Ancestral Seed Zones and Genetic Mixture of Tanoak[1]

Richard Dodd,[2] Zara Rafii,[2] and Wasima Mayer[2]

Abstract

Understanding the genetic structure of tanoak (*Lithocarpus densiflorus*) is necessary to pathologists seeking natural variation in resistance to *Phytophthora ramorum*, cause of sudden oak death (SOD), and to resource managers who need indications of conservation priorities for this species now threatened by this introduced pathogen. We investigated population genetic structure using nuclear and chloroplast DNA in 43 populations from throughout the range of the species. Our chloroplast DNA results revealed four major and two rare haplotypes. The 4 major haplotypes delineated ancestral seed pools from central and northern coastal California, extreme northern California and Oregon, Klamath Mountains and the Sierra Nevada. Diversity at nuclear microsatellite loci supported the chloroplast lineages and indicated some further divergence within them. We propose that at least six breeding zones should be recognized for disease resistance screening and for conservation management. These include 1. southern-most populations from Nacimiento, Lompoc, and Santa Barbara that are relatively low in genetic diversity; 2. central coastal populations from Big Sur to the San Francisco peninsula; 3. populations from north of the San Francisco Bay to Arcata; 4. extreme northern California and Oregon populations from north of Arcata to the northern limit of the species' range; 5. Klamath Mountains; and 6. the Sierra Nevada.

Introduction

The sudden oak death (SOD) epidemic, caused by *Phytophthora ramorum*, has caused considerable damage to natural woodlands of central and northern California. Keystone species such as tanoak (*Lithocarpus densiflorus* [Hook. & Arn.] Rehder, Fagaceae), recently attributed to a new genus (*Notholithocarpus densiflorus* [Hook. & Arn.] Manos, Cannon & S. Oh), and coast live oak (*Quercus agrifolia* Née) have suffered the highest mortality. Because of the dominance of these species in the coastal mixed evergreen forests, their loss will have significant impacts on the composition, function and management of these coastal ecosystems. Although the disease is believed to be absent from north-western California, the extensive woodlands of tanoak there and in southern Oregon are likely to be highly sensitive to the disease because of suitable climatic conditions for the pathogen and the dominance of tanoak.

Studies of host resistance to SOD indicate quantitative variation in susceptibility (Dodd and others 2005, 2008; Hayden and others, unpublished data), so it seems likely that variations in host tolerance will play a crucial role in the future of the coastal ecosystems. The observed variation in susceptibility indicates that genetic variation is present in natural populations and so resistance screening is a worthwhile

endeavour. Initial work on population genetic variation in tanoak (Nettel and others in press) has guided the choice of populations in an effort, currently underway, to detect resistance within a limited number of populations. An understanding of genetic structure of susceptible host species is crucial in implementing resistance screening. Where populations are at risk of extinction, as is the case for tanoak, discovery of divergent population lineages will also provide a rationale for conservation priorities.

Here we report on a detailed analysis of genetic variation in nuclear and chloroplast DNA covering almost the entire range of tanoak. Our choice of genetic markers was predicated on the need to obtain the maximum amount of information on levels of genetic diversity and to identify seed lineages that might be important in demarcating potential seed zones. In tanoak, chloroplast DNA is inherited clonally through a maternal lineage. Genetic variation in chloroplast DNA is typically low, but it provides an excellent system for detecting the limits of seed dispersal. On the other hand, nuclear microsatellites are highly polymorphic and provide an excellent means of assessing variations in levels of genetic diversity.

Materials and Methods

Sampling and DNA Analysis

Leaves from mature individuals of *L. densiflorus* ssp. *densiflorus* and one population of dwarf tanoak *L. densiflorus* ssp. *echinoides* were obtained from coastal populations from Coos County, Oregon to Santa Barbara, California, and from the southern Klamath Mountains and the Sierra Nevada (table 1). We sampled a total of 905 trees from 43 populations (fig. 1), with an average of 21 trees per population. Leaves were maintained frozen at –20 ºC. We extracted DNA using a simplified CTAB method (Cullings 1992).

Chloroplast DNA

After testing more than 40 putative loci, we detected five polymorphic cpDNA mononucleotide microsatellites in tanoak (table 2): *ccmp4* (*atpF* intron, Weising and Gardner, 1991), *μcd2_p14* (*trnC-ycf6*, Deguilloux and others 2003), *Cmcs5* (*ndhG-ndhI,* Sebastiani and others 2004), *KI* (*trnK*), and *BII* (*rpoB*) (Nettel and others 2009). These five loci were amplified for a total of 505 individuals from the 43 populations sampled. Because of haplotype monomorphism in populations, we amplified fewer individuals than were used in nuclear microsatellite analyses. PCR conditions of the cpDNA microsatellite loci, as well as sequencing procedures, are described elsewhere (Nettel and Dodd 2007). Amplifications were electrophoresed on an ABI 3730 automated sequencer (Applied Biosystems, Foster City, CA).

Nuclear DNA

Eight nuclear microsatellite loci developed for tanoak (LD1, LD3, LD5, LD7, LD8, LD10, LD12, LD14, Morris and Dodd 2006) were PCR amplified using a fluorescent labeled primer. Primers were multiplexed in five different groups: i) LD1, LD3, LD7, LD10, LD14; ii) LD5; iii) LD12; and iv) LD8. The PCR cocktail contained 1X PCR Buffer, 2.0mM MgCl2, 0.2 mM of each dNTP, 250nM of each reverse primer, 250nM of each fluorescently labeled (FAM or HEX) forward primer, 1 unit of Amplitaq Polymerase (Invitrogen, Carlsbad, CA) and approximately 5 ng template DNA in a 20 μl reaction. Forward and reverse primers for the locus LD3

were run at a final concentration of 400nM, exclusively. Touchdown PCR cycling conditions for all loci were as follows: one cycle at 95 ˚C for 10 minutes followed by 20 cycles of 45 seconds at 94 ˚C, 45 seconds at 58 ˚C (lowering 0.5 ˚C each cycle), and 45 seconds at 72 ˚C. The final amplification step consisted of 20 cycles of 45 seconds at 94 ˚C, 45 seconds at 48 ˚C, and 45 seconds at 72 ˚C and was followed by a final extension step at 72 °C for 45 minutes. All reactions were performed on a Techne (UK) Flexigene thermocycler. We mixed 0.75 µl of PCR product with 8 µl of formamide and 0.5 µl of 500 LIZ size standard (Applied Biosystems, Foster City, CA) and we electrophoresed this cocktail on an ABI 3730 automated sequencer (Applied Biosystems, Foster City, CA). We used GENESCAN 3.7 and GENOTYPER 3.7 (Applied Biosystems, Foster City, CA) to analyze ABI results. We sequenced Loci LD3 and LD5 because they showed odd-sized series in fragment analyses. The sequences showed that indels within the flanking regions were responsible for the different size series and so the fragment sizes were adjusted appropriately.

Table 1—Location of sampling sites

Nearest location	County/State	Lat.	Long.	Sample #	Subspp.
Santa Barbara	Santa Barbara/Ca	34.50	-119.82	20	*densiflorus*
Lompoc	Santa Barbara/Ca	34.58	-120.50	24	*densiflorus*
Nacimiento	Monterey/Ca	35.81	-120.75	41	*densiflorus*
Pfeiffer Big Sur	Monterey/Ca	36.25	-121.78	17	*densiflorus*
Palo Colorado	Monterey/Ca	36.39	-121.90	32	*densiflorus*
Soquel	Santa Cruz, Ca	37.05	-121.85	57	*densiflorus*
S. F. Peninsula	San Mateo/Ca	37.40	-122.28	31	*densiflorus*
Point Reyes	Marin/Ca	38.06	-122.80	33	*densiflorus*
Forestville	Sonoma/Ca	38.47	-122.90	10	*densiflorus*
Cazadero	Sonoma/Ca	38.49	-123.06	10	*densiflorus*
Salt Point	Sonoma/Ca	38.56	-123.32	10	*densiflorus*
Gualala	Sonoma/Ca	39.35	-123.52	10	*densiflorus*
Fort Bragg	Mendocino/Ca	39.43	-123.80	7	*densiflorus*
Jackson SF	Mendocino/Ca	39.35	-123.61	24	*densiflorus*
Sinkyone	Mendocino/Ca	39.78	-12.83	20	*densiflorus*
Hickey	Mendocino/Ca	39.88	-123.73	20	*densiflorus*
King Range	Humboldt/Ca	40.16	-124.08	15	*densiflorus*
Weott	Humboldt/Ca	40.34	-123.95	9	*densiflorus*
Hacketsville	Humboldt/Ca	40.50	-124.16	18	*densiflorus*
Grizzly Creek	Humboldt/Ca	40.84	-123.91	20	*densiflorus*
Redwood Creek	Humboldt/Ca	40.85	-123.81	21	*densiflorus*
Korbel	Humboldt/Ca	40.85	-123.92	20	*densiflorus*
Fickle Hill	Humboldt/Ca	40.85	-123.82	20	*densiflorus*
PrairieCrk	Del Norte/Ca	41.45	-124.02	20	*densiflorus*
Jedediah	Del Norte/Ca	41.81	-124.08	20	*densiflorus*
Gasquet	Del Norte/Ca	41.88	-123.85	20	*densiflorus*
Gasquet	Del Norte/Ca	41.89	-123.99	13	*echinoides*
Elk Valley	Del Norte/Ca	41.99	-123.72	17	*densiflorus*
O'brien	Josephine/Or	42.07	-123.72	25	*densiflorus*
Brookings	Curry/Or	42.09	-124.31	21	*densiflorus*
Williams	Josephine/Or	42.26	-123.38	16	*densiflorus*
Sebastien	Curry/Or	42.32	-124.40	20	*densiflorus*
Merlin	Douglas/Or	42.52	-123.62	20	*densiflorus*
Agness	Curry/Or	42.53	-124.22	20	*densiflorus*
Glendale	Douglas/Or	42.73	-123.53	20	*densiflorus*
Marial	Curry/Or	42.81	-124.21	21	*densiflorus*
Powers	Coos/Or	42.93	-124.08	21	*densiflorus*
Dora	Coos/Or	43.17	-124.02	21	*densiflorus*
Blodgett	El Dorado/Ca	38.73	-120.75	54	*densiflorus*
Plumas	Butte/Ca	39.93	-121.58	10	*densiflorus*
Weaverville	Trinity/Ca	40.69	-122.93	17	*densiflorus*

Data Analysis

We checked data quality for scoring errors in nuclear and cpDNA microsatellites using MICRO-CHECKER (Oosterhout and others 2004). We used the same program to estimate the probability of the presence of null alleles in nuclear loci.

Chloroplast microsatellite alleles were combined into haplotypes, where each amplified locus was considered an independent, but linked, character inherited as a unity.

To ascertain reliability of the estimations based on nuclear microsatellites, we tested

deviations from Hardy-Weinberg (HW) equilibrium within each population by the inbreeding fixation index, F_{IS}, with the software FSTAT ver. 2.9.3.2 (Goudet 2002). HW equilibrium could not be rejected for any of the populations. We estimated allelic richness with the rarefaction method (R_t) and expected heterozygosity (H_e) with FSTAT in order to analyze the level of within population genetic diversity. We used BAPS 5.2 (Corander and others 2008) to infer population structure from the nuclear microsatellite diversity. We inferred population division (number of K populations) by performing 20 independent runs of each K ($K= 1$ to $K= 40$).

Results

Four major, and two rare chloroplast DNA haplotypes were detected (fig. 1). The four major haplotypes included: 1. a central and northern coastal California haplotype (B) with the northern limit near Korbel, Humboldt County, California; 2. an extreme northern California and Oregon haplotype (D) with a southern limit at Hoopa, Humboldt County, California; 3. a Sierra Nevada haplotype (E) found in populations from El Dorado and Butte Counties; and 4. a Klamath haplotype (F) from east of Weaverville in Trinity County. The rare haplotypes (A and C) occurred as four individuals (three from Palo Colorado and one from Big Sur) and a single individual (from Cazadero) respectively. With the exception of populations with the two rare haplotypes, we detected a mixture of haplotypes in only one population - haplotypes B and D in approximately equal proportions in a population between Korbel and Hoopa, Humboldt County. Interestingly, the northernmost population from Dora, Coos County, Oregon had the B haplotype characteristic of central and northern coastal California populations, suggesting a possible non-indigenous source.
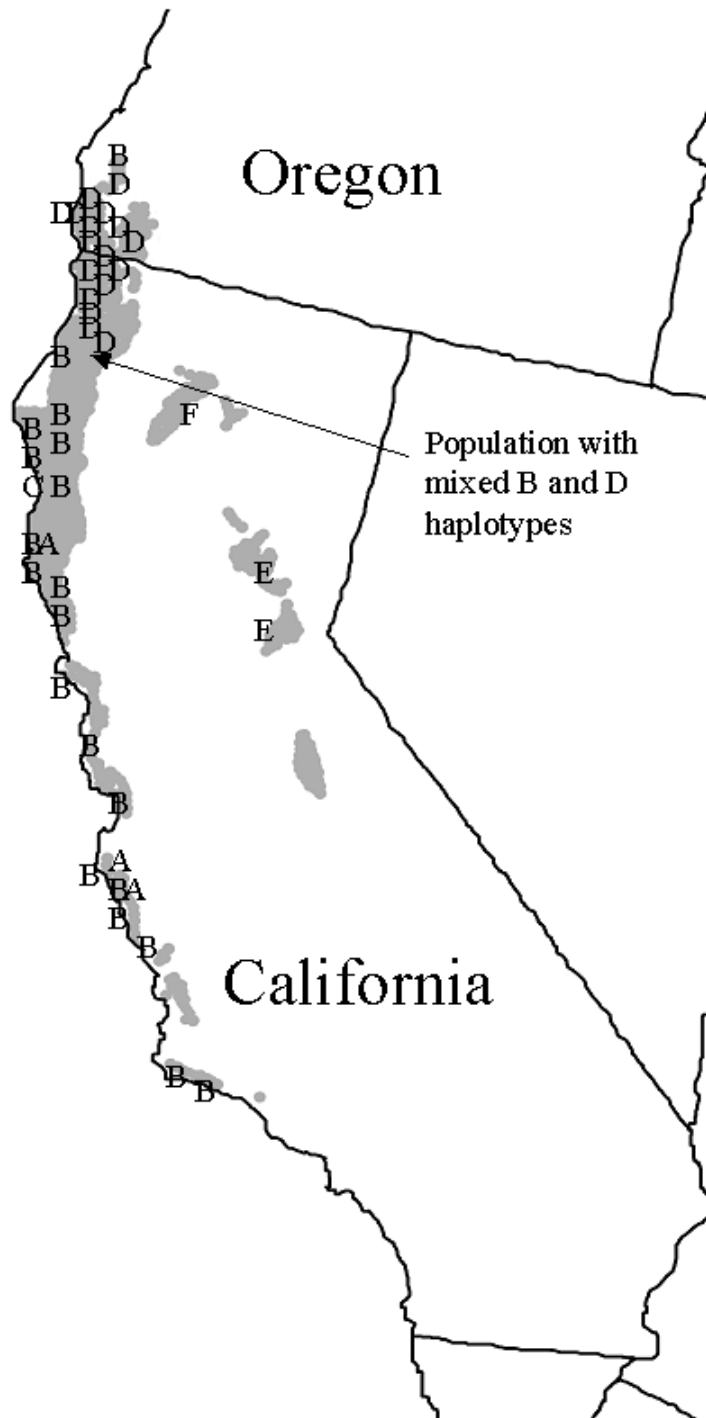
Figure 1—Distribution of chloroplast haplotypes. Four major haplotypes (B,D,E,F) and two rare haplotypes (A,C). Shading shows the distribution of tanoak from Tappeiner and others 1990.

Based on nuclear microsatellite diversity, the 905 individuals were assigned to 11

clusters with a posterior probability of 0.99. The partition of clusters is shown as a Voronoi tessellation in fig. 2. Significant breaks occurred: 1. between Hoopa and Korbel, consistent with the chloroplast haplotype lineage divergence; 2. near San Francisco between Point Reyes and the mid-San Francisco Peninsula; and 3. between coastal and interior (Weaverville, El Dorado, and Butte) populations. The three southernmost populations (Nacimiento, Lompoc, and Santa Barbara) and the three interior populations formed single population clusters. A single population of dwarf tanoak formed its own cluster suggesting lack of gene flow between it and nearby sources of the tree form. An unrooted tree based on Nei's genetic distances indicated major divergence between the two subspecies (fig. *2*). Interior populations of ssp. *densiflorus* were widely divergent from coastal populations. Among the latter, southern California populations from Nacimiento, Lompoc, and Santa Barbara were well differentiated, but remaining populations from central and northern California and Oregon showed relatively low levels of differentiation.
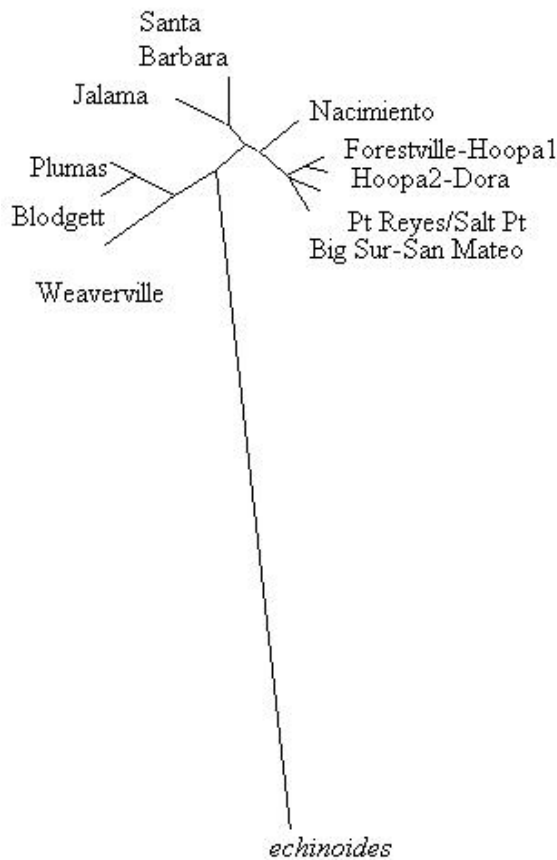

Figure 2—Unrooted tree based on Nei's genetic distances.

Genetic diversity, measured as allelic diversity, or as expected heterozygosity, increased significantly with latitude (fig. *3*). Using the same groupings detected by BAPS, we tested for differences in genetic diversity among selected groups (table 2). Overall, genetic diversity estimates were significantly different among groups. In pairwise comparisons, southernmost populations had significantly lower diversity than central coast populations and interior populations (Weaverville, Plumas, and Blodgett) had significantly higher diversity than all other populations.
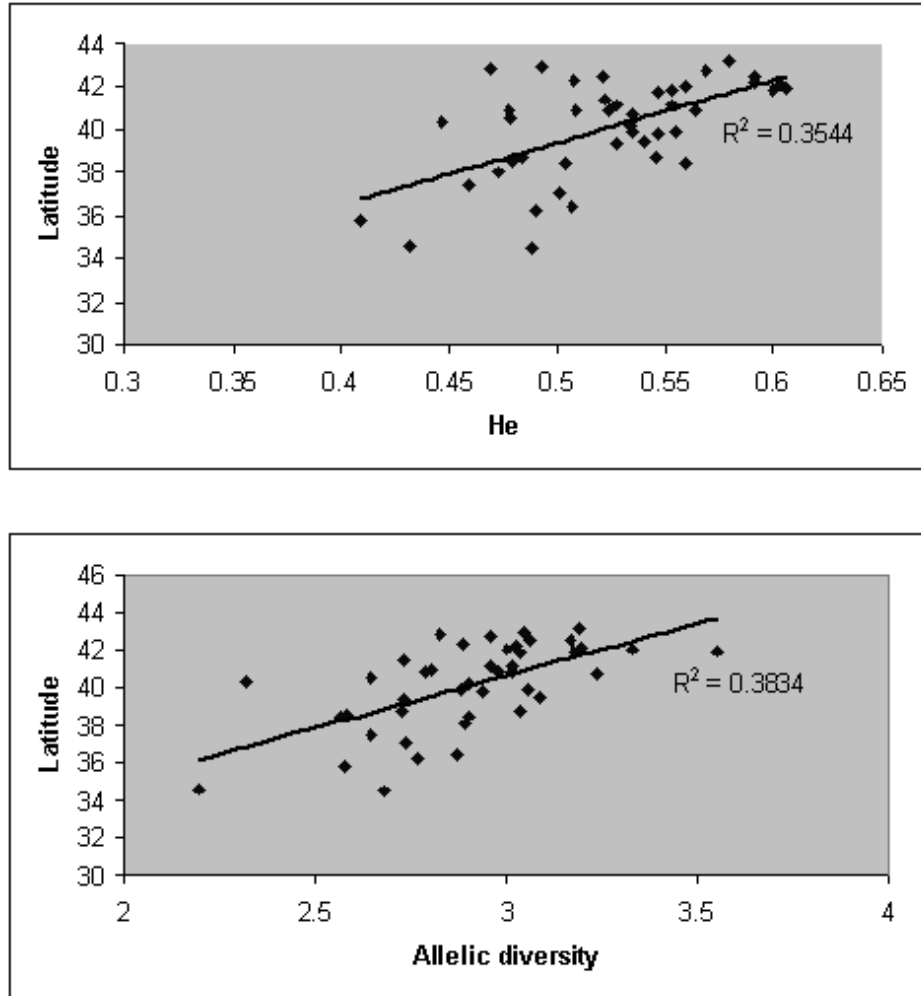

Figure 3—Plots of population genetic diversity, measured as expected heterozygosity (H_e) and allelic diversity, by latitude.

Table 2—Genetic diversity measured by allelic richness (*A*) and expected heterozygosity (*H*$_S$) for major clusters of populations of tanoak

	Group1	Group2	Group3	Group4	Group5
A	2.49	2.76*	2.82	3.00	3.11*
H$_S$	0.44	0.49	0.52	0.55	0.54

Groups 1 – 4 reflect increase in latitude, Group 5 is interior populations. Group1 – Nacimiento, Lompoc, Santa Barbara; Group 2 – Pfeiffer Big Sur to San Francisco Peninsula; Group 3 – Point Reyes to Korbel; Group 4 – Hoopa to Dora; Group 5 – Weaverville, Plumas, and Blodgett. Overall, two-sided probabilities after 10,000 permutations were $A = 0.006$ and $H_S = 0.005$. Asterisks indicate populations that showed significantly greater diversity than the group to the left in the table (one-tailed tests). All tests were carried out in FSTAT (Goudet 2002).

Discussion

Disease epidemics in forest ecosystems can have devastating consequences when the susceptible host is a keystone species. A cascade of biotic and abiotic effects can lead to transformation of the ecosystem, imposing permanent ecological, social and economic costs. Screening for resistance and protection of populations of conservation interest are two priorities in the early stages of the disease outbreaks. In wild ecosystems, the structure of genetic diversity in the threatened host is rarely known and must be one of the first avenues to be addressed for resource management. Screening for resistance is expensive and time-consuming and can be helped by knowledge of host genetic structure that might indicate the optimal sampling strategy. Conservation is aimed at maintaining existing evolutionary potential through identification of population lineages that may have unique adaptive potential (Moritz 1994, Crandall and others 2000). The recent SOD epidemic has indications of becoming a major transforming influence in the coastal forests and woodlands of California and Oregon (Barrett and others 2006, Meentemeyer and others 2004) and currently tanoak is the host that has suffered greatest mortality. Screening of tanoak for resistance has begun (Hayden and others, unpublished data), but the rationale for selecting populations for screening is hampered by lack of knowledge of genetic structure of this host species. This paper reports on variation in genetic diversity over an extensive range-wide sampling scheme.

Chloroplast haplotype diversity in tanoak is low, as expected for plants with heavy seeds and no specialized systems for long distance dispersal. The partition of haplotypes indicates the geographic limits to colonization by seed, and is therefore an excellent conservative estimate of seed zones based on the likely expansion of populations since the last glacial maximum (LGM). It is likely that the distribution of tanoak was restricted during the LGM, with some populations occupying sites where climatic conditions remained suitable. Four major haplotypes were detected indicating interior lineages in the Sierra Nevada and the Klamath Mountains, a coastal California lineage and a northern California-Oregon lineage. The distribution of tanoak today is discontinuous from the Sierra Nevada through the Klamath Mountains to coastal California, and so seed dispersal across these breaks is unlikely. However, the divergence of chloroplast haplotypes near to Arcata, Humboldt County, separating central and northern coastal California from extreme north coastal California and Oregon, is interesting. The significance of this break is supported by divergence in nuclear microsatellite diversity, indicating that gene flow by pollen has not been sufficiently invasive to break down the divergence between these two coastal chloroplast lineages. This would suggest that the northern and southern

179

lineages have only relatively recently met, or that there are partial barriers to mating. The latter could be realized by differences in average flowering times, a hypothesis that deserves further investigation.

Nuclear microsatellite divergence supports the lineages detected by chloroplast DNA, but also indicates some further differentiation among populations. Most significantly, analysis of nuclear DNA indicates a break between populations north and south of the San Francisco Bay and divergence among the three southernmost populations at Nacimiento, Lompoc, and Santa Barbara. In the southern range of tanoak, populations are dispersed and commonly relatively small. It is likely that lack of gene flow among these populations has led to differentiation through genetic drift. These southernmost populations exhibited the lowest levels of genetic diversity as would be expected if they were diverging through genetic drift. Levels of genetic diversity increase northwards, but interestingly the highest levels of diversity are in the Sierra Nevada/Klamath populations.

Tanoak Conservation

Many coastal tanoak populations have been heavily affected by the SOD epidemic in the last decade; mortality rates of more than 60 percent have been reported (Maloney and others 2005) and an exponential increase of mortality has been detected over a five-year period (Cobb and others 2008). Furthermore, the predicted trend for climate warming (IPCC 2001) is likely to affect tanoak adversely. The loss of tanoak trees throughout the different ecosystems where it is a major component will result in serious consequences for northern California and southern Oregon landscapes. Effects could be devastating, including a loss of wildlife that depend on acorns, a raise of fuel loads that would result in more extreme wildfires, and an increase in the risk of soil erosion in coastal ecosystems. Conservation efforts for habitat restoration and to preserve the evolutionary potential of this species are likely to be needed in the near future. Our study sheds light on the partition of genetic diversity at putatively neutral loci in both the nuclear and chloroplast genomes. Genetic variation and genetic differentiation throughout the range has to be taken into account to determine conservation priorities, seed sources for long-term preservation, and reforestation efforts.

Based on our chloroplast and nuclear DNA data, we identify six regions that should be considered as breeding groups that should be included in any resistance screening and should be represented in conservation management. We use the term breeding groups rather than seed zones because we cannot infer growth performance from our neutral DNA markers. 1. Extreme southern populations that in our sampling include Nacimiento, Lompoc, and Santa Barbara - These are identified as relatively depauperate genetically and should be considered as important in a conservation context. 2. Central coastal California from Pfeiffer Big Sur north to the San Francisco Peninsula - This region includes populations that have recently been ravaged by SOD disease and may therefore offer the opportunity for identifying tolerant individuals. 3. Northern coastal California from north of the San Francisco Bay to Arcata in Humboldt County - Some of the southern populations in this range have also been devastated by disease and will offer the possibility of identifying tolerant individuals. 4. Extreme north coastal California and Oregon - This is an area of extensive tanoak woodlands. Genetic diversity is relatively high and may therefore offer opportunities for finding diverse genotypes that may include resistant/tolerant traits. 5. Klamath

region - Our sampling in this region is relatively low, although ssp. *echinoides* has been sampled. This is an important region for conservation and screening and requires further investigation. 6. Sierra Nevada - The high genetic diversity of interior populations, relatively deep divergence from coastal populations, and lack of disease in this region make it an important component of resistant screening.

Acknowledgments

Literature Cited

Barrett, T.M.; Gatziolis, D.; Fried, J.S.; and Waddell, K. 2006. Sudden oak death in California: What is the potential? Journal of Forestry. 104: 61–64.

Cobb, R.C.; Lynch, S.C.; Meentemeyer, R.K.; and Rizzo, D.M. 2008. Five years of monitoring infection and mortality in redwood tanoak forests. In: Frankel, S.J.; Kliejunas, J.T.; and Palmieri, K.M., eds. Proceedings of the sudden oak death third science symposium. General Technical Report PSW-GTR-214, Albany, CA. U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station; 215–217.

Corander, J.; Marttinen, P.; Sirén, J.; and Tang, J. 2008. BAPS: Bayesian analysis of population structure. http://web.abo.fi/fak/mnf/mate/jc/software/baps.html. [August 18, 2009].

Crandall, K.A.; Bininda-Edmonds, O.R.P.; Mace, G.M.; and Wayne, R.K. 2000. Considering evolutionary processes in conservation biology: an alternative to 'evolutionary significant units'. Trends in Ecology and Evolution. 15: 290–295.

Cullings, K.W. 1992. Design and testing of a plant-specific PCR primer for ecological and evolutionary studies. Molecular Ecology. 1: 233–240.

Deguilloux, M.F.; Dumolin-Lapegue, S.; Gielly, L.; Grivet, D.; and Petit, R.J. 2003. A set of primers for the amplification of chloroplast microsatellites in *Quercus*. Molecular Ecology Notes. 3: 24–27.

Dodd, R.S.; Hüberli, D.; Douhovnikoff, V.; Harnik, T.Y.; Afzal-Rafii, Z.; and Garbelotto, M. 2005. Is variation in susceptibility to *Phytophthora ramorum* correlated with population genetic structure in coast live oak (*Quercus agrifolia* Nee)? New Phytologist. 165: 203–214.

Dodd, R.S.; Hüberli, D.; Mayer, W.; Harnik, T.Y.; Afzal-Rafii, Z.; and Garbelotto, M. 2008. Evidence for the role of synchronicity between host phenology and pathogen activity in the distribution of sudden oak death canker disease. New Phytologist. 179: 505–514.

Goudet, J. 2002. FSTAT, a program to estimate and test gene diversities and fixation indices. Ver. 2.9.3.2. Available from http://www.unil.ch/izea/softwares/fstat.html. [August 18, 2009].

IPCC [Intergovernmental Panel on Climate Change]. Climate change 2001: working group I: The scientific basis. Cambridge University Press. http://www.grida.no/climate/ipcc_tar/wg1/. [August 18, 2009].

Maloney, P.E.; Lynch, S.C.; Kane, S.F.; Jensen, C.E.; and Rizzo, D.M. 2005. Establishment of an emerging generalist pathogen in redwood forest communities. Journal of Ecology. 93: 899–905.

Meentemeyer, R.K.; Rizzo, D.M.; Walter, M.; and Lotz, E. 2004. Mapping the risk of establishment and spread of sudden oak death in California. Forest Ecology and Management. 200: 195–214.

Moritz, C. 1994. Defining "evolutionarily significant units" for conservation. Trends in Ecology and Evolution. 9: 373–375.

Morris V.R.F. and Dodd, R.S. 2006. Characterization of microsatellite markers for the tanoak tree *Lithocarpus densiflorus* Molecular Ecology Notes. 6: 706–708.

Nettel, A. and Dodd, R.S. 2007. Drifting propagules and receding swamps; genetic footprints of mangrove recolonization and dispersal along tropical coasts. Evolution. 61: 958–971.

Nettel, A.; Dodd, R.S.; and Rafii Z. [In press]. Genetic diversity, structure, and demographic change in tanoak, *Lithocarpus densiflorus* (Fagaceae), the most susceptible host to the sudden oak death disease in California. American Journal of Botany.

Oosterhout, C.V.; Hutchison, W.F.; Wills, D.P.; and Shipley, P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes. 4: 535–538.

Tappeiner, J.C.; Harrington, T.B.; and Walstad, J.D. 1984. Predicting recovery of tanoak and Pacific madrone after cutting and burning. Weed Science. 32: 413–417.

Weising, K. and Gardner, R.C. 1999. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome. 42: 9–19.