Report on Save the Redwoods League grant: Developing Conservation strategies for Giant Sequoia, an imperiled California endemic

Richard Dodd Principal Investigator

In the proposed research we planned to determine population genetic diversity and differentiation in giant sequoia to address some questions that would have consequences for its conservation: these included the following:

1. Is genetic diversity at neutral markers partitioned evenly among populations of giant sequoia? If so, does the architecture of diversity follow the common model for long-lived outcrossing woody species in which most genetic diversity is held within populations and population differentiation is low. The consequences of such a genetic architecture are that conservation strategies can be focused at the level of the species and not of individual populations.
   Hypothesis: Due to small population sizes, particularly in the northern range of giant sequoia, genetic diversity has been lost and populations have diverged through the stochastic effects of genetic drift.
2. Is there evidence of deep lineage divergence among populations consistent with long separation of population lineages, or is divergence among populations low suggesting that contemporary populations are the result of colonization from single source populations? If divergence is deep, this would suggest that different lineages of giant sequoia should be given particular attention in conservation management because they are more likely to have different adaptive gene complexes that could play an important role in adaptation under climate change.
   Hypothesis: The northern populations have a long history of separation from southern groves. Colonization during population demographic fluctuations has been from local populations.
3. Is present population size correlated with levels of inbreeding? If there is a strong inbreeding effect, assisted outcrossing by introducing plants from other populations may reduce risks of inbreeding depression and the possible extinction of populations.
   Hypothesis: Small populations, particularly in the northern disjunct range, will have high levels of inbreeding as a result of low levels of migration among populations.

Development of microsatellite markers

This necessitated the development of a suitable molecular marker system that was unavailable for this species. We chose to develop microsatellite markers as these are typically hypervariable and would most likely provide an adequate level of polymorphism and a system that would be suitable for further analyses including inferences on past demographic changes.

We identified 36 repetitive DNA sequences and developed primers for their amplification. Of these, 11 proved reliable for polymerase chain reaction (PCR) amplification and were variable (Table 1). The 11 loci have been reported in Conservation Genetics Resources (DeSilva and Dodd 2014. Development and characterization of microsatellite markers for giant sequoia,

Sequoiadendron giganteum (Cupressaceae) Conservation Genetics Resources 6: 173-174). A copy is attached to this report.

Table 1: Characteristics of microsatellite markers for *Sequoiadendron giganteum*

Primer ID	Sequence 5'-3'	Repeat type	Allele range (bp)	no. of alleles	H_O^*	H_E^*
GS_40527	F:GCTCACATCACATCCACAAGC R:TGTCGACCTAATCGATGTTTGAAG	$(CA)_{12}$	107-125	8	0.286	0.533
GS_7365	F:CAACCAACTGATCCTACATTGC R:TGTGCAGGTTGTTGTCTTGC	$(ACA)_9$	211-229	5	1.00	0.641
GS_29596	F:AAACCCCCGTTTTGGTGTTC R:TGCTCTATACTACTCTCTTAGCTCTC	$(AG)_{13}$	257-287	14	0.550	0.682
GS_31267	F:AGGGAAGGAGATGTAGACAAAGG R:CTCTCTACCCCCACTCTCTATG	$(GA)_{11}$	142-180	15	0.429	0.582
GS_34305	F:GACTTGTCTTGATTCCTTGACTG R:TCATCTCAAGTCATACACTGCC	$(GT)_{15}$	88-102	6	0.667	0.690
GS_17786	F:ACATAACGCAAACATGGGGG R:TTGGTCGATGAGTGCTAGGG	$(AC)_{11}$	104-114	5	0.476	0.717
GS_36493	F:TCCCTTCATCAGTCCCTACC R:GGAGAGGCATGCAGACAAAG	$(CT)_{11}$	135-149	8	0.381	0.466
GS_31670	F:TATGGTAGAGGGTAGAGGGG R:ACCACGCACACACTAACTC	$(GA)_{14}$	155-191	15	0.524	0.793
GS_39473	F:TCATGAGTAGTGGGTTCACAAG R:GAGAGAGAGATCGAGGTGTG	$(CA)_{16}$	111-161	25	0.714	0.854
GS_30133	F:ACACCATGCCTCTATCCGAG R:AGAGTGGGAAGCTGATGACC	$(AC)_{13}$	185-201	9	0.619	0.645
GS_33118	F:ACTCAGGGACAAGAACGTGG R:ACACAAGCAAGACCGAATATAGC	$(CA)_{12}$	163-169	3	0.190	0.177
M13	**Fam - TGCCATCCCTATACACAACCA**					

*H-obs and H-exp are based on the Cabin Creek Sequoia grove (n=21)

Rangewide study of genetic diversity

METHODS

We amplified the 11 microsatellite loci for a total of 357 individuals covering the geographic range of giant sequoia. The trees were from a clonal orchard at the University of California Russell Research station that was established in 1981 from seed collections made in 1974 to 1976. More detail on the clonal orchard can be found in Fins and Libby (1982). Twenty three of the natural giant sequoia groves were represented, including 8 northern disjunct groves from Placerville to McKinley and 15 southern groves to the southernmost Deer Creek grove (Table 2). Foliage was sampled for DNA extraction.

DNA Extraction and amplification

DNA was extracted using the cetyltrimethyl ammonium bromide (CTAB) method, following Cullings (1992). Individuals were genotyped at the eleven microsatellite loci, described above and in Table 1. All DNA amplifications were carried out using an 18 bp M13 tail attached to the forward primers and a universal fluorescent-labeled M13 primer, employing the method described by Schuelke (2000). Details of the PCR reactions are described in DeSilva and Dodd (2014) attached to this report. Amplified products were run through an ABI 3730 automated sequencer. Microsatellite fragments were analyzed with Genescan 3.7 and Genotyper 3.7 software (Applied Biosystems).

Table 2: Populations sampled in the Russell Research Station clonal orchard *Sequoiadendron giganteum* and

Grove	Latitude	Longitude	Grove area (ha)	Sample size	Mean Allelic diversity per locus*	H_O	H_E	F_{IS}
Placerville	39.057	-120.574	1	8	2.08	0.46	0.53	0.15
N. Calaveras	38.279	-120.302	24	29	3.20	0.47	0.56	0.16
S. Calaveras	38.247	-120.240	184	21	3.60	0.54	0.63	0.15
Tuolomne	37.769	-119.807	8	8	2.64	0.35	0.53	0.35
Merced	37.750	-119.839	8	8	3.05	0.61	0.60	-0.02
Mariposa	37.509	-119.604	101	15	3.61	0.51	0.65	0.22
Nelder	37.435	-119.590	195	29	3.73	0.58	0.69	0.16
McKinley	37.023	-119.105	22	19	3.19	0.58	0.57	-0.02
Cabin Creek	36.806	-118.941	40	21	3.39	0.53	0.62	0.15
Converse Basin	36.809	-118.977	1498	21	3.71	0.63	0.67	0.06
Lockwood	36.793	-118.841	40	10	3.45	0.55	0.62	0.12
Windy Gulch	36.766	-118.811	405	13	4.07	0.61	0.68	0.11
Grant	36.750	-118.984	130	2	-**	-**	-**	-**
Redwood Mtn	36.694	-118.916	1271	17	3.48	0.57	0.65	0.12
Giant Forest	36.565	-118.752	855	17	3.69	0.58	0.63	0.10
Atwell Mill	36.468	-118.674	383	18	3.70	0.60	0.63	0.05
South Fork	36.358	-118.706	97	8	3.68	0.65	0.65	0.01
Mtn Home	36.230	-118.681	1620	17	3.70	0.58	0.62	0.08
Wheel Meadow	36.140	-118.513	498	30	3.83	0.61	0.65	0.07
Black Mtn 1	36.118	-118.679	669	13	3.68	0.54	0.62	0.14
Black Mtn 2	36.102	-118.649	669	16	3.63	0.50	0.62	0.20
Packsaddle	35.929	-118.592	137	12	2.93	0.53	0.61	0.14
Deer Creek	35.872	-118.609	21	5	2.82	0.70	0.63	0.37

*Based on minimum population size of 5, **Too few individuals

Data analysis

The microsatellite data were analyzed in ARLEQUIN Vers. 3.5 (Excoffier et al 2005) to obtain genetic diversity indices within populations and departure from Hardy Weinberg equilibrium.

Isolation by distance

Because of the linear distribution of giant sequoia in the Sierra Nevada, we recognize that a pattern of isolation by distance (IBD) may confound discrete population genetic structure. Pritchard *et al.* (2000) cautions that where IBD is significant, interpretation of STRUCTURE results will be challenging. Therefore, we tested for IBD, first on the full data set and then on the

northern disjunct populations and the southern populations separately. The Placer population was excluded from the rest of northern populations because of small population size and very low genetic diversity. Deer Creek was excluded from the southern populations because of its small population size. Tests of IBD were performed in IBDWS Vers. 3.23 (Jensen *et al.* 2005), with 30,000 randomizations of Slatkin's similarity index [M= ((1-F_{ST})-1)/4)] and geographic distances based on population centroids of latitude and longitude coordinates.

Population structure

The program STRUCTURE v.2.3.3 (Pritchard et al. 2000, Falush et al. 2003), which implements a model-based clustering method, was used to infer the number of genetic clusters (K) present in the dataset, and to assign individuals to these clusters. We used the admixture model with correlated allele frequencies and did not include any a priori information about collection sites. For all final simulations performed in STRUCTURE, we used a ''burn-in'' period of $5x10^4$ followed by a run length of $5x10^5$ MCMC iterations, which were sufficient to give a stable α and estimate of the log probability of the data. Because genetic diversity of the total data set showed a significant pattern of isolation by distance, whereas, IBD was not significant for northern and southern groups separately (see IBD analyses), we ran STRUCTURE with reduced data sets (northern from Placer to McKinley and southern including all other populations) separately, using K=1–8 and K=1–12, for north and south respectively. We ran simulations ten times for each of the three STRUCTURE analyses. To estimate the most likely K value, we used the ΔK statistics (Evanno et al. 2005) by inputting STRUCTURE output into STRUCTURE HARVESTER (Dent et al 2012). Results are displayed using the program DISTRUCT (Rosenburg 2004).

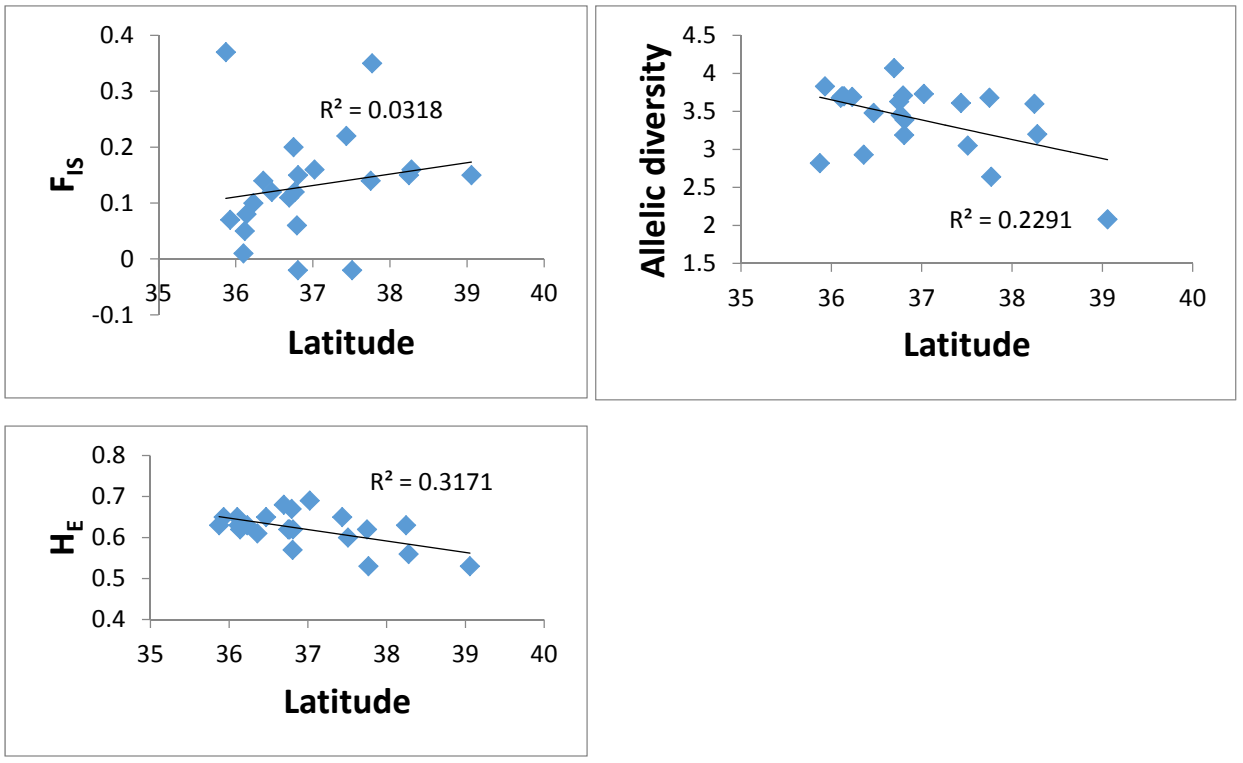
RESULTS

Genetic diversity

Mean allelic diversity was lowest at Placerville and greatest at Windy Gulch (Table 2). These estimates are adjusted for population size, with the minimum size of 5 individuals at Deer Creek and represent the average number of alleles at each of the 11 loci. Observed heterozygosities ranged from 0.35 at Tuolomne to 0.70 at Deer Creek. Ranking of expected heterozygosities did not follow that of observed heterozygosities and the departures from Hardy Weinberg equilibrium (HWE) due to an excess of homozygotes are indicated by F_{IS} values in Table 2. F_{IS} was high at Tuolomne and Deer Creek. Allelic diversity and expected heterozygosity were positively correlated with grove size (Mantel r= 0.12 and r= 0.11 respectively) and negatively correlated with latitude of origin (Mantel r= -0.64 and r= -0.54 respectively). The partial Mantel test for latitude of origin after removing grove size was r= -0.64, two tail Prob = 0.1 for allelic diversity and r= -0.54, two tail Prob =0.1 for expected heterozygosity. Although two small groves (Merced and Deer Creek) showed substantial departure from HWE evidenced by high

values of F_{IS}, there was no correlation between grove size, or latitude and inbreeding coefficient (Fig 1).

Fig 1. Plots of genetic diversity measures with latitude of origin of giant sequoia groves



We tested for partition of genetic diversity within populations and among populations within the northern and southern groups using the Analysis of Molecular Variance procedure in ARLEQUIN. As is common in trees species most of the genetic variance (88%) was within populations, with 12% attributable to among populations within the northern and southern groups of groves. Only 0.5% was attributable to the difference between the two groups of groves (Table 3).

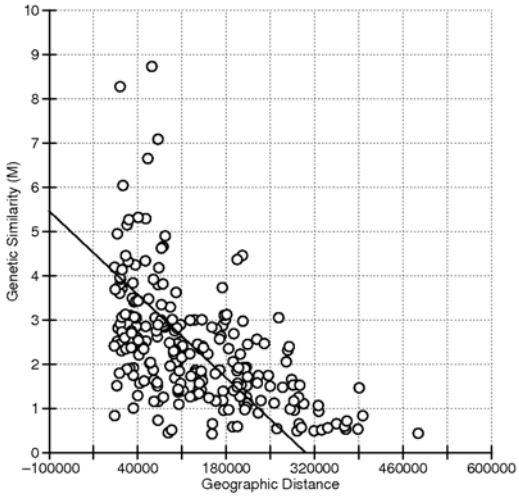
Table 3. Analysis of Molecular Variance of microsatellite diversity in giant sequoia

Source of variation	d.f.	Sums of squares	Variance components	Percentage variation
Among groups	1	26.663	0.0176	0.46
Among populations Within groups	21	348.477	0.43780	11.56
Within populations	691	2302.239	3.33175	87.98
Total	713	2677.380	3.78715	

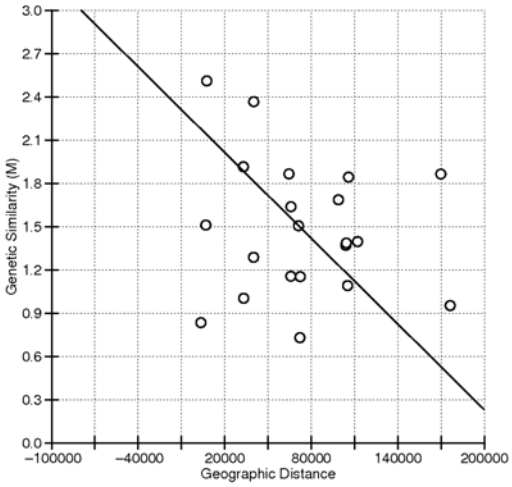
Isolation by distance

Isolation by distance was significant over the entire range of the sampled populations when a Mantel test was performed between raw genetic similarity indices and raw geographic distances Mantel *r* was significant (r = -0.54; Prob $r \leq 0 < 0.0001$) (Fig 2). However, separate analyses for northern groves (North Calaveras to McKinley) and southern groves (Cabin Creek to Deer Creek) were not significant (r = -0.15; Prob $r \leq 0$ = 0.69 for northern groves) (r = -006; Prob $r \leq 0$ = 0.68 for southern groves).

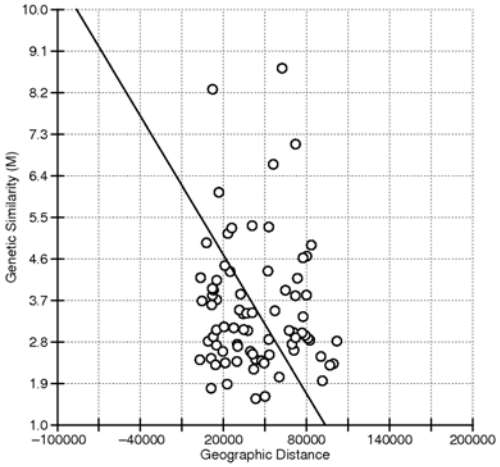
Fig 2 Plots of Isolation by Distance: Slatkins similarity and geographic distance



All groves



Northern Groves



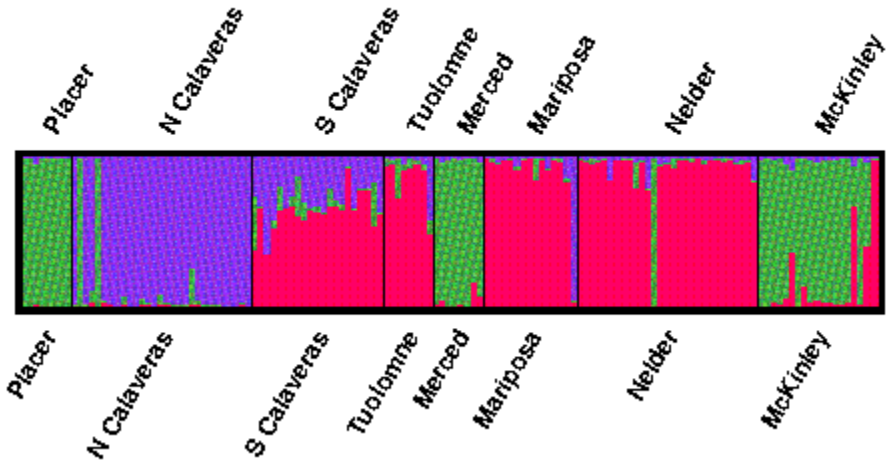
Southern Groves

Population structure

The Evanno method found an optimum of 3 groups from the STRUCTURE analyses for northern groves and 4 groups for southern groves. Populations in the north showed relatively low

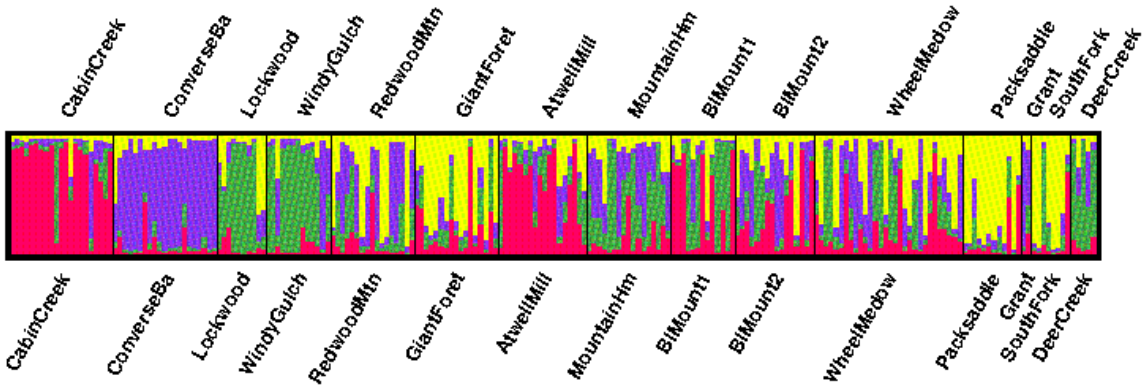
admixture within populations, or admixture within individuals, suggesting lack of gene flow among the groves (Fig3). South Calaveras showed the greatest admixture, sharing ancestry with North Calaveras and the Tuolomne to Nelder group. Merced and McKinley formed a cluster despite Merced being very close to the Tuolomne grove. The Placer grove was also attributed to this cluster. It seems unlikely that these three well-separated groves form a lineage distinct from the other northern populations. It is more likely that small population sizes at Placer and Merced contributed to poor assignments.

Fig 3. Assignment of northern populations of giant sequoia to 3 groups from STRUCTURE output. Vertical bars represent single individuals colour coded for the group to which they were assigned.



Southern populations showed much greater levels of admixture, both among populations and among individuals within populations. The two northernmost groves at Cabin Creek and Converse Basin of this group were the most uniform (Fig4).

Fig 4. Assignment of southern populations of giant sequoia to 4 groups from STRUCTURE output. Vertical bars represent single individuals colour coded for the group to which they were assigned.



Population size changes

We have not completed the analyses of demographic processes. However, our preliminary data indicate: 1. Populations have undergone contractions. This is supported by our preliminary runs of the coalescent model in LAMARC (Kuhner 2006) and in MSVAR (Beaumont 1999). The latter places the beginning of population contraction at about 5 Kya. These programs use coalescence simulators that require considerable computing memory and time. Therefore, it is impractical to run the software on all populations. Our final simulations will include two populations from the northern group and two populations from the southern group. At this time, our preliminary data shows no difference between populations drawn from northern and southern groups.

CONCLUSIONS

This is the first molecular report on genetic diversity in giant sequoia. Some of the earlier findings from isozymes (Fins and Libby 1982) are confirmed, but we are able to elucidate much more of the genetic structure of this species.

We set out address three questions. The first relates to the architecture of genetic variation in the species. We find that, like many wind-pollinated tree species, genetic diversity at microsatellite loci is partitioned mainly at the level of populations. Approximately 88% of genetic variation is attributable to populations with the remainder attributable to variation among populations. Analysis of molecular variance (AMOVA) suggests only a small genetic variance associated with divergence between northern disjunct and southern more continuous populations. However, our Bayesian analyses with STRUCTURE suggest that the northern populations are distinct. The divergence amongst northern populations probably confounds the AMOVA results, so that it is artificial to consider them as a single cluster of populations for comparison with southern groves. Genetic diversity was decreased with latitude whether measured as allelic diversity, or as the equilibrium expected heterozygosity. We were able to show that this was not a function of decreasing grove size in the north, but was a true latitudinal effect. Therefore, northern groves

fall into clusters that are both genetically distinct and also poorer in genetic diversity than southern groves. Conservation of genetic diversity of this species will require a grove-by-grove approach in the north, but a less systematic approach in the south.

Is there evidence of deep lineage divergence? As is common with tree species, very sharp divergence among populations was not detected in giant sequoia. However, the proportion of molecular variance associated with populations (~12%) was high compared with many conifer species. In the north, individuals and populations showed very little admixture suggesting migration has been limited locally. This may be less important among some of the southern groves. We conclude that the northern groves have been isolated for a considerable time. Indeed, our preliminary analyses of demographic changes would favour long separation of northern groves from one another and from the south. This could give enough time for adaptive gene complexes to have evolved differently among populations.

Is there any indication that current population size is associated with higher inbreeding? We found no significant correlation between grove size and the inbreeding coefficient. However, the inbreeding coefficient tells to what extent the population departs from equilibrium and is not a true measure of selfing, or of bi-parental inbreeding. We hope to understand this better in the current STRL proposal that was funded in 2014.

References

Beaumont MA (1999). Detecting population expansion and decline using microsatellites. Genetics 153: 2013–2029.

Cullings KW (1992). Design and testing of a plant-specific PCR primer for ecological and evolutionary studies. *Molecular Ecology* 1: 233–240.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4: 359-361.

DeSilva R, Dodd RS (2014). Development and characterization of microsatellite markers for giant sequoia, *Sequoiadendron giganteum* (Cupressaceae) Conservation Genetics Resources 6: 173-174.

Excoffier L, Laval G, Schneider S (2005). Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47–50.

Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. Molecular Ecology 14: 2611–2620.

Falush D, Stephens M, Pritchard JK (2003). Inference of population structure: Extensions to linked loci and correlated allele frequencies. Genetics, 164:1567–1587.

Fins L, Libby WJ (1982). Population variation in Sequoiadendron: Seed and seedling studies, vegetative propagation, and isozyme variation. Silvae Genetica 31:102-110.

Jensen JL, Bohonak AJ, Kelley ST (2005). Isolation by distance, web service. BMC Genetics 6: 13.

Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics 22: 768-770.

Pritchard J K, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics, 155:945–959.

Rosenburg NA (2004). DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 4: 137-138.

Schuelke M (2000). An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology, 233-234.