U.S. Citrus Genomics Initiative

A White Paper

Prepared by

U.S. Citrus Genomics Steering Committee

March 14, 2003

TABLE OF CONTENTS

1. Table of Contents ii
2. Executive Summary iii
3. Introduction 1
   - Citrus: a unique biology that complicates variety development 1
   - Citrus industry requirements 2
4. General Objectives 5
   - Exploitation of existing populations and germplasm resources 6
5. Specific Objectives 8
   - Genome characterization 8
   - Functional analysis of the citrus genome 11
   - Bioinformatics 14
   - Resource centers 16
   - International efforts 17
   - Workshops and Symposia 18
6. Conclusion 19
7. References 20
8. Appendices 22
   - Appendix A – Summary of Short-term Objectives (1-2 years) 23
   - Appendix B – Summary of Long-term Objectives (2-5 years) 24
   - Appendix C – Interim Steering Committee Members 25
   - Appendix D – Interim Advisory Committee Members 28

Cover Image – A citrus graphic of unknown authorship.

EXECUTIVE SUMMARY

This white paper presents the genomic tools required to solve the primary problems that limit the economical production and processing of citrus, the most economically significant fruit crop produced in the U.S. and world.

The U.S. citrus industry is faced with three constraints to economical production:

i. *Production areas are being severely challenged* – commercial rootstock and scion cultivars are susceptible to invasive pests, diseases and environmental problems.
ii. *Market share is decreasing* – competition from foreign producers.
iii. *Production costs are increasing* – declining farm labor force, encroachment of urban areas on the most productive land, and elevated environmental concerns and constraints that increase costs.

The primary solution to these constraints is new rootstock and scion varieties.

Citrus rootstock and scion varieties would contribute substantially to addressing these constraints since new rootstock and scion cultivars are the least expensive way to adapt citrus production to management schemes requiring less labor, fewer chemical inputs, and adapted to marginal sites. New citrus cultivars would also improve the competitiveness of U.S. citrus producers in local and international markets. However, there is a major problem with developing new citrus varieties - new variety development is slow and costly. The development of new citrus varieties has averaged 30 years.

Genomic technology can shorten the development time of new citrus varieties.

i. *Marker-assisted selection* – a technology to identify individual plants with certain characteristics prior to field evaluation. For example, a plant breeder could determine the color and flavor of fruit a newly germinated hybrid seedling would produce, or to identify its resistance or susceptibility to diseases. The seedling would not have to be planted in the field for 5-10 years, as is the case now, to determine mature plant characteristics.
ii. *Genetic engineering* – a technology to alter a single characteristic that is difficult, impractical, or impossible to alter through conventional methods, while still retaining variety integrity.
iii. *Global genomic expression* – a technology to determine rapidly and accurately overall tree health and fruit quality. The tool would assess the commercial potential of new varieties and provide growers, packers, and processors with information necessary to evaluate and intervene with the appropriate methods to achieve the maximum economic yield from the production grove or final citrus product.

Citrus genomic technology is absolutely essential for the sustainability and future viability of the U.S. citrus industry, and will enable citrus to remain the most significant fruit industry in the U.S., as it currently is today.

INTRODUCTION

Citrus is the most economically significant fruit crop produced in the United States (U.S.) and the world. U.S. production is localized to Florida, California, Arizona, Texas, Hawaii, Louisiana, and Alabama. It occupies an area of more than 1 million acres that represents greater than 50% of the area for all the other fruits produced in the U.S., including apple and grapes[1]. The total U.S. production of citrus fruits in the year 2000 was more than 17 million metric tons, compared to 5.3 million for apple, 7.6 million for grapes, and 6.2 million for all the other fruits. Total world citrus production is 90 million metric tons[2]. The annual on-tree value of citrus fruit production in the U.S. is estimated to be $2.2 billion[1].

The U.S. citrus industry is faced with at least three constraints to economical production of citrus. First, citrus production areas are being severely challenged by pest, disease, and environmental problems to which current commercial rootstock and scion cultivars are susceptible. Second, the U.S. citrus industry is experiencing a decrease in market share due to competition from foreign producers in the fresh and processed markets. Third, economical production of citrus in the U.S. is severely threatened by a declining farm labor force, encroachment of urban areas on the most productive land, and elevated environmental concerns and constraints that increase production costs.

Genomic technology will be an integral component to removing these constraints. Applications will include: 1) more rapid and efficient production of new rootstock and scion cultivars that meet changing consumer preferences; additionally, new cultivars provide the least expensive way to adapt citrus production to management schemes requiring less labor, fewer chemical inputs, and marginal sites; 2) to alter one or a few undesirable characteristics in an otherwise acceptable variety; 3) accurate and rapid assessment of plant health using global genomic expression of genes, proteins, and/or important metabolites, and closely associated methods to 4) rapidly identify pests and diseases.

CITRUS: A UNIQUE BIOLOGY THAT COMPLICATES VARIETY DEVELOPMENT

Citrus is grown as a combination tree composed of the fruit-producing scion variety bud-grafted onto a rootstock variety adapted to the soil and environment of the local production area. Rootstock and scion variety development is inherently costly in citrus because it is a tree crop requiring both large field acreage and many years to adequately evaluate field performance. For example, the time required to develop new citrus varieties developed through the USDA's variety development program has averaged 28 years for scions and 35 years for rootstocks[3]. In addition to the long-term nature of tree breeding, rootstock and scion variety development in citrus is difficult because of four characteristics of citrus biology. First, *Citrus* reproductive biology is

[1] National Agricultural Statistics Services-USDA, 2002.
[2] Food and Agriculture Organization of the United Nations, Economic and Social Department (ES), Commodities and Trade Division (ESC), October, 2002.
[3] Personal communication with Dr. C. Jack Hearn, USDA citrus scion breeder from 1962-1995.

unusual in that nearly all commercial citrus types such as sweet oranges, grapefruit, lemons, and some mandarins reproduce by facultative apomixis, specifically adventive embryony (Webber and Batchelor, 1943). Apomictic types produce from 5% to 100% (Frost and Soost, 1979) asexual seedlings and are very inefficient seed parents for use in conventional breeding. Second, self-and cross-incompatibility, as well as partial to complete pollen sterility, is common in *Citrus*. Third, juvenility of most citrus germplasm lasts 5-10 years prior to the first flower. Fourth, commercial citrus species (e.g., sweet oranges, grapefruit, lemons, and limes) are complex hybrids of the true citrus species (i.e., mandarin, pummelo, and citron). Most of the many individual varieties within these commercial citrus species arose via somatic mutations. This combination of large land areas, many years of evaluation, apomixis, sexual incompatibility, prolonged juvenility, and complex hybrid nature has hampered variety improvement efforts.

The development and availability of genomic tools is essential for increasing the efficiency of conventional citrus variety development methods. These tools, when combined with progress made in developing monoembryonic seed parents, an increased understanding of the interspecific hybrid origins of commercially important citrus types, and the development of *in vitro* technologies required for genetic transformation and somatic hybridization, indicate that citrus variety development is more feasible now than any time in the past.

CITRUS INDUSTRY REQUIREMENTS

Economics - Successful citrus industries are based on cultivars that are well adapted to the climate and other features of a growing region, such that productivity is high and the fruit exhibits those characteristics desired most by consumers. Citrus trees are capable of survival in a wide range of locations and environments, but the regions where adaptation of specific cultivars is greatest are the ones that come to dominate production and marketing. However, superimposed on this general rule are a number of economic factors that influence whether growers, processors, packers and shippers can make a profit, thus ensuring economic survival of localized industries. Production costs vary from region to region, based on such factors as the costs for land and labor, social and/or economic costs associated with environmental protection and the legal rights of the labor force, and the costs of material inputs such as fertilizers, pesticides and fungicides, and water availability. As the global competition for citrus markets increases, it becomes more critical for growers to have, and industries to be based upon, the most genetically advanced, superior cultivars possible. The survival of the great citrus industries in the U.S. is critically dependent on the availability of such improved plant materials in the future. The costs of producing citrus in the U.S. exceed those from nearly every other citrus producing country in the world. Genetically advanced cultivars can provide some advantage to U.S. citrus industries that may help to overcome the disadvantage of high production costs. Citrus production and marketing is a global enterprise, especially when processed sweet oranges are considered. However, the availability of refrigerated trans-oceanic transportation also allows fresh citrus fruit to be shipped throughout the world, and the citrus industries in some regions of the world are primarily focused on the fresh markets. The kinds of

genetic improvements desired for processed or fresh fruit are not always the same, but several commonalities can be found, such as resistance to diseases and pests, tolerance of adverse or less than optimum environmental conditions, fruit quality improvements, and extension of maturity and harvesting seasons. Following are brief descriptions of some of the kinds of improvements that are desired by, or would be valuable to, the U.S. citrus industries. Given the unique reproductive biology and the impediments to targeted citrus cultivar improvements, many of these requirements may be difficult or even impossible to achieve without the application of genomic technologies.

Pest and disease resistance – Pre- and postharvest host pest resistance may be a particularly difficult objective to achieve, because there are very few sources of genetic resistance among citrus species or related genera. The number of insect pests attacking citrus crops in the U.S. is increasing, and promises to continue to do so in the future. In recent years, for example, the Florida citrus industry has been plagued by the increasing proliferation and distribution of root weevils, most notably *Diaprepes*; this pest has recently been found in Texas as well. The larvae of this insect feed on citrus roots and dramatically depress yields, and in severe infestations can result in tree death. There seems to be no genetic resistance in the common citrus species, though there may be some in related genera. It may be possible to develop resistance to this insect by utilizing genes from other organisms. Brown citrus aphid invaded North America a few years ago, and its threat is related to its vectoring capacity for citrus tristeza virus (CTV). Since it has been in Florida, the increase of severe decline strains of this virus has essentially spelled the doom for sour orange rootstock, which has been one of the most widely adapted and productive citrus rootstocks in many growing regions of the world. Recently the citrus psyllid has made its way to Florida. It can cause some problems in the growth of trees, especially young trees. It represents a much greater threat, however, because it is the vector for citrus greening, a very serious and debilitating disease common to Asia, but not yet found in the U.S. Cultivars resistant to these and other insect pests, present or not yet introduced, would be a significant accomplishment that would both decrease crop protection costs and crop losses, as well as environmental contamination with pesticides.

The number of diseases present in the U.S. citrus industries likewise presents a production challenge, though genetic resources for resistance to many of the more serious diseases have been identified. Many diseases occur outside of the U.S. and represent serious threats to continued production. The fact that genetic resistance exists to many of the important diseases, both within citrus and among the closely related genera, might suggest that breeding for disease resistance is an option for the future. However, most commercial cultivar groups are not amenable to improvement through standard breeding techniques, and it is likely that the benefits of genomic science applied to citrus improvement may first be seen in this area using both host and non-host resistance strategies. Already, several research groups within the U.S. are engaged in cloning genes for resistance to CTV, citrus nematodes, and potentially scores of other diseases. Such work is costly and takes time, but in the end the ability to target improvements toward citrus cultivars genetically resistant to debilitating diseases

makes such work worthy of the investment. Given that some diseases present currently or in the future, such as CTV, citrus greening, citrus variegated chlorosis, and citrus canker, potentially could destroy the economic viability of the U.S. citrus industries; the application of genomics to develop solutions prior to the demise of our production base can fill a very critical need for the near future.

Tolerance of environmental conditions - As recently as thirteen years ago, nearly 25% of the Florida citrus industry and almost 100% of the Texas industry were eliminated within two days, not by diseases or insects but by freezing temperatures. Cold has always limited the regions in which citrus could be reliably produced, and is likely to cause serious crop losses again. No production area in the Continental U.S. is immune to this risk. There have been some recent advances in understanding the genetic control of cold tolerance within citrus and the related genera. These advances were made possible through the application of genetic mapping and genomic technologies. No solutions have yet been provided, but continued work in this area holds promise for the future. There are many other examples of tolerance to stresses caused by the physical environment, for which genomic science may provide information leading to novel and innovative approaches to minimizing or eliminating the challenges. Many of the good producing areas have been lost from production due to urbanization or other causes, and production has moved to less desirable locations where soil and water quality may be lower. Such areas may contain soils with excess calcium, high pH values, and mineral imbalances; they can be prone to flooding and/or drought depending on weather fluctuations. Likewise, the quantity and quality of water in many areas can become a limiting factor to economically viable crop production. Salinity in irrigation water is becoming a more serious issue in all the citrus production areas of the U.S., including Florida. Breeding for rootstock improvements is possible and has been actively pursued. However, genomic science offers methods to facilitate the selection of individuals possessing genes and alleles that are desirable because of their abilities to confer tolerance to these environmental deficiencies, thereby dramatically improving the efficiency of the breeding process and decreasing the time to release of genetically improved rootstock cultivars. Furthermore, the availability of cloned genes that could be used to transform already existing cultivars to withstand environmentally related stresses would improve overall productivity.

Fruit quality - There are many aspects of fruit quality, including physical attributes like fruit size, shape, color, texture, number of seeds, and peelability. Equally important are the chemical characteristics of the fruit, such as sugar and acid content, internal color, flavor and aroma compounds, enzyme content as it relates to processing limitations, and the nutritive and health-enhancing components as well. Most of these aspects relate to the overall consumer experience and perception of the citrus products, not to production challenges, and as such they can influence the decisions of consumers who are confronted with an increasingly diverse array of purchase options. Improvements in these aspects can translate directly into increased consumer demand, which ultimately drives the citrus industry to success, or failure if demand for product shifts to other options. In the past five years, the U.S. citrus industries have seen how important the demands of consumers have now become in the domestic and world citrus market. The

Spanish and Moroccan industries have capitalized on their ability to produce and market Clementine mandarins, because of their unique adaptation to the Mediterranean climate. More importantly though, the demand for Clementines has increased dramatically in the world marketplace because of the convenience they offer modern consumers as a small and portable, easy to peel, seedless, and acceptably delicious fruit. Spain currently exports over 60% of the world supply of fresh tangerines[4]. The introduction of the Clementine has upset the world's previous interest in navel oranges, grapefruit, and seedier and less easy to peel mandarins. Currently, the Florida fresh fruit industry is struggling to survive in a marketplace that no longer finds their product desirable, and growers are even abandoning crops on the tree. The effects are also being felt in Texas and California, where the decline in grapefruit and navel orange consumption has begun to depress prices and profit margins as well. Genomic science holds promise in improvements in fruit quality that are difficult, and in some instances, impossible to achieve by any other approach. Research to identify, clone and manipulate genes involved in rind peelability, easy segmentation, and metabolic pathways that lead to the production of sugars and acids, pigments, aromatic or other compounds associated with flavor improvements, and nutrient and phytonutrient elements is underway, and holds promise of increasing the organoleptic and nutritive desirability of citrus fruit and fruit products.

GENERAL OBJECTIVES

The effective use of genomic tools requires genetically appropriate and uniform tissues for analyses so that meaningful results are obtained. An effective and economical strategy for citrus requires both consideration of the constraints and advantages of its unique biology as an apomictic long-lived woody perennial. This includes, for example, selection of the appropriate developmental stages, treatments, and environmental conditions that might affect the traits of interest, ensuring that the tissues and populations are available at the time desired and identifying genomic information developed from other plant species that would augment the citrus effort. Some of these considerations are as follows:

EXPLOITATION OF EXISTING POPULATIONS AND GERMPLASM RESOURCES

Population development – The long juvenile phase and large land areas required for maintenance and evaluation makes the generation of large populations for specific experiments inefficient. Due to the highly heterozygous nature of citrus breeding germplasm, breeding populations are likely to segregate for most of the commercially important traits. Therefore, most of the populations used for analysis will come from segregating progeny generated by citrus breeding programs. Close communication between the genomics effort and the breeding efforts need to be maintained to ensure the generation of large populations. Populations segregating for particular traits can be generated and linkage analyzed, but the difficulty and expense of extensive population

[4] World tangerine exports (USDA, Foreign Agricultural Service, "World Horticultural Trade and U.S. Export Opportunities," August, 2000) may exclude some exports by minor producers.

development as well as the low resolution achievable by such an approach will require linkage disequilibrium analysis of existing germplasm sources.

Germplasm collections – The long lived nature and vegetative propagation of citrus genotypes has permitted the accumulation of a large number of spontaneous mutants for morphological, biochemical and developmental traits. This series of nearly isogenic genotypes represents an untapped resource for the genomics effort. For example, molecular markers capable of distinguishing between the 'Hamlin,' 'Valencia,' and 'Washington Navel' sweet oranges, three orange types that are dramatically different in virtually every horticultural characteristic are not currently available. Large germplasm collections of citrus have been established in several countries representing numerous genotypes. This germplasm resource could be exploited using linkage disequilibrium methods to search for genes controlling specific traits.

Leverage model plant systems – A large body of rapidly expanding genomic information is available from model systems, which can be utilized in citrus. For example, the carotenoid biosynthetic pathway has been extensively studied in tomato and pepper. Carotenoids are an important economic trait in citrus, conferring peel and flesh color and nutraceutical properties to citrus products. It is anticipated that utilizing the extensive information from the genomic efforts on, say, carotenoid biosynthesis and genetics in tomato could facilitate the identification and isolation of candidate genes controlling the production of these compounds in citrus.

Annual citrus model – The long juvenile phase of commercial citrus makes the assessment of phenotypes, particularly of adult phase traits unacceptably long. Studies in transformation, gene function, inheritance and expression across developmental stages in citrus would be best met by a fast cycling citrus genotype. It is therefore essential to identify and/or develop transformable citrus types that will flower and set fruit within 1 year. This can be accomplished by developing transformation methods for precocious citrus types that already exist in U.S. citrus germplasm collections, transforming commercial types to develop an early flowering genotype, or the development of efficient methods of transformation for mature adult phase tissue.

To directly address the various types of problems confronting the U.S. citrus industry, and in the context of the above considerations of citrus biology, the following five basic genomic tools are required.

1. Database of expressed genes.

A citrus EST database houses information on the number, expression profile, and chromosomal location of genes controlling commercial traits. A collection of expressed genes will be obtained by sequencing cDNA libraries generated from diverse plant tissues, organs, stages of development, and interactions with biotic and abiotic stresses. From the total EST sequence database, a representative set of "unigenes" will be derived and their functions defined by comparison to genes of known biochemical function from other organisms (see "Bioinformatics"). Unigene sequences

are of value for the design of arrays, which can then be used by investigators to observe the spatial and temporal expression profiles of all available citrus genes. From expression patterns and known biochemical functions, metabolic pathway maps can be generated and candidate genes identified that control specific economically important traits. Unigene sequences are valuable for the design of gene-specific oligonucleotides for several purposes including oligonucleotide-based expression arrays, real-time RT-PCR, genetic mapping, as probes of genomic libraries, and chip-based genotyping.

2. Database of proteins and metabolites.

A database of the proteins and metabolites produced under different times, tissues, and conditions is essential for validating gene expression studies which, when validated, will provide the knowledge to effectively coordinate gene transcription to produce economically important phenotypes. This objective describes the importance of generating proteomic and metabolomic data, and its subsequent integration with the results from genomic studies.

3. Pedigree independent genetic map.

Citrus germplasm is composed of numerous genotypes originating from several species. Current maps are hard to compare across families, pedigrees and laboratories. Few easy to assay codominant markers are available in citrus. Therefore it is difficult to look at homologous regions from maps generated in different laboratories. We propose to 1) utilize a large mapping population of *Poncirus trifoliata x Citrus sinensis* and *C. sinensis x P. trifoliata* hybrids from the various U.S. citrus research institutions, and 2) to generate a reference genetic map for citrus using sequenced confirmed markers such as SSRs and SNPs. These markers have several alleles per locus and are pedigree independent.

4. Physical map

Development of a physical map will form the basis for functional genomic studies to identify and isolate economically important genes with known metabolic functions, as well as providing the basis for larger scale genome sequencing. The basic approach will utilize the construction of two BAC libraries, one from *P. trifoliata*, the trifoliate orange, and one from *C. sinensis*, the sweet orange. Genetic data will then be integrated into the constructed physical maps.

5. Genome sequence

Initially, a genome sequence obtained from the coding regions of the citrus genome will complete the identification of most of the genes in citrus. It is not feasible to sequence all of the expressed genes by producing ESTs from cDNA libraries because some messages are so rare that they will simply escape detection even in libraries that are relatively enriched for rare messages by normalization. A complete genome sequence will eventually be required, since elements that specify gene expression patterns and

chromosome structure are located in the noncoding regions. Additional advantages of a complete genome sequence are an abundance of potential genetic markers, knowledge of intron positions, and availability of the sequences of regulatory elements that lie adjacent to the transcribed region of any gene.

SPECIFIC OBJECTIVES

GENOME CHARACTERIZATION

Model genotypes - Choosing a single genotype on which to focus most analyses will facilitate analysis of the various types of genome data to be collected in the proposed projects. Some analysis may be conducted on alternative genotypes, to identify SNP markers or to characterize gene expression for traits not present in the model genotype. However, we strongly encourage participants in this project to use the model genotype. Criteria to choose the model genotype include its commercial importance, extent of previous work, genome characteristics, and suitability for various manipulations such as transformation. We propose that an orange variety be chosen for the model genotype because this is the variety for which most EST sequence already exists (see below), it is the commercially most important variety worldwide, and its genome includes segments from both mandarin and pummelo (Nicolosi et al., 2000). Sweet orange can be transformed (Pena et al., 1995; Bond and Roose, 1998) although other genotypes are easier. It is not necessary to select a specific orange variety for all experiments because essentially all oranges derive from a single ancestor by mutation and therefore are nearly isogenic. A seedy sweet orange with high pollen fertility may be a good choice for physical and linkage mapping. The presence of two somewhat divergent genomes in this model genotype may complicate physical mapping and whole genome shotgun sequence assembly. However, these problems are not expected to be severe because such efforts have been successful in other hybrid plants such as sugarcane, wheat and cotton.

A second species that is very important in citrus breeding is the related genus *Poncirus trifoliata* or trifoliate orange. This species is used as a rootstock and many of the most successful citrus rootstocks are *Citrus* x *Poncirus* hybrids. It contributes several important disease resistance traits to citrus rootstocks, and trees grafted on trifoliate orange and trifoliate hybrid rootstocks are often more freeze tolerant and have better internal fruit quality than those on most other citrus rootstocks. Therefore, it will be valuable to study various aspects of the *Poncirus* genome and compare it with sweet orange.

A rapid cycling citrus type for transformation and gene expression studies will also be required. A citrus type that is morphogenetically competent, precocious, and seed propagated would facilitate the application of genetic transformation and other *in vitro* techniques for variety development and functional genomic studies. Efficient regeneration of plants from *in vitro* cultures is essential to the successful application of gene transfer technologies. Citrus transformation by *Agrobacterium* is efficient for some citrus types such as grapefruit, sweet orange and trifoliate orange or its hybrids (e.g.,

Carrizo). However, a long juvenile phase of four to six years complicates a rapid evaluation of gene constructs affecting adult traits such as fruit quality (e.g., rind and flesh color, flavors, maturity dates, seediness, abscission, peelability, rag content, acid and sugar levels), tree architecture, and other adult traits associated with productive mature trees. A rapid cycling citrus model would also permit the evaluation of genetic stability of new constructs over multiple sexual generations.

EST sequencing - EST sequencing should include both 5' and 3' end sequencing. The 5' sequence will usually include a substantial portion of the protein-coding region, which helps to categorize the biochemical nature of the gene product. The 3' end is the location of the 3' untranslated region (UTR), which is usually 200 to several hundred bases in length and contains the greatest variability between individual genes in multigene families (paralogs) and between alleles of individual genes. This variability makes the 3' end sequence most valuable for the differentiation of paralogs in unigene assemblies, and consequently recognition of SNPs and the design of gene- and allele-specific oligonucleotides. EST sequencing of sweet orange should identify many SNPs due to the presence of two somewhat divergent genomes (i.e., pummelo and mandarin).

Genetic markers and mapping - A number of different genetic linkage maps of citrus have been produced in various laboratories (Roose et al., 2000), but it is difficult to interrelate the linkage groups identified on different maps because few common markers have been used. It is essential to develop a reference map of the model genotype that includes a set of markers that are highly polymorphic so that they can be mapped in various populations. This will allow various maps to be compared and combined. Maps of the reference genotypes are also important to facilitate development of a physical map and ultimately a genomic sequence. We propose to combine the available *C. sinensis x P. trifoliata* and the reciprocal hybrid populations in the U.S. today into a large 300+ composite mapping population for generating reference maps of sweet and trifoliate oranges. Citrus researchers should also agree on a common nomenclature for linkage groups. SSR (microsatellite) markers are most suitable for development of a reference map because they are single locus markers that are heterozygous in many populations. Other classes of sequence tagged site (STS) markers with these properties are also suitable such as expressed sequenced tags (ESTs), resistance gene candidates (RGCs), or single nucleotide polymorphisms (SNPs). SSR markers are suitable for germplasm characterization and variety identification, although it has not yet been shown that they can distinguish mutationally derived varieties of citrus, as has been shown in grapes (Riaz et al., 2002). SNPs and/or other markers derived from ESTs should also be included on the map for synteny comparisons with other species. Additionally, it is not uncommon to find SSRs located within EST sequences. There are now over 16,000 examples of cereal plant ESTs that contain SSRs (http://wheat.pw.usda.gov/ITMI/2002/EST-SSR/). An effort to mine citrus EST sequences for SSRs therefore is justified in order to combine the advantages of SSRs with the relevance of expressed genes.

Physical mapping - Physical maps are required tools for efficient gene localization and isolation, studying the organization and evolution of genomes, and as an initial step for efficient whole genome sequencing. The highest resolution physical map of the citrus genome will be the complete DNA sequence of each chromosome, or in other words, one marker at each and every nucleotide of the citrus genome. However, lower resolution physical maps that are easier to construct would allow us to address important short-term goals that will then lead to the complete sequence.

A chromosomal or cytogenetic map is developed based on distinct banding patterns of stained chromosomes visualized under light microscopy and results in about one marker every 10 Mb (Casey, 2002). This type of physical map is of the lowest resolution, yet to date, one does not exist for citrus, in part because citrus chromosomes are very small and difficult to analyze. A cytogenetic map for citrus, at any resolution may prove useful for the citrus genomic community, but of greater utility to the citrus genomics community will be a medium to high-resolution physical maps of sweet orange and trifoliate orange that combines ordered sets of cloned DNA anchored by sequence-tagged connectors (STCs; Chen et al., 2002).

A bacterial artificial chromosome (BAC) library that represents 9X coverage of the *Poncirus* genome and a BAC library that represents 8X coverage of a *Poncirus* x *Citrus* hybrid are available for BAC fingerprinting to develop contigs and for BAC end sequencing for generating STCs data. Relatively low pass efforts should generate large ($\geq$ 20-30 Mb) contigs with a STC marker every 0.1Mb (Chen et al., 2002). Gap closure and higher resolution marker density will be obtained by linkage of the genetic and physical maps, and by placing ESTs on these maps.

Genome sequencing - Obtaining the sequence of all or a large and informative portion of the citrus genome appears to be a feasible objective because the citrus genome is relatively small (382 Mbp). Genome sequencing would facilitate positional cloning approaches to gene identification by allowing researchers to rapidly identify candidate genes in specific regions. While EST sequencing is expected to identify many citrus genes, many genes expressed at low levels, in few cells, or in response to specific environmental conditions may not be identified in EST sequencing projects. In rice, full genome sequencing identified about 55,000 genes, while an assembly of 87,842 ESTs gave only 24,776 unigene clusters (Yu et al., 2002). In comparison, a “stringent” assembly of 350,000 barley ESTs (paralogs separated) gave about 53,000 unigene clusters, while a “relaxed” assembly (paralogs often combined) gave 41,000 unigene clusters (http://harvest.ucr.edu), Together, these and similar numbers from other EST projects indicate that EST sequencing in citrus will continue to reveal new citrus genes in a cost-effective manner until well past the 200,000 EST mark, but at a reduced rate as the unigene number per EST begins to plateau.

Methods for genome sequencing continue to decrease in cost. The scale of the task could also be reduced by developing libraries from lower complexity DNA (Peterson et al., 2002) which would exclude possible large regions in the genome that have few genes, but much repetitive DNA that is difficult to assemble. While no detailed

characterization has been reported, staining *Citrus* metaphase chromosomes with DAPI and CMA shows that several chromosomes contain large blocks of terminal heterochromatin (Miranda et al., 1997). Based on experiments in other plant species, DNA in these regions is unlikely to contain many genes. The portion of the genome included in such regions is difficult to estimate, but may be as much as 50% based on the observed length and staining intensity. The small size of citrus chromosomes makes cytogenetic characterizations difficult. Therefore, analysis of the citrus genome for coding/noncoding regions should be done by BAC fingerprinting.

GENOME CHARACTERIZATION

Short term objectives (1-2 years)

1. Establish sweet orange (C. sinensis) and trifoliate orange (P. trifoliata) as model citrus types.
2. Identify a rapid cycling citrus type for genetic transformation.
3. Isolate DNA from the composite mapping population and distribute to all cooperating laboratories.
4. Develop a framework genetic map in the C. sinensis and P. trifoliata using SSR markers. About 200 SSR plus other EST and BAC derived markers should be used.
5. Determine a reference set of SSR markers for germplasm characterization and variety identification.
6. Develop efficient methods to map ESTs in citrus and implement these in the model mapping population.
7. Select an existing BAC library from Poncirus and/or construct a BAC library from sweet orange, the proposed model citrus genotype.
8. Complete physical map construction of the model genotype based on a combination of BAC fingerprinting and BAC end sequence (STCs) analysis.
9. Integrate the physical map with the genetic maps using SSRs and EST derived markers.
10. Construct high-resolution region specific physical maps of the Poncirus genome from areas already known to contain genes encoding important traits such as nematode resistance and cold tolerance.

Long term objectives (2-5 years)

1. Develop maps that include gene-based markers (ESTs, RGCs, and SNPs) in various populations suitable for analysis of pest and disease resistance, stress tolerance, fruit quality, tree size, and other traits of interest.
2. Analyze EST markers and traits to identify candidate genes responsible for naturally occurring variation in these traits.
3. Relate linkage maps to physical maps.
4. Integrate genetic, physical, EST sequences, and genome sequencing data.
5. Obtain the sequence of low-complexity portion of the citrus genome to at least 5X redundancy.
6. Analyze selected regions of high-complexity sequence to estimate the number of genes likely to be present in such regions.
7. Obtain sequence of selected gene-rich regions from specific resource genotypes including a mandarin, pummelo, citron, and trifoliate orange. These regions may correspond to clusters of disease resistance genes, genes involved in fruit quality traits, or other traits of interest.

FUNCTIONAL ANALYSIS OF THE CITRUS GENOME

Genomic coding is translated into tree productivity and fruit quality through a complicated system of gene transcription, protein expression and metabolism. The

previous sections have primarily described the necessity, utility and process of acquiring information on genomic DNA (genomics). This section describes the importance of generating mRNA (transcriptome), protein (proteomics) and metabolism (metabolomic) data, and its subsequent integration with the results from genomic studies. The utilization of genomic data in concert with transcriptome, proteomic and metabolomic data is the functional analysis of the citrus genome.

Citrus trees and fruit express a unique network of molecules that determine how the different parts of the plant function and how the plant responds to its environment. These changes determine plant resistance, productivity and quality attributes. One of the major objectives of this project is to create a profile of the different transcriptome, proteome and metabolome elements as they appear in different tissues, and as they change in response to development and stress (environmental, pathogenic, or insect induced).

Ultimately, all citrus genes will be cataloged in association with their location on the chromosome, the function of their product, how activity of the gene product fits into the complex metabolic pathways present in citrus, and how it is regulated developmentally and in response to environmental stimuli. Using high throughput screening methods, production problems will be rapidly screened to understand the underlying molecular basis for a specific problem, and the best solutions will be readily implemented. These solutions include changes in management/harvesting/ postharvest practices, or new variety development strategies. In this way, functional analysis will be accessible to those that can best use it to solve industry problems. There are three key elements to this research: 1) Transcript profiling; 2) High throughput identification of proteins; 3) High throughput identification of metabolites.

Transcript profiles - As cells change in response to developmental (flower development, fruit ripening) or environmental (pathogen or insect attack, or nutrient stress) stimuli, the composition and abundance of individual mRNA transcripts (transcriptome) change in response to the need for changes in protein and metabolite production (proteome/metabolome). Profiling how specific transcripts change in response to development or environmental stimuli will provide necessary information on the function of the genes corresponding to the individual transcripts. This information can then be used to describe the presence and function of specific metabolic pathways active in the tissues of interest. Gene array technologies provide a high throughput method for developing transcript profiles, and information from these profiles will allow the construction of a map of the metabolic pathways functioning in citrus. Array technologies are rapidly evolving and there are currently three major methods: 1) attachment of short oligonucleotides (~25 nucleotides) onto a solid substrate using photolithography and solid-phase chemistry techniques (e.g., Affymetrix Genechips®); 2) printing glass slide arrays with oligonucleotides synthesized based on EST/genomic sequence data; or 3) arraying of cDNA PCR products onto nylon membranes or glass slides. The techniques ultimately employed will depend on the status of array methods, the application, and available resources. These gene transcript profiles will be validated via protein and metabolite profiling discussed below.

Proteomics and metabolomics - The viability of the citrus industry is based on two factors, consumer demand and grower/producer profitability. Consumer desirability is based on such factors as flavor, aroma, texture, and easy of peeling. Profitability is dependent upon minimizing production costs through the use of cultivars that are pest and disease resistant, flourish in less than ideal environments, are easily processed, and meet consumer requirements. These desirable traits are the result of the proteins produced by the plant and the compounds (metabolites) subsequently generated by these proteins. The identification of those proteins and metabolites responsible for a given trait may be accomplished through proteomic and metabolomic approaches respectively. Potential benefits to the citrus industry from this approach would include increased profitability and consumer satisfaction through the shortening the time required to introduce new commercial cultivars, the development of strategies to counteract environmental pressures (e.g. pests, disease, climate) and the institution of post-harvest regimes that maximize citrus quality. In addition to the benefits to the citrus industry and citrus consumers, these studies would answer fundamental questions in plant biology and would be of significant interest to the scientific community.

Proteomic and metabolomic studies consist of three steps: separation, identification and profiling. The first step is essential to the success of the following two and consists of taking a sample derived from citrus and separating the complex mixture of proteins and metabolites into individual components. There are a myriad of separation techniques available and many including liquid chromatography (LC), SDS-PAGE, or capillary electrophoresis (CE) and may be used in tandem with the identification step in a high-throughput fashion. Identification is generally accomplished through a combination of mass spectrometry (MS) methods to determine identify of a protein and in the case of metabolites a combination of MS and nuclear magnetic resonance (NMR). Profiling consists of comparing the protein expression levels and metabolite concentrations found in samples derived from different cultivars, tissue types, cultivation or post-harvest conditions in order to understand what proteins or metabolites may be responsible for a trait of scientific or commercial significance (e.g. easy-peeling, disease resistance, etc.).

We propose to undertake the following targets for proteomic and metabolomic studies and that participating scientists agree upon a set of standard methods to facilitate the comparison of results and eliminate the duplication of work.

- *Model genotypes* - Proteomic and metabolomic characterization of the model citrus genotypes. Results from this project will be used to validate results derived from DNA microarrays and EST studies by verifying protein expression and allow the subsequent coordination of gene transcription with protein expression and metabolite production. Results will be used to establish baseline protein expression levels and metabolite concentrations, and to identify constituently expressed proteins and produced metabolites that will be used as standards for comparing results derived from different cultivars or cultivation conditions.

- *Rapid cycling citrus model* - Following proteomic and metabolomic characterization of the fast cycling citrus genotype, newly derived plants will be profiled against the unmodified plant to determine the effect of a newly introduced or modified gene(s) on protein expression and metabolite production. These studies will as aid in the validation of gene function and the coordination of phenotype with genotype.
- *Characterization of specific traits of economic importance* - Projects in this group are driven by the needs of the citrus industry and are conducted by the proteomic and metabolomic profiling of cultivars with and with out a desired trait to identify the protein(s) or metabolite(s) associated with the trait.

FUNCTIONAL ANALYSIS OF THE CITRUS GENOME

Short term objectives (1-2 years)

---

1. Develop a framework for the generation and sharing of data.
2. Establish standards for transcriptome, proteomic and metabolomic methods to facilitate the comparison of data generated from different laboratories and thereby reduce duplication.
3. Streamline methods.
4. Establish baseline protein expression levels and metabolite concentrations for a sample that complements genomic studies, a sample that will be used as a model system and on trait specific projects.
5. Establish a mechanism for the evaluation and incorporation of new methodologies as they become available, and for industry input into research targets.

Long term objectives (2-5 years)

---

1. Discovery of highly expressed proteins in the citrus proteome and coordination of these results with those from genomic studies.
2. Discovery of major metabolites associated fruit quality and plant survivability.
3. Significantly increased understanding of what factors will affect fruit quality and how the plant responds to environmental stresses.
4. Identify key proteins and metabolites that may be used as markers or targets for improving cultivars.
5. Evaluate progress and set new objectives.

BIOINFORMATICS

Bioinformatics is a hybrid discipline of computer science and biology. It includes the organization, storage, integration, and analysis of biological information. The core genomic data includes large data sets such as (1) chromatograms representing DNA sequences, genotype information, and fingerprints of DNA and protein fragments, (2) images derived from expression arrays, northern blots, southern blots, or stained gels, and (3) quantitative data derived from expression arrays, real-time RT PCR, and biochemical assays. In addition core biological data includes various other measurements of the biological material and the environment. Primary derivatives of

the core genomic data include DNA sequence strings and their associated quality values, tables of genotype scores, gene expression tables, and protein sequences. Further derivatives include assemblies of ESTs to form unigene sequences, complete gene sequences including positions of reading frames and splice junctions, gene-specific oligonucleotides, alignment of gene sequences with related genes from other organisms, groupings of genes according to expression pattern and function, genomic clone contig assemblies and physical maps, and genetic linkage maps including molecular markers and trait loci. In essence, the core genomics information spans all the biological information polymers (i.e., DNA, RNA, protein), metabolites encoded and synthesized by these polymers, and how these molecules result in the characteristics of the whole organism. The amount of information is vast and is being accumulated at increasing speed as advances in technology continue to automate and increase the throughput of data collection.

The first bioinformatics objective is to establish the means by which the various forms of the core data can be stored. The vast size of the data requires storage capacity in the range of 10's of terabytes. The diverse nature of the data requires staff that are knowledgeable in the biology of citrus and the nature of the data, as well as computer hardware. There is also a need for security of the core data, which requires restricted access and backup, and there is a need for users to be able to access the core data on-demand. The size and constantly expanding nature of such data in new genomics projects precludes the complete and routine replication of such data on personal computers, and this necessitates web accessibility.

The next bioinformatics objectives relate to the production of derivatives of core data that meet the needs of users. Generally, all derivatives are based on relationships between elements of the core information, and between elements of the core information and external data sources. A relational database is necessary in order to record and readily access these relationships. To keep costs to a minimum in a centralized databasing location while making the database amenable to a broad group of users, a SQL compliant database (e.g., MySQL or PostgreSQL) would be preferred for the relational database.

Several off-the shelf databasing and extraction tools can fulfill many of the initial needs in citrus genomics. EST assembly and the viewing of assemblies, consensus sequences and gene-specific oligonucleotides will be among the first needs in citrus bioinformatics. StackPack (http://www.sanbi.ac.za/CODES/STACKPACK_REQUEST/) and HarvEST: Citrus (http://harvest.ucr.edu) are two packages that can handle most of these needs. A more comprehensive genome annotation database will be required as data on genome sequence, gene expression, promoters, trait locations, splice junctions, and other data accrue. One model for citrus to consider is the GMOD (Generic Model Organism Database) (http://www.gmod.org/) which aims to provide complex annotation architecture that ultimately is framed around the complete genome sequence. GMOD is already being used for the two premiere model plants, Arabidopsis (TAIR) and rice (Gramene). No matter what database method is used, it is desirable that the database can be customized and improved to take into consideration preferences of citrus

investigators. For this reason, it will not be desirable to contract the work to any closed-sourced database service.

For many purposes, primary or subsequent derivatives of the core data (e.g. gene-specific oligonucleotides, genetic maps, and protein sequences) meet the most frequent needs of users. For these derivatives, storage space is not a limiting factor, being generally on the order of less than a few gigabytes, which is well within the range of today's entry level personal computers. In such cases, delivery of information can be achieved using centralized computers via the internet or through the installation of software and local databases on personal computers. The advantage of web-based delivery is that anyone with an internet connection can have access, but this advantage is counterbalanced by the risk of crashes of a single centralized facility and the vastly slower speed of information access and simple search and export operations through the internet compared with today's entry level personal computers.

The unifying purpose of bioinformatics ultimately is to simplify the process of finding pertinent formation so that it can be applied to the derivation of solutions to practical problems. We can expect the required user time to fall dramatically, in parallel with continued price improvements in the consumer computer market and a general increase in the bioinformatics work force.

BIOINFORMATICS

Short term objectives (1-2 years)

Build partnerships with established bioinformatics centers to:

1. Provide appropriate hardware and database methods for the core data generated in the previous objectives.
2. Provide adequate security measures to ensure the integrity of the core data.
3. Create derivative forms of the core data that organize and present it in such a way that users can readily extract meaning from it.
4. Develop an effective mix of web-dependent and web-independent user interfaces to access the core data and its derivatives.

RESOURCE CENTERS

Currently, citrus germplasm is maintained in the U.S. at the USDA-ARS National Clonal Germplasm Repository for Citrus and Dates, located on land provided by the University of California, Riverside. The primary objectives of the facility as regards citrus are to collect, maintain, evaluate, preserve, and distribute pathogen-free clonal germplasm of *Citrus* and related genera. However, an International Citrus Genome Program will require the maintenance of other materials, both plant stocks (mapping populations, mutants, transgenic plants) as well as collections of molecular tools (EST collections, BAC libraries, genomic and cDNA libraries, primer data, etc.). These resources should be freely available to all researchers at nonprofit agencies. The various centers should

be organized around a common portal (e.g., website) where each resource center's contact information, resources, and method of distribution are available.

RESOURCE CENTERS

Short term objectives (1-2 years)

---

1. Establish a common web-based portal to organize the resource centers.
2. Determine and catalogue what materials are available for general use.
3. Initiate establishment of stocks.

Long term objectives (2-5 years)

---

1. Make genomic and genetic resources freely available to the research community.
2. Attempt to minimize the space needed, the maintenance costs and labor, and the distribution costs in making these stocks available.

INTERNATIONAL EFFORTS

Most research groups working on citrus biotechnology in the several countries where citrus is an important commodity have not been organized around a citrus genomic research initiative model, but rather, in a more "subject" driven research. The following is a brief summary of current genomic efforts in citrus in other nations. The information identifies the potential partners in the organization of a larger international citrus genomics community, where resources and efforts could be shared in achieving a comprehensive citrus genomics research effort.

1. *Brazil* – the world's largest citrus producer. One of the first countries to start a genomic initiative on citrus; however, their focus has currently been on citrus pathogen genomics. They successfully sequenced the whole genome of the bacterium *Xylella fastidiosa* and are now starting functional genomic studies in the same organism. Additionally, they have sequenced *Xanthomonas axonopodis* pv. citri (citrus canker) and *X. campestris* pv. campestris (black rot in crucifers) and have conducted a genome comparison between these two *Xanthomonas* pathogens with differing host specificities (da Silva AC et al., 2002). The Brazilian citrus research community is currently well organized as a consortium, to face the challenge of citrus genomics.

2. *Spain* – an important citrus producer of fresh fruit with a high interest in *C. clementina*. Has contributed important information in the genetic transformation of citrus using *Agrobacterium.* A citrus genomic's effort has been initiated and includes the generation of EST libraries, cDNA microarrays, development of molecular markers for QTL identification, and construction of a BAC library from *C. clementina*.

3. *Japan* – primary interest is Satsuma mandarin. Several thousand ESTs from Satsuma mandarin produced. Additionally, several other genes coding for proteins

with similarity to sucrose phosphate synthase, endoxyloglucan transferase, expansin, extensin, pectinesterase, and others were isolated from the same species.

4. *France* – beginning development of STMS markers for gene mapping, assisted selection, and for the identification of somatic hybrids, in addition to CAPs for cytoplasmic genome analysis.

5. *United States* – most citrus molecular biology has emphasized cloning the gene (*Ctv*) for resistance to *Citrus tristeza virus* (CTV) using a map base cloning strategy. High resolution linkage mapping has identified several markers tightly linked to *Ctv*. BAC contigs encompassing *Ctv* have been constructed and the region that appears to be responsible for the resistance has been sequenced. Other resistance genes have been isolated using a PCR approach with degenerative primers designed from nucleotide binding site motifs from Arabidopsis, tobacco, and flax. Several linkage maps have been developed in citrus to improve genetic analysis of several traits, however, most of these are low density, and therefore, new maps are being developed using AFLP and ISSR. In addition to CTV resistance, other traits such as low juice acidity (*acitric* gene), cold acclimatation response QTLs, citrus nematode and *Phytophthora* resistance, etc, have also been targeted in mapping. Genes involved in important pathways such as flavonoid, carotenoid, and terpenoid have also been isolated. Most of the U.S. groups working on citrus molecular biology have also well developed systems for genetic transformation. Large-scale EST projects have been initiated over the past 2 years.

WORKSHOPS AND SYMPOSIA

Research Workshops and Symposia will play a critical role in the Citrus Genomics Consortium. The International Society of Citriculture meets every three to four years, and this meeting has been a good venue for research progress on citrus breeding, genetics, physiology, pathology, biotechnology and genomics. The yearly Plant and Animal Genome Meeting has also been a good venue for the above topics in the context of the Fruit and Nut Workshop. It is recognized that future, at least yearly, meetings should have a specific Workshop for discussions about the overall progress of Citrus Genomics and on the individual projects. It may also be required to have specific Workshops or Symposia to specifically present progress to Industry Representatives and Growers. These meetings can be held at the annual Plant and Animal Genome Meeting in San Diego, or they can be organized at another venue by the Steering Committee.

WORKSHOPS AND SYMPOSIA

Short term objectives (1-2 years)

---

1. Establishment of a calendar of events for the International Citrus Genomics Consortium.

Long term objectives (2-5 years)
1. Regular annual Citrus Genomics Consortium meetings.

CONCLUSION

A citrus genomics initiative is presented that includes both general and supporting specific objectives. The sweet orange is recommended as the primary model citrus type because of its commercial importance. However, to facilitate application to citrus rootstock improvement, genetic mapping of trifoliate orange is also recommended. These objectives were chosen to provide the genomic tools required for applied research to address the various constraints to the economic production of fresh and processed citrus. A primary solution to many of these problems is new rootstock and scion varieties. The reason is that varieties that have the appropriate characteristics are the least expensive way to adapt citrus production to management schemes requiring less labor, fewer chemical inputs, and adapted to marginal sites. New citrus varieties would also improve the competitiveness of U.S. citrus producers in local and international markets. However, developing new citrus varieties is a slow and costly endeavor and genomic science can assist in increasing both the speed and efficiency of new variety development. Because variety development in tree crops, and particularly citrus is slow, it is essential that U.S. institutions involved in citrus research coordinate their expertise and activities. This will ensure the effective utilization of these tools in variety development and release to the industry. Attainment of these objectives will result in the development of new varieties more rapidly and efficiently than current conventional methods. Also, it is understood that these tools will not be limited to variety development applications, but will contribute to other applied ancillary applications such as powerful diagnostics for monitoring overall tree health.

The U.S. citrus industry is represented by growers, packers, processors, and specialty marketers. Each industry component has constraints on economical production and marketing. Many kinds of improvements and many more newly conceived possibilities can be achieved by an integrated and nationwide genomics effort in citrus. These improvements can provide solutions to practical problems faced by the citrus industry, such as the control of pre- and postharvest diseases and pests, the management of soil fertility and water resources to maximize production, the ability to maintain or improve productivity in marginal growing conditions, and improved profitability by producing new value-added products (e.g., nutraceticals, pectin by-products, oils, and flavors). Such improvements likewise can protect environmental quality, by minimizing chemical inputs and protecting precious water resources. Genomic research may also lead to improvements in fruit or juice quality, with the result of increased consumption, enhanced eating experience, and improvements in the health value of this already significant source of human nutrition. The broad potential for improvement that can be realized through genomic research may hold the key to the sustainability and future viability of the citrus industries of the U.S. These improvements may enable citrus to be the most significant fruit industry in the U.S., as it currently is today.

REFERENCES

Bond J and Roose ML. 2000. Agrobacterium-mediated transformation of the commercially important citrus cultivar Washington navel orange. Plant Cell Rep. 18:229-234.

Casey, D. 2002. A primer on molecular genetics. http://www.ornl.gov/hgmis/publicat/primer/toc.html

Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, et al. 2002. An Integrated Physical and Genetic Map of the Rice Genome. Plant Cell 14: 537-545, 2002

da Silva AC, Ferro JA, Reinach FC, Farah CS, Furlan LR, Quaggio RB, Monteiro-Vitorello CB, Van Sluys MA, Almeida NF, Alves LM, do Amaral AM, Bertolini MC, Camargo LE, Camarotte G, Cannavan F, Cardozo J, Chambergo F, Ciapina LP, Cicarelli RM, Coutinho LL, Cursino-Santos JR, El-Dorry H, Faria JB, Ferreira AJ, Ferreira RC, Ferro MI, Formighieri EF, Franco MC, Greggio CC, Gruber A, Katsuyama AM, Kishi LT, Leite RP, Lemos EG, Lemos MV, Locali EC, Machado MA, Madeira AM, Martinez-Rossi NM, Martins EC, Meidanis J, Menck CF, Miyaki CY, Moon DH, Moreira LM, Novo MT, Okura VK, Oliveira MC, Oliveira VR, Pereira HA, Rossi A, Sena JA, Silva C, de Souza RF, Spinola LA, Takita MA, Tamura RE, Teixeira EC, Tezza RI, Trindade dos Santos M, Truffi D, Tsai SM, White FF, Setubal JC, Kitajima JP. 2002. Comparison of the genomes of two Xanthomonas pathogens with differing host specificities. Nature 23:459-463.

Frost HB and Soost RK. 1968. Seed reproduction: development of gametes and embryos. In: *The Citrus Industry*, Vol. II, ed. W Reuther, LD Batchelor and HJ Webber, Berkeley: University of California Press.

Miranda M, Ikeka F, Endo T, Morigushi T and Omura M. 1997. Comparative analysis on the distribution of heterochromatin in *Citrus*, *Poncirus*, and *Fortunella* chromosomes. Chr. Res. 5:86-92.

Nicolosi E, Deng ZN, Gentile A, La Malfa S, Continella G and Tribulato E. 2000. Citrus phylogeny and genetic origin of important species as investigated by molecular markers. Theor. Appl. Genet. 100:1155-1166.

Pena L, Cervera M, Juarez J, Navarro A, Pina JA, Duran-Vila N, and Navarro L. 1995. Agrobacterium-mediated transformation of sweet orange and regeneration of transgenic plants. Plant Cell Rep. 14:616-619.

Peterson DG, Schulze SR, Sciara EB, Lee SA, Bowers JE, Nagel A, Jiang N, Tibbitts DC, Wessler SR, Paterson AH. 2002. Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery. Genome Res. 12:795-807

Riaz S, Garrison KE, Dangl GS, Boursiquot JM, Meredith CP. 2002. Genetic divergence and chimerism within ancient asexually propagated winegrape cultivars. J. Am. Soc. Hort. Sci. 127:508-515.

Roose ML, Fang D, Cheng FS, Tayyar RI, Federici CT, and Kupper RS. 2000. Mapping the *Citrus* genome. Acta Horticulturae 535:25-32.

Webber HJ and Batchelor LD. 1943. "The Citrus Industry." Vol. I, ed. W Reuther, HJ Webber and LD Batchelor, Berkeley: University of California Press.

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, et al.

2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296: 79-92.

APPENDICES

Appendix A – Summary of Short-term Objectives (1-2 years).....23
Appendix B – Summary of Long-term Objectives (2-5 years) 24
Appendix C – Interim Steering Committee Members 25
Appendix D – Interim Advisory Committee Members 28

Appendix A – Summary of Short-term Objectives (1-2 years)

GENOME CHARACTERIZATION:

1. Establish sweet orange (*C. sinensis*) and trifoliate orange (*P. trifoliata*) as model citrus types.
2. Identify a rapid cycling citrus type for genetic transformation.
3. Isolate DNA from the composite mapping population and distribute to all cooperating laboratories.
4. Develop a framework genetic map in the *C. sinensis* and *P. trifoliata* using SSR markers. About 200 SSR plus other EST and BAC derived markers should be used.
5. Determine a reference set of SSR markers for germplasm characterization and variety identification.
6. Develop efficient methods to map ESTs in citrus and implement these in the model mapping population.
7. Select an existing BAC library from *Poncirus* and/or construct of a BAC library from sweet orange, the proposed model citrus genotype.
8. Complete physical map construction of the model genotype based on a combination of BAC fingerprinting and BAC end sequence (STCs) analysis.
9. Integrate the physical map with the genetic maps using SSRs and EST derived markers.
10. Construct high-resolution region specific physical maps of the Poncirus genome from areas already known to contain genes encoding important traits such as nematode resistance and cold tolerance.

FUNCTIONAL ANALYSIS OF THE CITRUS GENOME:

1. Develop a framework for the generation and sharing of data.
2. Establish standards for transcriptome, proteomic and metabolomic methods to facilitate the comparison of data generated from different laboratories and thereby reduce duplication.
3. Streamline methods.
4. Establish baseline protein expression levels and metabolite concentrations for a sample that complements genomic studies, a sample that will be used as a model system and on trait specific projects.
5. Establish a mechanism for the evaluation and incorporation of new methodologies as they become available, and for industry input into research targets.

BIOINFORMATICS:

Build partnerships with established bioinformatics centers to:

1. Provide appropriate hardware and database methods for the core data generated in the previous objectives.
2. Provide adequate security measures to ensure the integrity of the core data.
3. Create derivative forms of the core data that organize and present it in such a way that users can readily extract meaning from it.
4. Develop an effective mix of web-dependent and web-independent user interfaces to access the core data and its derivatives.

RESOURCE CENTERS:

1. Establish a common web-based portal to organize the resource centers.
2. Determine and catalogue what materials are available for general use.
3. Initiate establishment of stocks.

INTERNATIONAL EFFORTS:

1. Promote the organization of an international citrus genomics community.

WORKSHOPS AND SYMPOSIA:

1. Establishment of a calendar of events for the International Citrus Genomics Consortium.

Appendix B – Summary of Long-term Objectives (2-5 years)

GENOME CHARACTERIZATION:

1. Develop maps that include gene based markers (ESTs or SNPs) in various populations suitable for analysis of pest and disease resistance, stress tolerance, fruit quality, tree size, and other traits of interest.
2. Analyze EST markers and traits to identify candidate genes responsible for naturally occurring variation in these traits.
3. Relate linkage maps to physical maps.
4. Integrate genetic, physical, EST sequences, and genome sequencing data
5. Obtain the sequence of low-complexity portion of the citrus genome to at least 5X redundancy.
6. Analyze selected regions of high-complexity sequence to estimate the number of genes likely to be present in such regions.
7. Obtain sequence of selected gene-rich regions from specific resource genotypes including a mandarin, pummelo, citron, and trifoliate orange. These regions may correspond to clusters of disease resistance genes, genes involved in fruit quality traits, or other traits of interest.

FUNCTIONAL ANALYSIS OF THE CITRUS GENOME:

1. Discovery of highly expressed proteins in the citrus proteome and coordination of these results with those from genomic studies.
2. Discovery of major metabolites associated fruit quality and plant survivability.
3. Significantly increased understanding of what factors will affect fruit quality and how the plant responds to environmental stresses.
4. Identify key proteins and metabolites that may be used as markers or targets for improving cultivars.
5. Evaluate progress and set new objectives.

BIOINFORMATICS:

RESOURCE CENTERS:

1. Make genomic and genetic resources freely available to the research community.
2. Attempt to minimize the space needed, the maintenance costs and labor, and the distribution costs in making these stocks available.

INTERNATIONAL EFFORTS:

WORKSHOPS AND SYMPOSIA:

1. Regular annual Citrus Genomics Consortium meetings.

Appendix C – Interim Steering Committee Members

Andrew P. Breksa III
Research Chemist
USDA, ARS
Western Regional Research Center
800 Buchanan Street
Albany, CA 94710
Tel: (510) 559-5898
Fax: (510) 559-5849
E-Mail: apb3@pw.usda.gov

José Chaparro
Research Geneticist
U. S. Horticultural Research Laboratory
2001 South Rock Road
Ft. Pierce, Florida 34945
Tel: (772) 462-5800
Fax: (772) 462-5986
E-mail: JChaparro@ushrl.ars.usda.gov

Timothy Close
Professor of Genetics and Geneticist
Department of Botany
4130 Batchelor Hall
University of California
Riverside, CA 92521
Tel: (909) 787-3318
Fax: (909) 787-4437
E-mail: timothy.close@ucr.edu

Abhaya Dandekar
Professor
Department of Pomology
1059 Wickson Hall
University of California
Davis, CA 95616-8683
Tel: (530) 752-7784
Fax: (530) 752-8502
E-mail: amdandekar@ucdavis.edu

Fred G. Gmitter, Jr.
Professor of Citrus Genetics and Breeding
University of Florida
Citrus REC - Lake Alfred
700 Experiment Station Road

Lake Alfred, FL 33850-2299
Tel: (863) 956-1151
Fax: (863) 956-3579
E-mail: fgg@lal.ufl.edu

Eliezer Louzada
Associate Professor
Texas A&M
TAMU-Kingsville Citrus Center
312 N International Blvd.
Weslaco, TX 78596
Tel: (956) 968-2132
Fax: (956) 969-0649
E-mail: e-louzada@tamu.edu

Erik Mirkov
Associate Professor
Texas A&M
Texas A&M University Agricultural Research and Extension Center
Department of Plant Pathology
2415 East Highway 83
Weslaco, Texas 78596
Tel: (956) 968-5585
Fax: (956) 968-0641
E-mail: e-mirkov@tamu.edu

Gloria Moore
Professor Citrus Genetics
University of Florida
Horticultural Sciences Department
1117 Fifield Hall
PO Box 110690
Gainesville, FL 32611
Tel: (352) 392-1928, x221
E-mail: gam@mail.ifas.ufl.edu

Randall P. Niedz (Chair)
Research Leader and Geneticist
U. S. Horticultural Research Laboratory
2001 South Rock Road
Ft. Pierce, Florida 34945
Tel: (772) 462-5800
Fax: (772)-462-5986
E-mail: RNiedz@ushrl.ars.usda.gov

Mikeal L. Roose
Professor of Genetics
University of California – Riverside
Department of Botany
4125 Batchelor Hall
Riverside, CA 92521
Tel: (909) 787-4137
Fax: (909) 787-4437
E-mail: roose@citrus.ucr.edu

Robert Shatters
Research Molecular Biologist
U. S. Horticultural Research Laboratory
2001 South Rock Road
Ft. Pierce, Florida 34945
Tel: (772) 462-5800
Fax: (772)-462-5986
E-mail: Shatters@ushrl.ars.usda.gov

Appendix D – Interim Advisory Committee Members

Doug Cook
Professor,
206 Robbins Hall
Dept. of Plant Pathology
University of California
Davis, CA 95616
Tel: (530) 754-6561
Fax: (530) 752-5674
Email: drcook@ucdavis.edu

James J. Giovannoni
Research Molecular Biologist
USDA, ARS
U.S. Plant, Soil and Nutrition Research
429 Tower Road, Ithaca, NY 14853-2901
Tel: (607) 255-1414
Fax: (607) 254-1459
Email: jig33@cornell.edu

Barbara Baker
Research Plant Molecular Geneticist
USDA, ARS, Plant Gene Expression Center
800 Buchanan Street
Albany, CA 94710-3102
Tel: (510) 559-5912
Fax: (510) 559-5929
Email: bbaker@socrates.berkeley.edu

Alan Collmer
Professor
Department of Plant Pathology
Cornell University
Ithaca, NY 14083-4203
Tel: (607) 255-7843
Fax: (607) 255-4471
Email: arc2@cornell.edu

Ted Batkin
President
California Citrus Research Board
P.O. Box 230
Visalia CA 93279
Tel: (559) 738-0246
Fax: (559) 738-0607
E-mail: Ted@citrusresearch.org

Andrew W. LaVigne
Executive Vice President/CEO
Florida Citrus Mutual
302 South Massachusetts Avenue
Lakeland, FL 33801-5091
P.O. Box 89, Lakeland, FL 33802-0089
Tel: (863) 682-1111
Fax: (863) 682-1074
E-Mail: AndyL@flcitrusmutual.com

Ray Prewett
Executive Vice President
Texas Citrus Mutual
901 Business Park Drive Suite 400
Mission, TX 78572
Tel: (956) 584-1772
Fax: (956) 584-3307
Email: RPrewett@cngmail.com

Glenn Wright
Citrus Specialist
Yuma Mesa Agricultural Center
2186 W. County 15th Street
Somerton, AZ 85350
Tel: (928) 726-0458
Fax: (928)-726-1363
Email: gwright@ag.arizona.edu