

Finding the Mix: Using SSR and SNP Molecular Markers to Determine Genetic Relationships in Wheat



**Sam Schulte
West Des Moines, Iowa
The World Food Prize Foundation
2012 Borlaug-Ruan International Intern**

**International Maize and Wheat Improvement Center
(CIMMYT)
El Batán, México
June 2-July 31, 2012**

TABLE OF CONTENTS

Acknowledgements	3
Personal Statement	4
About CIMMYT.....	4
The Wheat Molecular Breeding Program	5
About Wheat	5
Molecular Marker Background.....	6
Project Overview	7
Lab Work and Methodology	9
I: Beginning Work.....	9
II: Preparing the DNA.....	9
III: SSRs.....	10
A: Polymerase Chain Reaction (PCR)	10
B: Gel Electrophoresis	11
1) Agarose	11
2) Acrylamide	11
C: Results Analysis	12
IV: SNPs.....	14
A: PCR	14
B: Testing the SNPs	14
C: Results Analysis	15
V: Forming Phylogenetic Trees.....	16
VI: Comprehensive Analysis	17
VII: Results- Phylogenetic Trees.....	18
Personal Reflection	25
References	27

Acknowledgements

The work and life of Dr. Norman E. Borlaug served as a catalyst for nearly all the opportunities with which I was provided throughout this internship. His devotion to inspiring a new generation of agriculturalists provided the foundation for the Borlaug-Ruan International Internship, and his work led to the success of the International Maize and Wheat Improvement Center (CIMMYT).

Next, I must extend my thanks to Ms. Lisa Fleming, Director of Global Education Programs at The World Food Prize, for her extraordinary work in providing me with a seamless internship. I could always count on Lisa for prompt answers to my many questions.

Thanks to Ambassador Kenneth Quinn, The World Food Prize Foundation President, for his leadership over and constant support of the youth programs at The World Food Prize.

At CIMMYT, I would like to thank everyone in the Wheat Molecular Breeding lab. The positive attitudes and generosity of everyone there made me feel welcome and at home.

Special thanks to Dr. Susanne Dreisigacker, head of the Wheat Molecular Breeding lab, for the time she took out of her schedule in order to provide me with a wonderful project.

Also at CIMMYT, I would like to thank Claudia Núñez Ríos for serving as my main mentor in the lab. Claudia showed me the techniques to make my project a success and happily helped me whenever I asked. ¡Muchas gracias por tu ayuda!

I must also thank Miguel Santillán Martínez of CIMMYT; Miguel served as a mentor in the lab but, on the weekends, he referred to himself as “the world’s best tour guide” as he showed me Mexico City and the surrounding area. I agree with his self-proclaimed title, and I greatly enjoyed all the new cultural experiences in Mexico.

Thanks also to Norma Hernández of CIMMYT, who helped to facilitate the logistics of my arrival to, stay in, and departure from Mexico.

From my high school, I would like to thank Mr. Brant Reif, who was my faculty mentor during the Global Youth Institute. His passion for the sciences was awesome, and it served as a motivation for me to apply for this internship. Also responsible for my application is Ms. Karen Downing, my Honors Seminar teacher who encouraged me to apply.

Finally, I must extend my thanks to my family, who supported me long before and during my journey to Mexico.



Personal Statement

Since I first learned of the Borlaug-Ruan International Internship three years ago, I knew that I wanted to participate. I read the reports of many students and could not wait to go on my own internship. However, I would have to wait a few years to get involved. In coming from the largest high school in Iowa, Valley High School in West Des Moines, many other older students had priority in being selected to attend the Global Youth Institute.

Finally, as a senior, I was able to attend. I wrote my research paper on Ethiopia, and I proposed a new method of irrigation which would utilize small wells throughout the country. Writing the paper was a great experience in itself, as I learned about Ethiopia and its complex situation in producing food for its people.

The Global Youth Institute provided me with the opportunity to hear from distinguished leaders regarding their experiences with alleviating hunger. I heard a mesmerizing speech from President Lula da Silva of Brazil and dozens of others from other leaders. It was difficult for me to believe that I was in the same room with some of these people, and I was slightly confused as to why they would spend their time to talk with young people who seemed to have little influence or power.

On the final day of the Global Youth Institute, I spent 25 minutes discussing my paper with a panel of experts, though we were only supposed to spend five minutes on each paper. I honestly did not realize that time had passed so quickly, and I must apologize to the other students in the room! However, in that moment, I realized that I was truly passionate about what I was discovering and conversing. I admired the panel of experts for their thoughtful commentary on my paper. I realized why these experts had spent their time with the young participants in the Global Youth Institute. Like Dr. Borlaug, they realized the importance of the next generation in suppressing hunger. It gave me a sense of responsibility and made me feel as though I could do something impactful despite my youth and relative lack of experience.

Thankfully, I would have that opportunity through the Borlaug-Ruan International Internship. I was thrilled to learn that The World Food Prize placed me at one of my requested locations, CIMMYT, in El Batán, México. The highly enjoyable lab work solidified my choice of career. I now know that I want to work in plant pathology, a field I had never even considered before becoming involved with the World Food Prize. The experience undoubtedly changed my life in many ways.

About CIMMYT

Before its founding, CIMMYT, a Spanish acronym for the International Maize and Wheat Improvement Center, was a research program founded by the Rockefeller Foundation and Mexico in 1943. From the beginning, CIMMYT was dedicated to an international approach to improving agricultural practices. Different experimental stations were set up across the globe in order to test the growth of corn and wheat in different environments.

In Mexico, Dr. Norman E. Borlaug and his team developed new, short wheat varieties, thereby galvanizing the success of this new research program. Mexico's own wheat production grew

substantially and served as a model for agricultural success. This model became imperative as wheat production fell in Asia in the mid 1960s. In 1966, CIMMYT was officially formed, and Dr. Borlaug's newly developed wheat varieties were sent to South Asia to help rectify the shortcomings in wheat. Wheat production doubled, and the Green Revolution began. This was the first step in Dr. Borlaug's work in saving the lives of over one billion people (CIMMYT, 2012).

Thus, since the beginning, CIMMYT has been committed to helping people across the globe. To this day, CIMMYT looks to empower smallholder farmers through improved wheat and corn crops. CIMMYT is now a part of the Consultative Group on International Agricultural Research (CGIAR) and serves as a headquarters for other CIMMYT locations both in Mexico and around the world.

CIMMYT and the World Food Prize have had a longstanding relationship. Dr. Norman E. Borlaug served as a leader for both programs, and Borlaug-Ruan International Interns have traveled to CIMMYT since the founding of the program. I hope that my work has served to further this prosperous relationship, one which has surely benefited many across the world.

The Wheat Molecular Breeding Program

While at CIMMYT, I worked in the Wheat Molecular Breeding program under Dr. Susanne Dreisigacker. Generally, the Wheat Molecular Breeding program uses molecular markers to help plant breeders better know the genetic content of their wheat plants, allowing for easier and more effective parental selection. This process, called marker assisted selection (MAS), provides breeders with information that can produce plants with more desirable alleles.

As a Borlaug-Ruan International Intern, I knew I wanted to work with plants and with the hard sciences; my placement in this program, therefore, was a perfect fit. In the lab, I worked mainly with a small group of people, though my main mentor was Ms. Claudia Núñez Ríos. This project would not have been possible without her continuous help and guidance.

About Wheat

Wheat occupies more land space than any other crop in the world and is traded more than all other crops combined (Dixon, 2009). Wheat also contributes more to human sustenance than any other crop (Dixon, 2009). Thus, wheat's importance is easy to see. Unfortunately, this vital crop may be overstretched in the future. The world's population is growing at approximately 1.2% annually (CIA Factbook, 2011), yet wheat production is expected to grow at only 0.9% annually through 2050 (Dixon, 2009). Aggregately, wheat prices, when adjusted for inflation, are expected to rise by 40% by 2050 (Dixon, 2009).

The production of wheat will continue to be an issue in the future, as the supply must meet the demands of an increasing population. In this concern to feed the world, the farmer must be placed at the forefront of our considerations. Ultimately, it is the farmer who must be able to grow the food and the one who has to be able to make a profit in doing so.

These farmers who choose to grow wheat can grow wheat of two colors: red or white. Both types of wheat have their own respective purposes. Generally, red wheat is used to make bread flour, while white wheat is used to make crackers, cereals, and breads with a white crust. Within those categories, wheat can further be broken down into a hard and soft variety and into a Winter and Spring variety (USDA ERS). However, for the purposes of this paper, I will solely be looking at the differences between red and white wheat.

These two types of wheat, therefore, serve different purposes. Each type has a distinctive taste and appearance. The two are not mutually exclusive; in fact, some use a hybrid of the two in baking. Farmers, however, must scrupulously separate the two types of wheat, as it is possible for the different colors of wheat to mix with each other.

When this occurs, the wheat can no longer be classified as red or white. Consequently, the farmer is unable to gain certification that the wheat is of a pure variety; it is simply labeled as a mixed type of wheat. This inability to gain certification puts the farmer at great financial risk. The ability to certify wheat as a pure, specific type is critical to gaining a healthy profit. Mixed wheat sells for significantly less money than pure wheat (Pasikatan, 2003). This impact is not an isolated issue, as the mix continues throughout generations.

After a mix occurs, it is difficult to know how or when the mix happened. There is no simple way to look at a seed and to be able to extrapolate the lines of wheat which mixed to produce the contaminated seed.

Molecular Marker Background

One effective way to discover this mix is through the use of molecular markers. A molecular marker can generally be defined as an identifier of a specific part of a genome. This identifier can be either a short sequence of DNA or a single nucleotide. These molecular markers describe a variance, or polymorphism, in the alleles of the genome.

Molecular markers of interest are typically those which are near located near certain genes on the string of DNA. When new generations of plants are made, traits that are located more closely to each other on the DNA are more likely to both be inherited. Therefore, when molecular markers are found near traits of interest on the DNA, the presence of that marker can also indicate if the trait is present, because the two are likely to be inherited together.

As was mentioned earlier, this marker can be either a short sequence of DNA or a single nucleotide. The first type of marker, a short sequence of DNA, is referred to as a microsatellite or single sequence repeat (SSR). Here, a short sequence of DNA repeats a specific number of times. Often, the number of repetitions varies between plants within the species, and the number of repeats can indicate the presence or absence of an allele. Therefore, in testing for SSRs, the number of repeats determines the presence or absence of a certain trait. SSRs can test for numerous different alleles, because the number of repeats can vary widely.

The other type of molecular marker is contingent upon a difference within a single nucleotide between different plants. This variation, called a single nucleotide polymorphism (SNP), is what causes different alleles. Therefore, just one nucleotide can make the difference between the

presence and absence of a trait. SNP tests are much faster than SSR tests, allowing for more tests to be done in a short period of time.

In the wheat molecular breeding lab at CIMMYT, SSRs have served as the main study in molecular markers, and many alleles have been identified through using these microsatellites (CIMMYT's Experience). As we continue to learn more about wheat, more thorough linkage maps will be constructed, allowing for more alleles to have associated microsatellites. Though not widespread now, it is expected that SNPs will become the predominant molecular marker in laboratories, as they are both less expensive and more efficient.

The use of these molecular markers has been seen as advantageous in many respects. Markers are able to detect recessive genes, eliminating many time consuming tests that are often done to determine what the recessive genes are. Markers can also be used very early in a plant's life, allowing for results to be generated quickly.

When marker based selection is included in a breeder's repertoire, this can help the breeder to more easily make decisions about backcrossing. Backcrossing is the integration of new, desirable traits into an existing line. Through the use of molecular markers, the desired genotype is usually able to be integrated within three backcrosses. Compare this to traditional approaches, in which approximately six backcrosses are needed. Molecular markers, therefore, increase the speed at which desirable traits can be integrated into wheat varieties.

There are, however, some obstacles to using molecular markers. There are relatively few markers that have been developed for wheat, and many lack confidence in the reliability of markers. Wheat's genome is especially complex, as it is allohexaploid (CIMMYT's Experience); modern wheat is a combination of three wild species of wheat: *Triticum uratu*, *Aegilops tauschii*, and *Aegilops speltoides*. This gives rise to three genomes: A, B and D. Each genome contains seven chromosomes. It also consists of 16,000 Mega bases, more than six times more bases than in corn. Of all these bases, approximately 99% of bases are simply repeated sequences that do not code for genes. Wheat also has low levels of polymorphism between linkage groups, meaning that finding distinguishing alleles between different variations of wheat is often difficult (William, 2007).

Project Overview

Through my research, I used the aforementioned molecular marker techniques to determine where a mix between red and white seed of wheat occurred. I evaluated 97 separate wheat samples; each of these 97 samples belonged to one of nine lines of wheat. Of these nine lines, the mixed line was suspected in line 4, named "Villa Juarez." This line differed from the original Villa Juarez line in color, so a mix was consequently expected within the Villa Juarez line. When I received the samples, it was not known with which other line line 4 had mixed.

The Villa Juarez line came from the National Institute for Forestry, Agriculture, and Livestock (INIFAP) in Obregón, a city in northern Mexico. The program was ready to disseminate the Villa Juarez line to local farmers for their use; this, however, did not occur due to the suspected mix. Consequently, three samples from line 4 Villa Juarez and seeds from the original Villa Juarez line were both sent to CIMMYT in order to be compared and to determine both if a mix

had occurred and, if so, where that mix had occurred. Seeds from other lines were sent as well, as those lines were used to attempt to discover where the mix had occurred.

To find this mix, I used molecular markers to test each sample for 25 different polymorphic loci, or sites on the genome, to compare which samples had which alleles. The similarity of loci between lines would indicate which lines were most closely related. The markers I chose to test were holistically irrelevant; the only stipulation I observed was choosing markers from many different chromosomes. Testing from only the same chromosome would not have provided me with comprehensive information about the genetic makeup of each line. Therefore, I was broad in my aim to compare the lines of wheat. I was not concerned with what those markers implied about the specific traits the samples had. Rather, I was concerned about how those loci varied between lines and samples.

The one set of markers for which I purposefully tested were markers for the color of wheat. New molecular markers have recently been developed which are able to detect the color of wheat. This is especially useful when the color of wheat cannot be determined by looking at the seed, as the color's appearance can often be ambiguous. Because of the relatively recent development of these markers, CIMMYT has never used these markers before; I was the first at CIMMYT to use these markers. Therefore, I had no prior experience to refer to when testing the markers. I had to manipulate many conditions in order to produce any results. Even after these efforts, I still did not obtain conclusive results. This, however, does not necessarily indicate that the color markers are flawed. Rather, this shows that more time is needed on these markers so that the conditions can be optimized. Because the results of the color markers were inconclusive, they were not considered in the comprehensive analysis of my data.

After gathering this data, I was able to form a phylogenetic tree, which shows which samples were most closely related by similarity in loci. Hypothetically, if a mix had not occurred, then the three samples from line 4 would have paired with the original Villa Juarez line in the phylogenetic tree.

My results indicated that one of the samples from line 4 still belonged to the original Villa Juarez line, indicating that no mix had occurred there. The other two samples, however, paired with line 7, "Tepahui," in the phylogenetic tree, indicating that those two samples came from a mix in line 7. This information was reported to INIFAP so that the mix can hopefully be rectified and avoided in the future.

The following section describes in detail how I came to these conclusions.



Lab Work and Methodology

I: Beginning Work

Before I could begin to work in the lab, however, the samples first had to be taken from the plants. Because of the considerable distance between the research station in El Batán and the research station in Obregón, I could not take part in the sample extraction.

I did, however, have the opportunity to partake in the extraction of samples at the Toluca research station, which is a more manageable two-hour drive from El Batán. The process of collecting samples began by placing labeled stakes in the ground with two numbers on them. One number indicated the row, while the other number indicated the number of the sample in that row. After placing the wooden stakes in the ground, we returned with plates which held 96 tubes. A small sample was cut from each plant and placed into a tube, each of which corresponded to a large chart which told which tube contained which sample, as denoted by the wooden stake. This process was somewhat complicated, because the plants were very young and difficult to distinguish from small weeds.

II: Preparing the DNA

After the samples were collected in Obregón, they were delivered to the research station in El Batán. There, the DNA had to be extracted from each sample. Again, I did not have the opportunity to extract the DNA from my samples; I did, however, assist in extracting the DNA from the samples we collected at Toluca.

First, the leaves were frozen in liquid nitrogen and stored at -80°C . Then, they were put into a lyophilizer, a chamber which dehydrates the samples. Through this process, the pressure inside the chamber is lowered, allowing for sublimation of the water.

After the material was dehydrated, a small metal ball was placed into each tube with a sample. Then, the plate with the tubes, samples, and metal balls was placed into a machine which violently shook back and forth at 30 times per second for three minutes. This process pulverized the samples into a powder.

Subsequently, 400 μl of Cetrimonium bromide (CTAB) was added to each tube through the use of a multi-pipette. Caps were then put on each tube. Throughout this process, it was essential that the same caps were used for the same tubes, as using a different cap could easily contaminate the DNA of two different samples. Then, the samples were placed in the oven at 65°C for 90 minutes. While in the oven, the samples were on a machine which shook them in a circular motion at approximately 100 rpm. Then, the samples were removed from the oven. 300 μl of a chloroform – Octanol (24:1) solution was then added to each tube. Then, the plates were placed on a slow, rotating axis for 15 minutes. Subsequently, the tubes were centrifuged at 3750 rpm for 30 minutes. This produced a supernatant, 300 μl of which was transferred to new tubes. This supernatant was then placed in an oven at 37°C for one hour. Then, 400 μl of Isopropyl alcohol was added to each tube. The tubes were then centrifuged at 3500 rpm for 30 minutes. The Isopropyl alcohol was then removed, and 400 μl of Isopropyl alcohol was added again. The samples were again centrifuged at 3500 rpm for 30 minutes. The Isopropyl was again removed,

and the remaining DNA in the bottom of the tube was left to dry. Then, 200 μl of TE buffer was added to the DNA.

Following the DNA extraction, each sample of DNA had to be quantified to determine the concentration of DNA in the TE buffer. To accomplish this, a spectrophotometer was used. 2 μl of the TE Buffer and DNA mixture was pipetted onto the sensors of the spectrophotometer, which told the concentration of the DNA. This information was then used to determine how much of the DNA and TE buffer mixture would need to be combined with water in order to bring the entire mixture to a 50 ng/ μl concentration.

After these steps were complete, I was left with extracted DNA from each of the 97 plant samples. In order to ensure that the DNA was extracted properly and was of sufficient quality, I ran a quality gel test in agarose gel by placing the DNA directly into the agarose gel wells and running the gels, a process to be described in more detail later. The results of the test verified that the DNA had been properly extracted; there were clear bands of DNA which appeared, indicating a high quality of DNA.

III: SSRs

A: Polymerase Chain Reaction (PCR)

This uniformity in DNA concentration was necessary to achieving a proper reaction in polymerase chain reaction (PCR), a process which makes copies of the DNA. In order to facilitate PCR, a reaction mix first had to be formulated. The reaction mix was based upon the number of samples which would be amplified through PCR. A few extra reactions were always prepared, because in transferring the reaction mix into the wells for PCR, some residual mixture would be left in the tube. For example, if I needed to prepare enough reaction mix for all 97 of my samples, I would prepare enough mix for 105 reactions. For each reaction, I added 1 μl of 10x Buffer, 0.3 μl of MgCl_2 , 0.8 μl of dNTP's (deoxyribonucleotides, which are single units of DNA), 2.5 μl of 1 μM Primer, and 0.4 μl of Taq Polymerase into a 1.5 mL tube. This tube was then placed on a vortex shaker for a few seconds in order to mix the contents evenly. After this, 5 μl of DNA from each of the samples was placed into a 384 well plate. To those DNA samples, 5 μl of the PCR reaction mix was added. Then, the plate was centrifuged at 1500 rpm for 4 minutes. The plate was then put into a thermal cycle, the machine which accomplishes PCR.

The thermal cycle works through three main phases of temperatures. It begins with a high temperature, which was 94° C in this case. This first phase, called denaturation, breaks apart the two strands of DNA. Then, at a lower temperature of 50-60° C, the primers join with the open DNA strands which have just been broken apart. This second phase is called the annealing phase. Here, the two primers have a set DNA sequence which only pairs with one location on the DNA; one primer pairs with the beginning part of the sequence to be replicated, while the other primer pairs with the end of the sequence to be replicated. This means that PCR only replicates one portion on the DNA sequence, not all the DNA. The final phase is called the extension phase. In this phase, which occurs at 72° C, the Taq DNA polymerase replicates the DNA strands by starting at the place on the DNA where the beginning primer annealed. The replication continues until the Taq DNA polymerase reaches the end primer. Taq DNA polymerase is used because it is stable at high temperatures. After this process has occurred for one cycle, there are two complete DNA strands. These cycles are used anywhere from 30-45 times. Sometimes, more

DNA needs to be replicated in order to obtain results; in this case, more cycles are used. Millions of copies of the short DNA sequence are produced after just 30 cycles.

When facilitating PCR, I chose different primers and Taqs. Each primer is complementary to a specific position on the DNA, so I chose primers that corresponded to specific loci on the DNA. Thus, each PCR would use the primer for the allele I was investigating at that time. Altering the choice of Taq would sometimes produce better results, as some primers work more effectively with certain Taqs.

B: Gel Electrophoresis

After the PCR was finished, gel electrophoresis was used. Gel electrophoresis is a process which shows if a sample has the allele in question. When PCR occurs, the primers attach to sequences on the DNA which is common to all the plants in that species. Between that beginning and ending sequence, however, there is variation between individuals that have an allele and individuals that do not have an allele. Thus, as an example, an individual with an allele might have 600 base pairs between the two primers, while an individual without an allele might have only 300 base pairs between the two primers. When placed into the gel for gel electrophoresis, these differences in the number of base pairs become apparent, as gel electrophoresis functions by separating bands of DNA by size. This occurs through the use of an electric charge; the negatively charged DNA is pulled toward the positive charge. Segments of DNA which have more base pairs have a more difficult time moving through the gel, because it is more difficult for more base pairs to fit through the pores in the gel. The shorter segments of DNA, therefore, move through the gel more quickly. Consequently, when looking to the example, the segment without the allele (300 base pairs) would have moved much farther in the gel than the segment with the allele (600 pairs). When comparing these distances with a control, which is known to have or not have the allele, it can be determined which samples have the allele.

To accomplish this gel electrophoresis, I used two types of gels: agarose and acrylamide.

1) Agarose

I used 3% agarose gels for some of my gel electrophoresis. To prepare the gel, I combined 6 grams of agarose and 200 mL of 1x TBE Buffer in a flask. I then added approximately 25 mL of water and microwaved the gel until enough water evaporated so that the liquid would again be at 200 mL. I then carefully poured the hot liquid into a mold. After pouring the liquid, I put combs with 50 teeth into the gel. These teeth would eventually create the wells into which I would load the PCR product. After the gel cooled, I added 5 μ l of blue dye to each of the PCR products; this dye allowed me to see when the gel electrophoresis had finished. I then loaded 15 μ l of the PCR product and blue dye solution into each of the wells. I also added 4.0 μ l of a molecular weight mixture to a well in each row, as this would allow me to see how many base pairs each segment of DNA had by comparing the bands with the molecular weight bands. I turned the battery on to 150 Volts and let the gel run for about 45 minutes. After it was finished, I took the gel out and carefully placed it in ethidium bromide for approximately 15 minutes in order to stain the gel. After the gel was stained, I took a picture of the gel by placing the gel over ultraviolet light.

2) Acrylamide

The fundamental principles behind acrylamide gels are the same as those of agarose gels, yet the procedure is somewhat different in order to achieve the end result. For each 12% acrylamide gel,

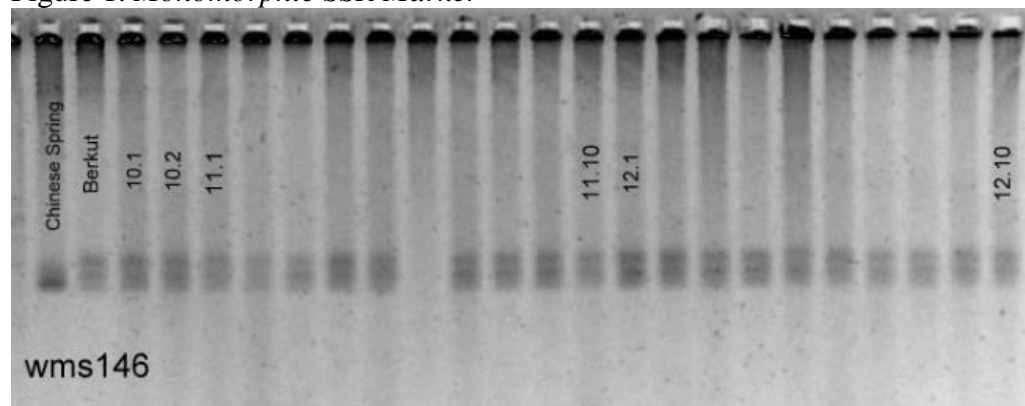
I combined 30 mL of 40% acrylamide, 50 mL of double distilled water (ddH₂O), 20 mL of TE Buffer, 350 µl of 25% APS (Ammonium persulfate), and 50 µl of TEMED (tetramethylethylenediamine) in a 250 mL flask. After combining these, I had to move relatively quickly, because the acrylamide gel begins to set within a few minutes. To make this gel, I poured the liquid acrylamide mixture between two glass plates. I then placed a comb with 62 teeth between the two glass plates. Once the comb was in place, I pipetted in any extra liquid acrylamide mixture to ensure the two glasses would be full of gel. I then let the gel sit. Often, I would prepare the gel the day before I needed it in order to conserve time the next day.

After the gel set, I placed the glass plates and gel into a structure which held the buffer. After adding the buffer, I removed the comb and added my samples to the wells. I then added 1.2 µl of a molecular weight to one well in each gel. I also added 1.2 µl of blue dye to one well, as this would allow me to see when the gel had finished running. I turned the battery on to 250 Volts and let it run for approximately 3.5 hours. After the gel had run, I removed the gel from the glass plates and placed it in fixer for five minutes. I placed the fixer (and all following solutions) and gel on a slow-moving shaker so that the fixer coated the entire gel. I then removed the fixer and placed the gel in a silver nitrate aqueous solution for approximately 15 minutes. I then replaced the silver nitrate aqueous solution with developer, and I let the gel sit in the developer until the bands of DNA were clearly visible. I then removed the developer and added a stopper solution. After this, I placed the gel on a visible light source and took a picture.

C: Results Analysis

As I mentioned earlier, I tested for a total of 25 markers through both SSRs and SNPs. However, I actually tested more than 25; I simply did not use some of the primers I tested, as I did not incorporate primers which were not polymorphic across the nine lines of wheat. To conserve materials and to expedite the process, I began my research by testing only one sample from each of the nine lines. Then, when comparing these samples after gel electrophoresis, I looked at which molecular markers yielded the same result across all nine lines. These results were denoted as monomorphic, meaning that all nine lines were genetically similar in this allele. Because the goal of my project was to use the differences in alleles to determine how similar or different the nine lines were, markers that yielded consistent results across the lines were not useful. Therefore, I narrowed down the number of markers used to 25, the number of polymorphic markers. The following picture is an example of a monomorphic marker.

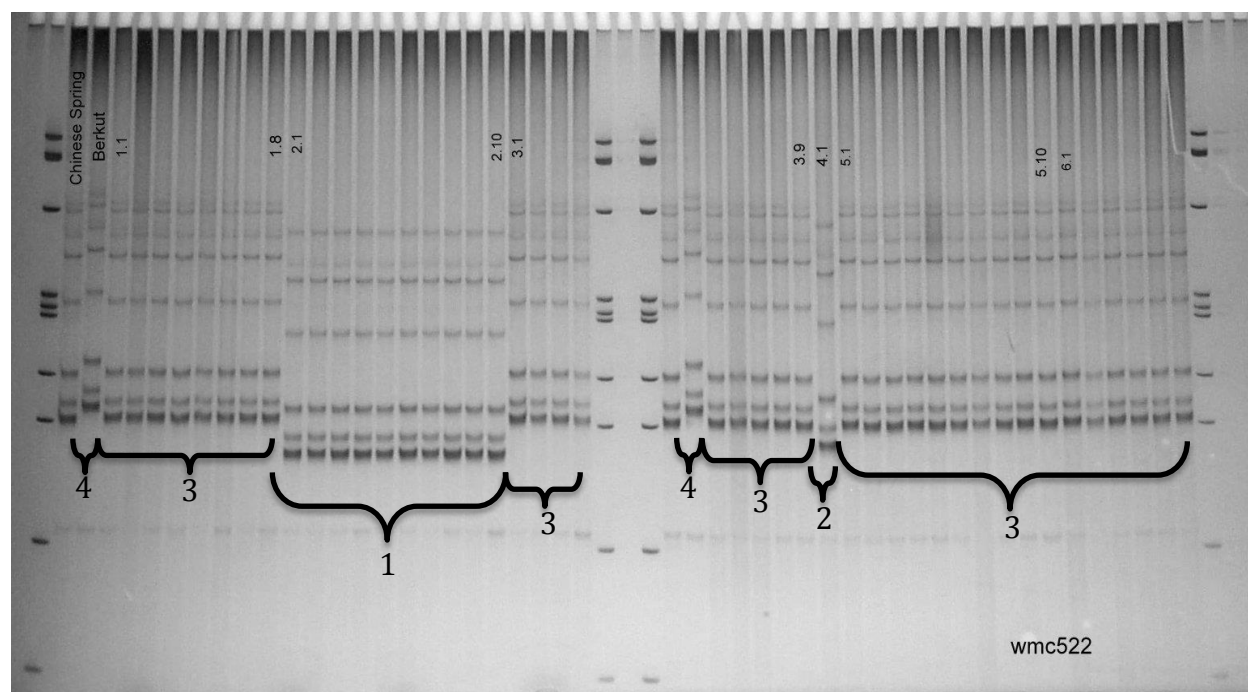
Figure 1. *Monomorphic SSR Marker*



Molecular markers which were polymorphic revealed information about the alleles present and absent within the different samples. For example, the following picture reveals which samples shared the same allele and which did not.

There are four alleles present within this acrylamide gel. The numbers on the gel indicate the wheat line and sample of each sample. Therefore, the first sample on the left is sample 1.1, meaning the sample is first sample from wheat line one. Then, 2.1 indicates that that sample is the first sample in wheat line two. The label “wmc522” is the marker for which I tested all the samples. Not all samples are present in this picture, because one acrylamide gel could not hold all 97 samples.

Figure 2. *Polymorphic SSR Marker*



In analyzing this picture, I found that there were four different alleles present in this test. I labeled each allele by starting with the one whose band was closest to the bottom of the picture. Therefore, the bands in line 2 (2.1-2.10) were listed as Allele 1. These bands all traveled the same distance from the top of the picture in gel electrophoresis, indicating that these samples have the same allele. The next allele, Allele 2, is in sample 4.1, as that band is slightly higher up the picture than the band of Allele 1. The next allele, Allele 3, is present in lines one, three, five, and six. The final allele, Allele 4, is only present in the control, Berkut. The two controls are samples with known presence or absence of an allele and are normally used to compare with the samples with unknown genetic material. Here, however, the controls were not used in the results, as the presence or absence of the allele was not my primary concern; rather, I was concerned with the differences in alleles between samples.

After analyzing this picture and the rest of the samples that were tested with the wmc522 marker,

I took these results and entered them into an Excel spreadsheet. I placed a “1” in the column if the allele was present and a “0” if the allele was absent. If the sample did not amplify, I placed an NA in the place of the sample. These results would later be used to make the phylogenetic tree.

IV: SNPs

After completing most of my SSR tests, I then moved on to testing SNPs. SNPs are a new form of testing at CIMMYT. In fact, everyone from the wheat molecular breeding lab had a meeting to discuss what SNPs are, as SSRs have served as the main focus. As I mentioned before, SNPs are a different type of molecular marker that looks to one nucleotide for variation. For example, take the nucleotide sequence GAAGCTGGAG[C/T]AGCTCGAGAT, a sequence which codes for plant height. In that sequence, the C/T in brackets shows a SNP. If the nucleotide is C, then the plant is a regular, wild type plant. However, if the nucleotide is T, the plant is a dwarf. This is, therefore, a single nucleotide polymorphism.

A: PCR

Preparing the PCR was relatively similar to preparing the PCR for the SSR tests, yet the reaction mix was slightly different. Prior to making the mix, 3.0 µl of DNA from each sample was placed into a 384 well plate. These plates were then placed in a 60°C oven until all liquid evaporated, leaving just the dried DNA. This process took about 50 minutes.

After that, I added 2.5 µl of 2x KASP Reaction Mix, a premade reaction mix which includes MgCl₂. Then, I added 2.4 µl of ddH₂O, though the reaction only calls for 0.2 µl of ddH₂O. However, the thermal cycles require a minimum of 5.0 µl of liquid per well in the 384 well plate, so more ddH₂O was added. Finally, I added 0.1 µl of Assay Mix, which contains the necessary primers and corresponding dyes, a topic to be addressed below.

After the mix for the PCR was ready, I put the plate into the thermal cycle. The temperatures are similar to those in PCR for SSR tests, yet there are a few differences. PCR for SNPs begins with a 94° C phase for 15 minutes. This is followed by 10 cycles of what is called a “touchdown” program. Each cycle involves an initial 94° C phase for 20 seconds. This is followed by a 65° C phase for 60 seconds. However, in the 65° C part of this cycle, the temperature decreases by 0.8°C in every subsequent cycle. After these 10 cycles, there are 26 cycles of 94°C for 20 seconds and 57°C for 60 seconds.

The actual multiplication of the DNA is also somewhat different when compared to the PCR for SSRs. After the primers cut the sequence of DNA needed and amplify it in PCR, the Taq polymerase enzyme operates differently, as they include dyed probes which attach only to specific regions of DNA. Therefore, two different dyed probes are put in the PCR reaction mix; one reacts with a certain nucleotide, and the other reacts with a different nucleotide. These two different dyes are called FAM and VIC. This results in two different dyes which were attached to the different nucleotides (and subsequently, alleles) in question. For example, the C nucleotide might attach to the FAM dye, while the T nucleotide might attach to the VIC dye. Thus, the two nucleotides (and alleles) can be distinguished by looking at which dye is present.

B: Testing the SNPs

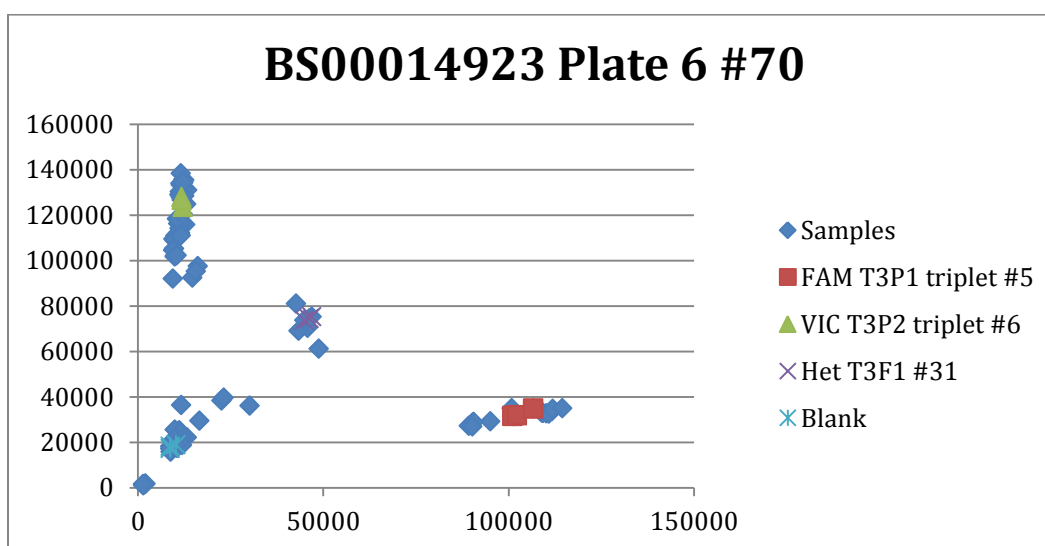
Once the PCR was finished, I put the plates into a machine that measures the entrance and exit wavelength of light in order to discriminate between the two different types of dyes. The two

dyes caused the light to have different wavelengths, thereby separating the two alleles into different groups and producing a graph which shows a distinction between which samples belong to which allele. Three FAM and VIC controls are also put onto the plate so that they will appear on the graph, identifying the two clusters that appear on the graph. Water is also measured in order to serve as a control. If a piece of DNA contains both alleles, it is heterozygous for that trait; the analysis also shows which samples are heterozygous by clustering them between the VIC and FAM clusters.

C: Results Analysis

The following graph of marker BS00014923 illustrates how the samples are separated into two groups based upon the presence of FAM or VIC dye. The FAM and VIC controls appear as the red boxes and green triangles, which correspond to the legend to the right of the graph.

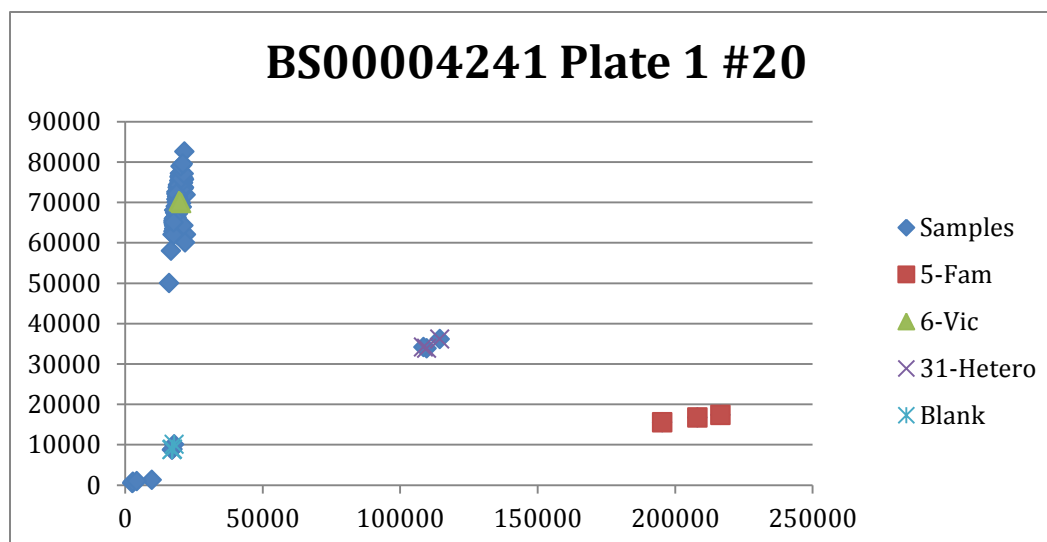
Figure 3. *Polymorphic SNP*



Thus, the cluster at the top left consists of VIC dye, while the cluster on the right consists of FAM dye. The samples on the bottom left are controls and water.

After obtaining these results, Microsoft Excel is then used to determine which samples pair with which dye. Once they were identified, they were then placed into the spreadsheet which also contained the identity of traits from the SSR tests. Again, a “1” on the spreadsheet indicated that the sample had either the FAM or VIC dye, while a “0” indicated that the sample did not have the FAM or VIC dye.

Like the SSR tests, the SNP tests also yielded some monomorphic results which were not of significant use for this project. The following graph is an example of a monomorphic SNP, which indicated that every sample had the same allele. Here, the only samples with FAM dye came from the FAM controls themselves. All actual samples had the VIC dye. A total of 13 SNP markers were polymorphic and used in my project.

Figure 4. *Monomorphic SNP*

V: Forming Phylogenetic Trees

After completing the spreadsheet with all the data about the presence and absence of alleles, I was able to compile the information to fulfill the mission of my project. The data was exported to software programs called PowerMarker and Tree View, programs which are able to compile information into phylogenetic trees. Essentially, the program takes the data from each sample and analyzes how closely that sample is related to all the other samples by the similarity of the alleles between the samples. After considering this information, the program produces a phylogenetic tree.

PowerMarker has the ability to produce trees which are contingent upon different methods of determining relationships between plants. I tested for three different mathematical methods of determining this relationship: Euclidean, Rogers, and Shared Allele. These three mathematical methods, however, generated the same phylogenetic tree each time, so the mathematical method of determining the relationships was irrelevant.

I also looked at two methods of putting together the tree, including Unweighted Pair Group Method Using Arithmetic Averages (UPGMA) and Neighbor Joining (NJ). The fundamental difference between these two methods is that UPGMA trees assume that the evolution of the tree occurred at a constant pace. NJ trees, however, recognize that this evolution may take place at alternative rates. Therefore, the NJ tree appears more uprooted than a UPGMA tree. The NJ tree is also more frequently used for data which is based upon DNA, something which was applied in my project. These two methods of trees did produce slightly different results.

In addition to looking at different mathematical methods, I also looked to different methods of gathering data. I constructed some trees using only data from the SSR tests and constructed other trees using only data from the SNP tests. Here, the real emphasis was placed upon the tree which

was based solely upon the SNPs, as this form of data collection is new at CIMMYT. My trees indicated that the data based upon SNPs and SSRs was mostly equal. There were a few variances, yet the two types of test produced very similar results. Therefore, my project helped to prove that SNPs are a data type which can effectively be used to determine the genetic relationship between different samples by using the same number of markers.

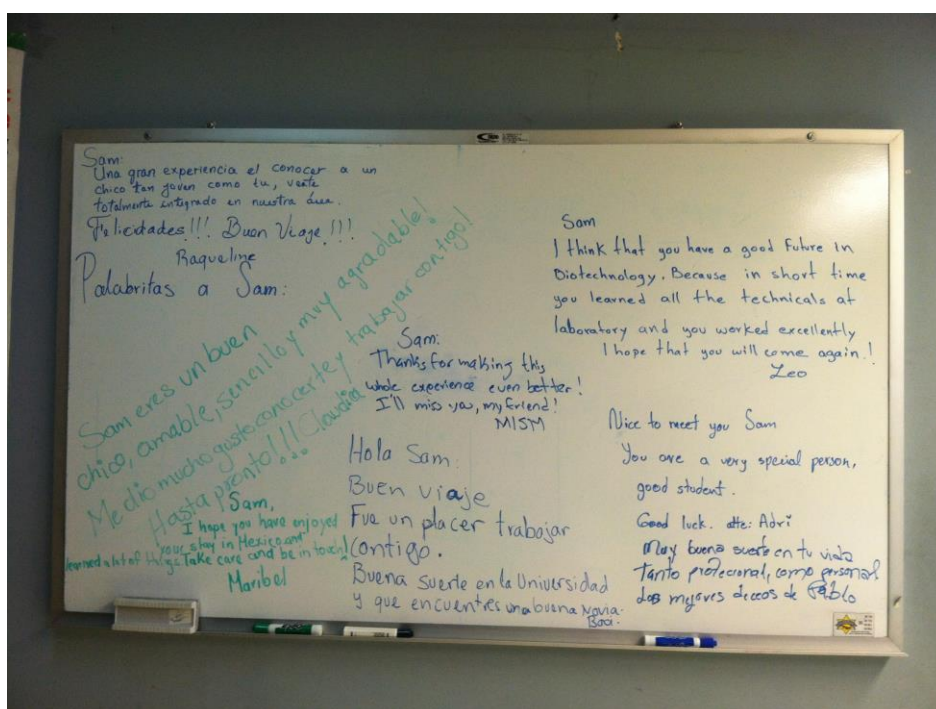
I also constructed a simpler tree by grouping the samples from each line of wheat together in order to form only nine data points on the tree (Figures 5 and 6). These trees help to more clearly illustrate how the lines are related to each other, though they may not be as accurate, as there were mixes present within the lines.

VI: Comprehensive Analysis

Most of the trees below show that the first sample in line 4 (4.1) grouped with the original Villa Juarez line, indicating that no mix occurred. The other two samples in line 4 (4.2 and 4.3) grouped with line 7. Thus, this indicates that these samples are not Villa Juarez but are a mix of line 7, “Tepahui.” INIFAP had expected that the Villa Juarez had mixed with line 9, “Kronstad.” However, the phylogenetic trees show that this was not the case. This project, therefore, confirmed that the seed of Villa Juarez was mixed and that that mix was derived from line 7.

When I started the work on my project, the aim was to have at least 40 polymorphic markers which would be used to construct these trees. However, because of time constraints, I was only able to find 25 polymorphic markers in total. Having more time to test more markers would have given more verisimilitude to the results, as this would have provided a more holistic view of the genetic makeup of each of the samples. However, the 25 markers have still provided definitive results.

Thus, this project found the mix that had occurred with line four through the usage of SSR and SNP molecular markers. Such data can be used by scientists at the Obregón research station to help avoid this same mix in the future, a mix which can end up costing the farmer immensely over time.



“Palabritas a Sam,” or “Short words to Sam.” The other members of the Wheat Molecular Breeding lab wrote these “short words” for me on my last day.

VII: Results- Phylogenetic Trees

Figure 5. All Results By Line; Euclidean; NJ

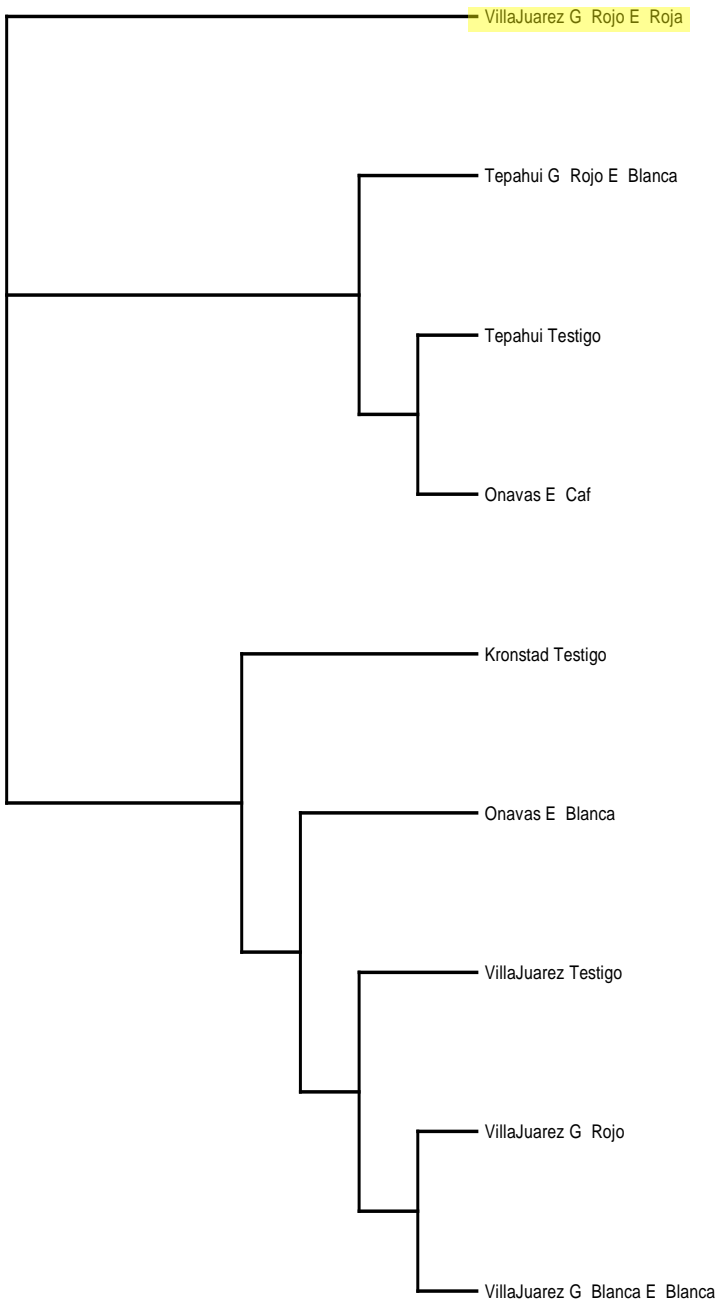


Figure 6. All Results By Line; Euclidean; UPGMA

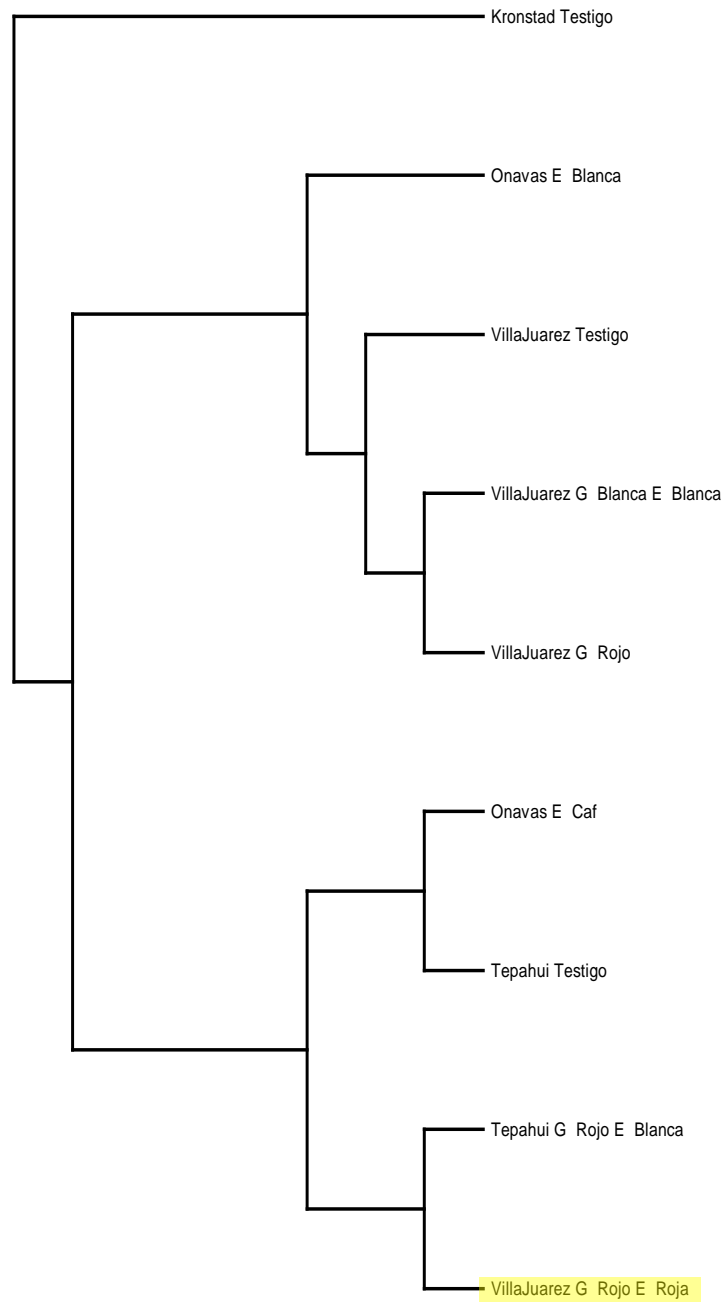


Figure 7. All Samples and Markers; Euclidean; NJ



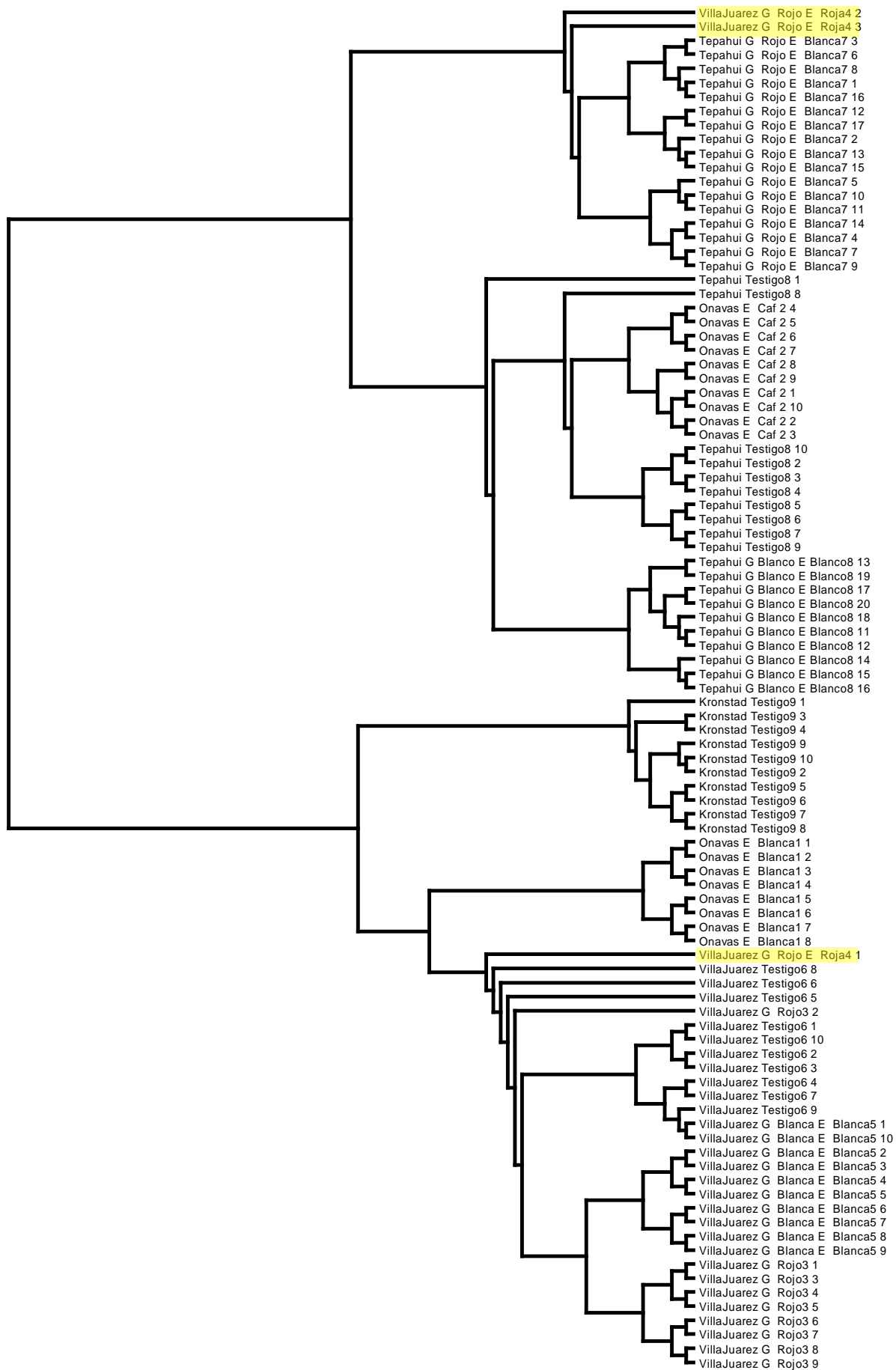
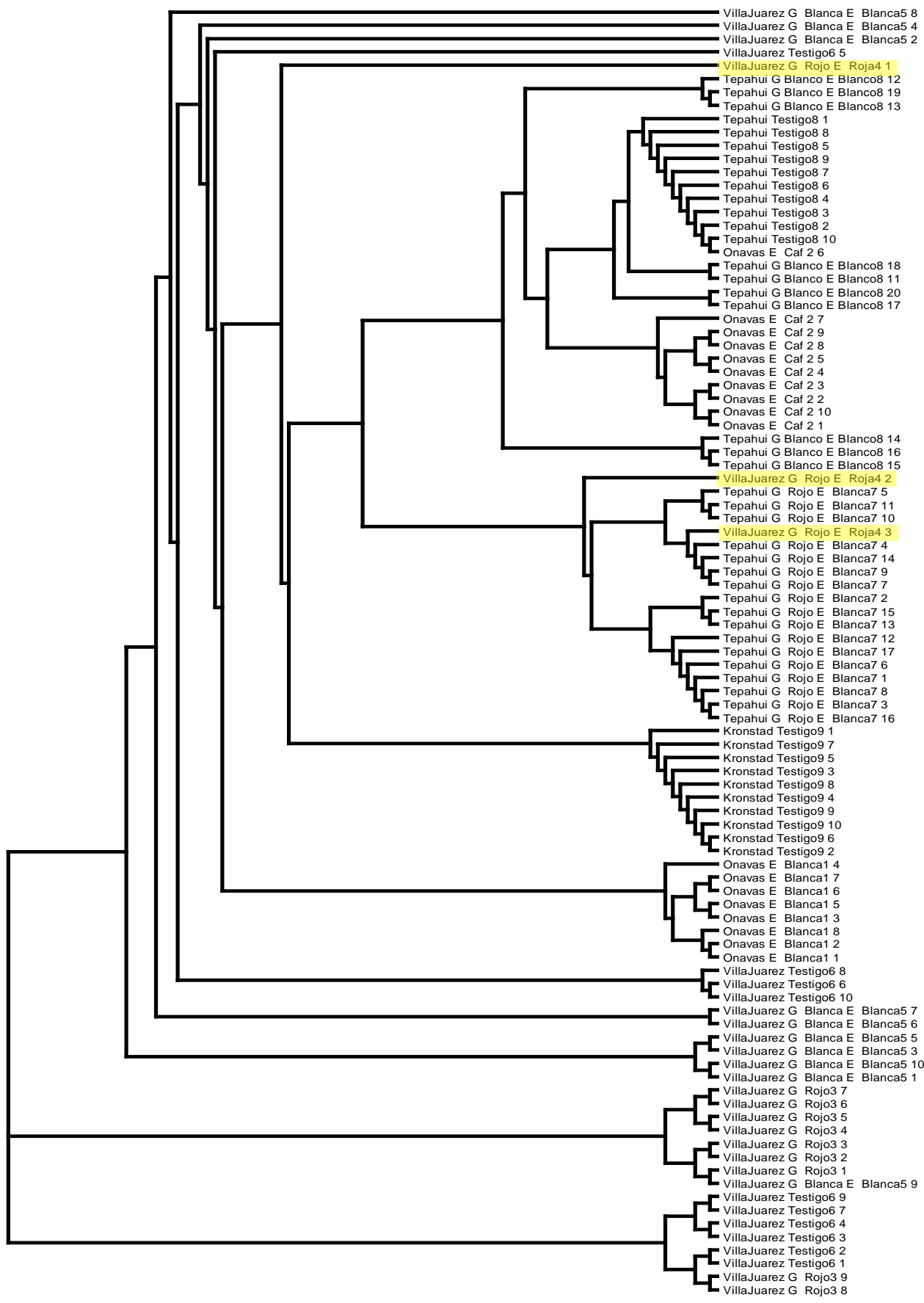


Figure 9. Only SSR Markers; Euclidean; NJ



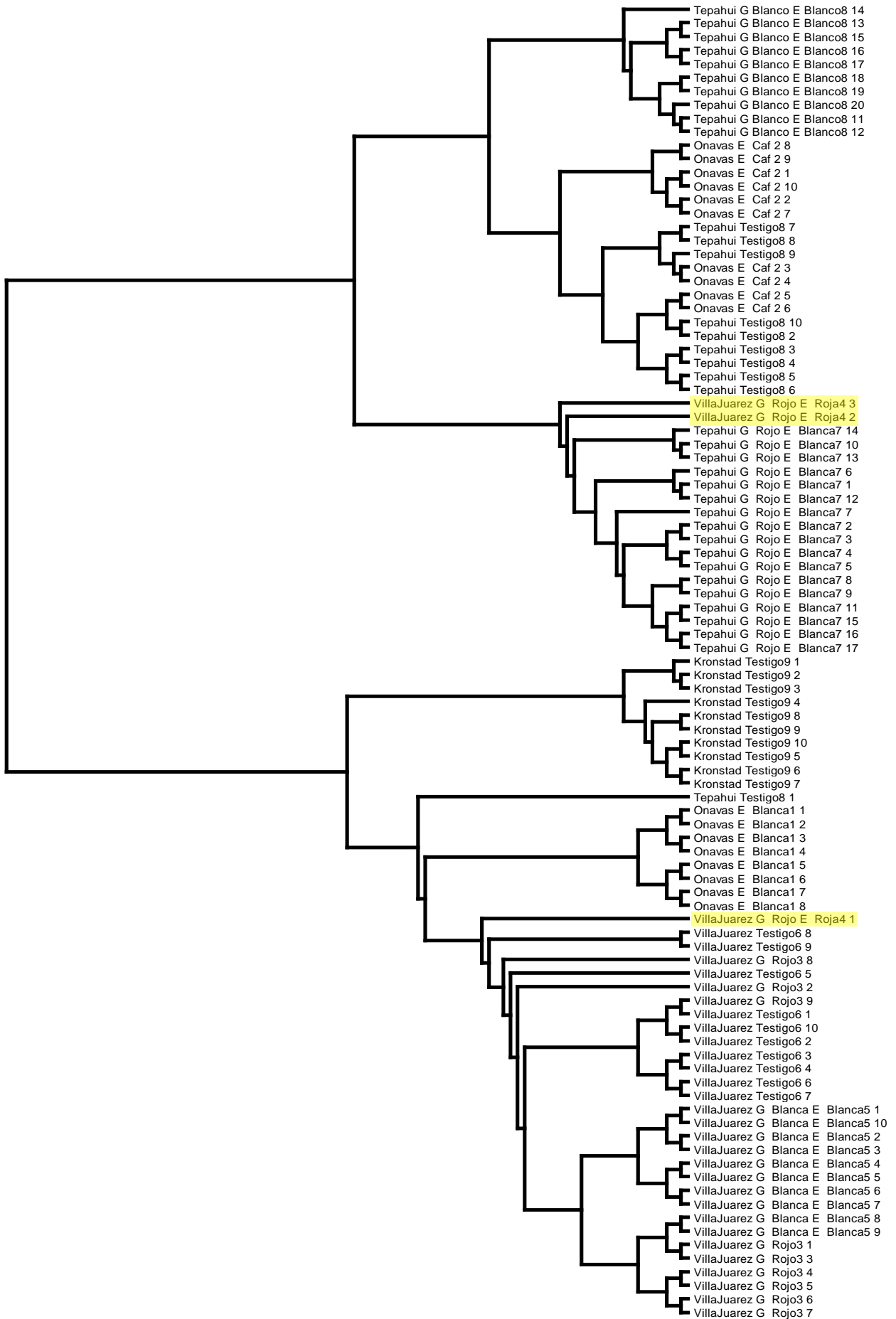
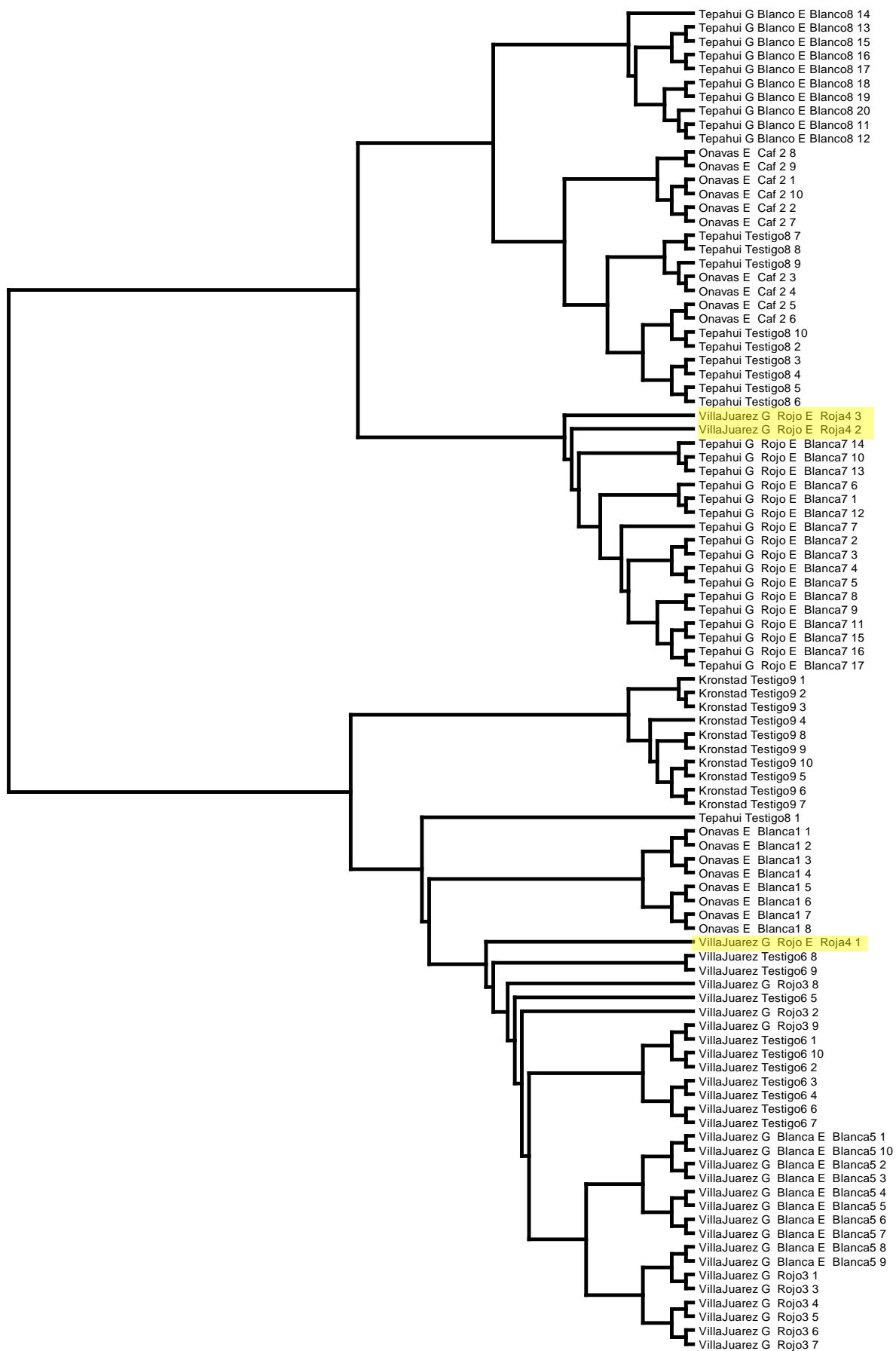




Figure 12. Only SNP Markers; Euclidean; UPGMA



Personal Reflection

In middle school, when studying an ancient civilization or a different group of people, we always completed what were known as PERSIA charts. The purpose of this chart was to fill out the respective areas (Political, Economic, Religion, Social, Intellectual, and Artistic) with characteristics of the studied group, nation, or people. Though I rather disliked completing the chart in middle school, ironically, I believe that the PERSIA chart is a perfect means to show just how incredible my internship was.

Political: I experienced the Mexican presidential election, which occurred during my first month in Mexico. Political signs were everywhere. Protests were frequent in Mexico City, as students in particular questioned the impartiality of the media. I was also asked difficult questions: why is it easy for Americans to come to Mexico but difficult for Mexicans to go to the US? Politics between Mexico and the United States is often topic to approach with trepidation, but what I discovered was just how passionate the Mexican people were about their nation. Politics was a common discussion during coffee breaks, and people scrupulously decided for whom to vote. Lines at the polling station in the town in which I lived stretched back for at least 100 people. Their passion was tremendous.

Economic: As I was going through Mexico City on the way to Toluca, one of the workers at CIMMYT pointed out a stark dichotomy. To the left of the highway, there were modern apartments which reached dozens of stories into the sky. To the right, small dilapidated houses clung to the side of the mountain. The highway split the two communities in half. It was a kind of poverty to which I was not accustomed. It was simply eye-opening to see such extremes within a few hundred feet of each other.

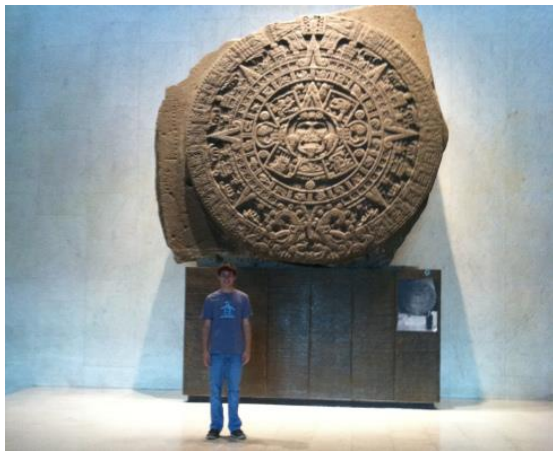
Religion: The rather monolithic Catholic religion in Mexico serves as a contrast to the more diverse views in the US. Many religious people would stop and acknowledge the cathedral in the town square whenever they passed it, something I don't see people do whenever they drive by their churches in the US. I also experienced the beauty of the Catholic churches, including the national cathedral in Mexico City.

Social: Culturally, I experienced many new things. From the obligatory topic of food to new and continuing friendships, I was truly transformed socially. I ate chinicuiles, which are Maguety worms. I had the wonderful opportunity to explore Mexico with my friend Miguel, who actually plans to visit me in the US sometime soon! I also became friends with many in the lab. Everyone there was incredibly friendly; all people in the lab greet each other with "buenos días" every morning, something which made me feel welcomed, especially on my first day.

Intellectual: I went into the internship with a vague idea of what career I wanted to pursue. I knew I was interested in the sciences, but I wasn't really sure where to go from there. Because of the internship, I now know that plant pathology is what I want to pursue. The internship was unparalleled in its ability to grant me this gift: to know what I want to do for the rest of my life.

Artistic: My friend Miguel and I had a mutual interest in different areas of art, so we had the fantastic opportunity to explore a lot of art in Mexico. We went to see a symphony perform at the National University in Mexico City. We also went to the contemporary art museum in the historic city center and to the national archaeology museum, which is unrivaled in its Aztec and Mayan artifacts. We explored the Teotihuacan pyramid, an enormous pyramid which I was actually able to climb! I was told that if you touch the top of the pyramid and look into the sun, you receive the blessing of the sun god.

Unfortunately, I did not have the opportunity to receive the blessing of the sun god, as I only discovered this ritual the day after I visited the pyramid. However, I don't think this will be a problem, as I know I must return soon; this is not only to receive the blessing of the sun god, but to be reunited with the friends and culture that have left an indelible impression upon my life.



References

- CIA Factbook: World [Internet]. 2011 [cited 2012 Sep 10]. Available from:
<https://www.cia.gov/library/publications/the-world-factbook/geos/xx.html>
- CIMMYT: About us [Internet]. [cited 2012 Sep 14]. Available from:
<http://www.cimmyt.org/about-us>
- CIMMYT [Centro Internacional de Mejoramiento de Maíz y Trigo]. 2005. Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory. Third Edition. Mexico, D.F.: CIMMYT.
- Dixon J, Braun HJ, Kosina P, Crouch J. 2009. Wheat Facts and Futures 2009. Mexico, D.F.: CIMMYT.
- Himi E, Maekawa M, Miura H, and Noda K. 2011. Development of PCR markers for *Tamyb10* related to R-1, red grain color gene in wheat. *Theor Appl Genet*, 122(8): 1561-76.
- Pasikatan MC, Dowell FE. 2003. Evaluation of a high-speed color sorter for segregation of red and white wheat. *Appl Eng Agric*, 19(1): 33-38.
- Robinson AD, Feil EJ, Falush D. 2010. Bacterial population genetics in infectious diseases. New York: J Wiley. 582 p.
- Robinson P, Holme J. 2011. KASP version 4.0 SNP genotyping manual. Herts, England: KBioscience.
- Sherman JD, Souza E, See D, Talbert LE. 2008. Microsatellite markers for kernel color genes in wheat. *Crop Sci*, 48: 1419-24.
- Source Bioscience: Taqman SNP genotyping [Internet]. [cited 2012 Sep 14]. Available from:
<http://www.lifesciences.sourcebioscience.com/genomic-services/faq/taqman-snp-genotyping.aspx>
- Takezaki N, Nei M. 2008. Empirical tests of the reliability of phylogenetic trees constructed with microsatellite DNA. *Genetics*, 178(1): 385-92.
- USDA ERS: Wheat Data [Internet]. [cited 2012 Sep 02]. Available from:
<http://www.ers.usda.gov/data-products/wheat-data.aspx>
- William HM, Trethowan R, Crosby-Galvan EM. 2007. Wheat breeding assisted by markers: CIMMYT's experience. *Euphytica*, (157)3: 307-319.