SEQUENCE-TAGGED SITE MARKERS IN GRAPEVINE AND CITRUS.

M. R. THOMAS, N. S. SCOTT, R. BOTTA[1] AND J. M. H. KIJAS[2]
CSIRO Plant Industry, Horticulture Unit, Adelaide, GPO Box 350, Glen Osmond, SA 5064, Australia; [1]Centro di Studio per il Miglioramento Genetico e la Biologia della Vite, CNR, Via Pietro Giuria 15 , 10126 Torino, Italy; [2]Swedish University of Agricultural Sciences, Uppsala, Sweden.

Abstract
DNA markers based on PCR are becoming increasingly important in a wide range of applications including cultivar identification, phylogenetics and the construction of linkage maps for marker assisted breeding and map-based cloning of genes. We have investigated the utility of the microsatellite sequence-tagged site marker type because of the high level of associated polymorphism and simple genetics. For grapevine cultivar identification the marker type has been very successful. For phylogenetic analysis, studies with *Citrus* and related genera suggests that microsatellite length variation may be better suited to construct relatedness over small genetic distances rather than large genetic distances. The integration of trinucleotide microsatellite markers into a linkage map of *Citrus* has demonstrated the utility of this marker type for genetic analysis within wide intergeneric crosses and the potential to act as "anchor loci" to align linkage maps from different crosses and laboratories.

Introduction
Fruit crops such as grapevine and citrus are similar in that they are long-lived perennials and once a cultivar is bred it is vegetatively propagated via cuttings. This has resulted in many cultivars having very long cultivation histories of often hundreds of years. The major impediment to genetic improvement of grapevine and citrus is the long juvenile period of 3-10 years before flowering. Evaluation of fruit and plant performance can extend for a number of years which can result in a 15-20 year period from the time of a cross to the release of a new cultivar. In addition some citrus exhibit polyembryony which results in seeds that contain both zygotic and apomictic embryos. DNA marker technology has the potential to reduce or remove these limitations to genetic improvement by providing accurate genetic identification, determination of genetic relatedness and marker-assisted breeding with desirable traits identified at the juvenile stage.

Microsatellite STS markers
The existence of microsatellite sequences in plants and algae was first reported in 1986 (Tautz et al. 1986). Grapevine (Thomas et al. 1993, Thomas and Scott 1993) and citrus (Kijas et al. 1994, Kijas et al. 1995) have been found to contain di-, tri, and tetranucleotide microsatellite sequences. These plant microsatellite DNA sequences are highly polymorphic, inherited in a co-dominant manner and easily analysed by the use of the polymerase chain reaction (PCR) when converted to STS markers (Kijas et al. 1995, Thomas et al. 1994, Thomas and Scott 1993).

A major barrier preventing more widespread use of sequence-tagged microsatellites is the time and expense associated with the conventional method of marker development. This becomes especially relevant within a small research program investigating a little studied genome. A method designed to speed marker development has been devised to reduce laborious screening of genomic libraries for microsatellite repeats. Biomagnetic separation using magnetic particles was applied to capture and enrich for microsatellite sequences (Kijas et al. 1994, Kijas et al. 1995). In our experience approximately 21% of clones from an enriched library contained a microsatellite sequence. Other enrichment methods have also been developed (Kandpal et al. 1994, Karagyozov et al. 1993, Kiyama et al. 1994). Attrition rates during conversion of sequence information to an STS marker were found to be similar for microsatellite enriched libraries from citrus DNA and non-enriched libraries from hops DNA (Brady et al. 1996) with the percent of useful sequences being approximately 25%. The major difference between the two libraries was that 29% of clones from the enriched library were rejected because they had identical sequence to a previously sequenced clone. Identical clones were not observed for the non-enriched library.

Cultivar Identification
Accurate grapevine identification is necessary because of the global problem which has arisen as a result of the long history of cultivation. Distribution of vegetative cuttings into new

geographical regions of diverse climate has caused changes in phenotypes resulting in many cultivars being wrongly identified and renamed. The spread across cultural boundaries has also increased the problem due to different countries or regions adopting different names for the same cultivar. Ampelography was formalised at the beginning of the twentieth century to describe and identify cultivars based on phenotypic traits (Viala and Vermorel 1909). However, the difficulties associated with this phenotype-based ampelographic system are due, in large part, to the subjectivity of the methodology and the global distribution of grape growing regions. There are upward of 24,000 names for a *Vitis vinifera* population thought to contain around 5,000 genuine cultivars (Truel et al. 1980).

Grapevine identification by DNA profiling using microsatellite STS markers (Thomas and Scott 1993) offers several advantages over ampelography. The method is objective, being based on genotypes rather than phenotypes and the analysis is rapid requiring only small amounts of tissue. The genetic stability of vegetatively propagated cultivars also makes a DNA identification system very attractive. A semi-automated system for long-term studies where DNA profile data can be collected and collated from many gels into a database over a period of years has been developed (Thomas et al. 1994). The semi-automated system has been used in our laboratory since 1992 and has been successful adopted by the Australian Wine Research Institute to provide a commercial service. DNA extracted from leaves, wood, roots, canes and berries can be used to identify cultivars. The ability to extract DNA from wood and roots is important when the identity of a rootstock needs to be determined after the grafting of a scion onto the rootstock. Wineries are interested in a quality control method that confirms that grapes delivered by growers are from the right cultivar and DNA profiling has been used to sample grapes prior to and after crushing.

The information in the DNA profile database has been used to objectively investigate cultivar identification errors that may have occurred as a result of mislabelling, multiple naming of a single cultivar, or incorrect ampelographic identification (Thomas et al. 1994). In Australia, incorrect identification has resulted in mixed plantings of Merlot and Cabernet Franc and mislabelling of Chenin Blanc and Crouchen as Semillon. DNA profile results for these cultivars show that they are all easily distinguishable with Merlot and Cabernet Franc differing at 6 of the 7 loci tested. A DNA profile study of grapevine cultivars from Italy has also confirmed that the cultivars Favorita, Vermentino and Pigato are actually only one cultivar (Botta et al. 1995).

From the data collected grapevines are very polymorphic at microsatellite loci and show high heterozygosity not only within *V. vinifera* but in other *Vitis* species as well (Botta et al. 1995, Thomas et al. 1994, Thomas and Scott 1993). This high degree of genetic diversity, especially within the species *V. vinifera* may be explained as a combination of the hypervariability of microsatellite DNA sequences, grapevines low tolerance of inbreeding and the genetic fixation of superior phenotypes by clonal propagation in many cases hundreds of years apart and from different geographical regions.

Phylogenetics and relatedness

DNA profile data has proved useful in parentage analysis to determine the origin of cultivars. Examples exist confirming and disproving parent assignment of grapevine cultivars (Thomas et al. 1994). In one of these cases the major cultivar in Germany, known as Müller Thurgau, was found by DNA profiling not to be a result of a Riesling X Sylvaner cross. Riesling was confirmed as the maternal parent but Sylvaner was not the paternal parent as it lacked alleles in common with Müller Thurgau at 4 of the 7 loci assayed.

Although citrus taxonomy has been well studied confusion remains regarding the number of true citrus species and their relationship with each other. This arises due to the ability of many species to interbreed freely and reproduce both sexually and asexually (apomictic progeny). A majority of taxonomists consider there to be six genera within the 'true citrus trees' (*Citrinae* subtribal group C), being *Citrus*, *Poncirus*, *Fortunella*, *Eremocitrus*, *Microcitrus* and *Clymenia* (Vardi and Spiegel-Roy 1978). A recent comprehensive morphological study across 43 biotypes of citrus proposed that there is only 3 true biological species (Barrett and Rhodes 1976). These are the pummelo (*C. grandis*), citron (*C. medica*) and mandarin (*C. Reticulata*).

Microsatellite STS length variation was compared and used for phylogenetic reconstruction at nine loci among eight *Citrus* species and five closely related genera. Two different programs were investigated for calculating genetic distance. The algorithm of (Nei 1972) which is formulated for an isoallele mutation system and the delta mu (Ddm) model of (Goldstein et al. 1995) which assumes the step wise mutation model for allele generation. The Neighbor-joining method (Saitou and Nei 1987) was used to produce a phenogram which grouped eight *Citrus* species into two distinct branches (Fig. 1). The membership of these groups correspond closely with two citrus subgroups proposed after analysis of 146 morphological characters (Barrett and Rhodes 1976). This indicates microsatellite allele variation has predictive phenetic value within the *Citrus* genus. When the *Citrus* species were reanalysed with microsatellite genotypes from five related genera, the *Citrus* associations were preserved with the exception of pummelo and citron. These, along with the mandarin, are proposed to represent true ancestral species. Members from each genus grouped together while genera split into separate branches, indicating microsatellite variation does contain taxonomic information at this evolutionary level. However, bootstrap analysis indicated confidence in phylogenetic prediction was not uniform with time, as higher values were observed for intra-*Citrus* associations than for genus level groupings. This suggests that plant microsatellite length variation may be more suitable for phylogenetic reconstruction over short evolutionary timescales.

Linkage analysis

The purpose of genetic maps is to identify markers closely linked to important traits such as those that effect disease resistance, plant performance and fruit quality. Of special importance for fruit crops such as citrus and grapevines which have very long generation times is the potential to use genetic maps to assist breeders by marker assisted breeding. To investigate the ability of microsatellite markers to map in very wide fruit crop crosses citrus microsatellite markers were tested and integrated into a RFLP map derived from an intergeneric cross made between *Citrus* and the related genera *Poncirus* (Kijas et al. 1997). The ability to map plant microsatellites in very wide crosses has not been fully investigated and is dependant on sufficient sequence conservation being present within priming sites to amplify homologous loci across the genera of interest. The extent to which microsatellite primers are able to amplify homologous loci between plants seems to vary. In grapevine primer sequence conservation exists across a range of *Vitis* species (Thomas and Scott 1993) similar to rice (Wu and Tanksley 1993), however, (Roder et al. 1995) in a study of wheat microsatellites found very low rates of successful amplification when rye and barley were tested.

Analysis of progeny from the *Citrus* x *Poncirus* cross resulted in seven out of fourteen trinucleotide microsatellite STS markers being added to a citrus genetic map. Their location, in every instance, was found to be on a different linkage group indicating a wide spread throughout the genome (Kijas et al. 1997). Two markers remained unlinked while null alleles, similar allele sizes or multiple locus amplification prevented linkage analysis of the remaining 5 markers. Analysis of allele segregation revealed two of the nine microsatellite STSs showed distorted segregation. Examination of the direction of segregation distortion (ie between *Citrus* and *Poncirus* parent alleles) showed in both cases the *Citrus* allele to be under represented (Kijas et al. 1997). A similar situation was reported for RFLP markers where the *Citrus* allele was under represented in eight of nine skewed RFLPs (Jarrell et al. 1992). This indicates the unequal allele segregation toward the *Poncirus* parent is independent of marker type and also independent of the type of polymorphism being assayed. It also suggests that the dominance of *Poncirus* should be considered when designing breeding crosses aimed at introgressing specific genes from this breeding parent. Plant breeders should also recognise that this phenomenon may be a general occurrence in wide plant crosses.

The codominance mode of inheritance and amplification within a wide breeding cross indicates microsatellite STS markers can be typed and mapped in different breeding crosses. This suggests that they are ideal markers to act as anchor loci to align RFLP, AFLP and RAPD based linkage maps generated in different mapping crosses from different laboratories.

ACKNOWLEDGEMENTS

This research was supported in part by grants from the Grape and Wine Research and Development Council, the Dried Fruits Research and Development Council, and the Department of Industry Science and Technology.

REFERENCES

Barrett HC, Rhodes AM (1976) A numerical taxonomic study of affinity relationships in cultivated *Citrus* and its close relatives. Systematic Botany **1**:105-136

Botta R, Scott NS, Eynard I, Thomas MR (1995) Evaluation of microsatellite sequence-tagged site markers for characterising *Vitis vinifera* cultivars. Vitis (in press):

Brady JL, Scott NS, Thomas MR (1996) DNA typing of hops (*Humulus lupulus*0 through application of RAPD and microsatellite marker sequences converted to sequence tagged sites (STS). Euphytica 91:277-284

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics **139**:463-471

Jarrell DC, Roose ML, Traugh SN, Kupper RS (1992) A genetic map of citrus based on the segregation of isozymes and RFLPs in an intergeneric cross. Theor Appl Genet **84**:49-56

Kandpal RP, Kandpal G, Weissman SM (1994) Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. Proc Natl Acad Sci USA **91**:88-92

Karagyozov L, Kalcheva ID, Chapman VM (1993) Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. Nucleic Acids Research **21**:3911-3912

Kijas JMH, Fowler JCS, Garbett CA, Thomas MR (1994) Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin coated magnetic particles. Biotechniques 16:656-662

Kijas JMH, Fowler JCS, Thomas MR (1995) An evaluation of sequence tagged microsatellite site markers for genetic analysis within *citrus* and related species. Genome 38:349-355

Kijas JMH, Thomas MR, Fowler JCS, Roose ML (1997) Integration of trinucleotide microsatellites into a linkage map of *Citrus*. Theor Appl Genet 94:701-706

Kiyama R, Nishikawa N, Oishi M (1994) Enrichment of human DNAs that flank poly(dA).poly(dT) tracts by triplex DNA formation. J Mol Biol **237**:193-200

Nei M (1972) Genetic distances between populations. Amer Nat 106:283-292

Roder MS, Plaschke J, Konig SU, Borner A, Sorrells ME, Tanksley SD, Ganal MW (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 246:327-333

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Bio Evol 4:406-425

Tautz D, Trick M, Dover G (1986) Cryptic simplicity in DNA is a major source of genetic variation. Nature 322:652-6

Thomas MR, Cain P, Scott NS (1994) DNA typing of grapevines: A universal methodology and database for describing cultivars and evaluating genetic relatedness. Plant Mol. Biol. 25:939-949

Thomas MR, Matsumoto S, Cain P, Scott NS (1993) Repetitive DNA of grapevine: classes present and sequences suitable for cultivar identification. Theor. Appl. Genet. 86:173-180

Thomas MR, Scott NS (1993) Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequenced-tagged sites (STSs). Theor. Appl. Genet. 86:985-990

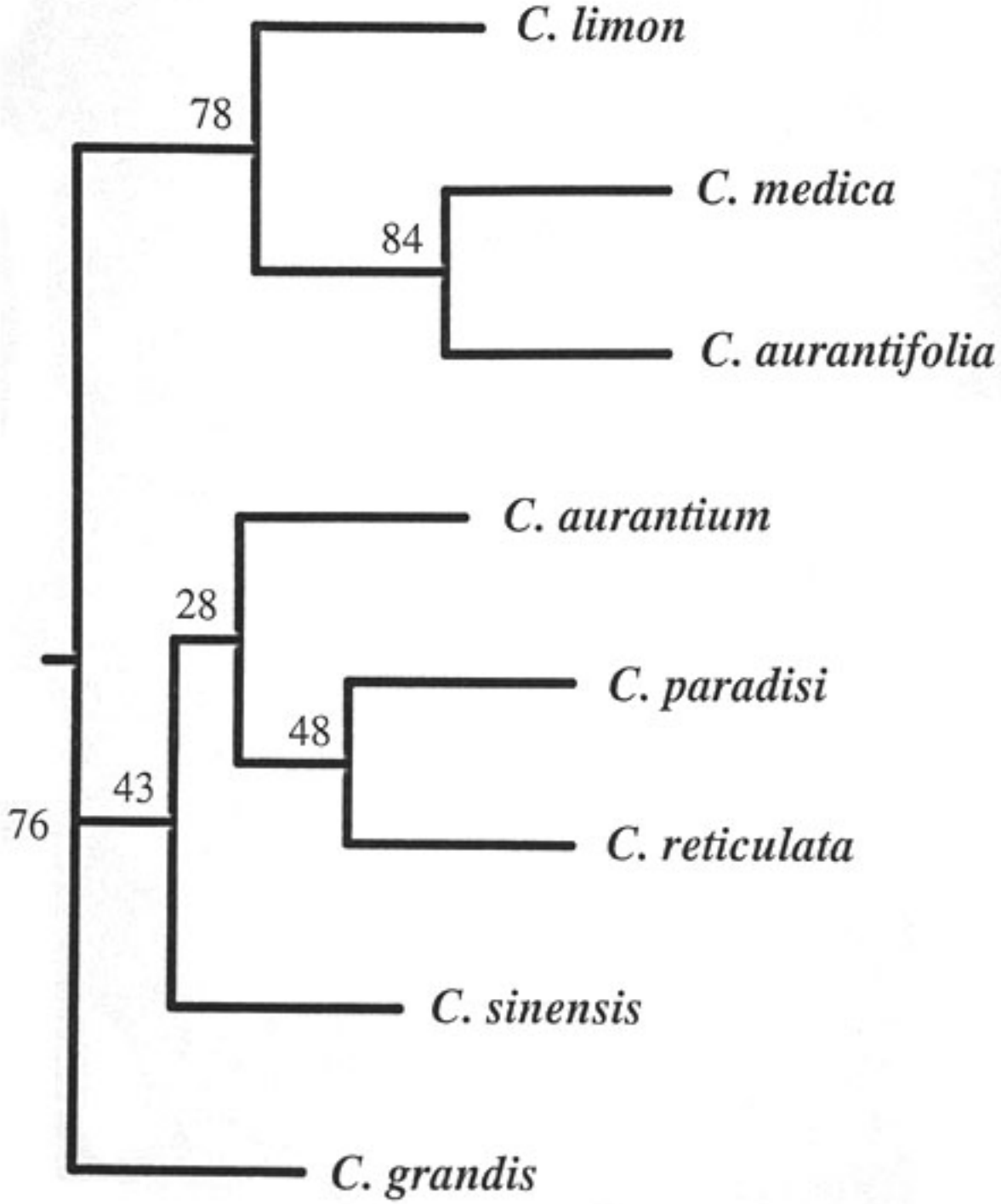
Truel P, Rennes C, Domergue P (1980) Identifications in collections of grapevines. Proc. 3rd Int. Symp. on Grape Breeding. Dept. of Vitic. and Enol. Univ. Calif., Davis, California., USA, pp 78-86

Vardi A, Spiegel-Roy P (1978) Citrus breeding, taxonomy and species problems. Proc Int Soc Citriculture 184:51-57

Viala P, Vermorel V (1909) Ampelographie. Masson, Paris

Wu K, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241:225-235

FIG. 1 - Neighbor-joining tree relating eight *Citrus* species based on variation at nine microsatellite loci. The outgroup option of Neighbor was used with *C. grandis* selected, causing the program to root the tree at the midpoint between the most distantly related species pair. Numbers at the nodes indicate the percentage of times the group consisting of species to the right of the fork occurred from 250 bootstrap replications.


C. limon
78
C. medica
84
C. aurantifolia
C. aurantium
28
C. paradisi
48
43
76
C. reticulata
C. sinensis
C. grandis