Massey University / SROS Genotyping Report Summary – 15 April 2020

Introduction

Massey University of New Zealand (Massey) and the Scientific Research Organisation of Samoa (SROS) were contracted by the Samoa Cocoa Industry Development Initiative (S.C.I.D.I.), to help identify suitable cocoa trees for scion wood and seed supply, using genetic identification.

The overarching goal of the S.C.I.D.I. Initiative is for Samoa to achieve a vibrant and sustainable cocoa export industry. One required aspect of this project is to understand the varieties of cocoa that have been grown.

S.C.I.D.I. intends to use the information gathered to identify healthy and productive trees of the Criollo, Trinitario, or Forastero varieties, to supply seed, scion wood and root stock to develop seedlings and grafted plants that are healthier, more disease resistant, and produce premium quality cocoa as mature trees.

Project Background

This genotyping project was undertaken to help S.C.I.D.I. identify parent trees to supply seeds for seedlings and bud wood for grafted seedlings at the S.C.I.D.I. nurseries. The parent trees for genotyping were selected in discussion with cocoa farmers. Each tree was picked for its history of high yield and disease resistance from a farmer's perspective. Farmers were also able to provide a varietal category they believed selected trees belonged to, i.e., Criollo, Trinitario, Forastero or Lafi 7 (a Trinitario sub-type unique to Samoa).

S.C.I.D.I. selected four separate plantations across Samoa with unique characteristics to ensure a healthy representation of Samoan cocoa farms.

- Two plantations were chosen on Upolu, and two on Savai'i.
- One plantation on each island had replaced the majority of the cocoa trees within the last 20 years.
- The second plantation on each island had a large majority of trees still present that were older than 20 years.
- All four farms are geographically separated across Samoa, covering the traditional cocoa growing districts from near Apia, across to Itu Asau.

The Cocoa Varieties

Cocoa has traditionally been identified by researchers, both in Samoa and internationally, as being grouped into four main sub-species: Criollo, Forastero, Trinitario, and Nacional. As Nacional has not been introduced to Samoa, the cocoa varieties present in Samoa are Criollo, Forastero and Trinitario.

The Criollo variety produces the finest flavour cocoa but has a low disease resistance and a low harvest yield, compared to Trinitario and Forastero. Due to its fine flavour value, the Criollo variety is used to make high value chocolate products and generally produces good market returns when marketed in varietal bean form.

The Forastero variety has high disease resistance and produces a large harvest. However, the taste of Forastero cocoa is not as complex/fine as that of Criollo and Trinitario.

Trinitario cocoa is more disease resistant than Criollo, and produces a yield similar to Forastero, with a fine flavour, similar to Criollo.

The Genetic Data

Genotyping is the process of comparing DNA samples from multiple individuals to determine how similar the DNA sequences are, and therefore how closely related the individuals are. This can be done across the entire DNA of each individual (genome mapping) or by using short sections of DNA called microsatellite markers.

Previous international research seeking to identify the varieties of cocoa trees has focused on using 15 microsatellite markers at specific points along the cocoa genome. This set of 15 markers has been used internationally since 2004 to aid in identifying cocoa trees as Trinitario, Criollo, Forastero, or Nacional.

Such genotyping approaches often use "reference samples" to aid in accurate identification. Reference samples are samples of a known genotype, that are compared against the unknown samples. Cocoa Samples close enough to the reference samples can be considered to be of the same variety as the reference sample, e.g. Trinitario.

Tree Selection

Once the four farms had been selected, 50 individual trees were selected from each farm. Suitable trees were identified by the farmer. In consultation with a trained staff member from the Samoa Trust Estate Corporation (STEC), farmers established the likely tree variety based on visual characteristics (phenotyping).

The selected trees were labelled with a unique identifying number, as well as fluorescent tape, to indicate selected trees at a distance. Tree locations were also mapped using GPS.

The data used to phenotype the tree, as well as additional data such as tree age, productivity, and GPS location were then entered into a STEC database to enable enhanced analysis of the genotyping results at a later date.

The Process

Fresh leaf samples were taken from each of the 200 selected trees by a team of SROS scientists and support staff. The leaf samples were sent to SROS for processing, with DNA successfully sampled from 197 of the 200 samples. These DNA samples were then sent to Massey University in New Zealand for genotyping and analysis.

Once the samples were received by Massey, the samples were subjected to Polymerase Chain Reactions (PCR) amplification. PCR is used to amplify small sections of DNA for analysis. The PCR used for the Samoa Genotyping project successfully amplified the 15 microsatellite markers required for this study.

Once amplified by PCR, the DNA samples were analysed by a genetic analyser, which reads DNA and converts the sampled DNA into genotype files that can be analysed by scientific software.

The Genotype Analysis

197 trees were successfully genotyped. These successful samples were analysed for genetic diversity and populational structure to determine how genetically close each sample was to the others. The data generated from all 197 trees was plotted into a "Neighbour-net" (Figure 3, Page 27 of the full report). Each tree is represented on the Neighbour-net in the form of a single line. Each line is identified by the code each tree was given at the time of sample selection, e.g. STC45.

The Neighbour-net provides certain insights. These are:

- The more similar trees are genetically, the closer they appear on the Neighbour-net.
- The length of each line on the Neighbour-net represents the amount of genetic variation ("genetic distance") between individuals.
- The merging of individual tree lines that occurs towards the centre of the Neighbour-net and which varies considerably amongst the tree lines indicates points where trees last shared a common ancestor in a previous generation.
- The short lines and visual blocks that are shown between individual tree lines close to the centre of the Neighbour-net indicate degrees of common genetics between various sub-sets of trees, due to inter-breeding in earlier generations.

Results Summary

The genotyping analysis identified 3 groups of ancestral DNA across the 197 samples through the use of a Structure Plot (Figure 2, page 26 of the full report). The Structure Plot identified that there was evidence for a minimum of two different ancestral DNA groups amongst all trees, and that the most likely number of ancestral DNA groups was three. The 197 samples were all found to be carrying varying amounts of DNA from each ancestral population. **In other words, none of the sampled trees could be classified as a pure-bred variety**. This indicates that cocoa trees have inter-bred (hybridised) since the initial introduction of each variety and ancestral population to Samoa.

Of the 197 trees, 83 were found to have 90% of their DNA matching to one of the identified Samoan ancestral DNA populations. To analyse the relationships between these trees, they were then placed into a second Neighbour-net (Figure 4, page 29 of full report). Two reference samples from international cocoa gene banks, one Criollo and one Amelonado (a Forastero sub-type) sample, were also included to enable comparison of the Samoan samples against known Genotypes. The Project was not able to source a Trinitario reference for this genotyping project.

One group of trees sampled showed a high degree of genetic similarity to the Amelonado reference sample, while a second group showed a somewhat high degree of genetic similarity to the Criollo reference sample, although much less than the first group to the Amelonado reference sample. A third group of trees showed a high degree of genetic similarity to each other, without there being a reference sample available to quantify likely variety in a scientific fashion. This 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 several possibilities arise:

1. The third genetic group are what is known as Samoan Trinitario, **OR**:
2. The third genetic group represents a further sub-species or cultivar of Forastero, with a significantly different genetic profile than that of the Amelonado reference sample, which in its own right is a cultivar of the Forastero variety, **OR**:
3. The third genetic group represents a cultivar of hybrids between Criollo and Forastero, with a significantly different genetic footprint to the Samoan Trinitario cultivar.

The Genotypes of the cocoa trees were also compared against the initial farmer provided phenotype data, to determine the congruence between phenotypes and genotypes of the cocoa trees. The comparison of genotype to phenotype showed that the phenotyped groupings identified did not match the genotyped groupings. While some trees from the phenotyped Criollo and Forastero groups were found to be in the genotyped Criollo and Amelonado groups respectively, the majority of trees from each phenotyped group were found spread across the three genotyped groups. Therefore, there was little match found between the genotyping and phenotyping identifications.

Each plantation included in the project had a different balance of the three identified genetic groups of trees.

Discussion

The genotyping project has, in addition to providing direct results, also generated associated learnings and insights, which require further examination.

The project's starting point was the observation that Cocoa trees in Samoa are planted either on the basis of phenotyped parentage (selected seeds from a well-performing tree), or without any selection criteria (parent trees not selected for performance).

The genotyping project had selected samples from trees that had been phenotyped by farmers as high-performing trees. The purpose of the genotyping project was to identify the genetic profile of trees selected through phenotyping, to provide certainty of the parentage of the selected trees. The generated knowledge would then aid the decisions made in relation to seed selection for the production of root stocks, and scion wood selection for the grafting programme.

The project did not expect to find "pure" Criollo/Forastero/Trinitario genetic lines, as cocoa is well known for its tendency for rapid hybridisation.

However, the low degree of congruence between farmer phenotyping and DNA analysis results came as a surprise. In trying to understand why this is the case, we have the following observations:

- Cocoa is a species that cross-fertilises very easily. Since the initial introduction of cocoa into Samoa in the 1883, there have been at least 3 further introductions of new genetic material by 1900. These have then been grown in the same plantations as the trees from earlier introductions, leading to extensive cross-fertilisation across all the Samoan cocoa plantations.
- In addition to this cross-fertilisation, nature and farmers tend to select for high-performing trees on plantations, regardless of type, meaning trees were likely to be descended from high performing trees regardless of variety.
- Farmers may attempt to breed trees on the basis of phenotype-based pod identification, with the belief certain pod visual characteristics are only shown by certain sub-species, e.g. Criollo. However, a farmer may actually be breeding Forastero trees that grow Criollo-like pods, as the farmer cannot visually tell the difference, and the tree providing the pods itself already being a hybrid between a Forastero and Criollo. To the farmer though, both trees would present as Criollo.
- Genotyping can only be used to compare the DNA similarity, expressed as a percentage. The methods used cannot determine what physical characteristics are shared by trees, but only what DNA is shared. As small sections of DNA can code for very different characteristics, trees that have very similar DNA, may have completely different pod and bean characteristics, if the DNA that codes for these characteristics is in the DNA that is different. However, if two trees of different varieties also share this section of DNA, they may have the same pod and bean characteristics that the farmer is looking for, despite not sharing much else in terms of DNA.
- The different balances of variety groups between plantations is to be expected. Early introductions of Criollo, the Amelonado variety of Forastero, as well Trinitario cocoa sub-species occurred only on the then Deutsche Handels and Plantagen Gesellschaft (DHPG) plantations during the German administration period, which ended in 1914. Later, multiple subsequent introductions of further Forastero sub-type, including the Amelonado variety, occurred. In the 1950s and 1960s, Trinitario propagating projects took place, including the Lafi-7 project, which saw thousands of trees being propagated through cuttings across Samoa. As these programmes all affected only certain regions of Samoa, and some projects occurred as recently as the 1990s, not enough time has elapsed for the associated genes to spread relatively evenly across Samoa.

Non-Genotyping Related Learning

The genotyping team conducted an independent literature review on cocoa varieties. The final report summarises not only genotyping related learnings drawn from the literature review, but also learnings related to the management of cocoa black pod disease, which are in line with, and confirm, the position S.C.I.D.I. is holding on this matter:

- Cocoa production losses to Black Pod is estimated to be in the region of 60 – 80% of all production.
- Black Pod shows as water-soaked spots on cocoa tissue that eventually mummify cocoa pods.
- Tree infection to symptoms occurs within 1 to 2 months of initial infection
- The process of pod infection to pod death can occur in as little as 14 days.
- Cuprous oxide and metalaxyl are two recommended spray fungicides for Black Pod in alternate seasons (cuprous oxide during dry periods, metalaxyl during wet periods).
- For a single use application, Potassium Phosphonate has been demonstrated to work at controlling Black Pod through trunk injection.
- Previous cocoa breeding in Samoa focused on producing Black Pod resistant cocoa resulted in Lafi7, but this project resulted in some genetic bottlenecks.

Concluding Comments

Based on the project results, along with some associated learnings from other parts of S.C.I.D.I.'s work, S.C.I.D.I. now intends to utilise this data to assist in selecting seed and scion wood supply trees for our nurseries. A further genotyping round on the next generation of seedlings is now also being considered, to determine the genetic similarity and drift of the seedling population against the parent trees in late 2020 or early 2021.

Without proper management by fungicides, Black Pod infections will cause further damage to Samoa cocoa plantations and yields. Additional focus will therefore need to be given to managing the spread of Black Pod by way of farmer education.