Samoa Cocoa Industry Development Initiative - final report - 2020

Prepared by V. Vaughan Symonds and M. Cristina C. Heredia (Massey University),
in collaboration with Seeseei Molimau-Samasoni (Scientific Research Organisation of Samoa)

Executive summary

To assist with the revitalization of the Samoa cocoa (*Theobroma cacao* L.) breeding program, a set of 200 cocoa plants from four geographical locations in Samoa were sampled to investigate their genetic variation and ancestry. Fifteen microsatellite markers were utilised to evaluate the genetic diversity of the cocoa individuals within and across the collection sites. With an average of seven alleles per locus, considerable variation was observed across all marker loci. Genetic clustering analyses revealed three predominant genetic groups, where 83 of the sampled cocoa individuals retain at least 90% of just one of the three ancestral genoptyes and the remaining individuals appear to be various mixtures of the three genetic groups. This finding was essentially supported by all analytical methods employed. With the inclusion of control genotypes from two classic cocoa varieties, it was found that one of the genetic groups clusters with Amelonado (one of several varieties classically referred to as 'Forastero') and another group of individuals cluster with a representative of the Criollo variety. The third genetic group, represented by relatively few individuals is distinct from the other two groups and may represent hybrids between the other two groups; if this is the case, then these individuals would logically fit the description for the Trinitario variety. While the identities inferred from the genetic data do not match the on-farm variety identifications (not analysed here specifically), the rich history of introductions of cocoa varieties to Samoa, followed by outcrossing and natural regeneration is expected to lead to recombination of phenotypes, thus

making identification based on morphology alone a difficult endeavor. Although DNA markers also undergo recombination, analytical methods appear to be sufficiently sophisticated to reconstruct ancestral genotypes that roughly match traditional cocoa variety genotypes. Initially, on-farm managers will likely be most interested in specific traits known to produce high quality cocoa; however, a cocoa breeding and horticulture/propagation program that considers both traits of interest and the genetic composition of individuals, relative to traditional variety performance standards (quality, yield, pest resistance) is likely to enjoy greater long-term success.

Review of Related Literature

I. The history of cocoa and its introduction to Samoa

Theobroma cacao L. has a long history of cultivation and distribution. Its centre of origin is known to be the north Amazon basin, with the wild species originating in the area from Mexico to Peru (Urquhart, 1961). The earliest known cultivation was practiced by the Maya peoples of South and Central America. However, the succeeding centuries saw worldwide spread of cocoa's popularity. In the beginning of the 15th century, which the Europeans called the "Age of discovery", cocoa was recognised as a valuable commodity. Not only was it recorded to have been used as money in Nicaragua, but it was also regarded as a luxury beverage in many European countries (Afoakwa, 2014; Wood, 1985). The spread of cocoa from the Amazon was brought about by several Western conquistadors, who started colonising other countries for economic and political gain. It is believed that the initial spread of cocoa was through the movement of early humans from the Amazon to Central America, and it eventually spread both northward and southward (Shultes, 1958). The development of the cocoa plantation industry was led by the Spaniards and was eventually followed by the Dutch, French, British and Germans (Urquhart, 1961). Further movement of cocoa cultivation to Asia and the Pacific rim occurred in the sixteenth and seventeenth centuries, where the Criollo cocoa was transported across the Philippines, Sulawesi and Java by the Dutch and the Spanish conquistadors (Urquhart, 1961; Wood, 1985; Young, 1994). In the early 1800s, the cultivation of cocoa had expanded from Central America to South America and to several islands in the Caribbean and some areas in the East Indies (Young, 1994). It is believed that cocoa was also introduced to Sri Lanka and India after it was brought to the East Indies (Wood, 1985). It was in the nineteenth century and early twentieth century that Germans started to establish colonies in the Pacific Islands. This opened opportunities for Germany to pioneer the cocoa industry in Samoa (Stolberg, 2013; Urquhart, 1961).

Samoan cocoa is a product of a long history of development, cessation, re-introduction and transplantation of cocoa plant materials. The Criollo cocoa, which originated from Venezuela, was the first introduced cocoa variety in Samoa (Eden & Edwards, 1952; Urquhart, 1952). Later introductions of cocoa were sourced from Sri Lanka and Indonesia (Urquhart, 1952). These Criollo varieties were introduced by a German company in 1883, as an effort to establish

plantation holdings for tropical fruits (Droessler, 2017; Slade, 1984). However, when the cocoa industry in Samoa was still in its initial phase of development, the infestation of canker caused by the oomycete, *Phytophtora palmivora,* resulted in the abortion of cocoa cultivation and production (Eden & Edwards, 1952). It was not until 1898 that the more resistant Forastero (Amelonado) variety from Sri Lanka was introduced to Samoa, as a replacement for the susceptible Criollo (Eden & Edwards, 1952; Slade, 1984; Urquhart, 1952). It was also during the same period in the 1880's when the Trinitario cocoa was transplanted in Samoa, after it was re-planted (after the first failed attempt in 1934) in Sri Lanka in 1880 (Pohlan & Pérez, 2010). In addition, due to mixed plantings of Amelonado and Criollo, natural hybridisation occurred, resulting in increased numbers of the hybrid-origin Trinitario (was termed Samoa Trinitario) (Slade, 1984). Currently, the predominant cultivated cocoa in Samoa are the open-pollinated Trinitario (now termed as Koko Samoa) which are planted with Amelonado (Bourke, 1992; Dillon, 2014). The Amelonado cocoas, however, are thought to comprise only about 12% of the total plantings (Bourke, 1992). The history of introducing new planting materials triggered further mixture of the varieties across Samoa (Figure 1).

The Samoa cocoa industry has also been largely impacted by natural disasters, specifically the cyclones Ofa and Val. Both events had significant effects on large plantations of cocoa, with estimated tree loss of 10-50% caused by cyclone Ofa alone (Fleming & Blowes, 2003). Since those disasters, efforts have been in the works to revive the Samoan cocoa industry through new rehabilitation and breeding programs (Cocoa Samoa Ltd. (CSL), 2015; Samoa Ministry of Agriculture Forests Fisheries and Meteorology, 2004). While it is important to increase cocoa production in the country, creating new stocks that can withstand different abiotic and biotic stresses and combine the unique quality of Samoan cocoa is a high priority. However, while doing so, it is critical to distinguish the taxonomy and morphology of cocoa first — identifying the morphological and physiological results of years of evolution and development.

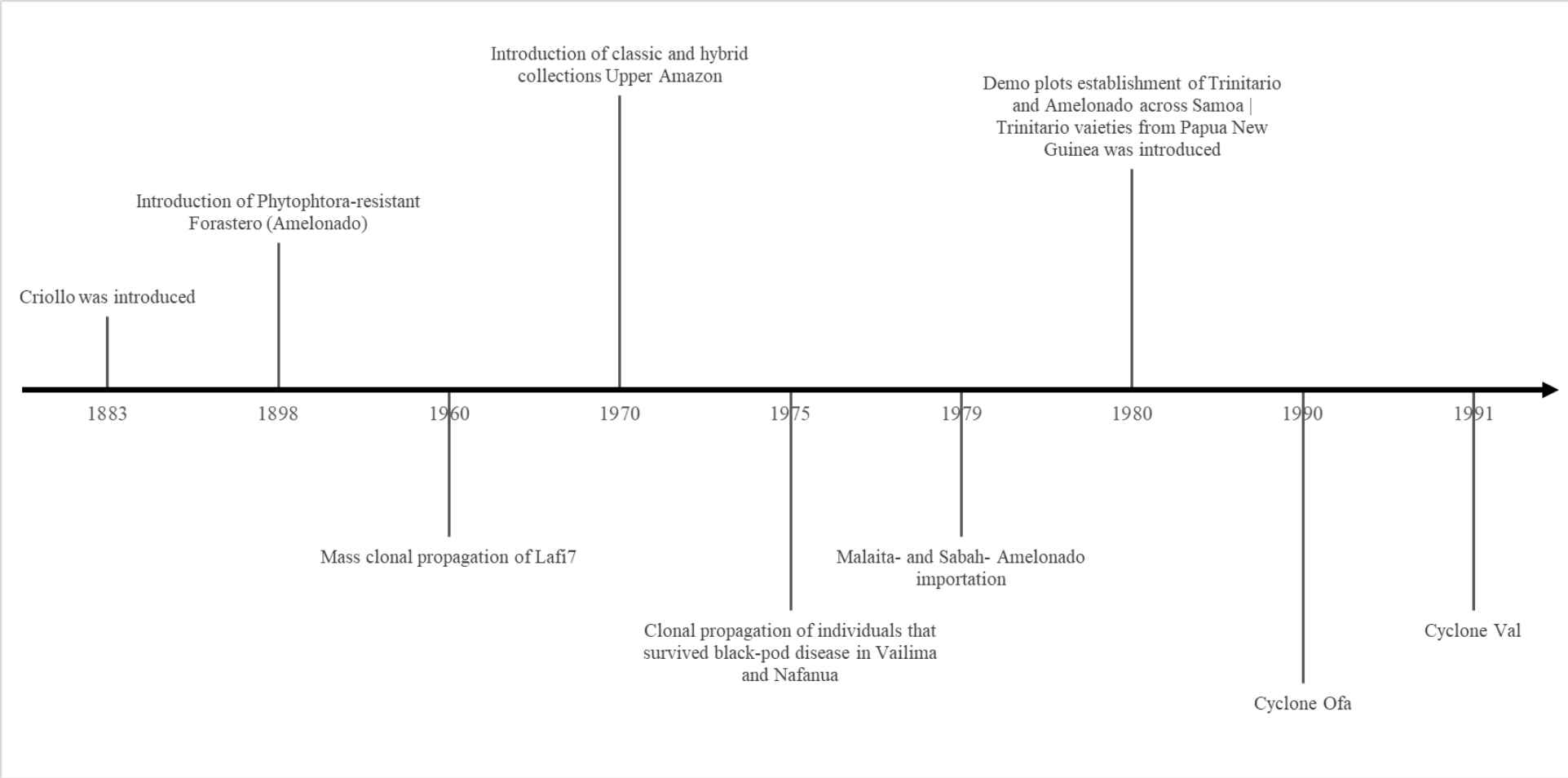


Figure 1. Timeline of a selection of the major events affecting the Samoan cocoa diversity.

II. The varieties of cocoa

Theobroma cacao, like many crops, is genetically diverse due to both natural diversity and many generations of breeding. The genus is classified under the family Sterculiaceae, together with the genera *Herrania, Guazuma* and *Cola* (Wood, 1985). *Theobroma cacao* L. is the largest known source of cocoa; however, two other species are also planted in Central America as another source of cocoa: *Theobroma bicolor* Humb et Bonpl. and *Theobroma grandiflorum* (Spreng) K. Schum (Rehm & Espig, 1991). The earliest classifications and nomenclature of cocoa varieties were published by Morris (1882) (as cited in Bhattacharjee, 2017). Two classes were identified with the following varieties: Class I: Cacao Criollo (Red); Class II: Cacao Forestero. A decade later, Hart (as cited in Bhattacharjee, 2017) published another classification of cocoa varieties under three classes: Class I: Criollo; Class II: Forastero; Class III: Calabacillo.

The two main founding groups of the traditional classification of cocoa were Criollo and Forastero with two subgroups (Amarillo and Colorado). The Calabacillo was then categorised as a separate class. However, the current classification of cocoa removed the class Calabacillo and replaced it with Trinitario as the third group. In recent years another group (Nacional) was included, which subsequently comprised the four principal varieties of *Theobroma cacao*: Forastero; Criollo; Trinitario; and Nacional (Afoakwa, 2014); however, there remain several different classification systems in use, depending on the criteria utilized.

Cocoa may have originated in a single area in the Amazon basin, but it has clearly diversified. Here we consider the four groups system and their distinguishing properties. The earliest recorded variety, the Criollo ('*native*') cocoa has long been established in Mexico, Central America and Columbia, and is characterised by red or yellow pods when ripe (Urquhart, 1961). It is classified into two major groups: the Central American Criollos and the South American Criollos (Toxopeus, 1985). In general, the fruit of Criollo is elongated and pointy with an uneven, warty and deeply furrowed surface (Rehm & Espig, 1991). This variety creates a more flavourful cocoa with a stronger aroma compared to Forastero (Rehm & Espig, 1991; Willson, 1999). However, due to constant selection, the population of the pure Criollo variety has become rare (Urquhart, 1961). Forastero ('*foreign*') cocoa, on the other hand, is a native of the Amazon region, but is widely cultivated in West Africa and Southeast Asia (Afoakwa, 2014).

Unlike Criollo, the fruit of Forastero are oval with a green-yellow colour and a smooth surface (Rehm & Espig, 1991). The pods are thick and have a woody layer, while the seeds are almost flat with a dark purple to black cotyledon (Urquhart, 1961). It is generally divided into two groups, the Upper Amazonian and the Lower Amazonian (also called Amelonado) (Young, 1994). In Ghana, Amelonado cultivars comprise 13.3% of the cultivated cocoa, while the remaining 34.4% and 52.3% are comprised of the Amazonia cultivars and hybrid cultivars respectively (Afoakwa, 2014). The first hybrid progenies of Criollo and Forastero, grown in Trinidad, resulted in the Trinitario ('cocoa of Trinidad') variety, which is hardier and has a greater yield than Criollo (Willson, 1999). However, its flavour is still inferior to that of Criollo (Urquhart, 1961). In addition, because of the high variation between its parents, the characteristics of the Trinitario pods and beans were found to be highly variable (Toxopeus, 1985). The later generations were also found to have lower vigour compared to the initial crosses from Criollo and Forastero. Generally, Criollo and Trinitario varieties are called the fine flavour cocoa while the Forastero variety is known as the bulk cocoa (van der Kooji, 2013). Nacional, a recently recognised variety, has also been considered as a fine flavour cocoa (Solorzano et al., 2012). The Nacional cocoa variety grows in Ecuador but only comprises ~5% of the world cocoa production, because of its limited cultivation (Afoakwa, 2014). The pure varieties of Nacional cocoa is considered rare, as the result of the introduction of foreign germplasms into Ecuador (Solorzano et al., 2012).

III. Major abiotic stresses affecting cocoa production in Samoa

The impact of climate change has led to detrimental effects on the production of cocoa. It is estimated that by the year 2050, some areas (Lagunes and Sud-Comoe) in the top-producing country Cote d'Ivoire will become un-suitable for cocoa farming due to climate change (Läderach, 2013). Likewise in Ghana and in the tropical states (Ekiti States) of Nigeria, cocoa farmers were found to be more vulnerable to decline in productivity due to climate change (Codjoe, 2013; Oyedokun & Oyelana, 2016). Like in many areas around the world, Samoa is also significantly affected by the climate change phenomenon. It is reported that the land- and sea-surface temperature in Samoa will continue to increase resulting in a projected increase of intense and frequent days of extreme heat and rainfall (Australian Bureau of Meteorology

and CSIRO, 2011). In addition, the same report projected a low incidence of drought and tropical cyclone. Even in the major cocoa-growing areas in West Africa, the projected decrease in production by 2050 is anticipated to mainly be caused by the increase in temperature and not by drought (Laderach et al., 2011; Schroth et al, 2016). With its estimated 2,792 ha of cocoa plantations (Rasmussen & McGoldrick, 1992), it is therefore crucial that the future cocoa breeding and management programs in Samoa be focused on heat and waterlogging stresses.

Cocoa's vegetative growth, reproductive growth, and fruit development are highly dependent on temperature (Alvim, 1977; de Almeida, 2007; Sale, 1969). The thermal damage brought by heat stress largely target the photosynthetic system of the plant (Allakhverdiev et al., 2008). This is induced by the oxidative stress that eventually impairs the plant's growth (De Silva et al., 2017). The reproductive stage of the plant is also heavily damaged due to reduced pollen viability, hampered fertilisation and fruit formation (Hafield et al., 2008). Hence it was suggested that management and strategies on increasing shade trees must be employed (Schroth et al., 2016). One strategy employed in Cote d'Ivoire is the integration of tree diversity around cocoa plantations to increase protection against heat stress (Smith Dumont, Gnahoua, Ohouo, Sinclair, & Vaast, 2014). Application of nutrients has also been shown to mitigate heat stress on crops. Macronutrients such as nitrogen (N), potassium (K), calcium (Ca), and magnesium (Mg) increase the antioxidants in crops which subsequently help against the reactive oxygen species (ROS) (Waraich, Ahmad, Halim, & Aziz, 2012). In addition, micronutrients such as boron (B), manganese (Mn) and selenium (Se) activates the physiological, biochemical and metabolic activities in crops, aiming towards heat stress alleviation.

On the other hand, extreme events such as cyclones, flash floods and heavy rainfall also have been a significant threat to Samoa's agriculture (Sesaga, 2014). In fact, in the early 1990's the cocoa export industry of the country ceased because of the devastating cyclones Ofa and Val (Dillon et al., 2014). Water stress such as flooding is detrimental to cocoa as it decreases leaf area and subsequently reduces stomatal conductance and photosynthetic activity of the plant (de Almeida, 2007). The occurrence of flooding in cocoa growing areas results in hypoxic soil conditions, wherein oxygen is being deprived (Sena et al., 1986). This is then followed by leaf epinasty, decay of roots and the development of adventitious roots and hypertrophied

lenticels on the waterlogged stem (Sena et al., 1986). In a study on flood-susceptible cocoa clones, damaged photosynthetic complex, oxidative stress, lead chlorosis and decreased levels of carbohydrate in the roots were observed (Bertolde et al., 2012). Despite the negative effects of waterlogging stress on crops, there is no continuous development of conserved strategies to mitigate it (Bailey-Serres et al., 2012).

Conclusively, abiotic stresses have intermittently destroyed large percentages of cocoa plantations in Samoa. However, other challenges caused by the effects of pests and diseases have been causing bigger problems as the consequences can be permanent if no complete eradication is accomplished.

IV. Major pests and diseases of cocoa

Compared to many crops which may be pest and disease-free when outside their centre-of-origin, cocoa can acquire new-encounter diseases from its relatives outside its centre-of-origin (Buddenhagen, 1977; Keane, 1992). The major new-encounter diseases that cause significant losses for cocoa production are: cocoa swollen shoot disease (CSSVD); *Phytophtora* pod rot and canker; vascular-streak dieback; and *Ceratocystis* wilt (Keane, 1992), while witches broom and *Moniliophthora* are the co-evolved parasites that have been affecting cocoa development in its centre-of-origin (Keane, 1992). However, in Samoa, the major diseases that cause large-scale losses are the fungal diseases caused by *Phytophtora palmivora*. Roughly 60-80% of losses are incurred by the pod borer pest during the wet season in Samoa, while 6-8 percent of annual death in the old Trinitario plantations are caused by the canker disease (Bourke, 1992).

CSSVD. Viral diseases, such as the CSSVD, have been a severe problem in the cocoa industry in West Africa, particularly in Togo, Ghana, Nigeria, and even in Cote d'Ivoire (Dzahini-Obiatey et al., 2010). Various strains of this disease have been classified, ranging from the virulent strain A to the mild strains B, C and D (Posnette, 1947). The aggressive strain has been confined to Ghana (West Africa), while the mild strains have been found in Trinidad, Sri Lanka and Indonesia (Keane, 1992; Kenten & Woods, 1976; Posnette, 1947). Symptoms are manifested differently depending on the part of the plant. In the stem, observed symptoms include

swelling, desiccation, and brittleness, while red-vein banding, vein-clearing, vein-banding, chlorosis, mottling, speckling, and the forming of yellow/white lesions are observed in the leaves (Keane, 1992; Posnette, 1947). The effects on the pods include mottling in the early stage, while in later stages, the pods develop dark red marbling, a smoother texture and rounded shape (Posnette, 1947). Its spread is relatively slow due to the sessile habit of its vector, mealy bugs (Tinsley, 1971). However, infections caused by the severe strains can result in about 70% yield loss, due to death within 2-3 years from the initial infection (Muller, 2008). Several management efforts, such as diversification, eradication, and vector control have been implemented to combat this problem. One of the strategies for diversification is to use CSSVD-resistant crops as barriers (Dzahini-Obiatey et al., 2006). Utilising citrus and oil palm as barriers between the cocoa trees, was found to be effective in preventing the spread of CSSVD, with only 1-5 incidences amongst 522 trees planted (Domfeh et al., 2016). Another approach utilises block planting in an area not surrounded by vegetation (Dzahini-Obiatey et al., 2006). This prevents further spread of the virus through isolation, which also helps avoid infections to other crops. Elimination also demonstrates potential control of the disease (Brunt, 2008). The complete eradication of infected trees has been practised by some growers in Ghana as their "cutting out" programme (George et al., 2015). However, it was observed that despite the efforts to remove the old infected cocoa trees, the disease still persisted due to inadequate treatment for the undetected and untreated cocoa trees (George et al., 2015). Although no reported cases of CSSV in Samoa has been published, the potential threat of the CSSV disease must be addressed in the future breeding programs of the country. One of the factors that must be anticipated is the slow spread of the disease. For instance, the aggressive strain was reported to be contained in West Africa, however the mild strain that has reached the Asian region still suggests a high risk of the disease reaching other regions of the world, particularly Samoa.

Phytophthora. The more destructive pathogen *Phytophthora,* which causes pod rot and canker, has resulted in higher global losses than any other cocoa disease recorded (Guest, 2007). The major diseases that are caused by *Phytophthora spp.,* such as the pod rot and canker, have resulted in about 20-30% of production losses worldwide (Erwin & Ribeiro, 1996). Pod rot alone causes 10-20% annual losses on worldwide cocoa production (Galindo, 1992). In Western Samoa alone, the estimated loss caused by the pod rot was 60-80% of cocoa production

(Keane, 1992). The major cause of pod rot in the centre-of-origin areas such as Ecuador is the *P. palmivora,* while areas in West Africa particularly in Nigeria and Cameroon are infested by *P. megakarya* (Keane, 1992). In addition, the less aggressive *P. capsici* and the more virulent *P. citrophthora* are predominant in Brazil (Brasier et al., 1981; Kellam & Zentmyer, 1986). The pod rot caused by *P. palmivora* manifests with the same symptoms as *P. megakarya*, but at a faster rate. The early symptoms are water-soaked spots that become necrotic during the later stage (Pereira, 1992). The necrotic spots show small, hard and dark lesions, which subsequently cause the pod to become shriveled and mummified (Guest, 2007). The process of infection could take only 10-14 days (depending on the species) for the whole pod to rot (J. L. Pereira, 1992). On the other hand, the early-stage detection of cocoa canker is more difficult as it is only manifested by slight discoloration (Leong, Hai, & Beng, 1977). Within 30-60 days, it will encircle the trunk and affects the cambium, causing it to appear watery to gummy with a dull brown-grey colour (Zaiger & Zentmyer, 1965). In many cases, an ooze is produced from the canker and eventually the infection causes the death of the tree (Firman & Vernon, 1970). The success of controlling the canker disease through excision of the lesion through fungicide was reported, by Leong et al. (1977), to be highly variable. Hence, without proper management and preventive action, it is not surprising that this disease will cause further damage on cocoa plantations.

One form of disease control undertaken in Brazil, led by the Comissao Executiva do Plano da Lavoura Cacaueira (CEPLAC), involves the underplanting of mixed hybrid plants to renovate a 52,000 ha cocoa plantation (J. L. Pereira, 1992). Other early attempts for controlling pod rot were through biological controls isolating the antagonists of Phytophthora. One study found that *Botryodiplodia theobromaea* was a potential biological control because of its myco-parasitic activity in the mycelia, sporangia and chlamydospores of *P. palmivora* (Okaisabor, 1968). Natural enemies were also found to be present in the air of cocoa plantations. The *A. giganteus* and *Penicillium purpurescens* were discovered to be highly aggressive against *P. palmivora,* suggesting the existence of natural microflora that prevents pathogens in cocoa plantations (Figueiredo et al. , 1977 (as cited by Galindo, 1992). Recent studies also evaluated and recommended the antagonist fungi *Trichoderma spp.* and bacteria *Pseudomonas spp.*, and *Bacillus spp.* to be used as microbial antagonists for cocoa's *P. palmivora* (Hernández-Rodríguez et al., 2014). However, despite the efforts for natural eradication of cocoa's fungal

diseases, the application of chemicals, such as fungicides, has remained the major control strategy around the world (Pereira, 1992).

Fungicide application has been proven to control pod rot. However, there is some difficulty in its application as application is hampered firstly, by the height of the tree and secondly, by abundant rainfall which compels growers to implement frequent reapplication (Acebo-Guerrero et al., 2012). Because of this, the risk of the development of fungicide-resistant *Phytophthora* is increased (Keane, 1992). To reduce this risk, the application of cuprous oxide must be in rotation with metalaxyl. Hence, the cuprous oxide spray is recommended during dry season, while metalaxyl is recommended during the wet season (McGregor, 1984). While a copper/metalaxyl combination has been effective, a more practical and long-term and cost-effective fungicide was found to have better performance. The application of potassium phosphonate has been demonstrated to control stem canker and reduce pod rot through trunk injection (Anderson & Guest, 1990; McMahon et al., 2010; Opoku et al., 2006). Lastly, the other minor diseases that have been recorded in Samoa are pink disease, tip die back, thread blight, brown pod and root rot (Bourke, 1992).

Although fungal diseases have the most damaging effect in Samoa, insect pests also cause severe damage in cocoa trees. In Samoa, the major insect pest is the rose beetle (*Adoretus versutus*), which damages the flush foliage of young trees (Bourke, 1992). The adults leave a lacy structure in the leaves by consuming the leaf parenchyma, but excluding the veins (Beaudoin et al. , 1995). In the later stages, the shoot growing points are increased, complete defoliation occurs, leading to stunted growth and eventual death (Bourke, 1992). The conventional measure employed in Samoa used stones as borders (one foot or more high) to protect the young plants against the beetles (Urquhart, 1961). Thirty to forty percent of shading was also recommended for cocoa plantations to reduce the damages from the beetles (Beaudoin, 1992). Other early control measures used sprays such as DDT emulsion or lead arsenate (Urquhart, 1961). While biological control, such as entomopoxvirus was observed to kill the rose beetle, concerns for its consequence in inducing epizooty was raised (Beaudoin et al., 1994). Hence with such consequences, it might be a better solution to use other biological agents that have no negative impact on coexisting with cocoa. Other minor insect pests of cocoa that are recorded in Samoa are longicorn borers (*Dihammus fasciatus, D. holotephrus,*

D. marianarum), aphids (*Aphiis gossypii, Toxopera aurati*), termites (*Neotermes* spp.), pod borer (*Acrocercops cramerella*), green looper (*Chrysodeixis eriosoma*), cutworm (*Spodoptera litura*), bagworm (*Mahasena* spp.), and thrips (*Selenthrips rubrocinctus*) (Bourke, 1992).

Aside from pest management efforts, there is also an issue of control of cocoa plant materials in Samoa. A strict regulation on the importation of cocoa beans and plant materials is being implemented to prevent infestation of pests and diseases that might endanger the local cocoa plants (Western Samoa Department of Agriculture Forests, 2005). In addition, new tools for breeding and biotechnology are paving new directions for reducing dependency on chemical usage (Dhawan & Peshin, 2009).

V. Breeding efforts for improving cocoa genotypes for biotic stress resistance and abiotic stress tolerance

The discriminating properties of different cocoa varieties have led to the utilisation of different plant resources (i.e. root stock) for vegetative propagation which was one of the earliest breeding efforts to produce clones (Richards, 1948). It is employed to select the superior germplasm and propagate them without losing the desired genes (Wiesman & Jaenicke, 2002). For instance, high-yielding and rot-resistant cocoa hybrid seeds are produced from the self-incompatible commercial clones (Essola, 2017). While the vast majority of rootstocks used for cocoa production are sourced from open-pollinated seeds, the shoot stock is sourced from the shoot of the mother plant (Sodré & Gomes, 2019). Although the desired genes are maintained in the progenies, vegetative propagation still impose high risk in the transmission of the acquired diseases and other inferior phenotypes of the mother plant. It is therefore inevitable that other methods and variations in producing superior cocoa genotypes are being explored.

Cocoa breeding is highly focused on developing cultivars with pest and disease resistance such as CSSV in West Africa, and Phytophthora pod rod in all areas (Rehm & Espig, 1991). Efforts to develop cocoa lines with resistance to diseases have been implemented. An early study on the use of gamma rays for mutation revealed induced genetic variability in cocoa, which subsequently produced CSSVD-resistant mutants (Adu-Ampomah et al., 1996). This study confirmed that Amelonado-Forastero have the highest susceptibility, while the Upper Amazon

Forastero showed the highest resistance. Moreover, somatic embryogenesis was also found to have a high rate of eliminating CSSV through the production of sterile clonal stock materials (Quainoo et al., 2008). Repeated inoculation of a large number of vectors (mealy bugs) was also performed to increase the CSSVD-resistance of the segregant population from Amazon x Trinitario crosses (Kenten & Lockwood, 1977). The role of phenolics on CSSVD-resistance was also investigated, but a very low correlation was found between the phenolic content and CSSV resistance (Ofori et al., 2015). However, despite the years of efforts to increase the resistance of cocoa to CSSVD, the genetic gain is still very low, due to the lack of major genes responsible for CSSVD-resistance (Padi et al., 2013).

On the other hand, due to the wide infestation of canker in Samoa, breeding strategies were initiated with breeding stock selection (Urquhart, 1952). The earliest recorded selection was the Trinitario variety, Lafi-7, which became the foundation parent for the succeeding generations of cocoa in Samoa (Eden & Edwards, 1952). However, this strong selection resulted in a diversity bottleneck (Slade, 1984). In addition, the Amelonado variety in Samoa was reported to exhibit resistance to *P. palmivora* canker and has fewer losses from pod rot (Bourke, 1992). Multiple disease-resistance to pod rot and canker was also studied. The genotype B36Xt65/326 from the cross Alpha B36xPa7/808 and Pa7/808xPound 7 was identified to have multiple-disease resistance (Nyadanu et al., 2012).

Aside from the pest and disease resistance breeding, cocoa breeding for early fruit bearing, high yield, and adaptability for varying environmental conditions are also being prioritised (Rehm & Espig, 1991). It is estimated that there are fewer than 100 new varieties of cocoa released since 2000 (Farrell et al., 2018). The same study concluded that the low number was caused by limited resources for breeding infrastructures rather than the source of germplasm. This shows that there has been limited studies and breeding programs that aim to develop heat tolerant and waterlogging tolerant cocoa varieties.

A molecular and physiological analysis of waterlogging-tolerant cocoa clone TSA 792 showed a survival index of 94% after 45 days of flooding (Bertolde et al., 2010). A significant finding from the study suggested that the absence of leaf chlorosis may be a relevant trait to assess for waterlogging tolerance. Moreover, physiological and gene expression was also shown to have

significant roles in selecting for waterlogging-tolerant genotypes. Physiological and biochemical factors such as chlorophyll and carbohydrates are not significantly reduced in tolerant clones (Bertolde et al., 2012). While increased in the expression of the alcohol dehydrogenase genes (ADH), which is an enzyme that helps maintain energy production even during anoxia (Bertolde et al., 2014).

On the other hand, while there appear to be no extensive studies on improving cocoa genotypes for heat-tolerance, future breeding efforts may also deal with selecting heat-tolerant cocoa genotypes by targeting traits such as stomatal sensitivity and conductance (Medina & Laliberte, 2017). Nevertheless, new advancements on molecular studies and biotechnology have allowed development of molecular markers for the identification of genotypes that could potentially be the parents for future breeding programs.

VI. Molecular markers for cocoa varietal identification

Cocoa is a diploid perennial species (2n = 2x = 20) with a small genome size between 411 Mb and 494 Mb (Argout et al., 2017; Argout et al., 2010). The first draft of a cocoa genome was sequenced from the Belizean Criollo genotype (B97-61/B2) in 2011 (Argout et al., 2010). The draft covered only ~76% of the estimated size of B97-61/B2 genome. The updated version (v2.0) of the Criollo genome, published recently, has a reduced scaffold number and improved anchored assembly (96.7% compared to 66.8% from v1.0) (Argout et al., 2017). It was also observed that the total size of the assembly was reduced to 75.5% of the estimated genome size of the Criollo genotype, mainly due to reduced gap size. While the sequence project is aimed at providing the reference genome for cocoa, it is uncertain if Criollo is the best representative of cocoa genotypes around the world, despite its high level of homozygosity as a result of years of domestication. Having considered that, a self-compatible homozygous Amelonado genotype, the Matina 1-6 clone traditional cultivar, was used as the reference for another cocoa genome sequencing project (Motamayor et al., 2013). In 2017, a more refined sequence (v1.1) was published using the same clone. The genome assembly of the Matina clone (445 Mbp) was comparatively larger than the Criollo genome (430 Mbp) (Argout et al., 2017; Motamayor et al., 2013). Currently, the Matina genome projects are handled in the cacao genome database (www.cacaogenomedb.org). This database is a collaborative project

between MARS Inc., ARS-USDA, IBM, Clemson University Genomics Institute, PIPRA, Hudson Alpha Institute for Biotechnology, National Centre for Genome Resources, Indiana University, and Washington State University.

Most cocoa trees are self-incompatible hence it is generally known to be outbreeding (de Schawe, Durka, Tscharntke, Hensen, & Kessler, 2013). The incompatible reproduction results in 25, 50 or 100 percent failure of the gamete fusion (Toxopeus, 1985). Subsequently, it is followed by ovary development failure and the death of the flower 3-4 days after pollination. A recent study on the wild and cultivated cocoa found an occurrence of 7-14% self-pollination (de Schawe et al., 2013). This is not surprising as there has been recorded variation for incompatibility in different populations of cocoa (Toxopeus, 1985). For instance, the Amazon and Trinitario varieties are self-incompatible while the Amelonado varieties are self-compatible. This reproductive condition has a significant effect not only on cocoa production but also on cocoa breeding. It also has a big influence on the characterisation of the cocoa genotypes in Samoa.

Utilisation of molecular markers to identify existing cocoa varieties in Samoa can provide conclusive and high-throughput results. Distinguishing varieties and landraces can be difficult due to their large number, hybridization, and because of the phenotypic plasticity they have acquired in an effort to acclimate to changing environments (Korir et al., 2013). In such situations, the conventional morphological characterisation may give false identification. Hence, the utilisation of molecular markers for DNA profiling provides more advantages in early-vegetative-stage analysis, cost-effectiveness for a high number of individuals, and more rapid results (Gosh, Sarkar, & Datta, 2001). Now that the cocoa genome has been sequenced and annotated (Argout et al., 2017; Argout et al., 2010), molecular studies have become more accessible and more advanced. Early studies on the use of molecular markers to survey cocoa genotypes used restriction fragment length polymorphisms (RFLP) and random amplification of polymorphic DNA (RAPD) markers (Laurent et al., 1994; N'Goran, Laurent et al., 1994). These markers, however, can suffer from low reproducibility and low efficiency. For instance, random amplified polymorphic DNA (RAPD) marker is time-consuming, laborious and has issues in terms of reproducibility (Bardakci, 2001). In addition, another marker, the restriction fragment length polymorphism (RFLP) is also laborious that required large quantity of DNA and tools for

cloning, while amplified fragment length polymorphism (AFLP) shows inconsistency in genetic maps and requires good primers (Budak et al., 2004). Nevertheless, these markers have been valuable in several MAB and linkage mapping in tomato, barley, soybean and rice (Budak et al., 2004; De Giovanni et al., 2004; Guan et al., 2014; Huang et al., 1997; Zhang et al. , 2010). To date, these markers have been replaced by highly-reproducible, co-dominant, locus-specific, PCR-based, multi-allelic and highly-polymorphic simple sequence repeat (SSR) markers (Budak et al., 2004; Jiang, 2013). It has been widely used in identifying cultivars and varieties across economically important crops such as potato (Rosa et al., 2010), cotton (Pereira et al., 2015) grapevine and olives (Mackay et al., 2008) and rice (Pourabed et al., 2015). The first isolation and characterisation of SSR markers in *Theobroma cacao* L. generated 23 polymorphic markers (Table 1) (Lanaud et al., 1999). Eleven of these markers, together with four additional markers were further screened and optimised for international standards for discriminating different accessions of *Theobroma cacao* (Saunders et al. , 2004). Saunders et al. (2004) concluded that eleven of the markers are more than sufficient to identify Forastero (Upper Amazon type, Lower Amazon type and unidentified type), Criollo, Trinitario, Hybrid, and cocoa that have an unknown or mixed origin. In addition, two hundred and one markers were developed for the construction of the linkage map of the crosses between an Upper Amazon Forastero and Trinitario clones (Pugh et al., 2004). While Risterucci et al. (2000) supplemented the SSR markers previously screened (Lanaud et al., 1999) to create a high-density molecular linkage map.

Subsequently, further studies in genetic diversity have employed these markers to differentiate international cocoa germplasms. For instance, from the markers screened by Lanaud et al. (1999) and Saunders et al. (2004), twelve SSR markers were utilised to assess the cocoa collections in West Africa (Aikpokpodion et al., 2009). When fourteen SSR markers (Lanaud et al., 1999; Saunders et al., 2004) were employed in the characterisation of cocoa genotypes in Vietnam, only eleven showed high allelic diversity to differentiate the genotypes (Everaert et al., 2017). Three main groups were obtained from the study: the Forastero (Upper and Lower Amazon type); Trinitario; and the hybrid cultivars from Forastero x Trinitario cross. Moreover, fifteen of the screened markers (Lanaud et al., 1999; Risterucci et al., 2000; Saunders et al., 2004) were employed to group the cocoa germplasm collected from Central and South America (Zhang et al., 2012; Zhang et al., 2009).

Furthermore, a population differentiation study was conducted in the Amazonian cocoa tree using the previously reported SSR markers from Pugh et al. (2004), Brown et al. (2005) and Lanaud et al. (1999) by Motamayor et al. (2008). Ninety-six SSR markers out of the initial 106 SSR markers used were retained for the analysis of 1241 individuals. That work identified 10 major genetic clusters: Amelonado, Contamana, Criollo, Curaray, Guiana, Iquitos, Marañon, Nacional, Nanay and Purus. This has been the only study that considers the 10 genetic clusters of cocoa; however, this system of identification has been taken up by germplasm centers. In the past decade since the release of the cocoa genome draft in 2009, only the 3-4 major groups (Criollo, Forastero, Trinitario and/or Nacional) seem to be considered in cocoa genetic diversity studies. Ultimately, this will need to be resolved by the cocoa community, as different classification schemes will only lead to confusion within the cocoa research and production communities.

SCIDI genotyping project

I. Introduction

Cocoa (*Theobroma cacao*) is widely utilised for chocolate production, and for popular products such as drinks, flavouring (pastries and ice cream), and butter for cosmetics and pharmaceutical industries (Carr & Lockwood, 2011). Its production has significantly increased over the last 30 years, with a global production of 1.3M tonnes in 1966 in contrast to 4.5M tonnes in 2016 (FAOSTAT, 2018). Countries situated between 10ºN and 10ºS of the equator are the most conducive for growing cocoa (Wiah & Twumasi-Ankrah, 2017). Among those countries, the largest producer of cocoa is Cote d'Ivoire, followed by Ghana, Indonesia, Nigeria and Brazil (FAOSTAT, 2018). Despite the fact that cocoa's centre of diversity is along the Amazon basin in the northern regions (lowland) of South America (Simmonds, 1998), cocoa is still mostly produced in West Africa (Cote d'Ivoire, Ghana, Nigeria), which provides 70% of the world's cocoa production (Wessel & Quist-Wessel, 2015). The increasing production in West Africa can be attributed to various government support and implementation schemes such as land area expansion for cocoa growing (Wessel & Quist-Wessel, 2015) and government subsidies (Emmanuel & Qineti, 2018; Wessel & Quist-Wessel, 2015). These efforts demonstrate how the interaction of the environmental conditions and the management system is significant for the cultivation of cocoa. Apart from the cocoa-producing regions mentioned, thriving areas in the Pacific Islands of Oceania also have a rich history of production for local and international markets.

The Oceania region, which comprises Australia, New Zealand and the Pacific Islands, supplies only 1.3% of the world cocoa production (FAO, 2018; FAOSTAT, 2018). Samoan cocoa is regarded as high-quality beans, which originated from the Criollo and Trinitario varieties that were introduced by German interests (Singh & Bhati, 2010). Together with 16 other countries, Samoa's cocoa beans were classified as fine flavour cocoa in the 1993 International Cocoa Agreement (United Nations, 1993). The high quality of Samoan cocoa has been recognised since the 1960s. It was in that period when its cocoa export industry was at its peak, producing over 5,000 tonnes of cocoa (Dillon et al., 2014). However, climate change, pests and diseases and natural disasters have significantly reduced Samoa's cocoa production (Cocoa Samoa Ltd.

(CSL), 2015). The decline in Samoa's cocoa production began in the 1970s, but the decline picked up pace in the 1980s (Dillon et al., 2014). At present, cocoa production in Samoa relies mainly on small household farmers (Samoa Bureau of Statistics, 2016). In recent years, international efforts from neighboring countries, such as New Zealand and Australia, have aided the revival of the Samoan cocoa industry (PHAMA, 2016). Local efforts are also being implemented for the conservation and improvement of local cocoa varieties. As a result, the number of households planting cocoa trees rose by 30% between 2009 and 2015 (Samoa Bureau of Statistics, 2016). However, years of untracked breeding and selection have made it difficult for local researchers to accurately identify the potential breeding materials in the country. One approach being explored is the utilisation of molecular tools to identify the cocoa varieties in the country. This will facilitate breeding programmes using the identified genotypes to help stimulate the cocoa industry in Samoa.

The current study utilizes a set of 15 microsatellite markers to identify genetic groupings within the Samoan cocoa collections, to quantify genetic diversity at collection sites and by genetic groupings, and associate the genetic diversity discovered with traditional cocoa varieties.

II. Materials and methods

Plant materials

Leaf material from two-hundred cocoa trees was sampled from four sites (50 trees per site) across Samoa by the Scientific Research Organisation of Samoa (SROS): Samoa Trust Estate Corporation (STC) in Mulifanua, Upol; Melzi plantation in Tafaigata, Upolu (MZ); Saipipi, Savaii (SV); Vaisala, Savaii (VS) (Supplementary Figure 1). From each tree, leaf material was collected fresh for immediate processing and also dried (in silica gel) for long-term preservation.

DNA extraction

Two DNA extraction protocols were investigated for the initial marker screening to identify which method would be optimum for genotyping. The first protocol was based on the Qiagen DNeasy Plant Mini Kit (Qiagen, USA) and the second protocol followed a modified STE-CTAB protocol, described below.

One mL of STE buffer was added to 50 mg dried ground cocoa leaf sample. The lysate was vortexed and centrifuged at 5000rpm for 10 minutes, discarding the supernatant. 600 µL of CTAB (pre-warmed at 65°C and added 2 µL of β-mercaptho-ethanol per mL of CTAB) was added to the pellet and the mixture was vortexed thoroughly. Incubation was done at 65°C for 1 hour after adding 100 µL of the BSA:NaCl (1:5) solution to the tube. The suspension was mixed and vortexed thoroughly after incubation. Then, after a thorough mix and vortex, 600 µL of chloroform was added, shaken to mix, and centrifuged at 14,000rpm for 5 minutes. The pellet was discarded and 600 µL of chloroform was added to ~400 µL supernatant which was transferred to a new tube, and the mixture was shaken and centrifuged at 14,000rpm for 5 minutes. ~300 µL of the supernatant was transferred to a final tube, and 500 µL of isopropanol was added. The mixture was inverted, incubated for 25 minutes and then centrifuged at 14,000rpm for 15 minutes. The pellet was washed with 500 µL of 70% ethanol, then centrifuged at 14,000rpm for 5 minutes. After removing the ethanol, the pellet was allowed to air-dry overnight. The dried pellet was resuspended in 100 µL of TE.

DNA quality was checked on 1% agarose gels through electrophoresis. The result was obtained by viewing the Ethidium Bromide-stained gel in a Bio-Rad Gel Documentation system (Bio-Rad, USA). The DNA concentration was quantified using a NanoDrop™ Spectrophotometer (Thermo Scientific, USA). High quality DNA samples were diluted at 1:5 with sterile water.

It was concluded that the Qiagen kit protocol was optimum for production of high quality genomic DNA from cocoa leaf. Hence, the total genomic DNA was extracted from freshly sampled young cocoa leaves using the Qiagen DNeasy Plant Mini Kit (Qiagen, USA). DNA quality was checked following the protocol above. The extractions were performed by colleagues at SROS in Samoa and posted to the Massey research group in New Zealand upon completion. In the end, three samples did not amplify well for any markers (subsequent sections) and were removed from analyses.

Microsatellite genotyping

Among the published cocoa SSR markers reviewed, fifteen markers were selected and sequenced for an initial trial (Supplementary Table 1). The forward primers were 5'-tailed with an M13 tag sequence (CACGACGTTGTAAAACGAC) (Brownstein et al., 1996). Polymerase chain reactions (PCRs) were performed to assess banding patterns and genotyping efficiency. Markers that showed a distinguished banding pattern and clear genotype results were to be used for the subsequent analysis of the cocoa varieties; in the end, all 15 initial markers performed well.

The PCR amplifications were carried out using a Biometra Ti thermal-cycler. The reaction mix contained 3.7 µL ultrapure water, 1 µL 10x buffer BD, 1 µL $MgCl_2$ (4.50 µM), 0.2 µL dNTP (10 µM), 1 µL forward primer (0.20 µM), 1 µL reverse primer (4.50 µM), 1 µL M13 primer (labelled with FAM, VIC or NED) (4.50 µM), 0.1 µL FirePol Taq Polymerase (5U/ µL) and 1 µL DNA (diluted to 1:5). Two PCR profiles were used to check the optimum amplification condition for the markers, since the universal fluorescent primer M13 was also incorporated in the mix. The first PCR profile is a standard used often utilised for M13-tailed microsatellite PCR. The PCR conditions were: 95ºC for 3 mins (initial denaturation), followed by 35 cycles of 95ºC for 30 sec (denaturation), 52ºC for 40 sec (primer annealing), and 72ºC for 40 sec (extension), and 72ºC

for 20 mins for the final extension. The second thermal cycling profile (Everaert et al., 2017) was: 94ºC for 4 mins (initial denaturation), followed by 35 cycles of 94ºC for 30 sec (denaturation), 46ºC or 51ºC for 1 min (primer annealing), and 72ºC for 1 min (extension), and the final extension at 72ºC for 15 mins. The protocol from Everaert et al. (2017) was used for all subsequent runs as it produced more prominent bands during electrophoresis trials. The PCR products were checked by resolving 5 µL of amplicon and 3 µL of 3X loading dye in 1.5% agarose gel electrophoresis for 100 minutes at 75 volts.

Marker pooling was carried out for three microsatellite markers, each labelled with a different dye in a ratio of 1.75 µL: 1.50 µL: 2.00 µL for markers labelled with FAM: VIC: NED, respectively. For genotyping, 1 µL of the pooled PCR products was then mixed with 9 µL of Hi-Di formamide mix (Applied Biosystems, California, USA). The Hi-Di mix was prepared by adding 14 µL of ROX-labelled CASS ladder (3.5 µL each of 100, 200, 300 and 400 bp ladder) (Symonds & Lloyd, 2004) and 86 µL of water to 1000 µL of Hi-Di formamide. The prepared genotyping plate was submitted to the Massey Genome Service (Palmerston North New Zealand) for microsatellite genotyping through fragment separation (capillary electrophoresis) on an ABI 3730 Genetic Analyzer (Applied Biosystems).

Data analysis

Allele calling

The raw genotype data were analysed using GeneMapper 5 Software (Applied Biosystems, USA) for initial allele calling. Two individuals (SV14 and SV49) had very low amplification rates so were removed from further analyses.

Data groupings

The data were analysed following two grouping schemes. The first scheme maintained individuals within collection sites. That is, all individuals from a particular site were treated as a group for comparisons of genetic diversity and differentiation (this configuration is referred to as *original site*, hereafter). However, this level of analysis does not consider one of the main objectives of the study, which is the identification of natural genetic groupings. To this end, analyses were conducted on another arrangement that disregarded the site-of-origin information and sought to identify genetic groupings, simply based on the genotype data. Once

genetic groupings were identified (using the STRUCTURE v.2.3.4 software, Pritchard et al. 2000), those groups or sub-sets thereof were further analysed and referred to as *inferred genetic groups*.

Genetic diversity

The GenAlEX ver. 6.51b2 software (Peakall & Smouse, 2005) was used to characterize allele frequency and heterozygosity metrics, generate summary statistics, genetic distances, including Fst (Wright 1968) and Principal Coordinate Analysis (PCoA) among the collected samples. The analysis of molecular variance (AMOVA) was also carried out using GenAlEX, based on the Euclidean distances among individuals. The marker data were permuted 999 times to test the significance of the results.

Population structure analysis

The population structure of the 197 cocoa samples was analysed with STRUCTURE v.2.3.4 software (Pritchard, Stephens, & Donnelly, 2000). The parameters applied were: the *admixture model* and *correlated allele frequencies*. Each run consisted of a burn-in length of 100,000 generations and MCMC (Markov chain Monte Carlo) length of 1,000,000 generations for each value of K (K = 1 to K = 5); 10 replicates were run for each K value. K represents the number of 'ancestral populations'. The 'ancestral populations' represent collections of alleles that make up a set of genotypes that either represent current observations of represent historical genotypes, remnants of which are found I the current samples. To determine the best value of K, the log likelihood Ln P(D) = L(K) was plotted against the K values. As K approaches its most liekly value, L(K) will start to increase in smaller increments until it plateaus. The best iteration is selected based on the inflection point in the data.

Utilizing cocoa accession controls

Despite best efforts, obtaining leaf or DNA material for particular cocoa accessions for use as positive controls was not possible in the study timeframe. We have found communication with germplasm certers, such as C.A.T.I.E., to be the primary issue, with weeks and months between email replies and still almost no progress. Instead, existing genotype data from the same markers used in the current study were obtained from authors of previous studies. In particular, the data from the Motamayor et al. (2008) paper was utilised. That study identified

10 genetic grouping within *Theobroma cacao*, utilising ~100 markers and >1,000 accessions, thus setting the standard for cocoa genotyping. The challenge with utilising such data lies with (1) run-to-run variability in genotype data and (2) variability due to genotype calling methodologies among research groups. We therefore focused on genetic groupings (identified by that group) that had very little variation (e.g., the Criollo group) to align our genotype data with theirs. Missing data for the control accessions was interpolated.

Cluster analysis

A genetic distance matrix using the Provesti model was created using R (R Core Team, 2013) and exported in the *nexus* file format**.** Cluster analysis was performed using the SplitsTree (v4.15.1) software (Huson & Bryant, 2005), employing the Neighbor-Net distance-based method. Neighbor-net graphs were generated for sub-sets of individuals based on the STRUCTURE results (further described in the Results section). One individual of each of the Criollo and Amelonado varieties were included in one such analysis to identify genetic affiliations. These two populations were selected because of the fairly uniform genotypes observed from their original allele calls.

III. RESULTS

Fifteen markers is sufficient to recover most cocoa genetic groupings

To confirm that the initial selection of microsatellite markers would be sufficient to differentiate cocoa varieties, we worked with the Motamayor (2008) data set. Of the 15 markers initially selected for the current study, 12 were used in the Motamayor study. That data set was pared down to those 12 markers and analysed using STRUCTURE (Pritchard et al., 2000); see Methods. The results at *K*=10 recovered nine of the ten groupings identified by the full dataset (not shown). Importantly, the set of 12 markers differentiated Criollo from all other varieties.

AMOVA and genetic diversity in the four Samoa collection sites (*site groups*)

All analyses indicate little to no genetic differentiation among the four Samoa cocoa collection sites (*site groups*). A summary of the analysis of molecular variance (AMOVA) shows little differentiation among the four Samoan cocoa sites and even among individuals (Table 1). This finding is supported by pairwise comparisons among sites using the traditional Fst metric, which also shows little differentiation among sites (Table 2). Fst measures differentiation among groups (in this case, collection sites) based on allele frequencies. The Fst metric ranges from 0 – 1, where a higher value represents greater differentiation. The pairwise comparisons among sites were all ≤ 0.023, again, indicating little to no genetic differentiation. The only differences among sites of note regard the increased number of effective alleles (Ne; 2.22) and elevated observed heterozygosity (0.42) and expected heterozygosity (0.48) at the STC site, relative to the other three sites (Table 3). This finding will be revisited in later sections.

Table 1. Summary of AMOVA analyses comparing the original *site groups* (*SG*) and the *inferred genetic groups* (*GG*) from STRUCTURE analyses at *K* = 2 and *K* = 3.

Source	df	SS	MS	Est. Var.	%
***SG* (n=197)**					
Among Pops	3	1.777	0.592	0.001	0.25%
Among Individual	193	92.322	0.478	0.008	1.74%
Within Individual	197	91.000	0.462	0.462	98.01%
Total	393	185.099		0.471	100%
***IGG-K=3* (*n*=83)**					
Among Pops	2	1.936	0.968	0.011	2.16%
Among Individual	80	36.371	0.455	0.000	0.00%
Within Individual	83	40.000	0.482	0.482	97.84%
Total	165	78.307		0.493	100.00%
***IGG-K=2* (*n*=123)**					
Among Pops	1	0.518	0.518	0.001	0.14%
Among Indiv	121	56.499	0.467	0.000	0.00%
Within Indiv	123	58.500	0.476	0.476	99.86%
Total	245	115.516		0.476	100.00%

Table 2. Pairwise Fst values for three different levels of comparison, including all samples compared among *site groups* (*SG*) and two levels of genetic structure based on results from the program STRUCTURE.

SG (n=197)					***IGG*** at K=3 (n=83)				***IGG*** at K=2 (n=123)		
MZ	**STC**	**SV**	**VS**		**1**	**2**	**3**		**1**	**2**	
0				**MZ**	0			**1**	0		**1**
0.016	0			**STC**	0.114	0		**2**	0.048	0	**2**
0.023	0.012	0		**SV**	0.112	0.108	0	**3**			
0.019	0.020	0.016	0	**VS**							

Table 3. Genetic diversity summary statistics for the original *site groups* and the *inferred genetic groups* based on STRUCTURE results: Na (average number of alleles per locus, Ne = average number of effective alleles per locus, Ho = Observed heterozygosity, He, expected heterozygosity, Fis = Wright's fixation index.

Population	Na	Ne	Ho	He	F_{is}
SG (n=197)					
MZ	5	1.92	0.40	0.47	0.14
STC	5	2.22	0.48	0.52	0.08
SV	5	1.99	0.42	0.48	0.12
VS	4	1.85	0.41	0.45	0.08
IGG at *K*=3 (n=83)					
1	3	1.61	0.34	0.34	0.02
2	3	1.81	0.42	0.42	0.01
3	5	2.84	0.53	0.59	0.08
IGG at *K*=2 (n=123)					
1	5.267	2.675	0.469	0.567	0.150
2	2.667	1.833	0.399	0.449	0.102

Patterns of genetic structure within the Samoa cocoa samples

The program STRUCTURE (Pritchard et al., 2000) uses a maximum likelihood approach to construct genetic groups based on genotype data from multiple individuals and an analytical model based on the Hardy-Weinberg equilibrium criterion. The allele frequency parameters of the hypothetical genetic groupings are then used to assign the genotypes of real samples back to one or more of the identified genetic groups. The assignments for each individual are represented by a stacked bar graph where different colors represent the different genetic groups, each showing the proportion of an individual's genotypes that are assigned to the genetic groupings. The analysis is run multiple times, assuming a numerical range of genetic groups. For example, when assuming two genetic groups, one would indicate *K*=2, etc.

STRUCTURE results for the SCIDI data set identify strong signals at *K*=2 and *K*=3 (Figure 2); beyond *K*=3, there is little support for any greater resolution. At *K*=2, the VS farm appears fairly uniform and the remaining three farms appear to have more of a mix of genotypes. At

K=3, for which there is greatest statistical support, the VS farm still has less variation than the other three farms but the STC farm has a high frequency of a genotype that is less common elsewhere. At each *K* value, most individuals appear to be mixtures of the two or three genetic groups revealed; however, at *K* = 3, a set of 83 individuals are comprised of 90% or more of just one of the ancestral genetic groups. To characterize the three genetic groups (*inferred genetic groups*), these individuals are analysed separately below, apart from the remaining admixed individuals.

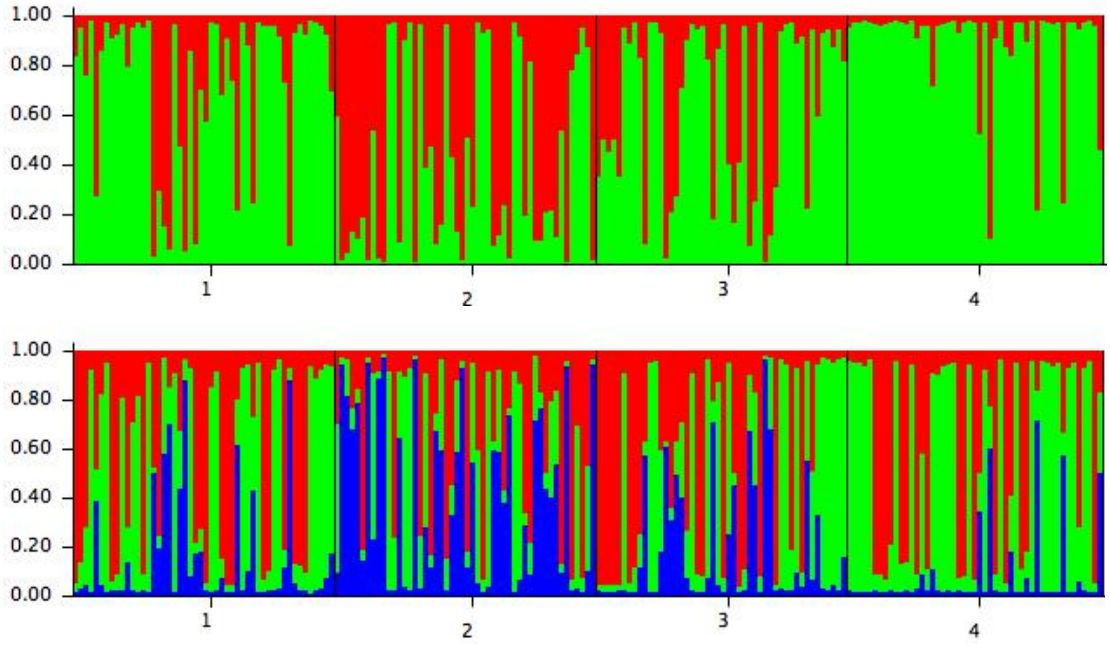


Figure 2. STRUCTURE plots with the greatest likelihood scores at *K*=2 (top) and *K*=3 (bottom), indicating the proportion of each individual's genotypes that belong to the two or three genetic groups assumed (indicated by red, green, and blue). The x-axis indicates the collection site for each individual; 1 = MZ, 2 = STC, 3 = SV, and 4 = VS. The y-axis indicates the coefficient of ancestry.

Another commonly employed approach to visualizing genetic structure within a set of samples is the generation of so-called Neighbor-nets. The program SPLITSTREE utilizes a genetic distance matrix based on genotype data to produce a Neighbor-net, which is similar to a phylogenetic tree, except that it represents all possible relationships among individuals (insofar as is possible to visualize within two dimensions). When dealing with intraspecific genetic relationships, where the lineage history is potentially complicated and reticulated (i.e., not bifurcating (simple branching)) due to admixture, Neighbor-nets should better represent relatedness among individuals. The term 'split' as used in the context of Neighbor-nets is a series of parallel lines that represent genetic signal in the data that distinguishes one group of individuals from others. The longer the split in a Neighbor-net, the more support there is for the pattern of distinction among groups.

We present two Neighbor-nets here. The first uses all 197 Samoa samples and identifies one primary split among individuals (Figure 3). This appears to roughly coincide with the STRUCTURE results at *K*=2, which is a good result, as these are very different types of analyses essentially identifying the same pattern.

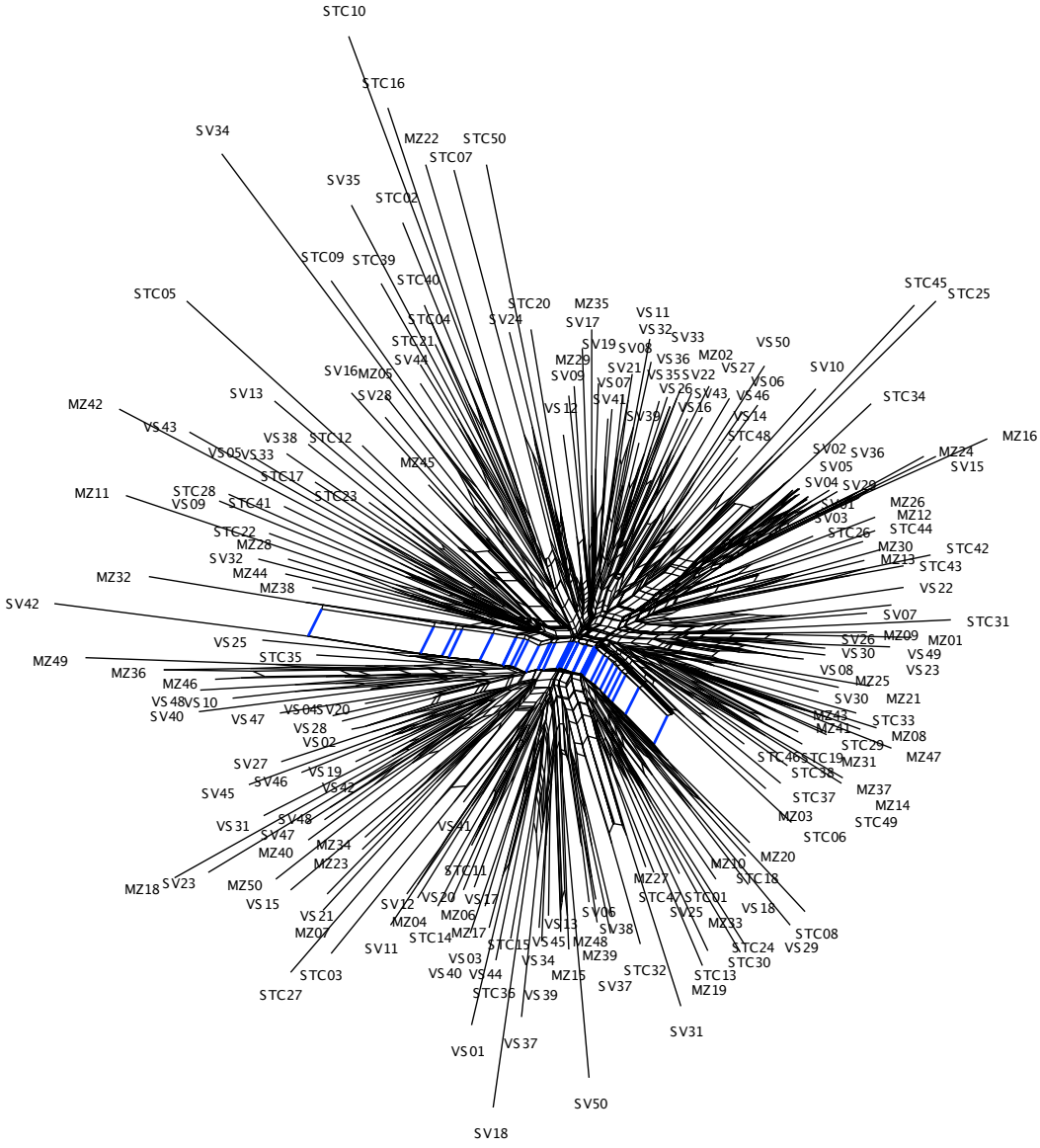


Figure 3. Neighbor-net generated from 15 microsatellite loci for 197 Samoa cocoa samples. The result identifies one primary split (highlighted here in blue) that divides samples into two main groups. There appears to be little correspondence between this pattern and farm-of-origin for samples (each group contains individuals from all four collection sites) but the pattern does appear to roughly correspond with the genetic structure results from STRUCTURE analyses at *K*=2.

The second Neighbor-net was generated using only individuals that were identified in STRUCTURE analyses as having 90% or greater genetic affinity with only one *inferred genetic group* (n=87). However, we also included one representative for each of two varieties, Amelonado and Criollo, in this analysis. The results (Figure 4) show three distinct genetic groups that essentially match the STRUCTURE *K*=3 results; here again, consistent results across different methods provides support for the observed patterns. The results of this analysis further show clear associations with Amelonado for what we have termed group I individuals and an association with Criollo for group II individuals. Each of those clusters of individuals have one variety embedded within them. The third cluster consists of comparatively fewer individuals, but, notably, all but two of them are from the STC site.

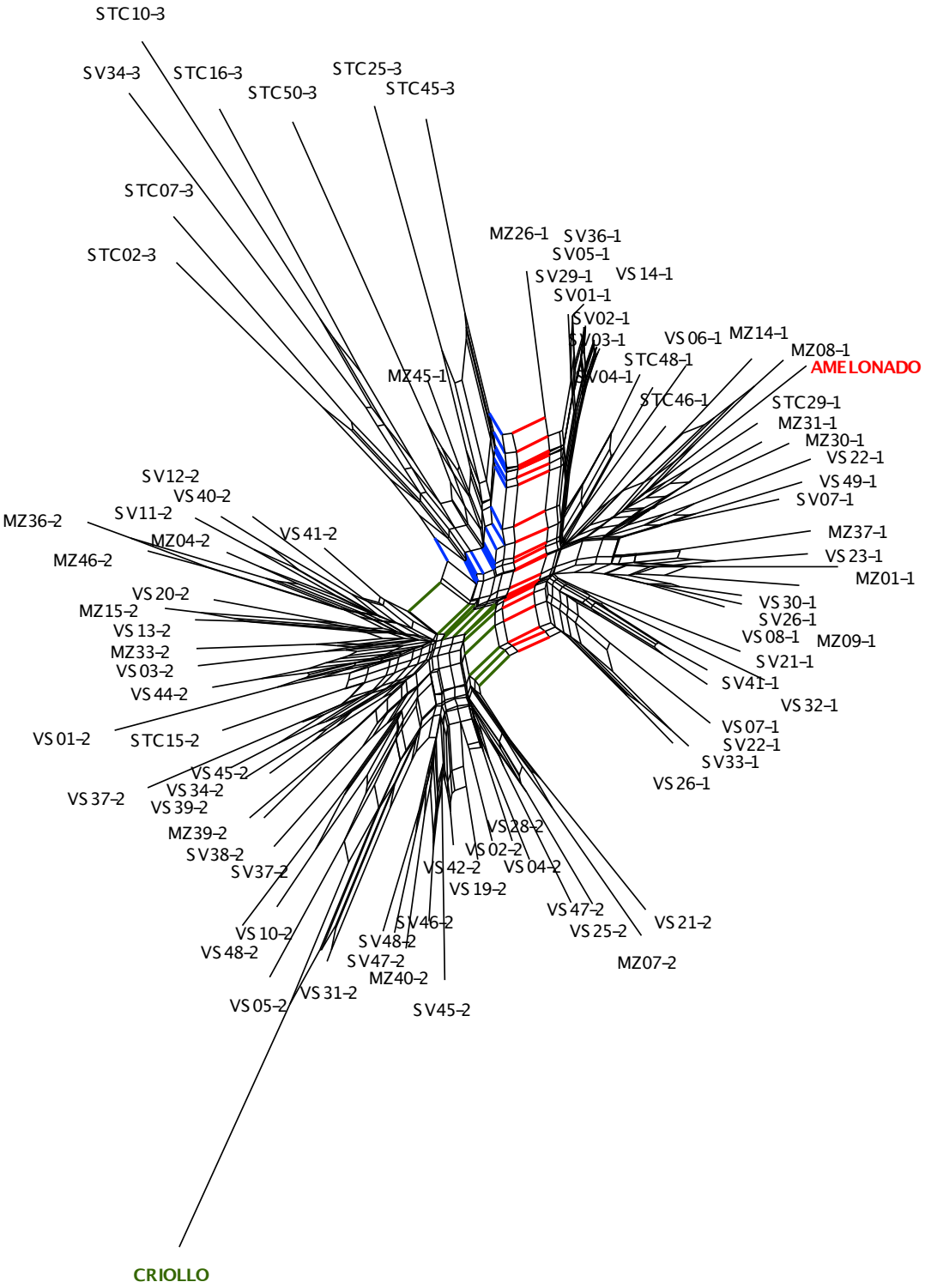


Figure 4. Neighbor-net of SCIDI individuals (n=83) identified as having genotype compositions that are 90% or greater from a single genetic grouping identified by the STRUCTURE *K*=3 analysis and a representative for the Amelonado and Criollo varieties. Using a very different analytical approach, this analysis identifies the same groupings as those revealed in the STRUCTURE analyses (-1, -2, and -3 in sample IDs indicate the clusters (*inferred genetic groups*) identified by the STRUCTURE analysis). Splits that differentiate the three groups are highlighted in blue, green, and red; note that the blue split here is unrelated to the blue split identified in Figure 3.

Finally, traditional PCoA analyses reveal essentially the same patterns described above but with an interesting reveal (Figure 5). A PcoA including all 197 Samoa cocoa samples and the representative varieties shows that the axis of greatest variation (Coordinate 1) essentially is defined by the Criollo sample at one end and very nearly by the Amolenado sample at the other extreme. Most of the Samoa samples sit within this range. Similarly, when only the individuals based on STRUCTURE *K*=3 results are examined (along with the variety controls), most individuals fit within the long axis defined by the two controls. Furthermore, the group III individuals appear intermediate to the goup I and group II individuals along the long axis, as expected based on the Neighbor-net results. The interesting feature in both plots is that the Criollo sample is not nested within the group II individuals but instead is set apart, raising an interesting hypothesis regarding Samoa cocoa samples that will be explored in the Discussion.

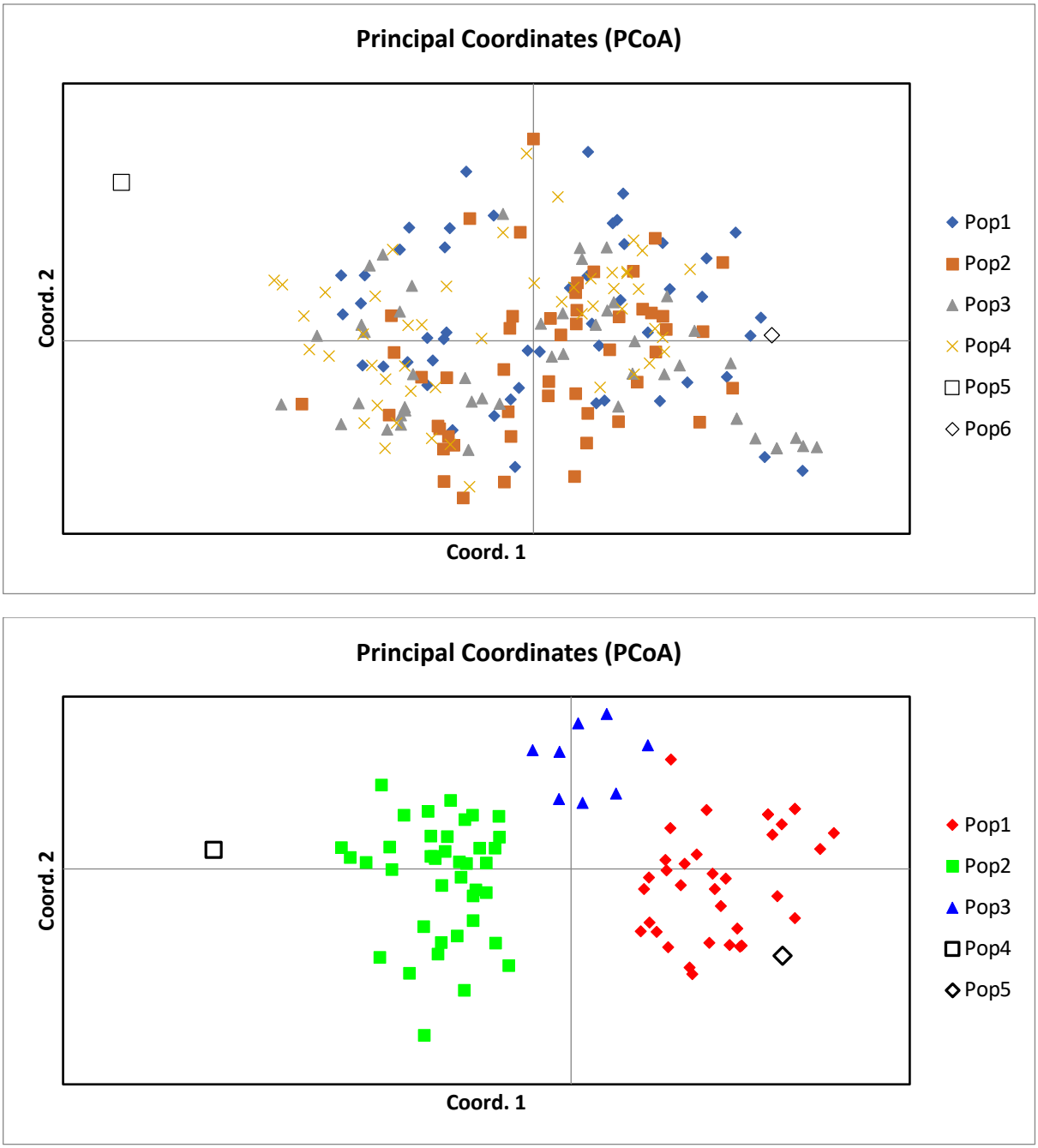


Figure 5. PcoA results for two levels of analysis. The top panel shows results for all 197 Samoa cocoa samples and the control varieties; Pop1 = MZ, Pop2 = STC, Pop3 = SV, Pop4 = VS, Pop5 = Criollo, Pop6 = Amelonado. The bottom panel shows results for the individuals with >90% genotype affinity from the STRUCTURE results; Pop1 = group I, Pop2 = group II, Pop3 = group III, Pop4 = Criollo, Pop5 = Amelonado.

AMOVA and the genetic parameters of the selected microsatellite markers for the inferred genetic groups identified by STRUCTURE and SPLITSTREE and PcoA analyses

In contrast to the *site group* analyses, the three groups of individuals identified by STRUCTURE and SPLITSTREE analyses (*inferred genetic groups*) show clear genetic differentiation. This is quantified by AMOVA results that show a higher degree of among group genetic variance than the *site group* data (Table 1). This is reiterated by the Fst results (Table 2), which reflect the greater differences among the three *inferred genetic groups*; all pairwise comparisons among these three groups have elevated but similar differentiation (Fst range: 0.108 – 0.114; Table 2). As noted previously, group III, largely comprised of individuals from the STC site, despite being represented by fewer samples, has an elevated number of alleles, effective alleles, observed heterozygosity, and expected heterozygosity than the other two groups. Also of note for group III is that it has the greatest number of private alleles (Supplementary Table 3).

IV. Discussion

Fifteen microsatellite markers resolve genetic differentiation within *Theobroma cacao* L.

Despite the large number of microsatellite markers developed and 33utilized in the last two decades for cocoa research, a subset of ~15 microsatellite loci has been shown to be sufficient for discriminating cocoa varieties (Aikpokpodion et al., 2009; Everaert et al., 2017; Zhang et al., 2012; Zhang, Mischke, Johnson, Philips-Mora, & Meinhardt, 2009). In our re-analysis of the large dataset of Motamayor (2008) using just 12 markers, we were able to recover nine of the 10 genetic groups originally described by the authors (not shown). In the current study, 15 markers were utilised and found to differentiate three primary genetic groups in the Samoan cocoa samples examined, with most individuals being combinations of those three groups.

Genetic data indicate little differentiation among Samoan collection sites

Among the four Samoa cocoa collection sites, little genetic differentiation was evident from any analytical method. Given the complex planting history and subsequent natural hybridization and regeneration, this is unsurprising. However, the standout at this level of analysis is the STC site. Although individuals from this site do not necessarily cluster together, there is clearly more genetic diversity in the set of samples from this site than from the others (Table 3). The literature indicates that the STC site sits within a geographic region with favourable rainfall for cocoa growth, especially compared with the MZ and SV sites (Ward & Ashcroft 1998) and would have been a primary production area during the peak of Samoa's cocoa industry, when 'Samoan Trinitario' was regarded as the premium cocoa (Slade 1984). The elevated genetic diversity at this site likely stems from the same source of variation as in the other areas (multiple varieties introduced, followed by outcrossing and regeneration) but why more diversity has been retained here than at the other sites is unknown. Given the high number of private alleles (genetic diversity that is only found in one location) at STC, this site may have simply had introductions of more cocoa varieties or from more geographic sources.

Multiple analytical approaches identify three genetic groups in Samoa cocoa samples

One of the strengths of the STRUCTURE software is its capacity to identify relictual genetic structure. Indeed, from a sea of genetic diversity in the Samoa cocoa samples that shows

almost no geographic pattern, STRUCTURE identified three primary genetic groups that were confirmed to be highly differentiated from one another by several other methods. The subsequent identification of 83 individuals that possess at least 90% of just one of the three genetic backgrounds allowed for the characterization of genetic diversity for each of the three groups. Despite differences in the sample sizes, there is a clear trend of increasing genetic diversity and heterozygosity moving from group I to group III. Our analyses indicate that the remaining individuals (of the 197) are various combinations of the three predominant genetic groups.

Samoa cocoa genetic groups appear to correspond with traditional cocoa varieties

The use of control genotypes from known varieties, Amelonado and Criollo, provided context for the three genetic groups identified here. Group I was found to cluster with the Amelonado genotype. This observation seems logical given the history of frequent planting of this variety in Samoa. The Criollo genotype has a clear affinity with group II individuals in the Neighbor-net analyses but does not nest within those individuals in the PCoA. It remains closest to them but sits just outside that group. The Criollo variety as defined by Motamayor (2008) has quite low genetic variability so it would not appear to be an issue of the representative genotype used. This leads to two hypotheses regarding the nature of the three genetic groups identified here.

Given the placement of the Amelonado and Criollo genotypes in the Neighbor-net analyses and the intermediate position of the group III individuals, it is tempting to interpret the results as follows: group I represents Amelonado/Forasteo, group II represents Criollo, and group III consists of a form of Trinitario. This would seem logical but would suggest that there remains a relatively high frequency of Criollo individuals on farms, which may or may not hold.

When considering the position of Criollo relative to other individuals in the PCoA, another interpretation of the data would suggest that the group II individuals and group III individuals represent different forms of 'Trinitario'. As Trinitario is generated by crossing Forastero and Criollo, multiple Trinitario variants that utilized different genotypes of the parent varieties would have been used. The individuals from different crosses would then have distinct genotypes, which should be discernible through genetic analysis. (Alternatively, while most

individuals within group II may be a form of Trinitario, Samoan plants that mostly closely affiliate with the Criollo representative (e.g., those from the VS site) may indeed be Criollo.). Given the long history of planting Trinitario, including selection for and widespread use of Lafi-7 in Samoa, this interpretation of two genetic groups of Trinitario may be the most likely. Upon brief examination, this scenario also may better match the on-farm identifications of individuals based on morphology.

Distinguishing between these alternative hypotheses will require better control samples for specific cocoa varieties, a clear issue in the current study. This problem is exacerbated by a lack of known pedigree control samples from Samoa specifically. Despite the completion of the contracted work, the Symonds group will continue to work on the issue of better resolving variety affiliations for the Samoa cocoa samples under study.

On-farm identification and genotype data are incongruent but not incompatible

The lack of a clear relationship between inferred cocoa variety IDs supplied by land managers and the genetic data is unfortunate but not completely surprising. With a long history of cocoa variety introductions and open pollination and natural regeneration, it should be expected that a good proportion of cocoa plants in Samoa are of mixed-variety ancestry. Indeed, this is what the microsatellite analyses indicate.

A land manager's interpretation of a given cocoa plant will be determined by individual characters and the aggregate of those characters (e.g., pod smoothness and seed color). It is entirely possible in a cocoa plant with a history of admixture that particular characters derived from one variety exist within a genetic background that is considerably mixed or even mostly of a different variety. Depending on the particular plant and its history, the degree of mixture can vary from a 50:50 mix of two varieties (if derived from recent inter-variety hybridization) to predominantly one variety with a small amount of genetic material from another (if there has been repeated backcrossing in one direction and introgression). Of course, these are the simplest scenarios. It can be much more complicated if there has been involvement from multiple varieties across multiple generations.

For the same reasons, it can be difficult to determine the varieties of cocoa plants based on genotype data. The history of admixture makes for a complicated mosaic of genotypes; however, while the genetic data confirm a fairly complicated history for many of the plants sampled, there is a clear signal in the data that supports three underlying genetic backgrounds. One would expect the individuals that are comprised of mostly one of the three genetic backgrounds to be more likely to match the variety descriptors based on phenotypes; however, this relies on both consistent variety assignment based on genetics and based no phenotypes. While the genetic and trait data are incongruous, they are not incompatible, and it will remain important to use both forms of data during selection of individuals for further propagation.

References

Acebo-Guerrero, Y., Hernández-Rodríguez, A., Heydrich-Pérez, M., El Jaziri, M., & Hernández-Lauzardo, A. N. (2012). Management of black pod rot in cacao (*Theobroma cacao* L.): A review. *Fruits, 67*(1), 41-48. doi:10.1051/fruits/2011065

Adu-Ampomah, Y., Owusu, G. K., Sackey, S., Padi, B., & Abdul-Karimu, A. (1996). Use of gamma rays to induce mutants resistant to cocoa swollen shoot disease in *Theobroma cacao* L. *Plant Breeding, 115*(1), 74-76. doi:10.1111/j.1439-0523.1996.tb00875.x

Afoakwa, E. O. (2014). *Cocoa production and processing technology*. Boca Raton: CRC Press.

Aikpokpodion, P. O., Motamayor, J. C., Adetimirin, V. O., Adu-Ampomah, Y., Ingelbrecht, I., Eskes, A. B., . . . Kolesnikova-Allen, M. (2009). Genetic diversity assessment of sub-samples of cacao, *Theobroma cacao* L. collections in West Africa using simple sequence repeats marker. *Tree Genetics & Genomes, 5*(4), 699-711. doi:10.1007/s11295-009-0221-1

Allakhverdiev, S. I., Kreslavski, V. D., Klimov, V. V., Los, D. A., Carpentier, R., & Mohanty, P. (2008). Heat stress: an overview of molecular responses in photosynthesis. *Photosynthesis Research, 98*(1), 541. doi:10.1007/s11120-008-9331-0

Alvim, P. T. (1977). Cacao. In P. T. Alvim & T. T. Kozlowski (Eds.), *Ecophysiology of tropical crops* (pp. 279-313). London: Academic Press.

Anderson, R. D., & Guest, D. I. (1990). The control of black pod, canker and seedling blight of cocoa, caused by *Phytophthora palmivora,* with potassium phosphonate. *Australasian Plant Pathology, 19*(4), 127-129. doi:10.1071/APP9900127

Argout, X., Martin, G., Droc, G., Fouet, O., Labadie, K., Rivals, E., . . . Lanaud, C. (2017). The cacao Criollo genome v2.0: an improved version of the genome for genetic and functional genomic studies. *BMC Genomics, 18*(730), 1-9. doi:10.1186/s12864-017-4120-9

Argout, X., Salse, J., Aury, J.-M., Guiltinan, M. J., Droc, G., Gouzy, J., . . . Lanaud, C. (2010). The genome of *Theobroma cacao*. *Nature Genetics, 43*(2), 101-109. doi:10.1038/ng.736

Australian Bureau of Meteorology and CSIRO. (2011). *Climate change in the Pacific: Scientific assessment and new research*. Samoa.

Bailey-Serres, J., Lee, S. C., & Brinton, E. (2012). Waterproofing crops: Effective flooding survival strategies. *Plant Physiology, 160*(4), 1698. doi:10.1104/pp.112.208173

Bardakci, F. (2001). Random amplified polymorphic DNA (RAPD) markers. *Turkish Journal of Biology, 25*, 185-196.

Beaudoin, L. (1992). *Feasibility study into the biological control of the rose beetle Adoretus versutus Harold within South Pacific*. Vanuatu: CIRAD-IRCC.

Beaudoin, L., Morin, J. P., Ngyen, C., & Decazy, B. (1995). Study of the undergrounf Adoretus versutus Har. (Col., Scarabaeidae) populations in Vanauatu: Detection of cohabitation with other white grubs. *Journal of Applied Entomology, 119*, 391-397.

Beaudoin, L., Robert, P., & Lal, S. N. (1994). An entomopoxvirus observed in *Adoretus versatus* Harold (Coleoptera; Scarabaeidae; Rutelinae) in Fiji. *International Journal of Pest Management, 40*(1), 66-68. doi:10.1080/09670879409371856

Bertolde, F. Z., Almeida, A.-A. F., & Pirovani, C. P. (2014). Analysis of gene expression and proteomic profiles of clonal genotypes from *Theobroma cacao* subjected to soil flooding. *PLOS ONE, 9*(10). doi:10.1371/journal.pone.0108705

Bertolde, F. Z., Almeida, A.-A. F. D., Corrêa, R. X., Gomes, F. P., Gaiotto, F. A., Baligar, V. C., & Loguercio, L. L. (2010). Molecular, physiological and morphological analysis of waterlogging tolerance in clonal genotypes of *Theobroma cacao* L. *Tree Physiology, 30*(1), 56-67. doi:10.1093/treephys/tpp101

Bertolde, F. Z., Almeida, A. A. F., Pirovani, C. P., Gomes, F. P., Ahnert, D., Baligar, V. C., & Valle, R. R. (2012). Physiological and biochemical responses of *Theobroma cacao* L. genotypes to flooding. *Photosynthetica, 50*(3), 447-457. doi:10.1007/s11099-012-0052-4

Bhattacharjee, R. (2017). Taxonomy and classification of cacao. In P. Umaharan (Ed.), *Achieving sustainable cultivation of cocoa : Genetics, breeding, cultivation and quality*: Cambridge : Burleigh Dodds Science Publishing, 2017.

Bourke, T. V. (1992). Pests and diseases of cocoa in Western Samoa and the Philippines. In P. J. Keane & C. A. J. Putter (Eds.), *Cocoa pest and disease management in Southeast Asia and Australasia* (pp. 195-198). Rome: Food and Agriculture Organisation of the United Nations (FAO).

Brasier, C. M., Griffin, M. J., & Maddison, A. C. (1981). The cocoa black pod phytphthoras. In P. H. Gregory & A. C. Maddison (Eds.), *Epidemiology of Phytophthora on cocoa in Nigeria*. Kew, England: Commonwealth Mycological Institute.

Brown, J. S., Schnell, R. J., Motamayor, J. C., Uilson, L., Kuhn, D. N., & Borrone, J. W. (2005). Resistance gene mapping for witches' broom disease in *Theobroma cacao* L. in an F2 population using SSR markers and candidate genes. *Journal of the American Society for Horticultural Science jashs, 130*(3), 366-373. doi:10.21273/JASHS.130.3.366

Brownstein, M. J., Carpten, J. D., & Smith, J. R. (1996). Modulation of non-templated nucleotide addition by Taq DNA polymerase: Primer modifications that facilitate genotyping. *BioTechniques, 20*(6), 1004-1010. doi:10.2144/96206st01

Brunt, A. A. (2008). The effects of cocoa swollen-shoot virus on the growth and yield of Amelonado and Amazon cocoa (*Theobroma cacao*) in Ghana. *Annals of Applied Biology, 80*(2), 169-180. doi:10.1111/j.1744-7348.1975.tb01618.x

Budak, H., Bölek, Y., Dokuyucu, T., & Akkaya, A. (2004). Potential uses of molecular markers in crop improvement. *KSU Journal of Science and Engineering, 7*(1), 75-79.

Buddenhagen, I. W. (1977). Resistance and vulnerability of tropical crops in relation to their evolution and breeding. *Annals of the New York Academy of Sciences, 287*(1), 309-326. doi:10.1111/j.1749-6632.1977.tb34249.x

Carr, M. K. V., & Lockwood, G. (2011). The water relations and irrigation requirements of cocoa (*Theobroma cacao* l.): A review. *Experimental Agriculture, 47*(4), 653-676. doi:10.1017/S0014479711000421

Cocoa Samoa Ltd. (CSL). (2015). *Reviving the cocoa industry in Samoa 2016*. Retrieved from http://www.academia.edu/31472204/Reviving_the_Cocoa_Industry_in_Samoa_2016

Codjoe, F. N. Y., Ocansey, C. K., Boateng, D. O., & Ofori, J. (2013). Climate change awareness and coping strategies of cocoa farmers in rural Ghana. *Journal of Biology, Agriculture and Healthcare, 3*(11), 19-29.

de Almeida, A.-A. F., & Valle, Raúl R. (2007). Ecophysiology of the cacao tree. *Brazilian Journal of Plant Physiology, 19*(4), 425-448. doi:10.1590/S1677-04202007000400011

De Giovanni, C., Dell'Orco, P., Bruno, A., Ciccarese, F., Lotti, C., & Ricciardi, L. (2004). Identification of PCR-based markers (RAPD, AFLP) linked to a novel powdery mildew resistance gene (ol-2) in tomato. *Plant Science, 166*(1), 41-48. doi:10.1016/j.plantsci.2003.07.005

de Schawe, C. C., Durka, W., Tscharntke, T., Hensen, I., & Kessler, M. (2013). Gene flow and genetic diversity in cultivated and wild cacao (*Theobroma cacao*) in Bolivia. *American Journal of Botany, 100*(11), 2271-2279. doi:10.3732/ajb.1300025

De Silva, C., Chandani, H., & Takashi, A. (2017). Effects of heat stress on growth, photosynthetic pigments, oxidative damage and competitive capacity of three submerged macrophytes. *Journal of Plant Interactions, 12*(1), 228-236. doi:10.1080/17429145.2017.1322153

Dhawan, A. K., & Peshin, R. (2009). Integrated pest management: Concept, opportunities and challenges. In A. K. Dhawan & R. Peshin (Eds.), *Integrated pest management: Inoovation-development process*. India: Springer Science.

Dillon, N., Hucks, L., Diczbalis, Y., & Kete, T. (2014). *Evaluation of molecular marker technology for the identification of elite cocoa germplasm in the South Pacific*. Australia: ACIAR.

Domfeh, O., Ameyaw, G. A., Dzahini-Obiatey, H. K., Ollennu, L. A. A., Osei-Bonsu, K., Acheampong, K., . . . Owusu-Ansah, F. (2016). Use of immune crops as barrier in the management of cacao swollen shoot virus disease (CSSVD) - Long-term assessment. *Plant Disease, 100*(9), 1889-1893. doi:10.1094/PDIS-03-16-0404-RE

Droessler, H. (2017). *Colonialism by deferral: Samoa under the tridominium, 1889–1899*: Emerald Publishing Limited

Dzahini-Obiatey, H., Ameyaw, G. A., & Ollennu, L. A. (2006). Control of cocoa swollen shoot disease by eradicating infected trees in Ghana: A survey of treated and replanted areas. *Crop Protection, 25*, 647-652. doi:10.1016/j.cropro.2005.09.004

Dzahini-Obiatey, H., Domfeh, O., & Amoah, F. M. (2010). Over seventy years of a viral disease of cocoa in Ghana: From researchers' perspective. *African Journal of Agricultural Research, 5*(7), 476-485. doi:10.5897/AJAR09.625

Eden, D. R. A., & Edwards, W. L. (1952). *Cocoa plantation management in Western Samoa*. Noumea: South Pacific Commission.

Emmanuel, B., & Qineti, A. (2018). *Cocoa production and export in Ghana*. Paper presented at the 162nd EAAE Seminar: The evaluation of new CAP instruments: Lessons learned and the road ahead, Budapest, Hungary.

Erwin, D. C., & Ribeiro, O. K. (1996). *Phytophthora diseases worldwide*. St. Paul, Minn: APS Press.

Esan, E. B. (1992). Micropropagation of cocoa (*Theobroma cacao* L.). In Y. P. S. Bajaj (Ed.), *High-Tech and Micropropagation II* (pp. 96-122). Berlin, Heidelberg: Springer Berlin Heidelberg.

Essola, E. E. J., Gusua, C. R., Tchapda, T. D., & Andre, O. N. P. . (2017). Vegetative propagation of selected clones of cocoa (*Theobroma cacao* L.) by stem cuttings. *Journal of Horticulture and Forestry, 9*(9), 80-90. doi:10.5897/JHF2017.0502

Everaert, H., Rottiers, H., Pham, P. H. D., Ha, L. T. V., Nguyen, T. P. D., Tran, P. D., . . . Messens, K. (2017). Molecular characterization of Vietnamese cocoa genotypes (*Theobroma cacao* L.) using microsatellite markers. *Tree Genetics & Genomes, 13*(5), 99. doi:10.1007/s11295-017-1180-6

FAO. (2018). Geographic and economic groups. Retrieved November 11, 2018 http://www.fao.org/countryprofiles/geographic-and-economic-groups/en/

FAOSTAT. (2018). Production/yield quantities of cacao in world. FAOSTAT Retrieved September 21, 2018 http://www.fao.org/faostat/

Farrell, A. D., Rhiney, K., Eitzinger, A., & Umaharan, P. (2018). Climate adaptation in a minor crop species: is the cocoa breeding network prepared for climate change? *Agroecology and Sustainable Food Systems, 42*(7), 812-833. doi:10.1080/21683565.2018.1448924

Firman, I. D., & Vernon, A. J. (1970). Cocoa canker caused by Phytophthora palmivora. *Annals of Applied Biology, 65*, 65-73.

Fleming, E., & Blowes, A. (2003). *Export performance in South Pacific Countries marginally endowed with natural resources: Samoa and Tonga, 1960 to 1999 (Working Paper No. 2003-8)*. Retrieved from University of New England, Graduate School of Agricultural and Resource Economics & School of Economics website http://ageconsearch.umn.edu/bitstream/12942/1/wp030008.pdf

Galindo, J. J. (1992). Prospects for biological control of black pod of cacao. In P. J. Keane & C. A. J. Putter (Eds.), *Cocoa pest and disease management in Southeast Asia and Australasia* (pp. 31-36). Rome: Food and Agriculture Organisation of the United Nations (FAO).

George, A. A., Henry, K. D.-O., Owusu, D., Francis, K. O., & Kwaku, A.-E. (2015). History and data analyses of 'cutting out' method for Cocoa swollen shoot virus disease (CSSVD) control in Ghana. *Journal of Plant Diseases and Protection, 122*(5), 200-206.

Gosh, S. K., Sarkar, C. K. G., & Datta, S. (2001). Molecular markers for cultivar identification and PBR. *Journal of Interllectual Property Rights, 6*, 377-388.

Guan, R., Chen, J., Jiang, J., Liu, G., Liu, Y., Tian, L., . . . Qiu, L.-j. (2014). Mapping and validation of a dominant salt tolerance gene in the cultivated soybean (*Glycine max*) variety Tiefeng 8. *The Crop Journal, 2*(6), 358-365. doi:10.1016/j.cj.2014.09.001

Guest, D. (2007). Black pod: Diverse pathogens with a global impact on cocoa yield. *Phytopathology, 97*(12), 1650-1653. doi:10.1094/PHYTO-97-12-1650

Hafield, J. L., Boote, K. J., Fay, P., Hahn, L., Izzauralde, R. C., Kimball, B. A., . . . Wolfe, D. W. (2008). Agriculture. In *The effects of climate change on agriculture, land resources, water resources and biodiversity in the United States*.

Hernández-Rodríguez, A., Ruíz-Beltrán, Y., Acebo-Guerrero, Y., Miguélez-Sierra, Y., & Heydrich-Pérez, M. (2014). Antagonistas microbianos para el manejo de la pudrición negra del fruto en *Theobroma cacao* L: Estado actual y perspectivas de uso en Cuba. *Revista de Protección Vegetal, 29*(1), 11-19.

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nature reviews. Genetics, 10*(9), 639-650. doi:10.1038/nrg2611

Huang, N., Parco, A., Mew, T., Magpantay, G., McCouch, S., Guiderdoni, E., . . . Khush, G. S. (1997). RFLP mapping of isozymes, RAPD and QTLs for grain shape, brown planthopper resistance in a doubled haploid rice population. *Molecular Breeding, 3*(2), 105-113. doi:10.1023/A:1009683603862

Huson, D. H., & Bryant, D. (2005). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution, 23*(2), 254-267. doi:10.1093/molbev/msj030

Jiang, G. L. (2013). Molecular markers and marker-assisted breeding in plants. In S. B. Andersen (Ed.), *Plant breeding from laboratories to field* (pp. 45-83): InTechOpen.

Keane, P. J. (1992). Diseases and pests of cocoa: An overview. In P. J. Keane & C. A. J. Putter (Eds.), *Cocoa pest and disese management in Southeast Asia and Australia* (pp. 1-12). Rome: Food and Agriculture Organisation of the United Nations (FAO).

Kellam, M. K., & Zentmyer, G. A. (1986). Morphological, physiological, ecological and pathological comparisons of Phytophthora species isolated from *Theobroma cacao*. *Phytopathology, 76*(2), 159-164.

Kenten, R. H., & Lockwood, G. (1977). Studies on the possibility of increasing resistance to cocoa swollen-shoot virus by breeding. *Annals of Applied Biology, 85*(1), 71-78. doi:10.1111/j.1744-7348.1977.tb00631.x

Kenten, R. H., & Woods, R. D. (1976). A virus of the cocoa swollen shoot group infecting cocoa in North Sumatra. *PANS, 22*, 488-490. doi:10.1080/09670877609414338

Korir, N. K., Han, J., Shangguan, L., Wang, C., Kayesh, E., Zhang, Y., & Fang, J. (2013). Plant variety and cultivar identification: advances and prospects. *Critical Reviews in Biotechnology, 33*(2), 111-125. doi:10.3109/07388551.2012.675314

Laderach, P., Eitzinger, A., Martinez, A., & Castro, N. (2011). *Predicting the impact of climate change on the cocoa-growing regions in Ghana and Cote d'Ivoire*. CIAT. Retrieved from https://www.eenews.net/assets/2011/10/03/document_cw_01.pdf

Läderach, P., Martinez-Valle, A., Schroth, G., & Castro, N. (2013). Predicting the future climatic suitability for cocoa farming of the world's leading producer countries, Ghana and Côte d'Ivoire. *Climatic Change, 119*(3), 841-854. doi:10.1007/s10584-013-0774-8

Lanaud, C., Risterucci, A. M., Pieretti, I., Falque, M., Bouet, A., & Lagoda, P. J. L. (1999). Isolation and characterization of microsatellites in *Theobroma cacao* L. *Molecular Ecology, 8*(12), 2141-2143. doi:10.1046/j.1365-294x.1999.00802.x

Laurent, V., Risterucci, A. M., & Lanaud, C. (1994). Genetic diversity in cocoa revealed by cDNA probes. *Theoretical and Applied Genetics, 88*(2), 193-198. doi:10.1007/BF00225897

Leong, C. C., Hai, T. C., & Beng, A. B. (1977). Control of cocoa stem canker caused by *Phytophthora palmivora* (Butler), Butler. *MARDI Research Bulletin, 5*(2), 21-29.

Loor Solorzano, R. G., Fouet, O., Lemainque, A., Pavek, S., Boccara, M., Argout, X., . . . Lanaud, C. (2012). Insight into the wild origin, migration and domestication history of the fine flavour Nacional *Theobroma cacao* L. variety from Ecuador. *PLOS ONE, 7*(11). doi:10.1371/journal.pone.0048438

Mackay, J. F., Wright, C. D., & Bonfiglioli, R. G. (2008). A new approach to varietal identification in plants by microsatellite high resolution melting analysis: Application to the verification of grapevine and olive cultivars. *Plant methods, 4*(8), 1-10. doi:10.1186/1746-4811-4-8

McGregor, A. J. (1984). Comparison of cuprous oxide and metalaxyl with mixtures of these fungicides for thecontrol of *Phytophthora* pod rot of cocoa. *Plant Pathology, 33*(1), 81-87. doi:10.1111/j.1365-3059.1984.tb00591.x

McMahon, P. J., Purwantara, A., Wahab, A., Imron, M., Lambert, S., Keane, P. J., & Guest, D. I. (2010). Phosphonate applied by trunk injection controls stem canker and decreases Phytophthora pod rot (black pod) incidence in cocoa in Sulawesi. *Australasian Plant Pathology, 39*(2), 170-175. doi:10.1071/AP09078

Medina, V., & Laliberte, B. (2017). *A review of research on the effects of drought and temperature stress and increased CO2 on Theobroma cacao L., and the role of genetic diversity to address climate change*. Costa Rica: Biodiversity International.

Motamayor, J. C., Lachenaud, P., da Silva e Mota, J. W., Loor, R., Kuhn, D. N., Brown, J. S., & Schnell, R. J. (2008). Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L). *PLOS ONE, 3*(10), e3311. doi:10.1371/journal.pone.0003311

Motamayor, J. C., Mockaitis, K., Schmutz, J., Haiminen, N., Livingstone, D., 3rd, Cornejo, O., . . . Kuhn, D. N. (2013). The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome biology, 14*(6), 1-24. doi:10.1186/gb-2013-14-6-r53

Muller, E. (2008). Cacao swollen shoot virus. In P. Rao Covind, Paul Khurana, S.M., Lenardon Sergio, L (Ed.), *Characterization, diagnosis and management of plant viruses*. USA: Studium Press.

N'Goran, J. A. K., Laurent, V., Risterucci, A. M., & Lanaud, C. (1994). Comparative genetic diversity studies of *Theobroma cacao* L. using RFLP and RAPD markers. *Heredity, 73*, 589. doi:10.1038/hdy.1994.166

Nyadanu, D., Akromah, R., Adomako, B., Dzahini-Obiatey, H., Akrofi, A. Y., Lowor, S. T., & Assuah, M. K. (2012). Breeding for multiple disease resistance in cocoa (T*heobroma cacao* L.). *International Journal of Plant Breeding and Genetics, 6*, 182-194. doi:10.3923/ijpbg.2012.182.194

Ofori, A., Padi, F. K., Ameyaw, G. A., Dadzie, A. M., Lowor, S., & Flachowsky, H. (2015). Genetic variation among cocoa (*Theobroma cacao* L.) families for resistance to cocoa swollen shoot virus disease in relation to total phenolic content. *Plant Breeding, 134*(4), 477-484. doi:10.1111/pbr.12282

Okaisabor, E. K. (1968). An active fungal parasite of *Phytophthora palmivora*. *Pest Articles & News Summaries Section B. Plant Disease Control, 14*(4), 353-360. doi:10.1080/05331846809432313

Opoku, I. Y., Akrofi, A. Y., & Appiah, A. A. (2006). Assessment of sanitation and fungicide application directed at cocoa tree trunks for the control of *Phytophthora* black pod infections in pods growing in the canopy. *European Journal of Plant Pathology, 117*(2), 167. doi:10.1007/s10658-006-9082-8

Oyedokun, F. O., & Oyelana, A. A. (2016). Effects of climate change on cocoa farmers' production. *Journal of Human Ecology, 53*(3), 233-244. doi:10.1080/09709274.2016.11906976

Padi, F. K., Domfeh, O., Takrama, J., & Opoku, S. (2013). An evaluation of gains in breeding for resistance to the cocoa swollen shoot virus disease in Ghana. *Crop Protection, 51*, 24-31. doi:10.1016/j.cropro.2013.04.007

Peakall, R. O. D., & Smouse, P. E. (2005). GenAlEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes, 6*(1), 288-295. doi:10.1111/j.1471-8286.2005.01155.x

Pereira, G. d. S., Cazé, A. L. R., Silva, M. G. d., Almeida, V. C., Magalhães, F. O. d. C., Silva Filho, J. L. d., . . . Hoffmann, L. V. (2015). Optimal use of SSR markers for varietal identification of upland cotton. *Pesquisa Agropecuária Brasileira, 50*(7), 571-581. doi:10.1590/S0100-204X2015000700007

Pereira, J. L. (1992). Cocoa and its pathogens in the region of origin: A continued risk. In P. J. Keane & C. A. J. Putter (Eds.), *Cocoa pest and disease management in Southest Asia and Australasia* (pp. 13-30). Rome: Food and Agriculture Organisation of the United Nations (FAO).

PHAMA. (2016). *Pacific cocoa export industry*. AECOM. Retrieved from http://phama.com.au/

Pohlan, H. A. J., & Pérez, V. D. (2010). Growth and production of cacao. In *Soils, plant growth and crop production* (Vol. III).

Posnette, A. F. (1947). Virus diseases of cacao in West Africa: I. Cacao viruses 1A, 1B, 1C and 1D. *Annals of Applied Biology, 34*(3), 388-402. doi:10.1111/j.1744-7348.1947.tb06372.x

Pourabed, E., Jazayeri Noushabadi, M. R., Jamali, S. H., Moheb Alipour, N., Zareyan, A., & Sadeghi, L. (2015). Identification and DUS testing of rice varieties through microsatellite markers. *International Journal of Plant Genomics, 2015*, 1-7. doi:10.1155/2015/965073

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945-959.

Pugh, T., Fouet, O., Risterucci, A. M., Brottier, P., Abouladze, M., Deletrez, C., . . . Lanaud, C. (2004). A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers. *Theoretical and Applied Genetics, 108*(6), 1151-1161. doi:10.1007/s00122-003-1533-4

Quainoo, A. K., Wetten, A. C., & Allainguillaume, J. (2008). The effectiveness of somatic embryogenesis in eliminating the cocoa swollen shoot virus from infected cocoa trees. *Journal of Virological Methods, 149*(1), 91-96. doi:10.1016/j.jviromet.2008.01.007

R Core Team. (2013). R: A Language and Environment for Statistical Computing. In.

Rasmussen, A., & McGoldrick, W. (1992). *Samoa's second national communication to the United Nations framework convention on climate change*. Samoa: Ministry of Natural Resources & Environment.

Rehm, S., & Espig, G. (1991). *The cultivated plants of the tropics and subtropics : Cultivation, economic value, utilization*. Weikersheim, Germany: Margraf.

Richards, D. A. (1948). Notes on the vegetative propagation of cacao by cuttings. *Journal of Horticultural Science, 24*(3), 192-199. doi:10.1080/00221589.1948.11513688

Risterucci, A. M., Grivet, L., N'Goran, J. A. K., Pieretti, I., Flament, M. H., & Lanaud, C. (2000). A high-density linkage map of *Theobroma cacao* L. *Theoretical and Applied Genetics, 101*(5), 948-955. doi:10.1007/s001220051566

Rosa, P. M., Campos, T. d., Sousa, A. C. B. d., Sforça, D. A., Torres, G. A. M., & Souza, A. P. d. (2010). Potato cultivar identification using molecular markers. *Pesquisa Agropecuária Brasileira, 45*(1), 110-113. doi:10.1590/S0100-204X2010000100015

Sale, P. J. M. (1969). Flowering of cacao under controlled temperature conditions. *Journal of Horticultural Science, 44*(2), 163-173. doi:10.1080/00221589.1969.11514294

Samoa Bureau of Statistics. (2016). *Report on Samoa Agricultural Survey*. Samoa: Samoa Bureau of Statistics.

Samoa Ministry of Agriculture Forests Fisheries and Meteorology. (2004). *The feasibility of reviving cocoa as a major industry : Strategic overview*. [Samoa]: Ministry of Agriculture, Forests, Fisheries & Meteorology.

Saunders, J. A., Mischke, S., Leamy, E. A., & Hemeida, A. A. (2004). Selection of international molecular standards for DNA fingerprinting of *Theobroma cacao*. *Theoretical and Applied Genetics, 110*(1), 41-47. doi:10.1007/s00122-004-1762-1

Schroth, G., Läderach, P., Martinez-Valle, A. I., Bunn, C., & Jassogne, L. (2016). Vulnerability to climate change of cocoa in West Africa: Patterns, opportunities and limits to adaptation. *Science of The Total Environment, 556*, 231-241. doi:10.1016/j.scitotenv.2016.03.024

Sena Gomes, A. R., & Kozlowski, T. T. (1986). The effects of flooding on water relations and growth of *Theobroma cacao* var. catongo seedlings. *Journal of Horticultural Science, 61*(2), 265-276. doi:10.1080/14620316.1986.11515700

Sesaga, T. S. (2014). *Samoa's 5th national report 2014 to the Convention on Biological Diversity*. Samoa: MNRE.

Shultes, R. E. (1958). A synopsis of the genus (*Herrania*). *Journal of the Arnold Arboreum, 39*, 217-275.

Simmonds, N. W. (1998). Tropical crops and their improvement. In C. C. Webster & P. N. Wilson (Eds.), *Agriculture in the Tropics* (3rd ed., pp. 257–293).

Singh, P. K., & Bhati, J. P. (2010). Development of agrientrepreneurship in Samoa: Opportunities and challenges. *Journal of South Pacific Agriculture, 14*(1), 6-13.

Slade, D. A. (1984). A review of small scale cocoa production in Western Samoa. *Alafua Agricultural Bulletin, 9*(1), 30-39.

Smith Dumont, E., Gnahoua, G. M., Ohouo, L., Sinclair, F. L., & Vaast, P. (2014). Farmers in Côte d'Ivoire value integrating tree diversity in cocoa for the provision of ecosystem services. *Agroforestry Systems, 88*(6), 1047-1066. doi:10.1007/s10457-014-9679-4

Sodré, G. A., & Gomes, A. R. S. (2019). Cocoa propagation, technologies for production of seedlings. *Revista Brasileira de Fruticultura, 41*(2). doi:10.1590/0100-29452019782

Solorzano, R. G. L., Fouet, O., Lemainque, A., Pavek, S., Boccara, M., Argout, X., . . . Lanaud, C. (2012). Insight into the wild origin, migration and domestication history of the fine flavour Nacional *Theobroma cacao* L. variety from Ecuador. *PLOS ONE, 7*(11), e48438. doi:10.1371/journal.pone.0048438

Stolberg, D. (2013). German in Samoa: Historical traces of a colonial variety. *Poznan Studies in Contemporary Linguistics, 49*(3), 321-353. doi:10.1515/psicl-2013-0012

Symonds, V. V., & Lloyd, A. M. (2004). A simple and inexpensive method for producing fluorescently labelled size standard. *Molecular Ecology Notes, 4*(4), 768-771. doi:10.1111/j.1471-8286.2004.00763.x

Tinsley, T. W. (1971). The ecology of cacao viruses. I. The role of wild hosts in the incidence of swollen shoot virus in West Africa. *Journal of Applied Ecology, 8*(2), 491-495. doi:10.2307/2402885

Toxopeus, H. (1985). Botany, types and populations. In G. A. R. Wood & R. A. Lass (Eds.), *Cocoa* (4th ed., pp. 80-92). New York: Longman Group Limited.

United Nations. (1993). *International Cocoa Agreement 1993*, Geneva.

Urquhart, D. H. (1952). *Cocoa growing in Western Samoa: A report on a survey made for the South Pacific Commission from 16th to 30th April, 1952*. Noumea: South Pacific Commission.

Urquhart, D. H. (1961). *Cocoa* (2nd ed.). London: Longmans.

van der Kooji, S. (2013). *Market study of fine flavour cocoa in 11 selected countries*. Amsterdam: Royal Tropical Institute.

Waraich, E. A., Ahmad, R., Halim, A., & Aziz, T. (2012). Alleviation of temperature stress by nutrient management in crop plants: a review. *Journal of soil science and plant nutrition, 12*(2), 221-244.

Ward, R. G., & Ashcroft, P. (1998). *Samoa : Mapping the diversity*: Institute of Pacific Studies, University of the South Pacific.

Wessel, M., & Quist-Wessel, P. M. F. (2015). Cocoa production in West Africa, a review and analysis of recent developments. *NJAS - Wageningen Journal of Life Sciences, 74*, 1-7. doi:10.1016/j.njas.2015.09.001

Western Samoa Department of Agriculture Forests, a. F. (2005). *The cocoa diseases ordinance 1961*. Western Samoa: University of South Pacific. Retrieved from www.extwprlegs1.fao.org/

Wiah, E. N., & Twumasi-Ankrah, S. (2017). Impact of climate change on cocoa yield in ghana using vector autoregressive model. *Ghana Journal of Technology, 1*(2), 32-39.

Wiesman, Z., & Jaenicke, H. (2002). Introduction to vegetative tree propagation: concepts and principles. In H. Jaenicke & J. Beniest (Eds.), *Vegetative tree propagation in agroforestry: Training guidelines and references* (pp. 148). Nairobi, Kenya: International Centre for Research in Agroforestry (ICRAF).

Willson, K. C. (1999). *Coffee, cocoa and tea*. New York: CAB International.

Wood, G. A. R. (1985). Cocoa. In G. A. R. Wood & R. A. Lass (Eds.), *Cocoa* (4th ed.). New York: Longman Group Limited.

Wright, S. (1968). *Evolution and the genetics of populations; a treatise* (Vol. 4). Chicago: University of Chicago Press.

Young, A. M. (1994). *The chocolate tree*. Washington: Smithsonian Institution Press.

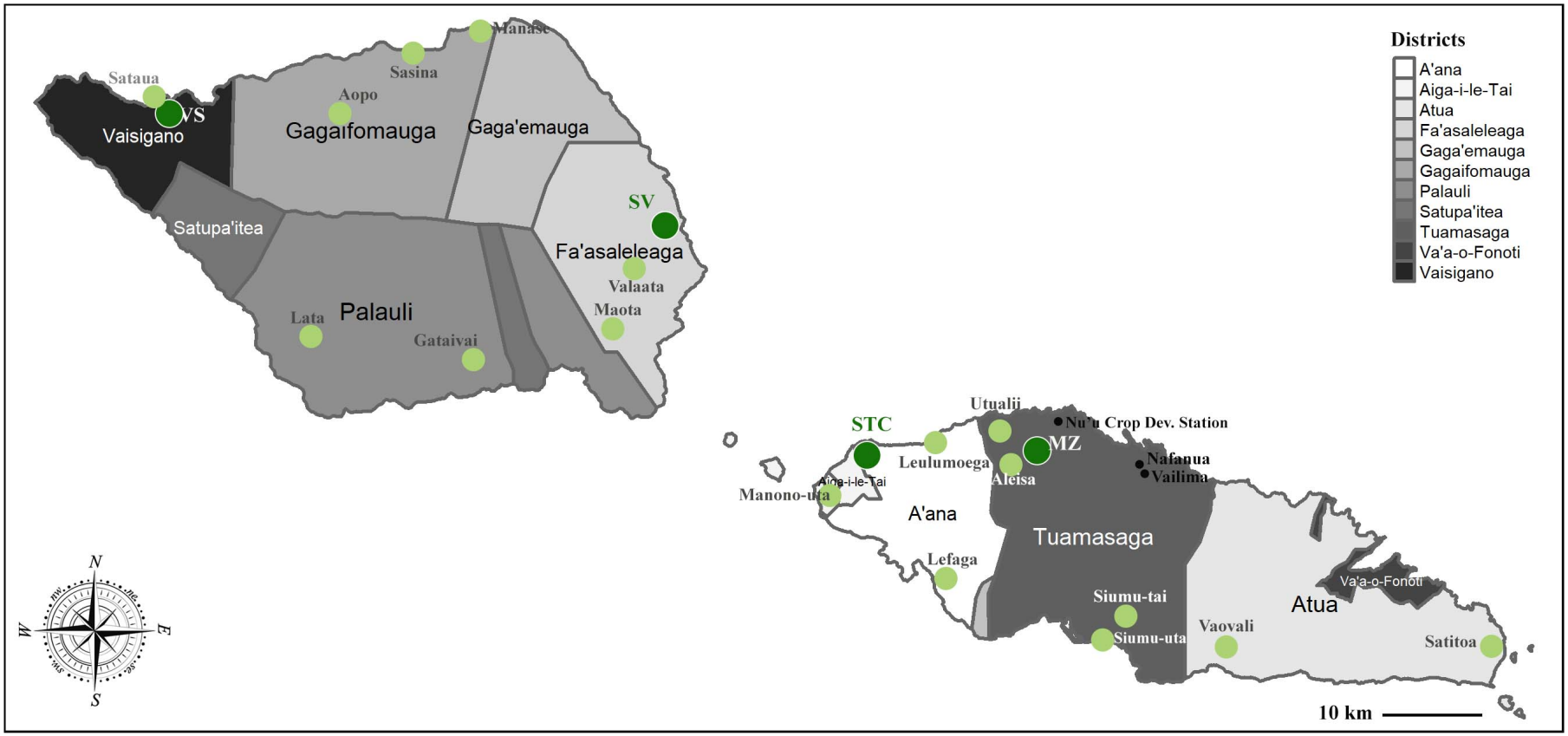
Zaiger, D., & Zentmyer, G. A. (1965). Phytophthora canker of cacao in the Caroline Islands. *Plant Disease Reporter, 49*(7), 565-567.

Zhang, D., Martínez, W. J., Johnson, E. S., Somarriba, E., Phillips-Mora, W., Astorga, C., . . . Meinhardt, L. W. (2012). Genetic diversity and spatial structure in a new distinct *Theobroma cacao* L. population in Bolivia. *Genetic Resources and Crop Evolution, 59*(2), 239-252. doi:10.1007/s10722-011-9680-y

Zhang, D., Mischke, S., Johnson, E. S., Phillips-Mora, W., & Meinhardt, L. (2009). Molecular characterization of an international cacao collection using microsatellite markers. *Tree Genetics & Genomes, 5*(1), 1-10. doi:10.1007/s11295-008-0163-z

Zhang, F., Chen, S., Chen, F., Fang, W., & Li, F. (2010). A preliminary genetic linkage map of chrysanthemum (*Chrysanthemum morifolium*) cultivars using RAPD, ISSR and AFLP markers. *Scientia Horticulturae, 125*(3), 422-428. doi:10.1016/j.scienta.2010.03.028

Zohary, D. (2004). Unconscious selection and the evolution of domesticated plants. *Economic Botany, 58*(1), 5-10.



Supplementary Figure 1. Distribution of the collection sites for the current study (dark green dots), demo plots (light green dots), the classic clones and hybrid plant materials (black dots) imported and established in Samoa.

Supplementary Table 1. Genetic parameters and primer details for *Theobroma cacao* L. screened microsatellite loci.

Marker Name	EMBL Acc. Number / Locus	Primer sequence (5'–3')	Chr	T_a(°C)	Size (bp)	Repeat structure	N_a	H_O	H_E	Screened/ selected				Source
										1	2	3	4	
mTcCIR1*	Y16883	GCAGGGCAGGCTCAGTGAAGCA TGGGCAACCAGAAAACGAT	8	51	143	$(CT)_{14}$	3	0.50	0.62	*	*			1, 2, 3
mTcCIR2	Y16978	CAGGGAGCTGTGTTATTGGTCA AGTTATTGTCGGCAAGGAGGAT	5	51	254	$(GA)_3\,N_5\,(AG)_2\,GG\,(AG)_4$	3	0.55	0.51	*				1
mTcCIR6*	Y16980	TTCCCTCTAAACTACCCTAAAT TAAAGCAAAGCAATCTAACATA	6	46	231	$(TG)_7\,(GA)_{13}$	8, 7	0.54	0.57	*	*		*	1, 2, 3, 4
mTcCIR7*	Y16981	ATGCGAATGACAACTGGT GCTTTCAGTCCTTTGCTT	7	51	160	$(GA)_{11}$	6, 4	0.42	0.75	*		*		1, 2, 3
mTcCIR8*	Y16982	CTAGTTTCCCATTTACCA TCCTCAGCATTTTCTTTC	9	46	301	$(TC)_5\,TT\,(TC)_{17}\,TTT\,(CT)_4$	5, 6	0.33	0.55	*	*	*		1, 2, 3
mTcCIR10	Y16984	ACAGATGGCCTACACACT CAAGCAAGCCTCATACTC	5	46	208	$(TG)_{13}$	4	0.56	0.71	*				1
mTcCIR11*	Y16985	TTTGGTGATTATTAGCAG GATTCGATTTGATGTGAG	2	46	298	$(TC)_{13}$	9, 11	0.46	0.81	*	*			1, 2, 3
mTcCIR12*	Y16986	TCTGACCCCAAACCTGTA ATTCCAGTTAAAGCACAT	4	46	188	$(CATA)_4\,N_{18}\,(TG)_6$	10, 11	0.62	0.87	*	*	*	*	1, 2, 3, 4
mTcCIR15*	Y16988	CAGCCGCCTCTTGTTAG TATTTGGGATTCTTGATG	1	46	254	$(TC)_{19}$	10, 11, 13	0.62	0.84	*	*	*	*	1, 2, 3, 4
mTcCIR18*	Y16991	GATAGCTAAGGGGATTGAGGA GGTAATTCAATCATTTGAGGATA	4	51	345	$(GA)_{12}$	8	0.46	0.72	*	*	*	*	1, 2, 3, 4
mTcCIR22*	Y16995	ATTCTCGCAAAAACTTAG GATGGAAGGAGTGTAAATAG	1	46	289	$(TC)_{12}\,N_{146}\,(CT)_{10}$	4, 8	0.29	0.43	*		*		1, 2, 3
mTcCIR24*	Y16996	TTTGGGGTGATTTCTTCTGA TCTGTCTCGTCTTTTGGTGA	9	46	198	$(AG)_{13}$	4, 5, 9	0.35	0.31	*		*	*	1, 2, 3, 4
mTcCIR26*	Y16998	GCATTCATCAATACATTC GCACTCAAAGTTCATACTAC	8	46	298	$(TC)_9\,C\,(CT)_4\,TT\,(CT)_{11}$	6, 8, 12	0.41	0.67	*	*	*	*	1, 2, 3, 4
mTcCIR33*	AJ271826	TGGGTTGAAGATTTGGT	4	51	265–348	$(TG)_{11}$	12	0.82	0.69		*			2, 3

		CAACAATGAAAATAGGCA										
mTcCIR37*	AJ271942	CTGGGTGCTGATAGATAA AATACCCTCCACACAAAT	10	46	136–187	$(GT)_{15}$	13	0.83	0.73	*	*	2, 3
mTcCIR40*	AJ271943	AATCCGACAGTCTTTAATC CCTAGGCCAGAGAATTGA	3	51	262–288	$(AC)_{15}$	10	0.71	0.65	*		2, 3
mTcCIR60*	AJ271958	CGCTACTAACAAACATCAAA AGAGCAACCATCACTAATCA	2	51	190–218	$(CT)_7(CA)_{20}$	9	0.80	0.80	*	*	2, 3

Genetic Parameters: N_a number of alleles, H_o observed heterozygosity, H_e expected heterozygosity;

Sources: (1) Lanaud et al. (1999); (2) Saunders et al. (2004); (3) Everaert et al. (2017); (4) Aikpokpodion et al. (2009);

* Screened and selected primers (from the cited literatures) with high quality profiles and polymorphism.

Supplementary Table 2. Summary statistics and genetic parameters of the SSR markersused to genotype the Cocoa individuals using the *site groups* and using the inferred *inferred genetic groups* from the STRUCTURE analyses.

Loci (n=197)	Allele		Ho	He	F_{is}	F_{st}	N_m	PIC
	Na	Ne						
mTcCIR 01	3	1.82	0.44	0.44	0.01	0.04	5.53	0.46
mTcCIR 12	4	2.06	0.52	0.50	-0.04	0.02	12.21	0.51
mTcCIR 15	7	2.59	0.57	0.60	0.05	0.01	20.40	0.61
mTcCIR 07	7	2.40	0.11	0.58	0.81	0.01	23.26	0.59
mTcCIR 06	9	2.27	0.42	0.55	0.24	0.02	9.86	0.56
mTcCIR 26	4	1.43	0.27	0.30	0.07	0.02	14.23	0.30
mTcCIR 11	7	1.93	0.51	0.47	-0.09	0.01	18.29	0.48
mTcCIR 24	2	1.53	0.29	0.35	0.15	0.00	62.81	0.35
mTcCIR 37	13	2.29	0.52	0.55	0.05	0.02	14.39	0.56
mTcCIR 22	5	1.86	0.48	0.46	-0.05	0.05	5.22	0.48
mTcCIR 33	8	1.85	0.42	0.45	0.07	0.02	14.41	0.46
mTcCIR 60	9	2.22	0.54	0.54	0.00	0.03	7.90	0.56
mTcCIR 08	4	1.75	0.33	0.42	0.21	0.06	4.16	0.45
mTcCIR 18	7	1.91	0.45	0.47	0.06	0.02	10.74	0.48
mTcCIR 40	7	2.06	0.53	0.51	-0.03	0.06	4.24	0.54
Total	96							
Mean	7	2.00	0.43	0.48	0.10	0.03	15.18	0.49

Loci *K*=3 (n=83)	Allele		Ho	He	F_{is}	F_{st}	N_m	PIC
	Na	Ne						
mTcCIR 01	2	1.71	0.37	0.39	0.07	0.15	1.47	0.45
mTcCIR 12	4	1.79	0.45	0.43	-0.03	0.07	3.59	0.45
mTcCIR 15	5	2.98	0.70	0.62	-0.13	0.08	2.74	0.60
mTcCIR 07	5	1.42	0.13	0.18	0.32	0.73	0.09	0.57
mTcCIR 06	7	2.42	0.54	0.56	0.02	0.13	1.64	0.55
mTcCIR 26	2	1.37	0.30	0.27	-0.14	0.01	32.43	0.26
mTcCIR 11	6	2.38	0.51	0.47	-0.07	0.13	1.65	0.43
mTcCIR 24	2	1.47	0.28	0.30	0.08	0.05	5.01	0.33
mTcCIR 37	8	3.03	0.54	0.55	0.01	0.13	1.69	0.52
mTcCIR 22	5	1.95	0.55	0.49	-0.13	0.05	4.96	0.50
mTcCIR 33	7	2.11	0.43	0.46	0.08	0.10	2.36	0.44
mTcCIR 60	6	2.08	0.55	0.52	-0.06	0.14	1.58	0.54
mTcCIR 08	4	1.84	0.23	0.45	0.48	0.08	2.87	0.44
mTcCIR 18	5	2.39	0.44	0.50	0.10	0.14	1.58	0.44
mTcCIR 40	6	2.36	0.41	0.56	0.28	0.08	2.81	0.55
Total	74							
Mean	5	2.09	0.43	0.45	0.06	0.14	4.43	0.47

Loci *K*=2 (n=123)	Allele		Ho	He	F_{is}	F_{st}	N_m	PIC
	Na	Ne						
mTcCIR 01	2	1.86	0.47	0.46	-0.02	0.02	10.22	0.47
mTcCIR 12	4	2.15	0.51	0.51	0.01	0.06	3.63	0.55
mTcCIR 15	7	3.12	0.58	0.62	0.07	0.09	2.44	0.69
mTcCIR 07	6	2.79	0.21	0.61	0.66	0.08	2.71	0.66
mTcCIR 06	8	2.45	0.46	0.58	0.21	0.06	3.62	0.62
mTcCIR 26	2	1.43	0.30	0.30	-0.02	0.00	142.84	0.30
mTcCIR 11	6	2.24	0.51	0.53	0.03	0.03	7.37	0.54
mTcCIR 24	2	1.47	0.28	0.31	0.10	0.03	7.05	0.32
mTcCIR 37	10	3.50	0.64	0.65	0.02	0.05	5.04	0.68
mTcCIR 22	5	1.74	0.43	0.41	-0.06	0.08	3.05	0.45
mTcCIR 33	7	2.06	0.44	0.50	0.13	0.02	10.01	0.51
mTcCIR 60	7	2.41	0.54	0.57	0.06	0.06	4.25	0.60
mTcCIR08	4	1.92	0.30	0.48	0.38	0.03	7.47	0.49
mTcCIR18	6	2.43	0.44	0.54	0.20	0.04	5.81	0.57
mTcCIR40	7	2.26	0.42	0.56	0.25	0.04	5.74	0.58
Total	83							
Mean	6	2.25	0.43	0.51	0.13	0.05	14.75	0.54

N_a = Number of different alleles; N_e = Number of effective alleles; H_o = Observed heterozygosity; H_e = Expected heterozygosity; Nm = Gene flow; PIC = Polymorphism Information Content

Supplementary Table 3. Summary table of the private alleles across original population classification (*site groups*) and across the inferred clusters (ancestral population at K=3 and K=2). Novel alleles are shaded in grey.

Orig	Locus	Allele	K=3	Locus	Allele	K=2	Locus	Allele
MZ	mTcCIR 01	157	Pop1	mTcCIR 22	294	Pop1	mTcCIR 12	218
MZ	mTcCIR 15	275	Pop2	mTcCIR 07	174	Pop1	mTcCIR 15	256
MZ	mTcCIR 26	301	Pop2	mTcCIR 06	267	Pop1	mTcCIR 15	260
MZ	mTcCIR 26	319	Pop2	mTcCIR 37	174	Pop1	mTcCIR 15	272
STC	mTcCIR 07	177	Pop2	mTcCIR 33	328	Pop1	mTcCIR 15	275
STC	mTcCIR 06	244	Pop2	mTcCIR40	299	Pop1	mTcCIR 07	164
STC	mTcCIR 06	257	Pop3	mTcCIR 12	218	Pop1	mTcCIR 07	176
STC	mTcCIR 11	307	Pop3	mTcCIR 15	256	Pop1	mTcCIR 07	177
STC	mTcCIR 11	327	Pop3	mTcCIR 15	260	Pop1	mTcCIR 07	178
STC	mTcCIR 37	179	Pop3	mTcCIR 07	176	Pop1	mTcCIR 06	243
STC	mTcCIR 37	183	Pop3	mTcCIR 07	177	Pop1	mTcCIR 06	244
STC	mTcCIR 60	206	Pop3	mTcCIR 07	178	Pop1	mTcCIR 06	256
STC	mTcCIR08	310	Pop3	mTcCIR 06	243	Pop1	mTcCIR 06	257
STC	mTcCIR40	285	Pop3	mTcCIR 06	244	Pop1	mTcCIR 11	307
STC	mTcCIR40	303	Pop3	mTcCIR 06	256	Pop1	mTcCIR 11	313
SV	mTcCIR 06	240	Pop3	mTcCIR 11	307	Pop1	mTcCIR 11	315
SV	mTcCIR 37	188	Pop3	mTcCIR 11	313	Pop1	mTcCIR 11	325
SV	mTcCIR 60	202	Pop3	mTcCIR 11	315	Pop1	mTcCIR 37	158
SV	mTcCIR18	346	Pop3	mTcCIR 11	325	Pop1	mTcCIR 37	162
SV	mTcCIR40	284	Pop3	mTcCIR 37	158	Pop1	mTcCIR 37	164
VS	mTcCIR 37	166	Pop3	mTcCIR 37	162	Pop1	mTcCIR 37	168
VS	mTcCIR 37	173	Pop3	mTcCIR 37	164	Pop1	mTcCIR 37	175
VS	mTcCIR 33	301	Pop3	mTcCIR 37	168	Pop1	mTcCIR 37	183
VS	mTcCIR18	350	Pop3	mTcCIR 37	175	Pop1	mTcCIR 22	291
			Pop3	mTcCIR 22	291	Pop1	mTcCIR 22	300
			Pop3	mTcCIR 22	300	Pop1	mTcCIR 33	305
			Pop3	mTcCIR 33	305	Pop1	mTcCIR 33	309
			Pop3	mTcCIR 33	309	Pop1	mTcCIR 33	327
			Pop3	mTcCIR 33	327	Pop1	mTcCIR 60	204
			Pop3	mTcCIR 60	204	Pop1	mTcCIR 60	206
			Pop3	mTcCIR 60	206	Pop1	mTcCIR 60	208
			Pop3	mTcCIR 60	208	Pop1	mTcCIR 60	220
			Pop3	mTcCIR 60	228	Pop1	mTcCIR 60	228
			Pop3	mTcCIR08	308	Pop1	mTcCIR08	308
			Pop3	mTcCIR08	310	Pop1	mTcCIR08	310
			Pop3	mTcCIR18	351	Pop1	mTcCIR18	346
			Pop3	mTcCIR18	355	Pop1	mTcCIR18	351
			Pop3	mTcCIR18	361	Pop1	mTcCIR18	355
			Pop3	mTcCIR40	285	Pop1	mTcCIR18	361
			Pop3	mTcCIR40	293	Pop1	mTcCIR40	284

Pop3	mTcCIR40	303	Pop1	mTcCIR40	285
			Pop1	mTcCIR40	293
			Pop1	mTcCIR40	303
			Pop2	mTcCIR 06	267
			Pop2	mTcCIR 37	179
			Pop2	mTcCIR 22	294
			Pop2	mTcCIR40	299