Spring 2019

Detection of Heteroplasmic Single Nucleotide Polymorphisms Using Melt Curve Analysis and Dual Labeled Fluorescent Probes

Emily Jezewski
University of Nebraska - Lincoln

Follow this and additional works at: https://digitalcommons.unl.edu/honorstheses

Part of the Biology Commons

Detection of Heteroplasmic Single Nucleotide Polymorphisms Using Melt Curve Analysis and Dual Labeled Fluorescent Probes

An Undergraduate Honors Thesis
Submitted in Partial fulfillment of
University Honors Program Requirements
University of Nebraska–Lincoln

Emily Jezewski
Biological Sciences
College of Arts and Sciences

March 8, 2019

Completed under the supervision of Alan Christensen, Ph. D. in the Department of Biological Sciences

Abstract:

Plant mitochondrial genomes are strange – they are unusually large, consist of huge amounts of non-coding DNA, and contain of several overlapping regions throughout the genome. The genome is made of several different sized linear and circular molecules and different mitochondria within a cell will have different pieces of the genome. Even if multiple mitochondria contain the same region of the genome, these sequences can differ by single nucleotide polymorphisms (SNPS). This is known as heteroplasmy. Heteroplasmy has been documented in *Arabidopsis thaliana* using whole-genome sequencing data. While heteroplasmy is well-documented, its existence in the mitochondrial genome is unexpected and it can be hard to quantify the degree to which heteroplasmy exists. This project sought to illustrate the existence of heteroplasmy in *A. thaliana* using melt-curve analysis and probe detection. Melt curve analysis takes advantage of the fact that DNA denatures at different temperatures depending upon the sequence; while probe detection quantifies different DNA sequences by the degree of fluorescence. The goal of this project was to show that both methods were successful in illustrating heteroplasmy in mitochondria and demonstrate heteroplasmy in DNA where one single base is different in a population. Both methods proved to be successful in illustrating heteroplasmy; however, there were drawbacks to each. Both methods required accurate knowledge of the presence of single nucleotide polymorphisms in the DNA sequences and melt curve analysis had a relatively high limit of detection.

Key Words: Heteroplasmy, melt-curve analysis, SNP detection, genetics, fluorescent probes, mitochondria, biological sciences

Introduction:

Plant mitochondrial genomes are unusually large, ranging in size from 200 kilobases (kb) to 10 Megabases (Mb). This large size is the result of the accumulation of "junk" DNA, or non-coding DNA. This point is best illustrated when comparing the mitochondrial genomes of *Arabidopsis thaliana* and a human's mitochondrial genome. A human's mitochondrial genome codes for 37 genes and is made up of 18kb; while *A. thaliana* has a mitochondria genome of 367kb and only codes for 58 genes.

In addition to their large size, plant mitochondrial genomes may also be mapped as a circle, which leads one to assume plant mitochondria possess a single DNA molecule. However, pulse field gel electrophoresis (PFGE) has shown that the genome is not actually circular (Klein et al 1994). Rather, the genome is made up of several different sized linear and circular molecules and different mitochondria within an individual cell will have different amounts of these genomic fragments. Additionally, many mitochondria do not have any DNA at all (Wang et al 2010). When DNA is present the different molecules have repeated and overlapping regions. The overlaps may be aligned when mapping the genome, which is likely why the genome may be mapped as a circle.

The repeated regions will often recombine with one another (Mackenzie 2007). In fact, these recombination events are typical within plant mitochondria and have become a recognizable characteristic of plants (Kmiec et al 2006). A consequence of these recombination events is that the mitochondrial DNA within a single mitochondrion and throughout all the mitochondria of a single plant are different (Barr et al 2005). This presence of multiple different genomes in a single cell or organism is known as heteroplasmy and has been found in *A. thaliana.* In *A. thaliana* there are two large repeated regions, and several other repeats of unusual

3

size (ROUS). The two large repeats recombine regularly throughout the mitochondria. The ROUS sequences may also recombine throughout the mitochondria and these recombination events produce heteroplasmic cells and plants (Wynn and Christensen 2019)

While ROUS recombination has been observed, these recombination events are rare and do not occur in every cell or every mitochondria. As such, some mitochondria may have two repeats that have been recombined, while others have no presence of recombination or the recombination has reverted. This is a type of "structural heteroplasmy" where mitochondria throughout a single plant and within a single cell have differently structured genomes due to recombination.

Evidence illustrates that this structural heteroplasmy exists, but less is known about the existence of heteroplasmic single nucleotide polymorphisms (SNPs). The presence of heteroplasmic SNPs would not be the expected pattern. Mitochondria are maternally inherited, but there is limited knowledge of how many mitochondria are passed on to offspring. This pattern of inheritance could be considered a type of bottleneck. The degree to which a bottleneck may exist contributes to the amount of heteroplasmy that exists (Roze et al 2005). For example, if maternal inheritance is tightly regulated and very few mitochondria are passed on, the likelihood of heteroplasmy may be diminished. However, data illustrates that the mitochondrial genomes do vary and are heteroplasmic. In fact, throughout the mitochondrial DNA of *A. thaliana*, there is evidence of DNA molecules SNPs at consistent loci (Wynn et al, in preparation).

There are two methods by which heteroplasmic SNPs may have arisen in plant mitochondria, despite the maternal bottleneck. The first method is paternal leakage. Paternal leakage occurs during fertilization when some paternal mitochondria DNA is taken up by the

maternal plant and reaches the offspring during seed production (Kmiec et al 2006). This paternal mitochondrial DNA is then apparent in the offspring plant. The second method through which heteroplasmy may arise is random mutation, which is a possible during both the DNA repair and replication processes (Kmiec et al 2006). These random mutations may be the source of heteroplasmy, but then they may also be passed onto offspring, depending on the degree of a bottleneck present during reproduction.

While both of these methods may give rise to heteroplasmy, detecting it and quantifying the degree to which heteroplasmy exists in plants is challenging. The most accurate method of illustrating heteroplasmy is currently considered to be whole genome sequencing (WGS). A major drawback to this method is cost – WGS is very expensive and requires a significant read depth in order to correctly identify a heteroplasmic SNP.

Another method of detection is melt curve analysis (Temesvári et al 2011). Melt curve analysis is reliant upon the fact that DNA denatures at slightly different temperatures, depending on its base sequences. Guanine and Cytosine pairs (G/C pairs) tend to have a slightly higher melting temperature than Adenine and Thymine pairs (A/T pairs) due to the number of hydrogen bonds; G/C pairs have three hydrogen bonds, while A/T pairs have two. Melt curve analysis is very accurate and can detect variations as minute as a single base pair.

One final method used to detect heteroplasmy utilizes probes (Lyon 2011). Probes anneal to specific DNA sequences and possess a fluorophore and a quencher. The quencher inhibits fluorescence until it is removed during DNA replication by DNA polymerase. The degree of fluorescence quantifies the amount of the specific DNA sequence within the sample.

Ultimately, the goal of this project was to illustrate and quantify heteroplasmy in *Arabidopsis thaliana* using melt curve analysis and probes. *A. thaliana* was chosen because it is

an ideal model organism – it has a short life cycle, it is easy to obtain DNA from, and our lab has previously shown the existence of heteroplasmy in *A. thaliana* mitochondria using NextGen Sequencing. While accurate, NextGen Sequencing is incredibly expensive. Thus, we sought to validate our bioinformatics findings using melt curve analysis and probe detection. We hypothesized that both methods would be accurate means of detection with the hope that they could detect SNPs occurring in *A. thaliana* that occurred with a frequency of 5% in the population.

Methods:

Experimentation was completed under the guidance of Dr. Alan Christensen at the University of Nebraska-Lincoln. *Arabidopsis thaliana* seeds were obtained from Lehle Seed Company (Round Rock, Texas, USA). All seeds were planted on soil and grown in a walk-in growth chamber with 16 hours light, 8 hours dark cycle days at 22°C. DNA was extracted using CTAB DNA extraction from mature plants after they began to flower (Allen et al 2006). Mitochondria DNA was purified using Strehle's prep (Strehle et al 2018).

Melt Curve Using Quantitative Polymerase Chain Reaction

Qualitative polymerase chain reaction (qPCR) was performed using a 96-well plate with a reaction volume of 20μl/well. Samples were run using the Bio-Rad CFX96 Real-Time System. qPCR amplifies DNA similarly to polymerase chain reaction (PCR); however, after each replication cycle, the amount of DNA in the reaction well is recorded using fluorescence. Each reaction well contained 10μl Sybr Green master mix (Bio-Rad), 4μls forward and reverse primers (2.5mM), and 2μl template DNA (0.5ng/ml). Plates were spun at 4,000rpm and then placed in the Real-Time System. The CFX96 Real-Time System program performed the

following cycling protocol: ten minute denaturing step at 95°Celsius (C), 45 cycles of 10s at 95°C, 15s at 60°C, and 16s at 72°C.

Sybr Green Master Mix fluoresces at a wavelength of 497nm and binds to DNA and stains it, effectively quantifying the DNA. Fluorescence was measured using the Sybr channel in the Real-Time system. qPCR also provides a Cq mean, which is the point at which the fluorescence curve enters the exponential phase. This point was calculated using multi-variate regression and using the average of three technical replicates. qPCR also recorded a melting temperature, which was the temperature at which the DNA denatured and fluorescence recordings dropped. This temperature was dependent on the bases in a particular DNA sequence. qPCR is sensitive enough to detect the melting temperature between two identical DNA sequences with a single base difference.

Primers were designed by Emily Wynn, graduate student in Dr. Christensen's lab, and obtained from Eurofins Genomics. Primers were diluted with double distilled water. Primers with greater G/C content than A/T were preferred and annealing temperature of all primers were between 50°C and 65°C. The efficiency of all primers used in experimentation was determined using a standard curve. The standard curve used a single sample of Col-0 DNA that was serially diluted using 1:10 dilutions. With each dilution it was expected that the Cq mean would increase by 3.2 because with 1:10 dilutions the DNA needs to be doubled 3.2 times to reach the same concentration as the previous dilution. An increase of 3.2 indicated an efficiency of 100%. Col-0 DNA was chosen as a positive control because it is the wild-type DNA.

A. thaliana DNA was amplified using PCR and primers flanking Repeat A, with an annealing temperature of 60°C and an extension time of 16 seconds. Amplicons obtained were A11 (sequence between the forward A1 primer and reverse A1 primer) and A22 (sequence

between A2 forward and reverse primers). All primer details can be found in Table 1 in the

supplemental data. A11 and A22 amplicons have a single nucleotide difference at position 509.

This is a cytosine (A11) to thymine (A22) change and results in a different melting temperature

between A11 and A22. DNA was cleaned up using a GeneJET PCR Purification Kit (Thermo

Scientific). Following purification, DNA was diluted to a concentration of 0.5ng/μl.

Concentration was recorded using a QuBit Fluorometer (Thermo Fisher).

Following dilutions, A11 and A22 samples were aliquoted into the following ratios of

A11:A22 DNA: 50:50, 70:30, 30:70, 80:20, 90:10, 95:5, all at 0.5ng/μl. These DNA samples

were run in qPCR using the A Internal 2 primers. The A Internal 2 primers amplified both the

A11 and A22 equally because the primer sequence was internal to the repeated sequences of A11

and A22. A Internal 2 primers amplified the specific region in A11 and A22 amplicons that

differed by a single base (C→T change). The entire region amplified was 75 bases in length

(including primers). A small amplicon region was required in order to increases the effectiveness

of the Real-Time system to detect a SNP in a DNA sequence. qPCR was performed with an

extension time of 16 seconds and an annealing temperature of 60°C. These experiments were

performed using Columbia-0 DNA.

Melt Curve Analysis Using Probe

Probe SNP detection was performed using qPCR with a 96-well plate (20μl reaction

volume). Experiments were performed using mitochondrial Col-0 DNA (.5ng/μl). There is a

frequently occurring SNP in the *A. thaliana* mitochondrial genome at position 70194, identified

by WGS (Wynn et al, in preparation). Two probes were designed to anneal to the wildtype SNP

(Guanine) and the mutant SNP (Adenine). Each reaction well contained 10μl SsoAdvanced

Universal Probes Supermix (Bio-Rad), 2μls forward and reverse Ung176 primer (2.5μM), 2μls

wild type probe and mutant probe (1.25μM), and 2μls template DNA. Thermocycler performed

the following cycling protocol: ten minutes at 95℃ for denaturing followed by 45 cycles of 95℃

for 10s, 60℃ for 15s, and 72℃ for 5s. Primers and probes were designed by Emily Wynn and

obtained from Integrated DNA Technologies. The primers were diluted with double distilled

water.

The probes were diluted with TE, stored at -20 degrees Celsius, and kept away from light

in order to prevent degradation. The wild type probe fluoresced at a wavelength of 518nm and

was read using the Hex channel on the Real-Time system. The mutant probe fluoresced at a

wavelength of 556nm and was read using the Fam channel on the Real-Time system. The

SsoAdvanced Universal Probes Supermix used did not fluoresce. This prevented inaccurate data

outputs by the Real-Time since the wild type mutant fluoresces in the same channel as Sybr

Green Master Mix. Four technical replicates were used and the Cq means of the individual probe

fluorophores were obtained by the cycle number in which fluorescence exceeded a set threshold.

Ratios of Cq means obtained from qPCR between wildtype probe and mutant probe allowed us

to determine the ratio of the wildtype SNP to mutant SNP at position 70194 of the *A. thaliana*

mitochondria genome.

Probes may be used in qPCR in varying ratios, when compared to the concentration of

primers. They may be used in any ratio ranging from 1:1 ratio of primer concentration to probe

concentration to 1:4 primer concentration to probe concentration. To determine the most efficient

ratio of primer to probe, we ran an assay with varying ratios of primer to probe from 1:1 to 1:4

using Col-0 DNA. Upon completion of the assay, the 1:1 ratio of primer:probe and 1:2 ratio of

primer:probe results were similar; while the 1:3 ratio and 1:4 ratio were not accurate. Probes are

also costly; thus, in order to preserve probes, the 1:2 ratio was chosen and used for following

experiments. Upon completion of experimentation, Cq means of probe fluorescence were obtained and analyzed. To determine the fold different between the mutant SNP and wildtype SNP in a sample the following equation was used: $2^{-(Cqmean\ FAM - Cqmean\ Hex)}$. This provided the fold ratio of mutant SNP to wildtype SNP.

Results:

Melt Curve Analysis

All primer efficiencies obtained were between 1.24 and 2.7. Efficiency above 1 could indicate the presence of PCR inhibitors; since we are not using these primers to quantify DNA, they should not interfere with the detection of our melt peaks (Table 1).

A11 had an average melting temperature of 75.8°C, A22 had an average melting temperature of 75.27°C. When A11 and A22 were in wells alone they exhibited a single melting temperature, indicating that all amplicons in the reactions had identical sequences. (Figures 1 and 2).

When A11 and A22 amplicons were mixed in varying ratios, Real-Time PCR detected two melting temperatures per reaction well, indicating the presence of two different amplicons that differed by a single base. Two melting temperatures were obtained for the 1:1 mix, the 70:30 mix of A11:A22, and the 30:70 mix of A11:A22. On the melt curves for the mixed samples, there is a dip in the melt peak that illustrates the reading of two different temperatures found in one, single sample of DNA (Figures 3 and 4).

A single melting temperature was obtained for 80:20, 90:10, and 95:5 mixes of A11 an A22, indicating that the Real-Time system could not detect an amplicon in a proportion of less than 30% (Figure 5).

Probe Detection of SNPS

The ratio of primer to probe was determined running an assay with the following ratios of primer to probe: 1:1, 1:2, 1:3, and 1:4. The 1:1 and 1:2 ratios produced accurate results; while the 1:3 and 1:4 ratios produced inaccurate results (Figures 6 and 7).

To determine if the SNP identified by the probe was localized in the mitochondrial DNA or the nuclear DNA, experiment was performed using mitochondrial Col-0 DNA. The SNP appeared in the mitochondrial DNA. The SNP occurred in 6.49% of the mitochondrial genomes. Vardict previously predicted around 10% of the genome to possess this SNP (Figure 8).

Discussion:

Overall, it is evident from sequencing data that *A. thaliana* have heteroplasmic mitochondria. This is clear and easy to identify when using whole-genome sequence data. However, experimentation has illustrated that other methods to detecting heteroplasmy, such as melt curve analysis and the use of probes, can detect SNPS, but that these methods require extensive knowledge of DNA sequences. Regardless, these methods are quick and inexpensive procedures that may be used to test samples for the presence or absence of various mitochondrial and nuclear SNPs.

With regards to identifying heteroplasmy and SNPs using melt curve analysis, we found the limit of detection to be 30%. In other words, Real-Time PCR had the abilities to detect a base change in a specific DNA sequence that occurred in 30% of the DNA within the sample. Due to this limit of detection, melt curve analysis would not have been able to detect the SNP at position 70194 that we used probe to measure. In addition to the SNP at position 70194, we know from our NextGen sequencing data that there are SNPs in lines of Col-0 that have SNPs occurring at a rate of 5%. These SNPs will not be able to be identified or illustrated using melt curve analysis. As such 30% as the limit of detection is a drawback to melt curve analysis.

Another limit to melt curve analysis is that it can only be used to identify a single SNP among a small sequence of DNA where users know the exact position of the SNP. This is because Real-Time PCR requires primers that result in the amplification of a specific sequence. This sequence needs to be small and contain a single base change between two identical DNA sequences for the results to be the most accurate. If there are multiple SNPs in the amplified sequence, the technique will not be as accurate and difficult to interpret. Therefore, another drawback of melt curve analysis is the requirement of accurate information regarding the DNA sequences and potential changes to those sequences.

Knowledge of the specific sequence is also a limitation when using probes. Probes rely on primers to amplify a small, specific sequence of DNA. Probes must also be specifically designed to anneal to an exact sequence of DNA. In order to use probes to identify heteroplasmy, two probes must be created to anneal to the wildtype and mutant DNA sequences. Despite the inherent drawbacks of the probes, there are several aspects of this method that ultimately make them more efficient than melt curve analysis. The probes have the ability to detect SNPs in a wide range of proportions and can identify SNPS present in small samples. Additionally, probes are useful for illustrating the presence of specific mutations located in mitochondrial genome.

Going forward, these experimental methods have applications in DNA repair. It may be helpful to investigate the outcomes of DNA repair when inducing DNA damage in *A. thaliana* using Cipro, or other adverse agents. We can also use melt curve analysis to try and understand what happens to the internal SNP within repeats A11 and A22 during recombination. For example, following recombination if A11 possesses the SNP, but A22 does not, what would happen when A12 and A21 recombine? This could provide insight into what is happening to a DNA strand during ROUS recombination. Additionally, probe analysis could be used to study

the presence of SNPs over the course of generations to help us understand the extent of a

bottleneck present during plant reproduction.

References:

Allen, G.C., FloresVergara, M.A., Krasynanski, S., Kumar, S. and Thompson, W.F. (2006) A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethyl-ammonium bromide. Nature protocols, 1: 2320-2325.

Barr, C. M., Neiman, M., Taylor, D. R., (2005). Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist.* 168: 39-50.

Klein M., et al, 1994. Physical Mapping of mitochondrial genome of *Arabidopsis thaliana* by cosmid and YAC clones. *The Plant Journal.* 6(3): 447-455.

Kmiec, B., Woloszynksa, M., Janska H., (2006). Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Current Genetics.*

Lyon, E., (2011). Mutation detection using fluorescent hybridization probes and melting curve analysis. *Expert Review of Molecular Diagnostics.* 1(1): 92-100.

Mackenzie, S. A., (2007). The unique biology of mitochondrial genome instability in plants, pp. 36–46 in *Plant Mitochondria*, edited by D. Logan. Blackwell Publishing, Oxford.

Roze, D., Rousset, F. J., Michalakis, Y. (2005). Germline Bottlenecks, Biparental Inheritance and Selection on Mitochondrial Variants: A Two-Level Selection Model. *Genetics.* 170: 1385-1399.

Strehle, M., Purfeerst, E., Christensen, A.C., (2018). A rapid and efficient method for enriching mitochondrial DNA from plants. *Mitochondrial DNA Part B.* 3(1): 239-242.

Temesvári, M., Pualik, J., Kóbori, L, Monostory, K. (2011). High-resolution melting curve analysis to establish CYP2C19*2 single nucleotide polymorphism: Comparison with hydrolysis SNP analysis. *Molecular and Cellular Probes.* 25: 130-133.

Wang DY, Zhang Q, Liu Y, et al. The levels of male gametic mitochondrial DNA are highly regulated in angiosperms with regard to mitochondrial inheritance. *Plant Cell.* 2010;22(7):2402-16.

Wynn, E. L., Purfeerst, E., Christensen, A. C. DNA repair in *Arabidopsis thaliana* uracil DNA *N*-glycosylase mutant. In Preparation.

Wynn, E. L., Christensen, A. C., (2019). Repeats of Unusual Size in Plant Mitochondrial Genomes: Identification, Incidence and Evolution. *Genes, Genomes, Genetics.* 9(2): 549-559.

Supplemental Data

Table 1. Primer and probe pairs used for PCR and qPCR, along with the annealing temperature (T_m) in degrees Celsius, and amplicon size (including primer length) in base pairs.

Primer Name	Sequence (5'-3')	T_m	Amplicon Size	Efficiency	R^2
A1 Forward Primer	CATGTACGTAGGGCCTTCCA	62.4	625	N/A	N/A
A1 Reverse Primer	ATTCGGTGACTTTCGCGGT	60.2	625	N/A	N/A
A2 Forward Primer	GATGCCCGTTTACGATGCAA	60.4	602	N/A	N/A
A2 Reverse Primer	AAACACCCTAACAGTGCCTC	60.4	602	N/A	N/A
A Internal 2 Forward Primer	TGGCACGAAAAGGAAATCCG	60.4	75	2.7	0.99
A Internal 2 Reverse Primer	CCCTCACTGAACCGACTTGA	62.4	75	2.7	0.99
Ung 176 Forward Primer	TATCCCGAAGATACAAGCGCC	62.6	81	1.24	0.958
Ung 176 Reverse Primer	CCACTCCTTCTGTGAGGCAA	62.4	81	1.24	0.958

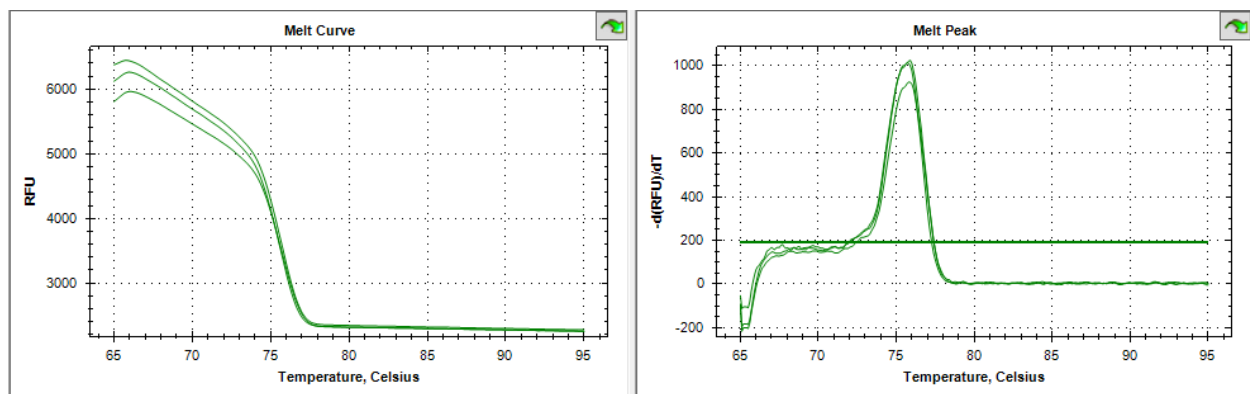| Wildtype Probe	5HEX/CCGGTCCTC/ZEN/ TTCTCTTGTGTCGAAGTGT/31ABkFQ	63.0	N/A	N/A	N/A
Mutant Probe	56-FAM/CCGGTCCTC/ZEN/ TTCTCTTGTGTCAAAGTGT/31ABkFQ	61.7	N/A	N/A	N/A

Figure 1: A11 Melt Curve Results. Ilustrates the melting temperature of A11 DNA sample (average melting temperature 75.8°C).
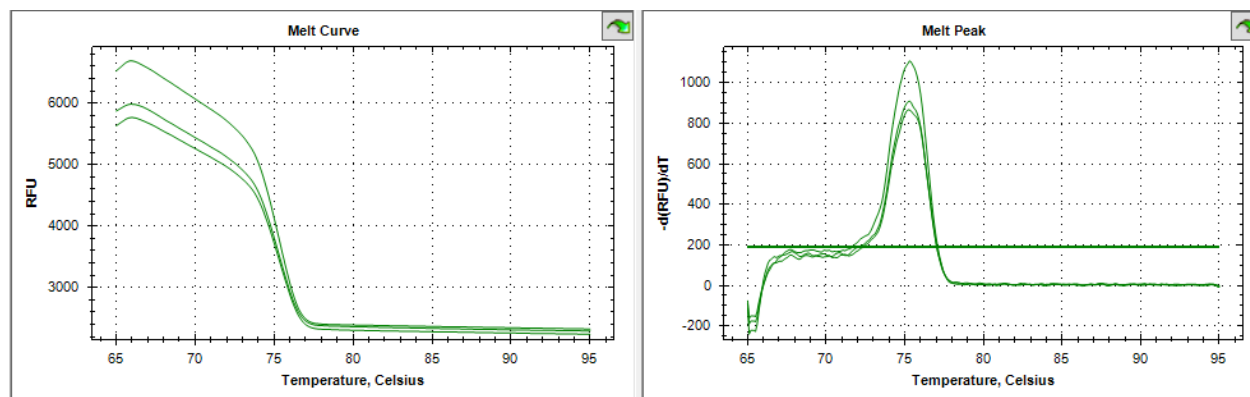

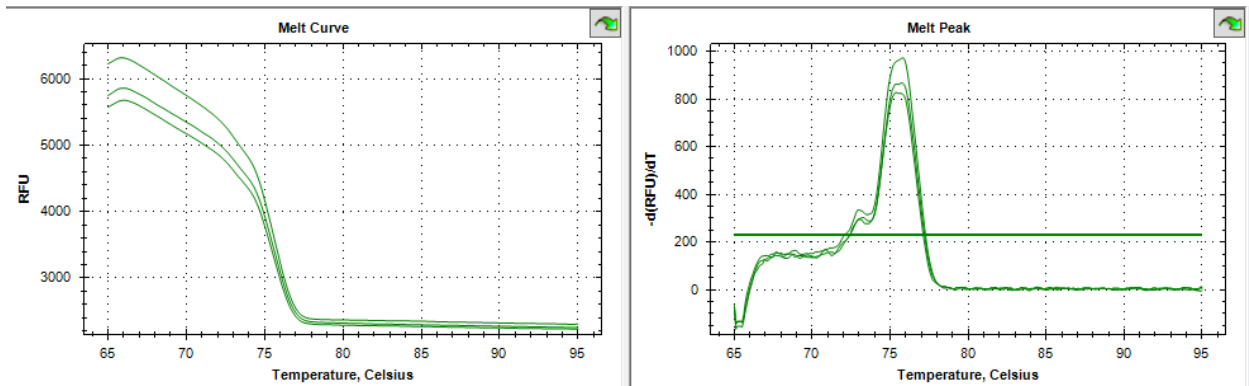Figure 2: A22 Melt Curve Results. Ilustrates the melting temperature of A22 DNA sample (average melting temperature 75.27°C).

Figure 3: 1 to 1 Mixture of A11 to A22 DNA. Illustrates the "dip" in the melt curve, which signals the presence of two different DNA sequences within a sample.
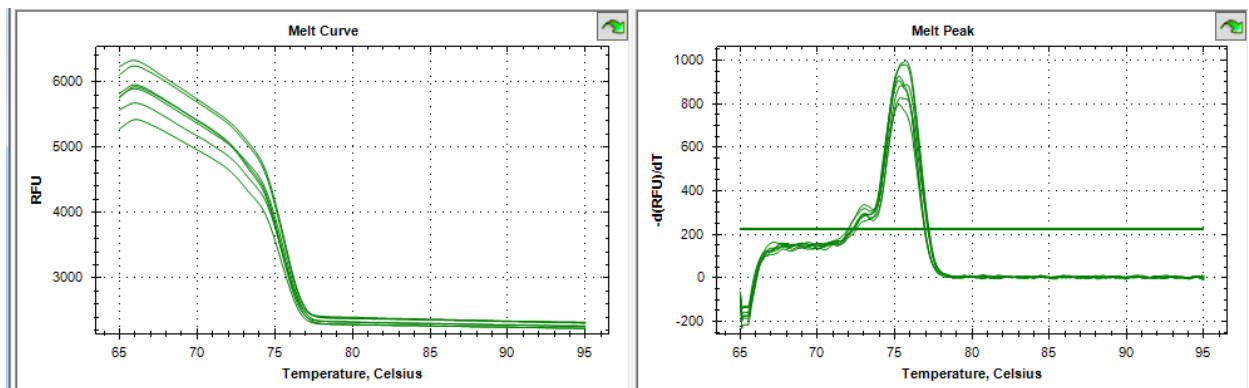
Figure 4: 70:30 and 30:70 Ratios of A11 to A22 Samples: Illustrates the ability of system to detect two different melting temperatures depending on the ratio of one DNA sequence, with one single base change from another sequence in a single sample
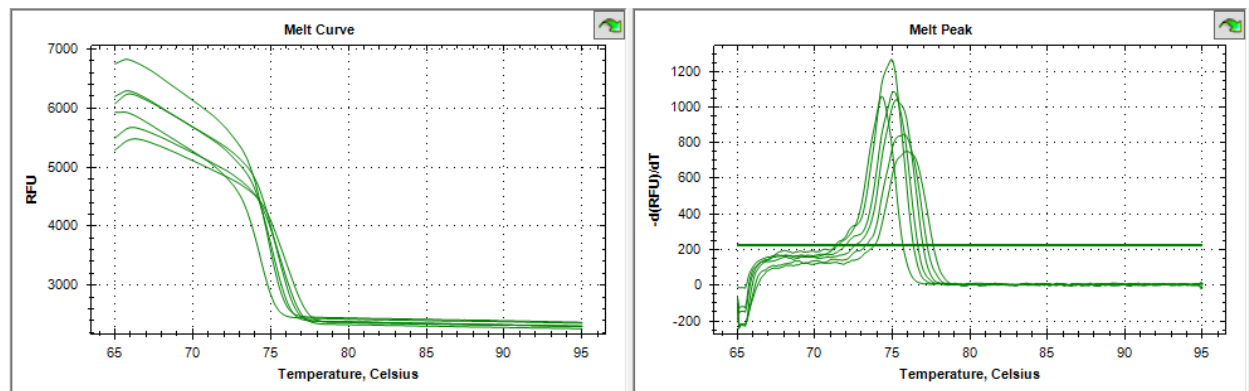

Figure 5: Limit of Detection for Real-Time qPCR. For ratios of A11:A22 above 70:30, only a single melting temperature was detected per sample.
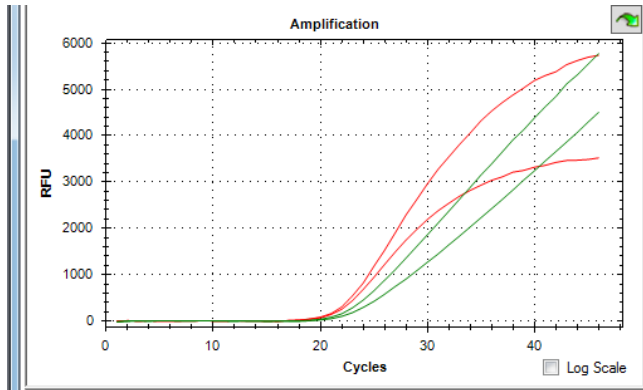
Figure 6: 1:1 and 1:2 Primer to Probe Comparison. Illustrates that the difference between the 1:1 ratio of primer to probe is not significantly different than the results of the 1:2 ratio.
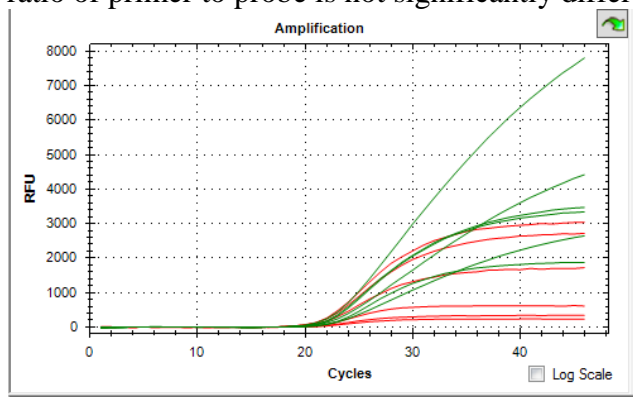

Figure 7: 1:3 and 1:4 Primer to Probe Comparison. Results are inaccurate and significantly different than those obtained from the 1:1 and 1:2 ratios.
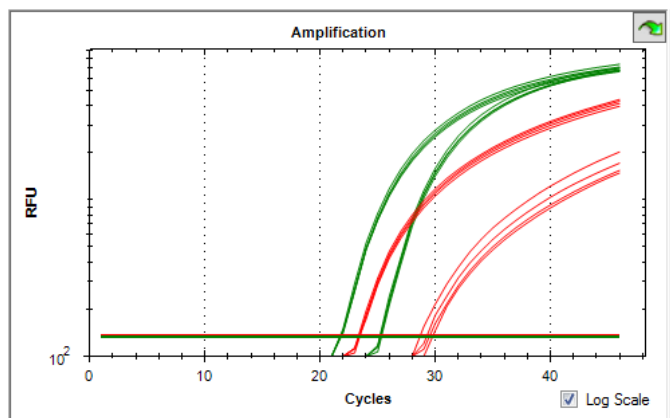

Figure 8: Illustrates the melt curve of nuclear DNA and mitochondria DNA, showing that the SNP is located within the *A. thaliana* mitochondria genome.