Proceedings of the California Avocado Research Symposium, November 4, 2006. University of California, Riverside. Sponsored by the California Avocado Commission. Pages 57-63.

Assessing the Genetic Determination of Valuable Avocado Traits Using Microsatellite (SSR) Markers and Quantitative Trait Locus (QTL) Analysis

Ongoing Project: Year 5 of 8

Project Leader: Michael T. Clegg (949) 824-4490, email: mclegg@uci.edu
Department of Ecology and Evolutionary Biology, UC Irvine, and
Department of Botany and Plant Sciences, UC Riverside

Co-operating Researchers:
Vanessa Ashworth (UCI), Haofeng Chen (UCR), Shizhong Xu (UCR)
vashwort@uci.edu, haofengchen@yahoo.com

Objectives

Varietal improvement in avocado (*Persea americana* Mill.) has long relied on multi-year field trials during which large numbers of seedlings are grown to maturity and compared for desirable characteristics. Inferior trees are removed from the breeding block as their deficiencies become apparent, leaving only the most promising genotypes. However, the time, land resource, and labor costs associated with growing trees to the appropriate stage of development are considerable. The pace of varietal improvement would accelerate substantially through the application of molecular markers that are detectable using DNA extracted from seedlings. If transmitted along with desirable traits, the markers can be used as surrogates for these traits and can be applied quickly to a large number of seedlings to enrich the initial pool of trees for traits that are of interest to the Industry. Our research is designed to identify markers that are co-transmitted with genetic factors conferring desirable characteristics in avocado. Our objectives are (1) to link avocado traits of interest to growers with molecular markers and (2) to harness this information via marker-assisted selection. This marker-guided method of variety improvement has the potential to increase selection intensity by several orders of magnitude.

Benefits to the Industry

- By gaining access to a less stochastic, faster, and less land- and labor-intensive method of breeding avocado, called marker-assisted selection (MAS), and to allied breeding strategies

- Our molecular marker data is a permanent resource: it can be combined with any trait of interest measured (non-destructively) in our population of experimental trees to identify marker-trait associations. This will furnish more detailed knowledge of the genetics of both complex traits (QTLs), as well as mono- and oligogenic traits

- Once the molecular framework is in place, other molecular studies can be more easily piggy-backed, thus magnifying the returns on initial investment

Methods

Unraveling the association between desirable traits and molecular markers relies on (1) the availability of a pool of molecular markers, and (2) a replicated experimental population of trees having a known genetic constitution. A summary of marker- and tree-related information is presented in Table 1.

Table 1. Summary of facts and figures relating to our markers and experimental trees.

1.	205 distinct genotypes of open-pollinated 'Gwen' progeny
2.	Four clones of each genotype at SCREC (Irvine) and Agricultural Operations (AgOps, Riverside); two clones planted at each location
3.	Grafted onto Duke 7 rootstock
4.	Trees planted out in 2001 (SCREC) and 2002 (AgOps)
5.	398 and 285 trees (= 683 trees) at each site, respectively
6.	127 microsatellite markers
7.	364 and 161 trees (525 trees) bore fruit this year
8.	34.7% of genotypes were sired by 'Bacon', 39.8% by 'Fuerte' and 25.5% by 'Zutano'
9.	Fruit dry weights (March–April): 15.6–43.8%, averaging 29.3 ± 5.0% DW at SCREC and 33.5 ± 5.4% at AgOps
10.	Fruit weights: <100 g to 799 g, with an average of 281 g (SCREC) and 255 g (AgOps) (preliminary data)
11.	Fruit load/tree: high (>100 fruit/tree) in 27.3% of trees, medium (50–99 fruit/tree) in 38.3%, low (1–49 fruit/tree) in 26%, and 8.5% of trees bore no fruit.

Key to this project is the maintenance of four clones of each experimental tree genotype growing at two different locations. Replicated field trials are essential because the environment exerts nongenetic effects on all traits. Classical quantitative genetics is used to partition the total variation between replicates into a genetic and a nongenetic component, whereby only the genetic component is heritable and hence of interest to the breeder. The juxtaposition of molecular and measurement data—against a controlled genetic background—is the basis of quantitative trait locus (QTL) analysis, a statistical procedure that detects trait-marker associations. Molecular markers associated with desirable avocado traits are employed in marker-assisted selection. Below, we describe two main areas of research that we have been pursuing over the past 12 months:

(1) *Statistical Analyses*: We completed the quantitative genetic analyses for our multi-year data on growth rate, flowering, and fruit load. Breeding advance can only be achieved if the traits of interest have an underlying heritable (genetic) component. Using Analysis of Variance (ANOVA), a statistical technique that splits total variation into genetic and nongenetic components, we determined heritability for three measures of growth rate (tree height, canopy diameter, and stem girth; data collected 2001–2004), flowering (2003–2005), and fruit load (2005 only; based on a visual determination of the number of fruit on a tree). Significance for deviation from the null hypothesis was assessed using *F*-ratios (from the ANOVA table of Proc GLM type III sum of squares) implemented in SAS, version 9.1.

(2) *Marker Data and Fruit Evaluation:* We also devoted considerable time to gathering additional marker data and linking it to traits measured on our experimental population of trees. The large fruit harvest this year has provided us with a generous amount of information on fruit-related traits. Preliminary information on the relationship between markers and fruit traits is presented; these relationships have not yet been statistically evaluated but serve to illustrate the principle of marker-assisted selection.

Results

Statistical Analyses

Our statistical analyses concentrated on the assessment of heritability, genotype x environment interactions, and trait correlations. These statistical measures shed light on the breeding potential that can be expected when selecting for a given trait. Having used our molecular markers to determine paternity for each tree genotype, we can also partition the total variation in a trait by pollen donor.

Broad-Sense Heritability: Of all the factors that influence growth rate—as expressed in terms of tree height, canopy diameter and stem girth—the heritable component accounts for ca. 30% (Table 2). The higher the heritability value, the faster the trait will respond to breeding. A value of 30% is respectable, but nonetheless suggests that environmental factors have significant impact on tree growth. Growth rate averaged 5–6 cm/month, attaining 14 cm/month for 'Gwen' progeny genotype 100. Although life history studies generally predict exponential growth in the early stages of plant development, the growth rates of our trees are linear over the three-year time interval examined.

Genotype x Environment Interaction: No genotype x location effect was noted for growth rate, indicating that none of the genotypes shows a marked preference for one site over the other. Flowering and fruit load showed a relatively weak effect (23.9 and 17.6%, respectively; Table 2). This means that different genotypes show differential flowering and fruit loads/tree depending on which location they are growing at (Irvine versus Riverside).

Table 2. Broad-sense heritability and genotype x environment interactions for three measures of growth rate (tree height, canopy diameter, stem girth), flower abundance, and fruit load per tree. These values are based on over 90,000 data points.

	Tree height	Canopy diameter	Stem girth	Flower abundance	Fruit load
Broad-sense heritability (%)	34.4	29.7	28.5	32.3	23.4
G x E interaction (%)	NS	NS	NS	23.9	17.6

Trait Correlations: Surprisingly, none of the growth rate measures was correlated with flowering abundance, and only a moderate correlation was found between growth rate and fruit

load. In practical terms, this means that selection for high fruit yields is not genetically tied to faster growth. In other words, breeding can focus on combining high fruit yields and short stature, rather than having to put up with large trees when selecting for high yields. This is a valuable property, given the trend toward breeding smaller avocado trees. Flowering abundance was not correlated with fruit load. One might expect higher fruit yields in response to abundant flowering, but the fact that only one in a thousand fruit attains maturity probably accounts for the lack of a correlation.

Pollen Donor Effect: Our molecular markers allow determination of the pollen parent of each experimental tree, enabling growth rates and flowering data to be linked to paternal origin. Specifically, we can ask whether the type of pollen donor has a measurable effect on selected growth parameters. Table 3 shows that 'Gwen' progeny sired by 'Fuerte' was significantly shorter than progeny sired by 'Bacon', 'Zutano', or mixed/unknown sources, and produced fewer flowers. 'Zutano'-sired progeny had a significantly higher fruit load than the other genotypes, combined with significantly smaller canopy diameter and stem girth. Moreover, progeny in the mixed category was significantly taller than 'Fuerte'- and 'Zutano'-sired trees and taller (nonsignificantly) than 'Bacon'-sired progeny. Similar analyses will be performed on this year's fruit evaluation data. Paternity-specific fruit attributes may be identified that would not be readily detected otherwise.

Table 3. Mean effects of pollen donor on growth rate (tree height, canopy diameter, stem girth; all in centimeters per month), flower abundance, and fruit load per tree.

	Tree height	Canopy diameter	Stem girth	Flower abundance	Fruit load
'Bacon'	5.931(a,b)	6.045(b)	0.226(a)	1.965(a)	1.410(b)
'Fuerte'	5.002(c)	6.482(a)	0.213(a)	1.418(c)	1.385(b)
'Zutano'	5.774(b)	5.241(c)	0.197(b)	1.846(a)	1.614(a)
Mixed	6.289(a)	6.484(a)	0.223(a)	1.604(b)	1.446(b)

Marker Data and Fruit Evaluation

Marker Data

Six markers were added to our data set between September 2005 and February 2006: AVT386, AVD003II, AVD006, AVD022, AVD010, AVD028. Since June 2006, we have added AVD037II, AVD026, and AVD036, taking the total to 16 markers.
By the end of 2005, our marker analyses had revealed that all but one of our 200 tree genotypes (whose maternal parent is 'Gwen') had been outcrossed (i.e., the pollen came from a different variety) and that 98 of the 200 genotypes had 'Bacon', 'Fuerte', or 'Zutano' as their male parent. For the remainder, the male parent did not match up with any of the varieties forming part of our molecular reference archive.

Fruit Weight and Shape

Here we present partial data sets on fruit weight and fruit shape in order to explore trends. Fruit weights are based on only 1–2 fruits/tree and about 50–60% of trees in respective locations, but

there is no reason to believe that the findings reported below would change substantially once complete data have been gathered.

For trees at SCREC and AgOps, fruit weight averaged 280 and 226 g, respectively. Table 4 presents a summary of fruit weights partitioned by pollen donor ('Bacon', 'Fuerte', and 'Zutano'), suggesting that 'Bacon' pollen sires somewhat heavier fruit and 'Zutano' pollen somewhat lighter fruit. At SCREC, 'Fuerte'-pollinated progeny produced smaller fruit than 'Bacon', but the reverse was true at AgOps, a possible consequence of the smaller sample size at the latter location. The larger 'Bacon'-sired fruits at SCREC could also signal better adaptation of 'Bacon' progeny to the coastal Orange County conditions than to the Riverside climate, or may be related to the larger size of 'Bacon' trees reported above. A comparison of the average fruit weights for these three pollen sources versus the overall orchard averages, suggests that at SCREC the genotypes in the "other"-category are likely to be larger-fruited than the fruit sired by 'Bacon', 'Fuerte' or 'Zutano', whereas the reverse is true at AgOps.

Table 4. Average fruit weights [grams], with sample numbers in parentheses.

	SCREC	AgOps	Both locations
'Bacon'	280.6 (48)	257.8 (41)	269.2 (89)
'Fuerte'	268.3 (46)	264.4 (27)	266.4 (73)
'Zutano'	266.8 (38)	248.2 (30)	257.5 (68)
Overall	271.9 (132)	256.8 (98)	264.4 (230)

Fruit shape was scored using the IPGRI descriptors and partitioned into four main shape categories representing spheroid or somewhat spheroid ("2"), obovate ("6"), narrowly obovate ("5"), and somewhat clavate (= elongated; "8"). Table 5 illustrates the relationship between pollen source and fruit shape score. A majority of genotypes are in the intermediate shape categories "5" and "6" that characterize 'Gwen' fruits and would be expected in 'Gwen' progeny trees. The more extreme round or elongate shapes arise much less frequently. Upon closer inspection, these data reveal a slight tendency for 'Bacon'-sired progeny to produce rounded fruit (mostly shape categories "2" and "6") and 'Zutano'-sired progeny to produce elongate fruit (mostly categories "5" and "8").

Table 5. Fruit shape, scored using IPGRI descriptors. Values are counts for AgOps and SCREC combined.

	"2"	**"6"**	**"5"**	**"8"**	**Totals**
'Bacon'	31	27	21	10	89
'Fuerte'	26	11	20	15	72
'Zutano'	10	9	30	18	67
Total	67	47	71	43	228

The Role of Alleles

The data presented so far are interpreted in terms of paternal influence acting in a 'Gwen' maternal genetic background. We can focus on one more level and examine relationships

between traits and particular alleles. Because a given tree genotype has two alleles at each genetic locus—one allele from each parent—and because each parent in turn has two alleles available to pass down to its offspring (e.g., A or B from the mother and C and D from the father), the same parental combination can result in four different progeny genotypes (AC, AD, BC, or BD). Table 6 illustrates a concrete example of fruit weight partitioned by alleles present at microsatellite marker locus AUCR418.

Table 6. Relationship between fruit weight and allelic composition for microsatellite locus AUCR418. Alleles "e" and "h" are present in 'Gwen' and 'Fuerte', alleles "h" and "c" in 'Zutano', and alleles "d" and "g" in 'Bacon'. All fruit weights [grams] are averaged for all trees (n = 111) possessing the allele in question.

	'Gwen' alleles:		
	Allele "e"	Allele "h"	Average
Paternal alleles:			
Allele "e"	334.42	271.53	302.98
Allele "h"	271.53	288.16	279.85
Allele "c"	249.69	240.00	244.85
Allele "d"	249.85	294.53	272.19
Allele "g"	274.20	291.91	283.06
Average	275.94	277.23	276.58

The marker locus itself is not (usually) synonymous with the gene affecting fruit weight. However, its proximity on the chromosome to a gene that *does* control fruit weight determines whether the marker can detect any signal. In the example illustrated in Table 6, genotypes having the allele pair e/e (two identical copies of allele "e") produce the heaviest fruit (334.42 g), whereas genotypes with allele pair c/h have the smallest fruit (240.00 g). Alleles "d" and "g" (from 'Bacon') produce larger fruit when combined with 'Gwen' allele "h" than in combination with 'Gwen' allele "e". If this trend were to prove statistically significant, then marker-assisted selection for large fruit size would recommend retention of seedling genotypes having the "e/e" genetic constitution but removal of seedlings with the "c/h" constitution.

This example serves to demonstrate the tremendous power of genetic markers in peeling away the oft-confusing veneer of phenotype to reveal the genetic underpinnings of any given trait for which sufficient data have been collected. The more genetic markers we can add to our database, the greater the likelihood of detecting a signal from genes affecting a trait of choice. The more numerous and diverse the traits measured on the trees, the more can be learned about the genetic structure of the avocado genome. Both our microsatellite markers and our trees represent a substantial resource that should be tapped. The permanence of the marker data means that, once completed, they will be available for referencing all future traits measured.

Conclusions and Timeline

The labor-intensive nature of this project is clear. Table 7 summarizes the number of data points that factored into the non-molecular matrix alone. In the coming year, we hope to boost the number of markers in our molecular data set.

Table 7. Updated table on labor requirements for collection of data on growth rates, flowering, and fruit evaluations. Readings = number of measurements taken as part of the fruit evaluations. Noninvasive evaluations include fruit shape, weight, width, length, horizontal circumference, vertical circumference, and skin texture. Invasive measurements include ripe fruit weight, 4 seed attributes, 4 skin attributes, and 4 flesh attributes.

Evaluation type	# Years or year	Readings per trait	# Fruits per tree	Tree #	Data points
Tree height	5	1	n/a	700	3,500
Canopy diameter	4	2	n/a	700	7,000
Trunk diameter	4	2	n/a	700	5,600
Flowering	3	6	n/a	700	75,600
				Subtotal:	**91,700**
Fruit[a]	2005	7	1–15	62	6,510
Fruit[b]	2005	13	1–3	62	2,418
				Subtotal:	**8,928**
Fruit[a]	2006	7	2–8	525	29,400
Fruit[b]	2006	13	1–3	525	20,475
				Subtotal:	**49,875**
				TOTAL:	**150,503**

Hours/evaluation: [a]Noninvasive: ca. 50 fruits in 3 hours = 16.7 fruits/h [3.6 min/fruit]
[b]Invasive: ca. 30 fruits in 3 hours = 10 fruits/h [6 min/fruit]

Accordingly, we have drawn up a tentative timeline for data collection and subsequent analyses. While tree growth and fruit data are adequate for preliminary studies (heritability and genotype x environment interaction effects), this is not true for the molecular data: a QTL analysis is not normally initiated until allele data are available for a threshold number of markers. This threshold is a function of the distribution of QTLs across the chromosomes and of an organism's chromosome number: the greater the number of chromosomes (avocado = 12), the greater the number of markers needed. The processing of 40 genetic marker loci over the next 12 months will enable a first pass-QTL analysis, plus additional time required for data formatting and manipulation. This translates into a time frame of ca. December 2007. The greater the number of markers processed, the greater the likelihood of detecting a signal. Our goal is eventually to run all 127 markers on all experimental trees.

Marker-assisted selection can start as soon as QTLs have been found. QTLs are visualized in terms of the associated alleles ("bands on a gel") and are ranked by efficacy, which assists the plant breeder's decision which QTLs to select on. Ranking depends on several factors, including strength of the marker-trait association, correlation between markers, and overall allelic composition of the breeding material. As a long-term strategy, we recommend that marker-assisted selection be applied to our Gwen progeny trees for two successive generations.