ASSESSING THE GENETIC DIVERSITY OF TEN CAMELLIA TAXA

by

Joshua H.L. Darfler

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Public Horticulture

Summer 2014

© 2014 Joshua H.L. Darfler
All Rights Reserved

UMI Number: 1567796

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI
Dissertation Publishing

UMI 1567796

Published by ProQuest LLC (2014). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code

ProQuest

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

ASSESSING THE GENETIC DIVERSITY OF

TEN CAMELLIA TAXA

by

Joshua H.L. Darfler

Approved: _____
Robert E. Lyons, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____
Blake C. Meyers, Ph.D.
Chair of the Department of Plant and Soil Sciences

Approved: _____
Mark Rieger, Ph.D.
Dean of the College of Agriculture and Natural Resources

Approved: _____
James G. Richards, Ph.D.
Vice Provost for Graduate and Professional Education

ACKNOWLEDGMENTS

I would to thank the many people who helped in various ways to make this thesis possible. First and foremost, I would like to thank the Longwood Graduate Program, The University of Delaware, and Longwood Gardens for their support of my research over the past two years. Notably, thank you to Dr. Robert Lyons, Dr. Matthew Taylor, and Dr. Antoni Rafalski for serving on my thesis committee. I would also like to thank DuPont Pioneer for use of their lab space and equipment, expertise, and technical assistance, which made this project possible.

Furthermore I would like to thank Ada Chang for all her help in the lab and with data analysis. Also thank you to Shawn Thatcher, Stephane Deschamps, Elmer Heppard, and Randy Parker for all their help and assistance with the RNA analysis section of this project. Finally, thank you to Benjamin Rhoe at the University of Delaware, and Joesph Rothleutner at the Morton Arboretum for their help with flow cytometry, and data analysis. All these individuals, and more, really made it possible to accomplish during my time in the Longwood Graduate Program and in the lab.

I would also like to thank my classmates, friends, and family for all their support outside of my research. Having such a great support network helped me get through the rough and tough times.

TABLE OF CONTENTS

LIST OF TABLES vi
LIST OF FIGURES viii
ABSTRACT x

Chapter

1 INTRODUCTION 1

2 LITERATURE REVIEW 5

2.1 Introduction and Overview of the Genus *Camellia* and Their Cultivation 5
2.2 Camellias at Longwood Gardens & Breeding Efforts 7
2.3 Genetics of Camellias 10
2.4 Classification of Camellias & Species of Interest 13

3 MATERIALS AND METHODS 18

3.1 Plant Material 18
3.2 Species Selection 18
3.3 Methods and Materials for DNA Analysis 19

3.3.1 Tissue Collection 19
3.3.2 Gene Sequences and Primer Design 19
3.3.3 DNA Extraction 20
3.3.4 PCR Amplification 20
3.3.5 PCR purification 21
3.3.6 DNA sequencing 22
3.3.7 Phylogenetic Tree Creation - DNA 22

3.4 Materials and Methods for RNA Analysis 23

3.4.1 RNA Sample Collection 23
3.4.2 RNA Extraction Protocol 23
3.4.3 DNA Digestion 25
3.4.4 RNA Sequencing 25

3.5 Materials and Methods for Flow Cytometry 26

4 RESULTS 27

4.1 Experiment Design 27
4.2 Results of DNA Analysis 27

4.2.1 Quantity and Quality of Extracted DNA 27
4.2.2 Analysis of DNA Sequences 29

4.3 Results of RNA Analysis 31

4.3.1 Quantity and Quality of Extracted RNA 31
4.3.2 Analysis of RNA Sequences 32

4.4 Results of Flow Cytometry 33

5 DISCUSSION 58

5.1 DNA Analysis Discussion 58

5.1.1 Camellia DNA Extraction 58
5.1.2 DNA Phylogeny 60
5.1.3 Recommendations for Future Work 65

5.2 – RNA Analysis Discussion 67

5.2.1 RNA Extraction 67
5.2.2 RNA Phylogeny 67
5.2.3 Recommendations for Future Work 68

5.3 Flow Cytometry Discussion 69

5.3.1 Recommendations for Future Work 70

REFERENCES 72

Appendix

A FLOWERS OF CAMELLIA SPECIES AND CULTIVARS MOST USED IN THIS RESEARCH 74
B COMMONLY USED ABBREVIATIOSN FOR CAMELLIA SPECIES 77
C LIST OF ALL PRIMERS USED FOR DNA SEQUENCING 78
D DETAILED TABLES OF POLYMORPHISMS BETWEEN ACCESSIONS 80

LIST OF TABLES

Table 1: Selection of camellia taxa used in this experiment. Samples with an * next to them indicate accessions not housed at Longwood Gardens. A † next to an accession indicates a sample only used for analysis with flow cytometry. 36

Table 2: Standard Deviation (SD) of DNA concentrations extracted per accession based on different data sets. (A) All samples included [n=45, mean=29.55 ng/µL], (B) All samples but two outliers removed from *C. amplexicaulis* [n=43, mean=25.84 ng/µL], (C) All samples larger than 10 ng/µL [n=37, mean=34.45 ng/µL], (D) All samples larger than 10 ng/µL, but one outlier removed [n=36, mean=32.04 ng/µL]. Where there is only a dash (-) present, there was only one sample and therefore standard deviation could not be calculated. 38

Table 3: Standard deviation (SD) of A260/280 per accession based on different data sets. (A) All samples included [n=45, mean=1.70], (B) All samples except the one outlier removed from *C. transarisanensis* [n=44, mean=1.66], (C) All samples larger than 10 ng/µL [n=37, mean=1.65]. Where there is only a dash (-) present, there was only one sample and therefore standard deviation could not be calculated. ... 40

Table 4: Column two shows the average A260/280 per accession when N=45 (all samples except one outlier). Column three shows the average when N=37 (all samples where the concentration of DNA was greater than 10 ng/µL). An asterisk (*) next to the average indicates it is significantly less than 1.8, the expected ratio of pure DNA..... 41

Table 5: Concentration and absorbance ratio for each sample submitted for RNA sequencing. 43

Table 6: List of genes sequenced in all accessions. Column one contains the gene code from NCBi's GeneBank. Column two contains a short description of the gene, and column 3 specifies if the GeneBank sequence originated from mRNA or DNA..... 44

Table 7: Number of basepairs (bp) sequenced per accession per gene. The length (bp) of the final assembled contig for each gene can be found in the final column. 45

Table 8: Number of differences between each accession (bp). 48

Table 9: Number of base pairs (both from single and paired reads) sequenced through Next Generation RNA Sequencing. These reads were then assembled into a number of transcripts, which represent a large number of genes. 50

Table 10: Mean and standard deviation of detected emission for each accession through flow cytometry. Samples were treated with PI stain and observed in the 545-625nm light spectrum. 51

Table 11: Table showing the difference in mean, median, and standard deviation (SD) of each sample when passed through a cell separator prior to analysis (filtered), and not (unfiltered). Only data from the major peak is shown. 54

LIST OF FIGURES

Figure 1: (A) Average amount of DNA extracted per accession using the Qiagen DNeasy Plant Maxi Prep Kit. [N=45]. (B) Average amount of DNA extracted per accession using the Qiagen kit with two statistical outliers removed (both from *C. amplexicaulis*) [N=43]. (C) Average amount of DNA extracted per accession using the Qiagen kit with all values less than 10 ng/µl removed from data set [N=37]. (D) Average amount of DNA extracted per accession using the Qiagen kit with all values less than 10 ng/µl removed along with statistical outlier from *C. amplexicaulis* [N=36]. 37

Figure 2: Average ratio of absorbance of light at 260nm and 280nm wavelength used to qualify DNA extracted using the Qiagen DNeasy Plant Maxi Prep Kit. (A) Average of all samples included [n=45, mean=1.70], (B) Average of all samples except the one outlier removed from *C. transarisanensis* [n=44, mean=1.66], (C) Average of all samples larger than 10 ng/µL [n=37, mean=1.65]. 39

Figure 3: Degradation of RNA samples from *C. sinensis* and *C. amplexicaulis*. Samples are run on 2% agarose gel post treatment with DNase. 42

Figure 4: (A) Descriptive statistics for number of polymorphisms identified across all genes sequenced and camellia accessions. One outlier was identified then removed, and all statistics were re calculated. (B) Box and whisker plot of data with outlier present. (C) Box and whisker plot with outlier removed. 46

Figure 5: Graph of contig length (bp) vs. the number of polymporhisms identified. The solid black linear regression line shows the regression for the whole data set. The dash line shows the positive linear regression when the outlier (indicated by red arrow) is removed..... 47

Figure 6: Evolutionary history of *C.* accessions inferred using the Neighbor-Joining Method. Evolutionary distances were computed using the Tamura-Nei method and are in the units of the number of base substitutions per site. Distances are placed next to their corresponding branch. 49

Figure 7: Fluorescence emission of C. sinensis (red) and C. sinensis var. assamica (gray). 52

Figure 8: Upper fluorescence regions, representing cells actively going through the G_2 stage of cell division, of all camellia accessions. *C. sinensis var. assamica* (gray) is noticeably larger and was classified as an outlier. Table indicates number of nuclei that (N=25,000) fell within this range. 53

Figure 9: All species accessions' fluorescence patterns. On the left, all accessions are overlapped, showing that only *C. transarisanensis* (purple) was significantly different. Graph on the right shows same samples, but graphs are offset vertically 56

Figure 10: All *C. japonica* cultivars fluorescences. *C. japonica* 'Dr. Tinsely' (pink) shows the only distinct difference, however it is not statistically significant. 57

ABSTRACT

Camellias have been cultivated for centuries beginning in ancient China, and have become important worldwide for both ornamental and economic reasons. The genus is the largest within the *Theacea*, and has been estimated to consist of 100-300 different species. The first director of Longwood Gardens, located in Kennett Square, Pennsylvania, identified camellias as one of the target genera for the newly created research department. The objective of this new department was to breed new cultivars of camellias that were consistently cold hardy in the region, and could be used for outdoors display. Today camellias are still a focus for the Longwood Gardens Research Department, and several cultivars have been released that exhibit exceptional cold resistance in the area.

Over the past several decades the science of plant breeding has undergone a drastic change. Available breeding technologies allow for the quick and efficient analysis of genetic material in order to better understand the genetic diversity of species, and to allow breeders to screen prospective new parent lines for their use in future crosses. This research focused on three such technologies – DNA analysis, next-generation RNA sequencing (NGS), and flow cytometry – and utilized them to help better elucidate the phylogenetic relationships between several *Camellia* taxa. A total of 19 camellia taxa, mostly housed at in Longwood Gardens' research collection, were used in this research including ten species and nine cultivar taxa.

The phylogenetic tree created through DNA sequencing confirms some results previously published by other authors, but does question how closely some other

species are related to each other. Primarily this research confirmed the treatment of the two subgenera *Camellia* and *Thea*, and suggests that *C. amplexicaulis* is the most distantly related taxa to all other taxa used in this research. These results could not be confirmed through the use of next generation RNA sequencing (NGS), as originally desired, but the data produced through NGS could be used later if additional research was done. This research indicated that DNA and RNA sequencing could provide deeper insight into the relatedness of different camellia species, and if desired this research could be expanded to include more *Camellia* species from the Longwood camellia collection. This genetic analysis requires specialized equipment, which may prove to be a large financial investment with potentially small practical returns.

The use of flow cytometry, however, could prove to be very beneficial for Longwood Gardens, allowing breeders to quickly assess the genome size of camellia accessions and to potentially screen seedlings for hybrid genome sizes. The data collected here shows that all taxa except for *C. transarisanensis* had very similar sized genomes. *Camellia transarisensis* showed a statistically significant increase in genome size, which may have negative implications in embryo development if this species is used for breeding.

Results from this study can be used to guide Longwood Gardens as they move forward with the camellia-breeding project, and provide insight into some of the pros and cons of different genetic analysis technologies. Furthermore some results can be used immediately to direct breeders in which species to include and exclude from future crosses.

Chapter 1

INTRODUCTION

The tea family (Theaceae Mirbel) is a family of angiosperms with near global distribution. The Plant List, an online collaboration between Kew, the Royal Botanic Gardens in the UK, Missouri Botanical Garden, and several other international organizations, lists 15 genera and 370 accepted species names. The largest genus in Theaceae is *Camellia* L. and consists of at least 250 accepted names ("The Plant List", 2013), with *Camellia japonica* L. and *Camellia sinensis* L. (Kuntze) two of the best known species. *Camellia sinensis* is the source of tealeaves and is now grown around the world as an economic crop (Durrant, 1982, p. 4). *Camellia japonica*, however, begins to display beautiful flowers anywhere from late winter to early spring. These flowers persist for about a month, and are prized for their use in the garden. In its native habitat, both *Camellia japonica* and *Camellia sinensis* are small trees, growing upwards of 18 feet (5.5 meters), and have evergreen leaves that are 2-4 inches (5-10 cm) long by 1-2 inches (2.5-5 cm) wide. The flowers of *C. japonica* are typically red with six or seven petals, however, many natural variants can be found, displaying a diversity of colors and flowering structures (Jiyin,, Parks, & Yueqiang, 2005, p. 43-44).

Though *Camellia sinensis* has never become a widely cultivated crop in the United States (Hume, 1955, p. 69), *Camellia japonica* cultivars, and some other select

ornamental species have enjoyed great popularity in the warmer sections of the country (Ackerman, 2002, p. 6). They have not become widely cultivated in cooler climates primarily due to their poor winter-hardiness (Hume, 1955, p. 5). Most camellia species originate from Southeast Asia and are unable to withstand the low temperatures and dry conditions of winters in most regions of North America. *Camellia japonica* is classified as hardy to USDA cold hardiness zone 7 or 8 (where temperatures do not go below -17ºC and -12ºC, respectively). This prevents camellias from growing outdoors on the East Coast much above Maryland and Southeastern Pennsylvania. Some organizations, including Longwood Gardens, and individuals are currently working to breed new camellia cultivars that demonstrate increased winter hardiness (Ackerman, 2002, p. 10).

Longwood Gardens is a public horticulture institution in Kennett Square, PA, renowned for their display gardens. Founded by industrialist Pierre S. du Pont, the property was originally purchased in 1906 to help preserve a stand of trees located near the old farmhouse situated on the property (Thomas, 1999, p. 2). After du Pont's death in 1954, the estate transitioned to a horticultural display garden, open to the public throughout the year ("Private to Public", n.d.). During this transition and overseen by Longwood Gardens' first director, Dr. Russell Siebert, the research department was formed, with *Camellia* being one of the original genera of interest. A research goal since the 1960s has been the creation and identification of new cold hardy camellia cultivars for outdoor display at Longwood.

When Longwood's research department first started breeding camellias, they relied primarily on controlled crosses and the keen eye of their research staff to select seedlings growing in their trials. This work was accomplished with the help of Dr. Clifford Parks, in California, as well as with the notorious camellia breeder Dr. William Ackerman on the East Coast. Dr. Parks in California, would make the crosses, and the seedlings were planted at Longwood Gardens. The first cultivar, *C x williamsii* 'Aida', was officially in selected in late 1970s after surviving several very cold winters with temperatures near -25.5ºC. Released in 1995, *C x williamsii* 'Aida' displayed semi-double flowers that were a delicate rose-magenta color, and proved to be one of the most cold hardy camellia selections (Thomas, 1999).

Over the past decades though, breeding technology has drastically changed. New genetic analysis techniques allow researchers to better understand the genetic relationships between parent plants and to better predict the resulting progeny (Brown & Caligari, 2008, p. 60). High throughput Polymerase Chain Reaction (PCR) and sequencing technologies allow genotyping of plants quickly and efficiently, and phylogenetic trees can subsequently be created to illustrate evolutionary distances between species. Flow cytometry allows for quick determination of the number of chromosome present in each individual. When combined, this information can help better inform breeders of which individuals to use in crosses to create the most diverse offspring possible.

In the past half-century of camellia breeding and selection at Longwood, several cultivars have been released. In addition to *C x williamsii* 'Aida', several

selections, such as *Camellia japonica* 'Longwood Valentine' and *Camellia japonica* 'Longwood Centennial', that display improved cold hardiness and single red cup-shaped flowers that bloom in early spring, were selected from a plant collecting expedition to South Korea in 1984 (Aiello, 2009, p. 28). Moving forward they hope to not only increase the cold hardiness of new cultivars, but also to improve and diversify the shape, color, and size of flowers available in these cold hardy camellias. To help with this project the research department is interested in incorporating new breeding technology available to better assess the diversity of their camellia collection and to streamline their breeding process.

The objective of this thesis was to investigate the potential use of new genetic techniques to assist the Longwood Gardens' camellia breeding program. The aim was to create standard protocols for nucleic acid extraction and purification, in addition to optimization PCR conditions for successful amplification of camellia genes. These tools can then be used to further the investigation into the genetic diversity of the camellia collection at Longwood Gardens.

Chapter 2

LITERATURE REVIEW

2.1 Introduction and Overview of the Genus *Camellia* and Their Cultivation

The genus *Camellia* (L.) is in the tea family and is named after the Reverend Georg Jossef Kamel, though the German physician, Andreas Cleyer, first described it to western civilizations during a trip to Japan in the 17th century (Ackerman, 2007, p. 2). All members of the genus are small trees and shrubs. Leaves are alternate, usually evergreen, and simple with either serrate, serrulate or, rarely, entire margins (Trehane, 1998, p. 43). Bloom time spans from late fall/early winter through late winter/early spring and spring, depending on the species and to some extent cultivar. Flower structure and location varies between sections. Today, camellia species are highly valued for their economic uses, as well as ornamental value. The modern agricultural crops of camellia include *Camellia sinensis,* from which tea is derived, and *C. oleifera* Abel, from which camellia oil is produced (Durrant, 1982, p. 4-8). The cultivation of camellias as an ornamental plant is relatively more recent, and started in the 3rd century C.E. China with two species, *Camellia japonica* and *Camellia reticulata*. Both are native to eastern and southwestern Asia, respectively, bare simple flowers with a single layer of petals, and natural variants are easily found growing in the wild (Trehane, 1998, p. 17). As selection occurred over the centuries, a variety of cultivars

displaying a range of color, flower structure and flowering time have been described (Trehane, 1998, p. 18).

Camellia cultivation as both an agricultural crop, and garden plant was well established prior to the British arriving in China in the late 17th century. The first specimens were returned to Great Britain in the late 1600s as dried specimens by way of the East India Company (Hume, 1955, p. 13). Carl Linnaeus is credited with introducing the first living camellia plant in the 1730s. He was attempting to start a tea plantation in Europe, and purchased what he thought was specimens of *Camellia sinensis*. Many of these plants, however, turned out to be *Camellia japonica*, and were some of the first ornamental camellia species introduced to the western hemisphere (Hume, 1955, p. 19). By 1747 camellia plants were growing in British glasshouses, and author George Edwards was one of the first to describe such plants:

> [It] is of a red Rose color, with stems in the Middle of a Yellow or Gold Colour…The green leaves were stiff, firm and smooth, like those of Evergreens…[T]this beautiful flowering Tree was raised by the late curious and noble Lord Petre, in his Stoves at Thorndon-Hall in Essex. (Trehane, 20-21)

Camellias in Europe took a couple more decades to really become established as a popular garden plant, but the introductions of *Camellia japonica* 'Alba Plena' and *C.* japonica 'Variegata' in 1792 ushered in a golden age for camellia growers (Hume, 1955, p. 21). By the turn of the century, live plants were being imported regularly into Great Britain, and could be found in the hothouses of the elite class throughout the UK. The popularity of camellias, and camellia breeding, quickly expanded throughout

Victorian England after it was discovered that some camellias could be grown outdoors in the cultivated landscape (Trehane, 1998, p. 21).

Almost simultaneous with the rise in popularity of camellias in England, came the fascination with camellias in the New World. Plants were first successfully introduced to the newly formed United States of America in the late 1790s. John Stevens' nursery, located in what is now Harlem, on New York City's Manhattan Island, was the first to successfully grow seeds of *Camellia japonica* in the U.S (Hume, 1955, p. 24). The popularity of indoor-grown camellias spread throughout the East Coast and were very popular among the affluent in Boston, New York City, and Philadelphia.

2.2 Camellias at Longwood Gardens & Breeding Efforts

Longwood Gardens' founder Pierre du Pont built the first part of the conservatory in 1921 to allow him to display many plants that could not normally grow in Southeastern PA. Some of the first plants to be installed in the Conservatory were selections from du Pont's camellia collection ("Camellia House", n.d.). Mister du Pont's fascination with camellias began in 1912 when he imported specimens from Georgia, and he imported bulk orders of named selections from France in 1916 (Thomas, 1999, p. 3). When the camellia collection was installed in the Conservatory, it reportedly had over 40 different varieties of *Camellia japonica* (Huttleston, 1960, p. 124). The first official inventory of camellias at Longwood Gardens was not conducted until 1937, at which time it was determined the collection consisted of 372

plants, representing 167 different cultivars. One of these was a specimen of *Camellia reticulata*, however all other accessions were *Camellia japonica* cultivars (Huttleston, 1960, p. 125).

After du Pont's death in 1954, as the estate transitioned to a public garden, *Camellia* became a species of interest for the newly formed research group. Over the years, Longwood Gardens has continued to accession new species and cultivars of camellia for both display and research purposes. One of the goals of the camellia breeding group and Longwood Gardens was to use their collection and expertise to help breed new winter-hardy camellia cultivars for display. This task was originally undertaken collaboratively with researchers around the U.S. (Thomas, 1999, p. 3), and was accomplished primarily through hybridizations and field trials. In the early 1960s, Clifford Parks, then at the Los Angeles State and County Arboretum in Arcadia, CA, made crosses between cultivars known to have superior cold hardiness (Ackerman, 2002, p. 15). Parks subsequently worked with Longwood's research department and the US Department of Agriculture to expand the accessions used in the crosses and to conduct field trials for selecting cold hardy cultivars (Ackerman, 2002, p.16). During this same time, William Ackerman was embarking on similar work, first through the U.S. Plant Introductions Station in Maryland, and then at the United States National Arboretum. Prior to the mid-1900s, many cultivars were created through open pollinated crosses, in which identifying the pollen parent was difficult and mostly speculative (Ackerman, 2002, p. 77). The use of controlled crosses during the 20th century created a drastic increase in the number of registered cultivars, all with careful

documentation (Ackerman 2002, 78). This practice relied heavily upon the identification of promising parent lineages to cross, and still reliant on subjective breeder evaluation. Furthermore, these breeding schemes required large trial areas, and compatible interspecific hybrids had to be discovered through trial and error.

More recently, the increased use of genetic knowledge in breeding schemes has helped streamline the process of making novel crosses, and has assisted breeders in predicting breeding outcomes more effectively (Brown & Caligari, 2008, p. 60). Being able to diminish the time between initial hybridization and readiness for the market (as a named cultivar) would be highly beneficial for the camellia breeding community. Furthermore, increasing species diversity for breeding schemes can help achieve breeding goals, such as cold-hardiness. Current efforts using only the major horticultural species (i.e. *C. japonica*, *C. reticulata*, and *C. sasanqua*) are showing a diminishing return, due to the restricted gene pool (Ackerman, 2007, p. 125). Using some of the lesser-represented species could help broaden the genetic material available to breeders, and thus make future crosses more productive (Ackerman, 2002, p. 78). Agricultural crop breeders often try to include wild relative of domesticated crops in breeding regimes in order to help widen the gene pool, and hopefully discover new genes for solving a variety of problems (Acquaah, 2002, p. 200). It can be argued that *Camellia japonica* has become a domesticated species through the centuries of breeding, and now modern camellia breeders must look outward to less utilized species to find continued success in their breeding practices (Ackerman, 2007, p. 126).

2.3 Genetics of Camellias

Beginning in the 1950s, the base chromosome number of *Camellia* was investigated and the subsequent ploidy numbers of different species were identified (Ackerman, 1971, p. 51). Much of this initial research was conducted through the creation of karyotypes from root tips (Ackerman, 1971, p. 9). Cells were treated and fixed, and photomicrographs were taken of the cells and chromosomes were counted visually (Ackerman, 1971, p. 11). The base number of chromosomes for *Camellia* is 15, meaning that gametes will carry a multiple of 15 chromosomes – 15, 30, 45, etc. The number of chromosomes varies among species, and in some instances varies within species (Huang, Tong, Zhang, & Gao, 2013, Table 4; Kondo, 1977, Table 1).

Creating interspecific hybrids, though beneficial, can be hindered by incompatibilities between species. Species differing in the number of sets of cellular chromosomes (ploidy) can prevent successful embryo development (Brown & Caligari, 2008, p. 164). In general, unequal gene dosage in a zygote – caused by unbalanced chromosome numbers – is lethal (Pierce, 2012, p. 259). Katsuhiko Kondo (1977) listed the ploidy number of various camellia species and cultivars previously determined (Kondo, 1977, Table 1). While the chromosomal base number for the genus *Camellia* is 15 (x=15), the data presented by Kondo actually showed a large variation in chromosome number among species, ranging between diploid (2n=2x) to some *C. sasanqua* cultivars that have been recorded as octoploids (2n=8x). There is, however, within-species variation, thus it is important to determine the ploidy numbers

of specific specimens within a breeding regimen to facilitate the most productive crosses (Ackerman, 1971, p. 51).

In recent years, the process of ploidy determination has become increasingly simplified through the use of flow cytometry. This process analyzes the fluorescence of small particles in a suspension as they are passed through a high power laser. This technology has been adapted to analyze nucleic acid content by staining the genomic DNA with a fluorescent dye and cells are individually passed through the laser at high speeds. The brighter the fluorescence created by a cell, the greater the amount of DNA is present, thus quickly and easily determining the relative ploidy number of each cell in a population. This has quickly become the preferred method of ploidy determination (Dolezel & Bartos, 2005). By using this technique, many more camellia species have been examined, however, the variation in chromosome number is still present between and within species (Huang, Tong, Zhang, & Gao, 2013). This genus has been described as "one of the most challengingly difficult taxa in plants, whose complexity is primarily governed by frequent hybridization…polyploidization and subsequent stabilization" (Huang, Tong, Zhang, & Gao, 2013). Breeders at Longwood Gardens have begun to understand the sexual compatibility of different cultivars and species in their collection, however the use of flow cytometry to quickly analyze accessioned individuals, may help expedite the breeding process by identifying which species are, and more importantly which are not, compatible for crossing.

Over the past decade, scientists have started to work at sequencing segments of the *Camellia* genome as well. This process has, at times, been difficult due to the high

concentrations of foliar polyphenols and polysaccharides in camellias. In *C. sinensis,* polyphenols can represent up to 25% of the dry weight of leaves, and polysaccharides can account for an additional 0.4% to 1.5%, depending on freshness and quality of leaves (Muoki et al., 2011). These compounds inhibit the polymerase chain reaction (PCR), degrade nucleic acids, and lead to hydrolysis of RNA, and therefore extraction and isolation of high quality DNA and RNA is essential for genetic research (Yockteng et al., 2013). Recent studies have shown though, that with some modification to standard DNA extraction protocols, high-quality DNA can be extracted from *Camellia sinensis* at sufficient concentrations (125µg/ml) and purity for further downstream analysis (Umar et al., 2011). Similarly, protocols have been published demonstrating the successful use of cetyltrimethylammonium bromide (CTAB) to extract sufficient yields of pure RNA at qualities suitable later use in Next-Generation RNA Sequencing (Yockteng et al., 2013).

GenBank®, the National Institute of Health's online genetic sequence database, lists 778 messenger RNA (mRNA) sequences derived from camellia species, 638 of which originate from *Camellia sinensis*, and 38 from *Camellia japonica*. GenBank® also contains eight fully sequenced chloroplast genomes, including both *C. japonica and C. sinensis*. ("GenBank", n.d.). The research focus on *C. sinensis* and *C. japonica* stems from their relatively high economic value as ornamentals. Though there are some sequences derived from lesser-known species of *Camellia*, to be able to build a phylogenetic tree of the species housed at Longwood Gardens, de novo sequencing would have to be conducted on many of their accessions.

In a study by Taniguchi et al. (2012), seven cDNA libraries were created from *Camellia sinensis* and 17,458 Expression Sequence Tags (ESTs) were generated from these. One hundred of these ESTs were selected, and primer sets were developed to amplify them in a *C. sinensis* model. Of these 100 primer-sets, 71 were able to produce well-amplified fragments, and of these, 70 were able to amplify segments in more than one species of *Camellia*. Fourteen different *Camellia* species were used to investigate the interspecific properties of these primers, and on average 81.7% of *Camellia sinensis* markers amplified in each of the different species (Tanigcuchi et al., 2012). In a similar experiment four primer sets were successfully used to amplify microsatellite regions across 22 different *Camellia* species and varieties (Caser, Torello, & Scariot, 2010). This research demonstrates that it is possible to create primer sets based on previously sequenced *Camellia sinensis* genes retrieved from GeneBank® and use them to create primers for other species. By sequencing identical genes across numerous species of *Camellia* mutation rate, and relative evolutionary distance can be estimated.

2.4 Classification of Camellias & Species of Interest

When camellias were first cultivated as an ornamental crop, there were only two described species: *Camellia japonica* and *Camellia reticulata.* By the 1920s, this number had grown to include a total of 40 species. In 1958, however, J. Robert Sealy published his book, *A revision of the Genus Camellia*, which described 82 species and 20 unclassified species. The number of described species continued to grow as

botanists explored Asian countries; in 1984, the book *Camellia* by Chang Hung Ta and Bruce Bartholomew was published and listed over two hundred different species divided into twelve sections (Ackerman, 2007, p. 5). This number has continued to grow and by the end of 2013 there were over 430 species listed in Tropicos, the Missouri Botanical Garden's online nomenclature database. Although there are some discrepancies among authorities regarding Camellia nomenclature, this thesis deferred to the work of Gao Jiyin, Clifford R. Parks, and Du Yueqiang entitled *The Collected Species of the Genus Camellia: an Illustrated Outline* (2005) for nomenclature references. The following species description was used in this research. When notable synonyms occur, they will be noted in the following descriptions of the selected species.

Section *Camellia* (L.) Dyer, is the largest section of the genus *Camellia*, and includes around 60 species, many of which have ornamental value. This section is characterized by large, usually red flowers with basal fusion of the flower parts (Jiyin, Parks, & Yueqiang, 2005, p. 27). Several species were selected from this section of *Camellia* to be used in this research; they include: *C. japonica*, and *C. azalea* (syn. *C. changii*). *Camellia azalea* Wei, is of notable interest due to its asynchronistic growth pattern, thus producing flowers throughout the entire year. Flowers are bright red, consist of 5-9 petals and are borne at shoot tips (Jiyin, Parks, & Yueqiang, 2005, p. 35). Breeders are now using this species to create a camellia hybrid that blooms continuously (Jiyin, Parks, & Yueqiang, 2005, p. 34).

Camellia amplexicaulis (Pitard) Coh. St. is a member of section *Archecamellia* Sealy of the genus *Camellia*. This section consists of only three taxa characterized by large leaves and single flowers borne at shoot tips (Jiyin, Parks, & Yueqiang, 2005, p. 22). *Camellia amplexicaulis* produces a purplish-red flower with fleshy concave petals and blooms in the early autumn. Because of its flowering characteristics and other phenotypic traits, Gao Jiyin and colleagues see potential benefits from using *C. amplexicaulis* as a parent (Jiyin, Parks, & Yueqiang, 2005, p. 23). Longwood Gardens is increasing the use of this plant in their breeding regime.

In the 1980s, research focused on bettering the understanding of the relationships within the genus *Camellia*. During this time, the section *Oleifera* Chang (1981) and the section *Paracamellia* Sealy (1958) were split. Both sections have white, slightly fragrant flowers. Chang separated the sections based on the elongated styles and anthers found in what is now section *Oleifera*; the oil content of species from section *Oleifera* is also higher than that of section *Paracamellia* (Jiyin, Parks, & Yueqiang, 2005, p. 168). This thesis research included two species from section *Oleifera* (*Camellia oleifera* and *C. sasanqua)*, and one species from section *Paracamellia* (*Camellia grijsii)*. *Camellia grijsii* is valued by some breeders for is proliferous flowering in late winter and early spring, though few hybrids have been successfully created (Jiyin, Parks, & Yueqiang, 2005, p. 182). *Camellia oleifera* is also of interest for its cold tolerance, yet less so for its ornamental properties (Jiyin, Parks, & Yueqiang, 2005, p. 170). Mainly grown in Asia as an economic oil crop,

Camellia oleifera has already proven instrumental in parenting more cold hardy camellia cultivars.

Camellia sinensis (L.) O. Kitze, the origin of tea as a beverage, is the most economically important species. As a member of the section *Thea*, *C. sinensis* has small, axillary, white flowers that bloom in late autumn and early winter. Though it is not widely grown as an ornamental, it can have a well-shaped habit, and has gained some popularity as a landscape plant (Jiyin, Parks, & Yueqiang, 2005, p. 225). This plant was included in this thesis research because of the surfeit of genetic information in scientific publications.

This research also strived to include species from some of the smaller, often underrepresented, sections of the genus. *Camellia gaudichaudii* orginates from section *Furfuraceae*, which is characterized by solitary white flowers having 6-12 petals and a slight fragrance. Interestingly, the species within this section are not graft-compatible with many other species or cultivars from other sections (Jiyin, Parks, & Yueqiang, 2005, p. 145). *Camellia granthamiana* belongs to the section *Protocamellia*, whose members have white flowers with many petals (10-14) with unfused stamens and styles (Jiyin, Parks, & Yueqiang, 2005, p. 201). There is much debate about whether *C. granthamiana* is synonymous with *Camellia albogigas*. They are reported as two species, although recognized that there has not been a documented, distinguishing characteristic to help discriminate the two species (Jiyin, Parks, & Yueqiang, 2005, p. 202). The 2010 Tropicos changed *C. albogigas* to a synonym of *C. granthamania* ("Tropicos", n.d.). Based on this recent research, the name *C. granthamania* will be

referenced this thesis research. Finally, *Camellia transarisanensis* (Hay.) Coh. St. is a member of the section *Theopsis*, which is most noted for its small leaves, and abundance of small fragrant flowers. In 2000, *C. transarisanensis* and *C. handelii* were grouped together, although currently, Longwood Gardens still labeled the accession as *C. handelii,* the name *C. transarisanensis* will be used in this thesis research. The fine-textured foliage and dense branching characteristic of this species is drastically distinct from the other camellia species included in this thesis research. A list of all species used during this research can be found in table 1.

Chapter 3

MATERIALS AND METHODS

3.1 Plant Material

All plant material was obtained from the University of Delaware Botanic Garden (UDBG), Longwood Gardens, and Camellia Forest Nursery in Chapel Hill, North Carolina. The University of Delaware Botanic Garden donated material from *Camellia japonica* (accession number 99-81*1) growing outdoors in a protected courtyard. *C. sinensis* var. *sinensis* was purchased from Camellia Forest Nursery in March 2013. All other samples were obtained from accessions growing at Longwood Gardens. Since the taxonomical classification of camellias is controversial, all species nomenclature was checked using the online Plant List (www.theplantlist.org) as well as The Collected Species of the Genus Camellia (Jiyin, Parks, & Yueqiang, 2005).

3.2 Species Selection

This research focused primarily on species accessioned at Longwood. A full inventory of the research greenhouses at Longwood (G36) was conducted to identify their accessions. Six species from Longwood were selected based on their health, ornamental interest, and potential use in future breeding efforts. Four more, outdoor-grown species were also included.

Two *C. japonica* cultivars were added to the list on the advice of Dr. Tomasz Anisko and Dr. Mathew Taylor (Longwood Gardens). These cultivars, *C. japonica* 'Kumasaka' and *C. japonica* 'Maiden of Great Promise' demonstrated similar flowering times, bloom color, and habit, and were included to investigate the possibility that they were genetic clones.

Dr. Matthew Taylor identified several new species and unique cultivars as plants of interest for use with flow cytometry. These had either been previously used in crosses at Longwood Gardens, or are expected to be used in the future, and were included as well.

3.3 Methods and Materials for DNA Analysis

3.3.1 Tissue Collection

Leaf tissue samples for DNA extraction were collected by pruning several small branches from each plant. Pruners were cleaned using alcohol wipes prior to each cut, and branches were chosen that contained relatively new growth that was neither deformed nor diseased. Samples were collected into new 1-quart resealable, plastic, storage bags, which were then sealed and labeled with accession number and collection date and placed in a styrofoam cooler with icepacks to prevent nucleic acid degredation. Samples were stored in a -20° C freezer.

3.3.2 Gene Sequences and Primer Design

A list of genes relevant to flower development and timing in *Arabidopsis thaliana* was compiled. This list was then used to find homologous genes in *C. japonica* and *C. sinensis* using NCBI's tblastn (NCBI, blast.ncbi.nlm.nih.gov). The resulting list of translated nucleotide sequences was then used to back query *A. thaliana* sequences. Camellia sequences that were highly similar to the original *A. thaliana* genes were used. Additional *Camellia* genes, known to be highly conserved across species, including Histone 3 gene and Ribosomal Spacer genes were selected. *Camellia* genes were identified from NCBI's GenBank (http://www.ncbi.nlm.nih.gov/genbank). NCBI's primer-BLAST was then used to create three to five PCR primer sets for each gene sequence. The expected product size

was set between 300 and 600 bp, and primer sets were selected to maximize coverage of the reference sequence.

All primers were initially tested using isolated DNA from *C. sinensis* and *C. japonica* using the conditions described below. PCR products were sequenced (see below for protocol), and successful primer sets that ensured the best coverage of the gene were used with other camellia accessions.

3.3.3 DNA Extraction

Approximately one gram of green leaf tissue was ground into a fine powder using a mortar and pestle and liquid nitrogen. DNA extractions were then preformed using Qiagen DNeasy Plant Maxi Kit (Qiagen; Venlo, Netherlands) according to manufacturer's recommended protocol with minor changes. Briefly, cells are lysed chemically and RNA contamination is removed using RNase. cell debris, and polysaccharides are physically removed and the sample is homogenized through a QIAshredder Maxi Spin Column. DNA is then adsorbed to a DNeasy Plant Maxi Spin Column while contaminates spin through. After rinsing the column, DNA is eluted with a low-salt buffer. The first elution was done with 1ml of Buffer AE, and collected into a 2.0ml microcentrifuge tube. The process was repeated with 0.75ml of Buffer AE and collected into a separate tube. The optical density (A260/280) of the samples was measured with a nanodrop spectrophotometer (Thermo Fisher Scientific; Waltham, MA). Samples were then diluted down to 10ng/µL using 1X Tris-EDTA buffer (10mM Tris, pH 8.0 and 1mM EDTA diluted into pure water) for use. One diluted sample was stored in a 4ºC refrigerator for easy use, while other samples were stored in a -20ºC freezer.

3.3.4 PCR Amplification

DNA amplifications were performed in 50µL or 20µL volume. The reactions contained 1x GoTaq Master Mix [a ready-to-use solution containing modified *Taq*

enzyme, dNTPs, $MgCl_2$ and reactions buffers (Promega Inc.; Madison, WI)], 5% Dimethyl sulfoxide (DMSO), 4.0µM of both the forward and reverse primer, 40ng of DNA and water to final volume.

Reactions were incubated in a Perkin Elmer 9700 thermocycler with the following conditions: 15 minutes at 95ºC to activate the *Taq* enzyme, then 40 cycles of the following three temperatures, 1 minute at 94ºC (denature), 1 minute at 55ºC (anneal), 1 minute at 72ºC (elongation), followed by a final extension incubation of 10 minutes at 72ºC.

PCR products were analyzed using the E-Gel® High-throughput DNA electrophoresis system (Life Technologies™, Thermo Fisher Scientific, Inc.; Waltham, MA). The 96-well gels were loaded with 8µL of PCR product and 12µL of water and run for 13 minutes. Gels were imaged using a UV-light, and then aligned using E-Editor™ 2.0 Software (Life Technologies™, Thermo Fisher Scientific, Inc.; Waltham, MA). All successfully amplified samples were quantified, purified, and submitted for DNA sequencing.

3.3.5 PCR purification

Successfully amplified PCR products were cleaned using ExoSAP-IT® (Affymetrix; Santa Clara, CA) to remove any unconsumed dNTPs and primers. For samples with higher concentration of DNA present (as determined by the agarose gel analysis), 8µL of PCR product was used, and for samples with lower concentrations of DNA present upwards of 15µL of PCR product was used. To this, 1µL of ExoSAP-IT® was added, and water was added to a final volume of 22µL. Samples were then incubated in a Perkin Elmer 9700 thermocycler with the following conditions: 20 minutes at 37ºC and 20 minutes at 80ºC (to deactivate the enzyme).

3.3.6 DNA sequencing

Once the PCR product was cleaned, 1µL of the forward primer was added to 10µL of clean PCR product, and 1µL of reverse primer was added to another 10µL of clean PCR Product. For some samples, 6.4µM of T3 and T7 primers were used as the forward and reverse primers, respectively. For some samples, however, the 6.4µM forward and reverse primers, originally used for the PCR, were utilized. Sequencing reactions were performed using the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit with AmpliTaq FS DNA polymerase (PE Applied Biosystems; Foster City, CA) and analyzed on ABI 377 (PE Applied Biosystems; Foster City, CA) sequencers. Any sequence ambiguities were resolved by repeated sequencing of the PCR products from both ends. All sequences were then aligned in Sequencher (Gene Codes Corp.; Ann Arbor, MI) using the original sequence downloaded from GenBank as a reference. Since many of the GenBank sequences were derived from cDNA, introns were not present, and gaps had to be added when PCR products contained intron sequences. Any ambiguous base calls were notated with conventional ambiguity codes, and the base changes at all polymorphic positions were identified by inspection and catalogued in an Excel (Microsoft Corporation; Redmond, WA) spreadsheet.

3.3.7 Phylogenetic Tree Creation - DNA

Once all sequences were aligned and analyzed in Sequencher (Gene Codes Corporation; Ann Arbor, MI), consensus sequences were created for each species at each gene. Sequences were then exported from Sequencher and imported into Microsoft Excel, where they were concatenated, with spacers inserted between genes, in order to create a single sequence for each species. Sequences were then imported into MEGA6 (MEGA; Tempe, AZ), with gaps left in and each accession was aligned. Insertion and deletion mutations where converted to a one letter polymorphism, so that they were equally weighted in the analysis. Aligned sequences were used to create a Neighbor-joining (NJ) phenetic tree using MEGA6 software using the Tamura-Nei

model. Most settings were left to default when constructing the tree; nucleotide substitution type was selected and included both transitions and transversions, and gaps were treated as a complete deletion. All heterozygous bases were left unchanged when imported into the MEGA6 software.

3.4 Materials and Methods for RNA Analysis

3.4.1 RNA Sample Collection

When possible, leaves were collected directly from living plants into liquid nitrogen to preserve freshness of tissue, and limit RNA degradation. When leaves could not to be collected directly from the plant, branches were collected from the plant and stored on ice until leaves could be removed in a lab setting (always less than an hour). Leaves were dropped into a plastic beaker containing liquid nitrogen, placed on a bed of dry ice. Leaves were crushed into a coarse powder and transferred to a 50ml tube that had been chilling on dry ice. The full 50ml tube was then packed on dry ice in an insulated cooler until it could be transferred to a -80ºC freezer. At a later date, samples were removed from the -80ºC and ground into a fine powder using mortar and pestle and liquid nitrogen. At all times samples were either kept on dry ice or in liquid nitrogen. The final product was again transferred to a clean 50ml tube, labeled and stored at -80ºC until needed for RNA extraction.

3.4.2 RNA Extraction Protocol

RNeasy Plant Mini-Kit (Qiagen; Venlo, Netherlands) was used as recommended to extract RNA from both *C. sinensis* and *C. ammplexicaulis.* Buffer RLC is first added to the samples to lyse cells and the mixture is then filtered through the QIAshredder columns to homogenize the samples. Ethanol is then added to the lysate, which is then transferred to the RNeasy Mini Spintube. The sample is spun down in a centrifuge, during which time the RNA binds to the silica membrane of the

RNeasy tube, and contaminants flow through. The RNA is then subsequently washed to remove remaining contaminants, and then eluted. RNA was then dried using an Eppendorf Vacufuge and suspended in 25µL RNase-free water. The optical density at 260 nm was measured and 2µL of product run on a 2% agarose gel to check for quantity and quality of final RNA.

RNA isolation protocol developed by Jones, Davies and Voelker (1995) using hexadecyltrimethylammonium bromide (CTAB) was used to extract RNA from all other species. Approximately one gram of tissue was ground in liquid nitrogen with 0.2g of Polyvinylpolypyrrolidone (PVPP) and then transferred to a 50ml Teflon FEP centrifuge tube, containing 12.5ml of extraction buffer (50mM Tris-HCl – pH 8.0, 10mM EDTA, 0.5% CTAB [w/v], 0.8M NaCl, and 1% [v/v] 2-mercaptoethanol). Sample was briefly vortexted then centrifuged at 15,000xg for 15 minutes to separate insoluble material. Supernatant was transferred to a new 50ml Teflon FEP centrifuge tube and deproteinized with 5ml of chloroform. Mixture was again vortexed and centrifuged at 10,000x g for 10 minutes. The top aqueous phase was transferred to a new tube and 1.5 volumes of precipitation buffer were added (50mM Tris-HCl – pH 8.0, 10mM EDTA, 0.5% CTAB [w/v], along with 0.2% 2-mercaptoethanol [v/v]). Samples were incubated at room temperature for 30 minutes then nucleic acids were separated by centrifugation at 15,000xg for 30 minutes. The supernatant was discarded and the RNA pellet was resuspended in 1M NaCl and 0.1% [v/v] 2-mercaptoethanol. The solution was then transferred to an RNase-free microfuge tube, and extracted twice with of phenol/chloroform (add 1 volume of phenol/chloroform, spin for 10 minutes at 13,000 rpm, transfer top aqueous layer, add 1 volume of phenol/chloroform, spin for 5 minutes at 13,000 rpm). The top aqueous layer was again transferred to a new RNase-free microfuge tube and 2 volumes of -80ªC 100% ethanol was added. Samples were incubated on dry ice for approximately 30 minutes. Isolated RNA was then sedimented by centrifuging samples in counter top centrifuge at maximum speed for 15 minutes. Liquid was discarded, and the remaining RNA pellet was washed with 80% ethanol. The pellet was dried and resuspended in 300µL

of RNase-free water. The A260/280 was then taken and 2µL of product run on a 2% agarose gel to check for quantity and quality of final RNA.

3.4.3 DNA Digestion

To treat RNA for DNA contamination, 5µL of 10x DNaseI reaction buffer along with 5µL of DNase was added to 250µL of each sample. Samples were incubated at room temperature for one hour, and then run on a new 2% agarose gel to verify DNA digestion. If DNA was still present, samples were allowed to incubate for an additional 30 minutes and samples were again run on a 2% agarose gel. No DNA contamination was noticed after this second incubation period. All samples were stored at -80ºC until needed for sequencing.

3.4.4 RNA Sequencing

RNA was sequenced based on a modified TruSeq® RNA Sample Preperation from Illumina (San Diego, CA). Briefly, Poly-A RNA was isolated from the total RNA mixture submitted through oligo dT magnetic beads. Then first strand cDNA synthesis was completed with SuperScript II Reverse Transcriptase and random hexamers, and second strand cDNA synthesis was preformed by adding Second Strand Master Mix (Illumina). RNA template was removed and double stranded cDNA molecules were isolated and purified. End repair, 3' end adenylation and index adapter ligation was performed according to published protocols (Illumina TruSeq® RNA Sample Preperation v2 Guide). Some samples were normalized using a modified Duplex-Specific Nuclease (DSN; Evrogen). Normalized cDNA was then enriched by PCR with the Illumina PCR Primer Cocktail under the following cycling conditions: 98ºC for 30 seconds, followed by 15 cycles of 98ºC for 10 seconds, 60ºC for 30 seconds and 72ºC for 30 seconds, a final extension step at 72ºC for 5 minutes was added at the end. The cleaned enriched DNA library was then evaluated using an Agilent Technologies 21000 Bioanalyzer. Cluster generation and sequencing of the

pooled library were performed on an Illumina cBot and HiSeq2500, respectively, according to manufacturer's instructions. Transcript discovery was carried out using Velvet and Oases software.

3.5 Materials and Methods for Flow Cytometry

Newly emerged leaf tissue was collected in mid-April and samples were prepared for analysis using the recommended procedure for the CyStain® PI absolute P kit from Partec (Münster, Germany). Briefly, samples were finely chopped using a razor blade in a petri dish with 500µL of Nuclei Extraction. Samples were then filtered through a Partec 50µm CellTrics disposable filter to remove leaf tissue, and 2.0mL of Staining Solution [2.0ml Staining Buffer, 12µL of Propidium Iodide, and 6µL RNase] was added. Samples were incubated at room temperature for roughly one hour before being analyzed using an Accuri C6 Flow Cytometer (BD Biosciences; San Jose, CA) to determine relative DNA fluorescence. Relative ploidy was then determined by comparing mean relative fluorescence of each sample to *C. sinensis*.

Chapter 4

RESULTS

4.1 Experiment Design

To study the genetic diversity of camellias, 19 different accessions were selected, mainly from Longwood Gardens' collection (Table 1). Of these, 21.05% were wild collected and an additional 11% were from known wild origin. There were two representatives from *Camellia* subg. *Thea* (*C. sinensis,* and *C. transarisanensis*), and the remaining species were from *Camellia* subg. *Camellia*. In total, eight different sections of the genus *Camellia* were represented in the chosen accessions (Jiyin, Parks, & Yueqiang, 2005). Several cultivars were also used and all but one were selections of *C. japonica*. The only hybrid not derived from *C. japonica* was *Camellia* 'Alice B. du Pont', which is an interspecific hybrid with *C. nitidissima* in its background.

4.2 Results of DNA Analysis

4.2.1 Quantity and Quality of Extracted DNA

When performing PCR, 10 ng/µL of DNA template is preferred. A total of 45 DNA extractions were performed and on average 29.55 ng/µL with a standard deviation of 25.73 ng/µL of DNA was extracted during each extraction. These values

were skewed upwards due to two values from *C. amplexicaulis*, which where abnormally high; 97.31 ng/µL and 120.98 ng/µL (Figure 1). The latter of these two concentrations was classified as an outlier (Grubbs' Test, α=.05, n=45), and once removed, the Grubbs' Test was performed again (α=.05, n=44), and the second value was determined to be an outlier as well. The corrected average concentration and standard deviation both decreased to 25.84 ng/µL and 19.27 ng/µL respectively (Table 2). The average concentration of extracted DNA was above the 10 ng/µL threshold required for successful downstream analysis. Eight extractions were below 10 ng/µL (17.78% of total extractions), including two from *C. transarisanensis*, two from *C. azalea,* and one from the following: *C. granthamiana, C. gaudichaudii*, *C. japonica* 'kumasaka', and *C. oleifera*. When these values were removed from the original data set, the average concentration of DNA extracted per accession increased to 30.18 ng/µL ± 18.79 ng/µL. With this data only one value (120.98 ng/µL) from *C. amplexicaulis* was identified as an outlier.

To measure the quality of the DNA, the ratio of optical density of the sample at 260nm and 280nm wavelength (A260/280) was measured. It is generally accepted that an A260/280 ratio of around 1.8 is pure DNA, values lower than 1.8 usually indicate the presence of protein contamination. The average ratio for DNA extracted during this experiment was 1.66 (a single outlier was removed from the data; Grubbs Test, α=.05, n=45) with a standard deviation of 0.22, and a fairly normal distribution around the mean. This average was statistically lower than the expected A260/280 ratio (Student's t-test, α=.05). Figure 2B shows the average A260/280 ratio for each

accession of camellia tested. When the extractions that yielded less than 10 ng/µL of DNA were removed (total of 8 samples), the average A260/280 decreased to 1.65 and the standard deviation remained at 0.22 (Table 3).

The average A260/280 ratio for four out of the twelve accessions was statistically less than 1.8 (T-test, n=44, α=.05) (Table 4). When all values less than 10 ng/µL were removed, *C. japonica* 'Kumasaka' was no longer identified as significantly less than 1.8 (Table 4).

4.2.2 Analysis of DNA Sequences

A total of 22 genes were successfully amplified in *C. sinensis*, which was used as an initial control for all primers sets created. Of these 22 genes, only 12 were successfully amplified across all accessions. The 12 genes that were ultimately used to create the phylogenetic tree are in Table 6. The complete list of primers used to amplify regions of the camellia genome can be found in Appendix C.

An average of 488 base pairs was sequenced per gene with a standard deviation of 227 bp, indicating some assembled contigs were much larger than others. On average, 5561 ± 263 bp were sequenced per camellia accession. Table 7 contains the number of base pairs sequenced per gene as well as the total number of base pairs sequenced per accession.

When total contigs were assembled (all genes concatenated from one accession) they ranged in lengths from 5170bp (*C. granthamiana)* to 5996bp (*C. sinensis*) with an average length of 5562bp sequenced (Table 7). A total of 762

polymorphisms were identified across all 12 contigs assembled. These included insertions, deletions, transitions, transversions, and small nucleotide repeat regions. The chalcone synthase gene (AY169403) had the highest rate of polymorphisms recorded at 29.44 polymorphisms per 100 base pairs sequenced. The apatela 1 gene (JX657332) was on the other side of the range with the lowest rate of polymorphism at only 3.50 polymorphisms per 100 base pairs sequenced. On average there was 13.03 ± 7bp polymorphisms identified per 100bp sequenced. Figure 4 shows the percent of polymorphisms averaged across all assembled contigs.

As mentioned, the Chalcone synthase gene had the highest percentage of identified polymorphisms and can be classified as an outlier (Grubbs Test for Outliers; α =.05, n=12). When removed, the average percent of polymorphisms dropped to 11.54 ± 5% (Figures 4A and 4C). The number of polymorphisms discovered per assembled contig was positively correlated with the number of base pairs ($R^2 = 0.22$), and the correlation increased in strength ($R^2 = 0.33$) with the outlier removed (Figure 5).

Table 8 shows the number differences between each accession. *C. amplexicaulis* and *C. sasanqua* had the most number of recorded differences (385bp), while *C. gaudichaudii* and *C. japonica* 'Maiden of Great Promise' share the most number of similar base pairs, and only had 95 discrepancies. All variance tables, broken down by gene, can be found in the Appendix D.

This data was then used to create the phylogenetic tree in MEGA6.0 (Figure 6). *C. japonica* and related cultivars grouped together, as expected, and *C.*

amplexicaulis represented the outgroup. *C. sinensis* and *C. transarisanensis* also grouped together, which was to be expected, as they are both from the *Camellia* sect. *Thea*.

4.3 Results of RNA Analysis

4.3.1 Quantity and Quality of Extracted RNA

Two methods were tested for extracting RNA from selected camellia accessions. The method using cetyl trimethylammonium bromide (CTAB) extracted significantly larger concentrations (ng/µL) of RNA than the protocol utilizing the Qiagen RNeasy kit (Difference of means T-test, $n_{(CTAB)=}15$, $n_{(RNeasy)=}20$, $\alpha=.05$). Prior to submitting samples for RNA sequencing, samples were run on a 2% agarose gel and optical density were again measured to ensure quality of samples. At this time, two samples (*C. sinensis,* and *C. amplexicaulis*), originally extracted using CTAB, appeared degraded, and so were discarded (figure 3). Previously extracted RNA, using the RNeasy Qiagen Kit, from both of these species showed no signs of degradation (not shown) and were of sufficient quality and concentration and so were submitted for sequencing instead.

All submitted samples had RNA concentrations greater than 40 ng/µL, sufficient for next-generation RNA sequencing, and most had concentrations above 100 ng/µL. The ratio of light absorbance at 260nm and 280nm wavelength (A260/280) can also be used to assess RNA quality, similar to DNA. An A260/280 ratio of around

2.0 signifies pure RNA. The average A260/280 ratio of the samples submitted was 2.1, with a standard deviation of 0.05, indicating no concern of protein contamination. The concentrations and A260/280 ratio for each sample submitted are in Table 5. Since only one sample was submitted for sequencing, these values are not an average but rather the actual concentration and absorbance ratio of the sample.

4.3.2 Analysis of RNA Sequences

Due to alternative splicing of genes, one DNA sequence can be translated into several different transcripts. Therefore it is important to look at both the number of genes sequenced by RNA and the number of cDNA transcripts. On average 143,117 transcripts were assembled from the samples that were normalized using the Duplex-specific Nuclease (DSN). For those samples that were not normalized with DSN, 209,501 transcripts were successfully assembled (Table 9). These values were two values were not significantly different from each other (two sample t-test, α=.05). The normalized transcripts represented 34,641 ± 4,152 genes on average, and the non-normalized transcripts represented 43,757 ± 9,210 genes on average. The number of genes present in the normalized and non-normalized samples were not significantly different from each other (two sample t-test, α=.05. The variability in transcripts assembled, and the genes represented by these transcripts, is significantly larger among the samples that were not normalized, than among the samples that were normalized (F-Test, α=.05). This is mostly due to the non-normalized sample of *C. sasanqua*, from which 329,420 transcripts were assembled representing 51,451 genes (Table 9).

On average, there were four transcripts sequenced per gene across all samples (both normalized and non-normalized). This number is again drawn slightly upwards by the large number of transcripts and genes sequenced from the non-normalized *C. sasanqua* sample, but is not significantly different from the mean (Grubbs outlier test, α=.05).

A smaller, and more manageable, Basic Local Alignment Search Tool (BLAST) database was created for each line that consisted of the longest transcript from each gene. These databases were then cross-referenced and a total of 17,441 transcripts had similar sequences (greater than 1e-10 significance score) present in more than one accession. Roughly one third of these transcripts (5,843 in total) had similar sequences in all ten accessions, and furthermore 1,736 sequences had an alignment region of 1kb or longer in all ten accessions.

The original genes sequenced through DNA sequencing (Table 6), where also used to search the RNA transcriptome database. All 12 genes had at least one transcript in at least one accession match at a high level of confidence, however only eight genes had matching transcripts from all ten accessions when cross-referenced. Many of these transcripts are quite long, so though a gene may have a matching transcript from each accession not all transcripts will align perfectly. Only transcripts representing six genes, originally sequenced, could also be aligned across all ten accessions. The results to date did not lead to additional conclusive results.

4.4 Results of Flow Cytometry

A total of 25,000 nuclei per accession were analysized for emissions in the 585/40nm range of light. The average height of and standard deviation were both exagagerated by the results derived from samples of *C. sinensis var. assamica* (Table

10). This species was included in the samples analyzed through flow cytometry because previously published literature showed that it had a similar sized genome to *C. sinensis* (Huang, 2013). However, upon analysis, the sample appeared to have a very unique flouresecence pattern (Figure 7) and was statistically classified as an outlier ($\alpha = .01$, Grubbs Test for Outliers). Note the large secondary peaks present in *C. sinensis* var. *assamica* that are not present in *C. sinensis*.

All samples did have a small second peak in this upper region, representing cells in the G_2 phase of cell divisions, but for all other samples this represented less than 2% of the entire sample, where as for *C. sinensis* var. *assamica* it represenetd over 18% of the the entire sample (Figure 8). For these reasons, *C. sinensis* var. *assamica* was removed from the data before further analysis. These secondary (G_2) peaks were also removed from all samples.

Tannic Acid (TA) and other secondary metabolites (SM) have been shown to interfere with flourochrome binding to DNA, and also may cause nuclei to aggregate while in solution. To lessen the negative effects of TA and SM, newly emerged leaves were selected when possible, which contain lower concentrations of TA and SM, and all samples were vortexed prior to analysis to homogenize the mixture, but a cell separator was also used on half of each sample to see if it affected results. On average, the cell separator increased the median height of fluorescence (two sample T-test, $\alpha = .05$), and more importantly, decreased the standard deviation of the sample (F-Test, $\alpha = .05$) (Table 11).

All samples showed mean values very close together, indicating that all were very similar in genomic content to *C. sinensis,* which was being used as a control (Figure 9). *C. transarisanensis* (purple) was the only peak that showed statistically signficant divergence from the group ($\alpha = .05$, Grubbs Test). When looking at only *C. japonica* and associated cultivars, all samples were very similar, and there was no indication of any genome duplication events (Figure 10).

Table 1: Selection of camellia taxa used in this experiment. Samples with an * next to them indicate accessions not housed at Longwood Gardens. A † next to an accession indicates a sample only used for analysis with flow cytometry.

#	Name	*Section* Author	Notes
1	*Camellia* 'Alice B. du Pont' †		*Camellia nitidissima* hybrid
2	*Camellia amplexicaulis*	*Archecamellia* Sealy	
3	*Camellia azalea*	*Camellia* (L.) Dyer	
4	*Camellia gaudichaudii*	*Furfuraceae*	
5	*Camellia granthamiana*	*Protocamellia*	syn. *Camellia albogigas*
6	*Camellia grijsii*	*Paracamellia* Sealy	
7	*Camellia japonica* *	*Camellia* (L.) Dyer	Growing outdoors
8	*Camellia japonica*	*Camellia* (L.) Dyer	Growing indoors
9	*Camellia japonica* 'Dr. Tinsely' †	*Camellia* (L.) Dyer	
10	*Camellia japonica* 'Kumasaka'	*Camellia* (L.) Dyer	
11	*Camellia japonica* 'Lady de Saumarez' †	*Camellia* (L.) Dyer	
12	*Camellia japonica* ‘Longwood Valentine’ †	*Camellia* (L.) Dyer	
13	*Camellia japonica* ‘Longwood Centennial’ †	*Camellia* (L.) Dyer	
14	*Camellia japonica* 'Maiden of Great Promise'	*Camellia* (L.) Dyer	
15	*Camellia japonica* ‘Tama-no-ura’ †	*Camellia* (L.) Dyer	
16	*Camellia oleifera*	*Oleifera* Chang	
17	*Camellia sasanqua*	*Oleifera* Chang	
18	*Camellia sinensis* *	*Thea* (L.) Dyer	
19	*Camellia sinensis var. assamica*†*	*Thea* (L.) Dyer	
20	*Camellia transarisanensis*	*Theopsis* Coh.St	syn. *Camellia handellii*

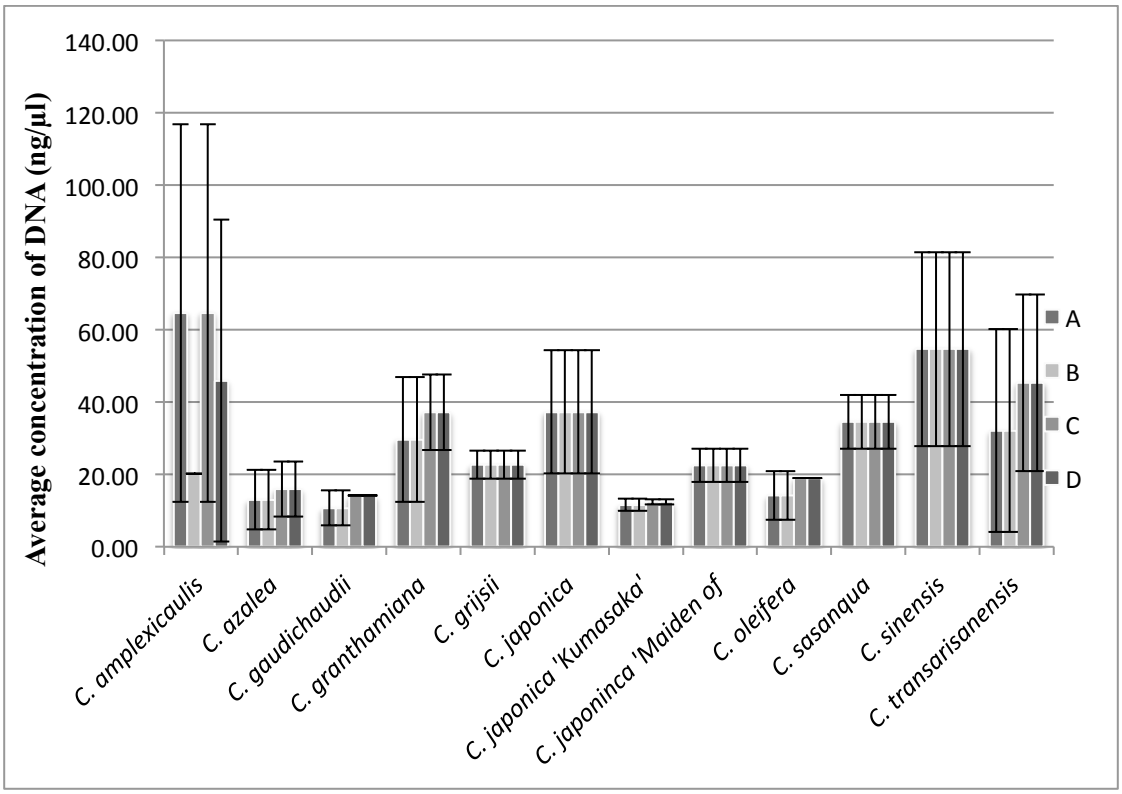


Figure 1: (A) Average amount of DNA extracted per accession using the Qiagen DNeasy Plant Maxi Prep Kit. [N=45]. (B) Average amount of DNA extracted per accession using the Qiagen kit with two statistical outliers removed (both from *C. amplexicaulis*) [N=43]. (C) Average amount of DNA extracted per accession using the Qiagen kit with all values less than 10 ng/µl removed from data set [N=37]. (D) Average amount of DNA extracted per accession using the Qiagen kit with all values less than 10 ng/µl removed along with statistical outlier from *C. amplexicaulis* [N=36].

Table 2: Standard Deviation (SD) of DNA concentrations extracted per accession using Qiagen DNeasy Maxi Prep kit. (A) All samples included [n=45, mean=29.55 ng/µL], (B) All samples but two outliers removed from *C. amplexicaulis* [n=43, mean=25.84 ng/µL], (C) All samples larger than 10 ng/µL [n=37, mean=34.45 ng/µL], (D) All samples larger than 10 ng/µL, with one outlier removed [n=36, mean=32.04 ng/µL]. Where there is only a dash (-) present, there was only one sample and therefore standard deviation could not be calculated.

	A	**B**	**C**	**D**
Species	**SD**	**SD**	**SD**	**SD**
C. amplexicaulis	52.25	0.03	52.25	44.52
C. azalea	8.25	8.25	7.66	7.66
C. gaudichaudii	4.81	4.81	-	-
C. granthamiana	17.25	17.25	10.42	10.42
C. grijsii	3.90	3.90	3.90	3.90
C. japonica	17.00	17.00	17.00	17.00
C. japonica 'Kumasaka'	1.68	1.68	0.69	0.69
C. japoninca 'Maiden of Great Promise'	4.57	4.57	4.57	4.57
C. oleifera	6.80	6.80	-	-
C. sasanqua	7.50	7.50	7.50	7.50
C. sinensis	26.83	26.83	26.83	26.83
C. transarisanensis	28.06	28.06	24.45	24.45
Average Standard Deviation	25.73	19.27	25.86	21.64

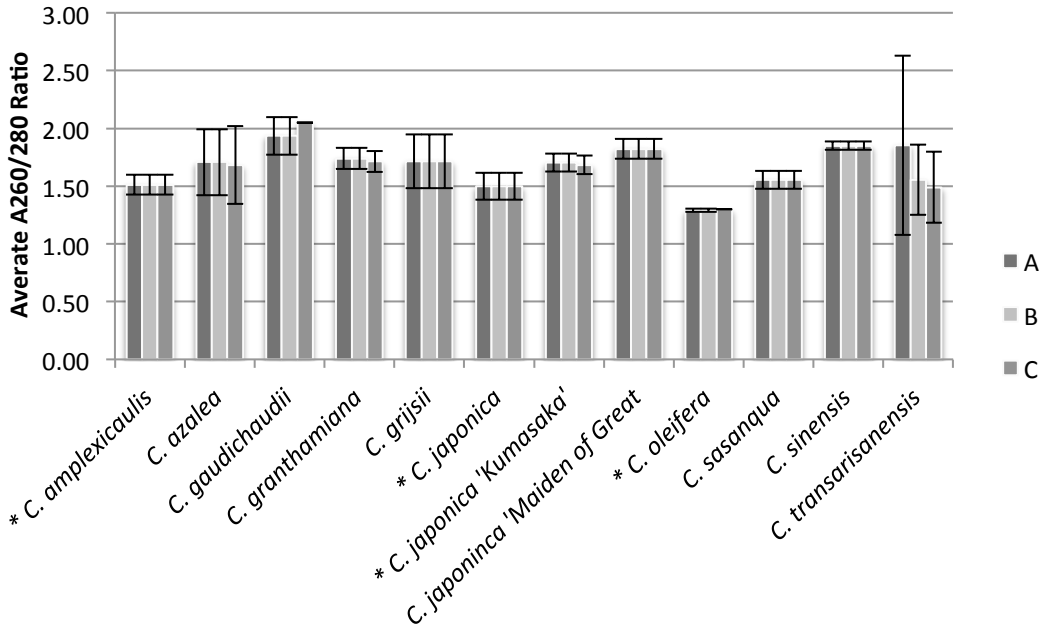


Figure 2: Average ratio of absorbance of light at 260nm and 280nm wavelength used to qualify DNA extracted using the Qiagen DNeasy Plant Maxi Prep Kit. (A) Average of all samples included [n=45, mean=1.70], (B) Average of all samples except the one outlier removed from *C. transarisanensis* [n=44, mean=1.66], (C) Average of all samples larger than 10 ng/µL [n=37, mean=1.65].

Table 3: Standard deviation (SD) of A260/280 per accession using Qiagen DNeasy Maxi Prep kit. (A) All samples included [n=45, mean=1.70], (B) All samples except the one outlier removed from *C. transarisanensis* [n=44, mean=1.66], (C) All samples larger than 10 ng/µL [n=37, mean=1.65]. Where there is only a dash (-) present, there was only one sample and therefore standard deviation could not be calculated.

	A	**B**	**C**
Species	**SD**	**SD**	**SD**
C. amplexicaulis	0.09	0.09	0.09
C. azalea	0.29	0.29	0.34
C. gaudichaudii	0.16	0.16	-
C. granthamiana	0.09	0.09	0.09
C. grijsii	0.23	0.23	0.23
C. japonica	0.12	0.12	0.12
C. japonica 'Kumasaka'	0.08	0.08	0.08
C. japoninca 'Maiden of Great Promise'	0.09	0.09	0.09
C. oleifera	0.01	0.01	-
C. sasanqua	0.08	0.08	0.08
C. sinensis	0.04	0.04	0.04
C. transarisanensis	0.78	0.30	0.31
Average Standard Deviation	0.33	0.22	0.22

Table 4: Average ratio of absorbance of light at 260nm and 280nm wavelength used to qualify DNA extracted using the Qiagen DNeasy Plant Maxi Prep Kit. (A) Average of all samples included [N=45], (B) Average of all samples larger than 10 ng/µL [N=37]. An asterisk (*) next to the average indicates it is significantly less than 1.8, the expected ratio of pure DNA.

	N = 45	N = 37
	Average of 260/280	Average of 260/280
C. amplexicaulis	1.51*	1.51*
C. azalea	1.71	1.68
C. gaudichaudii	1.94	2.05
C. granthamiana	1.74	1.71
C. grijsii	1.72	1.72
C. japonica	1.50*	1.50*
C. japonica 'Kumasaka'	1.70*	1.68
C. japoninca 'Maiden of Great Promise'	1.82	1.82
C. oleifera	1.29*	1.30*
C. sasanqua	1.56	1.56
C. sinensis	1.85	1.85
C. transarisanensis	1.56	1.49
Average	1.66	1.65

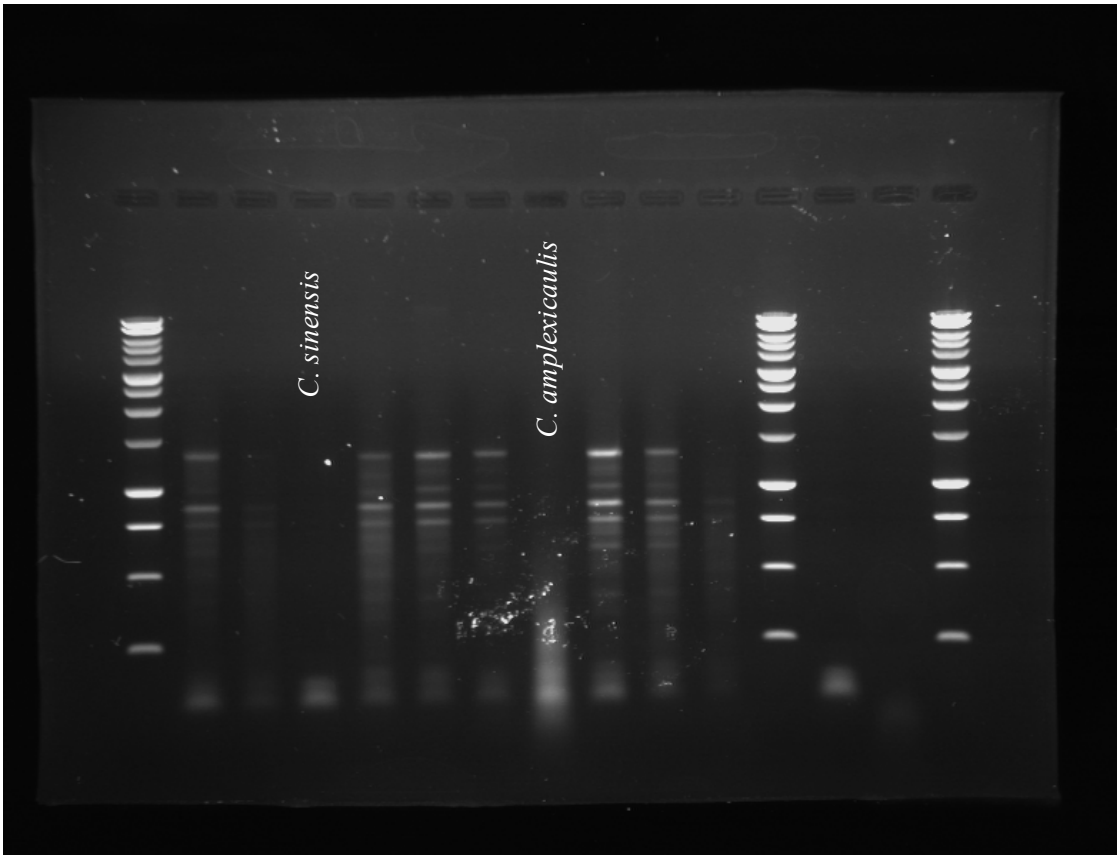


Figure 3: Degradation of RNA samples from *C. sinensis* and *C. amplexicaulis*. Samples are run on 2% agarose gel post treatment with DNase.

Table 5: Quantity (ng/µl) and quality (absorbance ratio of light at 260nm and 280nm wavelength) of RNA extracted from 10 camellia species. An A260/280 ratio of around 2.0 indicates pure RNA, no samples were statistically different from 2.0.

Species	ng/ul	A260	A280	260/280
C. amplexicaulis	40.25	1.006	0.511	1.97
C. azalea	210.48	5.262	2.462	2.14
C. gaudichaudii	110.81	2.77	1.307	2.12
C. granthamania	70.32	1.758	0.814	2.16
C. grijssii	305.81	7.645	3.673	2.08
C. japonica	199.16	4.979	2.35	2.12
C. oleifera	77.52	1.938	0.933	2.08
C. sasanqua	242.58	6.064	2.857	2.12
C. sinensis	125.45	3.136	1.486	2.11
C. transarisensis	301.88	7.547	3.531	2.14

Table 6: List of genes sequenced. The first column contains the gene code from NCBI's GeneBank. Column two contains a short description of the gene, and column 3 specifies if the GeneBank sequence originated from mRNA or DNA.

Gene Code (GenBank)	Short Description	Source
AF315492	5.8s Ribo	DNA
AY169403	Chalcone synthase	mRNA
AY787658	Histone 3	mRNA
DQ366599	Acetyl-CoA	mRNA
DQ887753	Pollen protein	mRNA
EU195885	lox1	mRNA
GQ214317	PPO	DNA
GQ461358	ELIP1	DNA
GU393024	ERF	mRNA
JQ398741	APETALA2	mRNA
JX657332	APL1	mRNA
KC149523	FL_T	mRNA

Table 7: Total length of sequenced region per accession by gene. All lengths are measured in base pairs (bp). See Appendix B for key to species abbreviations.

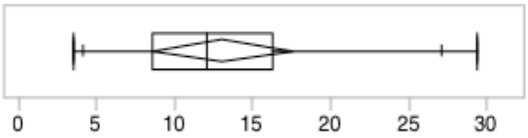
Gene	AMP	AZA	GAUD	GRA	GRI	JAP
AF315492	281	180	210	227	284	212
AY169403	496	673	615	647	623	662
AY787658	512	493	489	473	502	512
DQ366599	348	423	631	431	639	412
DQ887753	891	887	903	908	914	886
EU195885	536	735	465	352	352	536
GQ214317	380	379	395	379	374	419
GQ461358	286	266	335	278	255	278
GU399024	676	698	698	676	701	680
JQ398741	386	386	353	357	319	386
JX657332	198	200	200	205	295	198
KC149523	244	313	259	237	292	176
TOTAL	5234	5633	5553	5170	5550	5357

Gene	KAM	MAID	OLE	SAS	SIN	TRA	CONTIG
AF315492	224	220	221	210	311	215	196
AY169403	604	621	665	677	656	659	642
AY787658	512	463	507	512	512	512	512
DQ366599	419	419	640	385	735	683	431
DQ887753	911	941	891	929	909	912	887
EU195885	790	499	513	790	523	735	799
GQ214317	459	459	459	459	463	459	462
GQ461358	328	241	275	298	278	275	304
GU399024	698	698	701	698	701	699	701
JQ398741	386	365	267	277	395	387	398
JX657332	104	198	200	156	200	200	200
KC149523	219	176	300	300	313	228	329
TOTAL	5654	5300	5639	5691	5996	5964	

A

	ALL	Outlier Removed
MEAN	13.03	11.54
MAX	29.44	21.61
Q3	14.50	13.20
MEDIAN	12.01	11.27
Q1	9.79	9.20
MIN	3.50	3.50
Std.Dev	7.08	5.08

B



C

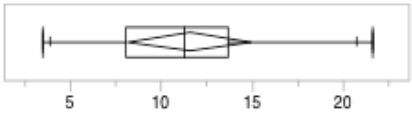


Figure 4: (A) Descriptive statistics for number of polymorphisms identified across all genes sequenced and camellia accessions. One outlier was identified then removed, and all statistics were re calculated. (B) Box and whisker plot of data with outlier present. (C) Box and whisker plot with outlier removed.

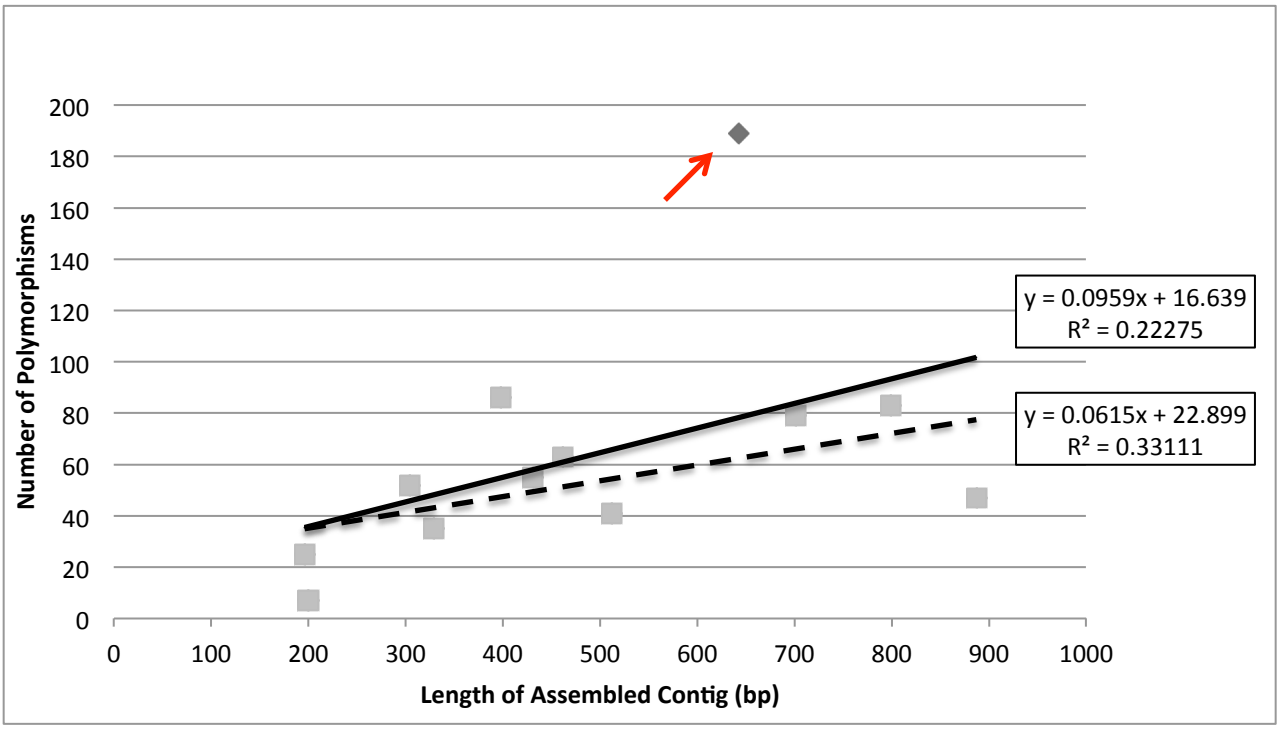


Figure 5: Graph of contig length (bp) vs. the number of polymporhisms identified. The solid black linear regression line shows the regression for the whole data set. The dash line shows the positive linear regression when the outlier (indicated by red arrow) is removed.

Table 8: Number of polymorphisms identified between camellia taxa. Numbers are color coated to from fewest differences present (white) to most differences present (red) between any two taxa.

	AMP	AZA	GAU	GRA	GRI	JAP	KAM	MAI	OLE	SAS	SIN	TRA
AMP												
AZA	311											
GAUD	309	183										
GRA	355	231	167									
GRI	374	245	213	180								
JAP	332	216	161	224	241							
KAM	293	266	178	284	332	228						
MAID	336	213	95	201	245	151	145					
OLE	334	202	135	174	221	186	238	155				
SAS	385	233	215	236	253	217	285	237	160			
SIN	315	223	204	242	253	246	306	234	225	289		
TRA	353	222	213	223	269	231	275	223	196	244	241	

n	66
Min	95.00
Q1	206.25
Mean	239.42
Q3	273.50
Max	385.00
Std.Dev	60.36

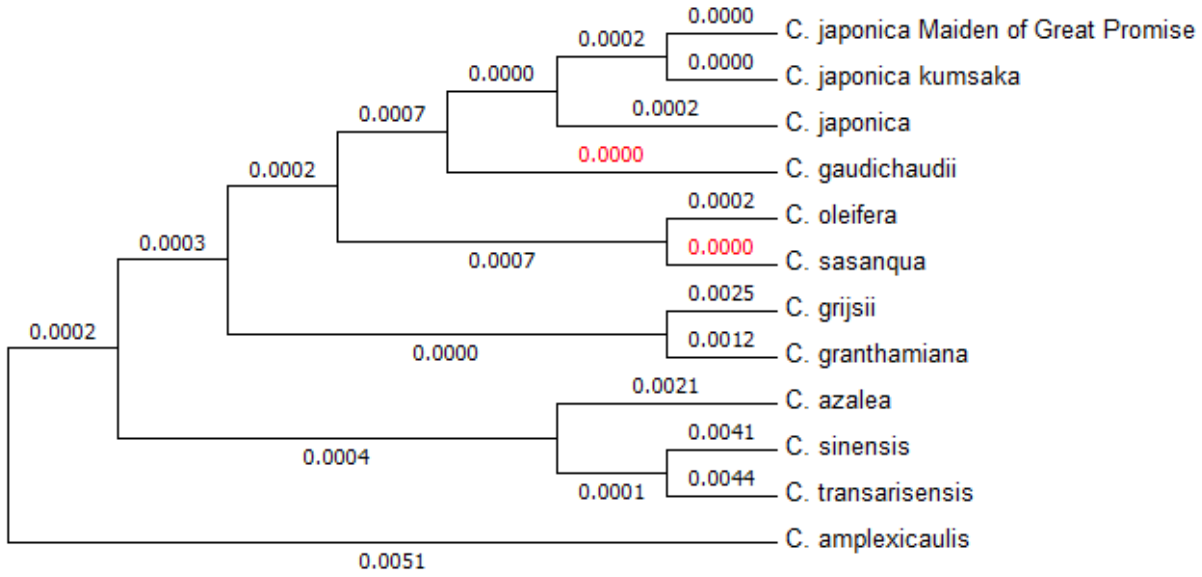


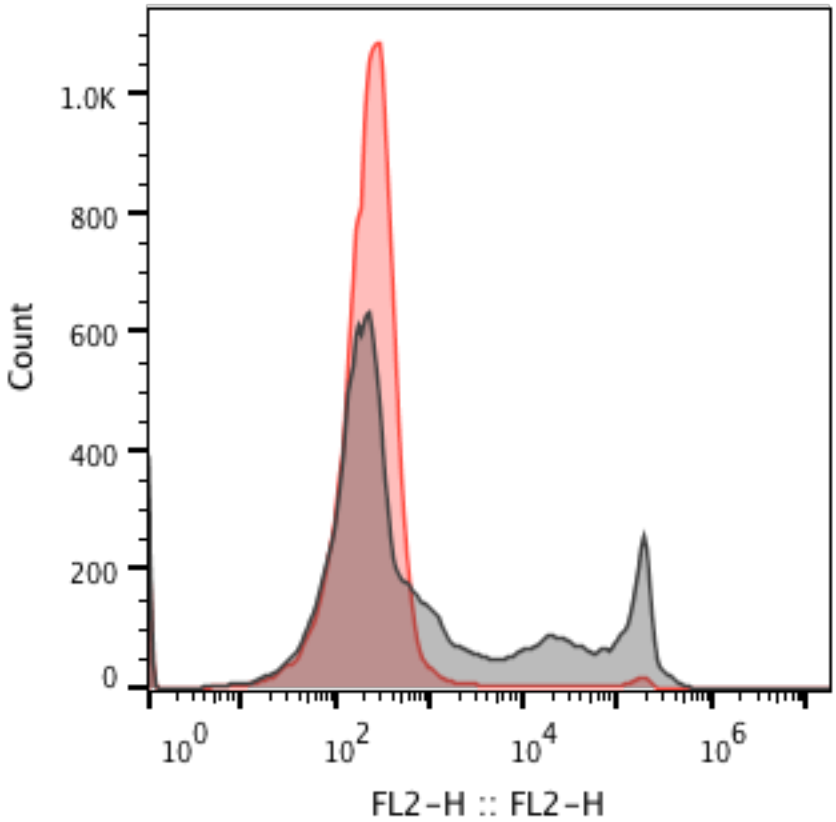
Figure 6: Evolutionary history of *C.* accessions inferred using the Neighbor-Joining Method. Evolutionary distances were computed using the Tamura-Nei method and are in the units of the number of base substitutions per site. Distances are placed next to their corresponding branch.

Table 9: Number of base pairs (both from single and paired reads) sequenced through Next Generation RNA Sequencing for ten camellia taxa. These reads were then assembled into a number of transcripts, which represent a large number of genes.

Accession	**Normalized**	**Single Read**	**Paired Read**	**Genes**	**Transcripts**
C. amplexicaulis	Y	1,873,224	24,921,664	33,592	118,467
C. azalea	Y	2,660,511	36,585,012	39,488	195,322
C. gaudicahuadii	Y	1,600,244	17,499,290	28,258	99,680
C. granthamiana	Y	1,002,020	13,391,572	34,562	111,309
C. grijsii	Y	2,008,635	25,605,822	37,074	171,338
C. grijsii	N	5,985,140	25,995,738	33,551	142,707
C. japonica	Y	1,588,398	19,835,314	34,602	117,150
C. oleifera	Y	1,213,438	16,520,210	26,805	104,372
C. sasanqua	Y	1,353,689	17,367,846	37,373	177,167
C. sasanqua	N	21,367,022	75,334,034	51,451	329,420
C. sinensis	Y	1,632,461	19,621,528	37,878	174,544
C. transarisanensis	Y	1,309,724	19,494,856	36,785	161,825
C. transarisanensis	N	9,052,433	66,358,968	46,269	156,377

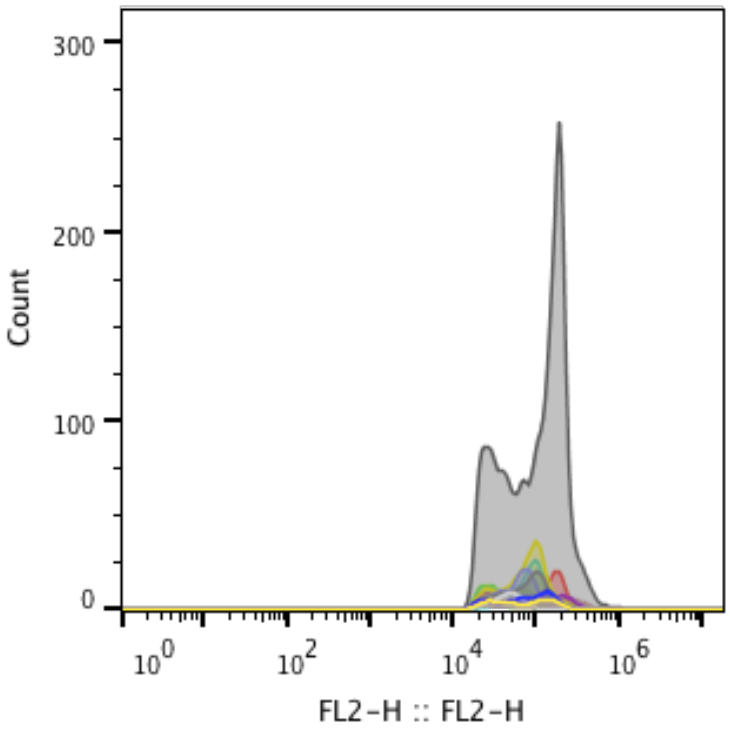
Table 10: Mean and standard deviation of detected emission for each camellia accession through flow cytometry. Samples were treated with PI stain and observed in the 545-625nm light spectrum.

	N	**Mean (FL2-H)**	**SD (FL2-H)**
C. sinensis	25000	2090	17742
C. sinensis var.assamica	25000	23730	64059
C. japonica	25000	1964	14488
C. azalea	25000	1169	7384
C. sasanqua	25000	847	12215
C. amplexicaulis	25000	1336	16021
C. transarisensis	25000	1812	14471
C. granthamania	25000	578	4387
C. japonica 'Longwood Centennial'	25000	2657	16073
C. japonica 'Longwood Valentine'	25000	1331	9352
C. japonica 'Tama-no-ura'	25000	1824	14986
C. japonica 'Lady de Saumarez'	25000	586	4022
C. 'Alice du Pont'	25000	1291	11088
C. japonica 'Dr. Tinsely'	25000	1151	8944
Mean		3026	15374
Standard Deviation		5988	14670



	Sample Name	Subset Name	Count
	B02 assamica-F.fcs	Ungated	25000
	B01 sinensis -F.fcs	Ungated	25000

Figure 7: Fluorescence emission of C. sinensis (red) and C. sinensis var. assamica (gray).



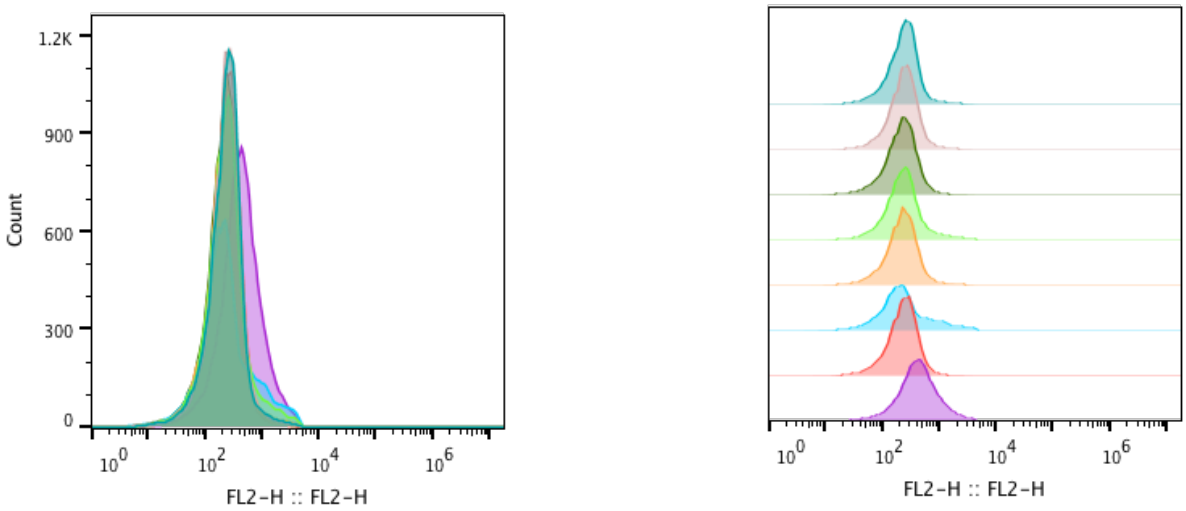
	Sample Name	Subset Name	Count
	E06 Dr. Tin-F.fcs	Peak 3	190
	E05 Alice-F.fcs	Peak 3	261
	E04 Lady de-F.fcs	Peak 3	161
	E03 Tama-no-ura-F.fcs	Peak 3	434
	B02 assamica-F.fcs	Peak 3	4656
	E02 Valentine-F.fcs	Peak 3	387
	E01 Centennial-F.fcs	Peak 3	681
	B08 Granth-F.fcs	Peak 3	107
	B07 Transar-F.fcs	Peak 3	284
	B06 amplex-F.fcs	Peak 3	165
	B05 sasanqua-F.fcs	Peak 3	117
	B04 azalea-F.fcs	Peak 3	326
	B03 japonica-F.fcs	Peak 3	483
	B01 sinensis -F.fcs	Peak 3	382

Figure 8: Upper fluorescence regions, representing cells actively going through the G_2 stage of cell division, of all camellia accessions. *C. sinensis var. assamica* (gray) is noticeably larger and was classified as an outlier. Table indicates number of nuclei that (N=25,000) fell within this range.

Table 11: The mean, median and standard deviation (SD) of the detected fluorescence of each taxa when passed analyzed on an Accuri C6 flow cytometer. Half of the samples were run directly (unfiltered), while the other half of the sample was passed through a cell separator prior to analysis (filtered).

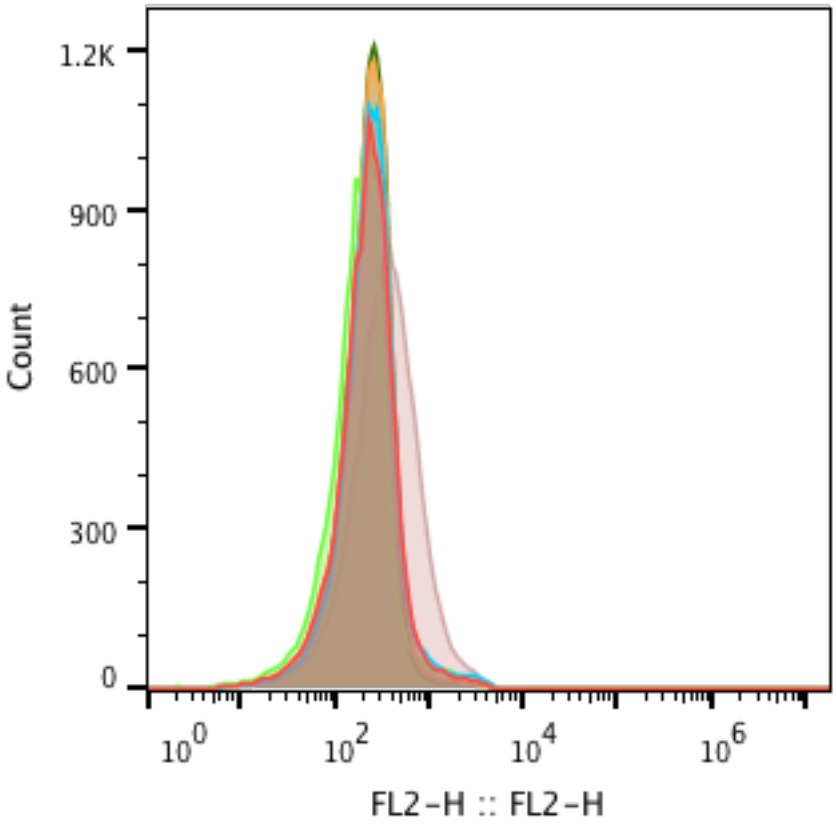
UNFILTERED					
	% of Total	Count	Mean (FL2-H)	Median (FL2-H)	SD (FL2-H)
C. sinensis	94.2	23560	215	176	259
C. japonica	91.1	22776	237	158	403
C. azalea	94.8	23700	293	209	389
C. sasanqua	90.9	22735	401	266	491
C. amplexicaulis	96.2	24047	282	224	312
C. transarisensis	93	23244	701	525	632
C. granthamania	96.8	24198	379	294	410
C. japonica 'Longwood Centennial'	94.2	23545	267	207	346
C. japonica 'Longwood Valentine'	95.9	23982	292	234	346
C. japonica 'Tama-no-ura'	94.6	23661	279	208	351
C. japonica 'Lady de Saumarez'	95.3	23818	239	207	222
C. 'Alice du Pont'	96.3	24065	323	253	350
C. japonica 'Dr. Tinsely'	94.2	23538	518	352	554
Mean	94.4	23605	340	255	390
SD	1.85	460	136	95.7	114

FILTERED					
	% of Total	Count	Mean (FL2-H)	Median (FL2-H)	SD (FL2-H)
C. sinensis	95.9	23964	269	228	256
C. japonica	94.7	23683	284	220	335
C. azalea	94.8	23691	337	224	462
C. sasanqua	96.6	24149	261	215	271
C. amplexicaulis	97	24249	287	233	301
C. transarisensis	94.7	23664	566	406	558
C. granthamania	96.8	24189	290	234	329
C. japonica 'Longwood Centennial'	94.1	23521	304	229	380
C. japonica 'Longwood Valentine'	96	24009	265	225	280
C. japonica 'Tama-no-ura'	93.6	23407	249	180	356
C. japonica 'Lady de Saumarez'	97	24244	255	222	254
C. 'Alice du Pont'	96.1	24020	294	230	334
C. japonica 'Dr. Tinsely'	95.4	23845	454	337	447
Mean	95.6	23895	317	245	351
SD	1.13	282	91.8	59.5	90.7



	Sample Name	Subset Name	Count
	B08 Granth-F.fcs	Gate 1	24189
	B06 amplex-F.fcs	Gate 1	24249
	B04 azalea-F.fcs	Gate 1	23691
	B03 japonica-F.fcs	Gate 1	23683
	B02 assamica-F.fcs	Gate 1	18251
	B01 sinensis -F.fcs	Gate 1	23964
	B07 Transar-F.fcs	Gate 1	23664
	B05 sasanqua-F.fcs	Gate 1	24149

Figure 9: All species accessions' fluorescence patterns. On the left, all accessions are overlapped, showing that only *C. transarisanensis* (purple) was significantly different. Graph on the right shows same samples, but graphs are offset vertically



	Sample Name	Subset Name	Count	Mean : FL2-H	SD : FL2-H
	B03 japonica-F.fcs	Gate 1	23683	284	335
	E01 Centennial-F.fcs	Gate 1	23521	304	380
	E02 Valentine-F.fcs	Gate 1	24009	265	280
	E03 Tama-no-ura-F.fcs	Gate 1	23407	249	356
	E04 Lady de-F.fcs	Gate 1	24244	255	254
	E06 Dr. Tin-F.fcs	Gate 1	23845	454	447

Figure 10: All *C. japonica* cultivars fluorescences. *C. japonica* 'Dr. Tinsely' (pink) shows the only distinct difference, however it is not statistically significant.

Chapter 5

DISCUSSION

5.1 DNA Analysis Discussion

5.1.1 Camellia DNA Extraction

The concentration of DNA extracted from accessions was much lower than previously published literature, which reported concentrations upwards of 125 ng/µL from only 0.3g of tissue (Umar, 2011). Since the extraction of DNA can be complicated by secondary metabolites, previous studies have been focusing on reducing the negative effects of these metabolites, and have developed more involved protocols for the extraction of DNA. This research, however, uses a much simpler kit, which comes with all materials necessary for extracting DNA. These kits may not be as efficient as more customized protocols, but do extract high enough concentrations of DNA for further analysis through PCR, and are much easier to use. *C. amplexicaulis* was the most variable in the amount of DNA extracted, as seen by the large standard deviation (Table 2A). However, once the outliers were removed, this standard deviation drastically decreased. The average concentration of DNA extracted by *C. transarisanensis*, however, was most effected by the removal of values less than 10 ng/µL. Though it was possible to extract good quality DNA in sufficient quantities

from both of these species, it shows that care must be taken with each extraction and that the protocol must be followed carefully to ensure consistency of extractions.The amount of DNA extracted from *C. grijssii, C. japonica, C. sasanqua,* and *C. sinensis* were all very consistent with standard deviations less than 8 ng/µL, no values less than the 10 ng/µL threshold, and no outliers present. This can bee seen in Figure 1 as the average concentration of DNA for each of these species does not change.

Previous studies have also reported higher ratios of absorbance (A260/280) than samples prepared for this research (Umar, 2011). Though samples used had a ratio statistically lower than the expected 1.8 (Table 4), all samples did amplify, and ultimately a phylogenetic tree could be created. The protocol from the Qiagen DNeasy kit could be modified in hopes of decreasing protein contamination, and increasing the quality of DNA for sequencing. The Qiagen Plant Mini protocols incudes an optional centrifugation step after the Buffer P3 is added. This was not tried with any samples, but could potentially increase the quantity and quality of DNA extracted from camellias. Alternatively more Buffer P3 could be added in hopes of precipitating more protein. Ultimately though, these recommendations may not be needed as pure DNA is not a necessity as shown in this research.

Only about 50% of primers created successfully amplified genomic regions across all accessions used. One benefit of extracting higher quality genomic DNA is that it may increase the number of primers that amplify regions across all accessions. It could be that other proteins and molecules are preventing the successful annealing of primers to template DNA. More recommendations for improving the compatibility

of primers are included in the following section. The low rate of primer success though is more likely due to the high rate of polymorphism within the genus. Primers were designed to be sequence dependent, and therefore can only bind to DNA segments with complimentary base pairs. Changes in nucleotide sequences, especially in the 5' region of the primer-binding site, can cause the primer to not anneal and therefore no replication will occur during PCR.

5.1.2 DNA Phylogeny

The Neighbor-joining (NJ) phylogenetic tree was selected because samples were relatively similar and there is no known evolutionary rate for camellias, thus ruling out the Unweighted Pair Group method with Arithmetic Mean (UPGMA). To verify results a Maximum Likelihood cladastic tree was also created and there were no differences noted between the two models. The phylogenetic tree that was created (Figure 4) does differ to some degree from previously published phylogenetic analysis of camellia. The two *C. japonica* cultivars did group closely with *C. japonica,* which is to be expected, and *C. gaudichaudii, C. sasanqua, C. oleifera, C. grijsii*, and *C. granthamiana* all grouped closer with *C. japonica,* as they are all in *Camellia* subg. *Camellia*, while *C. sinensis* and *C. transarianensis* grouped more closely together, which agrees with previous publications (Huang, Tong, Zhang, & Gao, 2013; Vijayan, Zhang, & Tsou, 2009). This research further affirms the findings of previously published phylogenetic trees that *C. sasanqua* and *C. oleifera* are closely related. Several key differences were present in the phylogenetic tree created here and with

other previously published attempts at understanding the genetic relationships within *Camellia.*

Results published by Huang, Tong, Zhang, and Gao (2013), the researchers grouped *C. grijsii* closer with *C. oleifera* and *C. sasanqua*, in *Camellia* sect, *Paracamellia*. Other studies however have shown that *C. grijsii* as actually more closely related to *C. granthamania* (Vijayan, Zhang, & Tsou, 2009). In the Collected Species of the Genus Camellia (Jiyin, Parks, & Yueqiang, 2005), *C. grijsii* is placed in *Camellia* sect. *Paracamellia*, while *C. granthamiana* is treated as a member of *Camellia* sect. *Protocamellia* and *C. olifera* is in the newly formed *Camellia* sect. *Olifera*, which was separated from section *Paracamellia*. The Flora of China (Min et al., 2010), treat *C. grijsii* and *C. oleifera* both as members of *Camellia* sect. *Paracamellia*, and *C. granthamiana* is listed in *Camellia* sect. *Heterogenea*. This research showed that *C. grijsii* was the most distantly related to all other accessions within the subgenus *Camellia*, and *C. granthamiana* was the closest relative with an evolutionary distance of 0.0037 base substitutions per site.

Camellia gaudichaudii has been shown to be evolutionarily close to *C. granthamiana;* both in *Camellia* sect. *Heterogenea* (Huang, Tong, Zhang, & Gao, 2013). This is in agreement with what is published in The Flora of China (Min et al., 2010). Alternatively the Collected Species of the Genus Camellia (2005) places *C. guadichaudii* in *Camellia* sect*. Furfuracea*, which is not recognized by The Flora of China. This research said it was most closely related to *C. japonica* (section *Camellia*). Though section *Camellia* and section *Heterogenea* are both in the subgenus

Camellia, pervious research suggests that section *Camellia* and section *Heterogenea* are the most distantly related within the subgenus (Huang, Tong, Zhang, & Gao, 2013).

One of the most prominent differences is the treatment of *C. azalea.* Past studies have grouped *C. azalea* closer to *C. japonica* (Huang, Tong, Zhang, & Gao, 2013; Vijayan, Zhang, & Tsou, 2009), and other treatments of genus *Camellia* agree that these two species are both in *Camellia* sect. *Camellia* (Jiyin, Parks, & Yueqiang, 2005;& Min, 2010). This research however placed *C. azalea* closer with *C. sinensis*, which is in a different subgenus as the other previously mentioned camellia sections. Since all previously published findings agree that *C. azalea* is closely related to *C. japonica*, this suggests there may have been an error in the data presented here. The plant from which these samples were collected is taxonomically identified as *C. azalea*, and since this species is phenotypically very unique, it is unlikely that the plant was misidentified. If this research is to continue, these data should be removed and the samples should be sequenced again and analyzed anew.

C. amplexicaulis was identified as the outgroup and was not closely related to any of the species used in this research. Interestingly, *C. amplexicaulis* The Flora of China (Min et al., 2010) does not list the species, though the Collected Species of the Genus Camellia does list the species as native to China, and list it as a member of *Camellia* sect. *Archecamellia*. *Camellia* sect. *Chrysantha* is sometimes grouped into section *Archecamellia* (Jiyin, Parks, & Yueqiang, 2005, p. 23), and it appears that The Flora of China does group these sections together. Included in section *Archecamellia*

is *C. chrysanthoides* and *C. flavida* (Min et al., 2010). These two species are both included in section *Chrysantha* in the Collected Species of the Genus Camellia (Jiyin, Parks, & Yueqiang, 2005). Recent publications, which include other species from *Camellia* sect. *Archecamellia*, have concluded that the section is in *Camellia* subg. *Thea*, but most distantly related within the subgenus (Huang, Tong, Zhang, & Gao, 2013). This would then suggest that *C. amplexicaulis* is more closely related to *C. sinensis* and *C. transarisanensis* than other species used in this research. This hypothesis is further confirmed by papers that do include *C. amplexicaulis* and place it close to *C. transarisanensis* (Vijayan, Zhang, & Tsou, 2009). This research supports the idea that *C. amplexicaulis* is more distantly related to all of the other selected accessions, but since there were no accessions from *Camellia* sect. *Chrysantha*, these results cannot conjecture to the relationship of section *Chrysantha* and section *Archecamellia*. The phylogenetic tree presented here, however places *C. amplexicaulis* outside of both subgenera, into its completely own section. If more accessions known to be from *Camellia* subg. *Thea* were included in this study though, it may resolve this problem.

The two *Camellia japonica* cultivars, 'Maiden of Great Promise' and 'Kumaska', were included to investigate the possibility that they were clones. This research showed that they were very similar, and essentially no evolution has occurred between the two species. However, there were some discrepancies between the sequences, and therefore it cannot be confirmed that they are clones. To definitively state that these two cultivars are in fact the same selection released under different

names would ultimately require the complete sequencing of their genomes. Though this is not possible at this time, it can be stated that these two cultivars display similar phenotypic characteristics, and genetically are very similar.

Overall, this research affirms the divisions of the subgenera within the genus *Camellia*, though it questions the distinctions between different sections. Species like *C. gaudichaudii* and *C. grijsii* were both found in different sections of the subgenus *Camellia*, than suggestedd by previous publications. The phylogenetic tree presented here also suggests that accessions in the subgenus *Camellia* are overall more closely related to each other, with a maximum evolutionary distance of 0.0037 base substitutions per site, compared to accessions within the subgenus *Thea*, maximum evolutionary distance of 0.0085 base substitutions per site.

This phylogenetic tree presents some information that could immediately be used by camellia breeders when setting up new hybridizations. *C. amplexicaulis* is the most distantly related species from *C. japonica* and therefore could potentially create the most diverse set of seedlings if crossed with *C. japonica.* Increased diversity between parents often leads to increased diversity among the progeny, and therefore a greater chancer of discovering distinctly different and new cultivar. Longwood Gardens has already begun to hybridize *C. japonica* with *C. amplexicaulis*, and the data presented here supports their use of *C. amplexicualis* in more crosses.

5.1.3 Recommendations for Future Work

Moving forward this research could be expanded to better understand how different species are related to each other. If Longwood Gardens does decide to do so, then it is recommended to increase the number of primers used in sequencing, and to possibly further prefect the DNA extraction protocol. Currently, Longwood Gardens does not have the facilities to do this work and it would either have to be done through partnerships with outside organizations, or the garden would have to invest in new lab equipment. Extraction and isolation of DNA, PCR replication of genomic regions, and PCR quality control could all be performed at Longwood, as long as the proper equipment is purchased. The technology used for analysis was mostly free, except for Sequencher™ DNA analysis software, which would have to be licensed. University of Delaware does have a limited license for the software, though it can only be used in specific locations on the Newark, Delaware campus. Longwood Gardens would have to invest in Sequencher™, or a similar program, in order to assemble and analyze sequences.

High-throughput DNA sequencing technologies are becoming more affordable, though still represent a large financial and intellectual capital investment. The purchasing of equipment and training of staff is not reasonable some organization such as Longwood Gardens, which would only get minimal use from the technology. Many educational institutions, such as the University of Delaware, offer sequencing services, which would be the most practical option for Longwood Gardens if they wish to continue this. Samples would be prepared at the garden, and then submitted to

a sequencing facility, and then results could be downloaded after sequencing was complete.

If more primers are to be developed for camellias at Longwood Gardens, it may be best to focus on more conserved regions of the genome, such as histone 3 or 40S ribosomal subunit, as these may have an increased chance of amplifying across all accessions. Conserved regions of the genome have a lower incident of mutation, and therefore the issues discussed above with sequence-specific primer binding would become less of a problem. Once a larger subset of primers have been selected, it would be relatively easy to increase the number of accessions used in sequencing. Sequencing more *C. japonica* cultivars could provide deeper insight into the species, and could help to further confirm or deny the allegation that the cultivars 'Maiden of Great Promise' and 'Kumasaka' are related. If more cultivars are included and more primers are developed in this research and these two cultivars continue to show a high degree of similarity, and exceptionally low level of evolutionary distance than it will further suggest that they may be indeed the same selection.

Continuing this research would place Longwood Gardens at the forefront of ornamental camellia research though may not have the most immediate impacts on the breeding program. This research affirmed the previously published results, except for the results derived from *C. azalea* that should be repeated, and so a close study of the literature could help elucidate phylogenetic relationships between different accessions without performing the experiments.

5.2 – RNA Analysis Discussion

5.2.1 RNA Extraction

Next generation RNA sequencing (RNA-seq) is still a relatively new and growing field of genetic research. At the time of this research only a few papers had been published using RNA-seq with camellia, and all of those have focused almost exclusively on *C. sinensis*. The RNA extracted from camellias during this research, using the CTAB method, is of comparable quantity and quality to other published articles (Muoki, 2011;& Yockteng, 2014). With an A260/280 near 2.0, there were no issues with contamination, and all sequences were easily cleaned of DNA through a DNase digest. It is essential to keep all samples as cold as possible at all times to ensure they do not degrade, and it is best to limit the time between extraction and sample sequencing to less than3 months. Overall, the protocol for extracting RNA takes longer than the protocol call to extract DNA, the results were more consistent and the product was of higher quality.

5.2.2 RNA Phylogeny

The large amount of data created through RNA-seq is exponentially larger than that derived from DNA sequencing. This therefore leads to a problem of data management, and data processing. Though there is a large amount of transcripts sequenced, not all of them are useful for various reasons. Therefore the challenge to parse out useful subsets of data is quite large. Some sets of transcripts that were similar to genes sequenced through DNA were able to be seperated, however this data

still required much further analysis if a phylogenetic tree was to be created. Ultimately a new phylogenetic tree could be created with these transcripts along with the many other transcripts that align across all species selections.

5.2.3 Recommendations for Future Work

The use of next generation RNA sequencing is unlikely to be utilized by Longwood Gardens in the future to better their breeding program. Of all experiments conducted during this research, RNA analysis took the most time and required the highest financial input. This would not have been possible without the many highly trained professionals who helped, specialized equipment, and high-powered computers used for data analysis.

There is a significant amount of data created from this research that is yet to be analyzed that could still provide interesting results. The large number of transcripts that have a large overlap (over 1,000 bp), could still be used to create a new phylogenetic tree that could then be used to compare to the results derived from the DNA research. If this is done then it may give better insight into how the species are related, and may also provide additional information about the abnormal treatment of *C. azalea* in the DNA Phylogenetic tree. Longwood Gardens would have to work closely with the researchers at DuPont Pioneer to continue the analysis of the RNA-seq data, as it requires advance bioinformatics.

5.3 Flow Cytometry Discussion

Due to time restraints, a control of known genome size, such as corn or pea, could not be grown and included with samples analyzed by flow cytometry. If a control were included it would allow for the estimation of each accessions genomic size, but since it was not included all data gained from flow cytometry was only comparative. Previously published papers include *Pisum sativum* (pea) as an internal standard to gauge (Contreras, 2011 and Hanson, 2001). Pea is a good standard as it has been heavily studied adn the genome size of several different cultivars has been published, furthermore it is easy to grow and only a few leaves would be needed for these experiments.

Almost all papers that report on *Camellia* genome size, or chromosome number, report a high amount of variability within the genus, and occasionally within the species. This research, though, shows very little variability between selected accessions, and *C. japonica* cultivars. Figure 9 shows that all accessions were comparatively very similar in genome size, and only *C. transarisanensis*, showed significant differences from the rest of the group. This species is rarely studied, and therefore there only are few previously published reports of its genome size. Though the results published here do not suggest a full doubling of its genome (2x to 4x), it does suggest that some event has occurred to enlarge the genome. *C. transarisanensis* (syn. *C. handelii*), along with other members of the *Camellia* sect. *Theopsis*, are known for their smaller leaves and fragrant flowers, and thus could be good a candidate for future breeding efforts (Jiyin, Parks, & Yueqiang, p. 262). The

difference in genome size, however, could prove to make it incompatible with at least the other accessions analyzed in this research.

5.3.1 Recommendations for Future Work

Moving forward, flow cytometry shows the highest opportunity to contribute significantly to Longwood Gardens breeding program. The kit purchased from Partec® is easy to use, and fairly inexpensive. Furthermore, small bench top flow cytometers are becoming increasingly affordable. If Longwood Gardens did not want to purchase a flow cytometer, an Accuri™ C6 flow cytometer is available for use at the University of Delaware's Center for Translocational Cancer Research (CTCR). Currently the equipment at CTCR requires the use of Propidium Iodide stain, which is considered a potential carcinogen, so extra care must be taken when preparing samples. However, with minor modifications, the Accuri™ C6 would be able to detect the emissions from 4',6-diamidino-2-phenylindole (DAPI) stain, which would allow for the use of Partec®'s DAPI kit (#05-5002). This kit is easier to use, and contains chemicals that are considered less dangerous.

If Longwood Gardens' research staff does decided to continue forward with the use of flow cytometry, it would be best to start including an internal standard, such as *Pisum sativum*. This would allow for a definitive, opposed to relative, estimate of genome size to be calculated. With Longwood's research greenhouse space, pea seeds could easily be grown and new leaves could be collected when other samples are collected. Researchers could also expand the number of accessions studied and due to

the ease of flow cytometry could easily determine the ploidy of all new camellia accessions acquired by the garden.

REFERENCES

1. Ackerman, W. L. (1971). *Genetic and cytological studies with camellia and related genera*. Washinton, D.C.: U.S. Agricultural Research Service, U.S. Dept. of Agriculture.

2. Ackerman, W. L. (2002). *Growing camellias in cold climates*. Baltimore, MD: Noble House.

3. Ackerman, W. L. (2007). *Beyond the camellia belt*. Batavia, IL, USA: Ball Publishign.

4. Acquaah, G. (2012). *Principles of plant genetics and breeding* (2nd ed.). West Sussex, UK: Wiley-Blackwell.

5. Aiello, A. (2009). Seeking cold-hardy camellias. *Arnolida, 67*(1), 20-30.

6. Brown, J., & Caligari, P. (2008). *An introduction to plant breeding*. Oxford, UK: Blackwell Publishing.

7. Caser, M., Torello, M. D., & Scariot, V. (2010). Microsatellite-based genetic relationships in the genus camellia: Potential for improving cultivars. *Genome, 53*, 384-399.

8. Dolezel, J., & Bartos, J. (2005). Plant DNA flow cytometry and estimation of nuclear genome size. *Annals of Botany, 95*(1), 99-110. doi:10.1093/aob/mci005

9. Durrant, T. (1982). *The camellia story*. London, UK: Heinemann.

10. Huang, H., Tong, Y., Zhang, Q., & Gao, L. (2013). Genome size variation among and within camellia species by using flow cytometric analysis. *PLOS One, 8*(5), e64981. doi:10.1371/journal.pone.0064981

11. Hume, H. H. (1955). *Camellias in america* (Revised ed.). Harrisburg, PA: J. Horace McFarland Co.

12. Jiyin, G., Parks, C. R., & Yueqiang, D. (2005). *Collected species of the genus camellia: An illustrated outline*. People's Republic of China: Zhejiang Science and Technology Publishing House.

13. Kondo, K. (1977). Chromosome numbers in camellia. *Biotropica, 9*(2), 86-94.

14. Min, T., Wu, Z., Li, D., Hong, D., Zhang, X., & et al. (2010). *Flora of china* Science Press.

15. Missouri Botanical Garden.Tropicos. Retrieved, 2014, Retrieved from http://www.tropicos.org

16. Muoki, R. C., Paul, A., Kumari, A., Singh, K., & Kumar, S. (2012). An improved protocol for the isolation of RNA from roots of tea (*camellia sinensis* (L.) O. kuntze). *Molecular Biotechnology, 52*, 82-88. doi:10.1007/s12033-011-9476-5

17. National Center for Biotechnology Information, U.S. National Library of Medicine. (2013). Genbank®. Retrieved, 2014, Retrieved from http://www.ncbi.nlm.nih.gov/genbank/

18. Pierce, B. A. (2012). *Genetics: A conceptual approach* (4th ed.). New York, NY: W.H. Freeman and Company.

19. The plant list (2013). (Version 1.1). Retrieved, 2014, Retrieved from http://www.theplantlist.org/

20. Thro, A. M., Parrott, W., Udall J.A., & Beavis W.D. (2004). Genomics and plant breeding. *Crop Science, 44*(6), 1893.

21. Trehane, J. (1998). *Camellias*. Porland, OR: Timber Press.

22. Vijayan, K., Zhang, W., & Tsou, C. (2009). Molecular taxonomy of camellia (theaceae) infered from nrITS sequences. *American Journal of Botany, 96*(7), 1348.

23. Yockteng, R., Almeida, A. M. R., Yee, S., Andre, T., Hill, C., & Specht, C. (2014). A method for extractiong high-quality RNA from diverse plants for next-generation sequencing and gene expression analyses. *Applications in Plant Sciences, 1*(12) doi:http://dx.doi.org/10.3732/apps.1300070

Appendix A

FLOWERS OF CAMELLIA SPECIES AND CULTIVARS MOST USED IN THIS RESEARCH

Camellia azalea

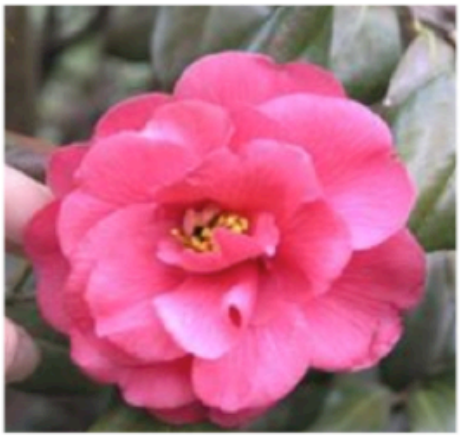
Camellia amplexicaulis

Camellia gaudichaudii

Camellia granthamiana

Camellia grijsii

Camellia japonica



Camellia japonica
'Maiden of Great Promise'

Camellia japonica
'Kumasaka'

Camellia oleifera

Camellia sasanqua

Camellia sinensis

Camellia transarisanensis

Appendix B

COMMONLY USED ABBREVIATIOSN FOR CAMELLIA SPECIES

Abbreviation	Species
AMP	*Camellia amplexicaulis*
AZA	*Camellia azalea*
GAUD	*Camellia gaudichaudii*
JAP	*Camellia japonica*
KAM	*Camellia japonica* 'Kumasaka'
MAID	*Camellia japonica* 'Maiden of Great Promise'
OLE	*Camellia oleifera*
SAS	*Camellia sasanqua*
TRA	*Camellia transarisanensis*

Appendix C

LIST OF ALL PRIMERS USED FOR DNA SEQUENCING

Gene Code (GenBank)	Forward Primer	Reverse Primer
AF315492	ACTAAACGACTCTCGGCAAC	TTTGTCAACCACCACTCGTC
AF315492	GATATCTCGGCTCTCGCATC	GTTTGTCAACCACCACTCGT
AY169403	CCAATTCCAGAGGTTGAGAA	GGATGTGCAATCCAAAACAA
AY169403	ACCAAAGCTAGGAAAAGAGG	TTCTCAACCTCTGGAATTGG
AY169403	CCTGCTATTTTGGATCAGGT	TAGGCTAGCAGCCATTTTAT
AY169403	ACATTGAAAAGAGTTTGGCG	ATAGGCTAGCAGCCATTTTA
AY787658	CCAAGTCTCAAATGGCTCGT	CAAGGTAAGACTCAGCAGCC
AY787658	TCGTACAAAGCAGACAGCTC	CGAGCAAGCTGCATATCCTT
AY787658	GACTTGTTCGTGAGATCGCT	AACACAGAACCGACAAAGCA
AY787658	TTGTTCGTGAGATCGCTCAG	CACAGAACCGACAAAGCACT
AY787658	CCGATCTCCGATTCCAGAGC	CTAAGCACGTTCGCCTCTGA
AY787658	AAGGCACCAAGGAAGCAACT	CAAACCCTAAGCACGTTCGC
AY787658	GATTCCAGAGCTCCGCTGTT	TAAGCACGTTCGCCTCTGAT
AY787658	AGCGTGTCACGATTATGCCT	ACACAGAACCGACAAAGCAC
DQ366599	GAACGAAGGAGCTTCTGGAG	TCCAGGCAAAGAAACGTTCA
DQ366599	GGCTAAGGGAAATGTCCTCG	AGCAATCTAATCAACCAAGCAC
DQ887753	AACTCGATGAGCCGTTTCTC	ACTGTGACCTTGGATCCTGA
DQ887753	CCATCAATCAGTTCCGCTCA	TGGTCTTCTCGTGGTTCTCT
DQ887753	GAACAATGTGGATGGCAAGC	TTGTTGTCCCAAGTAAGCCC
EU195885	AGAGTAGGTTGGCGCTTCTA	TGGCATTAAAGCATCGTGGT
EU195885	TGGTGCCATAAGCAAAGTGT	GTCCATCCACTGCATAAGGG
EU195885	CCATGGAAATGTCGGCTGTA	TTGGGAGGTACCCTGCATAA

Gene Code (GenBank)	Forward Primer	Reverse Primer
GQ214317	GATGCAAAACCGACCAACTG	AGACGGTTATAGAGTGGCGA
GQ214317	TCGCCACTCTATAACCGTCT	GAAATACGAGGACTGGGTCG
GQ214317	GGATCATCGAAAACTCGGCT	GGCACATTCACAAAGCTTCC
GQ214317	CGACCCAGTCCTCGTATTTC	CAGCACGCTCTCATCATCTT
GQ461358	TTGGCCCAGATATCCAACGG	GCCAATGCAACCAAACCCAA
GQ461358	GCGAGAGCAAGGTTCAGTCA	TGGCTTCGGTGGAGGAAAAG
GQ461358	CAGTCCATCCTGGCAAGACC	TGGCGGAGGAGTTGTAATGG
GU393024	ACAGCTTGTTCACTTGTTTG	GGAGGGGAAAATTAAGCAGA
GU393024	TCTGCTTAATTTTCCCCTCC	CATCCAAAATCAAGTCTCGC
GU393024	AAAATGATTTCGATGCGGAC	GGGAAAATTAAGCAGAGCAC
GU393024	GTGCTCTGCTTAATTTTCCC	ACGCATTGTCCAATACAGAT
JQ398741	GAGCTCCAAATATAGGGGTG	AATTCTCGGCTGTTTTTGTG
JQ398741	AAGGTCTAGAAGCTCTCAGT	CCCAGTACTCTGTCTCCTAA
JQ398741	ACAAAAACAGCCGAGAATTG	TGATGAACAGTCTTTGCTGA
JQ398741	GCTGATGAATCCGAAGTTTG	ACTGAGAGCTTCTAGACCTT
JX657332	GCCGTTCTGGACTTCTCAAA	TTTGCTCCTGCAATGCCTTA
JX657332	CGAGATCTCAATCCTCTGCG	GTAAGCACCACTTCCACCAA
JX657332	AGGTACTCGCATGCAGAAAG	GATCATGGTTTTGGTGCTGC
JX657332	AAGGCTAGGTTGGAGGTTCT	TGGTGGCATTACTGCATTGT
JX657332	AAAGTCCCCATAGTGTTTGG	TTCCTTTGGTGGAGAAGATG
KC149523	GGGTTGACATTGGTGGTGAT	TGTGATTCATCGTCTTCGGC
KC149523	AGATTGGGTTGTGTTGCATA	GTCCACCATAACAAGAGTGT
KC149523	AAACAAGTAGTCAGTCAGCC	TAAGGTCATCACCACCAATG

Appendix D

DETAILED TABLES OF POLYMORPHISMS BETWEEN ACCESSIONS

For research purposes only. Not intended for clinical use.

Date:		22-Apr-14
Project Name:	AF315492.SPF	
Contig Name:	AF315492.P	
Compare Bases to Consensus:	AF315492.P	
Comparison Range:	Unfiltered	
Base Positions:	1 to 196	
Options:	Large gap insertions (10 or more bases) excluded. Matches to ambiguous reference positions included.	

Position	AF315492.P	AMP_AF315492.P	AZA_AF315492.P	GAUD_AF315492.P	GRA_AF315492.P	GRI_AF315492.P	JAP_AF315492.P
42	C	C	Y	C	C	C	C
74	C	C	Y	C	C	C	C
78	A	A	W	A	A	A	A
89	T	Y	T	T	T	C	C
99	T	T	T	T	T	T	T
115	A	A	A	A	R	A	A
117	G	G	R	G	G	G	G
118	T	T	T	T	Y	T	T
119	T	T	T	T	Y	T	T
121	C	C	Y	C	C	C	C
137	T	T	T	T	T	Y	T
138	C	C	C	C	C	Y	C
140	A	A	A	A	A	A	A
141	A	A	:	A	A	:	G
146	T	Y	C	T	T	C	C
162	C	C	:	C	C	:	:
163	C	C	:	C	C	:	:
164	C	C	C	C	C	:	C
168	G	G	:	G	G	:	:
170	G	G	:	G	G	:	:
172	G	G	G	G	G	:	G
175	C	C	C	C	C	:	:
176	A	C	A	A	C	:	:
177	C	:	C	C	C	C	C
178	G	:	G	G	G	G	G
Total Differences		5	11	0	4	13	9

Position	AF315492.P	KAM_AF315492.P	MAID_AF315492.P	OLE_AF315492.P	SAS_AF315492.P	SIN_AF315492.P	TRA_AF315492.P	# Dif
42	C	C	C	C	C	C	C	1
74	C	C	C	C	C	C	C	1
78	A	A	A	A	A	A	A	1
89	T	T	T	T	Y	T	T	4
99	T	T	T	T	T	C	T	1
115	A	A	A	A	A	A	A	1
117	G	G	G	G	G	G	G	1
118	T	T	T	T	T	T	T	1
119	T	T	T	T	T	T	T	1
121	C	C	C	C	C	C	C	1
137	T	T	T	T	T	T	Y	2
138	C	C	C	C	C	C	Y	2
140	A	A	A	A	A	A	:	1
141	A	A	A	A	:	A	A	4
146	T	T	T	T	C	T	C	6
162	C	C	C	C	:	C	C	4
163	C	C	C	C	:	C	C	4
164	C	C	C	C	C	C	C	1
168	G	G	G	G	:	G	G	4
170	G	G	G	G	:	G	G	4
172	G	G	G	G	G	G	G	1
175	C	C	C	C	:	C	C	3
176	A	A	A	A	:	C	C	7
177	C	C	C	C	C	C	G	2
178	G	G	G	G	G	G	T	2
	Total Differences	0	0	0	9	2	7	

For research purposes only. Not intended for clinical use.

Date:		18-Apr
Project Name:	AY169403.SPF	
Contig Name:	AY169403.P	
Compare Bases to Consensus:	AY169403.P	
Comparison Range:	Unfiltered	
Base Positions:	1 to 642	
Options:	Large gap insertions (10 or more bases) excluded. Matches to ambiguous reference positions included.	

Position	AY169403.P	AMP_AY169403.P	AZA_AY169403.P	GAUD_AY169403.P	GRA_AY169403.P	GRI_AY169403.P	JAP_AY169403.P
24	A	X	A	X	A	A	A
26	C	X	C	X	C	C	Y
38	A	M	A	X	A	A	A
45	C	Y	C	X	C	C	C
56	C	G	C	C	C	C	C
59	G	C	R	G	G	G	R
66	G	G	G	G	G	G	G
69	C	C	C	C	C	C	C
75	C	C	Y	T	T	Y	Y
78	T	T	T	T	T	T	T
81	C	C	C	C	C	C	T
84	T	C	T	T	T	Y	Y
102	C	C	C	C	C	C	C
105	C	C	C	Y	C	C	C
106	G	A	G	G	G	G	G
108	C	C	C	C	C	C	C
120	A	G	A	A	A	A	A
132	C	T	C	C	C	C	C
141	T	C	T	T	T	T	Y
159	T	C	Y	T	T	T	Y
165	C	C	C	C	C	C	Y
168	G	G	G	G	G	R	G
171	T	T	T	T	T	T	T
177	C	G	C	C	C	C	C
178	C	C	C	C	C	C	C
189	T	T	T	T	T	T	T
195	C	C	C	C	C	C	C
207	A	A	A	A	A	A	A
210	T	T	T	T	T	T	T
217	G	G	G	G	G	R	G
223	G	G	G	G	G	G	G
225	C	T	C	C	C	C	C
240	C	C	C	C	C	C	C
244	G	A	G	G	G	G	G
246	G	G	G	G	G	G	G
248	T	T	K	T	T	T	K
252	A	A	A	A	A	A	A
255	C	C	C	C	C	C	C
261	G	G	G	G	G	G	G
271	A	G	R	A	A	A	A
282	C	C	C	C	C	C	C

Position	AY169403.P	AMP_AY169403.P	AZA_AY169403.P	GAUD_AY169403.P	GRA_AY169403.P	GRI_AY169403.P	JAP_AY169403.P
285	C	:	C	C	C	C	C
286	C	:	C	C	C	C	C
287	T	:	T	T	T	T	T
288	C	:	Y	C	C	C	C
289	G	:	G	G	G	G	G
290	T	:	T	T	T	T	T
291	A	:	A	A	A	A	A
292	G	:	G	G	G	G	G
293	G	:	G	G	G	G	G
294	C	:	C	C	C	C	C
295	C	:	C	C	C	C	C
296	A	:	A	A	A	A	A
297	A	:	A	A	A	A	A
298	T	:	T	T	T	T	T
299	C	:	C	C	C	C	C
300	C	:	C	C	C	C	C
301	C	:	C	C	C	C	C
302	T	:	T	T	T	T	T
303	A	:	A	A	A	A	A
304	T	:	T	T	T	T	T
305	T	:	T	T	T	T	T
306	T	:	T	T	T	T	T
308	G	:	G	G	G	G	G
309	G	:	G	G	G	G	G
310	T	:	T	T	T	T	T
311	G	:	G	G	G	G	G
312	A	:	A	A	A	A	A
313	T	:	T	T	T	T	T
314	G	:	G	G	G	G	G
315	G	:	G	G	G	G	G
316	A	:	A	A	A	A	A
317	G	:	G	G	G	G	G
318	C	:	C	C	C	C	C
319	A	:	A	A	A	A	A
320	G	:	G	G	G	G	G
321	C	:	C	C	C	C	C
322	C	:	C	C	Y	C	C
323	G	:	G	G	G	G	G
324	C	:	C	C	C	C	C
325	C	:	C	Y	C	C	T

326	A	:	A	A	A	A	A
327	T	:	T	T	T	T	T
328	T	:	T	T	T	T	T

Position	AY169403.P	AMP_AY169403.P	AZA_AY169403.P	GAUD_AY169403.P	GRA_AY169403.P	GRI_AY169403.P	JAP_AY169403.P
329	A	:	A	A	A	A	A
330	T	:	T	T	T	T	T
331	A	:	A	A	A	A	A
332	A	:	A	A	A	A	A
333	T	:	T	T	T	:	T
334	T	:	T	T	T	:	T
335	G	:	G	G	G	:	G
336	G	:	G	G	G	:	G
337	G	:	G	G	G	:	G
338	T	:	T	T	T	:	T
339	C	:	C	C	C	:	C
340	A	:	A	A	A	:	A
341	G	:	G	G	G	:	G
342	A	:	A	A	A	:	A
343	C	:	C	C	C	:	C
344	C	:	C	C	C	:	C
345	C	:	C	C	C	:	C
346	T	:	T	T	T	T	T
347	T	:	T	T	T	T	T
348	C	:	C	C	C	C	C
349	C	:	C	C	C	C	C
350	A	:	A	A	A	A	A
351	A	:	A	A	A	A	A
352	C	:	C	C	C	C	C
353	C	:	C	C	C	C	C
354	C	:	C	C	C	C	C
355	T	:	T	T	T	T	Y
356	T	:	T	T	T	T	T
357	G	:	G	G	G	G	G
358	G	:	G	G	G	G	G
359	G	:	G	G	G	G	G
360	C	:	C	C	C	C	C
361	A	:	A	A	A	A	A
362	T	:	T	T	T	T	T
363	C	:	C	C	C	C	C
364	T	:	T	T	T	T	T
365	C	:	Y	C	C	C	C

366	T	:	T	T	T	T	T
367	G	:	G	G	G	G	G
368	A	:	A	A	A	A	A
369	T	:	T	T	T	T	T
370	T	:	T	T	T	T	T
371	G	:	G	G	G	G	G

Position	AY169403.P	AMP_AY169403.P	AZA_AY169403.P	GAUD_AY169403.P	GRA_AY169403.P	GRI_AY169403.P	JAP_AY169403.P
372	G	:	G	G	G	G	G
373	A	:	A	A	A	A	A
374	A	:	A	A	A	A	A
375	C	:	C	C	C	C	C
376	T	:	T	T	T	T	T
377	C	:	C	C	C	C	C
378	T	:	T	T	T	T	T
379	C	:	C	C	C	C	C
380	T	:	T	T	T	T	T
381	G	:	G	G	G	G	G
382	T	:	T	T	T	T	T
383	T	:	T	T	T	T	T
384	T	:	T	T	T	T	T
385	T	:	T	T	T	T	T
386	G	:	G	G	G	G	G
387	G	:	G	G	G	G	G
388	A	:	A	A	A	A	A
389	T	:	T	T	T	T	T
390	T	:	T	T	T	T	T
391	G	:	G	G	G	G	G
392	C	:	C	C	C	C	C
393	A	:	A	A	A	A	A
394	C	:	C	C	C	C	C
395	A	:	A	A	A	A	A
396	T	:	T	T	T	T	T
397	C	:	C	C	C	C	C
398	C	:	C	C	C	C	C
399	T	:	T	T	T	T	T
400	G	:	G	G	G	G	G
401	G	:	G	G	G	G	G
402	T	:	T	T	T	T	T
403	G	:	G	G	G	G	G

Position	AY169403.P	AMP_AY169403.P	AZA_AY169403.P	GAUD_AY169403.P	GRA_AY169403.P	GRI_AY169403.P	JAP_AY169403.P
404	G	G	G	G	G	G	G
405	G	G	G	G	G	G	G
406	C	C	C	C	C	C	C
407	C	C	C	C	C	C	C
408	T	T	T	T	T	T	T
409	G	G	G	G	G	G	G
410	C	C	C	C	C	C	C
420	C	Y	C	C	C	T	C
425	T	T	T	T	T	T	T
426	G	G	A	G	G	G	G
429	A	A	G	A	A	A	A
435	A	A	A	A	A	A	A
439	G	K	G	G	G	G	G
453	G	T	G	G	G	G	G
466	A	A	M	A	M	A	M
468	G	G	G	G	G	A	G
475	G	G	G	G	G	G	S
489	C	C	C	T	C	C	T
495	T	T	T	T	T	C	T
506	C	C	C	C	C	C	C
507	G	G	G	G	G	A	G
510	C	C	C	C	C	T	C
516	G	G	G	G	G	A	G
561	C	C	C	C	C	C	C
563	A	A	G	A	A	A	A
573	T	K	T	T	T	T	T
582	C	C	C	C	C	C	C
609	A	T	A	A	A	A	A
610	C	C	C	C	C	C	Y
615	G	G	G	G	G	C	G
Total Differences		137	11	4	3	24	15

Position	AY169403.P	KAM_AY169403.P	MAI_AY169403.P	OLE_AY169403.P	SAS_AY169403.P	SIN_AY169403.P	TRN_AY169403.P	# Diff
24	A	A	A	A	A	R	A	1
26	C	C	C	C	C	C	C	1
38	A	A	A	A	A	A	A	1
45	C	C	C	C	C	C	C	1
56	C	C	C	C	C	C	C	1
59	G	R	R	G	R	G	G	6
66	G	G	G	G	G	K	G	1
69	C	C	C	C	C	C	Y	1
75	C	Y	T	C	Y	C	C	8
78	T	T	T	T	T	G	K	2
81	C	C	C	C	C	C	C	1
84	T	T	T	C	Y	C	C	7
102	C	C	C	C	C	M	C	1
105	C	C	C	C	C	Y	C	2
106	G	G	G	G	G	G	G	1
108	C	C	C	C	C	Y	C	1
120	A	A	A	A	R	A	A	2
132	C	C	C	C	C	C	C	1
141	T	Y	T	T	Y	Y	Y	6
159	T	Y	Y	Y	C	Y	Y	9
165	C	C	C	C	C	C	C	1
168	G	G	G	G	G	R	G	2
171	T	T	T	T	T	K	T	1
177	C	C	C	C	C	C	C	1
178	C	C	C	C	C	Y	C	1
189	T	T	T	T	T	Y	T	1
195	C	C	C	C	C	C	Y	1
207	A	A	A	A	A	R	A	1
210	T	T	T	T	T	A	T	1
217	G	G	G	G	G	G	G	1
223	G	G	G	G	G	K	G	1
225	C	Y	Y	C	C	C	Y	4
240	C	C	C	C	C	A	C	1
244	G	G	G	G	G	G	G	1
246	G	R	G	G	A	G	G	2
248	T	K	T	T	K	T	T	4
252	A	A	A	C	A	A	H	2
255	C	C	C	C	C	C	T	1
261	G	G	G	S	G	G	G	1
271	A	R	A	A	R	A	A	4

282	C	C	C	T	C	C	Y	2

Position	AY169403.P	KAM_AY169403.P	MAI_AY169403.P	OLE_AY169403.P	SAS_AY169403.P	SIN_AY169403.P	TRN_AY169403.P	Num of Var
285	C	C	C	C	C	C	C	1
286	C	C	C	C	C	C	C	1
287	T	T	T	T	T	T	T	1
288	C	Y	C	C	C	T	C	4
289	G	G	G	G	G	G	G	1
290	T	T	T	T	T	T	T	1
291	A	A	A	A	A	G	A	2
292	G	G	G	G	G	G	G	1
293	G	G	G	G	G	G	G	1
294	C	C	C	C	C	Y	C	2
295	C	C	C	C	C	C	C	1
296	A	A	A	A	A	A	A	1
297	A	A	A	A	A	A	A	1
298	T	K	T	T	K	K	K	5
299	C	C	C	C	C	C	C	1
300	C	C	C	C	C	C	C	1
301	C	C	C	C	C	C	C	1
302	T	T	T	T	T	T	T	1
303	A	A	A	A	A	A	A	1
304	T	W	T	T	T	T	T	2
305	T	T	T	T	T	T	T	1
306	T	T	T	T	T	T	T	1
308	G	G	G	G	G	G	G	1
309	G	G	G	G	G	G	G	1
310	T	T	T	T	T	T	T	1
311	G	G	G	G	G	G	G	1
312	A	A	A	A	A	A	A	1
313	T	T	T	T	T	T	T	1
314	G	G	G	G	G	G	G	1
315	G	G	G	G	G	G	G	1
316	A	A	A	A	A	M	A	2
317	G	G	G	G	G	G	G	1
318	C	C	C	C	C	C	C	1
319	A	A	A	A	A	G	A	2
320	G	G	G	G	G	G	G	1
321	C	C	C	C	C	C	C	1
322	C	C	C	C	C	Y	C	3
323	G	G	G	G	G	G	G	1
324	C	C	C	C	C	C	C	1

325	C	Y	Y	C	C	Y	Y	7
326	A	A	A	A	A	A	A	1
327	T	T	T	T	T	T	T	1
328	T	T	T	T	T	T	T	1

Position	AY169403.P	KAM_AY169403.P	MAI_AY169403.P	OLE_AY169403.P	SAS_AY169403.P	SIN_AY169403.P	TRN_AY169403.P	Num of Var
329	A	A	A	A	A	A	A	1
330	T	T	T	T	T	T	T	1
331	A	A	A	A	A	A	A	1
332	A	A	A	A	A	A	A	1
333	T	T	T	T	T	:	T	3
334	T	T	T	T	T	:	T	3
335	G	G	G	G	G	:	G	3
336	G	G	G	G	G	:	G	3
337	G	G	G	G	G	:	G	3
338	T	T	T	T	T	:	T	3
339	C	C	C	C	C	:	C	3
340	A	A	A	A	G	:	A	4
341	G	G	G	G	G	:	G	3
342	A	A	A	A	A	:	A	3
343	C	C	C	C	A	:	C	4
344	C	C	C	C	G	:	C	4
345	C	C	C	C	C	:	C	3
346	T	:	T	T	T	T	T	2
347	T	:	T	T	T	T	T	2
348	C	:	C	C	C	C	C	2
349	C	:	C	C	C	C	C	2
350	A	:	A	A	A	A	A	2
351	A	:	A	A	A	A	A	2
352	C	:	C	C	C	C	C	2
353	C	:	C	C	C	C	C	2
354	C	:	C	C	C	C	C	2
355	T	:	T	T	T	T	T	3
356	T	:	T	T	T	T	T	2
357	G	:	G	G	G	G	G	2
358	G	:	G	G	G	G	G	2
359	G	:	G	G	G	G	G	2
360	C	:	C	C	C	C	C	2
361	A	:	A	A	A	A	A	2
362	T	:	T	T	T	T	T	2
363	C	:	C	C	C	C	C	2
364	T	:	T	T	T	T	T	2

365	C	:	C	Y	C	C	C	4
366	T	:	T	T	T	T	T	2
367	G	:	G	G	G	G	G	2
368	A	:	A	A	A	A	A	2
369	T	:	T	T	T	T	T	2
370	T	:	T	T	T	T	T	2
371	G	:	G	G	G	G	G	2

Position	AY169403.P	KAM_AY169403.P	MAI_AY169403.P	OLE_AY169403.P	SAS_AY169403.P	SIN_AY169403.P	TRN_AY169403.P	Num of Var
372	G	:	G	G	G	G	G	2
373	A	:	A	A	A	A	A	2
374	A	:	A	A	A	A	A	2
375	C	:	C	C	C	C	C	2
376	T	:	T	T	T	T	T	2
377	C	:	C	C	C	C	C	2
378	T	:	T	T	T	T	T	2
379	C	:	C	C	C	T	C	3
380	T	:	T	T	T	T	T	2
381	G	:	G	G	G	G	R	3
382	T	:	T	T	T	T	T	2
383	T	:	T	T	T	T	T	2
384	T	:	T	T	T	T	T	2
385	T	:	T	T	T	T	T	2
386	G	:	G	G	G	G	G	2
387	G	:	G	G	G	G	G	2
388	A	:	A	A	A	A	A	2
389	T	:	T	T	T	T	T	2
390	T	:	T	T	T	T	T	2
391	G	:	G	G	G	G	G	2
392	C	:	C	C	C	C	C	2
393	A	:	A	A	A	A	A	2
394	C	:	C	C	C	C	C	2
395	A	:	A	A	A	A	A	2
396	T	:	T	T	T	T	T	2
397	C	:	C	C	C	C	C	2
398	C	:	C	C	C	C	C	2
399	T	:	T	T	T	T	T	2
400	G	:	G	G	G	G	G	2
401	G	:	G	G	G	G	G	2
402	T	:	T	T	T	T	T	2
403	G	:	G	G	G	G	G	2

Position	AY169403.P	KAM_AY169403.P	MAI_AY169403.P	OLE_AY169403.P	SAS_AY169403.P	SIN_AY169403.P	TRN_AY169403.P	Num of Var
404	G	:	G	G	G	G	G	1
405	G	:	G	G	G	G	G	1
406	C	:	C	C	C	C	C	1
407	C	:	C	C	C	C	C	1
408	T	:	T	T	T	T	T	1
409	G	:	G	G	G	G	G	1
410	C	:	C	C	C	C	C	1
420	C	Y	C	C	Y	T	Y	6
425	T	T	K	T	T	T	T	1
426	G	G	G	G	G	G	G	1
429	A	A	A	A	A	A	A	1
435	A	A	A	A	A	G	A	1
439	G	G	G	G	G	G	G	1
453	G	G	G	G	G	G	G	1
466	A	A	M	A	A	A	M	5
468	G	G	G	G	G	G	G	1
475	G	G	G	G	G	G	G	1
489	C	Y	T	C	C	C	C	4
495	T	T	T	T	T	T	T	1
506	C	C	C	S	C	C	C	1
507	G	G	G	G	G	G	G	1
510	C	C	C	C	C	C	C	1
516	G	G	G	G	G	G	G	1
561	C	C	C	C	C	C	T	1
563	A	A	A	A	A	A	A	1
573	T	T	T	T	T	T	T	1
582	C	T	T	C	C	C	C	2
609	A	A	W	A	A	A	T	3
610	C	C	C	C	C	C	C	1
615	G	G	G	G	G	G	G	1
Total Differences		80	10	7	14	41	17	0

For research purposes only. Not intended for clinical use.

Date:		18-Apr-14
Project Name:	AY787658.SPF	
Contig Name:	AY787658.P	
Compare Bases to Consensus:	AY787658.P	
Comparison Range:	Unfiltered	
Base Positions:	1 to 512	
Options:	Large gap insertions (10 or more bases) excluded. Matches to ambiguous reference positions included.	

Position	AY787658.P	AMP_AY787658.P	AZA_AY787658.P	GAUD_AY787658.P	GRA_AY787658.P	GRI_AY787658.P	JAP_AY787658.P
19	C	C	C	M	C	C	C
27	C	C	C	C	C	C	C
30	C	M	M	C	C	C	C
36	C	C	C	C	C	C	C
60	C	C	C	C	C	C	C
81	C	C	C	C	C	C	C
99	T	T	T	T	T	Y	Y
108	G	R	G	G	G	G	G
114	C	Y	Y	Y	C	C	C
138	G	G	G	G	G	K	G
141	C	C	C	C	C	C	C
151	C	C	C	C	C	C	C
168	C	C	C	C	C	C	C
181	A	A	A	A	R	R	R
183	C	M	M	C	C	C	C
195	A	W	A	A	A	A	A
198	C	Y	Y	C	C	C	C
204	A	A	A	A	A	A	A
213	T	T	T	T	T	T	T
228	C	C	C	C	Y	C	C
231	C	C	C	C	C	C	C
234	G	G	G	R	G	G	G
243	C	Y	Y	C	C	C	C
258	C	C	C	C	C	M	Y
261	T	T	T	T	T	T	T
267	G	K	K	G	G	G	G
285	T	T	T	T	T	T	T
289	G	G	G	G	G	G	G
330	T	T	T	T	T	T	T
345	T	T	T	T	T	T	K
366	T	T	T	T	T	T	Y
396	G	G	G	G	G	G	G
399	T	T	T	T	T	T	T
418	G	G	G	G	G	G	G
433	T	T	T	T	T	T	Y
435	T	T	T	T	T	Y	Y
440	C	C	C	C	C	Y	C
454	T	T	T	T	T	T	W
479	G	G	G	G	X	G	G
484	C	C	C	C	X	C	C
489	T	T	T	T	X	Y	T
Total Differences		8	6	3	2	7	8

Position	AY787658.P	KAM_AY787658.P	MAI_AY787658.P	OLE_AY787658.P	SAS_AY787658.P	SIN_AY787658.P	TRA_AY787658.P	# Dif
19	C	M	X	C	C	C	C	2
27	C	C	X	C	C	C	M	1
30	C	Y	X	C	C	C	C	3
36	C	C	C	Y	Y	C	T	3
60	C	C	C	C	C	C	Y	1
81	C	C	C	C	C	Y	M	2
99	T	T	T	T	Y	T	Y	4
108	G	G	G	G	G	G	G	1
114	C	T	T	C	C	C	Y	6
138	G	G	G	G	S	G	G	2
141	C	C	C	C	C	C	Y	1
151	C	C	C	C	C	C	M	1
168	C	Y	Y	C	C	C	C	2
181	A	A	A	A	R	A	A	4
183	C	C	C	C	Y	C	C	3
195	A	A	A	A	A	A	A	1
198	C	C	C	C	C	C	C	2
204	A	A	A	A	A	A	R	1
213	T	T	T	Y	T	T	T	1
228	C	C	C	C	C	C	C	1
231	C	C	C	C	C	C	Y	1
234	G	R	R	G	G	G	G	3
243	C	C	C	C	C	C	C	2
258	C	C	C	C	Y	C	C	3
261	T	T	T	T	Y	T	T	1
267	G	G	G	G	K	G	G	3
285	T	T	T	T	Y	T	T	1
289	G	G	G	G	G	K	G	1
330	T	T	T	Y	T	T	T	1
345	T	T	T	T	T	T	T	1
366	T	T	T	T	T	T	T	1
396	G	G	G	R	G	A	R	3
399	T	T	T	T	T	K	T	1
418	G	G	G	G	G	G	R	1
433	T	T	T	T	Y	T	T	2
435	T	T	T	T	T	T	T	2
440	C	C	C	C	C	C	C	1
454	T	T	T	T	T	T	T	1
479	G	G	G	G	R	G	G	1

484	C	C	C	C	C	C	Y	1
489	T	T	T	T	T	T	T	1
Total Differences			5	3	4	11	4	13	0

For research purposes only. Not intended for clinical use.

Date:	18-Apr-14
Project Name:	DQ366599.SPF
Contig Name:	DQ366599.P
Compare Bases to Consensus:	DQ366599.P
Comparison Range:	Unfiltered
Base Positions:	1 to 431
Options:	Large gap insertions (10 or more bases) excluded. Matches to ambiguous reference positions included.

Position	DQ366599.P	AMP_DQ366599.P	AZA_DQ366599.P	GAUD_DQ366599.P	GRA_DQ366599.P	GRI_DQ366599.P	JAP_DQ366599.P
4	G	X	X	X	G	X	X
49	A	X	A	A	A	A	A
62	C	X	C	Y	C	C	Y
63	G	X	R	G	G	G	G
72	G	X	G	G	R	G	G
86	T	T	T	T	T	T	T
87	C	C	C	C	C	C	C
88	T	T	T	T	T	T	T
89	G	A	G	G	G	G	G
105	A	A	A	A	A	R	A
109	T	T	T	Y	T	T	Y
120	A	A	A	A	A	A	A
124	G	G	G	G	A	G	G
126	A	A	T	W	T	T	A
133	A	A	A	A	A	A	M
150	T	T	T	T	T	T	T
153	C	C	C	C	C	C	C
157	G	G	G	G	G	G	G
166	A	A	A	A	A	A	A
167	G	G	R	G	G	G	G
174	A	A	A	A	A	A	A
175	C	C	C	C	C	C	C
177	A	A	A	A	A	A	A
188	A	A	A	A	A	A	A
195	C	C	C	C	C	C	C
198	C	C	C	C	C	C	C
199	G	G	G	G	G	G	G

Position	DQ366599.P	AMP_DQ366599.P	AZA_DQ366599.P	GAUD_DQ366599.P	GRA_DQ366599.P	GRI_DQ366599.P	JAP_DQ366599.P
207	T	T	T	T	T	T	T
214	A	A	A	A	A	A	A
218	G	G	G	G	G	G	G
222	T	T	T	T	T	T	T
223	G	G	G	G	G	G	G
226	C	C	C	C	C	C	C
230	A	A	A	A	A	A	A
232	T	T	T	T	T	T	T
247	C	C	Y	C	C	C	C
250	C	C	C	C	C	C	C
252	A	A	A	A	A	A	R
271	G	G	G	G	G	G	G
294	A	A	R	A	A	A	A
295	G	G	G	G	S	G	G
296	A	A	A	A	A	A	A
298	C	C	C	C	C	C	Y
308	G	G	G	G	G	G	G
318	A	A	A	A	A	A	A
325	G	G	G	G	G	G	G
330	A	A	A	A	A	A	A
334	A	A	A	A	A	A	A
342	C	T	C	C	C	C	C
345	A	A	A	A	A	G	A
350	A	A	A	R	A	A	R
374	A	A	A	A	A	A	A
399	C	C	C	Y	C	C	Y
400	G	G	G	G	R	G	G
417	G	G	G	G	G	G	G
Total Differences		2	5	5	5	3	7

Position	DQ366599.P	KAM_DQ366599.P	MAI_DQ366599.P	OLE_DQ366599.P	SAS_DQ366599.P	SIN_DQ366599.P	TRA_DQ366599.P	Number of Variants
4	G	X	X	G	X	G	S	1
49	A	A	A	A	G	A	A	1
62	C	Y	Y	C	C	C	C	4
63	G	G	G	G	G	G	G	1
72	G	G	G	G	G	G	G	1
86	T	T	T	T	Y	T	T	1
87	C	C	C	C	Y	C	C	1
88	T	T	T	T	K	T	T	1
89	G	G	G	G	G	G	G	1
105	A	A	A	A	A	A	A	1
109	T	Y	Y	T	T	T	T	4
120	A	R	A	A	A	A	A	1
124	G	G	G	G	G	G	G	1
126	A	W	W	T	R	A	A	8
133	A	A	A	A	G	A	A	2
150	T	T	T	T	W	T	T	1
153	C	C	C	C	Y	C	C	1
157	G	G	G	G	R	G	G	1
166	A	A	A	A	A	T	A	1
167	G	G	G	G	G	G	G	1
174	A	M	A	A	M	A	A	2
175	C	C	C	C	C	C	Y	1
177	A	A	A	A	W	A	A	1
188	A	A	A	A	R	A	A	1
195	C	S	C	C	S	C	C	2
198	C	S	C	C	C	C	C	1
199	G	G	G	G	R	G	G	1

Position	DQ366599.P	KAM_DQ366599.P	MAI_DQ366599.P	OLE_DQ366599.P	SAS_DQ366599.P	SIN_DQ366599.P	TRA_DQ366599.P	Number of Variants
207	T	T	T	T	W	T	T	1
214	A	A	A	A	R	A	A	1
218	G	G	G	G	K	G	G	1
222	T	K	T	T	T	T	T	1
223	G	G	G	G	G	K	G	1
226	C	C	C	C	Y	C	C	1
230	A	A	R	A	A	A	A	1
232	T	T	T	T	W	T	T	1
247	C	C	Y	C	C	Y	C	3
250	C	C	Y	C	C	C	C	1
252	A	A	A	A	A	A	A	1
271	G	G	G	G	G	R	G	1
294	A	A	A	A	A	A	A	1
295	G	G	G	G	R	G	G	2
296	A	A	A	A	M	A	A	1
298	C	C	C	C	C	C	C	1
308	G	G	R	G	G	G	G	1
318	A	A	A	A	M	A	A	1
325	G	G	G	G	G	G	T	1
330	A	A	A	A	M	A	A	1
334	A	A	A	A	W	A	A	1
342	C	C	C	C	C	C	C	1
345	A	A	A	A	A	A	A	1
350	A	R	R	A	A	A	A	4
374	A	A	A	A	A	A	M	1
399	C	Y	Y	C	C	C	C	4
400	G	G	G	G	G	G	R	2
417	G	G	G	G	X	R	G	1
Total Differences		10	9	1	24	5	5	0

For research purposes only. Not intended for clinical use.

Date:		18-Apr-14
Project Name:	DQ887753.SPF	
Contig Name:	DQ887753.P	
Compare Bases to Consensus:	DQ887753.P	
Comparison Range:	Unfiltered	
Base Positions:	1 to 887	
Options:	Large gap insertions (10 or more bases) excluded. Matches to ambiguous reference positions included.	

Position	DQ887753.P	AMP_DQ887753.P	AZA_DQ887753.P	GAUD_DQ887753.P	GRA_DQ887753.P	GRI_DQ887753.P	JAP_DQ887753.P
10	A	A	A	G	A	A	G
13	C	C	C	C	C	C	Y
58	C	C	C	C	C	C	C
70	C	C	C	C	C	C	T
106	C	Y	C	C	C	C	C
136	C	Y	T	C	C	C	C
154	G	G	G	G	G	G	G
172	C	Y	C	C	C	C	C
178	G	K	T	G	G	G	G
179	G	G	G	G	G	R	G
206	T	T	T	T	T	Y	T
207	A	A	A	A	A	A	A
251	A	A	A	A	A	M	C
259	G	K	G	G	G	G	G
268	C	C	C	S	C	C	C
269	A	A	A	R	A	A	A
270	G	G	G	G	C	G	G
277	C	C	C	C	C	C	T
283	C	C	C	C	T	C	C
316	G	G	G	G	G	G	G
349	G	G	G	G	G	G	A
372	C	C	C	C	C	C	C
426	C	C	C	T	C	Y	T
432	T	T	T	W	C	T	T

Position	DQ887753.P	AMP_DQ887753.P	AZA_DQ887753.P	GAUD_DQ887753.P	GRA_DQ887753.P	GRI_DQ887753.P	JAP_DQ887753.P
444	C	C	C	C	C	Y	C
453	C	Y	T	C	C	Y	C
462	T	Y	T	T	T	T	T
510	C	Y	T	C	C	C	C
579	G	G	G	G	T	G	G
591	C	C	C	C	C	C	Y
639	C	C	C	C	C	C	Y
645	C	C	C	C	C	Y	C
651	C	C	C	C	C	S	C
654	C	C	C	C	C	C	Y
672	C	C	C	C	C	C	Y
694	T	T	T	T	C	C	T
721	A	A	A	T	A	A	A
732	A	A	A	A	A	A	A
750	C	C	C	C	C	C	C
753	C	C	C	C	T	Y	C
768	C	C	C	C	C	C	C
771	G	R	G	G	G	G	G
816	C	C	C	C	T	C	C
862	A	R	A	A	A	A	A
870	C	C	C	C	C	C	C
882	C	Y	T	C	C	C	C
885	G	R	A	G	G	G	G
Total Differences		12	6	6	7	10	11

Position	DQ887753.P	KAM_DQ887753.P	MAI_DQ887753.P	OLE_DQ887753.P	SAS_DQ887753.P	SIN_DQ887753.P	TRA_DQ887753.P	Number of Variants
10	A	R	R	G	G	A	A	6
13	C	C	C	C	C	C	C	1
58	C	C	C	C	C	C	T	1
70	C	C	C	C	Y	C	C	2
106	C	C	C	C	C	C	C	1
136	C	C	C	C	C	C	C	2
154	G	R	R	G	G	G	G	2
172	C	C	C	C	C	C	C	1
178	G	G	G	G	G	A	G	3
179	G	G	G	G	G	G	G	1
206	T	T	T	T	T	T	T	1
207	A	A	A	A	A	A	T	1
251	A	A	A	A	M	A	A	3
259	G	G	G	G	G	G	G	1
268	C	C	C	G	S	C	C	3
269	A	R	R	A	A	A	A	3
270	G	S	S	G	G	C	G	4
277	C	C	C	C	Y	C	C	2
283	C	C	C	C	C	C	C	1
316	G	G	G	G	G	G	R	1
349	G	G	G	G	R	G	G	2
372	C	C	C	C	C	T	C	1
426	C	Y	Y	T	T	C	C	7
432	T	K	K	A	W	T	A	7

Posi

	DQ887753.P	KAM_DQ887753.P	MAI_DQ887753.P	OLE_DQ887753.P	SAS_DQ887753.P	SIN_DQ887753.P	TRA_DQ887753.P	# Dif
444	C	C	C	C	C	C	C	1
453	C	C	C	C	C	T	C	4
462	T	T	T	T	T	T	T	1
510	C	C	C	C	C	C	C	2
579	G	K	K	G	K	G	T	5
591	C	C	C	C	Y	C	C	2
639	C	C	C	C	C	C	C	1
645	C	C	C	C	C	C	C	1
651	C	C	C	C	C	C	C	1
654	C	C	C	C	Y	C	C	2
672	C	C	C	C	C	C	C	1
694	T	Y	Y	T	T	T	C	5
721	A	W	W	T	A	A	A	4
732	A	A	A	A	A	A	R	1
750	C	C	C	C	Y	C	C	1
753	C	C	C	C	Y	T	C	4
768	C	Y	Y	C	C	C	C	2
771	G	G	G	G	G	G	G	1
816	C	C	C	C	C	C	T	2
862	A	A	A	A	A	A	A	1
870	C	C	C	C	C	C	A	1
882	C	Y	Y	C	C	T	C	5
885	G	G	G	G	G	G	G	2
	Total Differences	11	11	5	13	6	9	0

For research purposes only. Not intended for clinical use.

Date:		22-Apr-14
Project Name:	EU195885.SPF	
Contig Name:	EU195885.P	
Compare Bases to Consensus:	EU195885.P	
Comparison Range:	Unfiltered	
Base Positions:	1 to 799	
Options:	Large gap insertions (10 or more bases) excluded. Matches to ambiguous reference positions included.	

Position	EU195885.P	AMP_EU195885.P	AZA_EU195885.P	GAUD_EU195885.P	GRA_EU195885.P	GRI_EU195885.P	JAP_EU195885.P
13	G	X	X	X	G	X	X
14	T	X	X	X	T	X	X
50	G	X	X	X	G	X	X
58	T	X	T	X	T	X	X
59	T	X	T	X	X	X	X
60	G	X	G	X	X	X	X
61	A	X	A	X	X	X	X
62	A	X	A	X	X	X	X
63	T	X	T	X	X	X	X
64	A	X	A	X	X	X	X
65	C	X	C	X	X	X	X
66	T	X	T	X	X	X	X
67	C	X	C	X	X	X	X
68	A	X	A	X	X	X	X
69	T	X	T	X	X	X	X
70	G	X	G	X	X	X	X
71	C	X	C	X	X	X	X
72	G	X	G	X	X	X	X
73	G	X	G	X	X	X	X
74	C	X	C	X	X	X	X
75	A	X	A	X	X	X	X
76	A	X	A	X	X	X	X
77	T	X	T	X	X	X	X
78	T	X	T	X	X	X	X
79	G	X	G	X	X	X	X
80	A	X	A	X	X	X	X
81	G	X	G	X	X	X	X
82	C	X	C	X	X	X	X
83	C	X	C	X	X	X	X
84	A	X	A	X	X	X	X
85	T	X	T	X	X	X	X
86	T	X	T	X	X	X	X
87	T	X	T	X	X	X	X
88	G	X	G	X	X	X	X
89	T	X	T	X	X	X	X
90	G	X	G	X	X	X	X
91	A	X	A	X	X	X	X
92	T	X	T	X	X	X	X
93	T	X	T	X	X	X	X
94	G	X	G	X	X	X	X

Position	EU195885.P	AMP_EU195885.P	AZA_EU195885.P	GAUD_EU195885.P	GRA_EU195885.P	GRI_EU195885.P	JAP_EU195885.P
95	C	X	C	X	X	X	X
96	A	X	A	X	X	X	X
97	A	X	A	X	X	X	X
98	C	X	C	X	X	X	X
99	A	X	A	X	X	X	X
100	A	X	A	X	X	X	X
101	A	X	A	X	X	X	X
102	T	X	T	X	X	X	X
103	A	X	A	X	X	X	X
104	G	X	G	X	X	X	X
105	A	X	A	X	X	X	X
106	C	X	C	X	X	X	X
107	A	X	A	X	X	X	X
108	G	X	G	X	X	X	X
109	C	X	C	X	X	X	X
110	T	X	T	X	X	X	X
111	G	X	G	X	X	X	X
112	A	X	A	X	X	X	X
113	G	X	G	X	X	X	X
114	T	X	T	X	X	X	X
115	G	X	G	X	X	X	X
116	T	X	T	X	X	X	X
117	G	X	G	X	X	X	X
118	C	X	C	X	X	X	X
119	T	X	T	X	X	X	X
120	T	X	T	X	X	X	X
121	C	X	C	X	X	X	X
122	A	X	A	X	X	X	X
123	C	X	C	X	X	X	X
124	C	X	C	X	X	X	X
125	C	X	C	X	X	X	X
126	G	X	G	X	X	X	X
127	A	X	A	X	X	X	X
128	T	X	T	X	X	X	X
129	A	X	A	X	X	X	X
130	C	X	C	X	X	X	X
131	A	X	A	X	X	X	X
132	C	X	C	X	X	X	X
133	A	X	A	X	X	X	X
134	A	X	A	X	X	X	X
135	G	X	G	X	X	X	X

Position	EU195885.P	AMP_EU195885.P	AZA_EU195885.P	GAUD_EU195885.P	GRA_EU195885.P	GRI_EU195885.P	JAP_EU195885.P
136	C	X	C	X	X	X	X
137	T	X	T	X	X	X	X
138	T	X	T	X	X	X	X
139	T	X	T	X	X	X	X
140	T	X	T	X	X	X	X
141	G	X	G	X	X	X	X
142	C	X	C	X	X	X	X
143	A	X	A	X	X	X	X
144	T	X	T	X	X	X	X
145	C	X	C	X	X	X	X
146	C	X	C	X	X	X	X
147	T	X	T	X	X	X	X
148	C	X	C	X	X	X	X
149	A	X	A	X	X	X	X
150	C	X	C	X	X	X	X
151	T	X	T	X	X	X	X
152	T	X	T	X	X	X	X
153	C	X	C	X	X	X	X
154	C	X	C	X	X	X	X
155	G	X	G	X	X	X	X
156	T	X	T	X	X	X	X
157	G	X	G	X	X	X	X
158	A	X	A	X	X	X	X
159	T	X	T	X	X	X	X
160	A	X	A	X	X	X	X
161	C	X	C	X	X	X	X
162	A	X	A	X	X	X	X
163	A	X	A	X	X	X	X
164	T	X	T	X	X	X	X
165	G	X	G	X	X	X	X
166	A	X	A	X	X	X	X
167	A	X	A	X	X	X	X
168	T	X	T	X	X	X	X
169	A	X	A	X	X	X	X
170	T	X	T	X	X	X	X
171	A	X	A	X	X	X	X
172	A	X	A	X	X	X	X
173	A	X	A	X	X	X	X
174	T	X	T	X	X	X	X
175	G	X	G	X	X	X	X
176	C	X	C	X	X	X	X

Position	EU195885.P	AMP_EU195885.P	AZA_EU195885.P	GAUD_EU195885.P	GRA_EU195885.P	GRI_EU195885.P	JAP_EU195885.P
177	T	X	T	X	X	X	X
178	T	X	T	X	X	X	X
179	T	X	T	X	X	X	X
180	T	X	T	X	X	X	X
181	G	X	G	X	X	X	X
182	C	X	C	X	X	X	X
183	T	X	T	X	X	X	X
184	A	X	A	X	X	X	X
185	G	X	G	X	X	X	X
186	A	X	A	X	X	X	X
187	C	X	C	X	X	X	X
188	A	X	A	X	X	X	X
189	A	X	A	X	X	X	X
190	A	X	A	X	X	X	X
191	T	X	T	X	X	X	X
192	C	X	C	X	X	X	X
193	T	X	T	X	X	X	X
194	T	X	T	X	X	X	X
195	G	X	G	X	X	X	X
196	A	X	A	X	X	X	X
197	T	X	T	X	X	X	X
198	T	X	T	X	X	X	X
199	A	X	A	X	X	X	X
200	A	X	A	X	X	X	X
201	T	X	T	X	X	X	X
202	G	X	G	X	X	X	X
203	C	X	C	X	X	X	X
204	C	X	C	X	X	X	X
205	G	X	G	X	X	X	X
206	A	X	A	X	X	X	X
207	T	X	T	X	X	X	X
208	G	X	G	X	X	X	X
209	G	X	G	X	X	X	X
210	A	X	A	X	X	X	X
211	A	X	A	X	X	X	X
212	T	X	T	X	X	X	X
213	T	X	T	X	X	X	X
214	C	X	C	X	X	X	X
215	T	X	T	X	X	X	X
216	T	X	T	X	X	X	X
217	G	X	G	X	X	X	X

Position	EU195885.P	AMP_EU195885.P	AZA_EU195885.P	GAUD_EU195885.P	GRA_EU195885.P	GRI_EU195885.P	JAP_EU195885.P
218	A	X	A	X	X	X	X
219	G	X	G	X	X	X	X
220	A	X	A	X	X	X	X
221	A	X	A	X	X	X	X
222	A	X	A	X	X	X	X
223	A	X	A	X	X	X	X
224	C	X	C	X	X	X	X
225	A	X	A	X	X	X	X
226	G	X	G	X	X	X	X
227	T	X	T	X	X	X	X
228	T	X	T	X	X	X	X
229	T	X	T	X	X	X	X
230	T	X	T	X	X	X	X
231	T	X	T	X	X	X	X
232	C	X	C	X	X	X	X
233	C	X	C	X	X	X	X
234	A	X	A	X	X	X	X
235	G	X	G	X	X	X	X
236	G	X	G	X	X	X	X
237	A	X	A	X	X	X	X
238	A	X	A	X	X	X	X
239	A	X	A	X	X	X	X
240	G	X	G	X	X	X	X
241	T	X	T	X	X	X	X
242	A	X	A	X	X	X	X
243	T	X	T	X	X	X	X
244	G	X	G	X	X	X	X
245	C	X	C	X	X	X	X
246	C	X	C	X	X	X	X
247	A	X	A	X	X	X	X
248	T	X	T	X	X	X	X
249	G	X	G	X	X	X	X
250	G	X	G	X	X	X	X
251	A	X	A	X	X	X	X
252	A	X	A	X	X	X	X
253	A	X	A	X	X	X	X
254	T	X	T	X	X	X	X
255	G	G	G	X	X	X	G
256	T	T	T	X	X	X	T
257	C	C	C	X	X	X	C
258	G	G	G	X	X	X	G
259	G	G	G	X	X	X	G

Position	EU195885.P	AMP_EU195885.P	AZA_EU195885.P	GAUD_EU195885.P	GRA_EU195885.P	GRI_EU195885.P	JAP_EU195885.P
260	C	C	S	X	X	X	C
261	T	T	T	X	X	X	T
262	G	G	G	X	X	X	G
263	T	T	T	X	X	X	T
264	A	A	R	X	X	X	A
265	G	G	G	X	X	X	G
266	T	T	T	X	X	X	T
267	C	C	C	X	X	X	C
268	T	T	T	X	X	X	T
269	A	A	A	X	X	X	A
270	C	C	C	X	X	X	C
271	A	A	A	X	X	X	A
272	A	A	A	X	X	X	A
273	G	G	G	X	X	X	G
274	A	A	A	X	X	X	A
275	A	A	A	X	X	X	A
276	C	C	S	X	X	X	C
277	T	T	T	X	X	X	T
278	G	G	G	X	X	X	G
279	G	G	G	X	X	X	G
280	G	G	G	X	X	X	G
281	T	T	T	X	X	X	T
282	T	T	T	X	X	X	T
283	T	T	T	X	X	X	T
284	T	T	T	X	X	X	T
285	T	T	T	X	X	X	T
286	C	C	C	X	X	X	C
287	C	C	C	X	X	X	C
288	C	C	C	X	X	X	S
289	G	G	G	X	X	X	G
290	A	A	A	X	X	X	A
291	G	G	G	X	X	X	G
292	C	C	C	X	X	X	C
293	A	A	A	X	X	X	A
294	A	A	A	X	X	X	A
295	G	G	G	G	X	X	G
296	C	C	C	C	X	X	C
297	A	A	A	A	X	X	A
298	C	C	C	C	X	X	C
299	T	T	T	T	X	X	T
300	T	T	T	T	X	X	T

Position	EU195885.P	AMP_EU195885.P	AZA_EU195885.P	GAUD_EU195885.P	GRA_EU195885.P	GRI_EU195885.P	JAP_EU195885.P
301	C	C	C	C	X	X	C
302	C	C	C	C	X	X	C
303	T	T	T	T	X	X	T
304	G	G	G	G	X	X	G
305	C	C	C	C	X	X	C
306	Y	C	C	T	X	X	C
307	G	G	G	G	X	X	G
308	A	A	A	A	X	X	A
309	T	T	T	T	X	X	T
310	C	C	C	C	X	X	C
311	T	T	T	T	X	X	T
312	C	C	C	C	X	X	C
313	A	A	A	A	X	X	A
314	T	T	T	T	X	X	T
315	G	G	G	G	X	X	G
316	A	A	A	A	X	X	A
317	A	A	A	A	X	X	A
318	G	G	G	G	X	X	G
319	A	A	A	A	X	X	A
320	G	G	G	G	X	X	G
321	G	G	G	G	X	X	G
322	T	T	T	T	X	X	T
323	A	A	A	A	X	X	A
324	A	A	A	A	X	X	A
325	A	A	A	A	X	X	A
326	A	A	A	A	X	X	A
327	A	A	A	A	X	X	A
328	C	C	C	C	X	X	C
329	A	A	A	A	X	X	A
330	T	T	T	T	X	X	T
331	T	T	T	T	X	X	T
332	A	A	A	A	X	X	A
333	T	T	T	T	X	X	T
334	T	T	T	T	X	X	T
335	T	T	T	T	X	X	T
336	C	C	C	C	X	X	C
337	C	C	C	C	X	X	C
338	T	T	T	T	X	X	T
339	T	T	T	T	X	X	T
340	A	R	A	A	X	X	A

Position	EU195885.P	AMP_EU195885.P	AZA_EU195885.P	GAUD_EU195885.P	GRA_EU195885.P	GRI_EU195885.P	JAP_EU195885.P
341	C	C	C	C	X	X	C
342	T	T	T	T	X	X	T
343	A	R	A	A	X	X	A
344	T	Y	M	T	X	X	T
345	A	A	A	A	X	X	A
346	G	G	G	G	X	X	G
347	T	T	T	T	X	X	T
348	A	A	A	A	X	X	A
349	A	A	A	A	X	X	A
350	A	A	A	W	X	X	A
351	A	A	A	A	X	X	A
352	T	T	T	T	X	X	T
353	T	T	T	T	X	X	T
354	:	:	:	:	X	X	:
355	:	:	:	:	X	X	:
356	C	C	C	C	X	X	C
357	T	T	T	T	X	X	T
358	G	G	G	G	X	X	G
359	T	T	T	T	X	X	T
360	T	T	Y	T	X	X	T
361	T	T	T	T	X	X	T
362	A	A	A	A	X	X	A
364	G	G	G	G	X	X	G
365	T	T	T	T	X	X	T
366	T	T	T	T	X	X	T
368	C	Y	Y	C	X	X	C
369	T	T	T	T	X	X	T
370	C	C	C	C	X	X	C
371	A	A	A	A	X	X	A
372	A	R	R	A	X	X	A
373	C	C	C	C	X	X	C
374	A	A	A	A	X	X	A
375	T	T	T	T	X	X	T
376	A	A	A	A	X	X	A
377	C	C	C	C	X	X	C
378	T	T	T	T	X	X	T
379	T	T	T	T	X	X	T
380	G	G	G	G	X	X	G
381	G	G	G	G	X	X	G
382	T	T	T	T	X	X	T
383	A	A	A	A	X	X	A
384	C	C	C	C	X	X	C

385	T	T	T	T	X	X	T

Position	EU195885.P	AMP_EU195885.P	AZA_EU195885.P	GAUD_EU195885.P	GRA_EU195885.P	GRI_EU195885.P	JAP_EU195885.P
386	A	A	A	A	X	X	A
387	A	A	A	A	X	X	A
388	A	A	A	A	X	X	A
389	A	A	A	A	X	X	A
390	C	C	C	C	X	X	C
391	A	A	A	A	X	X	A
392	A	A	A	A	X	X	A
393	A	A	A	A	X	X	A
394	C	C	C	C	X	X	C
395	T	T	T	T	X	X	T
396	T	T	T	T	X	X	T
397	G	G	G	G	X	X	G
398	T	T	T	T	X	X	T
400	A	A	A	A	X	X	A
401	G	G	G	G	X	X	G
402	C	C	C	C	X	X	C
403	G	G	G	G	X	X	G
404	A	A	A	A	X	X	A
405	A	A	A	A	X	X	A
406	T	T	T	Y	X	X	T
407	A	A	W	A	X	X	A
408	C	C	C	C	X	X	C
409	A	A	A	A	X	X	A
410	C	C	C	C	X	X	C
411	A	A	A	A	X	X	A
418	:	:	:	:	:	:	:
421	A	A	A	A	M	M	A
422	A	A	A	A	A	A	A
423	T	T	T	T	T	W	T
424	A	A	A	A	A	W	A
429	C	C	C	C	C	H	C
430	T	T	T	T	T	Y	T
433	C	C	C	C	C	S	C
436	G	G	G	G	C	Y	G
437	T	T	T	T	T	T	T

Position	EU195885.P	AMP_EU195885.P	AZA_EU195885.P	GAUD_EU195885.P	GRA_EU195885.P	GRI_EU195885.P	JAP_EU195885.P
439	C	C	C	C	C	C	C
445	T	T	T	T	W	A	T
446	T	T	T	T	W	T	T
466	G	G	G	G	G	G	R
478	C	C	C	C	R	T	C
481	C	C	M	C	Y	Y	C
484	T	T	T	W	T	T	T
486	C	C	C	C	C	A	C
490	T	T	T	T	T	T	T
493	C	C	C	C	M	C	C
509	C	C	C	C	C	C	G
523	T	T	T	T	T	T	T
527	G	G	G	G	G	G	S
556	T	T	T	T	C	C	T
574	C	C	C	C	C	C	C
577	C	C	C	C	T	T	C
592	C	C	C	C	C	Y	C
598	T	T	T	T	K	T	T
622	G	R	G	G	R	G	A
628	C	C	C	C	C	C	C
634	T	T	Y	T	T	T	T
664	T	T	T	T	Y	C	T
667	C	C	C	C	R	C	C
670	C	C	C	C	C	C	C
694	T	T	T	T	T	T	T
711	G	G	R	G	G	G	G
715	G	G	G	G	R	G	G
724	C	C	C	C	C	C	C
725	A	A	A	A	A	A	A
727	A	A	A	A	A	A	A
736	T	T	T	T	T	T	W
767	G	G	G	X	X	X	G
770	G	G	G	X	X	X	G
781	A	R	A	X	X	X	A
Total Differences		8	12	4	362	15	7

Position	EU195885.P	KAM_EU195885.P	MAI_EU195885.P	OLE_EU195885.P	SAS_EU195885.P	SIN_EU195885.P	TRA_EU195885.P	Number of Variants
13	G	G	X	S	G	X	X	1
14	T	T	X	W	T	X	X	1
50	G	G	X	K	G	X	X	1
58	T	T	X	X	T	X	T	1
59	T	T	X	X	T	X	T	2
60	G	G	X	X	G	X	G	2
61	A	A	X	X	A	X	A	2
62	A	A	X	X	A	X	A	2
63	T	T	X	X	T	X	T	2
64	A	A	X	X	A	X	A	2
65	C	C	X	X	C	X	C	2
66	T	T	X	X	T	X	T	2
67	C	C	X	X	C	X	C	2
68	A	A	X	X	A	X	A	2
69	T	T	X	X	T	X	T	2
70	G	G	X	X	G	X	G	2
71	C	C	X	X	C	X	C	2
72	G	G	X	X	G	X	G	2
73	G	G	X	X	G	X	G	2
74	C	C	X	X	C	X	C	2
75	A	A	X	X	A	X	A	2
76	A	A	X	X	A	X	A	2
77	T	T	X	X	T	X	T	2
78	T	T	X	X	T	X	T	2
79	G	G	X	X	G	X	G	2
80	A	A	X	X	A	X	A	2
81	G	G	X	X	G	X	G	2
82	C	C	X	X	C	X	C	2
83	C	C	X	X	C	X	C	2
84	A	A	X	X	A	X	A	2
85	T	T	X	X	T	X	T	2
86	T	T	X	X	T	X	T	2
87	T	T	X	X	T	X	T	2
88	G	G	X	X	G	X	G	2
89	T	T	X	X	T	X	T	2
90	G	G	X	X	G	X	G	2
91	A	A	X	X	A	X	A	2
92	T	T	X	X	T	X	T	2
93	T	T	X	X	T	X	T	2
94	G	G	X	X	G	X	G	2

Position	EU195885.P	KAM_EU195885.P	MAI_EU195885.P	OLE_EU195885.P	SAS_EU195885.P	SIN_EU195885.P	TRA_EU195885.P	Number of Variants
95	C	C	X	X	C	X	C	2
96	A	A	X	X	A	X	A	2
97	A	A	X	X	A	X	A	2
98	C	C	X	X	C	X	C	2
99	A	A	X	X	A	X	A	2
100	A	A	X	X	A	X	A	2
101	A	A	X	X	A	X	A	2
102	T	T	X	X	T	X	T	2
103	A	A	X	X	A	X	A	2
104	G	G	X	X	G	X	G	2
105	A	A	X	X	A	X	A	2
106	C	C	X	X	C	X	C	2
107	A	R	X	X	A	X	A	3
108	G	G	X	X	G	X	G	2
109	C	C	X	X	C	X	C	2
110	T	T	X	X	W	X	T	3
111	G	G	X	X	G	X	G	2
112	A	A	X	X	A	X	A	2
113	G	G	X	X	G	X	G	2
114	T	T	X	X	T	X	T	2
115	G	G	X	X	G	X	G	2
116	T	T	X	X	T	X	T	2
117	G	G	X	X	G	X	G	2
118	C	C	X	X	C	X	C	2
119	T	T	X	X	T	X	T	2
120	T	T	X	X	T	X	T	2
121	C	C	X	X	C	X	C	2
122	A	A	X	X	A	X	A	2
123	C	C	X	X	C	X	C	2
124	C	C	X	X	C	X	C	2
125	C	C	X	X	C	X	C	2
126	G	G	X	X	G	X	G	2
127	A	A	X	X	A	X	A	2
128	T	T	X	X	T	X	T	2
129	A	A	X	X	A	X	A	2
130	C	C	X	X	C	X	C	2
131	A	A	X	X	A	X	A	2
132	C	C	X	X	C	X	C	2
133	A	A	X	X	A	X	A	2

134	A	A	X	X	A	X	A	2
135	G	G	X	X	G	X	G	2

Position	EU195885.P	KAM_EU195885.P	MAI_EU195885.P	OLE_EU195885.P	SAS_EU195885.P	SIN_EU195885.P	TRA_EU195885.P	Number of Variants
136	C	C	X	X	Y	X	C	3
137	T	T	X	X	T	X	T	2
138	T	T	X	X	T	X	T	2
139	T	T	X	X	T	X	T	2
140	T	T	X	X	T	X	T	2
141	G	G	X	X	G	X	G	2
142	C	C	X	X	C	X	C	2
143	A	A	X	X	A	X	A	2
144	T	T	X	X	T	X	T	2
145	C	C	X	X	C	X	C	2
146	C	C	X	X	C	X	C	2
147	T	T	X	X	T	X	T	2
148	C	C	X	X	C	X	C	2
149	A	A	X	X	A	X	A	2
150	C	C	X	X	Y	X	C	3
151	T	T	X	X	T	X	T	2
152	T	T	X	X	T	X	T	2
153	C	C	X	X	C	X	C	2
154	C	C	X	X	C	X	C	2
155	G	G	X	X	G	X	G	2
156	T	T	X	X	T	X	T	2
157	G	G	X	X	G	X	G	2
158	A	A	X	X	A	X	A	2
159	T	T	X	X	T	X	T	2
160	A	A	X	X	A	X	A	2
161	C	C	X	X	C	X	C	2
162	A	A	X	X	A	X	A	2
163	A	A	X	X	A	X	A	2
164	T	T	X	X	T	X	T	2
165	G	G	X	X	G	X	G	2
166	A	A	X	X	A	X	A	2
167	A	A	X	X	A	X	A	2
168	T	W	X	X	T	X	T	3
169	A	A	X	X	A	X	A	2
170	T	T	X	X	T	X	T	2
171	A	A	X	X	A	X	A	2
172	A	A	X	X	A	X	A	2

173	A	A	X	X	A	X	A	2
174	T	T	X	X	T	X	T	2
175	G	G	X	X	G	X	G	2
176	C	C	X	X	C	X	C	2

Position	EU195885.P	KAM_EU195885.P	MAI_EU195885.P	OLE_EU195885.P	SAS_EU195885.P	SIN_EU195885.P	TRA_EU195885.P	Number of Variants
177	T	T	X	X	T	X	T	2
178	T	T	X	X	T	X	T	2
179	T	T	X	X	K	X	T	3
180	T	T	X	X	T	X	G	3
181	G	G	X	X	G	X	G	2
182	C	C	X	X	C	X	C	2
183	T	T	X	X	T	X	T	2
184	A	A	X	X	A	X	A	2
185	G	G	X	X	G	X	G	2
186	A	A	X	X	A	X	A	2
187	C	C	X	X	C	X	C	2
188	A	A	X	X	A	X	A	2
189	A	A	X	X	W	X	A	3
190	A	A	X	X	A	X	A	2
191	T	T	X	X	T	X	T	2
192	C	C	X	X	C	X	C	2
193	T	T	X	X	T	X	T	2
194	T	K	X	X	T	X	T	3
195	G	G	X	X	G	X	G	2
196	A	A	X	X	A	X	A	2
197	T	T	X	X	T	X	T	2
198	T	T	X	X	T	X	T	2
199	A	A	X	X	A	X	A	2
200	A	A	X	X	A	X	A	2
201	T	K	X	X	T	X	T	3
202	G	G	X	X	G	X	G	2
203	C	C	X	X	C	X	C	2
204	C	C	X	X	C	X	C	2
205	G	G	X	X	G	X	G	2
206	A	A	X	X	A	X	A	2
207	T	T	X	X	T	X	T	2
208	G	G	X	X	G	X	G	2
209	G	G	X	X	G	X	G	2
210	A	A	X	X	A	X	A	2
211	A	A	X	X	A	X	A	2

212	T	T	X	X	T	X	T	2
213	T	T	X	X	T	X	T	2
214	C	C	X	X	C	X	C	2
215	T	T	X	X	T	X	T	2
216	T	T	X	X	T	X	T	2
217	G	G	X	X	G	X	G	2

Position	EU195885.P	KAM_EU195885.P	MAI_EU195885.P	OLE_EU195885.P	SAS_EU195885.P	SIN_EU195885.P	TRA_EU195885.P	Number of Variants
218	A	A	X	X	A	X	A	2
219	G	G	X	X	G	X	G	2
220	A	A	X	X	A	X	A	2
221	A	A	X	X	A	X	A	2
222	A	A	X	X	A	X	A	2
223	A	A	X	X	A	X	A	2
224	C	C	X	X	C	X	C	2
225	A	A	X	X	A	X	A	2
226	G	G	X	X	G	X	G	2
227	T	T	X	X	T	X	T	2
228	T	T	X	X	T	X	T	2
229	T	T	X	X	T	X	T	2
230	T	T	X	X	T	X	T	2
231	T	T	X	X	T	X	T	2
232	C	C	X	X	C	X	C	2
233	C	C	X	X	C	X	C	2
234	A	A	X	X	A	X	A	2
235	G	G	X	X	G	X	G	2
236	G	G	X	X	G	X	G	2
237	A	A	X	X	A	X	A	2
238	A	A	X	X	A	X	A	2
239	A	A	X	X	A	X	A	2
240	G	G	X	X	G	X	G	2
241	T	T	X	X	T	X	T	2
242	A	A	X	X	A	X	A	2
243	T	T	X	X	T	X	T	2
244	G	G	X	X	G	X	G	2
245	C	C	X	X	C	X	C	2
246	C	C	X	X	C	X	C	2
247	A	A	X	X	A	X	A	2
248	T	T	X	X	T	X	T	2
249	G	G	X	X	G	X	G	2
250	G	G	X	X	G	X	G	2

251	A	A	X	X	A	X	A	2
252	A	A	X	X	A	X	A	2
253	A	A	X	X	A	X	A	2
254	T	T	X	X	T	X	T	2
255	G	G	X	X	G	X	G	2
256	T	T	X	X	T	X	T	2
257	C	C	X	X	C	X	C	2
258	G	G	X	X	G	X	G	2
259	G	G	X	X	G	X	G	2

Position	EU195885.P	KAM_EU195885.P	MAI_EU195885.P	OLE_EU195885.P	SAS_EU195885.P	SIN_EU195885.P	TRA_EU195885.P	Number of Variants
260	C	C	X	X	C	X	C	3
261	T	T	X	X	T	X	T	2
262	G	G	X	X	G	X	G	2
263	T	T	X	X	T	X	T	2
264	A	A	X	X	A	X	A	3
265	G	G	G	:	G	G	G	2
266	T	T	T	:	T	T	T	2
267	C	C	C	:	C	C	C	2
268	T	T	T	:	T	T	T	2
269	A	A	A	:	A	A	A	2
270	C	C	C	:	C	C	C	2
271	A	A	A	:	A	A	A	2
272	A	A	A	:	A	A	A	2
273	G	G	G	:	G	G	G	2
274	A	A	A	:	A	A	A	2
275	A	A	A	:	A	A	A	2
276	C	C	C	:	C	C	C	3
277	T	T	T	T	T	T	T	1
278	G	G	G	G	G	G	G	1
279	G	G	G	G	G	G	G	1
280	G	G	G	G	G	G	G	1
281	T	T	T	T	T	T	T	1
282	T	T	T	T	T	T	T	1
283	T	T	T	T	T	T	T	1
284	T	T	T	T	T	T	T	1
285	T	T	T	T	T	T	T	1
286	C	C	C	C	C	C	C	1
287	C	C	C	C	C	C	C	1
288	C	C	C	C	C	C	C	2
289	G	G	G	G	G	G	G	1

290	A	A	A	A	A	A	A	1
291	G	G	G	G	G	G	G	1
292	C	C	C	C	C	C	C	1
293	A	A	A	A	A	A	A	1
294	A	A	A	A	A	A	A	1
295	G	G	G	G	G	G	G	1
296	C	C	C	C	C	C	C	1
297	A	A	A	A	A	A	A	1
298	C	C	C	C	C	C	C	1
299	T	T	T	T	T	T	T	1
300	T	T	T	T	T	T	T	1

Position	EU195885.P	KAM_EU195885.P	MAI_EU195885.P	OLE_EU195885.P	SAS_EU195885.P	SIN_EU195885.P	TRA_EU195885.P	Number of Variants
301	C	C	C	C	C	C	C	1
302	C	C	C	C	C	C	C	1
303	T	T	T	T	T	T	T	1
304	G	G	G	G	G	G	G	1
305	C	C	C	C	C	C	C	1
306	Y	T	T	T	T	C	C	11
307	G	G	G	G	G	G	G	1
308	A	A	A	A	A	A	A	1
309	T	T	T	T	T	T	T	1
310	C	C	C	C	C	C	C	1
311	T	T	T	T	T	T	T	1
312	C	Y	C	C	C	T	C	3
313	A	A	A	A	A	A	A	1
314	T	T	T	T	T	T	T	1
315	G	G	G	G	G	C	G	2
316	A	A	A	A	A	A	A	1
317	A	A	A	A	A	A	A	1
318	G	G	G	G	G	G	G	1
319	A	A	A	A	A	A	A	1
320	G	G	G	G	G	G	G	1
321	G	G	G	G	G	G	G	1
322	T	T	T	T	T	T	T	1
323	A	A	A	A	A	A	A	1
324	A	A	A	A	A	A	A	1
325	A	A	A	A	A	A	A	1
326	A	A	A	A	A	A	A	1
327	A	A	A	A	A	A	G	2
328	C	C	C	C	C	C	C	1

329	A	A	A	A	A	A	A	1
330	T	T	T	T	T	T	T	1
331	T	T	T	T	T	T	T	1
332	A	A	A	A	A	A	A	1
333	T	T	T	T	T	T	T	1
334	T	T	T	T	T	T	T	1
335	T	T	T	T	T	T	T	1
336	C	C	C	C	C	C	C	1
337	C	C	C	C	C	C	C	1
338	T	T	T	T	T	T	T	1
339	T	T	T	T	T	T	T	1
340	A	A	A	A	A	A	A	2

Position	EU195885.P	KAM_EU195885.P	MAI_EU195885.P	OLE_EU195885.P	SAS_EU195885.P	SIN_EU195885.P	TRA_EU195885.P	Number of Variants
341	C	C	C	C	C	C	C	1
342	T	T	T	T	T	T	T	1
343	A	A	A	A	A	A	A	2
344	T	T	T	T	T	T	T	3
345	A	A	A	A	A	A	A	1
346	G	G	G	G	G	G	G	1
347	T	T	T	T	T	T	T	1
348	A	A	A	A	A	A	A	1
349	A	A	A	A	A	A	A	1
350	A	A	A	A	A	A	A	2
351	A	A	A	A	A	A	A	1
352	T	T	T	T	T	T	T	1
353	T	T	T	T	T	T	T	1
354	:	:	:	:	:	T	:	1
355	:	:	:	:	:	T	:	1
356	C	C	C	C	M	C	C	2
357	T	T	T	T	T	T	T	1
358	G	G	G	G	G	G	G	1
359	T	T	T	T	T	T	T	1
360	T	T	T	T	T	T	T	2
361	T	T	T	T	T	T	T	1
362	A	A	A	A	A	A	A	1
364	G	G	G	G	G	G	G	1
365	T	T	T	T	T	T	T	1
366	T	T	T	T	T	T	T	1
368	C	C	C	C	C	C	C	3
369	T	T	T	T	T	T	T	1

370	C	C	C	C	C	C	C	1
371	A	A	A	A	R	A	A	2
372	A	A	A	A	A	A	A	3
373	C	C	C	C	C	C	C	1
374	A	A	A	A	A	A	A	1
375	T	T	T	T	T	T	T	1
376	A	A	A	A	A	A	A	1
377	C	C	C	C	C	C	C	1
378	T	T	T	T	T	T	T	1
379	T	T	T	T	T	:	T	2
380	G	G	G	G	G	:	G	2
381	G	G	G	G	G	:	G	2
382	T	T	T	T	T	:	T	2
383	A	A	A	A	A	:	A	2
384	C	C	C	C	C	:	C	2
385	T	T	T	T	T	:	T	2

Position	EU195885.P	KAM_EU195885.P	MAI_EU195885.P	OLE_EU195885.P	SAS_EU195885.P	SIN_EU195885.P	TRA_EU195885.P	Number of Variants
386	A	A	A	A	A	A	A	1
387	A	A	A	A	A	A	A	1
388	A	A	A	A	A	A	A	1
389	A	A	A	A	A	A	A	1
390	C	C	C	C	C	C	C	1
391	A	A	A	A	A	A	A	1
392	A	A	A	A	A	A	A	1
393	A	A	A	A	A	A	A	1
394	C	C	C	C	C	C	C	1
395	T	T	T	T	T	T	T	1
396	T	T	T	T	T	T	T	1
397	G	G	G	G	G	G	G	1
398	T	T	T	T	T	T	T	1
400	A	A	A	A	A	A	A	1
401	G	G	G	G	G	G	G	1
402	C	C	C	C	C	C	C	1
403	G	G	G	G	G	G	G	1
404	A	A	A	A	A	A	A	1
405	A	A	A	A	A	A	A	1
406	T	T	T	T	T	T	T	2
407	A	A	A	A	A	A	A	2
408	C	C	C	C	C	C	C	1
409	A	A	A	A	A	A	A	1

410	C	C	C	C	C	C	C	1
411	A	A	A	A	M	A	A	2
418	:	:	:	:	:	A	:	1
421	A	A	A	A	A	A	A	2
422	A	A	A	A	W	A	A	1
423	T	T	T	T	T	T	T	1
424	A	A	A	A	A	A	A	1
429	C	C	C	C	C	C	C	1
430	T	T	T	T	T	T	T	1
433	C	C	C	C	C	C	C	1
436	G	G	G	G	G	:	A	4
437	T	T	T	T	T	:	T	1

Position	EU195885.P	KAM_EU195885.P	MAI_EU195885.P	OLE_EU195885.P	SAS_EU195885.P	SIN_EU195885.P	TRA_EU195885.P	Number of Variants
439	C	C	C	C	C	T	C	1
445	T	T	T	T	T	T	T	2
446	T	T	T	T	T	T	T	1
466	G	G	G	G	G	G	T	2
478	C	C	C	C	C	C	C	2
481	C	C	C	C	C	C	C	3
484	T	A	W	T	T	T	T	3
486	C	C	C	C	C	C	C	1
490	T	T	T	T	Y	T	Y	2
493	C	C	C	C	C	C	C	1
509	C	C	C	C	S	C	C	2
523	T	T	T	C	Y	T	T	2
527	G	G	G	G	G	G	G	1
556	T	T	T	C	Y	C	Y	6
574	C	C	C	C	C	G	C	1
577	C	C	C	C	Y	C	C	3
592	C	C	C	C	C	C	C	1
598	T	T	T	T	T	T	Y	2
622	G	G	G	G	R	A	R	6
628	C	C	C	C	Y	C	C	1
634	T	T	T	T	T	T	T	1
664	T	T	T	T	Y	T	T	3
667	C	C	C	C	C	C	Y	2
670	C	C	C	T	C	C	C	1
694	T	T	T	T	T	T	Y	1
711	G	G	G	G	G	G	G	1
715	G	G	G	G	G	G	G	1

724	C	C	C	C	C	C	Y	1
725	A	A	W	A	A	T	A	2
727	A	A	A	A	A	Y	A	1
736	T	T	T	T	W	T	T	2
767	G	S	G	G	G	G	G	1
770	G	G	X	G	G	K	G	1
781	A	A	X	A	A	A	A	1
Total Differences		8	3	226	19	22	12	0

For research purposes only. Not intended for clinical use.

Date:	25-Apr	2014
Project Name:	EU195885.SPF	NOTE: This is a segment that sequenced across all accessions but did not align with the GenBank Sequence
Contig Name:	EU195885.P	
Compare Bases to Consensus:	EU195885.P	
Comparison Range:	Unfiltered	
Base Positions:	1 to 799	
Options:	Large gap insertions (10 or more bases) excluded. Matches to ambiguous reference positions included.	

Position	EU195885.P	AMP_EU195885.P	AZA_EU195885.P	GAUD_EU195885.P	GRA_EU195885.P	GRI_EU195885.P	JAP_EU195885.P
107	A	X	A	X	X	X	X
110	T	X	T	X	X	X	X
136	C	X	C	X	X	X	X
150	C	X	C	X	X	X	X
168	T	X	T	X	X	X	X
179	T	X	T	X	X	X	X
180	T	X	T	X	X	X	X
189	A	X	A	X	X	X	X
194	T	X	T	X	X	X	X
201	T	X	T	X	X	X	X
260	C	C	S	X	X	X	C
264	A	A	R	X	X	X	A
276	C	C	S	X	X	X	C
288	C	C	C	X	X	X	S
306	Y	C	C	T	X	X	C
312	C	C	C	C	X	X	C
315	G	G	G	G	X	X	G
327	A	A	A	A	X	X	A
340	A	R	A	A	X	X	A
343	A	R	A	A	X	X	A
344	T	Y	M	T	X	X	T
350	A	A	A	W	X	X	A
354	:	:	:	:	X	X	:
355	:	:	:	:	X	X	:
356	C	C	C	C	X	X	C
360	T	T	Y	T	X	X	T
368	C	Y	Y	C	X	X	C
371	A	A	A	A	X	X	A
372	A	R	R	A	X	X	A
379	T	T	T	T	X	X	T
380	G	G	G	G	X	X	G
381	G	G	G	G	X	X	G
382	T	T	T	T	X	X	T
383	A	A	A	A	X	X	A
384	C	C	C	C	X	X	C
385	T	T	T	T	X	X	T
406	T	T	T	Y	X	X	T
407	A	A	W	A	X	X	A
411	A	A	A	A	X	X	A
418	:	:	:	:	:	:	:
421	A	A	A	A	M	M	A

422	A	A	A	A	A	A	A

Position	EU195885.P	AMP_EU195885.P	AZA_EU195885.P	GAUD_EU195885.P	GRA_EU195885.P	GRI_EU195885.P	JAP_EU195885.P
423	T	T	T	T	T	W	T
424	A	A	A	A	A	W	A
429	C	C	C	C	C	H	C
430	T	T	T	T	T	Y	T
433	C	C	C	C	C	S	C
436	G	G	G	G	C	Y	G
437	T	T	T	T	T	T	T
439	C	C	C	C	C	C	C
445	T	T	T	T	W	A	T
446	T	T	T	T	W	T	T
466	G	G	G	G	G	G	R
478	C	C	C	C	R	T	C
481	C	C	M	C	Y	Y	C
484	T	T	T	W	T	T	T
486	C	C	C	C	C	A	C
490	T	T	T	T	T	T	T
493	C	C	C	C	M	C	C
509	C	C	C	C	C	C	G
523	T	T	T	T	T	T	T
527	G	G	G	G	G	G	S
556	T	T	T	T	C	C	T
574	C	C	C	C	C	C	C
577	C	C	C	C	T	T	C
592	C	C	C	C	C	Y	C
598	T	T	T	T	K	T	T
622	G	R	G	G	R	G	A
628	C	C	C	C	C	C	C
634	T	T	Y	T	T	T	T
664	T	T	T	T	Y	C	T
667	C	C	C	C	R	C	C
670	C	C	C	C	C	C	C
694	T	T	T	T	T	T	T
711	G	G	R	G	G	G	G
715	G	G	G	G	R	G	G
724	C	C	C	C	C	C	C
725	A	A	A	A	A	A	A
727	A	A	A	A	A	A	A
736	T	T	T	T	T	T	W
767	G	G	G	X	X	X	G

770	G	**G**		**G**		X		X		X		**G**	
781	A	R		**A**		X		X		X		**A**	
Total Differences					8		12		4		14	15	7

Position	EU195885.P	KAM_EU195885.P	MAI_EU195885.P	OLE_EU195885.P	SAS_EU195885.P	SIN_EU195885.P	TRA_EU195885.P	Number of Variants
107	A	R	X	X	A	X	A	1
110	T	T	X	X	W	X	T	1
136	C	C	X	X	Y	X	C	1
150	C	C	X	X	Y	X	C	1
168	T	W	X	X	T	X	T	1
179	T	T	X	X	K	X	T	1
180	T	T	X	X	T	X	G	1
189	A	A	X	X	W	X	A	1
194	T	K	X	X	T	X	T	1
201	T	K	X	X	T	X	T	1
260	C	C	X	X	C	X	C	1
264	A	A	X	X	A	X	A	1
276	C	C	C	X	C	C	C	1
288	C	C	C	C	C	C	C	1
306	Y	T	T	T	T	C	C	10
312	C	Y	C	C	C	T	C	2
315	G	G	G	G	G	C	G	1
327	A	A	A	A	A	A	G	1
340	A	A	A	A	A	A	A	1
343	A	A	A	A	A	A	A	1
344	T	T	T	T	T	T	T	2
350	A	A	A	A	A	A	A	1
354	:	:	:	:	:	T	:	1
355	:	:	:	:	:	T	:	1
356	C	C	C	C	M	C	C	1
360	T	T	T	T	T	T	T	1
368	C	C	C	C	C	C	C	2
371	A	A	A	A	R	A	A	1
372	A	A	A	A	A	A	A	2
379	T	T	T	T	T	:	T	1
380	G	G	G	G	G	:	G	1
381	G	G	G	G	G	:	G	1
382	T	T	T	T	T	:	T	1
383	A	A	A	A	A	:	A	1
384	C	C	C	C	C	:	C	1
385	T	T	T	T	T	:	T	1
406	T	T	T	T	T	T	T	1
407	A	A	A	A	A	A	A	1
411	A	A	A	A	M	A	A	1

418	:	:	:	:	:	A	:	1
421	A	A	A	A	A	A	A	2
422	A	A	A	A	W	A	A	1

Position	EU195885.P	KAM_EU195885.P	MAI_EU195885.P	OLE_EU195885.P	SAS_EU195885.P	SIN_EU195885.P	TRA_EU195885.P	Number of Variants
423	T	T	T	T	T	T	T	1
424	A	A	A	A	A	A	A	1
429	C	C	C	C	C	C	C	1
430	T	T	T	T	T	T	T	1
433	C	C	C	C	C	C	C	1
436	G	G	G	G	G	:	A	4
437	T	T	T	T	T	:	T	1
439	C	C	C	C	C	T	C	1
445	T	T	T	T	T	T	T	2
446	T	T	T	T	T	T	T	1
466	G	G	G	G	G	G	T	2
478	C	C	C	C	C	C	C	2
481	C	C	C	C	C	C	C	3
484	T	A	W	T	T	T	T	3
486	C	C	C	C	C	C	C	1
490	T	T	T	T	Y	T	Y	2
493	C	C	C	C	C	C	C	1
509	C	C	C	C	S	C	C	2
523	T	T	T	C	Y	T	T	2
527	G	G	G	G	G	G	G	1
556	T	T	T	C	Y	C	Y	6
574	C	C	C	C	C	G	C	1
577	C	C	C	C	Y	C	C	3
592	C	C	C	C	C	C	C	1
598	T	T	T	T	T	T	Y	2
622	G	G	G	G	R	A	R	6
628	C	C	C	C	Y	C	C	1
634	T	T	T	T	T	T	T	1
664	T	T	T	T	Y	T	T	3
667	C	C	C	C	C	C	Y	2
670	C	C	C	T	C	C	C	1
694	T	T	T	T	T	T	Y	1
711	G	G	G	G	G	G	G	1
715	G	G	G	G	G	G	G	1
724	C	C	C	C	C	C	Y	1
725	A	A	W	A	A	T	A	2

727	A	A	A	A	A	Y	A	1
736	T	T	T	T	W	T	T	2
767	G	S	G	G	G	G	G	1
770	G	G	X	G	G	K	G	1
781	A	A	X	A	A	A	A	1
Total Differences		8	3	4	19	22	12	

For research purposes only. Not intended for clinical use.

Date:		18-Apr-14
Project Name:	GQ214317.SPF	
Contig Name:	GQ214317.P	
Compare Bases to Consensus:	GQ214317.P	
Comparison Range:	Unfiltered	
Base Positions:	1 to 462	
Options:	Large gap insertions (10 or more bases) excluded. Matches to ambiguous reference positions included.	

Position	GQ214317.P	AMP_GQ214317.P	AZA_GQ214317.P	GAUD_GQ214317.P	GRA_GQ214317.P	GRI_GQ214317.P	JAP_GQ214317.P
7	A	X	X	X	X	X	X
10	T	X	X	X	X	X	X
26	C	X	X	C	X	X	X
34	C	X	X	C	X	X	X
40	A	X	X	A	X	A	X
50	C	G	S	C	C	S	G
62	C	C	Y	C	C	C	C
63	G	G	G	G	R	G	S
73	C	C	Y	Y	C	C	C
74	G	G	G	G	R	G	G
96	G	G	G	G	G	G	G
98	T	T	T	T	T	T	T
104	T	T	T	T	Y	T	T
112	G	G	R	G	G	G	G
120	G	G	G	R	G	G	G
129	T	T	T	C	T	T	T
139	T	T	T	T	T	T	T
174	G	G	K	T	T	K	G
181	C	C	M	C	C	C	C
182	G	G	R	G	G	G	G
185	C	C	S	C	C	Y	C
188	A	C	M	A	A	C	C
194	C	C	C	C	C	C	C
195	G	G	G	R	G	G	G
200	G	G	R	G	G	G	G
202	A	A	A	A	A	A	A
206	G	G	R	G	G	G	G
211	A	T	A	A	A	A	A
219	G	G	G	G	G	G	G
220	G	G	G	G	G	G	G

Position	GQ214317.P	AMP_GQ214317.P	AZA_GQ214317.P	GAUD_GQ214317.P	GRA_GQ214317.P	GRI_GQ214317.P	JAP_GQ214317.P
222	G	G	G	G	G	:	G
223	G	G	G	G	G	:	G
224	A	A	A	A	A	:	A
225	:	:	G	:	:	K	:
226	:	:	G	:	:	G	:
227	:	:	A	:	:	A	:
236	T	T	T	T	T	Y	T
238	G	G	G	S	G	G	R
240	G	A	R	G	G	R	R
249	G	G	G	G	G	S	G
256	G	G	G	G	G	G	G
259	G	G	G	G	G	G	G
261	G	G	G	G	G	G	G
271	T	T	T	T	Y	T	T
272	C	C	C	C	C	Y	T
285	T	T	T	C	Y	T	T
300	C	C	C	C	Y	T	C
314	C	C	C	C	C	T	Y
315	G	G	G	G	G	G	G
316	A	A	A	A	M	A	A
340	G	G	G	R	G	G	G
342	C	C	C	Y	C	C	C
365	G	G	G	A	A	G	G
367	A	A	A	A	A	A	R
372	T	A	T	T	T	T	T
380	A	A	A	A	A	W	A
381	G	G	G	G	G	G	G
397	A	G	G	A	A	G	G
398	G	G	G	G	G	G	G
399	G	G	G	G	G	G	G
421	C	C	X	C	C	X	C
423	G	G	X	X	G	X	G
437	G	X	X	X	X	X	G
Total Differences		6	16	10	9	18	9

Position	GQ214317.P	KAM_GQ214317.P	MAI_GQ214317.P	OLE_GQ214317.P	SAS_GQ214317.P	SIN_GQ214317.P	TRA_GQ214317.P	Number of Variants
7	A	A	A	A	R	G	G	3
10	T	T	Y	T	T	T	T	1
26	C	C	C	C	Y	T	T	3
34	C	C	C	C	C	C	T	1
40	A	A	A	A	A	G	A	1
50	C	C	C	C	S	G	G	7
62	C	C	Y	C	C	C	C	2
63	G	G	G	G	R	A	G	4
73	C	T	T	C	Y	C	C	5
74	G	G	G	G	R	G	G	2
96	G	G	G	G	G	G	T	1
98	T	T	T	T	T	T	Y	1
104	T	T	T	T	T	T	T	1
112	G	G	G	G	G	G	G	1
120	G	G	G	R	R	G	G	3
129	T	C	Y	Y	Y	T	T	5
139	T	T	T	T	T	T	Y	1
174	G	K	K	K	K	G	G	8
181	C	C	C	C	C	C	C	1
182	G	G	G	G	G	G	G	1
185	C	C	C	C	C	C	C	2
188	A	A	A	A	M	C	A	6
194	C	C	C	C	C	C	T	1
195	G	R	R	G	G	G	G	3
200	G	G	G	G	G	G	G	1
202	A	A	A	A	A	G	A	1
206	G	G	G	K	G	G	G	2
211	A	A	A	A	A	A	A	1
219	G	G	G	G	G	:	G	1
220	G	G	G	G	G	:	G	1

Position	GQ214317.P	KAM_GQ214317.P	MAI_GQ214317.P	OLE_GQ214317.P	SAS_GQ214317.P	SIN_GQ214317.P	TRA_GQ214317.P	Number of Variants
222	G	G	G	G	G	G	G	1
223	G	G	G	G	G	:	G	2
224	A	A	A	A	A	A	A	1
225	:	:	:	:	:	G	:	3
226	:	:	:	:	:	G	:	3
227	:	:	:	:	:	A	:	3
236	T	T	T	T	T	T	T	1
238	G	S	S	G	R	G	G	5
240	G	G	R	G	R	A	G	7
249	G	G	G	G	G	G	G	1
256	G	G	G	G	G	A	G	1
259	G	G	G	G	S	G	G	1
261	G	S	S	S	G	G	G	3
271	T	T	T	T	T	T	T	1
272	C	C	C	C	Y	C	C	3
285	T	C	C	Y	Y	T	T	6
300	C	C	C	C	C	C	C	2
314	C	C	C	C	C	C	C	2
315	G	G	G	G	G	A	G	1
316	A	A	A	A	A	A	A	1
340	G	R	R	G	G	G	G	3
342	C	Y	Y	C	C	C	C	3
365	G	A	A	A	R	G	G	6
367	A	A	A	A	A	A	A	1
372	T	T	T	T	T	T	T	1
380	A	A	A	A	A	A	A	1
381	G	G	G	G	G	C	G	1
397	A	A	A	A	R	A	G	6
398	G	G	G	G	G	G	R	1
399	G	G	G	G	G	T	G	1
421	C	C	C	C	C	C	Y	1
423	G	G	G	G	G	A	G	1
437	G	G	G	G	G	A	G	1
Total Differences		10	13	7	17	20	11	0

For research purposes only. Not intended for clinical use.

Date:		18-Apr-14
Project Name:	GQ461358.SPF	
Contig Name:	GQ461358.P	
Compare Bases to Consensus:	GQ461358.P	
Comparison Range:	Unfiltered	
Base Positions:	1 to 304	
Options:	Large gap insertions (10 or more bases) excluded. Matches to ambiguous reference positions included.	

Position	GQ461358.P	AMP_GQ461358.P	AZA_GQ461358.P	GAUD_GQ461358.P	GRA_GQ461358.P	GRI_GQ461358.P	JAP_GQ461358.P
17	A	X	T	A	X	X	X
29	A	X	G	A	A	X	A
31	G	X	G	G	S	X	G
35	G	X	G	T	T	X	T
53	A	A	A	A	A	A	A
55	T	T	T	T	T	T	T
64	C	C	C	C	C	C	S
69	T	T	T	T	T	T	K
73	A	A	A	G	A	A	R
75	G	G	G	G	G	G	G
76	T	T	T	T	T	T	T
77	C	C	C	C	C	C	Y
79	G	G	T	G	G	G	G
82	G	G	G	G	G	G	G
84	G	G	G	G	G	G	G
85	G	G	G	G	G	G	G
90	C	C	C	C	C	C	T
116	G	G	G	G	G	G	G
128	G	G	T	G	G	G	G
133	C	C	C	C	C	C	Y
140	T	T	T	T	Y	T	T
141	T	T	T	T	T	T	T
153	G	G	G	G	G	G	G
156	T	C	C	T	Y	T	T
159	T	T	C	T	T	C	T

Position	GQ461358.P	AMP_GQ461358.P	AZA_GQ461358.P	GAUD_GQ461358.P	GRA_GQ461358.P	GRI_GQ461358.P	JAP_GQ461358.P
169	A	A	A	A	A	A	A
171	T	T	T	T	T	T	T
173	T	C	T	T	T	T	T
174	G	G	G	G	G	G	G
178	G	G	G	T	K	G	T
179	T	T	C	T	T	C	T
185	T	T	C	T	T	T	T
186	A	A	A	A	W	A	A
190	G	G	G	G	G	G	G
199	T	T	T	T	Y	T	T
200	T	T	T	G	T	T	G
207	A	A	G	A	R	A	A
211	G	G	G	A	G	G	A
213	G	G	G	A	S	G	A
215	G	G	G	G	G	G	G
220	C	C	C	C	C	C	C
221	T	T	T	A	T	T	T
225	C	T	C	C	Y	C	C
229	T	:	T	T	T	T	T
230	G	G	G	G	G	G	G
235	T	T	T	T	T	T	T
252	G	G	G	G	R	A	G
262	G	G	G	G	G	G	G
268	G	G	X	G	G	G	G
276	T	C	X	T	T	T	T
278	A	A	X	A	M	A	A
292	G	G	X	A	G	X	G
Total Differences		5	9	8	12	3	11

Position	GQ461358.P	KAM_GQ461358.P	MAI_GQ461358.P	OLE_GQ461358.P	SAS_GQ461358.P	SIN_GQ461358.P	TRA_GQ461358.P	Number of Variants
17	A	A	X	X	X	X	X	1
29	A	A	X	X	X	G	G	3
31	G	G	X	G	G	G	G	1
35	G	T	X	G	G	G	G	4
53	A	A	X	A	M	A	A	1
55	T	T	X	T	T	Y	T	1
64	C	C	C	C	C	C	C	1
69	T	T	T	T	T	Y	T	2
73	A	G	G	A	A	A	A	4
75	G	G	G	G	G	G	A	1
76	T	T	T	T	T	T	W	1
77	C	C	C	Y	C	C	C	2
79	G	G	G	G	G	T	T	3
82	G	G	G	G	G	R	G	1
84	G	G	G	G	G	G	R	1
85	G	G	G	R	G	G	R	2
90	C	C	C	C	C	C	S	2
116	G	G	R	R	G	G	G	2
128	G	G	G	G	G	G	R	2
133	C	C	Y	Y	C	C	C	3
140	T	T	T	T	T	T	K	2
141	T	T	T	T	T	T	W	1
153	G	G	G	T	T	G	G	2
156	T	T	T	C	C	T	T	5
159	T	T	T	T	T	C	C	4

Position	GQ461358.P	KAM_GQ461358.P	MAI_GQ461358.P	OLE_GQ461358.P	SAS_GQ461358.P	SIN_GQ461358.P	TRA_GQ461358.P	Number of Variants
169	A	A	A	A	R	A	A	1
171	T	T	T	A	W	T	T	2
173	T	T	T	T	T	T	T	1
174	G	G	G	G	G	R	G	1
178	G	T	T	G	G	G	G	5
179	T	T	T	T	T	T	T	2
185	T	T	T	T	T	T	T	1
186	A	A	A	A	A	A	A	1
190	G	G	G	G	R	G	G	1
199	T	T	T	T	T	T	T	1
200	T	G	G	T	T	T	T	4
207	A	A	A	A	A	A	A	2
211	G	A	A	G	G	G	G	4
213	G	A	A	G	G	G	G	5
215	G	G	G	G	G	G	R	1
220	C	C	C	C	C	C	Y	1
221	T	W	W	T	T	T	T	3
225	C	C	C	T	R	C	C	4
229	T	T	T	T	T	T	T	1
230	G	G	G	G	R	G	G	1
235	T	T	T	G	G	T	T	2
252	G	G	G	G	G	A	A	4
262	G	G	G	A	A	G	G	2
268	G	G	G	G	G	R	G	1
276	T	T	T	T	T	T	T	1
278	A	A	A	C	C	A	A	3
292	G	A	A	G	G	G	G	3
Total Differences		8	9	11	11	9	14	0

For research purposes only. Not intended for clinical use.

Date:		22-Apr-14
Project Name:	GU399024.SPF	
Contig Name:	GU399024.P	
Compare Bases to Consensus:	GU399024.P	
Comparison Range:	Unfiltered	
Base Positions:	1 to 701	
Options:	Large gap insertions (10 or more bases) excluded. Matches to ambiguous reference positions included.	

Position	GU399024.P	AMP_GU399024.P	AZA_GU399024.P	GAUD_GU399024.P	GRA_GU399024.P	GRI_GU399024.P	JAP_GU399024.P
6	G	X	G	G	X	G	X
15	G	X	G	R	X	G	X
22	C	C	C	Y	X	Y	C
46	G	G	G	R	G	G	G
51	G	G	G	R	G	R	G
73	A	R	A	A	A	A	A
81	G	G	G	R	G	G	G
88	A	A	A	A	A	A	R
90	T	T	T	T	T	T	T
92	G	G	G	R	G	R	G
96	C	C	C	Y	C	C	C
108	C	C	C	C	C	C	C
111	A	A	A	A	A	A	A
117	C	Y	T	C	C	C	Y
118	G	G	G	G	G	G	G
135	C	C	C	C	C	C	C
138	C	C	C	C	C	C	C
161	C	C	C	Y	C	C	Y
162	G	G	G	G	G	S	S
164	C	C	C	C	C	C	C
168	C	C	C	C	C	C	C
175	C	C	C	C	C	C	C
180	C	C	C	C	C	C	C
194	C	C	C	C	C	M	M
202	A	A	A	A	G	R	R
203	C	C	C	C	C	S	C
206	C	C	C	C	C	C	C
208	G	G	G	G	G	G	G
214	G	G	G	R	G	G	G
245	A	A	A	A	A	A	A
249	G	G	G	G	G	G	G
259	G	S	C	G	G	G	G
261	G	G	G	G	G	R	G
272	C	C	C	C	C	C	C
306	G	G	G	G	G	G	G
327	T	K	T	T	T	T	T
335	G	G	G	G	G	R	G
345	G	S	G	G	G	G	G
351	C	Y	C	C	C	C	C
363	T	T	T	T	K	Y	T

Position	GU399024.P	AMP_GU399024.P	AZA_GU399024.P	GAUD_GU399024.P	GRA_GU399024.P	GRI_GU399024.P	JAP_GU399024.P
369	C	C	C	C	C	M	C
381	G	G	G	G	G	R	G
383	C	C	C	C	C	M	C
399	T	Y	T	T	T	T	T
441	T	T	T	Y	T	T	Y
459	C	S	C	C	C	C	C
483	G	G	G	G	G	S	G
486	G	G	G	G	G	G	G
499	A	R	R	A	A	A	A
508	A	A	A	A	A	A	A
516	:	:	:	:	:	:	:
517	:	:	:	:	:	:	:
518	:	:	:	:	:	:	:
520	G	G	G	G	G	G	R
529	G	R	G	G	G	G	G
534	G	S	S	G	G	G	G
538	C	C	C	C	C	C	C
539	G	G	G	G	G	G	R
555	G	G	G	T	G	G	K
557	G	G	G	R	G	G	R
577	G	S	G	G	G	G	G
579	G	G	G	G	G	R	G
589	A	A	A	A	A	A	A
601	T	:	T	T	T	T	T
602	A	:	A	A	A	A	A
605	G	G	G	G	G	G	G
610	G	G	G	G	G	S	G
613	T	T	T	T	T	T	T
624	G	A	G	G	G	G	G
625	A	A	A	A	A	A	A
626	A	A	A	A	A	A	A
640	C	T	C	C	C	C	C
649	A	A	A	A	A	A	A
662	T	T	T	T	T	T	T
664	G	C	G	G	R	G	G
668	T	T	T	G	T	T	G
680	C	C	C	C	C	C	C
688	A	R	A	A	A	A	A
689	T	T	T	T	T	T	T
Total Differences		18	4	13	3	16	12

Position	GU399024.P	KAM_GU399024.P	MAI_GU399024.P	OLE_GU399024.P	SAS_GU399024.P	SIN_GU399024.P	TRA_GU399024.P	Number of Variants
6	G	G	G	G	G	R	G	1
15	G	G	R	R	R	G	G	4
22	C	C	Y	C	C	C	C	3
46	G	R	R	G	G	G	G	3
51	G	G	G	G	G	G	G	2
73	A	A	A	A	A	A	A	1
81	G	G	G	G	G	G	G	1
88	A	A	A	R	R	A	A	3
90	T	T	T	T	T	Y	T	1
92	G	R	R	G	G	G	G	4
96	C	Y	Y	C	C	C	C	3
108	C	C	C	Y	C	C	C	1
111	A	W	W	A	A	A	A	2
117	C	C	C	Y	Y	C	C	5
118	G	G	K	G	G	G	G	1
135	C	C	C	C	C	C	Y	1
138	C	C	C	C	C	C	T	1
161	C	Y	Y	Y	C	C	C	5
162	G	G	G	G	G	G	G	2
164	C	C	C	Y	C	C	C	1
168	C	C	C	C	S	Y	C	2
175	C	C	C	S	C	C	C	1
180	C	C	C	C	Y	C	C	1
194	C	C	C	C	C	C	C	2
202	A	A	A	A	A	A	A	3
203	C	C	C	S	S	C	C	3
206	C	C	C	Y	C	C	C	1
208	G	G	G	G	G	G	R	1
214	G	G	G	G	G	G	G	1
245	A	A	A	A	A	R	A	1
249	G	G	G	R	G	G	G	1
259	G	G	G	S	G	C	G	4
261	G	G	G	G	G	G	G	1
272	C	C	C	C	C	Y	C	1
306	G	G	G	G	G	R	G	1
327	T	T	T	T	T	T	T	1
335	G	G	G	G	G	G	G	1
345	G	G	G	G	G	G	G	1
351	C	C	C	C	C	C	C	1

363	T	T	T	T	T	T	T	2
369	C	C	C	C	C	C	C	1

Position	GU399024.P	KAM_GU399024.P	MAI_GU399024.P	OLE_GU399024.P	SAS_GU399024.P	SIN_GU399024.P	TRA_GU399024.P	Number of Variants
381	G	G	G	G	G	G	G	1
383	C	C	C	M	C	C	C	2
399	T	T	T	T	T	T	T	1
441	T	Y	Y	Y	T	T	T	5
459	C	C	C	C	C	C	G	2
483	G	G	G	G	G	G	G	1
486	G	G	G	G	G	G	T	1
499	A	A	A	A	A	A	A	2
508	A	A	A	A	A	A	M	1
516	:	:	:	:	:	:	R	1
517	:	:	:	:	:	:	G	1
518	:	:	:	:	:	:	C	1
520	G	G	G	R	M	G	G	3
529	G	G	G	G	G	G	G	1
534	G	G	G	G	G	G	G	2
538	C	C	C	C	C	C	G	1
539	G	G	G	R	M	G	G	3
555	G	T	T	K	G	G	G	5
557	G	R	R	R	G	G	G	5
577	G	G	G	G	G	R	G	2
579	G	G	G	G	G	G	G	1
589	A	A	A	W	A	A	A	1
601	T	T	T	T	T	T	:	2
602	A	A	A	A	A	A	:	2
605	G	G	G	G	G	G	R	1
610	G	G	G	G	G	G	G	1
613	T	T	T	T	T	T	Y	1
624	G	G	G	G	G	R	G	2
625	A	A	A	A	M	A	A	1
626	A	A	A	A	W	A	A	1
640	C	C	C	C	C	C	C	1
649	A	A	A	A	A	A	G	1
662	T	T	T	T	T	T	Y	1
664	G	G	G	G	G	G	G	2
668	T	G	G	T	T	T	T	4
680	C	C	C	C	C	C	Y	1
688	A	A	A	A	A	A	G	2

689	T	T		T		T		T		T		T		C		1
Total Differences			9		12		18		10		9		19			0

For research purposes only. Not intended for clinical use.

Date:		22-Apr-14
Project Name:	JQ398741.SPF	
Contig Name:	JQ398741.P	
Compare Bases to Consensus:	JQ398741.P	
Comparison Range:	Unfiltered	
Base Positions:	1 to 398	
Options:	Large gap insertions (10 or more bases) excluded. Matches to ambiguous reference positions included.	

Position	JQ398741.P	AMP_JQ398741.P	AZA_JQ398741.P	GAUD_JQ398741.P	GRA_JQ398741.P	GRIJ_JQ398741.P	JAP_JQ398741.P
8	C	C	C	X	X	X	G
13	C	C	C	X	X	X	C
17	C	C	C	X	X	X	C
21	G	G	G	X	X	X	S
24	C	C	C	X	X	X	C
25	C	C	C	X	X	X	C
26	A	M	M	X	A	X	M
31	G	G	G	X	G	X	G
40	C	C	C	C	Y	X	C
41	T	T	T	T	T	X	Y
54	T	T	T	T	T	X	K
55	T	A	T	T	T	X	T
56	A	A	A	A	A	X	A
62	G	G	G	G	G	X	R
71	G	G	G	G	G	G	S
79	A	:	A	A	A	A	A
80	G	:	G	G	G	G	G
81	A	:	A	A	A	A	A
82	:	T	:	:	:	:	:
83	:	G	:	:	:	:	:
85	G	A	G	G	G	G	G
86	:	G	:	:	:	:	:
91	:	:	:	:	:	:	:
92	:	:	:	:	:	:	:
93	:	:	:	:	:	:	:
94	:	:	:	:	:	:	:
95	:	:	:	:	:	:	:
96	:	:	:	:	:	:	:
97	T	C	C	T	T	T	T
100	C	C	C	C	C	C	C
112	C	C	C	C	C	C	C
113	G	G	G	G	S	G	G
116	C	C	C	C	M	C	C
122	C	C	C	C	S	C	C
124	G	G	G	G	S	G	G
125	G	T	A	G	G	G	G
135	T	T	T	T	T	T	T
143	T	K	T	T	T	T	T
150	T	K	T	T	T	T	K
151	T	G	T	T	T	T	T
153	C	S	C	C	C	C	S

154	T	T	Y	T	T	T	T
156	A	A	M	A	A	A	A
157	C	C	C	C	C	C	Y
Position	JQ398741.P	AMP_JQ398741.P	AZA_JQ398741.P	GAUD_JQ398741.P	GRA_JQ398741.P	GRIJ_JQ398741.P	JAP_JQ398741.P
160	C	S	C	C	C	C	S
161	G	G	G	G	G	G	G
163	C	S	C	C	C	C	C
165	A	A	A	A	A	A	A
168	A	A	A	A	A	A	A
170	T	T	K	T	T	T	T
172	C	C	G	C	C	C	C
180	A	R	A	A	A	A	A
189	C	C	C	C	C	C	C
190	G	G	G	G	G	G	R
195	T	T	T	T	T	T	K
197	A	A	A	A	A	A	A
202	G	A	A	G	A	G	G
203	C	C	C	C	C	C	C
204	A	A	A	A	A	A	M
224	G	G	G	G	G	G	A
227	A	A	M	A	A	A	A
238	C	S	C	C	C	C	S
241	T	T	T	T	T	T	K
245	G	G	T	G	G	G	G
246	C	A	C	C	C	C	C
248	A	A	A	A	A	A	A
259	G	R	G	G	G	G	G
298	G	T	G	G	G	G	G
301	T	T	C	T	T	T	T
303	T	C	T	T	T	T	T
308	G	G	G	G	G	G	S
311	G	A	G	G	G	G	G
315	G	G	G	G	G	G	G
316	A	A	A	A	A	A	A
317	G	G	G	G	G	G	G
318	G	G	:	G	G	G	G
319	A	A	:	A	A	A	A
320	G	G	:	G	G	G	G
321	:	:	R	:	:	G	:
322	:	:	A	:	:	A	:
324	:	:	G	:	:	G	:
326	A	A	A	A	A	A	A

360	G	G	G	G	K	G	G
373	T	T	T	T	Y	T	T
383	T	T	T	T	T	W	T
387	G	G	G	G	G	S	G
Total Difference		25	17	0	8	5	18

Position	JQ398741.P	KAM_JQ398741.P	MAI_JQ398741.P	OLE_JQ398741.P	SAS_JQ398741.P	SIN_JQ398741.P	TRA_JQ398741.P	Num of Var
8	C	G	X	X	X	C	C	2
13	C	S	X	X	X	C	C	1
17	C	S	X	X	X	C	C	1
21	G	G	X	X	X	G	G	1
24	C	Y	C	X	X	C	C	1
25	C	C	C	X	X	C	M	1
26	A	M	M	A	X	M	A	6
31	G	G	G	G	G	G	T	1
40	C	C	C	C	C	C	C	1
41	T	T	T	T	T	T	T	1
54	T	T	K	T	T	T	T	2
55	T	T	T	T	T	T	T	1
56	A	R	R	A	A	A	A	2
62	G	G	G	G	G	G	G	1
71	G	G	G	G	G	G	G	1
79	A	A	A	A	A	A	A	1
80	G	G	G	G	G	G	G	1
81	A	A	A	A	A	A	A	1
82	:	:	:	:	:	A	:	2
83	:	:	:	:	:	G	:	2
85	G	G	G	G	G	A	G	2
86	:	:	:	:	:	G	:	2
91	:	:	:	:	:	A	A	2
92	:	:	:	:	:	G	:	1
93	:	:	:	:	:	A	:	1
94	:	:	:	:	:	A	:	1
95	:	:	:	:	:	G	:	1
96	:	:	:	:	:	A	:	1
97	T	T	T	T	T	C	C	4
100	C	C	S	C	C	C	C	1
112	C	C	C	C	M	C	C	1
113	G	G	G	G	G	G	G	1
116	C	C	C	C	C	C	C	1
122	C	C	C	C	C	C	C	1
124	G	G	G	G	G	G	G	1
125	G	G	G	G	G	A	A	4
135	T	T	T	T	T	T	C	1
143	T	T	T	T	T	T	T	1
150	T	T	T	T	T	T	T	2
151	T	T	T	T	T	T	T	1

153	C	C	C	C	C	C	C	2
154	T	T	T	T	T	T	T	1
156	A	A	A	A	A	A	A	1
Position	JQ398741.P	KAM_JQ398741.P	MAI_JQ398741.P	OLE_JQ398741.P	SAS_JQ398741.P	SIN_JQ398741.P	TRA_JQ398741.P	Num of Var
157	C	C	C	C	C	C	C	1
160	C	S	S	C	C	C	C	4
161	G	G	G	G	G	A	G	1
163	C	C	C	C	C	C	C	1
165	A	A	A	A	A	G	A	1
168	A	A	A	A	A	A	G	1
170	T	T	T	T	T	T	T	1
172	C	C	C	C	C	C	C	1
180	A	A	A	A	A	A	A	1
189	C	C	C	C	C	G	C	1
190	G	G	G	G	G	G	G	1
195	T	K	T	T	T	T	T	2
197	A	R	A	A	A	A	A	1
202	G	G	G	G	G	G	G	3
203	C	S	C	C	C	C	C	1
204	A	A	A	A	A	A	A	1
224	G	R	R	G	G	A	A	5
227	A	A	A	A	A	A	A	1
238	C	S	S	C	C	C	C	4
241	T	K	K	T	T	T	T	3
245	G	G	G	G	G	G	G	1
246	C	C	C	C	C	C	C	1
248	A	A	A	R	A	A	A	1
259	G	G	G	G	G	G	G	1
298	G	G	G	G	G	G	G	1
301	T	T	T	T	T	T	T	1
303	T	T	T	X	T	T	T	1
308	G	G	G	X	G	G	G	1
311	G	G	G	X	G	G	G	1
315	G	G	G	X	:	G	G	1
316	A	A	A	X	:	A	A	1
317	G	G	G	X	:	G	G	1
318	G	G	G	X	X	G	G	1
319	A	A	A	X	X	A	A	1
320	G	G	G	X	X	G	G	1
321	:	:	:	X	X	:	:	2
322	:	:	:	X	X	:	:	2
324	:	:	:	X	X	:	A	3

326	A	R		R		X		X		A		:		3
360	G	G		G		X		X		G		G		1
373	T	T		T		X		X		T		T		1
383	T	T		T		X		X		T		T		1
387	G	G		G		X		X		G		G		1
Total			14		9		1		4		17		10	0

For research purposes only. Not intended for clinical use.

Date:		22-Apr-14
Project Name:	JX657332.SPF	
Contig Name:	JX657332.P	
Compare Bases to Consensus:	JX657332.P	
Comparison Range:	Unfiltered	
Base Positions:	1 to 200	
Options:	Large gap insertions (10 or more bases) excluded. Matches to ambiguous reference positions included.	

Position	JX657332.P	AMP_JX657332.P	AZA_JX657332.P	GAUD_JX657332.P	GRA_JX657332.P	GRI_JX657332.P	JAP_JX657332.P
23	A	G	A	A	A	A	A
36	G	G	R	G	G	G	G
72	G	G	R	G	G	G	G
93	G	G	R	G	G	G	G
124	A	A	A	A	A	A	A
151	A	G	A	A	A	A	A
160	C	C	C	C	Y	C	C
Total Differences		2	3	0	1	0	0

Position	JX657332.P	KAM_JX657332.P	MAI_JX657332.P	OLE_JX657332.P	SAS_JX657332.P	SIN_JX657332.P	TRA_JX657332.P	Number of Variants
23	A	X	A	A	X	A	A	1
36	G	X	G	G	X	G	G	1
72	G	X	G	G	G	G	G	1
93	G	X	G	G	G	G	G	1
124	A	R	A	A	A	A	A	1
151	A	A	A	A	A	A	A	1
160	C	C	C	C	C	C	C	1
Total Differences		1	0	0	0	0	0	0

For research purposes only. Not intended for clinical use.

Date:		22-Apr-14
Project Name:	KC149523.SPF	
Contig Name:	KC149523.P	
Compare Bases to Consensus:	KC149523.P	
Comparison Range:	Unfiltered	
Base Positions:	1 to 329	
Options:	Large gap insertions (10 or more bases) excluded. Matches to ambiguous reference positions included.	

Position	KC149523.P	AMP_KC149523.P	AZA_KC149523.P	GAUD_KC149523.P	GRA_KC149523.P	GRI_KC149523.P	JAP_KC149523.P
37	A	X	A	A	A	G	X
43	G	A	G	G	G	G	G
50	T	C	T	T	T	T	T
62	C	C	C	C	C	C	C
66	C	C	A	C	A	C	C
68	C	C	C	Y	C	C	C
69	C	C	C	C	C	C	C
72	C	C	C	C	Y	C	C
76	C	C	C	C	T	C	C
91	T	:	T	T	T	T	T
99	C	C	C	C	M	C	C
103	T	T	K	T	T	T	T
126	C	C	Y	C	C	C	C
128	T	T	T	T	Y	T	T
174	T	T	T	T	T	T	T
183	C	A	C	C	C	C	C
197	C	C	C	C	C	C	C
210	G	G	R	G	G	G	G
217	G	:	:	:	G	G	X
218	T	:	:	:	T	T	X
219	A	:	:	:	A	A	X
220	G	:	:	:	G	G	X
221	G	:	:	:	G	G	X
222	G	:	:	:	G	G	X
223	A	:	:	:	A	A	X
224	A	:	:	:	A	A	X
225	G	:	:	:	G	G	X
226	T	:	:	:	T	T	X
227	T	:	:	:	T	T	X
228	A	:	:	:	A	A	X
229	C	:	:	:	C	C	X
230	C	:	:	:	C	C	X
231	A	:	:	:	A	A	X
236	T	C	T	T	T	T	X
290	T	T	T	X	X	G	X
Total Differences		20	19	16	5	2	0

Position	KC149523.P	KAM_KC149523.P Incomplete	MAI_KC149523.P Incomplete	OLE_KC149523.P Incomplete	SAS_KC149523.P Incomplete	SIN_KC149523.P Incomplete	TRA_KC149523.P Incomplete	Number of Variants
37	A	A	X	A	A	A	X	1
43	G	G	G	G	G	G	T	2
50	T	T	T	T	T	T	T	1
62	C	C	C	C	C	Y	C	1
66	C	C	C	C	C	C	Y	3
68	C	C	C	C	C	C	C	1
69	C	C	C	C	C	Y	C	1
72	C	C	C	C	C	C	C	1
76	C	C	C	C	C	C	C	1
91	T	T	T	T	T	T	T	1
99	C	C	C	C	C	C	C	1
103	T	T	T	T	T	T	T	1
126	C	C	C	C	C	C	C	1
128	T	T	T	T	T	T	T	1
174	T	T	K	T	T	T	T	1
183	C	C	C	C	C	C	C	1
197	C	C	S	C	C	C	C	1
210	G	G	G	G	G	G	G	1
217	G	A	X	G	G	:	G	5
218	T	T	X	T	T	:	T	4
219	A	G	X	A	A	:	A	5
220	G	G	X	G	G	:	G	4
221	G	X	X	G	G	:	G	4
222	G	X	X	G	G	:	G	4
223	A	X	X	A	A	:	A	4
224	A	X	X	A	A	:	A	4
225	G	X	X	G	G	:	G	4
226	T	X	X	T	T	:	T	4
227	T	X	X	T	T	:	T	4
228	A	X	X	A	A	:	A	4
229	C	X	X	C	C	:	C	4
230	C	X	X	C	C	:	C	4
231	A	X	X	A	A	:	A	4
236	T	X	X	T	T	T	T	1
290	T	X	X	T	T	T	X	1
Total Differences		2	2	0	0	17	2	0