TRANSCRIPTOME ANALYSIS OF DOMESTICATED BREADFRUIT AND ITS WILD RELATIVES

A THESIS SUBMITTED TO THE FACULTY OF THE PROGRAM IN PLANT BIOLOGY AND CONSERVATION

BY KRISTEN LARICCHIA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN PLANT BIOLOGY AND CONSERVATION FROM NORTHWESTERN UNIVERSITY AND THE CHICAGO BOTANIC GARDEN

DECEMBER 5th, 2014

COMMENTS ON THESIS DEFENSE FOR MASTER OF SCIENCE IN PLANT BIOLOGY AND CONSERVATION NORTHWESTERN UNIVERSITY AND THE CHICAGO BOTANIC GARDEN

The student Kristen Laricchia defended his/her thesis, titled Transcriptome Analysis of Domesticated Breadfruit and Its Wild Relatives

on the date of 5 December 2014. The committee members have rendered the following decision:

☒ Pass unconditionally

☐ Pass conditionally with minor revisions and requires no re-examination (conditions detailed on the next page)

☐ Pass conditionally with major revisions and requires no re-examination (conditions detailed on the next page)

☐ Requires re-examination (conditions are detailed on the next page)

☐ Fail to pass (comments are attached on the next page)

Committee members:

Printed name	Signature	Department/Affiliation
NORMAN WICKETT Co-Major advisor		PBC
Nyree Zerega Co-Major advisor		PBC
Erik Andersen		Molecular Biosciences / IBiS

In the case of revisions or re-examination, specify details on the next page. Once all requirements have been met, the major advisor and program director should sign below.

Signature of major advisor confirming final approval — Date: 5 December 2014

This form, along with a pdf file of the final approved thesis, must be submitted to the PBC program assistant and the program director, at which point the latter will sign below and approval will be communicated to TGS.

Plant Biology and Conservation Program Director Signature — Date: 5 December 2014

ABSTRACT

Underutilized crops have the potential to economically benefit developing countries and to improve global food security. Breadfruit (*Artocarpus altilis*, Moraceae) is one such crop that can provide essential nutrients and requires relatively low-energy input to maintain compared to major crops. Humans have selected for many cultivars of breadfruit since its domestication began approximately 3,500 years ago. They selected for different traits and dispersed the new cultivars throughout the islands of Oceania as they migrated in an eastward direction from Melanesia to Polynesia and north into Micronesia. The goal of this research was to provide a better understanding of the genomic effects of the domestication process and, ultimately, to develop genomic resources that may facilitate the development of improved breadfruit cultivars in the future. In this study, a reference transcriptome of breadfruit was de novo assembled and characterized, and 24 transcriptomes of breadfruit and its wild relatives were analyzed to reveal genetic patterns correlated with the domestication gradient. This analysis suggested that signals of positive selection did not increase along a domestication gradient, but the presence of positive selection in certain genes varied according to ploidy level and taxon. In addition, over 1,000 genes were indicated as being driven by positive selection, many of which localized to plastids. Nucleotide sites and individuals under positive selection were also discovered in MADS-box genes and carotenoid biosynthesis genes. Overall, this research revealed several effects of the domestication process, provided candidate genes for further study, and produced new genomic resources for breadfruit.

ACKNOWLEDGMENTS

I would like to thank my advisers, Nyree Zerega and Norman Wickett, and committee member Erik Andersen, for their advice and encouragement throughout the development of this project. I am also grateful for the support from members of the Zerega and Wickett lab, and in particular for the bioinformatics assistance from Matt Johnson and Elliot Gardner. I would also like to thank Stuart Wagenius for his advice on statistical analyses. This research would not have been possible without the aid of the Breadfruit Institute, its staff members, and its collection manager, Ian Cole. The guidance from Stephanie Moorman and Sarah Owens at Argonne National Laboratory was also essential to this research. Finally, I would like to thank my family, friends, and peers for their constant support.

My graduate school education at Northwestern University and the Chicago Botanic Garden was partially funded by the Shaw Fellowship for Plant Biology and Conservation for which I am greatly appreciative. This project received funding from the Institute for Sustainability and Energy at Northwestern (ISEN) and the Plant Biology and Conservation Travel Award.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Breadfruit: An Underutilized Tree Crop of Oceania

Plant domestication occurs when humans alter a species' genetics, likely in preference for certain phenotypic characteristics, to the extent that a new species is developed which may be more reliant on humans for reproduction or better fits their needs (Doebley et al. 2006; Purugganan and Fuller 2009). Breadfruit (*Artocarpus altilis* (Parkinson) Fosberg, Moraceae) is a tropical tree crop that was domesticated in Oceania but is cultivated throughout the wet tropics today (Figure 1; Jones et al. 2011). Although it can serve as a valuable food source in the regions in which it grows, it is considered an underutilized crop and is listed by the Treaty on Plant Genetics Resources for Food and Agriculture as a crop that can positively impact food security (Jones et al. 2011). A large number of cultivars display selection for different traits in breadfruit (Jones et al. 2011, 2013b). The desire for certain traits, such as high micronutrient production and a longer fruiting period, has promoted many studies and efforts to conserve and better understand the diversity found in breadfruit (Jones et al. 2010, 2013a, 2013b; Zerega et al. in press). Many cultivars are seedless, and when seeds are present, they are recalcitrant and do not withstand drying, freezing, or prolonged periods of storage (Ragone 1997). Therefore, trees are often cultivated through vegetative propagation, and living collections, the largest of which is managed by the Breadfruit Institute of the National Tropical Botanical Garden (http://www.ntbg.org/breadfruit/), are essential for conserving and researching breadfruit. Breadfruit was also often vegetatively propagated during its domestication in Oceania. A better understanding of how this domestication has impacted the genetics of breadfruit can aid in its conservation and provide resources that can help promote its use. Human-necessitated dispersal of breadfruit throughout islands in Oceania provides a good opportunity to use transcriptomes to

study how the domestication process has impacted the various cultivars of breadfruit and its wild relatives, which is the focus of this research.

Breadfruit and its wild relatives, *Artocarpus camansi* Blanco and *Artocarpus mariannensis* Trecul, are monoecious trees that grow to approximately 20 meters and live for about 80 years (Ragone 1997, 2008). They are likely wind pollinated, but triploid cultivars produce fruit through parthenocarpy (Hasan and Razak 1992). Breadfruit has a false fruit structure that is technically a compound fruit or syncarp (Ragone 1997; Zerega et al. 2006). In seeded breadfruit, the "seeds" are actually an achene fruit, and the fleshy tissue around it is swollen perianth tissue. This fleshy perianth tissue is edible and valued as a source of starch. Although the false fruit structure consists predominantly of the perianth, it will be henceforth referred to here as a fruit. People usually harvest breadfruit when it is mature but not fully ripe (Ragone 1997). The fruit can then be boiled or roasted, or it can be dried and ground to be used as flour (Ragone 1997; Zerega et al. 2006). Breadfruit is also eaten raw when it has ripened. The "seeds" (technically the achene fruit), if present, are typically eaten either boiled or roasted. Trees can produce approximately 200 fruit a year, each weighing between one and four kilograms (Purseglove 1968; Marte 1986).

Although breadfruit is a valuable, high-yielding food source, it is considered an underutilized crop because it can offer many advantages if its use is increased. Several factors contribute to its lack of utilization. The fruits oxidize quickly and have a short shelf-life, the timing of fruit bearing varies among cultivars, and the seeds are recalcitrant, making them unable to survive drying or storage processes (Ragone 1997; Jones et al. 2010). Urbanization also threatens the use of breadfruit. Many regions are more heavily relying on imported crops, reducing variety in people's diets while contributing to a cultural loss of knowledge on preparing

and storing breadfruit (Zerega et al. 2006; Jones et al. 2013b). Breadfruit has also been stigmatized during its introduction to new areas. The first attempt to introduce breadfruit to the Caribbean in 1789 failed after numerous collections were thrown overboard during the infamous mutiny on the HMS Bounty. Although the Caribbean is now the second biggest producer of breadfruit after Oceania, the fruit was first introduced here as an alternative food for slaves; however, it was ill received and often used as livestock feed instead (Roberts-Nkrumah 2007). Breadfruit is a less popular choice of food in some countries, such as Trinidad and Tobago, because it can still be associated with poverty and slavery, and in other areas its consumption suffers from poor demand and marketing (Roberts-Nkrumah 2007; Roberts-Nkrumah and Legall 2013; Englberger et al. 2014).

Nonetheless, breadfruit can serve as a sustainable crop and alleviate problems associated with food insecurity. It grows best in tropical regions, and its range of suitable habitat overlaps many of the most food insecure regions (Figure 2; Lucas and Ragone 2012). Breadfruit requires little water, fertilizer, and technology to grow, and provides greater yields per hectare than rice, corn, and wheat (Jones et al. 2011). It is traditionally grown in agroforestry systems, often with other tree crops or root crops, and growing it locally is more economical than importing other foods (Ragone 1997; Roberts-Nkrumah and Legall 2013; Lincoln and Ladefoged 2014). To combat its short-shelf life, breadfruit can be processed into chips or ground into flour, and it has traditionally been fermented in undergrounds pits to make it into a sour paste that can last over a year (Pollock 1984; Ragone 1997). Other uses for breadfruit include lumber, livestock feed, insect repellent, shade, and erosion control (Zerega et al. 2006; Roberts-Nkrumah and Legall 2013). It has been a staple crop in the Pacific Islands for over 3,000 years, and its sustainability and nutritional value were recognized by many voyagers eager to introduce the fruit to new areas

(Ragone 1997). The fruit is mainly a source of starch, with the carbohydrate content of breadfruit flour ranging from 48-61 percent (Roberts-Nkrumah and Legall 2013). Breadfruit is gluten free, has an intermediate glycemic index, and can provide more protein than banana and cassava, although not as much as taro (Dan Ramdath et al. 2004; Dignan et al. 2004; Jones et al. 2011). People can also obtain many micronutrients from breadfruit, and flour made from breadfruit contains more calcium, manganese, potassium, and iron than wheat, rice, and corn flour (Table 1; Jones et al. 2011). The ability of breadfruit to provide such nutrients is particularly important because it grows in areas with high levels of micronutrient deficiency, such as the Federated States of Micronesia (Englberger et al. 2003b).

People living in the Federated States of Micronesia as well as other countries in the Pacific also suffer from high rates of vitamin A deficiency (VAD) (Englberger et al. 2003a, 2003c). Vitamin A, the active form of which is referred to as retinol in the human body, is an essential vitamin for proper functioning of the eye (Olson 1989; McLaren and Frigg 2001). A lack of vitamin A can lead to blindness in severe cases of VAD, and approximately 190 million preschool-aged children are affected by VAD, predominantly in Africa and south-east Asia (McLaren and Frigg 2001; WHO 2011). Certain carotenoids serve as provitamin A carotenoids, which are then converted to vitamin A. Carotenoids that can be converted to vitamin A are said to have provitamin A activity and include α-carotene, β-carotene, and β-cryptoxanthin (Olson 1989). Breadfruit provides provitamin A carotenoids and grows in areas afflicted by VAD, although the amount of carotenoids produced varies according to cultivar (Englberger et al. 2003a; Jones et al. 2013a). Lutein is the most abundant carotenoid found in the fruits, but it does not have provitamin A activity. β-carotene, which does have provitamin A activity, is the next most prominent carotenoid produced by breadfruit (Jones et al. 2013a). Carotenoid levels appear

to have decreased during the process of domestication, possibly in preference for white-colored fruit (Jones et al. 2013a, 2013b). However, *A. mariannensis* as well as certain cultivars, such as Samoan 1, have been noted to have superior carotenoid levels (Ragone 1997; Jones et al. 2013a, 2013b). Breadfruit also produces more carotenoids when ripe, and breadfruit that is sweeter and more yellow in color has been observed during seasons with drier weather (Englberger et al. 2003a, 2003b, 2014). However, coloration is not always consistent with carotenoid content and cannot be used to predict if provitamin A carotenoids are present (Englberger et al. 2003a).

The ability of breadfruit to serve as a food source has led to its domestication and dispersal throughout Oceania. Breadfruit cultivars exhibit a wide range of morphological and genetic diversity (Zerega et al. 2004, 2006; Jones et al. 2013b) and originated through two separate mechanisms involving two progenitor species. In Melanesia and Polynesia breadfruit is thought to have originated through propagation and selection of *A. camansi*, whereas in Micronesia, much of the breadfruit is of hybrid origin from *A. camansi*-derived *A. altilis* and *A. mariannensis*, which is native to the Mariana Islands and Palau (Zerega et al. 2004, 2005). *Artocarpus camansi* is native to New Guinea and the Moluccas (Jarrett 1959; Ragone 1997). The islands of Melanesia, including New Guinea, Fiji, Vanuatu, and the Solomon Islands, have been inhabited by humans for tens of thousands of years during which the seeds of *A. camansi* have likely been used as a food source (Zerega et al. 2004). Approximately 5,000 years ago Austronesian people, referred to as Lapita, came to the northern part of New Guinea and rapidly migrated eastward into the previously uninhabited islands of Polynesia (Kirch 2000). They brought food plants with them, such as *A. camansi*, likely transporting them as vegetative propagules for long journeys eastward on which recalcitrant seeds would not survive (Zerega et al. 2004). Over generations of vegetative propagation and selection, humans dispersed various

selected cultivars of breadfruit throughout Oceania (Figure 3). Several routes of migration into Micronesia are proposed with possible secondary migration routes and introductions (Zerega et al. 2004). Breadfruit was also introduced to the Caribbean in 1793, and Europeans have brought breadfruit to additional areas outside Oceania, such as West Africa, Southeast Asia, and Central and South America (Bligh 1792; Ragone 1997; Roberts-Nkrumah 2007). Regions vary in the ploidy level and varieties of breadfruit that are harvested. Breadfruit cultivars today are either diploid (typically with seeds) or triploid (seedless), with the propensity of seedless cultivars increasing from New Guinea to East Polynesia (Ragone 2001; Jones et al. 2013b). The diploid number in breadfruit is 56 chromosomes, whereas the triploids have 84 chromosomes (Ragone 2001). Triploid breadfruit is thought to have arisen through either spontaneous mutation, or through non-disjunction during meiosis. Aberrations in meiosis may have been caused by the accumulation of deleterious somatic mutations as a result of generations of clonal propagation (Ragone 2001; Zerega et al. 2004).

Both *A. camansi* and *A. mariannensis* are diploid species. *Artocarpus camansi* trees produce large fruits containing many seeds (Jones et al. 2013b). The fruit of *A. mariannensis* is typically smaller and also seeded. *Artocarpus altilis* x *A. mariannensis* hybrids and *A. altilis* exist in both diploid and triploid forms and can either have or lack seeds (Jones et al. 2013b). Early generation *A. altilis* x A. *mariannensis* hybrids resemble *A. mariannensis* morphologically and genetically, whereas later generation hybrids are larger and usually seedless, therefore aligning them more closely with *A. altilis* (Fosberg 1960; Zerega et al. 2004, 2005). Previous studies suggest that hybrids can better tolerate saline soils and contain higher amounts of iron and carotenoids (Jones et al. 2011). Additionally, Jones et al. (2013b) have conducted a comprehensive analysis of morphological variation in breadfruit cultivars throughout Oceania.

Along the west to east migration route, people appear to have predominantly selected for smooth-skinned varieties of breadfruit, which may have been easier to peel (Jones et al. 2013b). Although not selecting for larger fruit, people have bred the species to produce less seeds and a smaller fruit core, therefore increasing the relative amount of edible fleshy pulp and likely making it easier to ferment and subsequently knead the breadfruit into a paste (Pollock 1984; Jones et al. 2011). Breadfruit from Melanesia, where domestication began, tends to have larger fruit with more yellow-colored pulp and produces more latex than those of the other geographic regions (Figure 4; Jones et al. 2013b). In Western Polynesia, the fruits are distinguished by their smaller size and less yellow pulp. Human-mediated selection in Eastern Polynesia has favored fruits that are mostly seedless and are of an intermediate size between those of Melanesia and Western Polynesia. Breadfruit cultivars found in Micronesia, many of which are hybrids, produce the least amount of latex, have few to no seeds, are of more variable size, and have pulp that is more yellow than that of Eastern Polynesia, but less yellow than that of Melanesia. *Artocarpus mariannensis*, which is native to Micronesia, has pulp with particularly strong yellow coloration (Englberger et al. 2014).

Domestication has also impacted the genetic diversity found in breadfruit. Gene flow is greater in islands closer to New Guinea, where wild relatives, early-generation cultivars, and diploid breadfruit are often found (Xing et al. 2012). Genetic diversity is lower among the more eastern islands of Oceania compared to the western ones (Zerega et al. 2006, in press). The greatest genetic diversity is found in Micronesia, where many hybrids are found, as well as Melanesia, and the lowest diversity is found in Polynesia (Zerega et al. 2004; Jones et al. 2011). An evaluation of the largest breadfruit germplasm collection using microsatellites showed that 49% of the triploid cultivars are represented by a single genotype, but many other genotypes are

represented by only one sample (Zerega et al. in press). This same study suggests that triploid hybrids and diploid *A. altilis* display the greatest amount of genetic diversity. Wild relatives were found to harbor less genetic diversity than cultivated breadfruit; however, this result may be an artifact of a much smaller and less geographically diverse sampling of the wild relatives, or could be explained by the fact that breadfruit is a perennial species and therefore may experience less severe domestication bottlenecks (Miller and Gross 2011, Zerega et al. in press).

Domestication and Target Genes

Studies on the domestication of other species, such as those that have helped researchers understand genetic bottlenecks, can also aid in designing studies aimed at understanding the domestication process in breadfruit. Research suggests that domestication can occur both through conscious selection, in which people select for certain phenotypes in plants, or through unconscious selection, in which traits are changed unintentionally through cultivation (Meyer and Purugganan 2013). Many studies have focused on discovering the genes underlying these traits that were impacted during the domestication process (Doebley et al. 2006; Meyer and Purugganan 2013). These studies often use a top-down approach where genes linked to observed phenotypes are examined using methods such as quantitative trait loci mapping (Ross-Ibarra et al. 2007). With technological advances, a bottom-up approach is gaining popularity in which signals of positive selection are searched for within many genes, and discovered candidate genes are further examined to determine if they are linked to particular phenotypes (Doebley et al. 2006). Signals of positive selection are indicated by a higher rate ratio of nonsynonymous mutations (which change the amino acid sequence) to synonymous mutations (which do not alter the amino acid sequence) than expected by genetic drift and neutral selection (Yang and

Bielawski 2000). Studies using both the top-down and bottom-up approach in model and non-model crops have revealed that human-mediated selection often leads to desirable traits such as reduced seed dispersal, loss of seed dormancy, altered resource allocation, loss of photoperiod sensitivity, larger fruit size, and inhibited defense mechanisms (Olsen and Wendel 2013). Many studies have focused on the phenotypic differences in breadfruit that impact desirable traits, such as the increase in the proportion of edible starchy material and difference in skin texture, but studies analyzing the genetic differences that potentially affect these traits are lacking.

Certain genetic changes underlying phenotypes have been found to be common during the domestication of many species and could also have influenced breadfruit during its domestication. Domestication often impacts genes controlling transcription because transcription factors have regulatory functions that can modify complex traits, and they play a dominant role in regulating morphological development (Doebley et al. 2006; Century et al. 2008; Olsen and Wendel 2013). During the short time-span during which much domestication has taken place, mutations leading to the elimination of a gene or the creation of a new functional gene are less likely to occur than mutations that modify the function of a gene through regulatory means (Vaughan et al. 2007). One group of genes consisting predominantly of transcription factors is the MADS-box gene family. These genes are involved in controlling inflorescence morphology, fruit development, flowering timing, phase transitions, and seed development, as well as many other processes (Smaczniak et al. 2012). Positive selection has been detected in numerous MADS-box genes in a variety of crops, including tomato, maize, cauliflower, and pearl millet (Purugganan et al. 2000; Vrebalov et al. 2002; Liu et al. 2004; Zhao et al. 2011; Clotault et al. 2012; Xie et al. 2014). Following domestication, which results in a species distinguishable from it wild relative, human-mediated selection often results in new varieties of the species through

the diversification of genes (Meyer and Purugganan 2013). This selection for particular varieties likely involved a greater amount of conscious selection, and studies suggest that diversification is more likely to impact enzyme-encoding genes rather than transcription factors (Purugganan and Fuller 2009).

One pathway of enzyme-encoding genes that has been investigated because of its importance in the human diet is the carotenoid pathway, which is also of interest in breadfruit because of its role in producing provitamin A carotenoids. Carotenoids are 40-carbon isoprenoid molecules that include hydrocarbons, such as β-carotene, and xanthophylls, like lutein and zeaxanthin (Figure 5; Krinsky 1989; Ruiz-Sola and Rodríguez-Concepción 2012). Carotenoids help prevent oxidation reactions that can damage cellular components of the plants by absorbing excess energy rather than passing this energy on to oxygen (Krinsky 1989). Carotenoids also impart yellow, orange, and red coloration to plant parts, attracting potential pollinators (Grotewold 2006). They additionally serve as precursors for apocarotenoids, such as abscisic acid and strigolactone, which function in stress response pathways and interactions with mycorrhizal fungi (Nambara and Marion-Poll 2005; Xie et al. 2010; Stange and Flores 2012). The first substrate involved in the carotenoid biosynthesis pathway, geranylgeranyl diphosphate, is a product of the methylerythritol phosphate pathway. It is the substrate for the first enzyme acting in the carotenoid pathway, phytoene synthase (Lu and Li 2008). Many carotenoid enzymes, including phytoene synthase, localize to the thylakoid membranes of chloroplasts and to plastogloubles and crystalline structures of chromoplasts (Tevini and Steinmüller 1985; Cunningham and Gnatt 1998).

Gene duplication is thought to have played an essential role in the evolution of carotenoid genes, with some carotenoid genes existing in single copies whereas others have multiple

paralogs (Galpaz et al. 2006; Ruiz-Sola and Rodríguez-Concepción 2012). Differential expression and control of paralogous carotenoid genes allows for the functional separation of carotenoids localized to different tissues (Galpaz et al. 2006; Howitt and Pogson 2006). Many transcription factors also impact the production of carotenoids and their effects have been studied in multiple organisms; these transcription factors include PIF, DDB1, DET1, LAF1, HY5, HFR1, and COP1 (Toledo-Ortiz 2010; Stange and Flores 2012). Both carotenoid biosynthesis genes and genes regulating transcription of carotenoid-related genes may have been impacted during the domestication of breadfruit.

Importance of Understanding the Domestication Process in Breadfruit

Breadfruit can contain high levels of carotenoids and other essential vitamins or minerals, and it is a low-energy input tree, making it of particular interest to better establish in countries where malnutrition and food insecurity is prevalent. The recalcitrant seeds hinder efforts to increase its use while also making it more difficult to conserve cultivar diversity in germplasm collections (Ragone 2007). The need for living collections along with the increase of urbanization and use of non-traditional foods threatens the existence of unique cultivars in certain geographic areas (Jones et al. 2013b). These conditions also create the need for knowledge of the genetic basis of the cultivars, which can be gathered by better understanding the domestication process. Such an understanding may also aid studies in the development of superior cultivars, such as those that have a higher tolerance for salinity, are of lower stature, produce fruit more consistently throughout the year, and have a higher production of carotenoids. Research on the various breadfruit cultivars can reveal how domestication has affected the genetics of *A. altilis* and will provide useful information for those making conservation, breeding, and selection decisions.

Although it is thought that humans selected for fewer seeded, smoother-skinned fruit during the domestication of *A. altilis*, it is unclear which other less obvious or unintended traits may have been influenced by the process, such as those that developed through unconscious selection. A transcriptome analysis of the different cultivars of *A. altilis* and its wild relatives across a domestication gradient can help identify patterns in genes belonging to pathways that affect desirable traits.

Research Objective

The objective of this study was to produce, characterize, and analyze transcriptomes of *A. altilis* and its wild relatives, and to identify putative genes that have been affected by the domestication process. Taking a bottom-up approach, site and branch-site models were used to detect putative genes driven by positive selection. Positive selection was additionally explored in targeted genes of interest, which included MADS-box genes and genes related to carotenoid biosynthesis, using the branch-site test and the McDonald-Kreitman test. This study also tested the hypothesis that signals of positive selection in genes would increase along a longitudinal gradient representative of the domestication gradient and that positive selection would be associated with ploidy level, species, or region of origin for particular genes.

METHODS

Sample Collection

All samples were obtained from the National Tropical Botanical Garden germplasm collection, with the majority of the samples located at the Kahanu Garden in Maui, Hawaii. This collection represents *Artocarpus altilis*, *A. camansi*, *A. mariannensis*, and *A. altilis* × *A. mariannensis*

hybrids. Provenances of the collections include the Philippines, the Seychelles, Indonesia, Honduras, and 34 Pacific Islands, and data associated with these collections include morphology, seasonality, genetic diversity, and nutrition information (Breadfruit Institute: http://www.ntbg.org/breadfruit/, Zerega et al. 2004, 2005, 2006; Jones et al. 2010, 2011, 2013b; Witherup et al. 2013). At the site of the Kahanu Garden collection, the mean maximum temperature is 27.1°C, the mean minimum temperature is 19.7°C, and the mean annual precipitation is 2051mm. The soil is Hana Very Stony Silt Clay Loam, which is derived from volcanic ash, is well draining and slightly to moderately acidic, and has a surface horizon containing approximately 8% organic matter (http://wrcc.dri.edu/; http://websoilsurvey.nrcs.usda.gov; Jones et al. 2011).

Sample material from the Kahanu Garden was collected from 65 individuals during two days in October 2013, during which the majority of trees belonging to different seasonality groups shared an overlap in their timing of fruit production (Jones et al. 2010). To ensure adequate sampling along the domestication gradient, collected individuals include *A. altilis*, *A. camansi*, *A. mariannensis*, and *A. altilis* × *A. mariannensis* hybrids, and where appropriate represent both diploid and triploid trees from Melanesia, Micronesia, Western Polynesia, and Eastern Polynesia. Information on sample characterization was obtained from a recent study which genetically characterized the breadfruit germplasm collection and corrected misidentified accessions (Zerega et al. in press). The ploidy level of breadfruit accessions had previously been determined and confirmed using both chromosome counts and microsatellite data (Ragone 2001; Zerega et al. 2004). Material for RNA isolation was taken from young leaves still enclosed in or just emerging from the stipule and from the fleshy perianth tissue (the edible flesh of the multiple fruit structure – a syncarp) at the point in development when fruits are typically harvested, which

is when they are mature but not fully ripe. Tissue samples were stored in RNA Shield (Zymo Research, Irvine, CA, USA) and kept cool to stabilize and preserve the RNA for transport. An additional four samples were collected during November 2012 and were used for preliminary analyses. These samples, including a diploid *A. altilis* fruit, triploid *A. altilis* fruit, diploid *A. camansi* fruit, and diploid *A. camansi* vegetative sample (root, stem, leaf), were obtained from the National Tropical Botanical Garden's McBryde Garden and a private backyard, both located in Kauai, Hawaii (Table 2). Whole syncarp samples were harvested as described above and then shipped to the Chicago Botanic Garden over a 2-day period on dry ice. Upon arrival of the samples, the fleshy perianth tissue was cut into pieces weighing no more than 50mg, immediately immersed in liquid nitrogen, and stored in a -80°C freezer.

Extraction and Sequencing

RNA was extracted from all samples using a RNeasy® Mini Kit (Qiagen, Valencia, CA, USA) and quantified using a Qubit® RNA Assay Kit and Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). Samples with at least 10 ng/µl of RNA were sent to Argonne National Laboratory where the quality was determined using an Agilent 2100 bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA, USA). Twenty-two samples were chosen that both represented various stages along the domestication gradient and produced at least 1µg of high quality RNA upon extraction (Table 2). Ribosomal RNA was depleted from samples with the plant-specific Ribo-Zero rRNA Removal Kit (Epicentre, Madison, WI, USA). The ScriptSeq v2 RNA-Seq Library Preparation Kit (Epicentre, Madison, WI, USA) was used for polyA-selection, fragmentation of the RNA through heat-denaturation, cDNA synthesis, library construction, and size selection using AMPureXP beads (Beckman Coulter, Brea, CA, USA). Insert sizes ranged from 100 to

600 base pairs (bp). For the samples collected from the Kahanu Garden, paired-end sequencing was conducted using the Illumina HiSeq 2000 system with eight samples sequenced on one lane, eight samples sequenced on a second lane, and six samples sequenced on 2/3 of a third lane. For the four samples from Kauai, paired-end sequencing with an insert size of 180 bp was conducted using an Illumina HiSeq 2000 system across one lane.

Transcriptome Assembly

Raw reads were trimmed using TRIMMOMATIC v0.30 (Bolger et al. 2014) to remove low quality bases, adapter sequences, and other sequencing artifacts. Universal adapter and primer sequences, as well as their reverse complements, were trimmed using both simple and palindrome modes with the seed mismatch, palindrome clip threshold, simple clip threshold, and minimum adapter length set to 2, 30, 10, and 1 respectively. Parameters were set to trim the first 15 bases of each read which displayed evidence of random hexamer priming, leading and trailing bases with quality scores below 20, and bases following a four base sliding window with an average quality score below 20. TRIMEST vEMBOSS:6.6.0.0 (http://emboss.sourceforge.net/apps/release/6.6/emboss/apps/trimest.html) was then used to remove poly-A and poly-T tails with minlength and mismatch parameters set to 4 and 1 respectively. Only sequences longer than 50 base pairs were retained, and if only one end of a paired-end read was discarded, the other end was kept as an unpaired read. FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used prior and subsequent to read trimming to assess the quality of the reads and the effects of trimming. The trimmed paired and unpaired reads were assembled using TRINITY vr2013-02-25 (Grabherr et al. 2011), a pipeline designed for de novo RNA-seq assemblies. TRINITY consists of three processes:

Inchworm, Chrysalis, and Butterfly. Inchworm assembles sequences into unique contigs by k-mer extension. Chrysalis then clusters these contigs according to sequence similarity and generates de Bruijn graphs for them. Each cluster receives a component number, and isoforms should have the same component number, but not all sequences sharing a component number are isoforms. Butterfly determines the path taken along each graph for a read and generates the most likely full-length transcripts for alternatively spliced isoforms. It sometimes determines that a de Bruijn graph should be split into multiple graphs, resulting in multiple subcomponent numbers. Splice variants can commonly be assembled into different contigs but retain the same component/subcomponent name in the assemblies and be misinterpreted as paralogs. Therefore, only the longest isoform per component/subcomponent was retained to reduce redundancy in the assembly and better ensure that each gene was only represented by one transcript. Individual assemblies were produced for each sample. One reference assembly was also formed by pooling reads from the nine *A. altilis* individuals because a reference genome has not yet been assembled for any *Artocarpus* species.

The assemblies were assessed using the following quality statistics: number of contigs, total size of contigs, longest contig, contig mean, contig median, and contig N50. The average ortholog hit ratio and average collapse factor were calculated by blastx and tblastn comparisons (percent identity ≥ 80%) to the proteome of *Morus notabilis* C.K.Schneid (Chuansang), the most closely related species to breadfruit with a well annotated available genome sequence. The ortholog hit ratio gives an estimate of the completeness of a transcript and is calculated as the number of bases in an assembly contig that match a *M. notabilis* protein divided by the total length of that protein (O'Neil and Emrich 2013). The collapse factor measures erroneous collapsing of transcripts into a single contig and represents the number of proteins in the *M.*

notabilis proteome that had a best hit to a particular contig in the assembly (O'Neil and Emrich 2013).

Annotation

A suite of programs from the TRINOTATE vr2013-11-10 (http://trinotate.sourceforge.net/) pipeline was used to annotate the assemblies, therefore assigning functional information to each assembled transcript. Putative coding regions were determined with TRANSDECODER (http://transdecoder.sf.net) using the PfamAB database to assist in identifying protein domains. Protein domains were also separately searched against the PfamA database using HMMER because the sqlite database subsequently used was only linked to PfamA identifiers. Similarities to known sequences were determined using blastx and blastp searches against the UniProtKB/Swiss-Prot database with an e-value of .00001. All resulting annotations were uploaded to the sqlite database and when possible populated with GO (gene ontology) terms, which were extracted from blast matches. GO terms provide functional classification for genes, with the terms originating from a network of acyclic graphs. The terms are hierarchical, with higher level terms describing broad categories, such as the level 1 term "biological process", and lower level terms being more specific, such as the level 2 term "immune system process" (Primmer et al. 2013). The WEGO software (Ye et al. 2006) was then employed to visualize and quantify higher level GO terms.

Alignment and Variant Detection

Paired and unpaired reads were aligned to the reference assembly with BURROWS-WHEELER ALIGNER v0.6.2 (Li and Durbin 2010). Single nucleotide polymorphism (SNPs) and insertions or

deletions (indels) were identified separately for diploids and triploids following the RNA-seq best practices guideline provided by GATK v3.2-2 (McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013). These steps involved marking duplicates to be ignored for certain downstream tools, trimming the N's of cigar strings in reads to remove regions absent in the sequence alignment but present in the reference assembly, performing local realignment around indels to minimize mapping artifacts, and recalibrating base quality scores to improve accuracy of variant calling. Indel realignment was initially performed on individual samples because the tool bypasses regions with high read depth, which are those with greater than 20,000 reads. This realignment step was then performed cumulatively across all samples. To bootstrap the database of known SNPs and indels required for base recalibration, variants were first called using GATK's Unified Genotyper with the minimum emission and calling confidence thresholds set to 20. SNPs were hard filtered according to the following criteria: QualByDepth < 2.0, FisherStrand > 60.0, RMSMapping Quality < 40.0, HaplotypeScore > 13.0, MappingQualityRankSum < -12.5, ReadPosRankSum < -8.0. Clusters of 3 SNPs within 35 bases of each other were also filtered per recommendation by GATK. Indels were hard filtered according to the following criteria: QualByDepth < 2.0, FisherStrand >200.0, ReadPosRankSum < -20.0. SNPs and indels surviving hard filtering were used as the known sites for the base recalibration step. Base recalibration, variant calling, and hard filtering were repeated until the recalibration plots reached convergence and the resulting SNPs and indels of the subsequent round of hard filtering were chosen as the final set of variants. The average sequencing depth across the reference assembly for individual samples, including positions with a depth of zero, was calculated using GATK's DepthOfCoverage tool. The BaseCoverageDistribution tool was used to determine the

cumulative sequencing depth at each base across the reference assembly, excluding bases with a base quality below 17 and a mapping quality below 20.

Detection of Orthologs and Identification of Positively Selected Genes

Groups of orthologous genes, or orthogroups, were determined with PROTEINORTHO v5.10 (Lechner et al. 2011) using the translated sequences of assembled transcripts determined by TRANSDECODER as inputs. PROTEINORTHO was run on a separate set of assemblies, including the reference assembly, and the corresponding top blast hits, Pfam descriptions, and GO terms of the reference were used to annotate the orthogroups. MAFFT v7.130b (Katoh et al. 2002) was then used to align the amino acid sequences for each orthogroup and PAL2NAL v14 (Suyama et al. 2006) was used to produce codon alignments guided by the amino acid translations with gaps and stop codons omitted. A phylogeny for the codon alignment of each orthogroup was generated using RAXML v7.2.8 (Stamatakis 2006; Ott et al. 2007) with the GTRGAMMA substitution model.

Each ortholog group was searched for signals of positive selection using PAML v4.8 (Yang 2007). The value of omega (ω), the rate ratio of nonsynonymous to synonymous mutations, is often used to detect signals of selection from DNA sequence alignments. Conventionally, neutral selection is indicated by $\omega=1$, purifying selection by $\omega<1$, and positive selection by $\omega>1$ (Yang and Bielawski 2000). To identify genes containing sites influenced by positive selection, ω was calculated for each orthogroup under two models implemented in PAML, M1 and M2. These site models allow omega to vary among sites but not among branches. M1, a nearly neutral model, assumes a proportion (*p0*) of sites in one class where $\omega< 1$, and another proportion (*p1*) of sties in a second class where $\omega=1$. M2, a positive selection model,

allows for a third site class where ω >1 for a proportion (*p3*) of the sites. The log likelihood values given the models were compared using a chi square test with two degrees of freedom to test for significance ($p<0.05$). Orthogroups determined to have sites under positive selection were further tested by comparison with the M7 and M8 models, which are similar to M1 and M2, respectively, but which allow for a beta distribution of ω <1 in ten categories. Only orthogroups that displayed significance after a comparison of both M1vM2 and M7vM8 were retained as a subset of genes with sites that may have experienced recent, positive selection. Sites were considered to be under selection if they were indicated as such under the Bayes empirical Bayes analysis (Deely and Lindley 1981; Yang et al. 2005) of the M2 model with P>95%.

Genes containing sites under positive selection were additionally analyzed using the branch-site test of PAML. This test allows for ω to vary among sites and also allows the ω of a designated foreground branch to differ from the ω of the background branches. A null model A of the branch-site test was applied to the data in which four site classes are allowed, none of which include a class in which ω>1 (Table 3). This null model is compared to modified model A which allows for ω>1 in the foreground branch but not in the background branches in two of the site classes. The log likelihood values under these models were compared using a chi square test with one degree of freedom, which is a conservative test although the conditions of the regular chi square distribution are not met (Zhang et al 2005; Yang and Reis 2011), and a bonferroni correction was applied to adjust for multiple testing. Sites under positive selection were determined by the Bayes empirical Bayes approach of the modified model with P>95%.

A GO enrichment analysis can help identify whether a given set of genes related to a particular function was more highly selected for than would be expected by chance. To test for overrepresentation of particular hierarchical GO categories among positively selected genes, a

GO enrichment analysis was performed in BINGO v2.44 (Maere et al. 2005) using a hypergeometric test and false discovery rate (FDR) correction to test for significance ($p<0.05$). This analysis was performed separately for the gene set from the site test and the gene set from the branch-site test.

Selection Across a Domestication Gradient

To determine whether the number of genes influenced by positive selection increased along the domestication gradient, the proportion of genes under positive selection was calculated for each sample. To account for the uneven distribution of samples among ortholog groups, this proportion represents the ratio of the number of genes indicated to be driven by positive selection from the branch-site model to the number of ortholog groups included in the branch-site test in which that particular sample was present. To analyze the data across a domestication gradient, the eastward longitudinal distance between the island where a sample was collected and New Guinea, where the wild progenitor, *A. camansi*, originates, was obtained for each sample from Xing et al. (2012). The proportion of genes under positive selection was plotted as a function of longitudinal distance from New Guinea. To test for significance, a linear model was fitted to the data using R (R Core Team 2014). This test was conducted using the full set of samples. It was also conducted using only *A. camansi* and *A. altilis* samples because these species represent a more direct route of domestication, as Micronesian *A. mariannensis* is only thought to be involved in the hybrid breadfruit. The correlation between distance from New Guinea and the number of individual sites driven by positive selection per sample was investigated with similar methods, using a ratio of the total sites under positive selection as indicated by the modified branch-site model for a sample to the total number of bases of that sample among all the

orthogroups. Each orthogroup was also tested to determine whether signals of positive selection in that gene differed according to the region of origin, species, or ploidy level of the samples. A Fisher's exact test using Monte Carlo simulations of p-values with 2,000 replicates was used to analyze the presence or absence of positive selection signals per each gene across all samples in which that ortholog was present.

Genes of Interest

MADS-box genes were of particular interest in this study because transcription factors play an important role in developmental and regulatory processes, and changes in such genes could have influenced plant traits during the domestication process. Orthogroups of MADS-box genes were identified by searching the annotations of the corresponding reference transcripts for the terms "MADS" and "SEPALLATA" within the top blast hit description. The carotenoid pathway was also of interest because carotenoids are essential for the production of vitamin A, a vitamin in which many people are deficient. Orthogroups of carotenoid biosynthesis genes were selected by searching the GO annotations of the corresponding reference transcripts for the terms "carotenoid", "abscisic acid catabolic process", "abscisic acid metabolic process", and "abscisic acid biosynthetic process." Additional orthologous genes related to transcriptional control of carotenoid biosynthesis genes were identified by the presence of the terms "PIF", "DDB1", "DET1", "LAF1", "HY5", "HFR1", and "COP1" within the top blast hit description. The branch-site test of PAML was applied to these MADS-box and carotenoid orthogroups, and the PfamA annotations of the corresponding reference transcripts were used to determine whether discovered polymorphisms localized to particular protein domains. The McDonald-Kreitman test was also employed to test for positive selection. This test calculates the ratio of nonsynonymous

(P_n) to synonymous (P_s) polymorphisms within a species and compares it to the ratio of fixed nonsynonymous (D_n) to fixed synonymous (D_s) substitutions in that species (McDonald and Kreitman 1991). The presence of a fixed mutation is determined by utilizing at least one sequence from an outgroup species. The neutrality index (NI), where $NI = (P_n/P_s)/(D_n/D_s)$, is expected to equal one in a species under neutral selection, whereas a NI greater than one indicates purifying selection, and a NI less than one denotes positive selection. The value of alpha (α), given by $\alpha = 1-(D_sP_n)/(D_nP_s)$, signifies the proportion of polymorphisms under positive selection. The McDonald-Kreitman test was utilized in DnaSP v5 (Librado and Rozas 2009) using two different approaches. In one approach, all *A. mariannensis* and *A. camansi* samples were used as the outgroup and compared to the *A. altilis* and *A. altilis* × *A. mariannensis* hybrid samples. However, the domestication of *A. altilis* as resulting from selection on *A. camansi* without hybridization during human migration provides a more clear progression of domestication, so a second approach was used in which only *A. camansi* samples were used as an outgroup and *A. altilis* was used as the intraspecific group.

RESULTS

Sample Choice

Twenty-six samples of high quality RNA were initially chosen for sequencing that represent both diploids and triploids, both fruit and vegetative tissue, and all four taxa from the four major regions of Oceania (Figure 6). Subsequently, two *Artocarpus altilis* fruit samples were removed from the analysis because the sequence reads either predominantly consisted of adapter sequences or the assembly quality was too poor for downstream analyses. The remaining 24 samples consisted of one *A. mariannensis*, seven *A. camans*i, seven *A. altilis* × *A. mariannensis* hybrids, and nine *A. altilis* (Table 2); twelve were derived from leaf, eleven from fruit, and one

from combined root, stem, and leaf tissue. Melanesia was represented by six samples, Micronesia by nine, West Polynesia by four, and East Polynesia by five. Fourteen of the samples were diploid and ten were triploid. *Artocarpus mariannensis* was underrepresented because many trees of this species were not fruiting during the collection period and leaf extractions did not yield high quality RNA.

Assembly

A total of 484,991,615 paired end reads 101 base pairs in length were produced among the twenty-four samples. Prior to trimming, the majority of the reads contained a high representation of poly A/T tails and displayed evidence of sequencing artifacts, particularly from primers and adapters. Subsequent to trimming with TRIMMOMATIC and TRIMEST, clear evidence of primers and adapters no longer existed, and the majority of the poly A/T tails had been removed. Quality trimming did not remove all poly A/T tails, likely because TRIMEST parameters were set as to avoid aggressive trimming that may incidentally remove high quality sequences. The quality of the assemblies was measured using several different statistics. After reducing the assemblies so that only the longest isoforms were retained, the average number of contigs across samples was 44,651 contigs, average total size of contigs was 18,516,015 bp, average longest contig was 19,113,035 bp, average median contig size was 298 bp, average mean contig size was 417 bp, average N50 was 447 bp, average ortholog hit ratio was 0.28, and average collapse factor was 1.11 (Figure 7). Given the sequencing methods used, the assembly quality metrics fall within a range of metric values for reference transcriptomes generated by several other studies (Hyun et al. 2012; Hodgins et al. 2013; Liu et al. 2013).

Reference Assembly and Annotation

Pooled assemblies were created for each species, but the *A. altilis* assembly was chosen as the reference because it was the most complete and produced the best assembly quality metrics. This reference assembly contained 99,901 contigs comprised of 59,565,119 bp. It has a median contig size of 325 bp, mean contig size of 596 bp, N50 of 906 bp, average ortholog hit ratio of 0.45, and average collapse factor of 1.11. Using TRINOTATE for the annotation, 17,517 (17.5%) of the contigs received GO terms (Figure 8). A total of 50 level 2 GO terms could be applied to these contigs, 14 of which belonged to the cellular component category, 13 to the molecular function category, and 23 to the biological process category. Within these level 2 GO terms, cell (76.1%), cell part (76.1%), and organelle (51.6%) were the most highly represented terms among the cellular component annotations, catalytic activity (57.0%) and binding (63.6%) constituted most of the molecular function terms, and cellular process (61.8%) and metabolic process (54.1%) made up the majority of the biological function category.

Coverage and Variant Detection

The mean depth per sample, taking into account coverage, ranged from 0.73 bp to 13.24 bp (Figure 9). The cumulative depth at each base position of the reference assembly ranged from 0 to 23,270 reads (Figure 10). Taking into account depth and quality statistics, a total of 471,539 SNPs were detected among the samples. Out of this total, 187,306 were unique to the diploids, 147,042 were unique to the triploids, and 137,191 were common to both ploidy levels. The ratio of transitions to transversions was 1.47 in the diploids and 1.48 in the triploids.

Ortholog Detection and Signals of Positive Selection

Assembled transcripts from all the samples could be grouped into 25,878 orthogroups. After annotation, 8,691 of these ortholog groups could be described with GO terms. Running the M1vM2 site models of PAML on the full set of both annotated and unannotated orthologs suggested that 1,737 genes contained polymorphisms driven by positive selection. Further testing with the M7vM8 comparison indicated that 1,699 of these genes contained sites under positive selection using a beta distribution of ω values less than one, and influences of positive selection were therefore supported by both versions of the site model comparisons. GO terms could be assigned to 517 of these genes. A GO enrichment analysis on the annotated orthologs under positive selection compared to the full set of annotated orthologs produced 15 over-represented GO terms originating from 410 genes (Table 4). These terms were all "cellular component" ontologies and include "plastid part", "intracellular part", "chloroplast thylakoid lumen", "plastid thylakoid lumen", "intracellular", "chloroplast thylakoid", "plastid thylakoid", "chloroplast part", "thylakoid lumen", "thylakoid", "thylakoid part", "organelle subcompartment", "plastid stroma", "cell part", and "cell" (Figure 11). No significantly over-represented terms were discovered for the biological processes or molecular function categories of GO terms.

The branch-site test of PAML was additionally applied to determine if certain sites were impacted by positive selection in particular branches by individually designating each sample as a foreground branch in which ω was allowed to exceed one. Of the 1,699 genes, 1,011 displayed signals of positive selection in an indicated foreground branch, 300 of which could be annotated with GO terms. A GO enrichment analysis on this set of genes produced 219 genes that contributed to the over-represented terms "intracellular part and intracellular" within the cellular

component ontologies (Table 5). Biological process and molecular function terms were not over-represented.

Selection Across a Domestication Gradient

The proportions of genes and sites under positive selection were tested to determine whether these values increased along a longitudinal gradient representative of the domestication gradient. Considering the full sample set, the proportion of positively selected genes did not differ among distances from New Guinea according to a linear model ($F=0.20$ on 1 and 22 d.f., $p=0.66$), nor did the proportion of sites driven by positive selection (Figure 12; $F=0.22$ on 1 and 22 d.f., $p=0.64$). Including only *A. camansi* and *A. altilis* samples to represent a more clear progression of the domestication gradient, the proportion of positively selected genes also did not differ along longitudinal distance according to a linear model ($F=0.22$ on 1 and 14 d.f., $p=0.65$), nor did the proportion of positively selected sites ($F=0.22$ on 1 and 14 d.f., $p=0.65$).

Fisher's exact tests were applied to determine whether signals of positive selection differed according to ploidy level, region of origin, or species. One gene, which could not be annotated, was found to differ with ploidy level (Figure 13; $p<0.05$). This gene was present in eight of the samples, but only displayed signals of positive selection in three individuals, all of which were diploid. Signals of positive selection did not vary among samples from different regions. In two of the genes, ATP-dependent Clp protease proteolytic subunit-related protein 1(ClPR1) and F-box protein At5g49610, positive selection differed according to species ($p<0.05$). ClPR1 was identified in twenty of the samples, but was only found to be under positive selection in the *A. mariannensis* sample. F-box protein At5g49610 was present in nine samples,

but only two individuals, the only *A. camansi* samples included in this subset, appeared to be driven by positive selection in this gene.

Genes of Interest

Among the orthogroups, fourteen MADS-box genes were identified. Six of these genes had sites driven by positive selection in particular lineages, as was determined using the branch-site test of PAML (Table 6). Positive selection was indicated in a triploid *A. altilis* sample from East Polynesia for Agamous-like MADS-box protein AGL8 homolog, in a diploid *A. camansi* sample from Melanesia for MADS-box protein CMB1, in a diploid *A. camansi* sample from Micronesia for MADS-box protein AGL9 homolog, and in a diploid *A. mariannensis* sample from Micronesia for MADS-box transcription factor 6, MADS-box protein SVP, and MADS-box protein SOC1. Five out of six sites under positive selection that were found to be significant using a BEB analysis in SOC1 localized to a SRF-type transcription factor (DNA-binding and dimerisation domain), and one site localized to a K-box region. SNPs also localized to K-box regions for six out of six positively selected sites in AGL8, five out of nine positively selected sites in CMB1, and five out of seven positively selected sites in transcription factor 6. Sites of positive selection did not localize to particular domains in SVP or AGL9. A McDonald-Kreitman test comparing cultivated breadfruit, including *A. altilis* × *A. mariannensis* hybrids and *A. altilis*, to their wild relatives, *A. mariannensis* and *A. camansi*, did not reveal signals of selection within any of the explored MADS-box genes (Table 7). Additionally, the McDonald-Kreitman test using *A. altilis* as the intraspecific group and *A. camansi* as the outgroup did not suggest that any MADS-box genes were under positive selection.

Genes related to carotenoid biosynthesis were identified as belonging to 40 orthogroups. The branch-site test of PAML suggests that two of these genes, zeaxanthin epoxidase and 15-cise-phytoene desaturase, have undergone positive selection in certain branches (Table 8). The zeaxanthin epoxidase gene displays signals of positive selection in a triploid *A. altilis* sample from West Polynesia, and the 15-cis-phytoene desaturase gene contains sites under positive selection in a diploid *A. camansi* sample from Micronesia. The two sites under positive selection in PDS localized to a NAD(P)-binding Rossmann-like domain and flavin containing amine oxidoreductase domain. The positively selected sites in ZEP did not localize to any particular domains. The McDonald-Kreitman test indicated that two out of the 40 genes, carotene epsilon-monooxygenase and molybdenum cofactor sulfurase, are driven by positive selection when comparing cultivated breadfruit to an outgroup of its wild relatives (Table 9). Carotene epsilon-monooxygenase also showed signals of positive selection in a comparison of *A. altilis* samples to an *A. camansi* outgroup, whereas the molybdenum cofactor sulfurase orthogroup contained an insufficient number of *A. altilis* samples to conduct the McDonald-Kreitman test. Positive selection was not supported in any other genes related to carotenoid biosynthesis using this version of the McDonald-Kreitman test.

DISCUSSION

Breadfruit is an underutilized crop that can serve as an economical food source and provide essential micronutrients in food insecure regions. Cultivars of breadfruit display variation in nutritional content, seasonality, genetic diversity, and morphology because many varieties have been developed by human-mediated selection for over 3,000 years. Based on amplified fragment length polymorphisms (AFLPs) and isozymes, it is thought that humans domesticated *Artocarpus altilis* from *A. camansi* as they moved from Melanesia throughout the Polynesian

islands (Ragone 1991; Zerega et al. 2006). The history of the development of breadfruit in Micronesia is less clear, with multiple proposed introductions and hybridization events with different amounts of introgression with the wild relative *A. mariannensis* (Fosberg 1960; Zerega et al. 2004, 2006). Nonetheless, studies have shown that people selected for certain phenotypic characteristics in breadfruit, such as fewer seeds, a reduced center core, and smoother skin (Jones et al. 2013b). The dispersal history of breadfruit and its close ties with human migration routes make it a good candidate for studying the process of domestication (Zerega et al. 2004; Xing et al. 2012).

The ability to sequence nearly complete transcriptomes from non-model species and relatively little starting material through the use of next generation sequencing technologies provides a beneficial platform for studying the process of domestication in various crop species and producing genomic resources for these organisms (Parchman et al. 2010; Strickler et al. 2012; Howe et al. 2013; Liu et al. 2014; Wang et al. 2014). Transcriptomes can be particularly useful for non-model organisms, such as breadfruit, because only the expressed genes are sequenced. This method therefore remains cost effective while reducing the complexity involved in the de-novo assembly of species which lack a reference genome and avoids issues related to the c-value paradox, such as potentially large variability in the size of non-coding regions among species (Strickler et al. 2012; Hodgins et al. 2013). The focus of this study was to characterize transcriptomes of breadfruit, identify putative genes driven by positive selection, and analyze such genes to reveal patterns correlated with the domestication gradient.

The domestication process in breadfruit was studied using RNA-seq to generate transcriptomes for 24 individuals representing various taxa, ploidy levels, and regions of origin. A reference assembly constructed from only the *A. altilis* samples consisted of 50 level two GO

terms, providing the opportunity to study many different pathways and particular processes within breadfruit. Genes could be compared across samples within the 25,878 ortholog groups that were detected. These ortholog groups were analyzed for signatures of positive selection both across the entire transcriptome and within particular pathways of interest.

The site and branch-site test were utilized to screen for positive selection within all the orthogroups of the transcriptomes. The site model indicated that 1,699 genes contained sites driven by positive selection, and a GO enrichment analysis showed that particular GO terms were overrepresented in the cellular component category. Many of these terms were specific to the plastids and in particular to the thylakoids. No biological process or molecular function terms were overrepresented, possibly because a large variety of genes with different functions or belonging to different pathways were under selection, many of which localized to plastids. The overrepresented genes had a wide array of blast hits, including those related to translation factors, transcription factors, pathogen resistance, carotenoid biosynthesis, starch and sugar metabolism, DNA repair, and methyltransferase activity, amongst many others. The fact that many of these indicated genes or their protein products were localized to plastids is of particular interest because breadfruit is a starch crop and plastids are the location of starch biosynthesis and storage (Geigenberger 2011). Mutations in plastid specific genes could impact the quantity or composition of starch in breadfruit, and because breadfruit has traditionally been used as a main source of starch, humans may have paid particular attention to starch-related traits during domestication. Many carotenoid biosynthesis genes also localize to plastids (Figure 14). Studies suggest that carotenoid levels can vary considerably among cultivars (Jones et al. 2013a). Selection on genes that localize to the plastid or the carotenoid biosynthesis genes themselves may be impacting this variation. The branch-site model was subsequently utilized, and it

detected 1,011 genes with sites under positive selection in particular samples, and those related to intracellular GO terms were enriched. These terms encompass many of the overrepresented terms discovered with the site model, but not to the extent that they are enriched in GO categories more explicit than the “intracellular” and “intracellular parts” categories. Overall, a multitude of genes displayed signals of positive selection among both the site and branch-site test, suggesting that many genes may have been influenced either during the domestication of breadfruit or through natural selection.

Signals of positive selection were additionally screened using only a subset of the ortholog groups which were identified as MADS-box genes. MADS-box genes are considered to be primary targets of domestication because of their role in regulating gene expression and controlling complex traits. They affect plant parts as diverse as lateral roots, pollen tubes, leaf stomata, ovules, petals, and stamens (Smaczniak et al. 2012). MADS-box genes consist of a DNA-binding domain which allows it to interact with the gene it regulates and a K-box region which is thought to form a coiled-coil that promotes protein dimerisation (Shore and Sharrocks 1995). Using the branch-site test, MADS-box genes AGL8, CMB1, AGL9, SVP, SOC1, and transcription factor 6 can putatively be considered to be driven by positive selection in particular samples (Table 6). Sites under positive selection were found in the DNA-binding domain of SOC1, which may impact its interactions with the genes it regulates. SOC1, AGL8, CMB1, and transcription factor 6 contained sites under positive selection in K-box regions, potentially influencing protein-protein interactions. The six MADS-box genes under positive selection are involved in processes such as the timing of flowering and the development of floral organs (Gu et al. 1998; Mandel and Yanofsky 1998; Sung et al. 2000; Lee and Lee 2010; Li et al. 2010; Wu et al. 2012). Numerous MADS-box genes have been implicated as playing roles in the

domestication process, particularly in their control on fruit morphology and seed dispersal (Smaczniak et al. 2012). Their roles in domestication have been documented in species as various as tomato, wheat, cauliflower, maize, soybean, and sunflower (Purugganan et al. 2000; Vrebalov et al. 2002; Liu et al. 2004; Cockram et al. 2007; Blackman et al. 2011; Zhao et al. 2011; Kim et al. 2012; Xie et al. 2014). Given the strong selection for differing seediness and floral structure in breadfruit as evident in morphological differences, MADS-box genes are good candidates for genes that were impacted during domestication. MADS-box genes can also aid plants in adapting to new environments, such as by altering the timing of flowering in response to different climate conditions (Kane et al. 2005; Mariac et al. 2011; Meyer and Purugganan 2013). Breadfruit has spread from its site of origin to various islands in Oceania. This exposure to many different environments may have necessitated changes in seasonality and flowering control. The signals of positive selection in MADS-box genes were detected in various species, ploidy levels, and regions. The variety of selection amongst these genes may result from the need for different forms of adaptation or different human-mediated selection pressures in the species throughout Oceania.

The carotenoid pathway was also examined for evidence of positive selection (Figure 5). This pathway is responsible for photo-oxidative protection and produces pigments imparting colors to various plant organs, but it also plays an important role in producing substrates that act as precursors in stress response pathways. Carotenoids are of particular interest in breadfruit because they can serve as necessary precursors of vitamin A, and focus has been placed on promoting the use of cultivars with superior carotenoid production (Jones et al. 2013a; Englberger et al. 2014). A better understanding of the variation of carotenoid production among breadfruit cultivars can be reached by elucidating differences in selection pressures among the

cultivars. Utilizing the branch-site test, two genes in the carotenoid biosynthesis pathway, zeaxanthin epoxidase (ZEP) and 15-cis-phytoene desaturase (PDS), showed signs of positive selection in particular samples. ZEP is putatively under positive selection in a triploid *A. altilis* sample from West Polynesia (Table 8). ZEP catalyzes the conversion of zeaxanthin into antheraxanthin and subsequently into violaxanthin, which is further converted to neoxanthin by neoxanthin synthase (Figure 5; Ruiz-Sola and Rodríguez-Concepción 2012). In circumstances of high light intensity, violaxanthin can be converted back into zeaxanthin, which is more efficient at dissipating light energy as heat; ZEP therefore plays an important role in the adaptation of plants to different light conditions (Demmig-Adams et al. 1996; Ruiz-Sola and Rodríguez-Concepción 2012). ZEP is also an important factor in stress response because of its role in producing violaxanthin and neoxanthin, which can be converted into abscisic acid (ABA), a hormone which functions in abiotic stress response (Schwartz et al. 1997; Lu and Li 2008). The partial control by ZEP in producing ABA from carotenoid substrates can deplete carotenoid precursors (Vallabhaneni and Wurtzel 2009), and ZEP is therefore involved in both carotenoid biosynthesis and stress response pathways. Studies show that the overexpression of ZEP confers greater salt and drought tolerance in *Arabidopsis thaliana* (Park et al. 2008), the mutation of ZEP proteins leads to a 30% increase in carotenoid levels in tomato (Galpaz et al. 2008), allele differences in ZEP affect yellow coloration in potatoes (McCord et al. 2012), and pigment accumulation related to ZEP differs between yellow-colored cultivated carrots and white-colored wild carrots (Just et al. 2009). The inhibition of ZEP can lead to an increase in the total amount of carotenoids, and ZEP is thought to have wide-spanning regulatory effects on other processes of the carotenoid pathway in potatoes (Römer et al. 2002). A previous study by Jones et al. (2013a) analyzed the total and specific carotenoid levels in 95 samples of breadfruit from the

germplasm collection. The *A. altilis* sample for which the branch-site test indicated ZEP as being driven by positive selection ranked 18th for highest amount of total carotenoids. Nine individuals that ranked above this sample were designated as elite cultivars that produced superior levels of provitamin A carotenoids and many lower-ranking individuals produced considerably lower amounts of carotenoids, demonstrating that this sample produces a relatively high amount of carotenoids, potentially as a result of the effects of ZEP. The possibility that ZEP plays a large role in affecting total carotenoids levels makes this gene a good candidate for being a target of positive selection meriting further study.

Using the branch-site test, positive selection was also detected in a diploid *A. camansi* sample from Micronesia (Table 8) within the PDS gene. The protein encoded by this gene converts phytoene to ζ-carotene, which is subsequently converted by ζ-carotene desaturase (ZDS) to lycopene, a red-colored carotenoid which does not have provitamin A activity itself, but which is involved in further steps of the pathway in producing the provitamin A carotenoids β-carotene and α-carotene (Figure 5; Aluru et al. 2008). Inhibition of PDS can lead to an accumulation of phytoene and a decrease in carotenoid levels. Inactivation of PDS has been shown to lead to the production of white tissues in *A. thaliana* (Stange and Flores 2012), reduction of PDS expression has been documented in carrots displaying decreased pigmentation levels (Bowman 2012), and upregulation of PDS has been observed in the transition from the mature green to ripe orange stage in tomatoes (Fraser et al. 1994). Furthermore, the overexpression of bacterial homologs of PDS and ZDS have been shown to be necessary in increasing the levels of carotenoids produced in transgenic maize 34-fold (Aluru et al. 2008). The sites under positive selection in PDS localized to a NAD(P)-binding Rossmann-like domain and a flavin containing amine oxidoreductase domain. NAD(P), NAD, and FAD are potential

cofactors of phytoene desaturase that can act by binding to a Rossmann fold, and polymorphisms in the binding site for cofactors may alter the function of the enzyme (Chamovitz et al. 1991; Pecker et al. 1992; Schaub et al. 2012). Oxidoreductases catalyze the removal of hydrogen atoms, and desaturases specifically remove two hydrogen atoms from fatty acids. Polymorphisms in the oxidoreductase domain of PDS may therefore functionally change the enzyme. The *A. camansi* sample with this gene under positive selection ranked 53rd out of 95 samples in a study examining total carotenoid levels (Jones et al. 2013a). The PDS gene is of interest in this sample because it acts early in the carotenoid biosynthesis pathway and could impact many downstream genes, therefore altering total carotenoid levels found in breadfruit.

A second test for positive selection, the McDonald-Kreitman test, was also utilized to explore the carotenoid genes. This test was utilized both using *A. camansi* as the outgroup and *A. altilis* as the intraspecific group, and additionally including *A. mariannensis* in the outgroup and *A. altilis* x *A. mariannensis* hybrids in the intraspecific group. Both versions of the McDonald-Kreitman test used in this study showed carotene epsilon-monooxygenase, an ε-ring hydroxylase (LUT1), to be under positive selection. LUT1 predominantly hydroxylates α-carotene to produce lutein (Figure 5; Diretto et al. 2010). Its downregulation leads to an accumulation of α-carotene in transgenic "golden" potatoes (Diretto et al. 2010), it is considered a target for manipulation in producing transgenic maize with enhanced carotenoid levels (Owens et al. 2014), and the mutation of it leads to reduced amounts of lutein in *A. thaliana* (Pogson et al. 1996; Kim and DellaPenna 2006). Lutein is the predominate carotenoid produced by breadfruit (Jones et al. 2013a) and is likely responsible for most of the yellow coloration found in some fruits. Research suggests that lutein production varies among breadfruit cultivars, and along the domestication gradient, many breadfruit exhibit decreased yellow-coloration, which is thought to be an effect of

human-mediated selection (Jones et al. 2013a, 2013b). However, wild relatives, particularly *A. mariannensis*, have stronger yellow coloration (Ragone 1997; Jones et al. 2013a, 2013b). Mutations in domesticated breadfruit cultivars may have led to variation in lutein levels, making LUT1 a candidate target gene of domestication that should be further explored.

Using *A. camansi* and *A. mariannensis* as the outgroup and both *A. altilis* and *A. altilis* x *A. mariannensis* as the intraspecific group, molybdenum cofactor sulfurase (MOCOS) was indicated to experience positive selection. MOCOS regulates the last step of ABA biosynthesis by catalyzing the sulfurization of molybdenum cofactor, which is required by aldehyde oxidase to oxidize xanthoxin and produce ABA (Bittner et al. 2001). ABA then acts as a hormonal signal to activate stress-inducible genes, such as those that close stomata and reduce transpirational water loss during drought stress (Xiong et al. 2001). MOCOS therefore plays an essential role in the response to stress and the ability of plants to withstand such stresses. Transgenic maize overexpressing MOCOS and exposed to drought stress produce a higher amount of ABA compared to non-transgenic maize (Lu et al. 2013), barley with mutated MOCOS have reduced amounts of ABA when exposed to temperature stress compared to controls (Walker-Simmons et al. 1989), tomatoes lacking a sulfurated molybdenum cofactor exhibit a wilted phenotype (Sagi et al. 2002), and *A. thaliana* plants with mutated MOCOS have decreased tolerance of freezing, drought, and high salt conditions (Xiong et al. 2001). The dispersal of breadfruit throughout many different islands with varying environments has likely exposed the trees to different types and amounts of stresses, which may have required alterations in ABA biosynthesis and the MOCOS gene.

Overall, the branch-site test and McDonald-Kreitman test revealed four genes related to the carotenoid biosynthesis pathway that are putatively driven by positive selection, including

zeaxanthin epoxidase, 15-cis-phytoene desaturase, carotene epsilon-monooxygenase, and molybdenum cofactor sulfurase. The genes can be involved not only in producing carotenoid end products, but also in generating precursors essential to the stress response pathways, such as the ABA biosynthesis pathway, as well as precursors of hormones stimulating interactions with mycorrhizal fungi, such as strigolactone (Smith 2014). A previous study by Xing et al. (2012) indicated that along the domestication gradient, breadfruit roots exhibited a decrease in the colonization intensity, but not richness, of mycorrhizal fungi. Interactions with mycorrhizal fungi can affect water and nutrient absorption, stress tolerance, and pathogen resistance in breadfruit (Xing et al. 2012). Positive selection of carotenoid-related genes could have impacted the regulation of hormones such as strigolactone and influenced changes in the composition or intensity of root colonization along the domestication gradient. Although carotenoid biosynthesis genes were found to be under selection, a targeted search of transcriptional controls related to carotenoid biosynthesis, including PIF, DBB1, DET1, LAF1, HY5, HFR1, and COP1, did not reveal any indication of positive selection. Regulation of the carotenoid pathway is not fully understood, but positive selection on the enzyme-encoding genes rather than transcription factors of carotenoids could be responsible for the variation in carotenoid content observed across the domestication gradient. Studies on crop diversification suggest that conscious selection plays a strong role in this process and that the selection more often occurs on enzyme-encoding genes (Meyer and Purugganan 2013). The pigments that carotenoids produce impart a visible trait that may have been a target during conscious selection for certain breadfruit varieties and may explain the presence of positive selection in the enzyme-encoding genes.

Furthermore, patterns of positive selection were analyzed across a longitudinal gradient as well as among species, ploidy levels, and regions of origin. Regardless of whether *A.*

mariannensis and *A. altilis* × *A. mariannensis* hybrids were included in this gradient, neither the number of genes nor the number of sites driven by positive selection exhibited changes along the domestication gradient. Therefore, the hypothesis that signals of positive selection, as represented by counts of genes and sites, would increase along the domestication gradient was not supported. As people travelled throughout the islands of Oceania, they may have selected for different traits but not necessarily a greater number of traits as they developed various cultivars. Moreover, the domestication of a species from its wild relative often only necessitates changes in a small number of genes (Mayes et al. 2012). For example, the morphological differences between maize and teosinte are largely attributed to only five genes, and the domestication of wheat involved approximately seven domestication syndrome factors (Doebley and Stec 1991; Peng et al. 2013).

The presence of positive selection signals varied according to ploidy for one gene and differed among species for two genes. Such signals did not vary with region. Although the gene that differed among diploids and triploids did not receive any annotations, this gene could have been involved with the transition of heavily seeded diploid fruit to seedless triploid cultivars. The gene that was discovered to be under positive selection only in *A. mariannensis* is an ATP-dependent Clp protease that functions in the stroma to degrade proteins (Adam and Clarke 2002). Although these proteases are housekeeping genes and are constitutively expressed, they are thought to play an important role in stress tolerance by degrading damaged proteins produced by stress induction (Adam and Clarke 2002; Kidric et al. 2014). The mechanisms and specific substrates involved are largely unknown, but studies have shown variation in ClP protein and transcript abundance in drought tolerant wheat, heat stressed barley, low temperature stressed rice, and greening delayed *A. thaliana* lines (Cui et al. 2005; Sjögren et al. 2006; Chi et al. 2008;

Demirevski et al. 2008; Bazargani et al. 2011; Rollins et al. 2013). Wild relatives tend to be better adapted to surviving stress, whereas stress tolerance is often lost during the process of domestication, possibly from being linked to undesirable traits (Gorovits and Czosnek 2008). *Artocarpus mariannensis* is native to the harsh low lying atoll conditions of the Mariana Islands where salt spray is more common, and breadfruit without *A. mariannensis* traits do not grow well in atoll regions (Ragone 1988). Clp proteases could be involved in the higher stress tolerance observed in *A. mariannensis* as compared to the other taxa involved in this study, although results should be interpreted with caution as *A. mariannensis* was only represented by one sample.

Signals of positive selection within a F-box protein were found among the *A. camansi* samples, but not among any of the other species. Within SCF complexes (Skp1-Cullin-F-box proteins), which catalyze the ubiquitination of proteins thus marking them for degradation, F-box proteins are the subunits which confer specificity in targeting certain substrates (Lechner et al. 2006; Jain et al. 2007; Wang et al. 2009). A variety of F-box proteins have been discovered and are thought to be involved in processes including photomorphogenesis, flower development, the ethylene response pathway, and circadian clock regulation (Jain et al. 2007; Tacken et al. 2012; Wang et al. 2009). F-boxes have also been found to be differentially expressed in response to abiotic stresses (Calderón-Villalobos et al. 2007; Jain et al. 2007; Yan et al. 2011). F-box proteins were not previously considered targets of domestication, but have been found to be driven by positive selection during the domestication of maize in a study that used untargeted screenings for selection (Wright et al. 2005; Doebley et al. 2006). This study, as well as the indication of positive selection in a F-box protein within *A. camansi* samples, highlights the necessity of examining species for targets of positive selection without a priori assumptions in

addition to studying previously discovered candidate genes or screening specific pathways. Both the F-box protein and ATP-dependent Clp proteases displayed signals of positive selection exclusively in the wild relatives, suggesting that these genes may have been altered during the development of cultivated breadfruit.

Searches for positive selection were conducted using the McDonald-Kreitman test, site tests, and branch-site test. These tests vary in their advantages and disadvantages and do not necessarily converge on the same conclusions. The McDonald-Kreitman test benefits from its consideration of fixed polymorphisms within a species and the divergence between species, but errors can occur in the presence of slightly deleterious mutations or with differences in effective population sizes (Eyre-Walker 2002). This test was employed to test for positive selection by comparing wild and cultivated breadfruit. In contrast, the site-test was used to detect positive selection within codons of a gene across all breadfruit samples, and the branch-site test was utilized to search for positive selection in particular samples. The branch-site test is considered more powerful than the site-test and the branch-test, but this power is also dependent on the length of the sequences, the strength of positive selection, the sequence divergence, and the proportion of sites in the foreground branch under positive selection (Yang and dos Reis 2011). This test loses power when each branch is tested individually as the foreground branch as opposed to making a priori assumptions, but it is still valuable for identifying candidate genes for further examination. All of these tests take into consideration the ratio of nonsynonymous to synonymous mutations. Nonsynonymous mutations alter the amino acid sequence, making them more likely than synonymous mutations to alter the structure and function of a protein and potentially be selected against. These nonsynonymous mutations are of particular interest when

they are located in certain functional domains of a protein, such as a DNA-binding domain or active site, as was found to be the case for several genes in this study.

Although this research focused on identifying mutations underlying positive selection, many SNPs were discovered in the samples that can be used for other applications. The genetic diversity of breadfruit has previously been studied with isozymes, AFLPs, and microsatellites (Ragone 1991; Zerega et al. 2004, 2005, in press). Similar to these studies, the genetic diversity and differentiation in breadfruit can be utilized to analyze how the diversity of breadfruit has changed along the domestication gradient and to determine the presence or strength of potential genetic bottlenecks. Genetic markers for the SNPs discovered in the transcriptomes can be developed, adding to the diversity of markers sets available for breadfruit. These markers can then be used in studies targeting particular genes and will be useful for screening larger sample sets from the germplasm collection.

Overall, this transcriptome analysis provides a glance at the magnitude and diversity of effects that human-mediated selection can have on a species and aids in the understanding of domestication. Similar to what has been found in many other domesticated crops, several MADS-box genes were indicated to be impacted by positive selection in breadfruit. Carotenoid biosynthesis genes were also found to be under positive selection, and future studies could explore whether these particular genes are linked to the variation in carotenoid production among the cultivars. Although variation in breadfruit has been indentified for some traits that exhibit clear phenotypes, less obvious traits or traits that may have unintentionally been selected for may have been impacted during the domestication process. This study identified over 1,000 genes putatively under positive selection. Each of these genes can be studied in depth to determine whether they confer traits that may influence the conservation or use of breadfruit, such as

pathogen resistance or seasonality. In addition, species and ploidy specific signals of positive selection were indentified in breadfruit and its wild relatives, providing candidate genes that may underlie traits exclusive to these species and ploidy levels. No differences were observed in the number of positively selected genes or sites along the domestication gradient, but various genes driven by positive selection were discovered within all the samples. These results highlight the importance of conserving a large variety of individual breadfruit trees, as many may have genes that have been altered through domestication or natural selection with yet unknown affects on phenotypes. Therefore, germplasm collection efforts should focus on conserving breadfruit and its wild relatives from a variety of regions throughout Oceania so that those with unseen benefits are not overlooked.

CONCLUSION

To counteract people's reliance on only a few major crops, efforts are being taken to promote locally grown foods that will offer a more diverse and sustainable diet. Breadfruit is a low-energy input tree crop that provides many micronutrients and grows in food insecure regions. To identify superior cultivars, much focus has been placed on understanding the phenotypic variation among breadfruit, which has been highly influenced by the domestication process as people dispersed it throughout islands in Oceania. However, little is known about the genetic variation that may underlie this trait variation. This study has explored the transcriptomes of breadfruit to determine genetic patterns that correlate with the domestication gradient and to reveal putative genes under positive selection that could potentially affect desirable traits. It has also provided genomic resources that can be used in future studies on the diversity and domestication of breadfruit, which will be particularly useful for those aiming at making it a

more widespread crop. Greater utilization of breadfruit will provide a sustainable and nutritious crop to areas where increased costs associated with energy-inputs hinder economic development and food security.

FIGURES AND TABLES

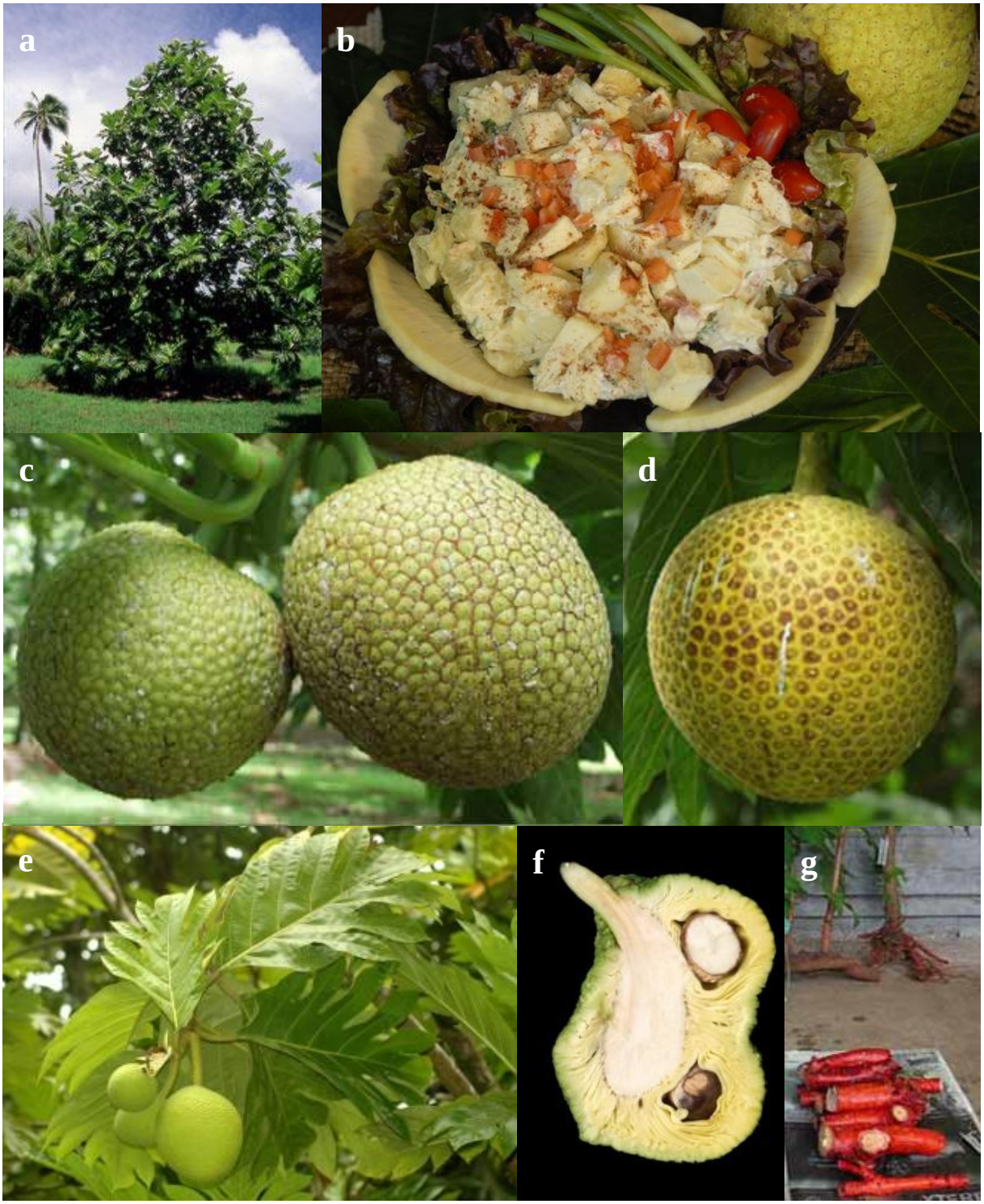
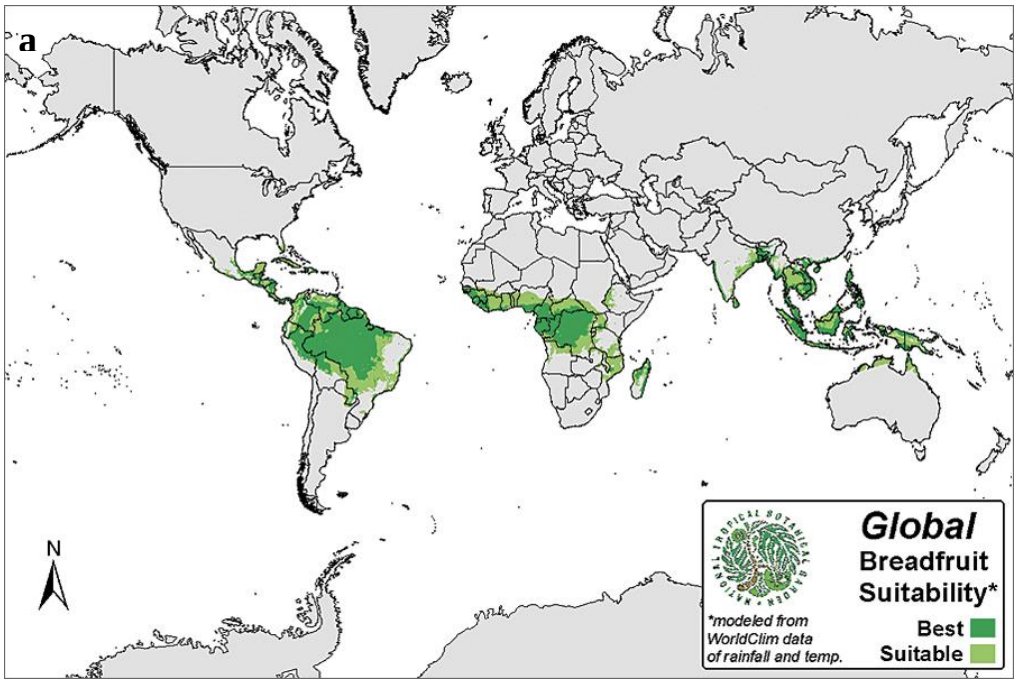
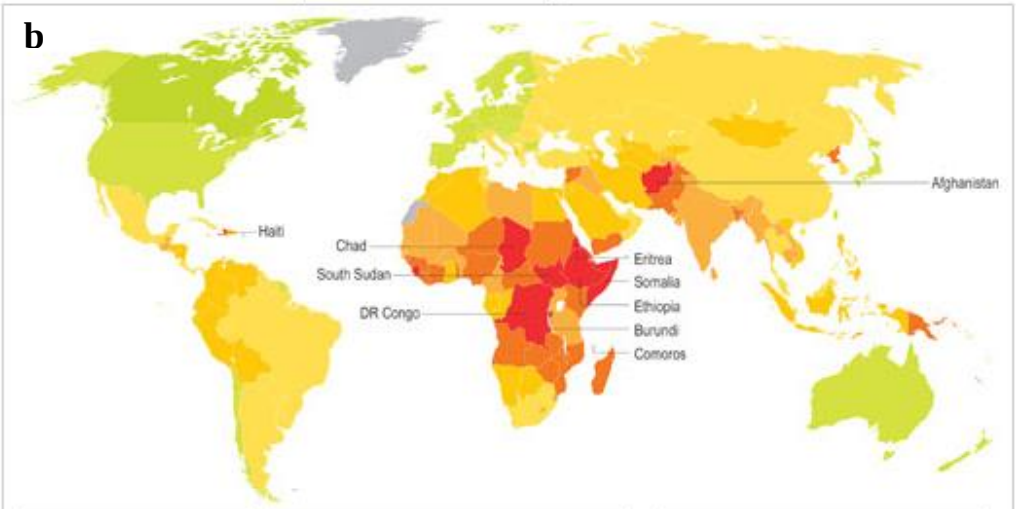


Figure 1. Breadfruit. a) Breadfruit tree, b) Salad prepared with breadfruit, c) Immature breadfruit on left and mature breadfruit on right, d) Ripe breadfruit, e) Fruit and leaves of Samoan 1 variety, f) Cross section of dugdug variety, g) Root cuttings of breadfruit for vegetative propagation. All images obtained from the Breadfruit Institute (http://www.ntbg.org/breadfruit/).



b



Legend	
Extreme risk	
High risk	
Medium risk	
Low risk	
No Data	

Rank	Country	Rating
1	Somalia	Extreme
2	DR Congo	Extreme
3	Haiti	Extreme
4	Burundi	Extreme
5	Chad	Extreme

Rank	Country	Rating
6	Ethiopia	Extreme
7	Eritrea	Extreme
8	Afghanistan	Extreme
9	South Sudan	Extreme
10	Comoros	Extreme

Figure 2. Suitable habitat range for breadfruit and regions of food insecurity. a) Best and suitable regions to grow breadfruit according to rainfall and temperature (http://www.ntbg.org/breadfruit/), b) Maplecroft's food security risk index in 2013 according to country (Maplecroft: http://www.maplecroft.com).

Table 1. Mineral content comparison between breadfruit, wheat, rice, and corn flour (table modified from Jones et al. 2011). All values are means with letters depicting significant differences with a type one error rate of 0.05 (nd=value below detection limit).

Species	Total Bulk Samples	Mineral Content(µg/g dry weight)									
		Ca	Co	Cu	Fe	K	Mg	Mn	Na	P	Zn
A. altilis	82	583^{b}	0.011^{ab}	2.6^{bc}	11^{c}	$11,081^{a}$	$1,022^{c}$	1.7^{g}	195^{b}	$1,319^{e}$	2.6^{f}
A. altilis × *A. mariannensis*	17	559^{bc}	0.010^{ab}	3.1^{d}	15.1^{b}	$11,399^{a}$	933^{c}	1.9^{f}	$1,222^{a}$	$1,390d^{e}$	3.1^{e}
A. camansi	2	658^{b}	0.007^{bc}	6.4^{a}	14.1^{b}	$10,876^{a}$	1337^{b}	2.9^{e}	75^{b}	$3,017^{b}$	6.9^{d}
A. mariannensis	1	$1,188^{a}$	0.014^{a}	5.0^{b}	17.5^{b}	$10,702^{a}$	1783^{a}	5.3^{d}	$1,288^{a}$	$1,583^{cd}$	$3.8e^{f}$
Whole wheat flour	1	249^{bcd}	0.006^{bcd}	$4.0b^{c}$	34.5^{a}	$2,828^{b}$	1694^{ab}	31.8^{a}	8^{b}	$4,802^{a}$	28^{a}
Fortified refined wheat flour	1	145^{cd}	0.002^{cd}	2.0^{ef}	38.5^{a}	982^{b}	388^{d}	5.7^{d}	12^{b}	$1,516^{cde}$	8.5^{cd}
Refined wheat flour	1	198^{d}	-	2.3^{ef}	7.9^{d}	899^{b}	342^{d}	8.5^{c}	-	$1,413^{de}$	10.9^{c}
Refined rice flour	1	61^{d}	0.010^{ab}	3.0^{cde}	5.8^{d}	$1,008^{b}$	490^{d}	18.2^{b}	56^{b}	$1,819^{c}$	18^{b}
Corn flour	1	17^{d}	nd	1.4^{f}	5.9^{d}	$1,108^{b}$	327^{d}	1.1^{g}	58^{b}	953^{f}	4.4^{e}

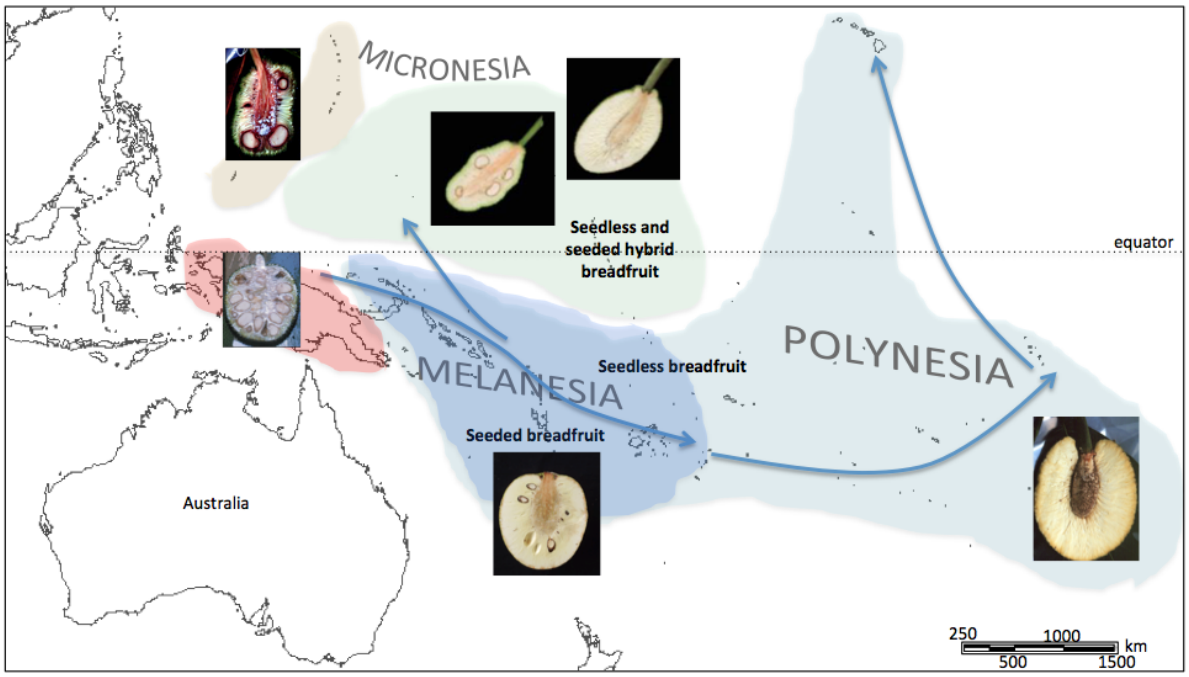


Figure 3. Dispersal and domestication of breadfruit throughout Oceania (image modified from Zerega et al. 2014). Arrows show direction of domestication. Images show the distribution of variation among wild relatives and cultivars that were sampled in this study.

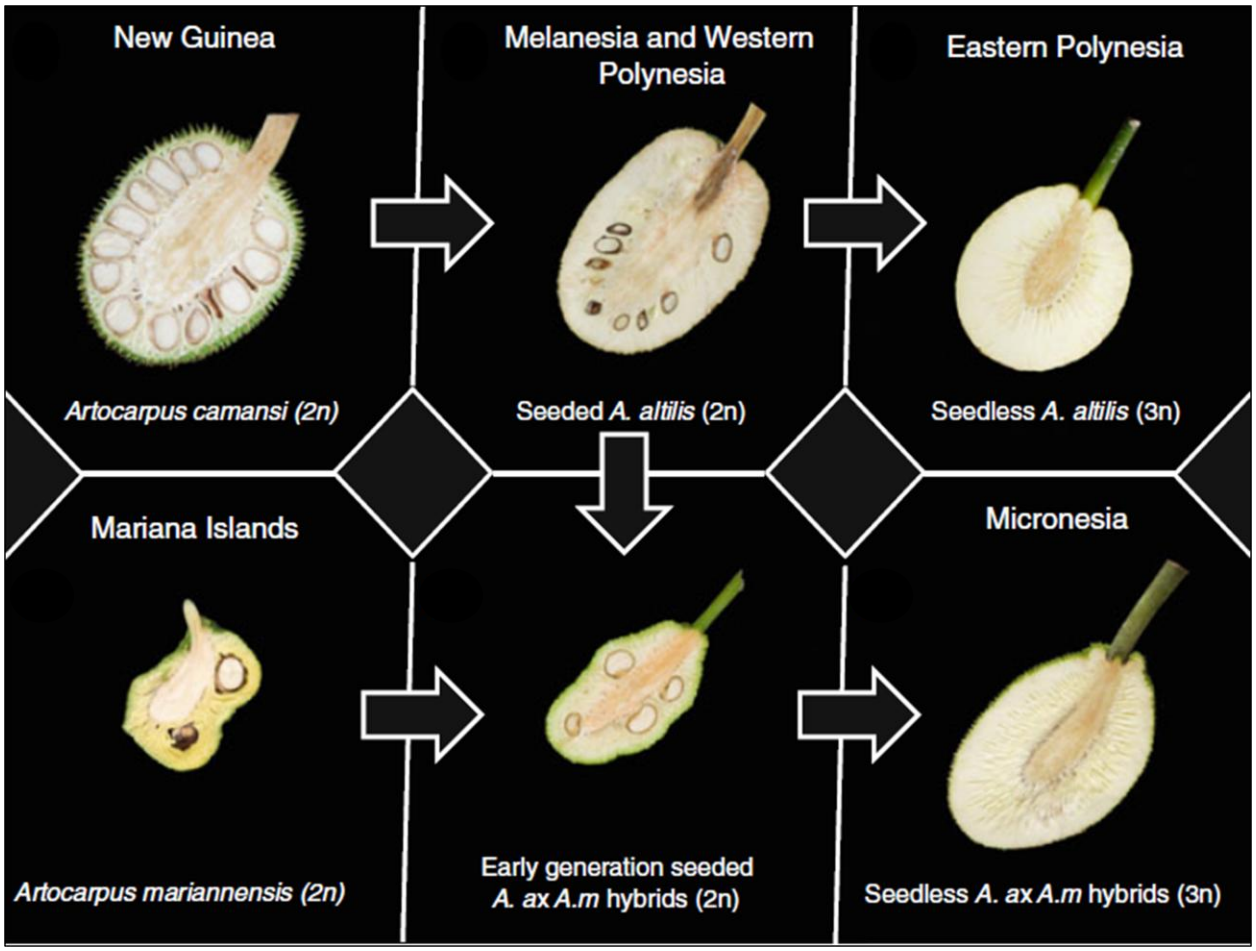


Figure 4. Differences in size and seediness of *A. altilis* and its wild relatives (Jones et al. 2013b). Arrows depict the direction of domestication.

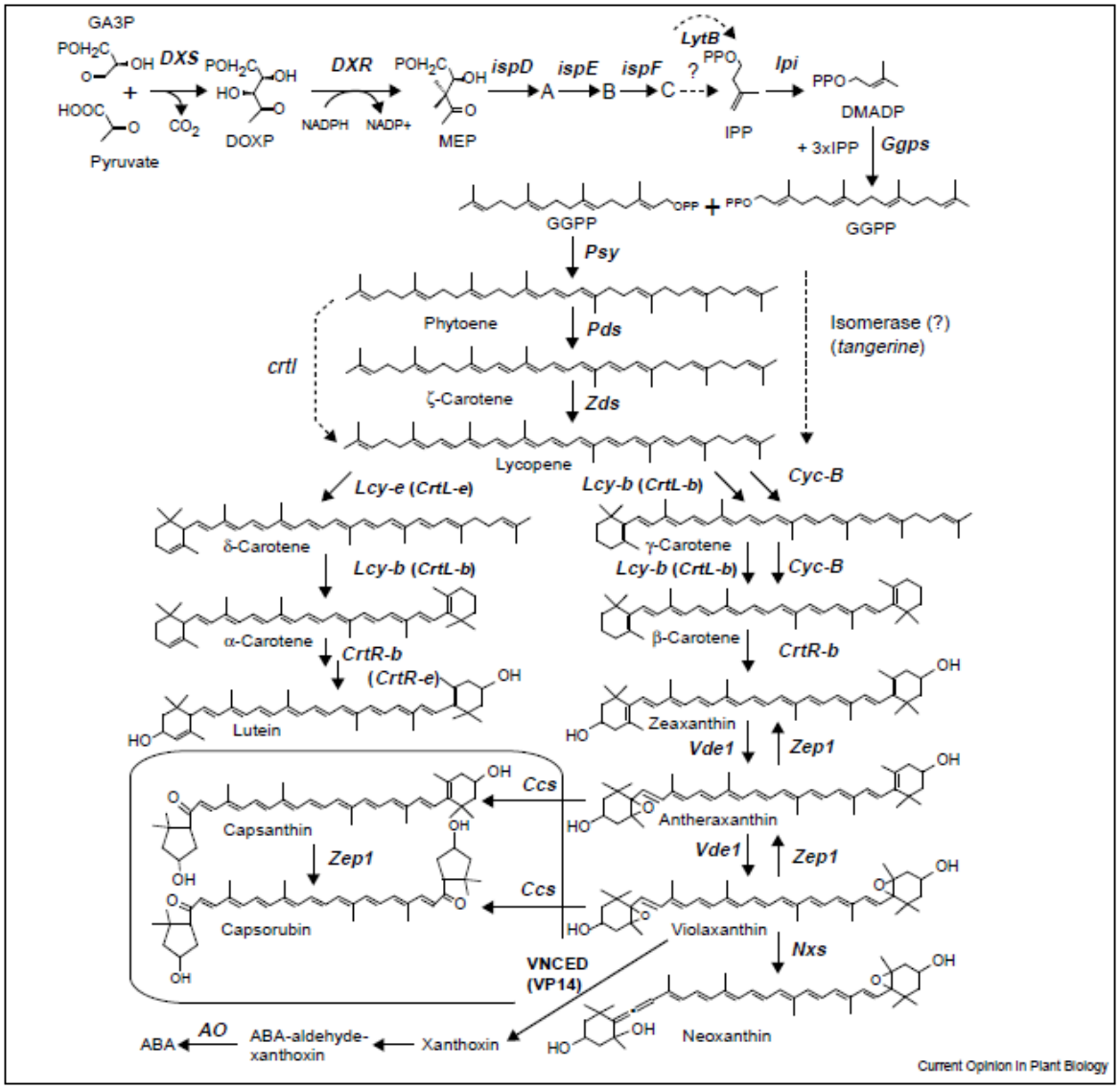


Figure 5. Carotenoid biosynthesis pathway in plants (Hirschberg 2001). A=DPME, B=4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate, C=2C-methyl-Derythritol 2,4-cyclodiphosphate, ABA=abscisic acid, AO=aldehyde oxidase, *CrtR-b*=beta-carotene 3-hydroxylase, *CrtR-e* (LUT1)=carotene epsilon-monooxygenase, Cyc-B=chromoplast-specific lycopene cyclase, DMADP=dimethylallyl diphosphate, DOXP=1-deoxy-D-xylulose 5-phosphate, DXR=DOXP reductoisomerase, DXS=DOXP synthase, GGPP=geranylgeranyl diphosphate, Ipi=IPP isomerase gene, IPP=isopentenyl diphosphate, ispD=DPME synthase, ispE DPME kinase, ispF=2C-methyl-D-erythritol 2,4 cyclodiphosphate, Lcy-b/CrtL-b lycopene β-cyclase, Lcy-e/CrtL-e lycopene ε-cyclase, MEP=2-C-methyl-Derythritol 4-phosphate, Nxs=neoxanthin synthase gene, Pds=phytoene desaturase, Psy=phytoene synthase, Vde=violaxanthin de-epoxidase, VNCED (VP14)=9-*cis*-epoxycarotenoid dioxygenase, ZDS=ζ-carotene desaturase, Zep1=Zeaxanthin epoxidase. The box designates a pathway that takes place in the chromoplast of pepper fruits.

Table 2. Sample choices and characteristics (veg.=root, stem, leaf; Pld=ploidy). More information on each sample can be obtained from the Breadfruit Institute of the National Tropical Botanical Garden (NTBG; http://www.ntbg.org/breadfruit/).

Species	Sample Name	NTBG Accession No.	Tissue	Pld	Region	Variety	Island of Origin
A. mariannensis	RNA7	900252.002	fruit	2	Micronesia	Dugdug	Mariana Islands
A. camansi	EW2[K]	501.005	fruit	2	Melanesia	Kapiak	Papua New Guinea
	EW4[K]	501*	veg.	2	Melanesia	Kapiak	Papua New Guinea
	RNA35	389.001	fruit	2	Melanesia	Kapiak	Papua New Guinea
	RNA2	910280.001	fruit	2	Micronesia	Meikole	Pohnpei, FSM
	RNA37	980212.001	fruit	2	Micronesia	Camansi	Palau
	RNA5	770444.001	fruit	2	East Polynesia	Camansi	Society Islands
	RNA34	890455.001	fruit	2	West Polynesia	Ulu fatu	Samoa
A. altilis × *A. mariannensis*	RNA48	890184.001	leaf	2	Micronesia	Luthar	Yap
	RNA16	890183.001	leaf	3	Micronesia	Midolab	Palau
	RNA40	910269.001	leaf	3	Micronesia	Faine	Chuuk,FSM
	RNA17	790487.001[c]	leaf	3	Micronesia	Unk (Huehue)	Pohnpei, FSM
	RNA32	890174.001	fruit	2	West Polynesia	Ulu afa	Tokelau
	RNA36	890173.002	fruit	2	West Polynesia	Ulu afa elise	Tokelau
	RNA10	790490.001[c]	leaf	3	East Polynesia	Rotuma	Society Islands
A. altilis	EW1[K]	970236**	fruit	2	Melanesia	Ulu fiti	Rotuma
	RNA21	900265.001	fruit	2	Melanesia	Karawa	Fiji
	RNA24	900261.001	leaf	2	Melanesia	Samoan 2	Fiji
	RNA25	890479.002[aG]	leaf	3	Micronesia	Meisei	Pohnpei, FSM
	RNA26	890167.002[aG]	leaf	3	Micronesia	Meisaip	Pohnpei, FSM
	RNA38	880690.001[b]	leaf	3	West Polynesia	Kea	Tonga
	EW3[K]	30042.001	fruit	3	East Polynesia	Toneno	Society Islands
	RNA39	790485.001[b]	leaf	3	East Polynesia	Puupuu	Society Islands
	RNA49	910265.001[a]	leaf	3	East Polynesia	Rotuma	Society Islands

*seed offspring of 501.005; **vegetative offspring of 970236.001; [a,b,c] three distinct lineage groups and [G]one genotype represented by multiple samples (Zerega et al. in press); [K]collections from Kauai, HI

Table 3. Parameters of the null and modified branch-site models implemented in PAML.

Site Class	Proportion	Background	Foreground	
			Null model	Modified Model
0	p_0	$0<\omega_0<1$	$0<\omega_0<1$	$0<\omega_0<1$
1	p_1	$\omega_1=1$	$\omega_1=1$	$\omega_1=1$
2a	$(1- p_0- p_1)\, p_0/(p_0+ p_1)$	$0<\omega_0<1$	$\omega_2=1$	$\omega_2>1$
2b	$(1- p_0- p_1)\, p_1/(p_0+ p_1)$	$\omega_1=1$	$\omega_2=1$	$\omega_2>1$

Figure 6. Sample Distribution. Islands from which samples originated are depicted. Tonga, Samoa, and Tokelau are encompassed in West Polynesia, and the Society Islands are part of East Polynesia. The number of samples originating from each island is noted in parentheses. FSM=Federated States of Micronesia.

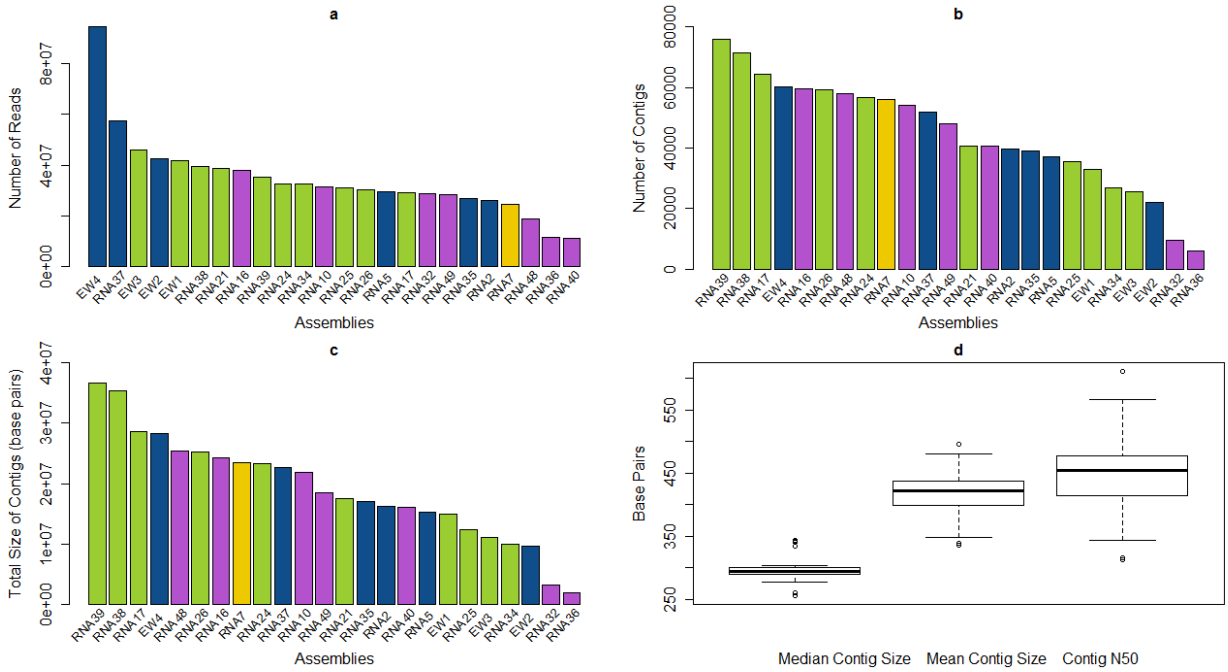


Figure 7. Assembly metrics. a) Number of trimmed reads used for each assembly b) Number of contigs across samples, c) Total size of contigs across samples, d) Median, mean, and N50 for all samples. Yellow=*A. mariannensis*, blue=*A. camansi*, purple=*A. altilis* × *A. mariannensis* hybrids, green=*A. altilis*.

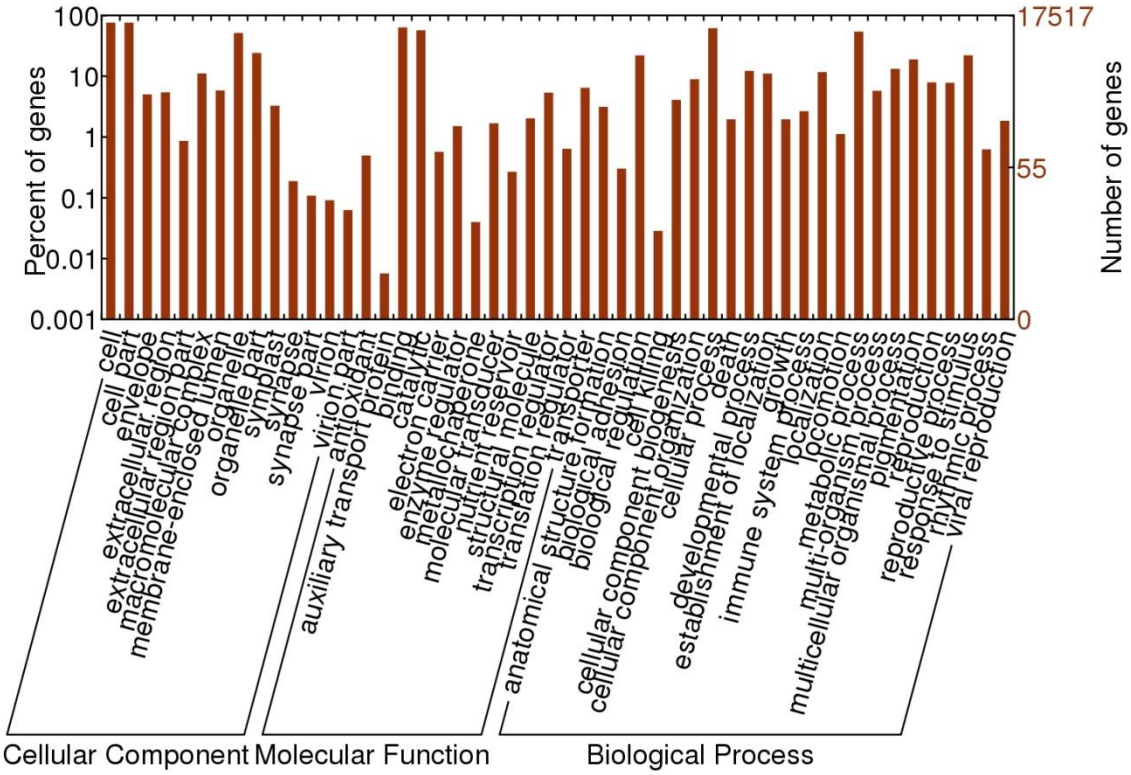


Figure 8. Wego output. Level 2 GO terms of cellular components, molecular functions, and biological process resulting from functional annotation of the pooled assembly consisting of all *A. altilis* samples.

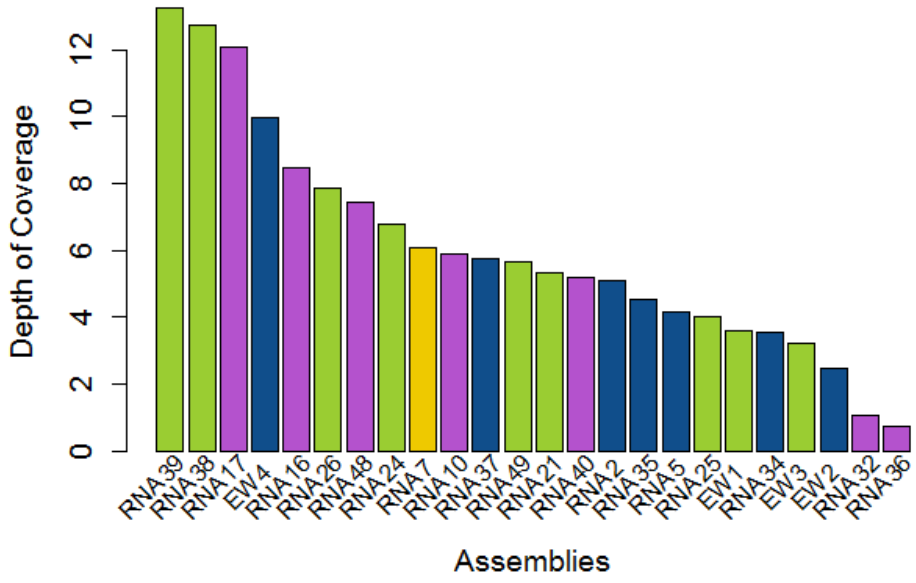


Figure 9. Depth of coverage per sample. Depth of coverage is measured as the average depth across all bases of the reference transcriptome, including those in which a sample did not have coverage. Yellow=*A. mariannensis*, blue=*A. camansi*, purple=*A. altilis* × *A. mariannensis* hybrids, green=*A. altilis*.

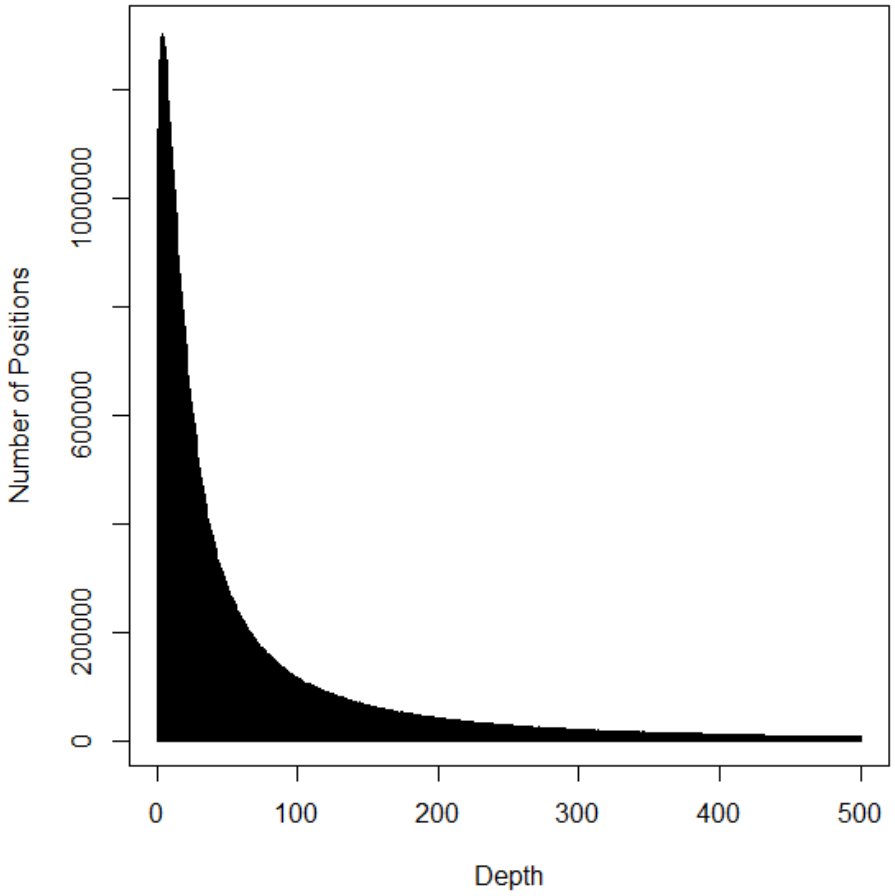


Figure 10. Distribution of read depths for bases with a minimum base quality of 17 and minimum mapping quality of 20. Base positions with a depth above 500 reads are not shown.

Table 4. Over-represented GO terms of genes indicated to be under positive selection using the M1vM2 and M7vM8 site model tests. All terms refer to cellular components. X=total number of genes in the subset of positively selected genes with cellular component annotations, x=number of genes with the indicated GO ID in the subset of positively selected genes, N=total number of genes with cellular component annotations, n=number of genes with the indicated GO ID in the total set of genes.

GO Description	GO ID	P-value	Corrected P-value	x/X	% x/X	n/N	% n/N
plastid part	44435	6.31E-05	7.93E-03	59/428	13.79	605/7254	8.34
intracellular part	44424	7.91E-05	7.93E-03	370/428	86.45	5765/7254	79.47
chloroplast thylakoid lumen	9543	1.34E-04	7.93E-03	7/428	1.64	21/7254	0.29
plastid thylakoid lumen	31978	1.34E-04	7.93E-03	7/428	1.64	21/7254	0.29
intracellular	5622	1.37E-04	7.93E-03	370/428	86.45	5785/7254	79.75
chloroplast thylakoid	9534	1.52E-04	7.93E-03	26/428	6.07	204/7254	2.81
plastid thylakoid	31976	1.52E-04	7.93E-03	26/428	6.07	204/7254	2.81
chloroplast part	44434	1.90E-04	8.69E-03	57/428	13.32	602/7254	8.3
thylakoid lumen	31977	2.54E-04	9.93E-03	8/428	1.87	30/7254	0.41
thylakoid	9579	2.71E-04	9.93E-03	28/428	6.54	235/7254	3.24
thylakoid part	44436	4.80E-04	1.60E-02	23/428	5.37	184/7254	2.54
organelle subcompartment	31984	7.65E-04	2.16E-02	26/428	6.07	226/7254	3.12
plastid stroma	9532	8.21E-04	2.16E-02	31/428	7.24	289/7254	3.98
cell part	44464	8.86E-04	2.16E-02	407/428	95.09	6606/7254	91.07
cell	5623	8.86E-04	2.16E-02	407/428	95.09	6606/7254	91.07

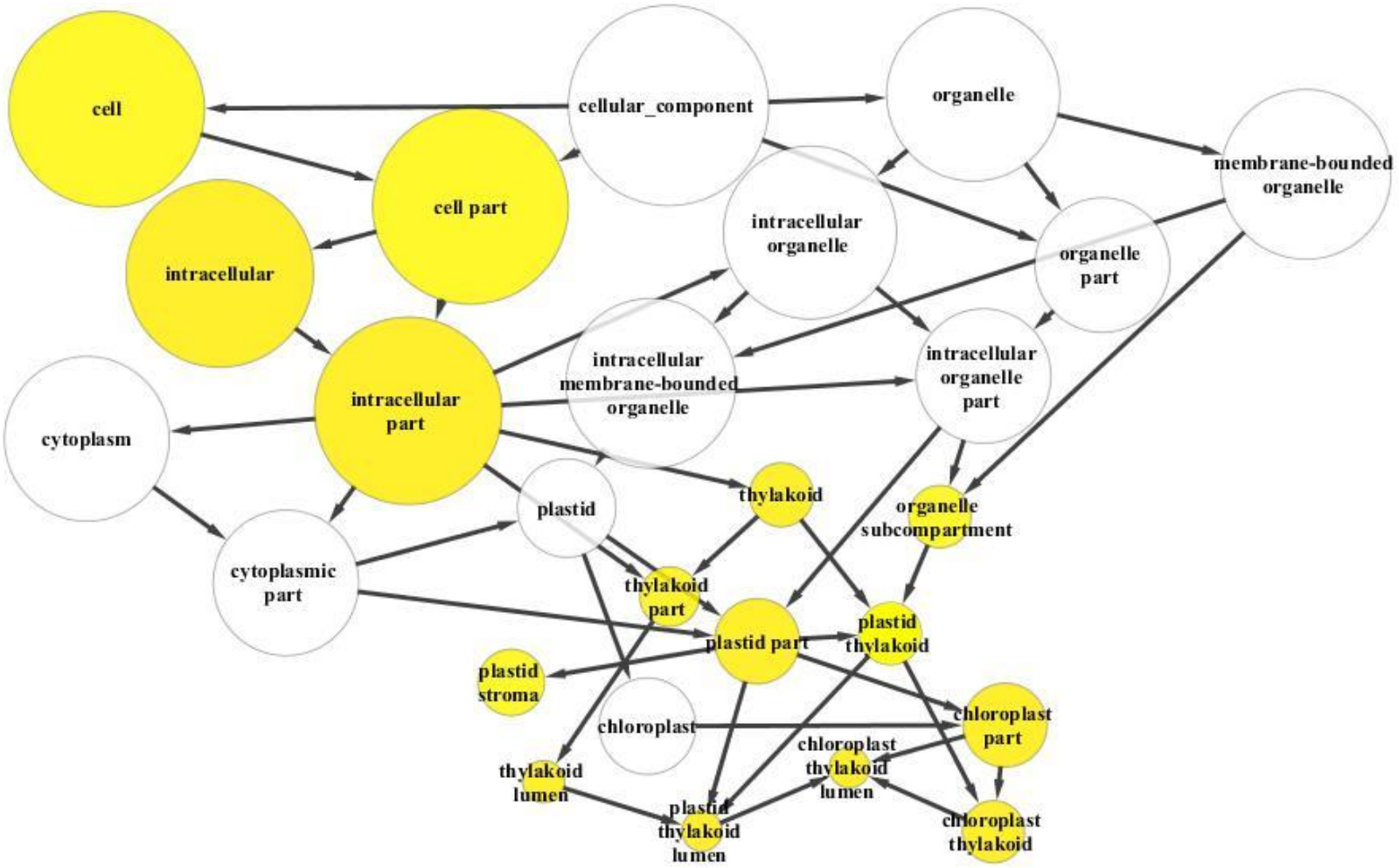


Figure 11. Network of overrepresented GO terms for genes under positive selection as determined by PAML's site models. Overrepresented GO categories are indicated in yellow and node side is proportional to the number of terms within that category.

Table 5. Over-represented GO terms of genes indicated to be under positive selection using the branch-site test. All terms refer to cellular components. X=total number of genes in the subset of positively selected genes with cellular component annotations, x=number of genes with the indicated GO ID in the subset of positively selected genes, N=total number of genes with cellular component annotations, n=number of genes with the indicated GO ID in the total set of genes.

GO Description	GO ID	P-value	Corrected P-value	x/X	% x/X	n/N	% n/N
intracellular part	44424	4.13E-05	8.83E-03	219/246	89.02	5765/7254	79.47
intracellular	5622	6.28E-05	8.83E-03	219/246	89.02	5785/7254	79.75

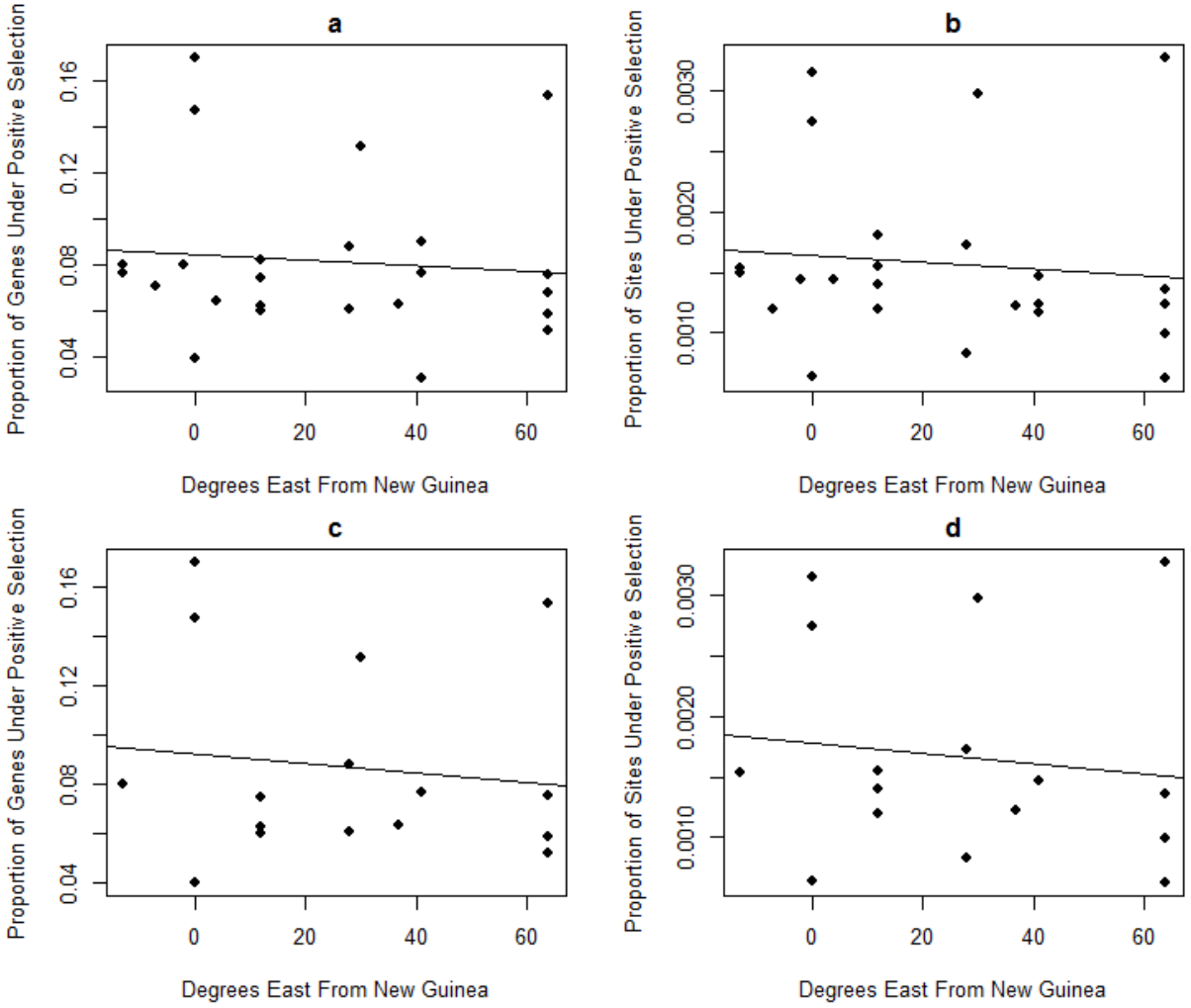


Figure 12. Signals of positive selection across a longitudinal gradient. a) Proportion of genes under positive selection for all samples along longitudinal distance from New Guinea (R^2=-0.04, F=0.20 on 1 and 22 d.f., p=0.66), b) Proportion of sites under positive selection for all samples along longitudinal distance from New Guinea (R^2=-0.04, F=0.22 on 1 and 22 d.f., p=0.64), c) Proportion of genes under positive selection along longitudinal distance from New Guinea with *A. mariannensis* and *A. altilis* × *A. mariannensis* hybrid samples omitted (R^2=-0.06, F=0.22 on 1 and 14 d.f., p=0.65), d) Proportion of sites under positive selection along longitudinal distance from New Guinea with *A. mariannensis* and *A. altilis* × *A. mariannensis* hybrid samples omitted (R^2=-0.06, F=0.22 on 1 and 14 d.f., p=0.65).

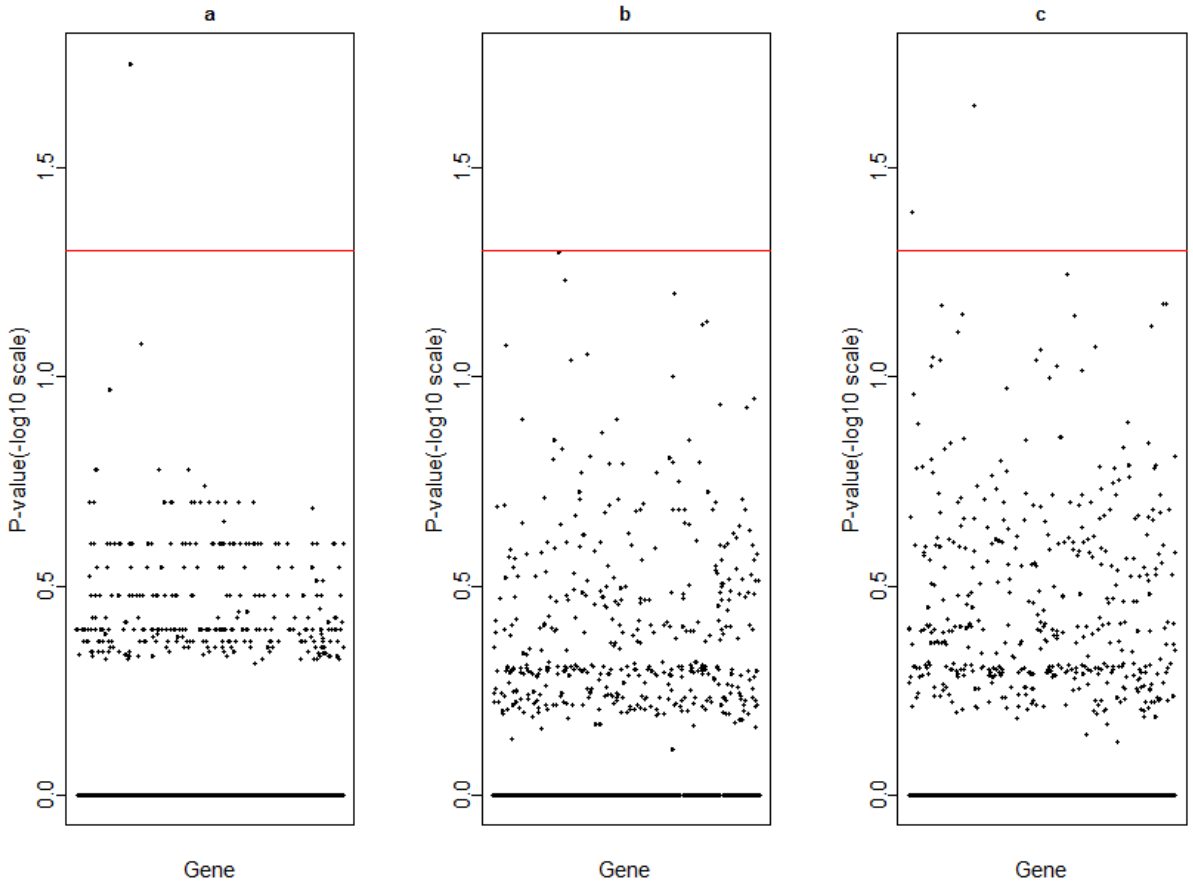


Figure 13. Association between signals of positive selection and ploidy, region, and species for 1,011 genes. a) P-values of Fisher's exact test to determine whether signals of positive selection differed according to ploidy level, b) P-values of Fisher's exact test to determine whether signals of positive selection differed according to region, c) P-values of Fisher's exact test to determine whether signals of positive selection differed according to species. The red line denotes the 0.05 significance level with points above the line representing p-values less than 0.05.

Table 6. Branch-site test results for MADS-box genes indicated to be under positive selection. In site class 2a, omega of the background branches is $0< \omega_0<1$ and omega of the foreground branch is $\omega_2>1$. In site class 2b, omega of the background branches is $\omega_1=1$ and omega of the foreground branch is $\omega_2>1$. Chi^2=chi-square test statistic, Pr=proportion of sites, Bω=omega of background branches, Fω=omega of foreground branch, Sites=number of sites driven by positive selection (P>95%), Sp=species (M=*A. mariannensis*, C=*A. camansi*, H= *A. altilis* × *A. mariannensis* hybrids, A=*A. altilis*), Pl=ploidy, Re=region of origin (M=Melanesia, m=Micronesia, W=West Polynesia, E=East Polynesia).

				Site Class 2a				Site Class 2b						
Ortholog ID	Branch	Blast ID	Blast Description	Chi^2	Pr	Bω	Fω	Pr	Bω	Fω	Sites	Sp	Pl	Re
9853	RNA7	SOC1_ARATH	MADS-box protein SOC1	20.30	0.04	0.00	999.00	0.02	1.00	999.00	6.00[ab]	M	2.00	m
1276	RNA49	AGL8_SOLTU	Agamous-like MADS-box protein AGL8 homolog	18.98	0.04	0.08	999.00	0.02	1.00	999.00	6.00[b]	A	3.00	E
2546	RNA7	SVP_ARATH	MADS-box protein SVP	19.21	0.06	0.08	999.00	0.03	1.00	999.00	7.00	M	2.00	m
25077	RNA37	AGL9_PETHY	Agamous-like MADS-box protein AGL9 homolog	28.81	0.11	0.00	999.00	0.06	1.00	999.00	17.00	C	2.00	m
1275	RNA7	MADS6_ORYSJ	MADS-box transcription factor 6	14.08	0.07	0.05	195.69	0.07	1.00	195.69	7.00[b]	M	2.00	m
17512	EW2	CMB1_DIACA	MADS-box protein CMB1	21.00	0.08	0.10	999.00	0.07	1.00	999.00	9.00[b]	C	2.00	M

[a]One or more SNPs localized to a SRF-type transcription factor (DNA-binding and dimerisation domain)
[b]One or more SNPs localized to a K-box region

Table 7. McDonald-Kreitman test for positive selection within MADS-box orthogroups. Version 1 uses *A. mariannensis* and *A. camansi* as the outgroup and all cultivated breadfruit, which includes *A. altilis* and *A. altilis* × *A. mariannensis* hybrids, as the intraspecific group. Version 2 uses *A. camansi* as the outgroup and *A. altilis* as the intraspecific group. G-values and P-values represent the respective values after applying William's correction. Metrics that could not be calculated are indicated by "NA" and the inability to conduct the test because of a lack of appropriate data is denoted by "--".

			Version 1				Version 2			
Ortholog ID	Gene	Description	Neutrality Index	Alpha	G value	P value	Neutrality Index	Alpha	G value	P value
1275	MADS6_ORYSJ	MADS-box transcription factor 6	0.418	0.582	1.188	0.27574	0.784	0.216	0.11	0.7399
1276	AGL8_SOLTU	Agamous-like MADS-box protein AGL8 homolog	1.142	-0.142	0.036	0.8504	1.141	-0.141	0.035	0.85146
1674	JOIN_SOLLC	MADS-box protein JOINTLESS	NA	NA	NA	NA	--	--	--	--
2546	SVP_ARATH	MADS-box protein SVP	NA	NA	NA	NA	NA	NA	NA	NA
3420	AGL9_PETHY	Agamous-like MADS-box protein AGL9 homolog	0	1	NA	NA	--	--	--	--
3465	SOC1_ARATH	MADS-box protein SOC1	NA	NA	NA	NA	NA	NA	NA	NA
4393	SVP_ARATH	MADS-box protein SVP	NA	NA	NA	NA	NA	NA	NA	NA
6312	CMB1_DIACA	MADS-box protein CMB1	NA	NA	NA	NA	NA	NA	NA	NA
9222	MAD16_ORYSJ	MADS-box transcription factor 16	NA	NA	NA	NA	NA	NA	NA	NA
9852	AGL14_ARATH	Agamous-like MADS-box protein AGL14	NA	NA	NA	NA	1.308	-0.308	0.027	0.87007
9853	SOC1_ARATH	MADS-box protein SOC1	1.17	-0.17	0.066	0.79719	--	--	--	--
17512	CMB1_DIACA	MADS-box protein CMB1	NA	NA	NA	NA	NA	NA	NA	NA
25077	AGL9_PETHY	Agamous-like MADS-box protein AGL9 homolog	0	1	NA	NA	0	1	NA	NA
25044	SEP2_ARATH	Developmental protein SEPALLATA 2	1.546	-0.546	0.972	0.32428	--	--	--	--

Table 8. Branch-site test results for carotenoid genes indicated to be under positive selection. In site class 2a, omega of the background branches is $0< \omega_0<1$ and omega of the foreground branch is $\omega_2>1$. In site class 2b, omega of the background branches is $\omega_1=1$ and omega of the foreground branch is $\omega_2>1$. Chi^2=chi-square test statistic, Pr=proportion of sites, Bω=omega of background branches, Fω=omega of foreground branch, Sites=number of sites driven by positive selection (P>95%), Sp=species (M=*A. mariannensis*, C=*A. camansi*, H= *A. altilis* × *A. mariannensis* hybrids, A=*A. altilis*), Pl=ploidy, Re=region of origin (M=Melanesia, m=Micronesia, W=West Polynesia, E=East Polynesia).

					Site Class 2a			Site Class 2b						
Ortholog ID	Branch	Blast ID	Blast Description	Chi^2	Pr	Bω	Fω	Pr	Bω	Fω	Sites	Sp	Pl	Re
6680	RNA38	ABA2_PRUAR	Zeaxanthin epoxidase, chloroplastic	25.37	0.20	0.52	999.00	0.00	1.00	999.00	12.00[ab]	A	3.00	W
7721	RNA2	PDS_ARATH	15-cis-phytoene desaturase, chloroplastic/ chromoplastic	13.34	0.04	0.00	999.00	0.00	1.00	999.00	2.00[ab]	C	2.00	m

[a]One or more SNPs localized to a NAD(P)-binding Rossmann-like domain

[b]One or more SNPs localized to a flavin containing amine oxidoreductase domain

Table 9. McDonald-Kreitman test for positive selection within orthogroups related to carotenoid biosynthesis. Version 1 uses *A. mariannensis* and *A. camansi* as the outgroup and all cultivated breadfruit, which includes *A. altilis* and *A. altilis* × *A. mariannensis* hybrids, as the intraspecific group. Version 2 uses *A. camansi* as the outgroup and *A. altilis* as the intraspecific group. G-values and P-values represent the respective values after applying William's correction. Metrics that could not be calculated are indicated by "NA" and the inability to conduct the test because of a lack of appropriate data is denoted by "--". Significant terms are signified by bold type.

			Version 1				Version 2			
Ortholog ID	Gene	Description	Neutrality Index	Alpha	G value	P value	Neutrality Index	Alpha	G value	P value
315	ABAH3_ORYSJ	Abscisic acid 8'-hydroxylase 3	1.022	-0.022	0.006	0.93781	--	--	--	--
544	HY5_ARATH	Transcription factor HY5	NA	NA	NA	NA	NA	NA	NA	NA
1210	ALDO2_ARATH	Indole-3-acetaldehyde oxidase	NA	NA	NA	NA	NA	NA	NA	NA
2388	CESA8_ARATH	Cellulose synthase A catalytic subunit 8 [UDP-forming]	NA	NA	NA	NA	NA	NA	NA	NA
3063	Probable polyamine transporter At3g19553	Probable polyamine transporter At3g19553	NA	NA	NA	NA	NA	NA	NA	NA
3149	CRTSO_SOLLC	Prolycopene isomerase, chloroplastic	NA	NA	NA	NA	NA	NA	NA	NA
3169	ZEP_ARATH	Zeaxanthin epoxidase, chloroplastic	NA	NA	NA	NA	--	--	--	--
4407	HMOX1_ARATH	Heme oxygenase 1, chloroplastic	NA	NA	NA	NA	NA	NA	NA	NA
4414	MOCOS_SOLLC	Molybdenum cofactor sulfurase	0	1	NA	NA	--	--	--	--
5203	DET1_YEAST	Broad-range acid phosphatase DET1	NA	NA	NA	NA	--	--	--	--
5267	MOCOS_SOLLC	Molybdenum cofactor sulfurase	--	--	--	--	--	--	--	--
5551	PSY_DAUCA	Phytoene synthase, chloroplastic	NA	NA	NA	NA	NA	NA	NA	NA

5821	GGPPS_HEVBR	Geranylgeranyl pyrophosphate synthase, chloroplastic	NA	NA	NA	NA	NA	NA	NA	NA
6239	BCH2_CAPAN	Beta-carotene hydroxylase 2, chloroplastic	NA	NA	NA	NA	0	1	NA	NA
6255	LCYB_CAPAN	Lycopene beta cyclase, chloroplastic/ chromoplastic	NA	NA	NA	NA	NA	NA	NA	NA
6536	CRTSO_SOLLC	Prolycopene isomerase, chloroplastic	NA	NA	NA	NA	NA	NA	NA	NA
6680	ABA2_PRUAR	Zeaxanthin epoxidase, chloroplastic	NA	NA	NA	NA	NA	NA	NA	NA
7425	LUT5_ARATH	Protein LUTEIN DEFICIENT 5, chloroplastic	NA	NA	NA	NA	NA	NA	NA	NA
7542	CCD1_PHAVU	Carotenoid 9,10(9',10')-cleavage dioxygenase 1	NA	NA	NA	NA	0.706	0.294	0.131	0.7179
7649	COP1_ARATH	E3 ubiquitin-protein ligase COP1	NA	NA	NA	NA	NA	NA	NA	NA
7650	COP1_ARATH	E3 ubiquitin-protein ligase COP1	NA	NA	NA	NA	NA	NA	NA	NA
7721	PDS_ARATH	15-cis-phytoene desaturase, chloroplastic/ chromoplastic	NA	NA	NA	NA	NA	NA	NA	NA
8609	GGPPS_HEVBR	Geranylgeranyl pyrophosphate synthase, chloroplastic	NA	NA	NA	NA	NA	NA	NA	NA
9210	LUT1_ARATH	Carotene epsilon-monooxygenase, chloroplastic	0.182	0.818	5.608	**0.01788**	0.182	0.818	5.608	**0.01788**
9549	LCYE_ARATH	Lycopene epsilon cyclase, chloroplastic	NA	NA	NA	NA	NA	NA	NA	NA
9910	PUB44_ARATH	U-box domain-containing protein 44	NA	NA	NA	NA	NA	NA	NA	NA
12904	MOCOS_ARATH	Molybdenum cofactor sulfurase	4.747	-3.747	17.725	**0.00003**	--	--	--	--

13577	DET1_YEAST	Broad-range acid phosphatase DET1	NA	NA	NA	NA	NA	NA	NA	NA
13824	MOCOS_SOLLC	Molybdenum cofactor sulfurase	NA	NA	NA	NA	NA	NA	NA	NA
14362	HMOX1_ARATH	Heme oxygenase 1, chloroplastic	--	--	--	--	--	--	--	--
15083	ABAH1_ARATH	Abscisic acid 8'-hydroxylase 1	NA	NA	NA	NA	NA	NA	NA	NA
15503	AOX4_ARATH	Ubiquinol oxidase 4, chloroplastic/ chromoplastic	NA	NA	NA	NA	NA	NA	NA	NA
16492	PSY_DAUCA	Phytoene synthase, chloroplastic	NA	NA	NA	NA	NA	NA	NA	NA
16997	NCED3_ARATH	9-cis-epoxycarotenoid dioxygenase NCED3, chloroplastic	NA	NA	NA	NA	NA	NA	NA	NA
17004	ABA2_ARATH	Xanthoxin dehydrogenase	1.257	-0.257	0.544	0.46066	1.29	-0.29	0.68	0.40942
17918	MOCOS_SOLLC	Molybdenum cofactor sulfurase	--	--	--	--	--	--	--	--
19064	ABAH3_ORYSI	Abscisic acid 8'-hydroxylase 3	--	--	--	--	--	--	--	--
19966	ABAH4_ARATH	Abscisic acid 8'-hydroxylase 4	NA	NA	NA	NA	NA	NA	NA	NA
20144	ABAH3_ARATH	Abscisic acid 8'-hydroxylase 3	--	--	--	--	--	--	--	--
21471	ABAH4_ARATH	Abscisic acid 8'-hydroxylase 4	NA	NA	NA	NA	NA	NA	NA	NA

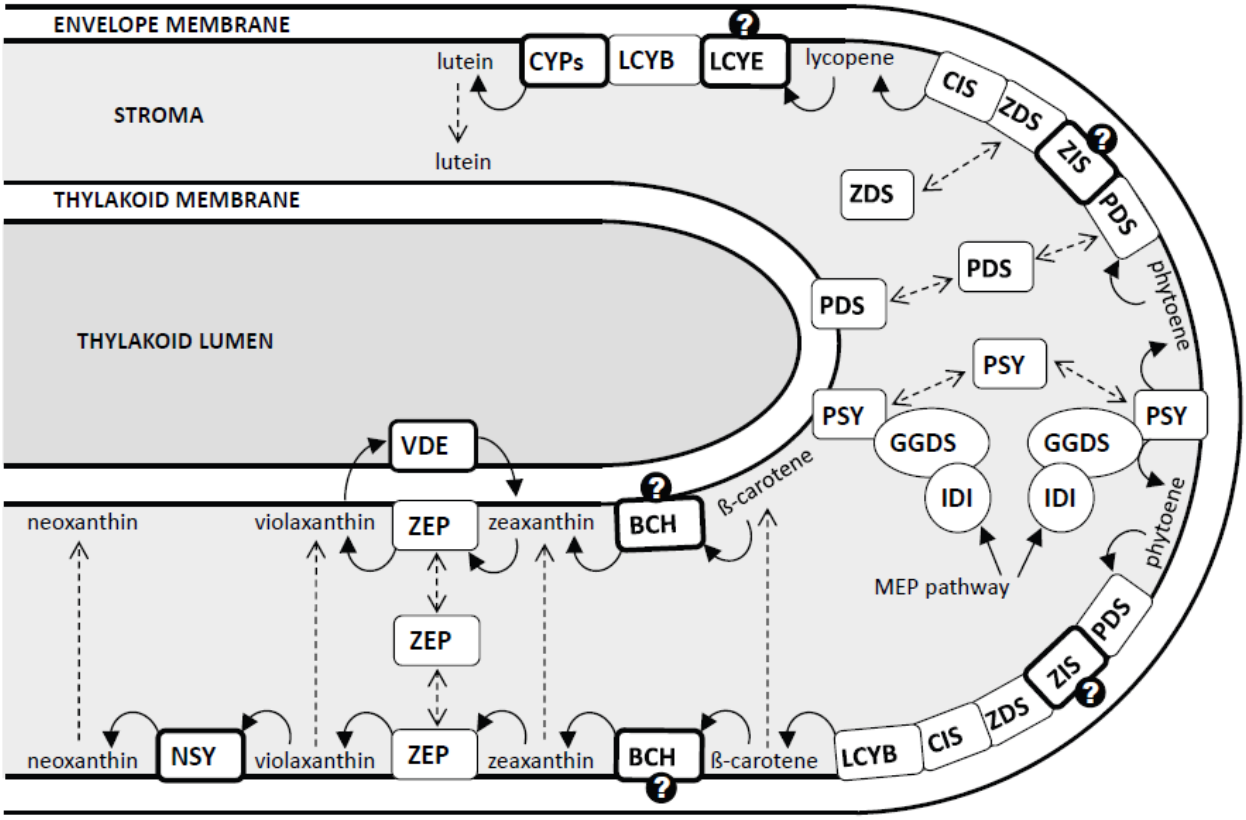


Figure 14. Localization of carotenoid biosynthesis in plastids (Ruiz-Sola and Rodríguez-Concepción 2012). Proteins with membrane attachment domains are indicated by bold boxes. Question marks indicate proteins without experimental evidence of subplastidial localization. BCH=non-heme di-iron carotenoid hydroxylase, CIS=carotenoid isomerase, CYP=cytochrome P450 carotenoid hydroxylase, GGDS=geranylgeranyl diphosphate synthase, IDI=isopentenyl diphosphate isomerase, LCYB=lycopene β-cyclase, LCYE=lycopene ε-cyclase, NSY=neoxanthin synthase, PDS=phytoene desaturase, PSY-phytoene synthase, VDE=violaxanthin de-epoxidase, ZDS=ζ-carotene desaturase, ZEP=zeaxanthin epoxidase, ZIS=15-*cis*-ζ-carotene isomerase.

REFERENCES

Adam Z. & Clarke A.K. (2002) Cutting edge of chloroplast proteolysis. *Trends in Plant Science*, **7**, 451–456.

Aluru M., Xu Y., Guo R., Wang Z., Li S., White W., Wang K., & Rodermel S. (2008) Generation of transgenic maize with enhanced provitamin A content. *Journal of Experimental Botany*, **59**, 3551–3562.

Van der Auwera G.A., Carneiro M.O., Hartl C., Poplin R., del Angel G., Levy-Moonshine A., Jordan T., Shakir K., Roazen D., Thibault J., Banks E., Garimella K. V, Altshuler D., Gabriel S., & Depristo M.A. (2013) From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, **43**, 11.10.1–11.10.33.

Bazargani M.M., Sarhadi E., Bushehri A.-A.S., Matros A., Mock H.-P., Naghavi M.-R., Hajihoseini V., Mardi M., Hajirezaei M.-R., Moradi F., Ehdaie B., & Salekdeh G.H. (2011) A proteomics view on the role of drought-induced senescence and oxidative stress defense in enhanced stem reserves remobilization in wheat. *Journal of Proteomics*, **74**, 1959–1973.

Bittner F., Oreb M., & Mendel R.R. (2001) ABA3 is a molybdenum cofactor sulfurase required for activation of aldehyde oxidase and xanthine dehydrogenase in *Arabidopsis thaliana*. *The Journal of Biological Chemistry*, **276**, 40381–40384.

Blackman B.K., Rasmussen D.A., Strasburg J.L., Raduski A.R., Burke J.M., Knapp S.J., Michaels S.D., & Rieseberg L.H. (2011) Contributions of flowering time genes to sunflower domestication and improvement. *Genetics*, **187**, 271–287.

Bligh W. (1792) *Voyage to the South Sea, Undertaken by Command of His Majesty, for the Purpose of Conveying the Bread-Fruit Tree to the West Indies, in His Majesty's Ship the Bounty, commanded by Lieutenant William Bligh.* George Nicol, London, UK.

Bolger A.M., Lohse M., & Usadel B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, btu170.

Bowman M.J. (2012) Ph.D. dissertation. University of Wisconsin-Madison, Madison, WI, USA.

Calderón-Villalobos L.I.A., Nill C., Marrocco K., Kretsch T., & Schwechheimer C. (2007) The evolutionarily conserved *Arabidopsis thaliana* F-box protein AtFBP7 is required for efficient translation during temperature stress. *Gene*, **392**, 106–116.

Century K., Reuber T.L., & Ratcliffe O.J. (2008) Regulating the regulators: the future prospects for transcription-factor-based agricultural biotechnology products. *Plant physiology*, **147**, 20–29.

Chamovitz D., Pecker I., & Hirschberg J. (1991) The molecular basis of resistance to the herbicide norflurazon. *Plant Molecular Biology*, **16**, 967–974.

Chi W., Ma J., Zhang D., Guo J., Chen F., Lu C., & Zhang L. (2008) The pentratricopeptide repeat protein DELAYED GREENING1 is involved in the regulation of early chloroplast development and chloroplast gene expression in *Arabidopsis*. *Plant physiology*, **147**, 573–584.

Clotault J., Thuillet A., Buiron M., De Mita S., Couderc M., Haussmann B.I., Maria C., & Vigouroux Y. (2012) Evolutionary history of pearl millet (*Pennisetum glaucum* R. Br.) and selection on flowering genes since its domestication. *Molecular Biology and Evolution*, **29**, 1199–1212.

Cockram J., Jones H., Leigh F.J., O'Sullivan D., Powell W., Laurie D. a, & Greenland A.J. (2007) Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. *Journal of Experimental Botany*, **58**, 1231–1244.

Cui S., Huang F., Wang J., Ma X., Cheng Y., & Liu J. (2005) A proteomic analysis of cold stress responses in rice seedlings. *Proteomics*, **5**, 3162–3172.

Cunningham F.X. & Gantt E. (1998) Genes and enzymes of carotenoid biosynthesis in plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, **49**, 557–583.

Dan Ramdath D., Isaacs R.L.C., Teelucksingh S., & Wolever T.M.S. (2004) Glycaemic index of selected staples commonly eaten in the Caribbean and the effects of boiling v. crushing. *British Journal of Nutrition*, **91**, 971–977.

Deely J.J. & Lindley D. V. (1981) Bayes Empirical Bayes. *Journal of the American Statistical Association*, **76**, 833–841.

Demirevska K., Simova-Stoilova L., Vassileva V., Vaseva I., Grigorova B., & Feller U. (2008) Drought-induced leaf protein alterations in sensitive and tolerant wheat varieties. *General and Applied Plant Physiology*, **34**, 79–102.

Demmig-Adams B., Gilmore A.M., & Adams III W. (1996) In vivo functions of carotenoids in higher plants. *The FASEB Journal*, **10**, 403–412.

DePristo M.A., Banks E., Poplin R., Garimella K.V., Maguire J.R., Hartl C., Philippakis A.A., del Angel G., Rivas M.A., Hanna M., McKenna A., Fennell T.J., Kernytsky A.M., Sivachenko A.Y., Cibulskis K., Gabriel S.B., Altshuler D., & Daly M.J. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–498.

Dignan, C., Burlingame, B., Kumar, S., Aalbersberg, W. (2004). The Pacific Islands food composition tables. South Pacific Commission, Noumea (New Caledonia).

Diretto G., Al-Babili S., Tavazza R., Scossa F., Papacchioli V., Migliore M., Beyer P., & Giuliano G. (2010) Transcriptional-metabolic networks in beta-carotene-enriched potato tubers: the long and winding road to the Golden phenotype. *Plant Physiology*, **154**, 899–912.

Doebley J. & Stec A. (1991) Genetic analysis of the morphological differences between maize and teosinte. *Genetics*, **129**, 285–295.

Doebley J.F., Gaut B.S., & Smith B.D. (2006) The molecular genetics of crop domestication. *Cell*, **127**, 1309–1321.

Englberger L., Aalbersberg W., Ravi P., Bonnin E., Marks G.C., Fitzgerald M.H., & Elymore J. (2003a) Further analyses on Micronesian banana, taro, breadfruit and other foods for provitamin A carotenoids and minerals. *Journal of Food Composition and Analysis*, **16**, 219–236.

Englberger L., Lorennij R., Taylor M., Saena V., Aalbersberg W., Dolodolotawake U., Tibon L., Tibon J., & Alfred J. (2014) Carotenoid content and traditional knowledge of breadfruit cultivars of the Republic of the Marshall I. *Journal of Food Composition and Analysis*, **34**, 192–199.

Englberger L., Marks G.C., & Fitzgerald M.H. (2003b) Insights on food and nutrition in the Federated States of Micronesia: a review of the literature. *Public Health Nutrition*, **6**, 5–17.

Englberger L., Schierle J., Marks G.C., & Fitzgerald M.H. (2003c) Micronesian banana, taro, and other foods: newly recognized sources of provitamin A and other carotenoids. *Journal of Food Composition and Analysis*, **16**, 3–19.

Eyre-Walker A. (2002) Changing effective population size and the McDonald-Kreitman test. *Genetics*, **162**, 2017–2024.

Fosberg F. (1960) Introgression in *Artocarpus* (Moraceae) in Micronesia. *Brittonia*, **12**, 101–113.

Fraser P.D., Truesdale M.R., Bird C.R., Schuch W., & Bramley P.M. (1994) Carotenoid biosynthesis during tomato fruit development. *Plant Physiology*, **105**, 405–413.

Galpaz N., Ronen G., Khalfa Z., Zamir D., & Hirschberg J. (2006) A chromoplast-specific carotenoid biosynthesis pathway is revealed by cloning of the tomato white-flower locus. *The Plant Cell*, **18**, 1947–1960.

Galpaz N., Wang Q., Menda N., Zamir D., & Hirschberg J. (2008) Abscisic acid deficiency in the tomato mutant high-pigment 3 leading to increased plastid number and higher fruit lycopene content. *The Plant Journal*, **53**, 717–730.

Geigenberger P. (2011) Regulation of starch biosynthesis in response to a fluctuating environment. *Plant physiology*, **155**, 1566–1577.

Gorovits R. & Czosnek H. (2008) Expression of stress gene networks in tomato lines susceptible and resistant to tomato yellow leaf curl virus in response to abiotic stresses. *Plant Physiology and Biochemistry*, **46**, 482–492.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., & Regev A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644–652.

Grotewold E. (2006) The genetics and biochemistry of floral pigments. *Annual Review of Plant Biology*, **57**, 761–780.

Gu Q., Ferrándiz C., Yanofsky M.F., & Martienssen R. (1998) The FRUITFULL MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development*, **125**, 1509–1517.

Hasan S.M.Z. & Razak A.R. (1992) Parthenocarpy in seedless breadfruit (*Artocarpus incisus*(Thunb.)L.). *Acta Horticulturae*, **321**, 648–652.

Hirschberg J. (2001) Carotenoid biosynthesis in flowering plants. *Current Opinion in Plant Biology*, **4**, 210–218.

Hodgins K.A., Lai Z., Oliveira L.O., Still D.W., Scascitelli M., Barker M.S., Kane N.C., Dempewolf H., Kozik A., Kesseli R.V., Burke J.M., Michelmore R.W., & Rieseberg L.H. (2013) Genomics of compositae crops: reference transcriptome assemblies and evidence of hybridization with wild relatives. *Molecular Ecology Resources*, **14**, 166–177.

Howe G.T., Yu J., Knaus B., Cronn R., Kolpak S., Dolan P., Lorenz W.W., & Dean J.F.D. (2013) A SNP resource for Douglas-fir: de novo transcriptome assembly and SNP detection and validation. *BMC genomics*, **14**, 137.

Howitt C.A. & Pogson B.J. (2006) Carotenoid accumulation and function in seeds and non-green tissues. *Plant, Cell and Environment*, **29**, 435–445.

Hyun T.K., Rim Y., Jang H.-J., Kim C.H., Park J., Kumar R., Lee S., Kim B.C., Bhak J., Nguyen-Quoc B., Kim S.-W., Lee S.Y., & Kim J.-Y. (2012) De novo transcriptome sequencing of *Momordica cochinchinensis* to identify genes involved in the carotenoid biosynthesis. *Plant Molecular Biology*, **79**, 413–427.

Jain M., Nijhawan A., Arora R., Agarwal P., Ray S., Sharma P., Kapoor S., Tyagi A.K., & Khurana J.P. (2007) F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. *Plant physiology*, **143**, 1467–1483.

Jarrett F.M. (1959) Studies in *Artocarpus* and allied genera, III. A revision of *Artocarpus* subgenus *Artocarpus*. *Journal of the Arnold Arboretum*, **40**, 113–155, 327–368.

Jones A.M.P., Murch S., & Ragone D. (2010) Diversity of breadfruit (*Artocarpus altilis*, Moraceae) seasonality: A resource for year-round nutrition. *Economic Botany*, **64**, 340–351.

Jones A.M.P., Ragone D., Aiona K., Lane W.A., & Murch S.J. (2011) Nutritional and morphological diversity of breadfruit (*Artocarpus*, Moraceae): Identification of elite cultivars for food security. *Journal of Food Composition and Analysis*, **24**, 1091–1102.

Jones A.M.P., Baker R., Ragone D., & Murch S.J. (2013a) Identification of pro-vitamin A carotenoid-rich cultivars of breadfruit (*Artocarpus*, Moraceae). *Journal of Food Composition and Analysis*, **31**, 51–61.

Jones A.M.P., Murch S.J., Wiseman J., & Ragone D. (2013b) Morphological diversity in breadfruit (*Artocarpus*, Moraceae): insights into domestication, conservation, and cultivar identification. *Genetic Resources and Crop Evolution*, **60**, 175–192.

Just B.J., Santos C.A.F., Yandell B.S., & Simon P.W. (2009) Major QTL for carrot color are positionally associated with carotenoid biosynthetic genes and interact epistatically in a domesticated x wild carrot cross. *Theoretical and Applied Genetics*, **119**, 1155–1169.

Kane N.A., Danyluk J., Tardif G., Ouellet F., Laliberte J.-F., Limin A.E., Fowler D.B., & Sarhan F. (2005) TaVRT-2 , a member of the St MADS-11 clade of flowering repressors, is regulated by vernalization and photoperiod in Wheat. *Plant Physiology*, **138**, 2354–2363.

Katoh K., Misawa K., Kuma K., & Miyata T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**, 3059–3066.

Kidrič M., Kos J., & Sabotič J. (2014) Proteases and their endogenous inhibitors in the plant response to abiotic stress. *Botanica Serbica*, **38**, 139–158.

Kim J. & Dellapenna D. (2006) Defining the primary route for lutein synthesis in plants: The role of Arabidopsis carotenoid β -ring hydroxylase CYP97A3. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 3474–3479.

Kim M.Y., Van K., Kang Y.J., Kim K.H., & Lee S.-H. (2012) Tracing soybean domestication history: From nucleotide to genome. *Breeding Science*, **61**, 445–452.

Kirch P.V. (2000) *On the Road of the Winds: An archaeological History of the Pacific Islands Before European Contact.* University of California Press, Berkley, CA, USA.

Krinsky N.I. (1989) Antioxidant functions of carotenoids. *Free Radical Biology & Medicine*, **7**, 617–635.

Lechner E., Achard P., Vansiri A., Potuschak T., & Genschik P. (2006) F-box proteins everywhere. *Current Opinion in Plant Biology*, **9**, 631–638.

Lechner M., Findeiβ S., Steiner L., Marz M., Stadler P.F., & Prohaska S.J. (2011) Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics*, **12**, 124.

Lee J. & Lee I. (2010) Regulation and function of SOC1, a flowering pathway integrator. *Journal of Experimental Botany*, **61**, 2247–2254.

Li H. & Durbin R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, **26**, 589–595.

Li H., Liang W., Jia R., Yin C., Zong J., Kong H., & Zhang D. (2010) The AGL6-like gene OsMADS6 regulates floral organ and meristem identities in rice. *Cell Research*, **20**, 299–313.

Librado P. & Rozas J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Lincoln N. & Ladefoged T. (2014) Agroecology of pre-contact Hawaiian dryland farming: the spatial extent, yield and social impact of Hawaiian breadfruit groves in Kona, Hawai'i. *Journal of Archaeological Science*, **49**, 192–202.

Liu Y., Roof S., Ye Z., Barry C., van Tuinen A., Vrebalov J., Bowler C., & Giovannoni J. (2004) Manipulation of light signal transduction as a means of modifying fruit nutritional quality in tomato. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 9897–9902.

Liu T., Tang S., Zhu S., Tang Q., & Zheng X. (2014) Transcriptome comparison reveals the patterns of selection in domesticated and wild ramie (*Boehmeria nivea* L. Gaud). *Plant Molecular Biology*, **86**, 85–92.

Liu Z., Ma L., Nan Z., & Wang Y. (2013) Comparative transcriptional profiling provides insights into the evolution and development of the zygomorphic flower of *Vicia sativa* (Papilionoideae). *PloS one*, **8**, e57338.

Lu S. & Li L. (2008) Carotenoid metabolism: biosynthesis, regulation, and beyond. *Journal of Integrative Plant Biology*, **50**, 778–785.

Lu Y., Li Y., Zhang J., Xiao Y., Yue Y., Duan L., Zhang M., & Li Z. (2013) Overexpression of *Arabidopsis* molybdenum cofactor sulfurase gene confers drought tolerance in maize (*Zea mays* L.). *PloS one*, **8**, e52126.

Lucas M.P. & Ragone D. (2012) Will breadfruit solve the world hunger crisis? *ArcNews*, Summer 2011.

Maere S., Heymans K., & Kuiper M. (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, **21**, 3448–3449.

Mandel M.A. & Yanofsky M.F. (1998) The *Arabidopsis* AGL9 MADS box gene is expressed in young flower primordia. *Sexual Plant Reproduction*, **11**, 22–28.

Mariac C., Jehin L., Saïdou A.-A., Thuillet A.-C., Couderc M., Sire P., Jugdé H., Adam H., Bezançon G., Pham J.-L., & Vigouroux Y. (2011) Genetic basis of pearl millet adaptation along an environmental gradient investigated by a combination of genome scan and association mapping. *Molecular Ecology*, **20**, 80–91.

Marte R. (1986) Nontraditional fruit crops in the Windward Islands. *Proceedings of the Interamerican Society for Tropical Horticulture*, **30**, 15–24.

Mayes S., Massawe F.J., Alderson P.G., Roberts J.A., Azam-Ali S.N., & Hermann M. (2012) The potential for underutilized crops to improve security of food production. *Journal of Experimental Botany*, **63**, 1075–1079.

McCord P., Zhang L., & Brown C. (2012) The incidence and effect on total tuber carotenoids of a recessive zeaxanthin epoxidase allele (Zep1) in yellow-fleshed potatoes. *American Journal of Potato Research*, **89**, 262–268.

McDonald J.H. & Kreitman M. (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, **351**, 652–654.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., & DePristo M.A. (2010) The genome analysis toolkit: A mapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

McLaren D. & Frigg M. (2001) *Sight and life manual on vitamin A deficiency disorders (VADD).* 2nd Ed. Basel: Task Force Sight and Life, Basel, Switzerland.

Meyer R.S. & Purugganan M.D. (2013) Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics*, **14**, 840–852.

Miller A.J. & Gross B.L. (2011) From forest to field: perennial fruit crop domestication. *American Journal of Botany*, **98**, 1389–1414.

Nambara E. & Marion-Poll A. (2005) Abscisic acid biosynthesis and catabolism. *Annual Review of Plant Biology*, **56**, 165–185.

O'Neil S.T. & Emrich S.J. (2013) Assessing De Novo transcriptome assembly metrics for consistency and utility. *BMC genomics*, **14**, 465.

Olsen K.M. & Wendel J.F. (2013) A bountiful harvest: genomic insights into crop domestication phenotypes. *Annual Review of Plant Biology*, **64**, 47–70.

Olson J.A. (1989) Provitamin A function of carotenoids : The conversion of β-Carotene into vitamin A. *The Journal of Nutrition*, **119**, 105–108.

Ott M., Zola J., Stamatakis A., & Aluru S. (2007) Large-scale maximum likelihood-based phylogenetic analysis on the IBM BlueGene/L. *Proceedings of the 2007 ACM/IEEE conference on Supercomputing*.

Owens B.F., Lipka A., Magallanes-Lundback M., Tiede T., Diepenbrock C., Kandianis C., Kim E., Cepela J., Mateos-Hernandez M., Buell C., Buckler E., DellaPenna D., Gore M., & Rochefors T. (2014) A foundation for provitamin A biofortification of maize: genome-wide association and genomic prediction models of carotenoid levels. *Genetics*, 114.169979.

Parchman T.L., Geist K.S., Grahnen J.A., Benkman C.W., & Buerkle C.A. (2010) Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics*, **11**, 180.

Park H.-Y., Seok H.-Y., Park B.-K., Kim S.-H., Goh C.-H., Lee B., Lee C.-H., & Moon Y.-H. (2008) Overexpression of Arabidopsis ZEP enhances tolerance to osmotic stress. *Biochemical and Biophysical Research Communications*, **375**, 80–85.

Pecker I., Chamovitz D., Linden H., Sandmann G., & Hirschberg J. (1992) A single polypeptide catalyzing the conversion of phytoene to ζ-carotene is transcriptionally regulated during tomato fruit ripening. *Proceedings of the National Academy of Sciences*, **89**, 4962–4966.

Peng J., Ronin Y., Fahima T., Roder M.S., Li Y., Nevo E., & Korol A. (2003) Domestication quantitative trait loci in *Triticum dicoccoides*, the progenitor of wheat. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 2489–2924.

Pogson B., Mcdonald K.A., Truong M., Britton G., & Dellapenna D. (1996) *Arabidopsis* carotenoid mutants demonstrate that lutein is not essential for photosynthesis in higher plants. *The Plant Cell*, **8**, 1627–1639.

Pollock N.J. (1984) Breadfruit Fermentation Practices in Oceania. *Journal de la Société des océanistes*, **40**, 151–164.

Primmer C.R., Papakostas S., Leder E.H., Davis M.J., & Ragan M.A. (2013) Annotated genes and nonannotated genomes: cross-species use of Gene Ontology in ecology and evolution research. *Molecular Ecology*, **22**, 3216–3241.

Purseglove J.W. (1968) *Artocarpus altilis*. In J.*W (ed.),Tropical Crops(2), Dicotyledons*, Longman, London, UK, 377–384.

Purugganan M.D., Boyles A.L., & Suddith J.I. (2000) Variation and selection at the CAULIFLOWER floral homeotic gene accompanying the evolution of domesticated *Brassica oleracea*. *Genetics*, **155**, 855–862.

Purugganan M.D. & Fuller D.Q. (2009) The nature of selection during plant domestication. *Nature*, **457**, 843–848.

R Core Team (2014) R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.

Ragone D. (1988) *Breadfruit varieties in the pacific atolls.* Integrated Atoll Development Project, UNDP, Suva, Fiji.

Ragone D. (1991) Collection, establishment, and evaluation of a germplasm collection of pacific island breadfruit. Ph.D. dissertation, University of Hawaii, Honolulu, HI, USA.

Ragone D. (1997) *Breadfruit, Artocarpus altilis (Parkinson) Fosberg, Promoting the Conservation and Use of Underutilized and Neglected Crops,* 10. International Plant Genetic Resources Institute, Rome, Italy.

Ragone D. (2001) Chromosome numbers and pollen stainability of three species of Pacific Island breadfruit (*Artocarpus*, Moraceae). *American Journal of Botany*, **88**, 693–696.

Ragone D. (2007) Breadfruit: Diversity, conservation and potential. *Acta Horticulturae*, **757**, 19–30.

Ragone D. (2008) Regeneration guidelines: Breadfruit. In Dulloo M.E., Thormann I., Jorge M.A. and Hanson J., editors. Crop Specific Regeneration Guidelines[CD-ROM]. CGIAR System-wide Genetic Resource Programme, Rome, Italy, 7.

Roberts-Nkrumah L.B. (2007) An overview of breadfruit (*Artocarpus altilis*) in the Caribbean. *Proceedings of the 1st International Symposium on Breadfruit Research and Development*, **757**, 51–60.

Roberts-Nkrumah L.B. & Legall G. (2013) Breadfruit (Artocarpus altilis, Moraceae) and chataigne (*A. camansi*) for food security and income generation : The case of Trinidad and Tobago. *Economic Botany*, **67**, 324–334.

Rollins J.A., Habte E., Templer S.E., Colby T., Schmidt J., & von Korff M. (2013) Leaf proteome alterations in the context of physiological and morphological responses to drought and heat stress in barley (Hordeum vulgare L.). *Journal of Experimental Botany*, **64**, 3201–3212.

Römer S., Lübeck J., Kauder F., Steiger S., Adomat C., & Sandmann G. (2002) Genetic engineering of a zeaxanthin-rich potato by antisense inactivation and co-suppression of carotenoid epoxidation. *Metabolic Engineering*, **4**, 263–272.

Ross-Ibarra J., Morrell P.L., & Gaut B.S. (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 8641–8648.

Ruiz-Sola M.Á. & Rodríguez-Concepción M. (2012) Carotenoid biosynthesis in *Arabidopsis*: a colorful pathway. *The Arabidopsis book / American Society of Plant Biologists*, **10**, e0158.

Sagi M., Scazzocchio C., & Fluhr R. (2002) The absence of molybdenum cofactor sulfuration is the primary cause of the flacca phenotype in tomato plants. *The Plant Journal*, **31**, 305–317.

Schaub P., Yu Q., Gemmecker S., Poussin-Courmontagne P., Mailliot J., McEwen A.G., Ghisla S., Al-Babili S., Cavarelli J., & Beyer P. (2012) On the structure and function of the phytoene desaturase CRTI from *Pantoea ananatis*, a membrane-peripheral and FAD-dependent oxidase/isomerase. *PloS one*, **7**, e39550.

Schwartz S.H., Tan B.C., Gage D.A., Zeevaart J.A.D., & McCarty D. (1997) Specific oxidative cleavage of carotenoids by VP14 of maize. *Science*, **276**, 1872–1874.

Shore P. & Sharrocks A. (1995) The MADS-box family of transcription factors. *European Journal of Biochemistry*, **229**, 1–13.

Sjögren L.L.E., Stanne T.M., Zheng B., Sutinen S., & Clarke A.K. (2006) Structural and functional insights into the chloroplast ATP-dependent Clp protease in *Arabidopsis*. *The Plant Cell*, **18**, 2635–2649.

Smaczniak C., Immink R.G.H., Angenent G.C., & Kaufmann K. (2012) Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development*, **139**, 3081–3098.

Smith S.M. (2014) Q&A: What are strigolactones and why are they important to plants and soil microbes? *BMC biology*, **12**, 19.

Stamatakis A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Stange C. & Flores C. (2012) Carotenoids and photosynthesis-regulation of carotenoid biosynthesis by photoreceptors. *Advances in Photosynthesis-Fundamental Aspects, Dr Mohammad Najafpour (Ed.), InTech*, .

Strickler S.R., Bombarely A., & Mueller L.A. (2012) Designing a transcriptome next-generation sequencing project for a nonmodel plant species. *American Journal of Botany*, **99**, 257–266.

Sung S.-K., Yu G.-H., Nam J., Jeong D.-H., & An G. (2000) Developmentally regulated expression of two MADS-box genes, *MdMADS3* and *MdMADS4*, in the morphogenesis of flower buds and fruits in apple. *Planta*, **210**, 519–528.

Suyama M., Torrents D., & Bork P. (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, **34**, W609–612.

Tacken E.J., Ireland H.S., Wang Y.-Y., Putterill J., & Schaffer R.J. (2012) Apple EIN3 BINDING F-box 1 inhibits the activity of three apple EIN3-like transcription factors. *AoB Plants*, **2012**, pls034.

Tevini M. & Steinmüller D. (1985) Composition and function of plastoglobuli. *Planta*, **163**, 91–96.

Toledo-Ortiz G., Huq E., & Rodríguez-Concepción M. (2010) Direct regulation of phytoene synthase gene expression and carotenoid biosynthesis by phytochrome-interacting factors. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 11626–11631.

Vallabhaneni R. & Wurtzel E.T. (2009) Timing and biosynthetic potential for carotenoid accumulation in genetically diverse germplasm of maize. *Plant physiology*, **150**, 562–572.

Vaughan D.A., Balázs E., & Heslop-Harrison J.S. (2007) From crop domestication to super-domestication. *Annals of Botany*, **100**, 893–901.

Vrebalov J., Ruezinsky D., Padmanabhan V., White R., Medrano D., Drake R., Schuch W., & Giovannoni J. (2002) A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (rin) locus. *Science*, **296**, 343–346.

Walker-Simmons M., Kudrna D.A., & Warner R.L. (1989) Reduced accumulation of ABA during water stress in molybdenum cofactor mutant of barley. *Plant Physiology*, **90**, 728–733.

Wang W., Feng B., Xiao J., Xia Z., Zhou X., Li P., Zhang W., Wang Y., Møller B.L., Zhang P., Luo M.-C., Xiao G., Liu J., Yang J., Chen S., Rabinowicz P.D., Chen X., Zhang H.-B., Ceballos H., Lou Q., Zou M., Carvalho L.J.C.B., Zeng C., Xia J., Sun S., Fu Y., Wang H., Lu C., Ruan M., Zhou S., Wu Z., Liu H., Kannangara R.M., Jørgensen K., Neale R.L., Bonde M., Heinz N., Zhu W., Wang S., Zhang Y., Pan K., Wen M., Ma P.-A., Li Z., Hu M., Liao W., Hu W., Zhang S., Pei J., Guo A., Guo J., Zhang J., Zhang Z., Ye J., Ou W., Ma Y., Liu X., Tallon L.J., Galens K., Ott S., Huang J., Xue J., An F., Yao Q., Lu X., Fregene M., López-Lavalle L.A.B., Wu J., You F.M., Chen M., Hu S., Wu G., Zhong S., Ling P., Chen Y., Wang Q., Liu G., Liu B., Li K., & Peng M. (2014) Cassava genome from a wild ancestor to cultivated varieties. *Nature Communications*, **5**, 5110.

Wang X., Kong H., & Ma H. (2009) F-box proteins regulate ethylene signaling and more. *Genes & Development*, **23**, 391–396.

WHO (2011) *Guideline: vitamin A supplementation in infants and children 6-59 months of age.* Geneva, Switzerland, World Health Organization.

Witherup C., Ragone D., Wiesner-Hanks T., Irish B., Scheffler B., Simpson S., Zee F., Zuberi M.I., & Zerega N.J.C. (2013) Development of microsatellite loci in *Artocarpus altilis* (Moraceae) and cross-amplification in congeneric species. *Applications in Plant Sciences*, **1**, 1200423.

Wright S.I., Bi I.V., Schroeder S.G., Yamasaki M., Doebley J.F., McMullen M.D., & Gaut B.S. (2005) The effects of artificial selection on the maize genome. *Science*, **308**, 1310–1314.

Wu R.-M., Walton E.F., Richardson A.C., Wood M., Hellens R.P., & Varkonyi-Gasic E. (2012) Conservation and divergence of four kiwifruit SVP-like MADS-box genes suggest distinct roles in kiwifruit bud dormancy and flowering. *Journal of Experimental Botany*, **63**, 797–807.

Xie X., Yoneyama K., & Yoneyama K. (2010) The strigolactone story. *Annual Review of Phytopathology*, **48**, 93–117.

Xie Q., Hu Z., Zhu Z., Dong T., Zhao Z., Cui B., & Chen G. (2014) Overexpression of a novel MADS-box gene SlFYFL delays senescence, fruit ripening and abscission in tomato. *Scientific Reports*, **4**, 4367.

Xing X., Koch A.M., Jones A.M.P., Ragone D., Murch S., & Hart M.M. (2012) Mutualism breakdown in breadfruit domestication. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 1122–1130.

Xiong L., Ishitani M., Lee H., & Zhu J. (2001) The *Arabidopsis* LOS5/ABA3 locus encodes a molybdenum cofactor sulfurase and modulates cold stress – and osmotic stress – responsive gene expression. *The Plant Cell*, **13**, 2063–2083.

Yan Y.-S., Chen X.-Y., Yang K., Sun Z.-X., Fu Y.-P., Zhang Y.-M., & Fang R.-X. (2011) Overexpression of an F-box protein gene reduces abiotic stress tolerance and promotes root growth in rice. *Molecular Plant*, **4**, 190–197.

Yang Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Yang Z. & Bielawski J.P. (2000) Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, **15**, 496–503.

Yang Z. & dos Reis M. (2011) Statistical properties of the branch-site test of positive selection. *Molecular Biology and Evolution*, **28**, 1217–1228.

Yang Z., Wong W.S.W., & Nielsen R. (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, **22**, 1107–1118.

Ye J., Fang L., Zheng H., Zhang Y., Chen J., Zhang Z., Wang J., Li S., Li R., Bolund L., & Wang J. (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research*, **34**, W293–297.

Zerega N.J.C., Ragone D., & Motley T. (2004) Complex origins of breadfruit (*Artocarpus altilis*, Moraceae): implications for human migrations in Oceania. *American Journal of Botany*, **91**, 760–766.

Zerega N.J.C., Ragone D., & Motley T. (2005) Systematics and species limits of breadfruit (*Artocarpus*, Moraceae). *Systematic Botany*, **30**, 603–615.

Zerega, N. J. C., Ragone, D., & Motley, T. J. (2006). Breadfruit origins, diversity, and human-facilitated distribution. *Darwin's Harvest: New Approaches to Origins, Evolution, and Conservation of Crop Plants. Columbia University Press, New York*, 213-238.

Zerega, N.J.C. (2014) Origins and development of breadfruit. In Encyclopedia of Global Archaeology. Smith C, editor. New York, USA: Springer; 2014, 1006-1009.

Zerega N.J.C., Wiesner-Hanks T., Ragone D., Irish B., Scheffler B., Simpson S., & Zee F. (in press) Diversity in the breadfruit complex (*Artocarpus*, Moraceae): Genetic characterization of critical germplasm. *Tree Genetics and Genomes*.

Zhang J., Nielsen R., & Yang Z. (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, **22**, 2472–2479.

Zhao Q., Weber A., McMullen M., Guill K., & Doebley J. (2011) MADS-box genes of maize: frequent targets of selection during domestication. *Genetics Research (Cambridge)*, **93**, 65–75.