DRAFT OF THE INTERNATIONAL CITRUS GENOME CONSORTIUM

WHITE PAPER

November 2003

**DRAFT OF THE INTERNATIONAL CITRUS GENOME CONSORTIUM
WHITE PAPER**

TABLE OF CONTENTS

Note about the cover.

INTERNATIONAL CITRUS GENOME CONSORTIUM STEERING COMMITTEE

SUMMARY

Citrus is the main fruit tree crop in the world. In addition to its agronomical and economical importance, the citrus industry has a great social and cultural interest as part of our heritage.

Traditional breeding as a genetic improvement strategy for citrus cultivar development faces many serious impediments. There are biological limitations associated with degrees of sexual sterility and incompatibility, extended juvenility, large individual plant size and consequent expenses of land, labor, and materials required to maintain and to grow progeny to maturity. The consequence to this limitations is reduced probability of acquiring and identifying superior individual offspring. In addition, it should be noticed that the most economically important citrus groups (sweet oranges, grapefruit, most lemons, Satsumas and Clementines) likely originated from unique individual seedlings. All subsequent diversification within these groups has been through the process of identification of spontaneously occurring somatic mutations. Two consequences follow from this facts. First, there is minimal genetic and allelic diversity within these cultivar groups which restricts opportunity for genetic advance. Second, it is extremely difficult if not impossible to obtain improved cultivars that will produce fruit which consumers recognize as typical citrus. To retain these essential characteristics and to achieve significant improvements in critical traits, demands the application of genetic improvement techniques other than traditional breeding through sexual hybridization an selection. Although some of these obstacles can be reduced by currently available methods such as marker-assisted selection, somatic hybridization and mutation breeding, such methods are currently useful for only a few traits and populations because the citrus genome is poorly characterized.

Genomic technology can provide citrus researchers with a new set of tools to address the problems and limitations described above. Genomics includes methods to rapidly identify and manipulate genes that influence desirable and undesirable traits. When the biological mechanisms responsible for these characters are understood, citrus breeders can develop improved varieties more rapidly and predictably than is

currently possible. Genomics also should provide tools that allow growers to optimize production practices and use more environmentally friendly pest and disease control strategies. Development of new varieties using genomic tools also will benefit consumers by providing fruit with enhanced flavor, quality, nutritional and biofunctional (health) value. Recently, the understanding of many aspects of plant biology has increased very rapidly because of collaborative research to sequence the genomes of a model dicot plant, *Arabidopsis thaliana* and a monocot crop plant, rice. The information from these projects is often difficult to apply to citrus for two reasons. First, these species are distantly related to citrus, and second, citrus is a perennial tree crop whose economic product is a fruit developmentally very different from those of Arabidopsis and rice. Citrus genome analysis will build on the foundation established by other plant genome projects, providing unique information applicable not only to citrus but also to other woody perennials. Overall, a citrus genome project will develop tools that make variety improvement, production, and post harvest management more efficient and predictable. These changes will benefit growers, consumers, and society.

The magnitude of the effort required for an integrated Citrus genomics project is too large for a single citrus research institute to fund and undertake. Therefore, we, researchers from both public and private institutions from Australia, Brazil, China, France, Israel, Italy, Japan, Spain, the USA and other citrus producing nations, and in order to ensure rapid scientific advantages without unnecessary duplication of effort, have agreed to organize an International Citrus Genomics Consortium which objective is to develop a set of genomic tools for use in citrus research worldwide. The organization will foster research cooperation, organization, and sharing of resources, enabling achievement of objectives that exceed the capacity of individual institutes. Information generated by the Consortium will be shared among its members for use in their research.

This document has been developed with the involvement of a multinational community of scientists. The Consortium is open to new members willing to be active participants in achieving our objectives.

I. INTRODUCTION

Citrus is the most economically important fruit crop in the world, with a total production of 105 million metric tons. The most important species are sweet oranges, mandarins, lemons and grapefruits. They are grown in developed and developing countries. However, the Citrus industry is faced with at least two constraints to economical production. First, production areas are being severely challenged by pest, disease, and environmental problems to which current commercial cultivars are susceptible. Second, economical production is severely threatened by a declining farm labour force, encroachment of urban areas on the most productive land, and elevated environmental concerns and constraints that increase production costs. In addition, fruit quality needs to be improved to meet the consumers demands.

Successful Citrus industries are based on cultivars that are well adapted to the environmental conditions, so that productivity is high and the fruit exhibits those characteristics desired most by consumers. Regions where adaptation of specific cultivars is greatest are the ones that 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 costs for land and labour, social and/or economic costs associated with environmental protection, the legal rights of the labour force, and the costs of material inputs such as fertilizers, pesticides, and water availability.

As the global competition for Citrus markets increases, the survival of the Citrus industry is critically dependent on the availability of genetically superior cultivars. Given the unique reproductive biology and the impediments for targeted Citrus cultivar improvements, many of these requirements may be difficult or even impossible to achieve without the application of genomic and other biotechnology procedures.

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. Variety development is inherently costly in Citrus because it requires both large field acreage and many years to adequately evaluate field performance (12-25 years). In addition to the long-term nature of tree breeding, rootstock and scion variety development in Citrus is difficult mainly because of the

following characteristics of Citrus biology:

- Citrus reproductive biology is unusual in nearly all commercial Citrus types: Sweet oranges, grapefruit, lemons, and some mandarins reproduce by facultative apomixis, (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.
- Self-and cross-incompatibility, as well as partial to complete pollen sterility, is common in Citrus.
- Juvenility of most Citrus germplasm lasts 5-10 years prior to the first flower.
- 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.

Genomic technology will be an integral tool to help to remove these and other constraints. Up to now, besides the studies carried out on model species, there is an increasing interest in initiatives on genomics applied to plants of direct economical interest [rice, maize, wheat, pine, Populus and grape, among others (Richmond & Somerville, 2001)]. In this regard, it is worthy to note that being Citrus a crop of great agronomical and economical interest, it has not been the subject of large scale genomics studies yet.

The development and availability of genomic tools is essential for increasing the efficiency of conventional Citrus variety development methods. Applications will include:

- more rapid and efficient production of new rootstock and scion cultivars that meet changing consumer preferences; these new cultivars would provide the possibility to adapt Citrus production to management schemes requiring less labour, fewer chemical inputs, and marginal sites;
- to alter one or a few undesirable characteristics in otherwise acceptable varieties;
- accurate and rapid assessment of plant health using global expression of genes, proteins, and/or important metabolites, and closely associated methods to;
- rapidly identify pests and diseases and
- rapid identification of genotypes.

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.

Among the problems concerning Citrus industry those of biological nature are highly important and susceptible of being solved. Following are brief descriptions of some of the kinds of improvements that are desired by, or would be valuable to, the Citrus industry:

1. *Vegetative and Reproductive Development.* Studies on vegetative and reproductive development are very important due to the decisive influence of the developmental factors on the size and quality of the fruit production. In this regard, fruit set has to be specifically mentioned as a trait to be improved in Citrus breeding. In general, the beginning of fruit development depends on the fertilisation of the ovules. However, fruit set can also be independent of fertilisation. It can be induced by exogenous factors (high levels of hormones) or by exogenous application of certain synthetic hormonal factors, in which cases parthenocarpic fruits (without seeds) are generated. Some Citrus varieties are autoincompatible, with spontaneous parthenocarpy, thus they do not produce seeds unless they cross- hybridize with other cultivars. Then, the control of both the capacity for fruit set and for producing seedless fruits are objectives to be approached with genomic tools.

2. *Pest and disease resistance.* With regard to the biotic stresses, it has to be noted the great losses produced by pathogens as bacteria, fungi, viruses and viroids. Citrus are sensitive to infections caused by several fungal *Phytophthora* species (mainly *P. citrophthora*) that produce the diseases known as "gummosis", "foot or collar rot" and "brown rot" of fruits. Concerning virus, it has to be said that the most important viral disease of Citrus in the world is tristeza, being *Citrus* tristeza virus (CTV) the causal agent. This virus affects specially to sour orange, what has provoked its abandon as a rootstock in most citrus growing areas of the world. This rootstock is very desirable for the Citrus industry, due to its rusticity and tolerance to Phytophthora and abiotic stresses. Thus, it would be interesting to introduce tolerance to Citrus tristeza virus into this rootstock. In addition, there are in many countries more aggressive strains of CTV that produce damages in the scion regardless the rootstock onto which is budded. All this strengthens the necessity of a better understanding of the Citrus defense mechanisms against tristeza and other virus to engineer plants with the capability to cope with the eventual entrance of any of these more harmful strains. Finally, exocortis is a Citrus disease produced by the infection of a viroid (*Citrus Exocortis Viroid, CEVd*) which produces bark scaling in the rootstock and reduces up to 30% the size of the tree canopy, with the corresponding crop reduction.

Citrus are affected by very severe diseases worldwide that have the potential to destroy the industry. Some examples are the Citrus tristeza virus (CTV) that cause the death of commercial varieties grafted on the sour orange rootstock. Only in Spain about 40 million trees has been killed by this virus. The consequence is that this high quality rootstock cannot be used in most citrus areas. Severe strains of CTV are present in several areas of the world producing direct damage on the scion varieties even if grafted on CTV tolerant rootstocks. Huanglonbing (ex greening), produced by the bacterium liberobacter sp., is the main limiting factor for the citrus industry in South East Asia. Citrus Variegated Chlorosis, caused by *xylella fastidiosa*, is producing very severe damage in several areas of South America. Another bacterial disease, Citrus canker, caused by *Xantomonas axonopodis patovar citri* (used to be known as *Xantomonas campestris pat citri*), produces important damage in South America, South East Asia, Africa and some spots in the USA and produce important limitations to free exchange of fruits due to quarantine regulations. Since these diseases may move into new areas, the production of resistant genotypes is of high priority for all

citrus growing countries.

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 is increasing. 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 may first be seen in this area using both host and non-host resistance strategies. Already, several research groups 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 Citrus industry, 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. Thus, it seems convenient to take advantage of the potential of new technologies to study host – pathogen interactions, with the aim of eventually improving the resistance of commercial cultivars against pathogens.

3. *Tolerance to environmental conditions.* There are many examples of tolerance to stresses caused by the physical environment, for which genomic science may provide information leading to novel and innovative approaches to minimize or eliminate these damaging stresses. 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. Both drought and salt excess, as well as cold stress and nutrient deficiency, affect to a great extent the size and quality of the production.

Citrus are mainly grown in semiarid zones, where rain is scarce, thus being

exposed to limited water supply. Citrus have also other characteristics that make irrigation necessary, as for example, a strong competence between vegetative and reproductive growth, a leaf surface with a high density of stomas and a scarcely profound and not much developed root system. Water deficit generates a diminished growth of leaves as well as alterations in the fruit quality. When the water stress occurs during the reproductive process, development of the inflorescence, the fertility and the fruit set and abscission can all be affected. Furthermore, the increasing demand of water for human and industrial consume is leading to an over-exploitation of the subterranean aquifers and to the use of low quality water for irrigation, for example with a salt concentration higher than desirable. Furthermore, depending on the soil and the method and frequency of irrigation, salt concentration in soil can be enhanced several times in the inter-irrigation periods. All this produces an important decrease in productivity that could be damped through knowledge of the basic mechanisms underlying the plant responses to these damaging stresses.

Iron deficiency, also named ferric chlorosis, is one of the main causes of abiotic and nutritional stress in fruit trees grown in calcareous soils. In these soils, the high pH causes the precipitation of the Fe in immobilized forms. As a consequence, despite the high amount of Fe in most soils, it is not available for plant absorption. Furthermore, this can be worsen by the utilization of irrigation water with a high bicarbonate content. All this means that every factor affecting the soil Fe availability, absorption, transport or metabolism can induce ferric chlorosis, and are potential targets for crop improvement in order to cope with this problem.

Cold has always limited the regions in which Citrus could be reliably produced, and is likely to cause serious crop losses again. 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.

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 challenges, 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 improve already existing cultivars to withstand environmentally related

stresses would improve overall productivity.

4. *Fruit quality:* There are many aspects of fruit quality, including physical attributes like fruit size, shape, colour, texture, number of seeds, peelability and period of durability. Equally important are the chemical characteristics of the fruit, such as sugar and acid content, internal colour, flavour and aroma compounds (organoleptic properties), enzyme content as it relates to processing limitations, and the nutritive and health-enhancing components as well. These traits are acquired along fruit development and maturation in the tree, and are related with physiological processes that regulate the fruit development, external maturation and abscission. 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. Genomic science holds promise of 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 the biochemical processes determining (key to determine) characteristics of the fruit such as rind peelability, easy segmentation, and the content of sugars and acids, pigments, aromatic or other compounds associated with flavour improvements, and nutrient elements is underway, and holds promise of increasing the organoleptic and nutritive desirability of Citrus fruit and fruit products.

On the other hand, post-harvest processes are determinant to ensure the quality of the fruit and its acceptance by the consumer. Storage at low temperature is necessary to retard the loss of quality by the fruits, once collected, as well as to reduce the amount of fungicides that are presently used to prevent fungal diseases (the attack by *Penicillium digitatum*), one of the most important causes of loses during post-harvest of Citrus fruits. However, many cultivars do not tolerate these temperatures and develop cold damage. These cultivars have to be stored at relatively high temperatures (10-14ºC) thus, having a very short commercial life. The mechanisms involved in cold tolerance or in response to fungal attack in Citrus fruits are practically unknown. The characterisation of the genes involved in both processes will allow

understanding of the underlying mechanisms and, consequently, will help to reduce the loss of quality of fruits, upon collection.

Genomic technology can provide all citrus researchers with a new set of tools to address the problems and limitations described above. Genomics includes methods to rapidly identify and manipulate genes that influence desirable and undesirable traits. When the biological mechanisms responsible for these characters are understood, citrus breeders can develop improved varieties more rapidly and predictably than is currently possible. Genomics also should provide tools that allow growers to optimize production practices and use more environmentally friendly pest and disease control strategies. Development of new varieties using genomic tools also will benefit consumers by providing fruit with enhanced flavor, quality, nutritional and biofunctional (health) value. Recently, the understanding of many aspects of plant biology has increased very rapidly because of collaborative research to sequence the genomes of a model dicot plant, *Arabidopsis thaliana* and a monocot crop plant, rice. The information from these projects is often difficult to apply to citrus for two reasons. First, these species are distantly related to citrus, and second, citrus is a perennial tree crop whose economic product is a fruit developmentally very different from those of Arabidopsis and rice. Citrus genome analysis will build on the foundation established by other plant genome projects, providing unique information applicable not only to citrus but also to other woody perennials. Overall, a citrus genome project will develop tools that make variety improvement, production, and post harvest management more efficient and predictable. These changes will benefit growers, consumers, and society.

The present Consortium is aimed to be the first large scale genomic study of these and other biological questions related to Citrus industry. Its general goal is to develop genomic tools for improving Citrus cultivars, paying specific attention to aspects relevant to the production process and fruit quality.

The magnitude of the effort required for an integrated genomics project is too large for a single citrus research institute to fund and undertake. Therefore, we have agreed to organize an International Citrus Genomics Consortium which objective is the development of a set of genomic tools for use in citrus research worldwide. The organization will foster research cooperation, organization, and sharing to leverage

resources, enabling achievement of objectives that exceed the capacity of individual institutes. Information generated by the Consortium will be shared among its members for use in their research. The Consortium has selected sweet orange (*Citrus sinensis* var. Pineapple) as the model for study because of its scientific and commercial importance. The use of this variety is not subjected to any legal restriction and is not patented.

The advantages of *Citrus* genome research are recognized by both public and private researchers from Australia, Brazil, China, France, Israel, Italy, Japan, Spain, the USA and other citrus producing nations. Given the magnitude of the effort required, and in order to ensure rapid scientific advantages without unnecessary duplication of effort, full cooperation of all nations involved in citrus genome research is essential. This document has been developed with the involvement of a multinational community of scientist.

II. SCIENTIFIC GOALS

The specific objectives of the Consortium are briefly outlined.

II.1. EST SEQUENCING AND DATABASE

Expressed Sequence Tags (ESTs) are short DNA sequences, usually based on a single sequencing run, representing genes expressed in an organism under particular conditions. This approach has been proven to be a successful tool for gene discovery, gene mapping, and the study of expression profiles in a wide variety of eukaryotic organisms. Plant species, and particularly crop plants, are not an exception, representing today more than 20% of the total number of entries in the NCBI dbEST database.

Several citrus EST projects are being conducted around the world using different species and cultivars. For that purpose, cDNA libraries obtained from a variety of tissues, organs, growth conditions, and different stages of development are being used. Therefore, the consolidation of such resources represents a unique opportunity to not only join efforts and share information, but also to foster collaboration among Consortium members.

As main goals, we propose to develop mechanisms for efficiently sharing EST data among participating groups, increase the coverage of all expressed genes, use EST data to develop molecular markers such as SNPs and SSRs for genetic and physical mapping, develop the bioinformatic tools necessary for the analysis and manipulation of EST data, develop strategies to minimize redundancy among participating groups, and provide those interested in functional genomics with a set of "unigene" sequences for the construction of microarrays for the analysis of temporal or spatial expression profiles. We strongly recommend sequencing ESTs from both 3' and 5' ends with the purpose of providing the information necessary to both categorize the biochemical nature of the gene product and differentiate individual genes from paralogs or alleles of the same gene.

Manipulating and sharing EST data also requires the use of an efficient and flexible interface that will make such information available to a large group of people without sacrificing quality or speed. Relational databases are a prerequisite for this purpose. We therefore propose to establish a central repository of EST data combining the use of relational databases and a Web-distributed interface. At least three mirror sites will be created and maintained up to date as the information becomes available.

A primary goal of public presentation of EST sequence data is the enable researchers to use this information to address and solve problems associated with

citrus production. It is therefore important to provide a database that presents the sequence with enough associated information that it is useful to a broad range of scientists and presented in an intuitive manner. It is also important to realize that unlike model organism EST projects, a citrus EST database will have sequences from multiple citrus types and therefore greater diversity in the sequence information for each gene loci. The ability to address this sequence diversity must be built into the database. To achieve these goals, we propose that the database present the EST sequence data with the following characteristics and information:

1. A mechanism for continual updating of the annotation of sequence information.

2. Annotation of sequences that includes Gene Ontology assignments consistent with the international Gene Ontology (GO) Consortium.

3. Presentation of sequence alignment into "contigs" that represent grouping of EST sequences from identical or near identical genes.

4. Graphical representation of alternate splicing processes of the same original gene transcript.

5. Ability to export FASTA files of select sequences or EST projects.

6. Posting of chromatograms and/or quality value date for each sequence.

7. On-line virtual northern analysis based on redundancy of sequence appearance within specific libraries.

8. A Unigene set for the pooled ESTs.

9. Ability to search based on GO annotation, for specific metabolic pathways or cellular functions.

10. Display of sequence alignments that facilitate the manual identification of SNPs, and automation for identification of high-confidence SNPs.

11. Operations that allow microarray chip contents to be readily explored using a list of query sequences. This is to address the recurring question of "which genes that were found to be interesting in study X are possible to examine on

my chip"?

12. BAC clone ID# corresponding to an EST, and vice-versa.

13. Nearest relative in fully sequenced genomes, and map location of the nearest relative on the model genome.

14. Key publications related to a given gene.

II.2. CONSENSUS LINKAGE MAP

Genetic markers and mapping

A number of different genetic linkage maps of citrus have been produced in various laboratories (Roose et al., 2000; Ruiz and Asins, 2003), but it is difficult to accurately interrelate the linkage groups identified on different maps because few common markers have been used. It is essential to develop reference maps of the model genotypes that include highly polymorphic markers so that they can be mapped in various populations. This will allow the many maps currently available or under development 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 currently available *C. sinensis x P. trifoliata* and the reciprocal hybrid populations into a large 300+ composite mapping population for generating reference maps of sweet and trifoliate oranges. These genotypes are sufficiently heterozygous for mapping, but it is possible that, due to past inbreeding, some chromosome regions are identical-by-descent and therefore cannot be mapped. To investigate this possibility, a supplementary population involving a *Citrus* x *Poncirus* hybrid as parent should also be identified and used to map a subset of at least 150 well-spaced markers shared with the consensus map and those markers that cannot be mapped on the consensus map. This strategy should define such most regions of the genome. This type of population should also be useful for mapping QTLs that differentiate Citrus and Poncirus, making the map particularly useful. Several such populations are available including both backcross and near-F2 configurations. 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 cleaved amplified polymorphic sequences (CAPS), expressed sequenced tags (ESTs), resistance gene candidates (RGCs), or single nucleotide polymorphisms (SNPs). It is essential that markers be single copy and have known sequences so that they can be easily integrated with the physical map and genome sequence. 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, CAPS and/or other markers derived from ESTs should also be included on the map for synteny comparisons with other species.

As citrus EST databases grow, it will be possible to identify SSR sequences within ESTs and use these to develop additional informative markers. Currently, of 10,134 sequences in the database that assemble into 1518 contigs and 3500 singletons, 210 contain SSRs (http://cgf.ucdavis.edu/). Thus, although only a small proportion of ESTs are likely to contain an SSR, EST sequences will provide a large number of potential SSR markers.

Initially, each laboratory should map those markers they already have available on the mapping population. Primer sequences for these markers should be deposited in a secure consortium web site, which can be accessed only by consortium members. Consortium members agree not to publish independently results derived from these markers without written permission of the primer developer. For EST markers, the consortium web site should include information on whether the marker has been confirmed by sequencing as amplifying the putative gene. This confirmation should eventually be provided for all markers.

Methods to relate the map to traits should be developed and implemented by individual groups involved in the consortium. Therefore laboratories involved in mapping are encouraged to investigate mechanisms for exchange of seed material (quarantine regulations make exchange of vegetative material prohibitively expensive) from the mapping populations so that QTL analyses can be conducted on this population in multiple environments.

Maps should also be used to facilitate studies of genetic diversity in citrus. This can be done by genotyping a subset of about 50 citrus cultivars and germplasm for a set of perhaps 100 highly polymorphic, efficiently spaced markers, probably genes likely to influence commercially important traits. The genotyping strategy should have sufficient power to identify evolutionary relationships between allele. Most likely, this will involve cloning and sequencing alleles from each targeted genotype. Essentially this approach will provide information on evolutionary relationships between genome segments in various cultivars, and also provide valuable information about functional genomics of these genes as the cultivars studied can be characterized for the traits putatively influenced by the genes.

Short term tasks:

1. To establish a large (over 300 hybrids) reference segregating population from an intergeneric cross for the production of a consensus map and develop a mechanism for the distribution of DNA and plant material to interested groups in order to replicate the common progeny at several locations differing for environmental variables.

2. Each group will genotype the common mapping population for their own two-primer-based markers. Marker genotypes will be shared through Excel files or a secure consortium web site. PCR based markers should eventually be verified by sequencing the mapped alleles. If five groups participate, the initial target for this effort will be to genotype at least 500 markers (in either or both parents) by the end of 2004, and 1500 markers by the end of 2005.

3. Each group will obtain linkage maps of the parents using their own software. All maps should be very similar and will be used to write a common paper. The consensus maps will be useful to study reliability, synteny and degree of genome coverage of previous linkage maps developed by each group for QTL analysis and breeding aims.

4. At least one group should develop a low-resolution (150-200 marker) map for a *Citrus* x *Poncirus* hybrid parent. This map should include well-spaced markers from the consensus map and also markers that cannot be mapped in the main *C. sinensis* x *P. trifoliata* population because they appear homozygous for different alleles in the parents. Such markers might represent genomic regions that are homozygous in one or the other of the parents, and which would otherwise be missing from the map.

Long term tasks:

5. The common mapping population will be used to map (non-redundant) ESTs in order to highlight EST-rich genomic regions. This information might be useful to focus sequencing efforts towards these regions. Each group should map a different set of ESTs and will proceed as in task 2.

6. Consortium members are encouraged to explore mechanisms to establish multi-location trials of the mapping population suitable for QTL analyses. This will allow studies of the extent of QTL x environment interactions in a common set of agronomic traits (i.e. yield components, general adaptation, alternate bearing...) taking into account data from hybrid evaluation and environmental variables at each location. Similar efforts should be encouraged using the ancillary Citrus x Poncirus mapping populations.

7. A subset of informative, efficiently spaced EST markers or sequences should be used to characterize a standard set of Citrus species, varieties, and related genera.

8. The linkage mapping group should coordinate mapping efforts with the physical mapping group so that as BAC ends are sequenced, markers derived from these sequences are developed and mapped. A target of perhaps 5000 mapped markers will be needed to fully support a physical map and genome sequence. Chip-based methods (Borevitz et al., 2003 - Genome Res. 13:513-523) to efficiently map very large numbers of markers should be investigated to facilitate this aspect of the project.

All results will be used for publications in common.

II.3. PHYSICAL MAP

The development of large insert DNA libraries has allowed the construction of genome-wide physical maps for a few species, including Arabidopsis, rice, human (International Human Genome Mapping Consortium 2001), mouse, and soybean. Development of a physical map of the sweet orange genome, and its integration with a genetic linkage map, will be useful for application of more rapid and efficient approaches to cloning economically important genes known only by their phenotype. Such an integrated map can be a powerful tool for marker-assisted selection by citrus breeders, and can be used to study and understand the physiological and molecular mechanisms underlying economically and biologically important traits. A physical map can also form the basis for functional genomic studies to identify and isolate economically important genes. Studies of the organization and evolution of genomes are facilitated through availability of integrated physical and genetic maps. Finally, a high quality physical map provides the essential basis for efficient larger scale genome sequencing, leading to a full-length genome sequence. The basic approach requires the construction of two sweet orange BAC libraries of sufficient size (number of clones and average insert size), produced using different restriction enzymes, to ensure adequate coverage (>7x genome coverage). Use of complementary libraries from different restriction enzymes has been shown to minimize the number of gaps between contigs, thus extending the lengths of BAC contigs. Several approaches have been developed for physical map construction from BACs, including those based on chromosome landing, subcloning, fingerprinting, or BAC end sequencing/fingerprinting. Recent research has demonstrated the feasibility of genome physical mapping by fingerprinting BACs on sequencing gels or capillary sequencers, and very efficient technologies using capillary sequencers, for example, are being developed that lead to greater automation, greater efficiencies, and improved resolution. Genetic data generated from mapping studies performed within the citrange families, described in the section on Linkage Mapping above, will then be integrated into the constructed physical map.

The highest resolution physical map of a genome is the complete DNA sequence of each chromosome, or in other words, one marker at each nucleotide of the citrus genome. However, lower resolution physical maps that are easier to

construct would enable addressing important short-term goals that will then lead to the complete sequence. Of greater utility to the citrus genomics community will be a medium to high-resolution physical map of sweet orange that combines ordered sets of cloned DNA inserts anchored by sequence-tagged connectors (STCs). Relatively low pass efforts should generate large contigs with an STC marker every 0.1Mb. Gap closure and higher resolution marker density will be obtained by linkage of the genetic and physical maps, and by placing ESTs, SSRs, and BAC end sequences on these maps.

A well-developed and characterized genetic map is crucial for orienting and anchoring BAC contigs assembled from physical mapping and for subsequent use of the physical maps in map-based gene isolation and cloning. The nature and types of citrus genetic linkage maps have been described in the section on Linkage Mapping above. Utilization of the citrange "super-family" described therein, as a common base linkage map for the international effort will enable the participant groups to work on a common set of biological materials, and through utilization of SSRs, ESTs, and BAC end sequences, it should be possible to efficiently integrate the genetic and physical maps. There are some inherent advantages to using BAC end sequences as part of the process of genetic map construction and integration with the physical map. Specifically, doing so can allow simultaneous achievement of an element of genetic mapping, anchoring and orienting BAC contigs, and providing contig coalescence sequence tags for contig mergence and local genome sequencing. Thus specific primers could be designed to amplify sequences by PCR to identify polymorphisms from *Citrus sinensis* and *Poncirus trifoliata*, the two parental species used to create the super-family. Because of the close relationship of *Citrus* with *Poncirus*, the majority of polymorphisms identified will be either differences in fragment length or internal restriction site differences. There also will likely be some presence/absence polymorphisms, though previous experience with certain limited genomic regions between *Citrus* and *Poncirus* has indicated that the former type will predominate. Fragment length differences should be obvious following amplification, but identification of restriction site differences requires digestion with a battery of different restriction enzymes after amplification, followed by electrophoresis. Again based on previous experience, it is reasonable to predict conservatively that a minimum of 50% of the BAC ends might reveal polymorphism between the two parents. Assuming that 1000

randomly selected BAC clones were end sequenced it would provide a minimum of 500 segregating markers that can be used simultaneously to produce a genetic map and to orient and anchor BAC contigs. If the fragment length and internal restriction site classes predominate, then the vast majority of the BAC end markers will serve as co-dominant markers. Assuming that the citrus genome extends 1500 to 1700 cM, genetic mapping of 500 BAC ends will locate one marker every 3 - 5 cM, with a genetic mapping resolution of approximately 0.5 cM. The development of a genetic map with this marker density, along with mapped SSRs and ESTs, and integrated with BAC contigs, will be invaluable for the genetic improvement and engineering of citrus rootstock and scion cultivars. Having the highest quality integrated physical and genetic map possible, then, will provide invaluable information to direct the full genome sequencing effort that is the ultimate objective of the International Citrus Genomics Consortium.

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.

II.4. WHOLE GENOME SEQUENCE

The sequencing of the citrus genome (380 Mb) will be an efficient approach for positional cloning and gene identification, and will complement EST sequencing, physical mapping and also, the development of functional genomic tools. In addition, the genome sequence will help to understand how the genome as a whole works to direct the growth, development and homeostasis of an entire organism.

The methodology proposed is based on a strategy that combines whole genome shotgun sequencing (WGS), a BAC clone approach and clone array pooled shotgun sequencing (CAPSS). This mixed approach offers several advantages since it combines the fastness of the WGS, the accuracy of the BAC clone approach and the cost and work reduction of the pooled array strategy.

The strategy suggested would be as follows:

1. Whole genome shotgun (WGS) with 3 kb, 10 kb and 50 kb libraries. Systematic sequencing of approximately 3,5 millions of readings.

2. Construction of several genomic BAC libraries with different partial digests of restriction enzymes to generate de inserts. The isolation of about 50.000 BAC clones would provide a 20-fold coverage of the genome.

3. Fingerprinting of the BAC clones, for the purposes of preparing contigs and checking the integrity of the clones. Considering the expected amount of repetitive DNA to be found, the information provided by the fingerprinting will be invaluable where repeated sequences make BAC sequences ambiguous.

4. BAC end sequencing of the isolated BACs. The BAC end sequencing should provide an STS every 3-5 kb on average, allowing the picking of clones with a minimum overlap, and providing further information for the physical map.

5. Determination of the BAC tiling path to minimize overlapping (physical map).

The construction of the physical map, constitutes an huge effort, that is justified because it will allow the BAC tiling path to be determined with minimal overlap at the outset of the sequencing.

6. Integration of the physical and the genetic maps. Whenever possible, a correspondence between the physical map and the linkage genetic maps, will be established.

7. Shotgun of the selected BAC clones (around 5000) using the pooled array strategy. The BACs are distributed in arrays of variable dimensions and the clones are pooled row-wise and column-wise. DNA preparations are made from each of the pools and shotgun libraries and sequencing is performed on the pools. The sequences are deconvoluted by co-assembling sequences from each row with each column and identifying those contigs formed by mixtures of row and column reads.

8. Assembly of WGS and BAC readings. Up to a 6x coverage is proposed. Approximately, a 1-2x coverage will be obtained with the BAC shotgun readings (around 1,5 million readings), and the remaining 4x coverage will come from the whole genome shotgun sequencing. The WGS reads will be assigned to the proper BAC and assembled using the ATLAS whole genome assembler, the method developed for assembling the rat genome sequence from a mixed WGS-BAC approach. The consensus sequence will be determined.

9. Annotation and sequence release. The most useful releases for the scientific community are large contiguous stretches of annotated sequence. The annotation should delineate potential open reading frames and identify regions of homology, especially with the EST sequences from citrus and other plant species.

10. WWW-Database. As a fundamental tool to co-ordinate the project, check the progress of the different steps and guarantee the access of the scientific community to the data even before the final assembly is completed, a database accessible via www will be constructed. The implementation of a suite with the

most common bioinformatic tools, as well as those necessary for clone fishing would facilitate the database search in a user friendly environment.

Being conservative, this approach would require about 5.2 million successful reads for a 6x sequence coverage (500 b per read). In addition, the project will require a few hundred DNA preparations and shotgun libraries. It may take two years of work with and 10 million cost.

II.5. FUNCTIONAL GENOMICS
II.5.1. RAPID CYCLING GENOTYPE

A short juvenile period genotype is necessary for the evaluation of transgenic genotypes. Commercially important traits of citrus are expressed primarily in fruit tissue. This requires that trees be capable of flowering and producing fruit in order to evaluate any genes affecting fruit traits. The long juvenile phase of four to six years of most citrus genotypes impedes the rapid evaluation transgenic trees modified to affect adult traits such as fruit quality (e.g., rind and flesh color, flavors, maturity dates, seediness, abscission, peelability, rag content, acid and sugar levels), and other traits associated with productive mature trees. A model rapid cycling citrus should be precocious and have a seed-to-seed generation cycle of less than one year. The genotype should be self-compatible to allow the production of selfed seed if necessary. The genotype should weakly apomictic producing 60-70% polyembryonic seed to allow true to type propagation via seed. It should produce fruit of sufficient size to allow the juicing and evaluation of fruit tissues for different traits. The genotype should be amenable to transformation, allowing the easy generation of transformed plants.

Several citrus genotypes have some of the characteristics described above. Key lime and some limequats have juvenile periods of under 1 year. They produce fruit of sufficient size for the evaluation of juice and fruit tissues. Sufficient female fertility exists in key lime to allow its use as a female parent. Transformation and regeneration systems have been reported for Key Lime. This information is not available for limequats.

Another potential model system are the early flowering transgenic genotypes generated in Spain. These plants expressing flowering genes from Arabidopsis have very short juvenile periods would be another good possibility.

II.5.2. MICROARRAYS

Transcript profiling is considered to be the first step in correlating gene expression with specific biological processes. As cells change in response to developmental or environmental stimuli, the composition and abundance of individual mRNAs change accordingly to reflect needed changes in protein and metabolite production. Profiling transcript composition will then allow us to gain insight into the characterization of the genes involved on the processes of interest. DNA microarrays have arisen as a largely reliable high throughput system for transcription profiling. The construction of microarrays and the sharing of expression data within the Consortium will promote rapid advances in Citrus functional genomics.

There are a number of choices to produce gene arrays. The Consortium will decide the definitive system based on availability of economical and technological resources of the member groups. Microarray design, including clone selection and elimination of redundancy will be discussed by microarray experts from every group and by the Steering Committee.

In any case, taking into consideration the present resources in the Consortium, cDNA arrays of EST clones will be produced in a first stage, and released to all members willing to use them. A considerable effort has been made by many groups of the Consortium in isolating ESTs from different Citrus species. Sharing those ESTs clones will be the key aspect to develop this common microarray and is greatly encouraged.

The construction of a gene-specific oligonucleotide chip based on EST/genomic sequence data will expand microarray applications in the second stage of the Microarray Program of the Consortium. The system of choice will depend on the availability of resources to produce an array set.

The data obtained by the use of cDNA microarrays provided by the Consortium will be released to a common Citrus expression database which will be tightly linked to the EST database. This database will be open to every member of the Consortium. The shared Consortium data will be presented in forms which would be MIAME compliant.

II.5.3. METABOLOMICS AND PROTEOMICS TOOLS

PRINCIPLES AND GENERAL AIMS

Metabolomics aims to separate, identify and profile compounds which determine traits, such as flavor, aroma, color, and tree responses to various stimuli. Proteomics aims to separate, identify and profile proteins that play a role in determining traits and specific processes. These also include those not regulated at gene expression or mRNA levels.

These approaches allow:

1. For a specific cultivar, to correlate between compounds and/or proteins and a given trait or process.
2. To compare between cultivars displaying different traits, and to identify the compound(s) and/or protein(s) associated with them.
3. To identify compounds of pharmaceutical, industrial and commercial values, such as antioxidants, and unique aroma and flavor molecules.

Metabolomics and proteomics of the model genotype would allow validating results derived from DNA microarrays and EST studies, by verifying protein expression, and allow the subsequent coordination of gene transcription with protein expression and changes in metabolite level. In addition, these studies would allow to validate the effect of newly introduced or modified genes in the model genotype on protein expression and metabolite level and profile.

METHODOLOGIES

Both metabolomics and proteomics are usually more useful when the complexity of a given organ or a tissue is reduced by fractionation to sub-organ, sub-tissue and sub-cellular fractions.

Metabolomics involves tandem techniques for the separation of an extract, and the identification of specific compounds. It involves the combination of liquid chromatography (LC)-mass spectroscopy (MS), Gas chromatography MS (GS-MS),

Fourier transform ion cyclotron resonance MS (FT-MS) and NMR technologies.

Peptide mapping, amino acids sequencing by MS-based techniques has become relatively inexpensive, allowing protein identification based on sequence homologies. Protein separations by 2D-gels, followed by matrix-assisted laser desorption ionization time of flight (MALDI-TOF)-MS or MS-MS for protein identification are widely used. Moreover, recent technologies also involve tandem techniques for separation and identification with two-dimensional LC coupled to MS-MS.

Outcome for Consortium members.

Released data should include the following:

1. 2D maps of proteins for a given cultivar, organ, fraction, etc.
2. Peptide sequences.
3. Profiles of compounds related to specific trait and cultivar.

II.5.4. TRANSIENT EXPRESSION TOOLS.

DEVELOPMENT OF A CITRUS VIRUS VECTOR.

Citrus breeding is limited by the lack of information about genes of potential interest to improve characters associated with the commercial value of fruits. The ability to sequence plant genomes has resulted in the identification of large numbers of novel open reading frames (ORFs). Large-scale functional genomic approaches are necessary for converting this sequence information into functional information. As an alternative to genetic transformation, virus vectors have been shown to be a useful strategy for expression or silencing of plant genes (Angell et al., 1999; Burton et al., 2000; Holzberg et al., 2002; Kjemtrup et al., 1998; Lacomme et al., 2003; Liu et al., 2002; Ratcliff et al., 2001; Ruiz et al., 1998; Yoshioka et al., 2003). Inoculation of plants with virus vectors is a direct way to assay the function of specific genes without the time consuming process of plant transformation and regeneration. This new approach is especially interesting in woody plants like citrus, with long juvenile periods (up to 6-8 years), in which transformation of adults plants is very difficult and long periods between transformation and fruit bearing are necessary.

Up to date, no virus vectors have been developed for citrus. Recently, the complete sequence of a new virus, the citrus leaf blotch virus (CLBV) has been obtained (Vives et al 2001). The properties of CLBV indicate that it can be used as a vector for expression or silencing genes in citrus plants because: i) It is symptomless all commercial cultivars assayed; ii) In contrast with other viruses like CTV which is phloem limited, CLBV replicates in all citrus tissues; iii) It accumulates mainly in meristematic tissues, thus offering an interesting model system to study genes involved in growth and development of leaves and fruits; vi) Its monopartite genome of 8747nt containing three ORFs probably will be easy to manipulate; v) It is mechanically transmissible to citrus, which would facilitate inoculation of engineered versions carrying the foreign genes .

OBJECTIVE

Development of a citrus virus vector as a tool for quick evaluation of the function of citrus or other plant genes.

BIBLIOGRAPHY

Angell, S. M. and Baulcombe, D. C. (1999) Potato virus X amplicon-mediated silencing of nuclear genes. *Plant Journal* **20**, 357-362.

Burton, R. A., Gibeaut, D. M., Bacic, A., Findlay, K., Roberts, K., Hamilton, A., Baulcombe, D. C. and Fincher, G. B. (2000) Virus-induced silencing of a plant cellulose synthase gene. *Plant Cell* **12**, 691-706.

Holzberg, S., Brosio, P, Gross, C. and Pogue, G. P. (2002) *Barley stripe mosaic virus*-induced gene silencing in a monocot plant. The Plant Journal **30**, 315-327.

Kjemtrup, S., **Sampson, K. S., Peele, C. G., Nguyen, L. V., Conkling, M. A., Thompson, W. F. and Robertson, D.** (1998) Gene silencing from plant DNA carried by a geminivirus. *Plant Journal* **14**, 91-100.

Lacomme, C., Hrubikova, K. and Hein, I. (2003) Enhancement of virus-induced gene silencing through viral-based production of inverted-repeats. The Plant Journal **34**, 543-553.

Liu Y., Schiff, M. and Dinesh-Kumar, P. (2002) Virus-induced gene silencing in tomato. The Plant Journal **31**, 777-786.

Ratcliff, F., Martín-Hernández, A. M. and Baulcombe, D. C. (2001) Tobacco rattle virus as a vector for analysis of gene function by silencing. *Plant Journal* **25**, 237-245.

Ruiz, M. T., Voinnet, O. and Baulcombe, D. C. (1998) Initiation and maintenance of virus-induced gene silencing. *Plant Cell* **10**, 937-946.

Vives, M. C., Galipienso, L., Navarro, L., Moreno, P., and Guerri, J. (2001) The nucleotide sequence and genomic organization of citrus leaf blotch virus: a candidate member of a new virus genus. Virology **287**, 225-233.

Yoshioka, H., Numata, N., Nakajima, K., Katou, S., Kawakita, K., Rowland, O., Jones, J. D. G. and Doke, N. (2003) *Nicotiana benthamiana* gp91[phox] homologs NbrbohA and NbrboB participate in H_2O_2 accumulation and resistance to *Phytophtora infestans*. The Plant Cell **15**, 706-718.

II.5.5. GENETIC TRANSFORMATION OF CITRUS

Efficient transformation systems are required for citrus for two reasons. The first reason is to identify gene function and the second, and most important, is the development of new and improved citrus varieties, or more precisely to modify otherwise acceptable citrus varieties that are deficient in one or a few characteristics. Transforming a plant with a gene with an unknown or putative function allows researchers to determine if a gene has an observable effect on the phenotype of the plant. Information on gene function is essential for the applied use of transformation for variety development, the primary objective in this white paper.

SCIENTIFIC CONSIDERATIONS:

To effectively utilize genetic transformation in citrus variety development requires the development of individual component and associated technologies. These technologies include the following:

1. Mature tissue transformation. An efficient method to transform mature tissue from commercial citrus types would bypass the long juvenility period and permit the direct testing of putative gene candidates.

2. Non-antibiotic selection systems. Systems to efficiently select transformed cells and tissues without the use of antibiotics should be developed for citrus. Antibiotic selection is a major concern of the public and nonantibiotic systems would help to address some of the public's concern over this aspect of genetic transformation. Negative (e.g., phosphinothricin) and positive (e.g., mannose) selection compounds should be developed for use in citrus transformation. Systems that require no selection (e.g., GFP) should also be considered.

3. Linkage group transformation. Transformation by large DNA inserts, such as BIBACs, containing multiple genes in linkage groups would be useful in both facilitating the positional cloning and functional analysis of candidate genes and for introducing multiple genes for engineering complex secondary product pathways into citrus.

4. Gene targeting. Targeting gene insertion by homologous recombination into the citrus genome would allow disruption of gene function (knock-outs), restoration of function of defective genes (gene therapy), or replacement of a gene with a new or altered version.

5. Promoter library. Implementation of transgenic technology in citrus variety development requires a repertoire of promoter and cis-acting elements to control gene expression. Such control is complex and requires the coordination of appropriate expression levels in specific tissues or cell types, at specific developmental stages, and under various environmental induction conditions (e.g., chemical sprays, insect feeding, pathogen infection). Constitutive regulation, though appropriate for some applications, is not suitable for effectively modifying complex metabolic processes that must be controlled in a highly regulated manner.

6. Gene stacking. Current transgenic technology typically utilizes a single gene. However, modifying multiple or complex characteristics of a tree will require the insertion of multiple genes. Because of a long juvenility combining multiple transgenes into a single tree by conventional breeding would take a long time. Therefore, gene stacking in citrus will require transformation with multiple genes or the sequential transformation of already transformed trees.

7. Gene removal. Removing genes once they are no longer useful would be beneficial in a number of applications. For example, the removal of selectable marker genes would permit both gene stacking and facilitate the public acceptance of transgenic trees. Or, the removal of genes that may be useful when the plant is small and juvenile, but may not be required in the mature tree; this would provide the benefits of transgenic traits but result in nontransgenic fruit.

8. Viral-mediated transformation. One problem of genetic transformation is that growers must plant the new transgenic trees. Current production areas would not immediately benefit from the technology. A technology utilizing viral vectors to transforming citrus would permit application of genetic transformation to existing grove plantings. Such a system would require a nonpathogenic virulent viral

vector and a method to disperse the virus, presumably an insect vector. One possible advantage of a nonintegrative viral transformation system is that the phenotype of the tree might be modified by simply deploying a different engineered form of the virus.

POLITICAL CONSIDERATIONS:

In addition to the scientific requirements of developing effective genetic transformation systems, there exists a strong political component to genetic transformation research. Two considerations that should be considered by both researchers and legislatures are as follows:

1. Field testing of transgenic trees. To fully develop this research requires the ability to adequately grow and test these transgenic trees in the field. This requires legislative action to provide appropriate protocols and safeguards, where none currently exist. Without this ability, the effort and resources used to develop these trees, including the resources in sequencing and mapping, is greatly reduced. Where no legislation exists or while legislation is being developed to field test transgenic trees, then collaborative field plantings should be encouraged between nations that have legislative approval and those that do not.

2. Transgenic rootstocks. The initial research and development of transgenic citrus trees should emphasize and encourage the modification of rootstock characteristics. A transgenic rootstock will have a greater public acceptance than a transgenic scion.

II.6. BIOINFORMATICS TOOLS

A crucial point in the success of a research project on genomics is the ability of extract and produce new biological knowledge from the vast amount of data generated in the project. Since the genomics data are in the order of terabytes of information, this task can only be accomplished by using the power of modern computers, originating the so called field of Bioinformatics. Therefore, the International Citrus Genomics Consortium has a special interest in the use and development of bioinformatics tools to achieve its goals. Furthermore, by means of the ever growing availability of high speed network connections on Internet, these bioinformatics tools, together with the Consortium database which they will operate upon, will constitute a very important joining element between the different members of the Consortium.

According to this, all the bioinformatics tools will be hosted together with the Consortium database, and mirrored in a few countries participating in the Consortium for security and efficiency reasons. In this manner, the bioinformatics tools developed by the different members of the Consortium to exploit the common database will be made available to the rest of participants. A close connection among the bioinformatics people in the different members of the Consortium is intended in order to avoid duplication of programming efforts. This connection could have the task groups of the Open Source community (http://sourceforge.net) as a model, and the software products developed will be (preferably) open source and available to the scientific community upon request. This will allow their customization for other species genomics projects, and to get the benefits of interaction with the global open source community.

A first step in the development of bioinformatics tools will be the implementation of a user friendly interface to access the database. This interface will mainly consist of web forms accessible with a web browser from anywhere by the Consortium members, who will log in to the database with a password. The database interface will be hosted in the International Citrus Genomics Consortium web page, also mirrored in different servers together with the database and the rest of bioinformatics tools. This interface will have to cope with the wide range of member's interests (EST data, linkage maps, breeding markers, genomic sequences, etc.), so it will need a global agreement and a clear and intuitive way of linking these different sections of the database. The

experience accumulated by other more advanced genomics projects (Arabidopsis, rice, human, Drosophila) will be of a great value and could be used as a model.

Another important set of bioinformatics tools to be developed is that concerning the annotation of the genomic sequences as they are released by the sequencing efforts of the Consortium. Again in this respect, the precise implementation of these tools could be inspired by the already finished and ongoing sequencing projects in other species.

Finally, other bioinformatic tools will be developed as needed by the Consortium members, and will address more specific research topics in the different Citrus genomics fields. As indicated above, they will be available from the International Citrus Genomics Consortium web page (and its mirrors) to the Consortium members and, upon request, to the rest of the scientific community, as web services or downloadable standalone programs. In this regard, a high level of participation of the Consortium members is encouraged in identifying new methods to exploit the data, improving their presentation, and facilitating their distribution to
the scientific community as a whole.

III. WORKSHOPS AND SYMPOSIA

Suggestion taken from the International Grape genome Program White Paper

(www. Vitaceae.org)

Research Workshops and Symposia are a very important part of the Program. In previous years there have been periodical meetings focused around Citrus Genetics and Breeding as well as Physiology and Biotechnology. These almost yearly meetings should in the future include a specific component, such as a workshop, for discussions about the progress of the Citrus Genome program and the individual projects. It is also recognized that more specific Workshops and Symposia will also be required either within the frame of these meetings, at the annual Plant and Animal Genome meeting in San Diego, or independently organized by the Steering Committee.

Short term goals:
Establishment of a calendar of events for the International Citrus Genome Project.

Long term goals:
Regular annual International Citrus Genome Project meetings.

IV. SOCIAL AND LEGAL ISSUES

Suggestion taken from the International Grape genome Program White Paper (www. Vitaceae.org)

Success of the International Citrus Genome Project strongly depends on the participant's agreement to share the results of their individual efforts. As the number of participants increase, the benefits of sharing results increases as well as the availability of public information. As a rule, the matter of intellectual property rights should be handled according to the legal convention of the country that provides for the individual research project in question. If there are any restrictions which potentially prevent free exchange of ideas, information and materials for research purposes, they should be identified and made widely known so that rights of individual investigators to intellectual property will not be inadvertently violated. When research projects are conducted under official joint agreements between nations, such agreements should contain clauses on the intellectual property rights in accordance with the international agreement on the intellectual property rights.

Participation in this International Citrus Genome Project does not necessarily require the exchange of funds between cooperating countries. Each country or entity will provide financial support tot he activities of its own scientists.

Release of genetically engineered plants into the environment for research purposes should follow the guidelines and applicable regulations of the country where the field tests are being conducted.

APPENDIX

List of Current Working Groups and Contact Person

I. INTRODUCTION

Vicente Conejero Tomás

Email: vconejer@ibmcp.upv.es

Fax: 34 96 387 78 79

II. SCIENTIFIC GOALS

II.1. EST SEQUENCING AND DATABASE

Gustavo Astua Monge

Email: gamo@centrodecitricultura.br

Fax: 55 19 3546 1399

II.2. CONSENSUS LINKAGE MAP

Mikael Roose

Email: roose@citrus.ucr.edu

Fax: 1 909 787 4437

II.3. PHYSICAL MAP

Fred Gmitter

Email: fgg@lal.ufl.edu

Fax: 1 863 956 4631

II.4. WHOLE GENOME SEQUENCE

Manuel Talón

Email: mtalon@ivia.es

Fax: 34 96 363 6301

II.5. FUNCTIONAL GENOMICS
II.5.1. RAPID CYCLING GENOTYPE

José X Chaparro

Email: jchaparro@ushrl.ars.usda.gov

Fax: 1 772 462 5986

II.5.2. MICROARRAYS

Jose Gadea Vacas

Email: jgadea@ibmcp.upv.es

Fax: 34 96 387 78 59

II.5.3. PROTEOMICS TOOLS

Avi Sadka

Email: vhasadka@volcani.agr.gov.il

Fax: 1 972 3966 9583

II.5.4. TRANSIENT EXPRESSION TOOLS

Jose Guerri   Email: jguerri@ivia.es

           Fax: 34 96 342 40 01

II.5.5. TRANSFORMATION

Randall P Niedz   Email: rniedz@ushrl.ars.usda.gov

           Fax: 1 772 462 5961

II.6. BIOINFORMATICS TOOLS

Javier Forment   Email: jforment@ibmcp.upv.es

           Fax: 34 96 387 78 59

III. WORKSHOPS AND SYMPOSIA

Vicente Conejero Tomás   Email: vconejer@ibmcp.upv.es

           Fax: 34 96 387 78 79

IV. SOCIAL AND LEGAL ISSUES

Vicente Conejero Tomás   Email: vconejer@ibmcp.upv.es

           Fax: 34 96 387 78 79