

Detecting the Limits of Northern and Southern Lineages of Tanoak in Northern California¹

Eduardo Sandoval-Castro² and Richard S. Dodd²

Abstract

Two chloroplast lineages of tanoak (*Notholithocarpus densiflorus*) meet between Korbel and Hoopa in the North Coast of California. Our earlier work suggests these lineages arose from southern and northern glacial refugia and this region represents their colonizing fronts. Earlier, we detected only one population of mixed lineages, suggesting that seedling establishment is limited locally, by short distance seed dispersal in tanoak. North western California and southern Oregon is recognized as a region of divergent lineages for a number of taxa, but it has not been confirmed whether the divergence is due to previous restriction in northern and southern refugia, or a colonizing front carrying rare alleles into new populations. Distinguishing the origin of this divergence is important, as the refugial scenario is likely to have greater implications for genetic adaptations to local environments and possibly for differences in disease resistance. Plant collections were made from the Hoopa Reservation for future resistance screening (Cobb and others 2008) and additional populations were sampled over the transitional zone. We use gene coalescence simulations to test among competing evolutionary scenarios. We detected minimal mixing of chloroplast haplotypes and a weak signal of nuclear DNA divergence across the break. Simulations confirmed the northern and southern refugial hypothesis.

Key words: chloroplast DNA, microsatellites, *Notholithocarpus densiflorus*, refugia, seed dispersal, tanoak

Introduction

Past events in the evolution of populations may lead to gene pools that have evolved independently in response to selective forces and to stochastic divergent processes. Although divergent processes may have occurred in the deep past, they can be detected in phylogeographic patterning among populations of living trees if population admixture is limited. In western North America a north-south divergence has been found in southern Oregon and northern California for a number of different taxonomic groups (Brunsfield and others 2001, Soltis and others 1997). Soltis and others (1997) attributed this divergence to one of two scenarios: 1) “North-south recolonization” following isolation in at least two Pleistocene refugia. 2) “Leading edge hypothesis” in which populations colonizing northwards underwent founder bottlenecks and a novel haplotype was carried forward on the colonizing front. Shafer and others (2010) extended the simple north-south recolonization hypothesis to take account of the complexity of refugia in the Pacific

¹ An abbreviated version of this paper was presented at the Seventh California Oak Symposium: Managing Oak Woodlands in a Dynamic World, November 3-6, 2014, Visalia, California.

² Post-doctoral Researcher and Professor, respectively, Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720. (sandovalcastro@gmail.com; dodd@berkeley.edu).

Northwest and emphasized the need to consider the ecological and evolutionary consequences of colonization from multiple refugia. Which of the two hypotheses is true in accounting for a phylogeographic divide within a species is important because long periods of isolation of populations in disjunct refugia would likely result in population groups with distinct gene combinations that could provide differential response to stress such as climate change and disease tolerance.

In previous work with *Notholithocarpus densiflorus*, we detected two major chloroplast lineages in coastal California (Dodd and others 2010, Nettel and others 2009). Spatially, these two lineages were composed of northern, or southern populations with a break occurring in Northern California at about 41 degrees North latitude between Korbel and Hoopa. Interestingly, we found only a single population in which chloroplast lineages were mixed over a distance of about 11 km. The sharp divergence between populations sharing the same chloroplast lineage suggests a single colonization route following mutations in the chloroplast genome given the leading edge hypothesis of Soltis and others (1997), or recent convergence of populations from northern and southern refugia with little opportunity for seed mixture (chloroplasts are maternally inherited in this species and therefore dispersed by seed).

Because nuclear and plastid DNA are differently inherited, the combination of plastid and nuclear molecular markers in DNA studies can provide a powerful complementary approach to detecting ancient to more recent population demographic changes. For heavy-seeded species such as tanoak, seed dispersal is likely to be quite limited, and so chloroplast DNA provides a good indication of ancient fragmentation of the species' distribution. On the other hand, highly polymorphic, co-dominant markers such as nuclear microsatellites allow for inferences on fluctuations in population size, including founder events and bottlenecks.

Here, we have taken a transect over the north-south chloroplast lineage break, to gain more insights on the transition zone and the degree of seed admixture. We test the hypothesis of northward colonization against the competing hypothesis of northern and southern refugia by comparing demographic simulations with data from nuclear microsatellites sampled across the transition zone.

Materials and methods

Sampling and DNA analysis

Leaves from mature individuals of *Notholithocarpus densiflorus* subsp. *densiflorus* were obtained from coastal populations situated between Korbel and Hoopa reservation in the North Coast of California. We sampled a total of 258 trees from 14 populations, with an average of 19 trees per population (fig. 1, table 1). Total genomic DNA was extracted from the fresh foliage using a simplified cetyltrimethyl ammonium bromide (CTAB) method (Cullings 1992).

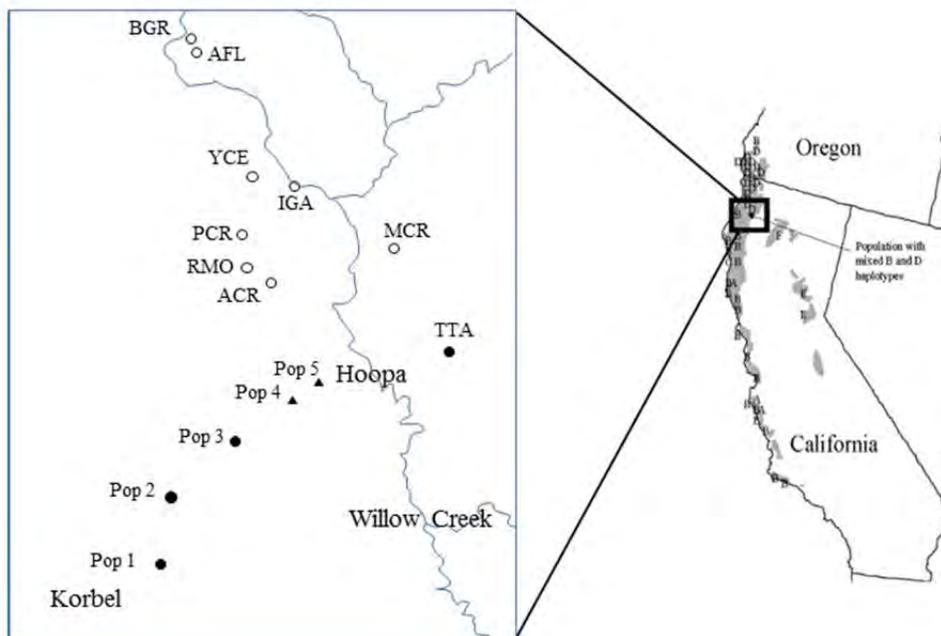


Figure 1—On right, map showing the distribution of chloroplast haplotypes and admixed area reported by Dodd and others (2010). On left, sampling sites analyzed in this study from coastal populations situated between Korbel and Hoopa reservation in the North Coast of California (Pop4 and Pop5 showed admixed haplotypes).

Chloroplast DNA

Four polymorphic cpDNA microsatellite loci previously used in tanoak (ucd2, ccmp4, cmcs5, and rpob; Nettel and others 2009) were PCR amplified using a fluorescent-labeled primer. PCR amplifications were performed following protocols described by Nettel and others (2009). To ensure reproducibility and consistency in PCR amplification, a negative and two positive samples were run as a control for each set of PCR reactions and genotyped. PCR cocktails were electrophoresed on an ABI 3730 automated sequencer (Applied Biosystems). Microsatellite alleles were visualized and scored in the program GeneMarker 2.2.0 (Softgenetics).

Nuclear DNA

Eleven nuclear microsatellite loci developed for tanoak (LD1, LD3, LD5, LD7, LD8, LD10, LD12, LD13, LD14, LD17, and LD19; Morris and Dodd 2006) were PCR amplified using fluorescent-labeled primers multiplexed in five different groups: 1) LD1, LD3, LD7, LD10, LD14; 2) LD5, LD17; 3) LD12, LD13; 4) LD19; and 5) LD8. PCR protocols followed those described by Nettel and others (2009). To ensure reproducibility and consistency in PCR amplification, a negative and two positive samples were run as a control for each set of PCR reactions and genotyped. PCR cocktails were electrophoresed on an ABI 3730 automated sequencer (Applied Biosystems). Microsatellite

alleles were visualized and scored in the program GeneMarker 2.2.0 (Softgenetics).

Data analysis

We checked data quality for scoring errors in nuclear microsatellites using MICRO-CHECKER (Oosterhout and others 2004) and estimated the probability of null alleles using the same program. 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 population genetic diversity as allelic richness (A_R) and expected heterozygosity (H_e) with FSTAT. Population structure was inferred using a Bayesian clustering algorithm implemented in STRUCTURE v2.2.3 (Pritchard and others 2000), assuming no a priori knowledge of population boundaries. STRUCTURE runs were based on 500,000 iterations after a burn-in of length 500,000 and assumed correlated allele frequencies and an admixture model with an estimated proportion α of admixed individuals. To check for Markov chain Monte Carlo (MCMC) convergence, we performed 10 replicates for each K value and checked the consistency of results. The most likely number of clusters (K) was considered to be the K value with the highest $\Pr(X|K)$ (Pritchard and others 2000). Also, the optimal K value was calculated after the ΔK method described by Evanno and others (2005). Output from STRUCTURE was post-processed using the programs CLUMP (Jakobsson and Rosenberg 2007) and DISTRUCT (Rosenberg 2004).

Demographic history models

We employed Approximate Bayesian Computations (ABC) to test the relative likelihood of a northward migration (leading edge hypothesis) against colonization from northern and southern refugia (north-south recolonization hypothesis). ABC tests were performed in DIYABC v. 2.0.3 (Cornuet and others 2013). In preliminary analyses, we ran a simple northward colonization scenario in which a northern group of populations (ACR, AFL, BGR, IGA, MCR, PCR, RMO, YCE) defined by the northern chloroplast haplotype (see table 1) colonized from admixed populations (Pop4 and Pop5) that in turn colonized from southern populations (Pop1, Pop2, Pop3 and TTA). This scenario was tested against two competing scenarios; 1) Colonization of the northern group of populations directly from the southern group followed by an admixture event to produce Pop4 and Pop5 and, 2) an admixed group (Pop4 and Pop5) that arose after contact of northern and southern groups expanding from northern and southern refugia. Parameter (two-population summary statistics: allele size variance, pairwise F_{ST} , Goldstein's $\delta\mu^2$ and shared allele distance) estimation was from a reference table, which consisted of 1,000,000

simulated datasets. We used the default generalized stepwise mutation model with a mean mutation rate across loci set at 10^{-3} to 10^{-4} per locus per generation. Uniform priors were used for current and past population sizes and for the timing of demographic events.

Results

Chloroplast diversity and structure

We detected two alleles each at loci *ccmcs5* and *rpob*. Loci *ccmp4* and *ucd2* were monomorphic in all populations analyzed. Interestingly, alternate alleles at these two latter loci were fixed in northern and southern populations giving the same two divergent chloroplast haplotypes reported earlier by (Dodd and others 2010, Nettel and others 2009). The resulting two haplotypes were detected admixed in only two populations, within a latitudinal band of about 7 km. Both populations were dominated by the haplotype of southern origin (table 1, fig.1). This finding supported previous studies on tanoak (Dodd and others 2010, Nettel and others 2009), suggesting this area as a recent contact zone between north and south haplotypes.

Table 1—Characterization of the 14 tanoak populations sampled in Northern California

Population	Code	Latitude	Longitude	N	Nuclear		Chloroplast Haplotype
					A _R	H _e	
100 Acre	ACR	41°06'	-123°46'	18	3.8	0.55	A
Acorn Flat	AFL	41°17'	-123°50'	15	4.5	0.66	A
Bear Grass	BGR	41°18'	-123°50'	33	4.0	0.62	A
Iron Gate	IGA	41°11'	-123°44'	10	4.0	0.56	A
Mill Creek	MCR	41°08'	-123°38'	18	3.9	0.56	A
Pine Creek	PCR	41°09'	-123°47'	20	3.9	0.57	A
Ranger Mountain	RMO	41°07'	-123°47'	18	3.8	0.55	A
Yurok Cemetery	YCE	41°11'	-123°46'	17	4.0	0.62	A
Tish Tang	TTA	41°03'	-123°35'	17	4.0	0.57	B
POP1	Pop1	40°52'	-123°52'	20	4.0	0.58	B
POP2	Pop2	40°56'	-123°51'	20	3.9	0.58	B
POP3	Pop3	40°58'	-123°47'	20	4.5	0.63	B(19), A(1)
POP4	Pop4	41°00'	-123°44'	20	4.3	0.61	B(12), A(8)
POP5	Pop5	41°01'	-123°43'	12	4.2	0.62	B(9), A(3)

N = sample size; H_e = expected heterozygosity; A_R = allelic richness detected with nuclear microsatellites; Haplotypes A and B detected with Chloroplast microsatellites Number in parenthesis is the number of individuals with corresponding haplotype A or B

Nuclear diversity and structure

Genetic diversity, measured as allelic richness ranged from an average of 3.8 alleles per locus in ACR and RMO to 4.5 in AFL and Pop3. Expected heterozygosity ranged from 0.55 in ACR and RMO to 0.66 in AFL (table 1). Using the same groupings detected by chloroplast DNA, we tested for

differences in genetic diversity among selected groups. Overall, genetic diversity estimates showed that admixed populations had slightly higher diversity than northern and southern groups (table 2), with differences at or close to the 5 percent level of significance in the contrast between northern and admixed groups.

Table 2—Genetic diversity measured by allelic richness (AR) and observed heterozygosity (HO) for major clusters of populations of tanoak. Overall, one-sided probabilities after 10,000 permutations were AR = 0.05 and HO = 0.07 for North vs Admixed and AR = 0.20 HO = 0.19 for South vs Admixed

Genetic Diversity	North group	South group	Admixed group
AR	3.98	4.10	4.28
HO	0.58	0.59	0.62

Nuclear microsatellites showed high levels of admixture among populations and the STRUCTURE analysis was unable to detect any signal of clustering (fig. 2). However, a chi-square test found a significant ($p > 0.001$) proportion of individuals from northern and southern chlorotypes had greater than 90 percent assignment to the alternate nuclear microsatellite clusters. This is seen as a subtle difference in fig. 2.

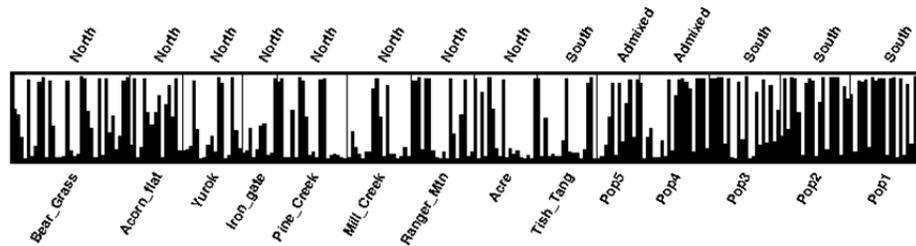


Figure 2—STRUCTURE analysis of sample locations across the transition zone from southern chloroplast to northern chloroplast haplotypes in tanoak. See fig. 1 for sample locations. Vertical bars represent proportion of an individual’s nuclear genome assigned to one of two clusters, white or black.

Demographic history in tanoaks

Among the competing scenarios that could account for colonization of this region of northern California, our DIYABC analyses found the refugial expansion model, with admixture to be the best fit to our nuclear microsatellite data (fig. 3).

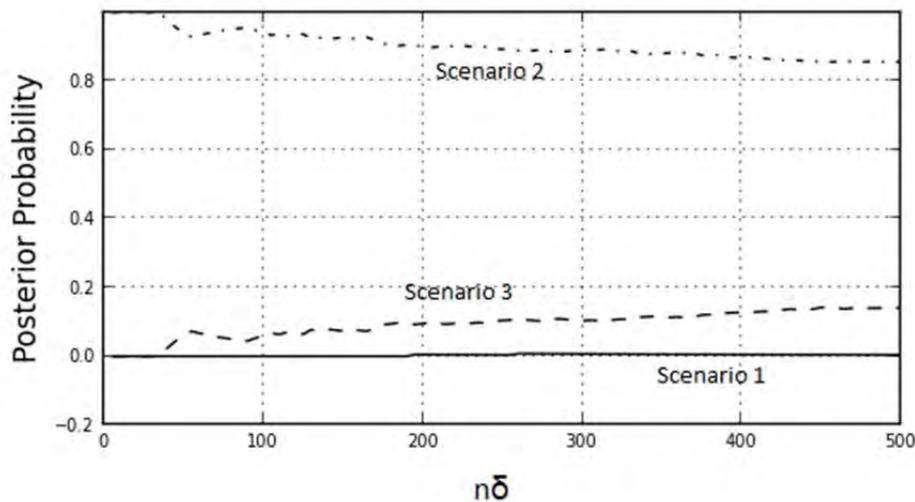


Figure 3—Posterior probabilities for three competing evolutionary scenarios: Scenario 1 simple northward colonization; Scenario 2 secondary contact from northern and southern refugia; Scenario 3 northward colonization from southern haplotype to northern haplotype groups, followed by admixture in region of mixed haplotypes. $n\delta$ is number of closest datasets.

Discussion

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 (Dodd and others 2010, Nettel and others 2009), indicating that gene flow by pollen has not been sufficiently invasive to fully break down the divergence between these two coastal chloroplast lineages. This would suggest that the northern and southern lineages have only relatively recently met, or that there are partial barriers to mating.

Here we analyzed populations of tanoaks situated in the area where previous studies have found a mixture of haplotypes from the Northern and Southern Coast of California (Dodd and others 2010, Nettel and others 2009). We found higher levels of genetic diversity in the sampling sites where both cpDNA haplotypes were detected. This finding supports our results from Bayesian modeling that these populations of mixed chloroplast haplotypes are a result of secondary contact between divergent groups rather than a stage in the northward colonization of tanoak.

We found very low levels of diversity and high differentiation in cpDNA compared with nuclear microsatellites. Since the chloroplast genome is maternally inherited, the apparent limitation of polymorphic populations to a

narrow band of about 7 km suggests that the distance of seed dispersal is very limited, or that the northern and southern refugial groups have only very recently come into secondary contact. Our earlier work detected a long-term demographic decline in tanoak, probably dating to the mid Holocene. Therefore, it seems most likely that the timing of secondary contact occurred only shortly before the onset of population decline.

In conclusion, our results showed an abrupt transition in cpDNA haplotypes, with the break occurring between Korbel and Hoopa. Nuclear microsatellites showed more genetic diversity with higher levels observed in sampling locations where both cpDNA haplotypes were found together, consistent with expectations in a zone of admixture among divergent lineages. The Bayesian modeling confirmed that the chloroplast divergence was most likely a result of northern and southern refugial populations. The results underline the importance of treating coastal California tanoak as two lineages with divergent evolution that could have important consequences for adaptive traits and must be taken into consideration in any restoration and conservation efforts.

References

- Brunsfeld, S.; Sullivan, J.; Soltis, D.; Soltis, P. 2001. **Comparative phylogeography of northwestern North America: a synthesis**. In: Silvertown, J.; Antonovics, J., eds. Integrating ecological and evolutionary processes in a spatial context. Oxford, UK: Blackwell Science: 31–9339.
- Cobb, R.C.; Lynch, S.C.; Meentemeyer, R.K.; Rizzo, D.M. 2008. **Five years of monitoring infection and mortality in redwood tanoak forests**. In: Frankel, S.J.; Kliejunas, J.T.; Palmieri, K.M., eds. Proceedings of the sudden oak death third science symposium, Gen. Tech. Rep. PSW-GTR-214. Albany, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station: 215–217.
- Cornuet, J.M.; Veyssier, J.; Pudlo, P.; Dehne-Garcia, A.; Gautier, M.; Leblois, R.; Marin, J.M.; Estoup, A. 2013. **DIYABC v2.0: a software to make approximate Bayesian computation 35 inferences about population history using single nucleotide polymorphism, DNA sequence and 36 microsatellite data**. *Bioinformatics* 30: 1187–1189.
- Cullings, K.W. 1992. **Design and testing of a plant-specific PCR primer for ecological and evolutionary studies**. *Molecular Ecology* 1: 233–240.
- Dodd, R.S.; Afzal-Rafii, Z.; Mayer, W. 2010. **Ancestral seed zones and genetic mixture of tanoak**. In: Frankel, S.J.; Kliejunas, J.T.; Palmieri, K.M., tech. coords. Proceedings of the sudden oak death fourth science symposium. Gen. Tech. Rep. PSW-GTR-229. Albany, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station: 171–182.
- Evanno, G.; Regnaut, S.; Goudet, J. 2005. **Detecting the number of clusters of individuals using the software structure: a simulation study**. *Molecular Ecology* 14: 261–12620.
- Goudet, J. 2002. **FSTAT, a program to estimate and test gene diversities and fixation indices, version 2.9.3.2 [computer program, documentation]**. <http://www.unil.ch/izea/software/fstat.html>. (12 February 2015).
- Jakobsson, M.; Rosenberg, N.A. 2007. **CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure**. *Bioinformatics* 23: 1801–1806.

- Morris, V.R.F.; Dodd, R.S. 2006. **Characterization of microsatellite markers for the tanoak tree, *Lithocarpus densiflorus*.** *Molecular Ecology Notes* 6: 706–708.
- Nettel, A.; Dodd, R.S.; Afzal-Rafii, Z. 2009. **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* 96: 2224–2233.
- Oosterhout, C. van; Hutchison, W.F.; Wills, D.P.; Shipley, P. 2004. **MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data.** *Molecular Ecology Notes* 4: 535–538.
- Pritchard, J.K.; Stephens, M.; Donnelly, P.J. 2000. **Inference of population structure using multilocus genotype data.** *Genetics* 155: 945–959.
- Rosenberg, N.A. 2004. **DISTRUCT: A program for the graphical display of population structure.** *Molecular Ecology Notes* 4: 137–138.
- Shafer, A.; Cullingham, C.I.; Cote, S.D.; Coltman, D.W. 2010. **Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America.** *Molecular Ecology* 19: 4589–4621.
- Soltis, D.E.; Gitzendanner, M.A.; Strenge, D.D.; Soltis, P.S. 1997. **Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America.** *Plant Systematics and Evolution* 206: 353–373.