69	19.8	49	14.1	35	248	2
Familiarity with the vegetable brand	7.6	11	20.7	30	38.6	56	33.1	48	145	3
Frozen edamame in the pods	16.2	19	21.4	25	29.9	35	32.5	38	117	4
Domestically produced (not imported)	11.6	13	22.3	25	21.4	24	44.6	50	112	5
Fresh (not frozen) edamame pods	21.8	22	29.7	30	26.7	27	21.8	22	101	6
Certified organic	18.5	17	16.3	15	34.8	32	30.4	28	92	7
Frozen shelled edamame beans (not in pods)	13.3	8	20.0	12	36.7	22	30.0	18	60	8
Non-GMO	21.4	12	28.6	16	25.0	14	25.0	14	56	9
Fresh (not frozen) shelled edamame beans	15.6	5	12.5	4	34.4	11	37.5	12	32	10
Imported only	50.0	1	0.0	0	0.0	0	50.0	1	2	11

1 = The most important factor or product characteristic. Only completed surveys answered by consumers living in the U.S. at the time the survey was taken were considered (n = 309).

breeding programs in Arkansas and Virginia (Carneiro et al., 2021).

In addition to the positive impression of domestic vegetables, participants expressed a positive impression of certified organic vegetables produced following USDA organic standards (mean = 4.2 ± 1.3), and a slightly positive impression of non-GMO vegetables, which are vegetables without any genetic modification/ bioengineering (mean = 3.9 ± 1.5). Although the statement was about vegetables in general, not only edamame, a stronger positive attitude was also expected by the researchers toward non-GMO. It is important to consider one limitation of our study that may have impacted the answer to this question was the use of a 6-point Likert scale, which did not offer a neutral choice to the respondents and forced them to choose between "somewhat disagree" (3) and "somewhat agree" (4). Additionally, our respondents slightly agreed they would pay more for domestically produced edamame (mean = 3.7 ± 1.2), but somewhat disagreed they would pay more for certified organic edamame (mean = 3.3 ± 1.5) or certified non-GMO edamame (mean = 3.0 ± 1.5). This result differs from a recent economics study conducted in Arkansas that reported significantly higher willingness to pay for edamame labeled as non-genetically modified (Wolfe et al., 2018). Therefore, further applied economics studies are suggested to investigate consumer willingness-to-pay for the different edamame products described above (domestic vs. imported, certified organic vs. non-certified organic, GMO vs. non-GMO).

Although our survey respondents somewhat agreed that domestic edamame is available at the grocery store where they most frequently purchase vegetables (mean = 3.6 ± 1.5), they somewhat disagreed that there is a wide range of domestically produced edamame available in the places where they do their grocery shopping (mean = 2.7 ± 1.2; Likert scale: 2 = "Disagree"). This perception is in accordance with the fact the U.S. market is mostly supplied by imports and the domestic production is still low (Neill and Morgan, 2021). Although a negative impression of frozen edamame was not confirmed, our survey respondents agreed that edamame beans sold encased in their pods are better than shelled edamame beans (mean = 4.0 ± 1.4). Moreover, respondents tended to agree that frozen edamame is reasonably priced (in pod = $1.69/10 oz package and shelled = $1.69/8 oz package) (mean = 4.6 ± 0.9). Thus, the prices above can serve as references for future economic studies and for growers and producers who aim to sell their edamame products directly to consumers.

Edamame Purchase Intent

Edamame is a vegetable with high nutritional value and a vegetarian/vegan-friendly source of protein (Carneiro et al., 2020). In fact, Kelley and Sánchez (2005) reported that most participants of their consumer studies expressed a positive likelihood to purchase edamame after they were informed about its health benefits. According to Nardi et al. (2019), consumers have a stronger tendency to turn their intention into consumption when the food choice regards healthy or hedonic products. Our participants ranked the four most important factors and product characteristics in their decision-making process to buy edamame and results are shown in **Table 4**. Price was the most important factor for the participants, and it was followed by availability and familiarity with the vegetable brand. Price was also identified by Montri et al. (2006) as one of the main factors that affect edamame consumers' decision to buy a new produce from supermarkets in the early 2000s; it had similar average score to "sample of the produce at supermarket," but it was rated below "friend's recommendation." Drugău-Constantin (2019) also reported "recommendation from a friend/family/known acquaintance" as the major factor that influences U.S. consumers' purchase decision. Although "friend's recommendation" was not a factor that we directly investigated in our study, our participants showed disagreement (mean = 2.9 ± 1.3; Likert scale: 2 = "Disagree" and 3 = "Somewhat disagree") with the statement "my choice of buying domestically

TABLE 5 | Correlation between edamame purchase factors or product characteristics ranked by consumers in the U.S.

Factor or product characteristic	By factor or product characteristic	Count (Pairs)	Spearman ρ	Prob > $\lvert\rho\rvert$ [a]
F2 = Familiarity with the vegetable brand	F1 = Price	125	−0.0545	0.5464
F3 = Availability	F1 = Price	219	−0.3456	**<0.0001**
	F2 = Familiarity with the vegetable brand	119	−0.2218	**0.0153**
F4 = Certified organic	F1 = Price	70	−0.3951	**0.0007**
	F2 = Familiarity with the vegetable brand	31	−0.3148	0.0846
	F3 = Availability	62	−0.2963	**0.0193**
F5 = Non-GMO	F1 = Price	42	−0.5763	<**0.0001**
	F2 = Familiarity with the vegetable brand	18	−0.2047	0.4151
	F3 = Availability	29	−0.4822	**0.0081**
	F4 = Certified organic	31	0.1209	0.5172
F6 = Fresh (not frozen) edamame pods	F1 = Price	91	−0.3099	**0.0028**
	F2 = Familiarity with the vegetable brand	28	−0.5038	**0.0063**
	F3 = Availability	68	−0.4600	**<0.0001**
	F4 = Certified organic	21	−0.4488	**0.0413**
	F5 = Non-GMO	13	−0.0840	0.7850
F7 = Fresh (not frozen) shelled edamame beans	F1 = Price	28	−0.4919	**0.0078**
	F2 = Familiarity with the vegetable brand	4	0.5774	0.4226
	F3 = Availability	20	−0.1953	0.4092
	F4 = Certified organic	5	−0.2294	0.7105
	F5 = Non-GMO	4	−0.5000	0.5000
	F6 = Fresh (not frozen) edamame pods	14	0.3457	0.2260
F8 = Frozen edamame in the pods	F1 = Price	97	−0.4236	**<0.0001**
	F2 = Familiarity with the vegetable brand	47	−0.5049	**0.0003**
	F3 = Availability	100	−0.3497	**0.0004**
	F4 = Certified organic	21	−0.3977	0.0742
	F5 = Non-GMO	7	−0.3268	0.4744
	F6 = Fresh (not frozen) edamame pods	30	0.0381	0.8414
	F7 = Fresh (not frozen) shelled edamame beans	1	-	-
F9 = Frozen shelled edamame beans (not in pods)	F1 = Price	50	−0.3224	**0.0224**
	F2 = Familiarity with the vegetable brand	24	−0.3289	0.1165
	F3 = Availability	51	−0.3498	**0.0119**
	F4 = Certified organic	10	−0.3651	0.2996
	F5 = Non-GMO	6	−0.6155	0.1934
	F6 = Fresh (not frozen) edamame pods	4	0.8165	0.1835
	F7 = Fresh (not frozen) shelled edamame beans	8	−0.5429	0.1644
	F8 = Frozen edamame in the pods	14	−0.1066	0.7168
F10 = Domestically produced (not imported)	F1 = Price	90	−0.4055	**<0.0001**
	F2 = Familiarity with the vegetable brand	38	−0.2596	0.1155
	F3 = Availability	74	−0.4845	**<0.0001**
	F4 = Certified organic	25	−0.3690	0.0695
	F5 = Non-GMO	18	−0.3694	0.1314
	F6 = Fresh (not frozen) edamame pods	34	0.1544	0.3832
	F7 = Fresh (not frozen) shelled edamame beans	12	−0.0632	0.8454
	F8 = Frozen edamame in the pods	32	0.0127	0.9448
	F9 = Frozen shelled edamame beans (not in pods)	13	−0.2842	0.3467
F11 = Imported only	F1 = Price	1	-	-
	F2 = Familiarity with the vegetable brand	1	.	-
	F3 = Availability	2	−1.0000	-
	F4 = Certified organic	0	-	-
	F5 = Non-GMO	0	-	-
	F6 = Fresh (not frozen) edamame pods	0	-	-
	F7 = Fresh (not frozen) shelled edamame beans	0	-	-
	F8 = Frozen edamame in the pods	2	−1.0000	-
	F9 = Frozen shelled edamame beans (not in pods)	0	-	-
	F10 = Domestically produced (not imported)	0	-	-

[a] *Significant correlation is indicated by numbers in bold.*

TABLE 6 | Rank of edamame products most likely to be purchased by consumers in the U.S.

Edamame product	Purchase intent—rank*								Total count	Product rank
	1		2		3		4			
	%	Count	%	Count	%	Count	%	Count		
Fresh (not frozen) edamame pods	39.9	91	22.4	51	20.2	46	17.5	40	228	1
Frozen edamame in the pods	37.6	76	26.7	54	19.3	39	16.3	33	202	2
Domestically produced (not imported)	14.1	28	24.2	48	37.4	74	24.2	48	198	3
Certified organic	24.9	43	17.3	30	30.6	53	27.2	47	173	4
Fresh (not frozen) shelled edamame beans	11.6	19	29.3	48	28.0	46	31.1	51	164	5
Frozen shelled edamame beans (not in pods)	25.2	37	27.9	41	20.4	30	26.5	39	147	6
Non-GMO	13.1	14	33.6	36	16.8	18	36.4	39	107	7
Imported only	5.9	1	5.9	1	17.6	3	70.6	12	17	8

**1 = The most likely to buy product. Only completed surveys answered by consumers living in the U.S. at the time the survey was taken were considered (n = 309).*

produced edamame (vegetable soybean) will be influenced by the opinions of my family and friends" (**Table 3**). Edamame in the pods was the most important product characteristic and frozen edamame products were ranked higher than similar fresh ones. Even though consumers may believe fresh fruits and vegetables are healthier or tastier, convenience is an important advantage of the frozen fruits/vegetable products that can affect consumer purchase decision (Storey and Anderson, 2018) and was possibly valued by our respondents. Convenience can be linked, for example, to easier selection, purchase, food preparation (e.g., prewashed), cooking, as well as cleaning before and/or after cooking (Storey and Anderson, 2018), and these components could be further investigated in future research. Moreover, among the 11 factors presented to rank, "domestically produced (not imported)" was ranked as the fifth most important factor, while "imported only" was ranked as the least important factor, which is consistent with the results presented in **Table 3** and confirms a preference for domestic products (produced in the U.S.). A significant negative correlation was observed between 18 pairs of factors (**Table 5**), which means that when the rank of one factor increased, the rank of the other factor decreased. For instance, the rank of price was negatively correlated with the rank of all other factors and product characteristics ($p < 0.05$), except for "familiarity with the vegetable brand" ($p > 0.05$) and "imported only" (insufficient pairs).

Next, our participants ranked the four edamame products they were most likely to buy on the day the survey was taken. Overall, shelled edamame products ranked lower than edamame in the pods (**Tables 4**, **6**). The four products chosen by most consumers were: (1) fresh (not frozen) edamame in the pods, (2) frozen edamame in the pods, (3) domestically produced (not imported), and (4) certified organic (**Table 6**). Even though consumers perceived a low availability of domestically produced edamame where they do their grocery shopping (**Table 3**, question 22), "imported only" was the least chosen purchase option, which reinforces previous suggestion that domestic edamame products (not imported) are valued by consumers in the U.S. (**Table 6**). Furthermore, our purchase intention results are aligned with the consumer study performed by Montri et al. (2006) in the metropolitan Philadelphia area. Authors reported that consumers would prefer to buy fresh edamame in-shell (in the pods) only (48.5%) or both shelled and in the pods (48.5%), instead of shelled (beans) only (3%). Also, most participants in their study reported they were more likely to buy edamame because it was produced in Pennsylvania. Shelled beans may not be the consumer most preferred option for edamame products, but they may be valuable for food service menus (e.g., salad bars). Consumer attitude toward food service retail selection options preferences was not investigated in our survey and future study can help identify alternative routes of sales for the different edamame products. As shown in **Table 3**, our participants slightly agreed that they would pay more for domestically produced edamame relative to non-domestically produced edamame (mean = 3.7 ± 1.2; Likert scale: 3 = "Somewhat disagree" and 4 = "Somewhat agree"). Thus, future studies are important to quantify how much consumers are willing to pay for different types of domestic edamame. For instance, Wolfe et al. (2018) reported consumers in the U.S. are willing to pay more (at least $0.42 more) for non-GMO labeled edamame in comparison with unlabeled or GMO labeled edamame. However, in our study, non-GMO edamame was ranked as the seventh edamame product most likely to be purchased by our participants from our list of 8 products (**Table 6**) and ranked nineth from our list of 11 factors considered during the purchase decision process. It suggests that when analyzed within a broader set of factors (not directly related), consumers may value other factors, such as convenience, over production characteristics, such as non-GMO, for example. A significant negative correlation was observed between 17 pairs of edamame products (**Table 7**). The rank of domestically produced (not imported) edamame was negatively correlated with the rank of all other factors and product characteristics ($p < 0.05$), except for "frozen shelled edamame beans (not in pods)" and "imported only" ($p > 0.05$). However, as "important only" was not chosen by many respondents, the low number of pairs possibly affected statistical analysis.

TABLE 7 | Correlation between edamame products ranked by consumers in the U.S.

Variable	By variable	Count (Pairs)	Spearman ρ	Prob>|ρ|[a]
P2 = Non-GMO	P1 = Certified organic	87	0.1837	0.0885
P3 = Fresh (not frozen) edamame pods	P1 = Certified organic	107	−0.4170	**<0.0001**
	P2 = Non-GMO	57	−0.3976	**0.0022**
P4 = Fresh (not frozen) shelled edamame beans	P1 = Certified organic	66	−0.5466	**<0.0001**
	P2 = Non-GMO	31	−0.4277	**0.0164**
	P3 = Fresh (not frozen) edamame pods	125	−0.0026	0.9770
P5 = Frozen edamame in the pods	P1 = Certified organic	90	−0.4184	**<0.0001**
	P2 = Non-GMO	51	−0.5164	**0.0001**
	P3 = Fresh (not frozen) edamame pods	153	−0.3423	**<0.0001**
	P4 = Fresh (not frozen) shelled edamame beans	86	−0.5249	**<0.0001**
P6 = Frozen shelled edamame beans (not in pods)	P1 = Certified organic	51	−0.2673	0.0579
	P2 = Non-GMO	28	−0.6971	**<0.0001**
	P3 = Fresh (not frozen) edamame pods	98	−0.6228	**<0.0001**
	P4 = Fresh (not frozen) shelled edamame beans	97	−0.2245	**0.0270**
	P5 = Frozen edamame in the pods	96	−0.1176	0.2539
P7 = Domestically produced (not imported)	P1 = Certified organic	113	−0.3013	**0.0012**
	P2 = Non-GMO	66	−0.4608	**<0.0001**
	P3 = Fresh (not frozen) edamame pods	133	−0.3311	**<0.0001**
	P4 = Fresh (not frozen) shelled edamame beans	84	−0.3447	**0.0013**
	P5 = Frozen edamame in the pods	117	−0.2113	**0.0222**
	P6 = Frozen shelled edamame beans (not in pods)	67	−0.1879	0.1278
P8 = Imported only	P1 = Certified organic	5	−0.5526	0.3340
	P2 = Non-GMO	1	-	-
	P3 = Fresh (not frozen) edamame pods	11	-	-
	P4 = Fresh (not frozen) shelled edamame beans	3	−1.0000	**<0.0001**
	P5 = Frozen edamame in the pods	13	0.0000	1.0000
	P6 = Frozen shelled edamame beans (not in pods)	4	−0.5000	0.5000
	P7 = Domestically produced (not imported)	14	0.1788	0.5407

[a] *Significant correlation is indicated by numbers in bold.*

CONCLUSION

Our study suggested there is a positive attitude toward domestically produced edamame in the U.S. market, as hypothesized, and identified a higher purchase intention for domestic edamame products. A higher preference and purchase intention were also identified for in-shell edamame products. Edamame is a specialty crop not yet largely produced in the U.S. and our findings provide valuable insights to support future studies that can help promote a sustainable growth of the edamame agrobusiness in the country. As price was identified as the major factor in the consumer decision-making process to buy edamame in the U.S., future economic studies that investigate willingness-to-pay for a diverse set of domestic edamame products (e.g., frozen vs. fresh, in-shell vs. shelled, certified organic vs. non-certified organic, non-GMO vs. GMO) can help local growers and processors develop their business strategies. Likewise, the market potential for organic production of edamame needs to be further investigated. The fact that consumers value a consistent sensory quality of edamame products reinforces the need for developing improved cultivars and standardizing production practices with the support of sensory data. Additionally, exploration of other value-added processed products with edamame as a major ingredient could possibly expand the interest in and motivation for edamame and further investigation is recommended. As a positive relationship between edamame consumption and gender identity was suggested, it would be valuable to further explore the impact of self-identity (e.g., self-reported gender, age, or ethnic/racial background) in the TPB variables as well as in the future consumption of edamame products in the U.S. (food choice behavior). Moreover, as a possible alternative to enrich the TPB model and complement our findings, complementary constructs such as risk perception, trust, and past behaviors could also be further explored in the context of investigating consumption of domestic edamame.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Virginia Tech Institutional Review Board (IRB) for research involving human subjects (IRB 20-023). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RC, TD, CN, and SD contributed with conception of the study and questionnaire design. RC collected and analyzed consumer data, wrote first draft, and edited manuscript. Manuscript was reviewed and approved by all authors.

FUNDING

This work was funded by USDA-NIFA (Grant No. 2018-51181-28384; Accession No. 1016465), and, in part, by the Virginia Agricultural Experiment Station.

ACKNOWLEDGMENTS

The authors thank all the volunteers who participated in this study.

REFERENCES

Ajzen, I. (1991). The theory of planned behavior. *Org. Behav. Hum. Dec. Proc.* 50, 179–211. doi: 10.1016/0749-5978(91)90020-T

Barjolle, D., Gorton, M., Đorđević, J. M., and Stojanović, Ž. (2013). *Food Consumer Science*. Berlin: Springer.

Carfora, V., Caso, D., and Conner, M. (2015). The role of self-identity in predicting fruit and vegetable intake. *Appetite* 106, 23–29. doi: 10.1016/j.appet.2015.12.020

Carneiro, R., Duncan, S., O'Keefe, S., Yu, D., Huang, H., Yin, Y., et al. (2021). Utilizing consumer perception of edamame to guide new variety development. *Front. Sustain. Food Syst.* 4:556580. doi: 10.3389/fsufs.2020.556580

Carneiro, R. C. V., Duncan, S. E., O'Keefe, S. F., Yin, Y., Neill, C. L., and Zhang, B. (2020). Sensory and consumer studies in plant breeding: a guidance for edamame development in the U.S. *Front. Sustain. Food Syst.* 4:124. doi: 10.3389/fsufs.2020.00124

Carson, L., Freeman, J., Zhou, K., Welbaum, G., and Reiter, M. (2011). Cultivar evaluation and lipid and protein contents of Virginia-grown edamame. *Horttechnology* 1, 131–135. doi: 10.21273/HORTTECH.21.1.131

Drugău-Constantin, A. (2019). Is consumer cognition reducible to neurophysiological functioning? *Econom. Manag. Finan.l Mark.* 14, 9–14. doi: 10.22381/EMFM14120191

Emanuel, A. S., Mccully, N., Gallagher, K. M., and Updegraff, J. A. (2012). Theory of Planned Behavior explains gender difference in fruit and vegetable consumption. *Appetite* 59, 693–697. doi: 10.1016/j.appet.2012.08.007

Flores, D., Giovanni, M., Kirk, L., and Liles, G. (2019). Capturing and explaining sensory differences among organically grown vegetable-soybean varieties grown in Northern California. *J. Food Sci.* 84, 613–22. doi: 10.1111/1750-3841.14443

Gliem, J. A., and Gliem, R. R. (2003). "Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales," in: *Midwest Research to Practice Conference in Adult, Continuing, and Community Education,* 82–88. Available online at: https://scholarworks.iupui.edu/bitstream/handle/1805/344/gliem+&+gliem.pdf?sequence=1

Kelley, K. M., and Sánchez, E. S. (2005). Accessing and understanding consumer awareness of and potential demand for edamame. *HortScience* 40, 1347–1353. doi: 10.21273/HORTSCI.40.5.1347

Mebrahtu, T., and Devine, T. E. (2008). Combining ability analysis for selected green pod yield components of vegetable soybean genotypes (*Glycine max*). *N Zeal. J. Crop Horticult. Sci.* 36, 97–105. doi: 10.1080/01140670809510225

Montri, D. N., Kelley, K. M., and Sánchez, E. S. (2006). Consumer interest in fresh, in-shell edamame and acceptance of edamame-based patties. *HortScience* 41, 1616–1622. doi: 10.21273/HORTSCI.41.7.1616

Nardi, V. A. M., Jardim, W. C., Ladeira, W., and Santini, F. (2019). Predicting food choice: a meta-analysis based on the theory of planned behavior. *Br. Food J.* 121, 2250–2264. doi: 10.1108/BFJ-08-2018-0504

Neill, C. L., and Morgan, K. L. (2021). Beyond scale and scope: exploring economic drivers of u.s. specialty crop production with an application to edamame. *Front. Sustain. Food Syst*. 4:582834. doi: 10.3389/fsufs.2020.582834

Nolen, S., Zhang, B., and Kering, M. K. (2016). Increasing fresh edamame bean supply through season extension techniques. *J. Horticult.* 3, 1–5. doi: 10.4172/2376-0354.1000170

Ogles, C. Z., Guertal, E. A., and Weaver, D. B. (2016). Edamame cultivar evaluation in Central Alabama. *Agron. J.* 108, 2371–2378. doi: 10.2134/agronj2016.04.0218

Pocol, C. B., Marinescu, V., Dabija, D.-C., and Amuza, A. (2021). Clustering Generation Z university students based on daily fruit and vegetable consumption: empirical research in an emerging market. *Br. Food J.* 123, 2705–2727. doi: 10.1108/BFJ-10-2020-0900

Steptoe, A., Pollard, T. M., and Wardle, J. (1995). Development of a measure of the motives underlying the selection of food: the food choice questionnaire. *Appetite* 25, 267–284. doi: 10.1006/appe.1995.0061

Storey, M., and Anderson, P. (2018). Total fruit and vegetable consumption increases among consumers of frozen fruit and vegetables. *Nutrition* 46, 115–121. doi: 10.1016/j.nut.2017.08.013

Vabø, M., and Hansen, H. (2016). Purchase intentions for domestic food: a moderated TPB-explanation. *Br. Food J.* 118, 2372–2387. doi: 10.1108/BFJ-01-2016-0044

Wolfe, E., Popp, M., Bazzani, C., Nayga, R. M., Danforth, D., Popp, J., et al. (2018). Consumers' willingness to pay for edamame with a genetically modified label. *Agribusiness* 34, 283–299. doi: 10.1002/agr.21505

Zhang, L., and Kyei-Boahen, S. (2007). Growth and yield of vegetable soybean (edamame) in Mississippi. *HortTechnology* 17, 26–31. doi: 10.21273/HORTTECH.17.1.26

Zhang, Q., Li, Y., Chin, K. L., and Qi, Y. (2017). Vegetable soybean: Seed composition and production research. *Italian J. Agron.* 12, 276–282. doi: 10.4081/ija.2017.872

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Sustainable Food Systems

ORIGINAL RESEARCH
published: 29 April 2022
doi: 10.3389/fsufs.2022.738453

Understanding the Role of Overall Appearance and Color in Consumers' Acceptability of Edamame

Renata Carneiro [1†], Kyle Adie [1], Dajun Yu [1], Mariah Beverly [2], Clinton Neill [2†], Bo Zhang [3], Thomas Kuhar [4], Steven Rideout [5], Mark Reiter [6], Haibo Huang [1], Sean O'Keefe [1] and Susan Duncan [1*]

[1] Department of Food Science and Technology, Virginia Tech, Blacksburg, VA, United States, [2] Department of Agricultural and Applied Economics, Virginia Tech, Blacksburg, VA, United States, [3] School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, United States, [4] Department of Entomology, Virginia Tech, Blacksburg, VA, United States, [5] Department of Plant Pathology, Physiology, and Weed Science, Virginia Tech, Painter, VA, United States, [6] School of Plant and Environmental Sciences, Eastern Shore Agricultural Research and Extension Center, Virginia Tech, Painter, VA, United States

Appearance properties of vegetables can affect consumers' acceptance toward them as well as purchase intent. Hence, appearance is highly associated with quality of edamame (*Glycine max* (L.) Merr.), a protein-rich vegetable that is experiencing increased sales in the USA. Edamame is a high-value specialty crop and its production has been promoted in Virginia and other states in the USA where the tobacco production has decreased in the last decade. To support current efforts to develop the edamame industry in the USA, consumer and color data of 10 edamame genotypes grown in Virginia were analyzed in this follow-up study to understand the role of overall appearance and color characteristics in consumers' acceptability of edamame beans. In two consecutive years, untrained adult volunteers used 9-point hedonic scales (1 = "dislike extremely", 9 = "like extremely") to evaluate appearance and overall liking of edamame samples (cooked and shelled edamame beans) and our researchers measured the reflective color of the samples with a chroma meter. In the first year, sensory panelists also completed a choice-based conjoint analysis to determine their willingness-to-pay (WTP) for dark vs. light green edamame beans in a salad. Edamame genotypes were significantly different in appearance and overall liking ($p < 0.05$) and the genotype R14-16195 was the most liked overall. Hedonic scores and color were significantly affected by "year" ($p < 0.05$), so intentional changes between years (e.g., sample preparation) should be avoided in future studies. Consumers showed higher WTP for dark green edamame beans. Additionally, green intensity (color index) and a* color coordinate were correlated to appearance liking scores ($p < 0.05$), which suggests color data can support breeding selection criteria and possibly predict consumer acceptability. Employing color measurement as quality control method can help improve harvest procedures, post-harvest handling, and define edamame quality standards for the USA market.

Keywords: sensory, willingness-to-pay (WTP), vegetable soybean, *Glycine max* (L) Merr, specialty crop

OPEN ACCESS

Edited by:
*Rakesh Bhardwaj,
National Bureau of Plant Genetic Resources (ICAR), India*

Reviewed by:
*Alexandru Rusu,
Biozoon Food Innovations GmbH, Germany*
*Kuan-Chi Wang,
Academia Sinica, Taiwan*
*Ram Kumar Choudhary,
Dr. Rajendra Prasad Central Agricultural University, India*

Correspondence:
*Susan Duncan
duncans@vt.edu*

† Present addresses:
*Renata Carneiro,
Department of Chemistry and Fermentation Sciences, Appalachian State University, Boone, NC, United States*
*Clinton Neill,
Dyson School of Applied Economics and Management, Cornell University, Ithaca, NY, United States*

Specialty section:
This article was submitted to Nutrition and Sustainable Diets, a section of the journal Frontiers in Sustainable Food Systems

Received: *08 July 2021*
Accepted: *04 April 2022*
Published: *29 April 2022*

Citation:
*Carneiro R, Adie K, Yu D, Beverly M, Neill C, Zhang B, Kuhar T, Rideout S, Reiter M, Huang H, O'Keefe S and Duncan S (2022) Understanding the Role of Overall Appearance and Color in Consumers' Acceptability of Edamame.
Front. Sustain. Food Syst. 6:738453.
doi: 10.3389/fsufs.2022.738453*

INTRODUCTION

Edamame (vegetable soybean; *Glycine max* (L.) Merr.) is a lucrative food commodity of major economic importance in Asian countries; it is a popular appetizer in Japan and has been largely produced and exported by China and Taiwan for the last five decades (Wang, 2018). More recently, edamame consumption has been increasing significantly in the USA, which has driven plant-breeding researches to improve this high-value specialty crop for domestic production (Carneiro et al., 2021). Moreover, as tobacco production decreases in the country, edamame has been suggested as an alternative crop to farmers in Virginia and other tobacco producing states (Carneiro et al., 2020; Neill and Morgan, 2021). Current efforts to breed new edamame varieties for sustainable production in the USA are also justified by the fact that most seeds available to domestic farmers that aim to grow this nutritious vegetable are expensive, came from other countries, or were not bred to be grown in major production regions of the country (Lord et al., 2021).

Color and appearance are often the initial criteria used by consumers to judge quality of fruits and vegetables and they can influence food consumption and purchases (Lawless and Heymann, 2010; Pathare et al., 2013). Accordingly, appearance is one of the major categories that define quality of edamame, which is frequently associated to green color and shape of pods and seeds, for example (Masuda, 1991; Moseley et al., 2021). The green color in edamame can be considered an indicator of freshness, in opposition to the yellow color. Yellowing indicates freshness decline in edamame and it is associated to degradation of free amino acids, sugars and ascorbic acid (Masuda, 1991). This information is relevant when handling both types of edamame products available in the market: fresh and frozen. Amilia et al. (2021) reported that brightness and green color of frozen salted edamame decreased during storage at room temperature while yellowness first increased then decreased when products turned brownish, which shows that color measurement can help determining shelf life, as well as best storage conditions and thawing practices.

Color of fruits and vegetables is associated with natural pigments (e.g., chlorophylls, carotenoids, anthocyanins, flavonoids) and this visual property can change as consequence of plant maturation, ripening, spoilage, processing, or packaging (Pathare et al., 2013). Visual analysis based on color and appearance standards can be useful tools to select genotypes, monitor post-harvest quality, determine best practices, and can also be used to define premium prices of edamame. Indeed, recent study suggested edamame genotypes that develop dark (black or brown) seed coat at maturity may not be the best option for production (Moseley et al., 2021). Besides color, other appearance properties that can be evaluated in foods include size, shape, visual surface texture, reflectance, glossiness, turbidity, and translucency (Lawless and Heymann, 2010). In addition to the green color of pods and beans, good shape and a spotless surface without defects are also valued in edamame (Carneiro et al., 2020). However, different than shape or size, color is an appearance property that can be an indicator of other sensory properties of foods, such as flavor and texture. Overall, color measurement requires simpler and cheaper equipment than flavor or texture analysis and does not require highly trained analysts. It is also faster and cheaper than sensory evaluation tests for quality control, so it can be a valuable and affordable tool to breeders, food producers and processors committed to the development of the edamame varieties that can perform well in the USA climate and environment, which is vital to assure a long-term sustainable production.

There are multiple coordinate systems that can be used to describe color of fruits and vegetables, such as Hunter *L a b*, RGB (red, green and blue), and the Commission Internationale de l'Eclairage (CIE) $L^*a^*b^*$, and colorimeters and spectrophotometers are examples of instruments that can be used to measure color coordinates (Pathare et al., 2013). The impact of color and other appearance properties in the overall acceptability of edamame still requires further investigation to guide best practices and the development of quality standards based on the needs and expectations of the growing USA market. Therefore, in this follow-up study we aimed to further understand the role of appearance liking and color characteristics of edamame beans in consumers' acceptability and willingness-to-pay (WTP) for edamame products. The information provided in this study complements our previous report on how consumer perception of edamame has been used to guide new variety development for sustainable domestic (USA) production in Virginia, Arkansas, Mississippi, and Missouri (Carneiro et al., 2021). This study provides support and guidance to current and future research on edamame.

MATERIALS AND METHODS

Edamame Samples

As reported in our previous publication (Carneiro et al., 2021), an initial set of 20 edamame genotypes (2 cultivars and 18 advanced breeding lines) was analyzed by consumers in the first years of our multistate breeding research program that aims to develop new edamame varieties for production in the USA. Sensory data (overall liking and appearance liking) of a subset of 10 edamame genotypes was further analyzed and its correlation with instrumental data (color measurement) was investigated in this follow-up study. The 10 selected genotypes were: 4 advanced breeding lines from the Virginia Tech breeding program (V10-3653, V16-0524, V16-0528, and V16-0547), 5 advanced breeding lines from the University of Arkansas breeding program (R14-16195, R14-6238, R14-6450, R15-10280, and R16-5336), and the cultivar UA-Kirksey (control). These 10 genotypes were selected from the initial set of 20 genotypes because they were the ones that successfully grew in two Virginia locations, Blacksburg (Virginia Tech's Kentland Farm) and Painter (Eastern Shore Agricultural Research and Extension Center), in the first 2 years of our research (2018 and 2019) and were analyzed by sensory panelists in both years.

Sample Preparation

Sample preparation for sensory panels was previously reported in Carneiro et al. (2021). Blanched edamame samples (boiling

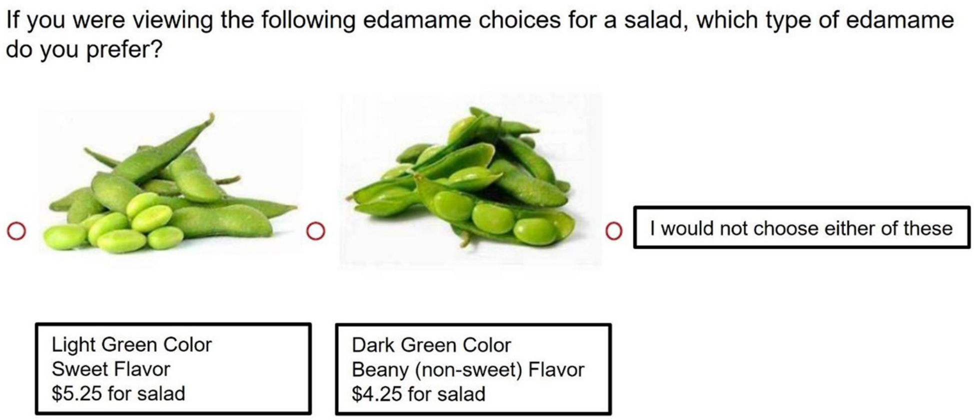

FIGURE 1 | Example of a choice question answered by consumers in the choice-based conjoint analysis to determine their willingness-to-pay (WTP) for edamame products.

distilled water; 98°C for 1 min) were manually shelled, and healthy beans were packed (re-sealable plastic bags) and frozen stored (−20°C) for sensory evaluation. The day before each sensory panel, frozen edamame beans were transferred to a refrigerator (4°C) to thaw (minimum of 4 h) then cooked in microwave oven (SHARP, model R-2W38, serial n°34328, 120 VAC, 60 Hz, Thailand, 1996). In the first year, edamame beans (200–250 g) were cooked for 4 min (high power) in the same re-sealable plastic (polyethylene) bags used for storage. In the second year, edamame beans (100 g) were cooked 1.5 min (high power) in a glass microwave-safe container covered with a paper towel sheet and remained covered in the container for another minute. In both years, cooked edamame beans cooled to room temperature (∼21°C), then a sample was separated for color measurement and the other beans were placed in 2-ounce plastic cups with lids for sensory evaluation. Samples were kept refrigerated until served.

Color Measurement

Reflective color of the edamame beans (as prepared for sensory evaluation) was measured in CIE L*a*b* (CIELAB) color scale [L* = darkness (0) or lightness (100); a* = greenness (–) or redness (+); b* = blueness (–) or yellowness (+)] using a Minolta CR-300 Chroma Meter (Minolta Co., Osaka, Japan). In the first year, 100 g of edamame beans were placed inside a clear plastic bag and lined on top of a plane surface (laboratory benchtop) for color measurement. In the second year, 50 g of edamame beans were placed in a plastic petri dish plate. In both years, the equipment was previously calibrated with a standard white plate (L* = 96.77, a* = 0.45, b* = 2.37), then it was positioned over the beans aligned on the top layer. Edamame beans were shuffled between measurements, and L*a*b* values were measured in triplicate, averaged, and recorded. Next, green intensity (GI = - a*/b*) was calculated for each sample (Flores et al., 2019) and correlations between color and sensory data were investigated.

Sensory Evaluation and Consumer Studies

Consumer studies were approved by the Virginia Tech Institutional Review Board (IRB) for Research Involving Human Subjects (IRB 18-310). Sensory tests were performed as described in our previous report (Carneiro et al., 2021). Overall liking (OL) and appearance liking (AL) of edamame samples were evaluated by untrained adult volunteers ($N = 50–53$ per session) that were not allergic to soy and participated in one or more sensory panels. Participants used 9-point hedonic scales ranging from "dislike extremely" (1) to "like extremely" (9) to evaluate shelled edamame beans prepared as described in section sample preparation (edamame pods were not evaluated by the sensory panelists). Next, correlation between OL and AL of edamame was investigated.

In the first year of the sensory evaluation study, participants ($N = 188$) were also asked to complete a choice-based conjoint analysis to determine their WTP for edamame products with different characteristics, such as color and taste. Color was varied at two levels (dark vs. light green) as well as flavor (beany vs. sweet). Price for a salad containing edamame beans was varied at three levels. Every choice question contained two options plus a "no buy" option. The experiment was designed using a fractional factorial which resulted in 10 choice questions per participant. To simulate the choice of color, pictures were used that had distinct color variation between light and dark green edamame beans. The color difference was also stated in words under each picture. An example of a choice question answered by consumers is shown in **Figure 1**.

Statistical Analysis

The effects of edamame genotype (10 genotypes listed in section edamame samples), growing location (Blacksburg and Painter, VA), year of research (1 and 2), and interaction factors ("genotype*location," "genotype*year," "location*year," "genotype*location*year") on color (L*a*b* values) and sensory evaluation responses (OL and AL) were tested. The following

TABLE 1 | Overall liking, appearance liking, and color characteristics of edamame genotypes.

Edamame sample			Sensory evaluation[a]		Color[b]			
Genotype	**Location**	**Year**	**Overall liking**	**Appearance liking**	**L***	**a***	**b***	**GI (–a*/b*)**
R14-16195	Blacksburg	1	6.44 ± 1.59	5.63 ± 1.72	53.73	−19.26	40.20	0.48
		2	6.20 ± 1.85	5.84 ± 1.53	49.06	−22.01	39.55	0.56
	Painter	1	6.02 ± 1.84	5.73 ± 1.86	50.05	−20.57	38.67	0.53
		2	6.92 ± 1.45	6.64 ± 1.60	50.54	−21.52	39.41	0.55
R14-6238	Blacksburg	1	5.67 ± 1.49	5.40 ± 1.67	55.13	−17.91	43.69	0.41
		2	5.88 ± 1.61	5.76 ± 1.86	49.42	−22.21	40.37	0.55
	Painter	1	6.10 ± 1.45	6.44 ± 1.45	59.17	−22.10	41.63	0.53
		2	6.56 ± 1.63	6.88 ± 1.48	48.29	−22.48	39.89	0.56
R14-6450	Blacksburg	1	5.92 ± 1.57	5.71 ± 1.81	50.61	−18.63	38.03	0.49
		2	6.02 ± 1.70	6.26 ± 1.50	49.23	−21.43	37.64	0.57
	Painter	1	5.88 ± 1.90	5.94 ± 1.58	49.66	−20.64	38.12	0.54
		2	6.22 ± 1.36	5.74 ± 1.79	48.84	−22.40	37.99	0.59
R15-10280	Blacksburg	1	5.98 ± 1.80	5.76 ± 1.57	52.89	−20.48	40.74	0.50
		2	6.64 ± 1.56	6.42 ± 1.67	48.74	−21.84	38.70	0.56
	Painter	1	6.38 ± 1.41	6.38 ± 1.56	50.22	−19.48	38.25	0.51
		2	6.04 ± 1.71	5.58 ± 1.55	50.18	−20.26	39.01	0.52
R16-5336	Blacksburg	1	5.04 ± 1.91	5.24 ± 1.71	54.85	−18.45	41.40	0.45
		2	6.08 ± 1.55	5.58 ± 1.64	51.25	−21.83	41.45	0.53
	Painter	1	6.16 ± 1.62	6.44 ± 1.54	54.40	−22.10	42.72	0.52
		2	6.74 ± 1.31	6.16 ± 1.63	49.33	−21.67	38.87	0.56
UA-Kirksey (check)	Blacksburg	1	6.62 ± 1.40	6.56 ± 1.47	54.47	−22.02	44.06	0.50
		2	5.93 ± 1.66	6.24 ± 1.42	46.87	−23.33	39.21	0.60
	Painter	1	5.80 ± 1.95	6.56 ± 1.72	56.52	−23.73	45.92	0.52
		2	5.80 ± 1.80	5.64 ± 1.79	49.85	−20.68	40.19	0.51
V10-3653	Blacksburg	1	5.67 ± 1.96	4.53 ± 2.09	51.61	−19.42	40.41	0.48
		2	5.72 ± 1.55	6.06 ± 1.58	48.04	−22.54	37.40	0.60
	Painter	1	5.24 ± 1.86	5.35 ± 1.72	55.57	−21.29	41.70	0.51
		2	5.84 ± 1.61	5.62 ± 1.58	49.40	−22.29	38.53	0.58
V16-0524	Blacksburg	1	6.22 ± 1.87	5.94 ± 1.82	58.70	−21.89	45.71	0.48
		2	6.40 ± 1.40	6.28 ± 1.51	49.09	−22.48	40.38	0.56
	Painter	1	5.96 ± 1.63	6.42 ± 1.55	52.65	−23.25	40.52	0.57
		2	5.60 ± 2.02	5.90 ± 1.61	49.60	−21.73	39.42	0.55
V16-0528	Blacksburg	1	5.48 ± 1.96	4.64 ± 1.85	54.10	−21.34	41.26	0.52
		2	5.28 ± 1.92	4.92 ± 1.68	52.97	−22.40	41.41	0.54
	Painter	1	5.42 ± 1.80	5.40 ± 1.76	56.48	−20.50	41.82	0.49
		2	6.12 ± 1.55	6.32 ± 1.66	49.69	−23.01	38.92	0.59
V16-0547	Blacksburg	1	5.56 ± 1.92	5.80 ± 1.94	53.80	−21.05	41.43	0.51
		2	6.54 ± 1.81	6.28 ± 1.54	50.71	−22.12	40.28	0.55
	Painter	1	6.44 ± 1.93	6.48 ± 1.61	53.97	−21.47	41.33	0.52
		2	6.44 ± 1.40	6.44 ± 1.61	50.68	−21.88	39.22	0.56

[a] *Arithmetic mean ± standard deviation; hedonic scale: 1 = "dislike extremely", 5 = "neither like nor dislike", 9 = "like extremely".*
[b] *CIELAB scale: L* = darkness (0) or lightness (100), a* = greenness (–) or redness (+), and b* = blueness (–) or yellowness (+); GI, green intensity (calculated).*

statistical analysis were performed in JMP® Pro 15.0.0: analysis of variance (ANOVA), Tukey's Honestly Significant Difference (HSD) tests, and Spearman's rank-order correlation (correlations multivariate analysis). WTP was estimated via random-parameters logit model in MATLAB. The standard errors for WTP were calculated via Daly et al. (2012). Significance level was set to 5% ($\alpha = 0.05$).

RESULTS

Edamame Color

Color characteristics of the 10 edamame genotypes grown in Virginia in both years of research are shown in **Table 1**. Color of edamame beans (L*, a*, b* values) was significantly affected by "year" ($p < 0.05$). Overall, GI increased from year 1

TABLE 2 | Consumer acceptability of edamame genotypes.

Edamame	Least squares means	
Genotype	Overall liking	Appearance liking
R14-16195	6.40[a]	5.96[a]
R15-10280	6.26[a]	6.04[a]
V16-0547	6.25[a]	6.25[a]
R14-6238	6.05[a,b]	6.12[a]
V16-0524	6.05[a,b]	6.14[a]
UA-Kirksey	6.04[a,b]	6.25[a]
R14-6450	6.01[a,b]	5.91[a]
R16-5336	6.01[a,b]	5.86[a,b]
V10-3653	5.62[b]	5.39[b,c]
V16-0528	5.58[b]	5.32[c]

[a–c]*LS-means followed by a superscript letter in common are not significantly different (p > 0.05).*

(range: 0.41–0.57) to year 2 (range: 0.51–0.60). The interaction "location*year" was significant for a* (greenness; $p < 0.05$). "Genotype" was a significant effect for b* (yellowness; $p < 0.05$), with variety R14-6450 having the lowest b* values when averaging the two locations: 38.1 (year 1) and 37.8 (year 2).

Consumer Acceptability of Edamame and Willingness-To-Pay (WTP)

Genotype, location, year, and all interaction factors significantly affected AL ($p < 0.05$). OL was affected by edamame genotype, year, and the interaction factors "genotype*location" and "genotype*location*year" ($p < 0.05$). Differences in consumer acceptability of edamame genotypes grown in Virginia are shown in **Table 2** (least squares mean; both locations, both years). Breeding lines V10-3653 and V16-0528 were the least liked edamame samples. Breeding lines R14-16195, R15-10280, and V16-0547 were the most liked genotypes overall, and UA-Kirksey (commercial check) and V16-0547 beans had the most liked appearance. Overall, hedonic mean scores (OL and AL) significantly increased from year 1 to year 2 ($p < 0.05$), and edamame genotypes grown in Painter had significantly higher AL mean scores than edamame grown in Blacksburg ($p < 0.05$). Furthermore, the choice-based conjoint analysis conducted in the first year of research showed that consumers are willing to pay \$0.77 more for dark green edamame beans in a salad, relative to beans with a light green color. The WTP value is significant at $\alpha = 0.05$, with a standard error of 0.32.

Relationship Between Edamame Color and Consumer Acceptability

Correlations between consumer acceptability scores (OL and AL) and edamame color characteristics (L*a*b* values and GI) were investigated using the complete dataset reported in **Table 1** (10 edamame genotypes, 2 VA locations, 2 years). Spearman's correlations (ρ) ranged between −1 (dark blue, cool) and +1 (dark red, warm), as shown in the heatmap

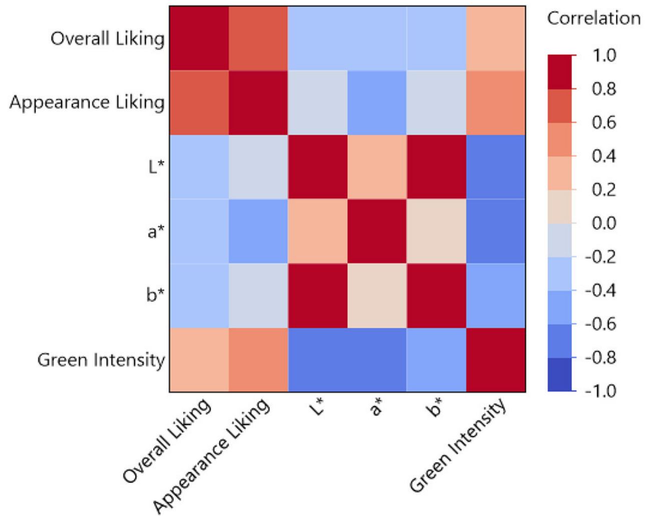

FIGURE 2 | Heatmap on Spearman's correlations between edamame color characteristics and consumer acceptability scores.

(**Figure 2**). A strong positive correlation was observed between OL and AL hedonic scores ($\rho = 0.68$), as well as between the L* and b* color coordinates ($\rho = 0.82$) (**Table 3**). A significant but less strong correlation was observed between the L* and a* color coordinates ($\rho = 0.32$). GI was negatively correlated with L* ($\rho = -0.76$), a* ($\rho = -0.74$), and b* ($\rho = -0.58$). A significant correlation was not observed between OL and color characteristics ($p > 0.05$). However, the GI color index and the red-green coordinate (a*) of the color space seemed to be significant predictors of AL scores ($p < 0.05$), which were correlated with OL. AL was negatively correlated with a* ($\rho = -0.51$) and was positively correlated with GI ($\rho = 0.40$).

DISCUSSION

Edamame appearance/color variability can be linked to genotype, planting date, harvest, processing conditions, such as blanching, and storage (Mozzoni et al., 2009; Amilia et al., 2021; Moseley et al., 2021). The blanching procedure followed in both years of our research was the same: immersion of edamame pods in boiling distilled water for 1 min (section sample preparation). Thus, color variation and differences in hedonic scores (OL and AL) were likely associated, for example, with changes in sample preparation for sensory tests (microwave cooking described in section sample preparation), environmental conditions of growing locations (e.g., temperature, precipitation), and/or harvest time and procedures. Nevertheless, the fact that OL and AL scores significantly increased from year 1 to 2 suggests that some changes were positive. Likewise, the overall increase in green intensity (GI) from year 1 to 2 suggests better quality of the samples (lower freshness decline). In fact, it is possible that in the first year, samples were overcooked during sample preparation for sensory tests; all edamame beans

TABLE 3 | Correlation between color characteristics and consumer acceptability (appearance and overall liking) of edamame.

Variable	By variable	Spearman ρ	Prob>$|\rho|$[a]
Appearance liking	Overall liking	0.6843	**<0.0001**
L*	Overall liking	−0.2802	0.0799
L*	Appearance liking	−0.1596	0.3254
a*	Overall liking	−0.2202	0.1722
a*	Appearance liking	−0.5147	**0.0007**
a*	L*	0.3166	**0.0466**
b*	Overall liking	−0.2266	0.1596
b*	Appearance liking	−0.0691	0.6719
b*	L*	0.8206	**<0.0001**
b*	a*	0.0025	0.9876
Green intensity	Overall liking	0.2848	0.0748
Green intensity	Appearance liking	0.4006	**0.0104**
Green intensity	L*	−0.7554	**<0.0001**
Green intensity	a*	−0.7357	**<0.0001**
Green intensity	b*	−0.5824	**<0.0001**

[a] *Numbers in bold indicate correlation is significant.*

had a "wrinkled" appearance that was much less noticed in the second year. Variability in size of the edamame beans (either within a same sample or among different ones) was not evaluated, but it may also have affected AL and/or OL scores. A previous study with peas reported that frozen pea samples were sorted into four sieves sizes with the purpose of reducing variation (Edelenbos et al., 2001). In this study, edamame samples were not selected by size due to limited amount of beans available for the sensory tests, but a similar approach could be considered to select edamame beans in future investigations. Nevertheless, as research on edamame continues in the USA, it is important to control as many sources of variability as possible to draw stronger conclusions and better guide edamame breeders, growers, processors, and consumers. Improving communication is also important to consolidate best production practices, including irrigation methods and standard harvest procedures (e.g., precooling strategies).

In our previous report (Carneiro et al., 2021), we described that the check-all-that-apply (CATA) answered by our sensory panelists investigated only flavor descriptors. In future sensory studies, the inclusion of appearance descriptors to the CATA questions could be valuable to better explain AL scores and the impact of specific appearance characteristics in overall acceptability of edamame. However, there is still a need for a standardized vocabulary (lexicon) that includes appearance terms to describe edamame pods and beans (Carneiro et al., 2020). Additionally, due to the limited amount of edamame pods, only shelled edamame beans were served in the sensory panels. Thus, samples tested by our panelists were closer to what consumers would see and consume in a salad, which was assessed in the WTP questions. Our statistical analysis showed that the a* axis of the color space was affected by the interaction factor "location*year," and significant differences among genotypes were observed in the b* color coordinate (yellowness). AL was significantly correlated with OL, a*, and GI, but no color characteristic was significantly correlated with overall acceptability scores. Descriptive sensory methods and focus group discussions are possible tools that can be applied in future research to guide further development of color standards. The need for further studies on color and appearance of edamame is also supported by the fact consumers in the USA are willing to pay more for dark green edamame beans in a salad ($0.77 more than light green edamame beans).

Although the appearance of pods is an important quality characteristic of edamame, it was not evaluated by our sensory panel. However, pod dimensions (length, width, and thickness) and pod pubescence (hairs per 2.4 cm^2) of all 10 genotypes evaluated in this study were analyzed by Lord et al. (2021). As reported by the authors, harvest occurred when plants reached the R6 stage and the immature edamame seeds (green color) reached 80–90% of their pod capacity. Genotype had a significant effect on pod length (range: 40–51 mm average) and pod pubescence, but did not have a significant effect on pod width (range: 10.9–12.2 mm average) and thickness (range: 6.5–8.1 mm average). Lord et al. (2021) also reported that genotype had a significant effect on other agronomic characteristics such as 10-pod sample weights and proportion of one-seeded pods and the advanced breeding lines R14-6450 and R15-10280 had the heaviest pods. In a recent study that analyzed edamame in the pods, Flores et al. (2019) measured color of pods of three edamame cultivars (Giant Midori, Kuroshinja, and Butterbean) and reported that only the a* color coordinate (greenness) was significantly different among genotypes. The authors also reported that higher AL scores were associated to greener color, as well as higher L* values (lightness).

CONCLUSIONS

Appearance is an important contributor to overall acceptability of edamame beans and it is often associated with quality of the product. GI (color index) and a* (red-green coordinate of the color space) showed significant correlation with AL scores and are suggested as indicators of edamame quality. Also, as consumers are willing to pay more for dark green edamame beans in a salad, color measurement can be used as a fast and affordable method to support breeding selection, monitor edamame quality (growers and processors), determine best storage conditions and preparation practices, and it also allows the selection of premium price products. Further research is needed to assist in the development of edamame color and appearance standards; focused discussions can be used in future studies to link subjective/visual measurement and green intensity (color index), for example. Further studies are also suggested to investigate the impact of other appearance attributes (e.g., size and glossiness of edamame beans) on consumer acceptability, and CATA questions containing appearance descriptors can be of great value to explain sensory evaluation data in future research. Additionally, minimizing possible sources of color and appearance variations pre- and post-harvest is of great importance as the edamame research continues in Virginia and other states in the USA.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Virginia Tech Institutional Review Board (IRB) for Research Involving Human Subjects (IRB 18-310). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RC wrote first draft. RC, CN, and SD contributed with conception and design of the study. RC, DY, and KA worked on edamame processing. RC and KA collected and analyzed consumer and color data. MB contributed with WTP data analysis. BZ led the development and selection of the edamame breeding lines. TK, SR, and MR led the field trials in the Virginia growing locations. All authors reviewed, edited, and approved the final version of the manuscript.

FUNDING

This work was supported by the USDA-NIFA (Grant No. 2018-51181-28384; Accession No. 436 1016465) and the Virginia Agricultural Experiment Station.

ACKNOWLEDGMENTS

The authors thank all the students and researchers of our edamame team for their hard work during the first two years of research. The authors also thank all volunteers that participated in the consumer studies, Drs. Hengjian Wang, Katherine Phetxumphou, and Jacob Lahne of the Virginia Tech Department of Food Science and Technology for their support throughout the project, and the Virginia Tech Statistical Applications and Innovations Group (SAIG) for their assistance with data analysis.

REFERENCES

Amilia, W., Wiyono, A. E., Ferzia, D., Rusdianto, A. S., Suryaningrat, I. B., Mahardika, N. S., et al. (2021). Physical, chemical, and sensory characteristics of frozen salted edamame during storage at room temperature. *Int. J. Food Agric. Nat. Resour.* 2, 9–18. doi: 10.46676/ij-fanres.v2i1.20

Carneiro, R., Duncan, S., O'Keefe, S., Yu, D., Huang, H., Yin, Y., et al. (2021). Utilizing consumer perception of edamame to guide new variety development. *Front. Sustain. Food Syst.* 4, 556580. doi: 10.3389/fsufs.2020.556580

Carneiro, R. C. V., Duncan, S. E., O'Keefe, S. F., Yin, Y., Neill, C. L., and Zhang, B. (2020). Sensory and consumer studies in plant breeding: a guidance for edamame development in the U.S. *Front. Sustain. Food Syst.* 4, 124. doi: 10.3389/fsufs.2020.00124

Daly, A., Hess, S., and Jong, G., De. (2012). Calculating errors for measures derived from choice modelling estimates. *Transportation Res. Part B* 46, 333–341. doi: 10.1016/j.trb.2011.10.008

Edelenbos, M., Thybo, A., Erichsen, L., Wienberg, L., Andersen, L., and Edelenbos, M. (2001). Relevant measurements of green pea texture. *J. Food Qual.* 24, 91–110. doi: 10.1111/j.1745-4557.2001.tb00594.x

Flores, D., Giovanni, M., Kirk, L., and Liles, G. (2019). Capturing and explaining sensory differences among organically grown vegetable-soybean varieties grown in northern California. *J. Food Sci.* 84, 613–622. doi: 10.1111/1750-3841.14443

Lawless, H. T., and Heymann, H. (2010). *Sensory Evaluation of Food Principles and Practices* (Second). New York, NY: Springer. doi: 10.1007/978-1-4419-6488-5

Lord, N., Kuhar, T., Rideout, S., Sutton, K., Alford, A., Li, X., et al. (2021). Combining agronomic and pest studies to identify vegetable soybean genotypes suitable for commercial edamame production in the Mid-Atlantic U.S. *Agric. Sci.* 12, 738–754. doi: 10.4236/as.2021.127048

Masuda, R. (1991). "Quality requirement and improvement of vegetable soybean," in *Vegetable Soybean: Research Needs for Production and Quality Improvement*, ed S. Shanmugasundaram (Taiwan: Asian vegetable research and Development Center), 92–102.

Moseley, D., Silva, M. P., Mozzoni, L., Orazaly, M., Florez-Palacios, L., Acuña, A., et al. (2021). Effect of planting date and cultivar maturity in edamame quality and harvest window. *Front. Plant Sci.* 11, 585856. doi: 10.3389/fpls.2020.585856

Mozzoni, L. A., Chen, P., Morawicki, R. O., Hettiarachchy, N. S., Brye, K. R., and Mauromoustakos, A. (2009). Quality attributes of vegetable soybean as a function of boiling time and condition. *Int. J. Food Sci. Technol.* 44, 2089–2099. doi: 10.1111/j.1365-2621.2009.02038.x

Neill, C. L., and Morgan, K. L. (2021). Beyond scale and scope: exploring economic drivers of U.S. specialty crop production with an application to edamame. *Front. Sustain. Food Syst.* 4, 582834. doi: 10.3389/fsufs.2020.582834

Pathare, P. B., Linus Opara, U., and Al-Said, F. A.-J. (2013). Colour measurement and analysis in fresh and processed foods: a review. *Food Bioprocess Technol.* 6, 36–60. doi: 10.1007/s11947-012-0867-9

Wang, K. C. (2018). East Asian food regimes: agrarian warriors, edamame beans and spatial topologies of food regimes in East Asia. *J. Peasant Stud.* 45, 739–756. doi: 10.1080/03066150.2017.1324427

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.