Assessment of the genetic diversity of Puerto Rican sweet potato (*Ipomoea batatas* (L.) Lam.)

by

Lorraine Rodriguez Bonilla

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In

Biology

UNIVERSITY OF PUERTO RICO

MAYAGÜEZ CAMPUS

2013

Approved by:

_____	_____
Carlos Ortiz Malavé, PhD Member, Graduate Committee	Date
_____	_____
Fernando Bird Picó, PhD Member, Graduate Committee	Date
_____	_____
Dimuth Siritunga, PhD President, Graduate Committee	Date
_____	_____
Nanette Diffoot Carlo, PhD Director, Department of Biology	Date
_____	_____
Dr. James Beaver, PhD Graduate Studies Representative	Date

ABSTRACT

Sweet potato (*Ipomoea batatas* L.) is the seventh most important food crop after maize, wheat, rice, barley, potato and cassava. Due to its many agricultural advantages, such as adaptability to different environmental conditions and its nutritional value, research focuses of sweet potato are increasing. Because the genetic diversity of sweet potato in Puerto Rico is poorly understood, there is a need to assess its diversity, especially among sweet potatoes cultivated by farmers in the island. A total of 137 samples of unknown origin from around the island were collected. This collection as well as 8 accessions from the Puerto Rican germplasm collection plus 8 accessions from the National Repository in Griffin GA, were subjected to a genetic diversity analysis with 23 SSR markers using an fluorescent PCR technique. The results of the analysis of the 23 loci showed a total of 205 alleles in the 155 samples, ranging from 2 to 20 alleles per locus with an average of 8.9 alleles per loci. Overall average observed heterozygosity (H_o) was high across populations with a value of 0.637 while measurements of total heterozygosity (H_t) revealed a large genetic diversity throughout the populations with a value of 0.731. The heterozygosity within populations (H_s) was 0.694 revealing high levels of genetic diversity in the populations. From the UPGMA clustering method two main clusters were depicted. Cluster I contained 19 unknown accessions from across the island while cluster II had the majority of unknown samples as well as the known accessions from PR and GA. Cluster II was subdivided into 4 smaller sub-clusters. In Sub-cluster 1 we had the majority of known samples, they clustered very closely together. Sub-cluster 2 samples from across the island grouped together and most samples were white fleshed accessions. Interestingly, two accessions in sub-cluster 3 were identified as clones (11W, 17W), both from the West but differing in flesh color. We can conclude that there is in fact a high level of genetic diversity across the island which can be related to genetic makeup of sweet potato, the ability of dispersal of a vegetatively propagated crop, human intervention and the outcrossing nature of sweet potato. High levels of diversity found in Puerto Rico and the history of domestication and dispersal of sweet potato turn this crop into an extremely valuable resource that needs to be protected and further studied.

RESUMEN

La batata (*Ipomoea batatas* L.) es el séptimo cultivo más importante después del maíz, trigo, cebada, papas arroz y yuca. Gracias a sus características y ventajas agrícolas, tales como la capacidad de adaptación a diferentes condiciones ambientales y su valor nutricional, la batata se ha convertido en foco de investigación. En Puerto Rico la batata y su variabilidad genética son desconocidas y hay una necesidad de evaluar la diversidad cultivada por los agricultores de la isla. Un total de 155 muestras de origen desconocido de alrededor de la isla fueron adquiridas. Esta colección compuesta de muestras desconocidas, así como 8 accesiones de la colección puertorriqueña y 8 del Depósito Nacional en Griffin GA, fueron sometidos a un análisis de la diversidad genética. Este trabajo conto con 23 marcadores moleculares o "SSR" los cuales se utilizaron mediante una técnica de PCR con fluorescencia. Los resultados del análisis de los 23 loci ayudaron a descubrir un total de 205 alelos en las 155 muestras, estos alelos van desde 2 hasta 20 por locus con un promedio de 8,9 alelos por locus. En general la heterocigosidad observada promedio (Ho) fue alta en todas las poblaciones con un valor de 0.637 mientras que los valores de heterocigosidad total (Ht) revelaron una gran diversidad genética en a través toda la población, con un valor de 0.731. La heterocigosidad dentro de las poblaciones (Hs) fue de 0.694 revelando altos niveles de diversidad genética en las poblaciones. A partir del método de agrupamiento UPGMA dos grupos principales fueron representados. El grupo I contiene 19 accesiones desconocidas de alrededor de la isla, mientras que el grupo II contiene la mayoría de las muestras desconocidas, así como las variedades conocidas de PR y GA. Este grupo se subdivide en 4 subgrupos más pequeños. En el sub-grupo 1 podemos observar que la mayoría de las muestras conocidas se agrupan juntas. También podemos observar como 3 accesiones desconocidas se unen también a este subgrupo. En el subgrupo 2 la mayoría de las muestras poseen raíces con pulpa blanca, lo que podría explicar este suceso. Curiosamente, en el subgrupo 3, se identificaron dos accesiones que m ser clasificadas como clones (11W, 17W), ambas muestras del Oeste pero difiriendo en color de la pulpa de la raíz. Podemos concluir que realmente hay un alto nivel de diversidad genética a través de la isla. Esta

diversidad puede estar relacionada con la composición genética de la batata, su capacidad de dispersión causada por el cultivo vegetativo de la misma, la intervención humana para escoger variedades mejores además de los altos niveles de cruzamiento que ocurren en la naturaleza. Los altos niveles de diversidad que encontrados en la isla, la historia de su domesticación y la dispersión de las batatas convierten este cultivo en un recurso muy valioso que debe ser protegido y estudiado más a fondo.

To Cruz, Fabian and Edith, without their trust and full support I wouldn't be where I am today.

"Caminante, son tus huellas el camino y nada más; Caminante, no hay camino, se hace camino al andar. Al andar se hace el camino, y al volver la vista atrás se ve la senda que nunca se ha de volver a pisar. Caminante no hay camino sino estelas en la mar."

-Antonio Machado

Acknowledgments

I have encountered many people in my life; from them I have learnt and grown. I want to thank Dr. Siritunga, for giving me an opportunity even when I didn't have the best academic standing. He gave me a chance and it changed my future, I'm a better person and professional for it. I thank him for the trust, support, confidence and all the tools necessary to be where I am right now.

This work could not be possible without the help of my graduate committee: Dr. Fernando Bird-Picó, thank you for your time and support with all the statistical data. Thank you to Dr. Carlos Ortiz for all the encouraging words, support and for allowing me the opportunity to freely use the samples at the Gurabo Experimental Station.

To my mentor and friend Milly Montero, thank you for teaching me all the tools I needed to succeed in the last five years, I could not have done it without you. Thanks for the long nights, the chats and the laughs when things were stressful, I look up to you as a professional and as a person. I also want to thank Gladys Toro for her help and words of support.

To my undergraduate students: Lorena and Derickson, thank you for giving me the opportunity to teach and share what I know and thank you for the support you gave to this project. Dianiris Luciano, thank you for being more than a student, a friend. I was blessed with the opportunity to teach you and work with you. You supported me a 100% throughout this journey and more than my student you were my complete equal. Thank you for all the time and work you put into making this possible, words can't even describe how much I thank you.

To my lab-mates: Kevin, thanks for being more than a work partner. You are most certainly the joy of the lab with all your occurrences. Thank you for all your help and for making the lab a fun place to work. Thanks to Shaimar, Gabriela, Clara, Cristina, Noraliz and Alejandro for everything you have done to help me achieve my goals.

I want to thank my brother Fabian for your time and helping me when I needed it the most, I will be forever be grateful. Thank you to my mom for always believing that I can do anything and for being there to do whatever she can to help me achieve my goals. To my aunt's Edith and Carmen, thanks for all the support thought the years, the time and the encouragement, I owe you so much.

Last but not least to my friends: Juliany, Ileana, Dagma, Fatima, Mario and Zorybel. All of you have been important part of my life, some of you have been forever in my life while others for a little less time, but I still have always felt each of your love and support. Even in my lowest you have been there for me, I love you.

TABLE OF CONTENTS

Table List

Figure List

Justification

Over 925 million people are undernourished today. Over 98% of these people live in developing countries mainly in Asia, the Pacific and in Sub Saharan Africa (SSA) (FAO 2012, Figure 1). According to UNICEF one person dies of starvation every 3.6 seconds and almost 75% of these are children. Developing countries depend on agriculture as an economic and food source. It is important for them to have some sustenance that can provide them with nutritional value and that can be easily grown in order to secure food for them and their families.

One way to assure the needs of this people can be met is by studying and safe guarding the genetic diversity of different crops, which are the source of nutrition for them. Some authors such as Allendorf (2013) have stated that there are three types of diversity: genes, species and ecosystems. All three of them are vital to us. Knowing and understanding these diversities can help the development of conservation programs. In agriculture and food security this can be available for the farmers and could also help with the discovery of new varieties. Farmers are the keepers of crop diversity and they maintain it without the knowledge of the importance of crop conservation. Farmers trade plant material and choose the ones with advantageous characteristics such as better yield, biotic stress resistance and abiotic stress resistance, thus safe-guarding the advantageous alleles in the population. Having diverse varieties growing at the same time can help with food security issues in case a virus or a plague affects one susceptible crop or variety in specific. If this is to happen then not all of the food sources will be scarce since the farmers can reestablish the crop with resistance cultivars.

Conservation of the different crop genetic resources is vital for food security. With the on-going abuse of urbanization and construction valuable areas to cultivate crops are lost and furthermore wild-species of important crops can also be lost forever through deforestation. These plants could have important traits that could be used in crop improvement or bio-fortification programs or for the conservation of genetic resources in order to maintain a healthy genetic pool, preventing genetic erosion which can impact economy and food production. There is a lot of uncertainty as to whether the increase in world food production could be met without diversity (FAO 2009).

Sweet potato (*Ipomoea batatas* L.) is the seventh most important food crop after wheat, rice, maize, potato, barley and cassava (FAOSTAT 2009). Sweet potato has recently received greater research-related attention due to its many agricultural advantages such as its adaptability to different environmental conditions and its nutritional value. With the abundance of undernourishment throughout the world, crops like sweet potato needs to be further studied and exploited.

In Puerto Rico, sweet potato is poorly understood and the genetic diversity is unknown. A collection of 8 accessions is maintained at the Gurabo Agricultural Experimental Station and due to its aforementioned importance it is necessary to increase the knowledge of the local varieties for future use to our advantage. In order to do this, an assessment of the genetic variability of the collection and unknown sweet potato varieties (grown by farmers) will be performed using molecular markers. Molecular markers are great tools that are known to be helpful in order to analyze the genetic diversity of different crops. They have proven to be very informative and successfully detect clones or duplicate accessions which are treated as different cultivars. Many farmers do not really know exactly what cultivar they are actually growing, this is because they only rely on visible characteristics of the plants. Cultivars are kept by farmers and since the plant is mostly vegetatively propagated information about the accession could be unknown or misguiding. Diversity can also be at risk, advantageous alleles might disappear from the population leaving it vulnerable to various abiotic or biotic stresses. For those reasons and more a study that can reveal the genetic diversity of Puerto Rican sweet potato is needed.

Chapter I

General Introduction

Unsustainable development is degrading the environment menacing the biodiversity that provides food and nutrition. With the increase of malnourishment in developing countries, which account for 98% of the world hunger (FAO, 2012), different crops are being targeted in research programs that aim to reduce the aforementioned issues. Sweet potato [*Ipomoea batatas* (L.) Lam.] is one of the most important staple crops. Sweet potato is grown in over 82 developing countries and is one of the top five most important crops in at least 40 of those countries (FAO STAT, 2009). On the other hand, sweet potato is also a key part in the daily diet of millions of people and it is produced in large quantities in industrialized countries such as the US. Despite its importance, sweet potato remains as an underexploited, poorly studied crop (Mukhopadhyay *et al.*, 2011; CIP, 1999), full of potential, which could help eradicate most of the undernourishment related issues today.

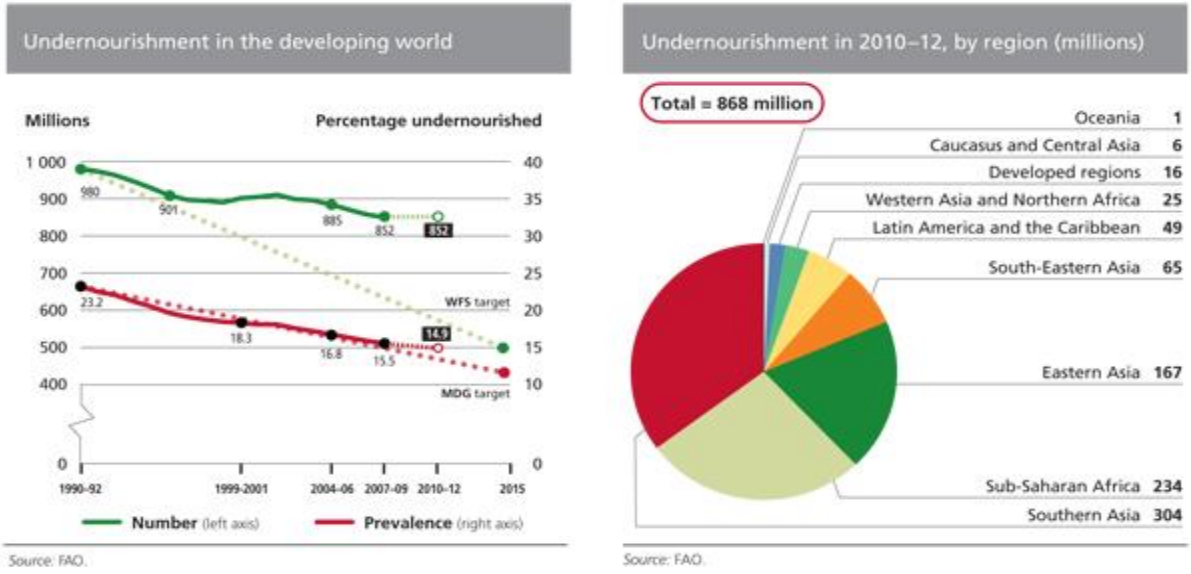


Figure 1. Undernourished people in 2010-12 (Source: FAO, 2013)

Botany of the plant

Sweet potato (2n=6x=90), a vine with storage edible roots and leaves is part of the family Convolvulaceae, genus: *Ipomoea* species *Ipomoea batatas* (L.). Linnaeus in 1753 described this species as *Convolvus batatas* but Lamarck in 1791 re-classified the species as a member of the *Ipomoea* genus based on pollen and stigma morphology.

Sweet potato is a dicotyledonous, tuberous-rooted, perennial plant that can also be annually produced. It is primarily grown via vegetative propagation of the stems or roots but can also be propagated using seeds. Sweet potato behaves as a short day photophilous plant which has a better performance growing at low elevations (Mukhopadhyay *et al.*, 2011) nonetheless it is known to grow at 2,500m as well. Although it is a very adaptable crop which can tolerate very high temperatures (Laurie *et al.*, 2013) it grows better at temperatures closer to 20° C (Mukhopadhyay *et al.*, 2011). Some varieties are known to be naturally able to resist virus and other diseases and be able to grow in very versatile environmental conditions such as drought and low fertility soil. These self-incompatible plants can show different morphologies, varying in color and shape of the leaves, roots and vines. Aside from varying in flesh color and shape the roots also have various levels of sweetness. Flower coloration also can be very diverse among varieties with colors fluctuating from green to purple. Sweet potato plants can grow either erect or spreading and, once planted, their roots can be harvested from 3 to 6 months after planting (CIP, 2000) even in marginally degraded soil. *Ipomoea batatas* (L.) is known to have 13 wild species as relatives (Austin and Huamán, 1996). The majority of these wild relatives are known to be native to the Americas with only *I. littoralis* which is known to be found in Australia and Asia. Sweet potato is openly pollinated but there are natural barriers preventing pollination (Mukhopadhyay *et al.*, 2011). Some of these barriers include its self-incompatibility and sometimes partial sterility. Bees and other insects are in charge of naturally pollinating sweet potato but for breeding purposes it's usually hand pollinated to control parentage (Rossel *et al.,* 2008).

Within the *Convolvulaceae,* the genus *Ipomoea* is known to have the biggest number of species. Austin and Huamán (1996) mentioned that the genus is comprised

by almost 700 species, with more than half of them being in the Americas where most of them are native while only a very small amount been introduced. Some of the known primary diversity centers for this crop are Central and South America specially Guatemala, Peru, Colombia and Ecuador. Secondary centers of diversity for sweet potato can be found in Asia and in the Sub Saharan Africa (SSA) (Yada *et al.*, 2010). Even though those are the centers for diversity, sweet potato can be found and grown between latitudes 42°N and 35°S and anywhere from sea level up to 3000 m (Veasey *et al.*, 2008) both in tropical and sub-tropical climates (Figure2).

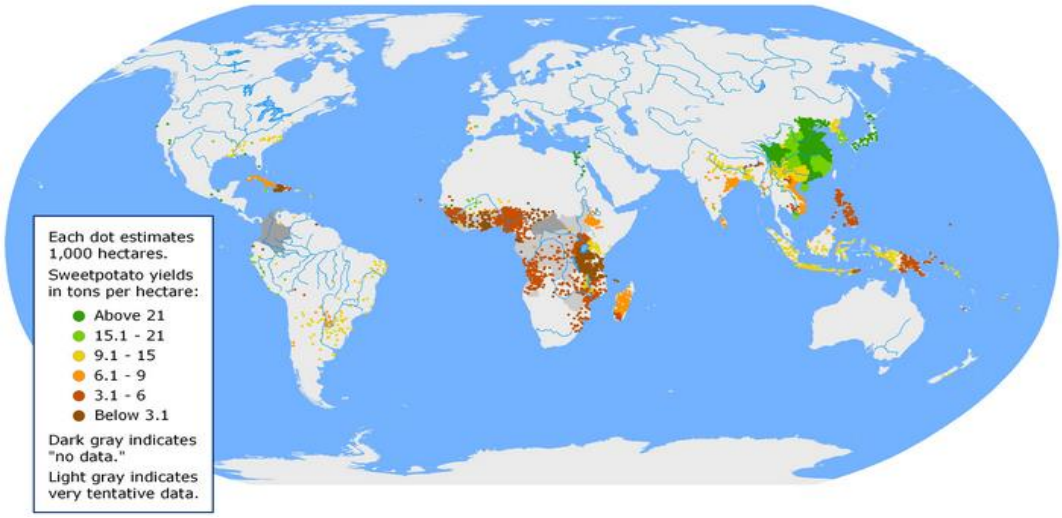


Figure 2: Areas of cultivation and yield of sweet potato (Source: CIP, 2011)

Origin and Distribution

Sweet potato is known to have been domesticated in America and according to O'Brien (1972) and Zhang *et al.*, (2004) this event took place somewhere between the mouth of the Orinoco River in Venezuela and the Yucatan peninsula more than 10,000 years ago. Linguistic and archeological findings are responsible for this affirmation. According to Austin 1988, sweet potato was then distributed throughout the Americas, due to the different migration routes of the New World inhabitants. Fossils of nearly 9,000 years found in Chilca Canyon in Peru prove that sweet potato was in fact one of the first crops to be domesticated (Engel, 1970)

It is known that this plant was introduced to Europe after Christopher Columbus first trip to America in 1942 via Spain (O'Brien, 1972). On his first trip to America Columbus visited different islands such as Cuba and Hispaniola (Dominican Republic and Haiti) from where many different plants were taken and other resources such as gold. The inhabitants of these islands, the Taínos, were descendants of the Arawakan-speaking people from northeastern Venezuela (Keegan, 1995) and the Guiana coast who colonized the West Indies and according to O'Brien (1972) the word "batata" came from the Arawak language.

Sweet potato was further distributed to other parts of the world such as Africa, India and the Pacific by Portuguese ships and it was documented in Asia approximately in year 1600. At the beginning these facts supported the domestication theory, there was no records or information of sweet potato growing anywhere in Asia, Africa, Australia or Europe before the Columbus trip to the New World, not even on the records of previous civilizations such as Egypt, Persia, Mesopotamia, Rome, Greece or India .

However new theories have arisen regarding where and how the domesticated form of sweet potato, we know today, came to be. Diversity studies done by Zhang *et al.,* (2000) using Amplified Fragment Length Polymorphism's (AFLP) showed that the genetic diversity found in Central America and the Peruvian one are very different from each other supporting in a way the previous theory of the domestication area. According to Roullier *et al.*, (2011) the results from Zhang *et al.,* (2000) could have confirmed this theory but another study with Random Amplification of Polymorphic DNA (RAPD) (Gichuky *et al.*, 2003) demonstrated the possibility of two different gene pools for American sweet potato.

Roullier *et al.*, (2011) established that it's known that other plants such as some from the genus Cucurbita were domesticated individually in different areas (Sanjur *et al.*, 2002) and that in some cases wild relatives of plants such as *Phaseolis vulgaris* were domesticated twice in different geographical areas suggesting that this might be a possibility for sweet potato as well. Roullier *et al.,* (2012) analyzed samples from different places in America and assessed their diversity using chloroplast (7) and nuclear SSR (13) markers and like Zhang *et al.,* (2000), a pattern of division was

observed. The samples segregated into two groups North and South, North being comprised by samples from Central American / Caribbean samples and the South comprised of samples from South America. These results give validity to the previously alluded theories, one of them being that sweet potato could have been domesticated two times one in South America and the other in the Central America/Caribbean region.

Economic Importance

Sweet potato is widely produced and consumed in different countries around the world. In 2011 more than 104,259,988 tones were produced worldwide according to FAOSTAT. Sweet potato, although is mainly used for human consumption, have many other uses in different countries. China is the top producer with more than 80% of the world production. In China sweet potato is extremely versatile, it is used as raw material for different commercial applications such as starch and alcohol production, livestock feed and as food security source in poor areas (ESEAP, CIP, 2011). Following China distantly are Tanzania, Uganda and Indonesia (Figure 3) accounting for only 3% of the total world production with primary use being as a staple or famine relief crop. In Africa the average annual per capita consumption of sweet potato 9Kg and this number may vary within countries and regions. In 2004, United States was the leading exporter of sweet potato with more than $22 million, while Canada was the main importer. In the US this crop is widely consumed in different ways, varying from main dishes to desserts with an average annual per capita consumption of 2kg. In Latin America sweet potato is the third most important crop following cassava and potato. The top producers of sweet potato in the America's are Argentina, Cuba, Haiti and Peru. Argentina ranks number 19 in the total world production of sweet potato in the latest report of FAOSTAT 2011.

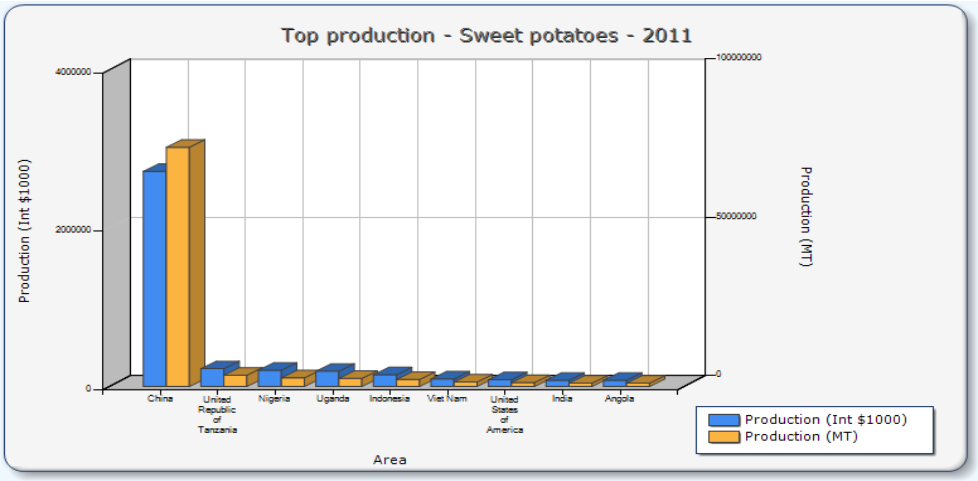


Figure 3: World top producers of sweet potato (Source: FAO, 2011)

Sweet potato production in Puerto Rico

In Puerto Rico, even though widely consumed, sweet potato is not one of the main crops produced. In fact it is considered a minor crop since it is mainly grown by small scale farmers and cultivated mostly by people 65 years and older (COA, 2007). Sweet potato is usually grown in home-gardens as a food source for family consumption. According to Moscoso (1955), there are two theories about the arrival of sweet potato to Puerto Rico. One state that is was introduced from Brazil by natives, and the other one suggest that sweet potato was part of the original flora of Puerto Rico. Taino indians cultivated 5 types of sweet potato: atibiunex, aniguamar, guaraca, guacarayca and guananagax (Moscoso, 1955). Today many more varieties of sweet potatoes are cultivated in Puerto Rico. In 2002 there were 220 farms that produced sweet potato in contrast with 438 farms in 2007. Even though the amounts of farms producing sweet potato have increased, the overall area being cultivated has decreased. In 2007 the census of agriculture reported near to 1,175,070 m^2 dedicated to the production of this crop with nearly 2327.4 tones, whereas in 2002 the amount of land was near to 2,534,850 m^2 and the production was of nearly 4277.8 tones. According to FAOSTAT in 2011 Puerto Rico sweet potato production increased from 2323 to 2974 tones.

Importance as a crop

Sweet potato is a highly nutritional crop being an excellent source of carbohydrates, dietary fiber, sugars, complex carbohydrates, proteins, iron and calcium. This crop is also an important source of vitamin A and C especially in the orange-fleshed varieties. Furthermore, sweet potato can produce more edible energy per area per day than other crops, such as wheat, rice or even cassava (Table 1.) (CIP, 2009). In 125g of fresh weight of its roots there is enough Beta-carotene to provide the necessary daily intake of a preschooler (CIP Facts 2009). This makes sweet potato a key to solve the vitamin deficiency issues which is one of the main concerns in developing countries, causing the death of more than 600,000 deaths per year (Burri, 2011).

Table1. Nutrients found on 100g fresh weight, raw, unpeeled sweet potato analyzed by Quality and Nutrition Lab, CIP

Minerals	mg
Iron	0.5
Zinc	0.2
Calcium	34
Potassium	298
Phosphorous	29
Total carotenoids)	15.5
Beta-carotene	13.1

Sweet potato is mainly grown for human consumption but even though it is a “subsistence” or “famine relief” plant, its uses have expanded. Sweet potato has starchy roots, and in some cases, immature leaves are the only part of the plant used by people. Mature leaves or bad roots are used for animal consumption, mainly for pig diets. Other edible and non-edible parts of the plant are used in the industry, for example storage roots are used as material for starch and glucose production. In

countries in which sweet potato is not used exclusively as a main food source like Brazil, China, Japan and the US you can find products such as chips, candies, sauces or simply as a base for many different desserts. In the SSA sweet potato still is a food security crop and a source of income that ranks third in importance after potato and cassava in countries like Uganda (Tumwegamire *et al.*, 2011) which have a daily per capita intake of about 240g (FAO, 2007). In those countries sweet potato becomes a seasonal staple during the dry season when other food sources are exhausted (Sweet Potato Facts, CIP 1999).

In developing countries sweet potato is a crop related with poverty, people that could not afford other crops such as maize were obligated to consume it. This relegated sweet potato as a secondary food crop even though the yellow-fleshed varieties of sweet potato are rich in beta-carotene (Tumwegamire *et al.*, 2011). Many people prefer not to use this crop because of the misconception of sweet potato/poverty relation. Although some do use sweet potato flour to make different types of food, and some countries even use it as a sweetener or just to accompany milk, the full exploitation of sweet potato in developing countries is constrained because of its perishability, bulkiness and most of all the low consumer acceptability.

Limitations to world sweet potato production

As any other crop, sweet potato has many constraints that sometime can prevent this crop from reaching its full potential. According to Mukhopadhyay *et al,* (2011), there are two main restrictions for this crop, the first one is a socioeconomic constraint and the second one is biotic. Urbanization reduces areas to grow this crop, increase in income changes food habits of society allowing them to afford other types of crops easier to cook. Another limitation is the fact that other low yielding crops are easier to maintain and adapt better to every type of soil.

Some of the biotic constraints that sweet potato suffers is caused by the sweet potato weevil (*Cylas fomicarius, brunneus, puncticolis)*, this insect is the biggest threat to this crop worldwide. Mukhopadhyay *et al.*, (2011) stated that the amount of losses in sweet potato caused by the weevil annually is approximately 90%. The roots are the

part of the plant more commonly damaged although the adult weevils can be found on the foliage (CIP, (2013). Some of the visual characteristics of the infected plants include feeding marks and oviposition holes. Some other threats are cause by different types of viruses like for example: Sweet potato virus disease (SVD), Sweet potato chlorotic stunt virus (SPCSV) and Sweet potato sunken vein virus (SPSVV) which can result in low yield (80% less root yield in some cases), chlorosis, rugosity and leaf strapping. SPVD is the most detrimental virus induced syndrome of sweet potato and it causes many economic losses because the plant produces non-usable tubers (Gwandu *et al.*, 2012). Fungi and bacteria can also cause problems to this plant, for example *Helicobasidium mompa* and *Rhizopus stolonifer* can cause the roots to rot while still in the soil.

Molecular Markers and genetic diversity

For many years farmers have relied only on morphological traits in order to distinguish between different varieties of a crop. It is very difficult in some cases for farmers to know exactly what plants they are growing or harvesting. In sweet potato cultivation constrains such as lack of space between varieties in the field, vine spreading and morphological variation can cause confusion regarding the exact identity of the cultivars being grown in the field. Such mix-ups can occur in part due to the fact that sweet potato's main reproduction method is vegetative, using the stem or roots. For conservation purposes relying only on morphological traits can be very dangerous. Diversity could be loss and a lot of important genes could disappear from the population preventing high diversity in the crop. This diversity is vital for food security in the case of staple crops.

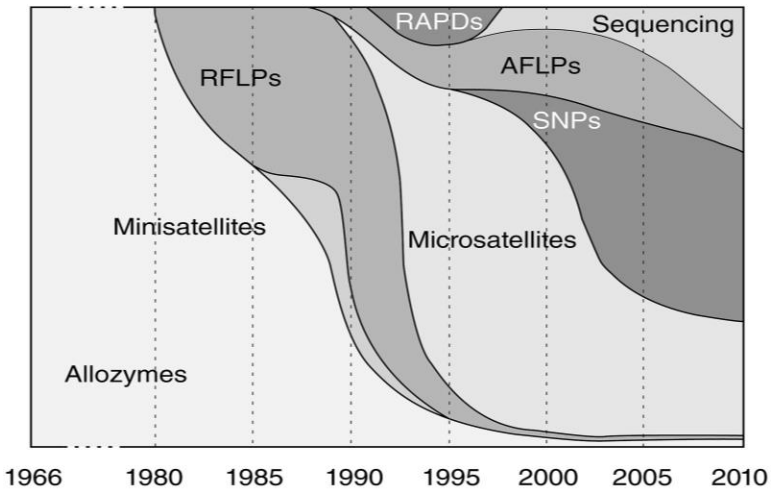


Figure 4: Subjective view of the changing or relative popularity of major molecular markers in conservation genetics (Source: Allendorf *et al.*, 2013)

Molecular markers can give us an insight to the genetic makeup of a population. These markers are crucial for conservation due their many advantages and the information they can give us about the genetic composition of a crop. Some markers such as Amplified Restriction Length Polymorphisms (AFLP), Random Amplified polymorphic DNA's (RAPD), Inter-simple Sequence Repeats (ISSR) and Simple Sequence Repeats (SSR) have been used in many genetic diversity studies (Table 2, Figure 4). But not all of them are equally informative. A good marker should not be affected by environmental or developmental pressures. It should be able to detect silent changes in coding regions and capable of identifying homologous changes in different lineages and must be reproducible.

According to Buteler (1999) an ideal molecular marker to study plants should disclose multiple alleles (co-dominant) have even distribution throughout the genome and can easily differentiate genetically similar individuals. In the case of AFLP and RAPD they show dominant instead of co-dominant information and they are not very reproducible. In contrast, SSR markers have all of the characteristics of a good informative marker (Table 2).

Table 2. Comparison of different molecular markers

	RFLP	RAPD	AFLP	Microsatellites
Abundance	High	Very High	High	Medium
Level of polymorphism	Medium	Medium	Medium	High
Dominance	Co-dominant	Dominant	Dominant	Co-dominant
Sequence information needed	No	No	No	No
Costs (development and application)	Medium	Low	High	High
Automation	Limited	Yes	Yes	Yes
Reproducibility	High	Low	Medium	High

Microsatellites or Simple sequence repeat (SSR) markers

SSR markers consist mainly of mono, di, tri, tetra and even hexa (Wang *et al.*, 2011) tandem repeat motifs that usually consist of 1 to 6 base pairs of nucleotides. These markers are highly polymorphic and co-dominant, reproducible and equally distributed throughout the plant genome. This repeats can be found between conserved areas in the genomes, which helps in the development of primers in order to access the information of these regions. SSRs can also be used between closely related species since they share homology of the flanking regions. New studies such as Wang *et al.*, (2011), suggest that both types of SSR, genomic (non-coding areas) and Expression Sequence Tags (EST) (transcribed sequences) can be effectively used in different analysis.

The main constraints of SSR markers can be found when a specific organism is not sequenced or previous sequences are not accessible. It is very expensive to start finding this markers because it is necessary to isolate, clone, sequence and

characterize the microsatellite loci first. Another constraint is the fact that SSR are relatively lower in plant genomes when compared to animals and humans (Hu *et al.,* 2004).

After experiments detecting these markers were successful and their capacity to examine diversity was exposed, many SSR markers have been identified in myriad of crops. Some of them had been used to analyze diversity in different plants such as apple (Hokanson *et al.*, 1998), wheat (Röder *et al.*, 1998), olives (Belaj *et al.*,2002) cassava (Chavarriaga-Aguirre *et al.*, 1998), Fregene *et al.*, 2003), Montero-Rojas *et al*., 2011), rice (Chakrararthi *et al.*, 2006), melon (Funiko *et al.*, 2007), Lima beans (Montero-Rojas *et al.,* 2013), Papaya [Carica papaya, L (Asudi *et al.*, 2013)], Wild Soybean (Bingrui *et al.,* 2013), Barley (Mikel *et al.,* 2008), Rice (Comertpay *et al.,* 2012) and in sweet potato (Buteler *et al.*, 1999; Hu *et al.*, 2004)

Literature review

Genetic diversity studies in Sweet potato

As mentioned before sweet potato is an understudied and poorly understood crop and it was not until the last decade that most of the studies regarding the genetic diversity of sweet potato started. Prakash and He (1996) studied the genetic relatedness of 30 United States sweet potato genotypes using DNA Amplification Finger Printing (DAF). This was one of the first studies to be done in order to evaluate genetic diversity in this crop. They found that the majority of the cultivars had distinctive fingerprinting patterns based on the high levels of polymorphisms found among the 30 cultivars examined. They noticed that the DAF technique was quite useful because it allowed them to differentiate between genotypes and because of this they developed specific profiles for the cultivars. For example, some cultivars named “Porto Rico” and “Creole” grouped separately from newly developed cultivars, these varieties were introduced to the US during the 19th century. New cultivars such as “Jewel” and “Carver” showed a very high degree of similarity evidencing their ancestral connection. They also noticed that cultivars which resulted of somatic mutations demonstrated high levels of genetic similarities with their parents. They concluded that the data extracted for DAF was consistent with the available pedigree information.

After it was demonstrated that DNA fingerprinting techniques using molecular marker were successful, many other studies were developed. Thompson *et al.,* (1997) studied the genetic linkages of RAPD markers in sweet potato. This study consisted of analyzing parents and their progeny in two sweet potato crosses in order to prove the feasibility of this method to study genetic linkage. In order to do this, they assessed 100 primers from which only 96 produced DNA fragments. The results showed that 134 polymorphic markers were observed and 74 of those 134 demonstrated to be useful for linkage studies because of the band pattern behavior they showed. These markers were the first linked molecular markers found in sweet potato.

Hu *et al.,* (2003) studied Inter-Simple Sequence Repeats (ISSR) in order to assess the genetic diversity present in 28 *I. batatas* cultivars and 4 wild species, part of

the CIP Series Batatas collection. They screened approximately 100 primers but only 24 of them produced scorable patterns. After analyzing the reproducibility of these markers only 8 where selected for their reproducibility and high polymorphisms. They scored a total of 81 polymorphic loci in both populations of sweet potato for the genetic variability analysis.

Other methods such as AFLPs have been used to analyze the genetic diversity of this crop. Zhang *et al.*, (2004) assessed the diversity and relationship of Latin America and the Pacific, in order to correlate them with the dispersal of the crop, testing this way the hypothesis related to the domestication of sweet potato. In places like New Zealand sweet potato is known as “kumara” which may be derived from the Quechua “kuma” leading to believe that they were introduced from Peru. According to the analysis Peru-Ecuador accessions share low genetic similarities with Pacific accessions leading to believe that the hypothesis of this distribution is incorrect and any similarities might be due to some seed transfer by birds. Nonetheless, Mesoamerican sweet potato and Pacific sweet potato share high genetic similarities suggesting a Mesoamerican origin. Elameen *et al.*, (2008) also used AFLP to analyze genetic diversity of sweet potato but using Tanzanian accessions. This study showed high diversity (0.709 mean heterozygosity) of the Tanzanian germplasm and furthermore allowed them to also to successfully discriminate duplicates.

Sweet potato is still poorly understood in part because of its polyploidy. Different studies such as those conducted by Jarret and Bowen (1994), developed some of the first SSR markers in order to analyze this crop. Still new markers were not developed until Wang *et al.*, (2013) studied the possibility of using EST SSR Markers. This type of SSR found in exons can be developed in a less expensive way than genomic SSRs because of the availability of EST libraries. Also since EST SSR are connected to coding sequences they can be helpful for targeted mapping of agriculturally important traits (Wang *et al.*, 2013). In order to develop these markers, more than 180,000 ESTs were analyzed and compared to known sequences to prevent redundancy. While testing these markers they encountered a total of 644 alleles per polymorphic loci for two varieties using 816 SSR primers. To confirm these results they added 8 more

accessions and tried the analysis for the 816 primers finding then only 342 primers to be polymorphic. These new EST SSR markers will be key to future studies of the genetic diversity, evolution and marker assisted selection in sweet potato.

Genetic diversity assessment of sweet potato trough microsatellites

The first work with SSRs in sweet potato was done by Buteler *et al.*, (1999). In this study they evaluated inheritance and nucleotide sequence profiles of the microsatellites found in sweet potato. They analyzed these markers in hexaploid sweet potato and possible ancestors of this plant which vary from diploid to tetraploids. They isolated 63 microsatellite loci and designed primers in order to characterize them. Of those 63 SSR isolates only 9 exhibited a clearly visible and hereditable pattern while the another 5 markers demonstrated Mendelian segregation. When analyzed, the sequences of the three species (*I. batatas, I. trifida (2x)* and *I.trifida (4x))* showed differences in the banding pattern. This might have been caused by mutations in the repeat regions or in the flanking regions of both polyploid and diploid species. Some of their findings demonstrated that variations found in the allele size were caused by changes in the repeats. This variation was triggered by insertion/deletion mutations in the flanking regions. Some polysomic segregation ratios were also noticed in four microsatellite loci, which was the first report of this kind of segregation in polyploids. They concluded that complex genetic mechanisms are the cause of the allele variations in the Ipomoea genus and that SSR markers are in tetrasomic inheritance in sweet potato.

Once SSR markers where developed in sweet potato, many other researchers started analyzing the diversity of this crop in different locations. Zhang *et al.*, (2000) used SSR markers to analyze the genetic diversity of sweet potato varieties from Latin America. Their aim was to understand the distribution and diversity of a germplasm available at Centro Internacional de la Papa (CIP). This collection hold 5526 cultivated accessions from 57 countries. In this study 113 accessions from different locations throughout Latin American, from Mexico to Peru, not previously studied were analyzed. In order to examine this diversity 12 SSR markers were used and each SSR variant was considered an allele. From the 12 markers only 6 displayed a reproducible pattern. A

total of 14 specific alleles were found per region, 8 of these alleles can be found in Mesoamerica, 4 in Colombia-Venezuela and the rest in Peru-Ecuador. From their results they observed that actual heterozygosity varies greatly among regions, Mesoamerica showed an actual heterozygosity of 0.714, Venezuela-Colombia 0.705 and Peru Ecuador 0.521 suggesting that Mesoamerica could be the possible primary center for diversity of this crop.

Since the use or SSR to analyze genetic diversity was very successful in sweet potato other studies such as the application of SSR in determining the genetic relationships of cultivars used for polycross breeding in Taiwan (Hwang *et al.,* 2002). This analysis used 8 primers to study diversity and relationship among Chinese and Japanese plants and cultivars obtained from different types of breeding in Taiwan. The SSR analysis showed a total of 20 alleles from which 17 were polymorphic, the average expected heterozygosity was 0.55, concluding that the plants used to do the polycross breeding possessed high levels of genetic variability.

Buteler *et al.*, (2002) used SSR markers to analyze paternity in polyploid sweet potato, evaluating two experimental populations, with known paternity using 2 markers. One of the populations (CIP) had more success than the second one Louisiana Agricultural Experimental Station (LAES), because in the CIP population 7 out of eight progeny were correctly matched with its parent while in the LAES population only 3 out of 12 worked. This could have happened because the LAES cultivars have a lesser genetic variability than the CIP cultivars which have more diverse samples. They concluded that even if you have few loci you can still allocate paternity demonstrating once again how useful are SSR markers in genetic studies.

Even though Buteler *et al.*, (1999) did a characterization of SSR markers in sweet potato more markers needed to be developed thus Hu *et al.*, (2004) in two different publications developed and characterized new SSR markers. First they used *I. trifida* sequences in order to develop new markers while wanting to test the transferability of these markers to *I. batatas* and other wild species (Hu *et al.*, (2004)). They identified 15 more microsatellites from which 12 were successful to amplify scorable loci. In addition they found that the microsatellites were a 100% transferable

between *I. trifida* and *I. batatas* but 83% or less to other wild relatives including *I. tiliacea* and *I. triloba*. This high transferability suggest high levels of conservation throughout evolution, also these results demonstrate that the transferability is due in part to genetic relatedness making them useful tools not only to analyze diversity also to study synteny. In the second study developing markers Hu *et al.*, (2004) screened a small-insert genomic library, then constructed microsatellites enriched libraries and lastly searched EST database. With the help of the small insert genomic library they found 46 microsatellites, from which they classified 20 as perfect, 8 as imperfect and 16 as compound. They built an enriched library of 800 clones and then screened for microsatellites. They found that 93 microsatellites could be detected and further analyzed. Lastly for the EST database they studied sequences from more than 4000 sweet potato ETS's and with this data they created 151 primer sets. Of these sets of primers only 75 of them were polymorphic and informational enough to distinguish between sweet potato cultivars.

Veasey *et al.*, (2008) assessed the genetic diversity of Brazilian sweet potato. They analyzed 78 sweet potato accessions from which 58 were landraces and 20 were putative clones from 19 different communities in Valle do Ribeira, Brazil, using 8 SSR markers. These markers were visualized on a 6% polyacrylamide gel stained with silver nitrate and genotyped based on absence/presence of bands. In order to analyze the results they performed an analysis of molecular variance (AMOVA), dendrograms and a Principal Component Analysis (PCA). With the 8 markers they scored a total of 46 polymorphic bands plus a total of 48 bands with an average of 6 bands per primer, which ranged from 3 to 10 bands. These results showed high levels of diversity, verified with the results of the AMOVA, indicating that the largest impact on the diversity was due to differences between varieties within households accounting for a 58%. Jaccard similarity coefficient indicated variation levels up to 69% demonstrating how traditional farmers maintain high levels of genetic variation in their households.

Arizio *et al.*, (2009) did a genotypic identification and diversity evaluation of a collection of 57 accessions of sweet potato found in a Gene Bank of INTA, Argentina. The samples used include commercial varieties, breeding clones, foreign material and

primitive materials from primary dispersion zones. For the analysis, they used sixteen set of primer amplifications separated in a 6% denaturing polyacrylamide gel and detected using silver staining. They created a similarity matrix using Jaccard coefficient, cluster analysis using UPGMA, genetic variation and diversity were also evaluated with the results. Seventy three polymorphic bands were identified which helped to discriminate 52 out of 57 accessions with a 92% of fidelity. Similarity analysis confirmed representative samples with a broad genetic base of well-preserved sweet potato germplasm. Regarding the cluster analysis, for each cluster, a predominant skin color can be observed and high genetic similarity was found (≥0.7). According to Shannon's diversity index values variation within regions accounted for 97% of the whole variance, demonstrating the varied composition of this Argentinian germplasm.

Africa, especially the Sub Saharan Africa (SSA), is known to be a second center for diversity of sweet potato (Zhang *et al.*, 1998). A characterization of Ugandan sweet potato was done in by Yada *et al.,* (2010) using 192 superior genotypes from a germplasm kept at The National Crops Resources Research Institute (NaCRRI) and analyzed with ten highly polymorphic SSR markers. They found an average of 4 alleles per locus and a rather high genetic variability (0.57). They found that some accessions were different morphologically but still grouped as clones ratifying once again the advantage of molecular characterization of a crop. Also Karuri *et al.,* (2010) evaluated the diversity of Kenyan sweet potato using 89 genotypes analyzed with 6 SSR markers and morphological analysis of the plant. This morphological analysis included characteristics from the vines, leaves and roots. An Analysis of Variance (ANOVA) of these characters showed that most characters analyzed were significantly different and the mean observed heterozygosity was of 0.75. The comparison between morphological traits and molecular markers were very vague. The authors suggest that this could be caused by the independent nature of molecular markers and morphological traits.

Tumwegamire *et al.*, (2011) studied the genetic diversity in white-(WFSP) and orange-fleshed (OFSP) sweet potato. They also found a large diversity between the accessions. Analyzing the data, they found some duplicates among the OFSP but all of them shared the similar flesh color, country or origin. In the cluster analysis some of the

accessions from the different groups clustered together even though some of them exhibit different flesh color. These results suggest the probability that OFSP accessions evolved from sisters WFSP as opposed to evolving only from introduced OFSP accessions. The researchers concluded that the varieties, regardless of flesh color, are separated based on location.

SSR markers are also useful to characterize genotypes resistant to many different biological and environmental stresses and one example of this is the Sweet potato virus disease (SPVD). In order to characterize these genotypes Karuri *et al.*, (2009) analyzed 89 genotypes with an average symptom severity using 6 SSR markers. The results demonstrated once again how polymorphic are microsatellite markers, 5 out of 6 markers were highly discriminatory therefore highly informative and the results showed that in spite of the SPVD there is a significant amount of genetic variability.

Gwandu *et al.*, (2012) characterized SPVD resistance in Tanzanian elite sweet potato. They evaluated 57 genotypes which were chosen by their promising reaction to SPVD infection and dry matter content with 4 SSR markers. These showed high polymorphisms and the number of alleles per locus range from 11 to 22 for the 4 SSR. These results suggest high genetic diversity which they attributed to the polyploidy and outcrossing nature of sweet potato.

Importance of the conservation of genetic resources

With more than 900 million people dying of famine and hunger today, actions need to be taken quickly to reverse this trend. Most people in developing countries do have food to eat, but most of these foods provide them with empty calories. Poor quality nutrition and lack of food diversity fuels hunger and malnourishment. Close to 850 million people suffered from undernourishment in 2006-07 and 870 million in 2010-12 according to the “State of food security of the world” 2012. The protection of different natural resources is extremely necessary in order to fulfill the needs of our growing population. According to FAO (2011) by 2050 due to the increase in population the need for food will be twice as much as today, while having less agricultural land and resources. One way to try and secure food sources for future generations is to protect

plant genetic resources which are one of the foundations for food security. The natural occurring diversity in plants is crucial to maintain food sources for millions of people. Plant diversity, to an extent, is kept by farmers. They share plant material, introduce new varieties and even select stronger and better genotypes. Also they keep, and in some cases without knowing protect wild relatives of many different species either for food or to feed their farm animals.

According to FAO 2010, the second report on “The state of the world’s plant genetic resources for food and agriculture” plant genetic resources and their adaptability can act as a defense against detrimental environmental conditions that might eradicate diversity. Genetic erosion, which can be defined as the loss of genetic diversity, either if it is genes from an individual or gene combinations or even the disappearance of a variety, can cause a reduction in diversity. This loss will be responsible for damaging the already delicate food security. To prevent this, conservation programs need to be established. Analyzing the diversity can lead us to find better information regarding the accessions/cultivars being grown in the field, what needs to be treated more effectively and which resources need to be conserved more diligently so that we can protect species that might be under-used and that could play an important role in the future of food security. The conservation of the diversity and resources is, and will be, crucial to meet the need for food that affects millions of people every year. It will give, not only food security, it will also help alleviate poverty in developing countries.

Considering all the above, the objectives of our research are:

I. To assess the genetic diversity of the current Puerto Rican germplasm. .

II. To analyze the genetic diversity of unknown sweet potato accessions from across the island.

III. To compare the genetic characteristics of known sweet potato collection from Puerto Rico, economically important varieties from the US and the unknown accessions grown by small-scale farmers and landowners from across the island.

Chapter II

MATERIALS AND METHODS

Plant Material

The plant material for this study, consisting of leaf tissue, was obtained from 3 different sources. A group comprising 8 known Puerto Rican accessions was collected from the Agricultural Experimental Station in Gurabo, PR. Ten samples, comprised of 9 known US commercial varieties and one known Puerto Rican landrace was obtained from The Plant Genetic Resources Conservation Unit (PGRCU) in Griffin, GA. A total of 137 samples from across the island were collected by undergraduate students as a part of a laboratory module on sweet potato which was implemented in the genetics laboratory of the University of Puerto Rico, Mayaguez Campus. The leaf samples (Table 3) of these unknown accessions were collected and information, such as owner name, geographical location (municipality, community), color and name of the sample (if known) were retrieved. Leaf material was kept at -20°C until DNA was extracted. Unknown samples were then separated geographically into 5 areas of Puerto Rico (North, South, West, East and Center) (figure 5).

DNA Extraction

Genomic DNA extractions were performed using leaf tissue according to CTAB method (Doyle and Doyle, 1990) modified in our laboratory. A small piece of leaf tissue, approximately 30 mg, of each sample was grounded using a sterile pestle and sea sand in a 2.0 mL tube for 3 minutes. After addition of 700µL of 2% CTAB extraction Buffer [20 mM EDTA, 0.1 M Tris-HCl pH 8.0, 1.4 M NaCl, 2%CTAB], leaf material was further ground for 2 minutes and mixed by inversion. The solution was then incubated at 70ºC for 30 min. After incubation, 500 µL of a chloroform-isoamylalcohol (24:1) solution was added to the tubes and gently mixed by inversion for 30 seconds. Samples were then centrifuged for 4 minutes at 13,200 rpm, 500 µL of the supernatant was transferred to a fresh 2.0 mL tube. To the supernatant 500µL of the chloroform-isoamylalcohol (24:1) and 200 µL 2% CTAB buffer were added and the solution was mixed gently by

inversion. The mixture was then centrifuged for 4 minutes at 13,200 rpm and the supernatant was transferred to a fresh 1.5mL tube with 700 µL of cold isopropanol (-20ºC). Samples were gently mixed by inversion and centrifuged at 13,200 rpm for 4 minutes. After centrifugation, the supernatant was discarded the resulting pellet was air-dried for 5 minutes. The pellet was then washed with 700 µL of cold 70% ethanol (-20ºC) to clean the DNA. The solution was then vortexed and centrifuged at 13,200 rpm for 4 min, the ethanol was discarded and the pellet was air-dried for 5 minutes. The DNA was then re-suspended in 150 µL TE 10:1 buffer (10 mM Tris**-**HCl pH 8.0, 1 mM EDTA pH 8.0) plus 4 µL of ribonuclease (RNAse 10 mg mL–1) in each tube and was incubated at 65ºC for 8minutes prior to storage at -20ºC. DNA was then quantified using a Nanodrop ND-1000 spectrophotometer. Samples were then diluted to 10ng/µL with deionized distilled water for SSRs amplifications.

SSR Markers and PCR Conditions

A total of 23 SSR markers (Table 4) were selected for this study taking into consideration their PIC value and use in previous related studies (Tumwegamire *et al.,* 2011; Yada *et al., 2010)*. In the PCR Reaction a 3 primer protocol is used: the forward (with and M13 tail), reverse and fluorescently labeled M13. In the first cycles the forward and reverse attach to the DNA and start amplification. After that the M13 primer takes the place of the forward primer, by this time M13 is incorporated to the PCR product which will be fluorescently labeled. The PCR reaction was prepared for a total volume of 25 µL. For 21 of the SSR, the reaction was as follows: 5 µL of 5X reaction buffer, 7 µL of 25 mM $MgCl_2$, 2 µL of 50 mM KCl, 1 µL of 10 mM ddNTP's, 1.25 µL of 2 µM from each forward and reverse primers, 0.5 µL M13 F/d 1pmol/µL fluorescent primer (LI-COR bioscience), 1 U Taq polymerase, 2 µL of 10 ng/µL DNA template and 4.8 µL of ddH_20. For SSR R03 the reaction consisted of: 5 µL of 5X reaction buffer, 9 µL of 25 mM $MgCl_2$, 2 µL of 50 mM KCl, 1 µL of 10 mM ddNTP's, 1.25 µL of 2 µM from each forward and reverse primers, 0.5 µL M13 F/d 1pmol/µL fluorescent primer (LI-COR bioscience), 1 U Taq polymerase, 1 µL of 30 ng/µL DNA template and 3.8 µL of ddH_20. For SSR R08 the reaction was: 5 µL of 5X reaction buffer, 8 µL of 25 mM $MgCl_2$, 2 µL of 50 mM KCl, 1 µL of 10 mM ddNTP's, 1.25 µL of 2 µM from each forward and reverse primers,

0.5 µL M13 F/d 1pmol/µL fluorescent primer (LI-COR bioscience), 1 U Taq polymerase, 2 µL of 40 ng/µL DNA template and 3.8 µL of ddH_20.

PCR conditions for amplification were: 94°C for 7 minutes followed by 35 cycles of 94°C for 30s, 50-63°C for 01:30s of annealing temperature depending on the SSR, 72°C for 1s, with a final extension time of 10 min at 72°C. The PCR products were then visualized on a 1.0% agarose gel for amplification confirmation and then visualized on a 6.5% denaturing polyacrylamide gel in a LI-COR 4300 automated DNA sequencer (Figure 6A). Product dilution, polyacrylamide gel preparation and electrophoresis conditions are explained in Appendix A.

Statistical Analysis

In order to analyze the data, band size information from each SSR was scored by hand based (Figure 6B) on the weight of the molecular marker 50-350 Size Standard IRDYE 700. The scored data for the 23 SSR and the 155 samples was then converted into a weight matrix in Microsoft Excel (Figure 6C). From this data, a binary matrix was constructed (Figure 6D) based on presence and absence of alleles (1, 0) which was used to analyze genetic dissimilarity among groups. To analyze these dissimilarities an Unweighted Pair Group Method with Arithmetic Mean (UPGMA) cluster method of Euclidean genetic distances from the NTSYS-PC program was used (Rholf, 2008).

Based on the Microsoft Excel matrix a new matrix was transcribed into a text format using TextPad ver. 6.2.2 with special codifications for GenoDive/GenoType (Figure 7 A-B) Software for analysis of population genetic data (Meirmans and Van Tienderen, 2004) in order to assess Genetic Diversity estimators, Allele Frequencies, Genetic Distances and Analysis of Molecular Variance (AMOVA). Preparation of the input matrix can be found in Appendix B.

Table 3: Name and information of the unknown samples

Id Number	Location	Given name	Flesh Color	Id by Location
1	Añasco	Batata Lila	Purple	1W
2	Mayaguez	Mameya	Yellow	2W
3	Isabela	N/A	White	3W
4	Guaynabo	Mameya	Yellow	1N
5	Camuy	Mameya	Yellow	2N
6	Vega Alta	Batata Lila	Purple	3N
7	Lares	N/A	White	1C
8	Lajas	N/A	White	4W
9	Jayuya	Batata Lila	Purple	2C
10	Añasco	N/A	White	5W
11	Caguas	N/A	N/A	3C
12	Aguadilla	Mameya	Yellow	6W
13	Aguadilla	N/A	White	7W
14	Moca	N/A	N/A	8W
15	Cidra	N/A	N/A	4C
16	Aguada	Mameya	Yellow	9W
17	Gurabo	N/A	White	1E
18	Aguadilla	Mameya	Yellow	10W
19	Cabo Rojo	N/A	White	11W
20	Aguada	N/A	White	12W
21	Aguada	N/A	White	13W
22	Aguada	N/A	White	14W
23	Aguada	Batata Blanca	White	15W
24	Sabana Grande	Batata Boba	White	16W
25	Isabela	Mameya	Yellow	17W
26	Barceloneta	Mameya	Yellow	4N
27	Moca	Mameya	Yellow	18W
28	San German	Mameya	Yellow	19W
29	Utuado	N/A	White	5C
30	Morovis	N/A	White	6C

31	Morovis	N/A	N/A	7C
32	Morovis	N/A	White	8C
33	Morovis	N/A	White	9C
34	Morovis	N/A	White	10C
35	Morovis	N/A	White	11C
36	Morovis	N/A	White	12C
37	Morovis	N/A	White	13C
38	Sabana Grande	Mameya	Yellow	20W
39	Añasco	N/A	N/A	21W
40	Lares	N/A	White	14C
41	Mayaguez	Mameya	Yellow	22W
42	Mayaguez	N/A	White	23W
43	Mayaguez	N/A	N/A	24W
44	Manati	N/A	N/A	5N
45	Isabela	Mameya	Yellow	25W
46	Rincon	N/A	N/A	26W
47	Mayaguez	N/A	N/A	27W
48	Bayamon	Mameya	Yellow	6N
49	Hormigueros	N/A	White	28W
50	Lajas	N/A	White	29W
51	San German	Mameya	Yellow	30W
52	Lares	N/A	N/A	15C
53	Lares	N/A	N/A	16C
54	Lares	N/A	N/A	17C
55	Lajas	N/A	N/A	31W
56	Lajas	N/A	N/A	32W
57	Quebradillas	Arecibeña	White	33W
58	Quebradillas	N/A	Yellow	34W
59	Quebradillas	N/A	Yellow	35W
60	Aguada	N/A	N/A	36W
61	Cabo Rojo	N/A	N/A	37W
62	Mayaguez	Del Pais	N/A	38W

63	Gurabo	N/A	Yellow	2E
64	Ciales	N/A	White	18C
65	Corozal	N/A	N/A	19C
66	Mayaguez	N/A	White	39W
67	Moca	Batata Blanca	White	40W
68	San Juan	N/A	White	7N
69	Aibonito	N/A	N/A	20C
70	Mayaguez	N/A	N/A	41W
71	Camuy	N/A	N/A	8N
72	Guanica	N/A	N/A	42W
73	Moca	Batata Lila	Purple	43W
74	Sabana Grande	N/A	N/A	44W
75	Mayaguez	N/A	N/A	45W
76	San Sebastian	Batata Blanca	White	46W
77	Hatillo	Boniato	White	9N
78	Corozal	La Santa	Yellow	21C
79	San German	N/A	N/A	47W
80	San Sebastian	Batata Blanca	White	48W
81	Barranquitas	N/A	White	22C
82	Moca	N/A	White	49W
83	Añasco	N/A	White	50W
84	Aguada	Mameya	Yellow	51W
85	Cabo Rojo	Mameya	Yellow	52W
86	Barranquitas	N/A	White	23C
87	Aguada	N/A	Purple	53W
88	Corozal	Mameya	Yellow	24C
89	Isabela	N/A	White	54W
90	Isabela	Mameya	Yellow	55W
91	Utuado	N/A	N/A	25C
92	Aguada	Mameya	Yellow	56W
93	Comerio	Mameya	Yellow	26C
94	Carolina	N/A	White	10N

95	Rio Grande	Mameya	Yellow	3E
96	Aguadilla	N/A	White	57W
97	Camuy	Mameya	Yellow	11N
98	Isabela	Mameya	Yellow	58W
99	Patillas	Mameya	Yellow	1S
100	San Sebastian	N/A	White	59W
101	Mayaguez	N/A	N/A	60W
102	Isabela	Mameya	Yellow	61W
103	San Sebastian	Candela	White	62W
104	Moca	N/A	White	63W
105	Mayaguez	N/A	White	64W
106	Mayaguez	Mameya	Yellow	65W
107	Jayuya	Batata Blanca	White	27C
108	Isabela	Batata Amarilla	Yellow	66W
109	Aguada	N/A	N/A	67W
110	Aguada	N/A	N/A	68W
111	Cabo Rojo	Mameya	Yellow	69W
112	Utuado	Mameya	Yellow	28C
113	Cabo Rojo	N/A	White	70W
114	Caguas	Mameya	Yellow	29C
115	Sabana Grande	Mameya	Yellow	71W
116	San German	Mameya	Yellow	72W
117	Isabela	N/A	N/A	73W
118	Moca	N/A	White	74W
119	Bayamon	N/A	N/A	12N
120	Bayamon	N/A	White	13N
121	San German	Mameya	Yellow	75W
122	Añasco	N/A	N/A	76W
123	Yauco	Mameya	Yellow	2S
124	Manati	Batata Blanca	White	14N
125	Camuy	Mameya	Yellow	15N
126	Moca	N/A	White	77W

127	Guanica	N/A	White	78W
128	Aguadilla	N/A	White	79W
129	San Sebastian	N/A	White	80W
130	Vega Baja	N/A	N/A	16N
131	Lares	Mameya	Yellow	30C
132	Yauco	N/A	N/A	3S
133	Coamo	N/A	N/A	4S
134	Barranquitas	N/A	N/A	31C
135	Camuy	Mameya	Yellow	17N
136	Las Marias	N/A	N/A	81W
137	San Juan	N/A	White	18N

Table 4.Summary of the 23 SSR Markers, M13 primer Tail, Annealing temperature, Motif and Allele size

Name	M13 5'-Tail	Forward Primer	Reverse Primer	Annealing	Motif	Allele Size
IB242	cacgacgttgtaaaacgac	gcggaacggacgagaaaa	atggcagagtgaaaatggaaca	58	(ct)3ca(ct)11	136-155
IB297	cacgacgttgtaaaacgac	gcaatttcacacacaaacacg	cccttcttccaccactttca	58	(ct)13	150-182
IB324	cacgacgttgtaaaacgac	tttggcatgggcctgtatt	gttcttctgcactgcctgattc	56	-	136-152
IBCIP-1	cacgacgttgtaaaacgac	cccacccttcattccattact	gaacaacaacaaaaggtagagcag	63	(acc)7a	155-167
IBS10	cacgacgttgtaaaacgac	ctacgatctctcggtgacg	cagcttctccactccctac	60	(ct)12	307-337
IBS11	cacgacgttgtaaaacgac	ccctgcgaaatcgaaatct	ggacttcctctgccttgttg	58	(ttc)10	254-305
IBR13	cacgacgttgtaaaacgac	gtaccgagccagacaggatg	cctttgggattggaacacac	58	(ttc)6	225-298
IBS17	cacgacgttgtaaaacgac	cagaagagtacgttgctcag	gcacagttctccatcctt	58	(gga)4	182-204
IBS18	cacgacgttgtaaaacgac	ctgaacccgacagcacaag	gggaagtgaccggacaaga	58	(tagc)4	296-298
IBR21	cacgacgttgtaaaacgac	gacagtctccttctcccata	ctgaagctcgtcgtcaac	58	(gac)5	181-207
IBC12	cacgacgttgtaaaacgac	tctgagcttctcaaacatgaaa	tgagaattcctggcaaccat	56	(ttc)6	110-134
J175	cacgacgttgtaaaacgac	atctatgaaatccatcactctcg	actcaattgtaagccaaccctc	58	(aatc)4	133-149
J10a	cacgacgttgtaaaacgac	tcaaccactttcattcactcc	gtaattccaccttgcgaagc	58	(aag)6	191-225
J67	cacgacgttgtaaaacgac	cacccatttgatcatctcaacc	ggctctgagcttccattgttag	58	(gaa)5	191-217
J116a	cacgacgttgtaaaacgac	tcttttgcatcaaagaaatcca	cctcagcttctgggaaacag	58	(cct)6	207-251
IB1809	cacgacgttgtaaaacgac	cttctcttgctcgcctgttc	gatagtcggaggcatctcca	60	(cct)6(ccg)6	144-155
IBJ544b	cacgacgttgtaaaacgac	agcagttgaggaaagcaagg	caggatttacagccccagaa	61-62	(tct)5	191-214
IBS01	cacgacgttgtaaaacgac	tcctccaccagctctgattc	ccattgcagagccatacttg	56	(aga)10	233-268
IB-R16	cacgacgttgtaaaacgac	gacttccttggtgtagttgc	agggttaagcgggagact	60	(gata)4	201-203
IB-R19	cacgacgttgtaaaacgac	ggctagtggagaaggtcaa	agaagtagaactccgtcacc	60	(cag)5	190-208
IB-R03	cacgacgttgtaaaacgac	gtagagttgaagagcgagca	ccatagacccattgatgaag	58	(gcg)5	243-258
IB-R08	cacgacgttgtaaaacgac	ggcgacaccttagagtat	caccccccctattcacaa	50	(t3a)4	204-216
S07	cacgacgttgtaaaacgac	gcttgcttgtggttcgat	caagtgaagtgatggcgttt	60	(tgtc)7	193-211

Figure 5: Map of the distrubution the island in groups. Yellow: North, West:Blue, South: Green, Center Red and East: Purple. Total samples per area 81W, 3E, 31C, 4S and 18N.

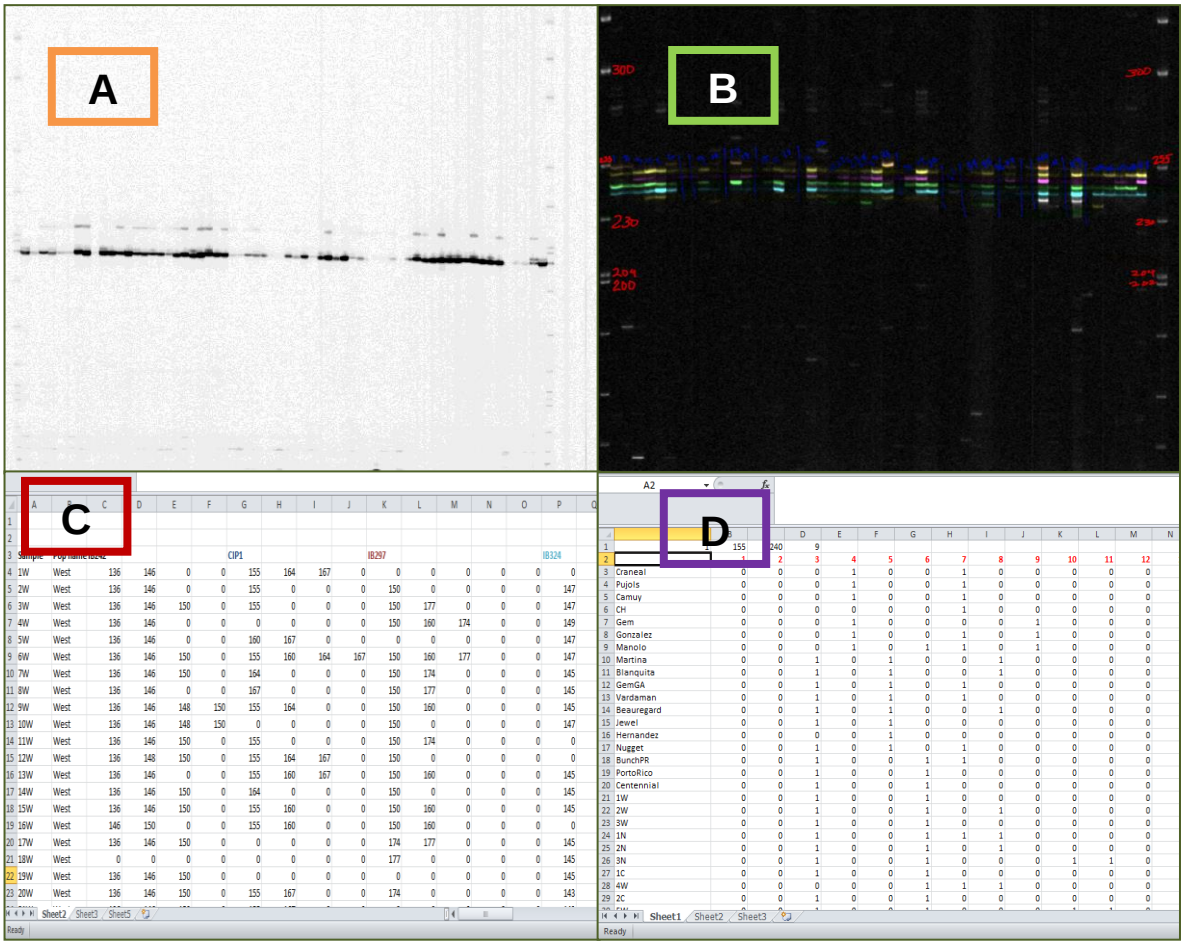


Figure 6 (A-D): Figure 6A, Polyacrylamide gel shown in LICOR 4300 DNA Analyzer. 6B, Polyacrylamide gel picture for band weight analysis. 6C, Weight Matrix. 6D, Binary Matrix.

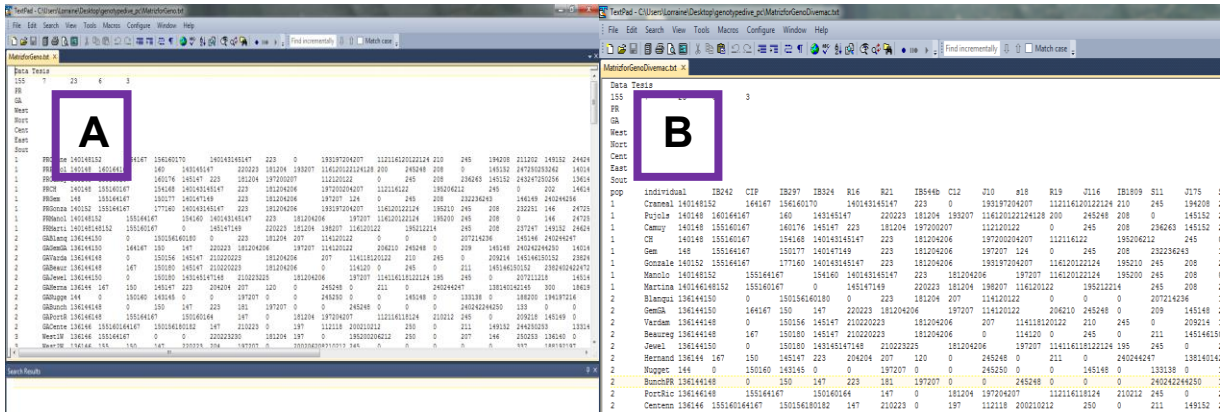


Figure 7(A-B): 7A, matrix prepared on TextPad for GenoType, 7B, matrix prepared on TextPad for GenoDive.

Chapter III

RESULTS

Genetic Diversity

Results of the analysis of the 23 loci showed a total of 205 alleles in 155 sweet potato samples, ranging from 2 to 20 alleles per locus (Table 5, Figure 8, Figure 9). The average of alleles per loci was 8.9 ± 1.034. The locus with the highest number of alleles was S11 with a total of 20 alleles while R19 had the lowest number of alleles of 2. The number of alleles per SSR with the highest frequency was 8 which were discerned by 5 SSR markers (Figure 9).

Table 5: Total alleles per locus

Locus	Allele Number	Locus	Allele Number
IB242	8	**IB1809**	9
CIP	4	**S11**	20
IB297	13	**J175**	10
IB324	8	**S10**	7
R16	6	**S17**	11
R21	4	**J67**	8
IB544b	8	**R13**	17
C12	10	**S01**	5
J10	13	**S07**	4
s18	8	**R08**	6
R19	2	**R03**	5
J116	19	**Total**	205

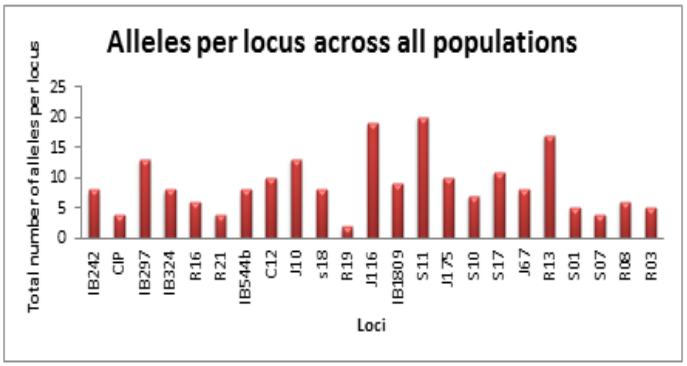


Figure 8: Total allele count per locus across populations

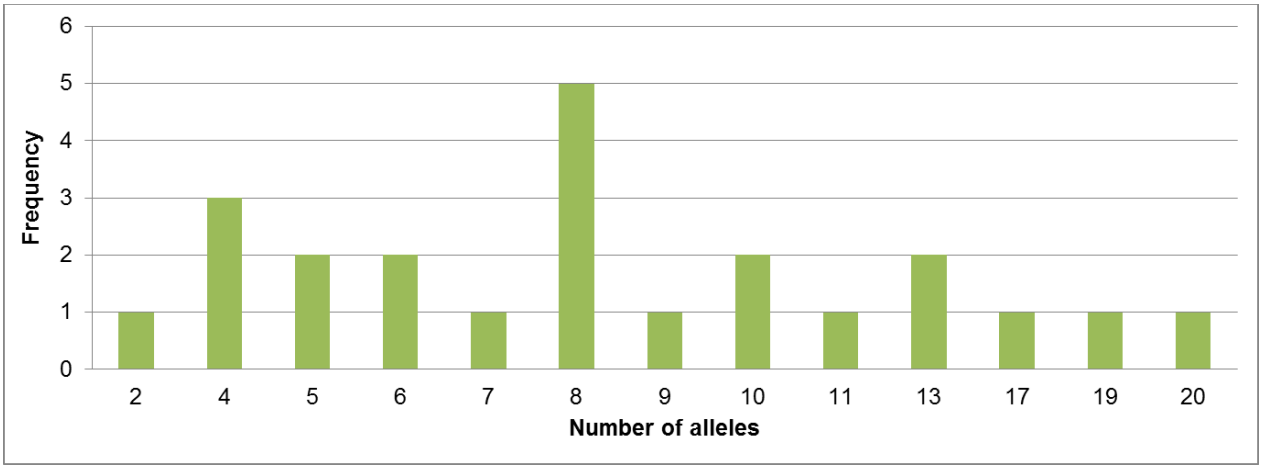


Figure 9: Frequency of the number or alleles across the 23 SSR Markers

Overall average observed heterozygosity (H_o) was high across populations with a value of 0.637 ± 0.05 (Table 6). Measurements of total heterozygosity (H_t) revealed a large genetic diversity throughout the populations with a value of 0.731 ± 0.031, which ranged from 0.076 to 0.875 per loci. The heterozygosity within populations (H_s) was 0.694 ± 0.034 revealing high levels of genetic diversity in the populations. The results for inbreeding coefficient (Gis) was low (0.082 ± 0.038) but suggest some level of inbreeding across the populations. Whereas the Gst fixation index of 0.1 ± 0.025 shows a moderate level of genetic differentiation between populations (Table 6).

Table 6: Summary of genetic diversity indexes

Statistic	Mean	Std.Err.	c.i.2.5%	c.i.97.5%	Description
Num	8.910	1.034	7.261	11.217	Number of alleles
Eff_num	3.132	0.286	2.617	3.708	Effective number of alleles
Ho	0.637	0.05	0.54	0.728	Observed Heterozygosity
Hs	0.694	0.034	0.625	0.756	Heterozygosity Within Pops
Ht	0.731	0.031	0.667	0.785	Total Heterozygosity
Gis	0.082	0.038	0.015	0.16	Inbreeding coefficient
Gst	0.1	0.025	0.057	0.15	Fixation index

Standard errors of F-statistics were obtained through jackknifing over loci, 95% confidence intervals of F-statistics were obtained through bootstrapping over loci, P-values were calculated using G'st(Nei)

The genetic diversity estimators per locus (Table 7) show very variable levels of H_o fluctuating from 0.1 to 0.9, although most loci gave high heterozygosity information. Levels of inbreeding coefficient fluctuated from negative values to 0.5 suggesting different levels of inbreeding assessed by varied loci.

Table 7. Genetic Diversity indexes per Loci

Locus	Ho	Hs	Ht	Gis	Locus	Ho	Hs	Ht	Gis
IB242	0.878	0.772	0.835	-0.137	**IB1809**	0.709	0.74	0.737	0.042
CIP	0.805	0.76	0.736	-0.058	**S11**	0.964	0.888	0.882	-0.086
IB297	0.712	0.792	0.806	0.102	**J175**	0.704	0.796	0.833	0.115
IB324	0.613	0.718	0.731	0.145	**S10**	0.258	0.625	0.787	0.587
R16	0.473	0.592	0.627	0.201	**S17**	0.933	0.887	0.889	-0.051
R21	0.71	0.648	0.605	-0.095	**J67**	0.92	0.825	0.814	-0.114
IB544b	0.713	0.704	0.684	-0.013	**R13**	0.64	0.818	0.859	0.217
C12	0.766	0.772	0.851	0.008	**S01**	0.343	0.522	0.619	0.343
J10	0.849	0.832	0.835	-0.021	**S07**	0.531	0.542	0.687	0.021
s18	0.451	0.563	0.705	0.198	**R08**	0.741	0.7	0.656	-0.059
R19	0.143	0.25	0.25	0.429	**R03**	0.151	0.362	0.516	0.582
J116	0.637	0.852	0.867	0.252	**Overall**	0.637	0.694	0.731	0.082

Genetic diversity in the populations

A total of 155 samples were divided into 7 groups: 1) PR group had 8 samples from a collection at the Gurabo Experimental Station; 2) 8 samples from USDA Repository in Griffin, GA; 3) Eighty one samples were collected from the West side; 4) 18 from the North; 5) 31 from the Center; 6) 3 from the East: and 7) 4 from the South (Figure 5). All populations showed high levels of observed heterozygosity with values ranging from 0.577 to 0.717 (Table 8).

The populations with the highest observed heterozygosity were the ones composed by samples from the East and Center areas of the islands, although these

results might be inflated because of a low amount of highly heterozygous samples. The population with the highest average number of alleles was the West (7.04) followed by the Center (6); both of them have the highest number of samples. The inbreeding coefficient was relatively low in most populations, aside from the North which had a high G_{is} value of 0.189.

Table 8: Indexes of genetic diversity per population

Population	Num	Eff_num	Ho	Hs	Ht	Gis
PR	4.957	3.98	0.7	0.679	0.679	-0.031
GA	4.682	3.59	0.709	0.72	0.72	0.014
West	7.045	4.258	0.66	0.73	0.73	0.096
North	5.091	3.652	0.577	0.711	0.711	0.189
Center	6	4.017	0.71	0.715	0.715	0.007
East	3.55	3.284	0.717	0.701	0.701	-0.023
South	2.875	2.588	0.599	0.663	0.663	0.097

The allelic information in the populations showed that the population with the highest percentage of alleles per locus was the West with 51% of the alleles and the lowest were the East and South both with 2% (Figure 10). Looking specifically at allele information per locus using Locus IB297 (Figure11), IB242 (Figure 12) and J116 (Figure 13) as examples we can see the fluctuation between populations. As expected, the population with the highest amount of samples has the highest number of alleles.

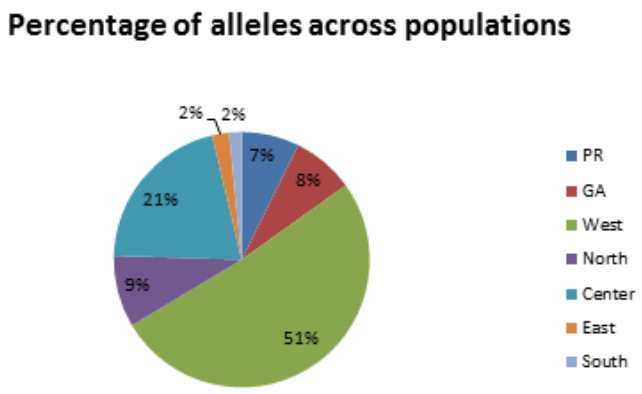


Figure 10: Percentage of alleles across populations

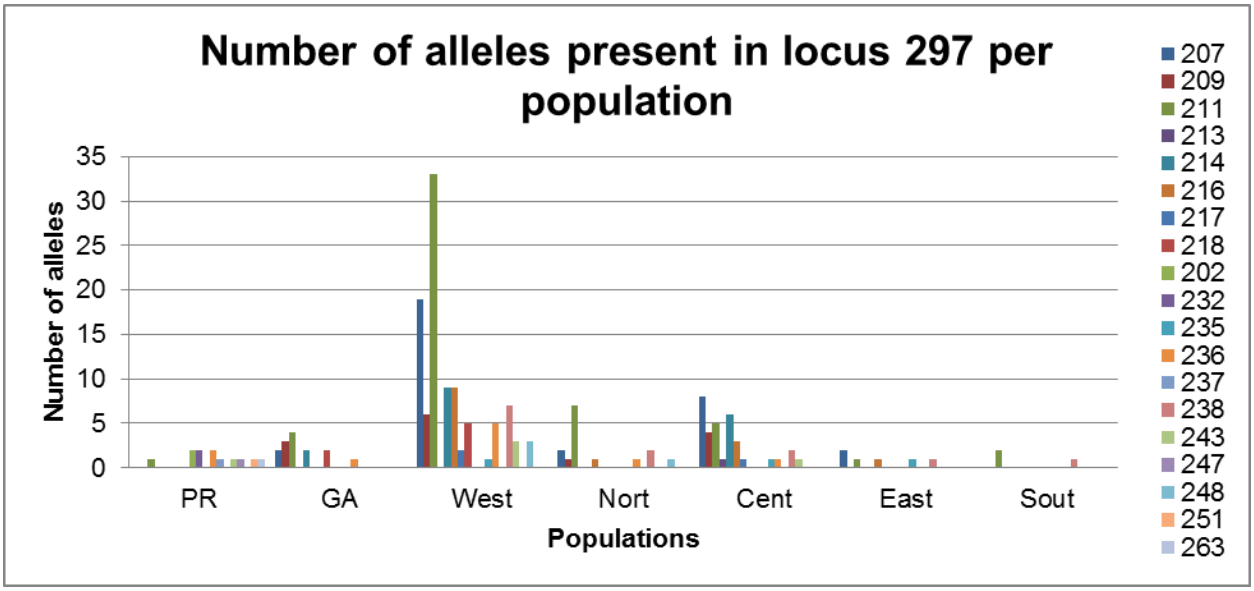


Figure 11: Alleles across populations in locus IB297

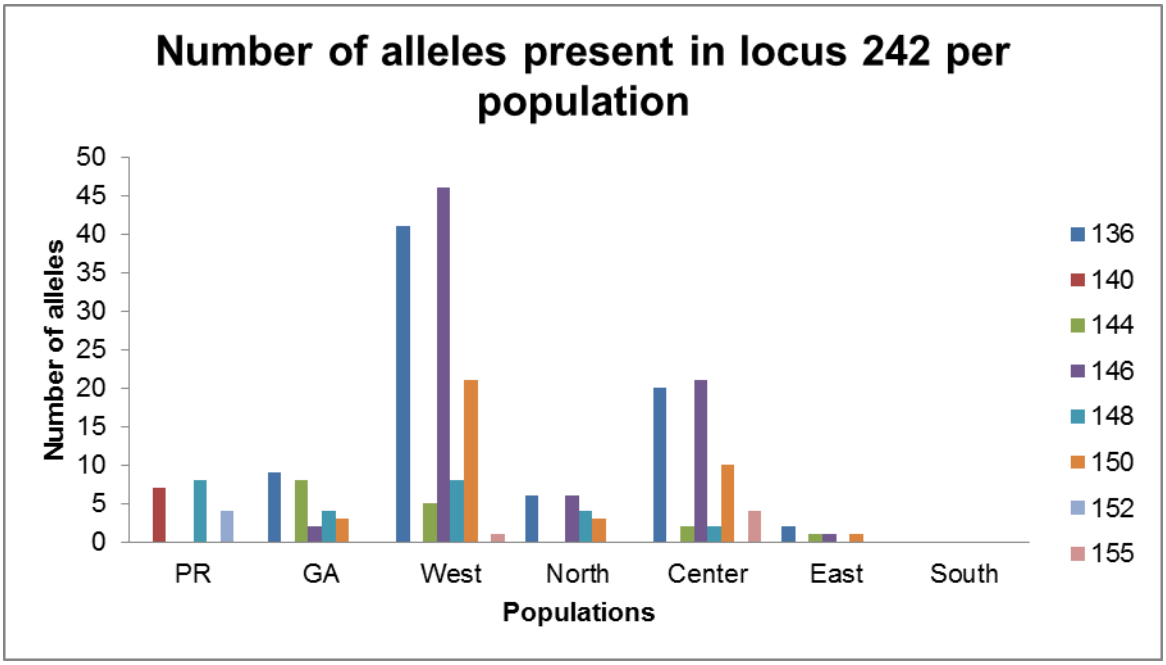


Figure 12: Percentage of alleles across populations in locus IB242

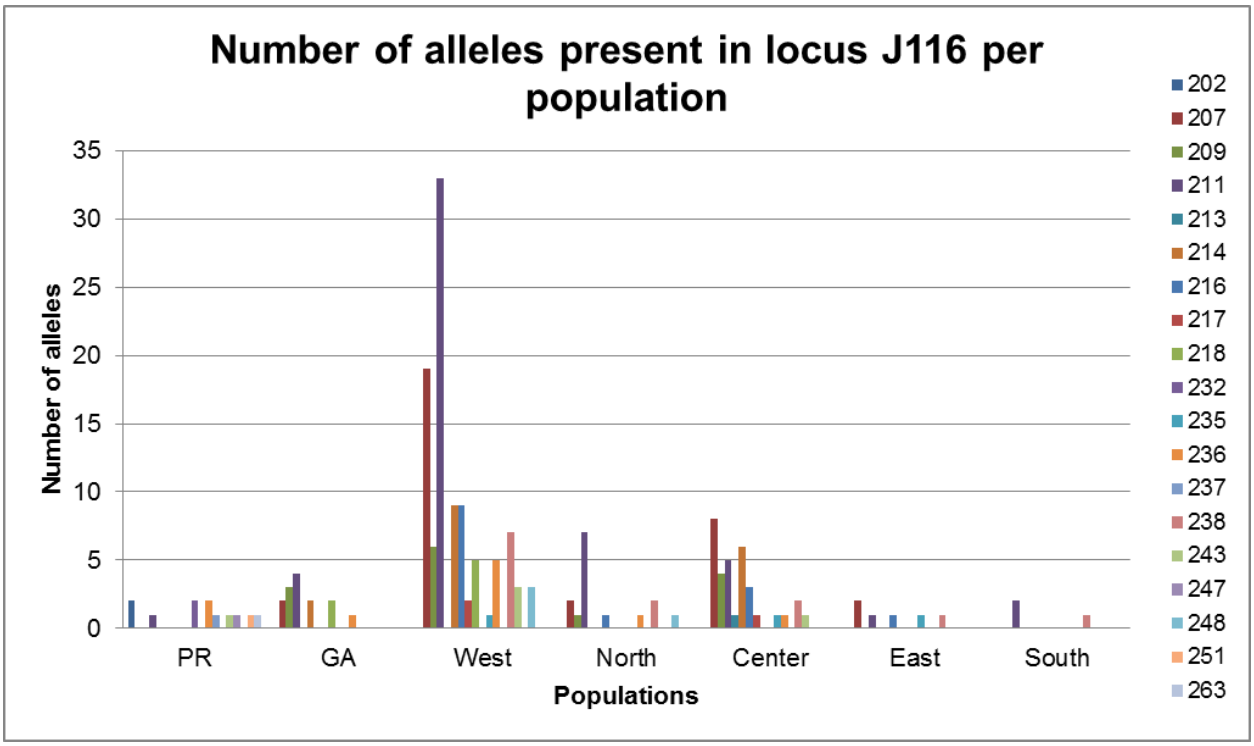


Figure 13: Percentage of alleles across populations in J116

An AMOVA was used to help differentiate between the 7 populations (Table 9). Based on the results the differences found within populations (96.9%) accounted for the great majority of the variation. The other 3.1% of the variation was caused by differences among populations. According to the p-value (0.005) the Phi-st value (differentiation) was significant.

Table 9: AMOVA* (Stepwise Mutation Model)

	SS	d.f.	var. comp.	perc.var
Within Pop	4426.79	148	29.91	96.9
Among Pop	769.04	6	0.95	3.1
Total	5195.83	154	30.86	
Phi-st:	0.031	999 permutations		
p(greater):	0.005			

The allelic weight information gave space for the construction of a genetic similarity analysis based on a binary matrix of presence (1) and absence (0) of alleles which was further used to construct a dendrogram based on the UPGMA clustering method. Two main clusters were depicted, group I contained 19 unknown accessions from across Puerto Rico and no trend could be defined as of why they grouped together (Figure 14 d). Cluster II have the gross of the unknown samples as well as the known accessions from PR and GA and can be subdivided into 4 smaller clusters (Figure14).

In sub-cluster 1 (Figure 14 a) we can see that most of the known samples grouped together. Even though the majority of the PR collection samples clustered alone together in a small group, other members of this population such as Martina and Gem grouped with GA accessions and 3 of the unknowns: 57W, 53W and 17C. Sub-cluster 2 had samples from across the island mixed together; this can be seen as well in sub-cluster 3. Although most samples in group 2 are white fleshed accessions which can be a reason for them to group together. Interestingly, in sub-cluster 3(Figure 14 b) we can identify two accessions that can be classified as clones which are 11W and 17W, both from the West but differing in flesh color.

Sub-cluster 4 (Figure 14 c) showed the same pattern of clustering observed in the previous groups, all samples blended together. We can see one of the GA accession present in this group and closely clustered with another sample which has yellow flesh as well.

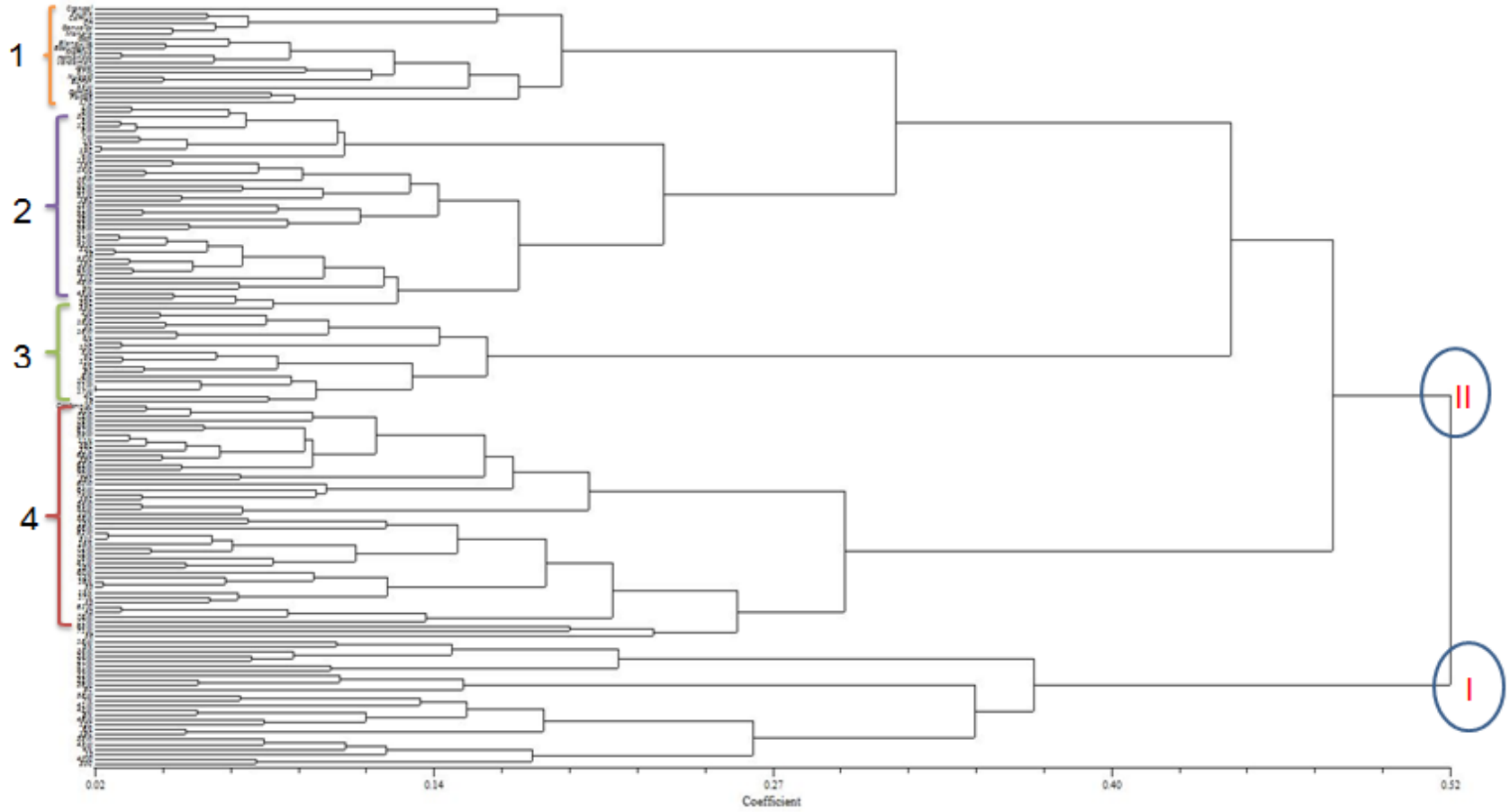


Figure 14: Dendrogram of the 155 samples

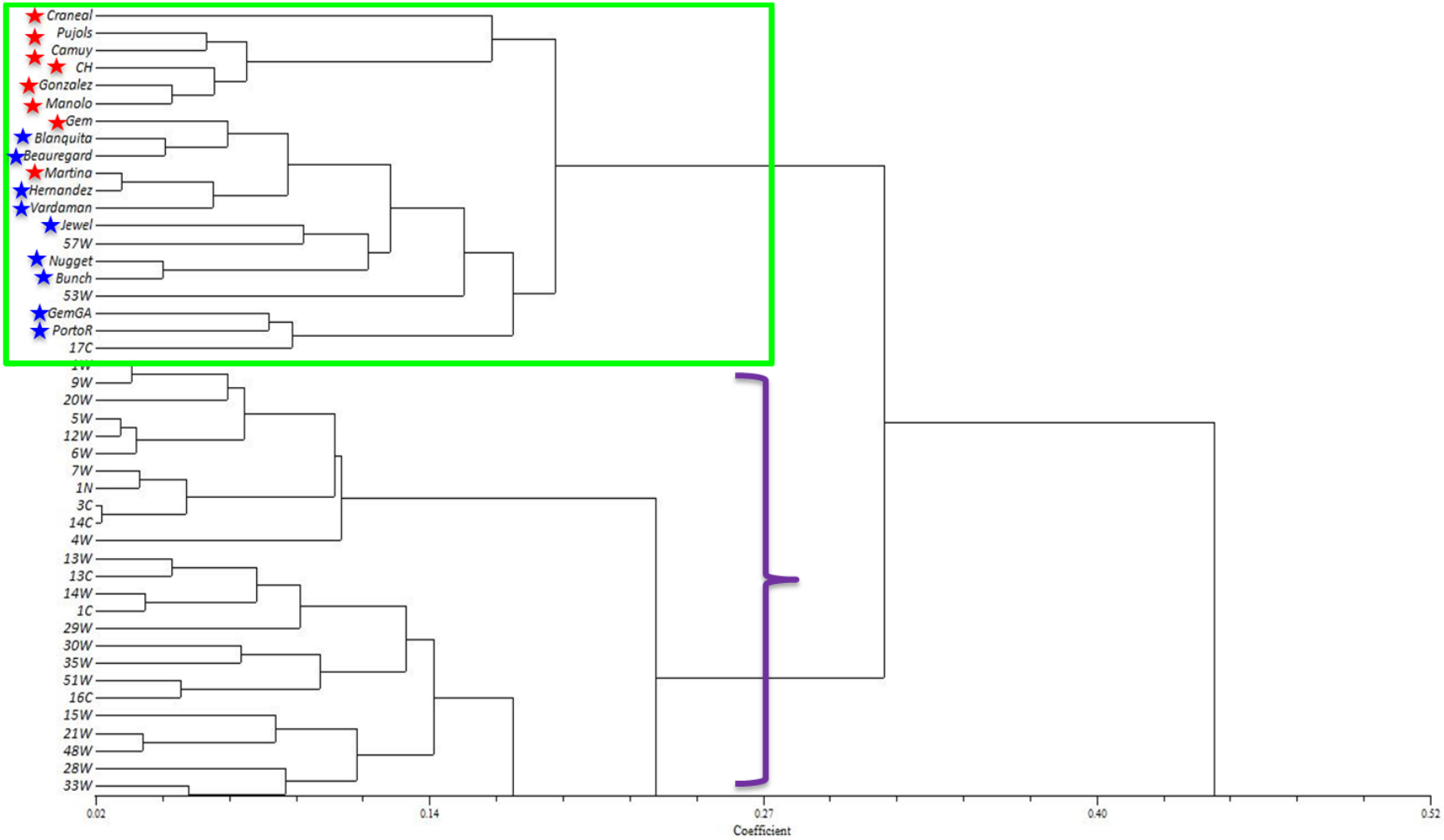


Figure 14a: Dendrogram depicting sub-clusters 1 and 2. ★ Shows PR samples, ★ shows GA samples

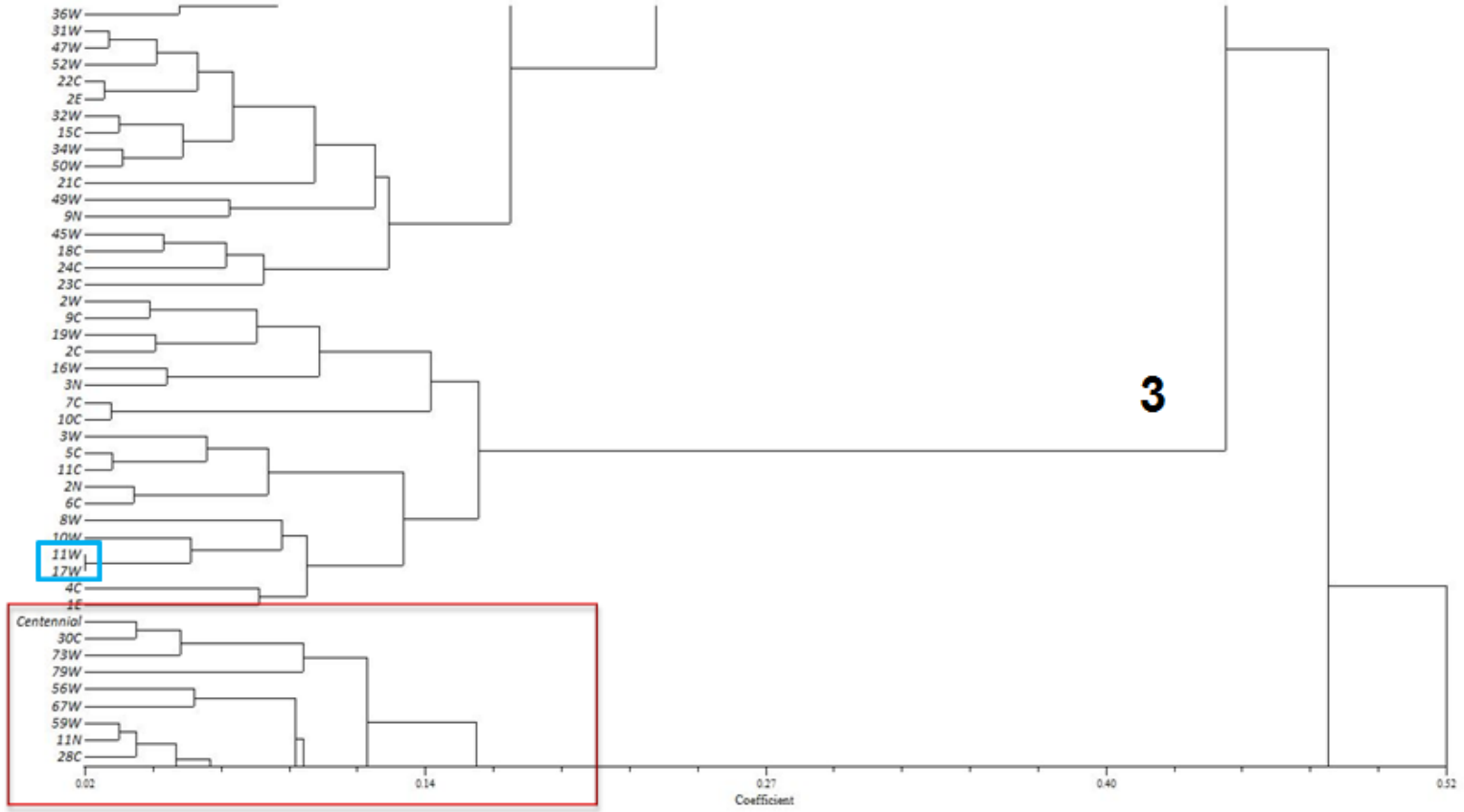


Figure 14b: Dendrogram depicting sub-cluster 3. Clone accession can be seen in the blue box. In the red box: Group containing Centennial (GA) and other unknown accession from PR.

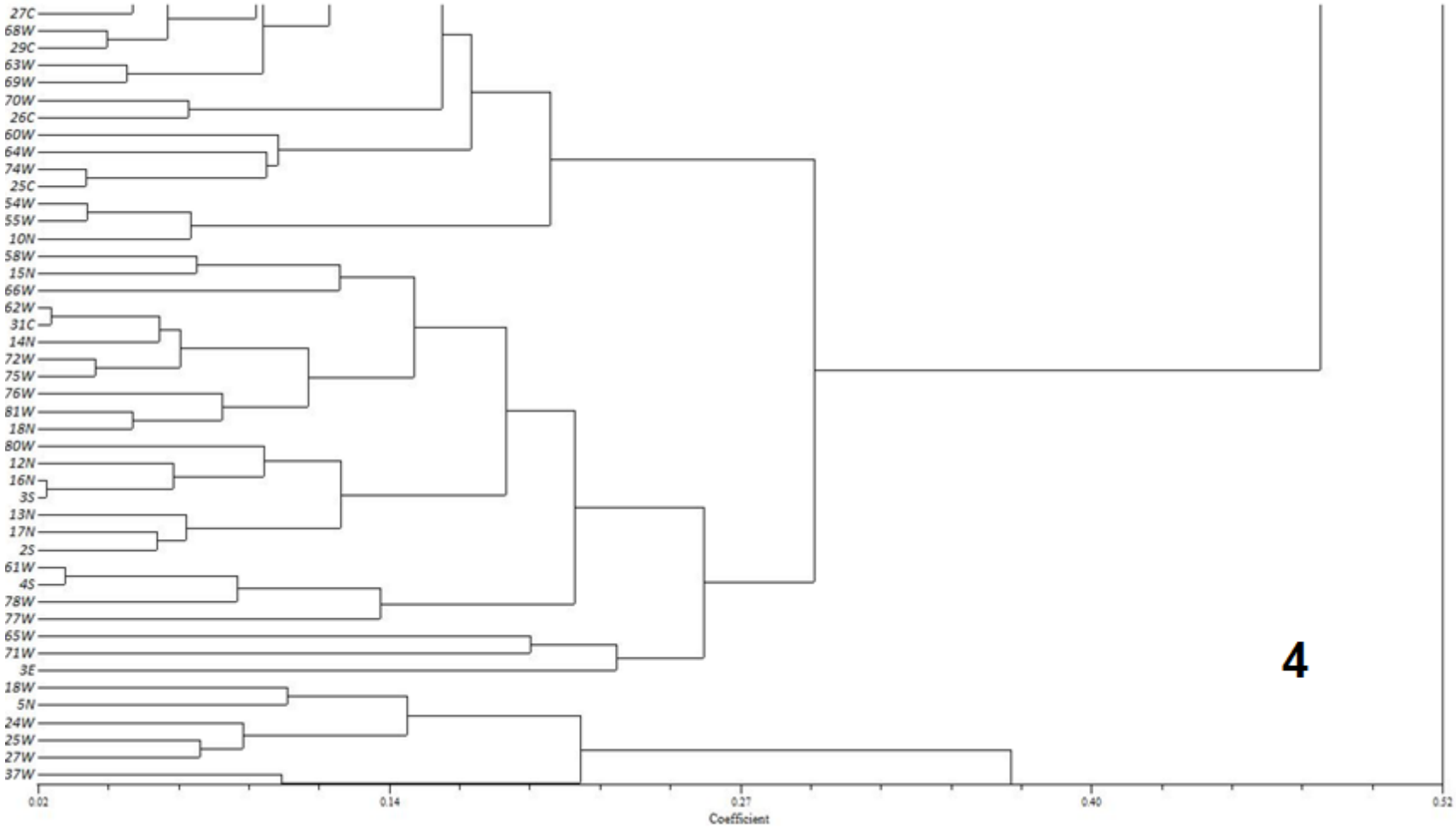


Figure 14c: Dendrogram depicting sub-cluster 4 with accessions from different areas of the island.

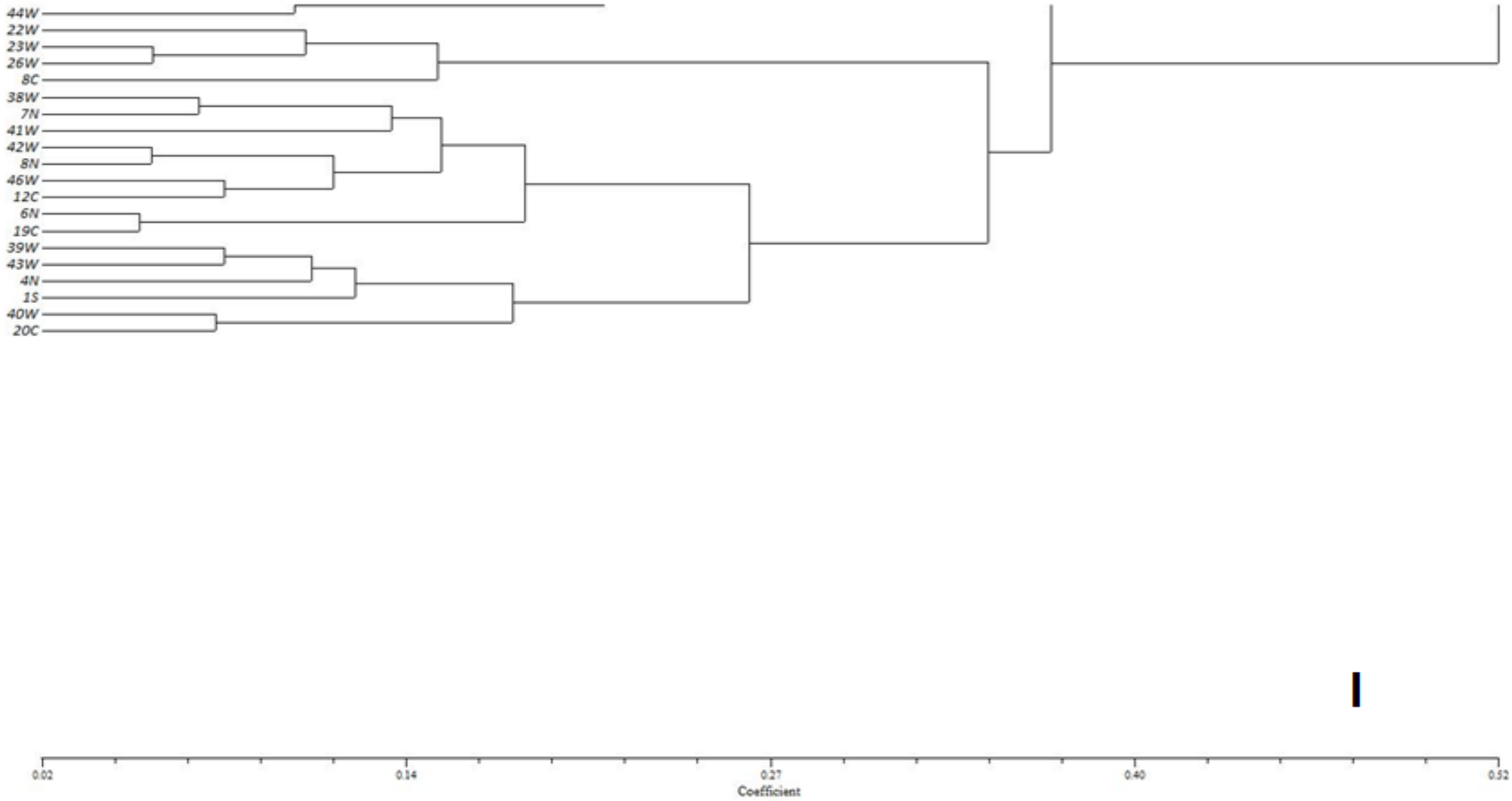


Figure 14d: Dendrogram depicting cluster I, this in one of the main groups observed.

Chapter IV

DISCUSSION

This is the first study assessing the genetic diversity of sweet potato in Puerto Rico. We found high genetic diversity with a mean value of 8.9 alleles per locus comparable to other studies (Hu *et al.*, 2004; Buteler *et al.*, 1999; Karuri *et al.*, 2009). When Tumwegamire *et al.,* (2011) analyzed 92 African accessions with 26 SSR markers they found a mean value of 6.1 alleles per locus ranging from 2 to 11. Gwandu *et al.*, (2012) analyzed 57 sweet potato genotypes in Tanzania with 4 SSR's finding high quantities of alleles as well fluctuating from 11 to 22. The high number of alleles can be explained by the hexaploidy nature of sweet potato.

The allelic composition of all 23 loci per population is fairly distributed across the populations. The West side group had the highest mean number of alleles 7.04 which can be linked to the amount of individuals in the group (81). Comparatively the populations of East and South, which had only 4 and 3 samples, had the lowest number of alleles at 3.55 and 2.87, respectively. All loci were highly polymorphic with more than 2 bands per locus. Allendorf *et al.,* (2013) states than more than two bands can suggest high genetic variation. Considering that the mean value of alleles per locus is 8.9 we can conclude with certainty that the sweet potato population in Puerto Rico is very diverse.

High levels of observed heterozygosity were found for all samples on all loci. Studies in other crops such as Montero-Rojas *et al.*, (2011) on cassava diversity in Puerto Rico found high levels of heterozygosity with values of 0.6705. There isn't previous data of this genetic diversity estimator in sweet potato to compare with. According to Harmrick and Godt (1997) a high level of genetic diversity can be defined as a value higher than 0.205 for an outcrossing monocot and 0.159 for a dicot. In sweet potato high levels of heterozygosity and genetic diversity could be explained by the outcrossing and self-incompatibility nature of the plant. This self-incompatibility in the field might result in chance seedlings from crossings (Yada *et al.*, (2010)) providing another path to increase genetic diversity.

High values of genetic diversity for each population were observed but in some cases might be overestimated. For example, it was expected seeing the West population having had the highest level of diversity due to large sample size. But the population with the highest genetic diversity was the East (H_o 0.701) while the West population was H_o 0.66. This is due to the low population size of only 4 samples in the East. These 4 samples were highly heterozygous confirmed by a negative value of the inbreeding coefficient (Allendorf *et al.*, 2013) (a negative value suggests an excess of heterozygotes). This has been previously reported in the assessment of the genetic diversity of Puerto Rican cassava where groups with small samples showed as well higher levels of H_o than the expected (Montero-Rojas *et al.,* 2011).

The GA population exhibited high values of observed heterozygosity. This group is mainly composed of sweet potato varieties that are of very economical importance in the US and are product of human selection which might have had enhanced genetic makeup. These cultivars are always propagated vegetatively in order to maintain specific desirable traits. The samples from the PR population show high levels of heterozygosity as well as the other populations from different regions in Puerto Rico. This due to the fact that sweet potato is highly heterozygous, with the capacity of multiple combinations of genes and chromosomes (Bruckner, 2004). The PR group and the unknown samples are mainly propagated vegetatively which is associated with continued mutations that contribute to diversity.

Analyzing other coefficients such as H_s demonstrated that most of the overall variability was focused within populations (Allendorf *et al.*, 2013) with a value of 0.73. This was further proved with the results from the AMOVA which suggested that 96.9% of the variation was from within populations rather than amongst (3.1%). Levels of differentiation as inferred from Phi-st value (0.031 p-value>0.005) suggest some level of genetic differentiation, a result comparable with the G_{st} value of 0.1, suggesting moderate genetic differentiation. Our results are comparable with those of Gwandu *et al.*, (2012) and Tumwegamire *et al.*, (2011) who found that the main source of variation is within populations and those of Zhang *et al.*, (2000) who calculated an among populations variation of 10% and a within population variation of 90%. All the genetic

variation and diversity in Puerto Rico could be explained too by the possibility that our sweet potato might descend directly from the first domesticated plants from Venezuela, brought to the West Indies by the Arawak Indians who populated the area as mentioned before. If indeed the plant was introduced to Puerto Rico and additional ecological factors such as climate, soil and also human practices might have added even more diversity to this already complex crop.

Clustering analysis of genetic distance yielded very interesting results. We saw two main clusters: Cluster I had 19 unknown samples from across the island and even though and trend couldn't be established this shows the great diversity present on the island.

Cluster II was subdivided into 4 smaller sub-clusters. Sub-cluster 1 contained the majority of the accessions from the PR and GA collection in addition to 3 samples from the unknown collection. This grouping might be explained by the fact that years ago sweet potato varieties from Puerto Rico were introduced to the US to be used in breeding programs for the enhancement of commercial varieties such as the ones analyzed (Dr. Carlos Ortiz, personal communication 2010). Also the PR collection samples are being maintained through vegetative propagation in the same location and there is no possibility of exchange with other plants that are not from the collection. Similarly the cultivars from GA are being maintained in vitro thus eradicating any possibilities of exchange. Even though they clustered separately from the majority of the unknown samples these accessions share more than 50% genetic information with them, which can allude to the possibility that the farmer-owned samples might be a subset of this group that became more heterozygous. This can be supported by the fact that sweet potato origin is still unknown but it's hinted that it could be the product of a tetraploid and a diploid mixture (Bruckner, 2004). It has 90 chromosomes and the fact that is vegetatively propagated which is known to add mutations to the populations greatly influence the diversity found in this study and previously reported.

Aside from the obvious grouping of the known samples from the unknown samples there was no other pattern observed. This phenomenon has been observed

previously in almost all genetic diversity analysis of sweet potato (Tumwegamire *et al.,* 2010; Gwandu *et al.*, 2012; Elameen *et al.,* 2007). Yada *et al.*, (2010) reported results similar in which samples clustered randomly and not following any distinguishable order, suggesting that a significant exchange of material is happening between landowners and farmers. Since it's so easy to propagate sweet potato either from the stem or the root, this can be a certain explanation for our results. According to Bruckner, (2004) accessions can group together but display very little if any regional similarity, suggesting an extremely high amount of genetic diversity and also a big amount of local diversity present in the sweet potato gene pool.

In sub-cluster 2 is interesting to notice that most of the accessions clustered in this group share white root flesh and most of these white accessions are from the west region of the island. In sub-cluster 3 we found two accessions that shared 100% identity (11W, 17W) although they differ in root color. Results like this are consistent with those found by Yada *et al.,* (2010) where he found duplicates morphologically different, he suggested that environmental pressures might be responsible of this event. It has been previously reported by Karuri *et al.*, (2010) that the high ploidy level in this crops may be responsible for the variability in qualitative traits caused by the increased mutation rates associated with polyploidy.

In sub-cluster 4 we found one of the accessions from GA, Centennial. This accession grouped closely with an unknown sample that has yellow fleshed color as well. Centennial is an important and preferred commercial variety in the US making possible that it was introduced to Puerto Rico for commercial purposes. Since Centennial is known to be a flavorful and a favorite cultivar, there is a possibility that it was grown all over the island for personal consumption, explaining its grouping with unknown varieties from across Puerto Rico.

It has been shown that there is a great amount of genetic diversity of sweet potato present in Puerto Rico. The genetic makeup of sweet potato itself allows for this genetic diversity to increase and be within healthy values. Farmers and landowners are a big part of the success of sweet potato in Puerto Rico. The material exchanges, as

well as human selection for better traits have influenced the genetic makeup of this important crop immensely. As mentioned before, the future of food security is uncertain. The genetic variability present in Puerto Rico for sweet potato might be a small piece in order to help solve the puzzle of food security and hunger. The high levels of heterozygosity and diversity in addition to new theories of the origin and domestication of this crop (Roullier *et al.*, 2011) might turn Puerto Rico into a fourth center for the diversity of this crop. As such, better efforts should be made to protect and conserve this valuable resource.

CHAPTER VI: CONCLUSIONS AND RECOMMENDATIONS

Conclusions

- This is the first successful assessment of genetic diversity of sweet potato in the Caribbean. After analyzing the results, a high level of genetic diversity was observed for sweet potato in Puerto Rico.

- From our results we can observe the ability of dispersal of a vegetatively propagated crop. Farmers and landowners are fundamental in the diffusion of crops. Without being conscious about it the farmers keep plant genetic diversity levels healthy. Whether it's selecting sweet potato, by color or flavor, exchanging material or introducing the plants to a new environment, variability is introduced to the crop.

- Most of the known samples clustered together indicating that they share high levels of genetic information. The unknown samples did not group in any region or root color pattern, alluding at the material exchange that could be on going in Puerto Rico. Some unknown samples clustered alone; there could be a possibility that sub-population from different ancestors might be present in Puerto Rico. Commercial varieties from other countries are available in the markets in Puerto Rico and might be favored by consumers for their consistency and flavor, enticing people to plant them for household consumption.

- High levels of diversity found in Puerto Rico and the history of domestication and the method dispersal of sweet potato turn this crop into an extremely valuable resource that needs to be protected and further studied.

Recommendations

- The present study found high levels of genetic diversity of sweet potato in Puerto Rico. This is the first such assessment conducted in the Caribbean. Taking into account our findings and the new theories of domestication of sweet potato we recommend collecting samples from the West Indies (Dominican Republic, Haiti and Jamaica) to analyze genetic diversity. If the levels of variety present in the other islands are high we could postulate the Caribbean as a new center of diversity of this crop.

- We saw in our result that there was a biased when analyzing the outcome of the samples collected in the East and in the South. More samples need to be collected from these areas as well as the North in order to try and reduce the error that the low sample numbers cause.

- As for the sweet potato present in Puerto Rico, conservation effort needs to be prepared. We are lucky to have such an important resource in our hands and should be protected.

References

Allendorf, F.W., Luikart, G., Aitken, S.N.(2013). Conservation and the Genetics of Populations 2nd Edition. Wiley-Blackwell.

Arizio, C.M., Hompanera, N., Suarez , E.Y., Manifesto, M.M. (2009). Genotypic identification and diversity evaluation of a sweet potato (Ipomoea batatas (L). Lam) collection using microsatellites. Plant Genetic Resources. 7: 135-138

Asudi, G.O., Ombwara, F.K., Rimberia, F.K., Nyende, A.B., Ateka, E.M., Wamocho, L.S. (2013). Evaluating diversity among Kenyan papaya germplasm using simple sequence repeats markers. African Journal of Food, Agriculture, Nutrition and Development, 13(1), 7307-7324

Austin, D.F., Huamán, Z. (1996). A synopsis of Ipomoea (Convolvulaceae) in the Americas. Taxon 45(1):3-38.

Belaj A., Satovic Z., Rallo L., Trujillo I. (2002) Genetic diversity and relationships in olive (*Olea europaea* L.) germplasm collections as determined by randomly amplified polymorphic DNA. Theor. Appl. Genet. 105:638-644.

Bingrui, S., Chongyun, F., Cunyi , Y., Qibin, M., Dajian, P., Hai, N. (2013). Genetic Diversity of Wild Soybeans from Some Regions of Southern China Based on SSR and SRAP Markers. American Journal of Plant Sciences, 2013, 4, 257-268.

Bruckner, Adam Wesley. AFLP-based Genetic Diversity Assessment of Global Sweetpotato (*Ipomoea batatas* (L.) Lam.) Germplasm Resources: Progress Toward The Development of a Sweetpotato Core Collection. Diss. North Carolina State University, 2004.

Burri, B. J. (2011). Evaluating Sweet Potato as an Intervention Food to Prevent Vitamin A Deficiency. Comprehensive Reviews in Food Science and Food Safety 10(2): 118-130.

Buteler, M.I., Jarret R.L., LaBonte D.R. (1999). Sequence characterization of microsatellite in diploid and polyploid Ipomoea. Theor. Appl. Genet. 99:123–132.

Buteler, M. I., LaBonte D. R., Jarret R.L., Macchiavelli R.E. I. (2002). Microsatellite-based paternity analysis in polyploid sweetpotato. J. Amer. Soc. Hort. Sci. 127(3): 392-396.

CIP. (2013). Available at: http://cipotato.org/vitaa/manual/09_TM%20Chapt%204b.pdf (verified 4 April 2013). Centro Internacional de la Papa, Peru.

Chakravarthi, B.K., Naravaneni, R. (2006). SSR marker based DNA fingerprinting and diversity study in rice (Oryza sativa. L). Afr. J. Biotechnol. 8: 684-688.

Chavarriaga-Aguirre, P., Maya, M., Bonierbale, M., Kresovich, S., Fregene, M., Tohme, J., Kochert, G. (1998). Microsatellites in cassava (Manihot esculenta Crantz): discovery, inheritance and variability. Theoretical and Applied Genetics. 97: 493-501.

Comertpay, G., Baloch, F.S., Kilian, B., Ulger, A.C., Ozkan, H. (2012). Diversity assessment of Turkish maize landraces based on fluorescent labelled SSR markers. Plant. Mol. Biol. Rep. 30:261-274.
Doyle, J.J., Doyle J.L. (1990) . Isolation of plant DNA from fresh tissue. Focus 12:13–15.

Elameen, A., Fjellheim, A., Larsen, O.A., Rognli, L., Sundheim, S., Msolla, E. Masumba, K., Mtunda, Klemsdal, S.S. (2008). Analysis of genetic diversity in a sweet potato (Ipomoea batatas L.) germplasm collection from Tanzania as revealed by AFLP. Genet. Resources Crop Evol. 55:397–408.

Engel, F. (1970). Exploration of Chilca-Canyon, Peru. Current Anthropology 11(1): 55-&.

ESEAP CIP. (2011). Available at http://www.eseap.cipotato.org/ (verified 7 Sept. 2011). Regional Office for East, Southeast Asia and the Pacific (ESEAP), Bogor, Indonesia

Food and Agriculture Organization of the United Nations. (2010). The state of the world's plant genetic resources for food and agriculture. FAO, Rome, Italy.

Food and Agriculture Organization of the United Nations. (2011). FAOSTAT database. Available at http://faostat.fao.org/ (verifi ed 10 Sept. 2011). Food and Agriculture Organization of the United Nations, Rome, Italy.

Fregene, M., Suárez, M., Mkumbira, J., Kulembeka, H., Ndedya, E., Kulaya, A., Mitchel, S., Gullberg, U., Rosling, H., Dixon, A., Dean, R., Kresovich S. (2003). Simple sequence repeat marker diversity in cassava landraces: genetic diversity and differentiation in an asexually propagated crop. Theoretical and Applied Genetics. 107: 1083-1093.

Fukino, N, Sakata, Y, Kunihisa, M, Matsumoto, S (2007) Characterization of novel simple sequence repeat (SSR) markers for melon (*Cucumis melo* L.) and their use for genotype identification. J Hort Sci Biotechnol 82(2):330–33

Gichuki, S.T., Berenyi, M., Zhang, D.P. (2003) Genetic diversity in sweetpotato [Ipomoea batatas (L.) Lam.] in relationship to geographic sources as assessed with RAPD markers. Genetic Resources and Crop Evolution, 50, 429–437.

Gwandu, C., Tairo, F., Mneney, E., Kullaya., A. (2012). Characterization of Tanzanian elite sweet potato genotypes for sweet potato virus disease (SPVD) resistance and high dry matter content using simple sequence repeat (SSR) markers. African Journal of Biotechnology Vol. 11(40), pp. 9582-9590.

Hamrick, J.L., Godt, M.J.W. Allozyme diversity in cultivated crop. Crop Sci. 1997;37:26–30.

Hokanson, S. C., Szewc-McFadden, A. K. (1998). Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a Malus x domestica borkh. core subset collection. Theoretical and Applied Genetics 97(5-6): 671-683

Hu, J., Nakatani M., Lalusin, A.G., Kuranouchi, T., Fujimura T. (2003). Genetic analysis of sweetpotato and wild relatives using inter-simple sequence repeats (ISSRs). Breed. Sci. 53:297–304.

Hu, J. J., Nakatani, M. (2004). New microsatellite markers developed from reported Ipomoea trifida sequences and their application to sweetpotato and its related wild species. Scientia Horticulturae 102(4): 375-386.

Hu J. J., Nakatani, M. (2004). Development and characterization of microsatellite markers in sweetpotato. Breeding Science 54(2): 177-188.

Hwang, S.H., Tseng, Y.T., Lo, H.F. (2002). Application of simple sequence repeats in determining the genetic relationships of cultivars used in sweetpotato polycross breeding in Taiwan. Sci. Hortic. (Amsterdam) 93:214–225.

International Potato Center, Asian Vegetable Research and Development Center and the International Board for Plant Genetic Resources. (1991). Descriptors for sweetpotato. In: Z.Huaman (ed.). Intl. Board for Plant Genetic Resources, Rome

International Potato Center. (1999). Sweetpotato Facts. CIP Peru

Jarret, RL, Bowen, N. Simple Sequence Repeats (SSRs) for sweetpotato germplasm characterization. Plant Genet Res Newslett. 1994;100:9–11.

Karuri, H.W., Ateka, E.M., Amata, R., Nyende, A.B., Muigai, A.W.T., (2009). Morphological markers cannot reliably identify and classify sweet potato genotypes based on resistance to sweet potato virus disease and dry matter content. J. Appl. Bio Sci., 15: 820–828.

Karuri, H.W., Ateka, E.M., Amata, R., Nyende, A.B., Muigai, A.W.T., Mwasame, E., Gichuki, S.T. (2010). Evaluating diversity among Kenyan sweet potato genotypes using morphological and SSR markers. Int. J. Agric. Biol., 12: 33–38.

Keegan, W.F. (1995). Modeling Dispersal in the Prehistoric West Indies. World Archaeology, 26(3): Colonization of Islands 400-420.

Laurie, S.M., Calitz, F.J., Adebola, P.O., Lezar, A. (2010). Characterization and evaluation of South African sweet potato (Ipomoea batatas (L.) LAM) land races. South African Journal of Botany 85 (2013) 10–16

Meirmans, P.G., Van Tienderen, P.H. (2004).GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms, Molecular Ecology Notes 4 p.792-794.

Mikel, M.A., Kolb, F.L. (2008). Genetic diversity of contemporary North American barley. *Crop Sci.* 48: 1399-1407.

Montero-Rojas M,, Correa, A,M,, Siritunga, D. 2011. Molecular differentiation and diversity of cassava (Manihot esculenta) taken from 162 locations across Puerto Rico and assessed with microsatellite markers. AoB PLANTS 2011 plr010

Montero-Rojas, M., Ortiz, M., Beaver, J.S., Siritunga, D. (2013). Genetic, morphological and cyanogen content evaluation of a new collection of Caribbean Lima bean (*Phaseolus lunatus* L.) Landraces. Genet Resour Crop Evol. In press

Moscoso, C. (1955). El cultivo de la batata en Puerto Rico. Repositorio Digital Agricola. Boletin No. 126.

Mukhopadhyay, S. K., Chattopadhyay, A., Chakraborty, I., Bhattacharya, I. (2011). Crops that feed the world 5. Sweetpotato. Sweetpotatoes for income and food security. Food Security 3(3): 283-305.

O'Brien, P.J. (1972). The sweetpotato: Its origin and dispersal. Am. Anthropol. 74:343–365.

Oetting W., Lee, H., Flanders, D., Wiesner, G., Sellers, T., King, R. (1995). Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13-tailed primers. Genomics. 30:450-458.

Prakash, C.S., He, G., Jarret, R. L. (1996). DNA marker-based study of genetic relatedness in United States sweetpotato cultivars. J. Am. Soc. Hortic. Sci. 121:1059–1096.

Rohlf, F. J. 2008. NTSYSpc: Numerical Taxonomy System, ver. 2. 02i. Exeter Publishing, Ltd.: Setauket, NY.

Röder M., Korzun, V., Wendehak, K., Plaschke, J., Tixier, M., Leroy, P., Ganal, M. (1998). A microsatellite map of wheat. Genetics.149:2007-2023.

Rossel, G., Espinoza, C., Javier, M. and Tay, D. 2008. Regeneration guidelines: sweet potato. In: Dulloo M.E., Thormann I., Jorge M.A. and Hanson J., editors. Crop specific regeneration guidelines [CD-ROM]. CGIAR System-wide Genetic Resource Programme, Rome, Italy. 9 pp.

Roullier, C., Rossel, G., Tay, D., MCKey, D., Levot, V. (2011). Combining chloroplast and nuclear microsatellites to investigate origin and dispersal of New World sweet potato landraces. Molecular Ecology 20, 3963–3977

Sanjur, O.I., Piperno, D.R., Andres, T.C., Wessel-Beaver, L. (2002) Phylogenetic relationships among domesticated and wild species of Cucurbita (Cucurbitaceae)

inferred from a mitochondrial gene: Implications for crop plant evolution and areas of origin. Proceedings of the National Academy of Sciences, USA, 99, 535–540.

Thompson, P.G,. Hong, L.L., Ukoskit, K., Zhu, S. (1997). Genetic linkage of randomly amplified polymorphic DNA (RAPD) markers in sweet potato. J. Am. Soc. Horticult. Sci. 122: 79-82.

Tumwegamire, S., Rubaihayo, P. R., LaBoonte, D.R., Diaz, F., Kapinga, R., Mwanga, R.O.M., Grüneberg, W.J. (2011). Genetic Diversity in White- and Orange-Fleshed Sweetpotato Farmer Varieties from East Africa Evaluated by Simple Sequence Repeat Markers." Crop Science 51(3): 1132-1142.

UNICEF. (2011). Available at http://www.unicef.org/mdg/poverty.html (verified 2 Sept. 2011). United Nations Children's Fund, New York, USA

USDA The Census of Agriculture. (2007). Available at http://www.agcensus.usda.gov (verified 29 Aug. 2011) United States Department of Agriculture

Wang *et al*.: Characterization and development of EST-derived SSR markers in cultivated sweetpotato (Ipomoea batatas). BMC Plant Biology, 2011, 11-139.

Veasey E.A., Borges, A., Silva-Rosa, M., Queiroz-Silva, J.R., Bressan, E. (2009). Genetic characterization of cassava (Manihot esculenta) landraces in Brazil assessed with simple sequence repeats .Genetics and Molecular Biology, 32, 1, 104-110.

Yada, B., Tukamuhabwa, P., Wajala, B., Kim, D.J., Skilton, R.A., Alajo, A., Mwanga, R. (2010). Characterizing Ugandan sweetpotato germplasm using fl uorescent labeled simple sequence repeat markers. HortScience 45(2):225–230.

Zhang, D. P., Ghislain, M. (1998). RAPD variation in sweetpotato (Ipomoea batatas (L.) Lam) cultivars from South America and Papua New Guinea. Genetic Resources and Crop Evolution 45(3): 271-277.

Zhang, D.P., Carbajulca, D., Ojeda, L., Rossel, G., Milla, S. , Herrera, C., and Ghislain, M. (2000). Microsatellite analysis of genetic diversity in sweetpotato varieties from Latin America. p. 295–301. CIP Program Report 1999–2000. International Potato Center, Lima, Peru.

Zhang, D., Rossel, G., Kriegner , A., Hijmans, R. (2004). AFLP assessment of diversity in sweetpotato from Latin America and the Pacific region: Its implications on the dispersal of the crop. Genet. Resour. Crop Evol. 51:115–120.

Appendix A. Polyacrylamide gel preparation and LICOR 4300 electrophoresis protocols

Reagents and preparation

Stock Bind Silane solution: Used to bind the gel to the Boroflex glass. It also helps to keep the gel in a good shape for continuous use.

- Add 50 µL of bind Silane to 10 mL of 100 % ethanol.
- Mix well and store at 4 °C in an amber colored bottle.

Work Bind Silane solution: To activate the bind silane solution mix 25 µL of stock bind silane solution and 25 µL of 10% acetic acid in a 1.5 mL microcentrifuge tube.

Ammonium persulfate (APS) solution: APS is a chemical which provides the source of free radicals needed for polymerization of the acrylamide. The APS solution should be prepared to have a 10% concentration.

- Add0.1g of APS powder to 1 ml ddH_2O in a 1.5 mL microcentrifuge test tube.

TBE 10X (1 L)
This solution is used as a running buffer for electrophoresis. Add the following to 1 L of ddH_2 O.
*1X dilution is the required one for running, it should be prepared fresh.

- Tris Base 107.8g
- EDTA 7.44g
- Boric Acid 55g

Polyacrylamide 6.5 % KB Plus Gel Matrix solution

- Meassure 20 mL of KB Plus Gel Matriz 6.5% to room temperature and let it rest for 20 minutes and prepare glass plates for gel preparation while KB Plus releases the oxygen. Note: If the oxygen is not released, formation of bubbles during the gel loading might occur.

- Add 150 µL of 10% APS and 15 µL TEMED when the gel sandwich is ready. With a small beaker load the solution between the glasses.

Cleaning plates

Prepare a 2% laboratory detergent solution (Micro 90 (International Products Corp. Burlington NJ).

1. If the gel is still in the glass, take a paper towel and cover the gel with it. Let the paper stick to the The gel should come off with the paper towel.

2. Pour a small amount of 2% Micro 90 solution onto the side of the plate that was in contact with the gel.

3. Thoroughly scrub the entire plate with a bristle brush. Remove any dried-on polyacrylamide and rinse. Repeat until smooth surface is seen and stand plates in a rack to air dry.

Preparation of the gel sandwich

1. Wipe both plates (gel side) with 70% isopropanol using Kimwipes as needed to remove dust particles on the plates. Allow the plates to dry for 3 min.

2. Apply the work Silane solution in both plates (gel side) in the area where wells will form. Allow the solution to dry on the plates.

3. Assemble the gel sandwich making sure that the rails fit tightly against the edges of both glass plates.

4. Prepare the polyacrylamide solution as described above and pour the gel solution into the sandwich inserting the shark tooth comb upside down.

5. Allow the gel to polymerize for 1. 5 hours (LI-COR manual, 2004)

Licor 4300 DNA analyzer running conditions.

Table 9. Show the standardized conditions necessary to run cassava SSR loci. **Parameter**	**Standardized Microsatellite Run**
Plate Length	25 cm
Spacer Thickness	0.25 mm
Gel Composition	6.5 % LI-COR KB Plus
Pre-Run time	25 min
Run Time	1.45 hours
Reload Gel	1. 45 hours
Voltage	1500
Power	40
Current	40 mA
Temperature	45 °C

Loading samples

1. Dilute 0.5 µl of the PCR product in 5 µl of Blue Stop solution from LI-COR. Denature it at 95 °C for 3 min. Denaturalization has to be done a few minutes before loading the samples in the gel.

2. Load on the well only 0.5 µl of the dilution and 1.0 µl of IRDye ladder.

APPENDIX B: MATRIX PREPARATION FOR GenoDive

Preparation of the Weight Matrix for GenoDive:

Once the weight matrix is in excel save it as .tex Unicode. Open with a text editor program such as Notebook, NotePad or TextPad.

Do as follows in GenoDive Manual (Meirmans, 2012):

- In the first line: Comments, a single line to give a short description of the data (optional, if no comments leave blank and enter)
- Second line, five numbers separated by tabs:
    - The total number of individuals
    - The number of populations
    - The number of loci
    - The maximum ploidy levels used
    - The number of digits used to code a single allele

- Next p lines: population names, on a separate line for every population. Optionally, the population name can be followed by the names of the groups to which this population belongs.
- Line p+3, column headers, separated by tabs:
    - The total number of individuals
    - Generic name for populations
    - Generic name for individuals
    - The name of every locus.
- Next lines, Individual data (separated by tabs):
    - The total number of individuals
    - The population number (not the population name) to which the individual belongs
    - The number of the clone to which the individual belongs (optional)
    - The name of the individual. This name should not begin with a number, not contain spaces and should be unique for every individual
    - The genotype of the individual at every locus

APPENDIX C

Table : Allele counts per locus

IB242										
Pop	Total	136	140	144	146	148	150	152	155	
PR	19	0	7	0	0	8	0	4	0	
GA	26	9	0	8	2	4	3	0	0	
West	122	41	0	5	46	8	21	0	1	
North	19	6	0	0	6	4	3	0	0	
Center	59	20	0	2	21	2	10	0	4	
East	5	2	0	1	1	0	1	0	0	
South	0	0	0	0	0	0	0	0	0	
Overall	250	78	7	16	76	26	38	4	5	
		0.31	0.02	0.06	0.30	0.10	0.15	0.01	0.02	1
CIP:										
Popn	Total	155	160	164	167					
PR	23	6	4	5	8					
GA	11	2	1	3	5					
West	129	45	22	35	27					
North	10	4	0	4	2					
Center	49	19	5	14	11					
East	5	2	1	1	1					
South	0	0	0	0	0					
Overall	227	78	33	62	54					
		0.34	0.14	0.27	0.23	1				
IB297										

Pop	Total	150	154	156	160	164	166	168	170	174	176	177	180	182	
PR	14	1	2	1	5	0	0	1	1	0	1	2	0	0	
GA	22	10	0	3	3	1	0	0	0	0	0	0	4	1	
West	114	42	0	13	25	0	2	0	0	18	0	10	4	0	
North	23	9	0	1	3	0	0	0	0	7	0	3	0	0	
Center	43	16	0	4	12	1	4	0	0	6	0	0	0	0	
East	7	2	0	0	2	0	0	0	0	2	0	1	0	0	
South	3	2	0	1	0	0	0	0	0	0	0	0	0	0	
Overall	226	82	2	23	50	2	6	1	1	33	1	16	8	1	
		0.36	0.00	0.10	0.22	0.00	0.02	0.00	0.00	0.14	0.00	0.07	0.03	0.00	1
IB324:															
Popul at	Total	136	140	143	145	147	148	149	150						
PR	27	0	5	5	7	8	0	2	0						
GA	16	0	0	2	5	8	1	0	0						
West	123	4	11	8	43	39	0	18	0						
North	17	2	2	0	6	4	0	3	0						
Center	52	4	5	4	13	18	0	6	2						
East	2	0	0	0	0	2	0	0	0						
South	6	0	1	0	2	2	0	1	0						
		0.04	0.09	0.07	0.31	0.33	0.00	0.12	0.00	1					
R16:															
Popul at	Total	210	220	223	225	230	235								
PR	10	0	2	8	0	0	0								
GA	16	4	3	8	1	0	0								
West	108	2	41	58	0	7	0								
North	20	0	7	13	0	0	0								
Center	58	2	20	29	1	6	0								

East	5	0	2	3	0	0	0			
South	2	0	1	0	0	0	1			
Overall	219	8	76	119	2	13	1			
		0.03	0.34	0.54	0.00	0.05	0.00	1		
R21:										
Popular	Total	181	190	204	206					
PR	18	7	0	7	4					
GA	19	7	0	8	4					
West	143	53	4	69	17					
North	17	6	0	9	2					
Center	58	24	2	29	3					
East	7	3	0	3	1					
South	1	0	0	1	0					
Overall	263	100	6	126	31					
		0.38	0.02	0.47	0.11	1				
IB544b:										
Popula	Total	193	194	197	198	200	204	207	209	
PR	23	3	0	6	1	2	3	8	0	
GA	15	0	0	6	0	0	1	8	0	
West	138	19	0	59	0	4	8	47	1	
North	27	2	1	11	0	2	1	10	0	
Center	63	7	0	27	0	4	1	24	0	
East	6	1	0	3	0	0	0	2	0	
South	8	0	0	4	0	1	1	2	0	
Overall	280	32	1	116	1	13	15	101	1	
		0.11	0.00	0.41	0.00	0.04	0.05	0.36	0.00	1
C12:										

Popula	Total	110	112	114	116	117	118	120	122	124	128				
PR	28	0	3	0	6	0	0	6	7	5	1				
GA	24	0	2	5	2	0	4	5	4	2	0				
West	153	10	38	0	14	1	16	34	17	20	3				
North	39	3	12	0	6	0	5	9	1	3	0				
Center	51	2	15	0	4	0	3	10	12	4	1				
East	2	0	0	0	0	0	0	0	0	2	0				
South	7	0	2	0	2	0	1	2	0	0	0				
Overall	304	15	72	5	34	1	29	66	41	36	5				
		0.04	0.23	0.01	0.11	0.00	0.09	0.21	0.13	0.11	0.01	1			
J10:															
Popula	Total	190	195	198	200	204	206	208	210	212	214	220	240	262	
PR	12	0	4	0	2	0	1	0	2	2	1	0	0	0	
GA	9	0	1	0	1	0	1	0	4	2	0	0	0	0	
West	220	1	41	1	42	1	49	37	21	26	0	0	1	0	
North	48	0	11	0	9	0	10	11	2	4	0	0	0	1	
Center	90	0	19	0	21	0	20	14	5	10	0	1	0	0	
East	9	0	2	0	2	0	2	1	0	2	0	0	0	0	
South	8	0	3	0	3	0	0	1	1	0	0	0	0	0	
Overall	396	1	81	1	80	1	83	64	35	46	1	1	1	1	
		0.00	0.20	0.00	0.20	0.00	0.20	0.16	0.08	0.11	0.00	0.00	0.00	0.00	1
s18:															
Popula	Total	230	240	242	245	248	250	252	260						
PR	9	0	0	0	8	1	0	0	0						
GA	13	0	0	0	8	3	2	0	0						
West	150	9	6	6	50	30	37	11	1						
North	27	1	1	1	9	7	7	1	0						
Center	62	1	7	1	23	10	16	4	0						

East	5	0	0	0	2	0	3	0	0											
South	3	0	0	0	0	3	0	0	0											
Overall	269	11	14	8	100	54	65	16	1											
		0.04	0.05	0.02	0.37	0.20	0.24	0.05	0.00	1										
R19:																				
Populat	Total	194	208																	
PR	8	1	7																	
GA	0	0	0																	
West	0	0	0																	
North	0	0	0																	
Center	0	0	0																	
East	0	0	0																	
South	0	0	0																	
Overall	8	1	7																	
		0.12	0.87	1																
J116:																				
Popula	Total	202	207	209	211	213	214	216	217	218	232	235	236	237	238	243	247	248	251	263
PR	12	2	0	0	1	0	0	0	0	0	2	0	2	1	0	1	1	0	1	1
GA	14	0	2	3	4	0	2	0	0	2	0	0	1	0	0	0	0	0	0	0
West	102	0	19	6	33	0	9	9	2	5	0	1	5	0	7	3	0	3	0	0
North	15	0	2	1	7	0	0	1	0	0	0	0	1	0	2	0	0	1	0	0
Center	33	0	8	4	5	1	6	3	1	0	0	1	1	0	2	1	0	0	0	0
East	6	0	2	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0
South	3	0	0	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Overall	185	2	33	14	53	1	17	14	3	7	2	3	10	1	13	5	1	4	1	1
		0.01	0.17	0.07	0.28	0.00	0.09	0.07	0.01	0.03	0.01	0.01	0.05	0.00	0.07	0.02	0.00	0.02	0.00	0.00
IB1809:																				

Populat	Total	145	146	147	148	149	150	152	154	155								
PR	14	2	4	0	0	4	0	4	0	0								
GA	20	7	3	0	3	2	2	3	0	0								
West	83	25	1	1	0	30	0	22	2	2								
North	16	6	0	0	0	5	0	5	0	0								
Center	28	10	0	0	0	6	0	12	0	0								
East	4	2	0	0	0	2	0	0	0	0								
South	5	2	0	0	0	1	0	2	0	0								
Overall	170	54	8	1	3	50	2	48	2	2								
		0.31	0.04	0.00	0.01	0.29	0.01	0.2	0.01	0.01	1							
S11:																		
Populat	Total	236	238	240	242	243	244	245	246	247	249	250	252	253	255	256	262	
PR	28	0	0	1	0	2	2	0	1	6	1	5	1	4	0	3	2	
GA	30	0	2	7	4	0	7	0	0	3	0	5	0	2	0	0	0	
West	140	4	4	20	0	6	26	5	5	21	0	23	0	16	7	3	0	
North	23	0	0	5	0	1	5	1	1	3	0	6	0	1	0	0	0	
Center	51	1	4	7	0	1	10	4	3	4	0	11	0	4	2	0	0	
East	5	0	0	0	0	0	1	1	1	1	0	1	0	0	0	0	0	
South	7	2	0	2	0	0	2	0	0	0	0	1	0	0	0	0	0	
Overall	284	7	10	42	4	10	53	11	11	38	1	52	1	27	9	6	2	
		0.02	0.03	0.14	0.01	0.0	0.1	0.03	0.03	0.13	0.00	0.18	0.00	0.09	0.03	0.02	0.00	1
J175:																		
Popula	Total	133	136	138	140	142	144	145	146	147	149							
PR	22	0	5	0	5	4	0	5	0	3	0							
GA	32	8	3	4	6	4	1	4	0	2	0							
West	172	24	45	24	27	16	0	28	5	3	0							
North	29	2	7	6	8	3	0	3	0	0	0							
Center	82	6	17	13	16	16	0	9	4	0	1							

East	7	1	1	1	1	1	0	1	0	1	0		
South	2	0	2	0	0	0	0	0	0	0	0		
Overall	346	41	80	48	63	44	1	50	9	9	1		
		0.11	0.23	0.13	0.18	0.12	0.00	0.14	0.02	0.02	0.00	1	
S10:													
Popul	Total	300	302	305	306	310	330	337					
PR	3	0	0	3	0	0	0	0					
GA	3	2	0	0	0	1	0	0					
West	45	9	4	0	1	1	9	21					
North	9	4	2	0	0	0	1	2					
Center	17	5	0	0	0	0	3	9					
East	4	2	0	0	0	0	1	1					
South	0	0	0	0	0	0	0	0					
Overall	81	22	6	3	1	2	14	33					
		0.27	0.07	0.03	0.01	0.02	0.17	0.40	1				
S17:													
Popul at	Total	183	186	188	189	192	194	197	198	200	202	204	
PR	28	2	1	3	2	6	1	1	4	5	2	1	
GA	30	2	4	6	1	0	2	8	0	5	0	2	
West	175	4	7	18	23	34	21	36	0	23	6	3	
North	47	2	3	4	7	9	8	5	0	6	2	1	
Center	71	2	6	8	9	10	7	20	0	4	4	1	
East	12	0	2	1	2	1	1	2	0	1	1	1	
South	11	0	2	2	3	1	2	1	0	0	0	0	
Overall	374	12	25	42	47	61	42	73	4	44	15	9	
		0.03	0.06	0.11	0.12	0.16	0.11	0.1	0.01	0.11	0.04	0.02	1
J67:													
Popul	Tota	192	194	195	197	204	206	215	216				

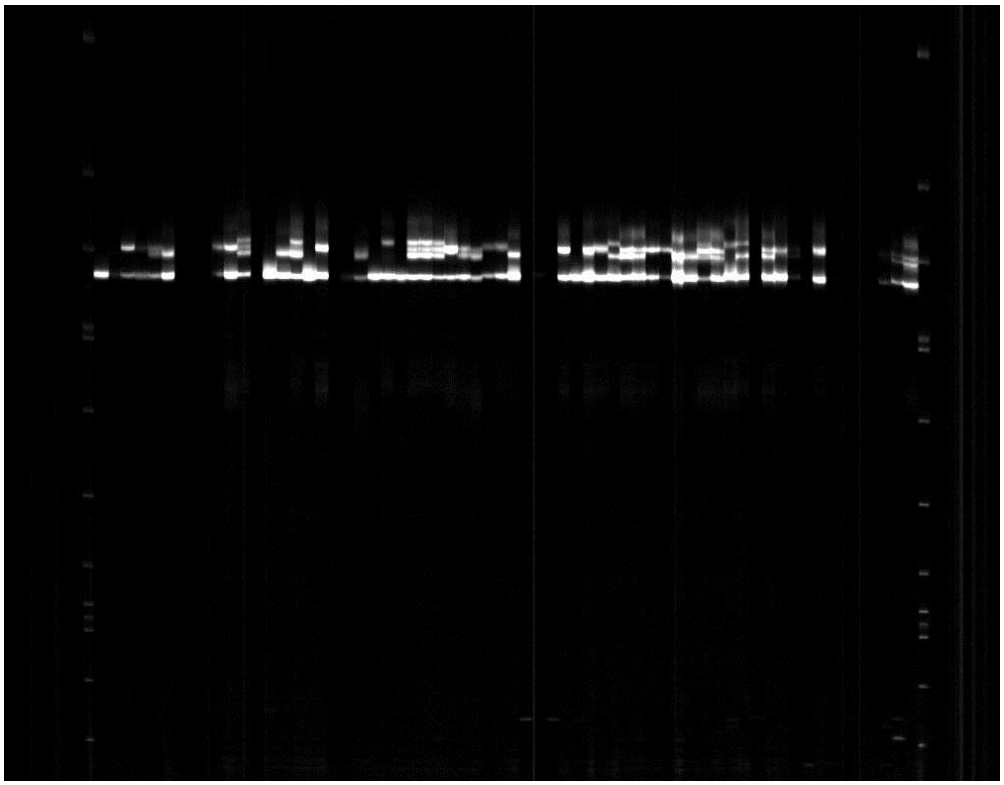
at	l																		
PR	22	0	9	1	4	0	4	3	1										
GA	27	2	10	0	8	1	0	0	6										
West	86	14	26	4	14	20	0	0	8										
North	17	6	5	0	2	3	0	0	1										
Center	30	4	7	2	5	7	0	0	5										
East	4	0	1	0	1	1	0	0	1										
South	6	2	2	0	0	1	0	0	1										
Overall	192	28	60	7	34	33	4	3	23										
		0.14	0.31	0.03	0.17	0.17	0.02	0.01	0.11	1									
R13:																			
Popula	Total	224	225	226	228	230	231	232	234	237	238	240	245	246	247	250	258	264	
PR	15	0	2	2	1	0	0	2	0	1	3	0	0	2	1	0	0	1	
GA	16	3	0	0	0	1	0	0	0	7	0	1	1	0	0	0	3	0	
West	68	0	17	0	4	8	11	0	0	16	0	0	7	0	0	2	3	0	
North	9	0	3	0	0	1	1	0	0	4	0	0	0	0	0	0	0	0	
Center	27	0	11	0	0	4	3	0	1	5	0	0	1	0	0	1	1	0	
East	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
South	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Overall	135	3	33	2	5	14	15	2	1	33	3	1	9	2	1	3	7	1	
		0.02	0.24	0.01	0.03	0.10	0.11	0.01	0.0	0.24	0.02	0.00	0.06	0.01	0.00	0.02	0.05	0.00	1
S01:																			
Popul at	Total	233	234	236	238	240													
PR	13	6	0	5	2	0													
GA	9	0	7	2	0	0													
West	52	0	35	13	3	1													
North	6	0	5	1	0	0													

Center	18	0	13	4	1	0		
East	1	0	1	0	0	0		
South	2	0	0	1	0	1		
Overall	101	6	61	26	6	2		
		0.05	0.60	0.25	0.05	0.01	1	
S07:								
Populat	Total	175	189	191	208			
PR	7	7	0	0	0			
GA	18	0	8	6	4			
West	128	0	69	25	34			
North	24	0	15	5	4			
Center	52	0	27	13	12			
East	5	0	3	1	1			
South	5	0	3	0	2			
Overall	239	7	125	50	57			
		0.02	0.52	0.20	0.23	1		
R08:								
Popul	Total	225	228	230	232	233	235	
PR	17	6	0	6	1	4	0	
GA	23	10	4	6	3	0	0	
West	138	63	26	36	11	0	2	
North	19	10	3	6	0	0	0	
Center	59	27	9	17	4	0	2	
East	6	3	2	1	0	0	0	
South	1	1	0	0	0	0	0	
Overall	263	120	44	72	19	4	4	
		0.45	0.16	0.27	0.07	0.01	0.01	1

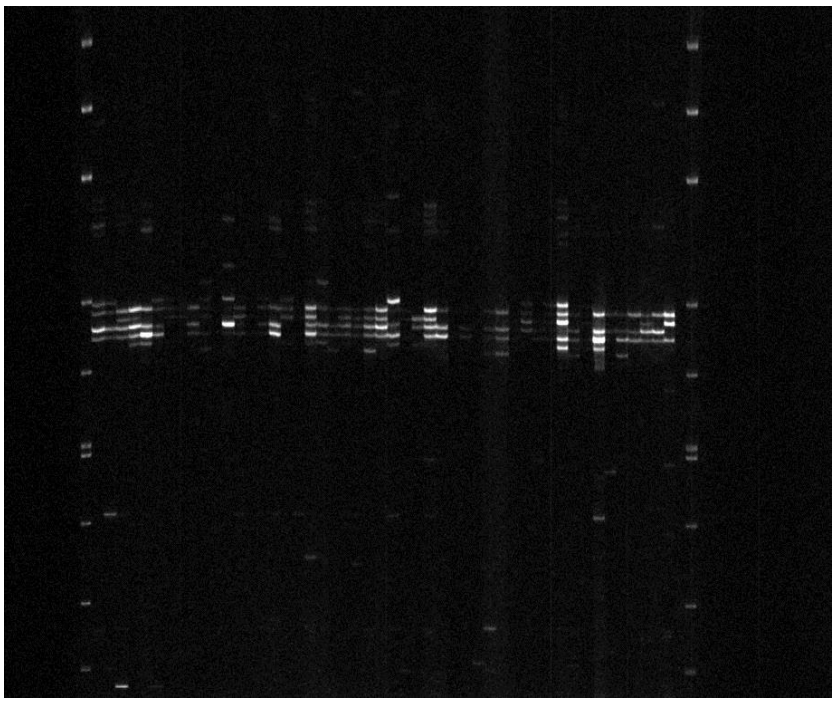
R03:							
Popula	Total	255	256	258	267	270	
PR	3	0	0	1	1	1	
GA	9	0	2	7	0	0	
West	48	1	32	15	0	0	
North	6	0	5	1	0	0	
Center	15	0	13	2	0	0	
East	3	0	3	0	0	0	
South	2	0	2	0	0	0	
Overall	86	1	57	26	1	1	
		0.01	0.66	0.30	0.01	0.01	1

APPENDIX D

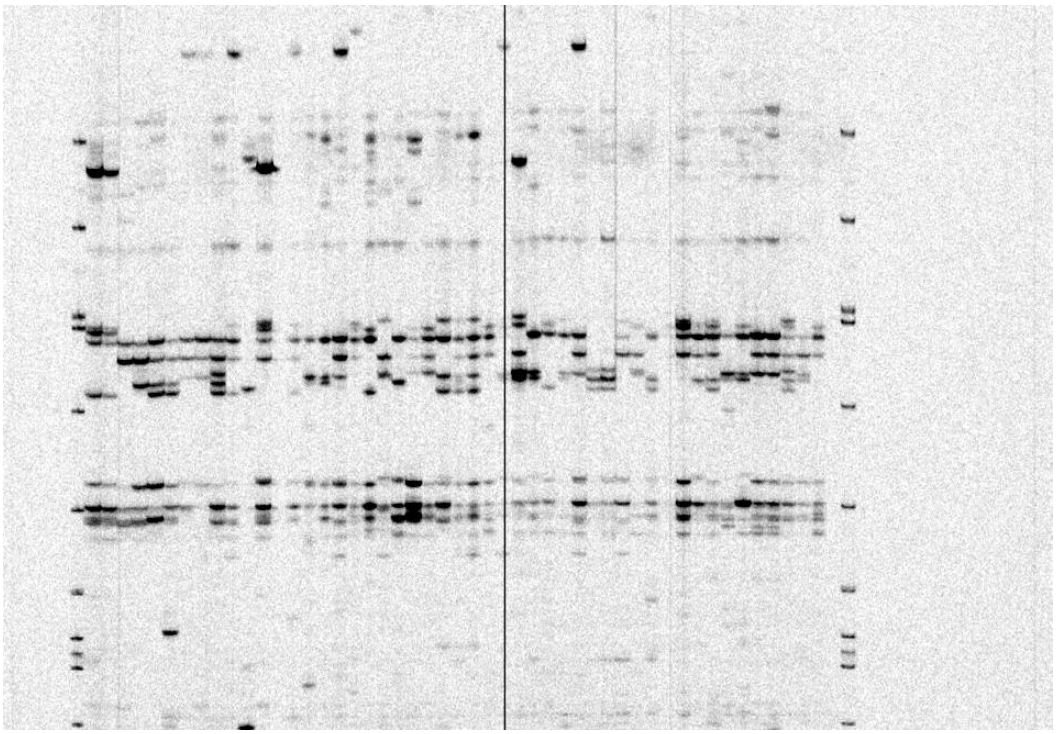
Gel Pictures



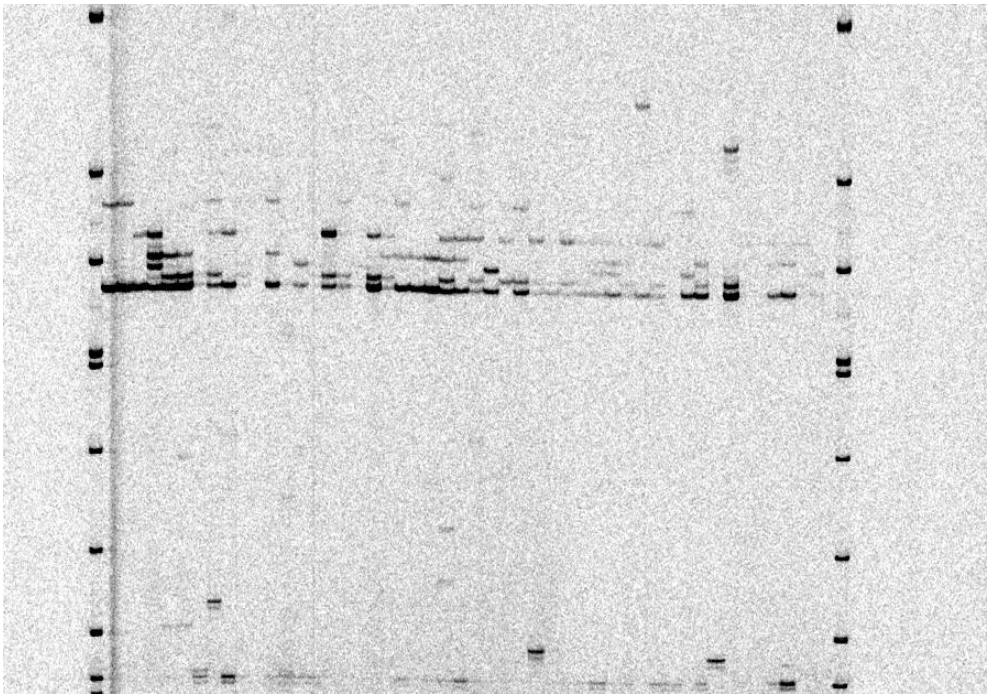
SSR R08



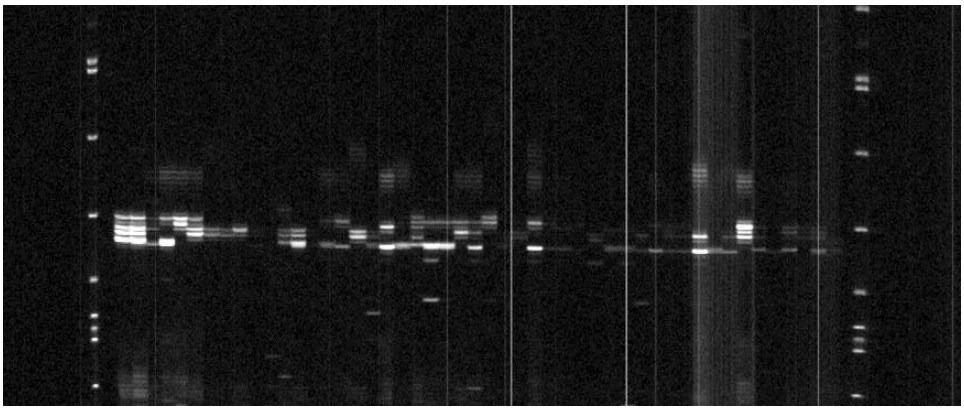
SSR S11



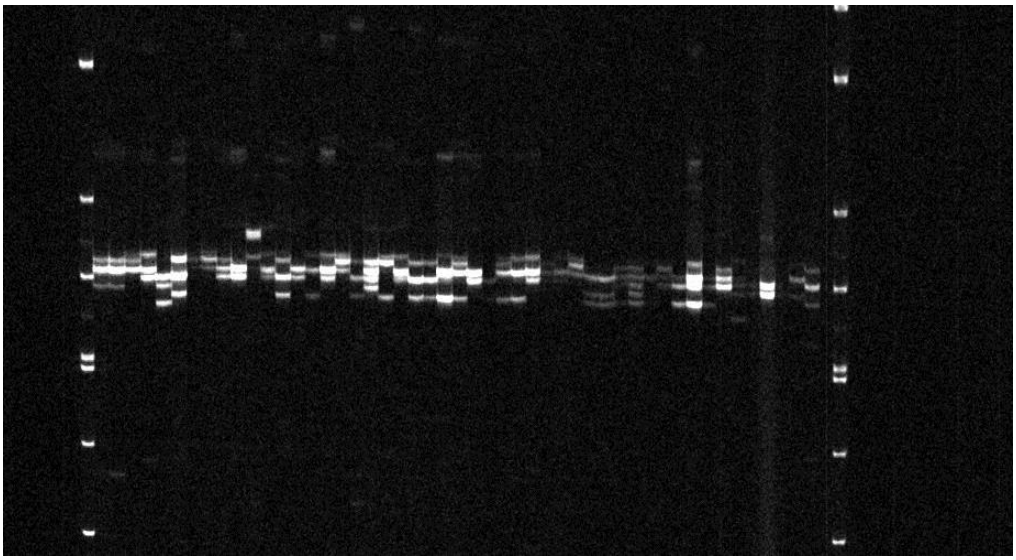
SSR S17



SSR R13



SSR J175



SSR S01

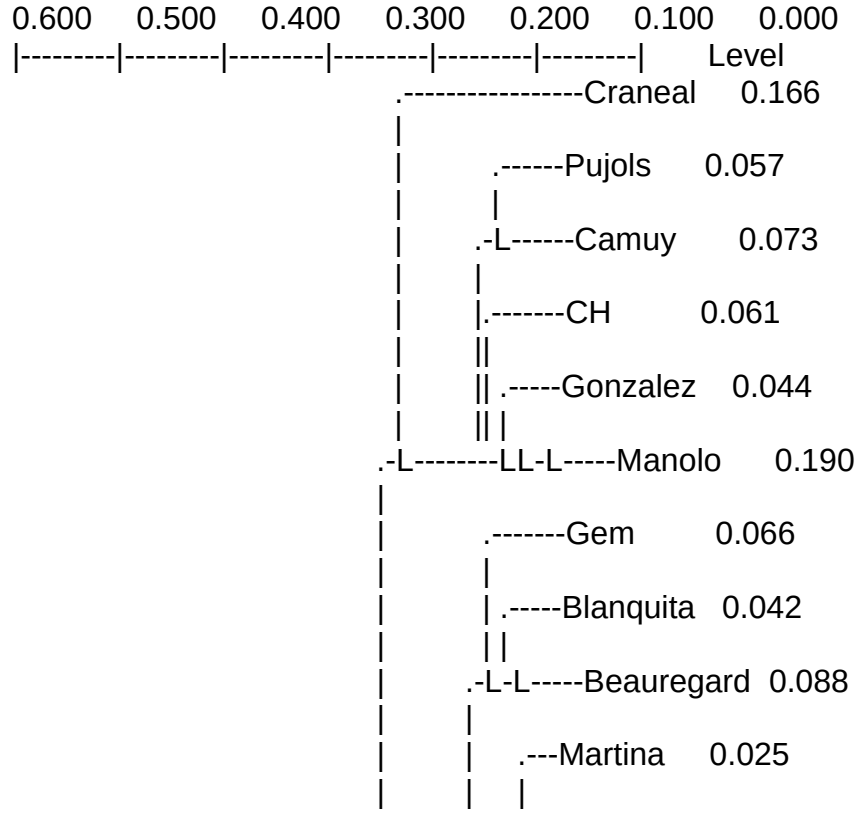
APPENDIX E

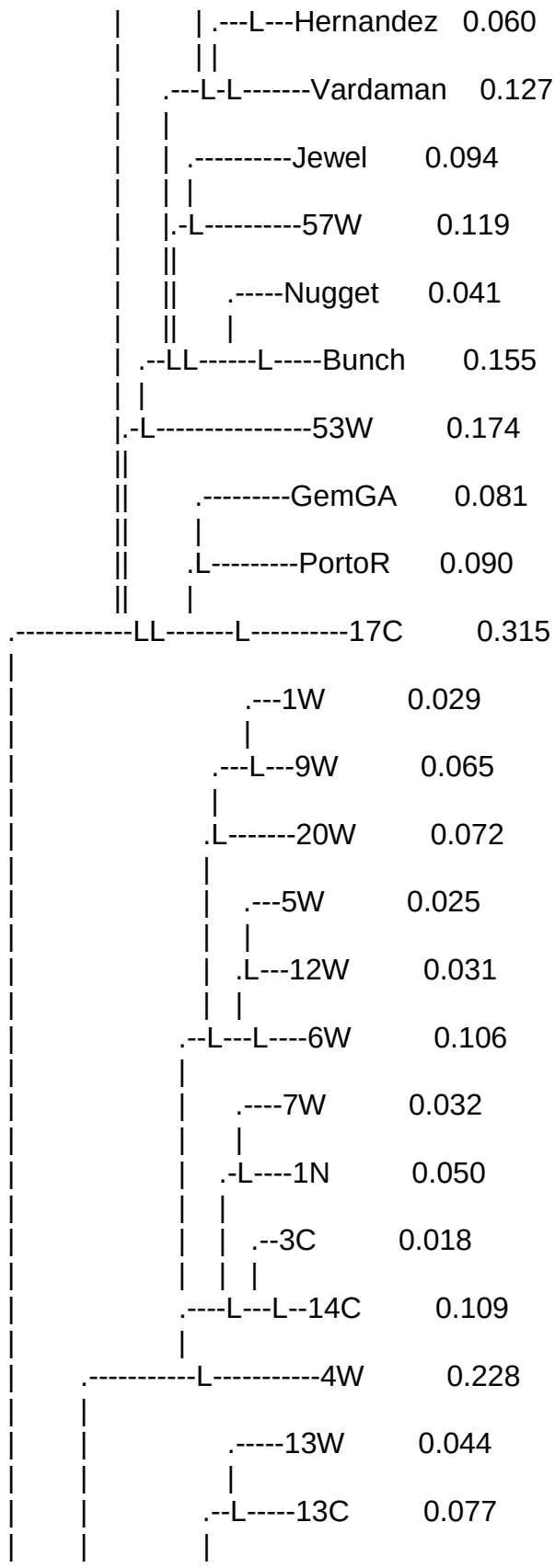
NTSYS Output

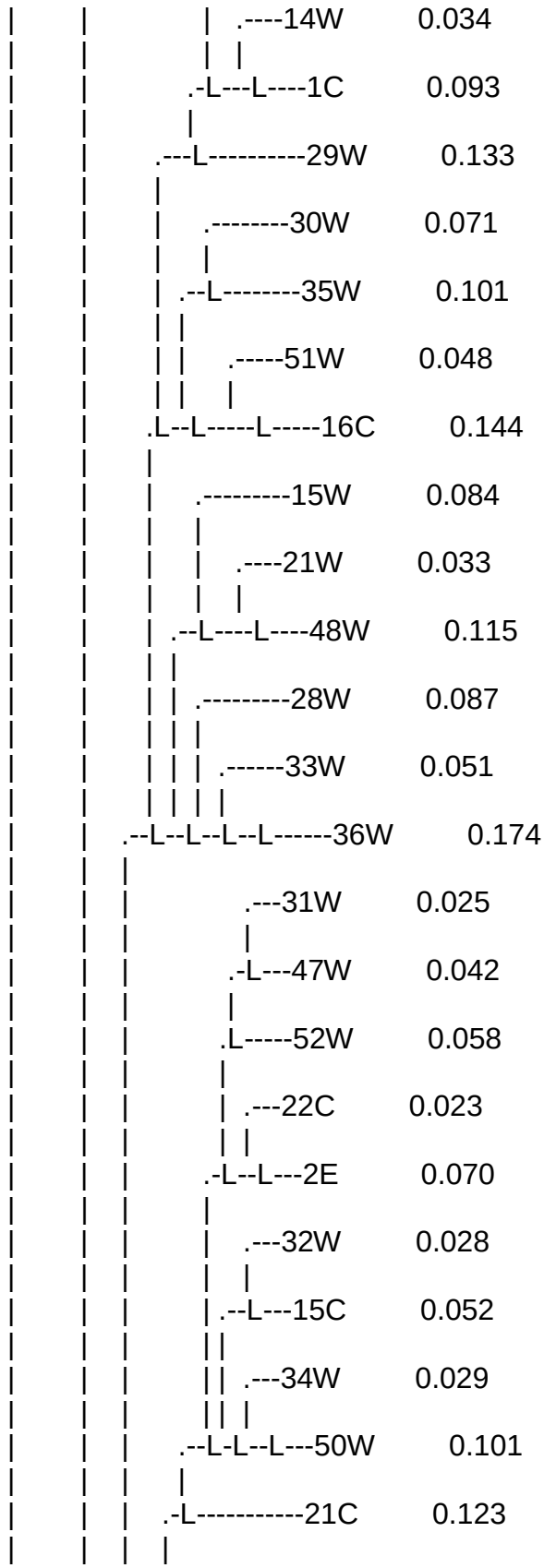
SAHN: NTSYSpc 2.02i, (C) 1986-1998, Applied Biostatistics Inc.
Date & time: 3/25/2013 9:27:53 PM

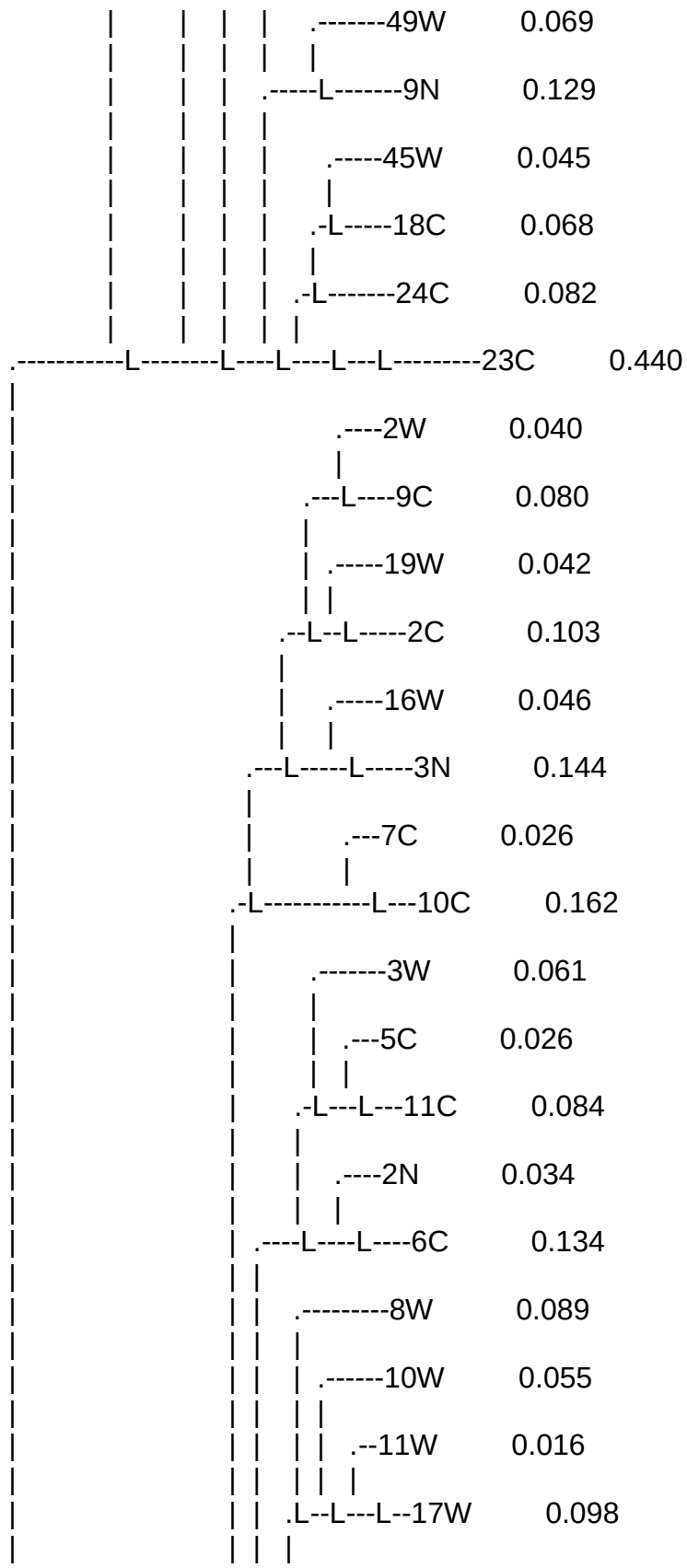
Input parameters
Read input from file: C:\Users\Lorraine\Desktop\Ntsyslast\MatSIMINT.NTS
Save result tree in output file: arbol
Clustering method: UPGMA
In case of ties: WARN

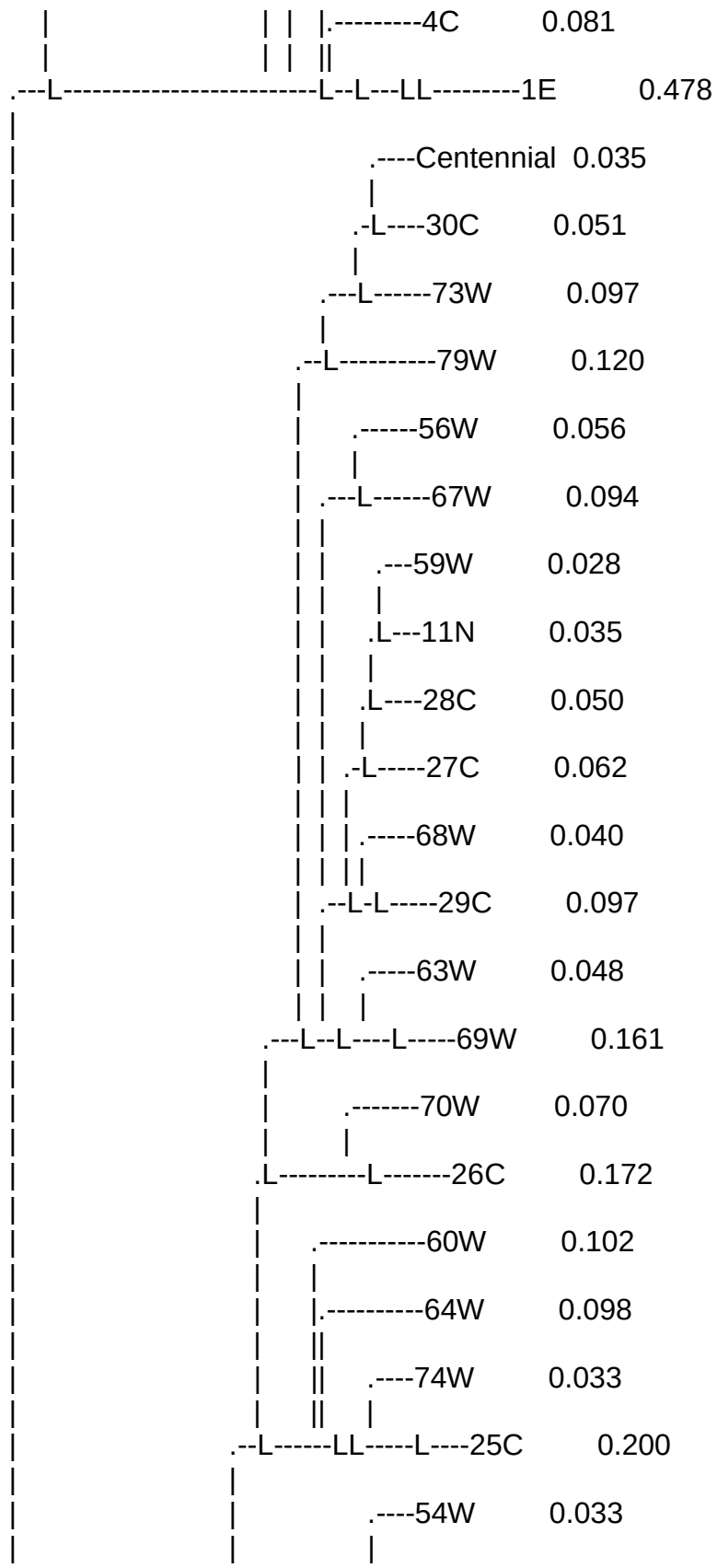
Comments:
SIMQUAL: input=C:\Users\Lorraine\Desktop\Ntsyslast\mat.NTS, coeff=J
by Rows, += 1.00000, -= 0.00000
DCENTER: input=C:\Users\Lorraine\Desktop\Ntsyslast\matSIM type was = 3
EIGEN: input=C:\Users\Lorraine\Desktop\Ntsyslast\matCen.NTS, k=3 vectors, length=SQRT(LAMBDA)
SIMINT: input=C:\Users\Lorraine\Desktop\Ntsyslast\MatVEC, coeff=EUCLID, direction=Rows
Matrix type =2, size =155 by 155, missing value code ="none" (dissimilarity)
Results will be stored in file: arbol

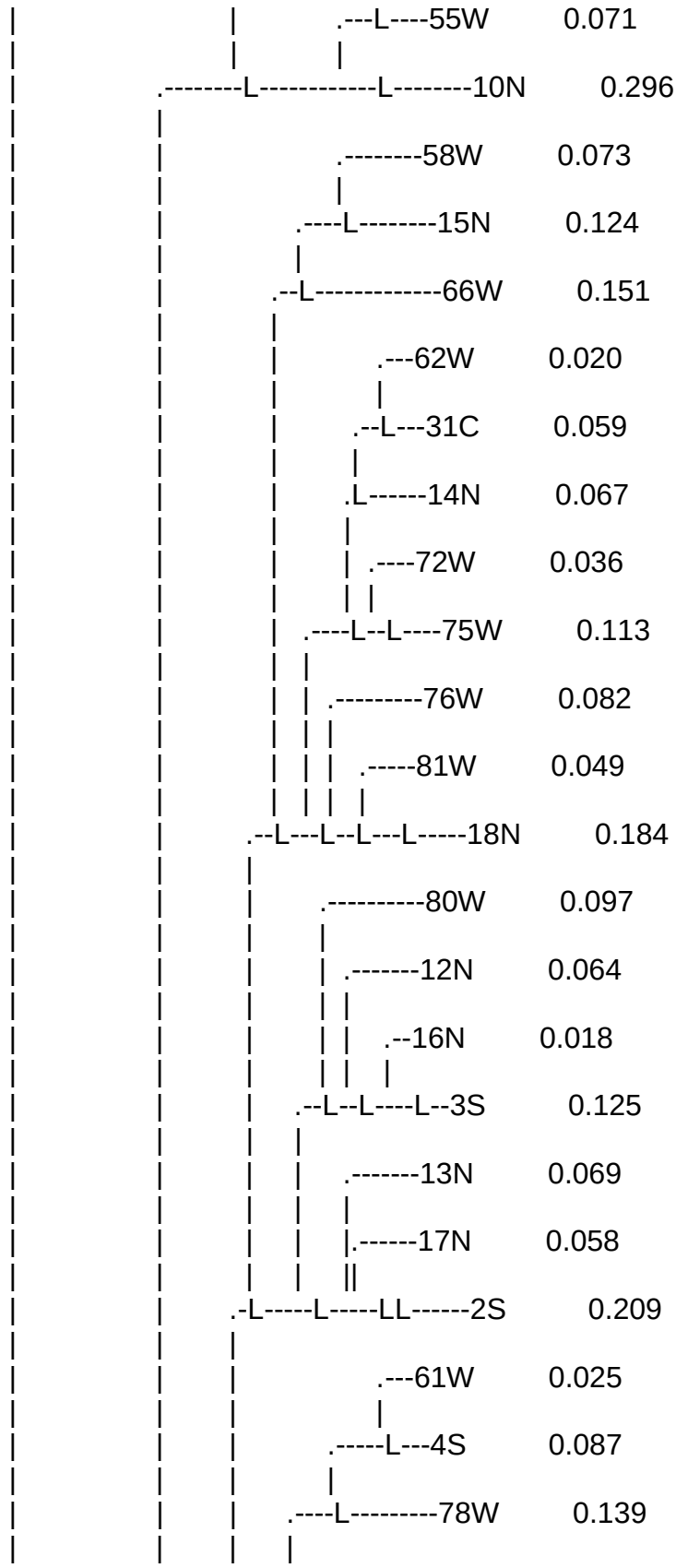


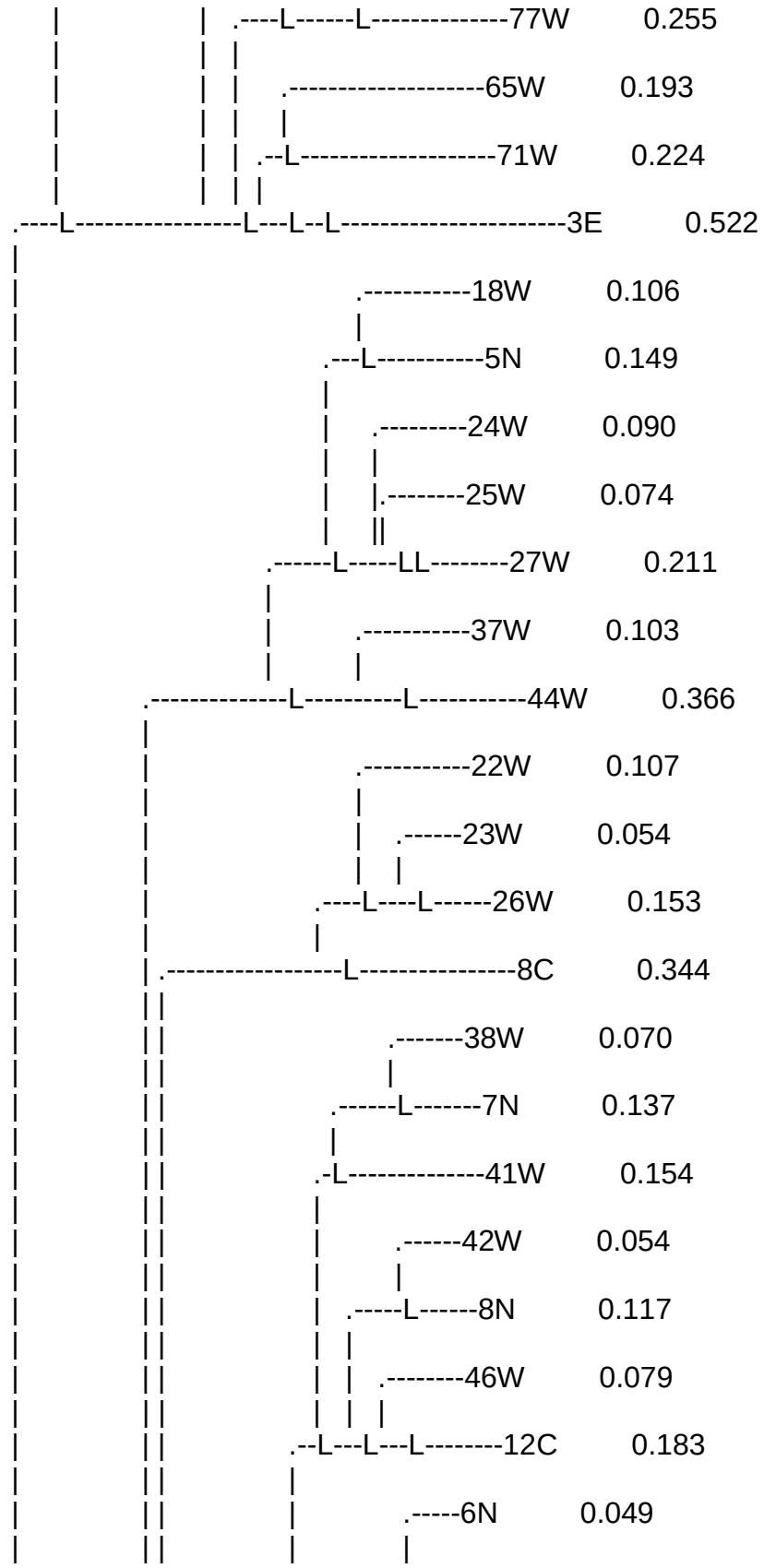

Hernandez 0.060
Vardaman 0.127
Jewel 0.094
57W 0.119
Nugget 0.041
Bunch 0.155
53W 0.174
GemGA 0.081
PortoR 0.090
17C 0.315
1W 0.029
9W 0.065
20W 0.072
5W 0.025
12W 0.031
6W 0.106
7W 0.032
1N 0.050
3C 0.018
14C 0.109
4W 0.228
13W 0.044
13C 0.077

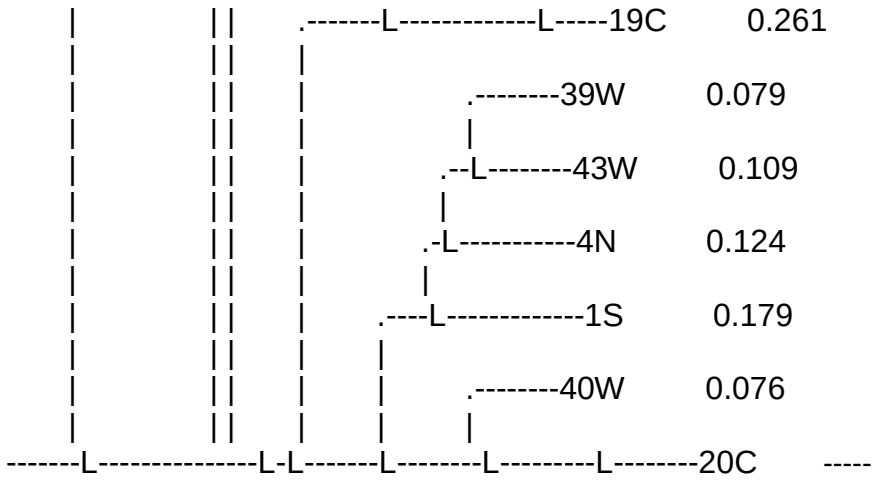

14W 0.034
1C 0.093
29W 0.133
30W 0.071
35W 0.101
51W 0.048
16C 0.144
15W 0.084
21W 0.033
48W 0.115
28W 0.087
33W 0.051
36W 0.174
31W 0.025
47W 0.042
52W 0.058
22C 0.023
2E 0.070
32W 0.028
15C 0.052
34W 0.029
50W 0.101
21C 0.123


49W 0.069
9N 0.129
45W 0.045
18C 0.068
24C 0.082
23C 0.440
2W 0.040
9C 0.080
19W 0.042
2C 0.103
16W 0.046
3N 0.144
7C 0.026
10C 0.162
3W 0.061
5C 0.026
11C 0.084
2N 0.034
6C 0.134
8W 0.089
10W 0.055
11W 0.016
17W 0.098


4C 0.081
1E 0.478
Centennial 0.035
30C 0.051
73W 0.097
79W 0.120
56W 0.056
67W 0.094
59W 0.028
11N 0.035
28C 0.050
27C 0.062
68W 0.040
29C 0.097
63W 0.048
69W 0.161
70W 0.070
26C 0.172
60W 0.102
64W 0.098
74W 0.033
25C 0.200
54W 0.033


55W 0.071
10N 0.296
58W 0.073
15N 0.124
66W 0.151
62W 0.020
31C 0.059
14N 0.067
72W 0.036
75W 0.113
76W 0.082
81W 0.049
18N 0.184
80W 0.097
12N 0.064
16N 0.018
3S 0.125
13N 0.069
17N 0.058
2S 0.209
61W 0.025
4S 0.087
78W 0.139


77W 0.255
65W 0.193
71W 0.224
3E 0.522
18W 0.106
5N 0.149
24W 0.090
25W 0.074
27W 0.211
37W 0.103
44W 0.366
22W 0.107
23W 0.054
26W 0.153
8C 0.344
38W 0.070
7N 0.137
41W 0.154
42W 0.054
8N 0.117
46W 0.079
12C 0.183
6N 0.049


19C 0.261
39W 0.079
43W 0.109
4N 0.124
1S 0.179
40W 0.076
20C