Genomic Characterization of *Citrus* Fruits
Analysis of the ITS1 region in a variety of *Citrus* fruits.

Mark Pan

Department of Biological Sciences, Fordham University
Bronx NY, 10458

Abstract:

The genomic characterization and the phylogeny of the genus *Citrus* is still an ongoing debate among biologists. *Citrus* species hybridize easily and their hybrid offspring also readily hybridize with one another, which generate many complications to the study of phylogeny. This project aims to establish a preliminary understanding of *Citrus* phylogeny. The genomic characterization was done using the ITS1 region, which is commonly used for taxonomy in higher plants. A phylogenetic tree was then constructed using sequences from eight samples along with published sequences.

Introduction:

Citrus is a genus of flowering trees and shrubs nested in the *Rutacea* family, which includes our common oranges, lemons, and other citrus fruits. It is estimated 131 million metrictons of citruses are produced each year world-wide (Curk, 2014). Although we frequently rely on citrus fruits for nutrition, very little is known about their genomic characterization. Even with our advancement in sequencing technology, *Citrus* phylogeny remain relatively unresolved and understudied.

Most citrus fruits are products of hybridization that are readily able to hybridize with one another (Ramadugu, 2013). In addition, many ancient and modern civilizations have inconsistent propagation methods for citrus fruits, which makes lineage and taxonomy even more difficult. However, through comparisons of SNPs on chromosome 2, four clusters of ancestor were

characterized (Curk, 2014). (Figure 1) These ancestor species are Mandarin (*C. reticulata*), Pomelo (*C. maxima*), Citron (*C. medica*), and Papeda (*C. micrantha*). The hybrid cross between these four ancestors is what gave rise to most citrus fruits today. Many citrus hybrid crosses were documented by cultivators (Table 1), but some were lost overtime and unknown to modern taxonomist.

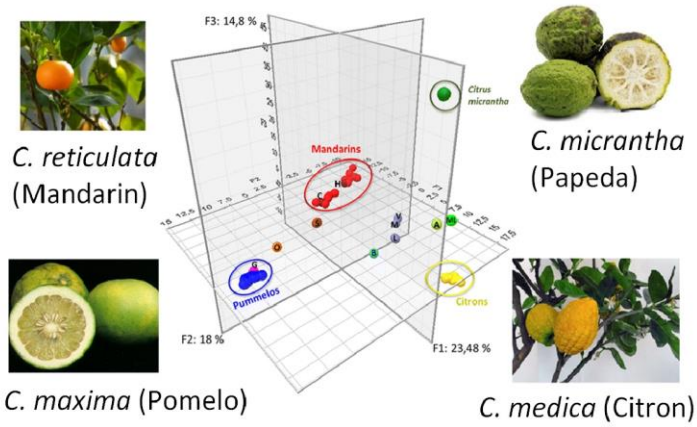

Figure 1: Comparison result of SNPs on chromosome 2 shows four distinctive ancestors.

Hybrid	Hybrid Cross
C. sinensis (Orange)	*C. reticulata* (Mandarin) x *C. maxima* (Pomelo)
C. clementina (Clementine)	*C. reticulata* (Mandarin) x *C. Sinensis* (Orange)
C. latifolia (Persian Lime)	*C. aurantifolia* (Key Lime) x *C. limon* (Lemon)
C. aurantifolia (Key Lime)	*C. micrantha* (Papeda) x *C. medica* (Citron)
C. limon (Lemon)	*C. maxima* (Pomelo) x *C. aurantium* (Bitter Orange)
C. meyeri (Meyer Lemon)	*C. limon* (Lemon) x *C. reticulate* (Mandarin)
C. paradisi (Grapefruit)	*C. sinesis (Orange)* x *C. maxima* (Pomelo)

Table 1: Known citrus hybrid crosses documented by cultivators.

The goal of this project is to gain a preliminary understanding of *Citrus* phylogeny and identify potential challenges in the field. Eight samples of commonly available citrus fruits are characterized in the internal transcribed spacer 1 (ITS1) region. The ITS1 region is an intronic sequence that is highly species specific. The sequence is located between the exon of 18s and

5.8s ribosomal gene, which is why this region is often utilized in barcoding studies and genomic characterization of plants (Tao, 2016).

Methods:

Obtaining samples and DNA extraction:

Eight citrus samples were purchased from local grocery stores on Arthur Ave. near Fordham University. The rinds of fruits were harvested and lysed by bead-beating (ceramic bead with garnet) in a Fast-Prep machine at 12.0 m/s for 40s. The DNA is then extracted and eluted using FastDNA SPIN Kit for Soil. The eight citrus fruits were orange, clementine, Persian lime, lemon/citron, grapefruit, mandarin, blood orange, and a repeat of orange from a different company. (Table 2)

	Samples	Company/Brand	Origin
A	Orange	River Pride	Florida
B	Clementine	LGS	Morocco
C	Persian Lime	Vision import group	Mexico
D	Lemon/citron	Sunkist	USA
E	Grapefruit		
F	Mandarin	Halos	USA
G	Orange	Homegrown Organic Farms	USA
H	Blood orange	Homegrown Organic Farms	USA

Table 2: Citrus samples, their corresponding company/brand, and the locations they were grown.

PCR and DNA visualization:

The ITS1 region of extracted DNA were amplified via PCR reaction with GoTaq Green Master Mix by Promega. Each PCR reaction contains 0.5µl (5 pmol) of forward and reverse primers, 8µl of dH_2O, and 10µl of GoTaq Green Master Mix. Degenerate primers ITS-p5 and

ITS-u2 from Tao et al. (2016) were used to amplify ITS1 (Figure 2) with 58℃ annealing temperature for 30s and 72℃ elongation temperature for 2min, because the expected product size is roughly around 482 base pairs. The primers that were used do not amplify fungal sequences. PCR products were then visualized on 1% agarose gel. (Figure 3)

Forward (ITS-p5): CCTTATCAYTTAGAGGAAGGAG

Reverse (ITS-u2): GCGTTCAAAGAYTCGATGRTTC

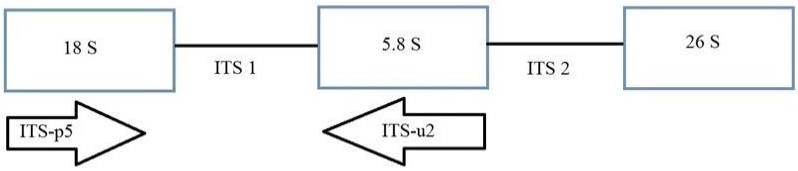

Figure 2: Schematic of ITS1 and the primers that used.

PCR purification and Bioinformatics:

PCR products were purified using QIAQuick PCR Purification kit (Qiagen, Venlo, Netherlands) then sequenced with T7 and Sp6 primers by Genewiz Inc (South Plainfield, NJ, USA). DNA sequences were edited and visualized using A Plasmid Editor (Davis, 2017). Sequences were also analyzed using Basic Local Alignment Search Tool (BLAST) nucleotide search on NCBI. Multiple sequence alignment and Neighbor-Joining Tree were generated using Clustal Omega (European Bioinformatics Institute, Cambridge, UK). Eight published ancestor sequences were added for comparison when generating phylogenetic tree.

Results:

The gel image of PCR products shows that all samples have expected band size. Unexpectedly, the Persian Lime (sample C) and the Lemon/Citron (sample D) samples have

double banding (Figure 3) and the trace file of their sequence also have double sequencing (not shown). This might due to the fact they are hybrids organisms.

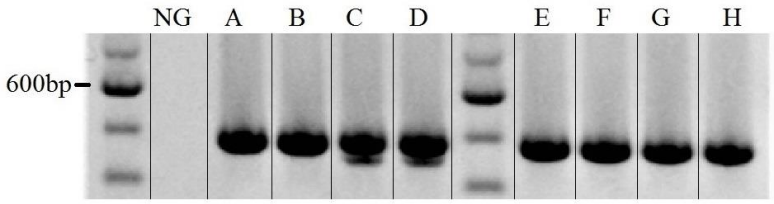

Figure 3: Gel visualization of PCR products. Note sample C (Persian Lime) and sample D (Lemon/Citron) exhibit double banding.

The resulting phylogenetic tree shows that sample C and D are closely clustered with the ancestor sequences of citron, which is expected since they are known hybrid products of citron and papeda or pomelo. Interestingly, sample F (Mandarin) did not cluster with the known ancestor sequences of mandarin, which would suggest sample F may not indeed be mandarin. (Figure 4) Although not much is known about the hybrid crosses of blood oranges, the tree suggests it is closely related to sweet orange and grapefruit.

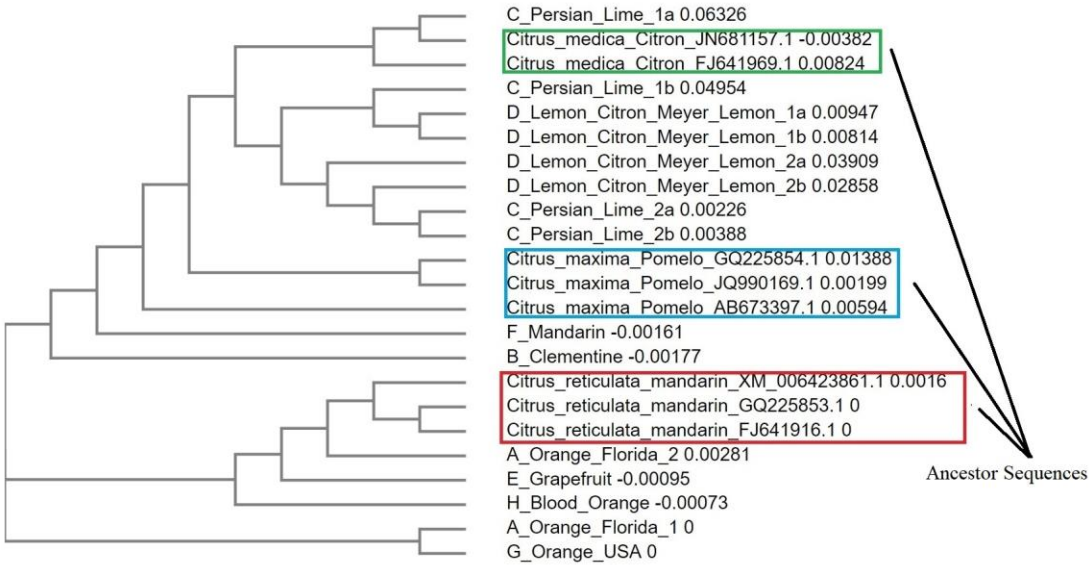

Figure 4: Neighbor-Joining Tree of samples and published ancestor sequences

Discussion:

In this project, the phylogenetic tree and PCR product gel are reflective of many difficulties in the field of *Citrus* phylogeny. The double banding of sample C and D suggests multiple alleles and their genetic characteristic as hybrid. However, the double bands may also reflect the ploidy of the fruit. Majority of citrus fruits are diploid, but limes and lemons are known to be tetra or polypoid due to propagation practices (Curk, 2016). It is impossible to know without further experimentation with digital PCR or molecular cloning.

Secondly, the Citrus Industry are also actively mislabeling and falsely distributing citrus fruits. Take example of the Lemon/Citron (D) sample, this name was given by the distributor which is not consistent with the known fact that citrons are ancestor species while lemons are hybrids. It is a clear indication that different varieties of same species can be mislabeled or distributed as another, which would explain why the Mandarin sample in the project did not cluster with the known sequences. According to literature, mandarins are often sold as tangerines if the variety is slightly more red or smaller (Krezdorn, 2014).

Lastly, study of citrus phylogeny is often complicated by false data and lack of documentation. For instance, many hybrid sequences such as sweet orange (*C. sinensis*) and lemon (*C. limon*) are being published as unique species sequences even though they are not unique species, this is problematic for phylogeny. In addition, citrus fruits like blood oranges are not well documented. It is known that blood oranges are hybrids but the exact hybrid crosses were lost overtime. Although blood oranges and sweet oranges are genotypically and phenotypically unique from one another, they are both categorized as *C. sinensis*. This is most likely due to the lack of study on blood oranges.

References:

Curk et al., 2014 Next generation haplotyping to decipher nuclear genomic interspecific admixture in *Citrus* species: analysis of chromosome 2. BMC Genetic (2014) 15:152

Tao et al., 2016 Barcoding the kingdom of Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. Molecular Ecology resources (2016) 16, 138-149

Ramadugu et al., 2013 A Six Nuclear Gene Phylogeny of *Citrus* (Rutacea) Taking into Account Hybridization and Lineage Sorting. PLOS 8:1

Xu et al., 2013 The draft genome of sweet orange (*Citrus sinensis*). Nature Genetics 45:59-66

Curk et al., 2016 Phylogenetic origin of limes and lemons revealed by cytoplasmic and nuclear markers. Annals of Botany 117(4):565-583

Krezdorn, 2014 Classification of Citrus 1-6